Michigan Technological University

Innovation and Industry Engagement

Advanced Technology Development Complex
1400 Townsend Drive
Houghton, Michigan 49931-1295
906-487-2228 • Fax 906-487-1979

Via Email: sff3@psu.edu

July 30, 2014

Sara Fitzsimmons
The American Chestnut Foundation

RE: Michigan Tech Proposal #1406017, "Identification of different ecotypes and centers of adaptive genetic diversity in American chestnut," submitted by Dr. Oliver Gailing

Dear Ms. Fitzsimmons:

On behalf of Dr. Gailing, Michigan Technological University is pleased to submit the attached proposal for your review and consideration.

For any questions or concerns of a legal or contractual nature please contact me at (906) 487-2228 or mkhaapap@mtu.edu. For technical questions please contact Dr. Gailing directly at either (906) 487-1615 or ogailing@mtu.edu.

Sincerely,

Marilyn Haapapuro
Associate Director for Industrial Contracting

Michigan Technological University

School of Forest Resources and Environmental Science
Noblet, Hesterberg, and Horner Halls
1400 Townsend Drive
Houghton, Michigan 49931-1295
906/487-1615
Fax 906/487-2915
ogailing@mtu.edu
Houghton, June 5, 2014

The American Chestnut Foundation - Call for Research Proposal 2014

To Whom It May Concern

Please find attached the proposal “Identification of different ecotypes and centers of adaptive genetic diversity in American chestnut” by Oliver Gailing, Brian McCarthy and Dana Nelson.

The proposal is a resubmission of a proposal with the same title that we submitted in 2013. As suggested by the reviewers we now have a stronger focus on candidate genes with putative role in disease-resistance (i.e. candidate genes that underlie QTL regions for chestnut blight and ink disease). Instead of $10,000 we request $ 6,000 for the genetic monitoring of EST-SSRs with putative role in biotic and abiotic stress resistance. In a preliminary analysis we have already identified a set of polymorphic candidate genes for American Chestnut, thus reducing the costs for the project. We do not request funds for labor costs and for part of the lab work. Our results will provide information on the genetic diversity in adaptive gene loci to be published in a peer-reviewed journal.

We hope that our proposal finds your interest.

Sincerely yours,

Oliver Gailing

(Associate Professor, Ecological Genetics)

a. Title: Identification of different ecotypes and centers of adaptive genetic diversity in American chestnut

b. Summary:

We will develop a set of new gene-based DNA markers with potential role in local adaptation to (1) identify centers of genetic diversity in American chestnut, (2) detect gene-based markers under divergent selection in contrasting environments and to (3) associate allele frequency distributions at gene markers with environmental variables. We will map the genes back to the *Castanea mollissima* linkage map (*in silico* mapping) to analyze the distribution of these genes on the linkage map and their potential association with QTL for phenotypic and resistance traits.

c. Principal investigators:

Oliver Gailing, Associate Professor for Ecological Genetics, Michigan Technological University, 1400 Townsend Drive, Houghton 49931, Michigan, ogailing@ mtu.edu.

Brian C. McCarthy, Professor of Forest Ecology & Chair, Dept. Env. & Plant Biology, 416 Porter Hall, Ohio University, Athens, OH 45701-2979 USA, mccarthy@ohio.edu

C. Dana Nelson, Project Leader/Research Geneticist, US Forest Service, Southern Institute Forest Genetics, Southern Research Station, Saucier, MS 39574 USA, dananelson@fs.fed.us, phone: 228-832-2747-201

d. Duration of project: 2 years

e. Total amount requested: We requested a total of $ 6,000. The estimated costs for the marker analyses are $ 9,500 excluding labor costs. The additional costs will be covered by Oliver Gailing's incentive account.

f. Short and long-term goals of the project: Short-term goals of the project are the identification of (1) center of genetic diversity and (2) of genes under divergent selection in contrasting environments.

The long-term goals are to (1) provide information for the selection of breeding material and reintroduction of hybrid American chestnut and (2) to use a combined outlier screening and QTL mapping approach to identify genes related to biotic and abiotic stress resistance for use in marker-assisted breeding.

g. Narrative

Background

Considerable progress has been made to develop blight resistant chestnuts for restoration purposes using mainly genetic marker-assisted back-cross breeding to incorporate blight resistance from *C. mollissima* into a *C. dentata* genetic background. Restoration programs are more likely to be successful if hybrid American chestnut populations are genetically diverse (James et al., TACF project 2010 - 2013) and locally adapted. For this purpose genetic variation of the American chestnut parents should be captured from many individuals originating from different geographic regions and climatic zones. The conservation of genetic variation in fitness-related traits (adaptive genetic variation) is crucial for the successful restoration of American chestnut since the species' reintroduction is threatened by other biotic (e.g. *Phytophtora cinnamomi*) and abiotic stressors. While growing genomic resources and gene-based markers are becoming available for American chestnut and related species (Barakat et al., 2012; Bodénès et al., 2012; Kubisiak et al., 2013; Nishio et al., 2011), genetic variation at these markers with annotated function (e.g. gene-based microsatellites) has not yet been analyzed in natural populations of American chestnut. In a preliminary study we have already identified and characterized a set of gene-based markers in American chestnut (Table 1).

We hypothesize (1) significant differences in the level of genetic variation for populations from different geographic regions and (2) significant differentiation among regions at some gene-based markers that reflect different local adaptations of the populations (ecotypes) across their distribution range.

These markers that are identified as under divergent selection between populations (outlier loci) from contrasting environments and/or associated with environmental variables across populations will be mapped back to the *Castanea mollissima* linkage map (Kubisiak et al. 2013) to test for a possible co-location with QTL regions. The expected results will be important to identify centers of genetic diversity and to select appropriate breeding material to produce locally adapted material for the reintroduction of American chestnut.

In the future, a combined outlier screening and QTL mapping approach based on nextgen sequencing markers will allow us to test for a co-location of genome-wide outliers with genomic regions that underlie QTL for traits related to biotic and abiotic stress resistance.

Work plan

We propose to characterize genetic variation within and among *C. dentata* populations covering the distribution range of the species using 16 gene-based microsatellite markers with annotated function (Expressed Sequence Tag- Simple Sequence Repeats, EST-SSRs). In this preliminary study we will focus on 10 populations that represent the five US climatic zones within the species' native range. Leaf material

has been sampled from about 30 trees per population recording the GPS position of each tree (Kubisiak and Roberds, 2005). In order to select the 16 gene-based microsatellite markers for the range-wide study, we will screen about 30 markers with annotated function in abiotic and biotic stress tolerance and known location on genetic linkage maps (Kubisiak et al., 2013) in two populations from different climatic zones. We will include orthologous candidate genes that had been identified as being under divergent selection in interfertile red oak species with different adaptations to drought (Lind and Gailing, 2013; Sullivan et al., 2013; Lind-Riehl et al. 2014). A total of 12 of these markers have already been tested in a panel of 8 American and Chinese chestnut individuals and 7 of them amplified polymorphic products in the expected size range (Table 1, Figure 1, unpublished results). Two of the loci, FIR013 and FIR039, were identified as putative outlier loci under strong divergent selection in the North American red oak species and have putative functions in drought stress response and the control of flowering time (Sullivan et al. 2013; Lind-Riehl et al. 2014). EST-SSR markers that display signatures of selection relative to potentially neutral genomic SSRs that had been analyzed in the same sample set (Kubisiak and Roberds, 2005) will be selected for the population analyses. Calculation of genetic variation within and among populations will follow standard procedures. Associations of genetic variation patterns with geographic distance and environmental parameters will also be tested. A detailed description of the methods is given below.

Methods

Marker selection

We have already identified 7 gene-based markers for American chestnut that were originally developed for oak (Durand et al. 2010, Lind-Riehl et al. 2014). Additional EST-SSRs that are genetically mapped in *Castanea mollissima* (Kubisiak et al., 2013) will be selected based on their annotated function in vegetative bud burst and biotic and abiotic stress tolerance, and genomic co-location with QTL for chestnut blight resistance. Our aim is to obtain at least 30 markers for the outlier analysis. Selection of the markers will be based on their reproducibility, absence of null alleles and potential importance in adaptation and disease resistance. Functional annotation will be assigned to microsatellite-containing ESTs using the BLASTx algorithm (Altschul et al., 1997) by comparison to homologous sequences in the NCBI database. Based on the initial outlier screen of the 30 polymorphic markers in two populations from different climate zones (outlier analysis), 16 EST-SSRs will be selected for the population analyses.

Marker analyses

DNA and leaf samples will be provided by Dana Nelson and Tom Kubisiak who performed a range-wide sampling for the assessment of neutral genetic variation patterns at genomic SSR and RAPD markers (Kubisiak and Roberds, 2005). We will be able to combine data sets and compare patterns of genetic variation at candidate gene associated EST-SSR markers with neutral variation (Kubisiak and Roberds,

2005). Samples will be amplified with fluorescent labeled forward primers using a Peltier Thermal Cycler (Geneamp® PCR system 2700, Applied Biosystems) and PCR products will be separated on an ABI PRISM® 3730 Genetic Analyzer (Applied Biosystems). PCR reactions will be adapted from Kubisiak et al. (2013). Even though species identity was tested with a chloroplast marker that differentiates between American chestnut and the native cogener species chinkapin (*Castanea pumila*) (Kubisiak and Roberds, 2005), the occurrence of interspecific hybrids cannot be excluded. We therefore use the generated marker information to assign individual samples to species and interspecific hybrids in the program STRUCTURE 2.3.4 (Pritchard et al., 2000). For this purpose we will include 20 *C. pumila* reference samples that were identified based on morphology and chloroplast marker information.

Outlier screens

A total of 30 EST-SSRs will be selected for the outlier screen and amplified in two populations from different adjacent climate zones to identify gene loci (outlier loci) that show a higher or lower differentiation between populations than expected under selective neutrality. We will use the program LOSITAN that implements the F_{ST}-based algorithms of FDIST to identify outliers that deviate significantly from a simulated neutral confidence envelope (Antao et al., 2008; Beaumont and Nichols, 1996). Loci with higher differentiation between populations than expected under neutrality are identified as potential outliers under divergent selection. Those falling below the lower bound of the neutral envelope might be under balancing selection. Since the confidence interval converges at extreme values for expected heterozygosity (H_e), candidate genes under balancing selection were not consistently identified in different simulations (Sullivan et al., 2013) while loci under divergent selection were highly reproducible (Sullivan et al., 2013). We will therefore run the simulations at least three times for each pairwise comparison. To identify signatures of selective sweeps we will also run the LnRH test statistic that estimates variability between populations at individual loci instead of population divergence to identify selection (Schlötterer, 2002).

Identification of genetic diversity centers

Genetic variation within and among populations and climatic regions will be calculated for all markers and separately for potentially adaptive (outlier markers) and neutral markers. Specifically the following genetic variation parameters will be calculated: number of alleles per locus (N_a), observed heterozygosity (H_o) and Nei's unbiased gene diversity (H_e) (Nei, 1973). Pairwise genetic differentiation between populations and corresponding significances will be calculated in GenePop4.1 (Raymond and Rousset, 1995). To visualize genetic distances among populations an unweighted pair-group method with arithmetic means (UPGMA) dendrogram (Sneath and Sokal, 1973) will be calculated in Populations 2.0 (Langella, 1999) using 1,000 bootstrap replicates. An Analysis of Molecular Variance (AMOVA,

(Excoffier et al., 1992) will be performed in Arlequin 3.5 (Excoffier and Lischer, 2010) in order to assess genetic variation within and among populations and climatic regions. To test for associations between geographic and genetic distances we will perform a Mantel test as implemented in GeneAlEx v.6.41 (Peakall and Smouse, 2006).

Association of allele frequencies with environmental variables

We will search for associations between geographic (longitude, latitude) and environmental variables (e.g. minimum and maximum temperature, precipitation) on the one hand and allele frequencies and genetic variation parameters on the other hand using stepwise regression analysis (Kubisiak and Roberds, 2005) and association mapping approaches. Associations between SSR alleles and environmental variables will be performed using a standard association mapping approach by substituting the phenotype with environmental variables according to Eckert et al. (2010b). Specifically, following Du et al. (2013) associations between SSR marker locus and climatic variables will be analyzed using a mixed linear model accounting for population structure and kinship (Yu et al., 2006) and P-values will be generated for each test using 10,000 permutations of genotypes with respect to environmental values. All analyses will be conducted using TASSEL v. 3.0 (Bradbury et al., 2007). Multiple testing corrections will be performed using the false discovery rate (FDR) method (Storey, 2002; Storey and Tibshirani, 2003). Data on six nuclear microsatellite loci and 19 RAPD loci available for the same populations (Kubisiak and Roberds, 2005) will be used to estimate pair-wise kinship coefficients in the software SPAGEDi (Hardy and Vekemans, 2002) and population structure coefficients in STRUCTURE 2.3.4 (Pritchard et al., 2000).

Climate data (annual minimum and maximum temperatures, monthly temperature, precipitation) will be obtained from the PRISM website (Oregon State University, http://www.prismclimate.org) and from an ESRI 30 arc-second grid file available at the WorldClim website (http://www.worldclim.org/current) according to Eckert et al. (2010a). Accumulated growing degree-days above 5 °C will be calculated and an aridity index will be calculated according to (Eckert et al., 2010a; Eckert et al., 2010b; Sork et al., 2010).

The location of all markers relative to QTL for phenotypic and disease-resistance traits will be assessed on the *C. mollissima* linkage map to analyze the genomic location of genes with signatures of selection.

h. Timeline:

	Year 1				Year 2			
Activity	**1-3**	**4-6**	**7-9**	**10-12**	**1-3**	**4-6**	**7-9**	**10-12**
Outlier screening	x	x						
Range-wide marker analyses			x	x	x	x		
Data analysis and publication		x	x	x	x	x	x	x

j. Budget: Total costs are estimated as $9,500. A total of $ 6,000 is requested from the American Chestnut Foundation. Based on our experience with these analyses we estimate $ 2,000 for the marker development and $ 7,500 for the population genetic analyses ($25 per sample x 300 samples, including DNA isolation, PCR, labeled primers and genotyping services). These estimates do not include labor costs. ***Requested funds:*** Year 1: $ 4,000. Year 2: $2,000.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research, **25**(17), 3389-3402.

Antao, T., Lopes, A., Lopes, R.J., Beja-Pereira, A., Luikart, G. (2008) LOSITAN: A workbench to detect molecular adaptation based on a F(st)-outlier method. BMC Bioinformatics, **9**.

Barakat, A., Staton, M., Cheng, C.H., Park, J., Yassin, N.B.M., Ficklin, S., Yeh, C.C., Hebard, F., Baier, K., Powell, W., Schuster, S.C., Wheeler, N., Abbott, A., Carlson, J.E., Sederoff, R. (2012) Chestnut resistance to the blight disease: insights from transcriptome analysis. BMC Plant Biology, **12**.

Beaumont, M.A. and Nichols, R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London Series B-Biological Sciences, **263**, 1619-1626.

Bodénès, C., Chancerel, E., Gailing, O., Vendramin, G.G., Bagnoli, F., Durand, J., Goicoechea, P.G., Soliani, C., Villani, F., Mattioni, C., Koelewijn, H.P., Murat, F., Salse, J., Roussel, G., Boury, C., Alberto, F., Kremer, A., Plomion, C. (2012) Comparative mapping in the Fagaceae and beyond with EST-SSRs. BMC Plant Biology, **12**, 153.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, **23**, 2633-2635.

Du, Q.Z., Pan, W., Xu, B.H., Li, B.L., Zhang, D.Q. (2013) Polymorphic simple sequence repeat (SSR) loci within cellulose synthase (PtoCesA) genes are associated with growth and wood properties in *Populus tomentosa*. New Phytologist, **197**(3), 763-776.

Eckert, A.J., Bower, A.D., Gonzalez-Martinez, S.C., Wegrzyn, J.L., Coop, G., Neale, D.B. (2010a) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). Molecular Ecology, **19**, 3789-3805.

Eckert, A.J., van Heerwaarden, J., Wegrzyn, J.L., Nelson, C.D., Ross-Ibarra, J., Gonzalez-Martinez, S.C., Neale, D.B. (2010b) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). Genetics, **185**, 969-982.

Excoffier, L. and Lischer, H.E.L. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, **10**, 564-567.

Excoffier, L., Smouse, P.E., Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, **131**, 479-491.

Hardy, O.J. and Vekemans, X. (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes, **2**, 618-620.

Kubisiak, T. and Roberds, J.H. (2005) Genetic structure of American Chestnut populations based on neutral DNA markers. In Restoration of American Chestnut to forest lands (K.C. Steiner & J.E. Carlson, eds): pp 1-13.

Kubisiak, T.L., Nelson, C.D., Staton, M.E., Zhebentyayeva, T., Smith, C., Olukolu, B.A., Fang, G.C., Hebard, F.V., Anagnostakis, S., Wheeler, N., Sisco, P.H., Abbott, A.G., Sederoff, R.R. (2013) A transcriptome-based genetic map of Chinese chestnut (*Castanea mollissima*) and identification of regions of segmental homology with peach (*Prunus persica*). Tree Genetics & Genomes, **9**, 557-571.

Langella, O. (1999) Populations, 1.2.30, CNRS UPR9034.

Lind, J. and Gailing, O. (2013) Genetic structure of *Quercus rubra* L. and *Q. ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. Tree Genetics & Genomes, **31**, 231-239.

Lind-Riehl J., Sullivan A., Gailing O. (2014) Evidence for selection on a CONSTANS-like gene between two red oak species. Annals of Botany, **113**, 967-975.

Nei, M. (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences USA, **70**, 3321-3323.

Nishio, S., Yamamoto, T., Terakami, S., Sawamura, Y., Takada, N., Nishitani, C., Saito, T. (2011) Novel genomic and EST-derived SSR markers in Japanese chestnuts. Scientia Horticulturae, **130**, 838-846.

Peakall, R. and Smouse, P.E. (2006) GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, **6**, 288-295.

Pritchard, J.K., Stephens, M., Donnelly, P. (2000) Inference of population structure using multilocus genotype data. Genetics, **155**, 945-959.

Raymond, M. and Rousset, F. (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity, **86**, 248-249.

Schlötterer, C. (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics, **160**, 753-763.

Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy* W.H. Freeman, San Francisco.

Sork, V.L., Davis, F.W., Westfall, R., Flint, A., Ikegami, M., Wang, H.F., Grivet, D. (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Nee) in the face of climate change. Molecular Ecology, **19**, 3806-3823.

Storey, J.D. (2002) A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B-Statistical Methodology, **64**, 479-498.

Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America, **100**, 9440-9445.

Sullivan, A., Lind, J., McCleary, T.S., Romero-Severson, J., Gailing, O. (2013) Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. Plant Molecular Biology Reporter, **31**, 231-239.

Yu, J.M., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, **38**, 203-208.

Table 1. Polymorphic EST-SSR in *Castanea dentata*

Locus	Forward primer (5'-3')	Reverse Primer (5'-3')	Linkage group*	BLASTX sequence description
FIR013	CGGGGAGGTTGATGAGTATT	AACACTGTCACCCCCATAGC	2	CONSTANS-like-1
FIR030	GGACATATTTATCTAGGAGACGAGGT	ATGTCCCATAGCACAGAGCA	7	NADH dehydrogenase
FIR039	GAGCCTCTTTCATCGCTCAC	TCAACACCCCAAAACTCCAT	1	Histon-Deacetylase 2c
FIR048	TGCACCAAAATTGGAGGATG	TTGATGCAAGGTGCAGTTTC	2	FtsH protease
FIR053	AGTTTCCCCACATTTGTTGC	TACCATGCACCAAGCAATTC	5	Glutaredoxin c9
VIT057	TCAGCAAAATCCCAACTTTGT	ACACTTCGCTGTTCCTCGAT	9	Ethylene-responsive transcription factor-like
POR003	CTCGCTCTCCTCCCCTAATC	AGCTTTGATCGAGTCCGAAA	3	uncharacterized protein

*: the location of markers on the *Quercus robur* linkage map is shown (Durand et al. 2010).

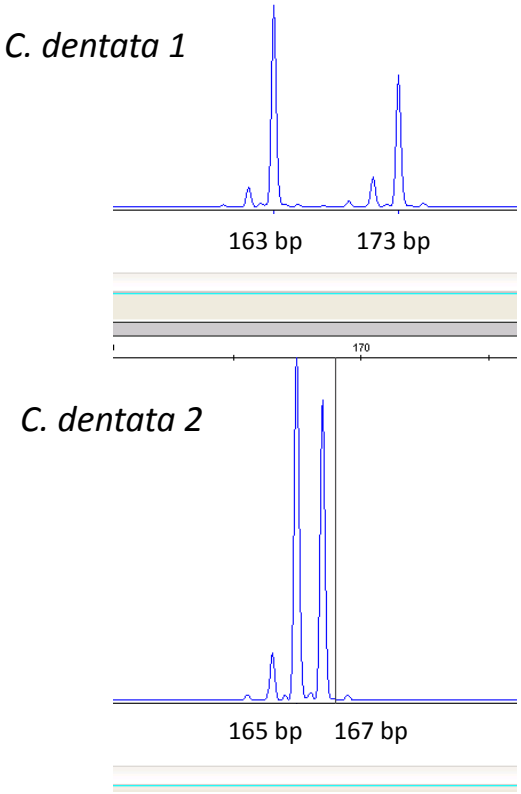

Figure 1: Amplification products of FIR030 in two *Castanea dentata* samples. Size is given in base pairs (bp).