Supporting Information Appendix

Divergence Island SNP (DIS) genotyping method

Single nucleotide polymorphisms reported to be fixed between M and S molecular forms were selected for genotyping from four different studies (1-4). We used Typer® AssayDesigner software (Sequenom, San Diego, CA) to devise a multiplex SNP genotype assay consisting of 15 SNPs that occur on all three chromosomes; 7 on the X, 5 on the 2L, 3 on the 3L (Fig. S1). Details of SNPs genotyped are provided in Table S1.

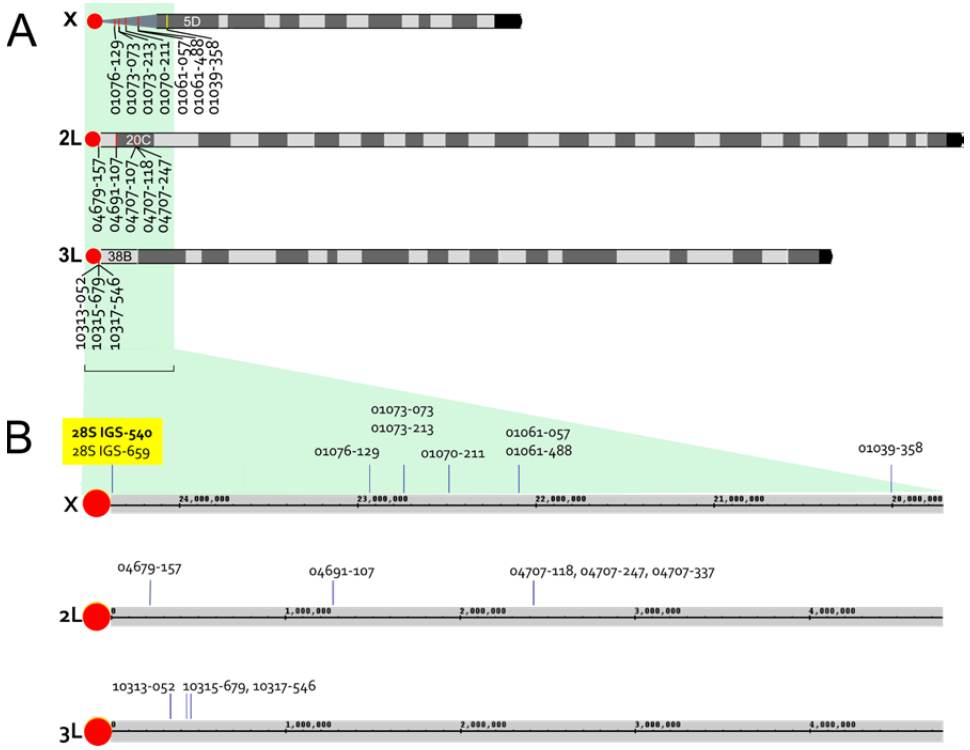


Fig. S1. Physical map of divergence island SNPs (DIS) used in this study. (A) locations on chromosome arms. (B) blowup of regions containing DIS illustrating the relative position of each. Red circles represent centromeres. 28S IGS-540 and 28S IGS-659, highlighted in yellow in (B), are the multi-copy 28S rDNA SNPs used to define the M and S forms in previous studies.

Table S1. Details of SNPs utilized for M, S and hybrid characterization.

SNP ID	Chromosome	Genome coordinate	M variant	S variant	reference
01076-129	X	22,944,682	T	G	Stump et al. 2005
01073-073	X	22,750,572	G	T	Stump et al. 2005
01073-213	X	22,750,432	G	A	Stump et al. 2005
01070-211	X	22,497,157	A	G	Turner et al. 2005
01061-057	X	22,105,860	T	C	Stump et al. 2005
01061-488	X	22,10,5429	A	T	Stump et al. 2005
01039-358	X	20,015,634	C	A	Stump et al. 2005
04679-157	2L	209,536	C	T	White et al. 2010
04691-107	2L	1,274,353	A	G	Turner& Hahn 2007; Turner et al. 2005
04707-118	2L	2,430,786	C	T	Turner et al. 2005
04707-247	2L	2,430,915	A	G	Turner et al. 2005
04707-337	2L	2,431,005	C	T	Turner et al. 2005
10313-052	3L	296,897	G	A	White et al. 2010
10315-679	3L	387,877	G	A	White et al. 2010
10317-546	3L	413,944	T	C	White et al. 2010

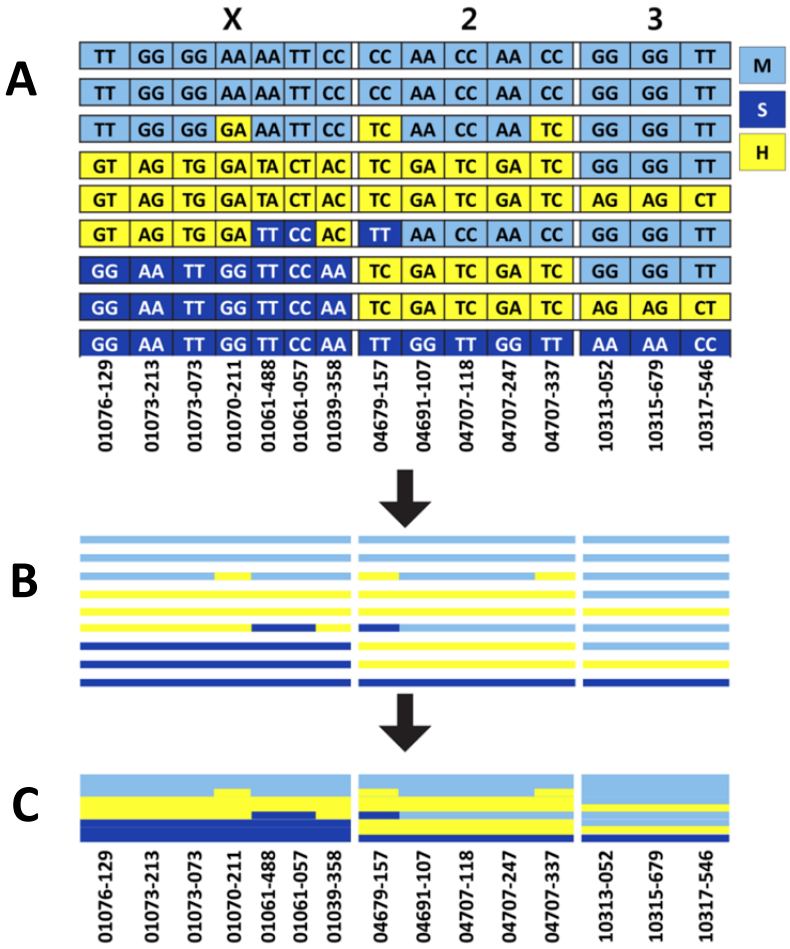
The Sequenom iPLEX®Gold assay was used for SNP genotyping following the manufacturer's protocol (5). This genotyping method utilizes the MALDI-TOF mass spectrometry to determine genotypes based on the mass of allele-specific fragments (6-8). Mass spectrogram visualization and genotype calls were conducted using TyperAnalyzer software version 4.0 (Sequenom, San Diego, CA).

A sample of multi-locus genotypes generated by the DIS method are presented for illustration in Table S2. In Table S2, each row represents an individual mosquito sample. Each column represents a locus. Loci are grouped by linkage group (X, 2L and 3L). Homozygotes for M alleles (MM) are marked in light blue, homozygotes for S alleles (SS) are marked in dark blue, and heterozygotes (MS) are marked in yellow. Consensus molecular form is determined based on all 15 locus genotypes. We allowed up to 2 mismatched calls. Individuals that have more than 13 MM genotypes were called M form. Individuals carrying more than 13 SS genotypes were called S form. Individuals carrying more than 13 MS genotypes were called F_1. And the rest were called as backcross (F_{1+n}). We employ the method described in Fig. S2 to illustrate the distribution of DIS genotypes among each study population, using what we refer to as “DIS maps”.

Table S2. A sample of multi-locus SNP genotyping results.

Sample	consensus molecular form	X							2L					3L		
		01076-129	01073-213	01073-073	01070-211	01061-488	01061-057	01039-358	04679-157	04691-107	04707-118	04707-247	04707-337	10313-052	10315-679	10317-546
06FOUM030	S	GG	AA	TT	GG	TT	CC	AA	TT	GG	TT	GG	TT	AA	AA	CC
06FOUM031	S	GG	AA	TT	GG	TT	CC	AA	TT	GG	TT	GG	TT	AA	AA	CC
02KON014	M	TT	GG	GG	AA	AA	TT	CC	CC	AA	CC	AA	CC	GG	GG	TT
02KON019	M	TT	GG	GG	AA	AA	TT	CC	CC	AA	CC	AA	CC	GG	GG	TT
06NGAL009	F_1	GT	AG	TG	GA	TA	CT	AC	TC	GA	TC	GA	TC	AG	AG	CT
06NGAL020	F_1	GT	AG	TG	GA	TA	CT	AC	TC	GA	TC	GA	TC	AG	AG	CT
03Tiko044	F_{1+n}	GG	AA	TT	GG	TT	CC	AA	TC	GA	TC	GA	TC	AA	AA	CC
03Tiko039	F_{1+n}	GG	AA	TT	GG	TT	CC	AA	TT	GG	TT	GG	TT	AG	AG	CT
09ABU007	F_1	GT	AG	TG	GA	TA	CT	AA	TC	GA	TC	GA	TC	AG	AG	CT
09ABU008	F_{1+n}	GG	AA	TT	GG	TT	CC	AA	CC	AA	CC	AA	CC	AA	AA	CC
09ABU019	F_{1+n}	GG	AA	TT	GG	TT	CC	AA	TT	GG	TC	GA	TC	GG	GG	TT

Fig. S2. Construction of DIS maps to describe the distribution of SNP genotypes among individuals within populations. (A) Individual multi-locus genotypes at each of the 15 loci are arranged in rows, individual loci in columns. Genotypes are organized by linkage groups, X chromosome, chromosome 2 and 3. Each genotype is color-coded, light blue for M form associated genotypes, dark blue for S and yellow for heterozygotes. (B) Each row is compressed into a series of 3 bars, each representing a linkage group. (C) Each bar is further compressed resulting in a thin line and lines are stacked into a series of rectangles facilitating visualization of the population with respect to the distribution of genotypes at the 15 SNPs.

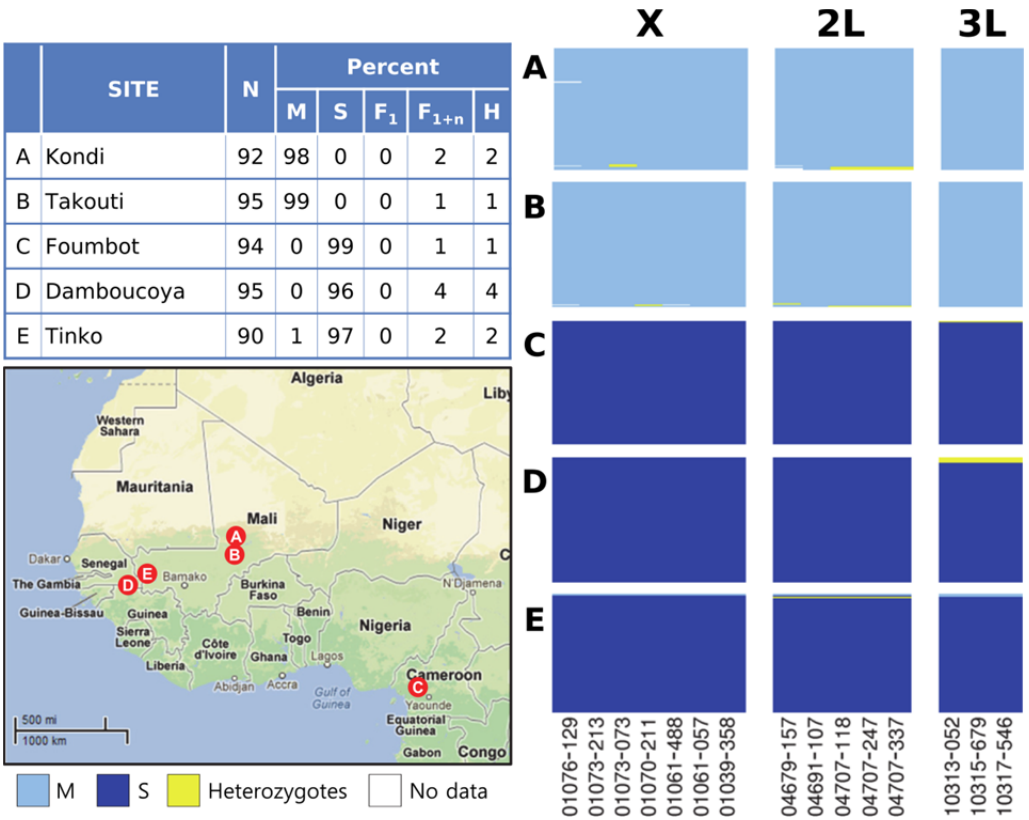


DIS maps for national populations

Analysis of DIS genotypes identified 5 classes of populations: (1) allopatric populations of the M and S forms (Fig. S3), (2) sympatric M and S populations with past introgression and no ongoing gene flow (Fig. S4), (3) populations with past introgression and loss of divergence on 2L (Fig. S5), (4) populations with ongoing gene flow and introgression (Fig. S6) and (5) populations with complete or nearly complete introgression (Fig. S7).

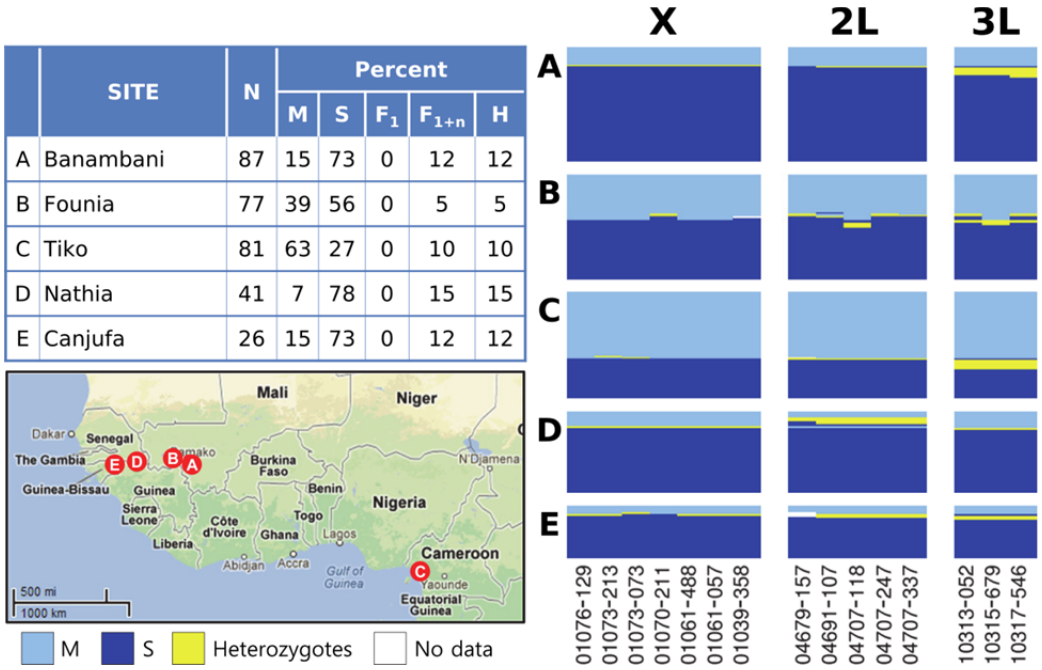
Collection site information is provided in Table S3. Summary of linkage disequilibrium (LD) results is provided in Table S4.

Fig. S3. Divergence Islands SNP (DIS) maps for allopatric populations of the M and S forms. Populations at the villages of (A) Kondi and (B) Takouti Mali were composed of M form individuals. Populations in (C) Foumbot, Cameroon and (D) Damboucoya, Senegal were made up exclusively of S form individuals and (E) Tinko, Mali consisted of 97% S form individuals.



	SITE	N	Percent				
			M	S	F_1	F_{1+n}	H
A	Kondi	92	98	0	0	2	2
B	Takouti	95	99	0	0	1	1
C	Foumbot	94	0	99	0	1	1
D	Damboucoya	95	0	96	0	4	4
E	Tinko	90	1	97	0	2	2

Fig. S4. DIS maps for sympatric M and S form populations with evidence for recent past introgression without evidence of ongoing hybridization (absence of F1 hybrids). These include populations at: (A) Banambani and (B) Founia in Mali, (C) Tiko, Cameroon, (D) Nathia, Senegal and (E) Canjufa, Guinea Bissau.



	SITE	N	Percent				
			M	S	F_1	F_{1+n}	H
A	Banambani	87	15	73	0	12	12
B	Founia	77	39	56	0	5	5
C	Tiko	81	63	27	0	10	10
D	Nathia	41	7	78	0	15	15
E	Canjufa	26	15	73	0	12	12

Fig. S5. DIS maps for sympatric M and S form populations with the loss of divergence at loci within the chromosome 2 island and no evidence of ongoing hybridization (no F1 genotypes, e.g. individuals that are simultaneously heterozygous for all ten SNPs on chromosome 3 and the X). These include the population at (A) Bioko, Equatorial Guinea and (B) Selinkenyi, Mali collected in 2012.

	SITE	N	Percent				
			M	S	F_1	F_{1+n}	H
A	Bioko	73	18	25	0	58	58
B	'12 Selinkenyi	83	8	2	0	90	90

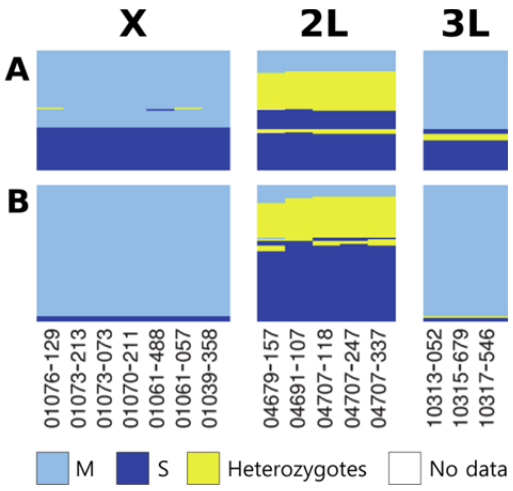


Fig. S6. DIS maps for sympatric M and S form populations with evidence of ongoing hybridization (presence of F1 genotypes) and introgression. Includes populations at (A) Bantinngoungou, Mali, (B) Njigalap, Cameroon and (C) the ENDO population from the villages of Goundry (and surrounding locations, as described in Riehle et al. 2011), Burkina Faso.

	SITE	N	Percent				
			M	S	F_1	F_{1+n}	H
A	Bantinngoungou	87	24	46	5	25	30
B	Njigalap	119	10	70	12	8	20
C	Endo	60	42	45	3	10	13

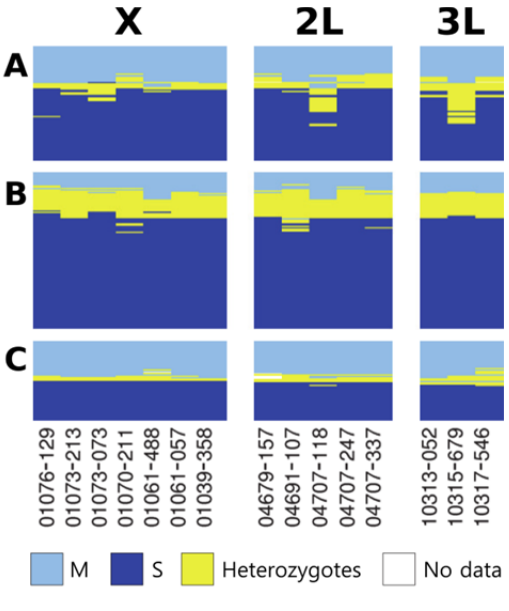
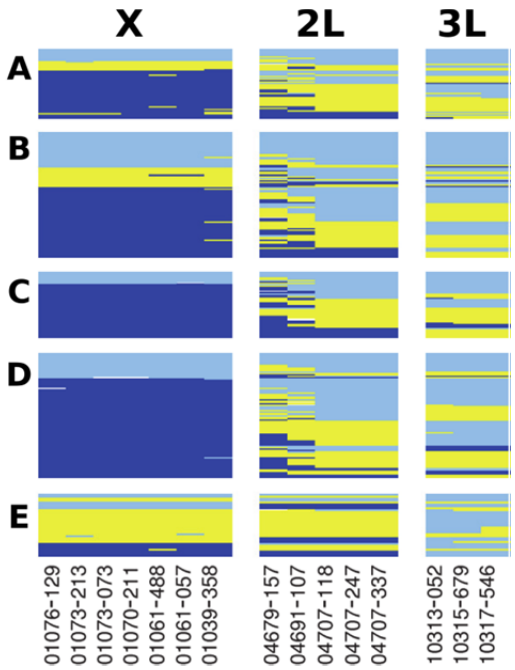


Fig. S7. DIS maps for sympatric M and S form populations with evidence of complete or nearly complete introgression. Includes populations at (A&C) females and males respectively from Abu, Guinea Bissau, (B&D) females and males respectively from Prabis, Guinea Bissau and (E) the GOUNDRY population from the villages of Goundry, Ouagadougou, Koupela, Kodougou and Soumousso, Burkina Faso.

	SITE	N	Percent				
			M	S	F_1	F_{1+n}	H
A	Abu ♀♀	50	16	2	0	82	82
B	Prabis ♀♀	92	25	0	1	74	75
C	Abu ♂♂	48	19	0	19	62	81
D	Prabis ♂♂	90	13	0	15	72	87
E	Goundry	32	3	0	0	97	97



M S Heterozygotes No data

Table S3. Collection site information. M, S, F1 and backcross frequencies are based on DIS assay result.

Site (collection)	Country	Latitude	Longitude	Collection Date	*A. gambiae*	%M	%S	%F1	%backcross
Endo	Burkina-Faso	12.5	-1.33333	Nov. 2006	60	42	45	3	10
Goundry	Burkina-Faso	12.5	-1.33333	Nov. 2006	32	3	0	0	97
Foumbot	Cameroon	5.48505	10.60005	Aug. 2006	94	0	99	0	1
Njigalap	Cameroon	5.8999	11.13387	Sep. 2006	119	10	70	12	8
Tiko	Cameroon	4.07860	9.36810	Sep.-Oct. 2003	81	63	27	0	10
Bioko	Equatorial Guinea	3.75000	8.78333	Sep. 2002	73	18	25	0	58
Abu (♀♀)	Guinea-Bissau	11.46144	-15.91411	Nov. 2009	50	16	2	0	82
Abu (♂♂)				Nov. 2009	48	19	0	19	62
Canjufa	Guinea-Bissau	12.43189	-14.12662	Nov. 2009	26	15	73	0	12
Prabis (♀♀)	Guinea-Bissau	11.80066	-15.74332	Nov. 2009	92	25	0	1	74
Prabis (♂♂)				Nov. 2009	90	13	0	15	72
Banambani	Mali	12.80000	-8.05000	July, 2005	87	15	73	0	12
Bantinngoungou	Mali	13.69735	-10.87117	Aug., 2006	87	24	46	5	25
Founia	Mali	12.89163	-9.46063	Oct. 2006	77	39	56	0	5
Kondi	Mali	16.36670	-3.38330	Oct.-Nov. 2002	92	98	0	0	2
Selinkenyi	Mali	11.70000	-8.28333	July, 1991	79	90	6	0	4
				July, 1996	92	74	23	0	3
				Aug., 2002	120	67	22	1	11
				Sep. 2004	107	60	35	0	5
				Aug., 2006	124	36	34	12	18
				Sep. 2010	51	6	39	0	55
				Sep. 2011	78	9	5	0	86
				Aug.-Sep. 2012	83	8	2	0	90
Takouti	Mali	15.46067	-3.49467	Sep. 2005	95	99	0	0	1
Tinko	Mali	13.49600	-10.75520	Aug., 2005	90	1	97	0	2
Damboucoya	Senegal	12.50707	-12.41138	Oct. 2009	95	0	96	0	4
Nathia	Senegal	12.50089	-12.34388	Oct. 2009	41	7	78	0	15

Table S4. Summary of linkage disequilibrium (LD) result. Strong LD ($r_2 > 0.5$) is indicated in gray font.

Site (collection)	Within X island	Within 2L island	Witihin 3L island	Between X and 2L	Between X and 3L	Between 2L and 3L
Populations with recent past introgression and no ongoing gene flow.						
Banambani	1.000	0.984	0.954	0.926	0.739	0.744
Founia	0.941	0.857	0.850	0.864	0.831	0.837
Tiko	0.985	1.000	1.000	0.967	0.820	0.799
Nathia	1.000	0.959	1.000	0.533	0.856	0.580
Canjufa	0.901	1.000	1.000	0.864	0.669	0.781
Populations with recent past introgression and loss of divergence on 2L.						
Bioko	0.986	0.971	1.000	0.233	0.824	0.184
'12 Selinkenyi	1.000	0.818	1.000	0.013	0.828	0.011
Populations with ongoing gene flow and introgression.						
Bantinngoungou	0.849	0.747	0.699	0.784	0.710	0.685
Njigalap	0.817	0.805	0.951	0.816	0.877	0.863
Endo	0.947	0.967	0.854	0.895	0.793	0.797
Populations with complete or nearly complete introgression.						
Goundry	0.948	0.981	0.604	0.001	0.048	0.051
Abu	0.882	0.634	0.898	0.132	0.104	0.067
Prabis	0.963	0.572	1.000	0.107	0.047	0.046
Temporal Variation in Hybridization at Selinkenyi, Mali						
1991	0.979	0.877	1.000	0.933	0.776	0.736
1996	1.000	0.959	1.000	0.964	0.973	0.938
2002	0.895	0.896	0.929	0.869	0.791	0.756
2004	1.000	0.977	0.987	0.957	0.917	0.876
2006	0.805	0.780	0.865	0.761	0.819	0.800
2010	0.978	0.927	1.000	0.236	0.879	0.207
2011	0.895	0.761	1.000	0.023	0.936	0.022
2012	1.000	0.818	1.000	0.013	0.828	0.011

References.

1. Stump AD*, et al.* (2005) Centromere-proximal differentiation and speciation in Anopheles gambiae. *Proc Natl Acad Sci U S A* 102(44):15930-15935.
2. Turner TL & Hahn MW (2007) Locus- and population-specific selection and differentiation between incipient species of Anopheles gambiae. *Mol Biol Evol* 24(9):2132-2138.
3. Turner TL, Hahn MW, & Nuzhdin SV (2005) Genomic islands of speciation in Anopheles gambiae. *PLoS Biol* 3(9):e285.
4. White BJ, Cheng C, Simard F, Costantini C, & Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of Anopheles gambiae. *Mol Ecol* 19(5):925-939.
5. Jurinke C, van den Boom D, Cantor CR, & Koster H (2002) Automated genotyping using the DNA MassArray technology. *Methods Mol Biol* 187:179-192.
6. Fu JF*, et al.* (2008) MassARRAY assay: a more accurate method for JAK2V617F mutation detection in Chinese patients with myeloproliferative disorders. *Leukemia* 22(3):660-663.
7. Wright WT*, et al.* (2008) Multiplex MassARRAY spectrometry (iPLEX) produces a fast and economical test for 56 familial hypercholesterolaemia-causing mutations. *Clin Genet* 74(5):463-468.
8. Zhang SJ, Qiu HX, Li JY, Shi JY, & Xu W (2010) The analysis of JAK2 and MPL mutations and JAK2 single nucleotide polymorphisms in MPN patients by MassARRAY assay. *Int J Lab Hematol* 32(4):381-386.