Evolution of hydra, a Recently Evolved Testis-Expressed Gene with Nine Alternative First Exons in Drosophila melanogaster

Shou-Tao Chen1, Hsin-Chien Cheng1, Daniel A. Barbash2, Hsiao-Pei Yang1,2*

1 Faculty of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, Taiwan, Republic of China; 2 Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, United States of America

We describe here the Drosophila gene hydra that appears to have originated de novo in the melanogaster subgroup and subsequently evolved in both structure and expression level in Drosophila melanogaster and its sibling species. D. melanogaster hydra encodes a predicted protein of ~300 amino acids with no apparent similarity to any previously known proteins. The syntenic region flanking hydra on both sides is found in both D. ananassae and D. pseudoobscura, but hydra is found only in melanogaster subgroup species, suggesting that it originated less than ~13 million y ago. Exon 1 of hydra has undergone recurrent duplications, leading to the formation of nine tandem alternative exon 1s in D. melanogaster. Seven of these alternative exons are flanked on their 3' side by the transposon DINE-1 (Drosophila interspersed element-1). We demonstrate that at least four of the nine duplicated exon 1s can function as alternative transcription start sites. The entire hydra locus has also duplicated in D. simulans and D. sechellia. D. melanogaster hydra is expressed most intensely in the proximal testis, suggesting a role in late-stage spermatogenesis. The coding region of hydra has a relatively high Ka/Ks ratio between species, but the ratio is less than 1 in all comparisons, suggesting that hydra is subject to functional constraint. Analysis of sequence polymorphism and divergence of hydra shows that it has evolved under positive selection in the lineage leading to D. melanogaster. The dramatic structural changes surrounding the first exons do not affect the tissue specificity of gene expression: hydra is expressed predominantly in the testes in D. melanogaster, D. simulans, and D. yakuba. However, we have found that expression level changed dramatically (~ -> 20-fold) between D. melanogaster and D. simulans. While hydra initially evolved in the absence of nearby transposable element insertions, we suggest that the subsequent accumulation of repetitive sequences in the hydra region may have contributed to structural and expression-level evolution by inducing rearrangements and causing local heterochromatinization. Our analysis further shows that recurrent evolution of both gene structure and expression level may be characteristics of newly evolved genes. We also suggest that late-stage spermatogenesis is the functional target for newly evolved and rapidly evolving male-specific genes.

Introduction

Much of the genetic novelty that accompanies speciation and organismal evolution is driven by reutilization of pre-existing genetic information. In an influential essay, Francois Jacob likened evolution to a process of tinkering [1]. After the primordial evolution of truly new macromolecules and mechanisms of replication, Jacob suggested that much of phenotypic novelty arises from reusing, recombining, and altering the function of available genes.

This view of evolution is strongly supported by studies of new-gene evolution. The vast majority of newly evolved genes can be attributed to duplication of pre-existing genes. These duplication events can range from single-gene events to duplications of entire genomes [2,3]. Much of protein evolution also conforms to Jacob's view, as new proteins are often generated from shuffling pre-existing protein domains [4].

The tinkering view of evolution does not rule out the occasional generation of novel protein sequences. One question then is whether and by what mechanisms such novel proteins evolve. Presumably novel proteins derive from noncoding sequences that acquire appropriate transcriptional and translational regulatory sequences. Experimental evolution studies suggest that random sequences of proteins can acquire biological functions at frequencies that are, surprisingly, greater than miniscule [5]. A recent study reported the exciting finding of several such candidate de novo genes in Drosophila that may be functional, based on RNA expression analysis [6]. A large number of new exons have also been identified in rodents that appear to have evolved in the DINE-1 element in the transcribed region 5' of the gene DVE4 in the mouse [7].

We are now encountering a growing number of cases in which alternative exons have evolved in a single gene. In these cases, it is often not clear whether the alternative exons are in cis or in trans. Here, we describe hydra, a testis expressed gene in Drosophila melanogaster that has at least four different alternative exons.

Citation: Chen ST, Cheng HC, Barbash DA, Yang HP (2007) Evolution of hydra, a recently evolved testis-expressed gene with nine alternative first exons in Drosophila melanogaster. PLoS Genetics 3(7): e107. doi:10.1371/journal.pgen.0030107
Author Summary

Similar groups of animals have similar numbers of genes, but not all of these genes are the same. While some genes are highly conserved and can be easily and uniquely identified in species ranging from yeast to plants to humans, other genes are sometimes found in only a small number or even in a single species. Such newly evolved genes may help produce traits that make species unique. We describe here a newly evolved gene called hydra that occurs only in a small subgroup of Drosophila species. hydra is expressed in the testes, suggesting that it may have a function in male fertility. hydra has evolved significantly in its structure and protein-coding sequence among species. The authors named the gene hydra after the nine-headed monster slain by Hercules because in one species, Drosophila melanogaster, hydra has nine potential alternative first exons. Perhaps because of this or other structural changes, the level of RNA made by hydra differs significantly between one pair of species. This analysis reveals that newly created genes may evolve rapidly in sequence, structure, and expression level.

derived from incorporation of intronic sequences into mRNAs [7].

A large fraction of eukaryotic genomes is composed of transposable elements (TEs). Because the TE component of genomes can evolve rapidly, TEs have a major impact on genome evolution [8–10]. TEs are widely considered to be selfish parasites that are deleterious to their host’s fitness [11], and most are likely eliminated quickly by natural selection. However, accumulating evidence has also shown that TEs can serve as an important source of genetic variation and genomic novelty [8,12–14].

TEs can generate genetic novelty by at least five different mechanisms. First, the reverse transcriptase enzyme from retrotransposons can generate duplicated retrotransposed genes [15–17] or generate chimeric fusion genes [15,18–21]. Second, TEs can change expression patterns of adjacent genes by providing novel regulatory elements or by disrupting host gene regulatory functions [14,22–26]. Third, coding regions from TEs such as envelope proteins and transposase enzymes can be incorporated by the host species to modify or even create new host protein-coding genes [13,27]. Fourth, TEs can contribute to noncoding regions (untranslated regions [UTRs]) of genes [12,24,28–30]. Fifth, TE insertions create multiple highly homologous sites throughout the genome that can cause duplications by ectopic recombination [9,31].

Comparisons between genomes have proven to be a powerful way to identify new genes [32] and to reconstruct the early history of new-gene evolution [2]. This approach has provided new data for models and mechanisms leading to new-gene evolution [6,20,33–35].

Such a comparative approach is used here to investigate the evolutionary history of a recently evolved gene in Drosophila. We report the characterization of the gene hydra, located in the pericentric region of the X chromosome in D. melanogaster. hydra was originally reported as predicted gene CG1338 [36]. We show that hydra originated in the melanogaster subgroup of Drosophila and subsequently experienced dramatic changes in its gene structure, including the formation of multiple alternative first exons, most of which are associated with the transposon DINE-1 (Drosophila interspersed element-1). We also investigate the history and impact of these gene-structure changes at both the DNA sequence and RNA expression levels by comparison among species of the melanogaster subgroup. We suggest that new genes will tend to evolve rapidly in coding sequence (CDS), gene structure, and gene-expression level. We also suggest that TE accumulation can cause local heterochromatinization that can in turn have a major impact on the evolution of gene structure and gene expression.

Results

The hydra Gene Is Restricted to the melanogaster Subgroup

CG1338 is located near the pericentric region of the X chromosome in D. melanogaster (cytological region 19E1). We first became interested in CG1338 during a genome-wide analysis of the transposon DINE-1 in the D. melanogaster genome. We present evidence below that CG1338 contains nine duplicated first exons in D. melanogaster. Because of these nine duplicated exon 1s (see Figure 1A), we propose to rename this gene hydra, after the nine-headed monster slain by Hercules.

We found that the gene region surrounding hydra in D. melanogaster contains dense repeats of DINE-1 (Figure 1A). DINE-1, also named INE-1 or DNAREP1DM, is a highly abundant transposon that is predominantly found in the heterochromatic regions of the D. melanogaster genome [37,38] and is believed to have invaded the Drosophila genome before the diversification of the melanogaster subgroup [37–40]. Further analysis showed that hydra is flanked by two blocks of tandem repeats, with duplicated fragments of DINE-1 in the 5′ end and an undefined 500-bp repeat in the 3′ end. Using available genomic contigs from various Drosophila species (http://flybase.net/blast), we found that the ~7-kb region surrounding hydra, including the adjacent gene CG1835, only exists in species of the melanogaster subgroup, but not in other species outside of the subgroup. BLAST searches using unique sequences from hydra identified no similar sequences in other regions of these genomes. We suggest that hydra evolved recently in the species of the melanogaster subgroup (Figure 1B). This conclusion was confirmed by Southern hybridization under low stringency using hydra exons 2–4 from D. melanogaster as the probe with DNA extracted from D. melanogaster, D. simulans, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, and D. virilis. Hybridization signals were only detected in flies of the melanogaster subgroup (i.e., D. melanogaster, D. simulans, D. yakuba, D. erecta), but not in the other species (unpublished data). TBLASTN searches to the nonredundant (nr) or whole-genome shotgun (wgs) databases found no hits to non-melanogaster subgroup Drosophila species with E values < 0.001.

Evolution of the Gene Structure of hydra

According to the Flybase genome annotation (http://www.flybase.org), hydra of D. melanogaster contains four exons, and two types of transcripts (type A and type B), which are transcribed using alternative first exons. Our analysis of the hydra gene region suggested that there are seven additional putative first exons similar to the two annotated first exons located in a region of dense DINE-1s in D. melanogaster (Figure 1A; Text S1). These seven additional potential first exons share similarity with the two annotated exon 1s in their CDSs (which are nine codons long) and in their 5′ upstream
regions. The first of these duplicated exons may be nonfunctional because it creates a stop codon when spliced to exon 2, assuming that it has the same structure and splicing pattern as the other 8 exons. Alternatively, this first exon 1 may use a different 3′ splice site, although no such evidence was found in our 5′ rapid amplification of cDNA ends (RACE) analysis described below (Table 1).

Our further analysis of multiple lines of D. melanogaster (and D. simulans) with Southern hybridization detected substantial length variation in the gene region, presumably due to length variation in the exon 1 region (Figure 2). Although we have not determined the sequence structure of these alleles, the length variation observed is consistent with the hypothesis that the number of alternative first exons varies in D. melanogaster. In contrast, there is only a single exon 1 in the orthologous region of hydra from the species D. yakuba and D. erecta (Texts S5 and S6), and there is no length polymorphism in this region in D. yakuba (unpublished data). This suggests that the dramatic gene structure change by duplication of exon 1 in hydra occurred within ~5–13 million y, after the divergence of D. melanogaster and D. yakuba [41,42]. By comparing the orthologous sequences of hydra using the released assembled genomic sequences of D. simulans and D. sechellia, we found that this region has undergone local duplications in both species (Figure 1B; Texts S2–S4). Determining the precise structure of the hydra region in D. sechellia will require further sequence analysis due to low sequence coverage of this region. We have, however, confirmed by Southern analysis that D. simulans does have two copies of hydra (Figure 2C). In D. simulans, we find that one of these copies has two potential exon 1s. In D. sechellia, we identified two unlinked contigs containing hydra-homologous sequences (Figure 1B). Scaffold 8 contains the syntenic genes run and cyp6V1 flanking a single copy of hydra. This copy of hydra contains two putative exon 1s. Scaffold 600 contains two tandem copies of hydra, each of which contains two or three putative exon 1s. These data suggest the possibility that D. sechellia has three copies of hydra, although further analysis will be required to determine the relative locations of these apparent duplicated copies that are on different scaffolds. Despite this uncertainty, we were able to investigate the phylogenetic relationship of these copies of hydra. Using either the sequence between exons 2–4, inclusive (Figure 3A; Text S7) or of intron 1 (unpublished data), we obtained similar phylogenetic trees that suggest that hydra duplicated independently in D. simulans and D. sechellia. Because both gene regions produce the same phylogeny and these sequences span over 1,300 bp in all species, we consider it unlikely that this phylogenetic signal of independent duplications could be due to gene conversion.

Upon further investigation of the intronic sequences next to exon 1, we found that seven out of the nine duplicated exon 1s in D. melanogaster and several of the duplicated exon 1s in both D. simulans and D. sechellia are immediately followed by DINE-1 sequences, while in other species of the melanogaster subgroup, including D. yakuba and D. erecta, there are no DINE-1s in this region. We thought it likely for two reasons that insertion of DINE-1 adjacent to hydra exon 1 occurred in the common ancestor of D. melanogaster, D. simulans, and D. sechellia. First, it seems highly unlikely that a TE would insert independently at the same location in different species. Second, DINE-1 is thought to have been active and then to have become transpositionally inactive before these species diverged [38,40]. To test this hypothesis and to understand the evolutionary history and mechanism that resulted in the exon 1 duplications of hydra, we performed phylogenetic analysis of all the available exon 1s (Figure 3B; Text S8). We found that D. simulans S2 and D. sechellia C3 are outgroups to all other exon 1s. Both of these exons are flanked by DINE-1 sequence, which is consistent with our above suggestion that DINE-1 insertion in this region is ancestral (at time B in Figure 1B). If this

Table 1. Relative Usage of Different Exon 1s of hydra, Determined by 5′ RACE

| Group A | 1 | 5 | 2 or 3 | Total in Group A | Group B | 6 | 9 | 6 or 9 | 4, 7, or 8 | 4, 6, 7, 8, or 9 | Total in Group B |
|---------|---|---|--------|----------------|---------|---|---|--------|------------|----------------|-----------------|
| Number of clones | 0 | 0 | 22 | 22 | 7 | 2 | 2 | 9 | 16 | 36 | 62.1% |
| Percentage of total clones | 37.9% | 37.9% | 12.1% | 3.4% | 3.4% | 15.5% | 27.6% | 62.1% |

All clones could be assigned to group A or B based on polymorphisms described in the Results. Alternative first exons 2 and 3 are identical, as are 4, 7, and 8, and thus cannot be distinguished. Three further subgroups of first exons in group B can be distinguished in longer clones by additional polymorphisms in their 5′ UTRs (subgroups 6, 9, and 4/7/8). In shorter clones from group B, these subgroups could not be uniquely distinguished and were thus classified as either 6 or 9, or as 4, 6, 7, 8, or 9.

doi:10.1371/journal.pgen.0030107.g001
inference of the ancestral state of exon 1 in these three
species is correct, it is interesting to note that DINE-1
does not flank every exon 1. Our data would then suggest
that either DINE-1 was lost next to some duplicated exon 1s or
that some exon 1s duplicated without duplicating the
adjacent DINE-1.

The relationships among the remaining D. simulans and D.
sechellia exon 1s suggest that exon 1 duplicated before the
speciation of D. simulans and D. sechellia to generate the exon 1
groups S1/S3 and C1/C4/C6, with another duplication event
generating exon 1 group C2/C5/C7 (Figure 3B). Both this tree
and the one made using the rest of the locus (Figure 3A)
suggest that after these exon 1 duplications, the entire locus
duplicated independently in D. simulans and D. sechellia.

Figure 3B also suggests that exon 1 duplicated independ-
ently in D. melanogaster. Based on the sequence similarity
of predicted exon 1 sequences, the duplicated first exons of D.
melanogaster can be divided into three groups: group A,
containing the first, second, third, and fifth duplicated exon
1s; group B, containing the fourth, seventh, eighth, and ninth
duplicated exon 1s; and group C, containing the sixth
duplicated exon 1. The first and the ninth exon 1s are the
previously annotated exon 1s for transcript types RA and RB,
respectively. Our phylogenetic analysis showed that exon 1s
adjacent to each other are generally more closely related
(Figure 3B). These results suggest that both the exon 1s and
their flanking DINE-1s were duplicated together, and that
unequal crossing-over is the major mechanism generating the
tandemly duplicated exons.

**hydra Is Expressed Predominantly in Adult Testes**

RT-PCR analysis demonstrated that hydra is expressed in
adult male testes in all three species tested: D. melanogaster, D.
simulans, and D. yakuba (Figure 4). This result demonstrates
that the expression pattern of hydra is conserved despite
having a very different gene structure in these species. In one
out of two stocks of D. melanogaster that we examined, we also
detected a low level of expression in the male carcass from
which testes were removed (Figure 4B). This expression
pattern is consistent with D. melanogaster EST data: among ten
ESTs in GenBank that match hydra, nine are from testis cDNA
libraries, and one is from a larval fat-body library. We further
investigated the pattern of expression of *hydra* in *D. melanogaster* by in situ hybridization. We found that *hydra* is expressed most intensely in the proximal region of the testis (Figure 5), where mature sperm are formed, suggesting that *hydra* functions in late stages of sperm development.

**Functional Consequences of the Evolving Gene Structure of hydra**

To study the impact of exon 1 duplication on gene expression, we asked if each of the duplicated exons could function as a transcriptional start site. To determine whether each putative exon 1 is functional in *D. melanogaster*, we needed to use a technique that could distinguish these very similar potential transcripts. We therefore performed 5′ RACE analysis using RNA extracted from adult male testes of *D. melanogaster* and analyzed the sequences of 58 clones. The previously described Ra and Rb transcripts differ by three nucleotides at positions 21 to 23 in their 5′ UTRs as well as by a nonsynonymous polymorphism in their second codon.

We designate alternative first exons 1, 2, 3, and 5 as class A (similar to Ra), and alternative first exons 4, 6, 7, 8, and 9 as class B (similar to Rb).

Some of these sequences can be further distinguished by additional polymorphisms further 5′ in the 5′ UTRs. Six different groups of alternative transcripts could be potentially distinguished, with two of these groups containing two or three identical first exons (Table 1). Among our 58 5′ RACE clones, four out of six of these groups were represented at least twice, demonstrating that at least four different alternative first exons in *hydra* are functional.

We also analyzed to see if each duplicated first exon contains core promoter sequences known to be important for the initiation of transcription of *Drosophila* genes [43]. We looked specifically for the TFIIB recognition element (−37 to −32 bp; GGGCGCC or CCACGCC), TATA box (−31 to −26 bp; TATAAA); initiator (−1 to +4 bp; TCA(G/T)T(T/C)), and downstream promoter element (+28 to +32 bp; (A/G)G(A/T)(G/T)(G/A/C)). We found that almost all duplicated exon 1s

---

**Figure 3. Phylogenetic Trees of hydra**

Trees are based on (A) exon 2–4 sequences and (B) exon 1, including putative 5′ UTR sequences. The trees were built using the neighbor-joining method according to the Jukes-Cantor substitution model. Branch lengths are proportional to the number of nucleotide substitutions. Percentage bootstrap values are shown at each node (500 replicates). In (B), sequences mel 1–9 correspond to the alternative exon 1s of *D. melanogaster* shown in Figure 1B. Sequences Ra and Rb are from the annotated transcripts of *D. melanogaster hydra* in FlyBase. *D. simulans* and *D. sechellia* exon 1s are labeled as shown in Figure 1B. Note that mel 6 is part of group B but clusters with group A, presumably because of having a divergent 5′ UTR.

doi:10.1371/journal.pgen.0030107.g003
contain all four core promoter sequences, with two exceptions: the sixth and the ninth duplicated exon 1s do not contain the downstream promoter element and the TATA box, respectively. Since both of these exons are used (Table 1), our results suggest that these two elements are not essential for transcription of hydra.

Using the sequences of the clones from 5' RACE, we also mapped the sites for transcription initiation. We found that most transcripts of hydra do not use a fixed site for transcription initiation. Instead, “slippery promoters” were found to be used in different exon 1s, as described for other Drosophila genes by Yasuhara et al. [44].

We also tested whether hydra structural evolution might lead to expression-level differences between sibling species by quantifying hydra expression levels in five wild lines of D. melanogaster and in four wild lines of D. simulans (Figure 6). Strikingly, hydra was expressed at a substantially higher level in all D. simulans lines compared with all D. melanogaster lines. Among all pair-wise comparisons, the fold increase in D. simulans ranged from ~4 to ~150, with the mean fold difference being ~22. These data indicate that hydra has lower expression in D. melanogaster, despite having multiple alternative transcription start sites.

Nonneutral Evolution of hydra

hydra contains an intact reading frame in all melanogaster subgroup species, encoding a predicted protein of 302, 280, and 308 amino acids in D. melanogaster, D. simulans, and D. yakuba, respectively. We calculated Ka/Ks ratios of the coding region of hydra among species in the melanogaster subgroup (Table 2). The ratio is <1 in all comparisons, suggesting that hydra is evolving under some degree of functional constraint.

To gain further insight into the evolution of hydra, we collected population samples from D. melanogaster and D. simulans, which have the highest interspecific Ka/Ks ratio of 0.75 (Texts S9 and S10). We found that the nucleotide diversity at both synonymous (\(\pi_s\)) and nonsynonymous (\(\pi_a\)) sites in D. melanogaster is similar to the average of other functional genes, with about ten times more polymorphism at synonymous sites relative to nonsynonymous sites (Table 3). In contrast, there is not much difference between \(\pi_s\) and \(\pi_a\) in D. simulans, where \(\pi_s\) is much lower than the average of other genes for this species [45]. In order to further investigate whether this unusual pattern of nucleotide diversity (similar levels of synonymous and nonsynonymous polymorphism) in D. simulans is lineage specific, we studied polymorphism in D. yakuba. \(\pi_s\) and \(\pi_a\) of hydra in D. yakuba are similar to those of D. melanogaster, and \(\pi_s\) is similar to the average value (0.0127) of six X-linked loci in D. yakuba [46], suggesting that the unusual pattern of nucleotide diversity of hydra is restricted only to D. simulans. The approximately equal values of \(\pi_s\) and \(\pi_a\) in D. simulans may suggest that hydra is under reduced functional
constraint in *D. simulans*. Two findings suggest, however, that *hydra* is not a pseudogene in *D. simulans*. First, it has an intact open reading frame, and second, *hydra* is expressed in *D. simulans* and has a similar expression pattern to other species (see above).

Results of the McDonald-Kreitman (MK) test [47] among these three species are shown in Table 4. Neutral evolution of *hydra* can be rejected between *D. melanogaster* and *D. yakuba*, which we suggest is caused by the large number of non-synonymous substitutions between these species. There is also a large amount of non-synonymous divergence between *D. melanogaster* and *D. simulans*. We do not, however, reject neutrality in this case, likely due to the unusually low πs in *D. simulans* discussed above (Table 3). Thus, these results suggest that *D. melanogaster* is the most likely lineage in which *hydra* has experienced positive selection.

To confirm this conclusion, we tested to see if the nonneutral evolution of *hydra* occurred in a lineage-specific pattern by performing polarized substitution analysis with the MK test (Table 4). Replacements were polarized to each species lineage using *D. yakuba hydra* as an outgroup sequence. The polarized MK test is significant only in the *D. melanogaster* lineage.

### Discussion

**hydra Is a Recently Originated Functional Gene**

We identified the gene *hydra* in *D. melanogaster*, but could only identify orthologs within other *melanogaster* subgroup species. Using both BLAST searches of full-genome sequences with the 6-kb region flanking *hydra* and Southern hybridization, we were unable to identify *hydra* from *D. ananassae*, *D. pseudoobscura*, or *D. virilis*. We were, however, able to identify the syntenic region in these species (Figure 1B). These data strongly suggest that *hydra* originated in the common ancestor of the *melanogaster* subgroup.

All available evidence also suggests that *hydra* is a functional gene. First, all *melanogaster* subgroup species contain intact open reading frames with no stop codons, even though there are many insertions and deletions in the *hydra* CDS among species. Second, Ka/Ks values between all species are less than 1, suggesting that *hydra* is under functional constraint (Table 2). Third, the level of polymorphism of *hydra* in *D. melanogaster* is similar to known functionally important genes (Table 3). Fourth, RT-PCR and in situ hybridization show that *hydra* is predominantly expressed in male testes (Figures 4 and 5).

### Table 2. Summary of Ka and Ks Analysis of *hydra* among Four Species of the *melanogaster* Subgroup

| Species Pair                 | Number of Synonymous Positions | Ks    | Number of Nonsynonymous Positions | Ka    | Ka/Ks |
|-----------------------------|--------------------------------|-------|-----------------------------------|-------|-------|
| *D. simulans/D. melanogaster* | 170.31                         | 0.17  | 576.69                            | 0.13  | 0.75  |
| *D. simulans/D. yakuba*     | 169.00                         | 0.37  | 578.00                            | 0.22  | 0.58  |
| *D. melanogaster/D. yakuba* | 168.97                         | 0.36  | 578.03                            | 0.23  | 0.63  |
| *D. simulans/D. erecta*     | 168.00                         | 0.42  | 579.00                            | 0.21  | 0.50  |
| *D. melanogaster/D. erecta* | 167.97                         | 0.39  | 579.03                            | 0.22  | 0.56  |
| *D. yakuba/D. erecta*       | 166.67                         | 0.32  | 580.33                            | 0.14  | 0.44  |

Based on sequences in both coding (exons 2–4) and noncoding (introns 2–3 and 3’ UTR) sequences.

Ks, nucleotide divergence at synonymous sites; Ka, nucleotide divergence at non-synonymous sites.

doi:10.1371/journal.pgen.0030107.t002
hydra Shares Common Features of New-Gene Evolution

Several other cases of new-gene evolution in Drosophila have been studied [2,48]. Examples include (1) Jingwei, a chimeric gene that evolved as a result of the insertion of an Adh retrosequence into a duplicated locus in the D. teissieri/D. yakuba lineage [19]; (2) Sdic, which originated through a complex set of rearrangements, including a gene fusion between the gene encoding the cell adhesion protein annexin X and a cytoplasmic dynein intermediate chain (Sdic encodes a novel sperm-specific dynein found only in D. melanogaster) [49]; (3) Sphinx, a recently evolved gene apparently created by both retroposition and exon shuffling, which produces a novel noncoding RNA gene found only in D. melanogaster [33]; (4) mkg, a gene family generated originally as retroposed duplicates about 1–2 million y ago that further evolved by individual members undergoing gene fission and fusion events [35]; (5) K81, a gene that arose through duplication by retroposition after the radiation of the melanogaster subgroup and acquired a new male germline–specific function [50]; (6) Dntf-2r, a rapidly evolving male-specific gene in D. melanogaster and sibling species, which originated from retroposition [34]; and (7) five protein-coding genes which appear to have originated in D. melanogaster from noncoding DNA and evolved male-biased expression [6].

Despite the diverse mechanisms involved in the creation of these young genes in Drosophila, there are two features common to most of them: (1) they evolve rapidly in sequence, including at nonsynonymous sites for protein-coding genes [19,20,34,49,51,52]; and (2) their expression is often male-specific, and in many cases testis-specific [6,52]. This expression pattern was also often found in a large survey of new retroposed genes [17].

hydra shares both of these characteristics. The Ka and Ks values between D. melanogaster and D. simulans were substantially higher than the average values of most other genes, and the Ka/Ks ratio ranged from 0.44 to 0.75 in all species compared (Table 2). These high values may be explained by accelerated functional divergence following the origin of new genes. However, the elevated Ka/Ks ratio could also result from an elevated mutation rate and reduced selective constraint. To further investigate the evolution of hydra, we analyzed the level of polymorphism and divergence among D. melanogaster, D. simulans, and D. yakuba. Our population genetics analyses rejected the null hypothesis of neutral evolution in D. melanogaster (Table 4). This rejection appears to be due to a significant excess of fixed differences at nonsynonymous sites in D. melanogaster, suggesting that positive selection has driven the evolution of hydra in D. melanogaster.

hydra Is Expressed in the Basal Part of the Testis

The male-specific expression pattern is a second feature that hydra shares with other new genes. For many newly evolved genes, male-specific or testis-specific expression has been demonstrated by RT-PCR, but the spatial expression patterns are unknown, making it difficult to refine our knowledge of what aspects of testis development or spermatogenesis may be under selection. We determined that hydra is expressed most intensely in the basal (proximal) part of the testis in D. melanogaster by in situ hybridization (Figure 5). This localization pattern suggests that hydra may have a function in late-stage spermatogenesis or sperm differentiation. We note that hydra is X-linked, which suggests that it escapes from the X inactivation that occurs during spermatogenesis [53]. Our results also stand in contrast to results from genome-wide studies that suggest that genes with male-biased expression as well as newly evolved testis-specific genes tend to be autosomal [17,45,55]. Interestingly, the newly evolved testis-expressed genes Sdic and CG15323 are also located in polytene region 19 of the X chromosome [6,49].

Several studies have shown that genes with male-biased expression have significantly faster rates of evolution than genes with female-biased or unbiased expression [56,57]. This difference is caused primarily by a higher Ka in the male-biased genes. Sexual selection is likely to be driving this acceleration of male-biased genes, but functional tests will be

Table 3. Summary of Polymorphism Data from hydra in D. melanogaster, D. simulans, and D. yakuba

| Species          | n  | ns   | na   |
|------------------|----|------|------|
| D. melanogaster  | 12 | 0.0123 | 0.0018 |
| D. simulans      | 16 | 0.0071 | 0.0072 |
| D. yakuba        | 13 | 0.0161 | 0.0055 |

Based on the same gene region of Table 2.

n: number of sequences; ns: nucleotide diversity at synonymous sites; na: nucleotide diversity at nonsynonymous sites.
doi:10.1371/journal.pgen.0030107.t003

Table 4. MK Tests for Nonneutral Evolution of hydra

| Species or lineage | Divergence | Polymorphism | MK Test |
|--------------------|------------|--------------|---------|
|                    | Nonsynonymous | Synonymous | Nonsynonymous | Synonymous | Fisher's Exact Test p (2-Tailed) |
| D. melanogaster and D. simulans | 60          | 25          | 19                | 14         | 0.196010 (ns) |
| D. melanogaster and D. yakuba    | 113         | 52          | 16                | 19         | 0.018517 * |
| D. simulans and D. yakuba        | 94          | 52          | 23                | 18         | 0.364027 (ns) |
| D. melanogaster lineagea         | 24          | 10          | 3                 | 6          | 0.049294 * |
| D. simulans linagaea             | 16          | 8           | 10                | 6          | 1.00000 (ns) |

*0.01 < p < 0.05, ns, nonsignificant.

aBased on polarized MK test of D. melanogaster and D. simulans using D. yakuba as an outgroup.

doi:10.1371/journal.pgen.0030107.t004
required to understand fully the role of selection. Our study of *hydra* suggests that late-stage spermatogenesis may be the key developmental process to investigate further for functional differences between species.

**Novel Aspects of *hydra*: *hydra* Has Undergone Recurrent Structural Evolution**

Our analysis of *hydra* reveals several aspects that are unusual or unique compared with other newly evolved genes. One novel feature of *hydra* is that it has experienced recurrent structural evolution in the lineages leading to both *D. melanogaster* and *D. simulans* (Figure 1). We can infer the likely occurrence of three independent structural changes to *hydra*. One is a duplication of exon 1 in the common ancestor of *D. simulans* and *D. sechellia*. Second is a duplication of the whole locus in *D. simulans* and *D. sechellia*. Although uncertainties remain about the structure of the *hydra* region in *D. sechellia*, our phylogenetic analyses suggest that *hydra* duplicated independently in these two species. Third is the further duplication of seven additional exon 1s in *D. melanogaster*. Our phylogenetic analysis suggests that these additional duplication events occurred only in *D. melanogaster*, rather than in the common ancestor of *D. melanogaster*, *D. simulans*, and *D. sechellia*, followed by loss in the latter two species. It is possible, however, that the phylogenetic signal could be erroneous if homogenization occurred by continual gene conversion and/or unequal crossing over among these short duplicated exons.

The duplicated exon 1 in *D. simulans* and *D. sechellia* and most of the additional exon 1 duplications in *D. melanogaster* are flanked by partial DINE-1 elements. DINE-1 is not associated with *hydra* in *D. yakuba* or *D. erecta*. Genome-wide analyses of DINE-1 suggest that DINE-1 was transpositionally active and then silenced in the common ancestor of *D. melanogaster* and *D. yakuba*, followed by another round of activity and silencing in *D. yakuba* [59]. These findings would suggest then that DINE-1 inserted ancestrally in *hydra*, followed by loss in the *D. yakuba* lineage but not the *D. melanogaster* lineage. An alternative scenario is that DINE-1 may have inserted in *hydra* during a period of activity in the *D. melanogaster/D. simulans* ancestor after divergence from *D. yakuba* that was too brief to have left a genome-wide footprint.

Regardless of the timing of the initial DINE-1 insertion into *hydra*, the further duplications of exon 1 in *D. melanogaster* were unlikely to have occurred by DINE-1 transposition because alternative exon 1s two through eight contain UTR and CDSs that are adjacent to rather than within the DINE-1 sequences. As noted above, the phylogenetic relationships among these alternative exons 1s suggest that they originated by unequal crossing over. DINE-1 may have promoted these duplications by increasing the homology among exon 1s. While we lack evidence for a direct role of DINE-1 in driving *hydra* structural evolution, our results do suggest that a property of transposable elements that is usually considered to be deleterious—their ability to cause ectopic recombination and unequal crossing over—may also contribute to structural evolution of genes. We have observed other cases of structural variation associated with DINE-1 insertions in the *melanogaster* species complex (unpublished data). For most genes, evolving under strong functional constraint, a DINE-1 insertion that causes structural changes in the locus is likely to be highly deleterious. However, for a newly evolved gene like *hydra*, there may be more flexibility in terms of tolerating gene structure changes. We also suggest that as a gene evolves in structure, it may accelerate the rate of CDS evolution.

More generally, our evidence that *hydra* has undergone multiple and independent structural changes suggests that gene structure as well as CDSs may evolve rapidly in new genes, even without any association with transposable elements. One other example is the *monkey king* gene family in *Drosophila*, which has undergone rapid structural changes caused by gene fission [35].

**hydra Expression Level Is Evolving**

A second novel aspect compared with other studies of newly evolved genes is that *hydra* expression level has dramatically changed between sibling species. Our quantitative RT-PCR results showed that the expression level of *hydra* between *D. melanogaster* and *D. simulans* is highly different (Figure 6). *hydra* in *D. melanogaster* is expressed at least 20-fold less on average than in *D. simulans*. Why is the expression level so different between these two closely related species?

Two hypotheses may explain the decreased *hydra* expression in *D. melanogaster*. First, it may be caused by competition or interference among the tandemly duplicated heads during the transcription initiation stage. In this model, incomplete transcription initiation complexes would form at multiple heads. Reduced transcription initiation could result from either sequestration of transcription factors or direct interference among promoters. Second, it may be the result of differences in the degree of heterochromatinization of the *hydra* region between *D. melanogaster* and *D. simulans*. *hydra* is located near the pericentromeric region of the X chromosome. This region may be under continuous pressure from encroachment of the nearby heterochromatic pericentromere, leading to species-specific differences in chromatin states. Increased numbers of TE insertions are one characteristic of heterochromatic regions, and the large number of DINE-1 insertions in *D. melanogaster* is consistent with the hypothesis that the *hydra* region is partially heterochromatic.

Our views on the evolution of genome structure and function have changed dramatically in the past decade as more whole-genome sequences have become available in various species. Repetitive elements are a major cause of structural changes and instability in genomes. At the cytological level, chromosome rearrangements often occur at repetitive sites [31,58,59]. Our identification of the *hydra* gene suggests that repetitive elements also contribute to the formation and evolution of new genes by causing structural changes in gene organization, driving the evolution of new exons and transcription start sites, and promoting divergence in gene expression.

**hydra Is a Rare Example of De Novo Gene Creation**

One of our most striking findings is that *hydra* appears to have been created de novo rather than being a duplicate of a pre-existing gene. Two pieces of evidence support this hypothesis. First, *hydra* (and CG1835) are found only in melanogaster subgroup species. The surrounding ~26 kb of DNA is found in other *Drosophila* species that lack *hydra* (Figure 1B). Second, *hydra* shares no homology with any known or predicted proteins. Arguing against this hypothesis of de novo origin is the fact that all forms of *hydra* contain
four predicted exons. One might expect that a recently evolved de novo gene would have a simpler one-exon structure, as has been observed for several candidate de novo genes in D. melanogaster [6]. Nevertheless, if hydra does have an ortholog in Drosophila outside of the melanogaster subgroup, it must have evolved so rapidly as to share no detectable similarity with any predicted proteins, including in multiple Drosophila genomes or other insect genomes that have been completely sequenced. If there is a highly diverged hydra parental gene or ortholog in other Drosophila, it also must have transposed to a novel genomic location in the melanogaster subgroup. Considering all the available data, we suggest instead that hydra evolved de novo from DNA sequence that inserted between run and cpeψ61 in the common ancestor of the melanogaster subgroup.

The Dynamic Nature of New Genes

The most common fate of newly formed genes is disappearance. For example, about 85% of the genes duplicated during the yeast whole-genome duplication have been lost in Saccharomyces cerevisiae [3,60]. New genes therefore can be viewed as analogous to mutations, with most being neutral or deleterious in effect and therefore subject to elimination. A common characteristic of successful new exons and new genes is a high rate of CDS evolution [7,20], and hydra shows this evolutionary pattern. Our discovery of hydra suggests that gene expression level and gene structure may also evolve rapidly and recurrently in new genes. We suggest that the successful retention of new genes requires maximal variation.

Materials and Methods

Fly strains. Flies used in this study include: (1) isofemale lines of D. melanogaster (WI) from Sergey Nuzhdin (University of California Davis, Winters, California, United States); (2) strains of D. melanogaster from worldwide populations: California (CA), Canton S (CS), Taiwan (TWN), Oregon R (ORC), France 3V1 (Fr3V1) (gift of S. C. Tsaur, Academia Sinica, Taiwan); (3) isofemale lines of D. simulans from African populations (mad, sz, su, zk lines; gift of John Pool, Cornell University, Ithaca, New York, United States), and one line each from France (Nantes), Japan (W60); (4) isofemale lines (cy lines) from Africa (gift of Peter Andolfatto, University of California San Diego, United States); and (5) strains of Drosophila species, including D. yakuba (Tai18E2, and 286.82-90), D. erecta, D. ananassae, D. virilis, and D. pseudoobscura (gift of S. C. Tsaur).

DNA preparation and analysis of DNA sequences. DNA was extracted from pools of 50 male flies from each line by a standard phenol-chloroform extraction followed by ethanol precipitation. Primers to amplify and sequence the hydra gene regions from D. melanogaster, D. simulans, and D. yakuba were designed using the homologous sequences obtained from BLAST searches against Flybase genome databases (http://www.flybase.org/blast). Primers for both D. melanogaster and D. simulans are IN1_F (5'-GGCTTTTGTA-CACGTGGCAAA) and 3UR (5'-TAGTGGAAATGGCCTTGC), and for D. yakuba yakF6_E2 (5'-CTTCTAGAACCCCCCAAC) and yake3_R (5'-CAGATTAATGCTCAGATG). In D. simulans, the primer pair specifically amplifies the second duplicate of hydra, which is adjacent to the S3 alternative exon 1 (Figure 1B). Other primer sequences and PCR amplification conditions are available upon request. PCR products were directly sequenced in both direction using ABI BigDye (Applied Biosystems, http://www.appliedbiosystems.com) technologies and modified the manufacturer's suggested protocols. All alleles were sequenced on both strands. Sequences are deposited in GenBank (http://www.ncbi.nlm.nih.gov/Genbank) under accession numbers EF596857–EF596877.

Southern hybridization. Genomic DNA was extracted from 60–70 flies of each line with phenol-chloroform, was ethanol precipitated, and was diluted to a final concentration of about 5 μg/μl. A total of 15 μg DNA was then digested separately with ScaII, separated on a 0.8% agarose gel, and transferred to a nylon membrane by Southern blotting. Probes were from PCR product using DNA template from D. melanogaster with primers hydra_rap500_82 (5'-CAGT(T/C) A G A C C T C T C T G A A T C ) and E2_146R (5'-GGTGGAAATCTCCTGGGTTTGG), and were prepared by DIG-labeling Nick translation kit (Roche, http://www.roche.com). Hybridizations were done at 50°C.

RNA preparation, RT-PCR, 5’-RACE, and molecular cloning. Testes of adult males aged 2–5 d old were dissected in PBS buffer and immediately transferred to PBS buffer on ice. RNAs were extracted with Trizol (Invitrogen, http://www.invitrogen.com). After a brief vortex mixing, the mixture was incubated at 4°C for 15 min and then centrifuged at low speed to obtain phase separation. The aqueous phase was transferred to a fresh tube, to which an equal volume of isopropanol was added in the presence of 3 μl glycolen. The mixture was kept at –20°C for 30 min, followed by centrifugation. The RNA pellet thus obtained was washed with 70% chilled ethanol, repelleted, and air dried. The final RNA pellet was dissolved in DEPC-treated distilled water and was used in at least two RT reactions.

RT was performed using the Superscript II RT kit obtained from Life Technologies (http://www.invitrogen.com) using oligo(dT)12-18 as the primer. The RT products were dissolved in 20 μl sterile distilled water and 1 μl was used for each PCR. For 5’-RACE, first strand synthesis using the lock-docking oligo(dT) primer and 35 cycles of PCR amplification were performed using the SMART RACE CDNA Amplification Kit purchased from Clontech (http://www.clontech.com). Oligos were designed on hydra-specific sequences used in this study (5'-TGGAATTTACGGCGCTCCT), 5’-RACE-derived PCR products were subcloned into the pGEM T vector (Promega, http://www.promega.com) using a standard cloning procedure, and 61 clones were picked randomly for DNA sequencing.

Quantitative PCR. A total of 90 testes of each line were collected from 5-d-old males. Testes were homogenized in Trizol (Invitrogen), and total RNA was phenol-chloroform extracted, ethanol precipitated, and cleaned up with RNasy mini kit (Qiagen, http://www.qiagen.com) and RNase free DNase Set (Qiagen). About 200 ng of total RNA was used for synthesizing cDNA with TaqMan Reverse Transcription Reagents (Applied Biosystems) using oligo-dT as the primer, following the manufacturer's instruction.

Quantitative real-time PCR (qPCR) was performed using the Lightcycler-FastStart DNA Master SYBR Green 1 kit (Roche), followed by the manufacturer's instructions. Primers were 5'-229F (5'-AGTGCCAAAAGTAGTGGAG) and 5'-110R (5'-CTGATTITACCCTACGCT), which are complementary to exon 2 and 3 of hydra, respectively. Note that this primer pair cannot distinguish transcripts derived from the two duplicates of hydra in D. simulans. PCR quality and specificity was verified by melting curve dissociation analysis and gel electrophoresis of the amplified products. Relative transcript abundance of hydra was calculated using the second derivative maximum values from the linear regression of cycle number versus log concentration of the amplified gene.

Amplification of the control gene GAPDH was used for normalization. Sequences of the used primer are: GAPDH F: 5’-AAGG-GAATCCCTGGGACTACG and GAPDH_R: 5’-CGGTTGGAGTACCAAGCTCC.

In situ hybridization. Digoxigenin-labeled antisense and sense control riboprobes were generated from a cDNA containing the sequence of exon 2 and 3 of hydra cloned into a pBluescript vector. Probes were synthesized using DIG RNA labeling mix (Roche), followed by treatment with 2 U DNase (Promega) and incubation in carbonate buffer for 40 min at 65°C as described in [61]. Probes were then ethanol precipitated, air dried, and resuspended in 200 μl hybridization buffer. Probes were used at a dilution of 1:50 during hybridization.

Testes were dissected in 1X PBS and kept on ice until fixation. Two fixations were done with 4% formaldehyde in 1X PBS for 20 min on ice. After washing twice in PBS/0.1% Triton-X, protease treatment and all following steps were done as described in [61]. Hybridizations were done at 60°C overnight.

Identification of hydra orthologs. Orthologs of hydra were from the following prepublication genome sequences: D. simulans and D. yakuba sequence from the Washington University Genome Sequencing Center (WUGSC), D. selysianus sequence from Broad Institute (http://www.broad.mit.edu), and D. erecta sequence from Agencourt Bioscience (http://www.agencourt.com). Sequences of hydra from these species were annotated by maximizing homology to the D. melanogaster sequence, and are shown in Text S1–S6.

Sequence analysis. Sequences were aligned using ClustalW [62], with some manual adjustment to keep gaps in-frame in coding regions. Phylogenetic analyses were performed using Mega 3.1 [63]. Sequence polymorphism and divergence analysis were done using

Evolution of hydra Gene in Drosophila

PLoS Genetics | www.plosgenetics.org July 2007 | Volume 3 | Issue 7 | e1071141
DnaSP 4.10 [64]. The effective number of synonymous and non-synonymous sites, Ka and Ks values, and pairwise divergence for synonymous sites based on Kimura's two-parameter model were estimated.

Supporting Information
Text S1. Annotation in GenBank Format of hydra Region Sequences in D. melanogaster
Found at doi:10.1371/journal.pgen.0030107.sd001 (54 KB TXT).

Text S2. Annotation in GenBank Format of hydra Region Sequences in D. simulans
Found at doi:10.1371/journal.pgen.0030107.sd002 (71 KB TXT).

Text S3. Annotation in GenBank Format of hydra Region Sequences in D. scellica scf 8
Found at doi:10.1371/journal.pgen.0030107.sd003 (42 KB TXT).

Text S4. Annotation in GenBank Format of hydra Region Sequences in D. scellica scf 600
Found at doi:10.1371/journal.pgen.0030107.sd004 (13 KB TXT).

Text S5. Annotation in GenBank Format of hydra Region Sequences in D. yakuba
Found at doi:10.1371/journal.pgen.0030107.sd005 (43 KB TXT).

Text S6. Annotation in GenBank Format of hydra Region Sequences in D. erecta
Found at doi:10.1371/journal.pgen.0030107.sd006 (43 KB TXT).

Text S7. Alignments in FASTA Format of hydra Exons 2–4 Used for the Phylogenetic Analysis in Figure 3
Found at doi:10.1371/journal.pgen.0030107.sd007 (7 KB TXT).

Text S8. Alignments in FASTA Format of hydra Exon 1 (Including Predicted 5' UTRs) Used for the Phylogenetic Analysis in Figure 3
Found at doi:10.1371/journal.pgen.0030107.sd008 (5 KB TXT).

Text S9. Alignment of All DNA Sequences Used in the Polymorphism and Divergence Analyses
Found at doi:10.1371/journal.pgen.0030107.sd009 (84 KB TXT).

Text S10. Alignment of Amino Acid Sequences Based on the DNA Alignment in Text S9
Found at doi:10.1371/journal.pgen.0030107.sd10 (12 KB TXT).

Acknowledgments
We thank Daven Presgraves and Henry Sun for helpful discussions, and Manyuan Long and the anonymous reviewers for insightful comments on the manuscript. We also thank Shun-Chern Tsaur, John Pool, Peter Andolfatto, and Sergeuz Nuzhdin for providing fly strains for population genetics studies, and Hong-Ya Liao for technical assistance.

Author contributions. HPY designed the research. STC, HCC, DAB, and HPY performed the experiments. STC, HCC, and HPY analyzed the data. DAB and HPY wrote the paper.

Funding. This work was supported by a grant (NSC 93–2311-B-010–008) to HPY from National Sciences Council, Taiwan.

Competing interests. The authors have declared that no competing interests exist.

References

1. Jacob F (1977) Evolution and tinkering. Science 196: 1161–1166.
2. Long M, Betran F, Thornton K, Wang W (2003) The origin of new genes: Glimpses from the young and old. Nat Rev Gen 4: 865–875.
3. Wolffe A, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387: 708–713.
4. Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425.
5. Hayashi Y, Sakata H, Makino Y, Umezawa H, Yomo T (2003) Can an arbitrary DNA sequence evolve towards acquiring a biological function? J Mol Evol 56: 162–168.
6. Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ (2006) Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A 103: 9935–9939.
7. Wang W, Zheng H, Yang S, Yu H, Li J, et al. (2003) Origin and evolution of new exons in rodents. Genome Res 15: 1258–1264.
8. Kidwell MG, Lisch DR (2000) Transposable elements and host genome evolution. Trends Ecol Evol 15: 95–99.
9. Kazarian HH Jr (2004) Mobile elements: Drivers of genome evolution. Science 303: 1626–1632.
10. Brookfield JF (2005) The ecology of the genome-mobile DNA elements and their hosts. Nat Rev Gen 6: 128–136.
11. Dawkins R (1976) The selfish gene. New York: Oxford University Press. 224 p.
12. Elliott DJ (2000) An evolutionarily conserved germ cell-specific hnRNP is a transspliced and exhibit testis-biased expression. Proc Natl Acad Sci U S A 103: 13427–13432.
13. Zhang XHF, Chasin LA (2006) Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. Proc Natl Acad Sci 103: 13427–13432.
14. Nozawa M, Aotsuka T, Tamura K (2005) A novel chimeric RNA gene in Drosophila hydra. Science 310: 523–527.
15. Betran E, Thornton K, Long M (2002) Retroposed new genes out of the X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A 103: 13427–13432.
16. Nozawa M, Aotsuka T, Tamura K (2005) A novel chimeric gene, jingwei, a young retroposed gene with 5' UTRs) Used for the Phylogenetic Analysis in Figure 3
Found at doi:10.1371/journal.pgen.0030107.sd007 (7 KB TXT).

Evolution of hydra Gene in Drosophila

37. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, et al. (2005) Evolution of the entire yeast genome. Nature 387: 708–713.
38. van de Lagaemaat LN, Dunn CA, Landry JR, Svenback D, et al. (2005) Impact of transposable elements on the evolution of mammalian gene regulation. Cytogenet Genome Res 110: 542–552.
39. van de Lagaemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet 19: 530–536.
40. Schlenke TA, Begun DJ (2004) Strong selective sweep associated with a transspliced insertion in Drosophila simulans. Proc Natl Acad Sci U S A 101: 1626–1631.
41. Jordan IK, Rogozin IB, Glazko GV, Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19: 68–72.
42. Malik HS, Henkoff S (2005) Positive selection of Iris, a retroviral envelope-derived host gene in Drosophila melanogaster. PLoS Genet 1: e44.
43. Britten RJ (2004) Coding sequences of functioning human genes derived entirely from mobile element sequences. Proc Natl Acad Sci 101: 16825–16830.
44. Nekrutenko A, Li WH (2001) Transposable elements are found in a large number of human protein-coding genes. Trends Genet 17: 619–621.
45. Aminetzach YT, Macpherson JM, Petrov DA (2005) Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. Science 309: 764–767.
46. Gray VH (2000) It takes two transposons to tango: Transposable-element-mediated chromosomal rearrangements. Trends Genet 16: 461–468.
47. Rubin GM, Vandell MD, Worman JR, Gabor Miklos GL, Nelson CR, et al. (2000) Comparative genomics of the eukaryotes. Science 287: 2204–2215.
48. Wang W, Brunet FG, Nevo E, Long M (2002) Origin of sphinx, a young chimeric RNA gene in Drosophila melanogaster. Proc Natl Acad Sci U S A 99: 4418–4423.
49. Betran E, Long M (2003) Dstf-2r, a young Drosophila retrospliced gene with specific male expression under positive Darwinian selection. Genetics 164: 977–988.
50. Wang W, Yu H, Long M (2004) Duplication-degeneration as a mechanism of gene fission and the origin of new genes in Drosophila species. Nat Genet 36: 525–527.
51. Adams M, Celniker S, Holt R, Evans C, Gocayne J, et al. (2000) The genome sequence of Drosophila melanogaster. Science 287: 2185–2195.
52. Quesneville H, Bergman CM, Andrieu O, Auner D, Nouaud D, et al. (2005) Combined evidence annotation of transposable elements in genome sequence. PLoS Comput Biol 1: e22.
53. Kapitonov VV, Jurka J (2005) Molecular paleontology of transposable...
elements in the Drosophila melanogaster genome. Proc Natl Acad Sci 100: 6569–6574.

39. Singh ND, Petrov DA (2004) Rapid sequence turnover at an intergenic locus in Drosophila. Mol Biol Evol 21: 670–680.

40. Yang HP, Hung TL, You TL, Yang TH (2006) Genome-wide comparative analysis of the highly abundant transposable element DINE-1 suggests a recent transpositional burst in Drosophila yakuba. Genetics 173: 189–196.

41. Li YJ, Satta Y, Takahata N (1999) Paleo-demography of the Drosophila melanogaster subgroup: Application of the maximum likelihood method. Genes Genet Syst 74: 117–127.

42. Tamura K, Subramanian S, Kumar S (2004) Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol Biol Evol 21: 36–44.

43. Ollet U, Liao GC, Niemann H, Rubin G (2002) Computational analysis of core promoters in the Drosophila genome. Genome Biol 3: research0087.0012–research0087.0012.

44. Yasuhara JC, DeCrease CH, Wakimoto BT (2005) Evolution of heterochromatic genes of Drosophila. Proc Natl Acad Sci U S A 102: 10958–10963.

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

46. Bachtrog D, Thornton K, Clark A, Andolfatto P (2006) Extensive introgression of mitochondrial DNA relative to nuclear genes in the Drosophila yakuba species group. Evolution Int J Org Evolution 60: 292–302.

47. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351: 652–654.

48. Pregnaghes DC (2005) Evolutionary genomics: New genes for new jobs. Curr Biol 15: R52–R53.

49. Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL (1998) Selective sweep of a newly evolved sperm-specific gene in Drosophila. Nature 396: 572–575.

50. Loppin B, Lepetit D, Dorus S, Couble P, Karr TL (2005) Origin and evolution of a chimeric fusion gene in Drosophila subobscura, D. madeirensis and D. guanche. Genetics 170: 207–219.

51. Arguello JR, Chen Y, Yang S, Wang W, Long M (2006) Origination of an X-linked testes chimeric gene by illegitimate recombination in Drosophila. PLoS Genet 2: e77.

52. Lifschytz E, Lindsley DL (1972) The role of X-chromosome inactivation during spermatogenesis. Proc Natl Acad Sci U S A 69: 182–186.

53. Parisi M (2003) Paucity of genes on the Drosophila X chromosome showing male-biased expression. Science 299: 697–700.

54. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL (2003) Sex-dependent gene expression and evolution of the Drosophila transcriptome. Science 300: 1742–1745.