The Origin of the *Jingwei* Gene and the Complex Modular Structure of Its Parental Gene, *Yellow Emperor,* in *Drosophila melanogaster*

Wen Wang, Jianming Zhang,* Carlos Alvarez,† Ana Llopart,* and Manyuan Long**

*Department of Ecology and Evolution, University of Chicago; and †Department of Molecular and Cellular Biology, Harvard University

Jingwei (*jgw*) is the first gene found to be of sufficiently recent origin in *Drosophila* to offer insights into the origin of a gene. While its chimerical gene structure was partially resolved as including a retrosequence of alcohol dehydrogenase (*Adh*), the structure of its non-*Adh* parental gene, the donor of the N-terminal domain of *jgw*, is unclear. We characterized this non-*Adh* parental locus, *yellow emperor* (*ymp*), by cloning it, mapping it onto the polytene chromosomes, sequencing the entire locus, and examining its expression patterns in *Drosophila melanogaster*. We show that *ymp* is located in the 96-E region; the N-terminal domain of *ymp* has donated the non-*Adh* portion of *jgw* via a duplication. The similar 5′ portions of the gene and its regulatory sequences give rise to similar testis-specific expression patterns in *ymp* and *jgw* in *Drosophila teissieri*. Furthermore, between-species comparison of *ymp* revealed purifying selection in the protein sequence, suggesting a functional constraint in *ymp*. While the structure of *ymp* provides clear information for the molecular origin of the new gene *jgw*, it unexpectedly casts a new light on the concept of genes. We found, for the first time, that the single locus of the *ymp* gene encompasses three major molecular mechanisms determining structure of eukaryotic genes: (1) the 5′ exons of *ymp* are involved in an exon-shuffling event that has created the portion recruited by *jgw*; (2) using alternative cleavage sites and alternative splicing sites, the 3′ exon groups of *ymp* produce two proteins with nonhomologous C-terminal domains, both exclusively in the testis; and (3) in the opposite strand of the third intron of *ymp* is an essential gene, *musashi* (*msi*), which encodes an RNA-binding protein. The composite gene structure of *ymp* manifests the complexity of the gene concept, which should be considered in genomic research, e.g., gene finding.

Introduction

The early history of a gene is of interest, because it addresses a general question about the origin of genes. A number of new genes with novel functions have been found which have revealed various evolutionary mechanisms underlying the origin of new genes (e.g., Long and Langley 1993; Martignetti and Brosius 1993; Ohta 1994; Long et al. 1996; Begun 1997; Chen, DeVries, and Cheng 1997). One of the major molecular processes that give rise to new genes is exon shuffling (Gilbert 1978). Many cases have been reported of new genes originating via exon shuffling (Patthy 1995; Long and Langley 1993; Long et al. 1996; Nurminsky et al. 1998). However, insight into the early evolution of such genes is dependent on the discovery of a young gene because of the rapid sequence evolution characteristic of new genes as revealed by several investigations (Long and Langley 1993; Long et al. 1996; Nurminsky et al. 1998).

Jingwei (*jgw*) was the first gene observed in *Drosophila* to have recently been created by exon shuffling, and its age is estimated at around 2 Myr. A portion of *jgw* was identified in *Drosophila yakuba* and *Drosophila teisseiri* in an in situ hybridization using *Adh* as a probe (Langley, Montgomery, and Quattlebaum 1982). By cloning and sequencing this portion of the gene, Jeffs and Ashburner (1991) observed that all *Adh* introns were lost, interpreting this as a processed pseudogene resulting from random insertion of a retrosequence into a region devoid of regulatory sequences.

Further molecular population genetic analysis, however, revealed strong purifying selection, as shown by the near limitation of nucleotide polymorphism to silent sites (Long and Langley 1993). This gene was observed to have specific RNA expression patterns, and its evolution was driven by ubiquitous Darwinian positive selection (Long and Langley 1993), which usually acts only on functional genes. These results suggest that *jgw* is a newly evolved functional gene. Furthermore, molecular characterization showed that the insertion of the *Adh* retrosequence recruited nearby preexisting exons and introns and thereby created a chimerical gene structure in a standard form of exon shuffling.

What is the source of the recruited exons and introns of the *jgw* gene? They could originate from a unique noncoding genomic sequence, as is approximately seen in the genes encoding BC1 RNA in rodents and BC200 RNA in primates (Brosius and Gould 1992). Alternatively, they could have originated from a preexisting gene or a duplicate of a gene. Although Long, Wang, and Zhang (1999) demonstrated that these recruited exons and introns are a portion of a duplicate of the gene *yellow emperor* (*ymp*), the structure of *ymp* itself was unclear, and the process by which the non-*Adh* portion originated remains to be investigated.

In this paper, we report the structure and some information concerning the function of the *ymp* locus in *Drosophila melanogaster*. We found that its structure is unique and not only offers a further explanation for the origin of the *jgw* gene, but also manifests the complexity of the concept of genes. The implication of these results will be discussed with respect to genomic research, such as gene-finding from genomic sequence data.

Key words: origin of new genes, exon shuffling, nested gene, alternative splicing.

Address for correspondence and reprints: Manyuan Long, Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637. E-mail: mlong@midway.uchicago.edu.

Mol. Biol. Evol. 17(9):1294–1301. 2000

Materials and Methods

Screening cDNA and Genomic Libraries

cDNA libraries of *D. melanogaster* and *D. yakuba* and a genomic library of *D. teissieri* were screened using a ^{32}P-labeled DNA fragment containing the first three exons of *D. teissieri jgw*, following standard procedure (Sambrook, Fritsch, and Maniatis 1989). Two distinct transcripts were isolated from the *D. melanogaster* cDNA library. Both strands of the inserts were sequenced using the sequencing kit of United States Biochemical (version 2). We named the transcripts *ymp*-1 and *ymp*-2, respectively, following Long, Wang, and Zhang (1999).

Drosophila yakuba cDNA and *D. teissieri* genomic libraries were made, using Lambda ZAP II and Lambda FIX II (Stratagene, San Diego), respectively, as vectors, using protocols provided by Stratagene and Sambrook, Fritsch, and Maniatis (1989). The RNA and genomic DNA were extracted from adult flies using modified procedures from Ashburner (1989). The *D. melanogaster* cDNA library (RNA from adult flies of the Oregon R strain) was a generous gift of Dr. Bruce A. Hamilton of the Whitehead Institute, Massachusetts Institute of Technology.

Mapping *ymp* in Polytene Chromosomes Using Fluorescence In Situ Hybridization

Digoxigenin-11-dUTP (DIG) (Roche Molecular Biochemicals) or Biotin-16-dUTP (Roche Molecular Biochemicals) labeled probes were constructed specifically for the shared three 5′ exons, *ymp*-1 3′ exons, and *ymp*-2 3′ exons, respectively, by PCR. Primers A747 and A698 (Long, Wang, and Zhang 1999) were used for amplifying the three shared exons, *ymp*1F (5′-GTGCCCATTATTGCGATTTCAT-3′) and *ymp*1R (5′-TCCCTGGCCTTTTATTTCCTTC-3′) were used for the *ymp*-1 3′ exons, and y43-3 (5′-TGGCATTGGTGAAGGACG-3′) and y43-1 (5′-AAAGAAGTAGCTACTCGGC-3′) were used for the *ymp*-2 3′ exons. Polytene chromosome slides for fluorescence in situ hybridization (FISH) were prepared according to the protocols of Ashburner (1989). DIG-labeled probes were detected with rhodamine-conjugated antibody, and biotin-labeled probes were detected with fluorescein-conjugated streptavidin. Single and double color FISHs were performed as described by Wiegant (1996) with modifications.

P1 Subcloning

Based on the results of polytene chromosome in situ hybridization, which show these genes located at 96E on the third chromosome of *D. melanogaster*, we screened *D. melanogaster* P1 clones (Hartl et al. 1994) around 96E by PCR amplifications using the same primers described in *Mapping* ymp *in Polytene Chromosomes Using Fluorescence In Situ Hybridization.* These P1 clones were from the laboratory of Dr. Spyros Artavanis-Tsakonas of Yale University.

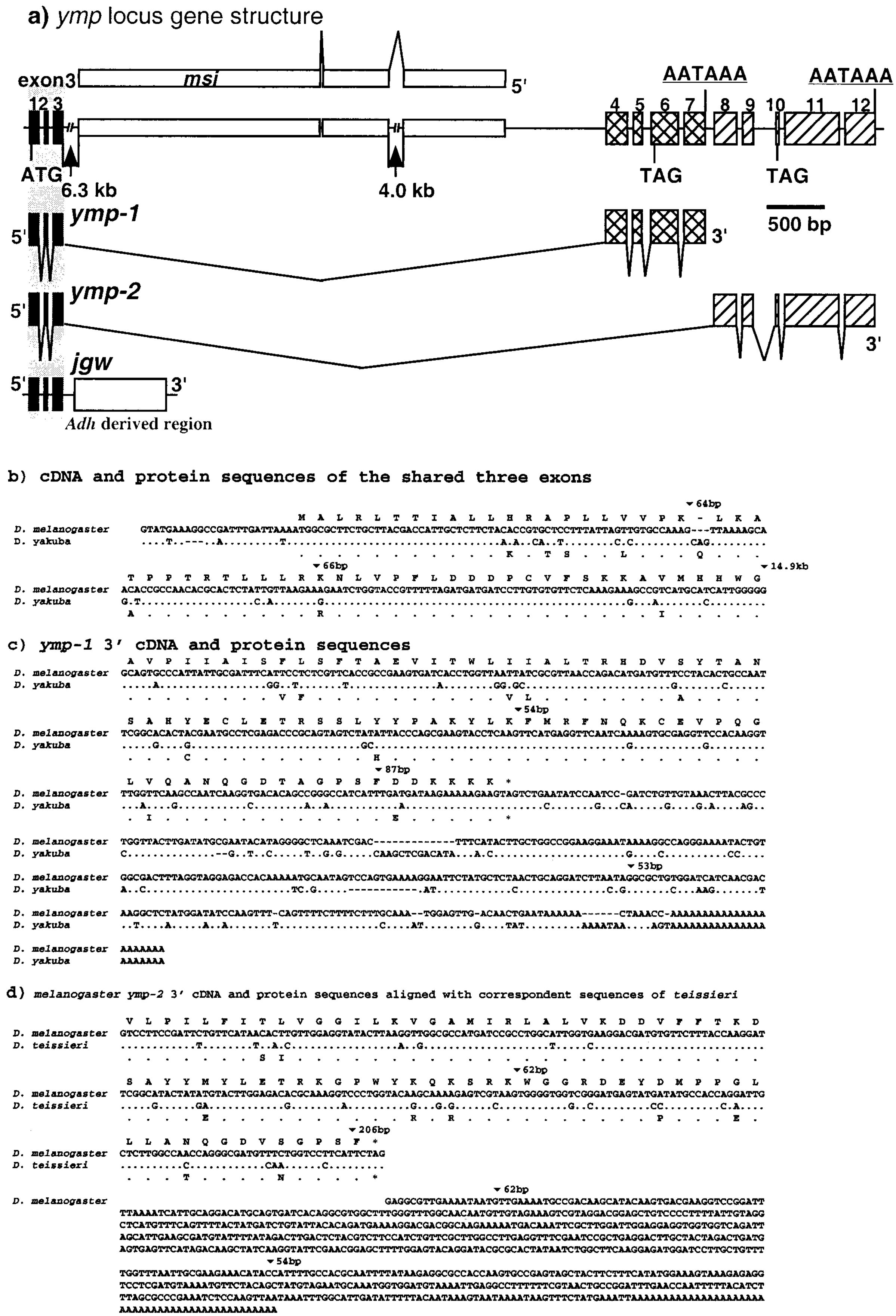
Two P1 clones (DS00423 and DS02160), each containing both the *ymp*-1 and the *ymp*-2 sequences, were identified by PCR amplification of both *ymp*-1 and *ymp*-2. DNA fragments from *Xho*I digestion of these two P1 clones were separated on a 0.7% agarose gel, and were then transferred to nylon membrane (Roche Molecular Biochemicals) by Southern blotting. The three DIG-labeled probes described in *Mapping* ymp *in Polytene Chromosomes Using Fluorescence In Situ Hybridization* were successively hybridized to the membrane. Almost identical hybridization patterns were found for these two independent P1 clones. All positive bands were purified from another agarose gel and subcloned into *Xho*I-cut Bluescript SK(+) plasmid (Stratagene, San Diego). These inserts were sequenced using an ABI automated sequencer. The contig for these subclones was established by PCR analyses using various primer-pairing strategies, together with the help of the *ymp*-1 and *ymp*-2 cDNA sequences.

Reverse Transcription Polymerase Chain Reaction

Poly (A) RNAs extracted from *D. melanogaster* whole heads, thoraces (male), abdomen (female), abdomen (male), eyes, brains, proboscis, gut, testis, and muscle were used for reverse transcription polymerase chain reaction (RT-PCR) in order to detect expression patterns of *ymp*-1 and *ymp*-2. PCR with *gapdh2* primers was used to provide an internal control for normalizing the cDNA concentration. The primers in the PCR reactions, at a concentration of 8 μM, are A691-internal/CD-3 (5′-TCCTGCAGTGAGAGCATAGA-3′) for *ymp*-1 and A691-internal (5′-TAGATGATGATCCTTGTGTG-3′)/Y43-4 (5′-CGGATTCGAAACCTCAAGGC-3′) for *ymp*-2. The expression of the *gapdh2* gene encoding glyceraldehyde-3-phosphate dehydrogenase-2 was chosen as an internal control because of its stable expression in various tissues (Tso, Sun, and Wu 1985). The primers for amplifying *gapdh2* that were added into the same PCR reactions used to amplify *ymp*-1 or *ymp*-2 were JCT.L (5′-CAAGCAAGCCGATAGATAAAC-3′) and t11.R (5′-GTCAAATCGACCACGGAAA-3′) at a concentration of 8 μM. The oligo JCT.L was designed to span an intron in order to rule out PCR amplification of genomic DNA. The detailed procedures for microdissecting flies, extracting RNA, synthesizing cDNA, and normalizing cDNA concentrations are in Alvarez, Robison, and Gilbert (1996).

Sequence Analyses

DNA alignments were conducted using the GeneJockeyII program package (BIOSOFT). The virtual translation of DNA sequences into protein sequences and alignment of the translated protein sequences were also carried out with the GeneJockeyII package. DNA and protein sequence similarity searches were conducted through the NCBI web site of the National Institutes of Health (http://www.ncbi.nlm.nih.gov). Estimation of synonymous substitution rates (K_s) and nonsynonymous substitution rates (K_a) and a test of deviation of the ratio K_a/K_s from unity were carried out using the *K*-estimator proposed by Comeron (1999).



Results

Gene Structure of the *ymp* Locus

Using the 5′ portion of *jgw* as a probe, we identified 77 positive plaques from a total 300,000 pfu of the *D. melanogaster* cDNA library. Among them, we identified two distinct classes of transcripts, *ymp*-1 and *ymp*-2. We also obtained the *ymp*-1 homologous sequence from the screening of a *D. yakuba* cDNA library and the *ymp*-2 homologous sequence of *D teissieri* by sequencing a *ymp*-positive phage clone identified from the *D. teissieri* genomic DNA library that we constructed. The cDNA sequences are shown in figure 1*b–d*. Subcloning and sequencing of the *D. melanogaster* P1 clones showed a complex genomic structure for these genes (fig. 1*a*).

We found that the two transcripts, *ymp*-1 and *ymp*-2, are transcribed from the same locus through alternative splicing. Putative protein sequences predicted from the cDNA sequence are 147 aa (amino acids) long for *ymp*-1 and 139 aa long for *ymp*-2, respectively. They share three small exons (totaling 174 bp coding sequence) that are similar to those recruited by the *Adh* retrosequence in the *jgw* gene. There are four more downstream exons in *ymp*-1 whose total length is 572 bp, and five more in *ymp*-2 whose total length is 986 bp. The intron fragment separating *ymp*-1 and *ymp*-2 exons is only 99 nt long. No standard donor splicing site is present at the 5′ side of this fragment, but at the 3′ side of this intron there is a receptor splicing site. We saw two adenylation sites at the end of *ymp*-1 and *ymp*-2, respectively. The cDNA sequences, splicing sites, and poly (A) signals showed that all four 3′ exons of *ymp*-1 are spliced out from the longer transcript terminated at the distal adenylation site, resulting in the *ymp*-2 transcript.

Strikingly, we found a well-characterized gene, *msi*, located in the big intron (14.9 kb) which separates the three small homologous exons from the downstream exons of *ymp*-1 and *ymp*-2. The *msi* gene is about 7.6 kb long, has two introns, and encodes a neural RNA-binding protein which is required for the development of adult external sensory organs (Nakamura et al. 1994). This gene is located on the DNA strand opposite the sense strand that encodes *ymp* genes (fig. 1*a*).

From these gene structure data, it appears that with two introns (3 and 7) separating three distinct exon groups of *ymp*, three novel proteins originated by recombination of these exon groups and the *Adh* retrosequence (fig. 2). That intron 3 also harbors a developmentally important gene indicates a unique role of introns in the evolution of genes. The following analyses will further show that the proteins YMP-1 and YMP-2 are not functionless.

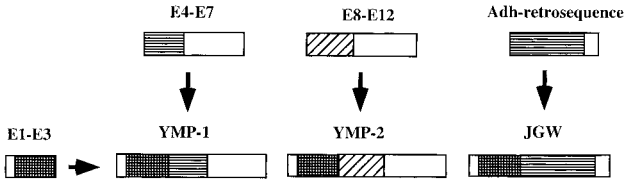


FIG. 2.—Origin of the three novel proteins, YMP-1, YMP-2, and JGW, as a consequence of exon recombination. E1–E12 represent exons 1–12 of *ymp* (for the origin of YMP-1 and YMP-2); E1–E3, making up JGW, are the first three exons of *ynd* (Long and Langley 1993), a duplicate copy of *ymp*. The hatched boxes are the regions encoding protein sequences, with different patterns showing different peptide sequences. The open boxes represent untranslated regions (UTRs) of mRNAs (the open boxes on the left represent 5′ UTRs; those on the right represent 3′ UTRs).

FISH visualized the *ymp* locus at 96E in chromosome 3R of *D. melanogaster* and in the corresponding chromosome regions in *D. yakuba* and *D. teissieri* (fig. 3). All three probes (the first three upstream homologous exons, with exons 4–7 forming the 3′ domain of *ymp*-1 and exons 8–12 forming the 3′ portion of *ymp*-2), hybridized to the same position. No signal was observed in other chromosome regions. It should be noted that we observed no extra signals in other chromosome regions in *D. yakuba* and *D. teissieri* (fig. 3*b–d*), although there are two copies of the three homologous exons, one in *ymp* and the other in *jgw* in their genomes. Thus, the position of the *ymp* locus must be close to *jgw*.

Expression Pattern

Tissue-specific RT-PCR shows that both *ymp*-1 and *ymp*-2 are expressed in *D. melanogaster* testis (fig. 4). The internal control cDNA, *gapdh2*, was amplified from every body part and tissue except the testis, where the signal of *gapdh2* is either weaker (the *ymp*-1 gel) or absent (the *ymp*-2 gel). This suggests that either the total cDNA amount in the testis used in the PCR was smaller or the concentration of *gapdh2* cDNA in the testis was low. The strong signals for the testis, however, show a higher concentration of *ymp*-1 and *ymp*-2 transcripts.

Sequence Analysis and Functional Constraint of *ymp* Proteins

A database search found no sequence homologous to either *ymp*-1 or *ymp*-2 at the DNA or protein level. There was no significant similarity between the downstream coding sequences of *ymp*-1 and *ymp*-2 (the regions outside of the first three shared exons). We cal-

←

FIG. 1.—*a*, Structure of the *ymp* locus. Two polyadenylation signals are present. The shared three 5′ exons and the *ymp*-1 downstream exons are indicated with black boxes. The five *ymp*-2 downstream exons are depicted by striped boxes. Coding regions of *msi* are indicated with blank boxes. The intron-exon structures of *ymp*-1 and *ymp*-2 are deduced from comparisons of the genomic sequence with the cDNA sequences of *ymp*-1 and *ymp*-2. *Msi* structure is from (Nakamura et al. 1994). *b*, The three shared 5′ exon cDNA sequence and deduced amino acid sequence. *c*, The downstream cDNA sequences of *Drosophila melanogaster* and *Drosophila yakuba ymp*-1 genes and deduced amino acid sequences. *d*, Downstream cDNA sequence of *D. melanogaster ymp*-2 and deduced protein sequence. In *b–d*, the filled triangles indicate positions and lengths of introns in *D. melanogaster*. The stop codons are indicated by asterisks.

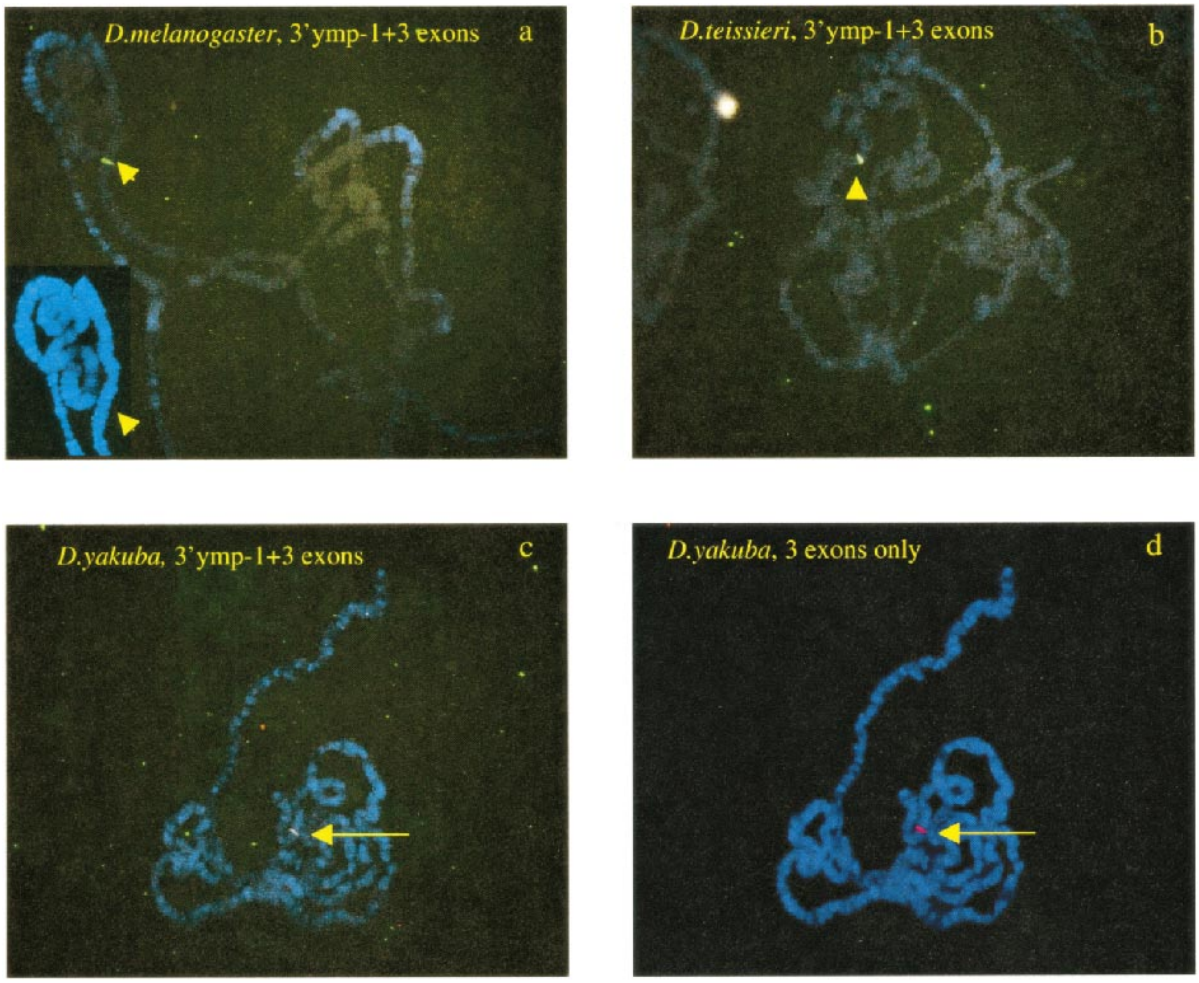


FIG. 3.—FISH results. *a*, *Drosophila melanogaster* and *b*, *Drosophila yakuba* polytene chromosomes hybridized with the DIG-labeled 5′ first three *ymp* exons (3 exons) as probe and the biotin-labeled *ymp*-1 downstream exons (3′*ymp*-1 exons) simultaneously; the red and green signals are overlapped. *c*, *Drosophila teissieri* polytene chromosome hybridized with the DIG-labeled 5′ first three *ymp* exons as probe and the biotin-labeled *ymp*-2 downstream exons as probes; again, signals are overlapped. *d*, The same *D. yakuba* polytene chromosome as in *c*, but only the red signal (the shared first three 5′ exons) is visualized.

culated synonymous and nonsynonymous substitutions between the *D. melanogaster* and the *D. yakuba ymp*-1 cDNA downstream coding sequences, and also between the *D. melanogaster* and the *D. teissieri ymp*-2 downstream coding sequences, as summarized in table 1.

Discussion

The origin of new genes includes two processes: the initial molecular assembly events and the subsequent population genetics. A processed retrosequence of the *Adh* gene is part of a young functional gene, *jgw* (Long and Langley 1993). The *Adh*-derived sequence was combined with three upstream exons about 2.5 MYA in the *yakuba–teissieri* lineage. Recently, Long, Wang, and Zhang (1999) demonstrated that there is another gene, dubbed *ymp*, containing the same structure as the recruited portion of *jgw*, which must have provided the donor for the exon-shuffling process that created *jgw*. However, the structure and function of *ymp*, as well as the portion of the donor gene that was involved in the shuffling process, remained unclear. The results of this study revealed that the *ymp* locus, the source of the recruited portion of *jgw*, has a remarkably complex gene structure.

The *ymp* locus produces two mRNAs, *ymp*-1 and *ymp*-2, resulting from the use of two adenylation sites and an alternative splicing process. The between-species comparison indicated significantly lower nonsynonymous substitution rates than synonymous substitution rates in the coding sequences of each protein (table 1), suggesting an evolutionary constraint on protein sequence typical of functional genes. These two transcripts share the three 5′ exons that are highly similar to the recruited portion of *jgw*, suggesting that the recruited portion of *jgw* arose from a duplication event of the *ymp* gene (Long, Wang, and Zhang 1999). Both *ymp*-1 and *ymp*-2 are specifically expressed in testes (fig. 3), suggesting that their functions may be related to reproduction. Thus, the interesting fact that *jgw* is specifically expressed in adult male *D. teissieri* is likely a consequence of a similar regulatory sequence inherited by the *jgw* gene from the *ymp* locus. It is remarkable that a

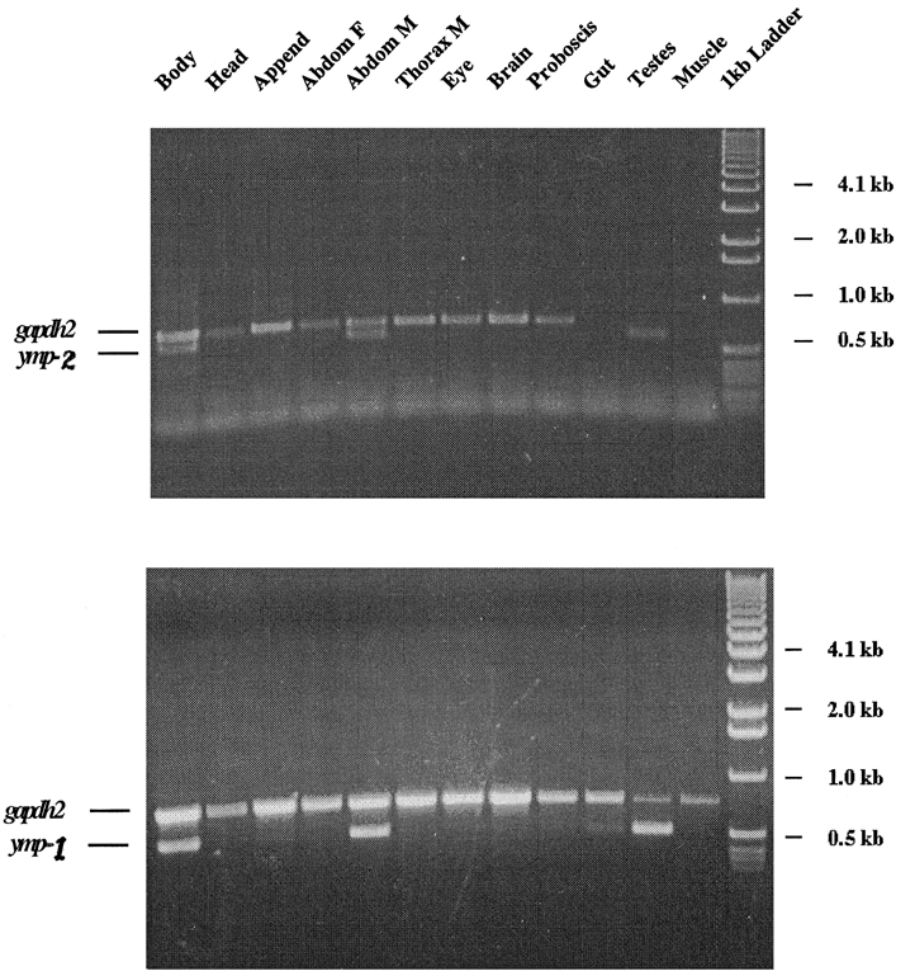


FIG. 4.—Tissue-specific RNA expressions of *ymp*-1 and *ymp*-2 detected by RT-PCR experiments. Gapdh2 is the amplification of cDNA of *gapdh2* as an internal control.

sibling species of *D. teissieri*, *D. yakuba*, which has been separated for a short time (2.5 Myr), evolved a different expression pattern in which the transcripts are also present in other developmental stages.

It becomes clear from this investigation and previous analysis (Long, Wang, and Zhang 1999) that the first three exons of the *ymp* locus are a donor for the recruited portion of *jgw*. Considering the hydrophobicity of the N-terminal peptide in JGW, YMP-1, and YMP-2 (Long, Wang, and Zhang 1999), the three small exons may encode a signal peptide, although this needs to be experimentally confirmed. Because there is no reported signal peptide homologous to this peptide, the target cellular membrane location of this signal peptide is unknown. YMP-1 and YMP-2 probably carry out different functions, since their sequences are not similar at the C-terminal ends. This feature, together with the shared promoter, makes the *ymp* locus different from other loci with multiple adenylation sites or alternative splicing, which usually produce isoforms with somewhat similar domains (with the exception of the unc-17/cha-1 locus [Alfonso et al. 1994]; the unc-17/cha-1 locus encodes two alternative forms, one of which contains only a noncoding first exon).

Table 1
Nonsynonymous Substitution Rates (K_a), Synonymous Substitution Rates (K_s), and K_a/K_s Ratios Between Different Species' *ymp*-1 and *ymp*-2 Downstream Coding Sequences

	K_a	K_s	K_a/K_s
melanogaster ymp-1 3′ vs. *yakuba ymp*-1 3′	0.050	0.243	0.206*
melanogaster ymp-2 3′ vs. *teissieri ymp*-2 3′	0.068	0.320	0.213*

* Significantly lower than unity, $P < 0.0001$.

The *ymp* locus is further complicated by the presence of the *msi* gene, nested in intron 3, which separates the first three exons from the rest of the downstream exons of *ymp* (fig. 1*a*). The nested structure of the *ymp* locus shows two unique features that differ from nested genes previously reported. The intronic *msi* is 7.6 kb long, making it the longest nested gene identified so far. The other nested genes are usually around 1 kb long (Henikoff et al. 1986; Chen et al. 1987; Furia et al. 1990, 1993; Levinson et al. 1990; Neufeld, Carthew, and Rubin 1991; McBabb, Greig, and Davis 1996; Valleix et al. 1999). Moreover, like the first reported nested cuticle

gene in the *Gart* locus (Henikoff et al. 1986), the *msi* gene is located on the strand opposite its host gene. Simultaneous transcription of both strands may lead to RNA interference (O'Hare 1986; Sharp 1999), as two recent experiments on *D. melanogaster* showed (Kennerdell and Carthew 1998; Misquitta and Paterson 1999). In the *ymp* locus, this interference, if any, may be avoided by a spatially differential expression of the *ymp* gene and the *msi* gene. The *msi* gene is expressed in sensilla (Nakamura et al. 1994), while the expression of the *ymp* gene is restricted to the testis. In the GART locus, however, simultaneous transcription of the purine gene and the intronic gene seems possible (Henikoff et al. 1986). Nested genes may not be uncommon gene structures. In a survey of a genomic region surrounding *Adh* gene of 2.9 Mb in *D. melanogaster,* Ashburner et al. (1999) identified 17 nested coding regions (CDS) using computer programs for gene prediction, although all of them except *Adh* and *Adh-r* have yet to be confirmed experimentally. How typical the different cases represented by GART and *ymp* are in their structures and their expression patterns and how nested genes are related to transcriptional interference are questions that remain to be clarified with further experimental data.

The *ymp* locus encompasses three phenomena pointing to an important role for introns: (1) an event of exon shuffling involving 5′ exons, (2) a long nested gene within an intron, and (3) alternative transcription termination associated with alternative splicing. A single locus combining this set of molecular properties has not previously been reported. This finding may add to the classical concept of genes, which has been modified with the discoveries of operons, introns, overlapping genes, alternative splicing, multiple polyadenylation sites, complex promoters, and nested genes. The complex structure and evolutionary history of *ymp* indicate the importance of introns in the origin of new genes, as the exon theory of genes has suggested (Gilbert 1978, 1989). Indeed, introns 3 and 7 in *ymp* facilitate the recombination of several exon groups and *Adh* retrosequences that led to the origin of three proteins (fig. 2). Meanwhile, the complexity of gene structure, as shown in the *ymp* locus, ought to be an important factor to consider in genomic research, such as the prediction of genes from genome data. In fact, the complex gene structure of the *ymp* locus, as described in this report, was not predicted from the genome sequences of *D. melanogaster* (Adams et al. 2000).

Acknowledgments

We thank Walter Gilbert of Harvard for support and discussion; Bruce Hamilton of MIT for his generous gift of *D. melanogaster* cDNA libraries; the laboratory of Spyros Artavanis-Tsakonas of Yale for maintaining and delivering the P1 clones of *D. melanogaster*; and Janice Spofford, A. Hon-Tsen Yu, and members of M.L.'s laboratory for critical reading and discussion of the manuscript. We also thank Josep Comeron for his *K*-estimator program. This project was supported by a Packard Fellowship in Science and Engineering and a grant from National Science Foundation to M.L.

LITERATURE CITED

ADAMS, M. D., S. E. CELNIKER, R. A. HOLT et al. (195 co-authors). 2000. The genome sequence of Drosophila melanogaster. Science **287**:2185–2195.

ALFONSO, A., K. GRUNDAHL, J. R. MCMANUS, J. M. ASBURY, and J. B. RAND. 1994. Alternative splicing leads to two cholinergic proteins in Caenorhabditis elegans. J. Mol. Biol. **241**:627–630.

ALVAREZ, C. E., K. ROBISON, and W. GILBERT. 1996. Novel Gq alpha isoform is a candidate transducer of rhodopsin signaling in a *Drosophila* testes-autonomous pacemaker. Proc. Natl. Acad. Sci. USA **93**:12278–12282.

ASHBURNER, M. 1989. *Drosophila,* a laboratory manual. Cold Spring Harbor Laboratory Press, New York.

ASHBURNER, M., S. MISRA, J. ROOTE et al. (25 co-authors). 1999. An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: the *Adh* region. Genetics **153**:179–219.

BEGUN, D. J. 1997. Origin and evolution of a new gene descended from *alcohol dehydrogenase* in *Drosophila.* Genetics **145**:375–382.

BROSIUS, J., and S. J. GOULD. 1992. On ''genomenclature'': a comprehensive (and respectful) taxonomy for pseudogenes and other ''junk DNA''. Proc. Natl. Acad. Sci. USA **89**: 10706–10710.

CHEN, C., T. MALONE, T. BECKENDORF, and R. L. DAVIS. 1987. At least two genes reside within a large intron of the dunce gene of *Drosophila.* Nature **329**:721–724.

CHEN, L. B., A. L. DEVRIES, and C. H. C. CHENG. 1997. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. USA **94**:3811–3816.

COMERON, J. M. 1999. K-estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics **15**:763–764.

FURIA, M., P. P. D'AVINO, S. CRISPI, D. ARTIACO, and L. C. POLITO. 1993. Dense cluster of genes is located at the ecdysone-regulated 3C puff of *Drosophila melanogaster*. J. Mol. Biol. **231**:531–538.

FURIA, M., F. A. DIGILIO, D. ARTIACO, E. GIORDANA, and L. C. POLITO. 1990. A new gene nested within the dunce genetic unit of *Drosophila melanogaster*. Nucleic Acids Res. **18**:5837–5841.

GILBERT, W. 1978. Why gene in pieces? Nature **271**:501.

———. 1989. The exon theory of genes. Cold Spring Harb. Symp. Quant. Biol. **52**:901–905.

HARTL, D. L., D. I. NURMINSKY, R. W. JONES, and E. R. LOZOVSKAYA. 1994. Genome structure and evolution in Drosophila: applications of the framework P1 map. Proc. Natl. Acad. Sci. USA **91**:6824–6829.

HENIKOFF, S., M. A. KEENE, K. FECHTEL, and J. W. FRISTROM. 1986. Gene within a gene: nested *Drosophila* genes encode unrelated proteins on opposite strands. Cell **44**:33–42.

JEFFS, P., and M. ASHBURNER. 1991. Processed pseudogenes in *Drosophila.* Proc. R. Soc. Lond. B **244**:151–159.

KENNERDELL, J. R., and R. W. CARTHEW. 1998. Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway. Cell **95**:1017–1026.

LANGLEY, C. H., E. MONTGOMERY, and W. F. QUATTLEBAUM. 1982. Restriction map variation in the Adh region of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **79**:5631–5635.

LEVINSON, B., S. KENWRICK, D. LAKICH, G. HAMMONDS, and J. GITSCHIER. 1990. A transcribed gene in an intron of the human factor VIII gene. Genomics **7**:1–11.

LONG, M., S. J. DE SOUZA, C. ROSENBERG, and W. GILBERT. 1996. Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. Proc. Natl. Acad. Sci. USA **93**:7727–7731.

LONG, M., and C. H. LANGLEY. 1993. Natural selection and origin of jingwei-a chimeric processed functional gene. Science **260**:91–95.

LONG, M., W. WANG, and J. ZHANG. 1999. Origin of new genes and source for N-terminal domain of the chimerical gene, *jingwei*, in *Drosophila*. Gene **238**:135–142.

MCBABB, S., S. GREIG, and T. DAVIS. 1996. The alcohol dehydrogenase gene is nested in the outspread locus of *Drosophila melanogaster*. Genetics **143**:897–911.

MARTIGNETTI, J. A., and J. BROSIUS. 1993. BC2000 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. Proc. Natl. Acad. Sci. USA **90**:11563–11567.

MISQUITTA, L, and B. M. PATERSON. 1999. Targeted disruption of gene function in *Drosophila* by RNA interference (RNA-i): a role for nautilus in embryonic somatic muscle formation. Proc. Natl. Acad. Sci. USA **96**:1451–1456.

NAKAMURA, M., H. OKANO, J. A. BLENDY, and C. MONTELL. 1994. Musashi, a neural RNA-binding protein required for *Drosophila* adult external sensory organ development. Neuron **13**:67–81.

NEUFELD, T. P., R. W. CARTHEW, and G. M. RUBIN. 1991. Evolution of gene position: chromosomal arrangement and sequence comparison of the *Drosophila melanogaster* and *Drosophila virilis* sina and Rh4 genes. Proc. Natl. Acad. Sci. USA **88**:10203–10207.

NURMINSKY, D. I., M. V. NURMINSKAYA, D. DE AGUIAR, and D. L. HARTL. 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. Nature **396**:572–575.

O'HARE, K. 1986. Genes within genes. Trends Genet. **2**:33.

OHTA, T. 1994. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics **138**:1331–1337.

PATTHY, L. 1995. Protein evolution by exon-shuffling. Springer-Verlag, New York.

SAMBROOK, J., E. FRITSCH, and T. MANIATIS. 1989. Molecular cloning—a laboratory manual. 2nd edition. Cold Spring Harbor Laboratory Press, New York.

SHARP, P. A. 1999. RNAi and double-strand RNA. Genes Dev. **13**:139–141.

TSO, J. Y., X.-H. SUN, and R. WU. 1985. Structure of two unlinked Drosophila melanogaster glyceraldehyde-3-phosphate dehydrogenase genes. J. Biol. Chem. **260**:8220–8228.

VALLEIX, S., J.-C. JEANNY, S. ELSEVIER, R. L. JOSHI, P. FAYET, D. BUCCHINI, and M. DELPECH. 1999. Expression of human F8B, a gene nested within the coagulation factor VIII gene, produces multiple eye defects and developmental alterations in chimeric and transgenic mice. Hum. Mol. Genet. **8**:1291–1301.

WIEGANT, J. 1996. Nonradioactive in situ hybridization application manual. 2nd edition. Boehringer, Mannheim, Germany.

EDWARD HOLMES, reviewing editor

Accepted May 16, 2000