The Molecular Evolution of the Testis TAF Basal Transcription Machinery Genes in Stalk-eyed Flies

Vera Pertsovskaya and Neyanel Vasquez

Mentor: Rick Baker

American Museum of Natural History, Sackler Institute for Comparative Genomics

Introduction

A common phenomenon seen throughout animals is sexual dimorphism--the condition that describes the phenotypic difference between males and females of the same species. A major force behind the evolution of this sexual dimorphism is gene duplication--a critical process in the creation of genes with novel functions and consequently the evolution of biological diversity [1,2]. Gene members of the same duplicate family can have differing expression patterns in males than in females. One area in which gene duplication has had a profound impact on sex differences is the generation of testis-specific genes that are required for spermatogenesis to work [3]. The transcriptomes of *Drosophila* and several vertebrates have shown that testis-specific genes are the most abundant class of tissue-specific genes [4]. One prominent example in *Drosophila* is the TBP-associated factor 5 (TAF5) gene. This gene is an essential component of the basal transcription machinery which controls transcription in all cells and has a duplicate copy in the genome. This duplicate copy is expressed only in the testis and is essential for the progression of the meiotic cell cycle in order for the spermatogenesis process to occur [3,5]. Our goal in this study was to investigate the diversification of TAF5 genes within stalk-eyed flies (*Diopsidae*); preliminary studies have identified several testis specific copies in one species of stalk eyed flies. Flies in this family are known for their elongation of the head into long stalks--a condition known as hypercephaly--and have become a model for studying sexual selection.

Basal Transcription Machinery

• The Basal Transcription Machinery regulates gene expression. This highly conserved mechanism operates in all cells and is constituted of 7 protein complexes which are RpII, TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. The number of genes in each complex varies from 1 to 14. Some of these genes are TAFs which are highly conserved and found in every TFIID complex. There are 12 TAF genes in every TFIID complex of every animal. Of these 12, 5 of them are duplicated genes in *Drosophila*. They are expressed exclusively in the testis and called tTAF 4,5,6,8, and 12 [6].

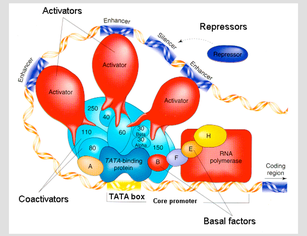

Fig 1. The process that genes must undergo to get expressed by the Basal Transcription Machinery which is represented by the TATA binding protein complex, the five complexes marked A-H and the RHA polymerase complex. TAF5 belongs to the TATA binding protein complex, also referred to as the TFIID complex.

Study Organism

• There are approximately 200-300 described species in 10-14 different genera within the Diopsidae family [7].

•Most species live in the tropics of Asia and Africa.

•Males and females can be distinguished apart by the size of their eye stalk. Males eye stalks are more exaggerated than females.

•Stalk-eyed flies are a model system for sexual selection. Some species are very promiscuous often mating more than 10 times a day. Consequently females tend to carry sperm from various different partners [7]. As a result, during fertilization there is competition between the sperm to fertilize the females eggs.

Fig 2. Females aggregate at night on grass and root hairs and males fight with one another for access to these harems. Matings occur primarily at dawn.

Fig 3. Aggressive interaction between males

Fig 4. A stalk-eyed fly on a leaf.

Questions

•What is the pattern of gene duplication gains and losses within family?

•What is the pattern of selection of tTAF genes ?

•Has there been any movements in the chromosomes for the tTAF genes?

Methods

PCR ANALYSIS
Individual flies (minus heads) were prepped to obtain their DNA using the Qiagen kit. Gene-specific primers were utilized to amplify 2 tTAF paralogs from several different species and populations. PCR products were cleaned using magnetic beads.

GENE FAMILY ANALYSIS
RNA-seq testes transcriptomes from 11 species had been sequenced previously. The results were assembled utilizing Trinity and we blasted the contigs to identify TAF5 genes from these flies. We cleaned the results by removing the untranslated regions. Afterwards we aligned the protein coding regions of the sstalk-eyed fly species and 5 other *Drosophila* and mosquito species using Muscle. Based on this alignment we constructed a phylogenetic tree using Geneious to distinguish the pattern of gene duplication in the family.

Teleopsis dalmanni (Gombak)
T. dalmanni (Soraya)
T. dalmanni (Langat)
T. whitei
T. Thaii
T. quinqueguttata
Diopsis apicalis
Diasemopsis signata
Diasemopsis meigenii
Sphyracephala becarrii
Sphyracephala europa

Table 1. List of stalk-eyed fly species used in the study. Different population of T. dalmanni from south-east asia are indicated in the parentheses.

PROTEIN STRUCTURE
Weused a program called Pfam which allowed us to analyze our protein sequences by letting us look at the domain organization of all of our protein sequences.

MOLECULAR EVOLUTION
To properly analyze the molecular evolution happening in stalk-eyed-flies, we selected the core TAF5 gene and two testis specific duplicates — tTAF5.1 and tTAF5.3. We obtained sequence, in addition to the NGS data, through PCR of other *Teleopsis* species. Introns were removed and the sequences were aligned using a translation alignment. Afterwards we examined the ratio of nonsynonymous and synonymous substitutions (dn/ds) in each paralog. We used SNAP to see their averaged pairwise dn/ds ratios, FUGE to determine the branch specific dn/ds values and DATAMONKEY to locate site specific dn/ds values.

CHROMOSOMAL LOCATION
Genomic hybridization (CGH) of male and female DNA to a microchip suggests the duplicate might have been Y-linked in the species T. thaii. We did PCRs on male and female of this species to see if the tTaf5 was Y-linked. We did this because we wanted to find out if this gene had moved from the X chromosome to the Y chromosome in T. thaii.

Results

Gene Family Evolution:

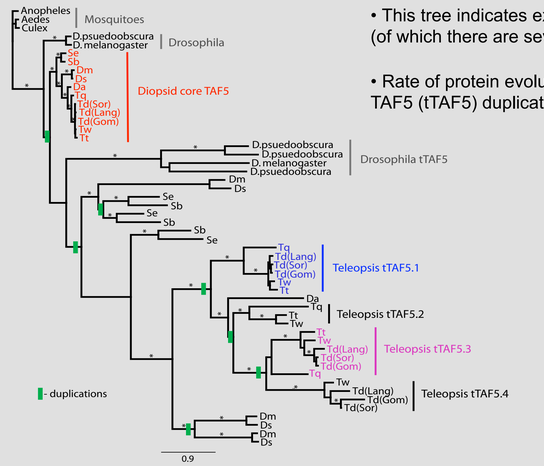

Fig 5. Phylogeny of TAF gene family among Diptera. Bootstrap values greater than 90 are indicated with an asterisk. Green tick marks are indicative of duplication events within Diopsidae. Refer to Table 1 for Diopsidae species names.

• This tree indicates extensive intra-generic duplication with the duplication events (of which there are seven) being concentrated within the *Teleopsis* clade.

• Rate of protein evolution (indicated by branch lengths) is much greater for testes TAF5 (tTAF5) duplicates than core TAF5.

Protein Structure:

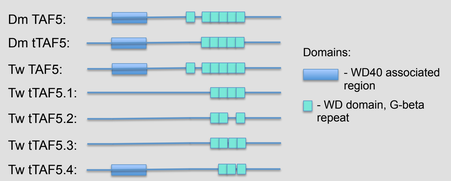

Fig 8. Protein domains of the TAF paralogs. Dm - *Drosophila melanogaster*, Tw – *Teleopsis whitei*.

• The domain organization of the TAF5 paralogs indicates that the testes-specific duplicates have lost several of their proteins domains and, therefore, may have different functional properties.

Molecular Evolution:

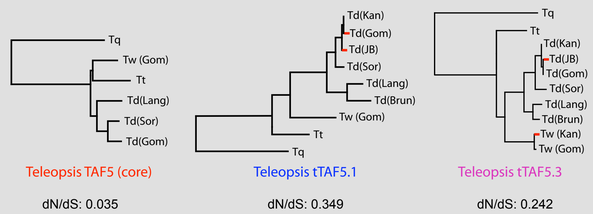

Fig 7. Protein evolution for TAF genes. The average pairwise dN/dS values within each monophyletic *Teleopsis* clade is presented. Branches with a dN/dS ratio greater than 1 are indicated in red.

• The evolution of the testis TAFs have a much lower level of stabilizing selection indicating a relaxation of selection pressures on these genes. The evolution of some branches may be driven by positive selection.

Chromosomal Location:

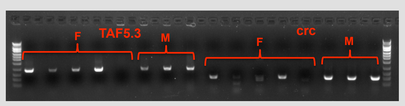

Fig. 8: PCRs done on female (F) and male (M) T. thaii flies for TAF5.3. The gene crc was used as a control.

• The tTAF5.3 PCR works well on females and, therefore, the gene cannot be located on the Y chromosome. This indicates the CGH results are incorrect.

Conclusions

• Our gene family tree shows that several gene duplications have occurred in the family and that there is a high amount of inter-generic gene duplication events occurring, especially in *Teleopsis*.

• The protein structure of all the TAF5 paralogs shows that the duplicates generally have lost a number of important functional domains.

• There is strong stabilizing selection on the TAF5 core gene while there is much weaker selection which allows protein evolution to occur more rapidly in the testes-specific paralogs.

• The TAF5.3 duplicate did not change it's chromosomal location in *T. thaii*.

References

1. Zhang J. 2003. *TREE*, Vol.18 No. 6, 292
2. 2. Demuth P. J, De Bie T, Stajich E. J, Cristianini N, Hahn W. M. 2006. *PloS ONE*, Issue 1.
3. White-Cooper, H. Bausek, N. 2010 *Phil. Trans. R. Soc. B*. 365, 1465-1480.
4. Mikhaylova, M. L, Nguyen, K. Nurminsky, I, D. *Genetics*, 2008, 179
5. Li C. V, Davis C. J, Lenkov K, Bolival B, Fuller T. M, Petrov A. Dmitri. 2009. *Oxford Journals*, Vol. 26. Issue 5.
6. Freiman N. R. *Biochim Biophys Acta*, 2009, Vol. 1789, 161
7. Baker H. R, Narechania A, Johns M. P, Wilkinson S. G. 2012. *Phil. Trans. R. Soc. B*, 367

Acknowledgements

•National Science Foundation
•American Museum of Natural History
•Richard H. Baker
•SRMP
•Hilleary, Nuala, Brian, Samara, Oscar, Sebastian,& Christine
•Our Parents