# SEQUENCING AND ASSEMBLY OF THE PRUNUS DOMESTICA CV. IMPROVED FRENCH GENOME.

Chris Dardick, USDA-ARS, Appalachian Fruit Research Station, Kearneysville, WV 25442

## OBJECTIVES

The objectives of this project are to enhance genome assembly of the Improved French genome in order to develop molecular markers useful for germplasm characterization and breeding.

## PROGRESS REPORT

Over the past two years, the USDA-AFRS Genetic Improvement Research Unit has been developing genome sequences for *Prunus domestica* to facilitate molecular characterization of prune genomes and enable marker assisted breeding. To date, extensive sequence coverage has been obtained for two prune cultivars; HoneySweet (a transgenic 'eastern' type prune) and 'Improved French'. Initial genome assemblies have been performed using the closely related peach genome as a template. These assemblies revealed that 87% of the plum genome is shared with peach. Over the shared regions, prune shows approximately 83.2% nucleotide identity to peach. By analyzing mismatched and broken paired-reads we have found that there are numerous local chromosomal rearrangements relative to peach resulting from large numbers of small insertions, deletions, and inversions. These differences suggest that many of the molecular markers developed for peach may not be transferrable to prune.

Using new assembly algorithms, we have made significant progress in assembling the 'Improved French' genome. The largest gains have come from the ability to "scaffold" *de novo* assembled contigs. Scaffolding has enabled assembly of approximately 50% of the prune genome into contigs >750 base pairs (vs. the prior 30%) and has had a significant improvement on average contig length.

We analyzed Single Nucleotide Polymorphisms (SNPs) across the entire prune genome. The relative level of sequence polymorphism in prune was found to be high. Prune was found to contain approximately 10 times the number of SNPs as peach (3.3 million vs 0.3 million). Increased polymorphism is expected given the hexaploid nature of prune. However, the extremely high level of polymorphism suggests that, unlike peach, prune cultivars have not been subject to a high degree of inbreeding.

A set of 30 prune-specific molecular markers were designed based on the prune genome. PCR based markers were designed to flank 3-10 base pair indels (insertion-deletion polymorphisms) at genomic intervals of 5 million base pairs for all 8 chromosomes (Figure 1).

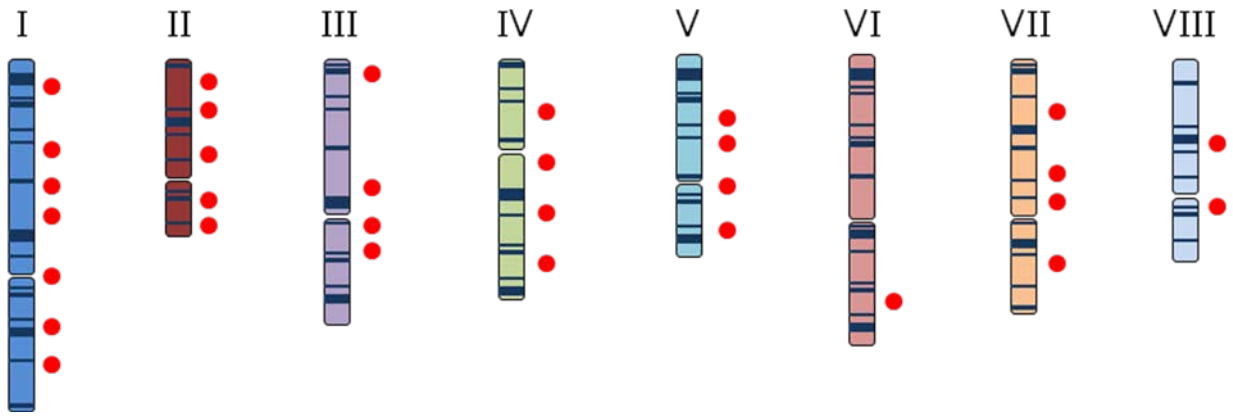


Figure 1. Distribution of molecular markers (red dots) across prune chromosomes.
The markers were chosen based on their ability to distinguish the 'HoneySweet' and 'Improved French' genomes. PCR fragments were distinguished using the High Resolution Melting technique. Resulting melt curves were categorized based on their shape. Marker data obtained from nearly 70 cultivars were analyzed to assess the ability of the markers to distinguish known prune pedigrees.

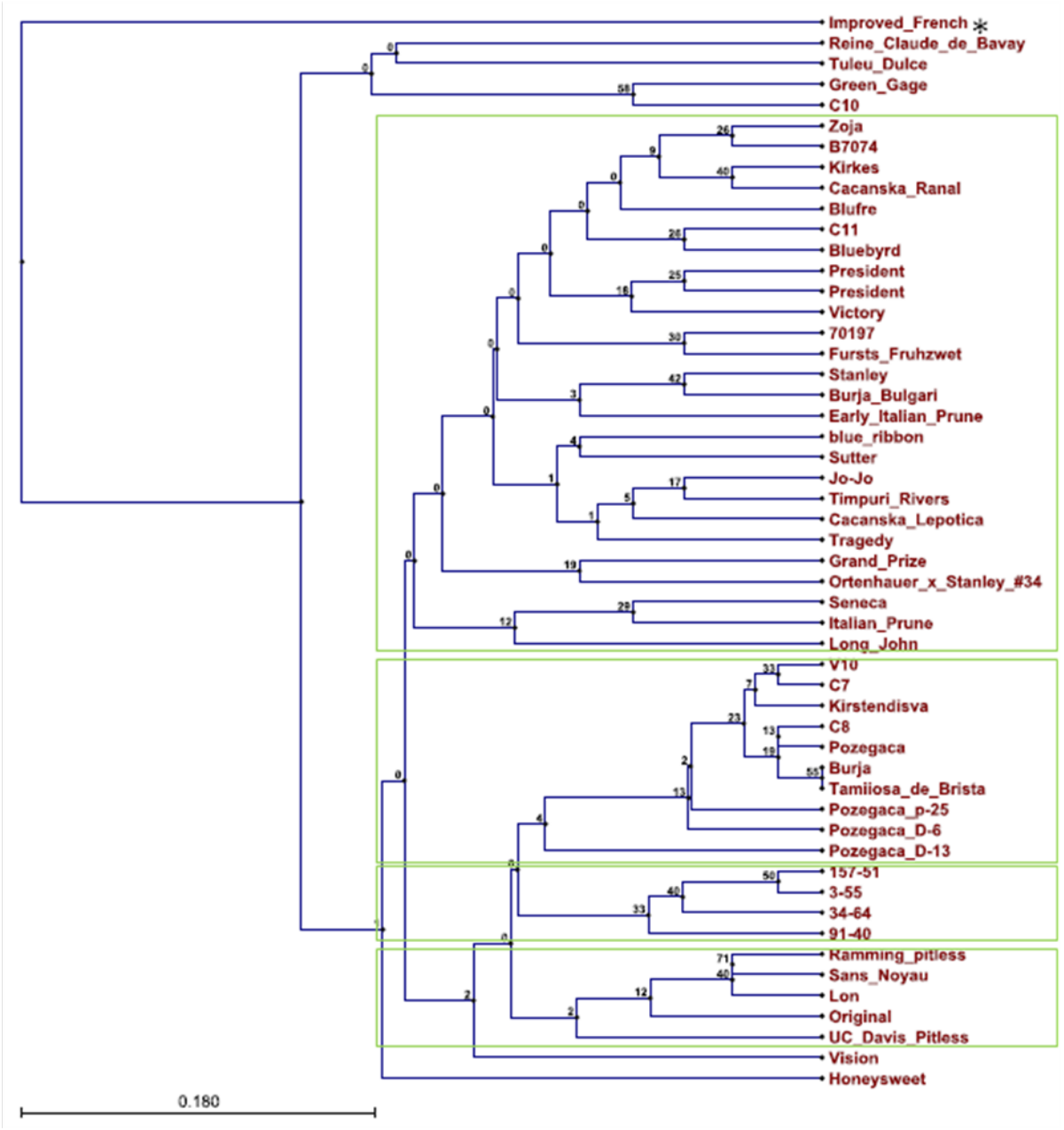


Figure 2. Phylogenetic tree derived from molecular marker data. Cultivar names are indicated to the right.

Results indicated that the markers successfully distinguished cultivars of known origins. For example, Posegaca types and related germplasm formed a distinct cluster. Stoneless types also produced a distinct clade. Surprisingly, 'Improved French' was found to be unrelated to the other cultivars examined. To confirm this finding, the experiment was repeated using a new set of cultivars that included all available 'French' types (Figure 3.)

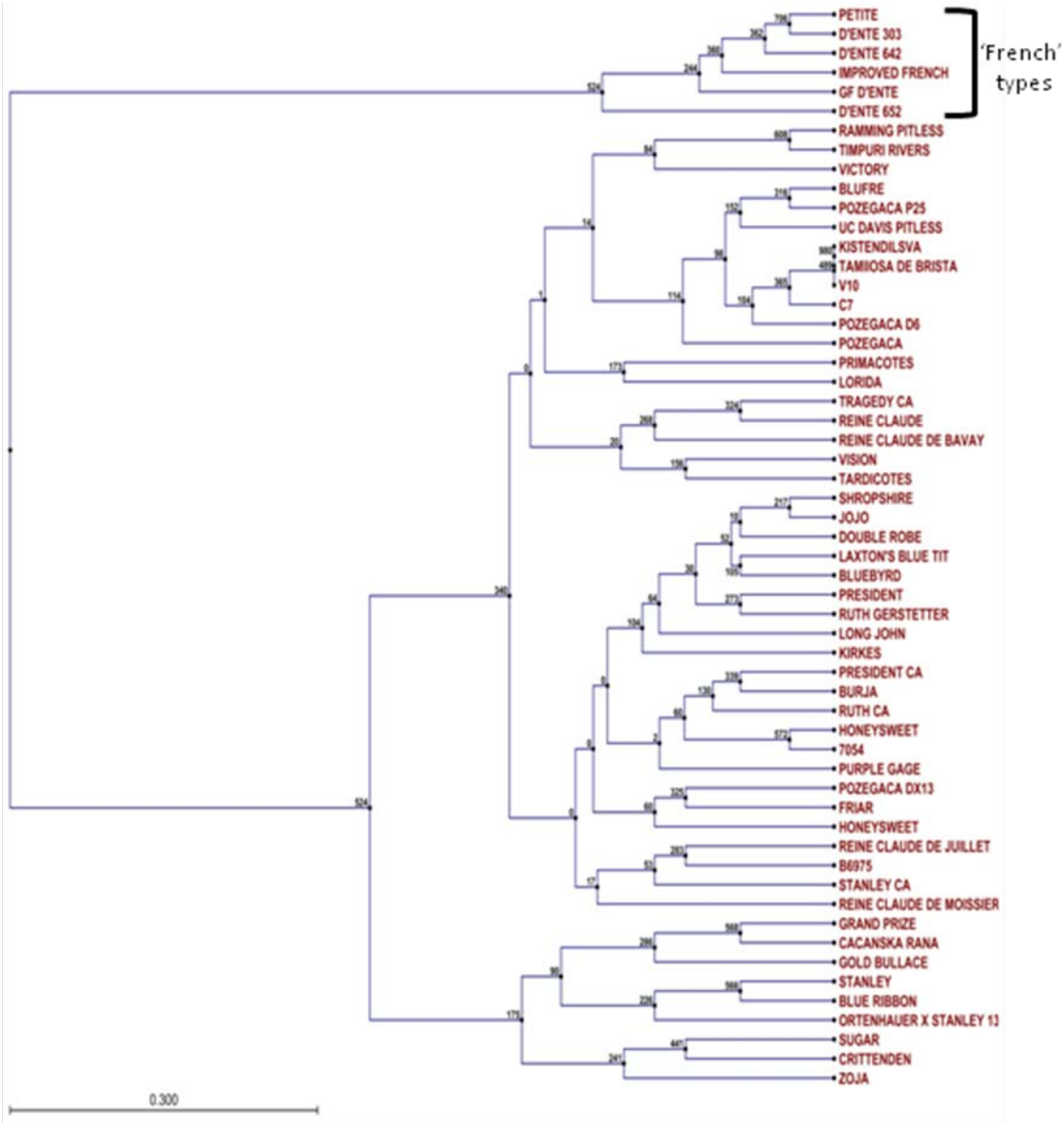


Figure 3. Phylogenetic tree of molecular marker data that includes six known 'French' cultivars (indicated).

The results confirmed our previous finding and suggest that 'French' germplasm is genetically distinct from other European and Eastern US cultivars.

We have also begun using the 'Improved French' genome data to determine the mode of inheritance within this hexaploid species. Prune is thought to have originated from a cross between an unknown tetraploid and a diploid *prunus* species resulting in a hexaploid. How the six chromosomal copies segregate is currently unknown- although genetic segregation data suggests it behaves as a diploid (4x2 or 3x3). We have identified genomic sites where all 6 chromosome copies contain unique sequences; allowing us to track the segregation of all six

copies in the progeny. By using PCR we have shown that the six different alleles can be distinguished.

ONGOING STUDIES

We are currently trying to improve the ‘Improved French’ genome. New DNA samples have been prepared from ‘Improved French’ germplasm and are being sequenced using Illumina mate-pair technology. This technology allows for the sequencing of longer DNA fragments and will facilitate better genome assembly. We anticipate receiving the new genome sequence data before the end of January, 2013. At that time, new genome assemblies will be performed; combining data from all sequencing methods available to date. This should lead to a substantially improved ‘Improved French’ genome assembly.

The molecular markers that we developed and validated are currently being analyzed within our FasTrack breeding program to confirm crosses between ‘HoneySweet’ and ‘Improved French’. These proto-typical studies will serve as the basis for future marker assisted breeding efforts that target additional prune traits of interest to the US prune industry.

Determination of the inheritance pattern for prune is currently underway. A molecular marker for this study has been developed and confirmed. We are now testing the allelic segregation within a population of selfed ‘Improved French’ seedlings. This information will be critical to future marker-assisted and conventional breeding efforts.