

**Analysis of Population Structure of *Simulium damnosum* sensu
lato In the Ecological Transition Zone of Central Ghana**

Submitted by

Ernest Tawiah Gyan

BSc. Environmental Science, 2006 University of Cape Coast

MPhil Zoology, 2013 University of Ghana

A thesis submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Life Sciences

College of Health Sciences and Engineering

La Trobe University

Victoria, Australia

October 2019

Table of Contents

Analysis of Population Structure of <i>Simulium damnosum</i> sensu lato In the Ecological Transition Zone of Central Ghana	
Table of Contents	I
Abstract.....	VIII
Statement of Authorship	IX
Dedication	X
Acknowledgement.....	XI
List of Abbreviations	XIII
List of Figures, Tables and Supplementary Sheets.....	XV
 List of Figures.....	XV
 List of figures of chapter 1	XV
 List of figures of chapter 2	XV
 List of figures of chapter 3	XVI
 List of figures of chapter 4	XVI
 List of figures of chapter 5	XVII
 List of Supplementary Figures.....	XIX
 Supplementary Figures A.....	XIX
 Supplementary Figures B	XIX
 Supplementary Figures C.....	XX
 List of Tables	XXI
 List of tables of chapter 2	XXI
 List of tables of chapter 3	XXI
 List of tables of chapter 4	XXII
 List of tables of chapter 5	XXII
 List of Supplementary Table.....	XXIII
 List of Supplementary Sheets	XXV
Chapter 1: General Introduction.....	1
 1.1 Onchocerciasis: a general background	1
 1.2 Onchocerciasis: Life cycle of <i>Onchocerca volvulus</i>	3
 1.3 Onchocerciasis: organization, control and elimination	5
 1.3.1 Onchocerciasis Control Program in West Africa (OCP)	5
 1.3.2 African Program for Onchocerciasis Control (APOC)	9
 1.3.3 Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) .	11
 1.3.4 Onchocerciasis Elimination Programme for the Americas.....	12
 1.3.5 Treatment	14

1.3.6 Ivermectin.....	15
1.3.7 Global geographical distribution of onchocerciasis	17
1.4 Life cycle and biology of <i>Simulium</i> blackflies.....	19
1.5 Distribution of <i>Simulium</i> blackflies	22
1.6 <i>Simulium</i> vector complexes and <i>Onchocerca-simulium</i> complexes.....	27
1.7 Blackfly diversity and evolution: a need to combine morphology, cytology and molecular genetics.....	29
1.8 Transmission Zone.....	36
1.9 Aims, scope and justification of the study.....	38
1.9.1 Justification (Rationale) of the study.....	38
1.9.2 Justification (Rationale) of sampling area	39
1.9.3 Scope of the study.....	40
1.9.4 Aims of the study.....	40
Chapter 2: Amplicon Sequencing.....	41
2.1 Introduction.....	41
2.1.1 Genetic diversity and its utility in population structure determination.....	41
2.1.2 Standard for estimating degree of genetic diversity	44
2.1.3 Purpose of study, specific objectives and mode of assessment.....	45
2.1.4 Hypotheses	46
2.2 Materials and Methods.....	48
2.2.1 Sample collection.....	48
2.2.2 Sample processing.....	49
2.2.3 DNA extraction with Qiagen's DNeasy Blood & Tissue Kit	49
2.2.4 DNA concentration determination with Qubit® Fluorometer	49
2.2.5 Amplification of short amplicons by polymerase chain reaction (PCR).....	50
2.2.6 Gel electrophoresis.....	50
2.2.7 Molecular phylogenetics analysis	51
2.2.8 Meta-analysis.....	52
2.2.9 Ethical clearance	54
2.3 Results	55
2.3.1 Meta-analysis assessment of diversity and divergence in different species	55
2.3.2 Sample and sequence characteristics	67
2.3.3 Polymorphism and diversity assessment of the mitochondrial and nuclear amplicon sequences.....	72
2.3.4 Polymorphism and divergence between populations using the mitochondrial and nuclear amplicon sequences	73
2.3.5 Determination of population structure	74

2.3.6 Genetic and geographic distance assessment of the mitochondrial and nuclear amplicon sequences	75
2.3.7 Gene flow and genetic differentiation of the mitochondrial and nuclear amplicon sequences.....	77
2.3.8 Neutrality test of the mitochondrial and nuclear amplicon sequences	78
2.3.9 Assessment of the degree of genetic diversity	80
2.3.10 Assessment of the relationships among defined DNA sequences and the relatedness of their genetic cluster units	81
2.4. Discussion.....	92
2.4.1 Meta-analysis, polymorphism and genetic diversity.....	92
2.4.2 Assessment of degree of genetic diversity	92
2.4.3 Gene flow, population structure and implications	93
2.4.4 Decision on general aims and hypothesis	94
2.5 Conclusion	95
Chapter 3: Long-Range amplification of <i>Simulium</i> genes.....	96
3.1: Introduction	96
3.1.1 Structure and function of tRNA and associated genes	96
3.1.2 Justification (Rationale) of the study in chapter three	99
3.1.3 Specific objectives, mode of assessment and hypotheses.....	99
3.2: Materials and Methods	102
3.2.1 Sample collection.....	102
3.2.2 DNA extraction with Isolate II Genomic DNA Kit	102
3.2.3 DNA concentration determination with Qubit® Fluorometer	102
3.2.4 Polymerase Chain Reaction	102
3.2.5 Gel Electrophoresis	103
3.2.6 Library Preparation	103
3.2.7 Molecular phylogenetic and statistical analysis	103
3.3: Results from long range amplicon sequencing.....	104
3.3.1 Amplification of targeted tRNA and associated genes.....	104
3.3.2 Polymorphism and diversity of tRNA's and associated genes from Ghana	104
3.3.3 Neutrality tests of tRNA's and associated genes from Ghana	105
3.3.4 Assessment of gene flow and genetic differentiation in Ghana using tRNA and associated genes	106
3.3.5 <i>A priori</i> assessment of genetic relatedness using the tRNA and associated genes	108
3.3.6 <i>A posteriori</i> assessment of genetic relatedness using the tRNA and associated genes	109
3.3.7 Nucleotide substitution model and phylogenetic test using tRNA and associated genes	111
3.4 Discussion.....	115

3.4.1 Genetic diversity of West African <i>Simulium</i> blackflies	115
3.4.2 <i>Simulium</i> population structure in West Africa	118
3.4.3 Implications on Delineation of Transmission Zones	121
3.4.4 Implications for Decisions to Stop Interventions	122
3.4.5 Decision on study objectives and hypothesis	122
3.5 Conclusion	124
Chapter 4: Whole Mitochondrial Genome Sequencing.....	125
4.1 Introduction.....	125
4.1.1 Background	125
4.1.2 Structure and function of mitochondrial genome	126
4.1.3 Benefits and demerits of mitochondrially encoded genes.....	127
4.1.4 Role of next generation sequencing in whole mitochondrial genome sequencing.....	130
4.1.5 Justification (Rationale) of the study in chapter four.....	131
4.1.6 Specific objectives, mode of assessment and hypotheses	132
4.2 Materials and Methods.....	134
4.2.1 Sample Collection.....	134
4.2.2 DNA Extraction.....	134
4.2.3 DNA concentration determination with Qubit® Fluorometer	134
4.2.4 Library Preparation Using Nextera DNA XT and DNA Flex Library Prep Procedures	135
4.2.5 Genomic DNA Tagmentation.....	135
4.2.6 Library Quantification and Quality Assessment	135
4.2.7 Normalization and Library Pooling	135
4.2.8 NextSeq Sequencing.....	136
4.2.9 Sequence Import	136
4.2.10 Trimming and removing duplicate sequences	136
4.2.11 Contamination removal.....	137
4.2.12 Assembly strategy and parameters.....	138
4.2.13 Mitochondrial Genome Validation	141
4.2.14 Mitochondrial Genome Mapping and Gap Filling	141
4.2.15 Extracting Consensus sequence from assembly	141
4.2.16 Sequence prediction, annotation and validation	141
4.2.17 Variants calling	142
4.2.18 Molecular phylogenetic and statistical analysis	143
4.3: Results.....	146
4.3.1 Results of library preparation and sequencing of whole mitochondrial genomes	146

4.3.2 Pre-processing of sequence reads	150
4.3.3 Contig measurement and assessment of nucleotide distribution	151
4.3.4 Characterization of the mitochondrial genome of <i>Simulium</i> blackflies	154
4.3.5 Polymorphism and diversity	155
4.3.6 Test of neutrality	156
4.3.7 Gene Flow and Genetic Differentiation	157
4.3.8 Determination of genetic relatedness	159
4.3.9: Phylogenetic relationship of <i>Simulium</i> blackflies	168
4.4 Discussion.....	173
4.4.1 Genetic diversity of West African <i>Simulium</i> blackflies from matrilineal lineage	173
4.4.2. <i>Simulium</i> population structure in West Africa	176
4.4.3 Implications of the <i>Simulium</i> blackfly genetic diversity, divergence and population structure on Delineation of Transmission Zones.....	179
4.4.4 Implications of the <i>Simulium</i> blackfly genetic diversity, divergence and population structure for Decisions to Stop Interventions	180
4.4.5 Decision on objectives and hypotheses	181
4.5. Conclusion	183
Chapter 5: Whole nuclear genome contribution to <i>Simulium</i> population genetics	184
5.1 Introduction.....	184
5.1.1 Organization and function of nuclear genome	184
5.1.2 Nuclear and cytoplasmic DNA considerations in phylogeographical works	185
5.1.3 Nuclear gene and intragenic recombination.....	186
5.1.4 Nuclear phylogeography using targeted loci to develop species tree.....	186
5.1.5 Justification (Rationale) of the study in chapter five	188
5.1.6 Specific objectives, mode of assessment and hypotheses	188
5.2 Materials and methods	190
5.2.1 Sample collection.....	190
5.2.2 DNA extraction.....	190
5.2.3 DNA concentration determination	190
5.2.4 Library preparation, quantification, normalization and pooling	190
5.2.5 NextSeq sequencing of libraries.....	190
5.2.6 Sequence import.....	190
5.2.7 Trimming and quality filtering	190
5.2.8 Assembly and mapping of reference nuclear genome of African <i>Simulium</i> blackfly	190
5.2.9 Molecular phylogenetic and statistical analysis	192
5.3 Results	193
5.3.1 Concatenated nuclear gene characteristics.....	193

5.3.2 Assessment of whole genome completeness of African <i>Simulium</i> blackflies.....	193
5.3.3 Nuclear mapping and reads coverage	195
5.3.4 African <i>Simulium</i> blackfly genome size determination	197
5.3.5 Polymorphism and diversity in the nuclear genome.....	198
5.3.6 Neutrality test in the nuclear genome.....	199
5.3.7 Assessment of gene flow and genetic differentiation in the nuclear genome	200
5.3.8 Assessment of genetic relatedness in the nuclear genome	202
5.3.9 Evolutionary relationship determination using nuclear genome.....	204
5.4 Discussion.....	208
5.4.1 African <i>Simulium</i> blackfly genome	208
5.4.2 Gene evolution events	209
5.4.3 Genetic diversity of <i>Simulium</i> blackflies from nuclear genome perspective	209
5.4.4 African <i>Simulium</i> blackflies phylogeny.....	212
5.4.5 Inferring <i>Simulium</i> population structure using nuclear genome.....	213
5.4.6 Implications of the <i>Simulium</i> blackfly nuclear genome diversity, divergence and population structure on transmission zone and elimination intervention	214
5.4.7 Decision on study objectives.....	214
5.4.8 Decision on study hypothesis.....	214
5.5 Conclusion	216
Chapter 6: General Discussion	218
6.1 Overview	218
6.2 Summary of chapters.....	218
6.2.1 Chapter 1: General introduction	218
6.2.2 Chapter two: Diversity, divergence and transmission status estimation from short amplicon sequencing of African <i>Simulium</i> blackflies	220
6.2.3 Chapter three: Confirmation of the diversity, divergence and transmission boundaries of <i>Simulium</i> blackflies from the Ghana ecological transition zone	221
6.2.4 Chapter four: Validation of the confirmed genetic diversity, divergence and transmission ranges of <i>Simulium</i> blackflies from Ghana study foci	222
6.2.5 Chapter five: Whole nuclear genome	223
6.3 Further discussion.....	224
6.4 Summary of additions to the body of knowledge	228
6.4.1 Chapter two	228
6.4.2 Chapter three	229
6.4.3 Chapter four	230
6.4.4 Chapter five	230
6.5 Data management and results dissemination	231

6.6 Technological innovations	231
6.7 Information for policy makers and decision implementers of global onchocerciasis control	233
6.8 Future Directions	234
Appendices.....	235
Supplementary Figures A.....	235
Supplementary Figures B.....	249
Supplementary Figures C.....	251
Supplementary Table.....	256
Supplementary Sheets	281
References.....	287

Abstract

Onchocerciasis is a parasitic infection commonly known as “river blindness”. It is the world’s second leading infectious cause of blindness. It is caused by the filarial nematode *Onchocerca volvulus* and transmitted from one person to another through the bites of the infected blackfly insect of the genus *Simulium*. The 2017 Global Burden of the disease was estimated at 20.9 million and 205 million people are at risk of infection in 34 endemic countries in Africa, Latin America and Yemen. Two of the 4 main WHO guides for decision making and implementation of vector control as Alternative Treatment Strategies for elimination of onchocerciasis are to define transmission zones and characterize *Simulium* blackflies. Little has been achieved on this front in most endemic regions of Africa, where 99% of the global disease burden of onchocerciasis occur. To contribute to solving these issues, this study used both short amplicon sequencing and high throughput next generation sequencing strategies to amplify 439 barcoding genomic regions, 91 long range markers, and 82 whole genomes of *Simulium* blackflies from West, Central, and South-Eastern Africa. Population structure analysis have demonstrated significant fly movement and gene flow due to interbreeding of blackflies across the studied river basins. Observation was made of the presence of 3 molecular units and that one of these may be genetically heterogeneous. There was no geographic population structure in the river basins. The West African *Simulium* blackflies formed a monophyletic group to those from Central and South-Eastern Africa; and all the African blackflies also formed a monophyletic group to those from Asia, Europe and the Americas. The central ecological transition zone of Ghana constitutes a single onchocerciasis transmission zone in the dry season, and this finding is concordant with the parasite data from the work of colleagues in La Trobe University.

Statement of Authorship

Except where reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis accepted for the award of any other degree or diploma.

No other person's work has been used without due acknowledgment in the main text of the thesis.

This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.



21/07/2019

Dedication

I dedicate this thesis to all those who have sacrificed significantly to the field of science in general and to the field of onchocerciasis to be specific. I also dedicate this work to my supervisor, Assoc. Prof. Warwick Grant, for his guidance and support; and to my boss, Dr. Mike Y. Osei-Atweneboana for providing a steppingstone for me to enter the field of molecular biology. Lastly but not the least, I dedicate this thesis to my family: Linda Zubaviel for her love, support, care and understanding; George S. B. Gyan for being both my first son and my best friend; my second son Michael Powell Gyan, for teaching me to be patient and persistent in life; my Mother, Georgina Sarkodie for her everlasting care; and to my sisters for being my representative back home in Ghana.

Acknowledgement

Similar to the difficulties of every mother who has been through natural childbirth labour, so has the challenges of my PhD journey been. Nevertheless, the cruel hands of quitting have not succeeded in dragging me down into the abyss of failure as a result of the contribution, guidance and support of several wonderful people, on whose shoulders I stand tall. Thanks be to my supervisor, Assoc. Prof. Warwick Grant, in whom I have found the exceptional quality of bringing out the best in all and sundry privileged to sit before him. I appreciate your guidance, encouragement and great effort in ensuring that I succeed in my PhD endeavour. I would not have made it this far without your holistic support, especially with the regular and additional scholarship extension you provided from TDR, the Special Programme for Research and Training in Tropical Diseases, co-sponsored by UNICEF, UNDP, The World Bank and WHO.

I am appreciative of the contribution of Prof. Mike Y. Osei-Atweneboana, the Director of Water Research Institute (WRI) of the Council for Scientific and Industrial Research (CSIR), for being my boss, father, helper and scientific-godfather in works relating to medical parasitology and molecular epidemiology. Without your benevolent hands, I would not have had samples for my studies, and of course without your pragmatic mind, pursuing both masters and PhD studies might have been nothing short of a mirage to a dreamer seeking to gather chaff in the wind. Thank you very much.

Getting academic tasks done has been better with the help of Dr. Shannon Hedtke, Postdoctoral Fellow in Grant Lab. Your support is nothing short of a supervisor. You are always there for us, even when we give short notice. Your scripts continue to make task accomplishments much faster than would have been. Keep up with the good work.

I am grateful for the timely support of my research progress panel chair, Dr. Brendan Rodoni, for his contribution to my study. You have always put in great effort to offer the best advice and guidance in this challenging journey of mine. I am grateful for the support of my research progress panel member, Professor Emeritus Terry Spithill, for his guidance and continuous inputs. Thank you for your time, for I would have been lost on the academic journey without your contributions.

I want to say thank you to my work colleagues, Dr. Gowtam Chalasani and Katie Crawford for their continuous support in the lab and office. The academic ideas we share will not be forgotten. I thank all members of Grant Lab, both past and present, for all contributions made.

I am grateful to all members of the Medical Parasitology and Molecular Epidemiology Lab – CSIR, especially those who have been in support of my field sampling: Francis Veriegh, Dr. Kwadwo Kyeremeh Frempong, Messrs. Alhassan Ahmed, Anthony Tetteh-Kumah, Edward Jenner Tettevi and Buhari Adamu Hamidu.

This work was supported by a La Trobe University Full Fee Research Scholarship (LTUFFRS) and La Trobe University Postgraduate Research Scholarship (LTUPRS). The supports of La Trobe International (LTI) and Graduate Research School (GRS) is acknowledged. I appreciate the assistance. Thanks to the management, security and reception of AgriBio and the School of Life Sciences of La Trobe University for hosting me and offering all the support needed.

Many thanks to my mother Ms Georgina Sarkodie and my sisters, Ophelia Gyan, Linda Gyan, Amanda Gyan and Maria Adorsu-Mendah. You are always there for me. It is my prayer that God continue to bless you.

My undying gratitude goes to my wife, Mrs. Linda Zubaviel Gyan for all the great things done and the little but very important things that I fail to recognize. Life in both Ghana and Australia would have been impossible without your love, care, support and endless guidance. Many thanks for being there for me in both thick and thin. You have always supported the house, especially in my absence. I am thankful for my two lovely boys, George S. B. Gyan and Michael P. Gyan for helping me get rid of academic stress. You came into our lives just at the best time anyone could ever wish for. You have all been my companion, without which I would not have had the strength to continue studying.

I pray for the continuous blessings of God in every aspects of your lives.

List of Abbreviations

AFRO	WHO Regional Office for Africa
AIC	Akaike information criterion
AICc	Akaike information criterion corrected
AMOVA	Analysis of molecular variance
ANOVA	Analysis of variance
APOC	African Programme for Onchocerciasis Control
ATP	Adenosine triphosphate
BIC	Bayesian information criterion
BlackV	Black Volta River Basin
Black-Volta	Black Volta River Basin
Bp	Base pair
CDC	Centre for Disease Control
CDTI	Community Directed Treatment with Ivermectin
CytB	Cytochrome
DALY	Disability-adjusted life years
DAPC	Discriminant Analysis of Principal Components
DNA	Deoxyribonucleic acid
DNTPs	deoxynucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
ESPEN	Expanded Special Project for Elimination of Neglected Tropical Diseases
gDNA	Genomic DNA
GTR	General-time reversible model
HKY85	Hasegawa-Kishino-Yano model
HPLC	High Performance Liquid Chromatography
iL3	Infective third stage larvae
INDEL	Insertion-deletion
ITS	Internal Transcribed Spacer
ivm	Ivermectin
JC69	Jukes-Cantor model
L1	First stage larvae
L2	Second stage larvae
L4	Fourth stage larvae

MDP	Mectizan Donation Programme
mf	Microfilariae
MTP	Monthly Transmission Potential
NADH	Nicotinamide adenine dinucleotide
ng	Nano gram
NGS	Next-generation sequencing
nM	Nano Molar
NTD	Neglected Tropical Diseases
OCP	Onchocerciasis Control Programme
OEPA	Onchocerciasis Elimination Programme for the Americas
Oncho	Onchocerciasis
<i>O. volvulus</i>	<i>Onchocerca volvulus</i>
PAHO	Pan American Health Organization
PCA	Principle Component Analysis
PCR	Polymerase chain reaction
qPCR	Real time quantitative polymerase chain reaction
RBF	River Blindness Foundation
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SNP	Single nucleotide polymorphism
SSR	Simple Sequence Repeats
TAE	Tris-acetate-EDTA
TDR	The Special Programme for Research and Training in Tropical Diseases
TRA	Trachoma
tRNA	Transfer RNA
UNDP	United Nations Development Programme
UNICEF	United Nations Children's Emergency Fund
UV	UltraViolet
WHO	World Health Organization
µg	Micro gram
µl	Micro litre
µM	Micro molar

List of Figures, Tables and Supplementary Sheets

List of Figures

List of figures of chapter 1

Figure 1.1: Life cycle of *Onchocerca volvulus*

Figure 1.2.1: The 11 countries covered by the OCP

Figure 1.2.2: Onchocerciasis prevalence map of West Africa prior to OCP control in 1974

Figure 1.2.3: Onchocerciasis prevalence of West Africa at the end of OCP control in 2002

Figure 1.2.4: Map showing APOC countries

Figure 1.2.5: Global map of onchocerciasis endemic countries

Figure 1.2.6: Map of 2017 distribution and status of onchocerciasis preventive chemotherapy

Figure 1.3: Life cycle of *Simulium* blackfly

List of figures of chapter 2

Figure 2.1: Map of study communities

Figure 2.3.1: Phylogenetic tree of *Caenorhabditis* sp. using whole mitochondrial genomes

Figure 2.3.2: Phylogenetic tree of *Simulium* sp. from Ghana, Mali and Liberia using IGS

Figure 2.3.3: Haplotype network of *Simulium* sp. from Ghana, Mali and Liberia using IGS

Figure 2.3.4: Pie chart of *Simulium* blackflies sequenced by different sequencing strategies

Figure 2.3.5: Gel images of short amplicons

Figure 2.3.6: *A priori* assessment of genetic relatedness from concatenated CO1, ND4, and 16S-rRNA gene sequences without population assignment

Figure 2.3.7: *A posteriori* assessment of genetic relatedness from concatenated CO1, ND4, and 16S-rRNA gene sequences after population assignment

Figure 2.3.8: *A priori* assessment of genetic relatedness in 90 CO1 gene sequences without population assignment

Figure 2.3.9: *A posteriori* assessment of genetic relatedness in 90 CO1 gene sequences after population assignment

Figure 2.3.10: *A priori* assessment of genetic relatedness in 90 ND4 gene sequences without population assignment

Figure 2.3.11: *A posteriori* assessment of genetic relatedness in 90 ND4 gene sequences after population assignment

Figure 2.3.12: Phylogenetic tree from concatenated CO1, ND4, and 16S rRNA gene sequences after population assignment

List of figures of chapter 3

Figure 3.3.0: Amplified region of *Simulium* mitochondrion from tRNA-Met to CO1

Figure 3.3.1: *A priori* assessment of genetic relatedness from 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences without population assignment

Figure 3.3.2: *A posteriori* assessment of genetic relatedness using 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences after population assignment

Figure 3.3.3: Unrooted phylogenetic tree from 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences

List of figures of chapter 4

Figure 4.2.1: Global map of *Simulium* whole genome sequencing

Figure 4.2.2: Flow diagram of the assembly process for generating whole mitochondrial genomes of African *Simulium* blackflies

Figure 4.3.1: Agarose gels showing NGS library from Africa

Figure 4.3.2: NGS library fragment analysis using the Agilent High Sensitivity D100 Screen Tape Assays

Figure 4.3.3: Quantitative real-time PCR (qPCR) of pooled library

Figure 4.3.4: Sequence read quality assessment and trimming

Figure 4.3.5: Annotated view of whole mitochondrion genome of African *Simulium* blackfly

Figure 4.3.6: Paired-end reads coverage of a mitochondrial genome

Figure 4.3.7: Mitochondrial contig selection

Figure 4.3.8: *A priori* assessment of genetic relatedness from 75 whole mitochondrial genomes in Ghana

Figure 4.3.9: *A priori* assessment of genetic relatedness from 75 mitochondrial genomes without DLOOP in Ghana

Figure 4.3.10: *A priori* assessment of genetic relatedness from 75 DLOOP sequences in Ghana

Figure 4.3.11: *A posteriori* assessment of genetic relatedness from 75 whole mitochondrial genomes without DLOOP

Figure 4.3.12: *A posteriori* assessment of genetic relatedness using 75 complete mitochondrial genomes from Ghana

Figure 4.3.13: Minimum spanning network using 75 whole mitochondrial genomes with DLOOPS

Figure 4.3.14: *A priori* assessment of genetic relatedness of 81 whole mitochondrial genomes of blackflies from Africa

Figure 4.3.15: *A posteriori* assessment of genetic relatedness of 81 whole mitochondrial genomes of blackflies from Africa

Figure 4.3.16: Maximum likelihood tree from 85 whole mitochondrial genomes from within and outside Africa

List of figures of chapter 5

Figure 5.2.1: Flow diagram of the assembly and mapping process for generating reference nuclear genome of African *Simulium* blackflies

Figure 5.3.1: Schematic diagram of a reference nuclear gene sequence

Figure 5.3.2: Quality distribution plot

Figure 5.3.3: *A priori* assessment of genetic relatedness in 75 blackflies using the concatenated nuclear gene sequences

Figure 5.3.4: *A posteriori* assessment of genetic relatedness in 75 blackflies using the nuclear gene reference panel

Figure 5.3.5: Rooted Maximum-Likelihood estimate showing the relationship between 81 *Simulium* sp. concatenated nuclear gene sequences in Africa

List of Supplementary Figures

Supplementary Figures A

Supplementary Figure 2.3.1: *Drosophila* sp. phylogenetic tree

Supplementary Figure 2.3.2: *Entomyzon* sp. phylogenetic tree

Supplementary Figure 2.3.3: *Simulium* sp. (ITS2) phylogenetic tree

Supplementary Figure 2.3.4: *Simulium* sp. (16S) phylogenetic tree

Supplementary Figure 2.3.5: *Simulium* sp. ND4 phylogenetic tree

Supplementary Figure 2.3.6: *Simulium* COI phylogenetic tree

Supplementary Figure 2.3.7: *Oreochromis* sp. haplotype network

Supplementary Figure 2.3.8: *Onchocerca* sp. haplotype network

Supplementary Figure 2.3.9: *Bos* sp. haplotype network

Supplementary Figure 2.3.10: *Chelodina* sp. haplotype network

Supplementary Figure 2.3.11: *Chlamydosaurus kingii* haplotype network

Supplementary Figure 2.3.12A: *Variabilichromis moorii* haplotype network

Supplementary Figure 2.3.12B: *Variabilichromis moorii* DAPC

Supplementary Figure 2.3.13: *Latimeria* sp. haplotype network

Supplementary Figure 2.3.14: *Crocodylus* sp. haplotype network

Supplementary Figure 2.3.15: *Mus* sp. and *Apodemus agrarius* haplotype network

Supplementary Figures B

Supplementary Figure 2.3.16: *A priori* assessment of genetic relatedness from 90 CO1 gene sequences

Supplementary Figure 4.3.1: DLOOP from 75 mitochondria genomes. A= DAPC, and B = Percentage of reassignment to cluster

Supplementary Figure 4.3.2: Scatterplot from 75 whole mitochondrial genomes with DLOOP

Supplementary Figure 4.3.3

Supplementary Figures C

Supplementary Figure C 2.3.12: *A priori* assessment of genetic relatedness in 32 ITS2 gene sequences without population assignment

Supplementary Figure C 2.3.13: *A posteriori* assessment of genetic relatedness in 32 ITS2 gene sequences after population assignment

Supplementary Figure C 2.3.14: *A priori* assessment of genetic relatedness in 32 IGS gene sequences without population assignment

Supplementary Figure C 2.3.15: *A posteriori* assessment of genetic relatedness in 32 IGS gene sequences after population assignment

Supplementary Figure C 2.3.16: *A priori* assessment of genetic relatedness in 90 16S rRNA gene sequences without population assignment

Supplementary Figure C 2.3.17: *A posteriori* assessment of genetic relatedness in 90 16S rRNA gene sequences after population assignment

List of Tables

List of tables of chapter 2

Table 2.3.1: Sequence characteristics of different organisms

Table 2.3.2 A: Polymorphism and diversity within and across different species

Table 2.3.2 B: Polymorphism and diversity within and across different species (continue)

Table 2.3.3: Neutrality Test of the sequences used in intra-species variation estimation

Table 2.3.4: Measures of population structure by analysis of molecular variance (AMOVA)

Table 2.3.5: Percentage identity of organisms across species

Table 2.3.6: sampling locations with coordinates used in this study

Table 2.3.7: Nucleotide sequence, polymorphism and genetic diversity for CO1, ND4 and 16S rRNA Genes

Table 2.3.8: Nucleotide sequence, polymorphism and genetic diversity for ITS2 and IGS

Table 2.3.9: Polymorphism and nucleotide divergence between populations

Table 2.3.10: Measures of population structure by analysis of molecular variance (AMOVA)

Table 2.3.11: Genetic distance between groups mean with geographical distances

Table 2.3.12: Gene flow and genetic differentiation with geographic distances

Table 2.3.13: Tajima's test results for the amplicon sequences

Table 2.3.14: Fu and Li's Test for the amplicon sequences

List of tables of chapter 3

Table 3.3.1: Polymorphism and diversity of *Simulium* sp. in Ghana using tRNA and associated genes

Table 3.3.2: Neutrality Test of *Simulium* sp. in Ghana using tRNA and associated genes

Table 3.3.3: Gene flow and genetic differentiation per river basin in Ghana using tRNA and associated genes

Table 3.3.4: Overall population differentiation test in Pru, Black Volta and Daka river basins using tRNA and associated genes

Table 3.3.5: Nucleotide substitution models test of 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences

List of tables of chapter 4

Table 4.3.1: Tapestation results showing some of the average library fragment sizes

Table 4.3.2: Summary of sequenced reads

Table 4.3.3: List of contamination reference sequences

Table 4.3.4: Contig measurements and nucleotide distribution of assembled sequences

Table 4.3.5: Pairwise percentage identity among blackflies from within and outside Africa

Table 4.3.6: Polymorphism and diversity in the mitochondrial genome of blackflies in Ghana

Table 4.3.7: Neutrality test in the mitochondrial genome of *Simulium* sp. in Ghana

Table 4.3.8: Gene flow and genetic differentiation per river basin in Ghana

Table 4.3.9: Overall population differentiation test in Pru, Black Volta and Daka river basins

Table 4.3.10: Nucleotide substitution models test of 75 whole mitochondrial genomes

List of tables of chapter 5

Table 5.3.0: Statistics of the completeness of some of the *Simulium* genomes sequenced

Table 5.3.1: Sequences used in the concatenated nuclear gene analysis

Table 5.3.2: Polymorphism and diversity in the nuclear genome of *Simulium* sp. in Ghana

Table 5.3.3: Neutrality test in the nuclear genome of *Simulium* sp. in Ghana

Table 5.3.4: Gene flow and genetic differentiation per river basin in Ghana

Table 5.3.5: Overall population differentiation test of the nuclear genome in the Pru, Black Volta and Daka river basins (from AMOVA test)

List of Supplementary Table

Supplementary Table 2.3.1 A: *Bos* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.1 B: *Bos* sp. accession numbers used in the meta-analysis (cont.)

Supplementary Table 2.3.2: *Caenorhabditis* sp. accession numbers for the meta-analysis

Supplementary Table 2.3.3: *Chelodina* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.4: *Chlamydosaurus* sp. accessions used in the meta-analysis

Supplementary Table 2.3.5: *Variabilichromis moorii* accession numbers for meta-analysis

Supplementary Table 2.3.6: *Latimeria* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.7: *Crocodylus* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.8: *Drosophila* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.9: *Entomyzon cyanotis* accession numbers used for meta-analysis

Supplementary Table 2.3.10: *Mus* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.11: *Apodemus* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.12: *Onchocerca* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.13: *Oreochromis* sp. accession numbers used in the meta-analysis

Supplementary Table 2.3.14A: *Simulium* sp. CO1 accession numbers used for meta-analysis

Supplementary Table 2.3.14B: *Simulium* sp. CO1 accession numbers used for meta-analysis

Supplementary Table 2.3.15A: *Simulium* sp. CytB accession numbers used for analysis

Supplementary Table 2.3.15B: *Simulium* sp. CytB accession numbers used for analysis

Supplementary Table 2.3.16: *Simulium* sp. ND4 accession numbers used for meta-analysis

Supplementary Table 2.3.17: *Simulium* sp. 16S accession numbers used for meta-analysis

Supplementary Table 2.3.18: *Simulium* sp. ITS2 accession numbers used for meta-analysis

Supplementary Table 2.3.19: *Simulium* sp. IGS accession numbers used for meta-analysis

Supplementary Table 4.3.2A: Reads summary

Supplementary Table 4.3.2B: Additional reads summary

Supplementary Table 4.3.3: Trimmed reads summary of whole mitochondrial genomes

Supplementary Table 4.3.4 A: Length and weight statistics of *Simulium* whole mitochondrial genome sequences from Africa, Asia and Europe

Supplementary Table 4.3.4 B: One-Sample T Test of AT/GC Ratio of whole mitochondrial genome sequences from Africa, Asia and Europe

Supplementary Table 4.3.5: Melting temperature statistics of *Simulium* whole mitochondrial genome sequences from Africa, Asia and Europe

Supplementary Table 4.3.6: Counts of nucleotides of *Simulium* whole mitochondrial genome sequences from Africa, Asia and Europe

Supplementary Table 4.3.7: Frequencies of nucleotides of *Simulium* whole mitochondrial genome sequences from Africa, Asia and Europe

Supplementary Table 4.3.8 A: *Simulium* accession number used for nuclear genome work

Supplementary Table 4.3.8 B: *Simulium* sp. accession numbers used for nuclear genome work

Supplementary Table B 2.3.10: Correlation of amplicon distance between group mean and geographic distance

Supplementary Table B 2.3.11: Genetic distance within group mean for mitochondrial and nuclear amplicon

Supplementary Table B 2.3.12: Pearson Product-Moment Correlation for the mitochondrial and nuclear amplicon

List of Supplementary Sheets

Supplementary sheet 1: Ethical approval document for the study protocol

Supplementary sheet 2: Independent Samples T-Test of nucleotide diversity for CO1 and ND4

Supplementary Sheet 3: Independent Samples T-Test of nucleotide diversity for CO1 and 16S-rRNA

Supplementary sheet 4: Independent Samples T-Test of nucleotide diversity for CO1 and ITS2

Supplementary sheet 5: Independent Samples T-Test of nucleotide diversity for CO1 and IGS

Supplementary sheet 6: Independent Samples T-Test of nucleotide diversity for CO1 and the Long-Range marker genes

Supplementary sheet 7: Independent Samples T-Test of nucleotide diversity for CO1 and the whole mitochondrial genome

Supplementary sheet 8: Independent Samples T-Test of nucleotide diversity for CO1 and the concatenated nuclear genes

Supplementary sheet 9: One -Sample T-Test of nucleotide diversity using tRNA and associated genes of *Simulium* blackfly from Ghana

Supplementary sheet 10: Sequence statistics for *Mus sp.* and *Apodemus agrarius*

Supplementary sheet 11: Sequence statistics for *Mus sp.* and *Apodemus agrarius*

Chapter 1: General Introduction

1.1 Onchocerciasis: a general background

Onchocerciasis, commonly referred to as “river blindness”, is known to be caused by the parasitic nematode *Onchocerca volvulus* and transmitted to humans through the repeated bites of the infected vector blackfly of the genus *Simulium*. The disease is referred to as river blindness due to the blood-feeding flies that transmit the blinding disease inhabiting fertile land alongside the rivers where they breed. Humans are the natural reservoirs. Some of the symptoms of infection and associated consequences include severe skin itching (pruritus), disfiguring skin lesions (dermatitis), blindness, epilepsy associated with heavy infection, abandonment of fertile river valleys, social ostracism, onchocercomata (subcutaneous nodules), lymphadenopathies, decreased life expectancy among sighted individuals who have high microfilarial load, and considerably excess loss of mortality of the blind (WHO, 1976; Whitworth *et al.*, 1993; Blanks *et al.*, 1998; Boussinesq *et al.*, 2002; Pion *et al.*, 2002; CDC, 2019; WHO, 2019e). Uncomfortable itching alone has been estimated to contribute to 60% of the disease burden, and onchocerciasis as a whole has been demonstrated to decrease life expectancy by up to 15 years (Daumerie and Savioli, 2010; WHO, 2010). Observation was made of increased relationship between mental conditions and onchocerciasis, especially in children. In the 1960s and many decades after that, onchocerciasis was associated with Nakalanga syndrome, which is a hyposexual dwarfism that mostly links with dental caries and mental disturbances. In other parts of Africa where onchocerciasis and epilepsy coexist, and where there is deep-rooted belief in spirits or other superstitious creatures, there is usually an anecdotal link between the two (Kipp *et al.*, 1996; Crump, 2003).

There were increased reports of the occurrence of Nodding syndrome, an invariably fatal paediatric neurologic condition, in parts of Sudan, Uganda and Tanzania. Its aetiology and pathophysiology remained a mystery for long time. Evidence now exist supporting the hypothesis that nodding syndrome is an autoimmune epileptic disorder that is caused by molecular mimicry with antigens of *O. volvulus*, hence patients will require immunomodulatory therapies and the disease itself may be prevented or treated with antiparasitic strategies like the use of ivermectin (Johnson *et al.*, 2017). It normally affects children aged 5-15 years old and makes them show epileptic-like seizures, cognition impairment as well as constant nodding of their heads when given food. Infected children live near fast-flowing rivers and streams and show the presence of *O. volvulus* antibodies. Although *O. volvulus* has been implicated, no direct association has been found between this

parasite and Nodding syndrome. Nodding syndrome is also linked with vitamin B₆ deficiency and malnourishment (Donnelly, 2012).

There have been reports demonstrating that the cost to individual families in parts of Africa can be catastrophic with poor subsistence farmers in countries like Nigeria spending half of their hard-earned annual income to combat blackfly associated diseases (Adeleke *et al.*, 2010). Presently, the over 99% of the infected people are living in 31 countries in Africa, with the remaining proportion of less than 1% living in some foci in Latin America and Yemen. Decades of control and elimination activities have reduced the prevalence, morbidity and mortality, and subsequently improved the disability adjusted life years. Globally, the prevalence of *Onchocerca volvulus* infections was estimated as 20.9 million in 2017 by the Global Burden of Disease Study (WHO, 2019e). The study further showed that 1.15 million of the infected people had vision loss while 14.6 million of them had skin disease. Different strategies for the disease elimination were adopted by the different continents of endemicity. Whereas biannual large-scale treatment using ivermectin was the main strategy for eliminating the disease in the Americas, the endemic regions of Africa adopted community-directed treatment with ivermectin as the principal strategy to eliminate the disease. In the Americas, four formerly endemic countries were verified by WHO as being free of onchocerciasis following decades of successful implementation of elimination activities, and the countries are Colombia, Ecuador, Mexico, and Guatemala. On the global roadmap to being declared free of onchocerciasis, 3 additional countries stopped mass drug administration by the end of 2017 and completed their 3 years post-treatment surveillance in at least one transmission locality. These countries are the Bolivarian Republic of Venezuela in the Americas, and both Uganda and Sudan of Africa. Approximately 1.8 million people are currently found in areas that do not require mass drug administration for onchocerciasis anymore. There is no proven vaccine or medication for the prevention of *O. volvulus* infection. Ivermectin is the main medication used for the control and elimination of the disease in mass drug administration programs (Cruz-Ortiz *et al.*, 2012; Katabarwa *et al.*, 2012; Richards Jr *et al.*, 2015; Rodríguez-Pérez *et al.*, 2015; Zarroug *et al.*, 2016b; Sauerbrey *et al.*, 2018; WHO, 2019e).

Among all international development initiatives, onchocerciasis control investments have provided one of the highest economic rates of return, some 15-20% (Hodgkin *et al.*, 2007). Furthermore, it has been tagged as ‘one of the most triumphant public health campaigns ever waged in the developing world’ (Hodgkin *et al.*, 2007). Perhaps, none of these heralded triumphant achievements could have been possible without much knowledge of the life cycles

of *O. volvulus* and *S. damnosum*, the discovery and donation of ivermectin, and the formation of efficient control programmes, whose relentless efforts fostered unimaginable partnership between wide range of stakeholders (Tsalikis, 1993; Boussinesq, 1997; Benton *et al.*, 2002; Vercruyse and Rew, 2002; Mehlhorn, 2008; Cupp *et al.*, 2011a; CDC, 2019; WHO, 2019b).

1.2 Onchocerciasis: Life cycle of *Onchocerca volvulus*

The life cycle of *Onchocerca volvulus* that causes human onchocerciasis consist of a definitive human host and an intermediate arthropod host (see Figure 1.1). When an adult female blackfly is ready to develop its eggs, it seeks a human host, bites the skin to create a pool of blood, and introduces third-stage filarial larvae onto the skin of the human host to penetrate the bite wound (1). At the period of taking the blood meal, the mandibles of the blackfly cut into the skin with fast scissor-like movements that causes blood to form a pool for the cibarial and pharyngeal pumps to suck the blood within 4-5 minutes of the blood feeding duration. The wounds from the bites become entry point for pathogens being carried by the vector or those found within the environment (Ubachukwu, 2004; Usip *et al.*, 2006). The larvae (2) develop into adult filariae in subcutaneous tissues that normally live in highly vascularised nodules in the subcutaneous connective tissues (3). The nodules are source of human blood for ingestion and digestion by the adult worms (Burnham, 1998), and also contain nutrients that are absorbed directly via the external cuticle of the worms (Howells, 1987; Smith *et al.*, 1988). The adult males and females mostly live in nodules for about nine to 14 years of their life span (Plaisier *et al.*, 1991; WHO, 1995), but some are occasionally able to reach 15 years (Roberts *et al.*, 1967; CDC, 2019). The maximum presumed life span of the adult worm is however 20 years (Richards Jr *et al.*, 2001). Nodules can contain different numbers of both sexes. In some instances, nodules are observed to contain 1-2 adult males and 2-3 adult females (Schulz-Key and Karam, 1986). The sizes of the females range from 33 to 50 cm in length and 270 to 400 µm in diameter. The sizes of the males have a range of 19-42 mm by 130-210 µm. Normally, once a female develops in a nodule, it does not go out. It can produce microfilariae for about 9 years. The sizes of the microfilariae range 220-360 µm by 5-9 µm. Generally, they have an average life span of approximately 2 years, but individual ages can range from 6 months to 3 years (Duke, 1968; Eberhard, 1986). The microfilariae are mostly found in the skin and in the lymphatics of connective tissues (4), but can sometimes be found in sputum, urine, and peripheral blood. Besides the skin, the microfilariae have high affinity for the eyes, where the inflammatory response against dying microfilariae during years of repeated infection leads to the gradual and ultimate blinding

sclerosis and opacification of the anterior eye by local inflammation and of the posterior eye by autoimmune mechanisms (Hall and Pearlman, 1999). Microfilariae enter almost all parts of the eye, except the lens, and cause bleeding, inflammation and many other complications. The visual impairment and complete blindness result from corneal opacities, chorioretinal degeneration, cataract and optic atrophy that has a tendency for bilateral symmetry to the extent of causing both eyes to be damaged (Crump *et al.*, 2012). Widespread skin lesions include the ‘leopard skin’ and ‘elephant skin’. When affected skin becomes dry and wrinkled, it loses its elasticity and causes the infected persons to look older than their age. The skin lesions and ocular complications result from the host immune response to dead parasites. Live microfilariae are not known to induce such responses. Unlike the live microfilariae, living adults are capable of stimulating reticuloendothelial response that aid in the formation of the fibrous nodules (Ridley, 1945).

It has been observed that the skin can contain about 90% of total microfilariae (WHO, 1995; Bradley *et al.*, 2010). When a blackfly takes a blood meal from an infected human (5), it ingests microfilariae (Blacklock, 1926a; Blacklock, 1927; Ham and Gale, 1984). Once inside the vector, the microfilariae migrate from the blackfly’s midgut through the hemocoel to the flight or thoracic muscles (6). The microfilariae develop into the first-stage larvae (7), through the second-stage larvae into the third-stage infective larvae (8). When the third-stage infective larvae succeeds in migrating to the proboscis (9), it can infect a person during another blood meal (1) (Duke *et al.*, 1966; Trpis *et al.*, 2001; CDC, 2019).

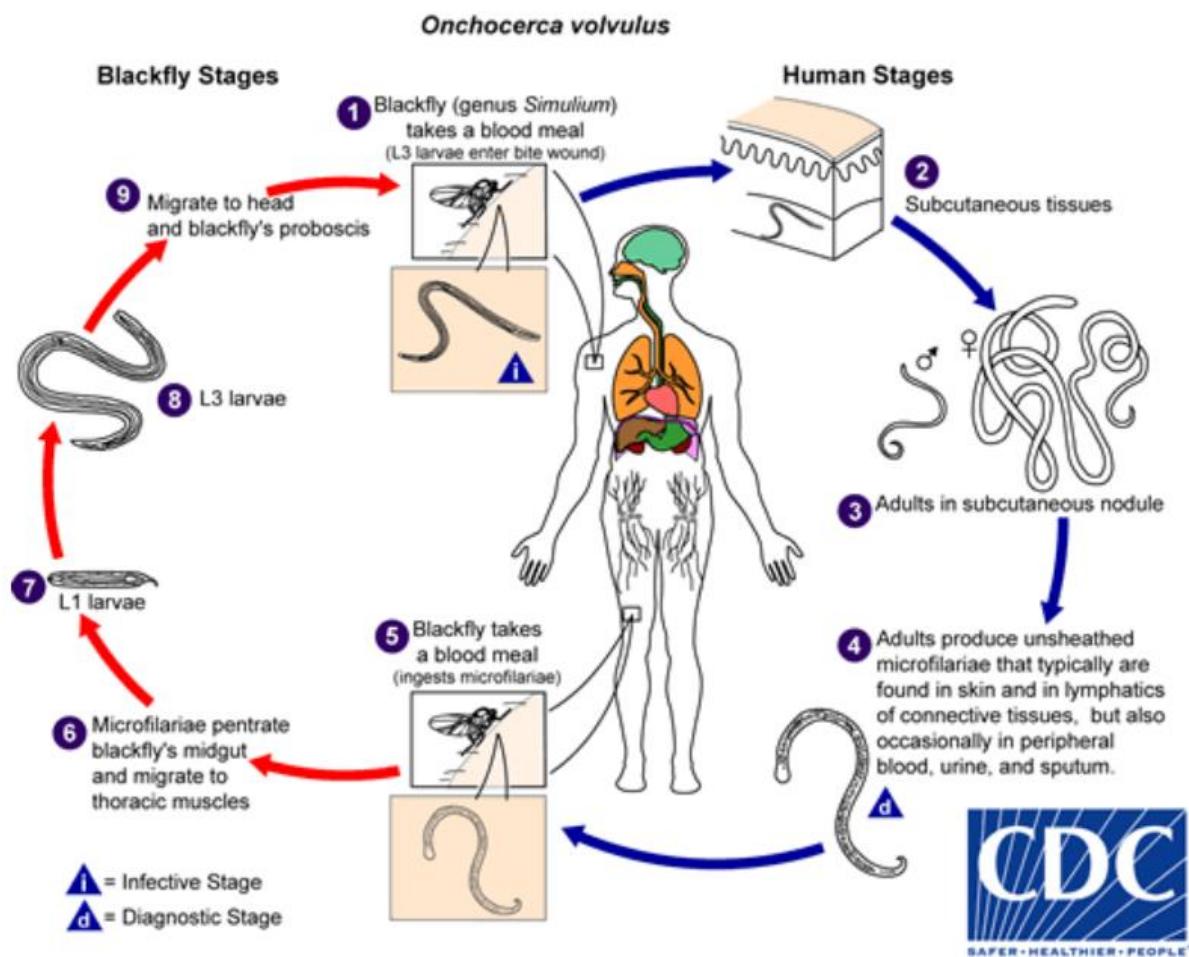


Figure 1.1: Life cycle of *Onchocerca volvulus*. The life cycle of *O. volvulus* (phylum: Nematoda; order: Spirurida; superfamily: Filarioidea; family: Onchocercidae) consist of the definitive vertebrate host stages 2-4 and the intermediate invertebrate host stages 1 and 5-9. The figure was obtained from the Centers for Disease Control and Prevention (CDC, 2019).

1.3 Onchocerciasis: organization, control and elimination

1.3.1 Onchocerciasis Control Program in West Africa (OCP)

By the efforts of four United Nations (UN) sponsoring agencies, there was the launch of the Onchocerciasis Control Programme in West Africa (OCP). These UN sponsoring agencies were the World Health Organization (WHO), United Nations Development Programme (UNDP), World Bank and the Food and Agriculture Organization of the United Nations (FAO). The number of people covered by the OCP was 30 million and they were in 11 countries. The most severe effects from *O. volvulus* infection are generally accepted as blindness and incapacitating skin lesions, which have the potential to affect about one third of adult population living in most highly affected communities. The quest to prevent this

blindness was the main reason for initiating the OCP in 1974 (Tsalikis, 1993; Boussinesq, 1997; Benton *et al.*, 2002).

The principal strategies for the control of onchocerciasis were initially centred on the killing of blackfly larvae. To accomplish this, the control programme used helicopters and aircraft to carry out weekly aerial spraying of insecticide over fast-flowing streams and rivers that constituted the main breeding sites of the blackfly vectors. In accordance with the life span of the worms, the spraying continued for 14 years to break the life cycle of the parasite. Through the strategy of using aerial application of selective insecticides on infested rivers to reduce the blackfly population at its larval stage (Hougard *et al.*, 1993), there was drastic reduction in the transmission of infecting *O. volvulus* larvae and the appearance of both the macrofilariae and microfilariae. The complete interruption of parasite transmission over a period that was above the reproductive life span of the adult worm was estimated to cause the gradual extinction of the parasite in human, thus leading to the potential elimination of the onchocerciasis in the regions under vector control (Hougard *et al.*, 2001).

The second strategy introduced, to augment the spraying of rivers and streams, was the use of ivermectin treatment, following its donation in 1989 by Merck & Co., Inc. (Vercruyse and Rew, 2002; Mehlhorn, 2008; Cupp *et al.*, 2011a).

The 11 countries covered by the OCP were: Benin, Burkina Faso, Côte d'Ivoire, Ghana, Guinea Bissau, Guinea, Mali, Niger, Senegal, Sierra Leone and Togo (see Figure 1.2.1) (WHO, 2019a).



Figure 1.2.1: The 11 countries covered by the OCP. The map was obtained from WHO (WHO, 2019a; WHO, 2019b).

Significant achievements were made at the end of the programme in 2002, a total period of 28 years, and these included at least 4 main core successes: (1) freed 18 million of children from the risk of blindness; (2) completely prevented 600,000 people from going blind; (3) succeeded in reclaiming 250,000 km² of land that was abandoned, and (4) successfully eliminated onchocerciasis as a public health problem in 10 out of the 11 endemic countries under its operation (see Figures 1.2.2 and 1.2.3). The target in the 11th country, Sierra Leone, could not be realized as a result of years of armed conflict, and prevalence of the disease still remained either equal to or above 60% in this country (see figure 1.2.3) (WHO, 2019c; WHO, 2019d).

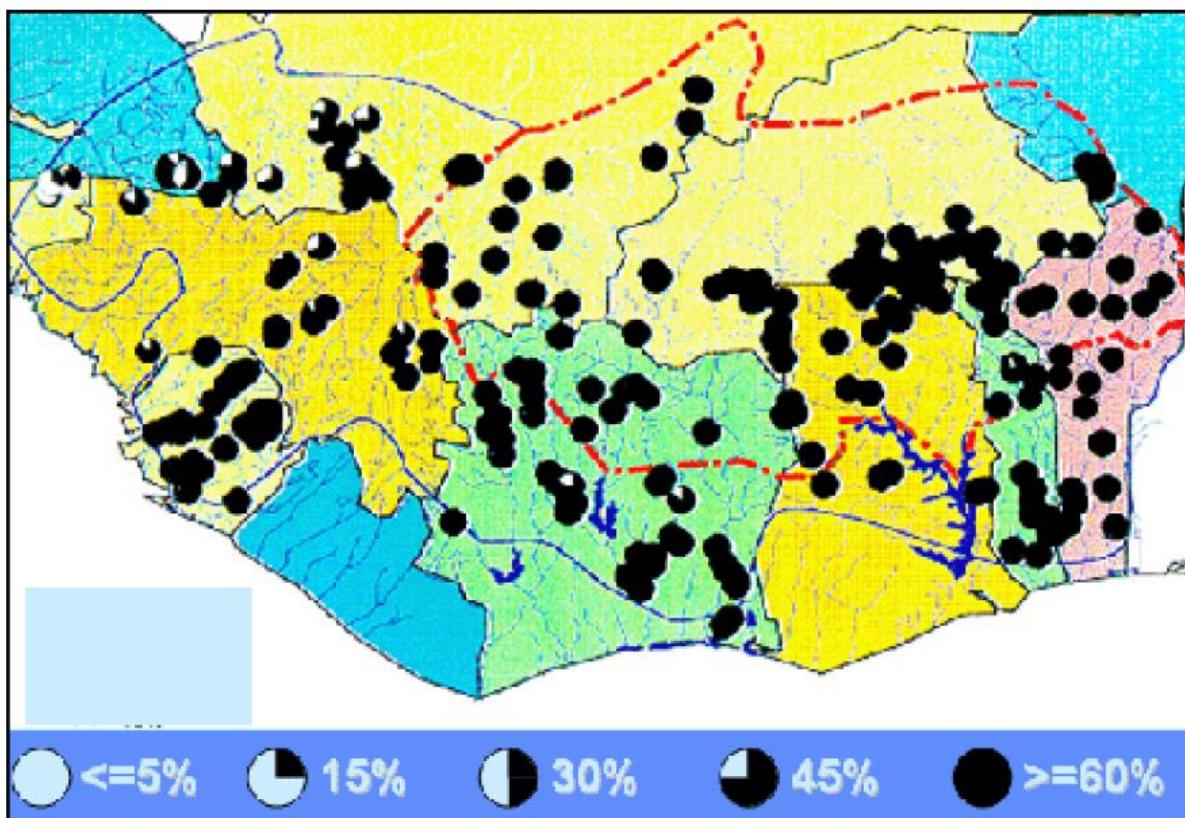


Figure 1.2.2: Onchocerciasis prevalence map of West Africa prior to OCP control in 1974.

The map was accessed from WHO (WHO, 2019d).

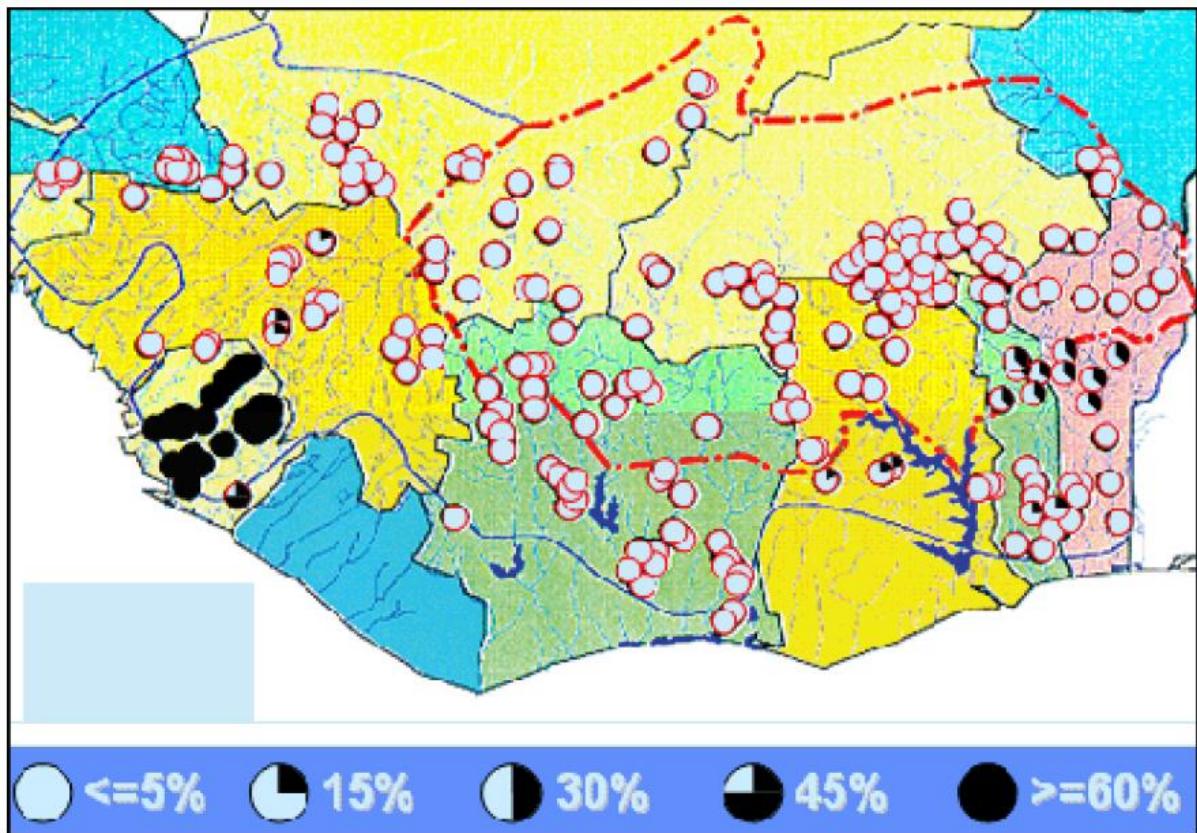


Figure 1.2.3: Onchocerciasis prevalence of West Africa at the end of OCP control in 2002.
The map was accessed from WHO (WHO, 2019d).

1.3.2 African Program for Onchocerciasis Control (APOC)

During the period of 1989 to 1994, the NonGovernmental Development Organizations (NGDO) for the first-time initiated mass distribution of ivermectin, known then as the Ivermectin Distribution Programme (IDP). It was introduced as an adjunct to vector control in the OCP area (Alley *et al.*, 1994). This led to the creation of the NGDO Coordination Group for Onchocerciasis Control in 1991 at the WHO Headquarters. Upon reflecting on the achievements and building on both the knowledge and experience of OCP, the NGDO Group and the sponsoring agencies launched a second programme in 1995 to control river blindness in the rest of endemic countries in Africa. The name of this second programme was the African Programme for Onchocerciasis Control (APOC). The goal of APOC was like that of the OCP, and that was to control onchocerciasis as a public health problem in the remaining endemic African countries. Unlike OCP, there was no viable option of carrying out

widespread aerial spraying in fast-flowing rivers in the countries covered by APOC (WHO, 2015; WHO, 2019c; WHO, 2019b).

APOC was considered a bigger partnership than OCP due to its involvement with 19 participating countries not previously covered by the OCP, their ministries of health, affected communities, many international and local NGDOs, United Nation agencies, donor countries, and the private sector (Merck & Co., Inc.). The World Bank was the fiscal agent of APOC while WHO was the executing agency. A notable feature of APOC was the use of the Community-Directed Treatment with ivermectin (CDTI) as the principal delivery strategy. Through the CDTI, local communities were empowered to combat river blindness in their own localities towards making significant progress in minimizing both suffering and transmission. APOC was further extended to 2015 to treat more than 90 million people per year in the participating 19 endemic countries, thus preventing more than 40,000 incidents of blindness annually and protecting an at risk population of 115 million (WHO, 2019b).

The APOC countries were: Angola, Burundi, Cameroon, Central African Republic, Chad, Congo, Democratic Republic of Congo, Ethiopia, Equatorial Guinea, Gabon, Kenya, Liberia, Malawi, Mozambique, Nigeria, Rwanda, Sudan, Tanzania and Uganda (see Figure 1.2.4) (WHO, 2019b).



Figure 1.2.4: Map showing APOC countries. The map was obtained from WHO (WHO, 2019b).

APOC and OCP were established with the purpose of controlling onchocerciasis as a public health problem rather than to attain elimination of the infection and transmission. Initially, there was great doubt that the disease could be eliminated with the prevailing resources, especially with only ivermectin. However, several studies demonstrated the feasibility of elimination of onchocerciasis in African setting by the use of only ivermectin treatment (Borsboom *et al.*, 2003; Diawara *et al.*, 2009; Tekle *et al.*, 2012; Traore *et al.*, 2012). Africa was then poised to tackle the disease head-on with the sole mission of elimination. During APOC's final year, over 119 million people were treated with ivermectin, and many countries had tremendously minimized the morbidity associated with onchocerciasis. In addition, an estimated number of over 800,000 people in Uganda and 120,000 people in Sudan did not need ivermectin treatment anymore at the period of APOCs closure (WHO, 2019e). APOC closed on December 31, 2015 after commencing the transition to onchocerciasis elimination. The Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) continued with the efforts in the elimination of onchocerciasis in the countries formerly covered by APOC (Tekle *et al.*, 2016; WHO, 2019e).

1.3.3 Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN)

The Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) was set up in May 2016 between WHO Regional Office for Africa (AFRO), Member States and NTD partners for the purpose of bringing together political, financial and technical resources to minimize the burden of the five most prevalent Neglected Tropical Diseases (NTDs) amenable to Preventive Chemotherapy (PC-NTDs) and occurring in Africa by the end of the five-year anticipated lifespan of the project. These five most prevalent NTDs are onchocerciasis (Oncho), lymphatic filariasis (LF), schistosomiasis (SCH), soil-transmitted helminthiasis (STH) and trachoma (TRA). The vital pillar of ESPEN, is country leadership and ownership of their control and elimination programmes and contributing to the overall WHO target of long-term sustainability and health systems strengthening. The four guiding principles of ESPEN are: national leadership and ownership, alignment with national priorities, coordination of stakeholders contributions, and mutual accountability (WHO, 2018). It further has four main objectives: (1) Increase treatments towards the attainment of 100% geographic coverage, (2) Decrease: stopping treatments once transmission has either been interrupted or control achieved, (3) Strengthen information systems for evidence-based

action, and (4) Enhance the efficient use of donated medicines through enhance supply chain management. ESPEN is situated at the WHO Regional Office for Africa (WHO, 2019e).

There are about 1.5 billion people affected by NTDs globally. Approximately 39% of the worldwide NTD burden occurs in Africa. There are still an estimated number of 600 million people requiring treatment in Africa from NTDs (ESPEN, 2019). ESPEN is committed towards making contributions to ensure that national NTD programmes get the financial resources, expertise and data needed to expedite the fight against the targeted NTDs through the coordination of partners and offering technical support. ESPEN is expected to safeguard the achievements made over the past two decades by integrating their successful approaches across all five PC-NTDs (AFRO, 2019).

1.3.4 Onchocerciasis Elimination Programme for the Americas

In the late 1980's, the America's population at risk of onchocerciasis was estimated at 5 million persons. However, a refinement mapping led to approximately 90% reduction of that estimate to an at-risk value of 500,000 living in 13 geographically isolated transmission zones or foci in six countries. The countries were Colombia, Guatemala, Brazil, Venezuela, Ecuador and Mexico (see Figure 1.2.5). The 3 biggest foci in the 3 countries of Mexico, Venezuela and Guatemala contained approximately 60% of the regional at-risk population. By 1990, there was the active distribution of ivermectin Mass-Drug-Administration (MDA) in 3 countries: Ecuador, Mexico and Guatemala. However, the treatment coverage was quite below what was required to interrupt transmission (Collins *et al.*, 1992; Cupp *et al.*, 2004).

The Pan American Health Organization (PAHO) in 1991 identified the donation of ivermectin as an opportunity to engage in regional campaign towards the elimination of onchocerciasis. A PAHO resolution (CD35.R14) was put forth in the same year. The resolution called for countries and partners to put together resources needed to use ivermectin MDAs to eliminate eye diseases that are of onchocerciasis-related, as well as transmission by 2007. Consequently, the Onchocerciasis Elimination Program for the Americas (OEPA) was established in 1993 as the main body achieve the Resolution CD35.R14. The River Blindness Foundation (RBF) provided a grant for the set-up of an OEPA secretariat in Guatemala. When the RBF closed, the Carter Center took over the administration of the OEPA secretariat in 1996 (Blanks *et al.*, 1998). The main approach employed by the OEPA to interrupt transmission is to use ivermectin MDA from twice to four times per year and a treatment

coverage of at least 85% of the eligible at-risk population in the entire endemic regions. The four transmission stages that each of the endemic focus passes through before attaining elimination is active, suppressed, interrupted and eliminated. A country can then request a national WHO verification when all its foci have reached the eliminated transmission phase (Sauerbrey, 2008; PAHO, 2009). The 6 endemic countries in the Americas that were covered by the OEPA, in relation to the remaining countries in the world, are shown in Figure 1.2.5 below.

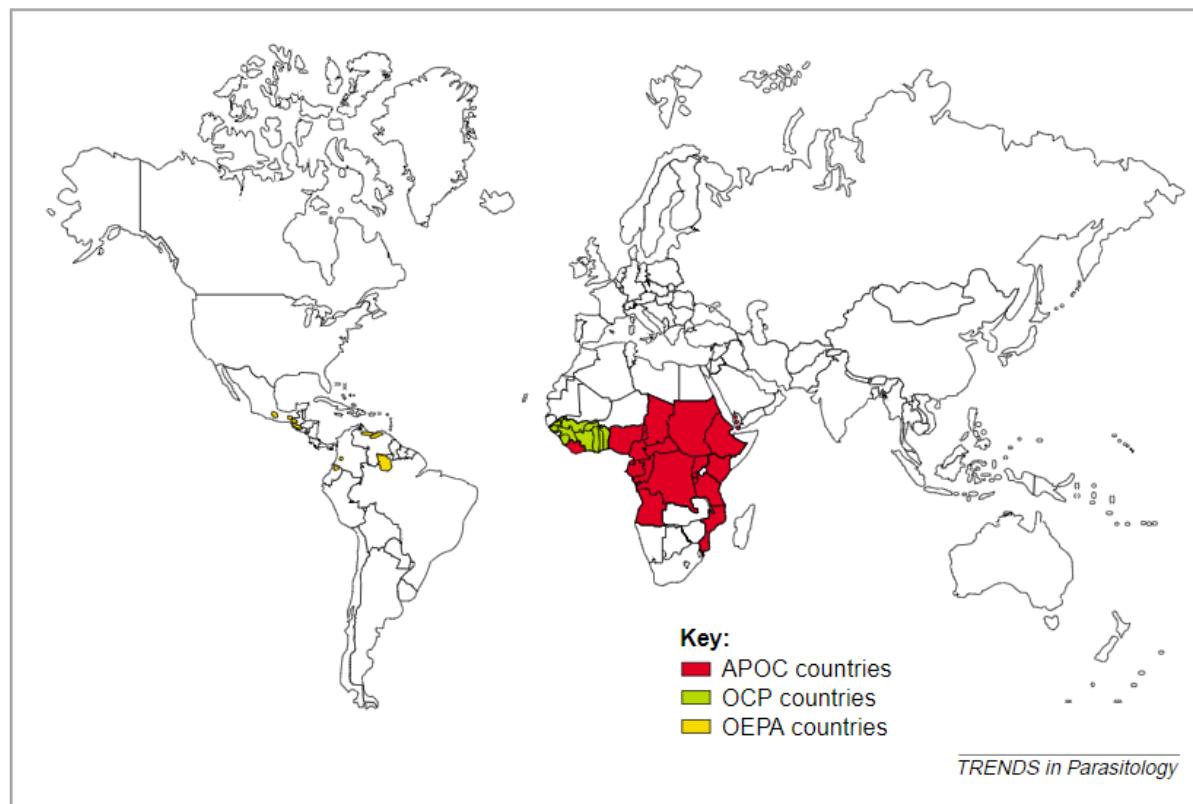


Figure 1.2.5: Global map of onchocerciasis endemic countries. The colour codes show the countries covered by APOC (red), OCP (green) and OEPA (yellow). The map was obtained from published article (Richards Jr *et al.*, 2001).

The PAHO has played major role in the process of verifying transmission interruption and elimination, and in ensuring a political will in each of the resolution it makes. Presently, a resolution has been made calling for the interruption of transmission in all the endemic regions in the Americas by 2022 (PAHO, 2008; PAHO, 2009; PAHO, 2016).

Colombia, with its single transmission focus and bi-annual ivermectin treatment, became the first country in the world to be verified and declared free of onchocerciasis transmission by the WHO in 2013. Its population under treatment was 1366 with 23 rounds of treatment and a treatment coverage reaching over 85% (WER, 2013).

1.3.5 Treatment

It has been recommended by the WHO to use ivermectin to treat onchocerciasis at least once annually for a period of 10-15 years to interrupt transmission. An adjustment in mode of treatment strategy is required for locations where *O. volvulus* co-exists with *Loa loa*. *L. loa* is a form of parasitic filarial nematode known to be endemic in the Cameroon, Congo, Nigeria, South Sudan, Central African Republic, and the Democratic Republic of the Congo. Ivermectin treatment of people with high levels of *L. loa* found in the blood tend to potentially lead to adverse reactions. The recommendation from WHO is to follow the Mectizan Expert Committee (MEC)/APOC recommendation required for the management of severe adverse conditions. Over 145 million people received treatment in Africa through the strategy of the Community Directed Treatment with Ivermectin (CDTI), and it represented over 70% coverage of the total number of people who needed to receive treatment worldwide. This achievement was made by the ESPEN towards its commitments to the attainment of its 4 core objectives (WHO, 2019e).

The donation of all the drugs, by the CEOs of the largest pharmaceutical companies, required to meet the goals needed to control, eliminate and eradicate the 10 NTDs by 2020 was a result of the world's commitment to the London Declaration on Neglected Tropical Diseases (WHO, 2018). A typical example of such donation that has made major contribution towards treatment of major global diseases is the wonder drug, ivermectin (Geary, 2005). Ecuador, with its population of 25,863 at risk in the 119 communities within the single transmission focus, got WHO verification of onchocerciasis transmission elimination in 2014, and became the second country in the world to be declared free of onchocerciasis (CDC, 2013; WHO, 2014). Unlike Colombia and Ecuador that had only 1 onchocerciasis transmission focus each, Mexico had three endemic foci and the second largest population at risk of infection in the Americas: an estimated value of 169,869 individuals living in 670 communities. Mexico finally received a WHO verification in 2015 of having eliminated onchocerciasis (Rodríguez-Pérez *et al.*, 2015; WER, 2015). Guatemala, with its largest at risk population in the Americas: 231,467 persons living in 518 communities, received WHO verification in 2016 of having eliminated onchocerciasis, and was also documented as having been the first country

in the world to use the revised WHO guidelines issued in that same year of being declared free from onchocerciasis transmission (WER, 2016). Venezuela and Brazil are the two countries left in the Americas yet to achieve onchocerciasis elimination. Venezuela has been successful at interrupting transmission in two of its three foci. The only focus left initially had the highest baseline rate of microfilariae in the skin for the whole of the Americas, a baseline prevalence of 72% in 1998. This South focus is contiguous with the Brazil's Amazonas focus with the Yanomami people living in that area. The single focus in Brazil had the second highest baseline microfilaria prevalence in the skin for the whole of the Americas, a value of 63% in 1995. The estimated number of 30,561 person at risk in the Amazonas border between north Brazil and south Venezuela constitute the Yanomami people. The foci still have ongoing transmission and the ongoing transmission is partly due to the people being largely migratory and hard to locate. The PAHO resolution expect the Yanomami foci in Brazil and Venezuela to also get WHO verification of onchocerciasis elimination by 2022. Onchocerciasis elimination in parts of the Americas have been inspiration to Africa, and the whole continent in the Americas is most likely to be declared free of onchocerciasis transmission once Brazil and Venezuela get verified by WHO (Amazigo, 2008; CDC, 2013; PAHO, 2016; Sauerbrey *et al.*, 2018).

1.3.6 Ivermectin

Ivermectin, registered for human use as Mectizan®, is a very safe and effective drug used for the mass treatment of many types of parasite infections (Omura, 2008; Cupp *et al.*, 2011a). These include onchocerciasis, scabies, head lice, trichuriasis, lymphatic filariasis and strongyloidiasis (Ottesen and Campbell, 1994; Crump and Omura, 2011; Gonzalez *et al.*, 2012; Ōmura and Crump, 2014). In some animals, it is used for the treatment and prevention of heartworm among other diseases (Papich, 2016). The drug is as effective and safe to use as albendazole or alternative antinematode drugs for the treatment of pinworm infection or enterobiasis (Ottesen and Campbell, 1994) Distribution of ivermectin for onchocerciasis and lymphatic filariasis has been made possible by the philanthropic donation of Merck & Co Inc in 1987 (Cupp *et al.*, 2011a), following its discovery in 1975 and acceptable medical usage in 1981 (Vercruyse and Rew, 2002; Mehlhorn, 2008). Global, national and regional programs depend on ivermectin for the control and elimination of *O. volvulus*. These programmes include the current Expanded Special Project for Elimination of Neglected Tropical Diseases

(ESPEN) and the Onchocerciasis Elimination Programme for the Americas (OEPA). Other programmes in the past include the African Programme for Onchocerciasis Control (APOC) and the Onchocerciasis Control Programme (OCP) (Remme, 1995; Ōmura and Crump, 2017; Rebollo *et al.*, 2018).

Merck & Co Inc has donated ivermectin for the past 33 years due to its effectiveness in slowly killing significant numbers of the microfilaria over time and it impairs female fecundity of the adult female *Onchocerca volvulus*. In 2011, it was observed that over 800 million doses had been given to more than 80 million people around that time. Consequently, the disease has been significantly minimized in over 25 countries and led to the interruption of transmission in at least 10 countries to the extent that onchocerciasis has no longer been seen in many formerly endemic countries (Cupp *et al.*, 2011b). The donation of ivermectin is to continue for as long as needed to control onchocerciasis and LF (Sturchio, 2001). Mass drug distribution programs were set up to empower community distributors in endemic countries and it was achieved due to the safety profile of the medicine (Amazigo *et al.*, 2002). Ivermectin-based programs have been successful at improving morbidity, mortality and productivity (Hotez *et al.*, 2007).

Ivermectin and avermectins are members of the macrocyclic lactones derived from the bacterium *Streptomyces avermitilis*. Its effectiveness results from its ability to interfere with the nervous system as well as muscle function by inhibiting neurotransmission. Preliminary work on the mechanism and action of this drug centred on its capability to open gamma-aminobutyric acid (GABA)-gated chloride channels (Campbell, 1985). In both mammals and invertebrates, ivermectin has strong activity at GABA receptors. GABA is the main inhibitory neurotransmitter in nematode somatic neuromuscular system. Further research however suggest possibility of glutamate-gated chloride channels to be the likely physiological targets of ivermectin and similar drugs (Cully *et al.*, 1994; Yates *et al.*, 2003). Thus, ivermectin has been proposed to act by paralysing the somatic and pharyngeal muscles that likely stops filaria worms from moving and feeding until they eventually die of starvation (Geary, 2005). Despite the safety profile of ivermectin, there are some known side effects like dry skin, red eyes and burning skin (Rothova *et al.*, 1989; Nontasut *et al.*, 2000). It is contraindicated for use in people with high burden of loiasis as a result of risk of ivermectin-associated adverse inflammatory issues (Keating *et al.*, 2014). It is also contraindicated for use in breastfeeding women, people with kidney or liver disease, and in children below the age of five or those weighing less than 15 kg (33 pounds) (Dourmishev *et al.*, 2005).

A single dose of the drug provides a quick and durable reduction in body burden of eyeworm (*L. loa*) for individuals with less than 20,000 microfilariae per ml of blood, and presents very low risk of the known ivermectin-associated adverse drug events (Pion *et al.*, 2019). All that is required for the treatment of onchocerciasis is a single dose of the drug to reduce microfilaridermia by 98-99% after about 1-2 months of medication (Basanez *et al.*, 2008). The protection of infected persons is to administer a single oral dose once or twice a year for a period of 10-15 years of the lifespan of the adult worms (Soboslay *et al.*, 1987). There is weak evidence of ivermectin's capability to reduce chorioretinal lesions and prevent vision loss among people infected with *O. volvulus* (Maheu-Giroux and Joseph, 2018).

1.3.7 Global geographical distribution of onchocerciasis

Parts of the information in this section have already been presented in earlier sections because of shared responsibilities, achievements and issues of the individual control/elimination programmes with those of ESPEN. This section is however necessary to present the overall geographical distribution of onchocerciasis at the worldwide level and to integrate the distributions at the individual sectoral levels under the coverage of the already presented individual control/elimination programmes. The worldwide distribution of onchocerciasis consists of all the countries covered by ESPEN, which was formerly covered by OCP and APOC, OEPA and a country in the Eastern Mediterranean Region. Until 2013, the disease was known to be endemic in 36 countries worldwide, which became 37 as a result of the formation of South Sudan on 9th July 2011. The distribution was globally divided into 31 countries in Africa, 6 countries in the Americas and the last one in the Eastern Mediterranean region (Scherr, 2012).

The 31 endemic countries in sub-Saharan Africa prior 2017 were Angola, Benin, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Republic of Congo, Côte d'Ivoire, Democratic Republic of the Congo, Equatorial Guinea, Ethiopia, Gabon, Ghana, Guinea, Guinea-Bissau, Kenya, Liberia, Malawi, Mali, Mozambique, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, South Sudan, Sudan, Togo, Uganda, and United Republic of Tanzania. The distribution of the disease in the endemic countries in the Americas prior to 2013 were Colombia, Ecuador, Mexico, Guatemala, Brazil and Venezuela (Bolivarian Republic of). The remaining country on the global endemicity list is Yemen in the Arabian Peninsula (WHO, 2019e).

A number of countries have achieved elimination of the disease. An announcement was made by WHO on 5th April 2013 confirming Colombia as the first country in the world to achieve

elimination of onchocerciasis. The other 3 countries in the Americas that have also been verified and declared free of onchocerciasis by WHO are Ecuador in September 2014, Mexico in July 2015, and Guatemala in July 2016. Presently, over 500,000 people in the Americas do not require ivermectin treatment anymore. All the 13 foci in the 6 countries in the Americas achieved coverage of over 85% in 2006. Subsequently, transmission got interrupted in 11 of the 13 foci by 2017. The remaining 2 foci in Brazil and Venezuela (Bolivarian Republic of) with ongoing transmission, where the Yanomami people live, became the primary focus of elimination efforts (Cruz-Ortiz *et al.*, 2012; Rodríguez-Pérez *et al.*, 2015; Nicholls *et al.*, 2018; Sauerbrey *et al.*, 2018; WHO, 2019e).

As at the end of 2017, three more countries stopped ivermectin mass drug administration and completed three years of post-treatment surveillance in at least one transmission location. These additional countries are the Bolivarian Republic of Venezuela, Sudan and Uganda (Cupp *et al.*, 2019). If these countries get confirmed by the WHO to be free of onchocerciasis, the only country in the Americas that will be left to complete the disease elimination will be Brazil. The net result of this outstanding achievement of decades combat against onchocerciasis is a total of 1.8 million people who currently live in areas that no more need mass drug administration for onchocerciasis (WHO, 2019e). See Figure 1.2.6 for the global distribution of countries requiring and not requiring preventive chemotherapy with ivermectin, and the status of endemicity.

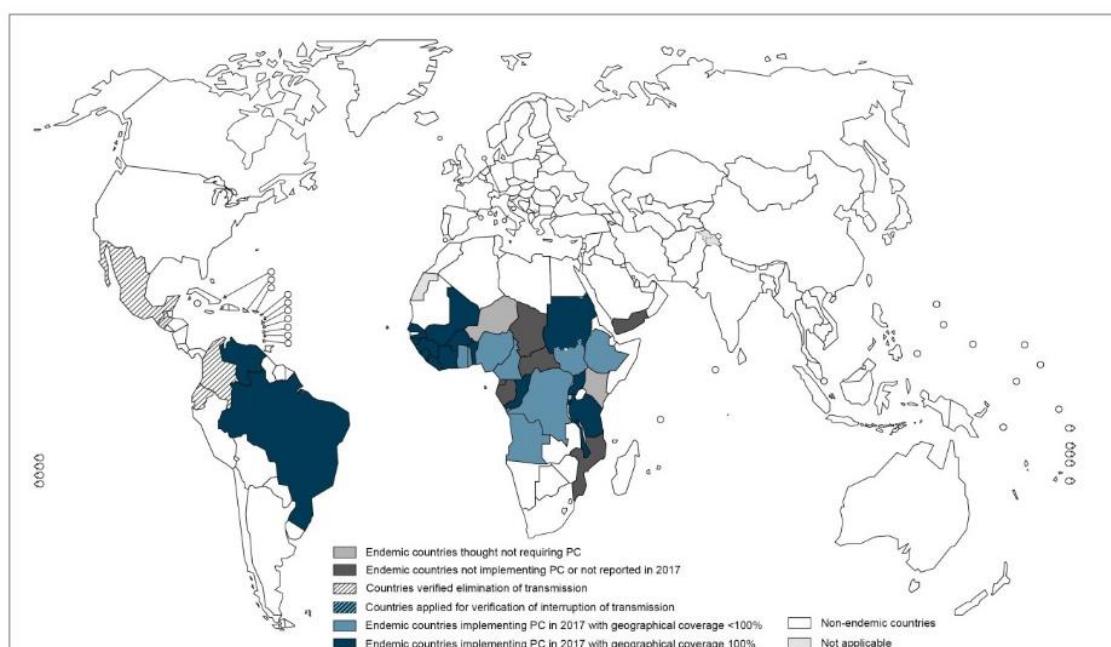


Figure 1.2.6: Map of 2017 distribution and status of onchocerciasis preventive chemotherapy. The source of the map is WHO (WHO, 2019e).

1.4 Life cycle and biology of *Simulium* blackflies

Members belonging to the dipteran genus *Simulium* have great veterinary and medical importance. Belonging to the suborder Nematocera and family Simuliidae and commonly known as blackflies or buffalo gnats, these insect vectors possess long segmented antennae of about 11 segments. All the immature stages are either aquatic or live in very moist habitats. Structurally, the family members possess a reduced mesothoracic segment that makes them look like having a humped back. The majority of the species are of the genera *Cnephia*, *Simulium*, and *Prosimulium* (Lawrence, 2008).

The earliest recognized simuliid fossils were dated to the late Jurassic times, but the fossil record of related families indicates that the blackflies likely originated much earlier than this (Kalugina, 1991; Currie and Grimaldi, 2000). The family Simuliidae most likely had a Pangean origin of approximately 300 million years ago. Blackflies are important organisms in both terrestrial and aquatic ecosystems, especially in the boreal biome of the palearctic and Nearctic regions (Malmqvist *et al.*, 2004). They are holometabolous insects of importance that undergo ecdysis several times in the course of their larval development. *Simulium* blackflies undergo complete metamorphosis characterized by the stages of egg, larva, pupa and adult (see Figure 1.3). The adult male and female feed on nectar and mate. The adult female takes a blood meal from a host to develop an egg mass. It lays a mass of 200 to 500 eggs of 0.2-0.5 mm long on substrates in fast flowing water, or deposit them on the water surface, where they settle into the sediments. The eggs hatch in 4-30 days into sausage-shaped larvae, but eggs of some species diapause. There are 4-9 larval stages with the last stage ranging 5-15 mm long, but most species have 7 larval stages in 1-6 months. The last larval stage usually molts into the non-feeding pupal stage. The 5-15 mm long pupal stage finally transforms into the winged adult in 4-7 days, and the cycle continues (Crosskey, 1990).

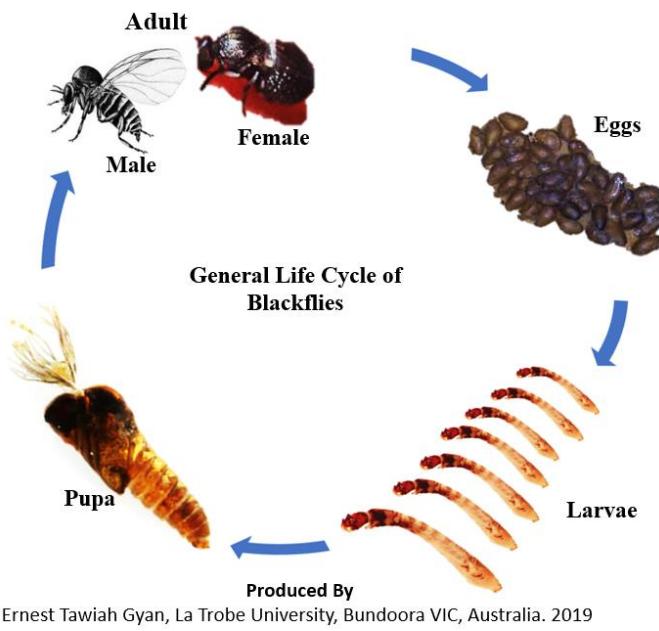


Figure 1.3: Life cycle of *Simulium* blackfly.

Some blackfly species possess mouthparts that are poorly developed and unable to cut flesh, hence become obligatory autogenous species to develop their eggs without blood (Crosskey, 1990). The females of the adult stage of most species are likely vectors of etiological agents, whose feeding activities can lead to diseases in humans. Besides filarial worms, their bites can also transmit other parasitic diseases such as protozoans and arboviruses to different types of domesticated animals (Adler, 2005). When large numbers of adult females bite their host, it can lead to economic loss to agriculture and tourism. It causes death in both wild and domestic mammals and birds (Currie and Adler, 2008). Major attacks by the livestock pests like *S. vampirum*, *S. luggeri* and *S. columbaschense* have led to mortality in pigs, horses, mules, sheep and cattle. Sublethal attacks can lead to reduced milk production, impotence, malnutrition and stress-associated phenomena (Adler *et al.*, 2004) Following a blood meal, adult blackfly females develop a single batch of 200 – 500 eggs. The majority of the species lay eggs either in or on flowing water. Some of the eggs get attached to wet surfaces like blades of aquatic grasses. There exists variation in different species in the duration an egg takes to hatch. The eggs of most species however hatch in about 4-30 days. Some species hatch in as little as couple of months whiles others hatch in as much as many months. Similar to the differences in the duration of egg hatching between species, there exists variation in the number of larval stages. The larvae are important part of the bioenergetic cycle in the lotic systems (Coscaron-Arias & Bramardi, 1996). They form an important food source for lots of invertebrates like plecopterans, and vertebrates like salmonids. Their filter-feeding

mechanisms have a vital role in the processing of organic matter in rivers and streams. Due to the low digestion efficiency of the larvae, their ingested organic matter gets egested as nutritious fecal pellets. These sink to the low parts of streams and rivers where they become food for most collector-gatherer functional feeding guild of aquatic invertebrates (Malmqvist *et al.*, 2001; Currie and Adler, 2008). The larval stages number 4-9 with a predominant observed number of 7. The larval development takes approximately 1-6 months, with both water temperature and food supply as important determinants of larval duration. The last stage larva is the final cycle stage which can pass through winter and attach to rocks, concrete surfaces like those of dams and man-made channels. The first instar larva usually has an egg burster while the seventh instar larvae has separated cervical sclerites. The instars in between the first and the last are normally separated by size (Fredeen, 1976).

The pupal stage is usually formed at the same location as the last stage larva. However, the pupal stage can sometimes be carried downstream due to water current carrying it. In some regions, this usually happen around spring or summer. Within about 4 to 7 days, the adults emerge from the pupal stage. They can survive for weeks before dying of old age. Most species reproduce greatly during the rainy season. Generally, this season normally occurs around March to May in East Africa (Camberlin and Philippon, 2002), May to October in West Africa, and July to September in Northern Africa (Fontaine *et al.*, 1998). Variations exist in the seasonal occurrence and durations of rainfall and temperature. This leads to variation in the number of generations completed in each year from one species to another with respect to geographical location of the species and their ability to adapt to the variations (Lough, 1986; Janowiak, 1988; Cheke *et al.*, 2015; Zarroug *et al.*, 2016a).

The larvae and pupae of blackflies develop in fast flowing water with high concentration of oxygen. Such appropriate habitats for development include fast flowing streams, rivers and waterfalls. The development of populations in such habitats are governed by mixtures of density-dependent and density-independent processes (Fredeen *et al.*, 1951; Davies, 1962; Lawton, 1991; Fonseca and Hart, 1996). Differences in morphology have been noticed from species to species and from one climatic zone to another. But generally, blackfly larvae have a large head with two prominent structures, the labral fans, projecting forward at the anterior end. The main function of the labral fans is to serve as feeding apparatus for filtering organic matter and other invertebrates out of the fast-flowing water currents. At the ends of the bulbous abdomen of blackfly larvae are found glands that produce silken threads. The silken threads have anchorage function by helping attach the larvae to stationary objects like rocks

in flowing water. The size of the mature larvae varies from about 5 to 15 mm in length based on the species being examined. The colours observed are grey, green, brown and greyish black. The larvae develop into pupae. Similar to the larvae, most pupae remain attached to the stationary objects and resist the currents of the fast-flowing water. The colour of the pupae is usually orange. While few species do not produce cocoon, most of the blackfly pupae produce cocoons of different density, size and nature of it being woven. The developing wings and legs become tightly held to the body with the shape of a mummy. Within 4 to 7 days, pupae develop into adults (Lewis, 1957; Crosskey, 1962; Craig, 1974; Barr, 1984; Braimah, 1987; Zettler *et al.*, 1998).

Rainfall and temperature are very important in the determination of the abundance and distribution of blackflies. This is because rainfall increases the amount and current of water bodies, concentration of oxygen in the fast-flowing water, and the abundance of food resource for the larvae. The temperature on the other hand creates conducive conditions for the development of the eggs, larvae, pupae and adults, and the transition from one stage to another. As noted above, there exist 2 clear distinct rainy seasons in East Africa. The heavier rains, known in Kenya-Tanzania (Somalia), as masika (gu) occur around March to May and last longer. The short rains, referred in Kenya-Tanzania (Somalia) as vuli (der) occur around October to November and the rain is usually not heavy (Griffiths, 1972; Hills, 1978; Nicholson and lakes, 1996; Camberlin *et al.*, 1997). In West Africa, rainfall can sometimes be unprecedented in terms of intensity, seasonal expression, duration and spatial character (Farmer and Wigley, 1985). The driest decade of the millennium in West Africa occurred in the 1980s (Nicholson, 1993) and this inevitably had drastic effect on the abundance and distribution of blackflies due to their dependence on fast flowing waters to repopulate. The annual rainfall is mostly constant along each latitude but decreases sharply from the south towards the north. The rainy season in the West Africa occurs in around June to September (Farmer and Wigley, 1985; Eltahir and Gong, 1996).

1.5 Distribution of *Simulium* blackflies

The distribution of blackflies gives an indication of their location, areas of their transmission of *O. volvulus*, and possible consequences to both transmission interruption decisions and ivermectin endpoint timelines. A *Simulium* species that has been very successful at inhabiting a wide range of geographical location is *S. ruficorne*. It is widespread throughout Africa, the

Arabian Peninsula, the Middle East, Madagascar, the southern Iberian Peninsula in southwest corner of Europe, Canary Islands in Spain, and Mascarene Islands in Mauritius, Réunion and Rodrigues (Crosskey, 1969; Crosskey and Büttiker, 1982). Perhaps the successful features that enable widespread adaptability to different geographical locations could be attributed to the species ability to adapt to arid areas in tolerating high water temperatures, very low rates of water flow, as well as much higher pollution levels than a lot of Simuliidae. This broad distribution and ecological tolerance, coupled with the presence of pupal gill are known to form polymorphisms (Crosskey, 1967), and make *S. ruficorne* possibly to be a sibling species complex observable by polytene chromosome studies (Bedo, 1989). Such cosmopolitan distribution could have implications for humans and domestic animals over broad geographical locations.

The smaller species known as *S. ochraceum* is the predominant vector of the parasite in the Americas, including Mexico, Central and South American countries. *S. meridionale*, *S. slossonae*, *S. articum* and *S. vittatum* are predominantly distributed in the United States and Canada. They are usually irritating pests of poultry and cattle, but have been observed to avidly attack humans on warm days following cold weather moments (Underhill, 1944; DeFoliart and Rao, 1965; Lawrence, 2008). *Simulium tuberosum* was first described by Lundstroem from Scandinavia in Europe (Landau, 1962). It is predominantly distributed in North America, from Alaska, North-west Territories, as well as Labrador (Stone, 1952; Shewell, 1957) south to Texas and Florida (Dyar and Shannon, 1927).

In the Australasian region, a number of blackfly species are found. *S. dycei* and *S. mackerrasorum* are widespread and abundant in the localities of north Queensland and the Northern Territory, with *S. mackerrasorum* extending all the way into Western Australia. Restricted distribution is observed in *S. standfasti* from Carnarvon Gorge in Queensland, and *S. lawnhillense* from the areas near both Lawn Hill in North Queensland and Arnhem Land in the Northern Territory (Colbo, 1976). *S. ornatipes* and *S. melatum* are distributed in New South Wales and Victoria (Bedo, 1975).

The genus *Simulium* Latreille s.l. is geographically distributed in the Oriental and Australasian Regions. A study in the region examined the patterns of distribution of this genus at the subgenus and species-group levels. It was observed that the five subgenera of *Byssodon*, *Nevermannia*, *Montisimulium*, *Eusimulium* and *Simulium* s.str were mainly Palaearctic and extending from the north to the Oriental Region. In addition, The six subgenera of *Hebridosimulium*, *Gomphostilbia*, *Inseliellum*, *Wallacellum*, *Morops* and

Himalayum were found to be almost endemic to the Oriental or Australasian Regions or both (Takaoka, 1996).

Simulium blackflies that serve as vectors of animals and humans are complexes of ecologically and evolutionarily distinct cryptic species. Accurate identification of such species is very necessary as a foremost thing to do to get a better understanding of the epidemiology of arthropod-borne diseases and their control (Procunier, 1989; Post *et al.*, 2007; Adler *et al.*, 2010; Post *et al.*, 2013). Cryptic species identification is difficult and even more challenging is the evaluation of species status of allopatric populations. This challenge becomes compounded when populations are predominantly disjunct on mainlands, between mainlands, or between a mainland and an island (Adler *et al.*, 2005). There are implications for assigning the same or different names to populations characterized by allopatric speciation, and this implication affects the understanding of vector potential and the development of control programs (Adler, 2009). A typical illustration of difficulties in determining species status for allopatric populations involves *S. nodosum*. The species is mainly distributed from India all the way across southern China, Thailand, Myanmar and Vietnam (Adler and Crosskey, 2015). It is morphologically similar to *S. shirakii*. *S. shirakii* is distributed in Taiwan, a location that is over 130 km from the Chinese mainland. As a result of the morphological similarity, the species status of *S. shirakii* as a different species from *S. nodosum* has been questioned (Takaoka and SuzuKi, 1984). *S. nodosum* transmits filaria to ruminants, and also feed on both humans and bovids (Datta, 1992; Takaoka *et al.*, 2003; Choochote *et al.*, 2005).

S. aureohirtum Brunetti has been observed in the Palearctic and Oriental regions, and it is believed to be widespread in geography (Adler and Crosskey, 2014). Three synonyms are used to refer to *S. aureohirtum*, and these are *S. geniculare* Shiraki from Taiwan, *S. philippinense* Delfinado from the Philippines, and *S. tuaranense* Smart and Clifford located in the Sabah, Malaysia (Adler and Crosskey, 2014). This species thrives in diverse locations, especially habitats that have been disturbed by humans. These habitats include manmade streams, and agricultural lands (Pramual and Kuvangkadilok, 2009a).

Different blackflies that transmit *O. volvulus* have been identified in Africa and the predominant of these vectors in West Africa are species of the *S. sanctipauli* subcomplex, *S. damnosum* subcomplex, *S. yahense*, and *S. squamosum*. In Ghana, there are many sibling species of blackflies and these include *S. sanctipauli*, *S. leonense*, *S. konkourense* and *S. soubrense* (Boakye *et al.*, 1993a). Their groupings include the *S. damnosum* subcomplex (*S.*

sirbanum, *S. damnosum* S. str; *S. dieguerense*), *S. sanctipauli* subcomplex, *S. squamosum* and *S. yahense*. The predominant among the sibling species in Ghana is *S. sanctipauli*. It has 3 chromosomal variants, or cytoforms, identified and these are the ‘Pra’ cytoform found in the Pra river basin, the ‘Djodji’ cytoform in South-East Ghana and South-West Togo, and the ‘Comoe’ cytoform that is located west in the Tano river (Surtees, 1986; Surtees *et al.*, 1988; Boakye *et al.*, 1999; Kutin *et al.*, 2004). The ‘Djodji’ cytoform may be a difficult to find because they may have been eliminated by onchocerciasis control activities (Fiasorbor *et al.*, 1992). There have been disagreements in the species names partly due to their close physical resemblance. For instance, the majority of the old records of *S. sanctipauli* were reassigned to *S. soubrense* and vice versa during the change in the nomenclature of the *Simulium* vectors after a revision in the field (Post, 1986). Despite this close resemblance, *S. sanctipauli* subcomplex shows divergent host preference, vector competency and adult behaviour. For instance, *S. soubrense* in Liberia is predominantly zoophilic (Garms, 1987) and is thought as being of similar populations of *S. konkourense* in Guinea and Guinea Bissau (Charalambous *et al.*, 1995). Unlike the zoophilic nature of the *S. soubrense* in Liberia, other members of the *S. sanctipauli* subcomplex are greatly anthropophilic and efficient vectors of the *O. volvulus* parasite. For example, great anthropophilic and vector competency characteristics have been observed in the ‘Djodji’ cytoform of *S. sanctipauli* and the ‘Beffa’ cytoform of *S. soubrense* (Garms and Cheke, 1985; Cheke and Denke, 1988). These two cytoforms can be found in Togo. Although *S. leonense* also shows similar increased anthropophilic and vector competency, members with such characteristics are found in Sierra Leone (Chavasse *et al.*, 1995).

A diversity of *S. damnosum* complex members have been identified in Ghana (Post *et al.*, 2013). *S. damnosum* complex has been estimated to consist of at least 60 sibling species and cytoforms (Adler *et al.*, 2010). The members demonstrate variation in their vector competence, anthropophagy, level of anthropophily, ecological characteristics, vectorial capacity for *O. volvulus* and geographical distribution (Basáñez *et al.*, 2009; Lamberton *et al.*, 2014; Adler *et al.*, 2010; Garms and Cheke, 1985; Lamberton *et al.*, 2015). They breed in many rivers such as the Kulpawn and Black Volta rivers in the dry season (Noamesi, 1966). Breeding and flight range of vectors are major factors to be understood when deciding the boundaries of onchocerciasis. Breeding places have been found to last for part or an entire season on rivers like Daka, Kulpawn, Oti and White Volta. A study of the flight ranges of *S. ochraceum*, *S. callidum* and *S. metallicum* in Guatemala using over 86,000 blackflies showed that the flies travelled a maximum distance of 15.5 Km (Dalmat, 1950, 1952). In a mark and

recapture study in the savanna zone of northern Ghana, a total of 3,886 *S. damnosum* adult blackflies were used to determine that the flies could travel a distance of 12.9 Km (Crisp, 1956). Similar study on the White Volta River using over 62,000 newly emerged *S. damnosum* indicated that the flies travelled a distance of 27 Km (Thompson, 1972). Blackflies have been observed to travel 83 Km along rivers from Yagaba to Nakong (river Sissili) in a timeframe of 4 months. Moreover, some blackflies were found to have travelled 182 Km from Yagaba to Nangodi (Red Volta) and lived 4 and a half months (Noamesi, 1966). The *S. quamosum* can travel at least 125 Km (Cheke and Garms, 1983). Members of the savanna cytospecies like *S. damnosum* s. str. and *S. sirbanum* have been observed to migrate distances that are over 400 Km (Garms *et.al.*, 1979). Despite these long flight ranges, *S. yahense* travels only a few km from its natal sites (Quillévéré *et al.*, 1977a; Garms and Walsh, 1987).

In the dry season, there is the displacement of *Simulium* adults on the drying harmattan or North East Trade winds that seasonally blow in December onward in the direction of south west across West Africa, including all sections of the ecological transition zone of Ghana (Gwynne-Jones, 1978) towards the Gulf of Guinea. During the rainy season, mostly occurring from June onward, a reverse effect takes place and it is characterized by the monsoon or South West Trade winds that blow and displace blackflies in the north eastern direction and behind the Intertropical Convergence Zone (Garms *et al.*, 1979; Walsh *et al.*, 1981a; Baker *et al.*, 1990). These phenomenon formed the basis of the western extension of the OCP eastern frontier in the 1980s and contributed to the repopulation of breeding sites in vector-controlled localities (Garms *et al.*, 1979; Walsh *et al.*, 1981a; Baldry *et al.*, 1985; Baker *et al.*, 1990).

The association of *S. neavei* with fresh-water crabs was first described in the onchocerciasis foci of Nyanza Province in Kenya (Someren and McMahon, 1950; McMahon, 1951). The blackfly was predominantly confined to the crab *Potamonautes niloticus* and was only occasionally found associating with other species (McMahon *et al.*, 1958). A similar distribution was observed in Uganda (Barnley and Prentice, 1958).

Unlike West and Central Africa where onchocerciasis is widespread with extensive merging foci into each other (WHO, 1995), there exist relatively few foci in East Africa, perhaps reflecting the more limited distribution of *Simulium* blackflies in the East Africa foci (Mustapha *et al.*, 2005).

In southern Malawi, which is the southern-most onchocerciasis focus in Africa, the predominant blackfly is *S. thyolense*. For instance, a study using a combination of molecular and cytotoxic techniques found *S. thyolense* as the most abundant species in that focus where it represented 91% of the larvae collected. *S. kilbanum* and cytoforms of the ‘Ketaketa’ subcomplex were identified to be the predominant members of the complex observed outside of the focus. All the biting female flies collected were found to be *S. thyolense* and this led to a conclusion that it is the main vector of *O. volvulus* in the focus, with the sampling locations consisting of Thyolo, Mwanza and Mulanje. The remaining relatively low numbers of *Simulium* species found were assumed to be zoophilic. The distribution of *Simulium* species in the ‘Thyolo-Mulanje’ focus was observed to be geographically isolated from those in the Mwanza focus (Mustapha *et al.*, 2005). Over 40 years ago, members of the *S. neavei* were the predominant groups of blackflies distributed over the Thyolo focus and there were only few anthropophilic *S. damnosum* s.l. (Lewis, 1961). As a result of human activities, such as the clearing of forest habitat of the crabs, there has been a disappearance of the *S. neavei* group (Raybould and White, 1979; Walsh *et al.*, 1993). Currently, about 99% of the anthropophilic *Simulium* species found are members of the *S. damnosum* complex (Roberts, 1990). Subsequently, there has been an association between the increase in the human-biting *S. damnosum* s.l. and the rise in the onchocercal skin disease prevalence in Malawi (Burnham, 1991).

1.6 *Simulium* vector complexes and *Onchocerca-simulium* complexes

A complex of ‘cryptic’ or ‘sibling’ species, popularly known as species complex refers to a group of closely related species that are not well distinguishable morphologically despite being reproductively isolated from one another. It requires the use of cytological, ecological or genetic attributes to differentiate between them (White *et al.*, 1978). A classic example of such vector complex is *S. damnosum* s.l. (Vajime and Dunbar, 1975). Cytological studies over decades have repeatedly showed that most nominal blackflies are complexes of sibling species (Kim, 1987; Brockhouse *et al.*, 1993) and polytene chromosome studies have confirmed multiple sibling species in the *Simulium damnosum* complex (Dunbar, 1969; Brockhouse *et al.*, 1993; Hadis *et al.*, 2008). Across members belonging to the Simuliidae, chromosomal speciation inferences are carried out and species resolved by approaches like cytotoxicity. For instance, observations of the prevalence of chromosomal rearrangement have been made possible through cytotoxic studies of population samples and interspecific hybrids (Rothfels, 1989; Livingstone and Rieseberg, 2004). The *S. damnosum*

complex is the largest sibling species complex with at least 55 distinct cytoforms (Post *et al.*, 2007) 9 of which serving as vectors for *O. volvulus* in the West African region with varying vector capacities (Boakye, 1993). These sibling species are the forest-dwelling group that consist of *S. squamosum* Enderlein and *S. yahense* Vajime & Dunbar; a savannah-dwelling group (*S. damnosum* sensu stricto and *S. sirbanum*; Vajime & Dunbar); and a transition -zone characterized by forest-savannah mosaic with the dwelling members of *S. sanctipauli* Vajime Dunbar, *S. leonense* Boakye, Post & Mosha, and *S. soubrense* Vajime & Dunbar (Boakye *et al.*, 1998). DNA sequence analyses and cytotaxonomy have assisted the division of the sibling species into 3 main groups. The cytotaxonomy have clearly demonstrated the occurrence of hybridisation and introgression between the siblings (Boakye and Mosha, 1988; Boakye and Meredith, 1993; Boakye *et al.*, 2000), calling into question the degree of reproductive isolation between them. Vector complexes also exist in Latin America (Shelley, 1991).

The term *Onchocerca-Simulium* complexes was first used in the context of human onchocerciasis to denote the well adapted parasite-vector combinations that normally lead to the development and transmission of local populations of *O. volvulus* populations (Duke *et al.*, 1966). The savannah form of *O. volvulus* develops successfully in savannah species of the vectors such as *S. damnosum* Theobald *sensu lato* (s.l.), but such parasites do not develop well within forest species, and vice versa. Different experimental cross-infection studies were carried out by feeding blackflies with microfilarial carriers of sympatric (the same) and allopatric (distant) regions. The studies of this parasite-vector complexes included *O. volvulus-Simulium* comparisons from within West Africa, between West Africa and northern Venezuela, West Africa and Guatemala, Guatemala and northern Venezuela, and also between northern and Amazonian foci in Venezuela (De Leon and Duke, 1966; Duke *et al.*, 1967; Duke, 1970; Takaoka *et al.*, 1986a; Takaoka *et al.*, 1986b; Basáñez *et al.*, 2000). The results indicated the existence of strong local adaptation between the vectors and parasites in each well-established endemic locality, and there was incompatibility between vectors and parasites originally taken from separate endemic localities. These results are important because they increase our understanding of how onchocerciasis disease can be transmitted outside its known endemic localities (Schiller *et al.*, 1984; Maia-Herzog *et al.*, 1999; Basáñez *et al.*, 2000). For this reason alone, and to manage economic costs and ecological impact of disease elimination activities, it is necessary to identify accurately the sibling species that transmit onchocerciasis in a given location (Crosskey, 1990; Brockhouse *et al.*, 1993) and to monitor their movements. Cryptic taxa differ in terms of geographical distributions, breeding

habitats, vectorial capacity, and some traits of importance to epidemiology, vector control and ecology (Bedo, 1976; Walsh *et al.*, 1981b; Crosskey, 1990; Brockhouse *et al.*, 1993).

1.7 Blackfly diversity and evolution: a need to combine morphology, cytology and molecular genetics

Separation of the effects of current factors from historical factors on the basis of current genetic structure and diversity of species can be achieved by using molecular markers (Templeton, 1998; Walton *et al.*, 2000). Cytotaxonomy has made many contributions to the understanding of *Simulium* blackflies' biodiversity mainly because cytotaxonomy is capable of detecting some levels of diversity that are not identifiable by the use of either morphotaxonomy or currently available molecular genetics (Rothfels, 1979; Adler *et al.*, 2004). Similarly, morpho-taxonomy and molecular genetics are able to detect some forms of diversity that cytotaxonomy is not capable of identifying effectively. For example, recent studies utilizing molecular genetic markers have unveiled hidden diversity among species that were not detectable at the cytological level (Pramual *et al.*, 2010; Pramual and Wongpakam, 2013; Pramual and Adler, 2014). Conversely, there has been reports of high diversity detectable at the cytological level but low diversity at the molecular level (Pramual *et al.*, 2011). Thus, no single method appears capable of assessing *Simulium* diversity under all circumstances and there is the need to integrate approaches and count on the synergies of cytology, morphology and molecular genetics to better understand the biodiversity, evolutionary relationships and ecological distribution of blackflies, (Thaijarern *et al.*, 2014).

Methods for the morphological identification of adult *Simulium* blackflies have been developed and are well documented (Wilson *et al.*, 1993). Similar morphological identification methods have been used to identify *S. yahense* with about 99% specificity (Garms and Zillmann, 1984; Wilson *et al.*, 1993). The methods enabled a 100% accuracy in the separation of the savannah cytospecies, consisting of *S. damnosum* s.s. and *S. sirbanum*, from the *S. sanctipauli* subcomplex (*S. sanctipauli*, *S. konkourense*, *S. leonense* and *S. soubrense*). This was achieved by using either the linear discriminant functions (Wilson *et al.*, 1993) or the thorax:antenna length ratio (Garms *et al.*, 1982). There was a decreased in the specificity with the identification of *S. squamosum*, and the methods resulted in about 5% misidentifications (Wilson *et al.*, 1993). The *S. squamosum* is better identified using isoenzyme electrophoresis but this method has not been adopted widely for field work in Africa due to the inconvenience of requiring either fresh or frozen-preserved samples and specialized equipment. For convenience, DNA-based molecular methods are preferred because of their

suitability for working well in the lab on samples preserved from field work (Wilson *et al.*, 1993).

Mitochondrial barcode sequences have been used as genetic markers for population genetic studies especially for phylogeography. Phylogeography has been used to identify features underlying species genetic structure. These features encompass: (1) contemporary factors like ongoing gene flow and ecological conditions (Pfenninger *et al.*, 2003); (2) historical occurrences such as Pleistocene climatic change (Randi *et al.*, 2004); and (3) their integrated effects (Whorley *et al.*, 2004). Molecular, chromosomal, and morphological characters are used to resolve and support evolutionary relationships among simuliid lineages (Rothfels, 1979; Moulton, 2000, Adler *et al.*, 2004).

Current data suggest that blackflies initially evolved in habitats characterized by mountains and cool climates. For instance, all of the oldest extant simuliid subfamilies and tribes such as *Prosimulium*, *Gymnopais*, *Helodon*, and *Parasimulium* inhabit small and cool parts of northern localities of the globe (Adler *et al.*, 2004). Study suggests that the oldest simuliid fossils dated to the Jurassic era (Currie and Grimaldi, 2000). It is however suggested that the family might have existed longer than that, with a Pangean origin, such that the Simuliini and the Prosimuliini possibly originated in Gondwanaland and Laurasia respectively (Currie and Adler, 2008).

For there to be an efficient vector system, there must at least be repetitive gonotrophic cycles and anautogeny, and these have been observed to have evolved early in an ancient dipteran ancestor (Downes and Colless, 1967). This efficient vector system is presently found in the Simuliidae (Adler *et al.*, 2004). *Simulium* as a genus consist of approximately 80% of all the species in the Simuliidae family. It is also made up of about 94% of nearly 50 of the world's main vectors and pests of importance to human welfare and health (Adler and McCreadie, 2019). Nearly all the vectors of human-disease organisms and over half of the world's species are multivoltine, thus consisting of several broods or generations of the organism in a breeding season. The evolution of multivoltinism in *Simulium* was a notable development for the transition of the blackflies from a mountainous habitat into warmer climatic regions, where humans eventually became concentrated. Annually, blackflies that serve as vectors of onchocerciasis and mansonellosis complete about 15-20 generations in a breeding season and thus increase the chances of their human hosts becoming repeatedly exposed to infection (Crosskey, 1990). A milestone in the evolution of blackfly in becoming recognized as modern-day vector and pest was its ability to colonize large streams and rivers to the extent of

being able to produce large populations to the magnitude of almost a billion flies in each day per kilometer of river or stream (Amrine, 1982). Furthermore, most species adapted to rivers are able to colonize disturbed and altered habitats, including those near hydroelectric dams (Fredeen, 1979). In this instance, the conditions of the water bodies near the dam area mimic those of large and naturally flowing rivers. A number of blackfly species breed and emerge from medium-sized to large rivers in numbers large enough to halt agricultural and other economic developments (Crosskey, 1990; Post *et al.*, 2007; Shelley *et al.*, 2010). Anthropophilic inclinations evolved independently in the subcomplexes of *S. damnosum* complex (Krueger and Hennings, 2006). Some vectors like blackflies have evolved to have preference for attacking large, abundant hosts rather than small ones, and ornithophilic species are less host specific than mammalophilic species (Malmqvist *et al.*, 2004). Probably, a necessary criterion for the evolution of anthropophily is the adaptation to feeding on large mammals, and an increase in host catholicity. In the *S. damnosum* complex, the most effective vectors are those with greater anthropophilic tendencies, while those with zoophilic tendencies are comparatively observed to be of lesser effective vectors (Post *et al.*, 2007).

The phylogenetic relationship existing between and among African *Simulium* blackflies serving as vectors of *O. volvulus* lack sufficient resolution in the Family simuliidae. Relationships within the *S. damnosum* complex have been inferred by the use of cytogenetic evidence (Dunbar and Vajime, 1981; Post *et al.*, 2007). Often, the cytodendrogram of such inferred relationships remains unrooted and is without directionality. Studies generating both molecular and cytogenetic data support the existence of two geographically defined African clades, one eastern and the other western (Krüger *et al.*, 2000). Despite some form of disagreement between molecular and cytological data, there is a robust acceptance of the existence of four western and two eastern *S. damnosum* subcomplexes, and further suggesting that savanna taxa are of more recent origin than forest taxa (Krueger and Hennings, 2006).

Low levels of genetic differentiation are observed by studying isoenzyme variation between closely related species and within species of blackflies (Snyder, 1982; Feraday and Leonhardt, 1989; Scarpassa and Hamada, 2003) and this general feature characterize the Family Simuliidae. A typical example is seen by comparing *Drosophila* with *Simulium*. There is lower variation between sibling species within the *S. damnosum* complex than between closely related species of *Drosophila*. As a result, sibling species of the *S. damnosum* complex show species diagnostic alleles at only 6% of isoenzyme loci (Meredith and Townson, 1981), compared with the relatively higher value of 10% between *D. pseudobscura*

and *D. persimilis* (Ayala and Powell, 1972). There is only 1% of larval proteins observed to be unique to individual species in the *S. damnosum* complex when compared with the relatively larger value of 14% in the *Drosophila virilis* group (Hubby and Throckmorton, 1965).

There is no strong data-driven evidence explaining this low level of genetic differentiation among *Simulium* species. However, a number of plausible suggestions have been put forth with provisional explanations. One of these hypotheses center on genetic introgression as a result of interspecific hybridization (Post, 1984; Boakye and Meredith, 1993). In agreement with this hypothesis is the parallel geographical variation in copy number associated with middle repetitive DNA sequences in species like *S. sanctipauli* and *S. squamosum* (Post and Flook, 1992). Although the transposable element pSO11 was not detected in these two species in Togo, an average copy number of 483 was detected in Ghana, and a relatively similar number, 418, was found in Cote d'Ivoire. A similar trend emanated from the uncharacterized middle repetitive DNA-sequence pSQ1 and this was explained by the genetic introgression between the two species. However, this could also have resulted from the parallel regional natural selection or genetic drift. Molecular phylogenetics lend backing support to the introgression hypothesis. Cytotaxonomic close relation exist between *S. squamosum* and *S. yahense* (Vajime and Dunbar, 1975). Their sister-group relationship has been confirmed by mitochondrial DNA phylogenies (Tang *et al.*, 1995; Krüger *et al.*, 2000). Using the intergenic spacer (IGS) region of the nuclear rDNA to produce phylogeny (Morales-Hojas *et al.*, 2002b) revealed a mix of haplotypes in the two species. This could have resulted from genetic introgression, persistence of ancient polymorphisms, or that they are not separate species when it comes to applying the biological species concept of being capable of mating and producing fertile offspring. Observation was made of a trend of intraspecific variation of ITS rDNA and this pattern was attributed to inter-breeding of sympatric sibling species (Tang *et al.*, 1996a).

Irrespective of the reasons given, it is clear that molecular variation existing between sibling species of *Simulium* is usually low. In addition, there have been challenges in the detection of species-specific molecular variation for the identification of species initially defined by cytotaxonomy. There have been studies focusing on the detection of species-specific isoenzyme variation in species that are closely related like *Prosimulium fuscum* and *P. mixtum* (Snyder and Linton, 1983) and in *S. verecundum* and *S. venustum*. However, there was no clear differentiation of the sibling species in the *S. verecundum/venustum*

supercomplex (Snyder, 1982). Many enzyme systems for species specific identification have been studied among the West African *S. damnosum* complex. Two such enzymes, phosphoglucomutase and trehalase, have shown species-diagnostic variation (fixed isoenzyme variants) and they were able to distinguish two sibling species from each other and also from the rest of the complex (Meredith and Townson, 1981). The species distinguished were *S. yahense* and *S. squamosum*. Another isoenzyme variant was able to distinguish these two species using only phosphoglucomutase (Thomson *et al.*, 1996). However, the fixed isoenzyme variation did not correlate with any particular cytospecies (Mebrahtu *et al.*, 1986).

Species-specific molecular weight variation has been identified in proteins found in salivary glands of blackflies (Brockhouse and Tanguay, 1996). However, more work needs to be done using the *S. damnosum* complex members. Similar to the use of salivary gland proteins, the use of DNA variation has been studied in the identification of sibling species of African blackflies. Some of such studies utilized species-specific variation in copy number of cloned repetitive DNA identified by DNA-DNA blot hybridization (Post and Crampton, 1988). This enabled the identification of variations between morphospecies of blackflies (Jacobs-Lorena *et al.*, 1988). Variation was found in the copy number of middle repetitive DNA sequences among three West African subcomplexes of the *S. damnosum* complex (Post and Flook, 1992). Despite this distinguishing capability, there was no species-specific variation detected, and an attempt to differentiate among members of the subcomplexes led to misidentification in some countries like Togo. Among the reasons identified for such discriminatory failure was the issue of sequences identified as transposable elements (Flook and Post, 1997). In the Nearctic, there have been DNA sequencing studies using 16S mitochondrial rDNA to differentiate among members of the *S. verecundum/venustum* supercomplex (Xiong and Kocher, 1993). However, there has been no concrete species-identification assay built by such DNA sequences as a result of failure of the studied species to demonstrate being monophyletic. The use of PCR has shown variation among morphospecies, but challenges still exist in attempts to differentiate between sibling species. Random Amplified Polymorphic DNA (RAPDs) techniques were used to differentiate between cytospecies in the Nearctic *S. vittatum* complex. However, none of the amplicons were found in all the cytospecies (Duncan *et al.*, 2004). Similar studies using the mitochondrial rDNA of 16S and 12S, ND4, and COII genes also failed to identify species-diagnostic traits (Tang *et al.*, 1996b; Zhu *et al.*, 1998; Pruess *et al.*, 2000). Species-specific variation in the mitochondrial COI and 16S rDNA established the species status of morphologically similar forms of *S. reptans* but

showed limited success with other morphologically similar forms of European species (Day *et al.*, 2008; Ilmonen *et al.*, 2009).

A major issue with studies of blackfly population structure is that there is no clarity if divergence between populations reflects undescribed sibling species or within-species population structure. Populations of *S. ochraceum* in different onchocerciasis foci demonstrate genetic divergence (Rodríguez-Pérez *et al.*, 2006). Although very little is known if this divergence is shown in transmission difference, the different populations have been identified chromosomally as cytotypes. Also, *S. tani* is a complex of species and cytotypes (Pramual *et al.*, 2005; Tangkawanit *et al.*, 2009). A study of North American cytospecies show that one cytospecies arose in the East while another arose in the West (Duncan *et al.*, 2004). Generally, mitochondrial and nuclear barcoding DNA sequences including ND2, CO1, 16S rDNA and IGS rDNA show limited species-diagnostic variation (Agatsuma *et al.*, 1993; Mank *et al.*, 2004). Unlike blackflies that show limited success in the identification of sibling species-specific DNA sequences, there has been better species-specific distinguishing traits in mosquitoes (Scott *et al.*, 1993). Nevertheless, the ITS1 rDNA demonstrated nine different amplicon sizes among 28 blackfly cytoforms in southern and eastern Africa (Krueger, 2006a) and this has enabled the identification of a large proportion of cytoforms across Africa. This system has been successfully used to detect anthropophilic species in some onchocerciasis endemic localities where multiple cytospecies breed in the local rivers (Krueger, 2006b). Cytospecies-specific variation in ITS1 was however limited in West African blackflies (Tang *et al.*, 1996a; Krüger *et al.*, 2000).

A number of reasons exist to account for the limited genetic differentiation and often lack of population structure associated with the use of short amplicon barcodes that are predominantly of mitochondrial source. A plausible reason could be a real panmictic population being studied. A second reason might be a problem with the barcode data, which only sample small proportion of the total mitochondrial genome such that other sections not sampled could be informative. This tends to constitute an ascertainment bias. SNP ascertainment bias refers to the systematic departure of genetic statistics, at the population level, from what is expected theoretically (Lachance and Tishkoff, 2013). It can arise from drawing samples from a nonrandom set of individuals, using part of a genome that does not capture all informative alleles, or by using biased SNP discovery procedures. SNP ascertainment bias will always occur when the entire genome of all individuals in a population are not sequenced due to the fact that a small sample size often leads to the capture

of common alleles than rare alleles (Gravel *et al.*, 2011). It however does not constitute a major issue when it comes to the identification of individuals, paternity analysis, and the assignment of individuals to different populations (Morin *et al.*, 2004; Bradbury *et al.*, 2011). Often, SNP density tends to be higher for whole genome sequencing than sequencing part of a genome (Lachance and Tishkoff, 2013), hence increasing the amount of genomic data increases the chances of identifying more informative sites

Generally, mitochondrial genes tend to be in linkage disequilibrium, hence sequencing more genes will not likely change the inference of the results with the exception of increasing the confidence in the results such as increasing the bootstrap values of phylogenetic trees. Although sequencing different genes from mitochondrial genomes does not generally lead to different inference, for rapidly evolving genomic sites however, the acquisition of private mutations followed by natural selection may lead to better resolution as more genetic data are sequenced in both sequence length and sample size. This is made possible because if populations have separated long enough, a mutation at one section or lineage may not be found in another section or lineage. The control region of mitochondrial genomes has become such a powerful species-specific identification tool partly because it is highly polymorphic (Bronstein *et al.*, 2018). Other benefits of using the control region of the mitochondria include fast evolutionary rate (McMillan and Palumbi, 1997; Meyer, 1993), selective neutrality as non-coding region (except as in Pereira *et al.*, 2008; Rech *et al.*, 2014), polymorphic nature (Ghatak *et al.*, 2016), maternal inheritance, and lack of recombination (Brown *et al.*, 1979; Avise, 1994). Moreover, the mitochondrial control region has been successfully used as a genetic marker in various phylogenetic studies such as the one used in invertebrate taxa (Diniz *et al.*, 2005; Shao *et al.*, 2005; Bronstein *et al.*, 2017; Zhang *et al.*, 2017), reptiles (Jiang *et al.*, 2011), fish (Jamandre *et al.*, 2014, Beltrán-López *et al.*, 2017), birds (Kryukov *et al.*, 2017), mammals (Firestone, 2000; Boyko *et al.*, 2009), and amphibians (Huang and Tu, 2016). Sequencing more genetic data by increasing the number of genes and sample size will lead to the detection, if present, of additional variable sites that could be informative. Although more sequencing data may solve the problem of ascertainment bias, it will not resolve the issue of linkage disequilibrium for the usage of mitochondrial genes. In this instance, the use of nuclear data could prove useful. Phenotypic, morphometric, cytogenetic and electrophoretic literature suggest that there is more variation in *Simulium* than is detected here by barcode sequencing (Vajime and Dunbar, 1975; May *et al.*, 1977; Feraday and Leonhardt, 1989; Wilson *et al.*, 1993; Krüger and Garms, 1999; Adler and Kachvorian, 2001; Pramual *et al.*, 2011), thus the inability for most barcoding genes and allozymes to detect

population structure in *Simulium* blackflies (Feraday and Leonhardt, 1989, Conflitti *et al.*, 2010; Conflitti *et al.*, 2012, Van Lun Low *et al.*, 2014) suggest the need for more sequence data generation from both mitochondrial and nuclear sources.

1.8 Transmission Zone

During a 2009 meeting of onchocerciasis experts, the concept of an onchocerciasis “transmission zone” was introduced and defined as “a geographical area where transmission of *O. volvulus* occurs by locally breeding vectors and which can be regarded as a natural ecological and epidemiological unit for interventions” (WHO-APOC, 2010). The delineation of a transmission zone is the first step recommended by the WHO when considerations are been made to stop onchocerciasis treatment. At the time of the meeting by the experts, the general view was that it was practically difficult to determine with some level of certainty that the vectors found in any given area will be exclusively locally breeding. Therefore, transmission zone was operationally defined as “a river basin, or a major section of a river basin, where onchocerciasis is endemic and where the river is the core of the endemic area, with communities with the highest prevalence of infection generally located close to the river and infection levels falling with increasing distance from the river till they become negligible or reach a neighbouring transmission zone” (WHO-APOC, 2010). There are 3 main classifications of transmission zones and they are zones characterized with: (1) active transmission, (2) suppressed transmission, and (3) interrupted transmission (WHO, 2016).

One of the major challenges faced by the APOC was the identification of the endemic communities where its mass ivermectin-treatment works were to take place in line with its stated objective of targeting the most highly endemic, affected and at-risk populations. To overcome this challenge, APOC initiated the rapid epidemiological mapping of onchocerciasis (REMO), which was a tool for the provision of data on the prevalence and distribution of onchocerciasis. The delineation of onchocerciasis transmission zones of various levels of endemicity, which was a vital step in the planning process for onchocerciasis control, was based on this REMO (Noma *et al.*, 2002). After mapping onchocerciasis in a country, it is expected of the programme implementers to select sentinel villages that are close to the breeding sites of the vectors. A decision to stop ivermectin MDA will depend on the completion of continuous MDA implementation with at least 80% therapeutic coverage in the transmission zone during the treatment phase, and getting an indication of being safe to stop the treatment (WHO, 2016).

Onchocerciasis transmission zones in Africa are believed to be extensive and span large geographical areas (Richards Jr *et al.*, 2015). Unlike Africa, onchocerciasis transmission in the Americas was found in small transmission zones or foci, and before control of the disease commenced, there were 13 transmission zones located in 6 countries in Latin America. Due to limited distribution of competent vectors, transmission zones were small, geographically delimited, and mostly maintained by high biting rates of the blackflies. The endemic countries were Colombia, Ecuador, Guatemala, Mexico, Brazil and Venezuela. The estimated human population at risk of infection that was associated with these transmission zones was approximately 500,000 persons (Sauerbrey, 2008; Gustavsen *et al.*, 2011; Richards Jr *et al.*, 2015). The 13 onchocerciasis transmission zones in the Americas were divided into 4 in Guatemala, 3 in Mexico, 1 in Brazil, 3 in Venezuela, 1 in Ecuador and the last 1 in Colombia (CDC, 2013).

The 2016 onchocerciasis Elimination guideline is centred around locally breeding vector sites (WHO, 2016). Population structure is different. It is the product of a combination of current patterns of transmission, which could be breeding site focal, and also the historical patterns of transmission (i.e. processes that have occurred in the past). If breeding sites are stable for long periods of time, such as tens to hundreds of blackfly generations, then the population structure may also be focal around such a perennial breeding site. However, if breeding sites are ephemeral, and move according to changes in local ecology, the population structure will be on a larger scale that reflects the scale of that breeding site mobility. In the transition zone, perhaps its “transitional” nature i.e. between forest and savannah, means that breeding sites do move around over large geographical distances, with the result that population structure is over the same sort of scale. Furthermore, the movement of breeding sites predisposes blackflies to population admixture. Ideally, blackflies that are able to interbreed will be more genetically similar than those that do not. So, population structure results from restrictions on blackfly mating. To mate, blackflies have to be in contact, and that will become possible only when they move. Thus, population structure is the result of blackfly movement. For a vector, movement allows transmission. This implies that movement for mating, that result in population structure, translates into movement for transmission. Consequently, population structure becomes equivalent to transmission zone. This perennial breeding site hypothesis versus the ephemeral breeding site or admixed population hypothesis need to be tested.

1.9 Aims, scope and justification of the study

1.9.1 Justification (Rationale) of the study

The world has fought tooth and nail using all resources at its disposal right from the time of discovery and realization of the effects of onchocerciasis on the general wellbeing of people, their livestock and on the development of nations. The journey to decreasing the debilitating effects of the disease has been laborious, and the quest to interrupt transmission has been of even greater challenge. A compounding effect to the fight against onchocerciasis is the diverse nature of the culture, interests and goals of the localities affected by the diseases. Yet, in the spirit of oneness, cooperation and partnership that is united on a common front and on a common goal, the world is still standing tall to completely root-out the disease and declare the world free of onchocerciasis transmission.

As former control and elimination programmes end their mandate and handover the vast wealth of experience, knowledge and success to the incoming program, assessments of performance of the combat against onchocerciasis are made and new strategies formulated. Decades of hard work have yielded lots of gains in controlling and eliminating river blindness. The disease has been eliminated in a number of countries, transmission has halted in other areas, and more countries are gearing up towards the elimination of onchocerciasis. The main and only drug used in mass drug administration is ivermectin. Resistance to ivermectin and related drugs remain an increasing problem for parasite control, especially for domestic animals (Prichard, 1994; Fiel *et al.*, 2001; Kaplan, 2002; Wolstenholme *et al.*, 2004; Bartley *et al.*, 2006; Craig *et al.*, 2007; Howell *et al.*, 2008; Molento *et al.*, 2008). Suboptimal response of the causative organism, *Onchocerca volvulus*, to ivermectin has been reported in parts of the world (Eng *et al.*, 2006; Bourguinat *et al.*, 2007; Churcher *et al.*, 2009; Gyan, 2013; Frempong *et al.*, 2016). Attempts have been made to explain and find solution to the suboptimal response by different workers across fields. More researchers are putting in resources in one way or another towards the onchocerciasis fight (Grant, 1994; Grant *et al.*, 2006; Osei-Atweneboana *et al.*, 2007; Osei-Atweneboana *et al.*, 2011; Doyle *et al.*, 2017; Crawford *et al.*, 2019). *Simulium* blackflies need careful study and monitoring to appropriately manage any form of ivermectin resistant parasites that might develop.

The process of drug development takes long time to complete before the drug becomes available for use in humans. Ivermectin must remain effective long enough until either onchocerciasis becomes completely eliminated from the world or another suitable drug of similar efficacy, safety, tolerance and broad-spectrum is found. This requires understanding

the evolutionary relationship of blackflies that transmit the disease and also of their ecological distribution in well-defined transmission zones. This in part will enable researchers and control programs know where to direct more resources to prevent the vectors from transmitting parasites that are showing signs of either sub optimal response or resistance, and also to safeguard the decades of success gained in fighting the disease. For onchocerciasis elimination to be effective, WHO recommends the characterization of the genetic diversity of the vectors that transmit the disease, and to define the transmission zones where they are found. The 2016 guidelines define a transmission zone -the area within which transmission occurs- by “locally breeding” vectors, but the guidelines do not include any suggestions as to how one determines what “locally breeding” really means. When one considers this in population genetic terms, “locally breeding” implies panmixia, which is an area (or zone) within which flies are able to interbreed freely. Consequently, characterizing blackfly genetic diversity and determining their population structure (i.e. the area over which flies are free to interbreed) will enable the delineation of onchocerciasis transmission zone of the central ecological transition zone of Ghana. Very little has been achieved in this area and it is probably due to limited availability of genetic data that does not encompass sufficient informative sites.

1.9.2 Justification (Rationale) of sampling area

The ecological boundary between the forest and savannah climatic locations in the middle belt of Ghana was chosen for the sampling area. This location will be referred to as the central ecological transition zone of Ghana, or simply the “transition zone”. There are good reasons of choosing the transition zone as sampling location. First and most important is that we have very detailed information on the parasite population structure in this transition zone. This is extremely important because it gives us a means to validate an important hypothesis of the study: if the hypothesis that vector population structure translate into parasite transmission zone is correct, then the parasites and vector population structures should be the same. Second, the reason we sampled parasites in this transition zone is because of the occurrence of ivermectin sub-optimal response (SOR). We want to know the risk that SOR may be more widespread. To do this, we need to define transmission zone boundaries. Moreover, because the sampling area is a transition zone, it might be ecologically more interesting than a more uniform forest or savannah ecological zone. One of the observations of evolutionary biology is that interesting things happen in ecologically unstable, or transitional, zones.

1.9.3 Scope of the study

The scope of the study will be restricted to studying the vectors only without any component for directly carrying out observations on the parasite or the human hosts of onchocerciasis. The specific objectives will be stated in each of the experimental chapters that seek to address specific questions. Except where indicated, the phrase ‘African *Simulium* blackflies’ refers to blackflies that vector onchocerciasis in Africa, either in part or in whole.

1.9.4 Aims of the study

This study is generally aimed at characterizing the genetic diversity of *Simulium* blackflies in the central ecological transition zone of Ghana and to define the transmission zones of the places they are found. In line with the WHO’s 2016 elimination guidelines definition of a transmission zone (locally breeding vectors), this study will determine if there is population structure amongst anthropophilic black flies captured by human landing at locations in the Pru, Daka and Black Volta river basins in the ecological transition zone of central Ghana. The secondary aim is to determine if the observed population structure of the blackflies is congruent with, and can explain, the observed population structure of the parasites they transmit. The study will review available literature to understand the genetic diversity expected in different organisms in order to know how comparable the genetic diversity of blackflies in the study area are to them. The study will also determine population structure of *Simulium* blackflies and use it to define transmission zone boundaries. Both amplicon and whole genome sequencing strategies will be explored. To accomplish this, genetic analysis will be carried out using nucleotide sequences of organisms such as the cichlid fishes that have experienced recent rapid speciation and morphological divergence, with the hope of getting a metric or yardstick against which to compare the genetic diversity of *Simulium*. Nucleotide sequences of other organisms of known population structure status will also be used to determine the genetic diversity associated with such known population structure status. These will lead to a better understanding in the characterization of molecular taxonomic unit of *Simulium* blackflies in the context of onchocerciasis transmission zone delineation. The study will utilize both amplicon and whole genome sequencing strategies to define the transmission zone boundaries.

Chapter 2: Amplicon Sequencing

2.1 Introduction

2.1.1 Genetic diversity and its utility in population structure determination

Caenorhabditis elegans reproduces by self-fertile XX hermaphrodites and facultative XO males. The males result from rare X chromosome nondisjunction in meiosis of hermaphrodite germline. They can also result from outcrossing that normally result in about 50% male progenies. Mostly, sperm production occurs before oocyte production in the hermaphrodite forms such that the number of sperms impose limits on self-progeny number (Maupas, 1900; Nigon, 1949). Self-fertilization, if any, can strongly minimize genetic diversity in species due to the fact that selection at one locus can drive to fixation the rest of the genome with it, in the presence of strong linkage disequilibrium across the whole genome (Charlesworth and Wright, 2001). Generally, the known genetic diversity of *C. elegans* is low (Koch *et al.*, 2000; Stewart *et al.*, 2005).

Fossil record shows that crocodiles originated in the late Permian or early Triassic era with those extant constituting a small order within the class Reptilia while those living only include 23 species (Li *et al.*, 2007). Among crocodylians, many molecular data across different studies have been used to examine intergeneric phylogenetic relationships ((Brochu and Densmore, 2001; Gatesy and Amato, 1992; Gatesy *et al.*, 1993, 2003, 2004; Harshman *et al.*, 2003; McAliley *et al.*, 2006; Ray and Densmore, 2002; White and Densmore, 2001; Willis *et al.*, 2007). Although there are controversies in the taxonomic status among the African slender-snouted crocodile (*Mecistops cataphractus*), the false gharial (*Tomistoma schlegelii*), *Gavialis gangeticus* and the dwarf crocodile (*Osteolaemus tetraspis*), there is a wide acceptance in the taxonomic status at the family level and this general acceptance recognizes the three Crocodylia families as the Crocodylidae, Alligatoridae and the Gavialidae (Li *et al.*, 2007). Among the well-studied and well understood crocodiles, to the species level, are the *Crocodylus niloticus* and the Saltwater crocodile, *C. porosus*, native in Australia (Bennett *et al.*, 1985; Fergusson, 2010; Schmitz *et al.*, 2003; Webb *et al.*, 1977). Other groups of crocodiles used for the interspecific genetic analysis included *Crocodylus acutus* (American crocodile), the endangered *Crocodylus intermedius* (Orinoco crocodile), the *Crocodylus moreletii* (fresh water Mexican crocodile), *Crocodylus rhombifer* (Cuban crocodile), *Crocodylus mindorensis* (Philippine crocodile, Mindoro crocodile, *Bukarot*, *Buwaya*, or commonly known as the Philippine freshwater crocodile), *Crocodylus novaeguineae* (New Guinea crocodile), *Crocodylus palustris* (Marsh crocodile, Broad-

snouted crocodile or the Mugger crocodile), the endangered freshwater *Crocodylus siamensis* (Siamese crocodile native to Indonesia, Brunei, East Malaysia, Laos, Cambodia, Myanmar, Thailand and Vietnam), and *Crocodylus johnsoni* (Australian freshwater crocodile).

Tilapia remains the common name for about 70 species of the perch-like fishes of the family Cichlidae, and these are mostly found in the fresh waters of tropical waters, especially in Africa (Trewavas, 1983; Stiassny, 1991). Among these are the substrate spawning Tilapia, and the mouthbrooding genera *Oreochromis* and *Sarotherodon* Tilapia. The poor management of genetic resources of tilapia has led to three main genetic problems globally and these are the loss of pure species through mismanagement of interspecific hybridization, high levels of inbreeding depression, and the contamination of genetically improved strains by introgression from feral species (Eknath *et al.*, 1993; Macaranas *et al.*, 1986; McAndrew, 1993). For the purpose of this study intraspecific genetic analysis in the *Oreochromis* sp. were carried out using *Oreochromis niloticus* from completely different geographical locations on two separate continents, and the locations were in the Democratic Republic of the Congo (DRC) and Myanmar. On the other hand, interspecific genetic analysis in the *Oreochromis* sp. was carried out using the *O. niloticus* (from Myanmar and the DRC), *O. leucostictus* from the DRC, *O. mossambicus* from Australia, and *O. aureus* from Australia (Decru *et al.*, 2016; Loh *et al.*, 2014; Kano *et al.*, 2016).

The source of the nucleotide sequences for the *Simulium* sp. used in the interspecific meta-analysis was selected from studies that sampled from Ghana, Mali and Liberia, and the study utilized the non-transcribed intergenic spacer of nuclear ribosomal DNA (Morales-Hojas *et al.*, 2002). The original study from which the nucleotide sequences were obtained from characterized the blackflies as *Simulium sirbanum* (Mali), *S. soubrense* (Liberia) and *S. squamosum* (Ghana). The remaining genetic materials used in the cross-country meta-analysis were from the CO1, CytB, 16S, ITS2 rDNA and ND4 of the blackflies.

The interspecific genetic comparison in the *Onchocerca* sp. was performed using *O. volvulus* from Brazil and *O. ochengi* from Cameroon. The species from Brazil were forest strain while those from Cameroon were Savannah strain (Crainey *et al.*, 2016).

It was generally thought and believed that coelacanths had gone extinct more than 80 million years ago (Mya) until the discovery of a surviving member of the lineage, *Latimeria chalumnae*, in 1938 (Thomson, 1991). From that time onward, more specimens have been caught off the Comoro archipelago close to the eastern cost of Africa in the Indian Ocean

(Forey, 1998). About 10,000 km east of the southwestern Indian Ocean, other coelacanths have also been captured off the coast of Manado located in Indonesia (Erdmann *et al.*, 1998). Studies have shown that the Indonesian coelacanth population are different from the Comoran population (Forey, 1998). The two main groups used for the interspecific genetic analysis among the coelacanths belonged to the *Latimeria chalumnae* and *Latimeria menadoensis* (Nikaido, 2011).

The blue-faced honeyeater, *Entomyzon cyanotis*, are also known as banabird. The birds used for the intraspecific genetic analysis in this study consisted of two populations of the same species selected from two different geographical locations in Australia (Peñalba *et al.*, 2019). The samples were selected from areas away from known or suspected contact zones to ensure that recent hybrids were not included in the sampling (Ford, 1987; Peñalba *et al.*, 2019).

The interspecific genetic analysis among the fruit flies involved 3 different groups of samples originally collected from Nairobi (*Drosophila simulans*), Côte d'Ivoire (*D. yakuba*), and Benin (*D. melanogaster*) (Clary and Wolstenholme, 1985; Wolff *et al.*, 2013).

Chlamydosaurus kingii is an agamid frillneck lizard and a conspicuous part of the arboreal fauna of northern Australia during the period of the wet season. It however becomes very scarce and rarely seen in the extended dry season (Shine and Lambeck, 1989; Christian and Green, 1994). The nucleotide sequences of *Chlamydosaurus* sp. used in this study were originally collected from three different sites in the wet-dry tropics of Australia in about 60 km Southeast of Darwin (Ujvari *et al.*, 2007).

Variabilichromis moorii is a species of freshwater cichlid endemic to Eastern Africa including fauna in the three East African Lakes of Tanganyika, Malawi and Victoria. In terms of ecology, morphology and behaviour, the cichlid fishes of Lake Tanganyika make up the most diverse species flock (Poll, 1986; Snoeks, 2000; Sturmbauer *et al.*, 2008;). Three different cichlid groups originally collected from the three water bodies of Malawi, Victoria and Tanganyika were selected for intraspecific genetic analysis in this study.

The most cytogenetically diverse *Simulium* blackflies in the Oriental Region is *S. tani* Takaoka & Davies, for which 11 cytoforms are known (Tangkawanit *et al.*, 2009; Adler *et al.*, 2013). Molecular phylogeographical study using the COX1 gene showed high genetic diversity and genetic differentiation in *S. tani* collected from different geographical parts of Thailand (Pramual *et al.*, 2005).

2.1.2 Standard for estimating degree of genetic diversity

There is currently no known formal standard for determining with certainty of what should constitute a high or low genetic diversity, especially relating to nucleotide diversity and haplotype diversity.

The use of CO1 sequences have proven useful in genomic studies, especially in the estimation of genetic diversity, in broad category of organisms including Arachnida (Boyer *et al.*, 2007), Chordata (Lara-Ruiz *et al.*, 2008; Rius *et al.*, 2008), crustacea (Raupach *et al.*, 2010; Goodall-Copestake *et al.*, 2010; Santos, 2006; Lejeusne and Chevaldonne, 2006; Pellerito *et al.*, 2009; Palero *et al.*, 2008; Haig *et al.*, 2010), hexapoda (Kim *et al.*, 2009; Jeong *et al.*, 2009), mollusca (Arruda *et al.*, 2009; Bird *et al.*, 2007; Kojima *et al.*, 2008; Cardenas *et al.*, 2009; Strasser and Barber, 2009; Ribeiro *et al.*, 2010), and porifera (Duran and Rutzler, 2006). Haplotype diversity estimation by the use of CO1 genes in this broad category of organisms resulted in values ranging from 0 to 0.98506 with a mean of 0.63388 that was a bit smaller than the median value of 0.70130. Nucleotide diversity on the other hand showed values ranging from 0. to 0.01993 with a mean of 0.00388 that was a bit larger than the median value of 0.00356. These averages for the nucleotide and haplotype diversities have been proposed to serve as standards for determining what should constitute low and high values of diversity (Goodall *et al.*, 2012).

While this proposed standard for genetic diversity might be considered for the CO1 gene, it becomes extremely challenging when attempting to infer degrees of diversity by using other genes and whole mitochondrial genomes. The standard proposed by Goodall and co-workers (2012) for estimating degrees of haplotype diversity and nucleotide diversity will be adopted for the CO1 of this study. An analysis of this is hypothesized to provide a means of inferring levels of high and low diversity in the remaining genes, as well as that of the whole mitochondrial genome.

Presently, there is a research gap on what should constitute an acceptable genetic diversity in *Simulium* blackflies for an inference to be made on what should constitute population and species. To determine the threshold for discriminating between different populations of the same species, and for differentiating one species from another, there needs to be a study to determine a commonality in genetic diversity for intra- and inter-population and species across diverse taxa. For the purpose of this study, meta-analysis using various organisms will

be used to estimate a threshold for inferring what should constitute an acceptable low or high genetic diversity that will correspond to population or species of *Simulium* blackfly.

2.1.3 Purpose of study, specific objectives and mode of assessment

2.1.3.1 Justification (Rationale) of the study in chapter two

In 1878, Manson made the discovery of the part played by insects in the transmission of diseases of humans during his study of the then *Filaria (Wuchereria) bancrofti* (Blacklock, 1927). The genus *Onchocerca* is part of the family Filariidae (Diesing, 1841). Since the confirmatory discovery that revealed blackflies to be the main vectors that transmit the filarial worms responsible for human onchocerciasis, blackflies have progressively gained recognition and become ranked among the top three or four of the world's most important groups of arthropod vectors (Blacklock, 1926a; Blacklock, 1926b; Adler and Crosskey, 2015). Almost 98% of the 2101 nominal species of blackflies are believed to feed on vertebrate blood, including those of mammals and birds. Most of these are likely to transmit blood-borne pathogens (Adler, 2009; Adler and Crosskey, 2015). Of certainty is the transmission of the etiological agent of onchocerciasis, *Onchocerca volvulus*, by *Simulium* blackflies and it was described by Leuckart in 1893. Following one of the initial descriptions of *O. volvulus* by Prout in Sierra Leone in 1901, and of the discovery of onchocerciasis in the Americas in 1915 by Rodolfo Robles Valverde; a lot of time, money, and other resources have been invested towards ensuring the control and elimination of river blindness. As a result, many achievements have been made and among such successes is the elimination of onchocerciasis as a public health problem (Richards Jr *et al.*, 2015; WHO, 2019c; WHO, 2019d). To safeguard such achievements, there is the need to monitor blackflies in predefined zones of their local transmission boundaries. The determination of such transmission boundaries requires characterizing the genetic diversity of the local breeding blackflies and determining their population structure.

Two of the 4 main WHO guides for decision making and implementation of vector control as Alternative Treatment Strategies for elimination of onchocerciasis involve defining transmission zones and characterizing *Simulium* blackflies. Such genetically characterized blackflies in well-defined geographical boundaries of their migration will enable control programmes keep track of the boundaries of onchocerciasis infection transmitted by locally breeding flies, and the location of migrating blackflies and the extent of their migration. This will further provide sufficient information on where to stop ivermectin mass drug treatment

following transmission interruption without risking blackflies travelling from another area of active transmission into the already treated area. It is generally assumed that a transmission zone will consist of a river basin and with transmission decreasing as one moves further away from the river basin. In Ghana and parts of Africa, there are many river basins but there is no information on whether those river basins consist of genetically identical and reproducing blackflies that are different from blackflies of other river basins. This study therefore seeks to characterize the genetic diversity of *Simulium* blackflies in the central ecological transition zone of Ghana and determine the boundaries of their infection transmission (transmission zone), and then compare their diversity with those of blackflies in other different geographical areas. A major research gap is not knowing what would constitute a high or low diversity in blackflies, hence the study will include a meta-analysis by using the nucleotide sequences from wide range of organisms to determine their diversity and subsequently identify how comparable the diversity of blackflies from the transition zone of Ghana are to those of these organisms.

2.1.3.2 Specific objectives

The specific objectives of this chapter are:

1. to use meta-analysis to determine the degree of genetic diversity existing in populations of a species and also between species
2. to use short amplicon barcoding genes to characterize the genetic diversity of *Simulium* blackflies in the transition zone of Ghana
3. to use short amplicon barcoding genes to determine the transmission zone status in the central ecological transition zone of Ghana

2.1.3.3 Mode of assessment

Similar mode of assessment as the one in section 3.1.3.2 of this thesis was used to assess the specific objectives of this chapter

2.1.4 Hypotheses

Null hypothesis 1

Meta-analysis of nucleotide sequences from public database will enable the determination of the degree of genetic diversity existing in populations of species and between species

Alternative hypothesis 1

Meta-analysis of nucleotide sequences from public database will not enable the determination of the degree of genetic diversity existing in populations of species and between species

Null hypothesis 2

Commonly used short amplicon barcoding genes will enable the characterization of genetic diversity of *Simulium* blackflies in Ghana

Alternative hypothesis 2

Commonly used short amplicon barcoding genes will not enable the characterization of genetic diversity of *Simulium* blackflies in Ghana

Null hypothesis 3

Commonly used short amplicon barcoding genes will enable the identification of transmission zone status of *Simulium* blackflies in the central ecological transition zone of Ghana

Alternative hypothesis 3

Commonly used short amplicon barcoding genes will not enable the identification of transmission zone status of *Simulium* blackflies in the central ecological transition zone of Ghana

2.2 Materials and Methods

2.2.1 Sample collection

Simulium blackflies were collected from the transition ecological zone of Ghana by the standard human-landing method (Walsh *et al.*, 1978; WHO, 1987) as part of the routine entomological surveillance work of the Ghana Council for Scientific and Industrial Research (CSIR). They were kept in 70% ethanol and transported to the laboratory. Among the 21 communities involved in the sampling across the river basins that included Pru, Daka and Black Volta (Figure 2.1); 8 were initially selected for genetic analysis.

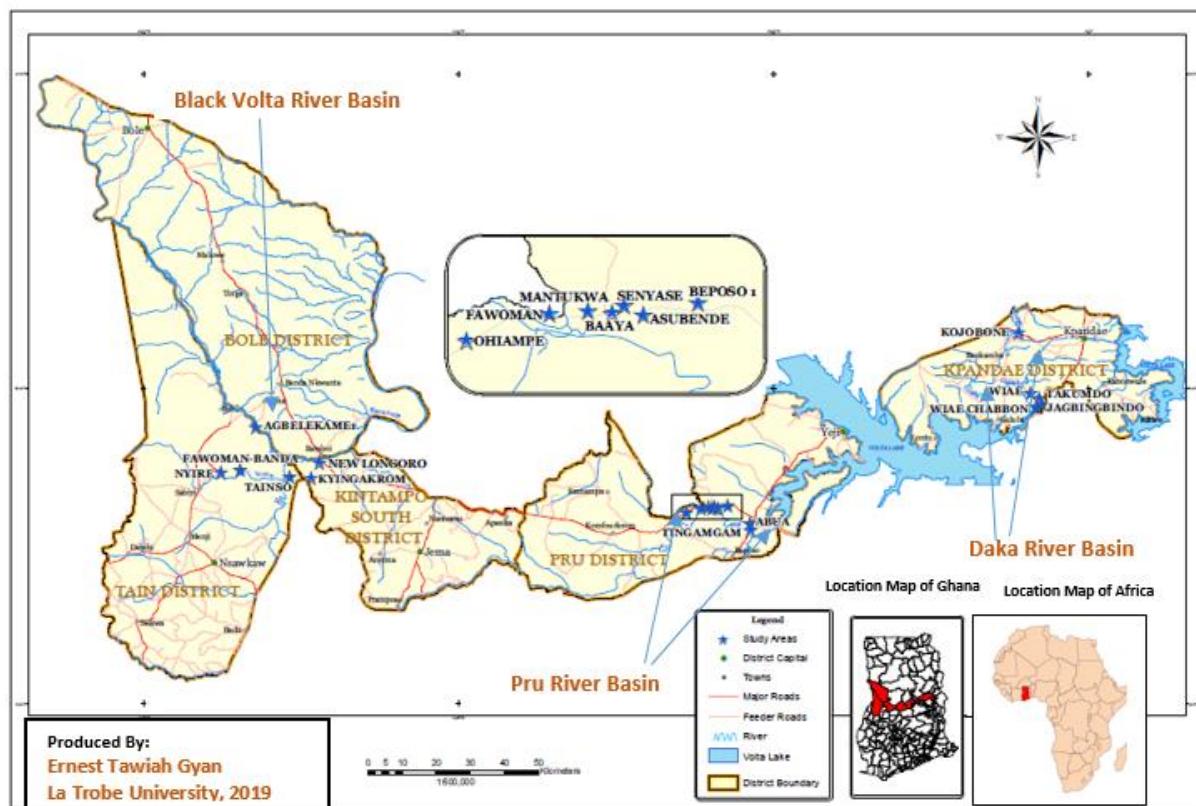


Figure 2.1: Map of study communities. This map was modified from Figure 2.3 of the study locations in the central ecological transition zone of Ghana (Gyan, 2013).

These communities in the ecological transition zone of Ghana for the genetic analysis were Wiae and Takumdo of Daka river basin in the East; Fawoman, Mantukwa, Beposo and Asubende of Pru river basin in the central part; and Agbekame and Kyingakrom of the Black Volta river basin in the West. The coordinates of the communities are found in Table 2.3.6. In a follow up work, additional 90 individual blackflies consisting of 30 blackflies from

each of the major river basins (Pru, Black Volta and Daka) were collected using similar procedure described in this section.

2.2.2 Sample processing

Each blackfly was taken from the 70% ethanol and dissected by separating the head, thorax, abdomen and legs into separate 1.5ml Eppendorf tubes. The thorax of each fly was thoroughly washed in distilled water and left in the water for 10 minutes to remove the ethanol. It was placed in -20°C freezer for 3 hours to freeze it.

2.2.3 DNA extraction with Qiagen's DNeasy Blood & Tissue Kit

DNA was extracted from 4 flies per each of the 8 communities using Qiagen's DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) by following the manufacturer's protocol with slight modifications to maximize the lysis efficiency and intactness of the DNA. An amount of approximately 20mg of -20°C frozen *Simulium* tissue was cut into small pieces and homogenized in a mixture of 75µl Phosphate-buffered saline (PBS) and 10µl of Buffer ATL. To each of the homogenized tissue was added an amount of 170µl of Buffer ATL and 20µl of proteinase K (600 mAU/ml). It was thoroughly mixed by 3 rounds of approximately 2.5 seconds pulse vortexing and incubated overnight at 56°C for 16 to 18 hours until the tissue completely lysed. The washing of the columns was followed by centrifugation at 2000 rpm for 30 seconds, and another 2 rounds of centrifugation with the first one at 4000 rpm for 60 seconds and the final one at 8000 rpm for 30 seconds. To elute the DNA, 50µl of a 70°C pre-heated buffer AE was added to the middle of the column that was earlier pre-heated at 70°C for 5 minutes. The membrane was incubated at room temperature for 60 seconds, centrifuged initially at 4000 rpm for 15 seconds and then finally centrifuged at 8000 rpm for 60 seconds to elute the final DNA.

2.2.4 DNA concentration determination with Qubit® Fluorometer

The Qubit® dsDNA High Sensitivity (HS) assay kits with the Qubit® Fluorometer (Thermo Fisher Scientific) were used to determine the concentration of DNA by following the manufacturer's protocol (Scientific, 2015) with minor modification. The prepared reagents containing the samples and standards were protected from sunlight using aluminium foil. All mixing prior to incubation were done by 3 rounds of pulse vortexing followed by 3-5 seconds of centrifugation at 14,000 rpm. After each incubation of the mixtures containing the samples and standards, the thin-walled PCR tubes were tapped gently to mix before quantification of the concentrations.

2.2.5 Amplification of short amplicons by polymerase chain reaction (PCR)

Primers flanking the 5¹ and 3¹ ends of each of the targeted genomic regions were used to amplify Cytochrome c oxidase I (CO1), Nicotinamide adenine dinucleotide dehydrogenase subunit 4 (ND4), 16S-rRNA (16S), Internal transcribed spacer 2 (ITS2) and Intergenic spacer (IGS) sections of the *Simulium* genome. The CO1, ND4 and 16S-rRNA genes were from the mitochondrial genome whiles the ITS2 and IGS were from the nuclear genome. Amplification of these short amplicons was carried out using Immolase™ DNA polymerase reagents (Bioline). The PCR was carried out in 10µl of reaction mix that consisted of 6.6µl of High-Performance Liquid Chromatography (HPLC) water, 1µl of 10x immoBuffer (Bioline), 0.4µl of 10mM dNTP, 0.6µl of 50mM MgCl₂, 0.1µl of 10uM sense primer, 0.1µl of 10uM antisense primer, and 1µl of 1ng DNA.

With the exception of the annealing temperatures that were different for each of the genes targeted for amplification, all aspects of the thermal cycling conditions were the same for all the genes. The thermal cycling conditions consisted of 95°C initial denaturation for 2 min, 30 cycles of: 95°C denaturing for 30 seconds; annealing of the primers to the DNA template for 30 seconds at temperatures of 50.1°C, 46 °C, 49 °C, 49.5 °C and 48.4 °C for CO1, ND4, 16S rRNA, ITS2 and IGS respectively; 72 °C extension for 45 seconds. These were finally extended for 10 minutes at 72 °C, and the products kept at 4 °C for 1 hour.

2.2.6 Gel electrophoresis

Electrophoresis of the PCR products were carried out using 1.5% agarose gel stained with 3X GelRed® Nucleic Acid Gel Stain (Biotium) by following the manufacturer's protocol. For each gel run, 5µl of PCR product was mixed with 2µl of 10X DNA loading dye (BlueJuice™ Gel Loading Buffer, Invitrogen™ by life technologies, U.S.A.), loaded into the well of the agarose gel and run in 1X TAE buffer at 100V for 45 minutes on the BioRad PowerPac 300 machine. All gel runs included 2µl of 10X 100bp DNA ladder (Invitrogen™) by ensuring approximately 0.1 µg of ladder per mm lane width and by following the manufacturer's protocol. Following electrophoresis, the gel was visualized and photographed using a 302nm UV lamp (BioRad Gel Doc™ EZ Imager). PCR products that showed clear single sharp band on the gel with the expected band size were further purified with the QIAquick PCR Purification Kit (QIAGEN, 2018) by following the manufacturer's protocol. The products were finally sent for Sanger sequencing.

2.2.7 Molecular phylogenetics analysis

The nucleotide sequences were imported into CLC Genomics Workbench 9.5.4 (<https://www.qiagenbioinformatics.com/>). They were aligned by Muscle (Edgar, 2004) and ClustalW (Thompson *et al.*, 1994). The chromatograms with quality scores showing as probabilities with 95% confidence and above were used to resolve nucleotide discrepancies and to avoid consensus base errors. The best nucleotide substitutional method was determined by computing the Bayesian information criterion (BIC), minimum theoretical Akaike information criterion (AIC), corrected minimum theoretical Akaike information criterion (AICc), and using the Hierarchical likelihood ratio test with a confidence level of 0.01 (Schwarz, 1978; Sakamoto *et al.*, 1986; Frati *et al.*, 1997; Huelsenbeck *et al.*, 1997; Posada and Crandall, 1998). The models tested included K80: Kimura 1980 model (Kimura, 1980), also known as K2P; JC69: Jukes-Cantor model (Jukes and Cantor, 1969); F81: Felsenstein 81 model (Felsenstein, 1981); HKY85: Hasegawa-Kishino-Yano model (Hasegawa *et al.*, 1985); and GTR: general-time reversible model (Tavaré, 1986), also known as REV. The model with the lowest BIC, AIC, or AICc value was chosen as the best-fit. The best-fit model was used to compute the final phylogenetic tree. Phylogenetic tree construction types used were Maximum likelihood tree, Maximum Parsimony tree, Minimum-Evolution tree, and Bayesian tree building with Mrbayes 3.2 (Ronquist *et al.*, 2012). Multiple phylogenetic trees that indicated similar blackfly ancestral relationships were critically examined and a representative tree with the overall best bootstrap values chosen. Relationships between individual genotypes at the population level were further examined by haplotype networks with Popart 1.7 (Bandelt *et al.*, 1999). Genetic distances, diversity indices and molecular variance were determined with Popart 1.7 (Bandelt *et al.*, 1999), MegaX 10.0.4 (Kumar *et al.*, 2018), and Arlequin 3.5.2 (Excoffier and Lischer, 2010).

2.2.8 Meta-analysis

2.2.8.1 Problem statement

The characterization of genetic diversity and delineation of onchocerciasis transmission zone partly depend on what is acceptable as a population and a species from a molecular point of view. Presently, there appears to be no concordance in the definition of species, in the context of *Simulium*, from the viewpoint of molecular taxonomy, morphotaxonomy and cytotaxonomy. This disagreement presents a great challenge in defining *Simulium* species based on observed groups of blackflies, and this lack of a commonality of what should constitute an interbreeding group of blackfly makes it extremely difficult drawing the geographical boundaries between different interbreeding group of blackflies.

2.2.8.2 Purpose of the meta-analysis

The main purpose of the meta-analysis was to determine the amount of genetic variability present in the broad taxonomic categories selected with a focus on the genetic diversity and divergence within and between populations of species, and of species within given genus. This was to provide information to enable an inference to be made on where the *Simulium* blackflies from the transition zone of Ghana fit on the broad genetic scale of the target organisms. It was also to enable a resolution of placing given groups of the studied *Simulium* blackflies, from the transition zone of Ghana, into appropriate levels of Molecular Operational Taxonomic Units (MOTU).

2.2.8.3 Inclusion criteria

The inclusion criteria were: (1) nucleotide sequences of the organism belonged to the kingdom Animalia; (2) nucleotide sequences that were either complete mitochondrial genomes or sequenced from one of the short amplicon genes that was targeted for amplification in this study; (3) the presence of nucleotide sequences for at least 2 different populations or species of the same organism in the same study; (4) members of a population or species were selected by the authors with justifiable reason for putting the members into the defined population or species groupings; (5) sufficient information to verify the source of the populations or species; and a (6) good number of nucleotide sequences, without missing loci, that would form good alignment to estimate genetic diversity.

2.2.8.4 Sample selection

Out of the 11,234 nucleotide sequences screened, 1,989 met the inclusion criteria. Thus, meta-analysis was carried out using the 1,989 nucleotide sequences consisting of whole mitochondrial genomes and short amplicon barcoding sequences. The nucleotide sequences

were obtained from public database including the nucleotide database of the National Center for Biotechnology Information (NCBI). They were used to assess the genetic diversity and divergence in different species. The amplicon sequences used were guided by the published amplicons of *Simulium* blackflies. The organisms were selected from the broad taxonomic categories of the wild and domestic mammals (*Bos* sp.), free living nematodes (*Caenorhabditis* sp.), parasitic nematode of humans (*Onchocerca volvulus*), parasitic nematode of animals (*Onchocerca ochengi*), lizards (*Chlamydosaurus kingii*), turtle (*Chelodina* sp.), crocodiles (*Crocodylus* sp.), fruit fly (*Drosophila* sp.), the fossil ancestor family of coelacanths (*Latimeria* sp.), fishes (*Oreochromis* sp. and *Variabilichromis moorii*), field and house mouse (*Mus* sp. and *Apodemus* sp.), and honeyeater birds (*Entomyzon*).

2.2.8.5 Diversity and divergence estimation

The diversity and divergence within populations of the same species were estimated with *Entomyzon cyanotis*, *Variabilichromis moorii*, *Chlamydosaurus kingii*, and the *Oreochromis niloticus* populations in both Myanmar and DRC.

The diversity and divergence within species of the same genera were estimated with *Caenorhabditis* sp., *Onchocerca* sp., *Chelodina* sp., *Crocodylus* sp., *Drosophila* sp., *Latimeria* sp., and *Oreochromis* sp. The house and field mice *Mus* sp. and *Apodemus* sp. were added to this category to serve as the highest boundary of diversity and divergence expected in this study of organisms within the same genera because although both mice look strikingly alike in most ways of mice comparison, they are actually different in belonging to different genera of *Mus* and *Apodemus*.

2.2.8.6 Molecular phylogenetic and statistical analysis

Similar molecular phylogenetic and statistical analysis were carried out as in section 4.2.18 of this thesis.

2.2.8.7 Limitations of the meta-analysis

Major limitations of the meta-analysis included: (1) fewer number of organisms that met the inclusion criteria for the intra-species genetic variation estimation; (2) Unequal sample size for the *Bos* sp.; (3) dependence on the classification and grouping of organisms into the intra-species and inter-species categories without having any idea of the competence of the workers doing the work and whether any assumptions were violated; (4) all the organisms were not collected in the same year and may not reflect diversity from similar generations; (5) the collections of the organisms were made solely for the objectives of the authors and may not

have included those samples of the population that would have likely provided greater accuracy in the estimation of the genetic diversity of this current study; and (6) some of the *Bos* sp. were domesticated, providing the advantage of capturing the expected variation within natural and artificial settings but with the disadvantage of not representing a truly natural group for comparison. In view of these limitations, interpretation of the results based on these data needed to be made with caution.

2.2.9 Ethical clearance

Ethical approval for the study protocol was granted by the Institutional Review Board of the Ghana Council for Scientific and Industrial Research (CSIR-IRB), as part of an existing study titled: “Development of diagnostic genetic markers to detect sub-optimal response to ivermectin treatment”. A copy of the approval document is shown in Supplementary sheet 1.

2.3 Results

2.3.1 Meta-analysis assessment of diversity and divergence in different species

A total of 1,989 nucleotide sequences obtained from public database (NCBI) were used to assess the diversity and divergence in different species (Table 2.3.1). These consisted of 40.2% (799) whole mitochondrial genomes, with the remaining 59.8% (1190) representing amplicon sequences of ND4, CO1, ND2, cytochrome b, 16S-rRNA, ITS2 and IGS. The alignment length of the complete mitochondrial genome sequences ranged from 13,582 bp (in *Caenorhabditis* sp.) to 16,978 bp (in *Crocodylus* sp.), while that for the amplicon sequences ranged from 196 (in *Simulium* sp.) to 1,005 bp in *Entomyzon cyanotis*. The genomic regions involved in this current study were from the mitochondrial regions of CO1, ND4, 16S-rRNA; and the nuclear regions of ITS2 and IGS. The accession numbers of all the organisms used are shown in Supplementary Tables 2.3.1- 19.

Out of the 13 main taxonomic groups of organisms involved in this meta-analysis, 3 consisted of individual populations within a single species and were used to estimate intra-species variation of each of the corresponding organism. Each of the remaining 10 consisted of different species in a single genus. The 3 taxonomic groups of organisms for the intra-species diversity work were *Entomyzon cyanotis*, *Variabilichromis moorii* and *Chlamydosaurus kingii*. Two populations were identified among *Entomyzon cyanotis*, and their proportions were 25.7% (9) and 74.3% for population 1 and 2 respectively. Three populations were identified among *Variabilichromis moorii* and their proportions were 35.2% (56), 25.2% (40) and 39.6% (63) for populations 1, 2 and 3 respectively. *Chlamydosaurus kingii* consisted of 2 populations and the respective proportions for population 1 and 2 were 56.3% (18) and 43.7% (14) respectively.

The *Onchocerca* spp. consisted of *Onchocerca volvulus* and *O. ochengi* with 3 number of individuals each. This total number of sequences for the *Onchocerca* sp. appear small but they were the only sequences available and did not significantly change the expected result. Among the *Drosophila* spp. existed *D. simulans*, *D. melanogaster* and *D. yakuba* with proportions of 35.6% (32), 31.1% (28) and 33.3% (30) respectively. Within *Latimeria* sp. were *L. chalumnae* and *L. menadoensis* with 80% (12) and 20% (3) proportion of individuals. The *Chelodina* spp. were made up of *C. oblonga* and *C. parkeri* with 60% (3) and 40% (2) of individuals. The *Bos* spp. had 3 species made up of *B. frontalis*, *B. grunniens* and *B. taurus* with 1% (4), 23.4% (95) and 75.6% (307) individuals. The *Crocodylus* sp. was composed of 3 (13%) individuals of *C. palustris*, and 2 (8.7%) individuals each of the remaining 10 species:

C. acutus, *C. intermedius*, *C. moreleti*, *C. rhombifer*, *C. niloticus*, *C. mindorensis*, *C. novaeguineae*, *C. siamensis*, *C. porosus* and *C. johnsoni*. There were 13 species of *Caenorhabditis* with *C. elegans* and *C. remanei* being composed of 3 (10.7%) individuals each whiles the remaining 11 were made up of 2 (7.1%) individuals each: *C. doughertyi*, *C. wallacei*, *C. brenneri*, *C. nouraguensis*, *C. macrosperma*, *C. castelli*, *C. angaria*, *C. tropicalis*, *C. virilis*, *C. afra* and *C. plicata*. Two populations of *Oreochromis niloticus* were found in Myanmar and DRC with 16 (48.5%) and 3 (9.1%) individuals each. The remainder of the *Oreochromis* spp. were made up of 2 (6.1%) individuals of *O. leucostictus*, and 6 individuals (18.2%) each of *O. mossambicus* and *O. aureus*. The *Mus* sp. and *Apodemus* sp. used as the potential source of highest genetic variability among these studied taxonomic organisms were made up of 149 (67.1%) and 73 (32.9%) individuals respectively. The 931 amplicon sequences of the *Simulium* sp. from the NCBI database were selected to consist of CO1, Cytb, ND4, 16S-rRNA, ITS2 and IGS genes.

The intra-species genetic variation was estimated by examining the variation existing in groups within the same species of organism (blue border in Table 2.3.1). The inter-species variation on the other hand was examined by looking at the genetic variation between different species of the same organism (green border in Table 2.3.1).

Table 2.3.1: Sequence characteristics of different organisms

Organism name	Sequence type	Alignment length (bp)	Number of sequences
<i>Chlamydosaurus kingii</i>	Mitochondrial genome	16,761	4
<i>Chlamydosaurus kingii</i>	ND4	627	32
<i>Variabilichromis moorii</i>	CO1	661	159
<i>Entomyzon cyanotis</i>	ND2	1,005	35
<i>Latimeria</i> sp.	Mitochondrial genome	16,452	15
<i>Crocodylus</i> sp.	Mitochondrial genome	16,978	23
<i>Drosophila</i> sp.	Mitochondrial genome	14,978	90
<i>Bos</i> sp.	Mitochondrial genome	15,798	406
<i>Caenorhabditis</i> sp.	Mitochondrial genome	13,582	28
<i>Chelodina</i> sp.	Mitochondrial genome	16,620	5
<i>Onchocerca</i> sp.	Mitochondrial genome	13,774	6
<i>Oreochromis</i> sp.	CO1	621	33
<i>Simulium</i> sp. (Group 1)	CO1	831	131
<i>Simulium</i> sp. (Group 2)	CO1	586	217
<i>Simulium</i> sp.	Cytochrome b	313	376
<i>Simulium</i> sp.	ND4	196	32
<i>Simulium</i> sp.	16S-rRNA	467	55
<i>Simulium</i> sp.	ITS2	332	80
<i>Simulium</i> sp.	IGS	615	40
<i>Mus</i> sp.	Mitochondrial genome	16,371	149
<i>Apodemus</i> sp.			73

In the table, blue border = organisms used to estimate intra-species genetic variation; green = organisms used to estimate inter-species genetic variation; and red border = field and house mouse, different genera but morphologically similar.

Among all the organisms studied, the most diverse was the CO1 of *Simulium* sp., set 2, ($\pi = 0.12906$) whiles the least diverse were *Entomyzon cyanotis* population 1 ($\pi = 0$) and *Chelodina parkeri* ($\pi = 0$). The nucleotide diversity in the *Mus* sp. and *Apodemus agrarius* was 0.015055 (Supplementary sheet 10). Overall, there was moderate nucleotide diversity but high haplotype diversity across all groups of organisms. The total nucleotide diversity ranged from 0.00463 to 0.12906 with an average of 0.049503. The total haplotype diversity ranged from 0.56061 to 0.9998 with an average of 0.898233. Details of the polymorphisms and diversity are found in Tables 2.3.2 A and B.

Table 2.3.2 A: Polymorphism and diversity within and across different species

Organism Name	M	π	H	Hd	M_T	π_T	H_T	Hd_T
<i>Latimeria chalumnae</i>	14	0.00026	12	1	667	0.01385	14	0.99048
<i>Latimeria menadoensis</i>	3	0.00012	2	0.66667				
<i>Entomyzon cyanotis</i> population 1	0	0	1	0	45	0.01676	10	0.86050
<i>Entomyzon cyanotis</i> population 2	14	0.00454	9	0.85538				
<i>Chelodina parkeri</i>	0	0	1	0	2373	0.07895	3	0.8
<i>Chelodina oblonga</i>	966	0.03912	2	0.66667				
<i>Chlamydosaurus kingii</i> population 1	5	0.00179	6	0.75817	15	0.00741	10	0.86694
<i>Chlamydosaurus kingii</i> population 2	4	0.00161	5	0.76923				
<i>Onchocerca volvulus</i>	28	0.00136	2	0.66667	539	0.02123	4	0.86667
<i>Onchocerca ochengi</i>	157	0.00763	2	0.66667				
<i>Crocodylus niloticus</i>	191	0.01141	2	1	4202	0.07475	14*	0.96443
<i>Crocodylus porosus</i>	233	0.01392	2	1				
<i>Variabilichromis moorii</i> population 1	7	0.00057	8	0.35065	22	0.00463	21	0.83059
<i>Variabilichromis moorii</i> population 2	6	0.00191	7	0.80128				
<i>Variabilichromis moorii</i> population 3	6	0.00156	6	0.50179				
<i>Caenorhabditis elegans</i>	42	0.00208	3	1	6228	0.12074	15	0.966
<i>Caenorhabditis remanei</i>	308	0.01529	2	0.66667				

M = Total number of mutations, π = Nucleotide diversity, H = Number of haplotypes, Hd = Haplotype diversity, M_T = Overall total number of mutations among groups compared, π_T = Total nucleotide diversity, H_T = Total number of haplotypes, Hd_T = Total haplotype diversity

*= Estimates of the total haplotype and total haplotype diversity were made using 14 different species of the *Crocodylus* genus to reflect the number of haplotypes expected among the species of the genus.

Table 2.3.2 B: Polymorphism and diversity within and across different species (continue)

Organism Name	M	π	H	Hd	M_T	π_T	H_T	Hd_T
<i>Oreochromis niloticus</i> (Myanmar)	32	0.00644	2	0.12500	76	0.03345	6	0.56061
<i>Oreochromis niloticus</i> (DRC)	51	0.05475	2	0.66667				
<i>Simulium</i> sp. (Mali), IGS	26	0.00648	13	1	69	0.01175	38	0.997
<i>Simulium</i> sp. (Liberia)	18	0.00611	10	1				
<i>Simulium</i> sp. (Ghana)	26	0.00925	16	0.99265				
<i>Simulium</i> sp. (ITS2)	70	0.05101	28	0.926	70	0.05101	28	0.926
<i>Simulium</i> sp. (CO1), set 1 (131 Sequences)	493	0.08528	129	0.9998	493	0.08528	129	0.9998
<i>Simulium</i> sp. (CO1), set 2 (217 Sequences)	394	0.12906	130	0.9847	394	0.12906	130	0.9847
<i>Simulium</i> sp. (16S), (55 Sequences)	81	0.01819	25	0.921	81	0.01819	25	0.921
<i>Simulium</i> sp. (ND4), (32 Sequences)	58	0.08088	14	0.893	58	0.08088	14	0.893
<i>Simulium</i> sp. (CYTB), (376 Sequences)	132	0.04410	135	0.9440	132	0.04410	135	0.9440

M = Total number of mutations in a given population within species or in a given species within genus, π = Nucleotide diversity in a given population within species or in a given species within genus, H = Number of haplotypes in a given population within species or in a given species within genus, Hd = Haplotype diversity in a given population within species or in a given species within genus, M_T = Overall total number of mutations among groups compared, π_T = Total nucleotide diversity, H_T = Total number of haplotypes, Hd_T = Total haplotype diversity.

The Tajima's D test was used to determine if the known population sequences involved in the meta-analysis fitted the neutral theory model at equilibrium between mutation and genetic drift (Table 2.3.3). As a population genetic statistic, the Tajima's D was computed for the intra-species sequences that constituted populations or members of the same species. It was not computed for the inter-species sequences that did not constitute individual populations. The Tajima's D from each of *Entomyzon cyanotis*, *Chlamydosaurus kingii*, and *Variabilichromis moorii* were not statistically significant ($p > 0.05$). While the Tajima's D value of the fish was below zero with the indication of a high frequency of rare alleles; that of the lizard and bird were all above zero, thus suggesting the lack of rare alleles or the presence of a low frequency of rare alleles. However, the lack of statistical significance by the use of the Tajima's D in all the species tested indicated that the allelic distribution and/or level of variability did not violate the neutrality assumption. Hence, the sequences were all evolving randomly (neutrally). To test if there existed sequences that had departures from neutrality due to genetic hitchhiking and population expansion, Fu and Li's statistics were used. The neutrality test for *Variabilichromis moorii* was generally inconclusive due to the fact that while both Tajima's D and Fu & Li's F statistical test indicated that the sequences were evolving randomly ($p > 0.05$), the Fu and Li's D test suggested a likely departure from neutrality due to possible hitchhiking or population expansion. All the neutrality test carried out suggested that the *Chlamydosaurus kingii* sequences were evolving randomly as expected ($p > 0.05$). Both Fu and Li's D and F statistical test suggested that the *Entomyzon cyanotis* sequences were not evolving randomly ($p < 0.05$) due to possible genetic hitchhiking or population expansion, an inference that the Tajima's D test's less sensitivity to such evolutionary phenomenon could not detect.

Table 2.3.3: Neutrality Test of the sequences used in intra-species variation estimation

Organism	Tajima's D	Tajima's p	FLD*	FLD* p	FLF*	FLF* p
<i>Entomyzon cyanotis</i>	1.9584	$p > 0.05$	1.52236	$p < 0.05$	1.97824	$p < 0.05$
<i>Chlamydosaurus kingii</i>	0.8217	$p > 0.05$	-0.06246	$p > 0.05$	0.25649	$p > 0.05$
<i>Variabilichromis moorii</i>	-0.5973	$p > 0.05$	-2.33650	$p < 0.05$	-1.98286	$p > 0.05$

The Tajima's D was calculated using the total number of mutations. In the table, FLD* = Fu and Li's D* test statistic, FLF* = Fu and Li's F* test statistic, p = Statistical significance

The proportion of variation within and among populations in a given species was assessed by the Analysis of Molecular Variance (AMOVA). The fixation index (F_{ST}) was used to measure the population differentiation resulting from genetic structure. All the genetic variation observed in *Oreochromis* sp. was found among populations ($VAP = 100\%$) with none within populations, and its populations were fixed with no interbreeding between them ($F_{ST} = 1, P < 0.001$). A similar observation of fixed populations was observed among the *Latimeria* sp. with strong indication of assortative mating. *Chelodina* sp. showed strong population structure with 77.8% of its genetic variation observed among population, with 22.2% of the variation found within population ($F_{ST} = 0.8, p < 0.001$). A similar strong population structure was found in *Variabilichromis moorii* with 87.6% of the genetic variation observed among population, with 12.4% of the variation found within population ($F_{ST} = 0.9, p < 0.001$). The taxonomic groupings of *Simulium* found in Ghana, Mali and Liberia were treated as testable population units without assigning the original taxonomic names given by the authors of those sequences. With the exception of the *Simulium* populations in Mali, Ghana and Liberia, the populations of all the remaining species not yet described on Table 2.3.4 had very strong population structures (F_{ST} between 0.943 and 0.999, $p < 0.001$) with 94.3 % to 99.9% of the genetic variation found among populations while the remaining 0.1% to 5.7% of the genetic variation was found within populations. A moderate population structure was found among the *Simulium* populations in Mali, Ghana and Liberia ($F_{ST} = 0.56, p < 0.001$), with 56.2% of the genetic variation found among the populations while the remaining 43.8% of the variation was found within the populations. A Mantel test utilizing Spearman rho correlation demonstrated a moderate isolation by distance ($r = 0.5$) of the *Simulium* populations in Mali, Ghana and Liberia. However, the test was not statistically significant ($p = 0.5$) possibly due to inadequate number and locations of samples. The majority of the genetic variation in *Mus* sp. and *Apodemus agrarius* was found among groups (98.76117%) with little of the variation found within group (1.23883%) (see Supplementary sheet 11).

Table 2.3.4: Measures of population structure by analysis of molecular variance (AMOVA)

Organism	VAP (%)	VWP (%)	Phi _{ST}	Phi _{ST} p
<i>Simulium</i> sp. IGS (Mali, Ghana & Liberia)	56.2	43.8	0.56	$p < 0.001$
<i>Oreochromis</i> sp.	100	0	1	$p < 0.001$
<i>Caenorhabditis</i> sp.	99.7	0.3	0.997	$p < 0.001$
<i>Chelodina</i> sp.	77.8	22.2	0.77756	$p < 0.001$
<i>Chlamydosaurus kingii</i>	94.7	5.3	0.947	$p < 0.001$
<i>Variabilichromis moorii</i>	87.6	12.4	0.876	$p < 0.001$
<i>Boss</i> sp.	99.6	0.4	0.996	$p < 0.001$
<i>Crocodylus</i> sp.	99.5	0.5	0.995	$p < 0.001$
<i>Drosophila</i> sp.	94.3	5.7	0.943	$p < 0.001$
<i>Entomyzon cyanotis</i>	99.6	0.4	0.996	$p < 0.001$
<i>Mus</i> sp. and <i>Apodemus agrarius</i>	99.9	0.1	0.999	$p < 0.001$
<i>Latimeria</i> sp.	100	0	1	$p < 0.001$

VAP= Percentage of Variation among population; VWP= Percentage of Variation within population; Phi_{ST} = Fixation index; Phi_{ST} p = Significance of the Fixation index (1000 permutations)

Nucleotide sequence similarity was measured by percentage identity. Generally, from Table 2.3.5, nucleotide sequence similarity was greatest within group than among group, with the least similarity within groups observed among the *Drosophila* sp. (PWS = 97.49 - 100%). The most similar within groups was observed among *Oreochromis* sp. with no variation seen within the same type of group examined (PWS = 100%). Overall, the least identical of the studied groups of organisms was *Mus* sp. and *Apodemus* sp. (PAS = 83.62% - 83.87%), followed by the *Caenorhabditis* sp. (PAS = 84.56 – 93.18). The least identical group of the studied organisms was the *Variabilichromis moorii* population (PAS = 98.94% - 99.39%), followed by the *Chlamydosaurus kingii* population (PAS = 98.415 – 98.88%). Among the African *Simulium*, the similarity within and among groups was generally great (> 97%), with the within group similarity ranging from 98.83% to 99.7% while the among group similarity ranged from 97.34% to 98.09%. In order of increasing magnitude, the geographical distances between sampling locations of the studied African *Simulium* blackflies were 696.61 km, 1149.14 km, and 1204.68 km for the pairwise location comparison of Mali with Liberia, Mali with Ghana, and Ghana with Liberia respectively. These distances are beyond the flight range of blackflies.

Table 2.3.5: Percentage identity of organisms across species

Organism	PWS (%)	PAS (%)
<i>Latimeria</i> sp.	99.95 - 100	95.92 – 95.96
<i>Crocodylus</i> sp.	98.52 - 100	90.37 – 99.47
<i>Onchocerca</i> sp.	98.86 - 100	96.36 – 96.38
<i>Variabilichromis moorii</i> (populations within a species)	99.39 - 100	98.94 - 99.39
<i>Oreochromis</i> sp.	100	92.17 - 96.94
<i>Chlamydosaurus kingii</i> (populations within a species)	99.36 - 100	98.41 – 98.88
<i>Drosophila</i> sp.	97.49 - 100	92.9 – 95.81
<i>Caenorhabditis</i> sp.	99.58 - 100	84.56 – 93.18
<i>Entomyzon cyanotis</i> (populations within a species)	99.6 - 100	96.12 – 99.42
<i>Bos</i> sp.	99.2 - 100	93.47 – 94.35
<i>Mus</i> sp. and <i>Apodemus</i> sp. (between genera)	98.75 - 100	83.62 - 83.87
<i>Simulium squamosum</i> AF421568 ^a (Ghana) ^a	99.7 ^{ad}	96.91 ^{ab}
<i>Simulium sirbanum</i> AF421594 ^b (Mali) ^b	99.41 ^{ce}	97.34 ^{ac}
<i>Simulium soubrense</i> AF421619 ^c (Liberia) ^c	98.83 ^{bf}	98.09 ^{bc}
<i>Simulium squamosum</i> AF421574 (Ghana) ^d		
<i>Simulium soubrense</i> AF421618 (Liberia) ^e		
<i>Simulium sirbanum</i> AF421589 (Mali) ^f		

PWS = Percentage identity within group (either population or species), PAS = Percentage identity among group (either population or species). The compared variations of groups are appropriately indicated in parenthesis with green = populations within a species, blue = between genera; the rest show comparisons between species of the same genera.

Generally, across well-defined species (Figure 2.3.1), members of a species formed distinct clades with good bootstrap support greater than 80%. Among the *Caenorhabditis* sp., *C. plicata* formed a distinct clade that was separate from the clades of all the other species as though it was outgroup. With the exception of *C. plicata* of the *Caenorhabditis* sp., members of *C. virilise* constituted a monophyletic group while all the members of the other species also formed another monophyletic group. However, the divergence was such that members of each taxonomic group acquired sufficient unique mutations that separated them on molecular basis from members of other taxonomic group. There was good (82%) to excellent (100%) bootstrap support for the separation of each group (either population or species) from one

another, thereby making subsequent percentage identity inference based on these taxonomic groupings to be most likely a true representation of taxonomic group level separation within the *Caenorhabditis* spp.

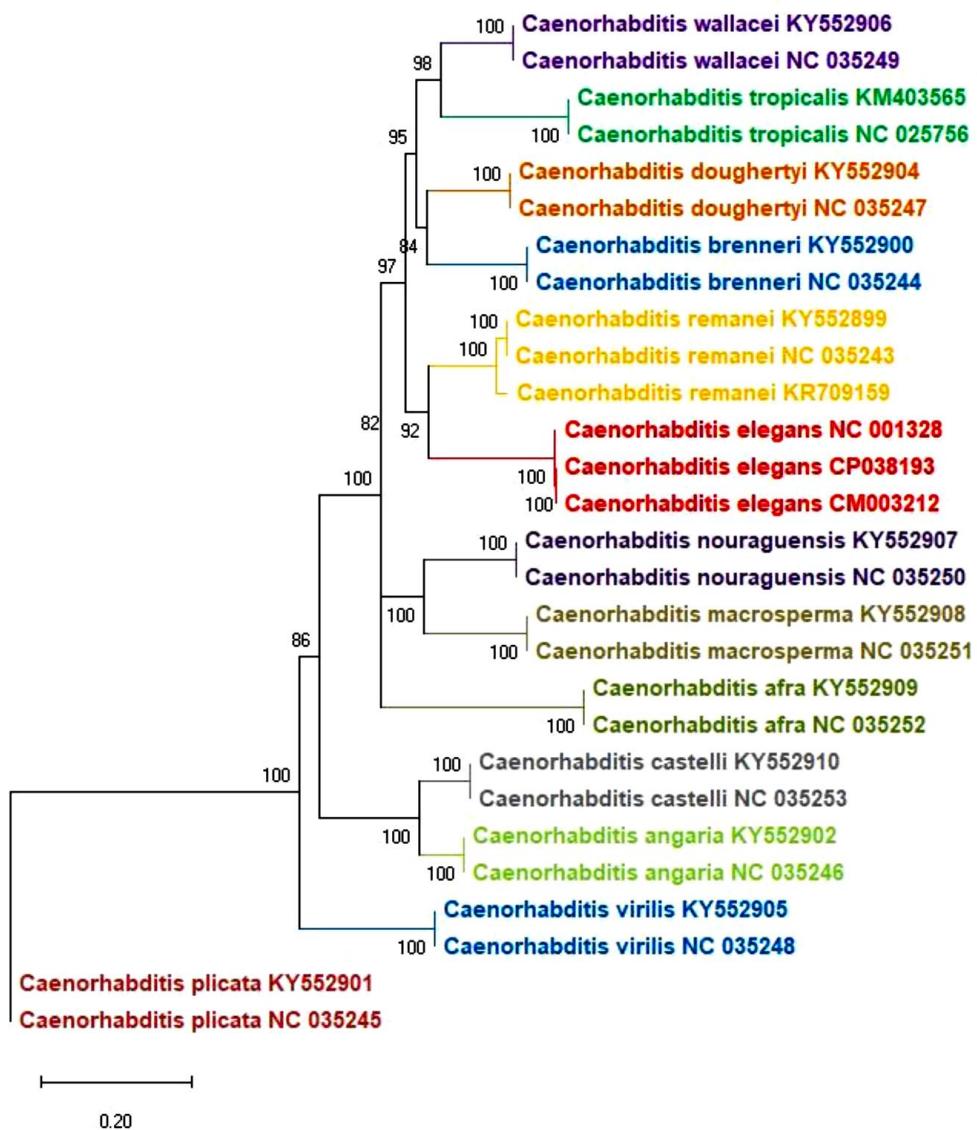


Figure 2.3.1: Phylogenetic tree of *Caenorhabditis* sp. using whole mitochondrial genomes.

Scale bar represents substitutions per site.

Among the fruit flies, *Drosophila yakuba* formed a clade completely different from those of *D. simulans* and *D. melanogaster* as if *D. yakuba* was an outgroup (Supplementary Figure 2.3.1). *D. melanogaster* and *D. simulans* shared most common recent ancestor and their members independently formed separate monophyletic groups. *D. yakuba* was much more

diverse than *D. simulans* and *D. melanogaster*, and this great divergence captured the wide range of variation among the fruit flies. Nevertheless, the divergence between these 3 species of fruit flies had excellent (100%) bootstrap support, thereby making any form of percentage identity inference based on their species level separation to be a true representation of species level percentage identity among the fruit flies. Using phylogenetic, DAPC, haplotype network and population structure analyses, similar observations were made on the species level separation among *Oreochromis* sp., *Onchocerca* sp., *Bos* sp., *Chelodina* sp., *Latimeria* sp., *Crocodylus* sp., *Mus* sp. and *Apodemus* sp. (Supplementary Figures 2.3.7-10, and Supplementary Figures 2.3.13-15).

Population level separation was determined by studying the diversity, divergence, and percentage identity at the population level. Among the *Entomyzon cyanotis*, 2 major populations were identified (Supplementary Figure 2.3.2) and they contained sufficient diversity to infer population level differences. Also, the number of distinct major populations independently identified in the *Chlamydosaurus kingii* and *Variabilichromis moorii* were 2 and 3 respectively (Supplementary Figures 2.3.11, 2.3.12A and 2.3.12B).

In contrast to the species trees described above, the species level phylogenetic analysis of *Simulium* shows extremely moderate divergence per informative site and correspondingly poor bootstrap support using the nuclear encoded IGS amplicon (Figure 2.3.2). This is despite the strong inference of population structure when these data are used to construct a haplotype network based on country of origin (Figure 2.3.3) rather than on morphological species definition (Figure 2.3.2; and Supplementary Figures 2.3.3-6).

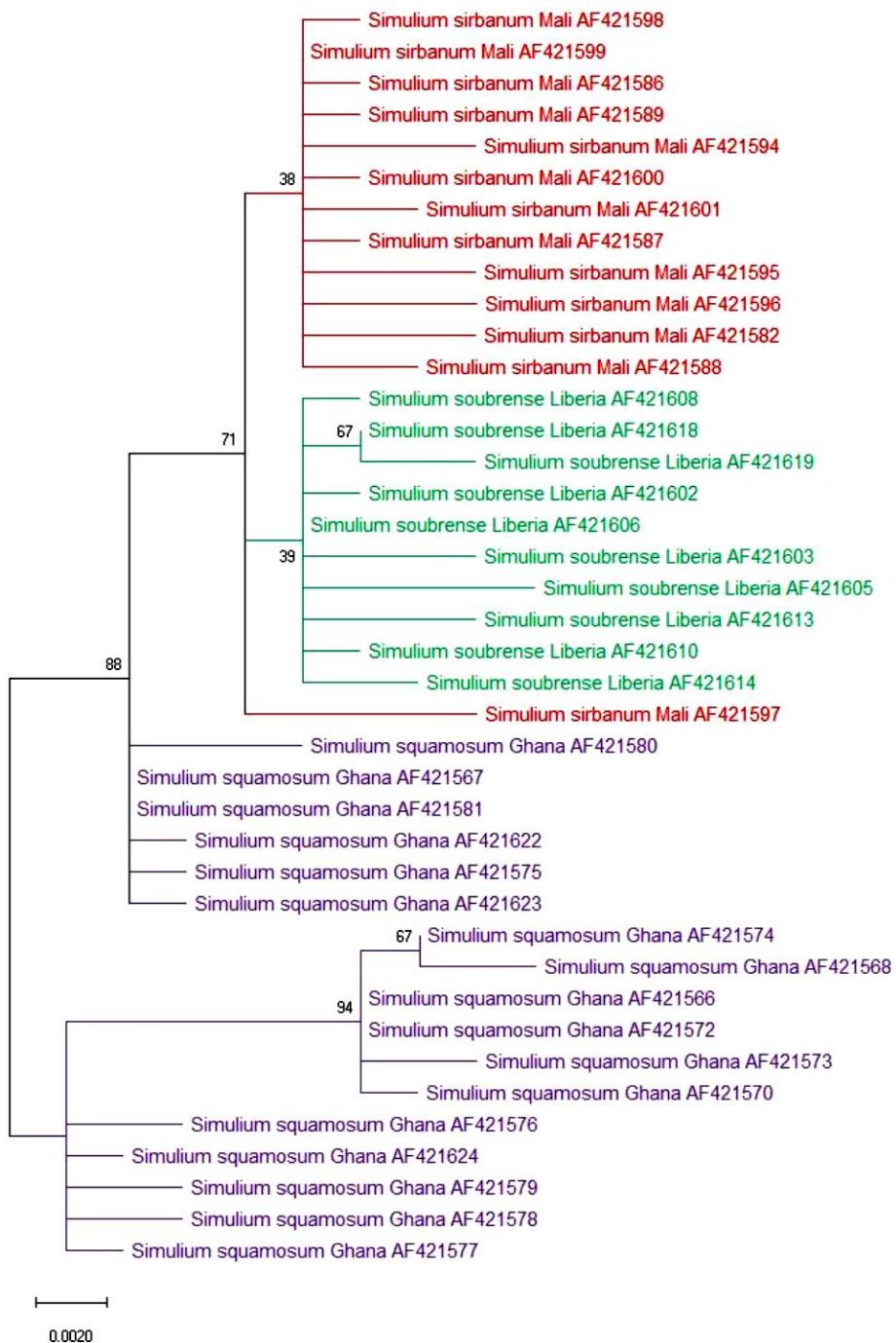


Figure 2.3.2: Phylogenetic tree of *Simulium* sp. from Ghana, Mali and Liberia using IGS

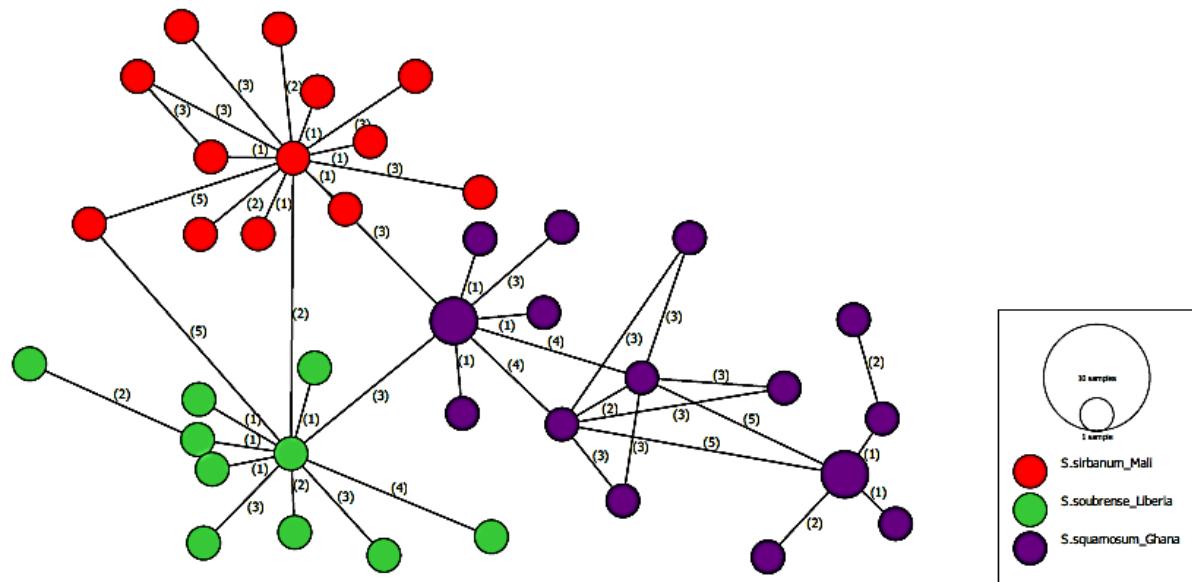


Figure 2.3.3: Haplotype network of *Simulium* sp. from Ghana, Mali and Liberia using IGS

2.3.2 Sample and sequence characteristics

A total of 612 individual blackflies were successfully sequenced. These consisted of: (a) 160 (26.14%) of the commonly used barcoding genes less than 1 kbp in the Sanger sequencing of 4 blackflies from all 8 communities; (b) 279 (45.59%) amplicon resequencing of some of the commonly used barcoding genes in the MiSeq run using 30 blackflies per each of the 3 communities; (c) 91 (14.87%) relatively long range-amplicons of 2.95 kbp; and (d) 82 (13.4%) whole genome sequences (Figure 2.3.4). The short amplicons amplified sequences from the COX1, ND4, 16S-rRNA, ITS2, and IGS genes from 32 blackfly DNA's from each of these 5 genes and with the samples selected from 8 different communities distributed in the 3 river basins of Pru, Black Volta and Daka as indicated in the sample collection section 2.2.1 of this thesis. The amplicon re-sequencing work used larger numbers of individual blackflies (279) from the 3 representative sampling locations of Pru, Black Volta and Daka river basins to re-amplify COX1, ND4 and 16S rRNA while the long-range sequencing amplified the contiguous mitochondrial region encompassing the tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 genes using 30 individual blackflies from the same river basins as in the short amplicon work, and 1 additional sampling location in the Pra river basin of South-Western Ghana. The whole genome sequencing strategy sequenced and assembled whole mitochondrial genomes, and concatenated nuclear genes that consisted of 18S rRNA, ITS1, 5.8S rRNA, 2S rRNA, ITS2, and 28S rRNA. It is worth emphasizing that with the exception of the additional locations of Malawi and Cameroon to the whole genome sequencing work,

the blackflies used in all the 4 types of sequencing were collected from the same geographical locations of the Pru, Black Volta and Daka river basins in the central ecological transition zone of Ghana.

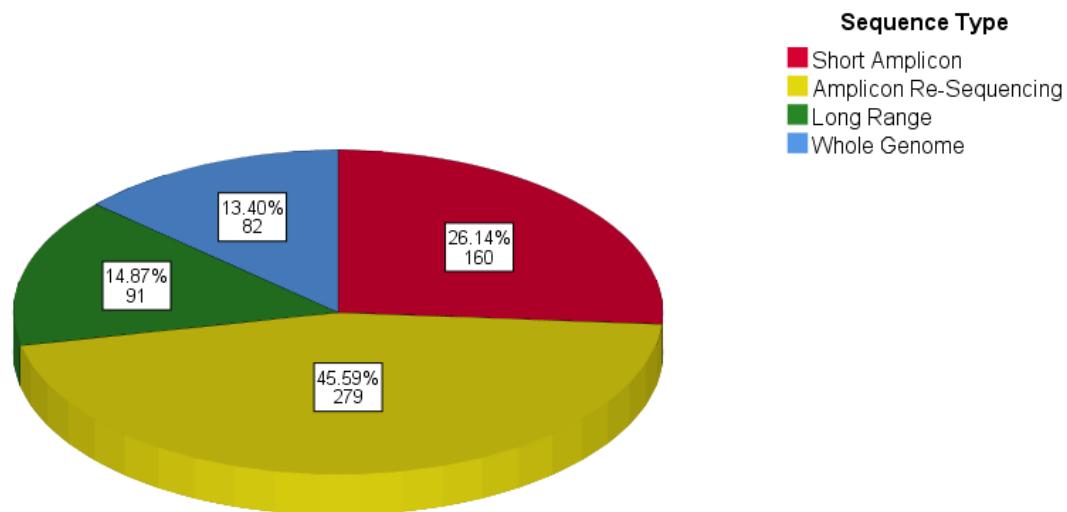


Figure 2.3.4: Pie chart of *Simulium* blackflies sequenced by different sequencing strategies. The segments of the pie chart represent the proportion of sequences produced from the corresponding sequencing strategy: Sanger for the short amplicon (red); MiSeq for the amplicon re-sequencing (yellow) and the long-range (green); and both MiSeq and Nextseq sequencing for the Whole Genome Sequencing.

The samples of the blackflies sequenced were all collected from the transition ecological zone of Ghana. Out of these 19 communities, 8 (green colour) were selected to represent the whole transition ecological zone of Ghana. The comparative outgroups were selected from South Western Ghana (Tofoi), Cameroon and Malawi. The coordinates of the communities are shown in Table 2.3.6.

Table 2.3.6: sampling locations with coordinates used in this study

River Basin	Community	Geographic Coordinates
Pru (BAR)	Asubende	8°01'00.4"N 0°58'53.2"W
	Baaya	08.01'11.9"N 000.59'44.4"W
	Ohiampe	08° 00'26.2N 001° 03'49.8"W
	Senyase	08.01'24.7"N 000.59'25.1"W
	Abua	07.58'38.7"N 000.53'48.5"W
	Fawoman	8°00'40.0"N 1°00'46.9"W
	Beposo	8°00'09.4"N 0°34'26.2"E
Tombe (BAR-Banda)	Mantukwa	8°00'47.3"N 1°00'30.5"W
	Fawoman-Banda	08.07'10.8"N 002.14'41.0"W
Black Volta (BAR)	Nyire	08.06'44.0"N 002.17'46.0"W
	Kyingakrom	8°03'20.8"N 2°01'57.6"W
	Agbelekame	8°14'32.5"N 2°12'38.6"W
Tain	New Longoro	08.08'15.8"N 002.02'09.6"W
	Tainso	08.06'03.0"N 002.06'53.8"W
Daka (NR-Kpandai)	Jagbenbendo	08.17'39.8"N 000.07'32.1"W
	Takumdo	8°18'55.3"N 0°07'53.3"W
	Wiae	8°17'09.8"N 0°08'40.1"W
	Kojoboni	08.29'07.7"N 000.10'59.3"W
	Wiae Chabbon	08.17'15.3"N 000.08'11.3"W
Nkam	Timté (littoral province)	4°34'12.7"N 10°09'49.0"E
Mbam	Ébebda II (Centre region)	4°22'44.0"N 11°15'50.0"E
Lingoni Falls	Thyolo	16°08'00.0"S 35°09'00.0"E
Pra	Tofoi	5°35'55.8"N 1°33'06.6"W

In order of increasing magnitude, amplification of the genomic regions of ND4, ITS2, 16S rRNA, IGS, COX1, and SIM4 (tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1) produced respective amplicons of 250bp, 600bp, 650bp, 700bp, 720bp, and 2.95Kbp (See Figure 2.3.5). The identity of each PCR product was confirmed by Sanger sequencing. The results from a BLAST search against the NCBI nucleotide database, by using the Sanger sequences as query sequences, indicated the highest sequence identity match (>96%) of each of the produced sequences to the respective targeted genomic regions, with various *Simulium* sp. as the topmost organisms found. These confirmed that the right genomic regions were successfully amplified. The scope of this study did not include morphological and cytological identification of the blackflies. Hence, it can only be inferred that the blackflies identified in the study area only belong to the *Simulium damnosum* complex widely accepted by authors of the field as the main species in the central ecological transition zone of Ghana. This conclusion was also arrived partly because of the great similarity of the studied genes greater than 96% identity match to only members of the *Simulium damnosum* complex.

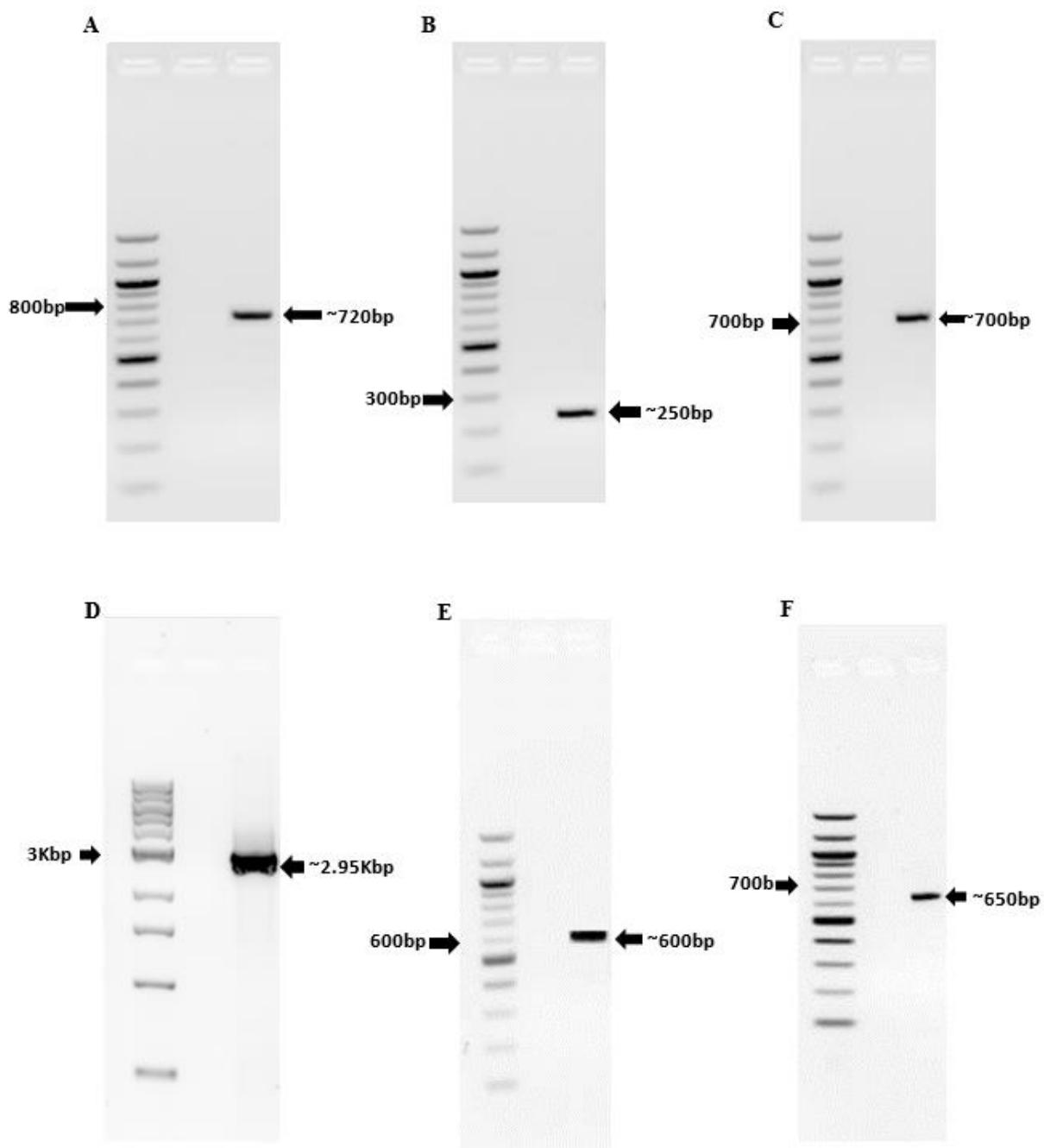


Figure 2.3.5: Gel images of short amplicons. The images are 1.5% Agarose gels showing PCR products of (A) COX1, (B) ND4, (C) IGS, (D) SIM4, (E) ITS2, and (F) 16S rRNA. In each gel, Lane 1 is a 100bp DNA Ladder (Invitrogen™) (except Lane 1 of gel D), Lane 2 is a No Template Control (NTC), and Lane 3 is an amplified *Simulium* blackfly gene from the transition zone of Ghana of the corresponding genomic region. Gel D was produced from 1.2% agarose. Lane 1 of gel D is a 1 kbp DNA ladder (New England Biolabs®).

2.3.3 Polymorphism and diversity assessment of the mitochondrial and nuclear amplicon sequences

Generally, there was high haplotype diversity and moderate nucleotide diversity across all the genes. The haplotype diversity in the mitochondrial genes ranged from 0.662, occurring in 16S-rRNA of Pru, to the highest value of 0.957, occurring in CO1 of Pru (Table 2.3.7). The average haplotype diversity was moderately high with a value of 0.838. The haplotype diversity in the nuclear genes had similar range with the lowest value of 0.714 occurring in IGS of Daka whiles the highest value of 0.917 occurred in IGS of Black Volta. The average haplotype diversity was equally moderately high with a value of 0.854. The nucleotide diversity on the other hand ranged from 0.0015, occurring in 16S-rRNA of Daka, to the highest value of 0.0126 that occurred in ND4 of Black Volta. The average nucleotide diversity was moderately low with a value of 0.0079. The nucleotide diversity in the nuclear genes was low with a range of 0.0026 occurring in IGS of Daka as the lowest value, and 0.0362 that occurred in ITS2 of Pru as the highest value (Table 2.3.8). In addition, the average nucleotide diversity was also moderately low with a value of 0.0142.

Table 2.3.7: Nucleotide sequence, polymorphism and genetic diversity for CO1, ND4 and 16S rRNA Genes

Gene	Community	Seq	N	S	H	Hd	K	π	M
CO1	Daka	31	550	26	19	0.95484	5.01290	0.00911	31
	Pru	31		32	19	0.95699	6.09462	0.01108	37
	Black Volta	31		28	22	0.95699	6.08817	0.01107	34
TDE	3 Populations	93		40	40	0.95208	5.67625	0.01032	46
				18	13	0.84731	2.88817	0.01229	19
ND4	Daka	31	235	12	13	0.89032	2.49892	0.01063	12
	Pru	31		15	12	0.90323	2.95914	0.01259	15
	Black Volta	31		22	22	0.87751	2.78495	0.01185	23
TDE	3 Populations	93		6	7	0.665	0.826	0.00149	6
				6	6	0.662	0.886	0.00160	6
16S-rRNA	Daka	31	555	8	9	0.716	0.968	0.00174	8
	Pru	31		12	13	0.669	0.881	0.00159	12
	Black Volta	31							

Seq = Number of Sequences used, N= Number of nucleotide sites, S = Number of polymorphic/indel/missing sites, H = Number of haplotypes, Hd = Haplotype diversity, K = Average number of nucleotide differences, TDE = Total Data Estimates, π = Nucleotide diversity, M = Total number of mutations

Table 2.3.8: Nucleotide sequence, polymorphism and genetic diversity for ITS2 and IGS

Gene	River Basin	Seq	N	S	h	Hd	K	π	M
ITS2	Black Volta	9	565	15	6	0.88889	2.50000	0.00448	9
	Daka	7		52	5	0.85714	18.09524	0.03229	53
	Pru	16		105	10	0.900	20.183	0.03617	121
TDE	3	32		114	17	0.871	15.192	0.02722	138
	Black Volta	9	625	10	7	0.917	2.389	0.00387	10
IGS	Daka	7		5	4	0.714	1.619	0.00262	5
	Pru	16		14	9	0.858	2.058	0.00333	14
TDE	3	32		24	15	0.827	2.044	0.00331	25

RB = River Basin, Seq = Number of Sequences used, N= Number of nucleotide sites, S = Number of polymorphic/indel/missing sites, h = Number of haplotypes, Hd = Haplotype diversity, K = Average number of nucleotide differences, TDE = Total Data Estimates, π = Nucleotide diversity (P_i), M = Total number of mutations.

2.3.4 Polymorphism and divergence between populations using the mitochondrial and nuclear amplicon sequences

Similar proportions of mutations were shared in the CO1 among the blackfly populations in the 3 river basins, with individual values of: 32.53% (27) shared in Pru and Daka; 33.73% (28) shared in Black Volta and Daka; and an equal proportion of 33.73% (28) mutations shared between the populations in Pru and Black Volta (Table 2.3.9). Although a relatively lower proportion of mutations were shared in the rest of the genes, ITS2 demonstrated unusually high range of shared mutations characterized by the lowest value of 2.17% (1) mutation shared among the populations of Black Volta and Daka whiles the highest value of 86.96% (40) was shared among the populations of Pru and Daka.

There was generally moderate nucleotide divergence occurring among populations in the 3 river basins. Although the average nucleotide divergence was 0.0106, the lowest value of 0.0015 was observed in 16S-rRNA between Pru and Daka whiles the highest value of 0.0349 was seen in ITS2 between Pru and Daka.

Table 2.3.9: Polymorphism and nucleotide divergence between populations

Gene	Population 1	Population 2	Dxy	Kp	Ms	Sub	1π2	1K2
CO1	Daka	Pru	0.01001	5.506	27	37	0.01108	0.01001
	Daka	Black Volta	0.00994	5.469	28	34	0.01107	0.00994
	Pru	Black Volta	0.01086	5.973	28	37	0.01108	0.01086
ND4	Daka	Pru	0.01130	2.656	10	19	0.01229	0.01130
	Daka	Black Volta	0.01251	2.940	12	19	0.01229	0.01251
	Pru	Black Volta	0.01176	2.764	8	12	0.01063	0.01176
16S-rRNA	Daka	Pru	0.00151	0.836	4	6	0.00160	0.00151
	Daka	Black Volta	0.00159	0.880	3	8	0.00174	0.00159
	Pru	Black Volta	0.00164	0.908	4	6	0.00160	0.00164
ITS2	Pru	Black Volta	0.02129	11.882	5	121	0.03617	0.02129
	Pru	Daka	0.03493	19.491	40	121	0.03617	0.03493
	Black Volta	Daka	0.02133	11.905	1	9	0.00448	0.02133
IGS	Pru	Black Volta	0.00352	2.174	3	14	0.00333	0.00352
	Pru	Daka	0.00302	1.866	1	14	0.00333	0.00302
	Black Volta	Daka	0.00321	1.984	1	10	0.00387	0.00321

Ms = Shared mutation, Dxy = Average number of nucleotide substitution per site between populations, Kp= Average number of nucleotide differences between populations; 1π2= Nucleotide Diversity (Total) between population 1 and 2; 1K2= Nucleotide Divergence (Total) between population 1 and 2; Sub= Total number of substitutions

2.3.5 Determination of population structure

The analysis of molecular variance across all the mitochondrial and nuclear genes from Table 2.3.10 showed that there was very little to none of the variation occurring among populations, with the majority to all of the variation occurring within populations. The variation among populations ranged from 0 to 2.5% whilst the variation within populations ranged from 97.5% to 100%. The measure of population differentiation due to genetic structure, the fixation index, across all the genomic sequences was approximately zero with no statistical

significance ($\Phi_{ST} p > 0.05$). The individual fixation indices however ranged from 0 to 0.025 across all the genes.

Table 2.3.10: Measures of population structure by analysis of molecular variance (AMOVA)

Gene	VAP	VWP	Φ_{ST}	$\Phi_{ST} p$
CO1	-1.09546	101.09546	-0.01095	0.763
ND4	0.97923	99.02077	0.00979	0.219
16S rRNA	-0.76116	100.76116	-0.00761	0.618
ITS2	0.49161	99.50839	0.00492	0.341
IGS	2.47804	97.52196	0.02478	0.22

VAP= Percentage of Variation among population; VWP= Percentage of Variation within population; Φ_{ST} = Fixation index; $\Phi_{ST} p$ = Significance of the Fixation index (1000 permutations).

2.3.6 Genetic and geographic distance assessment of the mitochondrial and nuclear amplicon sequences

The genetic distance between group mean of river basin population among the mitochondrial gene sequences of CO1, ND4 and 16Sr-RNA ranged from 0.0015 to 0.0208, while that for the nuclear gene sequences of ITS2 and IGS ranged from 0.0030 to 0.0464 (Table 2.3.11). The smallest genetic distance within group mean value of 0.0015 occurred in the 16S-rRNA of Daka river basin whiles the largest value of 0.0497 was found in the ITS2 of Pru river basin, leading to an average value of 0.0141 in all the genes studied across the 3 river basins (Supplementary Table B 2.3.11). In order of increasing magnitude, the geographic distances between paired sampling river basin locations for the mitochondrial gene sequences were 99.38 km, 115.95 km and 211.47 km for the respective paired locations of Pru and Daka, Pru and Black Volta, and between Black Volta and Daka. The pairwise geographic distances between sampling locations in the mitochondrial genes were similar to that of the nuclear genes. Thus, the geographic distances between paired sampling river basin locations for the nuclear gene sequences of ITS2 and IGS were 126.63 km, 126.99 km and 249.53 km for the respective paired locations of Pru and Daka, Pru and Black Volta, and between Black Volta and Daka. Although the geographic distance between Black Volta and Daka for the CO1 gene

(211.47 km) was greater than that between Pru and Daka (99.38 km), the same genetic distance between group means (0.0148) was observed for these two pairwise river basin sampling locations. Nevertheless, the geographic distance between the Pru and Black Volta river basins for CO1 gene (115.95 km), was lesser than that between Black Volta and Daka river basins (211.47 km) yet it yielded greater genetic distance of 0.0161 between the group mean of the former location and the latter (0.0148). Thus, the genetic distance and geographic distance for the CO1 gene were not correlated ($p > 0.05$). For the ND4 genomic region, the genetic distance between group means of the river basin populations showed a similar pattern in respect of magnitude to that of the geographic distance. Thus, in order of increasing magnitude, the genetic distances between the paired river basin locations for the ND4 gene were 0.0176, 0.0182 and 0.0208 for the paired locations of Pru and Daka, Pru and Black Volta, and between Black Volta and Daka. This resulted in a very strong positive correlation ($r = 0.999$, $p = 0.025$) between the group mean distance and geographic distance for ND4 (Table 2.3.12). In order of increasing magnitude, the genetic distances between group means for the 16S-rRNA gene were 0.00151 for Pru and Daka, 0.00159 for Black Volta and Daka, and 0.00164 for Pru and Black Volta. Thus, the genetic distance between group mean and geographic distance for the 16S-rRNA gene were not correlated ($p = 0.894$). Similarly, there were no correlation between the genetic distances and geographic distances for both ITS2 and IGS gene sequences ($p > 0.05$). In a nutshell, except for ND4, there was no correlation between the genetic distances and geographic distances for all the mitochondrial and nuclear genes studied ($p > 0.05$) as shown in Supplementary Tables B 2.3.10 and B 2.3.12.

Table 2.3.11: Genetic distance between group means with geographical distances

Gene	Location 1	Location 2	Distance (Between Group Mean)	Geographic Distance (Km)
CO1	Daka	Pru	0.0148	99.3800
CO1	Daka	Black Volta	0.0148	211.4700
CO1	Pru	Black Volta	0.0161	115.9500
ND4	Daka	Pru	0.0176	99.3800
ND4	Daka	Black Volta	0.0208	211.4700
ND4	Pru	Black Volta	0.0182	115.9500
16S	Daka	Pru	0.00151	99.3800
16S	Daka	Black Volta	0.00159	211.4700
16S	Pru	Black Volta	0.00164	115.9500
ITS2	Pru	Black Volta	0.0275	126.990000
ITS2	Pru	Daka	0.0464	126.630000
ITS2	Black Volta	Daka	0.0264	249.530000
IGS	Pru	Black-Volta	0.003515	126.990000
IGS	Pru	Daka	0.003021	126.630000
IGS	Black-Volta	Daka	0.003213	249.530000

2.3.7 Gene flow and genetic differentiation of the mitochondrial and nuclear amplicon sequences

The within population nucleotide diversity between river basins ranged from 0.86 to 19.58 with an average of 5.14 (Table 2.3.12). The lowest average number of nucleotide differences, a value of 0.84, occurred in 16S-rRNA between Pru and Daka while the largest value of 19.55 occurred in ITS2 between these same river basins. This supports evidence for the lack of population structure across all the mitochondrial and nuclear genes investigated.

Table 2.3.12: Gene flow and genetic differentiation with geographic distances

Gene	POP1	POP2	Hs	Ks	Kxy	GammaSt	Fst	Geographic Distance (Km)
CO1	Daka	Pru	0.9559	5.5538	5.5057	0.0121	-0.0087	99.3800
CO1	Daka	Black Volta	0.9559	5.5505	5.4693	0.0090	-0.0149	211.4700
CO1	Pru	Black Volta	0.9570	6.0914	5.9730	0.0066	-0.0198	115.9500
ND4	Daka	Pru	0.8688	2.6936	2.6556	0.0093	-0.0143	99.3800
ND4	Daka	Black Volta	0.8753	2.9237	2.9397	0.0191	0.0054	211.4700
ND4	Pru	Black Volta	0.8968	2.7290	2.7638	0.0227	0.0126	115.9500
16S	Daka	Pru	0.6634	0.8559	0.8356	0.0044	-0.0243	99.3800
16S	Daka	Black Volta	0.6903	0.8968	0.8803	0.0071	-0.0187	211.4700
16S	Pru	Black Volta	0.6893	0.9269	0.9084	0.0063	-0.0203	115.9500
ITS2	Pru	Black Volta	0.9407	13.8493	11.9097	0.0447	0.0456	126.9900
ITS2	Pru	Daka	0.9378	19.5826	19.5536	0.0519	0.0199	126.6300
ITS2	Black Volta	Daka	0.8757	9.3229	11.9048	0.1569	0.1350	249.5300
IGS	Pru	Black Volta	0.9833	2.1773	2.1736	0.0328	-0.0230	126.9900
IGS	Pru	Daka	0.9816	1.9246	1.8661	0.0474	0.0147	126.6300
IGS	Black Volta	Daka	1.0000	2.0521	1.9841	0.0588	-0.0100	249.5300

POP1 = Population 1 location, POP2 = Population 2 location, Hs = Average within-population diversity, Ks = within-population nucleotide diversity, kxy = the average number of nucleotide differences between population 1 and population 2, GammaSt = pairwise genetic distance, and Fst = fixation index.

2.3.8 Neutrality test of the mitochondrial and nuclear amplicon sequences

Across all the river basins, the Tajima's D test for all the mitochondrial and nuclear genes was below zero, thus suggesting a high frequency of rare alleles (Table 2.3.13). However, the lack of statistical significance in all the genes across all the river basins, with the exception of the nuclear genes for Pru river basin, indicated that the allelic distribution and/or level of variability did not violate the neutrality assumption. A follow up with both Fu & Li's D and F statistics showed that the ITS2 in Pru river basin also did not violate the neutrality assumption ($p > 0.05$). However, the IGS for Pru river basin was not evolving randomly ($p < 0.05$) as shown in Table 2.3.14 (see explanation in the discussion section).

Table 2.3.13: Tajima's test results for the amplicon sequences

Gene	Location	Tajima's D Test	
		D	p
CO1	Pru	-1.24647	> 0.10
	Black Volta	-1.03180	> 0.10
	Daka	-1.27473	> 0.10
ND4	Pru	-0.54462	> 0.10
	Black Volta	-0.70896	> 0.10
	Daka	-1.35228	> 0.10
16S-rRNA	Pru	-1.16076	> 0.10
	Black Volta	-1.55790	> 0.10
	Daka	-1.27425	> 0.10
ITS2	Pru	-1.93263	< 0.10
	Black-Volta	-1.12777	> 0.10
	Daka	-0.95920	> 0.10
IGS	Pru	-1.99160	< 0.10
	Black-Volta	-1.63274	0.10 > p > 0.05
	Daka	-1.02379	> 0.10

D= Tajima's D; p= Statistical significance; Time= Estimates of the Divergence Time

Table 2.3.14: Fu and Li's Test for the amplicon sequences

Gene	Location	FLD	p	FLF	p
CO1	Pru	- 0.63857	> 0.10	-0.98926	> 0.10
	Black Volta	- 0.04126	> 0.10	-0.43130	> 0.10
	Daka	- 0.43108	> 0.10	-0.83305	> 0.10
ND4	Pru	- 0.42448	> 0.10	-0.54149	> 0.10
	Black Volta	- 0.44608	> 0.10	-0.62125	> 0.10
	Daka	- 1.34490	> 0.10	-1.58442	> 0.10
16S-rRNA	Pru	- 1.91137	> 0.10	-1.96497	> 0.10
	Black Volta	- 3.15110	< 0.05	-3.11351	< 0.05
	Daka	- 2.69214	< 0.05	-2.64167	p < 0.05
ITS2	Pru	- 2.13769	0.10 > p > 0.05	-2.40456	0.10 > p > 0.05
	Black-Volta	- 1.27936	> 0.10	-1.38788	> 0.10
	Daka	- 1.11419	> 0.10	-1.05595	> 0.10
IGS	Pru	- 2.49021	< 0.05	-2.70833	< 0.05
	Black-Volta	- 1.73089	> 0.10	-1.90744	> 0.10
	Daka	- 0.96867	> 0.10	-1.06618	> 0.10

FLD= Fu and Li's D test statistic; FLF= Fu and Li's F test statistic; p= Statistical significance.

2.3.9 Assessment of the degree of genetic diversity

An independent sample t-test was used to ascertain if the nucleotide diversity of CO1 and those of the remaining short amplicon resequencing genes, long-range marker genes, whole mitochondrial genome, and the concatenated nuclear genes were statistically different (See Supplementary sheets 2 to 8). The results indicated that there was no statistical difference between the nucleotide diversity of CO1 and those of ND4 ($p = 0.189$), ITS2 ($p = 0.298$), long-range marker genes ($p = 0.894$), and whole mitochondrial genome ($p = 0.05$). However, the

independent sample t-test showed that there was a statistical difference between the nucleotide diversity of CO1 and those of the concatenated nuclear genes ($p = 0.005$), IGS ($p = 0.001$), and 16S-rRNA ($p = 0.005$).

2.3.10 Assessment of the relationships among defined DNA sequences and the relatedness of their genetic cluster units

Discriminant analysis of principal components (DAPC) assigns individual sequences to clusters based on their relationships to each other and can be carried out with or without a prior hypothesis of what those relationships might be. In *a priori* analysis, there is no prior hypothesis and individuals are assigned to clusters according to their genetic relationships alone. In *a posteriori* analysis, the individuals are first assigned to a particular group (usually indicated by colour coding in the plot) and then their positions relative to each other plotted according to their relationships. In both analyses, the closer two individuals are on the DAPC plot, the more closely they are related. In *a priori* assessment of genetic relatedness from the concatenated COX1, ND4 and 16S-rRNA gene sequences, 15 optimum principal components were chosen with 7 discriminant functions. This resulted in 8 clusters that formed 3 major groups as shown in Figure 2.3.6, A. Cluster 3 formed one major group, cluster 6 formed the second major group, while the rest of the clusters (1, 2, 4, 5, 7 and 8) formed the third major group. Thus, the cluster groups were assigned the names Group 1 for cluster 3, Group 2 for cluster 6, and Group 3 for the remaining clusters. The numbers of individual blackflies in each of the clusters were 8, 12, 7, 7, 19, 12, 6 and 22. The proportions of the blackflies in each of the three major cluster groups were 7.5%, 12.9% and 79.6% for Group 1, Group 2 and Group 3 respectively. One of the three groups may be genetically heterogeneous of the three major cluster groupings. The percentage of reassignment of individual blackflies to their original clusters indicated 100% of successful reassignment for all the clusters except for clusters 2, 5 and 7 that had 92%, 80% and 75% of successful reassignment respectively (Figure 2.3.6, B). Hence, all members of Group 1 and Group 2 were successfully assigned to their respective clusters 100% of the time. This suggested a complete distinction of Group 1 and Group 2 from each other and also from Group 3. The data suggested complete distinction among Group 1, Group 2 and Group 3.

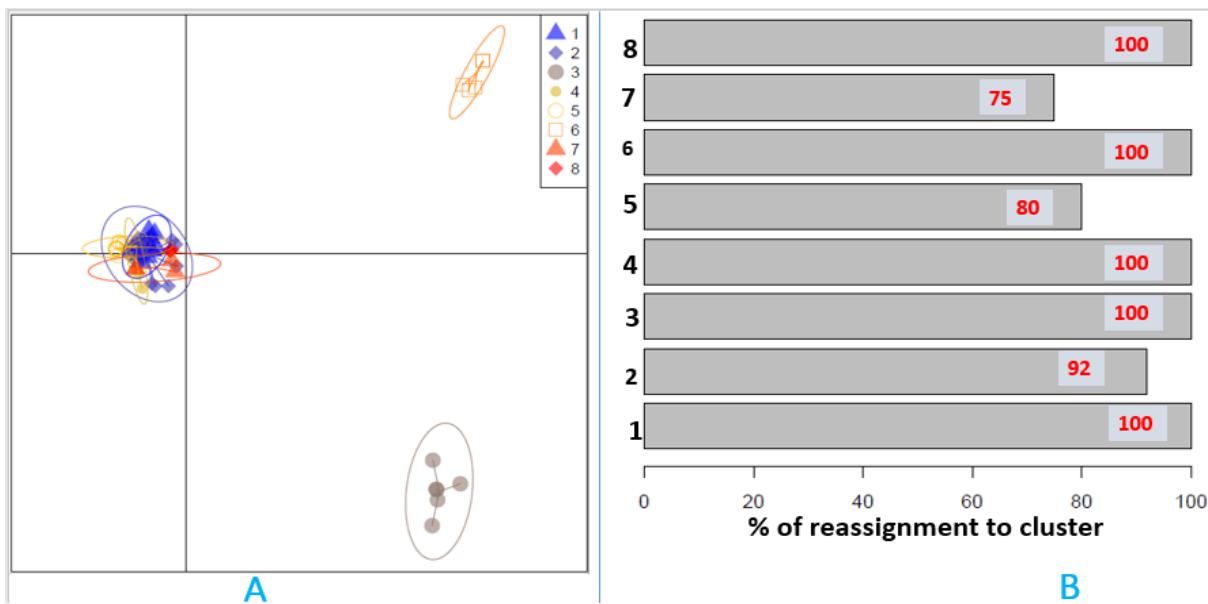


Figure 2.3.6: *A priori* assessment of genetic relatedness from concatenated CO1, ND4, and 16S-rRNA gene sequences without population assignment. A: DAPC plot, ellipses represent different genetic clusters; dots represent the individual blackflies. B: bar chart showing percentage of successful reassignment of individual blackflies to genetic clusters. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

In *a posteriori* assessment of genetic relatedness from the concatenated CO1, ND4 and 16S-rRNA gene sequences, individual blackflies were assigned to the original river basin of collection first, and then genetic differentiation between river basins analysed. The DAPC *a posteriori* assessment (Figure 2.3.7, B) showed that blackflies from all 3 river basins shared genetic characteristics, as indicated in the overlaps of the 3 river basin ellipses. The DAPC is supported by the haplotype analysis, which is another method of comparing genetic and geographic relationships. The haplotype network (Figure 2.3.7, A) shows 12 haplotypes with a frequency greater than one. Each of these haplotypes had members from all three river basins and very small mutation step differences between haplotypes. Thus, haplotypes from each river basin were closely linked, and distributed across all three river basins. This feature of close links between haplotypes of the river basins to one another, coupled with the shared haplotypes among all 3 river basins, suggested shared genetic characteristics. Although these tests do not directly quantify gene flow, the shared genetic characteristic most likely occurred as a result of gene flow and interbreeding of blackflies between river basins. There was reduction in the accuracy of reassignment of individual blackflies to the river basin from

which they were collected (Figure 2.3.7, C). This poor reassignment implies that genetic relatedness is correlated poorly with geographic origin.

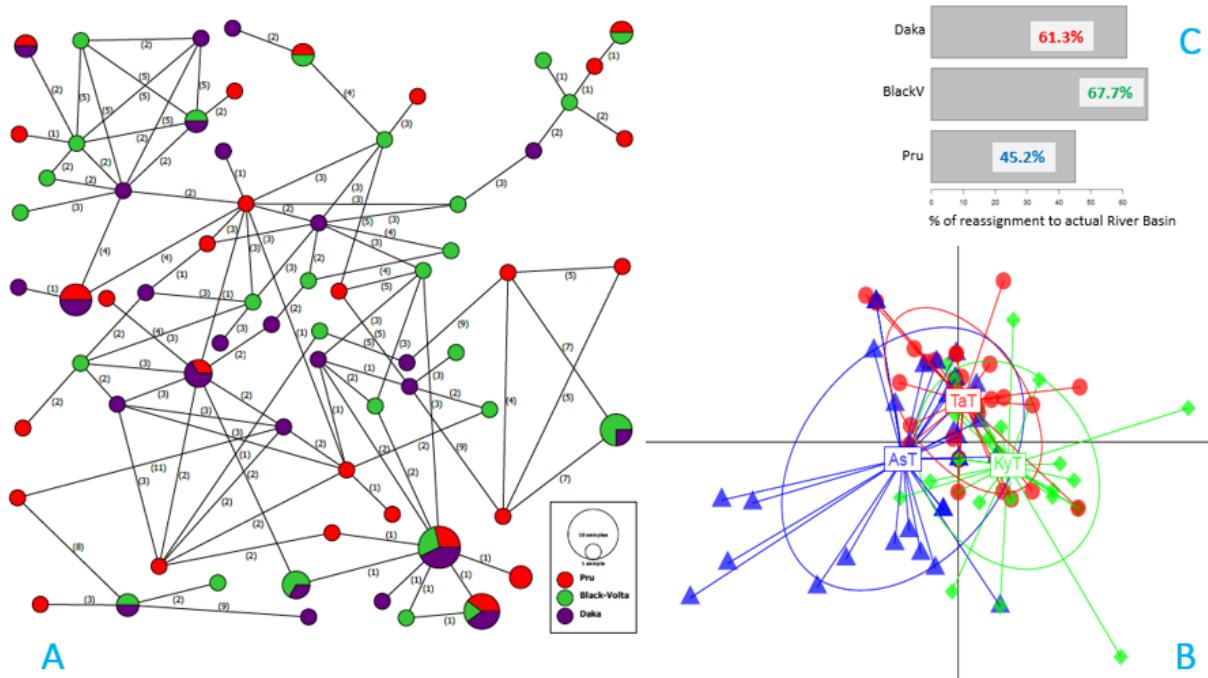


Figure 2.3.7: *A posteriori* assessment of genetic relatedness from concatenated CO1, ND4, and 16S-rRNA gene sequences after population assignment. Graph A is a haplotype network produced by minimum spanning network using Epsilon of 0 in Popart 1.7 (Bandelt *et al.*, 1999). Each circle in the haplotype network corresponds to one haplotype, and the size is proportional to its frequency among the samples. Colours of the circles correspond to river basin sampling locations. The numbers indicate the mutational step difference(s) between haplotypes. In section “B” of the figure, the ellipses represent the different river basins while the dots represent the individual blackflies collected. In the label of the ellipses, sample AsT = Pru river basin (blue), sample TaT = Daka river basin (red), and sample KyT = Black Volta river basin (green). Section “C” of the figure show the percentage of reassignment of individual blackflies to their original river basin of collection. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

The assessment of genetic relatedness in the 90 CO1, 90 ND4, 90 16S-rRNA, 32 ITS2, and 32 IGS mitochondrial and nuclear gene sequences using DAPC and haplotype networks showed similar pattern as in the concatenated gene sequences of CO1, ND4 and 16S-rRNA (Figures: 2.3.8 – 11, and Supplementary Figures C 2.3.12 - 17). There was the observation of

3 distinct blackfly molecular units and one of these may be genetically heterogeneous. Shared genetic characteristics characterized by shared haplotypes among all 3 river basins were observed and these indicated possible gene flow and interbreeding of blackflies among all 3 river basins.

From *a priori* assessment of the genetic relatedness using the 90 CO1 gene sequences without assigning any population attributes like river basin of collection (Figure 2.3.8, A and Supplementary Figure 2.3.16); 7 molecular clusters were observed, and these further formed 3 distinct groups. Five of the 7 clusters formed one large group whiles the remaining 2 clusters formed 1 small group each. Although there was clear separation between the 3 main molecular groups, members within a group did not show much variation from each other. This inadequate variability between members of a group made the clusters of the smaller group (5 and 6) overlay on each other to the point of looking like a single individual. The percentage of reassignment of individual blackflies to a cluster was 100% for clusters 1, 3, 5 and 6, thus supporting the distinctiveness of clusters 5 and 6 as independent groups which were separated from the remaining clusters.

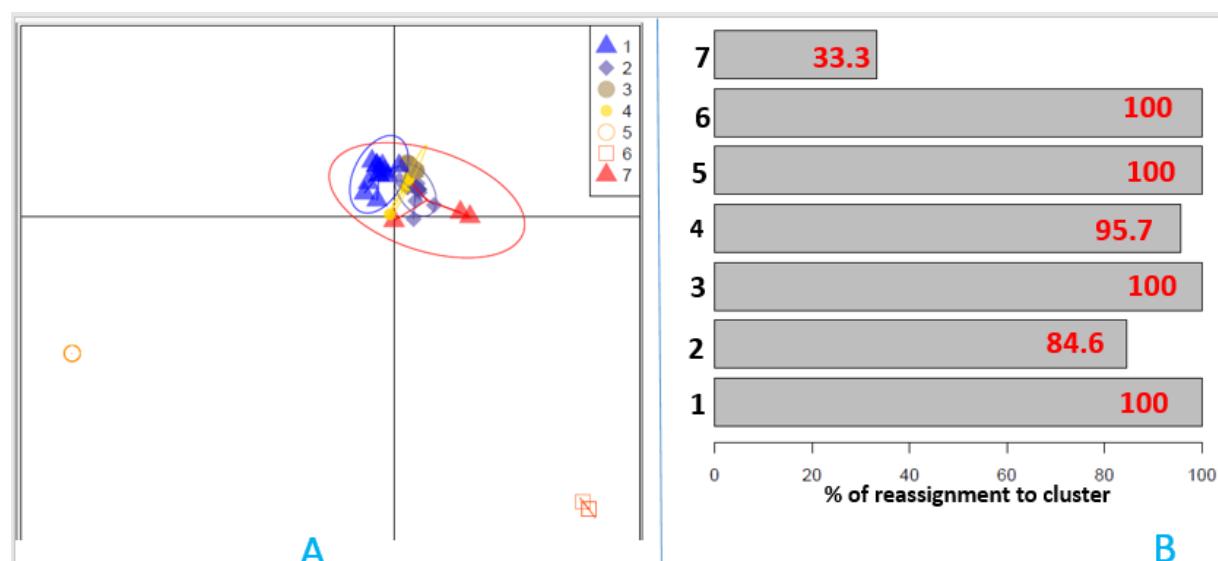


Figure 2.3.8: *A priori* assessment of genetic relatedness in 90 CO1 gene sequences without population assignment. In section “A” of the figure, the ellipses of the DAPC graph represent the different genetic clusters while the dots represent the individual blackflies collected. It was plotted by retaining 12 principal components, 7 clusters and 3 discriminant functions. In section “B” of the figure, the bar chat represents the percentage of successful reassignment of individual blackflies to their original genetic clusters identified. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

A posteriori genetic comparison of 90 blackflies using the CO1 gene from the 3 river basins of Pru, Black Volta and Daka were made and are shown in Figure 2.3.9. Section A of the figure showed 6 major haplotypes shared between blackflies in all 3 river basins. The rest of the haplotypes were either shared between 2 river basins or constituted a unique haplotype of a single river basin. The mutational step differences between haplotypes were very small, 1 to 3, in the majority of the haplotypes compared. The overlap of the 3 DAPC ellipses in section B of the figure support the observation of shared genetic characteristics among blackflies from all 3 river basins. The percentage of reassignment of individual blackflies to actual river basin in section C of the figure were all below 80% with individual values of 41.9%, 71% and 51.6% for Pru, Black Volta and Daka respectively. This suggested that the blackflies were not located in a single river basin all the time. Hence, an indication of movement between river basins.

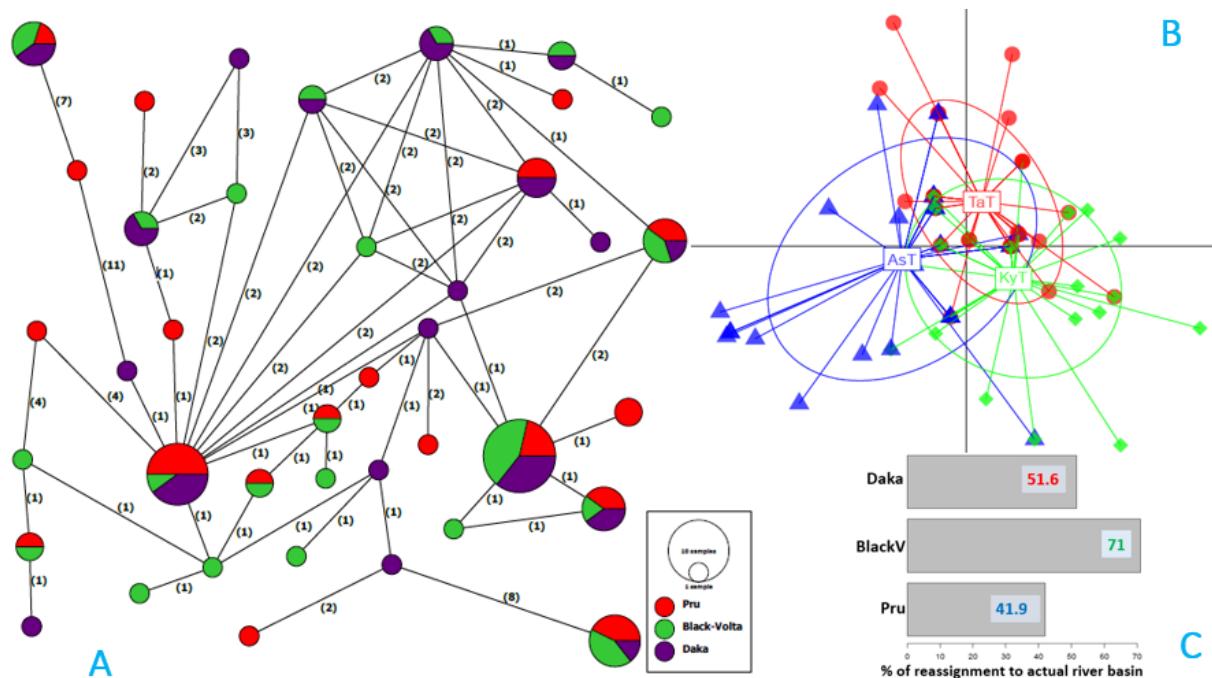


Figure 2.3.9: *A posteriori* assessment of genetic relatedness in 90 CO1 gene sequences after population assignment. A= haplotype network inferred with minimum spanning network using CO1 gene sequences from the river basins of Pru (red colour), Black Volta (green) and Daka (purple). This network was inferred using Epsilon of 0 in Popart 1.7 (Bandelt *et al.*, 1999). The circles (nodes) represent haplotypes and the size show the frequency and corresponding to the total number of individuals within the haplotype. The numbers indicate the mutational steps difference(s) between haplotypes. Section “B” of the figure shows the DAPC plot for the 90 CO1 gene sequences sampled from the river basins of Pru (AsT = blue

colour), Black Volta (KyT = green colour) and Daka (TaT = red colour). The ellipses represent the river basins while the dots represent the individual blackflies. Section C of the figure shows a bar chart of the percentage of reassignment of individual blackflies to their original river basin of collection. The DAPC graph and bar chart were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

A priori assessment of genetic relationship among the 90 blackflies was made by using 90 ND4 gene sequences without any form of population assignment (Figure 2.3.10). Four main clusters were observed with clusters 3 and 4 forming a single overlapping group while clusters 1 and 2 formed separate groups (see section A of Figure 2.3.10). Thus, the 4 clusters were genetically organized into 3 distinct molecular groups. This observation of the presence of 3 separate molecular group is supported by the 100% values of being able to reassign individual blackflies to clusters 1 and 2, with 100% and 87.5% of being able to assign the individual blackflies to clusters 3 and 4 respectively (see section B of Figure 2.3.10).

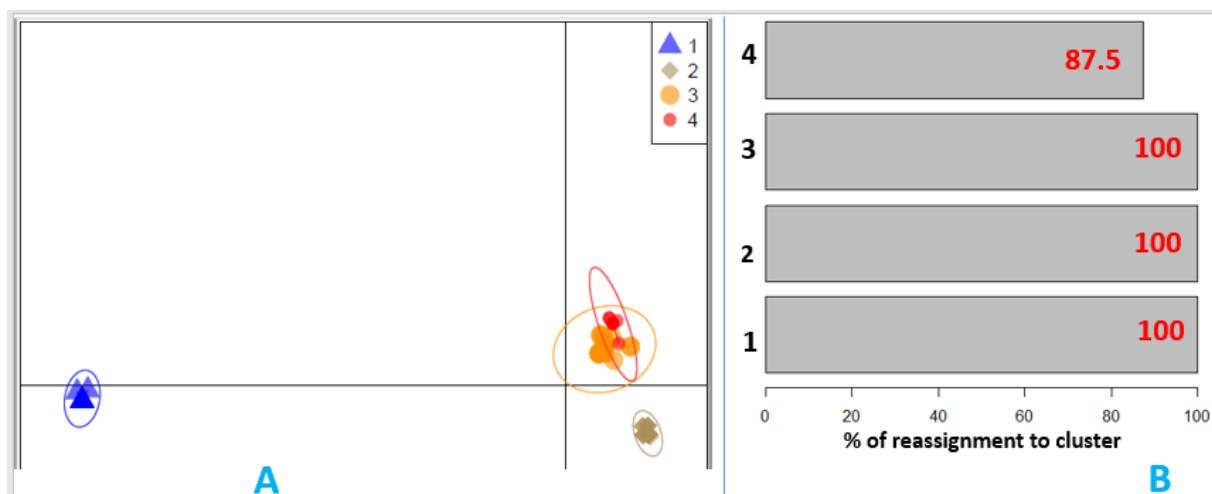


Figure 2.3.10: *A priori* assessment of genetic relatedness in 90 ND4 gene sequences without population assignment. In section “A” of the figure, the ellipses of the DAPC graph represent the different genetic clusters while the dots represent the individual blackflies collected. It was plotted by choosing and retaining 9 principal components, 4 clusters and 2 discriminant functions. The use of 3 discriminant functions yielded similar results as 2. In section “B” of the figure, the bar chart represents the percentage of successful reassignment of individual blackflies to their original genetic clusters identified. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

A posteriori comparison was made of the genetic relationship between blackflies in the 3 river basins using 90 ND4 gene sequences (Figure 2.3.11). In section A of the figure, 4 major haplotypes were shared by blackflies in all 3 river basins with very small mutational steps differences between them (average of 1). The remaining haplotypes were either shared between blackflies in 2 river basins or were unique haplotypes of single river basins. From the 3 overlapping ellipses of the DAPC in section B of the figure, individual blackflies from one river basin shared genetic characteristics with those of the remaining river basins. The percentage of reassignment of blackflies to the river basin of collection yielded values less than 80% with individual values being 25.8%, 48.4% and 77.4% representing Pru, Black Volta and Daka respectively (see section C of figure 2.3.11).

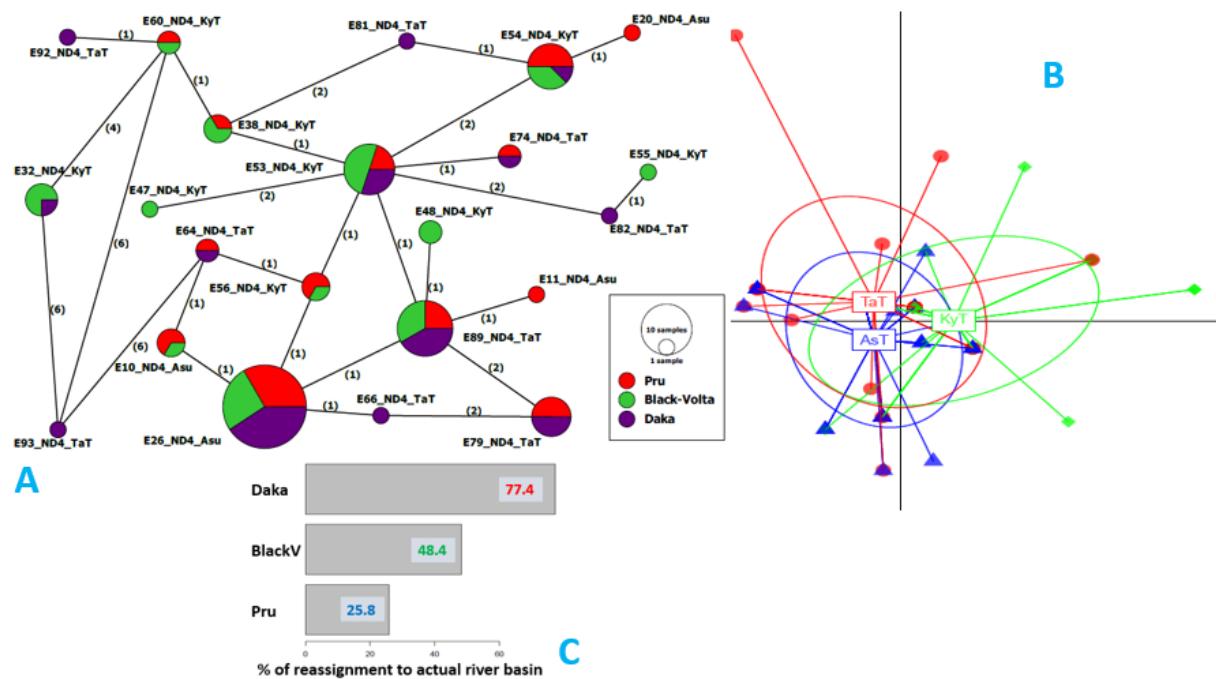


Figure 2.3.11: *A posteriori* assessment of genetic relatedness in 90 ND4 gene sequences after population assignment. In the figure, “A” = Haplotype network produced by using Epsilon of 0 in Popart 1.7 (Bandelt *et al.*, 1999) and inferred from 90 ND4 gene sequences from the river basins of Pru, Black Volta and Daka. The circles (nodes) represent haplotypes and the size show the frequency corresponding to the total number of individuals within the haplotype. The numbers indicate the mutational steps difference(s) between haplotypes. Section “B” of the figure shows the DAPC plot for the 90 ND4 gene sequences sampled from the river basins of Pru (AsT = blue colour), Black Volta (KyT = green colour) and Daka (TaT = red colour). The ellipses represent the river basins whiles the dots represent the individual blackflies. Section “C” of the figure shows a bar chart of the percentage of reassignment of

individual blackflies to their original river basin of collection. Graphs B and C were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

In summary, the DAPC and haplotype analyses of either concatenated or individual mitochondrial amplicons do not support population structure among blackflies according to their river basin of origin. The *a priori* analyses do, however, indicate that genetic structure does exist, separating the blackflies into 3 distinct groups. The correlation between these groups and river basins is not tested explicitly in these analyses, but the combined *a priori* and *a posteriori* analyses suggest that it is unlikely that there is a correlation between genetic relatedness and river basin of origin.

DAPC is a multivariate method that identifies and describes genetically related individuals, either in *a priori* clusters (groups) without prior knowledge about the group, or in *a posteriori* groups of known attributes. Haplotype network is a widely used method for analysing and visualizing the relationships existing among DNA sequences within a population or species. While DAPC and haplotype analyses are explicit tests of population structure and genetic relationships, phylogenetic analyses are based on hypotheses of evolutionary relationships.

The phylogenetic analysis showed that the blackflies from Ghana were monophyletic to the blackflies from either Central or South-Eastern Africa (Figure 2.3.12). All the African blackflies formed a monophyletic group to those from Asia and Europe. On a geographical large-scale comparison, there were strong bootstrap supports at the divergent points separating the regions of Western, Central, and South-Eastern Africa blackflies from each other and from those of Asia and Europe. However, from geographically small-scale comparison, there were poor bootstrap support separating clades from within each of the regions at the country level. This might get resolved by sequencing more genomic data and also by using larger sample size. The tree with the highest log likelihood (-4987.34) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.6223)). The rate variation

model allowed for some sites to be evolutionarily invariable ([+I], 38.93% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 99 nucleotide sequences. There were a total of 1341 positions in the final dataset.

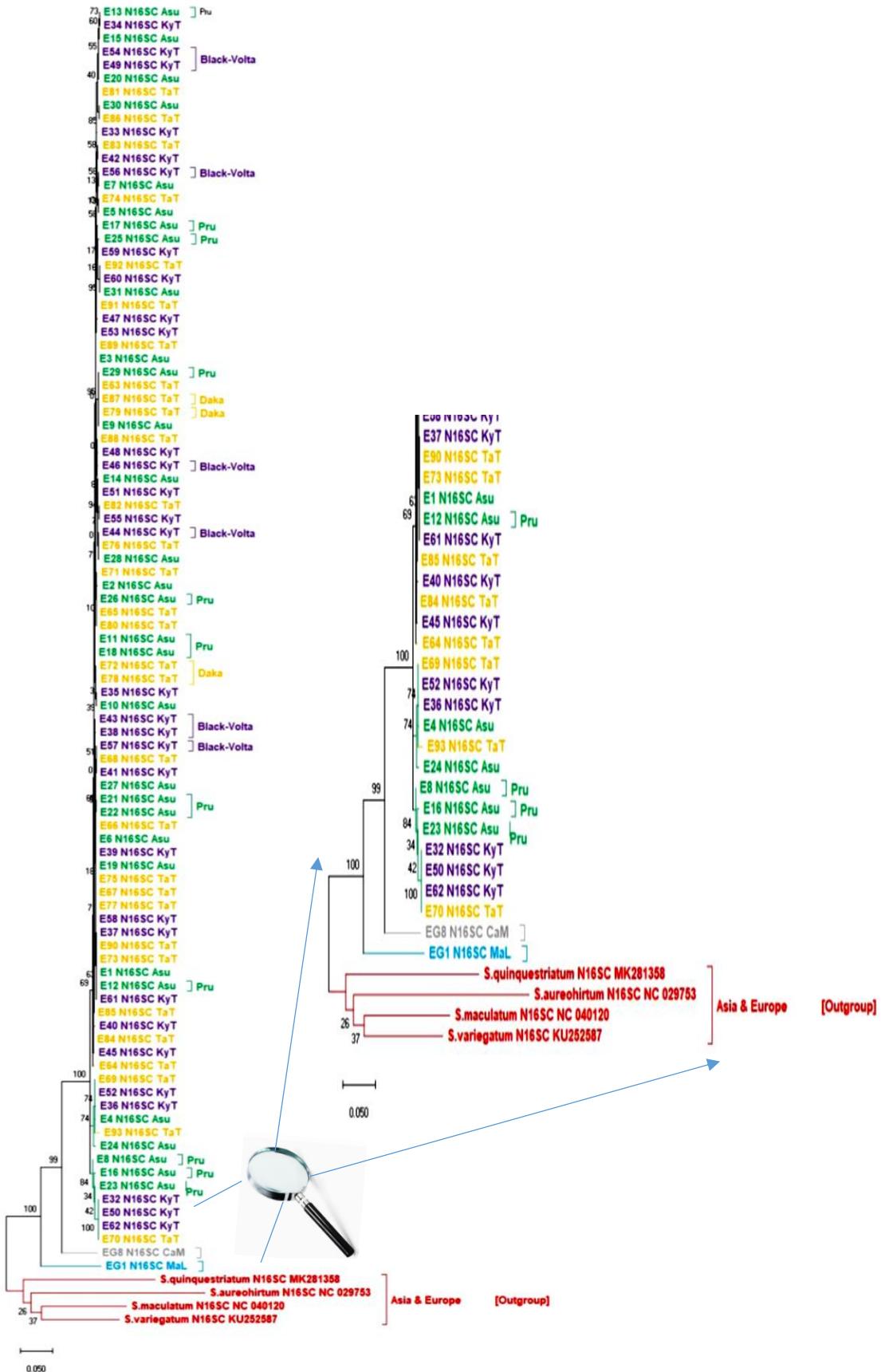


Figure 2.3.12: Phylogenetic tree from concatenated CO1, ND4, and 16S rRNA gene sequences after population assignment. The figure is a maximum likelihood tree showing the relationship among 93 concatenated CO1, ND4 and 16S rRNA gene sequences from the transition ecological zone of Ghana; one sequence each from Cameroon and Malawi; and 4 sequences from Asia and Europe. The phylogenetic tree is colour coded by the river basin or locality of collection: green = Pru river basin in West Africa; purple = Black Volta river basin in West Africa; yellow = Daka river basin in West Africa; blue = Lingoni Falls in South-Eastern Africa; grey = Mbam river basin in Central Africa; red = outlier from Asia and Europe. The scale bar (0.05) shows the average substitution per site. The values near the branches show the bootstrap supports. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model (Tamura, 1992). Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018).

2.4. Discussion

2.4.1 Meta-analysis, polymorphism and genetic diversity

From the neutrality test of the intraspecies, *Entomyzon cyanotis* and *Chlamydosaurus kingii* were evolving randomly (Tajima's $p > 0.05$). However, the use of Fu's tests showed that there was selection due to hitchhiking (Fu's $p < 0.05$). The neutrality test of the *Variabilichromis moori* was however inconclusive as the p -values of the Tajima's D and Fu's F were both not statistically significant ($p > 0.05$) but that of the Fu's D was statistically significant. This could be attributed to uneven sample size of the genetic groupings within the species.

Among the *Simulium* CO1 (Set 2), there was high nucleotide diversity ($\pi = 0.12906$) and high haplotype diversity ($Hd = 0.9847$). A possible reason could be because of a large population size of the blackflies that allowed the accumulation of individual mutations. Despite this, many of the clades of the phylogenetic tree (Figure 2.3.2) showed polytomies. Either there was not sufficient genetic data to make this inference or the polytomies could have resulted from adaptive radiation of the blackflies that does not allow sufficient lineage sorting of the flies. This will need to be tested by carrying out further sequencing to generate more data.

2.4.2 Assessment of degree of genetic diversity

The statistical difference between the nucleotide diversity of CO1 and those of 16S-rRNA could be attributed to the 16S-rRNA gene not being very informative. This was because, it had very few polymorphic sites (6) in both Pru and Daka, and 8 in Black Volta (See Table 2.3.13). On the other hand, the number of polymorphic sites in the CO1 gene was 26, 32, and 28 in Daka, Pru and Black Volta respectively. These led to a total number of polymorphic sites of 12 in the 16S-rRNA gene while in the CO1 gene it was 40. Such variation possibly contributed to the observation of statistical difference between them. All the measured parameters of polymorphism and genetic diversity were consistently lower in the 16S-rRNA gene while the corresponding values in the CO1 gene were all very high. For instance, the total number of mutations across all the 3 river basins using the 16S-rRNA gene was 12, while that of the CO1 gene was 46. These consequently produce very low overall nucleotide diversity value of 0.00159 for the 16S-rRNA gene, while that of the CO1 gene produce a relatively higher value of 0.01032 also across the 3 river basins. Similarly, the IGS gene was not very genetically informative and it was characterized by very low total numbers of polymorphic sites (24), haplotypes (15), mutations (25), average number of nucleotide

differences (2), and nucleotide diversity (0.00331). The concatenated nuclear genes had similar values to that of the CO1 gene for most of the polymorphism and genetic diversity parameters measured. A major contributing factor to the variation in the nucleotide diversity could be because of the vast variation in the overall number of haplotypes (CO1=40; Nuclear genes = 72) coupled with cumulative effect of the slight variations of the measured parameters between CO1 and the concatenated nuclear genes. These ultimately produced a moderate nucleotide diversity value of 0.00102 in the concatenated nuclear genes, while that in CO1 was relatively larger ($\pi = 0.01032$).

The lack of statistical difference in the nucleotide diversity between the CO1 gene and the majority of the genes assessed possibly suggest that the diversity in the CO1 gene could represent the diversity of the genes in the other parts of the genome, more especially, the mitochondrial genome. By adopting the threshold of nucleotide diversity proposed by Goodall and colleagues (2012), a mean value of 0.00388, it will imply that the nucleotide diversity of African *Simulium* blackflies is relatively larger than those of related organisms.

2.4.3 Gene flow, population structure and implications

The majority of the blackflies' genetic variation occurred within populations (97.5- 100 %, $p > 0.001$), with the remaining variation occurring among populations (0-2.5%, $p > 0.001$). This also suggests a lack of geographic population structure by river basin. The test for the overall measurement of population differentiation due to genetic structure, the overall Fixation Index (F_{ST}), was small (0-0.025) and close to zero. This small fixation index value closer to zero was similar to that of a panmictic population with random mating and greater gene flow between river basins. The force of gene flow in the population for a given generation (Nm), that influences the tendency of population differentiation because of genetic structure, was however large. Both the negligible fixation index and the large Nm that is greater than 1 (Wright, 1951) suggest that there is a lack of geographic population structure in the river basins across the central ecological transition zone of Ghana.

The presence of gene flow between blackflies in all the 3 river basins indicate that if suboptimal response of *O. volvulus* to ivermectin develops at any particular location, it can be transmitted to other locations. Communities at any particular river basin cannot stop ivermectin treatment without the risk of blackflies moving from an infected area in the transition ecological zone to re-infect people in the already treated area.

2.4.4 Decision on general aims and hypothesis

Generally, there was moderate nucleotide diversity but high haplotype diversity across all groups of the studied organisms. The total nucleotide diversity ranged from 0.00463 to 0.12906 with an average of 0.049503. Also, the total haplotype diversity ranged from 0.56061 to 0.9998 with an average of 0.898233. Thus, meta-analysis enabled the determination of the degree of genetic diversity in populations and species of the studied organisms. Therefore, the null hypothesis 1 was accepted and the alternative hypothesis 1 was rejected.

The haplotype diversity in the mitochondrial genes ranged from 0.662 to 0.957, with an average of 0.838. The haplotype diversity in the nuclear genes ranged from 0.714 to 0.917, with an average of 0.854. The nucleotide diversity in the mitochondrial genes ranged from 0.0015 to 0.0126, with an average of 0.0079. The nucleotide diversity in the nuclear genes ranged from 0.0026 to 0.0362, with an average of 0.0142. Thus, the genetic diversity of *Simulium* blackflies in the central ecological transition zone of Ghana were characterized by using CO1, ND4, 16S, ITS2 and IGS genes. Therefore, the null hypothesis 2 was accepted and the alternative hypothesis 2 was rejected.

It was observed that there was no population structure of the blackflies by river basin, and this was characterized by blackfly movement and gene flow between river basins. The entire central ecological transition zone of Ghana constitutes a single onchocerciasis transmission zone. Since the commonly used short amplicon barcoding genes of CO1, ND4, 16S, ITS and IGS enabled the determination of the transmission zone status of the study area, the null hypothesis 3 was accepted and the alternative hypothesis 3 was rejected.

The general aim of the study was achieved by characterizing the genetic diversity of the *Simulium* blackflies, determining their population structure and using it to determine the transmission zone status of the study area.

2.5 Conclusion

Within the *Simulium* blackflies, there was high diversity in the mitochondrial genes and moderate diversity in the nuclear genes. My data showed similar levels of diversity as the ones from the NCBI database. Comparatively, there is a high diversity of the blackflies in Africa than realized in other related species. Nearly all the diversity is within rather than between groups. The data showed that the *Simulium* genetics do not correlate with river basin type of transmission as currently assumed by onchocerciasis control programmes. The results point to what appears to be a strong relationship between diversity and geographical distance over large distances but very low differentiation within countries, and most likely a complex of either very closely related species, or populations within species, that are poorly differentiated and between whom gene flow may continue.

The presence of gene flow between blackflies in the 3 river basins, coupled with the lack of population structure by river basin suggest that the whole central transition ecological belt of Ghana may constitute a single onchocerciasis transmission zone. More data, possibly long-range sequences, will be needed to verify this inference and to find out if the observed phylogenetic polytomies resulted from phenomena like adaptive radiation of the blackflies or from inadequate genetic data.

Chapter 3: Long-Range amplification of *Simulium* genes

3.1: Introduction

Several attempts were made in the previous chapter to sequence the complete mitochondrial genome by PCR by using long-range amplification techniques. Upon numerous attempts, many of the primers either failed or amplified non-specific targets whose size or gene sequence were different from the primary mtDNA loci of interest. Through primer pairing process, a sense and anti-sense primer initially being part of independent primer pairs that were originally designed to amplify large chunks of the mitogenome became successful at amplifying the mtDNA loci used in the chapter three study. Validation of the sequence content and size by Sanger sequencing confirmed the mtDNA loci to be the sequence region between the tRNA-Gln and COX2 of the *Simulium* mitogenome, just as it was calculated and predicted. This relatively long-range amplicon was generated across all the target samples in the methodology alongside further attempts to amplify the complete mitogenome. The main purpose of exploring this tRNA mtDNA and spacer DNA loci was to attempt to increase the confidence in the inference made in the amplicon sequencing of the chapter two of this study.

This section reviewed the literature of mitochondrial ND2, CO1 and 4 tRNAs found between tRNA-Gln and COX2 in the *Simulium* mitochondrial genome, their role in characterizing genetic diversity, and their contribution in defining onchocerciasis transmission zone. These genes targeted for amplification in African *Simulium* blackflies were found to be above 1000 bp in size in *Simulium maculatum* (Wang and Min, 2019) and also in both *Simulium aureohirtum* and *Simulium variegatum* (accessions NC_029753.1 and NC_033348.1). For this thesis, the amplicons of these genes that were above 1000 bp in size were referred to as Long range amplicons. The estimated size of these genes was approximately 3 kbp. The genes were tRNA-Met, ND2, tRNA-Trp, tRNA-Tyr, tRNA-Cys, and COX1.

3.1.1 Structure and function of tRNA and associated genes

A transfer RNA (tRNA) is generally accepted as an adaptor molecule made up of RNA of about 75 to 95 nucleotides in length and that act as temporary carriers of amino acids to bring the appropriate amino acids to the ribosome based on the sequence of the mRNA (Sharp *et al.*, 1985). All tRNAs possess a secondary structure that consist of three hairpin loops and a terminal helical stem (cloverleaf), and these fold into an L-shaped tertiary structure. The focal functional parts in tRNA are the anticodon triplets that read the messenger RNA (Crick, 1968; Sprinzl and Vassilenko, 2005; Goodenbour and Pan, 2006). Thus, it performs a core

function of being the physical link between the amino acid sequences of protein and mRNA. During the process of a nucleotide sequence of an mRNA determining which amino acids get incorporated into a protein product of a gene from which the mRNA is transcribed, tRNA performs the role of determining which genetic code sequence corresponds to that particular amino acid (Crick, 1968).

Structurally, tRNA's are categorized into primary, secondary or tertiary structures with their characteristic L-shaped 3D structure that enable them to fit into the A and P sites of ribosomes. Although all tRNA's share this similar L-shaped 3D structure, they show variation in both the loop 'diameter' and the lengths of each arm in different species (Goodenbour and Pan, 2006; Itoh *et al.*, 2013). Generally, tRNA's possess different structural components. These include a 5'-terminal phosphate group, a cytosine-cytosine-adenine (CCA) tail and an arm consisting of a stem and a loop. The acceptor stem consists of 7 to 9 bp formed from the base pairing of the 5'- and 3'-terminal nucleotides. It may contain non-Watson-Crick base pairs. The arms of the tRNA are the Acceptor arm, Variable arm, Anticodon arm, DHU or D arm and the T ψ C or T arm (Jahn *et al.*, 1991; Itoh *et al.*, 2013).

The aminoacyl tRNA synthetases (AATS) catalyse the loading of amino acid onto tRNA's to form the aminoacyl-tRNA. This conjugate becomes covalently bonded to the 3'-hydroxyl group on the CCA tail and besides its sequence being necessary for enzymes to recognize tRNA's, it is of vital importance to translation (Sprinzel and Cramer, 1979; Green and Noller, 1997; Ibba and Söll, 2000).

Functionally, the D-arm is a highly variable region with the important role of stabilizing the tertiary structure of RNA's as well as influencing the accuracy and kinetics of translation at the ribosome. The T arm specifies the presence of pseudouridine, thymidine and cytidine residues, thus being involved in the interaction of tRNA's with ribosomes. The variable arm has a role in recognizing AATS of tRNA in some species. Interestingly, tRNA and tRNA-like aggregates are known to have a catalytic influence on replication, hence making them to be referred to as 'molecular or chemical fossils' of the RNA world (Maizels and Weiner, 1999; Itoh *et al.*, 2013).

The mitochondrially encoded nicotinamide adenine dinucleotide hydride (NADH) dehydrogenase 2 (ND2) is a protein that in most eukaryotes like humans is encoded by the ND2 gene and functions in proton translocation (Birrell and Hirst, 2010; Efremov *et al.*, 2010). The ND2 protein is found in the mitochondrial inner membrane and it is believed to be

the largest among the 5 complexes of the electron transport chain (Voet *et al.*, 2013). Various studies have shown that variants of ND2 are associated with diverse health issues including Leigh syndrome, mitochondrial encephalomyopathy with stroke-like episodes (MELAS), Leber's hereditary optic neuropathy, an increase in adult BMI, obesity, diabetes and hypertension (Flaquer *et al.*, 2014; La Morgia *et al.*, 2014; Thorburn *et al.*, 2017). Being one of the 7 mitochondrially-encoded subunits of NADH dehydrogenase enzyme, ND2 is the largest of the respiratory complexes. Its structure is observed as an L-shape with a long, hydrophobic transmembrane domain as well as a hydrophilic domain for the peripheral arm. Together with the rest of the mitochondrially encoded subunits, ND2 forms the core of the transmembrane region and is the most hydrophobic of the Complex 1 subunit (Voet *et al.*, 2013). Being a subunit of the respiratory chain Complex 1, ND2 is part of the minimal assembly of proteins needed to catalyze electron transfer and NADH dehydrogenation to coenzyme Q10 or ubiquinone despite the realization that only 3% of mitochondrial protein is attributable to complex 1 (RAGAN, 1987).

Cytochrome C Oxidase is known to be a complex metalloprotein with the provision of key functions in cellular respiration in eukaryotes and prokaryotes. Cytochrome C Oxidase 1 is called COX1, CO1 or COI in other eukaryotes (Kosakyan *et al.*, 2012). There are 3 mitochondrial DNA encoded subunits of the respiratory complex IV and CO1 is one of them. The third and final enzyme of the electron transport chain of mitochondrial oxidative phosphorylation is the complex IV. Together with COXII and COXIII, COX1 serve as a catalytic core of eukaryotic enzymes (Capaldi, 1990). Subunits I, II, and III are hydrophobic polypeptides in both eukaryotes and prokaryotes. COX 1 has 12 putative transmembrane (α -helical) regions in different organisms (Capaldi *et al.*, 1986; Raitio *et al.*, 1987). Whereas the orientation of some of the subunits of Cytochrome C are known, that of COX1 is yet to be determined with certainty. For instance, while the arrangement of COXII in Cytochrome C is such that both of its N and C termini are on the C side of the mitochondrial inner membrane, that of COXIII is arranged such that it is folded with its loop containing CYS₁₁₅ on the C side (Malatesta *et al.*, 1982; Capaldi, 1983; Capaldi, 1990). Within COX1 are found 7 conserved His but no conserved Cys. In addition, hemes and coppers can be localized to it (Ludwig and Schatz, 1980; Ludwig, 1987). Mutations in CO1 are associated with colorectal cancer, sensorineural deafness, acquired idiopathic sideroblastic anemia, recurrent myoglobinuria, complex IV deficiency, and Leber's hereditary optic neuropathy (Brown *et al.*, 1992; Gattermann *et al.*, 1997; Bröker *et al.*, 1998; Pandya *et al.*, 1999; Varlamov *et al.*, 2002; Greaves *et al.*, 2006; Lucioli *et al.*, 2006; Namslauer and Brzezinski, 2009).

Molecular tools are needed to aid the monitoring and evaluation of onchocerciasis elimination, especially in the characterization of genetic diversity of *Simulium* blackflies and the identification of the boundaries of their migration. Such efforts can be achieved with mitochondrial genes, especially tRNA-Cys, tRNA-Tyr and CO1 that have been used widely for species identification and phylogenetic work on closely related species, including blackflies (Hoy, 2002; Hebert and Gregory, 2005; Savolainen *et al.*, 2005; Ilmonen *et al.*, 2009; Rivera and Currie, 2009; Pramual *et al.*, 2010).

3.1.2 Justification (Rationale) of the study in chapter three

The diversity of the short amplicon barcoding genes from chapter two was generally similar or higher than the diversity in most species. The phylogenetic tree showed discrimination of the blackflies between countries but lack of differentiation within country. Although the *a priori* DAPC analysis demonstrated the presence of 3 distinct molecular groups, the data further demonstrated the lack of population differentiation by river basin. Although the fixation index, analysis of molecular variance and the haplotype network together suggested the lack of population structure, there were very short branches in the phylogenetic trees. It was therefore not very clear whether the lack of population differentiation by river basin was because of the absence of structure by river basin or because of inadequate genetic data to detect one. It became necessary to attempt to generate more molecular data to test this assumption. For this test to be feasible, an attempt was made to generate larger size sequences greater than 1000 bp.

3.1.3 Specific objectives, mode of assessment and hypotheses.

3.1.3.1 Specific objectives

The specific objectives of this chapter are:

1. To characterize the genetic diversity in tRNA rich sections of the *Simulium* mitochondrial genome in the transition zone of Ghana and assess if it is the same as the genetic diversity in other genes
2. To use tRNA-rich sections of the mitochondrial genome to determine if there is migration of *Simulium* blackflies between river basins in the central ecological transition zone of Ghana

3.1.3.2 Mode of assessment of specific objectives

1. Neutrality will be determined with Tajima's D test. Fu and Li's tests will be used to augment Tajima's D test in determining, if any, departures from neutrality due to genetic hitchhiking and population expansion (Fu, 1997; Zeng *et al.*, 2006).
2. Polymorphism and genetic diversity will be determined using the total number of mutations, the number of polymorphic (segregating) sites, total number of singleton mutations, number of haplotypes, haplotype diversity, nucleotide diversity and the average number of nucleotide differences.
3. The estimation of gene flow and genetic differentiation within river basins will be carried out using the average within-population diversity (H_s), within-population nucleotide diversity (K_s), average number of nucleotide differences between observed genetic groups (K_{xy}), pairwise genetic distance (GammaSt), and the fixation index (F_{st}). To augment the population structure status inferred with the fixation index, the analysis of molecular variance (AMOVA) test will be used to serve as a confirmation of the overall population differentiation status by determining the proportion of variation within and between observed genetic groups (Excoffier *et al.*, 1992). In all cases, if the fixation index estimation is accurate, it should be reflected in the corresponding value of the force of gene flow in the population for the given generation (Nm). Thus, the fixation index and the force of gene flow should be concordant. Generally, if $Nm < 1$, it will lead to local differentiation of populations; and if $Nm > 1$, then it will lead to little or no differentiation among populations (Wright, 1951; McDermott and McDonald, 1993).
4. The nucleotide diversity and haplotype diversity in African *Simulium* blackflies are expected to fall in the range of similar index values in related organisms. Using the proposed threshold for inferring the degree of nucleotide diversity and haplotype diversity in CO1 sequences; (a) haplotype diversity greater than a mean of 0.63388 or a median of 0.70130 will represent a high haplotype diversity, whiles haplotype diversity lower than these values will represent low haplotype diversity; and (b) nucleotide diversity greater than a mean of 0.00388 or a median of 0.00356 will represent high nucleotide diversity, whiles nucleotide diversity lower than these values will represent low nucleotide diversity (Goodall-Copestake *et al.*, 2012).
5. If there is migration of blackflies between river basins: (a) the fixation index is expected to be closer to zero, with the closeness to zero of the fixation indices reflecting the degree of migration and extent of lack of population differentiation such

that a zero value will indicate complete panmixia; (b) either all or the majority of genetic variation should occur within genetic groups with little or none occurring between the genetic groups. The reverse of these processes is expected to be true for the non-existence of migration between blackflies of different river basins such that the closeness of the fixation index is to 1, the more likely that the groups do not migrate but will show population differentiation due to genetic structure until a value of 1 is reached for a fixed population.

3.1.3.3 Hypotheses of the study in this chapter

Null hypothesis 1

The genetic diversity in the tRNA-rich section of the Ghanaian *Simulium* mitochondrial genome will be similar to those in the other genes in other parts of the blackfly

Alternative hypothesis 1

The genetic diversity in the tRNA-rich section of the Ghanaian *Simulium* mitochondrial genome will be different from those in the other genes in other parts of the blackfly

Null hypothesis 2

The tRNA-rich section of the mitochondrial genome will demonstrate migration of *Simulium* blackflies between river basins in the central ecological transition zone of Ghana

Alternative hypothesis 2

The tRNA-rich section of the mitochondrial genome will not demonstrate migration of *Simulium* blackflies between river basins in the central ecological transition zone of Ghana

3.2: Materials and Methods

3.2.1 Sample collection

A total of 90 different adult *Simulium* blackflies were collected from the same sampling locations and by the same method of collection and processing as the ones in section 2.2.1 of this thesis.

3.2.2 DNA extraction with Isolate II Genomic DNA Kit

Isolate II Genomic DNA Kit (Bioline) was used to extract DNA by following the manufacturer's protocol with some modification to maximize the lysis process and increase the amount of intact DNA. An amount of 25mg of -20°C frozen *Simulium* tissue was cut into small pieces with a scalpel and placed in a 1.5ml microcentrifuge tube. It was homogenized in a mixture of 75µl phosphate-buffered saline (PBS) and 10 µl of lysis buffer GL. An amount of 170µl lysis buffer GL and 25µl Proteinase K (600 mAU/ml) solution were added to the tube, followed by pulse vortexing. The mixture was incubated at 56°C overnight for 16 to 18 hours with occasional mild shaking. All centrifugations were carried out at 6,000 x g for 30 seconds, and then again at 11,000 x g for another 30 seconds. To elute the DNA, 50µl of 70°C preheated elution buffer G was directly added onto the silica membrane, incubated at room temperature for 1 minute and centrifuged as described above. The eluent was collected in a 1.5ml microcentrifuge tube and the elution step repeated a second time.

3.2.3 DNA concentration determination with Qubit® Fluorometer

The concentration of the extracted DNA was assessed using the procedure described in section 2.2.4 of this thesis.

3.2.4 Polymerase Chain Reaction

PCR Amplification of the amplicons of *Simulium* greater than 1000 bp was achieved with the Takara Gradient PCR thermal cycler. This was carried out in 10µl of reaction mix that consisted of 3.2µl of High Performance Liquid Chromatography (HPLC) water, 5µl of 2x GoTaq® Long PCR Master Mix (Promega Corp, Wisconsin, USA), 0.4µl of 10uM sense primer, 0.4µl of 10uM antisense primer, and 1µl of 1ng DNA.

The optimum annealing temperature was determined by using a gradient annealing temperature between 45°C and 65°C. After determining the optimum annealing temperature, the cycling conditions used for all the samples consisted of 95°C initial denaturation for 2 minutes, 30 cycles of: 95°C denaturing for 30 seconds; annealing of the primers to the DNA

template for 1 minute at a temperature of 56°C; and 72 °C extension for 3 minutes. These were finally extended for 10 minutes at 72 °C, and the products kept at 4 °C.

3.2.5 Gel Electrophoresis

Electrophoresis of the PCR products was carried out using 1.2% agarose gel stained with 3X GelRed® Nucleic Acid Gel Stain (Biotium) by following the manufacturer's protocol. For each gel run, 5µl of PCR product was mixed with 2µl of 10X DNA loading dye (BlueJuice™ Gel Loading Buffer, Invitrogen™ by life technologies, U.S.A.), loaded into the well of the agarose gel and run in 1X TAE buffer at 100V for 60 minutes. All gel runs included 2µl of 10X 100bp DNA ladder (Invitrogen™) by ensuring approximately 0.1 µg of ladder per mm lane width and by following the manufacturer's protocol. Following electrophoresis, the gel was visualized and photographed using a 302nm UV lamp (BioRad Gel Doc™ EZ Imager). The PCR products were purified with the QIAquick PCR Purification Kit (QIAGEN, 2018) by following the manufacturer's protocol. Portions of the PCR products that showed clear single sharp band on the gel with the expected band size were sent for Sanger sequencing for confirmation. The Sanger sequences were used as query sequences in BLASTn against the nucleotide database of NCBI. The samples that showed significant identity match to the continuous mitochondrial genomic regions of tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 of mitochondria genes of related *Simulium* sp. were selected for library preparation.

3.2.6 Library Preparation

The Nextera XT DNA Library Prep Kit (Illumina) was used to prepare libraries by following the manufacturer's protocol. Briefly, an amount of 5µl normalized long-range amplicon DNA (0.2 ng/µl per sample) was fragmented and amplified. Approximately 50µl of the amplified library was cleaned, normalized, quantified and then sequenced on an Illumina MiSeq® using v2 2×300bp chemistry.

3.2.7 Molecular phylogenetic and statistical analysis

Similar molecular phylogenetic and statistical analyses were carried out as in section 4.2.18 of this thesis.

3.3: Results from long range amplicon sequencing

3.3.1 Amplification of targeted tRNA and associated genes

Sequencing of the long-range amplicons produced nucleotide sequences of tRNA-Met, ND2, tRNA-Trp, tRNA-Tyr, tRNA-Cys, and COX1 Genes. The size of the nucleotide sequences was approximately 3 kbp as targeted (see Figure 3.3.0). The amplified region, blue colour, is shown and labelled in the annotated mitochondrial genome of African *Simulium* blackfly below.

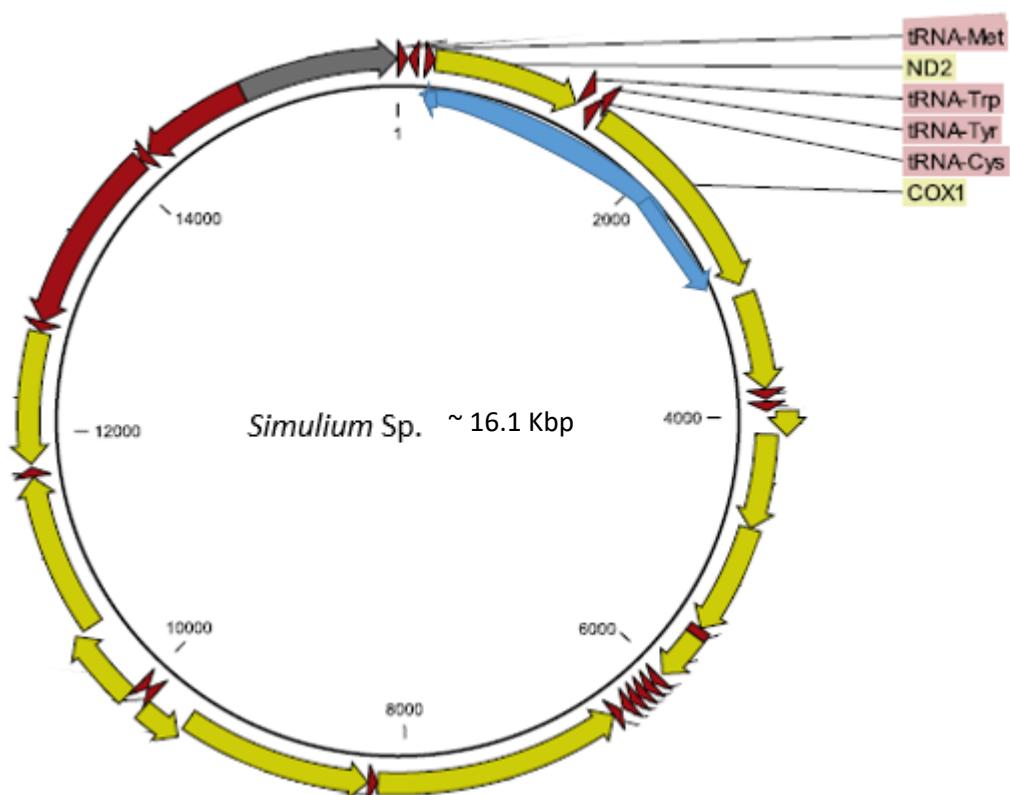


Figure 3.3.0: Amplified region of *Simulium* mitochondrial genome from tRNA-Met to CO1. The amplified region is coloured in blue with the genes labelled. The figure was produced in CLC Genomic Workbench 9.5.4 (<https://www.qiagenbioinformatics.com/>).

3.3.2 Polymorphism and diversity of tRNA's and associated genes from Ghana

The number of haplotypes was highest in Pru (27), slightly lower in Daka (26), and lowest in Black Volta (23). Likewise, the number of haplotypes resulted in the highest haplotype diversity in Pru (0.991), slightly lower in Daka (0.984), and least in Black Volta (0.975). However, these variations were negligible, leading to similar haplotype diversity with an

approximate value of 1 in each of the river basins. Thus, there was generally high haplotype diversity and high nucleotide diversity across the 3 river basins when compared with the diversity in other organisms (Goodall-Copestake *et al.*, 2012). The nucleotide diversities within each of the river basins were very similar to each other and statistically not different from 0.01 ($p = 0.590$) when tested with the one sample t-test using 0.01 as the test value (See supplementary sheet 9). Also, the overall nucleotide diversity across all 3 river basins was similar to each of the individual river basin values ($\pi = 0.01065$). Details of the polymorphism and diversity is found in Table 3.3.1.

Table 3.3.1: Polymorphism and diversity of *Simulium* sp. in Ghana using tRNA and associated genes

River basin	M	S	Eta (s)	H	Hd	π	k
Pru	163	159	62	27	0.991	0.01025	29.048
BlackV	175	166	68	23	0.975	0.01230	34.862
Daka	156	151	89	26	0.984	0.00919	26.051
All 3 river basins	257	242	94	66	0.986	0.01065	30.174

M = Total number of mutations; S = Number of polymorphic (segregating) sites; Eta (s) = Total number of singleton mutations; H = Number of Haplotypes; Hd = Haplotype (gene) diversity, Hd; π = Nucleotide diversity; k = Average number of nucleotide differences; BlackV = Black Volta.

3.3.3 Neutrality tests of tRNA's and associated genes from Ghana

In the neutrality test (Table 3.3.2), Tajima's D test was used to determine whether or not blackfly tRNA and associated sequences existed that did not fit the neutral theory model at equilibrium between mutation and genetic drift. Fu's F_s is shown to be more sensitive than Tajima's D towards detecting departures from neutrality due to genetic hitchhiking and population expansion (Fu, 1997; Zeng *et al.*, 2006), so Fu and Li's statistics were used to complement Tajima's D test statistic. Commonly across all the river basins, all the neutrality tests resulted in values below zero. This likely indicated the presence of a high frequency of rare alleles. But, the lack of statistical significance in all the river basins for all the neutrality tests ($p > 0.05$) indicated that the allelic distribution and/or level of variability did not violate the neutrality assumption. Hence, the blackfly tRNA and associated genes sequences were evolving randomly (neutrally).

Table 3.3.2: Neutrality Test of *Simulium* sp. in Ghana using tRNA and associated genes

River Basin	Tajima's D	Tajima's <i>p</i>	FLD*	FLD* <i>p</i>	FLF*	FLF* <i>p</i>
Pru	-1.13440	<i>p</i> > 0.05	-0.88920	<i>p</i> > 0.05	-1.14978	<i>p</i> > 0.05
Black Volta	-0.81424	<i>p</i> > 0.05	-0.95234	<i>p</i> > 0.05	-1.07241	<i>p</i> > 0.05
Daka	-1.30497	<i>p</i> > 0.05	-2.30401	<i>p</i> > 0.05	-2.33047	<i>p</i> > 0.05
All 3 river basins	-1.37609	<i>p</i> > 0.05	-2.12282	<i>p</i> > 0.05	-2.16442	<i>p</i> > 0.05

Tajima's D was calculated using the total number of mutations. FLD* = Fu and Li's D* test statistic, FLF* = Fu and Li's F* test statistic, *p* = Statistical significance.

3.3.4 Assessment of gene flow and genetic differentiation in Ghana using tRNA and associated genes

Nucleotide diversity, π , is defined as the average number of nucleotide differences per nucleotide site between two randomly chosen sequences in a population (Nei and Li, 1979). Table 3.3.3 shows the gene flow and genetic differentiation among the populations in the 3 river basins of Pru, Black Volta and Daka. Generally, there was high average number of nucleotide differences ($K_{xy} > 1$) between the pairwise river basin populations compared. In order of increasing magnitude, the average number of nucleotide differences occurred between populations in: Pru and Daka ($K_{xy} = 27.7$), Black Volta and Daka ($K_{xy} = 30.6$), and in Pru and Black Volta ($K_{xy} = 32.5$). The nucleotide diversity within populations also followed similar pattern as the average number of nucleotide differences between the pairwise populations in the 3 river basins. Thus, in order of increasing magnitude, the within-population nucleotide diversity occurred between populations in: Pru and Daka ($K_s = 27.5$), Black Volta and Daka ($K_s = 30.5$), and in Pru and Black Volta ($K_{xy} = 32.0$). In each of the populations compared using both the pairwise genetic distance, and the fixation index; the values were approximately zero. The individual values were however 0.005 for the populations in Black Volta and Daka, 0.006 for the populations in Pru and Daka, and 0.02 for the populations in Pru and Black Volta. Such concordant results demonstrated gene flow between the populations in the 3 river basins being compared and therefore support to the conclusion in Chapter Two of the lack of population structure across the 3 river basins.

Table 3.3.3: Gene flow and genetic differentiation per river basin in Ghana using tRNA and associated genes

Population 1	Population 2	<i>Hs</i>	<i>Ks</i>	<i>Kxy</i>	<i>GammaSt</i>	<i>Fst</i>
Pru	BlackV	0.98276	31.95517	32.47778	0.02506	0.01609
Pru	Daka	0.98736	27.54943	27.70333	0.01973	0.00556
BlackV	Daka	0.97931	30.45632	30.61222	0.01950	0.00509

The table shows various parameters of gene flow and genetic differentiation between population 1 and 2 in the 3 river basins of Pru, Black Volta and Daka. *Hs* = Average within-population diversity, *Ks* = within-population nucleotide diversity, *Kxy* = the average number of nucleotide differences between population 1 and population 2, *GammaSt* = pairwise genetic distance, and *Fst* = fixation index.

From Table 3.3.4, most of the blackflies' genetic variation occurred within populations (99.1%, $p > 0.05$), with the remaining variation occurring among populations (0.9%, $p > 0.05$). The test for the overall measurement of population differentiation due to genetic structure, the overall Fixation Index (*Fst*), was very small (0.009). The force of gene flow in the population for a given generation (*Nm*), that influences the tendency of population differentiation as a result of genetic structure, was however very large (*Fst*'s *Nm* = 54.04). The overall pairwise genetic distance, the *GammaSt* statistic (0.02853), was equally small with an *Nm* value of 17.02.

Table 3.3.4: Overall population differentiation test in Pru, Black Volta and Daka river basins using tRNA and associated genes

Variation	%Variation	<i>Fst</i>	<i>Fst's p</i>	<i>Fst's Nm</i>	<i>GammaSt</i>	<i>GammaSt's Nm</i>
Among Population	0.9	0.009	0.164	54.04	0.02853	17.02
Within Population	99.1					

N = Effective population size, m = proportion of migrants in a population, Nm = the force of gene flow in the population for a given generation, FST = Fixation Index across the river basins, $FST's p$ = Significance of the Fixation index (1000 permutations), $GammaST$ = overall pairwise genetic distance, % Variation = the percentage of variation observed either within or among given populations.

3.3.5 *A priori* assessment of genetic relatedness using the tRNA and associated genes

The *a priori* method of DAPC was used to identify genetic clusters within the dataset (Figure 3.3.1, A). It was prepared from 90 mitochondrial genomic sequences composed of tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1, amplified and sequenced as a single amplicon with all intervening sequences. The optimum number of principal components was 20 and it accounted for more than 90% of the variation within the data. Further discriminant analysis of principal component investigation by the use of 7 discriminant functions identified 8 clusters. The 8 clusters formed 3 major groups: clusters 2, 3, 5, 6, 7 and 8 formed the first and largest group; cluster 1 formed the second group; and cluster 4 formed the third and last group. The largest cluster group, consisting of 6 of the clusters, formed overlaps and a cline, while the other 2 remaining cluster groups formed independent groups that were distinctively separated from each other cluster group. One of the 3 distinct groups may be heterogeneous.

From figure 3.3.1 B, the percentage of reassignment of blackflies to their original clusters showed that clusters 2, 3, 4, 5, and 7 were assigned to their original clusters with 100% accuracy. The values for clusters 1, 6 and 8 however were 93.75%, 57.14%, and 83.33% respectively.

Figure 3.3.1 C shows the haplotype network produced using the minimum spanning network. A total of 8 haplotype groups were observed with the haplotype network supporting the DAPC and showing a strong association of haplotype with cluster assignment. With the exception of haplotype groups 2 (MOTU-C2, green colour) and 5 (MOTU-C5, pink colour); all the haplotype groups had few mutational steps differences (1-20) separating any 2 individual haplotypes closely joined together. The haplotype groups 2 and 5 were distinctively different from the others with relatively greater number of mutational step differences (47 and 55 respectively) separating them from all the other groups. These suggested that the 8 haplotype groups were further organized into 3 main broader groups, and this observation was consistent with the 3 main groupings of the DAPC investigated.

Hence, both the DAPC and haplotype network results were concordant, revealing the presence of 3 major genetic cluster or haplotype groups across the 3 river basins. This finding is also consistent with all the results obtained from the analysis carried out by the use of the mitochondrial and nuclear short amplicon sequences (CO1, ND4, 16S-rRNA, ITS2 and IGS genes).

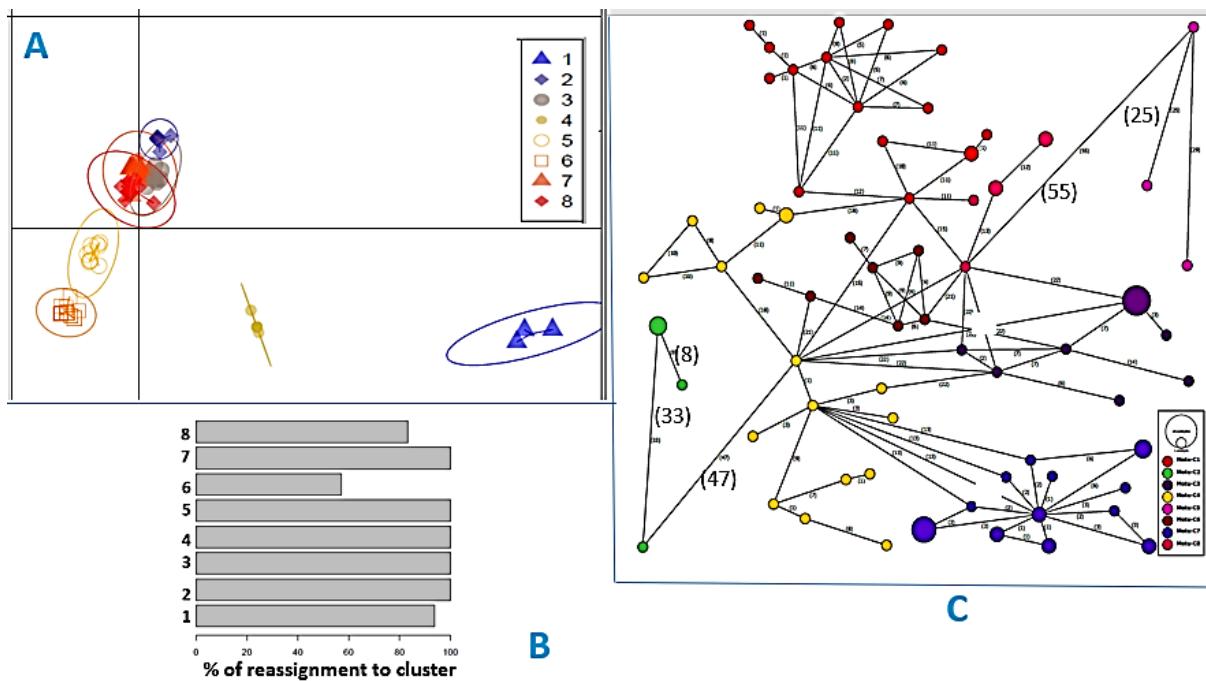


Figure 3.3.1: *A priori* assessment of genetic relatedness from 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences without population assignment. In the figure, plot A = DAPC scatterplot and colour coded by cluster, B = bar chart showing the percentage of successful reassignment of individual blackflies to the 8 original clusters identified, C = haplotype network of the tRNA's and associated genes that were colour coded by using the major clades identified in the phylogenetic tree of figure 3.3.3. The DAPC was produced with the *a priori* method without assigning population information and by choosing 20 principal components, 8 clusters with 7 discriminant functions. The DAPC graph and bar chart were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010)

3.3.6 *A posteriori* assessment of genetic relatedness using the tRNA and associated genes

The *a posteriori* method of DAPC was used to assess the genetic relatedness between the 90 tRNA and associated genes. From the overlaps of the 3 ellipses, individual blackflies among the 3 river basins were found to be related to each other as one would expect from populations with some form of gene flow. The overlap of the ellipses was however not absolute, with the presence of some of the blackflies being unique to each river basin while others were found to share the genetic characteristics with blackflies from the other river basins. This generally suggested the lack of geographic population structure by river basin.

Figure 3.3.2 C showed the percentage of reassignment of individual blackflies to the actual river basin of collection. Generally, there was moderate percentage of reassignment with values of 66.7% in Pru, and 53.3% realized in both Black Volta and Daka. This moderate value possibly resulted from the gene flow of migrant blackfly individuals across the 3 river basins either in the present or ancestral generations, in the recent or distant past.

The haplotype network, colour coded by river basin, showed that there were few shared haplotypes with the majority being from singletons (Figure 3.3.2 A). The haplotypes from each river basin were closely linked to those of the other river basins such that no cluster of haplotypes consisting of only a single river basin type of pattern was observed. From the number of mutational step differences between closely linked haplotypes, 3 main haplotype groups were observed. Thus, there existed a relatively small number of related groups that had diverged, such as one would see from a small, recent founding population composed of those haplotypes that had since diverged.

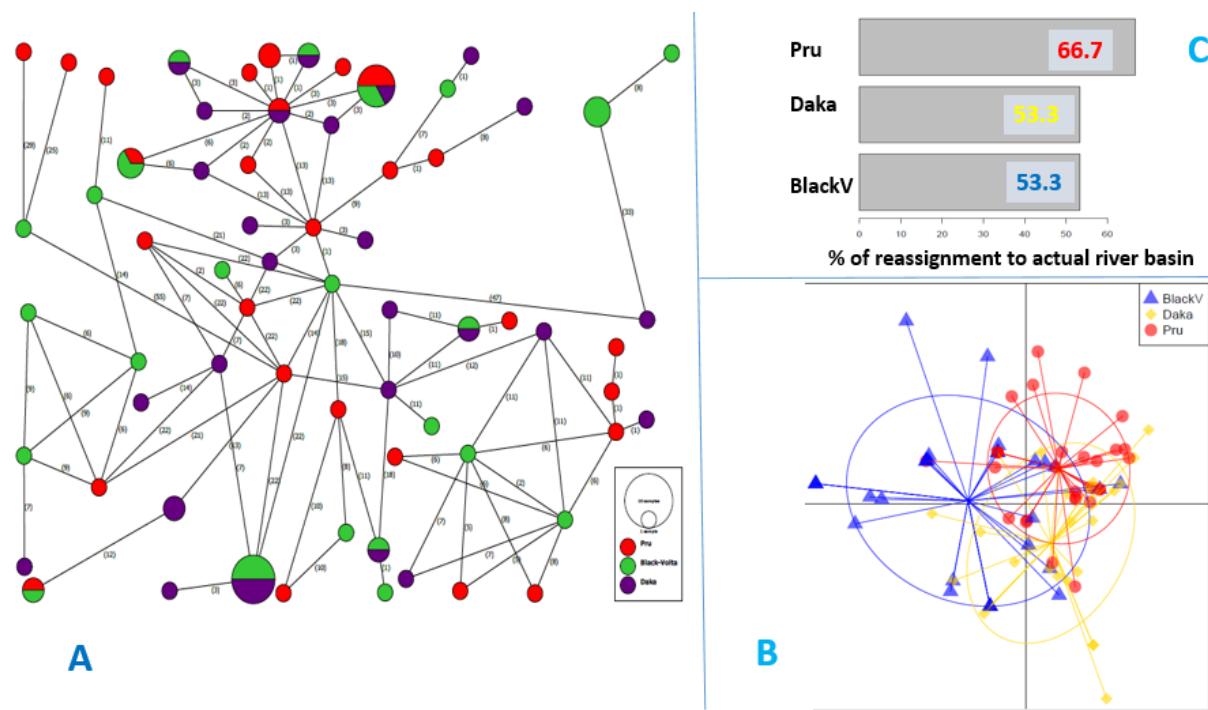


Figure 3.3.2: *A posteriori* assessment of genetic relatedness using 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences after population assignment. The *Simulium* blackflies were collected from 3 river basins in Pru, Black-Volta and Daka. In the figure, section: “A” = haplotype network, “B” = DAPC scatterplot, and “C” = percentage of reassignment of individual blackflies to their original river basins of collection. The

haplotype network was produced using minimum spanning network and Epsilon of 0 in Popart 1.7 (Bandelt *et al.*, 1999). Each circle in the haplotype network corresponds to one haplotype, and the size is proportional to its frequency among the samples. Colours of the circles correspond to river basin sampling locations. The ellipses of the DAPC represent the different river basins from which the sample's sequences were obtained while the dots represent the individual blackflies collected. Both the DAPC and bar chart were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

3.3.7 Nucleotide substitution model and phylogenetic test using tRNA and associated genes

Maximum Likelihood fits were used to assess 24 different nucleotide substitution models (Table 3.3.5). The Bayesian Information Criterion, corrected Akaike Information Criterion, and Maximum Likelihood respectively ranged from 15187.53 to 17105.99, 13265.14 to 15256.72, and -6448.44 to -7451.24. The Tamura-Nei model with gamma distribution and invariant sites was considered to describe the substitution pattern the best because it had the lowest BIC, AICc, and the largest maximum likelihood.

Figure 3.3.3 shows an unrooted phylogenetic tree produced from 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences. The tree building method was maximum likelihood with 1000 replicates. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites. The rate variation model allowed for some sites to be evolutionarily invariable. There were 3 main clades. Two clades had few members, whiles the third had the majority (90%). The first clade had blackflies belonging to the Pru and Daka river basins. The second clade had blackflies that originated from the Black Volta and Daka river basins. The third major clade had blackflies belonging to all 3 river basins. There were strong bootstrap supports, 87% to 100%, at the points of blackfly divergence within clades. Members of each major clade shared most recent common ancestor and their clades were distinctively diverged from the clades of other blackflies. The presence of blackflies from the different river basins sharing common clades support the notion that blackflies from all 3 river basins share most recent common ancestor, exhibit gene flow, and interbreed among themselves.

Table 3.3.5: Nucleotide substitution models test of 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences

Model	#Param	BIC	AICc	lnL
TN93+G+I	184	15187.53367	13265.13803	-6448.435459
T92+G+I	181	15209.659	13318.6025	-6478.172003
HKY+G+I	183	15211.57529	13299.62602	-6466.680898
GTR+G+I	187	15217.70016	13263.96552	-6444.844824
TN93+G	183	15257.58802	13345.63874	-6489.68726
GTR+G	186	15288.31826	13345.02994	-6486.3785
T92+G	180	15293.29057	13412.68049	-6526.212418
HKY+G	182	15301.29562	13399.79272	-6517.765686
TN93+I	183	15593.86175	13681.91248	-6657.824127
T92+I	180	15613.09319	13732.4831	-6686.113727
HKY+I	182	15617.92841	13716.42551	-6676.08208
GTR+I	186	15625.9626	13682.67428	-6655.200672
TN93	182	15816.08057	13914.57768	-6775.158164
T92	179	15831.37799	13961.21433	-6801.480753
HKY	181	15835.88286	13944.82636	-6791.283935
GTR	185	15848.70661	13915.86462	-6772.797304
K2+G+I	180	15874.82798	13994.21789	-6816.981121
K2+G	179	15991.4964	14121.33275	-6881.539961
K2+I	179	16305.43159	14435.26793	-7038.507553
JC+G+I	179	16473.77459	14603.61094	-7122.679055
K2	178	16515.59534	14655.87813	-7149.814056
JC+G	178	16587.48164	14727.76443	-7185.757206
JC+I	178	16896.87235	15037.15513	-7340.452559
JC	177	17105.99077	15256.72001	-7451.236396

The table shows the Maximum Likelihood fits of 24 different nucleotide substitution models.

The model test was run without singleton sites but contained 2 variants parsimony informative sites. Models with the lowest BIC (Bayesian Information Criterion) scores are considered to describe the substitution pattern the best. For each model, AICc (Akaike Information Criterion, corrected) value, Maximum Likelihood (lnL) value are shown. In the table, GTR= General Time Reversible; HKY= Hasegawa-Kishino-Yano; TN93= Tamura-Nei; T92= Tamura 3-parameter; K2= Kimura 2-parameter; JC= Jukes-Cantor, G= Gamma distributed, and I= Invariant site.

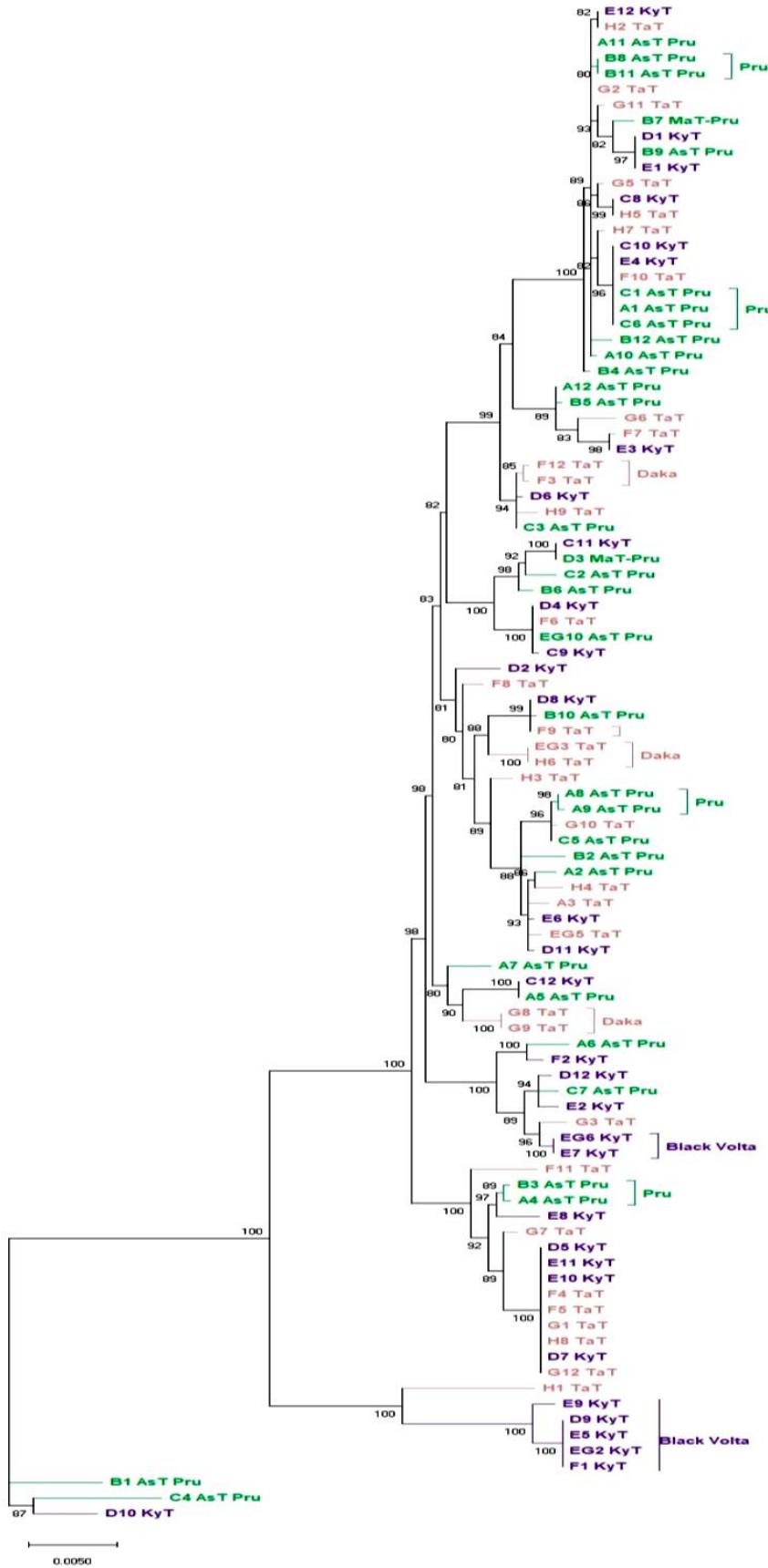


Figure 3.3.3: Unrooted phylogenetic tree from 90 tRNA-Met, ND2, tRNA-Trp, tRNA-Cys, tRNA-Tyr, and COX1 gene sequences. The evolutionary history was inferred by using the

Maximum Likelihood method based on the Tamura 3-parameter model (Tamura, 1992). The percentage of trees in which the associated taxa clustered together is shown next to the branches, thus showing the bootstrap values. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018). In the figure: green = Pru river basin, purple = Black Volta river basin, and pink = Daka river basin.

3.4 Discussion

3.4.1 Genetic diversity of West African *Simulium* blackflies

The blackflies in all the studied 3 river basins from Daka, through Pru to Black Volta along the East-West transect of the central ecological transition zone of Ghana were evolving randomly as one would expect from a neutrally evolving population (Tajima's $D < 0$, $p > 0.05$). There was similar observation made with *Simulium aureohirtum* with significantly negative Tajima's D values across 3 studied lineages that were undergoing demographic expansion (Thaijarern *et al.*, 2014). The main difference between our results and that of the *S. aureohirtum* is that whereas ours was statistically non-significant, *S. aureohirtum* showed a significant negative Tajima's value consistent with an expanding population.

In our study however, the lack of population differentiation and gene flow allows all populations to accumulate mutations without becoming differentiated. The large number of mutations, characterized by high diversity, implies perhaps that the populations might be old or, more likely, very large. Large populations are more genetically diverse. These river basins in the central ecological transition zone and even those in the whole of Ghana were hyperendemic for onchocerciasis prior to control of the vectors and parasites (De *et al.*, 1989; Remme *et al.*, 1989; Garms *et al.*, 2015). Moreover, transmission parameters many years after transmission control had commenced in the study area and in Ghana (Garms *et al.*, 2015), closely resembled those from other areas with hyperendemic onchocerciasis prior to disease intervention measures such as in Liberia (Garms, 1973; Garms, 1987), Sierra Leone (Bockarie *et al.*, 1990), Ivory Coast (Quillévéré *et al.*, 1977), and Togo and Benin (Garms and Cheke, 1985).

Hyperendemicity requires high biting rates, which in turn require large fly populations. So, the fact that these are hyperendemic zones (De *et al.*, 1989; Remme *et al.*, 1989) is consistent with large population size. Furthermore, sampling only 30 flies per river basin will result in an overestimate of haplotype diversity if the population is large simply because it is a small sample. The absence of recombination in these mitochondrial genes possibly made all nucleotide sites to evolve independently. The lack of recombination also results in high linkage disequilibrium, which again tends to inflate haplotype diversity. So, the data point to a large, near panmictic population with high levels of gene flow that would result in high transmission and hence hyperendemic onchocerciasis.

A large and old population of blackflies in the river basins of Pru, Black Volta and Daka that shows significant gene flow between the river basins could be the result of shared recent

history that could reduce the likelihood of accumulation of genetic differentiation between populations. Such blackfly species are capable of utilizing a wide range of ecological conditions as well as adapting to anthropogenic disturbed habitats including agricultural areas and artificial flowing water (Pramual and Kuvangkadilok, 2009b). The transition ecological zone of Ghana has similar agricultural areas and flowing waters, and these could be contributing factors to the observed population structure. While some blackfly species are confined to a given area such as forest or savannah and tend to have high levels of genetic structuring, others have continuous geographic distribution characterized by gene flow between populations (Pramual and Kuvangkadilok, 2012). The *Simulium* blackflies in Daka, Pru and Black Volta were genetically homogenous possibly due to being in the transition ecological zone characterized by forest and savannah climates. This might be a contributing factor to their continuous geographic distribution. This finding is also consistent with studies of other species such as *Simulium siamense* that is genetically homogenous due to occupying both forest and open areas (Pramual *et al.*, 2011).

Across the 3 river basins, the average number of nucleotide differences correlated with the total number of mutations and the number of polymorphic sites (shown in Table 3.3.1). Thus, the average number of nucleotide differences was relatively low in Daka ($k = 26.1$), moderate in Pru ($k = 29.0$) and high in Black Volta ($k = 34.86$). Despite these variation in nucleotide diversity, total number of mutations and the number of segregating sites, the nucleotide diversity was the same in all the river basins, and the haplotype diversities were very similar across the river basins.

Table 3.3.3 showed that the average within-population diversities of the blackflies across the 3 river basins from East to West of the transition ecological zone of Ghana were very similar and very high ($H_s = 0.98\text{--}0.99$). The within-population nucleotide diversities (K_s) of the blackflies were equally high ($K_s = 27.5\text{--}32.0$). The overall nucleotide diversity (π) across all the river basins in both the long range and the amplicon resequencing of CO1 and ND4 were the same ($\pi = 0.01$), but comparatively higher than those in other organisms (Goodall-Copestake *et al.*, 2012). Despite the observation that the nucleotide diversities were generally moderate, the haplotype diversities in both the long range and amplicon resequencing of CO1 and ND4 were very high (H_d of at least 0.8). Unlike the nuclear gene ITS2 and the nuclear spacer gene IGS that had high haplotype diversity (H_d 0.7–0.9) across all river basins, the long-range marker showed a different pattern of nucleotide diversity. Thus, the nucleotide diversity of the long-range marker ($\pi = 0.01$) was consistently higher than those of the nuclear

IGS across all river basins ($\pi = 0.003$). Comparatively, the nucleotide diversity of the nuclear ITS2 was moderately higher ($\pi = 0.03$) than that of the long range ($\pi = 0.01$) for the river basins of Pru and Daka, but that of the Black Volta was very low ($\pi = 0.004$). The similarity in the diversity of the blackflies across the river basins in the long-range genes and the amplicon resequencing coding genes of CO1 and ND4 indicated similar evolutionary processes and responses taking place in the African *Simulium* blackfly coding sequences of CO1, ND4, ND2, and also in the associated tRNA-Met, tRNA-Trp, tRNA-Tyr and tRNA-Cys.

The unique haplotypes further showed that the Ghanaian blackflies were genetically diverse such that the historical population size was possibly very large. The close relationship existing between the haplotypes gives the indication of a shared recent evolutionary origin that is consistent with the old and large population at equilibrium in Ghana. The nucleotide diversity $\pi = 4\mu Ne$; where μ is the mutation rate, and Ne is the effective population size. Assuming that the mutation rate is relatively constant, π is directly proportional to the population size, and higher values of π mean larger populations when those populations are at equilibrium. The observation of poor discrimination between the blackflies in the 3 river basins most likely is attributed to an old and very large population at panmixia over a long period of time, being characterized by lots of mutations, but no population subdivision. This old and large population at equilibrium appear to be related with parasite-vector interactions that enables the vectors to survive long enough to enable the parasites they carry to pass on to its definitive host. There have been many reviews and reports of parasites of medical importance manipulating their vectors, such as the occurrence of infection resulting in altered vector feeding patterns to such extent that more host contacts are made (Molyneux and Jefferies, 1986; Moore, 1993). It was proposed that natural selection tends to favour parasites capable of manipulating their vectors to enhance their transmission (Hurd, 2003). The compatibility between sympatric as opposed to allopatric parasite-vector combinations indicates local adaptation capable of resulting in evidence of co-evolution (Woolhouse *et al.*, 2002) between *Simulium* blackflies and *Onchocerca* sp. It is possible that the large population at equilibrium of the West African blackflies played a significant role in the speciation event that resulted in the *Onchocerca volvulus* and *O. ochengi* proposed host switch from their most recent common ancestor during the period of introducing domesticated cattle in Northeast Africa (Bain, 1981; Keddie *et al.*, 1998; Marshall and Hildebrand, 2002; Lefoulon *et al.*, 2017), and later to West Africa. Cattle were more abundant in the Northeast Africa than in West Africa and the proposed switch from cattle to human host by the parasite most likely

occurred in areas characterized by more humans than cattle. The local anthropophilic simuliids that were already there when cattle arrived, and already feeding on humans, started feeding on cattle too and acquired *O. ochengi* for the first time. *Onchocerca* parasites and simuliids have been implicated to have reciprocal effects on each other's phenotype and genotype. They further exert reciprocal effects on each other's survival at different stages of the life cycle of the parasite inside the blackfly, thus adapting to decrease deleterious effects on fitness while increasing transmission (Webster *et al.*, 2004; Basáñez *et al.*, 2009). Filarial parasites are known to switch vectors frequently. For instance, Many *Onchocerca* species are transmitted by vectors unrelated to black flies, and LF is transmitted by many mosquitos. Both malaria and LF can occur as concomitant human infections in many tropical regions while also sharing common vectors (Buck *et al.*, 1978). In addition, infective vectors are usually less fit than non-infective parasites and this could have led to population bottleneck in the parasites. Such parasites experiencing bottleneck most likely influenced the behaviour of their vectors to let them feed more frequently than normal to increase their chance of completing their life cycle. Subsequently, the bottleneck in the parasites might have been followed by population expansion in the course of time. This possibly led to co-evolution in the parasites and vectors. Due to the nature of the survival and co-evolution, the population expansion of the parasites could have led to similar survival of the vectors, making the vector population large, as suggested by this long-range mitochondrial data. This is consistent with the universally accepted principle that the success of parasites get improved if it developed faster or if its vector lived longer and one of these traits was under selective pressure (Koella, 1999).

3.4.2 *Simulium* population structure in West Africa

Table 3.3.3-4 showed that the majority of the genetic variation (99.1%) occurred within populations while a very small amount of it (0.9%) occurred among the populations. In addition, the fixation indices values were approximately 0 in all the river basins. Both the negligible fixation indices and the large force of gene flow in the populations (Fst 's $Nm = 54.04$), coupled with the majority of the genetic variation occurring within the populations rather than among, supported the idea of gene flow between populations in the 3 river basins, and a consequent lack of significant population differentiation and structure across the entire transition ecological zone of Ghana. The lack of population structure characterized by gene flow between the blackflies in the 3 river basins could be attributed to a number of reasons. A plausible contributing factor could be from the long-range vector migration of some blackflies. The *Simulium damnosum* subcomplex in the former OCP area showed changing

distributions. The *S. sirbanum* and *S. damnosum* s.s. were widely distributed, occurring primarily in Ghana, Togo, Benin and Cote d'Ivoire. During the dry season, they were capable of spreading to Sierra Leone (Garms and Vajime, 1975; Vajime and Quillévéré, 1978; Prod'hon *et al.*, 1982). Deforestation, climatic changes and high mobility have been noticed to be the major driving force of some species of *Simulium* spreading from their main habitats to other areas where they were never previously found. For instance, *S. sirbanum* was widespread in all OCP countries, but in Sierra Leone, it was only found in the extreme north of the country for a very long time (Prod'hon *et al.*, 1982; Post and Crosskey, 1985). Due to deforestation, climatic changes and high mobility, *S. sirbanum* later spread to the south of Sierra Leone and other forested parts of the OCP area, including Liberia (Garms *et al.*, 1991; Güzelhan and Garms, 1991). Vector control and flooding, such as from dams, could be a contributing factor affecting the distribution of blackflies. It could have caused them to decrease in their numbers and even disappear completely from their principal place of location, and some of their larvae washed to areas where they were not found previously. Such seasonal flooding activities potentially caused populations that were formerly isolated by large geographical distances to survive and breed with those of other areas over the transitional ecological zone. For instance, an observation was made that the proportion of *S. sirbanum* reduced in Sierra Leone following larvical treatments. *S. sirbanum* and *S. damnosum* vanished downstream of the Akosombo dam in Ghana as a result of flooding activities of the Senchi Rapids by the Kpong Dam along the Volta River (Bissan *et al.*, 1995). Most dams are constructed to provide benefits like hydroelectricity, flood control, waterway creation and the provision of water for irrigation in regions experiencing drought (Yüksel, 2009; Biswas, 2012; Terminski, 2014). The implementation of dam construction and operation however comes with both benefits and negative impacts (McManus *et al.*, 2010). The quest to increase the energy production of Ghana resulted in the construction of Bui dam at the Bui Gorge, a project that was on the planning boards since the 1920s (Miescher and Tsikata, 2009; Okoampa-Ahoofe, 2009; Hensengerth, 2011). Seasonal flooding of the river before dam construction occurred between June and November (Management, 2007). Flooding of the river became controlled to some extent by the management of the dam during periods of water level management for the operation of the dam. These activities usually flood the Black Volta river basin and cause variation in the population and distribution of blackflies in that locality. Seasonal changes could have caused the blackfly distribution to change to the extent of making them occupy localities that they would have previously not done so. Over the course of time, inter-breeding could have caused gene flow, decreased

genetic diversity and ultimately reduced or completely eliminated population structure. During the dry season in Ghana, populations became more restricted in the north and moved south of their originally permanent distribution range. This is seen in the shortage of dry season breeding sites because most of the rivers, such as the Sudan savanna belt and some of the Guinea savanna dry out entirely. The only exception to this is the few perennial rivers. A major contributing factor could possibly be the displacement of the *Simulium* adults on the drying harmattan winds that seasonally blow in December onwards in the direction of south-west across West Africa, including the transition zone of Ghana, (Gwynne-Jones, 1978) towards the Gulf of Guinea. A reverse event takes place during the rainy season, usually occurring from June onwards, characterized by the monsoon winds blowing and displacing blackflies in the north-eastern direction (Garms *et al.*, 1979; Walsh *et al.*, 1981a; Baker *et al.*, 1990). This was the major contributing factor that resulted in the repopulation of breeding sites in previously vector-controlled areas (Garms *et al.*, 1979; Walsh *et al.*, 1981a; Baldry *et al.*, 1985; Baker *et al.*, 1990). The long-term epidemiological trend across the 3 major river basins were likely determined by both short-range and long-range events. The short-range migration pattern is consistent with the observation that onchocerciasis infection is greatest when closer to a river basin but progressively decreases as one moves away from the river basins. The long-range migration pattern on the other hand explains why there is gene flow between the two extreme river basins by distance, which are Black Volta and Daka. Such variations in the seasonal wind movement caused members of the *S. damnosum* complex, such as the savanna species, to migrate distances of 400-500 km and above on the wind (Garms *et al.*, 1979; Cheke and Garms, 1983). In areas of overlap, such as the transition ecological zone, different forms of blackflies that would otherwise form distinct population structures in a non-overlapping area, are noted to interbreed and this would most likely lead to gene flow, decreased genetic diversity and reduced or complete absence of population structure (Boakye *et al.*, 1993b). Such overlap could sometimes lead to the occurrence of an intraspecific stepped cline (Surtees, 1988; Boakye *et al.*, 1993b). Another possible reason going hand in hand with the seasonal climatic changes and bi-directional wind movements is the search for breeding site. Changes in the vegetation of blackfly habitats influence their distribution and migration (Boakye *et al.*, 1998). The movement of natural water bodies such as rivers and streams in the transition zone could be in the direction of south-east or south-west towards the Gulf of Guinea, and seldomly moving up north. This movement would potentially carry larvae from one point to another. Also, adult blackflies follow the water movement in that south-east or south-west direction in search of fast flowing water areas to

breed, and this is consistent with the observation that capture points at slow-flowing river areas hardly produce migrating blackflies (Baker *et al.*, 1990). These movements in search of breading sites progressively make one form of blackfly move into the habitat of other blackflies, thereby making it rare to have any particular region constituting a permanent dwelling place for only one form of blackfly. Adult *Simulium* spp. often spread short distances (Wenk, 1981) within the average distance range as low as 7-15 Km to as high as 20-35 Km to cover the popularly cited total travel distance of 400-500 Km (Johnson *et al.*, 1985; Garms *et al.*, 1989; Baker *et al.*, 1990). Such short distance movements that ultimately cover large total distances can lead to gene flow in blackflies occupying different river basins and subsequently cause a reduction in the genetic diversity and population structure. Moreover, the average distance between any two river basins studied in the transition ecological zone of Ghana was within the 400 - 500 Km blackfly flight range, thereby making it possible for blackflies to migrate in the course of time from one river basin to another (Garms *et al.*, 1979; Magor and Rosenberg, 1980; Garms, 1981; Garms, 1982; Johnson *et al.*, 1985; Baker *et al.*, 1990; Post *et al.*, 2013).

3.4.3 Implications on Delineation of Transmission Zones

In the central ecological transition zone of Ghana, geographical distances between the river basins do not have any significant association with genetic distances or diversity of the blackflies.

Although nucleotide diversity was generally lower across the river basins, the haplotype diversities were very large and consistent with that of a panmictic population characterized by significant gene flow. There is a small probability value under panmictic population structure for any given large observed value of K_{ST} or the proportion of the ratio of the diversity within a population to that in the entire population (Hudson *et al.*, 1992).

River basins characterized by the absence of gene flow between their blackflies, but with the existence of significant population structure, have greater likelihood of being separate transmission zones of onchocerciasis. *A priori* DAPC analysis revealed the existence of 3 distinct molecular groups of blackflies and one of these may be genetically heterogeneous. The cause of this genetic differentiation under such shared large-scale geographical distance location need further studies. It could possibly result from a reproductive barrier because migratory movements are not only a function of geographical distance but are also affected by the existence of ecological and behaviour barriers (Barbujani and Sokal 1990, 1991, 1991; Sokal and Oden 1988; Sokal *et al.*, 1989, Dupanloup de Ceuninck *et al.*, 2000; Rosser *et al.*,

2000). Thus, 3 molecular groupings of blackflies across the central ecological transition zone of Ghana most likely resulted from barriers of reproduction existing within the 3 river basins but influencing the different blackflies in different ways. Although these distinct molecular groupings of blackflies share a single onchocerciasis transmission zone, their role in transmission needs studying and there are many techniques available to identify the presence of genetic barriers as well as to determine their role in population differentiation (Barbujani *et al.*, 1989; Barbujani and Sokal 1990; Stenico *et al.*, 1998; Simoni *et al.*, 1999).

3.4.4 Implications for Decisions to Stop Interventions

This study demonstrated that the entire central ecological transition zone of Ghana constitutes a single onchocerciasis transmission zone and this finding does not agree with the transmission zone definition based on *Simulium* vector breeding sites (WHO, 2016). Based on the current river basin definition of transmission zone, if ivermectin and other onchocerciasis interventions stop in communities of a given river basin, our study suggests that infected blackflies from the other river basins would potentially travel to reinvoke the communities that had stopped ivermectin treatment. There needs to be a comprehensive assessment of the migration, diversity, divergence and population structure of *Simulium* blackflies in the onchocerciasis endemic countries. This will lead to a better definition of onchocerciasis transmission zone, delineation of a reliable transmission zone, greater accuracy in the estimation of timelines to stop MDA using ivermectin, and a greater chance of guarantee to sustain the progress made in decades of onchocerciasis.

3.4.5 Decision on study objectives and hypothesis

The genetic diversity in the tRNA rich section of the West African *Simulium* mitochondrial genome was characterized as 0.986 and 0.01065 for the haplotype diversity and nucleotide diversity respectively across the entire central ecological transition zone of Ghana. The genetic diversity was comparatively higher than expected from related genes of similar organisms.

The nucleotide sequences of the tRNA-rich section of mitochondrial genome have demonstrated significant gene flow, lack of population structure by geographical place of origin, and are further characterized by the existence of significant migration of the blackflies between river basins in the central ecological transition zone of Ghana.

The genetic diversity in the tRNA-rich section of the *Simulium* mitochondrial genome was similar to those in the other genes in the other sections of the West African *Simulium*

blackflies, with the approximate values of 1 and 0.01 for the haplotype diversity and nucleotide diversity respectively. Hence, the null hypothesis 1 was accepted and the alternate hypothesis 1 was rejected.

Using the 90 nucleotide sequences of the tRNA and associated genes from West-Africa, it was demonstrated that there existed significant gene flow between blackflies in the 3 river basins. This indicated that significant migration of *Simulium* blackflies between river basins was occurring across the entire central ecological transition zone of Ghana. Hence, the null hypothesis 2 was accepted and the alternate hypothesis 2 was rejected.

3.5 Conclusion

The genomic region between tRNA-Gln and COII were sequenced from 90 West African blackflies. The size of each nucleotide sequence for the entire genomic region within a single blackfly was approximately 3 kbp and the sequenced genes were tRNA-Met, ND2, tRNA-Trp, tRNA-Tyr, tRNA-Cys, and COX1. The haplotype diversity in the studied river basins in West Africa was highest in Pru (0.991), slightly lower in Daka (0.984), and least in Black Volta (0.975). Each of these values was higher than the average haplotype diversity (mean $h = 0.63388$) observed in other related organisms. Observation was made in the studied localities in West-Africa of a higher nucleotide diversity than observed in other related organisms (mean $\pi = 0.00388$), with individual values of 0.01025 in Pru, 0.01230 in Black Volta and 0.00919 in Daka river basin. Thus, the overall genetic diversity in the tRNA rich section of the West African *Simulium* mitochondrial genome was characterized as 0.986 and 0.01065 for the haplotype diversity and nucleotide diversity respectively.

The phylogenetic tree produced from the long-range sequences had better bootstrap support values than those from the short amplicon resequencing. This suggests that more sequence data are needed to generate a more reliable phylogenetic tree with better bootstrap supports. The sequences within the 3 river basins were all evolving randomly as one would expect from a panmictic population with random mating and greater gene flow between the populations. There existed 3 molecular groups of blackflies and one of these may be genetically heterogeneous. There was significant gene flow between blackflies in the 3 river basins, reflecting significant migration between river basins throughout the central ecological transition zone of Ghana. The entire central ecological transition zone of Ghana constitutes a single onchocerciasis transmission zone.

Chapter 4: Whole Mitochondrial Genome Sequencing

4.1 Introduction

4.1.1 Background

Attempts to sequence whole mitogenome of *Simulium* blackflies collected from the central ecological transition zone of Ghana, Malawi and Cameroon became laborious and time consuming when the PCR method was the main choice. Some of the molecular techniques attempted were the Long-range PCR amplification, Touch-Up Gradient Amplification, Touchdown PCR, and Nested PCR (Coyle *et al.*, 2004; Davies and Gray, 2002; Massung *et al.*, 1998; Rowther *et al.*, 2012). When it was realised that the process was not yielding the intended amplicons of large size, an alternate method was sought. Upon realizing that insects generally have a higher mtDNA:nucDNA copy number ratio, it was hypothesized that a library prepared with whole insect DNA will bias the sequence amplification process in favour of more mitochondrial genes (rather than nuclear genes). This was partly because of the exponential amplification principle of PCR that enables the generation of a far greater quantity of any component of a starting material that was in greater proportion to the remaining components of the sample at the start-up amplification process. Although it was thought possible, there were uncertainties as to the possibility of greater success because this was not the usual way of generating whole genome sequences. A pilot trial with libraries prepared with two blackfly samples from Ghana became successful in a MiSeq run aimed at producing low sequencing depth reads. A follow-up next-generation sequencing, to replicate the success of the initial work, was carried out with 10 additional samples and it was also successful. A final validation with a much higher number of samples (a total of 25) in the NGS was again successful and in all these amplifications, whole genomes were generated for each sample with sufficient reads across all the mtDNA loci. The major finding of this work was that relatively low depth whole genome sequencing was sufficient to recover whole mtDNA sequence at sufficient depth to call variants.

Organisms of the metazoan group are known to have a mitochondrial genome composed of a circular molecule with the size range of about 14.5 to 19.5 kbp (Altman and Katz, 1976). Over the course of time, the complete nucleotide sequences and gene contents have been researched and sequenced for the mitochondrial DNA molecules of many organisms including *Drosophila* (Yu *et al.*, 2007), mouse (Van Etten *et al.*, 1980; Bibb *et al.*, 1981; Slonimski *et al.*, 1982), humans (Crews and Attardi, 1980; Anderson *et al.*, 1981; Montoya *et al.*, 1981; Ojala *et al.*, 1981) and cows (Anderson *et al.*, 1982). Although at the beginning of

this study, there was not a single mitochondrial genome of *Simulium* known to have been sequenced in any part of the world, 4 species of this haematophagous genus that do not vector the parasites of the human onchocerciasis have been sequenced recently from Asia and Europe (Day *et al.*, 2016; Wang *et al.*, 2016; Wang and Huang, 2019; Zhang *et al.*, 2019). The gene content of the complete mitochondrial genome of mammals and insects to a large extent contains 22 tRNAs, 2 rRNAs, 13 protein coding genomic units that are made up of 6 protein coding genes (Cytochrome b, cytochrome c oxidase subunits I, II and III, and ATPase subunits 6 and 8) and 7 protein encoding open reading frames (Chomyn *et al.*, 1983; Mariottini *et al.*, 1983; Michael *et al.*, 1984).

The relative locations of the tRNA genes, rRNA genes, replication origin and polyadenylated RNAs have been sequenced and mapped on the mitochondrial DNA molecule of *Drosophila yakuba* (Clary and Wolstenholme, 1985), the amphibian *Xenopus laevis* (Ohi *et al.*, 1978; Ramirez and Dawid, 1978; Rastl and Dawid, 1979), and in a handful of *Simulium* blackflies (Day *et al.*, 2016; Wang *et al.*, 2016; Wang and Huang, 2019; Zhang *et al.*, 2019). The gene content and order seem to be the same to a large extent in the mammalian mitochondrial genome, and also in the dipteran genomes (Clary and Wolstenholme, 1985).

4.1.2 Structure and function of mitochondrial genome

The mitochondrion is the powerhouse of a cell and has vital functions in respiration, ageing, genetic illness and self-destruction of a cell (Krauss, 2001; Alexeyev *et al.*, 2004; Wang and Youle, 2009; Tuppen *et al.*, 2010). Six genome types can be found in the mitochondrial genome depending on size, structure, presence of plasmid-like structures or introns, the presence or absence of molecules that are singular, homogeneous or heterogeneous (Kolesnikov and Gerasimov, 2012). It can be linear, as in unicellular organisms, or it can be circular as in animal cells (Nosek *et al.*, 1998; Kolesnikov and Gerasimov, 2012). Mitochondrial genome sequence and structure have provided information on patterns of gene flow and molecular evolution, as well as information on population genetics and phylogenetics (Wilson *et al.*, 2000; Salvato *et al.*, 2008). Like the mitochondrial genome of animals, those of insects are double stranded molecules with a size range of 14,503 bp to 19,517 (Lewis *et al.*, 1995). There are 37 genes encoding the large and small subunit ribosomal RNAs, 13 protein coding genes, and 22 tRNAs needed to translate the protein-coding genes (Faure and Casanova, 2006; Faure *et al.*, 2011). There is a core subset of mitochondrial genes that are found in all mitochondria sequenced to date, but there is

considerable heterogeneity in the total number of genes and in their arrangement. There are also some animals in which the mitochondrial genome is on more than one molecule. Notwithstanding, there appears to be high gene order conservation in the mitochondrial genome across most animal phyla. *Simulium* blackflies like most insects are exceptional to this gene order conservation because they have highly variable gene orders; and they are one of the 83 well-known insect species belonging to 11 orders that demonstrate rearrangement of the mitochondrial genes (Clary and Wolstenholme, 1985; Macey *et al.*, 1997; Shao and Barker, 2003; Thao *et al.*, 2004). Mitochondrial DNA are transmitted maternally, have a high mutation rate as a result of the limited repair system of about 5-10 times that of the nuclear DNA, and are characterised by the presence of a simple conserved structure (Brown *et al.*, 1979; San Mauro *et al.*, 2005). It is this faster evolution characteristic that makes animal mitochondrial genomes the main focus of phylogenetics and evolutionary works (Boursot and Bonhomme, 1986; Delsuc *et al.*, 2003; Hassanin *et al.*, 2013).

4.1.3 Benefits and demerits of mitochondrially encoded genes

It is generally believed that short amplicon barcoding genes are informative in molecular and phylogenetic analysis relating to species identification and delimitation. This include the use of the mitochondrially-encoded genes like CO1 and COII for diverse investigations in Simuliidae (Rivera and Currie, 2009; Pramual *et al.*, 2011; Hernandez-Triana *et al.*, 2012). Not only mitochondrial genes, but also nuclear genes being used alone or in combination with mitochondrial genes have been proven to have great use in Simuliidae (Krüger *et al.*, 2000; Thanwisai *et al.*, 2006; Phayuhasena *et al.*, 2010; Pramual and Nanork, 2012). Although the use of mitochondrial genes or genomes alone could be less informative, there are different ways of improving the likelihood of their utility. These include, but are not limited to, the combination of concordant data sets to enhance molecular resolution (Silva-Brandao *et al.*, 2005). In the worst case scenario, the lack of congruency between mitochondrial and nuclear data sets are best handled by excluding the mitochondrial data to enable the remaining nuclear data to yield more robust phylogenies (Sota and Vogler, 2001; Mallarino *et al.*, 2005). Through DNA fingerprinting, it becomes possible to identify species by sequencing and by phylogenetically analysing a single DNA fragment (Senatore *et al.*, 2014). The hallmark of this DNA fingerprinting is due to the stability and durability of DNA to such extent of making millions of years' fossil DNA useable, the ease of detecting damaged samples, and the timely determination of samples in any developmental stage of an organism (Harvey *et al.*, 2003). Mitochondrial genes have become the recognized standard fingerprinting marker

for lots of insects (Caterino *et al.*, 2000; Hebert and Gregory, 2005; Savolainen *et al.*, 2005), including *Simulium* blackflies.

Despite these merits of mitochondrial gene or genome utility, there has been criticism of its usage as a tool of significance in the delimitation of species (Rubinoff *et al.*, 2006). There are several reasons for these criticisms. These include the issues with its characteristic disproportionate level of amino acid conservation (Russo *et al.*, 1996), gene flow characterized by male-biased maternal inheritance (Ballard, 2000b; Moritz and Cicero, 2004), horizontal gene flow among and between species (Ballard, 2000b; Ballard, 2000a), problems of heteroplasmy (Frey and Frey, 2004) and the well-known retention of ancestral polymorphisms (Moritz and Cicero, 2004). Other reasons for the criticisms include the nuclear integration of mitochondrial DNA (numts) (Richly and Leister, 2004; Song *et al.*, 2008), bacterial infection like the presence of *Wolbachia* sp. that bias mitochondrial DNA variation (Hurst and Jiggins, 2005), and duplication found in the mitochondrial genome that is known to contribute to challenges in mitochondrial genome assembly (Campbell and Barker, 1999).

Decades of mitochondrial DNA (mtDNA) research has led to improvements in making inferences in molecular ecology and phylogeography. Despite the invaluable information derived from this organelle, there are issues worth considering with the potential to cause bias in their utilization when it comes to inferring demographic features of populations and the evolutionary history of species. Unlike plant and fungal mtDNA, most animal mtDNA do not usually undergo recombination (Birky, 2001) and this lack of recombination coupled with the absence of physical linkage of all sites often lead to major challenges in traditional genetics. Not too distant studies have suggested the possibility of recombination and rearrangements in the mitogenome (Awadalla *et al.*, 1999; Eyre-Walker & Awadalla 2001; Innan & Nordborg 2002; Dowton *et al.*, 2003). Some species in the Animal kingdom, such as the mussel *Mytilus galloprovincialis*, have showed mitochondrial recombination (Ladoukakis and Zouros, 2001).

On the average, mtDNA has lower effective population size, but this observation is not always true for many species (Ballard and Whitlock, 2004). Most species with strong sexual selection however have a higher effective population size, and for all species there exist instances such as during selective sweeps on the mtDNA, when the relative effective size of the mtDNA becomes very low (Ballard and Whitlock, 2004).

Mutational biases have the tendency to influence the evolution of mtDNA partly because there exists a strong heterogeneity of mutation rates in the hypervariable region (Penny *et al.*,

1995; Malyarchuk *et al.*, 2002), as well as in other parts of the mtDNA (Ballard, 2000c). There is the occurrence of a strand-specific substitution bias (Anderson *et al.*, 1981; Clarey & Wolstenholm, 1985; Garesse, 1988; Rand & Kann 1998; Ballard 2000a). Such bias can potentially cause issues in phylogenetic and phylogeographical interpretations, especially when pooling coding sequences from different strands and when conducting sliding window statistical analysis by the use of the maximum likelihood. There also exists biases toward A/T-ending codons in the mtDNA of many species, and it may be the result of factors that impinge on rates of DNA damage (Martin, 1995) and the availability of each nucleotide present in the cellular medium of the mitochondrion (Xia *et al.*, 1996). It is recognised that most mtDNA studies estimate the phylogenetic and demographic history of species (Ballard and Whitlock, 2004), and some of the interpretations have errors such as the error in inferring gene genealogy from available data, error in determining a typical genealogy from that of a single molecule, and error in inferring demographic history from gene genealogies. Many people have attempted to solve these problems, especially in differentiating between species and gene trees (Edwards & Beerli, 2000; Nichols, 2001, and Hudson & Turelli, 2003). Ballard and Whitlock (2004) pointed out that gene genealogies inference always have error, but the error in reconstructing the true mtDNA genealogy tends to become small when one uses enough sequence. Their review further points out that it can take more than a single gene to infer a gene tree with accuracy (Cummings *et al.*, 1995).

The genealogy of any given molecule, such as the mitogenome, is merely a random selection from an extremely variable distribution of genealogies, and it may have many possible outcomes given the actual history of the species or population (Pamilo and Nei, 1988; Edwards and Beerli, 2000; Nichols 2001).

The mitochondrial genome is relatively small in comparison to the nuclear genome. It is usually not likely to obtain a well-resolved gene or species tree even with complete mitogenome data (Degnan, 1993; Slade *et al.*, 1994). It is recommended that nuclear DNA can provide an alternative, in addition to the mitogenome data, due to its statistical near-independence of unlinked sites. Notwithstanding the need to draw inference from both mtDNA and nDNA sources of genetic data, there are always the usual numerous conflicts between mitochondrial and nuclear data genealogically-derived inferences, which are often incongruent especially at the species level (Ferris *et al.*, 1983; Powell 1983; Bernatchez *et al.*, 1995; Taylor & McPhail 2000; Lu *et al.*, 2001; Shaw 2002; Sota 2002; Rognon & Guyomard 2003; Seehausen *et al.*, 2003). Obviously, mtDNA and nDNA result in genealogical inferences that may or may not be concordant and this remain a constant debate.

4.1.4 Role of next generation sequencing in whole mitochondrial genome sequencing

Amplification techniques in the polymerase chain reaction (PCR) have been greatly used to amplify targeted sections of both recently extracted DNA and fossil specimens of ancient past (Shapiro *et al.*, 2002; Dalén *et al.*, 2007; Poulakakis *et al.*, 2008). Researchers sometimes need to use a small quantity and poor-quality DNA for their research works due to limited resource availability and challenges with sample management including issues with storage. They often attempt to amplify and sequence various overlapping short fragments of the highly degraded and fragmented DNA to enable them assemble nucleotide sequences of sufficient lengths for downstream analysis. The consequence of using such a small quantity and poor quality of DNA often leads to high rates of PCR failure, and a waste of both time and resource (Rowe *et al.*, 2011). As a result of these challenges with short amplicon PCRs, mitochondrial genomes become a major focus for PCR-based amplification and analysis due to their high copy numbers when compared with those of nuclear genomes. In addition, there are more published references available for designing primers for the mitochondrial genomes than for the nuclear genes (Rohland *et al.*, 2004; Ho and Gilbert, 2010). Despite these advantages, there are other shortcomings that hinder mitochondrial genome amplification by the strategy of using overlapping short amplicons or in a single amplification attempt. The presence of nuclear copies of mitochondrial genes (numts) (Lopez *et al.*, 1994) and contamination by exogenous DNAs potentially lead to the amplification of non-target PCR products (Ho and Gilbert, 2010). Universal primers developed from conserved genomic regions could preferentially amplify numts rather than the targeted mitochondrial fragments (Collura and Stewart, 1995). Also, attempts to use a conventional PCR strategy to amplify mitochondrial genomes from long-term stored DNA samples can potentially lead to miscoding lesions during the PCR amplification process (Gilbert *et al.*, 2003), and this bias can mislead the interpretation of demographic histories in DNA studies (Axelsson *et al.*, 2008).

The high-throughput sequencing technologies that are popularly referred to as next generation sequencing (NGS) are able to produce millions of short sequences simultaneously, help overcome challenges associated with DNA that are generally considered problematic, and are ideal for sequencing both fragmented DNA and whole genomes (Millar *et al.*, 2008; Knapp and Hofreiter, 2010). Eukaryotic cells contain high copy numbers of mitochondrial DNA that suggest that a few gigabases of high-throughput data would potentially be sufficient to

recover a complete mitochondrial genome at great sequencing depth. In addition, the high sequencing depth of NGS was proposed to be capable of reducing the effect of numts or other non-target products that usually produce incorrect sequences. Also, high-throughput haploid mitochondrial genomes are known to be appropriate for reducing the confounding effect of heterozygosity in diploid genomes (Gilbert *et al.*, 2007; Ho and Gilbert, 2010; Hung *et al.*, 2013).

4.1.5 Justification (Rationale) of the study in chapter four

The WHO guidelines suggest that river basins constitute onchocerciasis transmission zones. Despite the high genetic diversity observed in chapter three, the samples could still not be divided into distinct groups by river basin. There are two possible reasons. First, there might not be any structure. Second, and the reason for attempting to use whole mitochondrial genomes, is that there may be structure but that both the short amplicon barcoding gene sequences and the long-ranged amplicon sequences were not capable of detecting it. So, although the long amplicon data did not reveal any population structure by river basin, perhaps sequencing the most informative part of the mitochondrial genome might reveal that structure.

Generally, mitochondrial markers have phylogenetic advantages because of their fast rate of evolution, lack of recombination and maternal inheritance (Brown *et al.*, 1979; Avise, 1994). Being part of the mitochondrial genome, the control region or D-loop has the additional benefit of being used as a marker sequence for genetic diversity analysis due to its exceptionally fast evolutionary rate (even when compared to the rest of the mitochondrial genome) (Meyer, 1993; McMillan and Palumbi, 1997), presumed selective neutrality as a non-coding region (Pereira *et al.*, 2008; Rech *et al.*, 2014) and polymorphic nature (Ghatak *et al.*, 2016). The control region of the mitochondrial genome has had major phylogenetic significance in its use as a genetic marker in various vertebrates like reptiles (Jiang *et al.*, 2011), fish (Jamandre *et al.*, 2014; Beltrán-López *et al.*, 2017), birds (Kryukov *et al.*, 2017), mammals (Firestone, 2000; Boyko *et al.*, 2009) and amphibians (Huang and Tu, 2016), as well as in invertebrate taxa (Shao *et al.*, 2005; Diniz *et al.*, 2005; Bronstein *et al.*, 2017; Zhang *et al.*, 2017). Despite the benefits of the D-loop, its amplification and sequencing is often difficult due to a number of reasons such as the presence of repeat regions found in some species (Ludwig *et al.*, 2000; Irwin *et al.*, 2009), the occurrence of segmental duplications in some species (Bensch, 2000; Shao *et al.*, 2005; Nittinger *et al.*, 2005), likely homogenization between duplicated copies of the D-loop (Eberhard *et al.*, 2001; Cadahía *et*

al., 2009), and the formation of pseudogenes that become co-amplified by PCR with the control region (Singh *et al.*, 2008; Cadahía *et al.*, 2009).

This kind of anticipated challenge in attempting to amplify the control region of the mitochondrial genome was encountered during the initial stages of this study. Compounding the issue was the absence in any known public database of D-loop nucleotide sequences of *Simulium* blackflies that had significant sequence length, and the absolute lack of complete mitochondrial genome sequences in any public database at the start of this study in 2015. This made primer design difficult. It was therefore necessary to experiment with the use of total DNA to prepare libraries for whole genome amplification utilizing Next Generation Sequencing strategies. The success of this whole genome sequencing would contain the D-loop sequences, and probably both the mitochondrial genome and some nuclear gene sequences. Furthermore, such nucleotide sequences with anticipated greater diversity could enable the characterization of the genetic diversity of *Simulium* blackflies, as well as the delineation of their transmission boundaries.

4.1.6 Specific objectives, mode of assessment and hypotheses

4.1.6.1 Specific objectives

The specific objectives of study in this chapter are:

1. to sequence and assemble whole mitochondrial genomes of *Simulium* blackflies from Africa
2. to characterize the genetic diversity of blackflies in the central ecological transition zone of Ghana
3. to determine if there is population structure of the blackflies in the central ecological transition zone of Ghana
4. to use the complete mitochondrial genomes of the blackflies in the transition zone of Ghana to determine the transmission zone status of the study areas

4.1.6.2 Mode of assessment of specific objective

A similar mode of assessment as used in section 3.1.3.2 was used to assess the specific objectives of this chapter. In addition, the assembled mitochondrial genomes were compared to the mitogenomes of other blackflies and closely related organisms in public databases.

4.1.6.3 Hypotheses of the study in this chapter

Null hypothesis 1

The use of a library prepared from total DNA and low coverage sequencing will enable the assembly of complete mitochondrial genomes for blackflies in Africa

Alternative hypothesis 1

The use of a library prepared from total DNA and low coverage sequencing will not enable the assembly of complete mitochondrial genomes from blackflies in Africa

Null hypothesis 2

Complete mitochondrial genomes from blackflies in the transition zone of Ghana will enable the characterization of genetic diversity of the *Simulium* blackflies

Alternative hypothesis 2

Complete mitochondrial genomes from blackflies in the transition zone of Ghana will not enable the characterization of genetic diversity of the *Simulium* blackflies

Null hypothesis 3

The mitochondrial genomes will show population structure by river basin in the central ecological transition zone of Ghana

Alternative hypothesis 3

The mitochondrial genomes will not show population structure by river basin in the central ecological transition zone of Ghana

Null hypothesis 4

Complete mitochondrial genomes of blackflies will enable the identification of transmission zone status of the *Simulium* blackflies in the central ecological transition zone of Ghana

Alternative hypothesis 4

Complete mitochondrial genomes of blackflies will not enable the identification of transmission zone status of the *Simulium* blackflies in the central ecological transition zone of Ghana

4.2 Materials and Methods

4.2.1 Sample Collection

A total of 82 *Simulium* blackflies were selected for whole genome sequencing. The samples were collected from Daka (25) in the East, Pru in the Central (25), and Black Volta (25) river basins in the West of Ghana transition ecological zone, and Pra (1) river basin in South-Western Ghana, all in West Africa; Lingoni Falls river basin in Malawi of South-Eastern Africa (1); and Nkam and Mbam river basins in Cameroon of Central Africa (5). These study sites are shown in Figure 4.2.1.

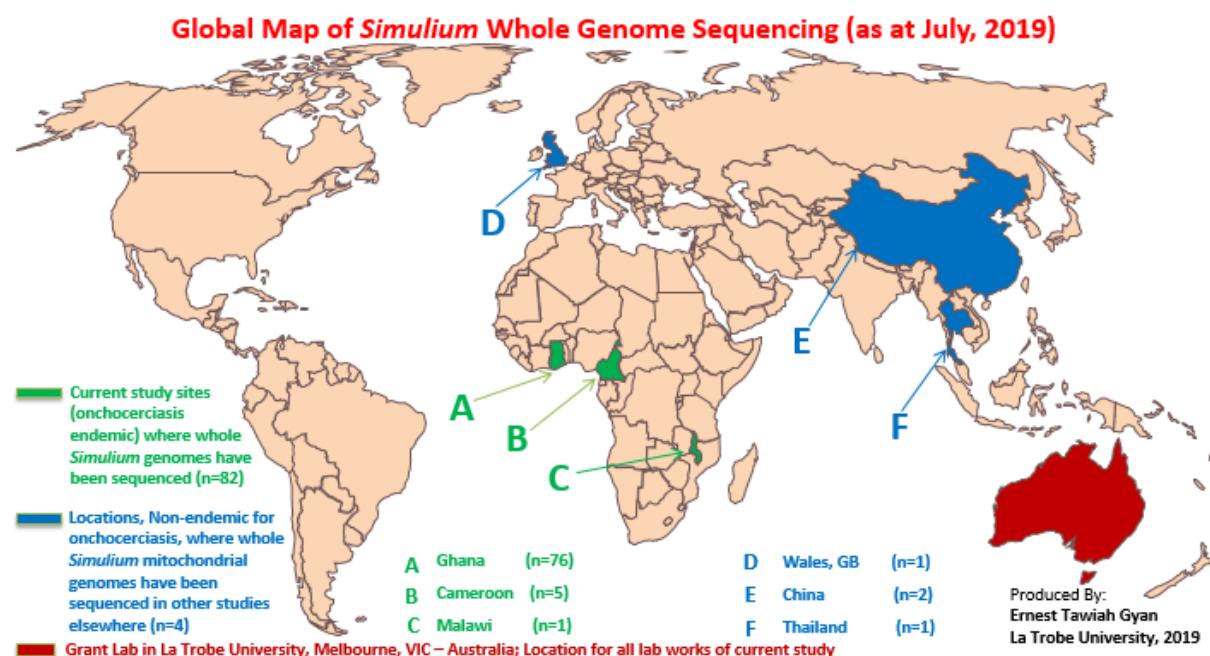


Figure 4.2.1: Global map of *Simulium* whole genome sequencing. The map shows the locations where *Simulium* blackflies were collected and their genomes sequenced in the past (blue) by other researchers elsewhere, and the present (green) in this study. The maps were developed with Microsoft PowerPoint 2016, Microsoft Word 2016 and Microsoft OneNote 2016.

4.2.2 DNA Extraction

DNA was extraction using the Isolate II Genomic DNA Kit (Bioline) by following the protocol described in section 3.2.3 of this thesis.

4.2.3 DNA concentration determination with Qubit® Fluorometer

The concentration of the extracted DNA was assessed using the procedure shown in the section 2.2.4 of this thesis.

4.2.4 Library Preparation Using Nextera DNA XT and DNA Flex Library Prep Procedures

The Nextera DNA XT Library preparation procedure described in Section 3.2.6 of this thesis was used to prepare 12 *Simulium* libraries from gDNA. These library preparations were carried out in 2 separate MiSeq runs with the first one consisting of 2 samples and the output named as MiSeq1, while the second run consisted of 10 samples and the output named as MiSeq 2.

The Nextera DNA Flex Library Prep Kit was used to prepare a further 70 libraries by following the manufacturer's protocol.

4.2.5 Genomic DNA Tagmentation

Briefly, an amount of 2-30 μ l DNA was added to each well of a 96-well PCR plate such that the total input amount was 100 ng. Where the initial DNA was less than 30 μ l, it was adjusted to 30 μ l using HPLC water. The Bead-Linked Transposomes (BLT) were resuspended by vortexing vigorously for a total of 20 seconds using approximately 2.5 seconds pulse vortexing technique to ensure the preservation of the efficacy of the transposomes. The rest of the protocol was followed without any modification.

4.2.6 Library Quantification and Quality Assessment

The concentration of the libraries was quantified using the procedure described in section 2.2.4 of this thesis. Quantitative Polymerase Chain Reaction (qPCR) was used to assess libraries likelihood of being able to generate clusters by measuring templates that have both adaptor sequences on either end which will subsequently form clusters on a flow cell (Illumina, 2019).

Library size and concentration quality control assessment was carried out by electrophoresis using Tapestation, the Agilent High Sensitivity D5000 Screen Tape System, and following the manufacturer's protocol.

4.2.7 Normalization and Library Pooling

After quantifying the libraries individually, they were each normalized to the same 10nM concentration and pooled by combining 5 μ l of each library in a 1.7 ml microcentrifuge tube.

The final library was quantified a final time using the procedure described in section 2.2.4 of this thesis.

4.2.8 NextSeq Sequencing

The 70 libraries were sequenced on an Illumina NextSeq sequencer using the 2x150bp paired end reads chemistry. It was carried out in 2 sequencing runs, with the first one named NextSeq 1 (24 sequences), and the second one named NextSeq 2 (46 sequences).

4.2.9 Sequence Import

The Illumina High-Throughput Sequencing Import tool option in the CLC Genomics workbench 9.5.4 was used to import the next generation sequences produced for each of the MiSeq and NextSeq illumina sequence runs. The paired reads option was chosen for the general setting to allow the import and easy identification of paired read sequences. The original read names and quality scores were not discarded. For each import, the Paired-end (forward-reverse) option was selected whiles the minimum distance and maximum distance were set to 1 and 1000 respectively. As a quality measure, failed reads were set to be removed. The NCBI/Sanger or Illumina Pipeline 1.8 and later option was chosen for the Quality score setting, to enable comparability of quality scores between Sanger and high-throughput sequences of other studies. For the batch processing of multiple individuals such that each individual sequence reads will be clearly identified and well organized, the subfolders per batch unit option was set to be created under the Result handling section.

4.2.10 Trimming and removing duplicate sequences

Duplicated sequences resulting from amplification processes prior to sequencing, such as sequence enrichment, were removed with the NGS core tools duplicate sequence removal. As a stringency for merging reads with sequencing errors for removal, a 20% maximum representation of minority sequence was used because the majority of the sequences (>98.5%) had Phred quality above 20; thus being capable of collapsing reads at that threshold that would only have been distinguished apart by sequencing errors such that anything else would be attributed to true biological variation (or PCR errors in the early cycles that might be indistinguishable from true variation). The CLC's NGS core tools duplicate sequence removal required the use of mapped sequences. An alternate way of removing the duplicated sequences was to complete an initial *de novo* assembly, followed by the use of the list of duplicated sequences generated from the assembly to remove such sequences from the original raw reads, and then restarting the *de novo* assembly once again with the raw reads that has the duplicated reads removed.

Different trim options were used for each of the illumina high-throughput sequences in a pilot sequence training process, followed by draft mitochondrial genome assembly before choosing the best trim options to use for the rest of the sequences. Following the adapter trimming, the following trimming options were explored: quality trimming, length trimming, ambiguity trimming to trim off stretches of Ns, and base trimming to remove a specified number of bases at either the 3¹ or 5¹ ends of the reads.

Although the adapters were directly removed on the sequencing machine, a validation step was put in place by carrying out adapter trimming process with the hypothesis that if the adapter removal process by the sequencing machine was efficient, then the subsequent adapter trimming process using CLC Genomics trimming tools should yield no additional trimmed sequences. Firstly, an adapter list that consisted of all the adapter sequences used for the sequencing was created and supplied to the trim tool. The adapter trimming parameters were as follows. Alignment scores costs of 2 for the mismatch cost, and 3 for the gap cost were used. The match thresholds were set to allow internal matches with a minimum score of 10, whiles allowing end matches with a minimum score of 4 at the end. The Remove adapter option was set as the action to perform when an adapter match was found, following the Workbench performing a Smith-Waterman alignment (Smith and Waterman, 1981).

For the length trimming, nucleotide bases at the 5¹ and 3¹ terminal nucleotides were not removed because poor quality reads at such ends were set to be taken care of automatically by both the quality trimming and the ambiguity trimming. The phred quality scores (Q), which is defined as: $Q = -10\log_{10}(P)$, where P is the probability of base-calling error; were converted to error probabilities. The error probabilities were then used to set the limit for the bases to be trimmed. For every base, a new value was set to be calculated by finding the difference between the expected limit and the base-calling error probability. The running sum of this value was set to be calculated. Where the sum dropped below zero, it was set to zero. Thus, the parts of the sequence not trimmed were the sequence regions ending at the highest value of the running sum and beginning at the last zero value just before this highest score value. Every sequence before and after this region was trimmed. Thus, reads whose scores were not above zero were completely removed. Ambiguous nucleotides were trimmed using 2 as the maximum number of ambiguities.

4.2.11 Contamination removal

To test for the presence of contaminants and remove them, the reads were mapped to genome sequences of bacteria, virus, eukaryote (human and nematodes), fungi and protozoa (Table

4.3.3). The organisms were *Wolbachia pipiensis* (accession: CP042444.1), *Wolbachia phage* (accession: AB036666.1) *Onchocerca ochengi* (accession: NC_031891.2), *Onchocerca volvulus* (accession: KT599912.1), *Homo sapiens* (accession: JX669269.1), *Fusarium bambusae* (accession: MH684411.1), and *Plasmodium falciparum* (accession: KY923425.1). The un-mapped reads were collected and used for downstream analysis. The similarity fraction was chosen as a proportion to correspond to one value (the upper 1 standard deviation) above the threshold of percentage identity of each of the corresponding organism with *Simulium sp.* (accessions NC_029753.1, NC_040120.1, and NC_033348.1). Such a choice ensured that only sequences of the organism in the contamination sequence reference list mapped and left out all other sequences as unmapped reads for the downstream analysis. To test the sensitivity of the use of the organisms on the contamination sequence reference list as representative of their respective taxonomic group, 3 methods were used. The first method was to align 300 *Homo sapiens* sequences and to extract the consensus sequence to be used as a query sequence in a blastn alignment with each of the mitochondrial genomes of 3 *Simulium sp.* (accession numbers NC_029753.1, NC_033348.1, and NC_040120.1). The percentage identity from each of the 3 alignments with *Simulium sp.* were compared. The second method was to use the longest sequence of the mitochondrial genome of *Homo sapiens* (accession JX669269.1) as the query sequence in the blastn alignment as done above with the first method. The percentage identity was noted. The third method was to randomly choose 10 sequences from the list of the 300 *Homo sapiens* whole mitochondrial genome sequences. Each of these 10 sequences were individually used as query sequences for blastn alignment with the 3 whole mitochondrial genome sequences of *Simulium sp* as done in the previous 2 methods. The percentage identity values from all the 3 methods were compared and the lowest value selected. A 1 standard deviation above and below this lowest value of percentage identity was used for the threshold validation for the similarity index of the reads mapping to remove contamination. When all the 3 methods yielded values not significantly different from each other ($P > 0.05$), the use of the 3 methods for the identification of the threshold of percentage identity was not repeated with the remaining organisms on the contamination sequence reference list. Instead, each of the organisms on the contamination sequence reference list were used as representative of their respective taxonomic groups. All other mapping parameters were as used in the previous sections.

4.2.12 Assembly strategy and parameters

The de Bruijn graph algorithm, employed in the CLC Genomics workbench 9.5.4 was used for the *de novo* assembly and it is similar to how most new *de novo* assembly algorithms

work (Zerbino and Birney, 2008, Zerbino *et al.*, 2010, Gnerre *et al.*, 2011). The *de novo* assembly was done individually for the individual fly sequences due to the need to vary the assembly parameters for the sequences with varying average lengths. In the assembly graph parameters, automatic word size was used to enable the de Bruijn graph algorithm to calculate the appropriate word size from the reads. Similarly, automatic bubble size detection was used due to all the trimmed samples having more than 95% of reads with a phred score above 30, and also to enable the assembly algorithm to select the appropriate bubble size. A minimum contig length of 200 was used initially for the assembly. The assembly was repeated several times with each succeeding assembly having a minimum contig length with an increased value by 100-200 to that of the previous assembly (e.g. 200, 400, 600, 700 ... 13000, etc.) until the optimum value of 13000 was obtained as the best value to assemble a complete mitochondrial genome at a faster rate in less than 30 minutes (when run on a Windows 10 Enterprise 64-bit laptop, with 8192MB RAM, 2.30GHz [4 CPUs]). The paired reads were set to auto-detect paired distances, and to perform scaffolding for ease of obtaining valuable paired distance statistics and for identifying the contigs with length similar to the theoretical size of the *Simulium* blackfly mitochondrial genome. The assembly mapping options used were: map reads back to contigs (slow), which is known to create assemblies with higher rate of accuracy despite requiring longer time to complete, mismatch cost of 2, insertion cost of 3, deletion cost of 3, length fraction of 0.5, and similarity fraction of 0.86. The 0.86 value for the similarity fraction was used because an NCBI nucleotide sequence blast with *Simulium* whole mitochondrial genomes indicated that the lowest percentage identity for *Simulium* sp. is 85.98%, which corresponded to 0.86 similarity fraction. The use of anything below this 0.86 similarity fraction threshold value was inappropriate for the *Simulium* species. Mitochondrial genomes were identified by screening contigs of consensus sizes between the hypothesized genome size of most insects (somewhere between 13000 and 17000), and picking the contig with the greatest number of both total read count and average coverage (as shown in figure 4.2.2) for further assessment. For instances that the contigs being screened resulted in numerous contig numbers greater than 10, the screening range was gradually reduced from the upper and lower boundaries until the contig with the greatest number of both total read count and average coverage was found. Whereby the size of this contig was in the size range of the expected mitochondrial genome, the consensus sequence of the contig was extracted and used as query sequences in BLASTn against the nucleotide database of NCBI. A percentage identity match of 86% and above to only *Simulium* whole mitochondrial genomes in the NCBI nucleotide database but not to the genomes of any other species with such high

level of percentage identity, led to the inference of the query contig being likely of the mitochondrial genome of a *Simulium* blackfly from the sampling area. When the mitochondrial genome was found complete, such a sequence was followed up with further *Simulium* identity confirmation. The 2 assemblies in the 1st MiSeq run produced mitochondrial genomes that were of similar sizes to other complete *Simulium* mitochondrial genomes originating from *Simulium aureohirtum*, *Simulium maculatum*, and *Simulium variegatum* (NCBI accession numbers NC_029753.1, NC_033348.1, and NC_040120.1). The gene contents, order and orientations of the 2 assemblies from the 1st MiSeq run were analysed and compared with those of *S. aureohirtum*, *S. maculatum*, and *S. variegatum*. Flow diagram of the assembly process can be found in Figure 4.2.2.

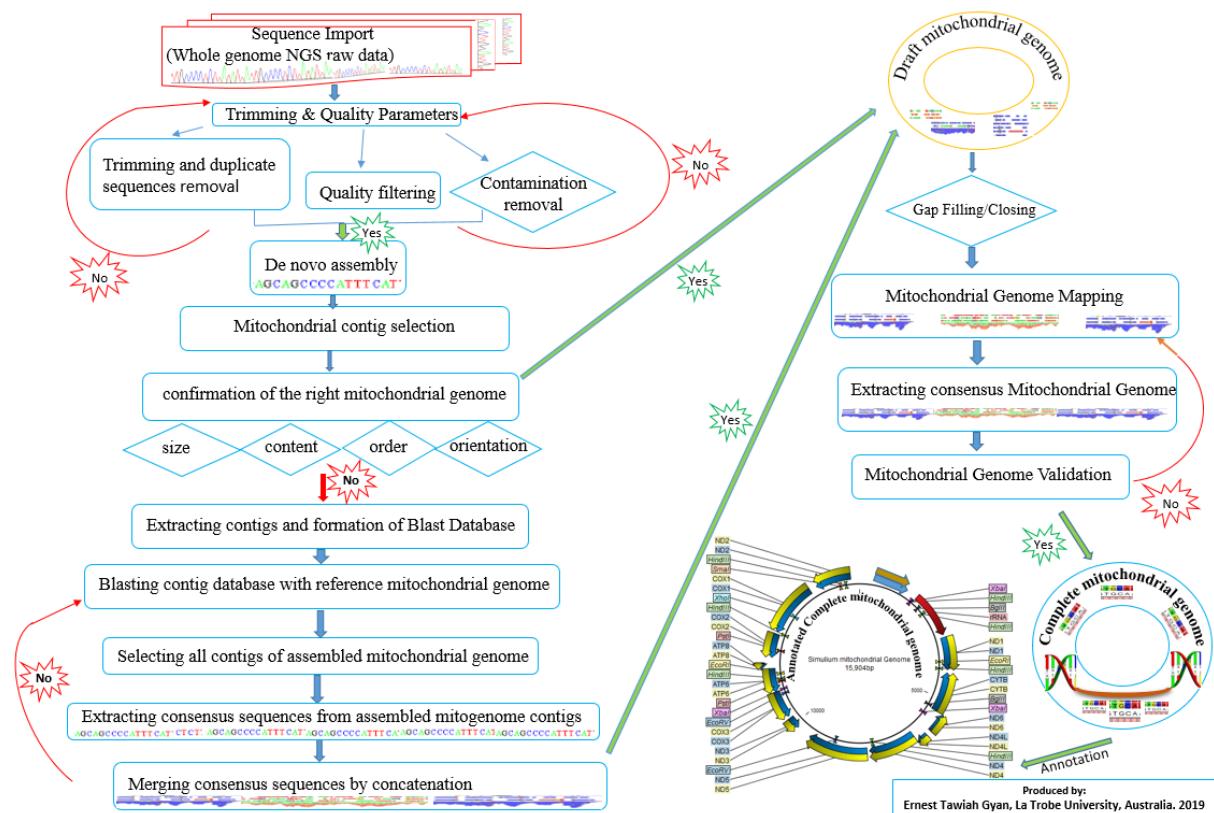


Figure 4.2.2: Flow diagram of the assembly process for generating whole mitochondrial genomes of African *Simulium* blackflies. The assembly begins with importing NGS sequences into CLC genomics workbench 9. After setting appropriate trimming and quality parameters, the raw reads were trimmed, and quality filtered. Duplicate sequences and those from non-*Simulium* organisms (contaminants) were removed. The subsequent reads underwent *de novo* assembly. Mitochondrial contigs were selected. After confirming the size, content, order and orientation of the selected mitochondrial contig, its resulting consensus sequence became draft mitochondrial genome. Two of the assemblies produced draft

mitochondrial genomes without any gaps, hence became the first reference mitochondrial genomes after their validation. The complete mitochondrial genomes were finally annotated.

4.2.13 Mitochondrial Genome Validation

The draft *Simulium* mitochondrial genome sequences were validated by paired end read mapping. Thus, the paired end reads that were used for the *de novo* assembly were mapped onto the mitochondrial genome sequences. Coverage and depth of mapped reads were studied using the CLC Genomics Workbench 9.5.4 and the associated calculations performed in Excel 2016. Coverage analysis was carried out using minimum length of 50 and *p*-value threshold of 0.0001. A confirmation was made of the consistency and connectivity of the mapped reads. The junctions between seed contigs were also confirmed. Additional validation by PCR may be needed and may be carried out in a follow up study.

4.2.14 Mitochondrial Genome Mapping and Gap Filling

The confirmed whole mitochondrial genomes of the *Simulium* sp. from the 1st MiSeq run were used as reference sequences for mapping of raw reads of other individual samples that had low coverage and gaps from their *de novo* assemblies. The mapping parameters were as used in the *de novo* assembly. The sections of the relatively low coverage assembled sequences with gaps were completely filled by mapping the raw reads of that individual fly to a reference sequence that was produced from a blackfly sampled from Ghana. The mitochondrial genomes were validated using the method described in the section 4.2.15 of this thesis.

4.2.15 Extracting Consensus sequence from assembly

From each *de novo* assembly was extracted consensus sequence for each individual blackfly. The parameters used were low coverage threshold of zero, insertion of ‘N’ ambiguity symbols for low coverage sections and resolving nucleotide base conflicts using both quality score and voting for the nucleotide base with the highest count.

4.2.16 Sequence prediction, annotation and validation

The mitochondrial genomes of the *Simulium* sp. were annotated with MITOS (Bernt, M., *et al.*, 2013) using the settings: genetic code of 05-Invertebrate, BLAST E-value Exponent of 2, cutoff of 50, maximum overlap of 20, clipping factor of 10, fragment overlap of 20, fragment quality factor of 10, start/stop range of 6, and final maximum overlap of 35. They were

validated with Geneious 8.1.3 (Biomatters, Auckland, New Zealand) by using the mitochondrial genomes of *Simulium aureohirtum* (NC_029753.1), *Simulium maculatum* (NC_040120.1), and *Simulium variegatum* (NC_033348.1) as references. Although the tRNA genes were predicted with tRNAscan-SE 2.0 (Chan and Lowe, 2019), the MITOS Web Server was used to identify both the tRNA and rRNA genes. All coding genes were predicted with the ORF Finder in NCBI (<https://www.ncbi.nlm.nih.gov/orffinder>) and later modified by comparing with previously published mitochondrial genomes of Simuliidae.

4.2.17 Variants calling

Variants were called on individual blackfly sequences using 3 main methods in CLC Genomic Workbench 9: (1) the Basic Variant Detection (BVD); (2) Fixed Ploidy Variant Detection (FPVD); and (3) the Low Frequency Variant Detection (LFVD). Concordant variants between the 3 methods of variant detection were identified and used for downstream analysis. The Basic Variant Detection finds variants in read mappings by simple counting without utilizing any statistical model. The Fixed Ploidy Variants Detection finds variants in read mappings by utilizing (1) a model based on the specified ploidy of the sample and (2) a sequencing error model with parameters estimated from the data. The Low Frequency Variant Detection finds variants in read mappings using (1) a model which allows for the presence of low frequency variants in the sample and (2) a sequencing error model with parameters estimated from the data.

For all the 3 variant detection methods, the variant parameter was chosen such that a ploidy of 1 for haploid was used for all the mitochondrial genes and genome, while a ploidy of 2 for diploid was used for all the nuclear genes. A 90% required variant probability was used for the FPVD. Broken pair reads, and non-specific read matches were excluded in all the methods. A minimum reads coverage of 50 with a minimum count of 2 were used. A minimum frequency of 1%, 20% and 35% were used for the LFVD, FPVD and BVD methods respectively. Noise filters were set through quality filters, direction and position filters, and technology specific filters. The quality filters utilized 5 neighborhood radius, 20 minimum central quality and 16 minimum neighbourhood quality. The direction and position filters used: read direction filter with 5% direction frequency, relative read direction filter of 1% significance, and read position filter of 1% significance. The technology specific filters were set to remove pyro-error variants: (a) in homopolymer regions with minimum length of 3, and

(b) with frequency below 0.8. The pyro-error variants removal tool is used to remove insertions and deletions in the reads that are likely to be due to pyro-like errors in homopolymer regions. The neighborhood radius is the region size for quality filter, and it is defined as the distance to both sides of the central nucleotide. The minimum neighbourhood quality deals with the minimum average quality score for the region around the central nucleotide.

4.2.18 Molecular phylogenetic and statistical analysis

File conversion between different file formats were carried out using PGDSpider 2.1.1.2 (Lischer and Excoffier, 2012). The best nucleotide substitution method was determined by computing and comparing the Bayesian information criterion (BIC), minimum theoretical Akaike information criterion (AIC), corrected minimum theoretical Akaike information criterion (AICc), and using the Hierarchical likelihood ratio test with a confidence level of 0.01. Generally, the model with the lowest BIC, AIC, or AICc value was chosen as the best-fit. The best-fit model was used to compute the final phylogenetic tree. Phylogenetic tree construction types used were Maximum likelihood tree, Maximum Parsimony tree, Minimum-Evolution tree, and Bayesian tree building with Mrbayes. Where applicable, the bootstrap method was used with 1000 replications.

The substitution models tested were Jukes-Cantor (JC), Felsenstein 81 (F81), Kimura 80 (K80), Hasegawa-Kishino-Yano (HKY), General Time Reversible (GTR), Kimura-2-parameter, Tajima-Nei, Tamura 3-parameter, and Tamura-Nei. The occurrence of +G in the model name denotes rate variation (4 categories), whiles +T in the model name denotes topology variation. The intention of the hierarchical likelihood ration test (hLRT) was that the best-fit model would be used to compute the final phylogenetic tree by allowing topology variation. In the model testing, 2 models were assessed at each particular time by comparing the log likelihood of the current best-fit model (the null hypothesis) with that of the alternative hypothesis model (the alternative hypothesis), such that the one with the highest statistically significant ($p < 0.01$) log likelihood was chosen and the other rejected. Thus, the null hypothesis was rejected when the p -value was not above 0.01. The chosen model then became the new null hypothesis to be compared with another model (the next alternative hypothesis) and the process repeated until all the test models were rejected and only the one with the highest log likelihood became accepted as the best-fit model.

In the Bayesian information criterion (BIC) model test, the BIC was computed using the formula $BIC = -2 \ln(L) + K \ln(n)$. The Bayesian weights were computed as $\exp(-\Delta/2) \sum_i (\exp(-\Delta_i/2))$, where \sum_i is the sum over all the models. Models were sorted in increasing order of BIC value, and the model with the lowest BIC value became the best-fit model.

From the Akaike information criterion (AIC) model testing, models were sorted in increasing order of AIC value. The model with the lowest AIC value was chosen as the best-fit. Where the ratio between the length of alignment and the number of parameters was less than 40, the corrected Akaike information criterion (AICc) results were preferred over the AIC results. The AIC was computed using the formula $AIC = -2\ln(L) + 2K$. The Akaike weights value was computed as $\exp(-\Delta/2)/\sum_i (\exp(-\Delta_i/2))$, where \sum_i is the sum over all the models.

Multiple phylogenetic trees that indicated similar blackfly ancestral relationships were critically examined and a representative tree with the overall best bootstrap values chosen. Relationships between individual genotypes at the population level were further examined by haplotype networks with Popart 1.7 (Bandelt *et al.*, 1999). Genetic distances, diversity indices and molecular variance were determined with Popart 1.7, MegaX 10.0.4, and Arlequin 3.5.2 (Excoffier, 2015).

Muscle (Edgar, 2004) and ClustalW (Thompson *et al.*, 1994) were used to align the nucleotide sequences. For Muscle alignment, the following settings were used: 16 maximum iterations, 1000 maximum hours, and 1000 maximum memory. For ClustalW, Guide Tree Algorithm was set to normal. The aligned sequences were exported to R x64 3.5.0, MegaX (Kumar *et al.*, 2018), DNAsP version 6 (Librado and Rozas 2009), and Popart-1.7 (Bandelt *et al.*, 1999).

Phylogenetic trees were constructed in MegaX (Kumar *et al.*, 2018), CLC Genomics Workbench 9.5.4, and the High-Performance Computing (HPC) of La Trobe University. For each alignment, the phylogenetic trees constructed were Maximum Parsimony, Maximum Likelihood, Minimum-Evolution, Raxml and Phym (Guindon and Gascuel, 2003). Bayesian tree building were carried out with MrBayes (Ronquist *et al.*, 2012) on the La Trobe HPC. Haplotype networks were produced using Minimum Spanning Network. Basic population genetics analysis carried out in PopART included: nucleotide diversity (Nei, 1987),

segregating sites, parsimony-informative sites, Tajima's D (Tajima, 1989), and Analysis of Molecular Variance (AMOVA). The Tajima's D was applied to know if the population was evolving neutrally, and to seek evidence of population expansion or bottleneck. Gene flow (Nm) was evaluated from GammaSt (Nei, 1982). The estimation of the genetic diversity within defined groups of taxonomic units were carried out with the nucleotide diversity, whiles the genetic distances between populations was estimated by calculating the number of net nucleotide substitution between populations (Nei, 1987). Moreover, the genetic distance estimation between populations was augmented by the use of Nei's standard genetic distances (Nei, 1972) between all pairs of populations. The purpose of using Nei's standard genetic distance is because this measure was shown to recover a greater proportion of correct trees in the study of Takezaki & Nei (1996). To find out the proportion of sequence divergence attributed to population differentiation, the sequence statistic K_{ST} (Hudson *et al.*, 1992; Roff & Bentzen 1992) was used. The value of K_{ST} is dependent on both the nucleotide diversity of the total sample (K_T) and the nucleotide diversity found within populations (K_s). As K_s becomes smaller when compared with K_T , the subdivision between populations become stronger. The Monte Carlo simulation was used to determine the statistical significance of observed K_{ST} by ensuring that individuals were sampled without replacement and they were randomly assigned to populations. This allowed the assessment of statistical significance by using the distribution of K_{ST} values that were over 1000 simulations. It was further noted that there is a small probability value under panmictic population structure for any given large observed value of K_{ST} (Hudson *et al.*, 1992).

Pairwise F_{ST} were computed with Arlequin version 3.5.2.2 (Excoffier, 2015). Correlation analysis were performed with IBM SPSS 25 (Ibm, 2017). One-sample T Test in the IBM-SPSS 25 was used to test if there was significant difference in the AT/GC ration of all the whole mitochondrial genome sequences by testing the difference of the AT/GC ratio values from the Test Value of the blackflies in the central transition ecological zone of Ghana. The value 3.0 was the AT/GC Ratio for the whole mitochondrial genome sequences of the blackflies in the transition ecological zone of Ghana. Bootstrapping was performed with 1000 number of samples and 95% level of confidence intervals.

4.3: Results

4.3.1 Results of library preparation and sequencing of whole mitochondrial genomes

The agarose gel electrophoresis results (Figure 4.3.1) show libraries prior to cleaning with fragment sizes of which some were below 100 bp and others above (Figure 4.3.1, A), clean library from DNA extracted from a blackfly sampled from Ghana (Figure 4.3.1, B), and another clean library from DNA extracted from a blackfly sampled from Cameroon (Figure 4.3.1, C). The use of the Agencourt AMPure XP PCR Purification system was able to selectively bind the library fragments 100 bp and larger to the paramagnetic beads, thereby preferentially eliminating all contaminants whose sizes were below 100 bp. The cleaned libraries had fragment sizes ranging from about 150 to 1000 bp, with greater concentration between 200 and 400 bp.

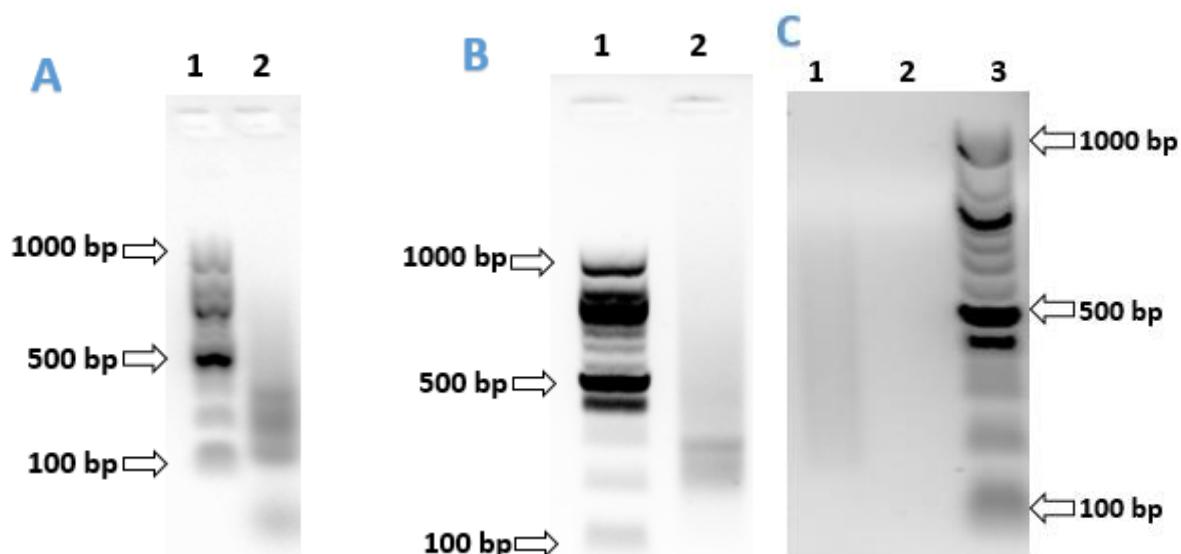


Figure 4.3.1: Agarose gels showing NGS library from Africa. The figure shows 3 gel images prepared using 2% agarose. In the figure, gel “A” shows NGS library before clean up, whilst gels “B” and “C” show libraries after size selection clean up. From gel A, well 1 indicates a 100bp ladder (InvitrogenTM), and well 2 shows an unclean library. From gel B, well 1 shows 100bp ladder (InvitrogenTM), whilst well 2 indicates an NGS library from Ghana. From gel C, well 1 shows an NGS library produced from blackfly from Cameroon, well 2 shows a no template control whilst well 3 shows a 100 bp ladder (InvitrogenTM).

The tapestation gel result (Figure 4.3.2, A and B) was consistent with that of the agarose gel. The strong band intensity was seen between the 200 to 1000 bp regions of the gel. Although

the library fragment lengths ranged from approximately 150 to 1000 bp with very few of the larger fragments beyond 700 bp, the bulk of the fragments peaked at an approximate range of 300 to 480 bp across all individual blackflies studied. All the samples used for the sequencing were successfully indexed with sequencing adapters as seen in the graphs of the amplifications, melt curves and melt peaks (Figure 4.3.3). The average library fragment sizes across the river basins fell between 400 and 500 bp (Table 4.3.1).

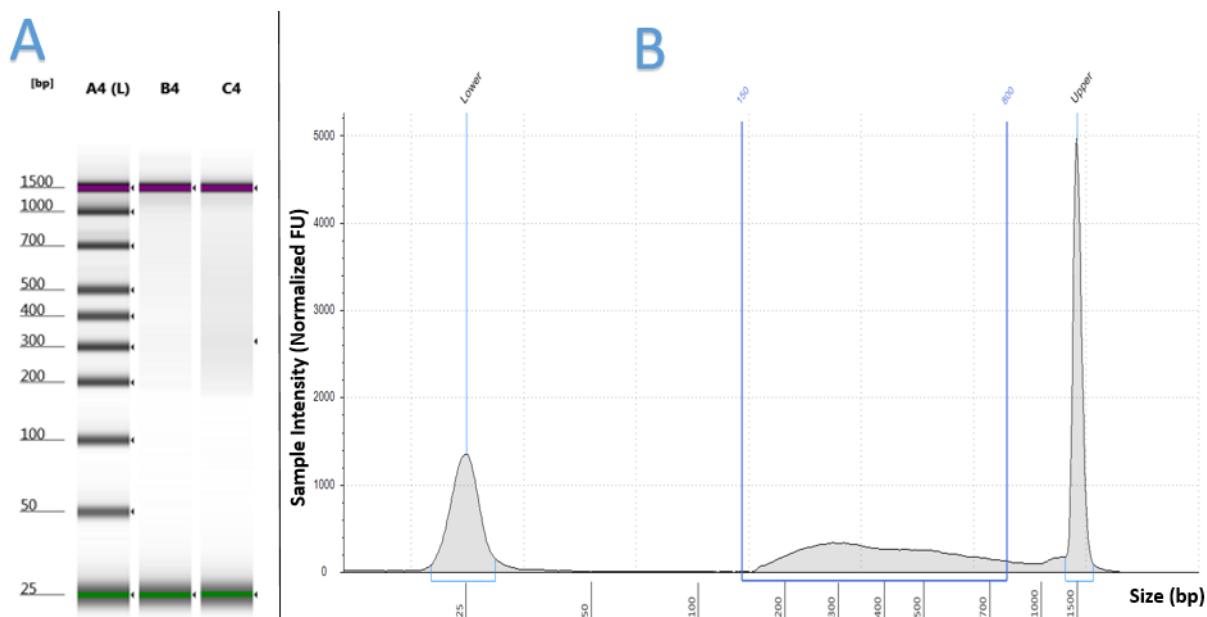


Figure 4.3.2: NGS library fragment analysis using the Agilent High Sensitivity D100 Screen Tape Assays. Figure 4.3.2 A is a Tapestation gel image of the DNA analysis of libraries for some of the African samples prior to whole genome sequencing. From left to right, column 1 is 100bp ladder; columns 2 and 3 are libraries from Pru and Daka river basins respectively. They were produced by using the Nextera DNA XT library kit. Figure 4.3.2 B is an electropherogram of the analysis of the NGS libraries. The analysis was carried out using the High Sensitivity D1000 Screen Tape Assays for the Agilent 2200 TapeStation System. The x axis represents the library fragment size in bp whiles the y axis represents the sample intensity (normalized FU).

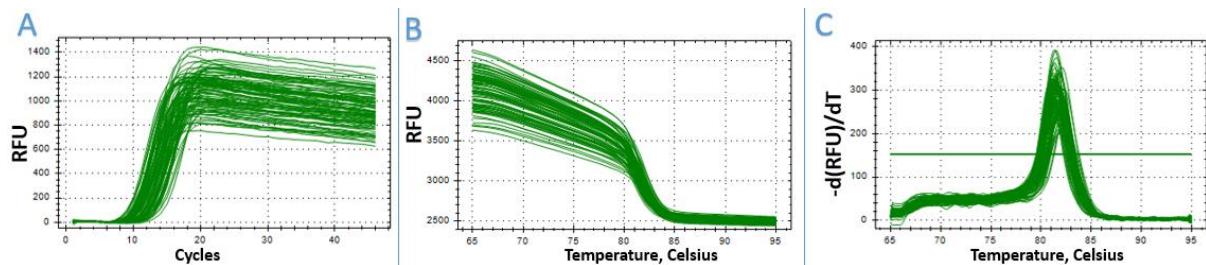


Figure 4.3.3: Quantitative real-time PCR (qPCR) of pooled library. In the above figure, A = Amplification, B = Melt curve, C = Melt peak.

Table 4.3.1: Tapestation results showing some of the average library fragment sizes

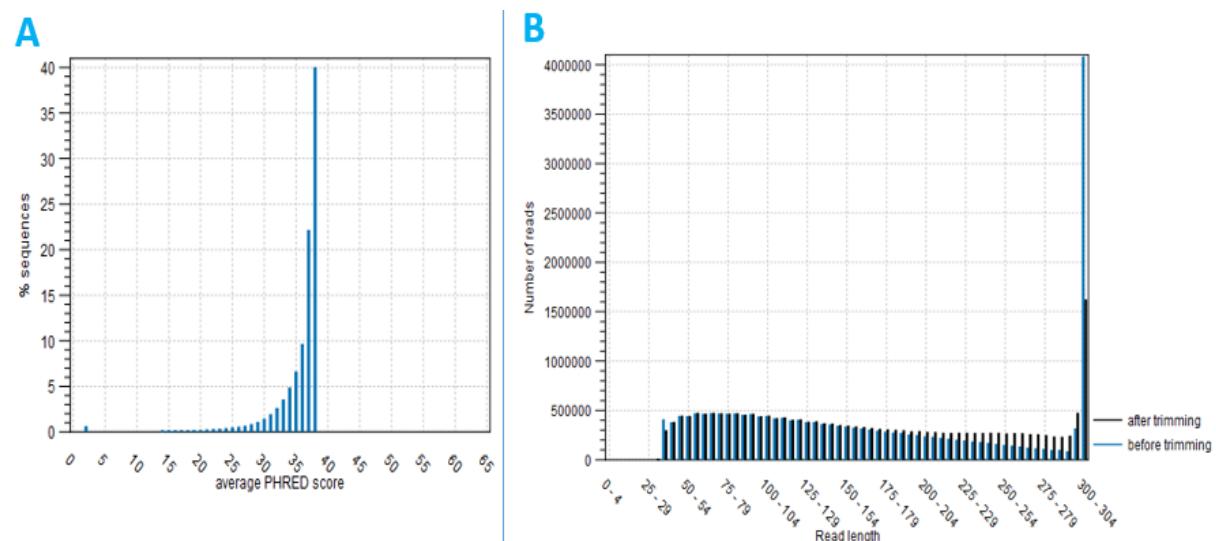
Sample Name	From (bp)	To (bp)	Average Size (bp)
SIMA1	150	800	478
SIMA2	150	800	431
SIMA3	150	800	475
SIMA4	150	800	425
SIMA5	150	800	457
SIMA6	150	800	449
SIMA7	150	800	452
SIMA8	150	800	414
SIMA9	150	800	431
SIMA10	150	800	405

The libraries were prepared with the Nextera DNA XT kit

A total of 83 libraries were sequenced and they consisted of individual blackfly numbers of 2 in the MiSeq1, 10 in the MiSeq 2, 25 in the NextSeq1 and 46 in the NextSeq2 (Table 4.3.2). The overall number of sequencing reads generated was 391,393,090. The proportion of the number of reads in the MiSeq1, MiSeq2, NextSeq1 and NextSeq2 out of the overall number of reads generated was 9.5% (37,310,772), 6.1% (23,964,252), 40.2% (157,462,998) and 44.1% (172,655,068) respectively. Thus, the sequencing reads generated were greatest in the NextSeq2, followed by the NextSeq1, and low in the MiSeq1 with the MiSeq2 having the lowest value. In order of decreasing magnitude, the lowest number of reads generated per individual blackfly occurred in MiSeq1 (17,158,958), MiSeq2 (2,109,848), NextSeq2 (1,870,762) and NextSeq1 (9,868). In order of increasing magnitude, the highest number of

reads sequenced per individual blackfly occurred in MiSeq2 (2,817,006), NextSeq2 (14,330,426), MiSeq1 (20,151,814), and NextSeq1 (38,585,502). Nevertheless, the average numbers of reads in the sequence types were 18,655,386 (MiSeq1), 2,396,425 (MiSeq2), 6,298,520 (NextSeq1) and 3,753,371 (NextSeq2).

From Figure 4.3.4A, the majority of reads sequenced had a quality score of 30 and above with a minimum of 99.9% base call accuracy. Reads with a quality score below 30 were filtered out. The proportion of reads per individual blackfly that survived the trimming process ranged from 99.39% to 99.96% (Figure 4.3.4 B). The average read length before trimming ranged from 148.8bp to 199.6bp, while that after trimming ranged from 148.8bp to 194.3bp (Supplementary Table 4.3.3).



42512, 18515, 100910, and 84991. However, the averages of the assembled sequences were 35,527 (MiSeq1), 11,357 (MiSeq2), 18,050 (NextSeq1), and 13,197 (NextSeq2). One out of the total of 83 blackfly libraries failed to assemble in the *de novo* assembly process, leading to an overall sequence success of 98.8%. Details of the sequenced and assembled reads can be found in the Supplementary Table 4.3.2 A and B.

Table 4.3.2: Summary of sequenced reads

Sequence	N	Total Reads	Mapped Reads	% Mapped Reads	mDNA:nDNA
MiSeq 1	2	37310772	71054	0.1904383	524
MiSeq 2	10	23964252	113571	0.473918401	210
NextSeq 1	25	157462998	433191	0.275106536	362
NextSeq 2	46	172655068	607077	0.351612615	283

Sequence = Sequence type, N = Total number of sequences, mDNA:nDNA = mitochondrial to nuclear read ratio as a measure of the mitochondrial genome copy number

4.3.2 Pre-processing of sequence reads

As shown in Table 4.3.3, the organisms used as references to detect and remove contamination sequences among those of *Simulium* were *Wolbachia pipiensis* (CP042444.1), *Wolbachia phage WO* (AB036666.1), *Onchocerca ochengi* (NC_031891.2), *Onchocerca volvulus* (KT599912.1), *Homo sapiens* (JX669269.1), *Fusarium bambusae* (MH684411.1), and *Plasmodium falciparum* (KY923425.1). These organisms belong to the taxonomic groups of bacteria, a eukaryotic nematode parasitic in cattle, a eukaryotic nematode parasitic in humans, eukaryotic mammal, fungi, and protozoa. The choice of this broad category of organisms was to enable the detection, if present, of contamination from the broad spectrum of organisms of active research in our lab. After the determination of the percentage identity between the organisms on the list of contamination reference sequences and the 3 *Simulium* whole mitochondrial genomes (accessions: NC_029753.1, NC_033348.1, and NC_040120.1), a 1 standard deviation to the lowest value of the percentage identity was allowed for the similarity index for the blast analysis. This decrease in the stringency for the blast analysis was to enable the detection, if present, of other contamination sequences that belonged to the same taxonomic grouping as those shown on the list of contamination reference sequences.

Table 4.3.3: List of contamination reference sequences

No.	Name of Organism	Accession Number	Size (bp)	Genome type	Type of Organism
1	<i>Wolbachia pipiensis</i>	CP042444.1	1,269,137	Complete Genome	Bacteria
2	<i>Wolbachia phage WO</i>	AB036666.1	32,987	Linear DNA	Virus
3	<i>Onchocerca ochengi</i>	NC_031891.2	13,744	Mitogenome	Eukaryote (Nematode)
4	<i>Onchocerca volvulus</i>	KT599912.1	13,769	Mitogenome	Eukaryote (Nematode)
5	<i>Homo sapiens</i>	JX669269.1	16,659	Mitogenome	Eukaryote (Human)
6	<i>Fusarium bambusae</i>	MH684411.1	63,593	Mitogenome	Fungi
7	<i>Plasmodium falciparum</i>	KY923425.1	5,884	Mitogenome	Protozoa

4.3.3 Contig measurement and assessment of nucleotide distribution

A total of 82 whole mitochondrial genomes were individually assembled and annotated (Figure 4.3.5 and Table 4.3.4). The largest assembly had an N25 length of 2964bp, N50 of 1525bp, N75 contig length of 660bp, average contig length of 901, and a total contig length of 209,932,758bp. Its nucleotide distribution for the Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) was 31.5%, 18.6%, 18.6%, and 31.4% respectively. There was sufficient read coverage across the entire length of each assembly and mapping (Figure 4.3.6) that resulted in the final mitochondrial genome development. The use of a minimum contig length of 200 (blue box in Table 4.3.4) for the assembly led to assemblies consisting of both nuclear and mitochondrial genes such that the nuclear gene contigs were numerous. However, the use of a minimum contig length of 13,000 (green box in Table 4.3.4) resulted in assemblies that consisted of either only whole mitochondrial genome contigs, or an assembly with contigs of few nuclear genes together with complete mitochondrial genomes. By using the 13,000 minimum contig length, the assemblies completed in as little as few minutes to an average time of 25 minutes. Moreover, such assemblies made it easier to identify the complete mitochondrial genomes amidst the few contigs of nuclear genes (Figure 4.3.7). In all the

assemblies, the adenine (A) and thymine (T) were similar and higher than the guanine (G) and cytosine (C). The proportion of A-T to G-C ranged from 1.49 to 3.05 with a mode of 1.5.

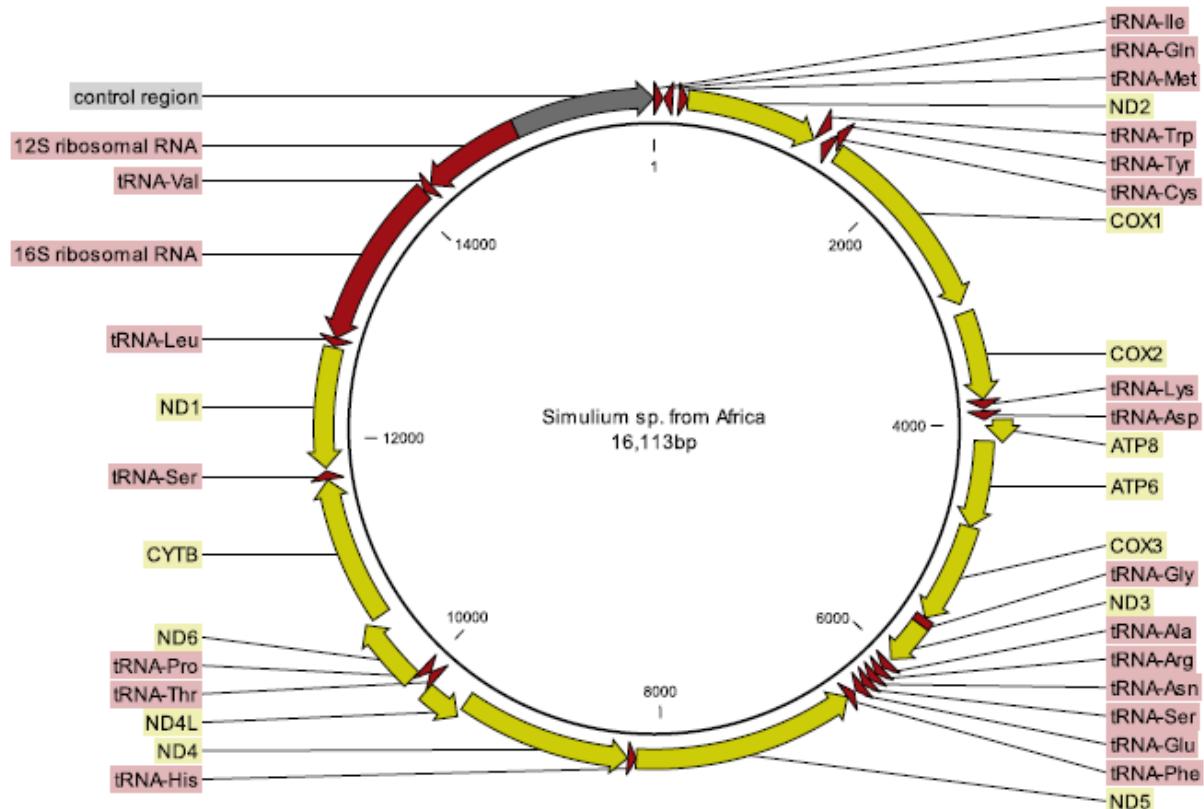


Figure 4.3.5: Annotated view of whole mitochondrion genome of African *Simulium* blackfly

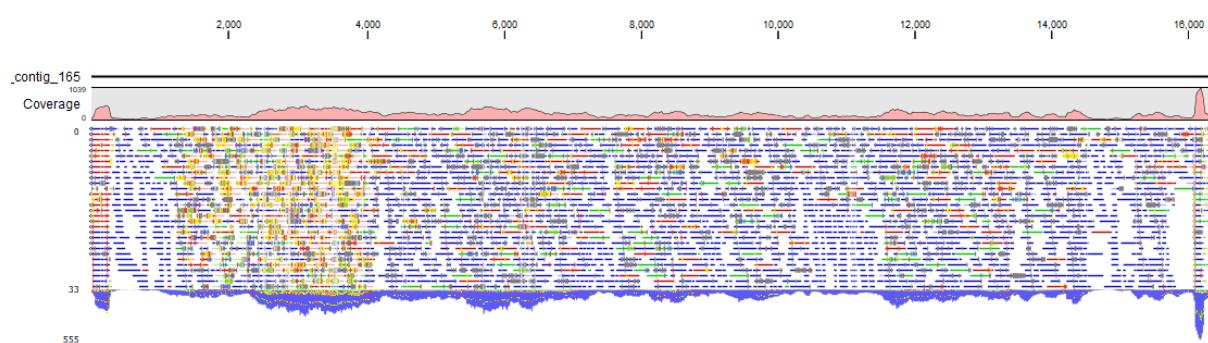


Figure 4.3.6: Paired-end reads coverage of a mitochondrial genome. It can be seen that the entire length of the contig mapping has reads with no loci missing any reads.

Rows: 7 / 247,290

Match any Match all

Consensus length	>	13000	<input type="button" value="+"/> <input type="button" value="x"/>		
Consensus length	<	17000	<input type="button" value="+"/> <input type="button" value="x"/>		
Name	Consensus length	Total read c...	Single reads	Reads in pairs	Average cover...
B7_S19_L001_R1_001_(paired)_trimmed_contig_636_mapping	15765	34331	2435	31896	278.05
B7_S19_L001_R1_001_(paired)_trimmed_contig_24416_mapping	14199	1514	248	1266	17.98
B7_S19_L001_R1_001_(paired)_trimmed_contig_4961_mapping	16720	1439	225	1214	14.37
B7_S19_L001_R1_001_(paired)_trimmed_contig_1426_mapping	15345	1403	163	1240	14.57
B7_S19_L001_R1_001_(paired)_trimmed_contig_21373_mapping	13738	1279	165	1114	15.23
B7_S19_L001_R1_001_(paired)_trimmed_contig_2853_mapping	13223	1154	128	1026	13.67
B7_S19_L001_R1_001_(paired)_trimmed_contig_4057_mapping	13027	1106	182	924	14.92

Figure 4.3.7: Mitochondrial contig selection. The use of a minimum contig length of 13,000bp produced a single contig assembly for the majority of samples, and in others, produced a single contig with few other contigs of nuclear gene origin. The use of 13,000 consensus length to sort through the contigs of assemblies produced using 200bp as the minimum contig length, led to easy identification of the whole mitochondrial genome (mostly found at the top of the contig list).

The percentage identity between the Ghanaian blackflies from West Africa and: (1) those from outside Africa ranged from 86.72 to 87.94; (2) the one from Malawi ranged from 91.78 and 91.98; and (3) the blackflies from Cameroon in Central Africa ranged between 93.19 and 93.42; The percentage identity among the blackflies from the 3 river basins in Ghana ranged from 97.59 and 99.10. The 3 blackflies used in the percentage identity test were chosen from 3 separate clusters having the greatest diversity among themselves. Details of the percentage identity test can be found in Table 4.3.5.

Table 4.3.4: Contig measurements and nucleotide distribution of assembled sequences

Sample ID	Length from contig measurements (bp)					Frequency (%)			
	N25	N50	N75	Average	Total	A	T	G	C
D3	2,964	1,525	660	901	209,932,758	31.5	31.4	18.6	18.6
B7	2,235	1,175	544	759	187,644,536	31.8	31.9	18.1	18.2
EG1	793	514	347	463	59,565,253	30.3	30.0	19.9	19.9
EG2	717	477	323	433	58,247,631	30.4	30.2	19.7	19.7
EG3	810	512	343	463	58,956,620	30.3	30.1	19.8	19.8
EG4	784	500	335	451	67,841,703	30.7	30.4	19.4	19.4
EG5	898	522	336	468	63,851,221	30.1	29.8	20.0	20.1
EG6	836	526	350	471	62,339,126	30.4	30.2	19.7	19.7
EG7	720	474	321	432	64,204,841	30.6	30.4	19.5	19.5
EG8	750	479	320	436	69,662,625	30.4	30.2	19.7	19.7
EG9	870	535	354	479	64,945,926	30.5	30.2	19.7	19.7
EG10	849	519	342	466	63,831,566	30.3	30.1	19.8	19.8
SIM1_NS	23,037	17,114	14,593	18,033	2,398,423	33.0	33.1	16.6	16.7
SIM2_NS	18,838	16,123	14,693	16,887	709,262	33.5	33.8	16.2	16.2
SIM3_NS	15,949	15,949	15,949	15,949	15,949	38.8	36.5	10.2	14.6
SIM4_NS	15,948	15,948	15,948	15,948	15,948	36.5	38.7	14.6	10.2
SIM5_NS	16,058	16,058	16,058	16,058	16,058	38.7	36.6	10.2	14.5
SIM6_NS	15,946	15,946	15,946	15,946	15,946	38.8	36.5	10.2	14.6
SIM7_NS	15,938	15,938	15,938	15,938	15,938	38.7	36.5	10.2	14.5
SIM8_NS	16,120	16,120	16,120	16,120	16,120	38.7	36.6	10.2	14.6
SIM10_NS	15,921	15,921	15,921	15,921	15,921	36.4	38.7	14.6	10.2
SIM11_NS	15,939	15,939	15,939	15,939	15,939	36.5	38.7	14.6	10.2

The table shows the length from contig measurements, including scaffolded regions, in bp; and frequency (nucleotide distribution) in percentage. The 2 samples in orange colour were the first to be sequenced at greater depth (over 17 million total reads each) and used to develop the initial draft reference mitochondrial genomes of African *Simulium* blackflies. The samples in purple colour each had total number of reads ranging 30-40 million. The rest of the samples in this category each had total number of reads less than 5 million.

4.3.4 Characterization of the mitochondrial genome of *Simulium* blackflies

The complete mitochondrial genomes of *Simulium* sp. in Africa are circular molecules of 16108 to 16232 nucleotides, and each of them has 2 rRNA genes, 22 tRNA genes, 13 protein coding genes and a non-coding control region (D-LOOP). The sequence characteristics of 3 distinct molecular cluster groups, whose members were distributed across all 3 major river basins in the ecological transition zone of Ghana, were assessed (Supplementary Table 4.3.4-7). Although it is not very common to study and describe genomic differences by molecular weight, a trend that was worth noting was observed in the molecular weights of the blackflies

in West and Central Africa, as well as in Asia and Europe; and this could be of operational significance to onchocerciasis (details are provided in the discussion section). Generally, the molecular weights from the *Simulium* blackflies in Asia and Europe were smaller than those in Africa. Comparatively, the single-stranded weights of the nucleotide sequences in: West Africa were higher, ranging from 4,963.39 kDa to 4,967.55 kDa; Central Africa were moderate, ranging from 4,957.32 kDa to 4,958.45 kDa; non-African blackflies were lower, ranging from 4,733.55 kDa to 4,900.44 kDa. A similar trend was observed in the double-stranded weights. At a salt concentration range of 0.1M to 0.4M, the melting temperatures of the *Simulium* whole mitochondrial genome sequences in West Africa were consistently lower than those from outside West-Africa. For instance, at a salt concentration of 0.1M, the melting temperatures of the sequences in West Africa were below 75°C while those from outside West Africa were above 75°C. The AT/GC ratio of the blackfly sequences from the West-African sampling locations of Pru, Black Volta and Daka were 3.0486, 3.0816 and 3.0650 respectively. These values were slightly higher than those in the Central African sampling locations of Mbam (2.9683) and Nkam (2.9683) as well as those from Asia and Europe, whose values ranged from 2.7037 to 2.9841 (See Supplementary Table 4.3.4 A). The One-sample T test shown in Supplementary Table 4.3.4 B however demonstrated that the slight difference in the AT/GC ratios were not statistically significant ($p = 0.513$).

Table 4.3.5: Pairwise percentage identity among blackflies from within and outside Africa

	Sa	Sv	Sm	Sc	Sma	Sp	Sb	Sd
<i>Simulium aureohirtum</i> (Sa)	100	86.77	87.28	86.52	87.02	87.09	87.19	87.11
<i>Simulium variegatum</i> (Sv)	86.77	100	88.12	86.24	86.75	86.86	86.88	86.72
<i>Simulium maculatum</i> (Sm)	87.28	88.12	100	87.23	88.83	87.79	87.94	87.69
<i>Simulium</i> from Cameroon (Sc)	86.52	86.24	87.23	100	91.56	93.24	93.42	93.19
<i>Simulium</i> from Malawi (Sma)	87.02	86.75	88.83	91.56	100	91.91	91.98	91.78
<i>Simulium</i> from Pru (Sp)	87.09	86.86	87.79	93.24	91.91	100	99.10	97.64
<i>Simulium</i> from Black Volta (Sb)	87.19	86.88	87.94	93.42	91.98	99.10	100	97.59
<i>Simulium</i> from Daka (Sd)	87.11	86.72	87.69	93.19	91.78	97.64	97.59	100

The sequences from Ghana were chosen from 3 distinct molecular cluster units that did not have geographical structure by river basin (see below).

4.3.5 Polymorphism and diversity

Polymorphism and diversity were compared among the 3 river basins in the transition ecological zone of Ghana (Table 4.3.6). In order of increasing magnitude, the total number of mutations was 794 (Pru), 801 (Black Volta), and 836 (Daka). The number of polymorphic (segregating) sites showed similar trend as that of the total number of mutations with the

lowest value occurring in Pru (773), moderate (784) in Black Volta, and highest in Daka (816). These led to a relatively higher number of total number of mutations (1351) and total number of polymorphic sites (1301) across all the river basins. Generally, the total number of mutations was greater than the number of polymorphic sites in each river basin. A different pattern was realized with the total number of singleton mutations across river basins with Daka showing the highest value (505), relatively moderate value occurring in Pru (488) and lowest in Black Volta (319). The number of haplotypes (25) was the same in each of the river basins, leading to an overall number of 75 haplotypes across all 3 river basins. The haplotype diversity was also the same, a value of 1, in each of the river basins. However, the nucleotide diversity demonstrated a different pattern with Pru having the smallest value of 0.00787, while Daka had a relatively moderate value of 0.00830, and Black Volta having the highest value of 0.00898. The average number of nucleotide differences demonstrated similar pattern as that of the nucleotide diversity with Pru having the lowest value of 126.8, a relatively moderate value of 133.6 value in Daka, and the highest value of 143.5 in the Black Volta.

Table 4.3.6: Polymorphism and diversity in the mitochondrial genome of blackflies in Ghana

River basin	<i>M</i>	<i>S</i>	<i>Eta (s)</i>	<i>H</i>	<i>Hd</i>	π	<i>k</i>
Pru	794	773	488	25	1	0.00787	126.8
BlackV	801	784	319	25	1	0.00898	143.5
Daka	836	816	505	25	1	0.00830	133.6
All river basin	1351	1301	505	75	1	0.00853	136.4

M = Total number of mutations; *S* = Number of polymorphic (segregating) sites; *Eta (s)* = Total number of singleton mutations; *H* = Number of Haplotypes; *Hd* = Haplotype (gene) diversity; π = Nucleotide diversity; *k* = Average number of nucleotide differences; BlackV = Black Volta River.

4.3.6 Test of neutrality

To determine whether or not there existed blackfly sequences that did not fit the neutral theory model at equilibrium between mutation and genetic drift, Tajima's D test as a neutrality test was used (Table 4.3.7). Tajima's D were all below zero and they were not statistically significant ($P > 0.05$). The individual values however ranged from as low as -1.598 in Pru to as high as -1.3 in Black Volta, with the overall value across all 3 river basins (-1.77) being relatively lower than any individual locations. The negative values indicated a high frequency of rare alleles. However, the lack of statistical significance in all the river

basins indicated that the allelic distribution and/or level of variability did not violate the neutrality assumption. Hence, the blackfly sequences were evolving randomly (neutrally). Fu and Li's statistical test was explored because it is generally more sensitive than Tajima's D towards identifying departures from neutrality as a result of genetic hitchhiking and population expansion (Fu, 1997; Zeng *et al.*, 2006). Both Fu and Li's D statistic, and Fu and Li's F statistic, support the idea of a non-departure from neutrality of the blackfly data due to the lack of statistical significance ($P > 0.05$). Similar to Tajima's D test results, Fu and Li's D test gave negative values for all the river basins, with the values ranging from as low as -2.35 in Pru to as high as -0.85 in Black Volta. The overall value of -2.03 was also not statistically significant ($P > 0.05$). A similar trend was observed with the Fu and Li's F test in the various river basins with the highest value being observed in Black Volta ($FLF^* = -1.17$) and the lowest value found in Pru ($FLF^* = -2.48$). All the neutrality tests agreed on the idea of the blackflies exhibiting variability in each and across all the 3 river basins and they were evolving with a similar distribution and pattern as with a neutrally evolving population.

Table 4.3.7: Neutrality test in the mitochondrial genome of *Simulium* sp. in Ghana

River Basin	Tajima's D	Tajima's Stat Sig	FLD*	FLD* Stat Sig	FLF*	FLF* Stat Sig
Pru	-1.59763	> 0.05	-2.35281	> 0.05	-2.48448	> 0.05
Black Volta	-1.30056	> 0.05	-0.85009	> 0.05	-1.17259	> 0.05
Daka	-1.59609	> 0.05	-2.28020	> 0.05	-2.42597	> 0.05
All 3 river basins	-1.77418	> 0.05	-2.02737	> 0.05	-2.30665	> 0.05

Tajima's D, Fu and Li's Tests were calculated using the total number of mutations. $FLD^* =$ Fu and Li's D* test statistic, $FLF^* =$ Fu and Li's F* test statistic, Stat Sig = Statistical significance

4.3.7 Gene Flow and Genetic Differentiation

The average within-population diversity was the same ($H_s=1$) when populations in the 3 river basins were assessed in a pairwise comparison. The within-population nucleotide diversity was relatively highest in the populations of Black Volta and Daka ($K_s= 153.05$), moderate in the populations of Pru and Black Volta ($K_s= 149.53$), and lowest in the populations of Pru and Daka river basins. The average number of nucleotide differences between populations in the 3 river basins showed a similar trend as that of the within-population nucleotide diversity with the highest value observed in the populations of Black Volta and Daka ($K_{xy}= 162.62$),

relatively moderate in the populations of Pru and Black Volta ($K_{xy} = 148.91$), and lowest in the populations of Pru and Daka ($K_{xy} = 147.47$). In order of increasing magnitude, the pairwise genetic distance was between populations in the Pru and Black Volta ($\text{GammaSt} = 0.018$), Pru and Daka ($\text{GammaSt} = 0.037$), and in Black Volta and Daka ($\text{GammaSt} = 0.051$) river basins. The fixation index realized by pairwise comparison of the populations among the 3 river basins was not significantly different from zero. The individual values however differed in order of decreasing magnitude in the compared populations of; Black Volta and Daka ($F_{ST} = 0.06$), Pru and Daka ($F_{ST} = 0.03$), and in populations of Pru and Black Volta ($F_{ST} = 0$). The results from the average within-population diversity, within-population nucleotide diversity, and average number of nucleotide differences indicate high genetic diversity but low genetic differentiation in the compared populations from the 3 river basins. Furthermore, the results from the fixation index demonstrate gene flow between the populations in the 3 river basins. These results agree with the notion of a lack of geographic population structure across the river basins, but with the presence of a single onchocerciasis transmission zone in the central ecological transition zone of Ghana.

Table 4.3.8: Gene flow and genetic differentiation per river basin in Ghana

Population 1	Population 2	<i>Hs</i>	<i>Ks</i>	<i>K_{xy}</i>	<i>GammaSt</i>	<i>F_{ST}</i>
Pru	BlackV	1	149.52830	148.90720	0.01833	-0.00417
Pru	Daka	1	142.71170	147.46560	0.03678	0.03224
BlackV	Daka	1	153.04670	162.61920	0.05070	0.05886

The table shows various parameters of gene flow and genetic differentiation between population 1 and 2 in the 3 river basins of Pru, Black Volta and Daka. Due to the potential for the existence of indels, especially in the control region, sites with alignment gaps were considered with gap as the fifth state. The Permutation test was performed with 1000 number of replicates and 1854547 pseudorandom number seed. H_s = Average within-population diversity, K_s = within-population nucleotide diversity, k_{xy} = the average number of nucleotide differences between population 1 and population 2, GammaSt = pairwise genetic distance, and fixation index.

From Table 4.3.9, the majority of the blackflies' genetic variation occurred within population (97.9%, $P > 0.001$), with the remaining variation occurring among populations (2.1%, $P >$

0.001). This also suggests a lack of geographic population structure by river basin. The test for the overall measurement of population differentiation due to genetic structure, overall Fixation Index (F_{ST}), was small (0.03) and approximately equal to zero. This small fixative index value closer to zero was similar to that of a panmictic population with random mating and greater gene flow between river basins. The force of gene flow in the population for a given generation (Nm), that influences the tendency of population differentiation because of genetic structure, was however large (F_{ST} 's $Nm = 16.25$). Both the negligible fixation index and the large Nm that is greater than 1 (Wright, 1951) support the idea of a lack of geographic population structure in the river basins across the central transition ecological zone of Ghana. The overall pairwise genetic distance, GammaSt statistic, was equally small (0.047) with an Nm value greater than the recommended value of 1 ($Nm = 10.17$) that is needed as the minimum value for gene flow between population to occur. This further supports the idea of a lack of population structure across the 3 river basins, thus adding support to the inference of a single onchocerciasis transmission zone in the central ecological transition zone of Ghana.

Table 4.3.9: Overall population differentiation test in Pru, Black Volta and Daka river basins

Variation	% Variation*	F_{ST}	F_{ST} 's Nm	GammaSt	GammaSt's Nm
Among Population	2.1	0.02986	16.25	0.04685	10.17
Within Population	97.9				

In the table, N = Effective population size, m = proportion of migrants in a population, Nm = the force of gene flow in the population for a given generation, F_{ST} = Fixation Index across the river basins, GammaST = overall pairwise genetic distance. The GammaSt was computed according to the consideration in Nei (1982). The F_{ST} was computed using the considerations in Hudson *et al.* (1992). * = The significance of the percentage variation = Phi_ST : $\text{Pr}(\text{random value} > \text{observed Phi_ST}) = 0.032$. Sites with alignment gaps were considered with gap as the fifth state. The Permutation test was performed with 1000 number of replicates and 1854547 pseudorandom number seed.

4.3.8 Determination of genetic relatedness

There are two main types of DAPC analysis, and these are *a priori* or *a posteriori* (Jombart, 2008; Jombart *et al.*, 2010). In the context of this study, *a priori* DAPC is defined as the use of the multivariate method to identify and describe clusters of genetically related individuals

without prior knowledge of the existence of any form of population grouping information. A *posteriori* DAPC is likewise defined as the use of the multivariate method to assess the genetically relatedness of individuals with existing knowledge by assigning known population parameters such as the sampling location to groups of the individuals. Without assigning any form of geographic information such as river basin, *a priori* DAPC analysis was carried out using complete mitochondrial genomes of 75 *Simulium* blackflies sampled from the transition ecological zone of Ghana (Figure 4.3.8 and Supplementary Figure 4.3.2). The optimum number of PCs retained was 30 and it accounted for over 95% of the total genetic variation observed. By examining the lowest BIC and carrying out optimization, the optimum number of clusters was identified as 8 and its associated discriminant function of 7 was used. With such a discriminating power, the 8 clusters resulted in the identification of 3 main blackfly groups. Clusters 1 and 3 formed one distinct group. Cluster 6 formed the second distinct group. The remaining clusters (2, 4, 5, 7 and 8) formed the third distinct molecular group (See section A of Figure 4.3.8). These 3 distinct molecular units need to be tested for the likelihood of being of operational significance in onchocerciasis. Additional data is required to ascertain their taxonomic status and significance. At this stage, it is proposed to refer to these distinct molecular units or groupings of blackflies as *Simulium* Unit Group (MOTU) 1, 2 and 3. In the assignment of individuals to population clusters, the prior group size was the same as the post group size. This led to a 100% of reassignment of individual blackflies to each of the 8 population clusters (figure 4.3.8, B). From clusters 1 to 8, the group sizes were respectively 8, 4, 3, 9, 26, 4, 13 and 8. By assigning clusters 1 and 3 to MOTU 1, cluster 6 to MOTU 2, and the remaining clusters to MOTU 3; the proportions of MOTU 1-3 thus became 14.7% (11), 5.3% (4) and 80% (60) respectively.

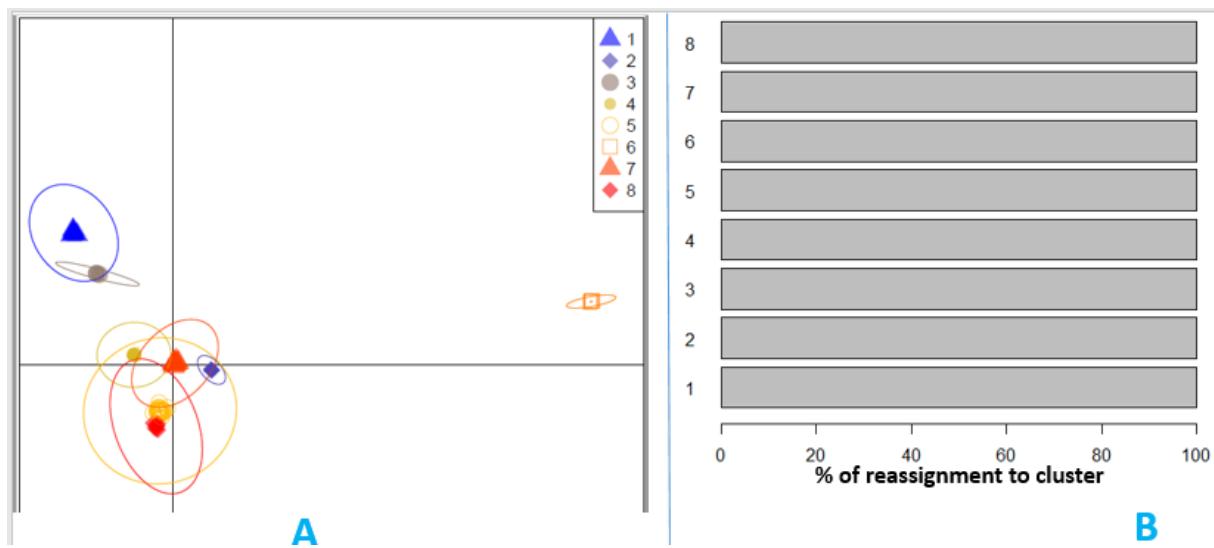


Figure 4.3.8: *A priori* assessment of genetic relatedness from 75 whole mitochondrial genomes in Ghana. It was plotted using 30 principal components, 8 clusters, and 7 discriminant functions. The ellipses represent the DAPC clusters whiles the dots represent the individual blackflies collected.

A repeat of the analysis using whole mitochondrial genomes without the control region (Figure 4.3.9), DLOOP only (Supplementary Figure 4.3.1) and a separate analysis using only the control region (Figure 4.3.10) from the same individual samples as above, showed similar 8 clusters organized into 3 distinct molecular groupings

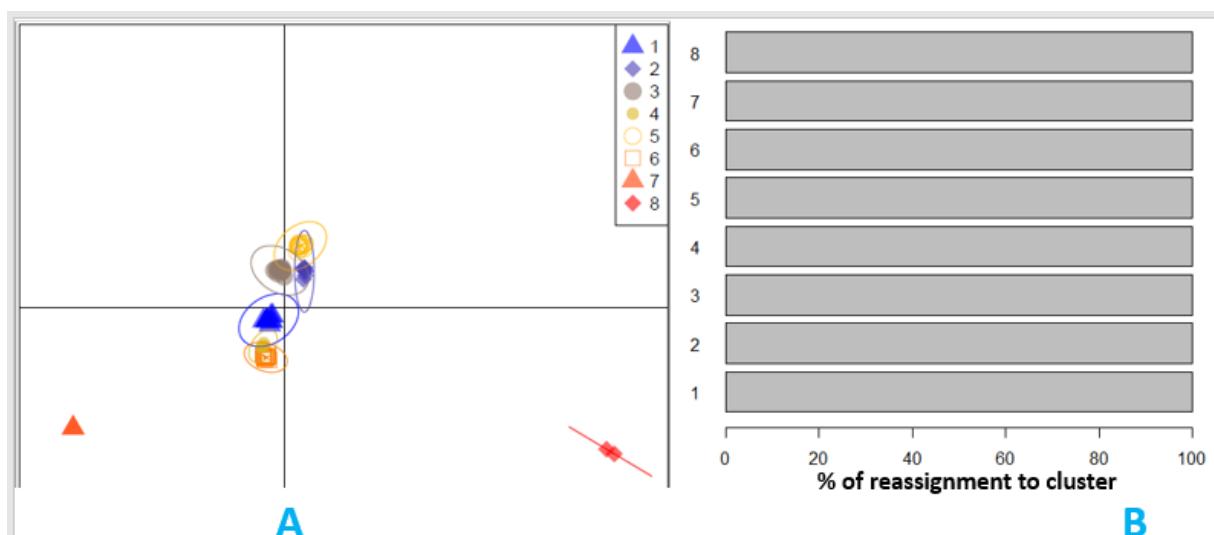


Figure 4.3.9: *A priori* assessment of genetic relatedness from 75 mitochondrial genomes without DLOOP in Ghana. It was plotted using 30 principal components, 8 clusters, and 7

discriminant functions. The ellipses represent the DAPC clusters whiles the dots represent the individual blackflies collected.

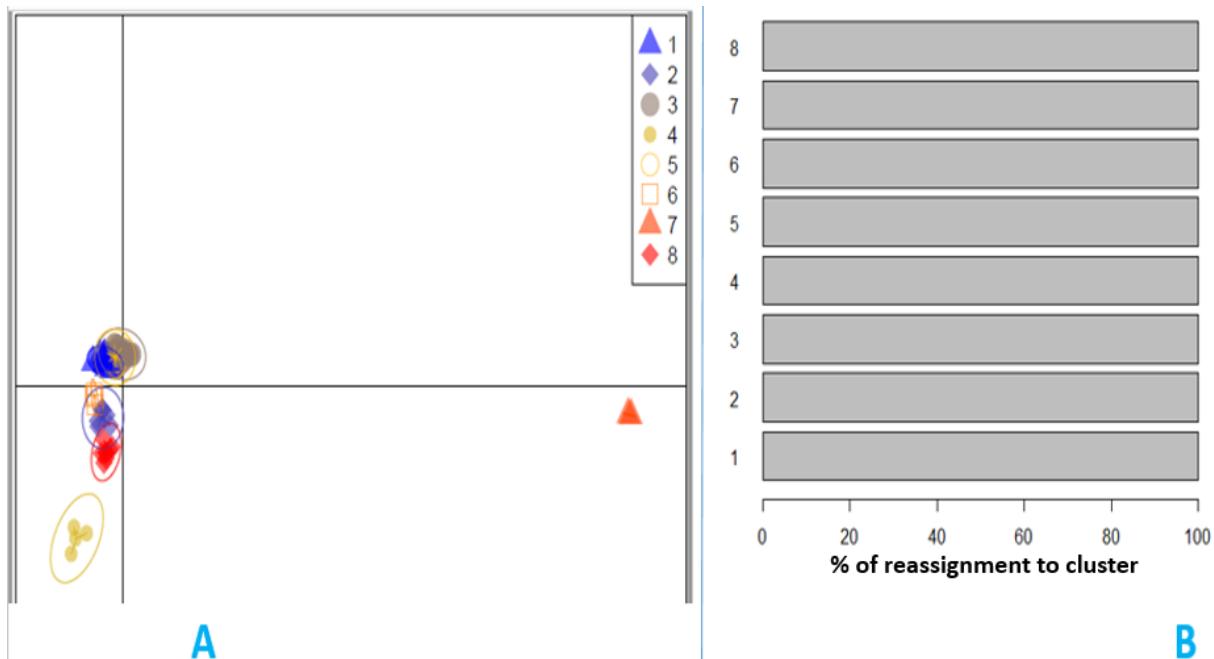


Figure 4.3.10: *A priori* assessment of genetic relatedness from 75 DLOOP sequences in Ghana. It was plotted using 30 principal components, 8 clusters, and 7 discriminant functions. The ellipses represent the DAPC clusters whiles the dots represent the individual blackflies collected.

A posteriori DAPC analysis was carried out using the same 75 samples from the central transition ecological zone of Ghana as in *a priori* analysis. The use of the 75 mitochondrial genomes without DLOOPS indicated overlap of individual membership among the 3 river basins, with 40% (Black Volta), 48% (Pru) and 60% (Daka) probability of successful reassignment of blackflies to the river basins from which they were originally collected (Figure 4.3.11, A and B). These corresponded to the local blackflies found in their original river basins from which they were collected. There were 60% (Black Volta), 52% (Pru) and 40% (Daka) of blackflies that could not be successfully assigned to the river basins from which they were originally collected due to the blackflies sharing genetic characteristics to those from other river basins different from where they were collected from and which may

have been migrants. From the haplotype network (Figure 4.3.11 C), there were shared haplotypes among members of all 3 river basins. Although most of the blackflies shared common haplotypes, there were 2 other groups that were independently distinct from the rest. The first group from Black Volta had 209 mutational step difference from the closest individual of shared haplotype from Daka. The second group had 294 mutational step difference from the closest individual of shared haplotype. There were a relatively small number of related groups that had diverged, such as one would see from a small, recent founding population composed of those three haplotypes that had since diverged, with some gene flow between groups. This is like the pattern seen with the parasite data from the work of colleagues in our lab and suggests that *Onchocerca volvulus* worms and *Simulium* blackflies may share a common history of colonisation and divergence. Thus, the DAPC and haplotype network support the hypothesis of the presence of blackfly divergence from a small founding population within a single onchocerciasis transmission zone with 3 molecular groups.

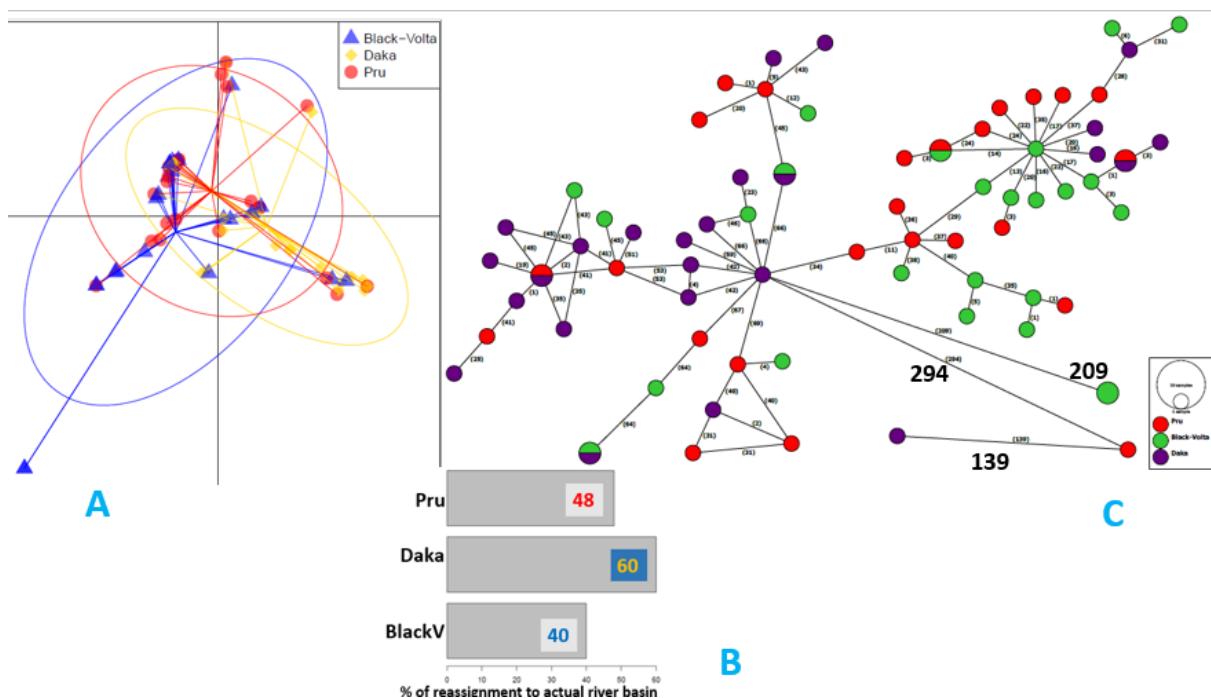


Figure 4.3.11: *A posteriori* assessment of genetic relatedness from 75 whole mitochondrial genomes without DLOOP. In the figure, “A” is DAPC scatter plot after assigning individual river basin information; “C” is haplotype network produced with the minimum spanning network; “B” represents the percentage of successful reassignment of individual blackflies to the river basins of collection. All assessments were made using 75 Whole mitochondria

genome sequences without DLOOP. The ellipses in “A” represent the different river basins from which the samples were obtained while the dots represent the individual blackflies collected. Each circle in the haplotype network corresponds to one haplotype, and the size is proportional to its frequency among the samples. Colours of the circles correspond to river basin sampling locations.

Similar results were obtained from using whole mitochondrial genomes with DLOOP (Figure 4.3.12-13), from the same set of 75 individual blackflies from Ghana. Thus, there were shared genetic characteristics by blackflies in all 3 river basins that were observed by the overlapping ellipses in Figure 4.3.12; and shared haplotypes by blackflies from all 3 river basins as shown in Figure 4.3.13.

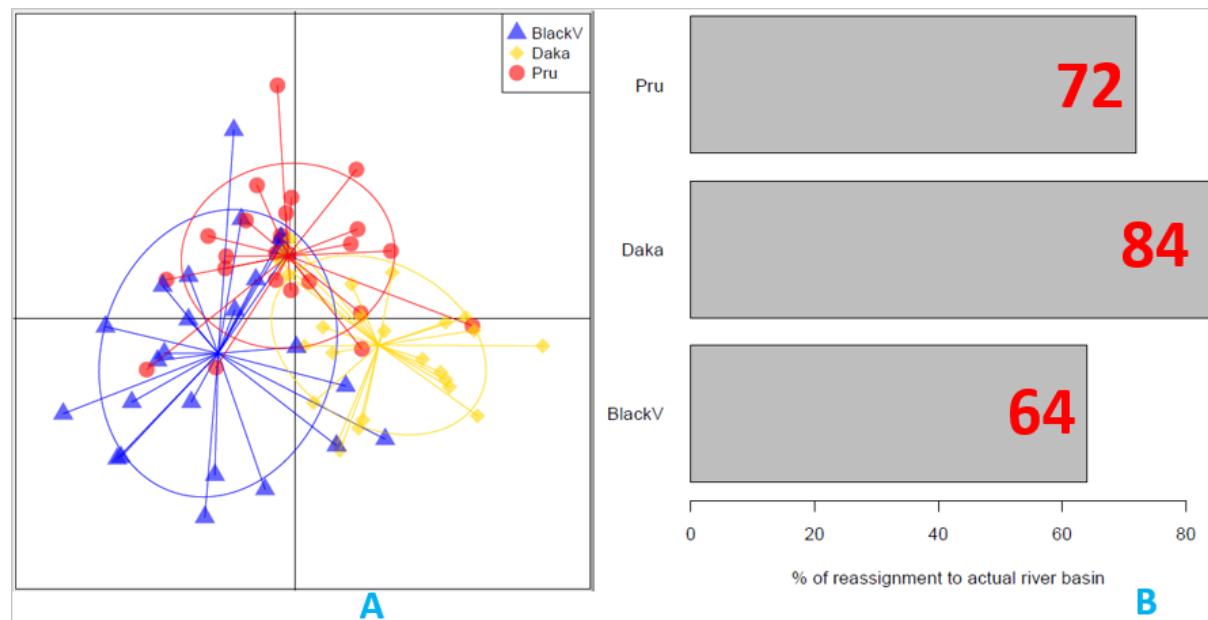


Figure 4.3.12: *A posteriori* assessment of genetic relatedness using 75 complete mitochondrial genomes from Ghana. It was plotted using 20 principal components and 2 discriminant functions. In the figure, “A” represents the DAPC scatter plot. The ellipses represent the different river basins from which the samples were obtained while the dots represent the individual blackflies collected. “B” represents the percentage of successful reassignment of individual blackflies to the river basins from which they were collected.

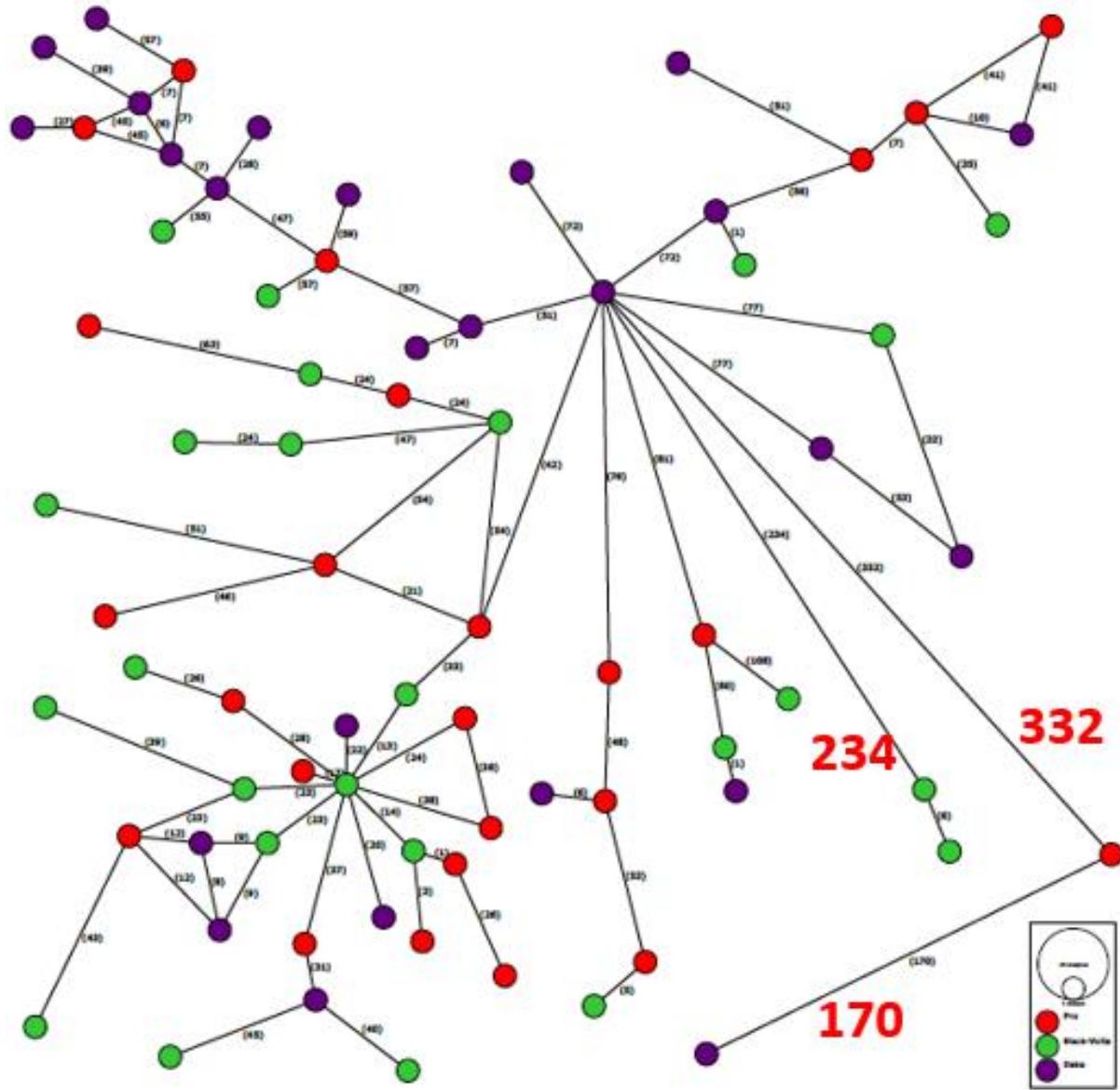


Figure 4.3.13: Minimum spanning network using 75 whole mitochondrial genomes with DLOOPs. The samples were collected from Pru, Black Volta and Daka river basins of Ghana. Each circle in the haplotype network corresponds to one haplotype, and the size is proportional to its frequency among the samples. Colours of the circles correspond to river basin sampling locations.

From Figure 4.3.14, the genetic relatedness among 81 whole mitochondrial genome sequences from Africa were assessed by using a *priori* type of DAPC analysis. In the graph, cluster 5 represented the single blackfly from Malawi (Lingoni Falls river basin); cluster 2 represented blackflies from Cameroon (Nkam and Mbam river basins); whiles the rest of the

blackflies were distributed in the 3 main groupings (consisting of clusters 1, 3, 4, 6-8), with members from each of these 3 latter groupings belonging to all the 3 Ghanaian river basins. The blackflies from West Africa (Ghana) were more closely related to themselves than to those from Central Africa (Cameroon) and South-Eastern Africa (Malawi). All the individual blackflies from Ghana formed 3 main distinct cluster groups and one of these may be genetically heterogeneous. Thus, the complete mitochondrial genomes differentiated among the various *a priori* clusters in Ghana, Malawi and Cameroon with 100% of successful reassignment of individual blackflies to each of the 8 clusters.

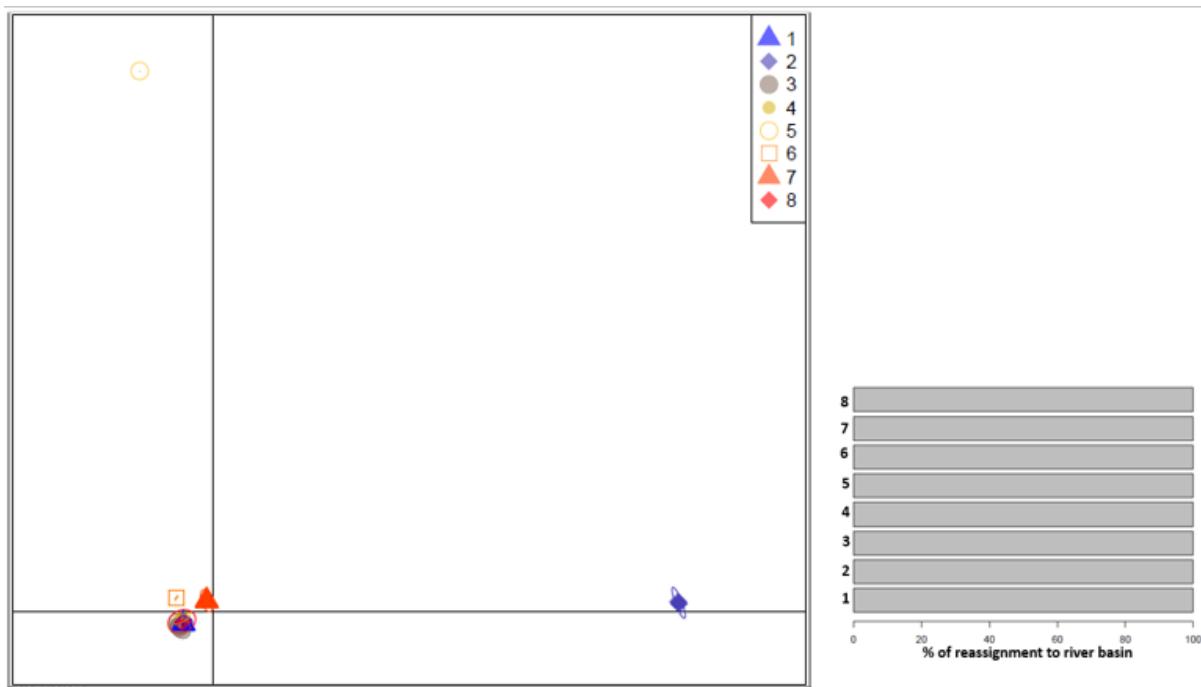


Figure 4.3.14: *A priori* assessment of genetic relatedness of 81 whole mitochondrial genomes of blackflies from Africa. The figure on the left represent the scatter plot of the DAPC analysis of 81 whole mitochondrial genomes from West, Central and South-Eastern Africa, using *a priori* strategy without population data assignment. It was plotted by retaining and choosing 30 principal components (PCs), 8 clusters and 7 Discriminant Functions. The samples consisted of: (1) 25 individual blackflies from each of the river basins of Pru, Black Volta and Daka, all located in the ecological transition middle zone of Ghana in West-Africa; (2) Five blackflies from Cameroon, in Central Africa; and (3) 1 blackfly from Malawi, located in South-Eastern Africa. The figure on the right represent percentage of reassignment of individual black flies to each of the 8 clusters. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

Figure 4.3.15 further confirmed the sufficiency of molecular resolution of the complete mitochondrial genomes to discriminate the Ghanaian blackflies from those of Cameroon and Malawi. There was 100% successful reassignment of blackflies to the river basins in Cameroon and Malawi, but not Ghana. These further supports the idea of blackfly movement between river basins in Ghana, in a single onchocerciasis transmission zone.

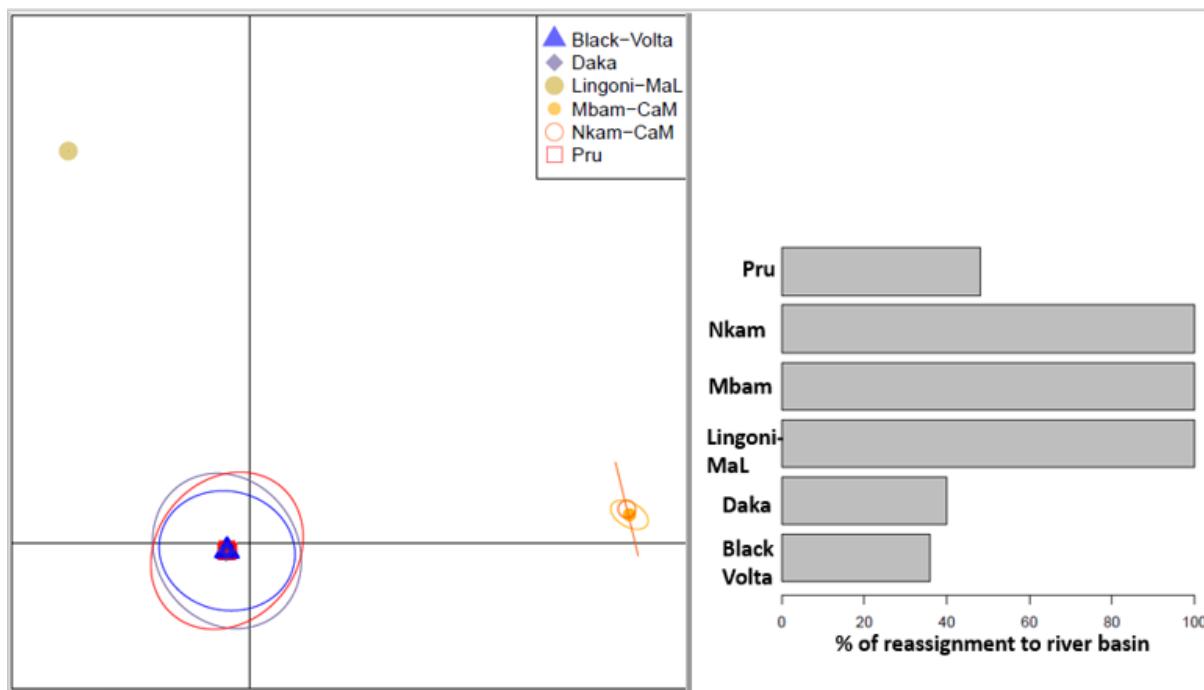


Figure 4.3.15: *A posteriori* assessment of genetic relatedness of 81 whole mitochondrial genomes of blackflies from Africa. The ellipses represent the different river basins from which the samples were obtained while the dots represent the individual blackflies collected. The figure on the left represent scatter plot of the DAPC analysis of 81 whole mitochondrial genomes from West, Central and South-Eastern Africa, using a *posteriori* strategy with population data assignment. It was plotted by retaining and choosing 30 principal components (PCs), 8 clusters and 7 Discriminant Functions. The samples consisted of: (1) 25 individual blackflies from each of the river basins of Pru, Black Volta and Daka, all located in the ecological transition middle Zone of Ghana in West-Africa; (2) Five blackflies from Cameroon, in Central Africa; and (3) 1 blackfly from Malawi, in South-Eastern Africa. The figure on the right represent the percentage of reassignment of individual black flies to each

of the river basins in Africa. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

4.3.9: Phylogenetic relationship of *Simulium* blackflies

The nucleotide substitution model tests for the 75 whole mitochondrial genomes in the ecological transition zone of Ghana indicated that the best model was the General Time Reversible Model with variation in gamma distributed and invariant sites (Table 4.3.10). Thus, it has a variation in rate and topology. The model of best fit (GTR+G+I) had the lowest BIC value of 76256, the lowest AICc value of 74371.3, and the highest maximum likelihood value of -37028.6 when compared with all the other models in the test. The estimated value of the shape parameter for the discrete Gamma Distribution was 0.7373. The proportion of sites estimated to be invariant was 83%. Substitution patterns and rates were estimated under the General Time Reversible model (+G+I) (Nei and Kumar, 2000). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+G]). Mean evolutionary rates in these categories were 0.06, 0.27, 0.61, 1.19, 2.87 substitutions per site. The nucleotide frequencies were A = 38.77%, T = 36.57%, C = 14.52%, and G = 10.14% (see Supplementary Figure 4.3.3). For estimating maximum likelihood values, a tree topology was automatically computed. The maximum Log likelihood for this computation was -37019.771. The analysis involved 75 nucleotide sequences. There was a total of 16168 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018).

Table 4.3.10: Nucleotide substitution model tests of 75 whole mitochondrial genomes

Model	Parameter	BIC	AICc	lnL
GTR+G+I	157	76256.04276	74371.28652	-37028.62273
TN93+G+I	154	76380.4992	74531.75664	-37111.85857
HKY+G+I	153	76442.0092	74605.27121	-37149.61611
T92+G+I	151	76542.12414	74729.3953	-37213.67866
GTR+G	156	77162.10959	75289.35791	-37488.65869
TN93+G	153	77255.52852	75418.79053	-37556.37577
HKY+G	152	77378.10914	75553.37572	-37624.66862
T92+G	150	77472.01225	75671.28799	-37685.62525
GTR+I	156	79262.23868	77389.48699	-38538.72323
TN93+I	153	79291.06139	77454.3234	-38574.14221
HKY+I	152	79384.45798	77559.72456	-38627.84304
T92+I	150	79463.79612	77663.07185	-38681.51719
GTR	155	80673.45098	78812.70386	-39251.33192
TN93	152	80719.56382	78894.8304	-39295.39596
HKY	151	80796.04559	78983.31675	-39340.63938
T92	149	80871.29482	79082.57514	-39392.26908
K2+G+I	150	81729.1158	79928.39154	-39814.17703
K2+G	149	82786.1226	80997.40292	-40349.68297
JC+G+I	149	83486.68432	81697.96464	-40699.96383
JC+G	148	84523.66157	82746.94648	-41225.45499
K2+I	149	84793.40897	83004.68929	-41353.32615
K2	148	86169.31237	84392.59727	-42048.28039
JC+I	148	86530.10074	84753.38564	-42228.67457
JC	147	87902.86637	86138.15587	-42922.05993

The table shows the Maximum Likelihood fits of 24 different nucleotide substitution models. Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), and Maximum Likelihood value (lnL) are shown. In the table, GTR= General Time Reversible; HKY= Hasegawa-Kishino-Yano; TN93= Tamura-Nei; T92= Tamura 3-parameter; K2= Kimura 2-parameter; JC= Jukes-Cantor, G= Gamma distributed, and I= Invariant site.

From the maximum likelihood tree (figure 4.3.16), the African *Simulium* blackflies were monophyletic to those from Asia and Europe. The West African blackflies from Ghana were monophyletic to all the other blackflies from Central and South-Eastern Africa. Blackflies from all the 3 river basins in the ecological transition zone of Ghana shared a single most recent common ancestor relative to other African and to Asian flies. There was interbreeding among flies across the river basins of Pru, Black Volta and Daka. The sharing of most recent

common ancestor by members of all 3 river basins without any clade consisting of members from a single river basin indicate support for the DAPC with a conclusion that there are 3 distinct but relatively recently diverged Clades within Ghana: SIM23, SIMWiT16 and EG-9-ToF form one clade, with a possible 2nd clade composed of SIM12KyT and EG-2Kyt, then the 3rd consisting of the remainder (the majority of the flies). The membership of these clades is consistent with the membership of the DAPC clusters. There was strong support, 100%, in the divergence of the Ghanaian blackflies from those in Cameroon; the Cameroon blackflies from the one from Malawi; and the African blackflies from all the non-African ones. This further support the inference of a single onchocerciasis transmission zone in the central ecological transition zone of Ghana, and these are distinct from all other blackflies from Central, South-Eastern, Asia and Europe.

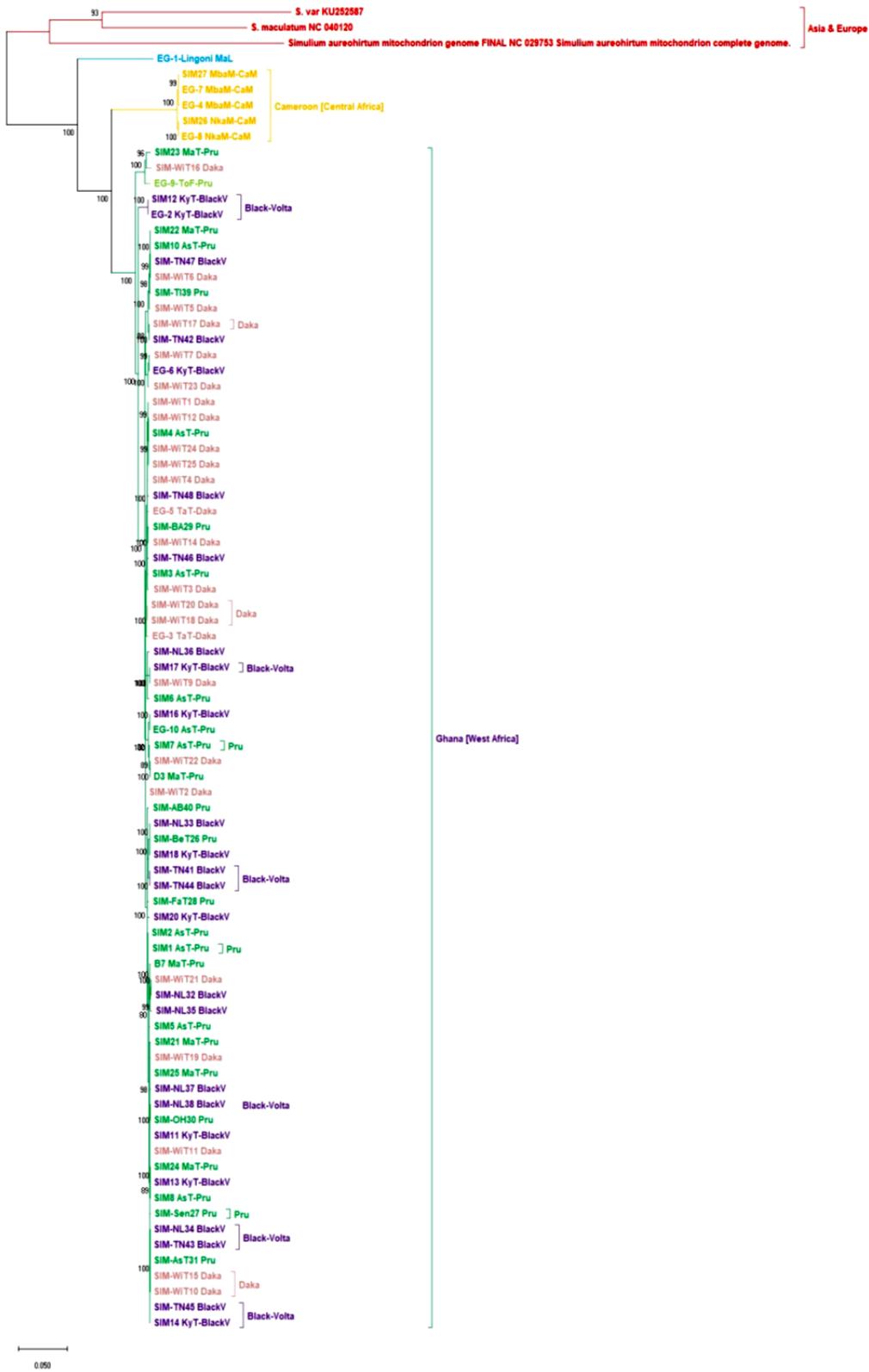


Figure 4.3.16: Maximum likelihood tree from 85 whole mitochondrial genomes from within and outside Africa. In the figure, Red = blackflies from Asia and Europe; Blue = blackfly from South-Eastern Africa (Malawi); Orange = blackflies from Central Africa (Cameroon); Green = blackflies from Pru river basin in West Africa (Ghana); Purple = blackflies from Black-Volta.

Black Volta river basin in Wester Africa (Ghana); Lemon Green = blackfly from Pra river basin in West Africa (Ghana); and Orange accent = blackflies from Daka river basin in West Africa (Ghana). The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible (GTR) model with Gamma distribution, and Invariant site (Nei and Kumar, 2000). The numbers on the branches indicate bootstrap values (expressed as percentages based on 1000 replicates). The scale bar (0.05) shows the average substitution per site.

4.4 Discussion

4.4.1 Genetic diversity of West African *Simulium* blackflies from matrilineal lineage

The blackflies in all the 3 river basins studied from Daka, through Pru to Black Volta along the East-West transect of the central ecological transition zone of Ghana were evolving randomly as one would expect from a neutrally evolving population. However, there was a high frequency of rare alleles (Tajima's $D < 0$) and this suggested recent selective sweep, population expansion after a recent bottleneck, linkage to swept genes. A similar observation was made with *Simulium aureohirtum* with significantly negative Tajima's D values across 3 studied lineages that were undergoing demographic expansion (Thaijarern *et al.*, 2014). Population expansion by blackflies at similar zone like those in Pru, Black Volta and Daka could be the result of shared recent history that could reduce the likelihood of accumulation of genetic differentiation between populations. Such blackfly species are capable of utilizing a wide range of ecological conditions as well as adapting to anthropogenic disturbed habitats including agricultural areas and artificial flowing water (Pramual and Kuvangkadilok, 2009b). The central transition ecological zone of Ghana has similar agricultural areas and flowing waters. While some blackfly species were observed to be confined to a given area such as forest or savannah and tend to have high levels of genetic structuring, others had continuous geographic distribution characterized by gene flow between populations (Pramual and Kuvangkadilok, 2012). The *Simulium* blackflies in Daka, Pru and Black Volta were genetically homogenous, possibly due to being in the transition ecological zone characterized by forest and savannah climate. This might be a contributing factor to their continuous geographic distribution. This finding is also consistent with studies of other species such as *Simulium siamense* that is genetically homogenous due to occupying both forest and open areas (Pramual *et al.*, 2011).

The blackflies in all the 3 river basins compared were evolving randomly as shown by the results of all the neutrality tests being statistically not significant. The absence of recombination in these mitochondrial genes possibly made all nucleotide sites evolve independently.

Both the total number of mutations and the number of segregating sites showed highest values in Daka, moderate in Black Volta and slightly lower in Pru. In the transect from East to West of the central ecological transition zone of Ghana, Pru river basin is found in between Black Volta and Daka. Consequently, Pru tends to share more mutations with Black Volta and Daka than Black Volta shares with Daka. This inevitably causes Pru to have the least

accumulated mutations that is not shared with any of the other river basins. A slightly different trend was observed with the number of singleton mutations across river basins with Daka showing the highest value, a relatively moderate value occurring in Pru, and the lowest observed in Black Volta. This is likely due to the blackfly migratory pattern in the direction of Daka, through Pru to Black Volta as a result of possibly following the course of river movement in search of breeding sites and food when the rivers attempt to move toward the Gulf of Guinea in one season, usually the dry season. On the return trip, usually the rainy season, the blackflies tend to move back to their previous habitat in the direction of Black Volta, through Pru to Daka. The net result is that there is more sharing of newly acquired mutations in Black Volta than Pru, and least in Daka.

The number of haplotypes (25) was the same in each of the river basins and, leading to an overall number of 75 haplotypes across all 3 river basins. The haplotype diversity was also the same and extremely high in each of the river basins ($Hd = 1$). The nucleotide diversity demonstrated a different pattern with Pru having the smallest value of 0.00787, while Daka had a relatively moderate value of 0.00830, and Black Volta had the highest value of 0.00898. The average number of nucleotide differences demonstrated a similar pattern as that of the nucleotide diversity with Pru having the lowest value of 126.8, a relatively moderate value of 133.6 value in Daka, and the highest value of 143.5 in the Black Volta.

The unique haplotypes further showed that the Ghanaian blackflies were genetically diverse such that the historical population size was possibly very large. The close relationship existing between the haplotypes in the 3 river basins give the indication of a shared evolutionary recent origin that is consistent with the rapid and large population expansion following an evolutionarily recent bottleneck that could have resulted from the relative recent arrival of the blackflies into Ghana. The high nucleotide and haplotype diversity showed that although the blackflies might have arrived recently in evolutionary time, there had been sufficient time to allow population structure to develop, if there had been significant force of barrier to reproduction among blackflies in the 3 river basins. The extremely high values of average within-population diversity ($Hs = 1$), within-population nucleotide diversity ($Ks = 142.7 - 153.0$), and the average number of nucleotide differences between populations in the 3 river basins, coupled with the majority of the genetic variation being found within populations (97.9%) than among population (2.1%), and the fixation index being closer to 0; all suggest significant gene flow between populations in the 3 river basins. Thus, indicating

significant blackfly movement among the 3 river basins and an inference that the blackflies might have been in the river basins for a bit longer than would have been expected from the suggestion of the founding population.

This is consistent with a small recent founding population, with those starlike related haplotypes (Figure 4.3.11 C), that had since diverged. There have been many reviews and reports of parasites of medical importance manipulating their vectors, such as the occurrence of infection resulting in altered vector feeding patterns to such extent that more host contacts are made (Molyneux and Jefferies, 1986; Moore, 1993). It was proposed that natural selection tends to favour parasites capable of manipulating their vectors to enhance their transmission (Hurd, 2003). The compatibility between sympatric as opposed to allopatric parasite-vector combinations indicates local adaptation capable of resulting in evidence of co-evolution (Woolhouse *et al.*, 2002) between *Simulium* blackflies and *Onchocerca* sp. It is possible that the population expansion of the West African blackflies took place following the speciation event that resulted in *Onchocerca volvulus* and *O. ochengi* due to the proposed host switch by their most recent common ancestor at the period of introducing domesticated cattle in Northeast Africa (Bain, 1981; Keddie *et al.*, 1998; Marshall and Hildebrand, 2002; Lefoulon *et al.*, 2017), and later to West Africa. Cattle were more abundant in the Northeast Africa than in West Africa and the switch from cattle to a human host by the parasite most likely occurred in areas characterized by more humans than cattle. Having been under selection pressure, the parasites might have initially reached a population bottleneck, and those few that survived in the human host might have possibly expanded in population following their survival and reproduction. *Onchocerca* parasites and simuliids have been implicated to have reciprocal effects on each other's phenotype and genotype. They further exert reciprocal effects on each other's survival at different stages of the life cycle of the parasite inside the blackfly, thus adapting to decrease deleterious effects on fitness while increasing transmission (Webster *et al.*, 2004; Basáñez *et al.*, 2009). Due to the nature of the survival and co-evolution, the population expansion of the parasites could have led to similar survival and population expansion of the vectors as suggested by this long-range mitochondrial data. This is consistent with the widely accepted principle that the success of a parasite is improved if it develops faster or if its vectors lived longer and one of these traits was under selective pressure (Koella, 1999).

4.4.2. *Simulium* population structure in West Africa

From Table 4.3.9, the majority of the blackflies' genetic variation occurred within populations (97.9%, $P > 0.001$), with the remaining variation occurring among populations (2.1%, $P > 0.001$). This also suggests a lack of geographic population structure by river basin. The test for the overall measurement of population differentiation due to genetic structure, overall Fixation Index (F_{ST}), was small (0.03) and approximately equal to zero. This small fixation index value closer to zero was similar to that of a panmictic population with random mating and greater gene flow between river basins. The force of gene flow in the population for a given generation (Nm), that influences the tendency of population differentiation because of genetic structure, was however large (F_{ST} 's $Nm = 16.25$). Both the negligible fixation index and the large Nm that is greater than 1 (Wright, 1951) support the idea of a lack of geographic population structure in the river basins across the central transition ecological zone of Ghana. The overall pairwise genetic distance, GammaSt statistic, was equally small (0.047) with an Nm value greater than the recommended value of 1 ($Nm = 10.17$) that is needed as the minimum value for gene flow between populations to occur. This further support the idea of a lack of population structure across the 3 river basins, thus adding support to the inference of a single onchocerciasis transmission zone in the central transition ecological zone of Ghana. The lack of population structure characterized by gene flow between the blackflies in the 3 river basins could be attributed to a number of reasons. A plausible contributing factor could be from the long-range vector migration of some blackflies. The *Simulium damnosum* subcomplex in the former OCP area showed changing distributions. The *S. sirbanum* and *S. damnosum* s.s. were widely distributed, occurring primarily in Ghana, Togo, Benin and Cote d'Ivoire. During the dry season, they were capable of spreading to Sierra Leone (Garms & Vajime, 1975; Vajime & Quillévéré, 1978; Quillévéré *et al.*, 1977). Deforestation, climatic changes and high mobility have been noticed to be the major driving force of some species of *Simulium* spreading from their main habitats to other areas they were never found. For instance, *S. sirbanum* was widespread in all OCP countries, but in Sierra Leone, it was only found in the extreme north of the country for very long time (Quillévéré *et al.*, 1977; Post & Crosskey, 1985). Due to deforestation, climatic changes and high mobility, *S. sirbanum* later spread to the south of Sierra Leone and other forested parts of the OCP area, including Liberia (Garms *et al.*, 1991). Vector control and flooding, such as from dams, could be a contributing factor affecting the distribution of blackflies. It could have caused them to decrease in number and even disappear completely from their principal location, and some of their larvae washed to areas they were not found previously. Such

seasonal flooding activities potentially caused populations that were formerly isolated by large geographical distances to survive and breed with those of other areas over the transition ecological zone. For instance, an observation was made that the proportion of *S. sirbanum* reduced among biting populations in Sierra Leone following larvicidal treatments. *S. sirbanum* and *S. damnosum* vanished downstream of the Akosombo dam in Ghana as a result of flooding activities of the Senchi Rapids by the Kpong Dam along the Volta River (Bissan *et al.*, 1995). The majority of dams are constructed to provide benefits like flood control, hydroelectricity, waterways creation and the provision of water for irrigation in regions experiencing drought (Yüksel, 2009; Biswas, 2012; Terminski, 2014). The implementation of dam construction and operation however comes with both benefits and negative impacts (McManus *et al.*, 2010). The quest to increase the energy production of Ghana led to the construction of Bui dam at the Bui Gorge, a project that was on the planning boards since the 1920s (Miescher and Tsikata, 2009; Okoampa-Ahoofe, 2009; Hensengerth, 2011). Seasonal flooding of the river before the dam construction occurred between June and November (Management, 2007). Flooding of the river became controlled to some extent by the management of the dam during periods of water level management for the operation of the dam. These flooding activities usually flood the Black Volta river basin and cause variation in the population and distribution of blackflies in that locality. Seasonal changes could have caused the blackfly distribution to change to the extent of making them occupy similar localities that they would have previously not done so. Over the passage of time, interbreeding could have caused gene flow, decreased genetic diversity and ultimately reduced or completely eliminated population structure. During the dry season in Ghana, populations became more restricted in the north and moved down south of their originally permanent distribution range. This is observed in the shortage of dry season breeding sites because most of the rivers, such as the Sudan savanna belt and some of the Guinea savanna, dry out entirely. The only exception to this is the few perennial rivers. A major contributing factor could be the displacement of the *Simulium* adults on the drying harmattan winds that seasonally blow in December onwards in the direction of south-west across West Africa, including the transition zone of Ghana, (Gwynne-Jones, 1978) towards the Gulf of Guinea. A reverse event occurs during the rainy season, usually occurring from June onwards, characterized by the monsoon winds blowing and displacing blackflies in the north-eastern direction (Garms *et al.*, 1979; Walsh *et al.*, 1981a; Baker *et al.*, 1990). This was the major contributing factor that resulted in the repopulation of breeding sites in previously vector-controlled areas (Garms *et al.*, 1979; Walsh *et al.*, 1981a; Baldry *et al.*, 1985; Baker *et al.*,

1990). The long-term epidemiological trend across the 3 major river basins were likely determined by both short-ranged and long-ranged events. The short-ranged migration pattern is consistent with the observation that onchocerciasis infection is greatest when closer to a river basin but progressively decreases as one moves away from the river basins. The long-range migration pattern on the other hand explains why there is gene flow between the two extreme river basins by distance, which are Black Volta and Daka. Such variations in the seasonal wind movement caused members of the *S. damnosum* complex, such as the savanna species, to migrate distances of 500 km and beyond on the wind (Garms *et al.*, 1979; Cheke and Garms, 1983). In areas of overlap, such as the transition ecological zone, different forms of blackflies that would otherwise form a distinct population structure in a non-overlapping area, are noted to interbreed and this most likely led to gene flow, decreased genetic diversity and reduced or complete absence of population structure (Boakye *et al.*, 1993b). Such overlap could sometimes lead to the occurrence of an intraspecific stepped cline (Surtees, 1988; Boakye *et al.*, 1993b). Another possible reason going hand in hand with the seasonal climatic changes and bi-directional wind movements is the search for breeding site. Changes in the vegetation of blackfly habitats influence their distribution and migration (Boakye *et al.*, 1998). The movement of natural water bodies such as rivers and streams in the transition zone could be in the direction of south-east or south-west towards the Gulf of Guinea, and seldomly moving up north. This movement potentially carries larvae from one point to another. Also, adult blackflies follow the water movement in that south-east or south-west direction in search of fast flowing water areas to breed, and this is consistent with the observation that capture points at slow-flowing river areas hardly produce migrating blackflies (Baker *et al.*, 1990). These movements in search of breading sites progressively make one form of blackfly move into the habitat of other blackflies, thereby making it rare to have any particular region constituting a permanent dwelling place for only one form of blackfly. Adult *Simulium* sp. often spread by short distance movements (Wenk, 1981) within the average distance range of as low as 7-15 Km to as high as 20-35 Km to cover the popularly cited total travel distance of 400-500 Km (Johnson *et al.*, 1985; Garms *et al.*, 1989; Baker *et al.*, 1990). Such short distance movements that ultimately cover large total distances can lead to gene flow in blackflies occupying different river basins to subsequently cause a reduction in the genetic diversity and population structure. Moreover, the average distance between any two river basins studied in the transition ecological zone of Ghana was within the 400 - 500 Km blackfly flight range, thereby making it possible for blackflies to migrate in

the course of time from one river basin to another (Garms *et al.*, 1979; Magor and Rosenberg, 1980; Garms, 1981; Garms, 1982; Johnson *et al.*, 1985; Baker *et al.*, 1990; Post *et al.*, 2013).

Another possible reason for the lack of population structure among the *Simulium* blackflies could be a result of introgression. The genetic introgression could be due to interspecific hybridization (Post, 1984; Boakye and Meredith, 1993). Interspecific hybrid individuals have been observed on several occasions among wild larvae of *Simulium damnosum* Theobald complex (Vajime and Dunbar, 1975; Post, 1984; Boakye and Mosha, 1988). There have been great challenges in determining if these hybrids can develop into fertile adults due to the difficulties in rearing or breeding blackflies in laboratories. Despite these challenges, it was observed in a laboratory cross-mating study that male hybrids from a cross between *S. soubrense* Vajime & Dunbar beffa form and *S. squamosum* Enderlein demonstrated the same mating propensity as beffa males (Meredith *et al.*, 1987). They were however not backcrossed to either of the parental forms to determine the status of their fertility. There was a report of the natural occurrence of progeny developed from a backcross between *S. sanctipauli* / *soubrense* V. & D. hybrid and *S. squamosum* (Post, 1984). This showed that there could be fertile hybrids. The estimated low frequency of hybridization (0.0009) led to the suggestion that it could not be considered as important for the occurrence of genetic introgression among the *S. damnosum* complex members. Nonetheless, there have since been records of individuals known to be backcross progeny in at least six places in the area that were covered by the OCP in West Africa (Post, 1984).

4.4.3 Implications of the *Simulium* blackfly genetic diversity, divergence and population structure on Delineation of Transmission Zones

Blackflies in a river basin that tend to interbreed more with each other will show less diversity and divergence while the reverse effect will be true for blackflies that interbreed less with each other. Geographical distances largely explain sections of genetic diversity in the populations of organisms (Barbujani & Sokal 1991; Excoffier *et al.*, 1991; Cavalli-Sforza *et al.*, 1994; Poloni *et al.*, 1995, 1997). In addition, genetic differences show inverse relationship to the amount of gene flow under the isolation-by-distance model. The amount of such gene flow is dependent on geographical proximity between populations (Wright 1943; Morton *et al.*, 1968; Malecot 1973). In the central ecological transition zone of Ghana however, geographical distances between the river basins do not have any significant association with genetic distances or diversity of the blackflies.

A population may be defined as “a group of individuals of the same species” (Berryman, 2002). Thus, population structure may refer to the existence of systematic variation in allele frequencies between subgroups within a group of individuals of the same species, possibly resulting from different ancestry. The cause of population structure may be due to non-random mating between groups of individuals as a result of some form of reproductive barrier, followed by genetic drift of allele frequencies in each group. Although nucleotide diversity was generally lower across the river basins, the haplotype diversities were very large and consistent with that of a panmictic population characterized by significant gene flow. There is a small probability value under panmictic population structure for any given large observed value of K_{ST} or the proportion of the ratio of the diversity within a population to that in the entire population (Hudson *et al.*, 1992).

River basins characterized by the absence of gene flow between their blackflies, but with the existence of significant population structure, have a greater likelihood of being separate transmission zones of onchocerciasis. *A priori* DAPC analysis revealed the existence of 3 distinct molecular groups of blackflies and one of these may be genetically heterogeneous. The cause of this genetic differentiation under such shared large-scale geographical distance location need further studies. It could possibly result from reproductive barrier because migratory movements are not only a function of geographical distance but are also affected by the existence of ecological and cultural barriers (Barbujani and Sokal 1990, 1991, 1991; Sokal and Oden 1988; Sokal *et al.*, 1989, Dupanloup de Ceuninck *et al.* 2000; Rosser *et al.*, 2000). Thus, the 3 molecular groups that might be genetically heterogeneous across the central ecological transition zone of Ghana most likely resulted from barriers of reproduction existing within the 3 river basins but influencing the different blackflies in different ways. Although these distinct molecular groupings of blackflies share a single onchocerciasis transmission zone, their role in transmission needs studying and there are many techniques available to identify the presence of genetic barriers as well as to determine their role in population differentiation (Barbujani *et*. 1989; Barbujani and Sokal 1990; Stenico *et al.* 1998; Simoni *et al.* 1999).

4.4.4 Implications of the *Simulium* blackfly genetic diversity, divergence and population structure for Decisions to Stop Interventions

This study shows that the entire central ecological transition zone of Ghana constitutes a single onchocerciasis transmission zone and this finding does not agree with the transmission zone definition based on *Simulium* vector breeding sites (WHO, 2016). Based on the current

river basin definition of transmission zone, if ivermectin and other onchocerciasis interventions stop in communities in any given river basin, our study suggests that infected blackflies from the other river basins will potentially travel to reinvoke the communities that have stopped ivermectin treatment. There needs to be a comprehensive assessment of the migration, diversity, divergence and population structure of *Simulium* blackflies in the onchocerciasis endemic countries. This can lead to a better definition of onchocerciasis transmission zone, delineation of a reliable transmission zone, greater accuracy in the estimation of timelines to stop MDA using ivermectin, and a greater chance of guarantee to sustain the progress made in decades of onchocerciasis.

4.4.5 Decision on objectives and hypotheses

A total of 82 complete mitochondrial genomes were sequenced, assembled and annotated. The genetic diversity of the blackflies in the central ecological transition zone of Ghana was characterized by using 75 complete mitochondrial genomes. Thus, the haplotype diversity was extremely large, a value of 1, in each of the river basins. The nucleotide diversity ranged from 0.00787 to 0.00898. Similar to the short amplicon sequence data, the whole mitochondrial genomes demonstrated that there was no population structure of the blackflies in the 3 river basins of Pru, Black Volta and Daka. A single onchocerciasis transmission zone in the entire central ecological transition zone of Ghana was confirmed and this was concordant with the parasite data from different research works in our lab.

A total of 82 libraries prepared from total DNA and sequenced at low depth enabled the assembly and annotation of 82 complete mitochondrial genomes of *Simulium* blackflies from Africa. Therefore, the null hypothesis 1 was accepted and the alternative hypothesis 1 was rejected.

The complete mitochondrial genomes of 75 blackflies enabled the characterization of the genetic diversity of the *Simulium* blackflies in the transition zone of Ghana. Hence, the null hypothesis 2 was accepted and the alternative hypothesis 2 was rejected.

The use of the 75 complete mitochondrial genomes showed that there was no population structure of the blackflies by river basin in the transition zone of Ghana. Therefore, the null hypothesis 3 was rejected and the alternative hypothesis 3 was accepted.

In agreement with the parasite data of the research work of colleagues in our lab, the 75 complete mitochondrial genomes demonstrated that the entire central ecological transition

zone of Ghana is made up of a single onchocerciasis transmission zone. All the objectives of this study were fulfilled.

4.5. Conclusion

The complete mitochondrial genomes of *Simulium* blackflies from Africa were found to be circular molecules of 16108 to 16232 nucleotides, and each of them had 2 rRNA genes, 22 tRNA genes, 13 protein coding genes and a non-coding control region (D-LOOP). This was like the mitochondrial genomes of other *Simulium* blackflies from Asia and Europe, as well as other dipterans.

A total of 82 whole mitochondrial genomes of African *Simulium* blackflies from West, Central and South-Eastern Africa were sequenced, assembled and annotated. the African *Simulium* blackflies were monophyletic to those from Asia and Europe. The West African blackflies from Ghana were monophyletic to all the other blackflies from Central and South-Eastern Africa. The whole mitochondrial genomes produced phylogenetic trees with better bootstrap support values than the short amplicon resequencing, long-range sequencing and the nuclear markers. The haplotype diversity was the same, a value of 1, in each of the river basins. This value of 1 was higher than the average haplotype diversity (mean $h = 0.63388$) expected in other related organisms. Observation was made in the localities studied in West-Africa of a higher nucleotide diversity than observed in other related organisms (mean $\pi = 0.00388$), with individual values of 0.00787 in Pru, 0.00898 in Black Volta and 0.00830 in Daka river basin. Thus, the overall genetic diversity in the mitochondrial genome of the West African *Simulium* blackfly was characterized as 1 and 0.00853 for the haplotype diversity and nucleotide diversity respectively.

The sequences within the 3 river basins were all evolving randomly as one would expect from a panmictic population with random mating and greater gene flow between the populations. Observation was made of the presence of 3 molecular groups of blackflies and one of these may be genetically heterogeneous. There was significant gene flow between blackflies in the 3 river basins, reflecting significant migration between river basins throughout the central ecological transition zone of Ghana. The entire central ecological transition zone of Ghana constitutes a single onchocerciasis transmission zone.

Chapter 5: Whole nuclear genome contribution to *Simulium* population genetics

5.1 Introduction

5.1.1 Organization and function of nuclear genome

Insect nuclear genomes are composed of 200 Mbp to 3 Gbp and they are arranged as a mosaic of 20 to 50 thousand genes and noncoding DNA. The association between its components, such as the relationship between heterogeneous nuclear RNA and mRNA, has been explored (Lewin, 1975; Spradling and Rubin, 1981). Genes may occur in single copies, moderately repeated copies, or highly repeated copies to the magnitude of hundreds to thousands of times when they code for products that occur abundantly in a cell. About 10 to 70 percent of arthropod nuclear DNA is made up of moderate to highly repetitive sequences. While many of the longer minisatellite DNA's are found in the centromeres and telomeres of the chromosomes, most of the smaller repeats, such as microsatellites, are spread throughout the genome (Lewin, 1990; Hartl, 1991; Palmer *et al.*, 1994).

The genomes of eukaryotic cells are packed several thousandfold into confines of the cell nucleus but at the same time maintain access to gene and chromatin structure required for essential roles in gene transcription, replication, and DNA repair. There is the organization of interphase chromosomes into distinct locations and these are nonrandomly distributed in relation with the nucleus and chromosomes. The placement of the interphase chromosomes is capable of affecting *trans* interactions and determining if a genomic locus will be in an active or repressive nuclear environment (Fraser and Bickmore, 2007; Cremer and Cremer, 2010). Within the nucleus can be found the nuclear bodies seen as the structures that enhance the efficiency of several nuclear processes. The nuclear bodies take part in vital processes like transcription, splicing, processing, as well as epigenetic regulation (Mao *et al.*, 2011). Active genes are known to relocate from chromosome territories (Chambeyron *et al.*, 2004; Branco and Pombo, 2006). They subsequently group into the subnuclear foci called transcription factories for the expression of genes (Chakalova and Fraser, 2010). Nuclear organization pertains to spatial arrangement of chromosomes and their location in relation to the nuclear periphery associates with chromatin structure and gene expression. The interactions of chromatin at the nuclear periphery have been mapped in humans (Guelen *et al.*, 2008) and *Drosophila* (Pickersgill *et al.*, 2006). The mapping showed large defined lamina-associated domains that are associated with transcriptional repression and low gene density (Guelen *et al.*, 2008; Zullo *et al.*, 2012).

5.1.2 Nuclear and cytoplasmic DNA considerations in phylogeographical works

Cytoplasmic DNA has for very long time been the primary choice in the consideration of phylogeographical works relating to the initial characterization of population structure, determination of maternal gene flow and testing population monophyly (Avise, 1995; Moore, 1995). Works of this kind with phylogeographical objectives often depend on the use of nuclear data to augment the initial results produced from cytoplasmic loci inputs. But, phylogeographical structures are theoretically expected to be less pronounced at the diploid nuclear loci in comparison with cytoplasmic loci because of the effective population sizes (McCauley, 1995; Moore, 1995). Although the effective population size of a locus associates with the number of breeding adults, it could either be larger or smaller than such a number based on the ploidy and mode of inheritance. The magnitude of phylogeographical structure in both uniparentally and biparentally inherited markers is affected by differential dispersal by gender (Palumbi *et al.*, 1994; McCauley, 1995). Also, stochastic variation associated with genealogical differentiation potentially creates challenges in resolving population processes from cytonuclear comparisons (Buonaccorsi *et al.*, 2001).

Phylogeographical application to the determination of onchocerciasis transmission zones must test for monophyletic groups because prolonged isolation usually leads to evolutionary distinctiveness or uniqueness of a population (Baker *et al.*, 2002). Genetic drift causes neutral gene monophyly in a population that has been isolated for a long time. The processes giving rise to monophyly at the nuclear autosomal loci (e.g. mutation and genetic drift) are often too slow to normally expect deep intraspecific partitions like the ones seen in animal mitochondrial DNA (Palumbi *et al.*, 2001; Baker *et al.*, 2002). It is generally expected that species or populations within species that diverged during the Pleistocene (2.5M to 12,000 years B. P.) would occupy a mixed-monophyly zone of divergence such that mitochondrial DNA will have a higher likelihood of monophyly while the average nuclear locus will not. The numerous nuclear and mitochondrial data from diverse taxa could potentially offer the initial null expectation for the level of gene tree congruency across loci (Palumbi *et al.*, 2001). A low expectation of nuclear monophyly therefore exists in recent population isolates. Most likely, there is limitation in the number of population processes that can be inferred from a monophyletic population by the age of the most recent common ancestral allele (the coalescent). Attempts to make inferences from ancient populations will inevitably need to consider the relatively deep coalescent times at the nuclear loci (Harding *et al.*, 1997; Fu and Li, 1999).

5.1.3 Nuclear gene and intragenic recombination

As the intragenic recombination rate approaches the nucleotide substitution rate at a given locus, haplotypes tend to have more than one immediate ancestor, and different segments within a haplotype tend to possess independent histories. Failure to consider recombination in a nuclear gene can therefore lead to the introduction of homoplasy and jeopardize phylogeny reconstruction, bias a gene tree shape to produce a pattern like that of population expansion, and ultimately ruin the molecular clock (Schierup and Hein, 2000). There have been advances in the techniques for estimating recombination rates and the detection of recombinants in aligned sequences (Crandall and Templeton, 1999; Kuhner *et al.*, 2000). Methods used in the reconstruction of gene trees in sequences characterized by recombination include: (a) the construction of a network that shows the reticulate haplotype relationships as a result of recombination (Posada *et al.*, 2001); (b) the building of a gene tree from each partition following the separation of linked clusters of polymorphic sites from recombination break points in a locus (Templeton *et al.*, 2000); and (c) the removal of noticeable recombinants that are few before (Harding *et al.*, 1997) or together with phylogenetic analysis (Templeton *et al.*, 2000). Recombination has characteristics that prevent genetic shuffling within populations from masking strong phylogeographical patterns. The effect of recombination, resulting from reduced gene flow between populations that are isolated, is usually confined to within each population. This subsequently enables dichotomous phylogenetic structure to develop in the isolated populations (Bernardi *et al.*, 1993). Another characteristic is that recombination rates are spatially heterogeneous on chromosomes, thereby enabling the option of selecting a preferred level of recombination for different purposes (Kaessmann *et al.*, 1999). Different conditions create the opportunity to select regions with either low or high recombination rates. If independent selective sweeps increase the genetic differentiation between different populations, then regions of the chromosomes characterized by low recombination can provide good phylogeographical resolution (Stephan *et al.*, 1998). On the other hand, historical demography studies can occur by capitalizing on the effects created by chromosome regions of high recombination (Wakeley and Hey, 1997).

5.1.4 Nuclear phylogeography using targeted loci to develop species tree

Phylogenetic interpretation that depends on DNA-sequence differences can lead to wrong conclusions even when the gene tree has already been correctly resolved. This kind of mistake results when an ancestral species is polymorphic for the gene and sorting of lineages produced from the alternate alleles finally leads to a gene tree that is not concordant with the species tree (Neigel and theory, 1986; Nei, 1987; Pamilo and Nei, 1988; Avise, 1989; Avise,

1990; Wu, 1991; Hudson, 1992). Mitochondrial sources of DNA have over the years been used in phylogenetic works due to their fast rate of evolving and their ability to give a lot of genotypic characters, either by analysis of restriction-fragment polymorphism (Brown, 1983; Moritz *et al.*, 1987; Dowling, 1990) or amplification by the use of PCR and nucleotide sequencing (Kocher *et al.*, 1989; Edwards *et al.*, 1991; Ruvolo *et al.*, 1991; Lanyon, 1992). However, lineage sorting minimizes the full potential of its usefulness as a result of the mitochondrial genes being inherited as a single linkage group. Consequently, it does not lead to independent estimates of the species tree. In comparison, nuclear genes can be taken from unique chromosomes in such a way that each gene tree provides an independent estimate of the species tree. In addition, a short internode between the first and second bifurcation in a three-species section of a phylogeny leads to high probability that a gene tree and a species tree will not be concordant (Nei, 1987; Pamilo and Nei, 1988; Wu, 1991). A better way to develop and interpret a species tree is to use many independent gene trees (Pamilo and Nei, 1988; Wu, 1991).

Observations have been made across numerous taxa of the existence of lower substitution rates in nuclear DNA that reduces the frequency of back and parallel mutations, and which can also minimize phylogenetic resolution in mitochondrial DNA data. Consequently, the presence of even a single fixed difference in a data with low homoplasy can create a statistically strong result at the intraspecific level irrespective of the bootstrap support (Harris and Hey, 1999). For this reason, there will be the need to use nuclear sequences that are longer than those sampled from mitochondrial DNA in order to be able to obtain sufficient sampling of phylogenetically informative characters.

More variation occurs in noncoding nuclear DNA than in coding sequence due to the absence (in most cases) of section constraints. This creates the possibility of creating primers in 2 relatively conserved exons to perform exon primed-intron crossing for the purpose of targeting variable intron DNA genes within the 2 conserved exons (Palumbi, 1996). The design and use of universal primers across different taxa have promoted the collection of many nuclear DNA data in vertebrates (Friesen *et al.*, 1997; Lyons *et al.*, 1997; Friesen *et al.*, 1999; Quattro and Jones, 1999), invertebrates (Palumbi, 1996) and plants (Strand *et al.*, 1997). The isolation of DNA fragments by PCR methods sometimes amplifies paralogs associated by gene duplication event and found at chromosome multi-loci, in addition to the orthologous alleles from each of those loci. PCR reaction conditions have a great role in determining whether or not multiple loci will be co-amplified (Hare *et al.*, 1996; Bagley and Gall, 1998). Such challenges can make nuclear DNA amplification by PCR methods to

become quite frustrating, expensive and time consuming. For this reason, whole genome amplification methods will be better options to generate more nuclear data.

5.1.5 Justification (Rationale) of the study in chapter five

The high genetic diversity of the short amplicon, long-range and complete mitochondrial genome sequence data did not identify population differentiation of the blackflies by river basin. The *a priori* DAPC was capable of identifying 3 distinct molecular groups or clusters of blackflies, thus suggesting that the mitochondrial markers were diverse enough to determine, if present, population structure by river basin. The inability of the mitochondrial markers to identify any population structure by river basin simply demonstrate that it does not exist, and it is not because there is not sufficient genetic diversity. All the markers used so far only show what is happening matrilineally without any indication of what is happening from the patrilineal or bi-parental side of the lineage.

It therefore became imperative to use more genetic data from nuclear sources to verify if they were indeed concordant with those of the mitochondrial genome. The general expectation was that if there was complete lineage sorting, then the mitochondrial markers must be concordant with the nuclear markers.

5.1.6 Specific objectives, mode of assessment and hypotheses

5.1.6.1 Specific objectives of the study

1. To characterize the genetic diversity of *Simulium* blackflies in the transition zone of Ghana by using nuclear genes
2. To determine onchocerciasis transmission zone status in the central ecological transition zone of Ghana by using nuclear genes

5.1.6.2 Mode of assessment

Similar mode of assessment of the specific objectives as used in the section 3.1.3.2 of this thesis was used to assess the specific objectives of this chapter

5.1.6.3 Hypotheses of the study

Null hypothesis 1

Concatenated nuclear genes of *Simulium* blackflies from Ghana will enable the determination of the genetic diversity of the blackflies

Alternative hypothesis 1

Concatenated nuclear genes of *Simulium* blackflies from Ghana will not enable the determination of the genetic diversity of the blackflies

Null hypothesis 2

Concatenated nuclear genes of *Simulium* blackflies will enable the determination of the onchocerciasis transmission zone status of the central ecological transition zone of Ghana

Alternative hypothesis 2

Concatenated nuclear genes of *Simulium* blackflies will not enable the determination of the onchocerciasis transmission zone status of the central ecological transition zone of Ghana

Mode of assessment of objectives

The assessment of the outcome of the objectives of this chapter will be made using similar criteria as in the “Mode of assessment of specific objective” in section 3.1.3.2 of this thesis.

5.2 Materials and methods

5.2.1 Sample collection

Samples were collected from the same locations and by the same sampling strategy as the mitochondrial genome work (see section 4.2.1).

5.2.2 DNA extraction

DNA was extracted using the Isolate II Genomic DNA Kit (Bioline) per the procedure outlined in section 3.2.2.

5.2.3 DNA concentration determination

The concentration of the extracted DNA was determined using the procedure shown in the section 2.2.4 of this thesis.

5.2.4 Library preparation, quantification, normalization and pooling

The libraries were prepared using the Nextera DNA XT and DNA Flex library prep procedure described in section 4.2.4. of this thesis.

5.2.5 NextSeq sequencing of libraries

The 2x150bp paired end reads chemistry was used to sequence the libraries (details are shown in section 4.2.8.

5.2.6 Sequence import

The Illumina High-Throughput Sequencing Import tool option in the CLC Genomics workbench 9.5.4 was used to import the next generation sequences by following the procedure used for the mitochondrial genome work (shown in section 4.2.9).

5.2.7 Trimming and quality filtering

The procedure described in section 4.2.10 was used to trim the reads and remove duplicate sequences. Possible contaminated sequences were removed by following the procedure in section 4.2.11 of this thesis.

5.2.8 Assembly and mapping of reference nuclear genome of African *Simulium* blackfly

The procedure for the *de novo* assembly to obtain the nuclear genes was the same as the one used for the whole mitochondrial genome (see section 4.2.12), except for a minor part indicated in this section of the thesis. After producing the assembly for the complete mitochondrial genome, its reads were completely excluded from subsequence analysis. All that remained, the nuclear reads, were retained for the nuclear genome analysis.

A nuclear reference composed of a panel of commonly used nuclear barcoding genes was constructed as follows (summarised in Figure 5.2.1). Individual sequences from Ghana and other parts of Africa were obtained from public database and used as a putative reference panel against which sequencing reads for individual flies were then mapped. Any of these putative reference sequences that demonstrated poor mapping coverage were excluded from the final list of gene sequences for developing the main nuclear reference panel. The preliminary mapping of the sequenced blackfly reads to nuclear genes from public database showed that 7 commonly used barcoding nuclear genes consistently had good read coverage (depth > 50) across all the individual blackflies sequenced, hence were used for the reference nuclear panel development. The 7 genes chosen for the reference panel were 18S-rRNA, ITS1, 5.8S-rRNA, IGS, 2S-rRNA, ITS2, and 28S-rRNA.

A total of 464 nucleotide sequences for the reference panel above were downloaded from the NCBI nucleotide database (accession numbers are shown in Supplementary Table 4.3.8 A and B). They were aligned gene by gene by Muscle (Edgar, 2004) and ClustalW (Thompson *et al.*, 1994) and a concatenated consensus reference sequence extracted. The nuclear reads of each individual blackfly were then mapped onto the concatenated nuclear consensus sequence. The assembly mapping options used were: map reads back to contigs (slow), which is known to create assemblies with higher rate of accuracy despite requiring longer time to complete, mismatch cost of 2, insertion cost of 3, deletion cost of 3, length fraction of 0.5, and similarity fraction of 0.86 that corresponds to the 86% identity threshold of African *Simulium* blackflies. From this mapping was extracted a consensus sequence representing the corresponding concatenated nuclear sequence of individual blackflies from the sampled area. This sequence was validated against closely related *Simulium* and other dipteran gene sequences on NCBI nucleotide database by comparing the gene size, content and percentage identity. The consensus sequence that had similar size, content and percentage identity to those of published *Simulium* gene sequences became the draft reference nuclear genome sequence of the African *Simulium* blackflies. Using this draft reference nuclear genome sequence of *Simulium* blackfly from the sampling location in Arica, the mapping process with each individual blackfly reads was repeated. The consensus sequence was once again extracted from each mapping and validated as before. The consensus sequences of each of the 82 individual blackflies from the sample locations were aligned and used for downstream molecular phylogenetic and population structure analysis similar to those carried out for the whole mitochondrial genome (shown in section 4.2.18). The assembly and mapping process

for generating the reference nuclear genome sequences of African *Simulium* blackflies are shown in figure 5.2.1 below.

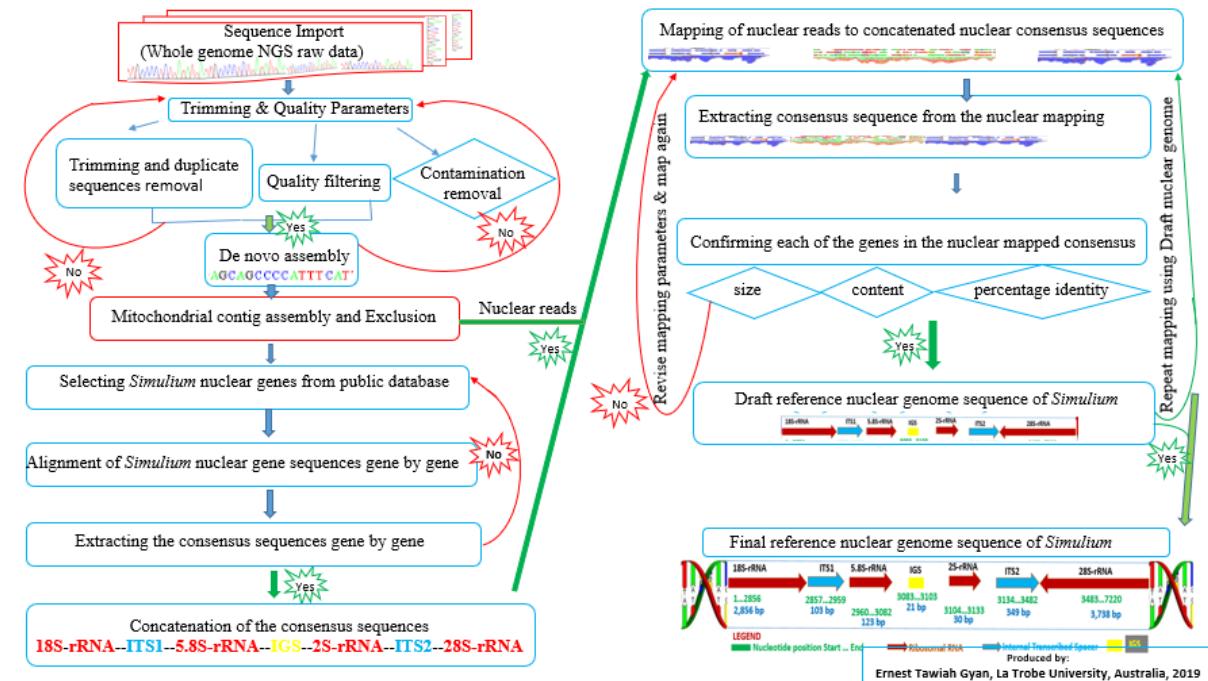


Figure 5.2.1: Flow diagram of the assembly and mapping process for generating reference nuclear genome of African *Simulium* blackflies.

5.2.9 Molecular phylogenetic and statistical analysis

The percentage completeness of the genome of each individual sample was assessed using the Core Eukaryotic Genes Mapping Approach (CEGMA) described in other studies (Parra *et al.*, 2007; Parra *et al.*, 2008).

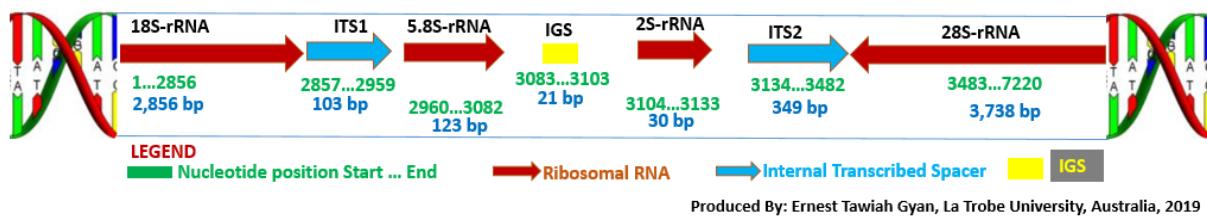
In the phylogenetic tree construction using the maximum-likelihood estimate, the initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.1000)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 49.28% sites).

Similar methods of the molecular phylogenetics and statistical analysis used in section 4.2.18 of the chapter four of this thesis was used in this chapter.

5.3 Results

5.3.1 Concatenated nuclear gene characteristics

The 7,220 bp reference sequence of the nuclear genome barcoding gene analysis consisted of a concatenation of the sequences of 18S-rRNA (2,856 bp), ITS1 (103 bp), 5.8S-rRNA (123 bp), IGS (21 bp), 2S-rRNA (30 bp), ITS2 (349 bp), and 28S-rRNA (3,738 bp). This is shown in figure 5.3.1.



Produced By: Ernest Tawiah Gyan, La Trobe University, Australia, 2019

Figure 5.3.1: Schematic diagram of a reference nuclear gene sequence. The reference nuclear gene sequence was produced by concatenating the consensus sequences of various *Simulium* sp. nucleotide sequence alignments of the represented genomic sequences. The initial *Simulium* sp. nucleotide sequences were extracted from the NCBI nucleotide database. The arrows show the orientation of the genomic sequence.

5.3.2 Assessment of whole genome completeness of African *Simulium* blackflies

The library and sequencing approach used in this study enabled the production of complete mitochondrial genomes as the primary targets, and some nuclear genes as minor targets. Thus, the main objective was to sequence and assemble a complete mitochondrial genome per sample. The sequencing of some nuclear genes was a minor target at this stage of work. It was never the intent of this project to achieve nearly 100% complete genome sequencing of each sample, as long as a complete mitochondrial genome was sequenced per sample, with sufficient numbers of nuclear genes that will enable the fulfillment of the primary aims of this study. The number of the complete 248 ultra-conserved core eukaryotic genes (CEGs) present in the 82 assembled genomes ranged from 2 to 197 (1.4 to 87.9%), with an average of 54.2 (37.6%). The number of the partial 248 ultra-conserved CEGs present across the genomes ranged from 4 to 239, with an average of 132.5. These led to percentage completeness range of 0.8 to 79.4 across the 82 genomes with an average of 21.8 for the presence of complete 248 ultra-conserved CEGs, while that for partial CEGs ranged from 1.6 to 96.4% with an average of 53.4. Thus, the overall percentage of completeness, taking into consideration both

complete and partial CEGs, ranged from 1.4 to 87.9% with an average of 37.6%. The total number of complete CEGs present, including putative orthologs, across the 82 assembled genomes ranged from 2 to 325 with an average of 101.8. However, the total number of partial CEGs present, including putative orthologs, across the 82 sequenced genomes ranged from 5 to 485 with an average of 283.4. The average number of orthologs per complete CEG ranged from 1 to 2.4 with an average of 1.9, while that for partial CEG ranged from 1.3 to 2.4 with an average of 2.1. The percentage of detected complete CEGs that had more than 1 ortholog ranged from 0 to 77.5% with an average of 55.8%, while that for partial CEGs that had more than 1 ortholog ranged from 25% to 74.2% with an average of 62.8%. The number of missing proteins across the 82 assembled genomes from the category of the complete CEGs assessment ranged from 51 to 246 with the average of 193.8, while that for the partial CEGs assessment category ranged from 9 to 244 with an average of 115.5. Details of the completeness assessment of the assembled genomes is shown in Table 5.3.0. While the percentage of completeness of the genomes varied, extraction of complete mitochondrial genome was achieved from all the 82 assemblies.

Table 5.3.0: Statistics of the completeness of some of the *Simulium* genomes sequenced

Sample ID	Status	#Prots	%Comp	Tot Com	#Total	Mean	%Ortho	MP
D3	Complete	197	79.44	87.905	318	1.61	44.67	51
	Partial	239	96.37		485	2.03	60.25	9
B7	Complete	193	77.82	86.49	325	1.68	47.67	55
	Partial	236	95.16		445	1.89	53.81	12
EG1	Complete	32	12.9	34.675	72	2.25	71.88	216
	Partial	140	56.45		303	2.16	72.14	108
EG2	Complete	28	11.29	29.635	60	2.14	71.43	220
	Partial	119	47.98		247	2.08	63.87	129
EG3	Complete	40	16.13	34.68	96	2.4	77.5	208
	Partial	132	53.23		305	2.31	74.24	116
EG4	Complete	30	12.1	33.67	69	2.3	73.33	218
	Partial	137	55.24		309	2.26	70.8	111
EG5	Complete	58	23.39	45.77	128	2.21	65.52	190
	Partial	169	68.15		374	2.21	67.46	79
EG7	Complete	13	5.24	22.58	28	2.15	53.85	235
	Partial	99	39.92		202	2.04	62.63	149
EG8	Complete	41	16.53	34.475	82	2	63.41	207
	Partial	130	52.42		289	2.22	72.31	118
EG9	Complete	31	12.5	34.88	64	2.06	61.29	217
	Partial	142	57.26		333	2.35	67.61	106
EG10	Complete	36	14.52	38.915	76	2.11	61.11	212
	Partial	157	63.31		348	2.22	64.97	91
SIM-1.NS	Complete	3	1.21	1.41	4	1.33	33.33	245
	Partial	4	1.61		5	1.25	25	244
SIM-WiT1	Complete	2	0.81	4.035	2	1	0	246
	Partial	18	7.26		39	2.17	61.11	230

The statistics of the completeness of the African *Simulium* genomes sequenced was based on 248 Core Eukaryotic Genes (CEGs). In the table, #Prots = number of 248 ultra-conserved CEGs present in genome; %Comp = percentage of 248 ultra-conserved CEGs present; Tot Com= Overall percentage of 248 ultra-conserved CEGs present; #Total = total number of CEGs present including putative orthologs; Mean = average number of orthologs per CEG; # %Ortho = percentage of detected CEGS that have more than 1 ortholog; and MP= Number of Missing Proteins

5.3.3 Nuclear mapping and reads coverage

Generally, the sizes of the consensus sequences of the *Simulium* blackflies from Africa that mapped to the reference nuclear gene sequence ranged from 7147 bp (in Nkam river basin in Cameroon) to 7172 bp (in Daka river basin in Ghana). The size of the consensus sequence produced from the reads that mapped to the blackfly reference sequence in the Lingoni River

basin in Malawi (7154 bp) was larger than those in Cameroon but smaller than those in Ghana. In all, there were a total of 82 individual samples with each having the complete concatenated nuclear sequences. The sampling locations of the 82 individual samples are as shown in section 5.2.1 of this thesis.

The quality distribution plots, with one representative below (Figure 5.3.2), showed that about 95% of the sequences had average of at least 30 PHRED score, and more than 80% had a PHRED score of over 35. This indicated that the chances that at least 95% of the bases used in this nuclear DNA sequence analysis were called incorrectly were 1 in 1000, with 99.9% base call accuracy. The percentage of this base call accuracy became better for 80% of the sequences beyond the 99.9% base call accuracy.

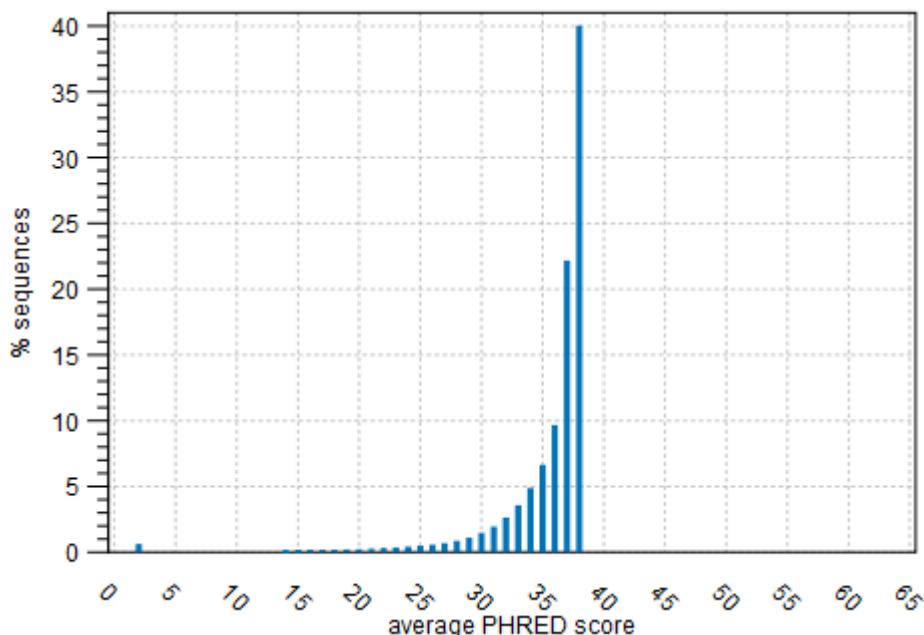


Figure 5.3.2: Quality distribution plot. The graph shows the distribution of average sequence quality scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities. x: PHRED-score. y: number of sequences observed at that quality score normalized to the total number of sequences.

Four (4) different next generation sequencing experiments were carried out, with 2 MiSeq (2x300bp paired end read chemistry) and 2 NextSeq (2x150bp paired end read chemistry) as shown in Table 5.3.1 below. Out of the 83 individual blackfly libraries prepared, 98.8% (82) sequenced successfully and yielded a total of 391,383,222 reads. The 1 library that failed to sequence properly produced 9,868 total reads, but these were predominantly of index sequences. Out of the total number of 391,383,222 reads generated across the 4 different

types of sequencing, the number of reads produced in the MiSeq 1, MiSeq 2, NextSeq 1, and NextSeq 2 were respectively 37,310,772 (9.5%), 23,964,252 (6.1%), 157,453,130 (40.2%), and 172,655,068 (44.1%). The total number of sequenced reads that mapped to the concatenated nuclear DNA reference sequence of 18S-rRNA, ITS1, 5.8S-rRNA, IGS, 2S-rRNA, ITS2, and 28S-rRNA were 0.43% (158,780), 0.60% (144,006), 0.41% (640,546) and 0.36% (618,225) for MiSeq 1, MiSeq 2, NextSeq 1, and NextSeq 2 respectively. Thus, the amount of the reads that mapped to the reference sequence was greatest in the NextSeq 1 (41.0%), followed by the NextSeq 2 (39.6%), and lower in the MiSeq 1 (10.2%), with the MiSeq 2 (9.2%) showing the least. In order of decreasing magnitude, the minimum reads coverage observed were 1,432 (MiSeq 1), 219 (MiSeq 2), 128 (NextSeq 1), and 76 (NextSeq 2). The average read coverage was however relatively high in MiSeq 1 (1,680), moderate in NextSeq 2 (531), low in MiSeq 2 (343), and lowest in NextSeq 2 (265). In order of increasing magnitude, the maximum read coverages were 3554, 7637, 13321, and 24013 for MiSeq 2, NextSeq 2, MiSeq 1, and NextSeq 1 respectively.

Table 5.3.1: Sequences used in the concatenated nuclear gene analysis

Sequence Type	Sequence Number	Total Reads	Mapped Reads	% Mapped Reads	Read Coverage		
					Min	Average	Max
MiSeq 1	2	37310772	158780	0.43	1432	1680	13321
MiSeq 2	10	23964252	144006	0.60	219	343	3554
NextSeq 1	24	157453130	640546	0.41	128	531	24013
NextSeq 2	46	172655068	618225	0.36	76	265	7637

In the table: Max = Maximum read coverage; Min = Minimum read coverage

5.3.4 African *Simulium* blackfly genome size determination

The total contig length of the largest genome assembly was 209,932,758 bp (Table 4.3.4). If the average contig length of this largest assembly is 901, and it has a total of 60 missing proteins, then the remaining size of the missing unassembled contig is the product of the average contig length (901) and the missing protein number (60), a total length of 54,060 bp. By adding this length to the contig length of the largest assembly, the estimated genome size of the African *Simulium* blackflies is 209,986,818 bp. The N75 contig length is 660, implying that 75% of the genome sequence is in contigs larger than or equal to this 660 bp N75 contig length. The minimum contig length of the unassembled missing proteins is the product of this 660 N75 contig length and that for the missing protein number of 60, producing a value of

39,600bp. Adding this value to the existing longest contig length of the largest assembly produces a value of 209,972,358 bp, and this is the minimum estimated size of the African *Simulium* blackflies. The N25 of the best assembly is 2,964 bp and this indicate that 25% of the genome sequence is in contigs larger than or equal to this value. The missing contig length of the upper boundary of the African *Simulium* blackfly genome size becomes the contig corresponding to the product of the missing protein number (60) and the 2,964bp N25 contig length, which produces a value of 177,840 bp. Adding this upper boundary length to the existing contig length of the best assembly gives 210,110,598 bp as the maximum expected genome size of African *Simulium* Blackfly. Hence, the contig lengths of the largest assembly suggest that the genome size of the African *Simulium* blackfly is estimated at 209,986,818 bp, but the true size may lie between 209,972,358 bp and 210,110,598 bp subject to variation in the genome size due to repetitive DNA contribution.

5.3.5 Polymorphism and diversity in the nuclear genome

Polymorphism and diversity were compared in the blackflies across the 3 river basins using the nuclear gene reference panel (see Table 5.3.2). Variant calling is described in section 4.2.17 in this thesis. There were 25 samples in each of the 3 river basins in Ghana, so the total number of mutations was 1 value higher than the corresponding number of polymorphic (segregating) sites for each river basin. The total number of mutations was highest in Daka (33), followed by Black Volta (32), and least in Pru (30) river basins, with a total of 43 mutations across all 3 river basins. The number of polymorphic sites also showed a similar pattern with Daka exhibiting the highest value of 32, followed by 31 in Black Volta, and least in the Pru (29) river basin, thus leading to an overall value of 41 across the entire river basins. The total number of singleton mutations was the same for Pru and Daka (9), while Black Volta had a higher value of 13. The number of haplotypes (25) was equal in all the 3 river basins, leading to corresponding equal haplotype diversities of 1, and the overall number of haplotypes across the 3 river basins was 72 rather than 75 as a result of 3 shared haplotypes (two haplotypes shared between Black Volta and Pru and a third haplotype shared between Black Volta and Daka. There were no haplotypes shared between Pru and Daka. This produced an overall haplotype diversity of 0.999 across the 3 river basins. Although haplotype diversity was equal among all river basins the nucleotide diversities were not, with Pru river basin showing a lower value (0.00103) than Daka river basin (0.00104), and Black Volta river basin showing the least (0.00099). This produced an average nucleotide diversity of 0.00102. The differences in nucleotide diversity are reflected in the average number of nucleotide differences between river basins, with Daka showing the highest ($k=7.440$),

followed by Pru ($k=7.357$), while Black Volta demonstrated the least ($k=7.107$). Thus, the overall average nucleotide difference was 7.332 across the 3 river basins. Collectively, the results demonstrate high mutation, high haplotype diversity, and moderate nucleotide diversity for nuclear genes that are sufficient to infer whether or not there is population structure (following further test by gene flow, genetic diversity and population structure).

Table 5.3.2: Polymorphism and diversity in the nuclear genome of *Simulium* sp. in Ghana

River basin	<i>M</i>	<i>S</i>	<i>Eta (s)</i>	<i>H</i>	<i>Hd</i>	π	<i>k</i>
Pru	30	29	9	25	1	0.00103	7.357
BlackV	32	31	13	25	1	0.00099	7.107
Daka	33	32	9	25	1	0.00104	7.440
All river basin	43	41	13	72	0.999	0.00102	7.332

M = Total number of mutations; *S* = Number of polymorphic (segregating) sites; *Eta (s)* = Total number of singleton mutations; *H* = Number of Haplotypes; *Hd* = Haplotype (gene) diversity; π = Nucleotide diversity; *k* = Average number of nucleotide differences; and BlackV = Black Volta river basin.

5.3.6 Neutrality test in the nuclear genome

In the neutrality test (see Table 5.3.3), Tajima's D test was used to identify whether or not there existed blackfly sequences that did not fit the neutral theory model at equilibrium between mutation and genetic drift. Generally, across all 3 river basins, considered individually and also as a group, Tajima's D were not statistically significant ($p > 0.10$). The lack of statistical significance in all the river basins indicated that the allelic distribution and/or level of variability did not violate the neutrality assumption. Hence, the blackfly sequences were evolving randomly (neutrally).

Zeng *et al.*, (2006) established that there are aspects of Fumio Tajima's data that the Tajima's D did not consider. The simulation of Fu (1997) indicates that Fu's *Fs* generally is more sensitive than Tajima's D towards detecting departures from neutrality due to genetic hitchhiking and population expansion. Consequently, Fu and Li's statistics were explored. The test statistics, Both the Fu and Li's D statistic, and the Fu and Li's F statistic (Table 5.3.3), support the idea of a non-departure from neutrality of the blackfly data due to the lack of statistical significance ($P > 0.10$). Thus, all the neutrality tests were concordant with the

idea of the blackflies evolving with similar distribution and pattern as with a neutrally evolving population.

Table 5.3.3: Neutrality test in the nuclear genome of *Simulium* sp. in Ghana

River Basin	Tajima's D	Tajima's <i>p</i>	FLD*	FLD* Stat Sig	FLF*	FLF* <i>p</i>
Pru	-0.27694	<i>p</i> > 0.10	-0.14830	<i>p</i> > 0.10	-0.22093	<i>p</i> > 0.10
Black Volta	-0.60637	<i>p</i> > 0.10	-0.80697	<i>p</i> > 0.10	-0.87291	<i>p</i> > 0.10
Daka	-0.55968	<i>p</i> > 0.10	0.01950	<i>p</i> > 0.10	-0.18960	<i>p</i> > 0.10
All 3 river basins	-0.54095	<i>p</i> > 0.10	-0.93622	<i>p</i> > 0.10	-0.93785	<i>p</i> > 0.10

The Tajima's D was calculated using the total number of mutations. FLD* = Fu and Li's D* test statistic, FLF* = Fu and Li's F* test statistic, *p* = Statistical significance

5.3.7 Assessment of gene flow and genetic differentiation in the nuclear genome

With the exception of the average within-population diversity, which was the same ($H_s = 1$) in all the 3 river basins; all the statistical tests to estimate gene flow and genetic differentiation (see Table 5.3.4) showed a similar pattern. The trend manifested in populations between Pru and Daka showing the highest values, followed by those between Pru and Black Volta, and the lowest value being realized in the populations between Black Volta and Daka river basins. Thus, in order of decreasing magnitude, the within-population nucleotide diversity was between Pru and Daka ($K_s = 12.3$), Pru and Black Volta ($K_s = 11.9$), and between Black Volta and Daka ($K_s = 11.6$). The average number of nucleotide differences between each of the 2 populations in the compared river basins decreased in order of magnitude for the values between: Pru and Daka ($K_{xy} = 12.6$), Pru and Black Volta ($K_{xy} = 11.9$), and between Black Volta and Daka ($K_{xy} = 11.4$). These results indicate high genetic diversity, but low genetic differentiation in the compared river basins. In order of decreasing magnitude, the pairwise genetic distance was between; Pru and Daka ($\text{GammaSt} = 0.031$), Pru and Black Volta ($\text{GammaSt} = 0.022$), and between Black Volta and Daka ($\text{GammaSt} = 0.014$). Finally, the largest fixation index was observed in the comparison between populations in Pru and Daka ($F_{st} = 0.020$), followed by that between Pru and Black Volta ($F_{st} = 0.002$), and lowest in the comparison between the populations of Black Volta and Daka ($F_{st} = -0.012$). In each of the populations compared using the both the pairwise genetic distance, and the fixation index; the values were approximately zero. Such concordant results demonstrate gene flow between the populations in the river basins being compared. Hence, an

additional support to the notion of the lack of population structure across the 3 river basins, but with the presence of a single onchocerciasis transmission zone in the central transition ecological zone of Ghana.

Table 5.3.4: Gene flow and genetic differentiation per river basin in Ghana

Population 1	Population 2	H_s	K_s	K_{xy}	$\text{Gamma}St$	F_{st}
Pru	BlackV	1	11.91167	11.93920	0.02156	0.00231
Pru	Daka	1	12.31500	12.56640	0.03051	0.02001
BlackV	Daka	1	11.56333	11.42400	0.01435	-0.01220

The table shows various parameters of gene flow and genetic differentiation between population 1 and 2 in the 3 river basins of Pru, Black Volta and Daka. H_s = Average within-population diversity, K_s = within-population nucleotide diversity, k_{xy} = the average number of nucleotide differences between population 1 and population 2, $\text{Gamma}St$ = pairwise genetic distance, and F_{st} = fixation index.

From Table 5.3.5, the majority of the blackflies' genetic variation occurred within populations (99.1%, $p > 0.001$), with the remaining variation occurring among populations (0.9%, $p > 0.001$). This suggests a lack of population structure by river basin. The test for the overall measurement of population differentiation due to genetic structure, overall Fixation Index (F_{ST}), was very small (0.00389) and indicates a panmictic population with random mating and gene flow between river basins. The force of gene flow in the population for a given generation (Nm) influences the tendency of population differentiation as a result of genetic structure (Wright, 1951), was correspondingly large (F_{st} 's $Nm = 64.09$), providing further support for the hypothesis of a lack of population structure between the river basins across the transition zones. Furthermore, the overall pairwise genetic distance, $\text{Gamma}St$ statistic, was small (0.02958) with an Nm value 8.20; values of Nm greater than 1 indicate that gene flow between populations does occur. This further supports the hypothesis of a lack of population structure across the 3 river basins and the inference of a single onchocerciasis transmission zone in the central transition ecological zone of Ghana.

Table 5.3.5: Overall population differentiation test of the nuclear genome in the Pru, Black Volta and Daka river basins (from AMOVA test)

Variation	% Variation	Fst	Fst's Nm	GammaSt	GammaSt's Nm
Among Population	0.85347	0.00389	64.09	0.02958	8.20
Within Population	99.14653				

N = Effective population size, m = proportion of migrants in a population, Nm = the force of gene flow in the population for a given generation, Fst = Fixation Index across the river basins, GammaST = overall pairwise genetic distance

5.3.8 Assessment of genetic relatedness in the nuclear genome

In *a priori* DAPC analysis without assigning any form of river basin population information, the optimum number of PCs retained was 20 and it accounted for over 95% of the total genetic variation in the concatenated nuclear DNA sequences. By examining the lowest BIC and carrying out optimization, the optimum number of clusters was identified as 6 (Figure 5.3.3 A). The highest discriminant function corresponding to a cluster of 6 was found to be 5. With such a discriminating power, the 6 clusters formed 3 main groups, with cluster 1 and 2 being distinct from each other and from the remaining 4 clusters. Thus, the remaining 4 clusters formed the 3rd group with a characteristic of a cline group of populations that had overlapping peripherals. These 3 distinct molecular units may be of operational taxonomic significance in onchocerciasis, hence will be assigned the status of MOTU (MOTU 1-3). In the assignment of individuals to a population cluster, the prior group size was the same as the post group size. This led to a 100% of reassignment of individual blackflies to each of the 6 population clusters (Figure 5.3.3 B). From cluster 1 to 6, the group sizes were respectively 10, 13, 11, 13 and 15. By assigning cluster 1 to MOTU 1, cluster 2 to MOTU 2, and the remaining clusters to MOTU 3; the proportions of MOTU 1-3 thus became 13.3%, 17.3% and 69.3% respectively.

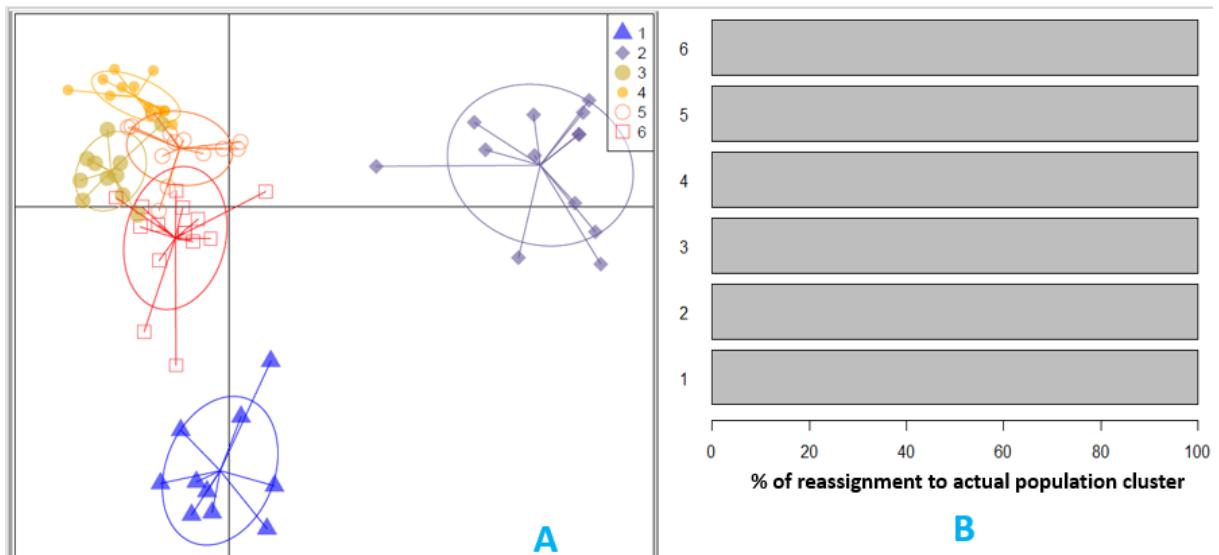


Figure 5.3.3: *A priori* assessment of genetic relatedness in 75 blackflies using the concatenated nuclear gene sequences. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010) with nuclear gene reference panel. Graph A is scatter plot of the DAPC of 75 concatenated nuclear DNA sequences. It was plotted by retaining and choosing 20 PCs, 6 clusters and 5 Discriminant Functions. The ellipses represent the different genetic clusters identified in the DAPC while the dots represent the individual blackflies collected. Graph B shows a bar chart of the percentage of successful reassignment of individual blackflies to the genetic clusters.

The *a posteriori* scatter plot of the DAPC showed that there was overlap of individual membership among the 3 river basins (Figure 5.3.4 B). There were individual blackflies that belonged exclusively to a particular river basin. However, there were other individuals collected from a given river basin but having genetic characteristics related to those of other blackflies from the other river basins.

Consistent with the DAPC, the haplotype network (figure 5.3.4 A) showed that the haplotypes from each river basin were closely related to each other with no obvious structure by river basin. There were 3 shared haplotypes. Two of the haplotypes were each shared by 2 blackfly individuals from Pru and Black Volta river basins. One haplotype was shared by 2 individuals with each individual blackfly belonging to Black Volta and Daka river basins. The remaining haplotypes consisted of singletons. The number of mutational step difference between any two haplotypes closest to each other ranged from 1 to 6.

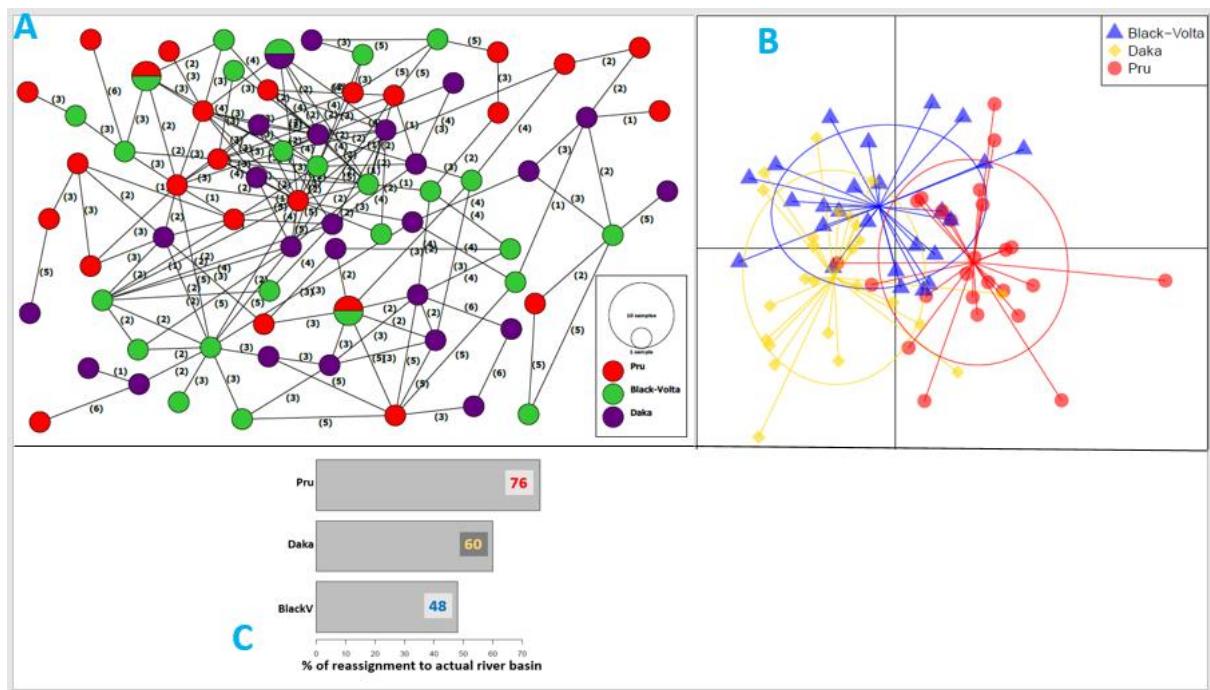


Figure 5.3.4: *A posteriori* assessment of genetic relatedness in 75 blackflies using the nuclear gene reference panel. Graph A is a haplotype network produced by minimum spanning network using Epsilon of 0 in Popart 1.7 (Bandelt *et al.*, 1999). Each circle in the haplotype network corresponds to one haplotype, and the size is proportional to its frequency among the samples. Colours of the circles correspond to river basin sampling locations. Graph B is a scatter plot of the DAPC of the same 75 sequences. The ellipses represent the different river basins from which the sample's sequences were obtained while the dots represent the individual blackflies collected. Graph C shows a bar chart of the percentage of successful reassignment of individual blackflies to the actual river basin from which they were collected. Both graph B and C were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

5.3.9 Evolutionary relationship determination using nuclear genome

The blackflies from Ghana (West Africa) formed a monophyletic group to those from Cameroon (Central Africa) and Malawi (South-Eastern Africa). Thus, blackflies from a single country formed monophyletic group to those from other countries (Figure 5.3.5). All the Central Africa blackflies shared most recent common ancestor with good bootstrap supports ranging from 71 to 96%. The Ghanaian blackflies however had poor to moderate bootstrap supports ranging from 0 to 78%. The blackflies from all the 3 river basins shared most recent common ancestor. The sequences from West Africa separated into 3 main clusters with well

resolved dichotomy clades that had short branch lengths and mostly poor bootstrap supports. The blackflies from South-Eastern Africa had the longest branch length, followed by those from Central Africa, and least in those from West Africa.

The Central Africa clade was divided between the 2 major river basins. There was good bootstrap support, 80%, separating the 2 blackflies from the Nkam river basin. Among the 3 blackflies in the Mbam river basin in Central Africa, 2 had polytomies with a bootstrap support of 71% (i.e. samples EG-7 and EG-8), while the third blackfly (SIM27) diverged earlier in time with strong bootstrap support of 96%. The 2 blackflies in the Nkam river basin clearly diverged from the 3 blackflies in the Mbam river basin. The blackflies from West, Central and South-Eastern Africa diverged from each other and underwent independent speciation.

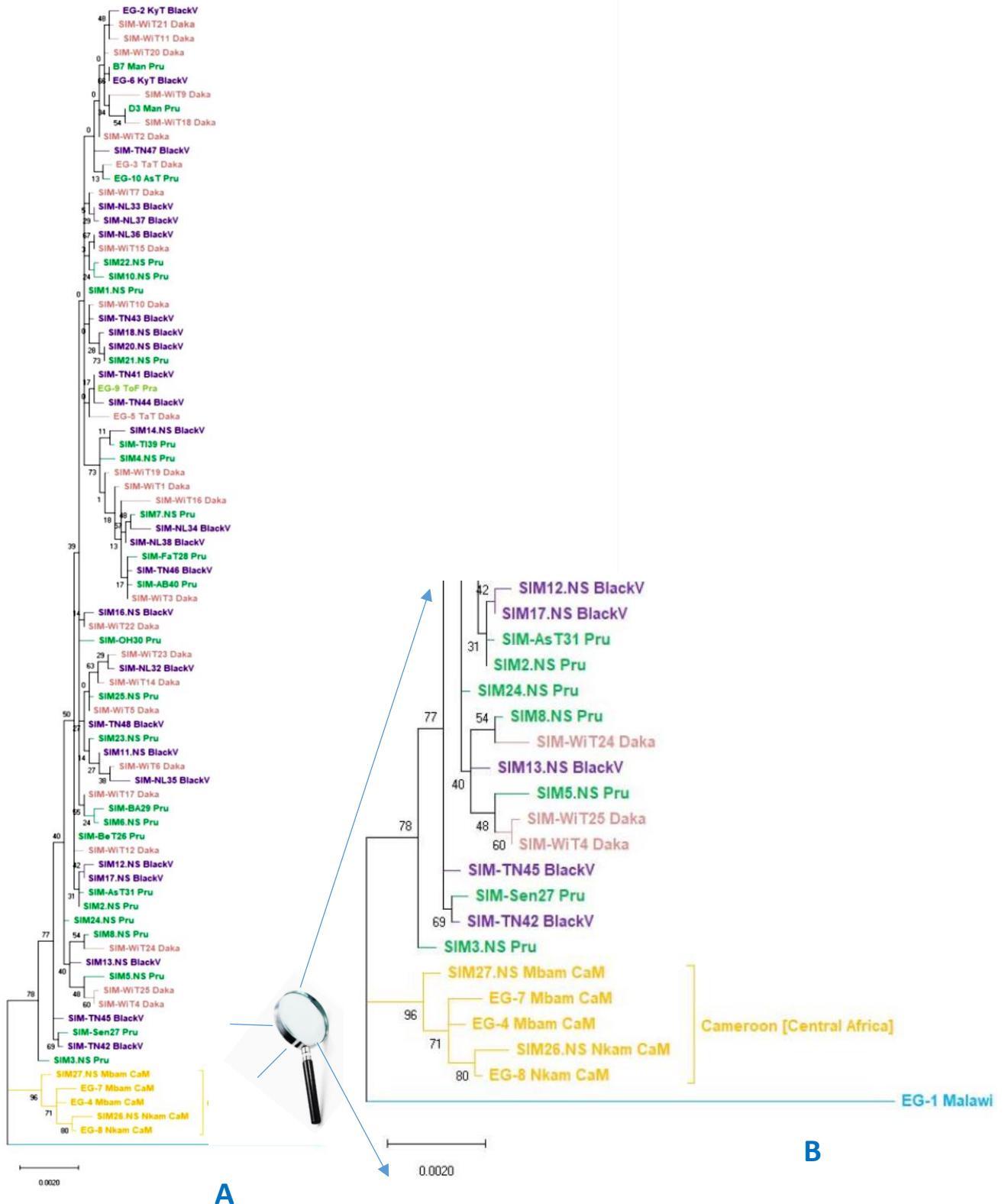


Figure 5.3.5: Rooted Maximum-Likelihood estimate showing the relationship between 81 *Simulium* sp. concatenated nuclear gene sequences in Africa. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Time Reversible model (Nei and Kumar, 2000). The numbers on the branches indicate bootstrap values

(expressed as percentages based on 1000 replicates). The scale bar (0.002) shows the average substitution per site. The tree with the highest log likelihood (-11838.55) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 82 nucleotide sequences. There was a total of 7186 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar *et al.*, 2018). Figure 5.3.5 B is a magnified section of the shown portion of the original tree (Figure 5.3.5 A). The different colours represent sequences from different river basins, i.e. blue – Malawi (South-Eastern Africa), gold - Nkam and Mbam river basins in Cameroon (Central Africa), green – Pru river basin, purple – Black Volta river basin, and hex colour # d99694 – Daka river basin. With the exception of the river basins in Cameroon and Malawi, all the river basins were located in Ghana (West-Africa).

5.4 Discussion

5.4.1 African *Simulium* blackfly genome

The contig length of the largest genome assembly indicated that the African *Simulium* blackflies have a genome size of 209,986,818 Mbp (see section 5.3.4). The bulk of the genome is composed of nuclear elements with the mitochondrial genome consisting of only 16,108 bp to 16,232 bp (see section 4.3.4). Determination of the difference between the sizes of the complete genome and that of the mitochondrial genome showed that the nuclear genome has a size of about 209,970,586 bp to 209,970,710 bp. Consequently, the complete genome or the nuclear genome is therefore estimated to have a size of approximately 210 Mbp. Thus, the size of the mitochondrial genome is negligible in comparison with that of the nuclear genome. The complete genome size of approximately 210 Mbp is consistent with the known insect nuclear genome size of 200 Mbp to 3 Gbp, and this will potentially enable further study of the relationship between the heterogeneous nuclear RNA and mRNA of African *Simulium* blackflies (Lewin, 1975; Spradling and Rubin, 1981). Although the genome size of 200 Mbp to 3 Gbp is the size of the majority, other smaller genome sizes have been identified too. For instance, the *Drosophila* genome has been estimated as 180 Mbp in size, and about one third of this is centric heterochromatin (Adams *et al.*, 2000). The genome size of *Arabidopsis thaliana* was estimated as 125 Mbp whiles that of *Caenorhabditis elegans* was estimated as 97-100.25 Mbp (Science, 1998). Although the genome of African *Simulium* blackflies are among the small genomes, the genome is larger than those of *Drosophila* sp., *Arabidopsis thaliana* and *Caenorhabditis elegans*. The mitochondrial DNA to nuclear DNA ratio ranged from 210 to 524 with an average of 345. The higher mitochondrial DNA copy number than that of the nuclear DNA made the total mitochondrial genome reads per gene to be exponentially greater than that of the nuclear genes and this was partly due to the exponential sequence amplification power of MiSeq and NextSeq acting on unequal numbers of DNA sources within the same reaction mix. With less than a million reads, it was possible to assemble a complete mitochondrial genome, leaving the remining reads for nuclear genome analysis. At the lowest coverage of 76, a complete mapped concatenated nuclear sequence was produced.

Estimation of the genome size of the *Simulium* blackflies in Africa was carried out using the largest *de novo* assembly that had a total genome completeness of 87.905% (see Table 5.3.0). The remaining *de novo* assemblies all had genome completeness less than 87% and were not used in the genome size estimation. The N50 and N90 in particular, combined with the

average contig size, suggest that more Illumina short reads are required, and that long reads like PacBio and Nanopore, or 10X perhaps, are required in order to scaffold short read contigs. A minimum of 5 of such assemblies should be able to give a better estimate of the genome size of the blackflies in Ghana, and Africa as a whole.

5.4.2 Gene evolution events

In addition to the orthologous genes, observation was made of the presence of paralogs that possibly resulted from gene duplication events. While this observation is less known or studied in *Simulium*, it is not uncommon in arthropods and other taxa. The occurrence of paralogs is common in the genomes of insects like *Drosophila* sp. (Malmanche *et al.*, 2003), and in fishes like zebrafish (Force *et al.*, 1999; Tong *et al.*, 2001). The presence of the paralogs in these insects and fishes were mainly attributed to gene duplication events. Inference from paralogs could sometimes offer insight into the interpretation of gene expressions. Moreover, the development of distinct expression patterns demonstrates vital mechanism for the retention of copies of duplicated genes (Zhang, 2003). For instance, retrotransposition generated 2 copies of amidophosphoribosyltransferase in *D. melanogaster* (Malmanche *et al.*, 2003), and these 2 paralogs were expressed at different developmental stages. Divergence of expression occurred in zebrafish for paralogs of *engrailed* (Force *et al.*, 1999) and for 2 other paralog genes (Tong *et al.*, 2001). Some paralogs could be expressed at different developmental stages and yet mediate the same enzymatic reaction (Ono *et al.*, 2006). PCR reaction conditions have been demonstrated to play a major role in the determination of whether or not multiple loci will be co-amplified (Hare *et al.*, 1996; Bagley and Gall, 1998), and these were major issue in the preliminary stages of attempting to amplify the genome of the African *Simulium* blackflies before the sequencing strategy was changed to whole genome amplification. Biological pathways may evolve from duplication of enzymes (Corbin *et al.*, 2004) and neofunctionalization during the moment when one of the duplicated genes acquires an advantage function (Li *et al.*, 2005).

5.4.3 Genetic diversity of *Simulium* blackflies from nuclear genome perspective

There was a very high haplotype diversity ($Hd = 1$) and a comparable level of nucleotide diversity ($\pi = 0.001$) in the *Simulium* blackflies' nuclear genome in all the 3 river basins (Table 5.3.2). If the observed diversity in the nuclear genome follows a similar pattern as that of the mitochondrial genome when compared with the genomes of related species, it would indicate that the haplotype diversity of the nuclear genome on the average is higher than those of related species and its nucleotide diversity is comparatively similar to those of other related

organisms (Hemmerter *et al.*, 2007). The observed diversity trend could be attributed to the numerous singletons that is common with recently acquired mutations that have not had sufficient time to spread across the populations.

The blackflies in all the 3 river basins of Pru, Black Volta and Daka were evolving randomly ($p > 0.10$) as one would expect from a neutrally evolving population (Tajima's $D < 0$, $p > 0.05$). This inference is concordant with that of the whole mitochondrial genome sequencing, which was also not significant, and evolving randomly. This was different from the observation made with *Simulium aureohirtum* at a different geographical location outside Africa that had significantly negative Tajima's D values across 3 studied lineages undergoing demographic expansion (Thaijarern *et al.*, 2014). A large population of blackflies in the central ecological zone of Ghana could have occurred due to the population being old and inhabiting river basins that have good conditions supporting the reproduction and survival of the blackflies. As a transition zone, there is a relatively short distance from the savannah climatic zone to the forest climatic area. During unfavourable ecological conditions in a particular part of the year, blackflies are able to migrate from one climatic area to the other that has better conditions of importance to their survival and reproduction. For instance, during a spatio-temporal study of *Simulium damnosum* in southern Ghana and south-western Togo from 1975 to 1997, it was observed that the percentage of savannah blackflies moving into the forest zone was progressively increasing since 1975 due to deforestation. It was further noted in the study that the savannah blackflies are capable of occurring further south in the dry season (Wilson *et al.*, 2002).

In addition, the stable population at equilibrium could have resulted from shared old history that possibly reduced the likelihood of accumulation of genetic differentiation between populations, and likely of adapting to a wide range of ecological conditions encompassing the adaptation to anthropogenically disturbed habitats including agricultural areas and artificial flowing water (Pramual and Kuvangkadilok, 2009b).

The *Simulium* blackflies in Daka, Pru and Black Volta were genetically homogenous when observed from the viewpoint of river basin. A possible reason could be because the sampled regions constitute an ecological transition zone of both forest and savannah vegetation (Afikorah-Danquah, 1997; Abu *et al.*, 2014) with blackflies being quite evenly distributed across the river basins irrespective of whether they were originally from a forest or savannah area. An alternate possible hypothesis is that the transition zone is ecologically variable and unstable in the sense that ecological conditions change through the year (e.g. change in river

flow) and due to human activity (e.g. vegetation). This results in habitat variability, so that blackflies are very mobile as they move to find suitable breeding sites, food and mates. The result, over time, is that blackflies move around as seasons and other ecological factors change, and populations are homogenised due to this movement. This most likely caused them to have a continuous geographic distribution. This finding is also consistent with studies of other species such as *Simulium siamense* that is genetically homogenous due to occupying both forest and open areas (Pramual *et al.*, 2011). The biallelic inheritance from maternal and paternal origins in such ecological transition zone, being generally distributed in the entire zone, could have contributed to the similarity in the genetic diversity across river basins.

While some blackfly species were observed in other studies to be confined to a given area such as forest or savannah and tend to have high levels of genetic structuring, others had continuous geographic distribution characterized by gene flow between populations (Pramual and Kuvangkadilok, 2012), and such observation of the latter is found in the central ecological transition zone of Ghana.

Migration intentions constitute the first step in the migration process (Macleod, 1996; Van Dalen and Henkens 2008), and the use of stated preferences like intentions have been used by social demographers, psychologists, and geographers as opposed to the preferences strategy mostly used by economist who are mostly interested in the actual movement of people (Borjas 1991; Hatton and Williamson 2004). The migration intention of diverse groups of people in Ghana influenced the actual movement of people into the central ecological zone of country. Within the transition zone, there is more internal movement of people than external due to locally shared socio-cultural factors, and such internal migration causes changes over time in the spatial distribution of people (Greenwood, 1997; Abu, 2014; Afikorah-Dankwah, 1997). The resulting social heterogeneity causes variation in the utilization of resources and ultimately leads to ecological heterogeneity. Such ecological heterogeneity characterized by differences in production regimes, technology, and different farming practices (Afikorah-Danquah, 1997) likely influences the local vegetation and the utilization of streams and rivers. This human induced ecological heterogeneity will likely amplify the natural ecological heterogeneity to give rise to unstable, patchy habitat for blackflies that then promotes mobility in the vectors as well as the human population. If such should occur, it would inevitably cause some form of local stratification in the river basin homogeneity of the blackflies' genetic diversity, and this would obviously not correlate with the river basin that is shared in the zone. The nuclear markers were robust enough to identify 3 distinct molecular

groupings of blackflies with 100% likelihood of being able to reassign individual blackflies to each DAPC cluster (see figure 5.3.3). The members of each of these distinct molecular groups were from all 3 river basins. This local variation in the blackfly genetics could be attributed to the effects of the socio-cultural heterogeneity of the host population in the central ecological transition zone.

5.4.4 African *Simulium* blackflies phylogeny

The West African blackflies were monophyletic to those of Central and South-Eastern Africa. All the African blackflies were also monophyletic to those from Asia and Europe. There were good bootstrap supports separating the Central and South-Eastern Africa blackflies from those in West Africa. There were however very poor bootstrap supports separating the individual blackflies in West Africa.

There was differentiation in the blackfly population in Cameroon, just as there was in the *O. volvulus* worm population studied by colleagues in our lab. In contrast, the poor bootstrap support in the blackflies from Ghana indicated poor differentiation and this was similar to the worm population from the same locations in Ghana. Caution is advised in the inference drawn from these limited data from Central Africa and it will require additional data, from at least 25-30 individuals per sampling site, to validate.

Hard polytomies were observed among the Central Africa blackflies (figure 5.3.5). There was possible radiation in the blackflies in Central Africa. All the blackflies in the Nkam river basin and 2 of those from the Mbam river basin (i.e. samples EG-7 and EG-4) constituted three simultaneous speciation events that resulted from the same common ancestor such that the subsequent daughter species were of equidistant to each other. Thus, the blackflies in the Nkam river basin diverged later in time from those in the Mbam river basin. Unlike soft polytomies resulting from insufficient phylogenetic information, such that lineages diverge at different times without the available data enabling the recognition of lineages that are closer relatives than others (Grafen, 1989; Purvis and Garland, 1993); the current observed hard polytomies constituted three simultaneous speciation events that resulted from the same common ancestor such that the subsequent daughter species were of equidistant to each other. A possible cause is blackflies that had rapidly expanded their range, or they were highly panmictic and underwent peripatric speciation in different localities. Often, complete lineage sorting may not take place between divergence events. Thus, the incomplete lineage sorting led to polytomies within gene trees in the instance that three or more daughter species acquired one or more common genotypes from the parents. The phylogeny of these hard

polytomies cannot be resolved with additional molecular data because as we move from short amplicons, through long amplicons to both whole mitochondrial and nuclear data, the polytomy did not resolve; thus the blackflies are genuinely poorly differentiated (Grafen, 1989; Walsh *et al.*, 1999). This is a common feature of cichlid fish, where the fish speciated quickly after their home lakes formed in Africa. The resulting daughter species were represented by several phylogenetic polytomies (Seehausen, 2006; Near *et al.*, 2008). However, unlike the cichlid fish, there is no evidence from morphology or ecology that speciation has occurred in the central ecological transition zone of Ghana. Perhaps, the only evidence for differentiation could be cytogenetics, but it is not known to what extent, if at all, the cytogenetic differences may constitute reproductive barriers. The results suggest that the West Africa blackflies are relatively recent and diverged from those from Central Africa, which also diverged from those in South-Eastern Africa. However, additional data is required to validate if it is indeed the case.

The low bootstrap support was not because of insufficient data because the nuclear genome markers were robust enough to identify 3 distinct molecular groups of blackflies in the central ecological transition zone of Ghana, and the members of these groupings were able to be reassigned to the clusters of each group with 100% likelihood. One of the 3 groups may be genetically heterogeneous.

5.4.5 Inferring *Simulium* population structure using nuclear genome

The majority of the blackflies' genetic variation occurred within populations (99.1%, $P > 0.001$), with the remaining variation occurring among populations (0.9%, $P > 0.001$). This confirmed the lack of geographic population structure by river basin as inferred from the short and long amplicon sequencing and whole mitochondrial genome analyses. The test for the overall measurement of population differentiation due to genetic structure, overall Fixation Index (F_{ST}), was very small (0.004) and approximately equal to zero. This small fixation index infers a panmictic population with random mating and gene flow between river basins. The force of gene flow in the population for a given generation (Nm), that influences the tendency of population differentiation because of genetic structure, was very large (F_{ST} 's $Nm = 64.09$). Both the negligible fixation index and the very large Nm that is greater than 1 (Wright, 1951) support the initial inference of a lack of geographic population structure in the river basins across the central transition ecological zone of Ghana. The overall pairwise genetic distance, GammaSt statistic, was equally small (0.03) with an Nm value greater than the recommended value of 1 ($Nm = 8.2$) that is needed as the minimum value for gene flow

between population to occur. This further supported the notion of a lack of population structure across the 3 river basins, thus adding support to the inference of a single onchocerciasis transmission zone in the central ecological transition zone of Ghana. The nuclear data is consistent with the mitochondrial data, and the same hypothesis pertains to the nuclear as to the mitochondria.

5.4.6 Implications of the *Simulium* blackfly nuclear genome diversity, divergence and population structure on transmission zone and elimination intervention

An assessment of the current river basin definition of transmission zone indicates that if ivermectin or other onchocerciasis elimination interventions stop in communities of any given river basin, there is the possibility of infected blackflies traveling from other river basins to reinvade the communities that have stopped ivermectin treatment. This calls for a comprehensive assessment of the migration, diversity, divergence and population structure of *Simulium* blackflies in the onchocerciasis endemic countries. This can lead to a better understanding and definition of onchocerciasis transmission zone, delineation of a reliable transmission zone, greater accuracy in the estimation of timelines to stop MDA using ivermectin, and a greater chance of guarantee to sustain the progress made in many years of onchocerciasis control and elimination activities.

5.4.7 Decision on study objectives

1. The genetic diversity of the *Simulium* blackflies in the study area was characterized by using 7,220 bp concatenated nuclear genes.
2. The onchocerciasis transmission zone status in the central ecological transition zone of Ghana was determined by using 7,220 bp concatenated nuclear genes
3. The study objectives were achieved

5.4.8 Decision on study hypothesis

1. The concatenated nuclear genes of *Simulium* blackflies from Ghana enabled the determination of the genetic diversity of the blackflies
2. The genetic diversity of the *Simulium* blackflies in Pru, Black Volta and Daka river basins were characterized as one with a high haplotype diversity of 1 and a nucleotide diversity of 0.001.
3. The null hypothesis 1 was accepted and the alternative hypothesis 1 was rejected

4. The concatenated nuclear genes of the *Simulium* blackflies enabled the determination of the onchocerciasis transmission zone status of the central ecological transition zone of Ghana
5. By using the concatenated nuclear genes, the onchocerciasis transmission zone status of the central ecological transition zone of Ghana was determined as a single onchocerciasis transmission zone characterized by free movement, interbreeding and gene flow between the blackflies in the studied 3 river basins
6. The null hypothesis 2 was accepted and the alternative hypothesis 2 was rejected

5.5 Conclusion

A 7,220 bp concatenated reference sequence was built from nuclear genome barcoding genes. These consisted of 18S-rRNA (2,856 bp), ITS1 (103 bp), 5.8S-rRNA (123 bp), IGS (21 bp), 2S-rRNA (30 bp), ITS2 (349 bp), and 28S-rRNA (3,738 bp). A total of 82 individual blackfly nuclear sequences were involved in the genetic analysis.

The largest assembled genome was 87.9% overall complete, consisting of 79.4% and 96.4% of the complete and partial ultra-conserved core eukaryotic genes respectively. The genome size of the African *Simulium* blackfly was estimated at approximately 210 Mbp. The complete mitochondrial genomes of *Simulium* blackflies from Africa were observed to be circular molecules of 16108 to 16232 nucleotides, and each of them had 2 rRNA genes, 22 tRNA genes, 13 protein coding genes and a non-coding control region (D-LOOP). It was like the mitochondrial genomes of other *Simulium* blackflies from Asia and Europe in terms of gene order and orientation.

The West African blackflies were monophyletic to those of Central and South-Eastern Africa. All the African blackflies were also monophyletic to those from Asia and Europe. There were good bootstrap supports separating the Central and South-Eastern Africa blackflies from those in West Africa. There were however very poor bootstrap supports separating the individual blackflies in West Africa (Ghana). *Simulium* blackflies in Africa are possibly undergoing radiation similar to those observed in cichlids.

There was a very high haplotype diversity ($Hd = 1$) but moderate nucleotide diversity ($\pi = 0.001$) in the *Simulium* blackflies' nuclear genome in all the 3 river basins. The blackflies in all the 3 river basins of Pru, Black Volta and Daka were evolving randomly ($p > 0.10$) as one would expect from a neutrally evolving population, and there was a high frequency of rare alleles (Tajima's $D < 0$). The nuclear markers were robust enough to identify 3 distinct molecular groupings of blackflies with 100% likelihood of being able to reassign individual blackflies to each DAPC cluster. The members of each of these distinct molecular groups were from all 3 river basins. The diversity, divergence and population structure data of the nuclear genome is concordant with those of the amplicon sequencing, amplicon resequencing and whole mitochondrial genome data. The additional data from both matrilineal and patrilineal lineages confirmed the lack of population structure by geographical place of origin, and the presence of continuous gene flow between blackflies in all 3 river basins in the dry season due to the interbreeding of the blackflies in the entire transition zone. The entire central ecological transition zone of Ghana constitutes a single onchocerciasis transmission

zone and this finding does not agree with the current transmission zone definition based on *Simulium* vector breeding sites (WHO, 2016). The presence of a single onchocerciasis transmission zone in the entire central ecological transition zone of Ghana implies that elimination interventions, such as community directed treatment with ivermectin, cannot stop in communities of any particular river basin without the risk of infected blackflies from the other river basins traveling to re-infect such localities. A new definition for onchocerciasis transmission zone based on the molecular genetics of the transmission determinants is recommended. Similar study as this one is needed to be carried out in the remaining onchocerciasis endemic countries to characterize the genetic diversity of the blackflies and delineate transmission zones.

Chapter 6: General Discussion

6.1 Overview

The purpose of this chapter is to summarise the initial objectives of the thesis and the principal findings, draw these investigations together in a broader discussion of the implications of the molecular point of view of the data results, discuss the technical aspect of the work, and to propose future directions.

6.2 Summary of chapters

6.2.1 Chapter 1: General introduction

Chapter one reviewed the literature on onchocerciasis: the disease, its treatment, transmission, control and elimination; life cycle of the parasites and vectors of onchocerciasis, and the distribution of the *Simulium* blackflies. It also examined vector complexes, and parasite-vector complexes. The chapter finally looked at the need for the integration of knowledge from morphology, cytology and molecular genetics towards meeting the broader aims of characterizing the genetic diversity of *Simulium* blackflies from Africa and using the available results to define the transmission zone of the disease.

The primary aim of this thesis was to characterise the genetic diversity of *Simulium* blackflies in the central ecological transition zone of Ghana, and to use the genetic data to define onchocerciasis transmission zones.

Onchocerciasis or “river blindness” is caused by the parasitic nematode *O. volvulus* and transmitted to humans through the repeated bites of the infected blackfly of the genus *Simulium*. Major consequences of its infection include severe itching, visual impairment, blindness, epilepsy, skin lesions, decreased life expectancy among the blind, social ostracism and the abandonment of fertile farmlands.

Ivermectin is still effective in killing significant numbers of the microfilaria and clearing them from the skin to make them unavailable to be picked up by an adult female blackfly during a blood meal. Four countries in the Americas (Colombia, Ecuador, Guatemala and Mexico) have successfully used the wonder drug, ivermectin, to eliminate transmission of the parasite. In the Bolivarian Republic of Venezuela (part of the Americas) and parts of Africa (Sudan and Uganda), mass drug administration with ivermectin has stopped due to elimination of transmission in at least one transmission location (Cupp *et al.*, 2019).

While some blackflies like *S. ruficorne* have a broad geographical distribution (e.g. from Africa, through Europe and the Arabian Peninsula to the Mascarene Islands) and can be found in arid areas and various locations characterized by extreme conditions like pollution, high water temperatures, and very low rates of water flow; other blackflies have very limited distribution. In the southern American countries like Mexico, *S. ochraceum* is the predominant vector of onchocerciasis. In the United States and Canada, the predominant vectors are *S. meridionale*, *S. slossonae*, *S. articum* and *S. vittatum*. Although they are usually irritating pest of poultry and cattle, they have been found to also attack humans (Underhill, 1944; DeFoliart and Rao, 1965; Lawrence, 2008). In the Australasia region, species like *S. dycei* and *S. mackerrasorum*, *S. melatum*, *S. standfasti*, and *S. lawnhillense* have been found (Bedo, 1975; Colbo, 1976). There is large number of named *Simulium* blackflies but not all of them appear to be distinct on the basis of reproduction. For instance, the species names *S. geniculare* Shiraki from Taiwan, *S. philippinense* Delfinado from the Philippines, and *S. tuaranense* Smart and Clifford located in the Sabah, Malaysia are all synonyms of *S. aureohirtum* (Adler and Crosskey, 2014).

In Africa, different *S. vectors* have been identified. These include the *S. sanctipauli* subcomplex, the *S. damnosum* complex, *S. yahense*, and *S. squamosum*. The concept of sibling species of blackflies show the existence of 4 main blackflies in Ghana: *S. sanctipauli*, *S. leonense*, *S. konkourense* and *S. soubrense* (Boakye *et al.*, 1993a). There needs to be an assessment of the various species of blackflies in Africa to determine the status of gene flow. This is necessary to accurately determine the localities of their parasite transmission, and consequently, of their transmission zones. In parts of Africa where *O. volvulus* resistance to ivermectin is suspected and suboptimal responses confirmed (Eng *et al.*, 2006; Bourguinat *et al.*, 2007; Churcher *et al.*, 2009; Gyan, 2013; Frempong *et al.*, 2016), there needs to be the monitoring of *Simulium* blackflies to ascertain their defined regions of migration and parasite transmission. The problem of recrudescence has two possible causes, both of which population genetics can elucidate. First, it may occur because treatment is stopped when there is in fact still local transmission (i.e. transmission has not actually been interrupted). If so, then the parasites that “reappear” will be from the same transmission zone as were present at the pre-treatment period. Second, it may occur due to invasion, in which case the invading parasites may be genetically different to those present prior to treatment. The problem of resistance is the risk it will spread from the population in which it has been selected to. Monitoring all these will enable a pragmatic decision to decide on where to stop ivermectin mass drug administration and where else to continue, following verification by the WHO of

transmission interruption. Without this measure, the cessation of ivermectin treatment in a country could cause infected blackflies from neighbouring countries to re-invade the already treated countries and render the decades success of control programmes fruitless. To test this concept and help disease elimination programmes make informed decisions on the end point for ivermectin cessation, this project aimed to characterize the genetic diversity of *Simulium* blackflies in the transition zone of central Ghana and define the nature of transmission zone at the collection points and surrounding communities.

Following the analysis of data from the transition zone of Ghana, the structure that was observed was not consistent with perennial, more or less permanent breeding sites, but it was consistent with the hypothesis that breeding sites are ephemeral and move around. In the transition zone, perhaps its “transitional” nature of being between forest and savannah, means that breeding sites do move around over large geographical distances, with the result that population structure is over the same sort of scale. Furthermore, the movement of breeding sites predisposes blackflies to population admixture, resulting in the DAPC and phylogenetic patterns observed. Hence, these are consistent with mobile, admixed populations in a single onchocerciasis transmission zone

6.2.2 Chapter two: Diversity, divergence and transmission status estimation from short amplicon sequencing of African *Simulium* blackflies

Chapter two initially used 1,989 available nucleotide sequence data from public database to estimate the degree of genetic diversity required to infer and differentiate a high genetic diversity from a low diversity; determined the genetic diversity expected in African *Simulium* blackflies, compared the diversity with those of other species; used amplicon sequencing to characterized the genetic diversity of African *Simulium* blackfly. The chapter further used 122 individual blackflies to sequence barcoding genes to characterize the genetic diversity of blackflies from Africa and utilized the available results to propose a type of onchocerciasis transmission zone in the central ecological transition zone of Ghana.

Generally, across the 3 river basins, the nucleotide diversity is high in the nuclear ITS2 gene (0.03), moderate in the mitochondrial genes of CO1 and ND4 (0.01), and low in IGS (0.003). The species level phylogenetic analysis of *Simulium* shows extremely low divergence per informative site and correspondingly poor bootstrap support using the nuclear encoded IGS amplicon (Fig 2.3.2). This is despite the strong inference of population structure when these

data are used to construct a haplotype network based on country of origin (Fig 2.3.3) rather than on morphological species definition (Fig 2.3.2).

Using the sequence data from the transition zone of Ghana, there were 3 groups, and that one of these may be genetically heterogeneous. The data showed (a) similar levels of diversity as the one in the NCBI database for *Simulium*, (b) suggested *Simulium* in Africa are either similar to or more genetically diverse than other taxa but that, (c) nearly all the diversity is within rather than between groups. In particular, the analyses pointed to what appeared to be a strong relationship between diversity and geographical distance over large distances but very low differentiation within countries, and most likely perhaps a complex of either very closely related species, or populations within species, that are poorly differentiated and between whom gene flow may continue. The nucleotide diversity of the African *Simulium* blackflies, especially of those in the transition zone of Ghana, are relatively larger than those of related organisms.

The population structure analysis of the Ghanaian transition zone also revealed high levels of nucleotide diversity with the majority of the genetic diversity occurring within, rather than between, groups. In this case, “group” was river basin. These data led to a conclusion that there was no genetic subdivision or population structure between river basins. Consequently, blackflies could move and interbreed freely between river basins. The data further showed that the *Simulium* genetics do not correlate with river basin type of transmission as already assumed by onchocerciasis control programmes. The presence of gene flow between blackflies in the 3 river basins, coupled with the lack of population structure by river basin suggest that the whole central ecological transition area of Ghana may constitute a single onchocerciasis transmission zone. More data, possibly long-range sequences, were needed to verify this inference and to find out if the observed phylogenetic polytomies resulted from phenomenon like adaptive radiation of the blackflies or from inadequate genetic data, hence the long-range amplification work of the chapter 3.

6.2.3 Chapter three: Confirmation of the diversity, divergence and transmission boundaries of *Simulium* blackflies from the Ghana ecological transition zone

The possibility that the failure to detect population structure with short amplicons might be the result of not having sampled sufficient genetic diversity to detect informative sites. The remedy to this problem was to sequence more of the mitochondrial genome. Thus, long amplicons, including tRNA genes that are likely to evolve faster than barcode genes could be the answer. Chapter 3 reviewed the literature of mitochondrial ND2, CO1 and 4 tRNAs

located between tRNA-Gln and COX2 within the *Simulium* mitochondrial genome, their role in determining genetic diversity, and their contribution in providing a definition for onchocerciasis transmission zone. Using Next Generation Sequencing technique, the targeted region of the mitochondrial genome that was reviewed in the literature was sequenced from each of the 90 individual blackflies. Molecular phylogenetic analysis of the long-range nucleotide sequence data was used to further explore the genetic diversity of the blackflies and to decide on the proposed transmission zone from Chapter two.

Similar results were obtained from this long-range data as in the short amplicon sequence data in chapter 1. It then appeared that the blackflies in Africa have a large and old population that might be undergoing adaptive radiation. It was then necessary to follow this work up with more data generation using complete mitochondrial genomes. This led to the whole mitochondrial genome sequencing work of chapter 4.

6.2.4 Chapter four: Validation of the confirmed genetic diversity, divergence and transmission ranges of *Simulium* blackflies from Ghana study foci

In Chapter 4, a complete shift was made in the approach of normally using amplicon sequencing by utilizing NGS technology to sequence and assemble complete mitochondrial genomes of 82 individual blackflies from onchocerciasis endemic localities in West, Central and South-eastern Africa. The increased amount of genetic data was used to decide on the genetic diversity of the blackflies and the defined transmission zone that was based on both short and long amplicon sequence data.

There is no significant population structure differentiation of the *Simulium* blackflies sampled from central ecological transition zone of Ghana

- The nucleotide diversity in the whole mitogenome across all 3 river basins (0.01) is the same as that of the short amplicon mitochondrial genes (CO1 and ND4)
- The population structures of the parasites and their likely vectors are, as expected, concordant
- The entire central Ghanaian ecological “transition zone”, which is largely hyperendemic for onchocerciasis, constitutes a single transmission zone
- Onchocerciasis control and elimination programs cannot cease transmission elimination efforts in communities in any particular river basin without risking

reinvasion of infection by blackflies migrating from non-treated communities in other river basins

- These inferences were limited to complete mitochondrial genomes that only reflect the genetics of blackflies maternally inherited. It was imperative to use relatively larger nuclear dataset to determine if the blackfly genetics from a mitochondrial source of data will be supported by that of a nuclear source. This led to the use of the concatenated nuclear genes data of chapter five

6.2.5 Chapter five: Whole nuclear genome

In Chapter 5, commonly used nuclear barcoding genes were concatenated to form a reference nuclear gene sequence. After individually mapping nuclear reads from 82 individual flies to this nuclear reference, the consensus sequences were used to carry out molecular phylogenetic analyses similar to those performed with the whole mitochondrial genome of Chapter 4. The genetic diversity from the nuclear analysis was compared to those of the whole mitochondrial genome and the remaining amplicon data. A decision on the transmission zone status of the blackflies was made based on the concatenated nuclear data from all 82 individual blackflies. There was a very high haplotype diversity ($Hd = 1$) and relatively moderate nucleotide diversity ($\pi = 0.001$) in the *Simulium* blackflies' nuclear genome in all the 3 river basins. The relatively lower diversity in the nuclear genome is expected due to the high mutation rate associated with mitochondrial genes. The high haplotype diversity could be because of the presence of numerous singletons that is common with recently acquired mutations that have not had sufficient time to spread across the populations. The nucleotide diversity of the *Simulium* blackflies was similar to that of mosquitoes found in Cameroon ($\pi = 0.00–0.002$) (Kengne *et al.*, 2007).

The general expectation is that mitochondrial population structure would be congruent with that of nuclear. The question of mitochondrial versus nuclear population structure may be particularly important for *Simulium* because cytotoxic polymorphism may create barriers to hybridisation between cytotypes that do not apply to mitochondria, leading to discordance between mitochondrial and nuclear population genetics. The implication of the concordance between the nuclear and mitochondrial population structure observed here is that the chromosomal polymorphism on which cytotoxicity is based is not correlated with restrictions on mating between cytotypes (i.e. cytotypes are not species), or even subspecies, by any genetic criterion. They are simply chromosomal variants that show some presumably unstable regional patterning. The inference of the vector population structure and

transmission from the nuclear data was similar to that of the mitochondrial genome, long-range amplicon sequencing and the short amplicon sequencing. The lack of population structure by river basin, and the presence of gene flow in blackflies from all 3 major river basins of Pru, Black Volta and Daka became much clearer by drawing inference from this nuclear genome data. From all the mitochondrial and nuclear data sources, a single onchocerciasis transmission zone in the central ecological transition zone of Ghana is therefore confirmed and concluded.

6.3 Further discussion

The general consensus from the complete mitochondrial genome and nuclear genome data, which are concordant, suggest that the *Simulium* blackflies from Africa show high frequency of rare alleles and with random evolution such that the allelic distribution and/or level of variability do not violate the neutrality assumption. Comparatively, this observation was similar to that of the *Entomyzon cyanotis* birds, *Chlamydosaurus kingii* frilled-necked lizards and the *Variabilichromis moori* fishes (Tajima's $p > 0.05$). Undetected by the Tajima's test, the Fu's test, which is more sensitive to evolutionary processes like genetic hitchhiking and population expansion (Fu, 1997; Zeng *et al.*, 2006); showed that the *Entomyzon cyanotis* were not evolving randomly. This could be due to alleles in the birds either undergoing some form of natural selection or because they are near another gene that is undergoing a selective sweep such that both the alleles and those undergoing the selective sweep are on the same DNA chain. The *Variabilichromis moori* result was similar to that of the *Simulium* blackflies from Africa due to similarity in the statistical significance (both Tajima's D and Fu & Li's F had $p > 0.05$), which was not statistically significant and had concordance in evolving randomly.

Having established that the nucleotide diversity in the *Simulium* blackflies' mitochondrial genome are fairly similar with an approximated value of 0.01 when compared gene by gene, except for 16S with lower value observed in the short amplicon data; it therefore becomes reasonable to compare a given gene in *Simulium* blackfly with similar genes of other related organisms. By comparing the nucleotide diversity of the blackflies with those of other organisms at the population level of the species (which indirectly becomes a comparison at the population level within the same organism, but at the general level across species); the diversity of the blackflies ($\pi = 0.008$) from Ghana in this study was higher than those of *Latimeria chalumnae* ($\pi = 0.00026$), *Latimeria menadoensis* ($\pi = 0.00012$), *Entomyzon cyanotis* population 1 and 2 ($\pi = 0$ and 0.00454 respectively), *Chelodina parkeri* ($\pi = 0$),

Chlamydosaurus kingii population 1 and 2 ($\pi = 0.00179$ and 0.00161 respectively), *Onchocerca volvulus* ($\pi = 0.00136$), *Variabilichromis moori* population 1, 2 and 3 ($\pi = 0.00057$, 0.00191 and 0.00156 respectively), *Caenorhabditis elegans* ($\pi = 0.00208$), and *Oreochromis niloticus* from Myanmar ($\pi = 0.00644$).

Interestingly, the nucleotide diversity of the blackflies from Ghana in this study ($\pi = 0.008$) was similar to those in the NCBI database that were also collected from Ghana even in different collection spatial time by different workers ($\pi = 0.00925$); as well as from other vastly separated geographical distant locations like Mali ($\pi = 0.00648$) and Liberia ($\pi = 0.00611$). This could suggest that either: (1) the blackfly populations in Ghana from this study are similar in size to those of other study from different spatial times and different geographical locations such that the populations have remained fairly stable over relatively long periods of time without significantly increase in population size; or (2) the blackflies collected from Ghana in this study and in the studies of other workers (as deposited in the NCBI nucleotide database) from the different times and different geographical locations have similar mutational rates with constant population size. The nucleotide diversity of the blackflies from Ghana was also similar to those of *Crocodylus niloticus* and *C. porosus* ($\pi = 0.01141$ and 0.01392 respectively), *Caenorhabditis remanei* ($\pi = 0.01529$) and *Onchocerca ochengi* (0.00763). Overall, it can be deduced that there is a great possibility that either the blackflies in Ghana from this study have similar evolutionary process acting on them as compared with those from other studies in different times or different geographical; or they are responding similarly to those evolutionary forces.

Nevertheless, the nucleotide diversity of the blackflies from Ghana was comparatively lower than those of *Chelodina oblonga* ($\pi = 0.03912$), *Oreochromis niloticus* from the Democratic Republic of Congo (DRC) ($\pi = 0.05475$); and also, lower than those of blackflies in the NCBI nucleotide database collected from areas whose exact locations were not disclosed by the other workers ($\pi = 0.05101 - 0.12906$). Such unusually high nucleotide diversity of blackflies in the NCBI database (higher than those from Ghana) that were collected from the undisclosed geographical locations, possibly different from Ghana, are similar to those of other organisms containing multiple species per organism. Given the difficulties of *Simulium* taxonomy, species misidentification is very likely and so this conclusion is plausible.

These discrepancy in the observed higher nucleotide diversity of some of the blackflies in the NCBI database compared with those in this study could be attributed to a number of reasons that may include: (1) some of the blackflies in the public dataset that were collected from the

undisclosed geographical locations might have been mis-identified and could belong to different species since their values are similar to those of other organisms containing multiple species (Tables 2.3.3A and B); (2) the blackflies in this study collected from Ghana could consist of single species diverging from each other across the 3 major river basins of the central ecological transition zone to the extent that there could be some form of population structure but with sufficient gene flow between the diverging groups; and (3) the blackflies from the Ghana study are genuinely less diverse than those from the undisclosed geographical locations. The second possible explanation, ‘2’, could be justified with the observation from our *a priori* DAPC results that consistently identified 3 distinct molecular groups of blackflies using all sequenced genes (short amplicons, long range, complete mitochondrial genomes, and nuclear genome). However, the first explanation, ‘1’, cannot be ruled out due to the greater similarity of such observation with those of other organisms containing multiple species. The latter explanation, ‘3’, seems possible but less likely to occur because the nucleotide diversity of the blackflies collected in this study from Ghana were similar to those in the NCBI nucleotide database that were also collected from Ghana and were well labelled by other workers. It is possible the variation in results of this study with those of the other workers who did not indicate the sources of the sample location either resulted in the use of sequences from multiple species, or that the species names were mis-identified in the studies of the other workers. Additional data may be needed to explain this speciation observation and discrepancies, and this could demystify some of the discordance in observations of the field between morpho-taxonomy, cyto-taxonomy and molecular taxonomy of African *Simulium* blackflies.

From the blackfly study in West Africa, the lack of population structure by river basin was characterized by almost all the variation occurring within population with very little occurring among populations (Table 4.3.9). Comparatively, there was moderate population structure between the *Simulium* blackflies collected from Mali, Liberia and Ghana in other studies (NCBI data) as shown in Table 2.3.4. In this, the proportions of the genetic variation occurring within population ($VWP = 43.8$) was similar to among population ($VAP = 56.2$). The vast geographical distances between such blackfly locations in their respective different countries could have resulted in mutations being acquired locally and there has not been sufficient time to enable them to distribute over such vast geographical distances. The *Simulium* blackflies from this study however showed different population structure to those of the other organism in this study because the remaining organisms showed very high significant population structure such that the greatest proportion of variation occurred among

population ($VAP = 77.8 - 100\%$) with very little occurring within population ($VWP = 0 - 22.2\%$). Both *Oreochromis* sp. and *Latimeria* sp. had fixed population with all the proportion of variation occurring among population and none occurring within population ($VWP = 0$).

Implication of this moderate isolation by distance of the blackfly's genetics observed between flies in Ghana, Mali and Liberia suggest that there is possibility of gene flow between the blackflies, maybe not directly in a straight flight pattern, but possibly by successive smaller migration movements. Moreover, this could be the result of the accrual of local genetic variation under geographically limited dispersal, leading to differences between population at opposite ends of a more or less continuous group of populations, if for instance the dispersal distances are relatively low (Wright, 1978). Further study is needed to determine if there are long range blackflies in Ghana, Mali and Liberia that could be travelling in longer distance to localities in between the 3 countries such that the gene flow eventually occur between the 3 countries through the isolation by distance model. If such should occur, it will become very difficult for any of these countries or those between them, as well as the countries closer to them, to cease ivermectin mass drug administration in the event that any of such countries observes onchocerciasis transmission interruption and gains WHO verification of being free from onchocerciasis transmission. This is because blackflies from untreated geographical locations in any of these countries could potentially carry parasites in the course of time to the other areas that observes transmission interruption. Within 3 years, there is the possibility that with such re-invasion and re-establishment of onchocerciasis transmission in a formerly treated geographical area, the 3 years may not be sufficient for the parasites to establish in significant numbers of the population and produce sufficient microfilaria to be detected at the population level. A major reason for this possible occurrence is that unlike the onchocerciasis-parasite-vector situation in the Americas that is attributed to foci transmission with vectors that are less competent in transmitting *O. volvulus* (exception with *Simulium exiguum* in Ecuador); the situation in Africa is characterised by widespread distribution of parasites and vectors, and with the situation compounded by blackflies that have long-range flight and possess very efficient transmission competency (Yaméogo *et al.*, 2004). This could also be partly due to the average life span of the microfilaria being approximately 2 years, but with individual ages ranging from 6 months to 3 years (Duke, 1968; Eberhard, 1986; Collins *et al.*, 1995).

6.4 Summary of additions to the body of knowledge

6.4.1 Chapter two

- With the exception of 16S-rRNA gene whose sequence had very little segregating sites, the nucleotide diversity of any of the mitochondrial genes can represent those of the other genes in the same mitochondrial genome because they are not statistically different
- With the exception of the nuclear ITS2 gene of *Simulium* blackflies in Africa whose sequence had little segregating sites, the nucleotide diversities of the nuclear genome are statistically different from those of the mitochondrial genome
- The mitochondrial genome of the *Simulium* blackflies from Africa evolve at a different rate than the nuclear genome
- *Simulium* population genetic inference by the use of any individual mitochondrial gene produces similar results, except for the level of confidence, as any other mitochondrial genomic region, irrespective of whether they are individual gene, concatenated or of the complete genome
- Although the short amplicon sequences of *Simulium* blackflies from Africa were generally observed to contain lesser diversity, they produced similar inference as those of the whole mitochondrial genome combined with concatenated nuclear genes
- The greatest genetic variation of *Simulium* blackflies from Africa is found in the control region of the mitochondrial genome, followed by the tRNA genes and least in the coding sequences
- The nucleotide diversity of CO1 (set 2), $\pi = 0.12906$, of the *Simulium* blackflies from Africa was greater than those of all the other organisms examined in this study. This indicate that the poor discrimination between blackflies that is characterized by the usually low bootstrap support values of phylogenetic analysis realized by different authors, contrary to the general view of the field, is not due to the lack of sufficient genetic diversity of blackflies in Africa. This can be attributed to an old and large panmictic population of blackflies
- There is a moderate population structure between *Simulium* blackflies in Ghana, Mali and Liberia, and this is due to moderate isolation by distance. Although this was not statistically significant, it is possible that either there were insufficient individual blackflies used in the analysis or moderate gene flow occurs over such large geographical distance between blackflies in Ghana, Mali and Liberia, and such a gene flow might be better explained by an isolation by distance model.

- The genetics of the *Simulium* blackflies in Africa correlate with geographical place of origin of the blackflies, but it is not concordant with the current species definition based on morpho-taxonomy.
- Despite the fact that the population genetics of the blackflies enable the detection of historical gene flow between geographical locations, and allowing geographical boundaries to be drawn; it does not allow any conclusions to be drawn on whether the gene flow is between members of a single species or whether it might be gene flow between sub-species or recently diverged species that are not yet completely isolated genetically
- Due to the earliest recognized Simuliid fossils dating to the late Jurassic times and the fossil record of related families suggesting blackflies to have originated much earlier than this date inferred from the earliest fossils (Kalugina, 1991; Currie and Grimaldi, 2000), it is most likely that Simuliidae had Pangean origin. The current data provide evidence that blackflies have been a feature of African habitats for a very long time and certainly longer than humans. It further shows that current African species have evolved locally such that they are not of recent introduction and have adapted to feed on new species like humans or cattle.

6.4.2 Chapter three

- A total of 90 individual blackfly sequences with each sequence made up of tRNA-Met, ND2, tRNA-Trp, tRNA-Tyr, tRNA-Cys, and COX1 have been produced. Once submitted to public database, it will become easily available for others to use as secondary data for their work
- Compared with the diversity in related species, *Simulium* blackflies from Africa have high nucleotide diversity and high haplotype diversity. This could be because of a large population size, suggesting an old population that has existed in the river basin for long evolutionary time
- The blackflies at the moment were found to be evolving randomly
- There was significant gene flow between blackflies in all 3 river basins, thus confirming the results of the short amplicon sequence data
- The blackflies in West Africa are monophyletic to those of Central and South-Eastern Africa
- There was no population structure by geographical location or river basin with the majority of variation occurring within population (99.1%) and very little occurring among population (0.9%)

- Three genetically distinct group of blackflies were observed in the transition ecological zone of Ghana that could be heterogeneous with gene flow
- The long-range data suggest that the presence of polytomies is not due to the lack of genetic divergence but possibly due to *Simulium* blackflies undergoing radiation that does not allow lineage sorting at the present time
- A single onchocerciasis transmission zone has been confirmed with these long-range data

6.4.3 Chapter four

- The unique haplotypes further showed that the Ghanaian blackflies were genetically diverse such that the historical population size was possibly very large
- The blackflies in all the studied 3 river basins from Daka, through Pru to Black Volta along the East-West transect of the central ecological transition zone of Ghana were evolving randomly as one would expect from a neutrally evolving population
- the majority of the blackflies' genetic variation occurred within population (97.9%, P > 0.001), with the remaining variation occurring among populations (2.1%, P > 0.001)
- A total of 82 whole mitochondrial genomes have been sequenced and assembled. They will be submitted to public database to enable other researchers have access to use as secondary data
- The mitochondrial genome of *Simulium* blackflies from Africa has been determined to be circular like those of related dipteran species. Its size ranges from 16,108 bp to 16,232 bp and consist of 2 rRNA genes, 22 tRNA genes, 13 protein coding genes and a non-coding control region (D-LOOP)
- The overall genetic diversity in the mitochondrial genome of the West African *Simulium* blackfly was characterized as 1 and 0.00853 for the haplotype diversity and nucleotide diversity respectively
- This study demonstrated that the entire central ecological transition zone of Ghana constitutes a single onchocerciasis transmission zone and this finding does not agree with the transmission zone definition based on *Simulium* vector breeding sites (WHO, 2016)

6.4.4 Chapter five

- The contig length of the largest genome assembly indicated that the African *Simulium* blackflies have a genome size of 209,986,818 bp. The complete genome size is now estimated at approximately 210 Mbp

- There was a very high haplotype diversity ($Hd = 1$) but relatively moderate nucleotide diversity ($\pi = 0.001$) in the *Simulium* blackflies' nuclear genome in all the 3 river basins
- The blackflies in all the 3 river basins of Pru, Black Volta and Daka were evolving randomly ($p > 0.10$) as one would expect from a neutrally evolving population, and there was a high frequency of rare alleles (Tajima's $D < 0$)
- The *Simulium* blackflies in Daka, Pru and Black Volta were genetically homogenous when observed from the viewpoint of river basin
- The West African blackflies were monophyletic to those of Central and South-Eastern Africa. All the African blackflies were also monophyletic to those from Asia and Europe
- Hard polytomies were observed among the Central Africa blackflies
- Most of the blackflies' genetic variation occurred within population (99.1%, $P > 0.001$), with the remaining variation occurring among populations (0.9%, $P > 0.001$).
- the additional data from both matrilineal and patrilineal lineages confirmed the lack of population structure by geographical place of origin, and the presence of continuous gene flow between blackflies in all 3 river basins due to the interbreeding of the blackflies in the entire transition zone
- the entire central ecological transition zone of Ghana constitutes a single onchocerciasis transmission zone and this finding does not agree with the current transmission zone definition based on *Simulium* vector breeding sites (WHO, 2016)

6.5 Data management and results dissemination

The nucleotide sequences generated and used in this study will be submitted to the NCBI nucleotide database. The results will be disseminated through conferences, workshops and article publications.

6.6 Technological innovations

- An optimal protocol of library preparation has been developed that successfully targets the sequencing of large numbers of complete mitochondrial genomes of blackflies at low depth as well as a recovery of significant proportion of the nuclear

genome, and it can be used for sequencing from other insect genomes with similar mitochondrial – nuclear genome copy number ratio

- It was demonstrated that the use of total DNA in library preparation and low sequencing depth produces complete mitochondrial genomes and significant amounts of nuclear genes in *Simulium* blackflies at a faster rate with lesser resources than conventional PCR method, and the procedure can be replicated in sequencing of other genomes of animals. This method of generating whole mitochondrial genome is faster and utilizes lesser money and resources than amplicon sequencing that uses PCR and gel electrophoresis. It took over 3 years of my PhD time to generate little amplicons for sequencing. But with the use of this proven and robust method of preparing libraries and sequencing, it took less than a month to generate more than 10 times sequencing data than the entire 3 years of PCR-Gel works
- The library-sequencing approach used in this study will enable accelerated sequencing of most mitochondrial genomes that have significant variation or range in the copy number ratio of mitochondrial DNA to nuclear DNA, especially those of insects and worms. With the demonstrated library-sequencing approach, the average number of individual complete mitochondrial genomes that can be generated in one iSeq is 4-5, MiniSeq and MiSeq are 23 each, and NextSeq is 384 (the maximum read length or chemistry used in this extrapolation is 2 x 150 bp for iSeq; 2 x 150 bp for MiniSeq; 2 x 300 bp for MiSeq; and 2 x 150 bp for NextSeq). If similar approach is followed, there will be accelerated development in the amount of sequences to be generated in the next few years to come. This should equally translate into accelerated research and development in all areas of endeavour.
- Goodbye to PCR method: The use of total DNA to prepare libraries for sequencing at lower depth, a method demonstrated and perfected in this PhD work, will enable researchers save at least 3 years of laborious PCR works in attempts to generate complete mitochondrial genomes of large numbers of individual samples. Students pursuing masters and PhD, that require large numbers of complete mitochondrial genomes or any part of it, can now complete their work faster without having to spend over 3 years trying to produce large amounts of genomic data using PCR method. We now say goodbye to the less efficient and time-consuming PCR method of mitochondrial genome sequencing and embrace the future of a faster and comparatively less expensive library-sequencing method

- A pipeline for a faster whole mitochondrial genome assembly within 5-20 minutes has been developed
- Eighty-two (82) complete mitochondrial genome sequences of African *Simulium* blackflies from West, Central and South-Eastern Africa have been sequenced, assembled and annotated. These are ready to be submitted to public database to support onchocerciasis elimination efforts. At the commencement of this PhD study in 2015, there was not even a single complete mitochondrial genome from any part of the world in any known public database. Years later, there has been the sequencing of only 4 complete mitochondrial genomes of blackflies and subsequently submitted to public database: 2 from China, 1 from Great Britain, 1 from Thailand, and none from Africa. This PhD work has produced at least 82 complete mitochondrial genomes from Africa in a relatively short period of time and serve as a foundation for much more sequencing (hundreds to thousands) of complete mitochondrial genomes of blackflies in the field in the near future. In the entire history of onchocerciasis, a neglected tropical disease that has infected over 37 million people worldwide and with 123 million people at risk; there has never been a single whole mitochondrial genome of African *Simulium* blackfly found in any known public database to support the research community. If submitted in time, our work from the African endemic foci, which has 99% of the disease burden, will be the first in history to make it to the general public.
- A world map has been produced that shows the most updated status of countries that are onchocerciasis endemic. The map will be updated from time to time to capture subsequent whole mitochondrial genome works in the world, and publicly made available with details of the sequenced genomes

6.7 Information for policy makers and decision implementers of global onchocerciasis control

- The observation that there is just a single onchocerciasis transmission zone across the entire central ecological transition zone of Ghana may have significant impact on timelines of onchocerciasis elimination from the foci, as well as the major decisions and strategies that need to be taken to that effect
- Onchocerciasis control and elimination efforts cannot be stopped in communities from any given river basin without risking blackflies coming from the other river basins to reinfect people

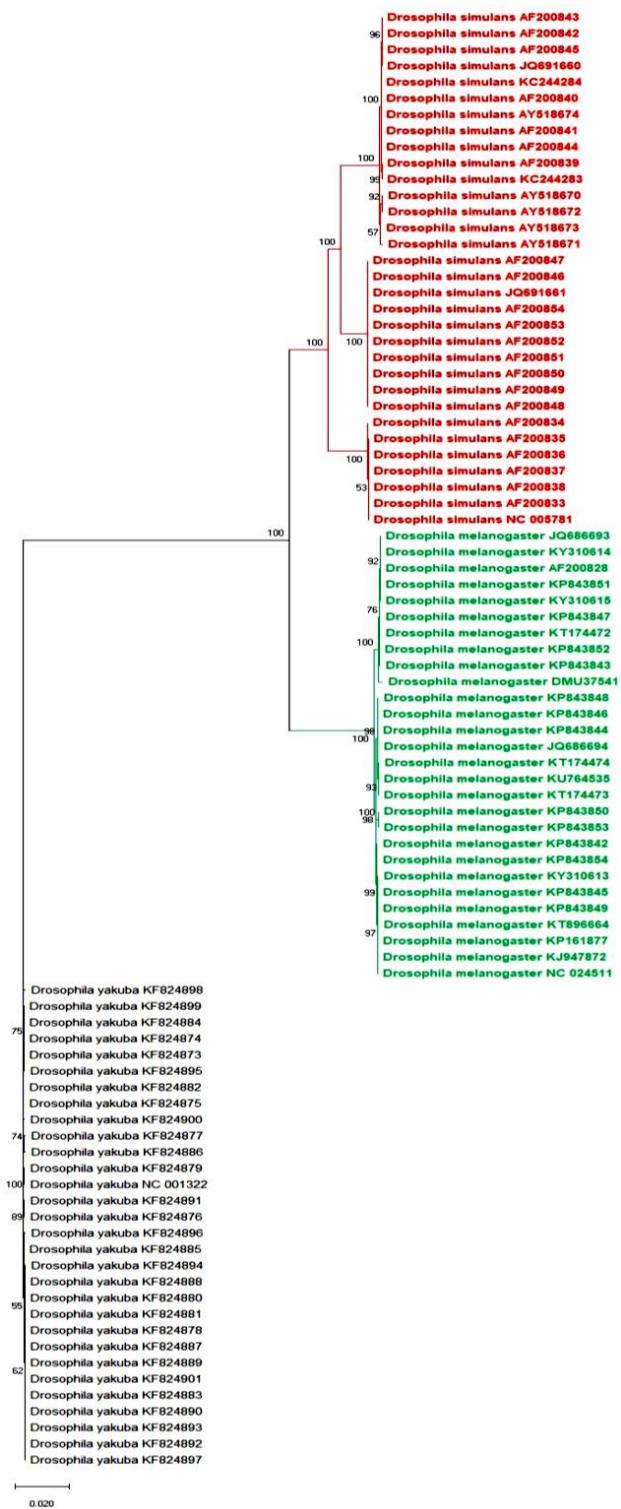
- This calls for assessment and mapping of onchocerciasis transmission zones across all African endemic countries to meet disease elimination targets
- Assessment of re-introduction of some form of vector control at critical endemic foci to break the continuous gene flow of *Simulium* blackfly over long geographical ranges, thereby creating patches of onchocerciasis transmission zones, similar to that of the Latin American foci, that can be better controlled towards effective and early disease elimination

6.8 Future Directions

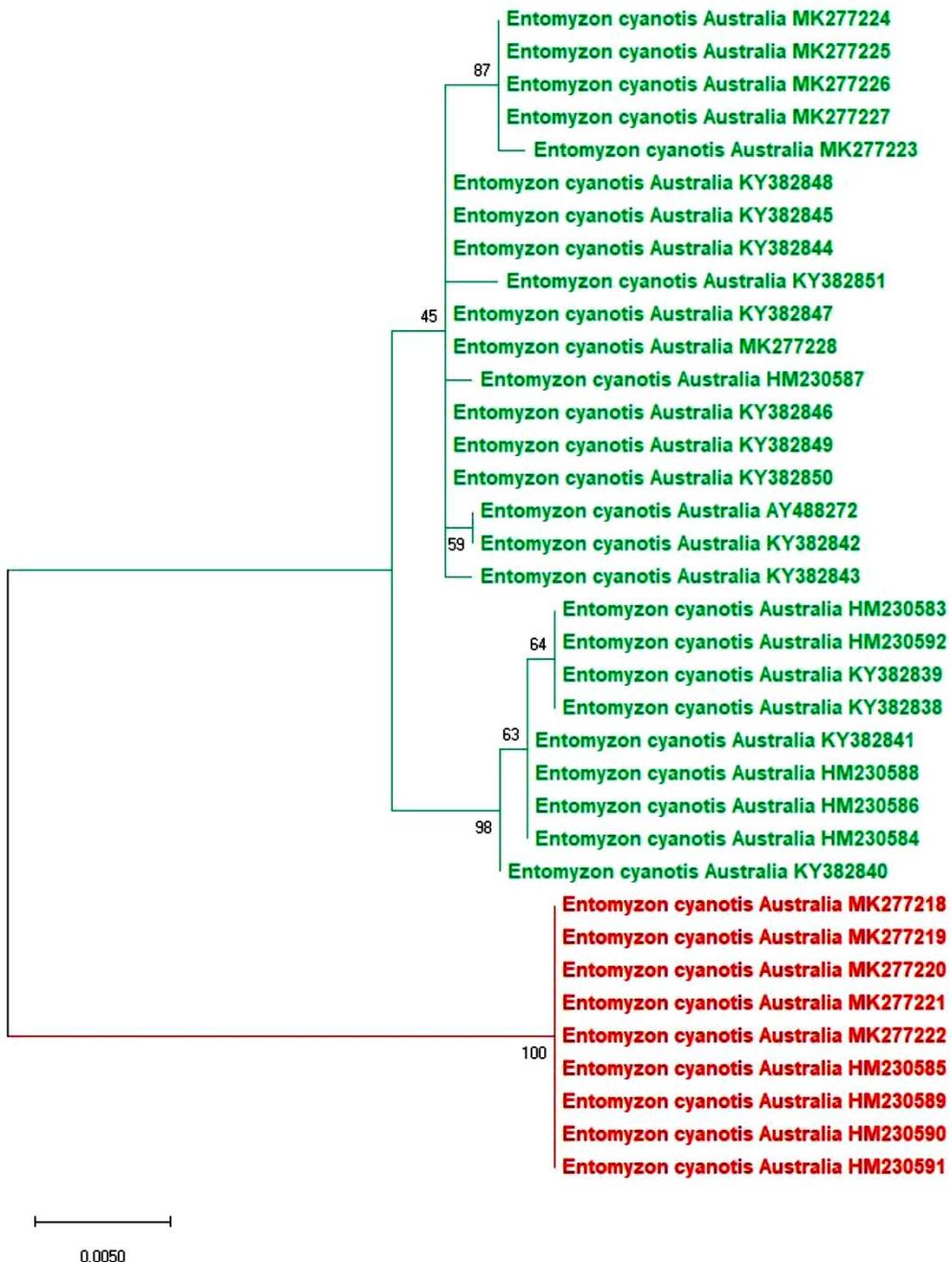
- The assembly of whole nuclear genome data is required to assess the extent of lineage sorting and to determine if the nuclear data will be congruent with that of the whole mitochondrial genome.
- Large number of individual blackflies from the study area will need to be sequenced to increase the proportion of the 3 distinct molecular groups observed in the *a priori* DAPC analysis
- Sequencing of cytotyped larvae would help correlate genomic variation with cytological variation, and elucidate the genetic basis of “cytotypes”, and whether cytotypes are isolated genetically from each other
- Similar molecular phylogenetic analysis will need to be performed in the remaining onchocerciasis endemic countries to find out the status of onchocerciasis transmission zone. This will enable control programmes to adequately determine the foci that can stop ivermectin mass drug administration without risking reinfection of the people in those regions already treated

Appendices

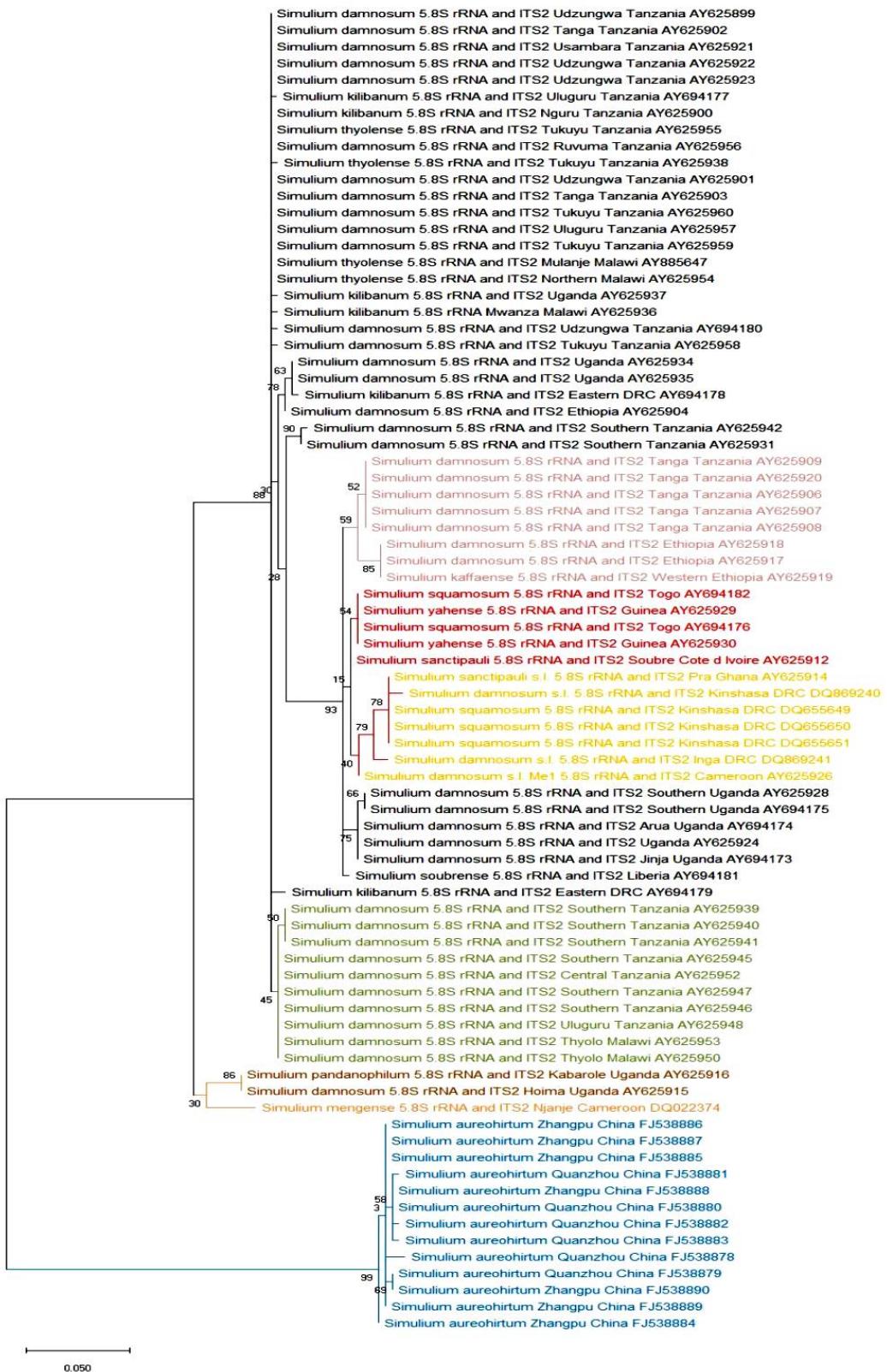
Supplementary Figures A



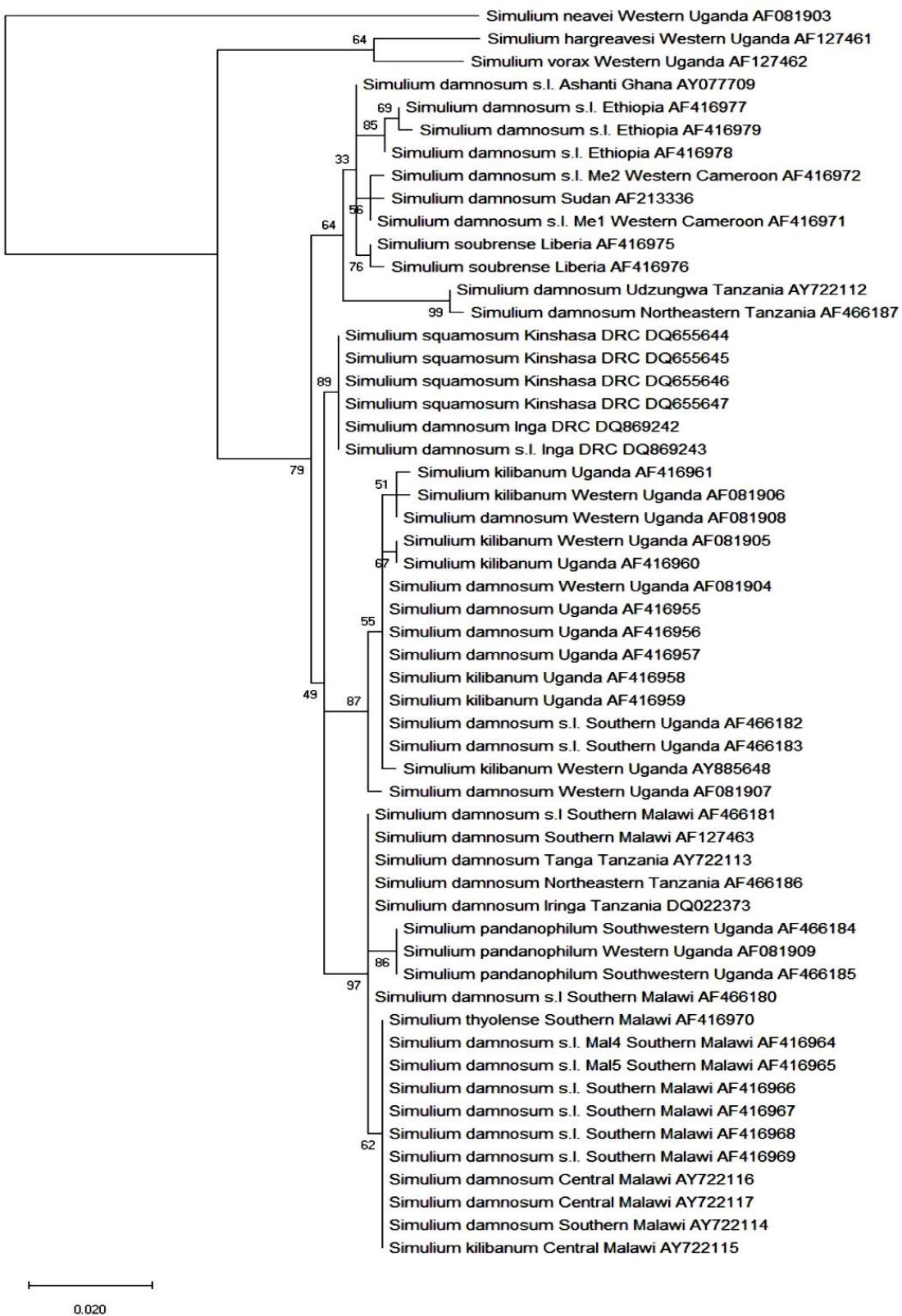
Supplementary Figure 2.3.1: *Drosophila* sp. phylogenetic tree



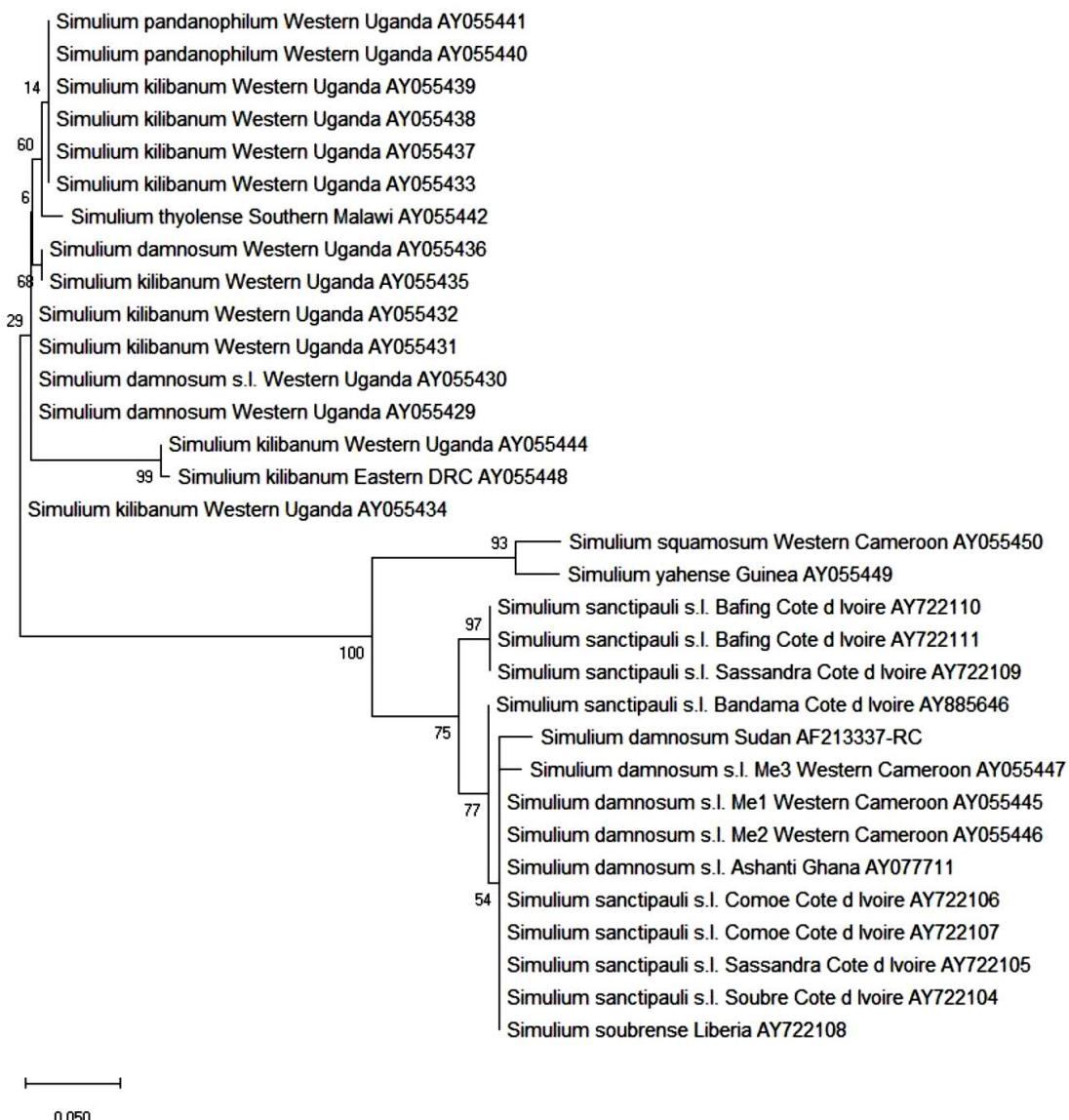
Supplementary Figure 2.3.2: *Entomyzon* sp. phylogenetic tree



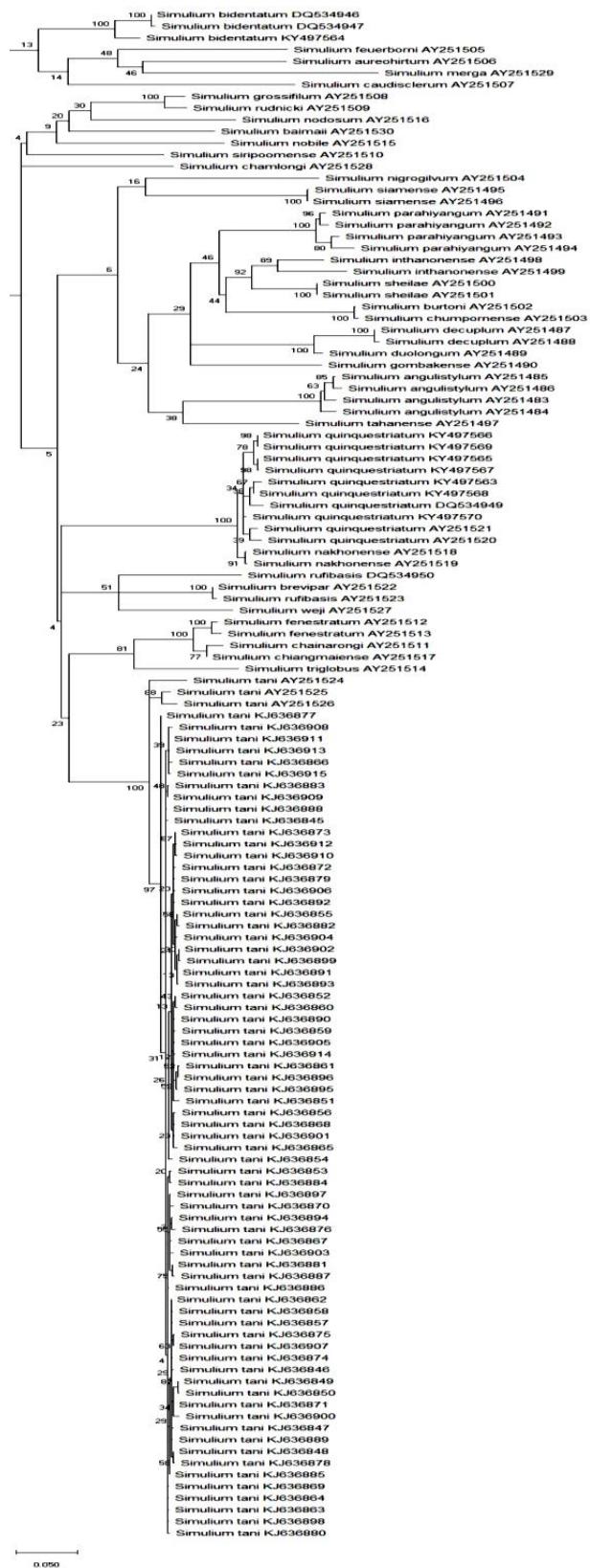
Supplementary Figure 2.3.3: *Simulium* sp. (ITS2) phylogenetic tree



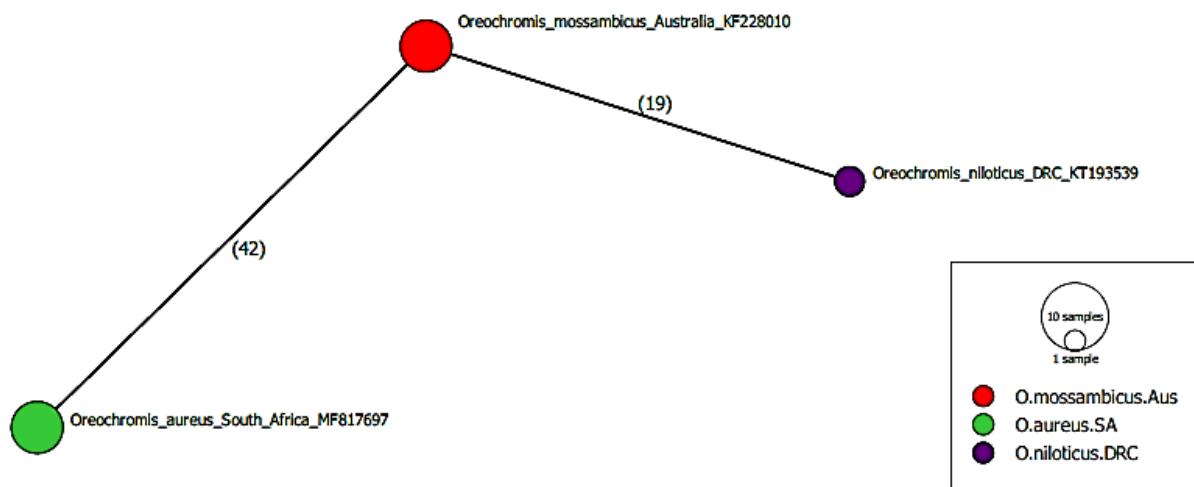
Supplementary Figure 2.3.4: *Simulium* sp. (16S) phylogenetic tree



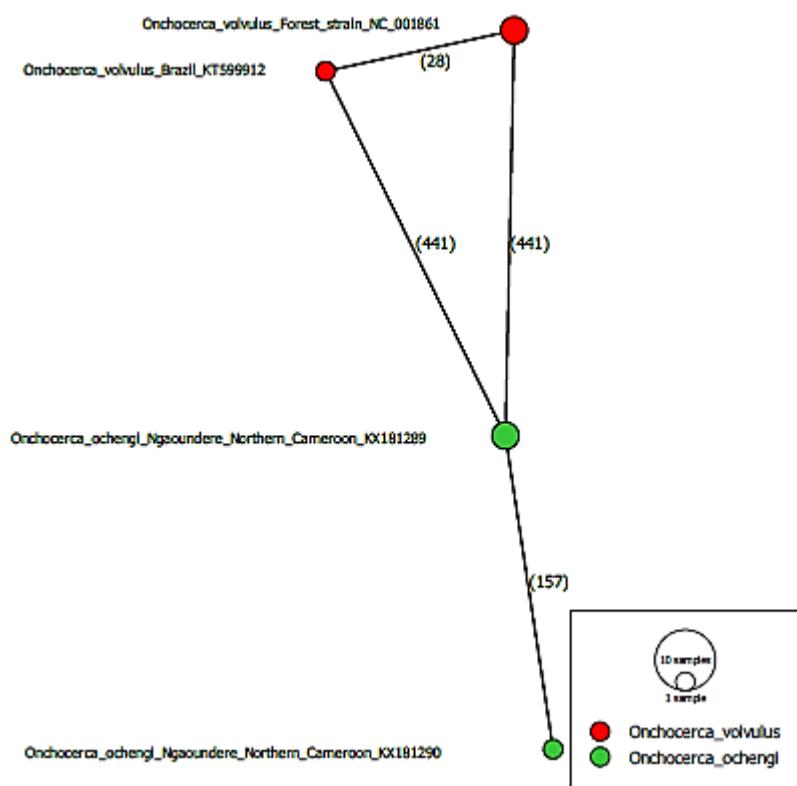
Supplementary Figure 2.3.5: *Simulium* sp. ND4 phylogenetic tree



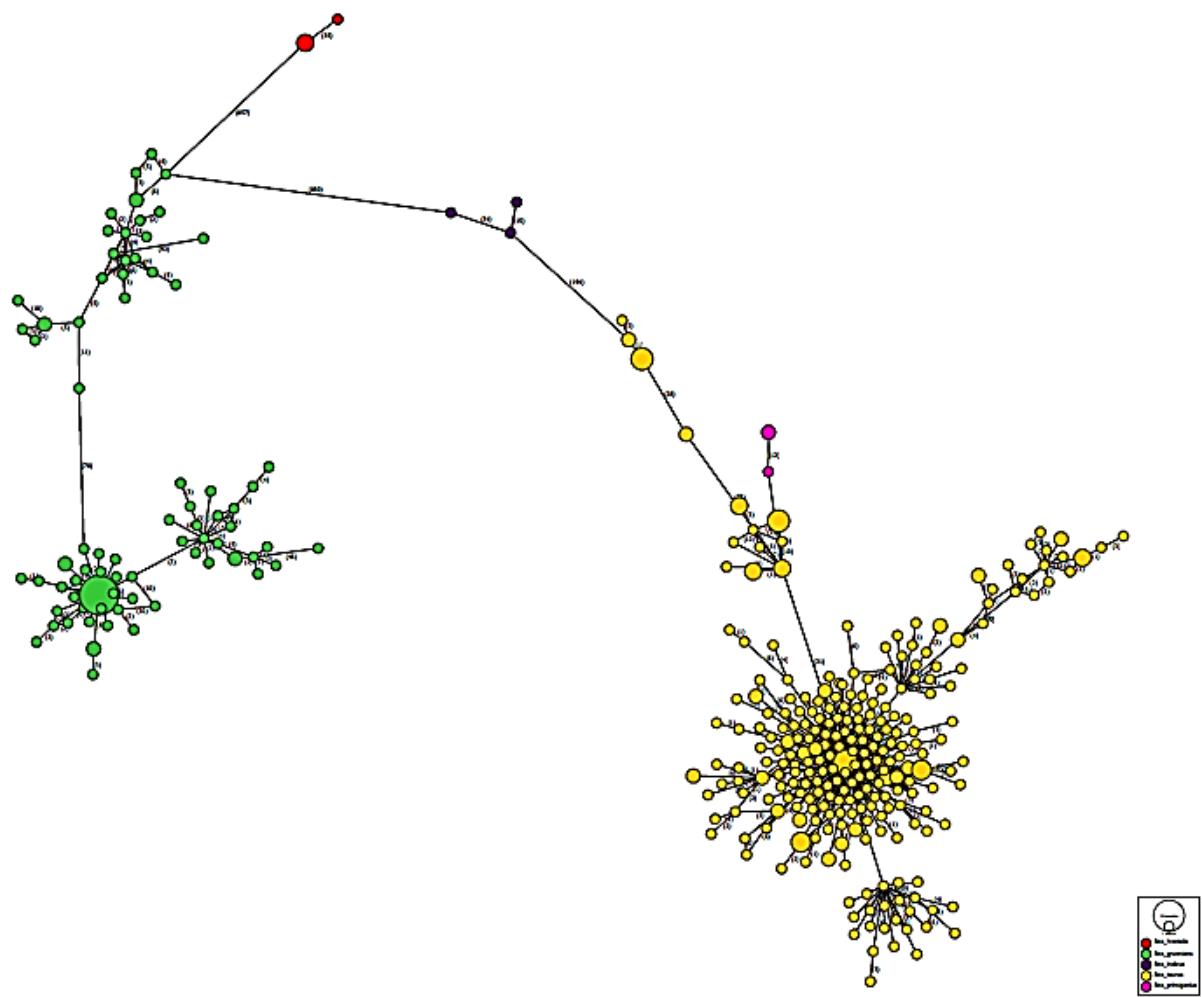
Supplementary Figure 2.3.6: *Simulium* COI phylogenetic tree



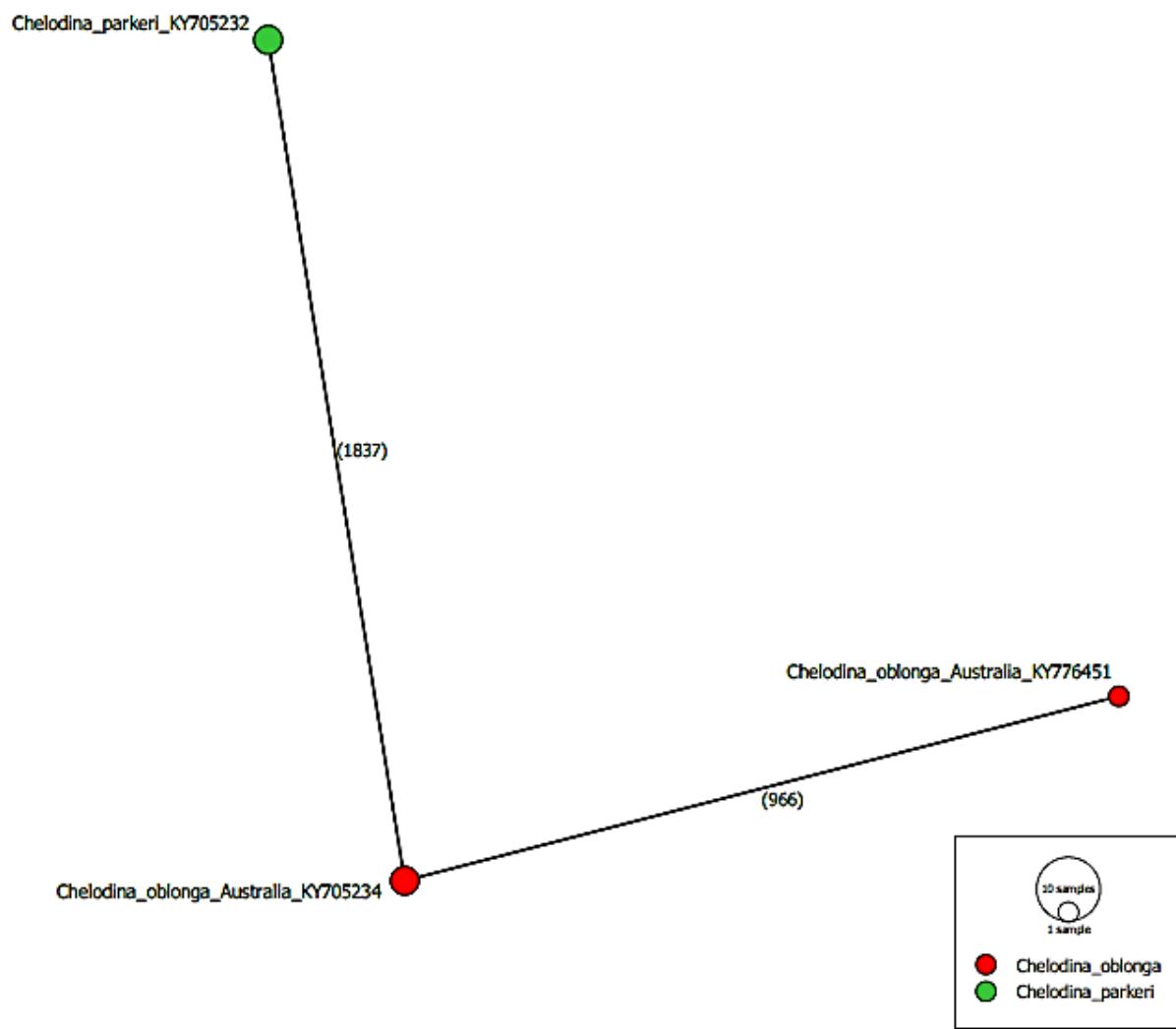
Supplementary Figure 2.3.7: *Oreochromis* sp. haplotype network



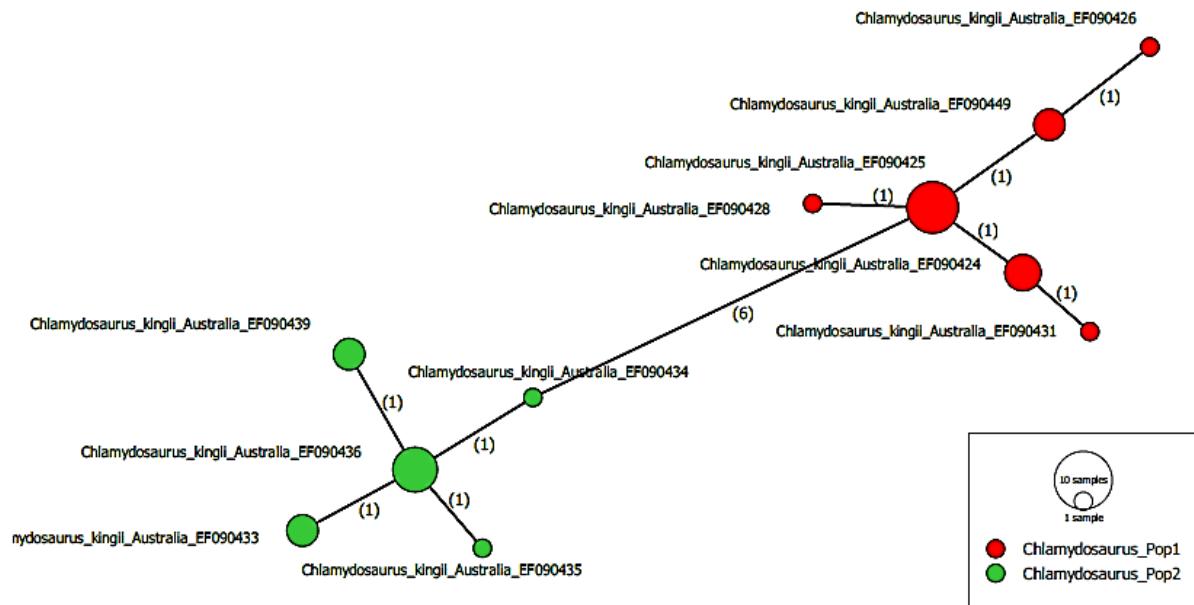
Supplementary Figure 2.3.8: *Onchocerca* sp. haplotype network



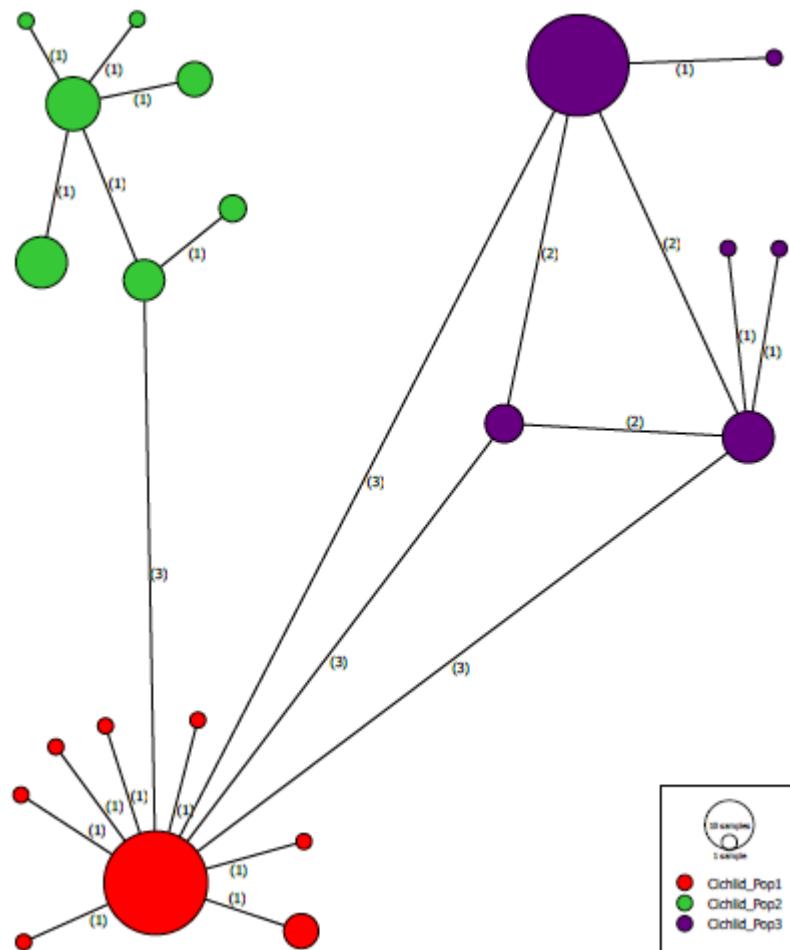
Supplementary Figure 2.3.9: *Bos* sp. haplotype network



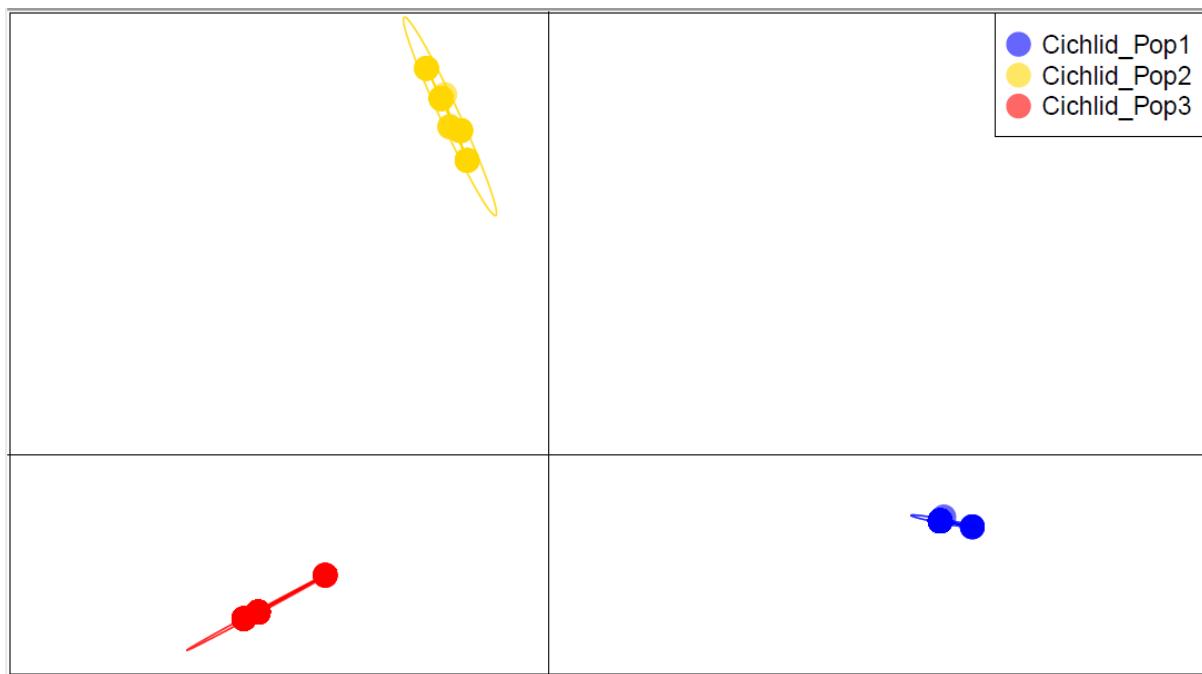
Supplementary Figure 2.3.10: *Chelodina* sp. haplotype network



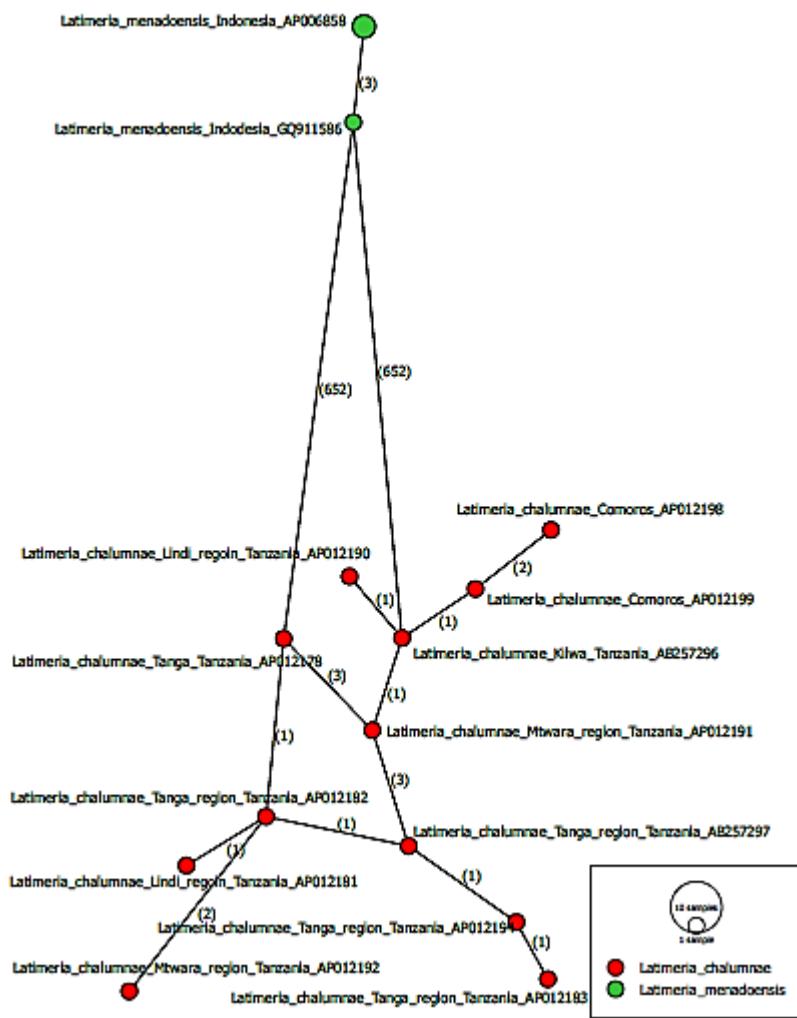
Supplementary Figure 2.3.11: *Chlamydosaurus kingii* haplotype network



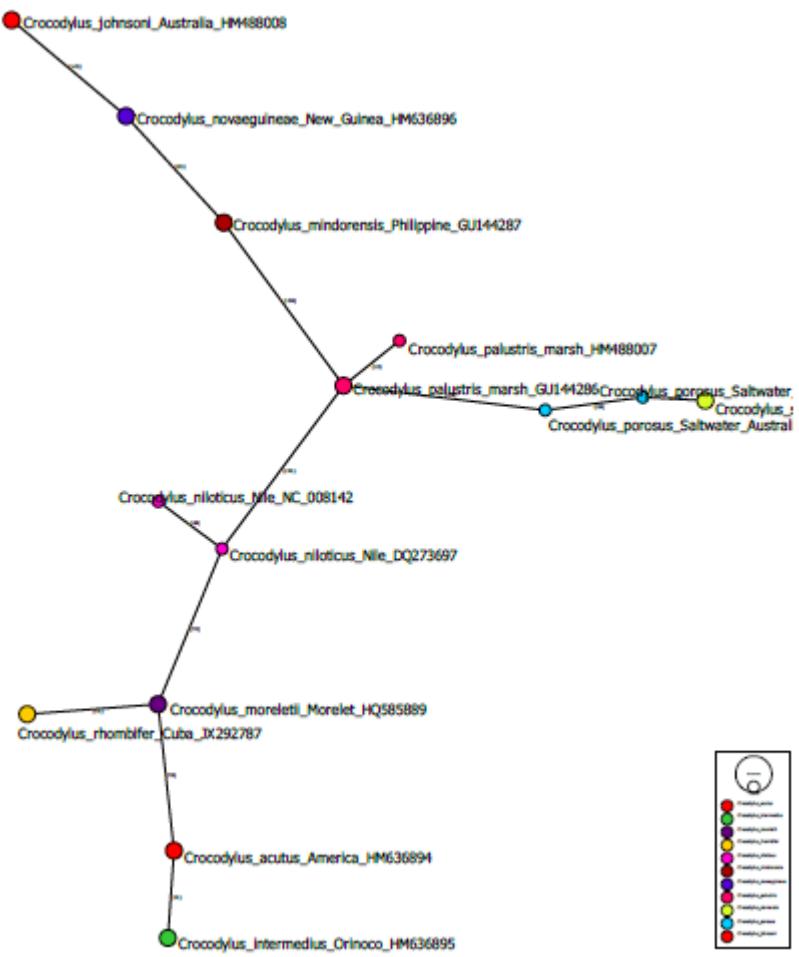
Supplementary Figure 2.3.12A: *Variabilichromis moorii* haplotype network



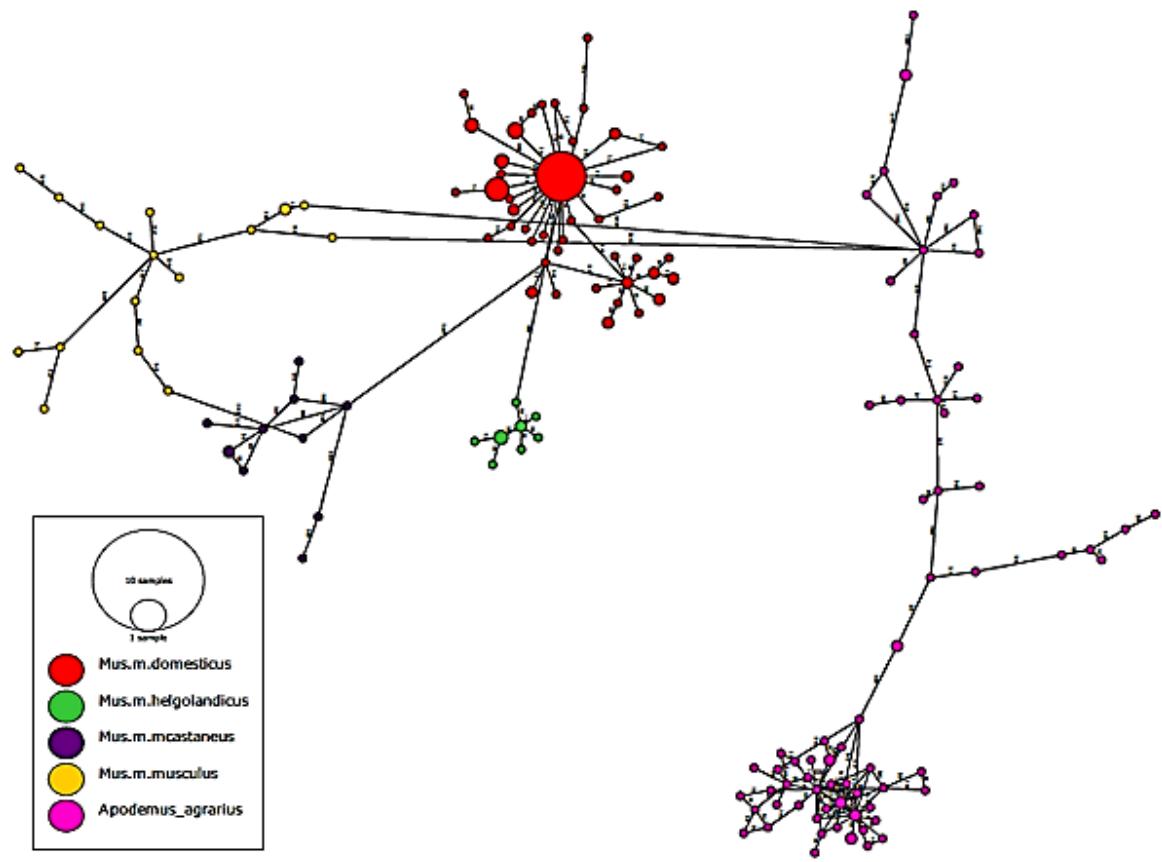
Supplementary Figure 2.3.12B: *Variabilichromis moorii* DAPC



Supplementary Figure 2.3.13: *Latimeria* sp. haplotype network

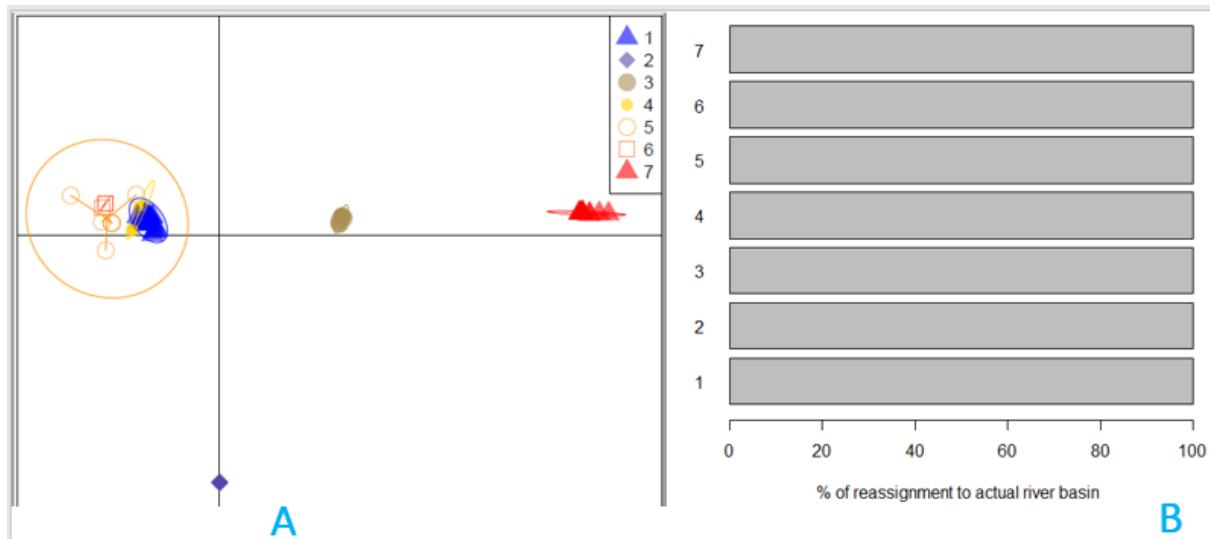


Supplementary Figure 2.3.14: *Crocodylus* sp. haplotype network

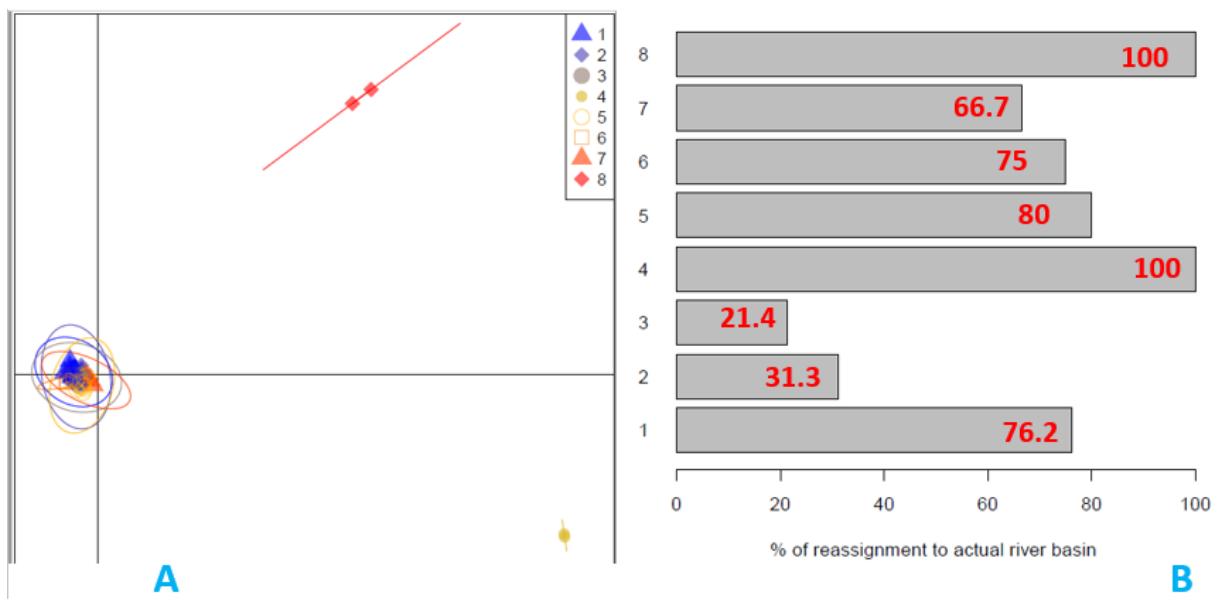


Supplementary Figure 2.3.15: *Mus* sp. and *Apodemus agrarius* haplotype network

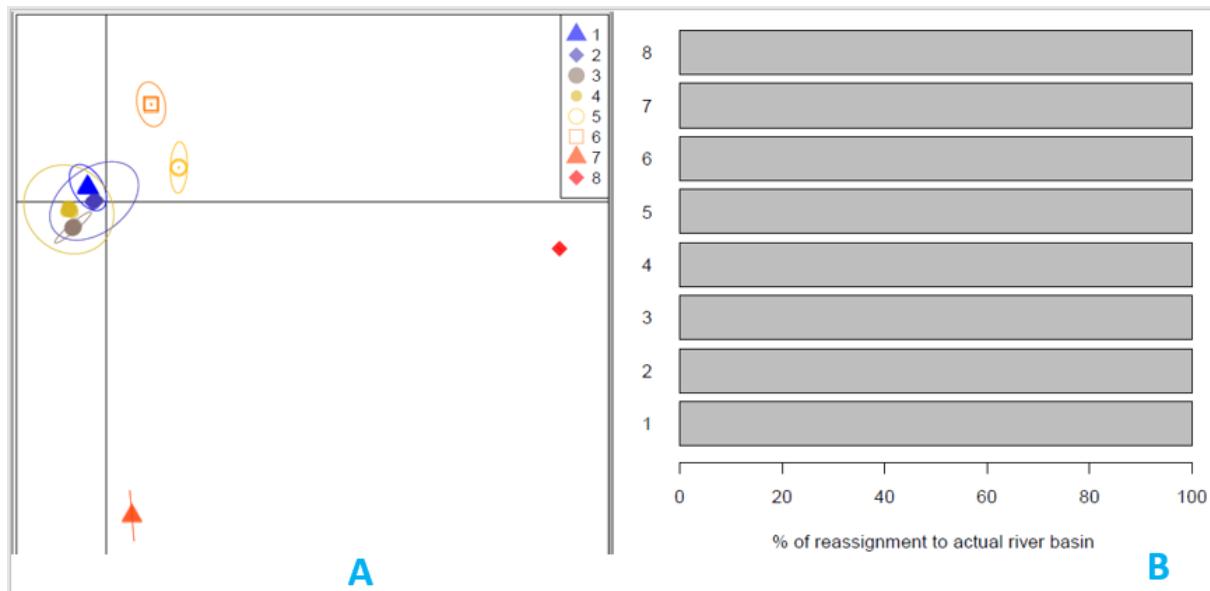
Supplementary Figures B



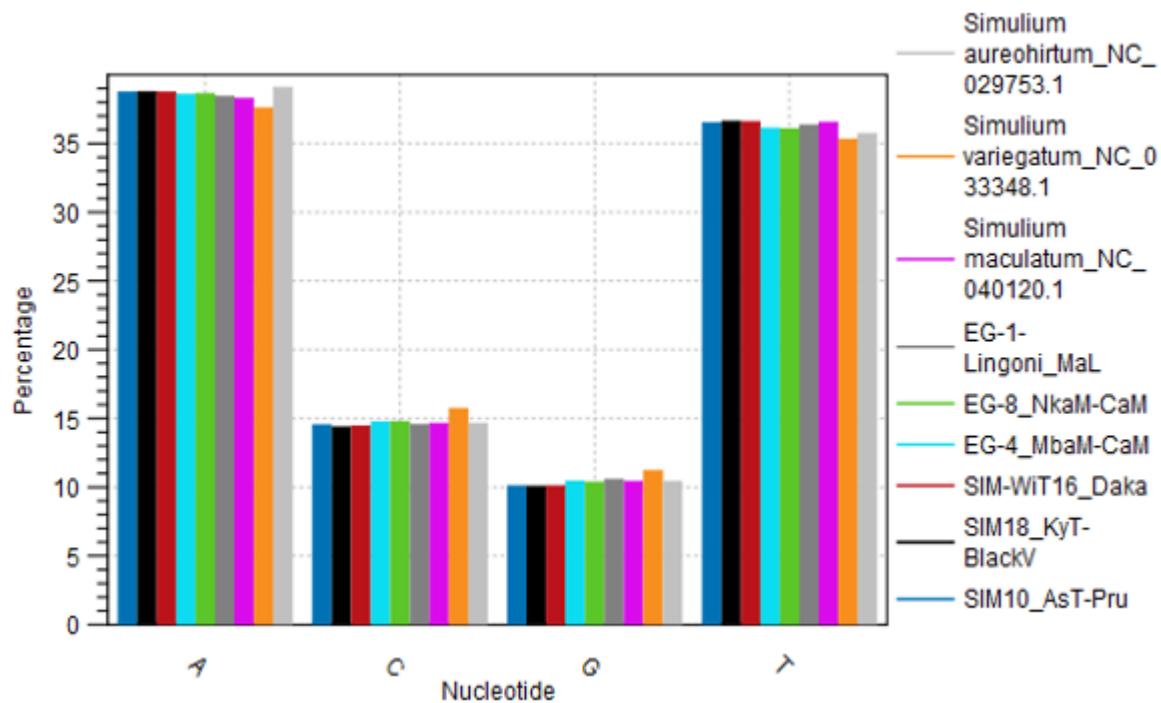
Supplementary Figure 2.3.16: *A priori* assessment of genetic relatedness from 90 CO1 gene sequences. A= DAPC, and B = Percentage of reassignment to cluster. The DAPC was plotted using 12 principal component, 7 clusters and 6 discriminant functions.



Supplementary Figure 4.3.1: DLOOP from 75 mitochondria genomes. A= DAPC, and B = Percentage of reassignment to cluster. The DAPC was plotted using 30 principal component, 8 clusters and 2 discriminant functions.

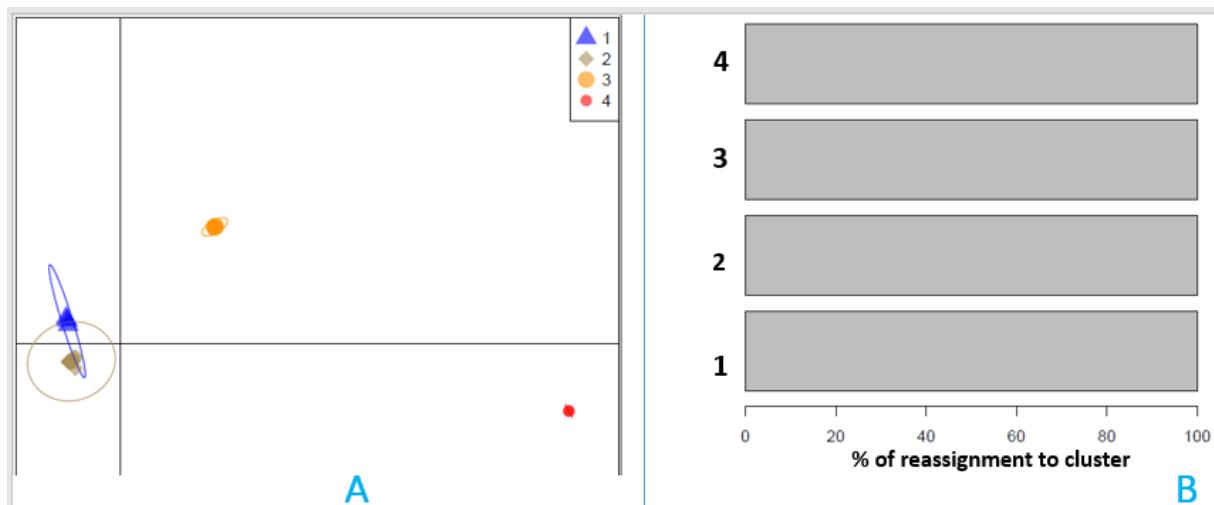


Supplementary Figure 4.3.2: Scatterplot from 75 whole mitochondrial genomes with DLOOP. A= DAPC, and B = Percentage of reassignment to cluster. The DAPC was plotted using 30 principal component, 8 clusters and 2 discriminant functions.

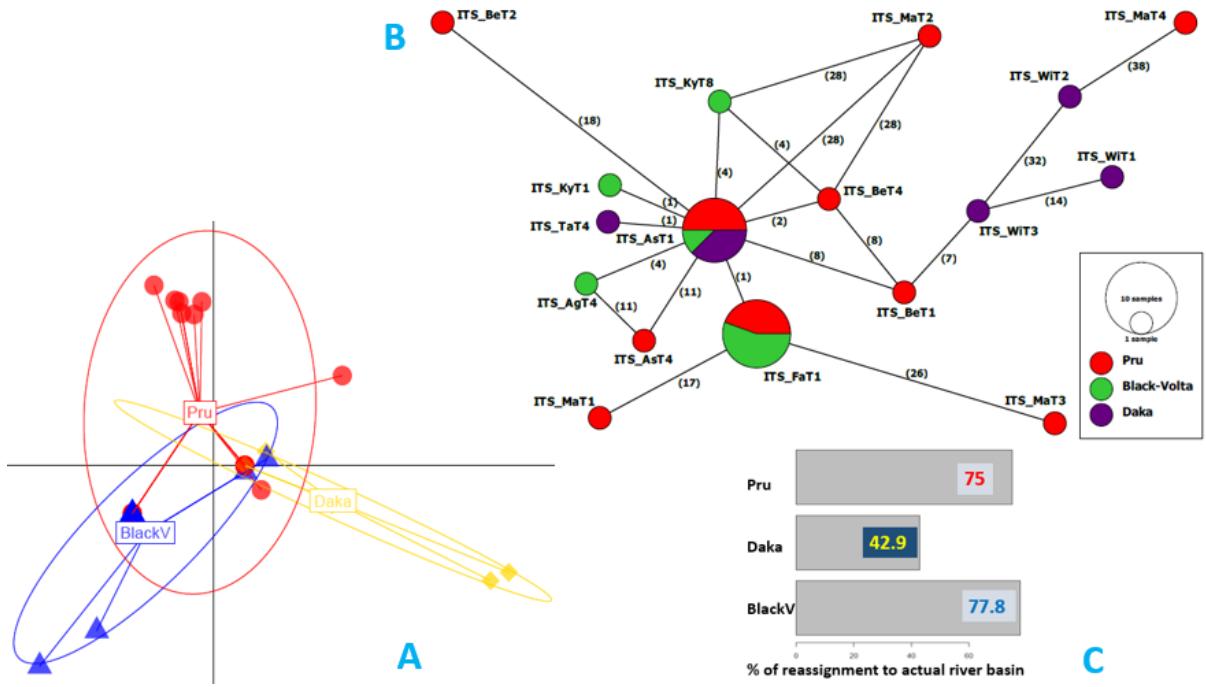


Supplementary Figure 4.3.3: Histogram of nucleotide frequencies of whole mitochondrial genome sequences. The sequences are from blackflies sampled from Ghana, Cameroon, Malawi, and compared with those from public database (NCBI accessions NC_040120.1, NC_033348.1, and NC_029753.1)

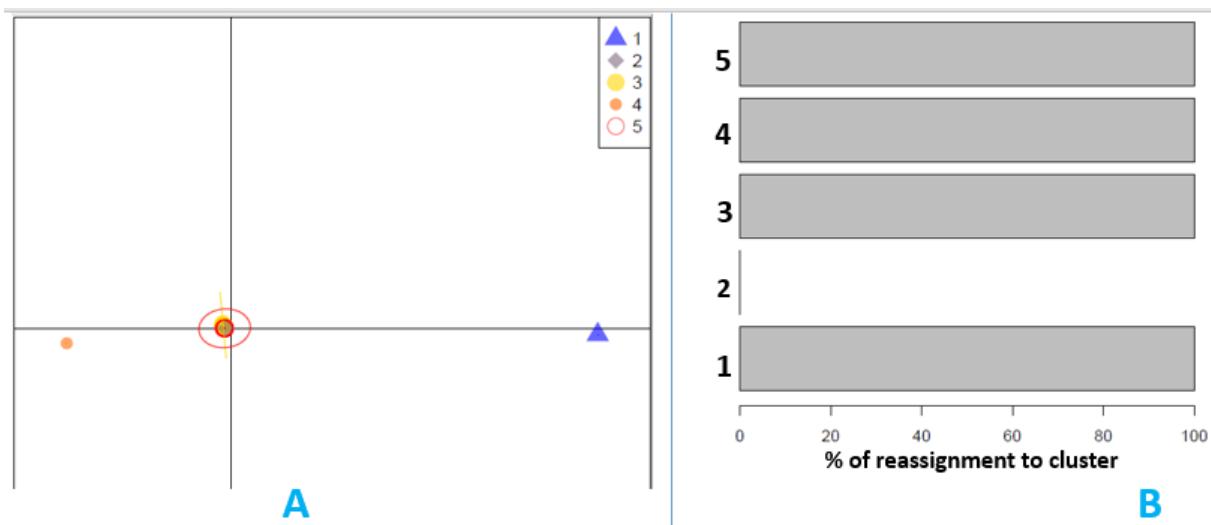
Supplementary Figures C



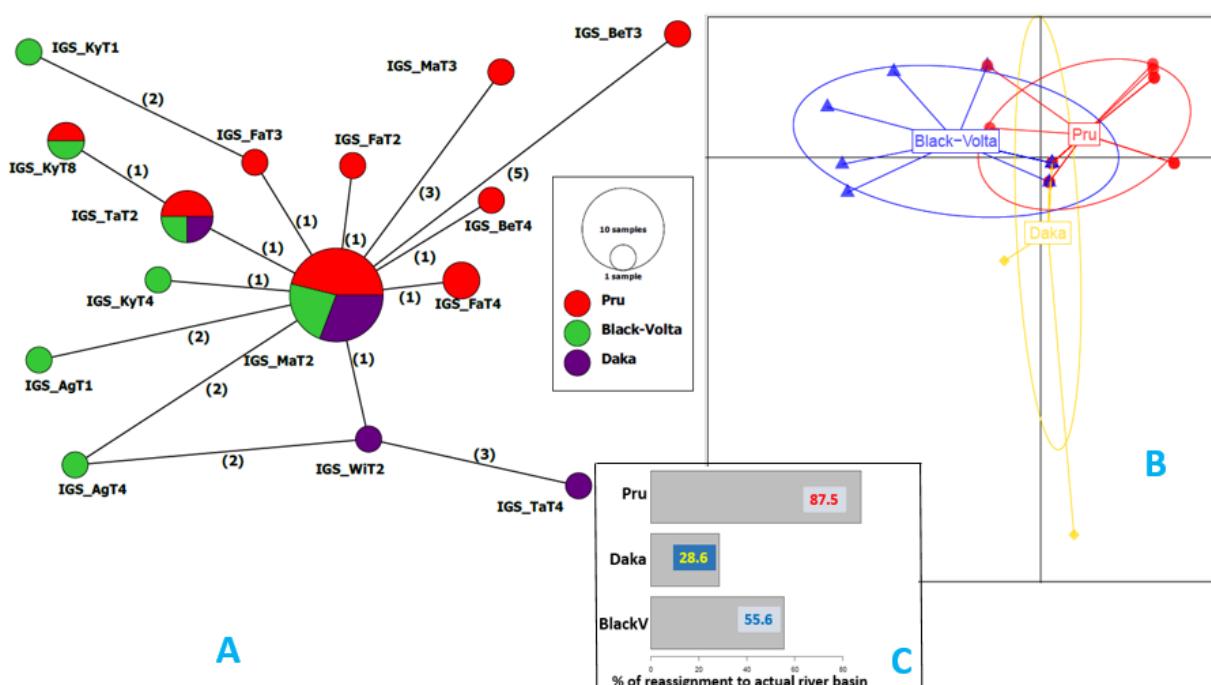
Supplementary Figure C 2.3.12: *A priori* assessment of genetic relatedness in 32 ITS2 gene sequences without population assignment. In section “A” of the figure, the ellipses of the DAPC graph represent the different genetic clusters while the dots represent the individual blackflies collected. It was plotted by choosing and retaining 10 principal components, 4 clusters and 2 discriminant functions. The use of 3 discriminant functions yielded similar results as 2. In section “B” of the figure, the bar chart represents the percentage of successful reassignment of individual blackflies to their original genetic clusters identified. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).



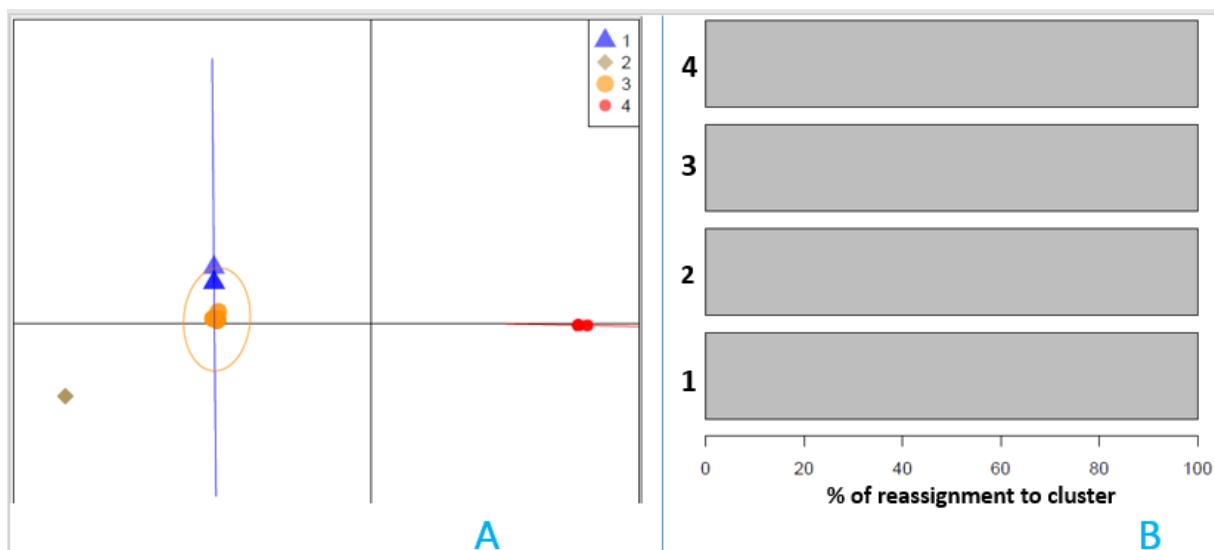
Supplementary Figure C 2.3.13: *A posteriori* assessment of genetic relatedness in 32 ITS2 gene sequences after population assignment. Section “A” of the figure shows the DAPC plot for the 32 ITS2 gene sequences sampled from the river basins of Pru (red colour), Black Volta (blue colour) and Daka (yellow colour). The ellipses represent the river basins whilst the dots represent the individual blackflies. Section “C” of the figure shows a bar chart of the percentage of reassignment of individual blackflies to their original river basin of collection. Graphs A and C were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010). In the figure, section “B” = Haplotype network produced by using the minimum spanning network and Epsilon of 0 in Popart 1.7 (Bandelt *et al.*, 1999). It was produced from 32 ITS2 gene sequences from the river basins of Pru (red colour), Black Volta (green colour) and Daka (purple colour). The circles (nodes) represent haplotypes and the size show the frequency corresponding to the total number of individuals within the haplotype. The numbers indicate the mutational steps difference(s) between haplotypes.



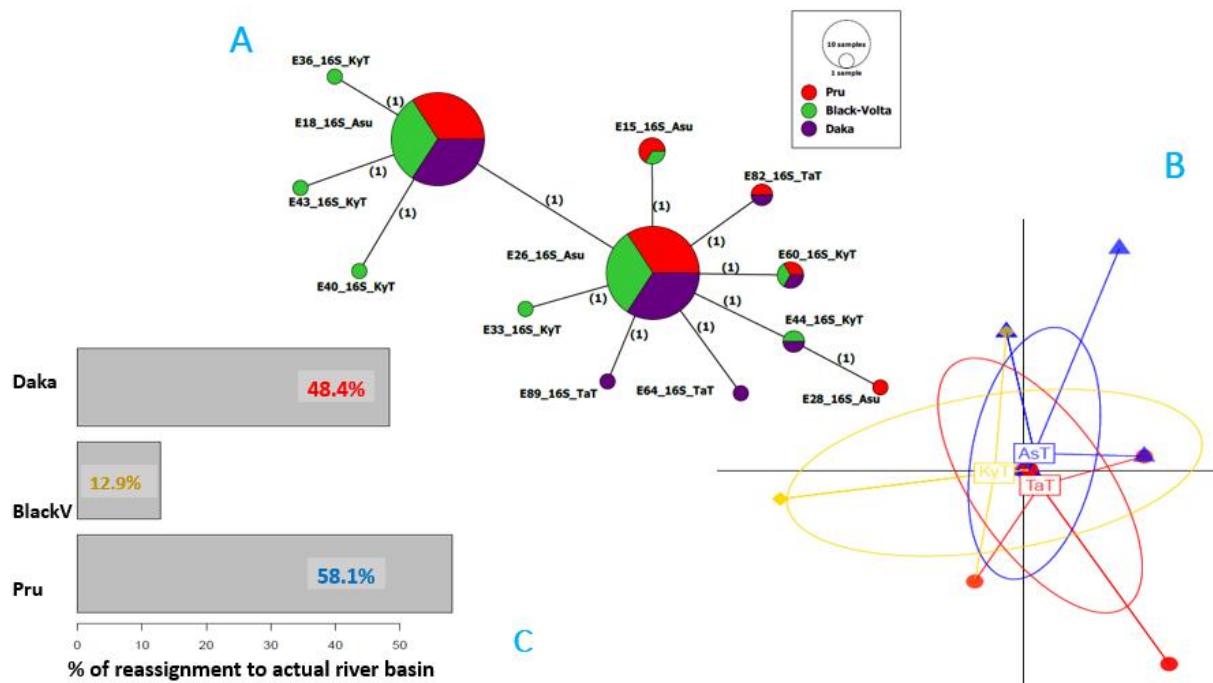
Supplementary Figure C 2.3.14: *A priori* assessment of genetic relatedness in 32 IGS gene sequences without population assignment. In section “A” of the figure, the ellipses of the DAPC graph represent the different genetic clusters while the dots represent the individual blackflies collected. It was plotted by choosing and retaining 6 principal components, 5 clusters and 2 discriminant functions. The use of 4 discriminant functions yielded similar results as 2. In section “B” of the figure, the bar chart represents the percentage of successful reassignment of individual blackflies to their original genetic clusters identified. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).



Supplementary Figure C 2.3.15: *A posteriori* assessment of genetic relatedness in 32 IGS gene sequences after population assignment. In the figure, section “A” = Haplotype network produced by using the minimum spanning network and Epsilon of 0 in Popart 1.7 (Bandelt *et al.*, 1999). It was produced from 32 IGS gene sequences from the river basins of Pru (red colour), Black Volta (green colour) and Daka (purple colour). The circles (nodes) represent haplotypes and the size show the frequency corresponding to the total number of individuals within the haplotype. The numbers indicate the mutational steps difference(s) between haplotypes. Section “B” of the figure shows the DAPC plot for the 32 IGS gene sequences sampled from the river basins of Pru (red colour), Black Volta (blue colour) and Daka (yellow colour). The ellipses represent the river basins whiles the dots represent the individual blackflies. Section “C” of the figure shows a bar chart of the percentage of reassignment of individual blackflies to their original river basin of collection. Graphs B and C were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).



Supplementary Figure C 2.3.16: *A priori* assessment of genetic relatedness in 90 16S rRNA gene sequences without population assignment. In section “A” of the figure, the ellipses of the DAPC graph represent the different genetic clusters while the dots represent the individual blackflies collected. It was plotted by choosing and retaining 4 principal components, 4 clusters and 3 discriminant functions. In section “B” of the figure, the bar chat represents the percentage of successful reassignment of individual blackflies to their original genetic clusters identified. The graphs were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).



Supplementary Figure C 2.3.17: *A posteriori* assessment of genetic relatedness in 90 16S rRNA gene sequences after population assignment. In the figure, section “A” = Haplotype network produced by using the minimum spanning network and Epsilon of 0 in Popart 1.7 (Bandelt *et al.*, 1999). It was produced from 90 16S rRNA gene sequences from the river basins of Pru (red colour), Black Volta (green colour) and Daka (purple colour). The circles (nodes) represent haplotypes and the size show the frequency corresponding to the total number of individuals within the haplotype. The numbers indicate the mutational steps difference(s) between haplotypes. Section “B” of the figure shows the DAPC plot for the 90 16S rRNA gene sequences sampled from the river basins of Pru (AsT = blue colour), Black Volta (KyT = yellow colour) and Daka (TaT = red colour). The ellipses represent the river basins whilst the dots represent the individual blackflies. Section “C” of the figure shows a bar chart of the percentage of reassignment of individual blackflies to their original river basin of collection. Graphs B and C were produced in R x64 3.5.0 (Jombart, 2008; Jombart *et al.*, 2010).

Supplementary Table

Supplementary Table 2.3.1 A: *Bos* sp. accession numbers used in the meta-analysis

MF614103, NC_036020, MK279400, MK279401, GQ464260, GQ464251, KM233416, GQ464311, GQ464313, GQ464257, GQ464262, GQ464250, GQ464261, JQ846022, GQ464293, GQ464300, GQ464284, JQ846021, GQ464269, GQ464278, KJ704989, GQ464288, GQ464287, GQ464303, GQ464267, MF973066, GQ464291, GQ464296, KX232524, GQ464249, GQ464304, AY684273, GQ464266, GQ464247, GQ464246, GQ464264, GQ464306, GQ464282, KX232526, GQ464298, GQ464254, GQ464255, MG837551, GQ464283, GQ464276, GQ464310, JQ692071, KM658599, EF494179, NC_006380, GQ464275, GQ464280, KU891851, KR052524, GQ464305, GQ464309, GQ464292, GQ464279, JQ846020, KX232527, KJ463418, GQ464307, GQ464302, GQ464299, GQ464290, GQ464263, GQ464281, GQ464301, GQ464312, GQ464273, GQ464271, GQ464274, GQ464277, GQ464285, GQ464308, GQ464294, GQ464295, GQ464289, GQ464265, GQ464259, KX232522, GQ464252, GQ464253, KX232521, GQ464270, GQ464272, JQ437480, GQ464297, GQ464256, KR011113, GQ464258, GQ464248, GQ464286, KX232525, GQ464314, EF494178, KX232523, GQ464268, EF494177, GU256940, KX759625, MF667929, FJ971087, HQ184045, FJ971084, FJ971085, HQ184040, HQ184041, HQ184042, HQ184043, HQ184044, FJ971086, GU985279, NC_013996, JQ437479, EU177866, EU177867, HQ184036, HQ184037, HQ184038, HQ184039, KT184471, KT184472, KP637147, FJ971080, FJ971081, HQ184033, HQ184030, HQ184031, HQ184032, FJ971082, FJ971083, HQ184034, HQ184035, EU177858, EU177857, KT184457, EU177860, EU177861, EU177853, HQ025805, KT184456, DQ124393, DQ124396, EU177854, DQ124383, EU177851, EU177856, EU177859, KT184461, KT184458, AY676856, EU177850, EU177852, EU177849, EU177855, KT184459, KT184460, KF163068, KF163085, KF163064, KF163077, KF163065, KF163066, KF163083, KF163078, KF163071, KF163069, KF163070, KF163091, KF163090, KF163067, KF163062, KF163080, KF163079, KF163086, JN817348, KF163072, JN817343, JN817336, KF163081, JN817300, JN817301, KT184468, JN817334, KT184467, KF163075, JN817349, KF163089, KF163094, KF163084, KT184469, JN817320, KF163092, KF163087, KF163061, KF163076, KF163074, KF163088, KF163073, KF163063, EU177842, JN817350

Supplementary Table 2.3.1 B: *Bos* sp. accession numbers used in the meta-analysis (cont.)

JN817351, EU177848, JN817338, JN817329, GU947020, EU177847, JN817331, KF163082, KF163093, AY676867, AY676870, JN817317, JN817340, JN817316, JN817347, JN817308, JN817333, JN817337, JN817342, JN817321, JN817312, KT184466, JN817339, KT184463, KT184464, JN817318, EU177841, JN817306, JN817314, EU177843, EU177846, EU177845, JN817344, JN817345, JN817335, JN817315, JN817332, KT184462, KT184465, EU177844, JN817313, JN817341, JN817323, GU947021, EU177863, EU177862, EU177864, EU177865, MG736676, JN817307, KT184453, DQ124382, DQ124402, DQ124380, DQ124388, DQ124394, DQ124398, DQ124381, AY676860, EU177818, AB074965, KT184455, EU177827, GU947007, GU947011, GU947013, GU947016, GU947017, EU177826, AY676864, JN817322, JN817309, JN817310, JN817311, GU947012, GU947015, GU947008, GU947014, GU947009, GU947010, GU947018, EU177820, DQ124376, AY676862, EU177828, DQ124374, AB074966, AY526085, NC_006853, JN817346, KT184452, EU177823, EU177836, EU177837, DQ124385, DQ124390, KJ709686, KJ709683, KT343749, AY676858, JN817326, AY676865, KC153975, KC153973, KJ789953, KF926377, EU177822, DQ124407, DQ124415, JN817325, DQ124404, DQ124405, EU177816, DQ124372, AY676868, AY676869, EU177829, DQ124386, KC153974, GU947019, JN817328, EU177833, EU177834, DQ124410, EU177839, KU891849, EU177830, JN817327, DQ124377, KJ709685, KJ709682, EU177831, EU177815, EU177817, CM008198, EU177840, DQ124409, KC153977, AY676873, AY676872, EU177819, KT184454, DQ124413, AY676866, KT184451, EU177824, AY676855, DQ124414, JN817324, KT184470, GQ129207, DQ124375, DQ124400, KY766256, AY676857, AY676859, AY676861, AY676871, AB074967, AB074968, AB074962, AB074964, AB074963, KC153976, DQ124417, DQ124418, DQ124401, DQ124412, DQ124392, KY766258, EU177838, MG837552, JQ967333, DQ124408, DQ124378, DQ124397, DQ124391, DQ124387, DQ124395, EU177821, KJ709684, KT343748, AY676863, DQ124371, DQ124406, DQ124416, KY766257, GQ129208, DQ124411, JN817319, MF169214, HM045018, KJ709681, KC153972, EU177835, EU177825, EU177832, DQ124384, DQ124373, DQ124379

Supplementary Table 2.3.2: *Caenorhabditis* sp. accession numbers for the meta-analysis

KY552899, NC_035243, KR709159, KY552904, NC_035247, KY552906, NC_035249, KY552900, NC_035244, CP038193, CM003212, NC_001328, KY552907, NC_035250, KY552908, NC_035251, KY552910, NC_035253, KY552902, NC_035246, KM403565, NC_025756, KY552905, NC_035248, KY552909, 035252, KY552901, NC_035245

Supplementary Table 2.3.3: *Chelodina* sp. accession numbers used in the meta-analysis

Organism name	Accession Number
<i>Chelodina oblonga</i>	KY705234
<i>Chelodina oblonga</i>	NC_037387
<i>Chelodina oblonga</i>	KY776451
<i>Chelodina parkeri</i>	KY705232
<i>Chelodina parkeri</i>	NC_037385

Supplementary Table 2.3.4: *Chlamydosaurus* sp. accessions used in the meta-analysis

Organism name	Accession Number
<i>Chlamydosaurus kingii</i>	EF090422
	EF090423
	EF090421
	NC_009421

Supplementary Table 2.3.5: *Variabilichromis moorii* accession numbers for meta-analysis

JF825855, JF825854, JF825853, JF825852, JF825851, JF825850, JF825848, JF825847, JF825846, JF825844, JF825842, JF825841, JF825840, JF825839, JF825843, JF825838, JF825837, JF825836, JF825835, JF825834, JF825832, JF825830, JF825829, JF825828, JF825827, JF825826, JF825825, JF825824, JF825823, JF825822, JF825821, JF825820, JF825819, JF825818, JF825817, JF825815, JF825814, JF825813, JF825811, JF825810, JF825809, JF825808, JF825807, JF825806, JF825849, JF825845, JF825833, JF825831, JF825816, JF825812, JF825805, JF825803, JF825802, JF825801, JF825797, JF825796, JF825795, JF825785, JF825784, JF825779, JF825777, JF825800, JF825792, JF825804, JF825798, JF825794, JF825791, JF825790, JF825789, JF825788, JF825787, JF825786, JF825783, JF825782, JF825780, JF825778, JF825776, JF825775, JF825774, JF825772, JF825771, JF825770, JF825769, JF825767, JF825766, JF825764, JF825761, JF825760, JF825759, JF825758, JF825757, JF825756, JF825754, JF825753, JF825752, JF825751, JF825750, JF825749, JF825748, JF825747, JF825745, JF825744, JF825743, JF825742, JF825741, JF825740, JF825739, JF825737, JF825799, JF825793, JF825781, JF825773, JF825768, JF825763, JF825755, JF825746, JF825765, JF825762, JF825738, JF825736, JF825734, JF825733, JF825726, JF825725, JF825717, JF825716, JF825714, JF825709, JF825707, JF825705, JF825703, JF825735, JF825732, JF825727, JF825721, JF825719, JF825713, JF825711, JF825710, JF825706, JF825704, JF825702, JF825698, JF825731, JF825728, JF825724, JF825723, JF825715, JF825712, JF825700, JF825718, JF825722, JF825720, JF825708, JF825701, JF825699, JF825730, JF825729, JF825697

Supplementary Table 2.3.6: *Latimeria* sp. accession numbers used in the meta-analysis

Organism name	Accession Number
<i>Latimeria chalumnae</i> (from Mtwara region, Tanzania)	AP012192
<i>Latimeria chalumnae</i> (from Tanga region, Tanzania)	AP012182
<i>Latimeria chalumnae</i> (from Tanga region, Tanzania)	AP012194
<i>Latimeria chalumnae</i> (from Tanga region, Tanzania)	AP012183
<i>Latimeria chalumnae</i> (from Mtwara region, Tanzania)	AP012191
<i>Latimeria chalumnae</i> (from Lindi region, Tanzania)	AP012181
<i>Latimeria chalumnae</i> (from Kilwa, Tanzania)	AB257296
<i>Latimeria chalumnae</i> (Lindi region, Tanzania)	AP012190
<i>Latimeria chalumnae</i> (from Tanga Tanzania)	AP012178
<i>Latimeria chalumnae</i> (from Tanga region, Tanzania)	AB257297
<i>Latimeria chalumnae</i> (from Comoros)	AP012198
<i>Latimeria chalumnae</i> (from Comoros)	AP012199
<i>Latimeria menadoensis</i> (from Indonesia)	AP006858
<i>Latimeria menadoensis</i> (from Indonesia)	NC_006921
<i>Latimeria menadoensis</i> (from Indonesia)	GQ911586

Supplementary Table 2.3.7: *Crocodylus* sp. accession numbers used in the meta-analysis

Organism name	Accession Number
<i>Crocodylus acutus</i> (from America)	HM636894
<i>Crocodylus acutus</i> (from America)	NC_015647
<i>Crocodylus intermedius</i> (Orinoco)	HM636895
<i>Crocodylus intermedius</i> (Orinoco)	NC_015648
<i>Crocodylus moreletii</i> (Morelet)	HQ585889
<i>Crocodylus moreletii</i> (Morelet)	NC_015235
<i>Crocodylus rhombifer</i> (Cuba)	JX292787
<i>Crocodylus rhombifer</i> (Cuba)	NC_024513
<i>Crocodylus niloticus</i> (Nile)	NC_008142
<i>Crocodylus niloticus</i> (Nile)	DQ273697
<i>Crocodylus mindorensis</i> (Philippine)	GU144287
<i>Crocodylus mindorensis</i> (Philippine)	NC_014670
<i>Crocodylus novaeguineae</i> (New Guinea)	HM636896
<i>Crocodylus novaeguineae</i> (New Guinea)	NC_015651
<i>Crocodylus palustris</i> (marsh)	GU144286
<i>Crocodylus palustris</i> (marsh)	NC_014706
<i>Crocodylus palustris</i> (marsh)	HM488007
<i>Crocodylus siamensis</i> (Siamese)	DQ353946
<i>Crocodylus siamensis</i> (Siamese)	NC_008795
<i>Crocodylus porosus</i> (Saltwater, Australia)	DQ273698
<i>Crocodylus porosus</i> (Saltwater, Australia)	NC_008143
<i>Crocodylus johnsoni</i> (Australia)	HM488008
<i>Crocodylus johnsoni</i> (Australia)	NC_015238

Supplementary Table 2.3.8: *Drosophila* sp. accession numbers used in the meta-analysis

JQ691660, AF200839, AF200840, KC244283, AF200844, AF200841, AF200845, AF200843, AF200842, KC244284, AY518674, AY518670, AY518673, AY518672, AY518671, AF200849, AF200847, AF200846, JQ691661, AF200854, AF200853, AF200852, AF200851, AF200850, AF200848, AF200835, AF200833, NC_005781, AF200836, AF200834, AF200837, AF200838, KP161877, KJ947872, NC_024511, KT896664, KY310613, KP843845, KP843850, KP843852, JQ686694, JQ686693, KP843843, DMU37541, KT174472, KY310614, KU764535, KP843844, KT174473, KP843848, KP843853, KP843849, KP843854, AF200828, KP843842, KP843846, KP843847, KY310615, KP843851, KT174474, KF824892, KF824888, KF824894, KF824901, KF824895, KF824899, KF824880, KF824881, KF824878, KF824889, KF824891, KF824883, KF824890, KF824898, KF824893, KF824900, KF824874, KF824887, KF824884, KF824877, KF824896, KF824875, KF824897, KF824882, KF824886, KF824876, KF824879, KF824873, KF824885, NC_001322

Supplementary Table 2.3.9: *Entomyzon cyanotis* accession numbers used for meta-analysis

MK277224, MK277225, MK277226, MK277227, MK277223, MK277228, AY488272, HM230587, KY382844, KY382845, KY382846, KY382847, KY382848, HM230583, HM230584, HM230586, HM230588, HM230592, KY382839, KY382841, KY382840, KY382849, KY382850, KY382842, KY382843, KY382838, MK277218, MK277219, MK277220, MK277221, MK277222, HM230585, HM230589, HM230590, HM230591
--

Supplementary Table 2.3.10: *Mus* sp. accession numbers used in the meta-analysis

CM004179, KF937875, KF937874, KF937873, JQ003190, HQ586004, KR020497, AY999076, GQ871745, GQ871744, FJ374665, FJ374663, FJ374660, FJ374659, FJ374657, FJ374655, FJ374653, FJ374652, FJ374651, FJ374648, FJ374647, FJ374642, FJ374640, EF108341, EF108337, EF108334, EF108332, NC_006914, EF108330, FJ374643, JX945967, JX945966, CM004189, JX945972, EF108344, HQ675031, JX945968, CM004276, FJ374654, JX945979, EU312160, HQ675030, JX945974, JX945970, JX945969, AY533105, FJ374641, HQ877407, DQ106412, AY172335, FJ374661, EF108336, NC_005089, FJ374664, FJ374656, CM004214, FJ374644, JX945976, JX945973, HQ675029, HQ675028, HQ675027, HQ675026, CM004184, GQ871746, EF108338, AY339599, CM004255, EU450583, JX945964, FJ374645, EU315229, EU315228, JX945978, JX945971, EU312161, CM004185, CM004182, CM004180, AY533108, AY533106, AY466499, FJ803909, FJ374662, FJ374658, KP090294, EF108335, DQ106413, EF108339, FJ374639, AY533107, JX945977, JX945975, JF286601, JX945965, D83491, FJ374650, CM004277, KC663622, CM004183, CM004178, KC663618, FJ374649, FJ374646, EF108340, EF108333, EF108331, KC663619, KP260514, KP877616, KP877611, KP877619, KP877620, KP877615, KP877618, KP877613, KP877612, KP877610, KP877617, KP877614, CM004186, CM004181, KF781657, EF108342, KF781659, NC_012387, KF781658, KF781661, KF781660, KF781662, KF781664, KF781663, KF781649, KF781648, KF781650, KF781647, KF781646, KF781651, CM004187, DQ874614, NC_010339, EF108343, KC663621, KF781645, KF781652, KF781653, KF781654, KF781656, KF781655,

Supplementary Table 2.3.11: *Apodemus* sp. accession numbers used in the meta-analysis

KY851926, KY851916, KY851935, KY851929, KY851921, KY851906, KY851925, KY851904, KY851900, KY851919, KY851895, KY851936, KY851933, KY851912, KY851930, KY851899, KY851922, KY851913, KY851902, KY851934, KY851927, KY851901, KY851959, KY851914, KY851910, KY851896, KY851924, KY851915, KY851903, KY851920, KY851911, KY851909, KY851897, KY851907, KY851905, KY851932, KY851961, KY851908, KY851923, KY851928, KY851898, KY851946, KY851943, KY851937, KY851931, KY851950, KY851894, KY851892, KY851957, KY851952, KY851953, KY851948, KY851941, KY851939, KY851955, KY851918, KY851917, KY851958, KY851942, KY851960, KY851947, KY851956, KY851944, KY851893, KY851945, KY851949, KY851954, KY851940, KY851951, KY851938, JN629047, NC_016428, HM034866

Supplementary Table 2.3.12: *Onchocerca* sp. accession numbers used in the meta-analysis

Organism name	Accession Number
<i>Onchocerca volvulus</i> (Forest strain)	NC_001861
<i>Onchocerca volvulus</i> (Forest strain)	AF015193
<i>Onchocerca volvulus</i> (Brazil)	KT599912
<i>Onchocerca ochengi</i> (Ngaoundere Northern, Cameroon)	KX181289
<i>Onchocerca ochengi</i> (Ngaoundere Northern, Cameroon)	NC_031891
<i>Onchocerca ochengi</i> (Ngaoundere Northern, Cameroon)	KX181290

Supplementary Table 2.3.13: *Oreochromis* sp. accession numbers used in the meta-analysis

<i>Oreochromis niloticus</i> (from Myanmar)	LC189956
<i>Oreochromis niloticus</i> (from Myanmar)	LC189955
<i>Oreochromis niloticus</i> (from Myanmar)	LC189954
<i>Oreochromis niloticus</i> (from Myanmar)	LC189953
<i>Oreochromis niloticus</i> (from Myanmar)	LC189952
<i>Oreochromis niloticus</i> (from Myanmar)	LC189951
<i>Oreochromis niloticus</i> (from Myanmar)	LC189950
<i>Oreochromis niloticus</i> (from Myanmar)	LC189949
<i>Oreochromis niloticus</i> (from Myanmar)	LC189948
<i>Oreochromis niloticus</i> (from Myanmar)	LC189946
<i>Oreochromis niloticus</i> (from Myanmar)	LC189945
<i>Oreochromis niloticus</i> (from Myanmar)	LC189944
<i>Oreochromis niloticus</i> (from Myanmar)	LC189943
<i>Oreochromis niloticus</i> (from Myanmar)	LC189942
<i>Oreochromis niloticus</i> (from Myanmar)	LC189957
<i>Oreochromis niloticus</i> (from Myanmar)	LC189947
<i>Oreochromis mossambicus</i> (from Australia)	KF228010
<i>Oreochromis mossambicus</i> (from Australia)	KF228009
<i>Oreochromis mossambicus</i> (from Australia)	KF228008
<i>Oreochromis mossambicus</i> (from Australia)	KF228007
<i>Oreochromis mossambicus</i> (from Australia)	KF228006
<i>Oreochromis mossambicus</i> (from Australia)	KF228005
<i>Oreochromis aureus</i> from (from Australia)	MF817697
<i>Oreochromis aureus</i> from (from Australia)	MF817698
<i>Oreochromis aureus</i> from (from Australia)	MF817699
<i>Oreochromis aureus</i> from (from Australia)	MF817700
<i>Oreochromis aureus</i> from (from Australia)	MF817701
<i>Oreochromis aureus</i> from (from Australia)	MF817702
<i>Oreochromis niloticus</i> (from DRC)	KT193329
<i>Oreochromis niloticus</i> (from DRC)	KT193539
<i>Oreochromis niloticus</i> (from DRC)	KT193494
<i>Oreochromis leucostictus</i> (from DRC)	KT193493
<i>Oreochromis leucostictus</i> (from DRC)	KT193495

Supplementary Table 2.3.14A: *Simulium* sp. CO1 accession numbers used for meta-analysis

KJ636862, KJ636874, KJ636846, KJ636848, KJ636878, KJ636885, KJ636897, KJ636849, KJ636850, KJ636871, KJ636857, KJ636858, KJ636875, KJ636907, KJ636889, KJ636863, KJ636898, KJ636864, KJ636867, KJ636869, KJ636880, KJ636881, KJ636886, KJ636903, KJ636883, KJ636909, KJ636888, KJ636845, KJ636847, KJ636853, KJ636870, KJ636877, KJ636887, KJ636884, KJ636908, KJ636911, KJ636913, KJ636894, KJ636873, KJ636912, KJ636872, KJ636879, KJ636892, KJ636855, KJ636882, KJ636902, KJ636891, KJ636899, KJ636904, KJ636893, KJ636856, KJ636868, KJ636901, KJ636865, KJ636852, KJ636860, KJ636890, KJ636859, KJ636905, KJ636906, KJ636861, KJ636896, KJ636910, KJ636895, KJ636914, KJ636915, KJ636851, KJ636854, KJ636866, KJ636900, KJ636876, AY251525, AY251526, AY251524, KY497566, KY497569, KY497565, KY497567, KY497563, KY497568, KY497570, DQ534949, AY251521, AY251520, AY251518, AY251519, DQ534946, DQ534947, KY497564, DQ534950, AY251522, AY251523, AY251527, AY251512, AY251513, AY251511, AY251517, AY251514, AY251508, AY251509, AY251530, AY251510, AY251504, AY251515, AY251491, AY251492, AY251493, AY251494, AY251498, AY251499, AY251500, AY251501, AY251487, AY251488, AY251489, AY251502, AY251503, AY251490, AY251485, AY251486, AY251483, AY251484, AY251497, AY251516, AY251528, AY251495, AY251496, AY251505, AY251506, AY251507, AY251529

Supplementary Table 2.3.14B: *Simulium* sp. CO1 accession numbers used for meta-analysis

EU025957, EU025961, EU025962, EU025956, EU025958, EU025959, EU025954, EU025964, EU025968, EU025963, EU025969, EU025948, EU025960, EU025966, EU025967, EU025950, EU025955, EU025965, EU025970, EU025942, EU025945, EU025947, EU025949, EU025946, EU025951, EU025952, EU025953, EU025943, EU025944, KX673602, KX673603, KX673601, KX689258, KX673600, KX673590, KX673594, KX673593, KX673592, KX673591, KX673597, KX673595, KX673596, KX673598, KX673582, KX673599, EU025971, JQ034312, KF155177, KF990256, JX441377, KF990249, KX759638, KX759641, JQ034311, KF155172, KX759640, KF990254, JX441381, JQ034310, KF155169, KF990252, JQ030883, KF155170, KF990258, KF990250, JQ034313, KF155168, KF990259, KF990257, JX441379, KF155176, KX759639, JQ034309, KF155175, KF990253, KF990263, KF990251, JQ034308, KF155171, JX441376, JX444910, GU203464, GU203462, GU203458, GU203463, GU203459, GU203460, KF990255, JX441380, KF155173, KF155174, KX673584, KX673585, KX673587, KX673589, KX673588, KX673586, GU203467, GU203470, GU203468, GU203473, GU203471, GU203466, GU203469, GU203474, GU203472, KX759633, KX759635, KX759634, KX673580, KX673579, KX673578, KX673581, KJ649638, KJ649643, KJ649645, KJ649639, KJ649640, KJ649641, KJ649642, KJ649644, KY751929, KY751930, MF101842, MF101843, MF101844, MF101845, KY751928, MF101846, KJ649623, KJ649624, KJ649625, KJ649626, KJ649627, KJ649628, KJ649629, KJ649630, KJ649631, KJ649632, KJ649633, KJ649634, KJ649635, KJ649636, KJ649637, KJ649646, KJ649648, KJ649649, KJ649650, KJ649652, KJ649651, KJ649647, JQ663455, JQ663446, JQ663456, JQ663457, JQ663447, JQ663458, JQ663459, JQ663460, JQ663461, JQ663462, JQ663464, JQ663465, JQ663467, JQ663469, JQ663470, JQ663474, JQ663471, JQ663473, JQ663475, LC034966, LC034967, LC034954, LC034962, LC034968, LC034969, LC034970, LC034971, LC034972, LC034964, LC034965, LC034955, LC034956, LC034957, LC034958, LC034973, LC034974, LC034959, LC034960, LC034961, LC034963, LC034978, LC034979, LC034975, LC034980, LC034976, LC034977, LC034981, LC034982, JQ663480, JQ663481, JQ663482, JQ663483, JQ663484, JQ663485, JQ663486, JQ663487, JQ663488, JQ663489, JQ663490, JQ663491, JQ663492, JQ663493, JQ663495, JQ663494, JQ663496, JQ663498
--

Supplementary Table 2.3.15A: *Simulium* sp. CytB accession numbers used for analysis

KJ663153, KJ663254, KJ663247, KJ663124, KJ663253, KM498070, KJ663244, KJ663243, KJ663195, KJ663239, KJ663240, KJ663248, KJ663256, KJ663237, KJ663191, KJ663189, KJ663109, KJ663233, KJ663230, KJ663197, KJ663228, KJ663236, KJ663238, KJ663241, KJ663194, KJ663113, KJ663252, KJ663187, KJ663138, KJ663114, KJ663115, KJ663145, KJ663152, KJ663116, KJ663118, KJ663122, KM498063, KJ663112, KJ663193, KJ663188, KM498067, KM498065, KJ663110, KM498060, KM498062, KJ663147, KJ663251, KJ663137, KJ663196, KJ663136, KJ663150, KJ663184, KM498059, KJ663245, KJ663135, KJ663121, KM498061, KJ663117, KJ663119, KJ663120, KJ663125, KJ663127, KJ663130, KJ663131, KJ663133, KJ663144, KJ663235, KM498068, KJ663141, KM498058, KJ663111, KJ663185, KJ663257, KM498066, KJ663190, KJ663123, KM498064, KJ663255, KJ663246, KJ663242, KJ663231, KJ663229, KJ663129, KJ663140, KJ663142, KJ663146, KJ663149, KJ663151, KM498072, KM498071, KM498069, KJ663234, KJ663250, KJ663232, KJ663143, KJ663192, KJ663148, KJ663134, KJ663126, KJ663249, KJ663132, KJ663139, KJ663186, KJ663128, KJ663212, KJ663211, KJ663210, KJ663208, KJ663207, KJ663205, KJ663204, KJ663203, KJ663200, KJ663199, KJ663198, KJ663168, KJ663165, KJ663158, KJ663157, KJ663156, KJ663108, KJ663107, KJ663106, KJ663103, KJ663098, KJ663096, KJ663094, KJ663078, KJ663076, KJ663073, KJ663071, KJ663067, KJ663066, KJ663064, KM497741, KM497739, KM497738, KM497736, KM497734, KM498042, KM498040, KM498039, KM498034, KM498030, KJ663183, KJ663182, KJ663180, KJ663179, KJ663175, KM497755, KM497753, KM497744, KM497743, KM497742, KJ663093, KJ663092, KJ663090, KJ663089, KM497726, KM497724, KJ663088, KM497722, KM497723, KJ663086, KM497727, KJ663083, KJ663079, KM497812, KM497811, KM497809, KM498083, KM498082, KM498079, KM498073, KM498056, KM498055, KM498052, KM498051, KM498050, KM498049, KM498045, KM498043, KJ663097, KJ663065, KM498041, KM497810, KM497800, KM497783, KM497771, KM497759, KM497769, KM497767, KM498081, KM498074, KJ663164, KJ663163, KJ663077, KM498028, KJ663177, KJ663173, KJ663172, KJ663170, KJ663169, KJ663091, KJ663081, KM497780, KM497772, KM497766, KM497765, KM497764, KM497752
--

Supplementary Table 2.3.15B: *Simulium* sp. CytB accession numbers used for analysis

KM497750, KM497749, KM497728, KJ663082, KM497725, KM497721, KM497720, KM497719, KM497718, KM498076, KM498053, KM498044, KJ663176, KJ663085, KM497781, KM497778, KM497777, KJ663160, KJ663209, KJ663201, KJ663166, KM497737, KM498037, KM498035, KJ663155, KJ663104, KJ663102, KJ663100, KJ663099, KJ663095, KJ663075, KJ663072, KJ663068, KM497735, KM498036, KM497733, KM498032, KM497732, KM498031, KM497758, KM497754, KJ663181, KM497763, KM497768, KM497760, KM497770, KM497761, KM498078, KM497802, KM498077, KM498048, KJ663087, KJ663084, KJ663080, KM497757, KM497756, KJ663178, KM498084, KM498057, KM498054, KM498080, KJ663202, KJ663159, KM497731, KM498038, KM498029, KM497807, KM497808, KM497806, KM497805, KM497801, KM497798, KM497779, KM498087, KM498086, KM498085, KM498075, KM498047, KM498046, KM497799, KJ663162, KJ663206, KJ663070, KJ663069, KM497740, KJ663167, KJ663154, KJ663161, KJ663105, KJ663101, KJ663074, KM498033, KM497729, KM497730, KM497748, KM497773, KJ663227, KJ663226, KJ663225, KJ663224, KJ663223, KJ663222, KJ663221, KJ663220, KJ663219, KJ663217, KJ663216, KJ663215, KJ663214, KM497797, KM497796, KM497795, KM497794, KM497792, KM497791, KM497790, KM497789, KM497787, KM497786, KM497785, KM497784, KM497782, KM497776, KM497775, KM497774, KM497747, KM497746, KJ663171, KM497745, KJ663174, KM497751, KJ663218, KM497793, KJ663213, KM497788, KM497804, KM497803, KJ663258, JX500928, JX500895, JX500889, JX500799, JX500797, JX500795, JX500794, JX500788, JX500786, JX500782, JX500761, JX500781, JX500780, JX500779, JX500777, JX500775, JX500773, JX500768, JX500767, JX500765, JX500764, JX500763, JX500762, JX500792, JX500770, JX500796, JX500766

Supplementary Table 2.3.16: *Simulium* sp. ND4 accession numbers used for meta-analysis

AY055441, AY055440, AY055439, AY055438, AY055437, AY055436, AY055435, AY055433, AY055432, AY055431, AY055430, AY055429, AY055442, AY055434, AY055444, AY055448, AY055445, AY055446, AY055447, AY055450, AY077711, AY885646, AY722106, AY722107, AY722105, AY722104, AY722110, AY722111, AY722109, AY722108, AF213337, AY055449

Supplementary Table 2.3.17: *Simulium* sp. 16S accession numbers used for meta-analysis

AF081904, AF081905, AF081906, AF416955, AF416956, AF416957, AF416958, AF416959, AF416960, AF466182, AF466183, AY885648, AF081908, AF081907, AF416961, DQ655644, DQ655645, DQ655646, DQ655647, DQ869242, DQ869243, AF416964, AF416965, AF416966, AF416967, AF416968, AF416969, AY722116, AY722117, AY722114, AY722115, AF416970, AF466180, AF466181, AF127463, AY722113, AF466186, DQ022373, AF081909, AF466184, AF466185, AF416971, AF416972, AY077709, AF213336, AF416975, AF416976, AF416977, AF416978, AF416979, AY722112, AF466187, AF127461, AF127462, AF081903

Supplementary Table 2.3.18: *Simulium* sp. ITS2 accession numbers used for meta-analysis

AY625899, AY625902, AY625921, AY625922, AY625923, AY694177, AY625900, AY625938, AY625955, AY625956, AY625901, AY625903, AY625960, AY625957, AY625959, AY625958, AY694180, AY625945, AY625952, AY625947, AY625946, AY625948, AY694179, AY694178, AY885647, AY625954, AY625953, AY625950, AY625936, AY625904, AY625934, AY625935, AY625937, AY625939, AY625940, AY625941, AY625942, AY625931, AY625909, AY625920, AY625906, AY625907, AY625908, AY625912, AY625914, DQ655649, DQ655650, DQ655651, AY625926, DQ022374, DQ869241, AY694182, AY694176, AY625929, AY625930, AY694181, DQ869240, AY625928, AY694175, AY694174, AY625924, AY694173, AY625916, AY625915, AY625919, AY625917, AY625918, FJ538880, FJ538881, FJ538885, FJ538888, FJ538889, FJ538886, FJ538887, FJ538879, FJ538890, FJ538884, FJ538882, FJ538883, FJ538878

Supplementary Table 2.3.19: *Simulium* sp. IGS accession numbers used for meta-analysis

Organism name	Accession number		
<i>Simulium sirbanum</i> (from Mali)	AF421598,	AF421599,	AF421586,
	AF421587,	AF421588,	AF421589,
	AF421595,	AF421596,	AF421594,
	AF421600,	AF421582,	AF421601,
	AF421597		
<i>Simulium soubrense</i> (from Liberia)	AF421608,	AF421618,	AF421602,
	AF421610,	AF421614,	AF421603,
	AF421605,	AF421619,	AF421606,
	AF421613		
<i>Simulium squamosum</i> (from Ghana)	AF421567,	AF421581,	AF421623,
	AF421622,	AF421580,	AF421575,
	AF421574,	AF421568,	AF421566,
	AF421570,	AF421572,	AF421573,
	AF421576,	AF421624,	AF421577,
	AF421579, AF421578		

Supplementary Table 4.3.2A: Reads summary

Sample Name	Sequence Type	Total Reads	Assembled Reads
SIM-WiT1	NextSeq2	1,870,762	16,407
SIM-WiT2	NextSeq2	14,330,426	84,991
SIM-WiT3	NextSeq2	3,790,902	26,381
SIM-WiT4	NextSeq2	3,448,592	29,090
SIM-WiT5	NextSeq2	3,851,994	30,944
SIM-WiT6	NextSeq2	4,745,176	18,483
SIM-WiT7	NextSeq2	3,380,762	7,037
SIM-WiT9	NextSeq2	4,209,502	15,860
SIM-WiT10	NextSeq2	3,371,642	9,647
SIM-WiT11	NextSeq2	2,949,740	12,228
SIM-WiT12	NextSeq2	3,392,570	24,347
SIM-WiT14	NextSeq2	3,139,782	12,085
SIM-WiT15	NextSeq2	3,994,810	20,357
SIM-WiT16	NextSeq2	2,788,674	16,238
SIM-WiT17	NextSeq2	3,492,358	16,335
SIM-WiT18	NextSeq2	4,289,508	24,854
SIM-WiT19	NextSeq2	2,722,304	17,930
SIM-WiT20	NextSeq2	3,125,878	19,867
SIM-WiT21	NextSeq2	3,009,718	23,963
SIM-WiT22	NextSeq2	3,137,974	12,270
SIM-WiT23	NextSeq2	2,928,730	4,270
SIM-WiT24	NextSeq2	2,779,262	10,466
SIM-WiT25	NextSeq2	3,576,940	17,283
SIM-BeT26	NextSeq2	3,715,246	2,742
SIM-Sen27	NextSeq2	3,972,144	10,484
SIM-FaT28	NextSeq2	4,137,766	4,276
SIM-BA29	NextSeq2	2,651,412	10,852
SIM-OH30	NextSeq2	4,492,386	13,862
SIM-AsT31	NextSeq2	2,400,150	4,068
SIM-NL32	NextSeq2	2,929,110	1,046
SIM-NL33	NextSeq2	2,566,004	1,304
SIM-NL34	NextSeq2	3,115,952	2,588
SIM-NL35	NextSeq2	3,856,848	5,283
SIM-NL36	NextSeq2	2,936,856	1,937
SIM-NL37	NextSeq2	3,585,650	2,069
SIM-NL38	NextSeq2	3,653,600	1,673
SIM-TI39	NextSeq2	3,575,800	2,632
SIM-AB40	NextSeq2	3,117,158	10,629
SIM-TN41	NextSeq2	3,533,186	1,624
SIM-TN42	NextSeq2	3,793,670	21,508
SIM-TN43	NextSeq2	3,912,978	10,818

Supplementary Table 4.3.2B: Additional reads summary

Sample Name	Sequence Type	Total Reads	Assembled Reads
SIM-TN44	NextSeq2	5,370,244	5,300
SIM-TN45	NextSeq2	5,241,994	8,713
SIM-TN46	NextSeq2	3,722,270	6,284
SIM-TN47	NextSeq2	2,989,624	1,135
SIM-TN48	NextSeq2	5,057,014	4,917
EG1	MiSeq2	2,233,440	6,552
EG2	MiSeq2	2,226,640	11,133
EG3	MiSeq2	2,109,848	9,624
EG4	MiSeq2	2,578,700	17,885
EG5	MiSeq2	2,675,588	6,222
EG6	MiSeq2	2,220,342	13,777
EG7	MiSeq2	2,553,072	18,515
EG8	MiSeq2	2,817,006	17,895
EG9	MiSeq2	2,298,008	4,907
EG10	MiSeq2	2,251,608	7,061
B7	MiSeq1	17,158,958	42,512
D3	MiSeq1	20,151,814	28,542
SIM1	NextSeq1	38,585,502	98,155
SIM2	NextSeq1	32,976,666	100,910
SIM3	NextSeq1	4,415,734	7,314
SIM4	NextSeq1	2,961,840	4,691
SIM5	NextSeq1	3,760,108	9,041
SIM6	NextSeq1	3,423,700	3,551
SIM7	NextSeq1	3,765,150	5,626
SIM8	NextSeq1	4,106,142	13,205
SIM10	NextSeq1	3,871,376	5,011
SIM11	NextSeq1	4,224,460	8,762
SIM12	NextSeq1	3,995,350	14,087
SIM13	NextSeq1	4,582,158	12,240
SIM14	NextSeq1	7,132,364	20,268
SIM15	NextSeq1	9,868	0
SIM16	NextSeq1	3,764,628	11,782
SIM17	NextSeq1	4,142,208	16,909
SIM18	NextSeq1	3,288,626	11,349
SIM20	NextSeq1	4,343,544	15,868
SIM21	NextSeq1	3,304,590	9,504
SIM22	NextSeq1	4,027,174	13,695
SIM23	NextSeq1	3,433,602	13,288
SIM24	NextSeq1	3,424,762	18,447
SIM25	NextSeq1	3,728,358	7,184
SIM26	NextSeq1	3,153,310	6,045

Supplementary Table 4.3.3: Trimmed reads summary of whole mitochondrial genomes

Sample Name	Percentage trimmed	Average length before trim	Average length after trim
B7	99.39%	165.8	160.4
D3	99.43%	170.7	164.5
EG5	99.56%	181.3	177.4
EG2	99.56%	182.2	178
EG4	99.62%	189	184.9
EG1	99.64%	195.7	188.5
EG3	99.68%	196.7	191.1
EG9	99.71%	199.6	194.3
EG8	99.72%	181.5	178
EG7	99.74%	184.2	180.2
EG6	99.74%	196.9	191.7
EG10	99.81%	196.4	191.4
SIM2	99.95%	148.9	148.9
SIM3	99.96%	148.8	148.8
SIM1	99.96%	148.8	148.9
SIM4	99.96%	149.1	149.1
SIM27	99.96%	149.5	149.5

The table shows the read summary of some of the individual blackfly sequences

Supplementary Table 4.3.4 A: Length and weight statistics of *Simulium* whole mitochondrial genome sequences from Africa, Asia and Europe

Info	Pru	Black V	Daka	MbaM	NkaM	MA	SM	SV	SA
Leng th (bp)	16,113	16,114	16,126	16,095	16,092	16,111	15,799	15,367	15,904
WSS	4,963.3 9	4,963.9 79	4,967.5 51	4,958.4 5	4,957.3 21	4,964.0 64	4,866.9 43	4,733.5 48	4,900.4 36
WDS	9,952.1 34	9,952.7 28	9,960.1 48	9,941.1 04	9,939.2 47	9,950.9 76	9,758.2 59	9,491.6 9	9,823.0 78
AT/ GC Ratio	3.0486	3.0816	3.0650	2.9683	2.9683	2.9683	2.9841	2.7037	2.9841

The table shows the length and weight information of African *Simulium* whole mitochondrial genomes compared with those from Asia and Europe. In the table; Info = Information, Pru = *Simulium* blackfly from Pru river basin in Ghana, BlackV = *Simulium* blackfly from Black Volta river basin in Ghana, Daka = *Simulium* blackfly from Daka river basin in Ghana, Mbam = *Simulium* blackfly from Mbam river basin in Cameroon, Nkam = *Simulium* blackfly from Nkam river basin in Cameroon, MA = *Simulium* blackfly from Lingoni Falls in Malawi, SM = *Simulium maculatum* (accession NC_040120.1), SV = *Simulium variegatum* (accession

NC_033348.1), SA = *Simulium aureohirtum* (accession NC_029753.1), WSS = Weight (single-stranded - kDa), and WDS = Weight (double-stranded - kDa).

Supplementary Table 4.3.4 B: One-Sample T Test of AT/GC Ratio of whole mitochondrial genome sequences from Africa, Asia and Europe

	Test Value = 3.0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
AT/GC Ratio	-0.684	8	0.513	-0.0253333	-0.110749	0.060082

Supplementary Table 4.3.5: Melting temperature statistics of *Simulium* whole mitochondrial genome sequences from Africa, Asia and Europe

[Salt]	Pru	BlackV	Daka	MbaM	NkaM	MA	SM	SV	SA
0.1M	74.99	74.93	74.96	75.22	75.21	75.19	75.17	75.95	75.17
0.2M	79.99	79.93	79.95	80.22	80.21	80.19	80.16	80.95	80.17
0.3M	82.91	82.85	82.88	83.14	83.13	83.11	83.09	83.87	83.09
0.4M	84.99	84.93	84.95	85.21	85.20	85.19	85.16	85.94	85.16
0.5M	86.60	86.53	86.56	86.82	86.81	86.80	86.77	87.55	86.77

The table shows the melting temperature statistics of African *Simulium* whole mitochondrial genomes compared with those from Asia and Europe. In the table; Pru = *Simulium* blackfly from Pru river basin in Ghana, BlackV = *Simulium* blackfly from Black Volta river basin in Ghana, Daka = *Simulium* blackfly from Daka river basin in Ghana, Mbam = *Simulium* blackfly from Mbam river basin in Cameroon, NkaM = *Simulium* blackfly from NkaM river basin in Cameroon, MA = *Simulium* blackfly from Lingoni Falls in Malawi, SM = *Simulium maculatum* (accession NC_040120.1), SV = *Simulium variegatum* (accession NC_033348.1), and SA = *Simulium aureohirtum* (accession NC_029753.1)

Supplementary Table 4.3.6: Counts of nucleotides of *Simulium* whole mitochondrial genome sequences from Africa, Asia and Europe

Nucleotide	Pru	BlackV	Daka	MbaM	NkaM	MA	SM	SV	SA
Adenine (A)	6,248	6,253	6,251	6,212	6,224	6,196	6,051	5,783	6,223
Cytosine (C)	2,347	2,326	2,335	2,378	2,383	2,352	2,319	2,424	2,336
Guanine (G)	1,632	1,629	1,632	1,685	1,675	1,705	1,649	1,730	1,660
Thymine (T)	5,886	5,906	5,908	5,820	5,810	5,858	5,780	5,430	5,685
C + G	3,979	3,955	3,967	4,063	4,058	4,057	3,968	4,154	3,996

The table shows the nucleotide counts of African *Simulium* whole mitochondrial genomes compared with those from Asia and Europe. In the table; Pru = *Simulium* blackfly from Pru river basin in Ghana, BlackV = *Simulium* blackfly from Black Volta river basin in Ghana, Daka = *Simulium* blackfly from Daka river basin in Ghana, MbaM = *Simulium* blackfly from MbaM river basin in Cameroon, Nkam = *Simulium* blackfly from Nkam river basin in Cameroon, MA = *Simulium* blackfly from Lingoni Falls in Malawi, SM = *Simulium maculatum* (accession NC_040120.1), SV = *Simulium variegatum* (accession NC_033348.1), and SA = *Simulium aureohirtum* (accession NC_029753.1)

Supplementary Table 4.3.7: Frequencies of nucleotides of *Simulium* whole mitochondrial genome sequences from Africa, Asia and Europe

Nucleotide	Pru	BlackV	Daka	MbaM	NkaM	MA	SM	SV	SA
Adenine (A)	0.388	0.388	0.388	0.386	0.387	0.385	0.383	0.376	0.391
Cytosine (C)	0.146	0.144	0.145	0.148	0.148	0.146	0.147	0.158	0.147
Guanine (G)	0.101	0.101	0.101	0.105	0.104	0.106	0.104	0.113	0.104
Thymine (T)	0.365	0.367	0.366	0.362	0.361	0.364	0.366	0.353	0.357
C + G	0.247	0.245	0.246	0.252	0.252	0.252	0.251	0.270	0.251
A + T	0.753	0.755	0.754	0.748	0.748	0.748	0.749	0.730	0.749

The table shows the nucleotide counts of African *Simulium* whole mitochondrial genomes compared with those from Asia and Europe. In the table; Pru = *Simulium* blackfly from Pru river basin in Ghana, BlackV = *Simulium* blackfly from Black Volta river basin in Ghana, Daka = *Simulium* blackfly from Daka river basin in Ghana, MbaM = *Simulium* blackfly from MbaM river basin in Cameroon, Nkam = *Simulium* blackfly from Nkam river basin in Cameroon, MA = *Simulium* blackfly from Lingoni Falls in Malawi, SM = *Simulium maculatum* (accession NC_040120.1), SV = *Simulium variegatum* (accession NC_033348.1), and SA = *Simulium aureohirtum* (accession NC_029753.1)

Supplementary Table 4.3.8 A: *Simulium* accession number used for nuclear genome work

U48383, U36210, U36209, U36208, U36207, U36206, U36205, U36204, KY288011, KY288010, KP710744, KP710743, KP710742, KP710741, KP710740, KP710739, KP710738, KP710737, KP710736, KP710735, KP710734, KP710733, KP710732, KP710731, KP710730, KP256364, KP256363, KM502581, KM502580, KM502579, KM410271, KM410270, KM410269, KM410268, KM410267, KM410266, KM410265, KM410264, KM410263, KM410262, KM410261, KM410260, KM410259, KM410258, KM410257, KM410256, KM410255, KM410254, KM410253, KM410252, KM410251, KM410250, KM410249, KM410248, KM410247, KM410246, KM410245, KM410244, KJ817817, KJ817816, KJ636959, KF550280, KF550279, KF550278, KF550277, KF550276, KF550275, KF550274, KF550273, KF550272, KF550271, KF550270, KF550269, KF550268, KF550267, KF550266, KF550265, KF550264, KF057788, KF057787, KF057786, KF057785, JQ793863, JQ793862, JQ793861, JQ793860, JQ793859, JQ793858, JQ793857, JQ793856, JQ793855, JQ793854, JQ793853, JQ793852, JQ793851, JQ793850, JQ793849, JQ793848, JQ793847, JQ793846, JQ793845, JQ793844, JQ793843, JQ793842, JQ793841, JQ673515, JQ673514, JQ673513, JQ673512, JQ673511, JQ673510, JQ673509, JQ673508, JQ673507, JQ673506, JQ673505, JQ673504, JQ673503, JQ673502, JQ673501, JQ673500, JQ673499, JQ285864, JQ285863, JQ285862, JQ285861, JQ285860, JQ285859, JN547777, JN547776, JN547775, JN547774, JN547773, JN547772, JN547771, JN547770, JN547769, JN547768, JN547767, JN547766, JN547765, JF916849, JF916848, JF916847, JF916846, JF916845, JF916844, JF916843, JF916842, JF916841, JF916840, JF916839, JF505387, HE577316, FJ538890, FJ538889, FJ538888, FJ538887, FJ538886, FJ538885, FJ538884, FJ538883, FJ538882, FJ538881, FJ538880, FJ538879, FJ538878, FJ538877, FJ538876, FJ538875, FJ538874, FJ538873, FJ538872, FJ538871, FJ538870, FJ538869, FJ538868, FJ538867, FJ538866, FJ437567, FJ437566, FJ437565, FJ437564, FJ436353, FJ436352, FJ436351, FJ436350, FJ436349, FJ436348, FJ231203, EU930244, EU779824, EU429933, EU429932, EU429931, EU429930, EU429929, EU429928, EU429927, EU429926, EU429925, EU429924, EU429923, EU429922, EU429921, EU429920, EU429919, EU429918, EU429917, EU429916, EU429915, EU429914, EU429913, EU429912, EU429911, EU429910, EU429909, EU429908, EU429907, EU429906, EU429905, EU429904, EU429903, EU429902, EU429901, EU429900, EU429899, EU429898, EU429897, EU429896, EU429895, EU429894, EU429893, EU429892, EU429891, EU429890, EU429889, EU429888, EU429887, EU429886, EU429885
--

Supplementary Table 4.3.8 B: *Simulium* sp. accession numbers used for nuclear genome work

EU429884, EU429883, EU429882, EU429881, EU429880, EU429879, EU429878, EU429877, EU429876, EU429875, EU429874, EU429873, EU429872, EU429871, EU429870, EU429869, EU429868, EU429867, EU429866, EU429865, EU429864, EU429863, EU429862, EU429861, EU429860, EU429859, EU429858, EU429857, EU429856, EU429855, EU429854, EU429853, EU429852, EU429851, EU429850, EU429849, EU429848, EU429847, EU429846, EU429845, EU429844, EU429843, EU429842, EU429841, EU429840, EU429839, EU429838, EU429837, EU429836, EU429835, EU429834, EU429833, EU429832, EU429831, EU429830, EU429829, EU429828, EU429827, EU429826, EU429825, EU429824, EU429823, EU429822, EU429821, EU429820, EU429819, EU429818, EU429817, EU429816, EU429815, EU429814, EU429813, EU429812, EU429811, EU429810, EU429809, EU429808, EU429807, EU429806, AY944195, AY944194, AY944193, AY944192, AY944191, AY594327, AY594326, AY594325, AY594324, AY594323, AY594322, AY594321, AY594320, AJ493253, AJ493252, AJ493251, AJ493250, AJ493249, AJ493248, AF427961, AF427960, AF427959, AF427958, AF421640, AF421639, AF421638, AF421637, AF421636, AF421635, AF421634, AF421633, AF421632, AF421631, AF421630, AF421629, AF421628, AF421627, AF421626, AF421625, AF421624, AF421623, AF421622, AF421621, AF421620, AF421619, AF421618, AF421617, AF421616, AF421615, AF421614, AF421613, AF421612, AF421611, AF421610, AF421609, AF421608, AF421607, AF421606, AF421605, AF421604, AF421603, AF421602, AF421601, AF421600, AF421599, AF421598, AF421597, AF421596, AF421595, AF421594, AF421593, AF421592, AF421591, AF421590, AF421589, AF421588, AF421587, AF421586, AF421585, AF421584, AF421583, AF421582, AF421581, AF421580, AF421579, AF421578, AF421577, AF421576, AF421575, AF421574, AF421573, AF421572, AF421571, AF421570, AF421569, AF421568, AF421567, AF421566, AF403825, AF403824, AF403823, AF403822, AF403821, AF403820, AF403819, AF403818, AF403817, AF403816, AF403815, AF403814, AF403813, AF403812, AF403811, AF403810, AF403809, AF403808, AF403807, AF403806, AF403805, AF403804, AF403803, AF403802, AF403801, AF403800, AF403799, AF403798, AF403797, AF403796, AF403795, AF403794, AF403793, AF403792, AF403791, AF403790, AF403789, AF403788, AF403787, AF403786, JQ285865, JQ285866, JQ285867, JQ285868, JQ285869, JQ285870

Supplementary Table B 2.3.10: Correlation of amplicon distance between group mean and geographic distance

Correlations			
		Distance Between Group Mean	Geographic Distance
Distance Between Group Mean	Pearson Correlation	1	-.011
	Sig. (2-tailed)		.970
	N	15	15
Geographic Distance	Pearson Correlation	-.011	1
	Sig. (2-tailed)	.970	
	N	15	15

Supplementary Table B 2.3.11: Genetic distance within group mean for mitochondrial and nuclear amplicon

Gene	Sample Location	Distance (Within Group Mean)
CO1	Daka	0.013584073
CO1	Pru	0.016283451
CO1	Black Volta	0.016557572
ND4	Daka	0.020663043
ND4	Pru	0.014805788
ND4	Black Volta	0.020558486
16S	Daka	0.001492381
16S	Pru	0.001601615
16S	Black Volta	0.001749608
ITS2	Pru	0.049700812
ITS2	Black Volta	0.004616598
ITS2	Daka	0.040683022
IGS	Pru	0.003328046
IGS	Black Volta	0.003863509
IGS	Daka	0.002623011

Supplementary Table B 2.3.12: Pearson Product-Moment Correlation for the mitochondrial and nuclear amplicon

Gene	Correlation Between		Variable 1 Mean	Variable 2 Mean	r	p
CO1	Genetic Distance (Between Group Mean) for CO1	Geographic Distance (Km) for CO1	.0152333	142.2666667	-0.377	0.754
	Fst (CO1)	Geographic Distance (Km)	-.0144700	142.2666667	-0.195	0.875
ND4	Genetic Distance (Between Group Mean) ND4	Geographic Distance (Km) for ND4	.0188667	142.2666667	0.999*	0.025
	Fst (ND4)	Geographic Distance (Km) for ND4			.391	.744
16S-rRNA	Genetic Distance (Between Group Mean) for 16S	Geographic Distance (Km) for 16S	.0015800	142.2666667	0.267	0.828
	Fst (16S)	Geographic Distance (Km) for (16S)			.811	.398
ITS2	Genetic Distance (Between Group Mean) for ITS	Geographic Distance (Km) for ITS	.0334333	167.7166667	-0.544	0.634
	Fst from ITS	Geographic Distance of ITS (Km)			0.133	0.500
IGS	Genetic Distance (Between Group Mean) for IGS	Geographic Distance (Km) for IGS	.0032497	167.7166667	-0.125	0.920
	Fst from IGS	Geographic Distance of IGS (Km)			0.124	.529

r = Pearson Correlation; p = Sig. (2-tailed); * = Correlation is significant at the 0.05 level (2-tailed).

Supplementary Sheets

Supplementary sheet 1: Ethical approval document for the study protocol



COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH HEAD OFFICE

P. O. BOX M. 32
ACCRA
GHANA
WEST - AFRICA

TEL: 233-30-2777651-4 (4 Lines)
FAX: 233-30-2777655
E-MAIL: headoffice@csir.org.gh
WEBSITE: www.csir.org.gh

Our Ref: CSIR/IRB/AL/VOL1.....

Date: 30th July, 2015.....

ETHICAL CLEARANCE

RPN 002/CSIR-IRB/2015

The Institutional Review Board of the Council for Scientific and Industrial Research (CSIR-IRB) has reviewed and given approval for the implementation of your study titled: "Development of diagnostic genetic markers to detect sub-optimal response to ivermectin treatment"

PRINCIPAL INVESTIGATOR : Dr. Mike Yaw Osei-Atweneboana

COLLABORATORS : Prof. Warwick Grant
Aime Gilles Adjami

Please note that a final review report must be submitted to the Board at the completion of the study. Your research records may be audited at any time during or after the implementation. Any modification of this research project must be submitted to the IRB for review and approval prior to implementation.

Please report all serious adverse events related to this study to CSIR-IRB within seven days verbally and fourteen days in writing.

This certificate is valid till 26th May, 2016.



Okyere Boateng
(CSIR-IRB, Chairman)

Cc: Director General, CSIR

Supplementary sheet 2: Independent Samples T-Test of nucleotide diversity for CO1 and ND4

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		.						Lower	Upper
Equal variances assumed	.050	.835	-1.583	4	.189	- .00141667	.00089473	- .00390084	.00106751
Equal variances not assumed			-1.583	9	.189	- .00141667	.00089473	- .00390591	.00107257

Supplementary Sheet 3: Independent Samples T-Test of nucleotide diversity for CO1 and 16S-rRNA

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		.						Lower	Upper
Equal variances assumed	12.590	.024	13.369	4	.000	.00881000	.00065899	.00698035	.01063965
Equal variances not assumed			13.369	2.049	.005	.00881000	.00065899	.00603833	.01158167

Supplementary sheet 4: Independent Samples T-Test of nucleotide diversity for CO1 and ITS2

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		.						Lower	Upper
Equal variances assumed	12.470	.024	-1.389	4	.237	- .01389333	.01000119	- .04166109	.01387443
Equal variances not assumed			-1.389	7	.298	- .01389333	.01000119	- .05657466	.02878800

Supplementary sheet 5: Independent Samples T-Test of nucleotide diversity for CO1 and IGS

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
		.						Lower	Upper
Equal variances assumed	2.286	.205	9.550	4	.001	.00714667	.00074836	.00506888	.00922445
Equal variances not assumed			9.550	7	.002	.00714667	.00074836	.00481494	.00947839

Supplementary sheet 6: Independent Samples T-Test of nucleotide diversity for CO1 and the Long-Range marker genes

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
	.334	.594	-.142	4	.894	- .00016000	.00112351	- .00327935	.00295935
Equal variances assumed			-.142	3.628	.894	- .00016000	.00112351	- .00340964	.00308964

Supplementary sheet 7: Independent Samples T-Test of nucleotide diversity for CO1 and the whole mitochondrial genome

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
	3.099	.153	2.789	4	.049	.00203667	.00073037	.00000883	.00406451
Equal variances not assumed			2.789	2.919	.071	.00203667	.00073037	- .00032461	.00439795

Supplementary sheet 8: Independent Samples T-Test of nucleotide diversity for CO1 and the concatenated nuclear genes

Independent Samples Test									
	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
	15.262	.017	14.347	4	.000	.00940000	.00065518	.00758092	.01121908
Equal variances assumed			14.347	2.002	.005	.00940000	.00065518	.00658390	.01221610

Supplementary sheet 9: One -Sample T-Test of nucleotide diversity using tRNA and associated genes of *Simulium* blackfly from Ghana

	One-Sample Test						
	Test Value = 0.01						
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference		
					Lower	Upper	
Nucleotide diversity	.635	2	.590	.00058000	- .0033475	.0045075	

Supplementary sheet 10: Sequence statistics for *Mus sp.* and *Apodemus agrarius*

π	Seg Site	Pars Sites	Tajima's D	Tajima's p
0.015055	3598	3186	-1.90309	0.0286027

In the table above, π = nucleotide diversity; Seg Site = number of segregating sites; Pars Sites 3186; Tajima's D: -1.90309; Tajima's p = 0.0286027

Supplementary sheet 11: Sequence statistics for *Mus sp.* and *Apodemus agrarius*

Organism	VAP (%)	VWP (%)	Phist	Phist p
<i>Mus sp.</i> and <i>Apodemus agrarius</i>	98.76117	1.23883	0.98761	$p < 0.001$

VAP= Percentage of Variation among population; VWP= Percentage of Variation within population; Phist = Fixation index; Phist p = Significance of the Fixation index (1000 permutations)

References

- Abu, M., *et al.* (2014). "Climate change and internal migration intentions in the forest-savannah transition zone of Ghana." Population and Environment **35**(4): 341-364.
- Adams, M. D., *et al.* (2000). "The genome sequence of *Drosophila melanogaster*." Science **287**(5461): 2185-2195.
- Adeleke, M., *et al.* (2010). "Socio-economic implications of *Simulium damnosum* complex infestation in some rural communities in Odeda Local Government Area of Ogun State, Nigeria." Journal of Public Health and Epidemiology **2**(5): 109-112.
- Adler, P. and R. Crosskey (2015). "World black flies (Diptera: Simuliidae): A comprehensive revision of the taxonomic and geographical inventory [2015]. 123 pp." South Carolina, Clemson University.
- Adler, P. H. (2005). Black flies, the Simuliidae. In "Biology of Disease Vectors" 2nd ed Ed by WC Marquardt, Elsevier Academic Press, San Diego.
- Adler, P. H. (2009). "Biodiversity of biting flies: implications for humanity." Insect Biodiversity: Science And Society 523-545.
- Adler, P. H. and E. A. Kachvorian. (2001). "Cytogenetics of the Holarctic black fly *Simulium noelleri* (Diptera: Simuliidae)." Canadian Journal of Zoology **79**(11): 1972-1979.
- Adler, P. H. and J. W. McCreadie. (2019). "Black flies (Simuliidae)." Medical and Veterinary Entomology. 2nd ed. Elsevier, San Diego, CA, pp. 237-259.
- Adler, P. H. and R. W. Crosskey (2014). "World blackflies (Diptera: Simuliidae): a comprehensive revision of the taxonomic and geographical inventory. 2016." Accessed at: www.clemson.edu/cafls/departments/esps/biomia/pdfs/blackflyinventory.pdf.
- Adler, P.H., *et al.* (2004). "The Black Flies (Simuliidae) of North America." Cornell University Press, Ithaca, NY xv + 941 pp. +24.
- Adler, P. H., *et al.* (2005). "Insular black flies (Diptera: Simuliidae) of North America: tests of colonization hypotheses." Journal of Biogeography **32**(2): 211-220.
- Adler, P. H., *et al.* (2010). "Evolution, epidemiology, and population genetics of black flies (Diptera: Simuliidae)." Infection, Genetics and Evolution **10**(7): 846-865.

Adler, P. H., *et al.* (2013). "Macrogenomic evidence for the origin of the black fly *Simulium suzukii* (Diptera: Simuliidae) on Okinawa Island, Japan." *PLoS One* **8**(8): e70765.

Afikorah-Danquah, S. J. I. B. (1997). "Local resource management in the forest-savanna transition zone: The case of Wenchi District, Ghana." *IDS Bulletin* **28**(4): 36-46.

AFRO (2019). "Expanded Special Project for Elimination of Neglected Tropical Disease." World Health Organization Regional Office for Africa (WHO-AFRO). Accessed at: <https://www.afro.who.int/health-topics/expanded-special-project-elimination-neglected-tropical-disease>. Access date: 18th September, 2019.

Agatsuma, T., *et al.* (1993). "Typing of *Simulium* using mitochondrial DNA PCR products with restriction endonuclease digestion." *Annals of Tropical Medicine & Parasitology* **87**(3): 307–309

Alexeyev, M. F., *et al.* (2004). "Mitochondrial DNA and aging." *Clinical Science* **107**(4): 355-364.

Alley, E., *et al.* (1994). "The impact of five years of annual ivermectin treatment on skin microfilarial loads in the onchocerciasis focus of Asubende, Ghana." *Transactions of the Royal Society of Tropical Medicine and Hygiene* **88**(5): 581-584.

Altman, P. L. and D. D. Katz (1976). "Cell biology, Federation of American Societies for Experimental Biology" Bethesda 1976: 367-368.

Amazigo, U. (2008). "The African programme for onchocerciasis control (APOC)." *Annals of Tropical Medicine & Parasitology* **102**(sup1): 19-22.

Amazigo, U. V., *et al.* (2002). "The challenges of community-directed treatment with ivermectin (CDTI) within the African Programme for Onchocerciasis Control (APOC)." *Annals of Tropical Medicine Parasitology* **96**(sup1): S41-S58.

Amrine, J. W. (1982). "The New River connection to the black fly problem in southern West Virginia." West Virginia University, Agricultural and Forestry Experiment Station Bulletting No 678: 1–30.

Anderson, S., *et al.* (1981). "Sequence and organization of the human mitochondrial genome." *Nature* **290**(5806): 457–465.

Anderson, S., *et al.* (1982). "Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome." *Journal of molecular biology* **156**(4): 683-717.

Arif I. A., *et al.* (2012): "Inferring the Phylogeny of Bovidae Using Mitochondrial DNA sequences: Resolving power of Individual Genes Relative to complete Genomes." *Evolutionary Bioinformatics* **8**(2012): EBO-S8897 139–150.

Avise J. C. (1994). "Molecular markers, natural history and evolution." Chapman and Hall, New York Springer-Science+Business Media Dordrecht, Boston **1994**:1-511.

Avise, J. C. (1995). "Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation." Conservation Biology **9**(3): 686-690.

Avise, J. C. J. E. (1989). "Gene trees and organismal histories: a phylogenetic approach to population biology." Evolution **43**(6): 1192-1208.

Avise, J. (1990). "Principles of genealogical concordance in species concepts and biological taxonomy." Oxford Surveys in Evolutionary Biology **7**: 45-67.

Awadalla P., *et al.* (1999) "Linkage disequilibrium and recombination in hominid mitochondrial DNA." Science **286**(5449): 2524–2525.

Axelsson, E., *et al.* (2008). "The effect of ancient DNA damage on inferences of demographic histories." Molecular Biology and Evolution **25**(10): 2181-2187.

Ayala, F. J. and J. R. Powell. (1972). "Allozymes as diagnostic characters of sibling species of *Drosophila*." Proceedings of the National Academy of Sciences. **69**(5): 1094–1096.

Bagley, M. and G. J. M. E. Gall (1998). "Mitochondrial and nuclear DNA sequence variability among populations of rainbow trout (*Oncorhynchus mykiss*).". **7**(8): 945-961.

Bain, O. (1981). "Species of the genus *Onchocerca* and primarily *O. volvulus*, considered from an epidemiologic and phylogenetic point of view." Annales de la Societe belge de medecine tropicale. **61**(2):225-31

Baker, R., *et al.* (1990). "Progress in controlling the reinvasion of windborne vectors into the western area of the Onchocerciasis Control Programme in West Africa." Philosophical Transactions of the Royal Society of London. B, Biological Sciences **328**(1251): 731-750.

Baker, R., *et al.* (2002). "Avise JC (2000) Phylogeography: the history and duration of species. Harvard University Press, Cambridge, MA. Avise J. C, Arnold J., Ball RM, Bermingham E., Lamb T., Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics." Annual Review of Ecology **56**: 7.

Baldry, D., *et al.* (1985). "Measures aimed at controlling the invasion of *Simulium damnosum* Theobald sl (Diptera: Simuliidae) into the Onchocerciasis Control Programme Area. I. Experimental aerial larviciding in the Upper Sassandra Basin of South-eastern Guinea in 1985." *International Journal of Pest Management* **31**(4): 255-263.

Ballard, J. W. O. (2000a). "Comparative genomics of mitochondrial DNA in *Drosophila simulans*." *Journal of Molecular Evolution* **51**(1): 64-75.

Ballard, J. W. O. (2000b). "When one is not enough: introgression of mitochondrial DNA in *Drosophila*." *Molecular Biology and Evolution* **17**(7): 1126-1130.

Ballard J. W. O. (2000c). "Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup." *Journal of Molecular Evolution* **51**: 48-63.

Ballard J. W. O. and Whitlock M. C. (2004). "The incomplete natural history of mitochondria." *Molecular Ecology* (2004) **13**(4): 729-744.

Bandelt, H.-J., *et al.* (1999). "Median-joining networks for inferring intraspecific phylogenies." *Molecular Biology and Evolution* **16**(1): 37-48.

Barnley, G. and M. Prentice (1958). "*Simulium neavei* in Uganda." *East African Medical Journal* **35**(8).

Barr, W. B. (1984). "Prolegs and attachment of *Simulium vittatum* (sibling IS-7)(Diptera: Simuliidae) larvae." *Canadian Journal of Zoology* **62**(7): 1355-1362.

Bartley, D., *et al.* (2006). "A small scale survey of ivermectin resistance in sheep nematodes using the faecal egg count reduction test on samples collected from Scottish sheep." *Veterinary Parasitology* **137**(1-2): 112-118.

Basáñez, M., *et al.* (2000). "*Onchocerca-Simulium* complexes in Venezuela: can human onchocerciasis spread outside its present endemic areas?" *Parasitology* **120**(2): 143-160.

Basanez, M.-G., *et al.* (2008). "Effect of single-dose ivermectin on *Onchocerca volvulus*: a systematic review and meta-analysis." *The Lancet Infectious Diseases* **8**(5): 310-322.

Basáñez, M. G., *et al.* (2009). "*Onchocerca-Simulium* interactions and the population and evolutionary biology of *Onchocerca volvulus*." *Advances in Parasitology* **68**: 263-313.

Bedo, D. (1976). "Polytene chromosomes in pupal and adult blackflies (Diptera: Simuliidae)." *Chromosoma* **57**(4): 387-396.

Bedo, D. (1989). "A cytological study of *Simulium ruficorne* (Diptera: Simuliidae) and its relationship to the *S. ornatipes* species complex." *Genome* **32**(4): 570-579.

Bedo, D. G. (1975). "C banding in polytene chromosomes of *Simulium ornatipes* and *S. melatum* (Diptera: Simuliidae)." *Chromosoma* **51**(3): 291-300.

Beltrán-López R. G., *et al.* (2017). "Phylogeny and taxonomy of the genus *Ilyodon Eigenmann*, 1907 (Teleostei: Goodeidae), based on mitochondrial and nuclear DNA sequences." *Journal of Zoological Systematics and Evolutionary Research* **55**(4): 340–355.

Bensch, S. (2000). "Mitochondrial genomic rearrangements in songbirds." *Molecular Biology and Evolution* **17**(1): 107-113.

Benton, B., *et al.* (2002). "Partnership and promise: evolution of the African river-blindness campaigns." *Annals of Tropical Medicine and Parasitology* **96** (*Suppl. 1*): S5–S14.

Bernardi, G., *et al.* (1993). "Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*." *Proceedings of the National Academy of Sciences* **90**(20): 9271-9274.

Bernatchez L, *et al.* (1995). "Introgression and fixation of Arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*)."*Canadian Journal of Fisheries and Aquatic Science* **52**(1): 179–185.

Berryman, A. A. J. O. (2002). "Population: a central concept for ecology?" *Oikos* **97**(3): 439-442.

Bibb, M. J., *et al.* (1981). "Sequence and gene organization of mouse mitochondrial DNA." *Cell* **26**(2): 167-180.

Birky C. W. (2001). "The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models." *Annual Review of Genetics* **35**(1): 125–148

Birrell, J. A. and J. J. F. I. Hirst (2010). "Truncation of subunit ND2 disrupts the threefold symmetry of the antiporter-like subunits in complex I from higher metazoans." *FEBS letters* **584**(19): 4247-4252.

Bissan, Y., *et al.* (1995). "Drastic reduction of populations of *Simulium sirbanum* (Diptera: Simuliidae) in central Sierra Leone after 5 years of larviciding operations by the Onchocerciasis Control Programme." *Annals of Tropical Medicine & Parasitology* **89**(1): 63-72.

Biswas, A. K. (2012). Impacts of large dams: Issues, opportunities and constraints. *Impacts of large dams: A global assessment*, Springer: 1-18.

Blacklock, D. (1926a). "The development of *Onchocerca volvulus* in *Simulium damnosum*." Annals of Tropical Medicine Parasitology **20**(1): 1-48.

Blacklock, D. (1926b). "The further development of *Onchocerca volvulus* Leuckart in *Simulium damnosum* Theob." Annals of Tropical Medicine Parasitology **20**(2): 203-218.

Blacklock, D. (1927). "The insect transmission of *Onchocerca volvulus* (Leuckart, 1893): the cause of worm nodules in man in Africa." British Medical Journal **1**(3446): 129.

Blanks, J., et al. (1998). "The Onchocerciasis Elimination Program for the Americas: a history of partnership." Revista Panamericana de Salud Pública **3**: 367-374.

Boakye, D., et al. (1998). "Sibling species distributions of the *Simulium damnosum* complex in the west African Onchocerciasis Control Programme area during the decade 1984-93, following intensive larviciding since 1974." Medical Veterinary Entomology **12**(4): 345-358.

Boakye, D., et al. (1999). "The cyt taxonomy of *Simulium sanctipauli* sensu stricto (Diptera: Simuliidae) in Ghana and the Côte d'Ivoire." Bulletin of Entomological Research **89**(1): 25-33.

Boakye, D., et al. (2000). "Evidence of multiple mating and hybridization in *Simulium damnosum* sl (Diptera: Simuliidae) in nature." Journal of Medical Entomology **37**(1): 29-34.

Boakye, D. and F. Mosha (1988). "Natural hybridization between *Simulium sanctipauli* and *S. sirbanum*, two sibling species of the *S. damnosum* complex." Medical Veterinary Entomology **2**(4): 397-399.

Boakye, D. and S. Meredith (1993). "Introgression between members of the *Simulium damnosum* complex: larvicidal implications." Medical Veterinary Entomology **7**(4): 393-397.

Boakye, D., et al. (1993a). "Cytotaxonomic revision of the *Simulium sanctipauli* subcomplex (Diptera: Simuliidae) in Guinea and the adjacent countries including descriptions of two new species." Bulletin of Entomological Research **83**(2): 171-186.

Boakye, D. (1993b). "A pictorial guide to the chromosomal identification of members of the *Simulium damnosum* Theobald complex in west Africa with particular reference to the Onchocerciasis Control Programme Area." Tropical Medicine Parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft of Deutsche Gesellschaft für Technische Zusammenarbeit **44**(3): 223-244.

Bockarie, M., et al. (1990). "The transmission of onchocerciasis at a forest village in Sierra Leone. I. *Simulium damnosum* sl biting densities and infection with *Onchocerca volvulus* at five representative sites." Annals of Tropical Medicine Parasitology **84**(6): 587-597.

Borsboom, G. J., *et al.* (2003). "Impact of ivermectin on onchocerciasis transmission: assessing the empirical evidence that repeated ivermectin mass treatments may lead to elimination/eradication in West-Africa." Filaria Journal **2**(1): 8.

Bourguinat, C., *et al.* (2007). "Genetic selection of low fertile *Onchocerca volvulus* by ivermectin treatment." PLoS Neglected Tropical Diseases **1**(1): e72.

Boursot, P. and F. Bonhomme (1986). "Génétique et évolution du génome mitochondrial des Métazoaires." Génétique, Sélection, Évolution **18**(1): 73.

Boussinesq, M. (1997). "L'onchocercose humaine en Afrique." Médecine Tropicale **57**(4): 389-400.

Boussinesq, M., *et al.* (2002). "Relationship between onchocerciasis and epilepsy: a matched case-control study in the Mbam Valley, Republic of Cameroon." Transactions of the Royal Society of Tropical Medicine and Hygiene **96**(5): 537-541.

Boyko, A. R., *et al.* (2009). "Complex population structure in African village dogs and its implications for inferring dog domestication history." Proceedings of the National Academy of Sciences **106**(33): 13903-13908.

Bradbury I. R., *et al.* (2011). "Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*." Molecular Ecology Resources **11**(Suppl.1): 218–225.

Bradley, J. E., *et al.* (2010). "Onchocerciasis." Topley Wilson's Microbiology Microbial Infections.

Braimah, S. A. (1987). "The influence of water velocity on particle capture by the labral fans of larvae of *Simulium bivittatum* Malloch (Diptera: Simuliidae)." Canadian Journal of Zoology **65**(10): 2395-2399.

Branco, M. R. and A. J. P. b. Pombo (2006). "Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations." PLoS Biology **4**(5): e138.

Brockhouse, C., *et al.* (1993). "Molecular identification of onchocerciasis vector sibling species in black flies (Diptera: Simuliidae)." Biochemical and Biophysical Research Communications **194**(2): 628-634.

Brockhouse, C. L. and R. M. Tanguay. (1996). "A rapid method for the detection of sibling species in samples of black fly larvae (Diptera: Simuliidae)." Journal of Agricultural Entomology **13**(4): 339–347.

Bröker, S., et al. (1998). "MtDNA mutations associated with sideroblastic anaemia cause a defect of mitochondrial cytochrome c oxidase." European Journal of Biochemistry **258**(1): 132-138.

Bronstein, O., et al. (2017). "Cryptic speciation in pan-tropical sea urchins: a case study of an edge-of-range population of *Tripneustes* from the Kermadec Islands." Scientific Reports **7**(1): 5948.

Bronstein, O., et al. (2018). "Mind the gap! The mitochondrial control region and its power as a phylogenetic marker in echinoids." BMC Evolutionary Biology **18**(1): 80.

Brown, M. D., et al. (1992). "A mitochondrial DNA variant, identified in Leber hereditary optic neuropathy patients, which extends the amino acid sequence of cytochrome c oxidase subunit I." American Journal of Human Genetics **51**(2): 378.

Brown, W. M. (1983). "Evolution of Animal Mitochondrial DNA." Pp. 62-88 in M. Nei and R. K. Koehn, eds. *Evolution of genes and proteins*. Sinauer, Sunderland, Mass."

Brown, W. M., et al. (1979). "Rapid evolution of animal mitochondrial DNA." Proceedings of the National Academy of Sciences **76**(4): 1967-1971.

Buck, A., et al. (1978). "Epidemiology of poly-parasitism. I. Occurrence, frequency and distribution of multiple infections in rural communities in Chad, Peru, Afghanistan, and Zaire." Tropenmedizin und Parasitologie **29**(1): 61-70.

Buonaccorsi, V., et al. (2001). "Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*)."Molecular Ecology **10**(5): 1179-1196.

Burnham, G. (1998). "Onchocerciasis." The Lancet **351**(9112): 1341-1346.

Burnham, G. M. (1991). "Onchocerciasis in Malawi. 1. Prevalence, intensity and geographical distribution of *Onchocerca volvulus* infection in the Thyolo highlands." Transactions of the Royal Society of tropical Medicine and Hygiene **85**(4): 493-496.

Cadahía, L., et al. (2009). "Repeated sequence homogenization between the control and pseudo-control regions in the mitochondrial genomes of the subfamily Aquilinae." Journal of Experimental Zoology Part B: Molecular Developmental Evolution **312**(3): 171-185.

Camberlin, P. and N. J. J. O. C. Philippon (2002). "The East African March-May rainy season: Associated atmospheric dynamics and predictability over the 1968-97 period." Journal of Climate **15**(9): 1002-1019.

Camberlin, P., *et al.* (1997). "Intraseasonal wind anomalies related to wet and dry spells during the “long” and “short” rainy seasons in Kenya." Theoretical and applied climatology **58**(1-2): 57-69.

Campbell, N. and S. C. Barker (1999). "The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region." Molecular Biology and Evolution **16**(6): 732-740.

Campbell, W. (1985). "Ivermectin: an update." Parasitology Today **1**(1): 10-16.

Capaldi, R. A., *et al.* (1986). "Sequence homologies and structural similarities between the polypeptides of yeast and beef heart cytochrome c oxidase." FEBS Letters **207**(1): 11-17.

Capaldi, R. A., *et al.* (1983). "Structure of cytochrome c oxidase." Biochim. Biophys. Acta **726**(135-38).

Capaldi, R. A. (1990). "Structure and function of cytochrome c oxidase." Annual Review of Biochemistry **59**(1): 569-596.

Caterino, M. S., *et al.* (2000). "The current state of insect molecular systematics: a thriving Tower of Babel." Annual Review of Entomology **45**(1): 1-54.

CDC (2013). "Progress toward elimination of onchocerciasis in the Americas—1993–2012." MMWR. Morbidity mortality weekly report. Accessed at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4604938/pdf/405-408.pdf>. Access date: 13th September, 2019; 62(20):405–408. **62**(20): 405.

CDC (2019). "(Centers for Disease Control and Prevention). 2019. Onchocerciasis: Life Cycle. Parasite Biology, Accessed At: <https://www.cdc.gov/dpdx/onchocerciasis/index.html> . Access Date: 9th October, 2019." Parasite Biology.

Chakalova, L. and P. Fraser (2010). "Organization of transcription." Cold Spring Harbor Perspectives in Biology **2**(9): a000729.

Chambeyron, S., *et al.* (2004). "Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription." Genes & Development **18**(10): 1119-1130.

Charalambous, M., *et al.* (1995). "Cytogenetical analysis of the *Simulium damnosum* complex (Diptera: Simuliidae) in Guinea Bissau." Medical and Veterinary Entomology **9**(1): 34-42.

Charlesworth, D. and S. Wright (2001). "Breeding systems and genome evolution." Current Opinion in Genetics Development **11**(6): 685-690.

Chavasse, D., *et al.* (1995). "Low level ivermectin coverage and the transmission of onchocerciasis." Transactions of the Royal Society of Tropical Medicine and Hygiene **89**(5): 534-537.

Cheke, R. and A. Denke (1988). "Anthropophily, zoophily and roles in onchocerciasis transmission of the Djodji form of *Simulium sanctipauli* and *S. squamosum* in a forest zone of Togo." Tropical Medicine and Parasitology **39**(2): 123-127.

Cheke, R. and R. Garms, Biological Sciences (1983). "Reinfestations of the southeastern flank of the Onchocerciasis Control Programme area by windborne vectors." Philosophical Transactions of the Royal Society of London. B, Biological Sciences **302**(1111): 471-484.

Cheke, R. A., *et al.* (2015). "Potential effects of warmer worms and vectors on onchocerciasis transmission in West Africa." Philosophical Transactions of the Royal Society B, Biological Sciences **370**(1665): 20130559.

Chomyn, A., *et al.* (1983). "Identification of the polypeptides encoded in the ATPase 6 gene and in the unassigned reading frames 1 and 3 of human mtDNA." Proceedings of the National Academy of Sciences **80**(18): 5535-5539.

Choochote, W., *et al.* (2005). "Seasonal abundance and daily flying activity of black flies (Diptera: Simuliidae) attracted to human baits in Doi Inthanon National Park, northern Thailand." Medical Entomology Zoology.

56(4): 335-348.

Churcher, T. S., *et al.* (2009). "Identifying sub-optimal responses to ivermectin in the treatment of River Blindness." Proceedings of the National Academy of Sciences **106**(39): 16716-16721.

Clary, D. O. and D. R. Wolstenholme (1985). "The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code." Journal of Molecular Evolution **22**(3): 252-271.

Colbo, M. (1976). "Four new species of *Simulium Latreille* (Diptera: Simuliidae) from Australia." Australian Journal of Entomology **15**(3): 253-269.

Collins, R., *et al.* (1992). "Ivermectin: reduction in prevalence and infection intensity of *Onchocerca volvulus* following biannual treatments in five Guatemalan communities." The American Journal of Tropical Medicine and Hygiene **47**(2): 156-169.

Collins, R. C., *et al.* (1995). "Vector competence of *Simulium exiguum* for *Onchocerca volvulus*: implications for the epidemiology of onchocerciasis." The American Journal of Tropical Medicine Hygiene **52**(3): 213-218.

Collura, R. V. and C.-B. Stewart (1995). "Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids." Nature **378**(6556): 485.

Conflitti, I. M., *et al.* (2010). "Good species behaving badly: non-monophyly of black fly sibling species in the *Simulium arcticum* complex (Diptera: Simuliidae)." Molecular Phylogenetics and Evolution, **57**(1): 245-257.

Conflitti, I. M., *et al.* (2012). "A "complex" problem: delimiting sibling species boundaries in black flies (Diptera: Simuliidae)." The Canadian Entomologist **144**(2): 323-336.

Corbin, C. J., *et al.* (2004). "Paralogues of porcine aromatase cytochrome P450: a novel hydroxylase activity is associated with the survival of a duplicated gene." Endocrinology **145**(5): 2157-2164.

Coyle, P.V., *et al.* (2004). "A touchdown nucleic acid amplification protocol as an alternative to culture backup for immunofluorescence in the routine diagnosis of acute viral respiratory tract infections." BioMed Central Microbiol **4**(1): 41.

Craig, D. (1974). "The labrum and cephalic fans of larval Simuliidae (Diptera: Nematocera)." Canadian Journal of Zoology **52**(1): 133-159.

Craig, T., *et al.* (2007). "Evidence of ivermectin resistance by Parascaris equorum on a Texas horse farm." Journal of Equine Veterinary Science **27**(2): 67-71.

Crandall, K. A. and A. R. Templeton (1999). "Statistical methods for detecting recombination. In The Evolution of HIV (Crandall, K.A., ed.), pp. 153–176, Johns Hopkins University Press."

Crawford, K. E., *et al.* (2019). "Utility of the *Onchocerca volvulus* mitochondrial genome for delineation of parasite transmission zones." BioRxiv: 732446.

Cremer, T. and M. Cremer (2010). "Chromosome territories." Cold Spring Harbor Perspectives in Biology **2**(3): a003889.

Crews, S. and G. Attardi (1980). "The sequences of the small ribosomal RNA gene and the phenylalanine tRNA gene are joined end to end in human mitochondrial DNA." Cell **19**(3): 775-784.

Crick, F. (1968). "The origin of the genetic code." Journal of Molecular Biology **38**(3): 367-379.

Crisp, G. (1956). "Simulium and onchocerciasis in the Northern Territories of the Gold Coast." *Simulium and Onchocerciasis in the Northern Territories of the Gold Coast*. pp.xvi + 171 pp. ref.4 1/2 pp.

Crosskey, R. (1962). "The identification of the larvae of African *Simulium*." Bulletin of the World Health Organization **27**(4-5): 483.

Crosskey, R. (1967). "The classification of *Simulium latreille* (Diptera: Simuliidae) from Australia, New Guinea and the Western Pacific." Journal of Natural History **1**(1): 23-51.

Crosskey, R. (1990). "The Natural History of Blackflies." John Wiley & Sons Ltd, Chichester, England 711 pp.

Crosskey, R. and W. Büttiker (1982). "Insects of Saudi Arabia. Diptera: Fam. Simuliidae." Fauna of Saudi Arabia **4**(1982): 398-446.

Crosskey, R. W. (1969). "A re-classification of the Simuliidae (Díptera) of Africa and its islands." A Re-classification of the Simuliidae of Africa and its Islands. (Suppl. 14).

Crump, A. (2003). "Outstanding, innovative disease control solution meets a minor hitch." Appropriate Technology **30**(3): 13.

Crump, A., et al. (2012). "The onchocerciasis chronicle: from the beginning to the end?" Trends in Parasitology **28**(7): 280-288.

Crump, A. and S. Omura (2011). "Ivermectin, 'wonder drug' from Japan: the human use perspective." Proceedings of the Japan Academy, Series B **87**(2): 13-28.

Cruz-Ortiz, N., et al. (2012). "Elimination of *Onchocerca volvulus* transmission in the Huehuetenango focus of Guatemala." Journal of Parasitology Research **2012**.

Cully, D. F., et al. (1994). "Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*." Nature **371**(6499): 707.

Cummings, M. P, et al. (1995). "Sampling properties of DNA sequence data in phylogenetic analysis." Molecular Biology and Evolution **12**: 814–822.

Cupp, E., et al. (2019). "Elimination of onchocerciasis in Africa by 2025: the need for a broad perspective." Infectious Diseases of Poverty **8**(1): 50.

Cupp, E., et al. (2011a). "Elimination of human onchocerciasis: history of progress and current feasibility using ivermectin (Mectizan®) monotherapy." Acta Tropica **120**: S100-S108.

Cupp, E. W., *et al.* (2004). "The effects of long-term community level treatment with ivermectin (Mectizan®) on adult *Onchocerca volvulus* in Latin America." The American Journal of Tropical Medicine and Hygiene **71**(5): 602-607.

Cupp, E. W., *et al.* (2011b). "Importance of Ivermectin to human onchocerciasis: past, present, and the future." Research and Reports in Tropical Medicine **2**: 81.

Currie, D. and D. Grimaldi (2000). "A new black fly (Diptera: Simuliidae) genus from mid Cretaceous (Turonian) amber of New Jersey." Studies on Fossils in Amber, with Particular Reference to the Cretaceous of New Jersey. Backhuys Publishers, Leiden, The Netherlands 473-485.

Currie, D. C. and P. H. Adler (2008). "Global diversity of black flies (Diptera: Simuliidae) in freshwater." Hydrobiologia **595**(1): 469-475.

Dalén, L., *et al.* (2007). "Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox." Proceedings of the National Academy of Sciences **104**(16): 6726-6729.

Dalmat, H. T. (1950). "Studies on the flight range of certain Simuliidae, with the use of aniline dye marker." Annals of the Entomological Society of America **43**: 537-545.

Dalmat, H. T. (1952). "Longevity and further flight range studies on the blackflies (Diptera, Simuliidae), using dye markers." Annals of the Entomological Society of America **45**: 23-37.

Datta, M. (1992). "An overview of the Simuliidae (Diptera) of west Bengal, India." Bengal Nat. Hist. Soc., New Series **11**: 41-62.

Daumerie, D. and L. Savioli (2010). "Working to overcome the global impact of neglected tropical diseases: first WHO report on neglected tropical diseases." World Health Organization, Vol 1.

Davies, P. A., and Gray, G. (2002). "Long-range PCR. In: Theophilus B. D. M., Rapley R. (eds) *PCR Mutation Detection Protocols*." Methods in Molecular Biology, Humana Press 187: 51-55.

Davies, J. B. (1962). "Egg-laying habits of *Simulium damnosum* Theobald and *Simulium mcdusaeforme* form hargrevesi Gibbins in Northern Nigeria." Nature **196**(4850).

Day, J. C., et al. (2008). "Confirmation of the species status of *Simulium galeratum* Edwards, 1920 (Diptera: Simuliidae) in Britain using DNA barcoding." *Medical and Veterinary Entomology* **22**: 55–61.

Day, J. C., et al. (2016). "Sequence and organization of the complete mitochondrial genome of the blackfly *Simulium variegatum* (Diptera: Simuliidae)." *Mitochondrial DNA Part B* **1**(1): 799-801.

De, G. S., et al. (1989). "A community trial of ivermectin in the onchocerciasis focus of Asubende, Ghana. II. Adverse reactions." *Tropical Medicine Parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft of Deutsche Gesellschaft fur Technische Zusammenarbeit* **40**(3): 375-382.

DeFoliart, G. and M. R. Rao (1965). "The ornithophilic black fly *Simulium meridionale* Riley (Diptera: Simuliidae) feeding on man during autumn." *Journal of Medical Entomology* **2**(1): 84-85.

Degnan, S. M. (1993). "The perils of single-gene trees — mitochondrial versus single-copy nuclear-DNA variation in white-eyes (Aves, Zosteropidae)." *Molecular Ecology* **2**(4): 219–225.

De Leon, J. R. and B. Duke (1966). "Experimental studies on the transmission of Guatemalan and West African strains of *Onchocerca volvulus* by *Simulium ochraceum*, *S. metallicum* and *S. callidum*." *Transactions of the Royal Society of Tropical Medicine and Hygiene* **60**(6): 735-752.

Delsuc, F., et al. (2003). "Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes." *Molecular Phylogenetics and Evolution* **28**(2): 261-275.

Diawara, L., et al. (2009). "Feasibility of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: first evidence from studies in Mali and Senegal." *PLoS Neglected Tropical Diseases* **3**(7): e497.

Diesing (1841). "Onchocerca Diesing, 1841." IN: Gargominy, Olivier. (2019). TAXREF. Version 4.4. UMS PatriNat (AFB-CNRS-MNHN), Paris. Checklist dataset <https://doi.org/10.15468/vqueam> accessed via GBIF.org on 2019-10-15.

Diniz F. M., et al. (2005). "The hypervariable domain of the mitochondrial control region in Atlantic spiny lobsters and its potential as a marker for investigating phylogeographic structuring." *Marine Biotechnology* **7**(5):462–473.

Donnelly, J. (2012). "CDC planning trial for mysterious nodding syndrome." *The Lancet* **379**(9813): 299.

Dourmishev, A. L., *et al.* (2005). "Ivermectin: pharmacology and application in dermatology." International Journal of Dermatology **44**(12): 981-988.

Dowling, T. (1990). "Nucleic acid II. Restriction site analysis." 250-319.

Downes, J. A. and D. H. Colless. (1967). "Mouthparts of the biting and blood-sucking type in Tanyderidae and Chironomidae (Diptera)." Nature **214**(5095): 1355–1356.

Downton, M., *et al.* (2003). "Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction." Journal of Molecular Evolution **56**(5): 517– 526.

Doyle, S. R., *et al.* (2017). "Genome-wide analysis of ivermectin response by *Onchocerca volvulus* reveals that genetic drift and soft selective sweeps contribute to loss of drug sensitivity." PLoS Neglected Tropical Diseases **11**(7): e0005816.

Duke, B. (1968). "The intake and transmissibility of *Onchocerca volvulus* microfilariae by *Similium damnosum* fed on patients treated with diethylcarbamazine, suramin or Mel W." Bulletin of the World Health Organization **39**(2): 169.

Duke, B. (1970). "*Onchocerca-Simulium* complexes: VI.—Experimental studies on the transmission of Venezuelan and West African strains of *Onchocerca volvulus* by *Simulium metallicum* and *S. exiguum* in Venezuela." Annals of Tropical Medicine & Parasitology **64**(4): 421-431.

Duke, B., *et al.* (1966). "*Onchocerca-Simulium* complexes: I.—Transmission of forest and Sudan-savanna strains of *Onchocerca volvulus*, from Cameroon, by *Simulium damnosum* from various West African bioclimatic zones." Annals of Tropical Medicine & Parasitology **60**(3): 318-336.

Duke, B., *et al.* (1967). "*Onchocerca-Simulium* complexes: V.—The intake and subsequent fate of microfilariae of a Guatemalan strain of *Onchocerca volvulus* in forest and Sudan-savanna forms of West African *Simulium damnosum*." Annals of Tropical Medicine & Parasitology **61**(3): 332-337.

Dunbar, R. W. (1969). "Nine cytological segregates in the *Simulium damnosum* complex (Diptera: Simuliidae)." Bulletin of the World Health Organization **40**(6): 974.

Dunbar, R. W. and Vajime, C. G. (1981). "Cytotaxonomy of the *Simulium damnosum* complex. In: Laird, M. (Ed.), Blackflies: The Future for Biological Methods in Integrated Control." Academic Press, London, pp. 31–43.

Duncan, G. A., *et al.* (2004). "Molecular differentiation of two sibling species of the black fly *Simulium vittatum* (Diptera: Simuliidae) based on random amplified polymorphic DNA." *Genome* **47**(2): 373–379.

Dyar, H. G. and R. C. Shannon (1927). "The North American two-winged flies of the family Simuliidae." *Proceedings of the United States National Museum* **69**(10): 3612.

Eberhard, J. R., *et al.* (2001). "Duplication and concerted evolution of the mitochondrial control region in the parrot genus *Amazona*." *Molecular Biology and Evolution* **18**(7): 1330–1342.

Eberhard, M. (1986). "Longevity of microfilariae following removal of the adult worms." *Tropical Medicine and Parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft of Deutsche Gesellschaft fur Technische Zusammenarbeit* **37**(4): 361-363.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." *Nucleic Acids Research* **32**(5): 1792-1797.

Edwards, S. V. and Beerli P. (2000). "Perspective: gene divergence, population divergence, and the variance in coalescent time in phylogeographic studies." *Evolution*, **54**(6): 1839–1854.

Edwards, S. V., *et al.* (1991). "Mitochondrial resolution of a deep branch in the genealogical tree for perching birds." *Proceedings of the Royal Society of London. Series B: Biological Sciences* **243**(1307): 99-107.

Efremov, R. G., *et al.* (2010). "The architecture of respiratory complex I." *Nature* **465**(7297): 441-445.

Eltahir, E. A. and C. Gong (1996). "Dynamics of wet and dry years in West Africa." *Journal of Climate* **9**(5): 1030-1042.

Eng, J., *et al.* (2006). "Ivermectin selection on β-tubulin: Evidence in *Onchocerca volvulus* and *Haemonchus contortus*." *Molecular and Biochemical Parasitology* **150**(2): 229-235.

ESPEN (2019). "Accelerating elimination of NTDs - Towards 2020." *Expanded Special Project for Elimination of Neglected Tropical Disease (ESPEN-AFRO)*. Accessed at: <http://espen.afro.who.int/>. Access date: 16th September, 2019.

Excoffier, L. and H. E. J. M. e. r. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." *Molecular Ecology Resources* **10**(3): 564-567.

Excoffier, L., *et al.* (1992). "Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data." *Genetics* **131**(2): 479-491.

Eyre-Walker, A. and P. Awadalla. (2001). "Does human mtDNA recombine?" *Journal of Molecular Evolution* **53**(4-5): 430–435.

Farmer, G. and T. Wigley (1985). "Climatic trends for tropical Africa." *Sustainable Agriculture and Natural Resource Management (SANREM) Knowledgebase Collection No. 3996*

Faure, E. and J.-P. Casanova (2006). "Comparison of chaetognath mitochondrial genomes and phylogenetical implications." *Mitochondrion* **6**(5): 258-262.

Faure, E., *et al.* (2011). "Probable presence of an ubiquitous cryptic mitochondrial gene on the antisense strand of the cytochrome oxidase I gene." *Biology Direct* **6**(1): 56.

Felsenstein, J. (1981). "A likelihood approach to character weighting and what it tells us about parsimony and compatibility." *Biological Journal of the Linnean Society* **16**(3): 183-196.

Feraday, R. M. and K. G. Leonhardt. (1989). "Absence of population structure in black flies as revealed by enzyme electrophoresis." *Genome* **32**(4): 531–537.

Ferris S. D, *et al.* (1983). "Flow of mitochondrial DNA across a species boundary." *Proceedings of the National Academy of Science of the USA*, **80**(8): 2290–2294.

Fiasorgbor, G., *et al.* (1992). "Distribution of *Simulium yahense* after larviciding activities in the southern extension of the Onchocerciasis Control Programme." *Tropical Medicine and Parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ)* **43**(1): 72-73.

Fiel, C., *et al.* (2001). "Resistance of Cooperia to ivermectin treatments in grazing cattle of the Humid Pampa, Argentina." *Veterinary Parasitology* **97**(3): 213-219.

Firestone, K. B. (2000). "Phylogenetic relationships among quolls revisited: the mtDNA control region as a useful tool." *Journal of Mammalian Evolution* **7**(1): 1-22.

Flaquer, A., *et al.* (2014). "Mitochondrial genetic variants identified to be associated with BMI in adults." *PloS One* **9**(8): e105116.

Flook, P. K. and R. J. Post. (1997). "Molecular population studies of *Simulium damnosum* s.l. (Diptera: Simuliidae) using a novel interspersed repetitive DNA marker." *Heredity* **79**(5): 531–540.

Fonseca, D. M. and D. D. Hart (1996). "Density-dependent dispersal of black fly neonates is mediated by flow." Oikos 75: 49-58.

Fontaine, B., *et al.* (1998). "Evolution of the relationship between near global and Atlantic SST modes and the rainy season in West Africa: statistical analyses and sensitivity experiments." Climate Dynamics 14(5): 353-368.

Force, A., *et al.* (1999). "Preservation of duplicate genes by complementary, degenerative mutations." Genetics 151(4): 1531-1545.

Fraser, P. and W. J. N. Bickmore (2007). "Nuclear organization of the genome and the potential for gene regulation." Nature 447(7143): 413-417.

Frati, F., *et al.* (1997). "Evolution of the mitochondrial cytochrome oxidase II gene in Collembola." Journal of Molecular Evolution 44(2): 145-158.

Fredeen, F. (1976). "The seven larval instars of *Simulium arcticum* (Diptera: Simuliidae)." The Canadian Entomologist 108(6): 591-600.

Fredeen, F. (1979). "Blackfly species adapt to change." Can. Agric, Pascal and Francis Bibliographic Databases. 24(1): 15-18.

Fredeen, F., *et al.* (1951). "Egg-laying habits, overwintering stages, and life-cycle of *Simulium arcticum* Mall.(Diptera: Simuliidae)." The Canadian Entomologist 83(3): 73-76.

Frempong, K. K., *et al.* (2016). "Does increasing treatment frequency address suboptimal responses to ivermectin for the control and elimination of river blindness?" Clinical Infectious Diseases 62(11): 1338-1347.

Frey, J. E. and B. Frey (2004). "Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of Thrips tabaci (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration?" Hereditas 140(2): 92-98.

Friesen, V., *et al.* (1999). "Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates." Molecular Ecology 8(12): 2147-2149.

Friesen, V., *et al.* (1997). "Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms." Molecular Ecology 6(11): 1047-1058.

Fu, Y.-X. and W.-H. Li (1999). "Coalescing into the 21st century: an overview and prospects of coalescent theory." Theoretical Population Biology 56(1): 1-10.

Fu, Y.-X. J. G. (1997). "Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection." Genetics **147**(2): 915-925.

Garesse, R. (1988). "*Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary considerations." Genetics, **118**(4): 649–663.

Garms, R. (1973). "Quantitative studies on the transmission of *Onchocerca volvulus* by *Simulium damnosum* in the Bong Range, Liberia." Zeitschrift fur Tropenmedizin und Parasitologie **24**(3): 358-372.

Garms, R. (1987). "Infection rates and parasitic loads of *Onchocerca volvulus*, and other filariae, in *Simulium sanctipauli* sl and *S. yahense* in a rain-forest area of Liberia." Tropical Medicine and Parasitology **38**(3): 201-204.

Garms, R. and R. Cheke. (1985). "Infections with *Onchocerca volvulus* in different members of the *Simulium damnosum* complex in Togo and Benin." Zeit Ange Zool **72**: 479-495.

Garms, R. and C. G. Vajime. (1975). "On the ecology and distribution of the species of the *simulium damnosum* complex in different bioclimatic zones of Liberia and Guinea." Tropenmedizin und Parasitologie **26**(3): 375-380.

Garms, R. and J. F. Walsh. (1987). "The migration and dispersal of black flies: *Simulium damnosum* s.l. the main vector of human onchocerciasis." In: Kim, K.C., Merritt, R.W. (Eds.), "Black Flies: Ecology, Population Management, and Annotated World List." Pennsylvania State University Press, University Park pp. 201–214.

Garms, R. and U. Zillmann. (1984). "Morphological identification of *Simulium sanctipauli* and *S. yahense* in Liberia and comparison of results with those of enzyme electrophoresis." Tropenmedizin und Parasitologie, **35**(4): 217–220.

Garms, R., et al. (1979). "Studies on the reinvasion of the Onchocerciasis Control Programme in the Volta River Basin by *Simulium damnosum* s.l. with emphasis on the south-western areas." **30**(3): 345-362.

Garms, R., et al. (1982). "The occurrence and movements of different members of the *Simulium damnosum* complex in Togo and Benin." Inist-CNRS ; INRA Database **69**: 219-236.

Garms, R., et al. (1989). "Seasonal extension of the breeding range of *Simulium sanctipauli* from forest into savanna in eastern Ghana and Togo." Zeitschrift Für Angewandte Zoologie **76**(4): 457-467.

Garms, R., et al. (1991). "A temporary focus of savanna species of the *Simulium damnosum* complex in the forest zone of Liberia." Tropical Medicine and Parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ) **42**(3): 181-187.

Garms, R., et al. (2015). "Assessments of the transmission of *Onchocerca volvulus* by *Simulium sanctipauli* in the Upper Denkyira District, Ghana, and the intermittent disappearance of the vector." Parasitology research **114**(3): 1129-1137.

Garms, R. (1981). "The reinvasion of the onchocerciasis control programme area in the Volta River Basin by *Simulium damnosum* sl., the involvement of the different cytospecies and epidemiological implications." Ann Soc Belg Med Trop **61**: 193-198.

Gattermann, N., et al. (1997). "Heteroplasmic point mutations of mitochondrial DNA affecting subunit I of cytochrome c oxidase in two patients with acquired idiopathic sideroblastic anemia." The Journal of the American Society of Hematology **90**(12): 4961-4972.

Geary, T. G. (2005). "Ivermectin 20 years on: maturation of a wonder drug." Trends in parasitology **21**(11): 530-532.

Ghatak, S., et al. (2016). "Polymorphism in mtDNA control region of Mizo-Mongloid breast Cancer samples as revealed by PCR-RFLP analysis." Mitochondrial DNA Part A **27**(3): 2205-2208.

Gilbert, M. T. P., et al. (2003). "Characterization of genetic miscoding lesions caused by postmortem damage." The American Journal of Human Genetics **72**(1): 48-61.

Gilbert, M. T. P., et al. (2007). "Whole-genome shotgun sequencing of mitochondria from ancient hair shafts." science **317**(5846): 1927-1930.

Gonzalez, P., et al. (2012). "Ivermectin in human medicine, an overview of the current status of its clinical applications." Current pharmaceutical biotechnology **13**(6): 1103-1109.

Goodall-Copestake, W., et al. (2012). "On the comparison of population-level estimates of haplotype and nucleotide diversity: a case study using the gene cox1 in animals." Heredity **109**(1): 50.

Goodenbour, J. M. and T. Pan. (2006). "Diversity of tRNA genes in eukaryotes." Nucleic Acids Research **34**(21): 6137-6146.

Grafen, A. (1989). The phylogenetic regression. Philos, TR Soc. B. **326**: 119–157.

Grant, W. N. (1994). "Genetic variation in parasitic nematodes and its implications." International Journal for Parasitology **24**(6): 821-830.

Grant, W. N., *et al.* (2006). "Heritable transgenesis of *Parastrongyloides trichosuri*: a nematode parasite of mammals." International Journal for Parasitology **36**(4): 475-483.

Gravel S, *et al.* (2011). "Demographic history and rare allele sharing among human populations." Proceedings of the National Academy of Sciences **108**(29): 11983–8.

Greaves, L. C., *et al.* (2006). "Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission." Proceedings of the National Academy of Sciences **103**(3): 714-719.

Green, R. and H. F. Noller (1997). "Ribosomes and translation." Annual Review of Biochemistry **66**(1): 679-716.

Griffiths, J. F. (1972). "Eastern Africa; Ethiopia highlands; the Horn of Africa. Climates of Africa, J. F. Griffiths, Ed.", Elsevier, 313–347.

Guelen, L., *et al.* (2008). "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions." Nature **453**(7197): 948.

Gustavsen, K., *et al.* (2011). "Onchocerciasis in the Americas: from arrival to (near) elimination." Parasites & vectors **4**(1): 205.

Güzelhan, C. and R. J. Garms (1991). "Cytogenetic comparison of *Simulium soubrense* populations in Liberia (Simuliidae, Diptera)." Zeitschrift für Zoologie **78**(2): 179-187.

Gwynne-Jones, D. (1978). "A new geography of Sierra Leone." Longman 1-212.

Gyan, E. T. (2013). "Parasitological Responses of *Onchocerca Volvulus* to Ivermectin Treatment and Genetic Analysis of Beta Tubulin Gene". University of Ghana Digital Collections, 1-143.

Hadis, M., *et al.* (2008). "Cytotaxonomy of *Simulium soderense* sp. nov. and a redescription of the 'Kulfo' form (Dunbar, 1969) of *Simulium damnosum* sl (Diptera: Simuliidae) as *Simulium kulfoense* sp. nov. in Ethiopia." International Journal of Tropical Insect Science **28**(1): 3-11.

Hall, L. R. and E. Pearlman (1999). "Pathogenesis of onchocercal keratitis (river blindness)." Clinical Microbiology Reviews **12**(3): 445-453.

Ham, P. and C. Gale (1984). "Blood meal enhanced *onchocerca* development and its correlation with fecundity in laboratory reared blackflies (Diptera, Simuliidae)." Tropenmedizin und Parasitologie **35**(4): 212-216.

Harding, R. M., et al. (1997). "Archaic African and Asian lineages in the genetic ancestry of modern humans." American Journal of Human Genetics **60**(4): 772.

Hare, M. P., et al. (1996). "Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves." Molecular Biology and Evolution **13**(2): 334-345.

Harris, E. E., et al. (1999). "X chromosome evidence for ancient human histories." Proceedings of the National Academy of Sciences **96**(6): 3320-3324.

Hartl, D. L. (1991). "Basic Genetics." Jones and Bartlett, Boston: 131-139.

Harvey, M. L., et al. (2003). "Mitochondrial DNA cytochrome oxidase I gene: potential for distinction between immature stages of some forensically important fly species (Diptera) in western Australia." Forensic Science International **131**(2-3): 134-139.

Hasegawa, M., et al. (1985). "Dating of the human-ape splitting by a molecular clock of mitochondrial DNA." Journal of Molecular Evolution **22**(2): 160-174.

Hassanin, A., et al. (2013). "Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: Application to the tribe Bovini (Cetartiodactyla, Bovidae)." Molecular Phylogenetics and Evolution **66**(3): 766-775.

Hebert, P. D. N. and T. R. Gregory (2005). "The Promise of DNA Barcoding for Taxonomy." Systematic Biology **54**(5): 852-859.

Hemmerter, S., et al. (2007). "A curious coincidence: mosquito biodiversity and the limits of the Japanese encephalitis virus in Australasia." BMC Evolutionary Biology **7**(1): 100.

Hensengerth, O. (2011). "Interaction of Chinese institutions with host governments in dam construction: the Bui Dam in Ghana." Evolution of Dam Policies. Springer, Berlin, Heidelberg, 2014. 229-271.

Hernandez-Triana, L. M., et al. (2012). "DNA barcodes reveal cryptic genetic diversity within the blackfly subgenus Trichodagmia Enderlein (Diptera: Simuliidae: *Simulium*) and related taxa in the New World." Zootaxa **3514**: 43-69.

Ho, S. Y. and M. T. P. Gilbert (2010). "Ancient mitogenomics." Mitochondrion **10**(1): 1-11.

Hodgkin, C., *et al.* (2007). "The future of onchocerciasis control in Africa." PLoS neglected tropical diseases **1**(1): e74.

Hotez, P., *et al.* (2007). "Recent progress in integrated neglected tropical disease control." Trends in parasitology **23**(11): 511-514.

Hougard, J.-M., *et al.* (2001). "Eliminating onchocerciasis after 14 years of vector control: a proved strategy." The Journal of infectious diseases **184**(4): 497-503.

Hougard, J.-M., *et al.* (1993). "Criteria for the selection of larvicides by the Onchocerciasis Control Programme in West Africa." Annals of Tropical Medicine & Parasitology **87**(5): 435-442.

Howell, S. B., *et al.* (2008). "Prevalence of anthelmintic resistance on sheep and goat farms in the southeastern United States." Journal of the American Veterinary Medical Association **233**(12): 1913-1919.

Howells, R. (1987). Dynamics of the filarial surface. Filariasis (Ciba Symposium 127), Wiley Online Library.

Hoy, M. A. (2002). "DNA amplification by the polymerase chain reaction: molecular biology made accessible. In: Insect Molecular Genetics: An Introduction to Principles and Applications, second ed. Academic Press, pp. 206–253."

Huang, Z. H. and F. Y. Tu. (2016). "Characterization and evolution of the mitochondrial DNA control region in Ranidae and their phylogenetic relationship." Genetics and Molecular Research **15**(3): gmr8491.

Hubby, J. L. and L. H. Throckmorton. (1965). "Protein differences in *Drosophila*. II. Comparative species genetics and evolutionary problems." Genetics **52**(1): 203–215.

Hudson, R. (1992). "Gene trees, species trees and the segregation of ancestral alleles." Genetics **131**(2): 509.

Hudson R. R. and M. Turelli. (2003). "Stochasticity overrules the ‘three-times rule’: genetic drift, genetic draft, and coalescent times for nuclear loci versus mitochondrial DNA." Evolution **57**(1): 182–190.

Huelsenbeck, J. P., *et al.* (1997). "Phylogeny estimation and hypothesis testing using maximum likelihood." Annual Review of Ecology and Systematics **28**(1): 437-466.

Hung, C.-M., *et al.* (2013). "The de novo assembly of mitochondrial genomes of the extinct passenger pigeon (*Ectopistes migratorius*) with next generation sequencing." *PLoS One* **8**(2): e56301.

Hurd, H. (2003). "Manipulation of medically important insect vectors by their parasites." *Annual Review of Entomology* **48**(1): 141-161.

Hurst, G. D. and F. M. Jiggins (2005). "Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts." *Proceedings of the Royal Society B: Biological Sciences* **272**(1572): 1525-1534.

Ibba, M. and D. Söll (2000). "Aminoacyl-tRNA synthesis." *Annual Review of Biochemistry* **69**(1): 617-650.

Ibm, C. J. A., NY: IBM Corp (2017). "IBM SPSS Statistics for Windows, Version Q3 25.0."

Ilmonen, J., *et al.* (2009). "The *Simulium vernalis* group (Diptera: Simuliidae) in Europe: multiple character sets for assessing species status." *Zoological Journal of the Linnean Society* **156**(4): 847-863.

Innan H, and M. Nordborg. (2002). "Recombination or mutational hot spots in human mtDNA?" *Molecular Biology and Evolution*, **19**(7): 1122–1127.

Irwin, J. A., *et al.* (2009). "Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples." *Journal of Molecular Evolution* **68**(5): 516-527.

Itoh, Y., *et al.* (2013). "Tertiary structure of bacterial selenocysteine tRNA." *Nucleic Acids Research* **41**(13): 6729-6738.

Jacobs-Lorena, M., *et al.* (1988). "Identification of species-specific DNA sequences in North American blackflies." *Tropical Medicine and Parasitology: Official Organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft Fur Technische Zusammenarbeit (GTZ)* **39**(1): 31–34.

Jahn, M., *et al.* (1991). "Anticodon and acceptor stem nucleotides in tRNAGln are major recognition elements for *E. coli* glutaminyl-tRNA synthetase." *Nature* **352**(6332): 258-260.

Jamandre B. W., *et al.* (2014). "High sequence variations in mitochondrial DNA control region among worldwide populations of flathead mullet *Mugil cephalus*." *International Journal of Zoology* 2014:1-9.

Janowiak, J. E. (1988). "An investigation of interannual rainfall variability in Africa." Journal of Climate **1**(3): 240-255.

Jiang, Y., et al. (2011). "Comparison of complete mitochondrial DNA control regions among five Asian freshwater turtle species and their phylogenetic relationships." Genetics Molecular Research **10**(3): 1545-1557.

Johnson, C., et al. (1985). "The pattern and speed of displacement of females of *Simulium damnosum* Theobald sl (Diptera: Simuliidae) across the Onchocerciasis Control Programme area of West Africa in 1977 and 1978." Bulletin of Entomological Research **75**(1): 73-92.

Johnson, T. P., et al. (2017). "Nodding syndrome may be an autoimmune reaction to the parasitic worm *Onchocerca volvulus*." Science Translational Medicine **9**(377): eaaf6953.

Jombart, T. (2008). "adegenet: a R package for the multivariate analysis of genetic markers." Bioinformatics **24**(11): 1403-1405.

Jombart, T., et al. (2010). "Discriminant analysis of principal components: a new method for the analysis of genetically structured populations." BMC Genetics **11**(1): 94.

Jukes, T. H. and C. R. Cantor (1969). "Evolution of protein molecules." Mammalian Protein Metabolism **3**(21): 132.

Kaessmann, H., et al. (1999). "DNA sequence variation in a non-coding region of low recombination on the human X chromosome." Nature Genetics **22**(1): 78-81.

Kalugina, N. (1991). "New Mesozoic Simuliidae and Leptoconopidae and blood-sucking origin in lower dipterans." Paleontologichesky Zhurnal **1991**: 69-80.

Kaplan, R. M. (2002). "Anthelmintic resistance in nematodes of horses." Veterinary Research **33**(5): 491-507.

Katabarwa, M. N., et al. (2012). "Transmission of onchocerciasis in Wadelai focus of northwestern Uganda has been interrupted and the disease eliminated." Journal of Parasitology Research **2012**.

Keating, J., et al. (2014). "Lymphatic filariasis and onchocerciasis prevention, treatment, and control costs across diverse settings: a systematic review." Acta Tropica **135**: 86-95.

Keddie, E. M., et al. (1998). "The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis." Molecular and Biochemical Parasitology **95**(1): 111-127.

Kengne, P., *et al.* (2007). "Molecular differentiation of three closely related members of the mosquito species complex, *Anopheles moucheti*, by mitochondrial and ribosomal DNA polymorphism." *Medical and Veterinary Entomology* **21**(2): 177-182.

Kim, K. C. (1987). "Black fliesecology, population management, and annotated world list." *International Conference on Ecology*. No. 595.771045 I5

Kimura, M. (1980). "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences." *Journal of Molecular Evolution* **16**(2): 111-120.

Kipp, W., *et al.* (1996). "The Nakalanga syndrome in Kabarole district, western Uganda." *The American Journal of Tropical Medicine and Hygiene* **54**(1): 80-83.

Knapp, M. and M. Hofreiter (2010). "Next generation sequencing of ancient DNA: requirements, strategies and perspectives." *Genes* **1**(2): 227-243.

Koch, R., *et al.* (2000). "Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*." *Genome Research* **10**(11): 1690-1696.

Kocher, T. D., *et al.* (1989). "Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers." *Proceedings of the National Academy of Sciences* **86**(16): 6196-6200.

Koella, J. C. (1999). "An evolutionary view of the interactions between anopheline mosquitoes and malaria parasites." *Microbes and Infection*. **1**(4): 303-308.

Kolesnikov, A. and E. Gerasimov (2012). "Diversity of mitochondrial genome organization." *Biochemistry* **77**(13): 1424-1435.

Kosakyan, A., *et al.* (2012). "COI barcoding of Nebelid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity and redefinition of the Hyalospheniidae Schultze." *Protist* **163**(3): 415-434.

Krauss, S. (2001). "Mitochondria: Structure and role in respiration." eLS John Wiley & Sons, Ltd.

Krueger, A. (2006a). "Guide to blackflies of the *Simulium damnosum* complex in eastern and southern Africa." *Medical and Veterinary Entomology*. **20**(1), 60–75.

Krueger, A. (2006b). "Zoophily and animal filariae in *Simulium damnosum* s.l.: on the indicative property of tarsal claws in females." *Acta Entomol Serbica* **206**(Suppl): 39–50.

Krueger, A. and I. C. Hennings. (2006). "Molecular phylogenetics of blackflies of the *Simulium damnosum* complex and cytophylogenetic implications." Molecular Phylogenetics and Evolution. **39**(1): 83–90.

Krüger, A., *et al.* (2000). "Molecular identification and phylogeny of East African *Simulium damnosum* sl and their relationship with West African species of the complex (Diptera: Simuliidae)." Insect Molecular Biology **9**(1): 101-108.

Krüger, A. and R. Garms. (1999). "Morphometric characterization of members of the *Simulium damnosum* Theobald complex (Diptera: Simuliidae) from East and West Africa." Annals of Tropical Medicine & Parasitology **93**(7): 753-761.

Kryukov A. P., *et al.* (2017). "Deep Phylogeographic breaks in magpie *Pica pica* across the Holarctic: concordance with bioacoustics and phenotypes." Zoological Science. **34**(3):185–200.

Kuhner, M. K., *et al.* (2000). "Maximum likelihood estimation of recombination rates from population data." Genetics **156**(3): 1393-1401.

Kumar, S., *et al.* (2018). "MEGA X: molecular evolutionary genetics analysis across computing platforms." Molecular Biology and Evolution **35**(6): 1547-1549.

Kutin, K., *et al.* (2004). "Efficiency of *Simulium sanctipauli* as a vector of *Onchocerca volvulus* in the forest zone of Ghana." Medical and Veterinary Entomology **18**(2): 167-173.

Lachance, J. and S. A. Tishkoff. (2013). "SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it." Bioessays **35**(9): 780-786.

Ladoukakis E. D. and E. Zouros. (2001). "Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA." Molecular Biology and Evolution **18**(7): 1168–1175.

Lamberton P. H., *et al.* (2014). "Onchocerciasis transmission in Ghana: biting and parous rates of host-seeking sibling species of the *Simulium damnosum* complex." Parasites & Vectors **7**(1):511.

Lamberton P. H, *et al.* (2015). "Onchocerciasis transmission in Ghana: persistence under different control strategies and the role of the simuliid vectors." PLoS Neglected Tropical Diseases **9**(4): e0003688.

La Morgia, C., *et al.* (2014). "Association of the mtDNA m. 4171C> A/MT-ND1 mutation with both optic neuropathy and bilateral brainstem lesions." BMC Neurology **14**(1): 116.

Landau, R. (1962). "Four forms of *Simulium tuberosum* (Lundstr.) in southern Ontario: a salivary gland chromosome study." Canadian Journal of Zoology **40**(5): 921-939.

Lanyon, S. M. (1992). "Interspecific brood parasitism in blackbirds (Icterinae): a phylogenetic perspective." Science **255**(5040): 77-79.

Lawrence, P. O. (2008). "Simulium spp., Vectors of *Onchocerca volvulus*: Life Cycle and Control." Encyclopedia of Entomology :3380-3382.

Lawton, J. H. (1991). "Ecology as she is done, and could be done." Oikos **61**(3): 289-290.

Lefoulon, E., et al. (2017). "Whence river blindness? The domestication of mammals and host-parasite co-evolution in the nematode genus *Onchocerca*." International Journal for Parasitology **47**(8): 457-470.

Lewin, B. J. (1975). "Units of transcription and translation: the relationship between heterogeneous nuclear RNA and messenger RNA." Cell **4**(1): 11-20.

Lewin, B. J. (1990). "A continuum of sequences includes structural genes." Genes IV. Oxford University Press, New York **857**.

Lewis, D. (1957). "Aspects of the structure, biology and study of *Simulium damnosum*." Annals of Tropical Medicine & Parasitology **51**(3): 340-358.

Lewis, D. (1961). "The *Simulium neavei* Complex (Diptera, Simuliidae) in Nyasaland." The Journal of Animal Ecology: 303-310.

Lewis, O., et al. (1995). "*Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons." Insect Molecular Biology **4**(4): 263-278.

Li, W.-H., et al. (2005). "Expression divergence between duplicate genes." TRENDS in Genetics **21**(11): 602-607.

Lischer, H. E. and L. J. B. Excoffier (2012). "PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs." Bioinformatics **28**(2): 298-299.

Livingstone, K. and L. Rieseberg (2004). "Chromosomal evolution and speciation: a recombination-based approach." New Phytologist **161**(1): 107-112.

Lopez, J. V., et al. (1994). "Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat." Journal of Molecular Evolution **39**(2): 174-190.

Lough, J. (1986). "Tropical Atlantic sea surface temperatures and rainfall variations in Subsaharan Africa." Monthly Weather Review **114**(3): 561-570.

Lu, G., et al. (2001). "Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*): relevance for speciation." Molecular Ecology **10**(4): 965–985.

Lucioli, S., et al. (2006). "Introducing a novel human mtDNA mutation into the Paracoccus denitrificans COX I gene explains functional deficits in a patient." Neurogenetics **7**(1): 51-57.

Ludwig, A., et al. (2000). "Heteroplasmy in the mtDNA control region of sturgeon (Acipenser, Huso and Scaphirhynchus)." Genetics **156**(4): 1933-1947.

Ludwig, B. and G. Schatz (1980). "A two-subunit cytochrome c oxidase (cytochrome aa3) from Paracoccus denitrificans." Proceedings of the National Academy of Sciences **77**(1): 196-200.

Ludwig, B. (1987). "Cytochrome c oxidase in prokaryotes." FEMS Microbiology Reviews **3**(1): 41-56.

Lyons, L. A., et al. (1997). "Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes." Nature Genetics **15**(1): 47-56.

Macey, J. R., et al. (1997). "Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome." Molecular Biology **14**(1): 91-104.

Magor, J. and L. J. Rosenberg (1980). "Studies of winds and weather during migrations of *Simulium damnosum* Theobald (Diptera: Simuliidae), the vector of onchocerciasis in West Africa." Bulletin of Entomological Research **70**(4): 693-716.

Maheu-Giroux, M. and S. A. J. T. L. I. D. Joseph (2018). "Moxidectin for deworming: from trials to implementation." The Lancet Infectious Diseases **18**(8): 817-819.

Maia-Herzog, M., et al. (1999). "Discovery of a new focus of human onchocerciasis in central Brazil." Transactions of the Royal Society of Tropical Medicine Hygiene **93**(3): 235-239.

Maizels, N. and A. M. J. C. S. H. M. S. Weiner (1999). "The genomic tag hypothesis: what molecular fossils tell us about the evolution of tRNA." Cold Spring Harbor Monograph Series **37**: 79-112.

Malatesta, F., et al. (1982). "Localization of cysteine115 in subunit III of beef heart cytochrome c oxidase to the C side of the mitochondrial inner membrane." Biochemical And Biophysical Research Communications **109**(4): 1180-1185.

Mallarino, R., *et al.* (2005). "Molecular systematics of the butterfly genus *Ithomia* (Lepidoptera: Ithomiinae): a composite phylogenetic hypothesis based on seven genes." *Molecular Phylogenetics And Evolution* **34**(3): 625-644.

Malmanche, N., *et al.* (2003). "The PRAT purine synthesis gene duplication in *Drosophila melanogaster* and *Drosophila virilis* is associated with a retrotransposition event and diversification of expression patterns." *Journal Of Molecular Evolution* **56**(5): 630-642.

Malmqvist, B., *et al.* (2004). "Vertebrate host specificity of wild-caught blackflies revealed by mitochondrial DNA in blood." *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**(suppl_4): S152-S155.

Malmqvist, B., *et al.* (2001). "Suspension feeders transform massive amounts of seston in large northern rivers." *Oikos* **92**(1): 35-43.

Malyarchuk, B. A., *et al.* (2002). "Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region." *Human Genetics* **111**(1): 46–53.

Management, E. R. (2007). Environmental and social impact assessment of the Bui Hydropower Project, Environmental Resources Management London. Final Report - Annex Volume, ERM Reference 0042911

Mank, R., *et al.* (2004). "A molecular marker for the identification of *Simulium squamosum* (Diptera: Simuliidae)." *Annals of Tropical Medicine & Parasitology* **98**(2): 197–208.

Mao, Y. S., *et al.* (2011). "Biogenesis and function of nuclear bodies." *Trends in Genetics* **27**(8): 295-306.

Mariottini, P., *et al.* (1983). "Antibodies against synthetic peptides reveal that the unidentified reading frame A6L, overlapping the ATPase 6 gene, is expressed in human mitochondria." *Cell* **32**(4): 1269-1277.

Marshall, F. and E. Hildebrand (2002). "Cattle before crops: the beginnings of food production in Africa." *Journal Of World Prehistory* **16**(2): 99-143.

Martin, A. P. (1995). "Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA." *Molecular Biology and Evolution* **12**(6): 1124–1131.

Massung, R. F., *et al.* (1998). "Nested PCR assay for detection of granulocytic ehrlichiae." *Journal of Clinical Microbiology* **36**(4): 1090-1095.

Maupas, E. (1900). *Modes et formes de reproduction des nematodes*. *Archives De Zoologie Experimentale Et Generale*, 8, 463-624.

May, B., et al. (1977). "Biochemical genetic variation in the family Simuliidae: Electrophoretic identification of the human biter in the isomorphic *Simulium jenningsi* group." Annals of the Entomological Society of America **70**(5): 637-640.

McCauley, D. E. (1995). "The use of chloroplast DNA polymorphism in studies of gene flow in plants." Trends in ecology & Evolution **10**(5): 198-202.

McDermott, J. M. and B. A. McDonald (1993). "Gene flow in plant pathosystems." Annual Review of Phytopathology **31**(1): 353-373.

McMahon, J. (1951). "The discovery of the early stages of *Simulium neavei* in phoretic association with crabs and a description of the pupa and the male." Bulletin of Entomological Research **42**(2): 419-426.

McMahon, J., et al. (1958). "The eradication of *Simulium neavei* from Kenya." Bulletin of the World Health Organization **19**(1): 75.

McManus, D. P., et al. (2010). "Schistosomiasis in the People's Republic of China: the era of the Three Gorges Dam." Clinical Microbiology Reviews **23**(2): 442-466.

McMillan, W. O. and S. R. Palumbi (1997). "Rapid rate of control-region evolution in Pacific butterflyfishes (Chaetodontidae)." Journal of Molecular Evolution **45**(5): 473-484.

Mebrahtu, Y., et al. (1986). "Characterization of *Simulium* (Edwardsellum) *damnosum* s.l. populations from six river systems in Kenya by cellulose acetate electrophoresis." Transactions of the Royal Society of Tropical Medicine and Hygiene **80**(6): 914-922.

Mehlhorn, H. (2008). " Encyclopedia of parasitology. (3rd ed.). Berlin: Springer. p. 646. ISBN 978-3-540-48994-8."

Meredith, S. E. O. and H. Townson. (1981). "Enzymes for species identification in the *Simulium damnosum* complex from West Africa." Tropenmedizin und Parasitologie. **32**(2): 123-128.

Meredith, S. E. O., et al. (1987). "Experimental hybridization between members of the *Simulium damnosum* complex." Medical and Veterinary Entomology **1**(2): 193-199.

Meyer, A. (1993). "Evolution in mitochondrial DNA in fishes." Biochemistry Molecular Biology of Fishes, 2. Molecular Biology Frontier: Elsevier Science 1993: 1-38.

Michael, N. L., *et al.* (1984). "All eight unassigned reading frames of mouse mitochondrial DNA are expressed." The EMBO journal **3**(13): 3165-3175.

Miescher, S. F. and D. J. G. S. Tsikata (2009). "Hydro-power and the promise of modernity and development in Ghana: comparing the Akosombo and Bui dam projects." **12**(13): 15-53.

Millar, C. D., *et al.* (2008). "New developments in ancient genomics." **23**(7): 386-393.

Mills, R. C. (1978). "The organization of rainfall in East Africa." Journal of Tropical Geography **47**(12): 40-50.

Molento, M., *et al.* (2008). Anthelmintic resistant nematodes in Brazilian horses, Vet Rec. **162**(12):384-385.

Molyneux, D. and D. J. P. Jefferies (1986). "Feeding behaviour of pathogen-infected vectors." Parasitology **92**(3): 721-736.

Montoya, J., *et al.* (1981). "Distinctive features of the 5'-terminal sequences of the human mitochondrial mRNAs." Nature **290**(5806): 465.

Moore, J. (1993). "Parasites and the behavior of biting flies." The Journal of Parasitology 1-16.

Moore, W. S. J. E. (1995). "Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees." Evolution **49**(4): 718-726.

Morales-Hojas, R., *et al.* (2002b). "Assessment of rDNA IGS as a molecular marker in the *Simulium damnosum* complex." Medical and Veterinary Entomology. **16**(4): 395–403.

Morin P.. A, *et al.* (2004). "SNPs in ecology, evolution and conservation." Trends in Ecology & Evolution **19**(4): 208–216.

Moritz, C. and C. Cicero (2004). "DNA barcoding: promise and pitfalls." biology **2**(10): e354.

Moritz, C., *et al.* (1987). "Evolution of animal mitochondrial DNA: relevance for population biology and systematics." Annual Review of Ecology and Systematics **18**(1): 269-292.

Moulton, J. K. (2000). "Molecular sequence data resolves basal divergences within Simuliidae (Diptera)." Systematic Entomology **25**(1): 95–113.

Mustapha, M., *et al.* (2005). "Incrimination of *Simulium thyolense* (Diptera: Simuliidae) as the anthropophilic blackfly in the Thyolo focus of human onchocerciasis in Malawi." Annals of Tropical Medicine Parasitology **99**(2): 181-192.

Namslauer, I. and P. Brzezinski (2009). "A mitochondrial DNA mutation linked to colon cancer results in proton leaks in cytochrome c oxidase." Proceedings of The National Academy of Sciences **106**(9): 3402-3407.

Near, T. J., *et al.* (2008). "Phylogenetics of notothenioid fishes (Teleostei: Acanthomorpha): inferences from mitochondrial and nuclear gene sequences." Molecular Phylogenetics and Evolution **47**(2): 832-840.

Nei, M. (1987). "Molecular evolutionary genetics", Columbia University Press. pp. 1-512.

Nei, M. and W.-H. Li (1979). "Mathematical model for studying genetic variation in terms of restriction endonucleases." Proceedings of the National Academy of Sciences **76**(10): 5269-5273.

Neigel, J. E. (1986). "Phylogenetic relationships of mitochondrial DNA under various models of speciation." Evolutionary Processes and Theory 515-534.

Nichols, R. (2001). "Gene trees and species trees are not the same." Trends in Ecology & Evolution **16**(7): 358–364.

Nicholls, R. S., *et al.* (2018). "Elimination of onchocerciasis from Colombia: first proof of concept of river blindness elimination in the world." Parasites & vectors **11**(1): 237.

Nicholson, S. (1996). "A review of climate dynamics and climate variability in Eastern Africa." Malawi Fisheries Repository 25-56.

Nigon, V. (1949). "Les modalités de la reproduction et le déterminisme du sexe chez quelques nématodes libres." Ann. Sci. Nat. **11**: 1-132.

Nittinger, F., *et al.* (2005). "Out of Africa? Phylogenetic relationships between Falco biarmicus and the other hierofalcons (Aves: Falconidae)." Journal of Zoological Systematics and Evolutionary Research **43**(4): 321-331.

Noamesi, G. K. (1966). "Dry season survival and associated longevity and flight range of *Simulium dtunnosum* Theobald in Northern Ghana." Ghana Medical Journal, S: 95-102.

Noma, M., *et al.* (2002). "Rapid epidemiological mapping of onchocerciasis (REMO): its application by the African Programme for Onchocerciasis Control (APOC)." Annals of Tropical Medicine & Parasitology **96**(sup1): S29-S39.

Nontasut, P., *et al.* (2000). "Comparison of ivermectin and albendazole treatment for gnathostomiasis." Southeast Asian journal of tropical medicine and public health **31**(2): 374-377.

Nosek, J., *et al.* (1998). "Linear mitochondrial genomes: 30 years down the line." Trends in Genetics **14**(5): 184-188.

Ohi, S., *et al.* (1978). "Mapping of mitochondrial 4S RNA genes in *Xenopus laevis* by electron microscopy." Journal of molecular biology **121**(3): 299-310.

Ojala, D., *et al.* (1981). "tRNA punctuation model of RNA processing in human mitochondria." Nature **290**(5806): 470.

Okoampa-Ahoofe, K. (2009). "The under-told story of Akosombo-The Gbedemah factor (Part 1)." Statesman Newspaper. Retrieved from http://www.thestatesmanonline.com/pages/news_detail.php

Omura, S. (2008). "Ivermectin: 25 years and still going strong." International journal of antimicrobial agents **31**(2): 91-98.

Omura, S. and A. Crump (2014). "Ivermectin: panacea for resource-poor communities?" Trends in parasitology **30**(9): 445-455.

Omura, S. and A. Crump (2017). "Ivermectin and malaria control." Malaria journal **16**(1): 172.

Ono, H., *et al.* (2006). "Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera." Developmental Biology **298**(2): 555-570.

Osei-Atweneboana, M. Y., *et al.* (2011). "Phenotypic evidence of emerging ivermectin resistance in *Onchocerca volvulus*." PLoS Neglected Tropical Diseases **5**(3): e998.

Osei-Atweneboana, M. Y., *et al.* (2007). "Prevalence and intensity of *Onchocerca volvulus* infection and efficacy of ivermectin in endemic communities in Ghana: a two-phase epidemiological study." The Lancet **369**(9578): 2021-2029.

Ottesen, E. A. and W. Campbell (1994). "Ivermectin in human medicine." Journal of antimicrobial chemotherapy **34**(2): 195-203.

PAHO (2008). "Resolution CD48.R12. Towards the elimination of onchocerciasis (river blindness) in the Americas." Pan American Health Organization.

PAHO (2009). "Resolution CD49.R19. Elimination of neglected diseases and other poverty-related infections." Pan American Health Organization. Accessed at: <https://www.paho.org/hq/dmdocuments/2009/CD49.R19-Eng.pdf?ua=1>. Accessed at: <https://www.paho.org/hq/dmdocuments/2009/CD49.R19-Eng.pdf?ua=1>. Access date: 11th August, 2019.

PAHO (2016). "Resolution CD55.R9. Plan of action for the elimination of neglected infectious diseases and postelimination actions 2016–2022. ." Accessed at: <https://www.paho.org/hq/dmdocuments/2016/CD55-R9-e.pdf>. Access date: 19th September, 2019.

Palmer, M., *et al.* (1994). "Genome size and organization in the ixodid tick *Amblyomma americanum* (L.)." Insect Molecular Biology **3**(1): 57-62.

Palumbi, S. R., *et al.* (1994). "Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales." Molecular Biology and Evolution **11**(3): 426-435.

Palumbi, S. R., *et al.* (2001). "Predicting nuclear gene coalescence from mitochondrial data: the three-times rule." Evolution **55**(5): 859-868.

Palumbi, S. R. J. M. s. (1996). "Nucleic acids II: the polymerase chain reaction." Molecular systematics 205-247.

Pamilo, P. and M. Nei. (1988). "Relationships between gene trees and species trees." Molecular Biology and Evolution **5**(5): 568-583.

Pandya, A., *et al.* (1999). "Heterogenous point mutations in the mitochondrial tRNA Ser (UCN) precursor coexisting with the A1555G mutation in deaf students from Mongolia." The American Journal of Human Genetics **65**(6): 1803-1806.

Papich, M. G. (2016). "Saunders Handbook of Veterinary Drugs: Small and Large Animal (4 ed.)", Elsevier Health Sciences. p. 420. Accessed from: <https://books.google.co.jp/books?id=SJvmCgAAQBAJ&pg=PA420#v=onepage&q&f=false> . Access Date: 14th Oct 2019

Parra, G., *et al.* (2007). "CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes." Bioinformatics **23**(9): 1061-1067.

Parra, G., *et al.* (2008). "Assessing the gene space in draft genomes." Nucleic acids research **37**(1): 289-297.

Penny, D., *et al.* (1995). "Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*." Molecular Biology and Evolution **12**(5): 863–882.

Pereira, F., et al. (2008). "Evidence for variable selective pressures at a large secondary structure of the human mitochondrial DNA control region." Molecular biology evolution **25**(12): 2759-2770.

Pfenninger M., et al. (2003). "Ecological and morphological differentiation among cryptic evolutionary lineages in freshwater limpets of the nominal form-group *Ancylus fluviatilis* (OF. Müller, 1774)." Molecular Ecology **12**(10): 2731–2745.

Phayuhasena, S., et al. (2010). "Phylogenetic relationships among the black fly species (Diptera: Simuliidae) of Thailand based on multiple gene sequences." Genetica **138**(6): 633-648.

Pickersgill, H., et al. (2006). "Characterization of the *Drosophila melanogaster* genome at the nuclear lamina." Nature Genetics **38**(9): 1005.

Pion, S., et al. (2002). "Excess mortality associated with blindness in the onchocerciasis focus of the Mbam Valley, Cameroon." Annals of Tropical Medicine & Parasitology **96**(2): 181-189.

Pion, S. D., et al. (2019). "Effect of a single standard dose (150–200 µg/kg) of ivermectin on *Loa loa* microfilaremia: systematic review and meta-analysis." Open Forum Infectious Diseases, Oxford University Press US. Vol. 6, No. 4, p. ofz019

Plaisier, A., et al. (1991). "The reproductive lifespan of *Onchocerca volvulus* in West African savanna." Acta Tropica **48**(4): 271-284.

Posada, D. and K. A. Crandall (1998). "Modeltest: testing the model of DNA substitution." Bioinformatics (Oxford, England) **14**(9): 817-818.

Posada, D., et al. (2001). "Intraspecific gene genealogies: trees grafting into networks." Trends in Ecology & Evolution **16**(1): 37-45.

Post, R. J. and J. M. Crampton. (1988). "The taxonomic use of variation in repetitive DNA sequences in the *Simulium damnosum* complex." In: Service, M.W. (Ed.), "Biosystematics of Haematophagous Insects." Systematics Association Special Volumes, Oxford University Press **37**: 245–256.

Post, R. J. and P. Flook. (1992). "DNA probes for the identification of members of the *Simulium damnosum* complex (Diptera: Simuliidae)." Medical and Veterinary Entomology. **6**(4): 379–384.

Post, R. J. and R. Crosskey (1985). "The distribution of the *Simulium damnosum* complex in Sierra Leone and its relation to onchocerciasis." Annals of Tropical Medicine & Parasitology 79.2 (1985): 169-194.

Post, R. J. (1984). "Natural interspecific hybridisation of *Simulium sanctipauli* s.l. with *Simulium squamosum* and *Simulium yahense* (Diptera: Simuliidae)." Tropenmedizin und Parasitologie 35(1): 58–60.

Post, R. J. (1986). "The cytotaxonomy of *Simulium sanctipauli* and *Simulium soubrense* (Diptera: Simuliidae)." Genetica 69(3): 191-207.

Post, R. J., et al. (2007). "Taxonomy and inventory of the cytospecies and cytotypes of the *Simulium damnosum* complex (Diptera: Simuliidae) in relation to onchocerciasis." Tropical Medicine & International Health 12(11): 1342-1353.

Post, R. J., et al. (2013). "Stability and change in the distribution of cytospecies of the *Simulium damnosum* complex (Diptera: Simuliidae) in southern Ghana from 1971 to 2011." Parasites vectors 6(1): 205.

Poulakakis, N., et al. (2008). "Historical DNA analysis reveals living descendants of an extinct species of Galápagos tortoise." Proceedings of the National Academy of Sciences 105(40): 15464-15469.

Powell, J. R. (1983). "Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*." Proceedings of the National Academy of Sciences of the USA 80(2): 492–495.

Pramual, P. and P. H. Adler (2014). "DNA barcoding of tropical black flies (Diptera: Simuliidae) of Thailand." Molecular Ecology Resources 14(2): 262-271.

Pramual, P., et al. (2011a). "Phylogeography of *Simulium siamense* Takaoka and Suzuki complex (Diptera: Simuliidae) in Thailand." Entomological Science 14(4): 428-436.

Pramual, P. and C. Kuvangkadilok (2009a). "Agricultural land use and black fly (Diptera, Simuliidae) species richness and species assemblages in tropical streams, Northeastern Thailand." Hydrobiologia 625(1): 173-184.

Pramual, P., et al. (2005). "Phylogeography of the black fly *Simulium tani* (Diptera: Simuliidae) from Thailand as inferred from mtDNA sequences." Molecular Ecology 14(13): 3989-4001.

Pramual, P. and C. J. G. Kuvangkadilok (2012). "Integrated cytogenetic, ecological, and DNA barcode study reveals cryptic diversity in *Simulium* (Gomphostilbia) angulistylum (Diptera: Simuliidae)." Genome 55(6): 447-458.

Pramual, P. and C. J. H. Kuvangkadilok (2009b). "Agricultural land use and black fly (Diptera, Simuliidae) species richness and species assemblages in tropical streams, Northeastern Thailand." Hydrobiologia **625**(1): 173-184.

Pramual, P. and P. Nanork (2012). "Phylogenetic analysis based on multiple gene sequences revealing cryptic biodiversity in *Simulium multistriatum* group (Diptera: Simuliidae) in Thailand." Entomological Science **15**(2): 202-213.

Pramual, P. and K. Wongpakam (2013). "Population genetics of the high elevation black fly *Simulium (N evermannia) feuerborni* Edwards in Thailand." Entomological Science **16**(3): 298-308.

Pramual, P., et al. (2010). "Cryptic biodiversity and phylogenetic relationships revealed by DNA barcoding of Oriental black flies in the subgenus Gomphostilbia (Diptera: Simuliidae)." Genome **54**(1): 1-9.

Prichard, R. (1994). "Anthelmintic resistance." Veterinary parasitology **54**(1-3): 259-268.

Procunier, W. (1989). "Cytological approaches to simuliid biosystematics in relation to the epidemiology and control of human onchocerciasis." Genome **32**(4): 559-569.

Prod'hon, J., et al. (1982). "Influence of the human and vector migrations on the strategy of the onchocerciasis control programme in the Volta river basin. 1. Return of emigrants living in the forest zones into their native savanna zone [*Simulium damnosum* *Onchocerca vulvulus*; Mali]." Cahiers ORSTOM Serie Entomologie Medicale et Parasitologie **20**(1): v.285-298(4)

Pruess, K. P., et al. (2000). "Utility of the mitochondrial cytochrome oxidase II gene for resolving relationships among black flies (Diptera: Simuliidae)." Molecular Phylogenetics and Evolution **16**(2): 286-295.

Purvis, A. and T. Garland (1993). "Polytomies in comparative analyses of continuous characters." Systematic Biology **42**(4): 569-575.

Quattro, J. and W. J. C. Jones (1999). "Amplification primers that target locus-specific introns in actinopterygian fishes." Copeia **1999**(1): 191-196.

Quillévéré, D., et al. (1977). "Etude du complexe *Simulium damnosum* en Afrique de l'ouest: VII. Etude de la bioécologie et du pouvoir vecteur des femelles de *S. sanctipauli*, *S. soubrense* et de *S. yahense* en Côte d'Ivoire. Identification morphologique des femelles en Côte d'Ivoire." Cahiers ORSTOM Sér Entomol Méd Parasitol **15**: 301-329.

Ragan, C. I. (1987). "Structure of NADH-ubiquinone reductase (complex I)." Current Topics in Bioenergetics, Elsevier **15**: 1-36.

Raitio, M., et al. (1987). "Isolation and analysis of the genes for cytochrome c oxidase in *Paracoccus denitrificans*." The EMBO Journal **6**(9): 2825-2833.

Ramirez, J. L. and I. B. Dawid (1978). "Mapping of mitochondrial DNA in *Xenopus laevis* and *X. borealis*: the positions of ribosomal genes and D-loops." Journal of Molecular Biology **119**(1): 133-146.

Rand, D. M. and L. M. Kann. (1998). "Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA." In: Woodruff R.C., Thompson J.N. (eds) *Mutation and Evolution. Contemporary Issues in Genetics and Evolution*, Springer, Dordrecht **1998**(7): 393-407.

Randi E, et al. (2004). "Phylogeography of roe deer (*Capreolus capreolus*) population: the effects of historical genetic subdivisions and recent nonequilibrium dynamics." Molecular Ecology, **13**(10): 3071–3083.

Rastl, E. and I. B. Dawid (1979). "Expression of the mitochondrial genome in *Xenopus laevis*: a map of transcripts." Cell **18**(2): 501-510.

Raybould, J. N. and G. B. White (1979). "The distribution, bionomics and control of onchocerciasis vectors (Diptera: Simuliidae) in Eastern Africa and the Yemen." Tropenmedizin und Parasitologie **30**(4): 505-547.

Rebollo, M. P., et al. (2018). "Onchocerciasis: shifting the target from control to elimination requires a new first-step—elimination mapping." International Health **10**(suppl_1): i14-i19.

Rech, G. E., et al. (2014). "Natural selection on coding and noncoding DNA sequences is associated with virulence genes in a plant pathogenic fungus." Genome biology evolution **6**(9): 2368-2379.

Remme, J. (1995). The African programme for onchocerciasis control: preparing to launch, Elsevier **11**(11): 403-406.

Remme, J., et al. (1989). "A community trial of ivermectin in the onchocerciasis focus of Asubende, Ghana. I. Effect on the microfilarial reservoir and the transmission of *Onchocerca volvulus*." Tropical medicine parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft of Deutsche Gesellschaft fur Technische Zusammenarbeit **40**(3): 367-374.

Richards Jr, F., et al. (2015). "One hundred years after its discovery in Guatemala by Rodolfo Robles, *Onchocerca volvulus* transmission has been eliminated from the central endemic zone." The American journal of tropical medicine and hygiene **93**(6): 1295-1304.

Richards Jr, F. O., *et al.* (2001). "Control of onchocerciasis today: status and challenges." Trends in parasitology **17**(12): 558-563.

Richly, E. and D. Leister (2004). "NUMTs in sequenced eukaryotic genomes." Molecular Biology and Evolution **21**(6): 1081-1084.

Ridley, H. (1945). "Ocular Onchocerciasis, including an Investigation in the Gold Coast." Ophthalmol.

Rivera, J. and D. Currie (2009). "Identification of Nearctic black flies using DNA barcodes (Diptera: Simuliidae)." Molecular Ecology Resources **9**: 224-236.

Roberts, J., *et al.* (1967). "Onchocerciasis in Kenya 9, 11 and 18 years after elimination of the vector." Bulletin of the World Health Organization **37**(2): 195.

Roberts, M. (1990). "Vectors of onchocerciasis in the Thyolo Highlands and other onchocerciasis foci in Malawi." Acta Leidensia **59**(1&2): 45-47.

Rodríguez-Pérez, M.A., *et al.* (2006). "Analysis of genetic variation in ribosomal DNA internal transcribed spacer and the NADH Dehydrogenase subunit 4 mitochondrial genes of the onchocerciasis vector *Simulium ochraceum*." Journal of Medical Entomology. **43**(4): 701-706.

Rodríguez-Pérez, M. A., *et al.* (2015). "Elimination of onchocerciasis from Mexico." PLoS Neglected Tropical Diseases **9**(7): e0003922.

Rognon, X. and R. Guyomard. (2003). "Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa." Molecular Ecology **12**(2): 435–445.

Rohland, N., *et al.* (2004). "Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens." Biotechniques **36**(5): 814-821.

Ronquist, F., *et al.* (2012). "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space." Systematic Biology **61**(3): 539-542.

Rothfels, K. (1979). "Cytotaxonomy of black flies (Simuliidae)." Annual Review of Entomology **24**(1): 507-539.

Rothfels, K. (1989). "Speciation in black flies." Genome **32**(4): 500-509.

Rothova, A., *et al.* (1989). "Side-effects of ivermectin in treatment of onchocerciasis." The Lancet **333**(8652): 1439-1441.

Rowe, K. C., *et al.* (2011). "Museum genomics: low-cost and high-accuracy genetic data from historical specimens." Molecular Ecology Resources **11**(6): 1082-1092.

Rowther, F. B., *et al.* (2012). "TOUCH-UP gradient amplification method." Journal of Biomolecular Techniques JBT **23**(1): 1-3.

Rubinoff, D., *et al.* (2006). "A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification." Journal of heredity **97**(6): 581-594.

Russo, C., *et al.* (1996). "Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny." Molecular Biology and Evolution **13**(3): 525-536.

Ruvolo, M., *et al.* (1991). "Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence." Proceedings of the National Academy of Sciences **88**(4): 1570-1574.

Sakamoto, Y., *et al.* (1986). "Akaike information criterion statistics." Dordrecht, The Netherlands: D. Reidel **81**.

Salvato, P., *et al.* (2008). "The complete mitochondrial genome of the bag-shelter moth Ochrogaster lunifer (Lepidoptera, Notodontidae)." BMC genomics **9**(1): 331.

San Mauro, D., *et al.* (2005). "A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome." Molecular Biology and Evolution **23**(1): 227-234.

Sauerbrey, M. (2008). "The onchocerciasis elimination program for the Americas (OEPA)." Annals of Tropical Medicine & Parasitology **102**(sup1): 25-29.

Sauerbrey, M., *et al.* (2018). "Progress toward elimination of onchocerciasis in the Americas." International health **10**(suppl_1): i71-i78.

Savolainen, V., *et al.* (2005). "Towards writing the encyclopaedia of life: an introduction to DNA barcoding." Philosophical Transactions of the Royal Society B: Biological Sciences **360**(1462): 1805-1811.

Scarpassa, V. M. and N. Hamada. (2003). "Isozyme variation in four species of the *Simulium perflavum* species group (Diptera: Simuliidae) from the Brazilian Amazon." Genetics and Molecular Biology. **26**(1): 39-46.

Scherr, K. M. (2012). "Legal implications of Sudan's Separation: the Question of Citizenship." Separation: 100.

Schierup, M. H. and J. J. G. Hein (2000). "Consequences of recombination on traditional phylogenetic analysis." Genetics **156**(2): 879-891.

Schiller, E. L., *et al.* (1984). "Morphogenesis of larval *Onchocerca volvulus* in the Panamanian black fly, *Simulium quadriannulatum*." The American journal of tropical medicine and hygiene **33**(3): 410-413.

Schulz-Key, H. and M. Karam (1986). "Periodic reproduction of *Onchocerca volvulus*." Parasitology Today **2**(10): 284-286.

Schwarz, G. (1978). "Estimating the dimension of a model." The Annals of Statistics **6**(2): 461-464.

Science, T. C. (The *C. elegans* Sequencing Consortium) (1998). "Genome sequence of the nematode *C. elegans*: a platform for investigating biology." Science 2012-2018.

Scientific, T. F. (2015). "Qubit dsDNA HS Assay Kits—User Guide." **32851**.

Scott, J. A., *et al.* (1993). "Identification of single specimens of the *Anopheles gambiae* complex by polymerase chain reaction." The American Journal of Tropical Medicine and Hygiene **49**(4): 520–529.

Seehausen, O., *et al.* (2003). "Nuclear markers reveal unexpected genetic variation and a Congolese-Nilotic origin of the Lake Victoria cichlid species flock." Proceedings of the Royal Society of London Series B Biological Science **270**(1511): 129–137.

Seehausen, O. J. (2006). "African cichlid fish: a model system in adaptive radiation research." Biological Sciences **273**(1597): 1987-1998.

Senatore, G., *et al.* (2014). "Molecular systematics of the *Simulium jenningsi* species group (Diptera: Simuliidae), with three new fast-evolving nuclear genes for phylogenetic inference." Molecular Phylogenetics and Evolution **75**: 138-148.

Shao, R. and S. C. Barker (2003). "The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes." Molecular Biology and Evolution **20**(3): 362-370.

Shao, R., *et al.* (2005). "Evolution of duplicate control regions in the mitochondrial genomes of metazoa: a case study with Australasian *Ixodes* ticks." Molecular Biology and Evolution **22**(3): 620-629.

Shapiro, B., *et al.* (2002). "Flight of the dodo." science **295**(5560): 1683-1683.

Sharp, S. J., *et al.* (1985). "Structure and transcription of eukaryotic tRNA gene." Critical Reviews in Biochemistry **19**(2): 107-144.

Shaw, K. L. (2002). "Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets." Proceedings of the National Academy of Sciences of the USA **99**(25): 16122-16127.

Shelley, A. (1991). "Simuliidae and the transmission and control of human onchocerciasis in Latin America." Cadernos De Saude Publica **7**: 310-327.

Shelley, A. J., *et al.* (2010). "The blackflies (Diptera: simuliidae) of Brazil." Aquatic Biodiversity in Latin America **6**(2010): 821.

Shewell, G. (1957). "Interim report on distributions of the black flies (Simuliidae) obtained in the Northern Insect Survey." Canadian Department of National Defence, Defence Research Board Environmental Protection Technical Report No 7.

Silva-Brandaو, K. L., *et al.* (2005). "Phylogenetic relationships of the New World Troidini swallowtails (Lepidoptera: Papilionidae) based on COI, COII, and EF-1 α genes." Molecular Phylogenetics and Evolution **36**(3): 468-483.

Singh, T. R., *et al.* (2008). "Bird mitochondrial gene order: insight from 3 warbler mitochondrial genomes." Molecular Biology and Evolution **25**(3): 475-477.

Slade, R. W., *et al.* (1994). "Multiple nuclear-gene phylogenies: application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny." Molecular Biology and Evolution **11**(3): 341-356.

Slonimski, P., *et al.* (1982). "Mitochondrial Genes", Cold Spring Harbor Laboratory. 12: 477-481.

Smith, R., *et al.* (1988). "Vascular perfusion of *Onchocerca volvulus* nodules." Tropical Medicine and Parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft of Deutsche Gesellschaft fur Technische Zusammenarbeit **39**: 418-421.

Snyder, T. P. (1982). "Electrophoretic characterizations of black flies in the *Simulium venustum* and *verecundum* species complexes (Diptera: Simuliidae)." The Canadian Entomologist. **114**(6): 503-507.

Snyder, T. P. and M. C. Linton. (1983). "Electrophoretic and morphological separation of *Prosimulium fuscum* and *P. mixtum* larvae (Diptera: Simuliidae)." The Canadian Entomologist **115**(1): 81–87.

Soboslay, P., *et al.* (1987). "Ivermectin effect on microfilariae of *Onchocerca volvulus* after a single oral dose in humans." Tropical Medicine and Parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft of Deutsche Gesellschaft fur Technische Zusammenarbeit **38**(1): 8-10.

Someren, V. V. and J. McMahon (1950). "Phoretic association between Afronurus and *Simulium* species and the discovery of the early stages of *Simulium neavei* on freshwater crabs." Nature, Lond **166**: 350-351.

Song, H., *et al.* (2008). "Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified." Proceedings of the National Academy of Sciences **105**(36): 13486-13491.

Sota, T. and A. P. Vogler. (2001). "Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*." Systematic Biology **50**(1): 39-59.

Sota, T. (2002). "Radiation and reticulation: extensive introgressive hybridization in the carabid beetles *Ohomopterus* inferred from mitochondrial gene genealogy." Population Ecology **44**(3): 145–156.

Spradling, A. C. and G. M. Rubin (1981). "*Drosophila* genome organization: conserved and dynamic aspects." Annual Review of Genetics **15**(1): 219-264.

Sprinzl, M. and F. Cramer (1979). The-CCA end of tRNA and its role in protein biosynthesis. Progress in nucleic acid research and molecular biology, Elsevier. **22**: 1-69.

Sprinzl, M. and K. S. Vassilenko (2005). "Compilation of tRNA sequences and sequences of tRNA genes." Nucleic Acids Research **33**(suppl_1): D139-D140.

Stephan, W., *et al.* (1998). "A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*." Proceedings of the National Academy of Sciences **95**(10): 5649-5654.

Stewart, M. K., *et al.* (2005). "High genetic diversity in the chemoreceptor superfamily of *Caenorhabditis elegans*." Genetics **169**(4): 1985-1996.

Stone, A. (1952). "The Simuliidae of Alaska." Proc. Entomol. Soc. Wash **54**: 69-96.

Strand, A., et al. (1997). "Nuclear DNA-based markers for plant evolutionary biology." Molecular Ecology **6**(2): 113-118.

Sturchio, J. L. (2001). "The case of ivermectin: lessons and implications for improving access to care and treatment in developing countries." Community Eye Health **14**(38): 22.

Surtees, D. (1986). "A New Cytotype *Simulium sanctipauli* (Diptera, Simuliidae) From Togo." Transactions of the Royal Society of Tropical Medicine and Hygiene, 80(2): 1-343.

Surtees, D., et al. (1988). "The cytotaxonomy of the Djodji form of *Simulium sanctipauli* (Diptera: Simuliidae)." Tropical Medicine and Parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ) **39**(2): 120-122.

Surtees, D. P. (1988). "Multivariate morphometrics and cytotaxonomy of the *Simulium damnosum* complex." PhD Thesis, University of Liverpool, U.K.

Takaoka, H. (1996). "The geographical distribution of the genus *Simulium Latreille* in the Oriental and Australasian regions." Japanese Journal of Tropical Medicine and Hygiene **24**(2): 113-124.

Takaoka, H., et al. (2003). "Black flies (Diptera: Simuliidae) attracted to humans and water buffalos and natural infections with filarial larvae, probably *Onchocerca* sp., in northern Thailand." Parasite **10**(1): 3-8.

Takaoka, H. and H. SuuKi (1984). "The blackflies (Diptera: Simuliidae) from Thailand." Medical Entomology and Zoology **35**(1): 7-45.

Takaoka, H., et al. (1986a). "Experimental infections of three Guatemalan blackfly species with north Venezuelan *Onchocerca volvulus*." Medical Entomology and Zoology **37**(4): 319-323.

Takaoka, H., et al. (1986b). "cross-compatibility study of Guatemalan and north Venezuelan *Onchocerca vulvulus* to *Simulium metallicum* from two countries." Japanese Journal of Parasitology.

Tamura, K. (1992). "Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases." Molecular Biology and Evolution **9**(4): 678-687.

Tang, J., et al. (1995). "The *Simulium damnosum* species complex: phylogenetic analysis and molecular identification based upon mitochondrial encoded gene sequences." Insect Molecular Biology. **4**(2): 79-88.

Tang, J., et al. (1996a). "Intra-specific heterogeneity of the rDNA internal transcribed spacer in the *Simulium damnosum* (Diptera: Simuliidae) complex." Molecular Biology and Evolution. **13**(1): 244–252.

Tang, J., et al. (1996b). "Genotyping North American black flies by means of mitochondrial ribosomal RNA sequences." Canadian Journal of Zoology. **74**(1): 39–46.

Tangkawanit, U., et al. (2009). "Cytosystematics of the *Simulium tuberosum* group (Diptera: Simuliidae) in Thailand." Zoological Journal of the Linnean Society **155**(2): 289–315.

Tavaré, S. J. L. (1986). "Some probabilistic and statistical problems in the analysis of DNA sequences." Lectures on Mathematics in the Life Sciences **17**(2): 57–86.

Taylor, E. B. and J. D. McPhail. (2000). "Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*." Proceedings of the Royal Society of London Series B Biological Science, **267**(1460): 2375–2384.

Tekle, A. H., et al. (2012). "Impact of long-term treatment of onchocerciasis with ivermectin in Kaduna State, Nigeria: first evidence of the potential for elimination in the operational area of the African Programme for Onchocerciasis Control." Parasites & Vectors **5**(1): 28.

Tekle, A. H., et al. (2016). "Progress towards onchocerciasis elimination in the participating countries of the African Programme for Onchocerciasis Control: epidemiological evaluation results." Infectious Diseases of Poverty **5**(1): 66.

Templeton, A. R. (1998). "Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history." Molecular Ecology Resources **7**(4): 381–397.

Templeton, A. R., et al. (2000). "Cladistic structure within the human lipoprotein lipase gene and its implications for phenotypic association studies." Genetics **156**(3): 1259–1275.

Terminski, B. (2014). "Development-induced displacement and resettlement: Causes, consequences, and socio-legal context", Columbia University Press.

Thaijarern, J., et al. (2014). "Phylogeography of the black fly *Simulium aureohirtum* (Diptera: Simuliidae) in Thailand." Genome **57**(8): 411–417.

Thanwisai, A., et al. (2006). "Molecular phylogeny of black flies (Diptera: Simuliidae) from Thailand, using ITS2 rDNA." Genetica **128**(1–3): 177–204.

Thao, M. L., et al. (2004). "Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha)." BMC Evolutionary Biology **4**(1): 25.

Thompson, B. H., *et al.* (1972). "A marking and recapture experiment on *Simulium damnosum* and bionomic observations." (Unpublished document, WHO/ONCH0/72.98).

Thompson, J. D., *et al.* (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Research **22**(22): 4673-4680.

Thomson, M.C., *et al.* (1996). "The unusual occurrence of savanna members of the *Simulium damnosum* species complex (Diptera: Simuliidae) in southern Sierra Leone in 1988." Bulletin of Entomological Research. **86**(3): 271–280.

Thorburn D.R., *et al.* (2017). "Mitochondrial DNA-Associated Leigh Syndrome and NARP". In Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJ, Bird TD, Dolan CR, Fong CT, Smith RJ, Stephens K (eds.). GeneReviews [Internet]. Seattle (WA): University of Washington, Seattle 2017:1993-2020.

Tong, S.-K., *et al.* (2001). "Phylogeny, expression and enzyme activity of zebrafish cyp19 (P450 aromatase) genes." The Journal of Steroid Biochemistry and Molecular Biology **79**(1-5): 299-303.

Traore, M. O., *et al.* (2012). "Proof-of-principle of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: final results of a study in Mali and Senegal." PLoS Neglected Tropical Diseases **6**(9): e1825.

Trpis, M., *et al.* (2001). "Development of *Onchocerca volvulus* (Filarioidea: Onchocercidae) in the West African Black Fly *Simulium yahense* (Diptera: Simuliidae) in Liberia." Journal of Parasitology **87**(6): 1265-1273.

Tsalikis, G. (1993). "The Onchocerciasis Control Programme (OCP) in West Africa: a review of progress." Health Policy and Planning. **8**(4): 349–359.

Tuppen, H. A., *et al.* (2010). "Mitochondrial DNA mutations and human disease." Biochimica et Biophysica Acta -Bioenergetics **1797**(2): 113-128.

Ubachukwu, P. (2004). "Human Onchocerciasis: Epidemiological Status of Uzo Uwani LGA of Enugu State." Nig J Parasitol **25**: 93-99.

Underhill, G. W. (1944). "Blackflies found feeding on Turkeys in Virginia (*Simulium nigroparvum* Twinn and *Simulium slossonae* Dyar and Shannon)." Blacksburg, Va (94): 32-ref.5.

Usip, L., et al. (2006). "Longitudinal evaluation of repellent activity of *Ocimum gratissimum* (Labiatae) volatile oil against *Simulium damnosum*." Memórias do Instituto Oswaldo Cruz **101**(2): 201-205.

Vajime, C. and D. Quillévéré (1978). "The distribution of the *simulium damnosum* complex in West Africa with particular reference to the onchocerciasis control programme area." Tropenmedizin und Parasitologie **29**(4): 473-478I.

Vajime, C. G. and R. Dunbar (1975). "Chromosomal identification of eight species of the subgenus Edwardsellum near and including *Simulium* (Edwardsellum) *damnosum* Theobald (Diptera: Simuliidae)." Tropenmedizin und Parasitologie **26**(1): 111-138.

Van Etten, R. A., et al. (1980). "Precise localization and nucleotide sequence of the two mouse mitochondrial rRNA genes and three immediately adjacent novel tRNA genes." Cell **22**(1): 157-170.

Van Lun Low, P. H. A., et al. (2014). "Mitochondrial DNA markers reveal high genetic diversity but low genetic differentiation in the black fly *Simulium tani* Takaoka & Davies along an elevational gradient in Malaysia." PLoS One **9**(6): e100512.

Varlamov, D. A., et al. (2002). "Metabolic consequences of a novel missense mutation of the mtDNA CO I gene." 11(16): 1797-1805.

Vercruyse, J. and R. S. Rew (2002). "Macrocytic lactones in antiparasitic therapy." CABI Publishing Ed. 1: 1-464.

Voet D. J., et al. (2013). Chapter 18: Mitochondrial ATP synthesis, Fundamentals of Biochemistry (4th ed.). Hoboken, NJ: Wiley. pp. 581–620. ISBN 978-0-47054784-7.

Wakeley, J. and J. Hey (1997). "Estimating ancestral population parameters." Genetics **145**(3): 847-855.

Walsh, H., et al. (1999). "Polytomies and the power of phylogenetic inference." Evolution **53**(3): 932-937.

Walsh, J., et al. (1981a). "World Health Organization Onchocerciasis Control Programme in the Volta River Basin." Blackflies: The Future for Biological Methods in Integrated Control 1981: 85-103

Walsh, J., et al. (1981b). "Further studies on the reinvasion of the onchocerciasis control programme by *Simulium damnosum* sl: the effects of an extension of control activities into southern Ivory Coast during 1979." Tropenmedizin und Parasitologie **32**(4): 269-273.

Walsh, J., et al. (1978). "Standardization of criteria for assessing the effect of *Simulium* control in onchocerciasis control programmes." Transactions of the Royal Society of Tropical Medicine and Hygiene **72**(6): 675-676.

Walsh, J., et al. (1993). "Deforestation: effects on vector-borne disease." Parasitology **106**(S1): S55-S75.

Walton, C., et al. (2000). "Population structure and population history of Anopheles dirus mosquitoes in Southeast Asia." Molecular Biology and Evolution **17**(6): 962-974.

Wang, C. and R. J. Youle (2009). "The role of mitochondria in apoptosis." Annual Review of Genetics **43**: 95-118.

Wang, G. and M. J. Huang (2019). "Characterization of the complete mitochondrial genome of *Simulium* (Byssodon) *maculatum* (Diptera: Simuliidae) and its phylogenetic implications." **121**: 152-160.

Wang, Q., et al. (2016). "Sequence and analysis of complete mitochondrial genome of *Simulium aureohirtum*." Journal of Environmental Entomology **38**(2): 361-370.

Webster, J., et al. (2004). "Do hosts and parasites coevolve? Empirical support from the Schistosoma system." The American Naturalist **164**(S5): S33-S51.

Wenk, P. (1981). Bionomics of adult blackflies. In 'Blackflies: The Future for Biological Methods in Integrated Control'.(Ed. M. Laird.) pp. 259–279, Academic Press: London, UK.

WER (2013). "Progress towards eliminating onchocerciasis in the WHO Region of the Americas: verification by WHO of elimination of transmission in Colombia." Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire **88**(36):381–385.

WER (2015). "Progress toward eliminating onchocerciasis in the WHO Region of the Americas: verification of elimination of transmission in Mexico." Weekly Epidemiological Record **2015**;90(43):577–81.

WER (2016). "Progress towards eliminating onchocerciasis in the WHO Region of the Americas: verification of elimination of transmission in Guatemala." Weekly Epidemiological Record **91**(43):501–5.

White, M., et al. (1978). "Chromosomal models of speciation." Modes of speciation.–WH Freeman Co., San Francisco, CA: 169-276.

Whitworth, J., et al. (1993). "Visual loss in an onchocerciasis endemic community in Sierra Leone." British Journal of Ophthalmology **77**(1): 30-32.

WHO-APOC (2010). Conceptual and operational framework of onchocerciasis elimination with ivermectin treatment. Accessed at: https://www.who.int/apoc/oncho_elimination_report_english.pdf. Access date: 13-09-2019, African Programme for Onchocerciasis Control.

WHO (1976). "Epidemiology of onchocerciasis." Report of a WHO Expert Committee, Geneva, Switzerland. No. 597

WHO (1987). "WHO Expert Committee on Onchocerciasis [meeting held in Geneva from 21 to 29 April 1986]: third report", World Health Organization. No. 752: 1-167

WHO (1995). "Onchocerciasis and its Control: Report of the WHO Expert Committee on Onchocerciasis Control Technical Report Series No 852." Geneva: World Health Organization. (1995): 702-703

WHO (2010). Working to overcome the global impact of neglected tropical diseases: first WHO report on neglected tropical diseases, Geneva: World Health Organization. Vol. 1.

WHO (2014). "Elimination of onchocerciasis in the WHO Region of the Americas: Ecuador's progress towards verification of elimination." Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire. World Health Organization. **89**(37): 401-405.

WHO (2015). The WHO African Programme for Onchocerciasis Control: final evaluation report, October 2015, African Programme for Onchocerciasis Control. No. JAF21. 6.

WHO (2016). Guidelines for stopping mass drug administration and verifying elimination of human onchocerciasis: criteria and procedures, World Health Organization. No. WHO/HTM/NTD/PCT/2016.1.

WHO (2018). "The Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) 2017 Annual Report." Accessed at: <https://apps.who.int/iris/bitstream/handle/10665/272344/ESPEN2017Rep-eng.pdf>. Access date: 4th September, 2019.

WHO (2019a). "Blindness and vision impairment." Onchocerciasis Control Programme (OCP). Accessed from: https://www.who.int/blindness/partnerships/onchocerciasis_OCP/en/. Access date: 5th October, 2019.

WHO (2019b). "Blindness and vision impairment." African Programme for Onchocerciasis Control (APOC). Accessed at: <https://www.who.int/blindness/partnerships/APOC/en/>. Access date: 5th October, 2019.

WHO (2019c). "History and future of APOC: a timeline." [African Programme for Onchocerciasis Control \(APOC\)](https://www.who.int/apoc/about/history/en/). Accessed at: <https://www.who.int/apoc/about/history/en/>. Access date: 7th October, 2019

WHO (2019d). "Onchocerciasis Control Programme in West Africa (OCP)." [African Programme for Onchocerciasis Control \(APOC\)](https://www.who.int/apoc/onchocerciasis/ocp/en/). Accessed from: <https://www.who.int/apoc/onchocerciasis/ocp/en/>. Access date: 8th October, 2019.

WHO (2019e). "Onchocerciasis key facts . Accessed from: <https://www.who.int/news-room/fact-sheets/detail/onchocerciasis>, Retrieved 28 September 2019."

Whorley J. R., *et al.* (2004). "Genetic structure of desert ground squirrels over a 20-degree-latitude transect from Oregon through the Baja California peninsula." [Molecular Ecology](#) **13**(9): 2709–2720.

Wilson, M. D., *et al.* (1993). "Multivariate morphotaxonomy in the identification of adult female *Simulium damnosum* Theobald complex (Diptera: Simuliidae) in the Onchocerciasis Control Programme area of West Africa." [Annals of Tropical Medicine and Parasitology](#) **87**(1): 65–82.

Wilson, K., *et al.* (2000). "The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to brachiopods?" [Molecular Biology and Evolution](#) **17**(6): 863-874.

Wilson, M. D., *et al.* (2002). "Deforestation and the spatio-temporal distribution of savannah and forest members of the *Simulium damnosum* complex in southern Ghana and south-western Togo." [Transactions of the Royal Society of Tropical Medicine and Hygiene](#) **96**(6): 632-639.

Wolstenholme, A. J., *et al.* (2004). "Drug resistance in veterinary helminths." [Trends in parasitology](#) **20**(10): 469-476.

Woolhouse, M. E., *et al.* (2002). "Biological and biomedical implications of the co-evolution of pathogens and their hosts." [Nature Genetics](#) **32**(4): 569-577.

Wright, S. (1951). "The genetical structure of populations." [Ann. Eugen](#) **15**(323-54).

Wright, S. (1978). "The relation of livestock breeding to theories of evolution." [Journal of Animal Science](#) **46**(5): 1192-1200.

Wu, C.-I. J. G. (1991). "Inferences of species phylogeny in relation to segregation of ancient polymorphisms." [127](#)(2): 429-435.

Xia X., *et al.* (1996). "On transition bias in mitochondrial genes of pocket gophers." Journal of Molecular Evolution **43**(1): 32–40.

Xiong, B. and T. D. Kocher. (1993). "Intraspecific variation in sibling species of *Simulium venustum* and *Simulium verecundum* complexes (Diptera: Simuliidae) revealed by the sequence of the mitochondrial 16s rRNA gene." Canadian Journal of Zoology **71**(6): 1202–1206.

Yaméogo, L., *et al.* (2004). "Control of river blindness in West Africa: Case history of biodiversity in a disease control program." EcoHealth **1**(2): 172-183.

Yates, D. M., *et al.* (2003). "The avermectin receptors of *Haemonchus contortus* and *Caenorhabditis elegans*." International journal for parasitology **33**(11): 1183-1193.

Yu, D., *et al.* (2007). "The complete nucleotide sequence of the mitochondrial genome of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae)." **396**(1): 66-74.

Yüksel, I. (2009). "Dams and hydropower for sustainable development." Energy Sources Part B **4**(1): 100-110.

Zarroug, I. M., *et al.* (2016a). "Seasonal variation in biting rates of *Simulium damnosum* sensu lato, vector of *Onchocerca volvulus*, in two Sudanese foci." PLoS One **11**(3): e0150309.

Zarroug, I. M., *et al.* (2016b). "The first confirmed elimination of an onchocerciasis focus in Africa: Abu Hamed, Sudan." The American Journal of Tropical Medicine and Hygiene **95**(5): 1037-1040.

Zeng, K., *et al.* (2006). "Statistical tests for detecting positive selection by utilizing high-frequency variants." Genetics **174**(3): 1431-1439.

Zettler, J. A., *et al.* (1998). "Factors influencing larval color in the *Simulium vittatum* complex (Diptera: Simuliidae)." Invertebrate Biology: 245-252.

Zhang, J. (2003). "Evolution by gene duplication: an update." Trends in Ecology Evolution **18**(6): 292-298.

Zhang J. J., *et al.* (2017). "Genetic diversity of mitochondrial control region (D-loop) polymorphisms in *Coilia ectenes taihuensis* inhabiting Taihu Lake, China." Genetics and Molecular Research **16**(1): gmr16019457.

Zhang, X., *et al.* (2019). "Mitochondrial Genomes Provide Insights into the Phylogeny of Culicomorpha (Insecta: Diptera)." International Journal of Molecular Sciences **20**(3): 747.

Zhu, X., *et al.* (1998). "Mitochondrial DNA polymorphism in a black fly *Simulium vittatum* (Diptera: Simuliidae)." Canadian Journal of Zoology **76**(3): 440–447.

Zullo, J. M., *et al.* (2012). "DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina." Cell **149**(7): 1474-1487.