Allele-Specific Population Structure of *Drosophila melanogaster* Alcohol Dehydrogenase at the Molecular Level

Michel Veuille,* Véronique Bénassi,† Sylvie Aulard† and Frantz Depaulis*

**Laboratoire d'Ecologie, Ecole Pratique des Hautes Etudes, Unité de Recherche Associée 258, Université Paris-6, 75005 Paris, France and †Laboratoire Populations Génétique Evolution, CNRS, 91190 Gif-sur-Yvette, France*

Manuscript received April 5, 1997
Accepted for publication February 23, 1998

ABSTRACT

The history of the *Drosophila melanogaster* alcohol dehydrogenase (ADH) Fast/Slow polymorphism was studied by recording molecular variation and inversion polymorphism in 233 chromosomes from European and African populations. Silent molecular variation in the Slow allele was very different between *standard* chromosomes and chromosomes bearing the *In(2L)t* inversion. Within populations, inverted *Slow* haplotypes were more variable than standard *Slow* haplotypes. Between populations, geographical structure was almost nonexistent for inverted *Slow* haplotypes and highly significant for standard *Slow*. All *Fast* haplotypes occurred on standard chromosomes. They showed little variation within and between populations. They were highly significantly closer to *standard Slow* haplotypes from Europe. These results suggest that the current range of *Fast* and *In(2L)t Slow* haplotypes is recent and that an older genetic differentiation between populations was followed by allele-specific gene flow.

An important part of population genetics theory describes the differentiation of populations for selectively neutral alleles under the counteracting effects of genetic drift and migration (Wright 1943, 1951). Our understanding of population structuring for locally advantageous alleles, although critical for theories of evolution, is far less developed. It relies on empirical evidence that differentiation is extreme for some loci in some populations. A methodological assumption is that the neutral model is a null hypothesis and that natural selection can result in locally increased differentiation. An alternative is that newly evolved genetic systems, when successful, spread to the other populations and drive them toward a new adaptive peak. We thus do not know whether species are more, less or equally geographically structured for recently selected alleles than they are for more ancient neutral variation.

To answer this question, we studied natural populations of *Drosophila melanogaster*. According to Tsacas and Lachaise (1974), this species originated in Africa, where it lives on decaying fruit. In neolithic times, it became adapted to human environments (Lachaise *et al.* 1988) and colonized Europe. It is thought to have spread through the rest of the inhabited world in historical times (David and Capy 1988).

This history is likely to have caused recent genetical differentiation between populations. There is evidence that it was accompanied by increased resistance to ethanol (reviewed by David 1988). The alcohol dehydrogenase enzyme presents two electromorphs that differ in their biochemical properties. Fast may be broadly described as being more active, whereas Slow is more thermostable (reviewed by Chambers 1988). Allozyme frequencies have been recorded in America (Vigue and Johnson 1973; Johnson and Schaeffer 1976; Pipkin *et al.* 1976), Australia (Oakeschott *et al.* 1982), Japan (Watada *et al.* 1986), China (Jiang *et al.* 1989), and the Old World (David *et al.* 1989). They systematically show clines in which the Fast allele is predominant at higher latitudes than Slow.

The sequencing of 10 alleles has shown that *Fast* haplotypes are less variable than *Slow* haplotypes (Kreitman 1983). This was confirmed by restriction site polymorphism studies (Aquadro *et al.* 1986; Kreitman and Aguade 1986a,b; Aguade 1988; Simmons *et al.* 1989; Benassi *et al.* 1993; Berry and Kreitman 1993; Benassi and Veuille 1995), thus suggesting that the Fast allele is recent. Hudson *et al.* (1987) found excess polymorphism at the *Adh* locus. Kreitman and Hudson (1991) found excess polymorphism in a world sample of Fast and Slow alleles. This was consistent with a model of balancing selection on the amino acid polymorphism or a site closely linked to it. A similar but somewhat smaller peak was also observed for within Slow comparisons, indicating a more complex history for the *Slow* haplotypes. The Fast allele lacked polymorphism, indicating a recent origin for Fast/Slow or a recent selective sweep within *Fast* haplotypes. The next step was to compare geographical structuring for Fast/Slow and for silent variation among *Slow* haplotypes. Among Old World populations, most silent variation opposes West African to East African populations, whereas Fast/Slow

Corresponding author: Michel Veuille, Laboratoire d'Ecologie, C 237, Université Paris-6, 7 quai Saint Bernard, 75252 Paris cedex 05, France. E-mail: mveuille@snv.jussieu.fr

opposes European to African populations (Benassi and Veuille 1995).

The recent appearance of the Fast allele is suggested by its linkage disequilibrium with the nearby *In(2L)t* inversion. The proximal breakpoint of this chromosome arrangement is located one chromosome division (~1700 kb) away from the *Adh* locus. Despite the high frequency of this inversion in natural populations of Drosophila (Lemeunier and Aulard 1992), the alcohol dehydrogenase (ADH) Fast allele has almost always been reported to occur in the *standard* chromosomal arrangement (reviewed by van Delden and Kamping 1989). Knowledge of the recombination rate between the inversion and the gene, coupled with the high frequency of the *In(2L)t* inversion in West Africa, suggested that the Fast allele could not be older than a few hundred years, at least in this part of the species range (Benassi *et al.* 1993). This observation could be interpreted in two ways. According to one hypothesis, the Fast allele recently appeared in the *standard* chromosome arrangement and dramatically increased in frequency, its absence in *In(2L)t* chromosomes being due to history. According to a second hypothesis, the *standard-Fast* arrangement is adaptive and is maintained by natural selection. The purpose of this study is to find evidence relating to this question by determining the karyotype of *Adh* haplotypes previously reported from Europe and East Africa (Benassi and Veuille 1995), and comparing them to available data for West Africa (Benassi *et al.* 1993).

MATERIALS AND METHODS

Population sampling, restriction site polymorphism data, and cytogenetical techniques have been described previously (Benassi *et al.* 1993; Benassi and Veuille 1995). All population samples originate from natural populations from France, Ivory Coast, and Malawi. The distances between these populations are 4200 km (France-Ivory Coast), 4600 km (Ivory Coast-Malawi), and 7000 km (Malawi-France). African samples were caught in ecological stations and nature reserves. Isochromosomal lines were directly obtained either from wild-caught individuals or, for some West African samples, from recently established isofemale lines.

Chromosome and molecular data were obtained for 65 Malawi lines, 84 Ivory Coast lines, and 84 French lines. The Ivory Coast and Malawi samples were random samples. The French sample included a random sample from Cognac (of which 8 Slow lines were available for the cytogenetic study) and 31 additional Slow lines from Bordeaux to compensate for the predominance of the Fast allele in this area (see Benassi and Veuille 1995). This did not bias statistical analysis because the different allelic classes were examined separately below. At the other end of the ADH cline, only two Fast alleles were available for Malawi.

The numbering of indels and restriction sites follows their order along the sequence, as in Benassi and Veuille (1995). From 5′ to 3′: indel ID1, *Msp*I -346, *Dde*I 9, *Ban*I 102, indel ID2, *Hae*III 287, *Dde*I 321, *Alu*I 562, indels Δ1, Δ2, and ID3, *Cfo*I 573, *Msp*I 583, *Hae*III 688, *Hae*III 816, *Msp*I 818, *Cfo*I 864, indel ID5, *Sau*3AI 1354, *Hae*III 1425, *Sau*96I 1425, *Dde*I 1518, *Dde*I 1527, *Alu*I 1596, indel ID6, *Dde*I 1875, *Hae*III 1925.

This study is concerned with genetic distance between nine "groups" (three haplotype classes in three populations). The three haplotype classes will be defined in the results section.

Genetic distance was calculated after Nei (1987) as the average number of differences between pairs of chromosomes. Let D_x and D_y be the average genetic distance within groups *X* and *Y*, respectively (hereafter "molecular diversity"), and let D_{xy} be the average genetic distance between individuals from the *X* and *Y* groups; "molecular divergence" was calculated according to the expression $D_a = D_{xy} - (D_x + D_y)/2$. We used Nei's and Miller's (1990) RESTSITE program that takes into account probability differences in restriction sites and includes a jackknifing procedure over restriction enzymes to estimate standard errors.

A phylogenetic analysis was carried out on the 36 genetic distances calculated between the nine groups using the neighbor-joining and unweighted pair-group method with arithmetic mean (UPGMA) procedures from version 1.01 of Molecular Evolutionary Genetics Analysis (MEGA; Kumar *et al.* 1993). These two different methods were used in order to check that our distance data exhibited a reliable tree-like structure. Because haplotypes exchange polymorphisms through recombination and because populations exchange haplotypes through migration, internal nodes have little biological meaning, except the most distal ones, which represent the average value of each group. The phylogenetic analysis and the distance matrix were therefore used as two different ways to analyze the relationships between the nine groups. Their significance was assessed in the same way, by comparing distances in order to state if a given group was significantly closer to a second one than to a third one. Comparing measures of molecular divergence cannot be achieved by variance tests like a *t*-test, because divergence values are not normally distributed and because distances are not independent. We used a resampling procedure. We first established a table of divergence values (D_a) between pairs of haplotypes. We then estimated the average distance between groups of 100 haplotypes by resampling haplotypes from pooled samples. One-tailed probability level was estimated over 1000 repetitions as the percentage of departure from the observed value. Samples of 100 were used to equalize the power of the test over observations. The sample size was smaller for several comparisons involving *Fast* haplotypes from Malawi, for which only two Fast alleles were available. The number of comparisons was set at 81 for comparing inversions from Malawi and France, in each of which 9 haplotypes had been found. The power of the test is the same for all other comparisons. There is much redundancy in all the calculated pairwise comparisons between distances. These data are not shown. Data relating to the questions addressed in this study will be presented below.

RESULTS

Inversion polymorphism: Primary data are given in the appendix. *Slow* chromosomes were found to carry *In(2L)t* inversions in the following proportions: 52 out of 69 for Ivory Coast, 9 out of 63 for Malawi, and 9 out of 39 for France. In the later sample, all inversions were found in the Bordeaux subsample. We also found 2 *In(2R)NS* (52AB;56F) and 2 *In(2R)* (44F;51A; 54A; 55F) inversions in Malawi, and 1 *In(2L)* (26A;30F) inversion in France, the latter 2 being new endemic inversions. This will not be dealt with in any detail, inversions from

TABLE 1

Frequency of restriction site polymorphisms in each haplotype group (three haplotype classes in three populations)

	Fast			*Standard-Slow*			*In(2L)t-Slow*		
N	Mal	I.C.	Fra	Mal	I.C.	Fra	Mal	I.C.	Fra
	2	15	45	54	17	31	9	52	9
1	100	100	100	92	94	87	78	81	56
2	100	100	100	100	100	100	100	98	100
3	0	0	0	55	88	37	56	62	56
4	100	100	100	100	88	100	100	100	89
5	0	0	0	0	0	0	0	2	0
6	0	0	0	2	0	0	0	0	0
7	100	100	98	100	100	67	100	100	100
8	100	100	100	98	100	100	89	100	100
9	100	100	98	100	100	67	100	100	100
10	0	24	67	37	12	33	11	2	0
11	100	100	100	100	100	100	100	92	100
12	100	100	100	98	100	100	100	100	100
13	100	100	88	45	82	100	89	79	78
14	50	100	98	100	82	100	89	75	78
15	100	100	97	100	100	100	89	100	78
16	100	96	98	16	18	60	44	58	56
17	0	4	2	84	82	40	56	42	44
18	100	96	76	43	29	70	33	52	88
19	0	0	2	0	0	3	0	0	0
20	100	100	98	24	76	97	33	83	100

The numbering of restriction sites is: 1, *Msp*I-346; 2, *Dde*I 9; 3, *Ban*I 102; 4, *Hae*III 287; 5, *Dde*I 321; 6, AluI 562; 7, *Cfo*I 573; 8, *Msp*I 583; 9, *Hae*III 688; 10, *Hae*III 816; 11, *Msp*I 818; 12, *Cfo*I 864; 13, *Sau*3AI 1354; 14, *Hae*III 1425; 15, *Sau*96I 1425; 16, *Dde*I 1518; 17, *Dde*I 1527; 18, *Alu*I 1596; 19, *Dde*I 1875; 20, *Hae*III 1925. Mal, Malawi; I.C., Ivory Coast; Fra, France.

the right arm of chromosome *2* being beyond the scope of this study, whereas the *In(2L)* endemic inversion was pooled with the standard arrangement, to which it is evolutionarily closer than to *In(2L)t*. All *Fast* lines were *standard* for the *In(2L)t* inversion. We will thus only consider the three following classes of haplotypes:

"*Fast*" refers to *Fast* haplotypes in *standard* chromosomes for the *In(2L)t* polymorphism.
"*st-S*" refers to *Slow* haplotypes in *standard* chromosomes for the *In(2L)t* polymorphism.
"*in-S*" refers to *Slow* haplotypes in *In(2L)t* chromosomes.

These three haplotype classes are present in each population, making up the nine groups that will be compared below.

Distribution of molecular polymorphism within groups: Table 1 shows the frequency of 20 restriction sites in the nine groups. Eleven of them were polymorphic among the 62 *Fast* chromosomes, 16 from the 101 *st-S* chromosomes, and 15 from the 70 *in-S* chromosomes. The number of polymorphic sites was not very different in the three classes. Six of them were unique to one or the other *Slow* classes. They were at a low frequency ranging from 2 to 8%. This value was never significantly different from the 0% value observed in the other groups. Therefore, there is no evidence that there are diagnostic sites between alleles. None of the 11 polymorphisms found in *Fast* were unique to this allele. All of them were observed in one or both of the *Slow* alleles. This suggests that they originated from *Slow* haplotypes through recombination or gene conversion.

Linkage disequilibrium between inversions and molecular polymorphisms: Linkage disequilibrium between polymorphic sites and the inversion was estimated using Fisher's exact test. We included indels in this survey, because two indels from the first intron of *Adh* have been shown to affect expression levels (Laurie *et al.* 1991; Matthews *et al.* 1992). Linkage disequilibrium involved four significant cases in Ivory Coast, two in Malawi, and none in France, because *In(2L)t* was absent from the French random sample (Cognac subsample). The two significant cases for Malawi included sites *Hae*III 816 ($P = 0.026$) and *Sau*3AI 1354 ($P = 0.002$). In Ivory Coast, the linkage disequilibrium was highly significant ($P < 10^{-8}$) for the two sites that were characteristic of *Fast* haplotypes: the F/S amino acid polymorphism itself and the Δ1 indel (Kreitman 1983). The other two significant linkage disequilibria involved *Msp*I -346 and *Hae*III 816 sites.

Genetical distances between haplotype classes: Genetical distances between the nine groups are shown in Table 2. The diagonal (boldface numbers) indicates

TABLE 2

Molecular diversity, molecular distance, and molecular divergence[a] between *Adh*, *Fast*, and *Slow* haplotypes in standard and *In(2L)t* chromosomes

		In(2L)t Slow			*Standard Slow*			*Standard Fast*		
D_{xy}	D_a	Malawi	Ivory Coast	France	Malawi	Ivory Coast	France	Malawi	Ivory Coast	France
In(2L)t	Malawi	**6.55**	0.03	0.61	0.39	0.08	0.98	2.62	2.38	2.75
Slow	$N = 9$	**2.28**	0.09	0.67	0.30	0.11	0.63	1.23	1.04	1.31
	Ivory Coast	6.27	**5.91**	0.14	1.46	0.63	0.76	1.42	1.60	2.10
	$N = 52$	2.14	**1.98**	0.17	0.90	0.50	0.46	0.94	0.72	1.02
	France	7.19	6.41	**6.63**	2.16	1.21	0.70	1.09	1.31	1.96
	$N = 9$	2.30	1.98	**2.29**	1.14	0.61	0.47	0.85	0.64	1.00
Standard	Malawi	6.37	7.12	8.17	**5.40**	0.83	2.28	4.79	4.02	4.11
Slow	$N = 54$	1.97	2.48	2.69	**1.61**	0.54	1.03	2.35	1.94	2.31
	Ivory Coast	5.69	5.93	6.85	5.87	**4.66**	1.74	4.19	3.92	4.38
	$N = 17$	2.11	2.16	2.12	2.12	**1.86**	0.72	2.64	2.27	2.55
	France	7.24	6.71	7.00	7.96	7.02	**5.97**	1.36	0.89	1.13
	$N = 30$	2.59	2.25	2.30	2.70	2.59	**2.28**	0.60	0.40	0.56
Standard	Malawi	6.71	5.20	5.22	8.30	7.33	5.17	**1.63**	0.09	0.69
Fast	$N = 2$	2.59	1.85	2.00	3.45	3.23	2.27	**1.65**	0.03	0.63
	Ivory Coast	6.32	5.22	5.29	7.38	6.91	4.54	1.57	**1.32**	0.22
	$N = 15$	2.38	1.86	1.96	2.92	3.03	1.81	1.24	**0.75**	0.20
	France	6.95	5.99	6.21	7.73	7.64	5.05	2.44	1.82	**1.82**
	$N = 45$	2.63	2.13	2.24	3.11	3.28	1.85	1.84	0.92	**0.84**

[a] Diagonal (bold), molecular diversity; lower sector, molecular distance; upper sector, molecular divergence; the lower number in each cell is the standard error.

molecular diversity within each group. The remaining cells are divided into two halves. The lower left half shows genetic distance D_{xy} between groups. The upper right half indicates molecular divergence D_a.

Intragroup variation (diagonal) is very homogeneous within haplotype classes. *In-S* chromosomes differ by an average of 6.36 restriction sites (SD = 0.39), *st-S* chromosomes differ by 5.34 sites (SD = 0.65), and *Fast* chromosomes differ by 1.60 sites (SD = 0.27). Over all populations, *in-S* chromosomes are 1.20 times more variable than *st-S* (range: 1.11–1.26), and *Fast* chromosomes are 0.30 times as variable as *st-S* (range: 0.28–0.31).

The difference between *Slow* haplotypes from each chromosome arrangement is slight. Its significance was assessed by bootstrap. The n_1 *st-S* distances and the n_2 *in-S* distances were pooled, and two samples of the same size were randomly drawn. Over 1000 replicates, the 11% difference observed in France was nonsignificant ($P = 0.134$). The 21 and 26% differences observed in Malawi and in Ivory Coast were highly significant ($P = 0.003$ and $P < 0.001$, respectively).

The upper right sector of each of the squares crossing the diagonal of Table 2 shows the molecular divergence between chromosomes from the same haplotype class but from different populations. The three haplotype classes showed similar patterns of differentiation between populations. Malawi and France were always the most different populations. Ivory Coast was consistently, albeit never significantly, closer to Malawi than to France. An unexpected observation is the difference of orders of magnitude. Although divergence values between the three comparisons of populations were weak and very similar in *in-S* (0.03, 0.14, and 0.61) and in *Fast* (0.09, 0.22, and 0.69), the divergence in *st-S* was much larger (0.83, 1.74, and 2.28). Most of these estimates were not significantly different. However, Malawi *st-S* was significantly more different from France *st-S* than from Ivory Coast *st-S* ($P < 0.02$).

The divergence between *st-S* and *Fast* is shown on the middle right square of Table 2. The *Fast* haplotypes from all three populations were always highly significantly ($P < 10^{-3}$) closer to *st-S* from France (mean = 1.12, SD = 0.23; range: 0.89–1.36) than to *st-S* from Ivory Coast and Malawi (mean = 4.23, SD = 0.31; range: 3.92–4.79).

The divergence between *Fast* and *in-S* is shown in the upper right square of Table 2. It shows a less marked divergence, which is consistent with the fact that *in-S* haplotypes are geographically less divergent than *st-S* haplotypes. The *Fast* haplotypes from all three populations were significantly ($P < 0.005$) closer to *in-S* from France than to *in-S* from Malawi. Ivory Coast *in-S* were intermediate. Only *Fast* haplotypes from France were significantly more different from *in-S* from Ivory Coast than from French *in-S* ($P < 0.05$).

The divergence between *in-S* and *st-S* from the different populations is shown in the middle upper square

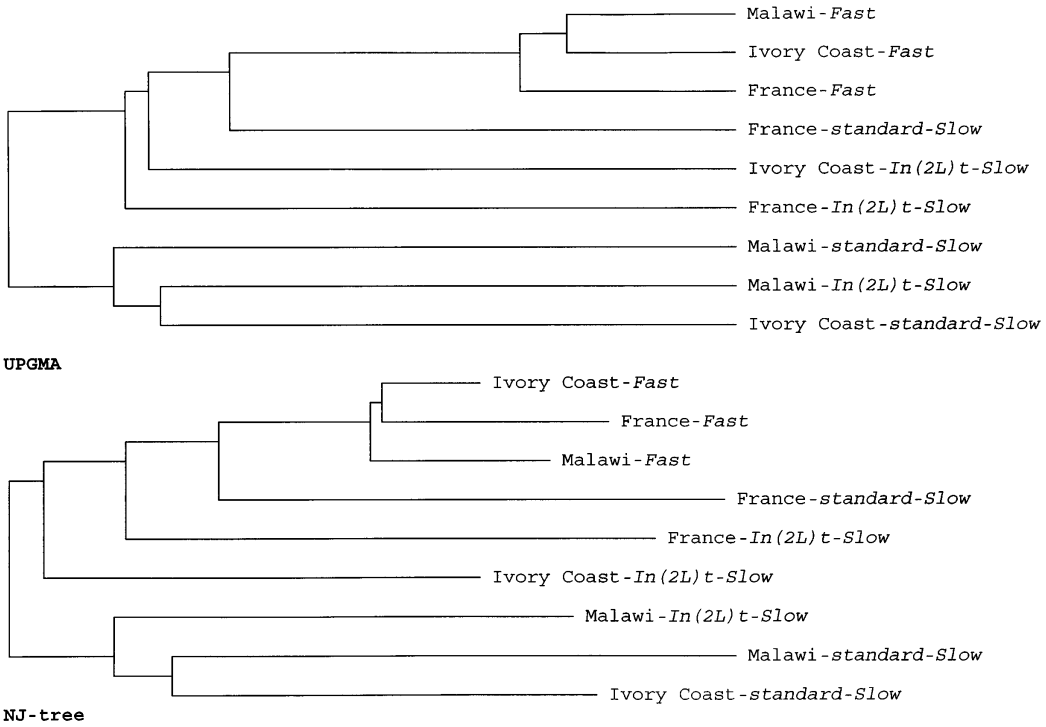

Figure 1.—UPGMA (top) and Neighbor-Joining (bottom) trees of nine *Adh* groups (three haplotype classes in three populations), based on the genetic distance (D_{xy}) matrix of Table 2.

of Table 2. The *st-S* haplotypes from France were not very different from *in-S* haplotypes from either population. The contrary was observed for *st-S* haplotypes from Malawi. Their divergence with Malawi *in-S* was significantly smaller than with Ivory Coast *in-S* ($P < 0.05$) and highly significantly smaller than with French *in-S* ($P < 0.005$). Ivory Coast *st-S* were intermediate because a significant result ($P < 0.05$) was observed between distances with European and Malawi *in-S*.

Phylogenetic analysis: The phylogenetic analysis is shown on Figure 1. Results from the UPGMA and the neighbor-joining trees are virtually identical. The three terminal branches corresponding to the Fast alleles from the three populations are the shortest. They join each other distally in the phylogeny. This reflects the low haplotype variability among Fast alleles because external branches contain the average genetic distance in each population. The two nodes joining *Fast* haplotypes are very close to each other. This reflects the close relationship of these alleles between populations. The three *in-S* groups corresponding to the three populations are at the opposite of *Fast*. They branch very deeply (proximally) in the tree. They stem from three neighboring nodes. These two aspects are complementary. They indicate the substantial variability of these haplotypes and their relatively low differentiation between populations. The pattern of distribution of the three *st-S* groups is very different from the former groups. The sample from France branches off on one side of the tree between *Fast* haplotypes and *in-S* haplotypes from France and Ivory Coast. The other two *st-S* samples, from West and East Africa, are at the other end of the tree. This is in agreement with the highly significant geographical differentiation observed among *st-S* haplotypes. This confirms that *Fast* haplotypes from any population are closer to European *st-S* than to African *st-S*. This also accounts for the fact that *in-S* haplotypes are intermediate between *st-S* haplotypes for their distance with *Fast* haplotypes.

The phylogenetic analysis provides a consistent interpretation of the distance matrix from which it is derived. For instance, the most rightward column of Table 2, which gives the distance of France *Fast* with the other groups, predicts exactly the order in which these groups connect to the phylogenetic line leading from France *Fast* to the most extreme group (Malawi and Ivory Coast *st-S*) on the NJ-tree. It should be noted that the trees of Figure 1 do not represent lines of descent between sequences, but rather distances between groups of haplotypes arranged in consideration of their chromosomal genotype, allozyme genotype, and population of origin. Each of these groups is polymorphic. These trees have no root. Previous studies have shown that the relationships between *Adh* haplotypes make up a multiconnected network as a result of intragenic recombination (see *e.g.*, Simmons *et al.* 1989; Benassi *et al.* 1993). They do not display a tree-like hierarchical structure. The coalescence of branches at the basis of the trees of Figure 1 represent the average number of differences in pairwise comparisons between the most extreme distal ends. Therefore, *Slow* haplotypes, which carry much recombined polymorphic sites, especially for *in-S*, con-

verge closer to the basis. The linear series connecting European *st-S* and *Fast* haplotypes on one branch is probably the only well-defined line of descent. Little recombination has altered the relationships between these groups.

A noteworthy feature of these trees is the closeness between Ivory Coast *in-S* and both French *in-S* and *st-S*, whereas Ivory Coast *st-S* is further away in the topology. This accounts for the small molecular divergence observed between French and West African *Slow* haplotypes by Benassi and Veuille (1995). This also confirms that geographical distances in Slow alleles are fundamentally heterogeneous at the chromosome level.

The phylogenetic analysis thus confirms and synthesizes what was already apparent from the distance analysis.

DISCUSSION

At least three haplotype classes can be distinguished in the *Adh* system: In a previous study (Benassi and Veuille 1995), the same sample was used for a different purpose. Fixation indices for the Fast/Slow allozymes were compared with fixation indices for *Slow* molecular variation. Genetic structuring was different in the two cases: the Fast/Slow distribution opposed Europe to Africa, and the *Slow* molecular variation opposed East Africa to West Africa and Europe. The purpose of this study was to demonstrate the action of selection on the amino acid polymorphism. A similar method was followed by Berry and Kreitman (1993) to study the ADH cline in North America.

This study assumed that silent variation in *Slow* was neutral and thus represented historical patterns, whereas changes in Fast/Slow frequencies represented selection. The present analysis shows that the silent variation in *Slow* is heterogeneous. Molecular variation is differently distributed in inverted and noninverted *Slow* chromosomes. Furthermore, the *In(2L)t* inversion frequency changes between populations: from 0.73 in Ivory Coast (SD = 0.05) to 0.23 in Malawi (SD = 0.06) and 0.14 in France (SD = 0.04). The inversion frequency changes as dramatically between populations as does the Fast allele. This variation is likely to have biased former estimates of genetical distance between populations (Benassi and Veuille 1995).

Genetic structuring of *Adh* between Old World populations: Three haplotype classes (*Fast*, *in-S*, and *st-S*) can be distinguished instead of two alleles (Fast and Slow). They induce discontinuities in the distribution of silent variation. Silent molecular variation in each of them can be provisionally considered neutral. This is substantiated by the close similarity of relative divergence between Malawi, Ivory Coast, and France in each class. Malawi and Europe are always the most distant populations. Assuming that the best estimate of the divergence between populations is provided by the average for the three classes, and setting the Malawi-France divergence at 1.00, we obtain 0.27 between Ivory Coast and Malawi and 0.59 between France and Ivory Coast. It should be noted that the distances between Ivory Coast and the other two populations are not significantly different among themselves. Moreover, the imperfection of this estimation is apparent in its violation of triangular inequality (*i.e.*, $AC \leq AB + BC$), which is generally considered a requirement of mathematical distances.

Variation patterns differ in the three haplotype classes: Levels of polymorphism within each haplotype class are remarkably similar between populations. *Inverted-Slow* haplotypes are 1.20 times more polymorphic than *st-S*, and *Fast* haplotypes are 0.30 times less polymorphic than *st-S*. Levels of divergence between populations, however, are different. There is more geographical structuring in *st-S* than in *in-S* or *Fast*. The three groups thus present three different patterns of polymorphism: *st-S* shows substantial polymorphic and substantial geographic differentiation; *in-S* is very polymorphic but shows little geographic differentiation; and *Fast* shows little polymorphism and little geographic differentiation.

Heterogeneity among *Slow* haplotypes: An unexpected result of this study is the heterogeneity among *Slow* haplotypes. The global amount of variation present in the two chromosome arrangements is similar. For instance, between the most extreme populations surveyed in this study (Europe and Malawi) the average number of restriction site differences between two *st-S* haplotypes (7.96) is close to that observed between two *in-S* haplotypes (7.19), or between an *st-S* haplotype from France and an *in-S* haplotype from Malawi (7.24). However, the geographical apportionment of this variation is different. The whole variation is shared by *in-S* chromosomes from all populations, which are therefore very similar among themselves. In *st-S*, each population is less variable than the pooled sample, whereas more variation segregates geographically. This differentiation is mostly between France and African populations. The fact that the two *Slow* classes share the same polymorphisms and have close heterozygosity levels indicates that they have been exchanging variation through recombination for a long time. Some event, more recent than the setting of this inversion polymorphism, occurred and produced the present pattern. We can imagine two possibilities. First, natural selection, acting on the *Adh* locus itself, could have affected *Slow* haplotypes on one chromosome arrangement, the other arrangement remaining unaffected because inversions inhibit crossovers. This hypothesis is not likely because polymorphism levels indicate no selective sweep event. Second, a selective sweep event could have occurred elsewhere on the chromosome. For instance, an advantageous mutant linked to *In(2L)t* could have appeared at a remote locus on the chromosome. It could have spread among world populations and have erased any

geographical structure without sweeping *Adh* variation away. If so, this phenomenon could be recorded in other parts of the second chromosome. This hypothesis is currently being tested.

Recent origin of *Fast* haplotypes: The low variability of *Fast*, its low differentiation between populations, and its absence from *In(2L)t* karyotypes together suggest the recent origin of *Fast* haplotypes. This may be because the *Fast* mutation appeared recently, or because it was subject to a recent selective sweep. The karyotype of this allele clearly indicates that it originated from a *standard* chromosome. It is genetically closer to *st-S* from Europe than from other populations. These facts are in favor of a European origin of *Fast*. Its broad distribution in Africa suggests that it subsequently spread throughout the Old World.

Origin of the *In(2L)t-Fast* linkage disequilibrium: A long-recognized property of the ADH system is the absence of *Fast* in *In(2L)t* chromosomes (reviewed by van Delden and Kamping 1989). Given known levels of recombination between the *Adh* locus and the proximal breakpoint of the *In(2L)t* inversion (Malpica *et al.* 1987), and assuming no epistatic selection between ADH and the inversion, the coexistence of the Fast allele and the inversion in Ivory Coast cannot be older than a few hundred years (Benassi *et al.* 1993). The genotype of *Fast/Slow vs. In(2L)t/standard* arrangements is available for several samples of *D. melanogaster* chromosomes from Raleigh, NC (Mukai *et al.* 1971), Eugene, OR (Voelker *et al.* 1977), and Carpenter, NC (Langley *et al.* 1977). In addition to the random samples from Ivory Coast and Malawi, these published studies make up 1002 chromosomes, including 328 Fast alleles (32.7%) and 150 inversions (14.9%). Only two *In(2L)t-Fast* chromosomes have been reported in this list (Langley *et al.* 1977), although 49 would be expected at random. According to Ashburner *et al.* (1979; cited in Aquadro *et al.* 1986), the recombination rate within and around the *Adh* locus is 10^{-5} events/kilobase/generation. According to Malpica *et al.* (1987), the recombination rate between the *Adh* locus and the *In(2L)t* proximal breakpoint in *In(2L)t/standard* heterozygotes is 5×10^{-4} (range: 1.9×10^{-3}, 6×10^{-5}). The recombination rate between *In(2L)t* and *standard* is thus an order of magnitude larger than the recombination rate within the gene in a *standard* homozygote. This makes sense because the distance between the gene and the breakpoint is $\sim$1000$\times$ larger than the gene. Given the substantial heterozygosity of this inversion, more recombination is expected between the F/S amino acid polymorphism and the inversion than between F/S and *Adh* silent polymorphisms. The contrary is observed. *Fast* haplotypes are polymorphic for eight restriction sites that are already present in *Slow*. These *Fast* haplotypes certainly originated by recombination, because they are polymorphic for substitutions that are also polymorphic in *Slow*. There are two possible explanations for this. One is that natural selection favors the *Fast-Standard* configuration. The other is historical and is suggested by the results of this study. According to this hypothesis, *Fast* evolved in Europe from a *standard* chromosome and remained isolated for some time. Then *Fast* and *In(2L)t* haplotypes spread and came into contact. The absence of recombinants would not indicate a recent origin of the haplotypes, but rather a recent gene flow between populations.

An hypothesis on the history of the *Adh* genetic system: A way to account for features of the *Adh* polymorphism is to assume that the history of the species combines geographic differentiation and gene flow. In a first stage, differentiation would have occurred between *standard Slow* from Europe and from Africa. The *In(2L)t* inversion would have been restricted to Africa, in agreement with Lemeunier's and Aulard's (1992) observation that *D. melanogaster* inversions mostly occur in tropical areas. *Fast* would have originated from a *standard Slow* chromosome in Europe, where it became predominant. In a second stage, world populations would have come into contact, thus favoring the spreading of *Fast* to Africa, and possibly that of *In(2L)t* to Europe.

According to Lachaise *et al.* (1988) and David and Capy (1988), *D. melanogaster* was long confined to Africa and Europe and only recently expanded its range to the rest of the world. This is in agreement with mitochondrial DNA variation, which shows that the founders of American and Far Eastern populations made up a few closely related lineages (Hale and Singh 1991). According to this classic scenario, gene flow occurred from the center of origin toward the periphery. The variation patterns being currently observed in the Old World would result from long-established equilibrium frequencies. An alternative scenario is that Fast appeared in Europe, increased in frequency, and secondarily spread to Africa. Its allele frequency probably rose from 0 to 17.6% as observed in the Lamto ecological reserve, showing that native fruitfly populations from remote field stations are not isolated. Driven by the ecological success of *D. melanogaster* in human-modified environments, the Fast allele may be invading old established African populations, possibly displacing them toward new adaptative peaks.

We thank Matthew Cobb for helpful criticism and careful reading of the manuscript, and Suzanne Mazeau and Olivier Langella for technical assistance. The article greatly benefited from the referees' comments on a previous version of the manuscript. The phylogenetic analysis was suggested by Andy Clark.

LITERATURE CITED

Aguade, M., 1988 Restriction map variation at the *Adh* locus of *D. melanogaster* in inverted and non inverted chromosomes. Genetics **119:** 135–140.

Aquadro, C. F., S. F. Desse, M. A. Bland, C. H. Langley and C. C. Laurie-Ahlberg, 1986 Molecular population genetics of the

Alcohol dehydrogenase gene region of *Drosophila melanogaster*. Genetics **114:** 1165–1190.

Ashburner, M., R. Camfield, B. Clarke, D. Thatcher and R. Woodruff, 1979 A genetic analysis of the locus coding for alcohol dehydrogenase, and its adjacent chromosome region, in Drosophila melanogaster, pp. 95–106 in *Eucaryotic Gene Regulation*, edited by R. Axel, T. Maniatis and C. F. Fox. Academic Press, New York.

Benassi, V., and M. Veuille, 1995 Comparative population structuring of molecular and allozyme variation of *Drosophila melanogaster Adh* between Europe, West Africa and East Africa. Genet. Res. **65:** 95–103.

Benassi, V., S. Aulard, S. Mazeau and M. Veuille, 1993 Molecular variation of *Adh* and *P6* genes in African populations of *Drosophila melanogaster* and its relation to chromosomal inversions. Genetics **134:** 789–799.

Berry, A., and M. Kreitman, 1993 Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the East coast of North America. Genetics **134:** 869–893.

Chambers, G. K., 1988 The *Drosophila Alcohol dehydrogenase* gene-enzyme system. *Advances in Genetics* **25:** 39–107.

David, J. R., 1988 Ethanol adaptation and alcohol dehydrogenase polymorphism in *Drosophila*: from phenotypic functions to genetic structures, pp. 163–172 in *Population Genetics and Evolution*, edited by G. de Jong. Springer-Verlag, Berlin.

David, J. R., and P. Capy, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

David, J. R, F. Alonso-Poraga Borai, P. Capy, H. Merçot, S. F. McEvey *et al.*, 1989 Latitudinal variation of *Adh* gene frequencies. *Drosophila melanogaster*: a Mediterranean instability. Heredity **62:** 11–16.

Hale, L. R., and R. S. Singh, 1991 A comprehensive study of genic variations in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history *vs.* selection in the genetic structure of geographic populations. Genetics **129:** 103–117.

Hudson, R. R., M. Kreitman and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Jiang, C., J. B. Gibson and H. Chen, 1989 Genetic differentiation in populations of *Drosophila melanogaster* from the People's Republic of China: comparison with patterns on other continents. Heredity **62:** 193–198.

Johnson, F. M., and H. E. Schaffer, 1973 Isozyme variability in species of the genus *Drosophila*. VII. Genotype-environment relationship in populations of *D. melanogaster* from the eastern United States. Biochem. Genet. **10:** 149–163.

Kreitman, M., 1983 Nucleotide polymorphism at the *alcohol dehydrogenase* gene region of *Drosophila melanogaster*. Nature **304:** 412–417.

Kreitman, M., and M. Aguade, 1986a Genetic uniformity of two populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide restriction enzyme digests. Proc. Natl. Acad. Sci. USA **83:** 3562–3566.

Kreitman, M., and M. Aguade, 1986b Excess polymorphim at the *Adh* locus in *Drosophila melanogaster*. Genetics **114:** 93–110.

Kreitman, M., and R. Hudson, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565-582.

Kumar, S., K. Tamura and M. Nei, 1993 MEGA: molecular evolutionary genetic analysis. Pennsylvania State University, University Village.

Lachaise, D., M.-L. Cariou, J. R. David, F. Lemeunier and L. Tsacas, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22:** 159–225.

Langley, C. H., K. Ito and R. A. Voelker, 1977 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Seasonal variation. Genetics **86:** 447–454.

Laurie, C. C., J. T. Bridgham and M. Choudhary, 1991 Association between DNA sequence variation and variation in expression of the *Adh* gene in natural populations of *Drosophila melanogaster*. Genetics **129:** 489–499.

Lemeunier, F., and S. Aulard, 1992 Inversion polymorphism in *Drosophila melanogaster*, pp. 339–405 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Cleveland.

Malpica, J. M., J. M. Vassallo, A. Frias and F. Fuentes-Bol, 1987 On recombination among *In(2L)t*, α-*Gpdh* and *Adh* in *Drosophila melanogaster*. Genetics **115:** 141–142.

Matthews, P., A. Agrostis, A. C. Taylor and W. McKechnie, 1992 An association between ADH protein levels and polymorphic nucleotide variation in the *Adh* gene of *Drosophila melanogaster*. Mol. Biol. Evol. **9:** 526–536.

Mukai, T., L. E. Mettler and S. I. Chigusa, 1971 Linkage disequilibrium in a local population of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **68:** 1065–1069.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M., and J. C. Miller, 1990 A simple method for estimating the average number of nucleotide substitutions within and between populations from restriction data. Genetics **125:** 873–879.

Oakeshott, J. G., J. B. Gibson, P. R. Anderson, W. R. Knibb, D. G. Anderson *et al.*, 1982 Alcohol dehydrogenase and Glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. Evolution **36:** 86–96.

Pipkin, S. B., E. Franklin-Springer, S. Law and S. Lubega, 1976 New studies of the *alcohol dehydrogenase* cline in *Drosophila melanogaster* from Mexico. J. Hered. **67:** 258–266.

Simmons, G. M., M. Kreitman, W. F. Quattelbaum and N. Miyashita, 1989 Molecular analysis of the alleles of *Alcohol dehydrogenase* along a cline in *D. melanogaster*. Evolution **43:** 393–409.

Tsacas, L., and D. Lachaise, 1974 Quatre nouvelles espèces de la Côte d'Ivoire du genre *Drosophila*, groupe *melanogaster* et discussion de l'origine du sous-groupe *melanogaster* (Diptera: Drosophiliidae). Ann. Univ. Abidjan **E7:** 193–211.

van Delden, W. V., and A. Kamping, 1989 The relation between the polymorphisms at the *Adh* and α*Gpdh* loci and the *In(2L)t* inversion in *Drosophila melanogaster* in relation to temperature. Evolution **43:** 775–793.

Vigue, C. L., and F. M. Johnson, 1973 Isozyme variability in species of the genus *Drosophila*. VI. Frequency-property-environment relationships of allelic *alcohol dehydrogenases* in *D. melanogaster*. Biochem. Genet. **9:** 213–227.

Voelker, R. A., T. Mukai and F. M. Johnson, 1977 Genetic variation in populations of *Drosophila melanogaster* from the western United States. Genetica **47:** 143–148.

Watada, M., Y. N. Tobari and S. Ohta, 1986 Genetic differentiation in Japanese populations of *Drosophila simulans* and *D. melanogaster*. I. Allozyme polymorphisms. Jpn. J. Genet. **61:** 469–480.

Wright, S., 1943 Isolation by distance. Genetics **28:** 114–138.

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Communicating editor: A. G. Clark

Appendix

Restriction site and indel genotype for each haplotype

In(2L)t-Malawi																										
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	1
0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1
0	0	1	0	1	1	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	0	1	0	0	0	0
0	0	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0	0
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0
0	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0	1	0	0	0	0
In(2L)t-Ivory Coast																										
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	1	0	0
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	0	1	0	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	0	1	0	1	1	1	1	0	0	1	0	1
0	1	0	0	0	1	1	0	0	0	0	1	1	1	0	0	1	0	1	1	1	1	0	0	1	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1	1	0	0	1
0	0	1	0	1	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1
0	0	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	0	1	0	1	0	0	0	1
0	1	1	0	1	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	0	0	0
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	0	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1	0	0	0	1
0	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	0	1	2
0	0	1	0	1	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	3
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1
In(2L)t-France																										
0	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	0	1	0	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1
0	1	1	1	0	0	0	0	0	0	0	1	1	1	0	1	1	0	0	0	0	0	1	1	1	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1
0	0	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1
0	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1
Standard-Slow-Malawi																										
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	1	0	0
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0
0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	0
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	0	0	0	0
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	0	1	0	0	0	0
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	0
0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	0	0	1	1	0	1	1	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1
0	1	1	0	0	1	0	0	0	0	1	1	1	1	0	1	1	0	0	1	1	0	1	0	0	0	1
0	0	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	0	0	0	0
0	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0
0	1	1	0	0	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0
0	0	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	1	0	0	1
0	1	1	0	0	1	0	0	0	0	1	1	1	1	0	1	1	0	0	1	1	0	1	0	1	0	0
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0

continued

Appendix continued

Restriction site and indel genotype for each haplotype

Standard-Slow-Malawi																											
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	0	0	0	1	1
0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	1	0	1	1	0	1	0	0	0	0	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	0	1	0	0	1
0	1	1	0	0	1	0	0	0	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	0	1
0	0	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	0	1
0	1	1	1	0	1	0	0	0	0	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	0	1	0	1	1	0	1	1	0	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	0	2
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0	3
0	0	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0	1	1	0	0	0	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	1	1	1	1	0	1	1	0	0	0	1
0	1	1	0	0	1	0	0	0	0	1	1	1	1	1	1	1	0	0	1	1	0	1	0	0	0	0	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0	1	0	0	0	0	1
0	1	1	0	0	1	0	0	0	0	0	1	0	1	0	1	1	0	1	1	1	0	1	0	0	0	0	1
0	1	1	0	0	1	0	0	0	0	1	1	1	1	0	1	1	0	1	1	1	0	1	0	1	0	0	1
0	1	1	0	1	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0	1	1	1	0	0	1
0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	0	0	1	1	1	0	1	0	0	1	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0	1	0	0	0	1	2
0	1	1	0	0	1	0	0	0	0	1	1	1	1	0	1	1	0	0	1	1	0	1	1	0	0	0	1
0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	0	0	1	1	0	1	1	0	0	0	1
Standard-Slow-Ivory Coast																											
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	1	6
0	0	1	0	1	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	1	0	1	1
0	1	1	1	0	0	0	0	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1	1	0	0	1	1
1	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	0	1	0	1	0	1	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1	0	1	0	1	1
0	1	1	1	0	0	0	0	0	0	0	1	1	1	0	1	1	0	0	0	1	0	1	1	1	0	1	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	0	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1	2
Standard-Slow-France																											
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1	2
0	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1	1
0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0	1	0	1	0	1	1
0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	0	1	3
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	0	1	0	1	2
0	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	1	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	1	0	1	2
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	1	0	1	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	1	1	1
0	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	0	1	0	0	0	1	1
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	6
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	0	0	0	1	1
0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0	1	0	1	0	1	2
0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	0	1	1	0	0	1	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	1
0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	1	0	1	0	0	1	3

continued

Appendix continued

Restriction site and indel genotype for each haplotype

Fast-Malawi																										
0	1	1	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1	1
0	1	1	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0	1	0	1	1	0	1	0	0	1	1
Fast-Ivory Coast																										
0	1	1	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1	10
0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1
0	1	1	0	0	1	0	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1	3
0	1	1	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0	1	1	1	1	0	0	1	0	1	1
Fast-France																										
0	1	1	0	0	1	0	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	1	1
0	1	1	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0	1	1	1	1	0	0	1	0	1	2
0	1	1	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	1
0	1	1	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1	4
0	1	1	0	0	1	0	0	1	1	0	1	1	1	1	1	1	0	0	1	1	1	0	0	1	0	1	4
0	1	1	0	0	1	0	0	0	1	0	1	1	1	0	1	1	0	1	1	0	1	0	1	0	0	1	1
0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	1
0	1	1	0	0	1	0	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	1	1
0	1	1	0	0	1	0	0	1	0	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	1
0	1	1	0	0	1	0	0	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	16
0	1	1	0	0	1	0	0	1	1	0	1	1	1	1	1	1	0	1	0	1	1	0	1	0	0	1	1
0	1	1	0	0	1	0	0	1	1	0	1	1	1	0	1	1	0	1	1	1	1	0	0	0	0	1	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	1	0	1	1
0	1	1	0	0	1	0	0	0	1	0	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	2
0	1	1	0	0	1	0	0	0	1	0	1	1	1	1	1	1	0	1	1	0	1	0	1	0	0	1	1
0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	1	7

Characters are, from 5′ to 3′: indel ID1, *Msp*I -346, *Dde*I 9, *Ban*I 102, indel ID2, *Hae*III 287, *Dde*I 321, *Alu*I 562, indels Δ1, Δ2 and ID3, *Cfo*I 573, *Msp*I 583, *Hae*III 688, *Hae*III 816, *Msp*I 818, *Cfo*I 864, indel ID5, *Sau*3AI 1354, *Hae*III 1425, *Sau*96I 1425, *Dde*I 1518, *Dde*I 1527, *Alu*I 1596, indel ID6, *Dde*I 1875, *Hae*III 1925. The last column is the number for each haplotype.