develops, phylogenetic analysis should become a standard tool to study behavioral adaptation in *G. aculeatus*.

Prospects for future research

Formal sessions and informal evening gatherings in Leiden were filled with excited discussions of how phenomena or populations used in one paper could be investigated using methods employed in another. Many studies elucidated behavioral mechanisms or mechanisms for the evolution of behaviors, but perhaps the most exciting aspect of the symposium was the potential that these studies revealed for future research: as the intraspecific phylogeny of the threespine stickleback improves, our understanding of behavioral evolution in this complex will expand. Computer-assisted manipulation of video images of sticklebacks will allow isolation of causal factors in stickleback behavior. The white stickleback, pairs of species from lakes, and other well-characterized, divergent, allopatric populations, will provide important insights into the evolution of behavioral and life history phenotypes. Access to populations in the Russian Far East and Eastern Europe also provide new possibilities for comparative studies. We do not anticipate having to wait another decade for the Third International Symposium on Stickleback Behaviour.

Acknowledgements

My research on stickleback evolution was supported by USA National Science Foundation grant BSR-8905758, and my participation in this symposium was partially supported by a van Dobberke Foundation grant to P. Sevenster. I thank C.M. Blair and C. Andrews for reading the manuscript. This is Contribution 926 from Ecology and Evolution at Stony Brook.

Michael A. Bell

Dept of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245, USA

References

1 Bell, M.A. and Foster, S.A. (1994) in *The Evolutionary Biology of the Threespine Stickleback* (Bell, M.A. and Foster, S.A., eds), pp. 1–27, Oxford University Press
2 Hagen, D.W. (1967) *J. Fish. Res. Board Can.* 24, 1637–1692
3 McPhail, J.D. (1969) *J. Fish. Res. Board Can.* 26, 3183–3208
4 van den Assem, J. and Sevenster, P. (1985) *Behaviour* 93, iii–iv + 1–277
5 Trivers, R.L. (1972) in *Sexual Selection and the Descent of Man* (Campbell, B., ed.), pp. 136–179, Aldine
6 McPhail, J.D. (1994) in *The Evolutionary Biology of the Threespine Stickleback* (Bell, M.A. and Foster, S.A., eds), pp. 399–437, Oxford University Press
7 Blouw, D.M. and Hagen, D.W. (1990) *Biol. J. Linn. Soc.* 39, 195–217
8 Rowland, W.R. (1994) in *The Evolutionary Biology of the Threespine Stickleback* (Bell, M.A. and Foster, S.A., eds), pp. 297–344, Oxford University Press
9 Reimchen, T.E. (1989) *Evolution* 43, 450–460
10 Brooks, D.R. and McLennan, D.A. (1991) *Phylogeny, Ecology, and Behavior*, University of Chicago Press
11 Bell, M.A. (1974) *Natural History Museum of Los Angeles County Contributions in Science* 257, 1–36
12 Haglund, T.E., Buth, D.G. and Lawson, R. (1992) *Copeia* 1992, 432–443
13 Ortí, G., Bell, M.A., Reimchen, T.E. and Meyer, A. (1994) *Evolution* 48, 608–622
14 Huntingford, F.A., Wright, P.J. and Tierney, J.F. (1994) in *The Evolutionary Biology of the Threespine Stickleback* (Bell, M.A. and Foster, S.A., eds), pp. 277–296, Oxford University Press
15 Foster, S.A. (1994) in *The Evolutionary Biology of the Threespine Stickleback* (Bell, M.A. and Foster, S.A., eds), pp. 381–398, Oxford University Press

Molecular insights into the evolution of an enzyme; esterase6 in *Drosophila*

J.G. Oakeshott, T.M. Boyce, R.J. Russell and M.J. Healy (1995)

It is still a suspicion among some evolutionary biologists that the incursion of molecular biology into their field will do little more than determine the DNA sequence differences underlying evolutionary changes already evident at the organismal level. Work on an esterase enzyme involved in the reproductive biology of *Drosophila* belies this view. Although it is already one of the most intensively studied gene–enzyme systems at an organismal level, recent molecular investigations reveal several unexpected, and, in some cases, still inexplicable phenomena in its evolutionary history.

J.G. Oakeshott, T.M. Boyce, R.J. Russell and M.J. Healy are at the CSIRO Division of Entomology, PO Box 1700, Canberra, ACT 2601, Australia.

To understand the molecular basis of evolutionary change, it is necessary to understand how new biochemical capabilities emerge from genetic variations within individual genes and across related genes in multigene families. We have gained some initial insights into these processes over the past 15 years through the use of gene cloning and sequencing techniques. We now know the scope and organization of a few multigene families, and, for some of these, we also know something about nucleotide variation within and among family members. The greater challenge from here is to determine the functional significance of the variation. This is harder because it requires detailed knowledge of the structures, both of the proteins encoded by the genes in the family, and of the promoters governing their expression. There is no multigene family for which such data are even close to complete. However, the data now accumulating for some systems reveal some remarkably simple links between genetic and biochemical differences, as well as many unexpected and a few inexplicable new phenomena. Here, we outline some of the insights and paradoxes emerging for the so-called β-esterase gene cluster in *Drosophila*.

Est6 and the β-esterase gene cluster

The β-esterase cluster is best characterized in *D. melanogaster*, where it consists of a tandem pair of genes called *Est6* and *EstP* (Ref. 1). The two genes span just 4 kb and their products show about 70% amino acid identity. Homologues have been identified across much of the genus *Drosophila*, as a tandem pair again in a few species but also as a cluster of three or four in many others (Fig. 1). In their sequences, the genes are all more closely related to each other than to other esterases for

© 1995, Elsevier Science Ltd 0169-5347/95/$09.50

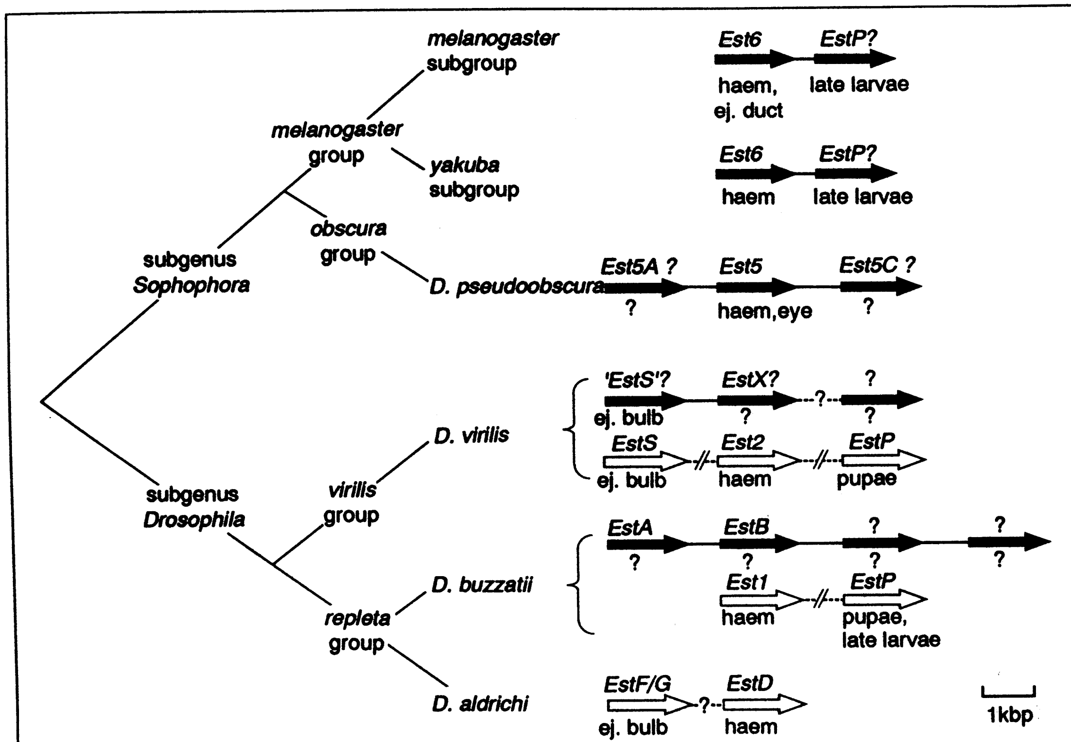


Fig. 1. Organization of β-esterase genes in various *Drosophila* species as inferred from molecular and classical genetics (black and white arrows, respectively). For the molecular analyses, the arrows indicate direction of transcription. Major expression phenotypes are indicated below the arrows. The question mark indicates unknown gene product (above the arrows), unknown linkage and orientation (between the arrows) or unknown expression (below the arrows). ..//.. indicates close linkage but precise location unknown. haem = haemolymph; ej. duct = ejaculatory duct; ej. bulb = ejaculatory bulb. For *D. aldrichi*, there are two isozymes, ESTF and ESTG, in the ejaculatory bulb but it is unknown whether or not they are encoded by separate genes. *Data are from Refs 1–16.*

which data are available. Nevertheless, they vary widely in their patterns of tissue-specific expression, both across the cluster and among species. Likewise, the biochemical characteristics of their protein products vary considerably, at the extreme acquiring new, non-esteratic functions.

Most of what we know to date about gene expression, protein biochemistry and physiological function concerns the 5′ gene in the pair, *Est6*. In *D. melanogaster* at least, the EST6 protein is expressed in many tissues but its two major sites are the sperm ejaculatory duct of the adult male and the haemolymph of all life stages[2]. The function of the haemolymph EST6 is unknown but the ejaculatory duct enzyme is transferred to the female during mating and there, by a mechanism still only partly understood, it stimulates egg-laying and reduces receptivity to remating[3]. *EstP* is essentially only expressed in late larvae, probably in epidermal tissue[1,15]. Its function is unknown but it clearly differs from *Est6*.

High level expression of *Est6* in the ejaculatory duct appears to be limited to just three sibling species in the *melanogaster* subgroup, *D. melanogaster, D. simulans* and *D. mauritiana* (although some other species show low levels of activity in this tissue)[2]. These three species are also unusual in expressing EST6 solely as a monomer; almost all other drosophilids analyzed express the enzyme as a homodimer of twice the molecular weight. This difference in structure may be critical to its function in reproductive behaviour in the *melanogaster* subgroup. Specifically, the smaller size of the monomer may enable transport across vaginal membranes, which is necessary for the effect on egg-laying and remating receptivity[3].

Est6 has also been subject to detailed study in another species in the subgenus *Sophophora*. This is *D. pseudoobscura*, in the *obscura* group, where the enzyme orthologous to EST6 on biochemical criteria (termed EST5 in this species)[16] is the central member of a tandem triplication spanning about 8 kb[5–7] (Fig. 1). Sequence analyses indicate that the proteins encoded by the other two genes also include all the motifs known to be required for esterase activity (see Box 1). Sequence similarities suggest that the 3′ member of the trio is orthologous to *EstP*, while the 5′ gene has arisen as a duplication of *Est5/6* subsequent to the divergence of the *obscura* and *melanogaster* species groups. EST5 accumulates in the haemolymph like EST6, but it is also found at significant levels in the eye, which is a novel site of activity studied in the species to date. Nothing is known about the expression of the other two genes in *D. pseudoobscura*.

Orthologues of EST6 have also been studied in the *virilis* and *repleta* species groups in the subgenus *Drosophila*. Classical genetics in *D. virilis* show a cluster of three genes encoding esterases related to EST6 and ESTP[8,9] (Fig. 1). One of these (EST2) is orthologous to EST6 on biochemical criteria and, again like EST6, it is abundant in the haemolymph. One of the others (ESTS) also resembles EST6, in so much as it is expressed in the male reproductive tract, albeit in the ejaculatory bulb rather than the duct. ESTS is also transferred to the female during mating but, unlike EST6, it is then essentially confined within her reproductive tract. It seems likely that the gene encoding ESTS has arisen as a duplication of *Est2/6* sometime since the divergence of the two subgenera, subsequently acquiring ejaculatory bulb expression and losing haemolymph expression. Little is known about the third gene in the *D. virilis* cluster, other than that it is mainly expressed in pupae. It may be orthologous to *EstP* in *D. melanogaster* or it may have arisen in an independent duplication event.

Molecular work on *D. virilis* has also so far revealed three genes related to *Est6* and *EstP* in *D. melanogaster*[9–11]. A 15 kb clone containing two of these has been isolated (Fig. 1) and shown to hybridize to the same chromosomal region to which the genes encoding the EST2, ESTS and ESTP isozymes have been mapped by classical genetics. One of the cloned genes is heavily expressed in the ejaculatory bulb and has therefore been nominated as the *EstS* gene. However, while subsequent sequence analyses confirm its close relationship to EST6 they also show that the protein encoded by the nominate *EstS* gene lacks two residues that have proven necessary for esterase activity in all enzymes related to EST6 so far characterized (Glu/Asp and His in the catalytic triad – see Box 1). It thus appears that the nominate '*EstS*' encodes a non-esteratic protein that is nevertheless closely linked and closely related to the 'real' *EstS* and shares its distinctive expression profile. The sequence of

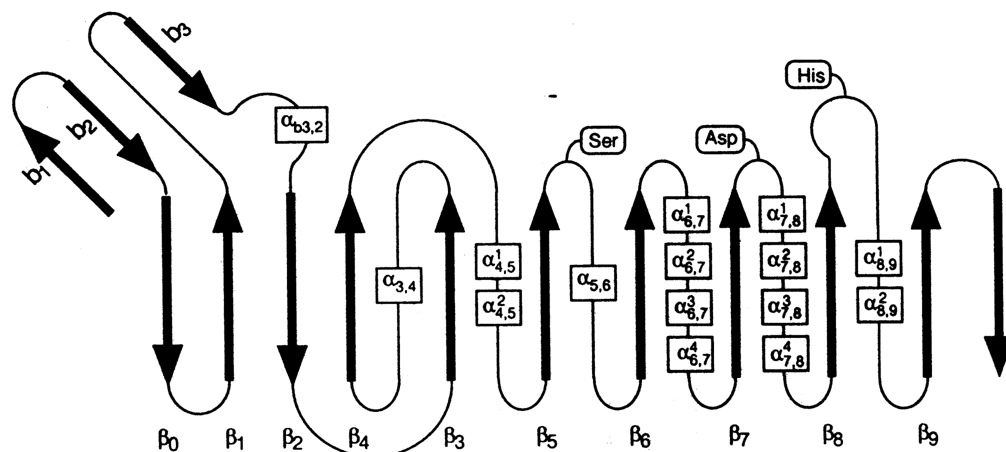
the other cloned *D. virilis* gene has not yet been reported.

The classical genetics of EST6 orthologues have been described in two *repleta* group species, *D. buzzatii* and *D. mojavensis*, both in the *mulleri* subgroup. The data from both species reveal a pair of tightly linked genes, one encoding a haemolymph esterase orthologous on biochemical grounds to EST6 and the other a late larval/early pupal esterase that may be orthologous to the ESTP of *D. melanogaster* and/or *D. virilis*[2,12,13] (Fig. 1). Some other *repleta* species also manifest an ejaculatory bulb esterase that may be orthologous to the ESTS of *D. virilis*, but the genetics of this enzyme have not been reported[14]. An *Est6* probe from *D. melanogaster* has been used to isolate a 16 kb genomic fragment from *D. buzzatii*, which contains four esterase-like genes, two of which have been sequenced[15]. The latter two show strong similarity with *Est6* and *EstP* but, like the cloned '*EstS*' of *D. virilis*, neither would encode a functional catalytic triad (in this case, all three residues in the catalytic triad are missing). Again, like '*EstS*', they have retained all the flanking and coding region motifs required for proper transcription and translation. They also lack the premature stop codons that are generally found to accumulate in non-functional genes. Thus, the *Est6* lineage appears to have diversified in species of the subgenus *Drosophila* to include genes whose products have acquired novel, non-esteratic functions.

What then are the molecular bases for this diversity of biochemical and physiological phenotypes across the cluster and the genus? In the analyses below, we trace many of the biochemical and physiological changes in esterase6 down to specific differences in the sequences of the EST6 protein and the *Est6* promoter. Both of these types of changes in primary sequence may be the foci for natural selection, within and between species.

Box 1. Functional analysis of the EST6 protein

EST6 shows 25% or more amino acid sequence identity with about 30 other carboxyl or cholinesterases from diverse eukaryotes[16,17]. Two of these other esterases (acetylcholinesterase from the electric eel and a fungal lipase) have been crystallized and their tertiary structures resolved. The two enzymes' structures are remarkably similar, even though primary sequence comparisons show them to be relatively distantly related within the carboxyl/cholinesterase multigene family. Furthermore, their structures show strong similarities with those of some other hydrolases that lie outside this family and with which they show negligible primary sequence similarity[17]. It is therefore proposed that the conserved tertiary structure, called an α/β hydrolase fold, will also apply to other carboxyl/cholinesterases like EST6 (Refs 16,17).



The backbone of the structure is termed the major β sheet. It is composed of about 10 slightly distorted planar structures called β strands (the arrows designated β_0–β_9 in the diagram) that are arranged in the form of a distorted ladder. The ladder is stabilized by several cysteine and salt bridges. The sequences encoding the 10 β strands are not contiguous in the primary sequence of the protein but they follow a characteristic order and orientation in the sequence. The intervening sequences are looped out in the tertiary structure so that the major β sheet is essentially hidden within the interior of the protein. The loops include several tightly wound helical structures called α helices (represented by boxes in the diagram) and one loop also contributes to a second, minor, β sheet containing three strands (b_1–b_3). Many regions in the loops are exposed on the protein surface but the precise functions of most such regions, including the minor β sheet, are unknown. Notable exceptions include four sites used for the attachment of carbohydrate side chains. The glycosylation at these sites enhances the enzyme's stability *in vivo*[16]. Other loop residues are also known to contribute to the active site and substrate binding sites of the enzyme.

The active site is a gorge made up of about 30 residues drawn from several different loops. The esteratic reaction is effected by three key residues called the catalytic triad at the base of the gorge, a serine (Ser) in the loop after strand β_5, a glutamic or aspartic acid (apparently Asp in EST6) after β_7, and a histidine (His) after β_8. The catalytic triad forms a charge relay system ultimately resulting in the transfer of a proton to the ester bond of the substrate. The function of the Ser requires its location at the apex of an extremely sharp turn, represented in the primary sequence by a highly conserved Gly-x-Ser-x-Gly motif. Three sites on the surface of the molecule at or near the lip of the gorge have been implicated in substrate binding. These three sites occur after strands b_3 and β_7, although other sites nearby, for example, around β_6 and β_2, may also be involved.

Variation and selection in the EST6 amino acid sequence

Surveys using high-resolution electrophoretic procedures and thermostability assays have detected over 20 variants of EST6 within each of *D. melanogaster*, *D. simulans* and the orthologous EST5 in *D. pseudoobscura*[18–24]. The enzyme is thus among the most polymorphic systems so far investigated at the protein level. Even more variation in the EST6 protein has been uncovered by molecular analyses. Sequence data for 13 independently isolated *Est6* genes from *D. melanogaster* and four from *D. simulans* reveal 16 amino acid replacement polymorphisms in the former and 11 in the latter[24,25]. The ratio of variation in replacement sites as opposed to the presumptively neutral silent sites (those that do not lead to amino acid replacements, because of the degeneracy of the genetic code) of the *Est6* gene (hereafter termed the R/S ratio) is 16% for *D. melanogaster* and 6% for *D. simulans* (although the *D. melanogaster* value is biased upwards because the sequenced lines were selected to represent different allozymes). While these data indicate some selective constraint against amino acid variation, which in itself is unsurprising, the R/S ratios for *Est6* are high compared to many other *Drosophila* genes (Table 1), and the inferred level of constraint against amino acid variation is correspondingly low.

Nevertheless, some of the amino acid polymorphism is apparently not just the result of the accumulation of selectively neutral changes. Deviations from expectations under neutral theory are evident in the frequency distributions of certain EST6/5 allozymes in *D. melanogaster*, *D. simulans* and *D. pseudoobscura*[4,21–24]. The most striking deviations are in *D. melanogaster* and *D. simulans*. In both species, there are two common groups of variants, collectively labelled EST6-F and EST6-S, within which electrophoretic mobilities are quite similar but between which mobilities are more

distinct. In each species, the frequency of the EST6-S group increases at the expense of EST6-F away from the equator. These latitudinal patterns recur across all continents so far investigated and both hemispheres. The data are most extensive for the east coasts of Australia and North America; both data sets reveal clinal changes of about 50% in EST6-S frequencies over 25–30° of latitude. The scale and repeatability of these patterns strongly suggest a latitudinally dependent process of natural selection[21–24]. However, attempts to mimic the selection in the laboratory have so far been inconclusive[44–46], so it may be either inoperative or too weak to detect consistently in the laboratory.

The pattern of change in the EST6 amino acid sequence across species is also indicative of some selection. As well as the multiple *D. melanogaster* and *D. simulans* sequences above, there are also data for single sequences of *Est6* from *D. mauritiana*, and the orthologous *Est5* in *D. pseudoobscura*. Pairwise comparisons of the other three species' genes back to *Est6* in *D. melanogaster* yield R/S values which, like the intraspecific values above, are large compared to many other similarly characterized genes (Table 1). This confirms the notion that the overall level of constraint against amino acid change in *Est6* is relatively weak. However, all three interspecific R/S values are larger than the intraspecific R/S for *Est6* in *D. simulans* (the intraspecific *D. melanogaster* value being precluded from comparison because it is biased upwards; see above)[24]. Moreover, the largest R/S value is obtained from the comparison of *D. melanogaster* to *D. pseudoobscura*, which, quaternary structure and expression data (Fig. 1) would suggest, involves the greatest divergence in function. This suggests that the larger R/S values for the interspecific comparisons are at least partly the result of divergent selection pressures on the protein sequence associated with some divergence in its function.

What amino acid changes are the objects of selection, both within and among species? Part of the difficulty in answering this question is the very abundance of amino acid variation, presenting a plethora of possible candidates for selection. For example, the EST6-F and EST6-S allozyme groups in *D. melanogaster* and *D. simulans* are each distinguished by two amino acid differences[16,24]. One of the two is charge-nonconservative in each species and is assumed to be the cause of the electrophoretic difference; however it need not be the target for selection. The discriminating polymorphisms are the charge-altering Asp/Asn at residue 237 plus the charge-conservative Thr/Ala-247 in *D. melanogaster*, and the charge-altering Thr/Asn-237 plus the charge-conservative Asp/Val-487 in *D. simulans* (EST6-F residues first in each case). To further investigate which amino acid differences may be targets for selection we now consider what is known about sequence–structure–function relationships within the EST6 protein.

Targets for selection in the EST6 amino acid sequence

Much of the tertiary structure of EST6 can be inferred quite precisely from the known structures of other members of the carboxyl/cholinesterase multigene family (Box 1). This information enables us to partition the EST6 primary sequence into several functional categories. Polymorphic and interspecific differences in *D. melanogaster*, *D. simulans* and *D. mauritiana* are rare in about 51% of the primary sequence represented by six of these functional categories (Table 2). Five of these – the major and minor β sheets, α helices and the cysteine and salt bridges – provide and stabilize the overall α/β hydrolase fold structure that characterizes the carboxyl/cholinesterase multigene family. The sixth, the active site gorge, is critical for catalysis. Thus, some of the elements of the enzyme most critical to its overall structure and its catalytic function are tightly conserved within the *melanogaster* subgroup.

Much of the variation within this subgroup occurs in the 41% of the sequence for which no specific structural motif or function has yet been identified. Many of these sequences lie on or near the protein surface where functional constraints might be relatively relaxed, so some of the variation observed in these categories of sequences could be neutral to natural selection. On the other hand, some of the highest levels of variation occur in two categories of sequence with very specific functions. These are the substrate binding sites and the glycosylation sites.

Remarkably, the variation in the substrate binding sites includes both the polymorphisms in *D. melanogaster* and one of the two polymorphisms in *D. simulans* that are associated with the major EST6-F/EST6-S allozyme difference (the differences in residues 237 and 247 above.) This could explain the observation that purified EST6-F and EST6-S differ in their affinities for various substrates *in vitro*[47,48]. It is clearly now important to determine whether these *in vitro* differences reflect *in vivo* differences. If they do, then ecological work will clearly be needed to test the hypothesis that the selection between EST6-F and EST6-S inferred from their latitudinal clines is, at least in part, mediated by differences in their use of various substrates.

Another candidate for selection may be one of the four glycosylation sites in *D. melanogaster* EST6. The site is missing in a minority of the isolates sequenced from this species, as well as all four isolates sequenced from *D. simulans*, and the one sequenced from *D. mauritiana*. The lack of this site prevents the attachment of a carbohydrate side chain to the EST6 protein. This in turn may affect fitness because physiological studies show that the *in vivo* properties of the enzyme are altered in laboratory EST6 mutants lacking all four glycosylation sites. Specifically, the lifespan of male-donated EST6 in the female after mating is shorter for these mutants than for wild-type EST6 containing all four glycosylation sites (Box 1 and Ref. 16).

While comparisons of EST6 variants within the *melanogaster* subgroup are presumably comparisons of proteins with similar functions, comparisons of EST6 to EST5 in *D. pseudoobscura* are between proteins with distinctly different expression patterns and presumed physiological roles. None of the functional categories in the amino acid sequence is invariant in the latter comparison. Both the active site gorge and, to a lesser extent, the salt bridges differ. Predictably, given that both molecules are functional esterases, the variation is still excluded from the catalytic triad and the Gly-x-Ser-x-Gly pentapeptide at the base of the gorge, which are critical for any esterase activity. However, the variation now found elsewhere in the gorge could change the nature of the molecules' esterase activities. Synthetic mutations at some of these sites in EST6 and other esterases produce kinetic and substrate specificity differences (probably because they affect the guidance and positioning of the substrates down the gorge)[49,50]. The variation between EST6 and EST5 at such sites is consistent with qualitative differences in the kinetics and substrate specificities of the purified enzymes *in vitro*[47,48,51], as well as the functional differences *in vivo* inferred from their differences in expression patterns.

Finally, and at the extreme of functional change, comparisons of EST6 with the presumptively non-esteratic products of the *EstS* clone from *D. virilis* and the *EstB* clone from *D. buzzatii* show widespread change in all sequence categories. In both comparisons, the variation includes

Table 1. Ratios of variation in amino acid replacement sites versus silent sites (R/S), and in the adjacent 5′ flanking DNA (up to the first 350 bp before the start of transcription) versus silent sites (P/S) for *Est6*[a] and some other *Drosophila* genes[b]

	R/S and P/S						
	Intraspecific comparisons				Interspecific comparisons		
Gene	*mel*[c]	*sim*	*mau*	*pseudo*	*mel* vs *sim*	*mel* vs *mau*	*mel* vs *pseudo*
Est6[5,16,24,25,28]	0.16 & –	0.06 & 0.06	–	–	0.14 & 0.16	0.17 & 0.15	0.29 & –[d]
Adh[29–32]	0.02 & 0.31	–	–	0.02 & 0.42	0.07 & 1.16	0.24 & –	0.07 & –
Adh-dup[31,32]	0.14 & –	–	–	0.08 & –	0.07 & –	–	0.03 & –
Xdh[33]	–	–	–	0.08 & 0.39	–	–	0.03 & –
pgd[34]	0.04 & 0.22	–	–	–	0.05 & 0.40	–	–
ci^D[35]	0.00 & –	0.00 & –	–	–	0.23 & –	–	–
per[36,37]	0.04 & –	0.04 & –	0.01 & –	–	0.03 & –	0.03 & –	0.37 & –
zeste[38]	0.00 & –	0.00 & –	0.00 & –	–	0.01 & –	0.01 & –	–
yp2[38]	0.09 & –	0.00 & –	0.14 & –	–	0.14 & –	0.15 & –	–
Sod[39]	–	–	–	–	0.00 & 1.17	–	–
hsp82[40]	–	–	–	–	0.00 & 1.34	–	0.03 & –
ac[41]	–	–	–	–	0.07 & 0.13	–	–
tra[42]	–	–	–	–	0.28 & –	–	–
sal[43]	–	–	–	–	0.17 & –	–	–

[a]Transcription starts 42 bp 5′ of the coding region of *Est6*.
[b]The variables used in the ratios are the standard polymorphism and divergence statistics $\hat{\pi}$ and $\hat{d}$, for the intra- and interspecific comparisons respectively[26]. Values of $\hat{d}$ are corrected for multiple hits[27].
[c]*mel* = *D. melanogaster*; *sim* = *D. simulans*; *mau* = *D. mauritiana*; *pseudo* = *D. pseudoobscura*.
[d]P/S cannot be calculated for this comparison although the promoter sequences are available. The sequences are so different that only about 100 bp of the first 350 bp can be aligned.

Table 2. Amino acid variation[a] in structurally defined categories of sequence[b] in EST6 and its homologues across various *Drosophila* species[c]

	Major β sheets	Minor β sheets	Cysteine, salt bridges	Active site gorge	Substrate binding sites	Glycosylation sites	Remaining α helices	Remaining residues
Number of residues aligned	67	13	16	28	34	8	132	204
Polymorphism for EST6								
– in *mel*	0	0	0.001	0	0.03	10.04	0.01	0.01
– in *sim*	0	0	0	0	0.05	0	0	0.02
Divergence of *mel* EST6 from the orthologous								
– *sim* EST6	10.02	0	0	0	0.06	0.13	0.02	0.05
– *mau* EST6	10.03	0	0	0	0.06	0.13	0.03	0.06
– *pseudo* EST5	0.25	0.26	0.06	0.20	0.39	0.69	0.26	0.32
Divergence of *mel* EST6 from the non-orthologous								
– virilis *EstS* product	0.42	0.26	0.47	0.44	0.96	2.08	0.65	0.68
– buzz *EstB* product	0.54	0.26	0.29	1.25	0.75	0.98	0.70	0.73

[a]Variation is calculated as $\hat{\pi}$ or $\hat{d}$ as in Table 1, and average pairwise divergence as in Ref. 26.
[b]Categories of sequence are summarized in Box 1, and detailed in Refs 16 and 17.
[c]Data are from Refs 7,10,15,16,24 and 25. Species abbreviations are as in Table 1, plus virilis = *D. virilis* and buzz = *D. buzzatii*.

changes to the catalytic triad and the Gly-x-Ser-x-Gly pentapeptide. Cysteine and salt bridges also differ, suggesting large-scale restructuring of the molecule. Detailed biochemical and physiological studies of these non-esteratic proteins are now needed to elucidate their functions.

Variation and selection in the *Est6* promoter

No EST6-null variants have been recovered from over 1000 field-derived isochromosomal lines of *D. melanogaster* that have been assayed electrophoretically for EST6 activity[16]. This suggests a selective premium on at least some EST6 activity. However, quantitative study of about 40 lines of this species has uncovered up to threefold activity differences in larvae, pupae, adult males and adult females[45,52]. Furthermore, the differences in these four activity measures are only weakly correlated with each other or with EST6 allozyme status. This suggests that the variation in the four measures is because of several distinct regulatory polymorphisms that have largely independent effects on EST6 protein levels in the different life stages and sexes.

Unlike the allozyme work, laboratory studies on EST6 activity differences reveal some relatively clear associations with fitness. Lines of *D. melanogaster* with low larval EST6 activities also show slower larval development times[45]. Several components of male reproductive fitness are also correlated with adult male activities[46], although the direction of the correlation varies between the components measured. Optimum EST6 activity levels might therefore vary in field populations with differences in population structures, mate competition pressures, and so on.

No allelic sequence data have been published for the regulatory regions of the *Est6* gene in *D. melanogaster*, but the 40 lines of this species assayed for EST6 activity variation above have been surveyed for restriction fragment length polymorphisms (RFLPs) in and around *Est6*[45,53]. Three 5′ polymorphisms are associated with activity differences and, from the evidence above, would be candidates as targets for selection. One polymorphism which is scored by the loss of a *Rsa*I site disrupts a 16 bp palindrome situated about 530 bp 5′ of the *Est6* coding region. Loss of this site is associated with about 25% less EST6 activity in adult males. The other two involve relics of what appears to have been a transposable element (the retrotransposon 17.6) inserted about 1.4 kb 5′ of *Est6*. These two relics are associated with 50% reductions in larval EST6 activity.

Sequence data for four isolates of *Est6* from *D. simulans*[24] show a highly non-random distribution of polymorphism along the first 1.15 kb of 5′ flanking DNA. The level of polymorphism in the first 350 bp immediately 5′ of the transcriptional start site is only 6% of the presumptively unconstrained level of silent site variation within the gene, while the corresponding ratio for the next 800 bp is 41% (Table 1, Ref. 24). The value for the first 350 bp 5′ (termed the P/S ratio hereafter) is very similar to the analogous R/S ratio for the coding region, suggesting that regulatory elements closely adjacent to the *Est6* gene are constrained to about the same extent as amino acid replacement sites within the gene. This P/S value is also substantially smaller than the corresponding P/S values for other *Drosophila* genes for which analogous data are available (Table 1). Thus, *Est6* is more polymorphic than most other genes studied in these species in terms of amino acid variation but relatively depauperate for polymorphisms in the first 350 bp of its promoter.

As was the case for the R/S values, comparisons across *D. melanogaster*, *D. simulans*, *D. mauritiana* and *D. pseudoobscura* reveal a trend for larger interspecific than intraspecific values for P/S (Table 1). Among the three *melanogaster* group species P/S values are about the same as the corresponding R/S values, as in the intraspecific data. However, the promoter for *Est5* in *D. pseudoobscura* is so divergent from the others that most of the first 350 bp cannot even be aligned with them. On one hand, such a high level of divergence is surprising, given that the majority of *Est6/5* expression sites are still shared across the two species[6]. On the other hand, some of the divergence presumably underlies the large differences in ejaculatory duct and eye expression between *D. melanogaster* and *D. pseudoobscura* (see above).

Targets for selection in the *Est6* promoter

Sequence–structure–function assignments are currently much less precise for the *Est6* promoter than they are for the EST6 protein. Nevertheless, about 1.15 kb of 5′ flanking DNA is known to be sufficient to direct most aspects of wild-type *Est6* expression, and this 1.15 kb has so far been partitioned into six regions with qualitatively different tissue-specific controls of expression (Box 2). Some novel insights into the promoter's evolution emerge when we map 5′ sequence variation within and among species onto the six different regions (Table 3).

We can now see why polymorphism and divergence within the *melanogaster* subgroup are much lower in the first 350 bp 5′ than in the next 800 bp. The first 350 bp 5′ contain the control elements for seven distinct sites of expression, as well as basic (e.g. the TATA box) signals that are required for expression in any tissue. By contrast, just two sites of expression are attributed to the next 800 bp. Of the three regions within the first 350 bp analyzed experimentally the first two, covering about 250 bp, are essentially invariant, and the third is nearly so. This suggests that a majority of the 350 bp may be constrained by control functions, either for the expression phenotypes so far assigned there, or for others yet to be identified. Conversely, the three regions in the next 800 bp show progressively higher levels of variation, suggesting that progressively less of the sequence is constrained by conserved control functions. One major site of expression, the ejaculatory duct, is controlled by the largest and most distal of these regions, but the latter is so large (about 600 bp) that much of it could still be irrelevant to function.

We can also comment further now on the three RFLPs associated with activity variation in *D. melanogaster*. One of these, a *Rsa*I site at about –530 bp, is associated with male activity variation and could be part of the sperm ejaculatory duct element assigned to the distal 600 bp region above. The other two are insertion polymorphisms at a site (–1.4 kb) beyond the 1.15 kb found to be qualitatively sufficient for wild-type *Est6* expression. The reason that these insertions disrupt larval *Est6* expression may not be that they have inserted into and inactivated a specific larval control element. Instead, their own transcription in larvae may interfere with *Est6* transcription at that stage, a phenomenon also reported for a *copia* element 5′ of the *Adh* gene[56].

As stated earlier, only about 100 bp of the first 1.15 kb 5′ of *D. melanogaster Est6* can be aligned with the 5′ region of *D. pseudoobscura Est5*. This suggests major differences in the organization of the two promoters and reinforces the anomaly noted above that the majority of expression sites are still shared by the two enzymes. Significantly, most of the alignable sequences are in the first two regions of the promoter. These regions include the TATA box and also, presumably, some of the elements responsible for expression sites like haemolymph that are shared by the two genes. One more distal element that can be aligned is the palindrome at about –530 bp, which is associated with male activity in *D. melanogaster* and aligns with sequences around –250 bp in *D. pseudoobscura*. This is consistent with the expression of the enzyme in the ejaculatory duct of *D. pseudoobscura*, albeit at much lower levels than in *D. melanogaster*. The species differences in enzyme level in this tissue could then be because of the few mismatches in the palindrome between the species, or effects of other elements in the distal region of the *D. melanogaster* promoter.

Conclusions

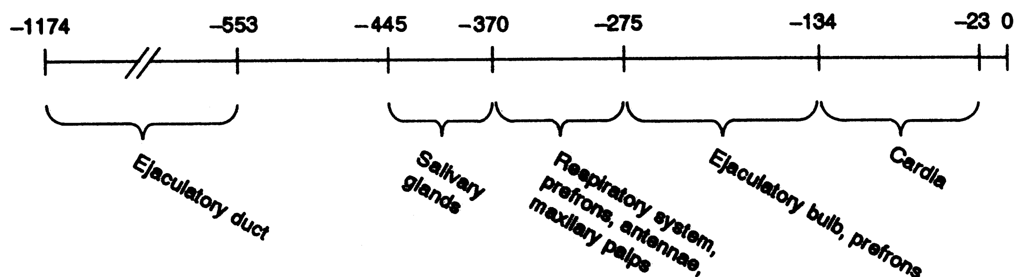
Molecular data confirm earlier biochemical evidence that *Est6* is a highly variable gene in a highly variable esterase gene cluster. As expected, the molecular data provide mechanistic insights into the biochemical variation and its possible adaptive significance. However, they also reveal some qualitatively new and unsuspected phenomena in the evolution of the cluster, the structure of the proteins encoded, and their expression.

Even within the genus, the cluster has acquired several qualitatively new functions. We do not know what these functions are but several may involve male reproductive biology. Haemolymph activity seems to be the major ancestral expression phenotype for EST6, and in a few species this activity is retained while high ejaculatory duct activity is acquired. In others, a duplicated copy of the gene has lost haemolymph activity and acquired ejaculatory bulb expression. Some of these species also possess a further copy whose product is again expressed at high levels in the ejaculatory bulb but which has lost its esterase activity. One priority now is to elucidate the reproductive functions, both of EST6 and of the latter, non-esteratic members of the cluster.

EST6 is one of the most variable enzymes characterized in terms of amino acid differences. Orthologous comparisons within a subgenus show differences in as many as a quarter of its residues and polymorphism within species is far greater than necessary to account for observed allozyme variation. Most of the amino acid differences lie in regions that appear not to be critical to the structure and activity of the molecule, and these may be inconsequential to selection and adaptation. However, a minority of the polymorphisms lie in regions of the protein with two very specific functions, namely the binding of substrates and attachment of carbohydrate side chains. Laboratory data suggest that the polymorphisms in these regions affect the enzyme's substrate specificities and the fate of the enzyme transferred from the male to the female during mating. Ecological work is needed to elaborate early demographic data that indicates that some of the allozymes corresponding to these amino acid differences are under selection in the field.

Box 2. Functional analysis of the *Est6* promoter

Our primary source of information on the *Est6* promoter is Ludwig *et al.*[54]. They made several hybrid genes *in vitro* by fusing various segments of 5′ flanking DNA from *D. melanogaster Est6* to the coding region of a reporter gene, *lacZ*, whose product is readily detected histochemically. The lengths of *Est6* flanking DNA used varied from just the first 134 bp 5′ of the coding region in the shortest construct, extending to over 1 kb in the longest. The hybrid genes were then transferred individually into the *D. melanogaster* genome and the resulting strains assayed qualitatively for the tissue-specific expression of the *lacZ* gene. By comparing *lacZ* expression across the various strains Ludwig *et al.* were able to attribute eight distinct sites of wild-type *Est6* expression to specific segments of 5′ flanking DNA.



The figure shows that seven of the eight sites were attributed to the first 450 bp 5′, although one major site, the ejaculatory duct, was due to sequences further upstream of the coding region. Their choice of constructs enabled Ludwig *et al.* to divide the first 450 bp into four roughly equal segments, each of which was found to be qualitatively responsible for at least one site of expression.

Our own results (M.J.H., M.M. Dumancic, A. Cao and J.G.O., unpublished) confirm and extend those of Ludwig *et al.* Our approach has differed in that we have left the various segments of 5′ flanking DNA from *Est6* attached to the *Est6* coding region, plus some of its 3′ flanking DNA. We have then put our constructs into a laboratory strain in which the endogenous *Est6* gene is a null allele and assayed the resulting strains quantitatively for EST6 activity. While Ludwig *et al.* had generally attributed qualitative control of individual expression sites to individual segments of 5′ DNA, we find that expression in many sites is also modified twofold or more by some of the other segments. Intriguingly, these other segments include regions 3′ of *Est6* but within the *EstP* gene. We also find that qualitative control of another major activity site, the haemolymph, is encoded in the first 171 bp of 5′ DNA.

Limited analyses of the *Est6* promoter have also been published for *D. simulans* and *D. mauritiana*[55]. Only a single construct bearing about 1.15 kb of 5′ DNA has been analyzed in each case. As with *D. melanogaster*, it is sufficient, at least qualitatively, for wild-type *Est6* expression. A single construct bearing only about 500 bp of 5′ DNA from the *Est5* of *D. pseudoobscura* has also been studied[5]. This proves to be sufficient for most aspects of wild-type *Est5* expression, including eye activity, which *Est6* lacks. It is also noteworthy that wild-type *Est5* is only expressed at low levels in the ejaculatory duct, the higher level of *Est6* expression in this tissue being inherited more distally in *D. melanogaster*.

Table 3. Nucleotide variation[a] in six regions of the 5′ flanking DNA of *Est6* analyzed for tissue-specific control functions in *D. melanogaster* (see Box 2)[b]

	–1174 to –554 ejaculatory duct	–553 to –446	–445 to –371 salivary glands	–370 to –276 respiratory system, prefrons, antennae maxillary palps	–275 to –135 ejaculatory bulb, prefrons	–134 to –23[c] cardia
					haemolymph	haemolymph
Polymorphism for *Est6* – in *sim*	0.04	0.02	0.01	0.02	0	0
Divergence of *mel Est6* from the orthologous						
– *sim Est6*	0.11	0.10	0.04	0.04	0.01	0
– *mau Est6*	0.12	0.11	0.04	0.03	0.01	0
– *pseudo Est5*	mostly unalignable	mostly unalignable	mostly unalignable	mostly unalignable	0.23 for 35 bp (otherwise unalignable)	0.29 for 63 bp (otherwise unalignable)

[a] Variation is calculated as $\hat{\pi}$ or $\hat{d}$ as in Table 1.
[b] Data are from Refs 1,5,7 and 24. Species abbreviations are as in Table 1.
[c] Coordinates for the *mel* sequence, where 0 is the first nucleotide in the coding sequence.

Patterns of sequence variation vary widely across the *Est6* promoter, with the first 350 bp 5′ of the gene much more conserved than more distant parts of the promoter. As with the coding region of the gene, some variation within and among the *melanogaster* subgroup species appears to lie in promoter elements with very specific functions, like the control of ejaculatory duct expression. We need more molecular work to elucidate these functional assignments, and more ecological work to understand the adaptive significance of the variation therein. A major paradox also emerges in respect of the *D. pseudoobscura* promoter, the sequence of which is so different from the *melanogaster* promoter that most of it is unalignable. Some of the sequence differences will explain two major differences in the tissue-specific expression of the gene across the two lineages. However, the anomaly remains that most aspects of its expression are shared by the different species. How this has been achieved remains a mystery.

References

1 Collet, C. *et al.* (1990) *Mol. Biol. Evol.* 7, 9–28

2 Oakeshott, J.G., Healy, M.J. and Game, A.Y. (1990) in *Ecological and Evolutionary Genetics of Drosophila* (Barker, J.S.F., Starmer, W.T. and MacIntyre, R.J., eds), pp. 359–378, Plenum

3 Richmond, R.C., Nielsen, K.M., Brady, J.P. and Snella, E.M. (1990) in *Ecological and Evolutionary Genetics of Drosophila* (Barker, J.S.F., Starmer, W.T. and MacIntyre, R.J., eds), pp. 273–292, Plenum

4 Keith, T.P. (1983) *Genetics* 105, 135–155

5 Brady, J.P., Richmond, R.C. and Oakeshott, J.G. (1990) *Mol. Biol. Evol.* 7, 525–546

6 Brady, J.P. and Richmond, R.C. (1990) *Proc. Natl Acad. Sci. USA* 87, 8217–8221

7 Brady, J.P. and Richmond, R.C. (1992) *J. Mol. Evol.* 34, 506–521

8 Korochkin, L., Ludwig, M.Z., Poliakova, E.V. and Philinova, M.R. (1987) *Sov. Sci. Rev. F. Physiol. Gen. Biol.* 1, 411–466

9 Korochkin, L. *et al.* (1990) in *Isozymes: Structure, Function and Use in Biology and Medicine* (Markert, C. and Scandalios, J., eds), pp. 399–440, Wiley-Liss

10 Sergeev, P.V. *et al.* (1993) *Nucleic Acids Res.* 21, 3545–3551

11 Enikolopov, G.N., Malevanchuk, O.A., Peunova, N.I., Sergeev, P.V. and Georgiev, G.P. (1989) *Doklady Akad. Nauk. SSSR* 306, 1247–1249

12 Zouros, E., van Delden, W., Odense, R. and van Dijk, H. (1982) *Biochem. Genet.* 20, 929–942

13 Schafer, D.J., Fredline, D.H., Knibb, W.R., Green, M.M. and Barker, J.S.F. (1993) *J. Hered.* 84, 188–194

14 Johnson, F.M. and Bealle, S. (1968) *Biochem. Genet.* 2, 1–18

15 East, P.D., Graham, A. and Whitington, G. (1990) in *Ecological and Evolutionary Genetics of Drosophila* (Barker, J.S.F., Starmer, W.T. and MacIntyre, R.J., eds), pp. 389–406, Plenum

16 Oakeshott, J.G., van Papenrecht, E.A., Boyce, T.M., Healy, M.J. and Russell, R.J. (1993) *Genetica* 90, 239–268

17 Cygler, M. *et al.* (1993) *Protein Sci.* 2, 366–382

18 Albuquerque, C.M.R. and Napp, M. (1981) *Genetics* 98, 399–407

19 Cooke, P.H., Richmond, R.C. and Oakeshott, J.G. (1987) *Heredity* 59, 259–264

20 Labate, J.A., Bortoli, A., Game, A.Y., Cooke, P.H. and Oakeshott, J.G. (1989) *Heredity* 63, 203–208

21 Oakeshott, J.G., Chambers, G.K., Gibson, J.B. and Willcocks, D.A. (1981) *Heredity* 47, 385–396

22 Anderson, P.R. and Oakeshott, J.G. (1984) *Nature* 308, 729–731

23 Jiang, C., Gibson, J.B. and Chen, H. (1989) *Heredity* 62, 193–198

24 Karotam, J., Boyce, T.M. and Oakeshott, J.G. *Mol. Biol. Evol.* (in press)

25 Cooke, P.H. and Oakeshott, J.G. (1989) *Proc. Natl Acad. Sci. USA* 86, 1426–1430

26 Nei, M. (1987) *Molecular Evolutionary Genetics*, Columbia University Press

27 Gojobori, T., Moriyama, E.N. and Kimura, M. (1991) *Methods Enzymol.* 183, 531–550

28 Karotam, J., Delves, A.C. and Oakeshott, J.G. (1993) *Genetica* 88, 11–28

29 Kreitman, M. and Hudson, R.R. (1991) *Genetics* 127, 565–582

30 Schaffer, S.W. and Miller, E.L. (1992) *Genetics* 132, 163–178

31 Bodmer, M. and Ashburner, M. (1984) *Nature* 309, 425–430

32 Jeffs, P.S., Holmes, E.C. and Ashburner, M. (1994) *Mol. Biol. Evol.* 11, 287–304

33 Riley, M.A., Kaplan, S.R. and Veuille, M. (1992) *Mol. Biol. Evol.* 9, 56–69

34 Begun, D.J. and Aquadro, C.F. (1994) *Genetics* 136, 155–171

35 Berry, A.J., Ajioka, J.W. and Kreitman, M. (1991) *Genetics* 129, 1111–1117

36 Kliman, R.M. and Hey, J. (1994) *Genetics* 133, 375–387

37 Colot, H.V., Hall, J.C. and Rosbash, M. (1988) *EMBO J.* 7, 3929–3937

38 Hey, J. and Kliman, R.M. (1993) *Mol. Biol. Evol.* 10, 804–822

39 Seto, N.O., Hayashi, S. and Tener, G.M. (1987) *Nucleic Acids Res.* 15, 10601

40 Blackman, R.K. and Meselson, M. (1986) *J. Mol. Biol.* 188, 499–515

41 Villares, R. and Cabrera, C.V. (1987) *Cell* 50, 415–424

42 O'Neill, M.T. and Belote, J.M. (1992) *Genetics* 131, 113–128

43 Reuter, D., Schuh, R. and Jackle, H. (1989) *Proc. Natl Acad. Sci. USA* 86, 5483–5486

44 Miekle, D.B. and Richmond, R.C. (1991) *Am. Nat.* 138, 762–767

45 Oakeshott, J.G., Saad, M., Game, A.Y. and Healy, M.J. (1994) *Heredity* 73, 160–169

46 Saad, M., Game, A.Y., Healy, M.J. and Oakeshott, J.G. *Genetica* (in press)

47 Danford, N.D. and Beardmore, J.A. (1979) *Biochem. Genet.* 17, 1–22

48 White, M.M., Mane, S.D. and Richmond, R.C. (1988) *Mol. Biol. Evol.* 5, 41–62

49 Myers, M.A., Healy, M.J. and Oakeshott, J.G. (1993) *Biochem. Genet.* 31, 259–278

50 Harel, M. *et al.* (1992) *Proc. Natl Acad. Sci. USA* 89, 10827–10831

51 Morton, R.A. and Singh, R.S. (1985) *Biochem. Genet.* 23, 959–973

52 Game, A.Y. and Oakeshott, J.G. (1989) *Heredity* 62, 27–34

53 Game, A.Y. and Oakeshott, J.G. (1990) *Genetics* 126, 1021–1031

54 Ludwig, M.Z., Tamarina, N.A. and Richmond, R.C. (1993) *Proc. Natl Acad. Sci. USA* 90, 6233–6237

55 Karotam, J. and Oakeshott, J.G. (1993) *Heredity* 71, 41–50

56 Strand, D.J. and McDonald, J.F. (1989) *Genetics* 117, 787–794

In the next issue of *TREE*:

- Do plants [illegible] SOS signals from their infested neighbours? *J. Bruin, M.W. Sabelis* and *M. Dicke*
- [illegible] change in temperate ecosystems – a southern perspective, *V. Markgraf,* [illegible] *G. Hope*
- [illegible] identification in biodiversity research, *M. Edwards* and *D.R. Morse*
- [illegible] cellular life, *J.P. Gogarten*
- [illegible] reproductive life history strategies, *M.M. Clark* and *B.G. Galef, Jr*