

RESEARCH GRANT REPORT TO
THE AMERICAN CHESTNUT FOUNDATION

**DEVELOPMENT OF *CASTANEA PUMILA* SPECIFIC SINGLE NUCLEOTIDE
POLYMORPHISMS TO DETECT THE OCCURRENCE OF HYBRIDIZATION
BETWEEN AMERICAN *CASTANEA* SPECIES**

Fenny Dane

Department of Horticulture, Auburn University, AL 36849

danefen@auburn.edu; 334-844-3046

OBJECTIVES

1. Design of high quality PCR primers based on 454 sequence analysis of *C. pumila* for the detection of single nucleotide polymorphisms (SNPs) in single copy genic regions.
2. Genotyping of DNA samples from each of the North American species (*C. dentata* and *C. pumila* var. *pumila* and var. *ozarkensis*) using primers to test for species specificity and hybridization between species
3. Analysis of population allele frequencies and genetic diversity of southern American *Castanea* populations.

MATERIALS AND METHODS.

Amino acid sequences of single copy genes of *C. pumila* (obtained via 454 transcriptome sequencing by Dane and Li) and *C. dentata* (downloaded from the fagaceae.org database) were aligned to their core orthologs from *Glycine*, *Oryza* and *Arabidopsis*. Unigene contig protein assemblies were compared and analyzed for sequence variations or putative mutations. Subsequently DNA contigs of protein assemblies with putative mutations were analyzed for the presence of SNP sites derived from either *C. dentata* or *C. pumila* reads. Contigs with SNPs were reblasted into the *C. pumila* data file and fagaceae.org database and verified via PCR and Sanger sequencing following the design of primers around putative SNP sites. Fragments were aligned using Vector NTI. Species specificity of the SNP sites was evaluated using an array of DNA isolated from many different American chestnut and Allegheny chinkapin trees. DNA was isolated from *Castanea* leaves or nuts using standard isolation procedures.

RESULTS AND CONCLUSIONS

The *C. pumila* cDNA library from the leaf transcriptome (following 454 sequencing) generated a total of >1.2 M reads (372 Mb of DNA) with an average read length of 305 bp and a total of 47565 contiguous sequences or contigs, with an average length of 670 bp (Li, 2011). Transcripts were annotated and classified into three general categories associated with cellular, molecular and biological functionalities. The biological processes constituted the most abundant component (42.7%) followed by the cellular component (28.1%) and the molecular function component (29.2%). While comparing the functionality of Allegheny chinkapin genes with those of the American and Chinese chestnut (Kremer et al., 2012), high levels of similarity in gene

function were detected. The *C. pumila* leaf transcriptome showed high transcript abundance for genes related to phytohormone signaling, induced systemic and systemic acquired resistance as well as several well-known disease resistance genes such as peroxidases and laccases. A similar set of defense related genes are thus expressed in *C. pumila* leaves as in *C. dentata* and *C. mollissima* healthy and canker tissues (Barakat et al., 2009; 2012).

A large fraction of the *C. pumila* contigs matched chloroplast genes (20.8%) and resulted in >450 x coverage of the chloroplast or cp genome. The cp genome of *C. pumila* was assembled using the *C. mollissima* cp genome (HQ336406; 160,799 bp) as a reference. For short sequence gaps, primers were designed from *C. pumila* reads, for long sequence gaps the reference cp genome of *C. mollissima* was used. Gaps were filled in using PCR amplification and Sanger sequencing. The *C. pumila* chloroplast genome of 160,596 bp was assembled. Very high sequence homology was detected between the *C. pumila* and *C. mollissima* cp genomes, with a mutation rate of only 0.251 %. Mapping of the *C. dentata* transcriptome, down loaded from the fagaceae.org website, resulted in a *C. dentata* cp genome with ~93% coverage. Comparative analysis between *C. mollissima*, *C. dentata* and *C. pumila* chloroplast genomes indicated that *C. pumila* and *C. dentata* share ~58% of the mutations, while 22% of the mutations are *C. pumila* specific, and 20% *C. dentata* specific, thus each species has accumulated uniquely different evolutionary fingerprints. Ancestral chlorotypes are defined as those which exhibit the smallest number of substitutions in a sample and are related most closely to an outgroup sequence. It is clear that the *C. dentata* D chlorotype has accumulated several unique deletions in intergenic regions of the cp genome which are not detected in either *C. pumila* (with the exception of one population in Georgia) or other *Castanea* species. Thus this *C. dentata* D chlorotype must be evolutionary recent.

Conserved single copy genes, which are truly shared in single copy throughout seed plants are ideal phylogenetic markers. Because of pervasive gene duplication in angiosperms, it is possible to obtain single copy genes shared among several different species using HaMStR core orthologs and *Arabidopsis* as the primer taxon. The resulting orthology groups were aligned and studied for putative informative mutations. Contigs with SNPs were reblasted into the *C. pumila* database (in linux) and fagaceae.org database for verification. Following potential SNP identification in *C. pumila* reads, more than 150 primer pairs were designed using Vector NTI software around SNP sites. PCR amplification and Sanger sequencing was used for verification. Problems were encountered either with PCR amplification, detection of homology or heterozygosity following sequencing, or discovery of pyro-sequencing errors either in the *C. pumila* 454 reads or *Castanea* reads submitted to the fagaceae.org database. Also, several SNPs turned out not to be species-specific.

One amino acid assembly C27816, which codes for a region of the ribonucleotide reductase protein 1 large subunit, contains a non-synonymous amino acid replacement as a result of one SNP (G-A) at an exon site. *C. dentata*, *C. mollissima*, *Quercus* have the AAA codon (lysine), *C. pumila* the AGA (arginine) codon. This SNP is consistent among different *C. pumila* and *C. dentata* populations.

Figure 1. Sequence alignment of a section of a ribonucleotide reductase gene showing the site of a non-synonymous mutation in an exon in *C. pumila* at position 164.

PUBLICATIONS

Dane, F., X. Li, K. Kocot, and L. Goertzen. 2013. Comparative analysis of the leaf transcriptome of *Castanea pumila* var. *pumila*. *Acta Horticulturae* (in press)

Li, X and Dane, F. 2013. Comparative chloroplast and nuclear DNA analysis of *Castanea* species in the southern region of the USA. *Tree Genetics and Genomes* 9: 107-116.

REFERENCES CITED

- Barakat, A, D.S. Diloreto, Y. Zhang, C. Smith, K. Baier, W.A. Powell, N. Wheeler, R. Sederoff, and J.E. Carlson (2009) Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biology* 9:51
- Barakat, A., Staton, M., Cheng, C.-H., et al., 2012. Chestnut resistance to the blight disease: insights from transcriptome analysis. *BMC Plant Biology* 12: 38.
- Kremer, A., Abbott, A.G., Carlson, J.E., Manos, P.S., Plomion C., Sisco, P., Staton, M.E., Ueno, S., Vendramin, G.G. 2012. Genomics of Fagaceae. *Tree Genetics and Genomes* 8: 583-610.
- Li, X. 2011. DNA fingerprinting of *Castanea* species in the USA. MS thesis submitted to the Graduate School of Auburn University, Auburn, AL
- Shaw, J. Craddock, J.H., Binkley, M.A. 2012. Phylogeny and phylogeography of North American *Castanea* Mill. (Fagaceae) using cpDNA suggests gene sharing in the southern Appalachian (*Castanea* Mill., Fagaceae). *Castanea* 77: 186-211.