

Differing Patterns of Selection on Repetitive DNA in Maize and its Relatives

Paul Bilinski¹, Jeremy J Berg^{2,3}, Patrice Albert⁴, Jim Birchler⁴, and Jeffrey Ross-Ibarra^{1,3}

¹Department of Plant Science, ²Department of Evolution and Ecology, Center for Population Biology and Genome Center, UC Davis, ⁴ Division of Biological Sciences, University of Missouri

Introduction

Historically, studies of adaptation to an environment have focused on the identification of genes, neglecting the role that repetitive elements play in adaptation. However, a growing body of research suggests that non-genic regions make important contributions to overall phenotype, through mechanisms such as expression regulation or determination of cell cycle length.

Here, we take a whole-genome approach to understanding how genome size and its largest components change and whether those changes show evidence of selection in response to environmental clines using the model system maize and its relatives. The maize genome is comprised primarily of three repetitive classes; transposable elements (TE), knobs, and centromere repeats (CentC) (Schnable et al. 2009). Fluorescent in-situ hybridization has shown vast structural variation within the genus (Albert et al. 2010), while other work has identified clinal variation in overall genome size (Poggio et al. 1998).

Our study investigates clinal patterns of structural variation in maize (*Zea mays* ssp. *mays*, *parviflora*, and *mexicana*) and identifies whether those regions show evidence of selection. We create mapping libraries from known maize repeats and align whole genome short read sequence against them to uncover relative abundance, correcting for overall genome size with flow cytometry data. We find striking patterns of repeat variation and genome size across taxa, and show that selection acts differently across the different taxa.

Materials and Methods

Line Selection, DNA Extraction, and Genome Size

Maize lines were selected from Kanizay et al. 2013, teosinte lines from across an altitudinal gradient. We filtered potential accessions based on their phylogenetic classification, ploidy, and location. Genomic DNA was isolated from leaf or seedling material using the standard CTAB DNA extraction and prepared for Illumina sequencing. Previous studies have shown both high and low coverage sequencing can accurately estimate repeat content in the genome (Tenaillon et al. 2011). For a subset of individuals from each of the subpopulations or accessions, leaf tissue was collected and analyzed using flow cytometry for genome size.

Sequencing and Repeat Abundance

The genetic relatedness of all *mays* populations was known from previously published data (van Heerwaarden et al 2011, Hufford et al. 2013). Uncharacterized accessions were genotyped by sequencing (GBS; Elshire et al. 2011). We made a reference library for each of the 3 repeats that allowed us to assign membership to each of the subgroupings in TEs, knobs, and CentC. Reads were mapped with BWA-MEM (Li and Durbin 2009), and repeat abundance was calculated from the percentage of reads mapping to each repeat class corrected for genome size.

Tests for Selection

In order to test for selection, we built the following model that identifies whether phenotypic change exceeds expectations set by genetic drift.

$$\mathbf{Y} = \mathbf{MVN}(\boldsymbol{\beta}\mathbf{X}, \mathbf{V}_a \mathbf{F})$$

The model states that our vector of phenotypes (\mathbf{Y}) is sampled from a multivariate normal distribution that incorporates information from a selection coefficient ($\boldsymbol{\beta}$), environmental data (\mathbf{X}), an estimate of additive genetic variance of the phenotypic trait (\mathbf{V}_a), and a relatedness matrix (\mathbf{F}). The model tests whether it is necessary to invoke selection in order to explain our phenotypes.

Results and Conclusions

Our findings show drastically different patterns of genomic composition and varying signals of selection across *Zea* taxa. Maize landraces show patterns of decreasing genome size along an altitudinal gradient in both geographic regions. Simulations for our model suggest that we have a low false positive rate with varying degrees of power to detect selection in many of our traits. The application of our model to landrace data shows that patterns of certain repetitive elements cannot be explained through a model of genetic drift and population structure (Figure 2). Knobs and overall transposable element content are being selected against in highland landrace populations, while patterns of increasing centromere repeats in highland varieties can be explained through neutral forces. The large variance in which transposable element subfamilies are at high abundance in the different accessions highlights the vast structural diversity present in our landrace panel (Figure 4). In applying our model of neutral evolution, we find evidence of selection on fewer subfamily tags

Figure 1. Boxplot of genome size in *Zea* accessions. For landraces, M=Mexico; SA=South America; H=Highlands; L=Lowlands.

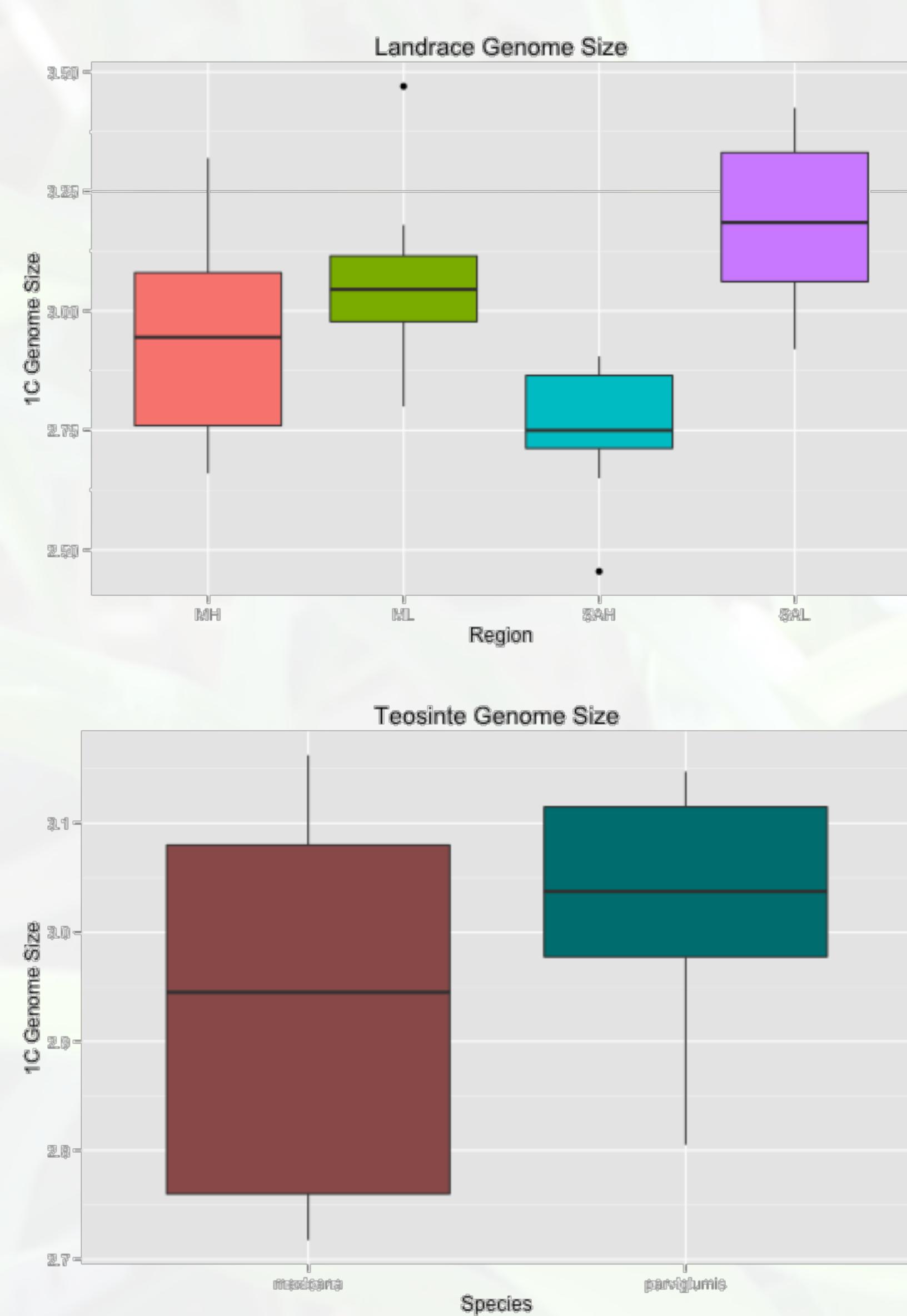


Figure 2a. Sensitivity test of our model of neutral evolution.

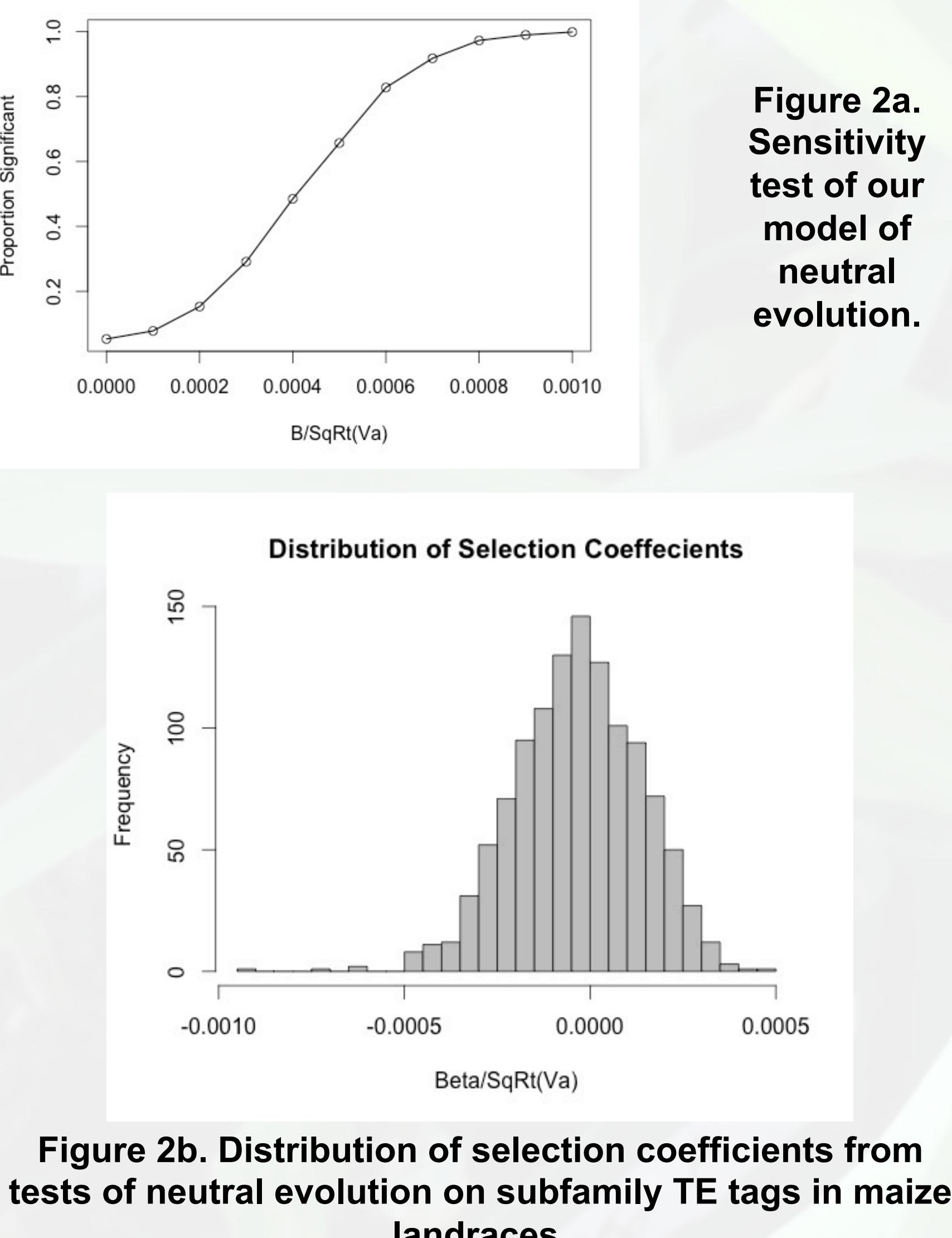


Figure 2b. Distribution of selection coefficients from tests of neutral evolution on subfamily TE tags in maize landraces.

Figure 3a. Fluorescent in-situ hybridization of *Zea mays* ssp. *mexicana*. Chromosomes are arranged from 1-10. Population names are indicated with altitude on the left, with counts for two variants of visible knobs on the right. For probe information, see Albert et al. 2010.

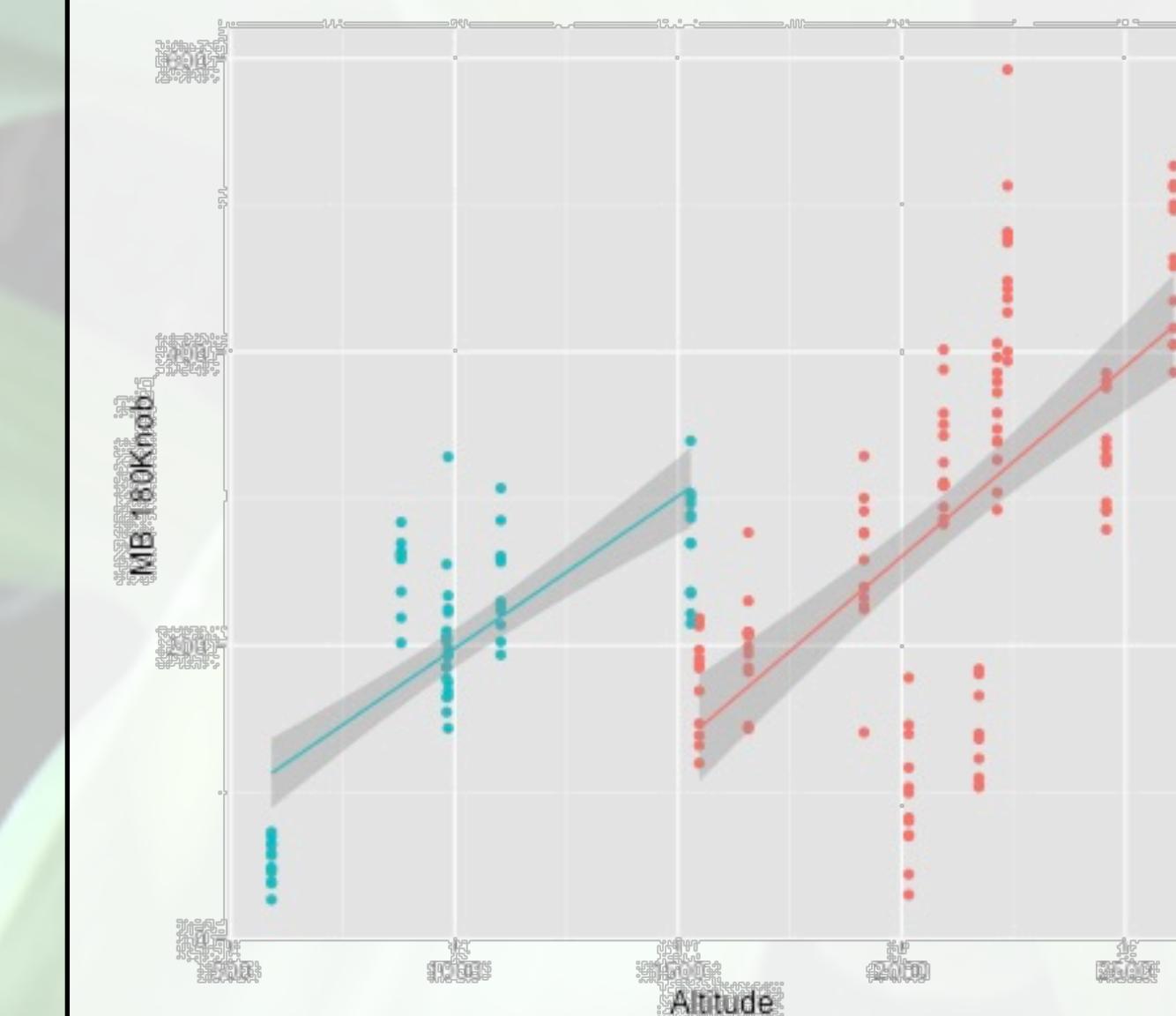
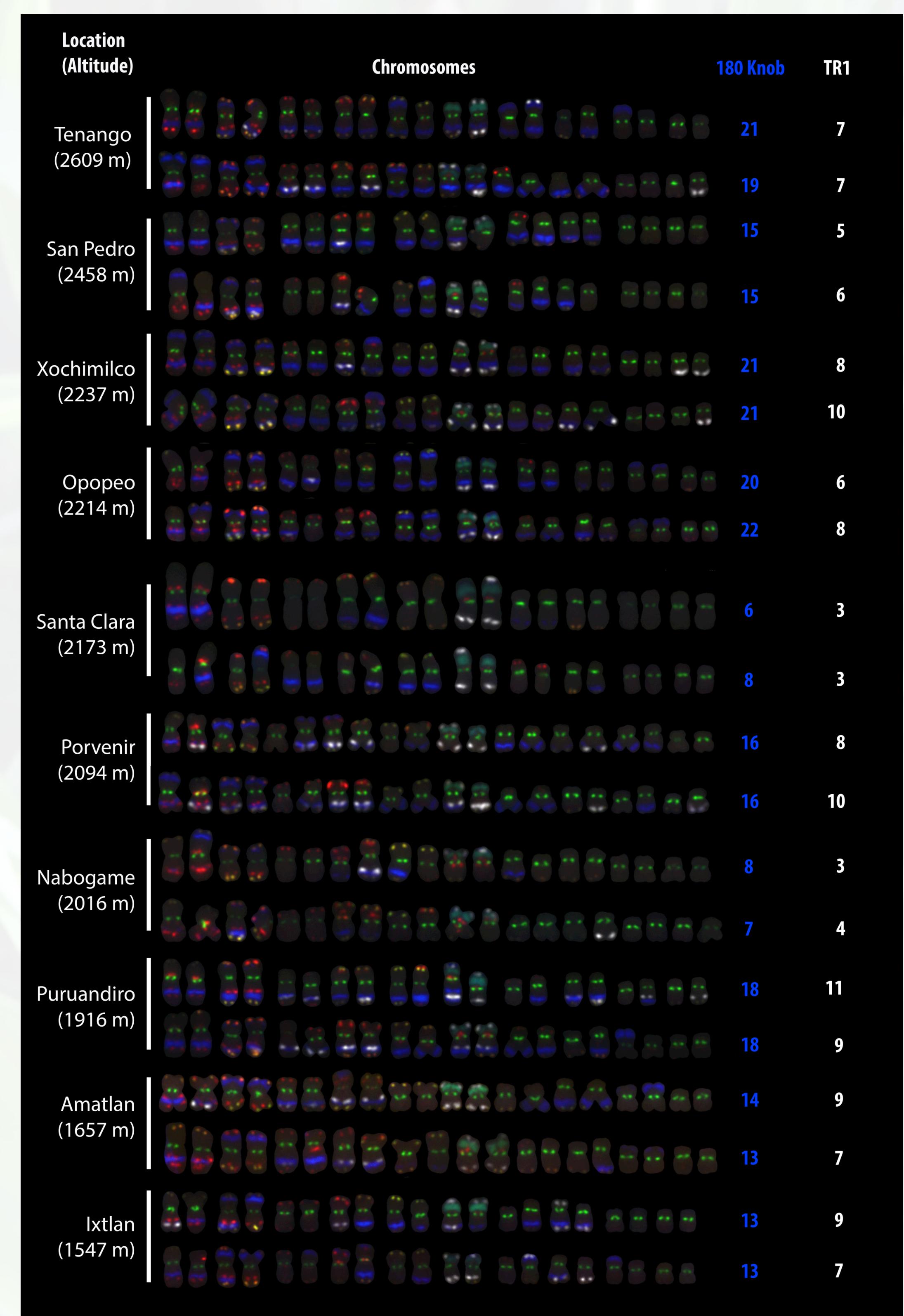


Figure 3b. Knob abundance from populations of *Zea mays* ssp. *mexicana* as measured through skim sequence data.

Scaled Genomic MB of DNA Elements

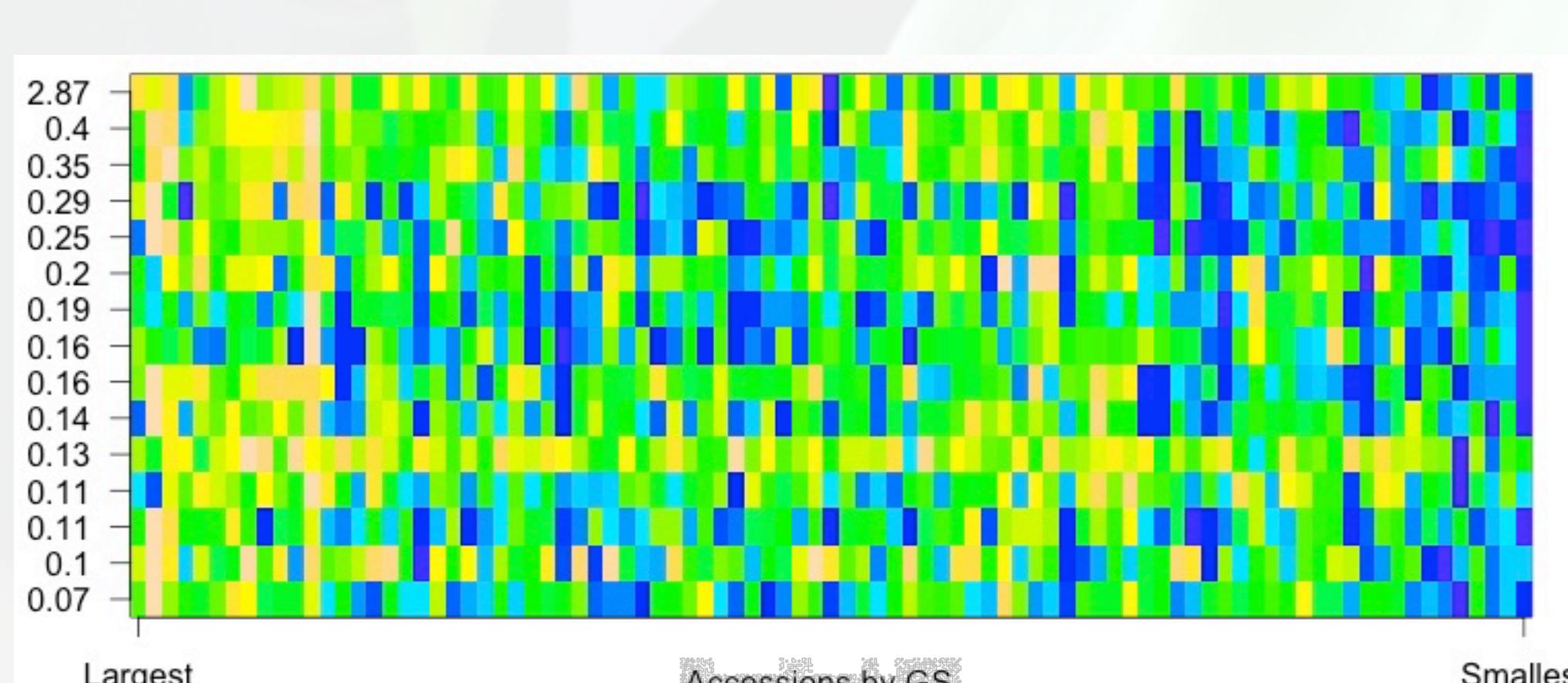
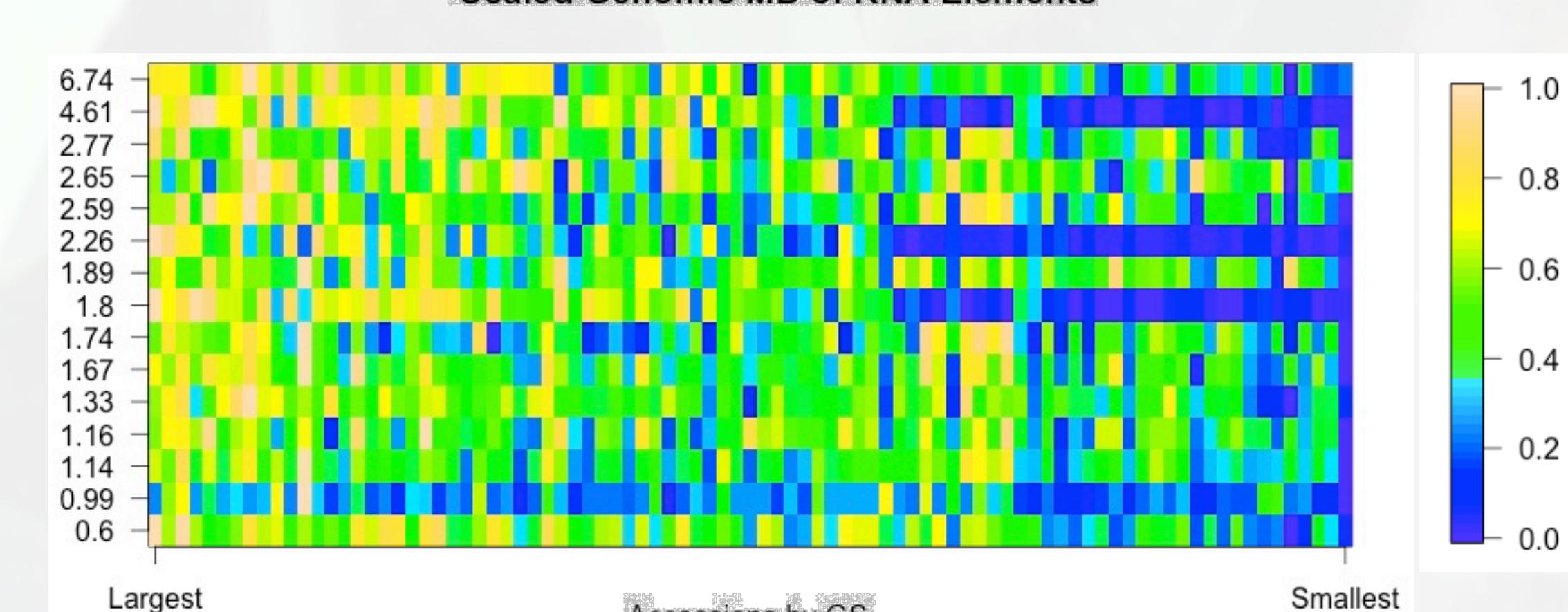


Figure 4. Transposable element diversity in maize landraces. Panel (a) shows RNA elements and Panel (b) shows DNA elements. MB contribution of each TE subfamily is shown on the Y axis, with contributions corrected for length of mapping tags and genome size. Accessions of maize landraces are ordered by genome size on the X axis.

Scaled Genomic MB of RNA Elements



than expected (20/1188) (Figure 2), contrasting our finding of selection against TEs genome wide. The dearth of selected TEs observed may indicate stabilizing selection, though the pattern could also result from other biological processes such as silencing new TE expansion or linked selection. Further investigations will help better understand these signals.

The teosinte also exhibit a pattern of lower average genome size in highland taxa, with *mexicana* having an average smaller genome than *parviflora*. However, in contrast to the patterns observed in landraces, trends within subspecies show increasing genome size along altitude, with 180bp knobs contributing greatly to the observed trend in both teosinte. TR-1 knobs were not found to increase along altitude in *mexicana*, leading to a notably different pattern of genomic composition between the teosinte subspecies (Figure 3). In the future, our expanding sample size will provide better insight as to whether the trend of increasing genome size along altitude is a result of selection in teosinte and in an antiparallel direction to the patterns observed in maize landraces.

Literature Cited

- P. S. Albert, et al., Cytogenet Genome Res 129, 6 (2010).
- R.J. Elshire, et al., PLoS One 6(5):e19379 (2011).
- J. van Heerwaarden et al., PNAS 108, 1088 (2011).
- M. B. Hufford, et al., PLoS Genetics 9, 5 (2013).
- L. B. Kanizay, et al., Heredity 110, 570 (2013).
- H. Li and R. Durbin, Bioinformatics, 25:1754-60 (2009).
- L. Poggio, et al., Annals of Botany 82, 107 (1998).
- P. Schnable, et al., Science 326, 1112 (2009).
- M.I. Tenaillon, et al., Genome Biol Evol 3, 219 (2011).

Acknowledgements

