Evolution of overlapping genes in *Drosophila* genomes

Marcin Jąkalski¹ ², Izabela Makałowska¹

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

Diagram:
- Partial:
  - A (tail to tail)
  - B (head to head)
- Complete:
  - C (nested)
  - D (nested)
## Numbers

<table>
<thead>
<tr>
<th>Species</th>
<th>Number of genes</th>
<th>Number of genes in overlaps</th>
<th>Number of overlaps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>22 291</td>
<td>2 978 (13.36%)</td>
<td>1 766</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>21 506</td>
<td>2 219 (10.32%)</td>
<td>1 276</td>
</tr>
<tr>
<td>Mouse</td>
<td>25 383</td>
<td>3 456 (13.67%)</td>
<td>2 053</td>
</tr>
<tr>
<td>Rat</td>
<td>22159</td>
<td>1 080 (4.87%)</td>
<td>607</td>
</tr>
<tr>
<td>Chicken</td>
<td>17 709</td>
<td>1 960 (11.07%)</td>
<td>1 135</td>
</tr>
<tr>
<td>Fugu</td>
<td>20796</td>
<td>993 (4.77%)</td>
<td>556</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>23 524</td>
<td>1 625 (6.99%)</td>
<td>1 026</td>
</tr>
</tbody>
</table>

Makałowska et al. 2005
Functions

- a strategy of microbial and viral genome organization

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

- 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**
- **overlaps check**
- **reconstrucion of pairs**
- **orthologs search in other species**
- **Identification of gene pairs conserved in all Drosophilas**
- **Identification of gene pairs conserved across all analyzed species**
Methods

A

B

C

chr 1

D

E

chr 1

chr 2
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 Drosophila species


<table>
<thead>
<tr>
<th>Species Chain</th>
<th>Orthologous Genes</th>
<th>Conserved Overlaps</th>
<th>Mismatched Chromosomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>3504 unique genes in 2001 pairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. simulans</td>
<td>2782</td>
<td>229</td>
<td>45</td>
</tr>
<tr>
<td>D. sechellia</td>
<td>2955</td>
<td>303</td>
<td>13</td>
</tr>
<tr>
<td>D. yakuba</td>
<td>3022</td>
<td>311</td>
<td>19</td>
</tr>
<tr>
<td>D. erecta</td>
<td>3017</td>
<td>316</td>
<td>2</td>
</tr>
<tr>
<td>D. ananassae</td>
<td>2882</td>
<td>263</td>
<td>32</td>
</tr>
<tr>
<td>D. pseudoobscura</td>
<td>2807</td>
<td>234</td>
<td>37</td>
</tr>
<tr>
<td>D. persimilis</td>
<td>2712</td>
<td>189</td>
<td>66</td>
</tr>
<tr>
<td>D. willistoni</td>
<td>2762</td>
<td>216</td>
<td>104</td>
</tr>
<tr>
<td>D. mojavensis</td>
<td>2736</td>
<td>191</td>
<td>61</td>
</tr>
<tr>
<td>D. virilis</td>
<td>2750</td>
<td>191</td>
<td>80</td>
</tr>
<tr>
<td>D. grimshawi</td>
<td>2714</td>
<td>176</td>
<td>91</td>
</tr>
</tbody>
</table>

Results - 6 model organisms

<table>
<thead>
<tr>
<th>Organism</th>
<th>Orthologous Genes</th>
<th>Conserved Overlaps</th>
<th>Chromosome Mismatches</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. sapiens</td>
<td>1709</td>
<td>0</td>
<td>419</td>
</tr>
<tr>
<td>M. musculus</td>
<td>1147</td>
<td>0</td>
<td>469</td>
</tr>
<tr>
<td>G. gallus</td>
<td>1548</td>
<td>0</td>
<td>379</td>
</tr>
<tr>
<td>D. rerio</td>
<td>1683</td>
<td>1</td>
<td>425</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>3504 unique genes in 2001 pairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. gambiae</td>
<td>2064</td>
<td>30</td>
<td>298</td>
</tr>
<tr>
<td>A. mellifera</td>
<td>1826</td>
<td>13</td>
<td>470</td>
</tr>
</tbody>
</table>
Ancient overlaps
Phylogram of the 12 sequenced species of Drosophila

Pairwise genomic mutation distances and the neighbour-joining method.

Matrix of attributes

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

<table>
<thead>
<tr>
<th></th>
<th>14</th>
<th>823</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dana</td>
<td>000000100010000000000011111111111111110000001111110110011011</td>
<td></td>
</tr>
<tr>
<td>Dere</td>
<td>000000100000011000000011111111111111110000001111110110011000</td>
<td></td>
</tr>
<tr>
<td>Dgri</td>
<td>000000100010000000000000000000000000100000011000011101100</td>
<td></td>
</tr>
<tr>
<td>Dmoj</td>
<td>000000100010010000000000000000000000000000000010111101100100</td>
<td></td>
</tr>
<tr>
<td>Dper</td>
<td>000000000000000000000000010000000000000000010111101100100</td>
<td></td>
</tr>
<tr>
<td>Dpse</td>
<td>0000000100000000000000000110111111111111100100000011111101100100</td>
<td></td>
</tr>
<tr>
<td>Dsec</td>
<td>0000000010011001100000011111111110100100000011111101100100</td>
<td></td>
</tr>
<tr>
<td>Dsim</td>
<td>000000000000000000000000010011000000100000000111101100100</td>
<td></td>
</tr>
<tr>
<td>Dvir</td>
<td>000000100000000000000000000000000000100000010111101100100</td>
<td></td>
</tr>
<tr>
<td>Dwil</td>
<td>0000000010011001100000001111111111111010000010111101100100</td>
<td></td>
</tr>
<tr>
<td>Dyak</td>
<td>00000000011001100000001011111111111010000001111110110010000</td>
<td></td>
</tr>
<tr>
<td>Dmel</td>
<td>1111111111111111111111111111111111111111111111111111111111111111</td>
<td></td>
</tr>
<tr>
<td>Agam</td>
<td>0000000000000000000000000000000000000000000000000000000000000000</td>
<td></td>
</tr>
<tr>
<td>Amel</td>
<td>0000000000000000000000000000000000000000000000000000000000000000</td>
<td></td>
</tr>
</tbody>
</table>
Orthologs
Nested overlaps

D. persimilis (184)
  D. grimshawi (174)
    D. virilis (189)
    D. mojavensis (186)
    D. willistoni (210)
      D. pseudoobscura (226)

D. simulans (224)
  D. sechellia (294)
    D. yakuba (304)
    D. erecta (306)
      D. ananassae (257)

D. melanogaster subgroup
  D. simulans
  D. sechellia
  D. yakuba
  D. erecta
  D. ananassae

D. pseudoobscura
D. persimilis
D. mojavensis
D. virilis
D. willistoni
D. grimshawi
D. melanogaster (823 nested overlaps)
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. sechelia (1467 pairs, 303 overlaps, 13 translocations)
- D. erecta (1525 pairs, 316 overlaps, 2 translocations)
- D. melanogaster (2001 overlaps)
  - D. ananassae (1415 pairs, 263 overlaps, 32 translocations)
  - A. gambiae (745 pairs, 30 overlaps, 298 translocations)
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

26/11/2008
References


Thank you for your attention.