Continuous rapid expansion of the mutually exclusive spliced exome in *Drosophila* species

Klas Hatje and Martin Kollmar

Abteilung NMR basierte Strukturbiologie, Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg 11, D-37077 Göttingen, Germany

mako@nmr.mpibpc.mpg.de

Abstract: Mutually exclusive splicing is an important mechanism in a wide range of eukaryotic branches to expand proteome diversity but the extent of its distribution within a single species and its evolutionary conservation is unknown. Here, we present a genome-wide analysis of mutually exclusive spliced exons (MXEs) in *Drosophila melanogaster* at unprecedented depth. Most of the new MXE candidates are supported by evolutionary conservation, transcriptome data analysis and identification of competing RNA secondary structural elements. The enrichment of the genes with MXEs in transmembrane transporters and ion channel activity is consistent with findings in human, although the MXEs appeared independently and in non-homologous genes, supporting the idea of a universal benefit of adapting ion channel and receptor properties by tandem exon duplications. The comparison of the mutually exclusive spliced exomes within the Drosophila clade shows high numbers of MXE gain and loss events implicating a role of these processes in speciation.

1 Introduction

Alternative processing of primary RNA transcripts is an important driver of increased proteome diversity and regulated gene expression in eukaryotes. Alternative splicing has been reported for alveolates and stramenopiles, green algae and plants, the cryptophyte *Guillardia theta* and the chlorarachniophyte *Bigelowiella natans*, fungi and metazoa, and has therefore been an essential characteristic of the last common ancestor of the eukaryotes. The prevalence of the splice types and the overall number of events strongly differ between branches and species. Mutually exclusive splicing is a particularly interesting

type of generating alternative transcripts: The *Drosophila Down Syndrome Cell Adhesion Molecule* (*Dscam*) gene contains 95 mutually exclusive spliced exons (MXEs) representing the most extensively alternatively spliced gene known. Mutations in MXEs and regions regulating their splicing cause human diseases like the Timothy syndrome, cardiomyopathy or cancer. Mutually exclusive splicing has been shown to be regulated by competing RNA secondary structures. We reported on the continuous gain and loss of MXEs across twelve *Drosophila* species [KH13].

2 Results and Discussion

The algorithm that has been developed for the search for MXEs is based on criteria derived from biological knowledge. A) MXEs must be translated in the same reading frame and the splice sites must be compatible. B) MXEs must have about the same length, because they code for the same structural region in the resulting protein, and length differences are only possible in loop regions. C) The protein sequences coded by the MXEs are supposed to be similar, because they code for the same region in the protein and developed most probably by exon duplication during evolution. As input the software requires the exon-intron structure of the gene. Subsequently the surrounding introns of each original exon are searched for candidates for MXEs. The new algorithm is fully integrated into WebScipio [FO08], the web interface to the Scipio software. Figure 1 shows clusters of MXEs as found in the *DSCAM* gene of *Drosophila melanogaster*.

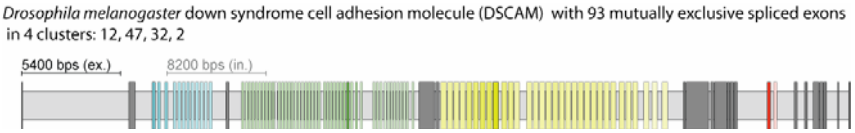

Figure 1. The *DSCAM* gene containing clusters of MXEs (coloured bars). Constitutive exons and introns are denoted by dark grey and light grey bars, respectively.

To characterize the mutually exclusive spliced exome of *Drosophila melanogaster*, we identified 1,297 exons that are mutually exclusive in annotated isoforms of the same gene. Of these 291 had similar length and sequence, including 218 internal MXEs. We predicted 539 exons of similar length and sequence that could be spliced in a mutually exclusive way (two times the annotated exons; Fig. 2). 419 of the MXE candidates were internal including 218 of the already annotated MXEs. Evidence for the predicted MXE candidates was obtained through additional data (Fig. 2): A) Mapping of EST and RNA-Seq data. B) Conservation of the MXE candidates in other

arthropods. C) Ab initio prediction of exonic regions in the respective introns using AUGUSTUS. D) Identification of competing RNA secondary structures. Of the internal MXEs 57% were supported by multiple data types, 21% were supported by EST data. Of the 44 newly predicted internal MXEs eight were supported by EST and/or RNA-Seq data. 94.5% of the annotated and reconstructed internal MXEs and 76.6% of the total predicted internal MXEs are evolutionarily conserved in at least one of the eighteen further analyzed species.

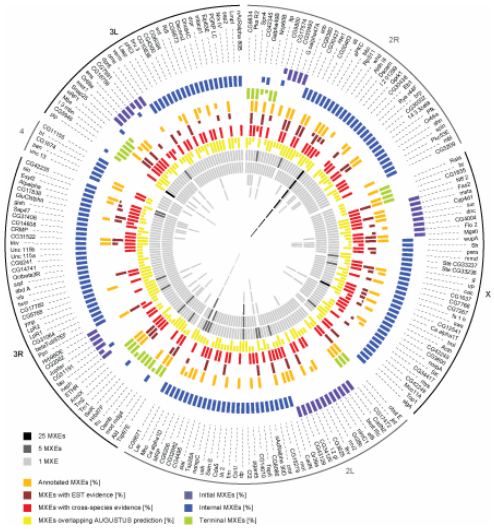

Figure 2. The mutually exclusive exome of *Drosophila melanogaster*. All genes containing predicted MXEs are listed.

In order to determine the extent of conservation within the Drosophila mutually exclusive spliced exomes we compared the data from *D.melanogaster* (dmel) with the reconstructed corresponding exomes of 11 further Drosophila species (Fig. 3A). In total, 2640 clusters were identified most of which are shared among several species, resulting in 770 unique clusters. Surprisingly, many of the clusters are unique to one of these groups like 164 clusters within the Drosophila subgenus group or 95 clusters within the obscura group. Only 68 clusters are conserved in all twelve species. To determine exon gain and loss during the evolution of the *Drosophila* species we counted these events based on maximum parsimony requiring the least exon loss events (Fig. 3B). The last common ancestor of the *Drosophila* species contained at least 186 clusters of MXEs. 456 clusters are unique to any of the *Drosophilas* and 111 clusters have been gained in certain branches.

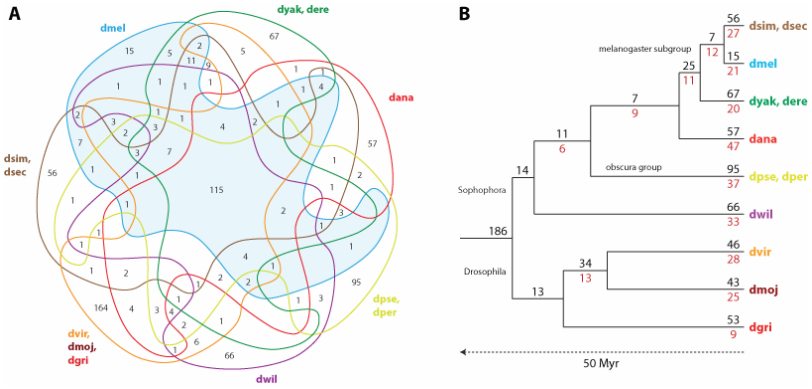

Figure 3. A) The Venn diagrams48 show the number of clusters of MXEs shared between species and subsets of species groups..B) The gain and loss of clusters of MXEs plotted onto the evolutionary tree of the Drosophila species.

3 Conclusion

Our analysis of the mutually exclusive exome of *D.melanogaster* considerably increased the number of mutually exclusive splicing events. Specifically, we have identified two times more internal MXE candidates than already annotated of which almost 80% are supported by evolutionary conservation or experimental transcript data.

References

[FO08] Florian Odronitz, Holger Pillmann, Oliver Keller, Stephan Waack, and Martin Kollmar. WebScipio: An online tool for the determination of gene structures using protein sequences. *BMC Genomics*, 9(422), 2008.

[HK13] Klas Hatje and Martin Kollmar. Predicting Tandemly Arrayed Gene Duplicates with WebScipio. *Nature Communications*, in revision.

[PH11] Holger Pillmann, Klas Hatje, Florian Odronitz, Björn Hammesfahr, and Martin Kollmar. Predicting mutually exclusive spliced exons based on exon length, splice site and reading frame conservation, and exon sequence homology. *BMC Bioinformatics*, 12(270), 2011.