Major sub-populations of *P. falciparum* in sub-Saharan Africa

Alfred Amambua-Ngwa[1], Lucas Amenga-Etego[2], Edwin Kamau[3], Roberto Amato[4, 14], Anita Ghansah[5], Lemu Golassa[6], Milijaona Randrianarivelojosia[7], Deus Ishengoma[8], Tobias Apinjo[9], Oumou Maiga[10], Ben Andagalu[5], Yavo William[11], Marielle Boyou[12], Oyebola Kolapo[1,13], Karim Mane[1], Archibald Worwui[1], David Jeffries[1], Vikki Cornelius[4,14], Umberto D'Alessandro[1], Dominic Kwiatkowski[4,14] and Abdoulaye Djimde*[15]

1. Medical Research Council Unit The Gambia at LSHTM, Banjul, The Gambia
2. West African for Centre for Cell Biology of Infectious Pathogens (WACCBIP), University of Ghana, Accra, Ghana
3. US Army Medical Research Unit Kenya - Walter Reed Project, Kisumu, Kenya
4. Wellcome Trust Sanger Institute, Hinxton, UK
5. Noguchi Memorial Institute for Medical Research (NMIMR), Accra, Ghana
6. Addis Ababa University, Ethiopia
7. Institut Pasteur of Madagascar, Antanarivo, Madagascar
8. National Institute for Medical Research (NIMR), Tanga, Tanzania
9. University of Buea, Buea, Cameroon
10. Bernhard Nocht Institute for Topical Medicine (BNITM), Germany
11. University Félix Houphouët-Boigny, Abidjan, Cote-D'Ivoire
12. Faculty of Medicine, University of Health Sciences, Gabon
13. University of Lagos, Lagos, Nigeria
14. MRC Centre for Genomics and Global Health, University of Oxford, UK
15. Malaria Research and Training Centre, University of Science, Techniques and Technologies of Bamako

* Corresponding author: adjimde@icermali.org

Abstract

Plasmodium falciparum remains a relevant global health pathogen with high levels of genomic variation and gene flow that could undermine malaria elimination strategies, especially in the high burden regions of Africa. Infections with *P. falciparum* remain complex across most of sub-Saharan Africa. SNP variants from 2263 isolates from 24 malaria endemic settings within 15 African countries classified into western, central and eastern ancestry, plus a divergent Ethiopian population. The parasite populations are interbred and share genomic haplotypes especially across drug resistance loci. Haplotypes across drug resistance associated loci showed the strongest recent identity-by-descent between populations and endogenous haplotypes have spread to and from all populations. A recent signature of selection on chromosome 12 with candidate resistance loci against artemisinin derivatives is evident in Ghana and Malawi. Such selection and emerging sub-structure may affect intervention strategies and the efficacy of drugs and vaccines for malaria elimination.

Introduction

The decline in global malaria prevalence is now reported to be stalling and elimination programmes will need additional knowledge, new tools and intervention strategies to meet global elimination milestones (*1*). The burden from *Plasmodium falciparum* in particular remains significant in Africa where it is prevalent across vast and varied vector, parasite and human populations (*2, 3*). While interventions drive down population sizes, reduce

transmission and disconnect parasite populations, they can also drive selection, adaptation and population fragmentation. Population fragmentation and reduced diversity are therefore important parameters for consideration in refining approaches or tools for elimination(*4*). As massive interventions are implemented against falciparum malaria in Africa, it is imperative to determine if this has impacted the structure of the parasite population, which until recently has been considered as highly diverse and homogeneously interconnected(*5*). With higher density of genomic variants from next generation sequencing technologies, a better resolution of African *P. falciparum* populations is now possible. SNP markers for specific local populations or across several sites within geographic blocs have enabled the description of genomic variation and signatures of selection in Africa(*6, 7*). However, the ancestry, current structure and gene flow between *P. falciparum* across Africa remains unclear.

The *Plasmodium* Diversity Network Africa (PDNA) has conducted *P. falciparum* genomic surveillance across Africa from the west Atlantic coast regions with high rainfall and perennial transmission; the Sahel with short rainy seasons and seasonal transmission; forest covered areas of Central Africa with perennial transmission; and Eastern Africa with perennial and seasonal transmissions to Ethiopia and the Island of Madagascar with co-transmission of *P.* vivax*(8).* With high resolution sequence variants of *P. falciparum* isolates across Africa, we reveal the population structure, admixture, markers of identity-by-decent (IBD), differentiation and signatures of selection.

Results

SNP variants (29,998) were extracted from the genomes of 2263 *Plasmodium falciparum* isolates sampled from 24 sites within 15 African countries (figure 1a, and table S1 & S2). At least 55% of infections were polygenomic with up to 9 clones in some samples from Ghana, Guinea and Malawi (figure S1). The proportions of complex infections (i.e. lower mean Fws) was highest for isolates from sites in Kenya and least in Ethiopia (figure 1b). Malaria transmission intensity around the sampling site in Kenya (Kisumu, Lake Victoria) is high and will be driving the high infection complexity(*9*). In West Africa, isolates from Gambia and Senegal where the least polygenomic, confirming earlier reports of a decline in complexity with drop in prevalence, probably due to scale up of interventions(*10*).

Standard principal component analysis (PCA) of genome haplotypes resolved 3 major groups of western (west Africa and the more central states of Cameroon and Gabon), eastern (DRCongo and all other sites in East Africa) and a distinct Ethiopian population (figure S2-S4). This substructure was refined to 6 clusters from dimension reduction of ancestral membership coefficients, assuming 6 or 15 ancestral populations (figure 1c & d respectively).

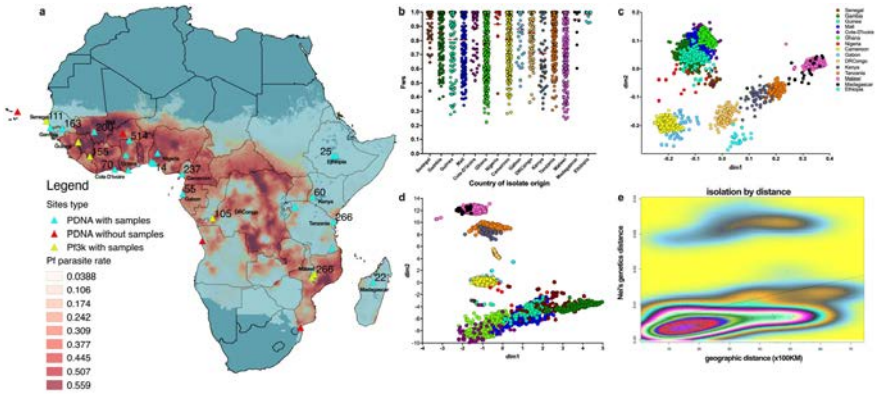

Figure 1. Sites, sample sizes and genetic groupings of *P. falciparum* isolates across PDNA and Pf3K studies in Africa. a) Sites*, P. falciparum* prevalence and studies from which SNP data of 2263 isolates were accessed. b) Complexity of infections from sampling sites (countries). c) Scatter plot of dimension 1 and 2 from multidimensional scaling of tess3r ancestry coefficients for 6 ancestral populations or d) 15 ancestral populations clustered with t-sne. e) Isolation-by-distance heatmap from distribution of genetic and geographic distances between isolates from the different sampling sites colour coded by site of sampling.

Six genetic clusters retained were west-African (WAF; Senegal, Gambia, Guinea, Mali, Cote D'Ivoire, Ghana, Nigeria), central-African (CAF; Cameroon and Gabon), south-central-African (SCAF; DRCongo), east-African (EAF; Kenya and Tanzania), south-east-African (SEAF; Malawi and Madagascar) and the horn-of-Africa (HAF; Ethiopia). Each cluster suggests an ancestral or transmission connectivity supported by geographic proximity and confirmed by significant isolation-by-distance (p=0.03), (figure 1e, figure S5). The major population continuums were within West and East with several fold difference in genetic distance (all F_{ST} values >0.1) between them and Ethiopia. This is indicative of differentiation due to factors beyond spatial separation and geographic barriers (western Cameroon forest, the equatorial forest, Congo basin rivers, and highlands of Ethiopia), (figure S6). These factors could include differences in human and vector populations as well as the history of interventions. Here, isolates from DRCongo and Ethiopia clustered away from geographically proximal sites in CAF and EAF respectively. Ethiopian human population has a unique ancestry from the rest of Africa, allowing sympatric transmission of *Plasmodium vivax(11). P. falciparum* isolates from neighbouring Djibouti also showed high genetic distances from other African populations (*12*). Like Madagascar, the human populations at the horn-of-Africa have higher frequencies of the duffy antigen and *P. vivax* is co-transmitted. However, isolates from Madagascar clustered with those from Malawi, indicating ancestry from mainland despite human populations originating from South East Asia and separation by 1,400 km of land and the Indian ocean (figure S5). Co-prevalence of *P. falciparum* and *P. vivax* thus have a minor effect on the structure observed.

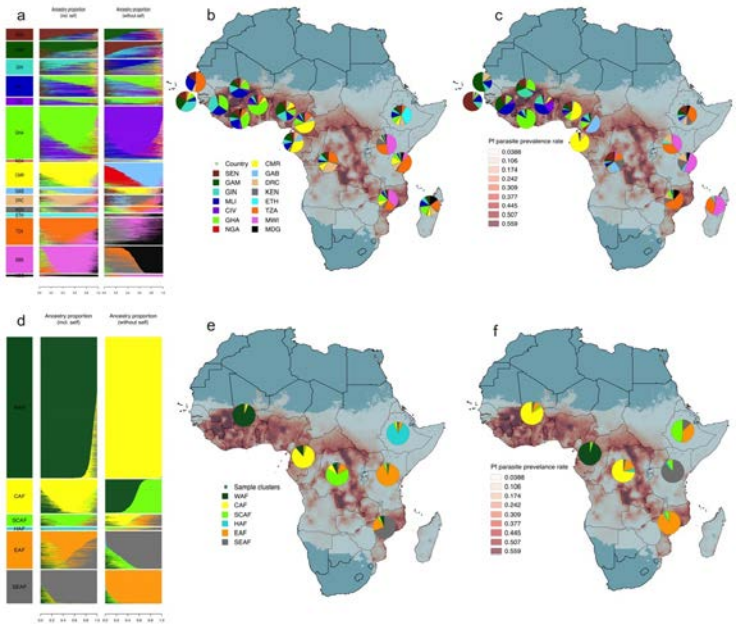

Figure 2. Genome-wide ancestry proportions for *P. falciparum* isolates (admixture-like barplots) or populations (pie-charts) modelled to include donors from all sites (inc self) or excluding isolates from recipient sampling site (without self). a) ancestry per isolate (rows) from each sampling site (column1). b) median ancestry from each sampling site located on map. c) median ancestry proportions between isolates from each sampling site, excluding donors from same site. d) ancestry proportions for genetic clusters (column1). e) ancestry proportions for each population including self-copying (a) and, (b) without self-copying.

Malaria parasites are believed to have jumped from western great apes to humans about 10,000 years ago, prior to major human migration events (*13*). The donation of ancestral genome chunks from CAF to both west and eastern populations in this study aligns with this suggested origin followed by the spread through historical and more recent human migration in Africa. Early human migration from central African was dominated by bantu populations moving westwards and south-eastwards, after the emergence of malaria in humans (*14*). These historical links prior to dispersal of humans and parasites west and east of Africa will account for the shared ancestry between all populations (figure 2). Recent human migration from colonization and slavery may also have contributed to this structure. Here, significant ancestral chunks are shared between distal French colonies like Cameroon, Mali and Senegal, while ancestry from WAF sites of Mali, Guinea and Senegal are present in DRCongo. T-SNE and FineStructure clustering of ancestral chunk matrices also maintained the major west and eastern groupings and sub-populations, as well as confirming isolates from DRCongo to be more related to eastern populations. (figure S7 & S8). Mixing of populations could facilitate gene flow, IBD signatures and spread of adaptive alleles between populations (*15*).

The proportions of isolates in IBD (<3%) was weak and uneven across the genome given intense recombination in Africa (figure 3a). However, high IBD proportions spanned twelve segments of the genome coding for candidate drug resistance loci; *Pfaat1* (PF3D7_0629500) on chromosome 6; known drug resistance genes, *Pfmdr1, Pfcrt* and *Pfdhps* and a cluster of genes on chromosome 12 (*Pfap2mu, PfATPase and Pfap2g2*) (figure 3a & 3b, figure S8). These genes are involved in drug responses, transportation and metabolism (figure 3b & 3c). These results confirm links between *Pfcrt* and *Pfaat1*, which together with *Pfap2g2* and *Pfatpase*2 have been identified as part of the malaria druggable genome (*16*). *Pfap2mu* in particular had been linked to artemisinin tolerance in Africa(*17*).

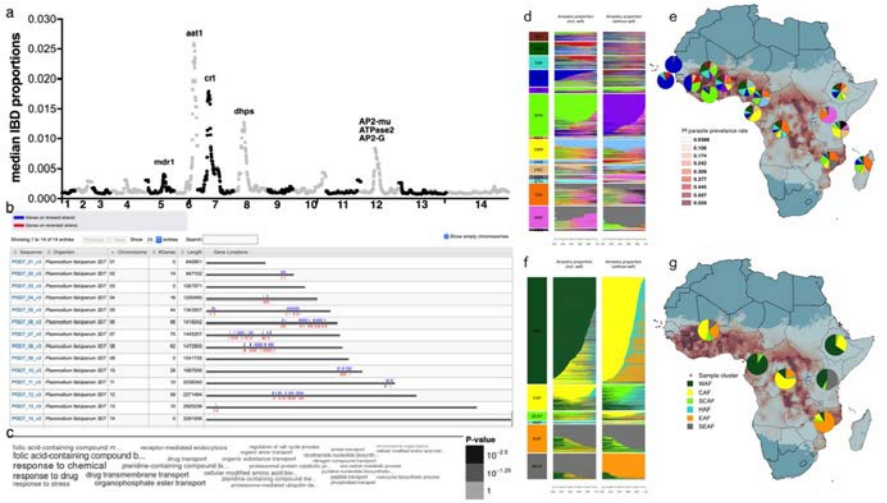

Figure 3. Haplotype sharing across drug resistance loci identified by pairwise identity-by-descent (IBD) and chromosome painting. a) IBD between pairs of *P. falciparum* isolates. IBD segment peaks highlighted at; multidrug resistance protein 1(mdr1), amino acid transporter 1(aat1), chloroquine resistance transporter (crt), dihydropteroate synthetase (dhps), AP2 domain transcription factors (AP2-G and AP2-mu) and aminophospholipid-transporting P-ATPase (ATPase2). b) Genomic segments with high IBD proportions. c) enrichment of ontology terms for genes within IBD segments. d) Admixture-like barplots of acquired ancestry proportions per isolates (rows) across drug resistance loci including donors from same site (inc self) or from different sampling sites (without self). e) median ancestry proportions (pie charts) donated to each site by isolates from other sampling sites. e) ancestry for isolates classified by genetic cluster defined in figure 1. f) Median proportions of acquired drug resistance haplotypes from different groups into each of the main population clusters.

With IBD detection limited within 25 generations, drug resistance haplotypes would have been recently and locally selected before sharing across Africa. Multiple local emergence was evident from the multiplicity of haplotype clusters for antifolate markers but also for *Pfmdr1,* which showed two ancestral lineages dominant in east and western populations respectively (figure S9). Multiple emergence of resistance against antifolates had been described, but not for a major quinoline resistance mediator such as *Pfmdr1 (18)*. Haplotype painting across drug resistance loci, confirms bidirectional sharing across drug resistance loci

(figure 3). Selection, emergence and spread of resistance to drugs is therefore feasible in all endemic sites across Africa. With treatment pressure from artemisinin derivatives and use of older drugs in preventive treatment, signs of differentiation and positive selection could inform future strategies to curb the process of drug adaptation.

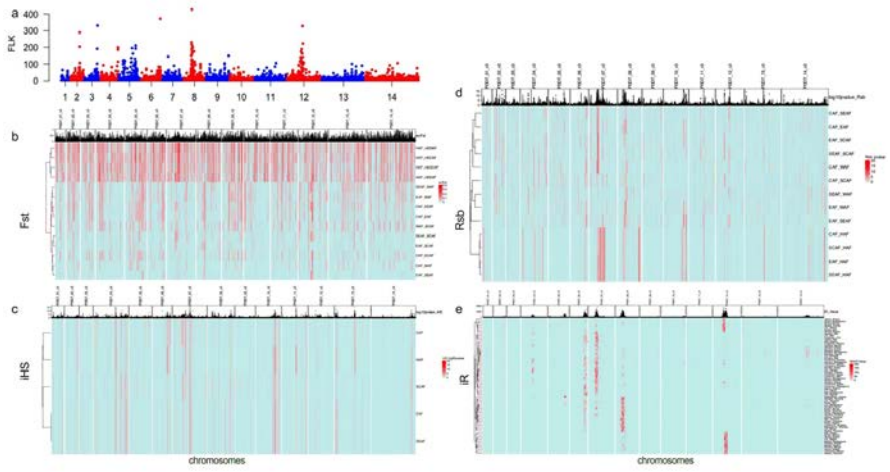

Figure 4. Signatures of differentiation and positive selection in *P. falciparum* populations from Africa. Heatmaps were clustered on rows for similar patterns within and between populations. SNP values are in columns separate by chromosomes for each population or pair of populations in rows. Low to high values are colour graded from azure to red on RGB colour wheel. a) Manhattan plot of median pairwise differentiation per SNP (FLK) between all genetic groups with isolates from Ethiopia defined as an outgroup. b) Weir and Cockerham's index of fixation (Fst). c) Integrated haplotype score (iHS) per population grouping. d) cross population extended haplotype homozygosity index (Rsb) between pairs of populations in rows. e) Significance of pairwise IBD index (iR) between isolates from different sampling sites (rows).

SNPs in drug resistance, erythrocyte invasion, gametocytogenesis, oocyst associated gene loci and antigenic loci were most differentiated between populations (figure 4a & 4b, table S5). These could be due to different histories but also selection from the different human and mosquito host across Africa. The drug loci (*Pfaat1, Pfmdr1, Pfcrt, Pfdhfr, Pfdhps*) and the IBD cluster on chromosome 12 showed signature of positive selection and haplotype differentiation across sampled populations (figure 4c-4e, table S5). Selection on the chromosome 12 cluster was strongest in Malawi and Ghana and these may have emerged independently. Further functional and clinical studies are urgent to determine if variants at these loci can compromise the efficacy of artemisinin combination treatments.

Conclusion

Plasmodium falciparum in Africa belong to major eastern, central and western subgroups, distant from a highly divergent Ethiopian population. These endogenous genomic lineages are the ancestral backbones on which adaptive loci such as drug resistance mutations may

have emerged, recombined and shared haplotypes across Africa in both eastern and western directions. This pattern may be repeated against artemisinin derivatives, which select for loci on chromosome 12 under directional selection. As sustained drug pressure are reducing prevalence, lowering recombination and increasing clonality and isolation of populations, there is need for increased molecular and phenotypic vigilance. This should be extended to large swath of endemic populations in central Africa, where civil strife and other global health pathogen epidemics could maintain malaria and threaten efforts at elimination.

References

1. W. H. O. (WHO), World Malaria Report. *WMR*, (2018).
2. C. Anopheles gambiae Genomes *et al.*, Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* **552**, 96-100 (2017).
3. D. Gurdasani *et al.*, The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327-332 (2015).
4. T. G. Anthony *et al.*, Fragmented population structure of plasmodium falciparum in a region of declining endemicity. *The Journal of infectious diseases* **191**, 1558-1564 (2005).
5. M. Manske *et al.*, Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* **487**, 375-379 (2012).
6. A. Amambua-Ngwa *et al.*, Consistent signatures of selection from genomic analysis of pairs of temporal and spatial Plasmodium falciparum populations from The Gambia. *Sci Rep* **8**, 9687 (2018).
7. C. W. Duffy *et al.*, Multi-population genomic analysis of malaria parasites indicates local selection and differentiation at the gdv1 locus regulating sexual development. *Sci Rep* **8**, 15763 (2018).
8. A. Ghansah *et al.*, Monitoring parasite diversity for malaria elimination in sub-Saharan Africa. *Science* **345**, 1297-1298 (2014).
9. A. Kapesa *et al.*, The current malaria morbidity and mortality in different transmission settings in Western Kenya. *PloS one* **13**, e0202031 (2018).
10. A. K. Bei *et al.*, Dramatic Changes in Malaria Population Genetic Complexity in Dielmo and Ndiop, Senegal, Revealed Using Genomic Surveillance. *The Journal of infectious diseases* **217**, 622-627 (2018).
11. J. A. Hodgson, C. J. Mulligan, A. Al-Meeri, R. L. Raaum, Early back-to-Africa migration into the Horn of Africa. *PLoS genetics* **10**, e1004393 (2014).
12. H. Bogreau *et al.*, Genetic diversity and structure of African Plasmodium falciparum populations in urban and rural areas. *The American journal of tropical medicine and hygiene* **74**, 953-959 (2006).
13. D. E. Loy *et al.*, Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. *Int J Parasitol* **47**, 87-97 (2017).
14. E. Patin *et al.*, Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* **356**, 543-546 (2017).
15. L. Henden, S. Lee, I. Mueller, A. Barry, M. Bahlo, Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. *PLoS genetics* **14**, e1007279 (2018).
16. A. N. Cowell *et al.*, Mapping the malaria parasite druggable genome by using in vitro evolution and chemogenomics. *Science* **359**, 191-199 (2018).
17. G. Henriques *et al.*, The Mu subunit of Plasmodium falciparum clathrin-associated adaptor protein 2 modulates in vitro parasite response to artemisinin and quinine. *Antimicrob Agents Chemother* **59**, 2540-2547 (2015).
18. T. Anderson, Mapping the spread of malaria drug resistance. *PLoS Med* **6**, e1000054 (2009).

Acknowledgments:

We thank the participants and local health workers from PDNA sites. DELGEME for supporting this analysis. Special thanks to Dr George Busby for discussion and advising on admixture analyses. Sequencing was done by the MalariaGEN program at the Wellcome Trust Sanger Institute. We thank the Pf3K for allowing access to non-PDNA data. The authors declare no competing interest. The short-read sequences used in this publication are available in the ENA and SRA databases. The views expressed are those of the authors and should not be construed to represent the positions of the U.S. Army or the Department of Defense. The investigators have adhered to the policies for protection of human subjects as prescribed in AR-70"

Supplementary Materials

Materials and Methods

Figures S1 – S8

Tables S1 – S3

Formatted: Font: +Body (Calibri)

Formatted: Line spacing: Multiple 1.15 li