GENOME EVOLUTION AFTER WHOLE GENOME DUPLICATION IN 32 BRASSICALES SPECIES

Tyler V. Kent
Stephen I. Wright
BMAP Consortium*

BACKGROUND & MOTIVATION

10 to 40% of all Angiosperms are polyploid [1] and most have had a recent whole genome duplication (WGD) event. Despite this the genomic consequences of WGD events are poorly understood and there are few comparative genomics resources available for closely related plants. The Brassicales Map Alignment Project (BMAP) has **sequenced 18 new species in the Brassiaceae and 1 new species in the Cleomaceae**, effectively doubling the reference genomes for the family. These represent the major lineages of the Brassicaceae, and cover multiple shared and independent WGD events. We are using this expanded comparative dataset to ask how genomes evolve following WGD events, and specifically how **relaxed selection following WGD can lead to increases in mutation load and transposable elements (TEs)**.

PHYLOGENETIC RECONSTRUCITON

- Species sequenced with Illumina and multi-aligned with progressiveCactus [2]
- Species tree constructed using 13,000 gene alignments in Astral & neutral branch lengths calculated in RAxML-ng [3] using 200,000 four-fold degenerate sites

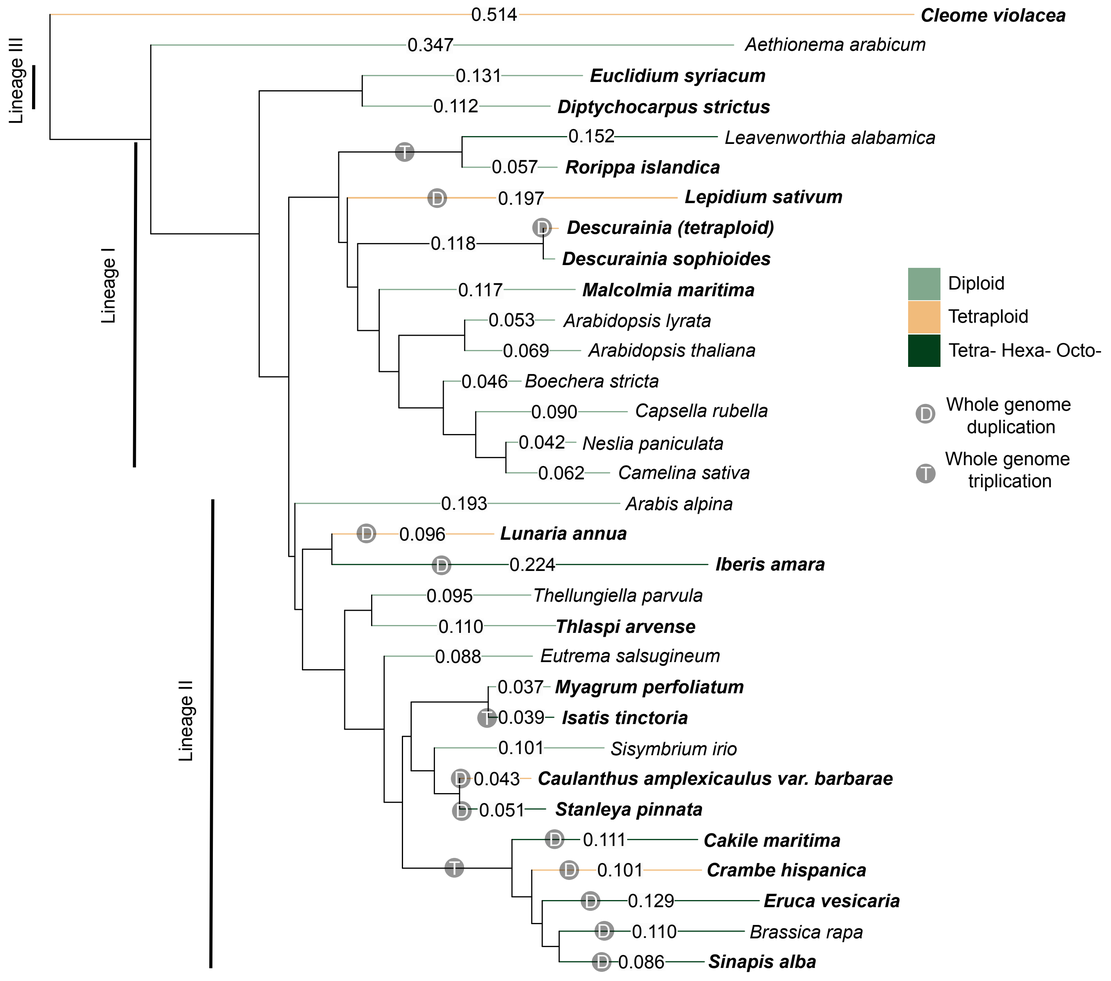

Figure 1: Species tree indicating ploidy level, WGD/WGT events, and neutral branch lengths. Bold indicates a newly-sequenced BMAP species.

POSITIVE GERP SCORES PREDICT INCREASING DELETERIOUS FITNESS EFFECTS

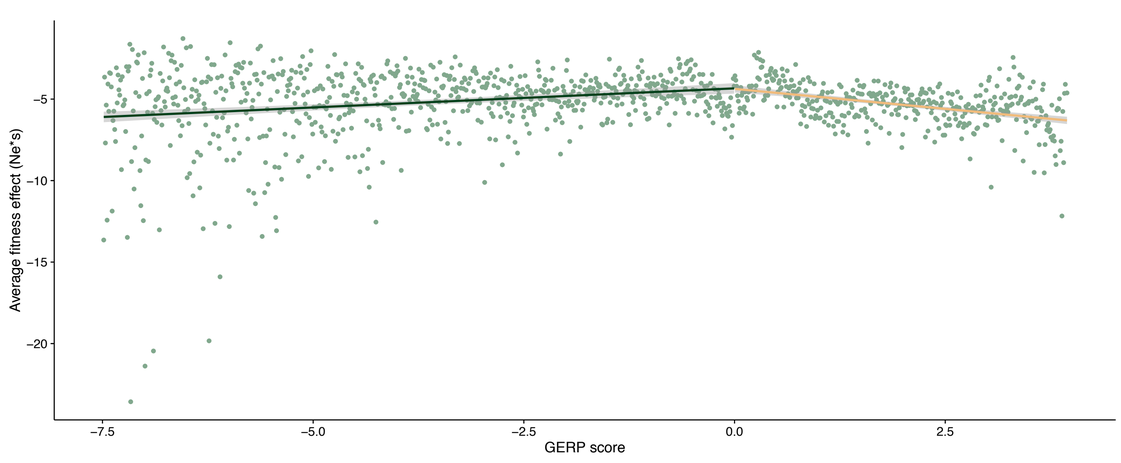

Figure 2: Mean Ne*s inferred from DFE-alpha for GERP bins of 0.01.

- We can connect Genomic Evolutionary Rate Profiling (GERP) scores [4] to fitness by binning scores, calculating site frequency spectra on each bin using 182 whole genome sequences of *Capsella grandiflora*, then estimating the distribution of fitness effects of each GERP score bin using DFE-alpha [5]
- Positive GERP scores show a significantly negative correlation with mean Ne*s, indicating a higher fitness effect for more highly conserved sites
- Mean Ne*s effect is small due to the gamma DFE with most of its mass at nearly neutral effects

MUTATIONS AT HIGHLY CONSERVED SITES HAVE LARGE DELETERIOUS FITNESS EFFECTS

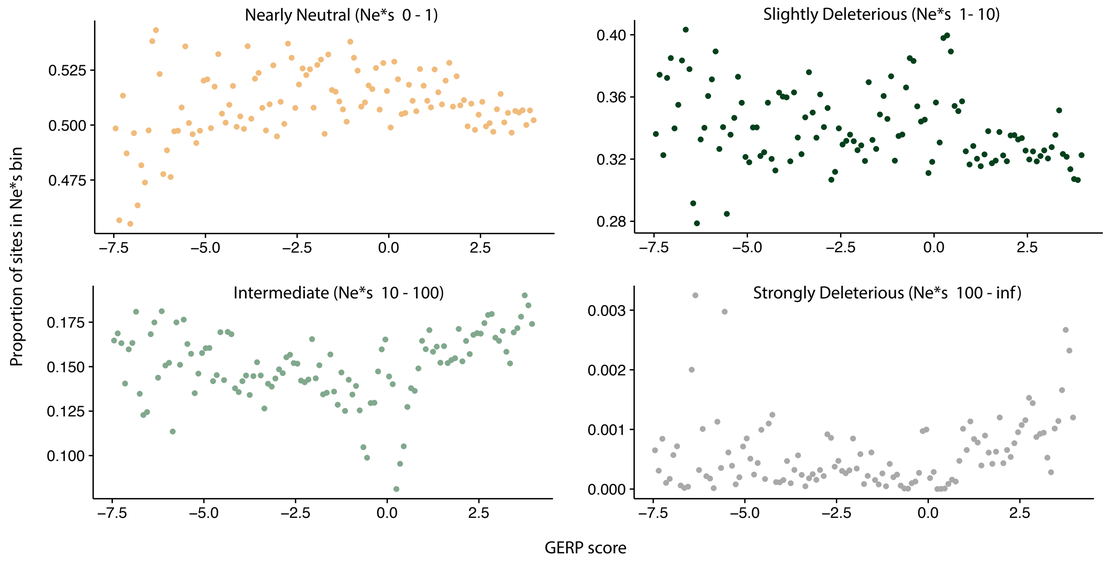

Figure 3: Proportion of sites within a GERP score bin which have DFE within an Ne*s bin

- The negative effect of mean Ne*s is driven by sites with large fitness effects (Ne*s > 10)
- Proportionally fewer sites with large fitness effects cause the small shift in mean Ne*s

PLOIDY BUT NOT TIME SINCE LAST WGD PREDICTS RELATIVE LOAD

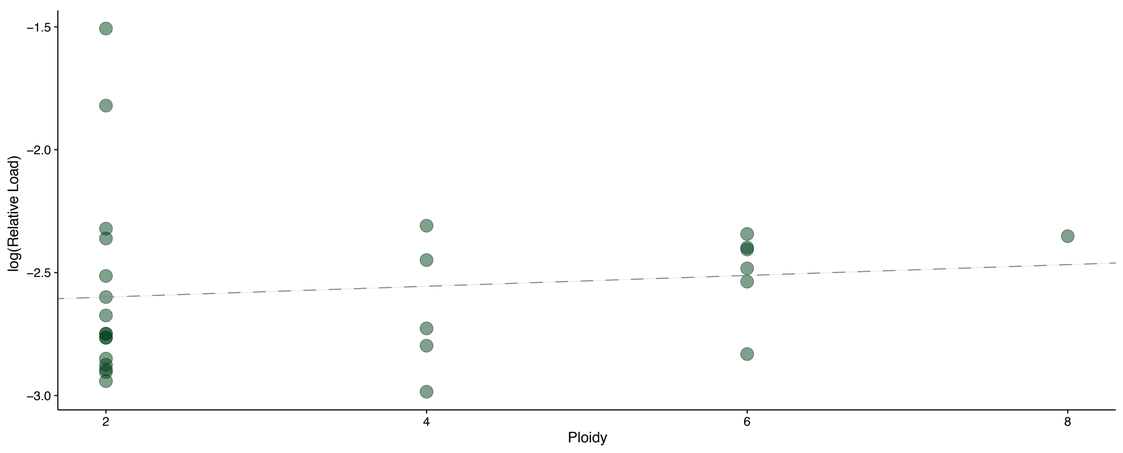

Figure 4: Relative load for each species as measured by the proportion of sites with GERP >0.45 and <3.65 which have mutated away from the inferred conserved nucleotide vs current ploidy level. The few sites with intermediate DFE drive this correlation (Kim Gilbert, unpublished [6])

- After phylogenetic correction, current ploidy level is a marginally significant predictor of relative mutation load (p=0.07, $R^2 = 0.08$)
- Time since last WGD (in units of log(Ks)) does not correlate with relative load, likely indicating a complex evolutionary process following WGD and diploidization

INTERACTIONS BETWEEN TRANSPOSABLE ELEMENT CONTENT, GENOME SIZE, AND TIME SINCE LAST WGD

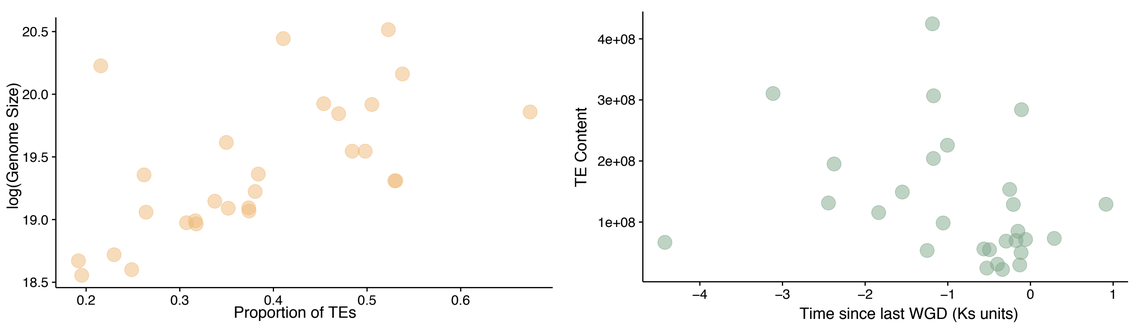

Figure 5: Relative load for each species as measured by the proportion of sites with GERP >0.45 and <3.65 which have mutated away from the inferred conserved nucleotide vs current ploidy level. The few sites with intermediate DFE drive this correlation (Kim Gilbert, unpublished [7])

- 68% of the variance in genome size can be explained by the proportion of TEs in the genome and the time since the last WGD (units of log(Ks))
- Recent WGD drives an increase in total TE content ($R^2 = 0.47$)
- Increases in TE activity following WGD and more rapid diploidization than loss of TEs may explain genome size differences

REFERENCES

[1] Otto and Whitton (2000) Annu Rev Genet; [2] Patten et al. (2011) Genome Res; [3] Stamatakis (2014) Bioinformatics; [4] Davydov et al. (2010) PLoS Comp Bio; [5] Keightley & Eyre-Walker (2007) Genetics; [6] Kim Gilbert: SMBE17 talk WEDNESDAY 2:15 Mutational Load

*BMAP Consortium: Adrian E. Platts, Patrick Edger, Florian Maumus, J. Chris Pires, Rod Wing, M. Eric Schranz, Thomas Mitchell-Olds, Detlef Weigel, Jeremy Schmutz, Jacob Washburn, Shengqiang Shu

UNIVERSITY OF TORONTO
ECOLOGY AND EVOLUTIONARY BIOLOGY

UNIVERSITY OF TORONTO

TYLER.KENT@MAIL.UTORONTO.CA
TVKENT.GITHUB.IO
@TYLERVKENT