# RESEARCH RESULTS FOR THE YEAR 2014: CLASSIFICATION OF WORLD-WIDE PRUNE CULTIVARS THROUGH GENOTYPING AND WHOLE GENOME SEQUENCING.

Chris Dardick[1], Ralph Scorza[1], Ann Callahan[1], Sarah Castro[2], and Ted DeJong[2].

[1]USDA ARS Appalachian Fruit Research Station, Kearneysville, WV.

[2]University of California, Davis CA.

## ABSTRACT

Over the last few decades, the US prune industry has successfully maintained strong and vibrant export markets, particular in Europe. But in recent years international competition has been steadily eroding the share of US world prune exports. US producers have sought to distinguish their prunes and prune products based on their high quality, consistency, and established health benefits. Thus far, consumers have responded positively to these improved labeling and packaging strategies. To maintain its edge, US industries are making significant investments into breeding new and superior varieties. Increased knowledge of the genetic relationships among worldwide prune germplasm would enable new breeding and marketing efforts that could even further distinguish US products. Prune cultivars are known to vary with respect to their fruit quality and in their abundance of health promoting compounds (like sugars and anti-oxidants) thus, marketing characteristics and approvals related to a one variety may or may not be applicable to other varieties. Knowledge of these traits and their genetic basis may help the US industry keep a preeminent position in a changing world marketplace that is experiencing increased competition, tougher regulatory standards, and shifting consumer preferences.

Most international sellers of dried plums typically market them as the ‘d’Agen’ variety. But d’Agen is not a single variety. d’Agen prunes are named for a town in France where they were first planted nearly 800 years ago. These plums, known at that time as ‘date plums’, were probably introduced from Persia (Doyle et al., 2012). Over the centuries, a number of varieties with the name d’Agen were propagated that had distinct characteristics and some of these were imported into the US. Many of these d’Agen types, including several numbered clones of d’Ente and petite d’Agen, are thought to be clonal material of an original d’Agen type; having arisen through mutation of clonally propagated d’Agen trees as opposed to being seedlings. The variety most commonly grown in the US, referred to as ‘Improved French’, is most commonly thought to have been derived from an open pollinated d’Agen seedling released by Luther Burbank around 1900. This variety may have been subsequently improved over the last century through selection of clonally propagated materials.

## OBJECTIVES

The objective of this study was to determine the genetic relationship of ‘Improved French’ to other commercial germplasm that is used worldwide. In addition, wild *Prunus* relatives were also included the study in attempt to determine the origins of *Prunus domestica* which could provide source material for new, valuable traits.

PROCEDURE

The objectives were accomplished using a new technique called Genotype-By-Sequencing (GBS). GBS is method that leverages the power of next generation DNA sequencing technologies to assess the genetic relationships among large numbers of individuals. This strategy reduces the costs of whole genome next generation sequencing using a technique that limits the genome sequencing to a smaller number of informative regions. In this way, the entire genome of each individual is not sequenced by rather a large number of snippets that typically include the regions containing genes. The platform is scalable to 47 or 96 individuals per sequencing run. Here, DNA from 96 individual plum samples was extracted and used for GBS analysis. A very general outline of the procedure is given below:

1) Obtain tissue samples from orchards and germplasm repositories.
2) Extract DNA.
3) Generate barcoded DNA libraries for sequencing.
4) Send libraries to reputable service provider (Illumina HiSeq instrument).
5) Deconvolute resulting data using sample barcodes.
6) Perform trimming and data quality control steps.
7) Assemble sequences from each sample to the reference genome (peach).
8) Identify sequence variations from each assembly (single nucleotide polymorphisms or SNPs).
9) Compare SNP profiles of all samples and generate relationship tree (ie. dendogram).

*The work was performed in cooperation with Dr. Tetyana Zhebentyayev and Dr. Chris Saski at Clemson University, Clemson University Genomics Institute (CUGI). It should be noted that Clemson University donated significant resources to this project in the form of supplies, personnel time, and computational time on CUGI servers.*

RESULTS AND CONCLUSIONS

The following specific accomplishments are reported for 2014:

- Approximately 500 tissue samples were obtained from The USDA ARS National Germplasm Repository, Davis CA, Ted Dejong, University of CA, Davis, Ralph Scorza, USDA ARS Appalachian Fruit Research Station, Kearneysville WV, and the French National Institute for Agricultural Research (INRA), Bordeaux, France.

- GBS libraries were prepared from 96 selected plum samples and used for sequencing.

- GBS data was processed and analyzed.

A summary of all prune samples obtained is shown in Table 1. A list of the 96 samples used for GBS sequencing is shown in Table 2.

### *GBS library preparation and sequencing*

In genotyping by sequencing experiments (GBS) in complex plant genomes, the enzyme choice for genomic selection is critical in collecting the sequencing depth of coverage necessary for dense marker distribution in the genome. Because of the close relationship with peach and the availability of a high quality of the reference assembly, we performed a virtual digest with several enzymes to predict the most appropriate enzyme for use in plum. Here, total genomic DNA was prepared for 96 plum accessions and digested with PstI. Restriction fragments were selected in the 300bp size range, individually indexed and tailed with Illumina sequencing adapters. Post library construction, 96 accessions were multiplexed together and run on two flow cell lanes of an Illumina HiSeq2500 (Illumina) on high output mode using a single-end 1x101bp run cycle.

### *Demultiplexing, filtering, and coverage*

A total of 165,327,337 sequencing reads were collected from 2 lanes of Illumina sequencing, where 250,420,858 (94%) were deemed good, barcoded reads. This suggested high quality DNA, sequencing library preparations, and efficient clustering and sequencing.

### *MergeCount Analysis*

The unique tags from each of the 96 accessions were aligned with each of the accessions to determine a set of unique tags amongst the 96 accessions. These tags were merged into a single file and counted where 1,086,556 tags were covered by 240,663,455 matching reads.

### *Reference alignment to Peach V. 2.0 Genome*

GBS tags were aligned to the peach reference assembly, and a total of 37,953 SNPs were scorable in at least 80% of the cultivars.

### *Data verification*

For data verification, first, we calculated the number of the Pst1 sites in the reference peach genome. The *Prunus persica* genome v1.0 (8 scaffolds) was digested '*in silico*' using the CLC Genomics Workbench software to produce a total of 46.004 restriction sites. Of these, 3,982 fragments that are compatible with the Illumina sequencing technology were predicted in the range of 100-400bp. So, the number of unique tags in individual cultivars (Table 1) is reasonable given a hexaploid configuration of the plum genome and its high level of heterozygosity. Second, using a BLAST analysis we confirmed that sample #64, the *Prunus persica* accession DPRU-520.4, produced 64-bp GBS reads that were identical to the *P. persica* v1.0 genome sequences. Likewise, GBS tags in sample #54 representing the American chestnut genotype KY115 had positive hits against the Fagacea EST database and no hits against peach genome.

The GBS tags derived from plum samples displayed significant similarity against peach genome sequences. Therefore, we concluded that GBS libraries indeed generated accurate tags compatible with the SNP calling pipeline and useful for downstream analyses.

***Dendogram***

SNP profiles for all 96 samples were compared and used to generate a complete dendogram (Figure 1.)

Preliminary analysis of the GBS data and the resulting dendogram produced from reference assembly to the peach genome reveal the following conclusions:

1) The GBS strategy successfully predicted most of the known genetic relationships among plum varieties.
2) All 'French' germplasm forms a single, distinguishable clade.
3) But many d'Agen plum types appear to be seedlings rather than clonally selected materials. This conclusion needs to be verified through the inclusion of additional samples.
4) The CA variety 'Improved French' is related but not identical to d'Agen germplasm.
5) The data supports the hypothesis that *Prunus domestica* originated from a hybridization between *Prunus spinosa* and *Prunus cerasifera*. However, a *Prunus spinosa* only origin cannot be ruled out until more diverse germplasm is sampled.

BUDGET NARRATIVE

Funds in the amount of $25,095 were used to support salary and benefits of a part-time technical position completely devoted to this project. Funds in the amount of $13,300 were used to pay for GBS sequencing and bioinformatic analyses performed by Clemson University researchers.

**Table 1. Summary of tissue samples collected.**

| Plum type | Repository | Number cultivars |
|---|---|---|
| Prunes – Improved French and d'Agen clones | INRA, France | 13 |
| | UC Davis, USA | 4 |
| | USDA, USA | 7 |
| Prunes – traditional cultivars | INRA, France | 120 |
| | UC Davis, USA | 9 |
| Prunes - new varieties/advanced selections | INRA, France | 28 |
| | UC Davis, USA* | 112 |
| | USDA, USA | 20 |
| Plums - Damsons | INRA, France | 15 |
| Plums - Mirabelles | INRA, France | 18 |
| Plums - Greengages | INRA, France | 34 |
| Table plums - new varieties/advanced selections | INRA, France | 105 |
| Related Prunus species | USDA, USA | 12 |
| | Nikita, Crimea | 2 |
| Total | | 500 |

**Table 2. Plum accessions used for GBS including total number of aligned reads and unique tags.**

| GBS #. | Sample Name | Unique Tags | Matching Reads |
|---|---|---|---|
| 1 | Pozegaca-P-24-row2-tree-8 | 72,272 | 959,903 |
| 2 | Blufre | 65,956 | 866,498 |
| 3 | French-prune-row12-tree1 | 105,108 | 1,406,256 |
| 4 | Grand-prize | 220,311 | 3,316,057 |
| 5 | DPRU-814 | 323,671 | 5,020,861 |
| 6 | B2 | 296,541 | 4,484,991 |
| 7 | Primacotes | 138,161 | 1,950,595 |
| 8 | D10S--8 | 130,264 | 1,834,252 |
| 9 | Sans Noyau | 336,795 | 5,339,098 |
| 10 | D18S--50 | 120,503 | 1,789,692 |
| 11 | G11N--39 | 65,296 | 821,869 |
| 12 | F13N--24 | 164,250 | 2,320,819 |
| 13 | Mariana-2624-row1-tree-6 | 121,656 | 1,652,508 |

| | | | |
|---|---|---|---|
| 14 | Prunus-Simonni-row6-tree12 | 137,936 | 2,022,777 |
| 15 | Burja-OZ1111a(Bulgaria) | 183,578 | 2,538,175 |
| 16 | P. spinosa | 156,872 | 2,206,768 |
| 17 | DPRU-848 | 256,856 | 3,876,778 |
| 18 | CI173 | 221,854 | 3,386,024 |
| 19 | Petite-3X | 171,545 | 2,350,873 |
| 20 | E11S--47 | 113,579 | 1,566,867 |
| 21 | G36S--57 | 174,228 | 2,506,714 |
| 22 | D6N-103 | 195,624 | 2,927,844 |
| 23 | 4-6E-6 | 126,157 | 1,839,728 |
| 24 | G5N--35 | 74,197 | 985,110 |
| 25 | Stanley-row1-tree16 | 119,059 | 1,654,480 |
| 26 | Cacanska-Rana | 185,270 | 2,717,929 |
| 27 | Blue-Ribbon | 124,008 | 1,707,711 |
| 28 | P. insititia | 323,636 | 4,828,952 |
| 29 | DPRU-2289.4 | 243,340 | 3,888,441 |
| 30 | Green-Gage | 12,300 | 107,304 |
| 31 | D3-39 | 136,505 | 1,750,677 |
| 32 | Empress | 232,073 | 3,292,175 |
| 33 | Sugar | 412,980 | 7,320,700 |
| 34 | 6--25--43 | 254,257 | 3,643,872 |
| 35 | D2-3-35 | 238,808 | 3,505,553 |
| 36 | F11S--65 | 124,624 | 1,809,223 |
| 37 | Reine-Claude-de-Bavay-row2-tr13 | 230,775 | 3,221,341 |
| 38 | Pearl-row18-tree30 | 154,678 | 2,152,336 |
| 39 | OrtenhauerxStanley-#34 | 263,909 | 4,001,792 |
| 40 | GF-D'ente-652 | 180,753 | 2,778,775 |
| 41 | DPRU-2399.6 | 290,286 | 4,250,254 |
| 42 | GF-d'Ente-2733-(707) | 272,365 | 4,140,480 |
| 43 | D3S-5 | 165,399 | 2,156,347 |
| 44 | Sutter | 230,101 | 3,395,663 |
| 45 | Imperial | 268,580 | 4,121,122 |
| 46 | 5--2--21 | 65,136 | 854,763 |
| 47 | 2-8E-11 | 559,332 | 8,821,437 |
| 48 | D2N-76 | 45,714 | 590,965 |
| 49 | Ruth-Gersletter-row1-tree11 | 159,452 | 2,327,070 |
| 50 | Cacanska-Lepotica-row9-tree3 | 42,917 | 529,857 |
| 51 | Tuleu-Gras-row12-tree1 | 300,598 | 4,586,330 |
| 52 | DPRU-473 | 184,000 | 2,690,331 |
| 53 | HoneySweet | 137,302 | 1,918,791 |
| 54 | KY115-chestnut (“blank “control) | 203,005 | 2,894,248 |
| 55 | 2-2E-38RR | 233,137 | 3,429,720 |
| 56 | E3S--44 | 194,137 | 2,809,663 |

| | | | |
|---|---|---|---|
| 57 | 2-3W-12 | 112,713 | 1,467,718 |
| 58 | P. cerasifera | 311,393 | 5,116,090 |
| 59 | 6--5--62 | 78,025 | 1,027,684 |
| 60 | Burton | 102,337 | 1,415,389 |
| 61 | Bluebyrd-row1-tree9 | 139,719 | 1,914,646 |
| 62 | Early-Italian-Runune-row2-tree3 | 238,058 | 3,376,927 |
| 63 | Anna-Spath | 206,599 | 2,957,176 |
| 64 | **Prunus persica -**DPRU-520.4 | 33,604 | 389,826 |
| 65 | FT-34 | 72,933 | 950,636 |
| 66 | P. brigantina | 393,405 | 4,080,072 |
| 67 | 1-11E-4C | 126,536 | 1,654,467 |
| 68 | Improved-French | 94,347 | 1,313,603 |
| 69 | D18S--12 | 130,696 | 1,862,371 |
| 70 | D8N--15 | 123,481 | 1,761,604 |
| 71 | 3-11W-13 | 150,424 | 2,076,399 |
| 72 | Ramming-Pitless | 108,792 | 1,504,703 |
| 73 | Stanley-Ca-row1-tree-14 | 134,482 | 1,872,881 |
| 74 | Tamjioasa-de-Bristrita-row1-tree4 | 134,637 | 1,998,053 |
| 75 | Italian-Prune-row11-tree6 | 112,780 | 1,606,214 |
| 76 | DPRU-571 | 315,871 | 5,418,161 |
| 77 | **IF-Bud-8** | 3,348 | 23,792 |
| 78 | P. brigantinaxP.cerasifera | 501,131 | 5,624,966 |
| 79 | 3-8E-46RR- | 239,246 | 3,454,000 |
| 80 | Moyer | 162,914 | 2,379,730 |
| 81 | D18S---7 | 189,095 | 2,731,276 |
| 82 | F2N--10 | 152,731 | 2,143,376 |
| 83 | D7N--92 | 137,431 | 1,913,432 |
| 84 | UC-Davis-Pitless-Bush | 151,528 | 2,254,465 |
| 85 | Reine-Claude | 53,001 | 636,739 |
| 86 | Pozegaca-P-25-row1-tree4 | 110,295 | 1,594,966 |
| 87 | Kinstendilsva | 223,495 | 3,265,819 |
| 88 | DPRU-795 | 174,951 | 2,660,578 |
| 89 | A1 | 62,826 | 779,838 |
| 90 | Muir-Beauty | 167,345 | 2,668,724 |
| 91 | Tulaire-Giant | 214,191 | 3,451,996 |
| 92 | Emperor | 282,679 | 4,065,919 |
| 93 | D18S--14 | 251,021 | 3,747,936 |
| 94 | G3N--16 | 94,500 | 1,296,074 |
| 95 | D4N--98 | 189,609 | 2,800,980 |
| 96 | original-pitless-Mitacle | 254,912 | 3,292,273 |

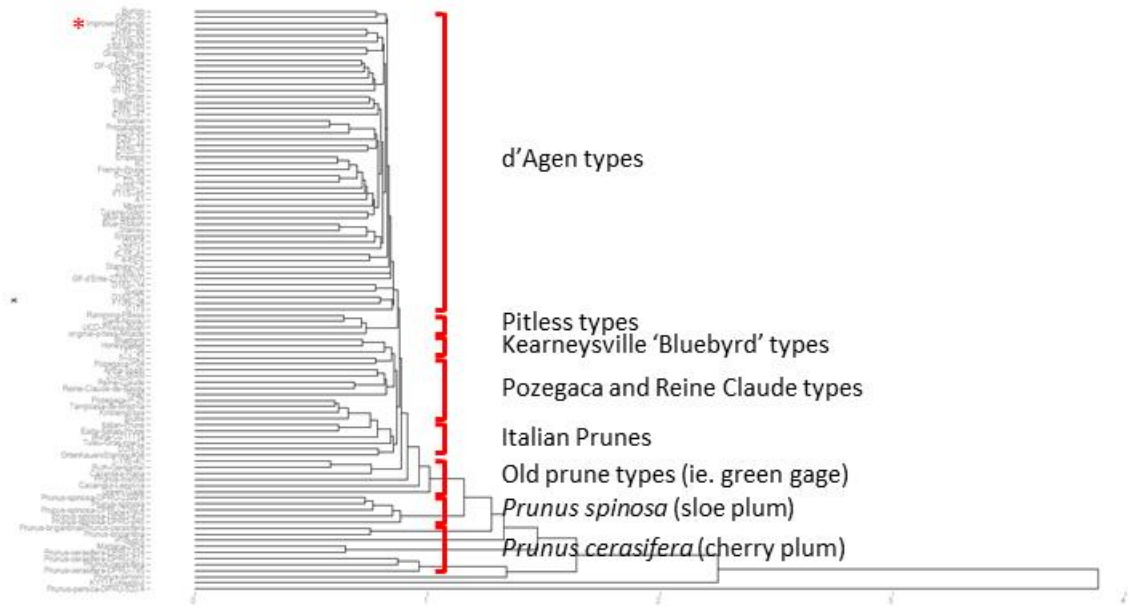


Figure 1. **Dendogram showing the genetic relationships of worldwide plum varieties.** The branching lines represent relatedness among all 96 samples tested. Prunus persica (peach) and *Prunus simonii* (apricot plum) form clear outgroups that are distantly related to *Prunus domestica* (prune plum). *Prunus spinosa* (sloe plum) and *Prunus cerasifera* (cherry plum) samples show a more recent split but both appear to be more closely related to *Prunus domestica* than other *Prunus* species. *Prunus spinosa* is the most closely related *prunus species* tested and may even be the sole progenitor of *P. domestica*. The commercial plum varieties show clear genetic separation according to known or anticipated lineages. 'Improved French' shows a surprising similarity to the cultivar 'Burton' which is an older CA variety selected a decade or two prior to 'Improved French'. The reason for this relationship (Burton being a parent or perhaps offspring having the same or closely related parents) will require additional investigation.