Molecular Evolution of the Metallothionein Gene *Mtn* in the *melanogaster* Species Group: Results from *Drosophila ananassae*

Wolfgang Stephan,* Virgen S. Rodriguez,*,† Bingjing Zhou*,† and John Parsch*,†

**Department of Zoology and †Graduate Program in Molecular and Cell Biology, University of Maryland, College Park, Maryland 20742*

Manuscript received December 8, 1993
Accepted for publication May 23, 1994

ABSTRACT

Three distinctly different alleles of the metallothionein gene *Mtn* have been identified in natural *Drosophila melanogaster* populations: Mtn^3, Mtn^1, and $Dp(Mtn^1)$, where the latter designates a tandem duplication of Mtn^1. In *Drosophila simulans*, only Mtn^3-type alleles have been found. It has been suggested that Mtn^3 is the ancestral allele and demonstrated that a presumed two-step transition from Mtn^3 to Mtn^1 to $Dp(Mtn^1)$ is accompanied by an approximate 5-fold increase in RNA levels. We analyzed the evolutionary genetics of the *Mtn* locus of *Drosophila ananassae*, a distant relative of *D. melanogaster* and *D. simulans* within the *melanogaster* species group. The *Mtn* gene of *D. ananassae* is most similar to Mtn^3: (i) it is identical with Mtn^3 at the amino acid level, but differs from Mtn^1 in its terminal codon; (ii) its 3′ UTR contains a characteristic extra DNA segment of about 50 bp which is present in Mtn^3, but lacking in Mtn^1; (iii) duplications of *Mtn* were not found in a worldwide sample of 110 wild *D. ananassae* chromosomes. However, the intron of the *Mtn* gene in *D. ananassae* is only 69 bp long, whereas the length of the Mtn^3 and Mtn^1 introns is 265 bp; and it lacks a polypyrimidine stretch upstream of the 3′ splice site in contrast to the much greater pyrimidine-richness found in the Mtn^3 and Mtn^1 introns. A short intron (67 bp) was also identified in a *D. pseudoobscura Mtn* allele, suggesting that the short intron is the ancestral form and that the transition from the short to the long intron occurred within the *melanogaster* species group. We discuss the significance of this observation with regard to the recently proposed classification of *D. melanogaster* introns into two groups: short introns (<90 bp) which tend to lack polypyrimidine stretches, and longer ones which have strong 3′ splice signals similar to mammalian introns. A database search revealed that this length dimorphism is an evolutionarily conserved feature of Drosophila introns; transitions from one size class to the other appear to be rare between closely related species (*e.g.*, within the *melanogaster* subgroup).

METALLOTHIONEINS (MTs) are small, metal-binding proteins which have been studied in vertebrates, invertebrates, plants, fungi and prokaryotes. Dual functions have been suggested for MTs: metal homeostasis and metal detoxification (KARIN 1985). MTs have the capacity to transfer copper and zinc to metalloenzymes *in vitro* (SEAGRAVE *et al.* 1986; CHURCHICH *et al.* 1989). The fact that MTs are induced by heavy metals such as cadmium, copper and zinc also suggests that MTs have a role in protecting organisms from the deleterious effects of metal contamination in their environments (BEACH and PALMITER 1981; FOGEL and WELCH 1982; KOROPATNICK *et al.* 1985). Both copper and zinc are toxic only in high concentrations. Cadmium which has no known biological function is toxic in very small amounts.

Certain commonalities among the various MT proteins, and among the genes that encode them, have been described. All vertebrate MTs are composed of 61 amino acids, 20 of which are cysteine residues, separated into two domains. Each domain contains a cluster of cysteines with which metal ions associate through thiolate bonds (HAMER 1986). The MTs of most invertebrates, such as Drosophila, appear to be single-domain proteins, containing too few cysteine residues to form two metal-binding clusters. In *Drosophila melanogaster*, two MT genes, designated *Mtn* (LASTOWSKI-PERRY *et al.* 1985) and *Mto* (MOKDAD *et al.* 1987), are known to exist. The *Mtn* protein contains 40 amino acids, the *Mto* protein 43. These two proteins share only 30% amino acid identity. Sequence identity between the *Mto* and *Mtn* proteins is mainly due to the fact that eight of the ten cysteines of the *Mtn* protein are conserved.

Two different single-copy alleles of the *Mtn* gene in *D. melanogaster* natural populations have been identified: Mtn^1 and Mtn^3 (MARONI *et al.* 1986; THEODORE *et al.* 1991). In *Drosophila simulans*, only Mtn^3 has been found (LANGE *et al.* 1990; THEODORE *et al.* 1991). Based on these observations, THEODORE *et al.* (1991) suggested that Mtn^3 is the ancestral allele. The two alleles differ by a single base substitution in the coding region that replaces the terminal Glu^{40} in Mtn^1 with Lys^{40}. Furthermore, a 49-bp segment in the 3′ UTR is deleted in Mtn^1. In natural *D. melanogaster* populations, the

Dedicated to the memory of ELAINE M. KASPER.

allele Mtn^1 (not Mtn^3) has also been found in duplicated form, designated $Dp(Mtn^1)$. These tandem duplications occur in low frequency (LANGE *et al.* 1990; THEODORE *et al.* 1991). *Mtn* RNA levels in *D. melanogaster* larvae homozygous for Mtn^3 are approximately 3-fold lower than in Mtn^1 larvae and almost 5-fold lower than in $Dp(Mtn^1)$ ones (THEODORE *et al.* 1991). Thus, the presumed two-step transition from Mtn^3 to $Dp(Mtn^1)$ is accompanied by a 5-fold increase in RNA production. There is some evidence to suggest that the two-step transition from Mtn^3 to Mtn^1 to $Dp(Mtn^1)$ would render flies a selective advantage in metal-contaminated environments. Four different tandem duplications found in natural *D. melanogaster* populations have been associated with increased metal tolerance in laboratory experiments (MARONI *et al.* 1987). A direct demonstration of a selective advantage of Mtn^1 duplications in nature, however, is difficult and has not yet been achieved (LANGE *et al.* 1990).

The fact that three different alleles have been found in the evolution of the *Mtn* gene in the *D. melanogaster* species complex suggests that this gene has undergone several changes in recent history. This observation and the rather detailed structural and functional information on this small gene make this system highly suitable for evolutionary studies. In this paper, we describe the molecular evolution of the *Mtn* gene in a wider phylogenetic range with the aim of identifying further important events in the history of this locus. We included *Drosophila ananassae* as an additional species. *D. ananassae* is a member of the *melanogaster* species group, but not of the *melanogaster* subgroup (TOMIMURA *et al.* 1993). We sequenced a copy of the *Mtn* gene in *D. ananassae* (designated Mtn^a) and screened a worldwide sample of 110 wild-caught chromosomes for duplications of this gene. We found that Mtn^a shows sequence characteristics similar to Mtn^3, yet has a much smaller intron. To understand the significance of this result, we screened the Mtn^a transcriptional unit for DNA polymorphisms using SSCP analysis in a subsample of 39 chromosomes, and conducted a database search of Drosophila introns. This provided some novel insights into the mode and time scale of size changes in Drosophila introns.

MATERIALS AND METHODS

Cloning and sequencing of the *Mtn* region of *D. ananassae*: A λ-phage library was constructed in λ-ZAP (Stratagene) from a complete *Sac*I-restriction digest of genomic DNA from a *ca; px* stock (HINTON 1984). The phage library was screened with the 1430-bp *Eco*RI-*Acc*I fragment (in pUC 19) containing the *Mtn* gene of *D. melanogaster* (MARONI *et al.* 1986; LANGE *et al.* 1990). A positive clone with a 3.1-kb long insert was obtained. Restriction mapping of this clone revealed that the insert had a unique *Bam*HI site and contained the entire transcription unit. Subsequent sequencing showed that the *Bam*HI site which is located in exon 2 at coordinate 417 was conserved between *D. melanogaster* and *D. ananassae* (see Figure 2 of LANGE *et al.* 1990). The sequence of both strands was determined by the dideoxy method as in LANGE *et al.* (1990).

TABLE 1

Population samples of *D. ananassae* used in the duplication survey

Locality	No. of lines	Year of collection
Brazil		
Ubatuba	19	1986
Fiji		
Lautoka	10	1981
India		
Hyderabad	20	1981
Myanmar		
Mandalay	19	1982
Nepal		
Bharatpur	6	1990
Godawari	4	1990
Hetauda	7	1990
Kathmandu	1	1990
Samoa		
Pagp Pago	10	1981
Tonga		
Tongatapu	10	1981
Vava'u	4	1981

Ten lines from India, 12 from Myanmar and 17 from Nepal were used in the SSCP analysis (see Table 2) and in experiment I of the duplication survey, the rest in experiment II. *D. ananassae* is polytypic with respect to abdominal pigmentation which ranges from pale yellow to black (Tobari 1993a). All lines included in this table belong to the light cosmopolitan form, except the samples from Tonga which belong to the dark form.

Duplication survey: The duplication survey was conducted in two separate six-cutter experiments. In experiment I, 39 *D. ananassae* strains from three natural populations (12 lines from Mandalay (Myanmar), 10 lines from Hyderabad (India) and 17 lines from four localities in Nepal; see Table 1) were inbred by brother-sister mating for 10 generations. DNA was extracted by the CsCl method (BINGHAM *et al.* 1984) and then subjected to single- and double-digests with several six-cutter enzymes. Using the *D. ananassae Mtn* clone as a probe, each single digest revealed one to three fragments. A restriction map of strain *ca; px* for three enzymes is shown in Figure 1. The construction of this map was facilitated because the positions of the *Hin*dIII site at coordinate 22 and that of the *Bam*HI site at 417 were known from sequencing. The *Bam*HI digest yielded two fragments of 2.7 and 3.0 kb in size which could be clearly separated on 1% agarose gels. As in our previous study, we chose *Bam*HI for the duplication survey, because the map of the *Bam*HI sites was similar to that of *D. melanogaster* (LANGE *et al.* 1990); in particular, the *Bam*HI site in the transcriptional unit (position 417) was conserved between the two species. Using this method, duplications are detected by the appearance on an autoradiograph of more than two bands (LANGE *et al.* 1990).

In experiment II, rather than using inbred lines as was done in experiment I, single flies from 71 isofemale lines from a worldwide collection (Table 1) which were not yet screened in experiment I, were used. DNA from these single flies was prepared according to STELLER (1986) and digested with *Bam*HI. From each of the 71 isofemale lines, three replicates were made.

SSCP analysis: Single strand conformation polymorphism (SSCP) followed by direct sequencing of stratified subsamples

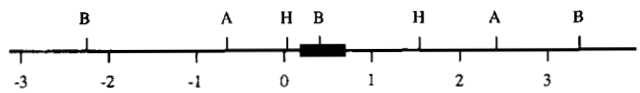

FIGURE 1.—Restriction map of the *Mtn* gene in *D. ananassae*. The box represents the transcriptional unit. The restriction sites are shown above the line: A, *Sac*I; B, *Bam*HI; H, *Hin*dIII. The coordinates are below the line; they correspond to those of Figure 2. The *Sac*I-*Sac*I fragment was used as a probe in the duplication survey (see text).

was used to survey DNA polymorphism in the transcriptional unit of the *Mtn* region in the three *D. ananassae* population samples from India, Myanmar and Nepal. The transcriptional unit was divided into two segments. Primers were placed at coordinates 134, 397 and 589 (see Figure 2). The protocol of AGUADÉ *et al.* (1994) was followed. The polymerase chain reaction (PCR) products were labeled with [α-^{35}S]dATP (DuPont). Final concentrations of 50 μM for each of cold dCTP, dGTP and dTTP, and 25 μM cold dATP and 0.05 μCi/μl of [α-^{35}S]dATP (>1000 Ci/mmol) were used. Reactions of 10 μl were set up for each line, containing 100 ng genomic DNA (for preparation, see below) and 20 ng of each primer. Amplification conditions for 30 cycles are: 92° (45 sec) for denaturing, 55° (45 sec) for annealing and 72° (45 sec) for extension. Prior to loading the samples, the machine was preheated at 90° for 2 min.

The products of the PCRs were run on a nondenaturing 0.6 × TBE, 0.5 × MDE gel (Hydrolink), with 0.6 × TBE in the upper and lower buffer reservoirs. A sample of 1.5 μl of each PCR product was mixed with 0.9 μl loading buffer. Samples were denatured at 95° for 5 min, chilled on ice and then loaded onto the gel. Gels were run at room temperature for 14 hr at 8 watts. Gels were dried and autoradiographed overnight on Kodak XAR films. The SSCPs were scored by dividing them into classes according to their mobility profile. To confirm this classification, PCR products were rerun such that samples of the same mobility class were grouped together. Representative lines from each mobility class were then sequenced using the direct method.

Direct DNA sequencing: A segment was PCR amplified such that one of the primers was kinased (HIGUCHI and OCHMAN 1989). The amplified DNA was then made single-stranded by treatment with λ exonuclease (Life Technologies, Inc.). Sequencing of single strands was done by the dideoxy method using Sequenase (Version 2.0; U.S. Biochemical Corp.). Both strands were sequenced.

***In situ* hybridization:** The cytogenetic location of *Mtn* in *D. ananassae* was determined using the biotinylated UTP-labeled *D. ananassae Mtn* clone as a probe (LANGER-SAFER *et al.* 1982). Conditions for hybridization and detection were chosen as described in MONTGOMERY and LANGLEY (1983). The coordinates of the chromosome map of TOBARI *et al.* (1993) were used to define the cytogenetic location of *Mtn*.

RESULTS

DNA sequence of the *Mtn* region in *D. ananassae*: Figure 2 shows the nucleotide sequence of the *Mtn* locus in the *ca; px* strain (designated Mtn^a). Relatively high sequence conservation was found in the functionally important parts of the *Mtn* gene. In the 5′-flanking region, the promoter elements involved in metal induction (STUART *et al.* 1984; MARONI *et al.* 1986) are well conserved between *D. melanogaster* and *D. ananassae*. Sequences which show more than 65% similarity with the mammalian consensus sequence 5′-CYTTTGCRYYCG-3′ are indicated by arrows. Here Y indicates pyrimidine and R purine. The coding region of Mtn^a is completely identical with Mtn^3 in *D. melanogaster* at the amino acid level but differs from Mtn^1 in that the terminal glutamic acid of Mtn^1 is replaced by lysine. At the DNA level, only three silent differences were found between the coding sequences of Mtn^3 and Mtn^a. The 3′-UTR of Mtn^a contains a 59-bp long segment which is homologous to the 49-bp segment of the Mtn^3 allele of *D. melanogaster* but is missing in Mtn^1. Both the 5′ and 3′ ends of this segment are conserved but otherwise there is relatively little sequence similarity. These observations show that the Mtn^a allele from *D. ananassae* is more similar to Mtn^3 than to Mtn^1, providing further evidence that Mtn^1 is derived from Mtn^3.

The overall degree of divergence between the *D. melanogaster* and *D. ananassae* alleles was determined by averaging over the silent sites (within the two coding regions) and the non-coding sites which could be aligned between Mtn^a and Mtn^3. The divergence was calculated to be 53/407 (or 13.0%). Using the Jukes-Cantor correction for multiple hits, this corresponds to an average divergence per nucleotide of 14.3%. For comparison, divergence in the homologous region between the *D. simulans Mtn* allele and Mtn^3 is 5.3%, and the average divergence per nucleotide is 5.5%. The highest levels of sequence divergence were found in the 5′- and 3′-UTRs.

The most dramatic differences between Mtn^a and Mtn^3 occur in the intron. While the introns of the Mtn^1 and Mtn^3 alleles in *D. melanogaster* are 265 bp long, that of the Mtn^a allele is only 69 bp. As indicated by the alignment of Figure 2, there appear to be short stretches of sequence conservation between these two introns, in particular at the 5′ splice site and between the putative branch point (coordinate 757) and the 3′ splice site. Putative branch sites in the *Mtn* sequences were determined by searching for a motif which matches the Drosophila branch point consensus sequence CTAAT and is at least 15 nucleotides away from the 3′ splice site and 38 nucleotides from the 5′ splice site (MOUNT *et al.*, 1992). Such a sequence was found in the *D. melanogaster* Mtn^3 allele 27 nucleotides upstream of the 3′ splice site. In Mtn^a the closest match is ATAAT which is 25 nucleotides upstream from the 3′ splice site.

In contrast to Mtn^3, the intron of Mtn^a lacks a polypyrimidine stretch upstream of the 3′ splice site. The consensus sequence of the Drosophila 3′ splice site is $Y_nNCAG|R$ (MOUNT *et al.* 1992), where N indicates any nucleotide and *n* the length of the polypyrimidine stretch (11 or 12; OHSHIMA and GOTOH 1987). In Mtn^a, only 5 of 11 (or 6/12) nucleotides are pyrimidines. Similarly, the entire segment between the putative branch site and the 3′ splice site ACAG|G contains only 11/18 (61%) pyrimidines which are clustered in two segments of 5 and 6 consecutive pyrimidines. For comparison, 10

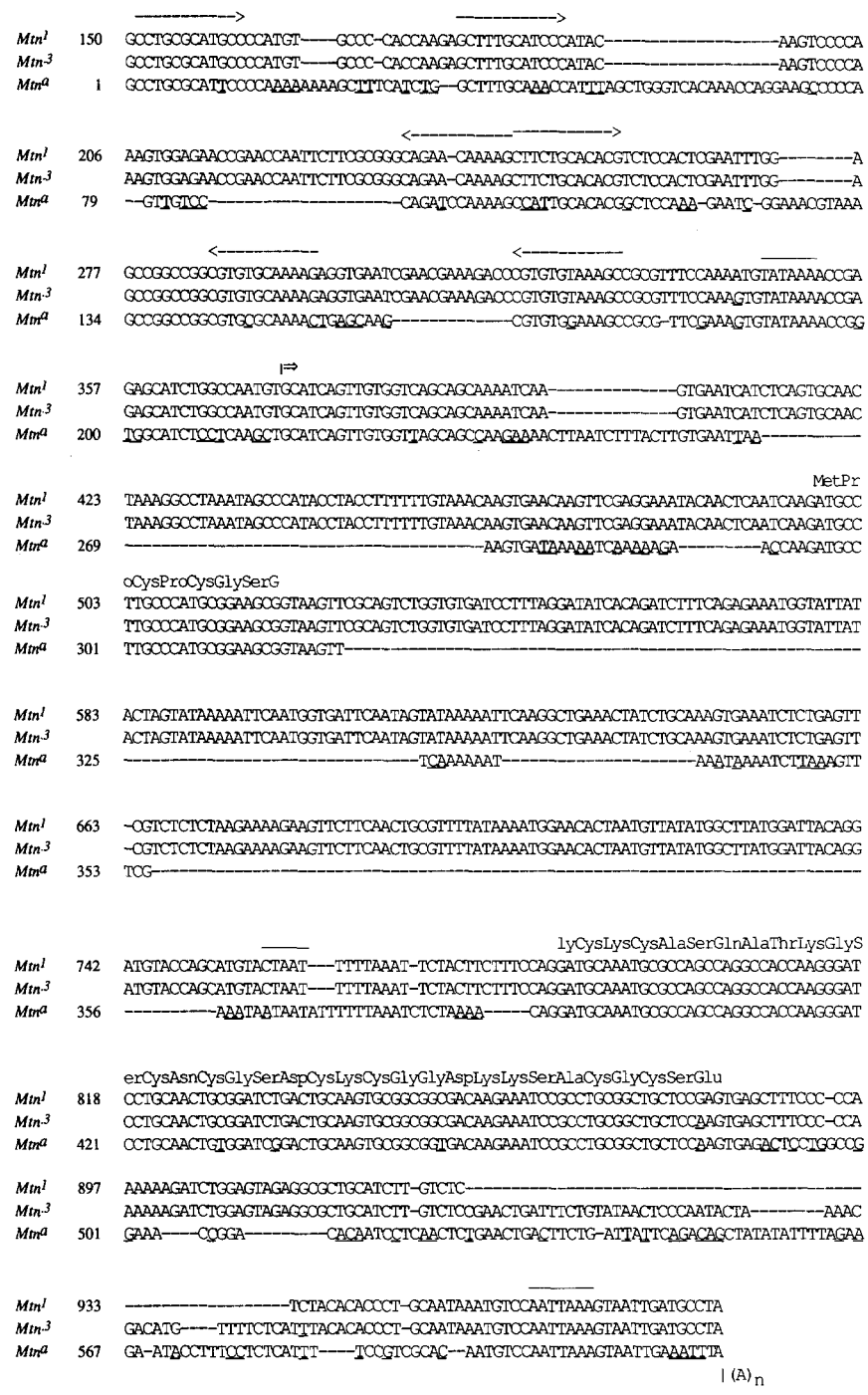

FIGURE 2.—Comparison of Mtn^1, Mtn^3 and Mtn^a sequences. The coordinates of the Mtn^1 sequence are according to GenBank (accession no. M12964). The TATA box and the putative capping and polyadenylation sites are indicated (above the sequences). The putative poly-A tail is also shown. Arrows indicate similarity of the Drosophila *Mtn* sequences with mammalian promoter elements involved in metal induction of the metallothionein gene (see text). Nucleotide differences are underlined in the Mtn^3 and Mtn^a sequences. Gaps in the alignment are denoted by dashes.

of 11 (or 11/12) nucleotides are pyrimidines upstream of the 3′ splice site in Mtn^3, and 15/19 (79%) nucleotides of the region between branch point and 3′ splice site are pyrimidines. Thus, while the intron of Mtn^3 has a strong 3′ splice signal similar to that of mammalian introns, the much smaller Mtn^a intron lacks this characteristic. This difference between short and long introns appears to be typical for *D. melanogaster*. Analyzing a database of 209 *D. melanogaster* introns, it has been found that larger introns have a greater pyrimidine content near the 3′ splice site than small ones (<90 bp) (MOUNT *et al.* (1992); S. MOUNT, personal communication). In this regard, large *D. melanogaster* introns are similar to mammalian introns, which typically have more than 80% pyrimidines in this region.

To determine whether the short intron is the ancestral form of the *Mtn* intron in the *melanogaster* species group, we used *Drosophila pseudoobscura* as outgroup (DESALLE 1992) and sequenced part of a copy of the *Mtn* gene from this species. The introns of *D. pseudoobscura* and *D. ananassae* show a high degree of sequence divergence. By contrast, their lengths are conserved. The

Mtn intron of *D. pseudoobscura* is 67 bp long and hence of similar size as the 69-bp intron of *D. ananassae*, suggesting that the short form is ancestral. The DNA sequence of the *D. pseudoobscura* intron (from the putative branchpoint CTAAT to the 3′ splice site) is: 5′-CTAATTGAGTCCATCTTCGATCTCCAG | G-3′. Thus, the pyrimidine content of the *D. pseudoobscura* intron is higher (8/11 or 9/12) than that of the Mtn^a intron (5/11 or 6/12), but lower than that of the Mtn^3 one (10/11 or 11/12).

No *Mtn* duplications in *D. ananassae*: In experiments I and II, wild chromosomes of a total of 110 isofemale lines from a worldwide sample were screened for duplications. No duplication was detected. As shown in our earlier study (LANGE *et al.* 1990), duplications of the gene should be readily detected with our method (see also Figure 1). Our negative result suggests that Mtn^a duplications in *D. ananassae* are present in very low frequency, if they exist at all. Since Mtn^a has sequence characteristics similar to Mtn^3, this result is consistent with previous reports that tandem duplications of Mtn^3 have not been observed in natural populations of *D. melanogaster* and *D. simulans* (THEODORE *et al.* 1991).

DNA polymorphism in the Mtn^a transcriptional unit: To assess whether there is variation in any of the characteristics of the *Mtn D. ananassae* allele, such as intron length, the terminal amino acid Lys^{40} and the 59-bp segment in the 3′-UTR, we used SSCP analysis and subsequent sequencing of subsamples of alleles. This analysis was done on 39 lines of three natural populations from India, Myanmar and Nepal. The results are displayed in Table 2. A region of approximately 480 bp was divided into two segments. The first segment (coordinates 134–413, corresponding to the first coordinate of the antisense primer and the last coordinate of the sense primer, respectively) contains part of the promoter region, exon 1 and the intron. In this segment, we observed five distinguishable SSCP morphs. Each morph was characterized by three major and one minor band. The upper major band was monomorphic, the two lower ones varied and could be used for classification of the SSCP morphs. Two representatives from each class were sequenced, unless a class consisted of only one member. In the second segment (coordinates 397–600), only two SSCP classes were detected.

Based on SSCP and sequencing of stratified subsamples, we found six silent site polymorphisms and one 1-bp insertion/deletion difference. All site polymorphisms were distributed across the first segment. The insertion/deletion difference occurred in the 59-bp extra segment in the 3′-UTR. We did not detect any insertion/deletion polymorphism in the intron. Average nucleotide sequence variation in this region can be estimated using the method of AGUADÉ *et al.* (1994). According to their formula, the average number of pairwise differences (per nucleotide site), π^*, is 0.004. This estimate compares well with the estimate of nucleotide diversity (NEI and LI 1979) of $\pi = 0.004$, which we obtained from a 6-cutter survey of the same 39 lines for an approximately 5-kb region around the *Mtn* gene (V. S. RODRIGUEZ and W. STEPHAN, unpublished). Both values compare also reasonably well with estimates of π from other loci in regions of normal recombination in *D. ananassae*, which range between 0.004 and 0.008 for the samples from Myanmar and India (STEPHAN and LANGLEY 1989; STEPHAN 1989). Indeed, our *in situ* hybridization data suggest that *Mtn* is located in a chromosomal region of normal recombination in *D. ananassae* (see below). Our estimates of nucleotide diversity are significantly higher than those at low recombination regions in *D. ananassae* (STEPHAN and LANGLEY 1989; STEPHAN and MITCHELL 1992). Furthermore, they are higher than the estimate of nucleotide diversity at *Mtn* in *D. melanogaster*, presumably because *Mtn* is located in a region

TABLE 2

DNA sequence variants in the *Mtn* region of *D. ananassae*

Lines	177	200	212	214	257	347	505+
Mtn^a	C	T	A	G	C	A	
M61	C	T	A	G	T	G	
M68	C	G	A	G	T	A	
M79	T	G	A	G	T	A	
M80	C	G	A	G	T	A	
M86	T	G	A	G	T	A	A
M89	T	G	A	G	T	A	
M90	C	G	A	G	T	A	
M91	T	G	A	G	T	A	
M97	T	G	A	G	T	A	
M99	T	G	A	G	T	A	
M117	T	G	A	G	T	A	
M119	T	G	A	G	T	A	
H3	T	G	A	G	T	A	
H11	C	G	G	G	C	A	
H15	T	G	A	G	T	A	
H23	T	G	A	A	T	A	
H31	T	G	A	G	T	A	
H36	T	G	A	G	T	A	
H39	C	G	A	G	T	A	
H41	T	G	A	G	T	A	
H47	C	G	A	G	T	A	
H50	T	G	A	G	T	A	
NB1	T	G	A	G	T	A	A
NB2	T	G	A	G	T	A	
NB5	T	G	A	G	T	A	
NB7	T	G	A	G	T	A	
NB9	T	G	A	G	T	A	A
NK8	C	G	A	G	T	A	
NG3	T	G	A	G	T	A	
NG4	T	G	A	G	T	A	
NG8	T	G	A	G	T	A	
NG14	T	G	A	G	T	A	
NH1	T	G	A	G	T	A	
NH5	T	G	A	G	T	A	
NH14	C	G	A	G	T	A	
NH17	C	G	G	G	C	A	
NH22	C	G	A	G	T	A	
NH23	C	G	G	G	C	A	
NH25	C	G	G	G	C	A	

The coordinates are from Figure 2. An "A" in column 505+ indicates that an A was inserted between coordinates 505 and 506. The lines are from Myanmar (M), India (H) and from four localities of Nepal (NB, NG, NH, and NK) (see Table 1).

TABLE 3

Comparison of small introns within the *melanogaster* subgroup

Gene	Intron	Size range	Species range	Accession numbers
Adh	1	63 67	mel, sim, mau, sec, tei, yak, ere, ore	M14802, X00607, M19264, X04672, X54118, X54120,
	2	61 76		X54116, Z00032
Adhr	1	51 54	mel, sim, mau, sec, tei, ere	
ci^D	1	64 64	mel, sim	X54360
	2	53 54		
G6pd	1	62 62	mel, sim	L13920, L13876
	2	69 85		
jgw	1	66 67	tei, yak	S57937, S57972
	2	54 65		
Mlc1	3	60 60	mel, sim	L08051
Mst26A	1	56 56	mel, sim, mau, sec	X70888, X70899, X70898, X72630
	2	61 61		
sal	1	57 61	mel, sim, ore	X57474, M21227, M21579
tra	2	53 62	mel, sim, ere	M17478, X66930, X66527
w	2	54 74	mel, sim, yak	X02974
yp2	1	63 68	mel, sim, mau, sec	L14422, L14426, L14417, L14424
z	2	62 62	mel, sim, mau, sec	Y00049

This table contains only introns which belong to the small size class (<90 bp) in at least one species of the *melanogaster* subgroup. All genes were drawn from GenBank/EMBL DNA sequence library (GenBank release 78.0) except the *Adhr* data (JEFFS *et al.* 1994), *D. simulans* ci^D (BERRY *et al.* 1991), *D. melanogaster Mlc1* (FALKENTHAL *et al.* 1985), *D. simulans* and *D. yakuba w* (D.A. KIRBY and W. STEPHAN, unpublished data), and the *D. simulans, D. mauritiana* and *D. sechellia z* sequences (HEY and KLIMAN 1993). The second column gives the intron number within a gene. The third column gives the intron size range (the smallest and largest values) for the species of the *melanogaster* subgroup shown in column 4. The species are abbreviated as follows: mel = *D. melanogaster*; sim = *D. simulans*; mau = *D. mauritiana*; sec = *D. sechellia*; tei = *D. teissieri*; yak = *D. yakuba*; ere = *D. erecta*; ore = *D. orena*. The *per* gene was not included because the physical organization of this locus is not sufficiently conserved within the *melanogaster* species subgroup (THACKERAY and KYRIACOU 1990). Note that the size of the introns available for this analysis was conserved within the small size class.

of restricted meiotic recombination in *D. melanogaster* [LANGE *et al.* (1990), and references therein].

Cytogenetic location of *Mtn* in *D. ananassae*: By *in situ* hybridization, we localized *Mtn* on the left arm of the second chromosome at position 25B. For comparison, *Mtn* in *D. melanogaster* is on the right arm of chromosome *3* (which corresponds to chromosome *2* in *D. ananassae*) at 85E10-15 (MARONI *et al.* 1986), near the centromere. Thus, while *Mtn* in *D. melanogaster* is located close to the centromere in a region of restricted recombination, its position in *D. ananassae* is in the middle of the left chromosome arm. There is no evidence that the genetic map is heavily contracted in the latter region (TOBARI 1993b). However, a more quantitative examination of the rate of recombination in the *Mtn* region in *D. ananassae* is not possible at present, because the density of known genetic markers in that part of the genome is low.

DISCUSSION

Molecular evolution of the *Mtn* gene in the *melanogaster* species group: This study and previous reports (MARONI *et al.* 1986, 1987; LANGE *et al.* 1990; THEODORE *et al.* 1991) suggest that three significant changes have occurred in the *Mtn* gene in the *melanogaster* species group: (1) a transition from a small intron to a large one prior to the *D. melanogaster-D. simulans* split (~5 mya); (2) a replacement of the terminal amino acid lysine by glutamic acid accompanied by the loss of an approximately 50-bp segment in the 3′ UTR in the lineage leading to *D. melanogaster*; and (3) a duplication of this latter allele in *D. melanogaster*. The alleles created in steps 1 to 3, Mtn^3, Mtn^1 and $Dp(Mtn^1)$, occur with varying frequencies in *D. melanogaster* populations. In contrast, no variation in the essential characteristics of the *Mtn* gene (*i.e.*, intron length, terminal amino acid, and copy number) in *D. ananassae* populations was detected in this study.

Laboratory experiments have utilized the occurrence of *Mtn* polymorphisms in natural *D. melanogaster* populations to elucidate the functional differences of the Mtn^3, Mtn^1 and $Dp(Mtn^1)$ alleles. It was demonstrated that a two-step transition from $Mtn^3$to Mtn^1 to $Dp(Mtn^1)$ leads in homozygotes to an approximately 5-fold increase in mRNA levels (THEODORE *et al.* 1991). A similar approach to the investigation of the functional significance of the intron change is not possible at present, because a polymorphism in *D. ananassae* populations was not detected. There are several lines of evidence, however, which suggest that the size change of the *Mtn* intron is an important evolutionary event. First, a survey of 209 *D. melanogaster* introns has revealed evidence for intron dimorphism in *D. melanogaster*, such that introns can be divided into two classes with respect to their length (MOUNT *et al.* 1992; S. MOUNT, personal communication). More than half of them belong to the small size class, ranging from less than 50 bp to approximately 90 bp (see their Figure 1B) with a mode around 63 bp. The large size class of completely sequenced introns has a large scatter and ranges up to 5,400 bp. Second, transitions from one size class to the other one appear to be rare events. A search of the database (GenBank release

TABLE 4

Between-group comparison of Drosophila introns

Gene	Intron	Species	Size (bp)		Pyrimidine content		Accession numbers	
			mel	other	mel	other	mel	other
Adh	Adult	p	654	795	0.55	0.73	M14802	M60982
	1	p	65	63	0.58	0.58		
	2	p	70	60	0.67	0.67		
bcd	1	p	559	64	0.82	0.64	X07870	X55735
	2	p	55	69	0.82	0.91		
	3	p	513	82	0.91	0.82		
boss	1	v	1393	2133	1.00	0.91	L08133	L08132
	2	v	304	510	0.82	0.64		
	3	v	57	63	0.64	0.73		
	4	v	121	53	1.00	0.55		
	5	v	61	55	0.73	0.82		
Cp15	1	v	71	68	0.91	0.64	X02497	X53421
Cp16	1	v	121	81	0.83	0.75	X16715	X53421
Cp18	1	v	176	67	0.73	0.64	X02497	X53421
Cp19	1	v	89	85	0.82	0.64	X02497	X53421
en	1	v	1133	1737	0.83	0.83		
	2	v	280	373	0.91	1.00		
Gld	1	p	2992	2813	0.73	0.64	M29298	M29299
	2	p	73	76	0.64	0.73		
	3	p	339	1206	0.64	0.73		
Hsp 82	1	p	1129	1063	0.55	0.91	X03810	X03812
Mlc1	1	p	153	139	0.67	0.67		L08052
	2	p	985	1163	0.82	0.64		
	3	p	60	70	0.73	0.55		
	4	p	450	330	0.27	0.64		
	5	p	110	138	0.82	0.91		
Mtn	1	p	265	67	0.91	0.73	M12964	
Rh1	1	p	63	57	0.64	0.82	K02315	X65877
	2	p	188	376	0.73	0.82		
	3	p	58	70	0.82	0.73		
	4	p	69	60	0.82	0.91		
Rh2	1	p	96	65	0.73	0.64	M12896	X65878
	2	p	82	62	0.45	0.73		
	3	p	54	63	0.91	0.82		
rp49	1	p	59	72	0.82	0.73	X00848	S59382
tll	1	v	121	99	0.73	0.82	L04954	L04955
tra	1	v	248	257	0.82	0.55	M17478	X66528
	2	v	57	63	0.55	0.73		
tub	1	v	67	60	0.73	0.82	L20448	L20449
Uro	1	v	69	55	0.75	0.67	X51940	X57114
Xdh	1	p	815	1024	0.82	0.73	Y00308	M33977
	2	p	274	62	0.58	0.50		
	3	p	65	67	0.64	0.73		
z	1	v	121	80	0.67	0.75	Y00049	M76700
	2	v	62	88	0.91	0.82		

All genes were drawn from GenBank/EMBL DNA sequence library (GenBank release 78.0) except *D. melanogaster Mlc1* (FALKENTHAL *et al.* 1985), *D. pseudoobscura Mtn* (this study) and *en* (KASSIS *et al.* 1986). "Species" refers to the species used in the between-group comparison: "p" denotes *D. pseudoobscura*, and "v" refers to *D. virilis*. Intron sizes for *D. melanogaster* (mel) and the "other" species are given in columns 4 and 5, respectively. Columns 6 and 7 contain the respective pyrimidine contents. Pyrimidine content in the region upstream of the 3′ splice site was determined as follows. The length n of the polypyrimidine stretch of the consensus 3′ splice site Y_nNCAG is somewhat ill defined (n = 11 or 12; OHSHIMA and GOTOH 1987). Therefore, we calculated first the pyrimidine content as the portion of pyrimidines in a stretch of 11 nucleotides (criterion 1). If this portion was equal for the two species in a given comparison, we determined the pyrimidine content for a stretch of n = 12 (criterion 2). Comparison between *D. melanogaster* and the other species revealed that 36 of the 45 introns available belonged either to the small size class (<90 bp) or to the large one in both species. In the remaining 9 introns, changes from one size class to other have occurred. In these 9 comparisons, the large introns have significantly more pyrimidines near the 3′ splice site than the small ones (paired t-test: $t = 2.723$, $P = 0.026$; see text). The assumptions of the paired t-test (*i.e.*, the homogeneity and normality of the distribution of the residuals) were met.

78.0) revealed no transitions from small to large introns (or vice versa) for introns within the *melanogaster* species subgroup (Table 3). The age of the *melanogaster* species subgroup has been estimated to be 17–20 million years (LACHAISE *et al.* 1988). If intron sequences were largely neutral, as it is often assumed, DNA polymor-

phisms within and between closely related species should occur frequently. This is indeed the case for nucleotide sequence variation; and large insertions and deletions occur frequently in large introns. In contrast, the size of small introns (<90 bp) appears to be conserved, even if the sequence is variable. Changes between the small and large size classes were found, however, when the *melanogaster* species group was compared with outgroups such as the *obscura* group. The divergence time between the *melanogaster* and *obscura* groups is approximately 46 million years (BEVERLEY and WILSON 1984). Nine of 45 introns available from the database or the literature showed transitions between the small and large size classes (Table 4).

Intron size changes and polypyrimidine content: The survey of *D. melanogaster* introns (MOUNT *et al.* 1992) also found that large introns have a greater pyrimidine richness near the 3′ splice site than small ones. To further investigate the relationship between intron length and pyrimidine content, we examined the changes in intron size between *D. melanogaster* and outgroup species found in the database or the literature (Table 4). Nine of 45 introns from Table 4 changed in size. Eight of these nine changes occurred such that the larger intron had the higher pyrimidine content, a significant proportion (paired *t*-test: $t = 2.723$, $P = 0.026$). In all nine comparisons, *D. melanogaster* has the larger intron. This may indicate that the larger pyrimidine content in large introns is not due to size changes but may be species specific. We did not, however, find evidence supporting this hypothesis. We tested whether the introns from Table 4 which did not change size class have a higher pyrimidine content in *D. melanogaster*. The differences between species were not significant (paired *t*-test: $t = 0.68$, $P > 0.5$).

Direct experimental evidence suggests that the efficiency of splicing depends on the content of the pyrimidine stretch upstream of the 3′ splice site (GUO *et al.* 1993). These experiments were done on the second intron of the *D. melanogaster white* gene which is 74 bp long and lacks a significant polypyrimidine stretch. Using Drosophila Kc cell nuclear extract as an *in vitro* system, it was found that (i) the wild-type second intron is accurately removed, (ii) increasing the pyrimidine content near the 3′ splice site enhances the removal of the intron, and (iii) increasing the size of the wild-type intron with insertions between the 5′ splice site and the branch point greatly reduced the efficiency of splicing of introns longer than 79 bp. These observations, along with our results, suggest that pyrimidine content may be an important parameter in the evolution of Drosophila introns such that increases in intron size are expected to be associated with greater pyrimidine richness near the 3′ splice site.

We thank SOICHI TANDA for help with *in situ* hybridization; DAVID KIRBY for assistance in the application of the PCR-based sequencing technique; CHARLES LANGLEY for the SSCP protocol; STEPHEN MOUNT for discussion of the intron dimorphism; LEROY POFF for assistance in the statistical analysis; STEVEN SCHAEFFER for *D. pseudoobscura* DNA; and two anonymous reviewers for comments on an earlier version of this paper. Furthermore, we are grateful to three undergraduates for their help in the initial phase of this project: ANGELIQUE BEST, JAQUELINE HARTZLER and ELAINE KASPER (deceased). This work was supported in part by a grant from the National Institutes of Health (GM 46233) to W.S. and by fellowships from the Graduate Program in Molecular and Cell Biology (University of Maryland) to J.P. and B.Z., and from the Summer Undergraduate Research Program for minority students to V.S.R.

LITERATURE CITED

AGUADÉ, M., W. MEYERS, A. D. LONG and C. H. LANGLEY, 1994 Single-strand conformation polymorphism analysis coupled with stratified DNA sequencing reveals reduced sequence variation in the *su(s)* and *su(w^a)* regions of the *Drosophila melanogaster* X chromosome. Proc. Natl. Acad. Sci. USA **91:** 4658–4662.

BEACH, L. R., and R. D. PALMITER, 1981 Amplification of the metallothionein-I gene in cadmium- resistant mouse cells. Proc. Natl. Acad. Sci. USA **78:** 2110–2114.

BERRY, A. J., J. W. AJIOKA and M. KREITMAN, 1991 Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

BEVERLEY, S. M., and A. C. WILSON, 1984 Molecular evolution in Drosophila and the higher Diptera. II. A time scale for fly evolution. J. Mol. Evol. **21:** 1–13.

BINGHAM, P. M., R. LEVIS and G. M. RUBIN, 1981 Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. Cell **25:** 693–704.

CHURCHICH, J. E., G. SCHOLTZ and K. KWORK, 1989 Activation of pyridox kinase by metallothionein. Biochem. Biophys. Acta **996:** 181–186.

DESALLE, R., 1992 The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol. Phylogenet. Evol. **1:** 31–40.

FALKENTHAL, S., V. P. PARKER and N. DAVIDSON, 1985 Developmental variations in the splicing pattern of transcripts from the Drosophila gene encoding myosin light chain result in different carboxyl-terminal amino acid sequences. Proc. Natl. Acad. Sci. USA **82:** 449–453.

FOGEL, S., and J. W. WELCH, 1982 Tandem gene amplification mediates copper resistance in yeast. Proc. Natl. Acad. Sci. USA **79:** 5342–5346.

GUO, M., P. C. H. LO and S. M. MOUNT, 1993 Species-specific signals for the splicing of a short Drosophila intron *in vitro*. Mol. Cell. Biol. **13:** 1104–1118.

HAMER, D. H., 1986 Metallothionein. Annu. Rev. Biochem. **55:** 913–951.

HEY, J., and R. M. KLIMAN, 1993 Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10:** 804–822.

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.

HINTON, C. W., 1984 Morphogenetically specific mutability in *Drosophila ananassae*. Genetics **106:** 631–653.

JEFFS, P. S., E. C. HOLMES and M. ASHBURNER, 1994 The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. **11:** 287–304.

KARIN, M., 1985 Metallothionein: proteins in search of function. Cell **41:** 9–10.

KASSIS, J. A., S. J. POOLE, D. K. WRIGHT and P. H. O'FARRELL, 1986 Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. EMBO J. **5:** 3583–3589.

KOROPATNICK, J., R. WINNING, E. WIESE, M. HESCHL, L. GEDAMU *et al.*, 1985 Acute treatment of mice with cadmium salts results in amplification of the metallothionein-I gene in liver. Nucleic Acids Res. **13:** 5423–5439.

LACHAISE, D., M.-L. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22:** 159–225.

LANGE, B. W., C. H. LANGLEY and W. STEPHAN, 1990 Molecular evo-

lution of Drosophila metallothionein genes. Genetics **126:** 921–932.

LANGER-SAFER, P. R., M. LEVINE and D. C. WARD, 1982 Immunological method for mapping genes on Drosophila polytene chromosomes. Proc. Natl. Acad. Sci. USA **79:** 4381–4385.

LASTOWSKI-PERRY, D., E. OTTO and G. MARONI, 1985 Nucleotide sequence and expression of a Drosophila metallothionein. J. Biol. Chem. **260:** 1527–1530.

MARONI, G., E. OTTO and D. LASTOWSKI-PERRY, 1986 Molecular and cytogenetic characterization of a metallothionein gene of Drosophila. Genetics **112:** 493–504.

MARONI, G., J. WISE, J. E. YOUNG and E. OTTO, 1987 Metallothionein gene duplications and metal tolerance in natural populations of *Drosophila melanogaster.* Genetics **117:** 739–744.

MOKDAD, R., A. DEBEC and M. WEGNEZ, 1987 Metallothionein genes in *Drosophila melanogaster* constitute a dual system. Proc. Natl. Acad. Sci. USA **84:** 2658–2662.

MONTGOMERY, E. A., and C. H. LANGLEY, 1983 Transposable elements in Mendelian populations. II. Distribution of three copia-like elements in a natural population of *Drosophila melanogaster.* Genetics **104:** 473–483.

MOUNT, S. M., C. BURKS, G. HERTZ, G. D. STORMO, O. WHITE *et al.*, 1992 Splicing signals in Drosophila: intron size, information content, and consensus sequences. Nucleic Acids Res. **20:** 4255–4262.

NEI, M., and W.-H. LI, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

OHSHIMA, Y., and Y. GOTOH, 1987 Signals for the selection of a splice site in pre-mRNA–computer ananlysis of splice junction sequences and like sequences. J. Mol. Biol. **195:** 247–259.

SEAGRAVE, J.-C., J. L. HANNERS, W. TAYLOR and H. A. O'BRIEN, 1986 Transfer of copper from metallothionein to nonmetallothionein proteins in cultured cells. Biol. Trace Elem. Res. **10:** 163–173.

STELLER, H., 1986 Rapid small scale isolation of Drosophila DNA and RNA, p. 10.1 in *Rubin Lab Methods Book.* The Rubin Lab, Berkeley, Calif.

STEPHAN, W., 1989 Molecular genetic variation in the centromeric region of the *X* chromosome in three *Drosophila ananassae* populations. II. The *Om(1D)* locus. Mol. Biol. Evol. **6:** 624–635.

STEPHAN, W., and C. H. LANGLEY, 1989 Molecular genetic variation in the centromeric region of the *X* chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics **121:** 89–99.

STEPHAN, W., and S. J. MITCHELL, 1992 Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae.* Genetics **132:** 1039–1045.

STUART, G. W., P. F. SEARLE, H. Y. CHEN, R. L. BRINSTER and R. D. PALMITER, 1984 A 12-base-pair DNA motif that is repeated several times in the metallothionein gene promoters confers metal regulation to a heterologous gene. Proc. Natl. Acad. Sci. USA **81:** 7318–7322.

THACKERAY, J., and C. P. KYRIACOU, 1990 Molecular evolution in the *Drosophila yakuba period* locus. J. Mol. Evol. **31:** 389–401.

THEODORE, L. A.-S. HO and G. MARONI, 1991 Recent evolutionary history of the metallothionein gene *Mtn* in Drosophila. Genet. Res. **58:** 203–210.

TOBARI, Y. N., 1993a Geographical distribution, pp. 19–22 in *Drosophila ananassae–Genetical and Biological Aspects,* edited by Y. N. TOBARI. Japan Scientific Societies Press, Tokyo.

TOBARI, Y. N., 1993b Linkage maps, pp. 49–52 in *Drosophila ananassae–Genetical and Biological Aspects,* edited by Y. N. TOBARI. Japan Scientific Societies Press, Tokyo.

TOBARI, Y. N., B. GONI, Y. TOMIMURA and M. MATSUDA, 1993 Chromosomes, pp. 23–48 in *Drosophila ananassae–Genetical and Biological Aspects,* edited by Y. N. TOBARI. Japan Scientific Societies Press, Tokyo.

TOMIMURA, Y., M. MATSUDA and Y. N. TOBARI, 1993 Polytene chromosome variations of *Drosophila ananassae* and its close relatives, pp. 139–151 in *Drosophila ananassae–Genetical and Biological Aspects,* edited by Y. N. TOBARI. Japan Scientific Societies Press, Tokyo.

Communicating editor: G. B. GOLDING