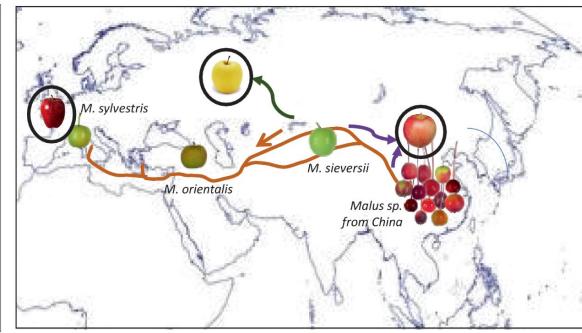
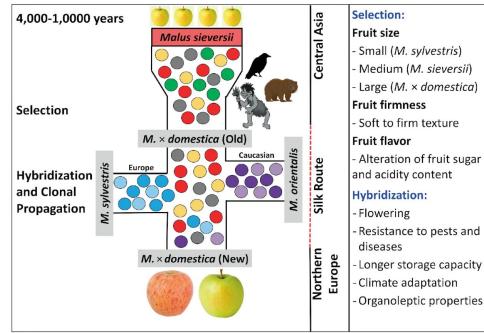


# How ‘Bout ‘Dem Apples?

Has Modern Agriculture Reduced The Genetic Diversity Of  
Apple Cultivars?



# *Malus Domestica* Originated In Central Asia And Spread To The Rest Of The World Through Vegetative Propagation Via The Silk ROad

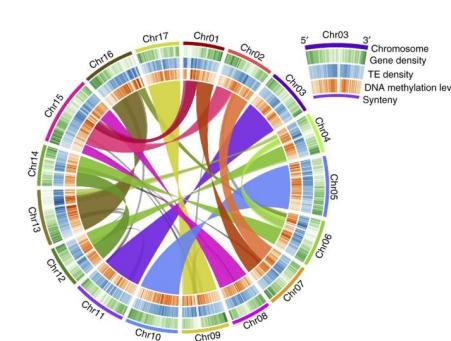
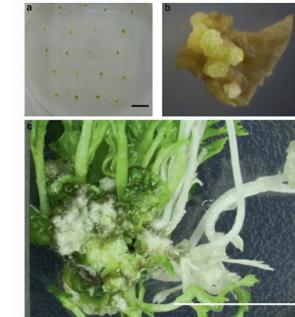
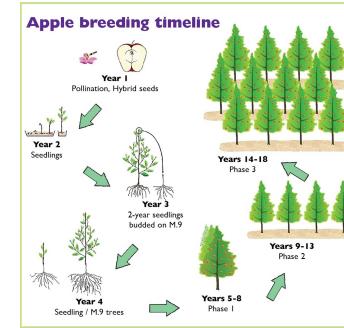
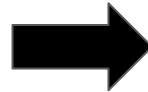


*M. Domestica* is typically:

- Diploid, 17 chromosomes
- Self-infertile
- Highly heterozygous



# Since ~1800, The Apple Industry Developed From Small Farms Selecting Chance Cultivars To Large Vertically Integrated Firms Using Increasingly Sophisticated Breeding Techniques



O'Rourke, D. (2021) Economic Importance of the World Apple Industry. In: Podwyszyńska, M., & Marasek-Ciołkowska, A. (eds.) (2021) The Apple Genome.

# Recent Research Has Shown That Today Many Cultivars Are Clonally Related To A Small Number Of Commercially Dominant Cultivars

nature > horticulture research > articles > article

Article | Open access | Published: 01 January 2021

## Genomic consequences of apple improvement

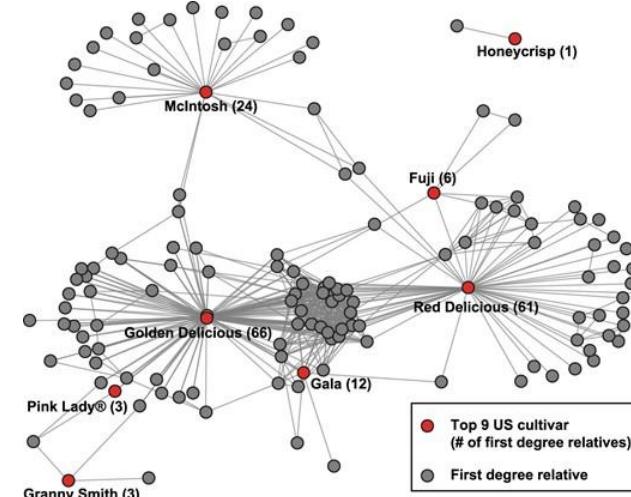
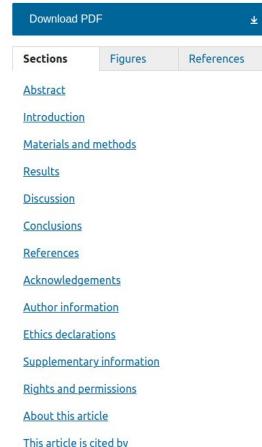
Zoë Migicovsky, Kyle M. Gardner, Christopher Richards, C. Thomas Chao, Heidi R. Schwaninger, Gennaro Fazio, Gan-Yuan Zhong & Sean Myles

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

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

### Abstract

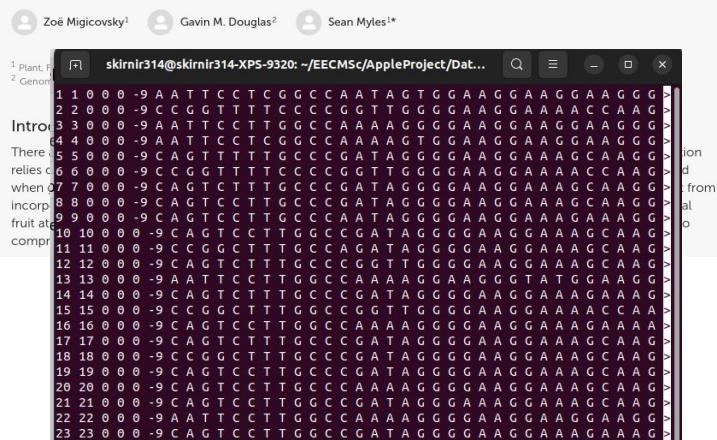
The apple (*Malus domestica*) is one of the world's most commercially important perennial crops and its improvement has been the focus of human effort for thousands of years. Here, we genetically characterise over 1000 apple accessions from the United States Department of Agriculture (USDA) germplasm 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.

# Combined Phenomic Description And SNP Data From Canada's Apple Biodiversity Collection

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



1,175 accessions (i.e., cultivars) across  
278,231 SNPs

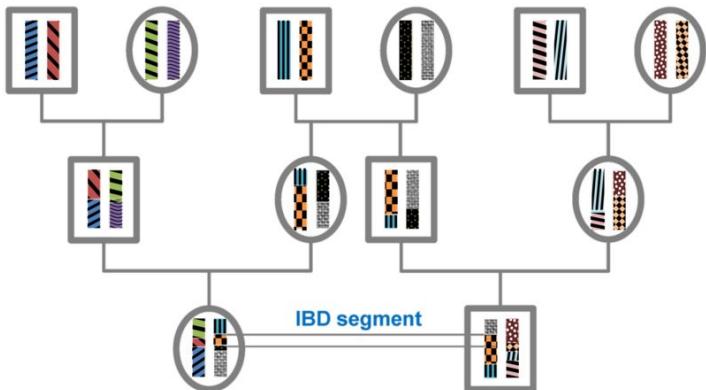
## Quantifying apple diversity: A phenomic characterization of Canada's Apple Biodiversity Collection

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: 22 June 2021 | <https://doi.org/10.1002/opp3.10214> | Citations: 17

# A tibble: 446 × 7	apple_id	Cultivar	Year	Country	Use	type	Period
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.

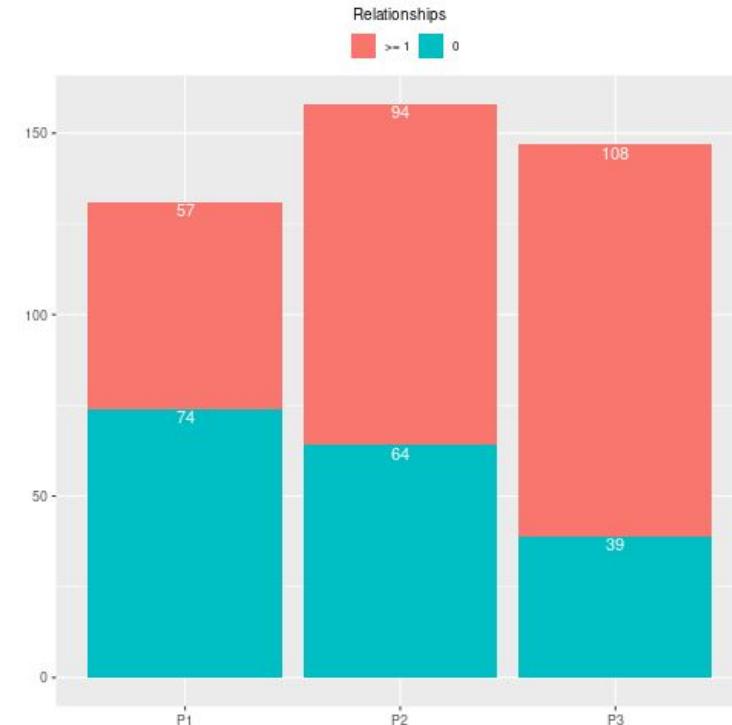
# Split Data Into Three Periods, Conducted Quality Control and Analysis In Plink and R

- Accessions split into three periods corresponding with different stages in apple industry development:
  - P1 (1800-1899): Early breeding techniques
  - P2 (1900-1959): Increased sophisticated
  - P3 (1960-Present): Vertical integration and modern genomics
- Filtering/cleaning:
  - SNPs filtered based on standard protocols (e.g., remove SNPs missing from many cultivars)
  - Include only *Malus domestica* with a known release year.
  - Remove non-diploid cultivars.
  - 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.

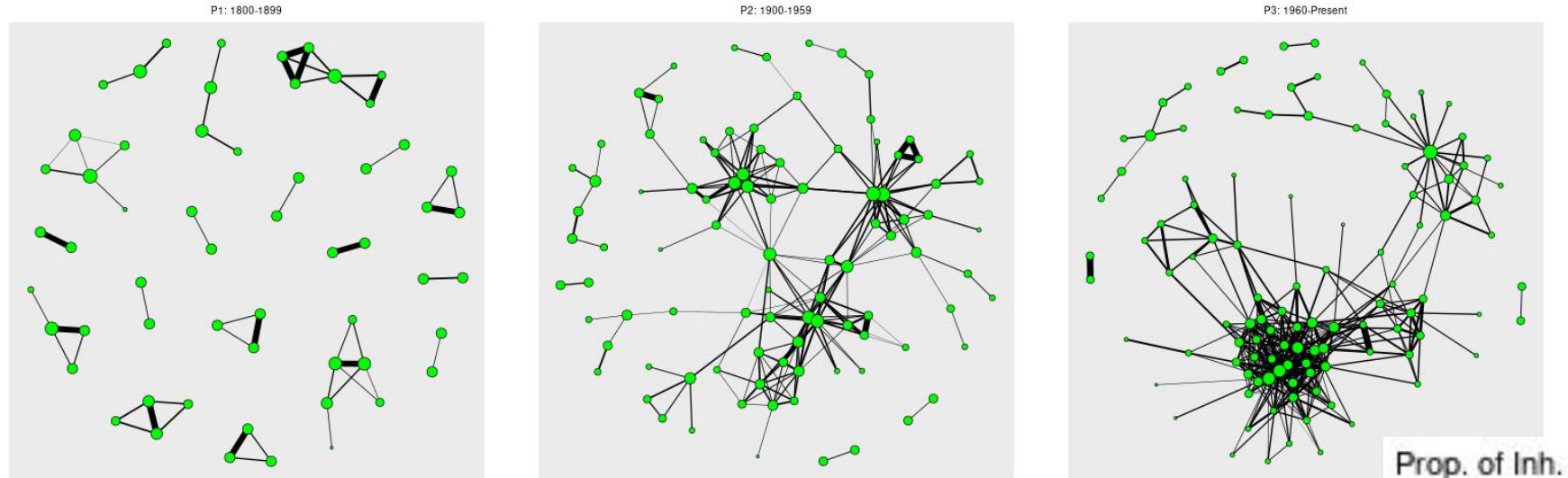
```
skirnir314@skirnir314-XPS-9320:~/EECMSc/AppleProject$ /usr/local/bin/Plink/plink  
--file ./Data/Snps/abc_combined_maf001_sort_vineland_imputed --keep ./Data/keep  
.p1.tsv --extract ./Data/Snps/plink.prune.in --allow-no-sex --geno .05 --mind .1  
--maf .01 --genome --out ./Data/Snps/p1_ibd  
PLINK 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 Relative Increased Over Time

- Used a  $\hat{\pi}$  threshold of  $\geq 0.125$  (cousins or closer) to determine 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 Also Showed Increased Relatedness Over Time

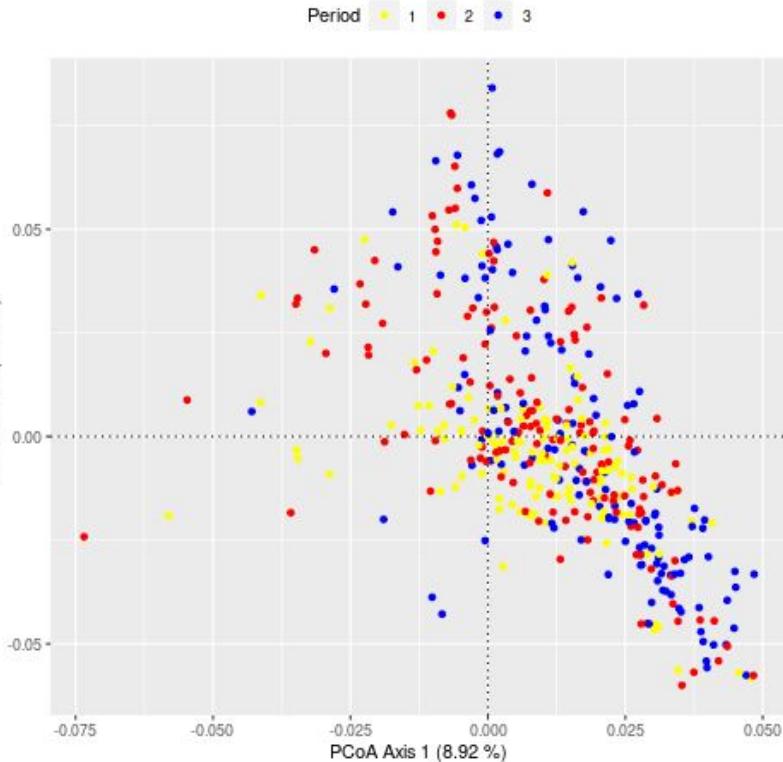


- Accessions with  $\hat{\pi} \geq 0.125$  were graphed using a Fruchterman-Reingold layout algorithm that placed accessions with higher  $\hat{\pi}$  values closer together.
- Average node degree increased from 1.9 in P1 to 4.7 in P2 and then to 7.9 in P3.
- Global efficiency (related to the inverse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

# But, a PCoA of Genetic Distance Showed No Change 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.

PCoA of apple cultivar SNPs



# Discussion

Caveats:

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

That said:

1. Time periods are somewhat informed by the literature
2. Uses a reasonably high number of accessions
3. Spot check confirms that Plink's IBD estimate is reasonable
4. Most cultivars are diploid

# Suggested Conclusions

- Newly developed cultivars probably are becoming increasingly related to each other.
- But, for now 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.

# References

- 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 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*. Springer Cham, pp. 1-18.
- Watts, S., et al. (2021) Quantifying apple diversity: A phenomic characterization of Canada's Apple Biodiversity Collection. *Plants People Planet*. 3 (6), 747-760.

# Appendix

## Yellow Bellflower



NJ, USA, mid 1700s

## Winesap (Maybe)



NJ, USA, first documented 1804

## Red Delicious



Iowa, USA, Jesse Hiatt, 1872

