Global-scale population genetic analysis of *Plasmodium falciparum* identifies country- and region-specific patterns of malaria parasite adaptation

Nina Billows[1]	nina.billows@lshtm.ac.uk
Jamille G. Dombrowski[2]	jamille.dombrowski@gmail.com
Joseph Thorpe[1]	Joseph.Thorpe@lshtm.ac.uk
Leen Vanheer[1]	Leen.Vanheer@lshtm.ac.uk
Sophie Moss[1]	Sophie.Moss@lshtm.ac.uk
Jesse Gitaka[4]	jgitaka@mku.ac.ke
Colin J. Sutherland[1,5]	colin.sutherland@lshtm.ac.uk
Claudio R. F. Marinho[2]	marinho@usp.br
Nguyen Thi Hong Ngoc [6]	ngocnimpe@gmail.com
Nguyen Thi Huong Binh[6]	nguyenhuongbinh74@gmail.com
Nguyen Quang Thieu[6]	thieunq@gmail.com
Susana Campino [1,*]	Susana.campino@lshtm.ac.uk
Taane G. Clark [1,7,*]	taane.clark@lshtm.ac.uk

[1] Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2] Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil.

[3] School of Medicine, Instituto de Medicina Tropical, University of São Paulo, São Paulo, Brazil.

[4] Directorate of Research and Innovation, Mount Kenya University, Gen. Kago Rd, Thika, Kenya.

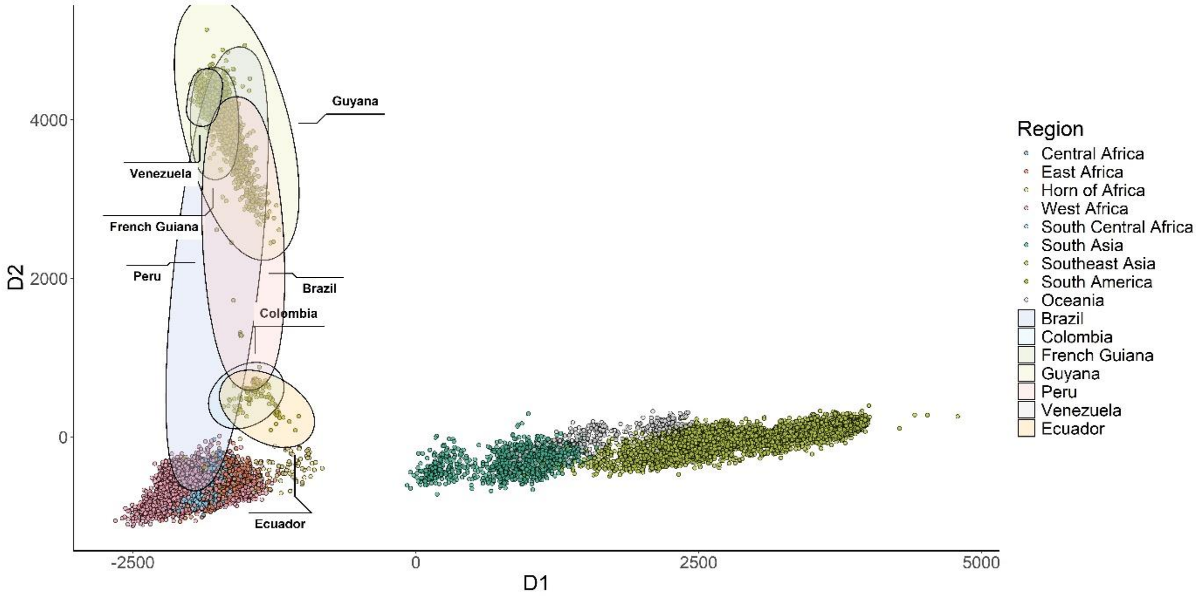
[5] UK Health Security Agency Malaria Reference Laboratory, LSHTM, London, WC1E 7HT, UK.

[6] Molecular Biology Department, Parasitology and Entomology, Vietnam National Institute of Malariology, Hanoi, Vietnam.

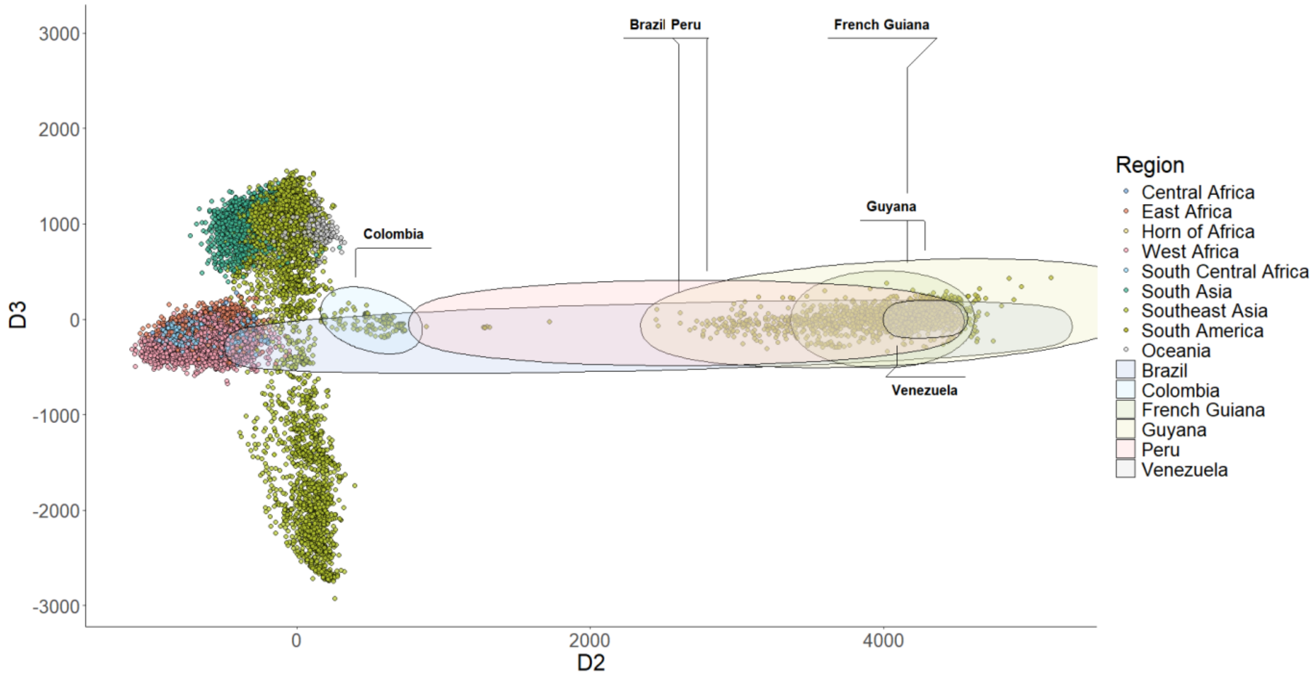
[7] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

SUPPLEMENTARY FIGURES

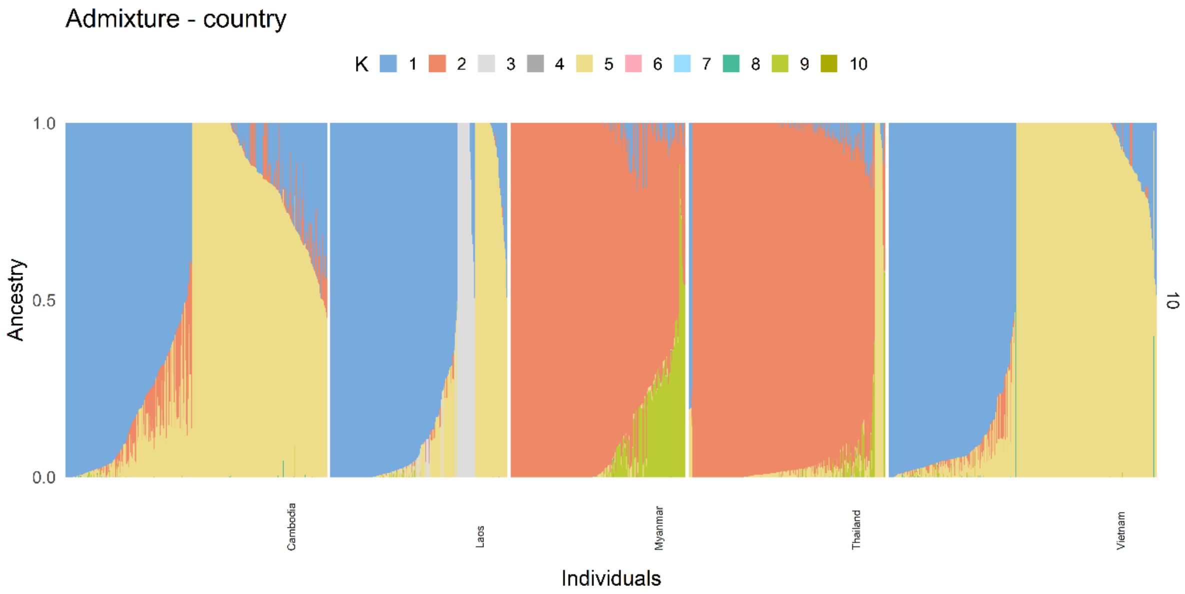
A



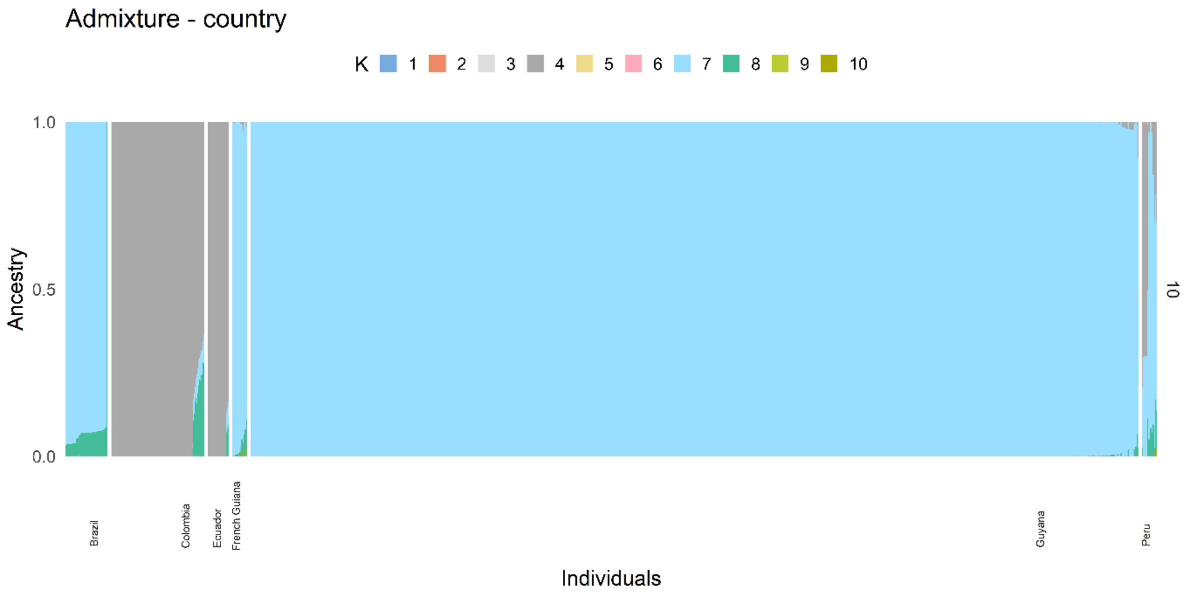
B



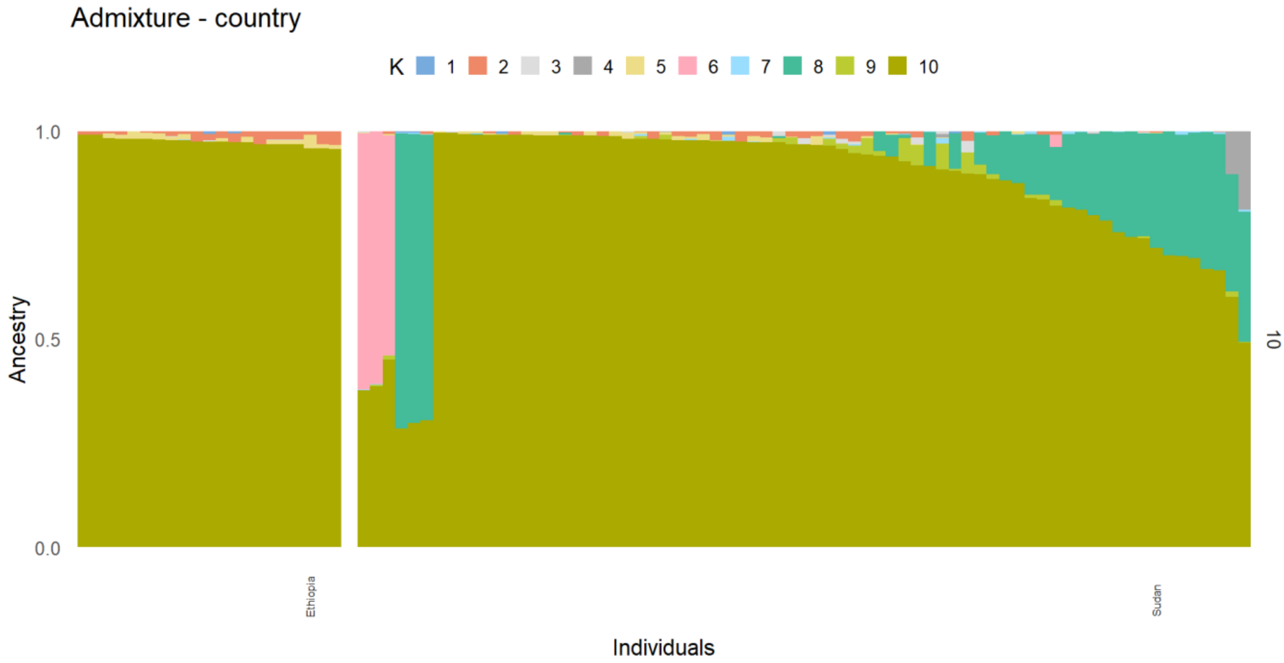
Supplementary Figure 1. Population structure of global *P. falciparum* samples showing separation amongst South American samples. Multidimensional scaling: D1 (19.7%) vs D2 (5.7%) (A) and D2 (5.7%) vs D3 (3.1%) (B) are shown and overlap between samples from South American countries are highlighted and labelled. Note for Panel B Colombia is an overlapping cluster between samples from Colombia and Ecuador.



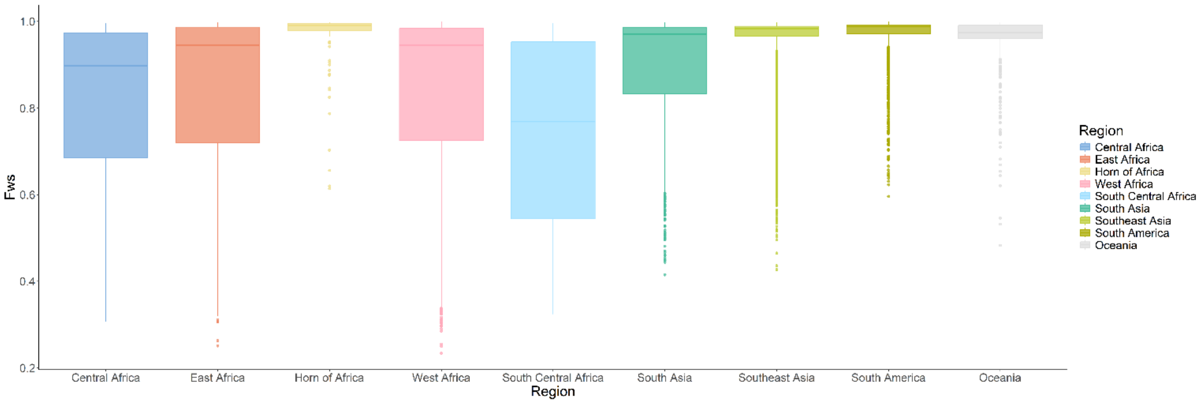
Supplementary Figure 2. Admixture within Southeast Asia. Samples underwent admixture analysis (K=10). Dominant clusters are shown for Southeast Asian samples, representing diverse ancestral populations.



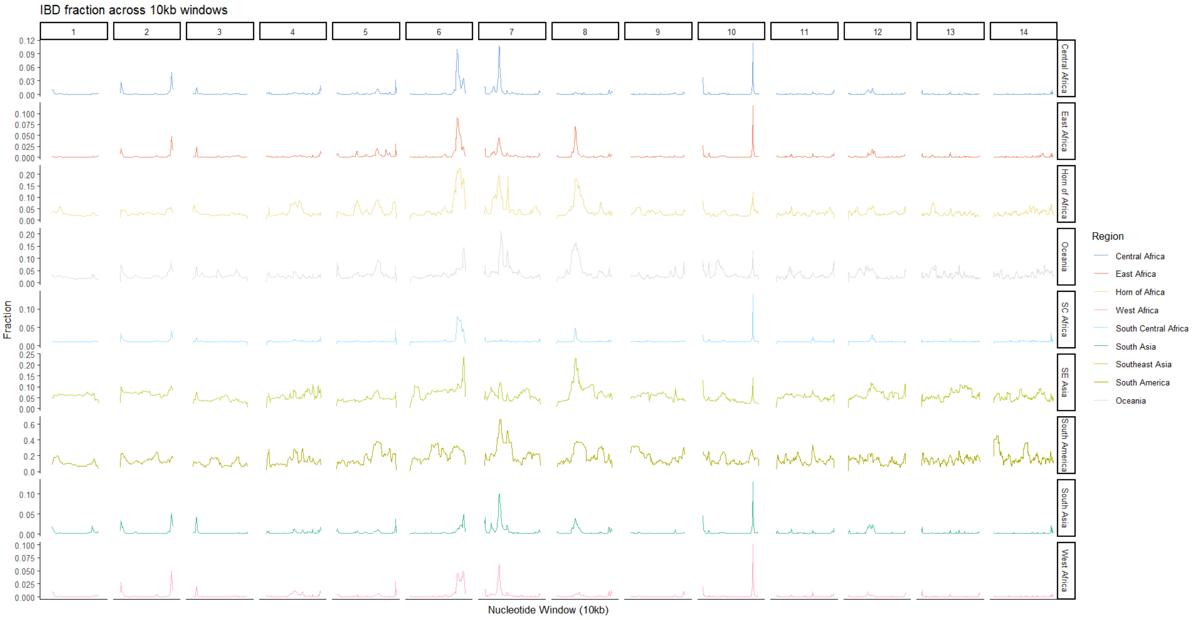
Supplementary Figure 3. Admixture within South America. Samples underwent admixture analysis (K=10). Dominant clusters are shown for South American samples, representing diverse ancestral populations with distinct clusters.



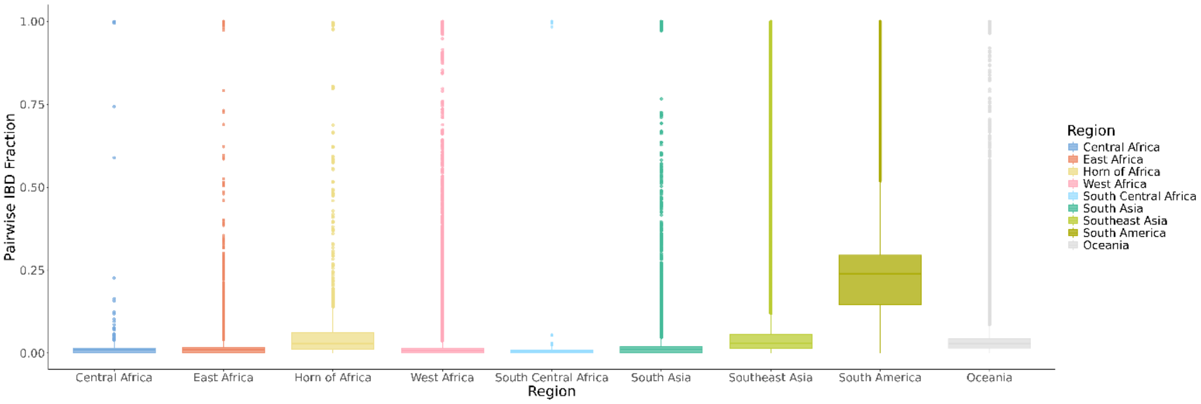
Supplementary Figure 4. Admixture within Horn of Africa. Samples underwent admixture analysis (K=10). Dominant clusters are shown for Horn of Africa samples, representing diverse ancestral populations with distinct clusters.



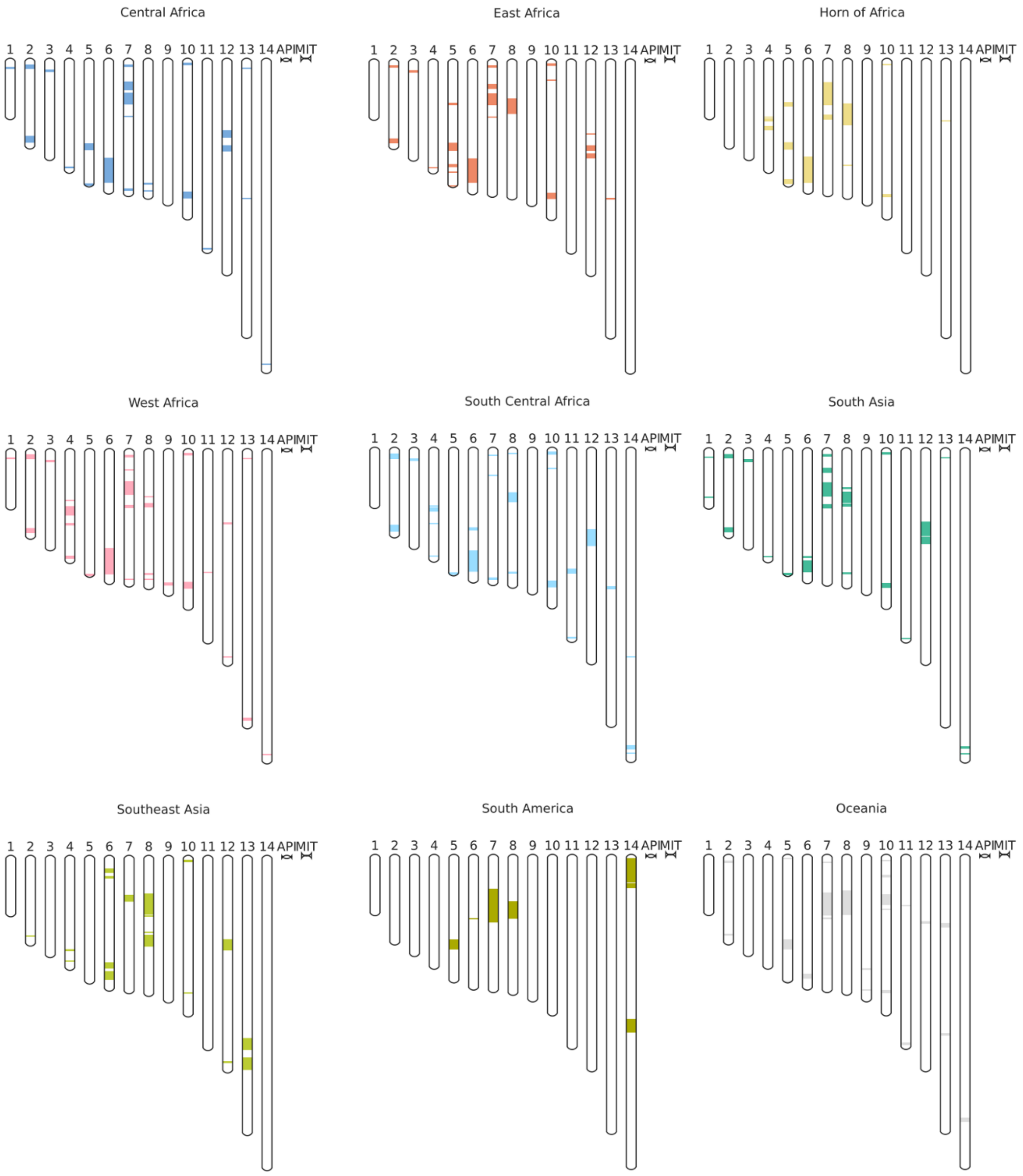
Supplementary Figure 5. Distribution of F_{WS} scores for *P. falciparum* per region. Boxplots showing the median, interquartile range and range of F_{WS} scores are highlighted per region.



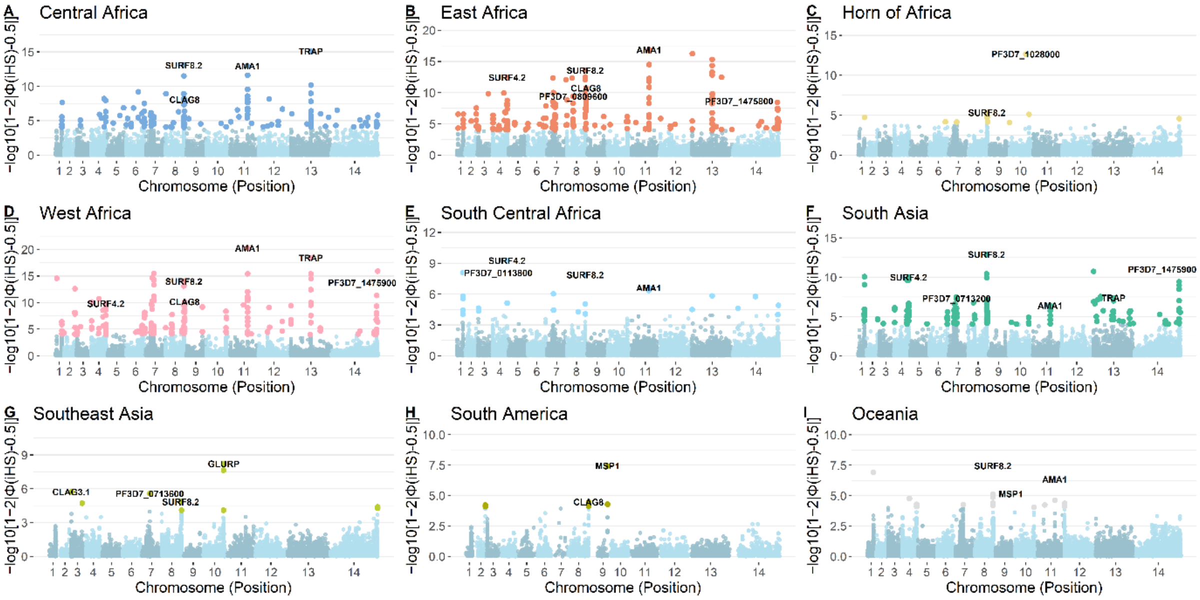
Supplementary Figure 6. Fraction of Identity by Descent (IBD) across 10kb windows. IBD fractions are shown across 14 nuclear chromosomes for 9 geographical regions. SC= South Central; SE= Southeast.



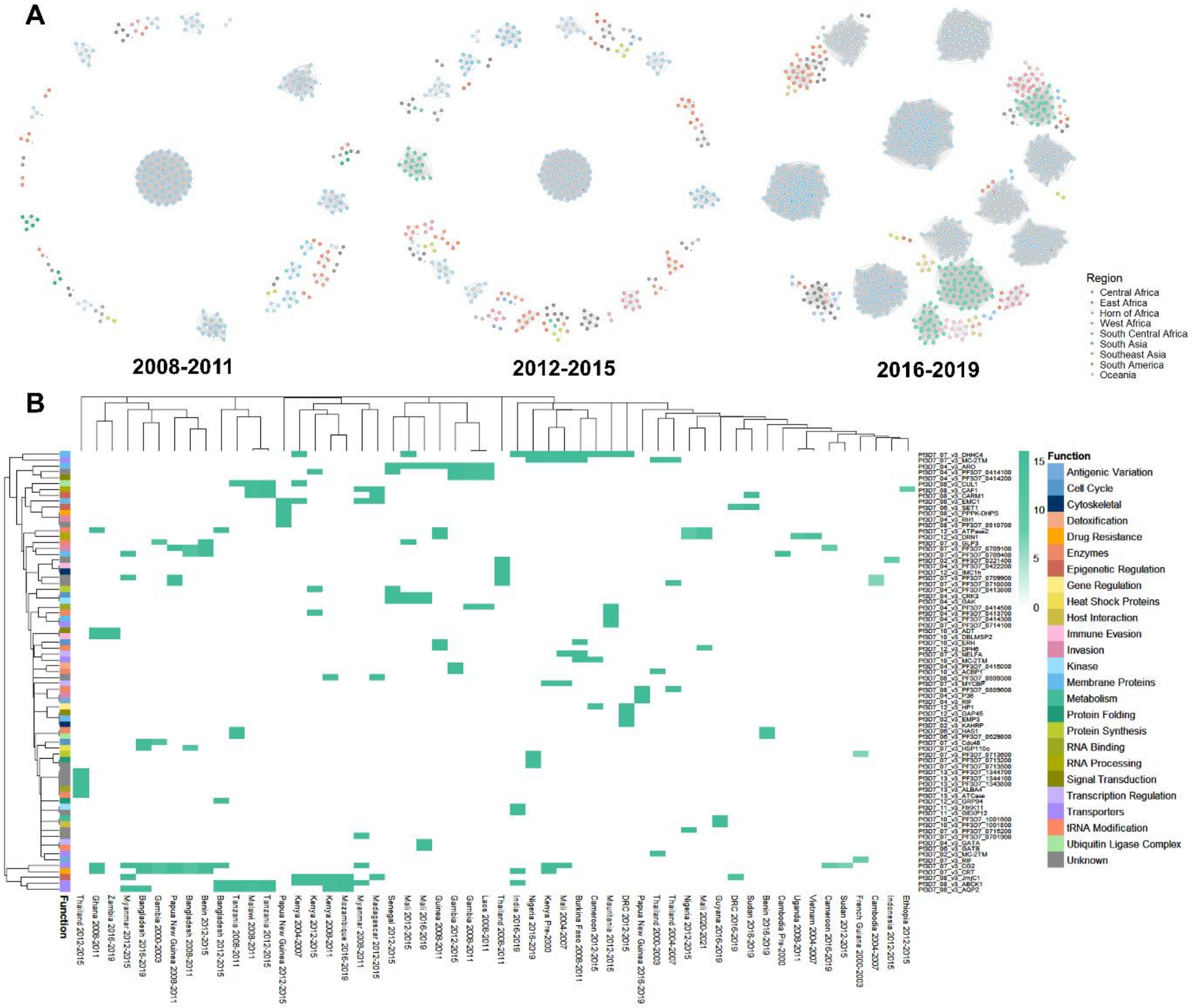
Supplementary Figure 7. Pairwise Fractions of Identity by Descent (IBD) across regions. Boxplots showing, the median and interquartile range of pairwise IBD fractions (sample-sample comparisons) are shown across 14 nuclear chromosomes for 9 geographical regions.



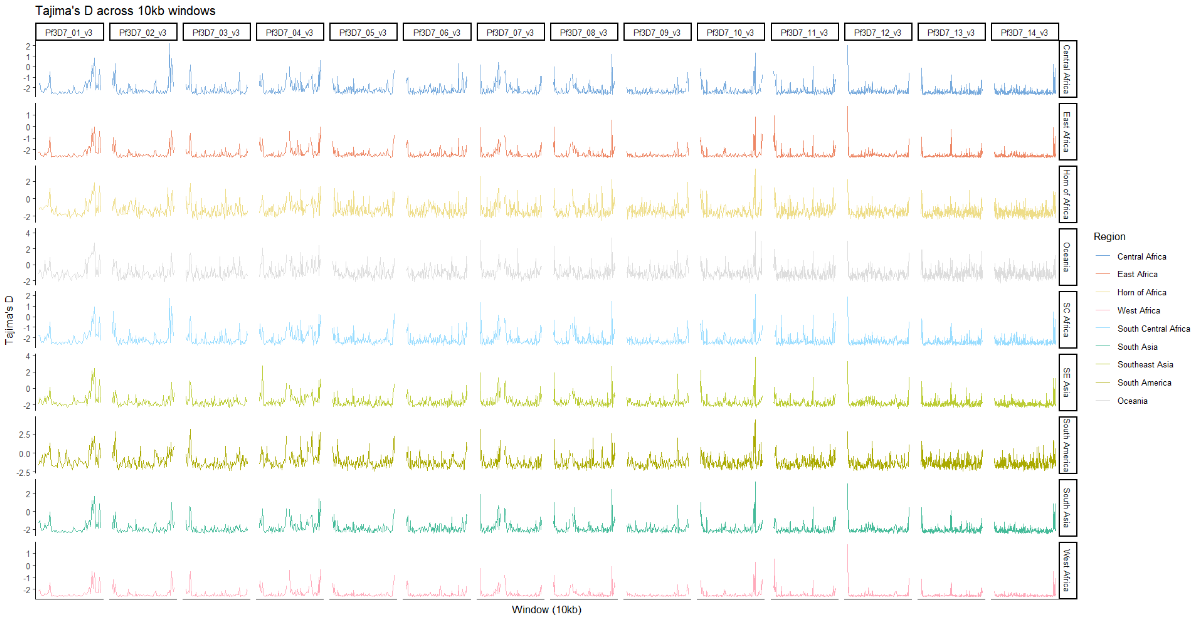
Supplementary Figure 8. Chromosomal regions with high fractions of Identity by Descent (IBD) for each region. Chromosome painting of regions across 14 nuclear chromosomes with high IBD fractions (0.95 quantile) are shown for 9 regions.



Supplementary Figure 9. Genomic regions under recent positive selection within single regions (iHS score). Genomic regions under recent positive selection across eight geographical regions, including Central Africa (A), East Africa (B), Horn of Africa (C), West Africa (D), South Central Africa (E), South Asia (F), Southeast Asia (G), South America (H), Oceania (I). Regions with $(-\log10[1-2\ |\Phi iHS-0.5\ |]) > 4.0$ are highlighted. Genes are annotated if they contain the top 5 hits or a sustained peak with greater than 5 significant SNPs.

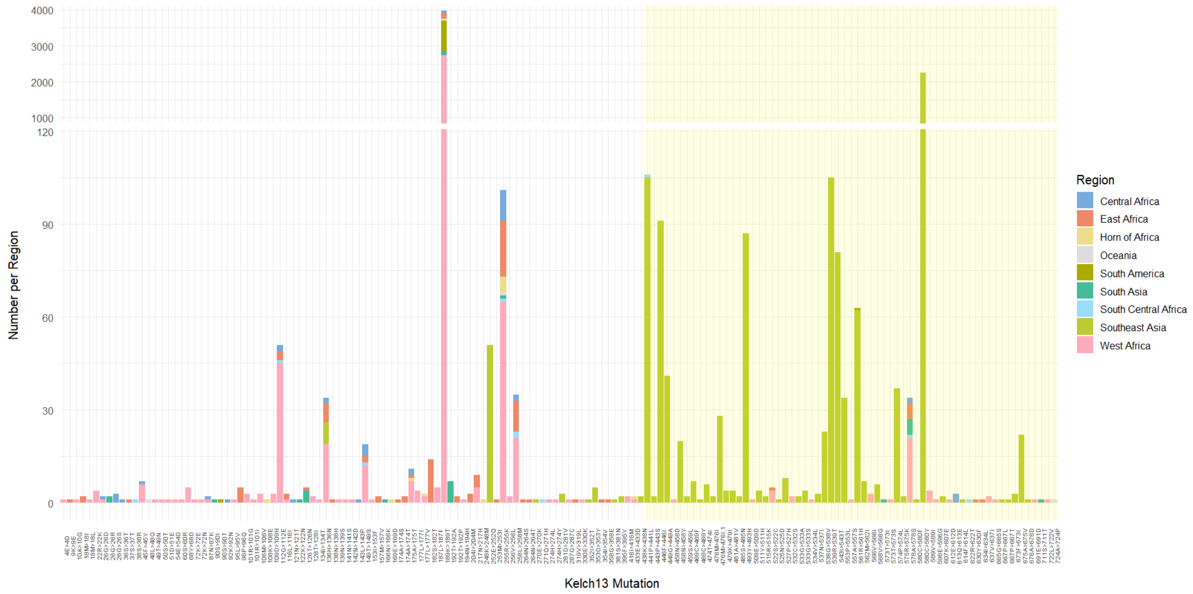


Supplementary Figure 10. Sample-level clustering based on IBD and loci under selection (iR) across countries and three-year time intervals. (A) Sample-level clustering based on identity-by-descent (IBD, proportion ≥ 0.95) across the three most common time intervals (2008–2011, 2012–2015, and 2016–2019), coloured by region. (B) Presence of loci under selection ('top hits') across countries and time intervals, with shade intensity reflecting $-\log_{10}$ P-values; 0 indicates no evidence of selection. Loci are annotated by their known or predicted function.

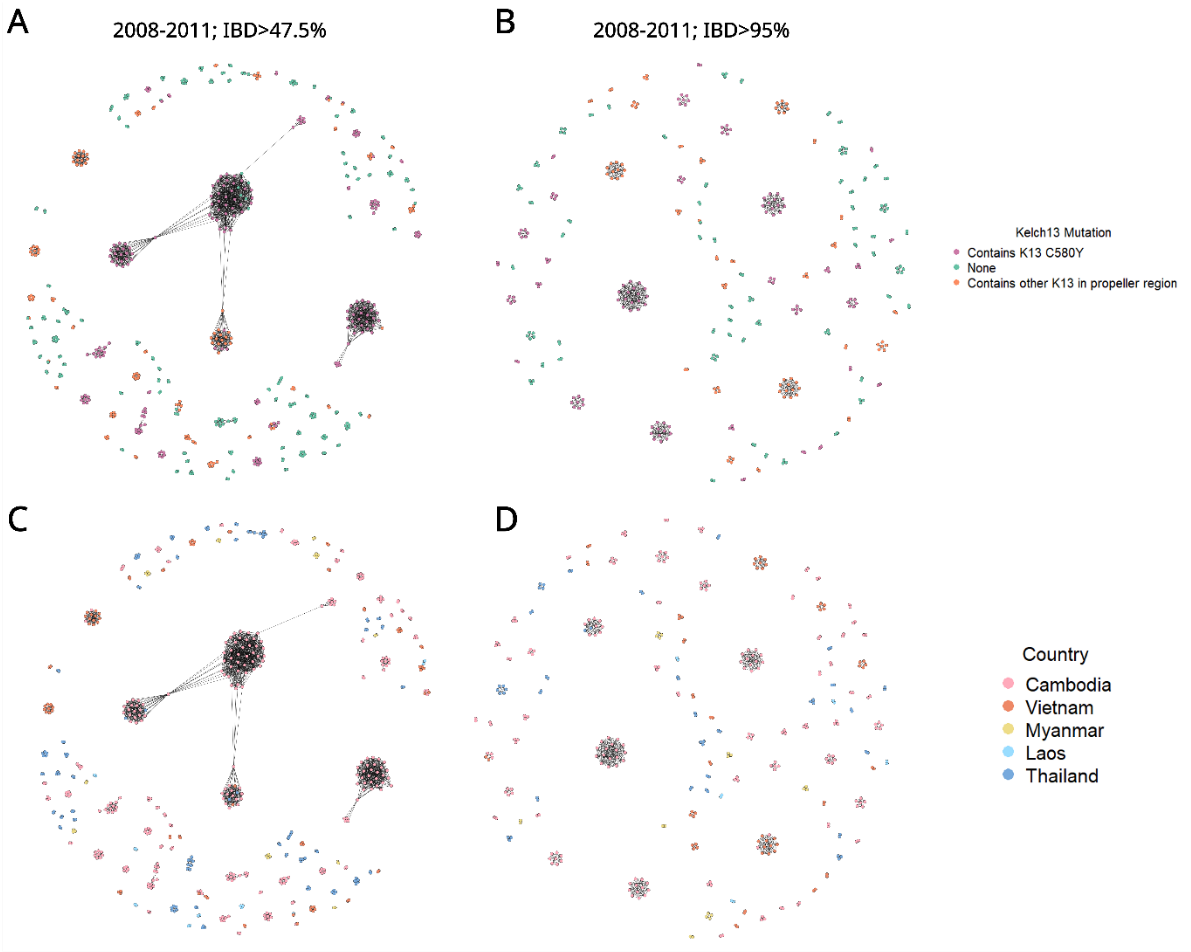


Supplementary Figure 11. Tajima's D calculations for 14 chromosomes and 9 geographic regions. Tajima's D calculations estimated using 10kb windows are shown for each chromosome and region. SC= South Central; SE= Southeast.

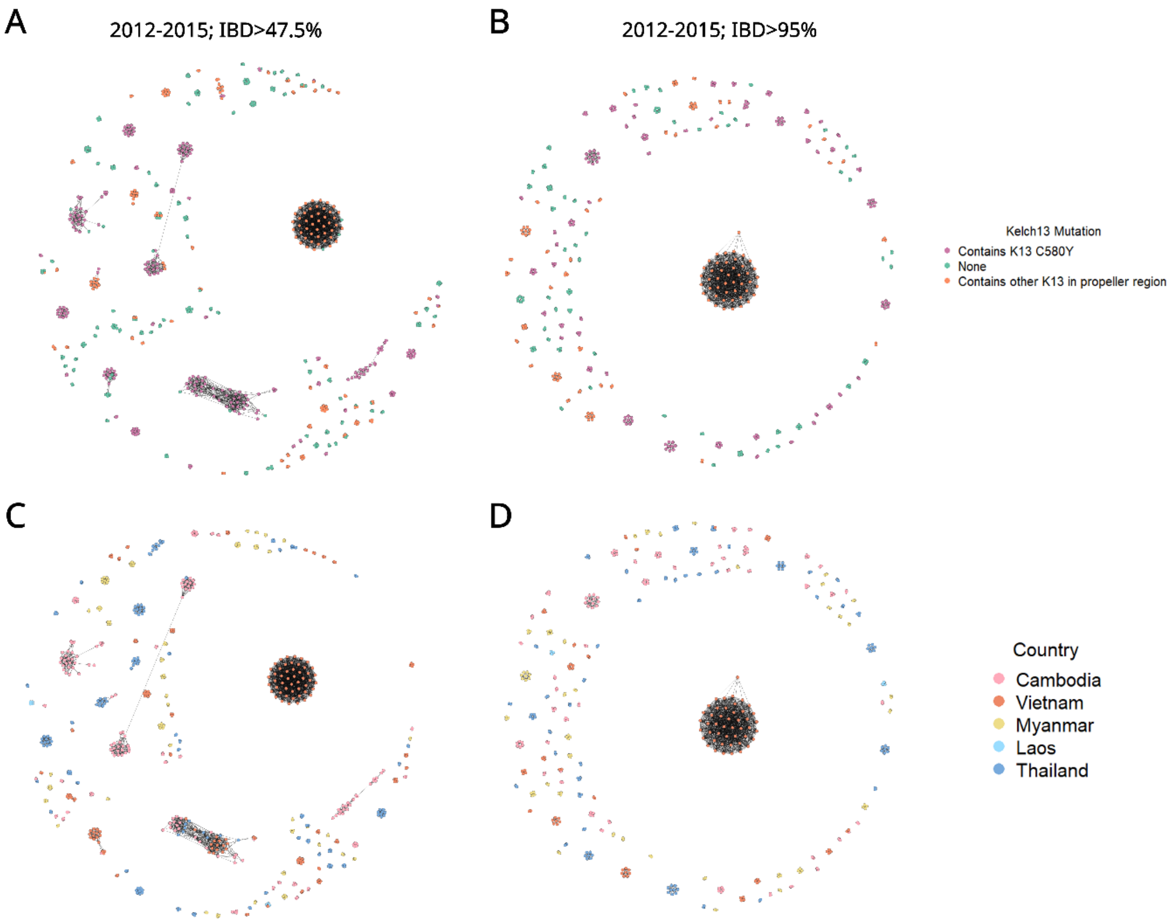
Supplementary Figure 12. Patterns of co-occurring drug-resistant mutations per region. A heatmap showing combinations of drug-resistant mutations across *pfcrt, pfmdr1, pfdhfr, pfdhps* and *pfk13* are shown with their relative numbers (-log(N)) amongst 9 geographical regions. Grey indicates site is missing due to poor genome coverage. *_2 indicates a second allele causes amino acid substitution.



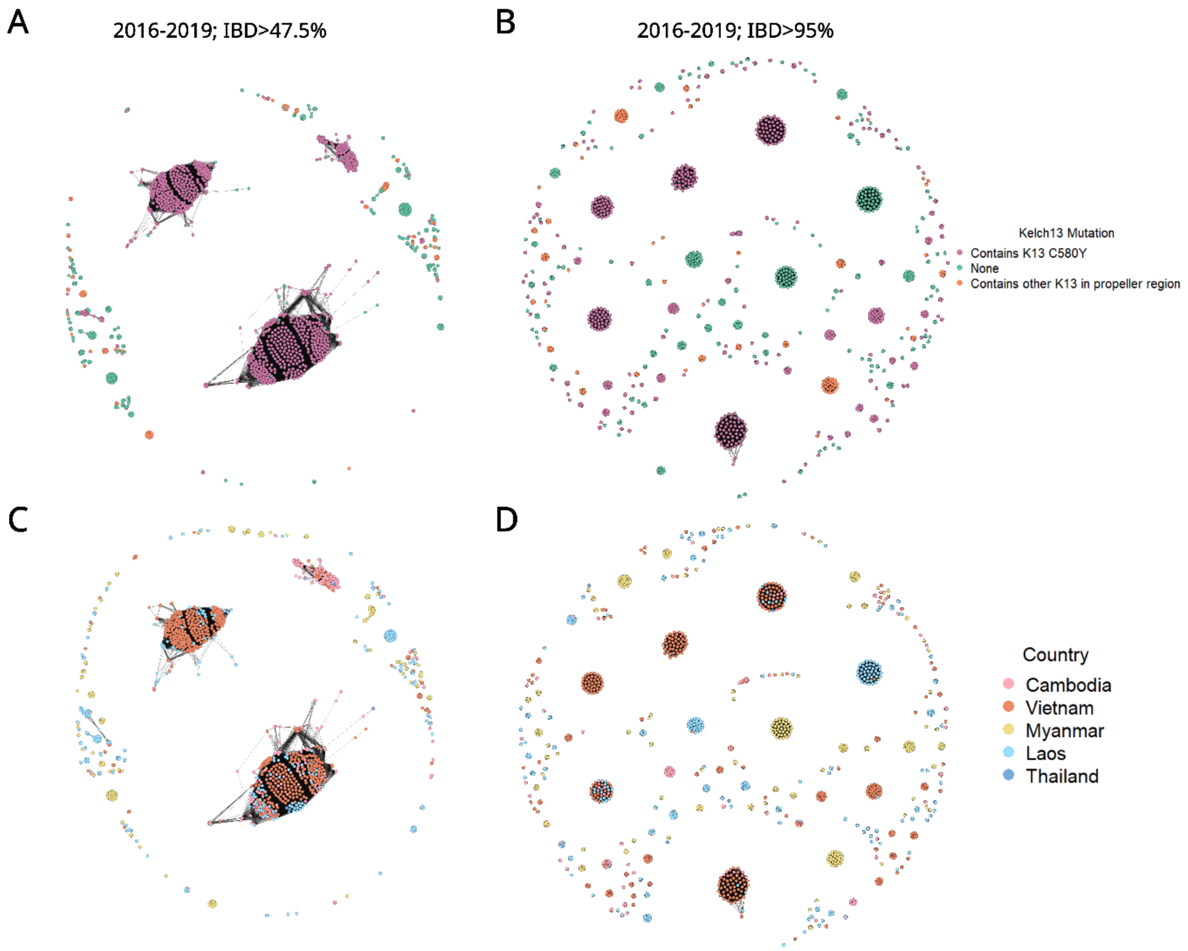
Supplementary Figure 13. Number of missense mutations across the *pfk13* gene by region. The propeller domain is highlighted to indicate its location relative to mutation density.



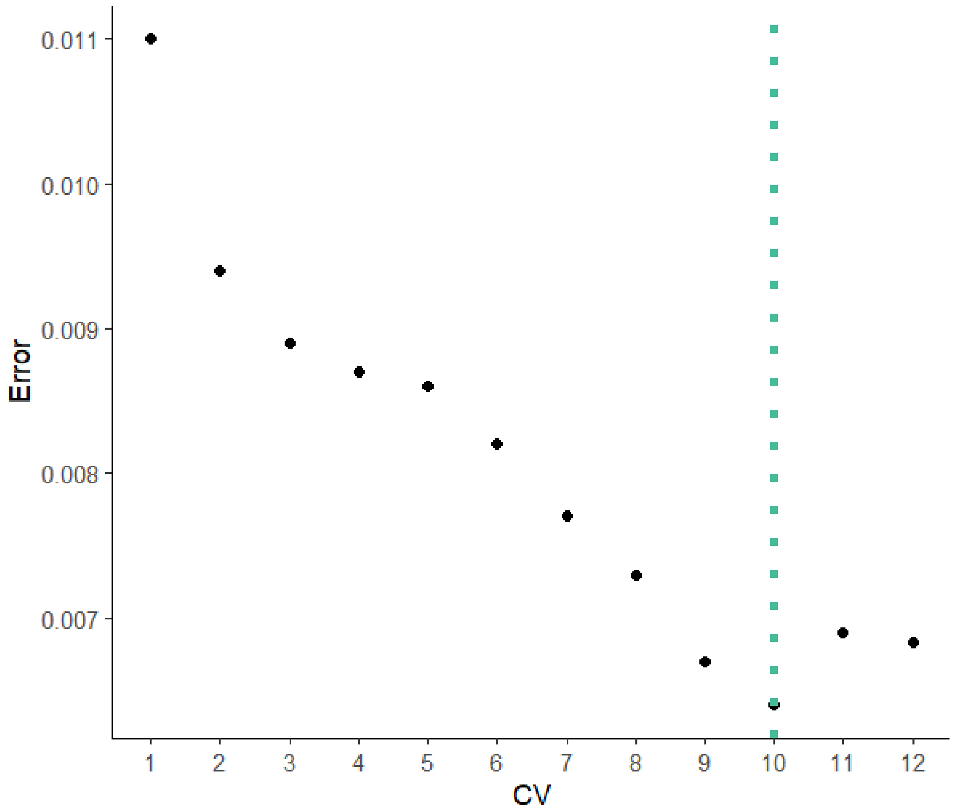
Supplementary Figure 14. Identity-by-descent (IBD) networks in Southeast Asia showing clusters at the 47.5% and 95% pairwise IBD thresholds (2008-2011). IBD networks were constructed at the 47.5% and 95% thresholds (2008–2011 samples), with nodes coloured by *pfk13* mutation status (A, B) and country of origin (C, D).



Supplementary Figure 15. Identity-by-descent (IBD) networks in Southeast Asia showing clusters at the 47.5% and 95% pairwise IBD thresholds (2012-2015). IBD networks were constructed at the 47.5% and 95% thresholds (2008–2011 samples), with nodes coloured by *pfk13* mutation status (A, B) and country of origin (C, D).



Supplementary Figure 16. Identity-by-descent (IBD) networks in Southeast Asia showing clusters at the 47.5% and 95% pairwise IBD thresholds (2016-2019). IBD networks were constructed at the 47.5% and 95% thresholds (2008–2011 samples), with nodes coloured by *pfk13* mutation status (A, B) and country of origin (C, D).



Supplementary Figure 17. Cross-validation results for optimal K selection for ADMIXTURE analysis. Cross-validation (CV) Errors are shown for K=1:12 and optimal K=10 is highlighted at the inflection point.