Evolution of transposable element composition and piRNA regulation across the Drosophila phylogeny

Jullien M. Flynn* & Andrew G. Clark
Cornell University, Molecular Biology and Genetics
*jmf422@cornell.edu

Background

Transposable elements (TEs) parasitize organisms' genomes. Animal hosts use a small RNA pathway, called piRNAs (23-29 nt) to silence TEs via transcriptional and post-transcriptional mechanisms.

piRNA cluster loci produce piRNAs

TEs and piRNA components evolve rapidly among species, potentially reflective of an arms race.

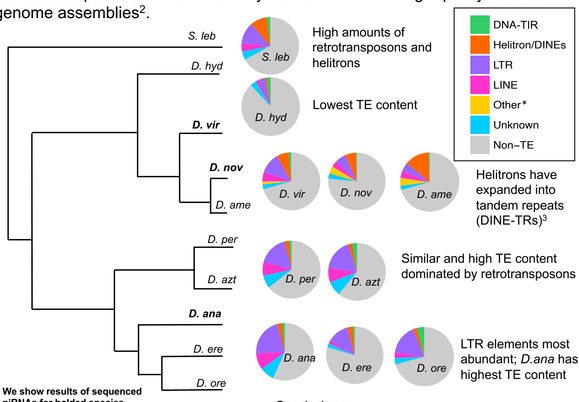
Across Drosophila species, TE compositions vary but also have similarities, allowing us to learn about how TEs and piRNAs coevolve.

In the Drosophila ovary, piRNAs are produced and passed to the next generation. In the testes, piRNAs are also produced but they may have divergent roles.

Question: How do TEs and piRNAs coevolve, and how does regulation differ between the ovaries and testes?

1. Evolution of TE composition across Drosophila

We used RepeatModeler2[1] to identify TEs de novo from high quality PacBio genome assemblies[2].



High amounts of retrotransposons and helitrons

Lowest TE content

Helitrons have expanded into tandem repeats (DINE-TRs)[3]

Similar and high TE content dominated by retrotransposons

LTR elements most abundant; *D.ana* has highest TE content

S.leb-Scaptodrosophila lebanonensis, D.hyd-hydei, D.vir-virilis, D.nov-novamexicana, D.ame-americana, D.per -persimilis, D.azt-azteca, D.ana-ananassae, D.ere-erecta, D.ore-orena

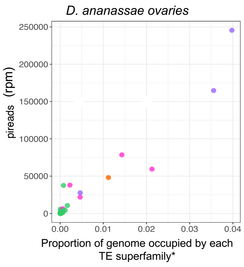
*Other- manually-curated interspersed tandem repeats such as pvB370[4] and 172-TR[5].

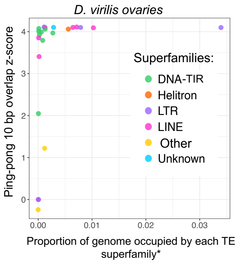
Conclusions:

- All species have high amounts of retrotransposons
- LTR elements are the most abundant subclass in most species
- Helitron abundance and evolutionary history varies among species
- Total abundance varies from 12-44% of genome

2. piRNA abundance and ping-pong cycle in relation to TE abundance

Presence of ping-pong indicates post-transcriptional regulation whereby piRNAs are amplified through cycles of hybridization and cleavage.





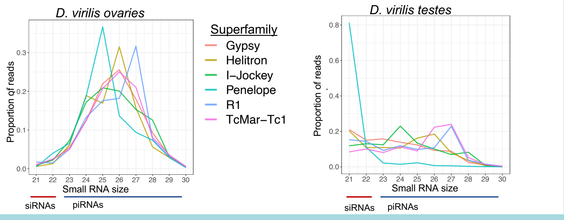
The abundance of piRNAs is generally correlated with TE genomic abundance. The ping-pong cycle is running for all subclasses of TEs in *D. virilis.*

**"Young" TE abundance included, <5% divergence from consensus

3. Ping-pong is running in the testes, but less than in ovaries

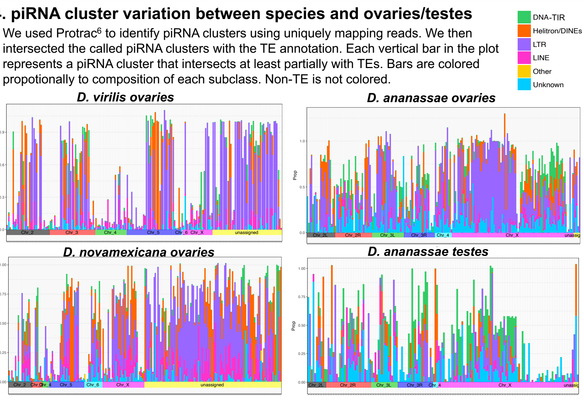
- Fewer overall piRNA reads in the testes
- Fewer families and superfamilies have a strong ping-pong signature

Penelope in *D. virilis* is regulated with piRNAs through ping-pong in the ovaries, but through siRNAs in the testes.



4. piRNA cluster variation between species and ovaries/testes

We used Protrac[6] to identify piRNA clusters using uniquely mapping reads. We then intersected the called piRNA clusters with the TE annotation. Each vertical bar in the plot represents a piRNA cluster that intersects at least partially with TEs. Bars are colored propotionally to composition of each subclass. Non-TE is not colored.



D. vir and *D.nov* have similar piRNA cluster compositions on some chromosomes (e.g. Chr_5), but differ in others (e.g. Chr_6, Chr_X)

The piRNA cluster composition differs greatly between ovaries and testes in *D. ana.* In particular, testes are depleted in LTR elements and enriched in DNA elements. Fewer piRNA clusters and reads overlap TEs in the testes also.

5. Conclusions

- piRNAs regulate TEs that are present in various abundances in the genome
- Small RNA regulation differs between the testes and ovaries
- Between closely related specis, piRNA clusters are conserved in some cases, and rapidly turn over in others
- piRNA cluster composition differs between the ovaries and testes, which may reflect different transcriptional activity of specific TEs between the gonads

References: 1-Flynn,Hubley et. al PNAS 2020, 2-https://www.ncbi.nlm.nih.gov/bioproject/?term=txid7214, 3-Dias et al. 2015 Chromosome Res, 4-Heikkinen et al. 1995 J. Mol, 5-Abdurashitov et al. 2013 BMC Genomics, 6-Rosenkranz & Zischler 2012 BMC Bioinf. **Acknowledgements** (PacBio genomes): NIH R01 GM116113 to R. Wing, M. Long and A.G. Clark.