Nucleotide Polymorphism in the *Est6* Promoter, Which Is Widespread in Derived Populations of *Drosophila melanogaster*, Changes the Level of Esterase 6 Expressed in the Male Ejaculatory Duct

Wendy A. Odgers,*,†,‡,[1] Charles F. Aquadro,† Christopher W. Coppin,*,‡ Marion J. Healy*,[2] and John G. Oakeshott*,[3]

**CSIRO Entomology, Canberra, 2601 Australia,* †*Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853 and* ‡*Division of Biochemistry and Molecular Biology, Australian National University, Canberra, 2601 Australia*

Manuscript received November 2, 2001
Accepted for publication July 8, 2002

ABSTRACT

Previous analysis of an Australian population of *D. melanogaster* revealed two predominant *Est6* promoter haplotypes, P1 and P7. These haplotypes, which differ at 14 sites over a 325-bp region, are associated with a 15–20% difference in male EST6 activity. Here we show that the P1/P7 sequence difference causes the male activity variation by recreating the activity difference among >60 independently transformed lines containing representative P1 or P7 promoter alleles fused to an identical *Est6* coding region. Furthermore we find that the whole fly difference reflects about a twofold difference in EST6 activity in the anterior sperm ejaculatory duct. EST6 activity variation in this tissue is known to affect reproductive fitness. Using a combination of RFLP analysis and DNA sequencing, we show that P1 and P7 are predominant in six populations from America, Asia, and Australia, albeit less frequent in a population from the presumptively ancestral east African range of the species. The sequence data show significant departures from neutral expectations for the derived American and Australian populations but not the presumptively ancestral Zimbabwean population. Thus the P1/P7 difference could be a major source of adaptively significant EST6 activity variation through much of the now cosmopolitan range of *D. melanogaster*.

IT is widely accepted that qualitative and/or quantitative differences in gene expression could be important sources of adaptive phenotypic variation. However, there are few empirical data on the population genetics of regulatory polymorphisms and their contribution to phenotypic variation (but see Choudhary and Laurie 1991; Laurie and Stam 1994; Stam and Laurie 1996 for notable exceptions). Here we investigate the relationships between 5′ nucleotide polymorphism and expression differences for the Esterase 6 (*Est6*/EST6) gene/enzyme system in *Drosophila melanogaster*.

In a survey of nucleotide sequence variation in the 5′ promoter region of *Est6* in *D. melanogaster* Odgers *et al.* (1995) found that most of the haplotypes in an Australian population fell into two distinct groups, termed P1 and P7. These haplotypes differ from one another at all 14 commonly polymorphic sites in a 325-bp region between 494 and 819 bp upstream of the *Est6* translation start site. The two haplotype groups are also associated with different EST6 activity phenotypes; adult males with the P1 haplotype show 15–20% lower EST6 enzyme activity than that of males with the P7 haplotype. Females do not show this difference.

Here we investigate the physiological basis for the different EST6 activity phenotypes associated with the P1/P7 promoter difference and examine the distribution of the P1 and P7 haplotypes across the species range. Representative P1 and P7 promoter alleles fused to an identical *Est6* coding region are transformed into an EST6 null background using *P*-element-mediated germline transformation. The resultant transformed lines are tested for EST6 activity in males, females, and male ejaculatory ducts. The male ejaculatory duct is part of the reproductive tract and EST6 activity in this tissue normally comprises 20–60% of the total EST6 activity in whole males and ~1% of total protein in this tissue (Myers *et al.* 1996). Our results show that the 14 nucleotide differences distinguishing P1 and P7 cause the 15–20% difference in whole male EST6 activity and that this difference is explained by about a twofold difference in the level of EST6 activity in the male ejaculatory duct. Furthermore, using a combination of restriction fragment length polymorphism (RFLP) analysis and DNA sequencing, we show that these two *Est6* promoter haplotypes are present at high frequencies in six presumptively derived populations from America, Asia, and Australia but are found only at low frequencies in a

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY058798–AY058828, L10670, and L34263–L34265.

[1]*Present address:* BioSecurity Australia, AFFA, GPO Box 858, Canberra, ACT, 2601 Australia.

[2]*Present address:* ANZFA, PO Box 7186, Canberra Mail Centre, ACT, 2610 Australia.

[3]*Corresponding author:* CSIRO Entomology, GPO Box 1700, Canberra, ACT, 2601 Australia. E-mail: john.oakeshott@csiro.au

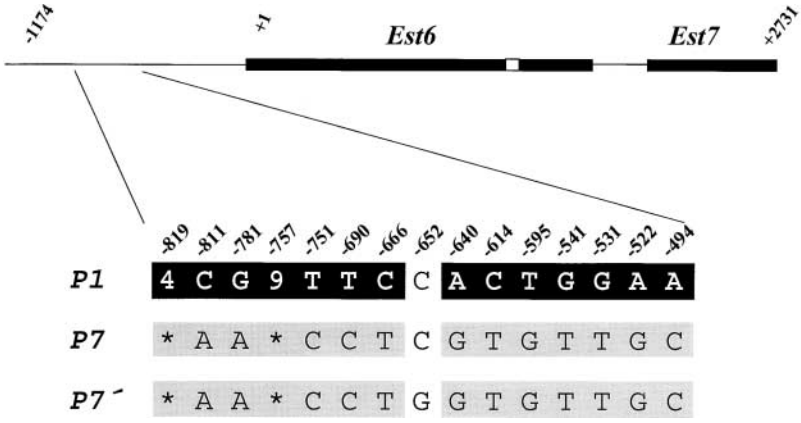

FIGURE 1.—*Est6* promoter constructs used for germline transformation. The genomic organization of the *Est6* and *Est7* region is shown. Only part of the *Est7* coding region is shown. The *Est6* translation start site is +1. Exons and introns are shown in black and white bars, respectively. The relative positions of the nucleotide differences that distinguish P1 from P7 are also shown. Constructs *P7*, *P7′*, and *P1* extended from ~1174 bp upstream to 2731 bp downstream of the *Est6* translation start site. The enlarged portion shows the nucleotide differences between the three constructs, with nucleotides diagnostic for P1 and P7 in black and gray, respectively. *P7* and *P7′* have 5- and 35-bp deletions (denoted indels 4 and 9, respectively; Figure 3) relative to *P1* at positions −819 and −757 bp, respectively. Refer to Figure 3 for full details of nomenclature.

population from the ancestral range of the species in East Africa.

MATERIALS AND METHODS

The transformation analysis of the P7 haplotype was initially performed using a construct we now call *P7′* (constructs are hereafter italicized to distinguish them from haplotypes) since it was subsequently discovered that *P7′* contained a nucleotide at −652 bp that differed from both the P1 and P7 sequences. This nucleotide difference involves a rare naturally occurring variant (see Figures 1 and 3 and ODGERS *et al.* 1995) present in the Dm145 (P7′) sequence. Therefore a smaller data set was subsequently collected for a construct (*P7*) containing the P7 sequence *sensu stricto*.

The insert fragment for all transformation constructs was a 3.9-kb genomic *Hin*dIII fragment comprising 1174 bp upstream of the *Est6* translation start site, the 1.68-kb *Est6* coding region, the 197-bp intergenic region, and 851 bp of the downstream *Est7* gene (DUMANCIC *et al.* 1998; Figure 1). For the *P7′* transformation construct the *Hin*dIII fragment was derived from the field-trapped North American strain Dm145 (COLLET *et al.* 1990; KAROTAM *et al.* 1993). The coding region of the Dm145 *Est6* gene specifies an EST6-8 Slow allozyme (COOKE and OAKESHOTT 1989) and its 5′ region belongs to the P7 group of promoter haplotypes as explained above. The insert fragment for transformation construct *P1* was identical to that of *P7′* except for the 14 nucleotide differences (12 substitutions, two indels) between −819 and −494 bp that define the P1 group of promoter haplotypes, plus the additional nucleotide difference at −652 bp described above. The insert fragment for transformation construct *P7* differed from that of *P7′* only at the −652-bp site. The *P1* 5′ and *P7* 5′ regions were obtained from representative lines with the P1 or P7 haplotype in ODGERS *et al.* (1995), using PCR with the primers and conditions described in ODGERS *et al.* (1995). The PCR-amplified regions (−1007/+80) from the two lines were then cloned and sequenced to verify the absence of PCR errors using standard procedures. Their respective *Nhe*I (−865)/*Bgl*II (−133) fragments were then submitted into the corresponding *Nhe*I and *Bgl*II sites of the *P7′* insert fragment in the vector pTZ19U (Pharmacia, Piscataway, NJ).

The insert fragments for the *P7′* and *P1* constructs were then subcloned into the transformation vector CaSpeR (ROBERTSON *et al.* 1988). Insert fragments were prepared by digestion with *Hin*dIII (recognition sites at −1174 and +2731 bp), purified, end filled using DNA polymerase I, ligated to *Bcl*I linkers, digested with *Bcl*I, and inserted into the *Bam*HI site of CaSpeR following standard procedures. The insert fragment for the *P7* construct was subcloned directly as a *Nhe*I (−865/+2703) fragment into the corresponding *Nhe*I sites of the *P1* transformation construct. The constructs chosen for germline transformation had the modified *Est6* gene and the marker gene *white*$^+$ transcribed in the same direction. The construct *P7′* was generated by HEALY *et al.* (1996; where it is termed −1174/+2731).

Transformation constructs were purified via two sequential cesium chloride gradients and microinjected into embryos following the method described in HEALY *et al.* (1996). Emergent adults were individually mated to the *w;Est6*$^{\emptyset}$ line (SHEEHAN *et al.* 1979) and flies in which the modified *Est6* gene had integrated were selected on the basis of eye color changes. Transformed lines were made homozygous for both the *Est6*$^{\emptyset}$ gene on the third chromosome (68F7-8; FLYBASE 1999) and a transformed *Est6* gene on either the X or the second chromosome (by using the partial dominance of the *miniwhite* marker gene in CaSpeR to distinguish homozyotes and heterozygotes for the transformation construct). For the *P7′* and *P1* constructs, 26 and 30 independent lines, respectively, were isolated for analysis. Eleven independent *P7* lines were obtained. Lines that were used for tissue dissections were analyzed by Southern blot hybridization to verify the structural fidelity of the inserted gene and to ensure a single-copy insertion event. All Drosophila stocks were cultured on standard agar-cornmeal-treacle medium at 25° in a 12-12 light-dark cycle and 60% humidity.

EST6 enzyme activity assays were performed on whole male and female adults and dissected male ejaculatory ducts. In all cases, newly eclosed virgin males and females from duplicate or triplicate cultures were aged separately for 3 days at 25°. For the whole adult assays, 10 individuals each from triplicate cultures were then frozen and stored at −80°. For the ejaculatory duct assays, 10 ducts from males reared in each duplicate culture were dissected into 70 μl sodium phosphate buffer (0.1 M, pH 7.0), frozen, and stored at −80°.

All samples were kept ice-cold during preparation. Whole adults and dissected ejaculatory ducts were homogenized in 200 and 70 μl of sodium phosphate buffer, respectively. Homogenates were centrifuged at 14,000 × *g* for 10 min and the supernatant was divided into aliquots and stored at −80°. Just prior to each assay, samples were centrifuged at 14,000 × *g* for 2 min and the supernatant was diluted in ice-cold sodium phosphate buffer.

The EST6 enzyme activity assay measured the production of β-naphthol from the substrate β-naphthylacetate in the

presence of eserine sulfate (10^{-5} M) and *p*-chloromercuribenzoate (10^{-4} M) to inhibit other esterases. The end-point spectrophotometric method of Sheehan *et al.* (1979) was modified following Healy *et al.* (1996) for use in 96-well microtiter plates. EST6 activity per fly was calculated as nanomoles β-naphthol produced/30 min at 28°. For whole adult homogenates, protein determinations were performed in 96-well microtiter plates using the Bradford assay (Bio-Rad, Richmond, CA, protein assay reagent, catalog no. 500-0006) and bovine serum albumin as a standard. EST6 specific activity was then calculated as micromoles β-naphthol produced per milligram protein per 30 min at 28°. Two-way factorial and single-classification analyses of variance were performed on logarithmically transformed line means for EST6 activity or specific activity using the Statview II software package (Abacus Concepts, Berkeley, CA, 1987).

To test whether the P1 and P7 promoter haplotypes were common in populations other than the one Australian population (Coffs Harbour, New South Wales) analyzed originally by Odgers *et al.* (1995), we analyzed the *Est6* promoter region of flies isolated from two other Australian populations and one population each from Asia, South America, North America, and Africa. Flies collected from Cairns (Queensland, Australia) and Tahbilk (Victoria, Australia) in May 1995 were gifts from Ary Hoffmann (La Trobe University) and Philip Batterham (University of Melbourne), respectively. Individual wild-caught males or males deriving from females inseminated in the field were crossed to either the TM3/TM6 third chromosome balancer stock or a stock carrying a deficiency for *Est6* [*Df(3L)vin7,ru*^{1}*h*^{1}*gl*^{2}*e*^{4}*ca*^{1}/TM3, breakpoints 68C8-11, 69B4-5; Bloomington Stock Center]. Individual progeny were then backcrossed to either the balancer or the deficiency stock and the backcross progeny were sib mated to yield flies that were either homozygous for a wild third chromosome or hemizygous over the deficiency. Flies were frozen at −80° for analysis. Flies from Beijing, China (Asia); Atacame, Ecuador (South America); Maryland (North America), and Sengwa, Zimbabwe (Africa) were isolated using a third chromosome balancer and were homozygous for wild third chromosomes (described in Begun and Aquadro 1993, 1994, 1995; Kindahl 1994).

DNA was extracted from individual flies using the method of Gloor *et al.* (1993). A total of 2 μl DNA was then used as template to amplify the *Est6* 5′ promoter region in duplicate 50-μl reactions using the PCR primers and conditions described in Odgers *et al.* (1995). Prior to PCR, a 5′ phosphate was added to one primer (−1007/−986) using polynucleotide kinase (Pharmacia). Duplicate PCR products were pooled and purified into 50 μl water using Centricon 100 filter units (Amicon, Beverly, MA; no. 4212).

The purified PCR products were then separately treated with two restriction endonucleases, *Rsa*I and *Ssp*I, which were diagnostic for two of the nucleotide polymorphisms that distinguish the P1 and P7 promoter haplotypes. The nucleotide at −531 bp in the P1 haplotype disrupts a *Rsa*I site that is present in the P7 haplotype. The nucleotide at −595 bp in the P7 haplotype disrupts a *Ssp*I site that is present in the P1 haplotype. Following digestion, DNA fragments were separated and visualized on a 2% agarose gel and scored for the presence or absence of each restriction site. The 35-bp insertion polymorphism found at −757 bp in the P1 haplotype could also be scored due to the size difference.

DNA that was to be sequenced was treated with λ exonuclease (Bethesda Research Laboratories, Gaithersburg, MD) following the method of Higuchi and Ochman (1989) and then purified once again into 50 μl water using the Centricon filter units. The resulting single-stranded DNA was then sequenced using the Sequenase kit (United States Biochemical, Cleveland). Primers used for sequencing were 5′CACCAGTCTTTG CTAGC3′ (−878 to −861), 5′GCGTGCACCTGCAAGTGC3′ (−675 to −658), 5′GGTTTCGTTTTAAAGTGC3′ (−456 to −439), 5′GCTATCGTTTTAATTCGC3′ (−238 to −221), and one of the PCR primers (−1007/−986; Odgers *et al.* 1995). The sequence data have been submitted to the GenBank Data Library (accession nos. AY058798–AY058828). DnaSP Version 3.50 (Rozas and Rozas 1999) was used to estimate nucleotide diversity (π) and F_{st} for each population and to carry out several population genetic tests including Tajima's *D* and Fu and Li's *D* and *F* (where appropriate assuming no recombination). Four sequences from *D. simulans* (Karotam *et al.* 1993; accession nos. L10670 and L34263–L34265) were used as outgroups where required.

RESULTS

Transformation experiments: EST6-specific activity was determined in whole males and females from 26 and 25 lines, respectively, for the *P7′* construct and in both sexes from 30 lines for the *P1* construct (Figure 2). Differences between the *P1* and *P7′* constructs in EST6-specific activity depended on gender ($F_{1,107} = 4.2$; $P < 0.05$) and were significant in males ($F_{1,107} = 11.3$; $P < 0.01$) but not females ($F_{1,107} = 0.1$; $P > 0.50$). EST6-specific activity was 20% lower in males with the *P1* construct (1.26 ± 0.12) than in those bearing *P7′* (1.58 ± 0.10; Figure 2) or 18% lower if values of EST6 activity per fly were used. Ten subsequently assayed *P7* lines that lacked the nucleotide difference at −652 bp yielded essentially the same specific activities in males as the *P7′* lines ($t_{34} = 0.17$; $P > 0.05$). These results indicate that the 14 nucleotide differences that characterize the P1 and P7/P7′ groups of *Est6* promoter haplotypes cause an ∼20% difference in EST6-specific activity in males but not in females.

Two experiments were carried out to measure the contribution of ejaculatory duct EST6 activity levels to the whole body differences between the *P1* and *P7/P7′* constructs. In the first EST6 activity was measured in ejaculatory ducts dissected from each of two lines for the *P1* and *P7′* constructs. The two lines for each construct had whole body activities within 10% of the respective means for all lines for that construct. Activity values in dissected ejaculatory ducts varied significantly between the *P1* and *P7′* constructs ($F_{1,2} = 219.8$; $P < 0.01$) and the effect was in the same direction as the whole male effect, but much larger; ejaculatory ducts taken from *P1* lines yielded less than one-half the EST6 activity of those taken from *P7′* lines (7.6 ± 0.4 and 26.9 ± 1.6 nmol β-naphthol produced/fly/30 min, respectively). There was no difference between the constructs due to the nonejaculatory duct component of whole male activity (40.4 ± 1.5 and 40.5 ± 1.0 nmol β-naphthol produced/fly/30 min for *P1* and *P7′*, respectively).

In the second experiment on ejaculatory duct activities EST6 activity was measured in ducts dissected from all 11 *P7* lines, plus 5 each of the original *P1* and *P7′* lines. All 10 lines in the *P1* and *P7′* samples were selected

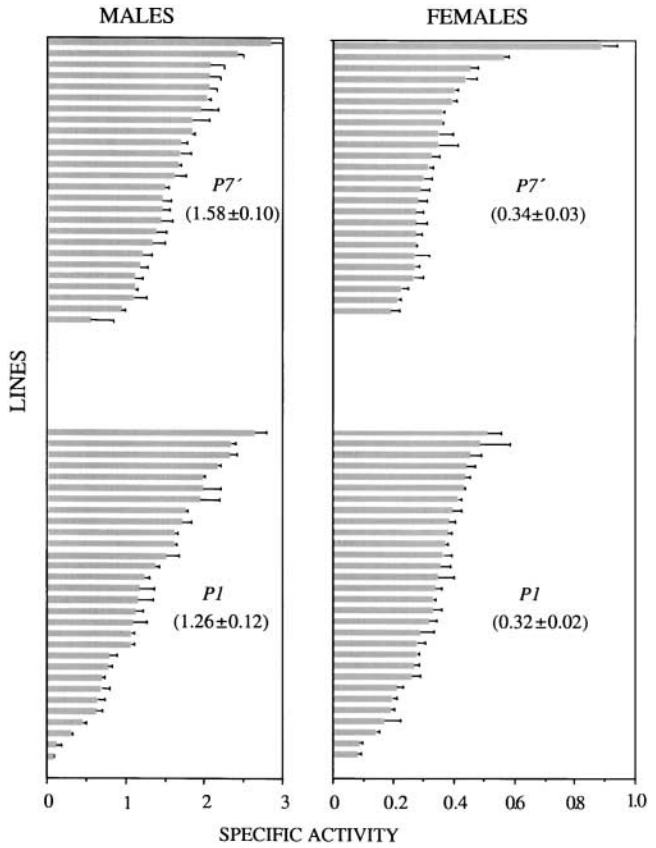

FIGURE 2.—Mean EST6-specific activity and standard error for adult males and females from each line for constructs *P1* and *P7′*. EST6-specific activity (micromoles β-naphthol produced per milligram protein per 30 min at 28°) was assayed in virgins selected from triplicate cultures and aged for 3 days at 25°. Construct means and standard errors are given in parentheses. Male and female data are presented using different scales. Note that the variance in EST6-specific activity among lines was 33% of the construct means in females and ranged from 21% for *P7* to 41% for *P1* in males. Similar variances have been found for *Adh* (15% for *Adh-F*, 43% for *Adh-S*) and *Xdh* (29%; LAURIE-AHLBERG and STAM 1987).

to be within 15% of the respective overall means for all lines for whole body activities but otherwise they were chosen at random. Two of the *P7′* lines and 1 of the *P1* lines had also been assayed in the first experiment. Ejaculatory duct EST6 activities (nanomoles β-naphthol produced per fly per 30 min) in this experiment were 31 ± 6 for *P7*, in good agreement with the 34 ± 5 for *P7′* ($P > 0.05$) and again greater than the 17 ± 5 for *P1* ($P < 0.05$). Across the two experiments we therefore conclude that the whole male EST6 activity differences between the P1 and P7 promoter classes are due to about a twofold difference in ejaculatory duct activity and that this difference is due to some number of the 14 nucleotide differences between P1 and P7.

RFLP analyses: To test whether the P1 and P7 promoter haplotypes were present in populations other than the Australian Coffs Harbour population (ODGERS *et al.* 1995) originally scored, we first analyzed the *Est6* promoter region by RFLP analysis in a total of 122 chromosomes isolated from two other Australian populations and one population each from Asia, South America, and Africa (Table 1; Figure 3). In addition, we reinterpreted the original RFLP survey of 42 chromosomes from Coffs Harbour, Australia (GAME and OAKESHOTT 1990) and a later RFLP study of 87 chromosomes from Maryland (KINDAHL 1994). All these population samples were randomly selected in respect of EST6 allozyme status.

In our RFLP survey, we used two restriction endonucleases to score the indel polymorphism, Ins(−757), and two nucleotide polymorphisms, *Ssp*I(−595) and *Rsa*I(−531), which were diagnostic for the P1 and P7 *Est6* promoter haplotypes (Figure 1; Table 1). However, in the sample from Zimbabwe there was so much size polymorphism that Ins(−757) was difficult to score (Figure 3). For this reason we discuss the Zimbabwe sample in a separate section below. In the reinterpreted RFLP surveys of the Coffs Harbour and Maryland populations Ins(−757) and *Rsa*I(−531) were scored but *Ssp*I(−595) was not. The RFLP haplotypes diagnostic for P1 in the non-African samples were + + − or + ns −, for Ins(−757), *Ssp*I(−595), and *Rsa*I(−531), respectively (where + is presence of the restriction site or indel and ns is not scored). The corresponding haplotypes for P7 were − − + or − ns +.

The RFLP analyses show P1 and P7 to be present in each of the three Australian, one Chinese, and two American populations (Table 1). Furthermore, in all but one case (Ecuador, P7 at 7%), these haplotypes occur at high frequencies, ranging from 25 to 80% for P1 and from 20 to 40% for P7. We confirmed that these RFLP-based haplotypes were in fact diagnostic for the same set of 14 polymorphisms that distinguished the P1 and P7 haplotypes in our original survey (ODGERS *et al.* 1995) by sequencing through the 325-bp polymorphic region from at least one putative P1 and one putative P7 chromosome from each population (except Zimbabwe, data not shown).

The non-African populations also yielded five RFLP-based haplotypes that were not classed as P1 or P7 (Table 1). One or more of these five haplotypes were present in all samples, except Beijing, at combined frequencies of between 10 and 40%. Three haplotypes were "intermediate" between P1 and P7 in that they had different combinations of the diagnostic RFLPs, while two others had a smaller insertion instead of Ins(−757) (denoted * in Table 1; sequence analysis indicated that this insertion was also located at −757 bp; it is insertion 10 in Figure 3). In Maryland and Ecuador, one intermediate haplotype (− ns − and − + −, respectively) was present at a similar frequency to P7, but in all other non-African cases, individual non-P1/P7 haplotypes were less common than either P1 or P7 in all populations.

Three of the non-P1/P7 haplotypes from Cairns (−

TABLE 1

Frequencies of *Est6* promoter haplotypes derived from RFLP analysis of seven populations

	P1	P7	Others
Cairns, Australia	3 (+ + −)	4 (− − +)	2 (− + +) 1 (− + −) 2 (* − +)[a]
Tahbilk, Australia	26 (+ + −)	15 (− − +)	1 (+ + +) 1 (− + +) 1 (− + −) 1 (* − +) 1 (* + −)
Coffs Harbour, Australia[b]	29 (+ ns −)	8 (− ns +)	3 (+ ns +) 2 (− ns −)
Maryland, USA[b]	40 (+ ns −)	18 (− ns +)	9 (+ ns +) 20 (− ns −)
Beijing, China	18 (+ + −)	12 (− − +)	—
Atacame, Ecuador	12 (+ + −)	1 (− − +)	2 (− + −)
Sengwa, Zimbabwe[c]	5 (ns + −)	6 (ns − +)	8 (ns + +)

+, presence of the site or indel; −, absence of the site or indel; ns, not scored. The sites scored were Ins(−757), *Ssp*I(−595), and *Rsa*I(−531), scores for which are shown in that order in the parentheses in the table.

[a] * denotes an insertion of smaller size, which subsequent sequence analysis (see text) showed to be at the same site as Ins(−757).

[b] The Coffs Harbour and Maryland data were taken from earlier RFLP surveys of Game and Oakeshott (1990) and Kindahl (1994), respectively, in which Ins(−757) and *Rsa*I(−531) were scored but not *Ssp*I(−595). Note the sequence difference that Game and Oakeshott (1990) attributed to *Taq*I(−630) is likely to be due to Ins(−757) (Odgers *et al.* 1995).

[c] Ins(−757) was not scored in the Zimbabwe sample because there were too many minor size differences in RFLP fragments in that population.

+ +, − + −, * − +) and one from Ecuador (− + −) were also sequenced across the 325-bp polymorphic region (Figures 3 and 4). One sequence from Cairns (− + −) and the Ecuador sequence (− + −) match those of the common intermediate from Maryland (− ns −; see below). This RFLP haplotype was also scored in Coffs Harbour but none from this population were sequenced by Odgers *et al.* (1995). The two other sequences from Cairns (− + + and * − +) are more similar to those found in Zimbabwe than to any found elsewhere to date (Figures 3 and 4 and see below) although both RFLP-based haplotypes were also scored in the third Australian population, Tahbilk.

Sequence analyses: We also sequenced 974 bp 5′ of the *Est6* initiation codon in 15 of the chromosomes scored above from Maryland and 12 of the chromosomes scored above from Zimbabwe (Figure 3). As with the RFLP survey above, the chromosomes sequenced were selected at random in respect to EST6 allozyme or other phenotypic information. There were strong similarities in the sequence data from Maryland and our previous sequence data for 17 chromosomes from Coffs Harbour, Australia (Odgers *et al.* 1995). We found 28 polymorphic nucleotide sites in Maryland, 19 of which were also seen in Coffs Harbour. Only 4 polymorphisms seen in Coffs Harbour were not found in Maryland. The two indels (−819 and −747 bp) found previously in Coffs Harbour were also found in Maryland and are diagnostic for haplotype P1. Two other indels were found in Maryland but not in Coffs Harbour, a 10-bp insertion at −917 bp and a 1-bp insertion at −683 bp. The two samples also showed similar distributions of polymorphism along the promoter, with 21 of the 23 sites in Coffs Harbour and all 28 sites in Maryland occurring in the region 5′ of −400 bp.

Nine sequence-based haplotypes were found in Maryland compared with 10 found in Coffs Harbour. Four haplotypes were shared between both populations, and these comprised P1, P7, and 2 haplotypes that were derived simply from P1 and P7 by additional nucleotide differences (P1- and P7-like). Of the 11 sequence-based haplotypes that were not shared between the two populations, 1 from Maryland and 2 from Coffs Harbour were P1-like and 2 from Coffs Harbour were P7-like (Figure 4). The remaining 6 sequence-based haplotypes that were not shared between the two samples were intermediate in that they had different combinations of the nucleotides that were diagnostic for P1 and P7.

As previously noted, one RFLP-based intermediate haplotype (− ns −) was particularly frequent in Maryland (20/87). This same haplotype or an equivalent (− + −) was found in all other non-African populations except Beijing but at lower frequencies (Table 1). Sequencing of five chromosomes with this RFLP haplotype

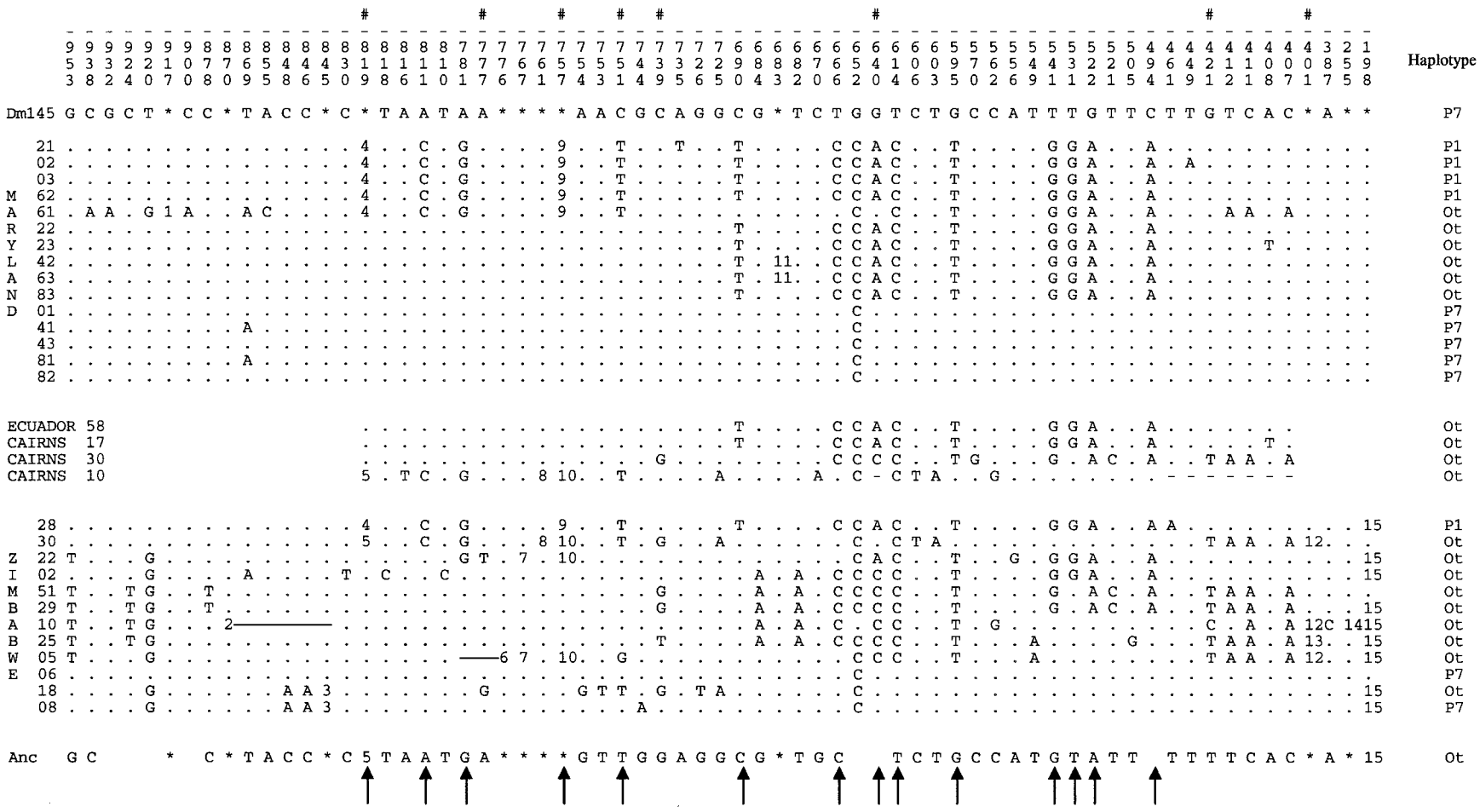

FIGURE 3.—*Est6* promoter nucleotide variation from several populations. The nucleotide numbering given at the top follows KAROTAM *et al.* (1993). # indicates that more than two alternative states are at this site. * denotes the absence of an indel. Indels are numbered, descriptions given below. Lines adjacent to numbers show the spans of indels that cover additional polymorphic sites. Missing data are indicated by a hyphen. "Anc" is the inferred ancestral haplotype representing the consensus between the *D. melanogaster*, *D. simulans*, and *D. mauritiana* sequences at each site polymorphic within the *D. melanogaster* lineage (KAROTAM *et al.* 1993). In six cases where the *D. simulans* and *D. mauritiana* sequences did not match any of the *D. melanogaster* sequences the ancestral state could not be inferred so the space was left blank. The right-hand column classifies the sequences as P1, P7 (including P7′), or other (Ot) according to the 14 diagnostic sites shown in Figure 1 and also marked † at the bottom. 1, 10-bp insertion ACTTTAAAGG; 2, 36-bp deletion between −870 and −835; 3, 3-bp insertion TTT; 4, 5-bp insertion CTTTT; 5, 5-bp insertion CTTTC; 6, 10-bp deletion between −776 and −785; 7, 1-bp deletion; 8, 1-bp deletion; 9, 35-bp insertion AGTAATTGTAATAATAATATAATAGTAATTTTAAT; 10, 16-bp insertion AGTAATAATTAA TAAT; 11, 1-bp insertion A; 12, 1-bp deletion A (could be between −401 and −410 in a homopolymer repeat); 13, 2-bp deletion A (could be between −401 and −410 in a homopolymer repeat); 14, 9-bp deletion between −255 and −249; 15, 3-bp deletion between −200 and −198.

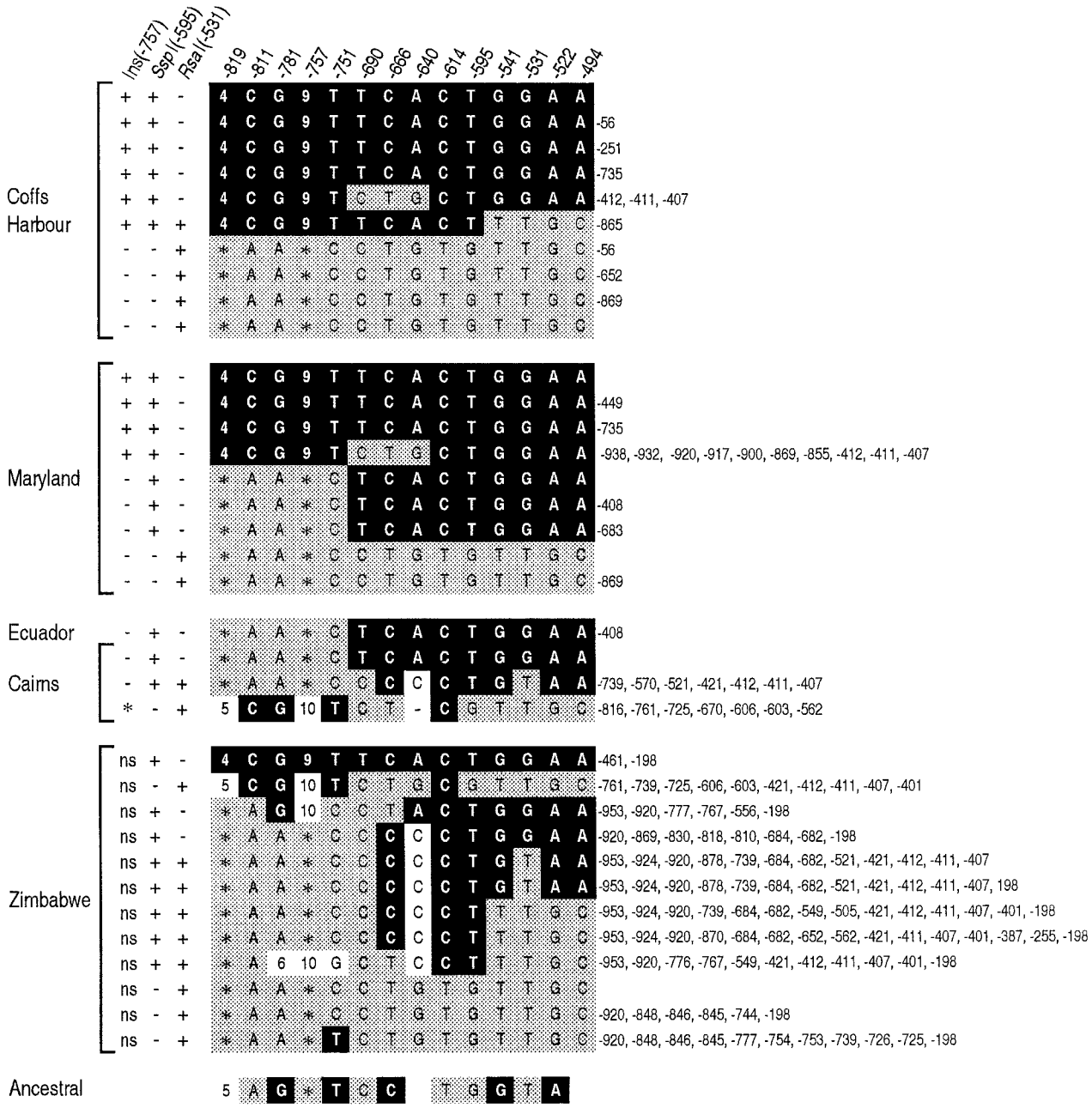

FIGURE 4.—*Est6* promoter haplotypes from several populations in relation to P1 and P7. Nucleotides diagnostic for the P1 and P7 haplotypes are in black and gray boxes, respectively. Nucleotides that are not boxed did not match either P1 or P7 at that site. Missing data are indicated with a hyphen. Nomenclature for indels and the "ancestral" haplotype is given in Figure 3. Sites for nucleotide polymorphisms in addition to those that characterize P1 and P7 are shown to the right of each haplotype. RFLP-based haplotypes (Table 1) are shown to the left. The data for Coffs Harbour are taken from ODGERS *et al.* (1995).

from Maryland revealed three sequence-based haplotypes that derived simply from one another by additional nucleotide differences (Figures 3 and 4). The consensus sequence of these haplotypes matched P7 5′ of −690 bp (five polymorphisms) and P1 3′ of −690 bp (nine polymorphisms). Sequence analysis of chromosomes from Maryland revealed a further intermediate haplotype (line 61 in Figure 3), which was indistinguishable from P1 using RFLPs but its sequence indicated that it matched P7 at three sites between −690 and −640 bp (similar to intermediate P5 in ODGERS *et al.* 1995; Figure 4). These intermediate haplotypes could be derived from P1 and P7 through simple recombination or gene conversion since the nucleotide sequences at the 14 diagnostic sites show blocks of similarity to one or the other of P1 or P7.

The higher frequency of intermediate haplotypes in Maryland is reflected in far less linkage disequilibrium

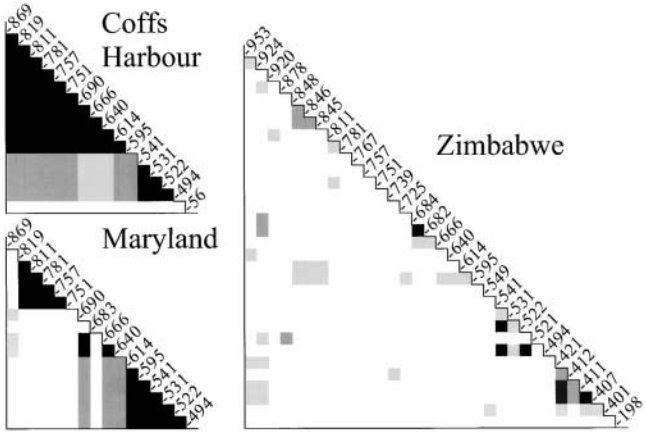

Figure 5.—Linkage disequilibrium between common nucleotide polymorphisms (frequencies 10–90%) in the *Est6* promoter in samples from Coffs Harbour, Australia; Maryland, North America; and Zimbabwe, Africa. Pairwise associations were assessed by Fisher's exact test. Nonsignificant associations are unshaded and significant associations are shown successively darker (light gray, $P < 0.05$; dark gray, $P < 0.01$; black, $P < 0.001$). For the Zimbabwe sample, where there were three alternatives for any one site, the least common alternative was excluded and the two more frequent sites were included. The overall pattern of disequilibrium changed only slightly when these sites were omitted from the analysis (not shown). The data for Coffs Harbour are taken from Odgers *et al.* (1995).

overall (49 out of 120, or 41% of comparisons were significant at the 5% level; Figure 5) but particularly between polymorphisms either side of −690 bp, compared with the Coffs Harbour sequences (105 out of 120, or 88% of comparisons were significant at the 5% level; Figure 5). However, the polymorphisms that distinguish P1 from P7 remain in strong disequilibrium across the regions −819 to −751 bp and −690 to −494 bp in both populations.

Zimbabwe results: The RFLP data from Zimbabwe (Table 1) showed what appeared to be P1 and P7 haplotypes (ns + − and ns − +, respectively) at appreciable frequencies (26 and 32% respectively). However, unlike other populations, the most common RFLP haplotype in Zimbabwe was an intermediate (ns + +, at 42%). Given also the evidence for additional variation from the numerous minor length differences in the RFLP data, we gave little weight to the RFLP results for this population and instead sequenced the region in most of the Zimbabwe lines.

Consistent with published sequence data for at least some other loci from Zimbabwe (Begun and Aquadro 1993, 1994, 1995; Eanes *et al.* 1996; but see also Andolfatto *et al.* 1999; Andolfatto and Kreitman 2000; Andolfatto 2001), we see differences between the Zimbabwe and non-Zimbabwe samples in the nature, level, and disequilibrium structure of the polymorphism. In the Zimbabwe sample we found 55 polymorphic sites across a 960-bp region compared to only 28 and 23 for the same region in the similarly sized samples from Maryland and Coffs Harbour, respectively (Figure 4). Nucleotide diversity (π) in Zimbabwe was also about twice that in the other two populations (Table 2). Twenty-eight of the Zimbabwe polymorphisms were found elsewhere and 27 were not. In contrast, only 15 of the 43 polymorphisms recorded elsewhere were not found in Zimbabwe and these polymorphisms tended to be far less common. F_{st} values were higher in comparisons of Zimbabwe with either Coffs Harbour (0.2000) or Maryland (0.1850) than in a comparison between the latter two populations (0.0002).

A similar pattern is seen in the indel data. Twelve polymorphic indels are found in Zimbabwe (at 10 sites) compared to a total of 7 polymorphic indels found elsewhere (at 5 sites). Five of these indels (at 3 sites) are shared between Zimbabwe and non-Zimbabwe populations. The polymorphisms at 8 positions in the Zimbabwe sequences consisted of three alternative states and four of these (at −819, −757, −751, and −640 bp) include polymorphisms that are diagnostic for the P1/P7 haplotype difference.

One of the few features of the Zimbabwe data that were consistent with the other populations was the distribution of polymorphisms along the *Est6* promoter region. In Zimbabwe, as elsewhere, most (52/55) of the polymorphisms occur in the region 5′ of −400 bp. This pattern is probably common across the *melanogaster* species subgroup: Interspecific comparisons across three species in the subgroup all show much greater variation 5′ compared to 3′ of −400 bp (Karotam *et al.* 1993; Oakeshott *et al.* 2001).

The 12 sequences from Zimbabwe all represented different haplotypes. No P1 haplotype was found, although a single P1-like haplotype (which contained 2 additional nucleotide polymorphisms) was sampled. Only one P7 and one P7-like (6 additional nucleotide polymorphisms) haplotypes were found. The remaining nine haplotypes have not been observed elsewhere to date and do not fit obviously into either the P1 or the P7 classifications, although 11 of the 14 polymorphisms that distinguish P1 from P7 are also highly polymorphic in Zimbabwe. On average, each non-P1/P7 haplotype from Zimbabwe contains 11 nucleotide polymorphisms in addition to those that distinguish P1 from P7 (compared to 1.4 and 1.7 from Coffs Harbour and Maryland, respectively). Furthermore, in contrast to the Coffs Harbour and Maryland samples, eight out of nine non-P1/P7 haplotypes from Zimbabwe cannot be derived from P1 and P7 simply through recombination or gene conversion. This is reflected in the far lesser degree of disequilibrium seen among Zimbabwe sequences (49 out of 496, or 10% of comparisons were significant at the 5% level; Figure 5) compared with non-Zimbabwe sequences (41 and 88% of comparisons were significant at the 5% level in Maryland and Coffs Harbour, respectively; Figure 5). Also, the few significant cases of disequi-

TABLE 2

Estimates of nucleotide diversity (π) and tests of departures from neutrality in the sequence data for −819 to −494 bp from Zimbabwe, Coffs Harbour, and Maryland

	Zimbabwe (12)[a]	Coffs Harbour (17)	Maryland (15)	Coffs Harbour and Maryland (32)
π[b]	0.0157 (0.0011)	0.0077 (0.0008)	0.0079 (0.0013)	0.0078 (0.0007)
Fu's *Fs*	−2.97	3.06	3.65	4.29
Strobeck's *S*	0.99	0.13	0.09	0.04
Fu and Li's *D*[c]	−0.14	1.22	1.25	1.62**
Fu and Li's *F*[c]	−0.24	1.88**	1.75*	2.44***
Tajima's *D*	−0.40	2.01**	1.93*	2.53**

*$0.10 > P > 0.05$, **$P < 0.05$, ***$P < 0.02$. Analyses were carried out using Version 3.50 of DnaSP (Rozas and Rozas 1999).

[a] Numbers in parentheses are numbers of sequences compared.

[b] Calculated using the Jukes and Cantor correction with standard deviations in parentheses.

[c] Four *D. simulans* sequences were used as outgroups (see MATERIALS AND METHODS).

librium in the Zimbabwe data generally involve physically close sites (close to the diagonal in Figure 5), whereas in the Maryland and in particular Coffs Harbour populations in which the P1 and P7 haplotypes predominate there are many more cases of disequilibrium between physically distant sites across the region.

Intriguingly, two sequences from Cairns, Australia, represented by lines 10 and 30 are similarly divergent from the Coffs Harbour and Maryland sequences (Figure 4) and show marked resemblance to several sequences observed in Zimbabwe. Like the Zimbabwe sequences, these two Cairns sequences are divergent both in terms of the number of additional polymorphisms and the fact that they cannot be derived from P1 and P7 by simple recombination or gene conversion. Line 30 from Cairns is identical to lines 51 and 29 from Zimbabwe at the 14 diagnostic sites. Similarly, line 10 from Cairns is the same as line 30 from Zimbabwe at 12 of the 14 diagnostic sites while the remaining 2 diagnostic sites are the same as several other haplotypes from Zimbabwe.

Tests for selection: Notwithstanding the two Cairns sequences above, the differences already noted between Zimbabwe and the other populations also lead to very different results in various tests for departures from neutrality in the sequence data for the −819- to −494-bp region (Table 2). Statistics assessing haplotype structure (Fu's *Fs*, Strobeck's *S*) confirm a pronounced deficit of haplotypes in the Coffs Harbour and Maryland samples compared to the Zimbabwe sample. Fu and Li's *D* and *F* tests and Tajima's *D* also reveal departures from neutrality in Coffs Harbour and Maryland but not Zimbabwe. The directions of the departures are consistent with an excess of high-frequency variants (locked up in the high-frequency haplotypes) in the two non-African samples. The same tests carried out on the data for the sequenced regions up- and downstream of −819 to −494 reveal no significant departures from neutrality (data not shown), albeit the smaller data sets for these regions reduce the sensitivity of the tests.

DISCUSSION

Analyses of the EST6 activities in a total of 66 independent lines transformed with either the *P1* or the *P7/P7′* constructs revealed an ~20% difference between the constructs in whole male-specific activity, which was not apparent in whole females: Males with the *P1* construct had 20% less EST6-specific activity (or 18% less activity per fly) than those with *P7* or *P7′*. We have shown that the 14 nucleotides that characterized the P1 and P7 groups of *Est6* promoter haplotypes are sufficient to cause this difference in male EST6-specific activity.

The size, direction, and sex specificity of this effect are consistent with previous observations in nontransformed, field-trapped flies. Odgers *et al.* (1995) showed that isoallelic lines from the P1 haplotype group have 15–20% less male EST6 activity than those from the P7 haplotype group. It is likely, therefore, that the 14 nucleotide differences that distinguish the P1 and P7 haplotype groups cause similar effects on male EST6 activity in both transformed and nontransformed flies. It also seems likely that the basis of the effect lies in the control of transcription rather than in some property of the transcript or its protein product because all 14 polymorphic nucleotides occur in a 325-bp region between −819 and −494 bp that is not transcribed.

Consistent with its male specificity, the whole male *P1/P7* activity difference is due largely to activity differences that occur in a male-specific tissue. Males with the *P1* construct had only about half as much EST6 activity in the ejaculatory duct as those with *P7* or *P7′*, and this difference was sufficient to explain the whole male effect. Therefore, the polymorphism(s) causing this tissue-specific effect on the level of EST6 activity probably lies in a region(s) that regulates *Est6* expres-

sion in the male ejaculatory duct. Germline transformation experiments with constructs bearing internal or sequential deletions in *Est6* 5′ flanking DNA have localized sequences required for qualitative and quantitative expression in the male ejaculatory duct to between −844 to −613 bp and −283 to −613 bp, respectively (Healy *et al.* 1996). This places all 14 nucleotide differences that differentiate P1 from P7 in regions that have been shown to regulate either qualitative or quantitative expression of *Est6* in the male ejaculatory duct (9 polymorphisms lie in the region −844 to −613 bp and 5 in the region −283 to −613 bp). Further experimentation is needed to uncover the precise molecular cause(s) of the *P1/P7* activity difference, but the congruence of the P1/P7 sequence differences with the promoter regions controlling ejaculatory duct expression clearly supports a causal relationship between some number of the P1/P7 nucleotide differences and the distinct ejaculatory duct expression patterns reported here.

It may be that only 1 of the 14 nucleotide differences that differentiate P1 and P7 causes the expression difference we have described but it is also possible that several of the polymorphisms are causally involved. At least four *Adh* polymorphisms contribute to ADH activity differences in *D. melanogaster* and there is also evidence of epistatic interactions between some of these polymorphisms (Choudhary and Laurie 1991; Laurie and Stam 1994; Stam and Laurie 1996). Functional dissections of other eukaryote promoters reveal complexes of interacting regulatory modules and imply high levels of coadaptation between, but particularly within, particular modules (Yuh *et al.* 1998; Fickett and Wasserman 2000; Ludwig *et al.* 2000). Notably, these findings extend to another Drosophila gene, *Gld*, whose expression is a complex mix of non-sex-specific and reproductive tissue activities (Keplinger *et al.* 2001). A transformation approach involving fine-scale promoter deletions and chimeric *P1/P7* promoters may be needed to elucidate the causal basis for the P1/P7 difference.

The difference between P1 and P7 in the level of EST6 activity found in the male ejaculatory duct represents a potential target upon which selection might act. Our tests on the available sequence data for the region indeed suggest significant departures from neutrality for the Coffs Harbour and Maryland populations where P1 and P7 are common. Several other lines of evidence further suggest that the presence and level of EST6 in the male ejaculatory duct affects reproductive fitness. First, we note that no EST6$^{\emptyset}$ mutants have been found in wild populations on three continents despite extensive searches (Cochrane and Richmond 1979; Voelker *et al.* 1980; Langley *et al.* 1981; Labate *et al.* 1989). Second, a laboratory null strain (Johnson *et al.* 1966; Sheehan *et al.* 1979) shows some differences in reproductive function when compared to EST6-active individuals (reviewed in Richmond *et al.* 1990). In particular, females mated to males with EST6 activity are less likely to remate in the short term but (at certain temperatures) are more likely to remate in the long term compared with those mated to EST6$^{\emptyset}$ males (Gilbert *et al.* 1981; Gilbert and Richmond 1982; Scott 1986). These effects seem likely to be due to EST6 that is transferred during copulation from the male ejaculatory duct to the female's reproductive organs and thence to her hemolymph (Richmond *et al.* 1980; Meikle *et al.* 1990).

It might be expected that the common P1 haplotype would have reproductive effects similar to those conferred by the *Est6*$^{\emptyset}$ gene since it lowers EST6 activity in the ejaculatory duct. In support of this proposal, experiments conducted by Saad *et al.* (1994) suggest that natural variation in the level of EST6 in the male ejaculatory duct affects reproductive fitness. Females mated to males with low EST6 activity are more likely to remate in the short term than are those mated to males with high EST6 activity. We have reanalyzed a subset of the data in Saad *et al.* (1994), which was composed of lines that can be identified as either P1 or P7 by RFLP or sequence analysis (Game and Oakeshott 1989; Odgers *et al.* 1995). The negative relationship between females' remating frequencies and their first mates' EST6 activities could be explained, at least in part, by the P1/P7 differences, although the effects did not reach statistical significance for the small subset of lines available for analysis (12 for P1 and only 2–3 for P7; analysis not shown).

Using a combination of RFLP analysis and DNA sequencing, we have also shown that the P1 and P7 promoter haplotypes are geographically widespread. P1 and P7 made up the majority of haplotypes found in the six non-African populations analyzed, with combined frequencies ranging from 60 to 90%. The average frequencies of P1 and P7 across these six populations are 55 and 26%, respectively. Interestingly, the low-activity P1 variant is more common than the high-activity P7 variant.

However, P1 and P7 (or simple derivatives thereof) are not the only haplotypes we found. In the non-African populations, most of the non-P1/P7 haplotypes that we sequenced could have originated through simple recombination or gene conversion between P1 and P7 sequences, hence the term intermediate haplotype. In Zimbabwe, however, *Est6* promoter polymorphism differs in nature, level, and disequilibrium structure from that observed elsewhere. In the Zimbabwe sample, there is twice as much polymorphism and far less disequilibrium than we observed in the non-African samples. Although we observed single P7, P7-like, and P1-like haplotypes in Zimbabwe, the majority of haplotypes were not observed elsewhere and could not be derived simply from P1 and P7 sequences through recombination or gene conversion (Figure 4). This result has intriguing parallels with those for some other loci that also show higher levels of polymorphism and less disequilibrium in Zimbabwe populations than elsewhere. This appears

to be the case for several loci on the X chromosome (Begun and Aquadro 1993, 1994, 1995; Eanes *et al.* 1996), albeit it is not so evident for a number of autosomal loci (Andolfatto *et al.* 1999; Andolfatto and Kreitman 2000; Andolfatto 2001).

One factor noted by others in this respect is that Zimbabwe lies within the ancestral, east African range of the species, whereas almost all the other populations analyzed in these surveys have been established much more recently (David and Capy 1988; Lachaise *et al.* 1988; Begun and Aquadro 1993; Bénassi *et al.* 1993; Bénassi and Veuille 1995; Schlötterer *et al.* 1997). Some authors have suggested that populations elsewhere may have lost variation through founder effects at establishment (an effect that in general might be more pronounced for X-linked than for autosomal genes; Begun and Aquadro 1995; Eanes *et al.* 1996; Irvin *et al.* 1998). On the other hand there are also a few cases of (autosomal) genes for which tropical populations outside of east Africa show levels of polymorphism and disequilibrium comparable to east African data (Aguadé 1999; Andolfatto and Kreitman 2000; Andolfatto 2001). The respective roles of demographic and latitudinally dependent adaptive processes in the geographic differentiation of *D. melanogaster* populations are thus unclear and likely to vary across loci (Andolfatto 2001).

Both historical factors and latitudinal differentiation may be relevant to the interpretation of our data on *Est6* promoter polymorphism. We have previously argued that the high level of difference of P1 and P7 both from each other and from an "ancestral" haplotype inferred by comparisons between *D. melanogaster*, *D. simulans*, and *D. mauritiana* was best explained by proposing that P1 and P7 arose independently, in geographic or genetic isolation from one another (Odgers *et al.* 1995). We can now reexamine this proposition with the benefit of several further pieces of information.

First, our data here show that P1 and P7 or closely related haplotypes co-occur widely throughout the species range, including the ancestral range where there are also many other more distantly related haplotypes. This suggests that both P1 and P7 could have originated in a diverse pool of ancestral haplotypes. Second, the data of Hasson and Eanes (1996) and Balakirev *et al.* (1999) show negligible disequilibrium between *Est6* and either the common cosmopolitan chromosome inversion *In(3L)P* or the closely linked *Sod* locus. This also argues against the origin of P1 and P7 in genetic isolation from one another. Third, such traces of disequilibrium as Balakirev *et al.* (1999) could find between the *Sod* and *Est6* coding regions were of a structure that was inconsistent with explanations of divergence during geographic isolation with subsequent admixture. Finally, we note that two *Est6* promoter alleles from wet tropical northern Australia (Cairns) showed haplotypes that are more similar to those found in Zimbabwe than to any found elsewhere to date [albeit a third tropical population (Ecuador, South America) showed no such haplotypes].

All these data are consistent with a scenario in which P1 and P7 arose in a pool of variously related haplotypes in an ancestral east African population and that much of this haplotype diversity was carried into descendant populations on other continents, but there has then been a presumptively selective increase in P1 in particular but also P7 at the expense of other haplotypes in many of the nontropical descendant populations. Our findings of nonneutral patterns of variation in the P1/P7 regions in Coffs Harbour and Maryland but not Zimbabwe certainly support this conclusion. Larger samples of *Est6* promoter sequences from transects leading away from the (wet) tropics in Africa and other continents are now needed to explore the proposition more fully. It would clearly also be of great interest to determine ejaculatory duct EST6 activity levels for some of the Zimbabwe promoter haplotypes.

Climatic associations are of further interest because a widespread allozyme polymorphism for EST6 shows latitudinal clines (Oakeshott *et al.* 1981). Odgers *et al.* (1995) found some linkage disequilibrium between RFLPs diagnostic for *Est6* promoter and coding region polymorphisms in the Coffs Harbour sample. The direction of the disequilibrium was such that the allozyme cline might also be associated with a promoter cline, P1 becoming more common away from the equator. However the disequilibrium was weak ($D' = -0.38$, $P < 0.05$) and there is no clear trend in our data for P1 frequency to increase away from the equator. A larger sample of populations from a defined latitudinal transect is needed to test for such a latitudinal relationship more rigorously, but it appears that the patterns of variation observed in the two regions could, at least partly, be shaped by different forces.

We thank Geoff Clarke for statistical assistance; Kerrie Medvecsky for technical assistance with flies collected in Australia; and members of the Oakeshott, Healy, and Aquadro laboratories for contributing valuable comments and discussion. Two anonymous referees suggested several significant improvements to the manuscript. This work was supported by an Australian Grains Research and Development Corporation postgraduate fellowship to W.A.O. and a grant from the U.S. National Institutes of Health to C.F.A.

LITERATURE CITED

Aguadé, M., 1999 Positive selection drives the evolution of the Acp-29AB accessory gland protein in Drosophila. Genetics **152:** 543–551.

Andolfatto, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **18:** 279–290.

Andolfatto, P., and M. Kreitman, 2000 Molecular variation at the *In(2L)t* proximal breakpoint site in natural populations of *Drosophila melanogaster* and *D. simulans*. Genetics **154:** 1681–1691.

Andolfatto, P., J. D. Wall and M. Kreitman, 1999 Unusual haplotype structure of the proximal breakpoint of *In(2L)t* in a natural population of *Drosophila melanogaster*. Genetics **153:** 1297–1311.

Balakirev, E. S., E. I. Balakirev, F. Rodriguez-Trelles and F. J. Ayala, 1999 Molecular evolution of two linked genes, *Est6* and *Sod*, in *Drosophila melanogaster*. Genetics **153:** 1357–1369.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550.

Begun, D. J., and C. F. Aquadro, 1994 Evolutionary inferences from DNA variation at the 6-phosphogluconate dehydrogenase locus in natural populations of *Drosophila*: selection and geographic differentiation. Genetics **136:** 155–171.

Begun, D. J., and C. F. Aquadro, 1995 Molecular variation at the vermilion locus in geographically diverse populations of *Drosphila melanogaster* and *D. simulans*. Genetics **140:** 1019–1032.

Bénassi, V., and M. Veuille, 1995 Comparative population structuring of molecular and allozyme variation of *Drosophila melanogaster Adh* between Europe, West Africa and East Africa. Genet. Res. **65:** 95–103.

Bénassi, V., S. Aulard, S. Mazeau and M. Veuille, 1993 Molecular variation of *Adh* and *P6* genes in African populations of *Drosophila melanogaster* and its relation to chromosomal inversions. Genetics **134:** 789–799.

Choudhary, M., and C. C. Laurie, 1991 Use of *in vitro* mutagenesis to analyse the molecular basis of the difference in *Adh* expression associated with the allozyme polymorphism in *Drosphila melanogaster*. Genetics **129:** 481–488.

Cochrane, B. J., and R. C. Richmond, 1979 Studies of esterase 6 in *Drosophila melanogaster*. II. The genetics and frequency distributions of naturally occurring variants studied by electrophoretic and heat stability criteria. Genetics **93:** 461–478.

Collet, C., K. M. Nielsen, R. J. Russell, M. Karl, J. G. Oakeshott *et al.*, 1990 Molecular analysis of duplicated esterase genes in *Drosophila melanogaster*. Mol. Biol. Evol. **7:** 9–28.

Cooke, P. H., and J. G. Oakeshott, 1989 Amino acid polymorphisms for esterase 6 in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **86:** 1426–1430.

David, J. R., and P. Capy, 1988 Genetic variation of *Drosophila melanogaster* populations. Trends Genet. **4:** 106–111.

Dumancic, M. M., J. G. Oakeshott, R. J. Russell and M. J. Healy, 1998 Characterisation of the EstP protein in *Drosophila melanogaster* and its conservation in Drosophilids. Biochem. Genet. **35:** 251–271.

Eanes, W. F., M. Kirchner, J. Yoon, C. H. Biermann, I.-N. Wang *et al.*, 1996 Historical selection, amino acid polymorphism and lineage-specific divergence at the *G6pd* locus *in Drosophila melanogaster* and *D. simulans*. Genetics **144:** 1027–1041.

Fickett, J. W., and W. W. Wasserman, 2000 Discovery and modelling of transcriptional regulatory regions. Curr. Opin. Biotech. **11:** 19–24.

FlyBase, 1999 The FlyBase database of the Drosophila genome projects and community literature. Nucleic Acids Res. **27:** 85–88 (http://flybase.bio.indiana.edu/).

Game, A. Y., and J. G. Oakeshott, 1989 Variation in the amount and activity of esterase 6 in a natural population of *Drosophila melanogaster*. Heredity **62:** 27–34.

Game, A. Y., and J. G. Oakeshott, 1990 Associations between restriction site polymorphism and enzyme activity variation for esterase 6 in *Drosophila melanogaster*. Genetics **126:** 1021–1031.

Gilbert, D. G., and R. C. Richmond, 1982 Esterase 6 in *Drosophila melanogaster*: reproductive function of active and null alleles at low temperature. Proc. Natl. Acad. Sci. USA **79:** 2962–2966.

Gilbert, D. G., R. C. Richmond and K. B. Sheehan, 1981 Studies of esterase 6 in *Drosophila melanogaster*. VII. The timing of remating in females inseminated by males having active or null alleles. Behav. Genet. **11:** 195–208.

Gloor, G. B., C. R. Preston, D. M. Johnson-Schlitz, N. A. Nassif, R. W. Phillis *et al.*, 1993 Type I repressors of *P* element mobility. Genetics **135:** 81–95.

Hasson, E., and W. F. Eanes, 1996 Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. Genetics **144:** 1565–1575.

Healy, M. J., M. M. Dumancic, A. Cao and J. G. Oakeshott, 1996 Localization of sequences regulating ancestral and acquired sites of Esterase6 activity in *Drosophila melanogaster*. Mol. Biol. Evol. **13:** 784–797.

Higuchi, R. G., and H. Ochman, 1989 Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17:** 5865.

Inoue, Y., and Y. Igarashi, 1994 On the category of naturally occurring inversions of *Drosophila melanogaster*. Jpn. J. Genet. **69:** 105–118.

Irvin, S. D., K. A. Wetterstrand, C. M. Hunter and C. F. Aquadro, 1998 Genetic variation and differentiation at microsatellite loci in *Drosophila simulans*: evidence for founder effects in new world populations. Genetics **150:** 777–790.

Johnson, F. M., B. B. Wallis and C. Dennison, 1966 Recessive esterase deficiencies controlled by alleles of *EstC* and *Est6* in *Drosophila melanogaster*. Dros. Inf. Serv. **41:** 159.

Karotam, J., A. C. Delves and J. G. Oakeshott, 1993 Conservation and change in structural and 5′ flanking sequences of esterase 6 in sibling *Drosophila* species. Genetica **88:** 11–28.

Keplinger, B. L., G. Xinmei, J. Qunie, Y. Feng and D. R. Cavener, 2001 Complex organization of promoter and enhancer elements regulate the tissue- and developmental stage-specific expression of the *Drosophila melanogaster Gld* gene. Genetics **157:** 699–716.

Kindahl, E. C., 1994 Recombination and DNA polymorphism on the third chromosome of *Drosophila melanogaster*. Ph.D. Thesis, Cornell University, Ithaca, NY.

Labate, J., A. Bortoli, A. Y. Game, P. H. Cooke and J. G. Oakeshott, 1989 The number and distribution of esterase 6 alleles in populations of *Drosophila melanogaster*. Heredity **63:** 203–208.

Lachaise, D., M.-L. Cariou, J. R. David, F. Lemeunier and L. Tsacas, 1988 Historical biogeography of the Drosophila melanogaster species subgroup. Evol. Biol. **22:** 159–225.

Langley, C. H., R. A. Voelker, A. J. Leigh Brown, S. Ohnishi, B. Dickinson *et al.*, 1981 Null allele frequencies at allozyme loci in natural populations of *Drosophila melanogaster*. Genetics **99:** 151–156.

Laurie, C. C., and L. F. Stam, 1994 The effect of an intronic polymorphism on alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics **138:** 379–385.

Laurie-Ahlberg, C. C., and L. F. Stam, 1987 Use of *P*-element-mediated transformation to identify the molecular-basis of naturally-occurring variants affecting *Adh* expression in *Drosophila melanogaster*. Genetics **115:** 129–140.

Ludwig, M. Z., C. Bergman, N. H. Patel and M. Kreitman, 2000 Evidence for stabilising selection in a eukaryotic enhancer element. Nature **403:** 564–567.

Meikle, D., K. Sheehan, D. Phillis and R. C. Richmond, 1990 Localisation and longevity of seminal-fluid Esterase 6 in mated female *Drosophila melanogaster*. J. Insect Physiol. **36:** 93–101.

Myers, M. A., M. J. Healy and J. G. Oakeshott, 1996 Mutational analysis of *N*-linked glycosylation of Esterase 6 in *Drosophila melanogaster*. Biochem. Genet. **34:** 201–218.

Oakeshott, J. G., G. K. Chambers, J. B. Gibson and D. A. Willcocks, 1981 Latitudinal relationships of esterase 6 and phosphoglucomutase gene frequencies in *Drosophila melanogaster*. Heredity **47:** 385–396.

Oakeshott, J. G., E. A. van Papenrecht, C. Claudianos, B. C. Morrish, C. Coppin *et al.*, 2001 An episode of accelerated amino acid change in Drosophila esterase-6 associated with a change in physiological function. Genetica **110:** 231–244.

Odgers, W. A., M. J. Healy and J. G. Oakeshott, 1995 Nucleotide polymorphism in the 5′ promoter region of *Esterase 6* in *Drosophila melanogaster* and its relationship to enzyme activity variation. Genetics **141:** 215–222.

Richmond, R. C., D. G. Gilbert, K. B. Sheehan, M. H. Gromko and F. M. Butterworth, 1980 Esterase 6 and reproduction in *Drosophila melanogaster*. Science **207:** 1483–1485.

Richmond, R. C., K. M. Nielsen, J. P. Brady and E. M. Snella, 1990 Physiology, biochemistry and molecular biology of the *Est-6* locus in *Drosophila melanogaster*, pp. 273–293 in *Ecological and Evolutionary Genetics of Drosophila*, edited by J. S. F. Barker, W. T. Starmer and R. J. MacIntyre. Plenum Press, New York.

Robertson, H. M., C. R. Preston, R. W. Phillis, D. M. Johnson-Solez, W. K. Benz *et al.*, 1988 A stable genomic source of *P*-element transposase in *D. melanogaster*. Genetics **118:** 461–470.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Saad, M., A. Y. Game, M. J. Healy and J. G. Oakeshott, 1994 Associ-

ations of esterase 6 allozyme and activity variation with reproductive fitness in *Drosophila melanogaster*. Genetica **94:** 43–56.

Schlötterer, C., C. Vogl and D. Tautz, 1997 Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. Genetics **146:** 309–320.

Scott, D., 1986 Inhibition of female *Drosophila melanogaster* remating by a seminal fluid protein (esterase 6). Evolution **40:** 1084–1091.

Sheehan, K., R. C. Richmond and B. J. Cochrane, 1979 Studies of Esterase 6 in *Drosophila melanogaster*. III. The developmental pattern and tissue distribution. Insect Biochem. **9:** 443–450.

Stam, L. F., and C. C. Laurie, 1996 Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics **144:** 1559–1564.

Voelker, R. A., C. H. Langley, A. J. Leigh Brown, S. Ohnishi, B. Dickinson *et al.*, 1980 Enzyme null alleles in natural populations of *Drosophila melanogaster*: frequencies in a North Carolina population. Proc. Natl. Acad. Sci. USA **77:** 1091–1095.

Yuh, C.-H., R. H. Bolouri and E. H. Davidson, 1998 Genomic cis-regulatory logic: experimental and computational analysis of a sea-urchin gene. Science **279:** 1896–1902.

Communicating editor: M. Aguadé