

1 **Title: Synthesising environmental, epidemiological, and genetic**  
2 **data to assist decision making for onchocerciasis elimination**

3 Himal Shrestha (1), Karen McCulloch (1), Rebecca H Chisholm (2, 3), Samuel Armoo (4),  
4 Francis Vierigh (4), Neha Sirwani (1), Katie E Crawford (1), Mike Osei-Atweneboana (4),  
5 Warwick N Grant (1), Shannon M Hedtke (1)

6 1. Department of Environment and Genetics, School of Agriculture, Biomedicine and  
7 Environment, La Trobe University, Bundoora, Australia

8 2. Department of Mathematics and Statistics, La Trobe University, Bundoora, Australia

9 3. Centre for Epidemiology and Biostatistics, Melbourne School of Population and  
10 Global Health, The University of Melbourne, Melbourne, Australia

11 4. Biomedical and Public Health Research Unit, CSIR-Water Research Institute, Accra,  
12 Ghana

13 Corresponding author/s: Dr Shannon M Hedtke, [S.Hedtke@latrobe.edu.au](mailto:S.Hedtke@latrobe.edu.au)

## 14 **Abstract**

15 **Background:** Population genetics is crucial for understanding the transmission dynamics of  
16 diseases like onchocerciasis. Landscape genetics identifies the ecological features that impact  
17 genetic variation between sampling sites. Here, we have used a landscape genetics framework  
18 to understand the relationship between environmental features and gene flow of the filarial  
19 parasite *Onchocerca volvulus* and of its intermediate host and vector, blackflies in the genus  
20 *Simulium*. We analysed samples from the ecological transition region separating the savannah  
21 and forest ecological regions of Ghana, where the transmission of *O. volvulus* has persisted  
22 despite almost half a century of onchocerciasis control efforts.

23 **Methods:** We generated a baseline microfilarial prevalence map from the point estimates of  
24 pre-ivermectin microfilarial prevalence from 47 locations in the study area. We analysed  
25 mitochondrial data from 164 parasites and 93 blackflies collected from 15 communities and  
26 four breeding sites, respectively. We estimated population genetic diversity and identified  
27 correlations with environmental variables. Finally, we compared baseline prevalence maps to  
28 movement suitability maps that were based on significant environmental variables.

29 **Results:** We found that the resistance surfaces derived from elevation ( $r = 0.793$ ,  $p = 0.005$ )  
30 and soil moisture ( $r = 0.507$ ,  $p = 0.002$ ) were significantly associated with genetic distance  
31 between parasite sampling locations. Similarly, for the vector populations, the resistance  
32 surfaces derived from soil moisture ( $r = 0.788$ ,  $p = 0.0417$ ) and precipitation ( $r = 0.835$ ,  $p =$   
33  $0.0417$ ) were significant. The correlation between the baseline parasite prevalence map and  
34 the parasite resistance surface map was stronger than the correlation between baseline  
35 prevalence and the vector resistance surface map. The central parts of the transition region  
36 which were conducive for both the parasite and the vector gene flow were most strongly  
37 associated with high baseline onchocerciasis prevalence.

38 **Conclusions:** We present a framework for incorporating environmental, genetic, and  
39 prevalence data for identifying when ecological conditions are favourable for onchocerciasis  
40 transmission between communities. We identified areas with higher suitability for parasite and  
41 vector gene flow, which ultimately might help us gain deeper insights into defining  
42 transmission zones for onchocerciasis. Furthermore, this framework is translatable to other  
43 onchocerciasis endemic areas and to other vector-borne diseases.

44 **Keywords:** onchocerciasis, *Onchocerca volvulus*, *Simulium damnosum*, population genetics,  
45 disease ecology, landscape genetics, transmission zones, persistence of transmission, Ghana

## 46 **Background**

47 Onchocerciasis is a neglected tropical disease caused by a filarial parasite, *Onchocerca*  
48 *volvulus*, and transmitted by the bites of black flies (*Simulium* spp.). The blackflies have a  
49 narrow range of ecological suitability, which leads to spatial heterogeneity in the prevalence  
50 and transmission of onchocerciasis [1–4]. The primary tool for onchocerciasis control is mass  
51 drug administration with ivermectin (MDAi) with an initial focus on mostly high endemic  
52 communities, i.e., there is also spatial heterogeneity in intervention history. Following the  
53 success of MDAi in controlling onchocerciasis as a significant public health problem in the  
54 majority of areas, almost all countries have switched their target from control to elimination.  
55 However, the target of onchocerciasis elimination with MDAi is impeded by some persistent  
56 onchocerciasis transmission foci despite decades of intervention [5–7].

57 Understanding the persistence of disease transmission requires spatial heterogeneity to be  
58 considered because of the risk that movement of infective vectors, and thus parasites, from the  
59 areas with different endemicity and MDAi history can re-initiate disease in areas where  
60 transmission of *O. volvulus* is thought to be eliminated. For instance, the migration of the  
61 parasites via infected humans has been linked to recrudescence in previously eliminated foci  
62 of Burkina Faso [8–10]. Similarly, the failure to achieve the elimination of onchocerciasis in  
63 West Africa with the onchocerciasis control program (OCP) was attributed to rapid insecticide  
64 resistance due to high vector gene flow and, thus, the spread of insecticide resistance alleles  
65 [11–14]. However, disease control programs have historically focused on government  
66 administrative units as the unit of intervention, which has led to a situation where treatment  
67 decisions are being made without much consideration of host- or vector-mediated movement  
68 of the parasites and, thus, the transmission zones.

69 The geographical unit in which parasite transmission occurs via locally breeding vectors is  
70 termed as a transmission zone [15]. Transmission zones form the biological basis of  
71 intervention units, and thus, a clear understanding of transmission zones and means to define  
72 their boundaries are crucial to ensure that the interventions are coordinated at the correct  
73 geographic scale. Onchocerciasis prevalence is high in the poorest of the poor nations of the  
74 world [12,16]. Therefore, the limited resources available in these areas must be judiciously  
75 allocated to the most essential areas to achieve the elimination of onchocerciasis transmission.  
76 The way forward to achieving the elimination goal is to align intervention units as closely as

77 possible to the natural transmission zones. However, delineating transmission zones is a  
78 challenging task, and several tools have been deployed so far to understand transmission  
79 zones.

80 We can gain some insights into the transmission zones based on prevalence mapping, where  
81 point prevalence data are interpolated spatially [4,17]. However, this is a static map and  
82 ignores the 'innate' connectivity between locations mediated by the movement of the human  
83 host and the vectors. Population genetics has been used to infer the movement of pathogens,  
84 whereby pathogen movement can be measured indirectly by the genetic relatedness of  
85 parasites across locations [18–28]. The dispersal, and thus gene flow, of parasites and vectors,  
86 are subject to influence by the environmental features of the landscape. Therefore, population  
87 genetics should be combined with spatial information and environmental data in order to  
88 provide a better picture of the transmission processes. This combination of spatial information,  
89 environmental data and population genetics is termed landscape genetics.

90 Landscape genetics explicitly quantifies the effects of landscape on evolutionary processes  
91 such as gene-flow, drift, and selection [29,30]. Spatial information can be added in the form  
92 of sampling location geographic coordinates and remote sensing satellite images of different  
93 environmental and climate variables such as elevation, slope, distance to the water bodies etc.  
94 There are then several steps required in order to use landscape genetics to infer transmission  
95 zones. First, the degree of genetic differentiation between sampling locations for parasites  
96 and/or vectors is measured. Second, the extent of correlation between a range of environmental  
97 variables and the measures of genetic differentiation estimated in step one is determined [31].  
98 Third, the most important environmental variables identified in step two are converted to  
99 resistance surface maps, which quantify the barriers to the gene flow of the study population  
100 in a pixel-level landscape map and are a proxy for the movement suitability of an organism in  
101 that particular landscape, i.e., high resistance implies low gene flow/mobility and low  
102 resistance implies high gene flow/mobility [32,33]. Resistance maps can be used to simulate  
103 the pattern of gene flow of the parasites and the vectors, giving insights into the predicted  
104 corridors of movement and, thus, the likelihood of transmission between locations [34,35].

105 We have implemented this technique in the ecological transition region of Ghana, an  
106 onchocerciasis hotspot of concern. Despite half a century of interventions, *O. volvulus*

107 transmission still persists in some communities [5,36,37], and there are also reports of  
108 suboptimal response (SOR) of infections to treatment with ivermectin [38–40]. A recent  
109 population genetic analysis by Crawford et al. [25], suggested a genetically homogeneous  
110 parasite populations in this area with the absence of isolation-by-distance, i.e., genetic  
111 connectivity of the parasite population not limited by the geographic distance between the  
112 population. This suggests cross-transmission of *O. volvulus* between communities, which may  
113 be contributing to the persistence of onchocerciasis transmission. With the hypothesis that the  
114 genetic connectivity is influenced by environmental factors, we used a landscape genetics  
115 framework to understand the spatial patterns of transmission in the ecological transition region  
116 of Ghana.

117 We have combined environmental data with the parasite genetic data (and have included  
118 additional vector genetic data from the ecological transition regions) with the objectives of:  
119 (i) determining the ecological factors affecting the spatial variation in the parasite and the  
120 vector population genetic estimates and; (ii) inferring the patterns and routes of gene-flow,  
121 and thus the likely transmission, for the parasite and the vector populations. We have identified  
122 key environmental variables that influence the population genetic structure of the parasite and  
123 the vector population and generated gene flow maps for the parasite and the vector population  
124 from the ecological resistance surface maps. This allowed us to identify potential corridors of  
125 parasite and vector movement between the sampling communities, which provides an  
126 evidence base for spatial delineation of transmission zones. Further, we have compared the  
127 movement suitability maps with the baseline microfilarial (mf) prevalence maps and discussed  
128 the immediate implications of the approach developed to aid elimination goals.

## 129 **Methods**

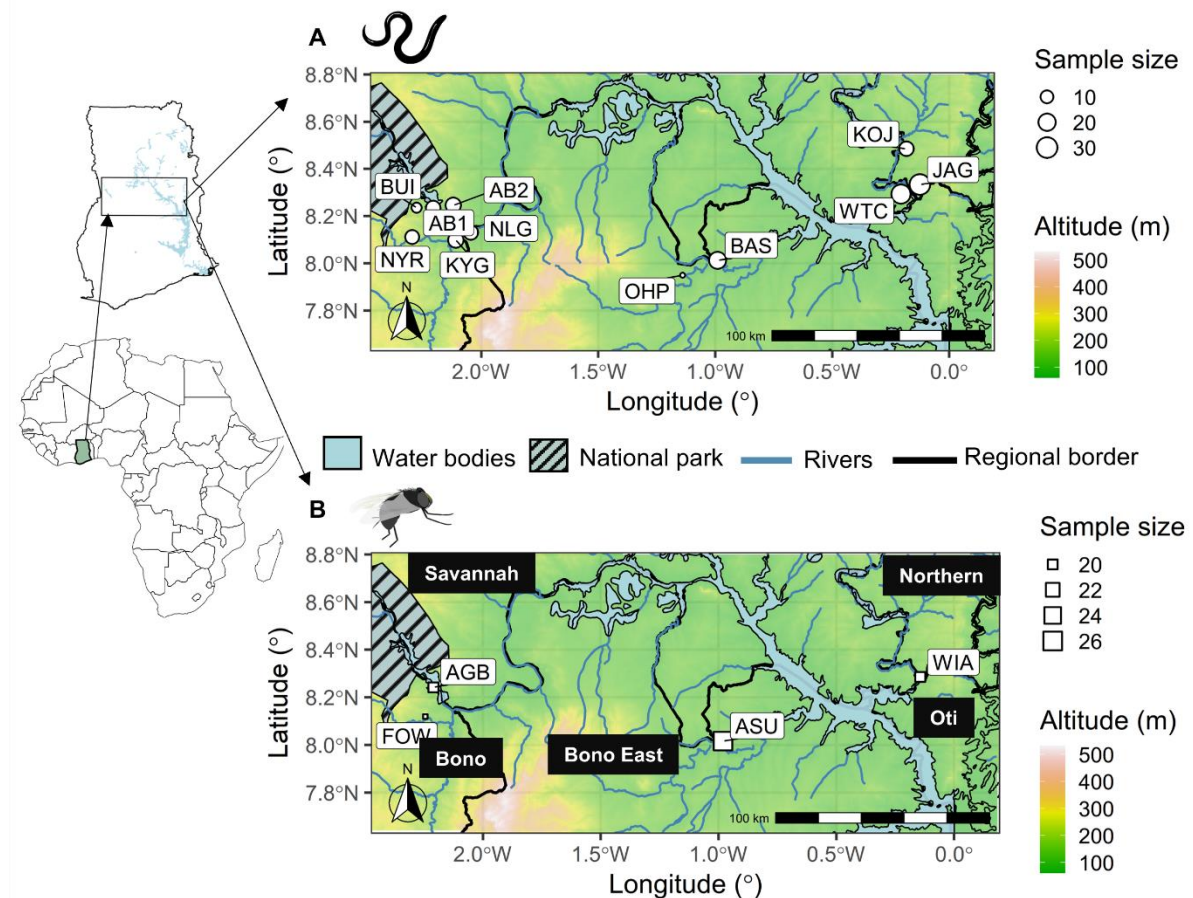
### 130 **Sampling locations**

131 The study area is a west-east transect in the ecological "transition zone" of Ghana: an area that  
132 includes the savannah ecotype in the north and the forest ecotype in the south [41–43], with  
133 the Lake Volta bisecting the eastern parts of the transition zone, and the Bui National Park in  
134 the west (Figure 1). We chose this area for the study as there is ongoing persistence of *O.*  
135 *volvulus* transmission despite decades of control efforts [36,40,44]. The elevation ranges from  
136 70–525 m above sea level, and mean annual temperature and precipitation range from 24–  
137 29°C and 1077–1355 mm, respectively [45,46].

138 The sampling locations belonged to four different government administrative regions, viz.,  
139 Bono, Bono East, Savannah, and the Northern regions (Figure 1; Additional file Table S3).  
140 Variant calls based on mitochondrial genome data from 164 female *O. volvulus* samples that  
141 had been isolated from 97 people from 15 communities, primarily in 2010–2012 (n = 107) and  
142 in 2006 (n = 34) and 2013 (n = 23), were obtained from Crawford et al. [25]. Ethics approvals  
143 for sampling parasites from people are reported in Crawford et al. [25]. Four communities  
144 from each of the four regions, viz., Bono, Bono East, Savannah and Northern region, were  
145 chosen for the sequencing of vector samples which were collected in 2013–2015. A total of  
146 93 *S. damnosum* samples collected in 2013 (n = 73) and 2015 (n = 20) by human landing catch  
147 were selected from four communities.

148 A bounding box formed based on the convex hull boundary (a boundary with a set of convex  
149 curves enclosing the sampling locations) around the sampling locations was used for the  
150 geospatial analysis. The dimension for the bounding box was 293.68×129.38 km (an area of  
151 37,995.59 km<sup>2</sup>). Geographic coordinates for all the communities were used to calculate the  
152 pairwise geographic distance between the communities (Additional file Table S3). We  
153 aggregated data from communities close to each other (less than 5 km) and used the centroid  
154 of the geospatial coordinates of the communities in close proximity for the merged  
155 communities. This brought the number of parasite sampling locations down to 11 but increased  
156 the sample size per community (Figure 1).





157

158 **Figure 1. The spatial context of the sampling locations of the *Onchocerca volvulus* and *Simulium***  
159 ***damnosum* in the transition region of Ghana.** Geographic coordinates are represented as the circle  
160 for parasites (A) and square for vectors (B), and their sizes correspond to the number of samples from  
161 the respective locations. The legend for the size is provided to the left of each figure. The communities  
162 are represented with community codes. The river lines and the government administrative borders are  
163 shown along with the water body (Lake Volta) and the Bui national park. The inset map shows the  
164 map of Africa and Ghana with the bounding box for our study area. More information about sampling  
165 locations and the number of samples are present in Additional file Table S3.

## 166 Sequencing and variant calling

167 Details on the genetic data generation and the parasite samples are available in Crawford et al.  
168 [25]. In brief, DNA was extracted from adult female *O. volvulus* from nodules using the  
169 Dneasy® Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions.  
170 Sequence libraries were generated based on either genomic DNA extracts or on amplicons  
171 targeting the mitochondrial genome and sequenced using Illumina MiSeq or HiSeq sequencing



172 platforms. Trimmed sequence reads were mapped to the *O. volvulus* (NC\_001861)  
173 mitochondrial reference genome and variants called using *GATK UnifiedGenotyper* [47].  
174 These data were submitted to the NCBI Short Read Archive under project PRJNA560089 [48].

175 For *S. damnosum*, the body of each fly was dissected and homogenised using a pestle.  
176 Extractions of total DNA were performed using the Isolate II Genomic DNA kit, following  
177 the manufacturer's instructions (Biolone, London, United Kingdom). Sequencing libraries  
178 were constructed and indexed using the Illumina DNA Prep tagmentation kit following the  
179 manufacturer's instructions (Illumina, San Diego, California, USA). Libraries were pooled and  
180 sequenced on one lane of a NovaSeq SP, 300 cycles (resulting in 150-bp paired-end reads) at  
181 the Australian Genome Research Facility (Melbourne, Victoria, Australia) (Additional file  
182 Table S1).

183 Sequenced reads were trimmed for quality and to remove adapter contamination using  
184 *trimmomatic* v.0.32 and keeping only those pairs where both pairs were >125 bp [49]. To  
185 assemble the genome, three flies with the largest number of paired reads were mapped using  
186 *bwa* v. 0.7.17 [50] to available *Simulium* spp. Complete or nearly complete mitochondrial  
187 genomes downloaded from NCBI (*Simulium variegatum*, NC\_033348; *Simulium noelleri*,  
188 NC\_050320; *Simulium quinquestriatum*, MK281358; *Simulium ornatum*, MT410845;  
189 *Simulium maculatum*, NC\_040120; *Simulium aureohirtum*, NC\_029753; *Simulium*  
190 *petricolum*, MT671497; *Simulium equinum*, MT920425; *Simulium angustipes*, MT628576;  
191 *Simulium lundstromi*, MT628562). Those reads that mapped to any *Simulium* genome were  
192 extracted and converted to fastq using *samtools* v.1.9 [51], and these were used to produce a  
193 preliminary assembly using *spades* v. 3.11.1 [52] and *velvetoptimiser* v. 2.2.5 [53,54]. These  
194 drafts were then improved using *pilon* v.1.23 [55]. Assemblies from the two different  
195 programs were aligned in Mesquite v.3.61 [56], and the consensus—defined as bases that were  
196 observed in both assemblies—was taken to produce a single consensus reference genome (i.e.,  
197 the consensus from two variant callers from one blackfly) for variant calling. Because  
198 mitochondrial genomes are circular, and thus the starting point for different linear assemblies  
199 differed, the assembly for each fly was oriented so that it began with tRNA-Ile to be consistent  
200 with *S. variegatum* (NC\_033348; [57]). The "AT-rich region" was variable in inferred length  
201 and sequence between different assemblers, different individual blackflies, and different  
202 species, and were difficult to align. Thus, this AT-rich, variable-length region was excluded.

203 All raw reads and assembled sequences were submitted to the European Nucleotide Archive  
204 (ENA) at EMBL-EBI under accession number PRJEB57094.

205 Variants were filtered to retain only those calls at positions with a minimum quality score of  
206 30 and a minimum depth of 20 using *vcftools* v.0.1.13 [50,58,59]. Individuals with more than  
207 75% missing data were excluded from the analysis. Variants were normalised using  
208 *bcftools* v.1.2. To ensure consistency between variant formatting, allelic primitives were  
209 called using the function *vcfallelicprimitives* implemented in *vcflib* [60]. The intersection of  
210 the two variant callers was then identified using *bcftools* v.1.2 [61]. For both parasite and  
211 vector data, we filtered the variants to remove indels, missing regions, and non-biallelic sites  
212 using *vcftools* v.0.1.13 [58]. The resulting dataset comprised 189 SNP loci for 164 individual  
213 *O. volvulus* and 632 SNP loci for 93 individual *S. damnosum*.

#### 214 **Prevalence data**

215 Pre-MDAi prevalence data for communities that fell within the study area bounding box and  
216 were based on observation of mf in a skin biopsy via microscopy were obtained from the  
217 Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) database  
218 [62]. The prevalence data collected for mapping, i.e., prior to MDAi was used. Duplicate  
219 observations were removed, and observations from the same geographic coordinates at  
220 different years were aggregated to calculate the average prevalence. There were 47 unique  
221 locations with prevalence data collected from 1976 to 2004 that fell within the study area used  
222 for the geospatial mapping of the baseline prevalence.

#### 223 **Environmental data**

224 We compiled different continuous environmental rasters which might be ecologically relevant  
225 to the onchocerciasis distribution based on the published literature, field experiments on  
226 blackflies [63,64] and ecological factors identified with previous geospatial modelling studies  
227 [2,17,65–67]. These environmental variables included distance to the nearest river, soil  
228 moisture, elevation, slope, temperature, and precipitation [2,65,67]. In addition, the dispersal  
229 capacity of the *Simulium* vector is dependent on the vegetation type and time of the year [68].  
230 Therefore, we included vegetation and seasonality-related variables in our analysis. In addition  
231 to environmental variables, we also included some sociodemographic aspects of the study  
232 area—for example, the human population density to consider the availability of human hosts

233 for disease transmission. We used the environmental variables corresponding to the year when  
234 the samples were collected for fitting the models to account for the differences in the time of  
235 sampling. For prevalence data, environmental variables before 2001 were used, and similarly,  
236 for the *O. volvulus* and *S. damnosum*, environmental variables from 2010–2012 and 2013–  
237 2015 were used respectively, as per the data availability. Our starting set of environmental and  
238 socio-economic datasets consisted of 32 continuous environmental rasters at a spatial  
239 resolution of 1 km from publicly available repositories via Earth Engine (Additional file Table  
240 S2) [69].

241 These variables were divided into six groups, viz., temperature, precipitation, topography,  
242 vegetation indices, hydrological and sociodemographic variables. We extracted the values for  
243 each sample location using the *raster* package in R v. 4.1.0 [70,71]. For testing the association  
244 of the landscape factors to the genetic differentiation or gene flow between the populations, a  
245 pairwise comparison of environmental characteristics between sampling locations is crucial  
246 [29,72]. Thus, we calculated the average of the values encountered by a pairwise straight path  
247 between each sampling site to account for the features in adjacent areas around sampling sites  
248 for all the environmental and sociodemographic variables. We generated a pairwise  
249 correlation matrix for all 32 variables to identify variables that are highly correlated with  
250 prevalence (Additional file Figure S2, S4). We included only those variables where Pearson's  
251 correlation coefficient between the ecological variable(s) and prevalence was less than  $< |0.6|$   
252 within each group of variables [72]. Further, we performed principal component analysis  
253 (PCA) to identify the variables that contributed most to the variance among the group of  
254 correlated variables (Additional file Figure S1, S3) [73]. For any given group of correlated  
255 variables, we selected the variable with the highest contribution score to the total variance in  
256 PCA analysis and the ease of interpretability of the variables. The environmental variables  
257 selected for the parasite sampling locations were also used for vector landscape genetics for  
258 easier comparison between the vector and the parasite landscape genetics.

### 259 **Prevalence mapping**

260 The mean of the posterior prevalence was obtained from the pre-MDAi mf prevalence data  
261 using the Bayesian approach with Integrated Nested Laplace Approximation (INLA) [74,75].  
262 The number of positive cases out of the total number of people tested in a location was  
263 assumed to follow a binomial distribution. The prevalence was modelled with different

264 environmental variables and a spatial random effect with a zero-mean Gaussian process  
265 following a Matérn covariance function. The Matérn field is represented with a finite element  
266 mesh formed of triangles around the sampling locations and adding vertices over the  
267 prediction region. Multiple triangulation meshes with different parameters for cut-off and  
268 length of triangles inside and outside the boundary were tested for the model fit and  
269 computational cost (Additional file Figure S5). We created a triangulation mesh with a 3 km  
270 cut-off; the maximum length of triangles inside and outside the boundary was set to 10 km  
271 and 100 km, respectively. Finally, we fitted the model and assessed the relationship of  
272 environmental variables with the prevalence data. The details of fitting a spatial model to the  
273 prevalence data for geospatial mapping are available in [2]. The prediction of the posterior  
274 prevalence was made at a 2 km resolution considering the high computational cost of  
275 prediction on a lower resolution.

## 276 **Population genetic analysis**

277 For the parasite and the vector samples, we carried out unsupervised *k*-means clustering  
278 analysis using the *adegenet* v. 2.1.6 package [76]. We inferred the optimal number of *k*  
279 (groups) for the population using unsupervised *k*-means clustering with the Bayesian  
280 Information Criterion (BIC). The vector results were consistent with the results of a haplotype  
281 network analysis using *PopART* [77] that identified outlier blackflies separated largely from  
282 the cluster of other samples. Given the taxonomic uncertainty of the species composition of  
283 the *S. Damnosum* complex, these outliers could not be assigned confidently as members of the  
284 same interbreeding population that we believe comprised the bulk of the black flies in the  
285 sample and were therefore excluded from the analysis. Then, we carried out a Discriminant  
286 Analysis of the Principal Components (DAPC) using communities as populations. DAPC is  
287 sensitive to the number of principal components retained. Therefore, we performed stratified  
288 cross-validated DAPC by varying the number of principal components using *xvalDapc*  
289 function in the *adegenet* v. 2.1.6 package. We calculated the membership probability of each  
290 sample, communities, and the posterior correct assignment probability for the communities.  
291 We calculated summary statistics for the genetic data, i.e., number of alleles, observed gene  
292 diversity, and the pairwise measure of genetic differentiation ( $F_{st}$ ) between sampling locations  
293 using the *Hierfstat* v. 0.5.11 package [78]. Similarly, mean allelic richness and number of  
294 haplotypes were calculated using *PopGenReport* v. 3.0.4 and *haplotypes* v. 1.1.2 package,

295 respectively [79,80]. The pairwise  $F_{st}$  matrix was adjusted for finite populations by linearising  
296 it with the equation  $F_{st}/(1 - F_{st})$  as suggested by [73,81,82].

### 297 **Landscape genetic analysis**

298 Landscape genetics analysis helps us understand how landscape features influence the spatial  
299 distribution of genetic variation. The simplest starting model is the isolation-by-distance  
300 model, where we test if there is a correlation between the pairwise genetic distance and the  
301 pairwise straight-path geographic distance between the sampling sites [30,83,84]. The  
302 geographic distance was calculated as the pairwise Euclidean distance between the geographic  
303 coordinates of the sampling sites using the *graph4lg* v. 1.6.0 package [85]. Geographic  
304 coordinates were converted to the Universal Transverse Mercator projection, a two-  
305 dimensional cartesian coordinate referencing system that is accurate when performing  
306 distance-related operations on spatial objects [86]. The coordinate referencing system used in  
307 our analysis for all the spatial objects was: `epsg-32630 (+proj=utm +zone=30`  
308 `+datum=WGS84 +units=m +no_defs)`. The pairwise linearised genetic differentiation  
309 between sites was considered a genetic distance. We performed Mantel tests between the  
310 geographic distance and the genetic distance matrix with the *vegan* v. 2.6.2 package, and the  
311 significance of the correlation was calculated based on 10000 permutations [87].

### 312 **Resistance surface maps**

313 In addition to geographic distances, we calculated ecological cost distances to assess the effect  
314 of intervening landscape features between the sampling sites on spatial genetic variation  
315 [31,88]. The ecological cost distances were calculated based on "resistance surface" maps.  
316 The values in each pixel of a resistance surface map reflect the extent to which the landscape  
317 feature on that pixel impedes or facilitates the movement or connectivity of the populations of  
318 interest between different locations [33,35]. We used *Circuitscape* implemented in Julia  
319 v. 1.6.1 to calculate the circuit distance, a proxy for the ecological cost distances, to generate  
320 connectivity maps and identify corridors for movement in the landscape [89].

321 The resistance surface maps were generated from the environmental variables using a search  
322 and optimisation method, where transformation parameters were explored to maximise the  
323 association between the pairwise genetic distance and the ecological cost distance using  
324 *ResistanceGA* v. 4.1.46 package [33]. The package uses a genetic algorithm to optimise

325 resistance surface parameters and offers eight transformations of ricker and monomolecular  
326 functions to a continuous surface. The following equations give the ricker and monomolecular  
327 transformation function:

328 Ricker transformation:  $resistance = raster \times e^{-magnitude \times shape}$

329 Monomolecular transformation:  $resistance = raster \times (1 - e^{-magnitude \times shape})$

330 The algorithm searches for the best combination of a transformation function, magnitude, and  
331 shape parameter. It provides a framework for optimising resistance surfaces from an  
332 environmental raster surface without any prior assumptions about the contribution of those  
333 surfaces on the resistance [33] and, therefore, provides an unbiased representation of the  
334 resistance surface based on genetic data.

335 The environmental variables selected for landscape genetic analysis were used to optimise the  
336 resistance surface maps. Linearised pairwise  $F_{st}$  genetic distance between sampling locations  
337 was used as the response parameter. The cost distance calculated from the transformed  
338 resistance surfaces was used as a predictor to find the best model that explains the genetic  
339 distance. A linear mixed-effects model with a maximum likelihood population effect (MLPE)  
340 was fitted to the data [90,91]. We optimised single surfaces of environmental variables and  
341 used the log-likelihood as the objective function for the MLPE model. Four replicates of 1000  
342 iterations each were run with the optimisation set to stop after 50 generations of no  
343 improvement. We set the maximum allowable resistance value to 100 during the optimisation  
344 process for easier rescaling and comparison of the resistance values of different environmental  
345 variables.

346 Each replicate of the resistance surface obtained via the optimisation process was tested using  
347 the circuit distance matrix obtained from those resistance surfaces. We used the partial Mantel  
348 test to assess the correlation between the genetic distance matrix and the pairwise circuit  
349 distance matrix accounting for the geographical distance matrix. The partial Mantel test is  
350 used frequently in landscape genetics analyses but has high type I error rates with spurious  
351 correlations [92]. Therefore, we used mixed matrix regression with randomisation (MMRR)  
352 as a confirmatory test. The MMRR was performed using the *lgMMRR* function in the  
353 *PopGenReport* v. 3.0.4 package based on Wang's (2013) method. The MMRR also gives us



354 the effect of the resistance surface on the genetic differentiation accounting for the geographic  
355 distances. To avoid spurious correlations, we took a conservative approach, and the resistance  
356 surfaces were deemed significantly associated with the genetic distance only if both the partial  
357 mantel and MMRR tests were statistically significant [73,94]. Significance for both the partial  
358 Mantel and MMRR were assessed based on 10,000 permutations.

### 359 **Composite resistance surface maps**

360 As landscape features and environmental gradients do not exist in isolation, the environmental  
361 resistance surfaces significantly associated with the genetic distance matrix were manually  
362 combined to form a composite resistance surface map. They were rescaled from 0 to 1, where  
363 the maximum resistance value among all the significant surfaces was considered as 1,  
364 preserving the relative contribution of each optimised surface to the composite resistance map.  
365 The composite resistance map was obtained by multiplying the rescaled significant resistance  
366 surfaces described in Schwabi et al. [31]. The composite resistance surfaces were used for  
367 connectivity mapping and identifying corridors of movement via Circuitscape v. 5.10.2  
368 [34,89].

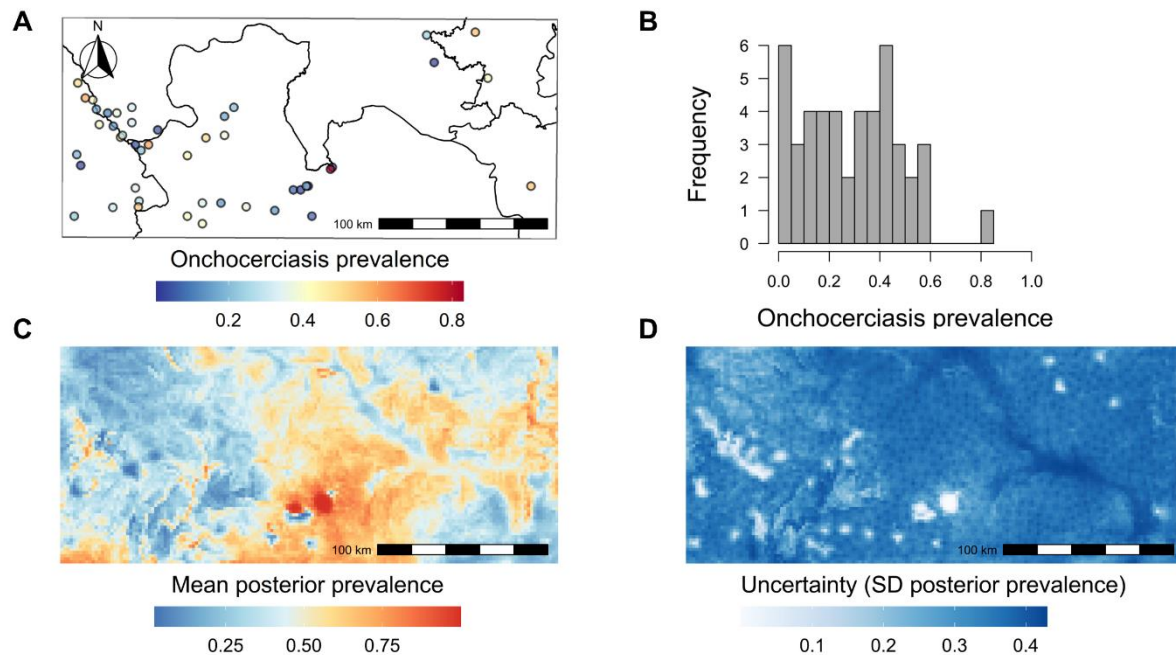
369 A bivariate map of posterior mean prevalence was plotted with composite resistance surface  
370 maps to visualise areas of varying prevalence and resistance. Correlation coefficients between  
371 the mean prevalence map and both the vector and parasite composite resistance surface maps  
372 were calculated. We also generated bivariate moving window correlation measures, their  
373 significance, and Moran's-I measure of spatial autocorrelation to measure the correlation  
374 between two spatial processes [95].



## 375 **Results**

### 376 **Prevalence mapping**

377 For the analysis of the prevalence data, the land surface temperature at night, temperature  
378 seasonality, minimum temperature of the coldest month, soil moisture, annual precipitation,  
379 slope, distance to the nearest river and prevalence of improved housing were selected. mf  
380 prevalence data ranged from 0.65% to 82.95% with a mean of 29.01% ( $\pm$  19.31% SD). Most  
381 of the data were from the western and south-central parts of the study area, with only five data  
382 points from the eastern parts (Figure 2a). The geostatistical interpolated map of baseline mf  
383 prevalence based on environmental data shows that the prevalence is higher, particularly in  
384 the south-central, central, and eastern areas of the transition Ghana (Figure 2c). The overall  
385 predicted prevalence is relatively low in the western areas of transition Ghana with scattered  
386 areas of high prevalence. As expected, the uncertainty map shows that the uncertainty was  
387 relatively lower in the actual sampling locations with varying levels of uncertainties in the  
388 interpolated areas (Figure 2d). Based on the regression coefficients, the soil moisture (mean  
389 coefficient: 0.043, 95% BCI: 0.004–0.084) and slope (mean coefficient: 2.126, 95% BCI:  
390 0.032–4.338) had a significant positive association with the mf prevalence while the  
391 temperature seasonality (mean coefficient: -0.022, 95% BCI: -0.044–0.001) had a significant  
392 negative association with the mf prevalence (Additional file Table S4). The spatial range of  
393 the mf prevalence map was estimated to be 4.4 km (95% BCI: 1.67–7.88 km).



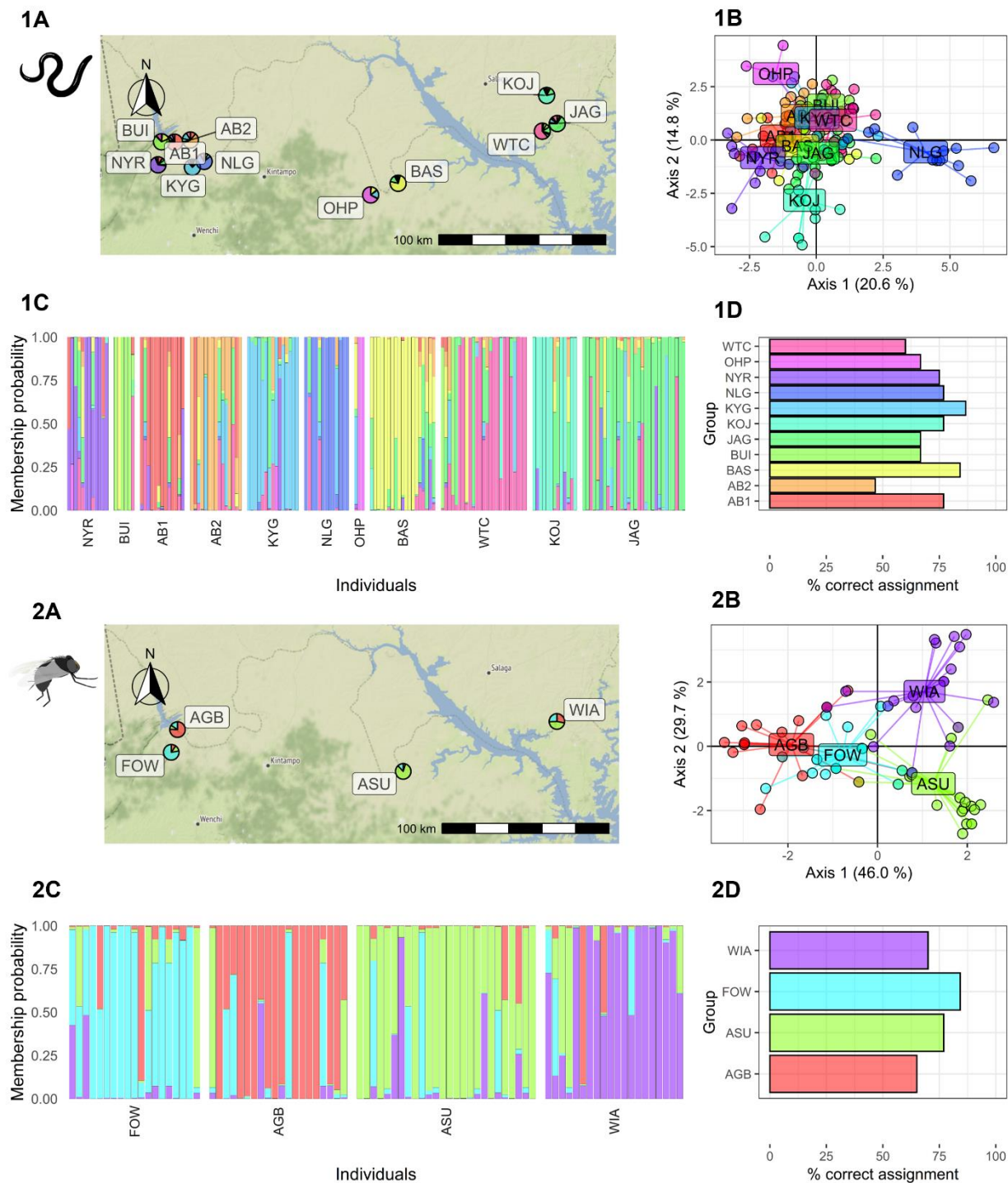
394

395 **Figure 2. Mapping baseline prevalence of *Onchocerca volvulus* infection in the transition region**  
396 **of Ghana.** Pre-MDAi point microfilarial prevalence data ( $n = 46$ ) (A), where circles represent  
397 sampling locations and the colours of the filled circles represent prevalence according to the heat bar  
398 below the figure. The solid line indicates the regional boundary. (B) shows the histogram of the pre-  
399 MDAi mf prevalence data. The model predicted estimate of the baseline prevalence of *O. volvulus*  
400 infection (C) in the transition region of Ghana and the uncertainty, i.e., the standard deviation (SD) of  
401 the posterior prevalence (D) is shown in the bottom row.

#### 402 **Population genetic analysis**

403 We carried out unsupervised  $k$ -means clustering analysis and visualised the haplotype  
404 network for both the parasite and the vector mitochondrial data separately to observe if there  
405 were any inherent clusters and if there were any outlier samples. We chose the minimum  
406 number of principal components that explained the highest cumulative variance. The number  
407 of principal components retained for the clustering analysis of the parasite and the vector was  
408 80 and 45, respectively. We chose the number of optimal clusters based on the BIC scores,  
409 i.e.,  $k = 8$  for the parasite data and  $k = 12$  for the vector data, as the decline in BIC saturated  
410 beyond these values (Additional file Figure S6). The clustering and haplotype network  
411 analysis on the *Simulium* data indicated the presence of outliers (groups 6 and 10; Additional  
412 file Figure S7) which were removed in the downstream analysis. For the parasite samples, the

413 number of alleles and the number of haplotypes corresponded to the sample size of the  
414 population, while the mean allelic richness and the gene diversity correlated with each other  
415 (Additional file Table S3). The number of principal components was optimised as 72 and 40,  
416 respectively. DAPC for the parasite genetic showed overlap between the clusters of the  
417 communities, except for a few communities like OHP and NLG (Figure 3). The average  
418 percentage of the correct assignment for parasites was 71.21% ( $\pm 11.45\%$  SD), which would  
419 generally be considered relatively poor. For vectors, DAPC also showed low overlap between  
420 clusters of the communities and an average % correct assignment of 74.03% ( $\pm 8.36\%$  SD).  
421 The mean percentage reassignment was not significantly different ( $p = 0.62$ ) between  
422 parasites and vectors, i.e., DAPC showed that the spatial distribution of parasite and vector  
423 genetic variation was similar.



424

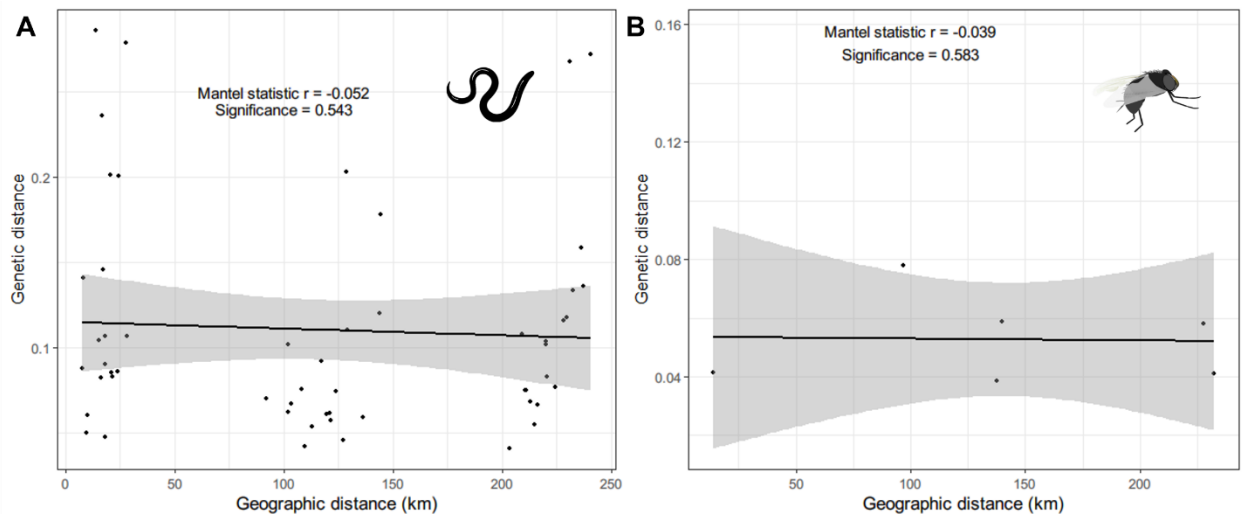
425 **Figure 3. Discriminant analysis of the principal components (DAPC) analysis for the parasite**  
 426 **and the vectors sampled from 11 and 4 communities respectively in the transition region of**  
 427 **Ghana.** The pie chart on the map (1A, 2A) indicates the community level of membership probability.  
 428 The DAPC analysis shows the community clusters (1B, 2B) and the individual level membership  
 429 probability (1C, 2C) with each block representing communities. The percentage of the samples

430 assigned correctly to their respective communities is shown for both the parasites (**1D**) and the vectors  
431 (**2D**). The community codes are presented in Additional file Table S3.

## 432 Landscape genetic analysis

### 433 Isolation-by-distance

434 The Euclidean distance matrix between sample locations and the matrix of linearised pairwise  
435  $F_{st}$  was used to test whether the parasites and vector population structure conformed to an  
436 isolation-by-distance model, in which the degree of genetic differentiation is correlated  
437 positively with geographic distance between sampling locations [84]. The Euclidean  
438 geographic distance between locations ranged from 2.2 km to 240.39 km. For the parasite  
439 sampling locations, six communities were less than 5km apart and were merged into two  
440 communities. The geographic distance for the parasites averaged 117.73 km ( $\pm 11.50$  SE;  
441 range: 7.86–240.43 km), and the genetic distance averaged 0.11 ( $\pm 0.009$  SE; range: 0.041–  
442 0.286). Similarly, for the vectors, the geographic distance for the parasites averaged 141.40  
443 km ( $\pm 33.61$  SE), and the genetic distance averaged 0.056 ( $\pm 0.007$  SE; range: 0.04–0.084). The  
444 Mantel test indicated a poor correlation between the genetic distance and the geographic  
445 distance for both the parasite (Mantel's  $r = -0.052$ ;  $p = 0.543$ ) and the vector data (Mantel's  $r$   
446  $= -0.039$ ;  $p = 0.583$ ) (Figure 4).



447

448 **Figure 4. The relationship between the genetic (linearised  $F_{st}$ ) and the Euclidean geographic**  
449 **distances.** Isolation-by-distance was tested by the Mantel test, and the significance and the strength of  
450 the relationship are shown for the parasite (**A**) and vector (**B**).

## 451 **Resistance surface optimisation and testing**

452 We selected five environmental variables for the resistance surface optimisation: elevation,  
453 isothermality, soil moisture, flow accumulation and annual precipitation. The values in the  
454 resistance surface represent the amount by which the movement is restrained by the given  
455 environmental variables. The ecological cost distances obtained for the respective resistance  
456 surfaces were used to determine whether the environmental variables could explain the genetic  
457 differentiation among parasite and vector sampling locations and performed four replicates of  
458 optimisation for 1000 iterations each, then chose the surface with the highest significance (i.e.,  
459 lowest p-value). For the parasites, we found that the inverse ricker transformation for elevation  
460 ( $r = 0.793$ ,  $p = 0.005$ ) and soil moisture ( $r = 0.507$ ,  $\beta = 0.002$ ,  $p = 0.022$ ) were significant  
461 (Table 1). The inverse reverse monomolecular transformations for elevation soil moisture  
462 were also significant, but the levels of significance were lower compared to the chosen  
463 resistance surfaces. Therefore, inverse ricker transformation surfaces for the elevation and soil  
464 moisture were used for the preparation of the composite resistance surface map for the parasite  
465 data.

466 The inverse ricker transformation was significant in both environmental layers with high  
467 resistance to gene flow in the low and high environmental values and lower resistance in the  
468 moderate range of environmental values, but with different scale parameters. The resistance  
469 to gene flow was lowest (< 30% of the total resistance) in areas with an elevation range of 90–  
470 150 m and in areas with soil moisture of 60–190 mm (Figure 5). A composite resistance  
471 surface map was prepared, which showed high resistance around the western parts of the study  
472 area, which are characterised by low soil moisture (i.e., Bui National Park in the west, a  
473 woodland Savannah zone [96]) and higher elevation. The areas around Lake Volta also have  
474 high resistance. Accordingly, the movement corridor map suggests that there is relatively  
475 lower connectivity of parasites in the northwestern part of the study area (Figure 6). The  
476 central parts of the study area are characterised by high connectivity, showing a potential route  
477 for the movement/transmission of parasites.

478 For the vector genetic data, resistance surfaces obtained from soil moisture ( $r = 0.788$ ,  $p =$   
479  $0.0417$ ) and precipitation ( $r = 0.835$ ,  $p = 0.0417$ ) were significant, with inverse reverse  
480 monomolecular and inverse ricker transformations, respectively. The lowest resistance (< 30%  
481 of the maximum resistance) for vector gene flow was in the areas with soil moisture of 22–

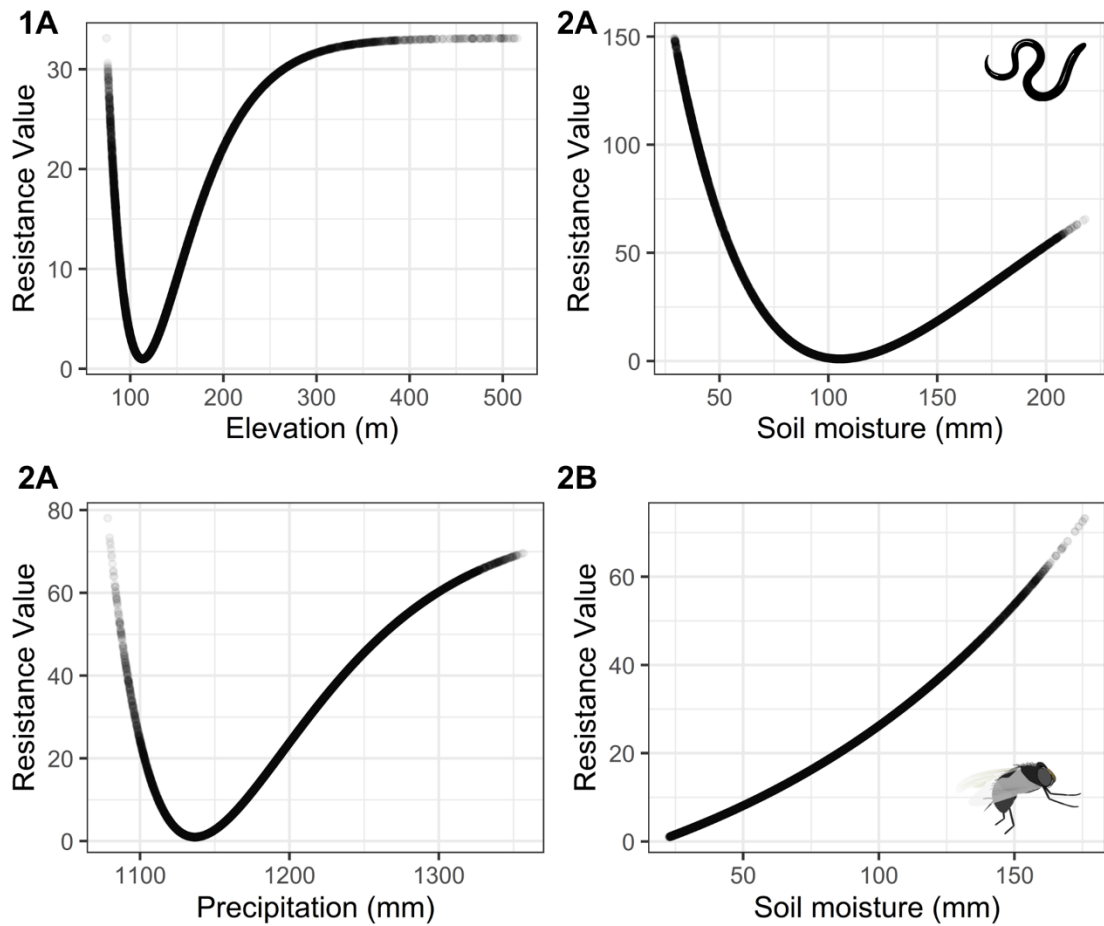
482 90 mm and precipitation of 110–120 cm. These two resistance surfaces were rescaled and  
483 merged to create a composite resistance surface as performed on the parasite data. The  
484 composite resistance surface for the vectors revealed that there was particularly low resistance  
485 for gene flow along the western and northwestern areas of the study area and a moderate level  
486 of resistance in the central region. The current density map also showed a higher level of  
487 connectivity (lower resistance) around the southwestern Savannah region (Figure 6).



**Table 1. 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 are 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.  $\beta_{geo}$  and  $\beta_{resist}$  represents the regression coefficients for the geographic distance and the cost distance due to the resistance surface respectively.

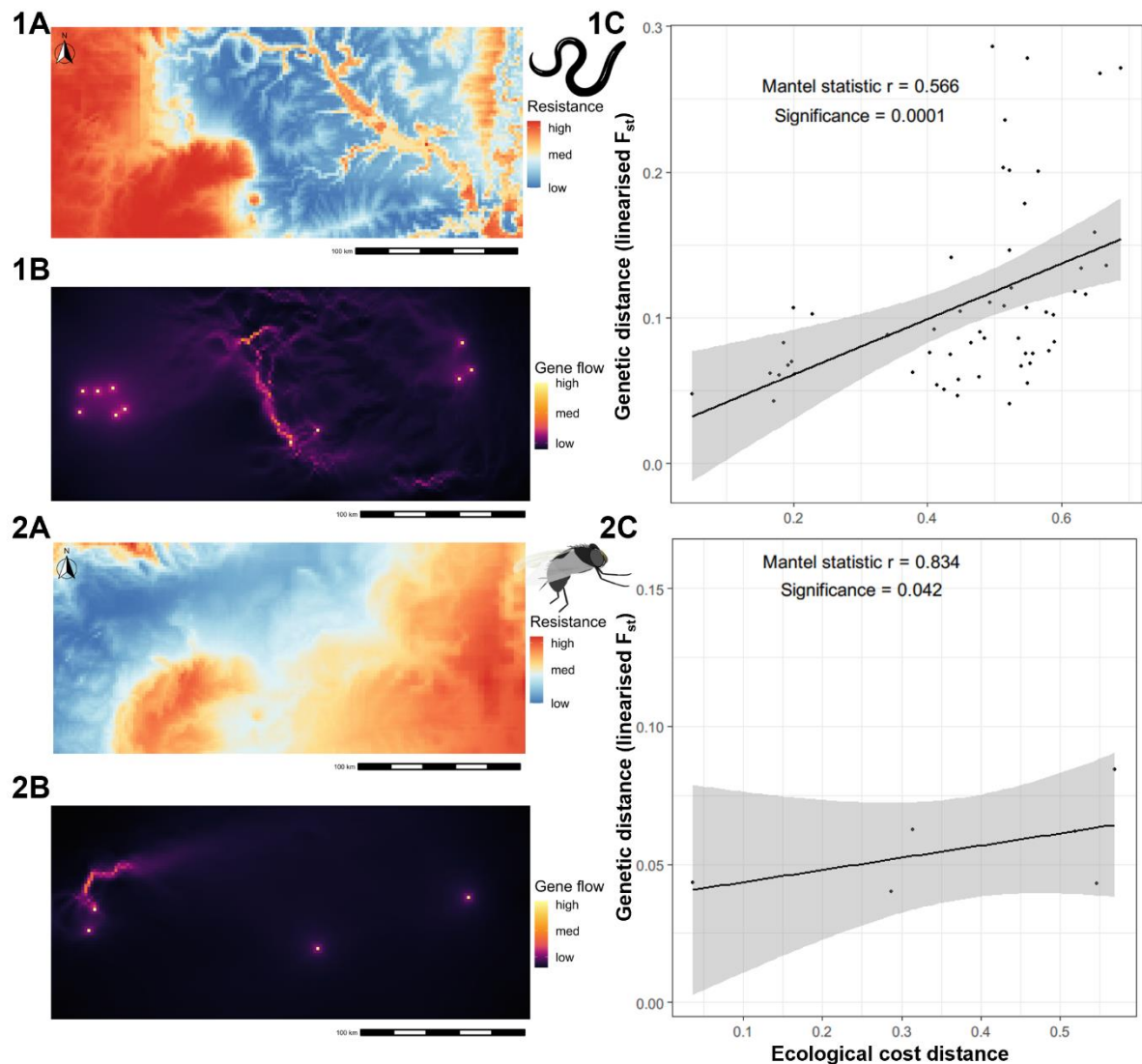
Organism	Covariates	# repli cates	Optimisation parameter for resistance surfaces			Genetic distance ~ resistance distance + geographic distance					
			Equation	Shape	Max	partial Mantel		MMRR			
						r	p	$\beta_{geo}$	p	$\beta_{resist}$	p
<b>Parasites</b> <i>(O. volvulus)</i>	<b>Elevation</b>	2	<b>Inverse Ricker</b>	<b>0.873</b>	<b>100.000</b>	<b>0.793</b>	<b>0.0002***</b>	<b>-0.00038</b>	<b>0.008*</b>	<b>0.022</b>	<b>0.0046*</b>
			Inverse-Reverse								
	Isothermality	3	Monomolecular	5.046	99.996	0.745	0.0002***	-0.00084	0.009*	0.046	0.0074*
			Inverse-Reverse Ricker	3.439	99.996	0.391	0.0640	-0.00035	0.131	0.004	0.2242
	<b>Soil moisture</b>	2	Ricker	0.936	99.999	0.337	0.1324	-0.00029	0.140	0.007	0.2748
			<b>Inverse Ricker</b>	<b>4.031</b>	<b>99.997</b>	<b>0.507</b>	<b>0.0002***</b>	<b>-0.00017</b>	<b>0.264</b>	<b>0.002</b>	<b>0.022*</b>
			Inverse Monomolecular	0.500	99.922	0.489	0.0135*	-0.00004	0.742	0.003	0.022*
			Inverse Ricker	5.000	99.976	0.439	0.1155	-0.00012	0.424	0.007	0.1364
<b>Vectors</b> <i>(S damnosum)</i>	Elevation	3	Inverse Monomolecular	0.500	99.835	0.804	0.0833	-0.00015	0.323	0.003	0.1229
			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
			<b>Inverse-Reverse Monomolecular</b>	<b>7.723</b>	<b>100.000</b>	<b>0.788</b>	<b>0.0417*</b>	<b>-0.00016</b>	<b>0.202</b>	<b>0.002</b>	<b>0.042*</b>
	Flow accumulation	3	Inverse Ricker	3.570	99.964	0.569	0.1250	-0.00019	0.250	0.001	0.2503
			Ricker	0.500	100.000	0.678	0.0833	-0.00020	0.334	0.039	0.3721
	<b>Precipitation</b>	<b>4</b>	<b>Inverse Ricker</b>	<b>2.096</b>	<b>99.984</b>	<b>0.835</b>	<b>0.0417*</b>	<b>-0.00018</b>	<b>0.161</b>	<b>0.002</b>	<b>0.0418*</b>

\*:  $p < 0.05$ , \*\*:  $p < 0.005$ , \*\*\*  $p < 0.0005$



1

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

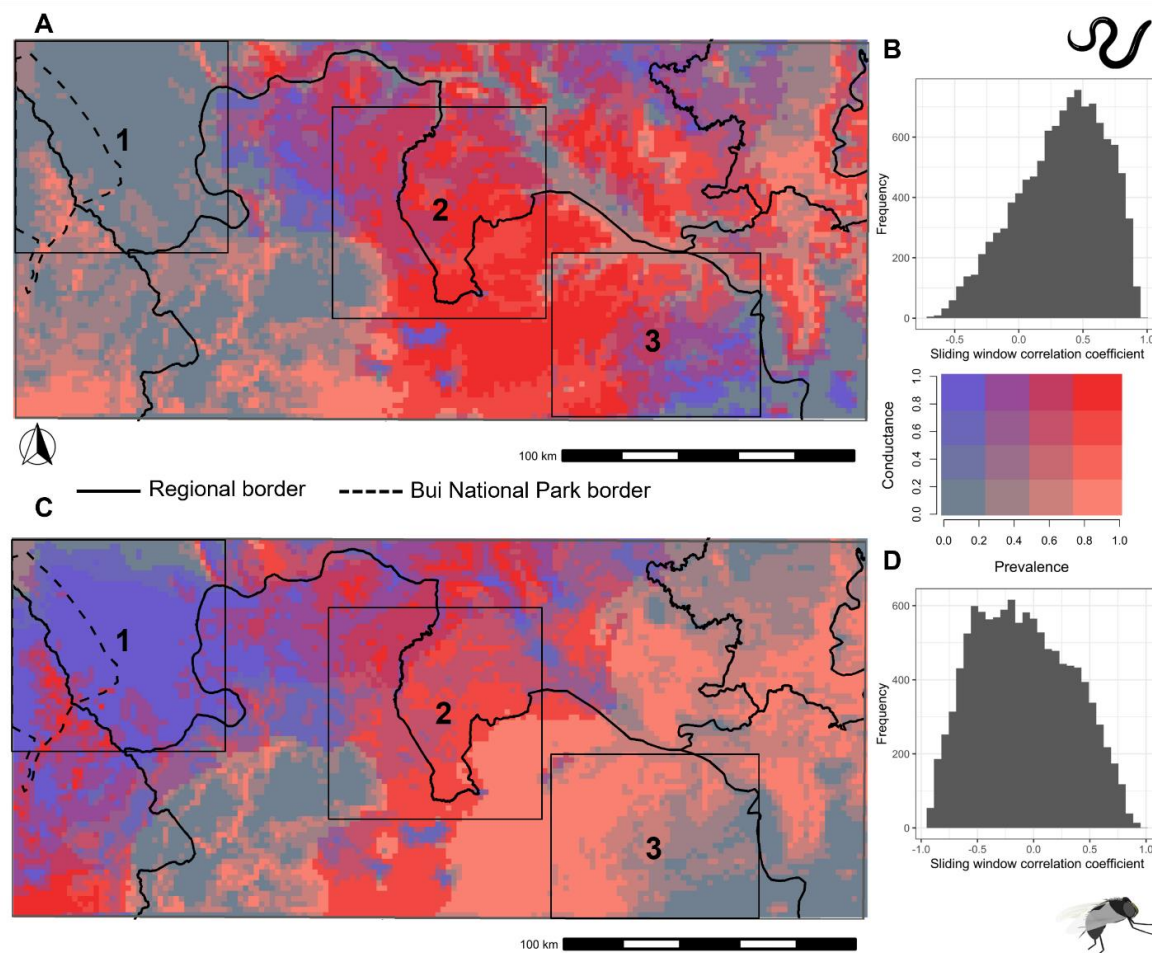


5

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

13 The bivariate map (Figure 7) obtained by combining mf prevalence map and the conductance  
14 surface (inverse of resistance surface maps where high conductance implies high suitability  
15 for movement) for the parasite shows that the area of high parasite conductance and high  
16 prevalence is in the central parts of the transition region of Ghana (Figure 7, Box 2). There is

17 a good correlation between the parasite's composite conductance surface and the *O. volvulus*  
18 infection prevalence map with the majority (57.34%) of sliding window correlation  
19 coefficients greater than 0.3 (Figure 7B). Therefore, the areas with high parasite conductance  
20 are also the areas of high *O. volvulus* infection prevalence and vice versa. Areas of high vector  
21 conductance and high prevalence are found in the central and southwestern parts of the study  
22 area. However, a substantial portion of the vector bivariate map has high conductance but low  
23 prevalence, particularly around the northwestern region of the study area (Figure 7, Box 1).  
24 As a result, the correlation between the conductance map for vectors and the mf infection  
25 prevalence is not as strong as the correlation for the parasite counterpart. Only 21.24% of the  
26 sliding window correlation coefficients are greater than 0.3 (Figure 7D). There are also the  
27 areas in the south-eastern parts of the area that have high prevalence and high parasite  
28 conductance; however, low vector conductance (Figure 7, Box 3).



29

30 **Figure 7. A bivariate map created using composite conductance surfaces and the *Onchocerca***  
31 ***volvulus* infection prevalence map.** The top row shows the bivariate map for the parasite (A) and the  
32 bottom row (C) for the vector. The legend for the bivariate map is shown on the right, where red colour  
33 indicates the areas with high prevalence and high conductance (represents high movement suitability),  
34 whereas blue colour indicates areas with high conductance but low prevalence. The histogram on the  
35 right of the respective map shows the frequency of the sliding window correlation coefficient between  
36 the conductance surface and the prevalence map for the *O. volvulus* infection prevalence map with the  
37 parasite (B) and the vector (D) conductance surface. The solid line represents the regional  
38 administrative border, while the broken line shows the border for Bui national park. The three boxes  
39 in the figure show contrasting patterns of conductance and prevalence: 1. High vector conductance but  
40 low parasite conductance and low *O. volvulus* infection prevalence; 2. High vector and parasite  
41 conductance and high *O. volvulus* infection prevalence; 3. Low vector conductance but high parasite  
42 conductance and high *O. volvulus* infection prevalence. The conductance and the prevalence on the  
43 map are rescaled from 0 to 1.

## 44 Discussion

45 For the first time in the context of onchocerciasis, we have integrated point location prevalence  
46 data, the population genetics of parasites and vectors (as a proxy for parasite and vector  
47 movement), and environmental data within a single landscape genetics framework. The visual,  
48 spatial representation of parasite and vector movement and infection prevalence shown in  
49 Figure 7 is a spatial representation of *O. volvulus* transmission and brings us a step closer to a  
50 quantitative, evidence-based method for "delineating" onchocerciasis transmission zones. We  
51 have transformed the metrics of genetic connectivity and landscape/ecological variables into  
52 a resistance/conductance surface, i.e., a spatial prediction of vector movement and parasite  
53 transmission suitability (high resistance or low conductance represents low suitability for  
54 movement and transmission and vice versa) which provides an evidence-based methodology  
55 by which it may be possible to define transmission zones. For example, geospatially explicit  
56 modelling of prevalence and landscape connectivity—can be used to identify reasons for  
57 ongoing transmission despite MDAi or newly arisen hot spots of transmission post-MDAi.

58 Just as the pre-MDAi prevalence is the product of the cumulative history of *O. volvulus*  
59 infection, so is the population genetic structure the product of events in the past. Both of these  
60 historical elements do not reflect the current transmission patterns of *O. volvulus*. However,  
61 using ecological data might enable us to better estimate current transmission as ecological and  
62 landscape data are 'current'. The timeframe over which the climate changes is long compared  
63 to the timeframe over which prevalence and population structure changes. Assuming that the  
64 ecological parameters are unlikely to have changed significantly over timeframes of either the  
65 cumulative infection history giving rise to the prevalence or the microevolutionary processes  
66 giving rise to the current population genetic structure, identifying environmental features  
67 associated with population genetics and the prevalence allows us to understand current and  
68 predict future transmission patterns.

69 For the ecological transition region of Ghana, the pre-MDAi infection prevalence was  
70 positively associated with slope and soil moisture. A likely explanation for this correlation is  
71 that greater topological slope results in faster river flow essential for vector breeding.  
72 Similarly, soil moisture was also identified to be significant in an analysis of Ethiopian  
73 *O. volvulus* nodule prevalence data, where areas with high soil moisture occur in arable land



74 that are usually inhabited by people who are exposed more to vector bites [2,64,97]. In  
75 contrast, temperature seasonality was negatively associated with mf prevalence (Additional  
76 file Table S4). This is likely because areas with higher fluctuations in temperature might not  
77 be favourable for *Simulium*. After all, different species of *Simulium* have different temperature  
78 ranges for breeding and biting activities [66], and activities of blood-seeking flies are limited,  
79 particularly in low temperatures [98]. Further, the significant relationship between mf  
80 prevalence to the temperature seasonality highlights the potential effect of global warming and  
81 alterations in annual temperature patterns on the distribution of onchocerciasis. We were not  
82 able to detect a significant association of the mf prevalence with the distance to the nearest  
83 river, which might be because all the communities surveyed happened to be close to rivers (<  
84 10 km). Therefore, the spatial coverage of the samples might influence the inferred  
85 relationship of the ecological variables with the prevalence and the genetic data.

86 The parasites themselves do not move, however, their movement between geographical  
87 locations is mediated either by infected blackflies or infected humans. Population genetics is  
88 able to provide insights into the migration of the parasites and the blackflies. The population  
89 genetic analyses of parasite and vector genetic data in the ecological transition region of Ghana  
90 were largely concordant: both parasite and vector showed low genetic differentiation or high  
91 genetic similarity between the sampled communities. Previous studies by Crawford et al. [25]  
92 and Gyan [99], suggested the same, i.e., both the parasite and the vector populations were  
93 largely genetically homogeneous. Consequently, there was no support for an isolation-by-  
94 distance population structure for either the parasites or their vectors in the ecological transition  
95 region of Ghana. This suggests that the gene flow of the parasite and the vector populations  
96 were not restricted by geographic distance in this study area. However, some degree of genetic  
97 differentiation between sampling locations was observed. In order to investigate the likely  
98 origins of this relatively weak population structure, we estimated an "ecological distance"  
99 parameter from local ecological data for each community, and observed a strong positive  
100 correlation. Thus, if "ecological distance" is substituted for "geographical distance" in the  
101 isolation-by-distance model, these data do show isolation-by-distance relationships driven by  
102 ecological rather than geographical proximity.

103 With the assumption that environmental factors could explain the resulting observed vector  
104 and the parasite genetic connectivity, we used a landscape genetics framework to (1) identify



105 the ecological factors influencing *S. damnosum* and *O. volvulus* population structure then (2)  
106 combine the resultant spatial correlation between inferred parasite/vector movement and  
107 ecology with the predicted spatial pattern of prevalence to produce an integrated map of likely  
108 transmission intensity. Landscape genetics methods combine ecological connectivity with  
109 genetic similarity. This allows us to identify the corridors of movement and, thus, the spatially  
110 explicit patterns of transmission. It is important to note that high vector connectivity might not  
111 necessarily mean high movement suitability, high vector density or high vector biting rates.  
112 These are the observed suitability for the movement of blackflies based on the genetic data.  
113 High biting rates are crucial for the high endemicity of the disease, whereas vector mobility  
114 might help maintain or even amplify onchocerciasis endemicity. Here, we assume that if the  
115 vector has high mobility in the areas of high prevalence, there is a likely possibility of high  
116 transmission events.

117 For the parasite population, resistance surfaces obtained from the elevation and soil moisture  
118 were significantly associated with the genetic distance. The resistance to parasite gene flow  
119 was low (i.e., genetic connectivity was high) in the areas of moderate elevation in the range  
120 of 90–150 m and in areas with moderate soil moisture, 60–190 mm. Our estimate of the range  
121 of elevation most strongly correlated with prevalence is essentially identical to the range  
122 reported proposed by Barro and Oyana [65]. The reason behind high resistance to the parasite  
123 gene flow in the areas of low soil moisture could be due to the un-arability of the land and,  
124 thus, the lack of human hosts. Soil moisture is reported to be an important environmental  
125 feature influencing the occurrence of onchocerciasis in other studies [2,67]. However, high  
126 soil moisture areas might also not be that suitable for onchocerciasis as those were around  
127 Lake Volta with non-flowing water and are generally unsuitable for vector breeding. Lake  
128 Volta is one of the biggest artificial lakes in the world. Lakes formed by river dams have been  
129 reported to affect vector breeding and decrease *O. volvulus* transmission [100–102].

130 Parasite connectivity indicates where parasite transmission can occur between locations.  
131 Blackfly connectivity, in contrast, indicates where transmission may occur between locations  
132 due to blackfly movement rather than, or in addition to, human movement. Therefore,  
133 differences in the blackfly resistance surface profile compared to the resistance surface for  
134 parasites represent the potential transmission mediated by human movement (Figure 6).  
135 Further, the blackfly resistance surface was not as strongly correlated as the parasite resistance

136 surface to the mf prevalence map, particularly in the western parts of the study region (Figure  
137 7, Box 1). There are several factors that may contribute to low concordance between blackfly  
138 and parasite resistance surfaces. One is the pattern of human population density. For example,  
139 the vector connectivity was high in the areas with low soil moisture, while parasite  
140 connectivity was low. Low soil moisture indicates lower suitability for agriculture, and they  
141 likely have lower human population density and thus appear unsuitable for parasite  
142 transmission. A similar case is Bui National Park in the west, where blackflies are present but  
143 there is a sparse human settlement and hence low parasite transmission. A second factor is the  
144 ratio of *O. volvulus* to *O. ochengi* (and potentially other *Onchocerca* species) in the blackflies.  
145 Doyle et al. [103] showed that the proportion of *O. volvulus* larvae in blackflies was lower in  
146 western communities compared to the communities in the central and eastern parts of the  
147 ecological transition region. The presence of a higher proportion *O. ochengi* has been proposed  
148 to impact the vectorial capacity for the *O. volvulus* due to the saturation of the vectors with  
149 *O. ochengi* [104,105].

150 The weak population structure observed across communities is consistent with the absence of  
151 isolation-by-distance observed (Figure 4). The strong correlation between gene flow and  
152 several ecological factors related to habitat suitability for black flies indicates that "ecological  
153 distance" explains the population genetic structure (Figure 6); i.e., there is a strong correlation  
154 between gene flow (genetic differentiation) and ecological connectivity. This strong  
155 relationship leads to two important conclusions. First, it provides an explanation for the strong  
156 correlation between gene flow and ecological parameters related to blackfly habitat. Second,  
157 it suggests a model in which blackfly connectivity is related to the degree to which "local"  
158 blackfly populations around discrete breeding sites overlap. What is perhaps surprising is that  
159 this proposed overlap between breeding sites extends to create continuous ecological corridors  
160 for blackfly movement and parasite transmission.

161 We produced a bivariate fusion map that combined the results of the mf prevalence and  
162 resistance surface mapping (Figure 7). The sliding window correlation coefficient between the  
163 surfaces showed a close overlap of the mf prevalence map with the parasite resistance surface,  
164 which further validates the landscape genetics output. The bivariate maps represent three  
165 different scenarios. Within box 1, there is a high suitability for vector mobility but low  
166 infection prevalence and low suitability for parasite mobility. Within box 2, the predicted

167 vector mobility seems to correlate well with parasite mobility and prevalence. In box 3, there  
168 is an apparent discordance between the parasite and the vector mobility. The high parasite  
169 mobility suggests that the spatial pattern of transmission is likely to be driven more by human  
170 movement than vector movement. Therefore, bivariate maps could help in drawing  
171 conclusions about what drives transmission in different epidemiological contexts.

172 Inferences like these might be vital in making spatially explicit onchocerciasis elimination  
173 decisions. For example, in the current study, we can hypothesise that communities in the  
174 central parts of the study areas (box 2) are one of the critical connecting areas with high  
175 suitability for the parasite and the vector gene flow and high onchocerciasis prevalence. The  
176 connectivity analysis using the composite resistance surface maps derived from the significant  
177 resistance surfaces for the parasites showed that the parasite gene flow was high in the central  
178 parts of the ecological transition region of Ghana, around communities from the Bono East  
179 (Figure 6). Therefore, MDAi alone might not be sufficient to eliminate onchocerciasis  
180 transmission in these areas, where alternative treatment strategies with vector control have to  
181 be implemented. However, in areas within box 3, where there is high infection prevalence due  
182 to high parasite mobility but low vector mobility, vector control might not be as effective as  
183 in the areas within box 2.

184 Other studies confirm that the communities within box 2 are characterised particularly by high  
185 biting rates, high vector density and high vector mobility [5,106] and were among the first to  
186 be targeted for both the vector control initially and MDAi later. In addition, this is the area  
187 where SOR against ivermectin was first reported [39,40]. Therefore, with the reports of SOR  
188 and the evidence of high gene flow from these areas, the possibility of spreading the SOR  
189 strains cannot be ignored. One can expect the consequences of SOR to be spread over an  
190 extensive geographical range as a result of the high gene flow of the parasites and the vectors.  
191 The approach outlined here might provide an indication of where different epidemiologically  
192 relevant phenotypes might likely spread and help design interventions accordingly.

193 Eliminating onchocerciasis transmission in areas of high connectivity might facilitate  
194 onchocerciasis elimination in surrounding areas of lower connectivity. However, it is not to  
195 say that the other areas might not act as the source of infection, particularly if the infection is  
196 well controlled in the high connectivity region. For example, recent modelling work suggests

197 that low endemic areas can act as a source to re-initiate transmission in MDAl-controlled  
198 onchocerciasis endemic areas [107,108]. Nevertheless, resistance surfaces and connectivity  
199 maps could be used to develop heterogeneous intervention strategies to address spatially  
200 heterogeneous transmission. Specifically, interventions should be coordinated across locations  
201 that are shown to be connected. The intensity of intervention should be varied according to  
202 connectivity so that locations of high connectivity receive more intensive interventions than  
203 regions of lower connectivity. The rationale is that transmission will be suppressed in a more  
204 coordinated fashion with less risk of hotspots of residual transmission even though initial  
205 prevalence and transmission may have been highly heterogeneous.

206 There are some caveats to the current study. First, the sampling density and spatial coverage  
207 of the samples in this study are low, and increasing sampling density, in particular, would  
208 increase the accuracy of the estimated resistance surfaces. Future landscape genetic studies  
209 should consider dense and stratified uniform sampling across space and environmental  
210 gradients [29,109]. Second, due to the unavailability of the nuclear genome sequence data, the  
211 genetic analyses utilised mitochondrial sequence data, which might underestimate gene flow  
212 [27], and we recommend using nuclear data in future landscape genetics studies. Nevertheless,  
213 this study serves as an important use case of the approach with the best data available. Third,  
214 the vector resistance surface maps we obtain with the current approach might not necessarily  
215 correspond with vector density or vector biting rates. Therefore, incorporating vector  
216 abundance data and annual biting rates might further enrich the insights from the approach.  
217 Nevertheless, this could be a powerful approach to spatially transforming population genetic  
218 connectivity estimates, accounting for ecological variables and predicting routes and  
219 geographical boundaries of transmission. Applying this approach to other geographic regions  
220 (such as persistent hotspots, cross-border transmission settings and others), and also to other  
221 filarial diseases, such as lymphatic filariasis, might prove valuable to the elimination endgame.

## 222 Conclusion

223 To meet onchocerciasis elimination goals, it is necessary to identify the areas that require  
224 intervention via "elimination mapping" (extending prevalence mapping to currently unmapped  
225 areas of unknown but probably low prevalence) and by better understanding the spatial  
226 patterns of transmission (delineation of transmission zones). We have shown previously how  
227 incomplete point prevalence data can be combined with ecological data to provide accurate,  
228 spatially continuous, predictions of prevalence [2]. Here we extend that work to provide a  
229 novel and promising approach to combine ecological parameters related to vector habitat with  
230 population genetic estimates of the vector and the parasite gene flow to produce spatial maps  
231 of movement suitability that identify the corridors of movement and give us insight into  
232 *O. volvulus* transmission. We demonstrated that the entire ecological transition zone was  
233 connected by corridors that are ecologically suitable for vector movement and hence parasite  
234 transmission. This leads to the conclusion that the entire ecological transition zone through  
235 which the Volta River flows should be treated as a single *O. volvulus* transmission zone. We  
236 conclude further that the persistence of transmission across this region, particularly in  
237 communities located in the central part of the region, is in part due to the high degree of  
238 transmission connectivity over large geographic distances via the "connectivity corridors" we  
239 have identified. The spatial pattern of transmission we describe suggests that interventions to  
240 interrupt transmission of *O. volvulus* in central Ghana must be coordinated over a large  
241 geographical area, particularly decisions to stop MDAi in communities in which local  
242 transmission may have been interrupted but which will be subject to re-invasion from  
243 surrounding areas in which transmission is yet to be suppressed. We also suggest that  
244 landscape genetics could be applied to other vector-borne diseases, particularly lymphatic  
245 filariasis, where instances of recrudescence following stop-MDA decisions are accumulating.

## 246 **Availability of data and materials**

247 The parasite sequence data are available at NCBI (<https://www.ncbi.nlm.nih.gov/> Accession  
248 #: PRJNA560089), and the blackfly sequence data have been uploaded to EMBL-EBI  
249 (<https://www.ebi.ac.uk/> Accession #: PRJEB57094). The onchocerciasis prevalence data were  
250 obtained from the ESPEN data portal ([https://espen.afro.who.int/tools-resources/download-](https://espen.afro.who.int/tools-resources/download-data)  
251 [data](https://espen.afro.who.int/tools-resources/download-data)), and the sources for the environmental data are provided in the supplementary  
252 information. The scripts for the analysis pipeline are uploaded to the GitHub repository  
253 ([https://github.com/himal2007/landscape\\_genetics\\_ghana](https://github.com/himal2007/landscape_genetics_ghana)).

## 254 **Abbreviations**

255 BCI: Bayesian credible interval; BIC: Bayesian information criteria; DAPC: Discriminant  
256 analysis of principal components; DNA: Deoxyribonucleic Acid; EBI: European  
257 Bioinformatics Institute; EMBL: European Molecular Biology Laboratory; ENA: European  
258 Nucleotide Archive; ESPEN: Expanded Special Project for Elimination of Neglected Tropical  
259 Disease; INLA: Integrated nested Laplace approximations; MDAl: Mass drug administration  
260 with ivermectin; MLPE: Maximum likelihood population effects; MMRR: Mixed matrix  
261 regression with randomisation; NCBI: National centre for biotechnology information; OCP:  
262 Onchocerciasis Control Programme; PCA: Principal component analysis; SD: Standard  
263 deviation; SE: Standard error; SNP: Single nucleotide polymorphism; SOR: Sub-optimal  
264 response

## 265 **Acknowledgements**

266 The authors wish to acknowledge Anusha Kode for assistance with the sequencing experiment  
267 and Dr Kwadwo Frempong for discussion with the results of the prevalence mapping.

## 268 **Funding**

269 This work was supported by funding from UNICEF/UNDP/World Bank/WHO Special  
270 Programme for Research and Training in Tropical Diseases (TDR) to SMH (P21-00481). HS  
271 was supported by Australian Government Research Training Program Scholarship, a La Trobe  
272 Graduate Research Scholarship and a La Trobe University Full-Fee Research Scholarship.

273 **Contributions**

274 Conceptualisation: Himal Shrestha, Shannon M. Hedtke, Warwick N. Grant

275 Data Curation: Himal Shrestha

276 Formal Analysis: Himal Shrestha

277 Funding Acquisition: Shannon M. Hedtke, Warwick N. Grant

278 Investigation: Neha Sirwani, Katie E Crawford, Samuel Armoo, Francis Vierigh

279 Methodology: Himal Shrestha, Shannon M. Hedtke, Karen McCulloch, Rebecca Chisholm

280 Project Administration: Warwick N. Grant, Shannon M. Hedtke

281 Resources: Samuel Armoo, Francis Vierigh

282 Supervision: Shannon M. Hedtke, Warwick N. Grant, Rebecca Chisholm, Mike Osei-

283 Atweneboana

284 Validation: Himal Shrestha, Karen McCulloch, Rebecca Chisholm, Warwick N. Grant,

285 Shannon M. Hedtke

286 Visualisation: Himal Shrestha

287 Writing – Original Draft Preparation: Himal Shrestha

288 Writing – Review & Editing: Himal Shrestha, Shannon M. Hedtke, Warwick N. Grant, Karen

289 McCulloch, Samuel Armoo, Rebecca Chisholm, Katie Crawford



## 290 **References**

- 291 1. Ngoumou P, Walsh JF, Mace JM. A rapid mapping technique for the prevalence and  
292 distribution of onchocerciasis: a Cameroon case study. *Ann Trop Med Parasitol*. 1994;88:463–  
293 74.
- 294 2. Shrestha H, McCulloch K, Hedtke SM, Grant WN. Geospatial modeling of pre-intervention  
295 nodule prevalence of *Onchocerca volvulus* in Ethiopia as an aid to onchocerciasis elimination.  
296 Cotton J, editor. *PLoS Negl Trop Dis*. 2022;16:e0010620.
- 297 3. Vieira JC, Brackenboro L, Porter CH, Basáñez M-G, Collins RC. Spatial and temporal  
298 variation in biting rates and parasite transmission potentials of onchocerciasis vectors in  
299 Ecuador. *Trans R Soc Trop Med Hyg*. 2005;99:178–95.
- 300 4. Zouré HG, Noma M, Tekle AH, Amazigo UV, Diggle PJ, Giorgi E, et al. The geographic  
301 distribution of onchocerciasis in the 20 participating countries of the African Programme for  
302 Onchocerciasis Control: (2) pre-control endemicity levels and estimated number infected.  
303 *Parasit Vectors*. 2014;7:326.
- 304 5. Lamberton PH, Cheke RA, Walker M, Winskill P, Osei-Atweneboana MY, Tirados I, et al.  
305 Onchocerciasis transmission in Ghana: biting and parous rates of host-seeking sibling species  
306 of the *Simulium damnosum* complex. *Parasit Vectors*. 2014;7:511.
- 307 6. Wanji S, Kengne-Ouafo JA, Esum ME, Chounna PWN, Adzemye BF, Eyong JEE, et al.  
308 Relationship between oral declaration on adherence to ivermectin treatment and  
309 parasitological indicators of onchocerciasis in an area of persistent transmission despite a  
310 decade of mass drug administration in Cameroon. *Parasit Vectors*. 2015;8:667.
- 311 7. Ekpo UF, Eneanya OA, Nwankwo EN, Soneye IY, Weil GJ, Fischer PU, et al. Persistence  
312 of onchocerciasis in villages in Enugu and Ogun states in Nigeria following many rounds of  
313 mass distribution of ivermectin. *BMC Infect Dis*. 2022;22:832.
- 314 8. Koala L, Nikiema A, Post RJ, Paré AB, Kafando CM, Drabo F, et al. Recrudescence of  
315 onchocerciasis in the Comoé valley in Southwest Burkina Faso. *Acta Trop*. 2017;166:96–105.
- 316 9. Koala L, Nikièma AS, Paré AB, Drabo F, Toé LD, Belem AMG, et al. Entomological  
317 assessment of the transmission following recrudescence of onchocerciasis in the Comoé  
318 Valley, Burkina Faso. *Parasit Vectors*. 2019;12:34.
- 319 10. Nikièma AS, Koala L, Post RJ, Paré AB, Kafando CM, Drabo F, et al. Onchocerciasis  
320 prevalence, human migration and risks for onchocerciasis elimination in the Upper Mouhoun,  
321 Nakambé and Nazinon river basins in Burkina Faso. *Acta Trop*. 2018;185:176–82.
- 322 11. Cheke RA, Garms R. Reinfestations of the southeastern flank of the Onchocerciasis  
323 Control Programme area by windborne vectors. *Philos Trans R Soc Lond B Biol Sci*.  
324 1983;302:471–84.

- 325 12. Cupp EW, Sauerbrey M, Richards F. Elimination of human onchocerciasis: History of  
326 progress and current feasibility using ivermectin (Mectizan®) monotherapy. *Acta Trop.*  
327 2011;120:S100–8.
- 328 13. Le Berre R, Garms R, Davies JB, Walsh JF, Philippon B, Johnson CG, et al. Displacements  
329 of *Simulium damnosum* and Strategy of Control Against Onchocerciasis [and Discussion].  
330 *Philos Trans R Soc Lond B Biol Sci. Royal Society*; 1979;287:277–88.
- 331 14. Service MW. Effects of wind on the behaviour and distribution of mosquitoes and  
332 blackflies. *Int J Biometeorol.* 1980;24:347–53.
- 333 15. APOC, WHO. Conceptual and operational framework of onchocerciasis elimination with  
334 ivermectin treatment. African Programme for Onchocerciasis Control; 2010.
- 335 16. Dunn C, Callahan K, Katarawa M, Richards F, Hopkins D, Jr PCW, et al. The  
336 Contributions of Onchocerciasis Control and Elimination Programs toward the Achievement  
337 of the Millennium Development Goals. *PLoS Negl Trop Dis. Public Library of Science*;  
338 2015;9:e0003703.
- 339 17. O’Hanlon SJ, Slater HC, Cheke RA, Boatman BA, Coffeng LE, Pion SDS, et al. Model-  
340 Based Geostatistical Mapping of the Prevalence of *Onchocerca volvulus* in West Africa.  
341 Soares Magalhaes RJ, editor. *PLoS Negl Trop Dis.* 2016;10:e0004328.
- 342 18. Adler PH, Cheke RA, Post RJ. Evolution, epidemiology, and population genetics of black  
343 flies (Diptera: Simuliidae). *Infect Genet Evol.* 2010;10:846–65.
- 344 19. Agatsuma T. Genetic differentiation among natural populations of the vector of  
345 onchocerciasis, *Simulium ochraceum* in Guatemala. *Int J Trop Insect Sci.* 1987;8:465–9.
- 346 20. Archie EA, Luikart G, Ezenwa VO. Infecting epidemiology with genetics: a new frontier  
347 in disease ecology. *Trends Ecol Evol.* 2009;24:21–30.
- 348 21. Barry AE, Waltmann A, Koepfli C, Barnadas C, Mueller I. Uncovering the transmission  
349 dynamics of *Plasmodium vivax* using population genetics. *Pathog Glob Health. Taylor &*  
350 *Francis*; 2015;109:142–52.
- 351 22. Blouin MS, Yowell CA, Courtney CH, Dame JB. Host movement and the genetic structure  
352 of populations of parasitic nematodes. *Genetics.* 1995;141:1007–14.
- 353 23. Charalambous M, Lowell S, Arzube M, Lowry CA. Isolation by distance and a  
354 chromosomal cline in the Cayapa cytospecies of *Simulium exiguum*, the vector of human  
355 onchocerciasis in Ecuador. *Genetica.* 2005;124:41–59.
- 356 24. Choi Y-J, Tyagi R, McNulty SN, Rosa BA, Ozersky P, Martin J, et al. Genomic diversity  
357 in *Onchocerca volvulus* and its *Wolbachia* endosymbiont. *Nat Microbiol. Nature Publishing*  
358 *Group*; 2016;2:1–10.
- 359 25. Crawford KE, Hedtke SM, Doyle SR, Kuesel AC, Armoo S, Osei-Atweneboana M, et al.  
360 Utility of the *Onchocerca volvulus* mitochondrial genome for delineation of parasite

- 361 transmission zones [Internet]. Evolutionary Biology; 2019 Aug. Available from:  
362 <http://biorxiv.org/lookup/doi/10.1101/732446>
- 363 26. Doyle SR, Bourguinat C, Nana-Djeunga HC, Kengne-Ouafo JA, Pion SDS, Bopda J, et al.  
364 Genome-wide analysis of ivermectin response by *Onchocerca volvulus* reveals that genetic  
365 drift and soft selective sweeps contribute to loss of drug sensitivity. Unnasch TR, editor. PLoS  
366 Negl Trop Dis. 2017;11:e0005816.
- 367 27. Hedtke SM, Kuesel AC, Crawford KE, Graves PM, Boussinesq M, Boussinesq M, et al.  
368 Genomic Epidemiology in Filarial Nematodes: Transforming the Basis for Elimination  
369 Program Decisions. Front Genet. 2020;10:1282–1282.
- 370 28. Small ST, Labbé F, Coulibaly YI, Nutman TB, King CL, Serre D, et al. Human Migration  
371 and the Spread of the Nematode Parasite *Wuchereria bancrofti*. Rogers R, editor. Mol Biol  
372 Evol. 2019;36:1931–41.
- 373 29. Balkenhol N, editor. Landscape genetics: concepts, methods, applications. Chichester,  
374 West Sussex, UK ; Hoboken, NJ, USA: Wiley Blackwell; 2016.
- 375 30. Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape  
376 ecology and population genetics. Trends Ecol Evol. 2003;18:189–97.
- 377 31. Schwabl P, Llewellyn MS, Landguth EL, Andersson B, Kitron U, Costales JA, et al.  
378 Prediction and Prevention of Parasitic Diseases Using a Landscape Genomics Framework.  
379 Trends Parasitol. 2017;33:264–75.
- 380 32. Hemming-Schroeder E, Lo E, Salazar C, Puente S, Yan G. Landscape Genetics: A Toolbox  
381 for Studying Vector-Borne Diseases. Front Ecol Evol. 2018;6:21.
- 382 33. Peterman WE. ResistanceGA: An R package for the optimization of resistance surfaces  
383 using genetic algorithms. Jarman S, editor. Methods Ecol Evol. 2018;9:1638–47.
- 384 34. McRae BH, Dickson BG, Keitt TH, Shah VB. Using circuit theory of model connectivity  
385 in ecology, evolution and conservation. Ecology. 2008;89:2712–24.
- 386 35. Spear SF, Balkenhol N, Fortin M-J, Mcrae BH, Scribner K. Use of resistance surfaces for  
387 landscape genetic studies: considerations for parameterization and analysis: resistance  
388 surfaces in landscape genetics. Mol Ecol. 2010;19:3576–91.
- 389 36. Otabil KB, Gyasi SF, Awuah E, Obeng-Ofori D, Tenkorang SB, Kessie JA, et al. Biting  
390 rates and relative abundance of *Simulium* flies under different climatic conditions in an  
391 onchocerciasis endemic community in Ghana. Parasit Vectors. 2020;13:229.
- 392 37. Walsh J, Davies J, Le Berre R, others. Entomological aspects of the first five years of the  
393 Onchocerciasis Control Programme in the Volta River Basin. Tropenmed Parasitol.  
394 1979;30:328–44.

- 395 38. Awadzi K, Attah SK, Addy ET, Opoku NO, Quartey BT, Lazdins-Helds JK, et al. Thirty-  
396 month follow-up of sub-optimal responders to multiple treatments with ivermectin, in two  
397 onchocerciasis-endemic foci in Ghana. *Ann Trop Med Parasitol*. 2004;98:359–70.
- 398 39. Awadzi K, Boakye DA, Edwards G, Opoku NO, Attah SK, Osei-Atweneboana MY, et al.  
399 An investigation of persistent microfilaridermias despite multiple treatments with ivermectin,  
400 in two onchocerciasis-endemic foci in Ghana. *Ann Trop Med Parasitol*. 2004;98:231–49.
- 401 40. Osei-Atweneboana MY, Awadzi K, Attah SK, Boakye DA, Gyapong JO, Prichard RK.  
402 Phenotypic Evidence of Emerging Ivermectin Resistance in *Onchocerca volvulus*. Lustigman  
403 S, editor. *PLoS Negl Trop Dis*. 2011;5:e998.
- 404 41. Boakye DA, Back C, Fiasorgbor GK, Sib APP, Coulibaly Y. Sibling species distributions  
405 of the *Simulium damnosum* complex in the West African Onchocerciasis Control Programme  
406 area during the decade 1984-93, following intensive larviciding since 1974: *Simulium*  
407 *damnosum* in West Africa. *Med Vet Entomol*. 1998;12:345–58.
- 408 42. Klutse NAB, Owusu K, Ntiama-Baidu Y. Assessment of Patterns of Climate Variables  
409 and Malaria Cases in Two Ecological Zones of Ghana. *Open J Ecol*. 2014;4:764–75.
- 410 43. WHO. Onchocerciasis control in the Volta river basin area: report of the mission for  
411 preparatory assistance to the governments of Dahomey, Ghana, Ivory Coast, Mali, Niger,  
412 Togo and Upper Volta. World Health Organization; 1973.
- 413 44. Yaméogo L. Special Intervention Zones. *Ann Trop Med Parasitol*. 2008;102:23–4.
- 414 45. Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, et al. The Shuttle Radar  
415 Topography Mission. *Rev Geophys*. 2007;45:RG2004.
- 416 46. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for  
417 global land areas. *Int J Climatol*. 2017;37:4302–15.
- 418 47. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The  
419 Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA  
420 sequencing data. *Genome Res*. 2010;20:1297–303.
- 421 48. Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources  
422 of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2020;49:D10–7.
- 423 49. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence  
424 data. *Bioinformatics*. 2014;30:2114–20.
- 425 50. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.  
426 ArXiv13033997 Q-Bio [Internet]. 2013 [cited 2022 Mar 20]; Available from:  
427 <http://arxiv.org/abs/1303.3997>
- 428 51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence  
429 Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.

- 430 52. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes:  
431 A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J*  
432 *Comput Biol.* 2012;19:455–77.
- 433 53. Seemann T. VelvetOptimiser: automate your Velvet assemblies [Internet]. 2012 [cited  
434 2023 Jan 18]. Available from: <https://github.com/tseemann/VelvetOptimiser>
- 435 54. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn  
436 graphs. *Genome Res.* 2008;18:821–9.
- 437 55. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an  
438 integrated tool for comprehensive microbial variant detection and genome assembly  
439 improvement. *PloS One.* 2014;9:e112963.
- 440 56. Maddison W, Maddison D. Mesquite: a modular system for evolutionary analysis. Version  
441 3.40. 2018. 2018.
- 442 57. Day JC, Gweon HS, Post RJ. Sequence and organization of the complete mitochondrial  
443 genome of the blackfly *Simulium variegatum* (Diptera: Simuliidae). *Mitochondrial DNA Part*  
444 *B.* Taylor & Francis; 2016;1:799–801.
- 445 58. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant  
446 call format and VCFtools. *Bioinformatics.* 2011;27:2156–8.
- 447 59. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing.  
448 ArXiv12073907 Q-Bio [Internet]. 2012 [cited 2022 Mar 20]; Available from:  
449 <http://arxiv.org/abs/1207.3907>
- 450 60. Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. Vcflib and tools for  
451 processing the VCF variant call format [Internet]. *Bioinformatics*; 2021 May. Available from:  
452 <http://biorxiv.org/lookup/doi/10.1101/2021.05.21.445151>
- 453 61. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of  
454 SAMtools and BCFtools. *GigaScience.* 2021;10:giab008.
- 455 62. ESPEN. Site level onchocerciasis prevalence data [Internet]. ESPEN; 2020. Available  
456 from: <https://espen.afro.who.int/diseases/onchocerciasis>
- 457 63. Cheke RA, Young S, Garms R. Ecological characteristics of *Simulium* breeding sites in  
458 West Africa. *Acta Trop.* 2017;167:148–56.
- 459 64. Opoku A. The ecology and biting activity of blackflies (Simuliidae) and the prevalence of  
460 onchocerciasis in an agricultural community in Ghana. *West Afr J Appl Ecol* [Internet]. 2006  
461 [cited 2023 Feb 15];9. Available from:  
462 <http://www.ajol.info/index.php/wajae/article/view/45689>
- 463 65. Barro AS, Oyana TJ. Predictive and epidemiologic modeling of the spatial risk of human  
464 onchocerciasis using biophysical factors: A case study of Ghana and Burundi. *Spat Spatio-*  
465 *Temporal Epidemiol.* 2012;3:273–85.



- 466 66. Cheke RA, Basáñez M-G, Perry M, White MT, Garms R, Obuobie E, et al. Potential effects  
467 of warmer worms and vectors on onchocerciasis transmission in West Africa. *Philos Trans R*  
468 *Soc B Biol Sci.* 2015;370:20130559.
- 469 67. Cromwell EA, Osborne JCP, Unnasch TR, Basáñez M-G, Gass KM, Barbre KA, et al.  
470 Predicting the environmental suitability for onchocerciasis in Africa as an aid to elimination  
471 planning. *PLoS Negl Trop Dis.* Public Library of Science; 2021;15:e0008824.
- 472 68. WHO. Report of the Third Meeting of the WHO Onchocerciasis Technical Advisory  
473 Subgroup Geneva, Switzerland, 26–28 February 2019 [Internet]. 2020. Available from:  
474 <https://www.who.int/publications-detail-redirect/9789240006638>
- 475 69. Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth  
476 Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens Environ.* 2017;202:18–  
477 27.
- 478 70. Hijmans RJ, Van Etten J, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, et al. Package  
479 ‘raster.’ R Package. 2015;734.
- 480 71. R Core Team. R: A Language and Environment for Statistical Computing [Internet].  
481 Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from:  
482 <https://www.R-project.org/>
- 483 72. Hemming-Schroeder E, Zhong D, Machani M, Nguyen H, Thong S, Kahindi S, et al.  
484 Ecological drivers of genetic connectivity for African malaria vectors *Anopheles gambiae* and  
485 *An. arabiensis*. *Sci Rep.* 2020;10:19946.
- 486 73. Saarman N, Burak M, Opiro R, Hyseni C, Echodu R, Dion K, et al. A spatial genetics  
487 approach to inform vector control of tsetse flies (*Glossina fuscipes fuscipes*) in Northern  
488 Uganda. *Ecol Evol.* 2018;8:5336–54.
- 489 74. Moraga P, Cano J, Baggaley RF, Gyapong JO, Njenga SM, Nikolay B, et al. Modelling  
490 the distribution and transmission intensity of lymphatic filariasis in sub-Saharan Africa prior  
491 to scaling up interventions: integrated use of geostatistical and mathematical modelling.  
492 *Parasit Vectors.* 2015;8:560.
- 493 75. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models  
494 by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol.*  
495 2009;71:319–92.
- 496 76. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers.  
497 *Bioinformatics.* 2008;24:1403–5.
- 498 77. Leigh JW, Bryant D. popart: full-feature software for haplotype network construction.  
499 *Methods Ecol Evol.* 2015;6:1110–6.
- 500 78. Goudet J. hierfstat, a package for r to compute and test hierarchical F-statistics. *Mol Ecol*  
501 *Notes.* 2005;5:184–6.

- 502 79. Adamack AT, Gruber B. POPGENREPORT: simplifying basic population genetic analyses  
503 in R. Dray S, editor. *Methods Ecol Evol.* 2014;5:384–7.
- 504 80. Aktas C. haplotypes: Manipulating DNA Sequences and Estimating Unambiguous  
505 Haplotype Network with Statistical Parsimony [Internet]. 2020 [cited 2022 Mar 21]. Available  
506 from: <https://CRAN.R-project.org/package=haplotypes>
- 507 81. Rousset F. Genetic Differentiation and Estimation of Gene Flow from  $F$ -Statistics Under  
508 Isolation by Distance. *Genetics.* 1997;145:1219–28.
- 509 82. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies.  
510 *Genetics.* 1995;139:457–62.
- 511 83. Rohlf FJ, Schnell GD. An Investigation of the Isolation-By-Distance Model. *Am Nat.* The  
512 University of Chicago Press; 1971;105:295–324.
- 513 84. Wright S. Isolation by Distance. *Genetics.* 1943;28:114–38.
- 514 85. Savary P, Foltête J, Moal H, Vuidel G, Garnier S. graph4lg: A package for constructing  
515 and analysing graphs for landscape genetics in R. Gaggiotti O, editor. *Methods Ecol Evol.*  
516 2021;12:539–47.
- 517 86. Diggle P. Model-based geostatistics for global public health: methods and applications.  
518 Boca Raton: Taylor & Francis; 2019.
- 519 87. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O’hara R, et al. Package  
520 ‘vegan.’ *Community Ecol Package Version.* 2013;2:1–295.
- 521 88. McRae BH. Isolation by resistance. *Evolution.* 2006;60:1551–61.
- 522 89. Kimberly R. Hall, Ranjan Anantharaman, Vincent A. Landau, Melissa Clark, Melissa  
523 Clark, Brett G. Dickson, et al. circuitscape in julia empowering dynamic approaches to  
524 connectivity assessment. *Land.* 2021;
- 525 90. Clarke RT, Rothery P, Raybould AF. Confidence limits for regression relationships  
526 between distance matrices: Estimating gene flow with distance. *J Agric Biol Environ Stat.*  
527 2002;7:361.
- 528 91. Fukuda Y, Moritz C, Jang N, Webb G, Campbell H, Christian K, et al. Environmental  
529 resistance and habitat quality influence dispersal of the saltwater crocodile. *Mol Ecol.*  
530 2022;31:1076–92.
- 531 92. Cushman SA, Landguth EL. Spurious correlations and inference in landscape genetics.  
532 *Mol Ecol.* 2010;19:3592–602.
- 533 93. Wang IJ. Examining the full effects of landscape heterogeneity on spatial genetic variation:  
534 a multiple matrix regression approach for quantifying geographic and ecological isolation:  
535 special section. *Evolution.* 2013;67:3403–11.



- 536 94. De Castro O, Di Maio A, Di Febbraro M, Imperato G, Innangi M, Véla E, et al. A Multi-  
537 Faceted Approach to Analyse the Effects of Environmental Variables on Geographic Range  
538 and Genetic Structure of a Perennial Psammophilous Geophyte: The Case of the Sea Daffodil  
539 *Pancreatum maritimum* L. in the Mediterranean Basin. Peruzzi L, editor. PLOS ONE.  
540 2016;11:e0164816.
- 541 95. Goslee SC, Urban DL. The **ecodist** Package for Dissimilarity-based Analysis of Ecological  
542 Data. J Stat Softw [Internet]. 2007 [cited 2022 Mar 21];22. Available from:  
543 <http://www.jstatsoft.org/v22/i07/>
- 544 96. Bempah G, Boama P. Effects of hydroelectric dam construction on land use land cover  
545 changes in Bui national park, Ghana. Mercat Fortaleza [Internet]. Universidade Federal do  
546 Ceará; 2021 [cited 2023 Jan 12];20. Available from:  
547 <http://www.scielo.br/j/mercator/a/PLxRFrG4YHhJpKNS7t7KZKF/>
- 548 97. Adeleke MA, Mafiana CF, Sam-Wobo SO, Olatunde GO, Ekpo UF, Akinwale OP, et al.  
549 Biting behaviour of *Simulium damnosum* complex and *Onchocerca volvulus* infection along  
550 the Osun River, Southwest Nigeria. Parasit Vectors. BioMed Central; 2010;3:1–7.
- 551 98. Renz A. Studies on the dynamics of transmission of onchocerciasis in a Sudan-savanna  
552 area of North Cameroon II: Seasonal and diurnal changes in the biting densities and in the  
553 age-composition of the vector population. Ann Trop Med Parasitol. 1987;81:229–37.
- 554 99. Gyan ET. Analysis of Population Structure of *Simulium damnosum sensu lato* In the  
555 Ecological Transition Zone of Central Ghana [PhD Thesis]. La Trobe University; 2019.
- 556 100. Katarawa MN, Zarroug IMA, Negussu N, Aziz NM, Tadesse Z, Elmubark WA, et al.  
557 The Galabat-Metema cross-border onchocerciasis focus: The first coordinated interruption of  
558 onchocerciasis transmission in Africa. Makepeace BL, editor. PLoS Negl Trop Dis.  
559 2020;14:e0007830.
- 560 101. Post RJ, Cheke RA, Boakye DA, Wilson MD, Osei-Atweneboana MY, Tetteh-Kumah  
561 A, et al. Stability and change in the distribution of cytospecies of the *Simulium damnosum*  
562 complex (Diptera: Simuliidae) in southern Ghana from 1971 to 2011. Parasit Vectors.  
563 2013;6:205.
- 564 102. Zarroug IMA, Elaagip A, Gumaa SG, Ali AK, Ahmed A, Siam HAM, et al. Notes on  
565 distribution of *Simulium damnosum s. l.* along Atbara River in Galabat sub-focus, eastern  
566 Sudan. BMC Infect Dis. 2019;19:477.
- 567 103. Doyle SR, Armoo S, Renz A, Taylor MJ, Osei-Atweneboana MY, Grant WN.  
568 Discrimination between *Onchocerca volvulus* and *O. ochengi* filarial larvae in *Simulium*  
569 *damnosum (s.l.)* and their distribution throughout central Ghana using a versatile high-  
570 resolution speciation assay. Parasit Vectors. 2016;9:536.
- 571 104. Renz A, Organization WH, others. Studies on the reinvasion by *Simulium damnosum s.l.*  
572 into the Eastern areas of Onchocerciasis Control Programme and on the vectorial capacity of  
573 different species of the *S. damnosum* complex in Togo and Benin 1982. World Health Organ

- 574 [Internet]. Onchocerciasis Control Programme in the Volta River Basin Area; 1982; Available  
575 from: <https://apps.who.int/iris/handle/10665/326643>
- 576 105. Wahl G, Enyong P, Ngosso A, Schibel JM, Moyou R, Tubbesing H, et al. *Onchocerca*  
577 *ochengi* : epidemiological evidence of cross-protection against *Onchocerca volvulus* in man.  
578 Parasitology. 1998;116:349–62.
- 579 106. Frempong KK, Walker M, Cheke RA, Tetevi EJ, Gyan ET, Owusu EO, et al. Does  
580 Increasing Treatment Frequency Address Suboptimal Responses to Ivermectin for the Control  
581 and Elimination of River Blindness? Clin Infect Dis. 2016;62:1338–47.
- 582 107. de Vos AS, Stolk WA, Coffeng LE, de Vlas SJ. The impact of mass drug administration  
583 expansion to low onchocerciasis prevalence settings in case of connected villages. PLoS Negl  
584 Trop Dis. Public Library of Science; 2021;15:e0009011.
- 585 108. McCulloch K, McCaw J, McVernon J, Hedtke SM, Walker M, Milton P, et al.  
586 Investigation into the effect of host migration on the transmission of *Onchocerca volvulus*  
587 using a patch model. Am J Trop Med Hyg. 2017. p. 564–564.
- 588 109. Leempoel K, Duruz S, Rochat E, Widmer I, Orozco-terWengel P, Joost S. Simple Rules  
589 for an Efficient Use of Geographic Information Systems in Molecular Ecology. Front Ecol  
590 Evol. 2017;5:33.
- 591