J Mol Evol (1990) 31:402–423

Journal of
Molecular Evolution

© Springer-Verlag New York Inc. 1990

Evolution of the Autosomal Chorion Cluster in *Drosophila*. IV. The Hawaiian *Drosophila*: Rapid Protein Evolution and Constancy in the Rate of DNA Divergence

Juan Carlos Martínez-Cruzado*

Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA

Summary. Autosomal chorion genes *s18*, *s15*, and *s19* are shown to diverge at extremely rapid rates in closely related taxa of Hawaiian *Drosophila*. Their nucleotide divergence rates are at least as fast as those of intergenic regions that are known to evolve more extensively between distantly related species. Their amino acid divergence rates are the fastest known to date. There are two nucleotide replacement substitutions for every synonymous one. The molecular basis for observed length and substitution mutations is analyzed. Length mutations are strongly associated with direct repeats in general, and with tandem repeats in particular, whereas the rate for an average transition is twice that for an average transversion.

The DNA sequence of the cluster was used to construct a phylogenetic tree for five taxa of the Hawaiian picture-winged species group of *Drosophila*. Assignment of observed base substitutions occurring in various branches of the tree reveals an excess of would-be homoplasies in a centrally localized 1.8-kb segment containing the *s15* gene. This observation may be a reflection of ancestral excess polymorphisms in the segment. The chorion cluster appears to evolve at a constant rate regardless of whether the central 1.8-kb segment is included or not in the analysis. Assuming that the time of divergence of *Drosophila grimshawi* and the *planitibia* subgroup coincides with the emergence of the island of Kauai, the overall rate of base substitution in the cluster is estimated to be 0.8%/million years, whereas synonymous sites are substituted at a rate of 1.2%/million years.

Key words: Chorion gene cluster — Homomorphism — Excess polymorphism — Comparative evolutionary rates — Molecular clock — Transitions and transversions

Introduction

The molecular clock hypothesis states that the rate of evolutionary divergence of any genetic unit remains constant over time. However, this rate of divergence is naturally dependent on mutation rate, and may depend on generation length or on the average number of replication rounds for any stem cell in germ lines of different lineages, as well as on the properties of replicases of different organisms (Jukes and Bhushan 1986) or organelles (Wilson et al. 1985). The evolutionary rate is also affected by the selective constraints specific to each genetic unit and may thus vary between genes (Dickerson 1971; Ticher and Graur 1989), intragenic functional units (Wilson et al. 1977; Hayashida and Miyata 1983), or even synonymous sites (Riley 1989; Ticher and Graur 1989).

It is clear then that, even if the molecular clock hypothesis holds, estimates of divergence times using molecular data should preferably be limited to phyletic groups with comparable mutational dynamics, and that the genetic units used must each be employed with its own previously determined divergence rate constant. Furthermore, the genetic

* *Current address:* Departamento de Biología, Recinto Universitario de Mayagüez, Universidad de Puerto Rico, Mayagüez, Puerto Rico 00708, USA
Offprint requests to: J.C. Martínez-Cruzado

units of choice must be appropriate for the divergence times of the studied phyletic groups: rapidly diverging units could reach saturation in less time than that corresponding to the divergence of distantly related phyletic groups (Kaplan and Langley 1979; Nei and Tajima 1983), whereas slowly diverging units would give no information for closely related phyletic group comparisons.

Hence, for the purpose of estimating divergence times it is essential to establish the constancy of the evolutionary rate of various genetic units in different phyletic groups as well as their particular rate constants. The picture-winged group of Hawaiian *Drosophila* is an ideal group in which to undertake this project. The high rate of speciation in this closely related group of species, combined with the certainty of their evolutionary relationships as determined by polytene chromosome banding patterns (Carson 1971, 1983; Carson and Kaneshiro 1976), help to construct phylogenetic trees devoid of the problem of multiple substitutions per site. This advantage is accentuated in the *planitibia* subgroup by their small populations and by the founder events that presumably gave rise to speciation events. Small populations and founder events minimize the problems that polymorphism in ancestral populations can cause when few chromosomes are investigated among closely related species (Coyne and Kreitman 1986). In addition, because most of these species are endemic to specific islands whose geological age is known, maximum divergence times, and consequently divergence rates, can be estimated. In particular, biogeographical and polytene chromosome banding pattern evidence (Carson and Yoon 1982) suggest that *Drosophila grimshawi* diverged from the *planitibia* subgroup in the island of Kauai. Potassium–argon ages measured on 16 Kauai samples lie in the range of 3.5–5.1 million years [Myr (McDougall 1979)] with a mean of 4.4 Myr. The rate of DNA divergence can be calculated using these species divergence times, and its constancy corroborated with the percent of varying positions between species endemic to the island of Hawaii and *Drosophila planitibia,* a species endemic to Maui, whose western and eastern slopes are 1.3 and 0.9 Myr old, respectively (McDougall 1964). The rate and its constancy can be further corroborated if the time of divergence between the two species in the island of Hawaii, *Drosophila silvestris* and *Drosophila heteroneura,* are approximated to the geological age of Hualalai [0.4 Myr (Funkhouser et al. 1968; McDougall 1979)], the volcano their founder(s) colonized (Carson and Bryant 1979; Carson 1982a; Kaneshiro and Kurihara 1982; DeSalle et al. 1986).

A detailed study of the evolution of a 4.1-kb *Pst*I fragment spanning autosomal chorion genes *s18, s15,* and *s19* in the Hawaiian *Drosophila* is begun here, to be integrated with previous observations entailing distantly related species of the same genus (Martínez-Cruzado et al. 1988; Fenerjian et al. 1989) in an attempt to acquire a complete picture of the mode of evolution of the cluster.

In this article, the linearity of molecular divergence of regions showing different divergence rates is addressed by assessing their relative rates of divergence in closely related taxa of the Hawaiian *Drosophila,* and comparing them to divergence levels previously observed between distantly related species. The relative observed rates between different coding regions, as well as between different noncoding sequences, seem to be the same for both sets of taxa. Nevertheless, the observed rates of coding versus noncoding regions are in contrast: the coding regions evolve at least as fast as the intergenic regions in the Hawaiian *Drosophila* but exhibit much more sequence similarity in the context of distantly related species. This is not necessarily the result of nonlinear divergence, but rather an expression of disparate modes of evolution of coding and noncoding regions. Unlike noncoding regions, the coding regions may evolve in circles, not gaining much differentiation despite a constant and rapid rate of substitutions as shown in this paper for the Hawaiian *Drosophila.* The constant but rapid rate of substitution comes in spite of evidence of excess polymorphism in and around the *s15* chorion gene.

In addition, evolutionary features of the cluster including length mutations and substitution types (transitions versus transversions) are explored here.

Materials and Methods

Isolation of Clones. Genomic libraries were the kind of gift of J. Hunt, University of Hawaii (*planitibia* subgroup taxa) and M. Kambysellis, New York University (*D. grimshawi*). All libraries were *Eco*RI partial except for *D. heteroneura,* a partial *Alu*I/*Hae*III library with *Eco*RI linkers. The origins of the libraries were as follows: the *D. heteroneura* library was generated from Q71G12, an isofemale line collected at Olaa Tract, Hawaii; *D. silvestris* from isoline U28T2 (here referred to as *D. silvestris*-Hilo, alluding to the eastern side of the island) collected at Kilauea Forest Reserve, Hawaii; *D. silvestris* from isoline U26B9 (here referred to as *D. silvestris*-Kona, or from the western side of Hawaii) collected at Kahuku Ranch, Hawaii; and *D. planitibia* from mass culture U84Y from Waikamoi, Maui. The libraries of *Drosophila differens* and *Drosophila picticornis* were constructed from strain U43V1 (Hanahilo, Molokai) and mass culture U71J (Kumwela Ridge, Kauai), respectively. The libraries were screened with a 7.6-kb *Eco*RI *D. grimshawi* fragment (Martínez-Cruzado et al. 1988), including genes *s15, s19,* and *s16* (Fig. 1), at Tm-33°C using standard procedures (Maniatis et al. 1982). Positive phage clones were purified, restriction mapped, and subcloned by established methods (Maniatis et al. 1982).

DNA Sequence Determination and Analysis. Restriction fragments were cloned into plasmid vectors pSDL12 or 13 (Levinson et al. 1984). Nested deletions were generated using the procedure

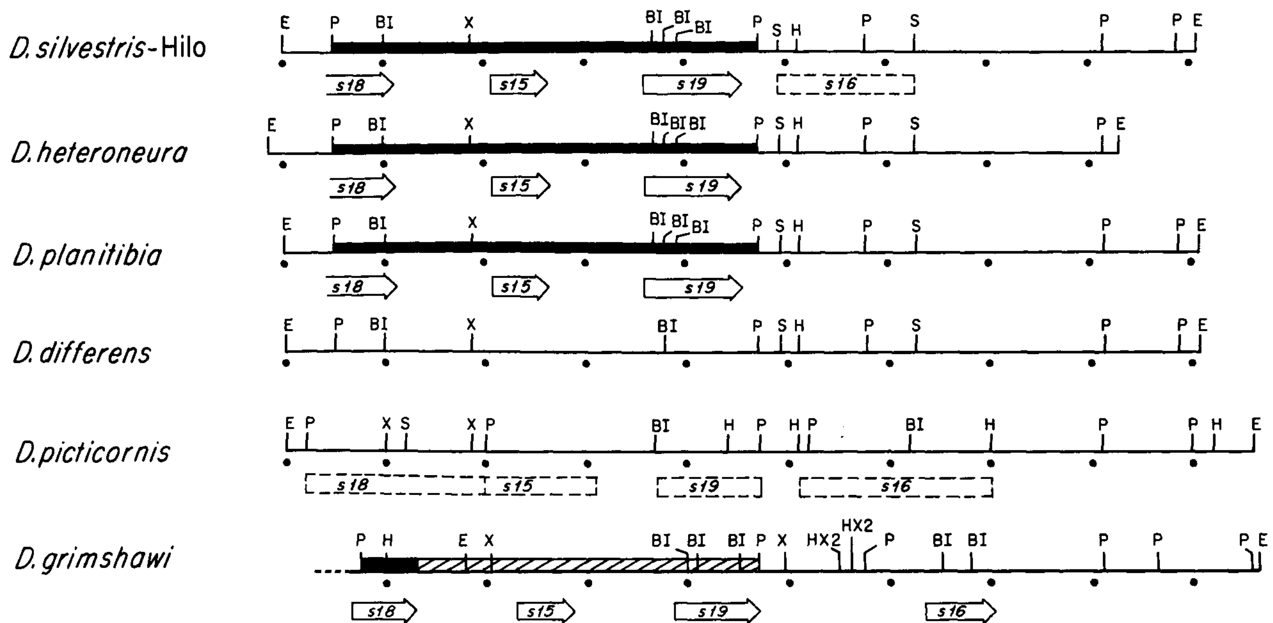


Fig. 1. Restriction map of the autosomal chorion gene cluster in six species of the Hawaiian *Drosophila.* Codes are: BI = *Bgl*I, E = *Eco*RI, H = *Hin*dIII, P = *Pst*I, S = *Sal*I, and X = *Xba*I. Maps are aligned by the common *Pst*I site located downstream of *s19,* and distances marked by dots in 1-kb intervals. Open boxes show gene locations and orientations if established by sequence analysis. Dashed boxes demarcate restriction fragments cross-hybridizing with specific gene probes. The *D. picticornis s15* gene is between a *Pst*I and a *Pvu*II site. The genes were not mapped in *D. differens.* Boxes on top of the maps represent sequences used for analyses that are reported here (solid) or elsewhere (hatched).

of Henikoff (1987) and the resulting clones sequenced by the dideoxy chain termination method (Sanger et al. 1977) using ^{35}S-thio-dATP (Biggin et al. 1983). Sequence reactions were run on buffer gradient polyacrylamide gels (Biggin et al. 1983). Each nucleotide was sequenced an average of 6.6 times and at least once in each direction. Computer programs were used to compile the resulting data (Staden 1982, 1984) and to analyze the final sequence (Pustell and Kafatos 1984).

Results

Organization of the Autosomal Chorion Gene Cluster in the planitibia Subgroup

For an initial characterization of the chorion gene cluster in the *planitibia* subgroup taxa, all genomic libraries, except for *D. silvestris*-Kona, were screened in parallel for clones hybridizing to the autosomal chorion gene cluster of *D. grimshawi.* The recovered phage clones were subjected to preliminary restriction analysis. All isolated clones contained all four chorion genes within single *Eco*RI fragments. The fragments were subcloned into a plasmid vector for further mapping. All taxa exhibited essentially identical restriction maps except for *D. picticornis* (Fig. 1). The latter appeared to share restriction sites with either *D. grimshawi* (*Bgl*I, around *s16*) or the *planitibia* subgroup species (*Eco*RI at the beginning of the fragment), or both (*Xba*I site upstream of *s15, Pst*I site downstream of *s19*), as well as displaying some unique sites (*Hin*dIII, downstream of *s16*). The 4.1-kb *Pst*I fragment containing genes *s18, s15,* and *s19* was chosen for sequence analysis because of its high content of coding sequences, its relatively large size, and because of its conserved ends in the taxa. The *D. silvestris*-Kona genomic clone, isolated later as above, was shown to possess the 4.1-kb *Pst*I fragment as well. The fragment was subcloned and subjected to sequence analysis.

The Evolution of Length Mutations

The nucleotide sequence was determined for the 4.1-kb *Pst*I fragment of *D. silvestris*-Kona, *D. silvestris*-Hilo, *D. heteroneura,* and *D. planitibia,* as well as for the *D. grimshawi* 574-bp segment composed of the last 6 bp of the intron and the second exon of *s18.* Sequences pertaining to *D. differens* and *D. picticornis* will be presented elsewhere. The rest of the *D. grimshawi* sequence used here for analyses and shown aligned to the *planitibia* subgroup sequences (Fig. 2) has been reported previously (Martínez-Cruzado et al. 1988).

In the alignment, nucleotide positions were numbered as their homologues in the *D. silvestris*-Kona sequence, which is the only taxon without unique length mutations and was taken as standard. Insertions were referred to as the *D. silvestris*-Kona position that follows the insertion. Deletions were numbered by the first missing *D. silvestris*-Kona nucleotide.

The sequence starts with the last 6 bp of the *s18* intron and ends at a *Pst*I site, which is the center of

a 12-bp palindrome that starts 128 bp downstream of *s19*. The palindrome is conserved in three other distantly related drosophilids (*Drosophila virilis, Drosophila melanogaster,* and *Drosophila subobscura*). Hence, the ends of the sequenced fragments can be aligned with certainty and define a homologous region. The length of this fragment is 4116 bp in *D. silvestris*-Kona, 4131 in *D. silvestris*-Hilo, 4092 in *D. heteroneura,* 4101 in *D. planitibia,* and 4009 in *D. grimshawi*; the length variation is only 3% (Fig. 3). With respect to *D. silvestris*-Kona, *D. silvestris*-Hilo shows 2 length mutations, *D. heteroneura* and *D. planitibia* 6, and *D. grimshawi* 31. Only one length mutation is potentially shared by two taxa: a 1-bp deletion (within an A run) between positions 2602 and 2607 (Fig. 2) in *D. grimshawi* and *D. planitibia.* In addition, several taxa have length differences encompassing a variable number of units in extensive tandem direct repeats located between 139 and 210 ($GGCGGCTATGGC_n$), 1475 and 1520 (TA_n), and between 2881 and 2898 (AAG_n).

Length Mutations Are Associated with Direct Repeats

As in prokaryotes (Farabaugh et al. 1978; Albertini et al. 1982) and many eukaryotes (Jones and Kafatos 1982), but unlike the *Adh* locus in Hawaiian *Drosophila* (Rowan and Hunt, unpublished), a large majority of the observed length mutations are associated with direct repeats, either tandem or nontandem. Of the 43 independent length mutations in the sample, only 3 (at positions 1790, 1857, and 2293) are clearly not associated with direct repeats. Of the remaining 40, 37 can be definitively assigned to one of the four categories proposed by Jones and Kafatos (1982) as typical of length mutations related to direct repeats. In particular, 30 involve tandem repeats and 5 involve the deletion/insertion of a repeat unit with or without the interrepeat unit sequence. Two others (at 1435 and 2626) belong to the fourth category of Jones and Kafatos, where one of the nontandem repeats along with only part of the interrepeat region is deleted/inserted. The remaining three length mutations (528, 758, and 2403) seem to have involved other rearrangements or more than one event, but it is notable that they share small tandem repeats within or close to the deleted/inserted segments.

Comparable Length Mutation Rates for Coding and Noncoding Regions

The rates for length mutations in coding and noncoding regions of the chorion gene cluster are comparable. Figure 3 shows the distribution of length mutations, which are just as frequent in coding as in noncoding regions: 28 length mutations (lm) in 2616 bp (0.011 lm/bp) for noncoding and 15 in 1500 bp (0.010 lm/bp) for coding sequences. It could be that the chorion coding regions are templates unusually prone to length mutations due to long and numerous tandemly repetitive sequences rich in glycine and tyrosine codons, and to the presence of stretches of alanine-coding triplets that are expected to enhance the probability of length mutations (Streisinger et al. 1966). However, chi-square and *G*-tests show that the ratio of length mutations associated with tandem direct repeats to the total number of length mutations in coding regions (12/15 = 0.80) is not significantly different ($P > 0.1$) than that in noncoding regions (18/28 = 0.64).

→

Fig. 2. See pages 406–407. DNA sequence of the 4.1-kb *Pst*I fragment encompassing the second exon of *s18* and the *s15* and *s19* genes in five taxa of the Hawaiian *Drosophila.* Abbreviations are SiK = *D. silvestris*-Kona, SiH = *D. silvestris*-Hilo, Het = *D. heteroneura,* Pla = *D. planitibia,* and Gri = *D. grimshawi. Drosophila silvestris*-Kona is taken as standard, with asterisks highlighting every 10 bp. Blank spaces represent nucleotides as in the standard sequence. Specific base differences are illustrated at the corresponding positions. Putative deletions are indicated by dots. The conceptual translations into protein pertain only to *D. silvestris*-Kona. Highlighted are the invariant sequences of the chorion-specific hexamer, TCACGT (Wong et al. 1985), the TATA boxes, the polyadenylation signals, and the cap sites, themselves distinguished by bent arrows. Vertical lines separate central intergenic regions from distal and proximal regions to their sides (see text). Three single nucleotide differences adjacent to a length mutation (e.g., 1860) were viewed as events related to the length mutation, and therefore were not utilized in base substitution analyses. Similarly, only a length mutation was scored at two stretches of DNA that may have undergone complex rearrangements (at positions 528–532, 758–763, and 2403). Base substitutions next to a tandem repeat of different multiplicity in different taxa (e.g., C–T change at position 835) were counted, as they probably do not border the length mutation. In the case of tandem direct repeats, once alignments were performed by the aforementioned rules, direct sequence comparisons were used to assign base substitutions even when all bases at a specific site may differ from their consensus repeat unit (e.g., 3400). In the case of positions 3649–3684, at which *D. grimshawi* has tandem direct repeats with a consensus sequence (GGTGGACACCTT) not identical to those of the *planitibia* subgroup taxa (GGTGGAATCCAT), it was assumed that one of the repeated units was used to convert others at some point in the evolution of each of the lineages. This evolutionary path yields the smallest number of length mutations and base substitutions in the repeats. To compare the repeats directly would only add three base substitutions to the end result. Elimination of the correction step in the *planitibia* subgroup would add two substitutions. The occurrence of a correction and expansion event in *D. grimshawi* is strongly suggested by the larger number of repeat units, of different consensus, in *D. grimshawi* than in *D. virilis* (Fenerjian et al. 1989) and the *planitibia* subgroup.

A

```
    Cys Leu Phe Ile Cys Ala Ile Ala Ala Val Ser Ala Ser Gly Tyr Gly Asn Val Gly Tyr Gly Gly Ala Arg Leu Gly Gly Leu Ala Tyr Gln Val Gln Pro Ala Leu Thr Val Ser Ser Ile Val Pro Val Gly Gly Tyr Gly
SIK  1 CTGCAG TGC CTT TTC ATT TGC GCC ATC GCC GCT GTT TCG GCT TCC GGC TAT GGT AAT GTT GGT TAT GGA GGT GCG CGC CTC GGT GGA TTG GCC TAT CAG GTG CAG CCT GCT CTG ACT GTT AGC TCT ATC GTG CCC GTT GGC GGA TAT GGT  SIK
SIH                                                                                                                                                                                                        SIH
Het                                                                                                                                                                                                        Het
Pla    C C T                                                                                                                                                                                               Pla
Gri    G T A G C T C G T T                                                                                                                                                                                 Gri

    Gly Gly Tyr Gly Gly Gly Asn Val Gly Gly Tyr Gly Gly Gly Tyr Gly Gly Gly Lys Gly Tyr             Gly Arg Ser Val Glu Val Pro Val Ala Ala Val Phe Thr Ser Asn Ser Arg Tyr Gly His Gly Val Ala Pro Ile
SIK 151 GGC GGC TAT GGC GGC GGC AAT GTC GGC GGC TAT GGC GGC GGC TAT GGC GGC GGC AAG GGC TAC ... ... ... ... GGT CGC AGC GTT GAG GTC CCC GTT GCC GCT GTG TTT ACT TCG AAC TCC CGT TAT GGT CAT GGT GTC GCA CCC ATC  SIK
SIH                                                                                                                                                                                                        SIH
Het    ... ... ... ... T                                                                                                                                                                                   Het
Pla                                                                                                                                                                                                        Pla
Gri    ... ... ... ... ... ... ... ... ... ... ... ... ... G GGC AGG GGC TAC GG C C ... ... G T                                                                                                          Gri

    Asp Arg Gln Ala Leu Gly Leu Ala Lys Leu Ser Leu Ala Ala Pro Gly Ala Gly Gly Pro Leu Val Tyr Asn Gln Pro Arg Arg Ile Val Asn Val Ser Ser Tyr Gly Pro Gln Gly Ser Tyr Lys Gln Pro Leu Gly Tyr Gly Ser Ile
SIK 289 GAT CGT CAA GCT CTC GGT CTG GCC AAG CTC AGT CTG GCT GCA CCC GGT GCC GGT GGT CCT CTG GTC TAC AAC CAG CCC CGT CGA ATT GTC AAT GTG TCC AGC TAT GGC CCA CAG GGC AGC TAC AAG CAG CCA TTG GGC TAC GGA TCA ATC  SIK
SIH                                                                                                                                                                                                        SIH
Het    G G                                                                                                                                                                                                 Het
Pla    C G                                                                                                                                                                                                 Pla
Gri    C C T C TG G G A A G G C G                                                                                                                                                                          Gri

    Glu Gln Ala Gln Gly Ala Ser Ala Ala Ala Ala Ser Ser Ser Val Ala Gly Lys Asn Lys Gly Tyr Gln Asn Ala Gly Tyr ---
SIK 439 GAG CAG GCT CAG GGC GCT TCG GCT GCC GCT GCC TCC AGC TCC GTG GCT GGC AAG AAC AAA GGA TAC CAG AAC GCC GGC TAT TAA GTGACTTTGAAAAGCAACAGTGATTAAGCACAGCTGAAAAGCTCTGGTTCTTATAGACGCATAAGGCAGTAGGAAGTAATAAAAAGCA  SIK
SIH    A T C A                                                                                                                                                                                             SIH
Het    C C                                                                                                                                                                                                 Het
Pla    C T C                                                                                                                                                                                               Pla
Gri    C G G CA AGC.. C                                                                                                                                                                                    Gri

SIK 611 ACGAAACATCAAAACTACAACAATTTTTTGTCATTTTGTGGGTCTTTGCCAATTGCTGATGAACTTTCAGTTTCTTGATTTTGTATGGCTGAATAAGCTGCATGCAAAGAATGATATTTATGGCTTGTGGGGAAACTGTCTAGAGAC...AGTGACAATTTTGCCATAATTTATGAAAATATTGCAAAGAGCTTTTTGAA  SIK
SIH    G                                                                                                                                                                                                   SIH
Het    T T                                                                                                                                                                                                 Het
Pla    T T T                                                                                                                                                                                               Pla
Gri    T T C T A A C G C C A GCTTAAATT G T G AG G A                                                                                                                                                        Gri

SIK 808 GACTTGCAGAGTTTAGTTTGTTGATTTT..TAAATAGACAAACTGTGAACTTAAAATAAATATAGAGCAACCTTTGTCAGCTAATAGTAAATCAATGTTGCTATTGAAGTATTGAATTTGAATCCAAGTCGAATCTATCGATAATGTATCTTAGTTATTTTGATTAAATGCTGATCGTCCTGTACGCTCATGAATTTATT  SIK
SIH    C                                                                                                                                                                                                   SIH
Het    C C                                                                                                                                                                                                 Het
Pla    C G A                                                                                                                                                                                               Pla
Gri    T TT CTA T C C A A G A G G                                                                                                                                                                          Gri
```

B

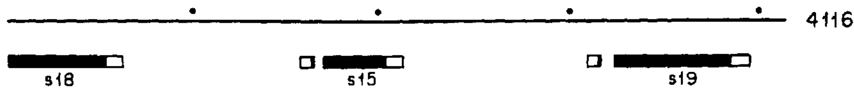
```
SIK 1086 TGAAGCATATCCGGTATCAAAAATCAACTGGCAACAGCTCCATTTGTGACTGCTCGATTGAGTGTTACTCGTTGGAATGTGAGTATAATTTCAAATCCAAAAATTCCAAATACATTGCAATAAATTGGGGCTACAAATGGC..GTGTTTCGTTTCGCGGATTCATTGAGAACTCACGAAGTGCTCGTACTTATTATTATT  SIK
SIH    C                                                                                                                                                                                                   SIH
Het    C C                                                                                                                                                                                                 Het
Pla    AC                                                                                                                                                                                                  Pla
Gri    T C GC G G GT                                                                                                                                                                                       Gri

SIK 1284 AT......TATATGTATATTTTTT.TTGTGCTTAGTATGTGTCGAACCCAATACGAGGGGCGGGTTTTTGCCTGGACATTATCAGTGGCCAGAAACAACGAAACATCGCGTAACCCAGCGATTTTATAATTGCTCGAGACACAACAACAAAACTACAAATGCAAGTACATAATAACAAGTTTCTCTTATGTACTTGGCAA  SIK
SIH    A A C                                                                                                                                                                                               SIH
Het    T ...... T                                                                                                                                                                                          Het
Pla    G T G A C                                                                                                                                                                                           Pla
Gri    ATTATT T A G T GT G C G                                                                                                                                                                             Gri

SIK 1397 TTATGAAATATATCACAGACATAGAAATTATACTAAATATAATTACCTGAAAAGTAACTACATAGATGTACACAGGTAATCTATATATATATTTATATATATATAAATAAACATATTAATATATGGAGGGTCACGTAAGTATCCCGTCTAATTTTTGCATATAAAAGCGAACGTTGCTGGCAAGTAAATCATAGTTTGAT  SIK
SIH    C                                                                                                                                                                                                   SIH
Het    A .... A                                                                                                                                                                                            Het
Pla    G A ...... A G                                                                                                                                                                                      Pla
Gri    T A A C G............ A ......... G A C C                                                                                                                                                           Gri

                                       Met Lys Phe Leu                                                   Ile Ala Phe Ala Thr Val Ala Val Phe Ala Cys Val Ser Ala Asn Pro Tyr Gly Gly
SIK 1597 TGATTACCACAACCAACAAAACTAAATATTCAAC ATG AAG TTC CTG GTAAGTAATCCTGGCGCTATTCACTGCTTGACAGACACACTAATCCTTTGGATTTTTATTTTGTTACTTGCAG ATC GCT TTT GCC ACC GTC GCC GTC TTC GCC TGC GTT AGC GCT AAT CCC TAC GGC GGT  SIK
SIH    G G                                                                                                                                                                                                 SIH
Het    T A                                                                                                                                                                                                 Het
Pla    C G                                                                                                                                                                                                 Pla
Gri    A G T G G T T G A T GG GGA T ...                                                                                                                                                                    Gri

    Asn Ile Gly Tyr Gly Gl  y Leu Gly Gly Gly Tyr Leu Gly Gly Leu Gly                              Gly His Gly Leu Gly Arg Val Gly Tyr Val Gln Asn Leu Gly Gly Gly Tyr Gly Arg Arg Ser Leu Leu Val
SIK 1773 AAC ATT GGA TAT GGC GG. ..C CTT GGT GGT GGC TAT CTG GGT GGT CTT GGC ... ... ... ... ... ... ... ... GGT CAT GGA CTG GGT CGT GTT GGC TAC GTT CAG AAC CTT GGC GGT GGT TAT GGA CGC AGA TCT CTC CTG GTC  SIK
SIH    A ... A                                                                                                                                                                                             SIH
Het    A A C GGT CTT GGC GGT CTT GGC GGT CTT GGC                                                                                                                                                           Het
Pla    T AA A ... ... ...                                                                                                                                                                                  Pla
Gri    ... ... ... ... . ... ... ... ... ... ... ... G T T AA C A C C A T A                                                                                                                                Gri

    Ser Gln Pro Ser Asn Pro Ser Ala Ala Ala Ala Ala Ala Ala Ala Ser Ser Gly Val Arg Pro Gly Leu Tyr Asn Gln Gly Gly Val Ile Gly Tyr Glu Leu Asp Gly Gly Tyr Ile Gly Arg Arg Gly Gly Tyr Gly Gly Gly Leu Gly
SIK 1893 TCC CAA CCA TCG AAC CCC AGT GCT GCT GCT GCC GCC GCT GCT GCC TCC TCT GGT GTC CGT CCC GGT CTA TAC AAC CAG GGT GGC GTC ATC GGA TAC GAG CTT GAT GGT GGC TAC ATT GGT CGT CGT GGT GGT TAT GGC GGA GGT CTG GGC  SIK
SIH    G                                                                                                                                                                                                   SIH
Het    G G A                                                                                                                                                                                               Het
Pla    G A C                                                                                                                                                                                               Pla
Gri    A G T T AA T C AT C G AC                                                                                                                                                                            Gri
```

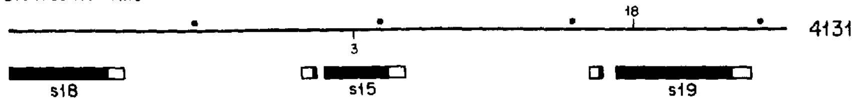
C

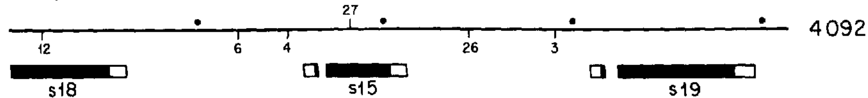
```
     Tyr ---
SlK  2063 TAT TAA TTCAGCCACCTCCAATTGAACGAATTGACAGTTTGCTCAATGGACCAGCAACAGCTACTGAAGTAACAGAAGCTGTTGCCAAAGCTGTGACTAAATAAAAACCGCTGAGTGCCGCGGAGCATGTTGAACAAAAAAATTTTACTACTTGTCTTTTGTCTACTGGGTAGCTAGAAAATGTGACATACTTTTGAGC  SlK
SlH       A C  SlH
Het       T A A C  Het
Pla       T G A C  Pla
Gri       A A T A A A A C CT C  Gri

SlK  2241 GTGGGCCTTAAATGGTTTTAGCTGGAATTACCGGGTGCCTCCATGTCAAGGAATATTCGTGGTTGAGCCCCACGTAAAATTAGACAGCCTTCAATTTACAAATAGATTATAAAGAATTGCTGATAGCAGCAAATTGAGAGGGATAGGTTTTTAGGTGCATAGCT.AATCAGCCAGGATACAATTTAGTATCCTCTTTATTG  SlK
SlH       G T G T A  SlH
Het       T A A  Het
Pla       G T C A A G  Pla
Gri       C G ......... T A C TG T G T A A A A AGC T C C G C  Gri

SlK  2448 GTCTTCA.AACAAATTTTTGACGAAACGCTTTTGAAGAAATTGCTACAAAGAGCTGGATAAAATCTAATTTGGCTGTTATAAAATATTTCTAAGAAAAAACCATACTTTGATAGTATTTAAGATAGCCTTACACAACAAATATAAAAGTTAAATAATATAAGTAAAAAACAAATAGAAACAGCCTTA............  SlK
SlH       A A G  SlH
Het       .........................A A TA T C T C  Het
Pla       A A TA T C C . G  Pla
Gri       C G A A A G C A G A ............ G G G A T T C .C T AGTTAAAAAGTCT  Gri

SlK  2626 ...GCCATTTCGTAGCTGGTCCAAAAACTAAAGTCTGTTTAGTTGTTTTGAGAGA.................GAATTGAGAGGCTGCTTCTTATCCAGCTGCATGCTCTCACGAAGCAGCAGCATCACCAGCTTCATTTAGTAGCACATATATTGGGAAGCAGTCCAACGAGAGATGAACACCCCCAATGTGGGGTTCA  SlK
SlH       C  SlH
Het       T  Het
Pla       T C T C  Pla
Gri       CTT G T G G A GAGAGAGAGAGAGAGAGT C G C G G C T  Gri

SlK  2885 AGTGTCGCCGATACGAGAGACGGTCGAACAAGCTGCTCAGGCGATATGCAGGTGCGATTTCGAAAGAAAAAAAA.CAAGAAGAAGAA..............GAAGAAGATGATGGAGAAGAAGAAGAACTGAGGCTGAAAGTGGGGCTGATGAACGCCCATCGTCATCGCCGAGAAGAAGAAAATTAAGCGGTGGTTA  SlK
SlH  SlH
Het       ... T  Het
Pla       T T . C  Pla
Gri       C G C C GAAGAAGAAGAAGAA A C C A  Gri

                                                                                                   Met Asn Thr Phe Ala
SlK  2989 AGAAGCTGAAGCTGAAGCTCAGCGCCGTGATAAATTCGCGGCGAGATCACGTTTCGAGTGCTTCAATTATTTTTGCTTATATAAAGAAGTGTGCTTGGCATTTGGTTTGTTAATTGTGCCAACTGTTCCAAGCAGCGAGCGCCCCCCAAATAACTAAAAGCATAACC ATG AAC ACA TTC GCC GTAAGTAGTCGA  SlK
SlH       A  SlH
Het  Het
Pla  Pla
Gri       A T T  Gri
```

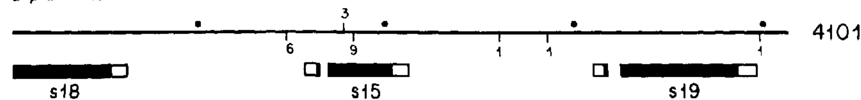
D

```
                                                                      Thr Leu Ala Ile Phe Ile Ser Ala Cys Leu Ala Leu Gly Ser Cys Gly Gly Tyr Gly Ser Gln Ile                     Gly Tyr Gly Ser
SlK  3183 TGACATTTGACGCCAATTGGCACAATTTCCGTTTGCCGCCATTAAGCATCGATTTCCACTCTCCATGTAG ACT CTT GCG ATC TTC ATT AGC GCC TGC CTG GCT CTT GGC AGC TGC GGC GGC TAT GGT AGC CAA ATC ... ... ... ... ... ... GGC TAT GGT AGC  SlK
SlH       A C GGC TAT GGT AGC CAA ATC  SlH
Het       C  Het
Pla       T C  Pla
Gri       G T C C T C G C ... ... ... ... ... ... ... ... ... ...  Gri

          Gln Ile Gly Tyr Gly Ser Gln Ile Gly Tyr Gly Ser Pro Ile Gly His Gly Ser Gln Ile Gly Tyr Gly Arg Pro Ile Asn Val Gly Leu Arg Arg Leu Ser Ser Val Gly His Gln Ser Gly Asp Gly Ala Ala Ala Ala Ser Ala Ala
SlK  3331 CAA ATC GGC TAT GGT AGC CAA ATC GGC TAT GGT AGC CCA ATC GGC CAT GGT AGC CAA ATC GGC TAT GGT CGC CCA ATC AAT GTT GGA CTC CGT CGC CTC TCC AGC GTT GGA CAC CAG TCC GGT GAT GGT GCC GCT GCT GCT TCG GCC GCT  SlK
SlH  SlH
Het       A  Het
Pla       C G T A G  Pla
Gri       ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... T CC G T G A G T  Gri

          Ala Ser Gly Gly Asn Lys Gly Pro Val Glu Ile Ile Ala Gly Gly Gly Thr Arg Tyr Gly Ala Ser Gln Asn Leu Arg Pro Ile Leu Leu Asn Ser Gly Tyr Asn Gly Gly Leu Asn Asp Asn Ile Gly Arg Ile Ala His Ile Val Gly
SlK  3481 GCT TCC GGT GGC AAT AAA GGC CCC GTT GAG ATC ATT GCT GGT GGT GGT ACC CGT TAT GGC GCC AGC CAG AAT CTG CGC CCA ATC CTC CTC AAC TCT GGT TAC AAT GGC GGA CTC AAC GAC AAC ATC GGT CGC ATC GCA CAC ATT GTC GGC  SlK
SlH       C  SlH
Het       C  Het
Pla       A C G C A  Pla
Gri       G C T C C T A C T C A  Gri

          Gly Gly Arg Ser Leu Val                                 Gly Gly Ile Asn Gly Gly Ile Tyr Gly Gly Gly His     Gly Gly Tyr Gly His Pro Arg Phe Ser Val Gln Pro Ala Ser Ala Thr Leu Leu Tyr Pro Gly Gln Asn
SlK  3631 GGT GGC AGG TCT CTC GTT ... ... ... ... ... ... ... ... GGT GGA ATC AAC GGT GGT ATC TAT GGT GGA GGC CAT ... GGG GGT TAT GGT CAT CCT CGT TTT TCT GTC CAG CCT GCC AGC GCC ACT CTC CTG TAT CCA GGC CAG AAC  SlK
SlH       A  SlH
Het       G  Het
Pla       A G  Pla
Gri       C ... GGT GGA CAC CTT GGT GGA CAT CTT CA CTT A CG AT AA T GGT C AT G C A T G  Gri

          Ser Tyr Arg Ser Ile Ser Ser Pro Val Glu Tyr Ser Lys Val Ile Leu Pro Val Arg Ala Ala Ala Pro Val Ala Lys Leu Phe Ile Pro Gln Asn Asn Tyr Gly Asn Tyr Gly Gly Ser Ala Gly Tyr ---
SlK  3756 AGC TAT CGT AGC ATT TCG TCG CCA GTC GAG TAC TCG AAG GTG ATT CTG CCC GTC CGT GCT GCT GCT CCA GTT GCC AAG CTG TTT ATA CCC CAG AAC AAC TAT GGC AAC TAT GGC GGC TCT GCT GGC TAC TAA AAAGTACCCTTCGAGTACTTTGCT  SlK
SlH  SlH
Het  Het
Pla       A T  Pla
Gri       C C A ... ... ... A CAA T T  Gri

SlK  3910 CTCTCCGAATCTCTGAGAACAATAACATGGAAGAGTTCGTTGTGCAAACAGCGCTAATAAAATTTT.AAAACAAGAAGAAAAAAAGAAAAAGAAAATGTTGTATTGATATTAAGAGGCATTGAGAGGGTGTAGATAAGTGGTTCCGGCCACCGCCAGTTGTTGTTCGCCAATGTTTGGCTGTTGTCGGTTGTCTGT..TGGCAGCTGCAG  SlK
SlH       G C  SlH
Het       G T C  Het
Pla       G . G C C  Pla
Gri       A C G A C.... C A T G T T GT  Gri
```









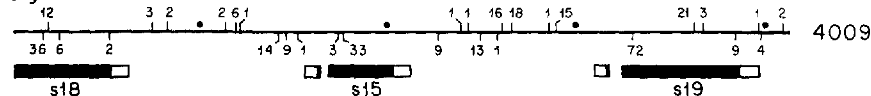


Fig. 3. Length mutation map of the sequenced DNA fragment in five taxa of Hawaiian *Drosophila.* All length mutations are positioned relative to *D. silvestris*-Kona. Relative insertions are denoted on top of the lines; deletions are placed under the lines. Number of base pairs in each length mutation are indicated. Filled boxes refer to the coding regions, empty boxes to untranslated regions of the messages. Dots mark distances in 1-kb intervals. The corresponding lengths (bp) are indicated to the right of each fragment.

Similar Substitution Rates in Intra- and Intergenic Sequences

The chorion cluster sequences are highly conserved in the Hawaiian taxa. In 3842 positions where alignments are possible for all five taxa, only 328 substitutions (8.5%) are observed, most of them (226) distinguishing *D. grimshawi* from the other four taxa. This allowed for a largely unambiguous alignment (Fig. 2). As criteria for the alignment, gaps were minimized in preference to base substitutions. Alternative specific ends of length mutations that are equivalent in terms of number of substitutions were decided arbitrarily. Within this framework, and in accordance with the parsimony rule, the alignment was constructed to imply the smallest number of base substitutions in the phylogenetic tree for sequences that were present in all the taxa (see below).

The coding regions are evolving as fast as the most divergent of the intergenic regions. To test whether different types (e.g., transitions vs transversions) and rates of substitutions could be detected in different regions in the evolution of these taxa, the fragment was divided into different, or potentially different, functional regions. Hence, coding regions were analyzed separately from noncoding regions. The noncoding regions were divided into transcribed and intergenic regions, which were themselves partitioned into proximal, distal, and central intergenic segments. The idea for the latter subdivision was to distinguish regions that are already known to have different base substitution rates in *Drosophila* evolution and that, for their different positions relative to structural genes, could be reasonably postulated to have different functions. Analysis of chorion intergenic sequences in *D. grimshawi* and other distantly related species shows that relatively conserved regions flank the genes at both sides (Martínez-Cruzado et al. 1988; Fenerjian et al. 1989). These were named "proximal" if found in the region flanking the 5′ end of a gene, and "distal" if found in the 3′ flanking region. The segments were defined as extending from the cap site (proximal) or polyadenylation signal (distal) of the genes out to the farthest three-way match of 5 bp or four-way match of 4 bp in alignments involving *D. melanogaster, D. subobscura, D. virilis,* and *D. grimshawi.* The proximal segments extend up to −436 for *s15* (in *D. grimshawi*; cf. Martínez-Cruzado et al. 1988; position 1136 in the numbering of Fig. 2), and up to −324 for *s19* (in *D. grimshawi*; cf. Fenerjian et al. 1989; position 2785). The distal sequences range from 48 to 166 nucleotides in length and extend to *D. silvestris*-Kona 653, 2276, and 4116 for *s18, s15,* and *s19,* respectively (Figs. 2 and 4). For showing different divergence rates between distantly related species, I shall refer to these sequences as paleoheterodiverging regions.

The frequencies of positions varying in any one of the taxa in each of the paleoheterodiverging regions are shown in Table 1. Because the intent is to compare evolutionary rates in different regions, only sequences present in all five taxa are included. Figure 5 depicts a histogram derived from the data presented in Table 1. Although double base substitutions in comparisons between more than two sequences usually make standard errors impossible to calculate, these are so few in the considered se-

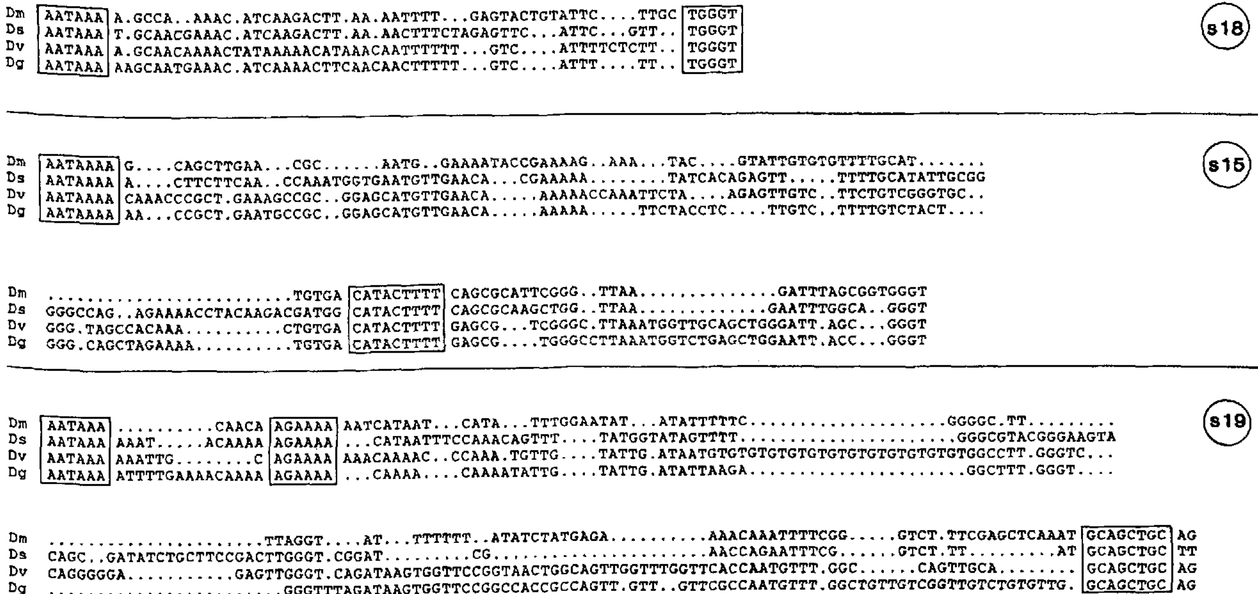


Fig. 4. Detailed alignments of relatively conserved sequences flanking the 3′ end of genes *s18*, *s15*, and *s19*. The distal sequences (see text) start with the first base following the polyadenylation signal (AATAAA), and extend to the farthest three-way match of 5 bp or four-way match of 4 bp. Alignments were performed as in Martínez-Cruzado et al. (1988). Invariant elements 5 bp or longer are boxed. Species included are *D. melanogaster* (Dm), *D. subobscura* (Ds), *D. virilis* (Dv), and *D. grimshawi* (Dg). *Drosophila grimshawi* was selected to represent the Hawaiian taxa for consistency with the previous 5′ flank alignments.

Table 1. Regional nucleotide divergence in the autosomal chorion gene cluster

Region	Length (bp) for five-way comparison	Substitutions[b]	% substitutions	Transitions[c]	Transversions[c]	Transition frequencies
Intron *s18*	6	0	0.0	0	0	
Second coding *s18*	471	40	8.5 ± 1.3	22	18	0.55 ± 0.08
3′ UT *s18*[a]	81	5	6.2 ± 2.7	4	3	0.57 ± 0.19
Distal *s18–s15*	48	4	8.3 ± 4.0	2	2	0.50 ± 0.25
Central *s18–s15*	476	41	8.6 ± 1.3	21	21	0.50 ± 0.08
Proximal *s18–s15*	420	24	5.7 ± 1.1	10	16	0.38 ± 0.10
5′ UT *s15*	44	0	0.0	0	0	
First coding *s15*	12	0	0.0	0	0	
Intron *s15*	73	7	9.6 ± 3.5	5	2	0.71 ± 0.17
Second coding *s15*	282	40	14.2 ± 2.1	23	23	0.50 ± 0.07
3′ UT *s15*[a]	102	7	6.9 ± 2.5	3	4	0.43 ± 0.19
Distal *s15–s19*	129	9	7.0 ± 2.2	6	3	0.67 ± 0.16
Central *s15–s19*	459	59	10.5 ± 1.2	39	27	0.59 ± 0.06
Proximal *s15–s19*	304	14	4.6 ± 1.2	5	8	0.38 ± 0.13
5′ UT *s19*	63	1	1.6 ± 1.6	1	0	1.00 ± 1.00
First coding *s19*	15	0	0.0	0	0	
Intron *s19*	82	6	7.3 ± 2.9	5	1	0.83 ± 0.15
Second coding *s19*	546	56	10.3 ± 1.3	29	30	0.49 ± 0.07
3′ UT *s19*[a]	88	6	6.8 ± 2.7	3	3	0.50 ± 0.20
Distal *s19–s16*	141	9	6.4 ± 2.1	4	6	0.40 ± 0.15
Total	3842	328	8.5 ± 0.5	182	167	0.52 ± 0.03

[a] The 3′ untranslated regions comprise sequences from the first base of the stop codon to the last base of the polyadenylation signal
[b] The number of substitutions refers to all base substitutions known to have occurred in the phylogenetic tree of sequences present in each of the five taxa
[c] The number of transitions and transversions includes all determined sequences as long as inference of base substitutions can be made. The only base substitution that could not be identified lies in a stretch of adenosines from 2871 to 2879

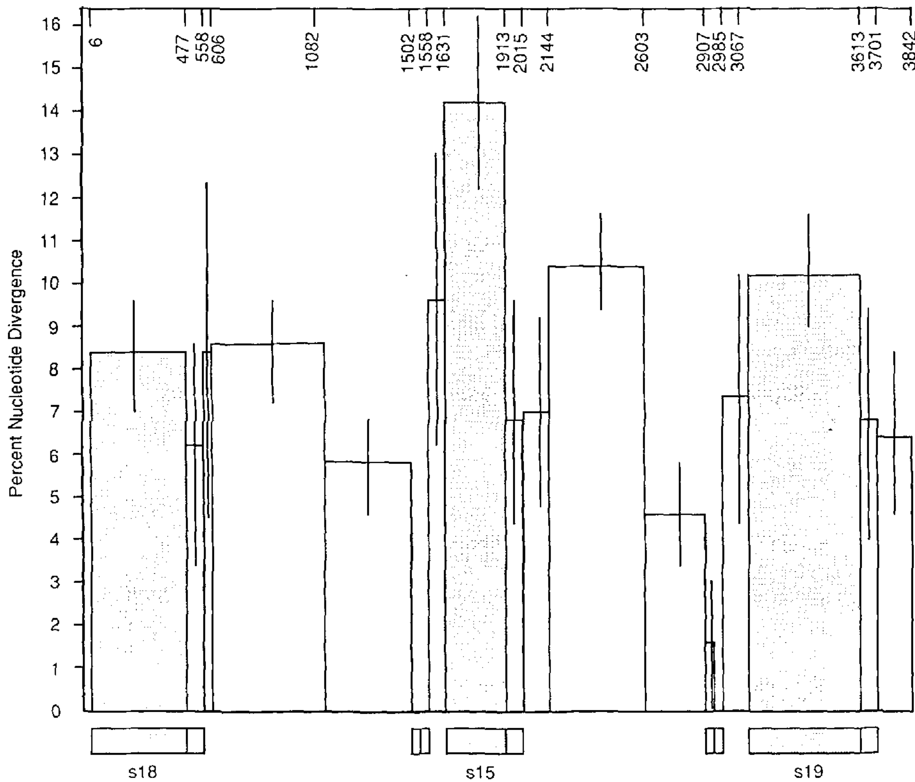


Fig. 5. Percent divergence in paleohetero diverging regions of the autosomal chorion gene cluster. Sequences reported in the abscissa are boxed if represented in mRNA. Coding regions and their corresponding histogram boxes are shaded. Bars represent one standard deviation from the mean. Last nucleotides of each region except for 5′ untranslated sequences (for aesthetic reasons) are indicated on the top. Intergenic regions are divided from left to right into distal, central, and proximal regions. Only sequences present in all five taxa are considered.

quences that standard errors could be calculated without any significant deviation from a hypothetical two-way comparison.

It is quite striking that the most divergent segment is the large coding region of *s15,* and that the divergence in the large coding regions of *s18* and *s19* are comparable to that for the least conserved central intergenic regions as well as the introns combined. This is interesting because the large coding regions clearly show a higher degree of conservation than central intergenic regions in distant species comparisons (Martínez-Cruzado et al. 1988; Fenerjian et al. 1989). However, the pattern of divergence is similar when comparisons are limited to coding or noncoding regions. Namely, as observed in distant species comparisons, introns and central intergenic regions of the Hawaiian taxa diverge at comparable rates, and faster than distal and proximal regions in that order. The 3′ untranslated sequences evolve as fast as distal regions, whereas the 5′ untranslated sequences are the most conserved. On coding sequences, those pertaining to the first exon are always more conserved than those of the second exon. The pattern of divergence in the Hawaiian taxa also reflects that of other *Drosophila* species in the order of divergence for the second coding regions: *s18* (Swimmer et al. 1990) is less divergent than *s19,* which is itself less divergent than *s15.*

Overall, the large (second) coding regions of all three genes show a divergence (10.5 ± 0.8) comparable to the central intergenic regions (10.7 ± 1.0) and the introns (8.4 ± 2.2).

It may be pointed out that only three sites show two unambiguously independent substitutions. These sites are occupied by three different nucleotides among the five taxa. In addition, 17 other sites show two different nucleotides that, according to the most probable phylogenetic tree (Fig. 6), would correspond to either parallel or reversed base substitutions in different evolutionary lineages, and would normally be scored as two substitutions for each site. However, 15 of these sites are scored here as having single base changes because they may be the result of polymorphisms maintained through evolutionary lineages (see below).

Phylogenetic Tree

The DNA sequence of the 4.1-kb *Pst*I fragment containing genes *s18, s15,* and *s19* was analyzed by a phylogenetic algorithm using parsimony (PAUP, Swofford 1985) for all five Hawaiian taxa: *D. grimshawi, D. planitibia, D. heteroneura, D. silvestris*-Kona, and *D. silvestris*-Hilo. The four shortest or more probable trees are shown in Fig. 6 (**a** to **d**, in descending order of probability). The three most probable trees are in agreement with what polytene chromosomes (Carson 1982a, 1983) could predict: *D. grimshawi* as the outgroup, and *D. planitibia* more closely related to, but outside of, the group of taxa in the island of Hawaii. The Fitch–Margoliash method (Fitch and Margoliash 1967) yielded the same most probable tree (Fig. 7) but reversed the order of the second and third most probable trees.

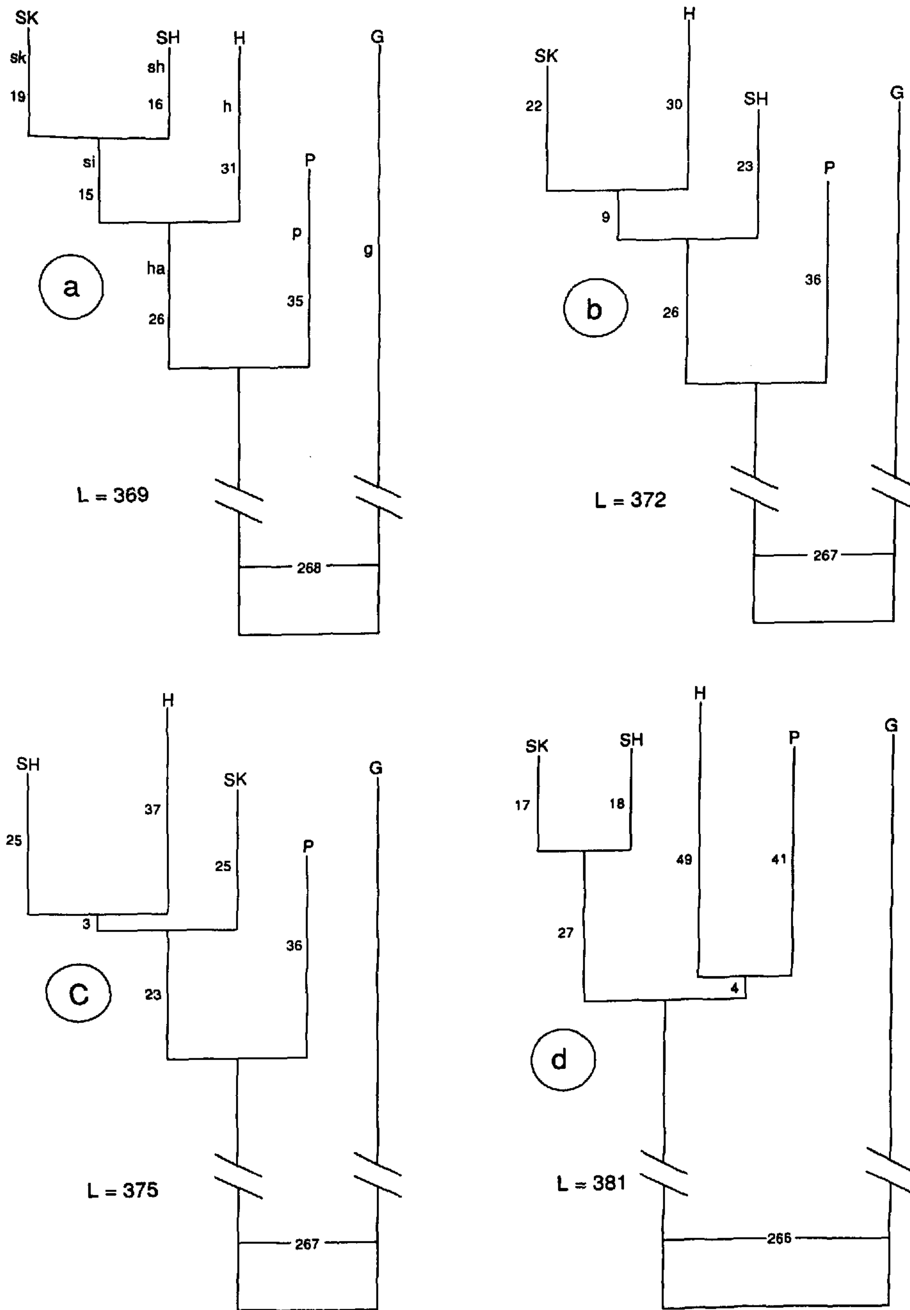


Fig. 6. Most probable gene trees of the 4.1-kb *Pst*I fragment including genes *s18, s15,* and *s19*. The trees were generated using the phylogenetic algorithm using parsimony (PAUP) (Swofford 1985). Total tree lengths (L) are shown to the lower left side of each tree. Branch lengths are scaled according to their distances, which are indicated on the corresponding branches. Deletions and gene correction events, regardless of length, were considered as single mutations. For clarity, with respect to the most probable tree (a), lineages leading to taxa were named after the taxa they lead to. Lineages leading to nodes were named after the lineages the nodes lead to. Hence, the lineage leading to both taxa of *D. silvestris* is called the *silvestris* lineage, and the lineage leading to the taxa in the island of Hawaii is called the Hawaii lineage. Nodes are named, from top to bottom, the *silvestris* node, the Hawaii node, and the *planitibia* node. Code for taxa and lineages: SK = *D. silvestris*-Kona, SH = *D. silvestris*-Hilo, SI = *silvestris*, H = *D. heteroneura*, HA = Hawaii, P = *D. planitibia*, G = *D. grimshawi*.

With respect to these taxa, the tree derived from mitochondrial DNA (mtDNA) restriction analyses (DeSalle et al. 1986) corresponds to the second (parsimony) or third (Fitch–Margoliash) most probable tree for nuclear DNA. The tree holding intraspecific taxa phylogenetically closer is the most probable for both methods. See the legend of Fig. 6 for names and lineages.

Homomorphisms and Excess Polymorphism

It is interesting that if the most probable tree (Fig. 6a) is accepted, only 3 of the 20 potential double base substitutions that are observed (23 are observed with the tree on Fig. 6b, and 26 with the tree on Fig. 6c) are unambiguous, as they involve more than two nucleotide identities (Table 2). The term *homomorphism* is defined here to represent the other 17 potential double base substitutions. A *homomorphism* is the observation of a character that exhibits two states that are distributed in such a way in taxa represented in a phylogenetic tree that would implicate either the independent emergence of one state in two lineages (homoplasy), or a change of one state to another in one lineage followed by the reverse in another. To take a specific example, the homomorphism at position 861 (Table 2) would implicate either an A-to-C change in both the *het-*

Table 2. Potential double base substitutions

	Position	SiK	SiH	Het	Pla	Gri
Homomorphisms						
1	861	A	C	C	A	A
2	1052	T	C	C	T	T
3	1209	A	A	T	T	A
4	1350	T	A	T	A	T
5	1355	A	C	A	C	C
6	1718	C	C	T	C	T
7	1959	C	C	A	A	C
8	2108	G	A	A	G	G
9	2212	T	C	T	C	C
10	2323	A	G	A	G	G
11	2328	C	T	C	T	T
12	2379	G	A	G	A	A
13	2506	A	A	T	T	A
14	2644	C	C	T	C	T
15	2679	A	C	A	C	C
16	3709	T	A	T	T	A
17	3923	T	G	G	G	T
Unambiguous double base substitutions						
18	531	G	T	C	T	–
19	1901	A	G	G	G	T
20	2905	G	G	T	G	A

Abbreviations as in Fig. 2

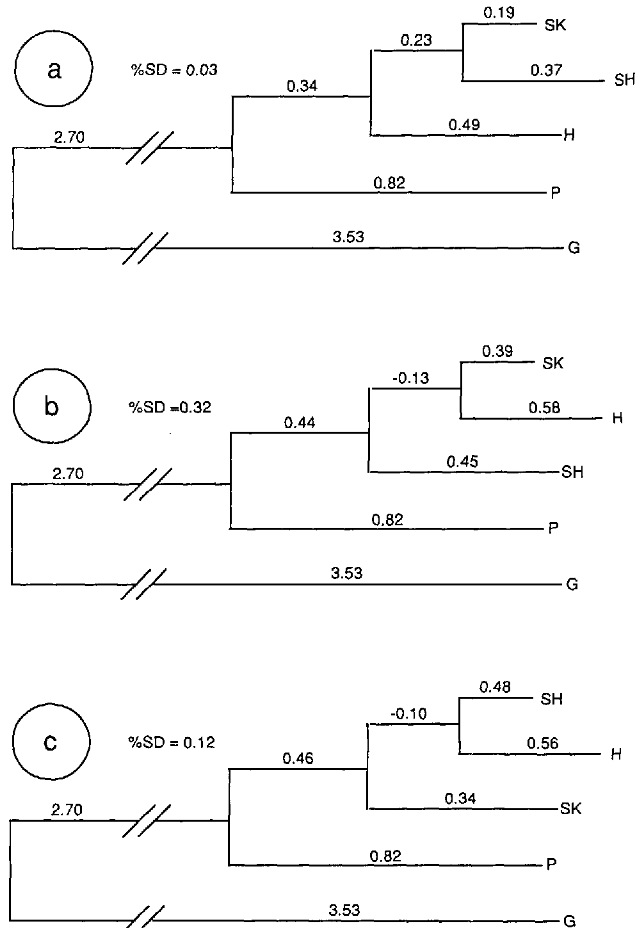


Fig. 7. Fitch–Margoliash most probable trees for the 4.1-kb *Pst*I fragment. Branch lengths are indicated, but scaled only for the most probable tree (a). Deletions and gene correction events were all considered as single mutations. Code for taxa and lineages are: SK = *D. silvestris*-Kona, SH = *D. silvestris*-Hilo, SI = *silvestris*, H = *D. heteroneura*, HA = Hawaii, P = *D. planitibia*, G = *D. grimshawi*.

eroneura and *silvestris*-Hilo lineages (Fig. 6a) or an A-to-C change in the Hawaii lineage followed by the reverse (C-to-A) in the *silvestris*-Kona lineage.

A very important aspect of these homomorphisms is that all nine homomorphisms involving exclusively the taxa in the island of Hawaii, as well as 15 of the 17 homomorphisms are clustered within a 1.8-kb segment (positions 861–2679 of *D. silvestris*-Kona, encompassing *s15* and flanking regions). This leaves only two homomorphisms in the remaining 2.3 kb of DNA, despite the relatively even distribution of point mutations along the whole sequenced fragment (Fig. 5).

Although it is entirely possible that homoplasies or reversals have occurred in the gene tree, the ratio of homomorphisms to unambiguous double base substitutions (17:3) as well as the clustering of these homomorphisms despite an even distribution of base substitutions along the 4.1-kb fragment, makes it highly unlikely that these are the only explanations for all homomorphisms.

In fact, only 2.1 homoplasies/reversals were expected along the 4.1-kb fragment. This expected number was calculated taking into account the proportion of positions substituted along each lineage and the higher rate of transitions over transversions. It was assumed that substitution rates were equal for all sites along the 4.1-kb fragment. Of the 2.1 expected homoplasies/reversals, 1.2 were expected outside, and 0.9 inside the 1.8-kb segment. The assumption of equal substitution rates for all sites does not seem to be a serious flaw, as only two homomorphisms (not far from the 1.2 expected) were observed in the 2.3-kb sequences.

Hence, approximately 14 homomorphisms in the 1.8-kb central segment are left to be explained by possibilities other than homoplasies or reversals. One possibility is the differential polling for a character that was (and may still be) polymorphic in ancestor populations. The clustering of homomorphisms in the 1.8-kb segment could be explained if ancestral polymorphisms were in excess in that segment. Another nonexclusive explanation for homomorphisms is the occurrence of natural interspecific hybridization prior to intraspecific recombination. If a segment within a sequenced fragment has its origin in another taxon, its sequences may present numerous homomorphisms. Thus, interspecific hybridization could also explain the clustering of homomorphisms in the 1.8-kb segment.

Natural hybridization between *D. heteroneura* and *D. silvestris* from the western side of the island of Hawaii has been previously reported (Kaneshiro and Val 1977; Carson et al. 1989). The proximity of four homomorphisms (four in 168 bp, only two in the remaining 3948; Table 2) shared by *D. silvestris*-Kona and *D. heteroneura* is suggestive. However, hybridization between *D. heteroneura* and *D. silvestris*-Kona followed by recombination could only account for up to six homomorphisms, and hybridizations between other pairs of taxa would have to be invoked if an excess of ancestral polymorphisms in the 1.8-kb segment is not to be considered.

Evidence of hybridizations between pairs of taxa other than *D. heteroneura–D. silvestris*-Kona is scant: only a single putative *D. heteroneura–D. silvestris*-Hilo hybrid has been found among 1900 specimens collected (Carson et al. 1989). Furthermore, to invoke hybridizations as the only explanation for the observed excess homomorphisms within the 1.8-kb segment would necessitate the coincidence that recovered extraspecific segments resulting from these hybridizations overlapped considerably with the 1.8-kb segment, but not, by recombination, with sequences out of it. In addition, more than two interspecific hybridizations would be needed to account for all observed homomorphisms in the 1.8-kb segment. Hybridizations between *D. silvestris*-Hilo and *D. heteroneura* or between *D. heteroneura* and *D. planitibia*, for example, could each account only for up to three additional homomorphisms, leaving five homomorphisms in the 1.8-kb segment to be explained by sample error. On the other hand, ancestral polymorphism can explain straightforwardly the uneven distribution of homomorphisms if ancestor populations contained relatively higher levels of polymorphism in a DNA segment that is selectively kept polymorphic.

The occurrence of excess polymorphism in the 1.8-kb segment in populations ancestral to the taxa in the island of Hawaii is suggested by the clustering of the homomorphisms in lineages stemming from the Hawaii node and the corresponding depletion of these homomorphisms in the Hawaii lineage. A test for the constancy of evolutionary rate of the sequenced fragment can be carried out by first constructing the most probable DNA tree, and then scoring the minimum number of base substitutions that must have occurred in each lineage. However, it is important to determine whether excess polymorphism existed in the 1.8-kb DNA segment in the ancestor population represented in the tree by the Hawaii lineage, or before, because in such a case, a large proportion of the base substitutions scored as occurring after the Hawaii lineage could have occurred in reality along or before the Hawaii lineage. This could bias the test of constancy of the

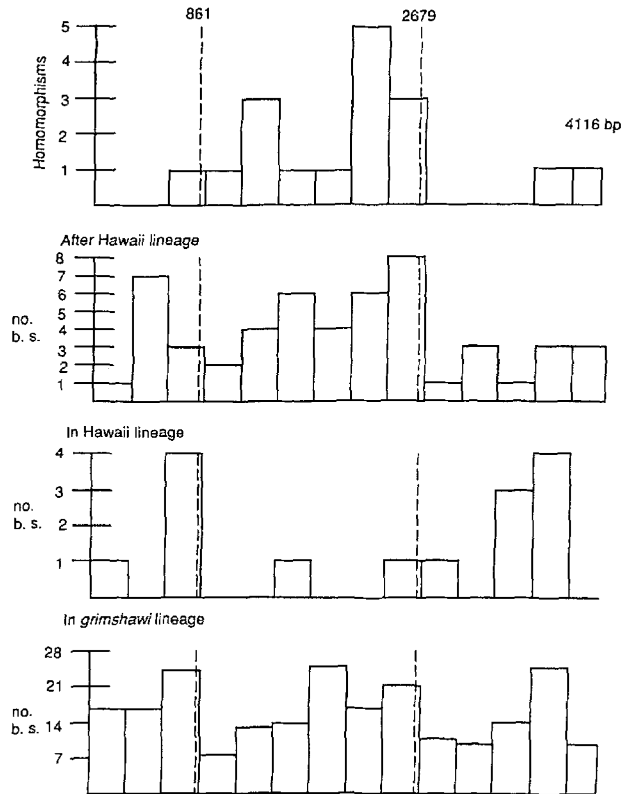


Fig. 8. Distribution of homomorphisms (top histogram) and of base substitutions scored for different lineages along the 4.1-kb chorion DNA *Pst*I fragment. Sites were numbered according to the number of their homolog in the standard *D. silvestris*-Kona fragment. Base substitutions in the lineages between the *planitibia* subgroup taxa were scored using those sequences not missing in any of the taxa, ignoring *D. grimshawi*. On the bottom, all 226 base substitutions between *D. grimshawi* and the *planitibia* subgroup taxa (*grimshawi* lineage) for homologous sequences present in all five taxa are diagrammed in blocks of 300 bp each. The last block is 216 bp long. Dashed lines enclose the 1.8-kb segment where 15 of the 17 homomorphisms are located.

rate of evolution, artificially increasing the rate in all lineages after the Hawaii lineage, and correspondingly decreasing it in the Hawaii lineage.

Such a prediction can be used to examine the possibility of excess polymorphism in the ancestor population. Accordingly, scored base substitutions in all four lineages stemming after the Hawaii lineage would be more frequent in the central 1.8-kb DNA segment, whereas those scored for the Hawaii lineage would have been mostly localized outside of the 1.8-kb segment, in the remaining 2.3 kb of sequence flanking it. No other possibility would generate such a prediction.

Figure 8 shows histograms depicting the positions on the 4.1-kb *Pst*I DNA fragments of base substitutions scored for different lineages. Serving as control, the *grimshawi* lineage shows a very even distribution of base substitutions: the relative proportion of sites that are substituted between the 2.3-kb sequences and the 1.8-kb central segment is

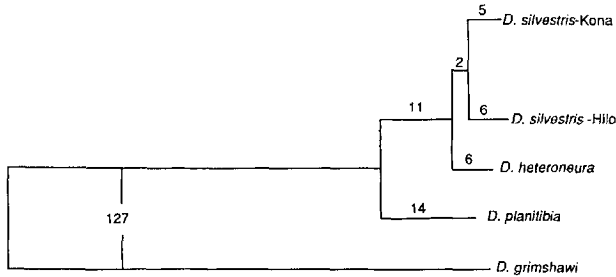


Fig. 9. Scaled gene tree of 2155 bp of DNA sequences including genes *s18* and *s19*. Numbers indicate the minimal number of base substitutions that must have occurred in every lineage for homologous sequences present in each of the five taxa.

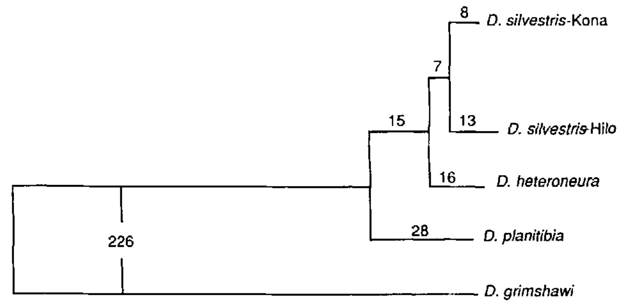


Fig. 10. Scaled gene tree of the 4.1-kb *PstI* fragment including *s18*, *s15*, and *s19* chorion genes. Only sequences present in each of the five taxa are included. Numbers relate to the minimal number of base substitutions that must have occurred in each lineage. Homomorphisms within the central 1.8-kb segment are excluded.

1.00:0.98. The lineages stemming after the Hawaii lineage (second panel from the top) show a relatively higher proportion of scored base substitutions (homomorphisms excluded) in the 1.8-kb central segment (1.00:1.72). Correspondingly, an apparent depletion of scored base substitutions is observed in the 1.8-kb segment for the Hawaii lineage, where the relative ratio of percent base substitution is 1.00: 0.46. Chi-square and *G*-tests show that the distribution of substituted sites is significantly different ($P < 0.025$ for both tests) between the Hawaii lineage and lineages stemming after the Hawaii lineage. No interspecific hybridization could cause such observations.

Thus, it is quite apparent that most of the observed homomorphisms are due to excess polymorphism around and including the *s15* gene in ancestor populations. This ancestral polymorphism may date prior to the *planitibia* subgroup radiation as more homomorphisms are observed involving *D. planitibia* than could reasonably be accounted for by homoplasies and reversals (calculations not shown).

Rate of Substitutions Among All Lineages Is Consistent with the Molecular Clock Hypothesis

The relative rate of base substitutions in 2.3 kb of nucleotide sequence outside of the excessively polymorphic DNA segment seems to be constant among all lineages. Because the excess polymorphism in the 1.8-kb central segment may alter the apparent number of base substitutions in different lineages of the tree, hence generating a misleading profile for constancy of evolutionary rate (see above), it was excluded from the constancy test. Figure 9 shows the minimal number of base substitutions that must have occurred in every lineage for homologous sequences present in each of the five taxa. Because it is not clear on which lineages homomorphisms should be scored, these are ignored. As can be seen, from any node, the number of base substitutions scored between the node and any of the taxa evolving from it is approximately the same. In fact, the observed ranges of base substitutions along lineages stemming from each one of the three nodes of the tree (14–19 from the *planitibia* node, 6–8 from the Hawaii node, and 5–6 from the *silvestris* node) are always smaller than those expected if the ratio of the variance to the mean of the number of substitutions per lineage were equal to two (11.2–22.8 from the *planitibia* node, 3.3–10.7 from the Hawaii node, and 1.7–9.3 from the *silvestris* node). Hence, the observed ranges are within the ranges expected for a molecular clock even when the topology of the tree does not approximate a star phylogeny. The ranges expected for a nonstar phylogeny are greater than for a star phylogeny (Gillespie 1986). It is noteworthy that the excess polymorphism does not significantly affect the rate constancy. If the whole 4.1-kb fragment is used to score the minimal number of base substitutions for each lineage (Fig. 10), the observed ranges (28–35 from the *planitibia* node, 15–20 from the Hawaii node, and 8–13 from the *silvestris* node) are still within the ranges expected for an acceptable molecular clock (20.1–38.9, 11.2–22.8, and 5.9–15.1, respectively). Furthermore, phylogenetic trees (using the 4.1-kb fragment or only the 2.3-kb sequences) constructed by the unweighted pair-group method with arithmetic mean (UPGMA), which assumes a constant rate of evolution, are in agreement with those constructed by the parsimony and Fitch–Margoliash algorithms, which do not. By these criteria, the 4.1-kb fragment, as well as the 2.3 kb of DNA, seems to be evolving at a constant rate in all lineages.

Rate of Nucleotide Substitutions and Species Divergence Times

Given that the rate of base substitutions seems to be relatively constant in all lineages, it could be used to estimate species divergence times. However, excess polymorphism could result in inappropriate

Table 3. Proportion of different nucleotides (above diagonal) and estimated number of substitutions per site (below diagonal) in the 2.3-kb sequences outside of the DNA segment exhibiting an excess of homomorphisms

Taxa	SiK	SiH	Het	Pla	Gri
SiK		5.2	7.0	15.7	66.2
SiH	5.2 ± 1.5		7.0	15.7	66.7
Het	7.0 ± 1.8	7.0 ± 1.8		15.8	66.3
Pla	15.9 ± 2.7	15.9 ± 2.7	16.0 ± 2.7		65.4
Gri	69.5 ± 5.9	70.0 ± 5.9	69.6 ± 5.9	68.5 ± 5.9	

All numbers multiplied by 10^3; abbreviations as in Fig. 2

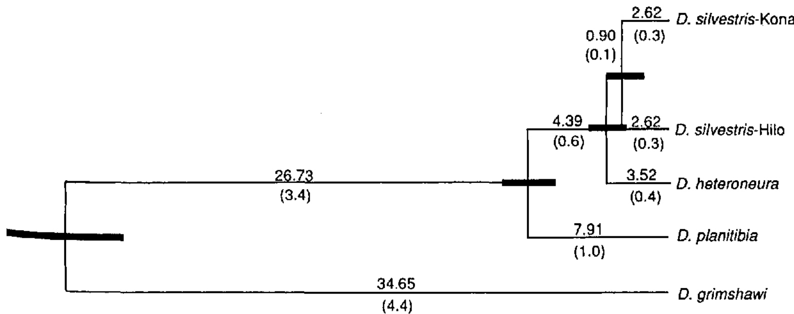


Fig. 11. Scaled gene tree of 2155 bp of DNA sequences containing chorion genes *s18* and *s19* with UPGMA genetic distances. Time distances in million year units are shown in parentheses. Thick lines represent one standard deviation to each side from the mean of the nodes.

scoring of base substitutions in different lineages (Fig. 8) and incorrect estimated divergence times. For this reason, only the sequences of the 2.3 kb surrounding the 1.8-kb excessively polymorphic segment were used for these calculations.

The proportion of different nucleotides between any pair of taxa, and the genetic distances (or estimated number of nucleotide substitutions per site) with standard errors that were used to construct the UPGMA tree for the 2.3-kb sequences are presented in Table 3. The two-parameter model (Kimura 1980) was used to calculate the genetic distances and their standard errors. It allows for two types of base substitutions to be of inhomogeneous rates, as is the case between transitions and transversions (see below). Figure 11 shows the resulting UPGMA tree, which is the same tree constructed for the 2.3 kb, or for the whole 4.1-kb fragment (Fig. 6a), using parsimony, or for the whole 4.1-kb fragment using the Fitch–Margoliash method (Fig. 7a). However, the *D. silvestris* taxa are not significantly closer to each other than to *D. heteroneura.*

Given the average genetic distance between *D. grimshawi* and members of the *planitibia* subgroup (Table 3), the nucleotide substitution rate in Hawaiian *Drosophila* was estimated to be 0.8%/Myr by assuming that these species groups diverged 4.4 Myr ago. This is the average for estimates for the geological age of the island of Kauai (McDougall 1979), where these species are though to have diverged (Carson and Yoon 1982).

Using this nucleotide substitution rate, the UPGMA tree was translated into time units (Fig. 11, in parentheses). The taxa in the island of Hawaii are estimated to have diverged from *D. planitibia* 1.0 ± 0.2 Myr ago. Because *D. planitibia* is endemic to the island of Maui, the estimate is in good agreement with the expected divergence time: good geological estimates for Maui exist, West Maui and East Maui being 1.3 and 0.9 Myr old, respectively (McDougall 1964). Similarly, *D. heteroneura* and *D. silvestris* must have emerged from the same founder event or from founder events originating from very closely related populations. This is so because they share a unique chromosomal inversion polymorphism in chromosone 3 [homologous to the ancestral chromosome element *B* (Müller 1940; Loukas and Kafatos 1986; Whiting et al. 1989)], and because mtDNA studies (DeSalle et al. 1986) show that if the species were the product of different founder events, they must have been close enough to hybridize to the extent required to homogenize their mitochondrial genomes. Thus, the time of divergence of *D. silvestris* and *D. heteroneura* can be approximated to the geological age of Hualalai [0.4 Myr (Funkhouser et al. 1968; McDougall 1979)], the volcano they colonized (Carson and Bryant 1979; Carson 1982a; Kaneshiro and Kurihara 1982; DeSalle et al. 1986). Remarkably, the UPGMA tree again agrees with the estimate, placing the time of divergence at 0.4 ± 0.1 Myr ago.

Transitions Are Twice as Probable as Transversions

Table 1 shows the number of transitions and transversions inferred for each one of the paleoheterodiverging regions. The transition frequencies are very close to one-half throughout the sequence, instead of one-third as would be expected on a random basis. The average transition frequency of noncod-

Table 4. Base frequencies of noncoding regions with more than 10 bp changes with expected and observed transition frequencies

	T	C	A	G	Transition frequencies	
					Expected	Observed
Central *s18–s15*	0.35	0.14	0.32	0.19	0.30	0.50 ± 0.08
Central *s15–s19*	0.29	0.17	0.36	0.19	0.31	0.59 ± 0.06
Proximal *s18–s15*	0.32	0.16	0.35	0.17	0.30	0.38 ± 0.10
Proximal *s15–s19*	0.21	0.17	0.33	0.29	0.37	0.38 ± 0.13
Intron *s15*	0.38	0.21	0.23	0.18		
Intron *s19*	0.31	0.26	0.24	0.20		
Pooled introns	0.34	0.23	0.24	0.19	0.34	0.77 ± 0.12

ing regions (51%) is not different from that of the coding regions (52%) and coincides with the 52% calculated for noncoding regions of closely related members of the A and B chorion multigene families of the wild silkmoth *Antheraea polyphemus* (Jones and Kafatos 1982). Only 5 of the 20 regions deviate by more than one standard error from a frequency of 50%: both introns and the distal and central *s15–s19* intergenic regions show an excess of transitions, whereas the *s18–s15* proximal intergenic region shows transversions and transitions occurring at frequencies closer to the expected 2:1 ratio.

The underlying rate of transitions seems to be the same as transversions. To estimate their ratio, the least constrained of the regions must be used to avoid the noise caused by natural selection, as seems to be the case for the *Adh* locus (Rowan and Hunt, unpublished). Although the chorion coding regions are the fastest evolving in the cluster (see below), they are inappropriate. Their substitutional pathways are ruled by the nature of the genetic code and heavily biased. Thus, it becomes necessary to resort to the use of the central intergenic regions.

Besides mutation rate, three possibly interdependent selected aspects of the sequence are likely to affect the transition/transversion rate: (1) the base composition, (2) the nonrandom distribution of nucleotides, and (3) base substitution limitations for the functional structure of the region. Because only the first feature can be determined with accuracy, the estimate for the expected frequency of transitions is based here solely on base composition and assumes that the base composition is in equilibrium, that all bases are equally stable, and that all mutational pathways are equally probable. Then, the expected frequency of transitions is

$$E = \Sigma (f_i f_j / 1 - f_i) \quad (1)$$

where f_i and f_j are the frequencies of each base and its transitional pair, respectively.

The expected and observed transition frequencies for noncoding regions of *D. silvestris*-Kona with more than 10 base substitutions are shown in Table 4. It is clear that in the central intergenic regions the transition frequency is at least 0.50 as opposed to the expected 0.33. The proximal intergenic regions are closer to a frequency of 0.33 than to 0.50, and the introns, in a class by themselves, favor transitions with a frequency of 0.77. The observed versus expected frequencies of changes of each base with every other base were analyzed in detail for each of these regions. For example, it can be deduced from Table 4 that Cs account for 35% of the non-T bases in introns. It would then be expected that 35% of the base changes involving a T would also involve a C and be a transition. These expectations can be built for T with respect to its two transversion complements, and for each base with respect to each of the other three bases in each one of the regions as well. Thus, four tests corresponding to the four bases can be undertaken for each one of the regions presented in Table 4, for a total of 20 (introns are pooled). For each test, there are two transversion entries (a given base can undergo a transversion toward any of two different bases) but only one transition entry to be tested. In 19 out of 20 entries, transitions have higher observed than expected frequencies, whereas transversions are higher than expected in only 5 out of 40 entries, 3 of these in the proximal *s15–s19* region. In central intergenic regions, all 8 transitional entries are favored but the same is true for only 1 of the 16 transversional entries.

As the largest and least constrained of the regions (see Discussion), the central intergenic regions are here taken to conclude that, mutationally speaking, transitions occur at least as often as transversions; selection might drive the observed ratios toward more transversions in the proximal regions and toward more transitions in the introns, notwithstanding the fairly small sample in these regions. Furthermore, the mutational bias toward transitions cannot be ascribed to only one of the transitional pairs (T/C or A/G).

Drosophila Autosomal Chorion Gene Products Are the Fastest Evolving Proteins Described

Figure 5 shows that the coding regions of chorion genes are as divergent as the central intergenic

Table 5. Percent of nucleotides substituted in replacement and synonymous sites present in each of the five Hawaiian taxa

Gene	Amino acids	Replacement sites[a]	Synonymous sites[a]	Replacement substitutions	Synonymous substitutions	Substitutions in replacement sites	Substitutions in synonymous sites
s18	157	335.75	135.25	25	15	7.4	11.1
s15	94	199.25	82.75	28	12	14.1	14.5
s19[b]	182	391.5	154.5	36	20	9.2	12.9
Total	433	926.5	372.5	89	47	9.6	12.6

[a] The number of replacement and synonymous sites was inferred from hypothetical coding sequences constructed for the phylogenetic tree node leading to all *planitibia* subgroup taxa

[b] The valine codon interrupted by an insertion in the *D. grimshawi s19* gene was not considered, as it stops being one in this species

Table 6. Aggregate number of replacement and synonymous substitutions inferred from all chorion sequences

Gene	Replacement substitutions	Synonymous substitutions	Repl : Syn
s18	25	15	1.7:1
s15	33	13	2.5:1
s19	38	21	1.8:1
Total	96	49	2.0:1

regions. To examine the differential rate of divergence between synonymous and nonsynonymous positions, hypothetical coding sequences of the second exon (those of first exons are perfectly conserved) of all three genes were constructed for the *planitibia* node (Fig. 6a) and aligned to all five taxa. The numbers of synonymous and replacement sites in matched segments were then calculated for the constructed sequences, presuming that any transition is mutationally twice as likely as any transversion (see above). Thus, third base positions in twofold degenerate codons were considered 50% synonymous, and in threefold degenerate codons, as well as in first positions of sixfold degenerate codons, were considered to be either 50% or 75% synonymous, depending on their identity in the constructed sequence. For example, the third base position of the isoleucine codon ATA is 50% synonymous, as only a transversion (50% chance of occurring) would maintain the codon synonymous, but for ATT or ATC, any transition (50% chance) or one of two possible transversions (25% chance) would be a silent substitution; therefore, that position is corrected as being 50 + 25 = 75% synonymous and 25% replacement.

From the node of the *planitibia* subgroup taxa, base substitutions were scored in every lineage. In case of multiple substitutions within codons, the smallest number of replacement substitutions was assumed to have occurred. Table 5 shows a grand total of 89 replacement substitutions and 47 synonymous substitutions in segments that can be matched for all five taxa. The corresponding numbers for all sequences are 96 and 49 (Table 6). Thus, replacement and synonymous substitutions occur at approximately a 2.0:1 ratio, close to the 2.5:1 (number of replacement sites : number of synonymous sites) ratio expected if synonymous and nonsynonymous substitutions were equally acceptable. If only the number of base substitutions between *D. silvestris*-Kona and *D. grimshawi* are considered, making a *G*-test possible, it can be demonstrated that the divergence rate in replacement sites is not significantly different for any of the three genes to that in synonymous sites.

The relative acceptability of replacement mutations is especially high for *s15*. In fact, the *s15* ratio for the total numbers of replacement to synonymous substitutions (2.5:1; Table 4) is virtually identical to that expected for replacement to synonymous sites (2.4:1).

To estimate the rate of divergence of chorion gene products, the amino acid replacement substitutions between *D. silvestris*-Kona and *D. grimshawi* in sequences aligned between these species were calculated (82) and were divided by the total number of amino acids compared (437) and twice the divergence time (4.4 Myr). According to these calculations, the average amino acid residue of chorion proteins is diverging at a rate of 2.1×10^{-8} replacement/site-year. This is twice the rate of 0.9×10^{-8} calculated for fibrinopeptides (Dickerson 1971) in mammals, the fastest evolving proteins described to date.

Discussion

Length Mutations Are Related to Direct Repeats

The vast majority of length mutations in the chorion gene cluster, 40 out of 43, are associated with direct

repeats. The association with tandem direct repeats is especially strong. Unlike previous in vitro observations with the large fragment of *Escherichia coli* DNA polymerase I (de Boer and Ripley 1988), all single base pair length mutations can be attributed to misaligned pairing (Streisinger et al. 1966) in homopolymer runs. Complex substitutions–deletions of adenine dimers for a single cytosine (de Boer and Ripley 1988) were not observed.

Although the strong association of length mutations with direct repeats could be explained by unequal crossing over (Lerman 1963; Magni 1963), slippage and mispairing of DNA strands during replication is a more likely explanation (Farabaugh et al. 1978; Albertini et al. 1982). Specific models of slippage and excision have been proposed to explain deletions (Efstratiadis et al. 1980), and alternative models have been suggested that explain both reiteration and deletion of tandem repeats, as well as deletion of nontandem repeats (Jones and Kafatos 1982).

Length Mutations in Coding and Noncoding Regions

The distribution of length mutations in the coding regions is reminiscent of the chorion genes in silkmoths, where length mutations are abundant in the glycine–tyrosine-rich arms of the proteins, but scarce in the central alanine–valine-rich region (Jones and Kafatos 1982). Specifically, 14 out of the 15 length mutations in coding sequences occur in glycine–tyrosine-rich repetitive regions, but none in the stretches of alanine residues present in these same genes.

Another aspect of the evolution of the cluster is that distant species comparisons demonstrate outstanding conservation in the size and gene order of the cluster (Martínez-Cruzado et al. 1988). In the Hawaiian *Drosophila* this length conservation is quite dramatic as well; although *D. silvestris*-Kona coding regions are 123 bp longer than in *D. grimshawi*, the difference in noncoding regions is only 16 bp. Conservation of intergenic regions may be related to the presence of elements that enhance specific replication of the cluster during choriogenesis (Delidakis and Kafatos 1987, 1989; Swimmer et al. 1989), or to requirements that this DNA be "polite," respecting functions for which it is not essential (Zuckerkandl 1986).

Excess Polymorphism Around s15

Excess ancestral polymorphism is certainly the best explanation for the observed excess of homomorphisms in the 1.8-kb segment. No other phenomenon could explain the uneven distribution among lineages of base substitutions in the 1.8-kb segment (Fig. 8). Furthermore, if the observed spatial clustering of homomorphisms were to be explained exclusively by interspecific hybridization and homoplasies or reversals, an unlikely combination of coincidences and sample errors would have to be invoked (see above).

In addition to the excess of ancestral polymorphisms in the 1.8-kb segment, it is very likely that the four very closely positioned homomorphisms between *D. silvestris*-Kona and *D. heteroneura* (at positions from 2212 to 2379, Fig. 2) are the result of interspecific hybridization followed by recombination. No nucleotide character identical for *D. silvestris*-Kona and *D. silvestris*-Hilo but different for the rest of the species was observed in between these clustered homomorphisms, as would be expected if these homomorphisms were the result of hybridization followed by recombination. By contrast, the other two *D. heteroneura*–*D. silvestris*-Kona homomorphisms (at positions 1355 and 2679) were separated from this cluster by nucleotide characters identical exclusively for both *D. silvestris* taxa (at positions 1454, 1489, 2084, and 2157 on one side; and at positions 2512 and 2515 on the other) and may thus be the result of ancestor polymorphisms.

Phylogeny of the Taxa in Hawaii

It is noteworthy that the most probable phylogenetic trees based on mitochondrial (DeSalle et al. 1986) and nuclear DNA (this report) disagree. This should not be surprising for many reasons. First, the distance between the Hawaii and *silvestris* node for the UPGMA tree of the 2.3-kb sequences is less than one standard deviation (Fig. 11). Although a more thorough study of polymorphism should be undertaken, two additional lines of evidence suggest that the nuclear DNAs of both *D. silvestris* taxa are more closely related to each other than to *D. heteroneura*. These are the number of length mutations separating the taxa (Fig. 3), and the fact that the sequenced *D. silvestris*-Kona and *D. heteroneura* isofemale lines are precisely the ones for which mtDNA studies suggest recent hybridization (DeSalle et al. 1986). The mtDNA restriction maps of these isofemale lines are more closely related to each other than to any other line from any of the three taxa examined. It sho uld then be expected that if other isofemale lines are sequenced, the observed distance between *D. silvestris*-Kona and *D. heteroneura* sequences would never be less than the one found here.

This would not be the first time that nuclear and mtDNA phylogenies are at odds. The mtDNA phylogeny places *D. differens*, another member of the *planitibia* subgroup, as ancestor to *D. planitibia* (DeSalle and Giddings 1986), whereas DNA:DNA

reassociation studies (Hunt and Carson 1982) suggest these species to be monophyletic.

Similarly, *Drosophila pseudoobscura* and *Drosophila persimilis* populations in hybrid zones show homogeneity in their mtDNA (Powell 1983). However, their allozymes are clearly distinct (Coyne 1976). This is explained by the effective fourfold smaller population size of the mitochondrial genome with respect to the nuclear genome (Birky et al. 1983), which exposes the mitochondrial genome to a higher degree of random drift and to a faster flow between the genomes of two populations (Takahata and Slatkin 1984). Inasmuch as mtDNA is maternally inherited, sexually asymmetric hybridization between two populations, as found in Hawaiian *Drosophila* (Kaneshiro 1976; Carson et al. 1989), will accelerate the mitochondrial genomes' homogenization process. In the *D. pseudoobscura–D. persimilis* hybrid zone, 1 out of 10,000 individuals collected are hybrids (Powell 1983). In Kahuku, when the *D. silvestris*-Kona specimen was collected, 1–2% of all individuals were showing morphological evidence of hybrid origin (Kaneshiro and Val 1977). It is thus clear that the potential for *D. heteroneura* and *D. silvestris*-Kona mtDNA homogenization exists and that in such a case other taxonomic data should be employed to yield a definitive evolutionary scheme for the taxa in Hawaii.

Rate of Chorion DNA Evolution

The rate of evolution of the chorion cluster as a whole seems to be quite representative of the genome. Hunt et al. (1981) estimated by DNA:DNA hybridization studies that the single copy genomes of *D. silvestris* and *D. heteroneura* would differ in 0.55% of their sites. This was assuming that a 1°C change in melting temperature is equivalent to a 1% difference in base sequences. There is now strong evidence that a 1°C change corresponds more properly to a divergence of 1.7% (Caccone et al. 1988). If so, the predicted divergence between *D. silvestris* and *D. heteroneura* becomes 0.94%, very close to the average of 1.00% observed here for the chorion cluster.

The rate of synonymous substitutions (1.2%/Myr for sequences present in each of the five taxa) also falls within the window of expectation. For example, Sharp and Li (1989) estimate the average silent substitution rate for *Drosophila* to be 1.6%/Myr. Independently, they show chorion genes to approximate the average rate in *Drosophila*. However, the generation length of Hawaiian *Drosophila* can be four times that of other drosophilids. The correlation between generation length and DNA substitution rate has been shown to be negative, if less than linear (Wu and Li 1985; Li et al. 1987), and it could explain perfectly the discrepancy between the observed rate and Sharp and Li's estimate. Similarly, if the long generation of Hawaiian *Drosophila* is taken into account along with the correction of Caccone et al. (1988) for DNA : DNA hybridization studies, the estimated divergence rate for the cluster (0.8%/Myr) approximates the estimated rate of 0.7%/Myr for genomic single-copy DNA in *Drosophila* (Britten 1986). Once again, in accordance with what could have been predicted (Sharp and Li 1989), the synonymous substitution rate of 1.2%/Myr is not less than half the 2%/Myr rate estimated for mtDNA (DeSalle et al. 1987) in *Drosophila*. Finally, the estimated divergence rate and estimated divergence times for Hawaiian species are mutually reinforceable. Specifically, East and West Maui are reliably estimated to have emerged 1.3 and 0.9 Myr ago (McDougall 1964). Using the estimated rate, *D. planitibia* concurrently diverged just over 1.0 Myr ago from the taxa in Hawaii (Fig. 11). Moreover, *D. silvestris* and *D. heteroneura* are estimated here to have diverged 0.4 Myr ago. This is precisely the geological age of Hualalai, the volcano their ancestors colonized.

In this context, it is not likely that the observed synonymous substitution rate of 1.2% is an overestimate. DeSalle and Templeton (1988) found that the species utilized here, belonging to the beta sublineage (Spieth 1982), have a threefold higher mtDNA substitution rate than those of the alpha sublineage. The hypothesis is that fixation rate in the beta sublineage is accelerated over the alpha sublineage by the occurrence of repeated founder events. Chorion DNA sequence analysis in the alpha sublineage should clarify whether nuclear DNA substitution rates exhibit the same evolutionary disparity.

A Rapidly Evolving Cluster

The rapid evolution of the coding regions of the genes is underscored not only by comparison with noncoding regions, but also by the substitution rates of replacement and synonymous positions. Two replacement substitutions occur for every synonymous one (Tables 5 and 6). By contrast, the silkmoth chorion genes, with essentially the same amino acid composition, exhibit three synonymous for every replacement substitution (Jones et al. 1979; Jones and Kafatos 1982).

As discussed previously, the *Drosophila* autosomal chorion genes are the fastest evolving proteins known. Unequal generation time has been invoked to explain differential substitution rates in mammals (Wu and Li 1985). Inasmuch as the evolutionary rate of selectively neutral DNA regions can be taken as a measure of relative average generation length,

the fact that the rate of divergence in synonymous sites (1.2%/Myr) is only two- or three fold that of vertebrates (Perler et al. 1980; Miyata et al. 1980; Li et al. 1987) suggests that the chorion genes may evolve as fast as vertebrate fibrinopeptides even per generation.

The observation of gene conversion in the well-known chorion multigene families of silkmoths (Iatrou et al. 1984; Rodakis et al. 1984; Eickbush and Burke 1985) would raise the possibility that the rapid evolution of the chorion genes in Hawaiian *Drosophila* would be the consequence of gene conversion (Dover 1982). In silkmoths, conversion is observed in regions that have common features within and even between gene families (Lecanidou et al. 1986). However, this does not hold true for *Drosophila* chorion genes. Each gene is clearly distinct, and shows distinct domains conserved in distantly related species (Martínez-Cruzado et al., unpublished). Possible similarities that could point to gene conversion were not found in searches using computer programs (Pustell and Kafatos 1984). Furthermore, gene conversion would not explain the observed high ratio of replacement to synonymous substitutions.

Evolutionary Forces Operating in the Chorion Cluster

In Hawaiian *Drosophila*, genetic drift seems to be playing a larger role in the evolution of the autosomal chorion gene cluster than positive selection. For one, the cluster is evolving at a very constant rate among all lineages examined. This is established by the small ratio of the variance to the mean number of base substitutions per lineage stemming from any one node; and by the generation of identical phylogenetic tree topologies from methods assuming or not assuming constant rates. The constancy of the rate of divergence is substantiated by the accurate fit of estimated species divergence times upon a priori predictions.

Because the neutral theory predicts that the rate of substitution would be equal to the rate of neutral mutations, it could be expected that the evolutionary rate of regions with different functions would be inversely porportional to the degree of selective constraints operating on them (Wilson et al. 1977). Under this light, distantly related species comparisons (Martínez-Cruzado et al. 1988; Swimmer et al. 1990) would indicate that the most constrained regions are generally, in descending order, the first exon, the proximal 5′ flanking region, the second coding region, and finally, the 3′ flanking regions and 3′ untranslated regions each with a minimal degree of conservation. Introns or sequences halfway between genes do not show any detectable conservation. If neutralism were to dominate chorion gene evolution throughout the genus, the order of evolutionary rate of these paleoheterodiverging regions would be expected to remain the same (Wilson et al. 1977) in the Hawaiian *Drosophila*.

These predictions are half fulfilled. The order of evolutionary rate in Hawaiian flies is the same, with the prominent exception of the second coding regions. Central intergenic regions evolve as fast as introns, followed by distal regions and 3′ untranslated sequences together. The proximal regions and the first intron are the most conserved.

It is possible to postulate that the second coding regions have been subjected to positive selection in the Hawaiian *Drosophila*. Positive selection models like founder-flush (Carson 1968, 1975, 1982b) or genetic transilience (Templeton 1980) have been invoked to explain the high frequency of speciation in Hawaiian *Drosophila*; although, if founder effects are always more intense for X-linked loci because of male hemizygosity (Carson and Templeton 1984), they should be particularly moderate for chorion genes as these should not be under selective pressure in males. The alternative hypothesis, which this author prefers, is that large segments within coding regions are undergoing circular evolution in the genus, evolving rapidly without becoming very different over time. Only conservative replacements would be permitted in these segments, but there would be many possible conservative replacements with almost identical selective value. Distant species divergences would be at saturation levels, showing conservation for only highly constrained sequences. In that way, the high rate of substitution of the lightly constrained sequences would only be detectable in close species comparisons. A detailed study of the one-step amino acid replacements occurring in Hawaiian *Drosophila* suggests that this may be the case (Martínez-Cruzado et al., unpublished). This model would predict that relative rates of divergence between genes would conform to the relative concentration of such lightly constrained segments in their sequences. Chorion genes having a large proportion of these segments would be more different between any pair of species than other genes. Then, distantly related species comparisons would predict, as observed, that *s15* would diverge faster than *s19* and than *s18*, in that order, in the Hawaiian *Drosophila*. Unfortunately, no predictions could be made for the positive selection model.

It may be pointed out that additional circumstantial evidence suggests that chorion genes may be highly tolerant of amino acid replacements in *D. melanogaster*. For example, both base differences found in the chorion coding regions of two Oregon-R chromosomes result in amino acid replacements (Levine and Spradling 1985; Wong et al. 1985). Fur-

ther intraspecific chorion sequence comparisons involving continental and Hawaiian species should discern the evolutionary forces operating on the cluster.

Selection and Transition/Transversion Mutation Rate

Several studies between relatively distantly related eukaryotic species have stated that transversions are as probable as transitions, both in nuclear (Jukes 1987) and mitochondrial (Clary and Wolstenholme 1987) DNA. However, in relatively distant interspecies comparisons where multiple base substitutions are expected to be common, the rate of transitions over transversions will be grossly underestimated (Brown et al. 1982; Shoemaker and Fitch 1989). The comparison presented here, comprising species whose sequences differ in not more than 6.8% of their sites, avoids that problem and concludes that any transition is about twice as likely as any transversion.

The large transitional bias in the introns is not unexpected as the distribution of nucleotides in their sequences is nonrandom: the region between the branchpoint and the acceptor sites is particularly pyrimidine-rich (Frendewey and Keller 1985). It is interesting that both proximal 5′ flanking regions show the lowest transition : transversion ratio in the whole fragment. This apparent transversional bias may be due to the particular functional constraints of these regions.

Acknowledgments. I thank J. Hunt and M. Kambysellis for the genomic libraries and for their interest; M. Power for screening one of the libraries; C. Swimmer and M.G. Fenerjian for allowing me to analyze unpublished data; F.C. Kafatos, R.C. Lewontin, J. Hey, J. Powell, R. Hudson, J. Hunt, J.C. Masters, A.C. Wilson, M. Kreitman, F. Bird, C. Swimmer, and Y. Grinblat for their interest and advice or useful comments on the manuscript; and B. Klumpar, M. Youk-See, and E. Fenerjian for assistance in preparation of the manuscript. This work was made possible by an NIH grant to R.C. Lewontin and an ACS grant to F.C. Kafatos.

References

Albertini AM, Hofer M, Calos MP, Miller JH (1982) On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell 29:319–328

Biggin MD, Gibson JJ, Hong GF (1983) Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 80:3963–3965

Birky CW, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics 103: 513–527

Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol 18:225–239

Caccone A, DeSalle R, Powell JR (1988) Calibration of the change in thermal stability of DNA duplexes and degree of base pair mismatch. J Mol Evol 27:212–216

Carson HL (1968) The population flush and its genetic consequences. In: Lewontin RC (ed) Population biology and evolution. Syracuse Univ Press, New York, pp 123–137

Carson HL (1971) Polytene chromosome relationships in Hawaiian species of *Drosophila* V. Additions to the chromosomal phylogeny of the picture-winged species. Univ Tex Publ 7103:183–191

Carson HL (1975) The genetics of speciation at the diploid level. Am Nat 109:83–92

Carson HL (1982a) Evolution of *Drosophila* on the newer Hawaiian volcanoes. Heredity 48:3–25

Carson HL (1982b) Speciation as a major reorganization of polygenic balances. In: Barigozzi C (ed) Mechanisms of speciation. Liss, New York, pp 411–433

Carson HL (1983) Chromosomal sequences and interisland colonizations in Hawaiian *Drosophila*. Genetics 103:465--482

Carson HL, Bryant PJ (1979) Genetic variation of Hawaiian *Drosophila*. VI. Change in a secondary sexual character as evidence of incipient speciation in *Drosophila*. Proc Natl Acad Sci USA 57:1280–1285

Carson HL, Kaneshiro KY (1976) *Drosophila* of Hawaii: systematics and ecological genetics. Annu Rev Ecol Syst 7:311–345

Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the founding of new populations. Annu Rev Ecol Syst 15:97–131

Carson HL, Yoon JS (1982) Genetics and evolution of Hawaiian *Drosophila*. In: Ashburner M, Carson HL, Thompson JN Jr (eds) The genetics and biology of *Drosophila*, vol 3b. Academic Press, New York, pp 298–344

Carson HL, Kaneshiro KY, Val FC (1989) Natural hybridization between the sympatric Hawaiian species *Drosophila silvestris* and *Drosophila heteroneura*. Evolution 43:190–203

Clary DO, Wolstenholme DR (1987) *Drosophila* mitochondrial DNA: conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J Mol Evol 25:116–125

Coyne JA (1976) Lack of genic similarity between two sibling species of *Drosophila* as revealed by varied techniques. Genetics 84:593–607

Coyne JA, Kreitman M (1986) Evolutionary genetics of two sibling species, *Drosophila simulans* and *D. sechellia*. Evolution 40:673–691

de Boer JG, Ripley LS (1988) An *in vitro* assay for frameshift mutations: hotspots for deletions of 1 bp by Klenow-fragment polymerase share a consensus DNA sequence. Genetics 118: 181–191

Delidakis C, Kafatos FC (1987) Amplification of a chorion gene cluster in *Drosophila* is subject to multiple *cis*-regulatory elements and to long-range position effects. J Mol Biol 197:11–26

Delidakis C, Kafatos FC (1989) Amplification enhancers and replication origins in the autosomal chorion gene cluster of *Drosophila*. EMBO J 8:891–901

DeSalle R, Giddings LV (1986) Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian *Drosophila*. Proc Natl Acad Sci USA 83:6902–6906

DeSalle R, Templeton AR (1988) Founder effects and the rate of mitochondrial DNA evolution in Hawaiian *Drosophila*. Evolution 42:1076–1084

DeSalle R, Giddings LV, Kaneshiro KY (1986) Mitochondrial DNA variability in natural populations of Hawaiian *Dro-*

sophila. II. Genetic and phylogenetic relationships of natural populations of *D. silvestris* and *D. heteroneura.* Heredity 56: 87–96

DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila.* J Mol Evol 26:157–164

Dickerson RE (1971) The structure of cytochrome c and the rates of molecular evolution. J Mol Evol 1:26–45

Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, Deriel JK, Forget BG, Weissman SM, Slightom JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot NJ (1980) The structure and evolution of the human β-globin gene family. Cell 21:653–668

Eickbush TH, Burke WD (1985) Silkmoth chorion gene families contain patchwork patterns of sequence homology. Proc Natl Acad Sci USA 82:2814–2818

Farabaugh PJ, Schmeissner U, Hoffer M, Miller JH (1978) Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lac* I gene of *Escherichia coli.* J Mol Biol 126:847–863

Fenerjian MG, Martínez-Cruzado JC, Swimmer C, King D, Kafatos FC (1989) Evolution of the autosomal chorion locus in *Drosophila.* II. Chorion gene expression and sequence comparisons of the *s16* and *s19* genes in evolutionarily distantly related species. J Mol Evol 29:108–125

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Frendewey D, Keller W (1985) Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell 42:355–367

Funkhouser JG, Barnes IL, Naughton JJ (1968) The determination of a series of ages of Hawaiian volcanoes by the potassium–argon method. Pac Sci 22:369–372

Gillespie JH (1986) Natural selection and the molecular clock. Mol Biol Evol 3:138–155

Hayashida H, Miyata T (1983) Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. Proc Natl Acad Sci USA 80:2671–2675

Henikoff S (1987) Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol 155:156–165

Hunt JA, Carson HL (1982) Evolutionary relationships of four species of Hawaiian *Drosophila* as measured by DNA reassociation. Genetics 104:353–364

Hunt JA, Hall TJ, Britten RJ (1981) Evolutionary distances in Hawaiian *Drosophila* measured by DNA reassociation. J Mol Evol 17:361–367

Iatrou K, Tsitilou SG, Kafatos FC (1984) DNA sequence transfer between two high-cysteine chorion gene families in the silkmoth *Bombyx mori.* Proc Natl Acad Sci USA 81:4452–4456

Jones CW, Kafatos FC (1982) Accepted mutations in a gene family: evolutionary diversification of duplicated DNA. J Mol Evol 19:87–103

Jones CW, Rosenthal N, Rodakis GC, Kafatos FC (1979) Evolution of two major chorion multigene families as inferred from clone cDNA and protein sequences. Cell 18:1317–1332

Jukes TH (1987) Transitions, transversions, and the molecular evolutionary clock. J Mol Evol 26:87–98

Jukes TH, Bhushan V (1986) Silent nucleotide substitutions and G+C content of some mitochondrial and bacterial genes. J Mol Evol 24:39–44

Kaneshiro KY (1976) Ethological isolation and phylogeny in the *planitibia* subgroup of Hawaiian *Drosophila.* Evolution 30:740–745

Kaneshiro KY, Kurihara JS (1982) Sequential differentiation of sexual behavior in populations of *Drosophila silvestris.* Pac Sci 35:177–183

Kaneshiro KY, Val FC (1977) Natural hybridization between a sympatric pair of Hawaiian *Drosophila.* Am Nat 111:897–902

Kaplan N, Langley CH (1979) A new estimate of sequence divergence of mitochondrial DNA using restriction endonuclease mapping. J Mol Evol 13:295–304

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Lecanidou R, Rodakis GC, Eickbush TH, Kafatos FC (1986) Evolution of the silkmoth chorion gene superfamily: gene families CA and CB. Proc Natl Acad Sci USA 83:6514–6518

Lerman LS (1963) The structure of the DNA–acridine complex. Proc Natl Acad Sci USA 49:94–102

Levine J, Spradling A (1985) DNA sequence of a 3.8 kilobase pair region controlling *Drosophila* chorion gene amplification. Chromosoma 92:136–142

Levinson A, Silver D, Seed B (1984) Minimal size plasmids containing an M13 origin for production of single-stranded transducing particles. J Mol Appl Genet 2:507–517

Li W-H, Tanimura M, Sharp PM (1987) An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J Mol Evol 25:330–342

Loukas M, Kafatos FC (1986) The actin loci in the genus *Drosophila*: establishment of chromosomal homologies among distantly related species by *in situ* hybridization. Chromosoma 94:297–308

Magni GE (1963) The origin of spontaneous mutations during meiosis. Proc Natl Acad Sci USA 50:975–980

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY

Martínez-Cruzado JC, Swimmer C, Fenerjian MG, Kafatos FC (1988) Evolution of the autosomal chorion locus in *Drosophila.* I. General organization of the locus and sequence comparisons of genes *s15* and *s19* in evolutionarily distant species. Genetics 119:663–677

McDougall I (1964) Potassium–argon ages from lavas of the Hawaiian Islands. Geol Soc Am Bull 75:107–128

McDougall I (1979) Age of shield-building volcanism of Kauai and linear migration of volcanism in the Hawaiian island chain. Earth Planet Sci Lett 46:31–42

Miyata T, Yasunaga T, Nishida T (1980) Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc Natl Acad Sci USA 77:7328–7332

Müller HJ (1940) Bearings of the *Drosophila* work on systematics. In: Hurley J (ed) New systematics. Clarendon Press, Oxford, pp 185–268

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105:207–217

Perler F, Efstratiadis A, Lomedico P, Gilbert W, Kolodner R, Dodgson J (1980) The evolution of genes: the chicken preproinsulin gene. Cell 20:555–566

Powell JR (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila.* Proc Natl Acad Sci USA 80:492–495

Pustell J, Kafatos FC (1984) A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res 12:643–655

Riley MA (1989) Nucleotide sequence of the *Xdh* region in *Drosophila pseudoobscura* and an analysis of the evolution of synonymous codons. Mol Biol Evol 6:33–52

Rodakis GC, Lecanidou R, Eickbush TH (1984) Diversity in

a chorion multigene family created by a tandem duplication and a putative gene-conversion event. J Mol Evol 20:265–273

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Sharp PM, Li W-H (1989) On the rate of DNA sequence evolution in *Drosophila.* J Mol Evol 28:398–402

Shoemaker JS, Fitch WM (1989) Evidence from nuclear sequences that invariable sites should be considered when sequence divergence is calculated. Mol Biol Evol 6:270–289

Spieth HT (1982) Behavioral biology and evolution of the Hawaiian picture-winged species group of *Drosophila.* Evol Biol 14:351–437

Staden R (1982) Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res 10:4731–4751

Staden R (1984) A computer program to enter DNA gel reading data into a computer. Nucleic Acids Res 12:499–504

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzhagi E, Inouye M (1966) Frameshift mutations and the genetic code. Cold Spring Harbor Symp Quant Biol 31:77–84

Swimmer C, Delidakis C, Kafatos FC (1989) Amplification-control element ACE-3 is important but not essential for autosomal chorion gene amplification. Proc Natl Acad Sci USA 86:8823–8827

Swimmer C, Fenerjian MG, Martínez-Cruzado JC, Kafatos FC (1990) Evolution of the autosomal chorion gene cluster in *Drosophila.* III. Comparison of the *s18* gene in evolutionarily distant species and interspecific control of chorion gene amplification. J Mol Biol (in press)

Swofford DL (1985) Phylogenetic analysis using parsimony (PAUP), version 2.4. Illinois Natural History Survey, Champaign IL

Takahata N, Slatkin M (1984) Mitochondrial gene flow. Proc Natl Acad Sci USA 81:1764–1767

Templeton AR (1980) The theory of speciation via the founder principle. Genetics 94:1011–1038

Ticher A, Graur D (1989) Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. J Mol Evol 28:286–298

Whiting JH, Pliley MD, Farmer JL, Jeffery DE (1989) *In situ* hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis.* Genetics 122:99–109

Wilson AC, Carlson SS, White SJ (1977) Biochemical evolution. Annu Rev Biochem 46:573–639

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol J Linn Soc 26:375–400

Wong Y-C, Pustell J, Spoerel N, Kafatos FC (1985) Coding and potential regulatory sequences of a cluster of chorion genes in *Drosophila melanogaster.* Chromosoma 92:124–135

Wu C-I, Li W-H (1985) Evidence for higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci USA 82:1741–1745

Zuckerkandl E (1986) Polite DNA: functional density and functional compatibility in genomes. J Mol Evol 24:12–27

Received June 14, 1989/Revised June 1, 1990