HortScience 59(6):873–880. 2024. https://doi.org/10.21273/HORTSCI17731-24

Relatedness of Luther Burbank's Plum (*Prunus* sp.) Introductions Based on Genotyping by Sequencing

Rachel A. Spaeth
Department of Horticulture & Agronomy, University of California, One Shields Avenue, Davis, CA 95616, USA

Dominique D.A. Pincot, Daniel Potter, P.J. Brown, and Tom Gradziel
Department of Plant Sciences, University of California, One Shields Avenue, Davis, CA 95616, USA

John E. Preece
National Clonal Germplasm Repository, USDA-ARS, 3201 Straloch Road, Davis, CA 95616, USA

Keywords. collections curation, genotyping by sequencing, Luther Burbank, pedigree analysis, plums, pomology

Abstract. **The renowned horticultural artist and plant breeder Luther Burbank worked with many species of plants. During his 50-year career, he introduced more than 800 cultivars, including more than 150 accessions of plums (*Prunus* spp.) in the late 1800s and early 1900s. Burbank preferred using wide, interspecific crosses to create a vast range of phenotypic variation and then artificially select from the extremes. Although a great artist, Burbank was a substandard scientist because he was derelict in pedigree note-taking. Although many of his introductions are extinct, hobbyists, enthusiasts, and international collections retain nearly a third of the economically viable cultivars he bred. For a century, many of his hybridizations remained inscrutable 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, 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 Burbank in his breeding experiments in the most comprehensive genetic survey of his work to date. Exact parent–offspring relationships between this population may be impossible to discern due to years of back crossing, sibling mating, and open pollination. However, the proportion of genomic similarity among these taxa provides information on the relatedness of the genotypes in Burbank's *Prunus* experiments, defining four primary lineages within his breeding population. These lineages comprised primarily *P. salicina* and *P. simonii* but also have influences from *P. americana*, *P. cerasifera*, *P. domestica*, and *P. rivularis*. The prevalence of *P.* simonii in Burbank's *Prunus* introductions appears to have been vastly underreported, indicating that some of the seedstock founders of his breeding population could have been *P. salicina* × *P. simonii* hybrids at the inception of his career. This research has implications for pedigree reconstruction and prioritizing conservation in collections curation for future studies.**

At the turn of the 19th–20th centuries, Luther Burbank (1849–1926) 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 (Dreyer 1993; Smith 2009). 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 et al. 1914), and as such produced hundreds of thousands of seedlings for evaluation (Topp et al. 2012). Using this method, Burbank revolutionized the human perception of what a plum could be and commercially introduced more than 150 cultivars of plums, prunes, and plumcots in 40 years (Brooks and Olmo 1952; Hedrick 1911; Howard 1945; Karp 2015) (Supplemental Appendix 1). His breeding population of plums included many intraspecific hybrids, interspecific hybrids, and bud sports with a huge range of phenotypic variation. Unfortunately, Burbank kept poor breeding records and relied instead on faded strips of clothing tied to branches and his memory to track the pedigrees of his plants (Dreyer 1993). 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 were 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 genome sequencing is cost-effective and covers much larger stretches of the genome for association mapping or discovering quantitative trait loci.

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 is often 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 sample the collection(s) broadly, establishing a genomic baseline for their target population. These data inform breeding choices and identify goals for cultivar selection or prioritizing conservation. Visualizing the relationships among the organisms in a breeding population is 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 performed using data from eight SSR markers of *Prunus salicina* and hybrids of *P. salicina* showed that interspecific hybrids tend to cluster together depending on their admixture of genotypes,

Received for publication 23 Jan 2024. Accepted for publication 2 Apr 2024.
Published online 20 May 2024.
R.A.S. is the corresponding author. E-mail: raspaeth@ucdavis.edu.

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 because 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). However, they are highly applicable to the reconstruction of pedigree data because the genomes with multiparental 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 intraspecific *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 *Prunus simonii*, *Prunus domestica*, *Prunus americana*, *P. salicina*, *Prunus rivularis*, and *Prunus 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 Burbank emerged (Hedrick 1911). Also missing from this list are the taxa that were patented after Burbank's death by his widow Elizabeth in collaboration with Stark Brothers. This target list was used to hunt for specific cultivars through ARS repository access, word of mouth, and local California Rare Fruit Growers scion exchanges. Once a cultivar was located, old literature was searched for claims Burbank made about their parentage as well as any historical images that may accompany them (Burbank et al. 1914; Brooks and Olmo 1952; Hedrick 1911; Howard 1945). Scions of material found at exchanges or through word-of-mouth were multigrafted onto mature trees with *Prunus cerasifera* 'Myrobalan 29C' as a universal rootstock at the Luther Burbank Home & Gardens (LBHG) in Santa Rosa, CA, USA. Historic maps were consulted for cultivar names at Luther Burbank's Goldridge Experiment Farm (GR) in Sebastopol, CA, USA. The GRIN Global Database and map of the US Department of Agriculture Agricultural Research Service National Clonal Germplasm Repository's (USDA-ARS-NCGR) 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 –80 °C. The DNA extraction protocol followed DNeasy Plant Kit from Qiagen (2006). Whole genome sequencing was completed using an ILLUMINA NextSEq. 500 with a run of 328.6M spots, 28.6G bases, and 11.6Gb downloads. Retrieved sequences were aligned to *P. salicina* 'Sanyueli' (Liu et al. 2020). Samples with more than 90% missing data were discarded. Approximately 50,000 single nucleotide polymorphisms (SNPs) were retrieved. All genotypes were set to a minimum depth of <5% missing, and then SNPs with >50% missing were discarded using TASSEL5 (Bradbury et al. 2007). The final imputed dataset contained 24,147 SNPs. Missing genotypes were imputed using Beagle (Browning et al. 2018). Sequence data were deposited in the Sequence Read Archive database at National Center for Biotechnology Information as a batch and can be accessed using the BioProject Accession number PRJNA1032951.

Principal component analysis. PCA was performed in TASSEL 5 using default parameters, and the resulting data were exported to R for visualization using the R packages "ggplot2" and "ggthemes" (Arnold 2021; Wickham 2016). Admixture cluster data were calculated using STRUCTURE for K values of 2, 3, 4, 5, and 6 (Pritchard et al. 2009). Diagnostic plots which maximized L(K) and minimized DIC (Ciofi et al. 2002; Evanno et al. 2005; Gao et al. 2011; Hampton et al. 2004; Tonkin-Hill and Lee 2016; Vernesi et al. 2003; Zeisset and Beebee 2001) were used to select the optimal value of K = 4. These thresholds show the maximum statistically relevant value of K, where the benefit of 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 and 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). Because these most of these taxa come from a population of related individuals, IBS is synonymous with IBD. However, six wild ancestors were included in this study to help tease apart genetic relationships. In their case, IBD was calculated from IBS following Bernardo et al. (1996).

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 (Fig. 1); yellow, green, or red mesocarp; and free or clingstone endocarps.

A kinship matrix based on IBD values of all taxa revealed degree of genetic similarity between the taxa (Fig. 2). Colors in this matrix correlated to the impact from shared IBS. Warm colors (yellow, orange, and red) showed genotypes that were more alike. The intensity of this color palate showed the genotypes that share more rare alleles. Cool colors (whites and blues) showed genotypes that were more likely to have opposing alleles, indicating that their genotypes were less alike. Somatic ploidy levels for *Prunus* in this study include diploid species (2n = 2x =16) (*P. americana*, *P. cerasifera*, *P. rivularis*, *P. salicina*, and *P. simonii*) and a hexaploid nm.species (2n = 6x = 48) *P. domestica* (Das et al. 2011; Glowaka et al. 2021).

Principal components (PCs) for the first five components are reported for each of the taxa (Supplemental Appendix 2). PCs 1 and 2, which account for 32.65% of the cumulative proportion (Table 1), were plotted against each other (Fig. 3). Ellipses representing admixture clusters were added to further visualize the relatedness of each taxon.

A phylogenetic tree was generated using Archaeopteryx (Han and Zmasek 2009), a tree-visualization package nestled in TASSEL 5 (Bradbury et al. 2007; Glaubitz et al. 2014).

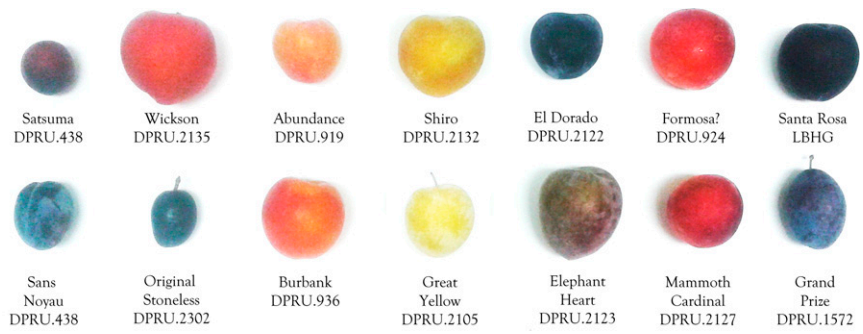

Fig. 1. Exocarp diversity in some of Burbank's plum (*Prunus* sp.) introductions grown at the Wolfskill Experimental Orchard in Winters, CA, USA, or the Luther Burbank Home & Gardens (LBHG) in Santa Rosa, CA, USA. Each fruit is labeled with their cultivar name and accession number where applicable below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

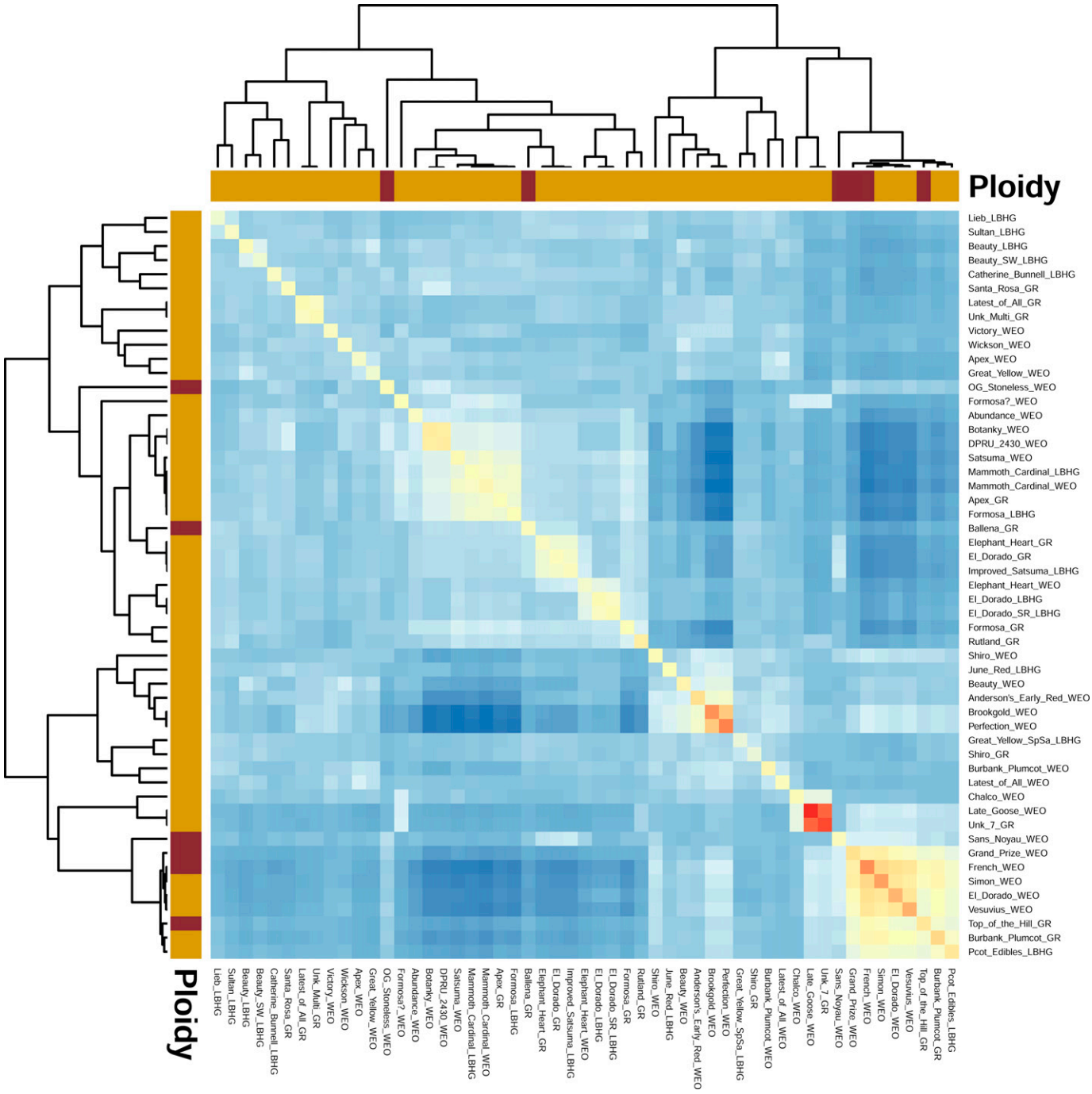

Fig. 2. A kinship heat map of 53 *Prunus* taxa introduced by Luther Burbank shows their degree of genetic similarity based on identity by state calculated in TASSEL 5. Ploidy of the taxa are indicated as gold (diploid, 2n) or maroon (hexaploid, 6n).

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 (Fig. 4). Clusters did not strictly adhere to species but instead refer to breeding population groups in general.

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 (Bradbury et al. 2007; Glaubitz et al. 2014) revealing the similarity and divergence among taxa (Fig. 5).

Table 1. Variation explained by the first five principal components (PCs) with proportion of total and cumulative total for 53 *Prunus* taxa of a Burbank breeding population generated using TASSEL 5.

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

Discussion

The combination of IBD kinship (Fig. 2), PCA (Fig. 3), phylogenetic admixture (Fig. 4), and genetic distance (Fig. 5) paints a congruent picture of relatedness among Burbank-introduced taxa. 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 and the proportions by which each taxa is connected in its relative breeding clusters. Paradoxically, these clusters did not directly correlate to individual species. Clusters 1 and 3 appear to be iterations of both *P. salicina* and *P. simonii*. Cluster 2 corresponds to the European plums *P. cerasifera* and *P. domestica*. Cluster 4 corresponds to the American plums *P. americana* and *P. rivularis*.

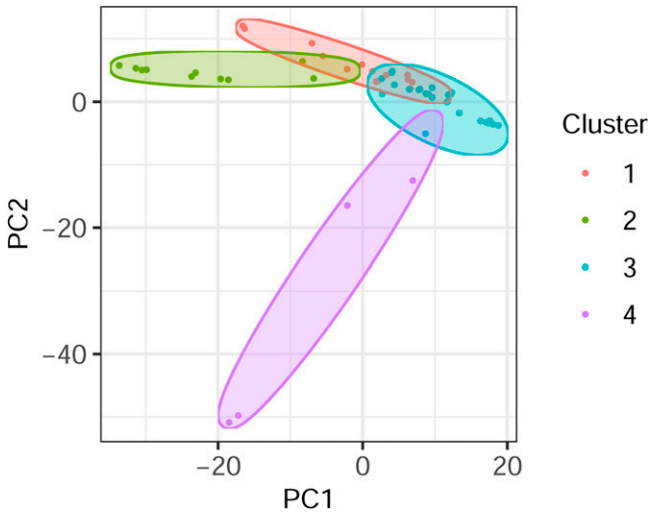

Fig. 3. Principal component (PC)1 and PC2 eigenvalues plotted for 53 Burbank-introduced *Prunus* taxa with admixture cluster groups (K = 4) highlighted by ellipses.

Including more wild types of both *P. salicina* and *P. simonii* is suggested to elucidate further the importance of these species in the Burbank *Prunus* breeding population.

The highest number of rare alleles are shared between 'Late Goose' (DPRU.546) (WEO) and Unknown 7 (GR) (Figs. 2, 4, and 5). This is also reflected in the purple ellipse on the PC plot (Fig. 3), 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, whereas 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 and a strong indication that Unknown 7 from Goldridge is a cultivar of *P. rivularis*.

Sister to the previous monophyletic group is 'Chalco' (DPRU.431) (WEO), which is listed as *Prunus* spp. in the USDA-ARS-NCGR GRIN Global database (Figs. 2, 4, and 5). 'Chalco' was bred and introduced by Burbank in 1898. He reported it to be a 'Simon' × 'Burbank' hybrid with perhaps some *P. americana* in its 12-year breeding history (Hedrick 1911), which is supported by the admixture analysis, containing about equal thirds of *P. simonii*, *P. salicina*, and a North American plum, represented in this study by 'Anderson's Early Red' (DPRU.843) (WEO) (Figs. 3 and 4).

P. rivularis 'Late Goose' (DPRU.546) (WEO) and Unknown 7 comprise mostly Cluster 4 in the admixture analysis (Figs. 3 and 4). Phenotypically, these taxa have nearly disease-free trees, small fruits with a yellow mesocarp, and a freestone endocarp. Interestingly, the exocarps of these fruits at maturity 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. The commercial value of these cultivars may be restricted to using them as rootstock (Volk 2019).

Another group with shared rare alleles is found in the taxa 'Brookgold' (DPRU.1736) and 'Perfection' (DPRU.1720), both from the Wolfskill collection (Figs. 2, 4, and 5). These two taxa have an admixture comprised entirely of Cluster 1 (Figs. 3 and 4). 'Perfection', once a synonym for 'Wickson', was arguably one of Burbank's best plum introductions (Karp 2015). However, both the phenotype and genotype for this 'Perfection' (DPRU.1720) accession are vastly different from 'Wickson' (DPRU.2135). 'Wickson' is a large, pointy-bottom *P. salicina* type (Fig. 1). 'Perfection' (DPRU.1720) from the Wolfskill is instead a small, rounded phenotype of a *P. salicina*. 'Brookgold' (DPRU.1736) is a *P. salicina* as well, which is supported by the genomic data from this study, including the IBD heatmap, PCA plot, and phylogenetic admixture dendrogram. It is possible that the 'Perfection' (DPRU.1720) is mistakenly sampled or propagated rootstock instead of the Perfection cultivar Burbank bred more than a century ago.

This pair of taxa is sister to 'Anderson's Early Red' (DPRU.843), the *P. americana* representative (Figs. 2, 4, and 5). 'Anderson's Early Red' (DPRU.843) has an admixture comprising approximately three-quarters of Cluster 1 and one quarter of Cluster 3 (Figs. 3 and 4). Both appear to have more disease resistance, smaller fruit size, and high perishability, making them limited in their degree of usefulness for a fruit consumption 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 well as a universally compatible plum rootstock to both diploid and hexaploid taxa despite limitations with its own fruit eating qualities.

The largest cluster of taxa with shared rare alleles contains a combination of hexaploid and diploid taxa (Figs. 2, 4, and 5). 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) (WEO) (*P. simonii*), 'El Dorado' (DPRU.2122) (WEO) (*P. simonii* × OP), 'Vesuvius' (DPRU.2108) (WEO) (*P. cerasifera*), 'Burbank' plumcot (historically reported as *P. salicina* × *P. armeniaca*, Goldridge accession) and Pcot Edibles (LBHG accession). These taxa all have a primary admixture of Cluster 2 (Figs. 3 and 4). 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, freestone endocarp. More work is needed to see the correlation between the genotype and phenotype in this group.

The only other taxa to have Cluster 2 present in its admixture is 'Sans Noyau' (DPRU.2419) (WEO) which appears on the opposite side of the unrooted circular dendrogram (Figs. 2–5). 'Sans Noyau' (DPRU.2419) and 'Original Stoneless' (DPRU.2302) (WEO) 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. A century later, researchers are still attempting to achieve true stonelessness in hexaploid plums, now with modern molecular techniques to assist them (Callahan et al. 2015; Galimba et al. 2020).

'Shiro' from Goldridge and 'Great Yellow' from LBHG share a considerable number of rare alleles (Figs. 2, 4, and 5). 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) (WEO), introduced by Burbank in 1911. 'Shiro' is a plum Burbank bred and introduced in 1899 (Supplemental Appendix 1) that is still readily available today through various tree nurseries. 'Great Yellow' is a Burbank plum that was bred by him but was introduced and patented posthumously by Stark Brothers in 1931 (US plant patent 13, Supplemental 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'. It may be that 'Great Yellow' from LBHG was a mislabeled accession of 'Shiro' when offered at a California Rare Fruit Growers scion exchange.

'Botanky' (DPRU.372) (WEO) and *P. simonii* (DPRU 2430) (WEO) share some rare alleles (Figs. 2, 4, and 5). Burbank had three introductions named with different iterations of the word "Botan." For this reason, the accession

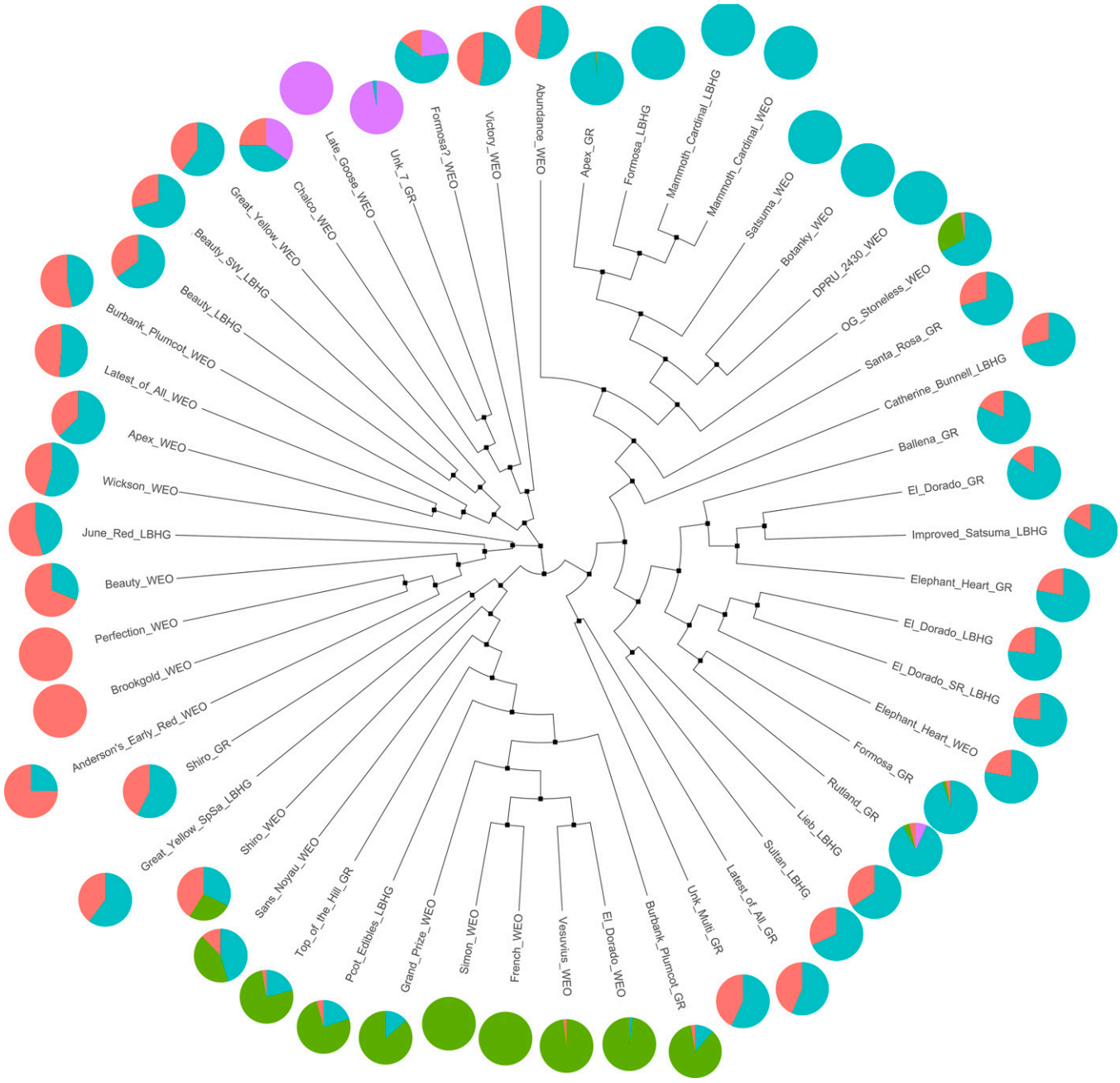

Fig. 4. A circular dendrogram representing the unrooted relationships among 53 Burbank *Prunus* taxa based on identity by descent calculations, combined with admixture pie charts from cluster analysis.

'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) (Figs. 2, 4, and 5). 'Mammoth Cardinal' was patented and introduced by Stark Brothers in 1934, 8 years after Burbank's death (US plant patent 16). 'Mammoth Cardinal' is characterized by its red exocarp, yellow mesocarp, and small, freestone endocarp, indicating visually that it might have a *P. simonii* ancestry (Fig. 1). 'Formosa' was bred by Burbank and then introduced in 1907. It resembles 'Mammoth Cardinal' but has a thicker bloom on its skin (Fig. 1). 'Apex' was touted as an early plumcot introduction, being bred by Burbank and introduced in 1911. 'Apex' is highly prolific, freestone, 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) (Figs. 2, 4, and 5). 'Abundance' and 'Satsuma' were both direct introductions of Burbank's, and all have an admixture comprising almost entirely Cluster 3 (Figs. 3 and 4). '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 an important genotype for increasing yield, producing bountiful crops, although the fruit is highly perishable and thus not ideal for commercial settings. 'Satsuma' was a directly introduction in 1886. The phenotypic influence of this taxon is visible in the red flesh of many of his other plums.

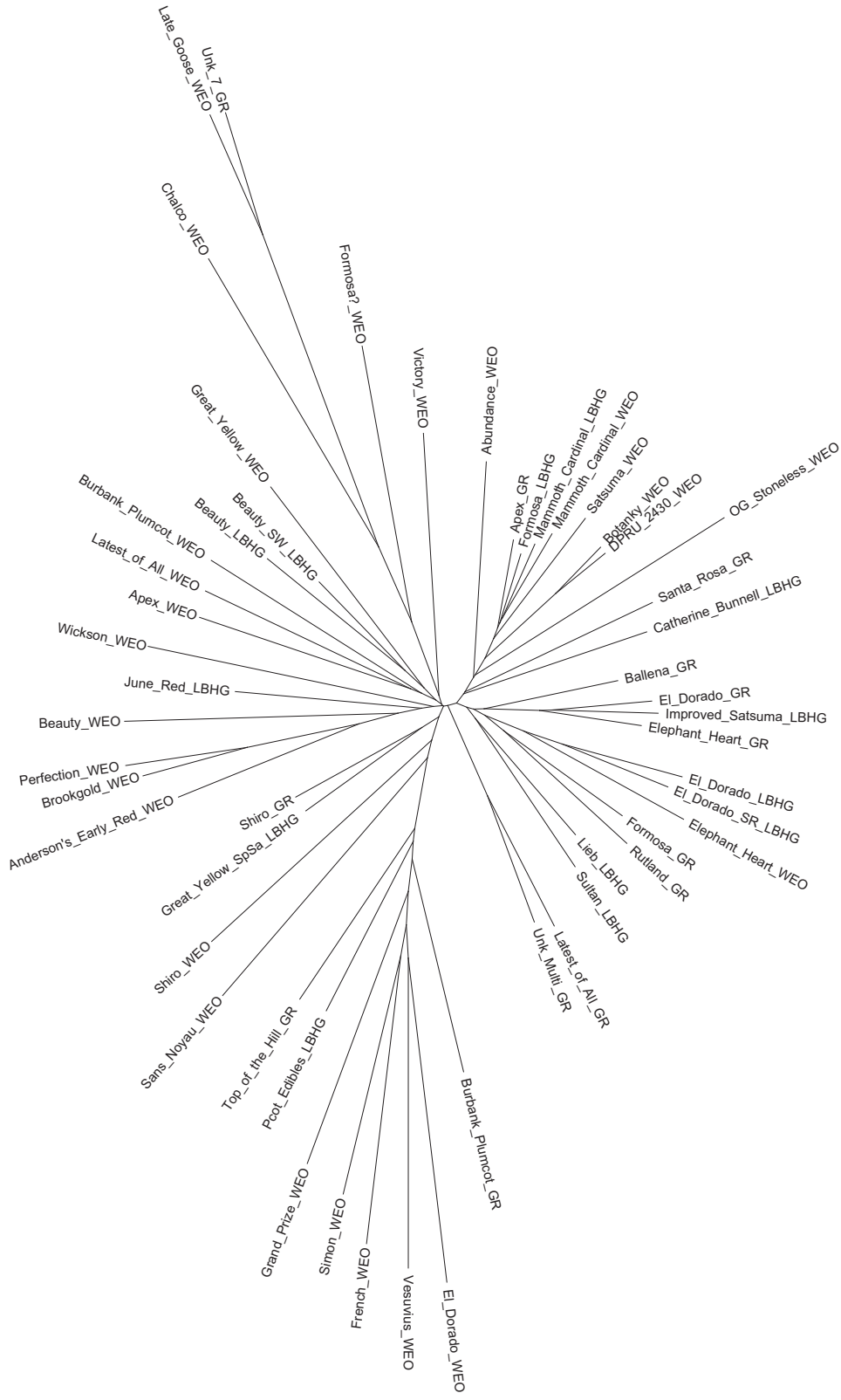

Fig. 5. An unrooted dendrogram of 53 *Prunus* taxa introduced or used in breeding experiments by Luther Burbank based on identity by descent values with branch lengths indicating genetic distance.

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, although 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* (Figs. 2, 4, and 5). 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 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 freestone, rounded plum with black exocarp (skin) and golden mesocarp (Fig. 1). The skin has quite a bit of waxy bloom. 'Elephant Heart' is a pointy-bottomed plum with a thickly waxy, purplish exocarp and a red mesocarp (Fig. 1). 'El Dorado' was introduced by Burbank in 1904 and continues to be a readily available cultivar through scion exchanges and heirloom fruit tree nurseries. It can be picked firm and ripened off-tree, making it ideal for commercial production. 'Elephant Heart' was introduced by Stark Brothers in 1929, 3 years after Luther Burbank died. Phenotypically it has a strong resemblance to 'Satsuma' but is considerably larger.

Conclusions

As a prominent horticultural artist who was derelict in his notetaking, 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 comprehensible approach. In some cases, these data support his claims of parentage; in others they refute his claims or leave more questions to be answered. Although this research is not a definitive pedigree map of Burbank's *Prunus* introductions, it is a valuable baseline to build from. Future work should include a broader panel of *P. salicina* and *P. simonii* to define their genetic differences, which could further parse out the prevalence of species composition in this breeding population.

Identity by descent, PCA, 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 Cited

Arnold JB. 2021. ggthemes: Extra themes, scales and geoms for 'ggplot2'. https://github.com/jrnold/ggthemes. [accessed Jun 2022].

Bernardo R, Murigneux A, Karaman Z. 1996. Marker-based estimates of identity by descent and alikeness in state among maize inbreds. Theor Appl Genet. 93:262–267. https://doi.org/10.1007/BF00225755.

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. https://academic.oup.com/bioinformatics/article/23/19/2633/185151. [accessed Aug 2020].

Brooks RM, Olmo HP. 1952. Register of new fruit and nut varieties 1920–1950. University of California Press, Berkley and Los Angles, CA, USA. http://reader.library.cornell.edu/docviewer/digital?id=chla3174378#mode/1up. [accessed Jun 2018].

Browning BL, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next generation reference panels. Am J Hum Genet. 103(3):338–348. https://doi.org/10.1016/j.ajhg.2018.07.015.

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

Callahan A, Dardick C, Tosetti, R, Lalli D, Scorza R. 2015. 21st century approach to improving Burbank's 'Stoneless' plum. HortScience 50(2):195–200. https://doi.org/10.21273/HORTSCI.50.2.195.

Carrasco B, Diaz C, Moya M, Gebauer M, Garcia-Gonzalez R. 2012. Genetic characterization of Japanese plum cultivars (*Prunus salicina* Lindl.) using SSR and ISSR molecular markers. Cien. Inv. Agr. 39(3):533–543. https://doi.org/10.4067/S0718-16202012000300012.

Ciofi C, Milinkovitch MC, Gibbs JP, Caccone A, Powell JR. 2002. Microsatellite analysis of genetic divergence among populations of giant Galapagos tortoises. Mol Ecol. 11(11):2265–2283. https://doi.org/10.1046/j.1365-294X.2002.01617.x.

Das B, Ahmed N, Singh P. 2011. *Prunus* diversity—early and present development: A review. Int J Biodivers Conserv. 3(14):721–734. https://doi.org/10.5997/IJBCX11.003.

Dreyer P. 1993. A gardener touched with genius: The life of Luther Burbank. 2nd ed. Luther Burbank Home & Gardens, Santa Rosa, CA, USA.

Endelman JB, Jannink J-L. 2012. Shrinkage estimation of the realized relationship matrix. G3. 11:1405–1413. https://doi.org/10.1534/g3.112.004259.

Evanno G, Regnaut S, Goudet J. 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.

Galimba K, Tosetti R, Loerich K, Michael L, Pabhakar S, Dove C, Dardick C, Callahan A. 2020. Identification of early fruit development reference genes in plum. PLoS One. 15(4):E0230920. https://doi.org/10.1371/journal.pone.0230920.

Gao H, Bryc K, Bustamante CD. 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. https://doi.org/10.1371/journal.pone.0090346.

Glowacka A, Sitarek M, Rozpara E, Poswyszynksa M. 2021. Pomological characteristics and ploidy levels of Japanese plum (*Prunus salicina* Lindl.) cultivars in Poland. Plants (Basel). 10(5):884. http://doi.org/10.3390/plants10050884.

Goudet J, Kay T, Weir BS. 2018. How to estimate kinship. Mol Ecol. 27(20):4121–4135. https://doi.org/10.1111/mec.14833.

Hampton JO, Spencer PBS, Alpers DL, Twigg LE, Woolnough AP, Doust J, Higgs T, Pluske J. 2004. Molecular techniques, wildlife management and the importance of genetic population structure and dispersal: A case study with feral pigs. J Appl Ecol. 41(4):735–743. https://doi.org/10.1111/j.0021-8901.2004.00936.x.

Han MV, Zmasek CM. 2009. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics. 10:356. https://doi.org/10.1186/1471-2105-10-356.

Hedrick UP. 1911. Plums of New York. http://chla.library.cornell.edu/cgi/t/text/text-idx?c=chla;cc=chla;idno=5519832. [accessed Jun 2018].

Howard WL. 1945. Luther Burbank's Plant Contributions. University of California College of Agriculture. Bulletin 691. Berkeley, CA, USA. https://www.biodiversitylibrary.org/item/121329#page/1/mode/1up.

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

Karp D. 2015. Luther Burbank's plums. HortScience. 50(2):189–194. https://doi.org/10.21273/HORTSCI.50.2.189.

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

Luby JJ, Howard NP, Tillman JR, Bedford DS. 2022. Extended pedigrees of apple cultivars from the University of Minnesota breeding program elucidated using SNP array markers. HortScience. 57(3):472–477. https://doi.org/10.21273/HORTSCI16354-21.

Qiagen. 2006. Dneasy plant handbook. https://dna.uga.edu/wp-content/uploads/sites/51/2013/12/DNeasy-Plant-Handbook-for-DNA-purification-Qiagen.pdf. [accessed Jan 2019].

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

Pritchard JK, Stephens M, Donnelly P. 2009. Inference of population structure using multilocus genotype data. Genetics. 155:945–959. https://doi.org/10.1093/genetics/155.2.945.

Smith JS. 2009. The garden of invention: Luther Burbank and the business of breeding plants. Penguin Press, London, UK.

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

Topp BL, Russell DM, Neumüller M, Dalbó MA, Liu W. 2012. Plum, p 571–621. In: Badenes M, Byrne D (eds). Fruit breeding. Handbook of Plant Breeding, vol. 8. Springer, Boston, MA, USA. https://doi.org/10.1007/978-1-4419-0763-9_15.

Van Tassel DL, DeHaan LR, Diaz-Garcia L, Hershberger J, Rubin MJ, Schlautman B, Turner K, Miller A. 2022. Re-imagining crop domestication in the era of high throughput phenomics. Curr Opin Plant Biol. 65. https://doi.org/10.1016/j.pbi.2021.102150.

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

Volk GM. 2019. Temperate tree fruits of North America: *Malus* Mill., *Prunus* L., *Diospyros* L., and *Asimina* Adans, p 353–375. In: Greene S, Williams K, Khoury C, Kantar M, Marek L

(eds). North American Crop Wild Relatives, Volume 2. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-97121-6_11.

Wickham H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, NY, USA. https://ggplot2.tidyverse.org. [accessed Jun 2022].

Zeisset I, Beebee TJC. 2001. Determination of biogeographical range: An application of molecular phylogeography to the European pool frog *Rana lessonae*. Proc R Soc Lond, B. 268:933–938. https://doi.org/10.1098/rspb.2001.1600.

Zhebentyayeva T, Shankar V, Scorza R, Callahan A, Ravelonandro M, Castro S, DeJong T, Saski CA, Dardick C. 2019. Genetic characterization of worldwide *Prunus domestica* (plum) germplasm using sequence-based genotyping. Hortic Res. 6:12. https://doi.org/10.1038/s41438-018-0090-6.