Poster C-46

In-paralogs analysis of Insecta

Authors:
Stanislav Vershenya (*Bioinformatics, University of Wuerzburg*)
Joerg Schultz (*Bioinformatics, University of Wuerzburg*)

Short Abstract: In this work the in-paralogs of fruitfly, mosquito and honeybee were compared with each other. For every species we were able to identify expanded groups of genes and thus characterized the trends in the adaptational process. Furthermore, we found in-paralogs reflecting adaptations on the genome level of feeding pattern, scent, vision.

Long Abstract:
Comparison of in-paralogs between species can reveal valuable information about the differences in their evolution since the moment of speciation till the present time We performed the genome analyses of three insects with vastly differing lifestyles, namely fruitfly (Drosophila melanogaster), mosquito (Anopheles gambiae) and honeybee (Apis mellifera).
Orthologs detection was based on Inparanoid algorithm. This approach is limited to two species and allows to define the evolutionary point of divergence and thus to separate out-paralogs from in-paralogs.
The clusters of orthologs with a single gene counterpart from each species (â€œone-to-oneâ€ orthologs) are representing 91-92% of the total number of clusters. These clusters represent the â€œcoreâ€ of the last common ancestor of both species. Compared to the total genes amount in Drosophila-Anopheles comparison single gene clusters would represent about 50% (6737 genes). For the comparison of Drosophila and Anopheles to Apis this number drops to about 35% percent of the total gene number. Fruitfly and mosquito are more closely related species and of course would have more gene homologs between themselves than with bee.
Clusters containing at least one species specific duplication can be subdivided into the following three groups:
1. Multiple genes in species one and single gene in species two.
2. Single gene in species one and multiple genes in species two.
3. Multiple genes in both species.
The speciation between Hymenoptera (Apis mellifera) and Diptera has happened about 250 million years (Myr) ago, whereas splitting Drosophila from Anopheles occurred 150 Myr ago. This timeframe allowed us to compare the duplication frequency in Anopheles and Drosophila before speciation and after. Therefore, we listed the genes duplicated in Drosophila compared to Apis and searched for these genes in orthologs groups between Anopheles and Drosophila. If the genes were encountered in groups with duplicated Drosophila genes then the duplications have happened in the last 150 Myr. Contrarily, if the genes were encountered in ortholog groups with a single Drosophila gene the genes duplicated between 250-150 Myr ago. Finally, if the gene was not found at all in Drosophila-Anopheles ortholog groups then orthologs of this gene were either lost in Anopheles or newly invented in Diptera. The same procedure was performed for the Anopheles genome. We found, that in Drosophila 40% (233 genes) of duplications happened

in the common ancestor and the remaining 60% (364 genes) after speciation. For Anopheles these numbers shift slightly to 30% (200 genes) and 70% (522 genes), respectively. In general, these percentages are proportional to the evolution time.
Further, the clusters of orthologs were classified through the Gene Ontology classification: cellular component, cellular process in which they are involved and molecular function. Caused by the annotation constraint, the most informative groups of clusters were those with multiple Drosophila and single Anopheles or Apis genes. These cases might reveal, how the fruitfly evolved, which new proteins and function it acquired in comparison to bee and mosquito, or what was of importance for fruitfly survival and its unique appearance. Similarly, looking at the group of singular Drosophila gene and multiple Apis or Anopheles genes might indicate how bee or mosquito evolved.
Generally, Cell communicating genes, as well as Electron transport genes, were under evolutionary pressure for all of the 3 species. The Electron transport genes play an important role in metabolizing different pathogens and all of three compared species seem to have modified those mechanisms in the past 250 Myr.
In Drosophila compared to Anopheles as well as vice versa duplicated genesâ€™ proteins preferentially belong to the extracellular space. Anopheles, for example, seemed to have gained within the process of evolving the ability to feed on blood a variety of mainly extracellular genes, which prevent platelet and clotting functions and modify inflammatory and immunological reactions in the vertebrate host.
High rates of genes duplications are observed among the genes of Antioxidant activity group. To investigate the molecular function of the duplicated genes in more detail we integrated them in to the cellular network via the KEGG database.
Lots of Drosophila in-paralogs correspond to carbohydrates metabolism: Glycolysis / Gluconeogenesis, Citrate Cycle (TCA), Pentose phosphate pathway, Fructose and Mannose metabolism, Galactose, Starch and Succrose metabolism. Duplication within these genes might reflect the sources of nutrients for Drosophila which are mainly the fruit juices and the yeast growing on rotting fruit.
Also we revealed the duplication of Drosophila genes leading to more precise ultraviolet light absorption. Comparably, another ortholog group with multiple genes from Drosophila and Apis contains genes encoding green-sensitive opsins. This duplication might reflect an adaptation to the insects' need for vision during the day.
The ability to discriminate and respond to chemical signals from the environment is prerequisite for survival and plays extremely important role in the life cycles of Apis, Drosophila and Anopheles. We found 4 ortholog groups containing multiple Drosophila and Anopheles genes encoding odorant receptors (or). The most massively expanded group contained 10 Drosophila and 17 Anopheles genes, encoding the whole range of odorant receptors. Such a relatively high number of duplication events in this gene category might indicate the importance of the odorant receptor genes in the last 150 Myr of evolution in these species. It was also shown that these genes have gone recent duplications in mammals where they are playing an important role in the process of feeding and mating habits.
As a conclusion, we found, that the gene duplication pattern is unique for each species and that this uniqueness is reflected through the differences in functional classes of duplicated genes. The preferences for some classes reflect the evolutionary trends of the last 250 million years and allow assumptions on the role of those genes duplications in the lifestyle of species. Furthermore, the observed gene duplications allowed us to find connections between genomic changes and their phenotypic manifestations. Despite these species

specific differences, we found high correlations between the independently duplicated genes between the species. This might hint for aâ€poolâ€ of genes preferentially duplicated. Taken together, the observed duplication patterns reflect the adaptational process and provide us another link to the field of genomic zoology.