Gene duplication and speciation in *Drosophila*: Evidence from the *Odysseus* locus

Chau-Ti Ting*†‡, Shun-Chern Tsaur‡, Sha Sun§¶, William E. Browne**‡‡, Yung-Chia Chen*, Nipam H. Patel†††§, and Chung-I Wu§

*Department of Life Science, National Tsing Hua University, Hsinchu 300, Taiwan; ‡Institute of Zoology and Research Center for Biodiversity, Academia Sinica, Taipei 105, Taiwan; and Departments of Ecology and Evolution, * Molecular Genetics and Cell Biology, and ††Organismal Biology and Anatomy and Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637

Edited by Margaret G. Kidwell, University of Arizona, Tucson, AZ, and approved June 30, 2004 (received for review March 22, 2004)

The importance of gene duplication in evolution has long been recognized. Because duplicated genes are prone to diverge in function, gene duplication could plausibly play a role in species differentiation. However, experimental evidence linking gene duplication with speciation is scarce. Here, we show that a hybrid-male sterility gene, *Odysseus* (*OdsH*), arose by gene duplication in the *Drosophila* genome. *OdsH* has evolved at a very high rate, whereas its most immediate paralog, *unc-4*, is nearly identical among species in the *Drosophila melanogaster* subgroup. The disparity in their sequence evolution is echoed by the divergence in their expression patterns in both soma and reproductive tissues. We suggest that duplicated genes that have yet to evolve a stable function at the time of speciation may be candidates for “speciation genes,” which is broadly defined as genes that contribute to differential adaptation between species.

Some genes diverge in function during or soon after speciation, leading to differential adaptation between closely related species and, often, reproductive incompatibilities in their hybrids. Such genes are sometimes referred to as “speciation genes” (1, 2). What makes speciation genes prone to diverge? Gene duplication is an attractive hypothesis because duplicated copies are indeed prone to differentiate into new functional niches (3–7), hence contributing to species differentiation (see Supporting Text, which is published as supporting information on the PNAS web site). However, little empirical evidence has become available despite the many theoretical inferences (5, 7). The hybrid-male sterility gene, *Odysseus* (*OdsH*), may fill in the gap (8).

*OdsH* is responsible for hybrid-male sterility when the *Drosophila mauritiana* allele is introgressed into an appropriate *Drosophila simulans* background (9), which carries other *D. mauritiana* genes that are known to be necessary for this sterility interaction. The genetic result has since been confirmed by transgenic experiments (10), which were also performed in properly controlled backgrounds. *OdsH* is most closely related to the paired-type homeodomain gene, *unc-4*, of *Caenorhabditis elegans* (8, 11) and its mammalian homologues (12, 13). Interestingly, the homeodomain of *OdsH* from the *Drosophila melanogaster* subgroup has experienced a 100- to 1,000-fold acceleration in amino acid substitution when compared with the rate of evolution between the *unc-4* genes from mouse and *C. elegans*. The functional homologue of *unc-4* has also been reported in *Drosophila* (14). We refer to this copy as *unc-4* in the sense of sequence, as well as expression, conservation.

Materials and Methods

Statistical Analysis. The pairwise distance between extant species was calculated by using the method of Li (15), as implemented in the GCG package. The method, in its intermediate derivation (15), gives A and S between each pair of genes, where A and S are the total number of nonsynonymous and synonymous substitutions, respectively. We decomposed the number of pairwise substitutions for both A and S into branch length by using the method of Fitch and Margoliash (16). The branch length among *Drosophila sechellia*, *D. simulans*, and *D. mauritiana* was resolved first. The branch length from their common ancestor to *D. melanogaster* was resolved later by using *D. sechellia* and *D. simulans*. The next branch length to *Drosophila yakuba* (purple line in Fig. 2) was resolved by using *D. melanogaster* and *D. simulans*. Finally, we used *unc-4* sequences to resolve the branch length indicated by the black line in Fig. 2. The choice of species in the resolution does not affect the conclusion of the analysis.

Whole-Mount in Situ Hybridization. Whole-mount in situ hybridization was carried out by using a digoxigenin-labeled RNA probe, as described (17). Briefly, embryos or tissues were fixed for 15–20 min in 3.7% formaldehyde after dissection in 1× PBS and dehydrated in methanol before being stored at −20°C. The hybridization was carried out at 65°C in the SDS hybridization solution overnight. The samples were then washed with several wash buffers before adding antidigoxigenin alkaline phosphatase. The hybridization results were detected by color reaction.

Quantitative RT-PCR. Total RNA of whole flies or relevant tissues was extracted with the TRIzol reagent according to the manufacturer’s instructions. A 1–2.5 μg aliquot was used as template for cDNA synthesis, employing the SuperScript First-Strand Synthesis system and oligo(dT) primers. Specific primers of *OdsH*, *unc-4*, and rp49 were used in the quantitative RT-PCRs with 1× SYBR Green PCR Master Mix. All reactions were performed in triplicate. Real-time quantitative PCR analysis was carried out according to the manufacturer’s instructions (Applied Biosystems).

Results

*OdsH* and *unc-4* are 13-kb apart in tandem, as shown schematically in Fig. 1a. They preserve the intron and exon structure but are very divergent in both the N terminus (including the entire exon 1) and the C terminus (most of the exon 4). At the extreme C terminus, the coding region contains a small conserved domain with 8 aa shared between *OdsH* and *unc-4* (Fig. 1b). In *C. elegans*, this domain is responsible for protein–protein interactions (18). Noncoding portion of the gene provides little clue of homology. We also cannot determine whether the exons 1 of...
among the sequences of Fig. 1. Evolutionary divergence may also be asymmetric. The level of divergence includes numerous changes. This extreme asymmetry in the sequence evolution of D. melanogaster can be unambiguously aligned (Fig. 1).

Table 1. Nonsynonymous ($K_s$) and synonymous ($K_a$) substitutions in *OdsH* and *unc-4* of the five species of *Drosophila*

<table>
<thead>
<tr>
<th>Species pair</th>
<th>$K_a \pm SD$</th>
<th>$K_s \pm SD$</th>
<th>$K_a/K_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>OdsH</em> vs. <em>unc-4</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yak (-mel, -sim, -sec, -mau)</td>
<td>0.033</td>
<td>0.122</td>
<td>0.025</td>
</tr>
<tr>
<td>mel (-sim, -sec, -mau)</td>
<td>0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(sim, sec, mau)</td>
<td>0.012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between <em>OdsH</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yak-mel</td>
<td>0.158 ± 0.035</td>
<td>0.317 ± 0.077</td>
<td>0.498</td>
</tr>
<tr>
<td>yak-sim</td>
<td>0.214 ± 0.041</td>
<td>0.409 ± 0.108</td>
<td>0.532</td>
</tr>
<tr>
<td>yak-mau</td>
<td>0.229 ± 0.043</td>
<td>0.425 ± 0.112</td>
<td>0.539</td>
</tr>
<tr>
<td>mel-sec</td>
<td>0.244 ± 0.046</td>
<td>0.412 ± 0.115</td>
<td>0.592</td>
</tr>
<tr>
<td>mel-sim</td>
<td>0.113 ± 0.027</td>
<td>0.157 ± 0.066</td>
<td>0.720</td>
</tr>
<tr>
<td>mel-mau</td>
<td>0.126 ± 0.029</td>
<td>0.194 ± 0.071</td>
<td>0.649</td>
</tr>
<tr>
<td>sec-sim</td>
<td>0.151 ± 0.032</td>
<td>0.186 ± 0.073</td>
<td>0.812</td>
</tr>
<tr>
<td>sec-mau</td>
<td>0.026 ± 0.016</td>
<td>0.041 ± 0.025</td>
<td>0.634</td>
</tr>
<tr>
<td>sim-mau</td>
<td>0.051 ± 0.021</td>
<td>0.040 ± 0.032</td>
<td>1.275</td>
</tr>
<tr>
<td>yaku-sim</td>
<td>0.061 ± 0.023</td>
<td>0.024 ± 0.023</td>
<td>2.542</td>
</tr>
</tbody>
</table>

*See text for the statistical tests. yak, D. yakuba; mel, D. melanogaster; sim, D. simulans; mau, D. mauritiana; sec, D. sechellia.*

*Averaged over the five species.

*Times that the divergence between D. yakuba and D. melanogaster on the basis of the $K_s$ values of *OdsH*. The duplication might have occurred in the genus *Drosophila*, likely in the lineage of the subgenus Sophophora after it split from the *Drosophila* subgenus. The recent genomic information on Anopheles and *Drosophila pseudoobscura* is consistent with this estimation. Anopheles has only one copy of *unc-4*-like gene, whereas *D. pseudoobscura* has both *unc-4* and *OdsH*.

Whereas *unc-4* is strongly conserved, *OdsH* shows an interesting trend in its sequence evolution. The $K_A/K_S$ ratio (*K_A* being the number of nonsynonymous changes per nonsynonymous site, and *K_S* being the number of synonymous changes per synonymous site) has increased appreciably in the lineages leading to *D. melanogaster* and its three sibling species. The ratio is highest between *D. mauritiana* and *D. simulans*, in which the effect of *OdsH* on hybrid-male sterility is manifested. To examine the trend statistically, we assign the number of nonsynonymous (A) and synonymous (S) substitutions to different parts of the phylogenetic tree of Fig. 2. Along the branch between the extant *unc-4* genes and the common ancestor of *D. melanogaster* and *D. yakuba*, indicated by the thin black line in Fig. 2, the A/S ratio is 51.5:90.8. From *D. yakuba* to the ancestor of the sibling species of *D. melanogaster* (purple line in Fig. 2), the ratio is 41.4:38.0, and among the *D. melanogaster* sibling species (solid red line),
The accelerated divergence is observed mostly in the last 1–2 million years, after a long period of “quiescence” since duplication. This recent episode of rapid sequence evolution is coupled with the different expression patterns of OdsH among extant sibling species (Fig. 3). The observation contrasts with the widely held view that duplicated genes diverge immediately after duplication (4, 5, 7, 19).

To study the divergence of OdsH among the sibling species, we surveyed the expression patterns by whole-mount in situ hybridization (Fig. 3). OdsH is expressed in the male reproductive tissues. The expression patterns are variable among the four sibling species, such that in D. melanogaster and D. yakuba, we observed relatively strong expression in the testis but not in the accessory gland (Fig. 3 a and d). In D. simulans, the expression is relatively low at the apical region of the testis, but it is high in the accessory gland (Fig. 3b). In D. mauritiana, the expression in the accessory gland is weaker than expression in the testis (Fig. 3c).

In C. elegans, unc-4 controls the motor-neuron differentiation (18). The published study on Drosophila unc-4 gene has revealed its expression in specific postmitotic neurons (14). As shown in Fig. 3, unc-4 is expressed (segment-wise) in a subset of ectodermal and neural cells in the CNS. This expression pattern is conserved across divergent species, from D. melanogaster, D. yakuba (Fig. 3i–l), and Drosophila ananassae to D. pseudoobscura (data not shown). In contrast, there is no detectable expression of OdsH in D. melanogaster or D. yakuba embryos (Fig. 3e and h). The ubiquitous expression of OdsH observed in the embryos of D. simulans and D. mauritiana does not show specific neuronal patterns (Fig. 3f and g). We further analyzed the expression levels of OdsH and unc-4 in embryos, larvae, adult males, and females of D. melanogaster by quantitative RT-PCR (Fig. 4a). Consistent with the in situ hybridization results, OdsH is most strongly expressed in adult males in D. melanogaster, whereas only a few-fold differences can be detected between males and females in D. mauritiana. The expression of OdsH in different tissues of the adult males was also assayed, and the result shows that the OdsH expression in males is mainly in the testes (Fig. 4b).

**Discussion**

Overall, OdsH has been evolving rapidly in both expression and DNA sequence. Presumably, the duplication of unc-4 has permitted it to evolve into a new functional niche. A recent experiment by sequence-dependent gene knockout has shown that OdsH functions in promoting the fertility of very young males (10). This new functional role of OdsH has been evolving away from the unc-4 expression and toward a testicular function.

---

Fig. 3. Gene expression in the four Drosophila species. (a–d) OdsH expression in male reproductive tissues. T, testis; AC, accessory gland. (e–h) OdsH expression in stage 14 embryos. (i–l) unc-4 expression in stage 14 embryos. D. melanogaster, D. mauritiana; D. melanogaster, D. simulans; D. yakuba, D. yakuba.

Fig. 4. Expression levels of OdsH and unc-4 in D. melanogaster, as quantified by real-time RT-PCR. All values are normalized against an internal control (rp-49), and the relative expression level is shown in the log scale. (a) Expression in different stages and sexes. The reference point of zero on the y axis is approximately the average expression of OdsH across the four measurements. (b) Expression of OdsH in different tissues in males.
The OdsH-induced hybrid-male sterility might be the consequence of such a shift in expression. We should note that other genetic factors from both species have been shown to be involved in this sterility interaction (9), making it more complex than the simplest two-locus Dobzhansky–Muller mechanism (20).

Without gene duplication, orthologous genes between the diverging species usually have to assume comparable functions, unless the environments have shifted drastically. However, if gene duplication occurred shortly before speciation, the diverging species would each inherit a duplicated gene whose functions are still prone to differentiation. The undifferentiated and redundant function is expected to evolve along different trajectories in the two diverging species. Although the connection between gene duplication and reproductive isolation has been suggested (4), our model is quite different in that it emphasizes functional divergence and differential adaptation (1).

The conventional view on gene duplication is that one of the two duplicated genes retains the original function, allowing the other to evolve a new one (neo-functionalization (3, 21)). Recent reports have suggested an alternative hypothesis that the ancestral functions are split between the duplicated genes [subfunctionalization (4, 6, 7)]. The duplicated pair of unc-4 and OdsH appears to have evolved in the former mode.

OdsH is also a “dispensable” gene because its deletion reduces male fertility only partially and only when males are <4 days old (10). The concept of dispensability is important because it makes a distinction between physiological assay and evolutionary interpretation. Dispensability indicates merely that a gene is not needed for normal viability and fertility under general conditions, without implying its fitness effect in the evolutionary context. Many other rapidly evolving genes that bear the signature of positive selection are also dispensable. The examples include the Acp26Aa and desaturase-2 genes in Drosophila and the glycoporins in human (22–24). OdsH is another example of a dispensable gene playing a significant role in evolution. In fact, such genes, being less constrained in evolution, may be more prone to diverge in function during speciation. It is the divergence in function, not the physiological importance of the gene function, that contributes to species differentiation.

Our results reveal a connection between species differentiation and gene duplication. Duplicated genes that are in the process of evolving into new functions at the time of species separation are likely to contribute to species differentiation. Many of the differentiated genes between recently diverged species may contribute to hybrid incompatibility, sexual isolation, or morphological differences. Molecular analyses of additional speciation genes may shed further light on the genetic basis of speciation that has been elusive for many decades (1).

We thank Justin Fay for helping with part of D. yakuba sequencing and Yu-Ping Poh for sequence analysis. This work was supported by grants from the National Institutes of Health and the National Science Foundation (to C.-I.W.) and the National Science Council of Taiwan (to C.-T.T. and S.-C.T.). N.H.P. is an Investigator of the Howard Hughes Medical Institute.