The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*

Aya Takahashi*†‡, Shun-Chern Tsaur*§, Jerry A. Coyne*, and Chung-I Wu*¶

*Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637; and †Laboratory of Applied Zoology, Faculty of Agriculture, Hokkaido University, Sapporo 060, Japan

The cuticular hydrocarbon (CH) pheromones in *Drosophila melanogaster* exhibit strong geographic variation. African and Caribbean populations have a high ratio of 5,9 heptacosadiene/7,11 heptacosadiene (the ''High'' CH type), whereas populations from all other areas have a low ratio (''Low'' CH type). Based on previous genetic mapping, DNA markers were developed that localized the genetic basis of this CH polymorphism to within a 13-kb region. We then carried out a hierarchical search for diagnostic nucleotide sites starting with four lines, and increasing to 24 and 43 lines from a worldwide collection. Within the 13-kb region, only one variable site shows a complete concordance with the CH phenotype. This is a 16-bp deletion in the 5′ region of a desaturase gene (*desat2*) that was recently suggested to be responsible for the CH polymorphism on the basis of its expression [Dallerac, R., Labeur, C., Jallon, J.-M., Knipple, D. C., Roelofs, W. L. & Wicker-Thomas, C. (2000) *Proc. Natl. Acad. Sci.* 97, 9449–9454]. The cosmopolitan Low type is derived from the ancestral High type, and DNA sequence variations suggest that the former spread worldwide with the aid of positive selection. Whether this CH variation could be a component of the sexual isolation between Zimbabwe and other cosmopolitan populations remains an interesting and unresolved question.

The molecular-genetic basis of phenotypic variations, either within or between species, is an important topic in genetics, evolution, agriculture, and medicine. Such differences may be morphological, physiological (1, 2), or behavioral (3) and could have either simple (4–7) or complex (8–11) genetic bases. Traits pertaining to species or racial differentiation are of special interest as they promise to shed light on the process of speciation (12–17). Despite a number of methodologies developed for the purpose of complex trait mapping, successes in molecular identification of quantitative trait loci or nucleotides (QTL or QTN, respectively) have been limited to a few cases (7, 18–20).

Naturally occurring polymorphism in the cuticular hydrocarbon (CH) pheromones in *Drosophila melanogaster* is a promising system for molecular cloning and characterization (21). In different geographical populations of this species, the ratio of the two major isomers of the main female cuticular hydrocarbon, 5,9 heptacosadiene (5,9-HD) and 7,11 heptacosadiene (7,11-HD), varies consistently. Many African and Caribbean populations show a high ratio of 5,9-HD/7,11-HD, whereas all other populations show a low ratio (1, 21). We shall refer to the former as the ''High'' and the latter as the ''Low'' phenotype. As this geographical variation exists in *D. melanogaster*, a species highly amenable to molecular genetic analysis, it can thus be a useful system for identifying QTNs for naturally occurring complex traits.

An early study (22) suggested that the delta-9 desaturase gene (*desat1*) is a strong candidate for the CH polymorphism because of its genetic location and the function in the desaturation of the fatty acid. Genetic mapping (1) showed that the phenotypic difference indeed maps to the 1.7-centimorgan (cM) interval where *desat1* resides. In this study, we further define the chromosomal segment responsible for the pheromonal difference and then carry out association mapping within the interval among worldwide strains. Specific QTNs for the trait can thus be delineated. Recently, Dallerac *et al.* (2) reported another desaturase gene (*desat2*) tandemly arrayed with *desat1* and, based on the expression pattern, concluded that *desat2* is responsible for the CH polymorphism. Here, we independently identify *desat2* as the responsible gene through genetic mapping, pinpoint a deletion in its 5′ region as the probable cause of the CH variation, and show that the cosmopolitan low phenotype probably arose from the high phenotype and spread by natural selection.

In *D. melanogaster*, there exists strong but asymmetric sexual isolation between populations in southern Africa surrounding Zimbabwe (Z) and other cosmopolitan populations (M; see ref. 3). Because several African populations still harbor a wide range of behavioral types, this system may be at the incipient stage of speciation (23). Previously, Coyne and colleagues (24, 25) showed an effect of the CH difference on male mate-choice between species of the *D. melanogaster* group, but found no detectable effect of the heptacosadiene polymorphism on mating frequencies within *D. melanogaster* (1). Because of the large number of genes involved in the Z–M behavior polymorphisms (refs. 11 and 26; C.-T. Ting, A. Takahashi, and C.-I Wu, unpublished work) and the incomplete transfer of CH in the rub-off experiments (1), it is still possible that the CH polymorphism may be one of the many determinants of the Z–M behavior differences (28). Knowing the genetic basis of CH will allow us to address this question directly by constructing and comparing lines that differ only at the CH locus.

With the identification of the molecular basis of this pheromonal difference, we wish to answer several questions: (*i*) What has been the direction of evolution: is the Low or High hydrocarbon phenotype the ancestral form? (*ii*) What forces drive the change and what maintains the geographical distribution?

Materials and Methods

***Drosophila* Stocks.** The *D. melanogaster* lines used are described in refs. 1, 3, and 23. Their geographical origins are indicated in Tables 1 and 2.

Table 1. DNA polymorphism in the 13-kb where the CH phenotype is mapped (see Fig. 1)

Positions of the polymorphic sites

					550-bp region surrounding 10294						630-bp region surrounding 12552					
Phenotype (geographic origin)	*D. melanogaster* lines	**4012**	**6500**	**9512**	**10294**	10387	10416	10597	10730	**11009**	12441	12452	12515	**12552**	12641	12748
D. simulans		C		C	**T**	+	T	CA	G		G	D	GT	+	C	G
D. mauritiana		C		*T*	**T**	+		CA	G		G	D	GT	+	C	G
D. sechellia		C		*T*	**T**	+	T	CA	G		G	D	GT	+	C	G
Low 5,9-HD (Cosmopolitan)	Oregon-R	*T*	T	–	***A***	–	*A*	AG	–	T	–	*I*	*AA*	**D**	–	*T*
	closs3, NY	–	T	–	***A***	–	*A*	AG	–	T	–	*I*	*AA*	**D**	–	*T/–*
	CAM110, NC	–	T	*A/–*	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	French V3-1	–	G	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	Yep11, Australia	–	T/G	*A/–*	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	cu kar	*T*	T	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	TWN24, Taiwan	–	T	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	QD12, Japan	*T*	G	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	QL4, China	–	T	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	Taishan, China	–	T	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
Low 5,9-HD (African/ Caribbean)	LA20, Zambia	–	T	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	OK17, Botswana	*T*	T	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
	ZH27, Zimbabwe	*T*	G	–	***A***	–	*A*	AG	–	G	–	*I*	*AA*	**D**	–	–
Intermediate (Cosmopolitan)	Hg, CA	–	T/G	*A/–*	***A/–***	–	*A*	AG	–	G	–	*I*	*AA*	**D/–**	–	–
	GF60, IN	–	G	*A/–*	***A/–***	–	*A*	AG	–	G	–	*I*	*AA*	**D/–**	–	–
High 5,9-HD (African/ Caribbean)	ZS53, Zimbabwe	–	G	*A*	–	–	*A*	AG	–	T	–	*I*	*AA*	–	–	–
	ZS2, Zimbabwe	–	G	–	–	–	–	AG	–	G	–	*I*	*AA*	–	–	–
	Caribbean A, St. Croix	–	G	*A*	–	D	*A*	AG	–	G	*A*	*I*	*AA*	+'	–	–
	Caribbean B, St. Croix	–	G	*A*	–	–	–	AG	–	G	–	*I*	*AA*	–	–	–
	Caribbean C, St. Croix	–	G	*A*	–	–	–	AG	–	G	–	*I*	*AA*	–	–	–
	Caribbean D, Dominican Republic	–	G	*A*	–	–	*A*	AG	–	G	*A*	*I*	*AA*	+'	–	–
	Caribbean E, Cuba	*T*	T	–	–	–	–	AG	A	G	–	*I*	*AA*	–	*T*	–
	Caribbean F, Dominican Republic	–	G	*A*	–	D	–	–	–	G	–	–	–	–	–	–
High 5,9-HD (Cosmopolitan)	Iso-1	–	T	–	–	–	–	AG	A	G	–	*I*	*AA*	–	*T*	–

Sequences are grouped by the CH phenotype of the chromosome and their geographic origin (in parentheses) as indicated in the left column. The collection site of each line is indicated in the second column (NY = New York state, NC = North Carolina, CA = California, and IN = Indiana; Iso-1 is a North American strain used in the *Drosophila* genome project). The two sites (10294 and 12552) that show a complete association with the CH phenotype are highlighted in bold. To avoid cluttering the table, singletons in these two stretches are omitted; these sites are site 10385 in Caribbean D, site 10649, 12634, 12900 in Caribbean F, site 12542 in OK17, and site 12685 in QL4. These singletons are used in the analysis of Table 3. Inferred derived nucleotides are italicized and the ancestral state is indicated by –. D, deletion; + or +' nondeletion (+ and +' differ by 1 bp).

Measuring Cuticular Hydrocarbon Components. Cuticular hydrocarbons were extracted from 4-day-old virgin female flies and analyzed by gas chromatography according to the method described in Coyne *et al.* (24). N-hexacosane was added to each sample as a standard. The hydrocarbon profiles were classified as either High type (high 5,9-HD) or Low type (low 7,11-HD) by the criteria used in Coyne *et al.* (1).

Genetic Mapping and DNA Sequencing. Recombinants between *cu* (86D1–4) and *kar* (87C8) were obtained by backcrossing F1 females from the cross between the *cu kar* (Low hydrocarbon phenotype) and Caribbean A (High hydrocarbon phenotype) lines to *cu kar* males. The recombinant chromosomes were extracted over balancers. Restriction fragment length polymorphism (RFLP) and SSCP (120 mM NaCl/5 mM sodium citrate/20 mM sodium phosphate, pH 6.8) markers were developed to assist the mapping. Sequences of the entire mapped region (13 kb; Fig. 1) were obtained from two Low strains, Oregon-R and *cu kar*, and two High strains, Z53 and Caribbean A. DNA sequences from the *Drosophila* sequence databank (29) were used for primer design. From these sequences, six candidate sites were identified and 20 more lines of either the High or Low CH phenotype were sequenced near the six sites. This process ruled out four of the six sites as the diagnostic site. Forty-three isofemale lines were genotyped at the remaining two diagnostic sites by RFLP or sequencing (Table 2).

Statistical Tests. Nucleotide variation, measured by the statistics θ_π and θ_ω, were calculated according to Tajima (30) and Watterson (31), respectively. The H test for detecting positive selection follows the description of Fay and Wu (32), which requires an estimate of divergence from the outgroup to correct for misinference of the ancestral state. We use the divergence level between *D. melanogaster* and *Drosophila simulans* from our own data (0.0557). Heterozygous nucleotides were counted as two different samples. In two instances, the mutant nucleotides at the two adjacent sites are always linked and we conservatively treat each as one single mutation event.

EVOLUTION

Table 2. Polymorphism data from 19 additional lines at the two sites with a complete association between genotype and phenotype in Table 1

Phenotype (geographic origin)	*D. melanogaster* lines	(Z/M behavioral phenotype)*	Diagnostic sites	
			10294	12552
Low 5,9-HD (Cosmopolitan)	Ten lines from Table 1	(Five lines are M)	A × 10	D × 10
	Miami M-103, FL		A	D
	Florida 13, FL		A	D
	Toonda, Australia		A	D
	EC SFS 2-11, Israel		A	D
	Ica, Peru		A	D
	Kolonia, Carolina Island		A	D
	Mysore, India		A	D
	Oahu, HI†		T	D
	Hg, CA‡	(M)	A	D
	GF60, IN‡		A	D
Intermediate (Cosmopolitan)	Hg, CA‡	(M)	A/T	D/+
	GF60, IN‡	(M)	A/T	D/+
Low 5,9-HD (African/Caribbean)	Three lines from Table 1	(M)	A × 3	D × 3
	LA47, Zambia	(M)	A	D
	LA66, Zambia	(M)	A	D
	Tunisia 8, Tunisia		A	D
Intermediate (African/Caribbean)	ZH34, Zimbabwe	(Z)	A/T	D/+
High 5,9-HD (African/Caribbean)	Two lines from Table 1 (Zimbabwe)	(Z)	T × 2	+ × 2
	Six lines from Table 1 (Caribbean)	(M)	T × 6	+ × 6
	ZS30, Zimbabwe	(Z)	T	+
	ZS56, Zimbabwe	(Z)	T	+
	LA69, Zambia	(Z)	T	+
	OK87, Botswana	(Z)	T	+
	OK59, Botswana	(Z/M intermediate)	T	+
	S. Africa 17, South Africa		T	+
	Kenya 48, Kenya		T	+
High 5,9-HD (Cosmopolitan)	One line from Table 1	(M)	T	+

See legends of Table 1 for details (FL = Florida, HI = Hawaii).
*The behavioral phenotype classified by Hollocher et al. (23) is given in this column.
†This line from Hawaii distinguishes the two sites.
‡Lines polymorphic for both phenotype and genotype.

Results

High-Resolution Genetic Mapping. The chromosomal region responsible for the CH polymorphism was previously mapped by recombination analysis to between *cu* and *kar* markers (placed closer to *kar*) by Coyne *et al.* (1). In this study, we generated 173 more recombinants between the *cu kar* (of the Low type) and Caribbean A (which has the High CH type) line. A series of DNA markers were developed to map where the breakpoints between *cu* and *kar* occurred. Among the 173 recombinants, eight have recombination breakpoints between *GstD1* and *Hsp70Bbc* and two of them have breakpoints even closer to each other (Fig. 1). The phenotypes of these latter two suggest that the allelic difference in CH maps to between the *Pp1–87Bb* and *desat1* markers, a distance of about 13 kb (Fig. 2). According to the genomic sequence data of *Drosophila* (29, 33), this region includes seven genes, but the previously implicated desaturase gene, *desat1*, falls outside of it. Significantly, there is a moderately divergent duplicate copy of *desat1* named *desat2* (2).

Association Between Nucleotide Changes and the CH Differences. To identify the QTNs responsible for the CH phenotype, DNA sequences of the entire 13-kb region were obtained from four different lines. Candidate nucleotides that may account for the CH variation could then be subsequently tested by using a larger number of lines. The four lines include two High-phenotype strains (Z53 and Caribbean A) and two Low-phenotype strains (Oregon-R and *cu kar*). There are six variable sites where the mutant exists in two lines and five of these sites separate the four lines by their CH phenotype. Singleton variations among these four lines are not associated with the CH difference and were not considered further. DNA sequences near these six nonsingleton sites were then obtained from 20 more lines of either the High or Low CH phenotype. We also sequenced one line each from the three sibling species, *D. simulans*, *Drosophila mauritiana*, and *Drosophila sechellia*. The results are shown in Table 1; sites of the six original polymorphisms are in boldface.

Table 1 shows two positions, 10294 and 12552, where the association between the DNA variation among strains and the CH phenotype is complete. The former is a single nucleotide polymorphism and the latter is an insertion/deletion polymorphism. These two sites are boldfaced. With the sequences from the outgroup, we can infer the ancestral vs. derived nucleotide. The ancestral type is indicated by a hyphen and the derived type is displayed in letters. (To avoid cluttering the table, singleton mutations are not included in the presentation.) Both CH phenotypes are present in most continents but in very different
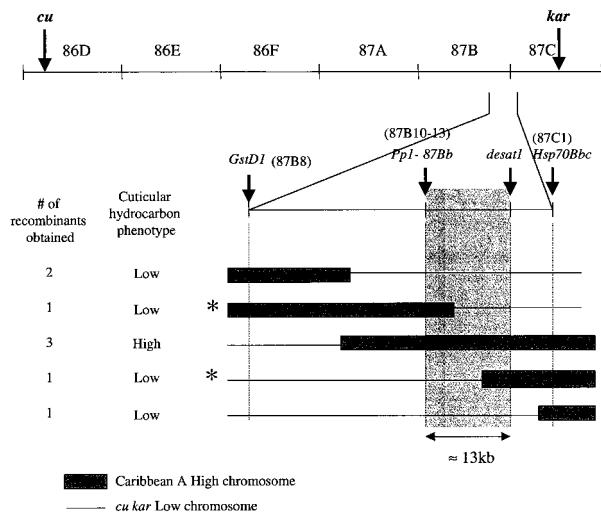
Takahashi *et al.*

Fig. 1. Results of the DNA marker assisted mapping. Recombinants between *cu* and *kar* morphological markers were obtained between the Caribbean A (High CH type) and the *cu kar* chromosome (Low CH type). Top ruler indicates the cytological locations of the four molecular markers, *GstD1*, *Hsp70Bbc*, *Pp1-87Bb*, and *desat1*. The two recombinants with ∗ indicate that the CH phenotype maps to the 13-kb region (shaded) between *Pp1–87Bb* and *desat1*.

Fig. 2. Genes in the 13-kb region where the CH phenotype is mapped are shown as horizontal line segments. Names of the genes, if known, are given. Nucleotide positions of the six polymorphic sites (excluding singletons) in the initial survey among the two Low and two High lines are shown. Triangle denotes the 16-bp deletions that is the cause of the CH polymorphism (see text).

frequencies. It is interesting to note that the cosmopolitan Low CH type has been derived from the High type and subsequently spread worldwide.

For the two sites that are diagnostic of the CH difference, 19 more lines were surveyed by PCR and restriction pattern analysis (see *Materials and Methods*), making the total number of lines 43. The results presented in Table 2 show a complete association between the deletion/insertion at site 12552 and the Low/High CH phenotype among the collection of 43 lines worldwide. In this expanded data set, one line from Oahu, Hawaii enabled us to reject site 10294 as the determinant of the CH variation. This line has the cosmopolitan Low type but has the High type nucleotide "T" at site 10294. (This unusual pattern has been confirmed twice more from the same line.) Therefore, site 12552 appears to be the determinant of the CH polymorphism. The High CH allele has the ancestral wild-type (inserted) state and the Low allele has the 16-bp deletion at this site. This site is about 150 bp upstream of the putative translation start of *desat2* and may be in the promoter region.

The "intermediate" phenotype observed in two North American lines, Hg of California and GF60 of Indiana, is of some interest. Because both intermediate lines are also polymorphic in the DNA region (Table 1), six females from each line were individually phenotyped and genotyped. For both lines, one individual was found to be of the cosmopolitan type in both phenotype and genotype, whereas the other five individuals remain intermediate in phenotype and were also heterozygous at site 12552 (see Table 2). One particular line from Africa (LA66; Table 2) with the Low CH phenotype has a deletion and a stop codon in the coding region of *desat2*, in addition to the deletion at site 12552. This result corroborates the loss-of-function interpretation of the Low CH phenotype (2), which permits further degeneracy of the coding sequence. By coincidence, the Iso-1 line used in the *Drosophila* genome project carries the African/Caribbean allele, which is rare among North American flies. Finally, we note that when the *desat2* genotype is superimposed on the behavioral observation of Hollocher *et al.* (23), in the African collections from LA, OK, and ZH, the High CH type is invariably more Z-like
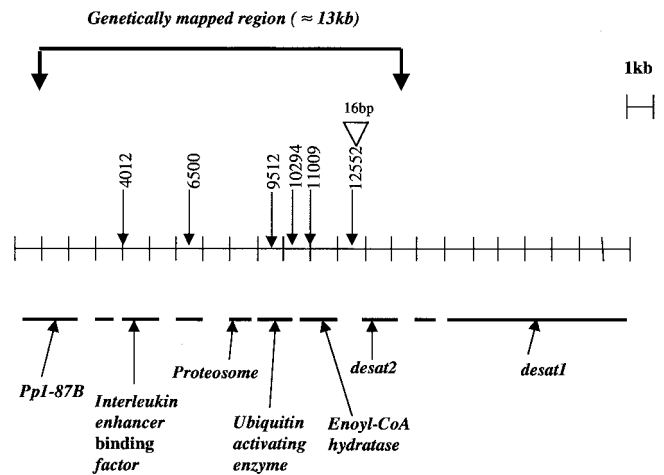
(Tables 1 and 2). The *desat2* genotype's potential importance will be discussed later.

Natural Selection and DNA Variation. Given that the cosmopolitan Low phenotype or, more specifically, the 16-bp deletion 5′ of *desat2* (site 12552) is derived from the ancestral High CH type, we ask whether natural selection has played a role in its spread worldwide, responsible for the geographical differentiation between African/Caribbean flies and those from other continents (Asia, Australia, Europe, and North America).

When a mutation is driven to fixation (or high frequency) by positive selection, nearby neutral variations hitchhike with it and variation around the site will be reduced (34, 35). After the reduction, the number of sites with the mutation at a very high frequency may in fact be increased from that before the hitchhike (32). Table 1 shows that the two regions surveyed between the two shaded sites (10294 and 12552) are virtually monomorphic among chromosomes with the cosmopolitan Low phenotype. This is not true among those with the ancestral (African/Caribbean) High type. Furthermore, though the Low-phenotype chromosomes have reduced variability in this region, many of the polymorphic sites carry the derived nucleotide instead of the ancestral type inferred from the outgroup.

For statistical tests, we combined the two surveyed regions (1180 bp from 24 lines; Table 1). It should be noted that the influence of selection on the CH phenotype is not likely to extend much beyond site 10294, because the three sites to its left do not show an association with the phenotype. Table 3 gives the frequency spectrum and summary statistics of the sequences of Table 1. We group the sequences by their geographical origin and analyze them both separately and jointly. The populations were analyzed separately to tease apart the effects of selection and population structure on DNA variation. The cosmopolitan populations are less polymorphic than the African/Caribbean populations, as is shown by the two estimates of nucleotide diversity, θ (30, 31). In this region, the African/Caribbean populations are 2.5–3 times as variable as the cosmopolitan populations. The large difference is unusual in light of a recent survey (P. Andolfatto, unpublished work) of DNA variation in this species. In that survey, 28 of the 60 autosomal loci show the

Table 3. Frequency spectra of DNA sequences between sites 10294 and 12748 of Table 1 (1180 bp in total)

Cosmopolitan lines ($n = 16$)		African and Caribbean lines ($n = 11$)	
Mutant frequency	Number of sites	Mutant frequency	Number of sites
1/16	3	1/11	7
2/16	1	2/11	3
13/16	2	3/11	2
15/16	1	6/11	1
(16/16)*	(3)	10/11	3
P(H)† = 0.0347	(0.001)	P(H) = 0.141	
P(D)† = 0.111		P(D) = 0.170	
θ_π = 1.38		θ_π = 4.22	
θ_w = 2.11		θ_w = 5.46	

Combined data ($n = 27$)			
Mutant frequency	Number of sites		
1/27	6	P(H) = 0.034	
2/27	5	P(D) = 0.085	
16/27	2	θ_π = 2.74	
21/27	1	θ_w = 4.41	
26/27	3		

The analyses are done by continents separately and collectively. Mutant denotes the derived variant. Singletons not shown on Table 1 are included in this analysis.

*Three sites (including two sites that are doublets; see *Materials and Methods*) are fixed in the cosmopolitan, but not the African/Caribbean collection. When these sites are included in the calculation of the H statistics, the corresponding P value is given in the parentheses.

$^†P$ values for the H and D statistics are given.

cosmopolitan populations to be more variable and, among the remaining 32 loci, only one approaches the ratio of 3:1.

To test whether there is an excess in the number of sites with high-frequency mutations, as is sometimes the case with hitchhiking (32), we use both Tajima's (30) D test and Fay and Wu's H test. The D statistic measures the excess in both the high and low frequency sites, whereas the H statistics measures the excess of only the high frequency sites. It should be noted that the H statistic is particularly useful when the number of segregating sites is small (sometimes as small as one; see Table 2 of ref. 32). The H statistic suggests that there is a significant skew of variation to high frequency in the cosmopolitan population and in the entire collection ($P < 0.05$). The African/Caribbean populations do not show a pattern as strong ($P > 0.05$). The excess may be more significant than indicated. There are at least three derived variants (including two pairs of doublets, AG and AA) that are fixed in the cosmopolitan, but not African/ Caribbean, populations. If these polymorphic sites are included in the H test, the P value for the cosmopolitan populations is 0.001. In conclusion, the immediate vicinity (<2.5 kb) of the site responsible for the cosmopolitan CH phenotype appears to have been driven to high frequency worldwide by positive selection.

Discussion

Genetic mapping and DNA sequence analysis have revealed a complete association between the deletion at site 12552 and the CH difference in *D. melanogaster* among 43 lines from a worldwide collection. This is also the *only* site within the 13-kb region defined by genetic mapping that shows complete concordance between genotype and phenotype. Another site at 10294 has a single exception for being a site with complete association in a line collected from Oahu, Hawaii, corroborating tight (but incomplete) linkage between site 10294 and the CH phenotype. It is still

relatively rare to be able to identify the QTNs of a common phenotypic polymorphism in nature by means of association study.

Site 12552 is a 16-bp deletion about 150 bp upstream of the translation start site of *desat2*. Dallerac *et al.* (2) recently showed that *desat2* is expressed only in females of the High CH phenotype, but not in females of the Low CH phenotype nor in males of either type. They suggest that the cosmopolitan type lacks a functional promoter of this gene. It seems likely that the deletion at site 12552 abolishes the expression of *desat2*. This interpretation also explains why the existence of a stop codon in the coding region of at least one line produces Low CH phenotype.

DNA-sequence data show that the cosmopolitan Low phenotype is derived from the ancestral High type. Our analysis suggests that the cosmopolitan type spread worldwide with the aid of positive selection because of two observations: (i) the overall level of variation among the cosmopolitan type is uncommonly low relative to the African collection (cf. ref. 37) and (ii) the residual variation in the cosmopolitan collection is skewed toward high-frequency variants, a pattern most consistent with the influence of positive selection (32). Several questions remain about the spread of the cosmopolitan CH type. Considering that Africa is probably the ancestral home of *D. melanogaster* (38), was this spread of the derived phenotype concurrent with the original spread of *D. melanogaster* throughout the world, or did it happen after the colonization? Why are Caribbean populations the only populations outside of Africa that have the high CH type? Are these populations the remnants of the early expansion of *D. melanogaster* before the spread of the low allele? Did they result from a secondary influx from Africa associated with recent human activities? Dating the spread of the cosmopolitan CH type by analyzing the linked neutral variation could resolve these issues and help date the initial colonization.

Our study reveals a simple genetic basis for a seemingly complex trait that not only segregates at high frequency in nature, but also differentiates populations geographically. There are cases of single-gene differences underlying naturally occurring traits (4–7), but species or racial differences could have a very complex genetic basis (8–11). The simplicity of the genetics and, furthermore, the observation of a loss-of-function mutation that spread are quite striking in this CH variation.

What would the selective forces be that drove the spread of the Low allele? Climatic adaptation against desiccation has been proposed (27, 39). Alternatively, if this phenotype is a component of the complex genetics of the Z–M sexual isolation (refs. 11 and 26; C.-T. Ting, A. Takahashi, and C.-I Wu, unpublished work), then the answer may partially lie in sexual selection. Although the behavior and CH phenotype are not correlated in interpopulation comparisons (Caribbean lines are not Z-like in behavior; also see ref. 1), Table 2 suggests a correlation within south central African populations. Could the pheromone be a necessary but not sufficient condition for the Z-type behavior? Many of the issues discussed so far will benefit from gene transformation experiments, especially site-directed gene targeting (36). It is certainly most desirable to be able to connect evolutionarily important phenotypes to their underlying molecular genetic bases and to analyze the forces driving the changes at both levels. The CH differentiation in *D. melanogaster* is a promising system in this respect.

Takahashi *et al.*

1. Coyne, J. A., Wicker-Thomas, C. & Jallon, J.-M. (1999) *Genet. Res.* **73,** 189–203.
2. Dallerac, R., Labeur, C., Jallon, J.-M., Knipple, D. C., Roelofs, W. L. & Wicker-Thomas, C. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9449–9454. (First Published August 1, 2000; 10.1073/pnas.150243997)
3. Wu, C.-I., Hollocher, H., Begun, D. J., Aquadro, C. F., Xu, Y. & Wu, M.-L. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2519–2523.
4. Orr, H. A. & Coyne, J. A. (1992) *Amer. Nat.* **140,** 725–742.
5. Osborne, K. A., Robichon, A., Burgess, E., Butland, S., Shaw, R. A., Coulthard, A., Pereira, H. S., Greenspan, R. J. & Sokolowski, M. B. (1997) *Science* **277,** 834–836.
6. Bradshaw, H. D., Wilbert, S. M., Otto, K. G. & Schemske, D. W. (1995) *Nature (London)* **376,** 762–765.
7. Sucena, E. & Stern, D. L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 4530–4534.
8. Wu, C.-I. & Palopoli, M. F. (1994) *Annu. Rev. Genet.* **28,** 283–308.
9. True, J. R., Weir, B. S. & Laurie, C. C. (1996) *Genetics* **142,** 819–837.
10. Nuzhdin, S. V., Dilda, C. L. & Mackay, T. F. (1999) *Genetics* **153,** 1317–1331.
11. Hollocher, H., Ting, C. T., Wu, M. L. & Wu, C.-I. (1997) *Genetics* **147,** 1191–1201.
12. Coyne, J. A. (1992) *Nature (London)* **355,** 511–515.
13. Ting, C.-T., Tsaur, S.-C. & Wu, C.-I. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5313–5316. (First Published April 25, 2000; 10.1073/pnas.090541597)
14. Werren, J. H. (1998) in *Endless Forms: Species and Speciation*, eds. Howard, D. & Berlocher, S. (Oxford Univ. Press, Oxford), pp. 245–260.
15. Palumbi, S. R. (1998) in *Endless Forms: Species and Speciation*, eds. Howard, D. & Berlocher, S. (Oxford Univ. Press, Oxford), pp. 271–278.
16. Markow, T. A. & Hocutt, G. D. (1998) in *Endless Forms: Species and Speciation*, eds. Howard, D. & Berlocher, S. (Oxford Univ. Press, Oxford), pp. 234–244.
17. Shaw, K. L. (1998) in *Endless Forms: Species and Speciation*, eds. Howard, D. & Berlocher, S. (Oxford Univ. Press, Oxford), pp. 44–56.
18. Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. (1999) *Nature (London)* **398,** 236–239.
19. Ting, C.-T., Tsaur, S.-C., Wu, M.-L. & Wu, C.-I. (1998) *Science* **282,** 1501–1504.
20. Frary, A., Nesbitt, T. C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B. & Tanksley, S. D. (2000) *Science* **289,** 85–88.
21. Ferveur, J. F., Cobb, M., Boukella, H. & Jallon, J. M. (1996) *Genetica (The Hague)* **97,** 73–80.
22. Wicker-Thomas, C., Henriet, C. & Dallerac, R. (1997) *Insect Biochem. Mol. Biol.* **27,** 963–972.
23. Hollocher, H., Ting, C. T., Pollack, F. & Wu, C.-I. (1997) *Evolution* **51,** 1175–1181.
24. Coyne, J. A., Crittenden, A. O. & Mah, K. (1994) *Science* **265,** 1461–1464.
25. Coyne, J. A. & Charlesworth, B. (1997) *Genetics* **145,** 1015–1030.
26. Takahashi, A. (2000) Ph.D. thesis (Hokkaido Univ., Sapporo, Japan).
27. Toolson, E. C. (1988) *Cuticular Permeability and Epicuticular Hydrocarbon Composition of Sonoran Desert Drosophila Pseudoobscura* (Wroclaw Technical Univ. Press, Wroclaw, Poland).
28. Alipaz, J. (2000) Ph.D. thesis (Univ. of Chicago, Chicago).
29. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) *Science* **287,** 2185–2195.
30. Tajima, F. (1989) *Genetics* **123,** 585–595.
31. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7,** 256–276.
32. Fay, J. C. & Wu, C.-I. (2000) *Genetics* **155,** 1405–1413.
33. Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., *et al.* (2000) *Science* **287,** 2204–2215.
34. Maynard Smith, J. & Haigh, J. (1974) *Genet. Res.* **23,** 23–35.
35. Stephan, W. & Mitchell, S. J. (1992) *Genetics* **132,** 1039–1045.
36. Rong, Y. S. & Golic, K. G. (2000) *Science* **288,** 2013–2018.
37. Andolfatto, P. & Kreitman, M. (2000) *Genetics* **154,** 1681–1691.
38. David, J. R. & Capy, P. (1988) *Trends Genet.* **4,** 106–111.
39. Howard, R. W. & Blomquist, G. J. (1982) *Annu. Rev. Entomol.* **27,** 149–172.

EVOLUTION