Evolution of exon–intron structure and alternative splicing in fruit flies and malarial mosquito genomes

Dmitry B. Malko,[1] Vsevolod J. Makeev,[1] Andrey A. Mironov,[1,2] Mikhail S. Gelfand[1,2,3,4]

[1] State Scientific Center GosNIIgenetika, Moscow 117545, Russia; [2] Department of Bioengineering and Bioinformatics, Moscow State University, Moscow 119992, Russia; [3] Institute for Information Transmission Problems RAS, Moscow 127994, Russia

Comparative analysis of alternative splicing of orthologous genes from fruit flies (*Drosophila melanogaster* and *Drosophila pseudoobscura*) and mosquito (*Anopheles gambiae*) demonstrated that both in the fruit fly genes and in fruit fly–mosquito comparisons, constitutive exons and splicing sites are more conserved than alternative ones. While >97% of constitutive *D. melanogaster* exons are conserved in *D. pseudoobscura*, only ~80% of alternative exons are conserved. Similarly, 77% of constitutive fruit fly exons are conserved in the mosquito genes, compared with <50% of alternative exons. Internal alternatives are more conserved than terminal ones. Retained introns are the least conserved, alternative acceptor sites are slightly more conserved than donor sites, and mutually exclusive exons are almost as conserved as constitutive exons. Cassette and mutually exclusive exons experience almost no intron insertions. We also observed cases of interconversion of various elementary alternatives, e.g., transformation of cassette exons into alternative sites. These results agree with the observations made earlier in human–mouse comparisons and demonstrate that the phenomenon of relatively low conservation of alternatively spliced regions may be universal, as it has been observed in different taxonomic groups (mammals and insects) and at various evolutionary distances.

[Supplemental material is available online at www.genome.org.]

Alternative splicing is one of the main mechanisms of creating protein diversity (Kriventseva et al. 2003). Indeed, at least half of human genes (Modrek and Lee 2002) and ~40% of fruit fly genes (Stolc et al. 2003) are alternatively spliced. On the other hand, about a quarter of human elementary alternatives, such as alternative sites and cassette exons, are not conserved in mouse (Modrek and Lee 2003; Nurtdinov et al. 2003), and because of that, about half of alternatively spliced human genes have non-conserved isoforms (Nurtdinov et al. 2003).

The human–mouse comparisons established variation in alternative splicing patterns as one of the main evolutionary driving forces. Indeed, it is a convenient mechanism of trying new proteins without sacrificing old ones: A young alternative isoform would retain a part of the old protein, and if the resulting protein turns out to possess valuable properties, the isoform can be gradually upregulated (Modrek and Lee 2003). Incorporation of introns leading to alternative splicing seems to be a constantly occurring process. For example, analysis of a young human gene family MAGE-A generated by retroposition of a processed, intronless mRNA (Chomez et al. 2001) demonstrated the presence of multiple alternative exons in the 5' region (Artamonova and Gelfand 2004). Similarly, all translated (primate-specific) *Alu* repeats in the human genome are subject to alternative splicing (Sorek et al. 2002; Lev-Maor et al. 2003).

However, until now all such studies relied on comparison of human and rodent (mouse and rat) genomes, and thus the observed trend could be considered as limited to mammalian genomes. Availability of the complete annotated genomes of the fruit fly *Drosophila melanogaster* (Adams et al. 2000; Misra et al. 2002) and the mosquito *Anopheles gambiae* (Holt et al. 2002; Mongin et al. 2004), as well as the genome of *Drosophila pseudoobscura* (Richards et al. 2005) makes it possible to study the evolution of exon–intron structure and alternative splicing in insect genomes. However, unlike human–mouse comparisons, where the average protein identity is ~85% (Makalowski et al. 1996; Waterston et al. 2002) and the exon–intron structure of genes is largely conserved (Batzoglou et al. 2000), and *D. melanogaster–D. pseudoobscura* orthologs, fruit fly and mosquito genes are much more divergent, both in terms of the protein identity, 56% on average, and exon–intron structure, with an average 4.67 introns per fruit fly gene and 3.47 introns per mosquito gene (Zdobnov et al. 2002).

This leads to two complications. First, in addition to simple conservation of alternatives, one needs to take into account possible changes in the exon–intron structure irrespective of alternative splicing. Second, the major technique of such analysis, alignment of isoforms of a relatively well studied organism (here, the fruit fly *D. melanogaster*) to the genome of a less studied species (mosquito), is not straightforward.

To offset this, we considered triples of orthologous genes from *D. melanogaster*, *D. pseudoobscura*, and *A. gambiae*, and mapped *D. melanogaster* protein isoforms to the *D. pseudoobscura* and *A. gambiae* genes using a spliced alignment algorithm. Then we compared annotated *D. melanogaster* exon–exon junctions with projections of predicted *D. pseudoobscura* and *A. gambiae* introns to the *D. melanogaster* isoforms. This allowed us to distinguish non-conserved alternative splicing from changes in the exon–intron structure due to lineage-specific intron insertion or deletion, and thus to describe separately evolution of the exon–intron structure and alternative splicing in the insect genomes.

[4] Corresponding author.
E-mail gelfand@iitp.ru; fax +7-095-2090579.

Results

Genome mapping and classification of *D. melanogaster* alternatively spliced isoforms

Two samples of alternatively spliced *D. melanogaster* (Dme) genes with orthologs in both *D. pseudoobscura* (Dps) and *A. gambiae* (Aga) were considered. The smaller curated sample contained 69 ortholog triples, and the larger non-curated sample contained 920 triples (see Methods).

For each Dme gene, all annotated protein isoforms were aligned to the corresponding genomic region. The regions between thus-defined splicing sites were called "coding segments." Thus, each exon was a union of one or more disjoint coding segments. Genomic regions that aligned to protein in all isoforms were called "constitutive segments," whereas regions that aligned to proteins for some, but not all isoforms, were called "alternative segments." Exons completely consisting of alternative segments 5′ or 3′ of the leftmost (resp., rightmost) constitutive segment were called left (resp., right) "marginal exons."

Exons present in all isoforms were called "constitutive"; by the above definition each such exon consisted of one constitutive coding segment. All other exons were classified by the type of elementary alternatives. This classification was based on comparison of isoform pairs (Supplemental Fig. S1). If an exon from one isoform overlapped with two or more exons of the other isoform, it was called a "retained-intron" exon (**I**). If an exon overlapped one exon from the other isoform, and the isoform-specific region was 5′ or 3′ of the common region, the exon was called an "alternative-acceptor" (**A**) or "alternative-donor" (**D**) exon, respectively. Finally, if an exon had no overlaps with the other isoform, it was called a "mutually exclusive" (**E**) or "cassette" (**C**) exon, depending on the existence (resp., absence) of exons in the other isoform between the common regions closest on both sides of the exon in question. Note that for the comparison of two isoforms, each exon was assigned to only one category, the only exclusion being alternative–donor and alternative–acceptor exons that could be combined. However, as there could be more than two isoforms corresponding to one gene, each exon could be ascribed to several categories.

Alignment of *D. melanogaster* isoforms to the *D. pseudoobscura* and *A. gambiae* genomes

Alignments of Dme mRNA isoforms with Dme, Dps, and Aga genes were used to map exon–exon boundaries (intron shadows) in the Dme isoforms. At that, positions of the annotated Dme intron shadows and projections of the Dps and Aga intron shadows were considered. Each exon and each coding segment (region between adjacent constitutive or alternative Dme intron shadows defined above) was assigned the value of its similarity in the spliced alignment of the Dme protein versus the Dps or Aga genome.

The Dme exons were classified according to the scores of alignments with homologous Dps and Aga genes. If the optimal alignment of an isoform had overall similarity <30%, all exons and coding segments specific for this isoform were considered missing in the respective genome. In addition, an exon or coding segment was considered missing if its similarity in alignments of all isoforms including this exon or segment was <35%, the threshold selected by the analysis of all alignments (see below). A coding segment was considered conserved (present in both ge-

nomes), if in at least some isoform alignment its similarity exceeded 35%. Finally, to account for a possibility of nested structure of alternative exons, exons that contained missing coding segments were treated in a special way, according to the arrangement of conserved and missing segments in these exons (Supplemental Fig. S2). Roughly speaking, an exon was considered missing if it contained missing segments and the part of an exon less these segments coincided with a different exon.

The observed changes in the exon–intron structure are not limited to loss (or birth) of complete exons. Other types of events are deletions (or insertions) of introns. We considered the following types of events, assuming that the basic structure is that of Dme (Supplemental Fig. S5). If a Dme exon coincided with a projection of a Dps or Aga exon, it meant that the exon had been "conserved exactly" (**e**). If a Dme exon aligned with two or more Dps or Aga exons, it could mean intron deletions in the Dme lineage or intron insertions in the Dps (resp., Aga) lineages, and if the similarity in at least two partial alignments of individual Dps (Aga) exons exceeds 35%, such Dme exons were called "divided" exons (**d**). Similarly, if several Dme exons aligned to a Dps or Aga exon, it could mean that Dme exons had merged into one exon in the Dps (Aga) gene by intron deletions, or the Dme exon had been created from one Dps (Aga) exon by intron insertions. A "joining event" (**j**) was recorded if at least two Dme exons aligned to the Dps (Aga) exon with similarity >35%. Finally, if parts of exons of Dme and Dps (Aga) aligned with similarity >35%, but no exon exactly corresponded to the union of exons in the other genome, and at least one remaining part also aligned with >35% similarity, a "mixed" event (**m**) was recorded. Since alignments of individual exons were constrained by alignments of the adjacent exons, definition of events for a particular exon depended on the isoform (Supplemental Fig. S6). However, such events were relatively rare (Supplemental Table S1).

Conservation of *D. melanogaster* isoforms in the *D. pseudoobscura* and *A. gambiae* genomes

Table 1 contains the data about conservation of Dme coding segments in the Dps and Aga genomes. In both the Dme–Dps and Dme–Aga comparisons, the results on the curated and non-curated samples are consistent. In the fruit flies (Dme–Dps), conservation of alternative segments is ~75%–80%, whereas 97% of constitutive segments are conserved, that is, present in both genomes. In the fruit fly–mosquito comparisons, the overall degree of conservation is understandably much lower, but the trend is the same: 77% of constitutive segments are present in both genomes compared with only ~45% of alternative regions.

One minor difference between the curated and non-curated samples is a larger number of left (5′) marginal alternative segments in the latter. One reason for that could be a larger fraction of isoforms incompletely sequenced at the 5′ in the non-curated sample. This would lead to the situation where the leftmost constitutive segment is located deep within the gene, and all segments 5′ of that are considered marginal by definition. In the curated sample in the Dme–Aga comparison, the fraction of non-conserved segments is larger among 5′-marginal segments than among internal alternative segments. This could be explained by faster evolution of protein N termini and problems with the alignment of gene termini, which is not constrained by adjacent regions.

The detailed information about exon fate, dependent on the type of elementary alternatives and the observed changes in exon

Table 1. Conservation of *D. melanogaster* coding segments in the *D. pseudoobscura* and *A. gambiae* genomes

Segments	Missing	Conserved	Total
Dme–Dps, small (curated) sample			
Left marginal (alternative)	11 (22%)	39 (78%)	50 (12%)
Internal alternative	5 (15%)	28 (85%)	33 (8%)
Internal constitutive	7 (2%)	280 (98%)	287 (70%)
Right marginal (alternative)	11 (28%)	29 (72%)	40 (10%)
Total			410 (100%)
Dme–Dps, large (non-curated) sample			
Left marginal (alternative)	244 (21%)	936 (79%)	1180 (23%)
Internal alternative	60 (22%)	212 (78%)	272 (6%)
Internal constitutive	82 (3%)	2853 (97%)	2935 (58%)
Right marginal (alternative)	159 (25%)	477 (75%)	636 (13%)
Total			5023 (100%)
Dme–Aga, small (curated) sample			
Left marginal (alternative)	37 (74%)	13 (26%)	50 (12%)
Internal alternative	18 (55%)	15 (45%)	33 (8%)
Internal constitutive	66 (23%)	221 (77%)	287 (70%)
Right marginal (alternative)	24 (60%)	16 (40%)	40 (10%)
Total			410 (100%)
Dme-Aga, large (non-curated) sample			
Left marginal (alternative)	649 (55%)	531 (45%)	1180 (23%)
Internal alternative	159 (58%)	113 (42%)	272 (6%)
Internal constitutive	705 (24%)	2230 (76%)	2935 (58%)
Right marginal (alternative)	342 (54%)	294 (46%)	636 (13%)
Total			5023 (100%)

(Dme) *D. melanogaster*; (Dps) *D. pseudoobscura*; (Aga) *A. gambiae*.

structure (intron insertion/deletion), is given in Supplemental Table S1 and in histograms (Fig. 1). The Dme–Dps comparison shows that most exons of the fruit flies are conserved, that is, present in both genomes: >97% of internal constitutive exons, 78% of left marginal alternative exons, 84% of internal alternative exons, and 80% of right marginal alternative exons. Exons tend to be completely missing or completely conserved, as very few conserved exons, both constitutive and alternative, have experienced intron insertion/deletion in either lineage (~5%). The least conserved type of elementary alternative is retained introns (~50%), whereas mutually exclusive exons are almost as conserved as constitutive ones (Fig. 1, top histogram). Alternative acceptor sites tend to be slightly more conserved than alternative donor sites.

As has been mentioned already, the overall conservation of the exon–intron structure is much lower in the Dme–Aga comparison. 77% of constitutive exons are conserved, and ~50% are conserved exactly, whereas ~40% of constitutive exons experienced intron insertion/deletion, and the remaining 10% had mixed types of events. At that, ~30% of conserved exons merge with adjacent exons, meaning that an intron was inserted in the fruit fly gene or deleted in the mosquito gene. The opposite type of event, a fruit fly exon divided into two or more in the mosquito gene, happens in ~9% cases, and if the mixed events are allowed, this fraction reaches 17%. This conforms to the observation that mosquito genes tend to have fewer introns than fruit fly genes (Zdobnov et al. 2002).

Left terminal alternative exons are the least conserved (34%), followed by right terminal exons (50%) and internal alternative exons (58%). Again, retained introns are the least conserved type of elementary alternative (Fig. 1, bottom histogram), and again, alternative acceptor sites are slightly more conserved than alternative donor sites. Remarkably, mutually

exclusive exons are almost as conserved as constitutive exons. Both mutually exclusive and cassette exons experienced almost no intron insertion/deletion, as in a vast majority of cases they are either completely missing or completely retained. At that, the fraction of completely conserved mutually exclusive exons is almost twice higher compared with that of completely conserved constitutive exons, although the absolute number of such exons is rather small, and thus the fraction is not very reliable.

Discussion

Most observations made in this study are robust in the sense that they coincide with the curated and non-curated samples of fruit fly genes. The exact reported numbers may depend on a number of arbitrary choices (alignment parameters, similarity threshold, etc.), but all data were studied in a uniform setting, and thus we believe that the main observed trends were captured correctly.

They are also consistent in the sense that the same trends are observed in the Dme–Dps and Dme–Aga comparisons. The most important of these trends is, of course, lower conservation of alternative segments or exons compared with constitutive ones. In both cases, the degree of conservation increases in the order:
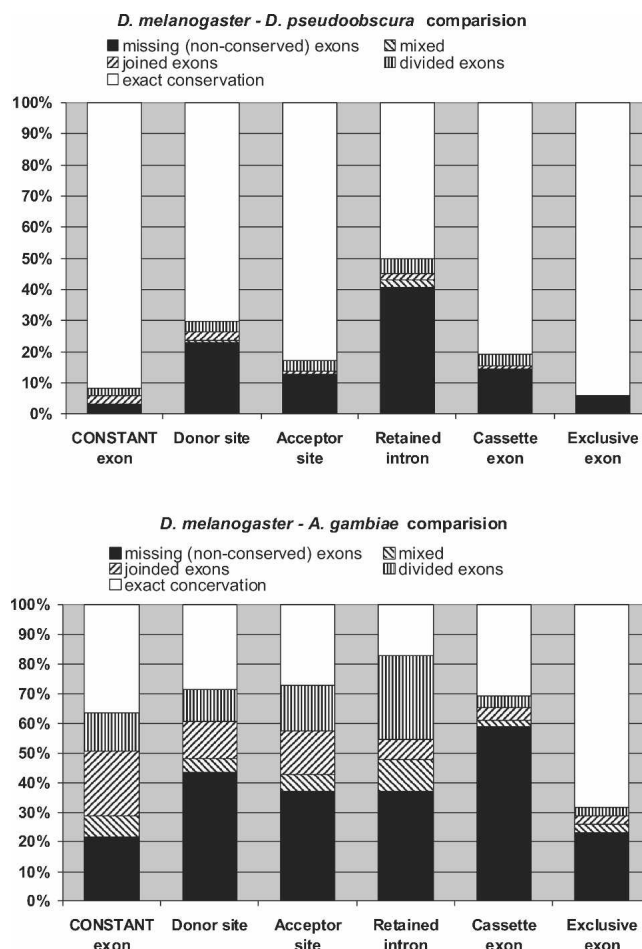

Figure 1. Conservation of various types of *D. melanogaster* elementary alternatives in the *D. pseudoobscura* and *A. gambiae* genomes. Shading in the histogram columns denotes types of events.

constitutive exons, internal alternative exons, right marginal exons, left marginal exons.

The difference in the conservation rate seems to increase with time, as it is more pronounced in the Dme–Aga comparison. The degree of conservation of alternative segments in the fruit flies, 75%–80%, is comparable to the observed degree of conservation of human elementary alternatives in the mouse genome (~75%) (Modrek and Lee 2003; Nurtdinov et al. 2003), and higher than <50% observed in the fruit fly–mosquito comparison. This is consistent with the evolutionary distances (250 million years for Dme–Aga and 75 million years for Dme–Dps and human–mouse genome pairs).

In several cases the observed changes in the exon–intron structure and alternative splicing patterns are quite dramatic, despite a very high degree of conservation of the protein sequence. The gene *CG1517* has the same structure in Dme and Dps. The cassette exon of the fruit flies corresponds to two variants of an exon differing by the choice of the acceptor site (Fig. 2A), and
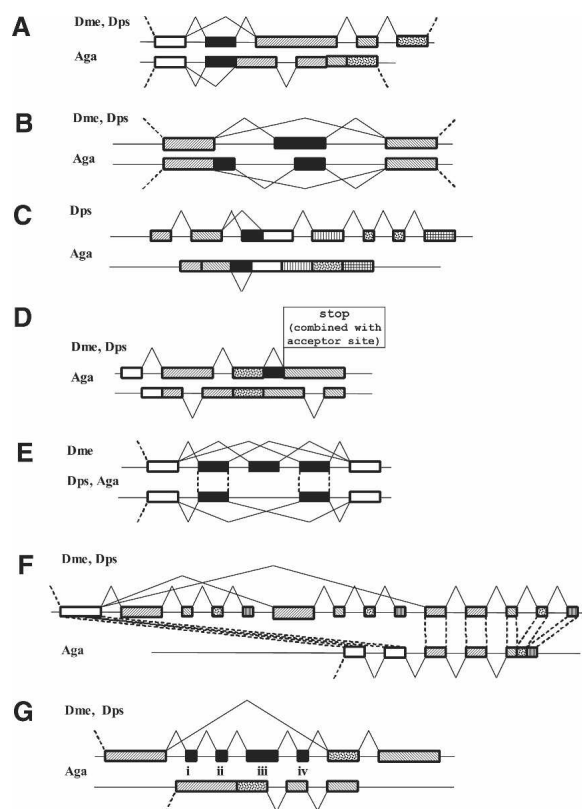
aside from relatively short non-conserved regions on the boundaries of the alternative segment (which may be the remnant of evolutionary events that created this situation), the rest of the protein, including the alternative segment, is highly conserved, although there are additional differences in the structure of the constitutive exons.

In the gene *CG31536*, the alternative is formed by a single cassette exon in the fruit flies, and by a combination of a shorter cassette exon and an alternative donor site in the mosquito gene (Fig. 2B). At that, all encoded protein isoforms are almost identical.

A segment arising from the use of two alternative acceptor sites in fruit flies in the gene *CG1587* corresponds to a retained intron in the mosquito gene (Fig. 2C). The latter contains no introns and, thus, may in fact be a product of retroposed processed mRNA substituting the original intron-containing gene (Weiner et al. 1986). However, the alternative region still may be spliced. This gene also has an additional intron in Dps, as compared with Dme.

These examples demonstrate that there is no major gap between different types of elementary alternatives. Intron deletion may easily convert a cassette exon into an alternative site without sacrificing any isoform: The alternative site would still work with the proximal site to splice an intron, the only difference being that after intron deletion the proximal sites for the two alternative sites coincide. An opposite scenario involving intron insertions seems somewhat less likely, as this insertion should occur exactly at the position of the alternative site. However, this could be the case for the gene *CG1517*, explaining nonconserved, unique regions mentioned above.

Species-specific (or, rather, taxon-specific) alternatives are exemplified by the fruit fly genes *CG1968* and *CG31116*, which contain exons completely missing in the mosquito genome, although the rest of the protein is highly conserved. Again, there are numerous differences in the exon–intron structure of the conserved constitutive part of the gene that can be easily explained by intron insertions and deletions (Fig. 2D,E). In the former gene, an intron inserted in the fruit fly genes may be retained, creating an in-frame stop codon and thus a protein isoform with the truncated C end, whereas in the latter gene, a chain of four cassette exons is inserted in the fruit fly genes (or, alternatively, deleted in the mosquito gene).

Finally, duplications of alternative regions may lead to genes with different sets of mutually exclusive exons. In the gene *CG30427*, three homologous four-exon chains in the fruit fly genes correspond to one such chain in the mosquito gene (Fig. 2F) (the most 3′ of the three copies contains an additional intron in the first exon). A somewhat simpler situation is that of the gene *14-3-3zeta*, which contains three copies of a mutually exclusive exon in Dme, but only two copies in Dps and Aga (Fig. 2G). These observations confirm the importance of exon duplications in the emergence of alternative splicing (Kondrashov and Koonin 2001).

Thus, we have described the major trends in the evolution of alternative splicing in insect genomes and also provided several examples of extreme plasticity of this mechanism of generating protein diversity. It will be instructive to repeat this study on other groups of related genomes (more mammals, birds, and maybe fishes, nematodes *Caenorhabditis elegans* and *C. briggsae*, plants), as well as to study in detail the evolution of alternative splicing of particular genes and multigene families (Artamonova and Gelfand 2004).

Figure 2. Schematic representation of the exon–intron structure and alternative splicing of the *D. melanogaster* genes and their orthologs in *D. pseudoobscura* and *A. gambiae*. Gene names and a brief description of events are given; for a more detailed discussion see the text. Homologous regions are shown by similar shading patterns. (*A*) CG1517: A cassette exon in the drosophilas corresponds to an alternative acceptor site in anopheles; (*B*) CG31536: A cassette exon in the drosophilas corresponds to a combination of a cassette exon and alternative donor site in the anopheles gene; (*C*) GC1587: An alternative acceptor site in the drosophilas corresponds to a candidate retained intron in the anopheles gene (the splicing sites are conserved in a retroposed gene); (*D*) GC1968: Retained intron in drosophilas, missing in the anopheles gene; numerous intron losses/insertions; (*E*) GC31116: Cassette exons in drosohilas that are missing in the anopheles gene; (*F*) GC30427: Duplication of a chain of cassette exons in drosophilas; (*G*) 14-3-3zeta: An additional mutually exclusive exon in Dme.

Methods

FlyBase release 3 (ftp://flybase.net/genomes/Drosophila_
melanogaster/) (Celniker et al. 2002) contained 13,656 *Drosophila
melanogaster* (Dme) genes, of which 1815 were annotated as al-
ternatively spliced in the coding region. Intersection of the latter
set with the list of 6089 orthologous gene pairs of Dme and *Ano-
pheles gambiae* (Aga) downloaded from http://komar.embl.de/
Orthology/ (Zdobnov et al. 2002) produced 1006 pairs; however, for
15 pairs the Aga counterpart was missing in ENSEMBL http://
www.ensembl.org/Anopheles_gambiae/ (Holt et al. 2002; Mongin
et al. 2004). Thus, the preliminary set of Dme–Aga gene pairs with
alternatively spliced Dme gene contained 991 (=1006 − 15) pairs.

The curated FlyBase (http://flybase.net/maps/) set of Dme
genes with alternatively spliced isoforms derived from the litera-
ture contained 75 genes. Of these, 69 had orthologs in Aga. Of
the latter set, 53 genes belong also to the preliminary set of 991
orthologous pairs. The remaining (non-curated) set contained
938 (=991 − 53) gene pairs.

Drosophila pseudoobscura (Dps) genes (Freeze 1) were down-
loaded from the Human Genome Sequencing Center at Baylor
School of Medicine (ftp://ftp.hgsc.bcm.tmc.edu/pub/data/
Dpseudoobscura). Only pairs that had orthologs in Dps were re-
tained (based on the bidirectional-best-hit criterion and the ab-
sence of close paralogs that could obscure the orthology relation-
ships), in particular, all pairs in the curated set. Thus, we
obtained 75 Dme–Dps–Aga ortholog triples in the curated set and
920 triples in the non-curated set.

Further filtering was applied to both sets as follows (Supple-
mental Fig. S7). A gene pair was excluded if the Dme gene did not
contain any constitutive segment (see below). A Dme isoform
was excluded if its annotation contained a non-canonical site
(non-ATG start; stop other than TGA, TAA, TAG; intron starting
at a dinucleotide other then GT or GC or ending at a dinucleotide
other than AG).

The Pro-Frame program (Mironov et al. 2001) was applied to
align Dme protein isoforms with the corresponding genome re-
gions. If the aligned genome region for an isoform contained a
frameshift or an in-frame stop codon, the entire orthologous pair
was excluded.

Then, Dme protein isoforms were aligned to the Dps and
Aga genomes also using Pro-Frame. The following parameters
were used: weight matrix = PAM120, matrix shift = 1, gap pen-
alty = 2, gap initiation penalty = 8, frame-shift penalty = 50, er-
ror site penalty = 50, exon initiation penalty = 20, minimal exon
length = 20, minimal intron length = 40. If no isoform of a Dme
gene could be aligned with the Dps or Aga genome at similarity
>30%, the orthologous pair was excluded. Finally, if the region of
the Dps or Aga genome covered by the constructed alignment
contained a run of N (unknown) nucleotides >20 symbols, the
orthologous pair also was excluded.

After this filtering, a smaller sample of 51 (of initial 69)
curated gene triples and a larger sample of 625 (of initial 920)
non-curated gene triples were obtained.

Acknowledgments

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et
al. 2000. The genome sequence of *Drosophila melanogaster*. *Science*
287: 2185–2195.

Artamonova, I.I. and Gelfand, M.S. 2004. Evolution of the exon–intron
structure and alternative splicing of the MAGE-A family of
cancer/testis antigens. *J. Mol. Evol.* **59:** 609–631.

Batzoglou, S., Pachter, L., Mesirov, J.P., Berger, B., and Lander, E.S. 2000.
Human and mouse gene structure: Comparative analysis and
application to exon prediction. *Genome Res.* **10:** 950–958.

Celniker, S.E., Wheeler, D.A., Kronmiller, B., Carlson, J.W., Halpern, A.,
Patel, S., Adams, M., Champe, M., Dugan, S.P., Frise, E., et al. 2002.
Finishing a whole genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence. *Genome Biol.*
3: research0079.1–0079.14

Chomez, P., De Backer, O., Bertrand, M., De Plaen, E., Boon, T., and
Lucas, S. 2001. An overview of the MAGE gene family with the
identification of all human members of the family. *Cancer Res.*
61: 5544–5551.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R.,
Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., et
al. 2002. The genome sequence of the malaria mosquito *Anopheles
gambiae*. *Science* **298:** 129–149.

Kondrashov, F.A. and Koonin, E.V. 2001. Origin of alternative splicing
by tandem exon duplication. *Hum. Mol. Genet.* **10:** 2661–2669.

Kriventseva, E.V., Koch, I., Apweiler, R., Vingron, M., Bork, P., Gelfand,
M.S., and Sunyaev, S. 2003. Increase of functional diversity by
alternative splicing. *Trends Genet.* **19:** 124–126.

Lev-Maor, G., Sorek, R., Shomron, N., and Ast, G. 2003. The birth of an
alternatively spliced exon: 3′ splice-site selection in *Alu* exons.
Science **300:** 1288–1291.

Makalowski, W., Zhang, J., and Boguski, M.S. 1996. Comparative
analysis of 1196 orthologous mouse and human full-length mRNA
and protein sequences. *Genome Res.* **6:** 846–857.

Mironov, A.A., Novichkov, P.S., and Gelfand, M.S. 2001. Pro-Frame:
Similarity-based gene recognition in eukaryotic DNA sequences with
errors. *Bioinformatics* **17:** 13–15.

Misra, S., Crosby, M.A., Mungall, C.J., Matthews, B.B., Campbell, K.S.,
Hradecky, P., Huang, Y., Kaminker, J.S., Millburn, G.H., Prochnik,
S.E., et al. 2002. Annotation of the *Drosophila melanogaster*
euchromatic genome: A systematic review. *Genome Biol.*
3: research0083.1–0083.22.

Modrek, B. and Lee, C. 2002. A genomic view of alternative splicing.
Nat. Genet. **30:** 13–19.

———. 2003. Alternative splicing in the human, mouse and rat
genomes is associated with an increased frequency of exon creation
and/or loss. *Nat. Genet.* **34:** 177–180.

Mongin, E., Louis, C., Holt, R.A., Birney, E., and Collins, F.H. 2004. The
Anopheles gambiae genome: An update. *Trends Parasitol.* **20:** 49–52.

Nurtdinov, R.N., Artamonova, I.I., Mironov, A.A., and Gelfand, M.S.
2003. Low conservation of alternative splicing patterns in the
human and mouse genomes. *Hum. Mol. Genet.* **12:** 1313–1320.

Richards, S., Liu, Y., Bettencourt, B.R., Hradecky, P., Letovsky, S., Nielsen,
R., Thornton, K., Hubisz, M.J., Chen, R., Meisel, R.P., et al. 2005.
Comparative genome sequencing of *Drosophila pseudoobscura*:
Chromosomal, gene, and *cis*-element evolution. *Genome Res.* **15:** 1–18.

Sorek, R., Ast, G., and Graur, D. 2002. *Alu*-containing exons are
alternatively spliced. *Genome Res.* **12:** 1060–1067.

Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M.F., Rifkin,
S.A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P.E., et al.
2003. A gene expression map for the euchromatic genome of
Drosophila melanogaster. *Science* **306:** 655–660.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F.,
Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.,
et al. 2002. Initial sequencing and comparative analysis of the
mouse genome. *Nature* **420:** 520–562.

Weiner, A.M., Deininger, P.L., and Efstratiadis, A. 1986. Nonviral
retroposons: Genes, pseudogenes, and transposable elements
generated by the reverse flow of genetic information. *Annu. Rev.
Biochem.* **55:** 631–661.

Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M.,
Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A.,
Subramanian, G.M., et al. 2002. Comparative genome and proteome
analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science*
298: 149–159.