

Generation of a large-scale genomic resource for functional and comparative genomics in *Liriodendron tulipifera* L.

Haiying Liang · Saravanaraj Ayyampalayam · Norman Wickett · Abdelali Barakat · Yi Xu · Lena Landherr · Paula E. Ralph · Yuannian Jiao · Tao Xu · Scott E. Schlarbaum · Hong Ma · James H. Leebens-Mack · Claude W. dePamphilis

Received: 30 August 2010 / Revised: 22 February 2011 / Accepted: 13 March 2011
© Springer-Verlag 2011

Abstract *Liriodendron tulipifera* L., a member of Magnoliaceae in the order Magnoliales, has been used extensively as a reference species in studies on plant evolution. However, genomic resources for this tree species are limited. We constructed cDNA libraries from ten different types of tissues: premeiotic flower buds, postmeiotic flower buds, open flowers, developing fruit, terminal buds, leaves, cambium, xylem, roots, and seedlings. EST sequences were generated either by 454 GS FLX or Sanger methods. Assembly of almost 2.4 million sequencing reads from all libraries resulted in 137,923 unigenes (132,905 contigs and 4,599 singletons). About 50% of the unigenes had significant matches to publically

available plant protein sequences, representing a wide variety of putative functions. Approximately 30,000 simple sequence repeats were identified. More than 97% of the cell wall formation genes in the Cell Wall Navigator and the MAIZEWALL databases are represented. The cinnamyl alcohol dehydrogenase (*CAD*) homologs identified in the *L. tulipifera* EST dataset showed different expression levels in the ten tissue types included in this study. In particular, the *LtuCAD1* was found to partially recover the stiffness of the floral stems in the *Arabidopsis thaliana* *CAD4* and *CAD5* double mutant plants, of the *LtuCAD1* in lignin biosynthesis. *L. tulipifera* genes have greater sequence similarity to homologs from other woody angiosperm species than to non-woody model plants. This large-scale genomic resource will be instrumental for gene discovery, cDNA microarray production, and marker-assisted breeding in *L. tulipifera*, and strengthen this species' role in comparative studies.

Communicated by R. Sederoff

Electronic supplementary material The online version of this article (doi:10.1007/s11295-011-0386-2) contains supplementary material, which is available to authorized users.

H. Liang (✉) · A. Barakat · Y. Xu · T. Xu
Department of Genetics and Biochemistry,
Clemson University,
100 Jordan Hall,
Clemson, SC 29634, USA
e-mail: hliang@clemson.edu

S. Ayyampalayam · J. H. Leebens-Mack
Department of Plant Biology, Plant Sciences Building,
University of Georgia,
Athens, GA 30602-7271, USA

N. Wickett · L. Landherr · P. E. Ralph · Y. Jiao · H. Ma ·
C. W. dePamphilis (✉)
Department of Biology and Huck Institutes of the Life Sciences,
Pennsylvania State University,
0201 Life Sciences Bldg.,
University Park, PA 16802, USA
e-mail: cwd3@psu.edu

L. Landherr
Department of Horticulture, Pennsylvania State University,
University Park, PA 16802, USA

S. E. Schlarbaum
Department of Forestry, Wildlife & Fisheries, Institute
of Agriculture, The University of Tennessee,
Knoxville, TN 37996-4563, USA

H. Ma
State Key Laboratory of Genetic Engineering, Institute of Plant
Biology, Center for Evolutionary Biology, School of Life
Sciences, Fudan University,
Shanghai 200433, China

H. Ma
Institutes of Biomedical Sciences, Fudan University,
Shanghai 200032, China

Keywords EST database · Xylogenesis · *Liriodendron* · Yellow-poplar · Magnoliaceae

Introduction

Liriodendron tulipifera L., commonly known as yellow-poplar, tulip tree, or tulip-poplar, is one of only two arborescent species in the genus *Liriodendron*. Yellow-poplar gained its name due to the uncanny similarity of its wood structure and density to true poplars (*Populus* species). However, these two species are from distinct evolutionary lineages: yellow-poplar is a member of Magnoliaceae in the order Magnoliales, whereas *Populus* species are in the core eudicot order Malpighiales. Magnoliaceae flowers usually possess stamens and pistils in a spiral pattern, which is distinct from most other angiosperm species with whorled floral organs and thought to be an ancestral trait for flowering plants (Soltis et al. 2004). Magnoliales and three other orders (Laurales, Piperales, and Canellales) comprise the magnoliids, which, along with Amborellales, Nymphaeales, and Illiciales, form a grade of “basal angiosperm” lineages that contain a wide diversity of floral and growth forms (Qiu et al. 2005; Soltis et al. 2005; Jansen et al. 2007). Among basal angiosperms, Magnoliales are the immediate sister to the species-rich clade including monocots and eudicots with ca. 97% of all angiosperm species (Qiu et al. 2005; Soltis et al. 2005; Jansen et al. 2007; Moore et al. 2007). Its special position in the plant phylogeny and “primitive” floral structure make *Liriodendron*, along with representatives of other basal angiosperm lineages, an ideal candidate for comparative studies of the evolution of form and process throughout flowering plant history (Wei and Wu 1993; Hunt 1998; Ronse de Craene et al. 2003; Zahn et al. 2005).

In addition to its important phylogenetic position, *L. tulipifera* has great economic and ecological values. This species is cultivated in many temperate parts of the world for wood production (Hunt 1998) and is one of the recommended species for waste landfill remediation (Kim and Lee 2005). As one of the largest and ornamentally coveted trees in North America, *L. tulipifera* can attain a height of 61 m with a trunk diameter of up to 152 cm. On good sites (site index=23 m) in the southern Appalachian mountains, *L. tulipifera* will grow faster than any associated species (Beck 1990). Compared with other commercially important species, *L. tulipifera* is remarkably free from damage by insects and diseases, does not require intensive stand management to grow well in dense stands, and is resistant to the damaging effect of metals (such as aluminum) (Klugh and Cumming 2003). The wood of *L. tulipifera* is commercially valuable and is a raw material

source for lumber, furniture, musical instruments, wooden wares, pulp, and many other industries (Moody et al. 1993; Hernandez et al. 1997; Williams and Feist 2004). *L. tulipifera* is also valued as a nectar source for honey production, as a source of wildlife food (mast), and as a large shade tree in urban settings. In addition, chemical extracts from *L. tulipifera* wood or leaves have proven useful for a variety of purposes, including anti-tumor effects and antifeeding activity for herbivores (sesquiterpenes) (Moon et al. 2007) and antimicrobial alkaloids (Bae and Byun 1987). Recently, there has been increased interest in conversion of biomass from *L. tulipifera* to biofuels, as evidenced by studies on ethanol production from this species (Xiang et al. 2004; Berlin et al. 2005; Çelen et al. 2008; Hwang et al. 2008; Koo et al. 2008, 2009).

Little genomic research has been conducted on this species, despite the use of *L. tulipifera* as a reference species in studies on plant evolution and its significant economic and ecological value. To date, only one *L. tulipifera* gene, encoding a laccase, has been functionally characterized (LaFayette et al. 1999). Laccases (EC 1.10.3.2) are copper-containing glycoproteins. Several studies have suggested the involvement of laccases in lignin biosynthesis (Ranocha et al. 2002 and references therein). The organization of two *L. tulipifera* chromosome regions (harboring a *GIGANTEA* and a *LEAFY* floral gene, respectively) was recently revealed (Liang et al. 2010, 2011). At present, there exists only one EST database (6,520 unigenes) developed from floral tissues by capillary sequencing (Albert et al. 2005; Liang et al. 2008) and one ca. 5X BAC library with 73,728 large-insert clones (Liang et al. 2007) available for *L. tulipifera*. This lack of genomic research has hindered the efforts to identify genes involved in traits of economic and ecological importance and limited *Liriodendron*'s role in comparative genomic studies. *L. tulipifera* is one of the species in the Magnoliaceae family with the lowest chromosome number ($2n=2x=38$). However, with a haploid genome size of 1,802 Mbp (Liang et al. 2007), sequencing and assembly of the *L. tulipifera* genome would be expensive, given currently available sequencing technologies. Thus, as with most forest tree species, large-scale sequencing and analysis of *L. tulipifera* ESTs remain a fundamental part of genomics research to enable gene discovery and functional investigations.

Here we report the generation and analysis of a deep transcriptome sequence resource for *L. tulipifera*. To maximize our ability to identify genes expressed in different tissues, extensive ESTs from ten different tissue types (premeiotic flower buds, postmeiotic flower buds, open flowers, developing fruit, terminal buds, leaves, cambium, xylem, roots, and seedlings) were isolated and sequenced. The unigenes from the newly built database

were compared to publically available plant protein sequence databases, and Gene Ontology (GO) terms were determined. Genes involved in wood formation were identified based on similarity to genes in available sequence databases. In particular, a *Liriodendron* cinnamyl alcohol dehydrogenase homolog (*LtuCAD1*) was characterized by overexpression in an *Arabidopsis* *CAD4/CAD5* double mutant. This dataset has also been mined for simple sequence repeats (SSRs) and microRNAs (miRNAs). The unigenes generated in this study will facilitate gene discovery and functional studies, support development of cDNA microarrays and assembly of short-read sequences, and thus allow expression profiling experiments to be integrated into investigations of xylem differentiation, reproductive development, insect and disease resistance, etc. in *Liriodendron*. The 29,289 gene-based SSRs identified in the unigene assemblies will enable marker-assisted breeding in the genus *Liriodendron*. The availability of this deep genomic resource will also strengthen the utility of *Liriodendron* in comparative studies of angiosperm evolution. Lastly, it is noteworthy that genomic resources are very limited for other species in the Magnoliaceae family, with a range of only two sequences in genus *Dugandiodendron* and 1,767 sequences in *Magnolia* deposited in GenBank (as February of 2011). Moreover, the majority of these publicly available sequences are from plastid genomes. Thus, the information developed in this study for *L. tulipifera* can serve as a reference in the Magnoliaceae family.

Materials and methods

Tissue source

Postmeiotic flower buds, open flowers, developing fruits, terminal buds, leaves, and cambium and xylem tissues were collected from mature ramets of clone 108 in the University of Tennessee's Tree Improvement Program *L. tulipifera* breeding orchard in Knoxville, TN and quick frozen with liquid nitrogen in the field. Clone 108 was selected from a pure *L. tulipifera* stand in eastern Tennessee in 1965. The ortet was 32 years of age with a height of 94 ft and a diameter (at 4.5 ft height) of 11.1 in. The bole (trunk) straightness of the ortet was rated as excellent and the pruning ability was good. Xylem and cambium tissues were obtained by removing a section of the bark at the height of 1.4 m from actively growing clone 108 ramets in April–June and scraping both exposed surfaces with RNA-free scalpels (Rnase-Zap, Ambion, Austin, TX). Open-pollinated *L. tulipifera* seeds (from ramets of clones 108, 7A, and 84A in the same orchard) were stratified by storing

4 months in the dark at 4°C, mixed with peat moss in 1 gal plastic bags. The seeds were then germinated by scattering them on top of Miracle-Gro® Potting Mix in covered flats (25×52 cm flats, approximately 400 seeds/flat) with a thin layer of potting mix sprinkled on top. Flats were kept under benches for shade at 25°C and ambient seasonal lighting (May and June) and watered as needed. Plastic coverers were used to keep the seeds moist. Young seedlings (emerging from seed coats) through late stage seedlings (with first true leaves emerging) were harvested (entire seedling) by removing seed coat (if needed), quickly rinsing in ddH₂O, blotting dry on toweling, and quick freezing in liquid nitrogen. For roots, young plants were grown in 6 cm square pots in the Penn State University Buckhout greenhouse (ambient light, 25°C) in Sun Gro Metro-Mix® 360 Growing Media or grown in the same growing media in mesh-bottom pots over a water reservoir for soil-free root collection. Fine, hairy roots and root tips were harvested and frozen as above.

RNA isolation, cDNA synthesis, and sequencing

Total RNA was extracted from younger tissues (seedlings, terminal buds, and postmeiotic flower buds) using the RNAqueous®-Midi kit (Ambion, catalog #1911) according to the manufacturer's protocol (http://www.ambion.com/techlib/prot/fm_1911.pdf) with modifications as described in Carlson et al. (2006). Total RNA was extracted from woody or mature tissues (cambium, xylem, roots, open flower, and fruit) using a modified version of the cetyl trimethyl ammonium bromide (CTAB) protocol developed by Chang et al. (1993), except that 2 to 3 g of frozen tissue was ground in a RNase-free, chilled mortar and pestle under liquid nitrogen and suspended in warm (65°C) CTAB buffer (made fresh same day using RNase-free stock solutions). Total RNA samples were DNase treated with amplification grade DNase I (Invitrogen, catalog #18068-015) and recombinant ribonuclease inhibitor, RNase Out (Invitrogen, catalog #10777-019), according to the manufacturer's recommendations. Purified RNA was recovered using the RNeasy Plant Mini kit (Qiagen, catalog #74104) RNA Cleanup protocol (sample concentrations adjusted to <100 µg in 100 µl RNase-free water) and checked on an Agilent 2100 Bioanalyzer (Agilent Technologies). Message RNA was then extracted from total RNA using the Poly(A) Purist™ mRNA Purification Kit (Ambion, catalog #1916) according to the manufacturer's protocol (http://www.ambion.com/techlib/prot/fm_1916.pdf), as described in Liang et al. (2008). mRNA from premeiotic flower buds was from a previous preparation for the floral cDNA library (*Ltu01*) (Liang et al. 2008) with an additional DNase treatment. The quality of the mRNA was determined using

an Agilent 2100 Bioanalyzer (Agilent Technologies) using the RNA 6000 nano chip and the mRNA Plant assay to ensure that the mRNA samples had no detectable DNA contamination and had less than 15% tRNA contamination.

cDNA was generated from mRNA samples by following the Joint Genome Institute (JGI) cDNA library creation protocol (version 1.0) (<http://my.jgi.doe.gov/general/index.html>) with modifications. An additional chloroform clean-up step was added after the phenol/chloroform/isoamyl alcohol purification and the protocol stopped after the precipitation step where multiple samples were combined to increase yield. cDNA was resuspended in DNA-RNase free water and quality control was performed on the Agilent 2100 Bioanalyzer (Agilent Technologies) using the DNA 7500 chip. cDNA samples were then taken through the Roche GS FLX Shotgun DNA Library Preparation procedure (Dec 2007 manual, catalog #04852265001). Libraries (454) were constructed and pyrosequenced as described previously (Poinar et al. 2006) at Penn State University. All 454 libraries sent for sequencing had mean fragment sizes between 300 and 800 bp and >10 ng of product. Additional sequencing was performed at Washington University in St. Louis, Missouri, for the premeiotic flower bud sample using the Sanger method.

Data processing, assembly, and annotation

Sequences from individual 454 libraries were extracted from SFF files and renamed to reflect the source material. The names of Sanger sequences also indicated the source library. After renaming, all sequences were combined into a single FASTA file. All sequences in the combined FASTA file were screened for contaminants and trimmed using SeqClean (<http://compbio.dfci.harvard.edu/tgi/software/>) with the Roche library adaptors, and the *Piper cenocladum* (C. DC.) chloroplast genome (NCBI accession NC_008326), mitochondrial gene sequences from magnoliids *Calycanthus floridus* (L.), *L. tulipifera*, *Laurus nobilis* (L.), *Piper betle* (L.), and *Asarum* spp. Qiu 96018, and the Univec database (<http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html>). After screening and trimming, the 454 and Sanger sequences were assembled using MIRA version 3.0.5 (<http://sourceforge.net/apps/mediawiki/mira-assembler>, Chevreaux et al. 2004) with default settings for EST sequences.

The resulting unpadding consensus sequences (i.e., unigenes) were assigned putative gene annotations from the PlantTribes 2.0 scaffold (Wall et al. 2008; <http://fgp.huckpsu.edu/tribe.php>). The PlantTribes 2.0 scaffold uses tribeMCL and orthoMCL to objectively classify the coding sequences of ten sequenced plant genomes (*A. thaliana* V7.0, *Chlamydomonas reinhardtii* V3.0, *Physcomitrella patens* V1.0, *Selaginella moellendorffii* V1.0, *Oryza sativa*

V5.0, *Sorghum bicolor* V1.0, *Vitis vinifera* V1.0, *Populus trichocarpa* V1.0, *Medicago truncatula* V1.0, *Carica papaya* V1.0) into Tribes and ortho groups (Orthos). Using custom perl scripts to parse the results of a BLASTx search (Altschul et al. 1990) against the inferred protein sequences of these ten genomes, unigenes were sorted into Tribes, which approximate gene families, and Orthos, which approximate putative orthologous gene sets. Each Tribe and Ortho in the PlantTribes database is annotated with a gene ontology (GO slim) term (Ashburner et al. 2000), conserved domain information (Marchler-Bauer et al. 2002), information from manually curated gene families, and common descriptive terms from the member sequences (Wall et al. 2008); accordingly, unigenes sorted into Tribes and Orthos are assigned the respective annotation. Unigenes with no significant (E value > 1e-5) hit to any of the ten sequenced genomes were searched against the GenBank non-redundant protein database.

GO enrichment analysis of the unigenes expressed in wood formation tissues was conducted using the DAVID Bioinformatics Resources 2008 with a False Discovery Rate (FDR) cutoff of 0.01 (Dennis et al. 2003; Huang et al. 2009). Simple sequence repeats (SSRs) were mined by using the scripts developed in-house in Clemson University Genomics Institute (CUGI). The minimum number of repeats was five for di-nucleotide repeats, four for tri-nucleotide repeats, three for tetra- and penta-nucleotide repeats, and two for hexa-nucleotide repeats. Primer3 was used to select candidate primers (Rozen and Skaletsky 2000). The single-copy gene coverage was calculated as the percent coverage of a *V. vinifera* reference gene (since *Vitis* represents the highest proportion of best hits from the annotation) by using the longest unigene in each tribe and ortho. The relative expression level is calculated as the percentage of the reads from each library in the overall reads from all libraries that were subjected to 454 sequencing. For comparative purposes (i.e., determining the most highly expressed unigenes), the expression level of each unigene was determined using the sum of the lengths of all reads assembled into the unigene over the length of that unigene.

Identification of conserved miRNA and prediction of their targets

Known miRNAs from the miRBase (release 14) were used to screen the *L. tulipifera* cDNA contig sequences using the program Patscan (Dsouza et al. 1997) with default parameters and two mismatches. Sequences with candidate miRNAs were first blasted against the *Arabidopsis* proteome; and sequences with hits to protein-encoding genes were removed. Filtered sequences were then checked for miRNA features using MIRcheck (Jones-Rhoades and

Bartel 2004). The targets of the identified miRNAs were searched in the *Liriodendron* cDNA dataset by using the approach previously described (Allen et al. 2005).

Expression of *LtuCAD1* in the *Arabidopsis CAD4/5* double mutant

The *LtuCAD1* gene was first cloned with *Bam*HI between the 35S promoter of the cauliflower mosaic virus (CaMV) and the nopaline synthase (*NOS*) gene terminator in a pBIN102-based binary vector. The *LtuCAD1* gene along with the 35S promoter and the *NOS* terminator were then cloned into the pCAMBIA1301 vector using the Gateway Cloning System (Carlsbad, California, US). The *Agrobacterium tumefaciens* strain GV31001 carrying the CAMBIA/*LtuCAD1* was used to transform *Arabidopsis CAD-C/D* double mutants (obtained from Dr. Armand Séguin in Canadian Forest Service, Canada) (Sibout et al. 2005) by the floral-dip method (Desfeux et al. 2000). *Arabidopsis* seeds transformed with the *LtuCAD1* were selected in Peter's plant food medium containing 25 µg/mL hygromycin.

Results and discussion

Sequencing of *Liriodendron* cDNA libraries from ten different tissue types and assembly

Non-normalized cDNA libraries were constructed for ten different types of *L. tulipifera* tissues: premeiotic flower buds, postmeiotic flower buds, open flowers, developing fruit, terminal buds, leaves, cambium, xylem, roots, and seedlings. All the libraries were sequenced with 454 GS FLX (one half plate each), except for Ltu01 (Liang et al. 2008) and Ltu01b, which were sequenced by the Sanger method (Table 1). Ltu01, Ltu01b, and Ltu19 (a 454 sequence library) were generated from the same mRNA preparation (for premeiotic flower buds). The average read length for 454 pyrosequencing was 235 bp, with the number of bases ranging from 45 to 63 Mb and the number of reads from 201,000 to 265,000. Assembly of all 12 libraries resulted in 137,923 unigenes (132,905 contigs and 4,599 singletons). The average unigene length was 478 bp, with 40 bp as the shortest and 5,807 bp the largest. Of the contigs in the final unigene set, 28,574 were 600 bp or longer and 17,020 were 800 bp or longer (Fig. 1). As indicated in Table 1, over 40,000 unigenes were expressed in each library, with 7,349 (size ranging from 103–4,931 bp) unigenes for broadly expressed genes (transcripts found in all surveyed tissue types). More than 2,000 unigenes were library-specific (i.e., tissue-specific), with premeiotic flower bud having the most unique unigenes (6,685), followed by root (4,994) and open flower (4,067).

All sequences and assemblies, as well as detailed information about each library, are available at <http://ancangio.uga.edu/content/liriodendron-tulipifera>. Sanger sequences were deposited in NCBI dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>) and 454 sequences were deposited in the NCBI Sequence Read Database <http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?>.

The average GC content for the 137,923 unigenes is 43.2% with a standard deviation of 6.5%, indicating that *L. tulipifera* genes tend to be slightly more AT-rich than annotated genes in currently sequenced genomes. The percentage GC composition in the *L. tulipifera* transcriptome is more similar to *A. thaliana* (42.7%) than to *O. sativa* (51.1%) (Kuhl et al. 2004). The codon usage in the translated sequences, generated by General Codon Usage Analysis (<http://bioinf.may.ie/gcua/index.html>; McInerney 1998), is represented in Online Resource 1. The pattern of codon preferences observed in the combined assembly was similar to *A. thaliana* (the Codon Use Database at <http://www.kazusa.or.jp/codon/>, GenBank Release 160.0, June 15, 2007), with only four different preferred codons. Only one amino acid (Leu) exhibits G or C at the degenerate third base of its preferred codon. This is consistent with the fact that dicots do not favor G and C in that position (Murray et al. 1989). Dinucleotides CG and TA are under-represented, which mirrors that of the *L. tulipifera* BAC and shotgun end sequence dataset (Liang et al. 2008) (Online Resource 2), as is common in eukaryotic sequences (Karlin et al. 1998).

Functional annotation and classification of the *Liriodendron* transcriptome

A BLASTX search of the 137,923 unigenes from the combined assembly, against ten sequenced plant genomes, revealed 68,464 matches (49.6% of the unigenes) with BLASTX (E value $\leq 10^{-5}$). Furthermore, a BLASTX search against the GenBank non-redundant protein database generated an additional 1,152 hits. Of all matches, 66.4% are either unknown, unnamed, hypothetical, or predicted proteins. The majority of the unigenes without similarity (76.0%) are less than 400 bp in length. When compared to model species with sequenced genomes, the *L. tulipifera* unigene set was most similar to *P. trichocarpa* (Torr. & Gray), with 46.9% of the unigenes having significant homology with *Populus* genes (BLASTX, E value $\leq 10^{-5}$). In contrast, only 43.0% and 42.3% of the *L. tulipifera* unigenes showed similarity to *Arabidopsis* and *Oryza* genes. Among the best BLASTX matches, woody angiosperm species have more hits (*V. vinifera* L. 39.8%, *P. trichocarpa* 20.9%, and *C. papaya* L. 14.0%) than the non-woody species (*M. truncatula* L. 9.2%, *A. thaliana* 5.8%, *O. sativa* 4.8%, and *S. bicolor* L. (Moench) 4.5%).

Table 1 Statistics for each *L. tulipifera* cDNA library

Tissue name	Library code	Sequencing method ^a	Number of reads	Number of bases (MB)	Average length (bp)	Unigenes in each library	Library specific unigenes
Assembly for 12 libraries			2,391,043	568.5	478 ^b	137,923 ^{b, c}	
Premeiotic flower bud	Ltu01 ^d	Sanger	9,442	3.96	418	59,393 ^e	6,685 ^e
Premeiotic flower bud	Ltu01b ^d	Sanger	14,601	8.29	566		
Premeiotic flower bud	Ltu19 ^d	454 FLX GS	264,000	63.00	238		
Postmeiotic flower buds	Ltu15	454 FLX GS	201,000	45.50	228	40,086	2,499
Open flower	Ltu14	454 FLX GS	263,000	62.80	239	50,812	4,067
Fruit	Ltu12	454 FLX GS	210,000	50.00	237	45,452	2,951
Terminal bud	Ltu11	454 FLX GS	258,000	59.31	229	47,391	2,540
Leaf	Ltu13	454 FLX GS	212,000	50.70	237	46,364	3,846
Cambium	Ltu10	454 FLX GS	265,000	57.44	216	43,642	2,931
Xylem	Ltu18	454 FLX GS	209,000	50.00	238	42,001	3,752
Root	Ltu16	454 FLX GS	264,000	63.50	240	51,285	4,994
Seedling	Ltu17	454 FLX GS	221,000	54.00	243	43,654	2,689

^a One half plate for all 454 FLX GS

^b Average unigene length and unigene number in the combined assembly

^c 7,349 unigenes were expressed in all surveyed tissue types

^d Generated from the same mRNA. Ltu01 was previously reported in Liang et al. 2008

^e These numbers were for the combination of three libraries (Ltu01, Ltu01b, and Ltu19) (for Unigenes in each library and Library specific unigenes)

BLASTX results with the *Arabidopsis* proteome can also be viewed through <http://ancangio.uga.edu/ng-genediscovery/liriodendron.jnlp> and the assembly can be searched using the Ancestral Angiosperm Genome Project blast interface at <http://ancangio.uga.edu/blast/blast.html>.

Detailed functional annotation of the unigenes was obtained by Gene Ontology (GO) slim terms: 41.2% of the unigenes can be assigned putative molecular functions, 39.7% have predicted cellular components, and 40.5% are given a biological process prediction. As seen in Fig. 2, a wide variety of putative functions are represented in the *Liriodendron* database. It is noteworthy that approximately 1.5% of the unigenes encode proteins with putative

transcription factor activity, 0.2% are related to the cell wall, and 0.5% are involved in developmental processes. In addition, single-copy genes are well represented in the dataset, with 98.2% of all single-copy tribes and 95.5% of all single-copy Orthos being populated by at least one *Liriodendron* unigene when compared to *V. vinifera* reference genes (Fig. 3).

Comparative genomics presents opportunities to study the dynamics of molecular evolutionary processes. However, the phylogenetic distribution of currently available genomic resources is not balanced, and this imbalance is even more acute in some clades, such as magnoliids (Jackson et al. 2006). This can lead to biasing evolutionary comparisons.

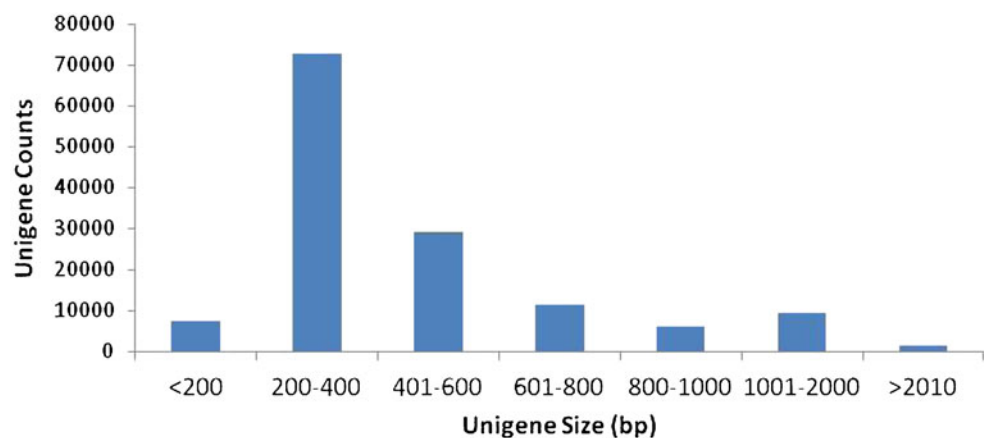
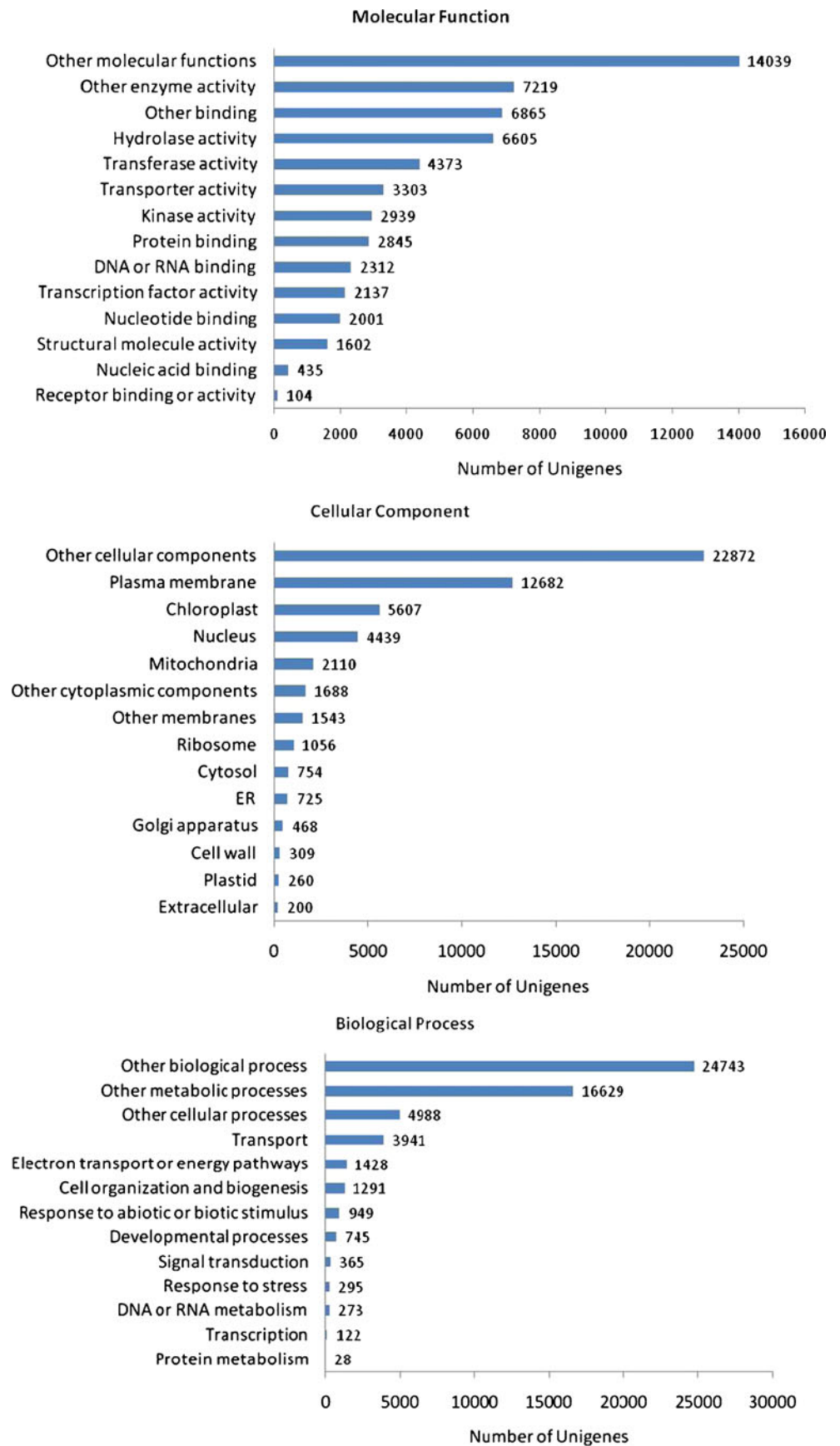
Fig. 1 *L. tulipifera* unigene size distribution

Fig. 2 GO-annotation classification of *L. tulipifera* unigene functions in terms of putative molecular functions, cellular components, and biological processes



Since the first *L. tulipifera* EST dataset became available, *Liriodendron* has been used a comparator to better understand the evolution of the origin and evolution of the flower (Zahn et al. 2005, 2006; Soltis et al. 2007; Chanderbali et al. 2010), as well as ancestral polyploidy in seed plants and angiosperms (CW dePamphilis, personal communication). Built from ten different tissue types, the new EST dataset is by far the most comprehensive genomic resource for *Liriodendron*. This resource will strengthen *Liriodendron*'s role in comparative studies of angiosperm evolution and facilitate molecular genetic and genomic investigations in *Liriodendron* and other species in the Magnoliaceae family.

In silico mining of simple sequence repeat markers

Simple sequence repeat (SSR) mining generated 29,289 repeats (dimers to pentamers), with 686 unique motifs. A total of 22,417 unigenes (16.3%) contain at least one SSR, with 53.1% of them having more than one SSR present. The number of SSRs identified in a unigene ranges from 1 to 15. This is consistent with the frequency of SSR-containing ESTs found in eudicotyledonous species, which ranges from 2.7% to 16.8% (Kumpatla and Mukhopadhyay 2005). Dimer repeats were the most commonly observed and constitute 41% of all the SSRs detected. The most common dimer, trimer, tetramer, and pentamer repeats are “ct,” “aag,” “tttc,” and “aaaag,” respectively. The SSR locations, forward and reverse primer sequences and their melting temperature (T_m) values, and expected amplicon sizes are listed in the Online Resources 3, 4, 5, and 6. After being validated, these SSRs can be applied in molecular breeding and investigations of candidate genes for traits of economic and ecological importance. These molecular markers may also be used to generate genetic maps for trait/gene association and refinement of candidate gene identification.

The genus *Liriodendron* contains only one other species, *Liriodendron chinense* (Hems1.) Sarg., which is native to China and Vietnam. This species is now considered an endangered species due to its limited seed production and small isolated populations (Xu et al. 2006). *L. tulipifera* and *L. chinense* are quite similar morphologically, except that the latter is smaller in stature. These two species are thought to have separated 10–16 million years ago (Parks and Wendel 1990), but hybridize readily (cf. Merkle et al. 1993). Preliminary data from Xu et al. (2006) indicated that 12 out of 15 single-locus SSR markers from the floral EST dataset of *L. tulipifera* (Albert et al. 2005; Liang et al. 2008) were found to be codominant and polymorphic in *L. chinense*, suggesting a high level of cross-species transferability. Thus, the SSRs developed from *L. tulipifera* can be applied in conservation of *L. chinense*. In a recent study

(Xu et al. 2010) using 132 SSR markers of the same source, 47.7% of the markers could be amplified in *Michelia maudiae* Dunn, 37.9% in *Manglietia maguanica* Chang et B.L. Chen, and 33.3% in *Magnolia amoena* Cheng. *Michelia*, *Manglietia*, and *Magnolia* are in the same Magnoliaceae family with *Liriodendron*. This suggests that the *L. tulipifera* SSRs can also be useful in related species of the same family, for which genomic resources are not available or very limited.

Conserved microRNA identification

MicroRNA (miRNAs) play an important role in plant development since they negatively control gene expression by cleaving or inhibiting the translation of mRNA of target

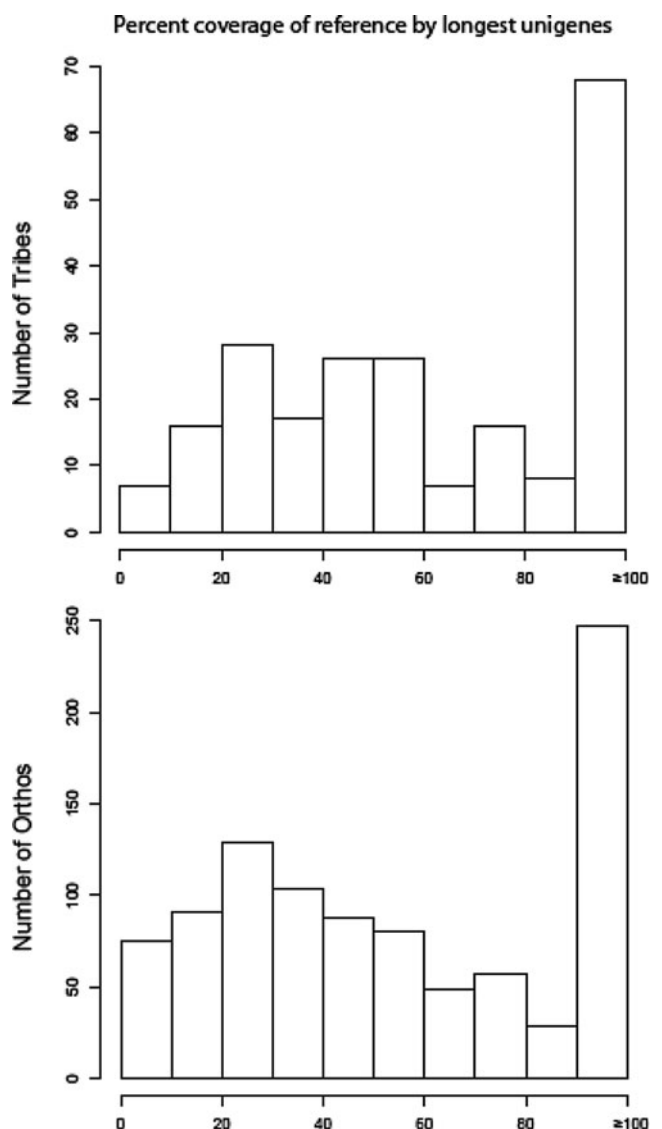


Fig. 3 The single-copy gene coverage in the *L. tulipifera* dataset. The percent coverage of a *Vitis vinifera* reference gene was calculated by using the longest unigene in each tribe and ortho

genes. Analysis of the *Liriodendron* transcript unigenes resulted in identification of 22 miRNA families from 53 unique miRNA precursor sequences (Online Resource 7). The number of sequence variants in each family varies between 1 and 9 bp. The number of miRNA families identified represents half of the number of conserved miRNA identified in plants. In a miRNA microarray study by Axtella and Bartel (2005), 13 out of the 23 families of *Arabidopsis* were found to be expressed in *L. tulipifera* leaves. We identified 8 of these 13 families of *Arabidopsis* miRNA in the *L. tulipifera* EST dataset.

The putative targets of these miRNAs are listed in the Online Resource 8. The miRNA target unigenes are involved in various molecular functions, cellular components, and biological processes. Molecular functions include DNA, RNA, nucleotide, or protein binding, hydrolase activity, kinase activity, structural molecule activity, transcription factor activity, transferase activity, and transporter activity. Endoplasmic reticulum (ER), Golgi apparatus, nucleus, ribosome, plastid, mitochondria, and chloroplast are among the cellular components. The biological processes include cell organization and biogenesis, developmental processes, response to abiotic or biotic stimulus, signal transduction, transcription, and transport. Among the 260 miRNA target unigenes being identified, 10% have hits in the Cell Wall Navigator database (Girke et al. 2004) and/or the MAIZEWALL dataset (Guillaumie et al. 2007), including one cellulose synthase gene and three monolignol biosynthesis-HCT (hydroxycinnamoyl CoA:shikimate/quinic acid hydroxycinnamoyltransferase). This resource provides an opportunity for functional and evolutionary studies of miRNAs in basal angiosperms.

Unigenes expressed in xylem and cambium tissues

Among the 137,923 *L. tulipifera* cDNA unigenes, 47% (64,247) were expressed in either cambium or xylem tissues. Table 2 reveals the significant GO enrichments of these unigenes. The most enriched GO terms in the biological process (BP) category include response to abiotic stimulus and post-embryonic development. In the cellular component (CC) category, plastid/plastid part, chloroplast/chloroplast part, plasma membrane, mitochondrion, (intracellular) non-membrane-bounded organelle, organelle membrane, and (organelle) envelope are the highly enriched terms, while in the molecular function (MF) category, helicase activity, nuclease activity, and nucleotide binding are highly enriched.

A total of 7,816 unigenes were found only in xylem and/or cambium tissues, with 3,752 unigenes specific to xylem tissue, 2,931 to cambium, and 1,132 common between these two tissue types (Fig. 4). A majority of the wood-specific unigenes (5,865, size ranging from 40 to 2,265 bp) did not have a match in the BLASTX search. The top seven most highly expressed unigenes are novel (no hits) (404–875 bp in length), followed by a *LOX3* (Lipoxygenase 3) and a *LSH1* (LIGHT-DEPENDENT SHORT HYPOCOTYLS 1) homolog. The wood tissue-specific unigenes expressed in both xylem and cambium tissues include genes expected to be involved in terpene synthesis (e.g., synthase, lupeol synthase 2, and allene oxide synthase homologs), cell wall formation (e.g., glycosyltransferase and polygalacturonate 4- α -galacturonosyltransferase), and lignin synthesis (e.g., cinnamoyl coa reductase 1 and caffeoyl-CoA 3-*O*-methyltransferase). This gene set will be

Table 2 The most enriched GO terms in the *L. tulipifera* xylem and cambium libraries

Category	GO term	Genes	Percent	FDR
BP	GO:0009628~response to abiotic stimulus	543	5.98	0
BP	GO:0009791~post-embryonic development	453	4.99	0
CC	GO:0009536~plastid	1,774	19.55	0
CC	GO:0009507~chloroplast	1,730	19.06	0
CC	GO:0005886~plasma membrane	961	10.59	0
CC	GO:0044435~plastid part	654	7.21	0
CC	GO:0044434~chloroplast part	635	7.00	0
CC	GO:0005739~mitochondrion	612	6.74	0
CC	GO:0043232~intracellular non-membrane-bounded organelle	530	5.84	0
CC	GO:0043228~non-membrane-bounded organelle	530	5.84	0
CC	GO:0031090~organelle membrane	514	5.66	0
CC	GO:0031975~envelope	471	5.19	0
CC	GO:0031967~organelle envelope	468	5.16	0
MF	GO:0004386~helicase activity	104	1.15	0
MF	GO:0004518~nuclease activity	108	1.19	0
MF	GO:0000166~nucleotide binding	1,303	14.36	0.003

BP biological process, CC cellular component, MF molecular function, FDR false discovery rate (cutoff=0.01)

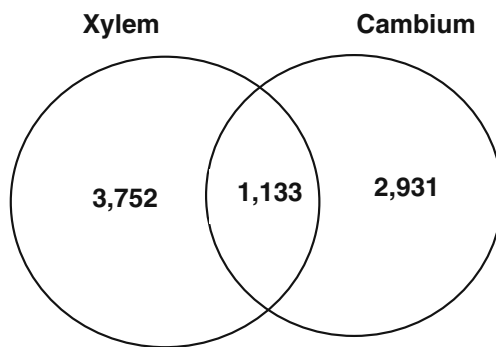


Fig. 4 Expression overlap between the 7,816 unigenes detected only in *L. tulipifera* xylem and/or cambium libraries. The Venn diagram shows that xylem and cambium share 1,133 of the “wood-specific” unigenes

a valuable resource for investigations aimed at improving and modulating wood properties.

Transcriptomes of wood-forming tissues have been sequenced for several tree species with economic importance, including *Eucalyptus* L'Hér, *P. trichocarpa*, loblolly pine (*Pinus taeda* L.), radiata pine (*Pinus radiata* D. Don), and white spruce (*Picea glauca* [Moench.] Voss.) (Li et al. 2009; 2010; Rengel et al. 2009). When sequences from these databases were used as queries in TBLASTX searches against the *Liriodendron* dataset, 52.1% (loblolly pine) to 89.2% (*Populus*) of the sequences found hit(s) at E value $\leq 1e-5$ (Table 3), suggesting most of the genes involved in *L. tulipifera* wood formation are well represented. When BLAST searches were performed against these publicly available xylogenesis databases, the percentage of hits in the *L. tulipifera* cDNA unigene set ranged from 14.6% to 32.6% (Table 3). The Cell Wall Navigator Database (Girke et al. 2004) is a primary wall gene database, including 661 sequences from *A. thaliana*, 641 from *O. sativa*, and 3,289 from UniProt. A total of 4,038 *L. tulipifera* unigenes matched the Cell Wall Navigator Database, representing all

categories (monosaccharide activation and interconversion, polysaccharide synthesis, reassembly, structural proteins, glycoprotein glycosyltransferases), 32 of the 35 protein families, and 97.4% of the 4,767 genes in the Cell Wall Navigator Database (Online Resource 9). Likewise, 99.3% of the 734 primary and secondary wall genes in the MAIZEWALL database (Guillaumie et al. 2007) were represented in the *L. tulipifera* EST resource. A total of 8,060 *L. tulipifera* cDNA unigenes (5.8%) matched the MAIZEWALL database, representing all 18 categories (Online Resource 10). These results are in line with the recent comparative investigation of Li et al. (2010), which suggested that vascular plants share a common ancestral xylem transcriptome, and while conifers have highly conserved xylem transcriptomes, angiosperm xylem transcriptomes are relatively diversified. The availability of the xylem transcriptome from a basal angiosperm species, such as *L. tulipifera*, not only provides a resource for molecular study of wood formation in basal angiosperm species, but also an opportunity to examine the evolution of the xylem genes in angiosperms in more detail.

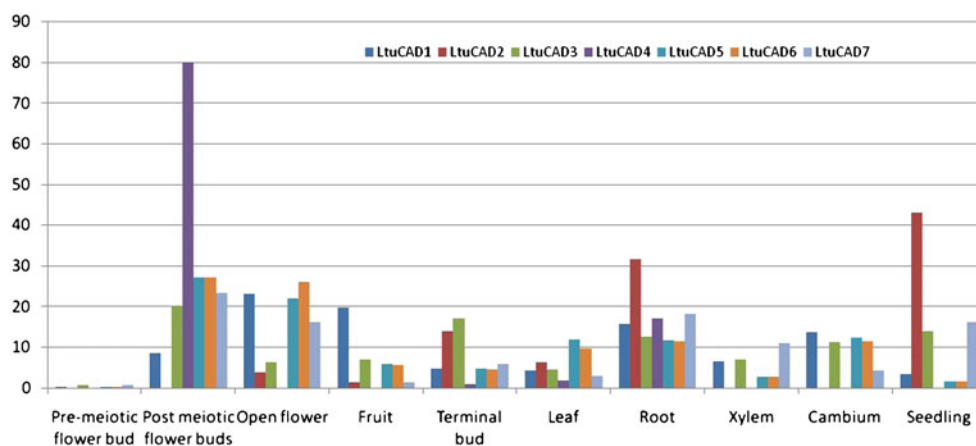
The cinnamyl alcohol dehydrogenase (CAD) is a key enzyme in lignin biosynthesis as it catalyzes the final step in the synthesis of monolignols. In *Arabidopsis*, CAD exists as a small multifamily consisting of nine genes (*AtCAD1* to *AtCAD9*) (Sibout et al. 2003). In *Oryza*, 12 CAD genes have been reported (Zhang et al. 2006), while there are 15, 18, and 18 CAD genes in *Populus*, *Vitis*, and *Medicago*, respectively (Barakat et al. 2009). There are seven CAD homologs (*LtuCAD1* to *LtuCAD7*) with full-length coding sequence present in this new *L. tulipifera* EST dataset. As can be seen in Fig. 5, the *L. tulipifera* CAD genes show different expression patterns in the ten tissue types included in this study, suggesting that they may be involved in different biological processes. For example, one CAD gene (*LtuCAD4*) was relatively highly expressed in postmeiotic flower buds, with ca. 80% of the reads for that gene coming

Table 3 Comparison of *L. tulipifera* transcriptome with publicly available xylogenesis and cell wall formation EST datasets

	No. of total unigenes in reference dataset	No. of hit unigenes in reference dataset	No. of hit unigenes in <i>Liriodendron</i> dataset
EUCAWOOD DB (<i>Eucalyptus</i>) (Rengel et al. 2009)	3,928	2,113 (53.8%)	21,786 (15.8%)
<i>Populus</i> ^a	7,991	7,126 (89.2%)	36,802 (26.7%)
Radiata pine Xylem DB (Li et al. 2009)	3,304	2,090 (63.3%)	20,069 (14.6%)
Loblolly pine Xylem DB ^a	18,320	9,541 (52.1%)	36,792 (26.7%)
White Spruce Xylem DB ^a	12,489	10,277 (82.3%)	44,939 (32.6%)
Cell Wall Navigator DB (Girke et al. 2004)	4,767	4,643 (97.4%)	4,038 (2.9%)
Maize Wall DB (Guillaumie et al. 2007)	734	729 (99.3%)	8,060 (5.8%)

^a Sequences were from Li et al. 2010

Fig. 5 Relative digital expression level of the *L. tulipifera* CAD family genes in ten different tissue types. The relative expression level is calculated from the number of reads for a given gene from each library expressed as a percentage of the total number of reads for that gene from all of the libraries that were subjected to 454 sequencing



from the postmeiotic flower bud library. Of the reads, 18% for *LtuCAD4* were detected in the root library, and a much smaller fraction of reads were drawn from several other libraries. All the *LtuCAD* genes were expressed at a very low level, if at all, in the premeiotic floral buds, which is consistent with the fact that premeiotic floral buds are a young tissue type with limited lignified walls (the bud outer scales were peeled off before the buds were homogenized for RNA extraction). *LtuCAD1*, which had a BLASTX *E* value of $3e-132$ to *AtCAD4* and $2e-130$ to *AtCAD5* (the two *Arabidopsis* CAD genes that have been shown to have major roles in lignin synthesis, (Sibout et al. 2005)), was expressed in all the ten tissue types being surveyed, with strongest expression in open flower, followed by fruit, root, and cambium tissue. Less *LtuCAD1* expression was detected in premeiotic floral bud and leaf. A similar expression pattern has been reported for *AtCAD4* and *AtCAD5* (Sibout et al. 2005).

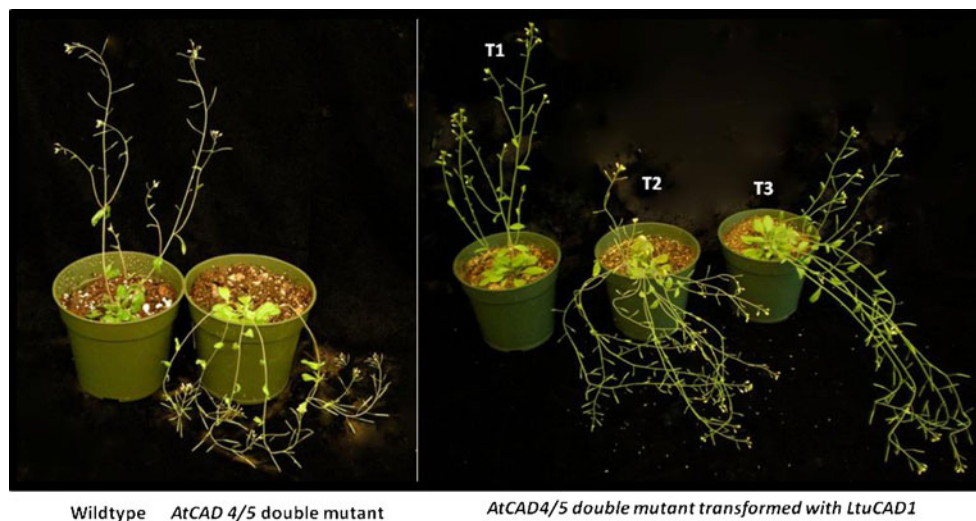
The *Arabidopsis* *CAD4* and *CAD5* double mutant plants have a phenotype with a limp floral stem at maturity

(Sibout et al. 2005), while no visual phenotypes were observed in *Atcad-4* and *Atcad-5* single mutants grown in the greenhouse (Sibout et al. 2003). When the *LtuCAD1* gene was over-expressed in the *Arabidopsis* *CAD4* and *CAD5* double mutant plants, stiffness of the floral stems was at least partially recovered (Fig. 6). This suggests the likely involvement of the *LtuCAD1* in lignin biosynthesis and also indicates that there has been sufficient conservation between the *Liriodendron* and *Arabidopsis* proteins to permit cross-species functional studies in these distantly related model species.

Conclusions

We report the sequencing, assembly, and annotation of 137,923 unigenes (132,905 contigs and 4,599 singletons, size ranging from 40 to 5,807 bp) derived from non-normalized cDNA libraries, which represented ten *L. tulipifera* tissue types: premeiotic flower buds, postmeiotic

Fig. 6 Transformation of *Arabidopsis* *CAD4/5* double-mutant with *LtuCAD1* shows partial phenotype recovery. T1, T2, and T3 are different transformed lines with *LtuCAD10*



flower buds, open flowers, developing fruit, terminal buds, leaves, cambium, xylem, roots, and seedlings. About 50% of the unigenes were significantly similar to publicly available plant protein sequences, representing a wide variety of putative functions. Putative BLAST-based homologs of most of the genes involved in cell wall construction are represented, including seven full-length cinnamyl alcohol dehydrogenase-encoding genes (*LutCAD1* to *LtuCAD7*). Approximately 50% of the unigenes did not match any sequence in the public databases, including the complete genomes of *Arabidopsis*, *Oryza*, and *Populus*. Some of these novel genes might be unique in basal angiosperm species and may be informative for understanding the origins of diverged gene families when characterized. In addition, about 30,000 simple sequence repeats (SSRs) have been identified. This new *Liriodendron* dataset currently provides the most comprehensive list of unigenes for any Magnoliaceae species. This large-scale genomic resource will facilitate gene discovery and cDNA microarray production in *L. tulipifera* and related species. The unigene sequences will become valuable in comparative and functional genomics of genes involved in the development of flowers, fruits, roots, buds, and wood formation, as well as in unraveling the molecular regulation of these important developmental stages in *Liriodendron*. This deep EST dataset will also further strengthen *L. tulipifera*'s role in comparative study as a basal angiosperm species.

Sanger sequences generated by this report are accessible in NCBI dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>) and 454 sequences are available in the NCBI Sequence Read Database (<http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?>). Assemblies and BLASTX results against the *Arabidopsis* proteome can be viewed through <http://ancangio.uga.edu/ng-genediscovery/liriodendron.jnlp>, and the assembly can be searched using the Ancestral Angiosperm Genome Project blast interface at <http://jlmwiki.plantbio.uga.edu/blast/blast.html>.

Acknowledgments We thank Stephan Schuster and Lynn Tomsho for their assistance in 454 sequencing, Yi Hu for RNA isolations, Denis S. Diloreto for seedlings, Stephen Ficklin for the mining of SSRs, and Xinguo Li for providing the pure xylem unigenes for *Populus*, loblolly pine, and white spruce. This study was mainly supported by the National Science Foundation grant, Ancestral Angiosperm Genome project (Award # DBI-0638595, PI: dePamphilis). A National Institute of Food and Agriculture, USDA grant to HL (project number SC-1700324, technical contribution No. 5832 of the Clemson University Experiment Station) contributed the sequencing of a one half 454 plate.

References

- Albert VA, Soltis DE, Carlson JE, Farmerie WG, Wall PK, Ilut DC et al (2005) Floral gene resources from basal angiosperms for comparative genomics research. *BMC Plant Biol* 5:5
- Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121:207–221
- Altschul SF, Gish W, Miller EW, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Axtella MJ, Bartel DP (2005) Antiquity of microRNAs and their targets in land plants. *Plant Cell* 17:1658–1673
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29
- Bae K, Byun J (1987) Screening of leaves of higher plants for antibacterial action. *Kor J Pharmacog* 8:1
- Barakat A, Bagniewska-Zadworna A, Choi A, Plakkat U, DiLoreto D, Yellanki P, Carlson J (2009) The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. *BMC Plant Biol* 9:26
- Berlin A, Maximenko V, Bura R, Kang KY, Gilkes N, Saddler J (2005) A rapid microassay to evaluate enzymatic hydrolysis of lignocellulosic substrates. *Biotechnol Bioeng* 93:880–886
- Beck DE (1990) *Liriodendron tulipifera* L. yellow-poplar. In: Burns RM, Honkala BH (tech. coords.) *Silvics of North America: 1. Conifers; 2. Hardwoods*. Agriculture Handbook 654, USDA, Forest Service, Washington, DC, 2:877
- Carlson JE, Leebens-Mack JH, Wall PK, Zahn LM, Mueller LA, Landherr LL, Hu Y, Ilut DC, Arrington JM, Choirean S, Becker A, Field D, Tanksley SD, Ma H, dePamphilis CW (2006) EST database for early flower development in California poppy (*Eschscholzia californica* Cham., Papaveraceae) tags over 6,000 genes from a basal eudicot. *Plant Mol Biol* 62:351–369
- Çelen I, Harper D, Labbé N (2008) A multivariate approach to the acetylated poplar wood samples by near infrared spectroscopy. *Holzforschung* 62:189–196
- Chanderbali AS, Yoo M-J, Zahn LM, Brockington SF, Wall PK, Gitzendanner MA, Albert VA, Leebens-Mack J, Altman NS, Ma H, dePamphilis DW, Soltis DE, Soltis PS (2010) Conservation and canalization of gene expression during angiosperm diversification accompany the origin and evolution of the flower. *PNAS* 107:22570–22575
- Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Rep* 11:113–116
- Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 14:1147–1159
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 24:P3
- Desfeux C, Clough SJ, Bent AF (2000) Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral-dip method. *Plant Physiol* 123:895–904
- Dsouza M, Larsen N, Overbeek R (1997) Searching for patterns in genomic data. *Trends Genet* 13:497–498
- Gerke T, Lauricha J, Tran H, Keegstra K, Raikhel NV (2004) The cell wall navigator database: a systems-based approach to organism-unrestricted mining of protein families involved in cell wall metabolism. *Plant Physiol* 136:3003–3008
- Guillaumie S, San Clemente H, Deswarte C, Martinez Y, Lapierre C, Murigneux A, Barrière Y, Pichon M, Goffner D (2007) MAIZE-WALL: database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. *Plant Physiol* 143:339–363
- Hernandez R, Davalos JF, Sonti SS, Kim Y, Moody RC (1997) Strength and stiffness of reinforced yellow-poplar glued laminated

- beams. Res Pap FPL-RP-554. US Department of Agriculture, Forest Service, Forest Products Laboratory, Madison
- Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc* 4:44–57
- Hunt D (1998) Magnolias and their allies. International Dendrology Society & Magnolia Society, pp 304
- Hwang SS, Lee SJ, Kim HK, Ka JO, Kim KJ, Song HG (2008) Biodegradation and saccharification of wood chips of *Pinus strobus* and *Liriodendron tulipifera* by white rot fungi. *J Microbiol Biotechnol* 18:1819–1826
- Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee S-B, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *PNAS* 104:19369–19374
- Jackson S, Rounsley S, Purugganan M (2006) Comparative sequencing of plant genomes: choices to make. *Plant Cell* 18:1100–1104
- Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 14:787–799
- Karlin S, Campbell AM, Mrázek J (1998) Comparative DNA analysis across diverse genomes. *Ann Rev Gen* 32:185–225
- Kim KD, Lee EJ (2005) Potential tree species for use in the restoration of unsanitary landfills. *Environ Manage* 36:1–14
- Klugh KR, Cumming JC (2003) Variation in organic acid exudates among mycorrhizal species colonizing *Liriodendron tulipifera* L. (yellow poplar) in the presence of aluminum. *Ecological Society of America Annual Meeting Abstracts* 88:186
- Koo B-W, Min B-C, Park N, Eom C, Yeo H, Ryu K-O, Choi I-G (2008) Chemical and physical characterizations of *Liriodendron tulipifera* on growth periods. Proceedings of the 30th Symposium on Biotechnology for Fuels and Chemicals, 4–7 May, New Orleans, LA, USA
- Koo B-W, Park N, Yeo H, Lee S-Y, Kim H-Y, Kim H, Choi I-G (2009) Organosolv pretreatment of *Liriodendron tulipifera* with acid and alkali catalysts. Proceedings of the 31st Symposium on Biotechnology for Fuels and Chemicals, 3–6 May, San Francisco, CA, USA
- Kuhl JC, Cheung F, Yuan Q, Martin W, Zewdie Y, McCallum J, Catanach A, Rutherford P, Sink KC, Jenderek M, Prince JP, Town CD, Havey MJ (2004) A unique set of 11,008 onion expressed sequence tags reveals expressed sequence and genomic differences between the monocot orders Asparagales and Poales. *Plant Cell* 16:114–125
- Kumpatla S, Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* 48:985
- Lafayette PR, Eriksson KE, Dean JF (1999) Characterization and heterologous expression of laccase cDNAs from the lignifying xylem of yellow-poplar (*Liriodendron tulipifera*). *Plant Mol Biol* 40:23–35
- Li X, Wu HX, Dillon SK, Southerton SG (2009) Generation and analysis of expressed sequence tags from six developing xylem libraries in *Pinus radiata* D. Don. *BMC Genomics* 10:41
- Li X, Wu HX, Southerton SG (2010) Comparative genomics reveals conservative evolution of the xylem transcriptome in vascular plants. *BMC Evol Biol* 10:190
- Liang H, Barakat A, Schlarbaum SE, Carlson JE (2011) Organization of the chromosome region harboring a *FLORICAULA/LEAFY* gene in *Liriodendron*. *Tree Genet Genomes* 7:373–384
- Liang H, Barakat A, Schlarbaum SE, Mandoli DS, Carlson JE (2010) Comparison of gene order of *GIGANTEA* loci in yellow-poplar, monocots and eudicots. *Genome* 53:533–544
- Liang H, Carlson JE, Leebens-Mack JH, Wall PK, Mueller LA, Buzgo M et al (2008) An EST database for *Liriodendron tulipifera* L. floral buds: the first EST resource for functional and comparative genomics in *Liriodendron*. *Tree Genet Genomes* 4:419–433
- Liang H, Fang EG, Tomkins JP, Luo M, Kudrna D, Kim HR et al (2007) Development of a BAC library for yellow-poplar (*Liriodendron tulipifera*) and the identification of genes associated with flower development and lignin biosynthesis. *Tree Genet Genomes* 3:215–225
- Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucl Acids Res* 30:281–283
- McInerney J (1998) GCUA (General Codon Usage Analysis). *Bioinformatics* 14:372–373
- Merkle SA, Hoey MT, Watson-Pauley BA, Schlarbaum SE (1993) Propagation of *Liriodendron* hybrids via somatic embryogenesis. *Plant Cell Tissue Organ Cult* 34:191–198
- Moody RC, Hernandez R, Davalos JF, Sonti SS (1993) Yellow poplar timber beam performance. Res. Pap. FPL-RP-520. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison
- Moon MK, Oh HM, Kwon BM, Baek NI, Kim SH KJS, Kim DK (2007) Farnesyl protein transferase and tumor cell growth inhibitory activities of lipiferolide isolated from *Liriodendron tulipifera*. *Arch Pharm Res* 30:299–302
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic. *PNAS* 104:19363–19368
- Murray EE, Lotzer J, Eberle M (1989) Codon usage in plant genes. *Nucl Acid Res* 17:477–498
- Parks CR, Wendel JF (1990) Molecular divergence between Asian and North American species of *Liriodendron* (Magnoliaceae) with implications for interpretation of fossil floras. *Am J Bot* 77:1243–1256
- Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A et al (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* 311:392–394
- Qiu YL, Dombrowska O, Lee J, Li L, Whitlock BA, Bernasconi-Quadroni F, Rest JS, Davis CC, Borsch T, Hilu KW, Renner SS, Soltis DE, Soltis PS, Zanis MJ, Cannone JJ, Gutell RR, Powell M, Savolainen V, Chatrou LW, Chase MW (2005) Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *Intl J Plant Sci* 166:815–842
- Ranocha P, Chabannes M, Chamayou S, Danoun S, Jauneau A, Boudet AM, Goffner D (2002) Laccase down-regulation causes alterations in phenolic metabolism and cell wall structure in poplar. *Plant Physiol* 129:145–155
- Rengel D, Clemente HS, Servant F, Ladouce N, Paux E, Wincker P, Couloux A, Sivadon P, Grima-Pettenati J (2009) A new *Eucalyptus* genomic resource dedicated to wood formation: a comprehensive survey. *BMC Plant Biol* 9:36
- Ronse de Craene L, Soltis DE, Soltis PS (2003) Evolution of floral structures in basal angiosperms. *Intl J Plant Sci* 164:S329–S363
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics methods and protocols in the series Methods in Molecular Biology*. Humana, Totowa, pp 365–386
- Sibout R, Eudesb A, Mouilleb G, Polletc B, Lapierrerc C, Jouaninb L, Séguin A (2005) *CINNAMYL ALCOHOL DEHYDROGENASE-C* and *-D* are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. *Plant Cell* 17:2059–2076

- Sibout R, Eudes A, Pollet B, Goujon T, Mila I, Granier F, Séguin A, Lapiere C, Jouanin L (2003) Expression pattern of two paralogs encoding cinnamyl alcohol dehydrogenases in *Arabidopsis*: isolation and characterization of the corresponding mutants. *Plant Physiol* 132:848–860
- Soltis PS, Soltis DE, Chase MW, Endress PK, Crane PR (2004) The diversification of flowering plants. In: Cracraft J, Donoghue MJ (eds) *Assembling the tree of life*. Oxford University Press, Oxford, pp 154–167
- Soltis DE, Soltis PS, Chase MW, Endress P (2005) Phylogeny, evolution, and classification of flowering plants. Sinauer Associates, Sunderland
- Soltis DE, Ma H, Frohlich MW, Soltis PS, Albert VA, Oppenheimer DG, Altman NS, dePamphilis CD, Leebens-Mack J (2007) The floral genome: an evolutionary history of gene duplication and shifting patterns of gene expression. *Trends Plant Sci* 12:358–367
- Wall PK, Leebens-Mack J, Müller KF, Field D, Altman NS, de Pamphilis CW (2008) PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic Acids Res* 36:D970–D976
- Wei ZX, Wu ZY (1993) Pollen ultrastructure of *Liriodendron* and its systematic significance. *Acta Bot Yunnanica* 15:163–166
- Williams RS, Feist WC (2004) Durability of yellow-poplar and sweetgum and service life of finishes after long-term exposure. *For Prod J* 54:96–101
- Xiang Q, Lee LYY, Torget RW (2004) Kinetics of glucose decomposition during dilute-acid hydrolysis of lignocellulosic biomass. *Appl Biochem Biotechnol* 114:1127–1138
- Xu M, Li H, Zhang B (2006) Fifteen polymorphic simple sequence repeat markers from expressed sequence tags of *Liriodendron tulipifera*. *Mol Ecol Notes* 6:728–730
- Xu M, Sun Y, Li H (2010) EST-SSRs development and paternity analysis for *Liriodendron* spp. *New For* 40:361–382
- Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis D, dePamphilis CW, Ma H (2005) The evolution of the *SEPALLATA* subfamily of MADS-box genes: a pre-angiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169:2209–2223
- Zahn LM, Leebens-Mack J, de Pamphilis CW, Ma H, Theissen G (2006) To B or not to B a flower: the role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. *J Hered* 96:225–240
- Zhang K, Qian Q, Huang Z, Wang Y, Li M, Hong L, Zeng D, Gu M, Chu C, Cheng C (2006) *GOLD HULL* and *INTERNODE2* encodes a primarily multifunctional cinnamyl-alcohol dehydrogenase in rice. *Plant Physiol* 140:972–983