ELSEVIER

Gene 238 (1999) 135–141

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Origin of new genes and source for N-terminal domain of the chimerical gene, *jingwei*, in *Drosophila* ☆

Manyuan Long *, Wen Wang, Jianming Zhang

Department of Ecology and Evolution, The University of Chicago, 1101 East 57th Street, Chicago, IL 60637, USA

Received 15 February 1999; received in revised form 4 May 1999; accepted 1 June 1999; Received by G. Bernardi

Abstract

This paper deals with a general question posed by the origin of new processed chimerical genes: when a new retrosequence inserts into a new genome position, how does it become activated and acquire novel protein function by recruiting new functional domains and regulatory elements? *Jingwei* (*jgw*), a newly evolved functional gene with a chimerical structure in *Drosophila*, provides an opportunity to examine such questions. The source of its exon encoding C-terminal peptide has been identified as an *Adh* retrosequence, which extends the concept of exon shuffling from recombination to retroposition as a general molecular mechanism for the origin of a new gene. However, the origin of 5′ exons remains unclear. We examined two hypotheses concerning the origin of these non-*Adh*-derived *jgw* exons: (i) these exons might originate from a unique genomic sequence that fortuitously evolved a standard intron–exon structure and regulatory sequence for *jgw*; (ii) these exons might be a duplicate of an unrelated previously existing gene. Genomic Southern analysis, in conjunction with construction and screening of a genomic bookshelf (sub-library), was conducted in a group of *Drosophila* species. The results demonstrated that there are duplicate genes containing the same structure as the recruited portion of *jgw*. We name this duplicate gene in *Drosophila teissieri* and *Drosophila yakuba* and its orthologous gene in *Drosophila melanogaster* as *yellow-emperor* (*ymp*). Thus, the 5′ exons/introns originated from a previously existing gene that provided new modules with specific sub-function to create *jgw*.

Keywords: Chimerical genes; Jgw; New gene evolution; Retroposition; Ymp; Ynd

1. Introduction

How new genes with novel functions originate is a fundamentally important but poorly understood evolutionary problem. The recent success in sequencing whole genomes clearly shows that organisms vary in the number and type of genes they possess (e.g., the bacterial genomes sequenced by Fraser et al., 1995 and Himmelreich et al., 1996). The roles of those lineage-specific genes in evolution, and the way those genes originated, are largely unexplained. Yet, any gene in an organism at some remote time had an origin and a period early in its evolution when it acquired new functions. The actual picture for this exciting early stage of genes is largely unknown. To study the origin of a gene in detail requires the discovery of a young gene, and in particular one that has retained significant features of its early stages. If a gene is old, signals of its early evolution will have been obscured by noise from later evolutionary processes. Furthermore, duplicated genes and shuffled exons, which may lead to the origin of new genes, are usually associated with rapid sequence evolution (Long et al., 1996; Ohta, 1994). This feature of gene duplication reinforces the need to study young genes.

The *jingwei* (*jgw*) gene in *Drosophila* provides an opportunity to investigate the early stage of evolution of genes, because of its young age and specific gene structure. *Jgw* exists only in two *Drosophila* sibling species, *Drosophila teissieri* and *Drosophila yakuba*, which were separated less than 2.5 million years ago (Lachaise et al., 1988). A portion of this gene was first observed in *D. teissieri* and *yakuba* by Langley et al. (1982) and cloned by Jeffs and Ashburner (1991). This portion of *jgw* was found to be a retrosequence from the gene encoding alcohol dehydrogenase (*Adh*) with all

Abbreviations: Adh, alcohol dehydrogenase; NIB, nuclear isolation buffer; PCR, polymerase chain reaction; ymp, yellow-emperor; ynd, yande.

☆ Presented at the International Society of Molecular Evolution Meeting, Puntarenas, Costa Rica, 11–16 January 1999.

* Corresponding author. Tel.: +1-773-7020557; fax: +1-773-7029740.

E-mail address: mlong@midway.uchicago.edu (M. Long)

0378-1119/99/$ – see front matter
PII: S0378-1119(99)00229-2

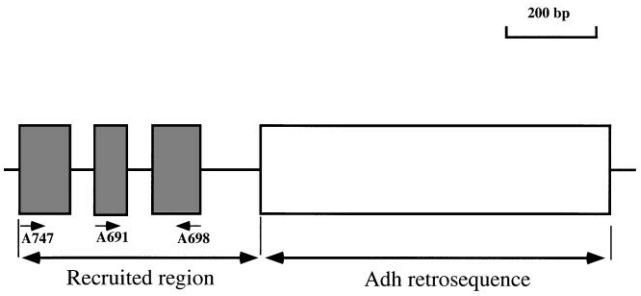


Fig. 1. The chimerical structure of the *jgw* gene in *D. teissieri* and *yakuba*. The grey region flanked by two oligonucleotide primers, A747 and A698, is the recruited exons.

three original introns lost (Jeffs and Ashburner, 1991), suggesting an important role of retroposition in the origin of new genes as discussed by Brosius (1991). In a survey of population genetic variation of *jgw* DNA sequences in *D. teissieri* and *yakuba*, Long and Langley (1993) found that most nucleotide polymorphisms are synonymous changes, suggesting a strong purifying selection acting at the protein sequence level and, thus, *jgw* is more likely a functional gene. The further identification of mRNA transcribed from *jgw* and its specific expression patterns are consistent with the conclusion that this newly originated genetic element is a functional gene, a notion different from a previous conclusion that it was a processed pseudogene (Jeffs and Ashburner, 1991). We also found that the origin of *jgw* is coincident with about the same period of the speciation of *D. teissieri* and *yakuba*. This revealed a young age of *jgw*, less than 2.5 million years, making it possible to directly observe the early evolution of *jgw* in protein sequence, a process driven by strong adaptive evolution.

Furthermore, we found that the previously observed DNA sequence does not represent a complete gene. Instead, the newly inserted *Adh* retrosequence recruited a group of upstream exons and introns into its transcription unit, which created a novel gene with chimerical structure. Fig. 1 is a sketch of this chimerical gene: three exons with small introns are joined with a fused large exon that originated from the *Adh* gene. Such a gene structure would give rise to a general question pertinent to the origin of chimerical processed genes: when a retrosequence lands on a new position of genome, how does it acquire new functional domains and new regulatory elements?

Although the origin of the *Adh*-derived exon of *jgw* is clear, the source for the recruited portion of *jgw* has not been certain. Two possible scenarios concerning the origin of these recruited *jgw* exons need to be investigated. (i) These exons might fortuitously originate from a unique non-coding genomic sequence, a possibility similar to the origin of two new RNA genes in mammals. The first gene encoding BC1 RNA in rodents, recruited a unique genomic sequence after the insertion of a tRNA retrosequence (Martignetti and Brosius, 1993a; Brosius and Gould, 1992); the second gene encoding BC200 RNA in primates recruited a unique region after the 7SL RNA-derived monomeric Alu element inserted into the genome (Martignetti and Brosius, 1993b; Brosius and Gould, 1992). (ii) Alternatively, the three 5′ *jgw* exons might be from an unrelated pre-existing gene. In this paper we will show that the second scenario is true, supporting the concept of exon shuffling (Gilbert, 1978), a mechanism for new gene evolution.

2. Materials and methods

2.1. Extraction of genomic DNAs

Genomic DNAs were extracted from 200 to 300 adult flies of *D. yakuba* (strain Y5), *D. teissieri* (T7), and *Drosophila melanogaster* (strain MK47) (for the sources of these strains, see Long and Langley, 1993; Richter et al., 1997). The anesthetized flies were homogenized in a 10 ml Wheaton tissue homogenizer with 5 ml of cold Nuclear Isolation Buffer (NIB) containing 10 mM Tris–HCl, 60 mM NaCl, 10 mM EDTA, 0.15 mM Spermidine, 0.15 mM Spermine, and 0.5% Triton X-100. The debris was isolated from the homogenate by brief centrifugation, and discarded. The nuclear pellets were washed several times using NIB by centrifugation before the pellets were resuspended in 4 ml NIB. 500 μg of proteinase K and 500 μl of 10% SDS were added to the

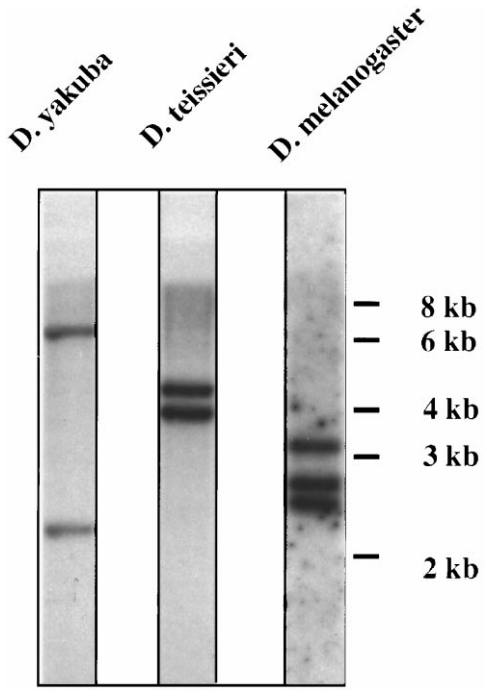


Fig. 2. The genomic Southern blots for the three *Drosophila* species, hybridized with the probe of the recruited portion of the *jgw* gene.

nuclear resuspension. The resuspension was incubated at 37°C for 50 min. DNAs were extracted using phenol-chloroform and were resuspended in 1 × TE buffer.

2.2. Southern analysis

One microgram of genomic DNAs was digested with 5–10 units of restriction enzyme *Eco*RI. We purposely chose this enzyme because there is no restriction site for *Eco*RI in the region of the DNAs we used as probe. The Southern transfer procedures (Southern, 1975) followed Sambrook et al. (1989). Hybond-N membrane (NEN Life Science) was used as transfer membrane.

The PCR product of a 314 bp DNA fragment of the recruited region of *jgw* was used as probe, which was amplified from *D. teisseiri* genomic DNA using primers A747 and A698 (for the oligonucleotide sequences, see Section 2.3) (Fig. 1). The probe was labeled using the random-priming method with the BRL kit and labeled by ^{32}P.

2.3. Construction of genomic bookshelves

20–30 μg of genomic DNA of *D. teissieri* (T7) and *D. melanogaster* (MK47) were digested with 200–400 units of *Eco*RI, and were electrophoresed in 0.8% agarose gel. The gel containing DNA of the expected sizes, as detected by Southern analysis, was sliced using a razor blade. The DNA (approx. 1 μg) was isolated and purified using GeneClean kit (Bio101).

Genomic bookshelves were constructed using a lambda ZAP II vector kit (Stratagene). The isolated genomic DNAs were ligated to the vector lambda Zap II and packaged with extract. These bookshelves (the sub-genomic phage libraries contained 10^7 pfu) were screened using the same probe as that used in the Southern analysis (see Section 2.2) for the phage clone that contains the Southern-detected genomic DNA fragments. Plasmid clones were made by in vivo excision of the positive phage clones following the protocol of the kit. The inserts in these plasmid clones were sequenced

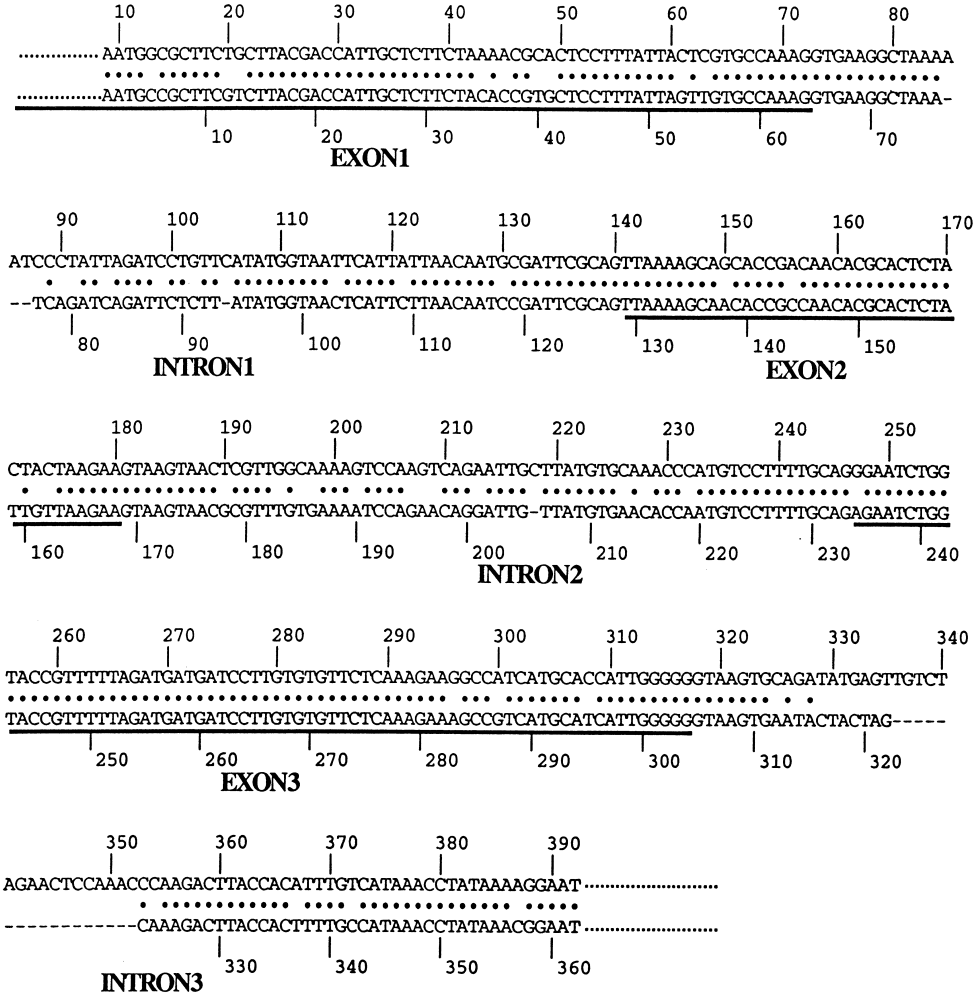
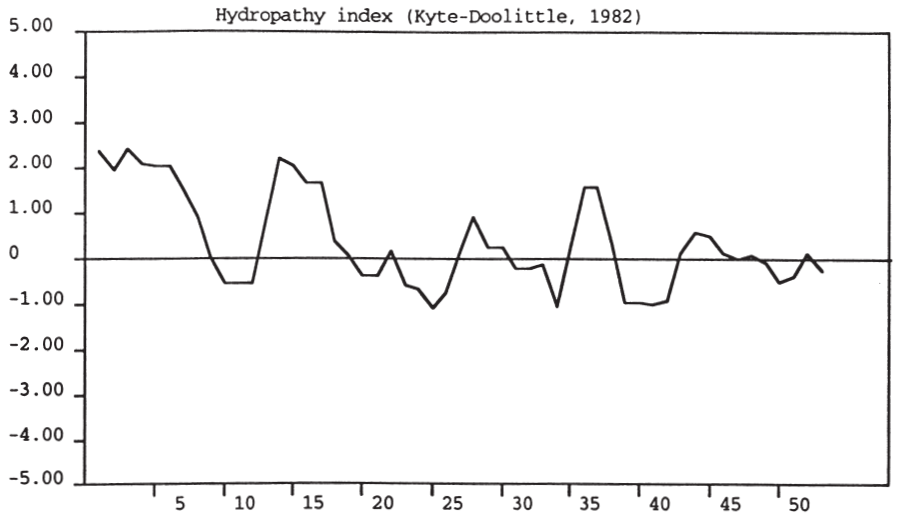


Fig. 3. The aligned sequences, similar to the recruited portion of *jgw*, of newly identified genes: the *ymp* gene in *D. melanogaster* (below), and the *ymp* gene in *D. teissieri* (top). The underlined regions indicate exons.

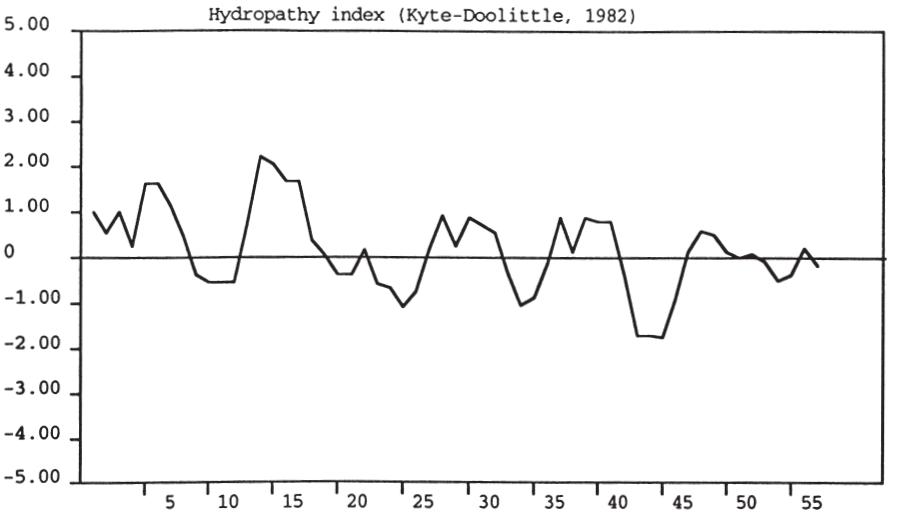
A

```
D.teis_jgw  MALRLTTITLLKRTPLLLVPKLKAAPTTRTLLLRSPFPGNLVPFSDDDPCVFSKKAIMHHGG
D.teis_ymp  MALLLTTIALLKRTPLLLVPKLKAAPTTRTLLLRR----NLVPFLDDDPCVFSKKAIMHHWG
D.mela_ymp  MPLRLTTIALLHRAPLLVVPKLKATPPTRTLLLRK----NLVPFLDDDPCVFSKKAVMHHWG
            •|• ••••|•• •|•••|••••••|• •••••••     ••••• ••••••••••|••• •
```

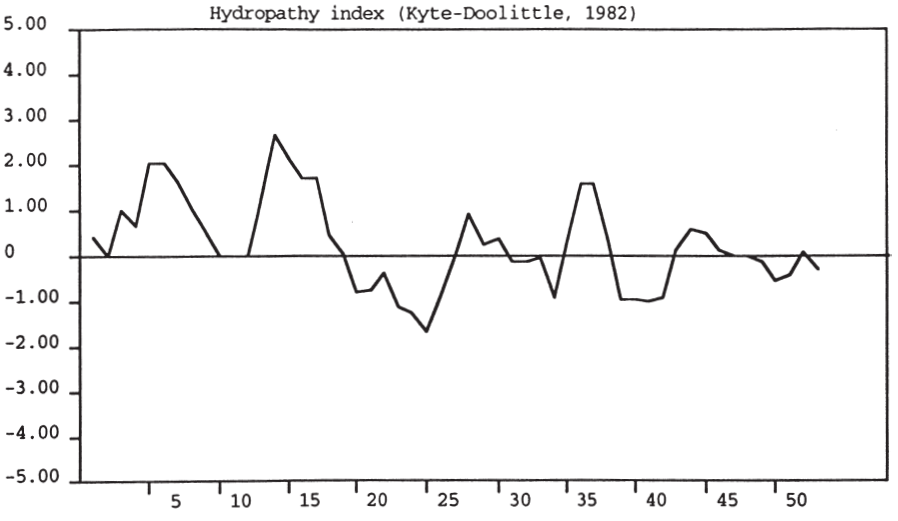
B



C



D



using a sequencing kit (United States Biochemical), with the following sequencing primers:

A747: 5′-CCAATTTGTTATAATGGCGCTTCG-3′;
A691: 5′-AAGCAGCACCGACAACACGCAC-3′;
A698: 5′-CATGGTGCATGATGGCCTTC-3′.

These primers were designed for sequencing double strands of the recruited region of *jgw*: A747 and A691 are for the sense strand, A698 for the antisense strand.

2.4. Computer analysis

The assembling and alignments of the DNA sequences of inserts from the clones isolated from the Bookshelves were done using the GeneJockeyII program package (BIOSOFT). The translation of DNA sequences to protein sequences and analysis of general properties and multiple alignment of the translated protein sequences were also carried out with the GeneJockeyII package. The BLAST search of GenBank was conducted in the web-site of NCBI of the National Institute of Health USA (www.ncbi.nlm.nih.gov).

3. Results

3.1. Genomic Southern analysis

Genomic Southern analysis indicated that there is duplicate copy paralogous to the recruited portion of *jgw*. Fig. 2 shows the Southern blot hybridized with the probe synthesized using the recruited portion of *jgw*. The restriction pattern using *Eco*RI generated two distinct bands. Because there is no *Eco*RI site in the recruited sequence of *jgw*, the two bands should be interpreted as representing duplicate copies paralogous to the template sequences, which was confirmed by sequencing of the DNA bands. This provides evidence to distinguish the two alternative hypotheses discussed in the Introduction: the activation of *Adh*-derived retrosequence does NOT occur because a single random DNA sequence mimics a structure of regulatory sequence, but because of a previously existing duplicate gene as target of the *Adh*-retrosequence.

The recruited portion of *jgw* is not restricted to the single lineage of *D. teissieri* and *yakuba*, like *jgw*. Fig. 2 shows that the *D. melanogaster* lane contains three bands with sizes of 2.5 kb, 2.7 kb, and 3.4 kb. Given the size of the probe (about 314 bp), the three bands at least represent two distinct genes except in some rare events, e.g., one single gene in *D. melanogaster* could contain large introns in this species, while those introns in *jgw* evolve to a smaller size.

We named this newly identified gene in *D. teissieri* and *yakuba* as *yellow-emperor* (*ymp*), a brother of the emperor yande in the Chinese legend that was used to name *jgw*, following Long and Langley (1993). The original suggestion to name this duplicate gene in *D. teissieri* and *yakuba* as *yande* (*ynd*) (Long and Langley, 1993) did not anticipate such a complex history, although *ynd* is indeed a gene in *D. melanogaster* structurally related to *jgw*. We also found that the phylogenetic distribution of this recruited portion of genes can be extended to several *Drosophila* subgroups, such as *Drosophila suzhuki* subgroup (>20 million years) and is associated with a complex evolutionary history (Wang and Long, unpublished data).

3.2. Bookshelf screening

Three positive plaques from *D. teissieri* bookshelf and five positive plaques from *D. melanogaster* bookshelf were identified by screening 10 000 pfu and 50 000 pfu for the two species separately. (The genomic DNAs of *D. melanogaster* that include all three bands were pooled to make a single bookshelf, for technical and economical reasons.) By sequencing the inserts in these clones, we identified the duplicate genes, *ymp* in *D. teissieri* and *D. melanogaster*. These DNA sequences and their hypothetical protein sequences are shown in Figs. 3 and 4 in a form of alignment, which is highly similar to the sequence of the recruited portion of *jgw*. All three sequences contain three exons and three introns in the same positions and same phases as the homologous portion of *jgw*.

Searching GenBank using BLAST and Psi-BLAST (Altschul et al., 1990, 1997) with nr option identified no similar sequences, suggesting that the protein sequences encoded by the recruited portion of *jgw* are novel.

3.3. Hydrophobicity analysis

Fig. 4B–D also shows hydrophobicity plots, defined by the method of Kyte and Doolittle (1982), of the hypothetical peptides of *ymp* and *jgw* in *D. teissieri* and *ynd* in *D. melanogaster*, which show a hydrophobic property. The hydrophobic property at this N-terminal peptide is similar to signal peptides in many proteins or transit peptide/presequence in the organellar-evolved

Fig. 4. (A) The hypothetical peptide sequences of *ymp*, and *ynd* corresponding to the recruited exons of *jgw* (its peptide sequence is also included). The dots show the identical residues, and the bars indicate similar residues. (B–D) Kyte–Doolittle hydropathy plots for *ymp* in *D. teissieri* (b); *jgw* in *D. teissieri* (C); and *ymp* in *D. melanogaster* (D).

nuclear-encoded protein (Long et al., 1996), possibly suggesting a related function carried by these peptides.

4. Discussion

We tested two alternative hypotheses about the origin of the recruited portion of *jgw*, in relation to a particular process of molecular origin of retroposition that created the special chimerical structure of *jgw*. These hypotheses address a most general question of how a processed gene becomes activated and acquires novel function by recruiting new protein domains and regulatory elements. The two evolutionary scenarios would bring different evolutionary consequences. If the recruited portion sequence just originated from a unique non-coding sequence, the elegant chimerical structure and regulatory sequence in the recruited portion would have to be evolved fortuitously, as shown in the origins of *BC1* in rodents (Martignetti and Brosius 1993a) and *BC200* in primates (Martignetti and Brosius, 1993b). If the single region represents a previously existing single-copy gene, then the insertion of the Adh retrosequence might demolish the established function and could only be fixed in species in some special conditions. For instance, the newly evolved function can just replace or improve the original one, because the advantages of the novel composite gene may outweigh the disadvantages of altering the resident gene. However, in the case of gene duplication, the insertion of new exon(s) would create new function without having to destroy original function.

We found that the recruited portion (*ynd*) of *jgw* has a duplicate copy. The sequence of this duplicate copy (*ymp*) in *D. teissieri* indeed indicates that it is highly similar to *jgw*. These two copies, *ynd* and *ymp*, are paralogous and are also likely to contain two and three additional exons, respectively, following the identified 5′ exons. In a common ancestor of *D. teissieri* and *D. yakuba*, an *Adh* retrogene integrated into the *ynd* gene following the third exon generating the chimerical gene, *jgw*. The *jgw* gene is a novel gene, now under selective pressure in the two *Drosophila* species. *Ymp* does also transcribe RNA (Wang, Zhang, and Long, unpublished data). The combined evidence rejects the hypothesis that the recruited portion of *jgw* originates from a unique non-coding sequence, but confirms its origin from a duplicate gene of pre-existing unrelated genes. Thus, the chimerical processed gene *jgw* differs from the other new chimerical neuronal RNA genes, *BC1* and *BC200* in mammals that recruited unique genomic sequences for a new function in the nervous system.

The origin of the recruited portion of *jgw* is a case for a general mechanism for the origin of new gene, exon shuffling (Gilbert, 1978), facilitated by gene duplication. It should be noted that the molecular mechanism of exon shuffling has been extended from illegitimate recombination to retroposition. Brosius (1991) emphasized the significance of retrosequence in generating novel gene structures. *Jgw* provide a clear case for such a process. Another new gene in *D. melanogaster*, *Sdic*, found by Nurminsky et al. (1998), was also clearly created by exon shuffling but with a different molecular process that did not involve retroposition (Nurminsky et al., 1998; Capy, 1998).

The hydrophobic property of the recruited gene region of *jgw* may reflect the shift of functional property of the original *Adh* gene from cytoplasmic locations to some unknown cellular membrane-related location. In fact, the different expression patterns in *D. teissieri* and *yakuba*, as detected by Long and Langley (1993), may further manifest functional divergence in these two species. The expression pattern of *D. teissieri* is male-specific. This is consistent with the testis-specific expression patterns of its parental gene *ymp* in *D. melanogaster* revealed by RNA-PCR experiments (Long, unpublished results), suggesting some distinct biological functions of these genes, although the database search has not yet found any matches to known genes.

Acknowledgements

We thank C.H. Langley, W. Gilbert, and R.C. Lewontin for their consistent support and valuable discussions. We thank S.J. De Souza for the discussion about the possible biochemical functions of the recruited portion of *jgw*. We also thank the Packard Foundation for a Packard Fellowship in Science and Engineering and National Science Foundation, USA (to M.L.) for the study of new gene evolution.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Brosius, J., 1991. Retroposons — seeds of evolution. Science 251, 753.

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

Capy, P., 1998. A plastic genome. Nature 396, 522–523.

Fraser, C.M., et al., 1995. The minimal gene complement of *Mycoplasma*. Science 270, 397–403.

Gilbert, W., 1978. Why gene in pieces? Nature 271, 501

Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C., Herrmann, R., 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucleic Acids Res. 24, 4420–4449.

Jeffs, P., Ashburner, M., 1991. Processed pseudogene in *Drosophila*. Proc. R. Soc. Lond. B 244, 151–159.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

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

Langley, C.H., Montgomerey, E., Quattlebaum, W.F., 1982. Restriction map variation in the *Adh* region of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 79, 5631–5635.

Long, M., de Souza, S.J., Rosenberg, C., Gilbert, W., 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., Langley, C.H., 1993. Natural selection and origin of *jingwei* — a chimeric processed functional gene. Science 260, 91–95.

Martignetti, J.A., Brosius, J., 1993a. Neural *BC1* RNA as an evolutionary marker: guinea pig remains a rodent. Proc. Natl. Acad. Sci. USA 90, 9698–9702.

Martignetti, J.A., Brosius, J., 1993b. *BC200* RNA: a neural RNA polymerase III product encoded by a monomeric Alu element. Proc. Natl. Acad. Sci. USA 90, 11563–11567.

Nurminsky, D.I., Nurminskaya, W.V., De Aguiar, D., Hartl, D.L., 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. Nature 396, 572–575.

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

Richter, B., Long, M.Y., Lewontin, R.C., Nitasaka, E., 1997. Nucleotide variation and conservation at the *dpp* locus, a gene controlling early development in *Drosophila*. Genetics 145, 311–323.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning — A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Southern, E.M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503–517.