UNDERSTANDING THE EVOLUTION OF *DOUBLESEX* GENE IN INSECT FAMILY LEPIDOPTERA

Alina Dyachenko, Elena Popova, Polina Trus, Yaroslav Lozinsky, Ekaterina Nuzhdina, Saurav Baral, Riddhi Deshmukh and Krushnamegh Kunte

INTRODUCTION

Doublesex (*dsx*) is a key gene involved in sex determination and sexual differentiation across phyla. dsx is a transcription factor that gets alternately spliced into male- and female-specific isoforms in response to sex-determining cues and regulates the expression of several genes in a sex-specific developmental cascade. The *doublesex* protein contains two domains, a DNA binding domain (characteristic of all DMRT genes) and a dimerization domain (that contains sex-specific and non-sex-specific sequence). *Dsx* regulates the expression of several genes involved in male and female reproductive functions such as, development of genitalia, yolk proteins, deposition of fat bodies in females, etc. However, in the past decade, several studies have found *doublesex* being co-opted in regulating sexually dimorphic phenotypes and behaviour in various insects. In *Papilio polytes* butterflies for example it is involved in formation of female-limited mimetic phenotypes, whereas, in *Bombyx mori* it is involved in pheromone production. *dsx* gene has achieved this versatility by means of a) alternate alleles, b) sex-, tissue-, and form-specific expression and c) differential expression levels in different tissues, at critical developmental stages. For a dynamic, yet essential gene like *dsx* to accommodate such varied roles while maintaining its original function, its critical domains must remain largely unaltered, while the rest of the sequence is malleable to evolutionary change. In an attempt to understand how *dsx* achieves this, we compared the sequences of *dsx* trancripts from 5 lepidopteran species (butterflies and moths) and found sites that have undergone synonymous and non-synonymous substitutions in each exon. We also performed a larger comparison of male-specific transcripts from 11 species to identify hotspots of mutations. Lastly we constructed a gene tree to understand the relationships between *dsx* transcripts from these 11 species and how they differ from their actual phylogenetic distances.

METHODS

Finding doublesex gene in the genome:

We downloaded genomes of lepidopterans with transcriptome information from NCBI. Using *doublesex (dsx)* protein sequence from *Papilio polytes* we extracted for the *dsx* gene from our downloaded genomes with BLAST 2.6.0+.

Extracting exons:

Using Integrative Genomics Viewer (IGV, ver. 2.3.94) we found coordinates and locations of each exon for different *dsx* transcripts. We aligned exon sequences using MEGA 7 (both DNA and protein alignments) and used the alignment to calculate the number of synonymous and non-synonymous substitutions with a Python script. We prepared the figure showing conserved and variable sites using Skylign tool.

Looking for doublesex across many species:

We downloaded additional male specific *dsx* sequences to identify synonymous and non-synonymous substitutions as mentioned above. We also used these sequences to construct a phylogenetic tree for these 11 species.

Demonstrative practicals on how to obtain sequence data:

- Dissection of wing tissue from developing pupae
- Extraction of DNA from tissues using ExtractMe DNA extraction kit
- PCR amplification of 16S ribosomal gene from extracted DNA.
- Visualization of amplified fragments using gel electrophoresis.

RESULTS

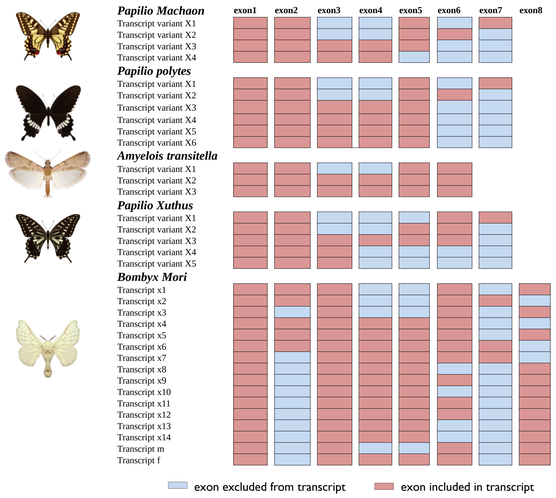

Fig.1 Exon usage in *doublesex* transcripts of 5 lepidopteran species

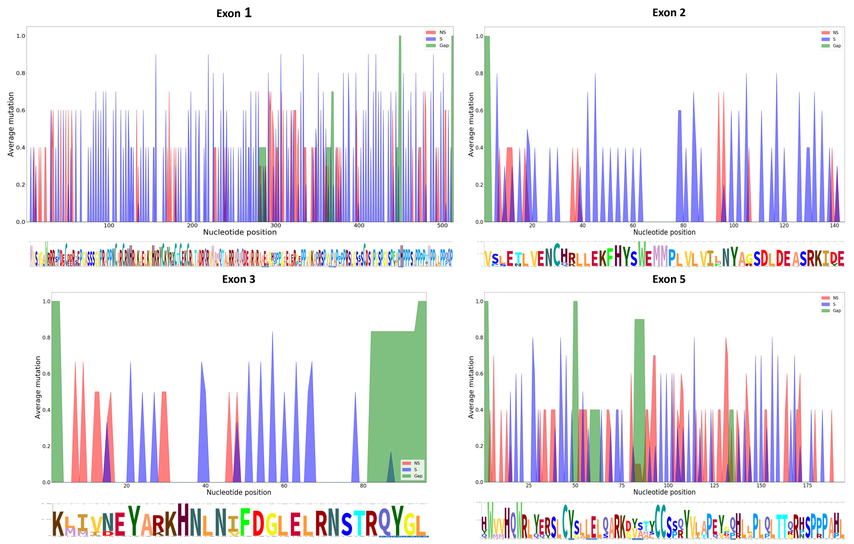

Fig.2 Average mutation in exons and amino acid conservation. X axis - nucleotide position, Y axis - average mutation

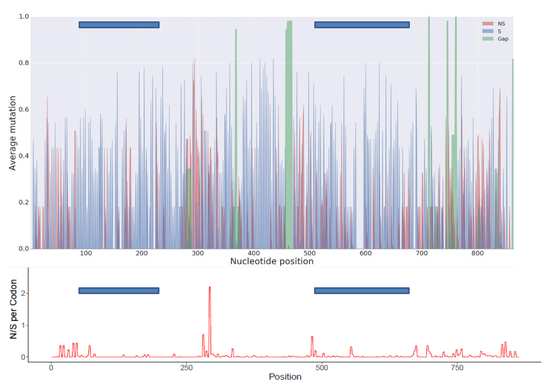

Fig. 3 Average mutation in male-specific *dsx* transcripts. Blue lines represent domains (DM and dimerization) of the *dsx* protein. The lower panel depicts the ratio of non-synonymous to synonymous substitutions in *dsx* proteins from 11 species of lepidopterans.

RESULTS

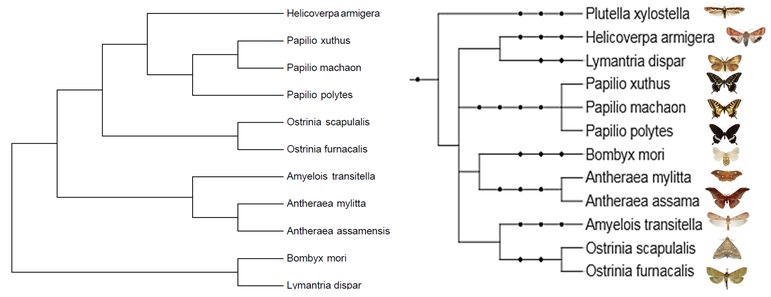

Fig. 4 Gene tree constructed using male-specific *dsx* sequences (left) and species tree depicting phylogenetic relationships between the 11 lepidopteran species.

	Ostrinia scapulalis	*Ostrinia furnacalis*	*Lymantria dispar*	*Helicoverpa armigera*	*Antheraea mylitta*	*Antheraea assama*	*Papilio xuthus*	*Papilio polytes*	*Bombyx mori*	*Papilio machaon*	*Amyelois transitella*
Ostrinia scapulalis	0										
Ostrinia furnacalis	0,007	0									
Lymantria dispar	0,093	0,095	0								
Helicoverpa armigera	0,038	0,044	0,088	0							
Antheraea mylitta	0,054	0,059	0,101	0,053	0						
Antheraea assama	0,053	0,057	0,101	0,052	0,009	0					
Papilio xuthus	0,054	0,056	0,106	0,061	0,072	0,072	0				
Papilio polytes	0,072	0,072	0,126	0,081	0,088	0,087	0,027	0			
Bombyx mori	0,112	0,113	0,14	0,119	0,131	0,123	0,12	0,13	0		
Papilio machaon	0,052	0,052	0,111	0,061	0,074	0,073	0	0,027	0,119	0	
Amyelois transitella	0,041	0,045	0,105	0,047	0,064	0,065	0,073	0,087	0,118	0,073	0

Fig. 5 Pairwise comparison of non-synonymous substitutions between male-specific *dsx* transcripts.

DISCUSSION

The *doublesex* gene has numerous transcripts across different species owing to alternative splicing. These transcripts may perform different functions in different organs and body parts of the butterfly, however, the differential expression of isoforms and their functions have not yet been studied. While both males and females use the first and second exon, the third and fourth exons are expressed only in females and the fifth exon is expressed only in males.

The gene is highly conserved in the regions that have an important functional role, i.e. the domains of the protein, owing to the higher number of synonymous substitutions compared to non-synonymous substitutions in these regions. The male-specific exon on the other hand, exhibits more non-synonymous substitutions. Whether this patterns has something to do with sexual selection in males remains to be seen.

The tree constructed from the *doublesex* sequences shows a different pattern compared to the to species trees created using phyloT. This gene may be undergoing selection in particular species (e.g. regulation of mimetic polymorphism in *Papilio polytes*), however, we cannot comment on the evolutionary history of this gene across different species groups without information from more sequences.

REFERENCES

1. Verhulst, E. C., & van de Zande, L. (2015). Double nexus—*Doublesex* is the connecting element in sex determination. *Briefings in Functional Genomics*, 14(6), 396–406. http://doi.org/10.1093/bfgp/elv005
2. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." *Journal of Molecular Biology*. 215:403-410.
3. James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. (2011), Integrative Genomics Viewer. *Nature Biotechnology* 29, 24–26
4. Kumar S, Stecher G, and Tamura K (2016) *Molecular Biology and Evolution* 33:1870-1874