Penn State University, Genome Resources and Tools Project (2009 – 2012)
John Carlson, PI

Project Objectives

1. Develop a complete reference genome for Chinese chestnut cv Vanuxem.
2. Identify all the genes in the three blight resistance QTL.
3. Provide gene sequences to the FHI Transgenics and Molecular Breeding groups.

Extended Tasks:

1. Prepare genomic DNA libraries from *Castanea mollissima* Vanuxem cultivar.
2. Sequence Vanuxem genomic DNA to 15X depth using the 454 platform.
3. Produce preliminary *de novo* assemblies of the reference genome sequence.
4. Sequence to optimal depth for assembly.
5. Produce BAC clone end sequences covering the genome.
6. Conduct "final" assembly of the genome sequence.
7. Predict and annotate genes in the assembled genome.
8. Map transcript unigenes to the assembled genome to identify expressed genes.
9. Sequence BAC contigs covering blight resistance QTL regions to great depth.
10. Identify all genes in QTLs and candidate blight resistance genes.
11. Conduct comparative genomics studies with peach and other model plant genomes to narrow list to most-likely candidate genes for testing.
12. Re-sequence American chestnut.
13. Submit publications on the Chinese chestnut genome.
14. Provide public access to the genome and tools through a public web portal.

The Chinese chestnut genome – summary:

Carlson and coworkers obtained two versions of the Chinese chestnut 'Castanea mollissima' during the Forest Health Initiative project. In the first version obtained with genomic DNA from clones of the TACF Chinese chestnut genotype Vanuxem maintained at Clemson University, we produced over 60 Gigabases of sequence data, which assembled into 587M bases in 51,766 scaffolds, covering app. 66% of the estimated 880Mb genome size. In this assembly we identified 66,662 possible genes, including 133 known disease resistance genes.

The second version of the Chinese chestnut genome was obtained with genomic DNA isolated directly from cuttings from a Vanuxem tree in the TACF breeding orchard. Over 55 Gigabases of new sequence data was prepared with updated "Next Gen" sequencing machines that provided longer sequences, matched ends, and low error rates. Using a new assembly program designed for heterozygous genomes, we assembled the new data into 41,270 scaffolds containing 724.4 Mp of sequence, covering app. 90% of the estimated genome size. Using new gene-prediction programs and RNA sequence validation, 38,268 genes were predicted in the new genome assembly, which is much closer to expectations than obtained with the first set of data.

In addition, the three blight-resistance QTL were separately sequenced to greater depth than the rest of the genome. BAC clones covering each of the three blight resistance QTL were selected from the Chinese chestnut physical map from which a total of 518.6 Mp of sequence was generated and assembled into scaffolds for each QTL. The QTL cbr1 was assembled into 214 scaffolds in which 432 genes were identified. The QTL cbr2 was assembled into 128 scaffolds containing in which 219 genes were identified. The QTL cbr3

was assembled into 53 scaffolds in which 131 genes were identified. The QTL genes were annotated for putative functions and the QTL were compared to disease resistance loci in the peach genome. Fifteen genes with the functional annotation of "defense response" genes were selected as the most likely genes for blight resistance. A manuscript on comparison of the structure of the Chinese chestnut genome and QTL sequences with the model peach genome has been prepared for publication. Genome Browsers for both the genome and the 3 QTLs were constructed.