

Pedigree-based approaches to identifying selection in modern maize

Kate Crosby¹, Oscar "Howie" Smith¹, & Jeffrey Ross-Ibarra^{1,2,3}

¹UC Davis, Department of Plant Sciences, ²UC Davis, Center For Population Biology, ³UC Davis, Genome Center



Abstract

Modern maize has resulted from generations of mass selection for fitness (yield, standability, maturity, pest resistance, etc). Yet, in terms of adaptation, data from genome scans have largely fallen short in identifying the molecular signals of selection in modern maize. This shortcoming may be due to the failure to account for population structure in the data. Here we explore whether inclusion of pedigree data allows for the identification of weak selection and selection on standing variation. We used reduced-representation genotype-by-sequence (GBS) data from more than 2000 maize inbred lines representing over 60 years of the worldwide maize germplasm. We initially reconstructed pedigree graphs with historical information alone, and then pruned them using identity-by-state (IBS) information. We intend to evaluate the power of our pedigree approach to identify loci targeted by selection during modern maize breeding. These loci will then be associated with phenotypic traits using novel population/quantitative genomics approaches.

Introduction

Modern maize inbreds have undergone repeated rounds of natural selection through time. Yield has increased through the years, and other traits such as leaf angle show dramatic directional change as well. Yet few signals of molecular selection have been found across and/or within heterotic groups, possibly due to the failure to account for population structure [1]. Pedigrees, built using a combination of historical parentage information and genetic data, can account for very minute population structure. While many pedigree records of historical parentage are incomplete or do not yet have affiliated genotypic data, there is still overlap between classic inbred accessions with parentage information that had GBS information.

Approaches

We collected over 2000 historical parentage records for US modern maize from:

- 1 Inbreds from USDA GRIN
- 2 Gerdes and Tracy [2]
- 3 Input from breeders and experts

This information was then paired with reduced representation sequence data from GBS 2.7 [3] aligned to B73 ref. genome version 3, and a concordant dataset was produced.

- 1 Pedigree graphs were constructed from historic data alone.
- 2 Pairwise kinship similarity was evaluated using Tassel [4] for known close relationships
- 3 Pedigree relationships were then weighted by kinship similarity, and edges pruned based on similarity values
- 4 Finally, the segregation of alleles from heterozygous parents were tracked to offspring.

Weighting pedigree graph

The occasional lack of concordance between the pedigree and the values produced from the kinship similarity matrix in Tassel meant that there were most likely errors in the pedigree records. To reconstruct the pedigree "graph" (G) we weighted "edges" (E) (connections between vertices (V)) by the kinship values in the similarity matrix :

$$G = (w_{ij} * E_{ij}, V)$$

For inbred parent to children kinship similarity values below 1.0 were discarded.

Pedigree graphs from historical records

We first used historical records to construct a series of directed, unweighted maize pedigree graphs (Figure 1).

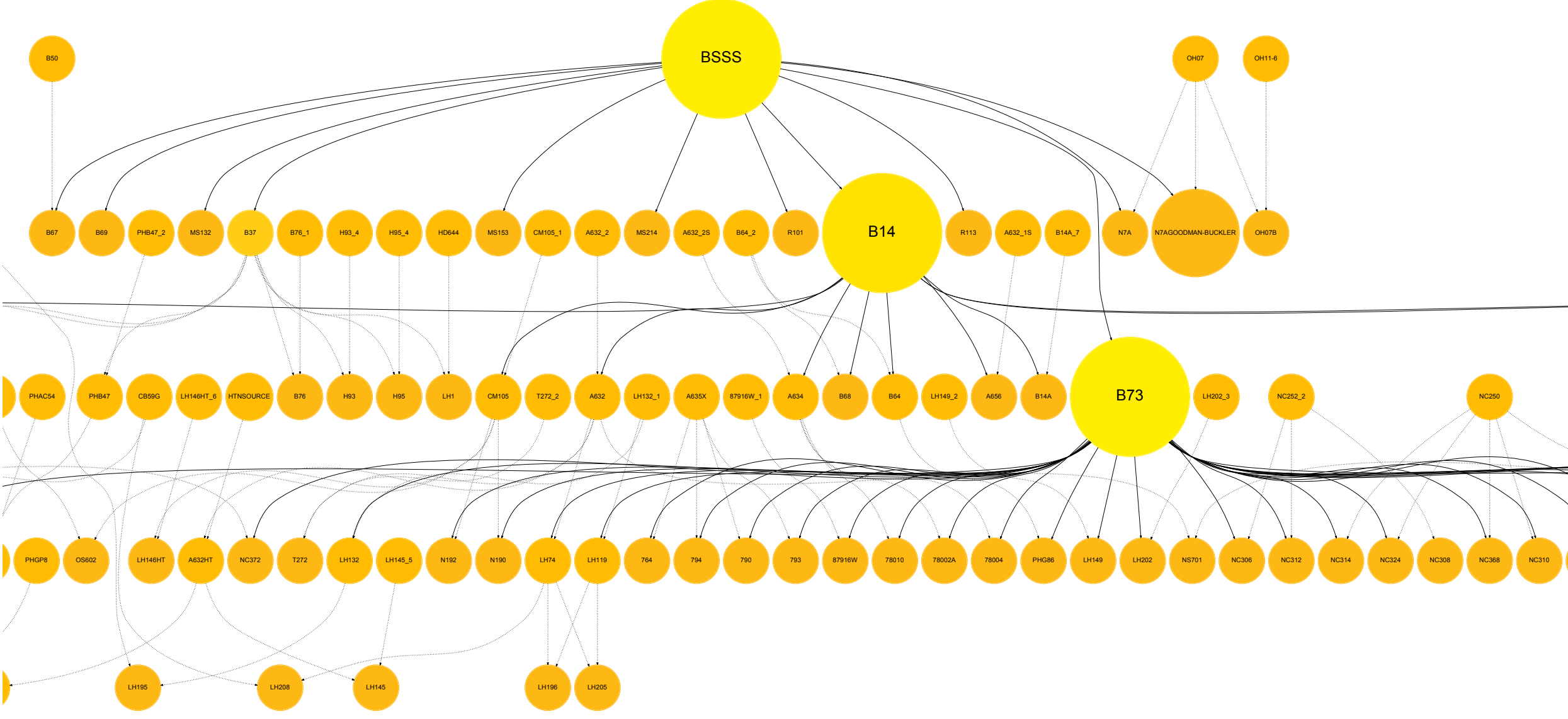


Figure 1: A portion of the modern maize pedigree with a focus on B73 (PI 550473). Length of connections do not indicate relatedness, nor does size of node indicate importance.

Kinship similarity as edge penalty

Subsets of this larger pedigree were used to examine directed parent-child relationships as well as full-sibling and half-sibling relationships. An example is shown below with B73 and its direct children. As some of these pedigrees were incomplete, cut-off values from kinship similarity matrices were used to validate historical records of relatives (parent-child and siblings) (Figure 2).

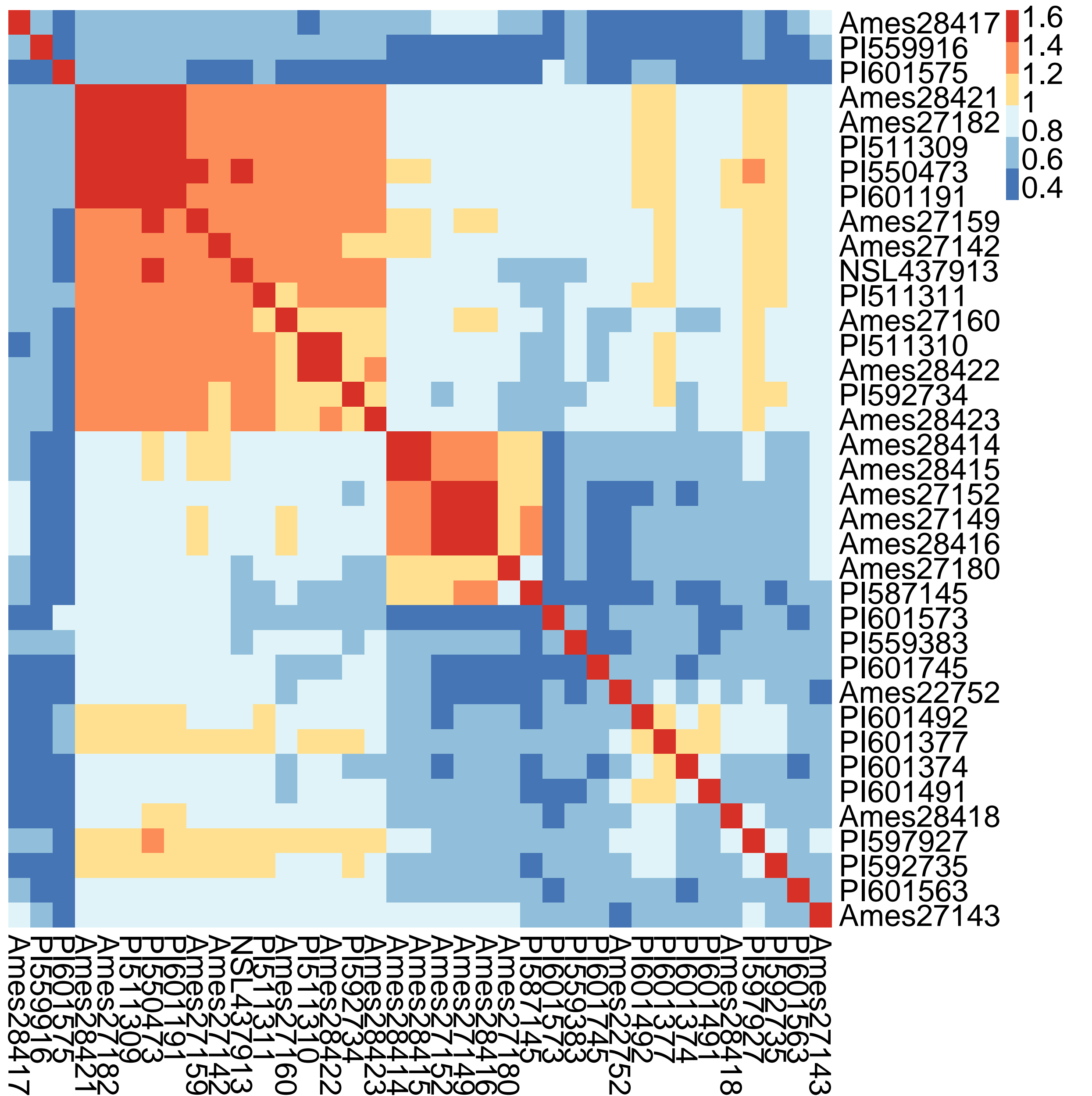


Figure 2: The pairwise kinship similarity matrix of B73 (PI 550473), and all its children from GBS 2.7 data

References

- [1] J P Gerke, J W Edwards, K E Guill, J Ross-Ibarra, and M D McMullen. The genomic impacts of drift and selection for hybrid performance in maize. *arXiv preprint arXiv:1307.7313*, 2013.
- [2] JT Gerdes, WF Tracy, JG Coors, JL Geadleemann, and Marian K Viney. *Compilation of North American maize breeding germplasm*. Crop Science Society of America Madison, Wisconsin, 1993.
- [3] J C Glaubitz, T M Casstevens, F Lu, J Harriman, R J Elshire, Qi Sun, and Edward S Buckler. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*, 9(2):e90346, February 2014.
- [4] P J Bradbury, Z Zhang, D E Kroon, T M Casstevens, Y Ramdoss, and E S Buckler. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics (Oxford, England)*, 23(19):2633–2635, October 2007.

Weighted pedigree

The pedigree of B73 with edges to children now weighted by kinship similarity (Figure 3).

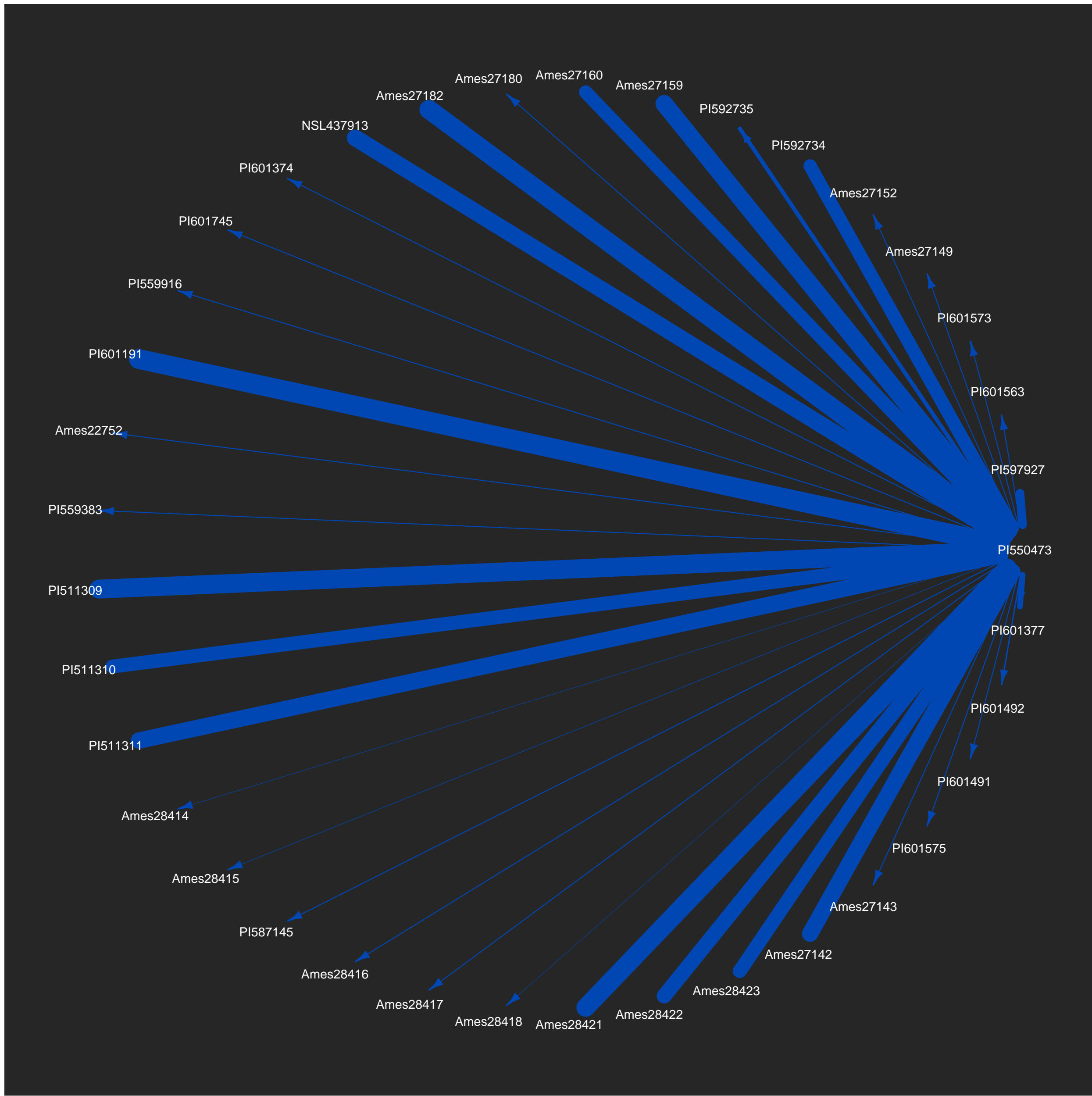


Figure 3: Weighted pedigree graph, where edge width indicates degree of similarity. Values of similarity below 1.0 were pruned - see BOX 1

Pruned pedigree

Once the pedigree is pruned, heterozygous loci in B73 (or any other parent of interest) may then be evaluated for distorted segregation in its children (Figure 4). For example, B73 is heterozygous at 379 sites (indels not included). If one allele is disproportionately represented in the next generation (Figure 4) then this may indicate directional selection at a locus either within a small pedigree or across the entire pedigree.

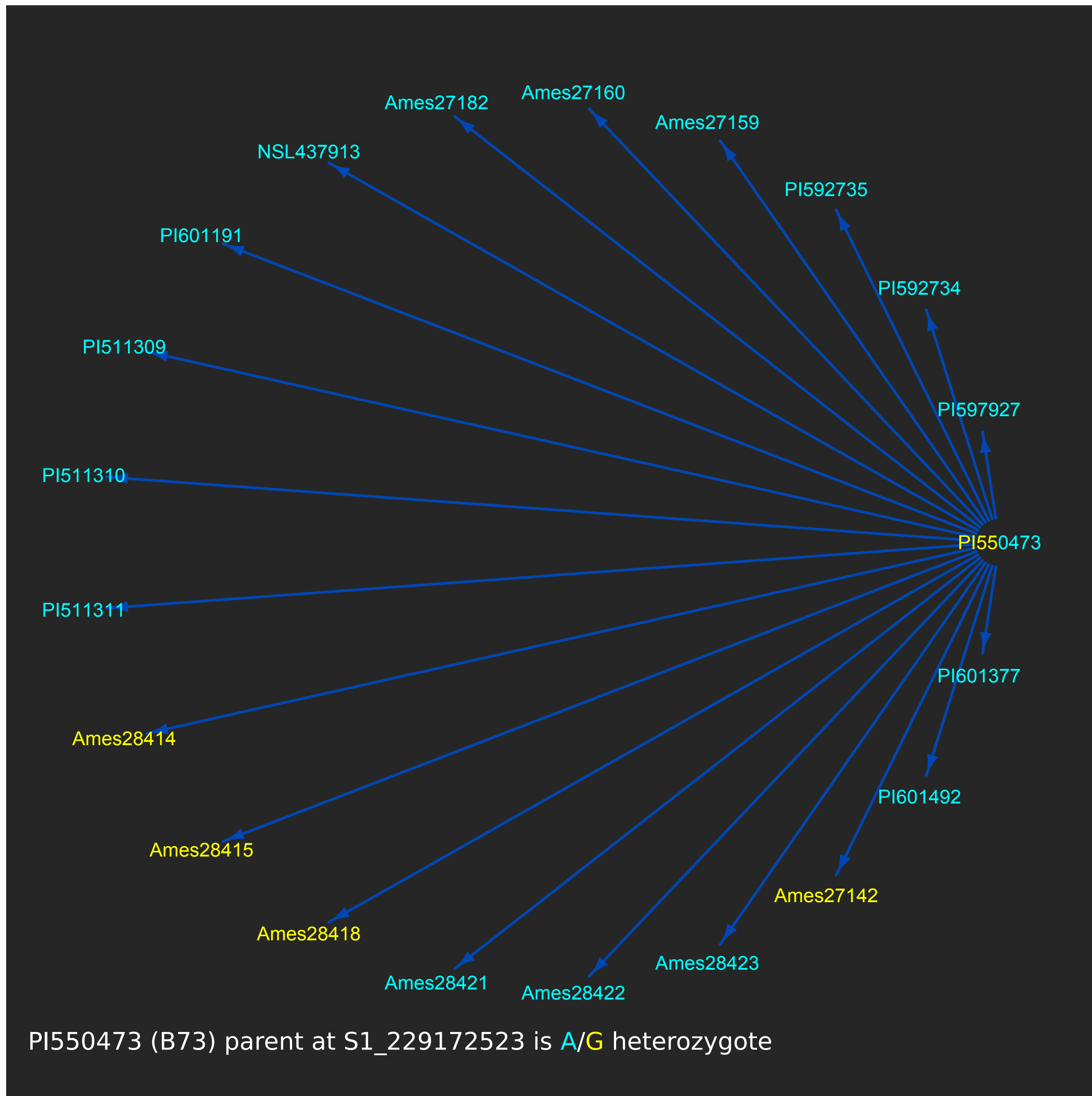


Figure 4: Pruned pedigree graph, exemplifying segregation of alleles at a heterozygous locus in B73 parent

Acknowledgements

We thank Dupont Pioneer and the USDA for funding this project. We also thank Justin Gerke for invaluable input and expertise.