

Genomics of forest and ecosystem health in the Fagaceae: meeting report

Antoine Kremer · Ronald Sederoff · Nicholas Wheeler

Received: 23 December 2009 / Accepted: 26 January 2010 / Published online: 25 February 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract A summary of 35 keynote, invited and volunteer papers delivered at a recent international conference is provided along with web links to PDFs of those presentations. Major conference themes targeted Genomic Tool Development for the Fagaceae and Application of Genomic Resources. The meeting provided a venue for reviewing the rapidly expanding knowledge base on Fagaceae genomics and for developing collaborations between scientists from Europe and North America.

Keywords Fagaceae · Genomics

Introduction

An International Conference entitled “Genomics of Forest and Ecosystem Health in the Fagaceae” was recently held at the Research Triangle Park, North Carolina (November 10th to 13th, 2009). More than 70 participants from nine countries and 19 US states attended the conference. The Fagaceae family comprises about 900 species in nine genera (Fig. 1) and represents many of the major hardwoods of the tem-

perate forests of the Northern Hemisphere. In Europe and North America, the major genera of interest are the oaks (*Quercus* spp.), chestnuts (*Castanea* spp.), and beeches (*Fagus* spp.), and in Asia, two additional genera, *Lithocarpus* and *Castanopsis*, are taxa being actively investigated. The meeting was stimulated by two main projects on Fagaceae genomics: in Europe, the EVOLTREE (<http://www.evoltree.org/>) project, which focuses on the oaks and in the USA, the National Science Foundation (NSF)-sponsored project “Genomic Tool Development for the Fagaceae” (www.fagaceae.org/), centered around the chestnuts.

The purpose of the conference was to bring together diverse groups working on related Fagaceae species with an emphasis on genomics research. One objective of the meeting was to present results on new genomic tools developed for species in the Fagaceae, and to explore ways in which these new methods and resources may be used to address existing problems. Focus areas were molecular breeding, particularly for disease resistance, the adaptation of natural populations, and the evolutionary history of the genera and the family. A second objective was to foster new projects including international collaboration. The main subject divisions of the meeting were the development of genomic tools and their applications.

The purpose of this report is to provide a summary of the 35 keynote, invited and volunteer papers given at the meeting for those who did not attend, but have interest in advances in this area of research. The meeting agenda, speaker abstracts, and most presentations (pdf format) are available at two project websites (<http://www.fagaceae.org/node/333339> ; <http://www.evoltree.org>). References in this report are to the authors of presented papers and posters, unless otherwise indicated.

Communicated by J. Dean

A. Kremer
INRA, UMR BIOGECO,
Bordeaux, France

R. Sederoff · N. Wheeler (✉)
Forestry and Environmental Resources,
North Carolina State University,
Raleigh, NC, USA
e-mail: nickwheeler@scattercreek.com

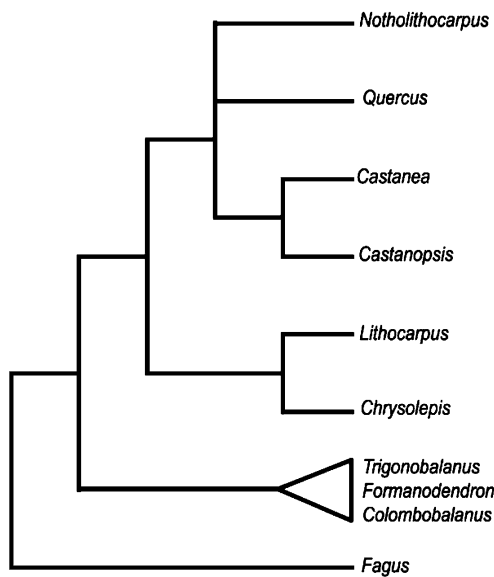


Fig. 1 An abbreviated phylogeny of the Fagaceae (adapted from Manos et al. 2008)

Development of genomic tools for the fagaceae

Gene discovery and expression (ESTs)

Bakarat et al. and Plomion et al. described characterizing the Fagaceae transcriptome through deep Expressed Sequence Tag (EST) sequencing, providing a resource for gene discovery, and a source of molecular markers for genetic mapping and linking genetic and physical maps. cDNA libraries and ESTs have been established by parallel projects in the USA and Europe. The US team combined 10K Sanger reads with 454 pyro-sequencing and produced over 2M reads for five Fagaceae species, resulting in 40,000 unigenes from American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*), over 19,000 unigenes from white oak (*Quercus alba*) and red oak (*Quercus rubra*), and ~7,000 unigenes from American beech (*Fagus grandifolia*). Thousands of the unigenes are full or nearly full-length cDNAs. These resources are available at the Fagaceae website (<http://www.fagaceae.com>). In silico comparisons between the transcriptome of canker and healthy stems in American and Chinese chestnut, as well as between canker tissues from both species led to the identification of candidate genes for resistance to the blight disease. In addition, the mining the cDNA sequences has allowed the identification of hundreds of chestnut microRNAs, of which several target defense-related genes.

As part of the EVOLTREE project, the European team obtained (1) over 145,000 Sanger ESTs from 20 cDNA

and SSH libraries made from various oak tissues (bud, leaf, differentiating xylem, root), developmental stages (bud phenology) as well as abiotically challenged trees, and (2) 2M 454 reads aiming at the discovery of genes differentially expressed between early and late flushing populations at both endo- and eco-dormancy stages, and between *Quercus petraea* and *Quercus robur*. A combined Sanger-454 unigene set is under construction and will provide a detailed catalog of the oak transcriptome, and also a source of single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) markers for linkage and association mapping. These resources are available through the *Quercus* portal at <https://w3.pierroton.inra.fr:8443/QuercusPortal>.

Genome sequence

One of the highlights of the meeting was the announcement (Nelson et al.) of the Forest Health Initiative (<http://foresthealthinitiative.org/>) which will support the production of a high quality reference genome sequence for Chinese chestnut (*C. mollissima*) as well as research on genetic engineering and marker technologies for breeding in American chestnut. The sequencing effort will be carried out at Penn State University (Carlson et al.) through deep “next generation” DNA sequencing. The reference genome will be assembled with the physical and genetic maps developed by the NSF Fagaceae Tools project. To facilitate the discovery of blight resistance genes, genome sequences may also be produced for American chestnut and blight-resistant hybrids by re-sequencing.

Mapping populations in the Fagaceae

A number of populations have been developed for mapping in Chinese, American and European chestnuts and their hybrids (Villani et al., Sederoff et al., Carlson et al., Kubisiak et al.). In the USA, two crosses have been made to establish mapping populations with three *C. mollissima* parents and two American by American crosses have been made. An F_2 population of a *C. dentata* x *C. mollissima* hybrid has recently been enlarged. The progeny from the crosses made for the NSF-sponsored project are available to the scientific community as tissue samples, scions, cuttings, or for study of the trees *in situ*. In Europe, five mapping populations in oak (three in *Q. robur*, one in *Q. petraea*, and one in inter-specific cross *Q. petraea* x *Q. robur*), one in European chestnut (*C. sativa*), and one in European beech (*F. sylvatica*) were constructed during the past decade and were the support for the comparative mapping activities conducted within EVOLTREE. Kremer et al. (2007) describe the status of mapping populations for the Fagaceae in considerable detail.

DNA marker development

Recent advances in high-throughput sequencing technologies have facilitated a striking increase in the number of DNA markers available for genetic studies in the Fagaceae (Carlson et al., Plomion et al.). Data for sequences from nuclear and organellar DNA of species representing several genera in the Fagaceae (*Fagus*, *Quercus*, and *Castanea*) are being mined for genetic markers; primarily SSRs, SNPs, and insertions or deletions (indels). Extensive EST resources are now available and proving extremely useful for identifying polymorphisms in large numbers of nuclear sequences from these genera (Kubisiak et al., Plomion et al., Lalague et al.). Marker development is focused on characterizing the inheritance and polymorphism information content of markers within and across species in the family. The transferability of EST-SSRs across the family is encouraging, at least for more closely related genera. Of particular interest is identification of polymorphisms in sequences that potentially have adaptive significance (Goicoechea et al.), including pathogen defense-response genes (Barakat et al.), or in sequences of potentially conserved orthologs (Lin et al.). Complete sequences of the chloroplast genomes of at least one species within *Fagus*, *Quercus*, and *Castanea* are also now available (Pinzauti et al.). Together, the growing numbers of both nuclear and organellar markers is providing a powerful framework for testing hypotheses related to the phylogenetics, biogeography, population genetics, and quantitative genetics of species across the family (Manos, Kremer, Finkeldey et al., Vazquez).

Genetic mapping and QTL analyses

High-density genetic maps are important tools that provide insights into the genome structure of organisms, identify quantitative trait loci (QTLs), facilitate the cloning of novel genes from bacterial artificial chromosome (BAC)-based physical maps (Fang et al., Faivre-Rampant et al.), and enable marker-assisted breeding. In North America, genetic maps currently are being constructed for *F. grandifolia*, *Quercus rubra*, *C. dentata*, and *C. mollissima* (Kubisiak et al.; Carey et al.) and in Europe, genetic maps are being constructed for *F. sylvatica*, *Q. robur*, *Q. petraea*, and *C. sativa* (Bodénès et al., Villani et al.). The identification and mapping of orthologous sequences across these species will greatly increase our understanding of genome organization and evolution in the family (Bodénès et al.). As a first step, consensus EST-SSR markers were developed from data mining in the OAK EST libraries and served for the comparative mapping in the three European genera. These efforts are currently extended by using COS markers. Quantitative trait locus (QTL) analysis is primarily focused on genes of adaptive significance (Gailing et al., Chancerel et al., Villani et al.) such as those

influencing bud break, species differentiation (Garnier-Gere et al.), water-use efficiency (Brendel et al.), and pathogen defense-response (Costa et al., Carey et al.). Eventual map-based cloning and verification of genes involved in these traits, as well as a thorough understanding of the underlying natural variation at these genes, is a major goal of both fundamental and applied research programs.

BAC libraries and physical mapping

Fang et al reported that three BAC libraries, each covering from 10 to 12× genome equivalents, were made from Chinese chestnut, and another BAC library with 12× genome coverage was constructed from American chestnut for comparative genomics studies. BAC clones from the Chinese chestnut libraries were analyzed by high information content fingerprinting. Further assembly of a total of 125,459 clones, representing 18× genome equivalents, by FPC v9.1 resulted in a physical map with 4,345 contigs and 17,132 singletons. A total of 579 markers have been integrated to the physical map with an eventual goal of anchoring ~1,500 genetically mapped markers. The map is accessible at <<http://www.fagaceae.org/cmmap/>> and is periodically updated.

A 12X BAC library has been constructed for one of the French reference *Q. robur* individuals (3P) for which both genetic and QTL maps are already available (Faivre-Rampant et al.). PCR screening with 60 genetic markers that spanned across the linkage map show that the library is useful for physical mapping, isolation of genes and genome sequencing. BAC end sequencing has provided a first glimpse of the genome sequence content and organization of the oak species.

Application of genomic resources

Genetics of biotic stress/disease resistance

Differential gene expression between Chinese and American chestnut (*C. mollissima* and *C. dentata*, respectively) at canker margins, paired with linkage to known blight resistance loci is being used to select candidate genes for chestnut transformation tests (Powell et al.). Resulting transgenic American chestnut plants will be tested for enhanced resistance to *Cryphonectria parasitica*, the chestnut blight pathogen. Since blight resistance is a quantitative trait and transformation procedures are difficult, it is important to develop a simple, early resistance assay that can detect resistance enhancement. Once resistance-enhancing genes are identified, they can be used as markers in breeding or placed in pyramid constructs. The knowledge gained on the quantitative resistance genes will benefit other tree species that are declining due to similar pathogens.

A common marker framework developed by the NSF-sponsored project “Genomic Tool Development for the Fagaceae” is being used for genotyping F₁ progenies coming from two inter-specific crosses established between European chestnut and Asian species (*Castanea sativa* × *C. crenata*, and *C. sativa* × *C. mollissima*) with the aim to construct genetic linkage maps for European chestnut and Asian chestnuts and to identify resistance QTL to *Phytophthora* sp. and to *C. parasitica* (Costa et al.).

Although hazelnut (*Corylus avellana*) is not a member of the Fagaceae, it shares a distant relationship (Order Fagales) and susceptibility to diseases similar to those that have decimated chestnut. Shawn Mehlenbacher of Oregon State University is routinely using marker-assisted selection for Eastern filbert blight (EFB) resistance in hazelnut <http://www.springerlink.com/content/mw161chwjhg6w2ll/>. Four single, dominant genes for EFB resistance have been mapped to four different linkage groups. BACs linked to the single dominant “Gasaway” gene are being sequenced to identify a gene for EFB resistance http://www.actahort.org/members/showpdf?booknrarnr=845_25. Molecular markers are also being used to survey genetic diversity within accessions of *Corylus avellana* at the USDA National Clonal Germplasm Repository at Corvallis, OR <http://www.springerlink.com/content/y29310k88515k617/>.

Genetics of abiotic stress/ecophysiology

Application of genomics in ecological genetics of Fagaceae have focused on adaptive traits such as drought, water use, and growth rhythm. Major QTLs for water-use efficiency (WUE) were detected in a *Quercus* pedigree (O. Brendel) and in a *C. sativa* pedigree (Villani et al.). Comparative location of QTLs within the *Quercus* genus showed genomic regions with more QTLs (Bodènès et al.). A candidate gene (ERECTA) potentially involved in the response of WUE has been analyzed in natural populations, showing 30% nucleotide differentiation value between *Q. petraea* and *Q. robur* (P. Garnier-Géré et al.). Results presented by V. Sork suggest genetic structure in valley oak (*Q. lobata*) associated with climatic variables, which will be tested in an association study using gene candidates related to cold and heat tolerance and osmotic stress. A similar project was presented by H. Lalagüe et al. which explored nucleotide diversity patterns along an altitudinal gradient in beech (*Fagus sylvatica*) for gene candidates involved in responses to abiotic stress. N. Schulz et al. presented a study monitoring gene expression and physiological responses where clonal seedlings developed from tissue culture of *Q. robur* were used in a 2-year drought stress experiment.

In France (Plomion et al., Kremer et al.), gene expression profiling and data mining with the existing EST resources were implemented to identify and catalog genes of adaptive

significance in European oaks and chestnut. Target traits were bud burst, bud set, length of growing season, and water-use efficiency. Nucleotide diversity was monitored in natural populations and SNP variation was recorded (Kremer). Additional SNPs were also inventoried in existing EST resources. A search for adaptive variation was extended by re-sequencing 1,000 genes in 23 oak trees (Garnier-Géré et al.). On-going efforts aim at mapping the whole set of genes in order to compare their location to QTL that were mapped in oak and chestnut pedigrees. QTLs for bud burst showed congruent location on oak and chestnut linkage groups (Kremer). SNP variation is currently being monitored in natural populations sampled along geographical gradients (latitude and altitude). SNP frequencies will be compared to the clinal variation observed for phenological traits. Finally SNP differentiation is scanned along linkage groups in order to identify hot spots of differentiation that will be compared with QTLs for adaptive traits. Similar approaches comparing QTL and F_{st} scanned along linkage groups are also implemented at the species level in order to identify genomic hotspots related to adaptive divergence among species, or speciation genes.

Transformation and vegetative propagation

Two transformation pipelines have been developed for American chestnut (*C. dentata*) to test for resistance-enhancing genes from Chinese chestnut (*C. mollissima*) and other sources (Powell et al; Merkle et al). Organogenic cultures or somatic embryos are target tissues for *Agrobacterium*-mediated transformation and both have been used successfully to produce transgenic whole plants. The system at UGA incorporates liquid culture selection and provides a rapid method for generating new transgenic events, though multiple bottlenecks between *Agrobacterium* inoculation and transgenic plant production would probably limit usable (i.e., transgenic plantlet-producing) events to about 40 per person per year. The system at SUNY-ESF uses semi-solid medium selection and has a similar production rate. Although this method takes longer to generate events, once a useful event is identified, it can be multiplied in shoot culture and shoots can be rooted to produce whole plants in just a few months. The shoot multiplication and rooting approach would also be useful in clonal propagation of trees from the backcross breeding program of the American Chestnut Foundation. There are currently 15 gene constructs in the SUNY-ESF transformation pipeline, mostly from non-chestnut sources, but three of the constructs contain a Chinese chestnut laccase gene. UGA has developed a cryostorage system for embryogenic cultures that can be used to help conserve chestnut germplasm and to hold transgenic and conventionally bred material indefinitely while propagules from the lines are tested.

Bioinformatics

In Europe (Fluch et al.): an e-Lab and central data access portal has been developed as an online resource linking various distributed and heterogeneous database systems all over Europe within the framework of EVOLTREE. At www.evoltree.eu online queries can be submitted to twelve databases which are linked within the e-Lab. Results are joined by the information system to a single result set, resolving overlaps and links in the underlying data automatically. The TDWG Access Protocol for Information Retrieval (TAPIR)—an XML-based protocol that can be used for information retrieval in distributed database architectures—is used to realize the decentralized database queries.

Currently, 103,141 EST, 1,677 nucleotide and 589 protein sequences from various Fagaceae species originating from the project are available and serve for Basic Local Alignment Search Tool (BLAST) queries. Four mapping pedigrees, five association populations, and geo-referenced information on 3,767 individuals from natural populations of oak are contained within the data sets. Two hundred twenty-four physical and QTL maps are also accessible via the portal. Information on *Fagus* and *Castanea* pedigrees and association populations, about microarray experiments, as well as further EST sequences will be included in the near future. SNP and SSR data produced will be available via the system as well.

In the USA (Cheng et al): The Fagaceae Genomics Web (FGW, <http://www.fagaceae.org>) is a bioinformatics resource that provides genetic and genomic information about chestnut, oak, beech, and other species in the Fagaceae. The site serves primarily as a public repository for data generated by the NSF-funded project “Genomic Tool Development for the Fagaceae” (NSF#0605135). These data include both 454 and Sanger EST sequences, unigene assemblies, SNPs, SSRs, physical maps, genetic maps, and BLAST homology results between unigene assemblies and large public databases such as ExPASy SwissProt and NCBI nr. The site also provides an online blast tool and fulltext searching capabilities. Most recently, unigene contigs have been annotated with GO terms; InterPro protein motifs and domains; and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, protein families and orthologs. Interactive expandable trees are available for browsing KEGG and Gene Ontology (GO) hierarchies as well as pie charts for GO summaries. Additionally, the project aided in the construction of two new open-source software tools: Hybdecon and Tripal. Hybdecon supports multi-dimensional overgo pooling strategies by providing an improved graphical interface for hybridization screening. The software then maps markers to BAC clones through a deconvolution process. Hybdecon was used in the creation of the Chinese chestnut (*C. mollissima*) physical map now accessible on the FGW site. Tripal, is a toolkit for construction of genomic websites. It combines

main-stream content-management software with standards for storage of genomic data.

European initiative

Another exciting highlight of the meeting was the announcement that the Directorate General for Research at the European Union (EU) is supporting a coordination and support action (CSA) called Foresttrac aimed at reinforcing EU and North American cooperation in the field of forest genomics and adaptation (Kremer, Final Session). The CSA will last two years and will deliver (1) a mapping of current research capacity in forest ecosystem genomics, (2) a validated research roadmap regarding adaptation to climate change, (3) a set of joint science plans, and (4) a collection of genomic resources laying the foundation for future whole genome sequencing of ecologically and economically important tree species. All project outcomes will be validated during the CSA by a wide group of stakeholders from Europe and North America and will be disseminated widely via the project website and key dissemination. The expected follow up of FORESTRAC is a future collaborative research project that should be jointly supported by funding agencies in Europe and in North America. Graham Harrison from the National Science Foundation reviewed the opportunities for seeking funding from NSF for international collaborations.

Opportunities for future research collaborations

At the conclusion of the conference, attendees discussed how scientists from the USA, Europe, and elsewhere (Asia) might engage in near-term and long-term research collaborations focused on genomics of the Fagaceae. The group defined a number of potential topics that might benefit from combined efforts some of which were initiated immediately.

Acknowledgements The success of this conference and the completion of this report would not have been possible without the assistance and contributions of many individuals and sponsors. The authors wish to thank the following for making this meeting a reality: remaining members of the organizing committee (John Carlson, Jeanne Romero-Severson, Paul Sisco, Jennifer Koch, William Powell, Susan McCord); Kelley McCarter, North Carolina State University (FEOP), for meeting planning and facilitation; ArborGen, Roche and 454 Sequencing, The Schatz Institute for Tree Molecular Genetics, the North Carolina Biotechnology Center, the Institute of Forest Biotechnology and EVOLTREE for their financial and in-kind support. The authors are indebted to the many individuals that contributed to the report, including: Scott Merkle, William Powell, John Carlson, Ali Barakat, Oliver Brendel, Silvia Fluch, Eric Fang, CH Cheng, Paul Sisco, Tom Kubisiak, Rita Costa, Fiorella Villani, Paul Manos, and Christophe Plomion. Our apologies to any omissions of contributors.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits

any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Kremer A, Casasoli M, Barreneche T, Bodénès C, Sisco P, Kubisiak T, Scalfi M, Leonardi S, Bakker E, Buiteveld J, Romero-Severson J, Arumuganathan K, Derory J, Scotti-Saintagne C, Roussel G, Bertocchi ME, Lexer C, Porth I, Hebard F, Clark C, Carlson J, Plomion C, Koelewijn HP, Villani F (2007) Fagaceae Trees. In: Kole C (ed) Genome mapping and molecular breeding in plants. Vol. 7: forest trees. Springer, Berlin, pp 161–18
- Manos PS, Cannon CH, Oh S-H (2008) Phylogenetic relationships and taxonomic status of the paleoendemic Fagaceae of western North America: Recognition of a new genus, *Notholithocarpus*. *Madrono* 55(3):181–190