**Title**

Genomic Characterization, Relatedness, and Analysis of Luther Burbank's Plum (Prunus sp) Introductions using Genotyping by Sequencing (GBS)

**Permalink**

https://escholarship.org/uc/item/5hw5t3ph

**Author**

Spaeth, Rachel

**Publication Date**

2022

Peer reviewed|Thesis/dissertation

Genomic Characterization, Relatedness, and Analysis of Luther Burbank's Plum
(*Prunus* sp) Introductions using Genotyping by Sequencing (GBS)

By

RACHEL SPAETH
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Horticulture & Agronomy

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____
Daniel Potter, Chair

_____
Thomas Gradziel

_____
Patrick J. Brown

_____
John Preece

Committee in Charge

2022

## Abstract:

Historic breeding populations often have limited pedigree notation and genomic data associated with them.  A prime example of this is the Plum Collection of over 150 cultivars introduced by Luther Burbank between the 1880's-1920's which consisted of many different *Prunus* species, the most prominent of which were combinations of the diploid species *P. salicina*, *P. simonii*, *P. americana*, and the hexaploid species *P. domestica*.  The primary goals of this study were to locate as many of Burbank's introductions as possible and to investigate their pedigrees and genetic diversity.  Fifty-three cultivars were located at three California sites; the USDA-ARS-National Clonal Germplasm Repository at the Wolfskill Experimental Orchard , Luther Burbank Home & Gardens, Gold Ridge Experiment Farm, and at numerous scion exchanges held by the California Rare Fruit Growers.  These taxa were analyzed using genotyping by sequencing (GBS) which retrieved over 24,000 SNPs spread over eight chromosomes with degrees of heterozygosity that ranged from 5.33% - 30.90%, with an average of 13.28%.  The transition/transversion rates of SNPs retrieved were 60.99% and 33.73% respectively  The remaining SNPs are indels.

SNP data were visualized using Identity by Descent (IBD) heatmaps, dendrograms, kinship matrices, and phylogenetic admixture plots. Eight distinct kinship groups emerged as significant using approximately unbiased bootstrapping on a dendrogram. Phylogenetic admixture analyses optimized at K=4 corresponded to *P. salicina*, *P. simonii*, European plums (*P. domestica* and *P. cerasifera*), and North American plums (*P. americana* and *P. rivularis*). The heatmap visualized shared rare alleles. Combined with phylogenetic admixture, the IBD analyses show how the taxa are related to each other.

Lastly, fruits scored for exocarp color, mesocarp color, free or cling-stone endocarps, and general shape were compared to genotypic data using genome-wide association studies (GWAS) and integrated haplotype scoring (iHS). Using Bonferroni-corrected and false discovery rate (FDR) p-values as a threshold of significance at alpha=0.05, GWAS showed exocarp color significantly correlated to a SNP on chromosome 2. Mesocarp color showed SNPs on chromosomes 2 and 3 that surpassed significance thresholds for association at FDR alpha=0.15. Similarly, free and cling-stone endocarps have a prominent, non-significant SNP on chromosome 7, and shape appears to be a multi-genic trait with SNP areas of interest on chromosomes 1, 3, 4,and 6. These results await confirmation in a larger sample. The iHS showed areas of improvement via long haplotypes on chromosomes 2, 4, and 7. Traits being selected against, as indicated by short haplotypes, were detected on chromosomes 1, 2, 3, 5, and 6. Taken together, these analyses provide new insights into the identities among Burbank's *Prunus* introductions and the potential of these valuable genetic resources for further development in the horticultural industry.

# Table of Contents

# Table of Contents (cont)

## Table of Contents (cont)

One of the most prolific contributors to the advancement of plum germplasm, Luther Burbank (1849-1926), utilized at least 12 different species of *Prunus* in his experiments, ultimately introducing over 150 new plum cultivars out of the over 800 cultivars of plants he introduced during his 50-year plant breeding career (Hedrick 1911, Howard 1945, Karp 2015).  Some of these introductions were directly acquired from other sources and then marketed by Burbank, rather than being bred by him, as was a common practice in the era before plant patents, while others were the result of hard work, patience, and artistic vision.  It could be said that he completely altered the human perception for what it meant to be a plum, strictly through conventional breeding methods, and his work caused the fresh-eating plum market to explode.  This was largely accomplished because of his keen observation skills, desire to import germplasm from throughout the world, and drive towards creating "better fruits and fairer flowers" for all to enjoy (Burbank 1902).  As a horticultural hobbyist and enthusiast with a high school-level of education, he was more willing to attempt crosses between distantly related species of plants than his classically trained, scientific contemporaries.  For Burbank, artificial selection was a sheer numbers game; plant a million seedlings, select two for cultivation, and burn the rest.

A significant downside to his often-haphazard breeding methods was his atrocious record-keeping.  It was so terrible that he once lost a Carnegie Grant for $10,000 because he would not write down the parentage for his experimental crosses (Dreyer 1993).  To his credit though, George H. Shull, a famous geneticist and botanist at the Carnegie Institute, said that if Luther had kept adequate records, he would not

have been nearly as prolific at generating new crosses (Dreyer 1993). His listed

parents for new introductions were based solely on his memory, and with thousands of

experiments happening simultaneously it stands to reason that his recollection would err

on occasion. Now, nearly 100 years after his death, genomic tools are available to aid

in the verification or refutation of his claims.

Burbank's foundational breeding experiments have been used extensively on a

global scale as a valuable source of high-quality material, launching the careers of other

great breeders such as Floyd Zaiger, creator of the Pluot (plum x apricot hybrid

backcrossed to plum), Aprium (apricot x plum hybrid backcrossed to apricot),

Nectaplum (nectarine x plum hybrid backcrossed to peach), and Pluerry (plum x cherry

hybrid backcrossed to plum). Many of Burbank's cultivars remain in existence today

because of conservation by hobbyists and enthusiasts from organizations like the

California Rare Fruit Growers (CRFG) and the North American Fruit Explorers (NAFEX).

Some are housed in formal collections in California at the Luther Burbank Home &

Gardens in Santa Rosa, California, the Luther Burbank Experiment Farm in Sebastopol,

California, and the USDA-ARS-NCGR's Wolfskill Experiment Orchard in Winters,

California (Figure 1). The genetic diversity of the Burbank accessions at these

collections encompasses different ploidy levels (2n, 6n) and species (*P. domestica, P.*

*salicina, P. armeniaca, P. ilicifolia, P. simonii, P. cerasifera, P. americana,* and *P.*

*rivularis*).
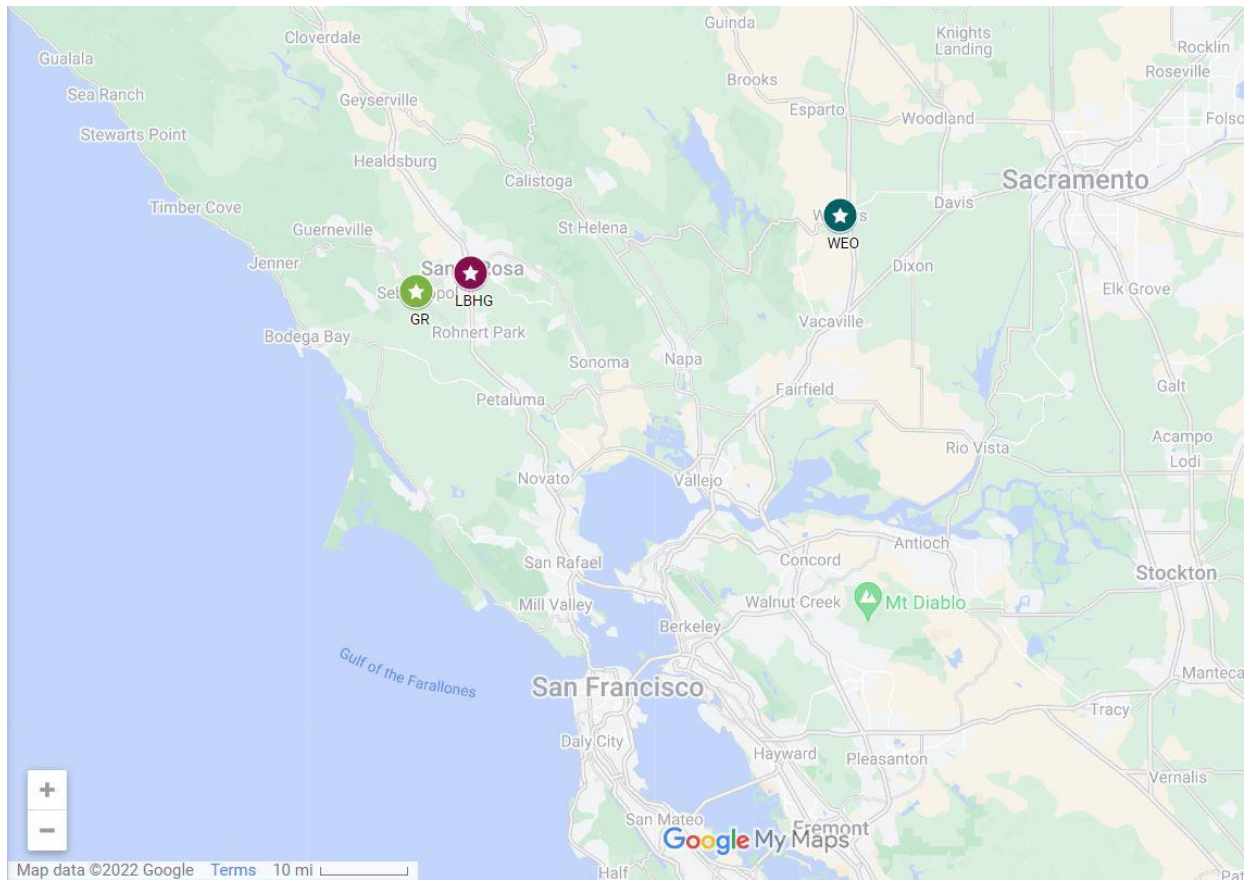
Fig 1. A map of northern California with the three study sites highlighted. The light green site is the Goldridge Experiment Farm in Sebastopol, CA (GR). The maroon site is the Luther Burbank Home & Gardens in Santa Rosa, CA (LBHG). The dark green site is the USDA-ARS-NCGR's Wolfskill Experiment Orchard in Winters, CA (WEO).

This dissertation covers the overall genetic composition of over fifty of Burbank's introductions in the context of four wild relatives using genotyping by sequencing (GBS), utilizing several marker statistics such as degree of heterozygosity, distribution of transitions and transversions, and chromosome-scale single nucleotide polymorphisms (SNPs) Then it uses phylogenetic techniques with principal component analysis (PCA), kinship analysis (K), and linkage disequilibrium (LD) to look at the relatedness among Burbank's taxa. Lastly, it compares single nucleotide polymorphisms (SNPs) to fruit marketability traits like exocarp and mesocarp color, overall shape, and free or cling stone endocarps using genome-wide association studies (GWAS) and integrated

haplotype scores (iHS).  Where applicable, the resulting genomic information will be

added to the Germplasm Resource Information Network (GRIN Global), RosBREED,

and NCBI databases to be accessible for future research.

# Chapter 1:  Genetic Characterization of Luther Burbank's Plum (*Prunus* sp) Introductions using Genotyping by Sequencing (GBS)

## Abstract:

Luther Burbank was a prolific plant breeder who revolutionized the human perception for what a "plum" could be by hybridizing several different species of *Prunus*. However, he was not a classically trained scientist and therefore had very poor record-keeping habits.  This study utilizes genotyping-by-sequencing (GBS) to analyze the diversity and genetic structure of 53 cultivars involved in Burbank's experiments including hybrids of *P. domestica, P. salicina, P. simonii, P. americana, P. rivularis,* and *P. cerasifera*.  GBS was successful at retrieving over 24,000 SNPs, and revealed a heterozygosity range from 5.33% - 30.90%, with an average of 13.28%, indicating some levels of inbreeding depression.  The transition/transversion rates of SNPs were 60.99% and 33.73% respectively.  STRUCTURE analysis revealed an optimal K=4, with clusters corresponding to North American species (*P. americana* and *P. rivularis*), European species (*P. cerasifera* and *P. domestica*), and unique clusters for two eastern Asian species (*P. simonii* and *P. salicina*).  A dendrogram was generated using Approximately Unbiased bootstrap resampling which revealed eight clades with AU support of over 95%.  These data will be useful for future *Prunus* analyses and provide a roadmap for surveying other breeding populations with limited pedigree information.

## Keywords:

## Introduction:

Luther Burbank was arguably one of the most prolific contributors to the advancement of *Prunus* germplasm, introducing over 150 cultivars of inter- and intra-specific hybrids nearly a century ago (Hedrick 1911, Howard 1945, Brooks and Olmo 1952, Karp 2015). He utilized conventional breeding methods to artificially select combinations of *P. domestica, P. salicina, P. armeniaca, P. ilicifolia, P. simonii, P. cerasifera, P. americana, P. rivularis* and others, pushing the boundaries of color, flavor, and shape (Figure 1). To date, some of his more important and long-lived introductions have been studied by researchers but have fallen short of capturing the breadth and depth of genetic variation in Burbank's *Prunus* breeding population (Minas et al., 2015, Sundouri et al., 2017, Marti et al., 2018, Salazar et al., 2018, Guerrero et al., 2021)



**Fig. 1**. A selection of plums collected at Burbank's Goldridge Experimental Farm in Sebastopol, California after he died showing the range of phenotypic diversity present in his breeding population. Image by J.B. Keil circa 1950.

Using genetic information in a breeding program saves time, space, and money because plants can be screened for favorable genotypes that correlate to desired phenotypes in the seedling stage.  In the past, sequencing genomes may have been cost-prohibitive, but today it is cheap, effective, and highly accessible, especially for perennial tree crops such as plums, cherries, peaches, and avocados which often take upwards of six years to develop fruits from seeds for analysis.  Instead of waiting six years to assess the fruit quality and marketability of hybrids grown from seed, one can observe fruit phenotypes within 2-3 years by grafting seedlings onto mature trees.  The resulting favorable cultivars can then be clonally propagated by grafting to preserve and perpetuate the genotype as has been the case for thousands of years for many perennial crops (Foster and Aranzana 2018).

Capturing accurate genomic data helps further speed up the process by allowing breeders to select genotypes that are correlated to phenotypes.  Previously, researchers utilized Simple Sequence Repeats (SSR's), also known as microsatellites, as markers to view the overall composition of the taxa in their study system.  This is helpful for identifying both beneficial and deleterious phenotypic traits that have been correlated with genotypes using Quantitative Trait Loci (QTL) but has some limitations.  For example, SSR markers do not have enough resolution to detect differences between bud sports (Fernandez I Marti et al. 2018) and are therefore less robust in looking at the differences in some perennial tree crops like citrus (Bernardi et al., 2014) and plums (Minas et al. 2015).  However, in some cases SSR's are still useful for comparing relationships in the *Prunus* subgenus if they are codominant and polymorphic (Guerrero et al. 2021).

Instead of using SSR markers for analysis, genotyping by sequencing (GBS) utilizes single nucleotide polymorphisms (SNPs) as the retrieval method for genetic information which helps resolve some of the ambiguity that arises from studies with closely related individuals.  Importantly, GBS has been used to effectively analyze genomic structure in populations that are both highly polyploid and heterozygous (Yang et al. 2017, Poland et al. 2012).  For example, *P. domestica*, an allohexaploid that segregates as a diploid and a tetraploid, is rather problematic from an analytical perspective if only microsatellite markers are used for identification or analysis (Zhebentyayeva et al., 2019, Gaši et al., 2020).  Since GBS takes a genome-wide approach to characterizing markers instead of focusing on a subset of the code, it is more useful for identification purposes (Pootakham et al., 2015, Salazar et al., 2018).  GBS is also better for tracing ancestry, especially in breeding populations with limited or inaccurate pedigree data (Velazco et al. 2019).

The thorough sampling method of GBS is applicable to agricultural crop improvement, historical specimen cataloging, taxonomy, archaeology, and phylogenetics studies.  The whole-genome approach is also a great way to establish a baseline in a breeding population whenever it is passed on from retiring researchers to the next generation or to reconstruct fragmented pedigree information. This experiment demonstrates the value of preserved collections as a rich source of genetic information and contributes to the relevance of maintaining these collections. These data are also important for marker-assisted selection (MAS) and transgenic modification (Aranzana et al. 2019, Guajardo et al. 2015).  Currently the reference genome information for the *Prunus* genus is highly limited with only seven of the 250-400 species: apricot (*P.*

*armeniaca*); sweet cherry (*P. avium*), almond (*P. dulcis*), Chinese plum (*P. mume*), peach (*P. persica*), Japanese plum (*P. salicina*), and Yoshino cherry (*P. yedoensis*) (Hodel et al. 2021).  Of these species, *P. persica* is the most useful for alignment of sequence data for inter-specific hybrids of *P. salicina* often found in breeding populations that require more scrutiny (Salazar et al. 2018).

Burbank cultivars used in current genetic characterization studies included 'Beauty,' 'Burbank,' 'Elephant Heart,' 'Satsuma,' and 'Wickson.' Using ten microsatellite markers, researchers were able to distinguish a genetic profile for these cultivars among 38 unique genotypes but could not distinguish between budsports of 'Santa Rosa.' This result did, however, indicate that the 'Santa Rosa' series are indeed budsports of each other, and not the product of a sexual cross (Minas et al. 2015).  When whole-genome sequencing was utilized on a similar subset of Burbank's breeding population, it was found that gene duplication events were responsible for either inducing or inhibiting the fruit's response to ethylene, a critical trait for ripening and post-harvest handling (Marti et al. 2018).

This study examines the overall genetic composition of an historic breeding population of plums introduced by Luther Burbank over a century ago, utilizing GBS to retrieve SNPs, examine heterozygosity, transition/transversion rates, phylogenetic admixture, and relatedness among individuals.  It covers 53 taxa Burbank either introduced or used in his breeding experiments (Appendix 1), the most comprehensive glimpse at his *Prunus* work to date.  The data collected for the taxa included in this project will be available for public access with the intention of allowing others to

experiment with inter & intraspecific plums, furthering Burbank's life goal of "creating better fruits and fairer flowers" for all to enjoy (Burbank 1902).

## **Materials and Methods:**

### Location of taxa and phenotype data:

A comprehensive list of Burbank's plum introductions is readily available (Howard, 1945).  However, by going back into the primary source material used by Howard to generate this list, the names of six more plums that were introduced by Luther emerged (Hedrick, 1904) (Appendix 1).  Also missing from this list are the taxa that were patented after Burbank's death by his widow Elizabeth in collaboration with the Stark Brothers.  This target list was used to hunt for specific cultivars through Agricultural Research Service (ARS) repository access, word of mouth, and local California Rare Fruit Grower (CRFG) scion exchanges.  Once a cultivar was located, historic literature published either during his lifetime or shortly thereafter was searched for claims Luther Burbank made about their parentage as well as any historical images that may accompany them (Burbank 1914, Hedrick 1911, Howard 1945, Brooks and Olmo 1952).

Scions of target material were multi-grafted onto mature trees with *P. cerasifera* 'Myrobalan 29C' as a universal rootstock at the Luther Burbank Home & Gardens (LBH&G) in Santa Rosa, California.  Historic maps were referenced for cultivar names at Luther Burbank's Goldridge Farm (GR) in Sebastopol California.  The GRIN Global Database and layout of the USDA-ARS-National Clonal Germplasm Repository at the Wolfskill Experimental Orchard (WEO) plum block were used for locating Burbank cultivars and their wild relatives (Appendix 2).

Genomic characterization:

Young leaf tissue was collected from mature trees in the early spring from all three study sites.  Leaves were stored in silica gel at room temperature until sufficiently dry, then frozen at -80ºC.  The DNA extraction protocol followed the DNeasy Plant Kit from Qiagin.  Retrieved sequences were aligned to *P. salicina* 'Sanyueli' (Liu et al. 2020).  Samples with more than 90% missing data were discarded, and this resulted in discarding 28/96 accessions. Duplicate samples were merged, leaving a total of 53 taxa.  Approximately 50,000 SNPs were retrieved.  Genotypes with a depth of <5 were set to missing, and then SNPs with more than 50% missing were discarded using TASSEL5.  Missing genotypes were imputed using Beagle 5.4 (Browning et al. 2018). TASSEL5 was used to look at genotype summary data, the number of SNP sites per chromosome, calculate LD per chromosome, and create PCA plots.

Chromosome Map:

The position of single-nucleotide polymorphisms (SNPs) along each chromosome was visualized using the R packages *'ggplot2'* and *'ggthemes'* (Wickham 2016, Arnold 2021)*.*

Dendrogram:

An identity-by-state (IBS) matrix  (Endelman & Jannink 2012) was generated with the *'Kinship'* analysis function using default parameters in TASSEL 5. To hierarchically cluster the accessions and obtain statistical values for said clustering, the *pvclust()* function from the *R* package 'pvclust' (Suzuki & Shimodaira 2006) was run on the IBS

matrix generated by TASSEL. The arguments for *pvclust()* were set to *method.dist = "cor", method.hclust = "complete",* and *nboot = 100.*

STRUCTURE:

A population structure analysis was performed using the software STRUCTURE (Pritchard et al., 2009).   Because six species were represented in this breeding population, K was set to 6, and the analysis was run for a 10,000 generation burn-in and the default 100,000 post burn-in periods over 5 reps.  To see if K=6 was optimal, K=3, K=4, and K=5 were also tested with the same burn-in criteria.  The results were exported to R for further analysis using the R package *'starmie'* (Tonkin-Hill and Lee 2016). The Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Deviance Information Criterion (DIC; Gao, Bryc, & Bustamante 2011), and the log posterior probability of K (L(K); Evanno, Regnaut, & Goudet 2005) were estimated for the different values of K determined that K=4 was optimal for this population because it maximized L(K) and minimized AIC.  These results were visualized using the R packages *'ggplot2'* and *'ggthemes'* (Wickham 2016, Arnold 2021)

## Results:

A total of 24,147 SNP sites were identified across all eight linkage groups (Table 1).  The overall genome size was approximately 272.95 Mbp.

| Chromosome | Length of Chromosome (Mbp) | Number of SNPsites |
|---|---|---|
| Chromosome 1 | 54.43 | 5419 |
| Chromosome 2 | 37.62 | 2826 |
| Chromosome 3 | 31.60 | 2788 |
| Chromosome 4 | 32.25 | 3027 |
| Chromosome 5 | 23.20 | 1987 |
| Chromosome 6 | 36.25 | 3169 |
| Chromosome 7 | 29.86 | 2542 |
| Chromosome 8 | 27.74 | 2389 |
| Total for all 53 taxa | 272.95 | 24,147 |

**Table 1.** The length of each chromosome and the number of SNP sites with a minor allele frequency of <5% removed and imputed using Beagle for in a multi-parental, inbred population of *Prunus* species introduced or used by Luther Burbank in his breeding experiments nearly a century ago.

SNP data were mined to determine the number of transitions and transversions that occurred in this population (Table 2).  5.28% of the total SNPS contained missing data and were removed for subsequent analyses.  Transitions made up 60.99% of all SNPs.  The prevalence of A/C and C/T transitions were nearly equal within each chromosome and across the entire genome.  Transversions were found in the remaining 33.73% SNPs.  A/C and G/T transversions were both just over 9% while A/T and G/C transversions were both slightly over 7%.

| SNP Type | Transitions | | Transversions | | | | Insertions or Deletions | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A/G | C/T | A/T | A/C | G/T | G/C | A | C | G | T |
| Chromosome 1 | 1638 | 1610 | 430 | 571 | 510 | 414 | 60 | 60 | 72 | 54 |
| Chromosome 2 | 848 | 846 | 201 | 255 | 263 | 211 | 41 | 59 | 58 | 44 |
| Chromosome 3 | 863 | 867 | 201 | 282 | 255 | 203 | 23 | 28 | 42 | 24 |
| Chromosome 4 | 931 | 918 | 220 | 290 | 269 | 214 | 30 | 56 | 55 | 44 |
| Chromosome 5 | 609 | 607 | 142 | 188 | 181 | 157 | 23 | 28 | 30 | 22 |
| Chromosome 6 | 964 | 973 | 235 | 267 | 301 | 227 | 44 | 65 | 63 | 30 |
| Chromosome 7 | 783 | 779 | 200 | 209 | 253 | 205 | 28 | 30 | 30 | 25 |
| Chromosome 8 | 734 | 757 | 152 | 237 | 201 | 200 | 28 | 38 | 25 | 17 |
| Totals | 7370 | 7357 | 1781 | 2299 | 2233 | 1831 | 277 | 364 | 375 | 260 |
| Frequency | 30.52 % | 30.47 % | 7.38 % | 9.52 % | 9.25 % | 7.58 % | 1.15 % | 1.51 % | 1.55 % | 1.08 % |
| | 60.99% | | 33.73% | | | | 5.28% | | | |

**Table 2.** The number of transitions, transversions, and insertions or deletions found across 24,147 SNPs for each chromosome in a dataset containing 53 *Prunus* taxa aligned to *P. salicina* 'Sanyueli' (Liu et al., 2020).

A chromosome map was generated to show the distribution of SNP sites (Figure 2) with a cutoff of a minor allele frequency (MAF) of 0.05% missing data removed.
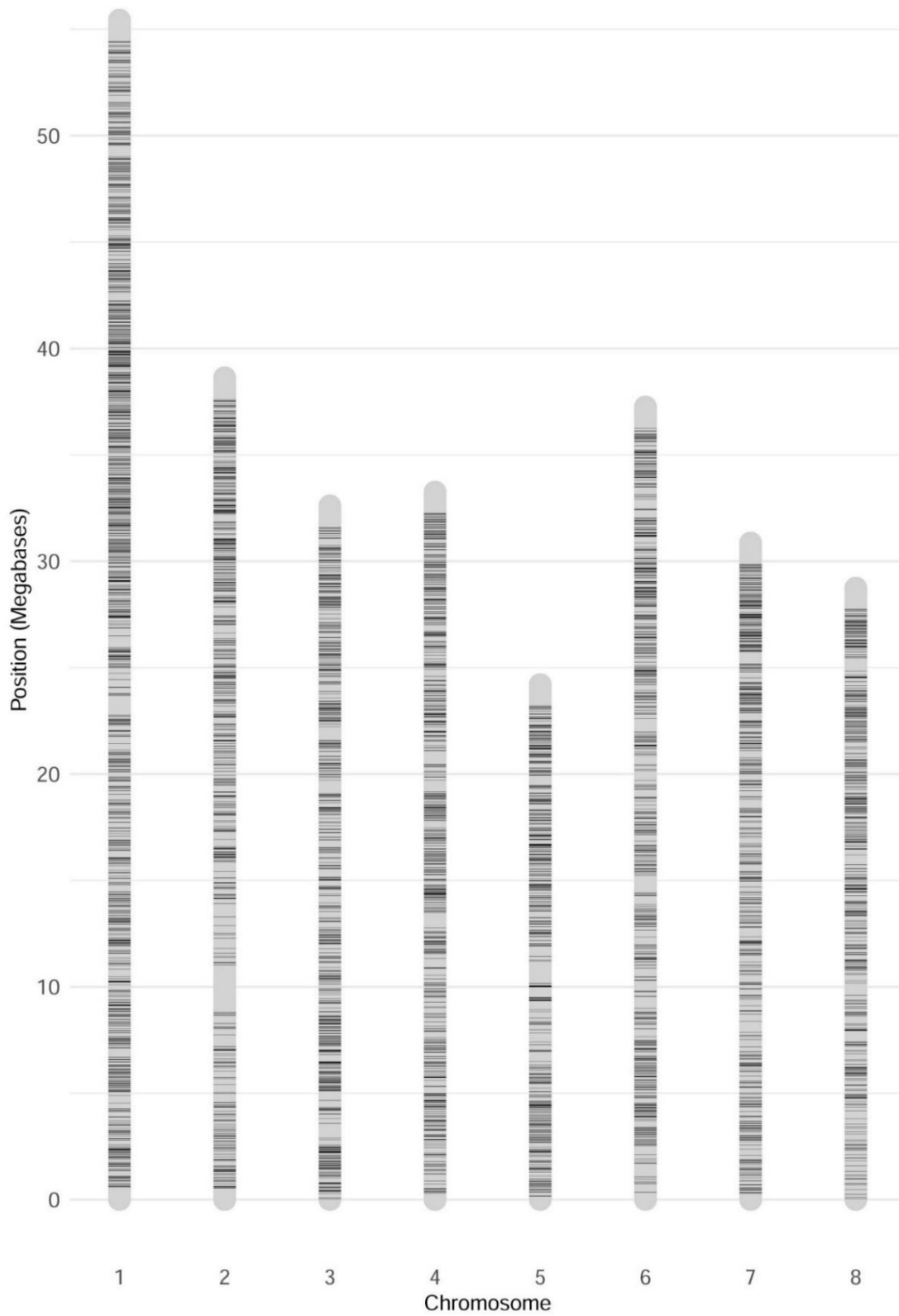
**Fig. 2.** A chromosome map illustrating the overall distribution of SNP sites (black bands) on all eight chromosomes for 53 Burbank *Prunus* taxa aligned to *P. salicina* 'Sanyueli' (Liu et al., 2020).

After removing taxa with a Minor Allele Frequency (MAF) of <0.5, the average MAF for the remaining taxa was 9.83%. The average proportion of heterozygosity found in this population was 13.28%, with a maximum of 30.90% and a minimum of 5.33% (Figure 3).  This is indicative of moderate to severe cases of inbreeding depression across the population (Sanchez-Perez et al. 2006) and is typical of breeding populations where there are limited founders and lots of back crossing events.  Eighteen of the taxa in this study had less than 10.0% heterozygosity, twenty-seven taxa had between 10.0 and 20.0% heterozygosity, and the remaining eight taxa had between 20.0 and 31.0% heterozygosity.  Some of the founders in Burbank's breeding population contained low levels of heterozygosity before he started experimenting with them.  These early introductions include 'Botanky - DPRU.372' (6.11%), 'Satsuma - DPRU.438' (11.05%), 'Simon – DPRU.545' (17.73%) which were released in 1888, 1886, and 1872 respectively.

The heterozygosity of some of his most famous and commercially viable cultivars include 'Wickson – DPRU.2135' (19.76%), 'Beauty' (LBHG.sw-9.03%, LBHG.c-12.23%, or DPRU.2120-19.22%), 'Santa Rosa' (12.22%), and 'Shiro – DPRU.2132' (20.53%). 'Wickson' was historically reported as the female parent of 'Shiro.'  The higher degree of heterozygosity in 'Shiro – DPRU.2132' could be attributed to its interspecific parentage, as Burbank reported it to also contain *P. munsoniana* (a North American native), and *P. cerasifera* (Hedrick 1911)*.*

The hexaploid *P. domestica* taxa ranged in heterozygosity from 11.99% to 23.45%.  'Top of the Hill,' which had the lowest degree of heterozygosity, was never formally introduced by Burbank, but is found on historical maps and continues to grow

at his Goldridge Experimental Farm in Sebastopol, California. 'Original Stoneless –

DPRU.2302' (18.99%), is an offspring of 'Sans Noyau – DPRU.2419' (23.45%) which

had been in cultivation since the late 1700's. These cultivars are unique because their

endocarp has been reduced to a single sliver that no longer encapsulates the seed in

the center of the fruit. The remaining piece, known as the funiculus, is responsible for

transporting nutrients to the developing embryo. Unfortunately, it needs to be lignified

to be fully functional. Burbank's breeding objective was to create a prune stuffed with

an almond, but alas the lignified funiculus prevented this endeavor from being

successful. 'French – DPRU.436' (19.70%) was introduced by S.D. Willard in 1889 and

then was used extensively by Burbank in his prune experiments. It is thought to be one

of the parents of 'Grand Prize – DPRU.1572,' a prune found on Burbank's property after

he died that was introduced by Stark Brothers in 1937. 'Grand Prize – DPRU.1572' has

a heterozygosity of 22.30%. These generally higher degrees of heterozygosity could be

attributed to these taxa simply having more sets of chromosomes.



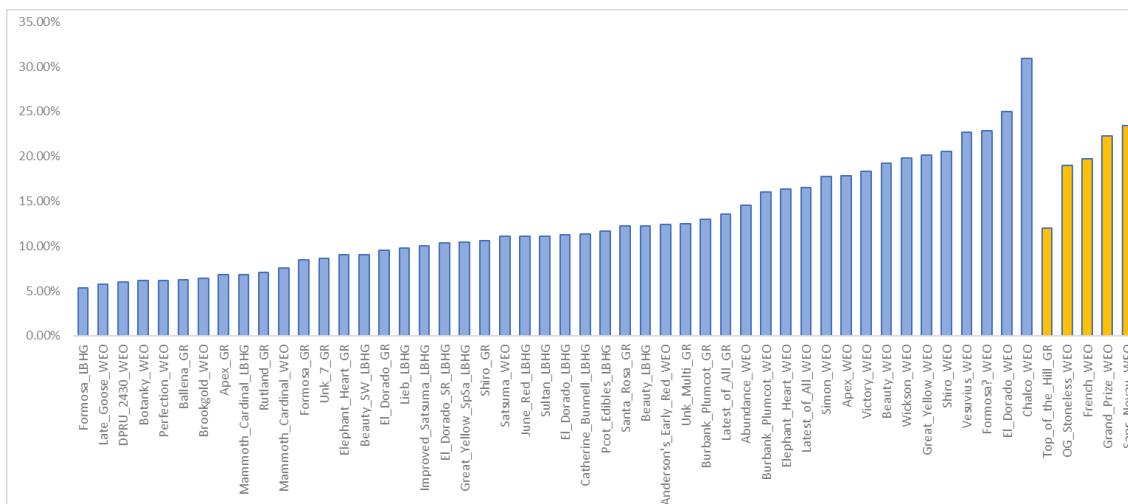**Fig. 3.** The proportion of heterozygosity found in 53 *Prunus* taxa used in Luther Burbank's breeding experiments. The blue bars are diploid (2n) and the gold bars are hexaploid (6n) cultivars. Cultivars specify the location where they are grown: WEO (Wolfskill Experiment Orchard, Winters, CA), LBHG (Luther Burbank Home & Gardens, Santa Rosa, CA), GR (Goldridge Burbank Experiment Farm, Sebastopol, CA).

Dendrogram

A dendrogram showing a hierarchical clustering of these taxa was generated using multiscale bootstrap resampling of p-values (Figure 4).  It is important to note that this dendrogram shows a relationship between how similar the alleles in their genomes are, not a strict measure of pedigree relatedness.  Eight major clades emerged with Approximately Unbiased (AU) values of larger than 95%.

Starting on the top of Figure 4, the first group, with an AU of 100% indicated that 'Latest of All' and the Unknown Multi-grafted individual (both from Goldridge Farm) are either incredibly closely related or are clones of the same individual.  The second group with an AU of 96% contains a monophyletic group of 'Victory – DPRU.791' (WEO) and 'Wickson – DPRU.2135' (WEO) with 'Apex – DPRU.1170' (WEO) and 'Great Yellow – DPRU.2105' (WEO) as sister taxa.  The third group is the largest, containing twelve taxa.  A pair of sister taxa in this clade are two accessions of 'Mammoth Cardinal' that were sourced from different locations (WEO – DPRU.2127 and LBHG), indicating they are likely the same accession.  Others in their clade are 'Apex' (GR), and 'Formosa' (LBHG) with 'Satsuma – DPRU.438' as their MRCA.  All these taxa have *P. salicina* indicated in their historically reported parentage.  'Apex' supposedly also contains some *P. armeniaca* (apricot) as well, but that cannot be confirmed from this data set.  The Goldridge and Wolfskill (DPRU.1170) samples of 'Apex' do not appear in the same cluster, so it is unlikely they are the same accession.  One would expect an interspecific hybrid to have a higher degree of heterozygosity.  'Apex – DPRU.1170' (WEO) has a heterozygosity of 17.86%, while 'Apex' GR has a heterozygosity of only 6.80%.  This

18

information shows support for the Wolfskill accession to be the correct version, though more data is required to be certain.

Also in group 3, one can find 'Botanky – DPRU.372' and '*P. simonii* - DPRU2430' as sister taxa with and AU of 100%. 'DPRU2430' is a wild accession of *P. simonii* from the Wolfskill collection that was included as a distant relative of individuals in this breeding population. This seems to suggest that 'Botanky – DPRU.372' is a *P. simonii* instead of the *P. salicina* it was reported to be. The admixture of 'Botanky – DPRU.372' shows it is primarily comprised of cluster 3, which is also evidence to suggest it is really *P. simonii*. Burbank received 'Botanky' in the collection of seeds that launched his *Prunus* experiments. It is entirely possible these seeds contained a mixture of both *P. salicina* and *P. simonii* and he assumed they were all *P. salicina* upon arrival. 'Abundance – DPRU.919' is listed as the MRCA for 'Botanky – DPRU.372' and '*P. simonii* - DPRU2430.' The reported parentage for 'Abundance' is *P. salicina* x *P. armeniaca*, but with the species of 'Botanky – DPRU.372' in question, it may be that 'Abundance – DPRU.919' is really a *P. simonii* hybrid instead. Sister to all of those individuals mentioned from group 3 is a clade that contains 'Ballena (GR),' 'Elephant Heart' (GR) 'El Dorado' (GR), and 'Improved Satsuma (LBHG).'

Interestingly, group 4 is comprised entirely of 'Elephant Heart – DPRU.2123' (WEO), and two accessions of 'El Dorado' (LBHG). This seems to indicate a relationship between 'Elephant Heart – DPRU.2123' and 'El Dorado' but more information is needed to tease out the specifics. 'Elephant Heart' was collected, patented, and released by Stark Brothers after Burbank's death. It has a clear 'Satsuma' influence with dark red flesh, and a thick, waxy coat. 'El Dorado' also has a

waxy coating, but is yellow on the inside instead, showing a *P. simonii* influence in its phenotype.

Group 5 is a sister group of 'Brookgold – DPRU.1736' and 'Perfection – DPRU.1720.'  'Brookgold – DPRU.1736' is a wild-type *P. salicina*.  'Perfection' was once a synonym for 'Wickson.'  This accession of 'Perfection – DPRU.1720' has a considerably different phenotype from 'Wickson' though, having much smaller, insipid fruits instead of the delicious, large fruits of 'Wickson' so it is highly unlikely they are the same accession.  Group 6 is a sister group of 'Great Yellow' (LBHG) and 'Shiro' (GR).  To the untrained eye, 'Great Yellow' may be mistaken for 'Shiro.'  Both are smallish, yellow fruits.  However, 'Shiro' has pointy-bottomed fruits, and 'Great Yellow' is fully round.  After seeing these fruits, 'Great Yellow (LBHG)' was mislabeled and is 'Shiro' instead.

Group 7 contains the taxa with the North American *Prunus* influence, 'Chalco – DPRU.431,' 'Late Goose – DPRU.546,' and 'Unknown 7' (GR).  Group 8 contains the most hexaploid taxa, 'Grand Prize – DPRU.1572,' 'French – DPRU.436,' and 'Top of the Hill (GR),' with 'Sans Noyau – DPRU.2419' as a distant MRCA with only a 58% AU.  Some anomalous taxa are found in this group as well such as 'Simon – DPRU.545,' 'El Dorado – DPRU.2122' (WEO), 'and 'Vesuvius – DPRU.2108' in one clade of the group and 'Burbank Plumcot' (GR) and 'Plumcot Edibles' (LBHG) in the other clade.  Phenotypically, the 'Burbank Plumcot' is a fruit that has a yellow mesocarp and yellow with a red blush exocarp.  The 'Plumcot Edibles' (LBHG) has a blue exocarp and a green mesocarp.  More information is needed to tease out the individual taxa in this group.
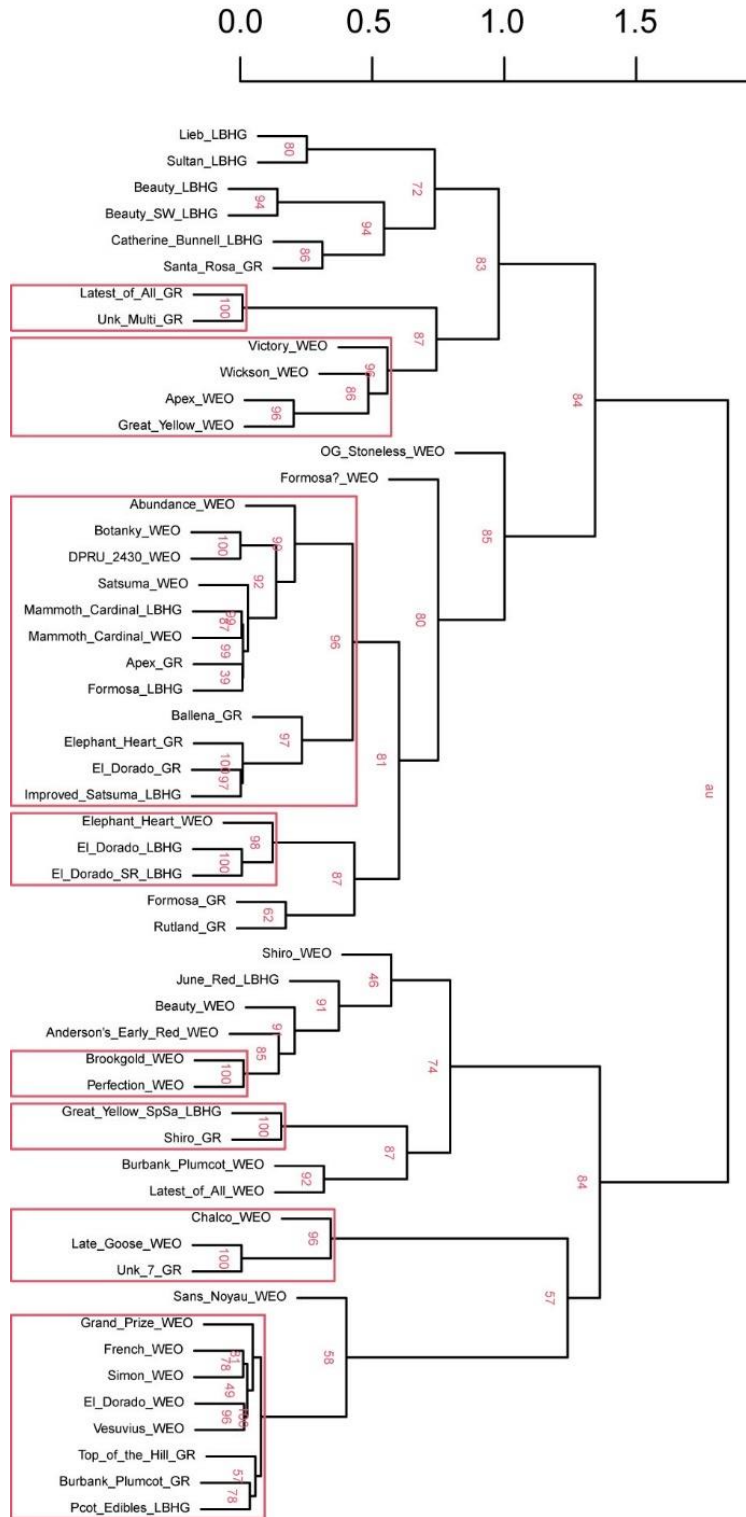
**Fig 4.** A dendrogram representing the genetic similarity between 53 *Prunus* taxa introduced by Luther Burbank. The branch length represents genetic distance calculated with Identity by Decent in TASSEL 5. The cluster method used is complete. Red values at the base of each node are Approximately Unbiased (AU), generated by multiscale bootstrap resampling p-values. Clusters with AU larger than 95% are highlighted by red rectangles.

Identity by Decent (IBD) values were calculated in Tassel 5. Initially, it was thought that an admixture of K= 6 would be optimal for this breeding population because six different species were included in the study. However, K=4 was considered optimal due to L(K) (logarithmic probability of K; Evanno et al. 2005) being maximized and Deviance information criterion (DIC; Gao et al. 2011) being minimized (Figure 5). These two diagnostic parameters concur that K=4 is the appropriate threshold for this breeding population.
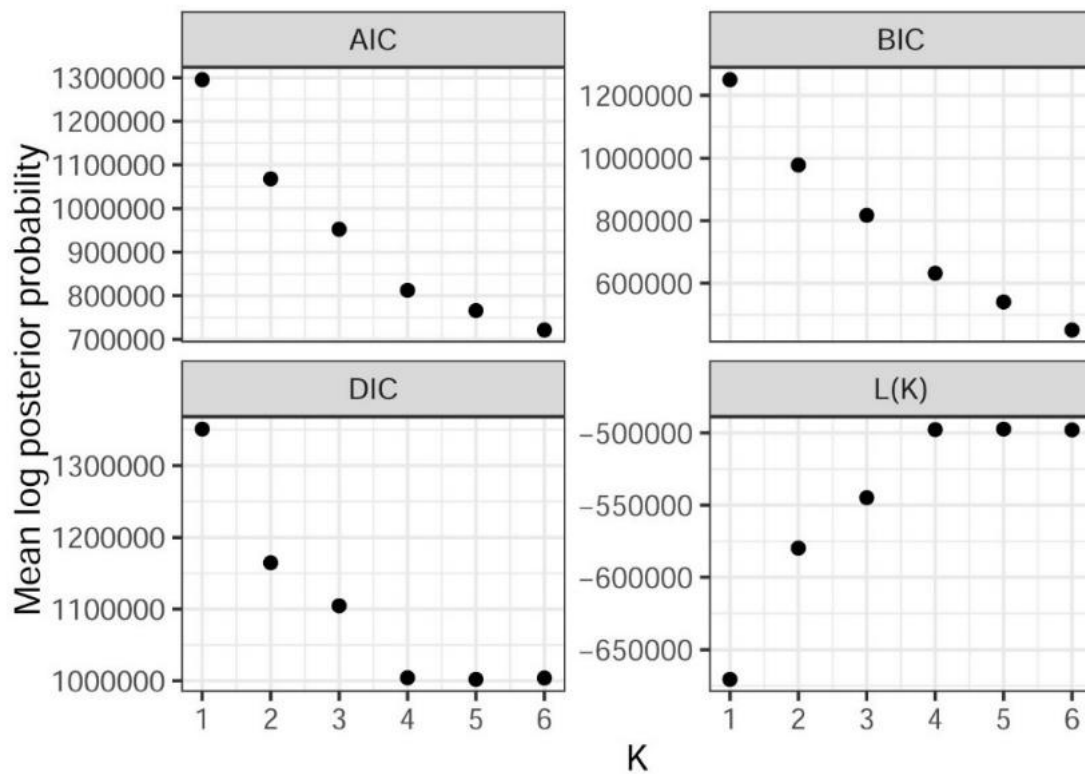


**Fig 5.** Diagnostic likelihoods using the mean log posterior probability to determine the optimal K value for STRUCTURE analysis showed a K=4 as being the optimal number of clusters in this population.

A STRUCTURE plot was generated using the optimized K=4 to observe the admixture of genotypes in this breeding population (Figure 6).
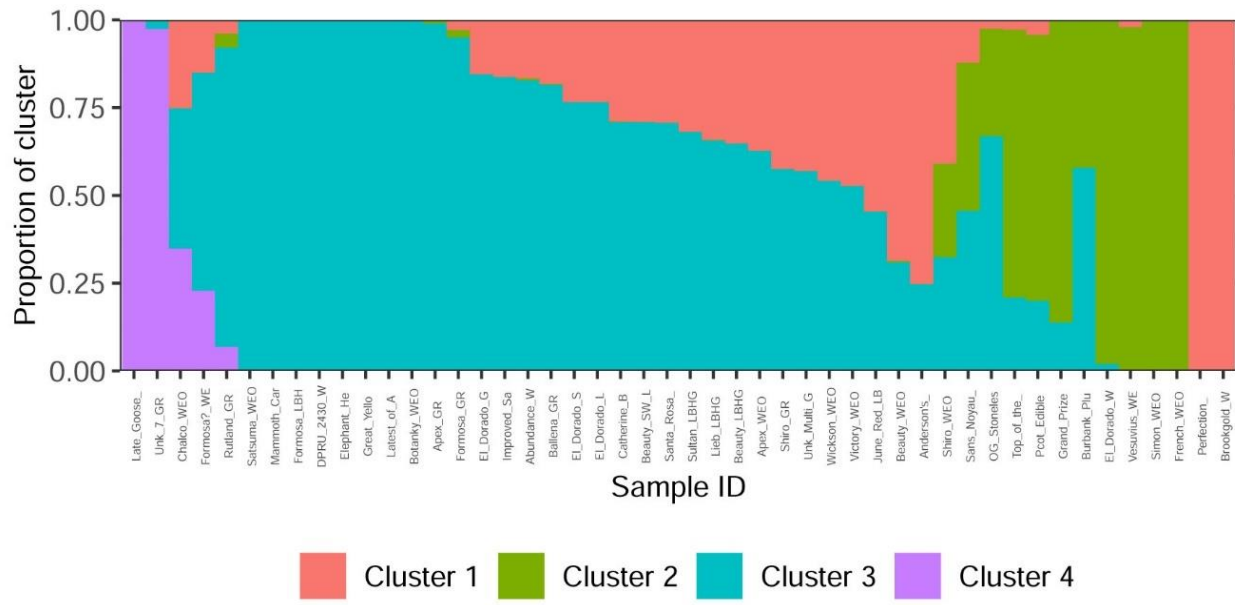
**Fig 6.** STRUCTURE analysis of admixture with a K=4 for 53 taxa of Luther Burbank's *Prunus* cultivars. Clusters 1 and 3 align with *P. salicina* and *P. simonii* respectively from Eastern Asia. Cluster 2 corresponds to European *Prunus* species. Cluster 4 represents North American *Prunus.*

The STRUCTURE clusters do not strictly adhere to species but do indicate some general trends with them. The bulk of this population is made up of Clusters 1 and 3. These two clusters most likely represent *P. salicina* and *P. simonii* respectively, both of which evolved in Eastern Asia. Cluster 2 is dominated by the hexaploid *P. domestica* taxa, but also is found in quite a few of the diploid individuals in the Burbank breeding population. *P. domestica* is reported to be an allopolyploid made up of a diploid *P. cerasifera* and a tetraploid *P. spinosa* (Gaši et al. 2020) so it appears that individuals with Cluster 2 share alleles with both *P. domestica* and *P. cerasifera*. Both of these species evolved in Europe. Cluster 4 corresponds to North American cherry-type plums with smaller, round fruit types, encompassing two species surveyed in this study: *P. americana* and *P. rivularis*. Interestingly, an historical photo of some of Burbank's parent stock corresponds to these four clusters (Figure 7)

23

**Fig 7**. A black and white image of four of the parents used in Burbank's breeding experiments that corresponds to the phylogenetic admixture analysis conducted in STRUCTURE with K=4. (Shaw 1910).

## Discussion:

Overall, these results indicate Luther Burbank frequently under- or over-estimated parental contributions in his *Prunus* introductions, going all the way back to founders in his breeding population. Between not taking adequate notes and failing to exclude pollinating insects from flowers he attempted to hybridize, his assumptions were at best based on color, shape, and flavor, and often did not accurately reflect true pedigree. That being said, his contributions to the advancement of interspecific *Prunus* breeding are still valid and valuable to modern breeders today now that the underlying genomic baseline in his *Prunus* population has been established.

When Luther was alive, scientists were on the forefront of establishing breeding principals related to the field of genetics without fully grasping the genetic basis of heritability. Burbank found inspiration in Darwin's "The Variation of Animals and Plants Under Domestication" (1868). Mendel's foundational work with pea breeding (Mendel 1866) was first published when Burbank was a teenager but wasn't translated to English and accessible to Burbank until 1901, over twenty years into Luther's plant breeding career. When learning of Mendel's work, Burbank was underwhelmed, possibly because his own experiences with plant breeding did not follow Mendelian ratios of genetic inheritance due to his usage of wide-crosses, disruptive artificial selection, and stubborn refusal to count phenotypes (Dreyer 1985). This is still perplexing because he should have observed basic segregation patterns such as the 3:1 dominant: recessive relationship achieved through self-fertilization or the 1:1 dominant: recessive relationships observed through backcrossing throughout his prolific career. He must have been intuitively aware of these patterns despite his lack of physically counting and documenting phenotype numbers.

He also (intentionally or unintentionally) frequently engaged in backcrossing, making strict parent-offspring relationships challenging to deduce. While dichotomous branching is observed in the dendrogram of Burbank's *Prunus* accessions, this pattern does not fully reflect a linear pedigree. *P. salicina* and *P. simonii* are undisputed founders of Burbank's plum breeding population and are largely responsible for his vast number of successes. *P. salicina* typically provided the shape and color of the fruit; *P. simonii* provided the free-stone endocarps and a drastic increase in flavonoids (Gomez and Ledbetter 1994). 'Simon' is not a direct Burbank introduction but was used

25

extensively in his hybridizations.  Both 'Botanky' and 'Satsuma' were directly introduced

by Burbank.   He received them as seeds, grew them, evaluated their fruits, and

artificially selected the most desirable of the lot before releasing them through his

catalogues.  In subsequent years, *Prunus* cultivars he introduced were accessions he

had bred with artistic vision and intention.

Creating successful *P. domestica* (prune) hybridizations were the most

challenging for Burbank.  He once said, "If I have engaged in a forty year long quest of a

perfect prune, without quite attaining the ideal, it is chiefly because this fruit shows such

a propensity to forget what it has learned and to revert to the standards of the ordinary

plum" (Burbank et al. 2014).  Burbank knew that his prune breeding populations were

not segregating as expected but did not comprehend why.  It was not until many years

later scientists discovered polyploidy (Stebbins 1947), which likely was the biggest

confounding factor in utilizing the hexaploid *P. domestica* accessions for breeding

experiments.

This research is a useful first step to surveying the genetic diversity found in an

historic breeding population.  The genetic diversity in the Burbank breeding population

sampled encompasses species from North America, Europe, and Eastern Asia.  They

are mostly homozygous individuals with inbreeding depression strongly apparent in a

third of the individuals sampled.  More information is needed to tease out some of the

anomalous relationships found in the dendrogram.  Adding accessions of apricot is

suggested to see if this will change the optimal K value for STRUCTURE by including

broader genetic diversity espoused by Burbank.

**Conclusion**:

Luther Burbank's foundational plant breeding work holds relevance today, nearly a century after his death because modern genomic tools allow us to peer into past. While there may be too many missing links in his breeding population encompassing 40 years of trial and error to compile definitive parent-offspring relationships, we can adequately surmise the interrelatedness of this population using IBD and phylogenetic admixture using SNPs retrieved with GBS. These once-popular cultivars of *Prunus* hold rich genetic resources that should be utilized in modern plant breeding efforts to combat current challenges like increased yield, shelf life, or disease resistance, and decreased sensitivity to chill hour requirements.

**References:**

Aranzana, M.J., V. Decroocq, E. Dirlewanger, I. Eduardo, Z.S. Gao, K. Gasic, A. Lezzoni, S. Jung, C. Peace, H. Prieto, R. Tao, I. Verde, A.G. Abbott, and P. Arus. (2019). *Prunus* genetics and applications after de novo genome sequencing: achievements and prospects. Horticultural Research. 6(58). https://doi.org/10.1038/s41438-019-0140-8 .

Arnold, J.B. (2021). ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. https://github.com/jrnold/ggthemes .

Bernardi, J., A. Marocco, P. Caruso, and C. Licciardello. 2014. Molecular characterization of citrus cultivars: Insight from recent studies. Citrus: Molecular Phylogeny, Antioxidant Properties, and Medicinal Uses. K. Hayat ed. Pgs 13-30

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Brooks, R.M. and H.P. Olmo. 1952. Register of New Fruit and Nut Varieties 1920-1950. University of California Press. http://reader.library.cornell.edu/docviewer/digital?id=chla3174378#mode/1up

Browning, B.L., Y. Zhou, and S.R. Browning. 2018. A one-penny imputed genome from next generation reference panels. American Journal of Human Genetics. 103(3): 338-348. https://doi.org/10.1016/j.ajhg.2018.07.015

Burbank, L.  1902.  Some of the fundamental principals of plant breeding.  Second International Conference on Plant Breeding and Hybridization.  New York.

Burbank L., J. Whitson, R. John, and H. Smith Williams (eds).  (1914).  Luther Burbank: his methods and discoveries and their practical application.  New York; London: Luther Burbank Press.

Ciofi, C., M.C. Milinkovitch, J.P. Gibbs, A. Caccone, and J.R. Powell.  2002.  Microsatellite analysis of genetic divergence among populations of giant Galapagos tortoises.  Molecular Ecology.  11(11): 2265-2283.  https://doi.org/10.1046/j.1365-294X.2002.01617.x .

Darwin, C. 1868.  The Variation of Animals and Plants under Domestication.  London: John Murray.

Dreyer, P. 1985.  A Gardener Touched with Genius: The Life of Luther Burbank.  University of California Press.

Endelman, J.B. and J-L. Jannink.  2012.  Shrinkage estimation of the realized relationship matrix.  G3(Bethesda) 11:1405-13.  https://doi.org/10.1534/g3.112.004259 .

Evanno, G., S. Regnaut, and J. Goudet.  2005.  Detecting the number of clusters of individuals using software STRUCTURE:  a simulation study.  Molecular Ecology.  https://doi.org?10.1111/j.1365-294x.2005.02553.x .

Foster, T. M. and M. J. Aranzana.  2018.  Attention sports fans!  The far-reaching contributions of bud sport mutants to horticulture and plant biology.  Horticulture Research 5(44).  https://doi.org/10.1038/s41438-018-0062-x .

Gao, H., K. Bryc, and C.D. Bustamante.  2011.  On identifying the optimal number of population clusters via the deviance information criterion.  PloS One. 6(6): e21014.  https://doi.org/10.1371/journal.pone.0021014 .

Gaši, F., Sehic, J., Grahic, J. et al. 2020.  Genetic assessment of the pomological classification of plum *Prunus domestica* L. accessions sampled across Europe. Genetic Resources and Crop Evolution. 67: 1137–1161. https://doi.org/10.1007/s10722-020-00901-y

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. (2014) TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS ONE 9(2): e90346

Gomez, E. and C.A. Ledbetter.  1994.  Comparative study of the aromatic profiles of two different plum species: *Prunus salicina* Lindl and *Prunus simonii* L.  Journal of the Science of Food and Agriculture.  65(1): 111-115.  https://doi.org/10.1002/jsfa.2740650116 .

Guajardo, V., S. Solis, B. Sagredo, F. Gainza, M. Munoz, K. Gasic, and P. Hinrichensen.  2014.  Construction of high density sweet cherry (*Prunus avium* L.) linkage maps using microsatellite markers and SNPs detected by genotyping by sequencing (GBS>).  PLoS One.  https://doi.org/10.1371/journal.pone.0127750 .

Guerrero, B.I., M.E. Guerra, S. Herrera, P. Irisarri, A. Pina, and J. Rodrigo.  2021.  Genetic diversity and population structure of Japanese plum-type (hybrids of *P. salicina*) accessions assessed by SSR markers.  Agronomy.  11: 1748.  http://doi.org/10.3390/agronomy11091748 .

Hampton, J.O., P.B.S. Spencer, D.L. Alpers, L.E. Twigg, A.P. Woolnough, J. Doust, T. Higgs, and J. Pluske. 2004. Molecular techniques, wildlife management and the importance of genetic population structure and dispersal: a case study with feral pigs. Journal of Applied Ecology. 41(4): 735-743. https://doi.org/10.1111/j.0021-8901.2004.00936.x .

Hedrick, U.P. 1911. Plums of New York. http://chla.library.cornell.edu/cgi/t/text/text-idx?c=chla;cc=chla;idno=5519832

Hodel, RGJ., E. Zimmer, & J.Wen. 2021. A phylogenomic approach resolves the backbone of Prunus (Rosaceae) and identifies signals of hybridization and allopolyploidy. Molecular Phylogenetics and Evolution. 160: 107118.

Howard, WL. 1945. Luther Burbank's Plant Contributions. University of California College of Agriculture. Bulletin 691. Berkeley, California.

Karp, D. 2015. Luther Burbank's Plums. HortScience. 50(2): 189-194.

Keil, J.B. 1950. Image of box of plums from Burbank's Goldridge Farm. Sonoma County Library Digital Collections: Luther Burbank Home & Gardens Archives. https://digital.sonomalibrary.org/Documents/Detail/burbank-plums/254975

Liu, C. D. Feng, W. Peng, J. Hao, J. Wang, J. Pan, Y. He. 2020. Chromosome-level draft genome of a diploid plum (*Prunus salicina*). GigaScience. 9(12). https://doi.org/10.1093/gigascience/giaa130 .

Marti, Fernandez I, A., Saski, C. A., Manganaris, G. A., Gasic, K., & Crisosto, C. H. 2018. Genomic Sequencing of Japanese Plum (*Prunus salicina* Lindl.) Mutants Provides a New Model for Rosaceae Fruit Ripening Studies. Frontiers in plant science, 9, 21. http://doi:10.3389/fpls.2018.0002 1

Mendel, G. 1866. Vershuche über plflanzenhybriden. Verhandlungen des Naturforschenden Vereines in Brünn. 4: 3-47.

Minas, I. S., Font I Forcada, C., Dangl, G. S., Gradziel, T. M., Dandekar, A. M., & Crisosto, C. H. 2015. Discovery of non-climacteric and suppressed climacteric bud sport mutations originating from a climacteric Japanese plum cultivar (*Prunus salicina* Lindl.). Frontiers in plant science, 6, 316. doi:10.3389/fpls.2015.00316

Poland, J.A., P.J. Brown, M.E. Sorrells, and J-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7(2): e32253. https://doi.org/10.1371/journal.pone.0032253.

Pootakham, W., P. Ruang-Areerate, N. Jomchai, C. Sonthirod, D. Sangsrakru, T. Yoocha, K. Theerawattanasuk, K. Nirapathpongporn, P. Romruensukharom, S. Tragoonrung, and S. Tangphatsornruang. 2015. Construction of a high-density integrated genetic linkage map of rubber tree (*Hevea brasiliensis*) using genotyping-by-sequencing (GBS). Frontiers in Plant Science: Plant Genetics and Genomics. https://doi.org/10.3389/fpls.2015.00367 .

Pritchard, J.K., M. Stephens, and P. Donnelly. 2009. Inference of population structure using multilocus genotype data. Genetics. 155:945-959.

Salazar, J.A., I. Pacheco, C. Silva, P. Zapata, P. Shinya, D. Ruiz, P. Martinez-Gomez, and R. Infante. 2018. Development and applicability of GBS approach for genomic studies in Japanese plum (*Prunus salicina* Lindl.). The Journal of Horticultural Science and Biotechnology. 94: 284-294.

Sanchez-Perez, R., P. Martınez-Gomez, F. Dicenta, J. Egea, and D. Ruiz. 2006. Level and transmission of genetic heterozygosity in apricot (*Prunus armeniaca* L.) explored using simple sequence repeat markers. Genetic Resources and Crop Evolution. 53: 763-770. https://doi.org/10.1007/s10722-004-4636-0 .

Shaw, G.T. 1910. Image: Plum comparison. Sonoma County Library Digital Collections: Luther Burbank Home & Gardens Archives. https://digital.sonomalibrary.org/Documents/Detail/plum-comparison/254016

Stebbins Jr., G.L. 1947. Types of polyploids: their classificationand significance. Advances in Genetics. 1: 403-429. https://doi.org/10.1016/S0065-2660(08)60490-3

Sundouri, A.S., Sk. Verma, M.K. Sharma, A. Kumar, N. Nazir, and A. Khalil. 2017. Characterization of Newly Introduced Exotic Plum Cultivars for Character Association and Genetic Improvement. Current Journal of Applied Science and Technology. 24(2) 1-10.

Suzuki, R. and H. Shimodaira. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 22(12): 1540-2. https://doi.org/10.1093/bioinformatics/btl117 .

Tonkin-Hill, G. and S. Lee. 2016. Starmie: population structure model inference and visualization. https://CRAN.R-project.org/package=starmie .

Velasco, J.G., M. Malosetti, C.H. Hunt, E. S. Mace, D.R. Jorday, and F.A. van Eeuwijk. 2019. Combining pedigree and genomic information to improve prediction quality: an example in sorghum. Theoretical and Applied Genetics. 132: 2055-2067. https://doi.org//10.1007/s00122-019-03337-w .

Vernesi, C., B. Crestanello, E. Pecchiolo, D. Tartari, D. Caramelli, H. Hauffe, and G. Bertorelle. 2003. The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): A microsatellite analysis. Molecular Ecology. 12(3): 585-595. https://doi.org/10.1046/j.1365-294X.2003.01763.x .

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Yang, X., J. Song, Q. You, D.R. Paudel, J. Zhang, and J. Wang. (2017). Mining sequence variations in representative polyploid sugarcane germplasm accessions. BMC Genomics. 18(594).

Zeisset I. and T.J.C. Beebee. 2001. Determination of biogeographical range: an application of molecular phylogeography to the European pool frog *Rana lessonae.* Proceedings of the Royal Society of London Britain . 268: 933-938. https://doi.org/10.1098/rspb.2001.1600 .

Zhebentyayeva T., V. Shankar, R. Scorza, A. Callahan, M. Ravelonandro, S. Castro, T. DeJong, C.A. Saski, and C. Dardick. (2019). Genetic characterization of worldwide *Prunus domestica* (plum) germplasm using sequence-based genotyping. Horticulture Research 6:12.

**Appendix:**

Appendix 1: Burbank *Prunus* cultivars and wild relatives sequenced in this study including the site where the cultivar is located, hypothetical pedigrees reported in historical sources, and the year the cultivar was introduced.  Any cultivar with an * is a wild relative, and not a Burbank introduction.  Cultivars are sorted by their historically reported pedigree.

| Cultivar | Site | Historically Reported Species | Year of Introduction | Literature Source |
|---|---|---|---|---|
| Catherine Bunnell | LBHG | 'Santa Rosa' x ? | 1908 | Never formally introduced |
| Shiro | GR, WEO | (*P. salicina* x *P. simonii*) x *P. cerasifera* | 1899 | New Creations in Fruits and Flowers Supplement |
| Anderson's Early Red* | WEO | *P. americana* | | Wild Relative |
| Rutland | GR | *P. armeniaca* x *P. salicina* | 1905 | Burbank 1914-1915 |
| Great Yellow | LBHG, WEO | *P. cerasifera* x ? | 1920 | Register of New Fruit and Nut Varieties 1920-1950 |
| Vesuvius | WEO | *P. cerasifera* x *P. salicina* | 1907 | Fancher Creek Nurseries 1907 |
| June Redskin | LBHG | *P. cerasifera* x *P. simonii* | 1922 | Register of New Fruit and Nut Varieties 1920-1950 |
| Grand Prize | WEO | *P. domestica* | 1932 | Register of New Fruit and Nut Varieties 1920-1950 |
| OG Stoneless | WEO | *P. domestica* | | Wild Relative |
| Sans Noyau* | WEO | *P. domestica* | 1768 | Hedrick 1911 |
| Top of the Hill | GR | *P. domestica* | | Never formally introduced |
| Botanky | WEO | *P. domestica* ssp. *domestica* | 1887 | Catalog of Fruit and Shade Trees 1887 |
| Latest of All | GR, WEO | *P. domestica* ssp. *domestica* | | Never formally introduced |
| French* | WEO | *P. domestica* ssp. *Insititia* | 1889 | Hedrick 1911 |
| Late Goose* | WEO | *P. rivularis* | | Wild Relative |
| Victory | WEO | *P. rivularis* | 1911 | Twentieth Century Fruits |
| Abundance | WEO | *P. salicina* | 1888 | Hedrick 1911 |
| Brookgold* | WEO | *P. salicina* | | Wild Relative |
| Elephant Heart | GR, WEO | *P. salicina* | 1920 | Register of New Fruit and Nut Varieties 1920-1950 |
| Improved Satsuma | LBHG | *P. salicina* | | Never formally introduced |
| Satsuma | WEO | *P. salicina* | 1886 | Burbank 1914-1915 |
| Sultan | LBHG | *P. salicina* | 1899 | Burbank 1914-1915 |
| Beauty | LBHG, WEO | *P. salicina* x ? | 1911 | Twentieth Century Fruits |

| Formosa | GR, LBHG | *P. salicina* x ? | 1907 | Nursery Catalog 1914 |
|---|---|---|---|---|
| Formosa? | WEO | *P. salicina* x ? | 1907 | Nursery Catalog 1914 |
| Lieb | LBHG | *P. salicina* x ? | 1914 | Burbank 1914-1915 |
| Mammoth Cardinal | LBHG, WEO | *P. salicina* x ? | 1919 | Register of New Fruit and Nut Varieties 1920-1950 |
| Apex | GR, WEO | *P. salicina* x *P. armeniaca* | 1911 | Twentieth Century Fruits |
| Burbank | GR, WEO | *P. salicina* x *P. armeniaca* | 1914 | Burbank 1914-1915 |
| El Dorado | GR, LBHG, WEO | *P. salicina* x *P. simonii* | 1904 | Twentieth Century Fruits |
| Wickson | WEO | *P. salicina* x *P. simonii* | 1892 | New Creations in Fruits and Flowers 1894 |
| Santa Rosa | GR | *P. salicina, P. americana,* and *P. simonii* | 1906 | Burbank 1914-1915 |
| DPRU 2430* | WEO | *P. simonii* | | Wild relative |
| Simon* | WEO | *P. simonii* | 1872 | Hedrick 1911 |
| Ballena | GR | *P. simonii* x 'Deleware' | 1906 | Rutland 1909 |
| Chalco | WEO | *P. simonii* x *P. salicina* | 1898 | New Creations in Fruits and Flowers Supplement |
| Perfection | WEO | | | No data |
| Unknown 7 | GR | | | No data |
| Unknown Multi | GR | | | No data |
| Unnamed Plumcot | LBHG | | | No data |

Appendix 2: Cultivars from the USDA-ARS-NCGR Wolfskill location with their accession numbers.

| Cultivar Name (or species if unnamed) | Accession Number |
|---|---|
| Abundance | DPRU 919 |
| Anderson's Early Red | DPRU 843 |
| Apex | DPRU 1170 |
| Beauty | DPRU 2120 |
| Botanky | DPRU 372 |
| Brookgold | DPRU 1736 |
| Burbank | DPRU 936 |
| Chalco | DPRU 431 |
| *P. americana* | DPRU 1250 |
| *P. simonii* | DPRU 2430 |
| El Dorado | DPRU 2122 |
| Elephant Heart | DPRU 2123 |
| Formosa? | DPRU 924 |
| French | DPRU 436 |
| Grand Prize | DPRU 1572 |
| Great Yellow | DPRU 2105 |
| Late Goose | DPRU 546 |
| Latest of All | DPRU 427 |
| Mammoth Cardinal | DPRU 2127 |
| Original Stoneless | DPRU 2302 |
| Perfection | DPRU 1720 |
| Sans Noyau | DPRU 2419 |
| Satsuma | DPRU 438 |
| Shiro | DPRU 2132 |
| Simon | DPRU 545 |
| Vesuvius | DPRU 2108 |
| Victory | DPRU 791 |
| Wickson | DPRU 2135 |

# Chapter 2:  Relatedness of Luther Burbank's Plum (*Prunus* sp) Introductions based on  Genotyping by Sequencing (GBS)

## Abstract:

Horticultural artist and plant breeder extraordinaire Luther Burbank worked with many different species of plants.  During his 50-year career, he introduced over 800 cultivars, including more than 150 accessions of plums (*Prunus* spp.) between the late 1800's and early 1900's.  Burbank preferred utilizing wide, inter-specific crosses to create a vast range of phenotypic variation and artificially selected from the extremes. While a magnificent artist, Burbank was a substandard scientist because he was derelict in pedigree note-taking.  Though many of his introductions are extinct, hobbyists, enthusiasts, and international collections retain nearly a third of the desirable cultivars he bred.  For a century, many of his hybridizations remained irreproducible mysteries until modern genomic and computational tools developed their resolution and statistical power.

Today, Genotyping by Sequencing (GBS) is a useful tool for pedigree reconstruction in the absence of reliable records.  GBS can inform Principal Component Analyses (PCA), Identity by Descent (IBD) kinship , and phylogenetic admixture, revealing complex relationships among taxa.  In this study, whole genome sequencing was performed on 53 *Prunus* taxa used by Luther Burbank in his breeding experiments This is the most comprehensive genetic survey of his work done to date.  It provides valuable information on the relatedness of the individuals in Burbank's *Prunus* experiments.  The research has implications for pedigree reconstruction and prioritizing conservation in collections curation for future studies.

## Introduction

During the turn of the 19th-20th century, Luther Burbank was one of the most prolific plant breeders of all time.  He was not a classically trained scientist, but rather a highly observant horticultural artist.  He had a keen interest in generating wide crosses between distant relatives in hopes of shuffling genomes and artificially selecting extreme or disruptive phenotypes (Burbank 1914), and as such produced hundreds of thousands of seedlings for evaluation (Topp et al. 2011).  Using this method, Mr. Burbank completely revolutionized the human perception for what a plum could be, commercially introducing over 150 cultivars of plums, prunes, and plumcots in the span of 40 years (Hedrick 1911, Howard 1945, Brooks and Olmo 1952, Karp 2015) (Appendix 1).  His breeding population of plums represents many intraspecific hybrids, interspecific hybrids, and bud sports with a huge range of phenotypic variation.  Unfortunately, Luther Burbank kept very poor breeding records and relied instead on faded strips of clothing tied to branches and his memory to track the pedigrees of his plants.  Taking good notes is essential for the reporting accuracy and reproducibility of any plant breeding program.

For thousands of years, humans used phenotypes to inform decisions in plant breeding, but eventually, this information reached a plateau in its usefulness due to phenotype by environment interactions.  When first applied as a technology, access to genomic data was either too expensive to be worthwhile or was limited in its predictive

capabilities.  Only small regions of Simple Sequence Repeats (SSRs) could be

processed computationally. Analyses were confounded by polyploidy, bud sports, non-

mendelian segregation, and the ability of plants to express multiple phenotypes from a

single genotype.  Today, high-throughput gene sequence is cost effective, and covers

much larger stretches of the genome for association mapping or discovering

Quantitative Trait Loci (QTL).

Filling in holes in pedigree notes with comprehensive genomic data provides a

rich resource that is much more accurate than relying solely on phenotypic breeding

notes alone (Luby et al. 2022).  This technology is valuable for the identification and

characterization of germplasm in current breeding lines in both conventional and

organic settings.  When a desirable phenotypic trait is linked to a genotype, plants can

be screened in the seedling stage through marker-assisted selection instead of waiting

until the plant reaches maturity, saving time and money for the breeding program.  Plant

breeders without this tool may have lower accuracy in parentage reporting, depending

on how meticulously they kept records and how precise they are with controlling the

parents of their specific crosses.  These data are also used as a tool to encourage the

stacking of favorable alleles while preventing inbreeding depression and linkage drag in

highly inbred populations (Imai and Kuniga 2021).  The identification and

characterization of germplasm in current breeding lines often is informed by genomic

data, in both conventional and organic settings.

As plant breeders or curators retire and pass their collections on to the next

generation, it is important for incoming researchers to broadly sample the collection(s),

establishing a genomic baseline for their target population.  This data informs breeding

choices and identifies goals for cultivar selection or prioritizing conservation.  Visualizing the relationships among the organisms in a breeding population are accomplished through Principal Component Analyses (PCA), Identity by Decent (IBD), and phylogenetic admixture.  These analyses provide information useful for partially reconstructing pedigree notes where no traditional notes and limited members of the breeding population exist.

Including some wild-type relatives or founders in the sampled population teases out the differences in cultivars from bottlenecked populations when visualizing the data through PCA.  For example, PCA done of *P. salicina* using eight SSR markers showed that interspecific hybrids tend to cluster together depending on their admixture of genotypes, each being pulled closer to their dominant ancestor by their shared components (Carrasco et al. 2012), but when wild types or founders of the population are included, the genetic differences among the population become easier to visualize. Kinship Matrices have limited application in annual crops with a highly structured population with controlled crosses, as the matrix can vary from generation to generation depending on the stability of the genome to phenome map for a trait of interest (Van Tassel et al. 2022), but they are highly applicable to the reconstruction of pedigree data because the genomes with multi-parental populations covering many generations are considered fixed traits, especially in perennial tree crops (Goudet et al. 2018). Phylogenetic admixture is useful in surveying the breadth of genetic diversity in germplasm collections, which in turn helps prioritize conservation choices where space is limited (Pikunova et al. 2022).

The primary goal of this research is to look at how a population of 53 inter and intra-specific *Prunus* taxa introduced by Luther Burbank nearly a century ago are related to each other using PCA, IBD, and phylogenetic admixture.  The population in this study is presumed to have been comprised of six *Prunus* species based on limited historical data such as Burbank's nursery catalogs (Howard 1945), Plums of New York (Hedrick 1911), and the Register of New Fruit and Nut Cultivars 1920-1950 (Brooks and Olmo 1952).  For this reason, wild-type accessions of *P. simonii*, *P. domestica*, *P. americana*, *P. salicina*, *P. rivularis*, and *P. cerasifera* were included to elucidate rather convoluted relationships.

**Materials and Methods:**

Location of taxa and phenotype data:

A comprehensive list of Burbank's plum introductions is readily available (Howard, 1945).  However, by going back into the primary source material used by Howard to generate this list, the names of six more plums that were introduced by Luther emerged (Hedrick, 1904).  Also missing from this list are the taxa that were patented after Burbank's death by his widow Elizabeth in collaboration with the Stark Brothers.  This target list was used to hunt for specific cultivars through ARS repository access, word of mouth, and local CRFG scion exchanges.  Once a cultivar was located, old literature was searched for claims Luther Burbank made about their parentage as well as any historical images that may accompany them (Burbank 1914, Hedrick 1911, Howard 1945, Brooks and Olmo 1952).  Scions of material found at exchanges or through word-of-mouth were multi-grafted onto mature trees with *Prunus cerasifera* 'Myrobalan 29C' as a universal rootstock at the Luther Burbank Home & Gardens

(LBH&G) in Santa Rosa, California.  Historic maps were referenced for cultivar names

at the Luther Burbank's Goldridge Experiment Farm (GR) in Sebastopol California.  The

GRIN Global Database and map of the USDA-ARS-National Clonal Germplasm

Repository's Wolfskill Experimental Orchard (WEO) plum block were used for locating

cultivars in their collection.

Genomic characterization:

Young leaf tissue was collected from trees in the early spring and stored in silica

gel at room temperature until sufficiently dry, then frozen at -80ºC.  .  The DNA

extraction protocol followed DNeasy Plant Kit from Qiagin.  Retrieved sequences were

aligned to *P. salicina* 'Sanyueli' (Liu et al. 2020).  Samples with more than 90% missing

data were discarded, and this resulted in discarding 28/96 taxa. Approximately 50,000

SNPs were retrieved.  All genotypes were set to a minimum depth of <5 missing, and

then SNPs with more than 50% missing were discarded using TASSEL5.  The final

imputed dataset contained 24,147 SNPs.  Missing genotypes were imputed using

Beagle (Browning et al. 2018).

Principal Component Analysis:

Principal component analysis was performed in TASSEL 5 using default

parameters, and the resulting data was exported to R for visualization using the R

packages *'ggplot2'* and *'ggthemes'* (Wickham 2016, Arnold 2021).  Admixture cluster

data was calculated using STRUCTURE for K values of 2, 3, 4, 5, and 6.  Diagnostic

plots which maximized L(K) and minimized DIC (Zeisset & Beebee 2001; Ciofi et al.

2002; Vernesi et al. 2003; Hampton et al. 2004; Evanno, Regnaut, & Goudet 2005; Gao,

Bryc, & Bustamante 2011) were used to select the optimal value of K=4.  These

thresholds show the maximum statistically relevant value of K where the benefit to

splitting the population into smaller pieces is no longer beneficial.  These data were

added to the PCA plot to visualize the relationships among Burbank's *Prunus* cultivars.

Heatmap:

An identity-by-state (IBS) matrix  (Endelman & Jannink 2012) was generated with

the *'Kinship'* analysis function using default parameters in TASSEL 5. The IBS matrix

was exported to R and was visualized using the function *pheatmap()* from the R

package '*pheatmap*' (Kolde 2019).  Since these taxa come from a population of related

individuals, IBS is synonymous with Identity by Decent (IBD).

Phylogenetic Trees:

Phylogenetic relationships were visualized as unrooted trees using the

Archaeopterix package within TASSEL 5 based on IBD values (Bradbury et al. 2007).

Pie charts were added to one of these trees to display admixture.

**Results:**

The phenotypic diversity in this population incorporated various combinations from

yellow, red, purple, or blue for the exocarp (Figure 1), yellow, green, or red mesocarp,

and free or cling-stone endocarps.

**Fig. 1**. Exocarp diversity in some of Burbank's plum (*Prunus* sp) introductions grown at the Wolfskill Experimental Orchard (WEO) in Winters, CA or the Luther Burbank Home & Gardens (LBHG) in Santa Rosa, CA.  Each fruit is labelled with their cultivar name and accession number where applicable below them.


A kinship matrix based on Identity by Decent (IBD) values of all taxa reveals degree of genetic similarity between the taxa (Figure 2).  Colors in this matrix correlated to the impact from shared Identity by State (IBS).  Warm colors (yellow, orange, and red) show genotypes that are more alike.  The intensity of this color palate shows the genotypes that share more rare alleles.  Cool colors (whites and blues) show genotypes that are more likely to have opposing alleles, indicating that their genotypes are less alike.
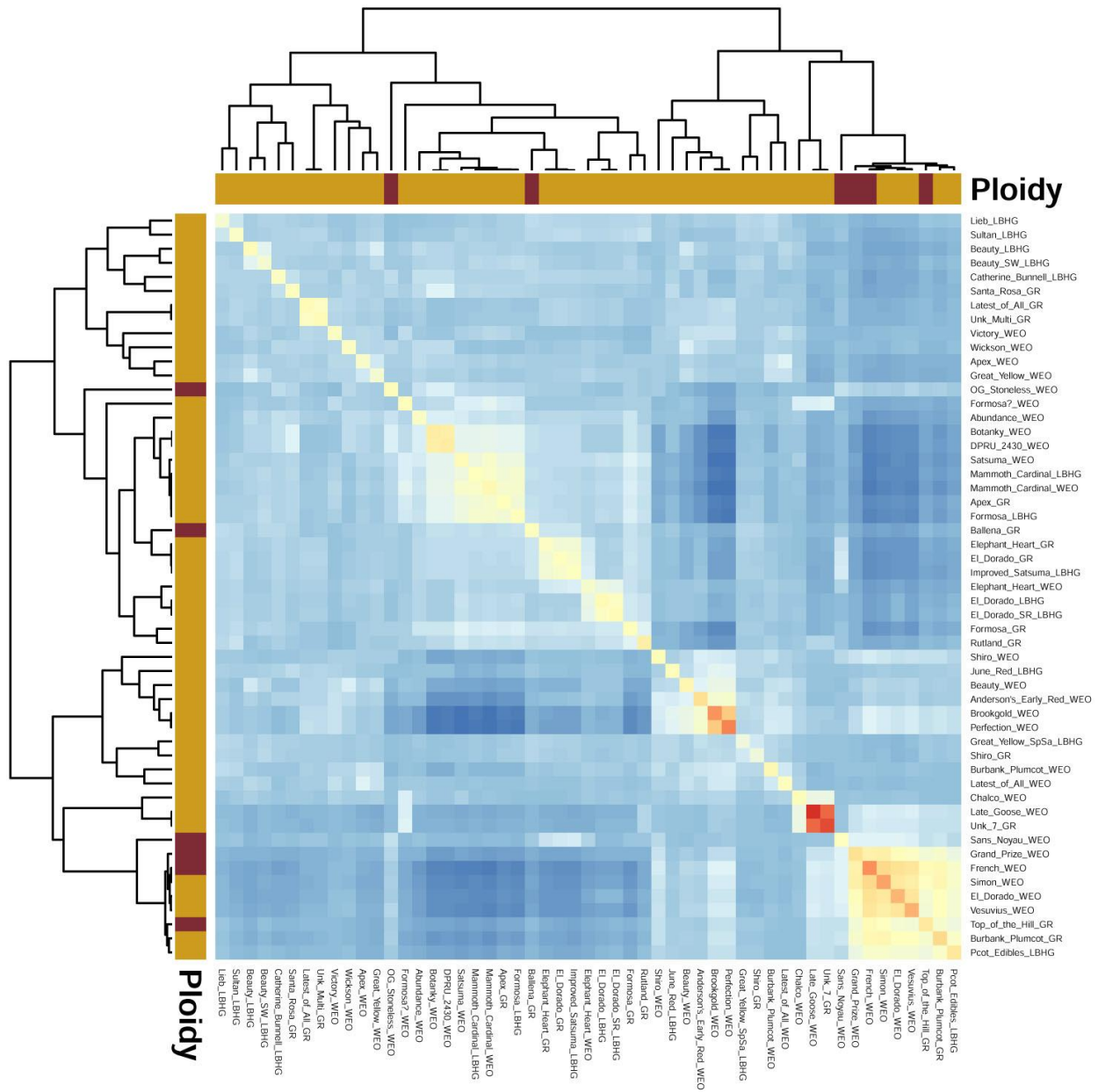
**Fig 2**. A kinship heat map of 53 *Prunus* taxa introduced by Luther Burbank shows their degree of genetic similarity based on Identity by State (IBS) calculated in TASSEL 5. Ploidy of the taxa are indicated as gold (diploid, 2n) or maroon (hexaploid, 6n).

PCs for the first five components are reported for each of the taxa (Appendix 2).

PC's one and two, which account for 32.65% of the cumulative proportion (Table 1),

were plotted against each other (Figure 3). Ellipses representing admixture clusters

were added to further visualize the relatedness of each taxon.

| PC | Proportion of Total (%) | Cumulative Proportion (%) |
|---|---|---|
| 1 | 20.47% | 20.47% |
| 2 | 12.18% | 32.65% |
| 3 | 9.50% | 42.15% |
| 4 | 4.68% | 46.83% |
| 5 | 3.64% | 50.47% |

Table 1.Variation explained by the first five principal components along with eigenvalues, proportion of total, and cumulative total for 53 *Prunus* taxa of a Burbank Breeding population generated using TASSEL 5.
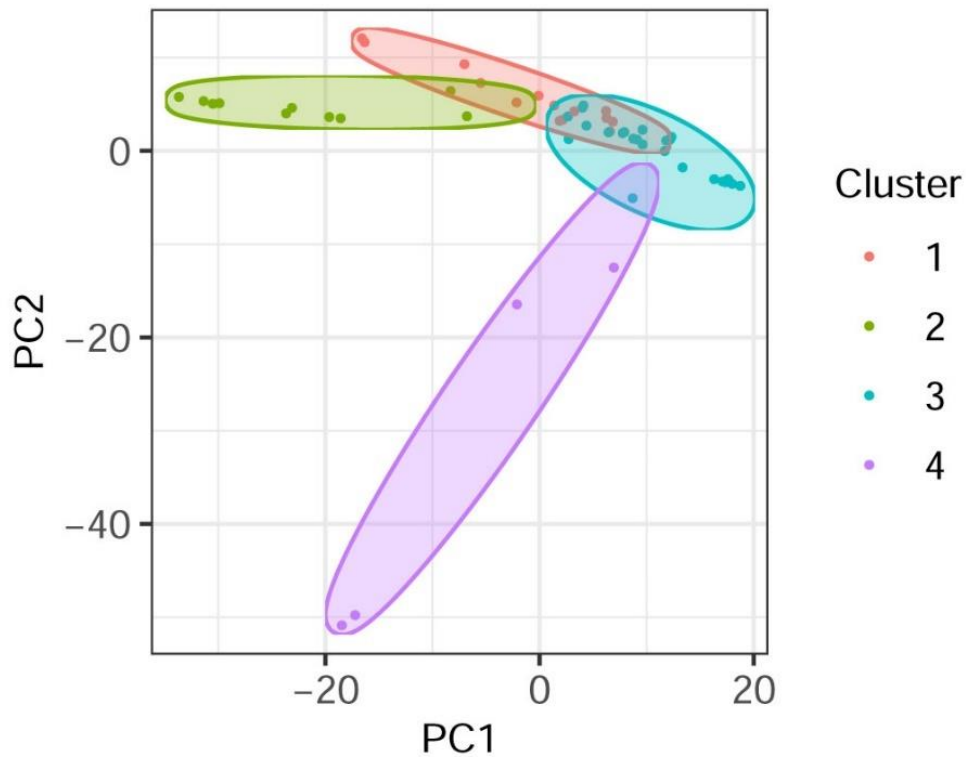


**Fig 3.** PC1 and PC2 eigenvalues plotted for 53 Burbank-introduced *Prunus* taxa with admixture cluster groups (K=4) highlighted by ellipses.

A Phylogenetic tree was generated using Archaeopteryx (Han and Zmasek 2009) a tree-visualization package nestled in TASSEL 5 (Bradberry et al.2007, Glaubitz et al. 2014).  Pie charts representing the admixture of each genomic cluster for K=4 were added to branches to illustrate the proportions of the genome shared among the taxa represented (Figure 4).
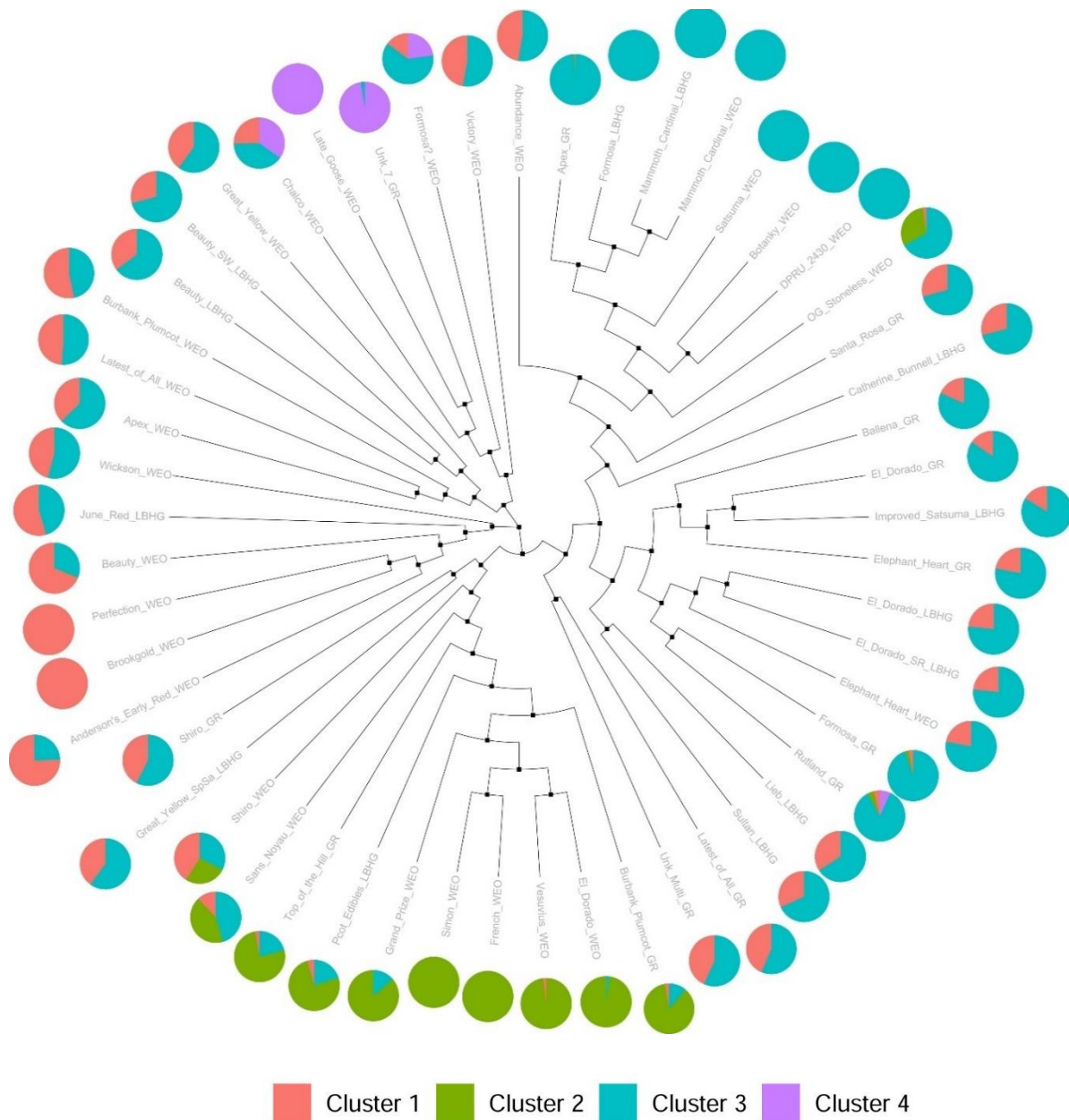
**Fig 4**. A circular dendrogram representing the unrooted relationships among 53 Burbank based on IBD calculations, combined with admixture pie charts from cluster analysis.

Another unrooted dendrogram with branch lengths representing genetic distance was generated using Archaeopteryx (Han and Zmasek 2009)  a tree-visualization package nestled in TASSEL 5 (Bradberry et al.2007, Glaubitz et al. 2014) revealing the similarity and divergence among taxa (Figure 5).
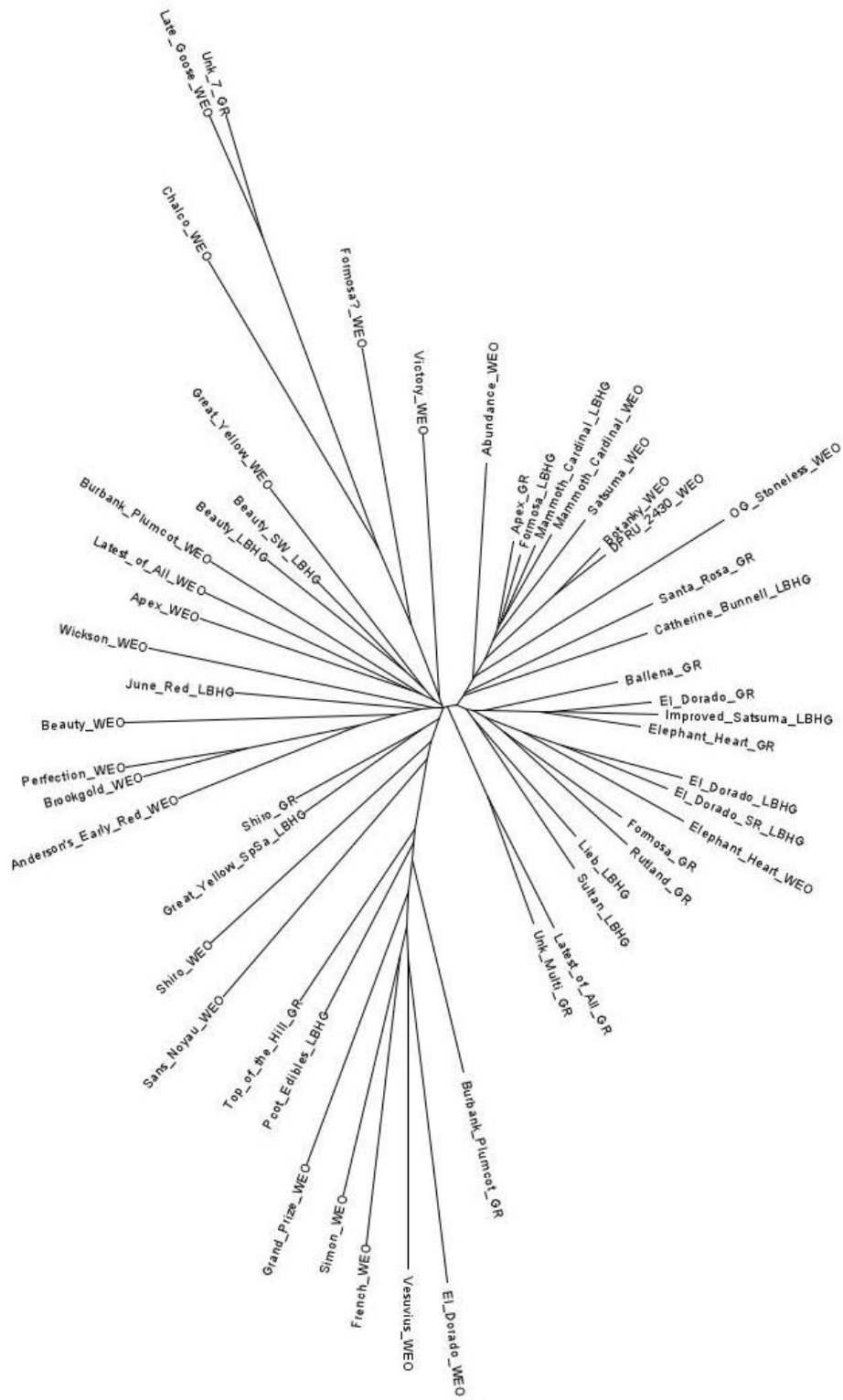
**Fig. 5.** An unrooted dendrogram of 53 *Prunus* taxa introduced or used in breeding experiments by Luther Burbank based on IBD values with branch lengths indicating genetic distance.

**Discussion:**

The combination of IBD kinship (Figure 1), PCA (Figure 2), phylogenetic admixture (Figure 3), and genetic distance (Figure 4) paints a congruent picture of relatedness among Burbank-introduced taxa, with eight primary groups of note emerging from the population sampled. Although these analyses are not direct indications of parent-offspring relationships in a strict pedigree sense, they do provide insight into the numbers of rare alleles shared among this breeding population.

The kinship matrix based on IBD values (Figure 1) indicates that the highest number of rare alleles are shared between 'Late Goose – DPRU.546' (WEO) and 'Unknown 7' (GR). This is also reflected in the branch lengths of the unrooted dendrogram (Figure 5), and the purple ellipse on the PC plot, covering the widest data range. 'Late Goose – DPRU.546' is one of the wild relatives (*P. rivularis* Scheele) from the Wolfskill Experimental Orchard that was included to help tease apart the tangled genotypes in this study, while 'Unknown 7' is an older tree that either was planted by Luther Burbank or is a seedling of a tree he planted located at the Goldridge Burbank Experiment Farm. This is not a confirmation of the identity of 'Unknown 7' but does show a strong connection between the two.

Sister to this group is 'Chalco – DPRU.431' (WEO), which is listed as *Prunus spp.* on the USDA-ARS-NCGR Wolfskill map. 'Chalco' was bred and introduced by Burbank in 1898. He reported it to be a 'Simon x Burbank' hybrid with perhaps some *P. americana* in its 12-year breeding history (Hedrick 1911), which is supported by the admixture analysis, containing approximately equal thirds of *P. simonii*, *P. salicina*, and a North American plum, represented in this study by 'Anderson's Early Red – DPRU.843'. *P. rivularis* 'Late Goose – DPRU.546,' and 'Unknown 7' are comprised

mostly of Cluster 4 in the admixture analysis (Figure 5). Phenotypically, these taxa

have nearly disease-free trees, and small fruits with a yellow mesocarp, and a freestone

endocarp. Interestingly, the exocarps of these fruits are quite different; 'Chalco –

DPRU.431' is black, 'Anderson's Early Red – DPRU.843' is red, and 'Late Goose –

DPRU.546' is yellow. In USDA hardiness zone 9b, these fruits ripen in early to mid-

June. The fruits are highly perishable, but also quite prolific. Commercial value of

these cultivars may be restricted to using them as rootstock.

Another group with shared rare alleles is found in the taxa 'Brookgold –

DPRU.1736' and 'Perfection – DPRU.1720,' both from the Wolfskill collection. These

two taxa have an admixture comprised entirely of Cluster 1. 'Perfection' was a synonym

for 'Wickson' once, arguably one of Burbank's plum introductions. However, both the

phenotype and genotype for this 'Perfection – DPRU.1720' accession are vastly

different from 'Wickson – DPRU.2135.' 'Wickson' is a pointy-bottom *P. salicina* type.

'Perfection – DPRU.1720' from the Wolfskill collection resembles the small, rounded

phenotype of a *P. cerasifera*. Notes from the Wolfskill plum block map suggest that

'Brookgold – DPRU.1736' may be a *P. cerasifera* as well, which is supported by the

genomic data from this study, including the IBD heatmap, PCA plot, and phylogenetic

admixture dendrogram. This pair of taxa is sister to 'Anderson's Early Red –

DPRU.843,' the *P. americana* representative. 'Anderson's Early Red – DPRU.843' has

an admixture comprised of approximately three-quarters Cluster 1 and one quarter of

Cluster 3. It would be quite interesting to revisit more taxa in *P. americana* and *P.*

*cerasifera* to look at the relationship of their shared rare alleles. Both appear to have

more disease resistance, smaller fruit size and high perishability, making them limited in

their degree of usefulness for a breeding program.  It is worth noting that diploid *P. cerasifera* is reported as one of the progenitors of the hexaploid species *P. domestica* (Zhebentyayeva et al. 2019).  As such, *P. cerasifera* works quite well as a universal plum rootstock to both diploid and hexaploid taxa despite its limitations for fruit quality.

The largest cluster of taxa with shared rare alleles contains a combination of hexaploid and diploid taxa.  The hexaploid *P. domestica* taxa in this group include 'Grand Prize – DPRU.1572,' 'French – DPRU.436,' and 'Top of the Hill (GR)."  The diploid taxa in this group are 'Simon – DPRU.545' (*P. simonii*), 'El Dorado – DPRU.2122' (*P. simonii* x ?, Wolfskill collection), 'Vesuvius – DPRU.2108' (*P. cerasifera*), 'Burbank Plumcot' (historically reported as *P. salicina* x *P. armeniaca*, Goldridge accession) and 'Pcot Edibles'(LBHG accession).  These taxa all have a primary admixture of Cluster 2.  The phenotype of 'Pcot Edibles' appears visually as an intermediate between a *P. domestica* with its dark blue exterior and greenish yellow mesocarp, and *P. simonii* with its fruit texture and small, round, free-stoned endocarp.  More work is needed to see the correlation between the genotype and phenotype in this group.  The only other taxa to have Cluster two present in its admixture is 'Sans Noyau – DPRU.2419' which appears on the opposite side of the unrooted circular dendrogram.  'Sans Noyau – DPRU.2419' and 'Original Stoneless – DPRU.2302' are both intriguing cultivars because their endocarp is reduced to a small lignified remnant of the funiculus.  Burbank's breeding goal was to make these like almond-stuffed prunes, but he was unable to accomplish that goal in his lifetime.

'Shiro' from Goldridge and 'Great Yellow' from LBHG share a considerable number of rare alleles.  The 'Shiro – DPRU.2132' accession from Wolfskill showed a

paraphyletic relationship with these sister taxa. The 'Great Yellow – DPRU.2105'
Wolfskill accession was located much more distantly on the dendrogram, appearing
genetically to be much more like the plumcot 'Apex – DPRU.1170,' introduced by
Burbank in 1911.  It may be that 'Great Yellow' from LBH&G was a mislabeled
accession of 'Shiro' when offered at a scion exchange.  'Shiro' is a plum Burbank bred
and introduced in 1899 (Appendix 1) that is still readily available today through various
tree nurseries.  'Great Yellow' is a Burbank plum that bred by him but was introduced
and patented posthumously by Stark Brothers in 1931 (USPTO #13, Appendix 1).
Given the decades-long distance between their introductions, one would presume their
genotypes to be more distinct than indicated by the clade of Goldridge 'Shiro' and LBHG
'Great Yellow.'  'Shiro' was reported to be an offspring of 'Wickson' – slightly smaller in
size than 'Wickson', with a similar pointy shape and incredibly prolific fruit sets (Hedrick
1911).  'Great Yellow' reportedly has a more rounded shape as is indicated by the
watercolor submitted for its patent (LBH&G Archives).

     'Botanky – DPRU.372' and '*P.simonii* - DPRU 2430' share some rare alleles as
indicated by the kinship matrix.  Burbank had three introductions named with different
iterations of the word 'Botan.'  For this reason, the accession 'Botanky – DPRU.372'
was included in this study.  These were direct introductions after Burbank received seed
from a bulb broker in Japan, with artificial selection from the seed population done as a
simple phenotyping followed by immediate cultivar release without additional breeding.
In the USDA-ARS-NCGR Wolfskill plum block map, 'Botanky – DPRU.372' is listed as
*Prunus spp.*, possibly *P. salicina*.  '*P. simonii* - DPRU 2430' is listed as a wild-collected
accession.

These two taxa are nested within a clade that contains rare allele signatures for 'Satsuma – DPRU.438' (WEO), two accessions of 'Mammoth Cardinal' (WEO – DPRU.2127 and LBHG), 'Apex' (Goldridge), and 'Formosa' (LBHG).  'Mammoth Cardinal' was patented and introduced by the Stark Bro's in 1934, eight years after Luther's death (USPTO #16).  'Mammoth Cardinal' is characterized by its red exocarp, yellow mesocarp, and small, free-stone endocarp, indicating visually that it would have a *P. simonii* influence.  'Formosa' was bred by Burbank and then introduced in 1907.  It resembles 'Mammoth Cardinal,' but has a thicker wax on the surface of its fruit.  'Apex' was touted as an early plumcot introduction, being bred then introduced by Burbank in 1911.  'Apex' is highly prolific, free-stone, and early ripening.  It has a complex flavor profile which may be attributed to either *P. simonii* or *P. armeniaca*.

All these taxa are sister to 'Abundance – DPRU.919' (WEO).  'Abundance' and 'Satsuma' were both direct introductions of Burbank's, and all of them have an admixture comprised almost entirely of Cluster 3.  'Abundance' was first released under the name 'Botan' in 1888 and is sometimes confused with the 'Abundance Plumcot' Burbank released many years later.  'Abundance' seems have been important in increasing yield, though the fruit is highly perishable and is not great for commercial settings.  'Satsuma' was a directly selected introduction in 1886.  The phenotypic influence of this taxon is visible in the red flesh of many of his other plums.  Again, it is important to note that the direct introductions were the result of growing a seed until it could be phenotyped and distributed, and not the products of Burbank's breeding experiments, though he is credited with their introduction as the person who artificially selected from the seedling population and then marketed them.

A paraphyletic group that has similar taxa present appears to have shared rare alleles from a genomic background containing both *P. salicina* and *P. simonii*. The first contains 'Elephant Heart' (Goldridge), 'El Dorado (Goldridge), and 'Improved Satsuma' (LBHG) as a monophyletic group. The second monophyletic group contains 'Elephant Heart – DPRU.2123' (WEO) and two accessions of 'El Dorado' (LBHG) that were uniquely collected at different scion exchanges. This indicates the two LBHG 'El Dorado' accessions are likely the same. In general, they differ drastically from 'Elephant Heart' in phenotype. 'El Dorado' is a free-stone, rounded plum with black exocarp (skin) and golden mesocarp. The exterior has quite a bit of wax. 'Elephant Heart' is a pointy-bottomed plum with a thickly waxy, purplish exocarp and a red mesocarp. 'El Dorado' was introduced by Burbank in 1904 and continues to be a readily available cultivar through scion exchanges and heirloom fruit tree distributers. It can be picked firm and ripened off-tree, making it ideal for commercial production. 'Elephant Heart' was introduced by the Stark Brothers in 1929, three years after Luther Burbank died, but was not patented. Phenotypically it has a strong resemblance to 'Satsuma,' but is considerably larger.

**Conclusions:**

As a prominent horticultural artist with a lack of note taking skills, Luther Burbank left many mysteries to solve. Nearly a century later, genotyping by sequencing provides us with a looking glass to view the relatedness of his introductions, making his haphazard style of breeding more accessible to those who prefer a more structured approach. In some cases, these data support his claims of parentage, in others they refute his claims or leave more questions to be answered.

Identity by descent, principal component analyses, and phylogenetic admixture are powerful tools that provide a roadmap for researchers deciphering inherited breeding populations, establishing a genomic baseline, guiding artificial selection decisions or, for curators of collections, prioritizing the conservation of rare and useful alleles in spaces that are accessible to the broader scientific community.

## **References:**

Arnold, J.B. (2021).  ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'.  https://github.com/jrnold/ggthemes .

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Brooks, R.M. and H.P. Olmo.  1952.  Register of New Fruit and Nut Varieties 1920-1950.  University of California Press.  http://reader.library.cornell.edu/docviewer/digital?id=chla3174378#mode/1up

Browning, B.L., Y. Zhou, and S.R. Browning.  2018.  A one-penny imputed genome from next generation reference panels.  American Journal of Human Genetics.  103(3): 338-348.  https://doi.org/10.1016/j.ajhg.2018.07.015

Burbank L., J. Whitson, R. John, and H. Smith Williams (eds).  (1914).  Luther Burbank: his methods and discoveries and their practical application.  New York; London: Luther Burbank Press.

Carrasco, B., C. Diaz, M. Moya, M. Gebauer, and R. Garcia-Gonzalez.  2012.  Genetic characterization of Japanese plum cultivars (*Prunus salicina* Lindl.) using SSR and ISSR molecular markers.  Plant Breeding, Cien. Inv. Agr.  39(3): 533-543.

Ciofi, C., M.C. Milinkovitch, J.P. Gibbs, A. Caccone, and J.R. Powell.  2002.  Microsatellite analysis of genetic divergence among populations of giant Galapagos tortoises.  Molecular Ecology.  11(11): 2265-2283.  https://doi.org/10.1046/j.1365-294X.2002.01617.x .

Endelman, J.B. and J-L. Jannink.  2012.  Shrinkage estimation of the realized relationship matrix.  G3(Bethesda) 11:1405-13.  https://doi.org/10.1534/g3.112.004259 .

Evanno, G., S. Regnaut, and J. Goudet.  2005.  Detecting the number of clusters of individuals using software STRUCTURE:  a simulation study.  Molecular Ecology.  https://doi.org?10.1111/j.1365-294x.2005.02553.x .

Gao, H., K. Bryc, and C.D. Bustamante.  2011.  On identifying the optimal number of population clusters via the deviance information criterion.  PloS One. 6(6): e21014.  https://doi.org/10.1371/journal.pone.0021014 .

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. (2014) TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS ONE 9(2): e90346

Goudet, J., T. Kay, and B.S. Weir.  2018.  How to estimate kinship.  Molecular Ecology 27(20): 4121-4135.  https://doi.org/10.1111/mec.14833 .

Hampton, J.O., P.B.S. Spencer, D.L. Alpers, L.E. Twigg, A.P. Woolnough, J. Doust, T. Higgs, and J. Pluske.  2004.  Molecular techniques, wildlife management and the importance of genetic population structure and dispersal:  a case study with feral pigs.  Journal of Applied Ecology.  41(4): 735-743. https://doi.org/10.1111/j.0021-8901.2004.00936.x .

Han, M.V. and C.M. Zmasek.  phyloXML: 2009.  XML for evolutionary biology and comparative genomics.  BMC Bioinformatics.  10(356).

Hedrick, U.P. 1911.  Plums of New York.  http://chla.library.cornell.edu/cgi/t/text/text-idx?c=chla;cc=chla;idno=5519832

Howard, WL.  1945.  Luther Burbank's Plant Contributions. University of California  College of Agriculture. Bulletin 691.  Berkeley, California.

Imai, A. and T. Kuniga.  2021.  Genome-wide estimation of pedigree haplotypes offers genetic compositions and founder origins in Japanese citrus breeding materials.  Scientia Horticulturae.  282. https://doi.org/10.1016/j.scienta.2021.110000 .

Karp, D. 2015.  Luther Burbank's Plums. HortScience.  50(2): 189-194.

Kolde, R.  2019.  pheatmap:  pretty heatmaps; Implementation of heatmaps that offers more control over dimensions and appearance.  https://CRAN.R-project.org/package=pheatmap .

Liu, C. D. Feng, W. Peng, J. Hao, J. Wang, J. Pan, Y. He.  2020.  Chromosome-level draft genome of a diploid plum (*Prunus salicina*).  GigaScience. 9(12).  https://doi.org/10.1093/gigascience/giaa130 .

Luby, James J., N.P. Howard, J.R. Tillman, and D.S. Bedford.  2022.  Extended pedigrees of apple cultivars from the University of Minnesota breeding program elucidated using SNP array markers. HortScience.  57(3).  https://doi.org/10.21273/HORTSCI16354-21 .

Pikunova, A., S. Goryunova, D. Goryunov, O. Golyaeva, M. Dolzhikova, and A. Pavlenko. C.O. Qualset (ed).  2022.  Genetic diversity and pedigree analysis of red currant germplasm.  Plants (Basel).  11(13): 1623.  https://doi.org/10.3390/plants11131623 .

Pritchard, J.K., M. Stephens, and P. Donnelly.  2009.  Inference of population structure using multilocus genotype data.  Genetics.  155:945-959.

Tonkin-Hill, G. and S. Lee.  2016.  Starmie:  population structure model inference and visualisation. https://CRAN.R-project.org/package=starmie .

Topp, B.L., D.M. Russell, M. Neumüller, M.A. Dalbó, and W. Liu. (2012). Plum (Handbook of Plant Breeding). 8, part 3. pp. 571–621. doi:10.1007/978-1-4419-0763-9_15.

Van Tassel, D.L., L.R. DeHaan, L. Diaz-Garcia, J. Hershberger, M.J. Rubin, B. Schlautman, K. Turner, and A. Miller. 2022. Re-imagining crop domestication in the era of high throughput phenomics. Current Opinion in Plant Biology. 65. https://doi.org/10.1016/j.pbi.2021.102150 .

Vernesi, C., B. Crestanello, E. Pecchiolo, D. Tartari, D. Caramelli, H. Hauffe, and G. Bertorelle. 2003. The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): A microsatellite analysis. Molecular Ecology. 12(3): 585-595. https://doi.org/10.1046/j.1365-294X.2003.01763.x .

Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Zeisset I. and T.J.C. Beebee. 2001. Determination of biogeographical range: an application of molecular phylogeography to the European pool frog *Rana lessonae.* Proceedings of the Royal Society of London Britain . 268: 933-938. https://doi.org/10.1098/rspb.2001.1600 .

Zhebentyayeva T., V. Shankar, R. Scorza, A. Callahan, M. Ravelonandro, S. Castro, T. DeJong, C.A. Saski, and C. Dardick. (2019). Genetic characterization of worldwide *Prunus domestica* (plum) germplasm using sequence-based genotyping. Horticulture Research 6:12.

## **Appendix**

Appendix 1: An updated list of all Burbank cultivars known through literature sources including their year of introduction and their reported parentage, where applicable. Any cultivar with an * was excluded from the Howard list of Burbank Plant Contributions (1945) but found in The Plums of New York (Hedrick 1911). Cultivars patented by Elizabeth Burbank and the Stark Brothers include United States Patent and Trademark Office (USPTO) numbers (Brooks and Olmo 1952). Cultivars known only from fruit prints that may have not been formally introduced are indicated as such (Luther Burbank Home & Gardens Archives).

| Burbank Plum & Plumcot List | Year of Intro | Suspected Parentage |
|---|---|---|
| A-248* | 1893 | P. munsoniana x P. triflora |
| Abundance Plumcot | 1914 | Satsuma x P. armeniaca |
| Ace | 1927 | Posthumous |
| Alhambra | 1905 | P. simonii, P. pissardii. P. domestica, P. triflora, P. americana, P. nigra |
| Allfruit | 1898 | P. simonii x P. triflora |
| America | 1898 | Robinson x Abundance (P. triflora x P. munsoniana) |
| Apex Plumcot | 1911 | P. salicina x P. armeniaca |
| Apple | 1898 | F2 Robinson x Satsuma |
| Aroma | 1914 | unknown |
| Ballena (The Whale) | 1906 | F1 Bartlett x OP |
| Bartlett | 1896 | P. simonii x Delaware |
| Beach (Improved) | 1897 | P. maritima |
| Bearer Plumcot | 1914 | (P. salicina x P. americana) x P. armeniaca |

| Burbank Plum & Plumcot List | Year of Intro | Suspected Parentage |
|---|---|---|
| Beauty | 1911 | unknown |
| Beauty Junior | 1920 | F1 Beauty x OP |
| Beejay | 1927 | Posthumous |
| Berckmans | 1887 | P. triflora |
| Best Black Blood* | 1900 | P. triflora x Simon |
| Blue Black | 1914 | fruit print |
| Blue Point | 1915 | fruit print |
| Botan (Abundance plum) | 1888 | Direct Import |
| Botankio | 1887 | Direct Import |
| Botankio No. 2 | 1887 | Direct Import |
| Bully | 1914 | unknown |
| Burbank "Nickle" Plumcot | 1914 | unknown |
| Burbank (plum) | 1888 | Direct Import |
| Burbank First* | 1906 | P. triflora |
| Burbank No. 1* | 1894 | P. triflora |
| Burbank No. 11* | 1896 | P. triflora x P. domestica |
| Burbank No. 7* | 1895 | P. triflora x P. domestica |
| Burbank plumcot | 1914 | Myrobalan x P. armeniaca |
| Burbank x Redick* | 1909 | P. triflora x P. americana |
| Callao | 1916 | unknown |
| Catherine Bunnell | 1908 | Santa Rosa x OP |
| Cazique | 1919 | unknown |
| Cel* | 1898 | Myrobalan x Wickson |
| Chabot | 1885 | Kelsey x OP |
| Chalco | 1898 | Simon x Burbank |
| Challenge | 1914 | P. simonii |
| Cherry Plumcot | 1914 | unknown |
| Choice | 1911 | America x OP |
| Climax | 1899 | Simon x Abundance |
| Climax's Brother* | 1900 | P. triflora x Simon |
| Coin | 1927 | Posthumous |
| Combination | 1901 | P. triflora, P. munsoniana, P. simonii |
| Conquest | 1911 | d'Agen or French x Sans Noyau |
| Corona Plumcot | 1914 | unknown |
| Cranberry | 1919 | P. cerasifera |
| Crimson Cluster | 1920 | Possibly 'Latest of All'; Burbank x OP |
| Delaware | 1893 | Satsuma x Kelsey |
| Discovery | 1915 | unknown |
| Dixie* | 1899 | Burbank x OP |
| Doris | 1894 | Satsuma x P. cerasifera |
| Duarte | 1900 | America x Climax |
| Early Crimson | 1914 | wild California, European, Japanese |
| Early Pale* | 1897 | possibly 'Lieb' |
| East | 1908 | Combination x Beach Plum |
| El Dorado | 1904 | P. triflora x P. simonii |

| Burbank Plum & Plumcot List | Year of Intro | Suspected Parentage |
|---|---|---|
| Elephant Heart | 1929 | Posthumous introduction by Stark Bros |
| Epoch | 1911 | P. besseyi x America |
| First | 1901 | Wild Goose x (Hawkeye, Hammer, Milton, Wyant, Wayland, Burbank) |
| Flaming delicious | 1932 | Posthumous USPTO #14 |
| Flickenger | 1921 | unknown |
| Formosa | 1907 | mix of P. triflora |
| Fourth of July | 1901 | F2 of d'Agen x (P. salicina & P. americana) |
| Garnet syn for Sultan | 1898 | Wickson x Satsuma |
| Gaviota | 1900 | P. triflora x P. americana |
| Geewhiz | 1911 | P. triflora x P. americana |
| Giant | 1893 | d'Agen x Pond |
| Giant Maritima | 1905 | beach plum x P. triflora hybr |
| Gigantic | 1914 | unknown |
| Globe | 1914 | P. triflora |
| Glow | 1911 | P. maritima x P. americana |
| Golden | 1893 | Robinson x Abundance (P. triflora x P. munsoniana) |
| Goldridge | 1926 | P. triflora (possibly Wickson x Climax) |
| Grand Prize | 1937 | Posthumous Patent |
| Great Yellow | 1931 | Posthumous USPTO #13 |
| Hale aka Late Blood* | 1893 | Kelsey x Satsuma |
| Heikes possibly Satsuma* | 1885 | P. triflora Direct Import |
| Hermosillo | 1906 | unknown |
| Highland* | 1897 | d'Agen x OP |
| Home Chestnut | 1915 | unknown |
| Honey Prune | 1894 | F1 d'Agen x OP |
| HTS 84,761* | 1902 | unknown |
| Hunn* | 1897 | P. triflora |
| Hybrid Plum No. 38,674 | 1899 | unknown |
| Improved French | | see Morganhill |
| Inca | 1919 | unknown |
| Japanese Plum Seedling* | 1893 | Kelsey x Satsuma |
| Japex* | 1893 | P. triflora |
| Jordan | 1914 | P. triflora |
| Juicy | 1893 | Robinson x Abundance (P. triflora x P. munsoniana) |
| July Fourth * | 1900 | P. domestica? x (P. triflora x P. americana) |
| June Redskin | 1934 | Posthumous USPTO #12 |
| Late Conical* | 1898 | P. triflora x Simon |
| Late Shipper | 1914 | P. simonii x P. triflora |
| Leib Sour* | 1901 | P. triflora x P. simonii |
| Leopard Prune | 1926 | P. domestica |
| Lieb | 1914 | Burbank, Satsuma, et al |
| Long Fruit | 1886 | P. triflora |
| Long Leaf Wonderful* | 1893 | P. domestica |
| Madeira | 1906 | unknown |
| Mammoth Cardinal | 1934 | Posthumous USPTO #16 |

| Burbank Plum & Plumcot List | Year of Intro | Suspected Parentage |
| --- | --- | --- |
| Maritima | 1899 | P. maritima x OP |
| Marketman | 1893 | |
| Masu (Maru) | 1885 | P. triflora Direct Import |
| Maynard | 1897 | P. triflora x P. simonii |
| McKevitt | 1926 | unknown |
| Midsummer | 1926 | unknown |
| Miracle | 1901 | F2 or F3 of Sans Noyau x d'Agen |
| Morganhill | 1908 | d'Agen x OP |
| Nikko* | 1898 | unknown |
| Nixie | 1911 | P. subcordata |
| Occident | 1899 | Wickson x Satsuma |
| October (Purple fruit) | 1892 | Satsuma x P. triflora |
| Odd | 1904 | fruit print |
| Othello | 1914 | P. pissardii x OP |
| Pasha | 1897 | P. triflora x OP |
| Peach | 1901 | P. maritima x OP |
| Pearl | 1898 | d'Agen x OP |
| Perfection | | see Wickson |
| President* | 1899 | Wickson x OP |
| Pride | 1908 | P. maritima x Combination |
| Prize | 1911 | Burbank x Satsuma |
| Prolific | | see Hale |
| Purple-leafed Hybrid K. P. 193 | 1893 | Kelsey x P. pissardi |
| Rajah | 1926 | P. domestica |
| Rice Seed (Gaviota) | | see Gaviota |
| Royal | 1898 | Simon x Abundance |
| Rubio | 1909 | unknown |
| Rutland Plumcot | 1905 | (P. triflora x P. armeniaca) |
| Sachem | 1919 | P. domestica |
| Santa Rosa | 1906 | P. triflora, P. simonii, P. americana |
| Satsuma | 1886 | P. triflora Direct Import |
| Sea-Egg* | 1906 | P. triflora |
| Shipper | | see Marketman |
| Shiro | 1899 | Robinson, myrobalan, Wickson |
| Silver Plumcot | 1919 | unknown |
| Sky-blue | 1926 | unknown |
| Sonoma | 1926 | P. triflora x OP |
| Splendor Prune | 1886 | Pond x d'Agen |
| Standard Prune | 1911 | Tragedy x Sugar |
| Sugar Prune | 1899 | d'Agen x OP |
| Sultan | | see Occident |
| Sweet Botan | | see Berkmans |
| Sweet Plumcot | 1914 | unknown |
| Three-string | 1914 | P. triflora |

| Burbank Plum & Plumcot List | Year of Intro | Suspected Parentage |
|---|---|---|
| Toyland | 1927 | myrobalan |
| Triumph Plumcot | 1911 | unknown |
| Turkey Egg | 1914 | P. domestica |
| Valleda | 1919 | unknown |
| Vesta | 1911 | unknown |
| Vesuvius | 1907 | P. pissardii x P. triflora |
| Victory | 1911 | P. munsoniana (America x OP) |
| Vulcan* | 1899 | Wickson x OP |
| Wickson | 1892 | Burbank x Simon |
| Zulu | 1916 | unknown |

Appendix 2:  Five Principal Components reported for 53 *Prunus* taxa introduced or used by Luther Burbank in his breeding experiments based on Genotyping by Sequencing (GBS) data, generated in TASSEL5.

| Taxa and Location | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| Abundance_WEO DPRU.919 | 11.684 | -0.019 | -3.439 | 3.942 | -2.637 |
| Anderson's_Early_Red_WEO DPRU.843 | -7.021 | 9.308 | 18.867 | -0.329 | -1.563 |
| Apex_GR | 16.302 | -3.034 | -9.4 | 5.206 | 4.734 |
| Apex_WEO DPRU.1170 | 6.241 | 3.53 | 4.447 | 5.111 | 0.769 |
| Ballena_GR | 9.617 | 0.689 | -2.198 | -4.296 | -5.426 |
| Beauty_LBHG | 6.84 | 3.129 | 5.616 | 6.603 | 1.213 |
| Beauty_SW_LBHG | 6.577 | 2.082 | 2.632 | 3.661 | 0.558 |
| Beauty_WEO DPRU.2120 | -5.502 | 7.288 | 14.373 | 6.189 | 3.425 |
| Botanky_WEO DPRU.372 | 17.631 | -3.087 | -10.541 | 13.585 | -9.397 |
| Brookgold_WEO DPRU.1736 | -16.339 | 11.654 | 25.197 | 3.66 | -0.118 |
| Burbank_Plumcot_GR | -23.132 | 4.614 | -10.415 | -0.323 | -0.307 |
| Burbank_(Plumcot)_WEO DPRU.936 | -0.097 | 5.92 | 10.409 | -3.671 | 4.282 |
| Catherine_Bunnell_LBHG | 7.759 | 1.903 | 1.385 | 0.029 | -4.841 |
| Chalco_WEO DPRU.431 | -2.117 | -16.465 | 8.985 | 4.192 | 3.704 |
| *P. simonii*_WEO DPRU.2430 | 17.595 | -3.031 | -10.513 | 13.419 | -9.492 |
| El_Dorado_GR | 12.192 | 1.226 | -2.056 | -15.034 | -11.988 |
| El_Dorado_LBHG | 9.095 | 1.219 | -1.074 | -10.539 | 17.689 |
| El_Dorado_SR_LBHG | 8.745 | 1.309 | -0.885 | -10.499 | 16.879 |
| El_Dorado_WEO DPRU.2122 | -30.521 | 5.049 | -15.169 | -0.79 | 4.001 |
| Elephant_Heart_GR | 11.834 | 1.094 | -2.034 | -13.84 | -10.975 |
| Elephant_Heart_WEO DPRU.2123 | 9.612 | 2.262 | 0.635 | -12.919 | 6.284 |
| Formosa_GR | 13.357 | -1.781 | -7.439 | -7.095 | 4.89 |
| Formosa_LBHG | 17.085 | -3.276 | -9.72 | 4.99 | 4.813 |
| Formosa?_WEO DPRU.924 | 6.93 | -12.512 | 0.101 | 8.561 | 6.122 |

| Taxa and Location | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| French_WEO DPRU.436 | -33.684 | 5.785 | -13.52 | 1.9 | 1.352 |
| Grand_Prize_WEO DPRU.1572 | -23.674 | 4.015 | -13.806 | 0.208 | 0.534 |
| Great_Yellow_SpSa_LBHG | 2.151 | 3.334 | 5.238 | -2.421 | -1.457 |
| Great_Yellow_WEO DPRU.2105 | 6.198 | 4.279 | 5.972 | 2.258 | 1.857 |
| Improved_Satsuma_LBHG | 12.315 | 1.495 | -1.854 | -15.271 | -12.333 |
| June_Red_LBHG | -2.151 | 5.194 | 9.605 | 3.003 | -2.806 |
| Late_Goose_WEO DPRU.546 | -18.477 | -50.892 | 9.355 | -2.32 | -2.866 |
| Latest_of_All_GR | 4.092 | 4.864 | 8.268 | -2.354 | -2.914 |
| Latest_of_All_WEO DPRU.427 | 1.358 | 4.861 | 8.992 | 4.929 | 2.777 |
| Lieb_LBHG | 4.372 | 2.711 | 3.081 | -6.807 | 1.013 |
| Mammoth_Cardinal_LBHG | 17.345 | -3.373 | -9.744 | 6.34 | 4.805 |
| Mammoth_Cardinal_WEO DPRU.2127 | 18.74 | -3.759 | -10.85 | 7.124 | 6.077 |
| OG_Stoneless_WEO DPRU.2302 | 2.706 | 1.255 | -10.501 | 6.61 | -9.008 |
| Pcot_Edibles_LBHG | -18.586 | 3.494 | -8.713 | -0.352 | 0.609 |
| Perfection_WEO DPRU.1720 | -16.598 | 12.067 | 25.982 | 3.866 | 0.396 |
| Rutland_GR | 8.685 | -5.072 | -5.725 | -9.42 | 6.286 |
| Sans_Noyau_WEO DPRU.2419 | -6.789 | 3.716 | -7.971 | -10.305 | -10.617 |
| Santa_Rosa_GR | 7.914 | 2.017 | 1.192 | 7.186 | -7.296 |
| Satsuma_WEO DPRU.438 | 17.989 | -3.538 | -10.048 | 4.017 | 6.625 |
| Shiro_GR | 1.877 | 3.217 | 5.376 | -1.614 | -1.527 |
| Shiro_WEO | -8.319 | 6.427 | 4.706 | -1.151 | -1.852 |
| Simon_WEO DPRU.545 | -31.383 | 5.335 | -14.471 | 0.959 | 0.68 |
| Sultan_LBHG | 6.44 | 1.983 | 2.725 | -8.44 | 0.99 |
| Top_of_the_Hill_GR | -19.634 | 3.633 | -9.252 | -0.278 | 0.157 |
| Unk_7_GR | -17.225 | -49.796 | 9.156 | -2.016 | -1.788 |
| Unk_Multi_GR | 3.994 | 4.631 | 7.993 | -2.756 | -3.178 |
| Vesuvius_WEO DPRU.2108 | -29.904 | 5.105 | -12.722 | 2.054 | 0.231 |
| Victory_WEO DPRU.791 | 2.651 | 3.698 | 7.76 | 7.993 | 3.593 |
| Wickson_WEO DPRU.2135 | 3.237 | 4.239 | 6.012 | 7.248 | -3.507 |

Appendix 3: Admixture cluster group Q matrix for K=4 generated in STRUCTURE with a burn-in period of 10,000 and 5 reps for 53 taxa of Burbank-introduced *Prunus* and 22,872 loci.

| Accession Name & Location | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Abundance_WEO DPRU.919 | 0.167 | 0.005 | 0.828 | 0 |
| Anderson's_Early_Red_WEO DPRU.843 | 0.753 | 0 | 0.247 | 0 |
| Apex_GR | 0.004 | 0.006 | 0.99 | 0 |
| Apex_WEO DPRU.1170 | 0.373 | 0 | 0.627 | 0 |
| Ballena_GR | 0.183 | 0.002 | 0.815 | 0 |
| Beauty_LBHG | 0.352 | 0 | 0.648 | 0 |
| Beauty_SW_LBHG | 0.291 | 0 | 0.709 | 0 |
| Beauty_WEO DPRU.2120 | 0.687 | 0.004 | 0.309 | 0 |
| Botanky_WEO DPRU.372 | 0 | 0 | 1 | 0 |
| Brookgold_WEO DPRU.1736 | 1 | 0 | 0 | 0 |
| Burbank_Plumcot_GR | 0.027 | 0.86 | 0.113 | 0 |
| Burbank (Plumcot) WEO DPRU.936 | 0.534 | 0 | 0.466 | 0 |
| Catherine_Bunnell_LBHG | 0.29 | 0.001 | 0.709 | 0 |
| Chalco_WEO DPRU.431 | 0.252 | 0 | 0.4 | 0.348 |
| *P. simonii* DPRU_2430_WEO | 0 | 0 | 1 | 0 |
| El_Dorado_GR | 0.155 | 0 | 0.845 | 0 |
| El_Dorado_LBHG | 0.234 | 0 | 0.766 | 0 |
| El_Dorado_SR_LBHG | 0.234 | 0 | 0.766 | 0 |
| El_Dorado_WEO DPRU.2122 | 0.002 | 0.978 | 0.02 | 0 |
| Elephant_Heart_GR | 0.167 | 0 | 0.833 | 0 |
| Elephant_Heart_WEO DPRU.2123 | 0.274 | 0 | 0.726 | 0 |
| Formosa?_WEO DPRU.924 | 0.149 | 0 | 0.622 | 0.228 |
| Formosa_GR | 0.029 | 0.021 | 0.95 | 0 |
| Formosa_LBHG | 0 | 0 | 1 | 0 |
| French_WEO DPRU.436 | 0.001 | 0.999 | 0 | 0 |
| Grand_Prize_WEO DPRU.1572 | 0 | 0.862 | 0.138 | 0 |
| Great_Yellow_SpSa_LBHG | 0.414 | 0 | 0.586 | 0 |
| Great_Yellow_WEO DPRU.2105 | 0.384 | 0 | 0.616 | 0 |
| Improved_Satsuma_LBHG | 0.163 | 0 | 0.837 | 0 |
| June_Red_LBHG | 0.546 | 0 | 0.454 | 0 |
| Late_Goose_WEO DPRU.546 | 0 | 0 | 0 | 1 |
| Latest_of_All_GR | 0.436 | 0 | 0.564 | 0 |
| Latest_of_All_WEO DPRU.427 | 0.489 | 0 | 0.511 | 0 |
| Lieb_LBHG | 0.342 | 0.001 | 0.657 | 0 |
| Mammoth_Cardinal_LBHG | 0 | 0 | 1 | 0 |
| Mammoth_Cardinal_WEO DPRU.2127 | 0 | 0 | 1 | 0 |
| OG_Stoneless_WEO DPRU.2302 | 0.025 | 0.306 | 0.669 | 0 |
| Pcot_Edibles_LBHG | 0.043 | 0.757 | 0.2 | 0 |
| Perfection_WEO DPRU.1720 | 1 | 0 | 0 | 0 |

| | | | | |
|---|---|---|---|---|
| Rutland_GR | 0.039 | 0.04 | 0.853 | 0.068 |
| Sans_Noyau_WEO DPRU.2419 | 0.122 | 0.422 | 0.456 | 0 |
| Santa_Rosa_GR | 0.293 | 0 | 0.707 | 0 |
| Satsuma_WEO DPRU.438 | 0 | 0 | 1 | 0 |
| Shiro_GR | 0.425 | 0 | 0.575 | 0 |
| Shiro_WEO DPRU.2132 | 0.41 | 0.266 | 0.324 | 0 |
| Simon_WEO DPRU.545 | 0.001 | 0.999 | 0 | 0 |
| Sultan_LBHG | 0.319 | 0 | 0.681 | 0 |
| Top_of_the_Hill_GR | 0.028 | 0.763 | 0.209 | 0 |
| Unk_7_GR | 0.001 | 0 | 0.025 | 0.974 |
| Unk_Multi_GR | 0.431 | 0 | 0.569 | 0 |
| Vesuvius_WEO DPRU.2108 | 0.021 | 0.979 | 0 | 0 |
| Victory_WEO DPRU.791 | 0.473 | 0 | 0.527 | 0 |
| Wickson_WEO DPRU.2135 | 0.458 | 0 | 0.542 | 0 |

# Chapter 3: Analysis of Luther Burbank's Plum (*Prunus* sp) Introductions using Genotyping by Sequencing (GBS), Genome-wide Association Studies (GWAS) and Integrated Haplotype Scores (iHS) for Phenotypic Traits Related to Fruit Marketability

## Abstract:

This study uses single nucleotide polymorphisms (SNPs) retrieved through genotyping by sequencing (GBS) to perform genome-wide association studies (GWAS) in an historic breeding population of Luther Burbank's inter- and intra-specific *Prunus* introductions. Phenotypic traits related to fruit marketability such as exocarp color, mesocarp color, free or cling-stone endocarps, and general shape were compared to SNPs. The most notable association was a SNP on chromosome 2 associated with fruit exocarp color, which is consistent with other studies looking at anthocyanin production in fruit skin color. Integrated haplotype scores (iHS) were calculated for each SNP to find evidence of positive and negative selection. Statistically significant iHS SNPs were not the same as the significant GWAS SNPs, indicating more research is needed to interpret the function of these signals.

**Introduction**:

The genus *Prunus* includes several commercially important crops such as peaches (*Prunus persica*), almonds (*Prunus dulcis*), cherries (*Prunus avium*), and apricots (*Prunus armeniaca*), prunes (*Prunus domestica*), and plums (*Prunus* spp). These species have been phylogenetically broken up into four main subgenera based on flower inflorescence structure and transcriptome data (Hodel et al. 2021). Species in these subgenera have high degrees of cross-compatibility within each group, and limited cross-compatibility between groups. Species with the highest degrees of cross-compatibility are found in the *Prunus* subgenus, which includes peach, almond, apricot, and plums. Plant breeders such as Luther Burbank (1849-1926) and Floyd Zaiger (1926-2020) utilized many species in the *Prunus* subgenus to create a plethora of morphological variation, including novelties like the plumcot, pluot, aprium, nectaplum, and pluerry.

Plums are fruits enjoyed by humans in fresh and dried forms globally. This fruit grossed around $124 million in the United States in 2020 (Iowa State University 2021). The term "plum" refers to a kind of fruit that encompasses between 19 and 40 different species depending on which taxonomic treatment is followed (Topp et al. 2012). Botanically speaking, a plum is a smooth-skinned drupe with an oblong seed. In general, plum trees are not very long-lived, often surviving for a mere 15-20 years (UC IPM 2017). Their growth form can be either upright or weeping. Plants take around 6 years to bear fruit if grown from seed. Plums vary in their self-compatibility with most fruit sets benefiting from or requiring outcrossing. They are typically propagated clonally through grafting for cultivar preservation. However, some cultivars produced through

wide crosses tend to throw bud sports that appear to be somatic mutations in meristematic cells (Foster and Aranzana 2018) leading to dramatic shifts in phenotypes such as the degree of response to ethylene for fruit ripening (Minas et al. 2015).

Often the mesocarp is sweet and pulpy and the skin is tart, though the range of flavonoids present varies dramatically depending on a cultivar's complex parentage (Gomez and Ledbetter 1994). These fruits have been popular for consumption by humans for thousands of years. The most common species for fresh eating are the diploid Japanese plums (*Prunus salicina*) while the dried fruits, typically marketed as prunes, are hexaploid European plums (*Prunus domestica*). Japanese cultivars tend to be less cold-hardy and have a lower chill-hours requirement for fruit set than their European prune relatives (UC IPM 2017), making them ideal for California cultivation.

*Prunus* species have eight chromosomes. Most of the *Prunus* species are diploid, with a few exceptions (Hodel et al. 2021). Currently the reference genome information for the *Prunus* genus is highly limited with only seven of the 250-400 species: apricot (*P. armeniaca*); sweet cherry (*P. avium*), almond (*P. dulcis*), Chinese plum (*P. mume*), peach (*P. persica*), Japanese plum (*P. salicina*), and Yoshino cherry (*P. yedoensis*), with *P. persica* being the most well studied genetically (Verde et al. 2012, Minas et al. 2015, Salazar et al. 2017, Carresco et al. 2018, Marti 2018, Salazar et al. 2018, Aranzana et al. 2019, Zhebentyayeva et al. 2019, Hodel et al. 2021). The recent availability of a fully annotated *P. salicina* genome represents a major advancement in the genomic characterization of many commercially and historically valuable hybrid fresh-eating plums as a means of identifying single nucleotide

polymorphisms (SNPs) that would not otherwise be discovered using the more distantly related *P. persica* genome  (Liu et al. 2020).

In any given stretch of DNA, SNPs occur as single base-pair changes (Bush and Moore 2012).  Because of the redundancy of DNA, sometimes these base pair changes do not change the functionality of the amino acids sequenced.  Other times, these SNPs change amino acids which in turn has significant effects on protein folding, altering the form and function of that protein.  Identifying the location of these SNPs can be incredibly powerful for pinpointing genetic bases for diverse phenotypes.

To uncover the association between the genotype and phenotype, one must first start with Linkage Disequilibrium (LD).  LD can be defined as a non-random association of alleles between loci either connected or broken apart through recombination (Kim et al. 2007; Abdurakhmonov and Abdukarimov 2008, Bush and Moore 2012).  Some of these loci in LD assort together in a predictable pattern instead of independently.  When blocks of SNPs are inherited together with no recombination occurring between them, they are referred to as haploblocks (Ge et al. 2010).  Finding loci in LD helps to define haploblocks, discover Quantitative Trait Loci (QTL), and conduct Genome wide Association Studies (GWAS), all tools that are essential for Marker Assisted Selection (MAS) in a breeding program.  Analyzing LD in controlled crosses for segregating populations has been successful at correlating QTLs in *P. persica* to traits like general shape (Tan et al. 2021), fruit quality (da Silva Linge et al. 2021), bloom date, ripening date, and fruit development period (Rawandoozi et al. 2021).  Similarly, fine mapping of candidate genes in *P. persica* and *P. avium* has revealed QTLs for traits like clingstone vs freestone (Peace et al., 2005), green vs purple leaves, flesh and skin color

(Sooriyapathirana et al., 2010, Bretó et al., 2017, Zhao et al., 2020), and fruit shape

(Aranzana et al. 2019). This has also been the case for correlating exocarp (skin) color

in Japanese Plum (*Prunus salicina* Lindl.) to a specific gene (Fiol et al. 2021).

While bi-parental crosses are important to demonstrate the biological basis for

the effect a QTL has on a subset of a population, using a multi-parental population

(MPP) instead of a bi-parental cross for SNP discovery is advantageous because it

eliminates any influences from population structure (Bahr et al. 2020, Scott et al. 2020)

and expands the total number of SNPs that can be detected. This is because tightly

controlled population structure with non-random mating eventually leads to higher

degrees of homozygosity, and ultimately linkage decay as recombination events break

up linked alleles in a contiguous chromosome (Bush and Moore 2012). Therefore,

using the LD of SNPs instead of QTLs to inform GWAS is especially useful in natural

populations (Cheng et al. 2013), curated wild collections (Cao et al. 2016, Guajardo et

al. 2020, Tan et al. 2021), and MPPs where the specific pedigree is unclear or unknown

(Navarro et al. 2020, Scott et al. 2020) for traits like fruit phenotypes related to

marketability (Zahid et al. 2022)

While GWAS and QTLs are powerful tools for locating candidate genes affiliated

with specific traits, they fall short when looking for regions of the genome undergoing

active selection in either breeding systems or natural populations. Haploblocking is a

useful method for detecting areas with extended haplotype homozygosity (EHH). The

decay of EHH can be utilized to generate integrated haplotype scores (iHS) (Gautier

and Vitalis 2012) to find evidence of positive selection, domestication, improvement,

and negative selection based on the extreme frequencies (high or low) of novel alleles

66

in a population (Ma et al. 2015, Zhao et al. 2022).  Longer haplotypes typically indicate

areas of improvement or positive artificial selection.  Shorter haplotypes indicate

diversifying or negative artificial selection.

This study uses GBS to compare the genomes of over fifty accessions from an

historic breeding population of *Prunus* introduced by Luther Burbank almost a century

ago and uses LD, GWAS, and iHS to compare genomic data with several traits related

to marketability including exocarp color, mesocarp color, free or cling-stone endocarp,

and general shape (Table 3).

## **Materials and Methods:**

Specimen Collection:

A comprehensive list of Burbank's plum introductions is readily available

(Howard, 1945).  However, by going back into the primary source material used by

Howard to generate this list, the names of six more plums that were introduced by

Luther emerged (Hedrick, 1904).  Also missing from this list are the taxa that were

patented after Burbank's death by his widow Elizabeth in collaboration with the Stark

Brothers.  This target list was used to hunt for specific cultivars through ARS repository

access, word of mouth, and local CRFG scion exchanges.  Once a cultivar was located,

old literature was searched for claims Luther Burbank made about their parentage as

well as any historical images that may accompany them (Burbank 1914, Hedrick 1911,

Howard 1945, Brooks and Olmo 1952).  Scions of material found at exchanges or

through word-of-mouth were multi-grafted onto mature trees with *Prunus cerasifera*

'Myrobalan 29C' as a universal rootstock at the Luther Burbank Home & Gardens

(LBHG) in Santa Rosa, California. Historic maps were referenced for cultivar names at the Luther Burbank Goldridge Farm (GR) in Sebastopol California. The GRIN Global Database and map of the USDA-ARS-National Clonal Germplasm Repository's Wolfskill Experimental Orchard (WEO) plum block were used for locating cultivars in their collection.

Tissue Preparation and Sequencing Methods:

Young leaf tissue was collected from trees in the early spring and stored in silica gel at room temperature until sufficiently dry, then frozen at -80ºC. The DNA extraction protocol followed DNeasy Plant Kit from Qiagin. Retrieved sequences were aligned to *P. salicina* 'Sanyueli' (Liu et al. 2020). Samples with more than 90% missing data were discarded, and this resulted in discarding 28/96 taxa. Approximately 50,000 SNPs were retrieved. All genotypes were set to a minimum depth of <5 missing, and then SNPs with more than 50% missing were discarded using TASSEL5. The final imputed dataset contained 24,147 SNPs. Missing genotypes were imputed using Beagle (Browning et al. 2018).

Linkage Decay:

Linkage decay was investigated using the '*Linkage Disequilibrium'* analysis function in TASSEL 5 using default parameters. The LD window size was set to 50. The resulting data was exported to R for further analysis. The mean and median values of $R^2$ were estimated for 100,000 basepair bins along the chromosomes. The original $R^2$ values, the mean $R^2$ values, and the median $R^2$ values were visualized using the R packages *'ggplot2'* and *'ggthemes'* (Wickham 2016, Arnold 2021)*.*

<u>Genome-wide Association Analysis (GWAS):</u>

Genome-wide association analyses were performed using the R package *'rrBLUP'* (Endelman 2011) for four phenotypes: Excocarp color, Mesocarp color, Cling versus Free Stone, and Shape. Color phenotypes were coded numerically for this analysis, from dark to light colors. A genomic relationship matrix (G) was first estimated using the *A.mat()* function. A QK-type GWAS (Yu et al. 2006) was run for each GWAS, using the G matrix and 3 principal components. The genomic inflation factors ($\lambda$) were calculated to be less than 1, and the quantile-quantile (QQ) plots showed no evidence of significant under- or over-inflation. Significance was determined using the Bonferroni-corrected and false discovery rate (FDR) corrected thresholds ($p = 0.05$).

<u>Integrated Haplotype Scores (iHS):</u>

Integrated haplotype scores (iHS) were calculated for each SNP using the *rehh* package (Gautier & Vitalis 2012, Appendix 2) and visualized in R.

## **Results:**

<u>Linkage Disequilibrium:</u>

Linkage disequilibrium (LD) was calculated for each chromosome. In most cases Generalized Additive Model (GAM) curves tended to decay within the first 150 kilobases of each chromosome which is consistent with inbred populations that have reached genetic bottlenecks in their life history (Figure 1). Occasional increases in LD($r^2$) values, common in outcrossing species, were also observed. In this study system, there are factors driving an increase in LD such as relatedness among the taxa and artificial

selection, while simultaneously having some attributes that decrease LD, namely an outcrossing breeding system with frequent wide crosses being made (Abdurakhmonov and Abdukarimov 2008, Cheng et al. 2013).  Because of this it is important to look closely at regions throughout each chromosome that are in LD rather than examining a genome-wide sweep of LD.
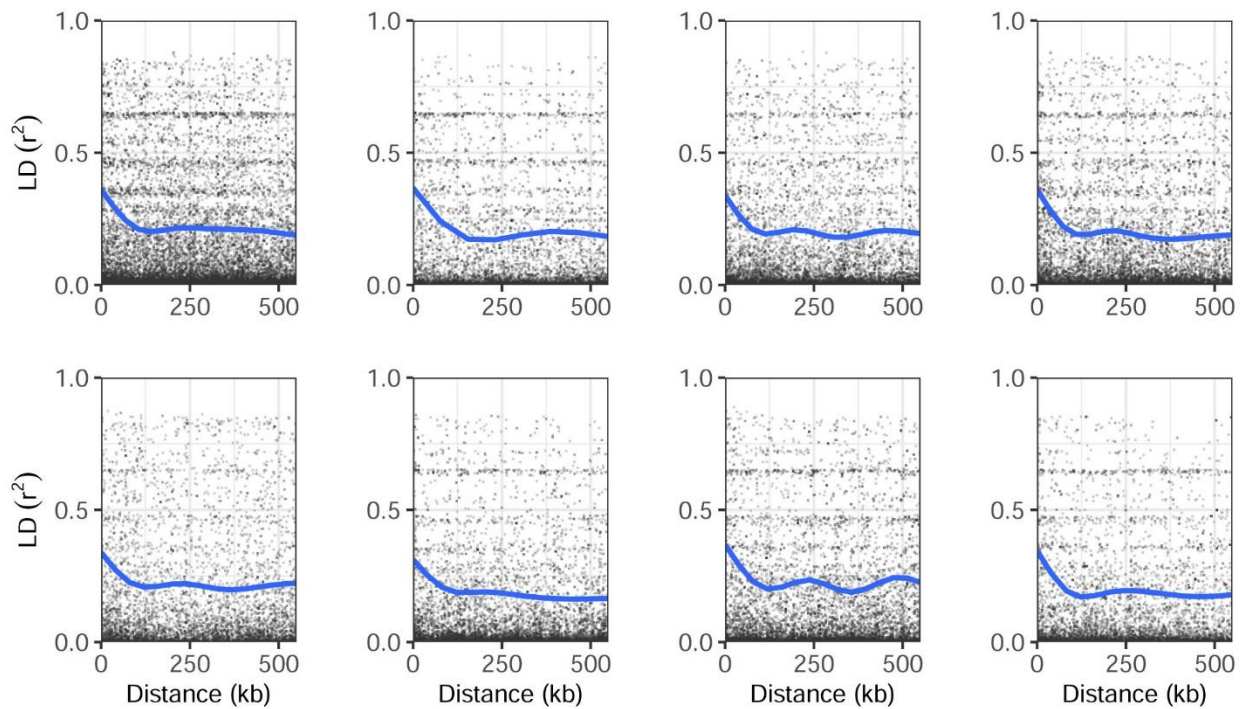


**Fig 1.** Linkage disequilibrium (LD) maps for the eight chromosomes found in 53 *Prunus* taxa of a Burbank breeding population with a Generalized Additive Model (GAM) curve in blue.  Top left is chromosome number 1.  Bottom right is chromosome number 8.

Mean and median Linkage Disequilibrium (LD) values were plotted against genetic distance (kilobases) for each chromosome (Figure 2).  Mean and median values were included to show overall trends in LD decay.  Maximum decay occurred between 125kbp and 200kbp for both the mean and the median LD ($r^2$) values.  Using a minor allele frequency (MAF) cutoff of 5% to reduce the inflation effect rare minor alleles have on LD was necessary because there were multiple founder species in this population.
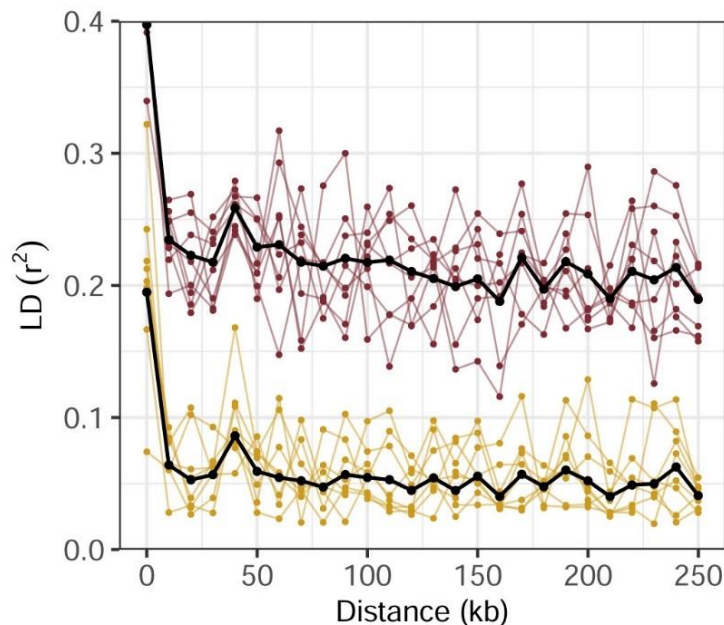


**Fig 2**.  Mean (maroon, top) and median (gold, bottom) values for LD across genetic distance in a Burbank breeding population of 53 *Prunus* taxa showed steep decay within 15-20kb.

Phenotyping:

Fruits were collected from all three locations to score discrete phenotypic traits related to marketability including exocarp color, mesocarp color, free or cling-stone endocarps, and general shape (Figures 3, 4, and 5; Appendix 1).
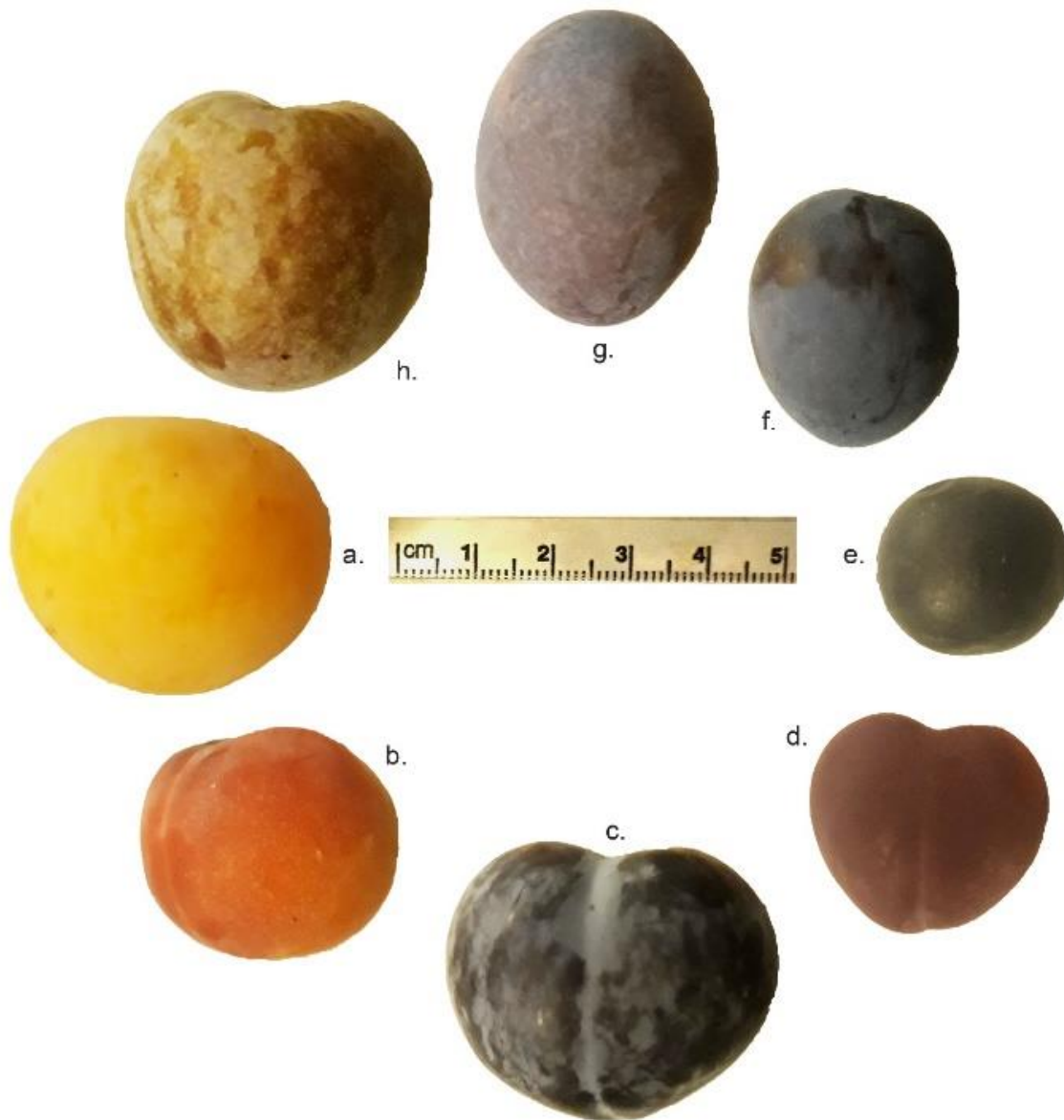
**Fig 3**. Exocarp color (yellow, red, black, purple, and blue) and overall shape (round, pointy, and oval) were scored for Burbank plum introductions. The cultivars shown are: a. 'Great Yellow DPRU.2105,' b. 'Abundance DPRU.919,' c. 'El Dorado DPRU.2122,' d. 'Satsuma DPRU.438,' e. 'Vesuvius DPRU.2108,' f. 'Sans Noyau DPRU.2419,' g. 'Grand Prize DPRU.1572,' and h. 'Elephant Heart DPRU.2123.' A summary of phenotypes can be found in Appendix 1.

**Fig 4**. The mesocarp colors of this Burbank *Prunus* collection are yellow (a, b, c), red (d, e, h), or green (f, g). The cultivars shown are: a. 'Great Yellow DPRU.2105,' b. 'Abundance DPRU.919,' c. 'El Dorado DPRU.2122,' d. 'Satsuma DPRU.438,' e. 'Vesuvius DPRU.2108,' f. 'Sans Noyau DPRU.2419,' g. 'Grand Prize DPRU.1572,' and h. 'Elephant Heart DPRU.2123.' A summary of phenotypes can be found in Appendix 1.
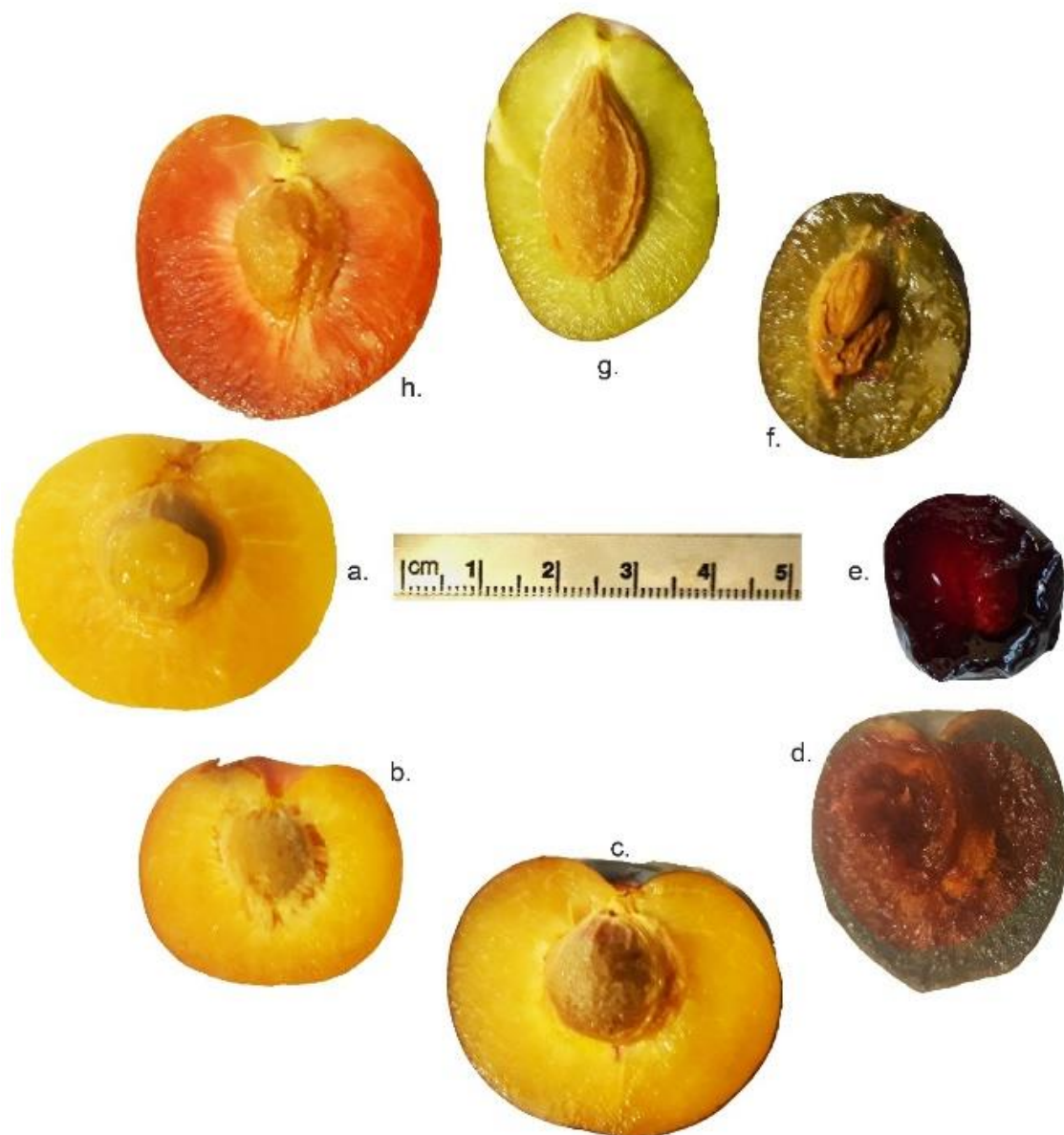
**Fig 5.** The endocarps found in Burbank *Prunus* collection are either free-stone (a, b, c, f, g,) or clingstone (d, e, h). Silhouettes of the fruits were included to look at the overall size of the fruits' mesocarp in relation to their endocarps. The cultivars shown are: a. 'Great Yellow DPRU.2105,' b. 'Abundance DPRU.919,' c. 'El Dorado DPRU.2122,' d. 'Satsuma DPRU.438,' e. 'Vesuvius DPRU.2108,' f. 'Sans Noyau DPRU.2419,' g. 'Grand Prize DPRU.1572,' and h. 'Elephant Heart DPRU.2123.' A summary of phenotypes can be found in Appendix 1.
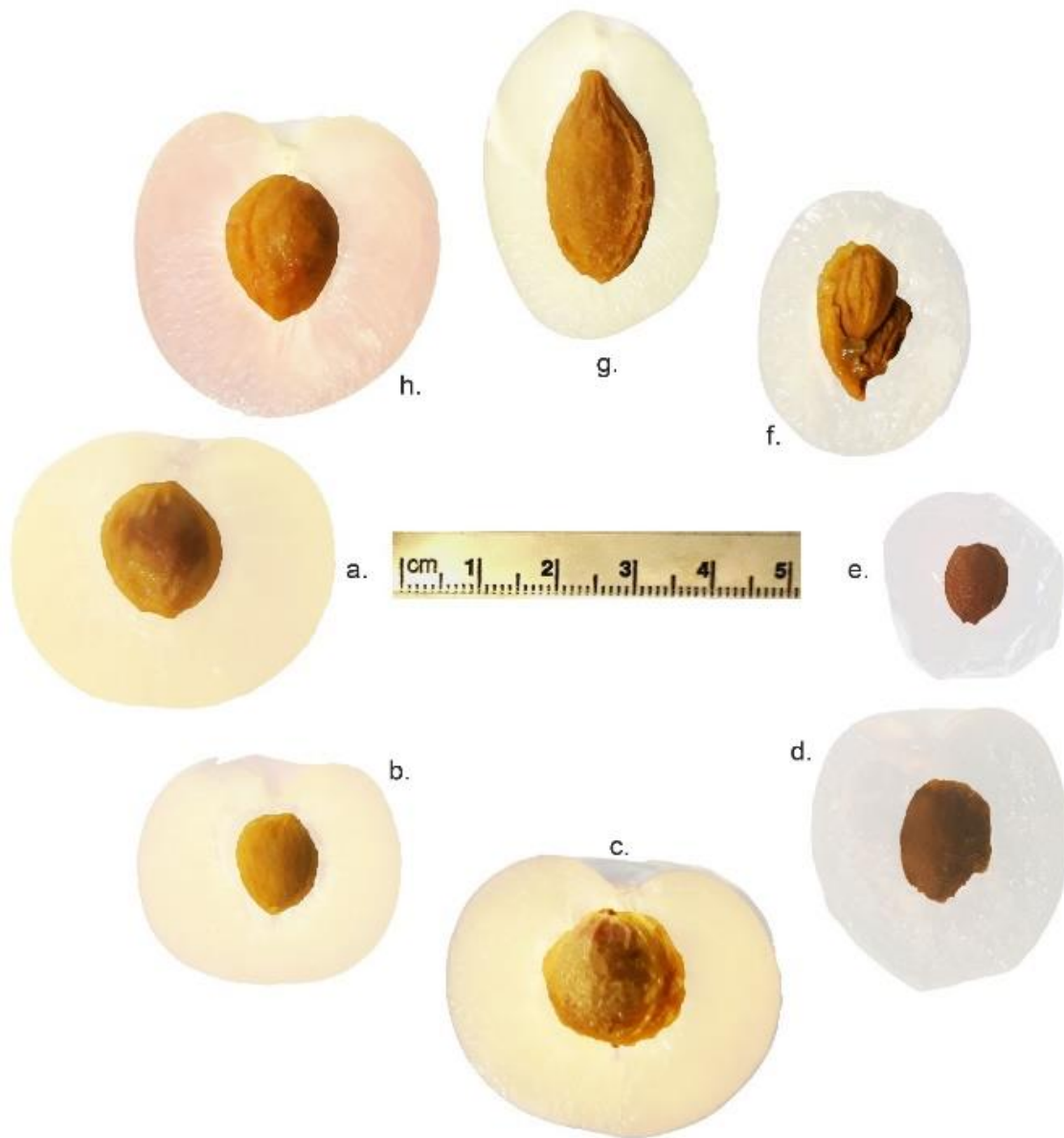
Phenotype by environment interaction is of concern for this population.  The taxa in this study system were grown at three locations in USDA hardiness zone 9b, which is characterized by having minimum temperatures around -3C, with rainy annual precipitation occurring in the winter months and no precipitation in the summer months. However, the study sites in Sebastopol, CA (GR) and Santa Rosa. CA (LBHG) have a marine influence due to their proximity to the ocean.  These locations have a fog bank that keeps night temperatures between 13C and 18C at night, and a maximum day temperature that is often 5C cooler than the study site at Winters, CA (WEO).  This has an influence on phenological traits like breaking dormancy, flowering time, fruit ripening time, and exterior fruit color.

In one particularly interesting case, *P. salicina* 'Wickson DPRU.2135'  plants at WEO often break dormancy, flower, and ripen 2-3 weeks earlier than those at GR and LBHG.  Plums of New York (Hedrick 1911) describes the fruit color of 'Wickson' as "dark red over a yellow ground, indistinctly splashed with darker red" (Figure 6). However, it was observed that this fruit is sweet and ripe when the exocarp is yellow at WEO and it rarely will acquire a pink blush at GR and LBHG.  Typically, when this fruit is marketed in stores, it is sold as a greenish-yellow plum picked slightly firm, but delightfully sweet.  If the fruit is refrigerated, its exocarp will turn red.

Temperature dependency of color pigmentation related to the anthocyanins responsible for red color in fruits has been observed in other fruits such as apple, grapes, persimmon, pomegranate, and tomato (Koshita 2014).  This is a conundrum when scoring fruit phenotypes.  A century ago, night temperatures were consistently cooler, leading to the red skin color historically depicted for 'Wickson.'  More work is

75

suggested to determine the temperature threshold and duration required for

anthocyanin development in this cultivar.  This phenomenon was not observed in the

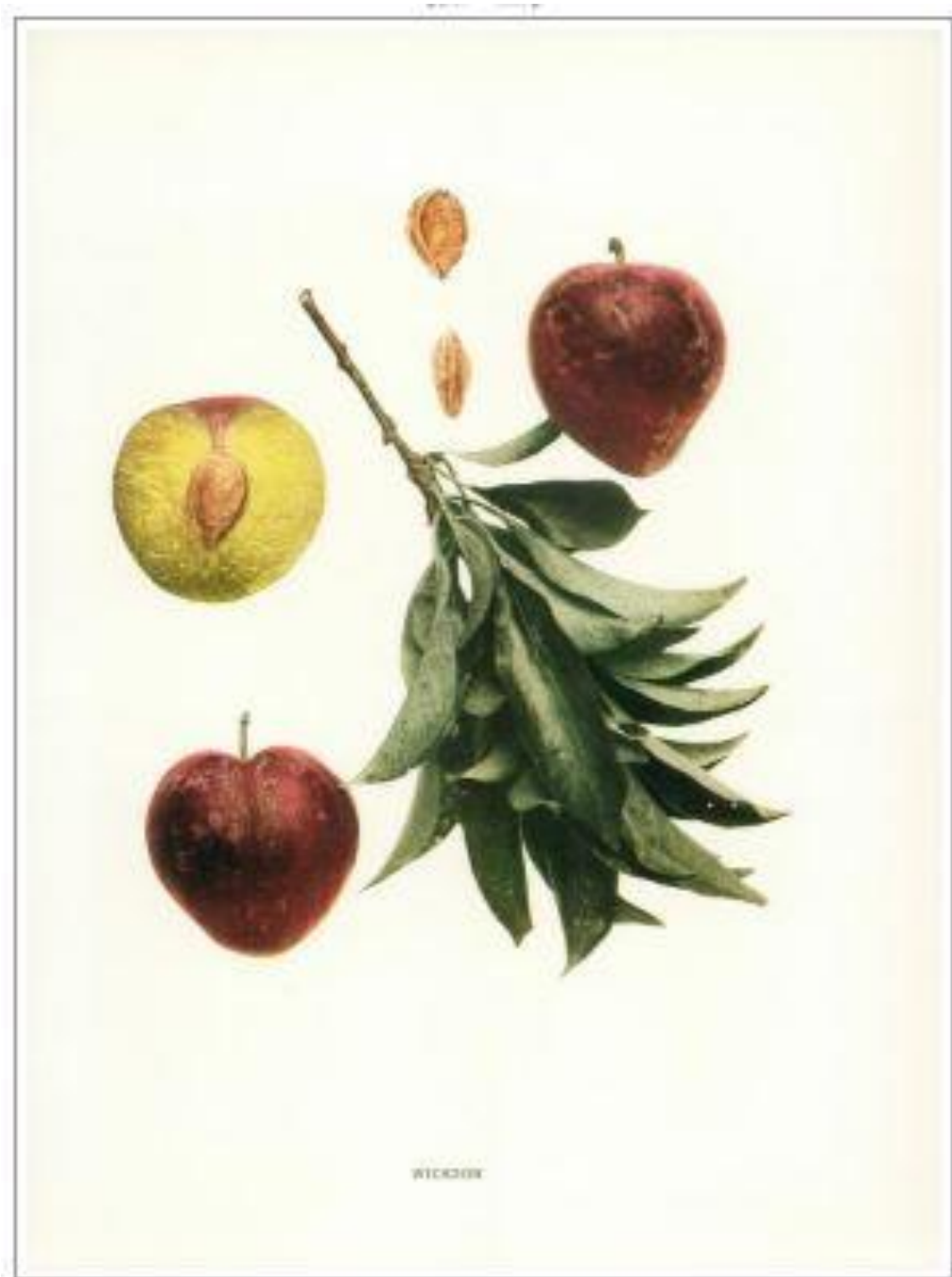other *Prunus* cultivars scored for this study.



**Fig. 6.** An historic watercolor illustration of *P. salicina* 'Wickson' from Plums of New York (Hedrick 1911) clearly exhibits a fruit with a dark red exocarp (skin).

Genome-wide association study (GWAS) was run for exocarp color, mesocarp color, free or cling-stone endocarps, and overall shape (round, pointy, or oval) using Bonferroni-adjusted p-values to control family-wise Type I error rate (Figure 6) and FDR to control for erroneously rejected null hypotheses among those rejected (Figure 7). Only one SNP exceeded the threshold of significance designated by these methods in both cases; a SNP on chromosome 2 at position 11,806,474 associated with exocarp color. Of these phenotypes, the strongest detectable correlation was observed with relation to exocarp color (red, yellow, purple, or blue), which was observed on chromosome 2 (Figure 7a, 8a). Mesocarp color (yellow, red, or green) showed some signal on chromosomes 2 and 3 (Figure 7b, 8b) though they did not exceed the threshold of significance set. Free or Cling-stone endocarps showed some signal on chromosome 7 (Figure 7c, 8c), and overall shape showed multigenic signals on chromosomes 1, 2, 3, 4, and 6 (Figure 7d, 8d).

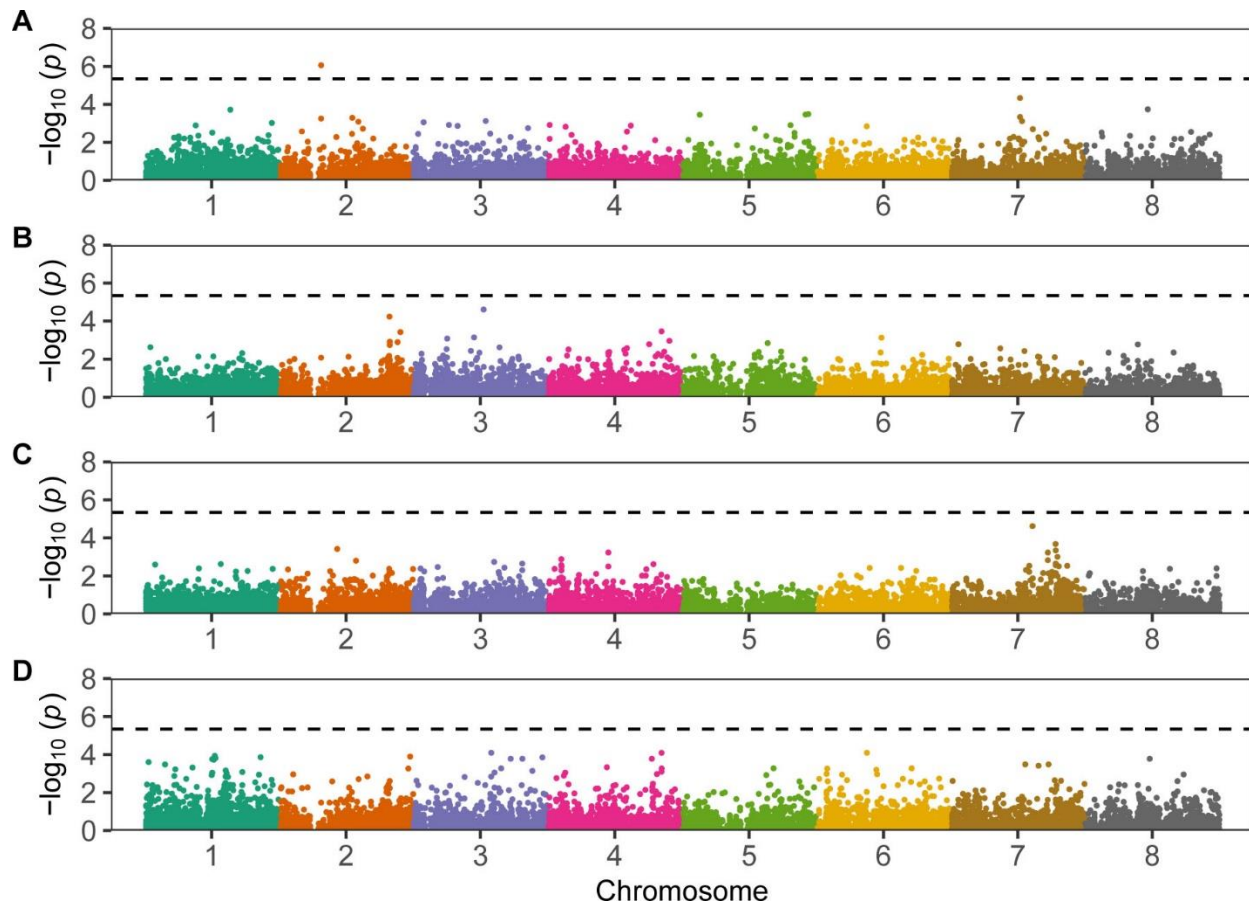**Fig. 7**. GWAS of phenotypes scored for 53 plum cultivars introduced or used in experiments by Luther Burbank using Bonferroni-corrected p-values. The dashed horizontal line at 5.34 represents the Bonferroni threshold of significance with alpha=0.05. Values above that threshold are SNPs correlated a phenotypic trait. These traits are (A) exocarp color; (B) mesocarp color; (C) endocarp adherence; and (D) general shape.
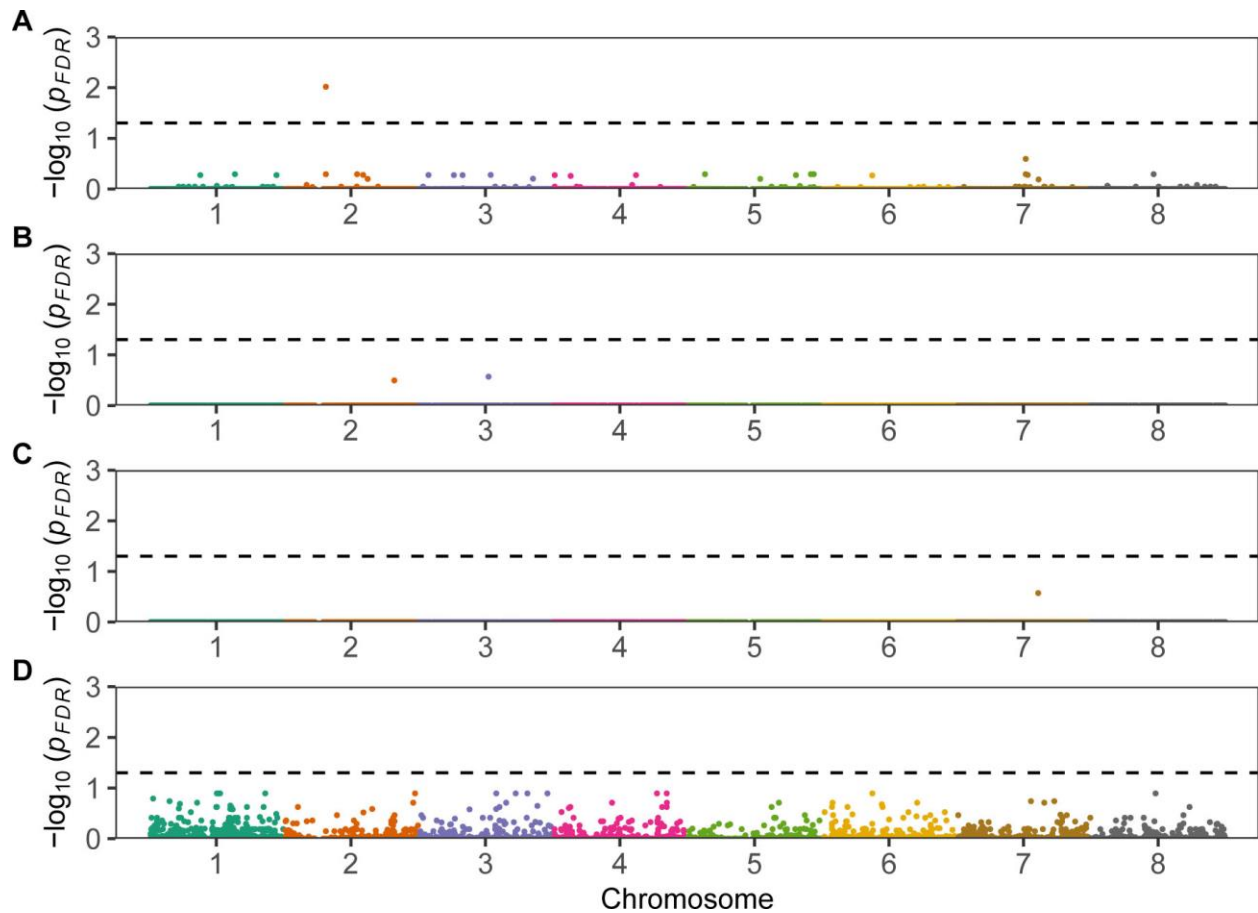
**Fig. 8.** GWAS of phenotypes scored for 53 plum cultivars introduced or used in experiments by Luther Burbank using False Discovery Rate (FDR)-corrected p-values. The dashed horizontal line at 1.30 represents the threshold of significance at alpha=0.05. These traits are (A) exocarp color; (B) mesocarp color; (C) endocarp adherence; and (D) general shape.

<u>iHS:</u>

Integrated haplotype scores (iHS) showed SNPs undergoing positive selection occurring on chromosomes 2, 4, 7, and 8 with most of them occurring on chromosome 7(Figure 9, Appendix 2). Evidence of alleles being selected against occurred on chromosomes 1, 2, 3, 5, and 6.
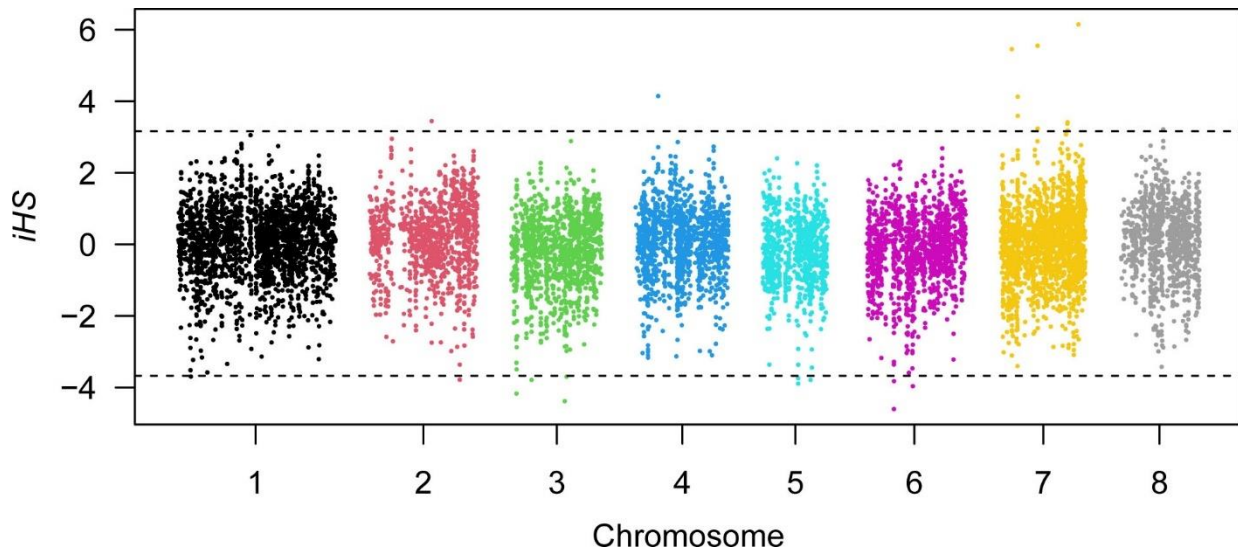
**Fig. 9.** Integrated haplotype scores (iHS) calculated using the *rehh* package in R showed evidence of positive and negative selection.  The iHS scores above the indicated threshold show genomic regions that are undergoing positive selection.  Values below the threshold show genomic regions that are being selected against.

When hexaploid taxa were removed from the dataset, iHS was re-calculated (Figure 10), resulting in a shift in the prevalence and SNP locations for indicators of positive selection (Appendix 3).  For diploid-only *Prunus*, SNPs undergoing positive selection were detected on chromosomes 1, 2, 4, 6, and 8.  Negative selection SNPs differed from those in the dataset with the hexaploid taxa by only two sites.
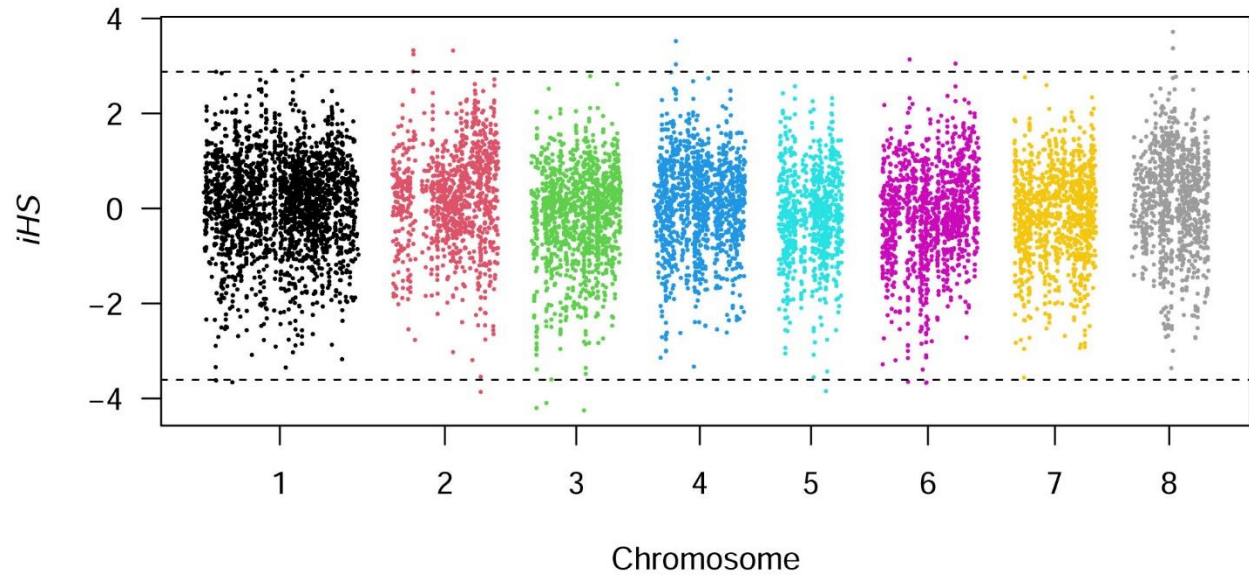
**Fig. 10.** Integrated haplotype scores (iHS) calculated with the hexaploid *P. domestica* taxa removed showed evidence of positive and negative selection. The iHS scores above the indicated threshold show genomic regions that are undergoing positive selection. Values below the threshold show genomic regions that are being selected against.

The SNPs with highest or lowest iHS were not the same as any of the SNPs identified as significant in the GWAS analysis for exo and meso carp color, shape, and free or cling-stone endocarps. This means there is selection occurring for traits that were not scored in this study and further research is needed to fully interpret these findings.

## Discussion:

This study covers the tip of the iceberg that is Burbank's breeding work. GWAS was successful at connecting SNPs to fruit exocarp color but fell short of expectations for other phenotypic traits. Previous studies indicate that fruit color is a multi-genic trait with regions on every chromosome coding for MYB transcription factors responsible for anthocyanin (red) pigmentation (Allan et al. 2008, Tuan et al. 2015, Salazar et al. 2017, Zhang et al. 2018, Liu et al. 2020, Fiol et al. 2921). Since a result indicating this association was not observed, the threshold set to detect SNPs for fruit exocarp and mesocarp color may either be too conservative, or the sample size may be too small. While relaxing this threshold does increase the chance of discovering false positive correlations, it may simultaneously punish them too stringently. Sequencing areas with the most significant SNPs to see if they are located within known MYB transcription factor regions would be useful to further tease out the relationship between fruit color and genotype.

In addition to fruit color, the intended use of the fruit is an important consideration when defining post-harvest breeding goals for a population. Burbank lived in an era before refrigeration when fruits were often consumed as preserved jellies, fermented products, or were dried for long term storage. Common breeding goals among the fresh and dried *Prunus* included increasing fruit weight, yield, shelf life, and disease resistance. Dried prunes also needed to have a high enough sugar content for them to dry without getting moldy.

Commercially, the phenotypic expression of fruit weight is manipulated by thinning and providing more water to an orchard. Consistent thinning of the taxa in this study was not feasible for all three locations, but a few cultivars stood out as being generally larger than the others. Notably, the cultivars 'Beauty DPRU.2120,' 'Wickson DPRU.2135,' 'Catherine Bunnell,' 'Elephant Heart DPRU.2123,' and 'Grand Prize DPRU.1572' are reliably the largest taxa included in this study system without thinning or supplemental irrigation. The fruit weight in each of these cultivars is problematic because the bearing load is often too great for the tree's branches. Size of endocarp can also influence the weight of a fruit and its overall size in *Prunus*. Similarly, increased yield is a characteristic Burbank desired. Some of his *Prunus* introductions produce fruit so thickly that the branches look like they are holding giant clusters of grapes. This trait was challenging to score because some cultivars did not receive enough chill hours for reliable fruiting or had adverse weather conditions like late freezes, early heat, or heavy, damaging precipitation that prevented fruit set.

Selecting for disease resistance or against disease susceptibility is another common goal for plant breeders. All the *Prunus* surveyed in the Burbank breeding population showed foliar symptoms of *Wilsonomyces carpophilus*, also known as shot-hole fungus (Adaskaveg et al. 2015). If this was a trait he was selecting against, he was unsuccessful. The taxa varied in their degree of susceptibility to *Brachycaudus helichrysi,* the aphid causing leaf curl (Bentley et al. 2009). Susceptibility to this insect pest was confounded by environmental interactions such as drought and abnormal temperature spikes during the winter dormant period. The disease pressure for this organism was high at all three locations, and little resistance was observed. The

biggest killer of plums in Sonoma County, CA is bacterial canker, caused by *Pseudomonas syringae* pv. *syringae* (Gubler et al. 2009).

Disease resistant rootstocks are essential for successful cultivation of any *Prunus*. Some of the taxa in this study are grafted onto *P. cerasifera* but rootstock documentation was incomplete for most, making it difficult to adequately score taxa susceptibility to this bacterium. If plants are genetically resistant to any of the above common diseases, we would expect to see significant SNPs in the GWAS results for disease resistance. Conversely, in an iHS plot, we would expect to see both areas of selection for disease resistance and against disease susceptibility, making this a highly valuable metric for plant breeders.

The difference in SNP number and location in the accessions with hexaploid *P. domestica* in their pedigree and those who do not have *P. domestica* illustrates Burbank's breeding goals differed between prunes and plums. Prunes typically feature attributes like higher sugar content to facilitate drying than is necessary for fresh-eating plums. Fruits needed to dry evenly without molding. Burbank mentioned selecting for fruits that would ripen and dry on the tree, then be harvested by shaking (Hedrick 1911, Burbank et al. 1914). Blue skinned plums were favored over yellow ones because the market preference during that time was for dark-colored Italian or French prunes, so genes for higher anthocyanin production would have been preferred. Stonelessness in prunes was a desired trait that Burbank was never able to fully attain. There was always a remnant sliver of the stony endocarp that formed, decreasing the attractiveness of this cultivar from a commercial standpoint.

Breeding goals for the fresh-eating plum market differ from those of the dried prune market.  Traits that would be preferred for the fresh eating plums are ones that increase shelf stability.  Therefore, climacteric fruits, ones that can ripen off the tree, would have been favored.  The ability for fruits to be climacteric can be altered depending on their receptor responses to ethylene.  'Santa Rosa,' arguably the Burbank *Prunus* cultivar with the most long-lived commercial success, tends to throw somatic budsports that vary in the copy number for a gene that responds to ethylene (Minas et al. 2015).

More samples would be useful for detecting genomic signal for mesocarp color, free and cling-stone endocarps, and fruit shape.  Signals were detected using iHS, but those SNPs did not correlate to the phenotypic traits scored.  This indicates there is evidence of positive and negative selection, but the phenotypes associated with these selection events is unknown.  Evidence from other research points towards loci on chromosome 6 controlling the trait of self-incompatibility, a feature that would be commonly selected against (Aranzana et al. 2019).  More time to understand seasonal variation will be required to fully phenotype some of the key attributes found in his *Prunus* collection because results can be incredibly variable between years depending on chill hour accumulation, annual rainfall, and mid-winter spikes in temperature.  These environmental cues have a strong influence on many relevant phenotypic traits such as increased fruit weight, yield, and disease susceptibility.

## Conclusion:

Though Burbank's note taking was minimal, modern genomic tools allow scientists to peer into the past and uncover the mysteries that tie genotype and phenotype together.  This research shows how GWAS and iHS can be powerful tools for surveying an inherited breeding population with little known pedigree information or genomic data.  Plum and prune breeders can find useful SNPs based in Luther's work to utilize in their own breeding experiments for traits related to marketability, leading to "better fruits…for all to enjoy" as Burbank intended.

## References:

Abdurakhmonov, I.Y. and A. Abdukarimov.  2008.  Application of Association Mapping to Understanding the Genetic Diversity of Plant Germplasm Resources.  International Journal of Plant Genomics. http://doi.org/10.1155/2008/574927 .

Adaskaveg, J.E., R.A. Duncan, J.K. Hasey, K.R. Day, W.D. Gubler, and B.L. Teviotdale.  2015.  Shot Hole Disease.  UC IPM Pest Management Guidelines:  Peach.  UC ANR Publication 3454. https://www2.ipm.ucanr.edu/agriculture/peach/shot-hole-disease/

Allan, A.C., R.P.Hellens, W.A. Laing.  2008.  MYB transcription factors that colour our fruit.  Trends in Plant Science.  13(3): 99-102.  https://doi.org/10.1016/j.tplants.2007.11.012 .

Aranzana, M.J., V. Decroocq, E. Dirlewanger, I. Eduardo, Z.S. Gao, K. Gasic, A. Lezzoni, S. Jung, C. Peace, H. Prieto, R. Tao, I. Verde, A.G. Abbott, and P. Arus.  (2019).  *Prunus* genetics and applications after de novo genome sequencing: achievements and prospects.  Horticultural Research.  6(58). https://doi.org/10.1038/s41438-019-0140-8 .

Bahr, T.L., K. Gasic, M.dC. Bassols Raseira, S. Bonow, and C. Marques Castro.  2020.  Genome-wide SNP discovery through genotyping by sequencing, population structure, and linkage disequilibrium in Brazilian peach breeding germplasm.  Tree Genetics & Genomes.  16(1).  https://doi.org/10.1007/s11295-019-1406-x .

Bentley, W.J., K.R. Day, and R.E. Rice.  2009.  Leaf Curl Plum Aphid.  UC IPM Pest Management Guidelines:  Plum.  UC ANR Publication 3462. http://ipm.ucanr.edu/PMG/r611301811.html#:~:text=The%20leaf%20curl%20plum%20aphid,near%20the%20base%20of%20buds.

Bradbury PJ, Z. Zhang, D.E. Kroon, T. M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Bretó, M.P., C.M. Cantín, I. Iglesias, and I. Eduardo. 2017. Mapping a major gene for red skin color suppression (highlighter) in peach. Euphytica 213, 14. https://doi.org/10.1007/s10681-016-1812-1

Brooks, R.M. and H.P. Olmo. 1952. Register of New Fruit and Nut Varieties 1920-1950. University of California Press. http://reader.library.cornell.edu/docviewer/digital?id=chla3174378#mode/1up

Browning, B.L., Y. Zhou, and S.R. Browning. 2018. A one-penny imputed genome from next generation reference panels. American Journal of Human Genetics. 103(3): 338-348. https://doi.org/10.1016/j.ajhg.2018.07.015

Burbank L., J. Whitson, R. John, and H. Smith Williams (eds). (1914). Luther Burbank: his methods and discoveries and their practical application. New York; London: Luther Burbank Press.

Bush, W.S. and J.H. Moore. 2012. Chapter 11: Genome-wide association studies. PLOS Computational Biology. 8(12). https://doi.org/10.1271/journal.pcbi.1002822 .

Carresco, B. M. Gonzalez, M. Gebauer, R. Garcia-Gonzalez, J. Maldonado, and H. Silva. 2018. Construction of a highly saturated linkage map in Japanese Plum (*Prunus salicina* Lindl.) using GBS for SNP marker calling. PLOS One. https://doi.org/10.1371/journal.pone.0208032 .

Cao, K., Z. Zhou, Q. Want, J. Guo, P. Zhao, G. Chu, W. Fang, W. Chen, X. Wang, X. Wang, Z. Tian, and L. Wang. 2016. Genome-wide association study of 12 agronomic traits in peach. Nature Communications. 7: 13246. https://doi.org/10.1038/ncomms13246 .

Cheng, Z. K. Gasic, and A. Wang. 2013. Population structure, linkage disequilibrium and selective loci in natural populations of *P. davidiana*. Journal of Agricultural Science. 5(6). https://dx.doi.org/10.5539/jas.v5n6p69 .

da Silva Linge, C., L. Cai, W. Fu, J. Clark, M. Worthington, Z. Rawandoozi, D.H. Byrne, and K. Gasic. 2021. Multi-locus genome-wide association studies reveal fruit quality hotspots in peach genome. Fronteirs in Plant Science. http://doi.org/10.3389/fpls.2021.644799 .

Endelman, J.B. 2011/ Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome. 4(3). https://doi.org/10.3835/plantgenome2011.08.0024 .

Fiol, A., B.E. Garcia-Gomez, F. Jurado-Ruiz, K. Alexiuo, W. Howard, and M.J. Aranzana. 2021. Characterization of Japanese plum (*Prunus salicina*) PsMYB10 alleles reveals structural variation and polymorphisms correlating with fruit skin color. Frontiers in Plant Science. https://doi.org/10.3389/fpls.2021.655267 .

Foster, T. M. and M. J. Aranzana. 2018. Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. Horticulture Research 5(44). https://doi.org/10.1038/s41438-018-0062-x .

Gautier, M. and R. Vitalis. 2012. *Rehh*: and R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics. 28(8): 1176-1177. https://doi.org/10.1093/bioinformatics/bts115 .

Ge, J. B. Budowle, J.V. Planz, and R. Chakraborty. 2010. Haplotype Block: a new type of forensic DNA markers. International Journal of Legal Medicine. 124: 353-361. https://doi.org/10.1007/s00414-009-0400-5 .

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES. (2014) TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLOS ONE 9(2): e90346

Gomez, E. and C.A. Ledbetter. 1994. Comparative study of the aromatic profiles of two different plum species: *Prunus salicina* Lindl and *Prunus simonii* L. Journal of the Science of Food and Agriculture. 65(1): 111-115. https://doi.org/10.1002/jsfa.2740650116 .

Guajardo, V., S. Solis, R. Almada, C. Saksi, K. Gasic, and M. Angeles-Moreno. 2020. Genome-wide SNP identification in *Prunus* rootstocks germplasm collections using Genotyping-by Sequencing: phylogenetic analysis, distribution of SNPs and prediction of their effect on gene function. Scientific Reports. 10: 1467. https://doi.org/10.1038/s41598-020-58271-5 .

Gubler, W.D., J.E. Adaskaveg, K.R. Day, and B.L. Teviotdale. 2009. Bacterial Canker. UC IPM Pest Management Guidelines: Plum. UC ANR Publication 3462. http://ipm.ucanr.edu/PMG/r611100111.html

Hedrick, U.P. 1911. Plums of New York. http://chla.library.cornell.edu/cgi/t/text/text-idx?c=chla;cc=chla;idno=5519832

Hodel, RGJ., E. Zimmer, & J.Wen. 2021. A phylogenomic approach resolves the backbone of Prunus (Rosaceae) and identifies signals of hybridization and allopolyploidy. Molecular Phylogenetics and Evolution. 160: 107118.

Howard, WL. 1945. Luther Burbank's Plant Contributions.

Iowa State University. 2021. Plums. Agricultural Marketing Resource Center. https://www.agmrc.org/commodities-products/fruits/plums

Kim, S., V. Plagnol, T.T. Hu, C. Toomajian, R.M. Clark, S. Ossowski, J.R. Ecker, D. Weigel, and M. Nordborg. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nature Genetics. 39: 1151-1155. https://doi.org/10.1038/ng2115 .

Koshita, Y. 2014. Effect of temperature on fruit color development. Abiotic Stress Biology in Horticultural Plants. 47-58. https://doi.org/10/1007/978-4-431-55251-2_4 .

Liu, C., D. Feng, W. Peng, J. Hao, J. Wang, J. Pan, Y. He. 2020. Chromosome-level draft genome of a diploid plum (*Prunus salicina*). GigaScience. 9(12). https://doi.org/10.1093/gigascience/giaa130 .

Liu, C., J. Hao, M. Qiu, J. Pan, Y. He. 2020. Genome-wide identification and expression analysis of the MYB transcription factor in Japanese Plum (*Prunus salicina*). Genomics. 112(6): 4875-4886. https://doi.org/10.1016/j.ygeno.2020.08.018 .

Ma, Y. X. Ding, S. Qanbari, S. Weigend, Q. Zhang, and H.Simianer. 2015. Properties of different selection signature statistics and a new strategy for combining them. Heredity. 115: 426-436. https://doi.org/10.1038/hdy.2015.42 .

Marti, Fernandez I, A., Saski, C. A., Manganaris, G. A., Gasic, K., & Crisosto, C. H. (2018). Genomic Sequencing of Japanese Plum (*Prunus salicina* Lindl.) Mutants Provides a New Model for Rosaceae Fruit Ripening Studies. Frontiers in Plant Science, 9, 21. https://doi:10.3389/fpls.2018.0002

Minas, I. S., Font I Forcada, C., Dangl, G. S., Gradziel, T. M., Dandekar, A. M., & Crisosto, C. H. (2015). Discovery of non-climacteric and suppressed climacteric bud sport mutations originating from a climacteric Japanese plum cultivar (*Prunus salicina* Lindl.). Frontiers in Plant Science, 6, 316. doi:10.3389/fpls.2015.00316

Navarro, A. T., G. Tumina, R.G.F. Visser, R.E. Voorrips, W.E. van de Weg, and C. Maliepaard. 2020. Multiparental QTL analysis: can we do it in polyploids? Acta Horticulturae. 1283. http://DOI:10.17660/ActaHortic.2020.1283.9 .

Peace, C.P., C.H. Crisosto, and T.M. Gradziel. 2005. Endopolygalacturonase: a candidate gene for *Freestone* and *Melting flesh* in peach. Molecular Breeding. 16: 21-31. https://doi.org/10.1007/s11032-005-0828-3 .

Rawandoozi, Z. J., T. P. Hartmann, S. Carpendo, K. Gasic, C. da Silva Linge, L. Cai, E. Van de Weg, and D. H. Byrne. 2021. Mapping and characterization of QTLs for phenological traits in seven pedigree-connected peach families. BMC Genomics. 22(187). https://doi.org/10.1186/s12864-021-07483-8 .

Salazar, J.A., I. Pacheco, P. Shinya, P. Zapata, C. Silva, M. Aradhya, D. Velasco, D. Ruiz, P. Martinez-Gomez, and R. Infante. 2017. Genotyping by sequencing for SNP-based linkage analysis and identification of QTLs linked to fruit quality traits in Japanese Plum (*Prunus salicina* Lindl.). Frontiers in Plant Science. 11. https://doi.org/10.3389/fpls.2017.00476 .

Salazar, J.A., I. Pacheco, C. Silva, P. Zapata, P. Shinya, D. Ruiz, P. Martinez-Gomez, and R. Infante. 2018. Development and applicability of GBS approach for genomic studies in Japanese plum (*Prunus salicina* Lindl.). The Journal of Horticultural Science and Biotechnology. 94: 284-294.

Scott, M.F., O. Ladejobi, S. Amer, A.R. Bentlye, J. Biernaskie, S.A. Boden, M.Clark, M. Dell'Acqua, L.E. Dixon, C.V. Filippi, N. Fradgley, K.A. Gardner, I.J. Mackay, D. O'Sulivan, L. Percival-Alwyn, M. Roorkiwal, R. Kumar SIngh, M. Thudi, R. Kumar Varshney, L. Venturini, A. Whan, J. Cockram, and R. Mott. 2020. Multi-parental populations in crops: a toolbox integrating genomics and genetic mapping with breeding. Heredity. 125: 396-416.

Sootiyapathirana, S.S., A. Khan, A.M. Sebolt, D. Want, J.M. Bushakra, K. Lin-Wang, A.C. Allan, S.E. Gardiner, D. Chagne, and A.F. Iezzoni. 2010. QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). Tree Genetics & Genomes. 6: 821-832. https://doi.org/10.1007/s11295-010-0294-x .

Tan, Q., L. Zhang, M. Chen, B. Wen, S. Jiang, X. Chen, X. Fu, D. Li, H. Wu, Y. Wang, W. Xiao, and L. Li. 2021. Chromosome-level genome assemblies of five *Prunus* species and genome-wide association studies for key agronomic traits in peach. Horticulture Research. 8(213). https://doi.org/10.1038/s41438-021-00648-2 .

Topp, B.L., D.M. Russell, M. Neumüller, M.A. Dalbó, and W. Liu. (2012). Plum (Handbook of Plant Breeding). 8, part 3. pp. 571–621. https://doi:10.1007/978-1-4419-0763-9_15

Tuan, P.A., S. Bai, H. Yaegaki, T. Tamura, S. Hihara, T. Moriguchi, and K. Oda. 2015. The crucial role of *PpMYB10.1* in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. BMC Plant Biology. 15(280). https://doi.org/10.1186/s12870-015-0664-5 .

University of California Agriculture and Natural Resources Statewide Integrated Pest Management Program (UC IPM). (2017). Management of Plums. Accessed 2022. http://ipm.ucanr.edu/PMG/GARDEN/FRUIT/plumsprunes.html

Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C. T., Gasic, K., et al. (2012). Development and evaluation of a 9K SNParray for peach by internationally coordinated SNP detection and validation in breeding germplasm. PLOS ONE 7:e35668. Doi:10.1371/journal.pone.0035668 .

Yu, J. G. Pressoir, W.H.Briggs, I.V. Bi, M.Yamasaki, J.F.Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics. 38(2):203-8. https://doi.org/10.1038/ng1702 .

Zahid, G. A. A. Kacar, D. Donmez, A. Kuden, and T. Giordani. 2022. Perspectives and recent progress of genome-wide association studies (GWAS) in fruits. Molecular Biology Reports. 49: 5341-5352.

Zhang, C., R. Ma, J. Xu, J. Yan, L. Guo, J. Song, R. Feng, and M. Yu. 2018. Genome-wide identification and classification of *MYB* superfamily genes in peach. PLOS ONE. 13(6). https://doi.org/10.1371/journal.pone.0199192 .

Zhao, J., C. Sauvage, F. Bitton, and M. Causse. 2022. Multiple haplotype-based analyses provide genetic and evolutionary insights into tomato fruit weight and composition. Horticultural Research. 9. http://doi.org.10.1093/hr/uhab009 .

Zhao, Y., W. Dong, Y. Zhu, A.C. Allan, K. Lin-Wang, and C. Xu. 2020. PpGST1, an anthocyanin-related glutathione S-transferase gene, is essential for fruit coloration in peach. Plant Biotechnology Journal. 18(5): 1284-1295. https://doi.org/10.1111/pbi.13291 .

Zhebentyayeva T., V. Shankar, R. Scorza, A. Callahan, M. Ravelonandro, S. Castro, T. DeJong, C.A. Saski, and C. Dardick. (2019). Genetic characterization of worldwide *Prunus domestica* (plum) germplasm using sequence-based genotyping. Horticulture Research 6:12.

# Appendix

<u>Appendix 1:</u>  A list of the Burbank taxa used in this study with ploidy, year of introduction, suspected parents, exocarp color (red, yellow, or blue), endocarp color (red or yellow), free or cling-stone endocarp, and general shape (round, pointed, or oval).

| Name &Location | Ploidy | Year of Introduction | Suspected Parents | Exocarp Color | Mesocarp Color | Endocarp Free(0) Cling(1) | Shape Round(0) Pointed(1) Oval(2) |
|---|---|---|---|---|---|---|---|
| Abundance_WEO DPRU.919 | 2 | 1914 | Satsuma x P. armeniaca | red | yellow | 0 | 0 |
| Anderson's_Early_Red_WEO DPRU.843 | 2 | no data | P. americana | red | yellow | 0 | 0 |
| Apex_GR | 2 | 1911 | P. salicina x P. armeniaca | red | yellow | 0 | 0 |
| Apex_WEO DPRU.1170 | 2 | 1911 | P. salicina x P. armeniaca | red | yellow | 0 | 0 |
| Ballena_GR | 6 | 1906 | (Simon x Delaware) x OP | blue | yellow | 0 | 2 |
| Beauty_LBHG | 2 | 1911 | | red | red | 1 | 1 |
| Beauty_SW_LBHG | 2 | 1911 | | red | red | 1 | 1 |
| Beauty_WEO DPRU.2120 | 2 | 1911 | | red | red | 1 | 1 |
| Botanky_WEO DPRU.372 | 2 | 1888 | P. salicina | purple | yellow | 1 | 0 |
| Brookgold_WEO DPRU.1736 | 2 | no data | P. salicina | blue | yellow | 1 | 0 |
| Burbank_Plumcot_GR | 2 | 1914 | P. salicina x P. armeniaca | red | yellow | 0 | 0 |
| Burbank_(Plumcot)_WEO DPRU.936 | 2 | 1914 | P. salicina x P. armeniaca | red | yellow | 0 | 0 |
| Catherine_Bunnell_LBHG | 2 | 1908 | Santa Rosa x OP | red | red | 1 | 1 |
| Chalco_WEO DPRU.431 | 2 | 1898 | Simon x Burbank | black | yellow | 1 | 0 |
| *P. simonii*_WEO DPRU.2430 | 2 | no data | P. simonii | red | yellow | 0 | 0 |
| El_Dorado_GR | 2 | 1904 | P. salicina x Simon | black | yellow | 0 | 0 |
| El_Dorado_LBHG | 2 | 1904 | P. salicina x Simon | black | yellow | 0 | 0 |
| El_Dorado_SR_LBHG | 2 | 1904 | P. salicina x Simon | black | yellow | 0 | 0 |
| El_Dorado_WEO DPRU.2122 | 2 | 1904 | P. salicina x Simon | black | yellow | 0 | 0 |
| Elephant_Heart_GR | 2 | 1929 | Satsuma x Wickson | red | red | 0 | 1 |
| Elephant_Heart_WEO DPRU.2123 | 2 | 1929 | Satsuma x Wickson | red | red | 0 | 1 |
| Formosa?_WEO DPRU.924 | 2 | 1907 | P. salicina | red | yellow | 0 | 1 |
| Formosa_GR | 2 | 1907 | P. salicina | red | yellow | 0 | 1 |
| Formosa_LBHG | 2 | 1907 | P. salicina | red | yellow | 0 | 1 |
| French_WEO DPRU.436 | 6 | 1889 | P. domestica | blue | yellow | 0 | 2 |
| Grand_Prize_WEO DPRU.1572 | 6 | 1937 | P. domestica | blue | yellow | 0 | 2 |
| Great_Yellow_SpSa_LBHG | 2 | 1931 | Shiro x Simon | yelow | yellow | 0 | 0 |
| Great_Yellow_WEO DPRU.2105 | 2 | 1931 | Shiro x Simon | yellow | yellow | 0 | 0 |
| Improved_Satsuma_LBHG | 2 | no data | Satsuma x OP | red | red | 1 | 1 |
| June_Red_LBHG | 2 | 1934 | Simon x P. americana | red | yellow | 0 | 0 |
| Late_Goose_WEO DPRU.546 | 2 | no data | P. rivularis | yellow | yellow | 0 | 2 |
| Latest_of_All_GR | 2 | no data | P. domestica | yellow | yellow | 1 | 0 |
| Latest_of_All_WEO DPRU.427 | 2 | no data | P. domestica | yellow | yellow | 1 | 0 |
| Lieb_LBHG | 2 | 1914 | Burbank x Satsuma | red | red | 0 | 1 |
| Mammoth_Cardinal_LBHG | 2 | 1934 | Simon x OP | red | yellow | 0 | 0 |
| Mammoth_Cardinal_WEO DPRU.2127 | 2 | 1934 | Simon x OP | red | yellow | 0 | 1 |
| OG_Stoneless_WEO DPRU.2302 | 6 | no data | P. domestica | black | yellow | 0 | 2 |
| Pcot_Edibles_LBHG | 2 | no data | | blue | yellow | 0 | 0 |
| Perfection_WEO DPRU.1720 | 2 | 1892 | Burbank x Simon | black | yellow | 1 | 0 |
| Rutland_GR | 2 | 1905 | P. salicina x P. armeniaca | red | red | 0 | 0 |
| Sans_Noyau_WEO DPRU.2419 | 6 | 1768 | P. domestica | blue | yellow | 0 | 2 |
| Santa_Rosa_GR | 2 | 1906 | (P. salicina x P. simonii) x P. america | red | red | 1 | 1 |
| Satsuma_WEO DPRU.438 | 2 | 1886 | P. salicina | red | red | 1 | 1 |
| Shiro_GR | 2 | 1899 | (Robinson x P. cerasifera) x Wickson | yellow | yellow | 1 | 1 |
| Shiro_WEO DPRU.2132 | 2 | 1899 | (Robinson x P. cerasifera) x Wickson | yellow | yellow | 1 | 1 |
| Simon_WEO DPRU.545 | 2 | 1872 | P. simonii | red | yellow | 0 | 0 |
| Sultan_LBHG | 2 | 1899 | Wickson x Satsuma | red | red | 0 | 0 |
| Top_of_the_Hill_GR | 6 | no data | P. domestica | blue | yellow | 0 | 2 |
| Unk_7_GR | 2 | no data | no data | | | | |
| Unk_Multi_GR | 2 | no data | no data | | | | |
| Vesuvius_WEO DPRU.2108 | 2 | 1907 | P. pissardii x P. triflora | red | red | 1 | 0 |
| Victory_WEO DPRU.791 | 2 | 1911 | (Robinson x Botanky) x OP | red | yellow | 1 | 1 |
| Wickson_WEO DPRU.2135 | 2 | 1892 | Burbank x Simon | red | yellow | 1 | 1 |

Appendix 2: Integrated haplotype scores (iHS) calculated using the *rehh* package in R shows evidence of positive and negative selection at certain SNPs throughout the genome in a breeding population of 47 diploid and 6 hexaploid *Prunus* taxa introduced or bred by Luther Burbank a century ago.

| Chromosome | Position | iHS | Log p-value |
|---|---|---|---|
| 1 | 5593723 | -3.69507 | 3.657913 |
| 2 | 21573899 | 3.445057 | 3.243411 |
| 2 | 30947868 | -3.78373 | 3.811081 |
| 3 | 2282699 | -4.17025 | 4.516739 |
| 3 | 7329326 | -3.78745 | 3.817576 |
| 3 | 18408151 | -4.38135 | 4.928317 |
| 3 | 18992023 | -3.69685 | 3.660971 |
| 4 | 8020352 | 4.144251 | 4.467339 |
| 5 | 12540270 | -3.89088 | 4.000524 |
| 5 | 12541601 | -3.74078 | 3.736476 |
| 5 | 16679749 | -3.78943 | 3.821045 |
| 6 | 11307706 | -4.60344 | 5.38136 |
| 6 | 11332690 | -3.82382 | 3.881412 |
| 6 | 17611891 | -3.96375 | 4.132047 |
| 7 | 4432039 | 5.455899 | 7.312241 |
| 7 | 6401357 | 4.125506 | 4.431891 |
| 7 | 6408505 | 3.590288 | 3.481074 |
| 7 | 13029276 | 3.237065 | 2.918056 |
| 7 | 13032712 | 5.550758 | 7.546019 |
| 7 | 22977539 | 3.372033 | 3.12717 |
| 7 | 23055350 | 3.411774 | 3.190161 |
| 7 | 23082767 | 3.184218 | 2.838196 |
| 7 | 26647050 | 6.148167 | 9.105774 |
| 8 | 15142522 | 3.212113 | 2.880209 |

Appendix 3: Integrated haplotype scores (iHS) calculated using the *rehh* package in R shows evidence of positive and negative selection at certain SNPs throughout the genome in a breeding population of 47 diploid only *Prunus* taxa introduced or bred by Luther Burbank a century ago.

| Chromosome | Position | iHS | Log p-value |
|---|---|---|---|
| 1 | 5459954 | -3.61379 | 3.5203515 |
| 1 | 5593723 | -3.62049 | 3.53158195 |
| 1 | 11088546 | -3.65869 | 3.59600998 |
| 1 | 25505507 | 2.902266 | 2.43124249 |
| 2 | 8025078 | 2.883595 | 2.4054269 |
| 2 | 8101319 | 3.330411 | 3.06189195 |
| 2 | 8135118 | 3.247243 | 2.93356687 |
| 2 | 21573899 | 3.325398 | 3.05407696 |
| 2 | 30947895 | -3.86096 | 3.94713617 |
| 3 | 2282699 | -4.19874 | 4.5712129 |
| 3 | 5652505 | -4.09426 | 4.37313138 |
| 3 | 18408151 | -4.25012 | 4.6702865 |
| 4 | 8020352 | 3.524921 | 3.37303466 |
| 4 | 8020388 | 3.034875 | 2.61864 |
| 5 | 16679749 | -3.84499 | 3.91880057 |
| 6 | 11332690 | -3.64841 | 3.57860937 |
| 6 | 11883321 | 3.13873 | 2.77036553 |
| 6 | 17476424 | -3.66942 | 3.61420725 |
| 6 | 17611891 | -3.66187 | 3.60140146 |
| 6 | 27451092 | 3.051613 | 2.64279792 |
| 8 | 15142522 | 3.719125 | 3.69915592 |
| 8 | 15142527 | 3.373011 | 3.12871251 |