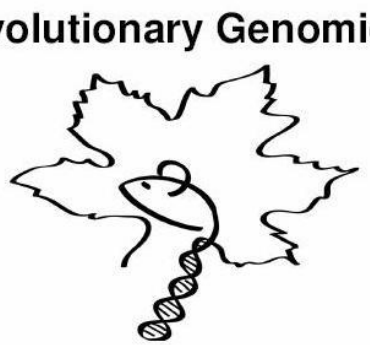


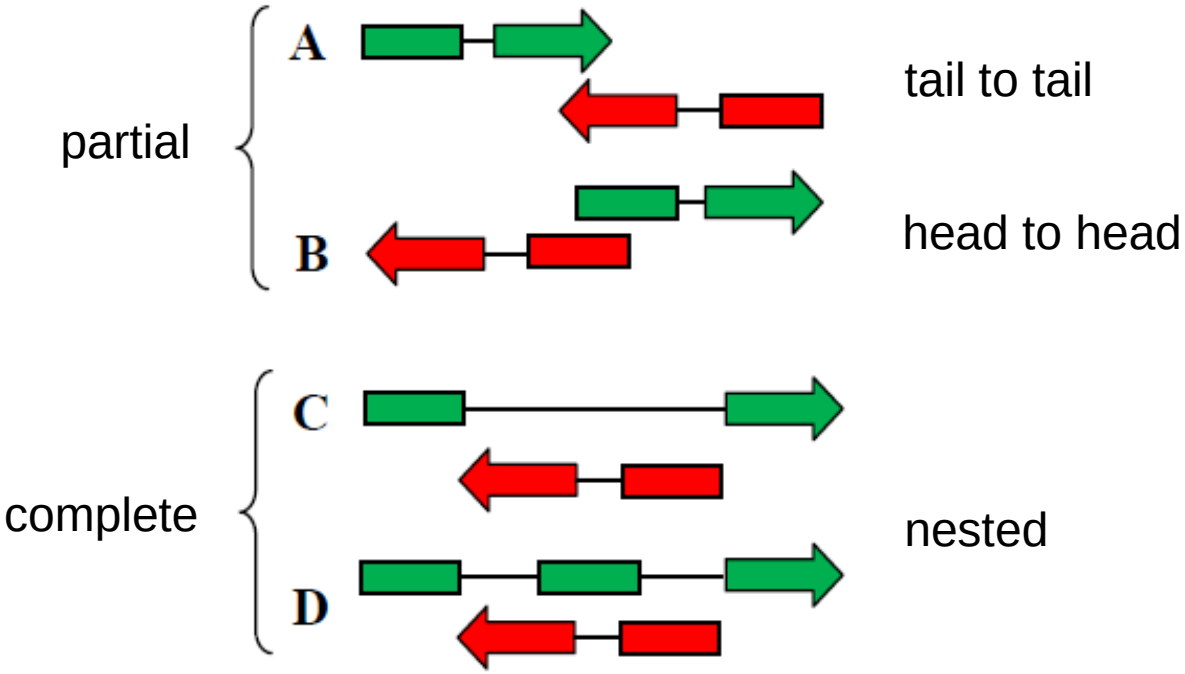
Evolution of overlapping genes in *Drosophila* genomes

Marcin Jąkalski[1,2], Izabela Makałowska[1]

1.Evolutionary Genomics Lab, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznań, Poland
2.Institute of Bioinformatics, Faculty of Medicine, University of Münster, Münster, Germany

Overlapping genes definition

- pairs of different genes, which genomic regions cover to some extent



Numbers

	Number of genes	Number of genes in overlaps	Number of overlaps
Human	22 291	2 978 (**13.36%**)	1 766
Chimpanzee	21 506	2 219 (**10.32%**)	1 276
Mouse	25 383	3 456 (**13.67%**)	2 053
Rat	22159	1080 (**4.87%**)	607
Chicken	17 709	1 960 (**11.07%**)	1 135
Fugu	20796	993 (**4.77%**)	*556*
Zebrafish	23 524	1 625 (**6.99%**)	1 026

Makałowska et al. 2005

Functions

- a strategy of microbial and viral genome organization

bacteriophage ΦX174

- regulation of key processes of gene expression in *Eukariota*

fruit fly

What makes overlapping genes interesting

- In human genome only 3% is occupied by protein coding genes. Many of them share genomic sequence.

- Majority of gene overlaps are not conserved and are lineage specific

Independent evolution or overlap loss?

Hypotheses on overlapping genes evolution

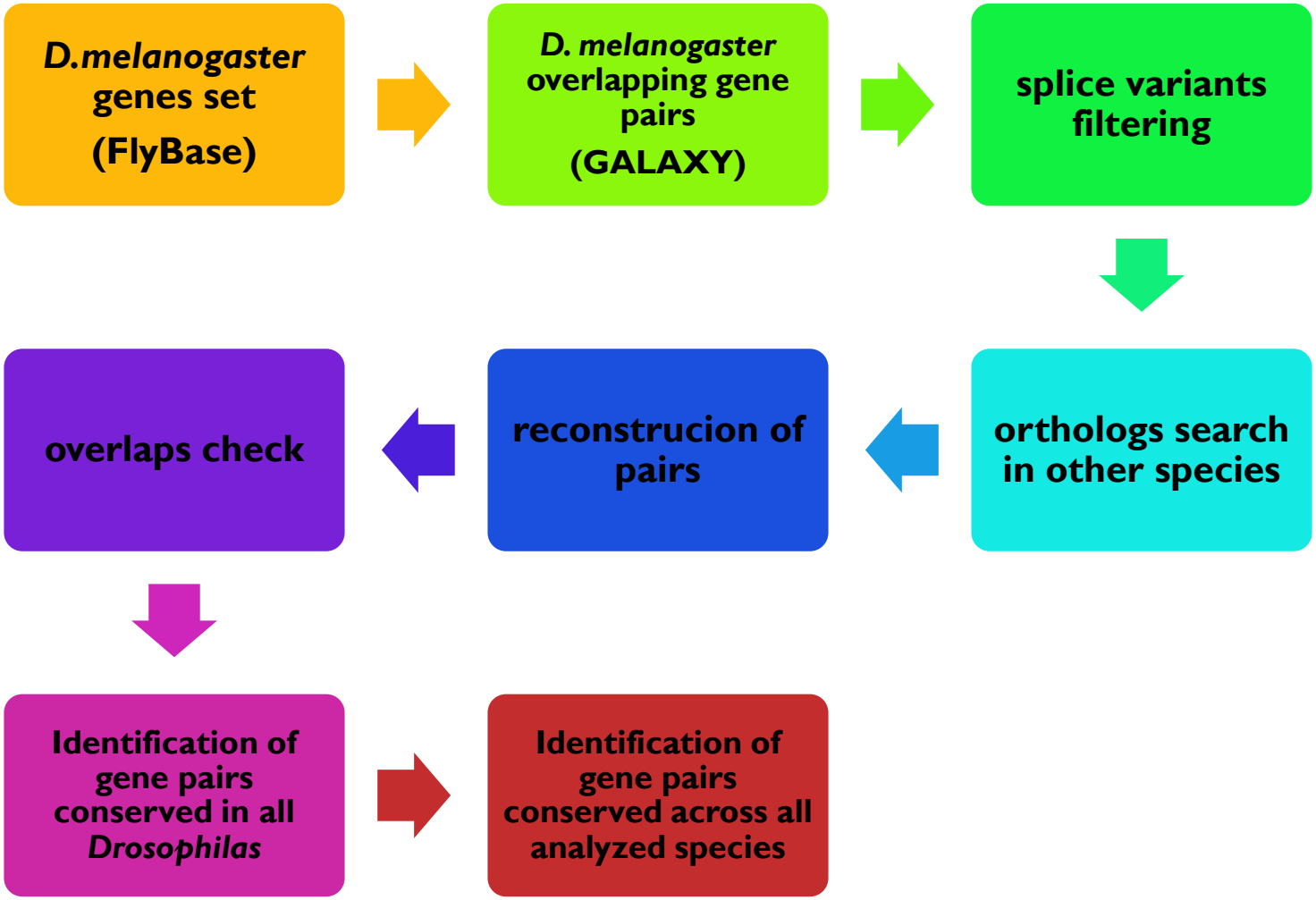
- Keese & Gibbs, 1992, **Origins of genes: "Big bang" or continuous creation?**
  - overprinting
- Shintani et. al., 1999, **Origin of gene overlap: The case of TCP1 and ACAT2**
  - translocations and signal adoption
- Dahary et. al., 2005, **Naturally occurring antisense: transcriptional leakage or real overlap?**

Aims of the project

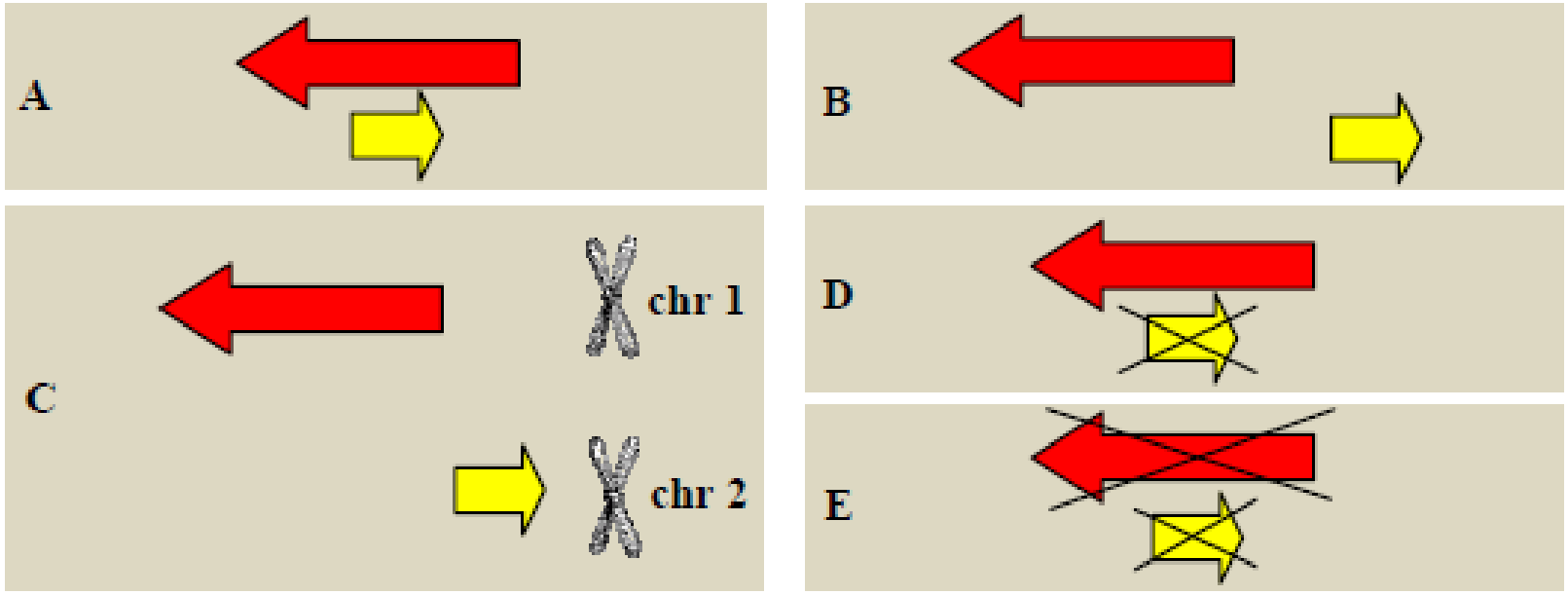
- identification of overlapping genes in 12 species of *Drosophila* genus
  - *D. melanogaster, D. yakuba, D. erecta, D. ananassae, D. willistoni, D. virilis, D. mojavensis, D. grimshawi, D. sechellia, D. persimilis, D. simulans, D. pseudoobscura*

- examination of both conservation of overlapping gene pairs and single genes, being a member of particular pair
  - all representatives of *Drosophila* genus
  - other insects (mosquito and bee)
  - vertebrates (human, mouse, chicken, zebrafish)

Methods


D.melanogaster genes set (FlyBase)
D. melanogaster overlapping gene pairs (GALAXY)
splice variants filtering
orthologs search in other species
reconstrucion of pairs
overlaps check
Identification of gene pairs conserved in all *Drosophilas*
Identification of gene pairs conserved across all analyzed species

Methods

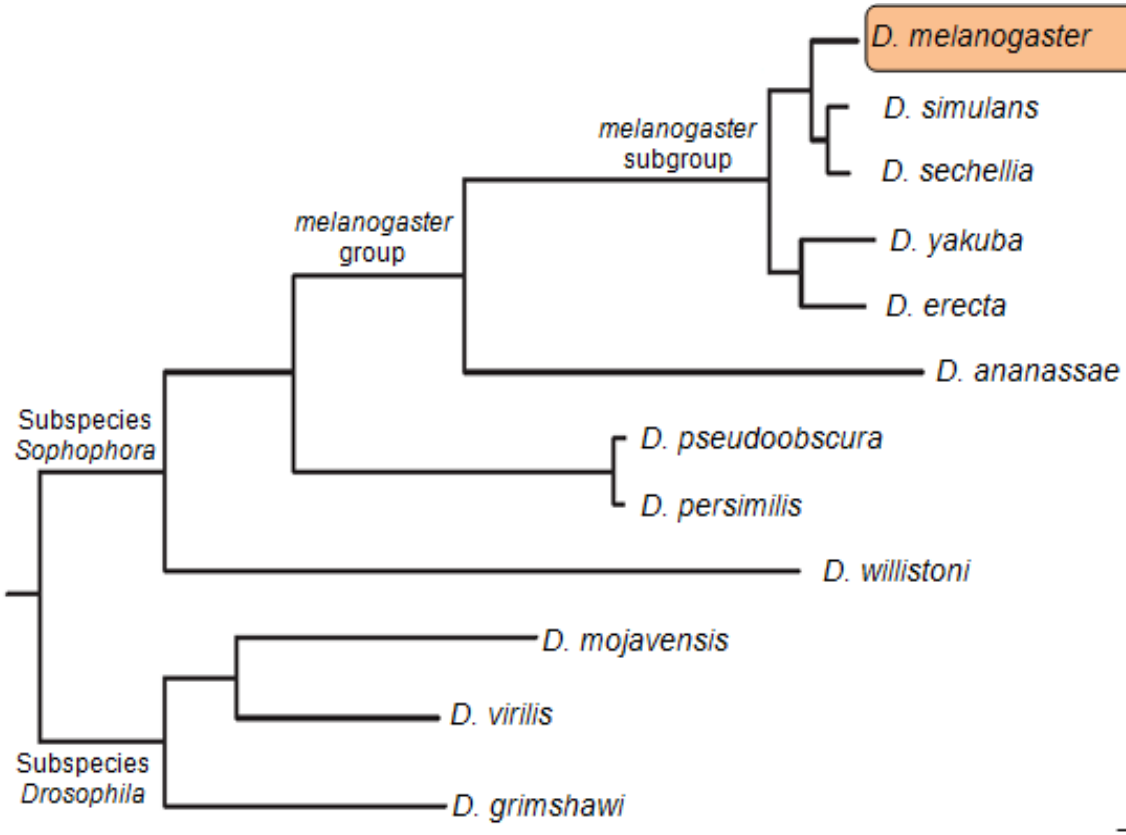


Results - *D. melanogaster*

- **2001** overlapping genes pairs

- **3504** unique genes overlapping in *D. melanogaster*
  - **16,5%** of fruit fly transcripts
  - **823** nested
  - **1007** tail to tail
  - **171** head to head

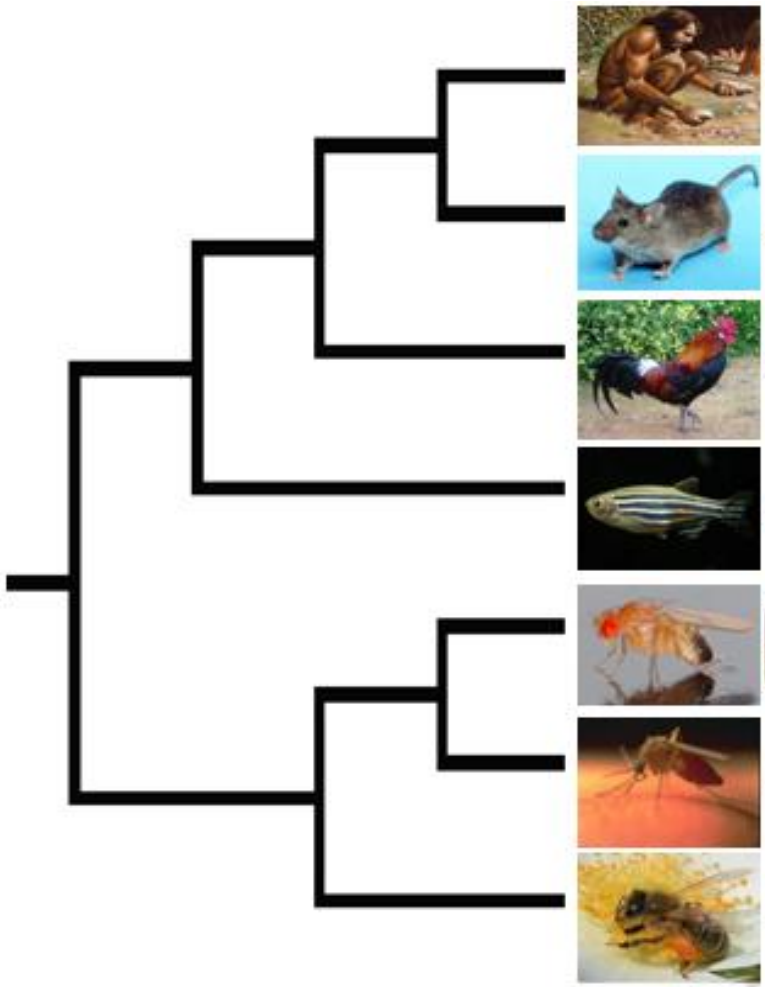
Results - 11 *Drosophilas*



	3504 unique genes in 2001 pairs		
D. simulans	2782	229	45
D. sechellia	2955	303	13
D. yakuba	3022	311	19
D. erecta	3017	316	2
D. ananassae	2882	263	32
D. pseudoobscura	2807	234	37
D. persimilis	2712	189	66
D. willistoni	2762	216	104
D. mojavensis	2736	191	61
D. virilis	2750	191	80
D. grimshawi	2714	176	91
	orthologous genes	conserved overlaps	mismatched chromosomes

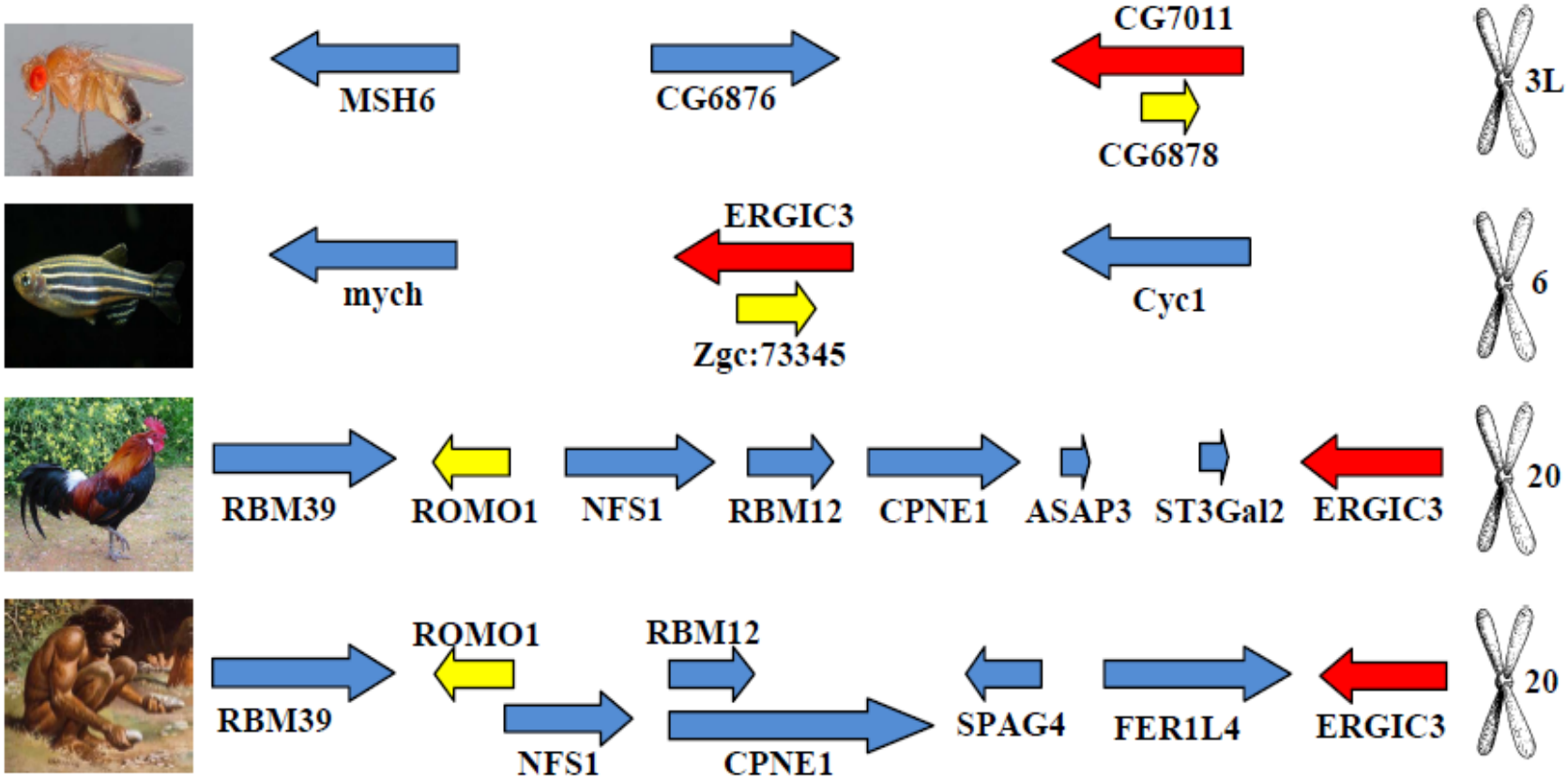
Phylogram from 'Evolution of genes and genomes on the Drosophila phylogeny'. Nature 2007.

Results - 6 model organisms

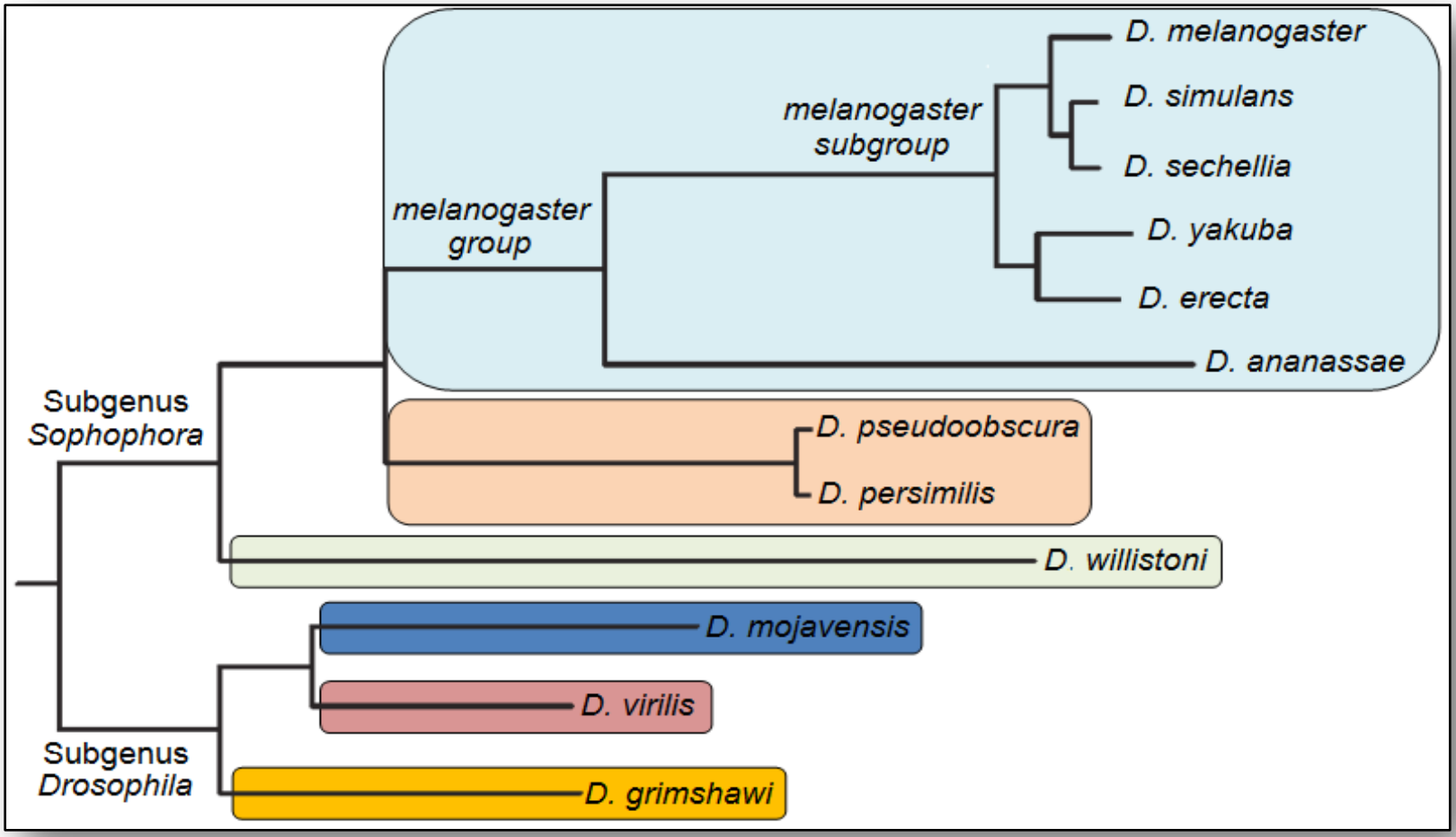


	orthologous genes	conserved overlaps	chromosome mismatches
H. sapiens	1709	0	419
M. musculus	1147	0	469
G. gallus	1548	0	379
D. rerio	1683	1	425
D. melanogaster	3504 unique genes in 2001 pairs		
A. gambiae	2064	30	298
A. mellifera	1826	13	470

Ancient overlaps



Phylogram of the 12 sequenced species of *Drosophila*



Drosophila 12 Genomes Consortium, Nature 2007

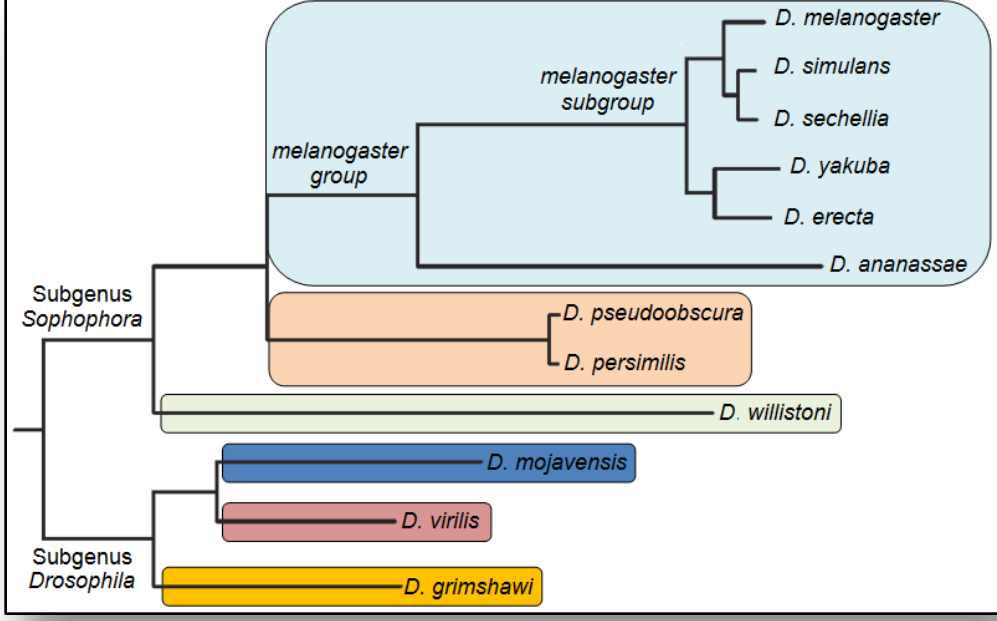
Pairwise genomic mutation distances and the neighbour-joining method.

Matrix of attributes

- presence of orthologs
- overlap conservation
- orthologs on different chromosomes

```
14    823
Dana    00000010001000000000001111111111111100000011111101100110 11
Dere    00000010000001100000001111111111111100000011111101100110 00
Dgri    00000010001000000000000000000000000100000011000011101100
Dmoj    00000010001000000000000000000000000000000010111101100100
Dper    00000000000000000000000100000000000000000010111101100100
Dpse    00000010000000000000011011111111111100010000111111011001 00
Dsec    00000010011001100000001111111111100100000000111111011001 00
Dsim    00000000000000000000000100110000001000000000111101100100
Dvir    00000010000000000000000000000000000100000010111101100100
Dwil    00000010001000000000010011111111110100000010111101100010 0
Dyak    00000000011001100000001011111111111010000001111101100100 00
Dmel    11111111111111111111111111111111111111111111111111111111
Agam    00000000000000000000000000000000000000000000000000000000
Amel    00000000000000000000000000000000000000000000000000000000
```

Orthologs



D. grimshawi (2714)

D. mojavensis (2736)

D. virilis (2750)

D. willistoni (2762)

D. pseudoobscura (2807)

D. persimilis (2712)

D. yakuba (3022)

D. melanogaster (3504)

D. simulans (2782)

D. sechellia (2955)

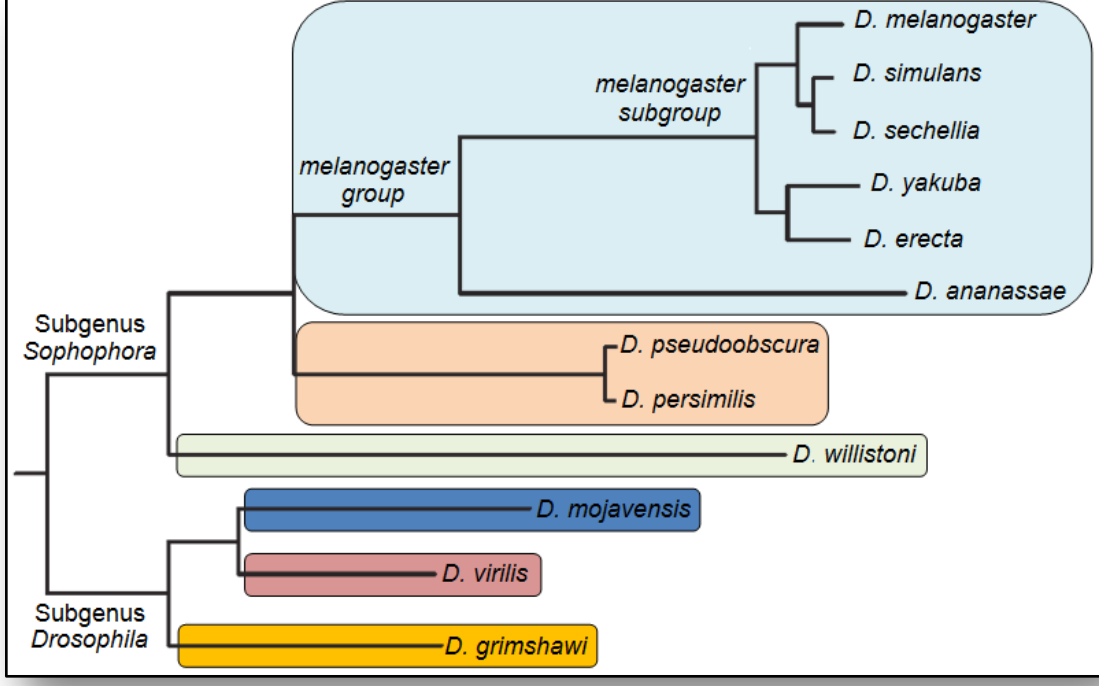
D. erecta (3017)

D. ananassae (2882)

A. gambiae (2064)

200.0

Nested overlaps



D. persimilis (184)

D. grimshawi (171)

D. virilis (189)

D. mojavensis (186)

D. willistoni (210)

D. pseudoobscura (226)

D. simulans (224)

D. sechellia (294)

D. melanogaster (823 nested overlaps)

D. yakuba (304)

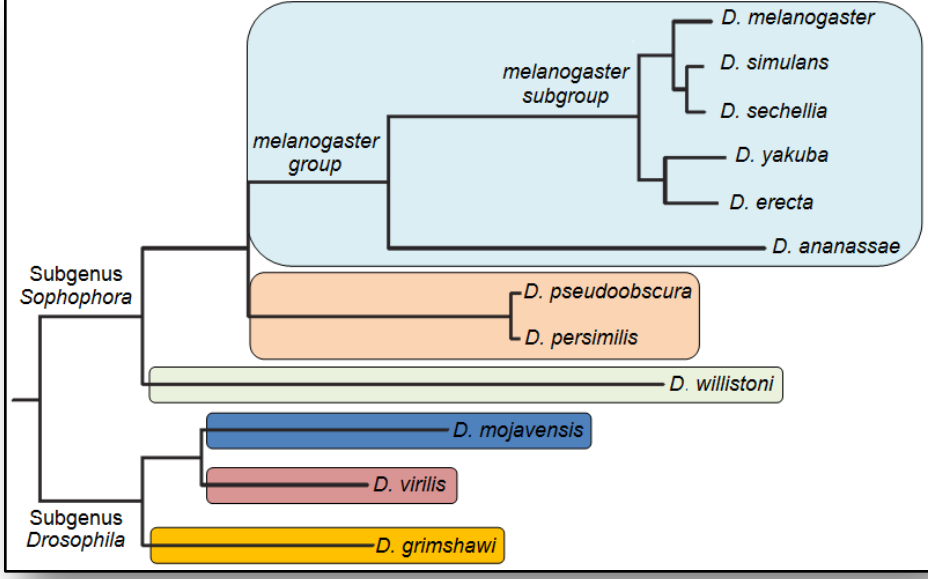
D. erecta (306)

D. ananassae (257)

A. gambiae (22)

80.0

Translocations



A. grimshawi (1270 pairs, 176 overlaps, 91 translocations)

D. mojavensis (1284 pairs, 191 overlaps, 61 translocations)

D. virilis (1301 pairs, 191 overlaps, 80 translocations))

D. willistoni (1302 pairs, 216 overlaps, 104 translocations)

D. pseudoobscura (1354 pairs, 234 overlaps, 37 translocations)

D. persimilis (1250 pairs, 189 overlaps, 66 translocations)

D. yakuba (1537 pairs, 311 overlaps, 19 translocations)

D. simulans (1300 pairs, 229 overlaps, 45 translocations)

D. melanogaster (2001 overlaps)

D. sechelia (1467 pairs, 303 overlaps, 13 translocations)

D. erecta (1525 pairs, 316 overlaps, 2 translocations)

D. ananassae (1415 pairs, 263 overlaps, 32 translocations)

A. gambiae (745 pairs, 30 overlaps, 298 translocations)

200.0

Conclusions - hypotheses evaluation

- high number of nonconserved overlaps
  - overlapping genes could rather be a lineage specific phenomenon - even among the same genus like *Drosophila* these genes are not conserved (overprinting)
- evolutionary old overlap
- orthologous genes on different chromosomes
  - translocations
- one or both genes from orthologous pairs missing
  - gene birth and death hypothesis

There is no single universal model explaining the origination of gene overlap phenomenon. Overlaps may not only be created but also lost!

Acknowledgements

Dr hab. Izabela Makałowska

Evolutionary Genomics Lab

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*

Thank you for your attention.