**Human Migration and the Spread of the Nematode Parasite *Wuchereria bancrofti***
**Supplementary Material and Methods**

***Samples***

Haiti samples were provided as cryopreserved peripheral blood mononuclear cells (PBMCs). Mali and Kenya samples were provided as cryopreserved whole blood and later filtered using a 2 Nm filter paper to reduce human cells (Erickson *et al.* 2013). Papua New Guinea (PNG) samples were provided as filtered microfilaria preserved in RNAlater (Life Technologies) and stored at −20°C. Further sample information can be found in supplementary table S2.

***Improvement of the W. bancrofti (Wb) draft genome sequence and gene annotation***

Pac Bio sequencing using chemistry RS II was performed on DNA isolated from the sample(s) described in Small *et al.* 2016. Following whole genome amplification (as noted in Small 2016), the whole genome amplified DNA was treated according to methods in Zhang *et al.* 2006 to reduce chimeric errors. DNA was cleaned using a 1.6:1 Agencourt AMPure XP bead concentration and then eluted in 20 µl ddH2O water. DNA was then incubated in a T100(TM) Thermocycler (Bio-Rad) with 1 µl of Phi29 enzyme with 25 µM concentration of dNTPs and 1x BSA for 30 minutes at 30 °C followed by 65 °C for 3 minutes to deactivate the Phi29 enzyme. DNA was once again cleaned with 1.6:1 Agencourt AMPure XP bead concentration and incubated with one unit of S1 Nuclease to cleave junctions of branched DNA molecules. DNA was again cleaned (see above) and nick-repaired using PreCR(R) DNA Repair (New England Biolabs, Inc.) and then quantified by Qubit Flourometric Quantitation (ThermoFisher Scientific). The DNA template was prepared and sequenced by the McGill University and Génome Québec. Ten PacBio SMRT cells were run for an estimated coverage of 20-40X. Whole genomes were aligned using progressiveCactus (*https://github.com/glennhickey/progressiveCactus*) with a guide tree reconstructed from whole mitochondrial genomes (Small *et al.* 2014). Ragout (Kolmogorov *et al.* 2014) was run allowing for repeat resolution with scaffolds named according to *B. malayi* reference genome. BUSCO v2.3 was used to evaluate genome completeness (Simão *et al.* 2015).

RNA was isolated from a single PNG sample preserved in RNALater. Template and library preparation followed the standard TruSeq RNA Kit TM protocol (Illumina, Inc.). The RNA library was sequenced at 100 base pairs as paired-end reads at the Case Western Reserve University Genomics Core on an Illumina HiSeq 2500. Resulting reads were mapped to the Human genome reference 19 (Hg19) as well as the Wb genome (PRJNA275548) using HISAT (Kim *et al.* 2015) to separate reads belonging to the human host and then Wb worm. Sequences mapping successfully to Wb and not human were used in Maker3 (Cantarel *et al.* 2008) along with Wb EST libraries (SAW95SjL-WbMf) and protein sequences curated for *B. malayi* (PRJNA10729) and *Loa loa* (PRJNA60051). Maker3 was run for three progressive iterations using initial gene prediction based on single copy orthologs identified using the program BUSCO (Simão *et al.* 2015).

***sWGA and Population Sequencing***

Primer sets were tested by amplifying DNA isolated from one infected-patient blood sample as well as from two single microfilaria (MF). DNA quality and ploidy were confirmed by amplifying and sequencing the mitochondrial cytochrome oxidase I (CO1) gene (Small *et al.* 2013). Primer sets were nearly indistinguishable using metrics of read depth and percent of the genome covered (Clarke *et al.* 2017; Leichty *et al.* 2014), so the larger primer set was selected to obtain more even coverage for fragmented DNA (supplementary table S5).

Isolated DNA was used directly for whole genome amplification following the protocol in Leichty *et al.* 2014, except with 1.5 µg of each of custom primers. The reaction was incubated for 8 hours at 30 C followed by a 70 °C for 15 min in T100(TM) Thermocycler (BioRad). Reaction time was chosen to maximize DNA yield while minimizing excessive duplication after analyzing reactions of 4, 6, 8, 10, and 12 hours. After 8-hour incubation, amplified samples were cleaned using a 1.6:1 Agencourt AMPure XP bead concentration to remove small fragments. 40 cycles of quantitative PCR (qPCR) using SYBR® (Life Technologies) in a Mastercycler® RealPlex2 (Eppendorf) were used to determine the relative proportions of host (human) and parasite (Wb) DNA. Human DNA was quantified using a custom designed primer pair to amplify a section of chromosome 1: HuQPCR1-F 5'-ACTTTGGGTCATTCCCACTG-3', HuQPCR1-R 5' -GCTCAGCTCCTTGCTGGATA -3'. Wb DNA was quantified using primers to

amplify isotype-1 of the β-tubulin gene (Hoti *et al.* 2003). Overall success rate of amplification, measured by final DNA concentration (> 500 ng), from single microfilaria (MF) was ~60%.

1.5-2 µg of amplified DNA was used in the TruSeq DNA PCR-Free kit (Illumina, Inc.) following steps for the 550 bp insert size. Finished libraries were quantified on a 2100 Bioanalyzer Instrument (Agilent). Samples were pooled in equal molarity and then initially sequenced on Illumina MiSeq for 35 bp to rigorously quantify the proportion of Wb and human DNAs in each library as well as library complexity before final sequencing. Resulting sequencing were mapped to the Wb genome using BWA (Li *et al.* 2009) and used to adjust pooling ratios between samples. A total of 12 pooled libraries (each containing 6 sample libraries) were sequenced on 12 lanes of the Illumina HiSeq X for 150 base pairs in paired-end mode at the McGill University and Génome Québec. Sequencing resulted in a total of 1.2 Tb of sequencing data from 42 sample libraries.

***Quality control***

All sequences were first trimmed for adapter sequences using TrimGalore (*http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/*) and then examined for base-quality using fastqc (*https://www.bioinformatics.babraham.ac.uk/projects/fastqc/*) and multiqc (*https://github.com/ewels/MultiQC*). Filtered read data were then mapped to the mitochondrial genome of Wb (Ramesh *et al.* 2012) to verify that each sample contained DNA from only a single genome. We calculated kinship value using the program KING (Manichaikul *et al.* 2010). We started with the full 56 sample data set and ran KING with the –relatedness option. Two related individuals were identified which were also noted in (Small *et al.* 2016). We removed one sample from a pair if the kinship value was > 3$^{rd}$ order within a host.

***Variant Calling***

GATK v3.1 HaplotypeCaller (McKenna *et al.* 2010) was run with the following parameters: "-newQual --emitRefConfidence GVCF --pcr_indel_model NONE". The resulting gVCF from HaplotypeCaller were converted to VCF format using GATK program GenotypeGVCFs. VCF positions were removed if they failed the following filter criteria: "QD < 5, QUAL < 30, DP < 14, MQ < 30, MQRankSum < -12.5, ReadPosRankSum < -8.0, FS > 60.0, ABHet < .30, ABHet

> 0.70, ABHom < 0.90". Positions passing all filters in each individual sample were merged to produce a final set of SNPs across all samples. As combining SNPs among samples can lead to missing data, homozygous or uncalled in original sample, the candidate SNP set was then used in freebayes to fill homozygous positions (Garrison *et al*. 2012). Putatively repetitive and paralogous sequences were identified using RepeatMasker (Tarailo-Graovac *et al.* 2009) and Snpable (*http://lh3lh3.users.sourceforge.net/snpable.shtml*).

***Simulations of Multiple Introductions***

We used the random-forest in abcrf (Pudlo *et al.* 2015) to test between models of 1 or more introductions into the Wb population of Haiti. We tested 4 models: 1) Single introduction from Mali; 2) Single introduction from Mali + 1 pulse gene flow event from a ghost population; 3) Single introduction from Mali + 2 pulse gene flow events from different ghost population; 4) Single introduction from Mali + 3 pulse gene flow events from different ghost populations. The sampled populations of Haiti, Mali, Kenya, and PNG were simulated under the best fit model as described in Table 1. msmove (*github.com/geneva/msmove*) was used to make 10,000 simulations under each model. We then calculated the following summary statistics: pi, theta, Tajima's D, distSkew, distKurtosis. Confusion matrix and classifications are shown below. The results suggest that we are not able to confidently differentiate between models of a single introduction and multiple introductions based on our currently available data.

| | M1 | M2 | M3 | M4 | Class Error |
|---|---|---|---|---|---|
| M1 | **3680** | 2156 | 2126 | 2007 | 63% |
| M2 | 3525 | **2128** | 2218 | 2108 | 78% |
| M3 | 3540 | 2196 | **2116** | 2118 | 78% |
| M4 | 3486 | 2150 | 2167 | **2177** | 78% |

| Best Model | M1 | M2 | M3 | M4 | PostProb |
|---|---|---|---|---|---|
| M3 | 128 | 119 | 129 | 124 | .494 |

***References***


Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., … Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188–196. http://doi.org/10.1101/gr.6743907

Clarke, E. L., Sundararaman, S. A., Seifert, S. N., Bushman, F. D., Hahn, B. H., & Brisson, D. (2017). Swga: A primer design toolkit for selective whole genome amplification. In J. Hancock (Ed.), *Bioinformatics* (Vol. 33, pp. 2071–2077). http://doi.org/10.1093/bioinformatics/btx118

Erickson, S. M., Thomsen, E. K., Keven, J. B., Vincent, N., Koimbu, G., Siba, P. M., … Reimer, L. J. (2013). Mosquito-parasite interactions can shape filariasis transmission dynamics and impact elimination programs. *PLoS Neglected Tropical Diseases*, *7*(9), e2433. http://doi.org/10.1371/journal.pntd.0002433

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv:1207.3907*.

Hoti, S. L., Subramaniyan, K., & Das, P. K. (2003). Detection of codon for amino acid 200 in isotype 1 β-tubulin gene of *Wuchereria bancrofti* isolates, implicated in resistance to benzimidazoles in other nematodes. *Acta Tropica*, *88*(1), 77–81. http://doi.org/10.1016/S0001-706X(03)00159-1

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. http://doi.org/10.1038/nmeth.3317

Kolmogorov, M., Raney, B., Paten, B., & Pham, S. (2014). Ragout - A reference-assisted assembly tool for bacterial genomes. *Bioinformatics*, *30*(12), i302–i309. http://doi.org/10.1093/bioinformatics/btu280

Leichty, A., & Brisson, D. (2014). Selective whole genome amplification for re-sequencing target microbial species from complex natural samples. *Genetics*, 1–30. Retrieved from http://www.genetics.org/content/early/2014/08/05/genetics.114.165498.short

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. http://doi.org/10.1093/bioinformatics/btp324

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-

generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. http://doi.org/10.1101/gr.107524.110

Ramesh, A., Small, S. T., Kloos, Z. A., Kazura, J. W., Nutman, T. B., Serre, D., & Zimmerman, P. A. (2012). The complete mitochondrial genome sequence of the filarial nematode *Wuchereria bancrofti* from three geographic isolates provides evidence of complex demographic history. *Molecular Biochemical Parasitology*, *183*(1), 32–41.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. http://doi.org/10.1093/bioinformatics/btv351

Small, S. T., Ramesh, A., Bun, K., Reimer, L., Thomsen, E., Baea, M., … Zimmerman, P. A. (2013). Population genetics of the filarial worm *Wuchereria bancrofti* in a post-treatment region of Papua New Guinea: insights into diversity and life history. *PLoS Neglected Tropical Diseases*, *7*(7), e2308. http://doi.org/10.1371/journal.pntd.0002308

Small, S. T., Tisch, D. J., & Zimmerman, P. a. (2014). Molecular epidemiology, phylogeny and evolution of the filarial nematode *Wuchereria bancrofti*. *Infection, Genetics and Evolution*, *28*, 33–43. http://doi.org/10.1016/j.meegid.2014.08.018

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *Chapter 4*, Unit 4.10. http://doi.org/10.1002/0471250953.bi0410s25

Zhang, K., Martiny, A. C., Reppas, N. B., Barry, K. W., Malek, J., Chisholm, S. W., & Church, G. M. (2006). Sequencing genomes from single cells by polymerase cloning. *Nature Biotechnology*, *24*(6), 680–686. http://doi.org/10.1038/nbt1214

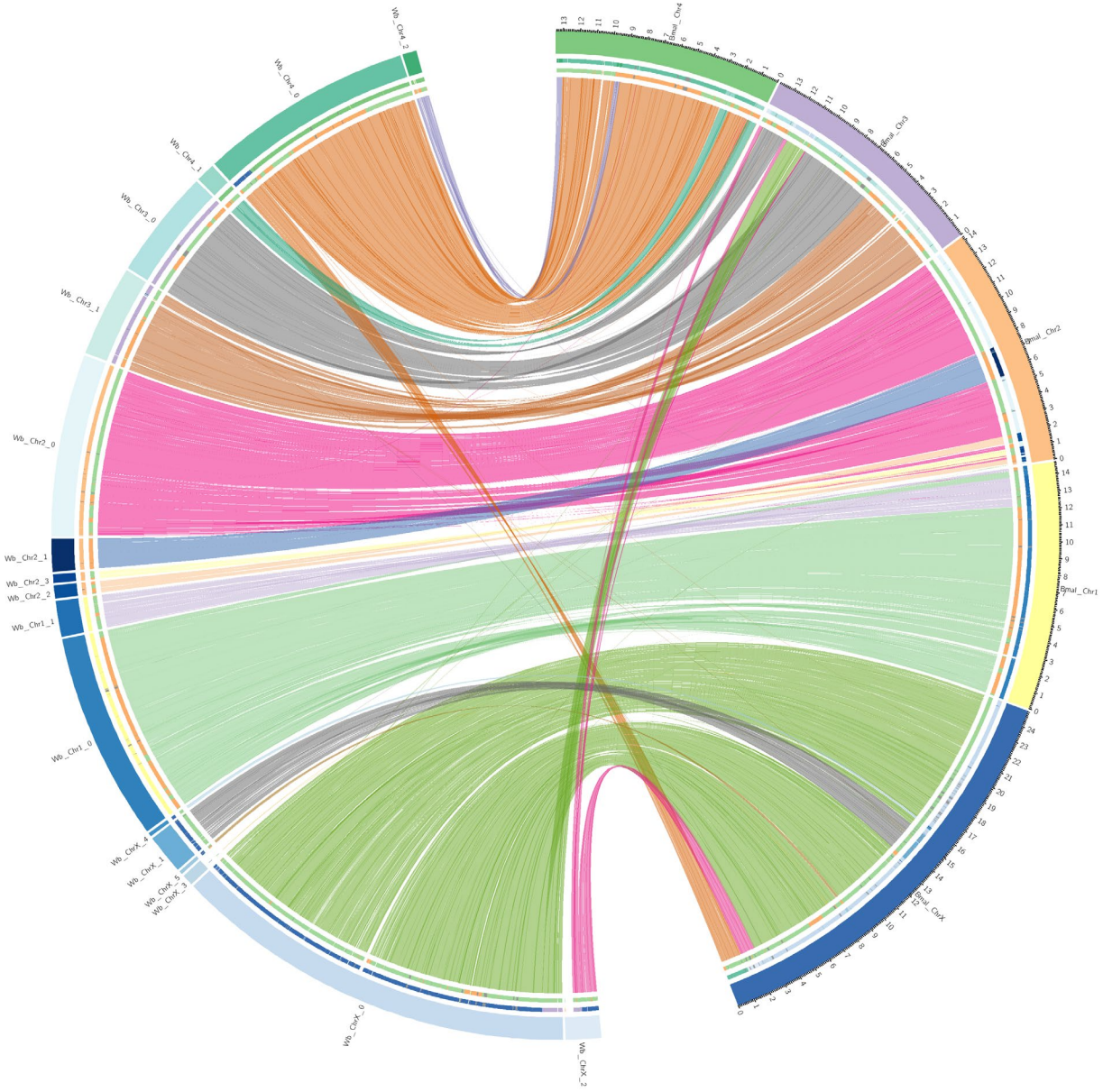

**Fig. S1.** A circos plot visualizing the alignment between the new *W. bancrofti* genome assembly (SAMN10411877) and chromosome scale assembly of *Brugia malayi*.

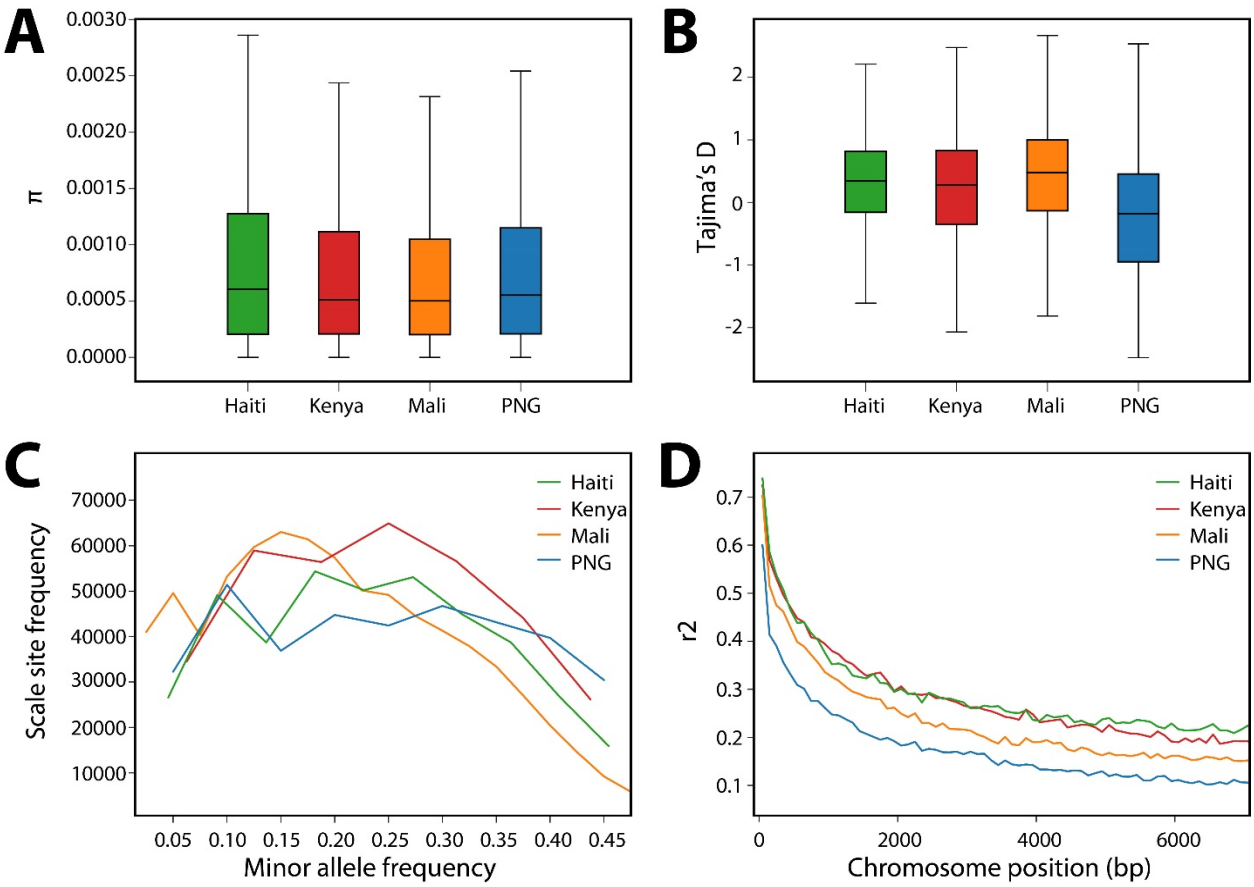


**Fig. S2.** Population genetic statistics for each *W. bancrofti* population. **A**) pairwise nucleotide diversity; **B**) distribution of Tajima's D; **C**) scaled site-frequency spectrum; **D**) decay in genotype correlation with distance.

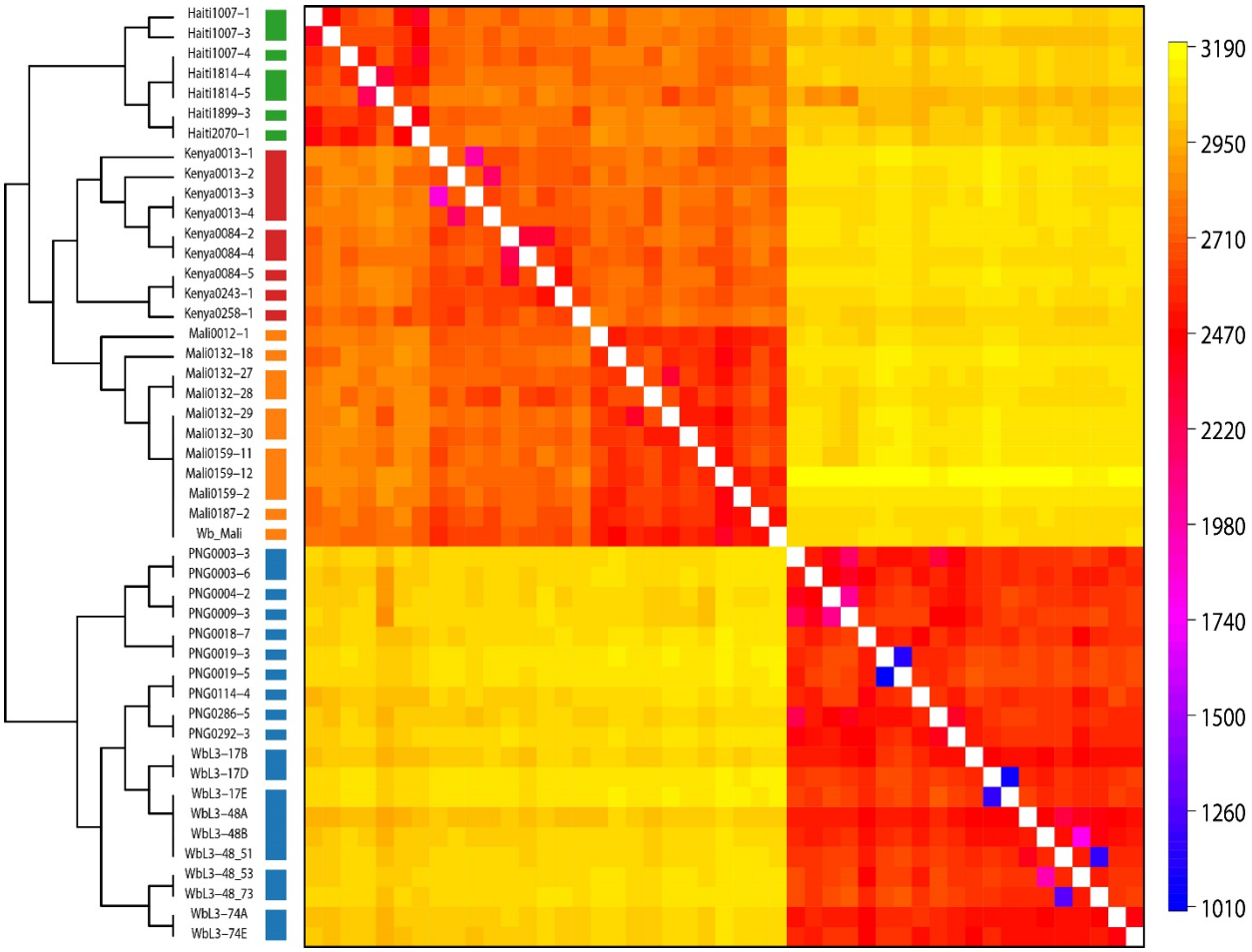


**Fig. S3.** Coancestry matrix (a summary of nearest neighbor haplotype relationships in the data set) of *W. bancrofti* populations using fineSTRUCTURE with added labels denoting sample IDs. Dendrograms relating individuals to the co-ancestry matrix are along the vertical axis with populations denoted by colors corresponding to the principal component analysis (PCA; Figure 1). Cooler colors represent higher co-ancestry. Fused boxes of similar color denote worms sampled from the same host infection.

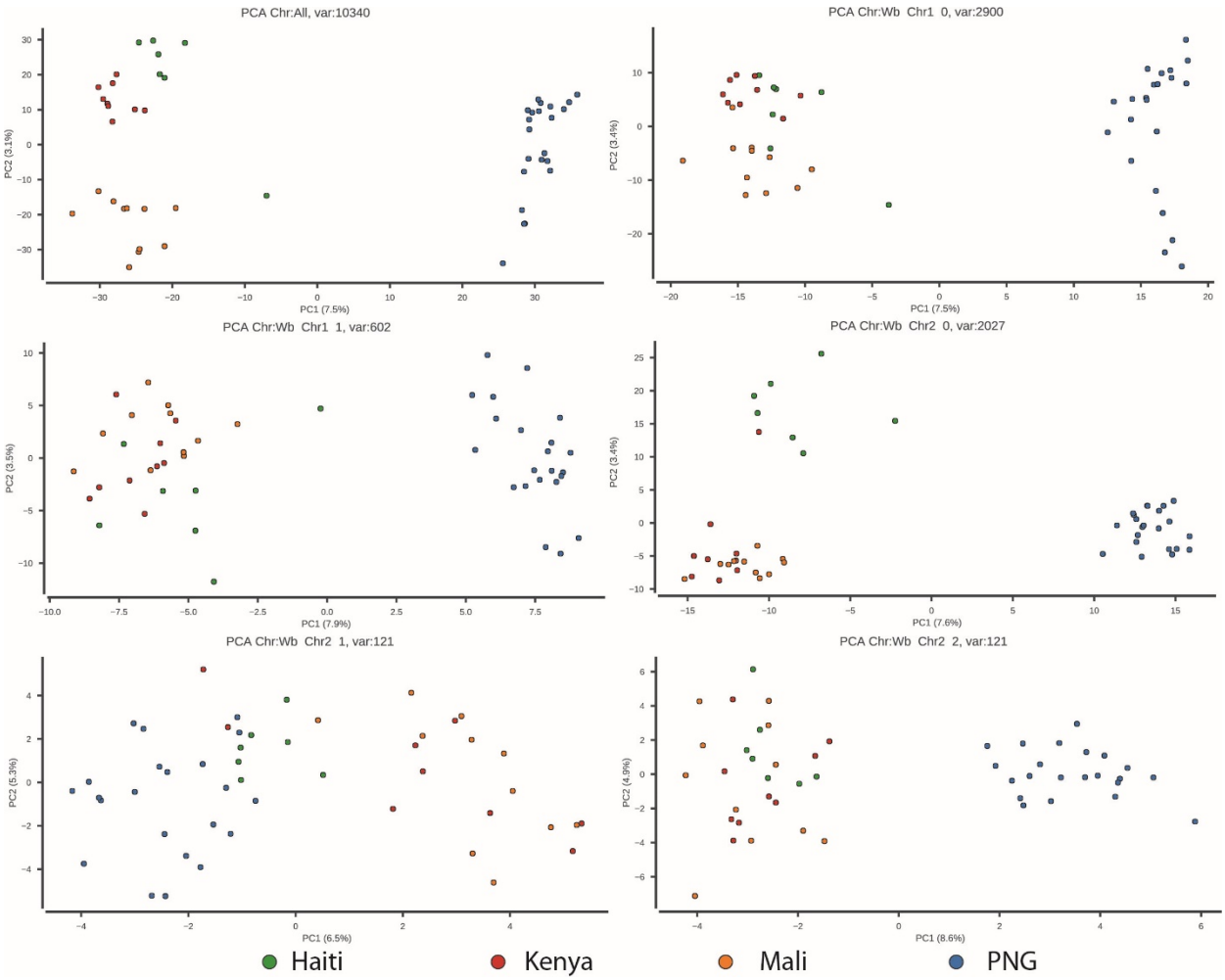


**Fig. S4-1.** Principal component analysis (PCA) for all scaffolds after removing sites in high linkage disequilibrium (LD). Scaffold names are listed at top center of each panel. Scaffold PCAs are split between two figures for readability.

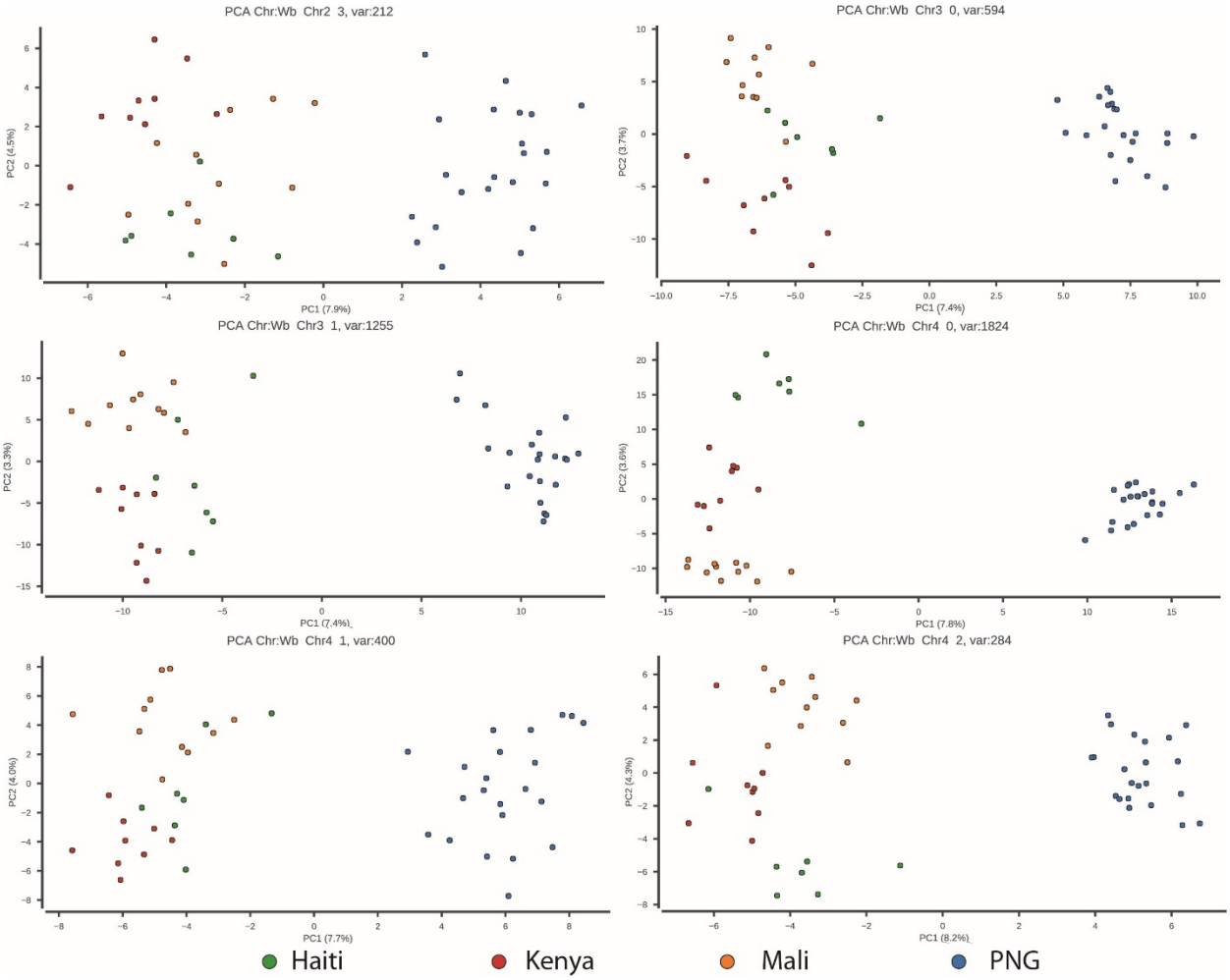


**Fig. S4-2.** Principal component analysis (PCA) for all scaffolds after removing sites in high linkage disequilibrium (LD). Scaffold names are listed at top center of each panel. Scaffold PCAs are split between two figures for readability.

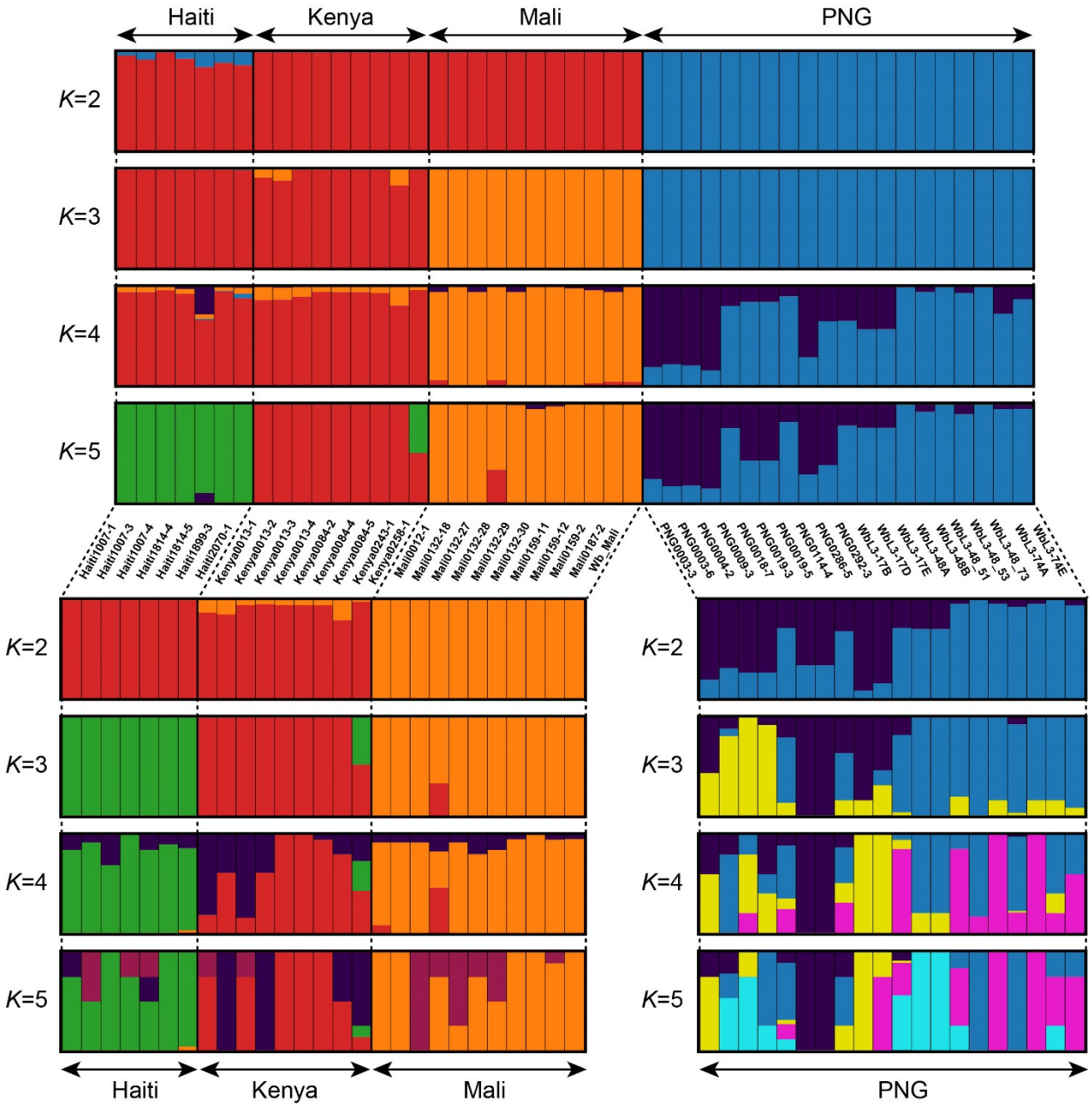


**Fig. S5.** ADMIXTURE results. Three panels represent results from running the program ADMIXTURE to investigate population structure and admixture. Top panel includes all sampled populations, while the right and left panels are specific to Haiti-Africa and Papua New Guinea (PNG).

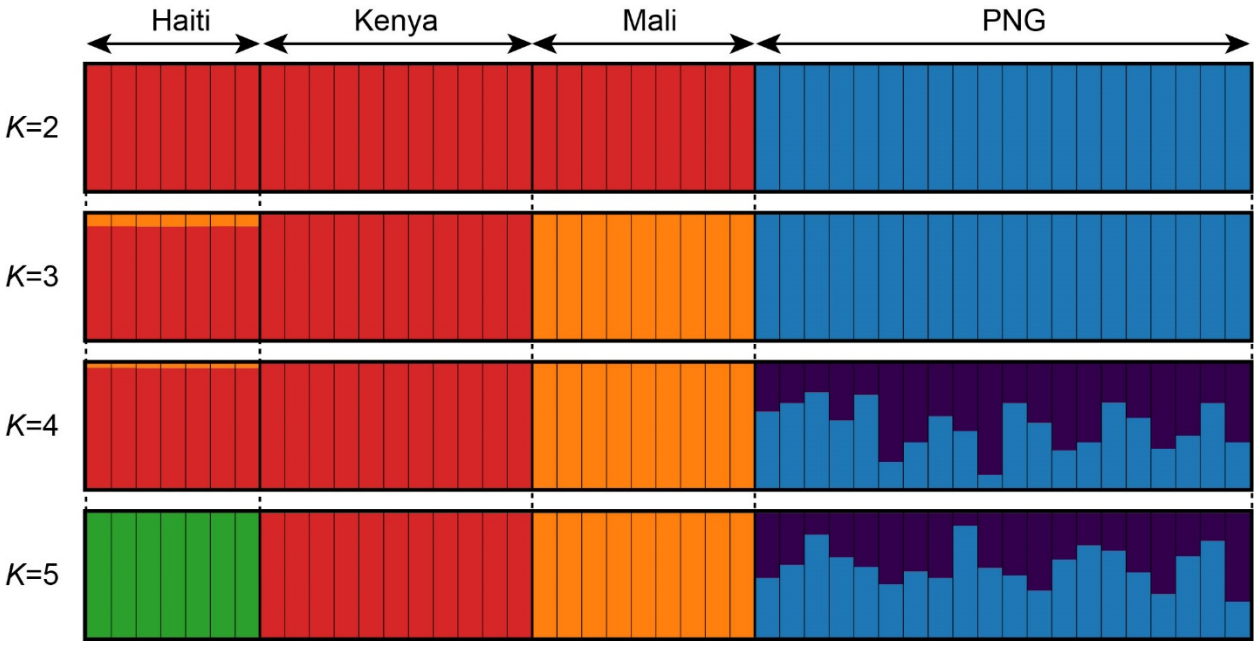


**Fig. S6.** ADMIXTURE results from simulations under best fit model for *W. bancrofti* population history. PNG: Papua New Guinea.

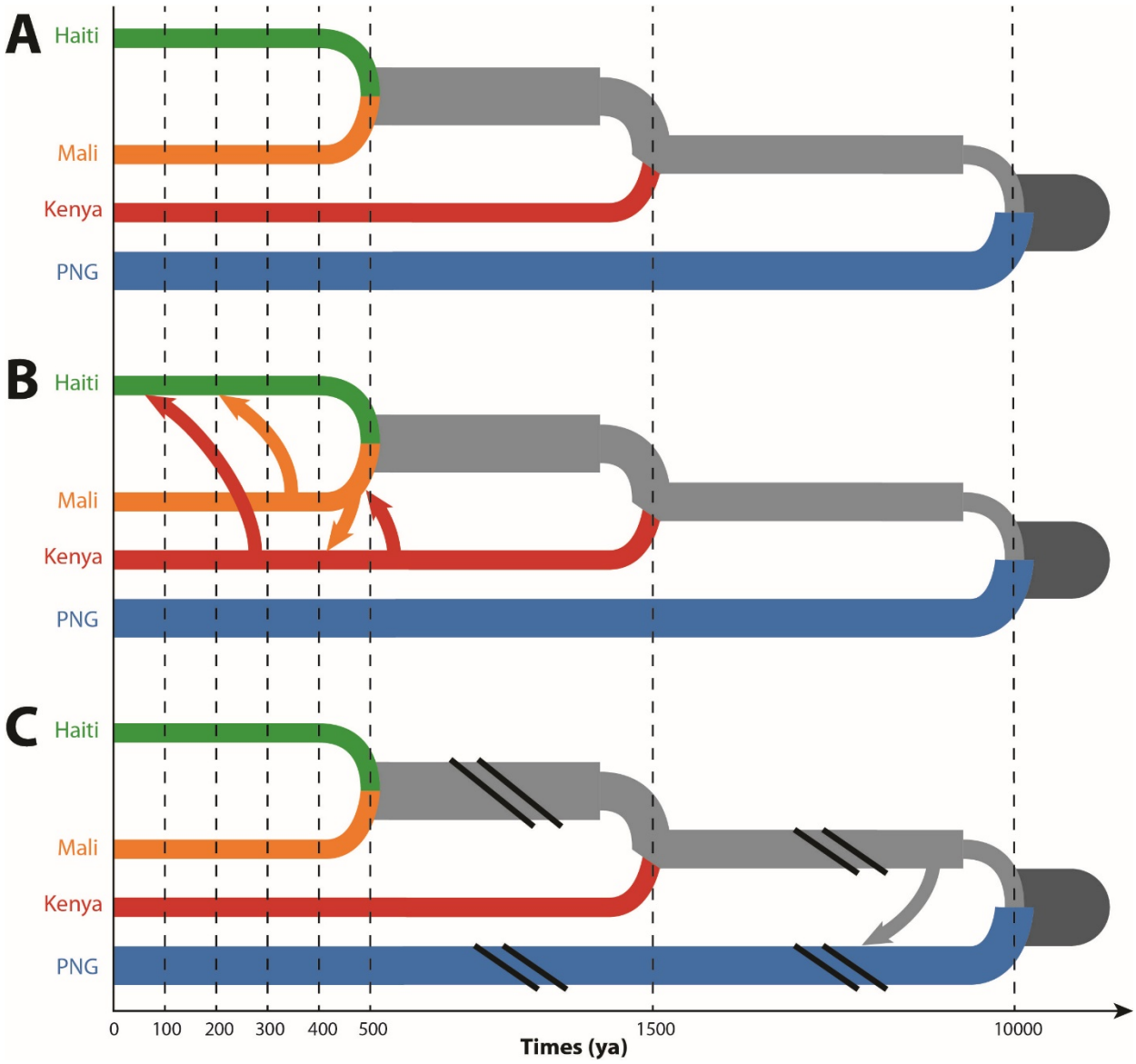


**Fig. S7.** Illustration of tested models for describing *W. bancrofti* population history. PNG: Papua New Guinea.

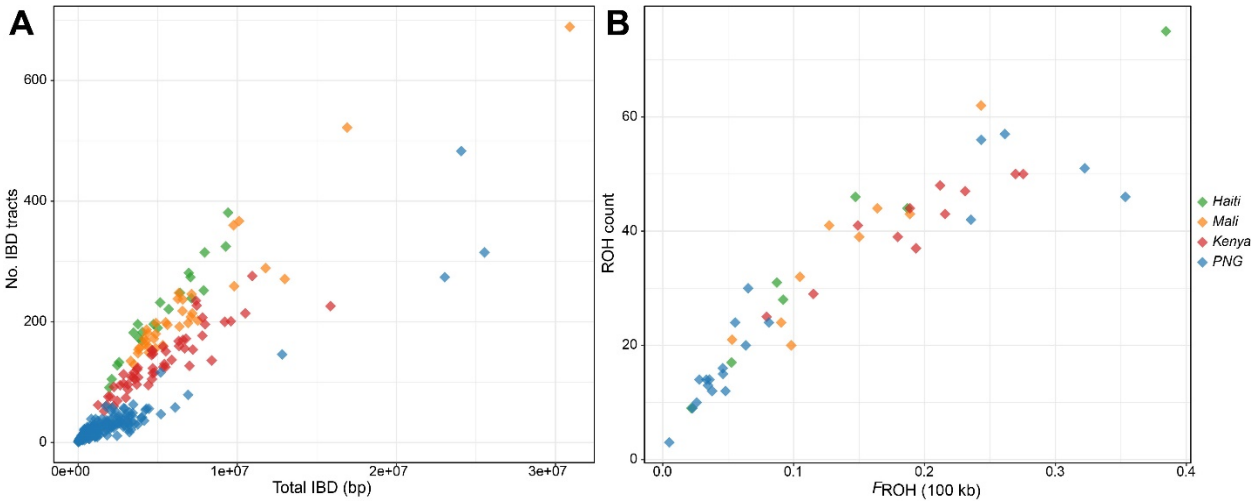


**Fig. S8.** Plots representing relatedness and inbreeding in *W. bancrofti* samples. **A**) Pairwise comparison of Identity by descent (IBD) tracts within *W. bancrofti* populations; **B**) Count of runs of homozygosity (ROH) > 100kb in the *W. bancrofti* samples. PNG: Papua New Guinea.

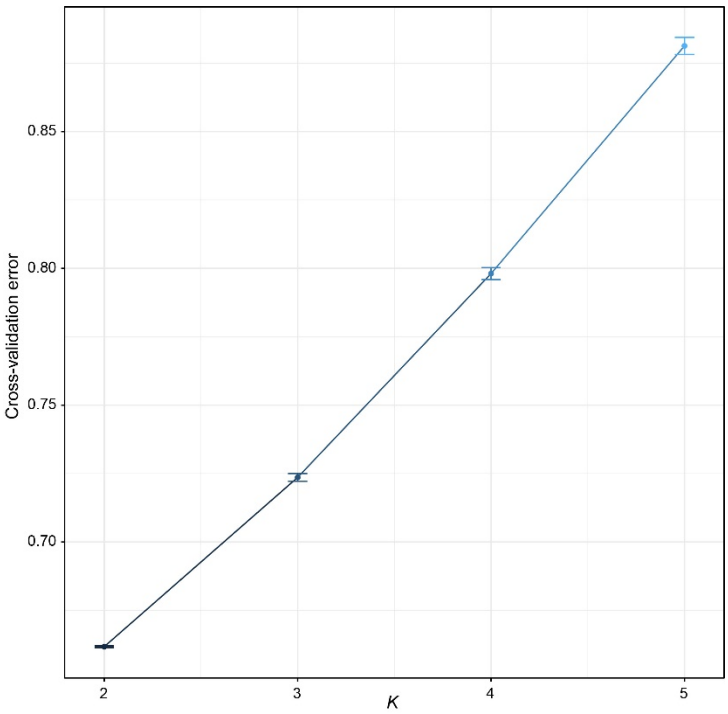


**Fig. S9.** Cross validation of ADMIXTURE results (± standard error; SE).

**Table S1.** Assembly statistics for *W. bancrofti* genome.

| Chromosome | Scaffold ID | Size | Ns | Genes | Gene density |
|---|---|---|---|---|---|
| 1 | Wb_Chr1_0 | 12,368,652 | 81,951 | 1,633 | 0.000132 |
| | Wb_Chr1_1 | 2,119,897 | 98,244 | 182 | 0.000086 |
| 2 | Wb_Chr2_0 | 10,841,590 | 87,666 | 1,143 | 0.000105 |
| | Wb_Chr2_1 | 1,918,629 | 415 | 253 | 0.000132 |
| | Wb_Chr2_2 | 750,293 | 0 | 106 | 0.000141 |
| | Wb_Chr2_3 | 490,737 | 2,531 | 58 | 0.000118 |
| 3 | Wb_Chr3_0 | 6,490,679 | 17,498 | 917 | 0.000141 |
| | Wb_Chr3_1 | 5,401,749 | 67,144 | 569 | 0.000105 |
| 4 | Wb_Chr4_0 | 12,704,880 | 207,068 | 1,369 | 0.000108 |
| | Wb_Chr4_1 | 1,128,067 | 610 | 98 | 0.000087 |
| | Wb_Chr4_2 | 813,737 | 109,849 | 45 | 0.000055 |
| X | Wb_ChrX_0 | 24,139,319 | 571,287 | 2,710 | 0.000112 |
| | Wb_ChrX_1 | 2,355,808 | 45,023 | 210 | 0.000089 |
| | Wb_ChrX_2 | 2,085,878 | 19,869 | 137 | 0.000066 |
| | Wb_ChrX_3 | 662,921 | 28,893 | 66 | 0.000100 |
| | Wb_ChrX_4 | 204,088 | 0 | 5 | 0.000024 |
| | Wb_ChrX_5 | 173,113 | 0 | 16 | 0.000092 |

**Table S2-1.** Sample Information for *W. bancrofti* worms included in this study. Table is splitted in two tables for readability.

| Population | Sample ID | Bleed code | Worm ID | Sex | Source material | Collection date | Hom ref | Hom alt | Het | Average genotype coverage | % genome > 10x | Ref |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Haiti | Haiti1007-1 | 1007 | 1 | F | MF | 1991 | 343,434 | 42,596 | 20,500 | 53 | 0.62 | This study |
| | Haiti1007-3 | 1007 | 3 | F | MF | 1991 | 373,040 | 44,294 | 28,002 | 201 | 0.76 | This study |
| | Haiti1007-4 | 1007 | 4 | M | MF | 1991 | 369,214 | 46,034 | 25,736 | 225 | 0.73 | This study |
| | Haiti1814-4 | 1814 | 4 | M | MF | 1991 | 373,049 | 44,704 | 33,048 | 253 | 0.88 | This study |
| | Haiti1814-5 | 1814 | 5 | M | MF | 1991 | 331,504 | 31,221 | 7,490 | 138 | 0.32 | This study |
| | Haiti1899-3 | 1899 | 3 | F | MF | 1991 | 369,653 | 37,121 | 54,958 | 286 | 0.81 | This study |
| | Haiti1899-6 | 1899 | 6 | M | MF | 1991 | 323,252 | 31,506 | 6,499 | 103 | 0.28 | This study |
| | Haiti1938-1 | 1938 | 1 | F | MF | 1991 | 350,786 | 35,012 | 8,923 | 124 | 0.38 | This study |
| | Haiti2070-1 | 2070 | 1 | F | MF | 1991 | 377,830 | 35,454 | 48,457 | 305 | 0.89 | This study |
| Kenya | Kenya0013-1 | 1040013 | 1 | F | MF | 2003 | 384,379 | 41,748 | 33,082 | 235 | 0.77 | This study |
| | Kenya0013-2 | 1040013 | 2 | F | MF | 2003 | 385,515 | 43,893 | 28,757 | 257 | 0.75 | This study |
| | Kenya0013-3 | 1040013 | 3 | F | MF | 2003 | 376,318 | 43,242 | 14,446 | 143 | 0.65 | This study |
| | Kenya0013-4 | 1040013 | 4 | F | MF | 2003 | 384,945 | 45,191 | 25,914 | 199 | 0.95 | This study |
| | Kenya0084-1 | 1100084 | 1 | F | MF | 2003 | 381,843 | 44,678 | 19,093 | 190 | 0.78 | This study |
| | Kenya0084-4 | 1100084 | 4 | F | MF | 2003 | 353,115 | 41,483 | 14,448 | 126 | 0.49 | This study |
| | Kenya0084-5 | 1100084 | 5 | F | MF | 2003 | 381,235 | 45,005 | 20,447 | 173 | 0.78 | This study |
| | Kenya0243-1 | 1110243 | 1 | F | MF | 2003 | 384,003 | 44,720 | 21,371 | 202 | 0.83 | This study |
| | Kenya01258-1 | 11201258 | 1 | F | MF | 2003 | 384,659 | 44,984 | 25,547 | 219 | 0.96 | This study |
| | Kenya0339-2 | 10802339 | 2 | F | MF | 2003 | 382,986 | 40,991 | 25,799 | 147 | 0.77 | This study |
| Mali | Mali0012-1 | 12A | 1 | M | MF | 2007 | 356,831 | 39,895 | 14,340 | 109 | 0.40 | This study |
| | Mali0132-18 | 132A | 18 | F | MF | 2007 | 364,937 | 42,789 | 19,537 | 161 | 0.56 | This study |
| | Mali0132-27 | 132A | 27 | F | MF | 2007 | 350,090 | 37,913 | 11,159 | 101 | 0.36 | This study |
| | Mali0132-28 | 132A | 28 | M | MF | 2007 | 336,995 | 36,090 | 9,463 | 106 | 0.31 | This study |
| | Mali0132-29 | 132A | 29 | M | MF | 2007 | 374,723 | 41,724 | 13,885 | 119 | 0.55 | This study |
| | Mali0132-30 | 132A | 30 | M | MF | 2007 | 370,633 | 41,266 | 11,877 | 115 | 0.52 | This study |
| | Mali0145-3 | 145A | 3 | M | MF | 2007 | 324,778 | 21,889 | 6,578 | 54 | 0.19 | This study |
| | Mali0159-11 | 159A | 11 | M | MF | 2007 | 364,535 | 40,545 | 9,743 | 111 | 0.48 | This study |
| | Mali0159-12 | 159A | 12 | F | MF | 2007 | 375,014 | 49,828 | 30,760 | 233 | 0.91 | This study |
| | Mali0159-2 | 159A | 2 | F | MF | 2007 | 374,456 | 46,680 | 31,718 | 207 | 0.81 | This study |
| | Mali0187-2 | 187A | 2 | M | MF | 2007 | 368,885 | 42,283 | 20,738 | 132 | 0.53 | This study |
| | WbMali | – | – | F | L4 | 2007 | 380,855 | 43,366 | 18,769 | 12 | 0.32 | Desjardins *et al.* 2013 |

**Table S2-2.** Sample Information for *W. bancrofti* worms included in this study. Table is splitted in two tables for readability.

| Population | Sample ID | Bleed code | Worm ID | Sex | Source material | Collection date | Hom ref | Hom alt | Het | Average genotype coverage | % genome > 10x | Ref |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Papua New Guinea | PNG00003-3 | ZXT0003 | 3 | M | MF | 2014 | 329,959 | 17,403 | 7,015 | 92 | 0.19 | This study |
| | PNG00003-6 | ZXT0003 | 6 | M | MF | 2014 | 389,143 | 30,702 | 25,022 | 255 | 0.80 | This study |
| | PNG00004-2 | ZXT0004 | 2 | F | MF | 2014 | 354,084 | 25,444 | 9,451 | 127 | 0.33 | This study |
| | PNG00009-3 | ZXT0009 | 3 | M | MF | 2014 | 315,878 | 14,037 | 5,793 | 99 | 0.17 | This study |
| | PNG0018-1 | ZXT0018 | 1 | M | MF | 2014 | 379,265 | 17,502 | 67,569 | 293 | 0.96 | This study |
| | PNG0018-7 | ZXT0018 | 7 | F | MF | 2014 | 396,366 | 20,088 | 44,767 | 198 | 0.92 | This study |
| | PNG0019-3 | ZXT0019 | 3 | M | MF | 2014 | 399,479 | 24,374 | 30,578 | 247 | 0.87 | This study |
| | PNG0019-5 | ZXT0019 | 5 | F | MF | 2014 | 399,967 | 27,095 | 24,057 | 238 | 0.81 | This study |
| | PNG0114-4 | ZXT0114 | 4 | F | MF | 2014 | 396,615 | 27,583 | 37,298 | 351 | 0.83 | This study |
| | PNG0286-5 | ZXT0286 | 5 | F | MF | 2014 | 361,395 | 21,915 | 9,072 | 141 | 0.36 | This study |
| | PNG0286-6 | ZXT0286 | 6 | F | MF | 2014 | 336,546 | 19,297 | 15,569 | 57 | 0.34 | This study |
| | PNG0292-3 | ZXT0292 | 3 | M | MF | 2014 | 370,779 | 25,067 | 8,985 | 130 | 0.49 | This study |
| | WbL3-17B | ZE40A | – | F | L3 | 2012 | 375,407 | 17,692 | 42,218 | 17 | 0.62 | Small *et al.* 2016 |
| | WbL3-17D | ZE40A | – | F | L3 | 2012 | 376,409 | 17,908 | 52,590 | 42 | 0.89 | Small *et al.* 2016 |
| | WbL3-17E | ZE40A | – | F | L3 | 2012 | 377,978 | 17,714 | 56,342 | 43 | 0.85 | Small *et al.* 2016 |
| | WbL3-36 | YUA086 | – | F | L3 | 2012 | 368,560 | 10,385 | 60,733 | 32 | 0.55 | Small *et al.* 2016 |
| | WbL3-48_51 | ZE39A | – | U | L3 | 2012 | 407,757 | 22,533 | 22,635 | 17 | 0.71 | Small *et al.* 2016 |
| | WbL3-48_53 | ZE39A | – | M | L3 | 2012 | 402,251 | 22,426 | 33,381 | 80 | 0.86 | Small *et al.* 2016 |
| | WbL3-48_73 | ZE39A | – | F | L3 | 2012 | 404,774 | 23,215 | 30,625 | 48 | 0.97 | Small *et al.* 2016 |
| | WbL3-48A | ZE39A | – | M | L3 | 2012 | 394,697 | 20,929 | 45,184 | 14 | 0.63 | Small *et al.* 2016 |
| | WbL3-48B | ZE39A | – | M | L3 | 2012 | 383,842 | 18,187 | 57,940 | 30 | 0.80 | Small *et al.* 2016 |
| | WbL3-17A | ZE40A | – | F | L3 | 2012 | 379,768 | 17,631 | 54,829 | 36 | 0.82 | Small *et al.* 2016 |
| | WbL3-74A | YUA086 | – | M | L3 | 2012 | 380,714 | 18,969 | 15,209 | 13 | 0.62 | Small *et al.* 2016 |
| | WbL3-74E | YUA086 | – | M | L3 | 2012 | 390,560 | 20,537 | 22,181 | 22 | 0.80 | Small *et al.* 2016 |
| | WbL3-17C | ZE40A | – | F | L3 | 2012 | 381,567 | 17,767 | 41,283 | 13 | 0.62 | Small *et al.* 2016 |

**Table S3.** Distribution of SNPs for each *W. bancrofti* population.

| Population | SNPs | Singletons | Doubletons | Private |
|---|---|---|---|---|
| Haiti | 165,412 | 43,168 | 33,663 | 24,776 |
| Mali | 152,458 | 31,910 | 27,479 | 23,201 |
| Kenya | 149,685 | 33,863 | 28,585 | 23,564 |
| PNG | 204,138 | 51,289 | 31,812 | 78,091 |

PNG: Papua New Guinea.

**Table S4.** Model selection summary and statistics.

| Parameter | Population | Model1-1 | Model1-2 | Model1-3 | Model2-1 | Model3-1 |
|---|---|---|---|---|---|---|
| Ne | Haiti | 282 | 282 | 282 | 282 | 282 |
| | Mali | 313 | 313 | 313 | 313 | 313 |
| | Kenya | 642 | 642 | 642 | 642 | 642 |
| | Papua New Guinea | 2,834 | 2,834 | 2,834 | 2,834 | 2,834 |
| T3 | Haiti, Mali | 289 | 289 | 289 | 429 | 3,875 |
| T2 | (Haiti, Mali), Kenya | 2,477 | 2,477 | 2,477 | 1,456 | 3,946 |
| T1 | (Haiti, Mali, Kenya), Papua New Guinea | 17,304 | 5,000 | 50,000 | 50,000 | 19,733 |
| NA | Haiti, Mali | 16,669 | 16,669 | 16,669 | 21,444 | 42,866 |
| | (Haiti, Mali), Kenya | 987 | 987 | 987 | 2,765 | 987 |
| | (Haiti, Mali, Kenya), Papua New Guinea | 107,255 | 107,255 | 107,255 | 98,317 | 215,517 |
| Admixture time | (Haiti, Mali, Kenya), Papua New Guinea | – | – | – | 4,558 | – |
| Admixture proportion | (Haiti, Mali, Kenya), Papua New Guinea | – | – | – | 0.001 | – |
| Migration | Time Mali – Haiti | – | – | – | – | 3,458 |
| | Rate Mali – Haiti | – | – | – | – | 5 |
| | Time Kenya – Haiti | – | – | – | – | 39 |
| | Rate Kenya – Haiti | – | – | – | – | 5 |
| | Time Mali – Kenya | – | – | – | – | 39 |
| | Rate Mali – Kenya | – | – | – | – | 5 |
| | Time Kenya – Mali | – | – | – | – | 39 |
| | Rate Kenya – Mali | – | – | – | – | 5 |
| Log Ln | – | -13.126 | -12.998 | -13.583 | -9.601 | -25.334 |
| AIC | – | 40.251 | 39.995 | 41.166 | 37.203 | 78.668 |
| Delta AIC | – | 3.048 | 2.792 | 3.963 | 0.000 | 41.465 |
| Probability of model | – | 0.218 | 0.248 | 0.138 | 1.000 | 0.000 |

**Table S5.** Primers used in the selective whole genome amplification (sWGA).

| Primer | Sequence |
|---|---|
| swga01 | AATCGATA*A*T |
| swga02 | ACGAATAA*T*T |
| swga03 | CGACGA*A*T |
| swga04 | CGATAC*G*A |
| swga05 | CGCGAA*A*A |
| swga06 | CGTAAA*C*G |
| swga07 | GACGAAAA*A*A |
| swga08 | TCGAAC*G*A |
| swga09 | TCGCGA*A*A |

**Table S6.** Pairwise population $F_{ST}$ with standard error (SE; upper diagonal).

| | Haiti | Mali | Kenya | PNG |
|---|---|---|---|---|
| Haiti | | 0.0107 | 0.0163 | 0.0176 |
| Mali | 0.103 | | 0.0073 | 0.0106 |
| Kenya | 0.112 | 0.098 | | 0.0047 |
| PNG | 0.198 | 0.255 | 0.241 | |

PNG: Papua New Guinea.

**Table S7.** mRNA ontology for gene identified during selection scan.

| Annotation | Panther | WBGene | UniProtKB | Gene | Function | Ortholog |
|---|---|---|---|---|---|---|
| snap_masked-Wb_Chr1_0-processed-gene-74.21-mRNA-1 | PTHR12243:SF17 | WBGene00011575 | G5EGG9 | Madf-4 | Alcohol dehydrogenase transcription factor Myb/SANT-like | *C. elegans* |
| maker-Wb_Chr1_0-snap-gene-74.10-mRNA-1 | PTHR13554:SF10 | WBGene00009445 | Q20058 | CELE_F35G12.12 | 26S Proteasome Non-atpase Regulatory subunit 5 | *C. elegans* |
| snap_masked-Wb_Chr4_0-processed-gene-101.20-mRNA-1 | PTHR28599:SF1 | WBGene00102523 | H3ET47 | SMIM-1 | Small Integral Membrane Protein 12 | *P. pacificus* |
| maker-Wb_Chr4_1-snap-gene-8.18-mRNA-1 | PTHR23128:SF136 | WBGene00008126 | O45306 | Sre-1 | Serpentine receptor class epsilon-21 | *C. elegans* |
| maker-Wb_Chr4_2-snap-gene-3.8-mRNA-1 | PTHR12411:SF316 | WBGene00007055 | O16454 | Tag-196 | Cathepsin F;cysteine protease | *C. elegans* |
| maker-Wb_Chr2_1-snap-gene-1.20-mRNA-1 | PTHR45975:SF2 | WBGene00009180 | Q6BER5 | Nurf-1 | Nucleocome-remodeling factor subunit BPTF | *C. elegans* |
| maker-Wb_Chr2_1-snap-gene-1.21-mRNA-1 | NA | NA | NA | HYP | NA | *C. elegans* |
| augustus_masked-Wb_Chr2_1-processed-gene-1.2-mRNA-1 | PTHR36562:SF1 | WBGene00013260 | Q9U213 | Rsr-2 | Serine/Arginine Repetitive Matrix 2 | *C. elegans* |
| maker-Wb_Chr2_1-augustus-gene-1.9-mRNA-1 | PTHR45975:SF2 | WBGene00009180 | Q6BER5 | Nurf-1 | Nucleocome-remodeling factor subunit BPTF | *C. elegans* |
| maker-Wb_Chr2_1-snap-gene-1.23-mRNA-1 | PTHR45975:SF2 | WBGene00009180 | Q6BER5 | Nurf-1 | Nucleocome-remodeling factor subunit BPTF | *C. elegans* |
| maker-Wb_Chr2_1-augustus-gene-1.14-mRNA-1 | PTHR11616:SF20 | WBGene00004902 | G5EBN9 | Snf-3 | Sodium- and chloride-dependent betaine transporter | *C. elegans* |
| maker-Wb_Chr4_0-augustus-gene-1.51-mRNA-1 | PTHR11550 | WBGene00012316 | G5EC98 | Ctps-1 | CTP Synthase | *C. elegans* |
| augustus_masked-Wb_ChrX_0-processed-gene-85.63-mRNA-1 | PTHR45624:SF4 | WBGene00000996 | Q27257 | Dif-1 | Congested-like trachea protein-related | *C. elegans* |
| maker-Wb_ChrX_0-snap-gene-85.5-mRNA-1 | PTHR13773 | WBGene00016384 | P53439 | Cdgs-1 | Phosphatidate Cytidylyltransferase | *C. elegans* |
| augustus_masked-Wb_ChrX_0-processed-gene-85.69-mRNA-1 | PTHR21723 | WBGene00004363 | Q22472 | Ric-3 | Resistance to inhibitors of cholinesterase protein 3 | *C. elegans* |
| snap_masked-Wb_ChrX_0-processed-gene-86.16-mRNA-1 | PTHR12322:SF57 | WBGene00019521 | O01582 | Dmd-7 | Doublesex/MAB-3 | *C. elegans* |
| snap_masked-Wb_ChrX_0-processed-gene-86.14-mRNA-1 | PTHR10844:SF21 | WBGene00000301 | Q94051 | Cav-1 | Caveolin-1 | *C. elegans* |
| snap_masked-Wb_ChrX_0-processed-gene-86.31-mRNA-1 | NA | NA | NA | HYP | NA | *C. elegans* |
| augustus_masked-Wb_Chr2_0-processed-gene-73.1-mRNA-1 | NA | NA | NA | HYP | NA | *C. elegans* |
| maker-Wb_ChrX_0-augustus-gene-91.50-mRNA-1 | PTHR10027:SF28 | WBGene00004830 | | Slo-1 | Large-conductance calcium-activated potassium channel | *C. elegans* |

**Table S8.** Results of enrichment test for genes from selection scan using Panther DB.

| **PANTHER GO-Slim Biological process** | ***Caenorhabditis elegans* REFLIST (19921)** | **Wb.panther. out (17)** | **Wb.panther. out (expected)** | **Wb.panther. out (over/under)** | **Wb.panther. out (fold Enrichment)** | **Wb.panther. out (raw *P*-value)** |
|---|---|---|---|---|---|---|
| Ribonucleoside triphosphate metabolic process (GO:0009199) | 2 | 1 | 0 | + | > 100 | 0.00256 |
| Ribonucleoside triphosphate biosynthetic process (GO:0009201) | 2 | 1 | 0 | + | > 100 | 0.00256 |
| Membrane assembly (GO:0071709) | 3 | 1 | 0 | + | > 100 | 0.00341 |
| Ribonucleotide biosynthetic process (GO:0009260) | 4 | 1 | 0 | + | > 100 | 0.00426 |
| Ribonucleotide metabolic process (GO:0009259) | 5 | 1 | 0 | + | > 100 | 0.00511 |
| Pyrimidine nucleotide metabolic process (GO:0006220) | 6 | 1 | 0.01 | + | > 100 | 0.00595 |
| Nucleobase-containing small molecule biosynthetic process (GO:0034404) | 7 | 1 | 0.01 | + | > 100 | 0.0068 |
| Nucleoside triphosphate metabolic process (GO:0009141) | 8 | 1 | 0.01 | + | > 100 | 0.00765 |
| Endomembrane system organization (GO:0010256) | 13 | 1 | 0.01 | + | 90.14 | 0.0119 |
| Pyrimidine nucleobase metabolic process (GO:0006206) | 13 | 1 | 0.01 | + | 90.14 | 0.0119 |
| Heterocycle metabolic process (GO:0046483) | 16 | 1 | 0.01 | + | 73.24 | 0.0144 |
| Pyrimidine-containing compound metabolic process (GO:0072527) | 16 | 1 | 0.01 | + | 73.24 | 0.0144 |
| Nucleobase-containing small molecule metabolic process (GO:0055086) | 18 | 1 | 0.02 | + | 65.1 | 0.0161 |
| Organonitrogen compound biosynthetic process (GO:1901566) | 47 | 1 | 0.04 | + | 24.93 | 0.0402 |

**Table S9.** Distribution of mean identity by descent (IBD) lengths within *W. bancrofti* populations.

| Population | IBD IN | IBD BETWEEN |
|---|---|---|
| Haiti | 6,163,370 | 5,068,997 |
| Mali | 5,281,618 | 5,188,573 |
| Kenya** | 11,721,362 | 5,511,120 |
| PNG** | 6,788,868 | 1,340,033 |

** significantly different.

PNG: Papua New Guinea.