

How 'Bout 'Dem Apples: Has Modern Agriculture Reduced The Genetic Diversity of Apple Cultivars?

Michael Jordan

Abstract

This paper uses a data set of single nucleotide polymorphisms (SNPs) from Canada's Apple Biodiversity Collection (ABC) to assess whether the genetic diversity of apple cultivars has decreased over time.

Introduction

The domestic apple (*Malus domestica*) is one of the oldest, most commonly cultivated, and most valuable fruit crops in the world, ranking second only to the banana in gross production value (O'Rourke, 2021). Although today *M. domestica* is produced by many countries – e.g., America, Chile, New Zealand – it originated in Central Asia and evolved from other *Malus* species, primarily *M. sieversii* (Volk et al., 2021). Following domestication, *M. domestica* spread from Asia along the Silk Road to Europe and beyond (Volk et al., 2021).

As *M. domestica* spread, trees with desirable traits were developed as cultivars through vegetative propagation (Volk et al., 2021). Asexual reproduction was necessary because apples are typically self-incompatible and highly heterozygous, meaning that cultivars must be cross-bred and offspring will exhibit traits not expressed in either parent tree (Volk et al., 2021). Therefore, asexual reproduction was required to maintain desirable traits.

Prior to the industrial era apple production was regional, with small farms choosing new cultivars to develop largely through trial and error or chance pollination events (O'Rourke, 2021). After industrialization the industry became increasingly structured, complex, and globally integrated as new technology allowed apples to be sold over greater distances, leading to large vertically integrated firms today (O'Rourke, 2021). As industry organization became more sophisticated, so to did apple breeding strategies. Controlled breeding programs were established by the early 1800s (Khan et al., 2021), during the 20th century the apple industry embraced genomics, and today it uses modern technologies like SNP arrays to develop cultivars (Volk et al., 2021).

Although these structural and technological developments facilitated a large globally connected industry, this may have come at the cost of genetic diversity. Research has shown that many accessions in collections today are clonally related to just a few commercially dominant cultivars (Migicovsky et al., 2021). I hypothesize that this is a recent phenomenon, with apple cultivars becoming more related and less genetic diverse over time due to increased industry organization and the use of more sophisticated breeding techniques. If so, this would support the continued maintenance of heritage cultivars as an important reservoir of potentially valuable genotypes.

Methods

To evaluate my hypothesis, I used a [data set](#) of 278,231 SNPs from 1,175 accessions in Canada's Apple Biodiversity Collection (ABC) (Migicovsky et al., 2022). To describe these accessions, I used a [data set](#) from a phenomic review of the ABC (Watts et al., 2022). Accessions were filtered using R version 4.3.2 to include only *M. domestica* with either a cultivation or a release year (earliest was used). Because SNP data did not account for non-diploid cultivars, I also removed triploid accessions by using the `het` function in Plink v1.90 to calculate the fixation index and dropping accessions exhibiting abnormally high heterozygosity (see Ploidism-

Exploration.R). This resulted in 446 remaining accessions, which I split into three time periods – 1800-1899, 1900-1959, and 1960-Present – reflecting stages in apple industry development.

SNP quality control was conducted in Plink. SNPs were pruned for linkage disequilibrium (indep-pairwise 10 3 .05). SNPs with a minor allele frequency of less than .01 and a missingness of greater than .05 were removed. Accessions with greater than .1 missingness were removed. This resulted in 154,167 SNPs and 131 accessions for P1, 148,692 SNPs and 158 accessions for P2, and 149,354 SNPs and 148 accessions for P3.

To evaluate accession relatedness, I estimated identity by descent (IBD) using Plink. Accessions with a $\hat{\pi}$ (proportion of inheritance) greater than .125 (a 3rd degree relationships or closer) were considered related. Related accessions were visualized using the tidygraph and ggraph R packages with a Fruchterman-Reingold (FR) layout algorithm using $\hat{\pi}$ as a weight. To evaluate accession genetic diversity, I calculated the genetic hamming distance between accessions using Plink and conducted PCoA with the cmdscale function in R.

Results

Per Figure 1a, cultivar relatedness increased in each period as the percentage of accessions with at least one 3rd degree or closer relationship increased from 44% in P1 to 59% in P2 to 73% in P3. The relatedness of accessions with at least one relationship also increased in each period. This can be seen visually in Figures 1b-d. The FR layout algorithm will place accessions that are more related closer together and increased node density is evident across periods. Extracting the global efficiency score – which is inversely related to the distance between nodes – of each graph confirms this. Global efficiency increased from 12% in P1 to 86% in P2 and to 115% in P3, indicating shorter distances between nodes. Figures 1b-d also show networks becoming more connected over time. Average node degree increased from 1.9 in P1 to 4.7 in P2 and 7.9 in P3.

However, despite relatedness increasing over time, a PCoA of the genetic distance between cultivars did not show any loss in genetic diversity. we can also see in Figure 1 that the first two principle components of the genetic distance PCoA did not suggest any difference in the genetic diversity of accessions between periods. This can be concluded because accessions from all periods occupy contiguous graph space.

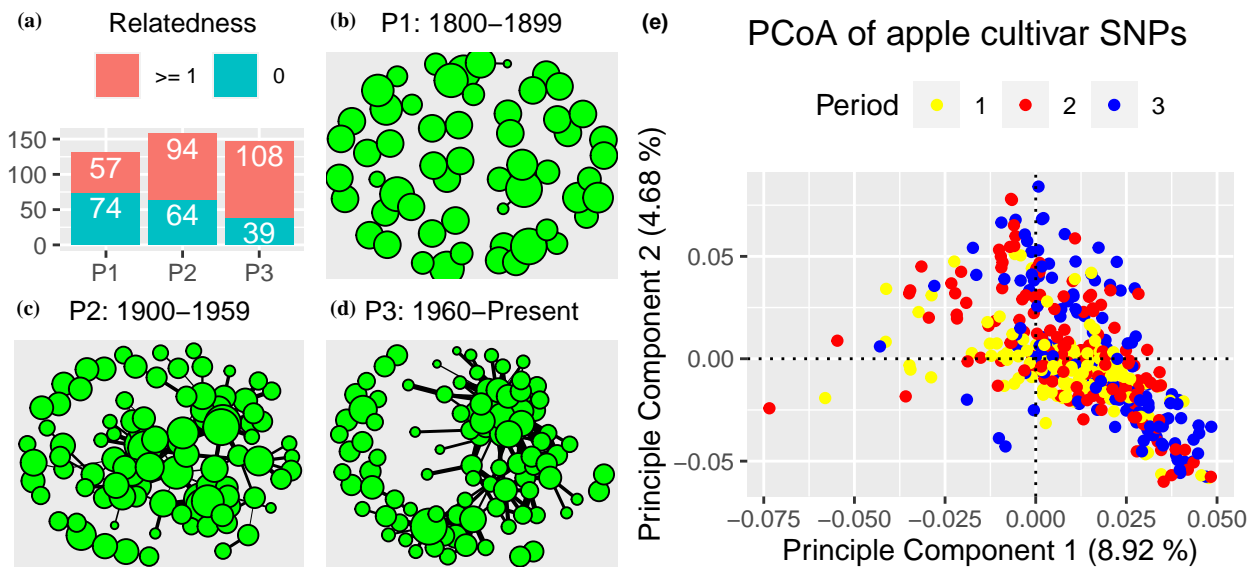


Figure 1: (a) Bar plot of accessions with 0 and >= 1 relationship by period, (b-d) Network graphs of related accessions for each period, (e) PCoA of Cultivar Genetic Distance.

Discussion

It's worth mentioning three caveats to this analysis. First, the accessions used in this study are not a random sample of cultivars developed across the different time periods, but instead represent cultivars that happened to be preserved and pass my filters. If preservation and filtering was non-random, then this analysis will be a biased measure of true cultivar relatedness and diversity across periods (and in fact the ABC is biased toward cultivars from Canada and the US).

Second, I picked time periods semi-arbitrarily. It would have been preferable to develop a more detailed understanding of exactly when the apple industry developed new breeding techniques and use that information to define period boundaries. But, this information did not arise in my reading of the literature.

Finally, Plink's estimation of IBD is imperfect. When researchers have compared Plink to other IBD methods, they have found some significant differences (Stevens et al., 2011). That said, there was also significant consistency and I was able to confirm that Plink's IBD results were reasonable by comparing them to known relationships. For example, Plink identified [Sir Prize](#) and [Clear Gold](#) as having a $\hat{\pi}$ of 56% indicating a 1st degree relationship. This makes sense, since Clear Gold is a mutation of a Golden Delicious and Sir Prize is a cross between Golden Delicious and another cultivar.

Caveats aside, I think these results are cause for both caution and peace of mind regarding apple cultivar genetic diversity. On the one hand, newly developed cultivars did become more related to each other over time, supporting the idea that the industry may be increasingly relying on a small number of cultivars to develop new apples whereas in the past it may have used a wider variety. On the other hand, a PCoA of genetic distance between cultivars did not show any corresponding loss in genetic diversity over time. My read of these two seemingly disparate results is that the extremely high heterozygosity of apples may have prevented a genetic bottleneck up to now despite the increasing relatedness of new cultivars.

Conclusion

While it is comforting to know that we are not at a stage of apple domestication where breeding has significantly constrained genetic diversity, I would still advocate for continued preservation of heritage apple cultivars on two bases. First, although the increased reliance on a small number of cultivars for breeding new apples has not yet significantly constrained apple genetic diversity, it may in the future and in the case that it does having heritage cultivars available to use for selecting new traits will be important. Second, heritage cultivars represent a window into what phenotypes people found valuable in the past, which is interesting historically and is in a sense a type of cultural heritage that in my opinion is worth preserving.

Reference List

- Khan, A., Gutierrez, B., Chao, C.T., & Singh, J. (2021) [Origin of the Domesticated Apples](#). In: Podwyszyńska, M., & Marasek-Ciołakowska, A. (eds.) (2021) *The Apple Genome*. Springer Cham, pp. 383-394.
- Migicovsky, Z., Douglas, G.M., & Myles, S. (2022) [Genotyping-by-sequencing of Canada's apple biodiversity collection](#). *Genomics of Plants and the Phytoccosystem*. 13.
- Migicovsky, Z., Gardner, K.M., Richards, C., Chao, C.T., Schwaninger, H.R., Fazio, G., Zhong, G., & Myles, S. (2021) [Genomic consequences of apple improvement](#). *Horticultural Research*. 8.
- O'Rourke, D. (2021) [Economic Importance of the World Apple Industry](#). In: Podwyszyńska, M., & Marasek-Ciołakowska, A. (eds.) (2021) *The Apple Genome*. Springer Cham, pp. 1-18.

- Stevens, E.L., Heckenberg, G., Roberson, E.D.O., Baugher, J.D., Downey, T.J., & Pevsner, J. (2011) [Inference of Relationships in Population Data Using Identity-by-Descent and Identity-by-State](#). Plos Genetics.
- Volk, G., Cornille, A., Durel, C., & Gutierrez, B. (2021) [Botany, Taxonomy, and Origins of the Apple](#). In: Podwyszyńska, M., & Marasek-Ciołakowska, A. (eds.) (2021) *The Apple Genome*. Springer Cham, pp. 19-32.
- Watts, S., Migicovsky, Z., & Myles, S. (2023) [Large-scale apple GWAS reveals NAC18.1 as a master regulator of ripening traits](#). *Fruit Research*. 3.

Working Style Assessment