J Mol Evol (1991) 33:321–331

Journal of
Molecular Evolution
© Springer-Verlag New York Inc. 1991

Evolution of the *dec-1* Eggshell Locus in *Drosophila*. I. Restriction Site Mapping and Limited Sequence Comparison in the *melanogaster* Species Subgroup

Stefan Andersson and Andrew Lambertsson

Department of Genetics, University of Umeå, S-901 87 Umeå, Sweden

Summary. We have analyzed ~18 kb of DNA in and upstream of the *defective chorion-1* (*dec-1*) locus of the eight known species of the *melanogaster* species subgroup of *Drosophila*. The restriction maps of *D. simulans, D. mauritiana, D. sechellia, D. erecta,* and *D. orena* are shown to have basically the restriction map of *D. melanogaster,* whereas the maps of *D. teissieri* and *D. yakuba* were more difficult to align. However, the basic amount of DNA and sequence arrangement appear to have been conserved in these species. A small deletion of varying length (65–200 bp) is found in a repeated sequence of the central transcribed region of *D. melanogaster, D. simulans,* and *D. erecta.* Restriction site mapping indicated that the *dec-1* gene is highly conserved in the *melanogaster* species subgroup. However, sequence comparison revealed that the amount of nucleotide and amino acid substitution in the repeated region is much larger than in the 5′ translated region. The 5′ flanking region showed noticeable restriction site polymorphisms between species. Based on calculations from the restriction maps a dendrogram was derived that supports earlier published phylogenetic relationships within the *melanogaster* species subgroup except that the *erecta–orena* pair is placed closer to the *melanogaster* complex than to *D. teissieri* and *D. yakuba.*

Key words: *Drosophila* — *dec-1* locus — Restriction map conservation — Sequence comparison — *Melanogaster* species subgroup — Phylogeny

Offprint requests to: S. Andersson

Introduction

During oogenesis in *D. melanogaster* the somatically derived follicle cells synthesize and secrete proteins in a developmentally specific fashion. These proteins assemble into the vitelline membrane, the innermost chorionic layer, endo-, and exochorion (King and Koch 1963; Quattroponi and Anderson 1969; Margaritis et al. 1980; for review of eggshell structure and morphogenesis see Margaritis 1985). Among these proteins are the major chorion proteins, which are highly regulated in terms of tissue and temporal specificity of expression (Petri et al. 1976; Waring and Mahowald 1979; Margaritis et al. 1980; Spradling and Mahowald 1980; Spradling et al. 1980; Griffin-Shea et al. 1982). Other proteins are the vitelline membrane proteins (Brennan et al. 1982). The major chorion genes are clustered in two chromosomal sites, one at 7F1–2 on the X-chromosome (Parks et al. 1986) and the other one at 66D11–15 on the third chromosome (Spradling 1981; Griffin-Shea et al. 1982). Both of these loci are amplified in the follicle cells during oogenesis, and the amplification is necessary for production of the normal chorion (Spradling and Mahowald 1980).

Apart from these major proteins the follicle cells also produce and secrete minor proteins of vital importance for normal choriogenesis. Of these, the most spectacular and intriguing ones are the fc proteins (*f*ollicle *c*ell) encoded by the *dec-1* locus. The developmental expression and genetic localization to region 7C1–9 on the X-chromosome of this gene were first reported by Lineruth and Lambertsson (1985, 1986) and Lineruth et al. (1985), and later by Bauer and Waring (1987) and Komitopoulou et

al. (1988). It has now been cytologically mapped to 7C3–4 (Hawley and Waring 1988; A. Lambertsson, T. Johansson, and S. Andersson, unpublished). Several female-sterile mutants belonging to a large complementation group at 7C on the X-chromosome, have been correlated to this locus (Lineruth and Lambertsson 1986; Komitopoulou et al. 1988). A screen of 130 *D. melanogaster* wild-type strains revealed four electrophoretic protein variants, *Fc1–Fc4* (Lineruth and Lindberg 1988), *Fc1* being by far the most common one. Furthermore, a similar set of proteins has been shown to be synthesized in *D. melanogaster* sibling species (Lineruth 1989).

The *dec-1* locus produces at least three transcripts by alternative RNA splicing: two 3.7-kb transcripts (a and b) accumulate in stages 9–10 in a 1:10 ratio, and a larger 5.7a-kb transcript accumulates in stages 11–12 of oogenesis; a and b designate two alternative splice pathways (for details see Waring et al. 1990). The protein products are posttranslationally modified (Lineruth and Lambertsson 1985; Bauer and Waring 1987; Lineruth 1987; Hawley and Waring 1988; Komitopoulou et al. 1988). The function(s) of the polypeptides encoded by the *dec-1* transcripts is as yet not known.

In the majority of *Drosophila* molecular evolutionary studies the alcohol dehydrogenase (*Adh*) and heat shock genes have been used (Leigh-Brown and Ish-Horowitz 1981; Langley et al. 1982; Zwiebel et al. 1982; Bishop and Hunt 1988). However, extensive work has also been done on the chorion genes in *Drosophila* (Fenerjian et al. 1989 and references therein) and on genes that control development in *Drosophila*, e.g., engrailed (Kassis et al. 1989 and references therein).

The fact that *dec-1* is involved in oogenesis, a fundamental biological process, suggests that the transcribed region might be conserved both within and between species. Having cloned the *dec-1* gene from several *D. melanogaster* wild-type strains (A. Lambertsson, T. Johansson, and S. Andersson, unpublished) provided us the opportunity to survey restriction map variation in this region in *D. melanogaster* and its sibling species. This paper is the first in a study of the structural and regulatory evolution of the *dec-1* gene. We report here the restriction maps of the eight species in the *melanogaster* species subgroup. These maps indicate that the *dec-1* gene is highly conserved whereas the 5′ flanking region is relatively nonconserved. We also compare DNA sequences from *D. melanogaster* and *D. erecta* that support the restriction analysis. Furthermore, although the position of an intron in the sequenced region is conserved its sequence shows great divergence between these two species. The phylogenetic relationships derived from these restriction site data support, on the whole, earlier proposed phylogenetic trees in the *melanogaster* species subgroup of *Drosophila*.

Materials and Methods

Drosophila Strains. The strains used are listed in Table 1. They were maintained at 20°C on standard *D. melanogaster* medium, except *D. mauritiana, D. sechellia, D. orena,* and *D. teissieri,* which were kept on malt–seminola–yeast–agar medium.

Restriction Mapping. The mapping of restriction sites was done by both single and double digestion of adult DNA prepared according to the method described by Jowett (1986). Six endonucleases were used that recognize 6 bp: *Bam*HI, *Eco*RI, *Hin*dII, *Pst*I, *Sac*I, and *Xba*I, bought from Boehringer Mannheim, FRG. The DNA was digested following the supplier's recommendations.

After digestion the genomic DNA was fractionated on 0.8% agarose gels and blotted onto GeneScreen Plus filters (Dupont, NEN Research Products). The blotting was performed by the VacuGene Vacuum Blotting System (Pharmacia LKB Biotechnology AB). The filters were then prehybridized in 1% SDS, 1 M sodium chloride, 10% dextran sulfate, and 100–200 μg/ml of denatured herring sperm DNA for at least 1 h at 65°C. Hybridization was for at least 12 h at 65°C and ^{32}P-labeled probes were used at 1–4 × 10^5 cpm/ml, specific activity 1–2 × 10^8 cpm/μg. The filters were washed in 2 × 200 ml of 2 × standard saline citrate (SSC: 0.15 M NaCl, 0.015 M sodium citrate) at room temperature for 30 min, 2 × 200 ml of 2 × SSC, and 1.0% SDS at 65°C for 30 min, and finally 2 × 200 ml 0.1× SSC at room temperature for 20 min, all with constant agitation.

Probes. DNA probes were obtained by restriction enzyme digestion, electrophoresis in low melting temperature agarose gels, excision of the desired fragment, and labeling of the DNA with [^{32}P]dCTP (specific activity 1–2 × 10^8 cpm/μg) using Promega Prime-a-Gene System (Promega).

Sequence Analysis. Two *Bam*HI restriction fragments (d and e in Fig. 1) from the *D. melanogaster* Samarkand wild-type strain were cloned into the pUC19 vector. To sequence the first part of the *melanogaster dec-1* gene we used the pUC/M13 forward and reverse primers and several internal primers obtained from the sequence published by Waring et al. (1990). The primers were as follows: Dm1 = 5′[bp+1]-GAGCTCCGGCGAACACAGAT C-3′[bp+21]; Dm2 = 5′[bp+736]-GAATTCTGCAGCTGGAG CTGG-3′[bp+721], here nucleotides GAATT were added at the 5′ end to create an *Eco*RI site; Dm3 = 5′[bp+731]-GAATTCT GCAGATCCTCCGGCAG]-3′[bp+752], again nucleotides GAATT were added at the 5′ end to create an *Eco*RI site; Dm4 = 5′[bp+1238]-GTCGACTGCAGCATTCTCAATTC-3′[bp+1220], nucleotides GTCGA were added to create a *Sal*I site; Dm5 = 5′[bp+1342]-GAATTCGTCAACAAGGAAGCG CAA-3′[bp+1359], nucleotides GAATTC were added to create an *Eco*RI site at the 5′ end; Dm6 = 5′[bp+2168]-G T C G A C G G A T C C T C T G T C C A C T G C-3′[bp+2151], nucleotides GTCGAC were added at the 5′ end to create a *Sal*I site; De1 = 5′[bp+147]-CTCTGAAGATCCTGCA-3′[bp+162], this primer is from *D. erecta.* Primer Dm6 lies in the repeated region but within a repeat that shows considerably less homology to the consensus sequence of the first four repeats.

To obtain the homologous fragments from *D. erecta* genomic DNA (1 μg) was subjected to polymerase chain reaction (PCR) amplification with *Taq* DNA polymerase (Promega) using primers Dm1–Dm2, Dm3–Dm4, and Dm5–Dm6, respectively. Exponential amplification was performed in a 100-μl reaction mix

containing 50 mM KCl, 10 mM Tris-HCl (pH 9.0 at 25°C), 1.5 mM $MgCl_2$, 0.01% gelatin (w/v), 0.1% Triton X-100, 0.2 mM for each of the dNTPs, 10 pmol each primer, 1 μg of genomic DNA (denatured by boiling for 5 min), and 2 units of *Taq* DNA polymerase (Promega). The reaction was overlaid with mineral oil. Thirty-five cycles (1 min at 94°C, 1 min at 55°C, 2.5 min at 72°C; the mix was finally left for 10 min at 72°C) were performed in a TempCycler (model 60, Coy Laboratory Products, Inc.). The PCR products were then digested directly or after LMT agarose electrophoresis with the appropriate restriction enzymes and cloned into pUC19.

The DNA was sequenced by the chain termination technique (Sanger et al. 1977) using the TaqTrack Sequencing system (Promega) and [^{35}S]dATP following the instructions of the supplier. Gels (5%) of 38 × 50 cm were used, and to maximize the number of sequenced nucleotides samples from each reaction were usually applied at intervals of 2 or 3 h. All sequences have been read at least twice.

Table 1. Strains of *Drosophila* used in this study and sources of the strains

Species	Locality	Source
D. melanogaster	Samarkand	Umeå Stockcenter, Sweden
	Israel	Umeå Stockcenter, Sweden
	Shahrinau	Umeå Stockcenter, Sweden
	Dilizhan	Umeå Stockcenter, Sweden
D. simulans	Barcelona	Umeå Stockcenter, Sweden
	St. Antiocio	Umeå Stockcenter, Sweden
D. mauritiana	Unknown	Umeå Stockcenter, Sweden
D. sechellia	Unknown	Dr. M. Solignac
D. erecta	Unknown	Umeå Stockcenter, Sweden
D. orena	Unknown	Dr. M. Solignac
D. teissieri	Unknown	Dr. K. Lineruth
D. yakuba	Malawi	Umeå Stockcenter, Sweden

Results

Restriction Site Analysis

Figure 1 shows the position of the various restriction sites, deletions, and the *dec-1* transcribed unit and its flanking regions of the species in the *melanogaster* species subgroup. The *Sac*I site at the start of the transcription unit in *D. melanogaster* (Hawley and Waring 1988) is apparently present in all stocks and species, and was therefore arbitrarily given a coordinate of 0.0.

Southern analysis of genomic DNA from the four variant forms *Fc1–Fc4* (Lineruth and Lindberg 1988) revealed that the polymorphism is due to a deletion in *Fc2, Fc3,* and *Fc4* of ~65, ~130, and ~200 bp, respectively (results not shown). The deletion is within the transcribed region in the 0.8-kb *Hin*dII–*Bam*HI fragment from +1.7 to +2.5 (Fig. 1). This fragment has recently been shown to contain five identical copies of a repeated sequence 78 bp in length followed by seven repeats of varying length and homology (Waring et al. 1990), and we may hypothesize that one 78-bp repeat is deleted in *Fc2,* and, consequently, in *Fc3* 2 × 78 bp (= 156) and in *Fc3* 3 × 78 bp (= 234) are deleted. This agrees well with the estimated length of the deletions we obtained on agarose gels (see above). Sequence analysis now in progress will reveal the location and nature of these deletions (S. Andersson and A. Lambertsson, unpublished). In addition, the *Fc1* strain (Samarkand) is lacking the *Bam*HI site at position +1.1, and this is the only intraspecific restriction site variation found within the 18 kb of DNA analyzed in four *D. melanogaster* strains.

In addition, we analyzed *D. simulans* (two lines, Barcelona and St. Antiocio), *D. mauritiana, D. sechellia, D. erecta, D. orena, D. teissieri,* and *D. yakuba,* and the DNA maps are also presented in Fig. 1. Except for a few sites the DNAs of the two *D. simulans* lines show the restriction map of *D. melanogaster* (Fig. 1). Furthermore, relative to the Barcelona line the St. Antiocio line is missing the *Xba*I and *Sac*I sites at positions +3.5 and −2.8, respectively, and also has a 200-bp deletion within the *Hin*dII–*Bam*HI fragment (see above and Fig. 1). The restriction maps of *D. mauritiana, D. sechellia, D. erecta,* and *D. orena* show very little divergence relative to the map of the *D. melanogaster* standard. It is also noteworthy that the *D. erecta* line analyzed here is heterogeneous for a ~100-bp deletion in the *Hin*dII–*Bam*HI fragment (Fig. 1). The *D. teissieri* map has no *Xba*I sites and only one *Bam*HI site (see Fig. 1). However, despite these missing sites notice that the gene and its sequence arrangement appear to have been conserved in these species. Even though the *Pst*I site at +3.8 is not present in *D. melanogaster,* this is particularly evident by the presence of the *Sac*I–*Pst*I fragment between −1.0 and +3.8 (Fig. 1). Note in particular that the *Bam*HI–*Hin*dII fragment, coordinates +2.5 to +3.6, is present in all species except *D. yakuba.* Admittedly, the size of this fragment appears not to be strictly conserved. Finally, it should be noted that the *Sac*I site at position +5.1 in *D. melanogaster* is missing in all the other species (Fig. 1).

Drosophila yakuba DNA shows a very divergent restriction site map relative to the species mentioned above. There are no sites for *Bam*HI and *Xba*I, whereas sites exist for *Sac*I at positions +2.1 and +4.1 (also in *D. teissieri*) that are not found in any other species in the *melanogaster* subgroup (Fig. 1). However, the *Sac*I and *Pst*I sites at positions 0.0, and +0.7 and +1.2, respectively, are obviously conserved.

The *dec-1* 5′ flanking region shows no divergence in the restriction map of the four *D. melanogaster* wild-type strains. In *D. simulans* and *D. mauritiana* the 5′ flanking DNA regions show a slight divergence relative to *D. melanogaster* DNA, whereas the *D. sechellia, D. erecta, D. orena, D. teissieri,* and

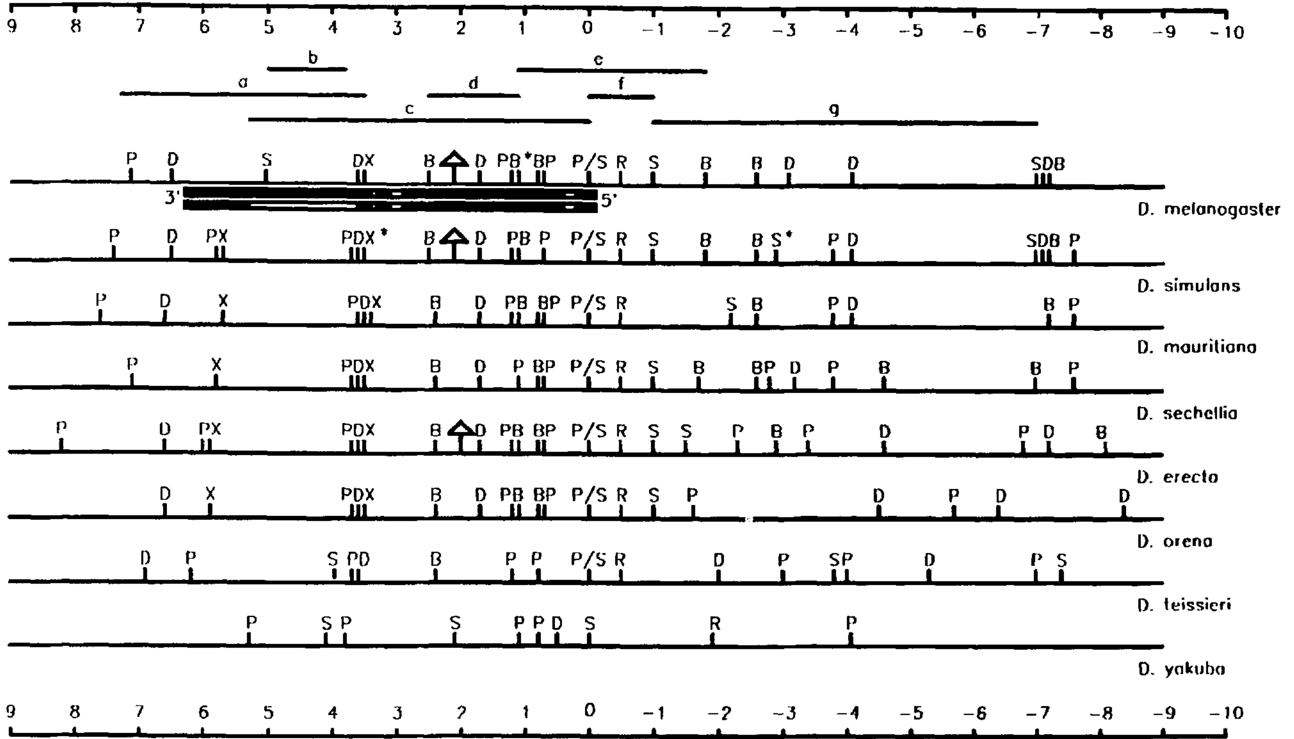

Fig. 1. Restriction maps of the *dec-1* region of *D. melanogaster* and seven other members of the *melanogaster* subgroup. Asterisked restriction sites were polymorphic within species. (Δ) denotes deletions (not to scale), filled and open boxes represent exons and introns, respectively, in the two *dec-1* transcripts in *D. melanogaster*—the upper is the 5.7-kb transcript and the lower is the 3.7-kb transcript. *a–g* indicate the different probes used. Restriction enzymes: B = *Bam*HI, D = *Hin*dII, P = *Pst*I, R = *Eco*RI, S = *Sac*I, X = *Xba*I.

Table 2. Nucleotide-distance matrix for the eight restriction maps in the *D. melanogaster* species subgroup

	mel	*sim*	*mau*	*sech*	*ere*	*ore*	*tei*	*yak*
D. melanogaster	15.5	13	11.5	13	9.5	9.5	6.5	3
D. simulans	4.71	19	14.5	15.5	11.5	10.5	10	5
D. mauritiana	5.76	3.60	17	15	11	11	8	5
D. sechellia	4.71	3.39	3.04	19	11	11	8	5
D. erecta	10.42	8.80	8.66	9.54	20	12	9	4
D. orena	7.33	7.53	5.72	6.76	5.81	14	7	4
D. teissieri	13.66	8.35	11.02	12.06	10.60	11.55	14	6
D. yakuba	23.45	17.16	15.93	17.16	21.46	17.60	10.84	9

Diagonal gives number of restriction sites in each map in Fig. 1 (except the *Hin*dII sites). The upper right half of the matrix gives the number of shared sites. The percentage nucleotide distance [calculated according to Nei and Li (1979)] are given in the lower half

D. yakuba regions show a large number of divergent sites (Fig. 1). However, it is interesting to note that the ~1.5 kb closest to the transcription start site appear to be conserved in the *melanogaster* species group, *D. teissieri* and *D. yakuba* excepted.

Evolutionary Distance Measurement

The nucleotide distance (δ) for each pair of *Drosophila* species was calculated from the number of sites mapped and the number of sites shared in the restriction maps shown in Fig. 1, according to Eqs. (8) and (10) of Nei and Li (1979). The polymorphic sites within the species were treated as ½-sites (Barrie et al. 1981). The *Hin*dII sites were not included because this enzyme is a variable six-cutter (5′GTPyPuAC3′). The results are given in Table 2 in matrix form. From this matrix the dendrogram in Fig. 4 was constructed by the neighbor-joining method (Saitou and Nei 1987).

Sequence Analysis

Except for 109 bp (between +1239 and +1348, Fig. 2) the sequence of the first 2151 and 1832 nucleotides (nt) of the *dec-1* gene of the Samarkand wild-

```
Dm        GAGCTCCGGCGAACACAGATCGATCTTTAGCGATTTTTCCGAGTTTTGCCCAAATACAGG
De        .........T.............G...........C.A..........A...CCC....A

Dm    61  ATGAGATTGTTTAGCCTTCTGCCGCTCCTGGCGCTTCTTGTCGTCCAGGTTGCCGGACAA
De        ......C...A.....A..................C..C..........C..T.......
Dm     1  MetArgLeuPheSerLeuLeuProLeuLeuAlaLeuLeuValValGlnValAlaGlyGln
De        ...............His..............................AlaVal......

Dm   121  AGTGAAGTAACCTCTGATGATCCTGCAACGGATGCTGGATCAACGACCAATTCCACCACG
De        ...C.G..G........A..........G.................G...C.....GG..
Dm    21  SerGluValThrSerAspAspProAlaThrAspAlaGlySerThrThrAsnSerThrThr
De        ...Gln.........Glu.........Arg...............Ser......ArgAla

Dm   181  GACACCAAACCAAGGATTCCCAGTCAGGATGAGgtgagtacgattagagagaattaatca
De        .C.....GG....................A...gtcagtacgcctagtacatccatcata
Dm    41  AspThrLysProArgIleProSerGlnAspGlu
De        Ala...Arg..................Glu...

Dm   241  ttgagtttgtttaatgcaatcctcttgattatatttttcaatcagATCCTTGGCCAGATG
De        gactctgctgtatgccagcctctggttcccat-ttatgtttgcag...............
Dm    52                                                 IleLeuGlyGlnMet
De                                                       ...............

Dm   301  CCGTCCATTAATCCCATCCGCACTGGCAATCCCCAGATGGACGCATTTTACATGATGTTC
De        ...C.....CCC................G...T..........TC...............
Dm    57  ProSerIleAsnProIleArgThrGlyAsnProGlnMetAspAlaPheTyrMetMetPhe
De        ...Pro...Pro...............Ser............Val...............

Dm   361  CCGGCGTTGGGCAGCCTGCTCAAGTGGGGCAGCCTTTTCCCAGCCTACTCGATTCTGGGC
De        .....................CGT.................C...C.G............
Dm    77  ProAlaLeuGlySerLeuLeuLysTrpGlySerLeuPheProAlaTyrSerIleLeuGly
De        .....................Arg.....................Gln............

Dm   421  GCCATTCCCGACAATCTACAGCCTACAGCGGCAGCCTCCAAGGTGGTCCTTGTCCTGGCC
De        .....G...........G.....CGA......C.......G......G..C.........
Dm    97  AlaIleProAspAsnLeuGlnProThrAlaAlaAlaSerLysValValLeuValLeuAla
De        ...Met..................Glu............Arg..................

Dm   481  GATGATGCGACGGCCAAGACTAGAGTTGCCCGTCAAAATCCGCCACCAAATCCACTTGGT
De        ..................G.CC......T...C..G........G.TC............
Dm   117  AspAspAlaThrAlaLysThrArgValAlaArgGlnAsnProProProAsnProLeuGly
De        ..................Ala......Val...............Leu............

Dm   541  CAGCTAATGAATTGGCCCGCTCTGCCGCAGGACTTCCAACTGCCTTCCATGGATCTGGGA
De        .....C......................................C........C......
Dm   137  GlnLeuMetAsnTrpProAlaLeuProGlnAspPheGlnLeuProSerMetAspLeuGly
De        ............................................................

Dm   601  CCGCAAGTGGGCTCGTTTTTGGCCCAACTGCCTGCTATGCCTACCGTGCCTGGTCTTCTG
De        .................................C.......CG..A.T..CA..A.....
Dm   157  ProGlnValGlySerPheLeuAlaGlnLeuProAlaMetProThrValProGlyLeuLeu
De        .................................Pro......AlaIso...SerIso...
```

Fig. 2. Sequence comparison of *dec-1* sequences. Dm, *D. melanogaster*; De, *D. erecta*. The intron is indicated by lowercase letters; potential lariat junction sites are underlined; the 5′ initial and 3′ terminal sequences (gt and ag, respectively) are in boldface. Dots indicate base or amino acid identity; a dash represents a deleted base or amino acid (three dashes). The stretch from +1240 through +1347 has not been sequenced. The repeats are indicated as Repeat 1, Repeat 2, etc. The sequence +1833 through +1850 at the end of the De sequence is primer Dm6 used in PCR (see Materials and Methods). The coordinates in this figure are not perfectly aligned with those in Fig. 1. Continued on pp. 326–327.

type strain of *D. melanogaster* and *D. erecta*, respectively, were sequenced. The difference in length of the sequenced DNA is due to the fact that the *D. erecta* DNA fragment obtained by PCR ends at +1832. Because the 3′ end of primer Dm6 is at +2151 this suggests that the *D. erecta* DNA is deleted between these coordinates. The *D. melanogaster* sequence from +1833 through +1863 is included in Fig. 2 to complete repeat 4.

The sequenced DNA contains 60 nt of the 5′ untranslated region (5′ UTR), exon 1, intron 1 (IVS1), and part of exon 2. In *D. melanogaster*, exon 2 harbors five identical copies of a repeating motif 78 nt in length followed by seven repeats of varying length and homology (Waring et al. 1990). Here we show that *D. erecta* has at least three 78-nt repeats, located at the same site as in *D. melanogaster*.

The sequence alignment in Fig. 2 clearly demonstrates that the 5′ untranslated region, exon 1, and the part of exon 2 that is sequenced are extensively conserved. Counting stretches of 9 or more base pairs, ca. 64% of the DNA is conserved between *D. melanogaster* and *D. erecta* (1106 bases of 1722, including the 72-nt-long intron 1). The longest conserved stretch is 102 bases, and several others are between 36 and 53 bases in length. There are 185 nt substitutions in the coding DNA, 67 silent and 118 replacement substitutions.

The first intron of the *dec-1* gene is 72 and 71 bp in length in *D. melanogaster* and *D. erecta*, respectively (Fig. 2). The sequence comparison reveals that this intron is poorly conserved. However, it is notable that the position of IVS1 is conserved. Furthermore, the sequence at the IVS1 5′ slice site (the donor site) in *D. melanogaster* and *D. erecta* matches the consensus sequence of (C/A)AGGT(G/A)AGT determined by Mount (1982) in eight and seven out of nine positions, respectively (Fig. 2). The 3′ splice

```
Dm   661 GGTGCCGCTGCTCCAGTTCCTGCTCCAGCTCCTGCTCCCGCTGCTGCTCCTCCTCCGGCG
De       ...................................C.........A.....---.T...A
Dm   177 GlyAlaAlaAlaProValProAlaProAlaProAlaProAlaAlaAlaProProProAla
De       .............................................Thr...---Leu...

Dm   721 CCAGCTCCAGCTGCAGATCCTCCGGCAGCACCTGTTCCAGATGCACCCCAACCAGCCATA
De   718 .............................................G....G.........
Dm   197 ProAlaProAlaAlaAspProProAlaAlaProValProAspAlaProGlnProAlaIle
De   196 .............................................Ala............

Dm   781 CTGGGACAAGCCGCTCTGCAGAACGCTTTCACCTTCTTTAACCCGGCTAACTTTGATGCC
De   778 ....................................C.G......T..............
Dm   217 LeuGlyGlnAlaAlaLeuGlnAsnAlaPheThrPhePheAsnProAlaAsnPheAspAla
De   216 ....................................Leu......Ser............

Dm   841 TCCAGTCTTCTGGGCCAGAGTGTACCCACATTTGCTCCTCCAAATCTTGATTTCGTCGCT
De   838 ...G.......A.....A..C.C..................C...T.C........G..G
Dm   237 SerSerLeuLeuGlyGlnSerValProThrPheAlaProProAsnLeuAspPheValAla
De   236 ...Gly...............Ala.....................Phe............

Dm   901 CAAATGCAAAGGCAATTCTTCCCAGGAATGACACCGGCA---CAACCTGCTGCCGCTGGC
De   898 .......................G...........A...CAG.........C........
Dm   257 GlnMetGlnArgGlnPhePheProGlyMetThrProAla---GlnProAlaAlaAlaGly
De   256 .......................................Gln.........Pro......

Dm   958 ACGGATGCCCAGGCCTCCGACATTTCCGAGGTAAGGGTACGTCCTGAGGATCCGTATTCG
De       ..........T.....................G................C..........
Dm   276 ThrAspAlaGlnAlaSerAspIleSerGluValArgValArgProGluAspProTyrSer
De       .........Leu....................................Ala.........

Dm  1018 CAAGAGGCACAGATGAAGATCAAATCGGCACTTGAAATGGAGCAGGAGAGGCAACAACAG
De       ............................................................
Dm   296 GlnGluAlaGlnMetLysIleLysSerAlaLeuGluMetGluGlnGluArgGlnGlnGln
De       ............................................................

Dm  1078 GCTCAGGTCAAGGATCAGGAGCAAGTGCCTCTCCTCTGGTTCCGAATGCCCACAACTCAG
De       ................................G..........AC........C......
Dm   316 AlaGlnValLysAspGlnGluGlnValProLeuLeuTrpPheArgMetProThrThrGln
De       ..........................................His...............

Dm  1138 AATCAGGATGCGACTGAAGAAAAGACTCTGGAGGATCTGCGGGTCGAGGCCAAATTGAGG
De       ................C...G.......................G........GC.....
Dm   336 AsnGlnAspAlaThrGluGluLysThrLeuGluAspLeuArgValGluAlaLysLeuArg
De       ...............Ala..........................................

Dm  1198 GCATTCGAGCGCCAGGTGATAGCCGAATTGAGAATGCTGCAG 1239
De       ..C..T...............T....................
Dm   356 AlaPheGluArgGlnValIleAlaGluLeuArgMetLeuGln  369
De       .....................Ser..................

Dm  1348 AAGGAAGCGCAACGAAGGGCTAGAAATTCCGGCATAAACACCCAGAAGGCAAATGCCTTG
De       ..............G.....AG.C..C........G....G................C..
Dm   406 LysGluAlaGlnArgArgAlaArgAsnSerGlyIleAsnThrGlnLysAlaAsnAlaLeu
De       .....................Gly.........Met...Ser..................
```

Fig. 2. Continued

site (the acceptor site) shows greater divergence in the surrounding sequence than does the 5′ splice site. However, both species possess the conserved AG sequence adjacent to the 3′ splice site. The lariat junction site is another characteristic sequence of the intron, and by comparing the sequences of 39 *Drosophila* introns the (C/T)T(A/G)A(T/C) consensus sequence was derived by Keller and Noon (1985). Potential lariat junction sequences are underlined in Fig. 2. It can be seen that *D. melanogaster* has two lariat junction sites that match the consensus sequence, whereas the potential sites in *D. erecta* match less well.

The most interesting part of the *dec-1* gene is the repeated region (Waring et al. 1990). The sequence alignment in Fig. 2 shows the four copies of the 78-nt repeating motif of *D. melanogaster* and the 3.5 copies of *D. erecta*. Except for a few stray substitutions, the four sequenced repeats of *D. melanogaster* are identical. These few substitutions do not result in any amino acid exchanges (Fig. 3a). As shown by Waring et al. (1990) the fifth repeat is followed by seven repeats of varying length and homology.

The structure of the *dec-1* repeats of *D. erecta* is somewhat different from those in *D. melanogaster* (Fig. 2). First, *erecta* repeat 1 differs from repeats 2 and 3 in that it is more similar to *melanogaster* repeats 1–4 (67 of 78 nt vs 57 of 78 nt, respectively). Second, repeats 2 and 3 begin and end with the consensus sequence 5′-CAAAGCCCCATGATGATGCAG-3′. In repeat 4 the central 36 bp are missing, which brings two copies of the consensus sequence adjacent to each other (Fig. 2). Further sequence comparison of the repeated region shows that only 39% of the DNA is conserved (110 bases

```
Dm  1408 AAGCGACAGGCCAAATCCCAGGATCAGACTCTGTCCAAGGAGGATATCGTCCAGATTATG
De       ..............GT.......G...G................C...............
Dm   426 LysArgGlnAlaLysSerGlnAspGlnThrLeuSerLysGluAspIleValGlnIleMet
De       .....................Glu...Arg..............................

Dm  1468 GCGTATGCCTATCGCATGGCCAACGAACAGATGGAGAGCGAGAAGGGCAAGCAGGACAAG
De       ......................................T.....................
Dm   446 AlaTyrAlaTyrArgMetAlaAsnGluGlnMetGluSerGluLysGlyLysGlnAspLys
De       ............................................................

Dm  1528 GTTTACGCGGCCTACAGGACGGAA
De       ........................
Dm   466 ValTyrAlaAlaTyrArgThrGlu
De       ........................

Dm  1552 CAGAATCCAATGATGATGCAGCAAAGACAATGGTCGGAG          Repeat 1
De       ....G........C....G....GC.G..G.....A...
Dm   474 GlnAsnProMetMetMetGlnGlnArgGlnTrpSerGlu
De       ...Ser......Thr...Glu..................

Dm  1591 GAGCAGGCCAAGATCCAACAGAATCAACAGCAGATCCAG
De       ..T.....................CT.............
Dm   487 GluGlnAlaLysIleGlnGlnAsnGlnGlnGlnIleGln
De       Asp.....................Pro............

Dm  1630 CAAAATCCAATGATGATGCAGCAACGACAATGGTCGGAG          Repeat 2
De       ...GGC.................GA.G....C.G.....
Dm   500 GlnAsnProMetMetMetGlnGlnArgGlnTrpSerGlu
De       ...Gly........................SerAla...

Dm  1669 GAGCAGGCCAAGATCCAACAGAATCAACAGCAGATCCAG
De       ..T...........G..G..A.GC.CCAT.AT...G...
Dm   513 GluGlnAlaLysIleGlnGlnAsnGlnGlnGlnIleGln
De       Asp.........Met......SerProMetMetMet...

Dm  1708 CAAAATCCAATGATGATGCAGCAAAGACAATGGTCGGAG          Repeat 3
De       ....GC..C.................G....C.G.....
Dm   526 GlnAsnProMetMetMetGlnGlnArgGlnTrpSerGlu
De       ...Ser........................SerAla...

Dm  1747 GAGCAGGCCAAGATCCAACAGAATCAACAGCAGATCCAG
De       ..T...........G..G..A.GC.CCAT.GT...G...
Dm   540 GluGlnAlaLysIleGlnGlnAsnGlnGlnGlnIleGln
De       Asp.........Met......SerProMetValMet...

Dm  1786 CAAAATCCAATGATGATGCAGCAAAGACAATGGTCGGAG          Repeat 4
De       ....GC..C.................C.CCAT.AT.AT.
Dm   553 GlnAsnProMetMetMetGlnGlnArgGlnTrpSerGlu
De       ...Ser..................SerProMetMetMet

Dm  1825 GAGCAGGCTAAGATCCAACAGAATCAACAGCAGATCCAG
De       C....AAGGCAGTGGACAGAGGATCC
Dm   566 GluGlnAlaLysIleGlnGlnAsnGlnGlnGlnIleGln
De       Gln...
```

Fig. 2. Continued

of 281, counting 9 or more bases). This should be compared with the 64% sequence conservation found for all the DNA sequenced (see above). This result indicates that repeated sequences in genes may be more prone to point substitutions and size changes, and also that the gene is more tolerable toward changes in these regions.

At the protein level, the amino acid alignment in Fig. 2 documents that 84% of the amino acid sequence is conserved between *D. melanogaster* and *D. erecta* (450 of 533 amino acids). The majority of the 83 amino acid substitutions are neutral. Comparing the amino acid sequence of the repeats (Fig. 3a) we found that 81, 61, and 61% of the 26 amino acids of repeats 1, 2, and 3, respectively, are conserved between *D. melanogaster* and *D. erecta*. Sequence comparison of all 94 amino acids or the consensus sequence in Fig. 3b showed that 65% of the sequence is conserved. Strikingly, approximately 40% of the amino acid substitutions (33 of 83 amino acids) has occurred in the repeated region.

The part of the *dec-1* gene sequenced here can be divided into a 5′ untranslated region (5′ UTR), a 5′ translated region (5′ TR), intron 1 (IVS1), and a repeated region (RR). Counting every nucleotide, amino acid, and deletion (three or more nucleotides), the amount of change observed in these parts of the *dec-1* homologous genes can be summarized as in Table 3. It is obvious from the sequence data that the intron sequence is the most rapidly evolving region. It is also apparent that the repeated portion of the coding region is evolving much more rapidly than the 5′ translated region, which is very conserved. A similar situation seems to exist in the *Sgs-3* glue gene (Martin et al. 1988).

Finally, comparing the *dec-1* sequence of the *D.*

a

```
De1     QSPMTMEQRQWSEDQAKIQQNPQQIQ
De2     QGPMMMQQRQSAEDQAKMQQSPMMMQ
De3     QSPMMMQQRQSAEDQAKMQQSPMVMQ
Dm1-5   QNPMMMQQRQWSEEQAKIQQNQQQIQ
```

b

```
De      QSPMMMQQRQSAEDQAKMQQSPMMMQ
Dm      QNPMMMQQRQWSEEQAKIQQNQQQIQ
```

Fig. 3. Repeating motif present in the *dec-1* proteins. **a** Sequence comparison of the three repeats of *D. erecta* (De1, 2, and 3) and the four identical repeats of *D. melanogaster* (Dm1–4). **b** Comparison of the two consensus repeat sequences in *D. erecta* (De) and *D. melanogaster* (Dm).

melanogaster Canton S wild-type strain reported by Waring et al. (1990) with the portion sequenced here, we find changes at the following positions: +787 (G → C; Glu → Gln), +939 (C → A), +948 (G → T), and +1171 (C → G; His → Asp). We are planning further work to study intraspecies polymorphism.

Table 3. Comparison of change in regions of *dec-1* homologous genes

	(Mismatches + deletions)/ (matches + mismatches + deletions) × 100 (%)			
	5′ UTR	5′ TR	Intron	RR
Dm vs De	15.0	8.6	63.9	24.6

Species abbreviations are as in Fig. 2. Change is defined as shown in the formula above the column headings (Martin et al. 1988). A mismatch is defined as a base that has changed between the two species; this does not include bases that have been deleted. A deletion is 3 or more contiguous bases. A match is any base that is identical in both species. The 5′ UTR (untranslated region) includes 60 bases upstream from the first codon of the protein. 5′ TR (translated region, from +61 to +1552 except between +1239 through +1347), IVS1 (intron 1, from +214 to +286), RR (repeated region, from +1552 to +1833)

Discussion

This paper is the first in a detailed study of the structural and regulatory conservation of the *dec-1* locus in the species of the *melanogaster* subgroup.

Restriction Site Mapping

One interesting feature of the *D. melanogaster dec-1* locus is that it shows four protein variants, *Fc1–4*, that are due to an increasingly larger deletion (*Fc2* < *Fc3* < *Fc4*) in the repeated regions of the 0.8-kb *Hin*dII–*Bam*HI fragment of the transcribed region in *Fc2–4* (Fig. 1). In a screen of 130 wild-type strains from all over the world *Fc1* was found to be by far the most common one (Lineruth and Lindberg 1988). The other variants are also recovered in natural populations in different parts of the world, and in the order of frequency *Fc2* > *Fc3* > *Fc4* (Lineruth and Lindberg 1988). The fact that *Fc1* is the most frequent and *Fc4* the least frequent variant form suggests that *Fc1* has a selection advantage over *Fc2–4* because of its higher number of repeats. This is presently being investigated in *D. melanogaster* (K. Lineruth, unpublished), and the deletions present in *D. simulans* and *D. erecta*, in the same 0.8-kb *Hin*dII–*Bam*HI fragment, make similar work possible in these species.

The data we have obtained from restriction mapping of the ~18 kb of DNA containing the *dec-1* gene and its 5′ flanking region indicate that the locus is highly conserved in the *melanogaster* species subgroup. This notion is supported by (1) the fact that the large intron in transcripts 3.7a and b in *D. melanogaster* is conserved in all species of the *melanogaster* species subgroup (S. Andersson and A. Lambertsson, unpublished); this intron is, however, an exon in the 5.7a transcript (Waring et al. 1990), (2) the presence of the *Bam*HI–*Hin*dII fragment, coordinates +2.5 to +3.6 in Fig. 1 (although not strictly conserved in size), in all species except *D. yakuba*, and (3) the spacing between shared sites being the same in most of the conserved region. These observations also indicate that most of the divergence in the *dec-1* transcribed region may be due to single-nucleotide substitutions. This idea is supported by the sequence comparison of the first ~1700 bp of *D. melanogaster* and *D. erecta*, which shows that most of the substitutions are single-nucleotide substitutions.

Although the 5′ flanking region is very conserved in different wild-type species of *melanogaster* this region shows poor conservation between species in the *melanogaster* subgroup. With the possible exception of the ~1.5 kb most adjacent to the transcription start site, the 5′ flanking region shows a large number of unshared restriction sites indicating that it is subject to extensive changes. This observation suggests that the *dec-1* transcribed sequence and its 5′ flanking region may have evolved at different rates. Therefore, our results may support the prediction from data of thermal elution studies on interspecies hybrids of single-copy genomic DNA that *Drosophila* genomes contain large interspersed blocks of rapidly evolving and more slowly evolving DNA (Zwiebel et al. 1982). Rapidly and slowly evolving DNA blocks have been shown to exist in the 68C glue gene clusters of several *Drosophila* species (Meyerowitz and Martin 1984). In this instance the nonconserved region, which is ~6 kb in length, contains the glue genes *Sgs3*, *Sgs7*, and *Sgs8*, where-

Table 4. Comparison of change in homologous *dec-1* genes with change in *Adh* and *Sgs-3*

	Nucleotide change coding (%)			Amino acid change coding (%)			Nucleotide change IVS (%)		
	dec-1	*Sgs-3*	*Adh*	*dec-1*	*Sgs-3*	*Adh*	*dec-1*	*Sgs-3*	*Adh*
Dm vs De	11.0	15.2	4.3	15.6	15.4	3.9	63.9	30.7	39.0

Species abbreviations are as in Fig. 2. Change is defined as in Table 3. The data for *Adh* and *Sgs-3* are from Martin et al. (1988). Coding (coding region, from +61 to +1833 except between +1239 to +1348), IVS (intron)

as the ~10 kb of conserved DNA consists of single-copy, nontranscribed sequence (Meyerowitz and Martin 1984; Martin and Meyerowitz 1986). In contrast, our results show that the *dec-1* transcribed region, which is ~6.0 kb in length, is very conserved, and that the 5′ flanking segment, ~9 kb, is relatively nonconserved.

Sequence Comparisons

We have sequenced and compared >1700 bp of the *dec-1* gene of *D. melanogaster* and *D. erecta*. Admittedly, this is not an extensive analysis, but the sequence comparisons presented here support the restriction site analyses and allow us to conclude that the *dec-1* gene appears to be highly conserved in the species of the *melanogaster* subgroup. However, our sequence data show that different regions of the *dec-1* gene show different amounts of nucleotide and amino acid substitutions (see Table 3). Thus, apart from the conserved 5′ splice site, consensus sequence IVS1 shows more than 63% nucleotide substitution between *D. melanogaster* and *D. erecta*. It is also noticeable that the amount of change, both nucleotide and amino acid, in the repeated region is approximately three times larger than in the 5′ translated region; 40% of the amino acid substitutions reported here have occurred in this part. This result suggests that the repeated region evolved much more rapidly. Further interspecies sequence comparisons of the whole *dec-1* gene will eventually reveal whether this is the case.

Without having more detailed sequence data from the other species it is premature to speculate as to how the repeated region originated and developed into its present form (unequal crossing-over may be one likely possibility however). Therefore, to map the deletions and to learn more about this DNA we are now sequencing the repeated region of variants *Fc2–4* in *D. melanogaster* (S. Andersson and A. Lambertsson, unpublished). To further study its evolutionary process this work will be extended to both closely and distantly related *Drosophila* species.

Other *Drosophila* genes that have been used in similar evolutionary studies are *Adh* (Bodmer and Ashburner 1984) and *Sgs-3* (Martin et al. 1988). Some of the data of these comparisons are summarized in Table 4. It can be seen that in both *dec-1* and *Sgs-3*, there is much more nucleotide and amino acid substitution in the coding regions as compared to the coding regions of *Adh*. This is even more pronounced if one only compares the *dec-1* repeated region (see Table 3). It thus appears that the *dec-1* coding sequences, and those of *Sgs-3*, are under much less rigorous selection pressures than is the *Adh* gene. We may conclude that the *dec-1* protein(s) can tolerate much change without losing its function. This is especially true for the repeated region, which can endure 26–78 amino acid deletions (S. Andersson and A. Lambertsson, unpublished). On the other hand, the 5′ translated region appears to be well conserved and is likely to be essential for proper function.

It is also clear that the nucleotide exchange in the introns of all three genes is considerably higher than in the coding portions (Table 4); notable also is that the amount of nucleotide substitution in the *dec-1* intron is about twice as large as compared to the introns of *Adh* and *Sgs-3*. We also note that the consensus sequences of the 5′ and 3′ splice sites are present and fairly well conserved. Potential lariat junction sites are also present but less well conserved. This suggests that introns, even small ones, are under little or moderate selection pressures but that the consensus sequences required for proper processing must be present and that this requirement could place constraints on the changes that would be allowable.

In summary, we conclude that at the restriction site level the *dec-1* gene appears to be extensively conserved, whereas sequence comparison reveals that the repeated sequence within the central part of the coding region has a much larger amount of nucleotide and amino acid substitutions than the 5′ translated portion. This may suggest that the repeated region is evolving much more rapidly than the 5′ part of the gene. The 3′ translated portion of the gene is presently being sequenced in order to see whether its rate of nucleotide and amino acid substitution is on par with that of the 5′ region.

Sequence comparisons of other genes have revealed that the upstream region of the engrailed (*en*) gene of *D. melanogaster* and *D. virilis* (Kassis et al.

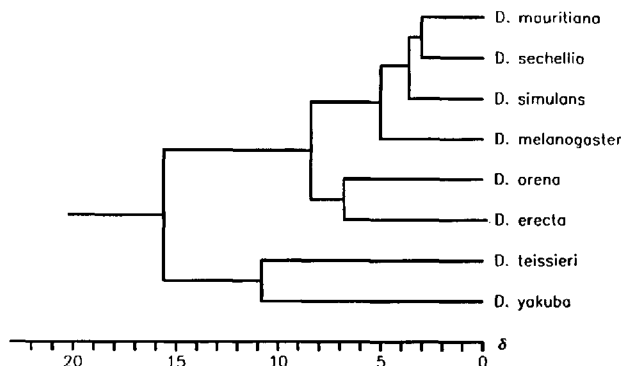

Fig. 4. Phylogenetic relationships among the eight species in the *D. melanogaster* species subgroup. The dendrogram was derived by the neighbor-joining method (Saitou and Nei 1987) from the nucleotide-distance matrix in Table 2. δ, nucleotide distance.

1989) and of the chorion genes of *D. melanogaster, D. subobscura, D. virilis,* and *D. grimshawi* (Martínez-Cruzado et al. 1988; Fenerjian et al. 1989) have small islands of conserved DNA in the 5′ flanking regions; as much as 30% of the 5′ *en* DNA is conserved in this way between *D. melanogaster* and *D. virilis.* One segment of conservation was even found downstream of the last gene in the autosomal chorion gene cluster (Fenerjian et al. 1989). In both cases, the small islands of conserved sequences are suggested to be involved in *cis*-acting functions; one of the conserved segments 5′ of the *en* gene was in fact shown to contain binding sites for homeodomain-containing proteins (Kassis et al. 1989).

The restriction maps of several species presented in this paper suggest that ca. 1.5 kb upstream of the *dec-1* gene may be conserved. By using sequence comparisons, we aim to identify *cis*-regulatory sequences important in the regulation of *dec-1* expression. Once identified these sequences will permit the use of P-element transformation to test whether they will support the developmentally correct expression of *dec-1.*

Evolutionary Distance

The dendrogram derived from the data obtained in this study shows three main clusters (Fig. 4). The three chromosomally homosequential species, *D. mauritiana, D. sechellia,* and *D. simulans,* and the closely related *D. melanogaster* form the *melanogaster* complex. The other four species in the subgroup are paired *D. erecta* and *D. orena,* and *D. yakuba* and *D. teissieri,* with the *erecta–orena* pair closer to the *melanogaster* complex. This indicates three main evolutionary lineages consistent with the historical biogeography of the *melanogaster* species subgroup (Lachaise et al. 1988).

Regarding the *melanogaster* complex, the phylogenetic relationships are congruent with those inferred from diverse characteristics: polytene chromosome banding patterns (Ashburner et al. 1984), allozymes (Daïnou et al. 1987), and DNA–DNA hybridization (Caccone et al. 1988). The branching points of the homosequential species are very close, which makes our clustering ambiguous; however, it seems as though the island species are closer to each other than to *D. simulans.* The relationships of these species have been the subject of several studies without any clear consensus of conclusions (reviewed in Lachaise et al. 1988).

In our dendrogram (Fig. 4) the *erecta–orena* pair is placed closer to the *melanogaster* complex than to *D. yakuba* and *D. teissieri.* This is in contrast to the phylogenetic relationships presented in Caccone et al. (1988) and Lachaise et al. (1988) where the *erecta–orena* and *yakuba–teissieri* species separate from the *melanogaster* complex before branching into two pairs. The sequence comparison presented in this paper indicates a high conservation of *dec-1* sequences between *D. melanogaster* and *D. erecta,* which may support our dendrogram. However, further interspecies sequence comparisons are necessary to elucidate this debate.

Acknowledgments. We thank Thore Johansson for expert technical assistance, Gail L. Waring for sharing data prior to publication, and an anonymous reviewer for helpful criticism and suggestions. S. Andersson was supported by the Sven and Lilly Lawski Foundation. This work was supported by the Erik Philip-Sörensen Foundation and the Swedish Natural Science Research Council.

References

Ashburner M, Bodmer M, Lemeunier F (1984) On the evolutionary relationships of *Drosophila melanogaster.* Dev Genet 4:295–312

Barrie PA, Jeffreys AJ, Scott AF (1981) Evolution of the β-globin gene cluster in man and primates. J Mol Biol 149:319–336

Bauer BJ, Waring GL (1987) 7C female sterile mutants fail to accumulate early eggshell proteins necessary for later chorion morphogenesis in *Drosophila.* Dev Biol 121:349–358

Bishop JG, Hunt JA (1988) DNA divergence in and around the alcohol dehydrogenase locus in five closely related species of Hawaiian *Drosophila.* Mol Biol Evol 5:415–431

Bodmer M, Ashburner M (1984) Conservation and change in the sequences coding for alcohol dehydrogenase in sibling species of *Drosophila.* Nature 309:425–430

Brennan MD, Weiner AJ, Goralski TJ, Mahowald AP (1982) The follicle cells are a major site of vitellogenin synthesis in *Drosophila melanogaster.* Dev Biol 89:225–236

Caccone A, Amato D, Powell JR (1988) Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics 118:617–683

Daïnou O, Cariou ML, David JR, Hickey D (1987) Amylase gene duplication: an ancestral trait in the *Drosophila melanogaster* species subgroup. Heredity 59:245–251

Fenerjian MG, Martínez-Cruzado JC, Swimmer C, King D, Kafatos FC (1989) Evolution of the autosomal chorion cluster in *Drosophila.* II. Chorion gene expression and sequence com-

parisons of the *s16* and *s19* genes in evolutionarily distant species. J Mol Evol 29:108–125

Griffin-Shea R, Thireos G, Kafatos FC (1982) Organization of a cluster of four chorion genes in *Drosophila* and its relationship to developmental expression and amplification. Dev Biol 91:325–336

Hawley RJ, Waring GL (1988) Cloning and analysis of the *dec-1* female-sterile locus, a gene required for proper assembly of the *Drosophila* eggshell. Genes & Dev 2:341–349.

Jowett T (1986) Preparation of nucleic acids. In: Roberts DB (ed) *Drosophila,* a practical approach. IRL Press Oxford, Washington DC, pp 276–285

Kassis JA, Desplan C, Wright DK, O'Farrel PH (1989) Evolutionary conservation of homeodomain-binding sites and other sequences upstream and within the major transcription unit of the *Drosophila* segmentation gene engrailed. Mol Cell Biol 9:4304–4311

Keller BK, Noon WA (1985) Intron splicing: a conserved internal signal in introns of *Drosophila* pre-mRNAs. Nucleic Acids Res 13:4971–4981

King RC, Koch EA (1963) Studies on the ovarian follicle cells of *Drosophila.* Q J Microsc Sci 104:297–320

Komitopoulou K, Margaritis LH, Kafatos FC (1988) Structural and biochemical studies on four sex-linked chorion mutants of *Drosophila melanogaster.* Dev Genet 9:37–48

Lachaise D, Cariou ML, David JR, Lemeunier F, Tsacas L, Ashburner M (1988) Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol Biol 22:159–225

Langley CH, Montgomery E, Quattlebaum WF (1982) Restriction map variation in the *Adh* region of *Drosophila.* Proc Natl Acad Sci USA 79:5631–5635

Leigh-Brown AJ, Ish-Horowitz D (1981) Evolution of the 87A and 87C heat-shock loci in *Drosophila.* Nature 290:677–682

Lineruth K (1987) An ovarian protein locus and its developmentally regulated expression in *Drosophila.* Doctoral thesis, University of Umeå, Sweden

Lineruth K (1989) Synthesis patterns of a set of follicle cell proteins in *Drosophila melanogaster* sibling species. Genetica 78:33–37

Lineruth K, Lambertsson A (1985) Stage specific synthesis of some follicle cell proteins in *Drosophila melanogaster.* Wilhelm Roux's Arch Dev Biol 194:436–439

Lineruth K, Lambertsson A (1986) Correlation between a female sterile mutation and a set of follicle cell proteins in *Drosophila melanogaster.* Mol Gen Genet 205:213–216

Lineruth K, Lindberg M (1988) Electrophoretic variant forms of a set of follicle cell proteins in *Drosophila melanogaster*: frequencies in wild-type strains and response in a selection experiment. Hereditas 108:59–64

Lineruth K, Lambertsson A, Lindberg M (1985) Genetic localization of a follicle cell protein locus in *Drosophila melanogaster.* Mol Gen Genet 201:375–378

Margaritis LH (1985) Structure and physiology of the eggshell. In: Kerkut GA, Gilbert LI (eds) Comprehensive insect physiology, biochemistry, and pharmacology, vol 1. Pergamon Press, Elmsford NY, pp 153–230

Margaritis LH, Kafatos FC, Petri WH (1980) The eggshell of *Drosophila melanogaster.* I. Fine structure of the layers and regions of the wild-type eggshell. J Cell Sci 43:1–35

Martin CH, Meyerowitz EM (1986) Characterization of the boundaries between adjacent rapidly and slowly evolving genomic regions in *Drosophila.* Proc Natl Acad Sci USA 83: 8654–8658

Martin CH, Mayeda CA, Meyerowitz EM (1988) Evolution and expression of the *Sgs-3* glue gene of *Drosophila.* J Mol Biol 201:273–287

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

Meyerowitz EM, Martin CH (1984) Adjacent chromosomal regions can evolve at very different rates: evolution of the *Drosophila* 68C glue gene cluster. J Mol Evol 20:251–264

Mount SM (1982) A catalogue of splice junction sequences. Nucleic Acids Res 10:459–472

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Parks S, Wakimoto B, Spradling AC (1986) Replication and expression of an X-linked cluster of *Drosophila* chorion genes. Dev Biol 117:294–305

Petri WH, Wyman AR, Kafatos FC (1976) Specific protein synthesis in cellular differentiation. III. The eggshell proteins of *Drosophila melanogaster* and their program of synthesis. Dev Biol 49:185–219

Quattroponi SL, Anderson E (1969) The origin and structure of the secondary coat of the egg of *Drosophila melanogaster.* Z Zellforsch 95:495–510

Saitou N, Nei M (1987) The neighbor-joining method: a new method for constructing phylogenetic trees. Mol Biol Evol 4: 406–425

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

Spradling AC (1981) The organization and amplification of two clusters of *Drosophila* chorion genes. Cell 27:193–202

Spradling AC, Mahowald AP (1980) Amplification of genes for chorion proteins during oogenesis in *Drosophila melanogaster.* Proc Natl Acad Sci USA 77:1096–1100

Spradling AC, Digan ME, Mahowald AP, Scott M, Craig EA (1980) Two clusters of genes for major chorion proteins of *Drosophila melanogaster.* Cell 19:905–914

Waring GL, Mahowald AP (1979) Identification and time of synthesis of chorion proteins in *Drosophila melanogaster.* Cell 16:599–607

Waring GL, Hawley RJ, Schoenfeld T (1990) Multiple proteins are produced from the *dec-1* eggshell gene in *Drosophila* by alternative RNA splicing and proteolytic cleavage events. Dev Biol 142:1–12

Zwiebel LJ, Cohn VH, Wright DR, Moore GP (1982) Evolution of single-copy DNA and the ADH gene in seven drosophilids. J Mol Evol 19:62–71

Received March 6, 1990/Revised March 8, 1991