Tree Genetics & Genomes (2008) 4:419–433
DOI 10.1007/s11295-007-0120-2

ORIGINAL PAPER

An EST database for *Liriodendron tulipifera* L. floral buds: the first EST resource for functional and comparative genomics in *Liriodendron*

Haiying Liang · John E. Carlson · James H. Leebens-Mack · P. Kerr Wall · Lukas A. Mueller · Matyas Buzgo · Lena L. Landherr · Yi Hu · D. Scott DiLoreto · Daniel C. Ilut · Dawn Field · Steven D. Tanksley · Hong Ma · Claude W. dePamphilis

Received: 6 September 2007 / Revised: 19 October 2007 / Accepted: 4 November 2007 / Published online: 20 December 2007

Abstract *Liriodendron tulipifera* L. was selected by the Floral Genome Project for identification of new genes related to floral diversity in basal angiosperms. A large, non-normalized cDNA library was constructed from pre-meiotic and meiotic floral buds and sequenced to generate a database of 9,531 high-quality expressed sequence tags. These sequences clustered into 6,520 unigenes, of which 5,251 were singletons, and 1,269 were in contigs. Homologs of genes regulating many aspects of flower development were identified, including those for organ identity and development, cell and tissue differentiation, and cell-cycle control. Almost 5% of the transcriptome consisted of homologs to known floral gene families. Homologs of most of the genes involved in cell-wall construction were also recovered. This provides a new opportunity for comparative studies in lignin biosynthesis, a trait of key importance in the evolution of land plants and in the utilization of fiber from economically important tree species, such as *Liriodendron*. Also of note is that 1,089 unigenes did not match any sequence in the public databases, including the complete genomes of *Arabidopsis*, rice, and *Populus*. Some of these novel genes might be unique in basal angiosperm species and, when better characterized, may be informative for understanding the origins of diverged gene families. Thus, the *Liriodendron*

Communicated by A.E. Jeffery Dean

Electronic Supplementary Material The online version of this article (doi:10.1007/s11295-007-0120-2) contains supplementary material, which is available to authorized users.

H. Liang · J. E. Carlson (✉) · D. S. DiLoreto
School of Forest Resources, the Huck Institutes of the Life Sciences, Pennsylvania State University,
University Park, PA 16802, USA
e-mail: jec16@psu.edu

J. H. Leebens-Mack · P. K. Wall · L. L. Landherr · Y. Hu · H. Ma · C. W. dePamphilis
Department of Biology, the Huck Institutes of the Life Sciences, Pennsylvania State University,
University Park, PA 16802, USA

L. A. Mueller · D. C. Ilut · S. D. Tanksley
Department of Plant Breeding, Cornell University,
Ithaca, NY 14853, USA

M. Buzgo
Department of Botany, University of Florida,
Gainesville, FL 32611l, USA

D. Field
Oxford Centre for Ecology and Hydrology,
Mansfield Road,
Oxford OX1 3SR, UK

Present address:
H. Liang
Department of Genetics and Biochemistry, Clemson University,
Clemson, SC 29634, USA

Present address:
J. H. Leebens-Mack
Department of Plant Biology, University of Georgia,
Athens, GA 30602, USA

Present address:
M. Buzgo
Department of Biology, University of Louisiana in Shreveport,
Shreveport, LA 71115, USA

expressed sequence tag database and library will help bridge our understanding of the mechanisms of flower initiation and development that are shared among basal angiosperms, eudicots, and monocots, and provide new opportunities for comparative analysis of gene families across angiosperm species.

Keywords EST database · Flower development · *Liriodendron tulipifera* · Magnoliaceae · Basal angiosperm

Abbreviations

ABI	Applied Biosystems
AP	*APETALA* gene
bp	base pairs
DEPC	diethylpyrocarbonate
EF-1-α	*Elongation Factor 1-alpha gene*
EST	expressed sequence tag
GO	Gene Ontology Consortium
K_s	rate of synonymous substitutions
MYB	*myeloblastosis*-like gene
PCR	polymerase chain reaction
RCA	rolling-circle amplification

Introduction

Commonly known as tuliptree, tulip-poplar, yellow-poplar, white-poplar, and whitewood, *Liriodendron* is a genus of the magnolia family (Magnoliaceae) in the order Magnoliales. Magnoliales and three other orders (Laurales, Piperales, and Canellales) comprise the magnoliids, which, along with Amborellales, Nymphaeles, Illiciales and several others, are together known as the "basal angiosperms" (Soltis et al. 2005). Recent molecular phylogenetic analysis of large datasets derived from chloroplast genome sequences identify the magnoliids as the likely sister group to a large clade of angiosperms containing monocots and eudicots (Jansen et al. 2007). *Liriodendron* consists of only two species, with one native to China and Vietnam [*Liriodendron chinense* (Hemsl.) Sarg] and another to eastern North America (*Liriodendron tulipifera* L.). These two species are thought to have separated 10–16 Ma (Parks and Wendel 1990). They are quite similar morphologically, although *L. chinense* is smaller in stature than *L. tulipifera*. The cup-shape flowers of *Liriodendron* are magnificent and highly fragrant, with a unique arrangement of floral organs. Unlike most angiosperms, whose flower parts are in whorls, *Liriodendron* has its stamens and pistils in spirals on a conical receptacle. This arrangement is found in early fossil angiosperms, and spiral arrangements may be a primitive angiosperm character (Hunt et al. 1998; Ronse de Craene 2003). The perianth of *Liriodendron* consists of three outer and six inner tepals, unlike the clearly differentiated sepals and petals in most other flowering plants. Such features place *Liriodendron* at a phylogenetic position that is ideal for comparative studies of the evolution of floral development. Thus, the Floral Genome Project included *L. tulipifera* in its study on the evolutionary genomics of flowers.

The Floral Genome Project has investigated the origin, conservation, and diversification of the genetic architecture of the flower. Dozens of genes that play important roles controlling normal flower development had been identified from model plants such as *Arabidopsis thaliana* and *Antirrhinum majus* (Zhao et al. 2001; Ma 2005; Zahn et al. 2006), while many genes with critical roles in the evolution of flower development remained undiscovered. To facilitate transferring knowledge from model systems to nonmodel organisms and to understand the origin and subsequent diversification of flowers, the Floral Genome Project systematically targeted phylogenetically critical lineages for which floral gene sequences were missing (Soltis et al. 2002).

Besides its unusual flower structure as a characteristic of basal angiosperms, *L. tulipifera* is also an economically valuable timber and landscaping tree. As the tallest hardwood species in eastern US forests, *L. tulipifera* grows rapidly and attains heights of 80–120 ft with a trunk diameter of 2 to 5 ft. *L. tulipifera* is amenable to genetic transformation and mass production of somatic clones (Wilde et al. 1992), making it a system of choice for biotechnology applications, including the engineering of trees for phytoremediation of soils contaminated by heavy metals (Merkle 2006). Because of its rapid growth and biomass accumulation, unusual resistance to insects and disease, and its niche as an aggressive pioneer tree species on disturbed lands, yellow poplar has great future potential to be domesticated as a source of fiber for biologically based products, for biofuels, and chemicals. The potential of yellow poplar to meet the increasing demand of renewable energy is becoming more apparent through recent research (Glasser et al. 2000; Hayward et al. 2000; Kim and Lee 2002; Kim et al. 2001; Nagle et al. 2002) on conversion of biomass from yellow poplar to biofuels such as ethanol.

Liriodendron has been used extensively as a key exemplar species in studies on plant evolution (Parks and Wendel 1990; Wen 1999; Endress and Igersheim 2000; Zahn et al. 2005a, b; Cai et al. 2006). Ironically, there is very little information about its nuclear genome. The present study reports the creation of an expressed sequence tag (EST) database for *L. tulipifera* to support evolutionary genomic studies of flowers and other developmental traits. This is the first EST data set for *Liriodendron*, and it contains over 6,500 unigenes, representing a wide variety of putative functions.

Materials and methods

Tissue source

Young floral buds of *L. tulipifera* were obtained from one ramet of clone 108 in the yellow poplar breeding orchard of the University of Tennessee at Knoxville's Tree Improvement Program. Flower buds were collected on the dates of August 14, 2001, August 30, 2001, October 4, 2001, November 5 and 6, 2001, and April 29, 2002 in the size range of 1.0 to 35.0 mm in length and approximately 2.0 to 5.0 mm diameter covering all of the premeiotic and meiotic stages of flower-bud development (Fig. 1a–e). Floral stages were defined by Buzgo et al. (2006) as organ initiation when the buds are below 1.2 mm in diameter [first internal tissue changes at 1.2 mm (Fig. 1f)]; microsporangia first recognized at stage 6 at 1.7 mm (Fig. 1g); and male meiosis starting at stage 8 at approximately 1.8 mm (Fig. 1h). Buds were immediately quick frozen in liquid nitrogen upon harvest and stored at −80°C. Before RNA extraction, leaves and sepals were removed and floral tissues verified, during which the buds were kept frozen. Floral buds were then

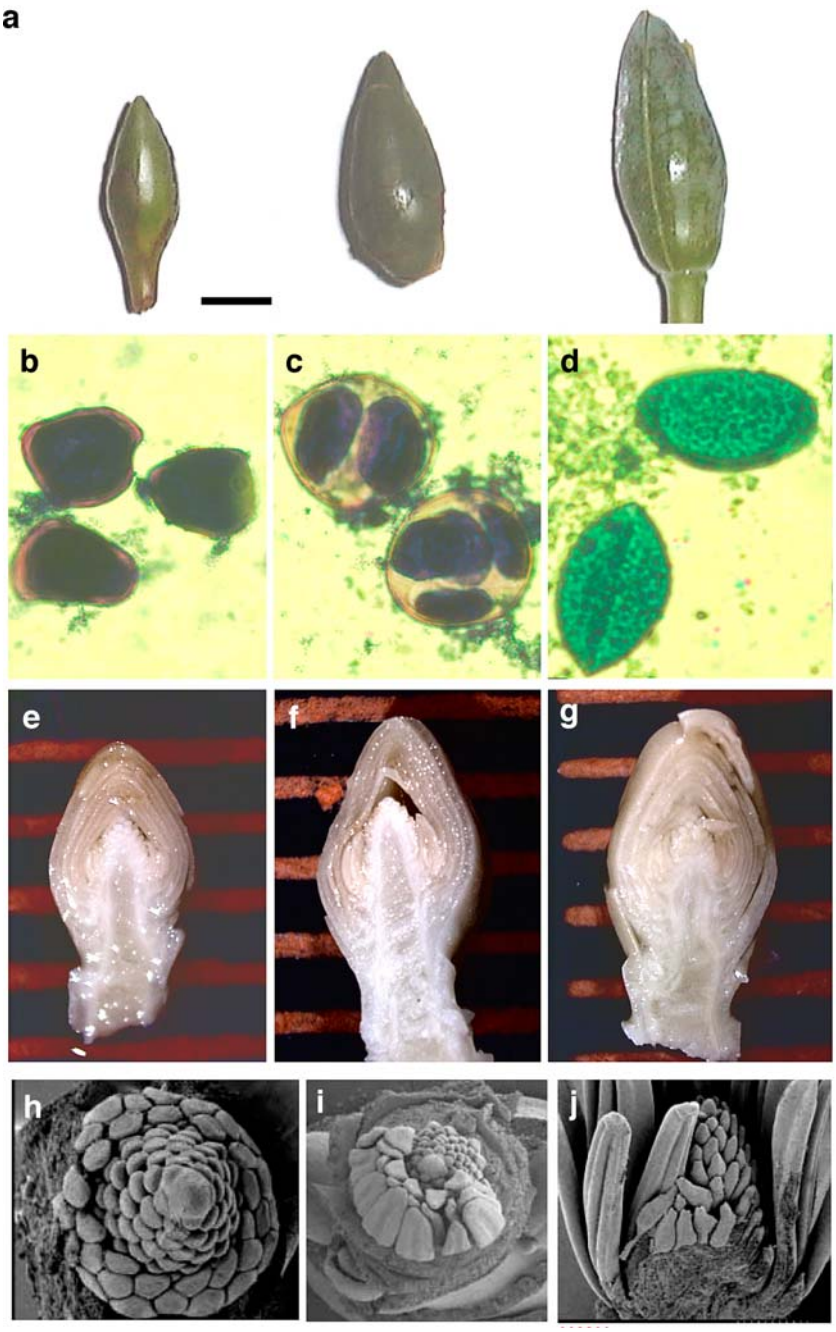

Fig. 1 Floral buds used for the cDNA library construction. **a** Three immature *Liriodendron* floral buds, with outer green bracts visible. The *left one* contains male meiotic cells (shown in **b**) at the time of meiosis. The *center one* contains tetrads (see **c**), the products of male meiosis. The *right one* has developing pollen (see **d**). *Bar* 5 mm. **b** Male meiocytes from the floral bud on the *left* in **a**. **c** Tetrads from the bud at the *center* in **a**. **d** Developing pollen grains from the bud on the *right* in **a**. **e**–**g** Dissected floral buds showing internal floral organ primordial and organs, with the green bracts already removed. These were the materials used for cDNA library construction. The *orange horizontal bars* in the background are 1 mm apart in each of **e**–**g**. **h**–**j** Electron scanning micrographs of the floral stages chosen for RNA isolation. **h** First internal tissue changes at 1.2 mm bud length (organs initiated when the buds are below 1.0 μm in diameter); **i** microsporangia externally recognized at stage 6 at 1.7 mm; **j** male meiosis starting at stage 8 putatively at 1.8 mm bud length

sorted by size to ensure adequate representation of the early stages of bud development. The inner parts of the buds, including floral meristematic regions, used for RNA isolation were 1.0 to 2.5 mm in length. Three sets of buds were assembled based on size and developmental stage (pool 1—August 14, 2001, August 30, 2001, and October 4, 2001 collections; pool 2—November 5 and 6, 2001 collections; pool 3—April 29, 2002 collection).

RNA extraction

RNA was extracted from each of the three sets of buds separately, according to the manufacturer's protocol (http://www.ambion.com/techlib/prot/bp_1911.pdf) for the RNAqueous®-Midi Kit (Ambion), with modifications as per Carlson et al. (2006). RNA was precipitated using 0.1 volumes sodium acetate and three volumes of 100% ethanol. RNA was dissolved into RNase-free [diethylpyrocarbonate (DEPC)-treated] water, and yields were determined by absorbance using an Eppendorf Biophotometer. Purified RNA was stored at −80°C. The amount of total RNA obtained was 2,711 μg (630 μg RNA from sample 1, 510 μg RNA from sample 2, and 571 μg RNA from sample 3) from a total of 2.65 g of buds (or approximately 1 mg RNA per gram of tissue).

mRNA isolation

Message RNA was extracted from total RNA according to the manufacturer's protocol (http://www.ambion.com/techlib/prot/bp_1916.pdf) for the Poly(A)Purist™ mRNA Purification Kit (Ambion, catalog # 1916). Approximately 500 μg of total RNA (after QC) was added to each of three Poly(A) Purist™ columns. After column purification, each aliquot of mRNA was ethanol-precipitated using the method provided with the kit (including glycogen). The RNA was then resuspended separately in THE™ RNA storage solution (Ambion). Resuspended mRNA was stored at −80°C. Quality and quantity of the mRNA were determined by micro-capillary electrophoresis on an Agilent Bioanalyzer, according to the manufacturer's suggested protocol. The mRNA was intact, had no detectable DNA contamination, and had less than 15% tRNA contamination. The amount of mRNA obtained was 32.9 μg, a yield of approximately 12.4 μg of mRNA per gram of bud tissues.

cDNA library construction

A directional cDNA library, designated Ltu01, was constructed from approximately 7 μg of mRNA using the ZAP-cDNA® Synthesis Kit (Stratagene), according to manufacturer's instructions (http://www.stratagene.com/manuals/200401.pdf), and following the modifications of Carlson et al. (2006). The pBluescript II SK(+/−) phagemid vector form of the library was excised from an aliquot of the lambda vector (into SOLR host cells) before library manipulation for DNA sequencing. The phagemid library was maintained under 100 mg/ml Ampicillin selection. The cDNA library construction yielded a primary library of approximately 3×10^6 total pfus and an amplified library with a titer of 2×10^{10} pfu per ml (with over 200 ml total volume). Monitoring of the library construction steps and gel electrophoresis were preformed nonradioactively using SYBR green stain. Analysis of the Ltu01 library by polymerase chain reaction (PCR) with M13F/M13R primers of 40 plasmids followed by agarose gel electrophoresis showed an average insert size of 1,346 bp. Approximately 50,000 colonies were picked from the excised Ltu01 library and replicated in triplicate using a QPix2® robot (Genetix) and then stored in 8% glycerol in 384-well plates at −80°C.

DNA sequencing

One replica of each set of plates was removed from storage at −80°C and entered into the sequencing queue. DNA template preparation and sequencing reactions were performed on 96-well plate format. DNA templates of cDNA inserts were prepared from overnight bacterial liquid cultures by rolling-circle amplification (RCA) of the pBluescript plasmids using TempliPhi™ DNA Amplification (Amersham) kits, following the manufacturer's protocol at one-quarter recommended volumes. Sequencing was conducted on RCA products using BigDye® Terminator v1.1 Cycle Sequencing Kits (Applied Biosystems) and T3 primer at one-sixth to one-quarter recommended reaction volumes. Sequencing reactions were purified using paramagnetic beads with CleanSEQ dye-terminator removal kits (Agencourt) according to the manufacturer's protocol but at one-half recommended volumes. Sequencing reaction products were automatically loaded and electrophoresed on an ABI PRISM® 3700 DNA Analyzer.

EST informatics

The Plant Genome Network at Cornell University provided automated processing, quality control, data archiving, unigene assembly, and library statistics as described in Carlson et al. (2006). The resulting EST database and unigene assemblies can be accessed at http://pgn.cornell.edu/library/lib_info.pl?lib_id = 14 and http://pgn.cornell.edu/unigene/unigene_info.pl?build_id=53, respectively. The *Liriodendron* sequences have been sorted into gene families which can be found in the PlantTribes database (http://www.floralgenome.org/tribe.php; Wall et al. 2007).

Repeat Masker Web Server (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) was used to detect simple sequence repeats (interspersed and simple repeats, $n \geq 20$).

In situ hybridization

Expression of the *LtuPNH* and *LtuFVE* genes were determined using terminal buds harvested from one yellow-poplar tree on the Pennsylvania State University campus on June 22, 2005. Fixation, dehydration, and clearing of samples were performed essentially according to Jackson (1991) with modifications as described in Zhang et al. (2005). Sections were made in 10 μm thickness. RNA probes were prepared using the DIG RNA labeling kit (Roche Diagnostics, Indianapolis, IN, USA) following the manufacturer's instructions. The *Ltu* Porin gene in sense and anti-sense orientations served as negative and positive controls, respectively. PCR primer pairs were designed such that a T7 promoter was introduced, and the transcription template was prepared by PCR (the T7 promoter sequence is given in lowercase letters in the primer sequences to follow). The *PINHEAD*-specific antisense primers used were PNH_F: 5′GCTACTGATTTGGGTATGAGGC3 and PNH_R: 5′taatacgactcactatagggAACTGACGCAAGT GAT3′. The *FVE*-specific antisense primers used were FVE_F: 5′ AATCACAACCTCGTCTGGCCTT3′ and FVE_R: 5′ taatacgactcactatagggACAAAGGGTGAGC GTG3′. The Porin-specific antisense primers used were Porin_F: 5′ TTTGACACTGCATCGGGCAACT3′ and Porin_R: 5′ taatacgactcactatagggATCTTCGCGCTCTTCT3′. The Porin-specific sense primers used were Porin_senF: 5′ taatacgactcactatagggTTTGACACTGCATCGG3′ and Porin_ senR: 5′ATCTTCGCGCTCTTCTCAATGG3′. Hybridization and detection steps were performed according to Drews et al. (1991) and Zhang et al. (2005).

Results

Single-pass sequencing of 12,937 cDNAs from *Liriodendron* floral bud library

A total of 9,531 high-quality ESTs were obtained from the sequencing of 12,937 cDNA inserts, resulting in a sequencing success rate of 74% (for stringency standards, see Carlson et al. 2006). The average length of ESTs was 466 bp. An assembly of overlapping, contiguous sequences yielded a total of 6,520 unigenes, of which 5,251 were singletons and 1,269 were in contigs (unigene build number 4, 2004-12-4, http://pgn.cornell.edu/). The average unigene length was 636 bp (singletons averaged 476 bp, and contigs averaged 797 bp). The average GC content for the Ltu01 data set was 45.0%.

The redundancy of sequences encountered by our random sequencing approached 55% when sequencing of the Ltu01 library was concluded. Overall, 81% of the sequences were unique. The two most abundant unigenes (one for a dormancy-associated protein and the other for a bulb-type mannose-specific lectin) were sampled 55 and 50 times (Table 1), respectively. The next 18 most abundant genes (Table 1) were sampled between 14 and 46 times. When compared to the proteome of *Arabidopsis*, it was found that unigenes with relative start sites at #0 and #1 had the highest frequencies, followed by unigenes with #2 and #3 relative start sites (Fig. 2). Relative start site was calculated by using this formula: (the position number of amino acid a unigene starts with in *Arabidopsis* homolog/ total number of amino acids in the *Arabidopsis* homolog)× 100. Unigenes with relative start site at #0 were likely to contain 5′ untranslated leader region (ULR) sequences. Since the cDNA library was constructed by oligo(dT) priming from the 3′-end of the mRNA, and 7.9 and 8.1% of the unigenes had relative start sites at the beginning of the coding sequence (positions #0 and #1), it was concluded that 16% of the unigenes contained full-length coding regions. Approximately 90% of the unigenes contained simple sequence repeats (SSR). The size of the SSRs ranged from dimers to hexamers, with dimers having the highest frequency (Fig. 3).

Codon usage

The codon usage in the Ltu01 translated EST sequences, generated by General Codon Usage Analysis (http://bioinf.may.ie/gcua/index.html; McInerney, 1998), is represented in Table 2. The pattern of codon preferences observed in Ltu01 dataset was more similar to *A. thaliana* than *Oryza sativa* (Nakamura et al. 2000; the Codon Use Database at http://www.kazusa.or.jp/codon/, GenBank Release 145.0, January 25, 2005), with four different preferred codons from *Arabidopsis* and 13 from rice. Only four amino acids exhibited G or C at the degenerate third base of their preferred codons. This is consistent with the fact that dicots do not favor G and C in that position. The preferred codons were marked with supescripted "a" in Table 2.

Classification of unigene sets to predicted functions

To identify *Liriodendron* unigenes that potentially encode homologs of known proteins, BLASTX (2.2.10) was conducted against the nonredundant protein database at GenBank. A total of 4,468 unigenes showed significant similarity (e value $\leq 1e-5$) with at least one published sequence (with known or unknown function), accounting for 69% of the 6,520 unigenes (see Supplementary Table S1). Approximately 13% of the protein homologs were

Table 1 The twenty most highly expressed *Liriodendron* unigenes in the Lt EST database (Ltu01 unigene build 4, http://pgn.cornell.edu/unigene/unigene_info.pl?build_id=53)

Unigene number	# ESTs (% total)	Contig length	*Arabidopsis* protein with best BLASTX match / Other species with strong BLASTX matches	BLASTX score
247536	55	811	Dormancy-associated protein (At1g28330);	3*e*−32
	(0.58)		Auxin-Repressed 12.5 kD Protein (*Fragaria* × *ananassa*)	6*e*−33
248480	50	675	No hit to *Arabidopsis* proteome;	–
	(0.52)		Bulb-type mannose-specific lectin (*Galanthus nivalis*)	8*e*−35
248418	46	764	No hit to *Arabidopsis* proteome;	–
	(0.48)		Dehydrin-like protein (*Prunus persica*)	9*e*−06
247464	45	1,863	Elongation factor 1-alpha (At5g60390);	0
	(0.47)		*Elaeis guineensis*	0
247669	30	1,256	Dehydrin Xero2 (At3g50970);	3*e*−06
	(0.31)		*Panax ginseng*	2*e*−09
247743	27	1,943	No hit to *Arabidopsis* proteome;	–
	(0.28)		*Cucumis sativus* male-specific 10-2 (M10-2) mRNA (BLASTN)	3*e*−15[a]
248102	26	675	No hit to *Arabidopsis* proteome;	–
	(0.27)		Hypothetical protein (*Oryza sativa* cv japonica)	4*e*−20
247943	25	845	Expressed protein (At5g48480);	6*e*−24
	(0.26)		Early tobacco anther 1 (*Nicotiana tabacum*)	6*e*−36
247512	25	1,958	Putative protein (At5g48390);	1*e*−127
	(0.26)		Putative ABI3-interacting protein 2 (*Oryza sativa*)	1*e*−156
248499	25	910	ADP-ribosylation factor 1 (At2g47170);	3*e*−99
	(0.26)		*Oryza sativa*	8*e*−99
247806	24	1,232	Cysteine proteinase RD19A (At4g39090);	1*e*−150
	(0.25)		*Sandersonia aurantiaca*	1*e*−154
247537	19	663	Putative protein, 68417.m03134 glycine-rich (At4g21620);	2*e*−12
	(0.20)		Expressed protein (*Oryza sativa*)	3*e*−10
248345	17	1,024	Putative porin (At3g01280);	1*e*−105
	(0.18)		Outer mitochondrial membrane protein porin, *Solanum tuberosum*	1*e*−126
247524	17	1,983	No hit to *Arabidopsis* proteome;	–
	(0.18)		*Cucumis sativus* male-specific 10-2 (M10-2) mRNA (BLASTN)	3*e*−15[a]
248022	16	1,520	Elongation factor 1-alpha (At5g60390);	0
	(0.17)		*Elaeis guineensis*	0
247337	16	1,297	Quinone oxidoreductase (At5g16990);	1*e*−125
	(0.17)		Allyl alcohol dehydrogenase (*Nicotiana tabacum*)	1*e*−122
247740	15	1,088	Ribonuclease, RNS3 (At1g26820);	3*e*−38
	(0.16)		*Hordeum vulgare*	1*e*−49
247988	15	703	Translationally controlled tumor protein (At3g16640);	1*e*−71
	(0.16)		*Hevea brasiliensis*	4*e*−77
247590	14	738	No hit to *Arabidopsis* proteome;	–
	(0.15)		No hits to GenBank (BLASTN or BLASTX)	–
247760	14	674	60S ribosomal protein L18A (At2g34480);	1*e*−89
	(0.15)		*Oryza sativa*	6*e*−91

[a] *e* Value from BLASTN

annotated as unknown, hypothetical, or expressed proteins. Detailed functional annotation of the unigenes was obtained as Gene Ontology Consortium structure according to Putative Cellular Components, Putative Biological Processes, and Putative Molecular Functions (Fig. 4), respectively. As seen in the pie charts of Fig. 4, a wide variety of putative functions are represented in the Ltu01 EST database. It is noteworthy that approximately 2.7% of the unigenes encoded proteins with putative transcription factor activity.

When aligned against nucleotide collections in GenBank (BLASTN 2.2.16), the sequences without significant matches (*e* value >1*e*−5) in BLASTX resulted in an additional 963 significant alignments (*e* value ≤1*e*−5; Supplementary Table S2). Therefore, a total of 5,431 unigenes had significant similarity to publicly available sequences, accounting for 83% of the total unigenes. Approximately 66% of the best-hit alignments were to eudicot sequences, while 31% were to monocots, 1% to Magnoliales, and less than 1% to conifer sequences.

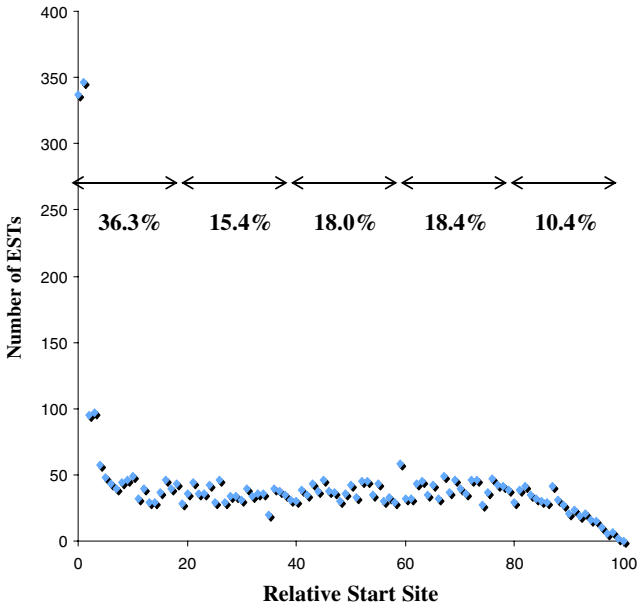

Fig. 2 Relative start site distribution for Ltu01 unigenes. The relative start sites are numbered in bp beginning with #0 in the first codon, as predicted from the *Arabidopsis* proteome

Possible contamination (best hits with microorganism and animal sequences) was found to be approximately 1% of the 5,431 hit unigenes. The ESTs from *Liriodendron* floral buds have also revealed a large panel of novel plant genes. A total of 1,089 unigenes (approximately 17% of the 6,520 unigenes discovered) did not match with any sequence available in GenBank, despite inclusion of the complete genome sequences of *Arabidopsis*, rice, and *Populus*. Some of these sequences may be novel genes, while others represent sequences from portions of cDNAs such as untranslated regions or nonconserved areas of protein where homology is not detected (Wang et al. 2004). Novel genes in *Liriodendron* could be unique in this species, or could be found in other magnoliid or basal angiosperm species; when better characterized, these genes may be informative for understanding the early divergene of flowering plants and the origins of diverged gene families.

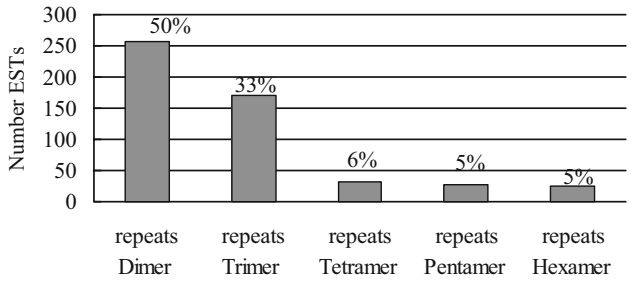

Fig. 3 Frequency of simple sequence repeats (SSR) in the Ltu01 unigene set

Highly expressed genes in *Liriodendron* floral buds

The number of ESTs assembled in the contigs gives an indication of the degree of expression of the respective gene in premeiotic immature floral tissues. The twenty most highly expressed *Liriodendron* unigenes are listed in Table 1. The top 20 unigenes accounted for almost 8% of the 9,531 high-quality sequences obtained. Comprised of 55 overlapping ESTs (0.84% of the total), the most highly expressed gene (unigene 247536) showed a high level of sequence similarity (*e* value=3*e*−32) to a dormancy-associated protein in *Arabidopsis*. There were also two putative homologs of dehydrin protein (unigenes 248418 and 247669) among the top five and one homolog of mitochondrial porin protein (unigene 248345) among the top 20 most-expressed unigenes. Dehydrin and porin

Table 2 Cumulative codon usage in Ltu01

AA Codon RSCU	*N* RSCU	AA Codon	*N*
Phe UUU	16,474 (0.97)	Ser UCU	17,905 (1.50)[a]
UUC	17,658 (1.03)[a]	UCC	11,119 (0.93)
Leu UUA	6,922 (0.54)	UCA	14,275 (1.20)
UUG	16,950 (1.33)	UCG	6,989 (0.59)
Tyr UAU	12,073 (1.06)[a]	Cys UGU	7,010 (0.96)
UAC	10,600 (0.94)	UGC	7,609 (1.04)[a]
ter UAA	1,108 (0.00)	ter UGA	1,562 (0.00)[a]
ter UAG	682 (0.00)	Trp UGG	11,020 (1.00)
Leu CUU	18,736 (1.47)[a]	Pro CCU	16,535 (1.38)[a]
CUC	14,129 (1.11)	CCC	8,532 (0.71)
CUA	7,349 (0.58)	CCA	16,512 (1.37)
CUG	12,460 (0.98)	CCG	6,490 (0.54)
His CAU	13,467 (1.24)[a]	Arg CGU	6,403 (0.84)
CAC	8,193 (0.76)	CGC	4,724 (0.62)
Gln CAA	14,950 (0.94)	CGA	5,065 (0.67)
CAG	16,803 (1.06)[a]	CGG	5,544 (0.73)
Ile AUU	18,466 (1.23)[a]	Thr ACU	12,839 (1.23)
AUC	16,663 (1.11)	ACC	10,118 (0.97)
AUA	9,910 (0.66)	ACA	13,122 (1.25)[a]
Met AUG	20,941 (1.00)	ACG	5,811 (0.55)
Asn AAU	19,706 (1.15)[a]	Ser AGU	9,976 (0.84)
AAC	14,648 (0.85)	AGC	11,119 (0.93)
Lys AAA	22,234 (0.85)	Arg AGA	12,104 (1.59)[a]
AAG	29,997 (1.15)[a]	AGG	11,743 (1.55)
Val GUU	19,096 (1.40)[a]	Ala GCU	23,582 (1.45)[a]
GUC	12,497 (0.91)	GCC	13,554 (0.84)
GUA	7,707 (0.56)	GCA	20,294 (1.25)
GUG	15,345 (1.12)	GCG	7,454 (0.46)
Asp GAU	30,624 (1.33)[a]	Gly GGU	16,479 (1.08)
GAC	15,559 (0.67)	GGC	12,828 (0.84)
Glu GAA	27,183 (0.99)	GGA	18,384 (1.20)[a]
GAG	27,661 (1.01)[a]	GGG	13,411 (0.88)

[a] The most frequently used codon;
N Sum of the frequencies of the codon; *RSCU* relative synomymous codon usage

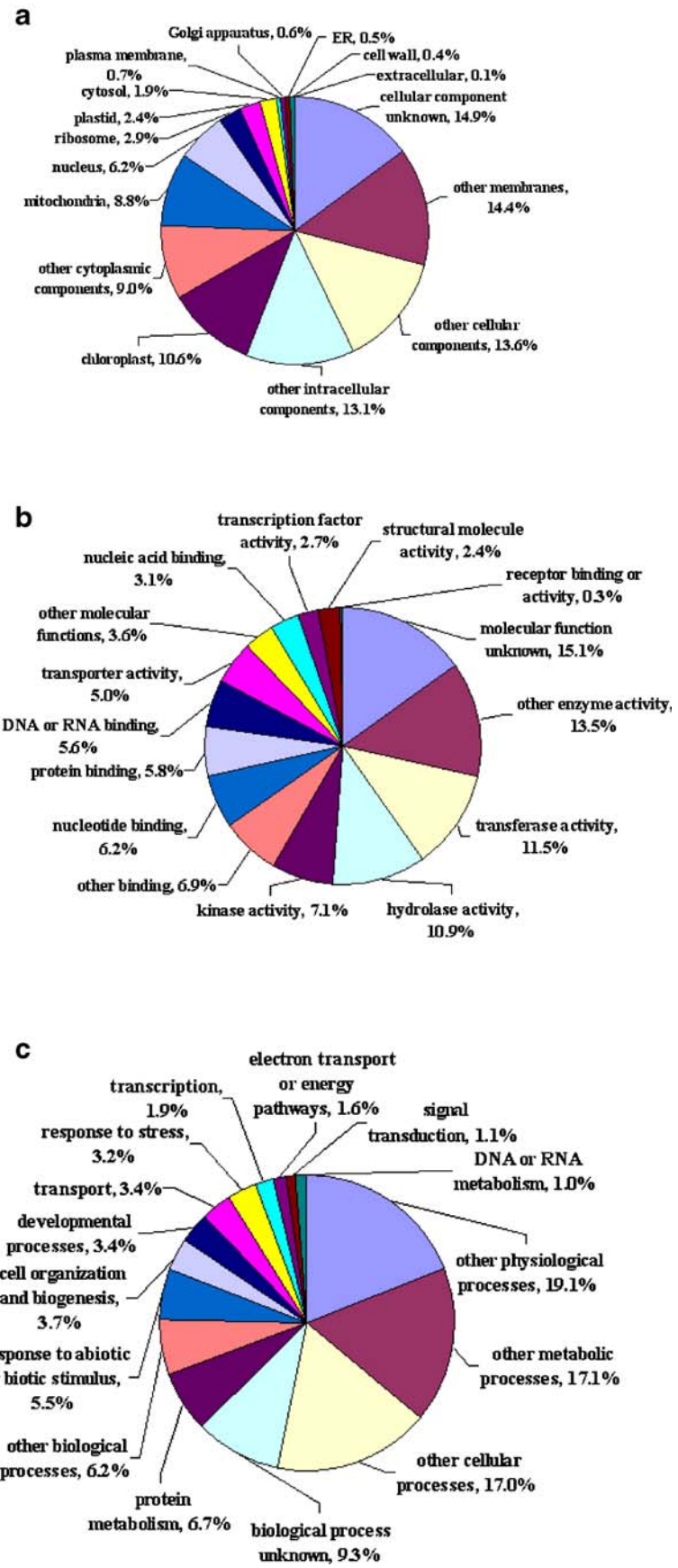

Fig. 4 Pie chart representation of GO-annotation classification of *Liriodendron* ESTs functions. **a** Putative cellular components; **b** Putative molecular functions; **c** Putative biological processes

Table 3 Floral gene family sequences identified in the Ltu01 database

Gene family name	Number of genes in			Number in Lut01		% of Ltu01 "transcriptome"
	ATH[a]	OSA[b]	PTR[c]	Unigenes	ESTs	
ABI1-like phosphatase	36	40	58	9	10	0.11
AP2	18	25	31	4	10	0.11
AP2-domain, other	54	51	67	5	20	0.21
ARGONAUTE	10	24	16	8	10	0.11
ASK1/SKP1	19	26	10	3	6	0.06
AUX/IAA proteins	25	28	32	6	8	0.08
Auxin response factors (ARFs)	23	28	44	14	20	0.21
BEL1-like KNOX homeodomain	13	14	20	1	1	0.01
B-Zip, Perianthia-like	10	18	14	2	2	0.02
B-Zip domain proteins, other	6	4	6	1	1	0.01
CLV1-like receptor-like protein kinases	712	1,252	1,565	85	114	1.20
CPS-like	1	4	2	1	1	0.01
Cullin	7	9	12	9	21	0.22
DEAD Box RNA helicases, CAF-like	6	8	7	2	2	0.02
DEAD-Box RNA helicases, other	54	54	78	19	37	0.39
DIVARICATA-like mybs	20	19	24	1	1	0.01
EIN3/EIL-like trans regulator	6	7	6	2	5	0.05
Elongation Factor	3	6	6	1	1	0.01
EXPANSIN	30	51	37	4	4	0.04
GIGANTEA	1	1	2	4	4	0.04
GL1-like MYB	132	117	210	7	7	0.07
GRAS-GAI, RGA, and SCARECROW	27	43	73	2	2	0.02
FVE	1	1	2	1	7	0.07
HEN1	2	1	2	1	1	0.01
HEN2	3	3	6	2	2	0.02
HEN4	8	5	14	3	3	0.03
Homeobox-leucine zipper	17	13	18	4	4	0.04
HUA1	12	8	11	1	1	0.01
ketoacyl-CoA synthase	21	27	36	5	8	0.08
KNOX family class 2 homeodomain	4	11	9	2	2	0.02
MADS-box	45	40	72	7	13	0.14
NAM/NAC	83	85	133	2	2	0.02
Phabulosa-like	5	10	10	7	14	0.15
Phytochrome	5	3	3	4	4	0.04
PINOID	22	26	25	6	6	0.06
Set Polycomb Zinc finger	3	2	4	1	1	0.01
Shaggy-like protein kinase	87	87	103	25	34	0.36
short-chain ADH	12	22	33	1	1	0.01
SPL	16	16	31	7	15	0.16
SPLAYED	17	20	36	7	9	0.09
SUESS	4	3	5	1	1	0.01
Tousled-like protein kinase	167	154	207	24	28	0.297
TSO1-like Transcription Factor	4	5	5	2	2	0.02
WD40 repeat, LEUNIG	2	6	6	3	5	0.05
WD40 repeat, MSI3-like	2	1	4	1	7	0.07
WD40 repeat, other	1	1	2	1	1	0.01
WRKY TFs	52	54	88	5	1	0.01
YABBY	5	7	13	3	4	0.04
ZF-HD family	15	11	21	2	2	0.02
Totals	1,828	2,451	3,219	318	465	4.88

[a] *Arabidopsis thaliana*
[b] *Oryza sativa*
[c] *Populus trichocarpa*

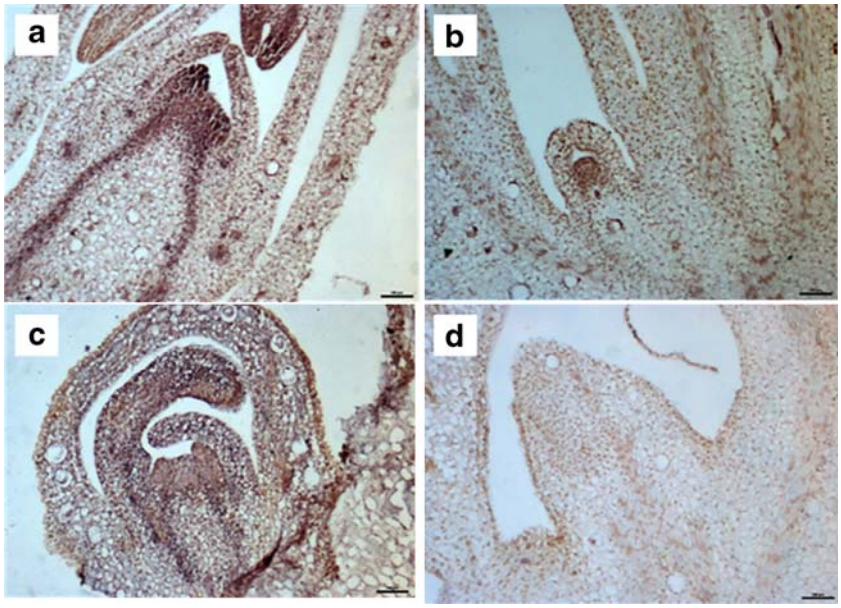

Fig. 5 Expression of the *Pinhead* and *FVE* flowering time genes in developing *Liriodendron* terminal buds. Bright field image results of *in situ* hybridization and colorimetric detection of **a** the *LtuPNH* transcripts; **b** Ltu*FVE* transcripts; **c** the outer mitochondrial membrane protein porin gene transcripts (positive control); and **d** outer mitochondrial membrane protein porin sense sequence (negative control). *Bar* 200 μm

proteins are also associated with dormancy. In total, these four unigenes accounted for 2.27% of the high-quality sequences. This finding is not entirely unexpected, as the majority of the floral buds, from which the cDNA library was made, were harvested in October and November when the buds were becoming cold acclimated and entering dormancy, before overwintering and then maturing and opening the following spring.

The eighth most highly expressed unigene (247943) was similar to a flower-specific gene *ETA1* (early tobacco anther 1, *e* value=6*e*−36). This was the first putative *ETA1* homolog found in *Liriodendron*. Proteins with functions in basic cellular processes, development and/or differentiation were also among the top hits, including lectin, elongation factor 1-alpha (two homologs), ADP-ribosylation factor 1, cysteine proteinase, and translationally controlled tumor protein (which in mammalian systems acts as a guanine nucleotide dissociation inhibitor on the translation elongation factor 1-alpha [Cans et al. 2003]). Unigenes 247743 and 247524 did not have any match in *Arabidopsis* proteome and protein database in GenBank. However, BLASTN search of these two sequences in GenBank resulted in significant alignment to a *Cucumis sativus* male-specific 10-2 (M10-2) mRNA (*e* value=3*e*−15). Among the top 20 highly expressed unigenes, consensus sequence 247590 did not match to any public sequences (BLASTN or BLASTX).

Flower-related genes expressed in *Liriodendron* floral buds

Among the 6,520 unigenes obtained from the Ltu01 floral bud database, 318 showed sequence similarity to genes known to be involved in flower development from *Arabidopsis* and rice (Table 3). These 318 unigenes represented 49 gene families or subfamilies that include members known to function in floral development, over half of such families (Zhao et al. 2001; Ma 2005; Zahn et al. 2006), and accounted for 4.9% of the high-quality ESTs. In most cases, the Ltu01 ESTs were the first evidence for these floral genes in *Liriodendron*. The floral genes were observed at levels of 0.01 to 1.20% of the sampled transcriptome. The proteins encoded by relatively highly expressed floral genes (greater than 0.1% of the ESTs) in *Liriodendron* included *CLV1*-like receptor-like protein kinases, *DEAD*-Box RNA helicases, shaggy-like protein kinases, a tousled-like kinase, cullin, *APETALA2* (*AP2*), other AP2-domain proteins, auxin response factors (ARFs), Squamosa-Promoter-Binding-like proteins, Phabulosa-like proteins, MADS-box proteins, *ARGONAUTE*, and *ABI1*-like phosphatase. However, most genes were observed in relatively low abundance (<0.1% of sequenced cDNA inserts). Also, there were 40 known floral gene families (Zhao et al. 2001; Ma 2005; Zahn et al. 2006) not detected in the *Liriodendron* EST data set, including *FLORICAULA/LEAFY* homolog, *FRIGIDA*-like hydroxyproline-rich glycoprotein, *CO*-like Zinc finger protein, and B-Box Zinc finger proteins. Since most of the missing floral genes are expressed at low levels in *Arabidopsis*, it is possible that their expression levels were also low in *Liriodendron*. Also, some floral homologs might not be transcribed in the bud stages used for the Ltu01 library construction. For instance, the flower meristem identity gene *FLORICAULA/LEAFY* may have stopped expressing by autumn when buds were harvested. In addition, over half of the missing gene

Table 4 Lignin biosynthesis genes identified in Ltu01

Gene	Contig # (length)	Best BLASTX hit species (GI number and *e* value)	Number of orthologs and close paralogs	
			Arabidopsis	*Oryza*
Phenylalanine ammonia-lyase (PAL)	249,339 (190 bp)	*Lycopodium tristachyum* (gi:58618140,05e25)	3	4
	250,674 (155 bp)	*Camellia sinensis* (51594297,1*e*−21)		
	251,239 (523 bp)	*Persea americana* (gi:1171999, 3*e*−79)		
Sinapyl alcohol dehydrogenase (SAD)	253,593 (194 bp)	*Medicago truncatula* (gi:92887576, 6*e*−11)	1	3
Caffeoyl-CoA *O*-methyltransferase (CCoAOMT)	248,434 (654 bp)	*Isatis tinctoria* (gi:74053616, 4*e*−87)	1	1
Cinnamoyl-CoA reductase (CCR)	249,484 (430 bp)	*Arabidopsis thaliana* (gi:15226134, 1*e*−33)	2	2
	251,716 (546 bp)	*Acacia mangium* × *Acacia auriculiformis* (gi:68159360, 4*e*−63)		
	251,771 (290 bp)	*Linum album* (gi:57282092, 2*e*−26)		
Caffeic acid *O*-methyltransferase (COMT)	249,797 (496 bp)	*Thalictrum tuberosum* (4808524, 1*e*−76)	1	1
	248,159 (516 bp)	*Medicago truncatula* (gi:92890207, 4*e*−35)		
Cinnamic acid 4-hydroxylase (C4H)	248,866 (658 bp)	*Cucumis sativus* (gi:109715482, 7*e*−47)	1	1
	248,941 (227 bp)	*Arabidopsis thaliana* (gi:1773287, 1*e*−33)		
4-Coumarate-CoA ligase (4CL)	249,731 (367 bp)	*Glycine max* (gi:18266852, 1*e*−54)	1	2
	250,082 (162 bp)	*Oryza sativa* (gi:56784511, 1*e*−15)		
Cinnamyl alcohol dehydrogenase (CAD)	249,253 (481 bp)	*Striga asiatica* (gi:109631192, 9*e*−71)	2	1
p-Hydroxycinnamoyl-CoA: shikimate/quinate p-hydroxycinnamoyltransferase (HCT)	252,332 (593 bp)	*Coffea arabica* (gi:116486991, 6*e*−54)	1	2
p-Coumarate 3-hydroxylase (C3H)[a]	247,690 (742 bp)	*Pinus taeda* (gi:54634267, 1*e*−46)	1	1
	248,232 (1,020 bp)	*Pinus taeda* (gi:54634267, 2*e*−43)		
	247,404 (685 bp)	*Pinus taeda* (gi:17978651, 2*e*−24)		
	247,606 (607 bp)	*Pinus taeda* (gi:17978651, 7*e*−24)		
	251,521 (452 bp)	*Pinus taeda* (gi:17978651, 1*e*−19)		
	248,941 (227 bp)	*Sesamum indicum* (gi:17978831, 3*e*−11)		
	249,703 (599 bp)	*Sesamum indicum* (gi:17978831, 5*e*−08)		
	252,870 (469 bp)	*Sesamum indicum* (gi:17978831, 5*e*−07)		
	249,148 (640 bp)	*Sesamum indicum* (gi:17978831, 1*e*−05)		
Ferulate 5-hydroxylase (F5H)[a]	248,232 (1,020 bp)	*Lycopersicon esculentum* × *Lycopersicon peruvianum* (gi:5002354, 5*e*−56)	2	1

Table 4 (continued)

Gene	Contig # (length)	Best BLASTX hit species (GI number and *e* value)	Number of orthologs and close paralogs	
			Arabidopsis	*Oryza*
	247,690 (742 bp)	*Lycopersicon esculentum* × *Lycopersicon peruvianum* (gi:5002354, 1*e*−50)		
	247,404 (685 bp)	*Arabidopsis thaliana* (gi:12578901, 6*e*−33)		
	247,606 (607 bp)	*Arabidopsis thaliana* (gi:12578901, 5*e*−24)		
	251,521 (452 bp)	*Medicago sativa* (gi:77744233, 1*e*−21)		
	251,533 (553 bp)	*Arabidopsis thaliana* (gi:45535205, 3*e*−17)		
	248,780 (353 bp)	*Broussonetia papyrifera* (gi:57470995, 1*e*−13)		
	249,148 (640 bp)	*Arabidopsis thaliana* (gi:12578901, 3*e*−11)		
	248,941 (227 bp)	*Oryza sativa* (gi:10140722, 5*e*−10)		
	249,703 (599 bp)	*Lycopersicon esculentum* × *Lycopersicon peruvianum* (gi:5002354, 1*e*−08)		
	252,870 (469 bp)	*Camptotheca acuminate* (gi:47933890, 4*e*−06)		

[a] Not best hits

families have only one or two copies in *Arabidopsis* and rice, suggesting that *Liriodendron* might have few paralogs, making them less likely to detect than those families with many members.

As examples, *in situ* hybridization was conducted with two *Liriodendron* floral genes: *LtuPNH* (unigene# 252439, BLASTX *e* value=2*e*−75) and *LtuFVE* (unigene# 247528, BLASTX *e* value=1*e*−153; Table 3 and Supplementary Table S1; Fig. 5). The *Arabidopsis PNH* gene (a member of the *Argonaut* floral gene family) is involved in establishing determinate versus indeterminate growth during vegetative and floral development and is expressed in the shoot apical meristem and the upper sides of lateral organ primordia (Lynn et al. 1999; Newman et al. 2002), while the *FVE* gene has a dual role in regulating the flowering time and cold response (Kim et al. 2004). Consistent with the *Arabidopsis PNH* gene, expression of *LtuPNH* was largely located in the meristem, leaf primordia, and vasculature. Expression of the *LtuFVE* gene was strong in meristematic cells but not detectable in organ primordia. These results suggest that *LtuPNH* and *LtuFVE* are indeed involved in *Liriodendron* flower development. The *in situ* results also demonstrate the value of the Ltu01 EST database as a source of candidate genes for the analysis of the role of gene expression in flower development in *Liriodendron*.

Cell-wall biosynthesis genes expressed in *Liriodendron* floral buds

Many of the genes known to be involved in cell-wall construction were represented among the 6,520 Ltu01 unigenes. The set of cell-wall biosynthesis genes in Ltu01 includes 27 unigenes for the lignin biosynthesis pathway (Table 4), six unigenes for cellulose synthase, three unigenes for pectinacetylesterase, one unigene for pectinesterase, two unigenes for high-pI laccase, six unigenes for peroxidase, and five expansin unigenes (Table 4 and Supplementary Table S3).

The lignin biosynthesis pathway has been studied intensively in trees (reviewed by Boerjan et al. 2003) due to the importance of lignin extraction in the pulp and paper industry. The biochemical pathway for lignin biosynthesis has undergone revisions in the past 10 years based on information from molecular genetics and genomics studies with model plant systems. Genomic information from a taxonomically basal organism such as *Liriodendron* will permit broader conclusions to be drawn about the structure of the lignin pathway. Among the 11 genes involved in the lignin synthesis pathway from phenylalanine ammonia-lyase (PAL) to cinnamyl alcohol dehydrogenase (CAD), nine putative orthologs were identified in the Ltu01 EST

database as defined by best hits by BLASTX alignment (Table 4). The BLASTX search did not result in "best hits" (top alignment scores) for ferulate 5-hydroxylase (F5H) and *p*-coumarate 3-hydroxylase (C3H). F5H and C3H are members of the large p450 superfamily of enzymes. Several *Liriondendron* members of a p450 family related to F5H or C3H were detected in the EST data set, although it is not clear whether they represent orthologs of known F5H and C3H genes. In addition, the Ltu01 EST data set produced six unigenes encoding laccases and two peroxidase unigenes, revealing that *Liriodendron* buds express the genes required for lignin polymerization, as well as lignin monomer synthesis (Supplementary Table S3).

Discussion

Expressed sequence tags (ESTs) are an important tool in functional genomics and have been a valid and reliable resource that can accelerate gene discovery (Somerville and Somerville 1999). The reported *Liriodendron* EST database contains a total of 9,531 high-quality ESTs. These sequences clustered into 6,520 unigenes, including 5,251 singletons and 1,269 contigs. Virtually all of the genes detected in this study were the first representatives of their respective gene family for *Liriodendron*, since there were only six entries for nuclear gene nucleotide sequences in GeneBank for *Liriodendron* before this project. In addition, this database has detected as many as 1,089 novel genes. This is a major boost for genomic-scale resources for *Liriodendron* and should draw greater interest in comparative genomics with basal angiosperms and in applying genomic approaches to improve economic properties of *Liriodendron*.

The *Liriodendron* EST database revealed putative members of over half of known floral gene families or subfamilies (Zhao et al. 2001; Ma 2005; Zahn et al. 2006), accounting for 4.8% of the ESTs and 317 unigenes. These results were quite similar to the floral EST database generated for the basal eudicot species California poppy by the Floral Genome Project (Carlson et al. 2006). Homologs of genes regulating many aspects of flower development were among the genes identified, including those for organ identity and development, cell and tissue differentiation, cell cycle control, and secondary metabolism. To date, the *Liriodendron* EST sequences for homologs of *SEPALLATA*, *DEFICIENS* and *GLOBOSA* (Zahn et al. 2005a, b), *SKP1* (Kong et al. 2007), and *GLYCOGEN SYNTHASE KINASE* (Yoo et al. 2006) homologs have been used in reconstruction of phylogenetic trees. The cDNA clones tagged in this study are available for further study on request at www.floralgenome.org, and the codon usage table developed from the EST data should aid scientists who attempt to amplify new genes from *Liriodendron* with PCR or reverse transcription-PCR, as was the case for a species-specific codon table in *Eschscholzia californica* (Carlson et al. 2006; Annette Becker, via personal communication to CWD).

The *Liriodendron* Ltu01 EST data set was based on a large, nonnormalized and nonsubtracted cDNA library constructed from premeiotic and meiotic floral buds. Avoiding normalization and subtraction of the library resulted in a redundancy rate of approximately 55% among the last sets of sequences generated. Our goal was to obtain an unbiased view of the early developmental floral transcriptome, which proved to be highly complex and possessed many copies of related paralogous genes. This approach yielded many novel gene sequences for a wide array of processes and an indication of the levels of expression of these genes in developing floral buds. Besides floral genes, this EST database has revealed many homologs of genes involved in other biological processes. For example, at least 9 of the 11 known genes in the lignin biosynthesis pathway have been identified in *Liriodendron* from the Ltu01 data set, including cinnamyl alcohol dehydrogenase (*CAD*), 4-coumarate:CoA ligase (*4CL*), and phenylalanine ammonia-lyase (*PAL*)). Using the EST sequences as probes, these three genes have been successfully cloned from a large-insert BAC library for *L. tulipifera* (Liang et al. 2007). The *Liriodendron* Ltu01 EST collection thus provides a resource not just for molecular study of flower development but also for other important biological processes.

This EST data set has also provided the opportunity to examine the genome organization of a basal angiosperm. The average GC content for the EST data set was 45.0%, which is similar to GC contents reported in the *Liriodendron* BAC and shotgun end sequencing data sets (42 and 41%, respectively; Liang et al. 2007). This indicates that *Liriodendron* has a relatively AT-rich genome. BLAST results of the EST unigenes also support the findings from *Liriodendron* BAC and shotgun end sequencing data sets that *Liriodendron* genome sequence aligned better to eudicot sequences than to monocot sequences. Cui et al. (2006) conducted K_s analyses of 92 paralogous pairs identified in the Ltu01 EST set and showed evidence for two rounds of ancient genome duplication in the lineage leading to *Liriodendron*. The *Liriodendron* EST data have already enabled the development of SSR markers that will be valuable tools for research on population genetics and mating-system analysis, conservation, and genome mapping (Xu et al. 2006).

Acknowledgements We thank David Oppenheimer and his lab for guidance on the *in situ* hybridization experiments. We thank the tree-improvement field crew at the University of Tennessee at Knoxville for collecting, freezing, and shipping the tissue samples used for RNA

extraction. We thank Sheila Plock for assistance with manipulations and curation of the cDNA library, and Marlin Druckenmiller in the Schatz Center for Tree Molecular Genetics at Penn State for assistance in high throughput sequencing. The Floral Genome Project was supported by a grant to C. dePamphilis and co-PIs from the NSF Plant Genome Research Program (DBI-0115684). Dr. Liang was supported by a post-doctoral fellowship from The Schatz Center for Tree Molecular Genetics.

References

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Ann Rev Plant Biol 54:519–549

Buzgo M, Liang H, Schlarbaum SE, DiLoreto DS, Carlson J, Soltis PS, Soltis DE (2006) Floral development in *Liriodendron tulipifera* L. (Magnoliaceae) and other basal angoisperms. In: Botany 2006, California State University, Chico, CA, USA, July 28–August 3, 2006; Presentation 617

Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson JE, dePamphilis CW, Boore JL, Jansen RK (2006) Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. BMC Evol Biol 6:77

Cans C, Passer BJ, Shalak V, Nancy-Portebois V, Crible V, Amzallag N, Allanic D, Tufino R, Argentini M, Moras D, Fiucci G, Goud B, Mirande M, Amson R, Telerman A (2003) Translationally controlled tumor protein acts as a guanine nucleotide dissociation inhibitor on the translation elongation factor eEF1A. Proc Natl Acad Sci U S A 100:13892–13897

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

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW (2006) Widespread genome duplications throughout the history of flowering plants. Genome Res 16:738–749

Drews GN, Bowman JL, Meyerowitz EM (1991) Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. Cell 65:991−1002

Endress PK, Igersheim A (2000) Gynoecium structure and evolution in basal angiosperms. Intl J Plant Sci 161:S211–S223

Glasser WG, Kaar WE, Jain RK, Sealey JE (2000) Isolation options for non-cellulosic heteropolysaccharides (HetPS). Cellulose 7:299–317

Hayward TK, Hamilton J, Tholudur A, Mc Millan JD (2000) Improvements in titer, productivity, and yield using solka-floc for cellulase production. Appl Biochem and Biotech 84–86:859–874

Hunt D (1998) Magnolias and their allies. International Dendrology Society & Magnolia Society, England

Jackson DP (1991) *In situ* hybridization in plants. In: Bowles DJ, Gurr SJ, McPhereson M (eds) Molecular Plant Pathology: A Practical Approach. Oxford University Press, Oxford, pp 163–174

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 J, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 chloroplast genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA (in press)

Kim HJ, Hyun Y, Park JY, Park MJ, Park MK, Kim MD, Kim HJ, Lee MH, Moon J, Lee I, Kim J (2004) A genetic link between cold responses and flowering time through *FVE* in *Arabidopsis thaliana*. Nat Genet 36:167–171

Kim KH, Tucker MP, Keller FA, Aden A, Nguyen QA (2001) Continuous countercurrent extraction of hemicellulose from pretreated wood residues. Appl Biochem Biotechnol 91/93:253–268

Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, dePamphilis CW (2007) Patterns of gene duplication in the plant *SKP1* gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. Plant J 50:873–885

Liang H, Fang EG, Tomkins JP, Luo M, Kudrna D, Arumuganathan K, Zhao S, Schlarbaum SE, Banks JA, dePamphilis CW, Mandoli DF, Wing RA, Carlson JE (2007) Development of a BAC library resource for yellow poplar *(Liriodendron tulipifera*) and the identification of genomic regions associated with flower development and lignin biosynthesis. Tree Genet Genomes 3:215–225

Lynn K, Fernandez A, Aida M, Sedbrook J, Tasaka M, Masson P, Barton MK (1999) The *PINHEAD/ZWILLE* gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the *ARGONAUTE1* gene. Development 126:469–481

Kim SB, Lee YY (2002) Diffusion of sulfuric acid within lignocellulosic biomass particles and its impact on dilute-acid pretreatment. Bioresour Technol 83:165–71

Ma H (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Annu Rev Plant Bio 56:393–434

Merkle SA (2006) Engineering forest trees with heavy metal resistance genes. Silvae Genet 55:263–268

Nagle NJ, Elander RT, Newman MM, Rohrback BT, Ruiz RO, Torget RW (2002) Efficacy of a hot washing process for pretreated yellow poplar to enhance bioethanol production. Biotechnol Prog 18:734–738

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucl Acids Res 28:292

Newman KL, Fernandez AG, Barton MK (2002) Regulation of Axis Determinacy by the *Arabidopsis PINHEAD*. Gene Plant Cell 14:3029–3042

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

Ronse de Craene L, Soltis DE, Soltis PS (2003) Evolution of floral structures in basal angiosperms. Intl J Plant Sci 164:S329–363

Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in *Arabidopsis*. Plant Cell 2:755–67

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL* and *atpB* sequences. Bot J Linn Soc 133:381–461

Soltis DE, Soltis PS, Albert VA, Oppenheimer DG, dePamphilis CW, Ma H, Frohlich MW, Theissen G (2002) Missing links: the genetic architecture of flower and floral diversification. Trends Plant Sci 7:22–31

Soltis DE, Soltis PS, Chase MW, Endress P (2005) Phylogeny, evolution, and classification of flowering plants. Sinauer Associates, Sunderland, MA

Somerville C, Somerville S (1999) Plant functional genomics. Science 285:380–383

Vavelidis M, Melfos V, Boboti-Tsitlakidis I (1997) A new occurrence of placer gold in the Marathousa area (central Macedonia, N. Greece). In: Papunen H (ed) Mineral deposits: research and exploration, where do they meet? Proceedings of the Fourth Biennial SGA

Meeting, Turku, Finland, 11–13 August, 1997. A.A. Balkema, Rotterdam, pp 339–342

Wall PK, Leebens-Mack JH, Muller K, Field D, Altman NS, dePamphilis CW (2007) PlantTribes: A gene and gene family resource for comparative genomics in plants. Nucl Acids Res (in press)

Wang J-PZ, Lindsay BG, Leebens-Mack J, Cui L, Wall K, Miller WC, dePamphilis CW (2004) EST clustering error evaluation and correction. Bioinformatics 20:2973–2984

Wen J (1999) Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. Annu Rev Ecol Syst 30:421–455

Wilde HD, Meagher RB, Merkle SA (1992) Expression of foreign genes in transgenic yellow-poplar plants. Plant Physiol 98:114–120

Xu M, Ruddall P (2006) Comparative floral anatomy and ontogeny in Magnoliaceae. Plant Syst Evol 258:1–15

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

Yoo MJ, Albert VA, Soltis PS, Soltis DE (2006) Phylogenetic diversification of glycogen synthase kinase *3/SHAGGY*-like kinase genes in plants. BMC Plant Biol 6:3

Zahn LM, Kong H, Leebens-Mack JH, Kim S, Soltis PS, Landherr LL, Soltis D, dePamphilis CW, Ma H (2005a) 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, dePamphilis CW, Ma H, Theissen G (2005b) To B or Not to B a flower: the role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. J Heredity 96:225–240

Zahn LM, Feng B, Fu MH (2006) Beyond the ABC model: Regulation of floral homeotic genes. In: Soltis DE, Soltis PS, Leebens-Mack JH (eds) Developmental Genetics of the Flower, Advances in Botanical Research. vol. 44. Elsevier, London, pp 163–207

Zhang X, Dyachok J, Krishnakumar S, Smith LG, Oppenheimer DG (2005) *IRREGULAR TRICHOME BRANCH1* in *Arabidopsis* encodes a plant homolog of the actin-related protein2/3 complex activator Scar/WAVE that regulates actin and microtubule organization. Plant Cell 17:2314–2326

Zhao D, Yu Q, Chen C, Ma H (2001) Genetic control of reproductive meristems. In: McManus MT, Veit B (eds) Meristematic Tissues in Plant Growth and Development. Sheffield Academic Press, pp 89–142