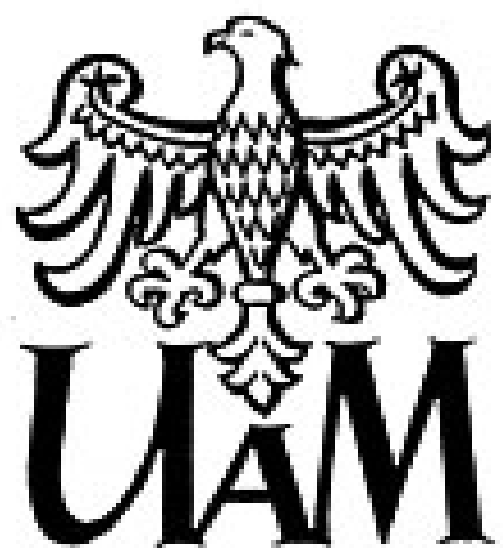# Evolution of overlapping genes in *Drosophila* genomes

Marcin Jąkalski, Izabela Makałowska

Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology,

Adam Mickiewicz University, Umultowska 89, 61-614 Poznań, Poland

Evolutionary Genomics

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.

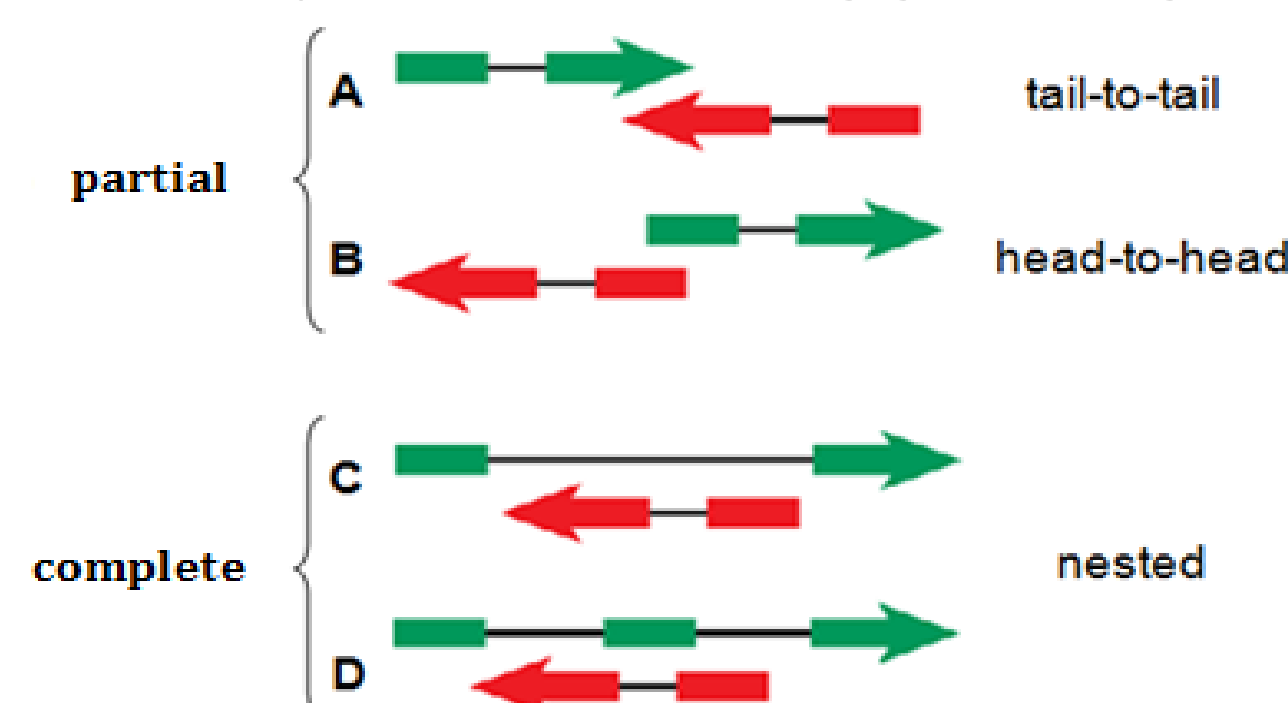There are few types of overlapping genes (Fig.1).

Fig. 1. Different types of overlapping genes. A. Tail-to-tail overlap in exonic region. B. Head-to-head overlap in exonic C.D. Nested genes. Green/red boxes mark coding sequence.

Despite large number of these genes their origin and evolution still remains unclear. Basically, there are several mechanism that explain origination 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 arose 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. virilis* 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 representing different evolutionary distances: for all representatives of *Drosophila* genus, other insects (mosquito and bee), and vertebrates (human, mouse, chicken, zebrafish).

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. Collecting *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.
6. Identification of genes pairs conserved in all 12 *Drosophila* genomes.
7. Identification of genes 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* genome – 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* genus were observed in *D. yakuba* – 1537, and least in *D. grimshawi* – 1270 (Fig. 2, Table 2). In comparison with vertebrates and insect from the group of model organisms these numbers are much lower. The highest proportion of conserved gene pairs is observed in mosquito – 747, and lowest in chicken – 437 genes pairs (Fig. 3).

D. melanogaster protein-coding genes	overlaps found	unique overlaps	unique genes in overlaps	nested overlaps	tail-to-tail overlaps	head-to-head overlaps
21243	3929	2001	3504	838	1017	173

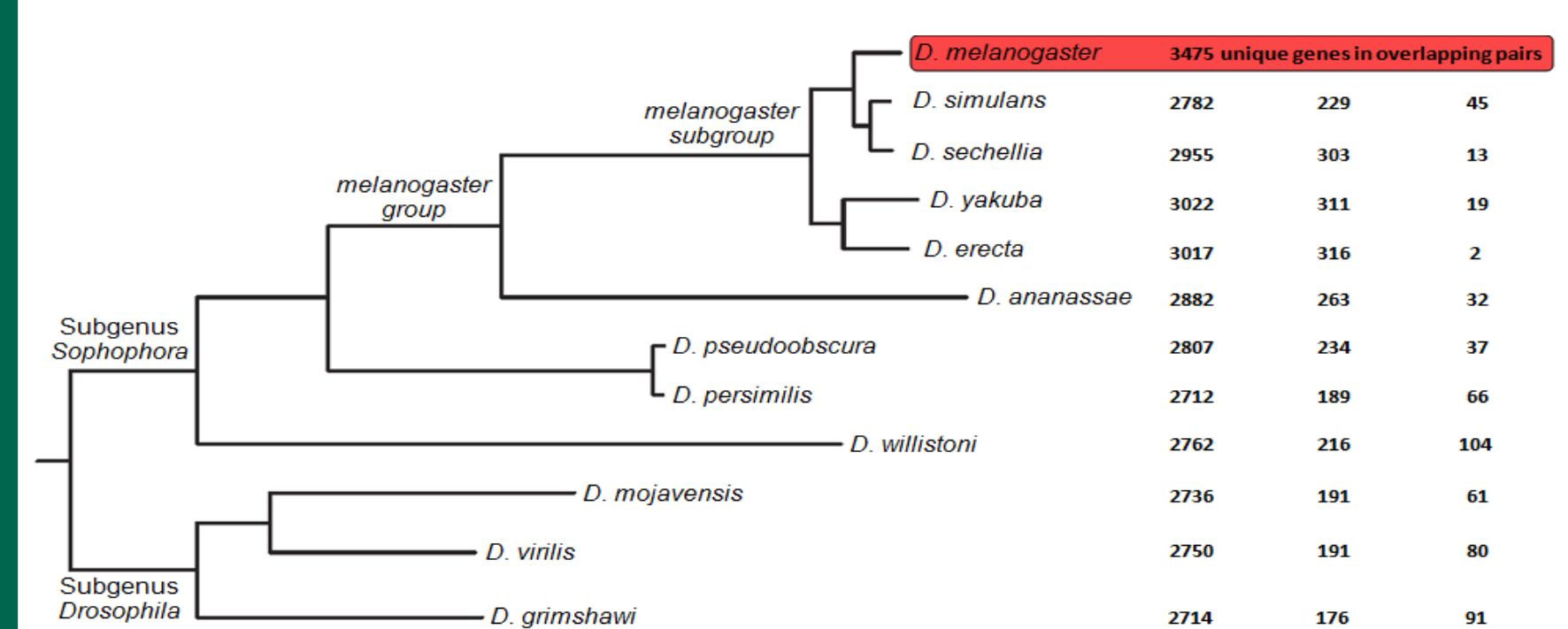*Table. 1. Summary of overlapping genes analysis in D. melanogaster.*

Fig. 2. Tree representation of overlapping genes analysis within Drosophila genus.

Species name	orthologous genes	conserved gene pairs	conserved overlaps	head-to-head overlaps	tail-to-tail overlaps	nested overlaps	chromosome mismatches
D. ananassae	2882	1415	263	4	2	257	32
D. erecta	3017	1525	316	4	6	306	2
D. grimshawi	2714	1270	176	2	3	171	91
D. mojavensis	2736	1284	191	5	0	186	61
D. persimilis	2712	1250	189	4	1	184	66
D. pseudoobscura	2807	1354	234	5	3	226	37
D. sechellia	2955	1467	303	4	5	294	13
D. simulans	2782	1300	229	4	1	224	45
D. virilis	2750	1301	191	0	2	189	80
D. willistoni	2762	1302	216	4	2	210	104
D. yakuba	3022	1537	311	4	3	304	19

Table 2. Summary of 11 Drosophilas overlapping genes analysis. The low number of identified head-to-head and tail-to-tail overlaps is probably due to poor annotation of UTR regions.

There were **61** gene pairs found to overlap within all 12 *Drosophila* species. These pairs are created by **116** unique genes.
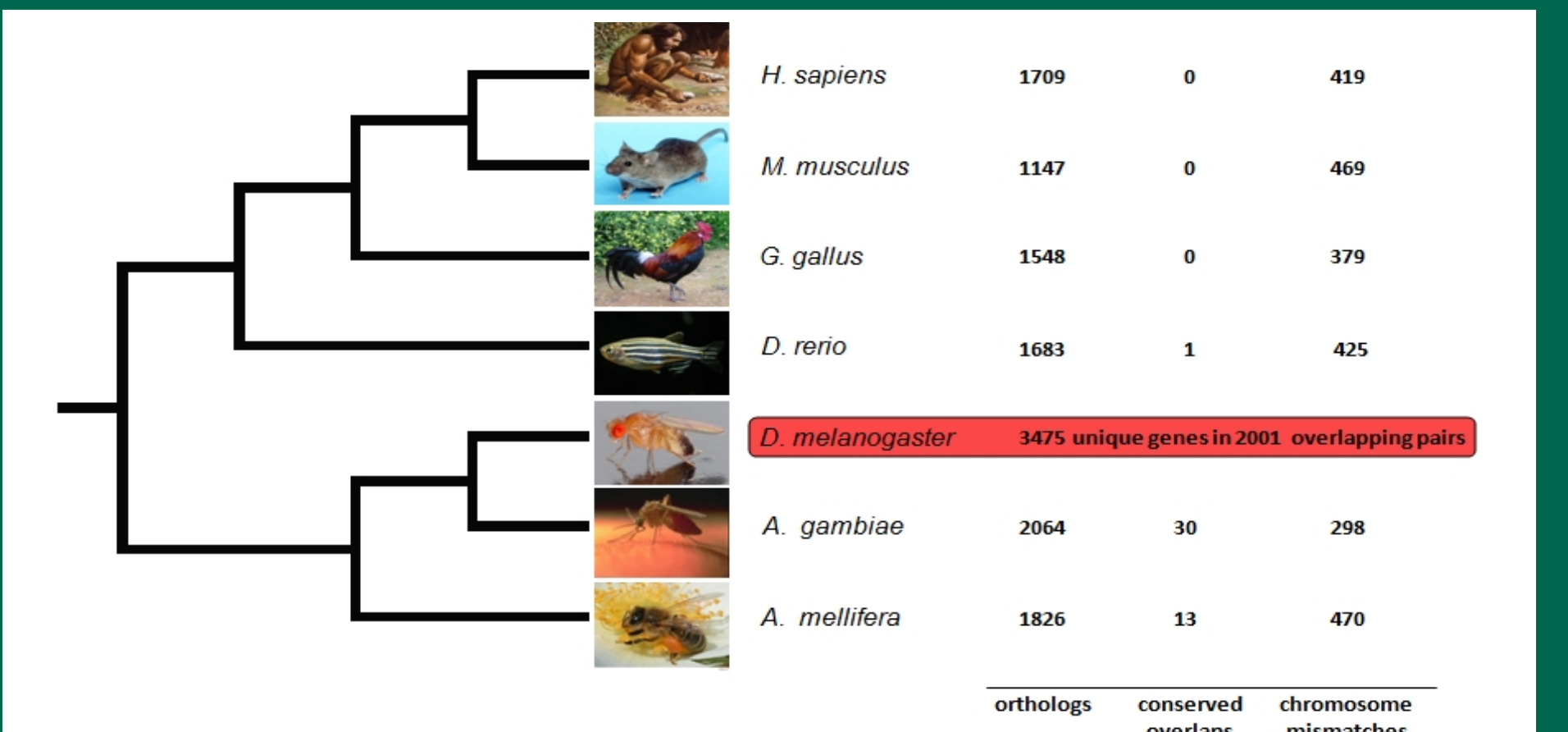

Fig. 3. Tree representation of D. melanogaster overlapping genes conservation across 6 analyzed model organisms.

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).
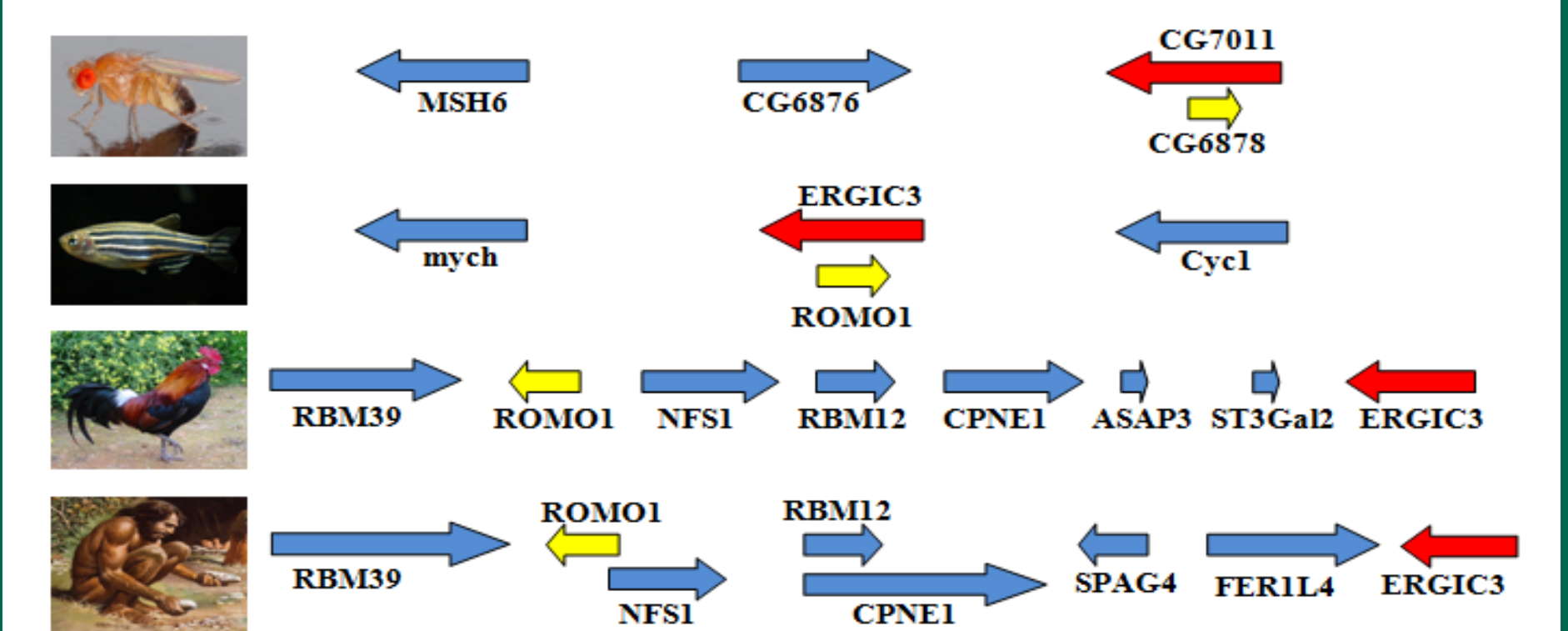

Fig. 4. D. melanogaster CG6878 and CG7011 genes overlap and their orthologs arrangement and overlaps conservation in zebrafish, chicken and human (genes located on chromosome 20).

CONCLUSION

All results show that within overlapping genes there are many formed relatively recently and even among the same genus 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. The low 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 translocation 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

1. Makałowska I, Lin CF, Hernandez K: Birth and death of genes overlaps in vertebrate. *BMC Evolutionary Biology* 2007, 7:193
2. Keese PK, Gibbs A: Origins of genes: "big bang" or continuous creation? *Proc Natl Acad Sci U S A* 1992, 89(20):9489-9493.
3. Shintani S, O'HUigin C, Toyosawa S, Michalova V, Klein J: Origin of gene overlap: the case of TCP1 and ACAT2. *Genetics* 1999, 152(2):743-754.
4. Dahary D, Elroy-Stein O, Sorek R: Naturally occurring antisense: transcriptional leakage or real overlap? *Genome Res* 2005, 15(3):364-368.
5. Drosophila 12 Genomes Consortium: Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 2007, 450:203-218