

FINDINGS:

NEW FINDINGS: June, July & August 2007 (Y1Q4).

1. New crosses were made to expand the size of the mapping populations for Chinese chestnut (Mahogany x Nanking) and for the cross of two F1 hybrids.
2. Eleven new cDNA Fagaceae libraries were constructed from different tissues and species and partially characterized to improve efficiency of ABI (Sanger) sequencing.
3. BAC high information content fingerprinting (HICF) has been carried out for 38,784 clones of the HindIII library of the Mahogany genotype. A new EcoRI library has been constructed.
4. 454 sequencing has now completed 144 million base pairs of Fagaceae sequence. BLAST analysis has been done and over 10,000 unique genes have been identified and annotated for the Fagaceae.
5. A panel of 16 trees from different individual and species of Fagaceae has been used to screen 153 SSR polymorphisms within and between species.
6. The chloroplast genome of Chinese chestnut has been sequenced.

MAPPING POPULATIONS: (Sandra Anagnostakis CAES/ Fred Hebard TACF)

During the summer, repeat crosses were made of 'Mahogany' Chinese chestnut with 'Nanking' Chinese chestnut pollen, and of two American x Chinese ('Mahogany') hybrids with each other. The nut yield will not be known until mid-September.

cDNA LIBRARY CONSTRUCTION (SUNY-ESF: William Powell / Kathleen Baier). Quarter 4 Progress: Eleven new cDNA libraries were constructed using directional cloning, non-normalization, and size fractionation of samples. These libraries typically yielded fragments ranging from 0.5 to 3Kb. These were fractionated on a column with a 400bp cut off, and selected fractions were pooled into 3 libraries before cloning for immature male flower, or 2 size classes for the rest of the libraries. Selected primary libraries were sampled for average insert size which ranged from 710bp to 910bp. Individual inserts ranged from a low of 400bp up to 1800bp. Our average insert sizes for the cDNA libraries fall within the range of mRNA sizes for plants. We will search our sequences for translational start sites to open reading frames that match known genes to confirm that the inserts are full length. Nine new libraries have been sent to Clemson for full-length cDNA sequencing: three primary libraries from CC (Chinese Chestnut) Mahogany immature flowers; two libraries from CC Mahogany stem (pooled unwounded, wounded, and infected stem); two libraries from CC seedling leaves; two libraries from CC seedling roots. We also have two American beech libraries (infected and uninfected) and one American chestnut completed, which will be sent soon to Clemson with the amplified Chinese chestnut libraries. All libraries are currently being amplified for permanent storage.

Proposed: Yr2Q1: This report concludes the primary responsibility on this project, with the exception of manuscript development, of our colleagues at SUNY-ESF.

**BAC LIBRARY CONSTRUCTION / FINGERPRINTING / cDNA SEQUENCING.
(Clemson: Jeff Tomkins)**

Quarter 4 Progress: BACs - The HindIII BAC library (10X coverage) has been completed and is currently serving as the substrate for fingerprinting. The second BAC library has been cloned using the restriction enzyme, EcoR1 and has been frozen off as glycerol stocks. We are awaiting the arrival of a -80 freezer to commence plating and picking of the library.

Fingerprinting: The project goal is to fingerprint approximately 150,000 clones, or ~75,000 clones per 10X BAC library. To date, 38,784 of the HindIII BAC clones (~26% of the total) have been fingerprinted putting us slightly ahead of schedule (see Gantt Chart, Supplemental Materials).

cDNA sequencing: Early testing of recently created, directionally cloned cDNAs show significantly improved success rates relative to the first cDNA libraries created during the first quarter of the project (based on test runs of 96 clones per library). We are still in the process of testing for insert size and homology. Preliminary success rates for HQ reads > 100 b are between 80 and 97%.

Proposed: YR2Q1: Individual cDNA libraries will be evaluated for quality. Ideally, equal numbers of clones will be sequenced during this quarter from each of a minimum of four libraries for a total of 10,000 clones. Fingerprinting will continue in full-scale production mode. Plating and picking of the second BAC library should be completed during the quarter.

454 SEQUENCING (Penn State: John Carlson)

FLX sequencer results for three additional tissue samples have been obtained this quarter. Quarter plates were completed for samples ACHS1 (American Chestnut healthy stem tree 1), ROB (Red Oak root cDNA) and WOB (White Oak root cDNA). For sample ROB, over 16 Mbases of high quality sequence were obtained. However for the WOB and ACHS1 samples sequence yields were not good indicating that these two cDNA preps were sub-optimal. Furthermore the yield and quality of cDNA for the samples ACHS2, ACWP1 (American chestnut whole plant 1), and ACWP2, was insufficient for sequencing. Problems appear to result from shipment of RNA. New whole tree American chestnut tissue samples were collected at the Connecticut Agricultural Research Station in July. Two American beech tree samples (ABWP1 and ABWP2) were provided by Jennifer Koch of the USFS Delaware Ohio Research Center. In the first 4 quarters of the project we have obtained over 144 Million base pairs of 454 sequence. The original target for 454 sequencing, using the GS20 model sequencer, was 216Mbp. With the FLX sequencer the target has been increased to 475Mbp of total 454 data comprised of 6.5 plates of FLX sequence, in total, plus the two GS20 runs conducted before the upgrade. By the end of this quarter we will have finished 3 ¼ plates of FLX 454 sequence plus the two GS20 plates. This will put us about one third of the way to our goal.

BLAST RESULTS FOR ALL FAGACEAE CONTIGS (Chris Smith, NCSU).

	Hits to Populus trichocarpa Predicted Proteins at an e-value 10-15
Chinese chestnut	7719
American chestnut	1067
Red oak	4985
White oak	3525
Total Fagaceae	10,562

Proposed: YR2Q1: New cDNA preps and 454 libraries are being prepared which will be ready for 454 sequencing in the first quarter of year two. We anticipate finishing the remaining 454 sequencing by the third quarter in year two.

PRIMER/MARKER DEVELOPMENT AND SCREENING (SIFG: Tom Kubisiak).

Plant materials for primer/marker screening: A panel of DNAs was assembled for use in all future primer/marker screening work. The panel consists of 16 trees from an array of Fagaceae species, including genotypes that are of direct interest to various researchers around the US and Europe:

- Two *Castanea mollissima* trees ['Nanking' and 'Mahogany', parents of an intraspecific Chinese chestnut mapping cross developed by The American Chestnut Foundation (TACF) and the Connecticut Agricultural Experiment Station (CAES)]
- Two *Castanea dentata* ('GMBig' and 'Horn', parents of an intraspecific American chestnut mapping cross developed by TACF)
- Two *Castanea sativa* trees ('Bursa' and 'Hopa', parents of an intraspecific European chestnut cross developed by CNR in Porano, Italy)
- Two *Quercus robur* trees ('3P' and 'A4', parents of an intraspecific pedunculate oak cross developed by INRA in Pierroton, France)
- Two *Quercus rubra* trees ('Sm1' and 'Sm2', parents of an intraspecific northern red oak cross developed by the University of Notre Dame)
- Two *Fagus grandifolia* trees ('Fg1504' and 'Fg1506', parents of an intraspecific American beech cross developed by the USDA Forest Service in Delaware, Ohio).

The intent of the screening panel is to investigate the transferability of previously published markers, as well as markers that will be developed throughout the course of this project, against a wide membership of the Fagaceae. We have selected genotypes that are parents of pedigrees currently being investigated by various researchers around the world, to facilitate future comparative analyses across the Fagaceae.

Plant materials for genetic mapping in *Castanea dentata* and *Castanea mollissima*: To date, plant materials (leaves) for both parents and 182 progeny from the intraspecific American chestnut cross being developed by TACF, as well as both parents and 176 progeny from the intraspecific Chinese chestnut cross being developed by TACF and CAES have been received for DNA extraction. The American chestnut samples have been processed into 96-well extraction plates and are being stored at -70C in preparation for DNA extraction. The Chinese chestnut leaf samples are also being stored at -70C. Once DNA has been extracted, a battery of 6-10 polymorphic SSRs will be used to assess the correct parentage of all progeny. Any progeny harboring non-parental alleles, i.e., possible pollen contaminants, will be discarded. Leaf material from more progeny of both interspecific and intraspecific crosses will be sent at a later date.

Primer/Marker Screening: To assess the usefulness, for this project, of SSR primer pairs developed for other members of the Fagaceae, a total of 142 previously published primer pairs were screened against the panel of DNAs described above. These primer pairs were developed from different Fagaceae species including: 30 from *Castanea sativa*; 15 from *Castanea crenata*; 32 from *Quercus robur*; 16 from *Quercus petraea*; 20 from *Quercus rubra*; nine from *Quercus myrsinifolia*; four from *Quercus macrocarpa*; six from *Fagus sylvatica*; and 10 from *Fagus crenata*. PCR was performed using fluorescently-tagged primers and amplified products were separated on an ABI 3130XL. Data for 120 of the 142 primer pairs/markers has been subject to automated analysis and is currently being manually edited and summarized. Product for the twenty-two remaining primers is in queue for separation and analysis.

Primer/Marker Development: In order to assess the potential usefulness of our 454 sequence data for identifying and developing polymorphic SSR markers for use in the intraspecific Chinese chestnut mapping cross, a small subset of 454 contigs (~30) containing SSR motifs were identified. From these 30 contigs, 11 had sufficient length and depth of read and appeared to be polymorphic within at least one of the Chinese chestnut mapping parents. Primer pairs were developed for all eleven of these repeat containing contigs. Seven pairs were synthesized and screened against the panel of DNAs described above. Product for these seven primers is currently in queue for separation and analysis. We hope that the results from this small preliminary primer screen will provide us with a rough estimate of our conversion rate from putative 454-based SSR polymorphism to confirmed SSR polymorphism based on fragment analysis of amplified PCR product. This will be useful for determining how many primer pairs may need to be developed and screened to obtain a set number (500) of useful polymorphic markers for genetic mapping.

Proposed: YR2Q1: We will complete screening of the remaining previously published primer pairs. The primer evaluation and marker development pipeline will be ramped up with a target of 50 new, polymorphic SSR markers by the end of the quarter.

CHESTNUT CHLOROPLAST GENOME SEQUENCED (Jeff Tomkins, Clemson).

This work was not in our original plan, but was carried out a part of another project in collaboration with Henry Daniell at the University of Florida, because this genome resource became available. The recently created HindIII BAC library was screened with a series of highly conserved chloroplast gene probes and the resulting positive clones were characterized and evaluated using BAC fingerprinting in addition to intact clone evaluation via PFGE. A single candidate BAC clone containing the complete chloroplast genome was shotgun sequenced to ~10 fold redundancy. Gaps were closed with primer walking. The sequence is currently being annotated and will be compared with chloroplast sequence from American chestnut and other plant taxa. The intent is to sequence the complete American chestnut chloroplast genome next for comparative purposes. Complete chloroplast genomic sequence is important for phylogeny based studies and applied biotechnology applications in chloroplast transformation. The genomic resources created by this project are already paying dividends.