



Estimation of conservation using divergence between maize and sorghum with the Andropogoneae tribe

Merritt B. Burch (mbb262@cornell.edu)¹, Baoxing Song², Cinta M. Romay², Elizabeth Kellogg³, Edward S. Buckler^{1,2,4}

1 Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853, 2 Institute for Genomic Diversity, Cornell University, Ithaca, NY USA 14853, 3 Donald Danforth Plant Science Center, St. Louis, MO USA 63132, 4 USDA-ARS; Ithaca, NY, USA 14853



Why use the Andropogoneae to find conservation?

- Finding evolutionary conserved regions is limited by experimental manipulation or existing variation between species
- The Andropogoneae tribe:
 - 1,200 species
 - NADP-ME C₄ photosynthesis
 - Common ancestor 16-20 Mya
- Sequencing ~900 Andropogoneae = >1 billion years of independent evolution

Goals

- Compare across Andropogoneae, Poaceae, and Angiosperm genomes:
 - How deleterious is any single mutation?
 - Do different evolutionary time depths provide equivalent estimates of deleterious or fitness mutation effects?

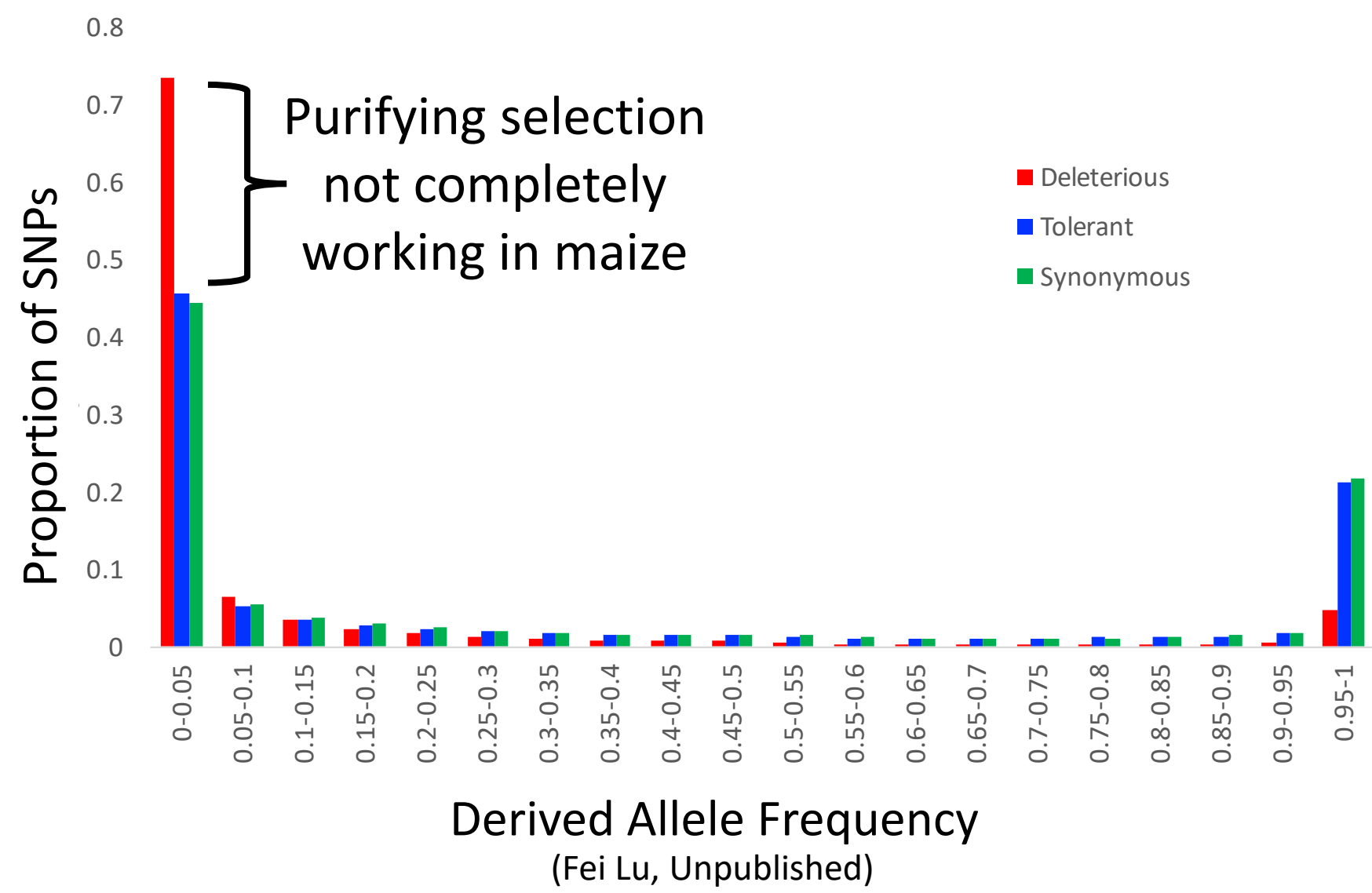
Calculating Genomic Rate Profiling (GERP) scores

- Uses sequence alignment to quantify constrained regions

GERP in Humans

Multiple Sequence Alignment

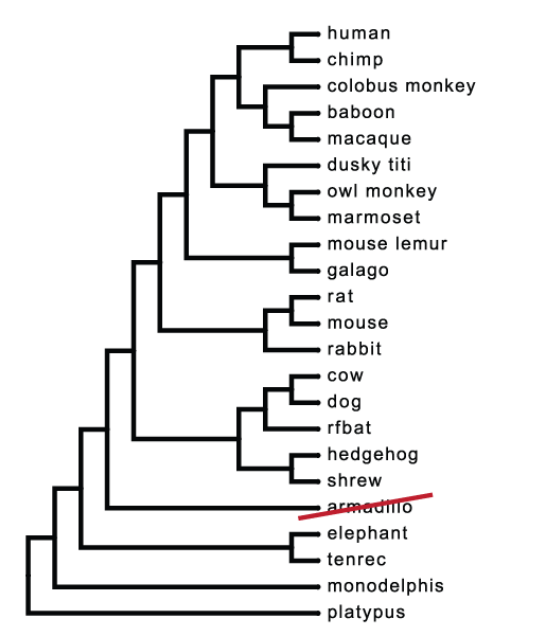
human	AATACGG	A	ACITTCATT	CATT
chimp	AATACGG	A	ACITTCATT	CATT
lobus monkey	AGTATGG	A	ACITTCATT	CATT
baboon	AGTATGG	A	ACITTCATT	CATT
macaque	AGTATGG	A	ACITTCATT	CATT
duky titi	AGTATGG	A	ACITTCATT	CATT
owl monkey	AGTATGG	A	ACITTCATT	CATT
marmoset	AGTATGG	A	ACITTCATT	CATT
mouse lemur	AGTACGG	A	ACITTCATT	CATT
galago	AGTACGG	A	ACITTCATT	CATT
rat	AGTATGG	A	ACATCATT	CATT
mouse	AGTATGG	A	ACATCATT	CATT
rabbit	AGTATGG	A	ACATCATT	CATT
cow	AGTATGG	A	ACATCATT	CATT
dog	AGTATGG	A	ACATCATT	CATT
rhesus	AGTATGG	A	ACATCATT	CATT
hedgehog	AGTATGG	A	ACATCATT	CATT
shrew	AGTATGG	A	ACATCATT	CATT
armadillo	AGTATGG	A	ACATCATT	CATT
elephant	AGTATGG	A	ACATCATT	CATT
tenrec	AGTATGG	A	ACATCATT	CATT
monodelphis	AGTATGG	A	ACATCATT	CATT
platypus	AGTATGG	A	ACATCATT	CATT



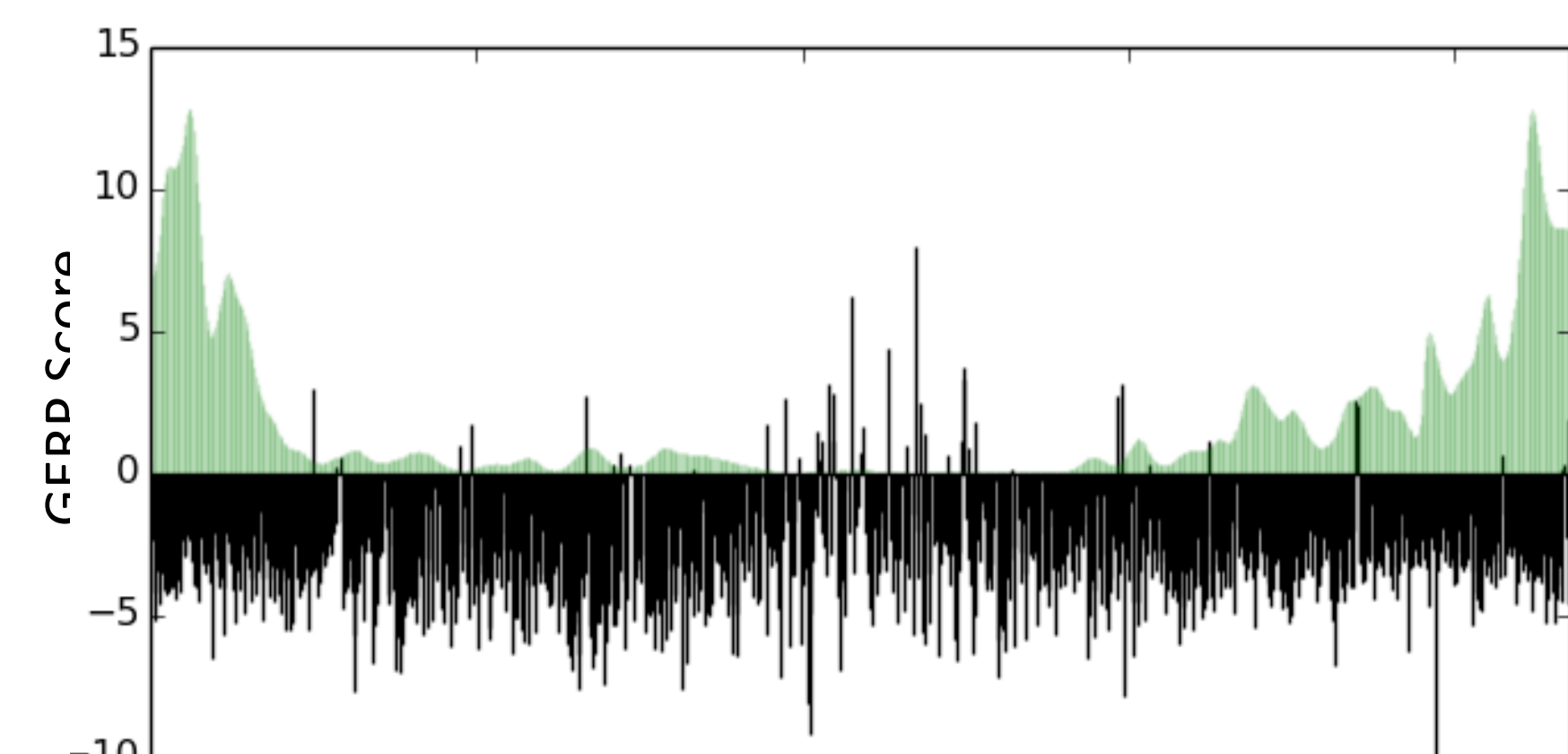
Derived Allele Frequency (Fei Lu, Unpublished)

Proportion of deleterious mutations is high under low allele frequencies

ML Tree Scaling Factor $r = 0.7$



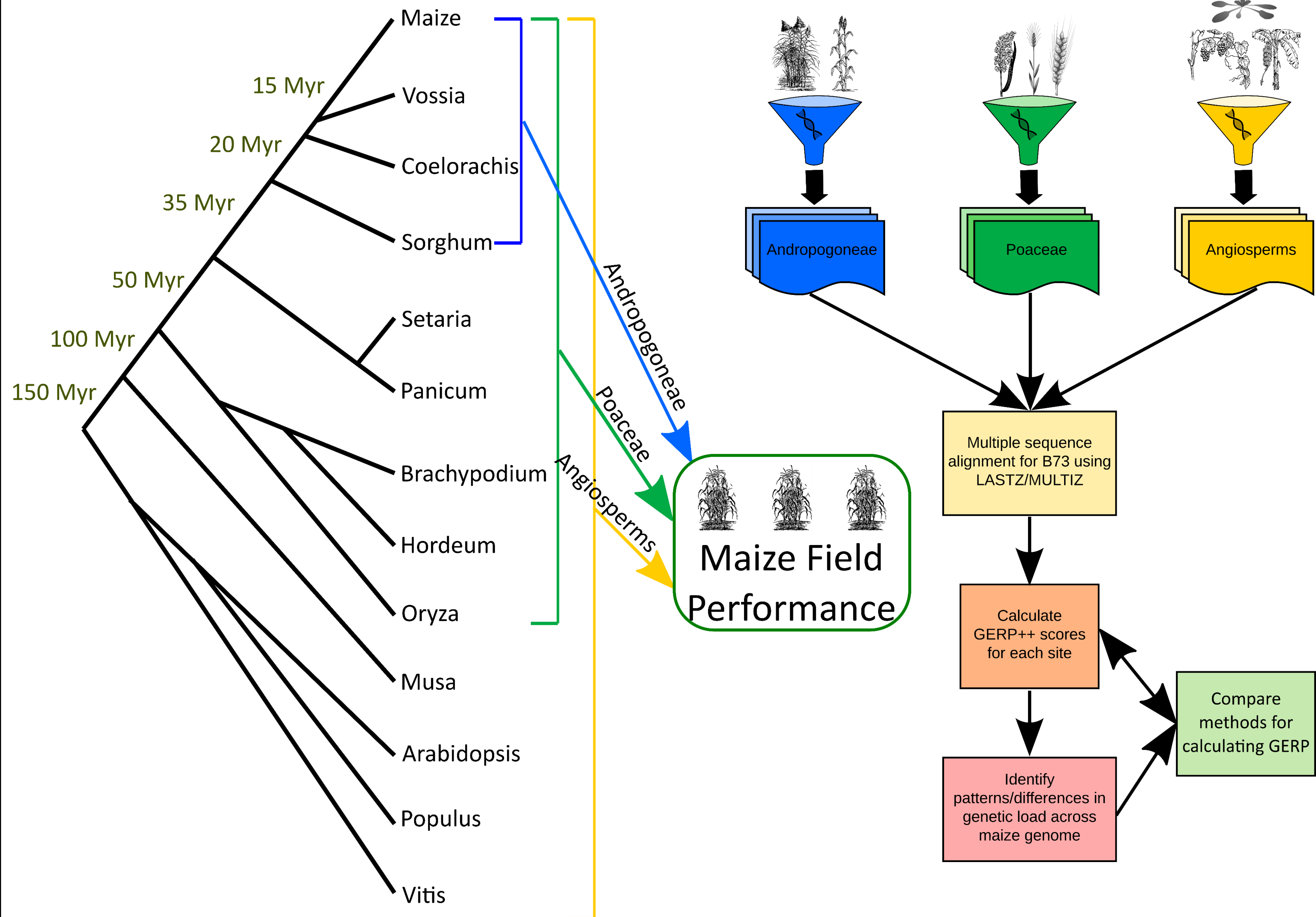
RS Score = 1.14 (Neutral - Estimated) (Davydov et al., 2010)



Chromosome 5 Position (Rodgers-Melnick et al., 2015)

Deleterious mutations are everywhere! Enriched in low recombination regions.

Different evolutionary time depths predict mutation effects



Next Steps

- Compare and contrast the power and sensitivity of GERP constraint by clade in:
 - Functional elements of genes
 - Protein families & domains
 - Intergenic regulatory regions

References

Davydov, E. V., et al., (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP+. *PLoS Computational Biology*, 6(12), e1001025.

Rodgers-Melnick, E., Bradbury, P. J., et al., (2015). Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences of the United States of America*, 112(12), 3823–3828.

Future Directions

- Use GERP estimates to design heterosis and find loci for gene editing
- Create machine learning models that predict transposable element families and hijacking into functional elements
- Examine patterns within ~1 billion years of independent transposable element activity