Title: Utility of the *Onchocerca volvulus* mitochondrial genome for delineation of parasite transmission zones

Authors and addresses:

Katie E Crawford[1,2], Shannon M Hedtke[1,2], Stephen R Doyle[1,3], Annette C Kuesel[4], Samuel Armoo[5], Mike Osei-Atweneboana[5], Warwick N Grant[1,2]

[1] Department of Animal, Plant and Soil Sciences, La Trobe University, Bundoora, Victoria, Australia

[2] Department of Physiology, Anatomy and Microbiology, La Trobe University, Bundoora, Victoria, Australia

[3] Wellcome Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

[4] UNICEF/UNDP/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases, World Health Organization, Switzerland

[5] Biomedical and Public Health Research Unit, Council for Scientific and Industrial Research-Water Research Institute, Accra, Ghana

Corresponding author:

Shannon Hedtke
S.Hedtke@latrobe.edu.au
Department of Physiology, Anatomy, and Microbiology
La Trobe University
Bundoora, Victoria 3086 AUSTRALIA

Abstract

In 2012, the reduction in *Onchocerca volvulus* infection prevalence through long-term mass ivermectin distribution in African meso- and hyperendemic areas motivated expanding control of onchocerciasis (river blindness) as a public health problem to elimination of parasite transmission. Given the large contiguous hypo-, meso- and hyperendemic areas with an estimated population of 204 million, sustainable elimination requires an understanding of the geographic, and in turn genetic, boundaries of different parasite populations to ensure interventions are only stopped where the risk of re-introduction of the parasite through vector or human migration from areas with ongoing transmission is acceptable. These boundaries, which define the transmission zones of the parasite, may be delineated by characterising the parasite genetic population structure within and between potential zones. We analysed whole mitochondrial genome sequences of 189 *O. volvulus* adults to determine the pattern of genetic similarity across three West African countries: Ghana, Mali, and Côte d'Ivoire. Population structure measures indicate that parasites from the Pru, Daka and Black Volta/Tombe river basins in central Ghana belong to one parasite population, showing that different river basins cannot be assumed to constitute independent transmission zones. This research forms the basis for developing tools for elimination programs to delineate transmission zones, to estimate the risk of parasite re-introduction via vector or human movement when mass ivermectin administration is stopped in one area while transmission is ongoing in others, to identify the origin of infections detected post-treatment cessation, and to investigate whether migration contributes to persisting prevalence levels during interventions.

Keywords: onchocerciasis elimination, population genetics, transmission zone, River blindness, *Onchocerca volvulus*

Declarations of interest: none

1. Introduction

Onchocerciasis, or river blindness, is a disease caused by the parasitic filarial nematode, *Onchocerca volvulus,* and transmitted by *Simulium* spp. blackflies, whose larvae develop in fast-flowing streams and rivers. The high level of onchocerciasis-related morbidity (in particular blindness, visual impairment, and skin disease) and resulting social and economic burden on communities led to the implementation of large-scale control and elimination programs. An estimated 45 million people would have been infected world-wide in 2014 without these programs (Remme et al., 2017). The World Health Organization (WHO) estimated that in 2017, 205 million people globally should be included in control and elimination programs, with over 99% (204 million) in sub-Saharan Africa (World Health Organization, 2017).

Mass drug administration of ivermectin (MDAi) has been the principal strategy for control of onchocerciasis since 1987, after MDAi was demonstrated to be safe and Merck & Co, Inc. (Kenilworth, NJ, USA) decided to donate ivermectin (Mectizan®) (De Sole et al., 1989). The Onchocerciasis Elimination Programme for the Americas (OEPA) targeted elimination of morbidity and transmission through biannual MDAi from its initiation in 1993, and the African Programme for Onchocerciasis Control (APOC, 1995-2015) targeted control of onchocerciasis as a public health problem through annual MDAi in areas with an infection prevalence exceeding around 40%, a prevalence which was previously shown to be associated with an increased risk of blindness (Prost et al., 1979; UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, 1992). Research studies and epidemiological evaluations (Diawara et al., 2009; Fobi et al., 2015; Traore et al., 2012) suggested that annual or biannual MDAi may have interrupted parasite transmission in many areas in Africa (Tekle et al., 2012; Tekle et al., 2016) and in 11 small foci across 5 countries in OEPA (World Health Organization, 2017). These successes led in 2012 to the decision to move from ensuring sustainable systems for control of onchocerciasis as a public health problem to elimination of onchocerciasis from 80% of endemic African countries by 2025; i.e., changing the focus from reducing the disease burden to permanent interruption of transmission in all areas.

This change means that countries face a number of new challenges including, but not limited to, delineating all areas that now need to be included in interventions, including areas of low prevalence, evaluating progress towards elimination using appropriate epidemiological and entomological evaluations,

and, crucially, making decisions on when and where to stop MDAi that take into account the need for collaborations between intervention projects or units within an individual country and between national intervention programs in neighbouring countries (Boakye et al., 2018; Bush et al., 2018; Rebollo et al., 2018; Unnasch et al., 2018). In 2016, WHO issued new guidelines which define procedures and entomological criteria for stopping MDAi and commencing post-treatment surveillance, and for subsequent verification of elimination of onchocerciasis and the start of post-elimination surveillance (World Health Organization, 2016). These guidelines also suggested a definition of an onchocerciasis transmission zone as the epidemiological unit within which these procedures and criteria should be applied.

In the context of onchocerciasis, elimination is "the reduction to zero of the incidence of infection caused by a specific agent in a defined geographical area as a result of deliberate efforts" (Dowdle, 1998) with post-elimination measures based on the outcomes of continued surveillance. Assuming these criteria are appropriate and have been met within the geographic area evaluated, there are, in principle, two different reasons elimination may nevertheless not have been achieved or may not be sustainable (Hedtke et al., submitted). First, residual infections not detected during the evaluations may be sufficient for the 'locally endemic' parasite population to recover. Second, infected individuals and/or infective vectors may migrate from areas of active transmission into areas where interventions were stopped, and lead to re-establishment of a parasite population via immigration. The magnitude of this risk depends on the alignment of the geographical boundaries for evaluating epidemiological and entomological indicators of transmission with the boundaries within which parasites (in humans or vectors) move and are transmitted, i.e., the parasite transmission zone. The 2016 WHO guidelines define a transmission zone as "a geographical area where transmission of *O. volvulus* occurs by locally breeding vectors and which can be regarded as a natural ecological and epidemiological unit for interventions". The experience from the OCP in West Africa has shown that epidemiologically significant transmission occurred over vast areas: re-invasion of geographic areas where larviciding had interrupted larval development by vectors from distant neighbouring areas was observed (Boatin et al., 1997; Dadzie et al., 2018) and demonstrated that long-range transmission poses a significant threat to elimination even if it is at low frequency. Thus decisions on the 'natural ecological and epidemiological unit for interventions' need to take into consideration the long-

term spatial density and migration patterns of both the vector and the human host (Barrett et al., 2008; Blouin et al., 1995; Criscione and Blouin, 2004; Criscione et al., 2005; Jarne and Théron, 2001; Nadler, 1995; Prugnolle et al., 2005). For onchocerciasis elimination programs to make evidence-based decisions on transmission zone boundaries, they need tools to generate this evidence. We have embarked on a research program to develop such tools.

Rapid Epidemiological Mapping of Onchocerciasis (REMO; based on nodule palpation (Noma et al., 2002)), shows that Africa is a complex mosaic of contiguous regions of low, moderate and high infection intensity (hypo- meso- and hyper-endemic, respectively; Figure 1) (African Programme for Onchocerciasis and World Health Organization, 2013). This continuous but heterogeneous pattern of endemicity suggests that parasite transmission occurs in a similarly continuous mosaic of overlapping "zones" of high, moderate and low transmission. These zones can be delineated using parasite population structure as a proxy for transmission zone boundaries: parasites from the same transmission zone are able to interbreed and thus are more closely related genetically than they are to parasites from different transmission zones, with which they are less likely to interbreed and thus are genetically less related. Consequently, the likelihood that parasites originate from the same transmission zone and the likelihood of parasite transmission between two locations (invasion) can be inferred from the degree of genetic relatedness between them, and defined quantitatively by estimating parameters of population structure (Archie et al., 2009; Real and Biek, 2007; Schwabl et al., 2017).

Early studies of *O. volvulus* found little genetic diversity between parasites isolated from different locations in Africa (Keddie et al., 1998; Zimmerman et al., 1994). However, recent whole genome studies identified large numbers of single nucleotide polymorphisms (SNPs) across the parasite nuclear, mitochondrial, and endosymbiotic bacterial (*Wolbachia* sp.) genomes, with extensive population genetic structure over large (>2000km) (Choi et al., 2016; Crainey et al., 2016) and moderate (>200km) geographic distances (Doyle et al., 2017). However, these genomic studies either focused on only a small number of samples, used pooled samples, or analysed only a small region of the genome. To date, there has been no large-scale survey of genetic diversity and population structure of *O. volvulus* at multiple spatial scales.

In the study reported here, we tested the utility of population genetic markers based on variation in the whole mitochondrial genome of *O. volvulus* for identifying transmission zones and developing genetics-based tools for onchocerciasis control and elimination programs. The generally high copy number of the *O. volvulus* mitochondrial genome makes it easier to amplify than the nuclear genome, and the lack of recombination makes it a useful marker for understanding population demographic history (e.g.,(Avise et al., 1987)). The use of the whole mitochondrial genome, rather than a specific gene, is important for avoiding ascertainment bias and thus underestimating genetic diversity and overestimating gene flow (Ingman et al., 2000; Torroni et al., 2006; Wakeley et al., 2001). We performed high throughput sequencing of the whole mitochondrial genome of 189 adult nematodes sampled from three countries in West Africa (Côte d'Ivoire, Ghana, and Mali) at two geographical scales. Spatial genetic structure between the three West African countries, and within one country, were used to propose genetically-defined transmission zones at the two spatial scales examined.

2. Materials and Methods

2.1 Sampling

Genomic DNA from 156 individual adult worms was used for long-range PCR, and 36 mt genome sequences were extracted bioinformatically from WGS (S. Hedtke, pers. comm.), for a total of 192 samples. Two of the 192 samples (ASU-7-15 and NLG-86-12) sequenced poorly and were excluded from the analysis, as was one sample (COT-34) that had been sequenced in duplicate. Of the 189 remaining samples, there were two adult male and 187 adult female worms. All adults were obtained by surgical excision of nodules: 13 from 1 sampling location in Côte d'Ivoire, 10 from 4 sampling locations in Mali, and 166 from Ghana (Figure 2). Sampling within Ghana occurred in 16 villages (Table S1, Supplementary Information) across three river basins, with a total of 74 samples from 6 villages in the Black Volta/Tombe (BVT) river basin, 22 samples from 4 villages in the Pru river basin, and 70 samples from 6 locations in the Daka river basin. Individual worms were isolated from nodules using collagenase digestion (Schulz-Key, 1988). Samples from the same river basin or country location are subsequently referred to as 'sample sets'. Details of the collection of the Ghanaian samples have been reported previously (Ardelli and Prichard, 2004, 2007).

2.2 DNA extraction and PCR Amplification

DNA was extracted from each worm using the DNeasy® Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. Whole mitochondrial genomes from each nematode sample were PCR amplified using one of three long range PCR strategies: (i) as a single amplicon of 13467 bp, 23 individuals (ii) as two amplicons of 6870 bp and 6912 bp, 95 individuals, or (iii) as three amplicons of 5171 bp, 5572 bp, and 5838 bp, 36 individuals (Figure 3; see Table S2 for primer sequences). PCRs using the GoTaq Long PCR Master Mix (Promega Corp, Wisconsin, USA) using 0.6-0.8 µM of each primer and a modified touchdown PCR protocol for the largest amplicon (Don et al., 1991) (Supplementary Information, Table S3). Each reaction was visualised on a 1 % agarose gel using a 1 kb DNA ladder (New England Biolabs) and a Lambda ladder (Roche) to confirm product size and specificity.

2.3 Mitochondrial Amplicon Re-sequencing

All PCR products were standardised to a concentration of 500 ng in 50 µl. When an individual sample had multiple PCR products (i.e., the two and three amplicon strategies), the products were combined at equimolar amounts to a total of 500ng in 50ul. Each standardised sample was sheared using a BioRuptor UCD-200 (Diagenode) at high power for 3 minutes with an interval of 30 seconds shearing and 30 seconds rest. Library preparation was performed using a modified Illumina TruSeq DNA HT Sample Prep kit protocol (https://nematodegenetics.files.wordpress.com/2014/05/truseq-ht-workflow-sept2013.docx) to allow Agencourt AMPure XP bead (Beckman Coulter) clean-up in a plate to minimise DNA loss from plate transfers. Samples were then standardised to 40 ng using a Qubit® 2.0 Fluorometer (Invitrogen Life Technologies). Each standardised sample was combined and cleaned using Agencourt AMPure XP beads. The library was size selected for a 400-600 bp insert using Pippin Prep™ (Sage Science) and then sequenced on an Illumina MiSeq® using v2 2x250bp chemistry.

Raw sequence reads had low-quality bases (below quality of 5) and any TruSeq3 Illumina adapters trimmed using Trimmomatic v. 0.32 palindrome trimming (minimum length = 100, 4-bp sliding window with average quality >20, palindrome clip threshold of 30, simple clip threshold of 10). Trimmed reads were then mapped to the *O. volvulus* mitochondrial reference genome (GenBank accession number NC_001861) using BWA-MEM v. 0.7.12 ((Li, 2013); see Supplementary Table S4). The mapping files were sorted and converted

to bam files using SAMtools v. 1.2 (Li et al., 2009) and indels were realigned using GATK v.3.3 RealignerTargetCreator and IndelRealigner (DePristo et al., 2011). Duplicate reads were removed using Picard v.1.115 MarkDuplicates (http://broadinstitute.github.io/picard/). Variants were called assuming a haploid genome using two different programs, GATK UnifiedGenotyper and freebayes v.0.9.14-17-g7696787 (with minimum allele count of 5; (Garrison and Marth, 2012)). Variants were filtered using VCFtools v.0.1.12a (Danecek et al., 2011) to exclude primer and missing regions, indels, and non-biallelic sites. Variants with a quality below 30 were removed using bcftools v.1.2 (Li, 2011) (https://samtools.github.io/bcftools/). As freebayes outputs multi-nucleotide polymorphisms, these were split using vcflib v.1 vcfallelicprimitives (https://github.com/ekg/vcflib). Concordant variants between UnifiedGenotyper and freebayes were identified, and discordant and singleton variants were excluded using VCFtools. Variants were annotated using SNPeff v3.6c using an invertebrate mitochondrial codon table (Cingolani et al., 2012) and the *O. volvulus* mitochondrial reference genome (GenBank accession number NC_001861).

2.4 Statistical Analysis

Files were converted from fasta to phylip and genepop formats using PGDSpider v.2.0.8.3 (Lischer and Excoffier, 2012). Consensus sequences were generated for each nematode sample using VCFtools vcf-consensus. Minimum spanning haplotype networks were generated using PopART v.1.7 (Leigh and Bryant, 2015). Polymorphic sites and haplotype and nucleotide diversity were calculated using DNAsp v.5.10.1 (Librado and Rozas, 2009). The number of haplotypes, Tajima's D (Tajima, 1989), Fu's F (Fu, 1997), pairwise φPT, M values (Slatkin, 1991), and Mantel test comparing genetic and geographic distance (Mantel, 1967) were calculated using Arlequin v.3.5.2.2 (Excoffier and Lischer, 2010). Discriminant analysis of principal components (DAPC), membership probabilities, and individual reassignment scores were generated using adegenet v.2.0.1, with the optimal number of PCs retained determined using the optimascore function (Jombart, 2008; Jombart and Ahmed, 2011; Jombart et al., 2010).

To examine changes in effective population size (N_e) in the Ghana population over time, we produced a Bayesian skyline plot. Best-fit models of sequence evolution for six possible partitions of the sequence alignment of the entire mitochondrial genomes were determined using the Bayesian Information

Criterion in PartitionFinder v1.1.1 (Lanfear et al., 2017), using PhyML for tree estimation (Guindon et al., 2010). Best-fit partitioning schemes placed intergenic regions and genes coding for tRNAs and rRNAs in the same partition, under the HKY+I+G model (Hasegawa et al., 1985), with each codon placed in an additional partition; codon 1 with the HKY+I+G model and codons 2 and 3 with the TrN+I+G model (Tamura and Nei, 1993). These models and partitioning scheme were used to produce a Bayesian skyline plot representing changes in effective population size in BEAST 2 (Bouckaert et al., 2014), with trees linked across partitions, estimating kappa with a starting value of 2.0 and a log normal prior, the proportion of invariant sites with a starting value of 0.2 and a uniform prior, and the gamma shape parameter with a starting value of 1.0 and an exponential prior given 4 rate categories. A strict clock model was used, also linked across partitions, with a clock rate of 9.7×10^{-8}, based on an estimated substitution rate from *Caenorhabditis elegans* (Denver et al., 2000) and assuming one *O. volvulus* generation per year - a potentially problematic assumption, as filarial nematodes are long-lived and have overlapping generations. The number of intervals within which coalescent events could occur was 5 (i.e., as many as 4 changes in effective population size). Two independent MCMC runs were performed for 150 million generations, storing trees and parameters every 1000 generations, and discarding 10% of each run as burn-in. Stationarity and convergence across runs were assessed using Tracer v1.7.1 (Rambaut et al., 2018), requiring the effective sample size (ESS) to be at least 125 for each run and for the two runs to have strongly overlapping marginal density plots and overlapping mean estimates with corresponding 95% highest posterior density intervals. Stationarity of alpha (the parameter that describes gamma, which in turn describes the shape of rate heterogeneity across sites) and the proportion of invariant sites were difficult to achieve for three of these partitions, with ESS values lower than 100 even after 100 million generations. Examination of the trace files indicated that this was being driven by interactions between these two parameters, as large fluctuations in parameter values were linked. When invariant sites make up a high proportion of the data and when variation at other sites is relatively low, as with these sequences, the algorithm vacillates between incorporating invariant sites into the gamma shape parameter or parameterizing invariant sites separately (Brown et al., 2010; Stamatakis, 2016). These fluctuations were solved by increasing the number of rate categories for the gamma

234 estimation from 4 to 5 and removing the proportion of invariant sites as an independent parameter. Tracer
235 v1.7.1 was used to generate Bayesian skyline plots.
236
237 **3. Results**
238 *3.1 Genetic Diversity*
239 Variants in the whole mitochondrial genome were identified from 189 samples. A total of 486 high
240 quality SNPs that passed stringent filtering were called, comprised of 297 singleton SNPs (i.e., those only
241 present in one individual) and 189 shared SNPs (i.e., those found in at least 2 samples; Table S5,
242 Supplementary Information). These 189 shared SNPs represented 155 haplotypes (i.e., a group of alleles
243 inherited from a single parent) of which only 19 were identical between 2 or more individuals (Table S5,
244 Supplementary Information). This high haplotype diversity of between 0.96 and 0.99 for each sample set
245 was observed in all communities and countries examined (Table 1). However, the corresponding nucleotide
246 diversity was low, ranging from 0.0006-0.0008 per population, due to the small number of polymorphic
247 sites per individual genome (average per sample 1.05-4.10, Table 1).
248 As singletons are not easily distinguishable from PCR error and thus would not be informative,
249 statistical analysis was performed only using SNPs found in more than one individual worm. Tajima's D, Fu's
250 F and Fu and Li's D are statistical tests of neutrality (Fu, 1997; Fu and Li, 1993; Tajima, 1989); i.e., they test
251 whether mutations are passed on to progeny randomly or are subject to selection or other forces that
252 restrict random inheritance by progeny. Non-zero values indicate divergence from neutral variation (when
253 mutation and genetic drift are at equilibrium); a positive value suggests balancing selection (multiple alleles
254 are maintained in a gene pool) or a population bottleneck (only a fraction of a population, and thus only a
255 fraction of the genetic diversity within the population contributed to the next generation), whereas a
256 negative value suggests the occurrence of a selective sweep (when individuals with an allele that provides a
257 survival/fecundity advantage have more offspring that inherit that beneficial allele than those that do not,
258 reducing the overall genetic diversity) or a recent population expansion (when common alleles increase
259 because the number of individuals in the population has expanded faster than new mutations can arise)
260 (Tajima, 1989). Tajima's D values (Table 1) were negative for the samples from Ghana ($p < 0.05$), whereas

those from Mali and those from Côte d'Ivoire were not significantly different from zero, which may be due to the low number of parasites in each of these sample sets (n = 10 and n = 13, respectively).

A positive or negative Fu's F, and Fu and Li's D, indicates a divergence from the expected number of alleles in a population: A positive value suggests balancing selection or a population bottleneck, whereas a negative value suggests an excess of alleles consistent with a population expansion or genetic hitchhiking (when genetic variation is physically linked on a chromosome, selection for an allele at one gene can increase the frequency of the "hitchhiker", and allele that does not impact survival (Fu, 1997). Fu's F values (Table 1) are consistent with the Tajima's D values found, with highly statistically significant ($p < 0.01$) negative values across the Ghana samples. Both Mali and Côte d'Ivoire sample sets had negative Fu's F values which were not statistically significant ($p > 0.05$), which again may be due to the low sample size.

The haplotype network shows a star-like pattern of many closely related haplotypes as indicated (Figure 4 A), consistent with the haplotype diversity statistics (Table 1). There is some clustering of haplotypes at a country level: most Côte d'Ivoire and Mali haplotypes group closely together in the networks (are more closely related to each other than they are to Ghanaian haplotypes), whereas Ghanaian haplotypes are spread throughout the whole haplotype network (Figure 4 B).

3.2 Population Structure

Population structure was analysed at 3 spatial scales: between countries, between river basins within a single country (Ghana) and between individual villages (Figure 2, Table 2 and Table 3). Pairwise comparisons of genetic sequence variation indicate statistically significant differences between Côte d'Ivoire and Ghana (Table 2; $\varphi_{PT} = 0.0180$, $p < 0.001$) and between Mali and Ghana ($\varphi_{PT} = 0.0112$, $p < 0.05$) but there was no difference found between Côte d'Ivoire and Mali ($\varphi_{PT} = 0$, $p > 0.05$). Consistent with these measures of differentiation, Slatkin's M values (Slatkin, 1991) indicate a high level of gene flow, i.e., historic or current migration of parasites followed by interbreeding, between Côte d'Ivoire and Mali, and much lower gene flow between Ghana and Côte d'Ivoire, and between Ghana and Mali (Table S6, Supplementary Information). A pairwise comparison of genetic sequence variation by river basin within Ghana (Table 2) showed a low level of genetic differentiation between the parasites from all three river basins (φ_{PT} values were not significantly different from zero; $p > 0.05$). Corresponding M values for the three river basins

(Table S6, Supplementary Information) showed higher gene flow between the Daka and Black Volta/Tombe river basins, relative to the gene flow between the Pru river basin and both the Daka and Black Volta/Tombe river basins. When samples were analysed pairwise by the village within Ghana where they had been collected, no measures of genetic differentiation were significantly different from zero ($p > 0.05$). Furthermore, M values between most villages suggest a high level of gene flow and thus past or current transmission between all villages (Table S7, Supplementary Information).

A Mantel test comparing genetic diversity and geographic distance indicates that there is a slight positive correlation ($r = 0.1989$, $p < 0.05$) between φ_{PT} values and the approximate distance (in km) between the sample sets from the three countries, which indicates isolation-by-distance (IBD) at this large geographic scale. There was no correlation between φ_{PT} values and distance in km between river basins within Ghana ($r = 0.0251$, $p > 0.05$), indicating interbreeding of parasites at this smaller spatial scale.

Discriminant analysis of principal components (DAPC (Jombart et al., 2010)), using the 45 principal components (PCs) that contribute most to genetic variation among the samples, demonstrates that variation within the data can be explained by population structuring at a large geographic scale (Figure 5 A) and at a small geographic scale (Figure 6 A) with some overlap between populations. Membership probabilities across the three countries (Figure 5 B-C) calculated by comparing the DAPC predicted country of origin with the actual country of origin show that the DAPC assign 96.4% of parasites from Ghana correctly to country of origin, but only 69.2% of parasites from Côte d'Ivoire and 50% of parasites from Mali assigned correctly. The lower percentages from Côte d'Ivoire and Mali may be due to the low number of parasites sampled and/or the lack of genetic structure between these two countries (as indicated by the parameters described in Table 2). Within Ghana, membership probabilities across the three river basins (Figure 6 B-C) indicate that there is some clustering of parasites by river basin but many parasites were assigned to the incorrect river basin: 25.7% of BVT-originating parasites, 28.6% of Daka-originating parasites, and 45.5% of Pru-originating parasites were not correctly assigned to their river basin of origin based on the genetic variation in their mitochondrial genome.

4. Discussion

315 *4.1 Genetic diversity and demographic history of O. volvulus*

316 Reconstruction of the whole genomes of 27 adult *O. volvulus* from Ecuador, Uganda, and forest and
317 savannah regions of West Africa (Choi et al., 2016) had pointed to greater levels of genetic diversity than
318 previously reported (Keddie et al., 1998; Zimmerman et al., 1994). Here, we describe the genetic diversity
319 and population structure of *O. volvulus* derived from complete mitochondrial genomes of 189 adult worms
320 sampled at two spatial scales in West Africa, thus (a) extending significantly the previously available data
321 and (b) analysing population structure at epidemiologically relevant spatial scales for the first time. The 155
322 unique haplotypes identified in this study (Table 1 & S5, Supplementary Information; Figure 4)
323 demonstrated conclusively that *O. volvulus* populations are extremely genetically diverse, and that the
324 historical population size of this parasite was likely very large. Furthermore, the close relationship between
325 the haplotypes suggests they share an evolutionarily recent origin, consistent with a rapid, large population
326 expansion following an evolutionarily recent host switch (Figure 4). The population expansion most likely
327 occurred after the speciation event that gave rise to *O. volvulus* and *O. ochengi* as a result of proposed host
328 switch by their most recent common ancestor following the introduction of domesticated cattle in
329 Northeast Africa approximately 2,500 - 10,000 years before present (Bain, 1981; Keddie et al., 1998;
330 Lefoulon et al., 2017; Marshall and Hildebrand, 2002). Following the host switch, the new species (*O.*
331 *volvulus*) presumably expanded rapidly into the naive human population of sub-Saharan Africa,
332 accumulating new mutations in its bottlenecked mitochondrial genome. The large population size (10^6)
333 suggested by these data is unsurprising when pre-control infection prevalence in the sampled regions is
334 considered (WHO Expert Committee on Onchocerciasis Control (1993 : Geneva and World Health
335 Organization, 1995; World Health Organization and Onchocerciasis Control Programme in West Africa,
336 1997).

337 *4.2 O. volvulus population structure in West Africa*

338 Our data indicate that there was no significant population differentiation between *O. volvulus* from
339 Côte d'Ivoire and Mali (Table 2) which implies that parasites from Mali and Côte d'Ivoire constitute a single
340 population. While investigation of a larger number of parasites from more areas in Mali and Côte d'Ivoire is
341 required, the lack of genetic structure between the parasite populations we examined could be explained

342 as a consequence of the periodic long-range vector migration with the Southwest-Northeast monsoonal
343 winds that required the expansion of the original OCP area (World Health Organization and Onchocerciasis
344 Control Programme in the Volta River Basin Area, 1985; World Health Organization and Onchocerciasis
345 Control Programme in West Africa, 1997). In this context, it is clear that this long-range vector migration is
346 epidemiologically significant and is most likely the primary determinant of the parasite's population
347 structure and hence of the spatial boundary of a parasite transmission zone, at the geographic scales
348 analysed here.

349 In contrast, our data indicate that *O. volvulus* populations are genetically structured (i.e., not
350 closely related) between Ghana and Côte d'Ivoire and between Ghana and Mali (Table 2). This implies
351 restricted gene flow east-west between Ghana and Côte d'Ivoire and Southeast-Northwest between Ghana
352 and Mali: parasites in Ghana mate infrequently with parasites from Côte d'Ivoire and Mali. While some
353 *Simulium* spp. subtypes can fly up to 250km (Garms et al., 1982; Post et al., 2013) and up to 500km when
354 aided by favourable wind conditions (Baker et al., 1990; Garms, 1981; Garms et al., 1979; Johnson et al.,
355 1985; Magor and Rosenberg, 1980). The maximum distances separating our sampling locations along the
356 transect between Mali and Ghana, and between Côte d'Ivoire and Ghana are larger (>500-1000km), and do
357 not align with the long-range vector migration between Côte d'Ivoire and Mali that follow the Southwest-
358 Northeast monsoon winds (Baker et al., 1990; Garms and Ochoa, 1979; Walsh et al., 1981; World Health
359 Organization and Onchocerciasis Control Programme in the Volta River Basin Area, 1985). This combination
360 of distance, wind direction, and *Simulium* spp. flight capabilities likely restricts Ghana/Mali and Ghana/Côte
361 d'Ivoire vector movement while facilitating Côte d'Ivoire/Mali migration, and hence also restricts parasite
362 gene flow between Ghana and Cote d'Ivoire and Ghana and Mali, accounting for the observed population
363 structure.

364 Thus, the parasite population structure we found across the sampling locations in the three
365 countries is consistent with known vector migration patterns and *Simulium* spp. flight capabilities. It is
366 noteworthy that the lack of mitochondrial population structure between Côte d'Ivoire and Mali we found is
367 in contrast to the O-150 population differences between the ecotypes of the sampling locations – forest in

Côte d'Ivoire and savanna in Mali (Duke et al., 1966), although we cannot exclude the possibility that diversity levels may have been underestimated due to low sampling size.

We found limited genetic differentiation among parasites from three river basins along as east to west transect of the savannah transition zone in Ghana, suggesting ongoing gene flow and hence parasite transmission, not only within but also between the three river basins sampled. The two furthest river basins (BVT and Daka) are separated by less than the maximum flight range of *Simulium* spp., suggesting that vector movement between river basins is sufficient to maintain parasite transmission and gene flow at this spatial scale. This in turn implies that long-term epidemiological patterns across the three river basins are determined by long-range transmission events rather than short-range focal transmission centred around breeding sites within a given river basin. Movement of infected humans could, of course, also contribute to parasite interbreeding (both within and across countries).

There are some notable exceptions to the lack of population structure between the sampled villages in Ghana. Parasites from inhabitants of Asubende (Pru River basin), Wiae (Daka River Basin) and Kojobone (Daka River Basin) are somewhat differentiated from those obtained from inhabitants of surrounding communities. Although not statistically significant, this weak genetic differentiation may be noteworthy because Asubende is one of the villages where sub-optimal response to ivermectin was first detected phenotypically (Awadzi et al., 2004; Frempong et al., 2016; Osei-Atweneboana et al., 2007). Parasites from Asubende were genetically characterized previously (Doyle et al., 2017), and those with a sub-optimal response phenotype were genetically distinct in nuclear genome comparisons, indicating that the unusual genetic structure of the parasite population from inhabitants of Asubende is not restricted to the mitochondrial genome. We suggest that the unusual genetic profile of Asubende parasites also influences the accuracy of parasite assignment to the Pru basin (which is poor relative to assignment to the Daka and Black Volta basins).

4.3 Implications of Population Structure on Delineation of Transmission Zones

A significant advantage of population genetic analysis is that population genetic parameters are an objective and quantitative measure of past transmission between any two locations. Large values for measures of population structure (such as the φ_{PT} statistic for mitochondrial data or the F_{ST} statistic for

nuclear genotypic data) imply restricted gene flow (i.e. restricted interbreeding) and hence little or no history of transmission between those locations. Conversely, values for measures of population structure that aren't statistically significantly different from zero, such as those observed between locations in 3 river basins in central Ghana (Tables 2, 3), imply unrestricted gene flow and hence a history of high transmission between river basins. In other words, population genetic structure is the product of restrictions on gene flow. In parasitological terms, gene flow can only occur when there is parasite transmission between two locations. Thus, the population structure reported here is a measure of the extent of transmission between the geographic areas where the parasites we analysed were collected. Consequently, our data suggest that parasite transmission in this part of West Africa occurs across large distances, consistent with known patterns of vector migration.

The observation that these data are consistent with an isolation-by-distance (IBD) model for parasite population structure and hence parasite transmission is important. Isolation-by-distance population structure implies a continuum of gene flow, and hence implies also a continuum of onchocerciasis transmission, such that locations that are relatively closer together will experience more transmission between them than locations further apart. The important feature of this analysis is the distance and direction over which this transmission occurs. Our analysis indicates that epidemiologically relevant transmission occurs over scales of at least 100 – 200 km, and may occur over greater distances under conducive wind conditions. This latter observation is consistent with the known behaviour of *S. damnosum* in West Africa, but the demonstration of epidemiologically relevant, east-west transmission between river basins in Ghana is novel and has significant implications for the onchocerciasis elimination.

4.4 Implications of Population Structure for Decisions to Stop Interventions

A clear implication of our work is that the transmission zones defined by the parasite genetic structure will very likely be much larger than 'MDAi project areas' or the focal area surrounding a vector breeding site (the WHO 2016 elimination guideline definition of an epidemiologically relevant transmission zone). For the area covered in this study (central Ghana, Mali and Côte d'Ivoire), transmission zones are measured in hundreds of kilometres and extend over several river basins. It is these large, population genetics-defined transmission zones which should be the units for assessing progress towards elimination

and criteria for stopping treatment, and for defining the extent of post-treatment cessation surveillance and estimating the risk of post-treatment cessation recrudescence (Hedtke et al., submitted).

Population genetics-defined transmission zones are based on quantitative measures, which can be used to derive estimates of the risk of post-treatment recrudescence due to immigration from other areas where criteria for stopping treatment have not yet been met. This risk estimate can be considered by the onchocerciasis elimination programs when making decisions to stop treatment in a specific geographic area. If the acceptable risk level is very low, our data would delineate two large transmission zones (North-South between Mali and Côte d'Ivoire, and East-West within Ghana) with little transmission between them. Conversely, cessation of treatment in one location in a single transmission zone while transmission continues in other locations within the same zone would carry a much higher risk, proportional to the extent of transmission between locations. For example, the lack of genetic structure between parasites sampled from three river basins in Ghana, and thus the high probability of transmission between these three river basins, indicates that a very high risk of recrudescence would have to be acceptable for treatment to be stopped in one river basin, even if the prevalence of infective (or infected) vectors meets WHO guideline criteria for stopping treatment, as long as the prevalence in at least one of the other basins does not. Our data illustrate that for elimination to be sustainable, transmission zones cannot be defined by river basin (or by MDAi project area or human infection prevalence patterns), and decisions on stopping treatment may be risky without population genetic data on the extent of interbreeding among river basins/MDAi project areas/REMO endemicity areas, including areas that are geographically very distant.

The current *O. volvulus* transmission models (EpiOncho and OnchoSim) have been used to predict the time to elimination for different levels of pre-intervention endemicity and duration and therapeutic coverage of MDAi (Basáñez et al., 2016; Stolk et al., 2015), and EpiOncho was used to model the time to measurable levels of recrudescence after cessation of MDAi in areas in Senegal and Mali under a range of assumptions concerning residual autochthonous transmission (Walker et al., 2017). Neither model is currently able to model spatial heterogeneity of endemicity or the impact of migration of a defined number of infected/infective vectors or infected people between different areas, i.e., heterochthonous transmission due to long-range movement. We are working to modify EpiOncho to include quantitative

449 population genetic measures and to allow simulation of different transmission zones (McCulloch et al.,
450 2017). Once available and validated, and provided the geographically relevant population genetic data are
451 available, such a model will allow to quantitate the risk associated with stopping interventions in one
452 geographic area meeting WHO guideline criteria when 'neighbouring' areas do not.

453 *4.5 The utility of population genetics for identifying the origin of parasites detected during post-treatment*
454 *and post-elimination surveillance.*

455 We have shown (Figure 5 A-C, Table 5) that the population genetic data permit identification of the
456 likely geographical origin of a single parasite. In the Comoé river valley in southwest Burkina Faso, where
457 transmission had been interrupted when the OCP ceased larviciding in 1989, *O. volvulus* infection
458 prevalence rates of up to 71% were identified in 2010-2011 and biannual MDAi was implemented.
459 Recrudescence could be attributable in some areas to long-range migration of infected flies from Côte
460 d'Ivoire, although recrudescence due to undetected infections in the Comoé river valley was not excluded
461 (Koala et al., 2017; Koala et al., 2019). With the methods established here, analysis of parasites from
462 Burkina Faso obtained before and after stopping of interventions could establish whether
463 infections/transmission detected now are due to parasites that were missed when decisions to stop
464 larviciding were made, or due to migration of parasites from other endemic regions via vectors or humans.
465 If these analyses indicate a migratory origin, analysis of parasites from potential source areas could
466 furthermore help identify the source and appropriate cost-effective strategies for addressing transmission
467 in the Comoé river valley.

468 *4.6 Conclusions*

469 As African onchocerciasis control programs move towards elimination of *O. volvulus* transmission,
470 they need tools for delineating transmission zones throughout which interventions can be stopped, with a
471 defined risk of resurgence they consider acceptable. We have examined the whole mitochondrial genome
472 sequences of 189 adult nematodes to investigate genetic diversity, demographic history and population
473 structure at three spatial scales in West Africa.

474 The *O. volvulus* mitochondrial genome was found to be more diverse than previously described,
475 with a demographic history consistent with rapid population expansion after a bottleneck likely associated

476 with host switching from cattle to humans. The data support delineation of (at least) two large transmission
477 zones (North-South between Mali and Côte d'Ivoire, and East-West within Ghana). The boundaries
478 between these large transmission zones, and the probability of transmission between any two points within
479 each zone, can be defined by population genetic measures which are the product of historical patterns of
480 transmission and are objective and quantitative. The population genetic parameters required to define a
481 transmission zone can be estimated cheaply and efficiently from genotyping of parasites caught during
482 entomological evaluations with appropriate sampling. The genetic polymorphisms obtained may be used to
483 estimate the risks of resurgence associated with stopping interventions in areas where stopping criteria are
484 met, when they are not met in other geographic areas, and to determine the origin of suspected post-
485 intervention recrudescence. Population genetic measures can therefore be the tools elimination programs
486 need to inform their decisions about cessation of MDAi and/or the frequency and strategy for post-MDAi
487 surveillance.

488 The data presented here are one of the largest samples of high-resolution population data for any
489 helminth, but are restricted to a relatively small segment of the sub-Saharan range of onchocerciasis. The
490 methods we used can be readily applied to identify *O. volvulus* genetic variation and population structure
491 across Africa to provide additional data to elimination programs facing the decision on where and when
492 stop interventions.

493

494 **Acknowledgements**

495 The authors wish to thank Andrew Robinson for his assistance with data analysis using the LIMS High
496 Performance Computing Cluster, and the members of the Grant lab for their assistance and support with
497 laboratory experiments and data analysis.
498 Funding: This work was supported by TDR, the UNICEF/UNDP/World Bank/WHO Special Programme for
499 Research and Training in Tropical Diseases, and an Illumina MiSeq grant to SRD. KEC was supported by an
500 Australian Government Research Training Program (RTP) Scholarship.

501

502

Tables

Table 1. Summary statistics including haplotype and nucleotide diversity, and neutrality tests by country and river basin in Ghana. N: number of samples, S: polymorphic sites, H: number of haplotypes, H_d: haplotype diversity, π: nucleotide diversity.

Location	N	S (average/ sample)	H	H_d (±SD)	π (±SD)	Tajima's D	Fu's F	Fu & Li's D
Côte d'Ivoire	13	30 (2.31)	11	0.962 (0.050)	0.0006 (0.0001)	-0.605	-2.604	-0.081
Mali	10	41 (4.10)	9	0.978 (0.054)	0.0008 (0.0001)	-1.327	-1.506	-0.994
Ghana	166	174 (1.05)	139	0.997 (0.001)	0.0006 (0.0000)	-2.283*	-24.431*	1.538
River basin								
Black Volta/Tombe	74	136 (1.84)	70	0.998 (0.003)	0.0007 (0.0000)	-2.336**	-24.697**	-2.960
Daka	22	56 (2.55)	20	0.991 (0.017)	0.0006 (0.0001)	-1.758	-10.345**	-1.956
Pru	70	123 (1.76)	60	0.995 (0.004)	0.0006 (0.0000)	-2.281**	-24.750**	-2.584

* $p < 0.01$

** $p < 0.001$

Table 2. Result of pairwise comparisons of genetic differentiation (φ_{PT}) between parasites from Ghana, Côte d'Ivoire and Mali and parasites from three river basins in Ghana (in italics)

Location	Ghana				Côte d'Ivoire
		Black Volta/Tombe	*Pru*	*Daka*	
Ghana					
Pru		0.0027			
Daka		0.0005	0.0031		
Côte d'Ivoire	0.0180***				
		0.0188***	0.0230***	0.0175*	
Mali	0.0112*				
		0.0097*	0.0150	0.0130***	0.0000

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

Table 3. Results of pairwise comparisons of genetic sequence variation (φ_{PT}) of between parasites from the inhabitants of 15 villages in Ghana.

Village	AB1	AB2	BUI	KYG	NLG	NYR	ASU	OHP	SEN	CHA	GHA	JAG	KOJ	TAK
AB2	0													
BUI	0	0												
KYG	0	0	0											
NLG	0	0	0	0										
NYR	0	0	0	0	0									
ASU	0.004	0.008	0	0.009	0.009	0.003								
OHP	0	0	0	0	0	0	0.011							
SEN	0	0	0	0	0	0	0.012	0						
CHA	0	0	0	0	0	0	0.004	0	0					
GHA	0	0	0	0	0	0	0.012	0	0	0				
JAG	0	0	0.003	0.0002	0	0	0.008	0.003	0	0	0			
KOJ	0.006	0.006	0.007	0.006	0.006	0.007	0.015	0.008	0.007	0.006	0	0.006		
TAK	0	0	0	0	0	0	0.012	0	0	0	0	0.003	0.009	
WIA	0.013	0.013	0.015	0.006	0.014	0.014	0.022	0.017	0.020	0.013	0.020	0.008	0.012	0.020

Figure Legends

Figure 1. Rapid epidemiological mapping of onchocerciasis (REMO) based on prevalence of nodules across communities with a history of onchocerciasis and covered by APOC. Figure from African Programme for Onchocerciasis Control and World Health Organization (2013).

Figure 2. Sampling locations for *O. volvulus* collected. A) Sampling across the African countries of Mali (yellow), Côte d'Ivoire (mauve), and Ghana (green). B) Sampling across the Black Volta/Tombe (red), Pru (purple), and Daka (blue) river basins within Ghana.

Figure 3. Design for amplification of *Onchocerca volvulus* mitochondrial genome using three possible primer combinations.

Figure 4. Haplotype networks based on whole mitochondrial genome sequences. A) Haplotype network for *O. volvulus* sampled in three different countries in West Africa: Mali, Côte d'Ivoire, and Ghana; B) Haplotype network for *O. volvulus* sampled across three river basins within Ghana

Figure 5. Discriminant analysis of principle components. Top: A) DAPC analysis by country. B) Membership probabilities for parasites from Ghana, Côte d'Ivoire and Mali. C) Percentage of parasites correctly assigned to country of origin. Bottom: D) DAPC analysis by river basins in Ghana. E) Membership probabilities for parasites from river basins in Ghana. F) Percentage of Ghanaian parasites correctly assigned to Ghanaian river basin of origin.

Figure 6. Changes in effective population size over time among West African *Onchocerca volvulus* represented by a Bayesian skyline plot under a coalescent model, assuming one generation per year. Blue line indicates the median estimates, red lines represent the 95% confidence intervals.

References

African Programme for Onchocerciasis, C., World Health Organization, 2013. Conceptual and operational framework of onchocerciasis elimination with ivermectin treatment. World Health Organization, Geneva.

Archie, E.A., Luikart, G., Ezenwa, V.O., 2009. Infecting epidemiology with genetics: a new frontier in disease ecology. Trends Ecol Evol 24, 21-30.

Ardelli, B.F., Prichard, R.K., 2004. Identification of variant ABC-transporter genes among *Onchocerca volvulus* collected from ivermectin-treated and untreated patients in Ghana, West Africa. Ann Trop Med Parasitol 98, 371-384.

Ardelli, B.F., Prichard, R.K., 2007. Reduced genetic variation of an *Onchocerca volvulus* ABC transporter gene following treatment with ivermectin. Trans R Soc Trop Med Hyg 101, 1223-1232.

Avise, J., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu Rev Ecol Syst 18, 489-522.

Awadzi, K., Boakye, D.A., Edwards, G., Opoku, N.O., Attah, S.K., Osei-Atweneboana, M.Y., Lazdins-Helds, J.K., Ardrey, A.E., Addy, E.T., Quartey, B.T., Ahmed, K., Boatin, B.A., Soumbey-Alley, E.W., 2004. An investigation of persistent microfilaridermias despite multiple treatments with ivermectin, in two onchocerciasis-endemic foci in Ghana. Ann Trop Med Parasitol 98, 231-249.

Bain, O., 1981. [Species of the genus Onchocerca and primarily O. volvulus, considered from an epidemiologic and phylogenetic point of view]. Ann Soc Belg Med Trop 61, 225-231.

Baker, R.H., Guillet, P., Seketeli, A., Poudiougo, P., Boakye, D., Wilson, M.D., Bissan, Y., 1990. Progress in controlling the reinvasion of windborne vectors into the western area of the Onchocerciasis Control Programme in West Africa. Philos Trans R Soc Lond B Biol Sci 328, 731-747, discussion 747-750.

Barrett, L.G., Thrall, P.H., Burdon, J.J., Linde, C.C., 2008. Life history determines genetic structure and evolutionary potential of host-parasite interactions. Trends Ecol Evol 23, 678-685.

Basáñez, M.G., Walker, M., Turner, H.C., Coffeng, L.E., de Vlas, S.J., Stolk, W.A., 2016. River blindness: mathematical models for control and elimination. Adv Parasitol 94, 247-341.

Blouin, M.S., Yowell, C.A., Courtney, C.H., Dame, J.B., 1995. Host movement and the genetic structure of populations of parasitic nematodes. Genetics 141, 1007-1014.

Boakye, D., Tallant, J., Adjami, A., Moussa, S., Tekle, A., Robalo, M., Rebollo, M., Mwinza, P., Sitima, L., Cantey, P., Mackenzie, C., 2018. Refocusing vector assessment towards the elimination of onchocerciasis from Africa: a review of the current status in selected countries. Int Health 10, i27-i32.

Boatin, B., Molyneux, D.H., Hougard, J.M., Christensen, O.W., Alley, E.S., Yameogo, L., Seketeli, A., Dadzie, K.Y., 1997. Patterns of epidemiology and control of onchocerciasis in west Africa. J Helminthol 71, 91-101.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10, e1003537.

Brown, J.M., Hedtke, S.M., Lemmon, A.R., Lemmon, E.M., 2010. When trees grow too long: investigating the causes of highly inaccurate bayesian branch-length estimates. Syst Biol 59, 145-161.

Bush, S., Sodahlon, Y., Downs, P., Mackenzie, C.D., 2018. Cross-border issues: an important component of onchocerciasis elimination programmes. Int Health 10, i54-i59.

Choi, Y.J., Tyagi, R., McNulty, S.N., Rosa, B.A., Ozersky, P., Martin, J., Hallsworth-Pepin, K., Unnasch, T.R., Norice, C.T., Nutman, T.B., Weil, G.J., Fischer, P.U., Mitreva, M., 2016. Genomic diversity in *Onchocerca volvulus* and its *Wolbachia* endosymbiont. Nat Microbiol 2, 16207.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6, 80-92.

Crainey, J.L., Silva, T.R., Encinas, F., Marin, M.A., Vicente, A.C., Luz, S.L., 2016. The mitogenome of *Onchocerca volvulus* from the Brazilian Amazonia focus. Mem Inst Oswaldo Cruz 111, 79-81.

Criscione, C.D., Blouin, M.S., 2004. Life cycles shape parasite evolution: comparative population genetics of salmon trematodes. Evolution 58, 198-202.

Criscione, C.D., Poulin, R., Blouin, M.S., 2005. Molecular ecology of parasites: elucidating ecological and microevolutionary processes. Mol Ecol 14, 2247-2257.

Dadzie, Y., Amazigo, U.V., Boatin, B.A., Seketeli, A., 2018. Is onchocerciasis elimination in Africa feasible by 2025: a perspective based on lessons learnt from the African control programmes. Infect Dis Poverty 7, 63.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Genomes Project Analysis, G., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156-2158.

De Sole, G., Remme, J., Awadzi, K., Accorsi, S., Alley, E.S., Ba, O., Dadzie, K.Y., Giese, J., Karam, M., Keita, F.M., 1989. Adverse reactions after large-scale treatment of onchocerciasis with ivermectin: combined results from eight community trials. Bull World Health Organ 67, 707-719.

Denver, D.R., Morris, K., Lynch, M., Vassilieva, L.L., Thomas, W.K., 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. Science 289, 2342-2344.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43, 491-498.

Diawara, L., Traore, M.O., Badji, A., Bissan, Y., Doumbia, K., Goita, S.F., Konate, L., Mounkoro, K., Sarr, M.D., Seck, A.F., Toe, L., Touree, S., Remme, J.H., 2009. Feasibility of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: first evidence from studies in Mali and Senegal. PLoS Negl Trop Dis 3, e497.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., Mattick, J.S., 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19, 4008.

Dowdle, W.R., 1998. The principles of disease elimination and eradication. Bull World Health Organ 76, 22-25.

Doyle, S.R., Bourguinat, C., Nana-Djeunga, H.C., Kengne-Ouafo, J.A., Pion, S.D.S., Bopda, J., Kamgno, J., Wanji, S., Che, H., Kuesel, A.C., Walker, M., Basáñez, M.G., Boakye, D.A., Osei-Atweneboana, M.Y., Boussinesq, M., Prichard, R.K., Grant, W.N., 2017. Genome-wide analysis of ivermectin response by *Onchocerca volvulus* reveals that genetic drift and soft selective sweeps contribute to loss of drug sensitivity. PLoS Negl Trop Dis 11, e0005816.

Duke, B.O.L., Lewis, D.J., Moore, P.J., 1966. *Onchocerca-Simulium* complexes. I. Transmission of forest and Sudan savanna strains of *Onchocerca volvulus*, from Cameroon, by *Simulium damnosum* from various West African bioclimatic zones. Ann Trop Med Parasitol 60, 318-336.

Excoffier, L., Lischer, H.E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10, 564-567.

Fobi, G., Yameogo, L., Noma, M., Aholou, Y., Koroma, J.B., Zoure, H.M., Ukety, T., Lusamba-Dikassa, P.S., Mwikisa, C., Boakye, D.A., Roungou, J.B., 2015. Managing the fight against onchocerciasis in Africa: APOC experience. PLoS Negl Trop Dis 9, e0003542.

Frempong, K.K., Walker, M., Cheke, R.A., Tetevi, E.J., Gyan, E.T., Owusu, E.O., Wilson, M.D., Boakye, D.A., Taylor, M.J., Biritwum, N.K., Osei-Atweneboana, M., Basáñez, M.G., 2016. Does increasing treatment frequency address suboptimal responses to ivermectin for the control and elimination of River blindness? Clin Infect Dis 62, 1338-1347.

Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147, 915-925.

Fu, Y.X., Li, W.H., 1993. Maximum likelihood estimation of population parameters. Genetics 134, 1261-1270.

Garms, R., 1981. The reinvasion of the onchocerciasis control programme area in the Volta River Basin by *Simulium damnosum* S.L., the involvement of the different cytospecies and epidemiological implications. Ann Soc Belg Med Trop 61, 193-198.

Garms, R., Cheke, R.A., Vajime, C., Sowah, S., 1982. The occurrence and movements of different members of the *Simulium damnosum* complex in Togo and Benin. Z Für Angew Zool 69, 219-236.

Garms, R., Ochoa, J.O., 1979. Further studies on the relative importance of Guatemalan blackfly species as vectors of *Onchocerca volvulus*. Tropenmed Parasitol 30, 120-128.

Garms, R., Walsh, J.F., Davies, J.B., 1979. Studies on the reinvasion of the Onchocerciasis Control Programme in the Volta River Basin by *Simulium damnosum* s.I. with emphasis on the south-western areas. Tropenmed Parasitol 30, 345-362.

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

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307-321.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22, 160-174.

Hedtke, S.M., Kuesel, A.C., Crawford, K.E., Graves, P.M., Boussinesq, M., Lau, C.L., Boakye, D.A., Grant, W.N., submitted. Genomic epidemiology in filarial nematodes: transforming the basis for elimination program decisions. Front Genet.

Ingman, M., Kaessmann, H., Paabo, S., Gyllensten, U., 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408, 708-713.

Jarne, P., Théron, A., 2001. Genetic structure in natural populations of flukes and snails: a practical approach and review. Parasitology 123 Suppl, S27-40.

Johnson, C.G., Walsh, J.F., Davies, J.B., Clark, S.J., Perry, J.N., 1985. The pattern and speed of displacement of females of *Simulium damnosum* Theobald s.l. (Diptera: Simuliidae) across the Onchocerciasis Control Programme area of West Africa in 1977 and 1978. Bull Entomol Res 75, 73-92.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403-1405.

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070-3071.

Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11, 94.

Keddie, E.M., Higazi, T., Unnasch, T.R., 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. Mol Biochem Parasitol 95, 111-127.

Koala, L., Nikiema, A., Post, R.J., Paré, A.B., Kafando, C.M., Drabo, F., Traoré, S., 2017. Recrudescence of onchocerciasis in the Comoé valley in Southwest Burkina Faso. Acta Trop 166, 96-105.

Koala, L., Nikiema, A.S., Pare, A.B., Drabo, F., Toe, L.D., Belem, A.M.G., Boakye, D.A., Traore, S., Dabire, R.K., 2019. Entomological assessment of the transmission following recrudescence of onchocerciasis in the Comoe Valley, Burkina Faso. Parasit Vectors 12, 34.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol 34, 772-773.

Lefoulon, E., Giannelli, A., Makepeace, B.L., Mutafchiev, Y., Townson, S., Uni, S., Verocai, G.G., Otranto, D., Martin, C., 2017. Whence river blindness? The domestication of mammals and host-parasite co-evolution in the nematode genus *Onchocerca*. Int J Parasitol 47, 457-470.

Leigh, J.W., Bryant, D., 2015. PopART: Full-feature software for haplotype network construction. Methods Ecol Evol 6, 1110-1116.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987-2993.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451-1452.

Lischer, H.E., Excoffier, L., 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298-299.

Magor, J.I., Rosenberg, L.J., 1980. Studies of winds and weather during migrations of *Simulium damnosum* Theobald (Diptera: Simuliidae), the vector of onchocerciasis in West Africa. Bull Entomol Res 70, 693-716.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res 27, 209-220.

Marshall, F., Hildebrand, E., 2002. Cattle before crops: the beginnings of food production in Africa. J World Prehistory 16, 99-143.

McCulloch, K., McCaw, J., McVernon, J., Hedtke, S.M., Walker, M., Milton, P., Basáñez, M.-G., Grant, W.N., 2017. Investigation into the effect of host migration on the transmission of *Onchocerca volvulus* using a patch model, American Journal of Tropical Medicine and Hygiene. Amer Soc Trop Med & Hygiene, Baltimore, Maryland, p. 564.

Nadler, S.A., 1995. Microevolution and the genetic structure of parasite populations. J Parasitol 81, 395-403.

Noma, M., Nwoke, B.E., Nutall, I., Tambala, P.A., Enyong, P., Namsenmo, A., Remme, J., Amazigo, U.V., Kale, O.O., Seketeli, A., 2002. Rapid epidemiological mapping of onchocerciasis (REMO): its application by the African Programme for Onchocerciasis Control (APOC). Ann Trop Med Parasitol 96 Suppl 1, S29-39.

Osei-Atweneboana, M.Y., Eng, J.K., Boakye, D.A., Gyapong, J.O., Prichard, R.K., 2007. Prevalence and intensity of *Onchocerca volvulus* infection and efficacy of ivermectin in endemic communities in Ghana: a two-phase epidemiological study. Lancet 369, 2021-2029.

Post, R.J., Cheke, R.A., Boakye, D.A., Wilson, M.D., Osei-Atweneboana, M.Y., Tetteh-Kumah, A., Lamberton, P.H., Crainey, J.L., Yameogo, L., Basáñez, M.G., 2013. Stability and change in the distribution of cytospecies of the *Simulium damnosum* complex (Diptera: Simuliidae) in southern Ghana from 1971 to 2011. Parasit Vectors 6, 205.

Prost, A., Hervouet, J.P., Thylefors, B., 1979. Epidemiologic status of onchocerciasis. Bull World Health Organ 57, 655-662.

Prugnolle, F., Liu, H., de Meeus, T., Balloux, F., 2005. Population genetics of complex life-cycle parasites: an illustration with trematodes. Int J Parasitol 35, 255-263.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67, 901-904.

Real, L.A., Biek, R., 2007. Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. J R Soc Interface 4, 935-948.

Rebollo, M.P., Zoure, H., Ogoussan, K., Sodahlon, Y., Ottesen, E.A., Cantey, P.T., 2018. Onchocerciasis: shifting the target from control to elimination requires a new first-step-elimination mapping. Int Health 10, i14-i19.

Remme, J.H.F., Boatin, B., Boussinesq, M., 2017. Helminthic diseases: onchocerciasis and loiasis, in: Quah, S.R., Cockerham, W.C. (Eds.), The International Encyclopedia of Public Health, 2 ed, pp. 576-587.

Schulz-Key, H., 1988. The collagenase technique: how to isolate and examine adult *Onchocerca volvulus* for the evaluation of drug effects. Trop Med Parasitol 39 Suppl 4, 423-440.

Schwabl, P., Llewellyn, M.S., Landguth, E.L., Andersson, B., Kitron, U., Costales, J.A., Ocana, S., Grijalva, M.J., 2017. Prediction and prevention of parasitic diseases using a landscape genomics framework. Trends Parasitol 33, 264-275.

Slatkin, M., 1991. Inbreeding coefficients and coalescence times. Genet Res 58, 167-175.

Stamatakis, A., 2016. The RAxML v8.2.X manual, in: Studies, H.I.f.T. (Ed.), Heidelberg, Germany.

Stolk, W.A., Walker, M., Coffeng, L.E., Basáñez, M.G., de Vlas, S.J., 2015. Required duration of mass ivermectin treatment for onchocerciasis elimination in Africa: a comparative modelling analysis. Parasit Vectors 8, 552.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585-595.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10, 512-526.

Tekle, A.H., Elhassan, E., Isiyaku, S., Amazigo, U.V., Bush, S., Noma, M., Cousens, S., Abiose, A., Remme, J.H., 2012. Impact of long-term treatment of onchocerciasis with ivermectin in Kaduna State, Nigeria: first evidence of the potential for elimination in the operational area of the African Programme for Onchocerciasis Control. Parasit Vectors 5, 28.

Tekle, A.H., Zoure, H.G., Noma, M., Boussinesq, M., Coffeng, L.E., Stolk, W.A., Remme, J.H., 2016. Progress towards onchocerciasis elimination in the participating countries of the African Programme for Onchocerciasis Control: epidemiological evaluation results. Infect Dis Poverty 5, 66.

Torroni, A., Achilli, A., Macaulay, V., Richards, M., Bandelt, H.J., 2006. Harvesting the fruit of the human mtDNA tree. Trends Genet 22, 339-345.

Traore, M.O., Sarr, M.D., Badji, A., Bissan, Y., Diawara, L., Doumbia, K., Goita, S.F., Konate, L., Mounkoro, K., Seck, A.F., Toe, L., Toure, S., Remme, J.H., 2012. Proof-of-principle of onchocerciasis elimination with ivermectin treatment in endemic foci in Africa: final results of a study in Mali and Senegal. PLoS Negl Trop Dis 6, e1825.

UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases, 1992. Methods for community diagnosis of onchocerciasis to guide ivermectin based control in Africa: report of an informal consultation held in Ouagadougou from 19-21 November 1991. World Health Organization, Geneva.

Unnasch, T.R., Golden, A., Cama, V., Cantey, P.T., 2018. Diagnostics for onchocerciasis in the era of elimination. Int Health 10, i20-i26.

Wakeley, J., Nielsen, R., Liu-Cordero, S.N., Ardlie, K., 2001. The discovery of single-nucleotide polymorphisms--and inferences about human demographic history. Am J Hum Genet 69, 1332-1347.

Walker, M., Stolk, W.A., Dixon, M.A., Bottomley, C., Diawara, L., Traore, M.O., de Vlas, S.J., Basáñez, M.G., 2017. Modelling the elimination of river blindness using long-term epidemiological and programmatic data from Mali and Senegal. Epidemics 18, 4-15.

Walsh, J.F., Davies, J.B., Garms, R., 1981. Further studies on the reinvasion of the onchocerciasis control programme by *Simulium damnosum* s.l.: the effects of an extension of control activities into southern Ivory Coast during 1979. Tropenmed Parasitol 32, 269-273.

WHO Expert Committee on Onchocerciasis Control (1993 : Geneva, S., World Health Organization, 1995. Onchocerciasis and its control : report of a WHO Expert Committee on Onchocerciasis Control. World Health Organization, Geneva.

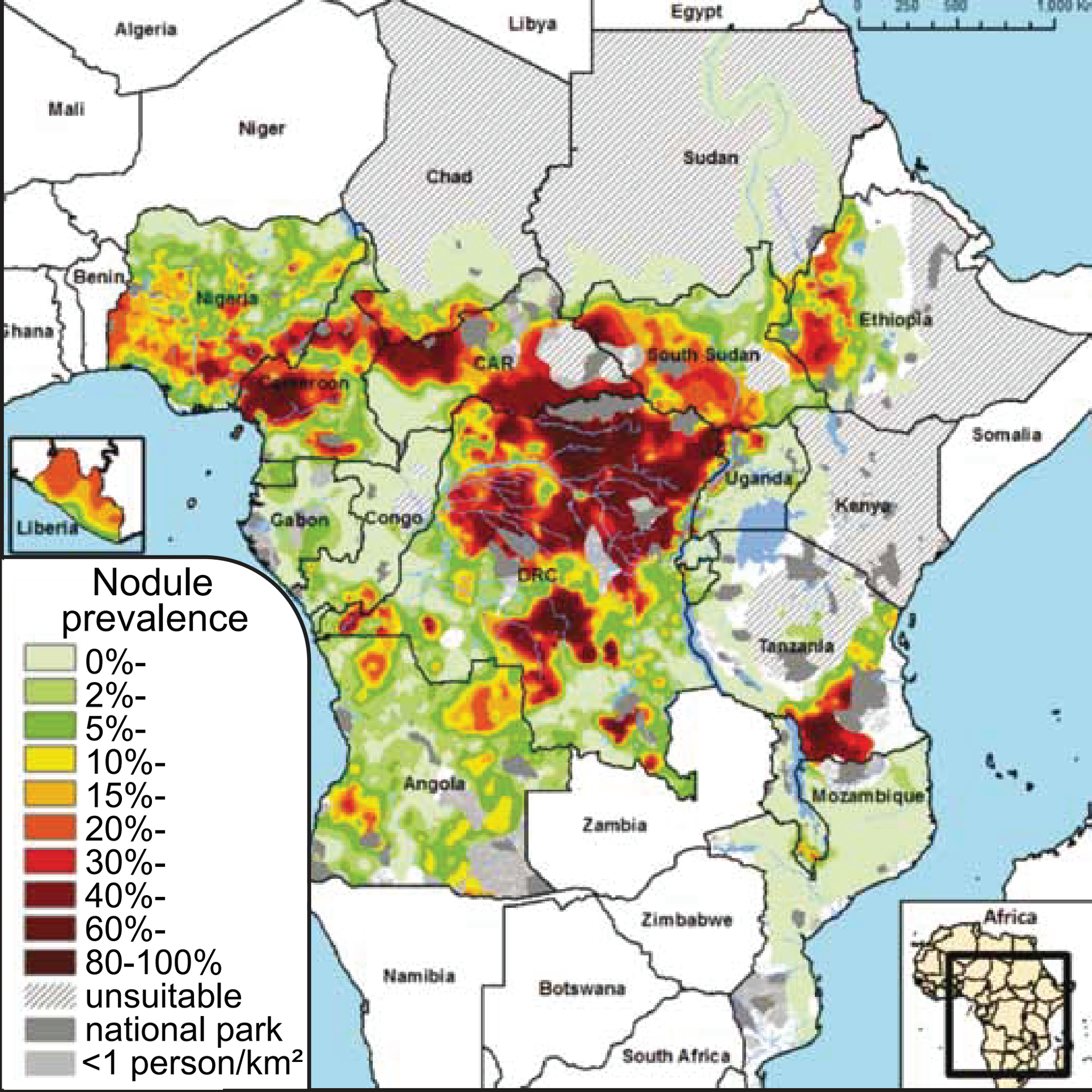
World Health Organization, 2016. Guidelines for stopping mass drug administration and verifying elimination of human onchocerciasis: criteria and procedures. World Health Organization, Geneva.

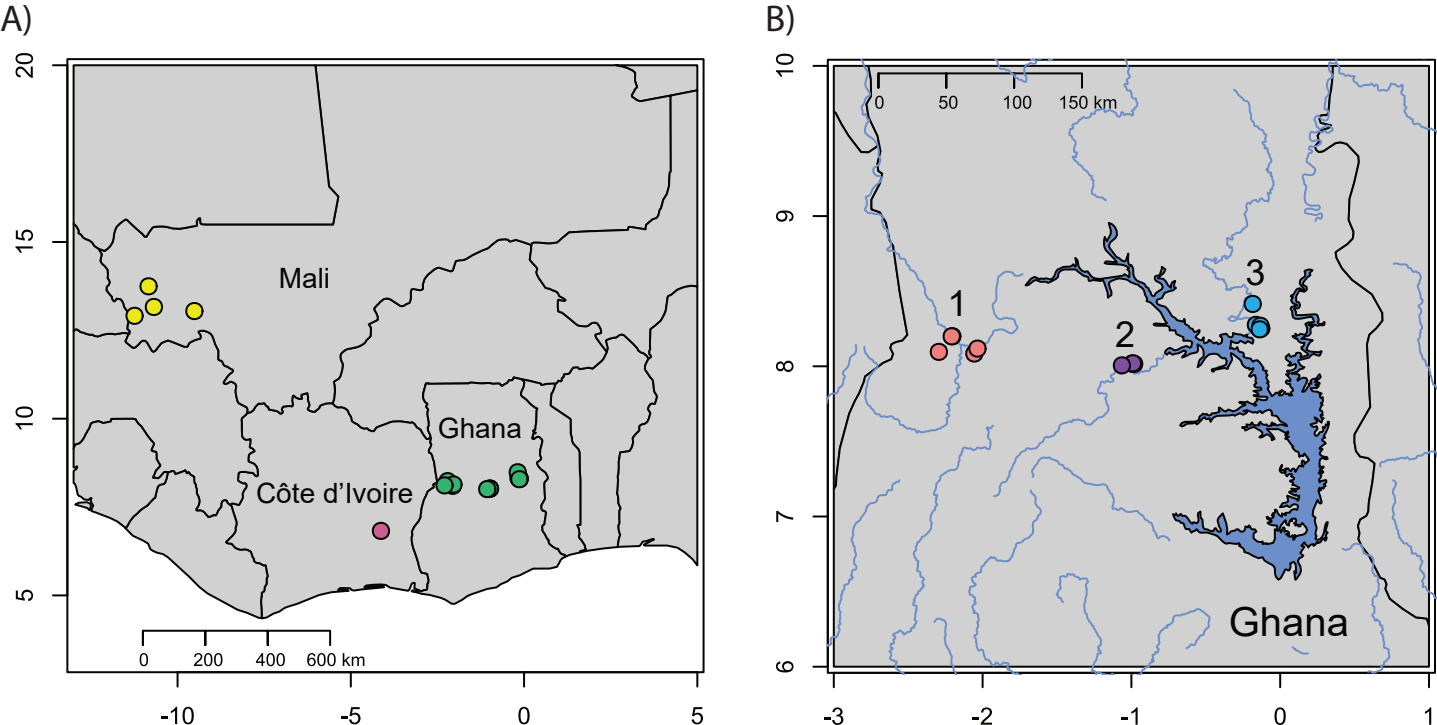
World Health Organization, 2017. Progress report on the elimination of human onchocerciasis, 2016-2017. Weekly Epidemiological Record 45, 681-700.

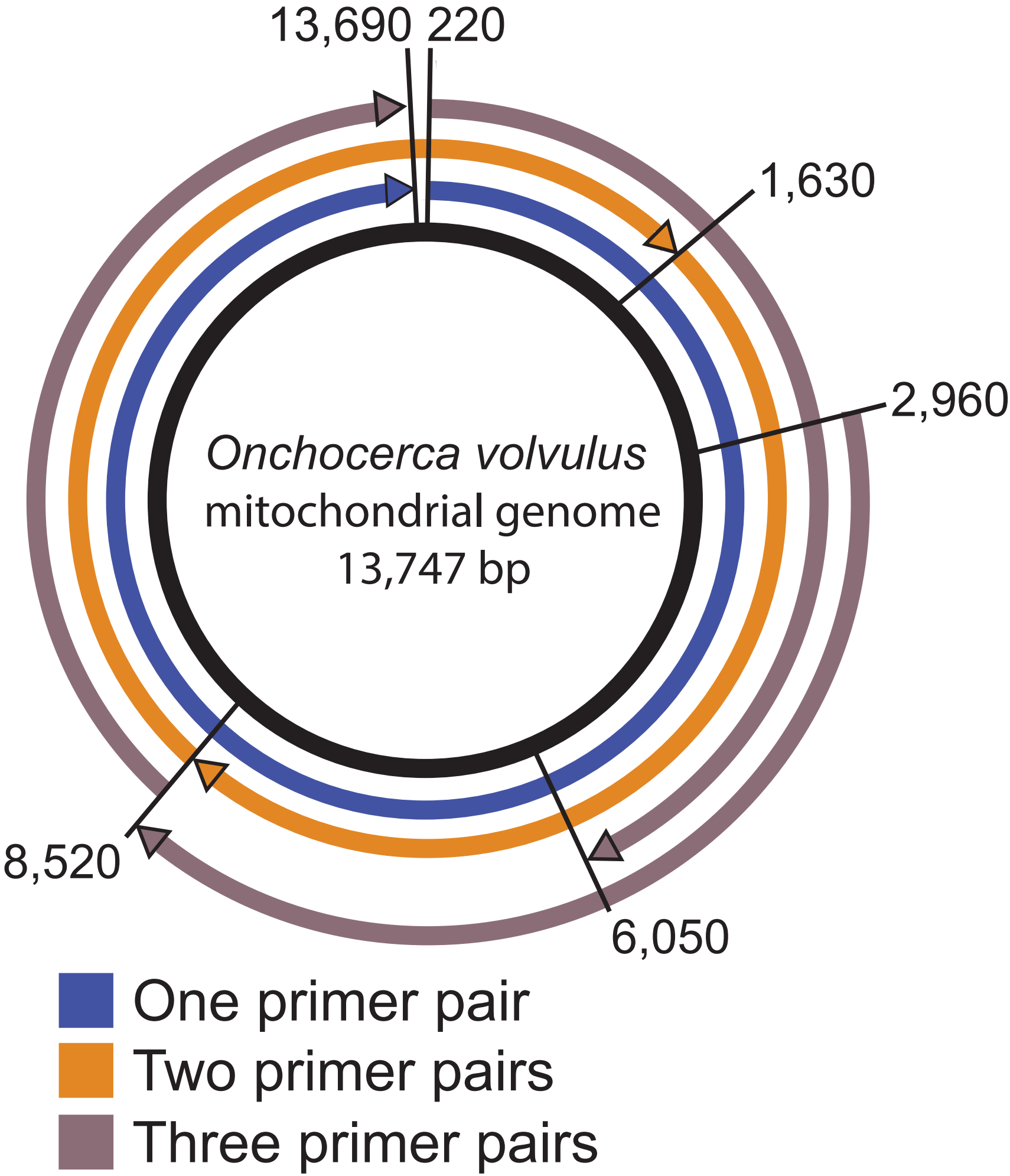
World Health Organization, Onchocerciasis Control Programme in the Volta River Basin Area, 1985. 10 années de lutte contre l'onchocercose en Afrique de l'Ouest: bilan des activités du Programme de lutte contre l'onchocercose dans la région du Bassin de la Volta de 1974 à 1984. World Health Organization, Geneva, Switzerland.

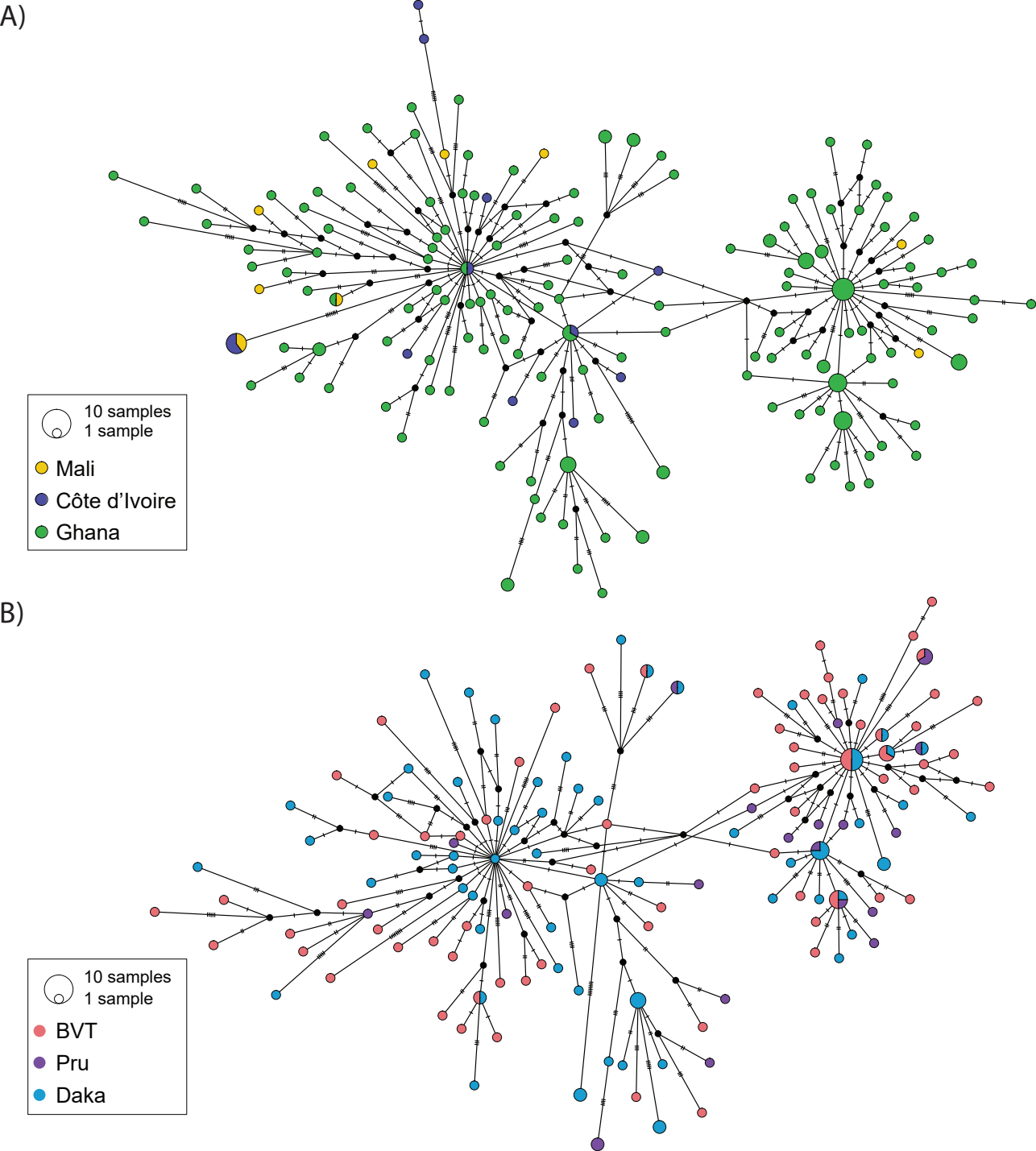
World Health Organization, Onchocerciasis Control Programme in West Africa, 1997. Twenty years of onchocerciasis control in West Africa: review of the work of the Onchocerciasis Control Programme in West Africa from 1974 to 1994. World Health Organization, Geneva, Switzerland.

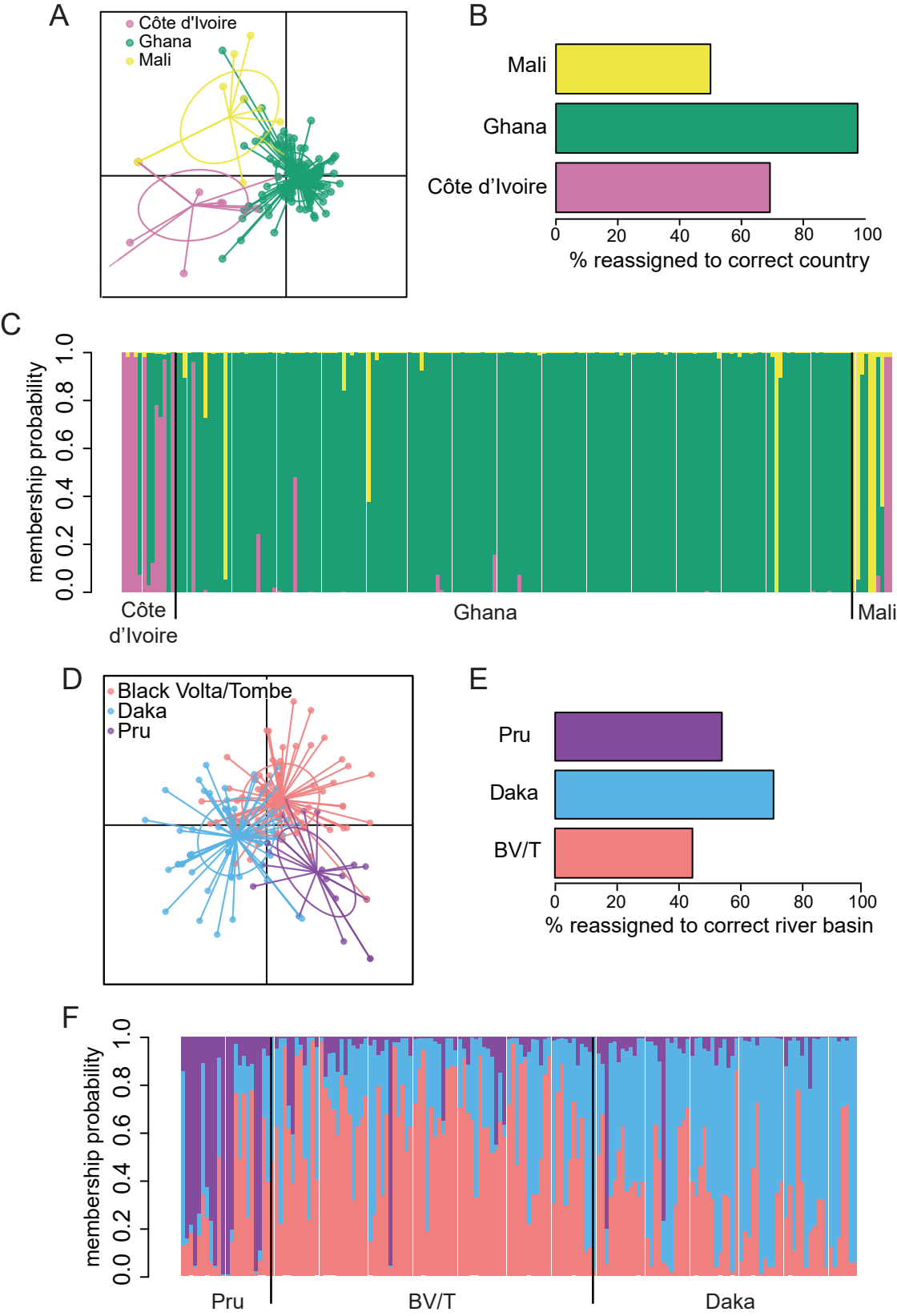
Zimmerman, P.A., Katholi, C.R., Wooten, M.C., Lang-Unnasch, N., Unnasch, T.R., 1994. Recent evolutionary history of American *Onchocerca volvulus*, based on analysis of a tandemly repeated DNA sequence family. Mol Biol Evol 11, 384-392.

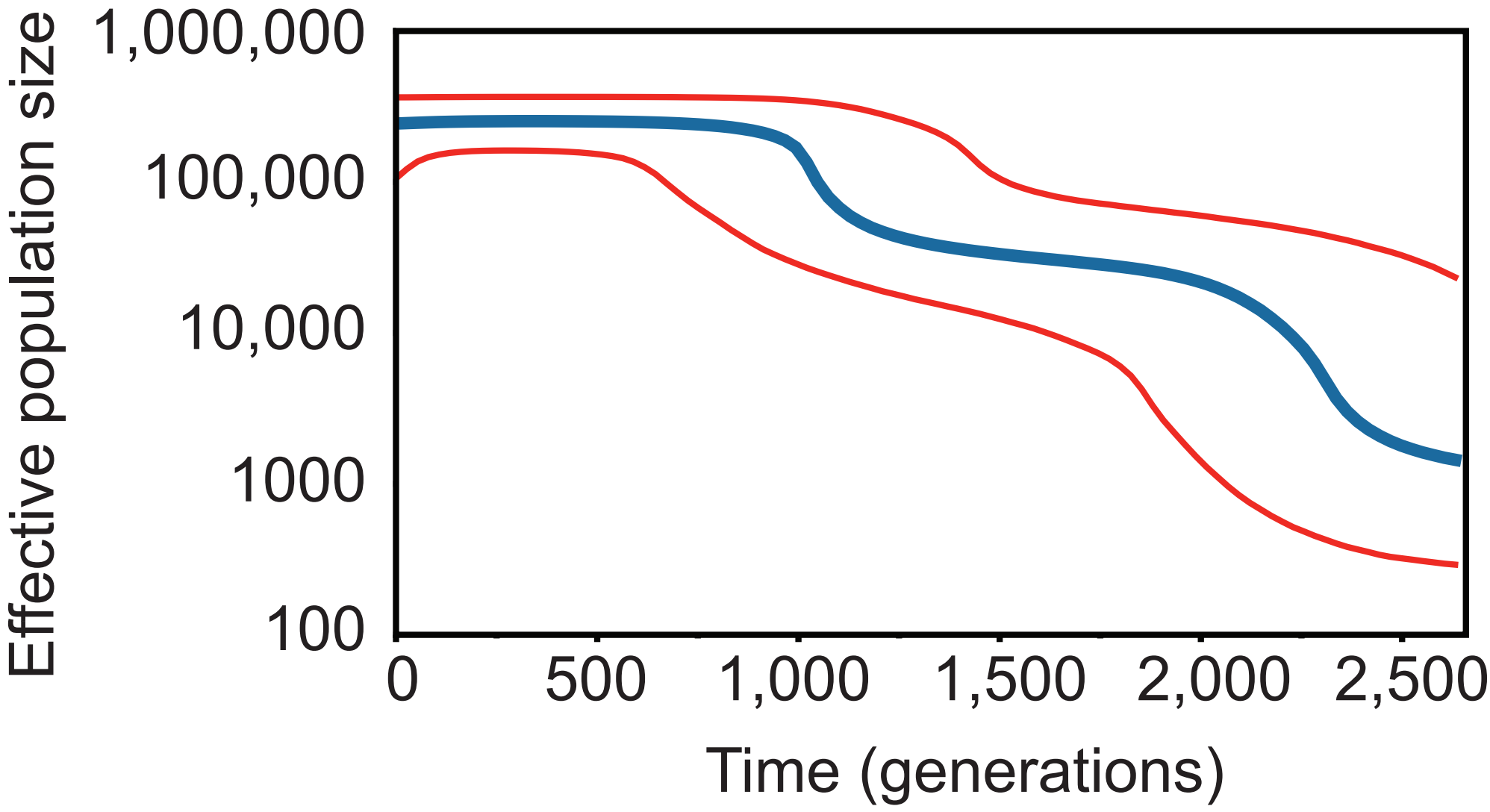

Algeria
Libya
Egypt
Mali
Niger
Chad
Sudan
Benin
Nigeria
Ghana
Ethiopia
CAR
South Sudan
Cameroon
Somalia
Liberia
Uganda
Kenya
Gabon
Congo
DRC
Tanzania
Angola
Mozambique
Zambia
Zimbabwe
Namibia
Botswana
South Africa
Africa
Nodule prevalence
0%-
2%-
5%-
10%-
15%-
20%-
30%-
40%-
60%-
80-100%
unsuitable
national park
<1 person/km²


A)
Mali
Ghana
Côte d'Ivoire
0 200 400 600 km
20
15
10
5
-10
-5
0
5
B)
0 50 100 150 km
1
2
3
Ghana
10
9
8
7
6
-3
-2
-1
0
1


13,690
220
1,630
2,960
Onchocerca volvulus
mitochondrial genome
13,747 bp
8,520
6,050
One primer pair
Two primer pairs
Three primer pairs


A)
10 samples
1 sample
Mali
Côte d'Ivoire
Ghana
B)
10 samples
1 sample
BVT
Pru
Daka


A
Côte d'Ivoire
Ghana
Mali
B
Mali
Ghana
Côte d'Ivoire
0
20
40
60
80
100
% reassigned to correct country
C
membership probability
0.0
0.2
0.4
0.6
0.8
1.0
Côte d'Ivoire
Ghana
Mali
D
Black Volta/Tombe
Daka
Pru
E
Pru
Daka
BV/T
0
20
40
60
80
100
% reassigned to correct river basin
F
membership probability
0.0
0.2
0.4
0.6
0.8
1.0
Pru
BV/T
Daka


Effective population size
1,000,000
100,000
10,000
1000
100
0
500
1,000
1,500
2,000
2,500
Time (generations)