

Population Genetics of the Developmental Gene *optomotor-blind* (*omb*) in *Drosophila polymorpha*: Evidence for a Role in Abdominal Pigmentation Variation

Jennifer A. Brisson,¹ Alan R. Templeton and Ian Duncan

Campus Box 1137, Department of Biology, Washington University, Saint Louis, Missouri 63130

Manuscript received June 5, 2004

Accepted for publication July 26, 2004

ABSTRACT

The developmental gene *optomotor-blind* (*omb*) encodes a T-box-containing transcription factor that has multiple roles in *Drosophila* development. Previous genetic analyses established that *omb* plays a key role in establishing the abdominal pigmentation pattern of *Drosophila melanogaster*. In this report we examine patterns of *omb* nucleotide variation in *D. polymorpha*, a species that is highly polymorphic for the phenotype of abdominal pigmentation. Haplotypes at this locus fall into two classes that are separated by six mutational steps; five of these mutational events result in amino acid changes. Two lines of evidence are consistent with a role for *omb* in the abdominal pigmentation polymorphism of *D. polymorpha*. First, we find that haplotype classes of *omb* are correlated with abdominal pigmentation phenotypes, as are microsatellite repeat numbers in the region. Second, tests of selection reveal that the two haplotype classes have been maintained by balancing selection. Within each class there is a significantly low amount of diversity, indicative of previous selective sweeps. An analysis including *D. polymorpha*'s closest relatives (members of the *cardini* group) provides evidence for directional selection across species. Selection at this locus is expected if *omb* contributes to variation in abdominal pigmentation, since this trait is likely of ecological importance.

THE gene *optomotor-blind* (*omb*) is one of the essential “toolkit” genes used to coordinate development in diverse animal taxa (CARROLL *et al.* 2001). The *omb* locus encodes a T-domain transcription factor (PFLUGFELDER *et al.* 1992a,b) that plays key patterning roles in the central nervous system (BRUNNER *et al.* 1992; PFLUGFELDER *et al.* 1992b; POECK *et al.* 1993) and leg and wing imaginal discs (GRIMM and PFLUGFELDER 1996; LECUIT *et al.* 1996; NELLEN *et al.* 1996) of *Drosophila melanogaster*. The T-domain is a highly conserved region that functions in DNA binding (PFLUGFELDER *et al.* 1992a; KISPERT and HERRMANN 1993; KISPERT *et al.* 1995). The T-box family is named after its founding member, the mouse Brachyury (T) gene (BOLLAG *et al.* 1994).

In addition to its early developmental roles, *omb* is necessary for patterning the pigment band in each adult abdominal segment: in the developing pupa, *omb* is expressed at the posterior edge of each abdominal segment in a stripe that prefigures the pigment band of the adult (KOPP and DUNCAN 2002). *omb* loss-of-function clones lose pigmentation and overexpression of *omb* leads to expansion of the pigment band (KOPP and DUNCAN 1997). The pattern and extent of the pigment band is a character that is highly variable in the genus

(reviewed in TRUE 2003; WITTKOPP *et al.* 2003a), differing among species groups (GARCIA-BELLIDO 1983; GOMPEL and CARROLL 2003), between closely related species (HOLLOCHER *et al.* 2000), between sexes within a species (KOPP *et al.* 2000), and geographically within a species (HEED 1963; LEE 1963; ROBERTSON *et al.* 1977; OHNISHI 1985). The control of pigmentation in *D. melanogaster* by *omb* suggests that *omb* may also be important in specifying this natural variation within and between species of *Drosophila*.

To explore *omb*'s role in abdominal pigmentation variation, we examined patterns of nucleotide variation at the *omb* locus in natural populations of a South American species of fruit fly, *D. polymorpha*. *D. polymorpha* is a member of the *cardini* group, members of which exhibit extensive abdominal pigmentation variation, both within and between species (STALKER 1953; HEED 1963; HEED and RUSSELL 1971; HOLLOCHER *et al.* 2000). *D. polymorpha* is one of the most polymorphic members of this group, with individuals that vary from having almost completely pigmented abdomens to having almost completely unpigmented abdomens (DA CUNHA 1949; MARTINEZ and CORDEIRO 1970). Although abdominal pigmentation is environmentally determined in some species (DAVID *et al.* 1990; GIBERT *et al.* 1996), early crossing experiments determined that the abdominal pigmentation polymorphism in *D. polymorpha* has a strong genetic basis. The trait was initially attributed to a single locus

¹Corresponding author: Campus Box 1137, Department of Biology, Washington University, St. Louis, MO 63130.
E-mail: brisson@biology.wustl.edu

with two alleles (DA CUNHA 1949), but later experiments indicated the action of at least one additional modifier locus (MARTINEZ and CORDEIRO 1970). The variation exhibited by this species is due in part to differences in the width of the posterior pigment band, the character that *omb* patterns in *D. melanogaster* (KOPP and DUNCAN 1997).

We approached the role of *omb* in controlling pigmentation variation in *D. polymorpha* from two complementary angles. First, we asked if there is a correlation between *omb* haplotypes and abdominal pigmentation phenotypes of *D. polymorpha*. Second, we investigated whether there is a signature of selection in *omb* haplotypes from natural populations of *D. polymorpha*. Abdominal pigmentation is likely to be an ecologically relevant character and has been hypothesized to play roles in sexual selection, crypsis, and/or thermal regulation (KETTLEWELL 1973; GIBERT *et al.* 1998; MAJERUS 1998; TRUE 2003). Consequently, any gene contributing to this character should be constrained or promoted by selection. Our findings from both approaches are consistent with a role for *omb* in the evolution of abdominal pigmentation variation in this species.

MATERIALS AND METHODS

Isolation and sequencing of haplotypes at the *omb* locus: Thirteen different species were used in this study. *D. parthenogenetica*, *D. similis*, *D. nigrodunni*, *D. procardinoides*, *D. cardinoides*, *D. cardini*, *D. dunni*, *D. neomorpha*, and *D. polymorpha* are all members of the cardini group, which is distributed throughout the Caribbean as well as the North, Central, and South American mainlands (HEED and KRISHNAMURTHY 1959; HEED and RUSSELL 1971). *D. parthenogenetica*, *D. nigrodunni*, *D. procardinoides*, *D. melanica*, *D. arawakana*, and *D. dunni* were obtained from the National Drosophila Species Resource Center, Tucson (stock nos. 15181-2221.00, 15182-2311.01, 15181-2241.00, 15030-1141.03, 15182-2261.00, and 15182.2301.00, respectively). *D. cardinoides*, *D. cardini*, *D. neomorpha*, and *D. polymorpha* were obtained from wild-caught collections in Brazil by J. A. Brisson in March through April of 2001 and March through May in 2002. *D. virilis* has been maintained in the laboratory of I. Duncan for a number of years. Sequences from *D. pseudoobscura* and *D. melanogaster* were obtained from their respective genome projects. Genomic DNA was extracted from individual flies using the DNeasy tissue kit (QIAGEN, Valencia, CA). *omb* haplotypes were obtained from all of these species and from 120 individuals of *D. polymorpha* collected from natural populations.

The *omb* locus in *D. melanogaster* consists of eight exons and encodes a mature transcript of 4.1 kb, as depicted in Figure 1 (PFLUGFELDER *et al.* 1992b). We used the *D. melanogaster omb* sequence to design a number of primer pairs. One pair, 370for (5'-CCAGGTTTCGCCAATGGGATG-3') and 1355rev (5'-GGT GATGACCATTTCGGTGC-3'), successfully amplified an ~1-kb N-terminal product in *D. virilis*, *D. melanica*, and *D. polymorpha*, homologous to exon 2 of *D. melanogaster*. This *D. polymorpha* sequence was then used to design the cardini group specific primers ombfor (5'-GACGACCCCATCGTCCTC-3') and ombrev (5'-CACTGACTTCAGCGTCTCGT-3'). Standard PCR using KlenTaq1 (BARNES 1992), which has a very low error rate (5.1×10^{-5} errors/bp), was performed with 30 cycles of 30 sec at 95°, 30 sec at 50°, and 3 min at 72° followed by a 10-

min final extension. For most species, amplified DNA was sequenced directly. However, to perform a more in-depth intraspecific study of the locus in *D. polymorpha* and to allow direct determination of haplotypes, we cloned amplified DNA into the pSTBlue-1 vector using the Acceptor Vector cloning kit (Novagen, Cambridge, MA). Cloned DNA was sequenced on both strands using the vector-specific primers T7 and U19 (Novagen) and the internal primers ombrev2 (5'-ATTCTG GAGCTGGCAATTATCG-3') and ombfor2 (5'-AGACCCG GATGATGCGGTCC-3'). Products of the sequencing reactions were purified using CentriSep spin columns (Princeton Separations) and electrophoresed on a Basestation automated sequencer (MJ Research, Watertown, MA). Divergent sequences were reamplified and sequenced directly to check for PCR artifacts.

The degenerate primers Tboxfor (5'-TGCAGACAAATGTT YCCNCARATG-3') and Tboxrev (5'-AAATGAAACAACYTT YTGCATCCA-3) were designed from an alignment of known *omb* orthologs in GenBank and used to amplify an ~250-bp segment of the *omb* T-domain homologous to portion of exons 3 and 4 in *D. melanogaster*. These exons are separated by a short (65 bp) intron. Touchdown PCR was used from 65° to 41° followed by 28 cycles of 94° for 1 min, 55° for 1 min, and 72° for 2 min. Species other than *D. polymorpha* were sequenced directly using the Tboxfor and Tboxrev primers. Alleles from 10 individuals of *D. polymorpha* were cloned to determine possible haplotypes. All remaining individuals were sequenced directly and haplotype subtraction was used to determine haplotypes (CLARK 1990). This was considered an appropriate technique since there was a high incidence of homozygosity and single-site heterozygosity.

For comparative purposes, we also used a segment of the 6-phosphogluconate dehydrogenase (*pgd*) locus on the X chromosome that encodes an enzyme involved in the pentose-phosphate shunt (CHRISTENSEN and LUCCHESI 1984). We used the exon-based degenerate primers pgdfor (5'-GGAGCCGACTC GCTNGARGAYATG-3') and pgdrev (5'-CGCGGCCTCGTG NCCNCCNGGCAT-3') designed from known *pgd* orthologs in GenBank to amplify an ~600-bp region of an intron in *pgd*. The amplified DNA from females was cloned into the pSTBlue-1 vector using the Acceptor Vector cloning kit and sequenced using the vector-specific primers T7 and U19. Divergent sequences were reamplified and sequenced directly to check for PCR artifacts. DNA from males was amplified and sequenced directly.

Interspecific analyses: For a wide phylogenetic perspective on variation at the *omb* locus, we sequenced the two *omb* regions in *D. melanica*, *D. polymorpha*, and *D. virilis* and obtained the *D. pseudoobscura* and *D. melanogaster omb* sequences from published work. DNA was translated into amino acid sequences for alignment. Sequences were aligned using the program PRETTYBOX (Genetics Computer Group, release 10.0). Isoelectric points were estimated using Seqman 4.00 (DNASTAR, Madison, WI).

To examine *omb* evolution among species of the cardini group, we amplified and sequenced the two *omb* regions for nine species: *D. cardini*, *D. cardinoides*, *D. dunni*, *D. neomorpha*, *D. nigrodunni*, *D. parthenogenetica*, *D. polymorpha*, *D. procardinoides*, and *D. similis*. Nucleotide sequences were aligned by hand. Several areas were unalignable due to microsatellite variation, and these areas were removed prior to analyses, leaving 654 included characters.

Analyses within *D. polymorpha*: For initial investigations, haplotype tree estimation using statistical parsimony (TEMPLETON *et al.* 1992) was implemented using the program TCS v. 1.13 (CLEMENT *et al.* 2000) for the *D. polymorpha* N-terminal *omb* data set culled of microsatellite variation. Following production of a haplotype tree without microsatellites, we attempted

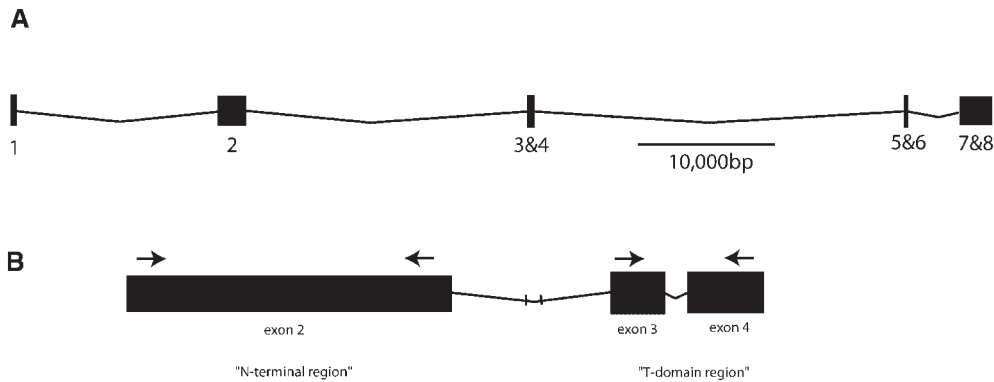


FIGURE 1.—(A) The *omb* locus of *D. melanogaster* showing the distribution of exons (numbered) in boxes and introns as lines. (B) Areas of *omb* used in this study, including portions of the N-terminal domain and the T-domain. Arrows represent locations of primers used for PCR amplification.

to add the microsatellite information in coded form (for example, a microsatellite with four variants was coded as A, G, C, or T). Two microsatellite regions, S_1 and S_2 , were successfully added in this manner to produce a second haplotype tree with more structure. Attempts to add the remaining microsatellites proved to be too ambiguous for reliably constructing haplotype relationships. The program DNAsp v. 3.5 (ROZAS and ROZAS 1999) was used to estimate heterozygosity (Θ) and nucleotide diversity (π) of the culled data set. Heterozygosity is derived from the proportion of segregating sites in the sequence (S; WATTERSON 1975), and nucleotide diversity is the average number of nucleotide differences per site between two sequences (NEI 1987).

Nested clade analysis (NCA) was used to examine the phylogeography of *D. polymorpha* N-terminal *omb* haplotypes (TEMPLETON *et al.* 1995). The haplotype tree was used to define a series of hierarchically nested clades following the nesting rules given in TEMPLETON *et al.* (1987) and TEMPLETON and SING (1993). The program GeoDis 2.0 (POSADA *et al.* 2000) calculated the NCA measures for each clade in each level of nesting. A total of 10,000 random iterations were used to permute the observations within a nesting clade over geography to test the null hypothesis of no association of a haplotype in a cladogram with geographical location, as described in TEMPLETON *et al.* (1995) and TEMPLETON (2004).

Nested clade analysis was also used to examine the correlation between haplotypes and the abdominal pigmentation phenotype. Abdominal pigmentation phenotypes of *D. polymorpha* from wild-caught collections were scored on a scale of one through nine according to a scheme developed specifically for this species (MARTINEZ and CORDEIRO 1970). We used nested clade analysis to look for nonrandom distributions of phenotypic categories across nested clades (TEMPLETON and SING 1993). In this approach, the different levels of nesting define groups of haplotypes. The phenotypes of individuals in each nesting level are then compared to one another via exact permutational tests. Exact permutational tests were also used to test for a nonrandom distribution of microsatellite repeats across phenotypic classes and to test for linkage disequilibrium between microsatellite repeat regions.

Tests of selection: We used DNAsp (ROZAS and ROZAS 1999) to implement a number of tests to determine whether selection has shaped genetic variation at the *omb* locus. Since most variation was in the form of single-site polymorphisms and thus there was little homoplasy in the data set, the infinite sites model was considered appropriate for tests of selection. First, we calculated Tajima's *D* statistic (TAJIMA 1989), a measure of whether the number of segregating sites at a locus is consistent with neutral expectations. Second, we calculated the D^* and F^* statistics of Fu and Li (1993). We also used an intraspecific contingency table test of homogeneity, comparing tip *vs.* interior polymorphisms, to test for selection (TEMPLETON 1996).

The two-by-two contingency table was analyzed using Fisher's exact test.

To investigate selection at the interspecific level, we again used a contingency test of neutrality (TEMPLETON 1996) using data from both the intraspecific *D. polymorpha omb* gene tree and the cardini group interspecific gene tree. This method categorizes silent *vs.* replacement mutations into three classes. The tip and interior classes refer to mutations in an intraspecific data set. The final class, fixed, refers to mutations that differ among species. The table was analyzed using an exact permutational test using the algorithm of ROFF and BENTZEN (1989) permuted randomly 1000 times.

RESULTS

Sequences obtained: Sequences of the N-terminal *omb* region were obtained for eight species of the cardini group and 120 individuals of *D. polymorpha*, as well as for *D. melanica*, *D. melanogaster*, *D. pseudoobscura*, and *D. virilis*. Sequences of the T-domain region were obtained for the same species, but only a subset (37) of *D. polymorpha* individuals were used. The *pgd* locus was sequenced for 88 individuals. These sequences can be found under GenBank accession nos. AY677667–AY677674, AY683324–AY683443, AY676329, AY676330, AY677630–AY677666, and AY687192–AY687279.

Across-species comparisons of *omb*: To determine broad patterns of evolution in both regions of *omb*, we aligned amino acid sequences from *D. melanogaster*, *D. melanica*, *D. pseudoobscura*, *D. polymorpha* and *D. virilis* (Figure 2). These species are representatives of most major clades in the Drosophila phylogeny (REMSEN and O'GRADY 2002). In the N-terminal alignment there is little similarity after the initial 20 amino acids. Nearing the end of this region (around amino acid 251 in Figure 2), amino acid similarity increases again as the sequence approaches the T-domain, which is immediately C-terminal to this region. There is also considerable length variation in the region, as exhibited most dramatically by *D. melanica*, which has 106 fewer amino acids than the longest sequence, that of *D. pseudoobscura* (Table 1). Some of this variation is due to differences in the length of amino acid runs, with asparagine (N) runs being the most prominent. Despite length variation and amino acid changes, the estimated isoelectric point of the

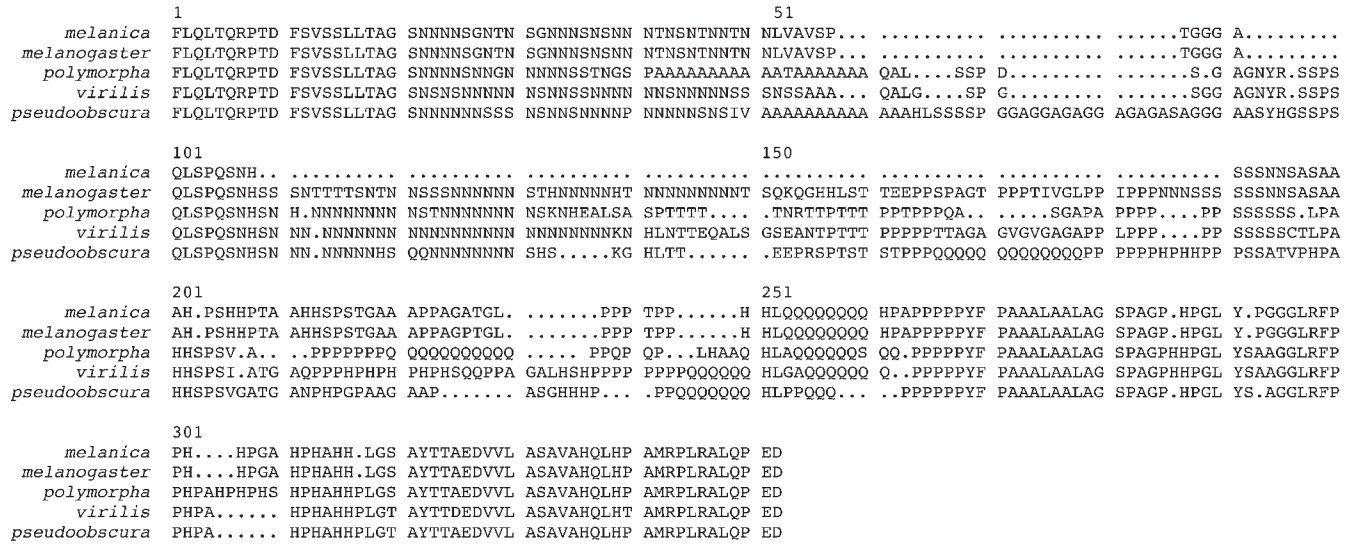


FIGURE 2.—Alignment of N-terminal region *omb* haplotypes from five divergent species of *Drosophila*. Letters are the common amino acid designations, and periods represent gaps introduced for alignment.

N-terminal portion of *omb* remains nearly constant among species (Table 1). The T-box region exhibited no amino acid variation and was well conserved at the nucleotide level as well, with 18 variable sites (of a total 178) among these highly divergent species.

Nucleotide diversity at *omb* within *D. polymorpha*: To analyze nucleotide variation in the *omb* N-terminal region of *D. polymorpha*, we performed separate analyses with a data set culled of microsatellites and another with information from two microsatellites included in coded form. After culling the microsatellites, 22 haplotypes and 23 polymorphic sites remained. To visualize the relationships among these haplotypes, we constructed a haplotype tree (Figure 3A). Haplotypes separated by up to 11 steps had >95% probability of being connected in a parsimonious fashion (TEMPLETON *et al.* 1992). The

network illustrates that the haplotypes are biallelic, falling into two classes separated by six mutational steps. Allele class A has 4 haplotypes, allele class B has 18 haplotypes, and their corresponding measures of nucleotide diversity are reported in Table 2. Twenty-six of the 120 individuals from 6 of the 19 populations exhibited an A-type allele.

In a second analysis, with information from two of the microsatellites included, 26 haplotypes and 25 polymorphic sites were observed. Again, haplotypes separated by up to 11 steps had >95% probability of being connected in a parsimonious fashion. The resulting network, which has more structure than the first, is presented in Figure 3B.

Five of the mutations that separate the two allele classes code for amino acid replacements and their loca-

TABLE 1
Species for which the N-terminal region of *omb* was sequenced

Species	Species group	Number of amino acids	Estimated isoelectric point
<i>D. melanogaster</i>	Melanogaster	280	7.264
<i>D. pseudoobscura</i>	Obscura	304	7.566
<i>D. melanica</i>	Melanica	198	7.438
<i>D. virilis</i>	Virilis	303	7.232
<i>D. ornatipennis</i>	Calloptera	275	7.437
<i>D. polymorpha</i>	Cardini	283	7.878
<i>D. similis</i>	Cardini	274	7.894
<i>D. nigrodunni</i>	Cardini	280	7.894
<i>D. neomorpha</i>	Cardini	272	7.894
<i>D. cardinoides</i>	Cardini	274	7.894
<i>D. cardini</i>	Cardini	288	7.894
<i>D. dunni</i>	Cardini	256	7.878
<i>D. procardinoides</i>	Cardini	284	7.894
<i>D. parthenogenetica</i>	Cardini	273	7.894

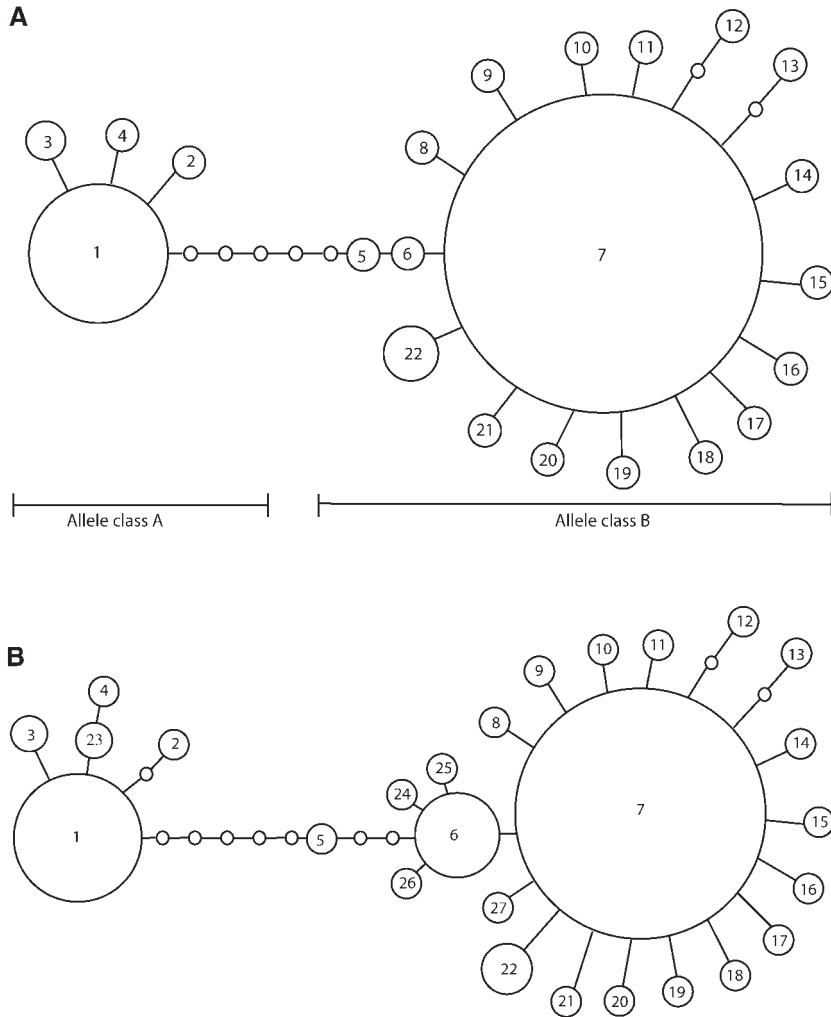


FIGURE 3.—(A) Tree constructed from *omb* N-terminal haplotypes culled of microsatellite repeats and (B) tree constructed from *omb* N-terminal haplotypes including information from two microsatellite repeats. Each circle represents a single haplotype, the circle is sized relative to the number of individuals of each observed haplotype, each line is a single mutational step, and empty circles are unobserved haplotypes.

tions within the sequence are illustrated in Figure 4. They were, for the allele classes A and B, respectively: threonine and isoleucine, proline and alanine, alanine and valine, alanine and serine, and alanine and glycine. One of these changes is conservative (between alanine and valine); two are changes between hydrophilic polar and hydrophobic nonpolar amino acids (threonine and isoleucine, serine and alanine). The alanine to proline

change, although conservative in the sense that both are hydrophobic nonpolar amino acids, is not conservative in the sense that they have very different structures. Finally, the alanine and glycine polymorphism is between a hydrophobic nonpolar amino acid (alanine) and a nearly neutral amino acid (glycine). These changes appear to be in complete linkage with each other such that a change at one site corresponds to a change in

TABLE 2
Measures of *omb* nucleotide variation

	bp	All		Nonsynonymous		Synonymous	
		π	Θ	π	Θ	π	Θ
All	750	0.0042	0.0067	0.0031	0.0020	0.0074	0.0202
Allele class A	750	0.0004	0.0011	0.0000	0.0000	0.0016	0.0041
Allele class B	750	0.0007	0.0050	0.0000	0.0004	0.0024	0.0182
T-box coding	178	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
T-box intron	65	0.0127	0.0058	NA	NA	NA	NA

π (NEI 1987) is the average pairwise difference and Θ is the number of segregating sites per site (WATTERSON 1975). NA, not applicable.

```

SVSLLTAGSNNNNNN--SNNGNNGN--SSSSTNGSPAAAAA---TAAAAAQAALSSPDSGAGNYRSSPSQLSPQSNHSHNNNNNNNNST
SVSLLTAGSNNNNNN---SNNGNNGN---SSS-INGSAAAAA---TAAAAAQAALSSPDSGAGNYRSSPSQLSPQSNHSHNNNNNNNN-ST
SVSLLTAGSNNNNNNNNNNNNNNNNNNNNN--SSSSINGSAAAAA---TAAAAAQAALSSPDSGAGNYRSSPSQLSPQSNHSHNNNNNNNN-ST
SVSLLTAGSNNNN-----SNNGNNGN--SS--TNGSPAAAAA---TAAAAAQAALSSPDSGAGNYRSSPSQLSPQSNHSHNNNNNN--ST
SVSLLTAGSNNNNNN---SNNGNNGN---SSSINGSAAAAA---TAAAAAQAALSSPDSGAGNYRSSPSQLSPQSNHSHNNNNNN--ST
      N1          S1 *      A          N2

NNNNNNNNNN---SKNHEALSASPTTTT-NRTTPTTPTPTPPQASGAPAPPPPPSSSS--LPAHHSVAVPPPPPPQQQQQQQ---PP
NNNNNNNNNNNNNNNNNNNNNNNNNNNN--SNNGNNGN---SSSINGSAAAAA---TAAAAAQAALSSPDSGAGNYRSSPSQLSPQSNHSHNNNNNNNN-ST
NNNNNNNNN-----SKNHEALSASPTTTT-NRTTPTTPTPTPPQASGAPAPPPPPSSSS--LPAHHSVAVPPPPPPQQQQQQQ---PP
NNNNNNNNNN---SKNHEALSASPTTTT-NRTTPTTPTPTPPQASGAPAPPPPPSSSS--LPAHHSVAVPPPPPPQQQQQQQ---PP
NNNNNNNNNN-----SKNHEALSASPTTTT-NRTTPTTPTPTPPQASGAPAPPPPPSSSS--LPAHHSVAVPPPPPPQQQQQQQ---PP
      N3          *      T          S2          Q

QP--LHAAQHLAQQQQQSQPPPPYFPAALAALAGSPAGPHHPGLYSAAGGLRFPHPHAPHSPHSHPHAHHPHLSAYTTAEDVVLASAVAHQ
QP--LHAAQHLAQQQQQQAQQPPPPYFPAALAALAGSPAGPHHPGLYSAAGGLRFPHPGHPHSPHSHPHAHHPHLSAYTTAEDVVLASAVAHQ
QP--LHAAQHLAQQQQQQAQQPPPPYFPAALAALAGSPAGPHHPGLYSAAGGLRFPHPGHPHSPHSHPHAHHPHLSAYTTAEDVVLASAVAHQ
QPPLHAAQHLAQQQQQSQPPPPYFPAALAALAGSPAGPHHPGLYSAAGGLRFPHPHAPHSPHSHPHAHHPHLSAYTTAEDVVLASAVAHQ
QP--LHAAQHLAQQQQQQAQQPPPPYFPAALAALAGSPAGPHHPGLYSAAGGLRFPHPGHPHSPHSHPHAHHPHLSAYTTAEDVVLASAVAHQ
      *
LHPAMRPLRALQPED
LHPAMRPLRALQPED
LHPAMRPLRALQPED
LHPAMRPLRALQPED
LHPAMRPLRALQPED

```

FIGURE 4.—Alignment of five representative *D. polymorpha omb* N-terminal amino acid sequences. Underlined areas are variable microsatellite regions and asterisks mark polymorphic amino acid sites.

all of the other polymorphic sites. As such, all five sites exhibited statistically significant nonrandom associations with each other ($P < 0.001$). The amino acid replacements that characterize the differences between the allele classes are also in linkage disequilibrium with the microsatellite repeats (Figure 5). Only one other amino acid substitution was found in this culled data set, in which a proline was replaced with an alanine.

We used the haplotype trees given in Figure 3 to implement nested clade phylogeographic analyses (TEMPLETON *et al.* 1995) to look for associations between haplotypes and geography. Haplotypes were nested as

illustrated in Figure 6. We found no significant inferences. Therefore, we are unable to reject the null hypothesis of no correlation between *omb* haplotypes and geographical location.

Within the coding region of the T-domain of *omb* (subsets of exons 3 and 4) we observed no nucleotide variation in 37 individuals sampled from 11 different populations. The small intron in this region exhibited two polymorphic sites and one indel of four nucleotides. These data were not considered in further analyses because of the lack of variation.

Correlation of *omb* haplotypes with abdominal pigmentation: Our main focus was on variation at the *omb* locus within a single species of the cardini group, *D. polymorpha*. Our primary goal was to determine whether the polymorphism in abdominal pigmentation in this species is correlated with variation at the *omb* locus. We used nested clade analysis to define haplotype groups to test for associations with the pigmentation categories, with the nesting design illustrated in Figure 6 for both haplotype trees, one excluding and one including some microsatellite information. The genotype and phenotype matrix is available at <http://www.genetics.org/supplemental>. An association between pigmentation and haplotype group was found in both trees. The data set culled of microsatellites had two nesting levels, with homogeneity rejected at the one-step level ($P = 0.019$) and the two-step level ($P < 0.001$). Similarly, the data set including information from two microsatellite loci had three nesting levels, with homogeneity rejected at the two-step level ($P = 0.003$) and the three-step level ($P = 0.003$). The association we find is due to an excessive number of lightly pigmented individuals (phenotypes one to three) in haplotype class A and a relatively higher proportion of medium (phenotypes four through six) and dark (phenotypes seven through nine) pigmented individuals in haplotype class B (Table 3). We

	N ₁	S ₁	A	N ₂	N ₃	T	S ₂	Q	AA
N ₁	-								
S ₁	**	-							
A	**	**	-						
N ₂	**	**	**	-					
N ₃	**	**	**	**	-				
T	ns	ns	ns	ns	ns	-			
S ₂	**	**	**	**	**	ns	-		
Q	ns	**	**	**	**	ns	*	-	
AA	**	**	**	**	**	ns	**	**	-

FIGURE 5.—Linkage disequilibrium between microsatellite runs in the N-terminal region of *omb* in *D. polymorpha*. “AA” refers to the amino acids that characterize the distinction between allele classes A and B. The association between pairs marked with two asterisks is $P < 0.01$, those with one asterisk are $0.01 < P < 0.05$, and “ns” indicates no association.

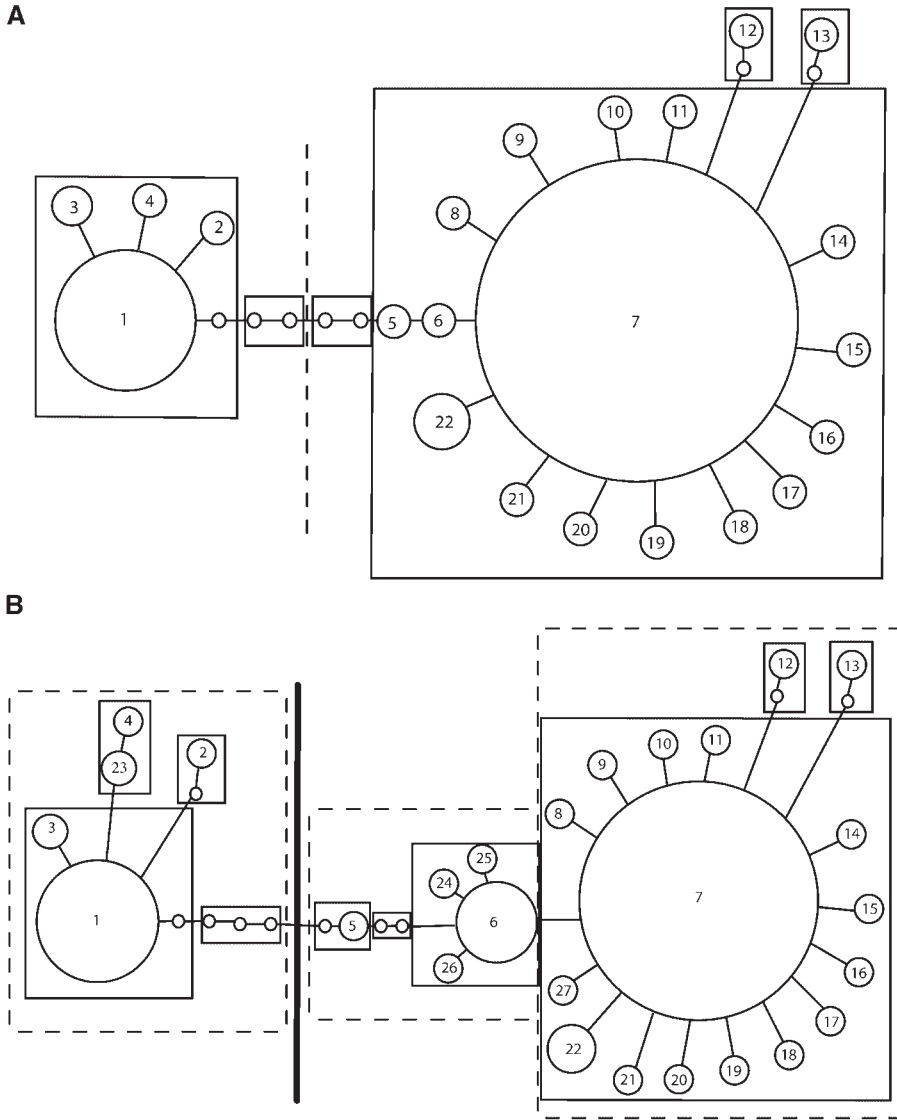


FIGURE 6.—Nesting design used for the nested clade analyses for the haplotype tree created from the *omb* N-terminal sequences (A) culled of microsatellites and (B) with information from two microsatellites. Solid lines outline one-step clades. Broken lines outline two-step clades. The heavy dark line separates three-step clades.

therefore conclude that there is an association between phenotypic categories and allele classes at the N-terminal region of the *omb* locus. In contrast, there was no association between nested haplotypes in the *pgd* tree and the pigmentation phenotypes (data not shown). Further, haplotypes at the *pgd* locus did not fall into two allelic classes.

The N-terminal region of *omb* is quite variable in length within *D. polymorpha* due to the presence of eight variable microsatellite runs. These runs resulted in length variation among individuals, ranging from 276 to 287 amino acids. Their distribution over the sequence is depicted in Figure 4, and the variation in the number of repeats at each run is reported in Table 4. Because the majority of these runs contained degenerate coding, we counted repeat number at the amino acid level.

We looked for an association between microsatellite repeat numbers and the pigmentation phenotypes of *D. polymorpha*. When all nine pigmentation categories are considered in the analysis, three of the eight micro-

satellite repeats at *omb* (N_1 , S_1 , and A) show a significant association with pigmentation. To increase statistical power, we collapsed the nine phenotypic categories into three categories of light (phenotypes one through three), medium (phenotypes four through six), and dark (phenotypes seven through nine), a common grouping of phenotypes for this species (DA CUNHA 1949; MACHADO *et al.* 2001). After this grouping, seven of the eight microsatellite repeats are associated nonran-

TABLE 3

Distribution of pigmentation phenotypes across the two allele classes

	Allele class A	Allele class B
Light	8	4
Medium	4	24
Dark	14	66

TABLE 4

Amino acid repeats in the N-terminal region of *omb* and number of individuals of each pigmentation phenotype bearing them

No. of repeats	N ₁ *			S ₁ **		A*		N ₂ **		N ₃ *				T		S ₂ **		Q*	
	4, 5	6, 7	8–12	2	4, 5	8–10	11–12	6, 7	8–10	5, 7, 8	9	10	11–14	4, 5	6, 7	4, 5	6, 7	5–8	9–12
Light	8	3	1	8	4	6	6	5	7	7	2	0	3	12	0	4	8	5	7
Medium	8	15	5	5	23	22	6	24	4	5	7	10	6	26	2	23	5	23	5
Dark	19	39	22	15	65	66	14	64	16	18	27	18	17	71	9	56	24	51	29

Amino acids are asparagine (N), serine (S), alanine (A), threonine (T), and glutamine (Q). Numbered subscripts designate distinct runs. Asterisks denote significant associations with the pigmentation phenotype of $P < 0.01$ (**) or $0.01 < P < 0.05$ (*). For some microsatellites, repeat categories were combined to avoid groups with zero entries.

domly with the three pigmentation categories (Table 4). Interestingly, the majority of the microsatellite repeats are in linkage disequilibrium with one another (Figure 5), indicating that the entire N-terminal region of *omb* is associated with the pigmentation phenotype, and therefore these tests are not independent.

We also found a microsatellite repeat at the *pgd* locus. This locus was sequenced for 88 individuals, a data set that mostly overlapped with the *omb* data set in terms of individuals used. In an additional 20 individuals, the sequence quality was inadequate for haplotype information, but sufficient for scoring microsatellite repeat number. This microsatellite repeat shows no association with the pigmentation phenotypes (Table 5).

Deviations from neutrality: excess of rare alleles and intraspecific replacement polymorphisms: We looked for deviations from neutrality in the N-terminal region of *omb* with the *D. polymorpha* data using TAJIMA's (1989) and FU and LI's (1993) tests. All three test statistics were negative (Tajima's $D = -1.09$, $P > 0.10$; $D^* = -4.47$, $P < 0.02$, $F^* = -3.77$, $P < 0.02$), indicating a lack of polymorphism except in the form of rare alleles (despite the presence of two major allele classes). We constructed a two-by-two contingency table by counting the number of replacement and synonymous mutational events on the tip and interior branches of the haplotype tree shown in Figure 3A. The contingency table is shown in Table 6, with the null hypothesis of homogeneity strongly rejected. This result indicates balancing or di-

versifying selection, reflecting the inordinately large number of replacement amino acid changes on interior branches of the haplotype tree.

We also performed a contingency test of neutrality (TEMPLETON 1996) using the *D. polymorpha* intraspecific gene tree in combination with the interspecific cardini gene tree. This, too, was significant (Table 6) in rejecting the null hypothesis of homogeneity. Neutral theory would predict the same proportion of nonsynonymous to synonymous mutations within and between species. However, replacements were more common in the interior than the tip classes within *D. polymorpha* and more common among species than among tip haplotypes within *D. polymorpha*.

DISCUSSION

Variation at the *omb* locus: At the distantly related species level (*i.e.*, across most of the *Drosophila* phylogeny) nucleotide variation is so high in the N-terminal region of *omb* that it is unalignable. The amino acid sequences encoded also have few alignable regions. However, it is notable that despite this large amount of variation, two domains, at the beginning and end of the sequence considered here, are nearly perfectly conserved and therefore likely to be of functional importance. A transcription repression domain lies upstream of the T-domain of *Tbx2*, the mouse homolog of *omb* (PAXTON *et al.* 2002). Whether or not a similar repression domain exists in *omb* will require functional studies.

TABLE 5

Amino acid repeats in a microsatellite at the *pgd* locus and number of individuals of each pigmentation phenotype bearing them

No. of repeats	<i>pgd</i> microsatellite			
	5, 6	7	8	9, 10, 11
Light	0	4	5	1
Medium	0	4	7	3
Dark	5	22	32	25

Exact permutational test: $P = 0.652$.

TABLE 6

Contingency analysis of silent vs. replacement mutations

	Silent	Replacement
Tip (intraspecific)	14	1
Interior (intraspecific)	2	5
Fixed (interspecific)	67	17

Permutational probability under the null hypothesis of homogeneity for intraspecific data is 0.003. Permutational probability under the null hypothesis of homogeneity for all data is < 0.001 .

Despite showing extensive amino acid and length variation, the remaining portion of the N-terminal region studied retains similar estimated isoelectric points. In marked contrast, there was very little variation among distantly related species in the T-domain. All nucleotide variation in this region produced synonymous codons. This result is not surprising, given the functional significance of this region (*i.e.*, its role in DNA binding).

The N-terminal domain contains within it a number of microsatellite runs. Microsatellite runs are common in the coding regions of important developmental genes, especially in the motifs flanking the DNA-binding domain (SUZUKI 1989). KARLIN *et al.* (2002) report that >80% of *Drosophila* proteins with multiple runs seem to function in the regulation of transcription. Among these are such well-known genes as *Ultrabithorax*, *period*, *mastermind*, and *nonA* (BEACHY *et al.* 1985; COLOT *et al.* 1988; NEWFELD *et al.* 1993; PEIXOTO *et al.* 1993; CAMPE-SAN *et al.* 2001). Further, KARLIN *et al.* (2002) report that alanine, serine, and glutamine account for a significant proportion of the runs. At the *omb* locus, we see runs of these three amino acids as well as a threonine and three asparagine runs. These runs are in close proximity to the DNA-binding region, the T-domain, as noted by PFLUGFELDER *et al.* (1992a,b) in their analysis of the *omb* locus in *D. melanogaster*. Length variation of these homopolymers may cause differences in structure that allow plasticity in protein-protein interactions (NEWFELD *et al.* 1991). The multiple asparagine (N) runs are particularly interesting, since N runs avoid the secondary structures of α -helices and β -strands and may provide binding sites for protein-protein interactions (BRENDDEL and KARLIN 1989; KARLIN *et al.* 2002).

Correlation between *omb* haplotypes and abdominal pigmentation: As previously mentioned, genetic studies have led to the expectation that *omb* might contribute to variation in the width of the abdominal pigment band among individuals of *Drosophila* (KOPP and DUNCAN 1997). Indeed, here we report a significant association between the A and B haplotype classes and other haplotype clades and abdominal pigmentation variation. We also report a significant association between microsatellite variation within the N-terminal *omb* coding region and the pigmentation phenotype. Seven of eight variable microsatellites in this region had repeat numbers that were nonrandomly distributed across the three phenotypic categories of light, medium, and dark pigmentation. These correlations indicate that *omb* (or a regulatory region of *omb* or another locus in linkage disequilibrium with this region of *omb*) may play a role in natural variation in this trait. Further, because *omb* haplotypes did not correlate with geography as determined by phylogeographic nested clade analysis, the correlation observed here is not a by-product of geographic sampling.

The association between *omb* haplotypes and pigmentation variation would be greatly strengthened by link-

age information from *omb* haplotypes segregating in an F₂ generation. Unfortunately, the only lines we have in the laboratory are light lines that have a dominant light allele and a single dark line. These light lines were established from flies collected in the northern part of *D. polymorpha*'s range, the only region in which the dominant light allele is known to exist (HEED and BLAKE 1963). Therefore, these are not appropriate for examining the polymorphism that we have detected in flies from southern Brazil, where the dominant light allele is not present. And, any results from these types of analyses would not be able to illuminate whether *omb* is the major locus underlying pigmentation variation for this species (DA CUNHA 1949) or a modifier locus (MARTINEZ and CORDEIRO 1970), because a cross between a light line with a dominant light allele and a dark line would produce only light or dark flies, never flies with intermediate levels of pigmentation. We therefore must estimate the amount of phenotypic variation explained by *omb* haplotypes on the basis of the data presented here. The average phenotype of an individual bearing an A-class allele is 6.0, while the average phenotype of an individual bearing a B-class allele is 6.9. The phenotypic differences between the two classes are thus on the order of one phenotypic class, suggesting *omb* may be a modifier locus.

Much recent effort has been applied to determining the number and identity of loci contributing to abdominal pigmentation variation either within or between species of *Drosophila* (KOPP *et al.* 2000, 2003; LLOPART *et al.* 2002; WITTKOPP *et al.* 2002, 2003b). One study specifically examined the potential contribution of *omb* to pigmentation evolution. WITTKOPP *et al.* (2003b) used markers within candidate genes to look for associations with pigmentation divergence between *D. americana* and *D. novamexicana*. They found that a marker within the *omb* gene was not linked to any QTL for this trait. However, these species differ primarily in the degree of melanization of each segment (*D. americana* has a uniformly darkly melanized abdominal cuticle, and *D. novamexicana* has a uniformly lightly melanized abdominal cuticle) and not in the width of the pigment band. Thus, *omb* is not expected to be involved in differences between these species. However, *omb* may be important in determining pigmentation differences between *D. yakuba* (pigmented like *D. melanogaster*) and *D. santomea* (completely unpigmented), as LLOPART *et al.* (2002) found that the X chromosome (*omb* is most likely on the X in the melanogaster species group) accounts for nearly 90% of the pigmentation differences between these species.

Biallelism at the *omb* locus: Variation at *omb* is best described as biallelism, with two main haplotype clusters (Figure 3), referred to here as allele class A and allele class B. Allele class A is associated with very little variation (fewer haplotypes) compared to allele class B. These haplotype counts argue that allele class B is the

older of the two classes. On the other hand, the amino acid states found in allele class A are identical to those found in the majority of the other species of the cardini group examined here. Thus, if placed within a phylogenetic context, it is apparent that allele class A is the ancestral state. However, the wide geographic spread (all populations sampled) and haplotype diversification (Figure 3) observed here for allele class B indicate that it has been in existence for some time. In fact, it was the only allele class found in 13 of the 19 populations, although the sparse sampling makes it possible that both alleles are found in the majority of the populations. It is not known whether the replacement substitutions that separate the two allele classes result in a functional change in the *omb* protein or whether they are in linkage disequilibrium with other changes in either the regulatory regions or the protein-coding region that are in some manner selectively advantageous.

Because molecular variation can be shaped by both demographic and selective processes, it is important to consider the demographic history of *D. polymorpha*. Certain scenarios could explain the biallelic structure observed in the N-terminal domain of *omb*. First, if *D. polymorpha* is fragmented into several smaller populations, genetic drift can cause populations to diverge from one another, resulting in a situation with very little polymorphism within populations compared to polymorphisms among populations. Second, the breakdown of a previous barrier between divergent populations or introgression from a related species may result in biallelism.

Insight into these scenarios is aided by previous work on the phylogeography of the species. *D. polymorpha* is a widespread species found throughout much of South America in a wide diversity of habitats and is therefore not prone to inbred, fragmented populations (HEED 1963; SENE *et al.* 1980; TIDON-SKLORZ and SENE 1992; ROHDE 1996b). The only structure among populations of the species from south and central Brazil (the same individuals used here), as determined by both a mitochondrial locus (*cytochrome b* and flanking regions) and a nuclear locus (*pgd*), is that of restricted gene flow via isolation by distance and one inference of a contiguous range expansion (J. A. BRISSON, unpublished data). Therefore, we can exclude population fragmentation as a causative factor in the maintenance of biallelism at this locus. Further, *omb* haplotypes from all the closely related species to *D. polymorpha* displayed nucleotide differences significantly different from *D. polymorpha* haplotypes, making it unlikely that introgression from a related species is contributing to variation at the *omb* locus in *D. polymorpha*.

Finally, a biallelic structure can be maintained by a chromosomal inversion. Indeed, previous work on *D. polymorpha* indicates that inversion polymorphisms are widespread in this species (ROHDE 1996a; DE TONI *et al.* 2001). To test whether the A or B alleles are associ-

ated with inversions, we created heterozygotes for allele classes A and B and examined their chromosomal morphology. Inversions were observed on the X chromosomes, but not on the autosomes (data not shown). In our analysis, we have determined that *omb* is not on the X chromosome in *D. polymorpha* (wild-caught males display two unique haplotypes), as it is in *D. melanogaster*. Therefore, we conclude that an inversion polymorphism is not responsible for maintenance of the two haplotype classes of *omb* in *D. polymorpha*.

Evidence for selection at the *omb* locus: Because *omb* is a candidate for abdominal pigmentation variation, we looked for the signature of selection at this locus within a single species, *D. polymorpha*. For TAJIMA's (1989) and Fu and Li's (1993) tests, we observed the pattern of a significantly large number of singleton mutations (rare haplotypes) in comparison to the neutral expectation. This pattern implies that a diversity-reducing event occurred in the past. In contrast, a contingency test of neutrality indicated balancing selection was acting to preserve both haplotype classes. Therefore, the history of selection at this locus within *D. polymorpha* appears to be selective sweeps within each allele class, while both classes are maintained by balancing selection.

The source of the balancing selection at *omb* within *D. polymorpha*, if it is indeed involved in the production of the pigmentation phenotype, could be the habitat of these generalist fruit flies. In southern Brazil, populations of *D. polymorpha* found in forested habitats have, on average, lighter pigmented abdomens than populations found in open, drier habitats (J. A. BRISSON, unpublished data). Therefore, alternative alleles could be selected in different environments and thus in different populations, maintaining a balanced polymorphism across the species.

The contingency test also revealed a higher number of fixed (interspecific) amino acid replacements relative to intraspecific tip amino acid replacements, suggesting that directional selection is favoring amino acid changes among species. These results, combined with the intraspecific selection tests, suggest the following picture: within *D. polymorpha* nucleotide replacement variation at *omb* has occurred with some frequency, but has been reduced by purifying selection within each allele class while both allele classes are maintained by balancing selection. In contrast, interspecific comparisons suggest that directional selection has been responsible for fixing differences between species. This behavior is consistent with that expected for a gene underlying an ecologically relevant trait such as abdominal pigmentation. A closer examination of *omb* variation in different species of the cardini group will elucidate the history of selection at this locus.

In the T-domain, the lack of amino acid variation among divergent species indicates purifying selection. Within *D. polymorpha*, we observed no nucleotide variation at all in the T-domain coding region. This total

lack of nucleotide polymorphism could be indicative of codon bias or functional constraint at the nucleotide level.

We thank two anonymous reviewers for their helpful comments on this work. J.A.B. is a Howard Hughes Medical Institute Predoctoral Fellow. This work was supported by National Science Foundation Doctoral Dissertation Improvement grant 0104977 to J.A.B.

LITERATURE CITED

- BARNES, W. M., 1992 The fidelity of *Taq* polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* **112**: 29–35.
- BEACHY, P., S. HELFAND and D. HOGNESS, 1985 Segmental distribution of *bithorax* complex proteins during *Drosophila* development. *Nature* **313**: 545–551.
- BOLLAG, R. J., Z. SIEGFRIED, J. A. CEBRATHOMAS, N. GARVEY, E. M. DAVIDSON *et al.*, 1994 An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T-locus. *Nat. Genet.* **7**: 383–389.
- BRENDEL, V., and S. KARLIN, 1989 Association of charge clusters with functional domains of cellular transcription factors. *Proc. Natl. Acad. Sci. USA* **86**: 5698–5702.
- BRUNNER, A., R. WOLF, G. O. PFLUGFELDER, B. POECK and M. HEISENBERG, 1992 Mutations in the proximal region of the *optomotor-blind* locus of *Drosophila melanogaster* reveal a gradient of neuroanatomical and behavioral phenotypes. *J. Neurogenet.* **8**: 43–50.
- CAMPESAN, S., D. CHALMERS, F. SANDRELLI, A. MEGIGHIAN, A. A. PEIXOTO *et al.*, 2001 Comparative analysis of the *nonA* region in *Drosophila* identifies a highly diverged 5' gene that may constrain *nonA* promoter evolution. *Genetics* **157**: 751–764.
- CARROLL, S. B., J. K. GRENIER and S. D. WEATHERBEE, 2001 *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Scientific, Oxford.
- CHRISTENSEN, A. C., and J. C. LUCCHESI, 1984 Cloning of the *Drosophila 6-phosphogluconate dehydrogenase* gene *pgd*. *Genetics* **107**: S20.
- CLARK, A. G., 1990 Inference of haplotypes from PCR amplified samples of diploid populations. *Mol. Biol. Evol.* **7**: 111–122.
- CLEMENT, M., D. POSADA and K. A. GRANDALL, 2000 TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**: 1657–1660.
- COLOT, H. W., J. C. HALL and M. ROSHBASH, 1988 Interspecific comparison of the *period* gene of *Drosophila* reveals large blocks of nonconserved coding DNA. *EMBO J.* **7**: 3929–3937.
- DA CUNHA, A. B., 1949 Genetic analysis of the polymorphism of color pattern in *Drosophila polymorpha*. *Evolution* **3**: 239–251.
- DAVID, J. R., P. CAPY and J. P. GAUTHIER, 1990 Abdominal pigmentation and growth temperature in *Drosophila melanogaster*, similarities and differences in the norms of reaction of successive segments. *J. Evol. Biol.* **3**: 429–445.
- DE TONI, D. C., F. DE OLIVEIRA and V. L. S. VALENTE, 2001 Chromosomal variability of *Drosophila polymorpha* populations from Atlantic Forest remnants of continental and insular environments in the State of Santa Catarina, Brazil. *Caryologia* **54**: 329–337.
- FU, Y. X., and W. H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- GARCIA-BELLIDO, A., 1983 Comparative anatomy of cuticular patterns in the genus *Drosophila*, pp. 227–255 in *Development and Evolution: Sixth Symposium of the British Society for Developmental Biology*, edited by B. C. GOODWIN, N. HOLDER and C. C. WYLIE. Cambridge University Press, Cambridge/New York.
- GIBERT, P., B. MORETEAU, J. C. MORETEAU and J. R. DAVID, 1996 Growth temperature and adult pigmentation in two *Drosophila* sibling species: An adaptive convergence of reaction norms in sympatric populations? *Evolution* **50**: 2346–2353.
- GIBERT, P., B. MORETEAU, J. C. MORETEAU, R. PARKASH and J. R. DAVID, 1998 Light body pigmentation in Indian *Drosophila melanogaster*: a likely adaptation to a hot and arid climate. *J. Genet.* **77**: 13–20.
- GOMPEL, N., and S. B. CARROLL, 2003 Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. *Nature* **424**: 931–935.
- GRIMM, S., and G. O. PFLUGFELDER, 1996 Control of the gene *optomotor-blind* in *Drosophila* wing development by *decapentaplegic* and *wingless*. *Science* **271**: 1601–1604.
- HEED, W. B., 1963 Density and distribution of *Drosophila polymorpha* and its color alleles in South America. *Evolution* **17**: 502–518.
- HEED, W. B., and P. R. BLAKE, 1963 A new color allele at the e locus of *Drosophila polymorpha* from Northern South America. *Genetics* **48**: 217–234.
- HEED, W. B., and N. B. KRISHNAMURTHY, 1959 Genetic studies on the cardini group of *Drosophila* in the West Indies. *Univ. Texas Pub.* **5914**: 155–179.
- HEED, W. B., and J. S. RUSSELL, 1971 Phylogeny and population structure in island and continental species of the *cardini* group of *Drosophila* studied by inversion analysis. *Univ. Texas Pub.* **7103**: 91–130.
- HOLLOCHER, H., J. L. HATCHER and E. G. DYRESON, 2000 Evolution of abdominal pigmentation differences across species in the *Drosophila dunni* subgroup. *Evolution* **54**: 2046–2056.
- KARLIN, S., L. BROCCIERI, A. BERGMAN, J. MRAZEK and A. J. GENTLES, 2002 Amino acid runs in eukaryotic proteomes and disease associations. *Proc. Natl. Acad. Sci. USA* **99**: 333–338.
- KETTLEWELL, B., 1973 *The Evolution of Melanism: The Study of a Recurring Necessity; With Species Reference to Industrial Melanism in the Lepidoptera*. Clarendon Press, Oxford.
- KISPERT, A., and B. G. HERRMANN, 1993 The *brachyury* gene encodes a novel DNA-binding protein. *EMBO J.* **12**: 3211–3220.
- KISPERT, A., B. KOSCHORZ and B. G. HERRMANN, 1995 The T-protein encoded by *brachyury* is a tissue-specific transcription factor. *EMBO J.* **14**: 4763–4772.
- KOPP, A., and I. DUNCAN, 1997 Control of cell fate and polarity in the adult abdominal segments of *Drosophila* by *optomotor-blind*. *Development* **124**: 3715–3726.
- KOPP, A., and I. DUNCAN, 2002 Anteroposterior patterning in adult abdominal segments of *Drosophila*. *Dev. Biol.* **242**: 15–30.
- KOPP, A., I. DUNCAN and S. B. CARROLL, 2000 Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* **408**: 553–559.
- KOPP, A., R. M. GRAZE, S. Z. XU, S. B. CARROLL and S. V. NUZHIDIN, 2003 Quantitative trait loci responsible for variation in sexually dimorphic traits in *Drosophila melanogaster*. *Genetics* **163**: 771–787.
- LECUIT, T., W. J. BROOK, M. NG, M. CALLEJA, H. SUN *et al.*, 1996 Two distinct mechanisms for long-range patterning by *decapentaplegic* in the *Drosophila* wing. *Nature* **381**: 387–393.
- LEE, T. J., 1963 Genetic analysis of the polymorphism of color pattern in *Drosophila auraria*. *Dros. Inf. Serv.* **37**: 97–98.
- LLOPART, A., S. ELWYN, D. LACHAISE and J. A. COYNE, 2002 Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. *Evolution* **56**: 2262–2277.
- MACHADO, M. X., D. C. DE TONI and P. R. P. HOFMANN, 2001 Abdominal pigmentation polymorphism of *Drosophila polymorpha* (Dobzhansky and Pavan, 1943) collected on Ilha de Santa Catarina and neighboring islands. *Biotemas* **14**: 87–107.
- MAJERUS, M. E. N., 1998 *Melanism: Evolution in Action*. Oxford University Press, Oxford.
- MARTINEZ, M. N., and A. R. CORDEIRO, 1970 Modifiers of color pattern genes in *Drosophila polymorpha*. *Genetics* **64**: 573–587.
- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- NELLEN, D., R. BURKE, G. STRUHL and K. BASLER, 1996 Direct and long-range action of a DPP morphogen gradient. *Cell* **85**: 357–368.
- NEWFELD, S. J., D. A. SMOLLER and B. YEDVOBNICK, 1991 Interspecific comparison of the unusually repetitive *Drosophila* locus *mastermind*. *J. Mol. Evol.* **32**: 415–420.
- NEWFELD, S. J., A. T. SCHMID and B. YEDVOBNICK, 1993 Homopolymer length variation in the *Drosophila* gene *mastermind*. *J. Mol. Evol.* **37**: 483–495.
- OHNISHI, S. A. T. K. W., 1985 Genetic analysis of color dimorphism in the *Drosophila montium* subgroup. *Jpn. J. Genet.* **60**: 355–358.
- PAXTON, C., H. H. ZHAO, Y. CHIN, K. LANGNER and J. REECY, 2002 Murine Tbx2 contains domains that activate and repress gene transcription. *Gene* **283**: 117–124.
- PEIXOTO, A. A., S. CAMPESAN, R. COSTA and C. P. KYRIACOU, 1993 Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. *Mol. Biol. Evol.* **10**: 127–139.
- PFLUGFELDER, G. O., H. ROTH and B. POECK, 1992a A homology domain shared between *Drosophila optomotor blind* and mouse

- brachyury* is involved in DNA binding. *Biochem. Biophys. Res. Commun.* **186**: 918–925.
- PFLUGFELDER, G. O., H. ROTH, B. POECK, S. KERSCHER, H. SCHWARZ *et al.*, 1992b The lethal(1) *optomotor-blind* gene of *Drosophila melanogaster* is a major organizer of optic lobe development: isolation and characterization of the gene. *Proc. Natl. Acad. Sci. USA* **89**: 1199–1203.
- POECK, B., A. HOFBAUER and G. O. PFLUGFELDER, 1993 Expression of the *Drosophila optomotor-blind* gene transcript in neuronal and glial-cells of the developing nervous system. *Development* **117**: 1017–1029.
- POSADA, D., K. A. CRANDALL and A. R. TEMPLETON, 2000 GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* **9**: 487–488.
- REMSEN, J., and P. O'GRADY, 2002 Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. *Mol. Phylogenet. Evol.* **24**: 249–264.
- ROBERTSON, A., D. A. BRISCOE and J. H. LOUW, 1977 Variation in abdomen pigmentation in *Drosophila melanogaster* females. *Genetica* **47**: 73–76.
- ROFF, D. A., and P. BENTZEN, 1989 The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. *Mol. Biol. Evol.* **6**: 539–545.
- ROHDE, C. A. V. L. S. V., 1996a Cytological maps and chromosomal polymorphism of *Drosophila polymorpha* and *Drosophila cardinoides*. *Braz. J. Genet.* **19**: 27–32.
- ROHDE, C. A. V. L. S. V., 1996b Ecological characteristics of urban populations of *Drosophila polymorpha* Dobzhansky & Pavan and *Drosophila cardinoides* Dobzhansky & Pavan (Diptera, Drosophilidae). *Revta Bras. Ent.* **40**: 75–79.
- ROZAS, J., and R. ROZAS, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SENE, F. M., F. C. VAL, C. R. VILELA and M. A. Q. R. PEREIRA, 1980 Preliminary data on the geographical distribution of *Drosophila* species within morphoclimatic domains of Brazil. *Pap. Avulsos Dep. Zool. Sec. Agric.* **33**: 315–326.
- STALKER, H. D., 1953 Taxonomy and hybridization in the cardini group of *Drosophila*. *Ann. Entomol. Soc. Am.* **46**: 343–358.
- SUZUKI, M., 1989 Spxx, a frequent sequence motif in gene regulatory proteins. *J. Mol. Biol.* **207**: 61–83.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- TEMPLETON, A. R., 1996 Contingency tests of neutrality using intra-/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the hominoid primates. *Genetics* **144**: 1263–1270.
- TEMPLETON, A. R., 2004 Statistical phylogeography: methods of evaluating and minimizing inference errors. *Mol. Ecol.* **13**: 789–809.
- TEMPLETON, A. R., and C. F. SING, 1993 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134**: 659–669.
- TEMPLETON, A. R., E. BOERWINKLE and F. SING, 1987 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117**: 343–351.
- TEMPLETON, A. R., K. A. CRANDALL and C. F. SING, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. 3. Cladogram estimation. *Genetics* **132**: 619–633.
- TEMPLETON, A. R., E. ROUTMAN and C. A. PHILLIPS, 1995 Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrinum*. *Genetics* **140**: 767–782.
- TIDON-SKLORZ, R., and F. D. M. SENE, 1992 Vertical and temporal distribution of *Drosophila* (Diptera, Drosophilidae) species in a wooded area in the state of Sao Paulo, Brazil. *Rev. Brasil. Biol.* **52**: 311–317.
- TRUE, J. R., 2003 Insect melanism: the molecules matter. *Trends Ecol. Evol.* **18**: 640–647.
- WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–275.
- WITTKOPP, P. J., K. VACCARO and S. B. CARROLL, 2002 Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* **12**: 1547–1556.
- WITTKOPP, P. J., S. B. CARROLL and A. KOPP, 2003a Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet.* **19**: 495–504.
- WITTKOPP, P. J., B. L. WILLIAMS, J. E. SELEGUE and S. B. CARROLL, 2003b *Drosophila* pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proc. Natl. Acad. Sci. USA* **100**: 1808–1813.

Communicating editor: G. GIBSON