

@Aeyocca



[https://github.com/Aeyocca/19\\_Botany](https://github.com/Aeyocca/19_Botany)



# Species level conserved noncoding sequence (CNS) variation

---

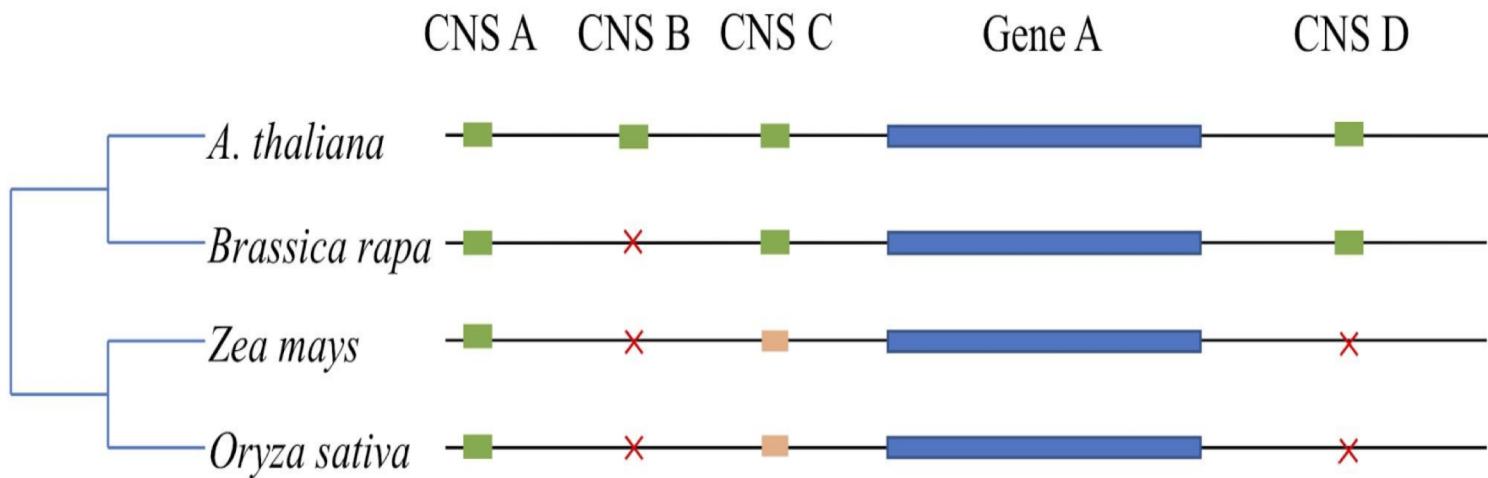
Alan E. Yocca  
Botany Conference July 30<sup>th</sup>, 2019



# Pangenome

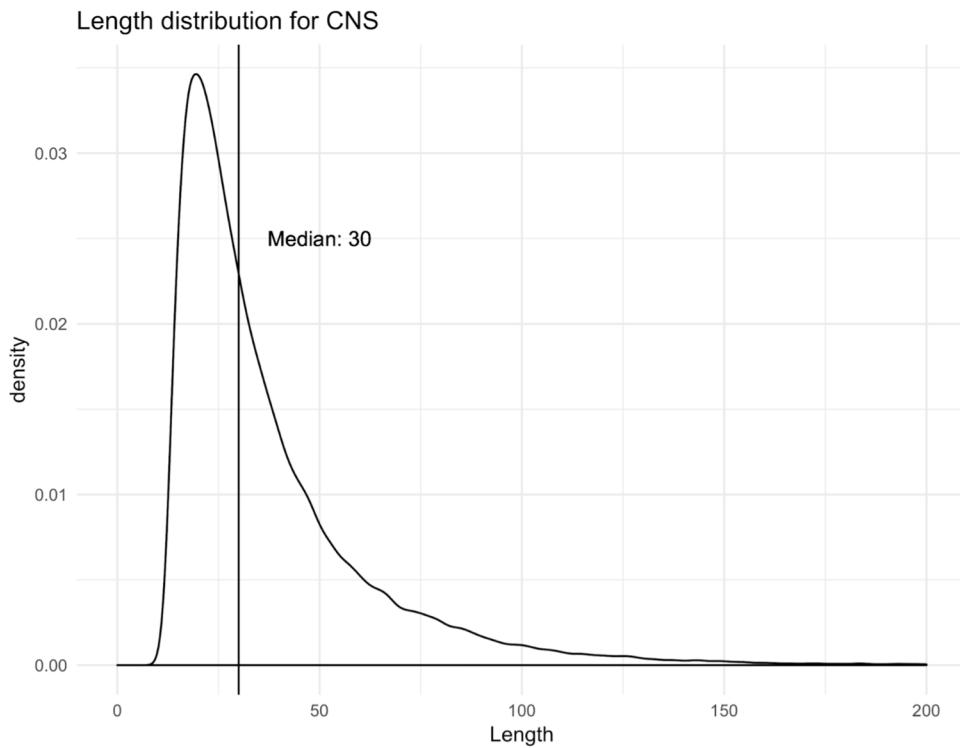


# Conserved Noncoding Sequences (CNS)

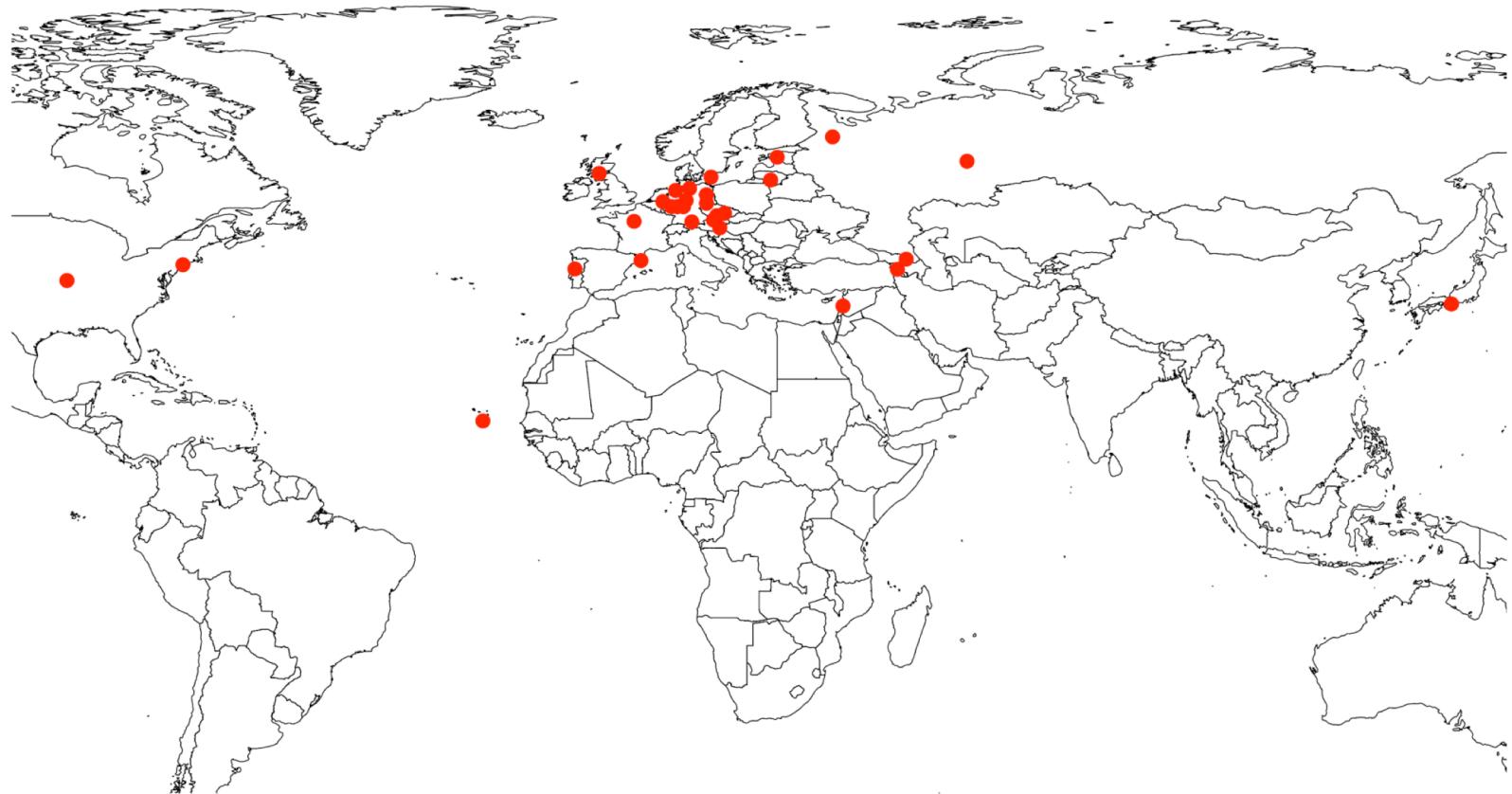


# Length distribution

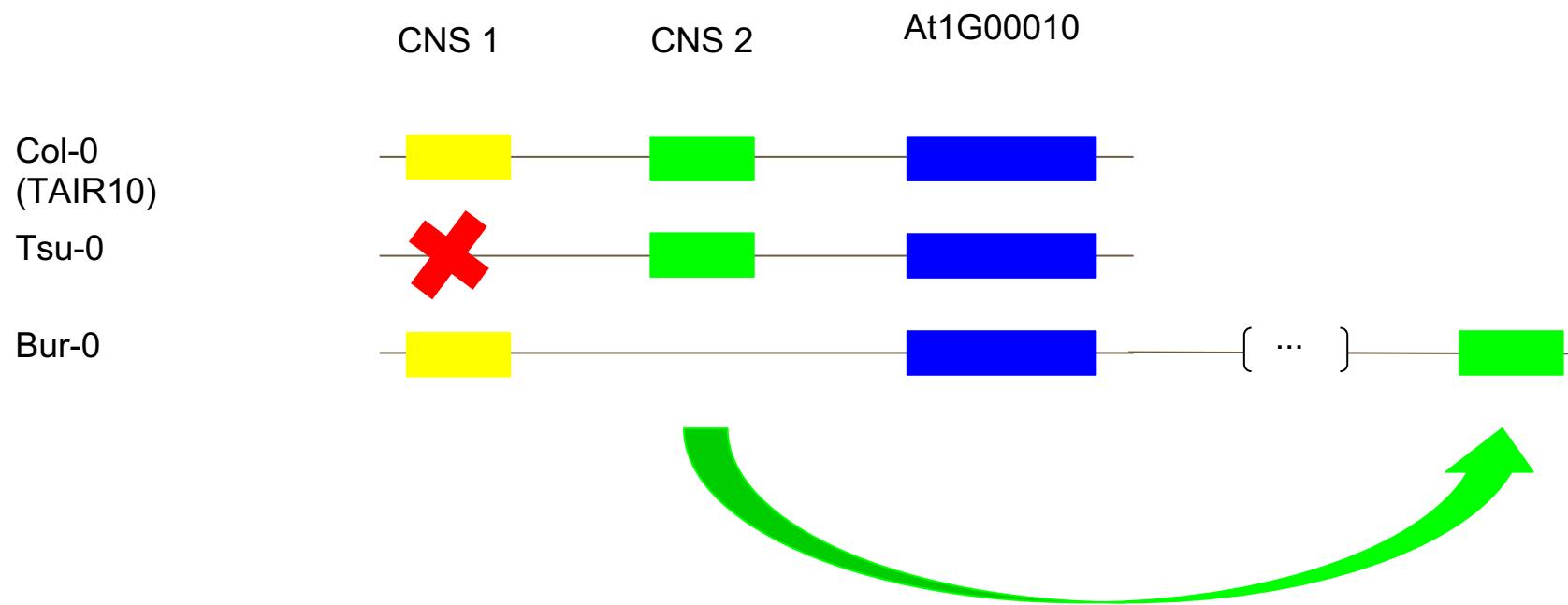
- Expected number of a given 6mer in *A. thaliana* genome:
  - >30,000
- Given 15mer:
  - 0.12



Why?  
What?  
How?



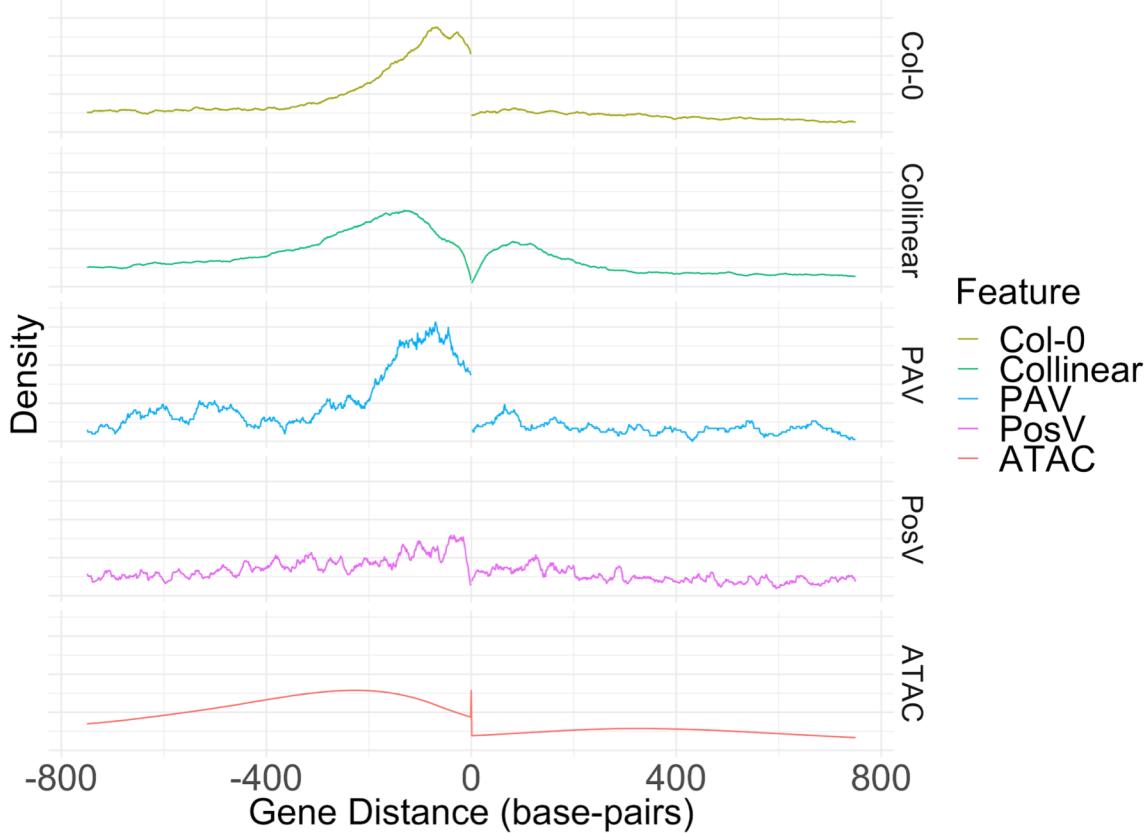
# Graphical representation of presence absence variation (PAV) & positional variation (PosV)



# Genomic Distribution

- 2.4% CNS display PAV in at least 1 accession
- 7.6% CNS display PosV in at least 1 accession

Feature density relative to the proximate gene



# CNS variability is associated with differential expression

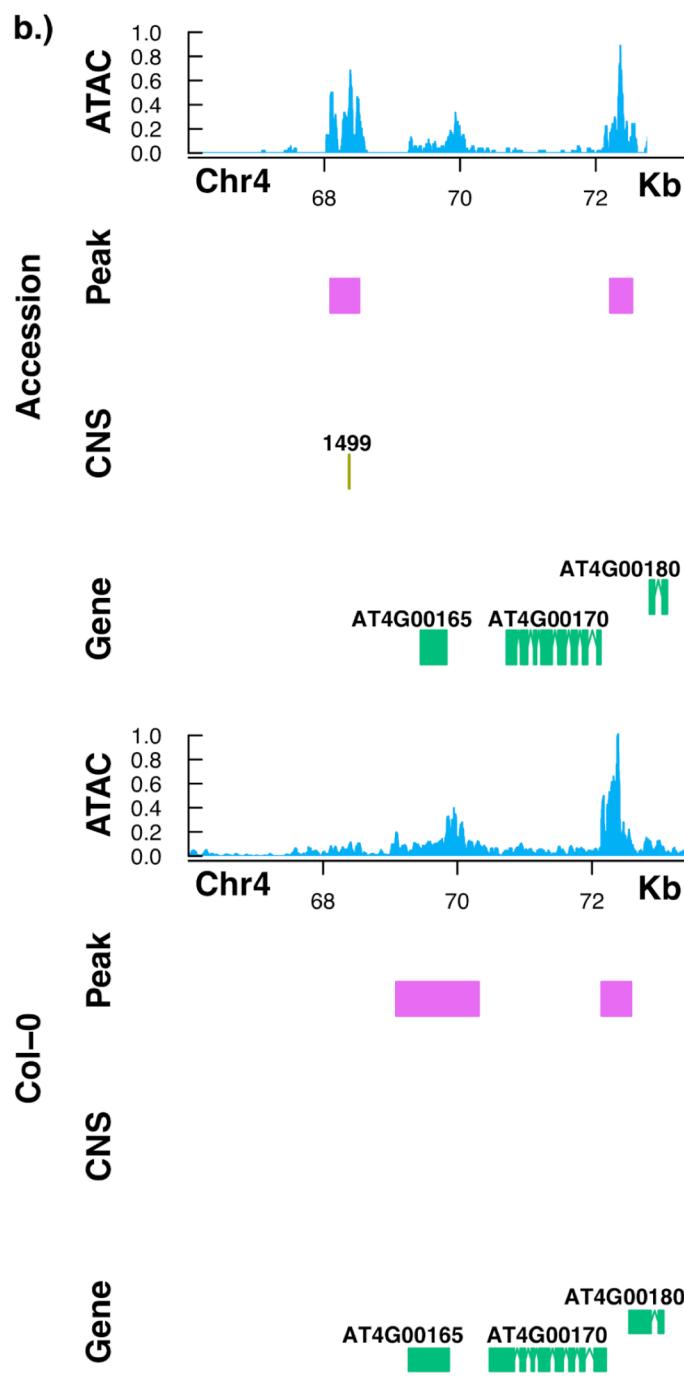
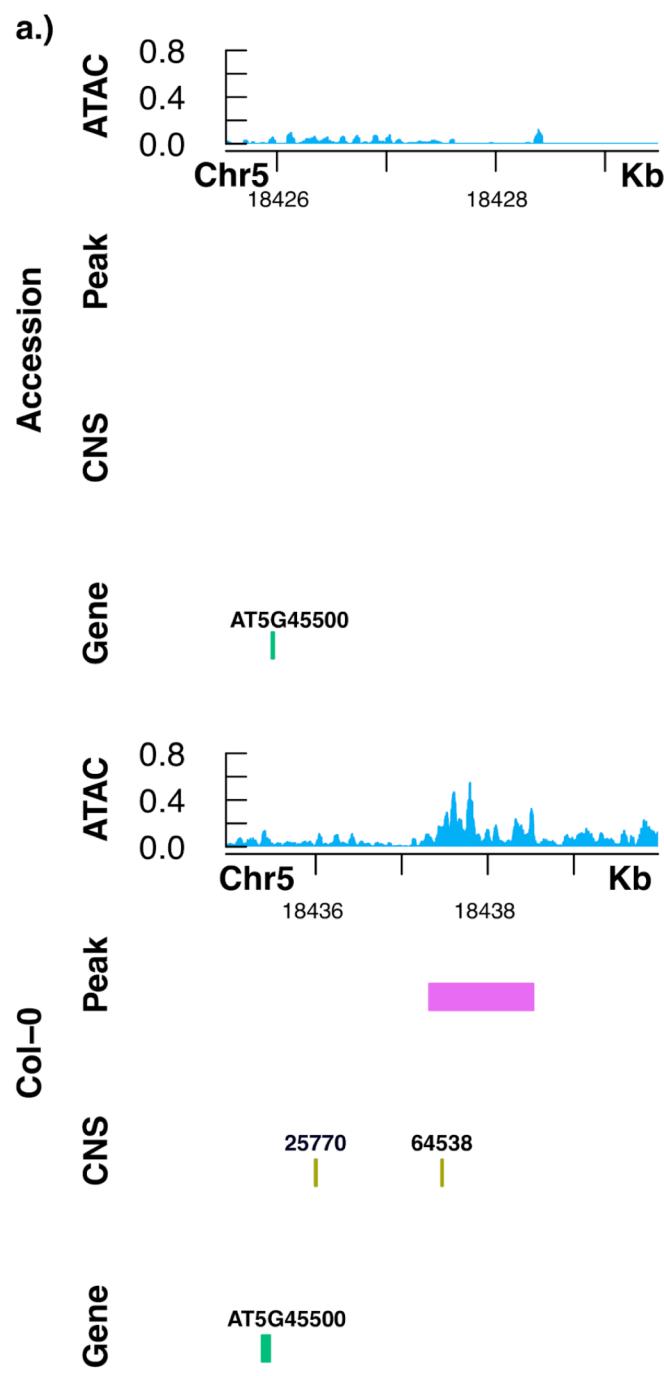
Gene Type	Upregulated Genes average per accession (%)	P-value (Hypergeometric test)	Downregulated Genes average per accession(%)	P-value (Hypergeometric test)
CNS Gain	208.5 (7.927)	< 1 x 10 <sup>-10</sup>	76.5 (2.908)	8.395x10 <sup>-06</sup>
CNS Loss	26.25 (2.449)	0.093*	121.75 (11.357)	< 1 x 10 <sup>-10</sup>
No CNS Change	489.75 (2.212)	< 1 x 10 <sup>-10</sup>	748.5 (3.380)	< 1 x 10 <sup>-10</sup>

\*Do not pass Bonferroni correction p-value 0.01 / 6 =

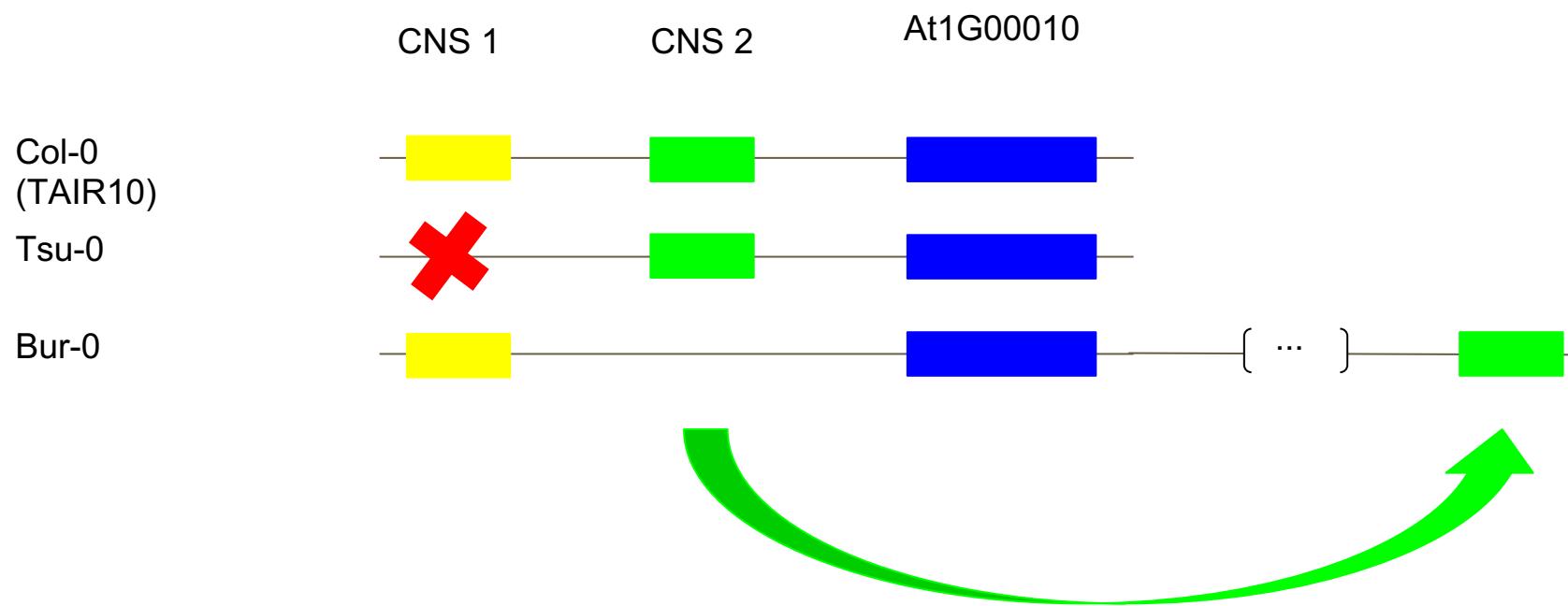
0.00167

Underrepresented

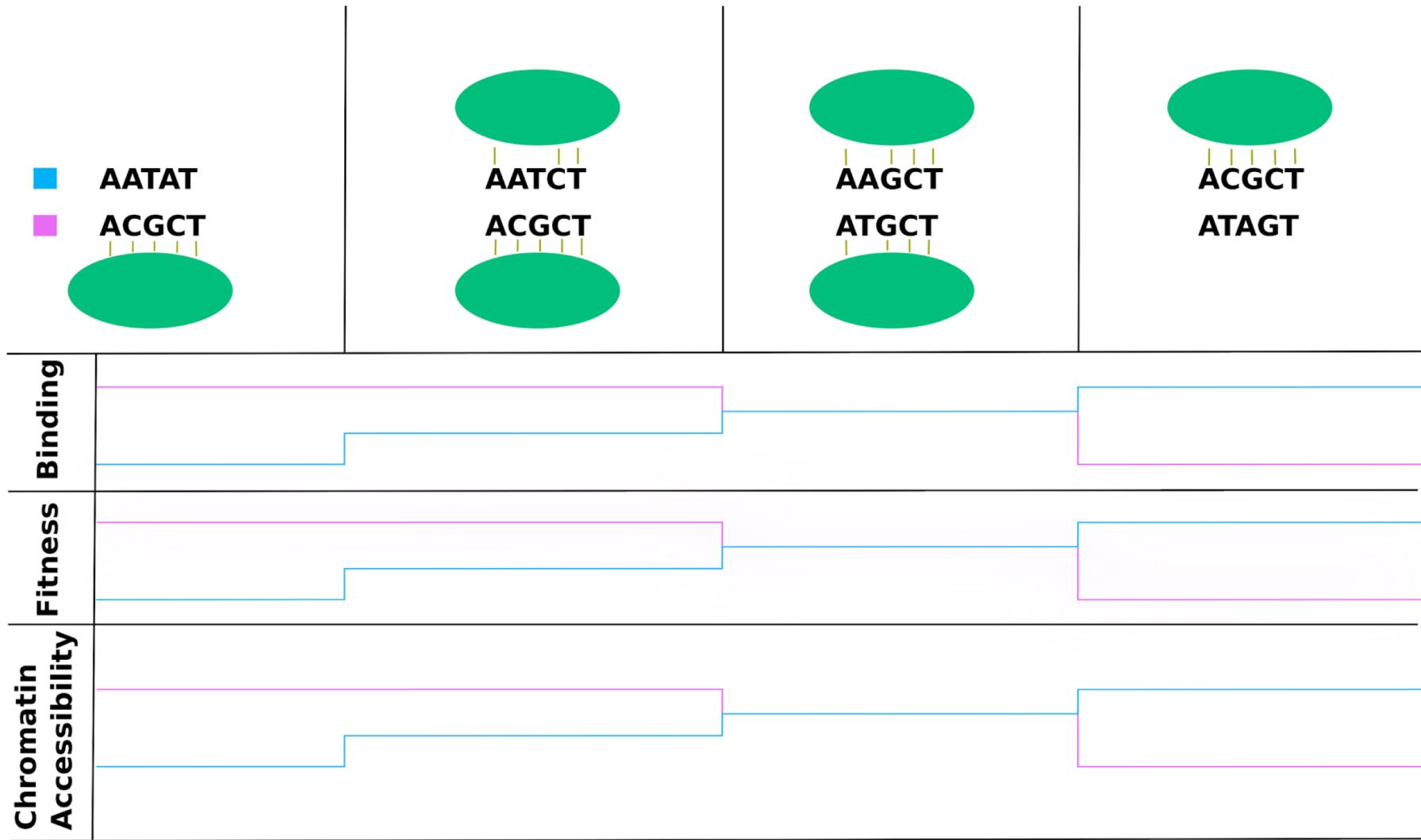
Overrepresented



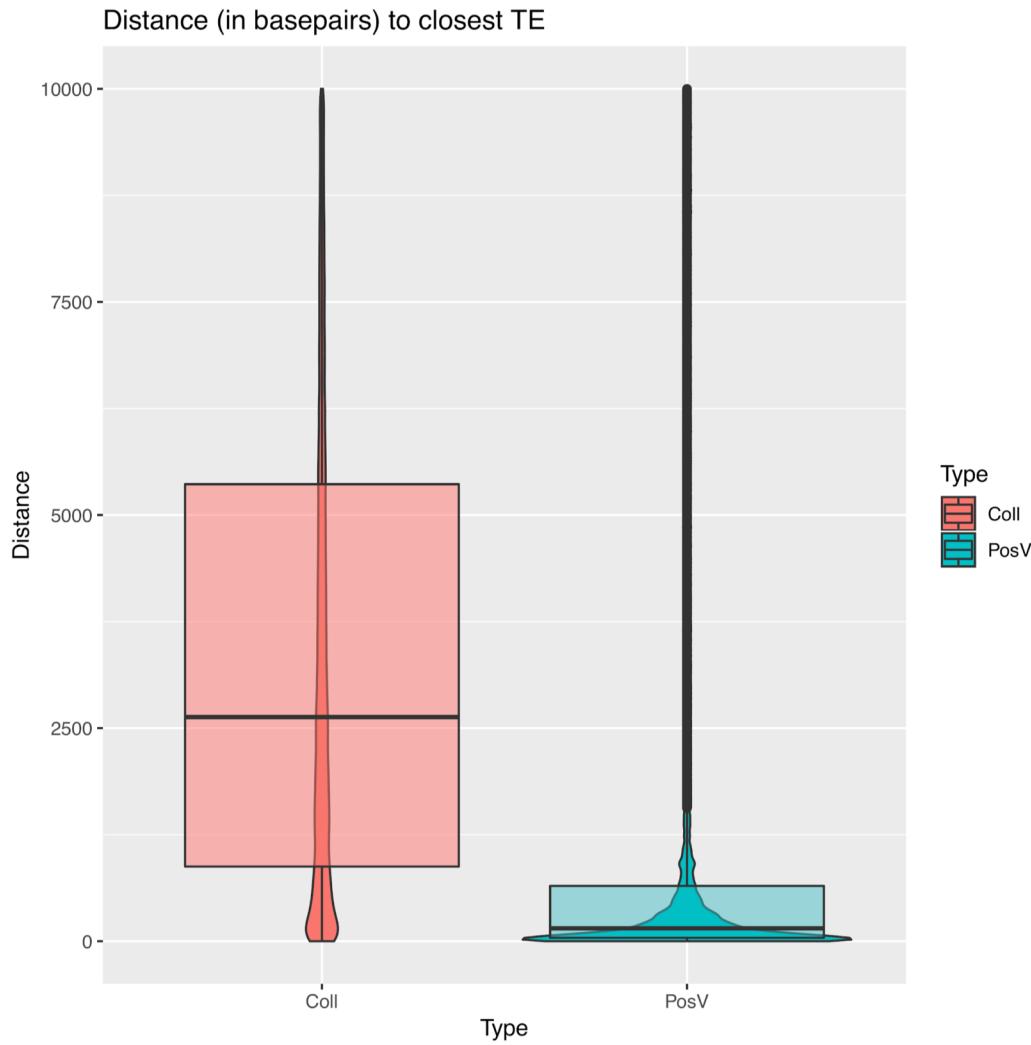
# Graphical representation of presence absence variation (PAV) & positional variation (PosV)



# Transposition mechanism



# Transposable Element involvement??



@Aeyocca



[https://github.com/Aeyocca/19\\_Botany](https://github.com/Aeyocca/19_Botany)



## Funding

- Michigan State University Plant Resilience Institute
  - <https://plantresilience.msu.edu/>
- Michigan State University NRT-IMPACTS
  - <https://impacts.natsci.msu.edu/>
  - NSF DGE-1828149
- BSA Genetics Section Student Travel Award



**Botanical Society  
of America**

@Aeyocca



[https://github.com/Aeyocca/19\\_Botany](https://github.com/Aeyocca/19_Botany)



## Acknowledgements

- **Patrick Edger**
- Kevin Bird
- Elizabeth Alger
- Yao Cao
- Scott Teresi
- Marivi Colle
- Jeremy Pardo
- **Bob Schmitz**
- **Zefu Lu**
- **Mike Freeling**

Committee:

- Robert VanBuren
- Jiming Jiang
- Emily Josephs

Michigan State  
University Department  
of Plant Biology

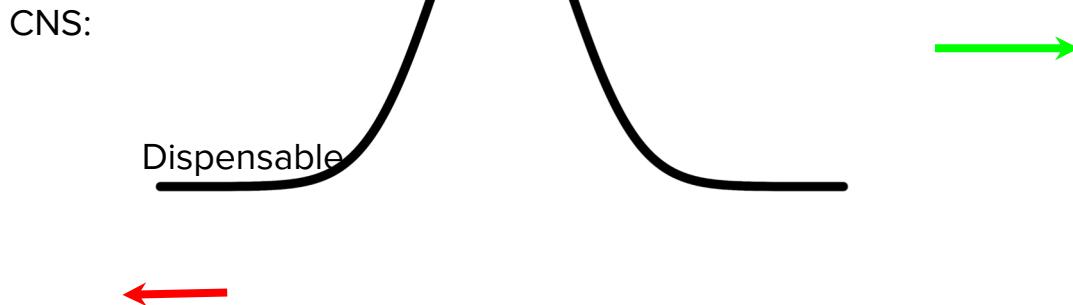
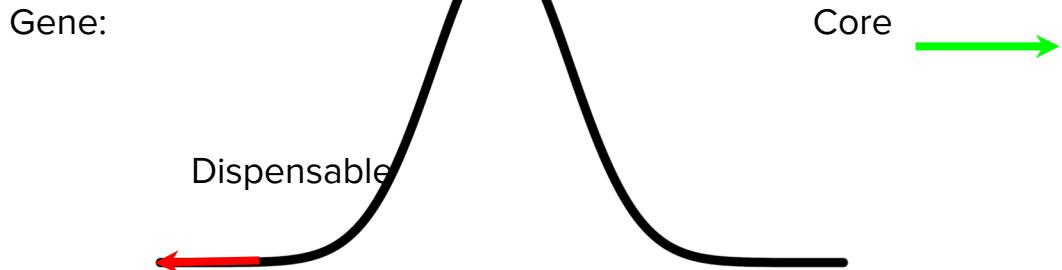


# CNS currently annotated for:

- Dicots/Eudicots
  - *Arabidopsis thaliana, Carica papaya, Glycine max, Malus domestica, Populus trichocarpa, Fragaria vesca, Medicago truncatula, Lotus japonicus, Theobroma cacao, Ricinus communis, Manihot esculenta, Vitis vinifera, Aquilegia, Prunus persica, Amborella trichocarpa*
- Monocots
  - *Zea mays, Triticum aestivum, Sorghum bicolor, Brachypodium distachyon, Musa acuminata, Hordeum vulgare*
- Brassicaceae
  - *A. thaliana, A. lyrata, Capsella rubella, Leavenworthia alabamica, Sisymbrium irio, Aethionema arabicum, Eutrema salsugineum, Schrenkiella parvula, Brassica rapa*

# Hypothesis for CNS function

- Known enrichment for Transcription Factor Binding Sites
- Regulatory fine-tuning
- Navigation of adaptive landscape
- Pools of adaptive sequences



# Hypothesis for CNS function

- Known enrichment of transcription factor binding sites (TFBS)
- Possible mechanism for regulatory fine tuning
- Sequence upon which selection can act freely
  - ie non-crucial of gene function, but may affect gene selective pressure
- Easily (comp. to genes) modulated sequence to play adaptive role over short evolutionary distances