

How About 'Dem Apples: Has Modern Agriculture Reduced Apple Cultivars Genetic Diversity?

Michael Jordan

Abstract

This paper uses a dataset of 278,231 SNPs from 1,175 accessions in Canada's Apple Biodiversity Collection (ABC) to assess whether the genetic diversity of newly developed 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 global gross production value (O'Rourke, 2021). Although today *M. domestica* is produced by many countries – e.g., America, Chile, New Zealand – it likely originated in Central Asia or Western China and evolved from other *Malus* species, primarily *M. sieversii* (Volk et al., 2021). Following initial 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, although apples can reproduce sexually, they are typically self-incompatible and highly heterozygous, meaning that a given cultivar must be pollinated by a different cultivar and that the seedlings will exhibit traits not expressed in either parent tree (Volk et al., 2021). Ergo, asexual reproduction was required to maintain desirable traits.

Prior to the industrial era apple production was regional, with small farms breeding cultivars for local tastes using trial and error (O'Rourke, 2021). After industrialization the apple industry became increasingly structured, complex, and globally integrated as new technology allowed apples to be sold over greater distances (O'Rourke, 2021). Today, the apple industry is dominated by large vertically integrated firms that handle all aspects from production to packing to marketing (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 uses technologies like genomic sequencing to develop cultivars with desirable traits (Volk et al., 2021).

These structural and technological developments have facilitated a large globally connected industry, but they may have come at the cost of genetic diversity. Pre-industrialization, small farmers would have presumably combined whatever cultivars were locally available to them and the fragmented nature of the market would have meant high global cultivar diversity. But, genetic research has shown that many cultivars in collections today are clonally related to just a few commercially dominant cultivars such as McIntosh (Migicovsky et al., 2021).

However, whether the genetic diversity of newly developed cultivars has actually changed over time is an open question. I hypothesize that the genetic diversity of new apple cultivars has decreased as breeding programs have increasingly exploited a small number of commercially successful cultivars. If so, this would support the continued maintenance of heritage cultivars as an important reservoir of potentially valuable genotypes.

Methods

To evaluate apple cultivar genetic diversity, I used a dataset of 278,231 SNPs from 1,175 accessions in Canada's Apple Biodiversity Collection (ABC) that were produced using genotype-by-sequencing and made publicly

available in [Dryad](#) (Migicovsky et al., 2022). To described these accessions, I used a dataset produced as part of Watts et al.'s 2022 phenomic characterization of the ABC and made available on [their lab's github](#).

Accessions were filtered using R version 4.3.2 to include only *M. domestica* accessions with either a cultivation year or a release year (earliest year was used), resulting in 446 remaining accessions. These were split into three time periods: prior to 1850, 1850-1950, and 1950-2024. These time periods were semi-arbitrary, intended to reflect stages in industry development from pre-industrialization to the modern era.

Accessions were also filtered to remove likely polyploid cultivars by using the `het` function in Plink v.190 to calculate the fixation index per accession. Accessions with a fixation index less than 0 were dropped (see `PloidismExploration.R` for details). SNPs were pruned for linkage disequilibrium using the `indep-pairwise` function in Plink. A window size of 10 SNPs, a window shift of 3 BPs, and an R^2 of .05 were used. SNPs with a minor allele frequency of less than .01 and a missingness of greater than .05 were also removed. Accessions with greater than .1 missingness were removed. This resulted in 154167 SNPs and 131 accessions for P1, 148692 SNPs and 158 accessions for P2, and 149354 SNPs and 148 accessions for P3.

I estimated IBD among cultivars launched in each period as well as the genetic distance (Hamming distance) among cultivars across all periods using Plink. IBD networks were visualized using the `tidygraph` and `ggraph` R packages with a Fruchterman-Reingold layout algorithm. Only cultivars with a $\hat{\pi}$ (proportion of inheritance) greater than .125 were considered related (i.e., I filtered for 3rd degree relationships or closer). Genetic distance PCA was completed in R using the same accessions and SNPs.

Results

Results showed that cultivar relatedness increased in each period. Looking at accession relatedness using IBD, the percentage of accessions that had at least one relationship increased from 44% in P1 to 59% in P2 to 73% in P3. Plotting networks of related accessions, we can see that cultivars became more related to each other in each period as indicated visually by increased number of connections and density of nodes. Extracting the global efficiency of each graph confirms this. From 12% in P1, it increases to 86% in P2 and then 115% in P3.

However, the PCA of the first two axes of ordination of genetic distance between accessions did not suggest any difference in the genetic diversity of accessions between periods as confirmed by the fact that SNPs from all periods occupied roughly the same space on the graph.

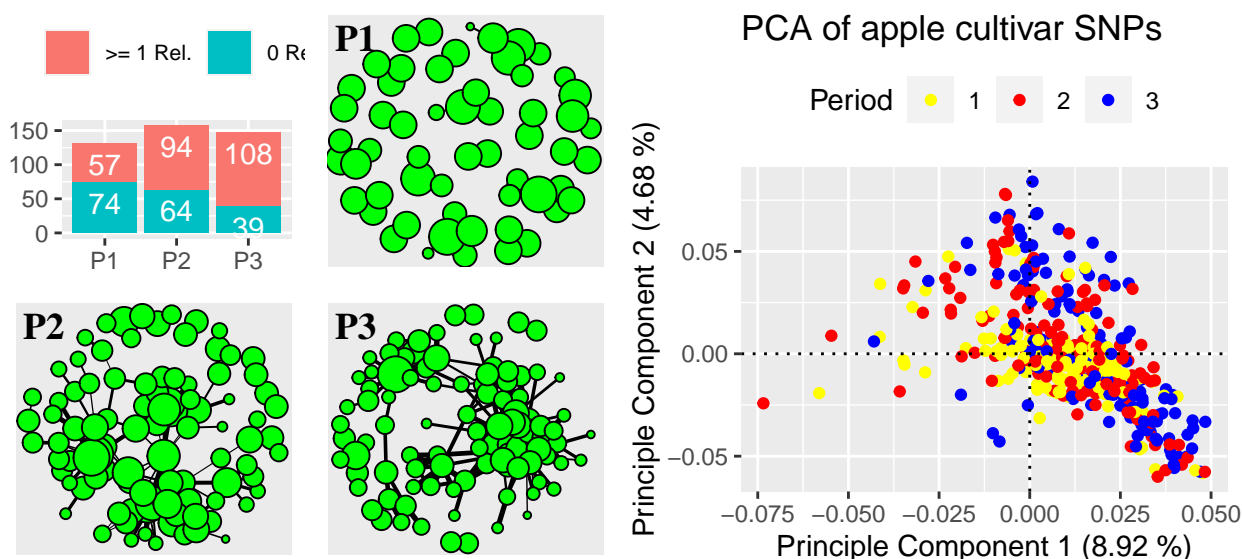


Figure 1: Network graphs related accessions and PCA of SNPs across all accessions.

Discussion

At first glance, these data suggest two disparate conclusions. On the one hand, new cultivars are becoming more closely related. On the other hand, they are no less genetically diverse. This circle can be squared, but, first, it's worth mentioning some caveats.

Notably, the accessions used in this study are not a random sample of cultivars developed in the different time periods, but instead represent cultivars that happened to be preserved. If preservation was non-random, then they may not reflect true cultivar diversity in the periods (and in fact the ABC is biased toward cultivars from Canada and the US).

Another important caveat is that Plink's estimate of IBD is imperfect as described in Stevens et al. (2011), where Plink was compared to another methods. That said, although those studies captured a difference between Plink and other methods, there was significant consistency between methods in both. Additionally, spot-checking Plink's results versus known relationships confirms that the results are reasonable. For example, in P3 Plink identified [Redspur Delicious](#) and [BisBee](#) as having a $\hat{\pi}$ of indicating a clonal relationship and these two cultivars are both known to be a mutation of the [Starking Delicious](#) cultivar.

Overall, I believe these results suggest that apple breeders have become increasingly reliant on a smaller number of cultivars. However, the high heterozygosity of apples has prevented this from resulting in a lack of genetic diversity.

Conclusion

Although we may not be at a stage of domestication with the apple where breeding has significantly constrained genetic variation in favor of a small number of genotypes that produce phenotypes valued by the global marketplace, I would still advocate for continued preservation of heritage apple cultivars. They represent not just a window into what phenotypes people found valuable in the past, but also a potential source of inspiration for the development of new cultivars. Although the global market may push breeders to develop apples with consistent phenotypes in terms of storage, acidity, flavor, etc., consumers also like novelty. Using heritage cultivars as a springboard may be the best way forward in terms of commercial breeding.

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, Edition 1*. 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 Phytoecosystem*. 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, Edition 1*. 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.M., Cornille, A., Durel, C.E., & Gutierrez, B. (2021) [Botany, Taxonomy, and Origins of the Apple](#). In: Podwyszyńska, M., & Marasek-Ciołakowska, A. (eds.) (2021) *The Apple Genome, Edition 1*. 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.

Bibliography

- Baric, S., Storti, A., Hofer, M., Guerra, W., & Dalla Via, J. (2020) [Molecular Genetic Identification of Apple Cultivars Based on Microsatellite DNA Analysis. I. The Database of 600 Validated Profiles.](#) *Erwerbs-Obstbau.* 62, 117-154.
- Mészáros, G., Mészárosová, M., & Geretschläger, A. (2020) [Comparison of identity by descent estimates with Plink and refinedIBD in dogs.](#) *Acta Fytotechn Zootechn.* 23 (4), 213-216.
- Watts, S., Migicovsky, Z., McClure, K.A., Yu, C.H.J., Amyotte, B., Baker, T., Bowlby, D., Burgher-MacLellan, K., Butler, L., Donald, R., Fan, L., Fillmore, S., Flewelling, J., Gardner, K., Hodges, M., Hughes, T., Jagadeesan, V., Lewis, N., MacDonell, E., MacVicar, L., McElroy, M., Money, D., O'Hara, M., Ong, Q., Palmer, L.C., Sawler, J., Vinqvist-Tymchuk, M., Rupasinghe, HPV., DeLong, J.M., Forney, C.F., Song, J., & Myles, S. (2021) Quantifying apple diversity: A phenomic characterization of Canada's Apple Biodiversity Collection. *Plants People Planet.* 3 (6), 747-760. <https://doi.org/10.1002/ppp3.10211>
- Xue, H., Zhang, B., Tian, J., Chen, M., Zhang, Y., Zhang, Z., & Ma, Y. (2017) Comparison of the morphology, growth and development of diploid and autotetraploid 'Hanfu' apple trees. *Scientia Horticulturae.* 225, 277-285. <https://doi.org/10.1016/j.scienta.2017.06.059>
- Podwyszyńska, M., Markiewicz, M., Matysiak, B., & Marasek-Ciołakowska, A. (2021) Apple Autotetraploids with Enhanced Resistance to Apple Scab (*Venturia inaequalis*) Due to Genome Duplication-Phenotypic and Genetic Evaluation. *International Journal of Molecular Sciences.* 22 (2), 527. <https://doi.org/10.3390/ijms22020527>
- Robinson, D.O., Coate, J.E., Singh, A., Hong, L., Bush, M., Doyle, J.J., & Roeder, A.H.K. (2018) Ploidy and Size at Multiple Scales in the Arabidopsis Sepal. *The Plant Cell.* 30 (10), 2308-2329. <https://doi.org/10.1105/tpc.18.00344>