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

## Supplementary information

**Table S1. The environmental and socio-economic variables that were included in the initial analysis.** The resolution, time period and sources of each of the variables are provided on the table.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Categories** | **Covariates** | **Resolution** | **Source** | **Time** | **Unit** | **References** |
| Temperature | BIO1 = Annual Mean Temperature | 30 arc second (~1 km) | WorldClim V1 Bioclim | 1970-2000 | °C | (Fick and Hijmans 2017) |
| BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)) |
| BIO4 = Temperature Seasonality (standard deviation × 100)+ |
| BIO5 = Max Temperature of Warmest Month |
| BIO6 = Min Temperature of Coldest Month+ |
| BIO7 = Temperature Annual Range (BIO5-BIO6) |
| BIO8 = Mean Temperature of Wettest Quarter |
| BIO9 = Mean Temperature of Driest Quarter |
| BIO10 = Mean Temperature of Warmest Quarter |
| BIO11 = Mean Temperature of Coldest Quarter |
| BIO3 = Isothermality (BIO2/BIO7) (×100)\* | % |
| Land surface temperature (day) | MOD11A1.006 Terra Land Surface Temperature and Emissivity Daily Global 1km | 2000-2001, 2010-2012, and 2013-2015 | °C (converted from K) | (Wan, Zhengming, Hook, Simon, and Hulley, Glynn 2015) |
| Land surface temperature (night)+ |
| Precipitation | BIO12 = Annual Precipitation\*+ | 30 arc second (~1 km) | WorldClim V1 Bioclim | 1970-2000 | mm | (Fick and Hijmans 2017) |
| BIO13 = Precipitation of Wettest Month |
| BIO14 = Precipitation of Driest Month |
| BIO16 = Precipitation of Wettest Quarter |
| BIO17 = Precipitation of Driest Quarter |
| BIO18 = Precipitation of Warmest Quarter |
| BIO19 = Precipitation of Coldest Quarter |
| BIO15 = Precipitation Seasonality (Coefficient of Variation) | coefficient of variation |
| Topographical | Digital Elevation Model (DEM)\* | 30 arc second (~1 km) | CGIAR-SRTM | 2000 | meters | (Farr et al. 2007) |
| Slope+ | topographic slope in degree |
| Vegetation indices | NDVI | 30 arc second (~1 km) | MOD13A2.006 Terra Vegetation Indices 16-Day Global 1km | 2000-2001, 2010-2012, and 2013-2015 | NA | (Didan, Kamel 2015) |
| EVI |
| Hydrological data | Flow accumulation\* | 30 arc sec (~1 km) | HydroSHEDS, WWF | 2000 | Number of cells | (Lehner, Verdin, and Jarvis 2006, 2008) |
| Distance to the nearest river (derived from water lines in DIVA-GIS) + | 2003 | km | (Hijmans et al. 2001) |
| Taselled cap wetness index | 5 km | Malaria Atlas Project | 2001-2015 | NA | (Lobser and Cohen 2007) |
| Soil moisture\*+ | TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces | 2000-2001, 2010-2012, and 2013-2015 | mm | (Abatzoglou et al. 2018) |
| Socio-demographic | Population density (UN - adjusted) | 30 arc second (~1 km) | Gridded Population of World Version 4 (GPWv4) | 2000-2001, 2010-2012, and 2013-2015 | density | (Center For International Earth Science Information Network-CIESIN-Columbia University 2017) |
| Improved housing (prevalence) + | 30 arc second (~1 km) | Malaria Atlas Project | 2001, 2001-2015, and 2015 | % | (Tusting et al. 2019) |
| Nighttime lights | 100 m | WorldPop (Resampled DMSP, OLS) | 2000-2001, 2010-2012, and 2013 | NA | (Hsu et al. 2015) |

\* Variables selected for the landscape genetics analysis; + Variables selected for prevalence mapping

Chart, histogram

Description automatically generated

**Figure S1. Principle component analysis on the environmental variables from the locations of prevalence data**. Variable correlation plot (A) showing the extent and the direction of the relation between variables. Scree plot (B) is showing the percentage of variance explained by the principal components and the first five dimensions explain >80% of the total variance. The contribution of the variables in explaining the variability (C) of the first five principal components was used to select the correlated variables.

Chart

Description automatically generated**Figure S2. Correlation matrix between the 32 environmental variables from the sampling locations used for the prevalence data analysis**. The name of the environmental variables is presented on their acronym format.

Chart

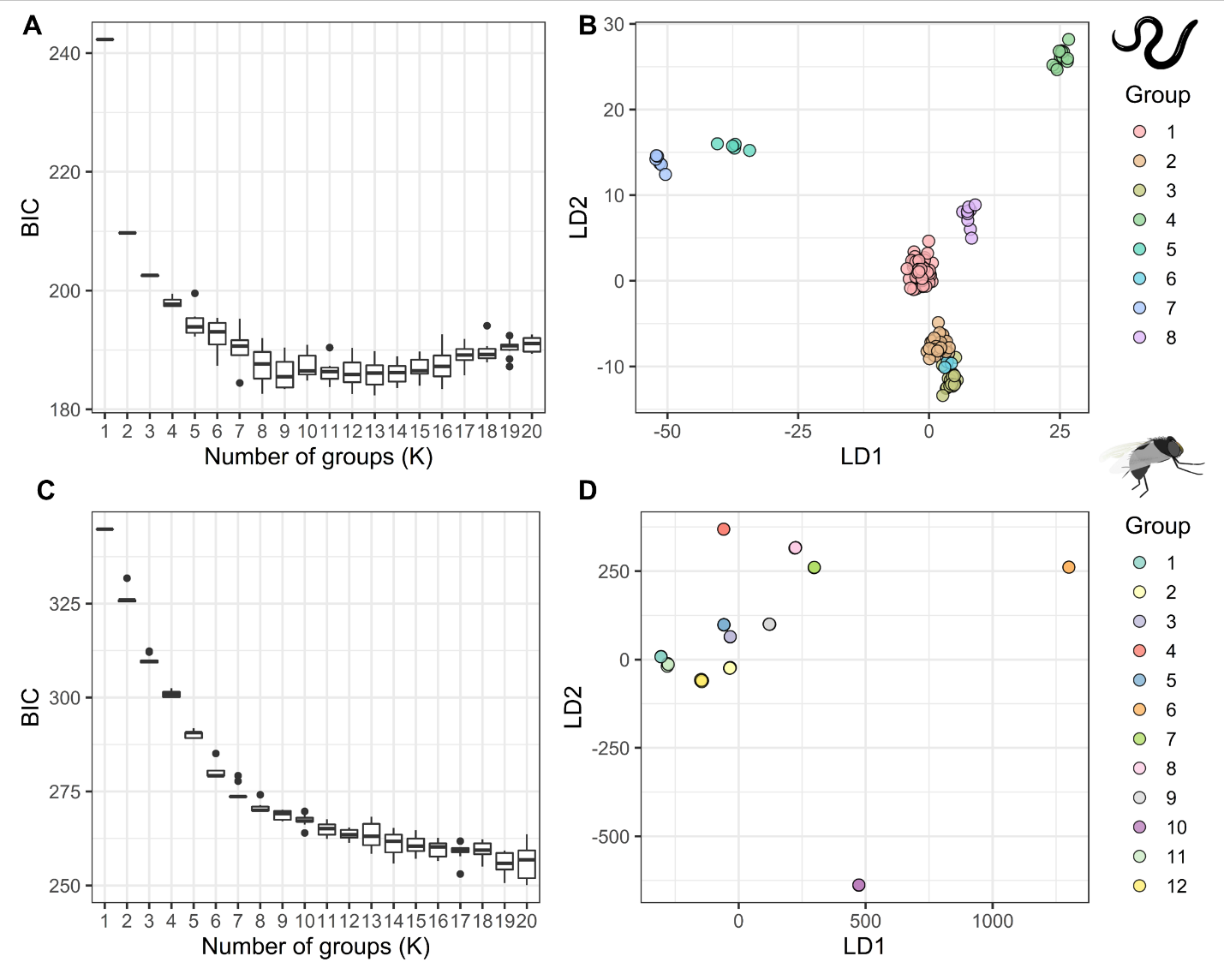
Description automatically generated

**Figure S3. Principle component analysis on the mean of the environmental variables along the straight path distance between the sampling locations for landscape genetics**. Variable correlation plot (A) showing the extent and the direction of the relation between variables. Scree plot (B) is showing the percentage of variance explained by the principal components and the first two dimensions explain >80% of the total variance. The contribution of the variables in explaining the variability (C) of the first five principal components was used to select the correlated variables.

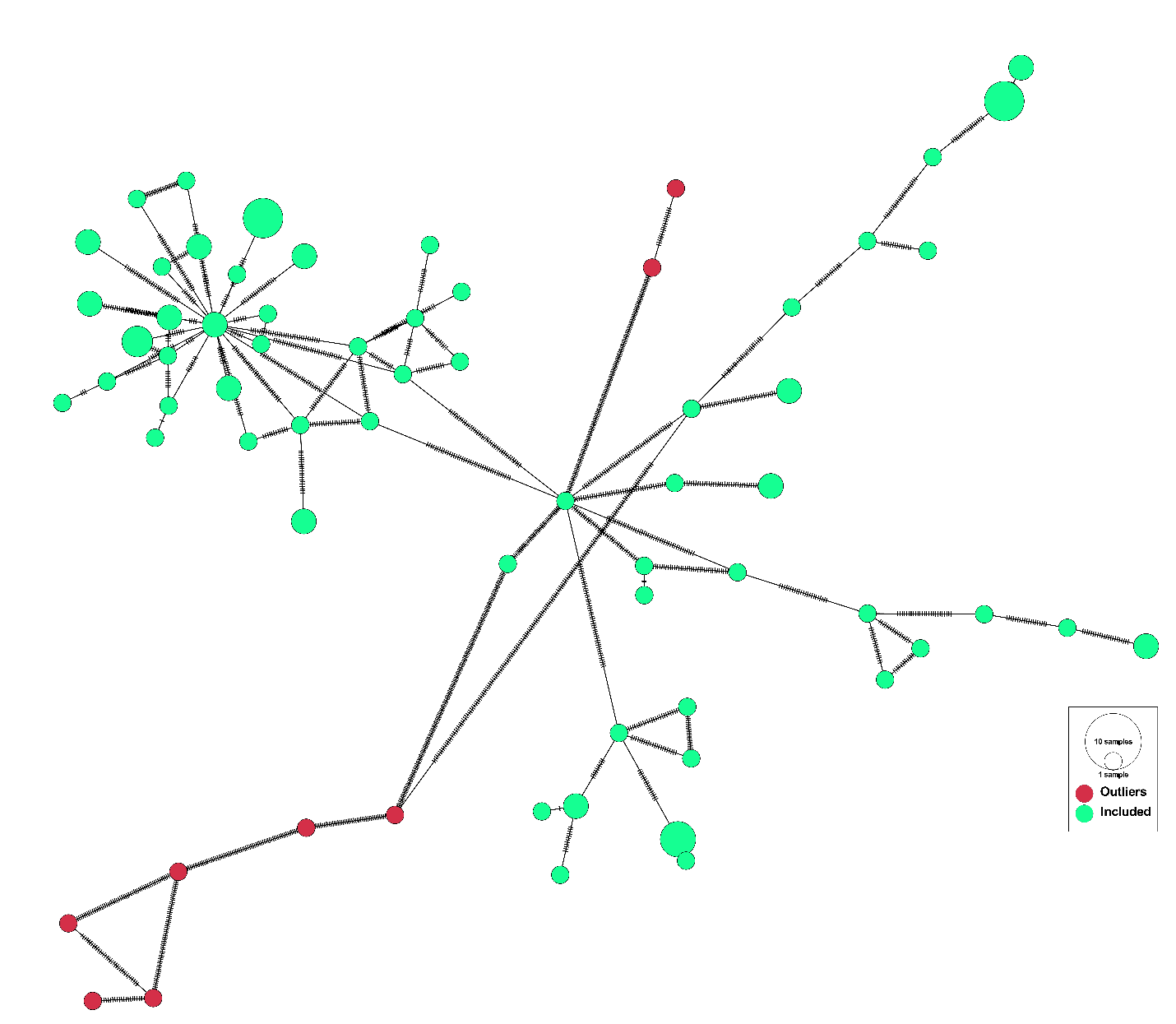
Chart

Description automatically generated

**Figure S4. Correlation matrix between the 32 environmental variables extracted as the mean value encountered along the straight path distance between the sampling locations used for the landscape genetics analysis**. The name of the environmental variables is presented on their acronym format.



**Figure S5. Clustering analysis on the *O. volvulus* and the *S. damnosum* genetic data.** The optimum number of clusters were determined with the help of BIC scores (**A, C**) and the discriminant analysis of the principle component (DAPC) was done on the inferred clusters (**B, D**) for both the parasite and the vector data.



**Figure S6. Haplotype network analysis of the *S. damnosum* samples showing the location of outliers in the haplotype network.** The red nodes are the outlier samples and the number of hatches between the nodes indicate variation between the nodes. The number of hatches for the edges connecting to outlier nodes are substantially high.

**Table S2. Mean estimates of the posterior regression coefficients for each environmental variables and the hyperparameters while mapping the baseline *O. volvulus* infection prevalence**. The bold environmental covariates are significant based on 95% BCI (Bayesian Credible Interval).

|  |  |  |
| --- | --- | --- |
| **Variables** | **Regression coefficients** | |
| **Mean** | **95% BCI** |
| **Slope** | **2.126** | **(0.032, 4.338)** |
| **Soil moisture** | **0.043** | **(0.004, 0.084)** |
| **Temperature seasonality** | **-0.022** | **(-0.044, -0.001)** |
| Improved housing prevalence | 14.300 | (-31.576, 59.597) |
| Minimum temperature coldest month | 0.249 | (-0.021, 0.535) |
| Distance to the nearest river | 0.174 | (-0.348, 0.701) |
| Annual precipitation | -0.024 | (-0.053, 0.005) |
| Land surface temperature night | -0.777 | (-2.428, 0.807) |
| Intercept | 18.770 | (-36.469, 73.334) |
| **Range** | **4,396.350** | **(1660.26, 7881.19)** |
| **Variance** | **30.905** | **(11.06, 85.812)** |
| 95% BCI includes 0.025 quantiles and the 0.975 quantiles of the posterior probability distribution of the coefficients | | |

## References

Abatzoglou, John T., Solomon Z. Dobrowski, Sean A. Parks, and Katherine C. Hegewisch. 2018. ‘TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958–2015’. *Scientific Data* 5(1):170191. doi: 10.1038/sdata.2017.191.

Center For International Earth Science Information Network-CIESIN-Columbia University. 2017. ‘Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11’.

Didan, Kamel. 2015. ‘MOD13A2 MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid V006’.

Farr, Tom G., Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, David Seal, Scott Shaffer, Joanne Shimada, Jeffrey Umland, Marian Werner, Michael Oskin, Douglas Burbank, and Douglas Alsdorf. 2007. ‘The Shuttle Radar Topography Mission’. *Reviews of Geophysics* 45(2). doi: 10.1029/2005RG000183.

Fick, Stephen E., and Robert J. Hijmans. 2017. ‘WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas’. *International Journal of Climatology* 37(12):4302–15. doi: 10.1002/joc.5086.

Hijmans, Robert J., Luigi Guarino, Mariana Cruz, and Edwin Rojas. 2001. ‘Computer Tools for Spatial Analysis of Plant Genetic Resources Data: 1. DIVA-GIS’. *Plant Genetic Resources Newsletter* 15–19.

Hsu, Feng-Chi, Kimberly E. Baugh, Tilottama Ghosh, Mikhail Zhizhin, and Christopher D. Elvidge. 2015. ‘DMSP-OLS Radiance Calibrated Nighttime Lights Time Series with Intercalibration’. *Remote Sensing* 7(2):1855–76.

Lehner, Bernhard, Kris Verdin, and Andy Jarvis. 2006. ‘HydroSHEDS Technical Documentation, Version 1.0’. *World Wildlife Fund US, Washington, DC* 1–27.

Lehner, Bernhard, Kristine Verdin, and Andy Jarvis. 2008. ‘New Global Hydrography Derived From Spaceborne Elevation Data’. *Eos, Transactions American Geophysical Union* 89(10):93. doi: 10.1029/2008EO100001.

Lobser, SE, and WB Cohen. 2007. ‘MODIS Tasselled Cap: Land Cover Characteristics Expressed through Transformed MODIS Data’. *International Journal of Remote Sensing* 28(22):5079–5101.

Tusting, Lucy S., Donal Bisanzio, Graham Alabaster, Ewan Cameron, Richard Cibulskis, Michael Davies, Seth Flaxman, Harry S. Gibson, Jakob Knudsen, Charles Mbogo, Fredros O. Okumu, Lorenz von Seidlein, Daniel J. Weiss, Steve W. Lindsay, Peter W. Gething, and Samir Bhatt. 2019. ‘Mapping Changes in Housing in Sub-Saharan Africa from 2000 to 2015’. *Nature* 568(7752):391–94. doi: 10.1038/s41586-019-1050-5.

Wan, Zhengming, Hook, Simon, and Hulley, Glynn. 2015. ‘MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006’.