

# HIGHLY ACCURATE FORENSIC APPROACHES FOR AUTHENTICATING PEDIGREES AND PROTECTING INTELLECTUAL PROPERTY IN OCTOPLOID STRAWBERRY USING HIGH-DENSITY SNP GENOTYPING ARRAYS

Mitchell J. Feldmann<sup>1</sup>, Dominique D. Pincot<sup>1</sup>, Thomas J. Poorten<sup>1</sup>, Christopher Heffelfinger<sup>2</sup>, Glenn S. Cole<sup>1</sup>, Michael A. Hardigan<sup>1</sup>, Charlotte B. Acharya<sup>1</sup>, Stephen Dellaporta<sup>2</sup>, and Steven J. Knapp<sup>1</sup>

<sup>1</sup>Department of Plant Sciences, University of California, Davis, Davis, CA 95616

<sup>2</sup>Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut, 06511

## Summary

Statistical approaches for genetically identifying parents and other close relatives in diploid organisms are extremely accurate when coupled with low-error-rate genome-scale approaches for genotyping DNA variants. The feasibility and application of such approaches have not been reported in strawberry, an allo-octoploid ( $2n = 8x = 56$ ). To study the problem, 1,287 individuals were genotyped with a 35,000-SNP genotyping array. The pedigree records for many individuals were either suspect or missing. Using exclusion analysis, parents were identified with 100% accuracy when DNA profiles for both parents were present. The accuracy slightly decreased when the DNA profile for one parent was missing. The sheer abundance of DNA variants in strawberry facilitated accurate parent identification even among closely related individuals. The authenticated pedigrees were incorporated into a pedigree database of more than 5,000 individuals developed to reconstruct the genealogy of strawberry from the 1800s to present.

## Methods

### Duo Exclusion Method

- How many non-mendelian loci exist between a focal individual and all other individuals?
- Comparing a focal individual to a single potential parent, this becomes counting the number of loci that two individuals have alternate homozygous genotypes.

$$S_i = fAA_{O_i}fBB_{P_i} + fBB_{O_i}fAA_{P_i}$$

$$F_d = \frac{1}{m} \sum_{i=1}^m S_i$$

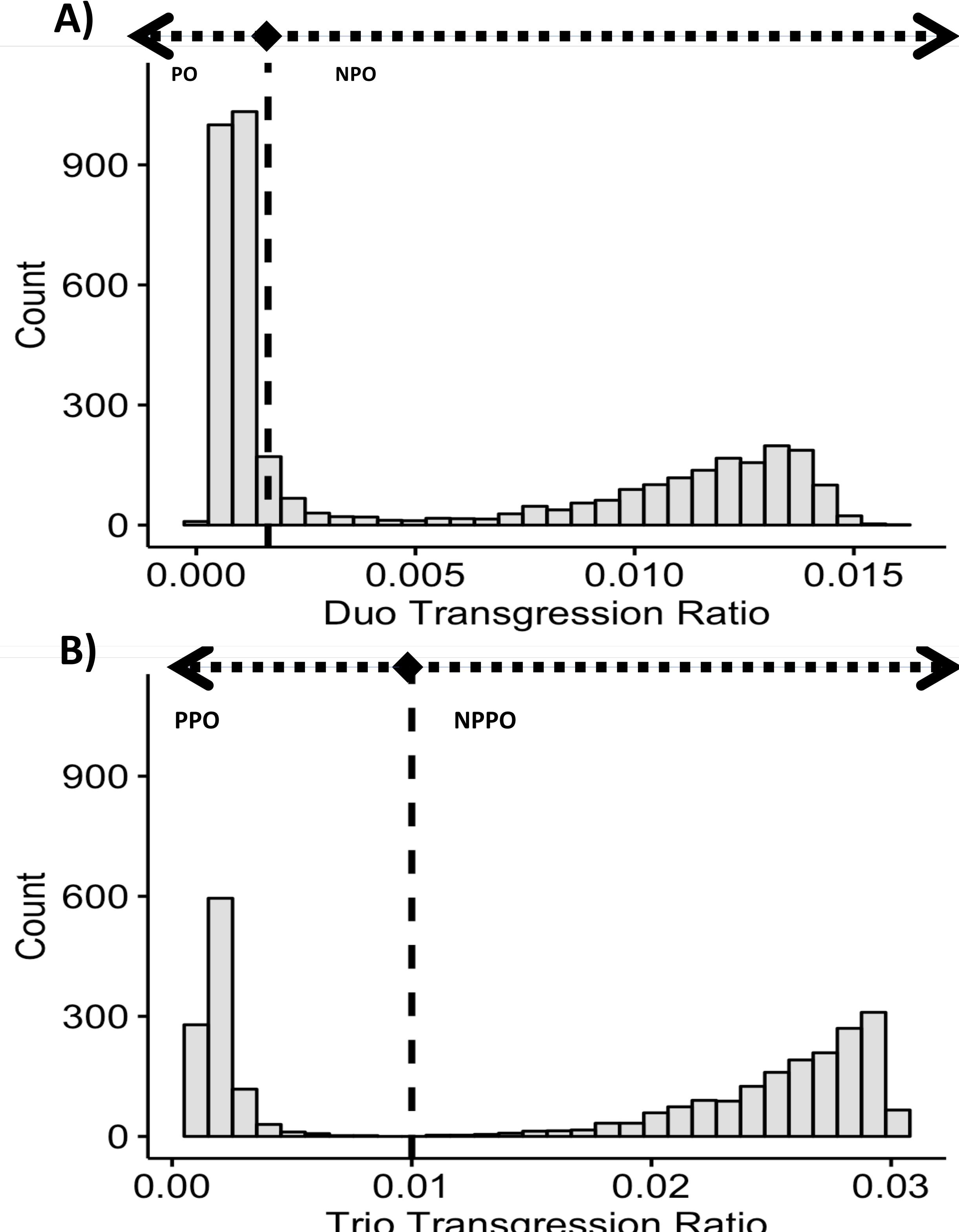
### Trio Exclusion Method

- How many non-mendelian loci exist between a focal individual and unique pairwise combinations of genotyped individuals?
- Comparing a focal individual to a potential parent pair, this includes any loci that are heterozygous in the focal individual and homozygous in both parents for the same allele.

$$T_i = fAB_{O_i}fAA_{P1_i}fAA_{P2_i} + fAB_{O_i}fBB_{P1_i}fBB_{P2_i}$$

$$F_t = \frac{1}{m} \sum_{i=1}^m T_i + S1_i + S2_i - S1_i S2_i$$

## Results



**Figure 1**

**A)** The distribution of transgression ratios when comparing a focal individual to single parents at a time. PO = inferred Parent-Offspring. NPO = Non Parent-Offspring. **B)** The distribution of transgression ratios when comparing a focal individual to a combination of potential parents. PPO = Inferred Parent-Parent-Offspring. NPPO = Non Parent-Parent-Offspring. Vertical dashed lines in each figure represent the statistical threshold for excluding potential PO or PPO combinations.

## Pedigree Network

The pedigree records of more than 5000 strawberry accessions have been obtained by thorough examination of breeding literature, governmental data, and other historical documents. Accessions from the University of California, Davis breeding program and their descendants constitute 29% of the database, and are depicted to the right as a network of relationships. Certain analyses found often in network theory may be applied. As an example, betweenness ("bottleneck-ness") an important topological property of a network, describing the number of shortest paths that pass through any single node ("accession"). Thus, nodes with the highest betweenness can control, and restrict, critical information flow ("genetics") [1]. The top 10 "bottlenecking" accessions are indicated in red.

## Application

### Problem

Our study was undertaken to statistically validate pedigree records, as well as reconstruct the pedigree database for an important germplasm collection with holdings spanning the history of the UC Davis Strawberry Breeding Program. As a demonstration, we assessed the statistical accuracy of exclusion analysis for identifying parents from genome-wide DNA profiles of 220 octoploid individuals with missing pedigree records.

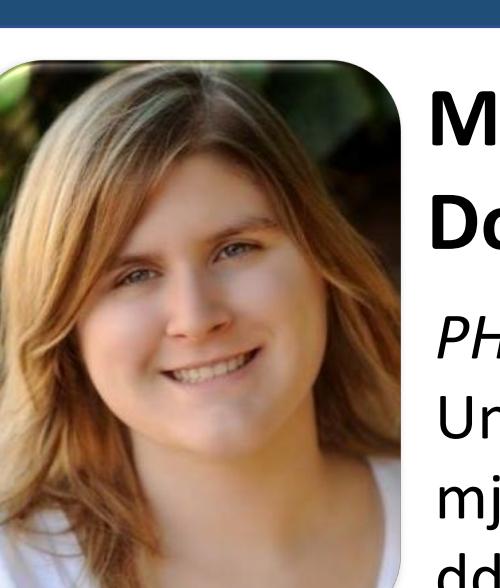
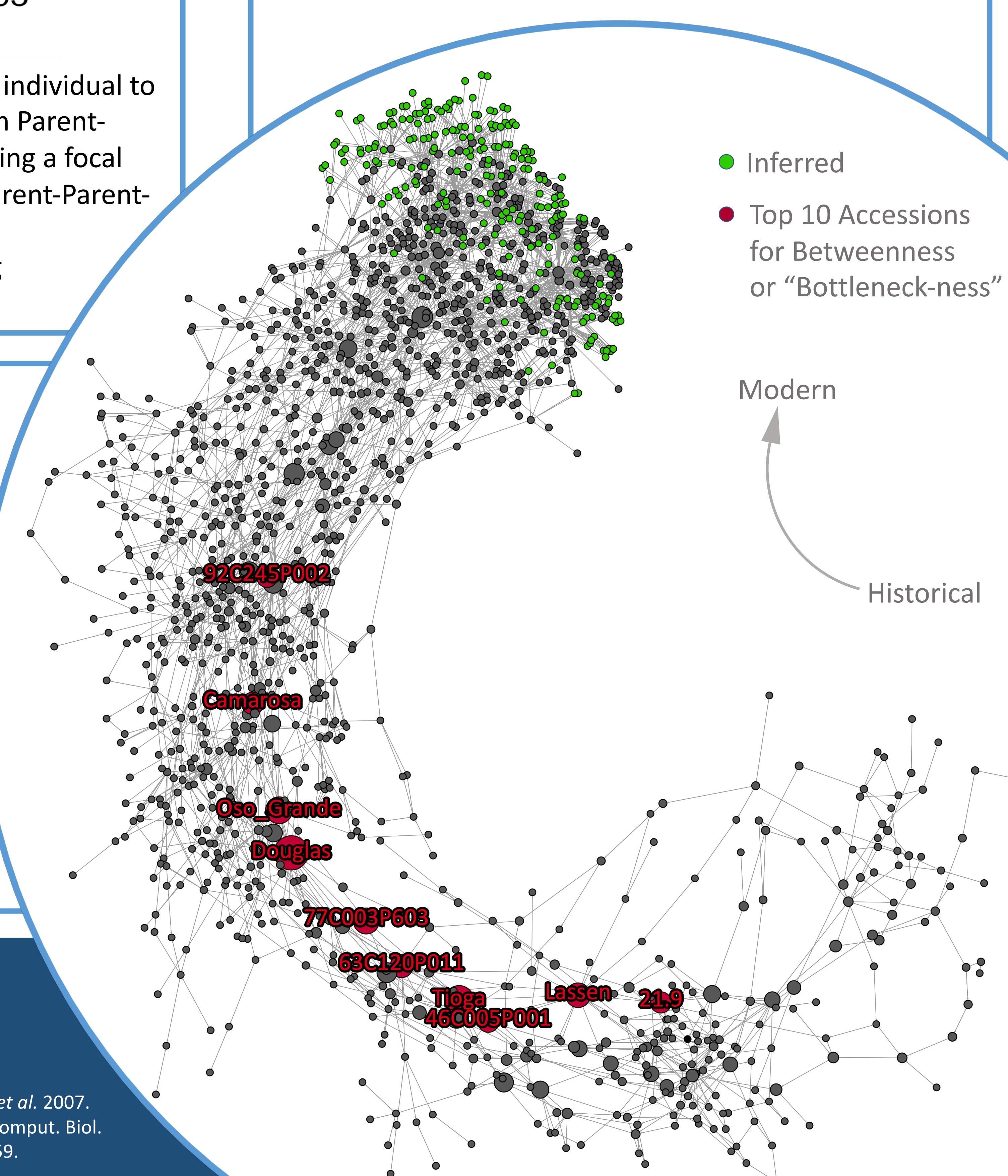
### Results

100% of individuals lacking parental information were placed in the pedigree network (depicted in green).

- Two parents were identified for 214 / 220 (97%).
- One parent was identified for final 6 / 220 (3%).

### Conclusions

- Molecular exclusion methods provide a powerful method for determining parentage by focusing on parent-offspring combinations that are most unlikely.
- A trio transgression ratio of 0.01 results in zero false positives and zero false negatives. A duo transgression ratio of 0.0016 results in zero false positives while minimizing false negatives (<2.5%).
- More broadly, we interconnected the pedigree records of cultivars worldwide with the pedigree records of holdings in the UCD collection.



**Mitchell Feldmann & Dominique Pincot**

PHD Students  
University of California, Davis  
mjfeldmann@ucdavis.edu  
ddpincot@ucdavis.edu

**Acknowledgments:** This research was supported by the University of California and grants to S.J.K. from the USDA National Institute of Food and Agriculture Specialty Crops Research Initiative (#2017-51181-26833) and California Strawberry Commission.



United States Department of Agriculture  
National Institute of Food and Agriculture

[1] Yu et al. 2007.  
PLoS Comput. Biol.  
3(4):e59.