Estimating the number of breeders from helminth larvae with genomic data.

Tristan. P. W. Dennis[1,2]✉, William Sands[2], Millicent Opoku[2,8,9], Alex Debrah3,4, Linda Batsa,

Kenneth Pfarr[6,7], Ute Klarmann-Schulz[6,7], Achim Hoerauf [6,7], Sabine Specht[10], Ivan Scandale[10],

Lisa C. Ranford-Cartwright[2]†, Poppy H. L. Lamberton[2]†✉

†Joint sr author
✉ Corresponding authors
1. Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool, UK
2. School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, Glasgow, UK
3. Kumasi Centre for Collaborative Research in Tropical Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
4. Department of Medical Diagnostics, Faculty of Allied Health Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
5. Department of Clinical Microbiology, School of Medicine and Dentistry, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
6. Institute for Medical Microbiology, Immunology and Parasitology, University Hospital Bonn, Bonn, Germany
7. German Center for Infection Research (DZIF), Partner Site Bonn-Cologne, Bonn, Germany
8. Department of Parasitology, Noguchi Memorial Institute for Medical Research, University of Ghana, Accra, Ghana
9. Department of Environment and Genetics, La Trobe University, Bundoora, Melbourne, Australia
10. Drugs For Neglected Diseases Initiative, Geneva, Switzerland

Abstract

Effective control of helminth infections requires the application of mathematical models to inform control efforts and policy, the development of product profiles for new drugs, and the monitoring of existing drugs for resistance. Key to the success of these approaches is accurately estimating the number of worms within a host, as well as distinguishing, in drug efficacy trials and monitoring, between adults surviving treatment and adults who have reinfected a host following drug treatment. In practice, observing adult worms is often extremely challenging in a live patient, as many adult helminths are embedded deep in host tissues. Transmission stages of the lifecycle can be more accessible. Genetic approaches to infer kinship between larvae or eggs offer a solution to establish adult worm burdens, and to distinguish between treatment failure or treatment success followed by reinfection. Here, we use low-coverage whole-genome (lcWGS), and mitochondrial sequencing, of *Onchocerca volvulus* larvae to estimate the number of adults contributing to pools of offspring of known and unknown parentage. lcWGS reconstructs full-sibling relationships, resolving the number of unique adult worms contributing to a pool of offspring. Mitochondrial genotyping is able to reconstruct maternal sib-ship, thus estimating the minimum number of adult females within a patient. Further development of these approaches will offer a useful technique for evaluating adult worm burden and drug success in monitoring and clinical trials.

Intro

Human helminthiases exert a devastating worldwide public-health burden. Many are neglected tropical diseases (NTDs), causing poor health, morbidity and mortality, disproportionately for the world's poorest people (1). Mass drug administration (MDA) with anthelmintic drugs remains the cornerstone of treatment for several NTDs. Monitoring and evaluation of these programmes is vital to support continued drug donation, to detect potential drug resistance, and to identify areas for improved interventions. The complex life cycle of many indirectly transmitted macroparasites makes elucidation of adult worm burden, the evaluation of control and drug trial success, and establishing baselines of infection levels in pre-clinical trials difficult, as adult worms of many species (e.g. *Schistosoma mansoni* (2), *Wuchereria bancrofti* (3), *Onchocerca volvulus* (4), *Opisthorchis viverrini*

49   (5) or *Trichuris trichiura* (6)) are either very difficult or impossible to retrieve from a living patient. As
50   such, inferring whether a trial macrofilaricide has successfully killed all adult worms, with any
51   subsequent infection being due to reinfection, or whether treatment failure has been followed by
52   repopulation of microfilariae (mf), or renewed reproduction and excretion of larvae, from surviving
53   adults, is impossible using current diagnostics. Moreover, estimating adult worm burden, a key
54   parameter for assessing infection severity and individual contribution to community transmission,
55   based on sampled offspring is problematic, as helminth larval abundance does not necessarily
56   linearly predict adult worm burden, and instead often follows a density-dependent pattern (5,7,8),
57   with adults only sporadically contributing to the larval pool at any given time (9). This gap introduces
58   a considerable level of uncertainty for mathematical models – key tools for informing policy and
59   control success.
60
61   *Onchocerca volvulus* is a filarial nematode parasite of humans, and is the agent of onchocerciasis,
62   also known as river blindness, a disease characterised by blindness and debilitating, disfiguring skin
63   disease. Onchocerciasis afflicts over 20 million people, with ~1 million affected by blindness (10).
64   The severe disease burden exerted by *O. volvulus* in Africa, South America and Yemen have led to
65   concerted efforts to eliminate onchocerciasis. Since 1997, Community Directed Treatment with the
66   microfilaricidal drug Ivermectin (Mectizan®) (CDTI) has been the mainstay of *O. volvulus* control and
67   is the current World Health Organization (WHO) recommended strategy, administered as yearly
68   MDA. Despite decades of CDTI, transmission and disease burden have remained stubbornly high
69   in many regions (11–14). This difficulty necessitates both improved monitoring of existing CDTI, and
70   the development and/or repurposing of anthelmintic drugs such as moxidectin (4), emodepside
71   (15,16), oxfendazole (17) or Corallopyronin A (18). Evaluation of anthelmintics for *O. volvulus* is
72   challenging. The adult worm is embedded in fibrous nodules, and whilst some of these nodules can
73   be detected if near the skin's surface, the adult worms themselves are accessible only through
74   surgical nodulectomy - an invasive procedure. Diagnostics therefore rely on the counting of larval
75   worms - microfilaria (mf) from skin biopsies, also an invasive procedure (albeit less so than a
76   nodulectomy).
77
78   Genetic approaches for evaluating sib-ship and relatedness enable the estimation of the number
79   of adults contributing to a sample of larvae (number of breeders, (*Nb)*), as well as establishing
80   whether the adults contributing to serially sampled larvae have changed, e.g. if treatment has cleared
81   the first set of parents, and a new set of parents have infected the patient. If full-siblings (FS) or
82   maternal half-siblings (HS) between pre- and post-treatment samples are identified, treatment failure
83   can be asserted due to the presence of surviving fecund females. Estimating the relationships and
84   parentage of helminth larvae in the absence of adult information has been performed with multilocus
85   genotyping methods such as microsatellite genotyping, with sib-ship and *Nb* reconstruction being
86   performed most commonly with likelihood-based methods such as COLONY (19) for *Schistosoma*
87   *mansoni* (20,21) and MASTERBAYES (22) for *Trichostrongylus tenuis*, methods which rely on well-
88   validated microsatellite, SNP or multilocus sequence typing (MLST) panels.
89
90   Estimation of pairwise relatedness with genome data makes it possible to infer close familial
91   relationships between individuals in the absence of validated multilocus data, and this is chiefly
92   performed by estimating the proportion of alleles between two individuals that share 0, 1 or 2 copies

93   identical-by-descent (IBD) (23). Whole-genome-sequencing is becoming increasingly affordable,
94   especially with the advent of low-coverage WGS (lcWGS) approaches (24). lcWGS is particularly
95   valuable if achieving high sequencing depth is problematic due to low-input and degraded samples
96   - common for many helminth larvae (25). WGS is now a widely used tool for parasite molecular
97   epidemiology (26,27), is possible from single helminth larvae (25), and can be used to determine
98   relatedness (28,29). Another approach toward assigning parentage is mitochondrial genome
99   sequencing. Mitochondria are inherited maternally, meaning that two individuals sharing identical,
100 or near-identical mitochondrial genomes, can be assigned as probable maternal siblings. This is
101 particularly useful in non-monogamous systems, such as many helminths including *O. volvulus*
102 *(*which is thought to be polyandrous) (30), where a combination of relatedness estimation and
103 mitochondrial genotyping can be used to identify maternal half-sibs. Due to higher copy number of
104 the mitochondrial genome in the cell opposed to the nuclear genome, it is often easier to selectively
105 amplify mitochondrial genomes than whole nuclear genomes for sequencing and analysis (24).
106
107   In this study, we perform a proof-of-concept with lcWGS, relatedness estimation, and
108 mitochondrial genotyping in a complex indirectly transmitted macroparasite, *O. volvulus.* Our aims
109 were to: 1) establish parentage and the *Nb* contributing to a sample of mf from which parentage is
110 known, dissected from the uterus of gravid adult females, and 2) identify parentage of a sample of
111 mf taken from a skin biopsy, of which parentage is unknown. Finally, we discuss the extent to which
112 the approaches taken here may be more widely used toward estimating larval relatedness, adult
113 worm burden, and treatment failure for any parasite where adults are inaccessible, but offspring,
114 (eggs or larvae) can be obtained.
115
116 **Results**
117
118 *1. Parentage of intrauterine mf*
119   We sequenced the genomes of 10 mf dissected from three adult females, from two separate
120 human patients (**Figure 1**). The mean number of reads ('X') covering each position was 10.1 over
121 the nuclear genome, and 47.6 over the mitochondrial genome. The exact number of per-sample
122 reads is given in **Supplementary Table 1**. To calculate relatedness between each mf, we used
123 genotype-likelihoods inferred from aligned reads as input for *NGSRelate,* which implements the
124 framework for estimating relatedness and relationship described by (28). The inferred R0, R1, and
125 KING kinship statistics from the intrauterine mf showed that mf collected from the same female
126 worms/broods all formed a discrete cluster within the kinship coefficient thresholds for full- or half-
127 sibling (28) , (**Supplementary Figures 1A & 1B**). One pair from the same female fell outside of this
128 (**Supplementary Figure 1A, 1B**). Based on these plots, we defined thresholds for full-sibling
129 relationships at R0 ≤ 0.1, R1 ≥ 0.5, and KING ≥ 0.25, and thresholds for half-sibling relationships as
130 R0 ≥ 0.1, R1 ≤ 0.5, and KING ≤ 0.25. These thresholds were then applied to our intrauterine mf data
131 (**Supplementary Figure 2**).
132
133   To further evaluate whether this method reconstructed sib-ship according to our expectations,
134 we drew networks from the relatedness data and plotted them (**Figure 2**). We found that individuals
135 dissected from the same female were at least full or half siblings. However, we found that the

136 assignments also predicted half-sibling relationships to mf dissected from different females. As it is
137 difficult to resolve relationships accurately below FS without nonsensical results (28,31), we
138 discarded our HS estimations and retained only FS relationships (**Figure 2**). The network showed
139 four FS groups, implying three female worms, and four male worms, contributed to the mf sample.
140
141 We extracted mitochondrial genome *data in silico* from the intrauterine mf WGS data and used
142 them to build a minimum-spanning network (**Figure 3**). The network showed a high-level of within-
143 brood sequence similarity, with members of the same brood (having been dissected out of the same
144 female uterus) being a maximum of three SNPs different to one another. This method reconstructed
145 mitochondrial group membership with all offspring corresponding to the mother that they were
146 dissected from (**Figure 3**). When the number of mitochondrial groups was considered alongside the
147 number of full-sibling groups inferred from the intrauterine mf sample, we estimated that seven
148 unique parents (three females and four males) had contributed to this sample of 10 mf.
149
150 *2. Maternal   assignment   of   skin-snip   microfilariae   from   the   same   patient*
151
152 To infer the number of breeders contributing to a sample of mf *for which the parentage was unknown*,
153 we attempted to sequence the genomes of 39 mf from two skin snips taken from the left and right
154 trochanter of a single patient. (**Figure 4**). Due to uneven and low nuclear genome coverage
155 (**Supplementary Table 2**) (mean 4.1), we were unable to obtain reliable relatedness data for these
156 samples. However, mitochondrial genome coverage across all samples was sufficient to genotype
157 whole mitochondria genomes (mean depth-of-coverage 216.1X) (**Supplementary Table 2**). The
158 medium-spanning network disclosed five major mitochondrial groups (**Figure 5**) indicating at least
159 five adult females contributing to the 39 microfilariae for which we successfully obtained sufficient
160 mitochondrial genome data from this patient. There was no structuring in mf haplotype by location
161 (left or right-hand trochanter) **(Figure 5).**
162

Discussion
164 For many helminth parasites, adult worms are not readily available for diagnosis, either at all from
165 living patients (e.g. *Schistosoma spp.)*, or only through invasive surgery (e.g. *O. volvulus)*, or
166 inefficient adult worm excretion methods (e.g. *Ascaris*, *Opisthorchis*). Despite many transmission
167 models and drug efficacy measurements being based on estimated adult numbers (32,33),
168 diagnostics often only focus on the offspring. Here, we use intrauterine- collected mf from known *O.*
169 *volvulus* adult females, as well as mf of unknown parentage collected from two skin-snips from one
170 patient, to test methods to predict sib-ship relationships, maternal parentage and number of breeders
171 (*Nb)*. We report the outcome of estimating helminth parentage using low-coverage WGS and
172 relatedness estimation, as well as mitochondrial genotyping. For the intrauterine mf, we were
173 successfully able to infer full-sibling relationships (including possible contribution of multiple fathers
174 to the mf pool (30)), and, in the absence of successful half-sibling inference, we combined
175 relatedness data with mitochondrial haplotypes to infer the *Nb* contributing to the offspring sample.
176 For the skin-snip mf, we were hampered by degraded, low-input DNA leading to uneven genome
177 coverage and unsuccessful relatedness estimation. However, good coverage over the mitochondrial
178 genome enabled us to infer a minimum of five females contributing to the offspring samples.
179

180   Identification of shared parentage and *Nb* in parasitic helminths is a vital tool for discerning adult
181   worm burdens, as well as for distinguishing between treatment failure followed by repopulation and
182   treatment success followed by reinfection. This has the potential to inform the outcome of
183   macrofilaricide clinical trials (e.g. (16,34)), as well as the establishment of key parasite infection
184   dynamic parameters, such as the minimum adult worm burden and density-dependence, with the
185   assistance of statistical models (e.g. (35,36)). Both relatedness estimation and mitochondrial
186   genotyping have the potential to help address these questions in *O. volvulus* and other indirectly
187   transmitted parasites. For *O. volvulus*, a chromosome level genome assembly has been generated
188   (37), but no microsatellite, SNP or MLST panels exist for parentage analysis. Generating these
189   resources typically require laboratory systems (20), as well as time, money, and access to samples.
190   For pathogens for which laboratory life cycles are problematic to achieve, such as *O. volvulus,* and
191   for which samples are in limited supply – for a variety of different reasons, including lack of funding,
192   or difficult-to-obtain samples - generating panels of markers is difficult. Thus, methods such as ours,
193   that rely on low-coverage WGS to infer relatedness, may be particularly suitable for studying infection
194   dynamics and drug trial outcomes. Furthermore, once siblings have been excluded, unrelated
195   samples can be used for parasite population genomic analysis without introducing biases from
196   siblings. This method can easily be generalised to several parasite systems, if a reference genome
197   assembly has been developed. Reference genomes are becoming increasingly common for even
198   the most neglected of human helminthiases. However, methods for SNP discovery such as RAD-
199   Seq also offer a potential stopgap for relatedness estimation in the absence of a reference genome,
200   using relatedness methods such as SEQOUIA (38).
201
202   The sequencing effort reported here aimed for 5X coverage per-sample, which cost £30 per-
203   sample (around $36 US at the time of writing). Options for improving the cost-effectiveness include
204   performing downsampling experiments to establish the minimum depth-of-coverage required to
205   identify kinships. Simulations suggest that *NGSRelate* may be robust to 1X coverage (39). Moreover,
206   library preparation costs can be greatly reduced by adapting commercial library preparation kits (e.g.
207   dilution) (24,40).
208
209   Another possibility toward reducing cost and increasing ease of access is to restrict parentage
210   analysis in this form to the resequencing of mitochondrial genomes alone, which could be performed
211   by designing long-range PCR primers to conserved mitochondrial regions (26). However, in the
212   absence of population mitochondrial haplotype frequencies, it is impossible to determine whether
213   two individuals with the same mitochondrial haplotype are truly maternal siblings, or whether their
214   mothers are siblings, or if they have unrelated mothers with the same mitochondrial lineage. Here,
215   our data indicate either there is homogeneous distribution of mf across an infected person despite
216   adult worm location, or that there are multiple females with the same mitochondrial haplotype on
217   either side of the body. Given extensive *O. volvulus* mitochondrial diversity and large population
218   sizes (26), mitochondrial genotyping can at least establish the minimum number of females
219   contributing to a pool of larvae, a valuable piece of information and a key parameter in models of
220   helminth transmission (8,32,36), but it cannot determine the maximum number of females
221   contributing to the offspring being assessed.
222

223   Despite technical improvements in single-larva DNA extraction and low-input library preparation
224   (25), single-mf WGS is still technically challenging, with often unreliable results. The skin-snip mf,
225   for example, sequenced very poorly, with highly heterogeneous coverage across the nuclear
226   genome, perhaps indicative of degraded genomic DNA. Further work is needed to improve single-
227   mf WGS sample storage and preparation both for the employment of this approach for use in
228   epidemiological investigations and clinical trials, but also for improving *O. volvulus* population
229   genomic analyses which hitherto has only been possible with adult worms (41,42). Population
230   genomics based on mf would make evolutionary analysis of *O. volvulus* much easier and more
231   ethical, as it removes the need for surgical nodulectomies to access adult worms, although the
232   invasiveness of skin-snips remains a problem.
233
234   Ultimately, development of this method to the point where it can be reliably used in clinical trials
235   will require optimisation of sample storage, DNA extraction and library preparation techniques, as
236   well as more extensive testing against samples of mf for which parentage is known, so that the
237   variation in the range of kinship coefficients corresponding to a given relationship and the error rates
238   can be quantified. This is possible by performing further sequencing studies where adult worms are
239   available.
240
241   **Conclusion**
242
243   The decreasing cost of WGS, combined with technical and theoretical advances that make
244   it possible to infer relationships between helminth larvae from low-coverage WGS data, are
245   facilitating the development of techniques that enable interrogation of fine scale helminth infection
246   dynamics, as well as identifying treatment failure, versus repopulation from surviving adult worms.
247   This is important because many adult helminths, such as *O. volvulus* are, often inaccessible, with
248   their offspring being comparatively easy to sample. Here, we employed lcWGS, relatedness
249   estimation, and mitochondrial genotyping to identify parentage, maternal parentage and *Nb*, showing
250   that, when sample data quality is sufficient, we are able to identify the number of males and females
251   contributing to a sample of offspring and, even when nuclear genome data quality is poor, the
252   minimum number of females may still be estimated. This method, when developed further has
253   potential use in investigating the outcome of macrofilaricide trials for *O. volvulus*, as well as for further
254   probing multiple parasite infection dynamics, in systems for which validated marker panels have not
255   yet been developed.
256
257   **Materials and Methods**
258
259   Data (sequencing metadata, relatedness data, mitochondrial group data) were parsed and
260   analysed in *R* version *4.0.2*, (43), particularly using the *dplyr, gpplot2,* (44), *RColorbrewer* (45)and
261   *cowplot* (46) packages**.**
262
263   *Sample Collection*
264   **Part 1:** Onchocercomata containing adult female worms were surgically removed from patients as
265   part of the field study "Comparison of doxycycline alone vs doxycycline plus rifampicin in their
266   efficacy against onchocerciasis' in Ghana in 2009.

267

268 **Part 2:** Skin-snips were collected in Oct 2019. All infected patients presented at least one palpable
269 nodule. For Mf analysis, two skin biopsies of 1±3 mg were taken from the buttocks using a
270 corneoscleral (Holth) punch. Each biopsy was immersed in 100 ml of 0.9% NaCl solution in a well
271 of a microtiter plate. The skin biopsies were incubated overnight at room temperature to enable mf
272 to emerge. The solution was then transferred onto a slide for microscopic examination. The biopsies
273 were weighed using an electronic balance and mf load was calculated per mg skin. The skin biopsies
274 and the solution containing the free mf were then stored at -20 degrees C for subsequent processing
275 and analyses.
276

277 *Sample Processing, DNA Extraction and Sequencing Library Preparation*
278 **Part 1: Intrauterine mf**: 10 mf were dissected from three gravid *O. volvulus* adult females, obtained
279 by surgical excision from two nodules. (**Figure 1**) Adult females were washed once in 70% ethanol
280 and three times in PBS. Insulin needles (29G) were used to dissect mf from the adult female uterus.
281 Individual mf were pipetted into a 96 well-plate, and DNA extracted using the Beckman Coulter
282 Agencourt DNAdvance gDNA purification kit following the manufacturer's instructions. DNA was
283 eluted into 35µl. The presence of *O. volvulus* was determined using O-150 primers (47). 26 µl of
284 DNA were fragmented and made into libraries using the New England Biolabs NEBNext Ultra FS II
285 kit with unique dual barcodes. Libraries were pooled, and each pool size selected using a Sage
286 Science Blue Pippin with a range of 150-500bp. Further QC was performed on an Agilent
287 Bioanalyzer 2100 High-Sensitivity Chip.
288

289 **Part 2: Skin-snip mf:** 39 mf were dissected into PBS with insulin needles from two skin-snips, one
290 of the left and one of the right trochanter of a single patient. (**Figure 5**). The remaining processing
291 followed the methods used for the intrauterine mf.
292

293 The intrauterine mf and skin-snip mf were then sequenced in two separate runs on an Illumina
294 NextSeq Nano and an Illumina NovaSeq 6000, aiming for 13^e6 paired-end reads (10X mean nuclear
295 genome coverage of 2x75 paired-end reads) and 3.3^e6 (5X coverage of 2x150 reads) respectively.
296

297 *Mapping and Genotype-Likelihood Calling*
298 Demultiplexed reads were trimmed using *Trimmomatic* v0.39 (48)**,** and aligned to the *O. volvulus*
299 reference genome vWBPS15 (37) using bwa-mem (49). PCR duplicates were identified and
300 removed with *GATK* (50). Samtools was used for SAM/BAM file manipulation, and calculation of
301 mapping statistics (51). BAM files were used as input for ANGSD (52), which was used to calculate
302 genotype likelihoods with the following filters: depth > 3, mapQ > 20, max-missingness <20% of
303 individuals, Hardy Weinberg Equilibrium (HWE) p-value 1e-6. See **Supplementary Tables 1 and 2**
304 for per-sample information, and [https://github.com/tristanpwdennis/onchogenome] for code.
305

306 *Relatedness Estimation*
307 The final genotype likelihood file was used as input for the *NGSRelate* software (39), which infers,
308 amongst others, the measures of relatedness: *R0, R1* and *KING-Robust-kinship*, as previously
309 described (28,29,39). The relationship between pairs of full siblings was based on previously defined
310 thresholds *.* All analyses downstream of *NGSRelate* were performed in R *4.0.2* (28,29) (see

311 https://github.com/tristanpwdennis/onchogenome). R0/R1 and R1/KING-Robust-kinships were
312 plotted with *ggplot2*. Graphs of related individuals were plotted in *iGraph* (53).

313

314 *Mitochondrial Genotyping*
315 Reads mapping to the *O. volvulus* mitochondrial genome were extracted from the BAM alignments
316 with samtools. SNPs were called using *Freebayes* (54) with a ploidy setting of 1, and filtered in
317 *bcftools* (51), based on depth > 3 reads, maximum missingness < 0.25, QUAL > 30. The resulting
318 filtered, multisample VCF file was converted into a FASTA alignment with bcftools *bcf-consensus.*
319 The alignment was used as an input for a minimum-spanning network in PEGAS (55), in R. Clusters
320 were defined based on the maximum level of divergence within-brood < 3 SNP differences.

321

Ethics Statements
323 Onchocercomata containing adult female worms were collected as part of a field study "Comparison
324 of doxycycline alone vs doxycycline plus rifampicin in their efficacy against onchocerciasis" in Ghana
325 (ISRCTN68861628). Approval for the study was obtained from the Committee on Human Research,
326 Publications and Ethics of the School of Medicine and Dentistry of the Kwame Nkrumah University
327 of Science and Technology (Ref: CHRPE/KNUST/KATH/22-07-08), Kumasi, Ghana and the Ethics
328 Committee of the University of Bonn, Medical Faculty (Nr. 186/08). Skin-snip collection was
329 approved by the Committee on Human Research, Publications and Ethics of the School of Medicine
330 and Dentistry of the Kwame Nkrumah University of Science and Technology (Ref:
331 CHRPE/KNUST/KATH/22-07-08), Kumasi, Ghana, Research Ethics Committee of the Liverpool
332 School of Tropical Medicine as well as the Ethical Committee of the University Hospital of Bonn,
333 Germany.

334

344

Conflict of Interest Declaration

346

347 The authors declare no competing interests.

348

Data accessibility

350

351 All raw read data used in this study have been deposited in ENA under study accession PRJEB63468.
352 All the code (analysis code as an RMarkdown file, sequence alignment/processing wrapping as shell
353 wrappers), sequencing metadata, relatedness data, ANGSD output and mitochondrial genome
354 alignments are available at https://github.com/tristanpwdennis/onchogenome.
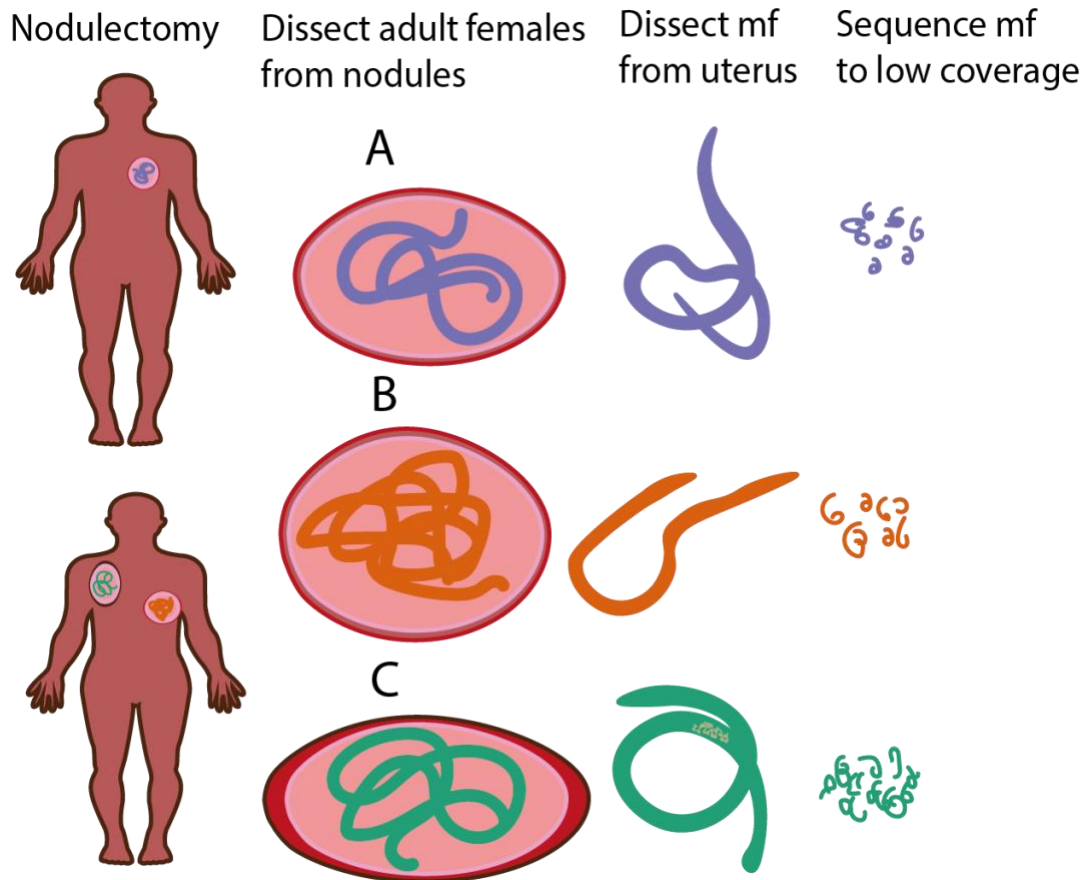
355
<u>Acknowledgements</u>

357

360

361

362 **Figures**
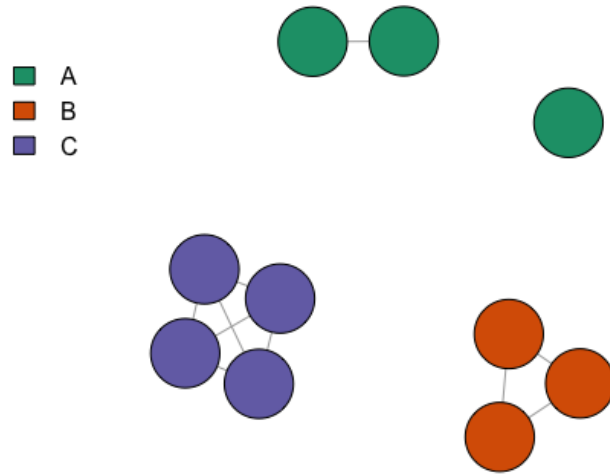363

364
365 **Figure 1:** Schematic of *Onchocerca volvulus* microfilariae (mf) dissections from known parents.
366 (Results Section 1) 10 mf were dissected from the uteri of 3 adult female worms, derived from 3
367 nodulectomies from 2 patients. mf were whole-genome-sequenced.

368

369
370 **Figure 2:** Network of full-sibling relationships between intrauterine *Onchocerca volvulus*
371 microfilariae. Nodes indicate intrauterine *O. volvulus* microfilariae and edges indicate full-sibling
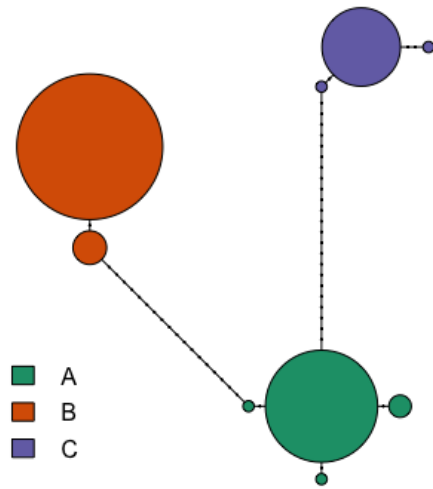372 relationships. Nodes are coloured by known maternal parentage.

373
374
375 **Figure 3:** Medium-spanning network of intrauterine *Onchocerca volvulus* microfilariae
376 mitochondrial haplotypes. Colour indicates shared known maternal parentage. Dots on
377 connecting lines indicate the number of single nucleotide polymorphisms (SNPs) between
378 haplotype groups.
379
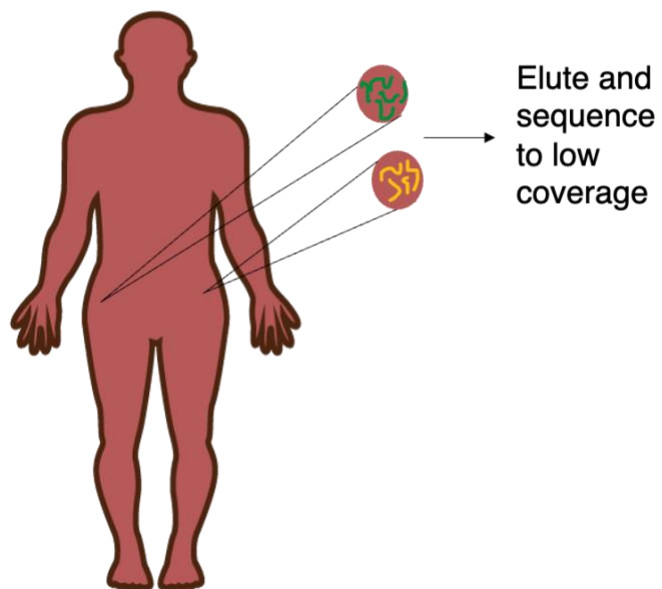
380
381
382   **Figure 4:** Schema of identifying parents of mf for which parentage is unknown. (Results Section 2)
383   39 mf were eluted from two skin-snips, from the left and right trochanter of a single patient.
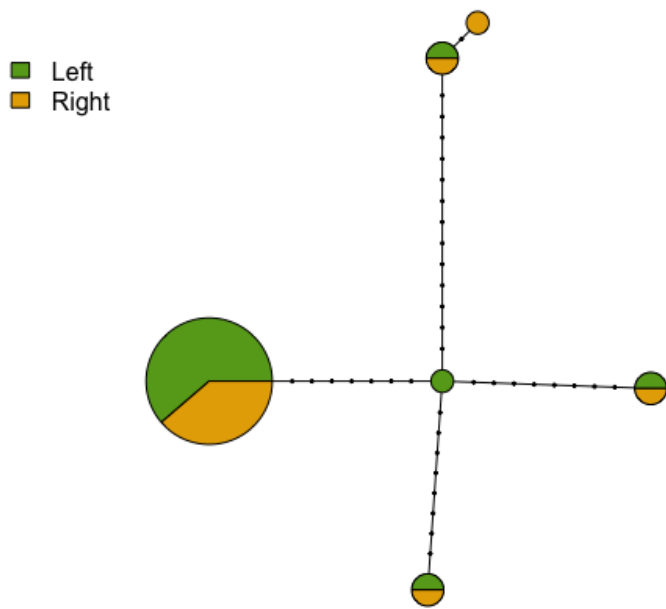
384
385 **Figure 5:** Medium-spanning network of mitochondrial haplotypes of skin-snip collected *Onchocerca*
386 *volvulus* microfilariae from two locations on the body of a single patient. Pie colours indicate the
387 proportion of samples taken from the left or right trochanter of the patient. Dots on connecting lines
388 indicate the number of single nucleotide polymorphisms (SNPs) between mitochondrial haplotype
389 groups.

Supplementary Figures
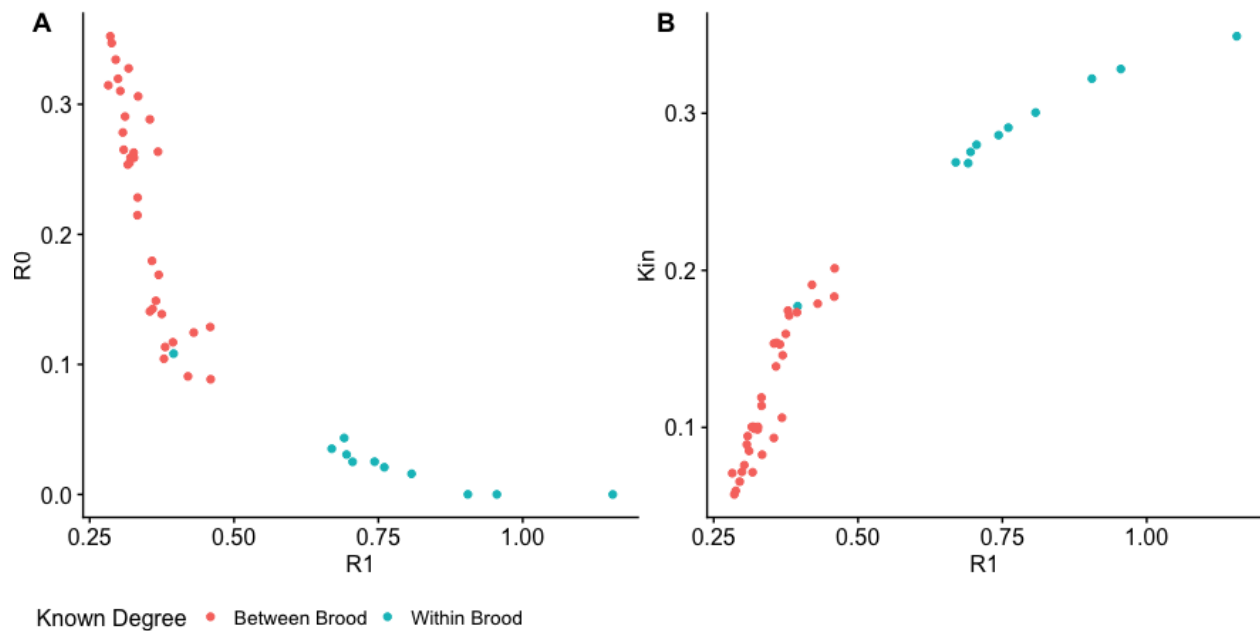
Supplementary Figure 1: Inferred kinship statistics for R0, R1 and KING-robust-kinship for pairs of intrauterine *Onchocerca volvulus* microfilariae, with points coloured by whether individuals in a pair were sampled from the same (within brood) or different (between brood) *O. volvulus* adult female worms.
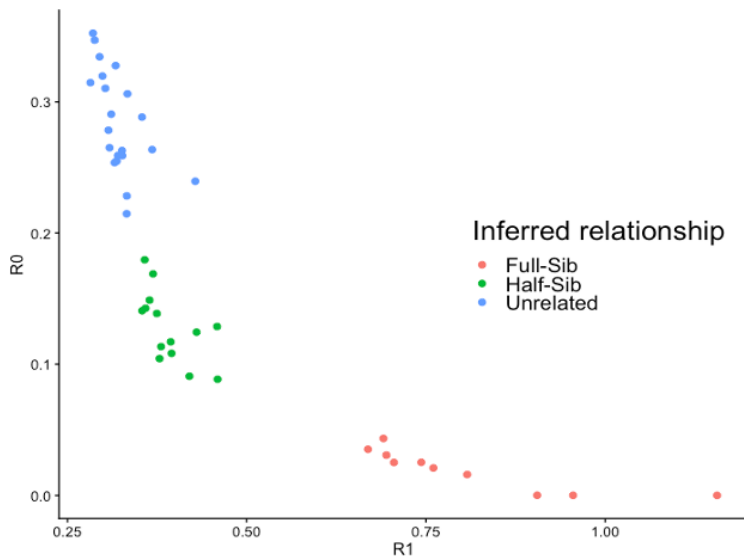
Supplementary Figure 2: Kinship statistics R0 and R1 for pairs of intrauterine *Onchocerca volvulus* microfilariae, coloured by the inferred relationship using thresholds for full-sibling (Full-Sib) relationships at R0 < 0.1, R1 > 0.5, and KING > 0.25 and thresholds for half-sibling (Half-Sib) relationships as R0 > 0.1, R1 < 0.5, and KING < 0.25, informed from Figure 1.

Bibliography

1. Hotez PJ, Alvarado M, Basáñez MG, Bolliger I, Bourne R, Boussinesq M, et al. The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. PLoS Negl Trop Dis. 2014 Jul 24;8(7):e2865.

2. Utzinger J, Booth M, N'goran EK, Müller I, Tanner M, Lengeler C. Relative contribution of day-to-day and intra-specimen variation in faecal egg counts of Schistosoma mansoni before and after treatment with praziquantel. Parasitology. 2001 May;122(5):537–44.

3. Mand S, Debrah A, Batsa L, Adjei O, Hoerauf A. Reliable and frequent detection of adult Wuchereria bancrofti in Ghanaian women by ultrasonography. Trop Med Int Health. 2004;9(10):1111–4.

4. Opoku NO, Bakajika DK, Kanza EM, Howard H, Mambandu GL, Nyathirombo A, et al. Single dose moxidectin versus ivermectin for Onchocerca volvulus infection in Ghana, Liberia, and the Democratic Republic of the Congo: a randomised, controlled, double-blind phase 3 trial. The Lancet. 2018 Oct;392(10154):1207–16.

5. Crellen T, Haswell M, Sithithaworn P, Sayasone S, Odermatt P, Lamberton PHL, et al. Diagnosis of helminths depends on worm fecundity and the distribution of parasites within hosts. Proc R Soc B Biol Sci. 2023 Jan 18;290(1991):20222204.

6. Hall A. Quantitative variability of nematode egg counts in faeces: a study among rural Kenyans. Trans R Soc Trop Med Hyg. 1981 Jan;75(5):682–7.

7. Duerr HP, Dietz K, Schulz-Key H, Büttner DW, Eichner M. Density-dependent parasite establishment suggests infection-associated immunosuppression as an important mechanism for parasite density regulation in onchocerciasis. Trans R Soc Trop Med Hyg. 2003 Mar;97(2):242–50.

8. Churcher TS, Filipe JAN, Basáñez MG. Density dependence and the control of helminth parasites. J Anim Ecol. 2006 Nov;75(6):1313–20.

9. Picq JJ, Jardel JP. Une methode d'evaluation des densites microfilariennes d'Onchocerca volvulus Leuckart, 1 893 chez des onchocerquiens.

10. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012 Dec;380(9859):2163–96.

11. Bakajika D, Senyonjo L, Enyong P, Oye J, Biholong B, Elhassan E, et al. On-going transmission of human onchocerciasis in the Massangam health district in the West Region of Cameroon: Better understanding transmission dynamics to inform changes in programmatic interventions. PLoS Negl Trop Dis. 2018 Nov;12(11):e0006904.

12. Koala L, Nikiema A, Post RJ, Paré AB, Kafando CM, Drabo F, et al. Recrudescence of onchocerciasis in the Comoé valley in Southwest Burkina Faso. Acta Trop. 2017 Feb;166:96–105.

13. Koala L, Nikièma AS, Paré AB, Drabo F, Toé LD, Belem AMG, et al. Entomological assessment of the transmission following recrudescence of onchocerciasis in the Comoé Valley, Burkina Faso. Parasit Vectors. 2019 Jan;12(1):34.

14. Eisenbarth A, Achukwi MD, Renz A. Ongoing Transmission of Onchocerca volvulus after 25 Years of Annual Ivermectin Mass Treatments in the Vina du Nord River Valley, in North Cameroon. PLoS Negl Trop Dis. 2016 Feb;10(2):e0004392.

15. Bah GS, Schneckener S, Hahnel SR, Bayang NH, Fieseler H, Schmuck GM, et al. Emodepside targets SLO-1 channels of Onchocerca ochengi and induces broad anthelmintic effects in a bovine model of onchocerciasis. PLOS Pathog. 2021 Jun;17(6):e1009601.

16. Krücken J, Holden-Dye L, Keiser J, Prichard RK, Townson S, Makepeace BL, et al. Development of emodepside as a possible adulticidal treatment for human onchocerciasis—

The fruit of a successful industrial–academic collaboration. PLOS Pathog. 2021 Jul;17(7):e1009682.

17.   Hübner MP, Martin C, Specht S, Koschel M, Dubben B, Frohberger SJ, et al. Oxfendazole mediates macrofilaricidal efficacy against the filarial nematode Litomosoides sigmodontis in vivo and inhibits Onchocerca spec. motility in vitro. PLoS Negl Trop Dis. 2020 Jul 6;14(7):e0008427.

18.   Schiefer A, Hübner MP, Krome A, Lämmer C, Ehrens A, Aden T, et al. Corallopyronin A for short-course anti-wolbachial, macrofilaricidal treatment of filarial infections. PLoS Negl Trop Dis. 2020 Dec 7;14(12):e0008930.

19.   Jones OR, Wang J. COLONY: a program for parentage and sibship inference from multilocus genotype data. Mol Ecol Resour. 2010;10(3):551–5.

20.   Steinauer ML, Christie MR, Blouin MS, Agola LE, Mwangi IN, Maina GM, et al. Non-Invasive Sampling of Schistosomes from Humans Requires Correcting for Family Structure. PLoS Negl Trop Dis. 2013 Sep;7(9):e2456.

21.   Gower CM, Gehre F, Marques SR, Lamberton PHL, Lwambo NJ, Webster JP. Phenotypic and genotypic monitoring of Schistosoma mansoni in Tanzanian schoolchildren five years into a preventative chemotherapy national control programme. Parasit Vectors. 2017 Dec;10(1):593.

22.   Johnson PCD, Hadfield JD, Webster LMI, Adam A, Mable BK, Keller LF. Bayesian paternity analysis and mating patterns in a parasitic nematode, Trichostrongylus tenuis. Heredity. 2010 Jun;104(6):573–82.

23.   Sticca EL, Belbin GM, Gignoux CR. Current Developments in Detection of Identity-by-Descent Methods and Applications. Front Genet. 2021;12:722602.

24.   Therkildsen NO, Palumbi SR. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Mol Ecol Resour. 2017;17(2):194–208.

25.   Doyle SR, Sankaranarayanan G, Allan F, Berger D, Jimenez Castro PD, Collins JB, et al. Evaluation of DNA Extraction Methods on Individual Helminth Egg and Larval Stages for Whole-Genome Sequencing. Front Genet [Internet]. 2019 [cited 2022 Dec 18];10. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2019.00826

26.   Hedtke SM, Kuesel AC, Crawford KE, Graves PM, Boussinesq M, Lau CL, et al. Genomic epidemiology in filarial nematodes: transforming the basis for elimination program decisions. Front Genet. 2020;10:1282.

27.   Durrant C, Thiele EA, Holroyd N, Doyle SR, Sallé G, Tracey A, et al. Population genomic evidence that human and animal infections in Africa come from the same populations of Dracunculus medinensis. Basáñez MG, editor. PLoS Negl Trop Dis. 2020 Nov;14(11):e0008623.

28.   Waples RK, Albrechtsen A, Moltke I. Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. Mol Ecol. 2019;28(1):35–48.

29.   Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010 Nov;26(22):2867–73.

30.   Hildebrandt JC, Eisenbarth A, Renz A, Streit A. Reproductive biology of Onchocerca ochengi, a nodule forming filarial nematode in zebu cattle. Vet Parasitol. 2014 Sep;205(1):318–29.

31.   Ackerman MW, Hand BK, Waples RK, Luikart G, Waples RS, Steele CA, et al. Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. Evol Appl. 2017 Feb;10(2):146–60.

32.   Anderson RM, Turner HC, Farrell SH, Truscott JE. Chapter Four - Studies of the Transmission Dynamics, Mathematical Model Development and the Control of Schistosome Parasites by Mass Drug Administration in Human Communities. In: Basáñez MG, Anderson RM, editors. Advances in Parasitology [Internet]. Academic Press; 2016 [cited 2023 Jun 27].

506   p. 199–246. (Mathematical Models for Neglected Tropical Diseases; vol. 94). Available from:
507   https://www.sciencedirect.com/science/article/pii/S0065308X16300720

508   33.   Minter A, Pellis L, Medley GF, Hollingsworth TD. What Can Modeling Tell Us About
509   Sustainable End Points for Neglected Tropical Diseases? Clin Infect Dis. 2021 Jun
510   15;72(Supplement_3):S129–33.

511   34.   Geary TG, Mackenzie CD, Silber SA. Flubendazole as a macrofilaricide: History and
512   background. Lustigman S, editor. PLoS Negl Trop Dis. 2019 Jan;13(1):e0006436.

513   35.   Neves MI, Gower CM, Webster JP, Walker M. Revisiting density-dependent fecundity in
514   schistosomes using sibship reconstruction. PLoS Negl Trop Dis. 2021 May;15(5):e0009396.

515   36.   Neves MI, Webster JP, Walker M. Estimating helminth burdens using sibship
516   reconstruction. Parasit Vectors. 2019 Dec;12(1):441.

517   37.   Cotton JA, Bennuru S, Grote A, Harsha B, Tracey A, Beech R, et al. The genome of
518   Onchocerca volvulus, agent of river blindness. Nat Microbiol. 2016 Nov;2(2):1–12.

519   38.   Huisman J. Pedigree reconstruction from SNP data: parentage assignment, sibship
520   clustering and beyond. Mol Ecol Resour. 2017;17(5):1009–24.

521   39.   Hanghøj K, Moltke I, Andersen PA, Manica A, Korneliussen TS. Fast and accurate
522   relatedness estimation from high-throughput sequencing data in the presence of inbreeding.
523   GigaScience [Internet]. 2019 May [cited 2023 Jan 1];8(5). Available from:
524   https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giz034/5481763

525   40.   Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive
526   Multiplexed Library Preparation for Megabase-Sized Genomes. PLOS ONE. 2015
527   May;10(5):e0128036.

528   41.   Choi YJ, Tyagi R, McNulty SN, Rosa BA, Ozersky P, Martin J, et al. Genomic diversity in
529   Onchocerca volvulus and its Wolbachia endosymbiont. Nat Microbiol. 2016 Nov;2(2):16207.

530   42.   Doyle SR, Bourguinat C, Nana-Djeunga HC, Kengne-Ouafo JA, Pion SDS, Bopda J, et
531   al. Genome-wide analysis of ivermectin response by Onchocerca volvulus reveals that
532   genetic drift and soft selective sweeps contribute to loss of drug sensitivity. Unnasch TR,
533   editor. PLoS Negl Trop Dis. 2017 Jul;11(7):e0005816.

534   43.   R Core Team. R: A language and environment for statistical computing. R Foundation
535   for Statistical Computing, Vienna, Austria. [Internet]. [cited 2023 May 3]. Available from:
536   https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-
537   rivers/r-development-core-team-2006

538   44.   Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to
539   the Tidyverse. J Open Source Softw. 2019 Nov;4(43):1686.

540   45.   Neuwirth E. RColorBrewer: ColorBrewer Palettes [Internet]. 2022 [cited 2023 Jan 1].
541   Available from: https://CRAN.R-project.org/package=RColorBrewer

542   46.   Wilke CO. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2' [Internet].
543   2020 [cited 2023 Jan 1]. Available from: https://CRAN.R-project.org/package=cowplot

544   47.   Detection of Onchocerca volvulus Infection by O-150 Polymerase Chain Reaction
545   Analysis of Skin Scratches on JSTOR [Internet]. [cited 2023 Jun 27]. Available from:
546   https://www.jstor.org/stable/30114150

547   48.   Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
548   data. Bioinformatics. 2014 Aug;30(15):2114–20.

549   49.   Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler
550   transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

551   50.   McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
552   Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA
553   sequencing data. Genome Res. 2010 Sep;20(9):1297–303.

554   51.   Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of
555   SAMtools and BCFtools. GigaScience. 2021 Jan 29;10(2):giab008.

556   52.   Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation

557   Sequencing Data. BMC Bioinformatics. 2014 Nov 25;15(1):356.

558   53.   Csardi G, Nepusz T. The igraph software package for complex network research.

559   54.   Garrison, Erik, Marth, Gabor. Haplotype-based variant detection from short-read
560   sequencing. arRXIV. 2012 Jul 4;1207.3907.

561   55.   Paradis E. pegas: an R package for population genetics with an integrated-modular
562   approach. Bioinformatics. 2010 Feb;26(3):419–20.