PROPOSAL TO
THE AMERICAN CHESTNUT FOUNDATION
RESEARCH PROGRAM

A. TITLE: "Assembly of the Mahogany genome using a new chromatin proximity approach."

B. SUMMARY

A new, quick, and inexpensive approach for developing reference genomes is available that uses DNA sequence data from genome-wide chromatin interactions to assemble high quality reference genomes with chromosome-length sequences. Although this technology is new, a comprehensive service based on this approach is already commercially available. The applicant has just received encouraging results for assembly of a tree species' genome using this new chromatin interactions approach by the Dovetail Genomics Company. We propose to test this approach for chestnut by working with Dovetail Genomics to build a reference genome for the important Chinese chestnut cultivar 'Mahogany.'

C. PRINCIPAL INVESTIGATORS AND INSTITUTION:

John E. Carlson, Professor
Department of Ecosystem Science and Management
The Pennsylvania State University
University Park, PA. 16802

D. ESTIMATED DURATION: 12 months from the time funding is received.

E. TOTAL AMOUNT REQUESTED:

Total Project Request: $10,000 to Penn State University.

F. SHORT AND LONG TERM GOALS OF THE PROJECT:

The **ongoing, long term goal** of our research is to provide tools and resources to help TACF advance its breeding program through marker-assisted selection at the whole-genome level, and to better understand mechanisms of blight and root rot resistance in TACF's breeding program.

Our **short-term goal** is to test a new technique for rapid construction of genomes that, if successful, could be applied to parental and hybrid genotypes across the TACF program. The Mahogany genome will also support detailed comparisons of resistance genes in TACF material.

G. NARRATIVE

Jared Westbrook and Jason Holliday are developing genome-wide selection (GWS) methods (Jannink et al 2010) to simultaneously accomplish selection at the DNA level for the specific QTLs responsible for blight resistance and Phytophthora resistance along with selecting for the overall American chestnut genetic background in individual progeny of current and future backcross generations. This will provide a rapid and precise approach to simultaneously select for the many loci for disease resistance and the American tree genome to augment phenotyping.

Genomic selection methods use genome-wide coverage with DNA markers to ensure that all regions of the genome are taken into account for effect on phenotype. DNA sequences from tissue samples of individual parent and progeny plants are mapped onto a reference genome and

then the results scored to reveal which parental alleles, and therefor traits, the progeny have inherited. Jared and Jason Holliday are developing a bioinformatics 'pipeline' to automate computing of the genotyping results. The success of this GWS approach depends on the quality and coverage of the reference genome.

We released to the public a draft reference genome providing whole genome coverage for the Chinese chestnut cultivar 'Vanuxem' in 2014 (Staton et al, 2014). The 800 million nucleotide genome was assembled into app. 40,000 large pieces, called scaffolds, and genes within each scaffold were identified. Such draft genomes are suitable for identifying genes and alleles, but not ideal for genome-wide assessment of introgression of the Chinese and American genomes among individuals in backcross families if the position of the scaffolds in the genome is uncertain. Over the past 3 years the Carlson lab was able to order the arrangement of genome pieces from our first draft of the Vanuxem genome assembly into contiguous sequences covering each of the 12 chromosomes, using the known positions of sequences from genetic linkage maps and physical maps for Chinese chestnut. The assembled 'pseudochromosome' sequences, ranging in size from 47 Million bases to 81 Mbases, collectively cover over 90% of the genome. In a current project funded by the USDA to TACF, long DNA sequences were collected from the Vanuxem genome using a technology abbreviated as "PacBio". This data allowed us to confirm the arrangement of sequences within each pseudochromosome, and to fill in most of the short gaps that remained between scaffolds.

The genome for Chinese chestnut cv Vanuxem, is now being used to map gene sequences and DNA markers, such as SNPs (Single Nucleotide Polymorphisms) from other chestnut genotypes, including other species and hybrids. Our genome for Chinese chestnut is based on only the Vanuxem cultivar however, which is not a key parent within the TACF backcross breeding scheme. Questions remain, such as: How similar is the overall structure of the Vanuxem genome to the genomes of other Chinese chestnut genotypes? Are there translocations of chromosome ends between chromosomes that differ among genotypes and that might not be diagnosed with the one reference genome? Do the blight resistance and root rot resistance QTLs contain the same blocks of genes in the same locations in different Chinese chestnut genotypes? Have there been any rearrangements or losses of genes within individual chromosomes in different Chinese chestnut genotypes that could not be identified with only one reference genome? Are there stronger resistance alleles from parents used in the TACF backcross breeding pedigree that we are missing using only the Vanuxem genome as reference? Such questions are best answered by constructing high quality genomes for additional Chinese chestnut parental genotypes and for hybrids used as parents within the TACF pedigree. However, the traditional approach that we took to produce the genome for cv Vanuxem, referred to as 'de novo' assembly followed by painstaking piecing together of scaffolds into chromosome-length sequences, seemed prohibitive even with access to recent DNA sequencing technologies, like PacBio (Claros et al 2012).

Fortunately, a new genome construction technology was recently announced that promises to reduce the time and cost of obtaining a high quality genomes. This new approach, called "chromatin proximity guided" genome assembly, is quicker and much less expensive for developing reference genomes (Burton et al 2013, Dekker et al 2013, Kaplan and Dekker 2013). It is based on the discovery that regions within chromosomes tend to reside in preferred locations within nuclei. This allows the location of DNA sequences within a chromosome to be reproducibly identified relative to the position of other genes or regions within the same chromosome. The technology involves using a 'fixative' to cross link chromatin (DNA and the proteins associated with that DNA) to be linked to each other and to other chromatin that is

adjacent to it in the nucleus. Extracting and then sequencing all of the cross-linked pieces of DNA from many nuclei provides a means to jump long linear distances to identify which individual sequences belong to the same chromosome and where they are located relative to each other within the chromosome. From the large data sets of genome-wide paired DNA sequences that are produced from the fixed chromatin, it is possible to proceed directly to assembling high quality chromosome-scale sequences as reference genomes.

Although this technology is new, commercial services based on this approach are already available, for very reasonable prices relative to the high costs of traditional genome sequencing and assembly approaches. The approach follows 3-4 steps. First, the customer provides a draft quality (preliminary) "de novo" assembly that was based on producing contiguous sequences of 150 bp to 5,000 bp in length by overlapping short-read DNA sequences (100 to 150 nucleotides). The customer also provides the company with partially purified high molecular weight DNA and/or leaf tissues from which nuclei will be prepared. The company prepares libraries from fragments of DNA obtained from chromatin cross-linking, and then the ends of the DNA fragments are sequenced, producing 'paired-ends' or 'mate-pairs'. The paired-end chromatin-interaction sequence data is then co-assembled, with the draft de novo assembly contigs provided by the customer, using proprietary software. The genome-wide chromatin interaction datasets generated by this protocol provide long-range information that permits genomic sequences to be assigned, ordered and oriented into chromosome-length assemblies.

Producing the cross-linked DNA from chromatin is still somewhat of an art form, however several research groups have reported impressive genome assembly results using chromatin-interaction DNA sequence data (eg. Xie et al, 2015 with Arabidopsis). An efficient, cost-effective and convenient alternative is provided by commercial services that have optimized protocols for making and sequencing chromatin-interaction DNA libraries, and have well-tested proprietary computer programs to assemble the data. Two companies – Phase Genomics and Dovetail Genomics - provide comprehensive chromatin proximity guided genome assembly services. I have had in-depth discussions with both companies, although I only have direct experience with Dovetail Genomics, Inc. Both companies produce genome-wide chromatin interactions DNA sequence data using a method known as "Hi-C." Dovetail Genomics also produces chromatin interaction DNA sequence data from a second method called "Chicago". The difference between "Chicago" and "Hi-C" data is whether the cross-fixing of the chromatin is conducted in nuclei *in vivo* or with purified DNA *in vitro*, respectively. Both companies have their own proprietary bioinformatics programs for chromatin proximity guided assembly.

I propose to test the chromatin proximity guided approach to assemble a genome for the Mahogany genotype, which has been so important in the TACF breeding program. We have previously produced 11.2 billion bases of genomic DNA sequence for Mahogany, which will serve as a starting point for the project. Our research approach will proceed stepwise, as follows:

1. First the Carlson lab will isolate several micrograms of high molecular weight DNA from Mahogany leaves. The DNA will be tested for purity and size using an Agilent capillary bioanalyzer. After QC, both the DNA and flash-frozen leaf tissues of Mahogany will be shipped by overnight courier to the Dovetail Genomics company.
2. Next, the Carlson lab will conduct a *de novo* assembly of the using existing 11 Gb of Illumina genomic sequence, using the short-read assembly program in the CLC Genomics Workbench bioinformatics software suit installed on a 32-core server in The Schatz Center. The assembled contigs (continuous genome sequences), will be filtered to remove any short or low quality sequences, and then delivered to Dovetail Genomics company electronically.

3. The Dovetail Genomics Company will produce both "Chicago" DNA libraries using the HMW DNA that we will provide, and "Hi-C" chromatin-interaction DNA libraries. They will sequence the two libraries to a depth of over 100-fold genome coverage.
4. The Dovetail Genomics company will use their proprietary "HiRise" bioinformatics software pipeline to assemble the *de novo* contigs together with the "Chicago" DNA and "Hi-C" DNA sequence into large scaffolds arranged into chromosome-scale sequences assemblies. The Dovetail Genomics company will provide detailed reports about the quality and size of the assembled scaffolds. It will also provide statistics and figures to describe how well the scaffolds were arranged into chromosome representations. The Dovetail Genomics will identify the genes and their locations within the chromosomes, and conduct a 'BUSCO' analysis to identify the number of full length single-copy genes within the assembly, relative to the expected number from model plant genomes.
5. The Carlson lab will then conduct validation studies of the HiRise genome assembly by comparisons to DNA marker order in high-density genetic linkage maps from pedigrees that contained Mahogany as parent in the past. The Mahogany genome assembly will be compared in detail to the Vanuxem reference genome to determine if and where any differences occur.
6. After QC and validation, the Mahogany chromosome sequences will be delivered to TACF and TACF-collaborators, along with locations of genes, and added to the chestnut genome browser by Dr. Margaret Staton at the University of Tennessee.

Preliminary results

We have not yet had the opportunity to test the chromatin proximity guided approach with chestnut. However, the Carlson lab has just received encouraging results from a trial that we conducted with Dr. Richard Buggs at Queen Mary University in London and the Dovetail Genomics company in Santa Cruz, CA. We tested the approach with one genotype of green ash (*Fraxinus pennsylvanica*), another highly heterozygous, and threatened, hardwood tree species with an estimated haploid genome size of 920-990Mb (Million bases), which is about 15% larger than the *Castanea* genome. A *de novo* assembly was first conducted with 35X depth (genome size equivalents) of Illumina 150 bp sequences from 3kb & 10kb paired end libraries, yielding a total length of 902.5 Mb. The *de novo* contigs plus 26.5 X depth of Chicago *in vitro* chromatin-proximity DNA sequence data were produced and then co-assembled by the Dovetail Genomics company using their High Rise® method, which improved the assembly of contigs scaffolds by app. 50-fold over the de novo assembly of the short Illumina sequences. Finally, Hi-C *in vivo* chromatin-proximity DNA sequence reads at 3,310 X coverage were produced and assembled with the Chicago data and *de novo* contigs to obtain a total assembly length of 961.22 Mb, and representing a 1,432-fold improvement in genome sequence continuity (Table 1). This resulted in 23 chromosome-length sequences, which is the haploid chromosome number for green ash.

Table 1. Comparison of de novo and proximity-guided genome assembly results.

Assembly Statistic	Starting de novo assembly of Illumina sequence data	HiRise assembly of de novo contigs + Chicago data	HiRise assembly of de novo contigs + Chicago + Hi-C data
Total Length	902.53 Mb	961.00 Mb	961.22 Mb
Number of contigs	555,484	496,215	495,002
Scaffold L50	11,707 scaffolds	275 scaffolds	14 scaffolds
Scaffold N50	19 Kb	935 Kb	27.1 Mb
Longest Scaffold	524,151 bp	6,592,681 bp	50,995,270 bp

The Dovetail Genomics company then conducted a search for genes within the green ash genome assembly using the Augustus gene-prediction algorithm (Stanke et al, 2006). The genes predicted by the Augustus program were annotated (provided with putative gene names) based on sequence similarity with known genes in the model plant Arabidopsis. With the predicted gene set, Dovetail Genomics conducted a BUSCO analysis to assess the completeness of both the genome assembly and the gene discovery. The BUSCO (Benchmarking Universal Single-Copy Orthologue) process identifies the number of full length single-copy genes within the assembly, based on a curated set of genes that are known to be single-copy across a broad range of model plant species with well characterized genomes. Relative to the expected number of single-copy genes in model plant genomes (303), BUSCO analysis found 149 complete and single-copy genes, 74 complete and duplicated single-copy genes, 36 fragmented single-copy genes, and 44 missing single-copy genes. This result of detection of 259 single-copy genes (85% of expected) is a good result for a first round of gene prediction, in a new genome. Find the complete set of genes will require the additional use of RNA sequence data, which the Augustus program does not take into account.

A graphic view of the Dovetail Genomics HiRise genome assembly for green ash is provided in figure 1. The "Link Density Histogram" dot plot compares the positions of 'mate pair' sequences (the pair of end sequences from each and every sequenced DNA fragment obtained by chromatin cross-linking) versus the positions of each individual DNA sequence within the genome assembly. The alignment produces a diagonal of lines from lower left to upper right in the plot that represent each of the 23 green ash chromosomes. Dots (sequences) that fall outside the diagonal are probably repetitive DNAs that occur in multiple chromosomes, although this has yet to be confirmed.

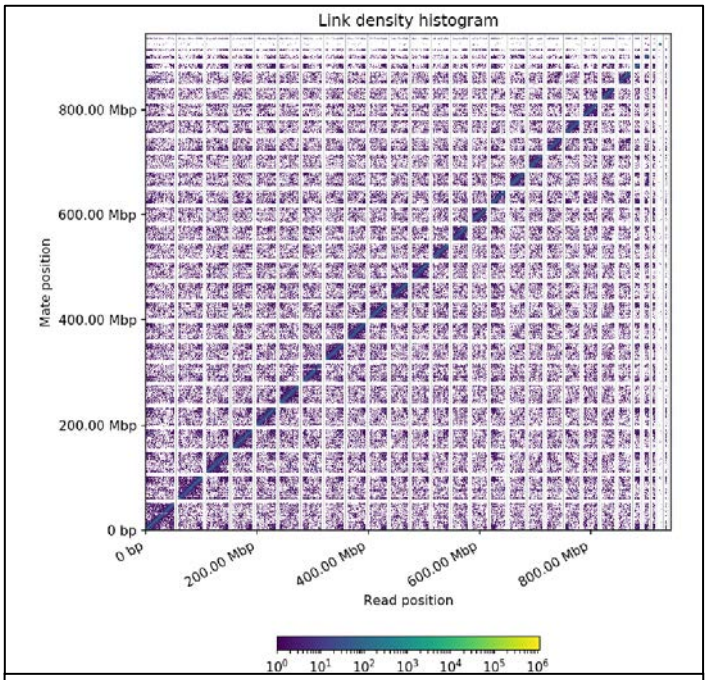

Figure 1. Dot plot alignment of the positions of 'mate pair' sequences against the positions of all individual DNA sequences within the green ash genome assembly.

The encouraging results from green ash provides confidence that we can obtain a high quality reference genome for the Mahogany genotype by contracting the Dovetail Genomics company to produce Chicago DNA and Hi- C DNA data and using their proprietary HiRise scaffolding pipeline to build a high quality genome starting with our *de novo* assembly of our existing set of Illumina reads from Mahogany. The assembly results will provide TACF and other chestnut genetics researchers with a valuable example of how well the chromatin-proximity guided assembly approach works with *Castanea*. We envision that a reference-quality genome that costs only app. $20,000 (including Illumina DNA sequencing and de novo assembly, but not post-assembly analyses) will make genomes for individual trees much more accessible, and will increase the confidence and accuracy of genomic selection within the TACF breeding program.

References

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. 31:1119–1125.

Claros, M.G., Bautista, R., Guerrero-Ferna´ ndez, D., Benzerki, H., Seoane, P., and Ferna´ndez-Pozo, N. (2012). Why assembling plant genome sequences is so challenging. Biology 1:439–459.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet. 14:390–403.

Felipe A. Simão, Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, Evgeny M. Zdobnov; (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31 (19): 3210-3212. doi: 10.1093/bioinformatics/btv351.

Jannink, J. L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in functional genomics*, *9*(2), 166-177.

Kaplan, N., and Dekker, J. (2013). High-throughput genome scaffolding from in vivo DNA interaction frequency. Nat. Biotechnol. 31:1143–1147.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006) AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Research*, *34* (Web Server issue), W435–W439. Doi: 10.1093/nar/gkl200.

Staton M, et al. The Chinese Chestnut Genome and QTL Sequences, V1.0, January 2014, http://www.hardwoodgenomics.org/chinese-chestnut-genome.

Xie, T., Zheng, J. F., Liu, S., Peng, C., Zhou, Y. M., Yang, Q. Y., & Zhang, H. Y. (2015). De novo plant genome assembly based on chromatin interactions: a case study of Arabidopsis thaliana. *Molecular plant*, *8*(3), 489-492.

H. PROJECT TIMELINE

Objective completion dates

Objective / milestone	Completion
1. Isolate high molecular weight DNA from Mahogany leaves, and deliver DNA and fresh leaf tissues to Dovetail Genomics company	**month 1**
2. Conduct *de novo* assembly using existing short-read genome sequence, and deliver results to Dovetail Genomics company	**month 2**
3. Dovetail Genomics company will produce “Chicago” and “Hi-C” chromatin-interaction DNA sequence data	**month 3**
4. Dovetail Genomics company will use the HiRise system to put together the *de novo* assembly contigs and the “Chicago” and “Hi-C” DNA sequence into chromosome-scale sequences assemblies	**month 6**
5. The HiRise genome assembly will be validated by comparison to high-density genetic linkage maps and the Vanuxem reference genome.	**month 9**
6. When validated, the HiRise genome assembly will be delivered to TACF and collaborators, along with locations of genes.	**month 12**

I. MEASUREMENT AND REPORTING OF RESULTS

Updates on progress in reaching milestones will be provided through frequent communications with Dr. Jared Westbrook whose feedback will be instrumental in guiding the project to an appropriate conclusion for optimal application to TACF breeding objectives.

J. BUDGET DETAILS

Funds will be used for molecular biology supplies, to purchase DNA sequencing services and for bioinformatics graduate student wages.

Personnel	
Salaries	
Fringe Benefits	$0
Subtotal Salaries, Wages, Fringe	**$0**
Total Personnel	**$0**
Other Direct Costs	
Materials & Supplies (DNA isolations and shipping)	$250
Travel	0
Publications	0
Other Direct Costs	
Fee for service paid to Dovetail Genomics company to produce Chicago DNA sequence and Hi-C DNA sequence, and to conduct HiRise assembly of the genome and identification of gene sequences and positions within the genome.	$8,000
Reimburse Dr. Meg Staton for creating a genome website	$750
Equipment	0
Total Direct Costs	**$10,000**

Indirect Costs (refer to budget notes above)	**0**
Total Costs	**$10,000**

K. BRIEF CV FOR PRINCIPAL INVESTIGATOR

John E. Carlson
The Pennsylvania State University
Department of Ecosystem Science and Management
323 Forest Resources Building
University Park, PA 16802

Office: (814) 863-7561
Lab: (814) 863-2513
Fax: (814) 865-3725
E mail: jec16@psu.edu

Education and Training:
B.Sc., 1974, Biochemistry, Honors, University of Pittsburgh, Pittsburgh, PA
M.Sc.**,** 1978, Agronomy & Plant Physiology, University of Illinois, Urbana – Champaign, IL
Ph.D., 1983, Genetics, University of Illinois, Urbana – Champaign, IL
Postdoctoral Research, 1982-1983, Plant Pathology, Kansas State University, Manhattan, KS

Professional Positions Held:

1983-1988	Research Scientist, Allelix, Inc., Mississauga, Ontario, Canada
1988-1994	Assistant Professor, University of British Columbia, Vancouver, B.C., Canada
1994-1997	Associate Professor, tenured, University of British Columbia, Vancouver, B.C., Canada
1995-1997	Chair**,** Genetics Graduate Program**,** University of British Columbia
1997-2007	Associate Professor, tenured, School of Forest Resources and The Life Sciences Consortium, Pennsylvania State University
2000-present	Director, The Louis W. Schatz Center for Tree Molecular Genetics at Penn State
2007-present	Full Professor, tenured, Department of Ecosystem Science and Management, Department of Plant Science, Huck Institutes fort the Life Sciences, Pennsylvania State University
2010-2013	Visiting Professor, Chonnam National University, South Korea

Awards
Alex and Jessie C. Black Faculty Award for Excellence in Research, College of Agricultural Sciences, Pennsylvania State University, 2012
Association of Public and Land-Grant Universities Excellence in Multistate Research Award for "Biological Improvement of Chestnut through Technologies that Address Management of the Species, its Pathogens, and Pests" project, 2010.

Selected Refereed Publications (total 136):

1. Konar A, Choudury O, Bullis R, Fiedler L, Kruser J, Stephens M, Gailing O, Schlarbaum S, Coggeshall MV, Staton ME, **Carlson JE**, Emrich S, Romero-Severson J. 2017. High-quality genetic mapping with ddRADseq in the non-model tree *Quercus rubra*. BMC Genomics, 18(1):417. BMC Genomics, 18(1):417, 12 pages.
2. Gailing O, Staton ME, Lane T, Schlarbaum SE, Nipper R, Owusu SA, **Carlson JE.** 2016. Construction of a Framework Genetic Linkage Map in *Gleditsia triacanthos* L. Plant Molecular Biology Reporter, pp.1-11.
3. Lane T, Best T, Zembower N, Davitt J, Henry N, Xu Y, Koch J, Liang H, McGraw J, Schuster S, Shim D, Coggeshall M, **Carlson JE,** Staton ME. 2016. The green ash transcriptome and identification of genes responding to abiotic and biotic stresses. BMC Genomics 17:702, 16 pages.
4. Staton ME, Best TO, Khodwekar SD, Owusu SA, Xu T, Yu Y, Jennings TN, Knaus BJ, Cronn RC, Arumuganathan AK, Coggeshall MV, Gailing O, Liang H, Romero-Severson J, Schlarbaum SE, **Carlson JE.** 2015. Preliminary genomic characterization of ten hardwood tree species from multiplexed low coverage whole genome sequencing. PLoS ONE 10(12): e0145031, 13 pages.
5. Staton M, Zhebentyayeva T, Olukolu B, Fang GC, Nelson D, **Carlson JE,** Abbott AG. 2015. Substantial genome synteny preservation among woody angiosperm species: comparative genomics of Chinese chestnut (*Castanea mollissima*) and plant reference genomes. BMC

Genomics. 16:744, 14 pages.
6. Khodwekar S, Staton M, Coggeshall MV, **Carlson JE**, Gailing O. 2015. Nuclear microsatellite markers for population genetic studies in sugar maple (*Acer saccharum* Marsh.). Ann. For. Res. 58(2) DOI:10.15287/afr.2015.360, 12 pp.
7. Zhang X, Carlson A, Tian Z, Staton M, Schlarbaum SE, **Carlson JE**, Liang H. 2015. Genetic characterization of *Liriodendron* seed orchards with EST-SSR markers. Journal of Plant Science and Molecular Breeding. 4:1. (http://dx.doi.org/10.7243/2050-2389-4-1).
8. Staton M, et al. The Chinese Chestnut Genome and QTL Sequences, V1.0, January 2014, http://www.hardwoodgenomics.org/chinese-chestnut-genome.
9. Nelson CD, Powell WA, Maynard CA, Baier KM, Newhouse A, Merkle SA, Nairn CJ, Kong L, **Carlson JE**, Addo-Quaye C, Staton ME, Hebard FV, Georgi LL, Abbott AG, Olukolu BA, Zhebentyayeva T. 2014. The forest health initiative, American chestnut (*Castanea dentata*) as a model for forest tree restoration: biological research program. Acta Hort. (ISHS) 1019:179-189.
10. Sha T, Liang H, Yan D, Zhao Y, Han X, **Carlson JE**, Xia X, Yin W. 2013. *Populus euphratica*: the transcriptomic response to drought stress. Plant Mol Biol 83(6): 539-557.
11. Owusu SA, Staton M, Jennings TN, Schlarbaum S, Coggeshall MV, Romero-Severson J, **Carlson JE**, Gailing O. 2013. Development of Genomic Microsatellites in *Gleditsia triacanthos* (Fabaceae) using Illumina Sequencing. Applications Plant Sciences 1(12): 4 pp.
12. Fang GC, Blackmon BP, Staton ME, Nelson CD, Kubisiak TL, Olukolu BA, Henry D, Zhebentyayeva Y, Saski CA, Cheng CH, Monsanto M, Ficklin S, Atkins M, Georgi LL, Barakat A, Wheeler N, **Carlson JE**, Sederoff R, Abbott AG. 2013. A physical map of the Chinese chestnut (*Castanea mollissima*) genome and its integration with the genetic map. Tree Genetics & Genomes, 9(2): 525-537.
13. Petit RJ, **Carlson J**, Curtu AL, Loustau ML, Plomion C, González-Rodríguez A, Sork V, Ducousso A. 2013. Fagaceae trees as models to integrate ecology, evolution and genomics. New Phytologist, 197: 369–371.
14. Kremer K, Abbott AG, **Carlson JE**, Manos PS, Plomion C, Sisco P, Staton ME, Ueno S, Vendramin GG. 2012. Genomics of Fagaceae (Review). Tree Genetics & Genomes, 8(3): 583-610.
15. Barakat A, Staton M, Cheng C-H, Park J, Buang N, Yassin M, Ficklin S, Yeh C-C, Hebard F, Baier K, Powell W, Schuster S, Wheeler N, Abbott A, **Carlson J**, Sederoff R. 2012. Chestnut resistance to the blight disease: insights from transcriptome analysis. BMC Plant Biology, 12:38, 14 pp.
16. Argout et al. 2011. The genome of *Theobroma cacao* L. Nature Genetics 43: 101–109.
17. Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell W, Wheeler N, Sederoff R, Carlson. JE 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. 11 pages. (doi:10.1186/1471-2229-9-51).
18. Tuskan GA, et al. 2006. The Genome of Western Black Cottonwood, *Populus trichocarpa* (Torr. & Gray ex Brayshaw). Science 313: 1596-1604.
19. Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis D, Doyle JJ, Soltis P, **Carlson JE,** Arumuganathan K, Barakat A, Albert V, Ma H, dePamphilis CW. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Research 16: 738 – 749.

Current Grants and Contracts:

"Identification of genes and alleles for blight resistance in Castanea spp,", USDA NIFA grant, $150,000 total, Subaward from American Chestnut Foundation, $41,144, 12-15-15 to 12-14-17.

"The Northeast Woody/Warm-season Biomass Consortium (NEWBio)", USDA Agriculture and Food Research Initiative Program, $9.9 Million total award, PI: T Richard (PD), 09-01-12 to 08-31-17.

H. CONFLICTS OF INTEREST

There are neither commercial nor IP conflicts of interest for the applicant with TACF or the Dovetail Genomics company,

Collaborators, Co-authors, Co-editors, & Other Affiliations presenting possible conflicts of interest:
Collaborators and Co-editors (last 48 months): Abbott A (U Kentucky); Ahn S-J (Chonnam National University); Barakat A (South Dakota); Barry KW (ORNL DOE); Brown N (Penn State); Brunner AM (Virginia Tech U); Chen C-C (Clemson University); Chen E (Penn State): Coggeshall M (U Missouri); dePamphilis C (Penn State); Fang EG (Clemson U); Gailing O (Michigan Tech University); Geib SM (USDA ARS); Han KY (Michigan State U); Hebard F (The American Chestnut Foundation); Hoover K (Penn State); Jacobson M (Penn State); Joshi C (Michigan Tech University); Kemanian A (Penn State); Kemble RJ (Syngenta Retired); Kim J (Chonnam National University); Koch J (USDA Forest Service); Kubisiak TL (USDA Forest Service); Lanaud C. (INRA France); Liang H (Clemson); Luan S (UC-Berkeley); Luthe D (Penn State); Miller W (Penn State); Neale D (UC-Davis); Nelson CD (USDA Forest Service); Pechanova O (Canadian Forest Service); Powell W (SUNY-ESF Syaracuse); Richard T (Penn State); Romero-Severson J (Notre Dame); Saski C (Clemson); Schlarbaum S (U Tennessee); Schuster S (Penn State); Shumaker K (University of West Alabama); Sisco P (American Chestnut Foundation); Smart L (Cornell); Staton M (University of Tennessee); Tien M (Penn State); Tomsho LP (Penn State); Welsh Scott (LignoLink Inc.); Tringe SG (ORNL, DOE); Wheeler N (Oregon State U); Xia X (Beijing Forestry University); Yin W (Beijing Forestry University); Yuceer C. (Mississippi State retired).
Co-authors only (last 48 months): Argout X (INRA France); Arumuganathan K (Virginia Mason Hospital); Axtell MJ (Penn State); Bagniewska-Zadworna A (Poznan University); Baier K. (SUNY-ESF Syaracuse); Chovitia M, (Lawrence Berkeley National Laboratory); Davis JM (U Florida). del Mar Jimenez-Gasco M (Penn State); Glavina del Rio T. (Lawrence Berkeley National Laboratory); Guiltinan M. (Penn State); Meilan R. (Purdue U); Sederoff R. (North Carolina State U); Strauss SH (Oregon State).
Graduate and Postdoctoral Advisors and Advisees. (total 54)
Graduate and Postdoctoral Advisors - Widholm J.M. (University of Illinois); Laughnan J.R. (deceased); Kemble RJ (Retired).
Current PhD Student Advisees (Penn State University)**:** Wang W; Weathers C: Wu D; Cannon N.