Structure and organization of amplicons containing the E4 esterase genes responsible for insecticide resistance in the aphid *Myzus persicae* (Sulzer)

Linda M. FIELD* and Alan L. DEVONSHIRE
IACR-Rothamsted, Harpenden, Herts. AL5 2JQ, U.K.

Insecticide resistance in the aphid *Myzus persicae* results primarily from the amplification of genes encoding the insecticide-detoxifying esterase, E4. Here we report the analysis of flanking DNA co-amplified with the E4 gene. The 5′ end of this gene has an untranslated leader sequence interspersed by two introns, and the promoter region lacks TATA and CAAT boxes. The DNA breakpoint involved in the generation of the amplification is just upstream (approx. 250 bp) of the putative E4 transcription start site; thus the E4 gene is very close to the 5′ end of the approx. 24 kb amplicon. PCR primers specific to the 'novel joint' generated during the amplification have been used to show that a wide range of aphid clones have the same amplicons, arranged as a series of head-to-tail direct repeats. Long-distance mapping has revealed the structure of these repeats. This has important implications for understanding both the generation of the amplified genes and the origin and spread of insecticide resistance in *M. persicae*.

INTRODUCTION

DNA amplification is an increase in the relative amount of a gene or DNA sequence that involves less DNA than a whole chromosome and which can arise either as part of a developmentally regulated programme or as an abnormal spontaneous event [1]. The latter has been studied extensively where it plays a major role in the development of drug resistance in both tumorigenic cells and cultured cell lines [2]. This occurs by the drug selecting for cells that carry units of amplified DNA (amplicons) containing genes that confer resistance by increasing the production of either the drug's target protein or a protein that detoxifies it. Besides the beneficial gene, the amplicon may also contain other genes in usually at least 50 kb of flanking DNA [2]. Indeed, newly amplified genes may be contained in very large regions of DNA, tens of megabases long, which subsequently become more condensed [3].

Studies of developmentally controlled gene amplification have shown that a range of different mechanisms can generate amplified DNA and this is also likely to be so in cultured and tumour cells [1]. When amplification occurs, a novel DNA sequence is generated between amplicons, and studies on the structure of these novel joints can give important information on the mechanisms involved in generating the amplification. The mechanisms can be divided into two major classes involving either over-replication within a cell cycle or unequal segregation [1,4]. Although amplified genes can be found on extra-chromosomal elements (double minutes or episomes), these tend to be unstable and the more stable forms of drug resistance involve tandem arrays of amplicons integrated into extended chromosomes [1]. These amplified units can be heterogeneous (i.e. contain numerous DNA rearrangements) or homogeneous (i.e. with very few or no detectable rearrangements) [2], and can be organized as head-to-tail direct repeats or contain inverted sequences [1]. Evidence for the former has come from studies of the breakpoints between amplicons containing multidrug resistance (*mdr*) genes in *Plasmodium falciparum* [5], whereas, in vaccinia virus, duplicated ribonucleotide reductase genes occurred as both direct and inverted repeats [6]. Sequencing of two joints linking units of amplified DNA containing CAD (the enzyme complex carbamyl phosphate synthetase–aspartate transcarbamylase-dihydro-orotase) genes in *Drosophila* cells gave rise to the hypothesis that in this case two illegitimate recombinations between sister chromatids led to the novel joints and then the amplifications occurred by a series of homologous or further illegitimate recombinations [7]. Indeed, one of the main hypotheses for how gene amplification occurs is by illegitimate recombination leading to unequal sister chromatid exchange and since this predicts that amplicons will be arranged as direct repeats with the selected gene spaced regularly [4], it can be tested by analysing arrangements of amplicons.

Amplification of chorion genes encoding eggshell proteins in insects is probably one of the best understood examples of developmentally regulated DNA amplification. In *Drosophila* two clusters of chorion genes amplify in the follicle cells of the ovary before their transcription. This occurs by multiple rounds of reinitiation of DNA replication within each gene cluster with progressive movement of replication forks on either side to give an 'onion-skin' structure [8].

Insects also provide some of the few examples, other than in cell cultures and tumours, of selection of abnormal spontaneous amplifications. In aphids and mosquitoes, amplifications containing genes encoding insecticide-detoxifying enzymes are responsible for resistance to insecticides [9]. The mosquito *Culex pipiens quinquefasciatus* has amplified B1 esterase genes at a single cytogenic locus [10], but the size and structure of the amplicons is unknown. However, the amplified E4 esterase genes that confer insecticide resistance in the peach–potato aphid, *Myzus persicae*, have been shown to occur as a tandem array of 24 kb repeats in one aphid clone [11]. These genes are situated at a single heterozygous chromosomal site on the short element of chromosome 3, which is involved in an autosomal translocation

Abbreviations used: IPCR, inverse PCR; RACE, rapid amplification of cDNA ends.

* To whom correspondence should be addressed.

commonly associated with resistance in this species [12]. Most other resistant aphid clones with the translocation have amplified E4 genes at the same locus, although the copy number may vary. However, one clone (4156) has additional amplified E4 genes at two other loci on chromosomes 2 and 5 [12].

The chromosomal location of amplified genes encoding the closely related FE4 esterase characteristic of resistant aphids of normal karyotype is more complex [11] and currently under investigation.

Here we report the cloning and sequencing of the 5′ flanking DNA of E4 genes in one typical translocated aphid clone, and the characterization of the novel joints generated during the amplification. This, combined with long-distance mapping, has determined the size, structure and arrangement of E4 amplicons in this clone and enabled a survey of the amplicons present in other aphid clones of wide geographic origin.

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 had been established as a clone from a single aphid collected from the field (Table 1), except for 246T and 271D, which were bred during aphid-crossing experiments involving clones 4156 and 944A [13]. The level of resistance [susceptible (S), resistant (R_1, R_2 or R_3)] for each clone was assessed by immunoassay for the E4 enzyme [14], and the presence of amplified E4 (as opposed to FE4) genes was inferred from the detection of 2.8 and 2.2 kb *Msp*I restriction fragments in the DNA [15].

Isolation and sequencing of aphid genomic DNA sequences

The approx. 2.3 kb 5′ fragment of amplified E4 DNA between *Sal*I and *Eco*RI (as shown in Figure 1) was cloned, after fragment enrichment, into pBluescript as described previously [16]. Positive clones were selected by probing colony blots and plasmid DNA with the 5′ *Kpn*I–*Eco*RI fragment cloned previously [16]. The 2.3 kb fragment was sequenced in both strands using M13 forward and reverse primers and custom-made oligonucleotides (16-mers) complementary to regions at approx. 300 bp intervals.

Table 1 Presence (√) or absence (—) of PCR products using primers 1, 2 and 3 as shown in Figure 2

The level of resistance was judged by immunoassay of E4 enzyme [14]; $R_{1/2}$ indicates E4 activity spanning that typical of R_1 and R_2 clones. S, susceptible; R, resistant.

Aphid clone	Origin	Level of resistance	Amplified E4 genes	PCR product using primers: 1-3	2-3	2-2	3-3
US1L	U.K. 1974	S	—	√*	—	—	—
944A†	Germany 1989	S	—	√*	—	—	—
1076A	U.K. 1992	S	—	√	—	—	—
1172D	Holland 1993	S	—	√	—	—	—
794J	U.K. 1982	R_3	+	√*	√*	—	—
1006A	U.S.A. 1991	$R_{1/2}$	+	√*	√*	—	—
4158	U.K. 1991	R_2	+	√*	√*	—	—
4156	U.K. 1991	R_2	+	√	√	—	—
246T	Bred from 4156	$R_{1/2}$	+	√	√	—	—
271D	Bred from 4156	$R_{1/2}$	+	√	√	—	—
1054D	Japan 1992	R_3	+	√	√	—	—
1134C	Greece 1993	R_3	+	√	√	—	—
1202C	Spain 1994	R_1	+	√	√	—	—
1259W	Jersey 1994	R_3	+	√	√	—	—
1262A	France 1995	R_2	+	√	√	—	—
1277I	Holland 1995	R_1	+	√	√	—	—
1282C	Chile 1996	R_3	+	√	√	—	—

* Fragment sequenced.

† This is the same clone as DS in [13].

Amplifying and sequencing of 5′ E4 sequences

Inverse PCR (IPCR) was used to amplify DNA segments upstream of the known sequences as described by Triglia et al. [17]. Briefly, this involved cutting genomic DNA with a restriction enzyme with sites outside and inside the known sequence, circularizing and then amplifying across the unknown region using PCR primers directed 'outwards' from near the ends of the known sequence. For sequences 5′ of the E4 gene, DNA from the aphid clone 794J was digested with an excess of *Spe*I which cuts approx. 900 bp upstream from the transcription start site (see Figure 1) and again in intron 7 (results not shown) to give an approx. 6.5 kb fragment. Ligation mixtures containing approx. 1 μg of DNA and 20 units of DNA ligase in a total volume of 800 μl were left overnight at 12 °C. The DNA was then purified by phenol/chloroform extraction and resuspended in sterile water (neither cleavage of the circles nor nicking at 94 °C as discussed in [17] was found to be necessary). Approx. 0.1 μg of DNA was used in a 50 μl PCR as described previously [18]. One 16-mer primer was designed from known 5′ sequence and pointed upstream, and the other directed downstream was complementary to a region approx. 300 bp upstream of the *Spe*I site in intron 7 used for the initial cutting. The IPCR products were sequenced in both strands with the primers used for the PCR amplification and oligonucleotides complementary to the new sequences obtained. Approx. 50 ng of each PCR product was sequenced directly using the ABI PRISM™ Dye Terminator Ready Reaction Kit and an ABI automated sequencer (type 373A).

Conventional PCRs using aphid genomic DNA (approx. 100 ng) were performed as described previously [18].

Mapping of DNA flanking the amplified E4 gene

Positions of restriction sites in the DNA flanking the amplified E4 gene have been reported previously [19]. We have now used the same techniques to map more distal sites, through to the adjacent amplicon, using single and double digests probed with two regions of cloned DNA as shown in Figure 4 (i.e. *Sal*I–*Eco*RI at the 5′ end and *Bgl*II–*Eco*RI at the 3′ end).

RESULTS AND DISCUSSION

The sequence of genomic DNA at the 5′ end of the amplified E4 gene was obtained by a combination of cloning and sequencing of the *Sal*I–*Eco*RI fragment (Figure 1) and direct sequencing of IPCR products. A comparison of these sequences with the E4 cDNA already reported [16] showed that upstream of the ATG start codon there is approx. 300 bp of untranslated leader sequence interrupted by two introns of approx. 100 and 1300 bp (Figure 1). This is in contrast with other insect esterase genes such as Est 5B in *Drosophila pseudo-obscura* [20], Est 6 in three species of *Drosophila* [21] and the amplified B1 esterase in *Culex pipiens quinquefasciatus* [21a], all of which have transcription start sites close to the ATG codon with no intervening intron regions.

Sequencing of products of rapid amplification of cDNA ends (RACE) obtained from E4 mRNA suggests that the start of E4

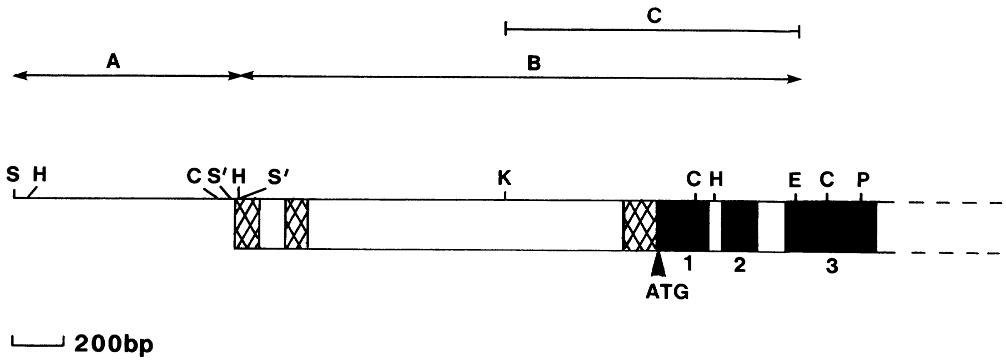

Figure 1 Structure of the 5′ end of the amplified E4 gene

□, Introns; ⊠, untranslated regions; ■, translated regions. 1, 2 and 3 indicate the first three of eight exons as defined in [19]. Region A was sequenced directly from an IPCR product and B by cloning and sequencing of genomic DNA cut with *Sal*I and *Eco*RI (see the Materials and methods section). C indicates the position of the probe used to identify colonies and plasmids containing 5′ E4 sequences. Letters show position of restriction sites: C, *Cla*I; E, *Eco*RI; H, *Hpa*I; K, *Kpn*I; S, *Spe*I; S′, *Sal*I; P, *Pvu*II.

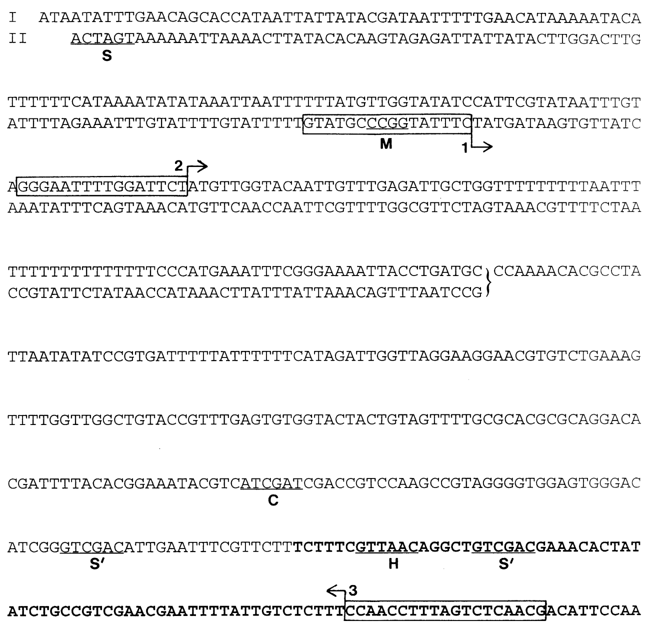

Figure 2 Sequences of the 5′ end of the amplified E4 gene and flanking DNA

Bold type indicates transcribed region (as judged from RACE product; see the text). I and II are alternative sequences upstream of }. Positions of PCR primers 1, 2 and 3 are boxed and their direction of priming indicated by arrows. Restriction sites are C, *Cla*I; H, *Hpa*I; M, *Msp*I; S, *Spe*I; S′, *Sal*I.

gene transcription is as indicated on Figure 2. However, there are no upstream putative TATA or CAAT boxes relative to this position and no typical arthropod initiation sequences [22]. It may be that the RACE product is not full length and the start of transcription is further upstream, or, quite possibly, E4, like other arthropod promoters, either lacks functional initiation sequences or has initiators that are not homologous to known sequences [22a]. The sequences shown in Figure 2 are part of a CpG island at the 5′ end of the E4 gene (L. M. Field and A. L. Devonshire, unpublished work) analogous with promoters of vertebrate housekeeping genes [23] and may indicate that E4 lacks tissue-specific expression. However, the presence of such a CpG island in an invertebrate gene and the reported correlation between DNA methylation and E4 gene transcription [24] is contrary to the generally held views on the role of DNA methylation in gene expression and is currently under investigation.

From our knowledge of the restriction map of E4 amplicons, the use of IPCR to clone 5′ E4 sequences from aphid clone 794J should have given an approx. 1.2 kb product (i.e. 900 bp of unknown sequence from the *Spe*I site to the *Hpa*I site plus approx. 300 bp from the internal primer in intron 7 to the *Spe*I site). This was indeed obtained, but a second product of approx. 750 bp was also produced. The sequences of both fragments, I for the larger product and II for the smaller, are given in Figure 2. Both have the same 222 bp sequence immediately upstream of the putative transcription start site, but thereafter show no homology. Thus two alternative ends of the E4 gene have been produced by IPCR, suggesting that one is the 5′ end of the wild-type unamplified E4 gene (presumably homozygous) and the other spans the novel joint created on one homologue by the amplification event. The creation of such a new breakpoint during DNA amplification has been described by Triglia et al. [5], and the way it would apply for E4 is given in Figure 3.

The *Spe*I and *Msp*I sites present in sequence II of Figure 2 have not been detected by mapping of the DNA flanking the amplified E4 gene [19]. During this mapping the amplified fragments were accompanied by a range of other faint bands, but it has not been possible to assign these to the single-copy gene. Taken together, these data suggest that sequence II corresponds to the 5′ region of both the single- (on one homologue) and the end-copy gene of the amplified array, and sequence I is across the novel joint between E4 amplicons (as shown in Figure 3). To test this hypothesis, primers specific to the two sequences were used for PCR; primers 1 and 3 (Figures 2 and 3) should give a product from a single (or end)-copy E4 gene, but primers 2 and 3 will only do so if the E4 gene is amplified, as shown in Figure 3. When tested on DNA from 17 aphid clones (Table 1), primers 1 and 3 always gave a product for resistant and susceptible aphids, but primers 2 and 3 only did so when the E4 gene was amplified. The authenticity of some of the PCR products was confirmed by direct sequencing (see Table 1). This provides strong support for our hypothesis that sequence I spans the novel joint between E4 amplicons. Furthermore, since PCR with primers 2 or 3 alone (Table 1) was not successful, all of the E4 amplicons must be in a tandem head-to-tail array [5], rather than head-to-head or tail-

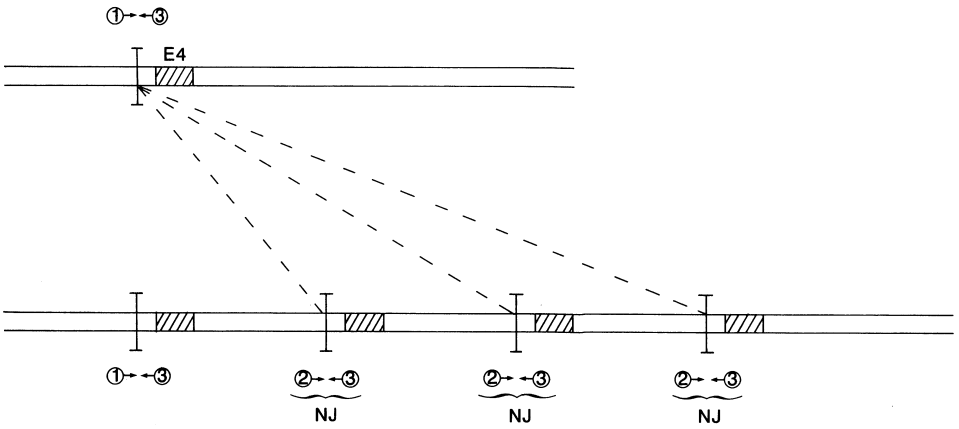

Figure 3 Model (not to scale) for the creation of a novel joint (NJ) during amplification of the E4 gene and flanking DNA

Top, Wild-type single-copy gene; bottom, E4 amplicons in tandem array. Position and direction of priming for primers 1, 2 and 3 are as given in Figure 2. Vertical bars correspond to the position of the } in Figure 2.

to-tail, which would permit amplification by single primers 2 or 3 respectively. Thus each of the resistant aphid clones studied, regardless of resistance level, has the same E4 amplicons in the same type of array.

The presence of identical novel joints between the amplicons of the 13 resistant aphid clones studied (11 of independent origin) shows that either the same event has occurred many times or, more likely, that the E4 amplification arose only once and subsequently spread throughout the world. This could result from migration, coupled with worldwide selection of the amplified genes, as has been suggested for the B1 genes in mosquitoes [25]. In aphids, spread could occur naturally by long-range aerial movement [26] or perhaps more likely as a consequence of the international trade in plants and produce which harbour insect pests [27]. The latter form of dispersal has been suggested to explain the sudden widespread appearance of the B-biotype of the whitefly *Bemisia tabaci* [28]. Insecticide-sensitive *M. persicae* are still common throughout the world, suggesting that selection for resistance by insecticide treatment is somehow counteracted by selection against the resistant phenotype. This has been demonstrated for U.K. populations, where resistant aphids showed poorer survival in the winter climate [29].

As described in the Introduction, clone 4156 has amplified E4 genes on chromosomes 2, 3 and 5 [12], from which 271D and 246T have inherited the amplicons on chromosomes 5 and 3 respectively. Thus, even when the E4 genes are at different loci, the same amplicon structure is retained. Although the mechanism by which the genes were relocated to different chromosomes is not known, an involvement of transposable elements is possible. There are inverted repeat sequences within the E4 amplicon, approx. 1 kb downstream of the E4 gene (L. M. Field and A. L. Devonshire, unpublished work), but any involvement with either transposition or the generation of the E4 gene amplification has not yet been established. Repeat sequences, which might be functionally related to transposable elements, have been found in amplicons containing the B1 gene of *Culex* [21a], but again their contribution to the amplification process is unclear. The wider role of transposable elements as indicators of insecticide resistance has been discussed by Wilson [30].

Mapping of DNA flanking the amplified E4 gene in aphid clone 794J has identified the sites shown between those marked with an asterisk in Figure 4. The three 5′-most sites within this 26 kb region (*Pvu*II, *Bam*HI and *Xba*I) are the same as the three 3′-most sites, and, since the repeat unit in 794J is known to be approx. 24 kb [1], it is likely that these sites are on adjacent repeat units. Taken together with the above evidence on the novel joints between amplicons, the proposed arrangement in Figure 4 is very well substantiated.

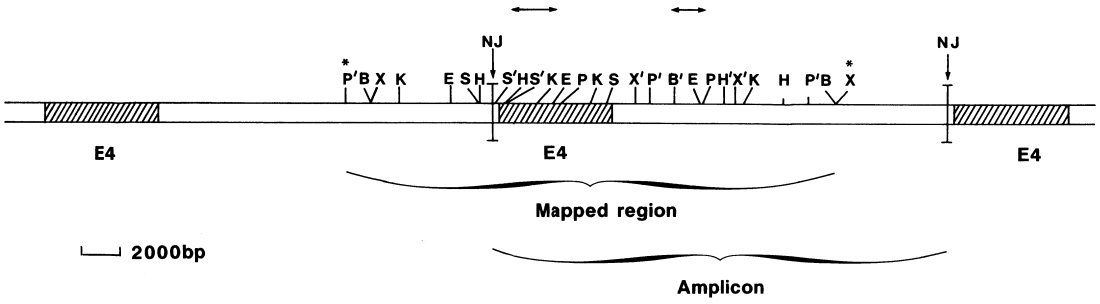

Figure 4 Proposed arrangement and restriction map of E4 amplicons

B, *Bam*HI; B′, *Bgl*II; E, *Eco*RI; H, *Hpa*I; H′, *Hin*dIII; K, *Kpn*I; P, *Pvu*II; P′, *Pst*I; S, *Spe*I; S′, *Sal*I; X, *Xba*I; X′, *Xho*I. ↔ indicates regions used as probes, NJ↓ is the novel joint between amplicons and * shows sites referred to in the text.

The finding that amplified E4 genes are on direct head-to-tail tandem repeats of approx. 24 kb of DNA is consistent with a model for the generation of amplified DNA by unequal sister chromatid exchange. There are no obvious homologies or strings of adenine bases sometimes found at breakpoints [5], which might allow misalignment, but the region around the joint is AT-rich and such regions have been associated with recombination points in amplification systems [7]. The unequal sister chromatid exchange model predicts an initial formation of chromatids with duplicated and deleted arms and then subsequent expansion by homologous but unequal exchanges leading to homozygous amplifications. In aphids, the E4 gene amplification is heterozygous [12,13], but this may be because the evolution of the amplification is still in the early stages. Most of the models for the generation of amplified DNA are based on studies of cell cultures and allow for the production of very large repeat units [7], which is not the case in aphids. This may indicate that there are more constraints to major genome rearrangement in a whole organism than there are in cell cultures.

We thank Diana Johnston and Zoë Harling for rearing the aphid clones and measuring their esterase content by immunoassay. IACR-Rothamsted receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the U.K.

REFERENCES

1 Stark, R. E., Debatisse, M., Wahl, G. M. and Glover, D. M. (1990) in Gene Rearrangement (Hames, B. D. and Glover, D. M., eds.), pp. 99–149, IRL Press, Oxford
2 Schoenlein, P. V. (1994) in Anticancer Drug Resistance: Advances in Molecular and Clinical Research (Goldstein, L. J. and Ozols, R. F., eds.), pp. 167–200, Kluwer Academic Publishers, Boston
3 Smith, K. A., Gorman, P. A., Stark, M. B., Groves, R. P. and Stark, G. R. (1990) Cell **63**, 1219–1227
4 Stark, G. R. (1993) Adv. Cancer Res. **61**, 87–113
5 Triglia, T., Foote, S. J., Kemp, D. J. and Cowman, A. F. (1991) Mol. Cell. Biol. **11**, 5244–5250
6 Slabaugh, M. B., Roseman, N. A. and Mathews, C. K. (1989) Nucleic Acids Res. **17**, 7073–7088
7 Azou, Y. and Laval, M. (1993) Biol. Cell **77**, 155–164
8 Orr-Weaver, T. L. (1991) Bioessays **13**, 97–105
9 Devonshire, A. L. and Field, L. M. (1991) Annu. Rev. Entomol. **36**, 1–23
10 Nance, E., Heyse, D., Britton-Davidian, J. and Pasteur, N. (1990) Genome **33**, 148–152
11 Field, L. M., Devonshire, A. L. and Tyler-Smith, C. (1996) Biochem. J. **313**, 543–547
12 Blackman, R. L., Spence, J. M., Field, L. M. and Devonshire, A. L. (1995) Heredity **75**, 297–302
13 Blackman, R. L., Spence, J. M., Field, L. M., Javed, N., Devine, G. J. and Devonshire, A. L. (1996) Heredity **77**, 154–167
14 Devonshire, A. L., Moores, G. D. and ffrench-Constant, R. H. (1986) Bull. Entomol. Res. **76**, 97–107
15 Field, L. M. and Devonshire, A. L. (1992) in Molecular Mechanisms of Insecticide Resistance: ACS Symposium Series (Mullin, C. A. and Scott, J. G., eds.), pp. 209–217, ACS, Washington D.C.
16 Field, L. M., Williamson, M. S., Moores, G. D. and Devonshire, A. L. (1993) Biochem. J. **294**, 569–574
17 Triglia, T., Peterson, M. G. and Kemp, D. J. (1988) Nucleic Acids Res. **16**, 8186
18 Field, L. M., Crick, S. E. and Devonshire, A. L. (1996) Insect Mol. Biol. **5**, 197–202
19 Field, L. M., Javed, N., Stribley, M. F. and Devonshire, A. L. (1994) Insect Mol. Biol. **3**, 140–148
20 Brady, J. P., Richmond, R. C. and Oakeshott, J. G. (1990) Mol. Biol. Evol. **7**, 525–546
21 Karotam, J., Delves, A. C. and Oakeshott, J. G. (1993) Genetica **88**, 11–28
21a Mouches, C., Pauplin, Y., Agarwal, M., Lemieux, L., Herzog, M., Abadon, M., Beyssat-Arnaouty, V., Hyrien, O., De Saint Vincent, B. R., Georghiou, G. P. and Pasteur, N. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 2574–2578
22 Cherbas, L. and Cherbas, P. (1993) Insect Biochem. Mol. Biol. **23**, 81–90
22a Böhm, S. K., Gum, J. R., Erickson, R. H., Hicke, J. W. and Kim, Y. S. (1995) Biochem. J. **311**, 835–843
23 Bird, A. P. (1995) Trends Genet. **11**, 94–100
24 Field, L. M., Devonshire, A. L., ffrench-Constant, R. H. and Forde, B. G. (1989) FEBS Lett. **243**, 323–329
25 Raymond, M., Callaghan, A., Fort, P. and Pasteur, N. (1991) Nature (London) **350**, 151–153
26 Loxdale, H. D., Hardie, J., Halbert, S., Footit, R., Kidd, N. A. C. and Carter, C. I. (1993) Biol. Rev. **68**, 291–311
27 Frey, J. E. (1993) in Plant Health and the European Single Market (Ebbels, D., ed.), pp. 157–165, British Crop Protection Council Monograph No. 54, Surrey
28 Denholm, I., Cahill, M., Byrne, F. J. and Devonshire, A. L. (1995) in *Bemisia*: 1995 Taxonomy, Biology, Damage, Control and Management (Gerling, D. and Mayer, T., eds.), pp. 577–604, Intercept, Andover
29 Foster, S. P., Harrington, R., Devonshire, A. L., Denholm, I., Devine, G. J. and Kenward, M. G. (1996) Bull. Entomol. Res. **86**, 17–27
30 Wilson, T. G. (1993) J. Econ. Entomol. **86**, 645–651

Received 9 September 1996/12 November 1996; accepted 12 November 1996