REVIEW

Genomics of Fagaceae

**Antoine Kremer · Albert G. Abbott · John E. Carlson ·
Paul S. Manos · Christophe Plomion · Paul Sisco ·
Margaret E. Staton · Saneyoshi Ueno ·
Giovanni G. Vendramin**

Abstract An overview of recent achievements and development of genomic resources in the Fagaceae is provided, with major emphasis on the genera *Castanea* and *Quercus*. The Fagaceae is a large plant family comprising more than 900 species belonging to 8–10 genera. Using a wide range of molecular markers, population genetics and gene diversity surveys were the focus of many studies during the past 20 years. This work set the stage for investigations in genomics beginning in the early 1990s and facilitated the application of genetic and quantitative trait loci mapping approaches. Transferability of markers across species and comparative mapping have indicated tight macrosynteny between *Quercus* and *Castanea*. Omic technologies were more recently developed and the corresponding resources are accessible via electronic and physical repositories (expressed sequence tag sequences, single-nucleotide polymorphisms, candidate genes, cDNA clones, bacterial artificial chromosome (BAC) libraries) that have been installed in North America and Europe. BAC libraries and physical maps were also constructed in *Castanea* and *Quercus* and provide the necessary resources for full nuclear genome sequencing projects that are currently under way in *Castanea mollissima* (Chinese chestnut) and *Quercus robur* (pedunculate oak).

A. Kremer · C. Plomion
INRA, UMR1202 Biodiversité Gènes et Communautés,
Cestas 33610, France
e-mail: christophe.plomion@pierroton.inra.fr

A. Kremer (✉) · C. Plomion
Université de Bordeaux,
Talence 33405, France
e-mail: antoine.kremer@pierroton.inra.fr

A. G. Abbott
College of Agriculture, Forestry & Life Sciences,
Department of Genetics and Biochemistry, Clemson University,
Clemson, SC 29634, USA
e-mail: aalbert@clemson.edu

J. E. Carlson
School of Forest Resources, Pennsylvania State University,
323 Forest Resources Building,
University Park, PA 16802, USA
e-mail: jec16@psu.edu

P. S. Manos
Department of Biology, Duke University,
Durham, NC 27708, USA
e-mail: pmanos@duke.edu

P. Sisco
The American Chestnut Foundation,
85 Stoney Hill Court,
Asheville, NC 28804, USA
e-mail: phsisco@gmail.com

M. E. Staton
Genomics Institute, Clemson University,
Clemson, SC 29634, USA
e-mail: mestato@yahoo.com

S. Ueno
Forestry and Forest Products Research Institute,
1 Matsunosato,
Tsukuba, Ibaraki 305-8687, Japan
e-mail: saueno@ffpri.affrc.go.jp

G. G. Vendramin
Plant Genetics Institute, National Research Council,
Via Madonna del Piano 10,
50019 Sesto Fiorentino, Italy
e-mail: giovanni.vendramin@igv.cnr.it

Keywords Fagaceae · Castanea · Fagus · Quercus · Genomics

Introduction

The Fagaceae is a large angiosperm family with species spread throughout the Northern Hemisphere (Fig. 1), from the tropical to the boreal regions. Oaks (*Quercus*), chestnuts (*Castanea*), and beeches (*Fagus*) are the only genera that are distributed in Asia, Europe, and North America where they cover very large continuous forests and constitute important forest resources that are harvested for a wide range of uses (biomass, fiber, wood products, and food). Besides their contributions to the economy, they are keystone species in forest ecosystems and are major drivers of terrestrial biodiversity. In most countries, they are also considered as major patrimonial and cultural resources (Logan 2005). Despite their economic, ecological, and cultural importance, research in genetics and genomics in the Fagaceae has received less interest than research in conifers because of biological constraints to genetic investigations (e.g., long generation and longevity). Earliest provenance experiments in oaks were initiated in the early 1900s by Cieslar (1923), and provenance research has continued throughout the twentieth century (Kleinschmit 1993). By the late 1980s, population genetics dominated genetic research in Fagaceae. Extensive genetic surveys were conducted using different markers in oaks (Kremer et al. 1993), beech (Comps et al. 1987), and chestnut (Villani et al. 1994; Kubisiak and Roberds 2006) and gene flow or hybridization was monitored in natural stands (Bacilieri et al. 1996; Dow and Ashley 1996). Genomics in the Fagaceae started with linkage mapping in chestnut (Kubisiak et al. 1997), oak (Barreneche et al. 1998), and beech (Scalfi et al. 2004) and later on with comparative mapping between oaks and chestnut (Barreneche et al. 2004; Casasoli et al. 2006). More recently, extensive research motivated by forest health and adaptation issues has been conducted mainly in American chestnut and European oaks (Kremer et al. 2007). Additionally, oaks are also model species to investigate genomic delineation of species, as hybridization frequently occurs between species of the same taxonomic sections of *Quercus*. A conference was recently held where the status of knowledge in Fagaceae genomics was reviewed (Kremer et al. 2010). This contribution is a follow-up synthesis of the conference and provides a general overview of past and ongoing genomic research in this important angiosperm family.

Taxonomy and phylogeny

The angiosperm family Fagaceae, comprising 8–10 genera and approximately 900 woody species, contains 2 highly diverse morphological groups that dominate a variety of habitats throughout the Northern Hemisphere. Temperate, subtropical, and semiarid floras are particularly rich in *Quercus* (415 species), whereas warmer, more mesic forests of Southeast Asia are comparably diverse in the castaneoid or chestnut-like genera *Lithocarpus* (300 species) and *Castanopsis* (120 species). Disparity in clade size is striking across Fagaceae as at least 4 genera have fewer than 12 species. Biogeographic patterns (see Fig. 1) combine several widespread genera (*Fagus*, *Castanea*, and *Quercus*) with narrow endemics (*Trigonobalanus* sensu lato, *Chrysolepis*, and *Notholithocarpus*), and there is clear evidence for the in situ diversification of regionally distributed subgroups within the larger genera (Axelrod 1983; Manos and Stanford 2001).

The wind-pollinated genus *Quercus* has been well-studied using molecular phylogenetics and two major clades have been identified (Manos et al. 1999; Bellarosa et al. 2005; Oh and Manos 2008; Denk and Grimm 2010). The larger of the two clades has been referred to as the New World oaks with sections *Lobatae* (red and black oaks), *Protobalanus* (golden cup oaks), and *Quercus s.s.* (white oaks). This concept emphasizes a putative origin and subsequent diversification in the Americas followed by at least one dispersal of an ancestral group of white oaks to Eurasia that gave rise to a small radiation of approximately 20 species placed in section *Quercus s.s.* (e.g., *Quercus robur*, *Quercus mongolica*). The smaller clade is strictly Eurasian and unites two groups: the Cerris oaks (including species of the *Ilex* group), a modest radiation of approximately 50 species of temperate (e.g., *Q. acutissima*) and semiarid (e.g., *Quercus ilex*, *Quercus suber*) regions (Denk and Grimm 2010), and the tropical, evergreen *Cyclobalanopsis* or cycle cup oaks (e.g., *Q. myrsinifolia*) with about 90 species (Deng 2007). This branching pattern challenges previous classification schemes that recognized the two oak subgenera, *Quercus* and *Cyclobalanopsis*. Subgeneric circumscription based on morphological characters is potentially at odds with the new pattern, such that the decurrent styles and expanded stigmatic surfaces previously used to define subg. *Quercus* may now require a more complex interpretation. Despite the apparent paraphyly of subg. *Quercus*, the genus *Quercus* in the broad sense remains united by floral synapomorphies and a circular single fruit or acorn surrounded by the conspicuous accessory referred to as the cupule or involucre.

These new phylogenetic alliances have clarified the origin of the genus *Quercus* and the evolution of its single-flowered pistillate inflorescence and developing fruit surrounded by a valveless cupule (Fig. 2). The sister group of the oaks appears to be the monotypic castaneoid genus *Notholithocarpus* (tanbark oak) of western North America that also bears an acorn-like fruit (Manos et al. 2008). In turn, this group is sister to a clade including the genera
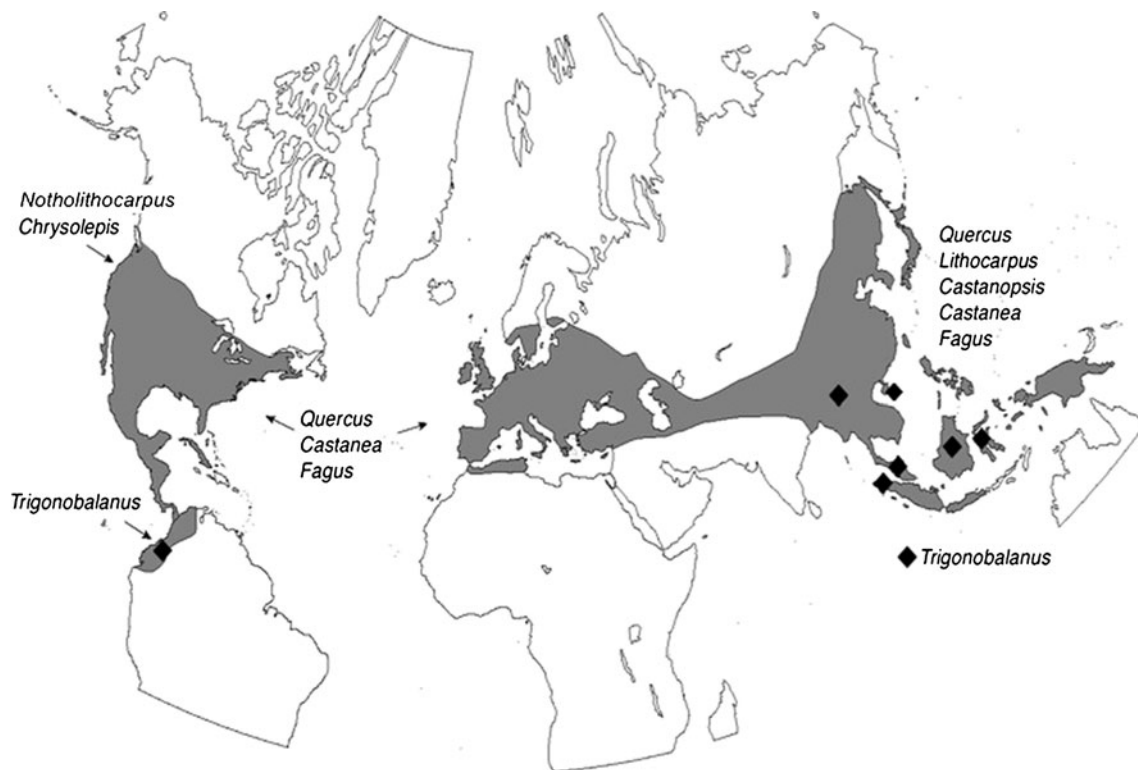
Fig. 1 Generalized distribution for Fagaceae (adapted from Manos and Stanford 2001). *Trigonobalanus* sensu lato is equivalent to *Colombobalanus excelsa* in South America; *Formanodendron doichangensis* in southwestern China and Thailand; *Trigonobalanus verticillata* in the Malay Archipelago and Hainan Island

Castanea and *Castanopsis*. Formerly recognized within the genus *Lithocarpus*, *Notholithocarpus densiflorus* is a paleoendemic tree species of the California Floristic Province (CFP), often occurring as a codominant in the redwood and mixed evergreen forests of the north coast ranges. The notion that *Notholithocarpus* is more closely related to the oaks and chestnuts than to its mistaken congeneric species of *Lithocarpus* may be a factor as the evolutionary dynamics of fungal infection and resistance become better understood.

Taken together, these results demonstrate that the oaks are more closely related to several castaneoid genera, confirming that the evolution of the acorn and the valveless cupule typically associated with *Quercus* is the result of the loss of lateral flowers in the dichasially arranged pistillate flowers, like those typical of the genus *Castanea*, and fusion of cupule valves. Interestingly, several independent losses of lateral flowers can be found in this clade (see Fig. 2), two in *Castanea* (Lang et al. 2007) and likely several more in *Castanopsis* (Manos et al. 2001), as well as elsewhere in the family (e.g., *Trigonobalanus*). This newly recognized clade of oaks and castaneoids brings together seemingly disparate morphological groups; however, the totality of vegetative, floral, and fruiting character states in the family is on display here, from evergreen and deciduous species to wind- and insect-pollinated pollination systems and singly

and dichasially arranged pistillate flowers with valveless and valvate cupules, respectively. In addition to the ecological importance of these species in the temperate zone, including the most drought- and cold-tolerant species in the family, several of them sit at the forefront of recent genetic and genomic initiatives (e.g., *Quercus alba*, *Q. mongolica*, *Quercus rubra*, *Q. robur*, and *Castanea mollissima*).

Recent studies also suggest that the sister group to the oaks and chestnuts includes an evolutionary branch containing the unlikely pairing of the two remaining castaneoid genera, *Lithocarpus* (stone oaks) and *Chrysolepis* (chinquapins). These two genera differ markedly in morphology, species number, habitat, and geography (Figs. 1 and 2). The two species of *Chrysolepis* occur mostly in redwood and mixed evergreen forest of the CFP of western North America. The most important morphological characteristic for defining *Chrysolepis* is a valvate cupule with internal valves separating the fruits (Hjelmquist 1948). This condition appears to be autapomorphic within Fagaceae, and the internal valves have been interpreted as vestiges of the branches of higher-order inflorescences (Nixon and Crepet 1989). In contrast, *Lithocarpus* is widespread in southeast Asia, with species occuring from Japan south throughout Indochina, Borneo, and New Guinea. In *Lithocarpus*, acorn-like fruits develop from each
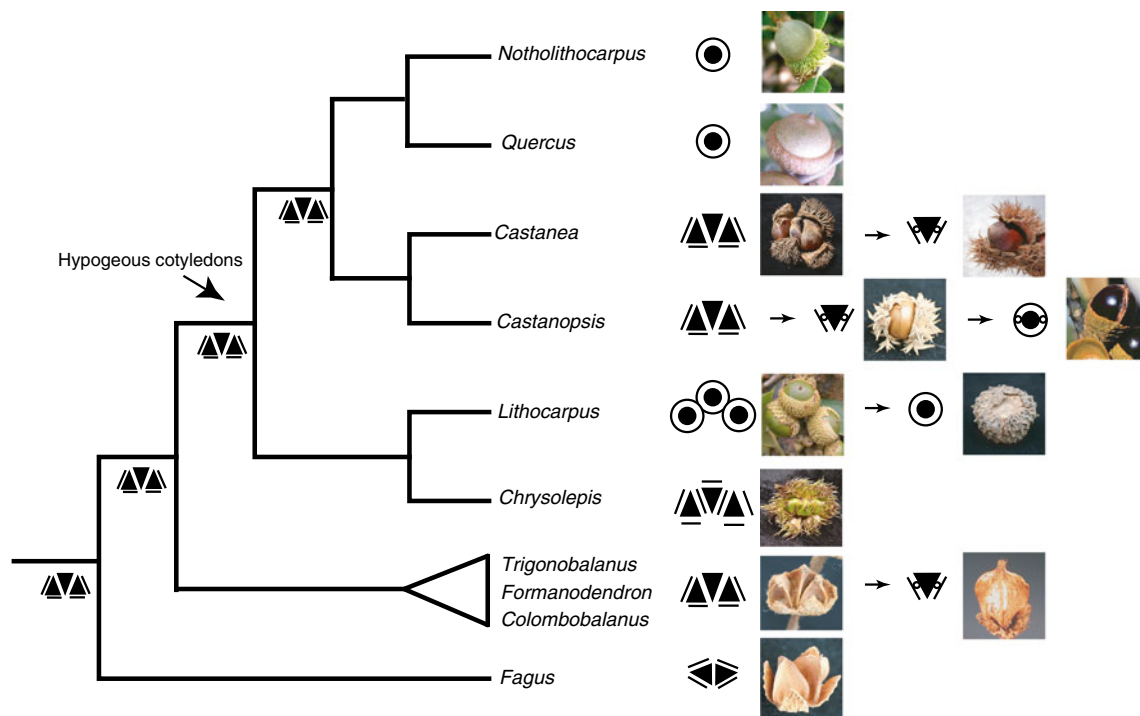
Fig. 2 Summary of the phylogeny of Fagaceae with a set of diagrams and images of representative cupule/fruit types (adapted from Manos et al. 2008). Cupule valves are indicated with *straight lines* and valveless cupules are indicated with *large circles*. Fruits are drawn with discs or triangles and aborted lateral flower position with small circles. For example, the uniloral cupule/fruit of *Quercus* is shown with a disc in a *large circle*. *Arrows* indicate observed transformations within that clade or within the inflorescence. Dichasium cupules are shown as the ancestral condition throughout the tree. Photographs are taken from the following species and listed from the top down: *Notholithocarpus densiflorus*; *Quercus coccinea*; *Castanea crenata* (*left*) and *Castanea pumila* (*right*); *Castanopsis tribuloides* (*left*) and *Castanopsis fissa* (*right*); *Chrysolepis chrysophylla*; *Lithocarpus fenestratus* (*left*) and *Lithocarpus rotundatus* (*right*); *Colombobalanus excelsa* (*left*) and *Formanodendron doichangensis* (*right*); *F. grandifolia*

pistillate flower of a dichasium, each one surrounded by a distinct valveless cupule. As a whole, *Lithocarpus* show abundant variation in inflorescence structure, sexual condition of the flower spikes, numbers of staminate and pistillate flowers in each dichasium, fruit type, and cupule type and ornamentation (Kaul 1987; Cannon and Manos 2000, 2001; Camus 1936-1954). Morphological support for *Chrysolepis* and *Lithocarpus* as sister genera may be found in their well-defined cupule/fruit apomorphies. One possible synapomorphy for this relationship is the flower cupule (sensu Forman 1966) or simply that each pistillate flower in a dichasium is surrounded by cupular tissue, although the origin and interpretation of forms within this subclade requires additional developmental investigation (Oh and Manos 2008).

As increased phylogenetic resolution has suggested new inferences on relationships and morphological transformations across Fagaceae, an important synapomorphy with clear ecological and evolutionary function has been identified. The evolution of hypogeous cotyledons occurred once in the evolutionary history of Fagaceae (Oh and Manos 2008). Given the novel arrangement of castaneoids + *Quercus*, only hypogeous cotyledons and the anatomy of the fruit wall (Soepadmo 1968) appear to unite this otherwise heterogeneous group of taxa relative to the early branching

genera, *Fagus* and *Trigonobalanus* sensu lato (see Fig. 2). The fossil record for Fagaceae shows that large, presumably hypogeous fruit evolved by the Middle Eocene, suggesting a relatively early transition to more specialized forms of animal dispersal (Crepet and Daghlian 1980; Manchester 1994). The animal dispersal syndrome, characterized by large nut size and lack of wings, also appears in the related Juglandaceae within the same time frame. The evolution of animal-dispersed fruit seems to reflect generalized coevolution between these families and rodents (Manchester 1987).

Early branching Fagaceae are associated with wind pollination, dichasially arranged pistillate flowers, valvate cupules, and smaller, passively dispersed trigonous fruit with epigeous cotyledons. This grade is comprised of relatively few species and often of limited geographic distribution (*Fagus*; *Colombobalanus*, *Formanodendron*, and *Trigonobalanus* = *Trigonobalanus* sensu lato). The monophyly of the continentally disjunct species of *Trigonobalanus* has been supported in most analyses. The great antiquity and relictual nature of these species is also evident from a variety of perspectives including apomorphies detected in pollen ultrastructure (Nixon and Crepet 1989), widespread distribution of fossil equivalents (Mai 1970; Crepet and Nixon 1989a, b; Kvacek and Walther 1989), and presence of anomalous morphology and polyploidy

(Chen and Sun 2010). Floral similarities with the oaks is remarkable among the three species; thus, *Trigonobalanus* sensu lato could also be defined by many of the floral and inflorescence features also found in *Quercus*, such as decurrent to dilated styles, expanded stigmatic surfaces, unisexual inflorescences, lax staminate inflorescences, and scabrate pollen exine structure. Despite unique apomorphies discovered for each species and a wealth of shared plesiomorphies, such as branched inflorescences, valvate cupules, and epigeous germination (Nixon and Crepet 1989), the recognition of a single genus seems appropriate based on molecular phylogenetic analysis and a unique base chromosome number of $x=7$ that is distinct from all other Fagaceae ($x=6$) (Chen and Sun 2010).

The genus *Fagus* is exceptional within the family because of its morphologically distinctive flowers, especially those of the pistillate inflorescence wherein only two flowers give rise to a pair of trigonous nuts surrounded by a valvate cupule. The most compelling interpretation of this structure is that two cupules, each with a single nut surrounded by two valves, are borne closely opposing one another. *Fagus* is likely of a high-latitude, Northern Pacific origin with a center of diversity in Asia and followed by bidirectional dispersion west to Europe and east through western North America and ultimately into eastern North America and Mexico (Denk and Grimm 2009).

In summary, recent analyses support the hypothesis that classifying Fagaceae based on reproductive character states is problematic because of the phylogenetic mixing of former elements previously placed in distinct subfamilies. For example, the four insect-pollinated castaneoid genera that had been treated as a cohesive taxonomic group, most often recognized at the subfamilial level (Castaneoideae), are clearly paraphyletic, forming a series of branches between *Fagus* at the base of the tree and *Quercus* + *Notholithocarpus*. Similarly, wind-pollinated taxa also have failed to be monophyletic. Thus, floral character states fail to accurately classify Fagaceae, much like fruit type (circular vs. trigonous) and cupule type (valveless vs. valvate). Instead, parallel evolution in key reproductive traits has dictated the major evolutionary themes in the family, and this has occurred within a relatively narrow spectrum of morphological possibility.

Organellar genomes

The size and complexity of the chloroplast genome is well-suited to short sequencing by synthesis reads. In this regard, during the past decade, the availability of complete organellar genome sequences increased significantly, largely fuelled by advances in high-throughput sequencing technology and the significant reduction in time and cost, resulting in more than 100 publicly available angiosperm chloroplast complete genomes at National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/genomes/). Until recently, the chloroplast genome of a single species (*C. mollissima*) was present in the database. In the last few years, within the EVOLTREE project, the complete chloroplast genome sequences of all European *Quercus* species and of *Castanea sativa* and *Fagus sylvatica* were obtained. For this purpose, a two-step approach was adopted:

- The full chloroplast genomes of *Q. robur*, *F. sylvatica*, and *C. sativa* (reference genomes) were generated using the traditional Sanger approach (Table 1) Multiple long PCRs were used to isolate chloroplast DNAs. Most of the cpDNA genome was amplified by using conserved primer pairs (Grivet et al. 2001). PCR fragments (about 140 kbp out of approximately 160 kbp) were pooled, nebulized into smaller fragments (approximately 900 bp), polished, and used to construct the libraries. Sequencing of gaps was performed designing internal primers.
- The chloroplast genomes of the other *Quercus* species and oak haplotypes (*Q. robur* and *Quercus petraea*) were produced and assembled using high-throughput resequencing approaches (Illumina platform).

Genome annotation was performed using the Dual Organellar GenoMe Annotator package and refined manually. As expected, the gene order and content of the newly sequenced chloroplast genomes are highly conserved. On the other hand, some structural differences were detected at four genes: *infA*, *accD*, *rps18*, *rpl22*, and *petA*. The gene *infA* is an example of chloroplast-to-nucleus gene transfer in land plants. The *infA* gene has been lost (either entirely or has become a pseudogene) in different lineages several times independently during angiosperm evolution. The analysis of the sequences of *infA* in three Fagaceae species (*Q. robur*, *C. sativa*, and *F. sylvatica*) revealed the presence of a functional gene in oak and of pseudogenes in beech and chestnut. The *accD* gene encodes acetyl-CoA carboxylase and is found in the chloroplast genomes of all analyzed plants. Their reading frame lengths vary not only between different genera but also within *Quercus* spp. The majority of *accD Quercus* spp. genes are composed of 553 codons length, while *Quercus calliprinos*, *Quercus crenata*, and *Q. ilex* (all of section Cerris) have 9 additional amino acids at the C termini of the protein. *F. sylvatica* and *C. sativa* show shorter *accD* genes, of 517 and 538 amino acids, respectively. In monocotyledonous angiosperms, the *accD* reading frame length is reduced and varies from 106 codons in *Oryza sativa* to 0 in *Zea mays* and this reduction is considered to be the cause of *accD* loss in monocot species. In contrast, the *accD* reading frame in coniferous species diversified in an opposite increasing direction (*Cryptomeria japonica*, 700 codons).

Table 1 Genome size of the chloroplast DNA molecule and its major components

Species	Total DNA (bp)	GC content (%)	LSC	SSC	IR
Q. robur	161,295	36.9	90,948	19,353	25,497
Q. petraea	161,322	36.8	90,977	19,349	25,498
Q. petraea ssp. *iberica*	161,402	36.8	91,036	19,366	25,500
Q. ilex	161,394	36.9	91,029	19,367	25,499
Q. suber	161,375	36.8	91,015	19,362	25,499
Q. lusitanica	161,436	36.8	91,054	19,376	25,503
Q. crenata	161,482	36.8	91,115	19,369	25,499
Q. macrolepis	161,405	36.8	91,048	19,357	25,500
Q. calliprinos	161,372	36.9	91,027	19,339	25,503
Q. dalechampii	161,392	36.8	91,016	19,373	25,500
Q. pubescens	161,356	36.8	91,000	19,358	25,499
Q. trojana	161,372	36.9	91,009	19,363	25,500
Q. virginiana	161,335	36.9	90,984	19,353	25,499
C. sativa	160,982	36.8	90,566	19,364	25,526
F. sylvatica	161,226	36.9	90,917	19,315	25,497

LSC large single copy, *SSC* small single copy, *IR* inverted repeat

The components of the ribosomal RNA are exclusively encoded by the plastid genome (plastome), while only a subset of the ribosomal proteins is encoded in the plastome. The chloroplast genomes of higher plants harbor a conserved set of 21 ribosomal protein genes which, with only few exceptions, are also found in the plastomes of all other plastid-containing taxa. A notable exception is the *rps18* gene which was recently found to be absent in the plastid genomes of several nongreen unicellular organisms. This gene was observed to be variable in length within Fagaceae, with *Q. rubra* and *F. sylvatica* displaying a length reduction of about 20 amino acids.

In all Fagaceae spp., *rpl22* consists of a pseudogene with multiple internal stop codons, suggesting that, in these species, the plastid-encoded copy may also not be functional. In other *Castanea* and *Quercus* species, a nuclear copy of *rpl22* has been identified, indicating a transfer to the nucleus. Finally, at the *petA* gene, a direct repeat of 370 nucleotides was observed in *Castanea* species and in a few *Q. petraea* individuals, suggesting a partial duplication of the gene.

Chloroplast genomes are excellent model systems in computational genomics studies, providing essential information for structural, functional, and comparative genomic studies. By comparing Fagaceae chloroplast genome sequences, 1,252 polymorphic sites, including single-nucleotide polymorphisms (SNPs), indels, and simple sequence repeats (SSRs), were detected which allowed discrimination of all the *Quercus* spp. haplotypes. The availability of new chloroplast markers opens up the possibility of developing innovative and valuable tools to assess the level of Fagaceae plastome variation for use in population genetic analysis, phylogeny construction, and to develop chloroplast-specific DNA barcodes for *Quercus* spp. and population identification.

Transcriptomics, proteomics, and metabolomics

Large scale transcriptomic investigations

Large-scale transcriptomic studies have been conducted for two Fagaceae species, aiming at elucidating genes associated with resistance against stem blight fungus infection for *Castanea* (Barakat et al. 2009) and to catalog genes in various tissues for *Quercus* (Ueno et al. 2010). These studies present the atlas of expressed genes in each species by massive sequencing of several libraries. Transcriptomic investigations were also implemented to identify cork-related genes for *Q. suber* and to develop microsatellite markers for some species in the Fagaceae.

The chestnut blight fungus (*Cryphonectria parasitica*) invaded North America from Asia before the turn of the twentieth century, infecting American chestnut stems and inducing the decline of the species. Genomic approaches to screen for candidate genes that confer resistance against the disease was recently initiated as one of the breeding and research efforts for recovery in the frame of a National Science Foundation (NSF)-funded project (NSF#0605135 "Genomic Tool Development for the Fagaceae"). Transcriptomes were collected from canker and healthy stems from susceptible American (*Castanea dentata*) and resistant Chinese (*C. mollissima*) genotypes. The pyrosequencing method using the Roche 454 platform (GS20 and FLX) produced 2.2 million reads with average read length of 198 bp to construct 93,018 contigs in total. From the canker transcript contigs, 5,636 genes from American chestnut and 8,369 from Chinese chestnut were tagged. The authors further compared the levels of gene expression between canker and healthy stem tissues as well as between Chinese and American chestnut transcriptomes

and found that genes related to stress resistance and response to biotic stimuli were upregulated in canker tissue and that more genes were identified in response to biotic and abiotic stimuli and stresses in Chinese chestnut tissues, suggesting that blight resistance might be associated with this difference (Barakat et al. 2009). This project also produced 8,000 Sanger reads from Chinese chestnut and about 510,000 Roche 454 platform reads from American beech (*Fagus grandifolia*), white oak (*Q. alba*), and red oak (*Q. rubra*) that could be assembled into 8,300, 22,000, and 28,000 contigs for *F. grandifolia*, *Q. alba*, and *Q. rubra*, respectively. Thousands of the unigenes are full-length or nearly full-length cDNAs. These resources are available at the Fagaceae website (http://www.fagaceae.org), a bioinformatic resource portal providing genetic and genomic information.

For *Q. petraea* and *Q. robur*, 20 libraries (9 suppression subtractive hybridization (SSH) and 11 cDNA libraries) sequenced by the Sanger method and 14 libraries sequenced by pyrosequencing method resulted in 146,000 Sanger and 1.9 million Roche 454 platform reads, respectively. These libraries originated from various tissues such as roots, buds, leaves, and xylem, some of which were submitted to different stresses. Global assembly produced 69,154 tentative contigs (TC) and 153,517 singletons, providing 222,671 nonredundant sequences: OakContigV1 (including alternative transcripts), 67.4 % of them had at least one hit in homology searches using protein and nucleotide databases. The expressed sequence tag (EST) catalog was examined for putative homologs of candidate genes involved in bud phenology, cuticle formation, phenylpropanoid biosynthesis, and cell wall formation, suggesting a good coverage of genes involved in these traits. This kind of genomic resource is a useful tool to discover genes of interest. All the data (contig and singleton sequences as well as annotations) are viewable and downloadable through the Oak Contig Browser (http://genotoul-contigbrowser.toulouse.inra.fr:9092/Quercus_robur/index.html) with user name: "oak" and password: "quercus33."

Transcriptome analysis for *Q. suber* was conducted (Soler et al. 2007) aiming at discovering a repertoire of genes in cork cells and to identify genes related to biosynthesis of suberin, which is a main component of cork. Although the number of unigene elements (UE) constructed was small (263 unique sequences including contigs and singletons), a set of candidate genes relevant to cork formation was identified by SSH and microarray techniques. The seasonal changes for transcript accumulation of selected genes were monitored by reverse transcription polymerase chain reaction (RT-PCR) and high levels of accumulation for suberin pathway and regulatory genes were identified (Soler et al. 2008). For *Q. suber*, a larger-scale transcriptome analysis has been conducted under the international framework "SuberGene." The sequences were registered in

dbEST, SRA as well as in the Oak Contig Browser described above.

Finally, a transcriptomic study was conducted in *Q. mongolica* var. *crispula* (Ueno et al. 2008), *Castanopsis sieboldii* (Ueno et al. 2009a), and *Fagus crenata* (Ueno et al. 2009b) to develop genic microsatellite (SSR) markers. Because SSRs are frequent within ESTs, they are a good source of genetic markers with functional annotation, though SSR frequency varies depending on the program and criteria used for SSR detection (12.8 % for *Q. mongolica* var. *crispula* (Ueno et al. 2008) vs. 23.7 % in *Q. robur* and *Q. petraea* UE (Ueno et al. 2010)). For *F. grandifolia*, *C. mollissima*, *C. dentata*, *Q. alba*, and *Q. rubra*, potential microsatellite markers can be found at the Fagaceae Genomics Web (FGW) site (http://www.fagaceae.org).

Gene expression studies

Gene expression studies that have been conducted so far can be grouped into two categories: (1) abiotic environmental adaptation, such as water use efficiency, bud dormancy and burst phenology, ozone stress, and wounding responses and (2) biotic interactions, such as those with ectomycorrhizae or *Phytophthora* spp.

Abiotic gene expression studies

Pedunculate oak (*Q. robur*) and sessile oak (*Q. petraea*) are the main European oaks and they inhabit contrasting ecological sites ranging from humid to semixeric conditions. They are frequently present in mixed stands under intermediate conditions, where they usually hybridize. Because the two oaks differ in their water use efficiency, transcriptomes from root tissue were collected from SSH libraries and differentially expressed genes between the two species under osmotically stressed environmental conditions were identified (Porth et al. 2005b). Genes upregulated after osmotic stress in sessile oak included betaine aldehyde dehydrogenase and oxygenase/oxidoreductase, which may confer water stress resistance to the species.

For sessile oak, Derory et al. (2006) constructed SSH libraries to enrich differentially expressed genes among early bud development stages and subsequently analyzed the expression differences by macroarray and RT-PCR experiments. They identified 233 transcripts that showed differential expression during early bud burst stages, many of which were in the cell rescue/defense-, metabolism-, protein synthesis-, cell cycle-, and transcription-related functional categories. To avoid frost and drought damage in the context of global climate change, an improved understanding of the signaling pathway for bud burst is crucial for biomass production and forest management. These differentially expressed genes are putative candidate genes for

future association studies between bud burst and gene variation.

Changes in gene expression through ozone exposure have been studied for *F. sylvatica*, which is the most important deciduous forest tree in temperate Europe. Because ozone is a harmful air pollutant, various effects (such as visible injuries and photosynthetic performance) of ozone on *F. silvatica* have been documented. After exposing leaves to ozone fumigation, Schneiderbauer et al. (1995) noticed the accumulation of extensin transcripts, which are involved in cell wall structure and plant defense against wounding or pathogen attack after ozone fumigation. Olbrich et al. (2005) used SSH and microarray techniques to identify 276 ozone-responsive ESTs belonging to different functional categories such as disease/defense, cell structure, signal transduction, and protein destination. Transcriptome analysis using a microarray for *F. sylvatica* adult leaves under field conditions with ozone fumigation following *Apignomonia errabunda* (endophytic fungus) infection showed transcriptional changes among sun/shade leaves, ozone treatments, and fungal infection, with largest transcriptional changes between sun and shade leaves and a few transcriptional changes after *A. errabunda* treatment (Olbrich et al. 2010).

Biotic gene expression studies

Gene expression analysis upon disease and/or symbiont infection is another topic of gene expression studies. Gene expression upon infection of ectomycorrhizae has been analyzed for pedunculate oak (*Q. robur*) with *Piloderma croceum* and European chestnut (*C. sativa*) with *Pisolithus tinctorius*. For *Q. robur*, Kruger et al. (2004) constructed an SSH library from inoculated roots to identify genes upregulated in the premycorrhizal phase. Genes relating to signal perception and transmission, stress response, metabolism, and growth, including pheromone receptor-like protein, formate dehydrogenase, metallothionein, and expansin, were identified. Frettinger et al. (2006) further found that one type of class III chitinase was upregulated in lateral roots but not in the principal roots during the premycorrhizal stage, suggesting that this gene may be associated with the initial steps of ectomycorrhiza establishment which occurs in lateral roots but not directly in principal roots in *Q. robur*. Frettinger et al. (2007) extended the previous study using cDNA microarray, RT-PCR, and quantitative real-time PCR analyses to identify transcriptional changes among preectomycorrhizal lateral and principal roots as well as mycorrhizae. They found that genes regulated in premycorrhizal lateral roots displayed nearly the same expression in mycorrhizae and that genes regulated in premycorrhizal principal roots often showed different expression from ectomycorrhizae. Furthermore, downregulation of genes classified in metabolism and upregulation of genes related to stress and

defense responses were identified, indicating the importance of these genes in the establishment of ectomycorrhizal symbiosis.

cDNA microarray expression analysis during initial contact (6 and 12 h) of *C. sativa* with *P. tinctorius* also demonstrated that genes associated with stress and defense response, signal transduction, and communication (cystatin, defensin, and receptor kinase-related protein) were upregulated (Sebastiana et al. 2009).

For *F. sylvatica* which is sensitive to the root pathogen *Phytophthora citricola*, Schlink (2009) collected transcriptomes from infected roots using the SSH technique. Library sequencing revealed pathogenesis-related genes, including cell wall proteins, protein kinases, and transcription factors responsible for pathogenesis and wounding. Schlink (2010) further analyzed root and leaf transcriptomes upon infection by *P. citricola* to gain insight into the change at the whole plant level using a cDNA microarray. She found downregulation of genes related to pathogenesis and upregulation of genes related to energy metabolism during the biotrophic stage.

Proteomics

Proteins that were identified in Fagaceae species are still limited (Abril et al. 2011). Jorge et al. (2005) described an optimized protocol for the *Q. ilex* leaf proteome. They analyzed 43 spots, most of which are related to photosynthesis and energy metabolism. Focusing on different developmental stages and drought conditions, they further analyzed changes in two-dimensional gel electrophoresis (2DE) profiles in four Spanish provenances (Jorge et al. 2006). Although the profiles of seedlings and adult trees were similar, qualitative differences were observed between watered and drought-stressed seedlings. They also found at least four spots differentiating provenances. Echevarria-Zomeno et al. (2009) analyzed 2DE profiles for *Q. ilex* seedlings during drought and recovery after drought stress. Acorn proteomics of *Q. ilex* analyzed among 10 populations in the Andalusia region showed that populations were clustered geographically as well as environmentally (Galvan et al. 2011). They suggested 56 spots as marker proteins that belonged to the storage and stress/defense categories.

For *Q. suber*, 17 proteins were differentially expressed between in vitro-derived haploid and somatic embryos. These proteins were involved not only in stress response and pollen development but also in cork biosynthesis (Gomez et al. 2009). Targeting cork formation in *Q. suber*, Ricardo et al. (2011) identified 54 proteins specifically associated with phellem activity. For *F. sylvatica*, Valcu and Schlink (2006a, b) developed an efficient method to recover proteins from leaves and roots for 2DE. Valcu et al. (2009) analyzed the proteome from leaves and roots that

were pathogen (*P. citricola*) infected or wounded. The authors indentified 68 protein spots which showed differential expression between infected and wounded tissue. Pawlovsky (2007) constructed a seed dormancy-breaking proteomic map for *F. sylvatica* nuts after imbibition in water, abscisic acid, or gibberellic acid (GA3) solutions. In total, 74 spots showed significant changes among the three treatments. Most of the proteins were involved in protein destination and energy metabolism with more proteins in the GA3 treatment, where defense-related proteins were also found. Expressional changes for *F. sylvatica* leaves after 3 years ozone exposure were analyzed using highly sensitive two-dimensional fluorescence difference gel electrophoresis (Kerner et al. 2011). Their results showed 142 protein spots (59 upregulated and 83 downregulated). Upregulated proteins were mainly involved in carbon metabolism/catabolism and mitochondrial electron transport chain, while downregulated proteins were related to the Calvin cycle and the photosynthetic electron transport chain, confirming the previous report on the effects of short-term ozone exposure for other plant species.

Metabolomics

Investigations of metabolomics are also limited in the Fagaceae. However, specific metabolites have been quantified along with transcriptome analysis. In an ozone response study of *F. sylvatica* leaves (Nunn et al. 2005), emission levels of ethylene and expression levels of genes that relate to ethylene biosynthesis were monitored. The authors found upregulation in both ethylene and mRNA levels of key enzyme for ethylene biosynthesis including ACC synthase and ACC oxidase. Betz et al. (2009) targeted shikimate pathway gene expression to analyze the effects of ozone on *F. sylvatica* saplings because the shikimate pathway is involved in the biosynthesis of various secondary metabolites. They identified a strong induction of some of the shikimate pathway transcripts and proteins after ozone treatments as well as elevated levels of conjugates of salicylic and gentisic acids which may be produced via the shikimate pathway. Although the simultaneous measurements of individual metabolites from many samples are still technically difficult, metabolomics along with transcriptome and proteome analyses may be a powerful tool to identify genes of interests and mechanisms of tree responses at molecular level.

Genetic mapping and QTL detection

Mapping populations were and are still being developed for three major genera of the Fagaceae: *Quercus*, *Castanea*, and *Fagus*. In this section, we will only report on the investigations of the last 3 years, since the paper by Kremer et al.

(2007) provides a comprehensive overview of the linkage mapping and quantitative trait loci (QTL) detection studies carried out before 2007. We will focus on new marker (EST-SSRs and SNPs) and mapping population development, as well as on the construction of second-generation linkage maps and their utilization to (1) study the genetic architecture of growth and adaptive traits, (2) learn about the nature and extent of divergence between related species, and (3) study the macrosynteny between Fagaceae species.

Available mapping populations in the Fagaceae

A number of intraspecific and interspecific full-sib (FS) families have been developed in chestnuts, oaks, and beeches (Table 2). In the USA, several crosses have been made with *C. mollissima* (Chinese chestnut) and *C. dentata* (American chestnut) parents to establish mapping populations. A three-generation outbred pedigree of a *C. dentata* × *C. mollissima* hybrid has also been produced. This material is currently being used to identify (by map-based cloning) gene(s) involved in the resistance to chestnut blight (*C. parasitica*), a necrotrophic ascomycete fungus that causes diffuse cankers (http://www.fagaceae.org). Two FS families from *F. grandifolia* have also been prepared for dissecting the major QTL involved in resistance to the insect that vectors bark disease (Koch et al. 2010, 2011). In *Q. rubra*, one FS family was a posteriori reconstructed from the open-pollinated progeny of a single tree using an exclusion method based on genomic SSRs. Acorns were harvested on the same maternal parent starting in 2000, yielding a total of over 600 F1s. The main goal of this mapping population is to detect QTLs for a series of traits related to heartwood color and resistance to the agent of the sudden oak death (*Phytophthora ramorum*). In Europe, five mapping populations for oak (three *Q. robur*, one *Q. petraea*, and one *Q. robur* × *Q. petraea*) were produced (https://w3.pierroton.inra.fr/QuercusPortal/index.php). One (*Q. robur*) has been used to identify genomic regions involved in species differentiation (Scotti-Saintagne et al. 2004a), to study the genetic architecture of quantitative traits (Saintagne et al. 2004; Scotti-Saintagne et al. 2004b; Scotti-Saintagne et al. 2005; Gailing et al. 2005; Casasoli et al. 2006; Parelle et al. 2007; Brendel et al. 2008; Derory et al. 2010), to map genes of interest (Porth et al. 2005a) and to examine the level of macrosynteny with European chestnut (Barreneche et al. 2004; Casasoli et al. 2006) for which one FS mapping population and linkage map was produced.

Recent DNA marker development in the Fagaceae and construction of second-generation linkage maps

During the past 2 years, the availability of ESTs for a number of Fagaceae species (Barakat et al. 2009; Ueno et al. 2010) has considerably increased the development of

Table 2 Summary of genetic material and linkage maps for *Castanea*, *Fagus*, and *Quercus* (updated from Kremer et al. 2007)

Genus	Species/country	Mapping pop	No. of offspring	No. of and types of mapped markers[a]	Genome size (cM)[b]	Reference
Castanea	*dentata*/USA	GMBig × Horn	166 (F1s)	224 SNP 37 SSR	803	Kubisiak, personal communication
	mollissima/USA	Mahogany × Nanking	179 (F1s)	885 SNP 207 SSR	743 (combined)	Kremer et al. (2007)
	mollissima/USA	Vanuxem × Nanking	158 (F1s)	871 SNP 217 SSR		
	mollissima × *dentata*/USA	R4T52 × R4T31	85 (F2s)	101 RAPD 13 isozymes 10 SSR 339 SNP	686	
	mollissima × *dentata*/USA	KYAdair1 × GL158	48 (BC1)	203 SNP	575	
Fagus	*grandifolia*/USA	1505×1504	46	258 AFLP[a] 10 SSR[a]	743 (combined)	Carey et al., personal communication
	grandifolia/USA	1506 × 1504	53			
Quercus	*rubra*/USA	SM1 × SM2	500 (F1s)	NA	762	
	robur/FR	3P × A4	14 to 96	81 AFLP 10 RAPD 51 gSSR 256 eSSR	753	Durand et al. (2010)
	robur/FR	3P × A4	46 to 96	386 SSR 53 SNP	825	Kremer et al. (2007)
	robur/NL	AltP1 × AltP2	101	108 eSSR	407	
	robur/DE	SL03 × EF03	96	100 AFLP 217 eSSR	680	
	petraea/FR	QS28 × QS21	78	353 eSSR 46 SNP	636	
	robur × *petraea*/FR	11P × QS29	84	279 eSSR 43 SNP	680	

[a] Sum over both parental maps

[b] Average over the two parental maps

gene-based markers, namely, EST-SSRs and SNPs/indels. SSRs constitute a marker of choice for establishing genetic linkage maps because of their high polymorphism, their abundance within genes, and their reproducibility and transferability across species. In oak, it was found that 23.7 % of the UE presented one or more SSR motifs (Ueno et al. 2010). Overall, these authors detected 52,834 SSRs (mainly dinucleotides and trinucleotides) in 38,653 UE. While tri-SSRs were more frequently found in coding regions, di-SSRs were more abundant in noncoding regions. Durand et al. (2010) designed primer pairs for a subset of 748 UE. Overall, 37.7 % (283) were found to amplify a single polymorphic locus in the reference mapping population of *Q. robur*. The usefulness of these loci for constructing a genetic map was then evaluated using a bin mapping approach (Vision et al. 2000). Bin maps were established for the male and female parents of this pedigree for which framework linkage maps based on amplified fragment length polymorphism (AFLP) markers were available (Scotti-Saintagne et al. 2004a, b). The bin set consisted of 14 highly informative offspring selected based on the number and position of crossover sites. A total of 256 EST-SSRs were finally assigned to bins and their map position was validated by linkage mapping using an extended set of 48 to 96 offspring (Durand et al. 2010).

The fact that SSRs were easy to handle and once developed were cost-effective in genotyping hundreds of individuals (also because of multiplexing possibilities) led the EVOLTREE consortium (http://www.evoltree.eu/) to genotype four additional mapping populations including intraspecific crosses between *Q. robur* and *Q. petraea* genotypes as well as an interspecific cross between these two species. Briefly, the set of SSR markers developed by Durand et al. (2010) was tested on the eight parental lines and a few offspring of these four FS pedigrees (Table 2). Individual maps were constructed with informative markers (93 gSSRs and 410 EST-SSRs) and a consensus map was generated for each species (Bodénès et al., submitted) including 398/275 markers in 933/767 cM for the *Q. robur/Q. petraea* consensus linkage maps. Evenly spaced SSRs were then selected to genotype the 4 different oak mapping pedigrees (in total 1,100 genotypes) for high-precision QTL mapping. In chestnut, over 200 EST-SSRs were also developed from EST data and integrated in two *C. mollissima* linkage maps. The EVOLTREE consortium also tested the transferability of EST-SSRs to other Fagaceae species. From an initial set of 198 tested *Quercus* primer pairs, rates of amplification ranged from very high (70–84 % in *Quercus* sp.) to medium (60 % in *C. sativa*) and low (16.5 % in *F. sylvatica*) (Bodénès et al., submitted), showing that SSR loci developed in oak could be transferred to chestnut with a relatively high success rate. By bringing together previously mapped anchor markers (12 EST-P from Casasoli et al. 2006 and 16

genomic SSRs from Barreneche et al. 2004) with 63 new EST-SSRs, Durand (2009) demonstrated the usefulness of EST-SSRs for comparative genomics in the Fagaceae. As illustrated in Fig. 3 for one linkage group, he identified orthologous regions between the oak and chestnut linkage maps, validating the high level of macrosynteny and colinearity previously reported by Casasoli et al. (2006) using fewer anchor markers.

More recent advances in high-throughput genotyping technologies have facilitated a remarkable increase in the number of DNA markers available for mapping in the Fagaceae. ESTs have been mined for SNPs, the most abundant form of genetic variation in the genome, to design SNP arrays. In chestnut, a 1,536 Illumina Goldengate SNP array was designed from the in silico mining of *C. mollissima* ESTs. Four mapping pedigrees were genotyped and from 203 to 885 SNPs were mapped. In oak, the potential of highly multiplexed SNP genotyping for genetic mapping was first evaluated using a 384 Illumina Goldengate SNP
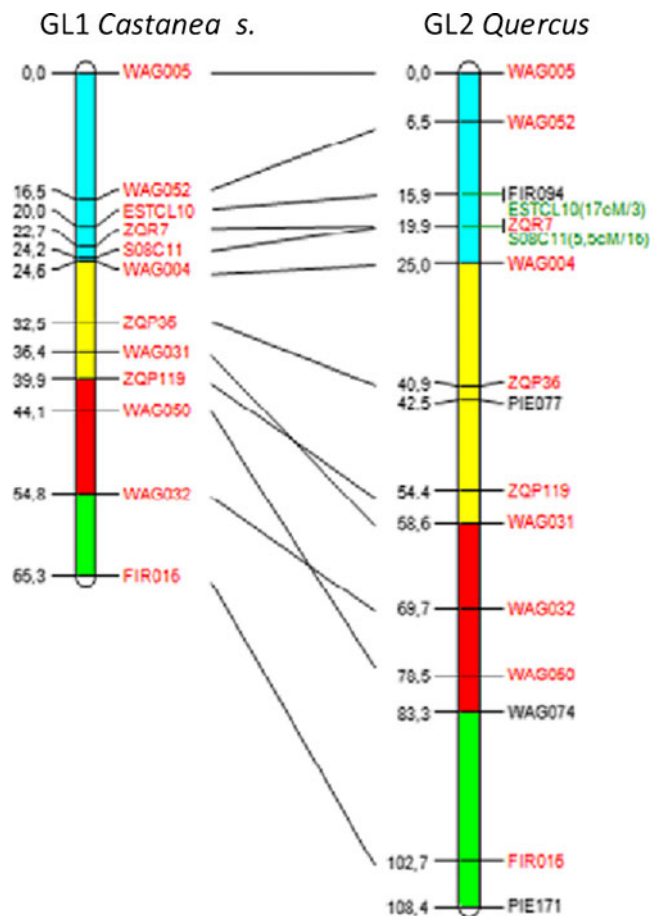

Fig. 3 Example of homologous linkage groups between *C. sativa* (*left*) and *Q. robur* (*right*). Linkage groups were named as in Durand et al. (2010) and Casasoli et al. (2006). Orthologous markers are linked by *plain lines*. Common intervals identified by orthologous markers were filled with corresponding color backgrounds in both oak and chestnut linkage groups (from Durand et al. 2010)

assay (Bodénès et al., submitted). SNPs were obtained in a set of 105 candidate genes for bud burst by either resequencing amplicons (Derory et al. 2010) or mining the EST collection (Ueno et al. 2010) following the strategy described by Lepoittevin et al. (2010). Genotyping was realized on 3 oak mapping populations, resulting in the mapping of 60 different loci. This first array was recently upgraded into an Infinium array (containing 8,000 SNPs) thanks to sequencing efforts using new generation sequencing technologies (Ueno et al. 2010) and resequenced amplicons. Four mapping pedigrees were genotyped and an ultradense linkage map is under construction. It will provide the raw material for the following: (1) to anchor future whole genome shotgun scaffolds to the oak linkage map, (2) for map-based cloning of major effect QTLs, (3) to study the genetic basis of incompatibility (intraspecific and interspecific crosses being available), and (4) to dissect the genetic architecture of adaptive divergence, part of these SNPs being already characterized in terms of between-species differentiation. This array also includes SNPs from comparative orthologous sequences defined within other dicotyledonous plants, providing a wide collection of anchor points for comparative genomics between oak and already sequenced plant genomes and, more specifically, to study the evolutionary history of the Fagaceae genome, as already illustrated for the Triticeae (Abrouk et al. 2010).

Update on QTL detection

Since the review by Kremer et al. (2007), three major QTL studies have been reported for *Quercus* using the reference cloned mapping population of pedunculate oak (3P × A4). These investigations explored the genetics of incipient speciation and adaptation in terms of number, position, and effects of major genes involved. Mapping and QTL data have been made available at the *Quercus* portal (https://w3.pierroton.inra.fr/QuercusPortal/index.php).

In the frame of two studies, Parelle et al. (2007) and Brendel et al. (2008) dissected the genetic architecture of traits related to species divergence between *Q. robur* and *Q. petraea*, two sympatric European white oak species. As mentioned earlier (see the "Gene expression studies" section), *Q. robur* is a pioneer species adapted to bottom lands and can withstand high moisture content of the soil. In contrast, *Q. petraea* requires drier and well-drained soils. A genomic scan based on conventional gene markers showed that a few genomic regions displayed higher interspecific divergence (Scotti-Saintagne et al. 2004a), suggesting the presence of "gene islands" of differentiation maintained by selection, despite extensive gene flow between the two species.

With respect to water logging, Parelle et al. (2007) detected six highly significant QTL regions for traits related to hypoxia responses: one for the rate of O_2 consumption by roots (an indirect index for root activity under hypoxia), two for the development of hypertrophied lenticels which are supposed to play a major role in hypoxia tolerance, and three for the degree of leaf epinasty, i.e., increase of the angle that petioles form with stems, which is known to be a highly specific hypoxia response. In their experiment, successive dates of sampling were used in a multienvironment QTL mapping model, which increased the power of detection of the QTL and the precision of mapping. Interestingly, the confidence interval of the position of a QTL for epinasty (in linkage group no. 8) coincided with a gene encoding an ethylene-responsive element-binding protein. Ethylene is known to play a role in epinastic growth in plants submitted to hypoxic stress (Jackson and Campbell 1976); therefore, this gene constitutes a relevant positional and functional candidate gene for this trait QTL. A bacterial artificial chromosome (BAC) clone containing this gene is being sequenced.

With respect to water use efficiency, Brendel et al. (2008) detected 10 QTLs for intrinsic water use efficiency (Wi) (the ratio of net CO_2 assimilation rate to stomatal conductance for water vapor), as estimated by leaf carbon stable isotope composition ($\delta^{13}C$). Repeated assessments over successive years clearly showed that one QTL of $\delta^{13}C$ (located on linkage group no. 11) contributed consistently to more than 20 % of the phenotypic variance of the trait, suggesting little influence by the environment. This region is the subject of intensive investigation, being the target of a map-based cloning approach. Four genomic regions in which colocalizing traits linked variation in Wi to variations in leaf chlorophyll and nitrogen content, leaf mass per surface area, and response of stomatal conductance to vapor pressure deficit were also identified. These localizations suggest a first set of physiological components underlying the diversity observed in the ratio of biomass accumulation to water used by transpiration in this pedunculate oak family. A detailed investigation on selected genotypes from this family (Roussel et al. 2009a, b) has shown that most of the diversity of Wi can be explained by differences in stomatal responses to environmental factors, mainly to incipient light. Regarding stomatal density, Gailing et al. (2008) detected 18 genomic regions on 10 linkage groups with a range of QTL effects varying from 3.6 to 15.9 % on male and female maps.

Derory et al. (2010) performed a meta-QTL analysis for bud burst assessed over 13 "year × site" observations. Initiation of the growing season, defined by bud flush date, has profound implications for adaptation of forest trees to their environment, with early flushing genotypes being susceptible to spring frost damage. A total of 19 QTLs were detected for bud flush (number of days necessary to reach stage 3, as described in Derory et al. 2006), most of them being expressed repeatedly over years or sites. Three QTLs

contributed repeatedly to more than 12 % of the phenotypic variation of the trait. Five candidate genes were found within the confidence intervals of the QTLs, three falling in the three major QTL regions. In particular, an α-amylase/ subtilisin inhibitor (ASI) overexpressed in quiescent bud (Derory et al. 2006) was located within the confidence interval of the strongest QTL for bud burst located on linkage group no. 9. A strong expression QTL for ASI was also identified in the same region (Le Provost, unpublished), suggesting that transcript abundance is controlled in *cis* by polymorphism(s) within or in proximity of the gene, reinforcing its functional, expressional, and positional attributes. The meta-QTL analysis of Derory et al. (2010) also showed how repeated phenotypic assessments either among clonal replicates or across years contribute to reduce substantially the confidence intervals of QTLs.

Perspectives

The genetic dissection of naturally occurring phenotypic variation (e.g., resistance to chestnut blight, bud burst in oak) has contributed to our understanding of the complex nature of quantitative traits during the last 15 years. However, we still lag behind in the exploitation of this useful information in breeding and/or forest tree management. At the dawn of the first Fagaceae genome sequences, that of *Castanea* (ongoing in the USA) and that of *Quercus* (starting in France), ultrahigh density linkage maps will allow the anchoring of whole genome shotgun scaffolds to the genetic maps and the reconstruction of chromosomes as long series of physical scaffolds. This resource will facilitate the map-based cloning of major genes and QTLs and shed new light into the molecular factors (including their allelic effects and epistatic and environmental interactions) involved in forest tree adaptation. Together with high-precision QTL mapping (that will require the development of appropriate plant materials and high-throughput phenotypic methods), the development of association mapping and reverse genetics approaches will also reinforce our understanding of the molecular basis of trait variation as illustrated in the other sections.

Physical mapping

For each plant family, it is important to develop model species and genetic and genomic resources for the identification, characterization, and cloning of important genes and for implementing the use of these resources for the purpose of sustaining plant production and species preservation. A portion of the current work on the genomes of chestnut and other Fagaceae species is focused on determining if important gene-containing regions are composed of evolutionarily

conserved genome blocks with specific marker associations and whether these genomic regions confer similar phenotypes in related species, suggesting conservation of function. In this regard, knowledge of the structural organization and genetic composition of trait-containing genomic regions of a model species genome may be directly useful for identifying the orthologous genomic regions of another species directly providing the tools necessary to genetically dissect these regions for the purpose of obtaining (1) more tightly linked markers for breeding, (2) genomic sequence contigs spanning the regions for gene identification and cloning studies, and (3) candidate genes located in these marked regions.

For the purposes above, the utility of having a complete genomics database (integrated genetic/physical map and mapped EST database) for a Fagaceae model genome species is clear. A physical map serves as an ideal tool to cross-compare maps of different species and to identify cloned genomic regions containing important gene loci, thus facilitating the process of gene marking and gene discovery in related species. In addition, physical maps provide marker resources that can serve to bridge the gap from the mapping of specific characters to the implementation of marker-assisted breeding schemes.

A first consortium was set up in the USA with teams of expertise focused on the development of an integrated genetic/ physical map of Chinese chestnut (*C. mollissima*) and the resources necessary for this purpose (i.e., BAC libraries, EST sequences, and a molecular marker set for bridging genetic map locations to the physical map contigs). A second consortium was developed in France focusing on oak. BAC libraries were developed for *Q. robur* and BAC end sequences obtained (Faivre-Rampant et al. 2011) to build long scaffolds in the framework of the oak genome sequencing project (funded by the French National Research Agency (ANR)). The large insert libraries produced by these two consortia are summarized in Table 3. This resource is currently being utilized for whole genome physical mapping in chestnut and as resources for marker-assisted genome walking for the identification of candidate genes for traits of importance to Fagaceae species arboriculture in chestnut and oak.

Two Chinese chestnut (*C. mollissima* "Vanuxem") BAC libraries that collectively represent 21× genome equivalents of the 794-Mb genome were used for physical map construction. Fingerprinting of the BAC clones using a high information content fingerprinting strategy followed by contig assembly resulted in a physical framework comprised of 126,445 clones in 1,377 contigs. Integration of the dense Chinese chestnut genetic map to the physical map was achieved via high-throughput overgo hybridization using probes derived from restriction fragment length polymorphism, SNP, and SSR markers. In total, approximately 1,500 molecular markers have been hybridized onto the physical map serving to

Table 3 Large insert library resources in the Fagaceae

Species/vector/accession/library designation	Average insert size/restriction enzyme	Number of clones	Genome coverage	Location
Castanea mollissima/BAC/Vanuxem/CMCMBb	128 kb/*Hin*dIII	73,728	11×	Clemson University Genome Institute
Castanea mollissima/BAC/Vanuxem/CMCMBd	115 kb/*Eco*RI	92,160	10×	Clemson University Genome Institute
Castanea mollissima/BAC/Vanuxem/CM_MBc	90 kb/*Hin*dIII	110,592	12×	Clemson University Genome Institute
Castanea dentata/BAC/GM Big/CDC_Ba	140 kb/*Hin*dIII	73,728	12×	Clemson University Genome Institute
Quercus robur/BAC/3P/3PEco	135 kb/*Eco*RI	92,160	12×	CNRGV INRA Toulouse
Quercus robur/BAC/3P/3PHind	120 kb/*Hin*dIII	98,304	14×	CNRGV INRA TOulouse

integrate this physical map with the molecular marker map for Chinese chestnut (Kubisiak et al., in preparation). The data for this genetic/physical map is publically available on the Fagaceae website at http://www.fagaceae.org. This physical map is already being utilized to identify chestnut blight and *Phytophthora* resistance QTL containing regions of interest for BAC contig sequencing and candidate gene discovery (Abbott, unpublished results).

In addition to the physical map contig information, BAC end sequences have been generated for the minimal tiling path of the Chinese chestnut physical map (Table 4). The BAC end sequences of the minimal tiling path are being used for merging physical map contigs, assisting in whole genome shotgun sequencing assemblies, and additional marker development.

Population genomics

Nucleotide diversity

There have been for the time being very few surveys of nucleotide diversity in Fagaceae species. While there are ongoing efforts to obtain the full genome sequence in *Castanea* and *Quercus* (see the "Full genome sequencing in the Fagaceae" section), there are no data on nucleotide diversity from previous investigations in the former genus and only a few reports for the latter genus. There are no published data on the other genera. Overall, the only reported studies concern the two oak species, *Q. petraea* and *Q. mongolica* var. *crispula* (Table 5). Assessment of nucleotide diversity is still limited to genes or portions of genes derived from ESTs catalogs that were constructed for identifying candidate genes of ecologically or economically important traits. As

estimates are limited to gene regions, they are likely to provide a downward biased picture of the whole genome. However, they may still permit comparisons with other studies (species or genera) conducted on the same genomic regions. Nucleotide diversity was monitored in 22 genes in 2 oak species (Table 5). The total mean diversity varies from 6.83×10^{-3} to 9.15×10^{-3} between the two species, with extreme variation among genes. Diversity in silent regions is fourfold larger than in nonsynonymous sites, most likely as a result of purifying selection. The genetic survey conducted in different populations (Table 5) also showed that the level of nucleotide diversity did not vary across sites and populations, except for *Q. mongolica* var. *crispula* where extreme values differed twofold (Quang et al. 2009). The values reported in Table 5 are among the highest that have been published so far for tree species that share similar life history traits (longevity, allogamy, and anemophily) such as conifers or other angiosperms (Table 5). Total and silent nucleotide diversity is, respectively, 4.3×10^{-3} and 7.5×10^{-3} in conifers. Earlier figures based on other molecular markers (Kremer and Petit 1993) already showed that oaks were among the most variable tree species and are close to those reported for maize (Wright et al. 2005) or wild rice (Zhu et al. 2007). These figures are in line with the very large effective population sizes of the oak species that have been studied so far, which cover very large and continuous ranges. These estimates may change once species with narrow and fragmented distributions are investigated. The level of nucleotide diversity (π) estimates the mutation drift equilibrium neutral parameter $4 N_e \mu$ (N_e being the effective population size and μ the mutation rate). None of the components of π has been so far estimated from either demographic investigation (for N_e) or divergence among related oak or Fagaceae species (for μ).

Table 4 Summary of BAC end sequencing details for Chinese chestnut physical map minimal tiling path

Species map	Number of BAC end sequences	Average sequence length (bp)	Total bases (Mb)	Genome spacing (kb)	Pairs	Genome coverage (%)
C. mollissima	43,143	622	26.8	18.4	19,005 (38,010 sequences)	3.4

Table 5 Nucleotide diversity statistics in oak species

Species	Gene	Gene acronym	N	N_p	Total length	π_{Tot}	π_{coding}	$\pi_{n\text{-}coding}$	$\pi_{Syn.}$	π_{Silent}	References
Quercus petraea	Galactinol synthase	GALA	44	9	505	14.7	11.3	17.4	39.01	20.66	1
	α-Amylase/subtilisin inhibitor	RASI	44	9	327	12.68	12.68	–	23.8	23.8	1
	Dof affecting germination 2	DAG2	42	9	360	1.56	1.56	–	5.43	5.43	1
	Histone H3	H3	52	9	374	3.56	2.3	4.2	8.34	4.65	1
	Seed maturation protein PM23	PM23	50	9	1,104	8.2	6.2	9.3	10.45	9.49	1
	Auxine-repressed protein	AUX-REPR	49	9	447	2.93	0.2	7	0.68	5.24	1
	Metal-nicotianamine transporter YSL1	YSL1	48	9	315	8.24	8.24	–	24.7	24.7	1
	GA3-β-hydroxylase	GA3	46	9	652	1.09	0.4	1.9	1.65	1.88	1
	GA20-oxydase	GA20	21	9	554	2.39	0.4	9.1	0	5.09	1
	Catalase2	CAT 2			336	3.16					2
	Auxin induced protein	AIP			556	6.61					2
	Cys-his-zinc finger protein	CHZP			385	2.09					2
	Glutathion-S-transferase	GST			477	8.27					2
	Cryptochrome 1	CRY1			367	4.20					2
	Dehydrin 1	DHN1	78	5	506	16.27	9.08	46.22	29.31	38.46	7
	Dehydrin 2	DHN2	96	5	554	8.46	8.05	9.02	23.0	11.43	7
	Dehydrin 3	DHN3	91	5	536	10.37	10.37	16.48	10.49	24.30	7
	Dehydrin 3 5′ UTR		88	5	377	64.62	/	64.62	/	64.62	7
	Mean values				8,227	9.15	4.96	16.78	11.49	16.85	7
Quercus mongolica sp. crispula	Methionine synthase		122	11	1,376	6.46				9.01	3
	Methionine synthase		120	9	1,385	6.67				7.11	4
	Cons 109		206	9	1,100	7.21				7.68	4
	Cons 14		184	9	1,015	6.99				7.15	4
	Mean values				4,876	6.83				7.74	
Populus tremula	77 genes		12 to 19	4	42,659	4.2		4.8	12		5
Populus balsamifera	590 genes		15	15	333,618	2.6			4.5		6

Silent sites include synonymous polymorphisms in coding regions and polymorphisms in non-coding regions

π nucleotide diversity ($\times 10^{-3}$), N total number of trees, N_p number of populations, 1 Derory et al. (2010), 2 Gailing et al. (2009), 3 Quang et al. (2009), 4 Quang et al. (2008), 5 Ingvarsson (2008), 6 Olson et al. (2010), 7 Vornam et al. (2011)

Skewed changes in the distribution of nucleotide diversity in a given genomic region (occurrence of rare alleles or maintenance of frequent alleles) may indicate footprints of natural selection. Indeed, after a selective sweep, mutations will create new alleles whose frequencies will be dependent of evolutionary forces as drift or selection. Such changes are captured by various statistical tests (Tajima 1989; Fu and Li 1993; Fay and Wu 2000). The reported studies on nucleotide diversity of oak species (Table 5) indicate very few significant departures from neutral expectations, but a general trend across genes for negative values of Tajima's D test, which would indicate potential directional selection or suggest a population expansion.

Linkage disequilibrium

Linkage disequilibrium (LD) that quantifies the nonrandom associations of SNPs is an important population property to consider when searching for genetic correlations with causal SNP variants of traits of interest. Besides demographic factors (mating system, population subdivision, and admixture) that may impact LD over the whole genome, selection has a local impact by extending the selection sweep and, therefore, enlarging LD. As for nucleotide diversity, there are only very few data on LD that are limited to a few genes. This limiting sampling may show a downward biased picture of LD, as genes exhibit usually higher recombination than other genomic regions. LD has been dissected in detail in Q. mongolica var. crispula within the methionine synthase locus (Quang et al. 2008, 2009) in different populations, supposedly with different population sizes. As for other forest species, there is decline in the level of LD at long physical distance. Even at short distances (<400 bp) very low LD was recorded in Q. mongolica var. crispula, suggesting higher recombination rates in oaks than in

conifers for example (Quang et al. 2008). Indeed, these authors estimate the minimum number of recombination events to be 0.32 in *Q. mongolica* var. *crispula*, while 0.25 was reported in *Pinus sylvestris* (Dvornyk et al. 2002). These results are confirmed by the distribution of LD in eight candidate genes of *Q. petraea* (Fig. 4) at short distances. Finally, the results in *Q. mongolica* var. *crispula* indicated systematically stronger LD in fragmented populations, which would be indicative of recent bottlenecks that have reduced population sizes.

Finally, a genome scan of interspecific differentiation between *Q. petraea* and *Q. robur* depicted hot spots of differentiation (Scotti-Saintagne et al. 2004a) that are likely to be maintained by reduced recombination rates. These results suggest that recombination rates are unevenly distributed across the genomes and that lower recombination rates are to be detected in hot spots of species differentiation.

Population differentiation

Gene differentiation has only been explored in very few species, corresponding to those investigated for nucleotide diversity and showing divergent results (Table 5). The study on *Q. mongolica* var. *crispula* found larger differentiation in SNPs than in microsatellites, while *Q. petraea* showed congruent results of F_{ST} values assessed by SNPs and microsatellites. The discrepancy may be related to different sampling schemes used for different markers within each species. In *Q. mongolica* var. *crispula*, different populations were sampled, whereas in *Q. petraea*, the same set of populations was used for measuring F_{ST} for SNPs and microsatellites.

A recent investigation explored in more detail the distribution of F_{ST} in the genome and among different sets of populations within *Q. petraea*. A systematic survey of nucleotide differentiation sampled along an altitudinal and

latitudinal gradient was conducted in the French Pyrénées and in Western Europe (Fig. 5). The study compared the value of population differentiation of bud burst assessed in provenance tests and nucleotide differentiation within 74 candidate genes (179 SNPs in total). Candidate genes had been previously selected based on expression studies, QTL mapping, or known function of the genes within model species (Derory et al. 2006, 2010). The mean value of F_{ST} across all SNPs was 0.023 along the altitudinal gradient and 0.02 along the latitudinal gradient, which are the same as earlier observed for microsatellites. However, the distribution was slightly skewed, with a few SNPs exhibiting values exceeding expectations under neutral evolution. These outlier loci were thus considered as closely linked to SNPs undergoing directional selection or being themselves the causal variants. There was, however, no clinal variation of SNP allele frequencies that would be parallel to the phenotypic cline observed for bud burst in the provenance tests. If genome-wide scans (GWS) of F_{ST} are to be implemented to detect genes undergoing divergent selection in oaks and other Fagaceae species, they would need to be based on very large numbers of SNPs due to the very low LD that has been so far observed. However, outliers detected by GWS are likely to be the causal variation of the phenotypic trait under investigation.

Breeding

In *Quercus*, classical tree improvement programs have been ongoing for many years (Kremer et al. 2004; Schlarbaum et al. 2008), and mapping populations have been developed to discover QTLs for traits that are important commercially and ecologically (Plomion—"Genetic mapping and QTL detection" section of this review). In *Fagus*, genotypes have been identified that are resistant to *Cryptococcus fagisuga*, the scale insect that vectors beech bark disease. Crosses have been made to study the inheritance of resistance, and molecular markers are being developed to identify genomic regions associated with resistance (Koch et al. 2010, 2011). But by far, the greatest effort to date has been to breed *C. dentata* (American chestnut) for resistance to *C. parasitica* (chestnut blight) and, more recently, to *Phytophthora cinnamomi* (ink disease). This review will focus on the breeding programs in *Castanea*.

Introduction of chestnut blight into the USA and Europe

The American chestnut tree and the introduction of chestnut blight from Asia have been reviewed in many places (Hepting 1974; Anagnostakis 1987; Smith 2000). By 1950, millions of American chestnut trees were dead in North America, and almost all of those that remained were short vegetative sprouts
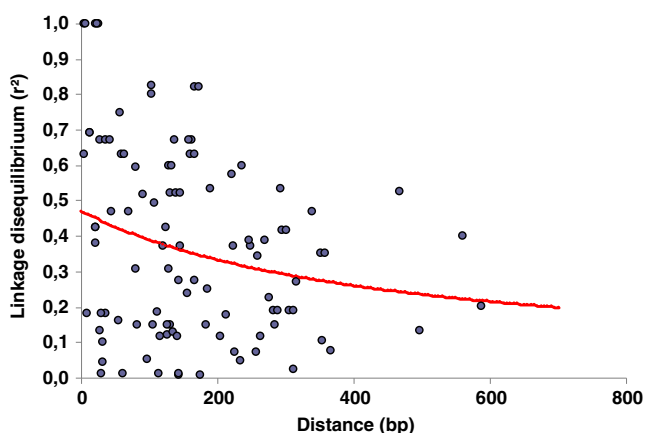

Fig. 4 LD (square of the correlation coefficient between two SNPs, r^2) in nine candidate genes of bud burst in *Q. petraea* (same genes as in Table 5) as a function of the distance separating the two SNPs

Fig. 5 Distribution of SNP differentiation values (F_{ST}) values between *Q. petraea* populations sampled along two geographic gradients. Data originate from 179 SNPs belonging to 74 candidate genes. *Blue bars* correspond to the F_{ST} values calculated between populations of the altitudinal gradient. *Purple bars* correspond to the F_{ST} values calculated between populations of the latitudinal gradient
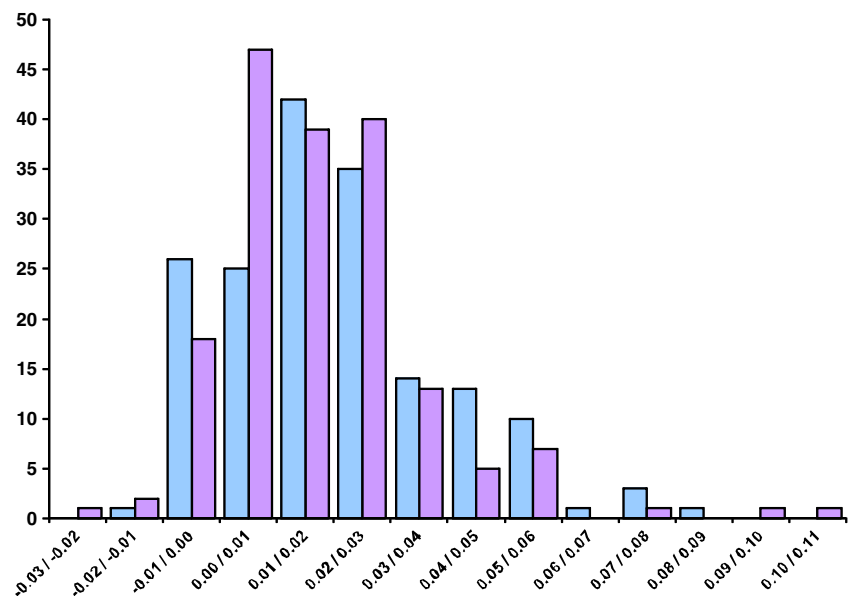
that did not flower. In 1938, the disease was discovered in Europe, and it continues to spread eastward on that continent wherever *C. sativa* (European chestnut) is found (Robin and Heiniger 2001). In Europe, however, some chestnut trees were found to be recovering from the effects of the disease, leading to the discovery of a virus that infected the blight fungus, weakening it in a phenomenon known as hypovirulence (Heiniger and Rigling 1994). Hypovirulence has not been as effective a biocontrol in North America (MacDonald and Double 2006).

Breeding for resistance to chestnut blight: the first efforts

In North America, it was quickly discovered that *C. parasitica* originated in Asia, that the Asian species of *Castanea* had resistance to the disease, and that both *C. mollissima* (Chinese chestnut) and *C. crenata* (Japanese chestnut) would make fertile hybrids with *C. dentata*. This led to the creation of several breeding programs designed to combine the blight resistance of Asian chestnut species with the timber qualities and forest competitiveness of American chestnut, primarily by making F1 hybrids between the species. Two of the most important of these early programs were conducted by R.B. Clapper of the US Department of Agriculture (USDA) and Arthur Graves at the Connecticut Agricultural Experiment Station (Clapper 1954; Anagnostakis 2007). Other breeding efforts, such as that by Albert Dietz using irradiated chestnut trees, have been reviewed by Burnworth (2002). By 1960, these programs had been abandoned, as interest in the chestnut waned and as the trees developed by these efforts were shown to be insufficiently blight-resistant as they matured.

Breeding for resistance to chestnut blight: the last 20 years

Interest in chestnut breeding was revived in the 1980s, primarily by the efforts of Gary Griffin of Virginia Polytechnic Institute and State University and Charles Burnham and colleagues at the University of Minnesota. Griffin (2000) developed a breeding program to enhance resistance within the pure *C. dentata* species by intercrossing a few large trees that managed to survive in the forest canopy before succumbing to the disease. Griffin's work has been supported by The American Chestnut Cooperators Foundation. A second breeding program was started by Burnham, who suggested that an American-type tree could be recovered by a series of crosses of Asian/American hybrids back to American to recover the American phenotype, followed by an intercross of moderately resistant trees to increase the resistance level (Burnham et al. 1986; Fig. 6).

Burnham's breeding scheme depended on his hypothesis that as few as two homozygous loci from Asian chestnut would be sufficient for blight resistance. His program, supported by The American Chestnut Foundation (TACF), was developed in more detail by Hebard (2006). A third major breeding effort was initiated by Robert Leffel, a retired USDA/ARS plant breeder living in Pennsylvania. Leffel became convinced that there were too many genes for blight resistance for backcrossing to work and that the goal should be a timber-type chestnut tree, regardless of how much *dentata* germplasm remained. His breeding scheme avoided much hand pollination by making use of cytoplasmic male sterility, often found in F1 American/Asian hybrids where the American parent was the female contributing the cytoplasm (Leffel 2004a, b; Sisco 2004). A fourth breeding program was begun by Joseph James, a retired physician in South Carolina, when he found that his farm was infested

Fig. 6 Breeding strategy for blight resistance in chestnut. The backcross breeding program proposed by Burnham et al. (1986) that is being used by TACF. In this diagram, petunia flowers are used to indicate the level of blight resistance as follows: *red* (high resistance), *pink* (moderate resistance), and *white* (low resistance). In each generation except the last, moderately resistant trees are chosen as parents. In the final generation, only highly resistant trees are left in the seed orchard. The diagram has been modified from the one designed by John Frampton at North Carolina State University, Raleigh, NC, USA
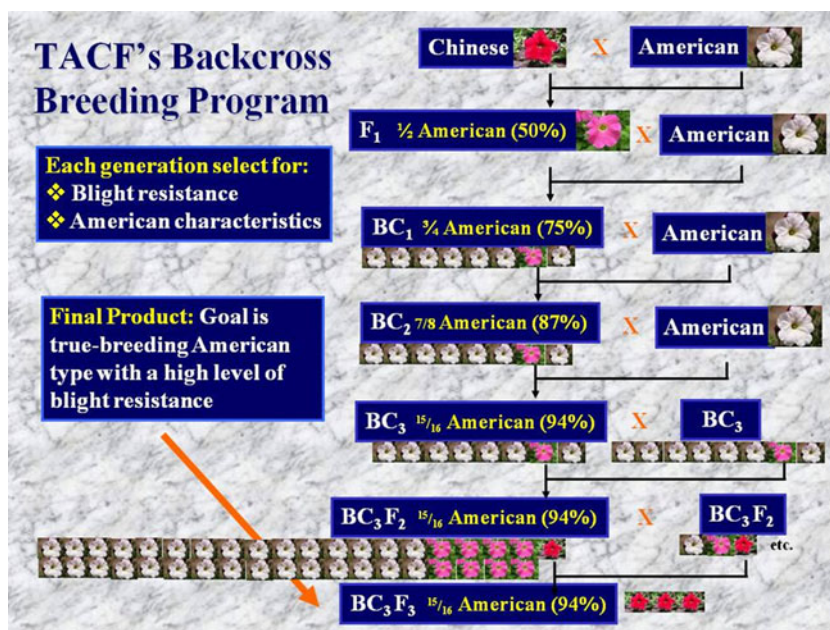
with *P. cinnamomi*, a causal agent of ink disease in chestnut (Crandall et al. 1945; Vannini and Vettraino 2001). James and Steve Jeffers of Clemson University teamed up to develop a screening method for resistance to ink disease, taking advantage of the fact that the Asian species of chestnut are resistant to the isolate of *P. cinnamomi* common in the southeastern USA (Jeffers et al. 2009).

Marker-assisted selection

Marker-assisted selection (MAS) has not been extensively used in chestnut breeding, but that may change in the next 3 years with the availability of more and better markers and as costs per data point decline. A first effort to map genes controlling chestnut blight resistance was published by Kubisiak et al. (1997), using a small F2 mapping population originating from a single Chinese chestnut cultivar, "Mahogany." Markers for three resistance QTLs revealed by that study are now being identified from a BAC library of Chinese chestnut cv. "Vanuxem" developed by a project funded by the US NSF (Wheeler and Sederoff 2009; http://www.fagaceae.org). More chestnut markers are being generated by a Forest Health Initiative (FHI; http://www.foresthealthinitiative.org). To date, the main use of markers has been to screen for American chestnut-specific alleles in advanced backcross progeny already selected for blight resistance by their canker phenotype (Hebard 2006; Shivanand Hiremath, personal communication). The purpose of this selection has been to eliminate as much of the Chinese chestnut genome as possible while retaining the phenotype for blight resistance. Sisco (2006a, b) discussed how markers might be used to select for blight resistance itself, as well as for pedigree verification, determination of the percentage

American chestnut in advanced backcross trees, investigation of chromosomal differences between Chinese and American chestnut (Islam-Faridi et al. 2009), and assessment of genetic diversity both in natural and backcross-derived populations (Kubisiak and Roberds 2006). At a recent planning meeting to discuss the future of TACF's breeding program, Jeanne Romero-Severson of Notre Dame University suggested using markers in the first backcross generation to accelerate the recovery of the American genotype (personal communication). This may be very useful as additional Asian chestnut cultivars are evaluated for possible incorporation into the breeding program.

Genomic selection

Grattapaglia and Resende (2011) have described a method they call "genomic selection." This may be used in the future for chestnut breeding as more markers are developed, as better correlations are determined between markers and traits, and as marker technology becomes less expensive.

Transgenic technology

For over 20 years, William Powell and Charles Maynard of the State University of New York—College of Environmental Science and Forestry and Scott Merkle of the University of Georgia have worked to perfect transgenic technology for the improvement of chestnut. Several technical problems have been overcome, such as establishing embryogenic regeneration systems, determining the best methods for transformation and screening of transformants, improving plantlet production, and acclimatizing plantlets to the natural environment

(Andrade and Merkle 2005; Powell et al. 2006; Merkle et al. 2007; Maynard et al. 2008; Andrade et al. 2009). Significant support for transgenic chestnut research is now being provided by the FHI, sponsored by the USDA Forest Service, Duke Energy, and the US Endowment for Forestry and Communities (http://www.foresthealthinitiative.org). A major focus is the transfer and testing of candidate genes for resistance to chestnut blight and ink disease. The project also includes groups working on shepherding the transgenic chestnuts through regulatory review by US federal agencies and on gauging public opinion with regard to releasing transgenic chestnuts into natural forests. Cross-pollination between transformed and nontransformed chestnuts in the forest would increase genetic diversity and adaptation to local environments of the transformed trees. Three questions being researched by the project are: (1) what is the most effective combination of transgenes for conferring resistance to chestnut blight and/or ink disease; (2) what is the best promoter for driving the transgenes (e.g., tissue-specific, wound-inducible, or constitutive); and (3) will either the transformation process itself, the construct, or the promoter have any undesired or unanticipated effects on the transformed tree or its environment (Ahuja 2011)? A matrix of constructs and promoters of many types will be tested in the next few years, with the goal of having at least one chestnut tree engineered for blight resistance deregulated by 2020. Transgenic technology has also been used on *C. parasitica* and its hypoviruses to permit hypovirulence to spread more quickly and more widely in a stand of chestnut trees (Nuss 2005).

Other pests and pathogens of chestnut that need to be addressed in the future

Dryocosmus kuriphilus, the Asian chestnut gall wasp, was introduced into North America in the mid-1970s (Payne et al. 1975) and has become a major pest of chestnut, causing defoliation, repression of blooming, and in the case of a severe infestation (Fig. 7), death of the tree (Rieske 2007). More recently, the same pest has been introduced into the chestnut-growing areas of northern Italy (Graziosi and Santi 2008). Two species of Ambrosia beetle, *Xylosandrus germanus* and *Xylosandrus crassiusculus*, have been introduced to North America and have become pests of chestnut by introducing pathogenic fungi into the stem. Asian and American species of chestnut are a preferred host of these insects in North America (Oliver and Mannion 2001). Without genetic resistance, pyrethrins must be sprayed at regular intervals on the trunks of chestnut trees where infestation is anticipated. This is one problem that might be addressed by genetic engineering of *Bt* genes into chestnut, as has been done with maize, potato, and cotton (Shelton et al. 2002; Brookes 2008; Rashid et al. 2008).

Fig. 7 Infestation of Asian chestnut gall wasp (*Dryocosmus kuriphilus*) on a backcross chestnut tree at the Biltmore Estate, Asheville, NC, USA in the spring of 2010. The tree should be fully leafed out, but instead most buds are infested with gall wasp larvae

Molecular marker resources and electronic resources

Online data repositories are the fastest and most convenient method for scientists to download public data and communicate ongoing research. The community of Fagaceae researchers has already recognized the need to connect and share diverse datasets spanning the fields of genetics, genomics, forestry, ecology, and evolution. Sharing raw data, analysis results, and ongoing work allows community members to efficiently find needed resources, to prevent duplication of effort, and to enhance collaborations across all forest tree science. Sequence data and electronically predicted or previously mapped SNPs and SSRs are particularly useful for comparative studies across species. Currently, three major online websites host Fagaceae data.

The EVOLTREE project (http://www.evoltree.eu) is a Europe-based network of researchers that are studying forest ecosystems across many forest tree species. The web portal acts as a clearinghouse for the diverse types of data emerging from the project, ranging across population, genetic, genomic, and mapping information. The eLab tool serves as an entry point for most types of data and allows users to filter for species of interest. EVOLTREE includes data from multiple genera in the Fagaceae family including *Castanea*, *Castanopsis*, *Fagus*, *Lithocarpus*, and *Quercus* (sensu lato). Metadata provided throughout the EVOLTREE sites allows users to identify what specific European partner is responsible for the given resource and how resources are being

used to contribute to the overall aims of the EVOLTREE network.

Oak species are a particular focus for the EVOLTREE project with the following resources described on the site: association populations, full-sib crosses as mapping pedigree crosses, natural populations, candidate genes, genetic maps, QTL maps, and ESTs. A separate website, *Quercus* Portal (https://w3.pierroton.inra.fr/QuercusPortal/index.php), has been developed to refer users directly to the oak resources within the EVOLTREE network. The *Quercus* Portal also has unique data not found through the main EVOLTREE website, including an updated unigene assembly and more genetic and QTL maps. The OakContigV1 transcriptome assembly of 1,704,117 ESTs can be searched or browsed via Gene Ontology terms, BLAST results, and tissue origin. Further analyses of the 69,154 OakContigV1 contigs yielded 52,834 putative SSRs and 36,411 putative SNPs, available for filtering and download in a variety of formats (Ueno et al. 2010).

Validated markers for mapping in *Quercus* can be found from either the EVOLTREE project or the *Quercus* Portal. Six hundred fifty-eight SSRs from *Q. petraea* (sessile oak) and *Q. robur* (pedunculate oak or English oak) are described with the locus name, motif string, repeat number, and primers, while 8,081 SNPs from the same two oak species are listed in the SNP database. SNP information includes the original sequence from which the SNP was derived, the putative function of that sequence, the target trait, and the mapping status. While a guest user may browse the SSR and SNP data, the ability to import and export the marker information in bulk is restricted to EVOLTREE members.

The FGW (http://www.fagaceae.org) houses genomic resources for five species of the Fagaceae family: Chinese chestnut (*C. mollissima*), American chestnut (*C. dentata*), red oak (*Q. rubra*), white oak (*Q. alba*), and American beech (*F. grandifolia*). Transcriptome sequences from each of the five species are available as unigene sets or as raw data downloads. For each unigene sequence, putative functional annotations from BLAST, KEGG, and InterProScan software results can be viewed with the alignment of the individual reads to the consensus sequence. The previously described physical map developed for Chinese chestnut is available for browsing. The FGW provides sets of in silico-mined SNPs and SSRs markers predicted from the transcriptome data of each species that may be downloaded as spreadsheets with predesigned primers and scores for likelihood of polymorphism. Of the five species, the deepest transcriptome sequencing was completed for the Chinese chestnut, yielding 12,539 SSRs and 25,904 SNPs predicted in silico. A subset of these markers has been mapped to yield a genetic map for Chinese chestnut. The FGW is in the process of being expanded into a more encompassing hardwood genomics website that will facilitate broader comparative genomics across all hardwood forest trees.

In addition to the specialized databases for Fagaceae species, three public databases (GenBank, EMBL, and DDBJ) are well-known and dbEST (http://www.ncbi.nlm.nih.gov/dbEST/), a part of GenBank, is focused on ESTs. Because ESTs are randomly collected partial sequences, database entries are redundant, with the same or mostly similar sequences resulting from the same transcripts and/or different splice variants of the same gene. The number of ESTs that are registered in dbEST increased exponentially with the advent of the genomic era where the development of new sequencing machines/techniques rapidly lowered the cost and time to collect sequences.

For Fagaceae species, the first ESTs in dbEST appeared in 31 July 2000 for *C. dentata* transcripts in the stem tissue (Connors et al. 2001) and the number of Fagaceae ESTs increased exponentially (Fig. 8), with the total number of 194,326 sequences from 11 species at present (31 March 2011). These sequences were collected by the Sanger sequencing method, with the average length of 569 bp calculated from all of the Fagaceae ESTs in the database. *Quercus* species, including *Q. petraea*, *Q. robur*, *Q. suber*, and *Q. mongolica* occupied most (77.1 %) of the total number of ESTs, followed by *Fagus* (16.2 %), *Castanea* (5.0 %), and *Castanopsis* (1.7 %) species (Table 6). It should be noted that 72.0 % of ESTs are from two European oak species (*Q. petraea* and *Q. robur*) in considering the database summary statistics. In terms of tissue from which ESTs were collected, ESTs from bud transcriptome are most popular (38.7 %), followed by root (26.3 %), bark (including phellem and xylem) (17.7 %), and leaf (15.4 %), with a minor portion of flower tissue (1.2 %). Some ESTs are also collected from callus by tissue culture.

Another useful resource for EST is the gene index project (http://compbio.dfci.harvard.edu/tgi/), where inventories of genes are constructed for representative species using a unified EST assembly protocol. Sequences are downloaded from dbEST and GenBank and subjected to quality control before they are assembled into TC using the Paracel Transcript
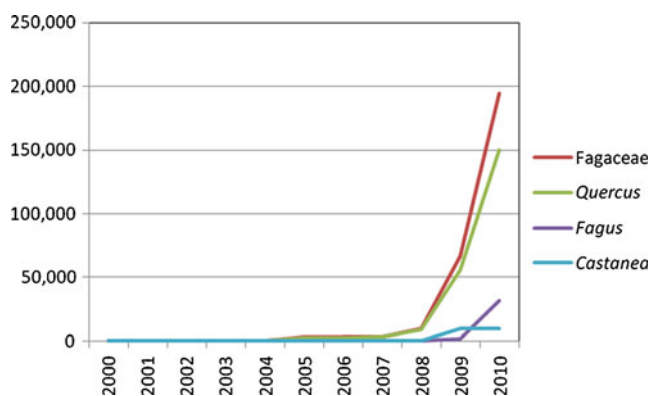

Fig. 8 Number of ESTs in the dbEST database. The *x*-axis indicates the registration date of sequences in the database, while the *y*-axis indicates the accumulated number of sequences for each category (Fagaceae, *Quercus*, *Fagus*, and *Castanea*)

Table 6 Fagaceae EST resources in dbEST (sequences collected by the Sanger method) and in SRA (sequences collected by pyrosequencing method)

Genus	Species	Number of sequences	Percentage
dbEST			
Quercus	Q. robur	81,646	42.0
	Q. petraea	58,173	29.9
	Q. suber	6,690	3.4
	Q. crispula	3,385	1.7
Fagus	F. sylvatica	31,309	16.1
	F. grandifolia	10	0.0
	F. crenata	87	0.0
Castanea	C. mollissima	9,480	4.9
	C. sativa	101	0.1
	C. dentata	91	0.0
Castanopsis	C. sieboldii	3,354	1.7
Total		194,326	100.0
SRA			
Quercus	Q. petraea	862	38.7
	Q. robur	466	20.9
	Q. suber	416	18.7
	Q. rubra	72	3.2
	Q. alba	52	2.3
Fagus	F. grandifolia	17	0.8
	C. mollissima	183	8.2
	C. dentate	161	7.2
Total		2,227	100.0

The data were downloaded 31 March 2011

Assembler Program. The oak gene index release 2.0 is now available, which resulted in 42,144 unique sequences (19,674 TC and 22,470 singleton ESTs) from 148,876 input sequences. Because of the unified methodology, the gene index has special value in comparative analysis, for example, the analysis of motif frequency of genic microsatellite sequences. When the distribution of SSR motifs were visualized by self-organizing map with nine plant species including oak, oak located near poplar (OakContigV1 close to PPLGI (*Populus* gene index)), which may reflect the phylogenetic position of oaks in the Eurosid I clade (Ueno et al. 2010).

Full genome sequencing in the Fagaceae

The value that a reference genome provides in accelerating research in a taxonomic group such as the Fagaceae is difficult to overemphasize. Reference genome sequences are the starting point and our current scientific currency for comparative genomics, translational genomics, functional genomics, evolutionary genomics, ecological genomics, molecular genetics, proteomics, and systems biology research. Great steps forward in genomic resources for the Fagaceae have been achieved through the development of expressed gene sequence databases, genetic maps, DNA libraries, and gene ecology studies. For research in the Fagaceae to continue to grow and flourish, reference genomes are now required. Reference genomes will propel research on the Fagaceae to the next level. Given the past investment in genetics and genomics research in the Fagaceae, the decreasing costs of high-throughput DNA sequencing, the depth of experience that exists for sequencing plant genomes, and all of the major challenges in forest ecosystem sustainability, now is the time to proceed with genome sequencing projects in the Fagaceae. There are currently two genome sequencing projects on the way.

The chestnut genome project

A project to create a reference genome for *C. mollissima* (Chinese chestnut) was initiated in 2009 with support from the FHI organization, a collaborative effort of the US Endowment for Forestry and Communities, the USDA Forest Service, and Duke Energy (http://www.foresthealthinitiative.org). The goals of The Chestnut Genome Project are (1) to develop a high-quality reference genome sequence for use by the greater scientific community, (2) to identify all of the genes for resistance to the chestnut blight fungus *C. parasitica*, and (3) to demonstrate the value of genomes in addressing forest health and ecosystem restoration.

The choice of a species and an individual genotype to serve as the reference genome for a genus or larger taxonomic group is an important first step in a genome sequencing project. For *Castanea*, Chinese chestnut (*C. mollissima*) was an obvious choice as species for the reference genome for several reasons. Chinese chestnut carries genes for resistance to two major pathogens responsible for extirpating American chestnut from its native range—the chestnut blight fungus *C. parasitica* and the ink disease fungus *P. cinnamomi*. Chinese chestnut is the source of resistance to these pathogens which is being used in the long-standing, multigenerational breeding program conducted by TACF (http://www.acf.org). The reference genome will provide DNA markers within the blight resistance genes to accelerate MAS. Also, opportunities exist for comparative and functional genomics within *Castanea* and for transfer of information from the *C. mollissima* genome to *C. dentata* (American chestnut) and related species. *Agrobacterium* and ballistics transformation systems are well-established for *C. dentata*, as are plant regeneration from tissue culture by organogenesis and somatic embryogenesis. This enables functional studies with candidate genes from the *C. mollissima* genome in a closely related, target species. *C. mollissima* and *C. dentata* can be induced to flower relatively quickly, enabling heredity studies. Furthermore, genome structure is relatively conserved in the Fagaceae as shown

by synteny among linkage groups of *Q. robur* and *C. sativa* (Casasoli et al. 2006) which indicates that comparative genomics across the Fagaceae family will be greatly enhanced by the chestnut reference genome. Finally, the estimated 1 C genome size of *C. mollissima* of 794 Mb is representative of Fagaceae genome sizes for which the mean size is ~750±90 Mb.

The cultivar Vanuxem was chosen for the reference genome due to the key role it has played in the Genomic Tools For The Fagaceae Project (http://www.fagaceae.org) sponsored by the NSF's Plant Genome Research Program. The Chestnut Genome Project builds upon the results of the Fagaceae Genomic Tools Project, in which a physical map, high-density genetic linkage maps, and EST databases were constructed for Chinese chestnut. The Fagaceae Genomic Tools Project provided 874 full-length coding sequences which are being used for training gene-finding algorithms in the Chestnut Genome Project, as well as over 40,000 Chinese chestnut unigenes for validating and mapping gene models. The integrated physical and genetic maps from the Fagaceae Genomic Tools Project provides over 740 genetic markers and 589 physical map contigs, representing ~857 Mb of sequence coverage, which is an invaluable resource for assembling the Chinese chestnut sequence into scaffolds and pseudochromosomes.

The Chestnut Genome Project is a collaboration that brings together expertise and facilities in genomics and bioinformatics from Pennsylvania State University and Clemson University and cooperators. The approach taken to develop the reference genome for *C. mollissima* relies heavily on next-generation sequencing technology. The project starts with 454/Roche shotgun sequencing of Vanuxem genomic DNA to a minimum depth of 20× coverage. Assemblies of the shotgun sequence are conducted at serial coverage (at 1×, 10×, 15×, 18×, etc.) to determine the actual depth of sequence required for and optimal assembly of the genome. The trial assemblies will inform the extent of paired-end or Sanger sequence required to supplement the shotgun sequence data for the assembly of the de novo genome sequence contigs into scaffolds. Larger scaffolds and finally pseudochromosomes can then be constructed by reference to loci order or BAC clone order in the *C. mollissima* genetic linkage maps and physical map. Gene models will be predicted using a combination of gene-finding algorithms and transcript mapping. Illumina genomic sequence data will be generated and used to correct gene models and to close sequence gaps. Finally, making use of the reference genome, the "resequencing" of additional chestnut genomes by Illumina technology is planned. The additional genome sequences will aid in the identification of SNPs and species-specific DNA markers and in the discovery of genes important in disease resistance through comparisons of the content of QTL regions across species.

At the time of this review, sequence coverage of the *C. mollissima* genome is at approximately 22× depth by 454/Roche sequencing technology and 47× depth by Illumina technology. The 454/Roche sequence includes 4× depth of paired-end sequences. Sanger sequences have also been produced from the ends of 43,143 BAC clones representing the minimum tiling path of the *C. mollissima* genome. De novo assemblies of the genome sequence data derived from the next-generation and Sanger sequencing technologies have incorporated over 925 Mb into contigs and approximately 587 Mb into 51,766 scaffolds, with and N50 length of 51,759 bp to date. The size of de novo assembly has not yet reached a plateau, suggesting that the high level of heterozygosity in the nondomesticated chestnut genome or other factors are contributing to a much larger genome assembly size in practice, than would be predicted based on the absolute 1 C value. To complete the genome assembly, both additional Roche 454 platform and Sanger sequence data is being produced.

However, even with the current build, over 97 % of the Chinese chestnut unigene transcripts (46,954) from the NSF Fagaceae Genomic Tools Project could be aligned to the genome sequence assembly. The Chinese chestnut transcripts aligned to the genome sequence at over 98 % identity, including many full length genes. The transcript mapping results indicate that the genome sequence data is of high quality and that the gene space is assembling properly. Gene ontology functional annotation of predicted genes from the current assembly cover all of the molecular, cellular, and biological categories (Fig. 9), including over 1,100 stress response genes. The mean coding sequence length of predicted genes is 2,761 bp, with a maximum gene coding sequence length of 43,203 bp which contained 71 properly assembled exons (annotated putatively as an AAA ATPase, von Willebrand factor type A domain-containing protein gene). The mean gene density detected in the current assembly (build 12) was 13 per scaffold, with a maximum of 58 genes in 1 scaffold.

A publicly accessible web portal for the genome sequence data is being created at the FGW site. The public web portal (http://www.fagaceae.org/FHI) will provide searchable databases for access to the genome sequences along with an interactive browser for viewing the assembled genome sequence, as soon as the assembly and annotation is finalized. A website for the Chestnut Genome Project has also been initiated at the NCBI to house the data and facilitate access to the final results (http://www.ncbi.nlm.nih.gov/genomeprj?term=chestnut).

The oak genome project

A second project "Genoak" to create a reference genome for *Q. robur* (pedunculate oak, $2n=2x=24$) has just started with
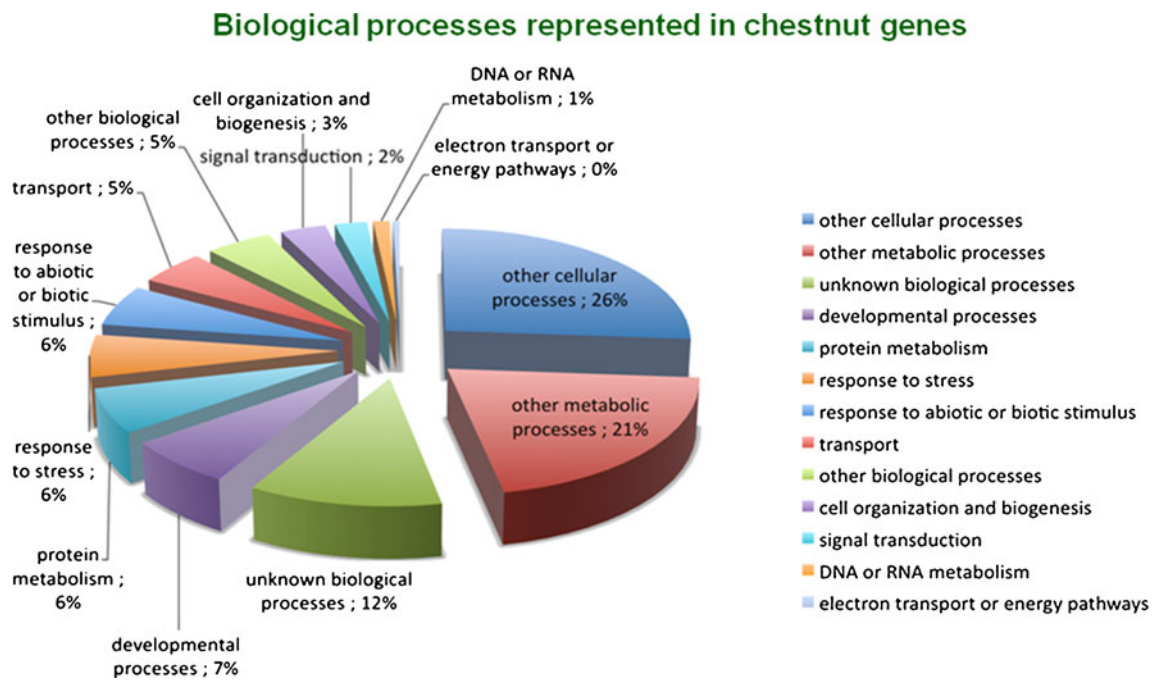
Fig. 9 Biological processes represented in chestnut genes

support from the French National Research Agency (ANR). The project is a collaborative effort between INRA and Genoscope (http://www.genoscope.cns.fr/spip/spip.php?lang=en). In France, oaks are by far the most important forest tree species. Pedunculate oak (*Q. robur*) and the closely related species sessile oak (*Q. petraea*) represent one third of the total forested area in France, i.e., approximately 5,000,000 ha. Oaks have a high socioeconomic value, providing various wood and biomass resources, barrels for wine, firewood, and timber for construction and furniture. The oak forest also provides important ecological services such as soil and water protection and maintenance of terrestrial biodiversity.

The objectives of the "Genoak" project are twofold: first, provide a reference genome sequence for pedunculate oak (*Q. robur* ~740 Mb/C; Kremer et al. 2007) and, second, use this information to discover ecologically relevant genetic variation occurring in genes of adaptive significance in contrasting environments. For the first objective, we will use a combination of new generation sequencing approaches. For the second objective, several sessile and pedunculate oak genotypes will be resequenced using the Illumina sequencing-by-synthesis technology. Then, by comparing their pattern of nucleotide diversity, we should reveal hot spots of adaptive divergence. More specifically, we aim at identifying those loci that have been important during incipient sympatric speciation as a by-product of ecological adaptation. Additionally, the project will offer public access of genomic resources/data/results via a user-friendly web-based interface.

The first steps forward in genomic resources development have been already achieved with (1) the development of an expressed gene sequence database (Ueno et al. 2010) continuously updated, (2) the construction of two BAC libraries (12× and 14× coverage; Faivre-Rampant et al. 2011) available at the CNRGV repository center (http://cnrgv.toulouse.inra.fr/), and (3) the establishment of a second-generation linkage map based on 400 EST-SSRs (Durand et al. 2010; Bodénès et al., submitted) that is presently being upgraded into an ultradense gene-based map using a 9 K Infinium genotyping array (Illumina Inc., San Diego, CA, USA). This high-density linkage map will enable the reconstruction of chromosomes as long series of physical scaffolds. Finally, insights into the composition of the oak genome have also been achieved through the sequencing of BAC ends (Faivre-Rampant et al. 2011) and clones from hypomethylated and genomic libraries (Lesur et al. 2011) and will be refined by the sequencing of a series of 20 BACs. Oak genomic resources have and will be made publicly accessible through the *Quercus* web portal: https://w3.pierroton.inra.fr/QuercusPortal/index.php.

Conclusion

The intense research activities recently implemented in European oaks and Chinese chestnuts will soon be witnessed by the full genome sequencing of *C. mollissima* and *Q. robur*. Large-scale genomic resources have already been developed and made available thanks to collaborative

projects. Interestingly, the comparative genomic analysis (macrosynteny between *Quercus* and *Castanea*) led to the same conclusions as phylogenetic investigations that the oaks are more closely related to chestnuts than to the beeches. These results demonstrate the need to reinforce the comparative genomics across the different genera of the Fagaceae. Gene sequences available in current libraries could easily constitute resources for constructing genetic markers and genomic tools for related genera such as *Notholithocarpus*, *Castanopsis*, and *Lithocarpus* that are rather close to oak and chestnut (Fig. 2). Conversely, pangenomic analysis of genus divergence may also improve the phylogenetic resolution within Fagaceae. Besides macroevolutionary issues, genomics may further contribute to solve a number of pending and compelling questions:

- *Taxonomic resolution*. Species identification is still an open area of research not only in oaks but also in *Castanopsis* and *Lithocarpus*, genera that each comprise more than 100 species (Cannon et al. 2010).
- *Hybridization and speciation*. Some of the Fagaceae genera exhibit frequent species hybridization. Full genome sequencing together with very dense genetic maps will allow for genome scanning to determine levels of species differentiation and the location of hot spots of speciation (Scotti-Saintagne et al. 2004a).
- *Adaptation to environmental changes*. Identifying and cataloging genes of adaptive significance remains an important stepping stone for understanding the evolutionary responses of tree population to selection pressures (Derory et al. 2006).
- *Community genomics*. Fagaceae species are dominant species of most temperate forest ecosystems and are drivers of terrestrial biodiversity. Interactions between trees and their associated species (mammals, fungi, and pathogens) can be tackled via gene–gene (or gene–species) interactions once candidate genes are known (Hersch-Green et al. 2011).
- *Breeding and genetic improvement*. Despite their long generation, breeding for pest resistance is an important goal in chestnuts and oaks and presumably in other genera as well. Rapid progress can be expected when nucleotide diversity is accessible at a pangenomic scale and can be used to improve the prediction of the breeding value of candidate trees for selection (Grattapaglia and Resende 2011).

References

Abril N, Gion JM, Kerner R, Muller-Starck G, Cerrillo RM, Plomion C, Renaut J, Valledor L, Jorrin-Novo JV (2011) Proteomics research on forest trees, the most recalcitrant and orphan plant species. Phytochemistry 72:1219–1242

Abrouk M, Murat F, Pont C, Messing J, Jackson S, Faraut T, Eric Tannier E, Plomion C, Cooke R, Feuillet C, Salse J (2010) Palaeogenomics of plants: synteny-based modelling of extinct ancestors. Trends Plant Sci 15:479–487

Ahuja MR (2011) Fate of transgenes in the forest tree genome. Tree Genet Genome 7:221–230

Anagnostakis SL (1987) Chestnut blight: the classical problem of an introduced pathogen. Mycologia 79:23–37

Anagnostakis S (2007) The chestnut plantation at Sleeping Giant: the legacy of Arthur Harmount Graves. J Amer Chestnut Found 21:34–39

Andrade GM, Merkle SA (2005) Enhancement of American chestnut somatic seedling production. Plant Cell Rep 24:326–334

Andrade GM, Nairn CJ, Le HT, Merkle SA (2009) Sexually mature transgenic American chestnut trees via embryogenic suspension-based transformation. Plant Cell Rep 28:1385–1397

Axelrod DI (1983) Biogeography of oak in the Arcto-Tertiary province. Ann Mo Bot Gard 70:629–657

Bacilieri R, Ducousso A, Petit RJ, Kremer A (1996) Mating system and asymmetric hybridization in a mixed stand of European oaks. Evolution 50:900–908

Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, 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 Biol 9:51

Barreneche T, Bodénès C, Lexer C, Trontin JF, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A (1998) A genetic linkage map of *Quercus robur* L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. Theor Appl Genet 97:1090–1103

Barreneche T, Casasoli M, Russel K, Akkak A, Meddour H, Plomion C, Villani F, Kremer A (2004) Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). Theor Appl Genet 108:558–566

Bellarosa R, Simeone MC, Papini A, Schirone B (2005) Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp. Mol Phylogenet Evol 34:355–370

Betz G, Knappe C, Lapierre C, Olbrich M, Welzl G, Langebartels C, Heller W, Sandermann H, Ernst D (2009) Ozone affects shikimate pathway transcripts and monomeric lignin composition in European beech (*Fagus sylvatica* L.). Eur J Forest Res 128:109–116

Brendel O, Le Thiec D, Saintagne C, Bodénès C, Kremer A, Guehl JM (2008) Detection of quantitative trait loci controlling water use efficiency and related traits in *Quercus robur* L. Tree Genet Genome 4:263–278

Brookes G (2008) The impact of using GM insect resistant maize in Europe since 1998. Int J Biotechnol 10:148–166

Burnham CR, Rutter PA, French DW (1986) Breeding blight-resistant chestnuts. Plant Breed Rev 4:347–397

Burnworth E (2002) A brief history of Stronghold, Inc. to restore the American chestnut. Stronghold, Inc, Sugarloaf Mountain, p 39

Camus A (1936-1954) Les chênes. Monographie du genre Quercus (et Lithocarpus). Encyclopédie économique de sylviculture, Vol. 6–8. Académie des Sciences, Paris

Cannon CH, Manos PS (2000) The Bornean Lithocarpus Bl. section Synaedrys (Lindley) Barnett (Fagaceae): discussion of its circumscription and description of a new species. Bot J Linn Soc 133:343–357

Cannon CH, Manos PS (2001) Combining and comparing morphometric shape descriptors with a molecular phylogeny: the case of fruit type evolution in Bornean Lithocarpus (Fagaceae). Syst Biol 50:1–21

Cannon CH, Kua CS, Zhang D, Harting JR (2010) Assembly free comparative genomics of short-read sequence data discovers the needles in the haystack. Mol Ecol 19(suppl 1):147–161

Casasoli M, Derory J, Morera-Dutrey C, Brendel O, Porth I, Guehl JM, Villani F, Kremer A (2006) Comparison of QTLs for adaptive traits between oak and chestnut based on an EST consensus map. Genetics 172:533–546

Chen G, Sun W-B (2010) Ploidy variation in Trigonobalanus verticillata (Fagaceae). Plant Syst Evol 284:123–127

Cieslar A (1923) Untersunchungen über die wirschaftliche Bedeutung der Herkunft des Saatgutes der Stieleiche. Centralbl Ges Forstwes 49:97–149

Clapper RB (1954) Chestnut breeding, techniques and results. J Hered 45:106–114

Comps B, Barrière G, Merzeau D, Letouzey J (1987) Alloenzyme variability of beeches in the Atlantic Range of western Europe. Can J Forest Res 9:1043–1049

Connors BJ, Maynard CA, Powell WA (2001) Expressed sequence tags from stem tissue of the American chestnut, Castanea dentata. Biotechnol Lett 23:1407–1411

Crandall BS, Gravatt GF, Ryan MM (1945) Root disease of Castanea species and some coniferous and broadleaf nursery stocks, caused by Phytophthora cinnamomi. Phytopathology 35:162–180

Crepet WL, Daghlian CP (1980) Castaneoid inflorescences from the Middle Eocene of Tennessee and the diagnostic value of pollen (at the subfamily level) in Fagaceae. Am J Bot 67:739–757

Crepet WL, Nixon KC (1989a) Earliest megafossil evidence of Fagaceae: phylogenetic and biogeographic implications. Am J Bot 76:842–855

Crepet WL, Nixon KC (1989b) Extinct transitional Fagaceae from the Oligocene and their phylogenetic implications. Am J Bot 76:1493–1505

Deng M (2007) Anatomy, taxonomy, distribution, and phylogeny of Quercus subgenus Cyclobalanopsis (Oersted) Schneid. (Fagaceae). Ph.D. dissertation, Graduate School of Chinese Academy of Sciences

Denk T, Grimm GW (2009) The biogeographic history of beech trees. Rev Palaeobot Palynol 158:83–100

Denk T, Grimm GW (2010) The oaks of western Eurasia: traditional classifications and evidence from two nuclear markers. Taxon 59:351–366

Derory J, Léger P, Garcia V, Schaeffer J, Hauser MT, Salin F, Luschnig C, Plomion C, Glössl J, Kremer A (2006) Transcriptome analysis of bud burst in sessile oak (Quercus petraea L.). New Phytol 170:723–738

Derory J, Scotti-Saintagne C, Bertocchi E, Le Dantec L, Graignic N, Jauffres A, Casasoli M, Chancerel E, Bodénès C, Alberto F, Kremer A (2010) Contrasting correlations between diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. Heredity 104:438–448

Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak Quercus macrocarpa. Mol Ecol 5:615–627

Durand J (2009) Contribution à la cartographie génétique chez les Fagacées. Ph.D. thesis. University Bordeaux I, 207p

Durand J, Bodénès C, Chancerel E, Frigerio JM, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn HP, Villani F, Mattioni C, Cherubini M, Goicoechea PG, Herrán A, Ikaran Z, Cabané C, Ueno S, Alberto F, Dumoulin P-Y, Guichoux E, de Daruvar A, Kremer A, Plomion C (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11:570

Dvornyk V, Sirvio A, Mikkonen M, Savolainen O (2002) Low nucleotide diversity at the pal1 locus in the widely distributed Pinus sylvestris. Mol Biol Evol 19:179–188

Echevarria-Zomeno S, Ariza D, Jorge I, Lenz C, Del Campo A, Jorrin JV, Navarro RM (2009) Changes in the protein profile of Quercus ilex leaves in response to drought stress and recovery. J Plant Physiol 166:233–245

Faivre-Rampant P, Lesur I, Boussardon C, Bitton F, Bodénès C, Le Provost G, Bergès H, Fluch S, Kremer A, Plomion C (2011) Analysis of BAC end sequences in oak, providing insights into the composition of the genome of this keystone species. BMC Genomics 12:292

Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Forman LL (1966) On the evolution of cupules in the Fagaceae. Kew Bulletin 18:385–419

Frettinger P, Herrmann S, Lapeyrie F, Oelmuller R, Buscot F (2006) Differential expression of two class III chitinases in two types of roots of Quercus robur during pre-mycorrhizal interactions with Piloderma croceum. Mycorrhiza 16:219–223

Frettinger P, Derory J, Herrmann S, Plomion C, Lapeyrie F, Oelmuller R, Martin F, Buscot F (2007) Transcriptional changes in two types of pre-mycorrhizal roots and in ectomycorrhizas of oak micro-cuttings inoculated with Piloderma croceum. Planta 225:331–340

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Gailing O, Kremer A, Steiner W, Hattemer HH, Finkeldey R (2005) Results on quantitative trait loci for flushing date in oaks can be transferred to different segregating families. Plant Biol 7:516–525

Gailing O, Langenfeld-Hyeser R, Polle A, Finkeldey R (2008) Quantitative trait loci affecting stomatal density and growth in a Quercus robur progeny: implications for the adaptation to changing environments. Glob Chang Biol 14:1934–1946

Gailing O, Vornam B, Leinemann L, Finkeldey R (2009) Genetic and genomic approaches to assess adaptive genetic variation in plants: forest trees as a model. Physiol Plant 137:509–519

Galvan JV, Valledor L, Cerrillo RM, Pelegrin EG, Jorrin-Novo JV (2011) Studies of variability in Holm oak (Quercus ilex subsp. ballota [Desf.] Samp.) through acorn protein profile analysis. J Proteomics 74:1244–1255

Gomez A, Lopez JA, Pintos B, Camafeita E, Bueno MA (2009) Proteomic analysis from haploid and diploid embryos of Quercus suber L. identifies qualitative and quantitative differential expression patterns. Proteomics 9:4355–4367

Grattapaglia D, Resende M (2011) Genomic selection in forest tree breeding. Tree Genet Genome 7:241–255

Graziosi I, Santi F (2008) Chestnut gall wasp (Dryocosmus kuriphilus): spreading in Italy and new records in Bologna province. B Insectol 61:343–348

Griffin GJ (2000) Blight control and restoration of the American chestnut. J Forest 98:22–27

Grivet D, Heinze B, Vendramin GG, Petit RJ (2001) Genome walking with consensus primers: application to the large single copy region of chloroplast DNA. Mol Ecol Notes 1:345–349

Hebard FV (2006) The backcross breeding program of The American Chestnut Foundation. In: Steiner KC, Carlson JE (eds)

Restoration of American chestnut to forest lands. Proceedings of a Conference and Workshop, May 4–6, 2004. The North Carolina Arboretum Natural Resources Report NPS/NCR/CUE/NRR 2006/001 National Park Service, Washington, DC, pp 61–78

Heiniger U, Rigling D (1994) Biological control of chestnut blight in Europe. Annu Rev Phytopathol 32:581–599

Hepting GH (1974) Death of the American chestnut. J Forest Hist 18:60–67

Hersch-Green E, Turley NE, Johnson MTJ (2011) Community genetics: what have we accomplished and where should we be going? Phil trans roy soc B Biol sci 366:1453–1460

Hjelmquist H (1948) Studies on the floral morphology and phylogeny of the Amentiferae. Bot Not 2(suppl):1–171

Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. Genetics 180:329–340

Islam-Faridi N, Nelson CD, Sisco PH, Kubisiak TL, Hebard FV, Paris RL, Phillips RL (2009) Cytogenetic analysis of American chestnut (*Castanea dentata*) using fluorescent in situ hybridization. Acta Hort (ISHS) 844:207–210

Jackson MB, Campbell DJ (1976) Waterlogging and petiole epinasty in tomato: the role of ethylene and low oxygen. New Phytol 76:21–29

Jeffers SN, James JB, Sisco PH (2009) Screening for resistance to *Phytophthora cinnamomi* in hybrid seedlings of American chestnut. In: Phytophthoras in forests and natural ecosystems. USDA-FS General Technical Report P8W-GTR-221, pp 188–192

Jorge I, Navarro-Cerrillo RM, Lenz C, Ariza D, Porras C, Jorrin J (2005) The holm oak leaf proteome: analytical and biological variability in the protein expression level assessed by 2-DE and protein identification tandem mass spectrometry, de novo sequencing and sequence similarity searching. Proteomics 5:222–234

Jorge I, Navarro RM, Lenz C, Ariza D, Jorrin J (2006) Variation in the holm oak leaf proteome at different plant developmental stages, between provenances and in response to drought stress. Proteomics 6(Suppl 1):S207–S214

Kaul RB (1987) Reproductive structure of *Lithocarpus* sensu lato (Fagaceae): cymules and fruits. J Arn Arb 68:73–104

Kerner R, Winkler JB, Dupuy JW, Gensen MJ, Lindermayr C, Ernst D, Mueller-Starck G (2011) Changes in the proteome of juvenile European beech following three years exposure to free-air elevated ozone. Forest - Biogeosciences and Forestry 4:69–76

Kleinschmit J (1993) Intraspecific variation of growth and adaptive traits in European oak species. Ann Sci For 50:166s–185s

Koch JL, Carey DW, Mason ME, Nelson CD (2010) Assessment of beech scale resistance in full- and half-sibling American beech families. Can J Forest Res 40:265–272

Koch JL, Carey DW, Mason ME, Nelson CD, Barakat A, Carlson JE, Neale D (2011) Development of molecular tools for use in beech bark disease management. In: McManus, Katherine A; Gottschalk, Kurt W., eds. 2010. Proceedings of the 21st US Department of Agriculture Interagency Research Forum on Invasive Species 2010; 2010 January 12–15; Annapolis, MD. Gen. Tech. Rep. NRS-P-75. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station, pp 38–40

Kremer A, Petit RJ (1993) Gene diversity in natural populations of oak species. Ann Sci For 50:186s–203s

Kremer A, Savill PS, Steiner KC (1993) Genetics of oaks. Ann Sci For 50(Suppl 1):1s–472s

Kremer A, Xu LA, Ducousso A (2004) Oak genetics. In: Burley J, Evans J, Youngquist J (eds) Encyclopedia of forest sciences. Elsevier, Amsterdam, pp 1501–1507

Kremer A, Casasoli M, Barreneche T, Bodénès C, Sisco P, Kubisiak T, Scalfi M, Leonardi S, Bakker EG, 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. In: Kole CR (ed) Genome mapping and molecular breeding in plants, vol 7, forest trees. Springer, Heidelberg, pp 165–187

Kremer A, Sederoff R, Wheeler N (2010) Genomics of forest and ecosystem health in the Fagaceae: meeting report. Tree Genet Genome 6:815–820

Kruger A, Peskan-Berghofer T, Frettinger P, Herrmann S, Buscot F, Oelmuller R (2004) Identification of premycorrhiza-related plant genes in the association between *Quercus robur* and *Piloderma croceum*. New Phytol 163:149–157

Kubisiak TL, Roberds JH (2006) Genetic structure of American chestnut populations based upon neutral DNA markers. In: Steiner KC, Carlson JE (eds) Restoration of American chestnut to forest lands. Natural Resources Report 2006/1. National Park Service, USA, pp 109–122

Kubisiak TL, Hebard FV, Nelson CD, Zhang J, Bernatzky R, Huang H, Anagnostakis SL, Doudrick RL (1997) Molecular mapping of resistance to blight in an interspecific cross in the genus *Castanea*. Phytopathology 87:751–759

Kvacek Z, Walther H (1989) Paleobotanical studies in Fagaceae of the European Tertiary. Plant Syst Evol 162:213–229

Lang P, Dane F, Kubisiak TL, Huang H (2007) Molecular evidence for an Asian origin and a unique westward migration of species in the genus *Castanea* via Europe to North America. Mol Phylogenet Evol 43:29–59

Leffel RC (2004a) Strategies for breeding blight-resistant, timber-type chestnuts. 2004 IUFRO Forest Genetics Meeting Proceedings, pp 273–284

Leffel RC (2004b) Cytoplasmic male sterility and chestnut breeding programs. J Amer Chestnut Found 18:54–59

Lepoittevin C, Frigerio JM, Garnier-Géré P, Salin F, Cervera M-T, Vornam B, Harvengt L, Plomion C (2010) In vitro vs in silico detected SNPs for the development of a genotyping array: what can we learn from a non-model species? PLoS One 5(6):e11034

Lesur I, Durand J, Sebastiani F, Gyllenstrand N, Bodénès C, Lascoux M, Kremer K, Vendramin GG, Plomion C (2011) A sample view of the pedunculate oak (*Quercus robur*) genome from the sequencing of hypomethylated and random genomic libraries. Tree Genet Genome. doi:10.1007/s11295-011-0412-4

Logan WB (2005) Oak: the frame of civilization. WW Norton & Company, New York

MacDonald WL, Double ML (2006) Hypovirulence: use and limitations as a chestnut blight biological control. In: Steiner KC, Carlson JE (eds) Restoration of American chestnut to forest lands. Proceedings of a Conference and Workshop, May 4–6, 2004. The North Carolina Arboretum Natural Resources Report NPS/NCR/CUE/NRR 2006/001 National Park Service, Washington, DC, pp 87–96

Mai DH (1970) Die tertiaren Arten von *Trigonobalanus* Forman (Fagaceae) in Europa. Jahrb Geol (Berlin) 3:381–409

Manchester SR (1987) The fossil history of the Juglandaceae. Mo Bot Gard Monogr 21:1–137

Manchester SR (1994) Fruits and seeds of the middle Eocene Nut Beds Flora, Clarno Formation, Oregon. Palaeontogr Am 58:1–205

Manos PS, Stanford AM (2001) The biogeography of Fagaceae: tracking the Tertiary history of temperate and subtropical forests of the Northern Hemisphere. Int J Plant Sci 162:S77–S93

Manos PS, Doyle JJ, Nixon KC (1999) Phylogeny biogeography and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). Mol Phylogenet Evol 12:333–349

Manos PS, Zhou Z-K, Cannon CH (2001) Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. Int J Plant Sci 162:1361–1379

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 55:181–190

Maynard CA, Powell WA, Polin-McGuigan LD, Viéitez AM, Ballester A, Corredoira E, Merkle SA, Andrade GM (2008) Chestnut. In: Kole C, Hall TC (eds) A compendium of transgenic crop plants: forest tree species. Blackwell Publishing, Oxford, pp 169–192

Merkle SA, Andrade GM, Nairn CJ, Powell WA, Maynard CA (2007) Restoration of threatened species: a noble cause for transgenic trees. Tree Genet Genome 3:111–118

Mi S, Figueiredo A, Acioli B, Sousa L, Pessoa F, Bald A, Pais MS (2009) Identification of plant genes involved on the initial contact between ectomycorrhizal symbionts (*Castanea sativa*—European chestnut and *Pisolithus tinctorius*). Eur J Soil Biol 45:275–282

Nixon KC, Crepet WL (1989) *Trigonobalanus* (Fagaceae): taxonomic status and phylogenetic relationships. Am J Bot 6:828–841

Nunn A, Anegg S, Betz G, Simons S, Kalisch G, Seidlitz H, Grams T, Haberle K, Matyssek R, Bahnweg G, Sandermann H, Langebartels C (2005) Role of ethylene in the regulation of cell death and leaf loss in ozone-exposed European beech. Plant Cell Environ 28:886–897

Nuss DL (2005) Hypovirulence: mycoviruses at the fungal–plant interface. Nat Rev Microbiol 3:632–642

Oh S-H, Manos PS (2008) Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear *CRABS CLAW* sequences. Taxon 57:434–451

Olbrich M, Betz G, Gerstner E, Langebartels C, Sandermann H, Ernst D (2005) Transcriptome analysis of ozone-responsive genes in leaves of European beech (*Fagus sylvatica* L.). Plant Biol (Stuttg) 7:670–676

Olbrich M, Knappe C, Wenig M, Gerstner E, Haberle KH, Kitao M, Matyssek R, Stich S, Leuchner M, Werner H, Schlink K, Muller-Starck G, Welzl G, Scherb H, Ernst D, Heller W, Bahnweg G (2010) Ozone fumigation (twice ambient) reduces leaf infestation following natural and artificial inoculation by the endophytic fungus *Apiognomonia errabunda* of adult European beech trees. Environ Pollut 158:1043–1050

Oliver JB, Mannion CM (2001) Ambrosia beetle (*Coleoptera*: Scolytidae) species attacking chestnut and captured in ethanol-baited traps in Middle Tennessee. Environ Entomol 30:909–918

Olson MS, Robertson AL, Takebayashi N, Silim S, Schroeder WR, Tiffin P (2010) Nucleotide diversity and linkage disequilibrium in balsam poplar (*Populus balsamifera*). New Phytol 186:526–536

Parelle J, Zapater M, Scotti-Saintagne C, Kremer A, Jolivet Y, Dreyer E, Brendel O (2007) Quantitative trait loci of tolerance to waterlogging in a European oak (*Quercus robur* L.): physiological relevance and temporal effect patterns. Plant Cell Environ 30:422–434

Pawlovsky TA (2007) Proteomics of European beech (*Fagus sylvatica* L.) seed dormancy breaking: influence of abscisic and gibberellic acids. Proteomics 7:2246–2257

Payne JA, Menke AS, Schroeder PM (1975) *Dryocosmus kuriphilus* Yasumatsu, (Hymenoptera: *Cynipidae*), an Oriental chestnut gall wasp in North America. USDA Cooperative Economic Insect Report 25:903–905

Porth I, Berenyi M, Scotti-Saintagne C, Barreneche T, Kremer A, Burg K (2005a) Linkage mapping of osmotic stress induced genes of oaks. Tree Genet Genome 1:31–40

Porth I, Koch M, Berenyi M, Burg A, Burg K (2005b) Identification of adaptation-specific differences in mRNA expression of sessile and pedunculate oak based on osmotic-stress-induced genes. Tree Physiol 25:1317–1329

Powell WA, Maynard CA, Boyle B, Seguin A (2006) Fungal and bacterial resistance in transgenic trees. In: Fladung M, Ewald M (eds) Tree transgenics, recent developments. Springer, Berlin, pp 235–252

Quang ND, Ikeda S, Harada K (2008) Nucleotide variation in *Quercus crispula*. Blume Hered 101:166–174

Quang ND, Ikeda S, Harada K (2009) Patterns of nucleotide diversity at the methionine synthase locus in fragmented and continuous populations of a wind pollinated tree, *Quercus mongolica* var. *crispula*. J Hered 100:762–770

Rashid B, Saleem Z, Husnain T, Riazuddin S (2008) Transformation and inheritance of Bt genes in *Gossypium hirsutum*. J Plant Biol 51:248–254

Ricardo CP, Martins I, Francisco R, Sergeant K, Pinheiro C, Campos A, Renaut J, Fevereiro P (2011) Proteins associated with cork formation in *Quercus suber* L. stem tissues. J Proteomics 74:1266–1278

Rieske LK (2007) Success of an exotic gallmaker, *Dryocosmus kuriphilus*, on chestnut in the USA: a historical account. OEPP/EPPO Bulletin 37:172–174

Robin C, Heiniger U (2001) Chestnut blight in Europe: diversity of *Cryphonectria parasitica*, hypovirulence and biocontrol. For Snow Landscape Res 76:361–367

Roussel M, Le Thiec D, Montpied P, Guehl JM, Brendel O (2009a) Diversity of water use efficiency in a *Quercus robur* family: contribution of related leaf traits. Ann For Sci 66:408–417

Roussel M, Dreyer E, Montpied P, Le Provost G, Guehl JM, Brendel O (2009b) The diversity of 13 C isotope discrimination in a *Quercus robur* full-sib family is associated with differences in intrinsic water use efficiency, transpiration efficiency and stomatal conductance. J Exp Bot 60:2419–2431

Saintagne C, Bodénès C, Barreneche T, Pot D, Plomion C, Kremer A (2004) Distribution of genomic regions differentiating oak species assessed by QTL detection. Heredity 92:20–30

Scalfi M, Troggio M, Piovani P, Leonardi S, Magnaschi G, Vendramin GG, Menozzi P (2004) A RAPD, AFLP, and SSR linkage map, and QTL analysis in European beech (*Fagus sylvatica* L.). Theor Appl Genet 108:433–441

Schlarbaum SE, Maxedon JS, Clark SL (2008) Oak research and technology transfer activities of the University of Tennessee's tree improvement program. In: Lockhart BR, Gardiner ES, Dey DC (eds) Tenth workshop on seedling physiology and growth problems in oak plantings; 2007 October 16–17; Jackson, MS. Gen. Tech. Rep. NRS-P-32. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station: 1

Schlink K (2009) Identification and characterization of differentially expressed genes from *Fagus sylvatica* roots after infection with *Phytophthora citricola*. Plant Cell Rep 28:873–882

Schlink K (2010) Down-regulation of defense genes and resource allocation into infected roots as factors for compatibility between *Fagus sylvatica* and *Phytophthora citricola*. Funct Integr Genomics 10 (2):253–264

Schneiderbauer A, Back E, Sandermann H Jr, Ernst D (1995) Ozone induction of extensin mRNA in Scots pine, Norway spruce and European beech. New Phytol 130:225–230

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodénès C, Burg K, Kremer A (2004a) Genome scanning for interspecific differentiation between two closely related oak species (*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.). Genetics 168:1615–1626

Scotti-Saintagne C, Bodénès C, Barreneche T, Bertocchi E, Plomion C, Kremer A (2004b) Detection of quantitative trait loci controlling bud burst and height growth in *Quercus robur* L. Theor Appl Genet 109:1648–1659

Scotti-Saintagne C, Bertocchi E, Barreneche T, Kremer A, Plomion C (2005) Quantitative trait loci mapping for vegetative propagation in pedunculate oak. Ann For Sci 62:369–374

Shelton AM, Zhao J-Z, Roush RT (2002) Economic, ecological, food safety, and social consequences of the employment of Bt transgenic plants. Ann Rev Entomol 47:845–881

Sisco PH (2004) Designing chestnut orchards to control pollen parent without hand pollination: the use of self-incompatibility and cytoplasmic male sterility. J Amer Chestnut Found 18:46–53

Sisco PH (2006a) An update on chestnut DNA projects. Part I: the genes for blight resistance from Chinese chestnut. J Amer Chestnut Found 20:21–29

Sisco PH (2006b) An update on chestnut DNA projects. Part II: other uses of DNA markers in the TACF breeding program. J Amer Chestnut Found 20:35–43

Smith DM (2000) American chestnut: ill-fated monarch of the Eastern hardwood forest. J Forest 98:12–15

Soepadmo E (1968) A revision of the genus *Quercus* L. subgen. *Cyclobalanopsis* (Oersted) Schneider. Malesia Gard Bull Sing 22:355–427

Soler M, Serra O, Molinas M, Huguet G, Fluch S, Figueras M (2007) A genomic approach to suberin biosynthesis and cork differentiation. Plant Physiol 144:419–431

Soler M, Serra O, Molinas M, Garcia-Berthou E, Caritat A, Figueras M (2008) Seasonal variation in transcript abundance in cork tissue analyzed by real time RT-PCR. Tree Physiol 28:743–751

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Ueno S, Taguchi Y, Tsumura Y (2008) Microsatellite markers derived from *Quercus mongolica* var. *crispula* (Fagaceae) inner bark expressed sequence tags. Genes Genet Syst 83:179–187

Ueno S, Aoki K, Tsumura Y (2009a) Generation of expressed sequence tags and development of microsatellite markers for *Castanopsis sieboldii* var. *sieboldii* (Fagaceae). Ann For Sci 66(5):509

Ueno S, Taguchi Y, Tomaru N, Tsumura Y (2009b) Development of EST-SSR markers from an inner bark cDNA library of *Fagus crenata* (Fagaceae). Conserv Genet 10(5):1477–1485

Ueno S, Le Provost G, Leger V, Klopp C, Noirot C, Frigerio JM, Salin F, Salse J, Abrouk M, Murat F, Brendel O, Derory J, Abadie P, Leger P, Cabane C, Barre A, de Daruvar A, Couloux A, Wincker P, Reviron MP, Kremer A, Plomion C (2010) Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. BMC Genomics 11:650

Valcu CM, Schlink K (2006a) Efficient extraction of proteins from woody plant samples for two-dimensional electrophoresis. Proteomics 6:4166–4175

Valcu CM, Schlink K (2006b) Reduction of proteins during sample preparation and two-dimensional gel electrophoresis of woody plant samples. Proteomics 6:1599–1605

Valcu CM, Junqueira M, Shevchenko A, Schlink K (2009) Comparative proteomic analysis of responses to pathogen infection and wounding in *Fagus sylvatica*. J Proteome Res 8:4077–4091

Vannini A, Vettraino AM (2001) Ink disease in chestnuts: impact on the European chestnut. For Snow Landscape Res 76:345–350

Villani F, Pigliucci M, Cherubini M (1994) Evolution of *Castanea sativa* Mill in Turkey and Europe. Genet Res 63:109–116

Vision TJ, Brown DG, Shmoys DB, Durrett RT, Tanksley SD (2000) Selective mapping: a strategy for optimizing the construction of high-density linkage maps. Genetics 155:407–420

Vornam B, Gailing O, Derory J, Plomion C, Kremer A, Finkeldey R (2011) Characterization and natural variation of a dehydrin gene in *Quercus petraea* (Matt.) Liebl. Plant Biol. doi:10.1111/j.1438-8677.2011.00446.x

Wheeler N, Sederoff R (2009) Role of genomics in the potential restoration of the American chestnut. Tree Genet Genome 5:181–187

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS (2005) The effects of artificial selection on the maize genome. Science 308:1310–1314

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol 24:875–888