Evidence that the *E4* and *FE4* esterase genes responsible for insecticide resistance in the aphid *Myzus persicae* (Sulzer) are part of a gene family

Linda M. FIELD[1] and Alan L. DEVONSHIRE
IACR-Rothamsted, Harpenden, Hertfordshire, AL5 2JQ, U.K.

The amplification of genes encoding the esterases E4 and FE4 is a widespread mechanism of insecticide resistance in the peach-potato aphid, *Myzus persicae* (Sulzer). We present evidence that in susceptible aphids the two genes are adjacent to each other in a head-to-tail arrangement with *E4* upstream of *FE4* and with approx. 19 kb of intervening sequence. There are also at least two other closely related sequences which might come from other members of an esterase gene family, in line with reports of other insect gene families encoding detoxifying enzymes. The close identity between *E4* and *FE4* genes indicates a recent duplication and divergence. The subsequent amplifications giving multiple copies of either *E4* or *FE4* must have involved two separate events, each probably occurring once and then being selected by insecticide exposure and spread by migration. The cloning of sequences upstream of the *FE4* gene suggest, by comparison with *E4*, that the two genes are regulated in different ways. *FE4* has sequences corresponding to a conventional promoter (TATA box and CAP site) that are not present in *E4*; on the other hand, *FE4* lacks the CpG island present 5′ of *E4* genes that may control expression through changes in DNA methylation. The differences are likely to have occurred by the duplication event that gave rise to *E4* and *FE4* leading to different 5′ sequences.

INTRODUCTION

A major cause of insecticide resistance in the peach-potato aphid, *Myzus persicae*, is the amplification of one of two closely related genes encoding the esterases E4 and FE4, which hydrolyse and sequester insecticides. The two enzymes differ slightly in their molecular mass (65 and 66 kDa for E4 and FE4, respectively) [1], resulting from the presence of different stop codons in their genes [2]. Their cDNAs are very similar, showing 99% identity in nucleotide sequences and predicting 98% identity in the amino acid composition of the enzymes [2]. Both forms of amplified gene have seven introns within the coding region that are of the same size and in identical positions in each form [3].

Upstream of the ATG initiation codon, *E4* genes have an approx. 300 bp untranslated leader sequence, interrupted by two introns of 95 and 1280 bp; thus the transcription start site is approx. 1700 bp upstream of the start of translation [4]. There are no TATA or CAAT boxes relative to this position, but the region is enriched in the dinucleotide CpG ([4] and L. M. Field and A. L. Devonshire, unpublished work), analogous to the CpG islands in the promoters of vertebrate housekeeping genes [5]. This indicates that *E4* expression might be influenced by DNA methylation, supported by the finding that the amplified genes are methylated at *Msp*I/*Hpa*II sites within and downstream of the gene [6] and that this is accompanied by over-expression of *E4* [7]. Furthermore and atypically, loss of methylation correlates with gene silencing [8]. In contrast to this detailed knowledge of the 5′ regions of the *E4* gene, sequences upstream of the ATG initiation codon of *FE4* have been unknown, but here we report these sequences and compare them with those of *E4*.

The seven introns common to amplified *E4* and *FE4* show 97% identity with most of the variation occurring in the last intron (L. M. Field, unpublished work), the sequences of which are given in [3] alongside a third, closely related, sequence that was detected in susceptible aphids. The relationship between this sequence and the *E4* and *FE4* genes has not been established, largely because the single-copy genes remain uncharacterized. This raises the question of how *E4* and *FE4* are related in wild-type, susceptible aphids. Their similarity suggests that they could either be alternative alleles at a single locus or that they have arisen by a duplication event with possibly the third sequence found in susceptible aphids coming from another member of a gene family. Here we report evidence for the latter.

MATERIALS AND METHODS

Aphid clones

Parthenogenetic cultures of *M. persicae* were reared on Chinese cabbage plants at 20 °C with a 16 h light:8 h dark regime. Each culture had been established as a clone from a single aphid collected from the field: clones US1L, 1076A, 945A, T1V, 4158, 794J, 4156 and 405D from the U.K.; 944A from Germany; 1202C from Argentina; 1006A from the U.S.A.; FrR from France; and 800F from Italy. The presence of amplified *E4* or *FE4* genes was inferred from the detection of either a 2.2 kb (*E4*) or 1.8 kb (*FE4*) *Msp*I restriction fragment from their DNA [9].

Amplifying and sequencing of 5′ *FE4* and *E4* sequences

Inverse PCR (IPCR) [10] was used to amplify DNA segments upstream of known sequences, as used previously to sequence upstream *E4* sequences [4]. In the present study, regions A and

Abbreviations used: IPCR, inverse PCR; FJ, junction associated with an *FE4* sequence.
[1] To whom correspondence should be addressed.

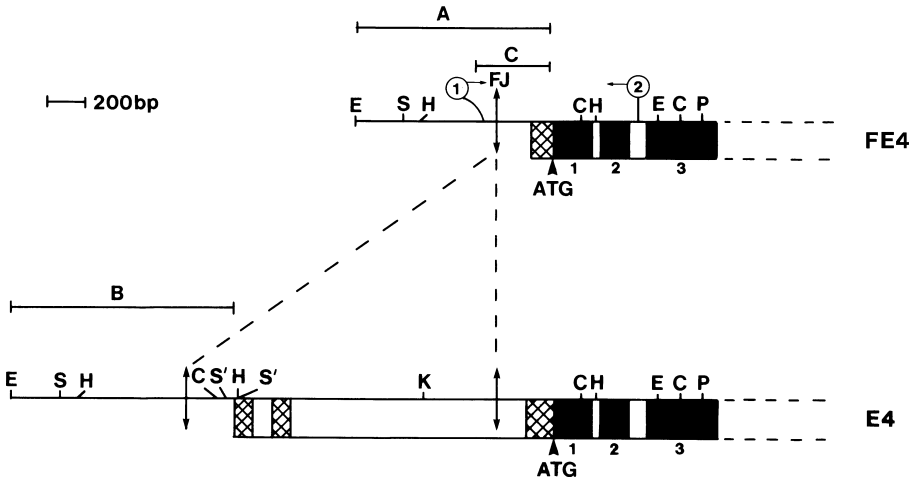

Figure 1 Structure of the 5′ ends of the amplified *E4* and *FE4* genes

□, Introns; ⊠, untranslated regions; ■, translated regions. 1, 2 and 3 indicate the first three of eight exons as defined in [4]. Regions A and B were sequenced directly from IPCR products and the sequence of region C is given in Figure 2. FJ denotes the junction upstream of *FE4* resulting from the absence of approx. 1.7 kb of sequence present in the *E4* gene (as shown). Numbers in circles denote the positions and direction of priming for oligonucleotides used for PCR. Letters show position of restriction sites: C, *Cla*I; E, *Eco*RI; H, *Hpa*I; K, *Kpn*I; S, *Spe*I; S^1, *Sal*I; P, *Pvu*II.

Table 1 Types of sequence detected in aphid DNA using primers specific for either *E4* genes (3 and 4), *FE4* genes (3 and 5) or unamplified sequences (3 and 6)

See Materials and methods for further details. S_2E4 is an unamplified sequence closely related to S_1E4.

		PCR product		
Aphid clone	Amplified gene	3–4	3–5	3–6*
US1L	None	*E4*	*FE4*	S_1E4+S_2E4
944A	None	*E4*	*FE4*	S_1E4
1076A	None	*E4*	*FE4*	S_2E4
4158	*E4*	*E4*	*FE4*	S_2E4
794J	*E4*	*E4*	*FE4*	S_1E4+S_2E4
4156	*E4*	*E4*	*FE4*	S_2E4
800F	*FE4*	*E4*	*FE4*	S_1E4
FrR	*FE4*	*E4*	*FE4*	S_2E4
405D	*FE4*	*E4*	*FE4*	S_2E4

* S_1E4, unamplified sequence previously detected in susceptible aphids [3].

B upstream of the amplified *E4* and *FE4* genes were obtained using ligations between the two *Eco*RI sites (upstream and in exon 3), shown in Figure 1, and 'outward-pointing' 16-mer PCR primers. For both genes a primer in exon 3 upstream of the *Eco*RI site and pointing downstream was used in conjunction with a primer pointing upstream; for *E4* this was the primer used previously to obtain the IPCR product out to the *Spe*I site, and for *FE4* the primer was just downstream from the ATG initiation codon. IPCR products A and B were sequenced directly in both strands, with the primers used for the PCR amplification and oligonucleotides complementary to the new sequences obtained. For *E4*, the sequences downstream of the *Spe*I site in region B were identical to those obtained previously [4].

PCR of gene-specific sequences

Published sequences of intron 7 of the amplified *E4* and *FE4* genes showed differences between them and the presence of another, closely related, sequence in susceptible aphids [3], referred to here as S_1E4. Three PCR primers, each one complementary to only one of the three sequences, were designed for use in conjunction with a common primer to amplify specifically only one sequence (*E4*, *FE4* or S_1E4). The common primer pointing downstream (primer 3 in Table 1) was 5′-GATGATCAAAACTATG-3′, complementary to a region in exon 7 (326 bp from the start of intron 7). The specific primers (38 bp from the end of intron 7) 4, 5 and 6 (see Table 1) were: *E4*, 5′-AAAACATCATCATAGAATA-3′; *FE4*, 5′-GATAGATCCTAGAATA-3′; and S_1E4, 5′-AATACAGCCTAGAATG-3′, respectively. The positions of these primers and the three expected sequences are shown in Figure 2 of [3].

PCR amplifications, using 100 ng of aphid genomic DNA, as described previously [3], gave products that were sequenced directly using an ABI PRISM™- Dye Terminator Ready Reaction Kit and an ABI automated sequencer (type 373A) (PE Applied Biosystems, Warrington, U.K.).

RESULTS

Direct sequencing of IPCR fragments for regions A and B (Figure 1) provided approx. 1.2 kb of sequence upstream of the *FE4* initiation codon and an additional upstream region of the amplified *E4* gene (the *Spe*I to *Hpa*I sequence was reported previously [4]). Only one IPCR product was obtained for *FE4* (using genomic DNA from aphid clone 800F, with amplified *FE4* genes), showing that all copies have the same 5′ end. This contrasts with previous results from clone 794J, which gave two IPCR products because the *E4* gene-amplification event generated repeat units (amplicons) with a novel joint close to the 5′ end of each gene, so that the single copy and end copy of the amplified array have a different 5′ sequence to the amplified copies [4].

```
TGTAGGGAATTTTGGATTCTATGTTGGTACAATTGTTTGAGATTTGCTGG

TTTTTTTTTTAATTTTTTTTTTTTTTTTTTTCCATGAAATTTCGGGAGA
            FJ
AGTACCGATGCCCGTGATGACAGACAAAGCGTGTACATACACTGTATTAT

TGTATAAATTACTTACATATATGATGTCATAAAATAATAATATTATCTAC

TAGGTATCTAATCTAACTGATTATTATTTTCAGATTTTGATTGTTATGAA

TATATACTCGTCCATGAATGAAATTCAGCAAACACATACTATACATTATA

ATTAAAAAATTCATATTTAAGAAATCTGAGAGTACTTAAGTTTTAAAATC

AACATAGGTCTTTTTTGAAAATG.........
```

Figure 2 Sequence at the 5′ end of the amplified *FE4* gene

Part of the sequence of 5′ IPCR product (region C on Figure 1). FJ indicates the *FE4* junction created by the absence of 1.7 kb of *E4* sequence as shown in Figure 1. Transcribed regions are underlined and the initiation codon and putative CAP site and TATA box are boxed.

Comparison of the 5′ sequences of amplified *E4* and *FE4* genes shows that approx. 1.7 kb of sequence present in amplified *E4* is absent in *FE4*, creating a junction (FJ) in the *FE4* sequence, as shown in Figure 1. Rapid amplification of cDNA ends has shown that transcription of *FE4* genes starts 94 bp upstream of the ATG initiation codon, demonstrating that *FE4* genes do not have the extended leader sequence present in *E4*. Part of the 5′ *FE4* sequence (region C in Figure 1) is given in Figure 2, a CAP site immediately 5′ of the transcription start site and a putative TATA box 19 bp upstream suggest that expression of *FE4* genes is controlled by this conventional promoter region.

The regions from the upstream *Eco*R1 sites to the location of the FJ (see Figure 2) are identical for amplified *E4* and *FE4* genes. Since the model for the amplification of *E4* [4] shows that this region is approx. 19 kb downstream of the single-copy *E4* gene (solid region of Figure 3), the 5′ copy of the amplified array of *FE4* genes must be downstream of a copy of the *E4* gene (Figure 3). This in turn suggests that wild-type aphids with unamplified genes have tandem copies of *E4* and *FE4*, with *E4* being 5′ of *FE4* and with approx. 19 kb of intervening sequence. Thus, when one gene amplifies, a single copy of the other gene must also be present, as shown in Figure 3. This model assumes that amplified *FE4* genes are in a tandem array of direct repeats, as already shown for *E4* genes [4].

If this model is correct, then PCR primers 1 and 2 (Figure 1 and Figure 3) should always generate a 865 bp fragment spanning the FJ and the first two exons of the *FE4* gene, whether or not *FE4* is amplified (the same primers should generate an approx. 2.5 kb fragment at the end of *E4* genes). However, less product would be expected in wild-type aphids or those with amplified *E4* genes where only one copy of the 865 bp region is present. Table 2 shows the results using these primers on DNA from 13 aphid clones. All gave the 865 bp product using the same amount of template DNA, but this was more intense (approx. 5–10-fold as judged by ethidium bromide staining) in aphids with amplified *FE4* genes, supporting the model. Some of the PCR products were sequenced (Table 2), confirming that the sequence already found in the 5′ region spanning the FJ was present in all clones and showing that the primers had primed on *FE4* sequence, which differs from *E4* by 1 bp in intron 1 and 2 bp in exon 2 [2].

To test the model further, PCR primers specific for either *E4* or *FE4* genes were used because non-specific, common primers are likely to detect only the amplified gene. This was possible using three primers in intron 7 specific to *E4*, *FE4* or a closely related sequence, S_1E4, found previously in susceptible-aphid DNA [3]. PCR with these and a common primer, followed by direct sequencing of products, showed that *E4* and *FE4* genes were both present in all clones, regardless of whether they had amplified genes (Table 1). The primer specific to the sequence from susceptible aphids also detected a fourth closely related sequence, S_2E4. Some clones had S_1E4, others S_2E4, and in two, i.e. US1L and 794J, separate PCR reactions gave S_1E4 and S_2E4.

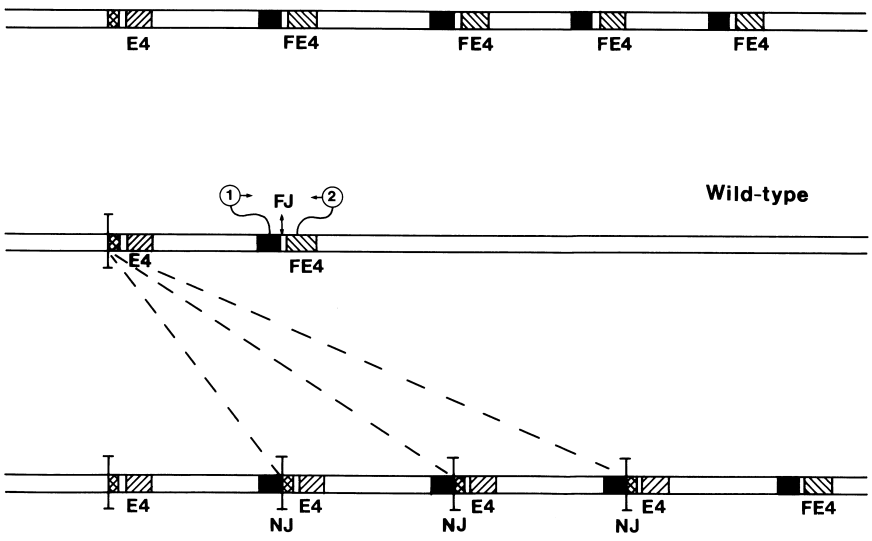

Figure 3 Model for the amplification of *E4* and *FE4* genes

The amplification of *E4* genes creating the novel joint (NJ) has been reported elsewhere [4]. FJ indicates the position of the junction present upstream of *FE4* genes (see Figure 1, Figure 2 and text). Solid, lined and hatched blocks represent 5′ regions with common sequences. Numbers in circles indicate positions of oligonucleotides used for PCR (see Figure 1). Not to scale.

Table 2 Presence of faint (+) or dark (++) PCR products, primed with oligonucleotides 1 and 2 as indicated on Figure 3 and specified in Materials and Methods

Aphid clone	Amplified gene	PCR product (865 bp)
US1L	None	+*
944A	None	+*
1076A	None	+
945A	None	+
1202C	None	+
1006A	*E4*	+*
4158	*E4*	+*
794J	*E4*	+*
4156	*E4*	+
T1V	*E4*	+
800F	*FE4*	++*
FrR	*FE4*	++*
405D	*FE4*	++*

* Fragment sequenced.

It seems likely that all clones have both sequences but that the one primed is random for each PCR. These results again support the proposed model for gene amplification in *M. persicae* and also show that there are at least two other closely related sequences which could be primed from other esterase genes, either alternative alleles or two further loci.

DISCUSSION

The sequence of DNA 5′ of the *FE4* gene suggests that its expression is controlled by a conventional promoter region, in contrast to *E4*, which has no TATA box and seems to be controlled by a 5′ CpG-rich region corresponding to the region missing in *FE4*. This suggests that very different control systems operate for the two genes and is consistent with the observation that the loss of DNA methylation and gene expression, which occurs for *E4* genes [6,8], has not been observed in aphids with amplified *FE4* genes. However, *FE4* genes are methylated and can be partially methylated in some field-collected samples [9]. The role of 5′ CpGs and DNA methylation in controlling aphid esterase gene expression is under further investigation.

The present data also show that susceptible, wild-type *M. persicae* have a copy of the *FE4* gene approx. 24 kb downstream of *E4* (Figure 3) and that there may be at least two other, closely related, genes (or pseudogenes), possibly all members of a gene family. The duplication and divergence of genes to give families is well established and depends on the need for the duplicated gene to acquire a useful function and therefore be retained by natural selection [11]. In the case of *E4* and *FE4* this criterion might be fulfilled by the consequent increase in the capacity to detoxify xenobiotic esters. There are precedents for such families of insect genes being associated with detoxification, such as the clusters of *P450* genes in housefly (*Musca domestica*), thought to have arisen by duplication and inversion events [12], and those associated with insecticide resistance in *Drosophila melanogaster* [13].

Again in *Drosophila*, tandem duplications of the metallothionein (*Mtn*) gene increase cadmium and copper tolerance and may represent the early stages in the evolution of a gene family [14]. There is also good evidence that esterase gene families in *Drosophila* have arisen by duplication and divergence. *D. pseudoobscura* has three tightly clustered paralogous genes (*Est-5A*, *Est-5B* and *Est-5C*) with homologues in *D. melanogaster* [15]. The latter also has a major cluster of 11 genes (the α-cluster), including three encoding the EST23, MCE and ALI esterases, which are homologues of genes involved in insecticide resistance in the sheep blow-fly, *Lucilia cuprina* and in *M. domestica* [16,17]. In turn, the *L. cuprina* gene is itself a member of an α-esterase multigene family [18], and a fourth gene ($\alpha E1$) has as its nearest relative, amongst published sequences, an esterase (*B2*) that when amplified is responsible for resistance in *Culex* mosquitoes [17]. In *Culex*, resistance can be conferred by the co-amplification of genes encoding two esterases (known as A2 and B2 [19] or Estα2-Estβ2 [20]) in a head-to-head configuration, thought to have arisen by a duplication event [20].

For most examples discussed, the duplicated genes have diverged considerably, suggesting that the duplications are ancient (e.g. amino acid identities of approx. 80% for the three *Est-5* genes in *Drosophila*, 37–60% for the genes in the α-cluster and approx. 75% for the *P450* genes), the only exception being the *Mtn* duplication which has a relatively recent origin [21]. The very close identity between *E4* and *FE4* also suggests a very recent duplication, and there must also have been subsequent amplification events to generate multiple copies of one or other gene. We have shown previously that the *E4* event probably occurred once and was spread by migration [4], and the finding of a common 5′ *FE4* sequence for a number of clones in the present study supports a similar view for *FE4*.

Amplified *E4* genes can occur at different chromosome locations, but retain the same amplicon structure [4]. Two of the three aphid clones, with amplified *FE4* genes, used in the present study have such multiple sites [22] and their common 5′ sequences suggest that *FE4* amplicons are also highly conserved. This is analogous to human cell lines, where gene amplification can involve interchromosomal transposition to generate a second locus, whereas the original one is maintained [23].

The selection of aphids with amplified genes that confer insecticide resistance is likely to have occurred within the last 50 years, consistent with the finding that the individual copies of each amplified gene, both within and between aphid clones, show no divergence. We have never found *E4* and *FE4* on the same amplified unit, which is in contrast to some *Culex* esterase genes where the paralogous genes have co-amplified [19,20].

It is notable that the junction (i.e. FJ) created during the ancestral duplication which gave rise to *E4* and *FE4* is only 3 bp downstream of the position of the novel joint created during the subsequent amplification of the *E4* genes, perhaps indicating a 'hotspot' for recombination. However, although the novel joint created during the *FE4* amplification has not been located, it cannot be in the same position since both single-copy and amplified genes have the same sequence over this region. We do know, from restriction mapping, that the size of the *FE4* amplicons is approx. 18 kb (L. M. Field, unpublished work) compared with the 24 kb for *E4* [4].

There is potential for further changes in copy number of aphid esterase genes and for sequence divergence, and with further selection new types of esterase-based resistance to insecticides may evolve. Such an event, known to be present for over 15 years, was recently characterized. A single point mutation in the *Lucilia* ALI gene changes a glycine to an aspartate and converts a carboxylesterase to an organophosphorus hydrolase that confers insecticide resistance [24].

We thank D. Johnston and Z. Harling for rearing well-characterized aphid clones. IACR-Rothamsted receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the U.K.

REFERENCES

1 Devonshire, A. L. (1989) Pestic. Sci. **26**, 375–382
2 Field, L. M., Williamson, M. S., Moores, G. D. and Devonshire, A. L. (1993) Biochem. J. **294**, 569–574
3 Field, L. M., Crick, S. E. and Devonshire, A. L. (1996) Insect Mol. Biol. **5**, 197–202
4 Field, L. M. and Devonshire, A. L. (1997) Biochem. J. **322**, 867–871
5 Bird, A. P. (1995) Trends Genet. **11**, 94–100
6 Field, L. M., Devonshire, A. L., ffrench-Constant, R. H. and Forde, B. G. (1989) FEBS Lett. **243**, 323–329
7 Field, L. M., Devonshire, A. L. and Tyler-Smith, C. (1996) Biochem. J. **313**, 543–547
8 Hick, C. A., Field, L. M. and Devonshire, A. L. (1996) Insect Biochem. Mol. Biol. **26**, 41–47
9 Field, L. M. and Devonshire, A. L. (1992) In Molecular Mechanisms of Insecticide Resistance: American Chemical Society Symposium Series (Mullin, C. A. and Scott, J. G., eds.), pp. 209–217, American Chemical Society, Washington, DC
10 Triglia, T., Peterson, M. G. and Kemp, D. J. (1988) Nucleic Acids Res. **16**, 8186
11 Fryxell, K. J. (1996) Trends Genet. **12**, 364–369
12 Cohen, M. B. and Feyereisen, R. (1995) DNA Cell Biol. **14**, 73–82
13 Maitra, S., Dombrowski, S. M., Waters, L. C. and Ganguly, R. (1996) Gene **180**, 165–171
14 Maroni, G., Wise, J., Young, J. E. and Otto, E. (1987) Genetics **11**, 739–744
15 Brady, J. P. and Richmond, R. C. (1992) J. Mol. Evol. **34**, 506–521
16 Robin, C., Russell, R. J., Medveczky, K. M. and Oakeshott, J. G. (1996) J. Mol. Evol. **43**, 241–252
17 Russell, R. J., Robin, G. C., Kostakos, P., Newcomb, R. D., Boyce, T. M., Medveczky, K. M. and Oakeshott, J. G. (1996) Insect Biochem. Mol. Biol. **26**, 235–247
18 Newcomb, R. D., Campbell, P. M., Russell, R. J. and Oakeshott, J. G. (1997) Insect Biochem. Mol. Biol. **27**, 15–25
19 Rooker, S., Guillemaud, T., Bergé, J., Pasteur, N. and Raymond, M. (1997) Heredity **77**, 555–561
20 Vaughan, A., Hawkes, N. and Hemingway, J. (1997) Biochem. J. **325**, 359–365
21 Theodore, L., Ho, A.-S. and Maroni, G. (1991) Genet. Res. Camb. **58**, 203–210
22 Blackman, R. L., Spence, J. M., Field, L. M. and Devonshire, A. L. (1995) Heredity **75**, 297–302
23 Roelofs, H., Nederlof, P. M., Tasseron-de Jong, J. G., van de Putte, P. and Giphart-Gassler, M. (1992) New Biol. **4**, 75–86
24 Newcomb, R. D., Campbell, P. M., Ollis, D. L., Cheah, E., Russell, R. J. and Oakeshott, J. G. (1997) Proc. Natl. Acad. Sci. U.S.A. **94**, 7464–7468

Received 26 August 1997/10 October 1997; accepted 17 October 1997