MALARIA

Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa

Alfred Amambua-Ngwa[1], Lucas Amenga-Etego[2], Edwin Kamau[3,4], Roberto Amato[5,6], Anita Ghansah[7], Lemu Golassa[8], Milijaona Randrianarivelojosia[9], Deus Ishengoma[10], Tobias Apinjoh[11], Oumou Maïga-Ascofaré[12], Ben Andagalu[3], William Yavo[13], Marielle Bouyou-Akotet[14], Oyebola Kolapo[1,15], Karim Mane[1], Archibald Worwui[1], David Jeffries[1], Vikki Simpson[4,6], Umberto D'Alessandro[1], Dominic Kwiatkowski[5,6], Abdoulaye A. Djimde[5,16]*

Understanding genomic variation and population structure of *Plasmodium falciparum* across Africa is necessary to sustain progress toward malaria elimination. Genome clustering of 2263 *P. falciparum* isolates from 24 malaria-endemic settings in 15 African countries identified major western, central, and eastern ancestries, plus a highly divergent Ethiopian population. Ancestry aligned to these regional blocs, overlapping with both the parasite's origin and with historical human migration. The parasite populations are interbred and shared genomic haplotypes, especially across drug resistance loci, which showed the strongest recent identity-by-descent between populations. A recent signature of selection on chromosome 12 with candidate resistance loci against artemisinin derivatives was evident in Ghana and Malawi. Such selection and the emerging substructure may affect treatment-based intervention strategies against *P. falciparum* malaria.

The worldwide decline in malaria prevalence is now stalling and additional knowledge, new tools, and intervention strategies will be needed for global malaria elimination and eradication (*1*). The burden of *Plasmodium falciparum* malaria in particular remains substantial in sub-Saharan Africa (sSA), where it involves various vectors and human populations (*2, 3*). Although interventions have reduced and disconnected malaria parasite populations, they may be driving selection, adapta-

tion, and population fragmentation. Population fragmentation and reduced diversity can be assessed for refining approaches or tools for elimination (*4*). Therefore, it is important to determine the effect of large-scale control interventions on the structure of the parasite population, which until recently was considered to be highly diverse and homogeneously interconnected in sSA (*5*). The ancestry, current structure, and gene flow between different *P. falciparum* populations across sSA remain unclear. Previous studies

have used single-nucleotide polymorphism (SNP) markers to characterize specific geographic populations and describe genomic variation and signatures of selection in sSA (*6, 7*). Recent higher-density genomic polymorphisms from next-generation sequencing technologies can further resolve African *P. falciparum* subpopulations and population-specific genomic signatures.

The Plasmodium Diversity Network Africa (PDNA) conducts *P. falciparum* genomic surveillance across sSA, from the West Atlantic coastal regions with their high rainfall and perennial transmission; the Sahel with its short rainy seasons and seasonal transmission; Central Africa with its forest-covered areas and perennial transmission; Eastern Africa with its perennial and seasonal transmission; to Ethiopia and the

[1]Medical Research Council Unit The Gambia at LSHTM, Banjul, The Gambia. [2]West African Centre for Cell Biology of Infectious Pathogens (WACCBIP), University of Ghana, Accra, Ghana. [3]United States Army Medical Research Directorate-Africa, Kenya Medical Research Institute/Walter Reed Project, Kisumu, Kenya. [4]Walter Reed Army Institute of Research, U.S. Military HIV Research Program, Silver Spring, MD, USA. [5]Wellcome Sanger Institute, Hinxton, UK. [6]MRC Centre for Genomics and Global Health, Big Data Institute, University of Oxford, Oxford, UK. [7]Noguchi Memorial Institute for Medical Research (NMIMR), Accra, Ghana. [8]Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa, Ethiopia. [9]Institut Pasteur of Madagascar, Antanarivo, Madagascar. [10]National Institute for Medical Research (NIMR), Tanga, Tanzania. [11]Department of Biochemistry and Molecular Biology, University of Buea, Buea, Cameroon. [12]Bernhard Nocht Institute for Tropical Medicine (BNITM), Hamburg, Germany. [13]Unite des Sciences Pharmaceutiques et Biologiques, University Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire. [14]Faculty of Medicine, University of Health Sciences, Libreville, Gabon. [15]Department of Zoology, University of Lagos, Lagos, Nigeria. [16]Malaria Research and Training Centre, University of Science, Techniques and Technologies of Bamako, Bamako, Mali.
*Corresponding author. Email: adjimde@icermali.org

Fig. 1. Sites, sample sizes, and genetic groupings of *P. falciparum* isolates across PDNA and Pf3K studies in Africa.
(**A**) Sites, *P. falciparum* (Pf) prevalence rate, and studies from which SNP data of 2263 isolates were accessed. Map was extracted from a malaria atlas showing *P. falciparum* prevalence as brown density within the ranges of the key (https://map.ox.ac.uk/explorer/#/). (**B**) Complexity of infections by inbreeding coefficient (Fws). (**C**) Scatter plot from multidimensional scaling of tess3r ancestry coefficients for six predicted ancestral populations.
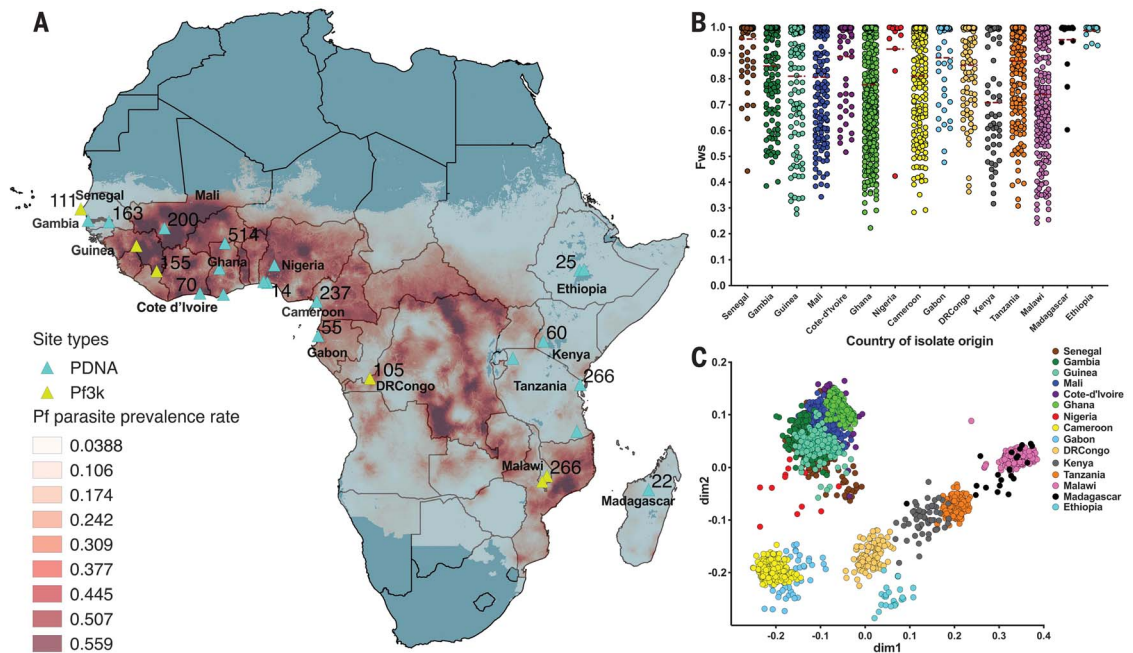
island of Madagascar with their cotransmission of *P. vivax* (*8*). Using high-resolution genome-wide SNP variants of *P. falciparum* isolates across sSA, we reveal the population structure, admixture, markers of identity-by-descent (IBD), differentiation, and signatures of selection.

SNP variants (29,998) were extracted from whole-genome sequences of 2263 *P. falciparum* isolates sampled from across 15 African countries (Fig. 1A and tables S1 and S2). At least 55%

of infections were polygenomic, with up to nine clones in some infections from Ghana, Guinea, and Malawi (fig. S1). The proportion of complex infections [i.e., lower mean inbreeding coefficient (Fws)] was highest in Kenya and lowest in Ethiopia (Fig. 1B). Malaria transmission around the sampling site in Kenya (Kisumu, Western Kenya) was stable and high (*9*), probably driving the high infection complexity. In West Africa, isolates from The Gambia and Senegal were the

least complex, confirming earlier reports of a decline in complexity with decreasing prevalence, probably due to the scale-up of interventions (*10*).

Standard principal components analysis, using imputed genome haplotypes (fig. S2), resolved three major groups: western (West Africa and the more-central countries of Cameroon and Gabon), eastern [Democratic Republic of the Congo (DR Congo) and all other sites in East Africa], and a
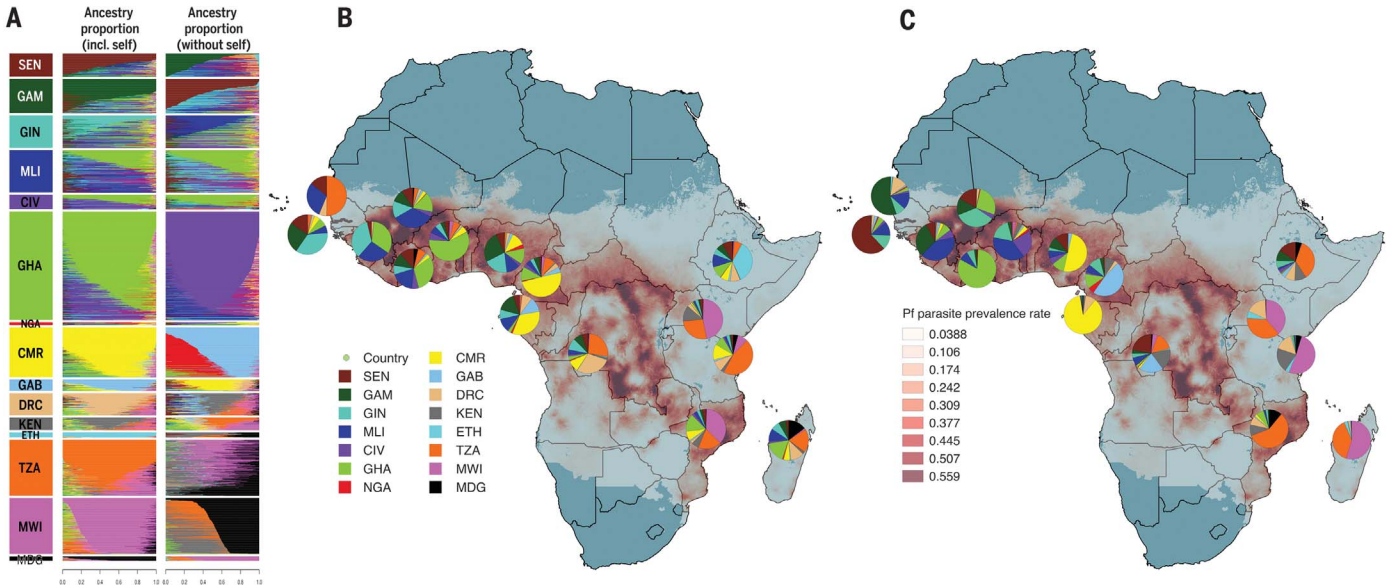

Fig. 2. Genome-wide ancestry proportions. Ancestry proportions for *P. falciparum* isolates (admixture-like bar plots) or populations (pie charts) modeled to include donors from all sites (incl. self) or excluding isolates from recipient sampling site (without self). (**A**) Ancestry per isolate (rows) from each sampling site (left column). (**B**) Median ancestry from each sampling site. (**C**) Median ancestry proportions between isolates from each sampling site, excluding donors from same site. Country colors are the same as in Fig. 1.
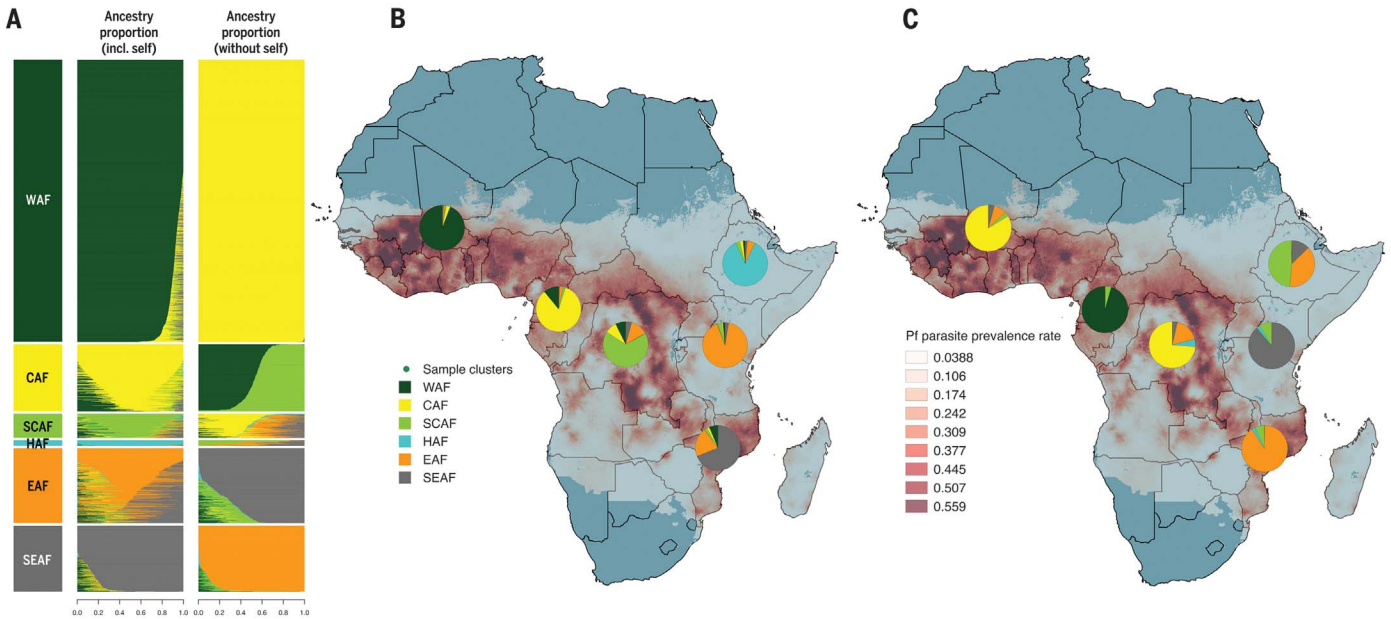

Fig. 3. Genome-wide ancestry proportions for *P. falciparum* populations in sSA. (**A**) Ancestry proportions for regional genetic blocs (left column). Ancestry proportions for each genetic cluster (**B**) including self-copying and (**C**) without self-copying.

distinct Ethiopian population (fig. S3). This sub-structure was refined to six distinct clusters from multidimensional scaling of ancestral membership coefficients, splitting DR Congo from East African populations (Fig. 1C and fig. S4). The six retained genetic clusters were West African (WAF; Senegal, Gambia, Guinea, Mali, Côte d'Ivoire, Ghana, and Nigeria), Central African (CAF; Cameroon and Gabon), South Central African (SCAF; DR Congo), East African (EAF; Kenya and Tanzania), Southeast 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 ($P = 0.03$, Mantel test) (fig. S5). The major population continuums were within West Africa and East Africa, with several-fold difference in genetic distance [all fixation index (F_{ST}) values > 0.1] between them and Ethiopia. Differentiation might also result from differences in human and vector populations, the history of interventions on spatial separation, and geographic barriers (e.g., western Cameroon forest, the equatorial forest, Congo Basin rivers, and highlands of Ethiopia). Isolates from DR Congo and Ethiopia clustered away from geographically proximal sites in CAF and EAF, respectively. Human populations from Ethiopia and other HAF sites, such as Djibouti, have a distinct ancestry from the rest of Africa, allowing sympatric transmission of *P. vivax,* with earlier reports of divergent *P. falciparum* populations (11, 12). As in Madagascar, HAF human populations have higher frequencies of the Duffy antigen, allowing *P. vivax* cotransmission. However, isolates from Madagascar clustered with those from Malawi, indicating mainland ancestry despite a high proportion of human populations originating from Southeast Asia and being separated by 1400 km of land and the Indian Ocean. Therefore, it is not likely that the divergence of HAF isolates is due to co-prevalence with *P. vivax* but might be driven by other factors such as differences in vector populations. This could also explain the differentiation between Congolese and other CAF isolates where vector populations differ, with *Anopheles funestus* being relatively dominant in DR Congo (13).

Recent studies have shown that *P. falciparum* from western great apes jumped into humans about 10,000 years ago, prior to major human migrations (14, 15). The donation of ancestral genome chunks from CAF to both western and eastern *P. falciparum* populations aligns with such an origin and the spread of malaria through historical and more recent human migration in Africa. Recent human migration brought on by colonization and slavery may have resulted in *P. falciparum* ancestral chunks shared between distal French colonies like Cameroon, Mali, and Senegal, whereas ancestry from WAF sites of Mali, Guinea, and Senegal are present in DR Congo (Fig. 2 and fig. S6). However, historical links prior to dispersal of humans and parasites to West and East Africa may also account for the shared ancestry between all major population blocs (Fig. 3). The early human migration from Central Africa, after the emergence of malaria in humans, was dominated by Bantu populations moving westward and southeastward (16). T-SNE and fineSTRUCTURE clustering of ancestral chunk matrices also maintained the major West and East African subpopulations, further indicating that isolates from DR Congo share more eastern ancestry (figs. S7 and S8). Human population mixing could have facilitated *P. falciparum* gene flow, IBD signatures, and spread of adaptive alleles across Africa (17).

The proportions of isolates sharing IBD (<3%) was weak and uneven across the genome, as expected for intensely recombining parasite populations (Fig. 4A and fig. S9). However, relatively high IBD proportions spanned 12 segments of the genome, including regions 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). These genes are involved in drug responses, transportation, and metabolism (fig. S10). These results confirm links between Pfcrt and Pfaat1, which together with Pfap2g2 and PfATPase have been identified as part of the malaria druggable genome (18). Pfap2mu in particular has been linked to artemisinin tolerance in Africa (19). Strong IBD around Pfap2mu in Ghana and Malawi (Fig. 4B) may have emerged independently and calls for increased vigilance against artemisinin-based combination therapy (ACT) efficacy. The introduction or local emergence and sharing of candidate drug resistance haplotypes would be recent, as IBD detection was limited to 25 generations. Haplotype painting across drug resistance loci (table S6) emphasized bidirectional gene flow across these loci (fig. S11). Multiple origins of antifolate markers were confirmed (20) but also seen for Pfmdr1, which showed two ancestral lineages dominant in West and East African populations, respectively (fig. S12). Multiple emergence for a major quinolone resistance mediator such as Pfmdr1
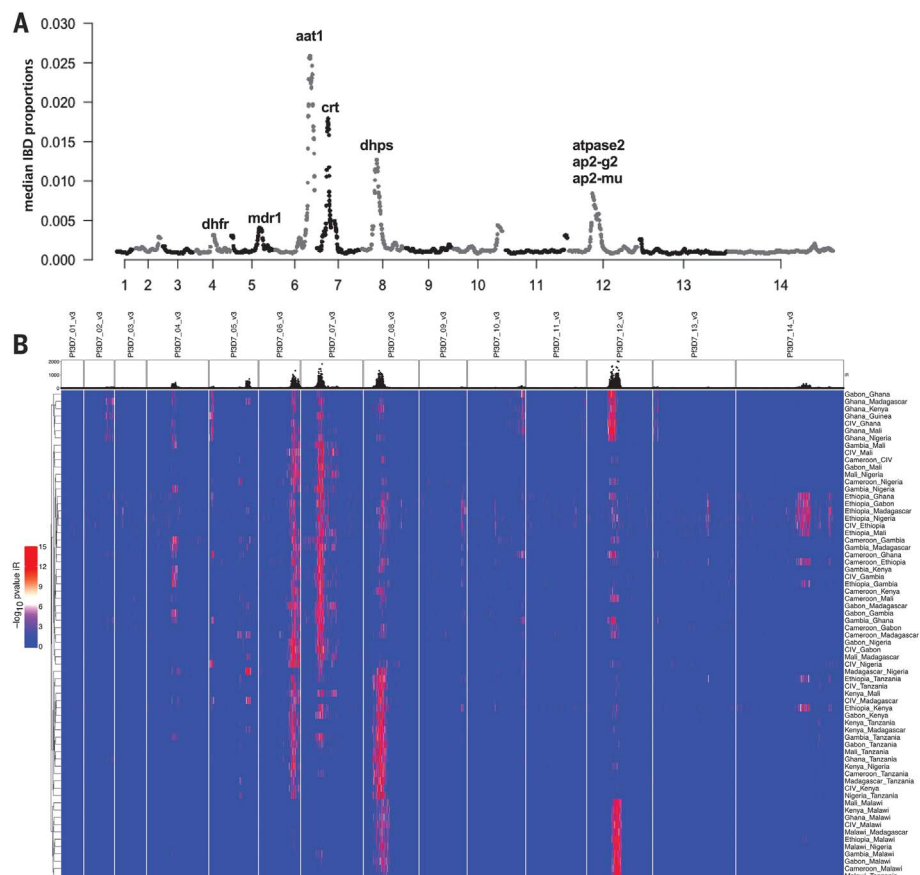

Fig. 4. Pairwise IBD between isolates across sites. (**A**) Manhattan plot of median IBD between pairs of *P. falciparum* isolates, showing each chromosome as numbered on the *x* axis. IBD segment peaks labeled for dihydrofolate reductase (dhfr), multidrug resistance protein 1 (mdr1), amino acid transporter 1 (aat1), chloroquine resistance transporter (crt), dihydropteroate synthetase (dhps), AP2 domain transcription factors (ap2-g2 and ap2-mu), and aminophospholipid-transporting P-ATPase (atpase2). (**B**) Heatmap of pairwise IBD between sampled populations clustered on rows for similar patterns between populations. SNP values are in columns separated by chromosomes for each pair of populations in rows. Low to high values are color graded from blue to red on RGB color wheel.

has not been previously reported. Selection, emergence, and spread of resistance to drugs is therefore possible in all malaria endemic sites across sSA. These findings are important because artemisinin resistance may emerge independently in sSA and not necessarily spread from Southeast Asia. This calls for careful surveillance of artemisinin resistance in sSA, where drug pressure from ACT and seasonal malaria chemoprevention with sulfadoxine-pyrimethamine and amodiaquine are being scaled up for elimination. These would also lead to population differentiation (fig. S13) and positive selection that could facilitate the development of clinical drug resistance.

SNPs related to drug resistance, erythrocyte invasion, gametocytogenesis, oocyst development, and antigenic loci were the most differentiated between populations (fig. S14, A and B, and tables S7 and S8). These could be due to different environmental conditions and varying human and mosquito populations. Known drug loci (Pfaat1, Pfmdr1, Pfcrt, Pfdhfr, and Pfdhps) and the IBD cluster on chromosome 12 showed signatures of positive selection and haplotype differentiation across sampled populations (figs. S14, C and D, S15, and S16, and tables S9 and S10). It would be important to determine whether variants at these loci can compromise the efficacy of artemisinins and/or ACTs.

P. falciparum in sSA is clustered into major western, central, and eastern subgroups and a highly divergent Ethiopian subpopulation. These endogenous genomic lineages are the ancestral backbone on which adaptive loci such as drug resistance mutations may have emerged, recombined, and been shared both westerly and easterly across sSA. This may occur again against current artemisinin-based treatments, which are already directionally selecting loci on chromosome 12. These signal the need for broader molecular

and phenotypic surveillance of *P. falciparum* in sSA, including the large swathes of endemic populations in Central Africa, where civil strife and other global health pathogen epidemics could maintain malaria and threaten elimination efforts.

REFERENCES AND NOTES

1. *World Malaria Report 2018* (World Health Organization, 2018).
2. Anopheles gambiae 1000 Genomes Consortium, *Nature* **552**, 96–100 (2017).
3. D. Gurdasani *et al.*, *Nature* **517**, 327–332 (2015).
4. T. G. Anthony *et al.*, *J. Infect. Dis.* **191**, 1558–1564 (2005).
5. M. Manske *et al.*, *Nature* **487**, 375–379 (2012).
6. A. Amambua-Ngwa *et al.*, *Sci. Rep.* **8**, 9687 (2018).
7. C. W. Duffy *et al.*, *Sci. Rep.* **8**, 15763 (2018).
8. A. Ghansah *et al.*, *Science* **345**, 1297–1298 (2014).
9. A. Kapesa *et al.*, *PLOS ONE* **13**, e0202031 (2018).
10. A. K. Bei *et al.*, *J. Infect. Dis.* **217**, 622–627 (2018).
11. J. A. Hodgson, C. J. Mulligan, A. Al-Meeri, R. L. Raaum, *PLOS Genet.* **10**, e1004393 (2014).
12. H. Bogreau *et al.*, *Am. J. Trop. Med. Hyg.* **74**, 953–959 (2006).
13. M. E. Sinka *et al.*, *Parasit. Vectors* **5**, 69 (2012).
14. D. E. Loy *et al.*, *Int. J. Parasitol.* **47**, 87–97 (2017).
15. D. A. Joy *et al.*, *Science* **300**, 318–321 (2003).
16. E. Patin *et al.*, *Science* **356**, 543–546 (2017).
17. L. Henden, S. Lee, I. Mueller, A. Barry, M. Bahlo, *PLOS Genet.* **14**, e1007279 (2018).
18. A. N. Cowell *et al.*, *Science* **359**, 191–199 (2018).
19. G. Henriques *et al.*, *Antimicrob. Agents Chemother.* **59**, 2540–2547 (2015).
20. T. Anderson, *PLOS Med.* **6**, e1000054 (2009).

Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa

Alfred Amambua-Ngwa, Lucas Amenga-Etego, Edwin Kamau, Roberto Amato, Anita Ghansah, Lemu Golassa, Milijaona Randrianarivelojosia, Deus Ishengoma, Tobias Apinjoh, Oumou Maga-Ascofar, Ben Andagalu, William Yavo, Marielle Bouyou-Akotet, Oyebola Kolapo, Karim Mane, Archibald Worwui, David Jeffries, Vikki Simpson, Umberto DAlessandro, Dominic Kwiatkowski, and Abdoulaye A. Djimde

Ebb and flow of parasite populations

The population genetics of the malaria parasite *Plasmodium falciparum* across Africa is poorly understood but important to know for grasping the risks and dynamics of the spread of drug resistance. Harnessing the power of genomics, Amambua-Ngwa *et al.* of the Plasmodium Diversity Network Africa found substantial population structure within Africa that is consistent with human and vector population divergence (see the Perspective by Sibley). Specific signatures of selection by antimalarial drugs were detected, along with indications of the effect of colonization and slavery. Furthermore, whole-genome sequencing showed that there is extensive gene flow among the different regions and that Ethiopia has a distinctive population of *P. falciparum*, which may be indicative of coexistence with another malaria parasite, *P. vivax.*

View the article online
https://www.science.org/doi/10.1126/science.aav5427
Permissions
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service