

Selection or Drift on Repetitive Elements Causes Large Genome Size Variation in *Zea* and *Tripsacum*

Paul Bilinski¹, Anne Lorant¹ Author¹, Jeffrey Ross-Ibarra^{1,2,*}

1 Dept. of Plant Sciences, University of California, Davis, CA, USA

2 Author2 Dept/Program/Center, Institution Name, City, State, Country

3 Author3 Dept/Program/Center, Institution Name, City, State, Country

* E-mail: rossibarra@ucdavis.edu

Abstract

Structural variation is rampant in the maize genome and is known to contribute greatly to large fluctuations in genome size, though little is known as to how selection acts on broad structural variation. Using next-gen sequence data, we investigate genome-wide changes in repetitive content in maize landraces, teosinte, and *Zea*'s sister genus *Tripsacum* across an altitudinal gradient in Mexico and South America to identify the repetitive content responsible for shifting genome sizes. We also use models akin to $F_{st} - Q_{st}$ analysis to search for evidence of selection across populations, treating repetitive content as our phenotype. We corrected for repetitive content abundance using genome size measures, and found incredible variation in individual repetitive tags across our altitudinal gradient. Surprisingly, different classes of repeats showed opposing clinal patterns after correcting for genetic relatedness.

Author Summary

Introduction

Last: table of taxa we are using

Results

Genome Size Variation and Repeats

Plot of genome size var in all taxa

Subsection 2

Discussion

Materials and Methods

Plant Material

Sequencing

snp chip data Skim Sequencing GBS of trip

Genome Size Estimates

Absolute genome size measures were performed by Plant Cytometry Services (JG Schijndel, NL). All maize and teosinte samples were grown in greenhouse conditions, and leaf samples were collected and

sent for measure at the same time. For populations of teosinte, two individuals per population were measured as previous studies have shown the greatest genome size variation across populations [1], while genome size measures were performed individually on each maize landrace accession. *T. dactyloides* experienced low germination rates, and therefore we were only able to perform genome size measure on XX individuals. These individuals spanned the altitudinal range of our *T. dactyloides* samples. We sequenced to low coverage our XX germinated *T. dactyloides* individuals and observed a strong correlation between genome size measures and relative genomic content mapping to maize cDNA (ref). To expand the number of *T. dactyloides* in our study, we extracted DNA from single embryos and performed low coverage sequencing and genotyping by sequencing (GBS). We used total genomic content mapping to maize cDNA to estimate relative genome size of our 96 *T. dactyloides* embryos.

Repetitive Content Mapping

Mapping references and sequence ID's were constructed separately each repeat and are available on our github repository (<https://github.com/paulbilinski/>). Reference sequences for canonical knobs, TR-1 knobs, and rDNA repeats were gathered from NCBI, and sequence identifiers are available in the reference files on github. We used the library of CentC repeats assembled in cite. Chloroplast DNA and cDNA were taken from the maize reference genome (5b, www.maizesequence.org). For the transposable element database, we began with the TE database consensus sequences from [2]. We BLASTed sequences against themselves and masked shared regions, only retaining unique tags that were at least 70bp in length in our mapping reference.

Read mapping was performed using the Burrows-Wheeler Aligner [3]. For repetitive sequences, we used parameters -B 2 -k 11 -a to store all hit locations with an identity threshold of approximately 80%. We decided upon a minimum seed length of 11 as it produced the most reads mapping against the full transposable element database. To test whether alignment algorithm played a large role in our repetitive content measures, we compared measures from BWA to measures from Mosaik, used for previously repetitive content mapping in cite. We observed nearly perfect correlations between total 180bp knob content across the panel with the two aligners ref. All data for this study was generated with BWA due to its speed of alignment and broader use in the community.

Genotype-Phenotype Evolution

Black box of ovaskienen.

Acknowledgments

shout out to the homeys.

References

1. Díez CM, Gaut BS, Meca E, Scheinvar E, Montes-Hernandez S, et al. (2013) Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytologist* 199: 264–276.
2. Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, et al. (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the b73 maize genome. *PLoS Genetics* 5: e1000732.
3. Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25: 1754–1760.

Figure Legends

Tables