Pf7: an open dataset of *Plasmodium falciparum* genome variation in 20,000 worldwide samples

Supplementary display items

Supplementary Table 1. Breakdown of analysis set samples by geography. Sites are divided into ten major sub-populations as described in the main text. Note that a) samples from Kisumu in western Kenya have been assigned to the Africa - Northeast (AF-NE) sub-population, whereas samples from Kilifi in coastal Kenya have been assigned to the Africa - East (AF-E) sub-population, b) samples from Odisha and West Bengal in India to the west of Bangladesh have been assigned to the Asia - South - East (AS-S-E) sub-population, whereas samples from Tripura in India to the east of Bangladesh have been assigned to the Asia - South - Far East (AS-S-FE) sub-population and c) samples from Ranong and Tak in western Thailand have been assigned to the Asia - Southeast - West (AS-SE-W) sub-population, whereas samples from Sisakhet in eastern Thailand have been assigned to the Asia - Southeast - East (AS-SE-E) sub-population

Major sub-population	Country	Admin level 1	Sequenced samples	Analysis set samples
South America (SA)	**Peru**	**Loreto**	21	21
	Colombia	**Nariño**	7	6
		Choco	3	3
		Cauca	146	123
		Valle del Cauca	3	3
	Venezuela	**Bolivar**	2	2
Africa - West (AF-W)	**Gambia**	**Western**	235	225
		North Bank	252	186
		Upper River	760	452
	Senegal	**Dakar**	93	91

		Sedhiou	62	59
	Guinea	**Faranah**	60	37
		Nzerekore	139	114
	Mauritania	**Guidimaka**	23	21
		Hodh el Gharbi	41	39
		Hodh ech Chargui	40	32
	Côte d'Ivoire	**Abidjan**	71	71
	Mali	**Kayes**	379	250
		Bamako	215	209
		Koulikoro	991	614
		Sikasso	161	57
		Segou	49	29
		Mopti	9	8
	Burkina Faso	**Haut-Bassins**	58	57
	Ghana	**Brong Ahafo**	69	50
		Ashanti	286	278
		Central	175	104
		Upper East	3300	2454
		Eastern	21	20
		Greater Accra	198	184
		Volta	41	41

	Benin	**Atlantique**	57	45
		Littoral	277	105
	Nigeria	**Lagos**	132	105
		Kwara	8	5
	Gabon	**Wouleu-Ntem**	59	55
	Cameroon	**Sud-Ouest**	294	264
Africa - Central (AF-C)	**Democratic Republic of the Congo**	**Kinshasa**	573	520
Africa - Northeast (AF-NE)	**Sudan**	**Khartoum**	124	67
		Blue Nile	66	0
		Kassala	13	9
	Uganda	**Apac**	15	12
	Kenya	**Kisumu**	64	63
	Ethiopia	**Amhara**	15	10
		Oromia	19	11
Africa - East (AF-E)	**Malawi**	**Chikwawa**	319	231
		Zomba	52	34
	Tanzania	**Kigoma**	199	143
		Kagera	61	52
		Morogoro	34	32
		Tanga	324	297
		Lindi	79	65

	Mozambique	**Gaza**	91	34
	Kenya	**Kilifi**	662	627
	Madagascar	**Mahajanga**	24	23
		Fianarantsoa	1	1
Asia - South - East (AS-S-E)	**India**	**Odisha**	122	114
		West Bengal	122	119
Asia - South - Far East (AS-S-FE)	**India**	**Tripura**	72	67
	Bangladesh	**Chittagong**	1658	1310
Asia - Southeast - West (AS-SE-W)	**Myanmar**	**Rakhine**	19	7
		Sagaing	93	38
		Mandalay	120	114
		Bago	124	89
		Kachin	28	26
		Kayin	760	631
		Shan	65	30
		Tanintharyi	51	50
	Thailand	**Ranong**	27	20
		Tak	967	875
Asia - Southeast - East (AS-SE-E)	**Thailand**	**Sisakhet**	112	59
	Laos	**Savannakhet**	452	411
		Champasak	218	208
		Salavan	147	144

		Attapeu	210	204
		Sekong	25	24
	Cambodia	**Pailin**	286	191
		Battambang	65	51
		Koh Kong	5	5
		Pursat	671	460
		Preah Vihear	216	150
		Stueng Traeng	60	52
		Ratanakiri	420	358
	Vietnam	**Bac Lieu**	4	1
		Binh Phuoc	751	657
		Quang Tri	40	35
		Dak Nong	73	70
		Quang Nam	95	75
		Binh Thuan	11	0
		Dak Lak	112	106
		Gia Lai	376	337
		Ninh Thuan	205	73
		Khanh Hoa	66	50
Oceania - New Guinea (OC-NG)	**Indonesia**	**Papua**	133	121
		East Sepik	166	149

	Papua New Guinea	**Madang**	55	43
		Milne Bay	30	29
Unverified identity	**-**	**-**	160	0
Total			***20864***	***16203***

Supplementary Table 2. Studies contributing samples. Information provided here is correct at the time of publication and to the best of our knowledge. For the most up to date partner study and contact information, please refer to the *Plasmodium falciparum* Community Project page on the MalariaGEN website: https://www.malariagen.net/projects/p-falciparum-community-project

Study ID	Study title	Contact	Samples	Sampling locations (first-level admin)
1001-PF-ML-DJIMDE	Developing the Community Project with partners in Mali	Abdoulaye Djimdé adjimde@icermali.org	96	Mali/Bamako, Mali/Koulikoro, Mali/Mopti
1004-PF-BF-OUEDRAOGO	Developing the Community Project with partners in Burkina Faso	Jean-Bosco Ouedraogo jbouedraogo.irssbobo@fasonet.bf	58	Burkina Faso/Haut-Bassins
1006-PF-GM-CONWAY	Genome-wide analysis of genetic variation in The Gambia	Alfred Amambua-Ngwa angwa@mrc.gm	79	Gambia/Western
1007-PF-TZ-DUFFY	Mother Offspring Malaria Study (MOMS) in Tanzania	Patrick Duffy duffype@niaid.nih.gov	50	Tanzania/Tanga, Tanzania/Morogoro
1008-PF-SEA-RINGWALD	Containment of artemisinin tolerant malaria parasites in South-East Asia (ARCE)	Pascal Ringwald ringwaldp@who.int	234	Vietnam/Binh Phuoc, Myanmar/Tanintharyi, Laos/Savannakhet
1010-PF-TH-ANDERSON	Genetic variation underlying drug resistance at the Thai-Burmese border	Tim Anderson tanderso@txbiomed.org	112	Thailand/Tak, Lab
1011-PF-KH-SU	Genome-wide scans of cultured adapted parasites in Cambodia	Thomas E Wellems	41	Cambodia/Pursat, Lab

		twellems@niaid.nih.gov		
1012-PF-KH-WHITE	Developing the Community Project with partners in Cambodia	White Nicholas nickw@tropmedres.ac	3	Cambodia/Pailin
1013-PF-PEGB-BRANCH	Developing the Community Project with partners in Peru	Julian C Rayner jcr1003@cam.ac.uk	16	Lab, Peru/Loreto
1014-PF-SSA-SUTHERLAND	Analysis of *Plasmodium falciparum* samples from UK travellers returning from malaria endemic countries	Colin Sutherland colin.sutherland@lshtm.ac.uk	8	Ghana/<unknown>, Mozambique/<unknown>, Uganda/<unknown>, Kenya/<unknown>
1015-PF-KE-NZILA	Genome-wide association study of in vitro drug resistance in Kenya	Irene Omedo io7@sanger.ac.uk	60	Kenya/Kilifi, Lab
1016-PF-TH-NOSTEN	Developing the Community Project with partners in Thailand	Francois Nosten francois@tropmedres.ac	21	Thailand/Tak
1017-PF-GH-AMENGA-ETEGO	Population genetics of natural populations in Northern Ghana	Lucas Amenga-Etego lucasmenga@gmail.com	409	Ghana/Upper East
1020-PF-VN-BONI	Measuring in vitro drug sensitivity in Vietnam	Thuy-Nhien Nguyen nhientt@oucru.org	24	Vietnam/Binh Phuoc
1021-PF-PG-MUELLER	Building a national repository of malaria isolates in Papua New Guinea	Ivo Mueller mueller@wehi.edu.au	56	Papua New Guinea/Madang, Papua New Guinea/East Sepik

1022-PF-MW-OCHOLLA	Genome variation and selection in clinical isolates from rural Malawi	Brigitte Denis bdenis@mlw.mw	371	Malawi/Chikwawa, Malawi/Zomba
1023-PF-CO-ECHEVERRI-GARCIA	Comparative analysis of permeome genes and drug resistance in Colombia	Diego F Echeverry difereg77@gmail.com	17	Colombia/Cauca, Colombia/Narino, Colombia/Valle del Cauca, Colombia/Choco, Lab
1024-PF-UG-BOUSEMA	FightMal - Correlating protection from malaria with immune profile of infected individuals in Uganda	Teun Bousema teun.bousema@radboudumc.nl	15	Uganda/Apac
1026-PF-GN-CONWAY	Effects of transmission intensity on population structure and signatures of selection in Guinea	David Conway david.conway@lshtm.ac.uk	199	Guinea/Nzerekore, Guinea/Faranah
1027-PF-KE-BULL	Genomics of severe malaria and low host immunity in Kenya	Irene Omedo io7@sanger.ac.uk	11	Kenya/Kilifi
1031-PF-SEA-PLOWE	Artemisinin Resistance Confirmation, Characterization and Containment (ARC3)	Pascal Ringwald ringwaldp@who.int	194	Thailand/Tak, Cambodia/Pailin, Cambodia/Battambang, Lab, Bangladesh/Chittagong
1044-PF-KH-FAIRHURST	Genomics of parasite clearance and recrudescence rates in Cambodia	Thomas E Wellems twellems@niaid.nih.gov	603	Lab, Cambodia/Pursat, Cambodia/Ratanakiri, Cambodia/Preah Vihear

1052-PF-TRAC-WHITE	Tracking Resistance to Artemisinin Collaboration (TRAC)	Elizabeth Ashley liz@tropmedres.ac	1174	Thailand/Tak, Thailand/Sisakhet, Thailand/Ranong, Cambodia/Ratanakiri, Cambodia/Preah Vihear, Cambodia/Pursat, Cambodia/Pailin, Bangladesh/Chittagong, Vietnam/Binh Phuoc, Myanmar/Bago, Myanmar/Mandalay, Myanmar/Kachin, Laos/Attapeu, Democratic Republic of the Congo/Kinshasa, Nigeria/Kwara
1062-PF-PG-BARRY	Understanding malaria parasite populations and outbreaks in Papua New Guinea	Alyssa Barry a.barry@deakin.edu.au	82	Papua New Guinea/Milne Bay, Papua New Guinea/East Sepik
1083-PF-GH-CONWAY	Alternative molecular mechanisms for erythrocyte invasion by *P. falciparum* in Ghana	Gordon Awandare gawandare@ug.edu.gh	117	Ghana/Brong Ahafo, Ghana/Upper East
1093-PF-CM-APINJOH	Population genetics of *P. falciparum* parasites in South-Western Cameroon	Tobias Apinjoh apinjohtoby@yahoo.co.uk	239	Cameroon/Sud-Ouest
1094-PF-GH-AMENGA-ETEGO	Population genetics of *P. falciparum* parasites in Northern Ghana	Lucas Amenga-Etego lucasmenga@gmail.com	256	Ghana/Upper East
1095-PF-TZ-ISHENGOMA	Genome variation and its effect on ACT treatment outcome in Tanzania	Deus Ishengoma deusishe@yahoo.com	300	Tanzania/Tanga, Tanzania/Lindi, Tanzania/Kagera

1096-PF-GH-GHANSAH	Population genetics of *P. falciparum* parasites in Southern Ghana	Anita Ghansah aghansah2013@gmail.com	101	Ghana/Central
1097-PF-ML-MAIGA	Detection of artemisinin-resistant *Plasmodium falciparum* parasites in Southern Mali	Abdoulaye Djimdé adjimde@icermali.org	138	Mali/Koulikoro
1098-PF-ET-GOLASSA	The prevalence of asymptomatic carriage; emergence of parasite mutations conferring anti-malaria drug resistance; and G6PD deficiency in the human population, as possible impediments to malaria elimination in Ethiopia	Lemu Golassa lgolassa@gmail.com	34	Ethiopia/Oromia, Ethiopia/Amhara
1100-PF-CI-YAVO	Drug resistance and *Plasmodium falciparum* diversity in forest zone of Côte d'Ivoire	William Yavo yavowilliam@yahoo.fr	71	Côte d'Ivoire/Abidjan
1101-PF-CD-ONYAMBOKO	Efficacy of three ACTs in treating *falciparum* malaria in the Democratic Republic of Congo	Caterina A Fanello caterina@tropmedres.ac	175	Democratic Republic of the Congo/Kinshasa
1102-PF-MG-RANDRIANARIVELOJOSIA	Genotyping *P. falciparum* and *P. vivax* in Madagascar	Milijaona Randrianarivelojosia milijaon@pasteur.mg	25	Madagascar/Mahajanga, Madagascar/Fianarantsoa
1103-PF-PDN-GMSN-NGWA	Population genetics of cross-border *P. falciparum* parasites in West Africa	Alfred Amambua-Ngwa angwa@mrc.gm	34	Nigeria/Lagos

1107-PF-KEN-KAMAU	Population genetics of *P. falciparum* parasites in Kenya	Ben Andagalu bandagalu@yahoo.com	64	Kenya/Kisumu
1108-PF-GAB-BOUYOU-AKOTET	Determining parasite genetic diversity in Gabon	Marielle Bouyou-Akotet mariellebouyou@yahoo.fr	59	Gabon/Wouleu-Ntem
1114-PF-PDN-DBS-GH-GHANSAH	Surveillance of kelch-13 mutation in Ghana	Anita Ghansah aghansah2013@gmail.com	82	Ghana/Central
1125-PF-TH-NOSTEN	Investigating artemisinin resistance emergence on Thai-Burmese border	Francois Nosten francois@tropmedres.ac	702	Thailand/Tak
1127-PF-ML-SOULEYMANE	Genetic analysis of *P. falciparum* before and after artemether-lumefantrine treatment in Mali	Abdoulaye Djimdé adjimde@icermali.org	164	Mali/Bamako
1131-PF-BJ-BERTIN	Identification of virulence factors in cerebral malaria in Benin	Gwladys Bertin gwladys.bertin@ird.fr	334	Benin/Littoral, Benin/Atlantique
1132-PF-K1000G-DBS-KE-BEJON	Using whole genome sequence data to analyse the spatio-temporal genetic diversity of malaria parasites in Kilifi, Kenya	Irene Omedo io7@sanger.ac.uk	620	Kenya/Kilifi, Lab
1134-PF-ML-CONWAY	Population Genetics of *P. falciparum* in West Africa	David Conway david.conway@lshtm.ac.uk	372	Mali/Kayes, Mali/Koulikoro
1135-PF-SN-CONWAY	Parasite adaption in Senegal at molecular, functional and population level	David Conway david.conway@lshtm.ac.uk	93	Senegal/Dakar

1136-PF-GM-NGWA	*Plasmodium falciparum* anti-malarial drug resistance in the Gambia: Identification of potential genetic markers by retrospective whole genome approaches	Alfred Amambua-Ngwa angwa@mrc.gm	123	Gambia/Upper River, Gambia/Western
1137-PF-GM-DALESSANDRO	Malaria transmission dynamics in The Gambia: Defining the spatial and temporal spread of malaria at micro-level (village)	Alfred Amambua-Ngwa angwa@mrc.gm	68	Gambia/Upper River
1138-PF-CD-FANELLO	Parenteral artesunate compared to quinine as a cause of late post-treatment anaemia in African children with hyperparasitaemic *P. falciparum* malaria (DHART)	Caterina A Fanello caterina@tropmedres.ac	160	Democratic Republic of the Congo/Kinshasa
1140-PF-ML-DUFFY	PfSPZ phase I trial in Doneguebougou, Mali	Jason Wendler jason.wendler@seattlechildrens.org	68	Lab, Mali/Koulikoro
1141-PF-GM-CLAESSENS	Genomic characterization of *P. falciparum* from asymptomatic infections in The Gambia	Antoine Claessens antoineclaessens@gmail.com	391	Gambia/Upper River, Lab
1145-PF-PE-GAMBOA	Genotype-phenotype study of erythrocyte invasion in Peruvian *P. falciparum* isolates	Dionicia Gamboa dionicia.gamboa@upch.pe	13	Lab, Peru/Loreto

1146-PF-MULTI-PRICE	Characterisation of drug resistance in Indonesian *P. falciparum* populations	Sarah Auburn sarah.auburn@menzies.edu.au	148	Indonesia/Papua, Sudan/Kassala
1147-PF-MR-CONWAY	Population genetics of *P. falciparum* parasites in Mauritania	David Conway david.conway@lshtm.ac.uk	104	Mauritania/Hodh el Gharbi, Mauritania/Guidimaka, Mauritania/Hodh ech Chargui
1148-PF-BD-MAUDE	Assessing the contribution of migration to the emergence and spread of antimalarial drug resistance in Southeast Bangladesh	Richard Maude richardmaude@gmail.com	1465	Bangladesh/Chittagong
1149-PF-MM-RINGWALD	Treatment Efficacy Studies in Myanmar	Pascal Ringwald ringwaldp@who.int	158	Myanmar/Sagaing, Myanmar/Shan
1151-PF-GH-AMENGA-ETEGO	Testing the effectiveness of selective whole genome amplification on samples collected in Northern Ghana	Lucas Amenga-Etego lucasmenga@gmail.com	196	Ghana/Upper East
1153-PF-Pf3KLAB-KWIATKOWSKI	Sequencing laboratory reference samples	Richard Pearson rp7@sanger.ac.uk	16	Lab
1162-PF-GM-NGWA-SM	Genomic variation and antimalarial resistance evolution in The Gambia	Alfred Amambua-Ngwa angwa@mrc.gm	407	Gambia/North Bank, Senegal/Sedhiou, Gambia/Western
1164-PF-ML-DJIMDE-SM	*Plasmodium falciparum* clearance times in Malian villages following artesunate monotherapy	Abdoulaye Djimdé adjimde@icermali.org	419	Mali/Sikasso, Mali/Koulikoro

1165-PF-CM-APINJOH-SM	Prevalence of gene polymorphisms in symptomatic and asymptomatic *Plasmodium falciparum* infected individuals from the Southwest region of Cameroon	Tobias Apinjoh apinjohtoby@yahoo.co.uk	56	Cameroon/Sud-Ouest
1167-PF-TZ-ISHENGOMA-SM	Surveillance of parasite populations and patterns of drug resistance, and associated parasite clearance or treatment failure in Tanzania	Deus Ishengoma deusishe@yahoo.com	347	Tanzania/Kigoma, Tanzania/Tanga
1168-PF-GH-AMENGA-ETEGO-SM	Genomic surveillance of *P. falciparum* in the Kassena-Nankana Districts, Ghana	Lucas Amenga-Etego lucasmenga@gmail.com	2440	Ghana/Upper East
1169-PF-CO-CORREDOR	Using genomic sequencing to diminish the malaria burden in the Pacific coast of Columbia	Vladimir Corredor vcorredore@unal.edu.co	151	Colombia/Cauca, Venezuela/Bolivar, Colombia/Narino
1180-PF-TRAC2-DONDORP	Tracking Artemisinin Resistance Collaboration (TRAC II) with SpotMalaria	Arjen Dondorp arjen@tropmedres.ac	249	Cambodia/Pailin, Thailand/Sisakhet, India/Tripura, India/West Bengal, India/Odisha
1181-PF-VN-THUYNHIEN	Monitoring the susceptibility of *P. falciparum* to antimalarial drugs in malaria endemic areas in southern Vietnam	Thuy-Nhien Nguyen nhienntt@oucru.org	36	Vietnam/Binh Phuoc
1182-PF-GM-DALESSANDRO-SM	Understanding the determinants of malaria heterogeneity and the	Alfred Amambua-Ngwa angwa@mrc.gm	246	Gambia/Upper River

	spatial and temporal spread of malaria in The Gambia			
1183-PF-GH-AWANDARE-SM	Alternative mechanisms for erythrocyte invasion by *Plasmodium falciparum*	Gordon Awandare gawandare@ug.edu.gh	196	Ghana/Greater Accra, Ghana/Volta, Nigeria/Lagos
1185-PF-KH-KYLE	Ancient Western Cambodian *P. falciparum* isolates	Dennis Kyle dennis.kyle@uga.edu	5	Cambodia/Koh Kong
1192-PF-ML-FAIRHURST-SM	Genomic surveillance of *Plasmodium falciparum* in Mali	Thomas E Wellems twellems@niaid.nih.gov	89	Mali/Koulikoro
1195-PF-TRAC2-DONDORP	Tracking Artemisinin Resistance Collaboration (TRAC II)	Arjen Dondorp arjen@tropmedres.ac	955	Thailand/Sisakhet, Cambodia/Pursat, Cambodia/Preah Vihear, Cambodia/Ratanakiri, Cambodia/Pailin, Bangladesh/Chittagong, Vietnam/Binh Phuoc, Myanmar/Mandalay, Myanmar/Bago, Laos/Sekong, Democratic Republic of the Congo/Kinshasa, India/Odisha, India/West Bengal, India/Tripura, Myanmar/Rakhine
1197-PF-ML-DIAKITE-SM	Multidisciplinary research for malaria control and prevention in Mali	Mahamadou Diakite mdiakite@icermali.org	450	Mali/Kayes, Mali/Koulikoro, Mali/Segou
1198-PF-METF-NOSTEN	Malaria Elimination Task Force	Francois Nosten	762	Myanmar/Kayin, Thailand/Tak

		francois@tropmedres.ac		
1199-PF-ML-LAWNICZAK	Developing SpotMalaria with partners in Mali	Mara Lawniczak mara@sanger.ac.uk	9	Mali/Koulikoro
1200-PF-GH-MAIGA-SM	Genomic surveillance of *Plasmodium falciparum* in the Ashanti region of Ghana	Oumou Maïga-Ascofaré maiga@bnitm.de	286	Ghana/Ashanti
1207-PF-KH-CNM-GENRE	Integrating genetic epidemiology as an intensified surveillance tool into the National Center for Parasitology Entomology and Malaria Control of Cambodia	Huch Cheah huch.cnm@gmail.com	154	Cambodia/Ratanakiri, Cambodia/Stueng Traeng
1208-PF-LA-CMPE-GENRE	Genetic epidemiology of *P. falciparum* malaria and associated antimalarial drug resistance in Lao PDR	Mayfong Mayxay mayfong@tropmedres.ac	910	Laos/Attapeu, Laos/Champasak, Laos/Salavan, Laos/Sekong, Laos/Savannakhet
1209-PF-VN-IMPEQN-GENRE	Genetic epidemiology of *P. falciparum* malaria and associated antimalarial drug resistance in Central Vietnam	Thuy-Nhien Nguyen nhienntt@oucru.org	652	Vietnam/Dak Lak, Vietnam/Dak Nong, Vietnam/Khanh Hoa, Vietnam/Ninh Thuan, Vietnam/Gia Lai, Vietnam/Binh Phuoc, Vietnam/Quang Tri
1223-PF-MZ-ROSANAS-URGELL	Evaluation of intermittent preventive treatment during pregnancy (IPTp) in Chókwè district, Southern Mozambique (acronym IPTpCHOKWE)	Anna Rosanas-Urgell arosanas@itg.be	92	Mozambique/Gaza, Lab

1224-PF-VN-ROSANAS-URGELL	Identification of molecular mechanisms of ACT treatment failure in Vietnam	Anna Rosanas-Urgell arosanas@itg.be	464	Vietnam/Binh Phuoc, Vietnam/Khanh Hoa, Vietnam/Ninh Thuan, Vietnam/Quang Nam, Vietnam/Quang Tri, Vietnam/Bac Lieu, Vietnam/Binh Thuan, Cambodia/Ratanakiri, Vietnam/Gia Lai
1233-PF-PG-MITA	Epidemiology of kelch13 mutants in Papua New Guinea	Toshihiro Mita tmita@juntendo.ac.jp	113	Papua New Guinea/East Sepik
1238-PF-VN-NIMPE-GENRE	Genetic epidemiology of *P. falciparum* malaria and associated antimalarial drug resistance in Vietnam	Thuy-Nhien Nguyen nhiennt@oucru.org	195	Vietnam/Binh Phuoc, Vietnam/Dak Nong, Vietnam/Gia Lai
1241-PF-GH-ANINAGYEI	Viability and pathogenicity of *Plasmodium* spp in infected blood donor units and immunological and genetic markers associated with malaria infections	Enoch Aninagyei eaninagyei@uhas.edu.gh	168	Ghana/Greater Accra, Ghana/Eastern
1247-PF-SD-HAMID-SM	Surveillance of antimalarial drug resistance related genes in *P. falciparum* in Sudan	Muzamil Abdel Hamid mahdi@iend.org	190	Sudan/Khartoum, Sudan/Blue Nile

Supplementary Table 3. Summary of discovered variant positions. We divide variant positions into those containing single nucleotide polymorphisms (SNPs) and non-SNPs (indels and combinations of SNPs and indels at the same position). We then further sub-divide each of these into those within exons (coding) and those in intronic or intergenic regions (non-coding). We further sub-divide SNPs into those containing only two alleles (bi-allelic) or those containing three or more alleles (multi-allelic). Discovered variant positions are unique positions in the reference genome where either SNP or indel variation was discovered by our analysis pipeline. Pass variant positions are the subset of discovered positions that passed our quality filters. Alleles per pass position shows the mean number of distinct alleles at each pass position; biallelic variants have two alleles by definition.

Type	Coding	Multi-allelic	Discovered variant positions	Pass variant positions	% pass	Alleles per pass position
SNP	Coding	Bi-allelic	2,215,203	1,668,246	75%	2.0
		Multi-allelic	476,423	395,788	83%	3.1
	Non-coding	Bi-allelic	1,360,243	845,642	62%	2.0
		Multi-allelic	345,932	216,045	62%	3.1
non-SNP	Coding		1,931,286	798,903	41%	3.5
	Non-coding		3,816,574	1,944,035	51%	3.8
Total			10,145,661	5,868,659	58%	2.9

Supplementary Table 4. Numbers of samples used to determine proportions in Table 2.

Marker	Associated with resistance to	South America (n=154-158)	Africa - West (n=5234-6233)	Africa - Central (n=397-520)	Africa - Northeast (n=120-170)	Africa - East (n=1373-1532)	Asia - South - East (n=164-233)	Asia - South - Far East (n=1212-1369)	Asia - Southeast - West (n=1657-1876)	Asia - Southeast - East (n=2059-3684)	Oceania - New Guinea (n=298-341)
crt 76T	Chloroquine	155	5660	397	157	1388	217	1326	1871	3665	333
dhfr 108N	Pyrimethamine	154	5589	517	170	1476	201	1369	1876	3684	333
dhps 437G	Sulfadoxine	154	5529	501	162	1424	220	1291	1875	3609	333
mdr1 2+ copies	Mefloquine	158	5515	478	123	1509	164	1268	1782	3461	314
kelch13 WHO list	Artemisinin	158	5595	505	144	1513	189	1341	1768	3475	305
plasmepsin 2-3 2+ copies	Piperaquine	158	5464	519	120	1512	172	1272	1790	3428	298
dhfr triple mutant	SP (treatment)	158	5234	440	167	1373	230	1212	1833	3619	339
dhfr and dhps sextuple mutant	SP (IPTp)	158	6233	510	170	1446	233	1217	1657	2059	341
kelch13 and mdr1	AS-MQ	158	5915	519	150	1532	203	1354	1798	3495	324
kelch13 and plasmepsin 2-3	DHA-PPQ	158	5823	520	148	1526	199	1355	1829	3442	316

Supplementary Table 5. Newly emerging Dd2 background mutations in *crt*. This table shows all *crt* haplotypes with a genetic background identical to the lab strain Dd2. Dd2 is derived from an isolate taken from a patient in Indochina in 1980. Numbers indicate the numbers of samples with the haplotype in each major sub-population. Note that only 4 samples with a mutation on a Dd2 background are from outside the eastern SE Asia region.

		Population										
Number of mutations on Dd2 background	Haplotype	SA	AF-W	AF-C	AF-NE	AF-E	AS-S-E	AS-S-FE	AS-SE-W	AS-SE-E	OC-NG	Total
0	Dd2						31	185	1566	756		2538
1	Dd2+R34L								2			2
	Dd2+C72Y									2		2
	Dd2+N88K									25		25
	Dd2+T93S									359		359
	Dd2+H97L								1	47		48
	Dd2+H97Y									77		77
	Dd2+M104K									1		1
	Dd2+F145I									103		103
	Dd2+L196P									2		2
	Dd2+I218F									203		203
	Dd2+T256I									4		4
	Dd2+L308S									1		1
	Dd2+M343I									20		20
	Dd2+M343L									4		4
	Dd2+G353V									37		37
	Dd2+A359S									1		1
	Dd2+G367C									18		18
	Dd2+V370E									1		1
	Dd2+R392H									1		1
2	Dd2+N88K+A195V									1		1
	Dd2+H97L+V141L						1					1
	Dd2+F145I+T93I									2		2
	Dd2+F145I+C171Y									2		2
	Dd2+F145I+A195V									1		1
	Dd2+F145I+C258W									13		13
	Dd2+F145I+T342A									3		3
	Dd2+F145I+R392G									2		2
	Dd2+L196P+I347V									1		1
	Dd2+I218F+A195V									16		16
	Dd2+I218F+T342S									1		1
	Dd2+I218F+S388F									1		1
	Dd2+T256I+T230I									1		1
3	Dd2+I218F+A195V+G302D									1		1

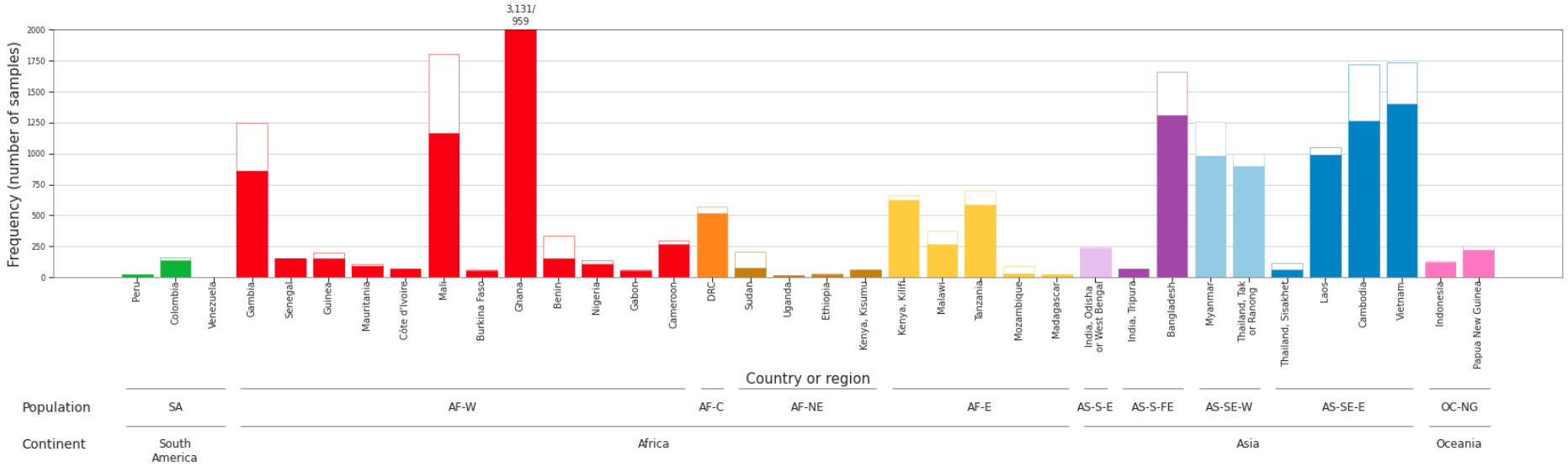
Supplementary Table 6. Frequency of HRP2 and HRP3 deletions by country. n=number of QC pass samples. Calls columns show number of samples for which an unambiguous deletion genotype (deleted or non-deleted) could be assigned.

Country	*hrp2* calls	% *hrp2* deletions	*hrp3* calls	% *hrp3* deletions	*hrp2* and *hrp3* calls	% *hrp2* and *hrp3* deletions
Bangladesh (n=1,310)	939	0%	850	0%	819	0%
Benin (n=150)	110	0%	104	0%	100	0%
Burkina Faso (n=57)	43	0%	32	0%	32	0%
Cambodia (n=1,267)	1,109	0%	1,091	3%	1,064	0%
Cameroon (n=264)	244	0%	240	0%	235	0%
Colombia (n=135)	124	0%	123	41%	118	0%
Côte d'Ivoire (n=71)	70	0%	71	0%	70	0%
DRC (n=520)	413	0%	392	0%	385	0%
Ethiopia (n=21)	20	0%	20	45%	20	0%
Gabon (n=55)	34	0%	38	0%	32	0%
Gambia (n=863)	517	0%	467	1%	460	0%
Ghana (n=3,131)	1,529	0%	1,448	0%	1,343	0%
Guinea (n=151)	121	0%	119	0%	119	0%
India (n=300)	75	0%	70	1%	68	0%
Indonesia (n=121)	117	4%	117	37%	115	2%
Kenya (n=690)	660	0%	647	0%	645	0%
Laos (n=991)	773	0%	717	2%	669	0%
Madagascar (n=24)	24	0%	22	0%	22	0%
Malawi (n=265)	265	0%	264	0%	264	0%
Mali (n=1,167)	709	0%	691	0%	669	0%
Mauritania (n=92)	79	0%	81	0%	79	0%
Mozambique (n=34)	10	0%	10	0%	6	0%
Myanmar (n=985)	645	0%	606	0%	585	0%
Nigeria (n=110)	34	0%	30	0%	30	0%
Papua New Guinea (n=221)	118	0%	106	0%	106	0%
Peru (n=21)	21	38%	20	75%	20	30%
Senegal (n=150)	142	0%	141	4%	138	0%
Sudan (n=76)	7	14%	7	86%	7	14%
Tanzania (n=589)	452	0%	470	0%	440	0%
Thailand (n=954)	855	0%	846	0%	823	0%
Uganda (n=12)	12	0%	12	0%	12	0%
Venezuela (n=2)	2	0%	2	0%	2	0%
Vietnam (n=1,404)	762	0%	740	1%	670	0%

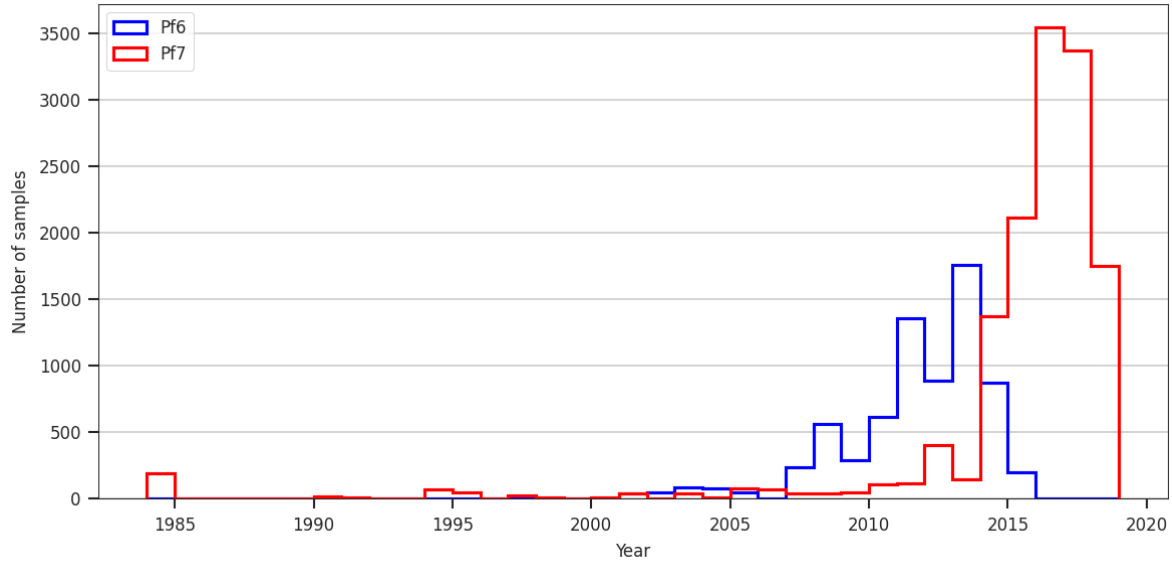
Supplementary Table 7. Summary of *hrp2* and *hrp3* deletion breakpoints. Telomere healing refers to the process whereby the end of a chromosome is deleted and a telomere repeat sequence attached to the breakpoint. Chromosome 11 recombination refers to a new hybrid chromosome being created by a recombination between chromosome 13 and 11 at a cluster of rRNA genes that appear to have orthologous copies on both chromosomes. Chromosome 5 recombination refers to a recombination between chromosome 13 and an inverted section of the middle of chromosome 5 containing the gene *mdr1*. For telomere healing an exact breakpoint position is given but for recombination events it is only possible to give a region in which the recombination has occurred.

Gene	Deletion type	Breakpoint coordinates	Countries	Samples with deletion
hrp2	Telomere healing	Pf3D7_08_v3:1373732	Cambodia	1
		Pf3D7_08_v3:1374280	Sudan	1
		Pf3D7_08_v3:1374462	Indonesia	5
		Pf3D7_08_v3:1374932	Peru	2
		Pf3D7_08_v3:1374986	Peru	6
hrp3	Chromosome 11 recombination	Pf3D7_13_v3:2800004-2807159	Thailand, Ghana, Indonesia, Peru, Bangladesh, Vietnam, Colombia, Ethiopia, Senegal, Laos, Cambodia, Sudan, Mali, Gambia	151
	Chromosome 5 recombination	Pf3D7_13_v3:2835587-2835612	Cambodia, Vietnam	21
	Telomere healing	Pf3D7_13_v3:2811525	India	1
		Pf3D7_13_v3:2812344	Sudan	1
		Pf3D7_13_v3:2815249	Tanzania	1
		Pf3D7_13_v3:2822480	Ghana	1
		Pf3D7_13_v3:2823645	Kenya	1
		Pf3D7_13_v3:2830952	Cambodia	7
		Pf3D7_13_v3:2832080	Democratic Republic of the Congo	1
		Pf3D7_13_v3:2834604	Vietnam	1
		Pf3D7_13_v3:2835532	Thailand	1
		Pf3D7_13_v3:2837145	Vietnam	7
		Pf3D7_13_v3:2837392	Cambodia, Laos	3
		Pf3D7_13_v3:2838654	Indonesia	2
		Pf3D7_13_v3:2841024	Thailand	1
		Pf3D7_13_v3:2841120	Indonesia	1

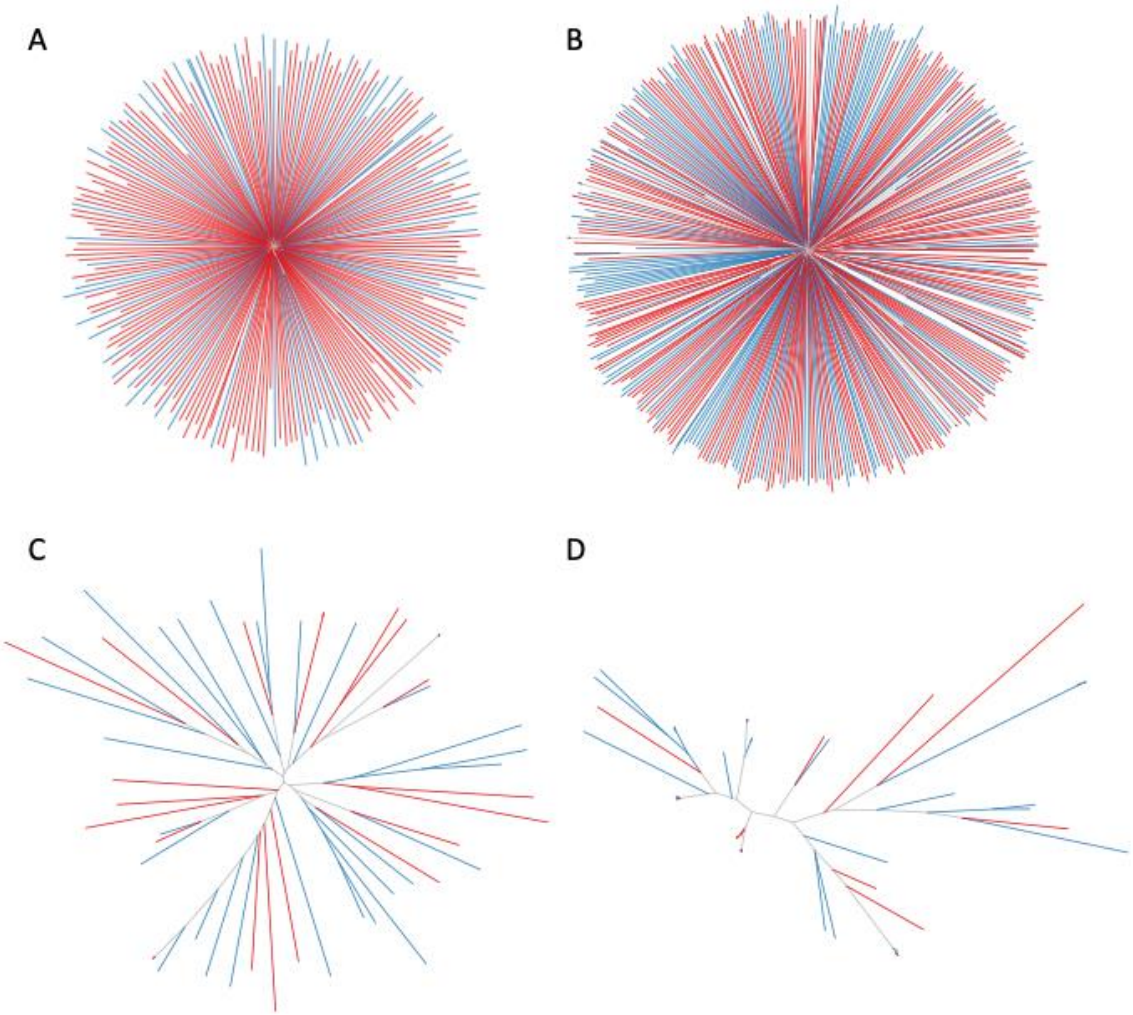
Supplementary Figure 1. Breakdown of samples by country. Solid bars indicate samples which passed QC. Unfilled bars represent samples that failed QC. The y-axis is truncated at 2,000 samples, with the numbers of QC pass/QC fail samples in Ghana shown above the bar. Bars are coloured according to the major sub-population to which the location is assigned.



Supplementary Figure 2. Distribution of samples by year of collection. The blue line shows samples in our previous Pf6 release. The red line shows samples that are newly released in Pf7



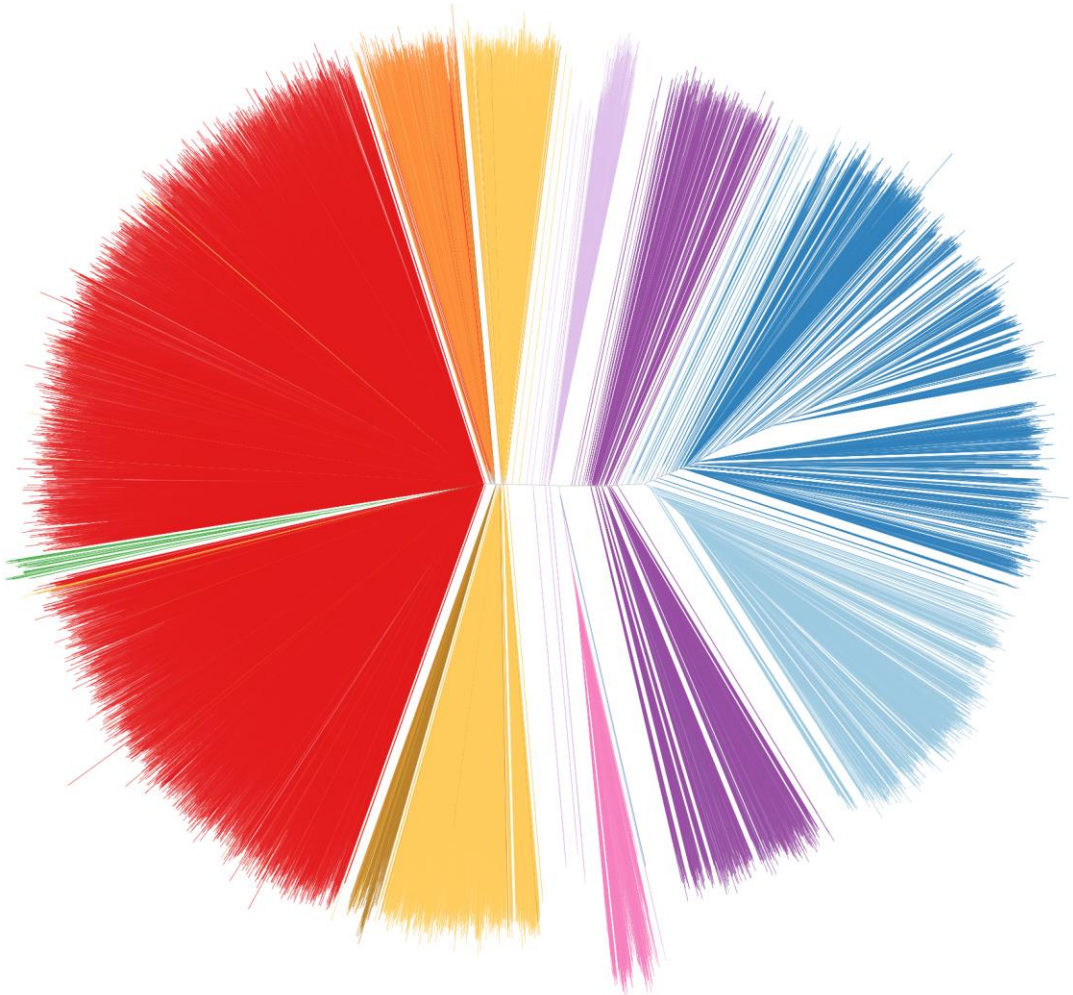
Supplementary Figure 3. Lack of bias in population structure due to use of sWGA. Genome-wide unrooted neighbour-joining trees showing population structure in samples from four locations for which one subset of samples were sequenced from genomic DNA (gDNA) material (shown as blue lines), and a second subset of samples were sequenced using sWGA material (shown in red). (**A**) Ghana, Upper East, 2015 (gDNA *n*=61, sWGA *n*=164). (**B**) Kenya, Kilifi, 2007-2012 (gDNA *n*=151, sWGA *n*=222). (**C**) Cambodia, Pursat, 2016 (gDNA *n*=30, sWGA *n*=22). (**D**) Vietnam, Binh Phuoc, 2016 (gDNA *n*=39, sWGA *n*=22). Note there are greater levels of population structure in the samples from SE Asia (**C** and **D**) than there are in African samples (**A** and **B**), though in all cases there is no obvious clustering by sample type.



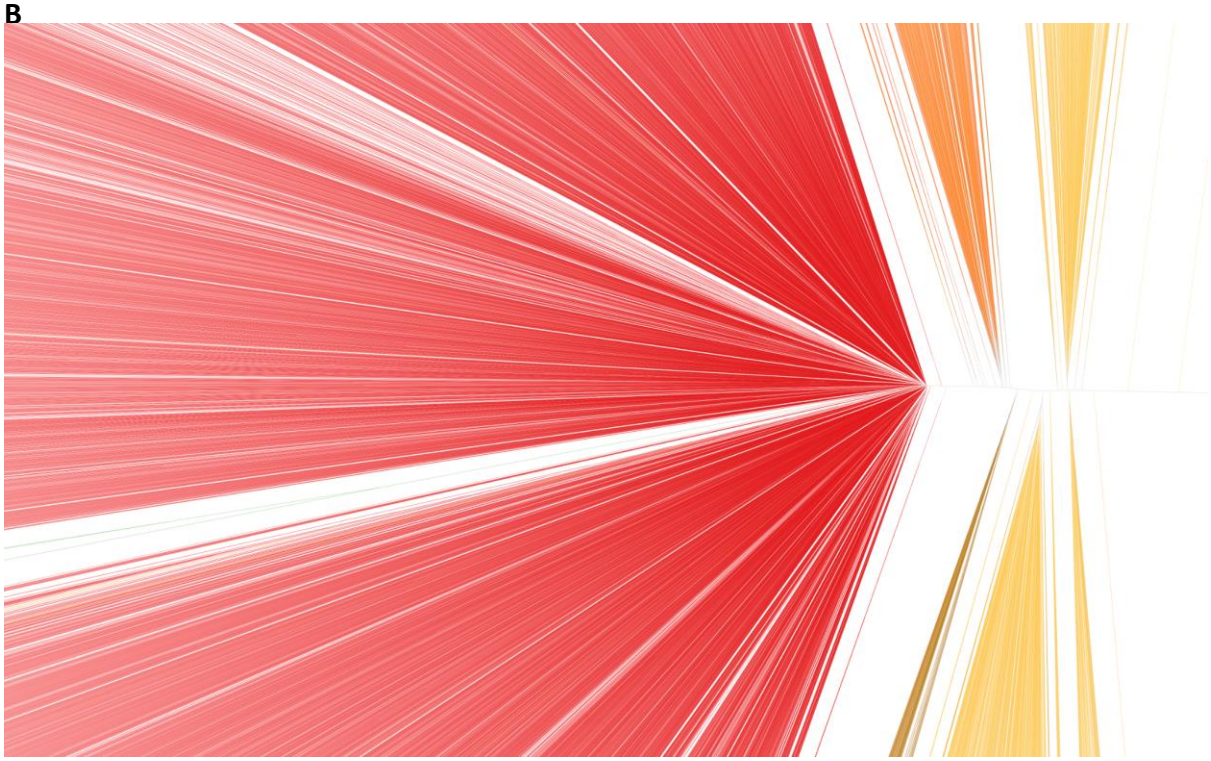
Supplementary Figure 4. Population structure from a neighbour-joining tree.

(**A**) Genome-wide unrooted neighbour-joining tree showing population structure across all locations, with sample branches coloured according to major sub-populations (Table 1): South America (green, *n*=158); Africa - West (red, *n*=6,262); Africa - Central (orange, *n*=520); Africa - Northeast (light brown, *n*=172); Africa - East (yellow, *n*=1,539); Asia - South - East (light purple, *n*=233); Asia - South - Far East (dark purple, *n*=1,377); Asia - Southeast - West (light blue; *n*=1,880); Asia - Southeast - East (dark blue; *n*=3,721); Oceania - New Guinea (magenta; *n*=342). (**B**) Magnified view of the part of the tree where the majority of samples from Africa coalesce, showing that the four African sub-populations are genetically close but distinct.

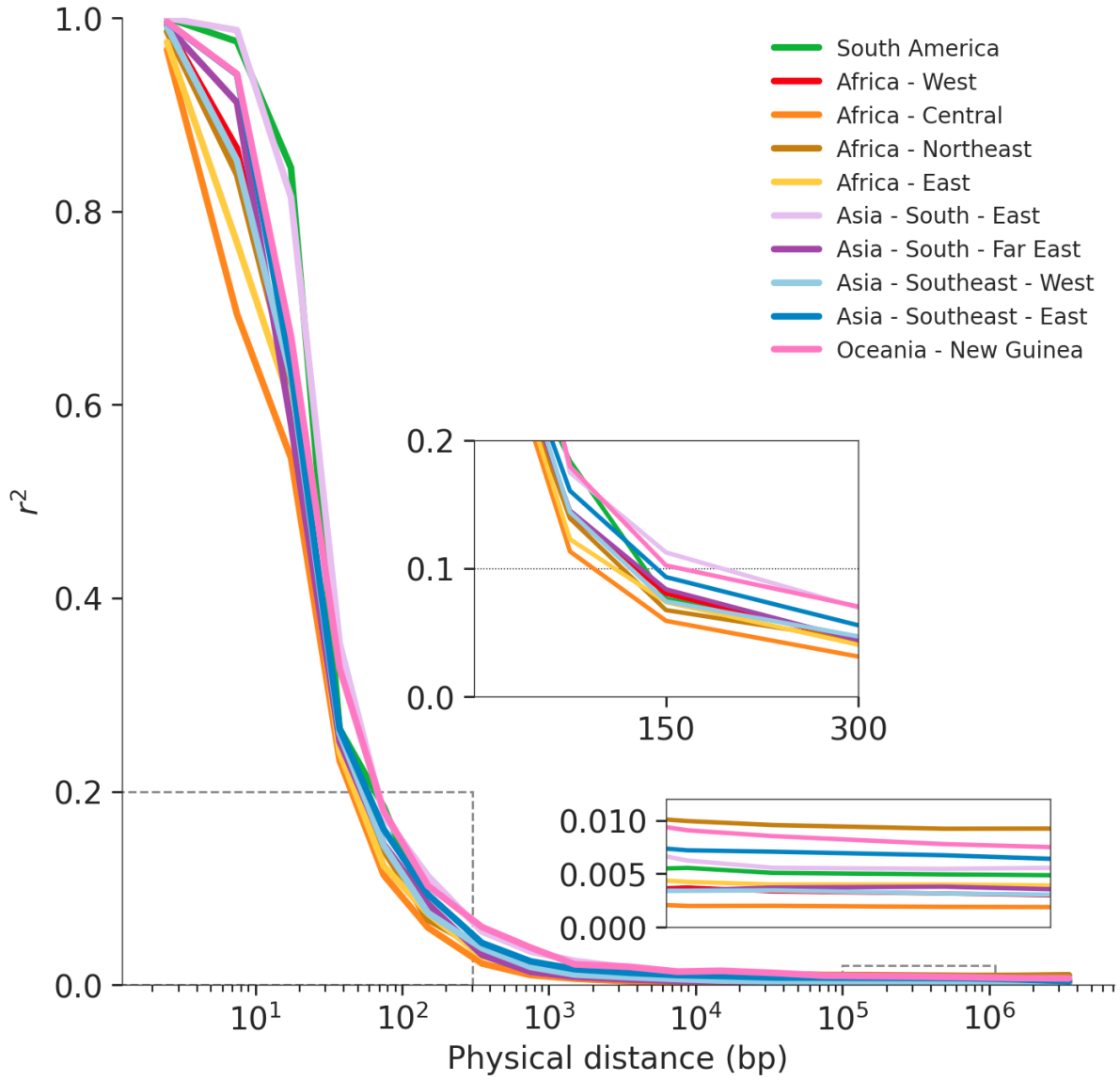
A



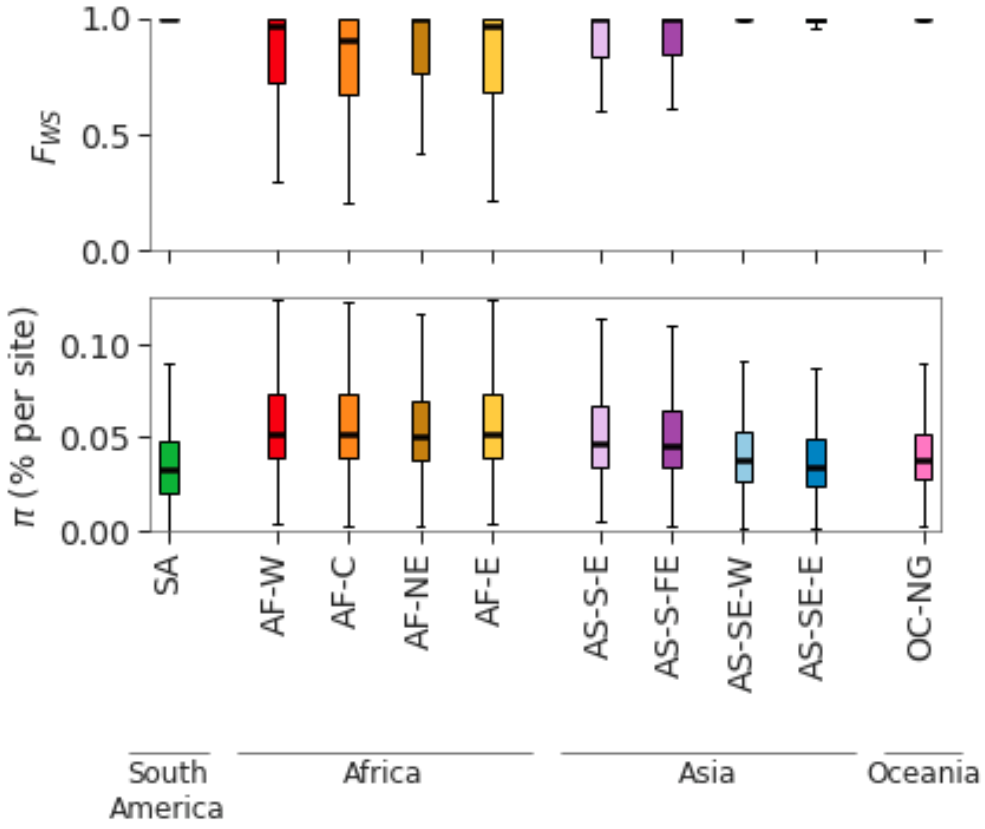
B



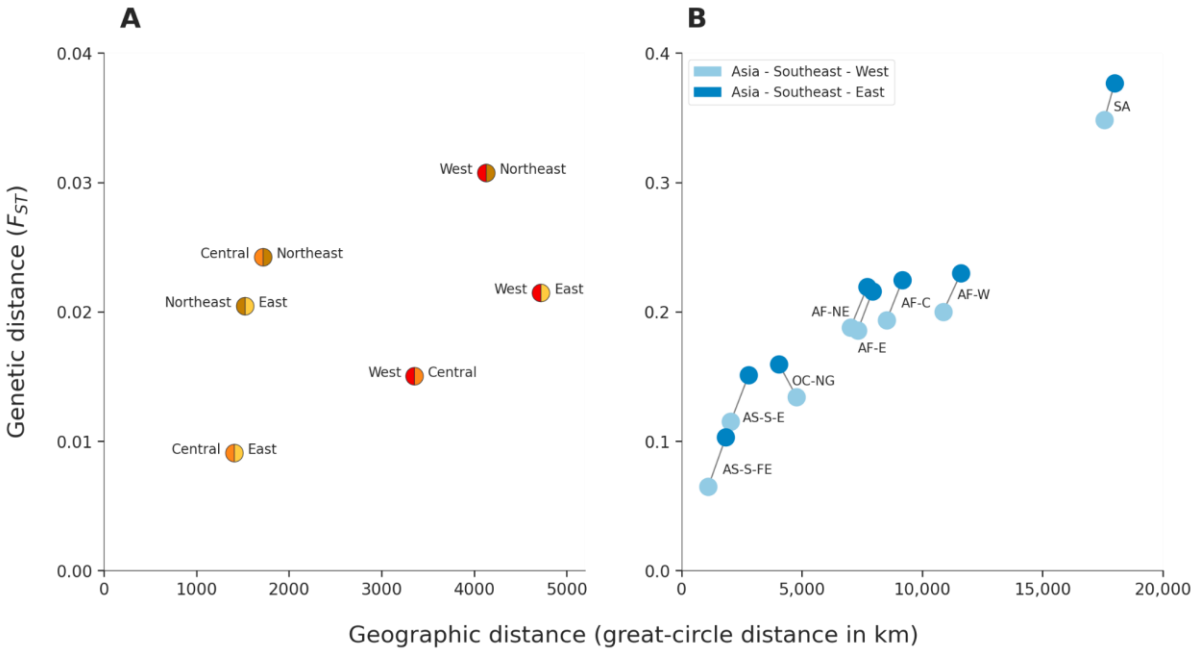
Supplementary Figure 5. Linkage disequilibrium decay in ten major parasite sub-populations. Genome-wide median LD (y-axis, measured by r^2) between pairs of SNPs as a function of their physical distance (x-axis, in bp), showing a rapid decay in all regional parasite sub-populations. The upper inset panel shows a magnified view of the decay, showing that in all sub-populations r^2 decayed below 0.1 (dashed horizontal line) within 250 bp. The lower inset panel shows r^2 for distance between 100 kbp and 1 Mbp, showing Northeast Africa has the highest level of long range LD, possibly reflecting the presence of highly related parasites in different samples.



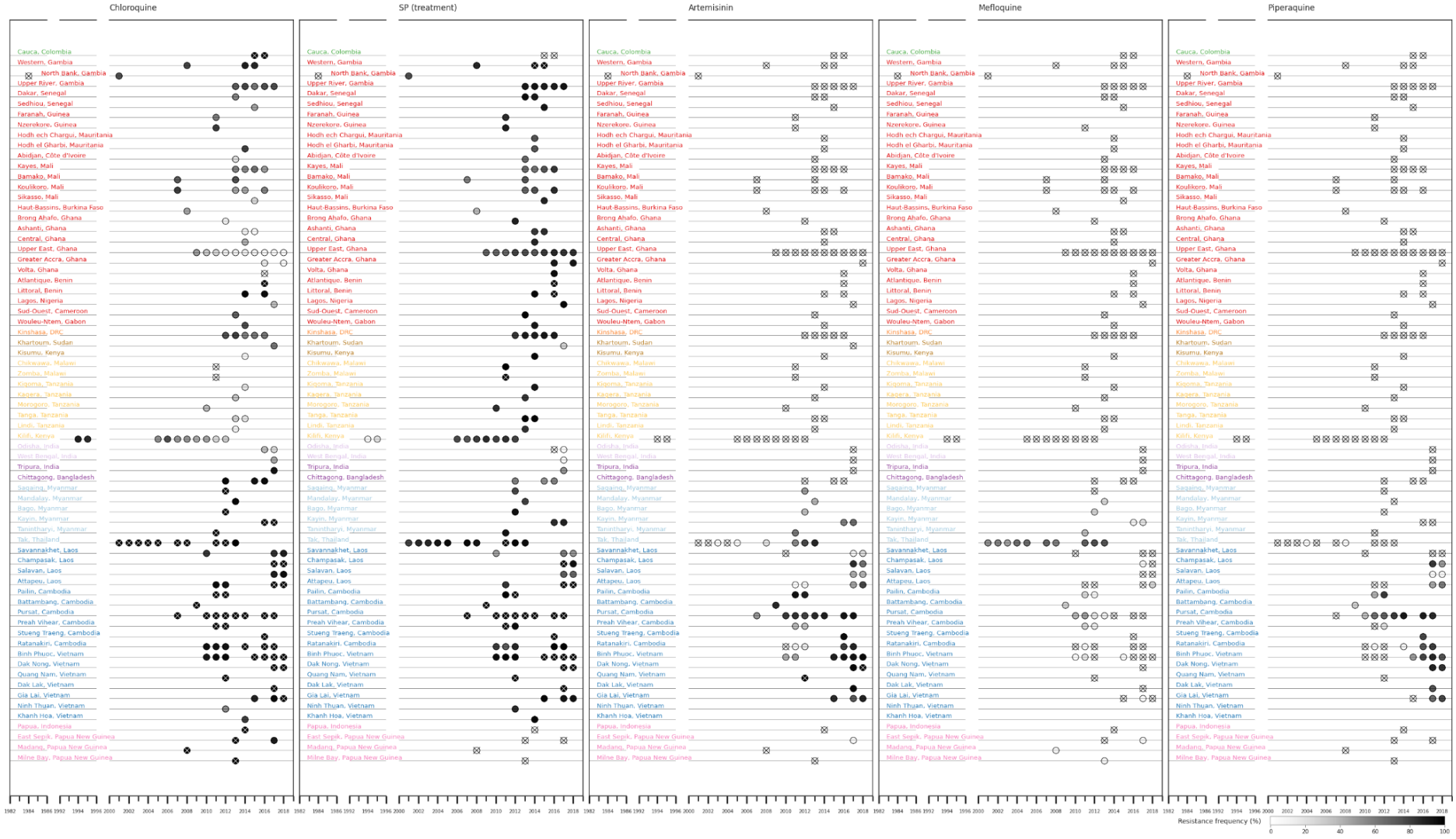
Supplementary Figure 6. Characteristics of the ten major parasite sub-populations. Upper panel shows distribution of within-host diversity, as measured by F_{WS}, showing that genetically mixed infections were considerably more common in Africa and to a lesser extent South Asia than other regions, consistent with the high intensity of malaria transmission in Africa. Lower panel shows distribution of per site nucleotide diversity calculated in non-overlapping 25kbp genomic windows. We only considered coding SNPs to reduce the ascertainment bias caused by poor accessibility of non-coding regions. In both panels, thick lines represent median values, boxes show the interquartile range, and whiskers represent the bulk of the distribution, discounting outliers.



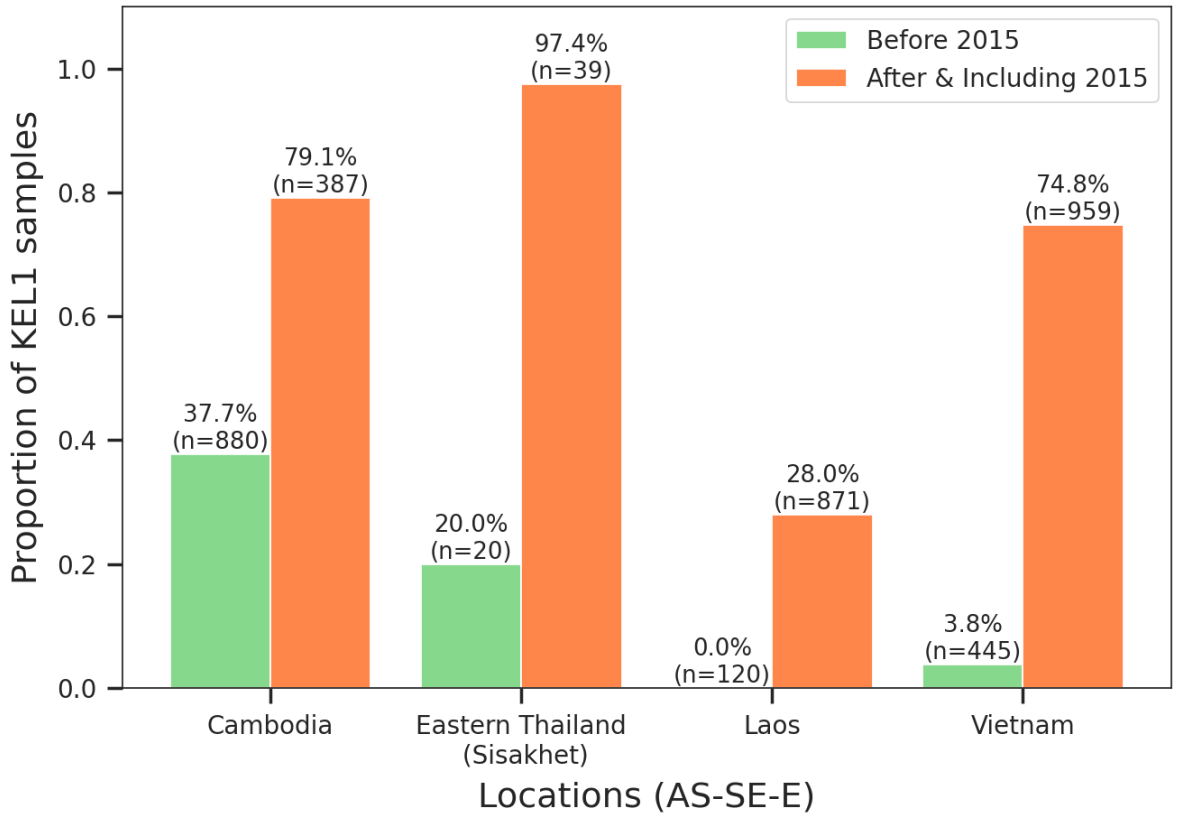
Supplementary Figure 7. Geographic patterns of population differentiation and gene flow. Each point represents one pairwise comparison between two regional parasite sub-populations. The x-axis reports the geographic separation between the two sub-populations, measured as great-circle distance between the centre of mass of each sub-population and without taking into account natural barriers. The y-axis reports the genetic differentiation between the two sub-populations, measured as average genome-wide F_{ST}. These two figures show that the sub-populations from northeast Africa and the eastern part of SE Asia are more genetically distinct from other sub-populations than might be expected due to their geographic separation. (**A**) Comparison of African sub-populations. Points are coloured based on the two sub-populations they represent. The distance from northeast Africa (AF-NE) to other African sub-populations is generally greater than that between the other African sub-populations. For example the central African sub-population (AF-C) is closer genetically to the west African sub-population (AF-W) than it is to the northeast African sub-population, despite being closer geographically to the latter. (**B**) Comparison of two SE Asian sub-populations against all other sub-populations. Compared to all other sub-populations, the sub-population from the eastern part of SE Asia (AS-SEA-E) generally has a greater genetic distance than that from the western part of SE Asia, and this difference in genetic distances is more than might be expected due to the extra geographic distance.



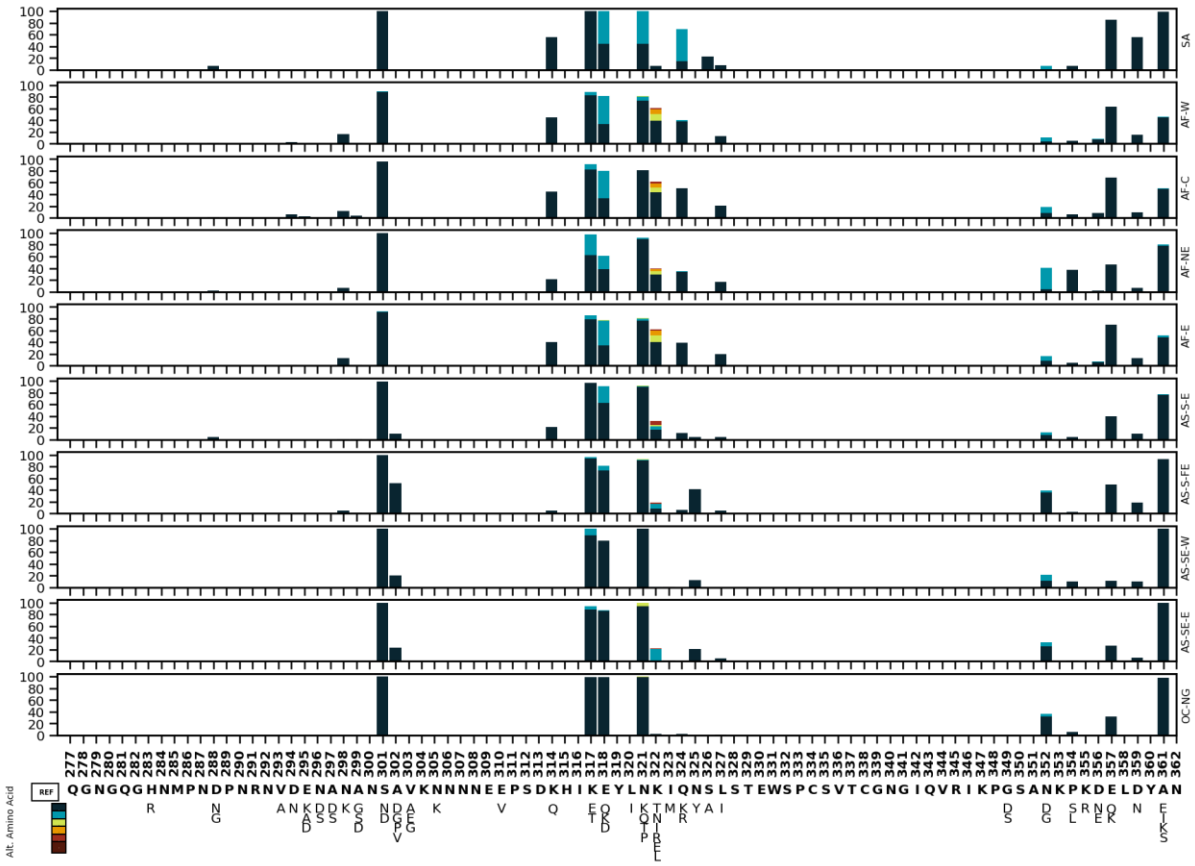
Supplementary Figure 8. Abacus plot of inferred drug resistance frequencies in location/year combinations. This shows each combination of location (first-level administrative division) and year for which we have at least 25 samples with an inferred drug resistance phenotype called. For each such combination we estimate the frequency of inferred resistance. Each line represents a first-level administrative division. Text on each line is coloured according to the major sub-population that location is within. Each point represents a year within that administrative division for which at least 25 samples have a phenotype call (resistant or sensitive) for the particular drug. The shade of the point represents the frequency of drug resistance in that year from white (0%) to black (100%). Where frequency is exactly 0% or 100% the point is marked with a cross to represent fixation.


Chloroquine
SP (treatment)
Artemisinin
Mefloquine
Piperaquine
Cauca, Colombia
Western, Gambia
North Bank, Gambia
Upper River, Gambia
Dakar, Senegal
Sedhiou, Senegal
Faranah, Guinea
Nzerekore, Guinea
Hodh ech Chargui, Mauritania
Hodh el Gharbi, Mauritania
Abidjan, Côte d'Ivoire
Kayes, Mali
Bamako, Mali
Koulikoro, Mali
Sikasso, Mali
Haut-Bassins, Burkina Faso
Brong Ahafo, Ghana
Ashanti, Ghana
Central, Ghana
Upper East, Ghana
Greater Accra, Ghana
Volta, Ghana
Atlantique, Benin
Littoral, Benin
Lagos, Nigeria
Sud-Ouest, Cameroon
Wouleu-Ntem, Gabon
Kinshasa, DRC
Khartoum, Sudan
Kisumu, Kenya
Chikwawa, Malawi
Zomba, Malawi
Kigoma, Tanzania
Kagera, Tanzania
Morogoro, Tanzania
Tanga, Tanzania
Lindi, Tanzania
Kilifi, Kenya
Odisha, India
West Bengal, India
Tripura, India
Chittagong, Bangladesh
Saqaing, Myanmar
Mandalay, Myanmar
Bago, Myanmar
Kayin, Myanmar
Tanintharyi, Myanmar
Tak, Thailand
Savannakhet, Laos
Champasak, Laos
Salavan, Laos
Attapeu, Laos
Pailin, Cambodia
Battambang, Cambodia
Pursat, Cambodia
Preah Vihear, Cambodia
Stueng Traeng, Cambodia
Ratanakiri, Cambodia
Binh Phuoc, Vietnam
Dak Nong, Vietnam
Quang Nam, Vietnam
Dak Lak, Vietnam
Gia Lai, Vietnam
Ninh Thuan, Vietnam
Khanh Hoa, Vietnam
Papua, Indonesia
East Sepik, Papua New Guinea
Madang, Papua New Guinea
Milne Bay, Papua New Guinea
1982 1984 1986 1992 1994 1996 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018
Resistance frequency (%)
0 20 40 60 80 100

Supplementary Figure 9. Increase in frequency of KEL1. The bars represent the proportions of all QC pass samples that are KEL1, with green bars showing samples collected prior to 2015 and orange bars samples from 2015 onwards. There is a dramatic increase in frequency of KEL1 across all four countries which make up the eastern part of SE Asia. n=number of QC pass samples.



Supplementary Figure 10. Variation in c-terminal of *csp*. The x-axis shows amino acid positions and the y-axis the proportion of non-reference alleles in different sub-populations. Different alternative amino acids are represented by different colours as represented by the legend below the x-axis.



Supplementary Figure 11. Proportion of C allele of *eba175* in different major sub-populations. Horizontal line shows 50% proportion.