## Title: Landscape genetics approach for onchocerciasis control with the parasite (*Onchocerca volvulus*)and the vector (*Simulium damnosum*) mitochondrial data from the transition region of Ghana

H Shrestha, SM Hedtke, K McCulloch, R Chisholm, K Frempong, WN Grant, fly collection collaborators, S Armoo, N Sirwani, Anusha, Osei’s group

## Background

Onchocerciasis is a neglected filarial disease transmitted by the bites of blackflies and occurs predominantly in Africa and some parts of the Americas (Hill et al., 2019). Upon infection, the human host exhibits a range of chronic clinical manifestations such as severe skin itching, skin depigmentation, blindness and epilepsy or nodding syndrome in children (Basáñez et al., 2006; Colebunders et al., 2019). Onchocerciasis has a huge socio-economic impact in the poorest of the poor nations of the world (Cupp et al., 2011; Dunn et al., 2015). Therefore, control of onchocerciasis has been a public health priority and has progressed through several stages since the commencement of the onchocerciasis control program (OCP) with black fly larviciding in 1975 (Boatin, 2008). Vector control was later complemented with the annual mass drug administration with ivermectin (MDAi) in 1987 (Richards et al., 2001). With OCP ending in 2002, semi-annual MDAi in most of the hyper- and meso-endemic villages has been the sole strategy to control onchocerciasis (Noma et al., 2002). MDAi has led to a significant reduction in the onchocerciasis transmission in the majority of the onchocerciasis endemic foci (with elimination in south American foci, Mali, Senegal), and therefore, the onchocerciasis elimination is now the primary goal (Lakwo et al., 2020; Tekle et al., 2016). Nevertheless, there are instances of persistence of onchocerciasis transmission despite repeated MDAi in some foci, which thwart the target of onchocerciasis elimination (Abong et al., 2021; Awadzi, Boakye, et al., 2004; Basáñez et al., 2006; P. H. Lamberton et al., 2014).

*Simulium damnosum*, the primary vector for the disease, has a specific ecological niche, where *Simulium* larvae need fast-flowing rivers with high oxygen saturation (Cheke et al., 2015, 2017). The narrow range of ecological suitability of blackflies leads to spatial heterogeneity in the prevalence and transmission of onchocerciasis where areas of varying endemicity are in close proximity to each other (Cromwell et al., 2021; Shrestha et al., 2022; Zouré et al., 2014). In addition, there has been spatial variation in interventions, i.e. not all communities (particularly hypoendemic communities) undergo MDAi. These untreated but hypoendemic communities might act as a source of infection in the areas where onchocerciasis is controlled with MDAi. This cross-transmission is usually facilitated by the migration of either infected human hosts or infected vectors, or both, as suggested by some modelling studies (Hedtke et al., 2020; Vos et al., 2021). The migration of the parasites via humans has been linked to recrudescence of onchocerciasis in previously eliminated foci of Burkina Faso (Koala et al., 2017; Nikièma et al., 2018). Similarly, failure to achieve elimination of onchocerciasis in West Africa with OCP was attributed to rapid insecticide resistance due to high vector migration and thus spread of insecticide resistance alleles (Cupp et al., 2011). However, disease control programs have historically focused on government administrative units as the unit of intervention which has led to a situation where treatment decisions are being made without much consideration of where transmission is actually occurring, i.e. the transmission zones.

Transmission zones can be defined as a geographical unit where the disease transmission occurs via locally breeding vectors and forms the basis of biological intervention units (African Programme for Onchocerciasis Control & World Health Organization, 2010). It is crucial to understand transmission zones to ensure that the intervention focus is at the correct scale. The control of onchocerciasis transmission depends on prioritising the limited resources to the most essential areas. The way forward to achieving elimination goals is to align intervention units as close as possible to the natural transmission zones. However, delineating a transmission zone is challenging, and several tools have been deployed to understand transmission zones.

We can gain some insights into the transmission zones based on prevalence mapping, where point prevalence data are interpolated spatially (O’Hanlon et al., 2016; Zouré et al., 2014). However, this is a static map and ignores the 'innate' connectivity between locations; therefore, prevalence map alone is insufficient for distinguishing if the locations belong to different transmission zones. The persistence of transmission is usually facilitated by the migration of pathogens which is challenging to quantify and thus, are rarely incorporated into prevalence mapping. Population genetics has been used to infer the movement of the pathogen where movement can be indirectly measured by the genetic relatedness of samples across locations (Crawford et al., 2019; Hedtke et al., 2020; Small et al., 2019). Genetic relatedness gives us an idea about how common the samples are based on the genetic traits they share, which might be the result of the movement of the fraction of the study population from one location to another.

Population genetics has been used to study the transmission dynamics of onchocerciasis (Adler et al., 2010; Agatsuma, 1987; Charalambous et al., 2005; Choi et al., 2016; Doyle et al., 2017; Hedtke et al., 2020). Population genetics can be used to quantify genetic relatedness between the samples, infer demographic history (e.g. host switching), the evolution of epidemiologically relevant traits like resistance, identify the origin population of the study samples etc. (Archie et al., 2009). However, we are not able to get a complete spatial picture of transmission processes of parasitic diseases with population genetics alone. The dispersal and thus, the geneflow of the parasites and the vectors are a subject of influence by the environmental features of the landscape. Since transmission zone is a spatial concept, population genetics estimates alone are not sufficient to gain insights without incorporating the spatial information and environmental data. Ecological variables can be incorporated into population genetics with the help of landscape genetics approach.

Landscape genetics combine population genetics, landscape ecology and spatial analytical techniques to explicitly quantify the effects of landscape on evolutionary processes like gene-flow, drift, and selection (Balkenhol, 2016). These techniques have traditionally been used in the field of species conservation, but have several potential implications for understanding the epidemiology of diseases and their control and elimination (Hemming-Schroeder et al., 2020; Lo et al., 2017; Saarman et al., 2018; Schwabl et al., 2017). With landscape genetics, we are able to add a spatial dimension to the utility of genetic data in understanding the disease processes, which is well suited to understanding onchocerciasis transmission zones. Spatial information can be added in the form of sample geographic coordinates, remote sensing satellite data of different environmental and climate variables such as elevation, slope, distance to the water bodies, mean annual temperature, mean annual precipitation etc. This allows us to understand how the physical environment influences the population genetic structure of the parasite and the vectors.

Landscape genetics involves a series of steps on how it could be used to infer transmission zones (Schwabl et al., 2017). First, we need to measure the spatial pattern of genetic differentiation of the parasite and the vector population. Second, we can use those parameters of genetic differentiation to see which environmental features might govern the spatial pattern of genetic differentiation. Third, we can transform the most important environmental maps to resistance surface maps based on the genetic connectivity optimisation algorithms. Resistance surface maps quantify the resistance of environmental features to a geneflow of the study population (Hemming-Schroeder et al., 2018; Peterman, 2018). Resistance maps can be used to simulate the pattern of gene flow, which gives us an idea about the migration routes of the parasites and the vectors and thus, the transmission zones (B. H. McRae et al., 2008).

We have implemented this technique to infer about onchocerciasis transmission in the transition ecological region of Ghana where there is persistence of onchocerciasis transmission despite onchocerciasis interventions for almost half a century. Onchocerciasis control started as a vector control in the transition region of Ghana as early as 1974 (Walsh et al., 1979), and ivermectin has been distributed for more than three decades (Alley et al., 1994). Initial population genetics analysis of the parasite samples from these rivers suggested a lack of isolation-by-distance (IBD) i.e. they were genetically homogeneous (Crawford et al., 2019). This suggests cross-transmission of onchocerciasis across river basins, the initially proposed transmission zones and a likely reason for the persistence of onchocerciasis. Not only were the samples collected at a greater distance similar, but some locations were geographically nearer but genetically isolated. This remains unexplained, and environmental factors might likely play a role in resulting to such patterns.

We have incorporated environmental data to the parasite genetic data and additional vector samples sequenced from the transition ecological regions with an objective to: i. determine ecological factors affecting the spatial variation in the parasite and the vector population genetic estimates; ii. infer patterns and routes of gene-flow for the parasite and the vector populations. The underlying hypothesis is that the genetic relatedness between parasites and vectors in different geographical locations allows us to quantify gene flow. For gene flow between these locations, there should be the movement of parasites and the vectors in space which can be inferred with the help of resistance surface maps. This will help us not necessarily to delineate transmission zones but infer about them. Further, we have compared the resistance surface maps with the baseline microfilarial prevalence maps and discussed the immediate implications of the pipeline developed to aid elimination goals.

## Methods

### Sampling locations

The study area is a west-east transect in the ecological “transition zone” of Ghana: an area that includes the Savannah ecotype in the north, the forest ecotype in the south (Klutse et al., 2014), the Volta Lake bissecting the study area in the east, and the Bui National Park in the west (Figure 1). The elevation ranges from 70–525 m above sea level, and mean annual temperature and precipitation ranges from 24–29°C and 1077–1355 mm, respectively (Farr et al., 2007; Fick & Hijmans, 2017).

Sampling locations belonged to four different regions, viz. Bono, Bono East, Savannah and the Northern regions (Figure 1; Table 1). Variant call data from 164 female *Onchocerca volvulus* samples that had been isolated from 97 people from 15 communities primarily in 2010-2012 were received from K. Crawford (Crawford et al. 2019; Supplemental Table X). Sampling locations for vectors were chosen as representative communities for a given river basin based on previous genetic analyses (Gyan, 2020). 93 *S. damnosum* samples collected in 2013-2015 by human landing catch were selected from four communities.

A bounding box formed based on the convex hull boundary with a buffer of 35 km around the sampling locations was used for the landscape analysis. The dimension for the bounding box was 293.68×129.38 km (an area of 37,995.59 km2). Geographic coordinates for all the communities were used to calculate the pairwise geographic distance between the communities (Table 1). We merged the communities near each other (less than 5 km) and used the centroid of the geospatial coordinates of the communities in closed proximity for the merged communities. This brought the number of parasite sampling locations down to 11 but increased the sample size per community (Figure 1).

We chose this area for the study as there is ongoing persistence of onchocerciasis transmission despite decades of control efforts (Osei-Atweneboana et al., 2007; Otabil et al., 2019; Yaméogo, 2008). The onchocerciasis control program has been running for almost half a century now in the transition region of Ghana, with vector control initiated in 1974 as a part of the Onchocerciasis Control Program (OCP) (Biritwum et al., 2021). The use of ivermectin to control onchocerciasis in Ghana commenced in the Pru and Black Volta in 1987 (Biritwum et al., 2021; Borsboom et al., 2003). Asubende (ASU) was among the first communities to receive ivermectin during the early clinical trials of ivermectin (Alley et al., 1994). Consequently, there have also been reports of suboptimal response in these areas (Awadzi, Attah, et al., 2004; Awadzi, Boakye, et al., 2004; Osei-Atweneboana et al., 2011).

Map

Description automatically generated

**Figure 1. The spatial context of the sampling locations of the *Onchocerca volvulus* and *Simulium damnosum* in the transition region of Ghana.** Geographic coordinates are represented as the circle for parasites and square for vectors, and their sizes correspond to the number of parasite samples from the respective locations. The communities are represented with community codes. The river lines and the administrative borders are shown along with the water body, which is Lake Volta. The inset map shows the map of Africa and Ghana with the bounding box for our study area. More information about sampling locations and the number of samples are present in Table 1.

### Sequencing and variant calling

Details on the genetic data generation and the parasite samples are available in Crawford et al (2019). In brief, DNA was extracted from adult female *O. volvulus* from nodules using the DNeasy® Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer’s instructions. Sequence libraries were generated based on either genomic DNA extracts or on amplicons targeting the mitochondrial genome and sequenced using Illumina MiSeq or HiSeq sequencing platforms. Trimmed sequence reads were mapped to the *O. volvulus* (NC\_001861) mitochondrial reference genome and variants called using *GATK UnifiedGenotyper* (McKenna et al., 2010). These data were submitted to the NCBI Short Read Archive under project PRJNA560089 (https://www.ncbi.nlm.nih.gov/sra/).

For *S. damnosum*, the thorax of each fly was dissected and homogenized using a pestle. Extractions of total DNA were performed using the Isolate II Genomic DNA kit, following the manufacturer’s instructions (Bioline, London, United Kingdom). Sequencing libraries were constructed and indexed using the Illumina DNA Prep tagmentation kit following the manufacturer’s instructions (Illumina, San Diego, California, USA). Libraries were pooled and sequenced on one lane of a NovaSeq SP, 300 cycles (resulting in 150-bp paired-end reads) at the Australian Genome Research Facility (Melbourne, Victoria, Australia) (Supplemental Table X). Raw reads were submitted to the NCBI SRA under project #########.

Sequenced reads were trimmed for quality and to remove adapter contamination using *trimmomatic* v.0.32 and keeping only those pairs where both pairs were >125 bp. To assemble the genome, three flies with the largest number of paired reads were mapped using *bwa* v. 0.7.17 [1] to available *Simulium* spp. complete or nearly complete mitochondrial genomes downloaded from NCBI (*Simulium variegatum*, NC\_033348; *Simulium noelleri*, NC\_050320; *Simulium quinquestriatum*, MK281358; *Simulium ornatum*, MT410845; *Simulium maculatum*, NC\_040120; *Simulium aureohirtum*, NC\_029753; *Simulium petricolum*, MT671497; *Simulium equinum*, MT920425; *Simulium angustipes*, MT628576; *Simulium lundstromi*, MT628562). Those reads that mapped to any genome were extracted and converted to fastq using *samtools* v.1.9 [2], and these were used to produce a preliminary assembly using *spades* v. 3.11.1 [3] and using *velvetoptimiser* v. 2.2.5 (https://github.com/tseemann/VelvetOptimiser; [4]). These drafts were then improved using *pilon* v.1.23 [5]. Assemblies from the two different programs were aligned in Mesquite v.3.61 [6], and the consensus—defined as bases that were observed in both assemblies—was taken. Because mitochondrial genomes are circular, and thus the starting point for different linear assemblies differed, the assembly for each fly was oriented so that it began with tRNA-Ile to be consistent with *S. variegatum* (NC\_033348; [7]). The “AT-rich region” was variable in inferred length and sequence between different assemblers, different individual blackflies, and different species, and were difficult to align. Thus, this AT-rich, variable-length region was excluded. The resulting partial genome was submitted to NCBI under accession number XXXXXXX.

[1, 8, 9]Variants were filtered to retain only those calls with a minimum quality score of 30 and a minimum depth of 20 using *vcftools* v.01.13 [10]. Individuals with more than 75% missing data were excluded from the analysis. Variants were normalised using *bcftools* v.1.2. To ensure consistency between variant formatting, allelic primitives were called using the function *vcfallelicprimitives* implemented in *vcflib* [11]. The intersection of the two variant callers was then identified using *bcftools* v.1.2 [12].

For both parasite and vector data, we filtered the variants to remove indels, missing regions, and non-biallelic sites using *vcftools* (Danecek et al., 2011). The resulting dataset included 189 SNPs for 163 *O. volvulus* and 632 SNPs for 93 *S. damnosum*.

### Prevalence data

Pre-intervention prevalence data for communities that fell within the study area bounding box and were based on observation of microfilariae in a skin biopsy via microscopy were obtained from the Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) database (ESPEN, 2020). Prevalence data with duplicate observations were removed, and the average across years was calculated when there were multiple collection time points at the same location. There were 47 unique locations with prevalence data that fell within the study area used for the geospatial mapping of prevalence.

### Environmental data

There is an inherent subjectivity to the ecological requirements of vector and parasite distribution. We tried to reduce this by including all the environmental variables relevant to vector breeding sites, parasite distribution, and disease ecology. We compiled different continuous environmental rasters which were ecologically relevant to the onchocerciasis based on the published literature, both field experiments (Cheke et al., 2017; Opoku, 2006) and geospatial modelling studies (Barro & Oyana, 2012; Cheke et al., 2015; Cromwell et al., 2021; O’Hanlon et al., 2016; Shrestha et al., 2022). These environmental variables included distance to the nearest river, soil moisture, elevation, slope, temperature, and precipitation (Barro & Oyana, 2012; Cromwell et al., 2021; Shrestha et al., 2022). In addition, the spatial distribution of onchocerciasis is highly contingent on the vector habitat and the ecological preference for their biting and breeding activities. The dispersal capacity of the *Simulium* vector is generally as high as around 27 km and is dependent on the vegetation type and time of the year (WHO, 2020). Therefore, we included vegetation and seasonality related variables into our analysis. In addition to environmental variables, we also included some socio-economic aspects of the study area—for example, the human population density to consider the availability of human hosts for disease transmission. We used the environmental variables from the corresponding year to account for the differences in the time period when the samples or data were collected. For prevalence data, environmental variables before 2001 were used, and similarly, for the *O. volvulus* and *S. damnosum*, environmental variables from 2010–2012 and 2013–2015 were used respectively as per the data availability. Our starting set of environmental and socio-economic datasets consisted of 32 continuous environmental rasters at a spatial resolution of 1 km from publicly available repositories via Earth Engine (Table S1) (Gorelick et al., 2017).

#### Selection of the environmental variables

We extracted the environmental and sociodemographic values for each prevalence sample location using the *raster* package in R (version 4.1.0) (Hijmans et al., 2015; R Core Team, 2021). For testing the association of the landscape factors to the genetic differentiation or gene flow between the populations, between site characteristics are crucial (Balkenhol, 2016; Hemming-Schroeder et al., 2020). Thus, we estimated the average of the values encountered by a pairwise straight path between each sampling site for all the environmental and sociodemographic variables.

We used principal component analysis (PCA) to identify variables that contributed most to the genetic variance (for landscape genetics) or to variance in prevalence (for prevalence interpolation) in the sampling sites (Figure S1, S3). For the *Simulium* landscape genetics, three out of four locations were also present in parasite sampling locations. Thus, we did not carry a separate variable selection step for the *Simulium* data. The environmental variables selected for the parasite sampling locations were also used for the vector landscape genetics for easier comparison between the vector and the parasite landscape genetics. For different categories of environmental variables, we generated the pairwise correlation matrix to identify the correlated variables. We included only those variables where Pearson's correlation coefficient was less than (Hemming-Schroeder et al., 2020). For any given group of correlated variables, we selected the variable with the highest? contribution score of each environmental variable for total variance in PCA analysis and considering the ease of interpretability of the variables. After carrying out the initial selection of variables from each category, a similar correlation analysis was done for the selected variables from all the categories combined.

### Population genetic analysis

For the parasite and the vector samples, we carried out unsupervised -means clustering analysis using the *adegenet* package (Jombart, 2008). We inferred the optimal number of (groups) for the population using unsupervised *k-*means clustering using the Bayesian Information Criterion (BIC) using the *R* package *adegenet* v.XX (Jombart, 2008). The vector results were consistent with the results of a haplotype network analysis using *PopART* (Leigh & Bryant, 2015) that identified outlier vector samples separated largely from the cluster of other samples. These outliers were removed from downstream analysis as they may represent different species within the *S. damnosum s.l.* species complex. Then, we carried out Discriminant Analysis of the Principal Components (DAPC) by assigning samples to their respective communities. DAPC is sensitive to the number of principal components retained. Therefore, we performed stratified crossvalidated DAPC by varying the number of principal components using *xvalDapc* function in the *adegenet* package. We calculated the membership probability of each sample, communities, and the posterior correct assignment probability for the communities.

We calculated summary statistics for the genetic data==number of alleles, observed gene diversity, and the pairwise measure of genetic differentiation () between sampling locations using the *Hierfstat* package (Goudet, 2005). Similarly, mean allelic richness and number of haplotypes were calculated using *PopGenReport* and *haplotypes* package, respectively (Adamack & Gruber, 2014; Aktas, 2020). Pairwise matrix was adjusted for finite populations by linearising it with the equation as suggested by (Rousset, 1997; Saarman et al., 2018; Slatkin, 1995).

#### Isolation-by-distance

To test whether the data fit an isolation-by-distance model, the correlation between genetic differentiation and geographic distance was calculated using the pairwise Euclidean distance between the geographic coordinates of the sampling sites and the pairwise linearised genetic differentiation between sites (Savary et al., 2021). Geographic coordinates were converted using the *graph4lg* package to the Universal Transverse Mercator projection, a two-dimensional cartesian coordinate referencing system which is accurate when performing distance-related operations on spatial objects (Diggle, 2019). The coordinate referencing system used in our analysis for all the spatial objects was: epsg-32630 (+proj=utm +zone=30 +datum=WGS84 +units=m +no\_defs). We performed Mantel tests between the geographic distance and the genetic distance matrix with *vegan* package, and the significance of the correlation was calculated based on 10000 permutations (Oksanen et al., 2013).

#### Cost distances

Cost distances reflect both the geographic distance and the hypothesised effect of the intervening landscape features between the sampling sites on the dispersal of the organism of interest (B. H. McRae, 2006; Schwabl et al., 2017). Cost distances are calculated based on “resistance surface” maps. Each pixel in a resistance surface map is assigned a value reflecting the extent to which the landscape feature on that the pixel impedes or facilitates the movement or connectivity of the populations of interest between different locations (Peterman, 2018; Spear et al., 2010). We used Circuitscape as implemented in Julia to calculate circuit distance to generate connectivity Maps and identify corridors for movement in the landscape (Kimberly R. Hall et al., 2021). In addition, connectivity maps were generated using Circuitscape. Circuit distances were further used to optimise resistance surfaces and test environmental variables' effect on genetic differentiation.

#### Resistance surface maps

To parameterise the resistance surfaces, we used ResistanceGA, which is an optimisation method based on a genetic algorithm that offers eight transformations of ricker and monomolecular functions to a continuous surface. The following equations give the ricker and monomolecular transformation function:

Ricker transformation:

Monomolecular transformation:

ResistanceGA searches for the best combination of transformation function, magnitude, and shape parameter. It provides a framework for optimising resistance surfaces from an environmental raster surface without any prior assumptions about the contribution of those surfaces on the resistance (Peterman, 2018) and therefore, provides an unbiased representation of the resistance surface based on the genetic data.

Five different environmental variables selected for the landscape genetic analysis were used to optimise the resistance surface maps. Linearised pairwise genetic distance between sampling locations was used as the response parameter. The cost distance calculated from the transformed resistance surfaces was used as a predictor to find the best model that explains the genetic distance. A linear mixed-effects model with a maximum likelihood population effect (MLPE) was fitted to the data (Clarke et al., 2002; Fukuda et al., 2022). We optimised single surfaces of environmental variables and used the log-likelihood as the objective function for the MLPE model. Four replicates of 1000 iterations each were run with the optimisation set to stop after 50 generations of no improvement. We set the maximum allowable resistance value to be 100 during the optimisation process for easier rescaling and comparison of the resistance values of different environmental variables.

#### Isolation-by-resistance

Each replicate of the resistance surface obtained via the optimisation process was tested using the circuit distance matrix obtained from those resistance surfaces. We used the partial Mantel test to assess the correlation between the genetic distance matrix and the pairwise circuit distance matrix accounting for the geographical distance matrix. Partial mantel has been dominant in landscape genetics analysis but are high in type I error rates with spurious correlations (Cushman & Landguth, 2010). Therefore, we used mixed matrix regression with randomization (MMRR) as a confirmatory test.

The MMRR was performed using the *lgMMRR* function in the *PopGenReport* package based on Wang's, (2013) method. MMRR also gives us the effect of the resistance surface on the genetic differentiation accounting for the geographic distances. To avoid spurious correlations, we took a conservative approach, and the resistance surfaces were deemed significantly associated with the genetic distance only if both the partial mantel and MMRR tests were statistically significant (De Castro et al., 2016; Saarman et al., 2018). Significance for both the partial mantel and MMRR were assessed based on 10,000 permutations.

#### Composite resistance surface maps

As landscape features and environmental gradients do not exist in isolation, the environmental resistance surfaces significantly associated with the genetic distance matrix were manually combined to form a composite resistance surface map. They were rescaled from 0 to 1, where the maximum resistance value among all the significant surfaces was considered as 1, preserving the relative contribution of each optimised surface to the composite resistance map. The composite resistance map was obtained by multiplying the rescaled significant resistance surfaces described in Schwabi et al. (2017). The composite resistance surfaces were used for the connectivity mapping via Circuitscape.

### Mapping prevalence data

The mean of the posterior prevalence was obtained from the pre-intervention microfilariae prevalence data using the Bayesian approach with Integrated Nested Laplace Approximation (INLA) (Moraga et al., 2015; Rue et al., 2009). The number of positive cases out of the total number of people tested in a location was assumed to follow a binomial distribution. The prevalence was modelled with different environmental variables and a spatial random effect with a zero-mean Gaussian process following a Matérn covariance function. The Matérn field is represented with a finite element mesh formed of triangles around the sampling locations and adding vertices over the prediction region. Multiple triangulation meshes with different parameters for cut off and length of triangles inside and outside the boundary were tested for model fit and the computational cost. We created a triangulation mesh with a 3 km cut off; the maximum length of triangles inside and outside the boundary was set to 10 km and 100 km, respectively. Finally, we fitted the model and assessed the relationship of environmental variables with the prevalence data. The details of fitting a spatial model to the prevalence data for geospatial mapping are available in Shrestha et al. (2022).

The prediction of the posterior prevalence was made at 2 km resolution considering the high computational cost of prediction on a lower resolution. A bivariate map of posterior mean prevalence was plotted with the composite resistance surface maps to visualise areas of varying prevalence and resistance. Correlation coefficient measures were calculated between the mean prevalence map and vector and the parasite composite resistance surface maps to test the association between them. We also generated bivariate moving window correlation measures, their significance, and Moran's I measure of spatial autocorrelation to measure the correlation between two spatial processes (Goslee & Urban, 2007).

## Results

We carried out unsupervised -means clustering analysis and visualised the haplotype network for both the parasite and the vector mitochondrial data separately to observe if there were any inherent clusters and if there were any outlier samples. We chose the minimum number of principal components that explained the highest cumulative variance. The number of principal components retained for the clustering analysis of parasite and the vector were 80 and 45, respectively. We chose the number of optimal clusters based on the BIC scores i.e. for the parasite data and for the vector data as the decline in BIC saturated beyond these values (Figure S5). The clustering and haplotype network analysis on the *Simulium* data indicated the presence of outliers (group 6 and 10; Figure S6) which were removed from the downstream analysis.

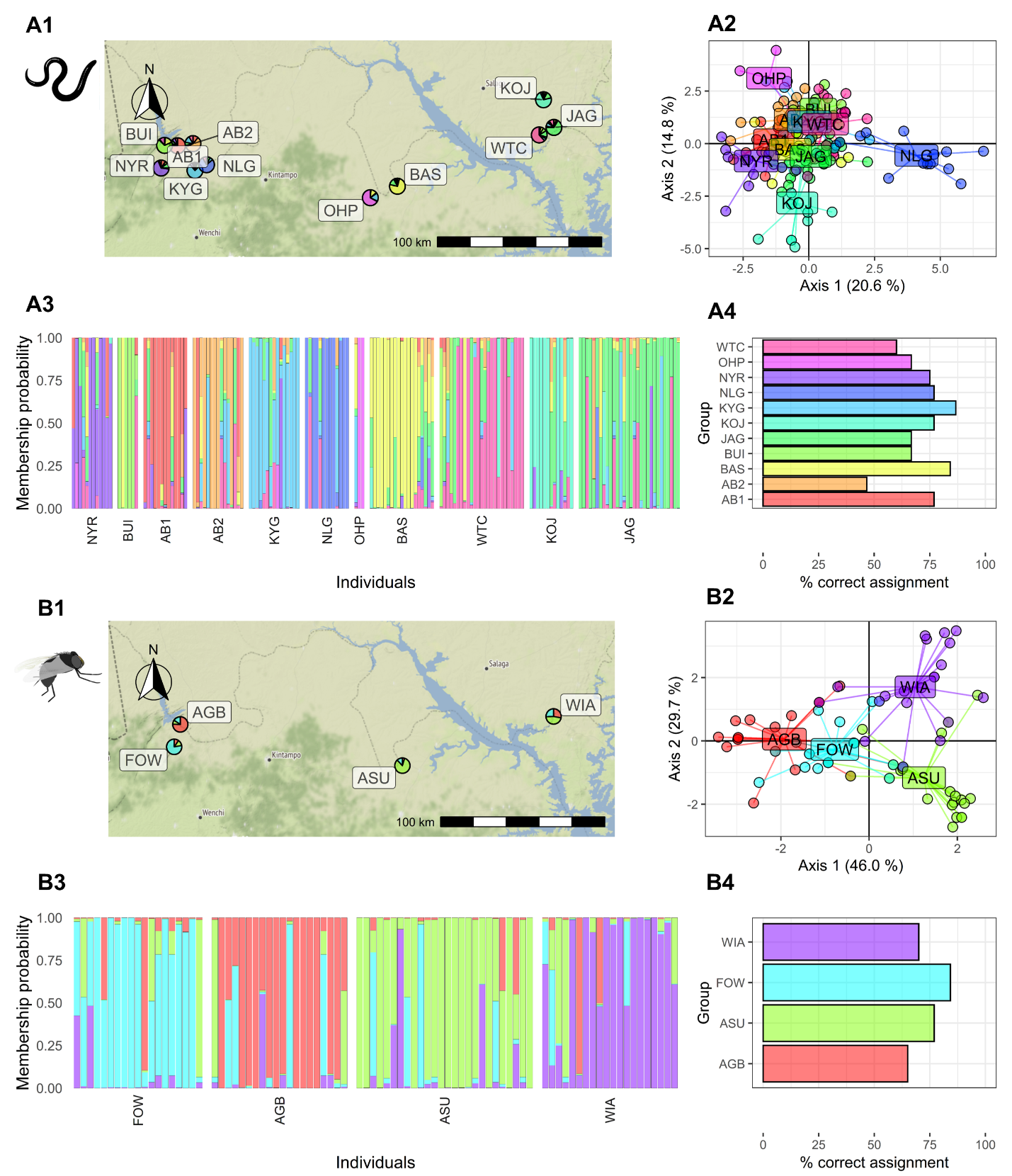
### Population genetic analysis

For the parasite samples, the number of alleles and the number of haplotypes corresponded to the sample size of the population, while the mean allelic richness and the gene diversity correlated with each other (Table 1). Across sampling locations, the parasite's number of alleles averaged 234.18 (±5.001 SE), mean allelic richness averaged 1.071 (±0.003 SE), gene diversity averaged 0.047(±0.002 SE), and the number of haplotypes average 14.27 (±2.13 SE). For the vector samples, the average number of alleles for the parasite populations across sampling locations was 941 (±15.54 SE), mean allelic richness averaged 1.438 (±0.023 SE), gene diversity and the number of haplotypes averaged 0.091 (±0.006 SE) and 18.5 (±1.708 SE), respectively.

**Table 1. Geographic coordinates of the sampling sites along with their, river basin, site code, sample size and population genetics summary statistics.**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Samples** | **River basin** | **Site name** | **Site code** | **Regions** | **Number of samples (n)** | **Longitude** | **Latitude** | **Number of alleles** | **Mean allelic richness** | **Gene diversity** | **Number of haplotypes** |
| Parasites | Black volta | Agborlekame/Agbelekame (1) | AB1 | Savannah | 13 | -2.21 | 8.23 | 227 | 1.0623 | 0.0415 | 13 |
| Agbelekame (2) | AB2 | Savannah | 15 | -2.12 | 8.25 | 227 | 1.0579 | 0.0386 | 14 |
| Bui | BUI | Bono | 6 | -2.28 | 8.24 | 216 | 1.0825 | 0.0550 | 6 |
| Kyingakrom | KYG | Bono | 15 | -2.11 | 8.10 | 240 | 1.0747 | 0.0498 | 15 |
| New Longoro | NLG | Bono East | 13 | -2.05 | 8.13 | 236 | 1.0790 | 0.0526 | 13 |
| Nyire | NYR | Bono | 12 | -2.30 | 8.11 | 236 | 1.0758 | 0.0505 | 12 |
| Daka | Wiae Chabbon\* | WTC | Savannah | 14 | 8.28 | -0.21 | 258 | 1.0699 | 0.0466 | 24 |
| Takumdo\* | 2 | 8.28 | -0.19 |
| Wiae\* | 9 | 8.32 | -0.22 |
| Jagbengbendo | JAG | Northern | 30 | -0.13 | 8.33 | 258 | 1.0637 | 0.0425 | 28 |
| Kojoboni | KOJ | Northern | 13 | -0.18 | 8.49 | 237 | 1.0804 | 0.0536 | 12 |
| Pru | Baaya\* | BAS | Bono East | 1 | 8.00 | -1.02 | 240 | 1.0693 | 0.0462 | 17 |
| Asubende\* | 16 | 8.02 | -0.96 |
| Senyase\* | 2 | 8.02 | -1.00 |
| Ohiampe | OHP | Bono East | 3 | -1.14 | 7.95 | 201 | 1.0635 | 0.0423 | 3 |
| Vectors | Black volta | Agborlekame/Agbelekame (1) | AGB | Savannah | 20 | -2.211 | 8.242 | 972 | 1.4964 | 0.1035 | 17 |
| Fawoman-Banda | FOW | Bono | 19 | -2.245 | 8.12 | 928 | 1.4388 | 0.0869 | 15 |
| Pru | Asubende | ASU | Bono East | 26 | -0.981 | 8.017 | 961 | 1.4248 | 0.0981 | 23 |
| Volta | Wiae | WIA | Northern | 20 | -0.144 | 8.286 | 904 | 1.3905 | 0.0773 | 19 |

We performed a stratified cross-validated DAPC for the parasite and the vector data, optimising the number of principal components used as 72 and 40, respectively. DAPC for the parasite genetic showed overlap between the clusters of the communities, with the exception of few communities like OHP and NLG (Figure 2). The average percentage of correct assignment was only 71.21% (±11.45% SD) for the parasites. For vectors, DAPC showed low but present overlap between clusters of the communities, with an average % correct assignment of only 74.03% (±8.36% SD) which was not signifcantly different compared to the parasite samples ().

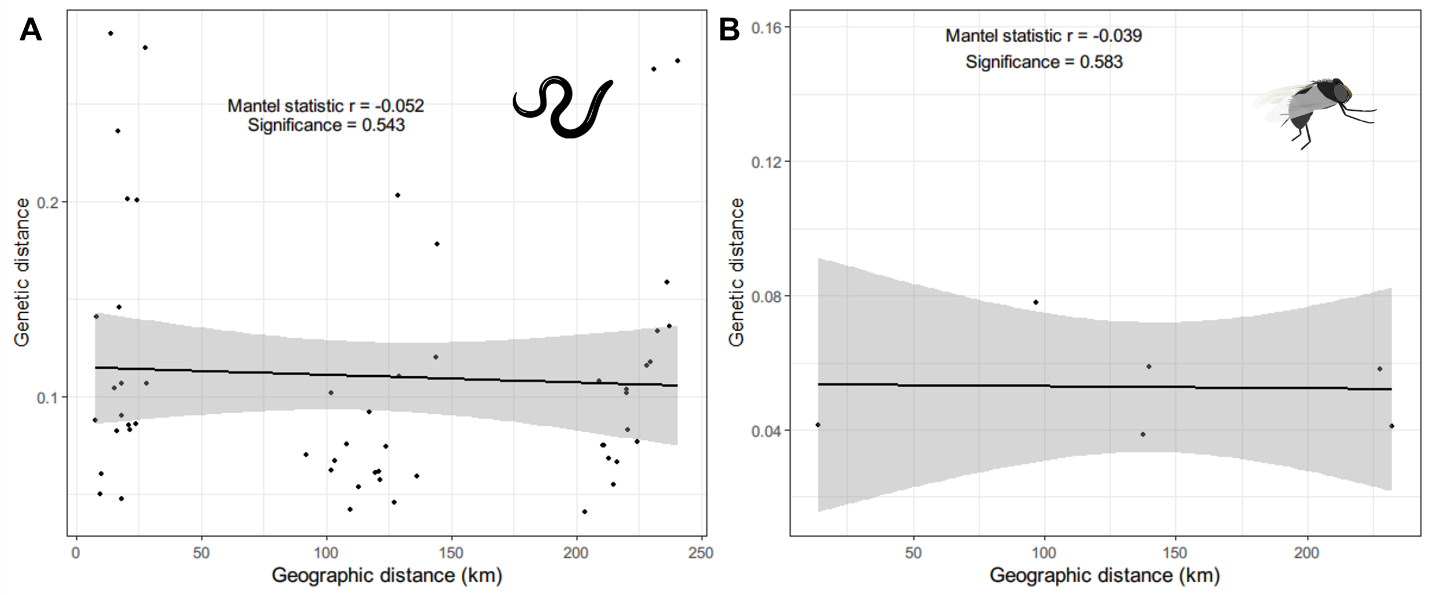


**Figure 2. Discriminant analysis of the principal components (DAPC) analysis for the parasite and the vector sample with respect to sampled 11 and 4 communities respectively in the transition region of Ghana.** The pie chart on the map (**A1, B1**) indicates community level of membership probability. The DAPC analysis showing the community clusters (**A2, B2**) and the individual level membership probability (**A3, B3**) with each block representing communities. The percentage of the samples assigned correctly to their respective communities are shown for both the parasites (**A4**) and the vectors (**B4**). Community codes: BAS: , WTC: ; all other community codes are presented in Table 1.

### Landscape genetic analysis

#### Isolation-by-distance

The Euclidean distance matrix between sample locations and the matrix of linearised pairwise was used to test isolation-by-distance (IBD). The Euclidean geographic distance between locations ranged from 2.2 km to 240.39 km. For the parasite sampling locations, six communities were less than 5km apart and were thus merged into two communities (Table 1). The geographic distance for the parasites averaged 117.73 km (±11.50 SE; range: 7.86–240.43 km), and the genetic distance averaged 0.11 (±0.009 SE; range: 0.041–0.286). Similarly for the vectors, the geographic distance for the parasites averaged 141.40 (±33.61 SE), and the genetic distance averaged 0.056 (±0.007 SE; range: 0.04–0.084). The partial Mantel test indicated a poor correlation between the genetic distance and the geographic distance for both the parasite (Mantel's r = -0.052; p = 0.543) and the vector data (Mantel's r = -0.039; p = 0.583) (Figure 3).



**Figure 3. The relationship between the genetic (linearised ) and the Euclidean geographic distance**. Isolation-by-distance was tested by Mantel test and the significance and the strength of relationship is shown.

#### Resistance surface optimisation and testing

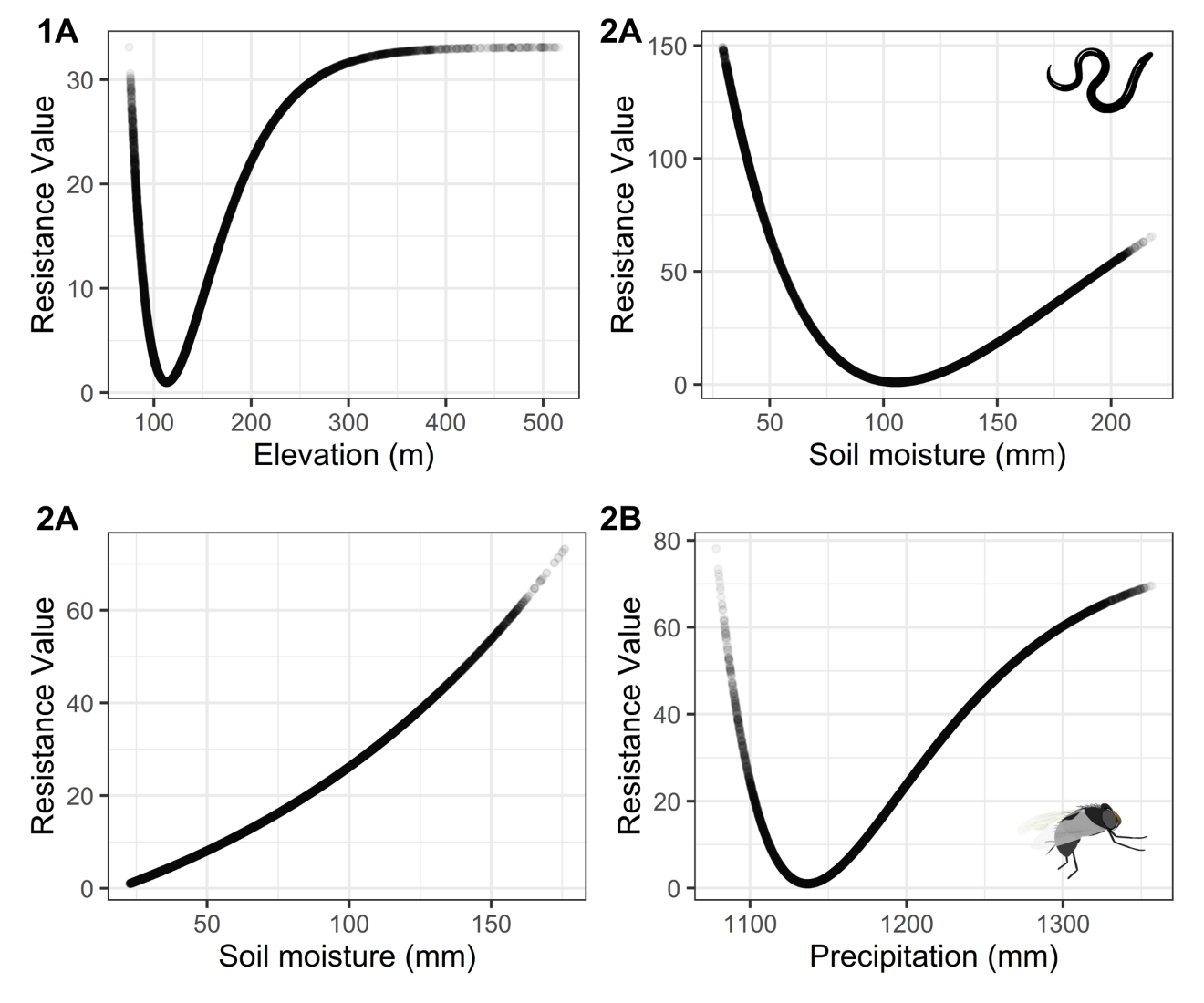
A suite of five environmental variables was selected for the landscape genetics analysis: elevation, isothermality, soil moisture, flow accumulation and annual precipitation. We tested 5 different environmental surfaces to determine whether they could explain the genetic differentiation among parasite and vector sampling locations. We performed four replicates of optimisation for 1000 iterations each and chose the surface with the highest significance (i.e., lowest p-value). For the parasites, we found that the inverse ricker transformation for both elevation (r = 0.793, p = 0.005) and soil moisture (r = 0.507, = 0.002, p = 0.022) to be significant (Table 2). The inverse reverse monomolecular transformation for elevation and inverse monomolecular transformation for soil moisture were also significant, but the level of significance was lower compared to the chosen resistance surfaces. Therefore, inverse ricker transformation surfaces for the elevation and soil moisture were used for the preparation of the composite resistance surface map for the parasite data.

In both the environmental layers, the inverse ricker transformation was significant i.e., high resistance to gene flow in the low and high environmental values, and low resistance in moderate range of environmental values, but with different the scale parameters. The resistance to the gene flow was the lowest i.e., < 30% of the total resistance for an elevation range of 90–150 m and for a soil moisture range of 60–190 mm(Figure 5). A composite resistance surface map was prepared which showed high resistance around the western parts of the study area which are characterised by low soil moisture (i.e., national parks in the west) and high elevation. The areas around Lake Volta also have high resistance. Accordingly the movement corridor map suggests that there is relatively lower connectivity of parasites in the northwestern part of the study area (Figure 6). The central parts of the study area are characterised by high connectivity, showing a potential route for the movement/transmission of parasites.

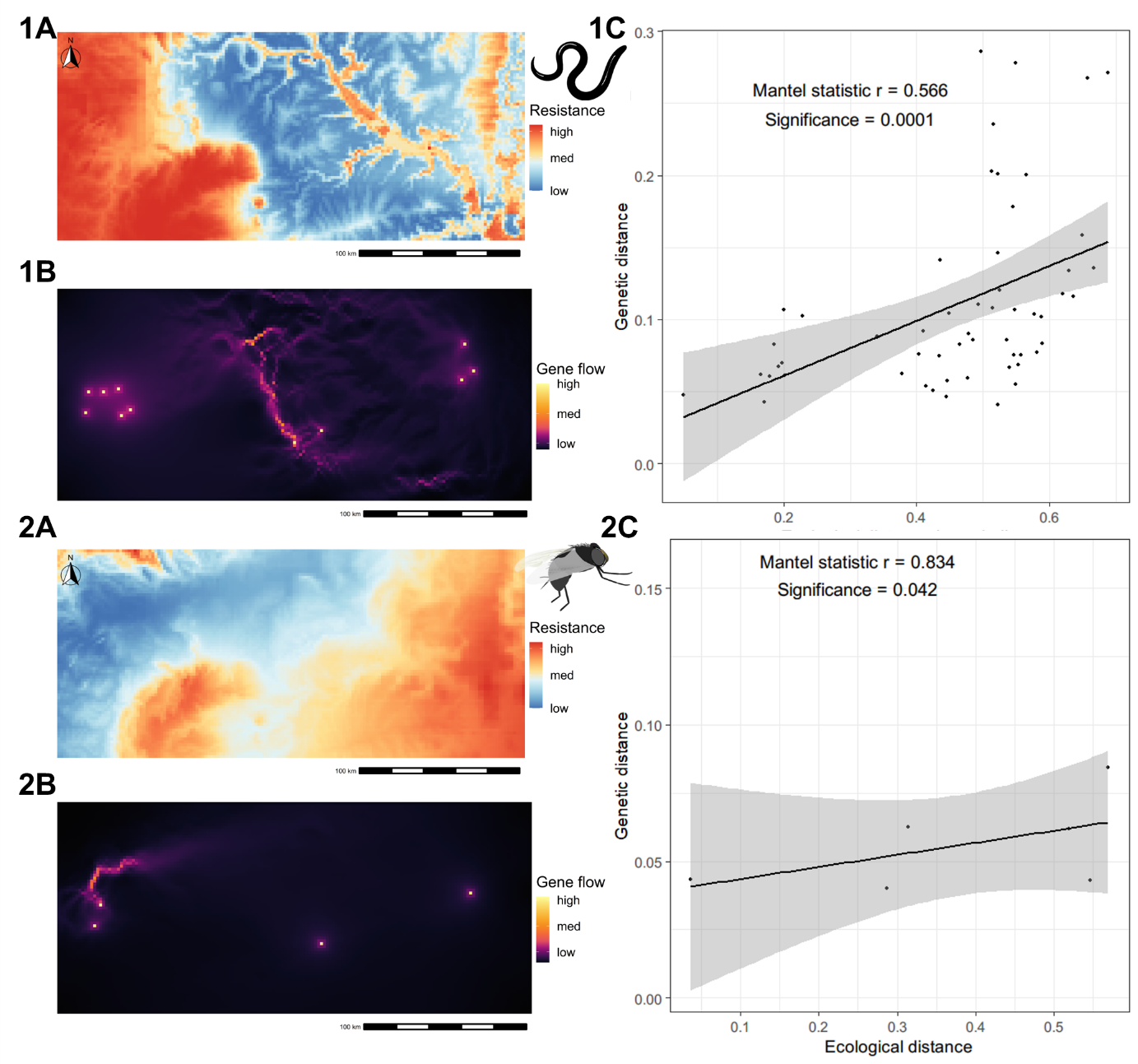
For the vector genetic data, resistance surfaces obtained from soil moisture (r = 0.788, p = 0.0417) and precipitation (r = 0.835, p = 0.0417) were significant, with inverse reverse monomolecular and inverse ricker transformations, respectively. The lowest resistance (<30% of the maximum resistance) for vector gene flow was in the areas with soil moisture 22–90 mm and precipitation of 110cm–120cm. These two resistance surfaces were rescaled and merged to create a composite resistance surface as performed on the parasite data. The composite resistance surface for the vectors revealed that there was particularly low resistance along the western and northwestern areas of the study area and a moderate level of resistance to the vector geneflow in the central region. The current density map showed a high level of connectivity around the communities in the western parts of the study area.

**Table 2. Transformation of environmental surfaces into resistance surfaces with an optimisation function available in *ResistanceGA*.** The strength and the direction of association of the resistance surface to the genetic distance is tested with the partial Mantel test and Multiple Matrix Regression with Randomisation (MMRR). The bold transformations are the selected resistance surfaces with the asterisks (\*) representing the significance of the coefficients. and represents the regression coefficients for the geographic distance and the cost distance due to the resistance surface respectively.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Organism** | **Covariates** | # **replicates** | **Optimisation parameter for resistance surfaces** | | | **Genetic distance ~ resistance distance + geographic distance** | | | | | |
| **partial Mantel** | | **MMMR** | | | |
| **Equation** | **Shape** | **Max** | **r** | **p** |  | **p** |  | **p** |
| **Parasites (*O. volvulus*)** | **Elevation** | **2** | **Inverse Ricker** | **0.873** | **100.000** | **0.793** | **0.0002\*\*\*** | **-0.00038** | **0.008\*** | **0.022** | **0.0046\*\*** |
|  | 2 | Inverse-Reverse Monomolecular | 5.046 | 99.996 | 0.745 | 0.0002\*\*\* | -0.00084 | 0.009\* | 0.046 | 0.0074\* |
| Isothermality | 3 | Inverse-Reverse Ricker | 3.439 | 99.996 | 0.391 | 0.0640 | -0.00035 | 0.131 | 0.004 | 0.2242 |
|  | 1 | Ricker | 0.936 | 99.999 | 0.337 | 0.1324 | -0.00029 | 0.140 | 0.007 | 0.2748 |
| **Soil moisture** | **2** | **Inverse Ricker** | **4.031** | **99.997** | **0.507** | **0.0002\*\*\*** | **-0.00017** | **0.264** | **0.002** | **0.022\*** |
|  | 2 | Inverse Monomolecular | 0.500 | 99.922 | 0.489 | 0.0135\* | -0.00004 | 0.742 | 0.003 | 0.022\* |
| Flow accumulation | 4 | Inverse Monomolecular | 0.500 | 99.998 | 0.120 | 0.4380 | -0.00010 | 0.560 | 0.000 | 0.8181 |
| Precipitation | 4 | Inverse Ricker | 5.000 | 99.976 | 0.439 | 0.1155 | -0.00012 | 0.424 | 0.007 | 0.1364 |
| **Vectors (*S. damnosum*)** | Elevation | 3 | Inverse Monomolecular | 0.500 | 99.835 | 0.804 | 0.0833 | -0.00015 | 0.323 | 0.003 | 0.1229 |
|  | 1 | Inverse Ricker | 2.873 | 99.998 | 0.777 | 0.0833 | -0.00017 | 0.284 | 0.002 | 0.1229 |
| Isothermality | 4 | Inverse Ricker | 3.678 | 100.000 | 0.647 | 0.1250 | -0.00009 | 0.453 | 0.004 | 0.2960 |
| **Soil moisture** | **4** | **Inverse-Reverse Monomolecular** | **7.723** | **100.000** | **0.788** | **0.0417\*** | **-0.00016** | **0.202** | **0.002** | **0.042\*** |
| Flow accumulation | 3 | Inverse Ricker | 3.570 | 99.964 | 0.569 | 0.1250 | -0.00019 | 0.250 | 0.001 | 0.2503 |
|  | 1 | Ricker | 0.500 | 100.000 | 0.678 | 0.0833 | -0.00020 | 0.334 | 0.039 | 0.3721 |
| **Precipitation** | **4** | **Inverse Ricker** | **2.096** | **99.984** | **0.835** | **0.0417\*** | **-0.00018** | **0.161** | **0.002** | **0.0418\*** |
| \*: p < 0.05, \*\*: p < 0.005, \*\*\* p < 0.0005 | | | | | | | | | | | |



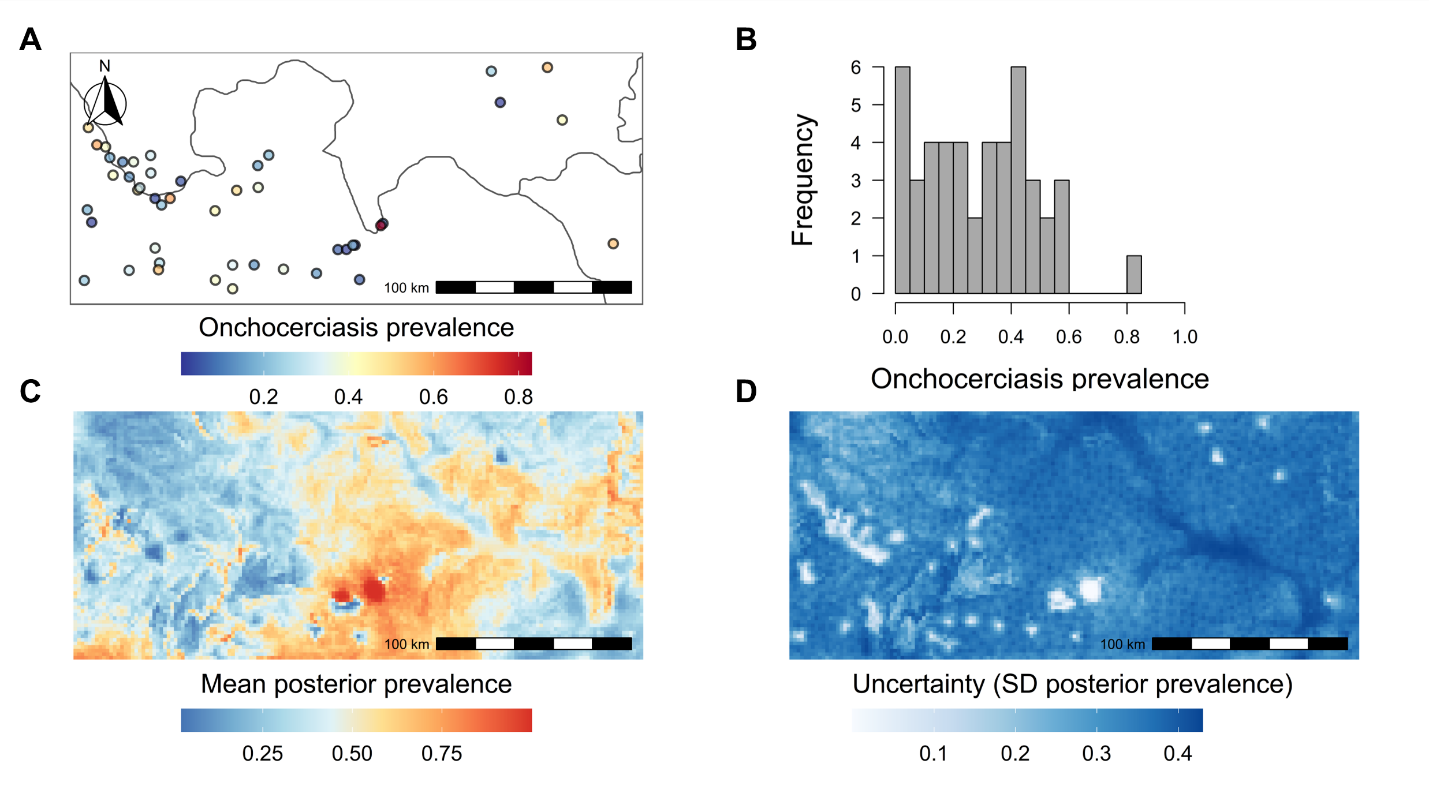
**Figure 5. Transformation functions for the significant environmental covariates.** The figure shows the relationship between the environmental variables with the resistance against gene flow of the *O.* *volvulus* (**1A, 1B**) and *S. damnosum* (**2A, 2B**).



**Figure 6. Composite resistance surface maps prepared from the significant environmental variables along with the gene flow map obtained based on the composite resistance surface map and its relationship with the observed genetic distance.** The resistance surface maps (**1A, 2A**) indicate the ease of movement for the parasite and the vector, and the gene flow map (**1B, 2B**) is obtained based on it with areas highlighted yellow showing the potential routes of movement/gene flow of the organism of interest. The relationship between the circuit distance (cost distance obtained based on the resistance surface) and the genetic distance (**1C, 2C**) is shown.

### Prevalence mapping and bivariate maps

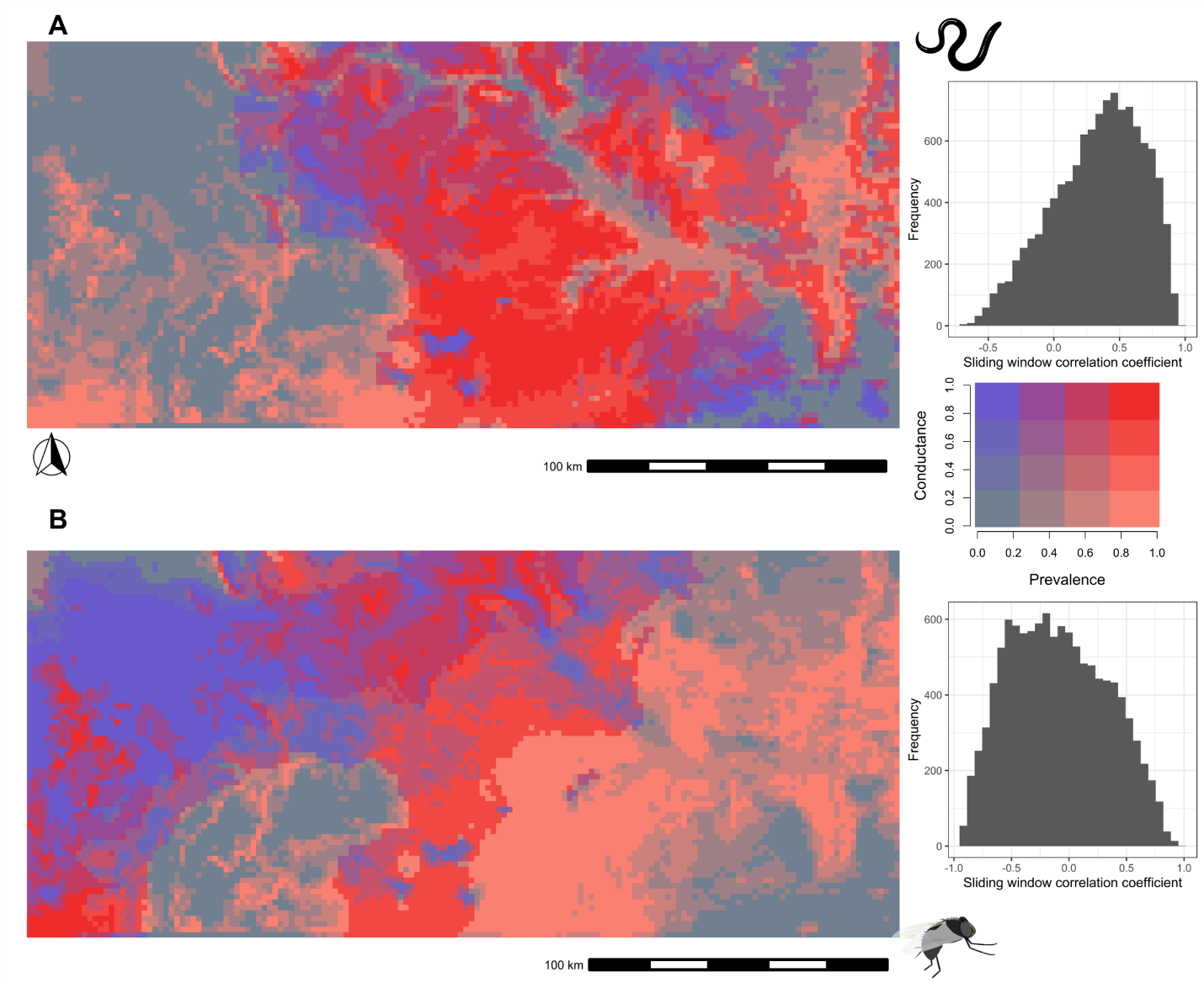
For the analysis of the prevalence data, eight different environemntal and socio-demographic variables were selected: the land surface temperature at night, temperature seasonality, minimum temperature of the coldest month, soil moisture, annual precipitation, slope, distance to the nearest river and prevalence of improved housing were selected. Microfilarial prevalence data ranged from 0.65% to 82.95% with a mean of 29.01% (± 19.31% SD). Most of the data were from the western and south-central parts of the study area, with only five data points from the eastern parts (Figure 7a). The geostatistical interpolated map of baseline microfilarial prevalence based on environmental data shows that the prevalence is higher particularly in the south-central, the central, and eastern regions of the transition Ghana (Figure 7c). The overall predicted prevalence is relatively low in the western regions of transition Ghana with scattered areas of high prevalence. As expected, the uncertainty map shows that the uncertainty was relatively lower in the actual sampling locations with varying level of uncertainties in the interpolated areas (Figure 7d). Based on the regression coefficients, the soil moisture (mean coefficient: 0.043, 95% BCI: 0.004–0.084) and slope (mean coefficient: 2.126, 95% BCI: 0.032–4.338) had a significant positive association with the microfilarial prevalence while the temperature seasonality (mean coefficient: -0.022, 95% BCI: -0.044–-0.001) had a significant negative association with the microfilarial prevalence (Supplementary Table 2). The spatial range of the microfilarial prevalence map was estimated to be 4.4 km (95% BCI: 1.67–7.88 km).



**Figure 7. Mapping baseline prevalence of *O. volvulus* infection in the transition region of Ghana.** Pre-intervention point microfilarial prevalence data () was used to estimate the baseline prevalence of *O. volvulus* infection in the transition region of Ghana. The histogram of the pre-intervention microfilarial prevalence data used in the model and the uncertainty associated with the prevalence map are shown.

The bivariate map for the parasite shows that the area of high parasite conductance and high prevalence are in the central parts of the transition region of Ghana (Figure 8). There is a good correlation between the parasite's composite conductance surface and the *O. volvulus* infection prevalence map with the majority (57.34%) of sliding window correlation coefficients greater than 0.3. Therefore, the areas with high parasite conductance are also the areas of high *O. volvulus* infection prevalence and vice-versa.

Areas of high vector conductance and high prevalence are found in the central and southwestern parts of the study area. However, a substantial portion of the vector bivariate map has high conductance but low prevalence, particularly around the northwestern region of the study area. As a result, the correlation between the conductance map for vectors and the microfilarial infection prevalence is not as strong as the correlation for the parasite counterpart. Only 21.24% of the sliding window correlation coefficients are greater than 0.3.



**Figure 8. Bivariate map created using composite conductance surfaces and the onchocerciasis prevalence map.** Top row shows the bivariate map for the parasite (**A**) and the bottom row (**B**) for the vector. The legend for the bivariate map is shown on the right where red colour indicates area with high prevalence and high conductance whereas blue colour indicates areas with high conductance but low prevalence. The histogram on the right of the respective map shows the frequency of sliding window correlation coefficient between the respective conductance surface and the prevalence map.

## Discussion

We used a landscape genetics framework to identify the ecological factors influencing *S. damnosum* and *O. volvulus* population structure and to infer potential spatial patterns of the vector and the parasite geneflow and thus, the onchocerciasis transmission. A suite of environmental, climate and socio-demographic variables was considered. We compared the output of landscape genetics, the resistance surface maps, with the baseline microfilarial prevalence map, which could be informative for the control and elimination of onchocerciasis in transition ecological region of Ghana. We sequenced the parasite and vector samples from the onchocerciasis endemic communities and vector breeding sites, respectively from the transition ecological region of Ghana. We did the population genetic analysis for the parasite and the vector samples and compared the population genetics estimates in a spatial context. Population genetic estimates have been discussed by Crawford et al., (2019) and Gyan, (2020), which suggest that both the parasite and the vector population were genetically homogeneous. Further, we did not observe any isolation-by-distance (IBD) for both the parasite and the vector populations at the scale of the transition region of Ghana,. This suggests that the geneflow of the parasite and the vector populations were not restricted by geographic distance in the transition ecological region of Ghana.

Historically, the ecological “transition zone” of Ghana has been composed of three river basins—viz. Black Volta/Tombe, Pru, and Daka—that were considered to be independent transmission zones. However, the analysis of the genetic data suggests that there is a single transmission zone in the transition ecological region of Ghana that spans multiple river basins. Both parasites and the vectors may relatively freely moving from one place to the other throughout a single Greater Volta river basin, which includes Lake Volta and its' several tributaries (Sam Armoo, *pers. comm.*). This would not be surprising given the ability of vectors to fly in the range of hundreds of km, particularly when assisted by seasonal winds (Baker et al., 1990; Garms et al., 1979; WHO, 2020). However, it is essential to note that despite being geographically near, some locations had high genetic separation, i.e., low genetic connectivity between locations and vice-versa.

With the assumption that environmental factors influence the vector and the parasite geneflow, we looked at different environmental variables that might influence the genetic connectivity using the landscape genetics framework. Using an optimisation algorithm, we created resistance surfaces from selected environmental features and tested if the resistance distance obtained from the corresponding environmental resistance surfaces is associated with the genetic distance. For the parasite population, resistance surfaces obtained from the elevation and soil moisture were significantly associated with the genetic distance. The resistance for the geneflow was low in the areas of elevation around the range of 90–150 m and, similarly, in the areas with soil moisture, 60–190 mm. This roughly corresponds to the reported range of elevation (95–142 m), which was hypothesised to be suitable for onchocerciasis for geospatial modelling study of onchocerciasis in Ghana (Barro & Oyana, 2012). The high resistance for the parasites in low soil moisture areas could be due to the un-arability of the land and, thus, the lack of human hosts. Soil moisture is reported to be an important environmental feature influencing the occurrence of onchocerciasis in several other studies (Cromwell et al., 2021; Shrestha et al., 2022). Similarly, high soil moisture areas might also not be suitable for onchocerciasis as those were around lake Volta with non-flowing water and generally unsuitable for vector breeding. Lake Volta is one of the biggest artificial lakes in the world. The lakes formed by river dams have been reported to affect the vector breeding sites decreasing onchocerciasis transmission (Katabarwa et al., 2020; Post et al., 2013; Zarroug et al., 2019).

Similarly, the resistance surface derived from soil moisture was also significantly associated with vector population genetics. However, the transformation parameter differed from the resistance surface obtained for the parasite population. The resistance was low for areas with low soil moisture 22–90 mm (vs high resistance for the parasite population in areas with low soil moisture), and the resistance increased almost linearly with the increase in soil moisture. This suggests that vector population geneflow occurred with relatively low resistance in areas with low soil moisture, unlike the parasite geneflow. The possible explanation for this could be the absence of human hosts in areas of low soil moisture around the western region of the transition ecological zone of Ghana. Some of those areas are a part of Bui national park, where there might be black flies but are particularly characterised by sparse human density.

Furthermore, in a study by Doyle et al., (2016) which discriminated the *O. volvulus* and *O. ochengi* larvae from blackflies collected in these regions, the proportion of *O. volvulus* larvae from flies were lower in western communities compared to the communities in the central and the eastern parts of the transition ecological region of Ghana. It was speculated that the spatial difference in the proportion of *O. ochengi* larvae was due to the seasonal increase (in the dry season) in the cattle population in the north-western regions. Therefore, the presence of *O. ochengi* in high proportion in these regions might have impacted the vectorial capacity for the *O. volvulus* as a result of saturation of the vectors with *O. ochengi* (Renz et al., 1982; Wahl et al., 1998), which might be a possible reason for high resistance for the *O. volvulus* populations but a low resistance for the vector populations.

The resistance surface derived from the mean annual precipitation was also significantly associated with the vector population genetic distance. The resistance was low in the areas with mean annual precipitation around 110–120 cm, and the resistance increased as the precipitation increased above 120 cm. While low precipitation decreases the river-flow, an essential feature for the breeding sites of the *S. damnosum*, a hefty rainfall was also reported not to be favourable for *Simulium* breeding in a study done in Ghana (Otabil et al., 2020). In a year-long longitudinal study, Otabil et al. (2020) found that heavy rainfall correlated with the decrease in the relative abundance of *Simulium*. Other studies also report a similar relationship between precipitation and the vector abundance, conducted in Nigeria (Ubachukwu & Anya, 2006) and Sudan (Zarroug et al., 2016). The possible reason behind the unfavorability of the high precipitation to vector abundance is that the heavy downpour might overflow the river banks, sweep away the *Simulium* larvae, and prove to be detrimental to the developmental stages in their lifecycle.

The connectivity analysis using the composite resistance surface maps derived from the significant resistance surfaces allowed us to identify likely geneflow patterns between sites and potential movement routes. The resistance distance obtained based on the connectivity analysis correlated well with the genetic distance. The connectivity map for the parasites showed that the parasite geneflow was high in the central parts of the transition ecological region of Ghana, around communities from the Pru river basin. Similarly, for the vector population, the intensity of the vector geneflow was higher in the western parts around the communities from the Black Volta-Tombe river basin. There were no clear pathways for the vector geneflow, which might be because of the lack of sampling sites. Nevertheless, we can use connectivity analysis as an exploratory tool to identify potential spatial patterns of gene flow of the parasite and vector populations and also hypothesise the source-sink dynamics.

Pre-intervention microfilarial prevalence was positively associated with slope and soil moisture. Areas with high slopes usually comprise fast-flowing rivers which are essential for breeding vector populations. Similarly, soil moisture was also identified to be significant in an analysis of the Ethiopian *O. volvulus* nodule prevalence data, where areas with high soil moisture are arable land are usually inhabited by people and are exposed more to vector bites (Adeleke et al., 2010; Opoku, 2006; Shrestha et al., 2022). In contrast, temperature seasonality was negatively associated with microfilarial prevalence. This is likely because areas with higher fluctuations in temperature might not be favourable for *Simulium*, because [[explain why – rivers drying up? Or only active at certain temperatures? Reduce abundance because only breeding in certain times of year?]] (Cheke et al., 2015; Renz, 1987). Further, the significant relationship of microfilarial prevalence to the temperature seasonality highlights the potential effect of global warming and alterations in annual temperature patterns on the distribution of onchocerciasis. Finally, it is worth noting that the distance to the nearest river was not significantly associated with the prevalence. This might be because almost all the communities surveyed for prevalence were near to the river (less than 10 km).

We produced a bivariate fusion map that combined the results of the microfilariae prevalence and connectivity mapping. Where onchocerciasis prevalence has historically been high, with ongoing transmission, there was also high connectivity for both the vector and the parasite sampling sites (Figure X, center). Despite being among the first communities targeted for both the vector control initially and MDAi later, the onchocerciasis transmission has persisted for quite a long time in the Pru river basin. The baseline prevalence in these areas was greater than 75% and returned to this level even after vector control (Alley et al., 1994; Lamberton et al., 2015).

We calculated the sliding window correlation coefficient between each resistance surface and the microfilarial prevalence map. As expected, there was a close overlap between the parasite resistance surface map and the microfilarial prevalence map. The concurrence between the parasite resistance surface maps and the microfilarial prevalence map also validates landscape genetics model output. However, the correlation between the vector resistance surface and the microfilarial prevalence map was not as strong as for the parasite resistance surface. The correlation breaks down mainly in the western parts of the transition ecological region, which had high conductance for the parasites but low microfilarial prevalence. This could be due to the combination of reasons mentioned above for contrasting high parasite resistance and low vector resistance in low soil moisture areas viz. lack of human population and the greater proportion of *O. ochengi* limiting the vectorial capacity for *O. volvulus*. Nevertheless, anticipating land-use changes that leads to the availability of human host in these regions might pose a risk of onchocerciasis transmission being established in currently low human population density areas in the western parts of the transition ecological region.

### Implications

For the first time in the context of onchocerciasis, we utilise the landscape genetics framework to incorporate the parasite and the vector genetic data with the environmental data. This approach takes us a step ahead in not necessarily "delineating" but inferring about onchocerciasis transmission zones. Here, we have transformed the metrics of genetic connectivity into a resistance surface and the geneflow map giving insights into transmission zones and the source-sink dynamics. Further, the bivariate fusion map can be used to visualise the areas with low resistance and high prevalence, which might act as residual infection pockets even after continuous interventions. Inferences like these might be vital in making spatially explicit onchocerciasis elimination decisions. For example, in the current study, we can hypothesise that communities in the Pru river basin are one of the critical connecting areas with low resistance for the parasite and the vector geneflow and high onchocerciasis prevalence. Since this region has a confluence of perfection for parasites and flies, MDAi alone might not be sufficient to eliminate onchocerciasis transmission in these areas. We might have to complement it with vector control interventions (like slash and clear strategy (Smith et al., 2019)).

Eliminating onchocerciasis transmission in the connecting areas might facilitate onchocerciasis elimination in other surrounding areas. However, it is not to say that the other areas might not act as the source of infection, particularly if the infection is well controlled in the Pru region. When there is a high transmission level in other peripheral communities, there is a high chance of infection being recruited to communities in Pru. Recent modelling work suggests that low endemic areas can act as a source to re-ignite transmission in MDAi-controlled onchocerciasis endemic areas (McCulloch et al., *pers comm,* Vos et al., 2021). Resistance surfaces could be used to prioritise interventions at a larger spatial scale with spatial heterogeneity in interventions. Specifically, areas with low parasite resistance and high prevalence should be prioritised for MDAis, areas with low vector resistance should be prioritised for vector interventions, and the areas with a low resistance to both the parasites and the vector should be prioritised for MDAi complemented with vector interventions. However, for the spatial scale of the current study, where all the communities are well-connected via areas of low resistance, a widespread MDAi needs to be maintained.

The absence of isolation-by-distance among the vector and parasite populations suggests that the connectivity between the river basins was maintained via ecological features elucidating the possibility of transmission across river basins. With the landscape genetics approach, we show that vectors are far more mobile through the landscape than would be suggested by just looking at breeding sites alone. Therefore, it is fair to say that the river basins, particularly in the context of the transition ecological region of Ghana, might not form the biological basis of the intervention unit. It is not unfair to propose a single and larger Great Volta river basin (Sam Armoo *pers. comm.*). Further, transmission zones or intervention units might not be isolated and static but rather dynamic. This further strengthens the fact that we need to have good MDAi coverage over a large geographical scale for it to be effective.

The first clinical trials of MDAi began in Asubende, a community in the Pru river basin, and unsurprisingly, SOR was reported first here (Awadzi, Attah, et al., 2004; Awadzi, Boakye, et al., 2004; Osei-Atweneboana et al., 2011). As shown by this study and other studies, Asubende is the ecologically favourable area for onchocerciasis, characterised particularly by high biting rates, vector density and vector mobility (Frempong et al., 2016; Lamberton et al., 2015). Therefore, with the reports of SOR in these regions and evidence of transmission from these areas, the possibility of spreading the SOR strains cannot be ignored. One can expect the consequences of SOR to be spread over an extensive geographical range than just the focus within which the MDAi is no longer effective. We have a prevalence source that is not controlled by MDAi, which will result in contamination of the gene pool outside of that focus by the SOR genotype. There is a double penalty, a short-term penalty where some areas act as a source of infection irrespective of SOR and a long-term penalty where the SOR genotype might disseminate more widely.

### Limitations and future directions

Despite the potential of landscape genetic approaches in understanding onchocerciasis transmission, some associated limitations exist. The vector or the parasite mobility inferred from the geneflow might not represent the current processes. However, these are the result of the vector and parasite migration that occurred in the recent past. Therefore, even if this is not definite proof of what is happening right now, this could happen in the future. Similarly, high vector mobility might not necessarily mean high vector density or high vector biting rates. High biting rates are crucial for the high endemicity of the disease, whereas the vector mobility might help maintain or even amplify onchocerciasis endemicity. Here we assume that if the vector has high mobility in the areas of high prevalence, there is a likely possibility of high transmission events. However, incorporating vector abundance data and annual biting rates might further enrich the insights from the approach.

There are some caveats specific to the current study that could be improved in future studies. First, the sampling density and the spatial coverage of the samples will increase the accuracy of the estimated resistance surfaces. Future landscape genetic studies should consider dense and stratified uniform sampling across space and environmental gradients (Balkenhol, 2016; Leempoel et al., 2017). Due to the unavailability of the nuclear sequence data, the analysis was done using the mitochondrial sequence data, which lacks recombination and thus might provide a low signal of gene flow (Hedtke et al., 2020). We recommend using nuclear data in future landscape genetics studies. Further, the analysis was done at a single spatial scale. Therefore, different environmental factors might prove to be significant at different spatial scales, either coarser or finer. Thus, the relationship pattern between the environmental variables and their resistance to the gene flow may differ in other regions.

There are seasonal shifts in the species distributions of black flies, which could be challenging to capture with samples from a single time frame Therefore, temporal sampling would be relevant to observing changes in connectivity due to seasonal fluctuations in blackfly densities. Further, blackflies could exist as a metapopulation with local extinction and re-colonisation dynamics (Hedtke et al., 2020), which may further complicate assessing connectivity when using a single year for sampling. Finally, it is essential to note that high resistance does not necessarily mean habitat unsuitability for the blackflies but rather observed unsuitability for the movement of the blackflies based on the genetic data. Nevertheless, this could be a powerful approach to spatially transforming population genetic connectivity estimates, accounting for ecological variables and gaining insights into transmission zones.

## Conclusion

We have demonstrated that the lack of isolation-by-distance, i.e., geographic distance failing to explain the genetic distance, in the transition ecological region of Ghana was well elucidated by considering the environmental variables. Both the parasite and vector populations from communities across the river basins in the transition ecological region of Ghana were connected through specific ecological features. We transformed population genetic estimates of the vector and the parasites into a spatial map which gives us insight into transmission zones and source-sink dynamics of onchocerciasis transmission. Environmental variables such as elevation and soil moisture were significantly associated with the parasite gene flow; similarly, the soil moisture and precipitation were significantly associated with the vector gene flow. In addition, the pre-MDAi microfilarial prevalence analysis found that environmental variables such as slope, soil moisture and temperature seasonality were significantly associated with the microfilarial prevalence. The fusion maps of the resistance surfaces and the prevalence map indicated the central Pru region as the area with low resistance values for both the parasite and the vector populations and high microfilarial prevalence. Therefore, in areas like Pru, which are also characterised by low vector resistance, MDAi alone might not be successful in eliminating transmission and are recommended to be complemented with vector control. Finally, we have used a novel landscape genetics framework for the first time in the context of onchocerciasis to add a spatial dimension to the population genetic estimates and gain insights into onchocerciasis transmission in the transition ecological region of Ghana. This approach could be translatable to any other vector-borne disease and other endemic regions around the world.

## References

Abong, R. A., Amambo, G. N., Hamid, A. A., Enow, B. A., Beng, A. A., Nietcho, F. N., Nji, T. M., Njouendou, A. J., Ritter, M., Esum, M. E., Deribe, K., Cho, J. F., Fombad, F. F., Enyong, P. I., Poole, C., Pfarr, K., Hoerauf, A., Carlow, C., & Wanji, S. (2021). The Mbam drainage system and onchocerciasis transmission post ivermectin mass drug administration (MDA) campaign, Cameroon. *PLOS Neglected Tropical Diseases*, *15*(1), e0008926. https://doi.org/10.1371/journal.pntd.0008926

Adamack, A. T., & Gruber, B. (2014). P op G en R eport: Simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution*, *5*(4), 384–387. https://doi.org/10.1111/2041-210X.12158

Adeleke, M. A., Mafiana, C. F., Sam-Wobo, S. O., Olatunde, G. O., Ekpo, U. F., Akinwale, O. P., & Toe, L. (2010). Biting behaviour of *Simulium damnosum* complex and *Onchocerca volvulus* infection along the Osun River, Southwest Nigeria. *Parasites & Vectors*, *3*(1), 1–7. https://doi.org/10.1186/1756-3305-3-93

Adler, P. H., Cheke, R. A., & Post, R. J. (2010). Evolution, epidemiology, and population genetics of black flies (Diptera: Simuliidae). *Infection, Genetics and Evolution*, *10*(7), 846–865. https://doi.org/10.1016/j.meegid.2010.07.003

Adriaensen, F., Chardon, J. P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003). The application of ‘least-cost’ modelling as a functional landscape model. *Landscape and Urban Planning*, *64*(4), 233–247. https://doi.org/10.1016/S0169-2046(02)00242-6

African Programme for Onchocerciasis Control, & World Health Organization. (2010). *Conceptual and operational framework of onchocerciasis elimination with ivermectin treatment*. African Programme for Onchocerciasis Control.

Agatsuma, T. (1987). Genetic differentiation among natural populations of the vector of onchocerciasis, Simulium ochraceum in Guatemala. *International Journal of Tropical Insect Science*, *8*(4-5–6), 465–469. https://doi.org/10.1017/S1742758400022499

Aktas, C. (2020). *haplotypes: Manipulating DNA Sequences and Estimating Unambiguous Haplotype Network with Statistical Parsimony* (1.1.2) [Computer software]. https://CRAN.R-project.org/package=haplotypes

Alley, E. S., Plaisier, A. P., Boatin, B. A., Dadzie, K. Y., Remme, J., Zerbo, G., & Samba, E. M. (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. https://doi.org/10.1016/0035-9203(94)90172-4

Archie, E. A., Luikart, G., & Ezenwa, V. O. (2009). Infecting epidemiology with genetics: A new frontier in disease ecology. *Trends in Ecology & Evolution*, *24*(1), 21–30. https://doi.org/10.1016/j.tree.2008.08.008

Awadzi, K., Attah, S. K., Addy, E. T., Opoku, N. O., Quartey, B. T., Lazdins-Helds, J. K., Ahmed, K., Boatin, B. A., Boakye, D. A., & Edwards, G. (2004). Thirty-month follow-up of sub-optimal responders to multiple treatments with ivermectin, in two onchocerciasis-endemic foci in Ghana. *Annals of Tropical Medicine & Parasitology*, *98*(4), 359–370. https://doi.org/10.1179/000349804225003442

Awadzi, K., Boakye, D. A., Edwards, G., Opoku, N. O., Attah, S. K., Osei-Atweneboana, M. Y., Lazdins-Helds, J. K., Ardrey, A. E., Addy, E. T., Quartey, B. T., Ahmed, K., Boatin, B. A., & Soumbey-Alley, E. W. (2004). An investigation of persistent microfilaridermias despite multiple treatments with ivermectin, in two onchocerciasis-endemic foci in Ghana. *Annals of Tropical Medicine & Parasitology*, *98*(3), 231–249. https://doi.org/10.1179/000349804225003253

Baker, R. H. A., Guillet, P., Sékétéli, A., Poudiougo, P., Boakye, D., Wilson, M. D., Bissan, Y., Garms, R., Cheke, R. A., Sachs, R., Howe, M. A., Lehane, M. J., Millest, A. L., Kone, T., Davies, J. B., Wilson, M. D., Rainey, R. C., Browning, K. A., Cheke, R. A., & Haggis, M. J. (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. https://doi.org/10.1098/rstb.1990.0141

Balkenhol, N. (Ed.). (2016). *Landscape genetics: Concepts, methods, applications*. Wiley Blackwell.

Barro, A. S., & Oyana, T. J. (2012). Predictive and epidemiologic modeling of the spatial risk of human onchocerciasis using biophysical factors: A case study of Ghana and Burundi. *Spatial and Spatio-Temporal Epidemiology*, *3*(4), 273–285. https://doi.org/10.1016/j.sste.2012.08.001

Basáñez, M.-G., Pion, S. D. S., Churcher, T. S., Breitling, L. P., Little, M. P., & Boussinesq, M. (2006). River blindness: A success story under threat? *PLoS Medicine*, *3*(9), e371. https://doi.org/10.1371/journal.pmed.0030371

Bauder, J. M., Peterman, W. E., Spear, S. F., Jenkins, C. L., Whiteley, A. R., & McGarigal, K. (2021). Multiscale assessment of functional connectivity: Landscape genetics of eastern indigo snakes in an anthropogenically fragmented landscape in central Florida. *Molecular Ecology*, *30*(14), 3422–3438. https://doi.org/10.1111/mec.15979

Biritwum, N.-K., de Souza, D. K., Asiedu, O., Marfo, B., Amazigo, U. V., & Gyapong, J. O. (2021). Onchocerciasis control in Ghana (1974–2016). *Parasites & Vectors*, *14*(1), 3. https://doi.org/10.1186/s13071-020-04507-2

Boatin, B. (2008). The Onchocerciasis Control Programme in West Africa (OCP). *Annals of Tropical Medicine & Parasitology*, *102*(sup1), 13–17. https://doi.org/10.1179/136485908X337427

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borsboom, G. J., Boatin, B. A., Nagelkerke, N. J., Agoua, H., Akpoboua, K. L., Alley, E. W. S., Bissan, Y., Renz, A., Yameogo, L., Remme, J. H., & Habbema, J. D. F. (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. https://doi.org/10.1186/1475-2883-2-8

Charalambous, M., Lowell, S., Arzube, M., & Lowry, C. A. (2005). Isolation-by-distance and a chromosomal cline in the Cayapa cytospecies of Simulium exiguum, the vector of human onchocerciasis in Ecuador. *Genetica*, *124*(1), 41–59. https://doi.org/10.1007/s10709-004-5491-9

Cheke, R. A., Basáñez, M.-G., Perry, M., White, M. T., Garms, R., Obuobie, E., Lamberton, P. H. L., Young, S., Osei-Atweneboana, M. Y., Intsiful, J., Shen, M., Boakye, D. A., & Wilson, M. D. (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. https://doi.org/10.1098/rstb.2013.0559

Cheke, R. A., Young, S., & Garms, R. (2017). Ecological characteristics of *Simulium* breeding sites in West Africa. *Acta Tropica*, *167*, 148–156. https://doi.org/10.1016/j.actatropica.2016.12.022

Choi, Y.-J., Tyagi, R., McNulty, S. N., Rosa, B. A., Ozersky, P., Martin, J., Hallsworth-Pepin, K., Unnasch, T. R., Norice, C. T., Nutman, T. B., Weil, G. J., Fischer, P. U., & Mitreva, M. (2016). Genomic diversity in Onchocerca volvulus and its Wolbachia endosymbiont. *Nature Microbiology*, *2*(2), 1–10. https://doi.org/10.1038/nmicrobiol.2016.207

Clarke, R. T., Rothery, P., & Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, *7*(3), 361. https://doi.org/10.1198/108571102320

Colebunders, R., Fodjo, J. N. S., Hopkins, A., Hotterbeekx, A., Lakwo, T. L., Kalinga, A., Logora, M. Y., & Basáñez, M.-G. (2019). From river blindness to river epilepsy: Implications for onchocerciasis elimination programmes. *PLOS Neglected Tropical Diseases*, *13*(7), e0007407. https://doi.org/10.1371/journal.pntd.0007407

Crawford, K. E., Hedtke, S. M., Doyle, S. R., Kuesel, A. C., Armoo, S., Osei-Atweneboana, M., & Grant, W. N. (2019). *Utility of the* Onchocerca volvulus *mitochondrial genome for delineation of parasite transmission zones* [Preprint]. Evolutionary Biology. https://doi.org/10.1101/732446

Cromwell, E. A., Osborne, J. C. P., Unnasch, T. R., Basáñez, M.-G., Gass, K. M., Barbre, K. A., Hill, E., Johnson, K. B., Donkers, K. M., Shirude, S., Schmidt, C. A., Adekanmbi, V., Adetokunboh, O. O., Afarideh, M., Ahmadpour, E., Ahmed, M. B., Akalu, T. Y., Al-Aly, Z., Alanezi, F. M., … Pigott, D. M. (2021). Predicting the environmental suitability for onchocerciasis in Africa as an aid to elimination planning. *PLOS Neglected Tropical Diseases*, *15*(7), e0008824. https://doi.org/10.1371/journal.pntd.0008824

Cupp, E. W., Sauerbrey, M., & Richards, F. (2011). Elimination of human onchocerciasis: History of progress and current feasibility using ivermectin (Mectizan®) monotherapy. *Acta Tropica*, *120*, S100–S108. https://doi.org/10.1016/j.actatropica.2010.08.009

Cushman, S. A., & Landguth, E. L. (2010). Spurious correlations and inference in landscape genetics: INFERENCE IN LANDSCAPE GENETICS. *Molecular Ecology*, *19*(17), 3592–3602. https://doi.org/10.1111/j.1365-294X.2010.04656.x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Castro, O., Di Maio, A., Di Febbraro, M., Imparato, G., Innangi, M., Véla, E., & Menale, B. (2016). A Multi-Faceted Approach to Analyse the Effects of Environmental Variables on Geographic Range and Genetic Structure of a Perennial Psammophilous Geophyte: The Case of the Sea Daffodil Pancratium maritimum L. in the Mediterranean Basin. *PLOS ONE*, *11*(10), e0164816. https://doi.org/10.1371/journal.pone.0164816

Diggle, P. (2019). *Model-based geostatistics for global public health: Methods and applications*. Taylor & Francis.

Doyle, S. R., Armoo, S., Renz, A., Taylor, M. J., Osei-Atweneboana, M. Y., & Grant, W. N. (2016). Discrimination between Onchocerca volvulus and O. ochengi filarial larvae in Simulium damnosum (s.l.) and their distribution throughout central Ghana using a versatile high-resolution speciation assay. *Parasites & Vectors*, *9*(1), 536. https://doi.org/10.1186/s13071-016-1832-7

Doyle, S. R., Bourguinat, C., Nana-Djeunga, H. C., Kengne-Ouafo, J. A., Pion, S. D. S., Bopda, J., Kamgno, J., Wanji, S., Che, H., Kuesel, A. C., Walker, M., Basáñez, M.-G., Boakye, D. A., Osei-Atweneboana, M. Y., Boussinesq, M., Prichard, R. K., & Grant, W. N. (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. https://doi.org/10.1371/journal.pntd.0005816

Dunn, C., Callahan, K., Katabarwa, M., Richards, F., Hopkins, D., Withers, P. C., Buyon, L. E., & McFarland, D. (2015). The Contributions of Onchocerciasis Control and Elimination Programs toward the Achievement of the Millennium Development Goals. *PLoS Neglected Tropical Diseases*, *9*(5), e0003703. https://doi.org/10.1371/journal.pntd.0003703

ESPEN. (2020). *Site level onchocerciasis prevalence data* [Data]. ESPEN. https://espen.afro.who.int/diseases/onchocerciasis

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, *45*(2). https://doi.org/10.1029/2005RG000183

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Frempong, K. K., Walker, M., Cheke, R. A., Tetevi, E. J., Gyan, E. T., Owusu, E. O., Wilson, M. D., Boakye, D. A., Taylor, M. J., Biritwum, N.-K., Osei-Atweneboana, M., & Basáñez, M.-G. (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. https://doi.org/10.1093/cid/ciw144

Fukuda, Y., Moritz, C., Jang, N., Webb, G., Campbell, H., Christian, K., Lindner, G., & Banks, S. (2022). Environmental resistance and habitat quality influence dispersal of the saltwater crocodile. *Molecular Ecology*, *31*(4), 1076–1092. https://doi.org/10.1111/mec.16310

Garms, R., Walsh, J. F., & Davies, J. B. (1979). Studies on the reinvasion of the Onchocerciasis Control Programme in the Volta River Basin by *Simulium damnosum s.I.* with emphasis on the south-western areas. *Tropenmedizin Und Parasitologie*, *30*(3), 345–362.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv:1207.3907 [q-Bio]*. http://arxiv.org/abs/1207.3907

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Goslee, S. C., & Urban, D. L. (2007). The **ecodist** Package for Dissimilarity-based Analysis of Ecological Data. *Journal of Statistical Software*, *22*(7). https://doi.org/10.18637/jss.v022.i07

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Graves, T. A., Beier, P., & Royle, J. A. (2013). Current approaches using genetic distances produce poor estimates of landscape resistance to interindividual dispersal. *Molecular Ecology*, *22*(15), 3888–3903. https://doi.org/10.1111/mec.12348

Gyan, E. T. (2020). *Analysis of Population Structure of Simulium damnosum sensu lato In the Ecological Transition Zone of Central Ghana* [La Trobe]. https://opal.latrobe.edu.au/articles/thesis/Analysis\_of\_Population\_Structure\_of\_Simulium\_damnosum\_sensu\_lato\_In\_the\_Ecological\_Transition\_Zone\_of\_Central\_Ghana/13180607

Hedtke, S. M., Kuesel, A. C., Crawford, K. E., Graves, P. M., Boussinesq, M., Lau, C. L., Boakye, D. A., & Grant, W. N. (2020). Genomic Epidemiology in Filarial Nematodes: Transforming the Basis for Elimination Program Decisions. *Frontiers in Genetics*, *10*, 1282. https://doi.org/10.3389/fgene.2019.01282

Hemming-Schroeder, E., Lo, E., Salazar, C., Puente, S., & Yan, G. (2018). Landscape Genetics: A Toolbox for Studying Vector-Borne Diseases. *Frontiers in Ecology and Evolution*, *6*, 21. https://doi.org/10.3389/fevo.2018.00021

Hemming-Schroeder, E., Zhong, D., Machani, M., Nguyen, H., Thong, S., Kahindi, S., Mbogo, C., Atieli, H., Githeko, A., Lehmann, T., Kazura, J. W., & Yan, G. (2020). Ecological drivers of genetic connectivity for African malaria vectors Anopheles gambiae and An. Arabiensis. *Scientific Reports*, *10*(1), 19946. https://doi.org/10.1038/s41598-020-76248-2

Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., Bevan, A., Racine, E. B., Shortridge, A., & others. (2015). Package ‘raster’. *R Package*, *734*.

Hill, E., Hall, J., Letourneau, I. D., Donkers, K., Shirude, S., Pigott, D. M., Hay, S. I., & Cromwell, E. A. (2019). A database of geopositioned onchocerciasis prevalence data. *Scientific Data*, *6*(1), 67. https://doi.org/10.1038/s41597-019-0079-5

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Katabarwa, M. N., Zarroug, I. M. A., Negussu, N., Aziz, N. M., Tadesse, Z., Elmubark, W. A., Shumo, Z., Meribo, K., Kamal, H., Mohammed, A., Bitew, Y., Seid, T., Bekele, F., Yilak, A., Endeshaw, T., Hassen, M., Tillahun, A., Samuel, F., Birhanu, H., … Richards, F. (2020). The Galabat-Metema cross-border onchocerciasis focus: The first coordinated interruption of onchocerciasis transmission in Africa. *PLOS Neglected Tropical Diseases*, *14*(2), e0007830. https://doi.org/10.1371/journal.pntd.0007830

Kimberly R. Hall, Ranjan Anantharaman, Vincent A. Landau, Melissa Clark, Melissa Clark, Brett G. Dickson, Aaron Jones, Aaron Jones, Jim Platt, Alan Edelman, & Viral B. Shah. (2021). Circuitscape in julia empowering dynamic approaches to connectivity assessment. *Land*. https://doi.org/10.3390/land10030301

Klutse, N. A. B., Owusu, K., & Ntiamoa-Baidu, Y. (2014). Assessment of Patterns of Climate Variables and Malaria Cases in Two Ecological Zones of Ghana. *Open Journal of Ecology*, *4*, 764–775. https://doi.org/10.4236/oje.2014.412065

Koala, L., Nikiema, A., Post, R. J., Paré, A. B., Kafando, C. M., Drabo, F., & Traoré, S. (2017). Recrudescence of onchocerciasis in the Comoé valley in Southwest Burkina Faso. *Acta Tropica*, *166*, 96–105. https://doi.org/10.1016/j.actatropica.2016.11.003

Lakwo, T., Oguttu, D., Ukety, T., Post, R., & Bakajika, D. (2020). Onchocerciasis Elimination: Progress and Challenges. *Research and Reports in Tropical Medicine*, *11*, 81–95. https://doi.org/10.2147/RRTM.S224364

Lamberton, P. H., Cheke, R. A., Walker, M., Winskill, P., Osei-Atweneboana, M. Y., Tirados, I., Tetteh-Kumah, A., Boakye, D. A., Wilson, M. D., Post, R. J., & Basáñez, M.-G. (2014). Onchocerciasis transmission in Ghana: Biting and parous rates of host-seeking sibling species of the *Simulium damnosum* complex. *Parasites & Vectors*, *7*(1), 511. https://doi.org/10.1186/s13071-014-0511-9

Lamberton, P. H. L., Cheke, R. A., Winskill, P., Tirados, I., Walker, M., Osei-Atweneboana, M. Y., Biritwum, N.-K., Tetteh-Kumah, A., Boakye, D. A., Wilson, M. D., Post, R. J., & Basañez, M.-G. (2015). Onchocerciasis Transmission in Ghana: Persistence under Different Control Strategies and the Role of the Simuliid Vectors. *PLOS Neglected Tropical Diseases*, *9*(4), e0003688. https://doi.org/10.1371/journal.pntd.0003688

Leempoel, K., Duruz, S., Rochat, E., Widmer, I., Orozco-terWengel, P., & Joost, S. (2017). Simple Rules for an Efficient Use of Geographic Information Systems in Molecular Ecology. *Frontiers in Ecology and Evolution*, *5*, 33. https://doi.org/10.3389/fevo.2017.00033

Leigh, J. W., & Bryant, D. (2015). popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lo, E., Hemming-Schroeder, E., Yewhalaw, D., Nguyen, J., Kebede, E., Zemene, E., Getachew, S., Tushune, K., Zhong, D., Zhou, G., Petros, B., & Yan, G. (2017). Transmission dynamics of co-endemic Plasmodium vivax and P. falciparum in Ethiopia and prevalence of antimalarial resistant genotypes. *PLOS Neglected Tropical Diseases*, *11*(7), e0005806. https://doi.org/10.1371/journal.pntd.0005806

Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in Ecology & Evolution*, *28*(10), 614–621. https://doi.org/10.1016/j.tree.2013.05.012

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, *18*(4), 189–197. https://doi.org/10.1016/S0169-5347(03)00008-9

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McRae, B. H. (2006). ISOLATION BY RESISTANCE. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1111/j.0014-3820.2006.tb00500.x

McRae, B. H., Dickson, B. G., Keitt, T. H., & Shah, V. B. (2008). USING CIRCUIT THEORY TO MODEL CONNECTIVITY IN ECOLOGY, EVOLUTION, AND CONSERVATION. *Ecology*, *89*(10), 2712–2724. https://doi.org/10.1890/07-1861.1

McRae, B., Shah, V., & Edelman, A. (2016). Circuitscape: Modeling landscape connectivity to promote conservation and human health. *The Nature Conservancy*, *14*.

Moraga, P., Cano, J., Baggaley, R. F., Gyapong, J. O., Njenga, S. M., Nikolay, B., Davies, E., Rebollo, M. P., Pullan, R. L., Bockarie, M. J., Hollingsworth, T. D., Gambhir, M., & Brooker, S. J. (2015). Modelling the distribution and transmission intensity of lymphatic filariasis in sub-Saharan Africa prior to scaling up interventions: Integrated use of geostatistical and mathematical modelling. *Parasites & Vectors*, *8*(1), 560. https://doi.org/10.1186/s13071-015-1166-x

Nikièma, A. S., Koala, L., Post, R. J., Paré, A. B., Kafando, C. M., Drabo, F., Belem, A. M. G., Dabiré, R. K., & Traoré, S. (2018). Onchocerciasis prevalence, human migration and risks for onchocerciasis elimination in the Upper Mouhoun, Nakambé and Nazinon river basins in Burkina Faso. *Acta Tropica*, *185*, 176–182. https://doi.org/10.1016/j.actatropica.2018.05.013

Noma, M., Nwoke, B. E. B., Nutall, I., Tambala, P. A., Enyong, P., Namsenmo, A., Remme, J., Amazigo, U. V., Kale, O. O., & Sékétéli, A. (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. https://doi.org/10.1179/000349802125000637

O’Hanlon, S. J., Slater, H. C., Cheke, R. A., Boatin, B. A., Coffeng, L. E., Pion, S. D. S., Boussinesq, M., Zouré, H. G. M., Stolk, W. A., & Basáñez, M.-G. (2016). Model-Based Geostatistical Mapping of the Prevalence of *Onchocerca volvulus* in West Africa. *PLOS Neglected Tropical Diseases*, *10*(1), e0004328. https://doi.org/10.1371/journal.pntd.0004328

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O’hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., Wagner, H., & others. (2013). Package ‘vegan’. *Community Ecology Package, Version*, *2*(9), 1–295.

Opoku, A. A. (2006). The ecology and biting activity of blackflies (Simuliidae) and the prevalence of onchocerciasis in an agricultural community in Ghana. *West African Journal of Applied Ecology*, *9*(1), Article 1. https://doi.org/10.4314/wajae.v9i1.45689

Osei-Atweneboana, M. Y., Awadzi, K., Attah, S. K., Boakye, D. A., Gyapong, J. O., & Prichard, R. K. (2011). Phenotypic Evidence of Emerging Ivermectin Resistance in Onchocerca volvulus. *PLoS Neglected Tropical Diseases*, *5*(3), e998. https://doi.org/10.1371/journal.pntd.0000998

Osei-Atweneboana, M. Y., Eng, J. K., Boakye, D. A., Gyapong, J. O., & Prichard, R. K. (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. https://doi.org/10.1016/S0140-6736(07)60942-8

Otabil, K. B., Gyasi, S. F., Awuah, E., Obeng-Ofori, D., Atta-Nyarko, R. J., Andoh, D., Conduah, B., Agbenyikey, L., Aseidu, P., Ankrah, C. B., Nuhu, A. R., & Schallig, H. D. F. H. (2019). Prevalence of onchocerciasis and associated clinical manifestations in selected hypoendemic communities in Ghana following long-term administration of ivermectin. *BMC Infectious Diseases*, *19*(1), 431. https://doi.org/10.1186/s12879-019-4076-2

Otabil, K. B., Gyasi, S. F., Awuah, E., Obeng-Ofori, D., Tenkorang, S. B., Kessie, J. A., & Schallig, H. D. F. H. (2020). Biting rates and relative abundance of Simulium flies under different climatic conditions in an onchocerciasis endemic community in Ghana. *Parasites & Vectors*, *13*(1), 229. https://doi.org/10.1186/s13071-020-04102-5

Peterman, W. E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, *9*(6), 1638–1647. https://doi.org/10.1111/2041-210X.12984

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Kling, D. E., Gauthier, L. D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M. J., Neale, B., MacArthur, D. G., & Banks, E. (2017). *Scaling accurate genetic variant discovery to tens of thousands of samples* [Preprint]. Genomics. https://doi.org/10.1101/201178

Post, R. J., Cheke, R. A., Boakye, D. A., Wilson, M. D., Osei-Atweneboana, M. Y., Tetteh-Kumah, A., Lamberton, P. H., Crainey, J. L., Yaméogo, L., & Basáñez, M.-G. (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. https://doi.org/10.1186/1756-3305-6-205

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Renz, A. (1987). Studies on the dynamics of transmission of onchocerciasis in a Sudan-savanna area of North Cameroon II: Seasonal and diurnal changes in the biting densities and in the age-composition of the vector population. *Annals of Tropical Medicine & Parasitology*, *81*(3), 229–237. https://doi.org/10.1080/00034983.1987.11812116

Renz, A., Organization, W. H., & others. (1982). *Studies on the reivasion by simulium damnosum sl into the Eastern areas of Onchocerciasis Control Programme and on the vectorial capacity of different species of the s. Damnosum complex in Togo and Benin 1982*.

Richards, F. O., Boatin, B., Sauerbrey, M., & Sékétéli, A. (2001). Control of onchocerciasis today: Status and challenges. *Trends in Parasitology*, *17*(12), 558–563. https://doi.org/10.1016/S1471-4922(01)02112-2

Rousset, F. (1997). Genetic Differentiation and Estimation of Gene Flow from *F* -Statistics Under Isolation-by-distance. *Genetics*, *145*(4), 1219–1228. https://doi.org/10.1093/genetics/145.4.1219

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Saarman, N., Burak, M., Opiro, R., Hyseni, C., Echodu, R., Dion, K., Opiyo, E. A., Dunn, A. W., Amatulli, G., Aksoy, S., & Caccone, A. (2018). A spatial genetics approach to inform vector control of tsetse flies ( *Glossina fuscipes fuscipes* ) in Northern Uganda. *Ecology and Evolution*, *8*(11), 5336–5354. https://doi.org/10.1002/ece3.4050

Savary, P., Foltête, J., Moal, H., Vuidel, G., & Garnier, S. (2021). graph4lg: A package for constructing and analysing graphs for landscape genetics in R. *Methods in Ecology and Evolution*, *12*(3), 539–547. https://doi.org/10.1111/2041-210X.13530

Schwabl, P., Llewellyn, M. S., Landguth, E. L., Andersson, B., Kitron, U., Costales, J. A., Ocaña, S., & Grijalva, M. J. (2017). Prediction and Prevention of Parasitic Diseases Using a Landscape Genomics Framework. *Trends in Parasitology*, *33*(4), 264–275. https://doi.org/10.1016/j.pt.2016.10.008

Shrestha, H., McCulloch, K., Hedtke, S. M., & Grant, W. N. (2022). *Geospatial modeling of pre-intervention prevalence of* Onchocerca volvulus *infection in Ethiopia as an aid to onchocerciasis elimination* [Preprint]. Epidemiology. https://doi.org/10.1101/2022.01.10.22269016

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*(1), 457–462. https://doi.org/10.1093/genetics/139.1.457

Small, S. T., Labbé, F., Coulibaly, Y. I., Nutman, T. B., King, C. L., Serre, D., & Zimmerman, P. A. (2019). Human Migration and the Spread of the Nematode Parasite Wuchereria bancrofti. *Molecular Biology and Evolution*, *36*(9), 1931–1941. https://doi.org/10.1093/molbev/msz116

Smith, M. E., Bilal, S., Lakwo, T. L., Habomugisha, P., Tukahebwa, E., Byamukama, E., Katabarwa, M. N., Richards, F. O., Cupp, E. W., Unnasch, T. R., & Michael, E. (2019). Accelerating river blindness elimination by supplementing MDA with a vegetation “slash and clear” vector control strategy: A data-driven modeling analysis. *Scientific Reports*, *9*(1), 15274. https://doi.org/10.1038/s41598-019-51835-0

Spear, S. F., Balkenhol, N., Fortin, M.-J., Mcrae, B. H., & Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: Considerations for parameterization and analysis: RESISTANCE SURFACES IN LANDSCAPE GENETICS. *Molecular Ecology*, *19*(17), 3576–3591. https://doi.org/10.1111/j.1365-294X.2010.04657.x

Tekle, A. H., Zouré, H. G. M., Noma, M., Boussinesq, M., Coffeng, L. E., Stolk, W. A., & Remme, J. H. F. (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. https://doi.org/10.1186/s40249-016-0160-7

Ubachukwu, P., & Anya, A. (2006). Studies on the diurnal biting activity pattern of  *Simulium damnosum* complex (Diptera: Simuliidae) in Uzo-Uwani local government area of Enugu state, NIgeria. *Nigerian Journal of Parasitology*, *22*(1). https://doi.org/10.4314/njpar.v22i1.37775

van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, *76*(1), 1–21.

Vos, A. S. de, Stolk, W. A., Coffeng, L. E., & Vlas, S. J. de. (2021). The impact of mass drug administration expansion to low onchocerciasis prevalence settings in case of connected villages. *PLOS Neglected Tropical Diseases*, *15*(5), e0009011. https://doi.org/10.1371/journal.pntd.0009011

Wahl, G., Enyong, P., Ngosso, A., Schibel, J. M., Moyou, R., Tubbesing, H., Ekale, D., & Renz, A. (1998). *Onchocerca ochengi*: Epidemiological evidence of cross-protection against *Onchocerca volvulus* in man. *Parasitology*, *116*(4), 349–362. https://doi.org/10.1017/S003118209700228X

Walsh, J., Davies, J., Le Berre, R., & others. (1979). Entomological aspects of the first five years of the Onchocerciasis Control Programme in the Volta River Basin. *Tropenmed Parasitol*, *30*(3), 328–344.

Wang, I. J. (2013). EXAMINING THE FULL EFFECTS OF LANDSCAPE HETEROGENEITY ON SPATIAL GENETIC VARIATION: A MULTIPLE MATRIX REGRESSION APPROACH FOR QUANTIFYING GEOGRAPHIC AND ECOLOGICAL ISOLATION: SPECIAL SECTION. *Evolution*, *67*(12), 3403–3411. https://doi.org/10.1111/evo.12134

Wang, I. J., Savage, W. K., & Bradley Shaffer, H. (2009). Landscape genetics and least-cost path analysis reveal unexpected dispersal routes in the California tiger salamander ( *Ambystoma californiense* ). *Molecular Ecology*, *18*(7), 1365–1374. https://doi.org/10.1111/j.1365-294X.2009.04122.x

WHO. (2020). *Report of the Third Meeting of the WHO Onchocerciasis Technical Advisory Subgroup Geneva, Switzerland, 26‒28 February 2019*. https://www.who.int/publications-detail-redirect/9789240006638

Yaméogo, L. (2008). Special Intervention Zones. *Annals of Tropical Medicine & Parasitology*, *102*(sup1), 23–24. https://doi.org/10.1179/136485908X337445

Zarroug, I. M. A., Elaagip, A., Gumaa, S. G., Ali, A. K., Ahmed, A., Siam, H. A. M., Abdelgadir, D. M., Surakat, O. A., Olamiju, O. J., Boakye, D. A., Aziz, N., & Hashim, K. (2019). Notes on distribution of *Simulium damnosum s. L.* along Atbara River in Galabat sub-focus, eastern Sudan. *BMC Infectious Diseases*, *19*(1), 477. https://doi.org/10.1186/s12879-019-4113-1

Zarroug, I. M. A., Hashim, K., Elaagip, A., Samy, A. M., Frah, E. A. M., ElMubarak, W. A., Mohamed, H. A., Deran, T. C. M., Aziz, N., & Higazi, T. B. (2016). Seasonal Variation in Biting Rates of Simulium damnosum sensu lato, Vector of Onchocerca volvulus, in Two Sudanese Foci. *PLOS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0150309

Zouré, H. G., Noma, M., Tekle, A. H., Amazigo, U. V., Diggle, P. J., Giorgi, E., & Remme, J. H. (2014). The geographic distribution of onchocerciasis in the 20 participating countries of the African Programme for Onchocerciasis Control: (2) pre-control endemicity levels and estimated number infected. *Parasites & Vectors*, *7*(1), 326. https://doi.org/10.1186/1756-3305-7-326

1. Li, H., *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.* 2013. **arXiv:1303.3997v2 [q-bio.GN]**.

2. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.

3. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* J Comput Biol, 2012. **19**(5): p. 455-77.

4. Zerbino, D.R. and E. Birney, *Velvet: Algorithms for de novo short read assembly using de Bruijn graphs.* Genome Res, 2008. **18**: p. 821-829.

5. Walker, B.J., et al., *Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement.* PLoS One, 2014. **9**(11): p. e112963.

6. Maddison, W.P. and D.R. Maddison, *Mesquite: a modular system for evolutionary analysis*. 2018.

7. Day, J.C., H.S. Gweon, and R.J. Post, *Sequence and organization of the complete mitochondrial genome of the blackfly Simulium variegatum (Diptera: Simuliidae).* Mitochondrial DNA B Resour, 2016. **1**(1): p. 799-801.

8. Garrison, E. and G. Marth, *Haplotype-based variant detection from short-read sequencing.* arXiv preprint, 2012. **arXiv:1207.3907 [q-bio.GN]**.

9. van der Auwera, G.A., et al., *From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline.* Current Protocols in Bioinformatics, 2013. **43**: p. 11.10.1-11.10.33.

10. Danecek, P., et al., *The variant call format and VCFtools.* Bioinformatics, 2011. **27**(15): p. 2156-8.

11. Garrison, E., et al., *Vcflib and tools for processing the VCF variant call format.* bioRxiv, 2021. **2021.05.21.445151**.

12. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2).