Evolution of overlapping genes in Drosophila genomes

Marcin Jakalski, Izabela Makalowska
Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland

INTRODUCTION

Overlapping genes can be defined as a pair of different genes which genomic regions cover to some extent. This phenomenon is often observed in viral, prokaryotic and also eukaryotic genomes. It is considered as a common strategy of genome organization and gene regulation in bacteria. Still increasing number of new evidences suggests that overlapping genes can regulate key processes of gene expression in Eukariota, including genomic imprinting, RNA interference, translational regulation and RNA editing. There are few types of overlapping genes (Fig.1).

Despite large number of these genes their origin and evolution still remains unclear. Basically, there are several mechanisms that explain origin of gene overlap phenomenon. Keese and Gibbs suggested that overlapping genes are created in an overprinting process – new genes are generated from previously existing nucleotide sequences. That’s why one of the genes from overlapping pair is representative of evolutionary and phylogenetically young protein coding genes. Their function, according to the hypothesis is adaptation to present life style of given organism in which particular pairs of overlapping genes are found. Shintani et al. says that overlaps could came into existence after the divergence of mammals from birds in one of two ways - translocation and the rearrangement accompanied by the loss of a part of the 3’ UTR from one gene or alternatively by rearrangement of two previously non overlapping genes. Dahary et al. found that human antisense genes have largely conserved linkage in torafugu so most vertebrates’ overlaps arse much earlier than suggested in previous theories.

AIM

Till now in studies treating of phenomenon of overlapping genes only relatively distant organism were compared (i.e. human and mouse, rice and A. thaliana). In our studies we focused on closely related species. We chose 12 species of Drosophila genus (D. melanogaster, D. pseudoobscura, D. sechellia, D. simulans, D. yakuba, D. erecta, D. ananassae, D. persimilis, D. willistoni, D. mojavensis, D. viridis and D. grimshawi) for their complete genomes became available relatively recently. Entirely studies were done with reference to a set of overlapping genes from Drosophila melanogaster.

We examined both conservation of overlapping genes pair and single genes, being a member of particular pair. Comparative analysis was done in three levels and single genes, being a member of particular pair. We examined both conservation of overlapping genes pair over Drosophila species. There were 61 gene pairs found to overlap within all 12 Drosophila species. These pairs are created by 116 unique genes.

MATERIALS AND METHODS

1. Identification of D. melanogaster overlapping genes pairs using GALAXY tool (http://g2.trac.bx.psu.edu/) - a platform for interactive large-scale genome analysis.
2. Comparative analysis of D. melanogaster and other 11 Drosophilas orthologous genes set (using FlyBase resources).
3. Identification of D. melanogaster and 11 Drosophilas orthologous genes pairs and overlaps conservation.
4. Similarity search (reciprocal BLAST approach) of Drosophila proteins encoded by overlapping genes to six model organisms’ proteomes (mosquito, bee, zebrafish, chicken, mouse, human).
5. Identification of fly-vertebrates and fly-other insects orthologs, orthologous genes pairs and overlaps conservation.
7. Identification of gene pairs conserved across all analyzed species (12 flies, bee, mosquito, zebrafish, chicken, mouse, human).

RESULTS

From originally found 3504 unique genes overlapping in Drosophila melanogaster (Table 1) the largest number of conserved genes (orthologs) was found within D. yakuba – 3022, and least in D. persimilis – 2712 (Fig. 2, Table 2). As compared to the group of model organisms, the highest number of orthologs were observed in mosquito – 2064 and least in mouse – 1147 genes (Fig. 3). Out of the 2001 overlapping genes pairs found in Drosophila melanogaster the highest number of conserved gene pairs in Drosophila genomes were observed in D. yakuba – 1537, and least in D. grimshawi – 1270 (Fig. 2, Table 2). In comparison with vertebrates and other species but are located on different chromosomes. This overlap was lost in higher vertebrates probably due to dosage effect – whole gene pair was first duplicated and then one of the genes from pair was lost.

CONCLUSION

All results show that within overlapping genes there are many formed relatively recently and even among the same species like Drosophila these genes are not conserved. This confirms Keese and Gibbs hypothesis and demonstrates that overprinting is one of the major mechanisms leading to gene overlaps.

There is relatively high number of conserved overlaps in 11 Drosophila genomes compared to 6 analyzed model organisms. There is also relatively high number of conserved overlaps found in mosquito and bee and basically no conserved overlaps found in chicken, mouse and human suggests that overlapping genes could rather be a lineage specific phenomenon and they are representatives of evolutionary young genes. The only gene pair found to be conserved in zebrafish and D. melanogaster could be a representative of ancestral overlap which is the only evidence supporting the Dahary’s hypothesis. This overlap was lost in higher vertebrates probably due to dosage effect – whole gene pair was first duplicated and then one of the genes from pair was lost.

Many genes overlapping in D. melanogaster are present in other species but are located on different chromosomes. This supports to some degree the translation hypothesis by Shintani.

Our results show there are multiple models of gene overlaps evolution and they may not only be created but also lost.

REFERENCES


Fig. 1. Different types of overlapping genes. A. Tail-to-tail overlap in exonic region. B. Head-to-head overlap in exonic C.D. Neural genes. Guaranteed bases mark coding sequence.

Fig. 2. Tree representation of overlapping genes analysis within Drosophila genus. Table 1. Summary of overlapping genes analysis in D. melanogaster.

Fig. 3. Tree representation of D. melanogaster overlapping genes conservation across 6 analyzed model organisms. Only 1 gene pair is conserved across evolutionary lineages. Drosophila melanogaster CG6878 and CG7011 gene pair is conserved in other 11 Drosophila genomes and in mosquito, bee and zebrafish. In all cases one gene is nested within the other (Fig. 4).