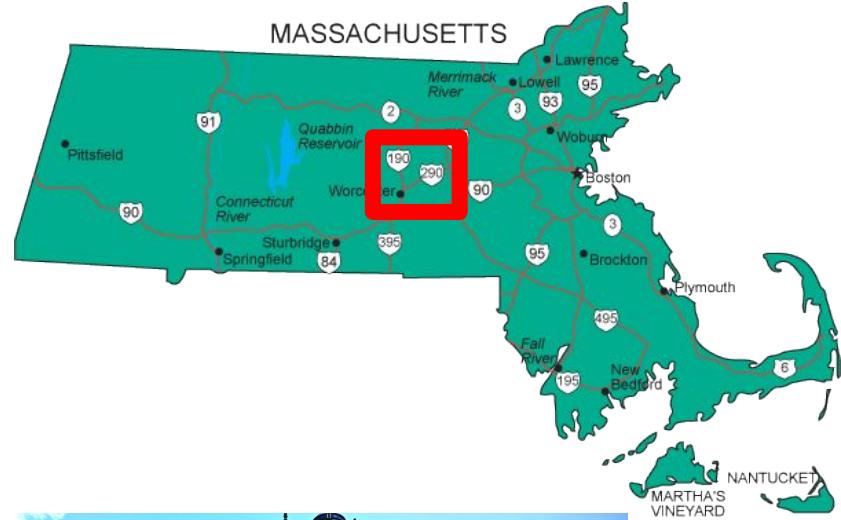
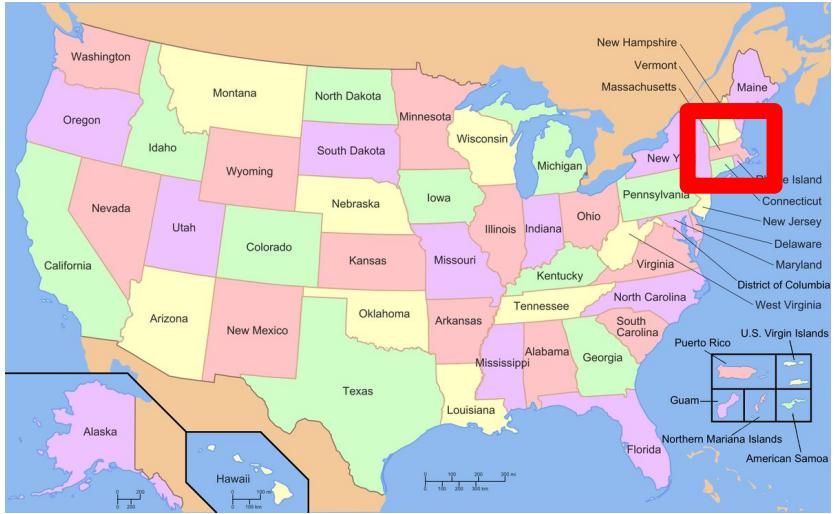


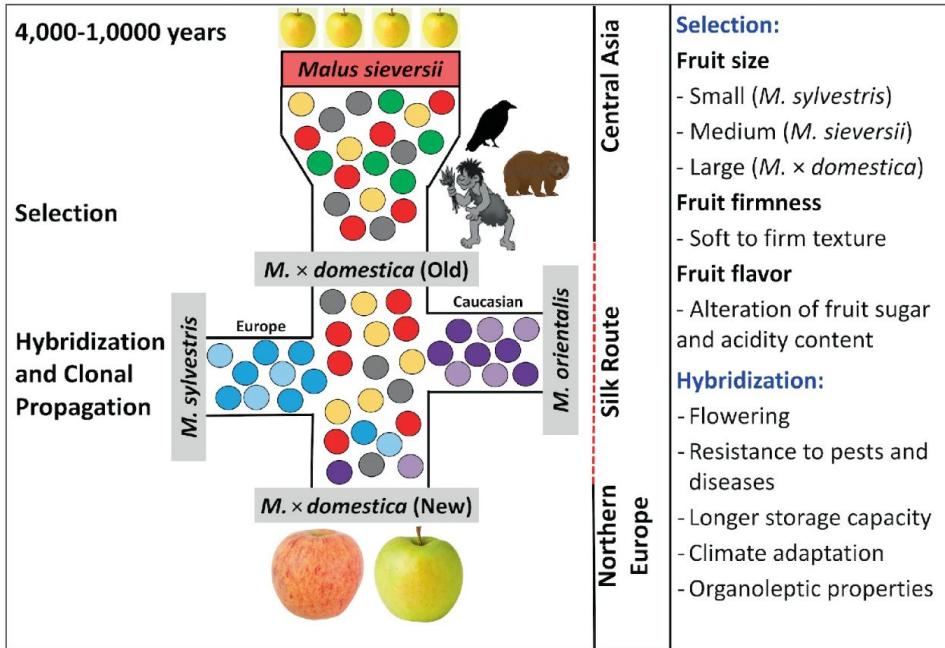
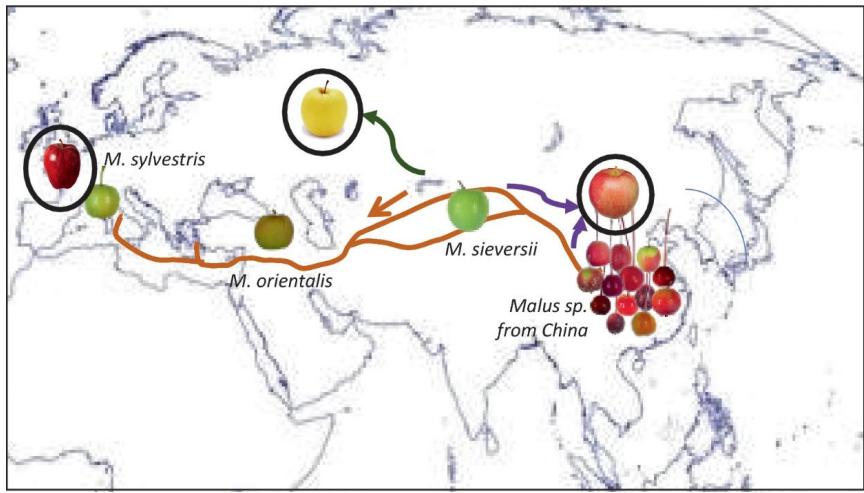
# How ‘Bout ‘Dem Apples?

Has Modern Agriculture Reduced The Genetic Diversity Of  
Apple Cultivars?









## Yellow Bellflower



NJ, USA, mid 1700s



## Winesap (Maybe)



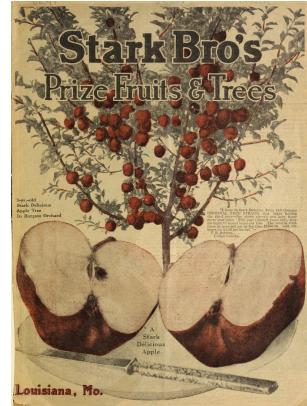
NJ, USA, first documented 1804



## Red Delicious

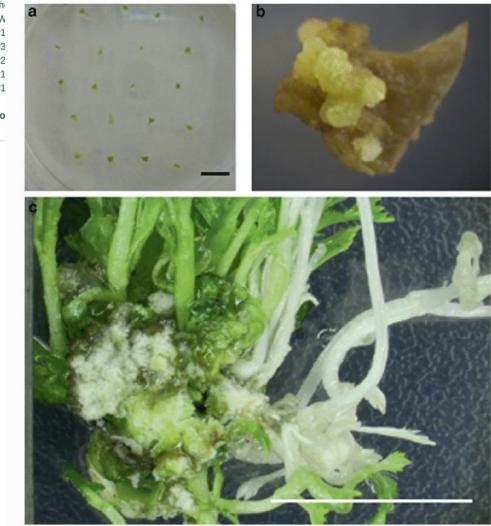
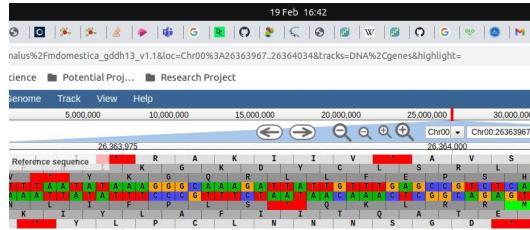


Iowa, USA, Jesse Hiatt, 1872





Screenshot of the GDR (Genome Database for Rosaceae) website for *Malus x domestica*. The page includes a navigation bar with links for Species, Data, Search, Tools, General, Help, and Community. A sidebar on the left lists various genomic resources like Overview, Description, Genes, Genetic Maps, Genomes, Germplasm, Markers, Publications, Sequences, SNP Arrays, Trait Loci, Transcripts, and Transposable Elements. The main content area displays a list of whole genome sequences and annotations for *Malus x domestica*, including Antonovka 172670-B Genome v1.0, Golden Delicious Genome NCBI annotation, 'Hanfu' v1.0 genome, Honeycrisp v1.0 genome, Gala haploid v1.0 genome, Gala diploid v1.0 genome, HFT1 Wh genome, GDDH13 W genome, and several v1.1 genomes. Below this is a section titled 'NCBI annotated gene vs. genes from'.



Transformation of CRISPR/Cas9 vector into 'JM21' apple rootstock. **a** Leaflet explants incubated on a kanamycin-containing redifferentiation medium after 1-week of co-culture with *Agrobacterium tumefaciens* harbouring a CRISPR/Cas9 vector. **b** Redifferentiating explants following 2 to 3 months of incubation. **c** Transformant genome-edited shoots (albino phenotypes) detected along basal regions of regenerated shoots; bar = 1 cm

## Genomic consequences of apple improvement

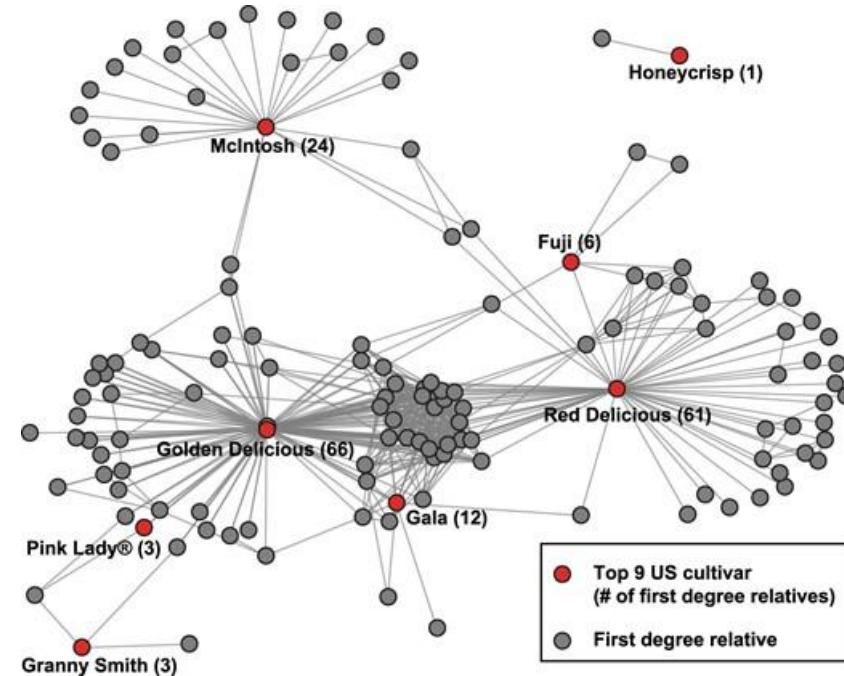
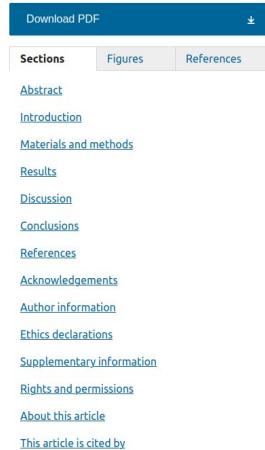
Zoë Migicovsky, Kyle M. Gardner, Christopher Richards, C. Thomas Chao, Heidi R. Schwaninger, Gennaro Fazio, Gan-Yuan Zhong & Sean Myles

Horticulture Research 8, Article number: 9 (2021) | [Cite this article](#)

8884 Accesses | 42 Citations | 117 Altmetric | [Metrics](#)

### Abstract

The apple (*Malus domestica*) is one of the world's most commercially important perennial crops and its improvement has been the focus of human effort for thousands of years. Here, we genetically characterise over 1000 apple accessions from the United States Department of Agriculture (USDA) germplasm collection using over 30,000 single-nucleotide polymorphisms (SNPs). We confirm the close genetic relationship between modern apple cultivars and their primary progenitor species, *Malus sieversii* from Central Asia, and find that cider apples derive more of their ancestry from the European crabapple, *Malus sylvestris*, than do dessert apples. We determine that most of the USDA collection is a large complex pedigree: over half of the collection is interconnected by a series of first-degree relationships. In addition, 15% of the accessions have a first-degree relationship with one of the top 8 cultivars produced in the USA. With the exception of 'Honeycrisp', the top 8



**Hypothesis:** Newly developed apple cultivars have become more related and less genetically diverse over time.

# Genotyping-by-sequencing of Canada's apple biodiversity collection

Zoë Migicovsky<sup>1</sup>

Gavin M. Douglas<sup>2</sup>

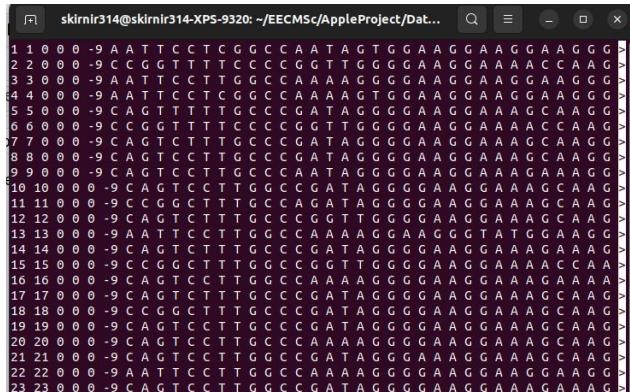
Sean Myles<sup>1\*</sup>

<sup>1</sup> Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University, Truro, NS, Canada

<sup>2</sup> Genome Centre, McGill University, Montréal, QC, Canada

## Introduction

There are over 10,000 named apple (*Malus X. domestica* Borkh) cultivars (Way et al., 1991), but most apple production relies on a small number of elite cultivars. These elite cultivars are also the primary source of breeding material used when generating new cultivars (Migicovsky et al., 2021a). Apple production and improvement could greatly benefit from incorporating more diverse cultivars for purposes including disease resistance (Khan and Korban, 2022) and unusual fruit attributes (Migicovsky and Myles, 2017). Before leveraging such potential benefits, it would first be necessary to comprehensively assess phenomic and genomic diversity across diverse apples.



1,175 accessions  
across 278,231 SNPs

# Quantifying apple diversity: A phenomic characterization of Canada's Apple Biodiversity Collection

Sophie Watts, Zoë Migicovsky, Kendra A. McClure, Cindy H. J. Yu, Beatrice Amyotte, Thomas Baker, David Bowlby, Karen Burgher-MacLellan, Laura Butler, Richard Donald, Lihua Fan ... See all authors ▾

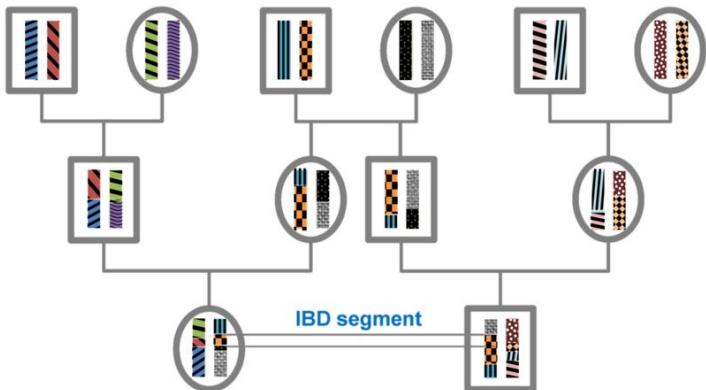
First published: 23 June 2021 | <https://doi.org/10.1002/ppp3.10211> | Citations: 17

## Funding information:

This research was supported in part by funding from the Canadian Horticultural Council (SM), Agriculture and Agri-Food Canada's AgriScience Program (SM), the Nova Scotia Department of Agriculture (SM), the Nova Scotia Fruit Growers Association (SM), Dalhousie University (SM), the Canada Research Chairs program (SM), the National Sciences and Engineering Research Council of Canada (SM), and A-Base funding (NOI-1767) from Agriculture and Agri-Food Canada (JS). ZM was supported by the National Science Foundation Plant Genome Research Program 1546869. SW was supported by a Vanier Scholarship from the National Sciences and Engineering Research Council of Canada.

| #  | A tibble: 446 × 7                         | apple_id | Cultivar            | Year      | Country | Use   | type  | Period |
|----|---|----------|---------------------|-----------|---------|-------|-------|--------|
|    |   | <dbl>    | <chr>               | <dbl>     | <chr>   | <chr> | <chr> | <dbl>  |
| 1  | 559 Adersleber Kalvill                    | 1870     | Germany             | Eating    | NA      | 1     |       |        |
| 2  | 411 Aivania                               | 1950     | Bulgaria            | Eating    | cider   | 2     |       |        |
| 3  | 245 Akane                                 | 1937     | Japan               | Eating    | dessert | 2     |       |        |
| 4  | 505 Alkmene                               | 1962     | Germany             | Eating    | NA      | 3     |       |        |
| 5  | 414 Alnarp 2                              | 1944     | Sweden              | Rootstock | NA      | 2     |       |        |
| 6  | 395 Alton                                 | 1938     | United States       | Eating    | NA      | 2     |       |        |
| 7  | 1216 Ambrosia                             | 1993     | Canada              | Eating    | dessert | 3     |       |        |
| 8  | 394 American Summer Pearmain              | 1817     | United States       | Eating    | dessert | 1     |       |        |
| 9  | 468 Ananas Berzenicki                     | 1900     | Lithuania           | Dessert   | dessert | 2     |       |        |
| 10 | 418 Anisim                                | 1870     | Former Soviet Union | Dessert   | NA      | 1     |       |        |
|    | # i 436 more rows                         |          |                     |           |         |       |       |        |
|    | # i Use `print(n = ...)` to see more rows |          |                     |           |         |       |       |        |

## Examined Relatedness Using Identity By Descent



- SNPs that are the same are all **identical by state**. If they are passed down from a common ancestor without recombination, they are **identical by descent**.
- IBD estimates return the proportion of inheritance  $\hat{\pi}$ :  $>0.125 = 3\text{rd degree relatives (cousin)}$ ,  $>0.25 = 2\text{nd degree(half siblings)}$ ,  $>0.5 = 1\text{st degree (full siblings)}$ , near 1 indicates twins/clones.

## Examined Genetic Diversity Using PCoA of Genetic Hamming Distance

```
[,1]      [,2]      [,3]      [,4]      [,5]      [,6]
[1,] 0.000000 0.176054 0.152459 0.150258 0.168394 0.188489
[2,] 0.176054 0.000000 0.178288 0.186578 0.181374 0.183655
[3,] 0.152459 0.178288 0.000000 0.131805 0.160806 0.190888
[4,] 0.150258 0.186578 0.131805 0.000000 0.167971 0.190529
[5,] 0.168394 0.181374 0.160806 0.167971 0.000000 0.191575
[6,] 0.188489 0.183655 0.190888 0.190529 0.191575 0.000000
> |
```

- Genetic Hamming distance is the percentage of SNPs that differ between accessions.
- Summarized in square distance matrix.

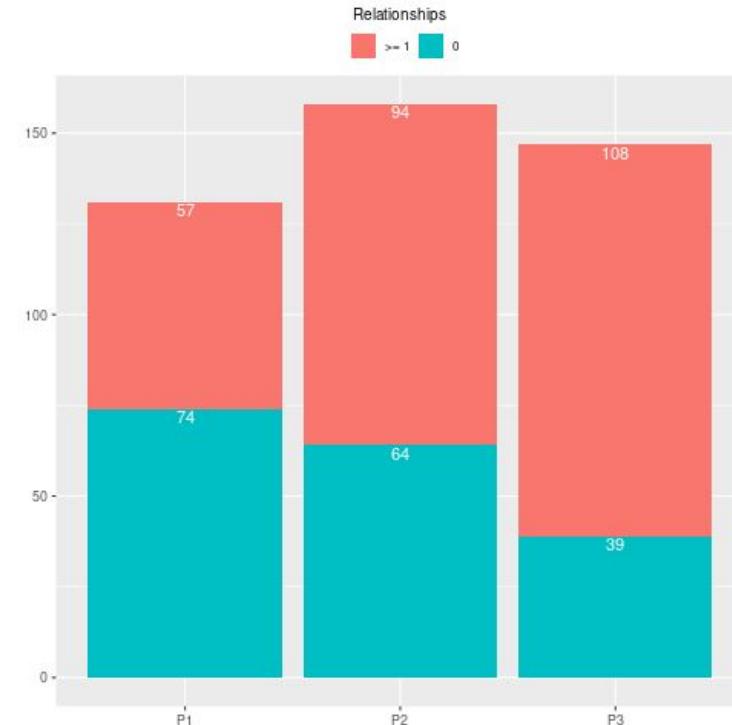
# Data were cleaned and analysis was run in Plink and R

- Filtering/cleaning:
  - Accessions were filtered to include only *Malus domestica* with a known release year.
  - Accessions further filtered to remove non-diploid cultivars (triploids would have presented as anomalously heterozygous diploids and biased results).
  - SNPs filtered based on standard protocols (e.g., remove SNPs missing from many cultivars)
  - Accessions split into three periods corresponding with different stages in apple industry development: P1 (1800-1899), P2 (1900-1959), P3 (1960-Present)
  - This left ~150k SNPs across 131 accessions in P1, 158 in P2, and 148 in P3.
- IBD estimates and genetic distance were calculated using Plink, PCoA in R.

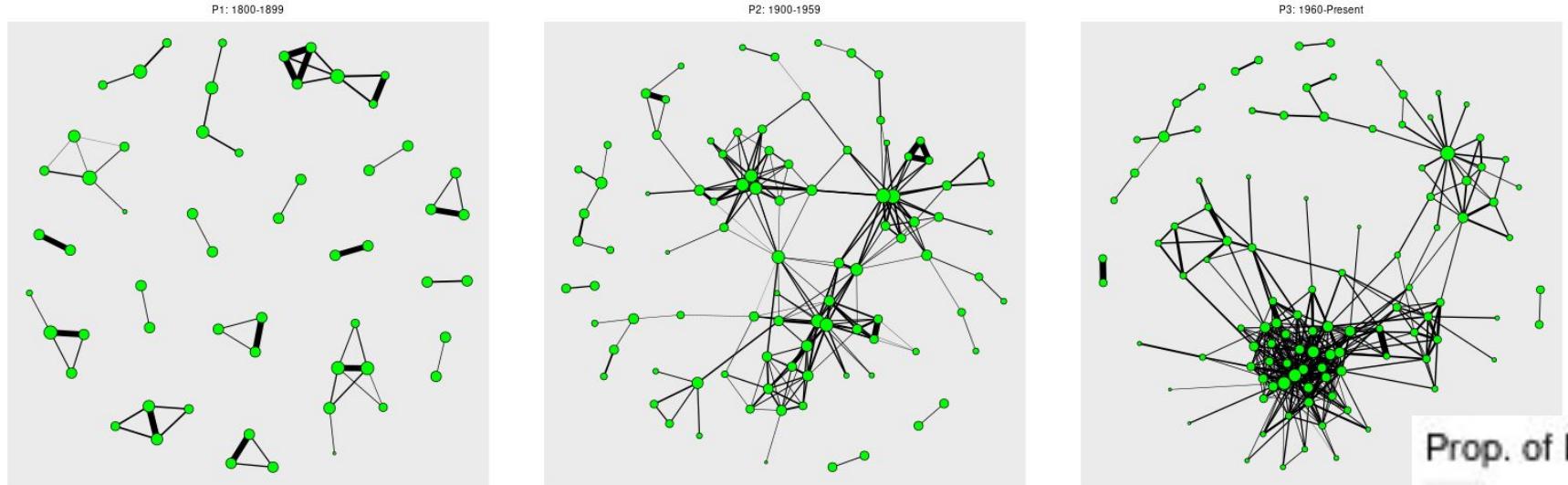
```
skirnir314@skirnir314-XPS-9320:~/EECMSC/AppleProject$ /usr/local/bin/Plink/plink  
--file ./Data/Snps/abc_combined_maf001_sort_vineland_imputed --keep ./Data/keep  
.p1.tsv --extract ./Data/Snps/plink.prune.in --allow-no-sex --geno .05 --mind .1  
--maf .01 --genome --out ./Data/Snps/p1_ibd  
PLINK v1.90b7.2 64-bit (11 Dec 2023)      www.cog-genomics.org/plink/1.9/  
(C) 2005-2023 Shaun Purcell, Christopher Chang   GNU General Public License v3  
Logging to ./Data/Snps/p1_ibd.log.  
Options in effect:  
  --allow-no-sex  
  --extract ./Data/Snps/plink.prune.in  
  --file ./Data/Snps/abc_combined_maf001_sort_vineland_imputed  
  --geno .05  
  --genome  
  --keep ./Data/keep.p1.tsv  
  --maf .01  
  --mind .1  
  --out ./Data/Snps/p1_ibd
```

# IBD Estimates Showed The Percentage Of Accessions With At Least One 3rd Degree Or Closer Relative Increased Over Time

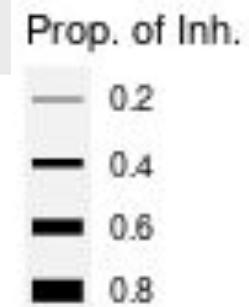
- Used a  $\hat{\pi}$  threshold of  $\geq 0.125$  (cousins or closer) to estimate relatedness.
- Based on this threshold, the % of accessions with at least one relative increased from 44% in P1 to 59% in P2 and then to 73% in P3.



# Network Graphs Of Related Accessions Showed Increased Relatedness Over Time

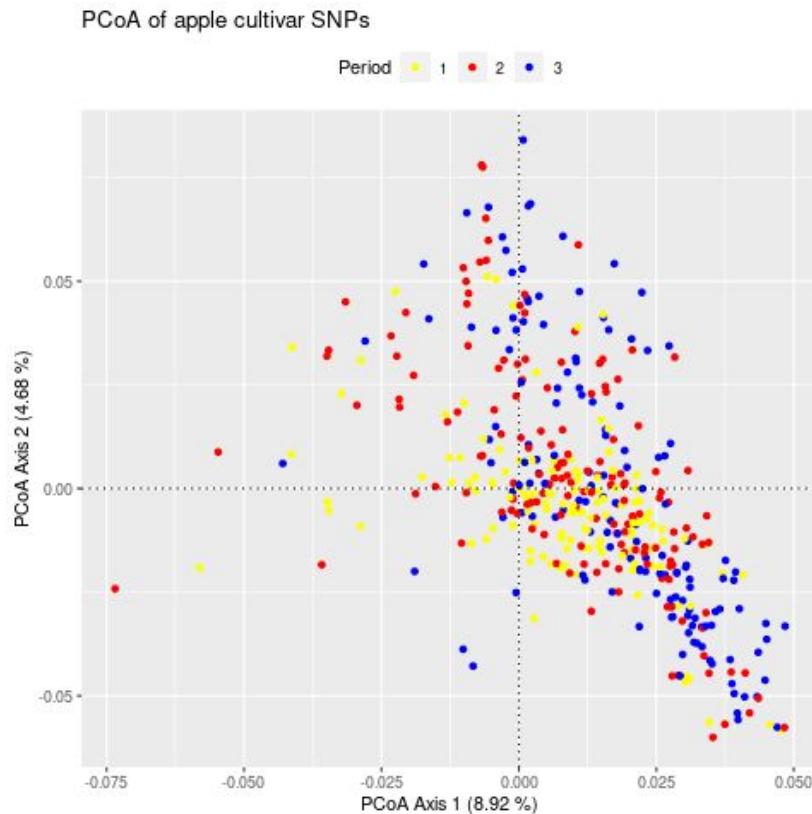


- Accessions with  $\hat{\pi} \geq 0.125$  were graphed using a Fruchterman-Reingold layout algorithm that placed accessions with higher  $\hat{\pi}$  values closer together.
- Average node degree increased from 1.9 in P1 to 4.7 in P2 and then to 7.9 in P3.
- Global efficiency (related to the inverse of distance between nodes) increased from 12% in P1 to 86% in P2 and then to 115% in P3.



# But, a PCoA of Genetic Hamming Distance Showed No Changes In Genetic Diversity Over Time

- Accessions with similar genetic distance from each other will be placed closer together on the plot.
- Plot doesn't indicate any changes in density or different groupings across periods, indicating no change in genetic distance.



# Caveats / Discussion

- Time periods were defined semi-arbitrarily
- Accessions were not a random sample from each time period
- Plink's IBD estimate is imperfect
- Excluding triploid cultivars excludes some major commercially successful cultivars
- Genetic Hamming distance is overall difference, doesn't focus on coding SNPs or SNPs that impact significant traits

That said:

- Time periods are somewhat informed by the literature
- Uses a reasonably large number of accessions
- Plink's IBD estimate checks out in a spot check

# Suggested Conclusions

- Accessions probably are becoming increasingly related to each other
- But, this does not seem to have come at the cost of decreased genetic diversity
- Regardless, we should continue to preserve heritage apple cultivars as a source of genetic diversity and a shared cultural heritage.