ORIGINAL PAPER

Development of a BAC library for yellow-poplar (*Liriodendron tulipifera*) and the identification of genes associated with flower development and lignin biosynthesis

Haiying Liang · Eric G. Fang · Jeffrey P. Tomkins ·
Meizhong Luo · David Kudrna · Hye Ran Kim ·
K. Arumuganathan · Shaying Zhao ·
James Leebens-Mack · Scott E. Schlarbaum ·
Jo Ann Banks · Claude W. dePamphilis ·
Dina F. Mandoli · Rod A. Wing · John E. Carlson

Abstract *Liriodendron tulipifera* L., a member of the Magnoliaceae, occupies an important phylogenetic position as a basal angiosperm that has retained numerous putatively ancestral morphological characters, and thus has often been used in studies of the evolution of flowering plants and of specific gene families. However,

Communicated by J. Dean

H. Liang · J. E. Carlson
School of Forest Resources, Pennsylvania State University,
University Park, PA 16802, USA

H. Liang · C. W. dePamphilis · J. E. Carlson (✉)
Huck Institutes of the Life Sciences,
Pennsylvania State University,
University Park, PA 16802, USA
e-mail: jec16@psu.edu

E. G. Fang · J. P. Tomkins
Department of Genetics and Biochemistry, Genomics Institute,
Clemson University,
Clemson, SC 29634, USA

M. Luo · D. Kudrna · H. R. Kim · R. A. Wing
Department of Plant Sciences, Arizona Genomics Institute,
University of Arizona,
Tucson, AZ 85721, USA

K. Arumuganathan
Benaroya Research Institute at Virginia Mason,
1201 Ninth Avenue,
Seattle, WA 98101, USA

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

S. Zhao
The Institute for Genomic Research,
Rockville, MD 20850, USA

S. Zhao
Department of Biochemistry and Molecular Biology,
University of Georgia,
Athens, GA 30602-7229, USA

J. Leebens-Mack · C. W. dePamphilis
Department of Biology, Penn State University,
University Park, PA 16802, USA

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

J. A. Banks
Department of Botany and Plant Pathology, Purdue University,
West Lafayette, IN 47906, USA

D. F. Mandoli
Department of Biology and Center for Developmental Biology,
University of Washington,
Box 355325, Seattle, WA 98195-5325, USA

Present address:
J. Leebens-Mack
Department of Plant Biology, University of Georgia,
Athens, GA 30602-7271, USA

genomic resources for these early branching angiosperm lineages are very limited. In this study, we describe the construction of a large-insert bacterial artificial chromosome (BAC) library from *L. tulipifera*. Flow cytometry estimates that this nuclear genome is approximately 1,802 Mbp per haploid genome (±16 SD). The BAC library contains 73,728 clones, a 4.8-fold genome coverage, with an average insert size of 117 kb, a chloroplast DNA content of 0.2%, and little to no bacterial sequences nor empty vector content clones. As a test of the utility of this BAC library, we screened the library with six single/low-copy genic probes. We obtained at least two positive clones for each gene and confirmed the clones by DNA sequencing. A total of 182 paired end sequences were obtained from 96 of the BAC clones. Using BLAST searches, we found that 25% of the BAC end sequences were similar to DNA sequences in GenBank. Of these, 68% shared sequence with transposable elements and 25% with genes from other taxa. This result closely reflected the content of random sequences obtained from a small insert genomic library for *L. tulipifera*, indicating that the BAC library construction process was not biased. The first genomic DNA sequences for *Liriodendron* genes are also reported. All the *Liriodendron* genomic sequences described in this paper have been deposited in the GenBank data library. The end sequences from shotgun genomic clones and BAC clones are under accession DU169330–DU169684. Partial sequences of *Gigantea*, *Frigida*, *LEAFY*, cinnamyl alcohol dehydrogenase, 4-coumarate:CoA ligase, and phenylalanine ammonia-lyase genes are under accession DQ223429–DQ223434.

Keywords Yellow-poplar · BAC library

Introduction

Liriodendron tulipifera L., also called tulip tree, tulip-poplar, yellow-poplar, white-poplar, and whitewood, is one of the most attractive and tallest tree species of the eastern United States (Harlow and Harrar 1969). *Liriodendron* has a large natural range, covering much of eastern North America (Little 1979). The wood of yellow-poplar is commercially valuable and is used in a diverse range of products, such as in furniture and framing construction as a substitute for increasingly scarce softwoods (Hernandez et al. 1997; Moody et al. 1993; Williams and Feist 2004). Yellow-poplar 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 plantings (McCarthy 1933). Because of its rapid growth at preferred sites, yellow-poplar may have additional future potential as a source of fiber for biologically based products, biofuels, and chemicals (Nagle et al. 2002).

Yellow-poplar is a member of Magnoliaceae in the order Magnoliales, which is one of the early branching angiosperm lineages commonly known as "basal angiosperms." Magnoliales and three other basal angiosperm orders (Laurales, Piperales, and Canellales) comprise an extremely diverse group that may be sister to the monocots, sister to the eudicots, or sister to both monocots and eudicots (Soltis et al. 2000). Thus, Magnoliales holds a phylogenetically critical position among the early angiosperms that ultimately gave rise to the largest modern angiosperm lineages (cf. Qiu et al. 2005). Floral and other structural features place yellow-poplar at a phylogenetic position that is ideal for comparative studies of the evolution of biological processes in land plants (Hunt 1998; Ronse de Craene et al. 2003; Wei and Wu 1993). In addition, the availability of a reliable transformation system and the ability to mass produce clonal trees from somatic embryos (Wilde et al. 1992) make the species a good model forest tree in which we can explore modern breeding techniques and population dynamics as well as functional genomics.

Despite the fact that yellow-poplar has been used extensively as a benchmark species in studies on plant evolution (Parks and Wendel 1990; Wen 1999; Endress and Igersheim 2000; Zahn et al. 2005), there is very little information about the nuclear genome. A PubMed search with "Liriodendron" resulted in only one publication related to gene structure and function (LaFayette et al. 1999) and none on genomics. Recently, the Floral Genome Project created a cDNA library from yellow-poplar floral tissues by Albert et al. (2005). Sequences from this cDNA library resulted in an EST dataset with over 6,500 unigenes representing a wide variety of putative functions (http://pgn.cornell.edu/). Toward a major step forward for genomic-scale resources for yellow-poplar, this floral EST data is likely to lead to an even greater interest in the structure of genes and their regulatory regions in *Liriodendron*.

Large-insert genomic libraries are invaluable for genome sequencing, physical mapping, positional gene cloning, and the analysis of gene structure and function. Both yeast artificial chromosome (YAC) and bacterial artificial chromosome (BAC) vectors greatly facilitate the cloning of large genome fragments. BAC technology has substantial advantages over YAC libraries (Shizuya et al. 1992; Woo et al. 1994; Yu et al. 2000 and references therein); hence, more and more genome libraries employ BAC vectors. In this paper, we report the construction of a yellow-poplar BAC library, the first large-insert DNA library constructed for *Liriodendron*. Our primary goal is to develop the resources required to fully implement yellow-poplar as a model tree species for evolutionary and comparative genomics. An understanding of the *Liriodendron* genome will also support genetic improvement of yellow-poplar, including the application of biotechnology, and the characterization of traits important to the physiology and ecology of the species. In addition to characterization of

the BAC library, we also provide new genome-size estimates and identification of three floral timing genes [*Gigantea*, *Frigida*, and *LEAFY* (*LFY*)] as well as three genes in the lignin biosynthesis pathway [cinnamyl alcohol dehydrogenase (*CAD*), 4-coumarate:CoA ligase (*4CL*), and phenylalanine ammonia-lyase (*PAL*)] from the library.

Materials and methods

Nuclear DNA content determination

Fresh leaves from nine yellow-poplar trees, including the plant from which the BAC library was constructed (clone 108 at the University of Tennessee), were sent to Benaroya Research Institute at Virginia Mason, Seattle, WA for flow cytometric estimation of nuclear DNA content. Nuclei from *Nicotiana tabaccum* cv. SR-1 served as a size standard and propidium iodide as a staining dye in the flow cytometric assays. The assays were conducted individually on one, two, or four leaves from each tree, with triplicate runs for each leaf. Preparation of samples and instrumentation techniques were performed as described previously (Arumuganathan and Earle 1991).

Nuclear DNA characterization

To examine the complexity of the yellow-poplar genome and the density of single/low-copy protein-coding genes in the BAC library, a small shotgun library of yellow-poplar genomic DNA was constructed in the *Eco*RV site of plasmid vector pBluescript II SK(+). The *Liriodendron* shotgun library contained 3,072 clones with an average insert size of 3 kb (the LT_Sa library is available at http://genome.arizona.edu/orders/). DNA sequences were obtained from both ends of 96 randomly selected clones from the plasmid shotgun library, using the vector T7 and M13 reverse primer sites. High-quality sequences were trimmed and then used as queries in BLASTX, BLASTN, and TBLASTX (2.2.11 version, Altschul et al. 1990). High-quality sequences were defined as those having >100 high-quality bases (Q20) after trimming of vector sequence. The BLAST search of NCBI was run using a Batch BLAST script (http://greengene.uml.edu/analysis/analysis.html). A probability cutoff value (*E* value) of 10^{-6} was used to assign putative identities to these sequences.

BAC library construction

Young floral buds were obtained from clone 108 in the yellow-poplar breeding orchard of the University of Tennessee's Tree Improvement Program based at Knox-ville, Tennessee. Outer bud layers were removed, and the innermost layers including meristematic regions were used for high molecular weight DNA isolation. BAC library construction was essentially the same as that described previously by Tomkins et al. (2001, 2002). The first size selection of *Hin*dIII fragments via pulsed field gel electrophoresis used switch times of 1–40 s in a linear ramp for 18 h (12°C, 0.5× TBE buffer). Fractions between 100 and 300 kbp were cut from the gel, inserted into a second gel, and run with a 3- to 5-s switch time (18 h at 12°C, 0.5× TBE buffer) to remove small trapped DNA fragments. After removing appropriate fractions from the second size selection, DNA was removed from the agarose by electro-elution (Model 422 Electro-Eluter, Bio–Rad). Size-selected DNA was ligated into the pCUGIBAC-1 vector (Luo et al. 2001) and electroporated (Cell-Porator, Gibco BRL) into *Escherichia coli* (strain DH10B) cells using 320–330 V. Transformed cells were plated on selective medium [Luria–Bertani (LB) medium] in 24×24 cm plates (Genetix) with 12.5 μg/ml chloramphenicol, 0.55 mM IPTG, and 80 μg/ml X-Gal. After a 20-h incubation at 37°C, the plates were placed at room temperature in the dark for an additional 20 h to allow stronger color development of nonrecombinant colonies. Plates were either stored at 4°C or used immediately for picking. Recombinant white colonies were picked robotically using the Genetix Q-BOT and arrayed as individual clones in 384-well microtiter plates (Genetix) containing 50 μl freezing broth. After incubation overnight, microtiter plates were stored at −80°C. Three copies of the library were made using the replicating function of the Genetix Q-BOT and stored in separate −80°C freezers. A total of 384 clones were selected randomly for characterization of average insert size, percentage of empty vectors, and the frequency of organellar DNA clones as previously described by Tomkins et al. (1999), Budiman et al. (2000), and Yu et al. (2000).

BAC end sequencing

Ninety-six BAC clones were randomly selected and sent to The Institute for Genomic Research for end sequencing. To determine the sequence composition of the BAC end sequences, sequences were quality-trimmed and then used as queries in BLAST searches as described above in "Nuclear DNA characterization".

BAC library screening for specific genes

Four filters containing the entire BAC library were screened by hybridization for *Gigantea*, *Frigida*, *LFY*, *CAD*, *4CL*, and *PAL* genes. Probe DNA for each gene was prepared either by restriction enzyme digest or PCR from yellow-poplar cDNA clones provided by the Floral Genome

Project (http://www.floralgenome.org/fgp/). For the yellow-poplar *LFY* gene, however, a genomic clone was generated by PCR amplification of *Liriodendron* genomic DNA using a degenerate-sequence primer pair (LFY737: 5′-CGAG GTGGCCCGGGSNA-3′ and RCLFY739: 5′-CCGTCACG TCCCGCATYGTYACNTG-3′) and confirmed by sequencing, as a corresponding cDNA clone did not exist. The *LFY* probe was 491 bp in length, containing 105 bp of coding sequence and 386 bases of intron sequence. Detailed information about the probes and the best BLASTX alignments to the probes are available in electronic supplementary material S1. Approximately 50 ng of template DNA was used for preparing labeled probes. Probes were labeled with ^{32}P-dATP (3,000 Ci/mmol, 10 mCi/ml) using Strip-EZ™ DNA Random Primed StripAble DNA Probe Synthesis and Removal Kit, according to manufacturer's instructions (Ambion). Probes were purified using Centri Spin-20 columns (Princeton Separations) before hybridization. After prehybridizing the filters at 65°C overnight, filters were hybridized with individual probes at 55°C overnight. Washing stringency was set at 55°C. The solutions used for hybridization and washing were based on Frank's (1997). Filters were stripped between hybridizations following the procedure of Maniatas et al. (1982). Positive BAC clones (those with strong hybridization signals) were ordered from the Clemson University Genomics Institute (LT_Ba library clones; http://www.genome.clemson.edu/groups/bac/). Clones were confirmed by Southern hybridization of restriction enzyme-digested BAC DNA using the same probe resources as for library screening. The *Gigantea* cDNA probe was labeled with ^{32}P-dATP, and hybridization and washing conditions were the same as those described above for library screening. For the *Frigida*, *LFY*, *CAD*, *4CL*, and *PAL* genes, however, positive clones were confirmed by labeling probes with digoxigenin-dUPT and filter hybridization using the DIG-High Prime DNA Labeling and Detection Starter Kit I (as described by the manufacturer Roche Diagnostics).

Subcloning of genes from the BAC clones and DNA sequencing

One positive BAC clone was chosen for each gene. The BAC clone inserts were fragmented by restriction enzyme digestion (*Hin*dIII was used for *Gigantea* and *4CL*, *Eco*RI for *Frigida*, *CAD*, and *PAL*, and *Pst*I for *LFY*) and subcloned into the pBluescript II SK(−) plasmid vector (Stratagene, La Jolla, CA, USA). Subclones containing fragments of the relevant genes were identified by dot-blot hybridization (following the method of Maniatas et al. 1982) using the probes prepared as above for BAC clone verification (i.e., the *Gigantea* cDNA probe was labeled

with ^{32}P-dATP, and the *Frigida*, *LFY*, *CAD*, *4CL*, and *PAL* genes were labeled with digoxigenin-dUPT). Positive subclones were sent to either the Nucleic Acid Facility at Pennsylvania State University or the Cornell BioResource Center for sequencing.

Genomic DNA hybridization for gene copy number estimates

Total genomic DNA was isolated from bark tissues of young yellow-poplar twigs (clone 108 at the University of Tennessee) using the Qiagen DNeasy Maxi Kit. Aliquots (10 μg) of genomic DNA were digested individually with five different restriction enzymes (*Eco*RI, *Bam*HI, *Eco*RV, *Nco*I, or *Xho*I) and then size-fractionated via electrophoresis in a 0.8% agarose gel before being transferred to Hybond N membrane (Amersham). The genomic DNA was hybridized individually with *Gigantea*, *Frigida*, *LFY*, *PAL*, *CAD*, and *4CL* probes (same probe resources as for library screening). The *Frigida*, *LFY*, *CAD*, *4CL*, and *PAL* probes were labeled with ^{32}P-dATP, and the *Gigantea* cDNA probe was labeled with digoxigenin-dUTP. Hybridization was performed as described above for BAC library screening and positive clone confirmation.

Results

Genome-size estimates

The approximate genome size of yellow-poplar was determined by flow cytometry using leaf nuclei from nine unrelated trees and tobacco as an internal standard. The genome sizes obtained ranged from 1,853.36 to 1,706.50 Mbp per haploid genome, with an average size of 1,802 Mbp per haploid genome (SD=16) (Table 1). Yellow-poplar thus has a medium-sized genome when compared to plants known to be at the small (*Arabidopsis* at 157 Mbp per haploid genome) (Bennett et al. 2003) or large (wheat at 15,966 Mbp per haploid genome) (Arumuganathan and Earle 1991) ends of genome-size distribution in angiosperms.

Nuclear DNA characterization

To estimate the complexity of the *Liriodendron* genome, a small shotgun library of yellow-poplar genomic DNA was constructed in the plasmid vector, pBluescript II SK(+). Ninety-six clones randomly selected from this shotgun library were end-sequenced, resulting in a total of 176 high-quality sequences. The average read length was 635 bp with a standard deviation of 130 bp. When searched against GenBank, TBLASTX (matching each translated end-sequence tag against the translated nucleotide database;

Table 1 Flow cytometry determination of the nuclear genome sizes of several representative yellow-poplar individuals

Tree ID	Sources	Mbp per haploid genome (mean±SD[a])
Clone #35	University of Tennessee, Knoxville, TN	1793.58±13.36
Clone #108	University of Tennessee, Knoxville, TN	1706.50±18.57
H522	University Park campus, Penn State University	1853.36±21.60
H483	University Park campus, Penn State University	1820.82±8.05
H482	University Park campus, Penn State University	1816.75±33.17
H164	University Park campus, Penn State University	1798.85±13.69
H295	University Park campus, Penn State University	1821.12±11.23
H9HT	University Park campus, Penn State University	1833.49±6.03
H11HT	University Park campus, Penn State University	1773.44±17.88
–	Average genome-size value	1801.99±15.95

[a] Average of 3, 6, or 12 assays (one, two, or four leaves run in triplicate) per tree

August 8, 2005) identified 54 matches with scores of $E \leq 10^{-6}$, whereas BLASTX (matching translated end sequences to the GenBank protein database) and BLASTN (matching nucleotide to nucleotide sequences) alignments resulted in 47 and 14 sequences, respectively. Among the 54 sequences with $E \leq 10^{-6}$, there was one sequence homologous to a chloroplast protein ($E \leq 8 \times 10^{-9}$), and there were three sequences identified as homologs of bacterial or fungal proteins ($E \leq 2 \times 10^{-27}$). For the remaining 50 sequences, 30 (60%) matched various transposable elements (17% of the total number of shotgun end sequences), while 16 sequences (32%) matched putative genes (or 9% of the total number of shotgun end sequences). There were four sequences with $E \leq 10^{-6}$ matching to sequences that could not be classified into any informative categories. Forty-one shotgun end sequences had best alignments to eudicot sequences (82%), whereas eight sequences (16%) with top hits had best alignments to monocot sequences. Finally, there were 122 shotgun end sequences (69%) that did not share significant sequence similarities to any DNA or protein sequences in GenBank. The AT content of the shotgun end sequences was 59%, whereas GC content was 41%. Detailed information about the BLAST results of the 54 shotgun end sequences can be found in supplemental Table S2.

Liriodendron BAC library construction and characterization

Our *L. tulipifera* BAC library, which was constructed from yellow-poplar clone 108 in the University of Tennessee's Tree Improvement Program, is suitable for physical mapping, map-based cloning, and high-throughput sequencing of selected genomic regions. The library consists of 73,728 clones stored in 192 384-well microtiter plates at the Clemson University Genomics Institute. All of the clones in the BAC library were gridded onto 22.5×22.5 cm filters at high density in 4×4 patterns of doubled spots for ease in spot identification after hybridization.

A total of 384 randomly sampled clones were digested with *Not*I to release the insert. Digestion typically generated vector plus one to two insert bands per BAC clone (Fig. 1) probably due to the facts that there are two *Not*I sites in pCUGI-1 flanking the multi-cloning site, and *Not*I recognizes an 8-bp GC sequence, while the yellow-poplar genome is relatively AT-rich (59% according to our BAC end and shotgun end sequences). To determine the size

Fig. 1 An analysis of 43 randomly selected *Liriodendron* BAC clones. Ethidium-bromide-stained CHEF gel (5–15 s switch time, 15 h) showing insert DNA above the common 7.5 kb pCUGI-1 vector band. Molecular weight marker in outside lanes is a 50-kb lambda ladder, and the molecular weight marker in the center of gel is a Midrange I (NEB) with bands at 15, 34, 49, 82, 97, 131, and 146 kb
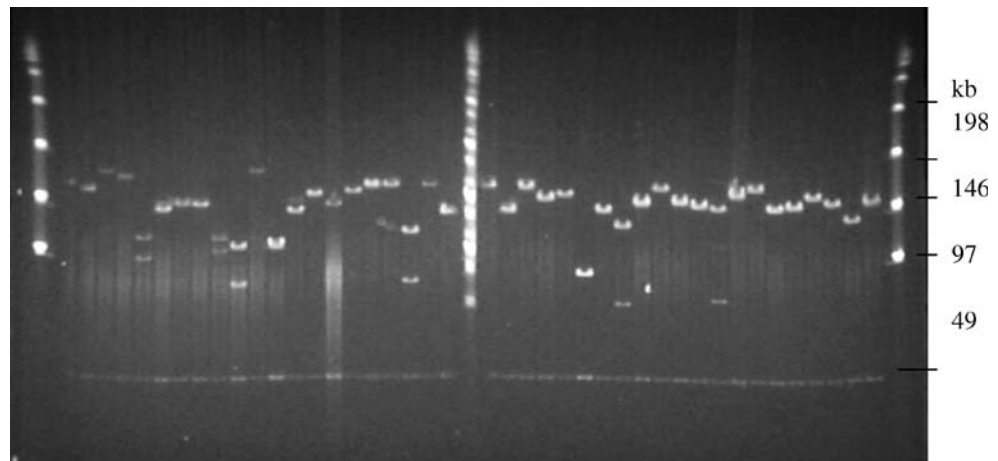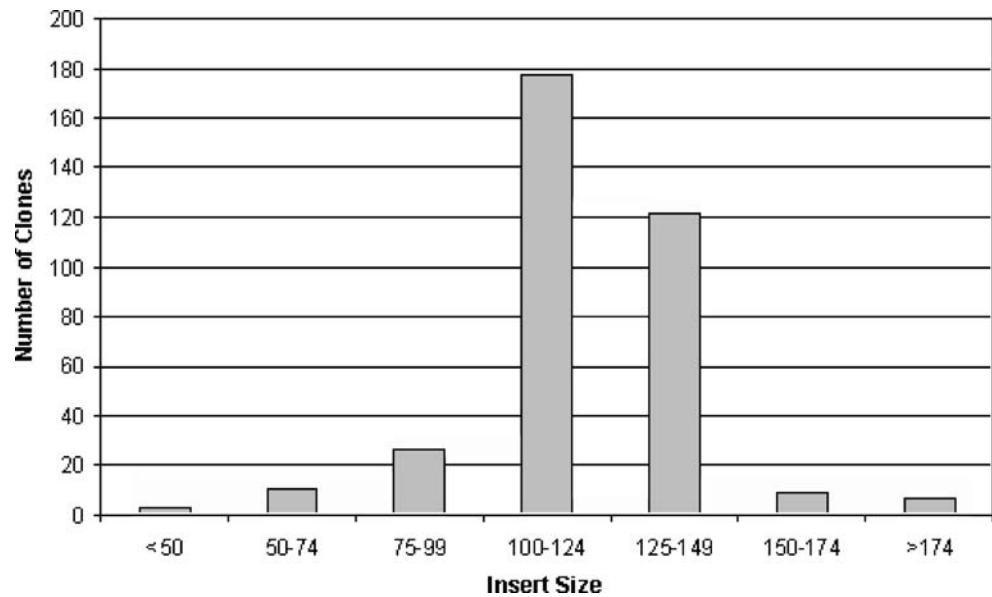
Fig. 2 Insert size distribution of clones from the yellow-poplar BAC library. To estimate insert size range, BAC DNA inserts from 384 randomly selected clones were analyzed using *Not*I digestion and pulsed field gel electrophoresis. The average insert size was 117 kb, and 89% of the clones are ≥100 kb

distribution of BAC clones in the library, the 384 BACs analyzed with *Not*I digests were grouped by insert size, and the insert size of each clone was plotted against the frequency of each group of clones represented in the library (Fig. 2). This analysis revealed that 89% of the clones in the library had an average insert size equal to or larger than 100 kbp. The average insert size of the BAC library was 117 kb, with a range of 40 to 280 kb. A very low percentage for the empty vector content clones (less than 1%) was found. Based on the average insert size and a haploid genome size of 1,802 Mbp, the coverage of the library is 4.8× haploid genome equivalents, resulting in 99.2% probability of recovering any specific sequence of interest (Clarke and Carbon 1976).

Approximately 0.2% of library sequences were determined to be of chloroplast DNA origin by screening the high-density colony filter arrays with four highly conserved chloroplast genes spaced equidistantly around the 154-kbp soybean chloroplast genome. We attribute this exceptionally low level of chloroplast sequence contamination to use of the inner layers of young floral buds that were not photosynthetically mature and probably had a low plastid content.

BAC end sequencing results

As one quality control, 96 BAC clones were end-sequenced. A total of 192 sequencing reactions were performed, resulting in a set of 182 high-quality sequences, for a sequencing success rate of 95%. The average read length, after trimming of vector sequence, was 761 bp with a standard deviation of 142 bp.

TBLASTX searches against GenBank resulted in 45 BESTs with alignment scores of $E \leq 10^{-6}$ or better, whereas BLASTX and BLASTN alignments resulted in 42 and nine

sequences, respectively. Among the BAC end-sequence tags, only one contained chloroplast DNA. This value, 0.5% of the total high-quality BAC end-sequence tags, is consistent with the estimate of 0.2% chloroplast DNA inserts obtained by hybridization screening of the library with probes from the chloroplast genome. Of the remaining 44 BESTs with matches in GenBank, 30 had sequences similar to transposable elements. This indicates that 68% of the BAC end-sequence tags with matches in GenBank and a minimum of 16% of the *Liriodendron* BAC end sequences may have been derived from transposable elements. In addition, there were 11 putative genes detected by BLAST searches, a value that represents 25% of sequences with matches in GenBank and 6% of the total high-quality *Liriodendron* sequences. There were two sequences with $E \leq 10^{-6}$ matching to sequences that could not be classified into any informative categories. The remaining BAC end-sequence tag with a significant alignment score was similar to a *Magnolia stellata* microsatellite marker (*E* value=7e−13). Finally, 136 high-quality BAC end-sequence tags (75%) lacked significant matches in GenBank. The putative assignments of the 45 BESTs by BLAST researches are available in S3. BLAST searches also revealed that 84% of the BESTs with $E \leq 10^{-6}$ had best hits to eudicot sequences, whereas 9% of the top alignments were to monocot sequences. The AT content of the BESTs was 59%, whereas GC content was 41%.

Screening of the BAC library for *Gigantea*, *Frigida*, *LEAFY*, *CAD*, *4CL*, and *PAL* genes

To test if our BAC library was useful for cloning homologues, we cloned three floral-timing genes, *Gigantea*, *Frigida*, and *LFY*, and three lignin biosynthesis pathway

**Table 2** Positive clones identified from screening the *Liriodendron* BAC library with homologous probes

Gene used as probe	Number of positive BAC clones	IDs[a] of sequenced BAC clones (GenBank #)	Length of DNA sequence obtained (bp)	Species with best BLASTX match (GenBank #) (*E*-value)
Gigantea	5	Lt__Ba172K7 (DQ223429)	8,108	*Arabidopsis thaliana* (gi:50660455) (5e-168)
Frigida	3	Lt__Ba20J4 (DQ223430)	7,340	*Arabidopsis thaliana* (gi:73915375) (3e-06)
LFY/FLO	2	Lt__Ba92I22 (DQ223431)	1,035	*Salvia coccinea* (gi:42795325) (2e-57)
CAD	4	Lt__Ba115C17 (DQ223432)	1,755	*Oryza sativa* (gi:50905799) (9e-18)
4CL	5	Lt__Ba135O8 (DQ223433)	1,996	*Fagus sylvatica* (gi:73665914) (4e-38)
PAL	8	Lt__Ba60I3 (DQ223434)	124	*Persea americana* (gi:1171999) (4e-15)

[a] IDs of BAC clones from the Lt__Ba *Liriodendron* BAC library at the Clemson University Genomics Institute's BAC/EST Resource Center

genes, *PAL*, *CAD*, and *4CL*, from the library. These genes are known to be low copy number in other plant species (Park et al. 1999; Johanson et al. 2000; cf. Dixon et al. 2001; cf. Wada et al. 2002; Dunford et al. 2005; http://fgp.huck.psu.edu/tribe.php). As shown in Table 2, at least two positive BAC clones were found for each of these genes, with *PAL* having the most clones (eight clones) and *LFY* the least (two clones). The fact that only two positive clones were found for *LFY* in the BAC library may have resulted from the small size of the probe that was used

(105 bp of coding sequence). Putative positive BAC clones were reconfirmed by Southern hybridization of restricted BAC DNA (Fig. 3), and then positive restriction fragments from the BAC inserts (those that hybridized in the Southern analysis) were cloned. Genomic DNA sequence was then obtained for each of the subclones to confirm that they contained the desired genes. BLASTX search of GenBank with the sequences of these BAC subclones yielded strong matches (*E*<5e−15) to homologues from other species for all of the genes (Table 2) except *Frigida*. When compared

Fig. 3 Examples of the BAC library screening for specific genes of interest and confirmation by Southern hybridization. An autoradiogram of BAC-library filter (*left*) and an autoradiogram of Southern blot (*Hin*dIII digest) (*right*) are shown

CAD BAC filter screening (filter C)

Gigantea positive clones confirmed by Southern hybridization
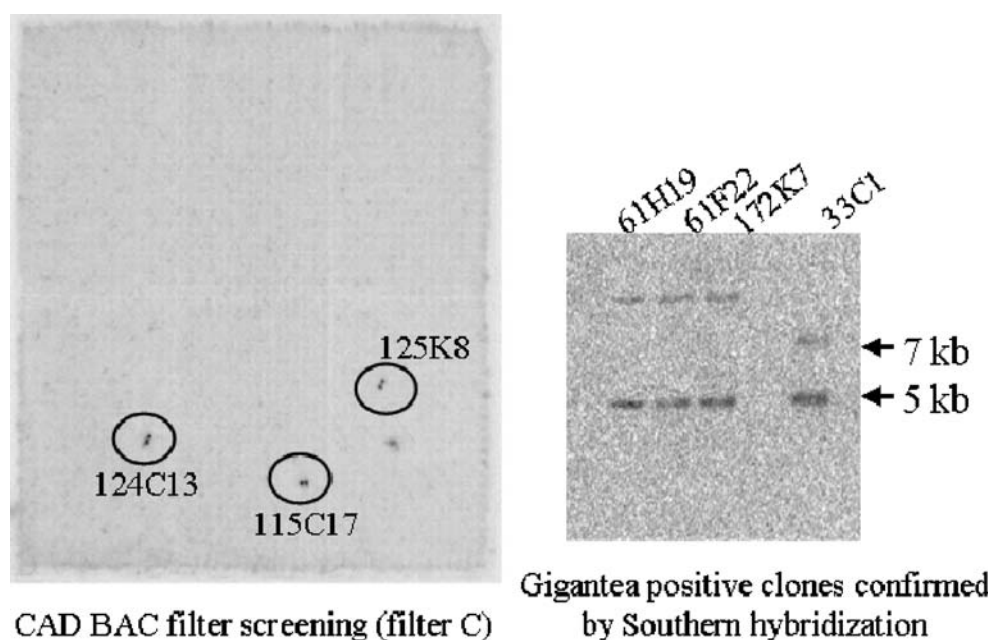
to the *Frigida* protein sequence from *Arabidopsis*, the *Liriodendron* Frigida sequence shared only 34% amino acid identity, with an *E* value for BLASTX at only 7e-06. This is not too surprising because the *ArabidopsisFrigida* gene is not well conserved in other plant species, including cereals, pea, and legumes (Izawa et al. 2003; Simpson 2003; Hecht et al. 2005).

Genomic Southern hybridization analysis

Southern hybridization was conducted to estimate the copy numbers of the genes used for screening the BAC library filter array. Five different restriction enzymes were employed individually to digest the genomic DNA. However, for each gene, only one or two of the restriction enzyme digests resulted in useful hybridization patterns with fragments smaller than 10 kb that could be accurately scored. As shown in Fig. 4, the *Gigantea*, *Frigida*, *LFY*, *PAL*, *CAD*, and *4CL* probes detected one to two fragments in the nuclear genome of *Liriodendron* by Southern hybridization. The only exception was the *CAD* probe which revealed five genomic fragments on the *Eco*RV digest blot. Based on the sequences included in the probes and the number of clones retrieved from the BAC library, most of the hybridization fragments were as expected in terms of size and number, indicating that the *Gigantea*, *Frigida*, *LFY*, *PAL*, and *4CL* genes appear to be either low or single copy genes in yellow-poplar, while *CAD* may be in a small, multigene family.

Discussion

We describe in this study the development and characterization of the first large-insert DNA library for the *Liriodendron* genus. To our knowledge, these are the first published data on the complexity of a *Liriodendron* genome or that of any other basal angiosperm species. This BAC library contains 73,728 clones with an average insert size of 117 kb. Contamination is very low in the library (0.2% of chloroplast DNA and less than 0.1% of empty vector clones). Taking 0.2% of chloroplast DNA into consideration along with a haploid genome size of approximately 1,802 Mb, we estimate that the library represents 4.8× haploid genome equivalents and calculate a 99.2% probability of including any specific sequence of interest. These features will facilitate the application of this library for many high-throughput genomic applications.

The genome-size estimates that we obtained from nine different yellow-poplar trees were over twice as large as the previously reported genome size for *Liriodendron* (790 Mbp, Bennett et al. 2000). Because the BAC library was constructed based on the previous genome-size estimate, the library coverage is lower than originally intended. However, at 4.8× haploid genome equivalents, it still provides sufficient genome coverage to permit low-copy genes to be isolated, as demonstrated by the success of retrieving sequences for *Gigantea*, *Frigida*, *LFY*, *CAD*, *4CL*, and *PAL* genes from the library.

We chose the *Gigantea*, *Frigida*, *LFY*, *CAD*, *PAL*, and *4CL* genes to test the utility of the library in identifying important regions of the genome for two reasons. First, these six genes are known to play important roles in timing floral initiation and in lignin biosynthesis in other plants. Functional genomics studies of these genes in *Liriodendron* and comparative studies of these genes with their homologues in other plants should be informative for further studies of floral development and lignin biogenesis. Second, these genes are low-copy genes in other plant species (in most cases, one to two copies per haploid plant genome for *Gigantea*, *Frigida*, and *LFY*, and one to four copies for *CAD*, *PAL*, and *4CL*; Park et al. 1999; Johanson et al. 2000; cf. Dixon et al. 2001; cf. Wada et al. 2002; Dunford et al. 2005). Our genomic Southern hybridization analysis resulted in simple hybridization patterns that indicate

Fig. 4 Genomic Southern hybridization. Aliquots of yellow-poplar genomic DNA were digested with restriction enzymes, separated by electrophoresis on a 0.8% agarose gel, and transferred to Hybond N membrane (Amersham). The genomic DNA blots were hybridized individually with *Gigantea*, *Frigida*, *LFY*, *PAL*, *CAD*, and *4CL* probes. *L* DNA Ladder I (GeneChoice)
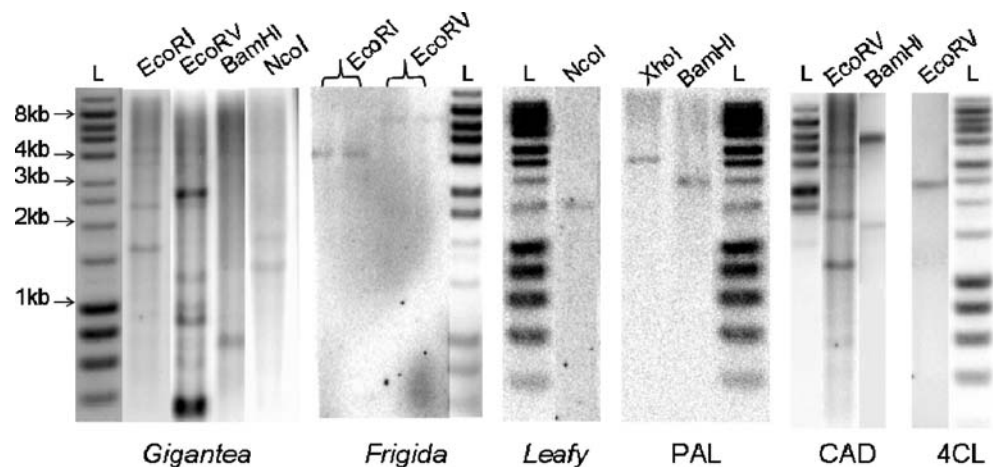
that these genes appear to be low copy in yellow-poplar as well. Assuming that *Gigantea*, *Frigida*, and *LFY* are present as single copy genes, and given the 4.8× depth of coverage for the library, we would expect to find four to five BAC clones in the library for each of these genes. Our primary screening of filter arrays of the library produced 4.3 clones per gene (on average; data not shown), which was very close to the expected number. After Southern hybridization analysis, used to confirm the clones, the average number of positive BAC clones was, however, reduced to 3.3 (ranging from two to five). The genomic sequences obtained from this study are the first to be reported for *Liriodendron* and among the first reported for any basal angiosperm species.

The BLAST searches of both the *Liriodendron* BAC end sequences and the small-insert random genomic clone shotgun end sequences produced at least twice as many matches in GenBank to transposons than to genes. Transposable elements are ubiquitous in the plant kingdom. They are present in high copy numbers in most plants, making them major constituents of plant genomes (Kumar and Bennetzen 1999). Plant species, like maize (San Miguel et al. 1996), *Pinus lambertiana* (Kossack and Kinlaw 1999), *Lilium speciosum* (Leeton and Smyth 1993), wheat (Moore et al. 1991), rice (Noma et al. 1997), *Vicia* species (Pearce et al. 1996a,b), *Allium cepa* (Pearce et al. 1996a,b), and tobacco (Yoshioka et al. 1993), contain more than 50% transposable elements in their genomes. Based on our genome-size estimate and BLAST results, we predict that over 50% of the *Liriodendron* genome is composed of transposable elements also. The number of transposable elements in the *Liriodendron* genome may be even greater than this estimate if the majority do not share sequence homology with any previously sequenced elements. Perhaps the relatively large genome size that we found for *Liriodendron* relative to other magnoliids can be explained by a larger number of transposable elements and/or other repetitive elements in *Liriodendron*, as opposed to a recent genome duplication. This interpretation is supported by our Southern hybridization results for single copy genes, which would suggest that either a recent genome duplication event has not taken place or that the low copy genes in *Liriodendron* behave as single copy on Southern blots because the genes have not yet diverged in sequence. Analysis of genes expressed during flower development (Albert et al. 2005) does show evidence for two rounds of ancient polyploidy in *Liriodendron* (Cui et al. 2006). However, the Ks approach used in Cui et al. (2006) cannot detect very recent genome duplication events in which sequences in multiple copies of a gene are so similar that they assemble into the same unigenes.

It is quite interesting that the majority of the *Liriodendron* BAC end sequences and shotgun end sequences with matches in GenBank aligned best to eudicot sequences (84 and 82%, respectively) despite the fact that the full rice genome was available. This pattern may support a phylogeny that places the magnoliids closer to eudicots than to monocots, or may be indicative of more similar composition patterns at the DNA and protein levels between *Liriodendron* and the diverse eudicots that have been identified (Leebens-Mack et al. 2005; Qiu et al. 2005). Sequences from BAC and shotgun end sequencing also indicate that *Liriodendron* has a relatively AT-rich genome.

Approximately 75% of BESTs and 68% of the genomic shotgun sequences did not share significant sequence similarities to any DNA or protein entries in GenBank, indicating that a significant portion of the yellow-poplar genome remains unknown and, when better characterized, may be more informative for understanding other basal angiosperms than sequences from those model species from which most genomic sequences in GenBank are derived. Because *L. tulipifera* occupies an important phylogenetic position as a basal angiosperm, genomic information from it will help provide important insights in the evolution of biological processes in angiosperms. When combined with the EST data that is available, the *LT__Ba* BAC library will be an important tool for gene targeting and unraveling gene functions in yellow-poplar. The yellow-poplar BAC library *LT__Ba* is available to the scientific community through the Clemson Genomics Institute (for ordering information, visit the CUGI BAC/EST Resource Center site http://www.genome.clemson.edu/cgi-bin/orders).

References

Albert VA, Soltis DE, Carlson JE, Farmerie WG, Wall PK, Ilut DC, Mueller LA, Landherr LL, Hu Y, Buzgo M, Kim S, Yoo M-J, Frohlich MW, Perl-Treves R, Schlarbaum S, Bliss B, Tanksley S,

Oppenheimer DG, Soltis PS, Ma H, dePamphilis CW, Leebens-Mack JH (2005) Floral gene resources from basal angiosperms for comparative genomics research. BMC Plant Biol 5:5

Altschul SF, Gish W, Miller EW, Myiers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–219

Bennett MD, Bhandol P, Leitch I (2000) Nuclear DNA amounts in angiosperms and their modern uses—807 new estimates. Ann Bot 86:859–909

Bennett MD, Leitch IJ, Price HJ, Johnston JS (2003) Comparisons with *Caenorhabditis* (~100 Mb) and Drosophila (~175 Mb): using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the *Arabidopsis* genome initiative estimate of ~125 Mb. Ann Bot 91:547–557

Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. Genome Res 10:129–136

Clarke L, Carbon J (1976) A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. Cell 9:91–100

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

Dixon RA, Chen F, Guo D, Parvathi K (2001) The biosynthesis of monolignols: a "metabolic grid", or independent pathways to guaiacyl and syringyl units? Phytochemistry 57:1069–1084

Dunford RP, Griffiths S, Christodoulou V, Laurie DA (2005) Characterisation of a barley (*Hordeum vulgare* L.) homologue of the *Arabidopsis* flowering time regulator GIGANTEA. Theor Appl Genet 110:925–931

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

Frank MB (1997) Southern blot hybridizations. In: Frank MB (ed) Molecular biology protocols. Oklahoma City, OK

Harlow WM, Harrar ES (1969) Textbook of dendrology. McGraw-Hill, New York, pp 512

Hecht V, Foucher F, Ferrandiz C, Macknight R, Navarro C, Morin J, Vardy ME, Ellis N, Beltran JP, Rameau C, Weller JL (2005) Conservation of *Arabidopsis* flowering genes in model legumes. Plant Physiol 137:1420–1434

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*, U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, WI

Hunt D (1998) Magnolias and their allies. International Dendrology Society & Magnolia Society, pp 304

Izawa T, Takahashi Y, Yano M (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. Curr Opin Plant Biol 6:113–120

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science 290:344–347

Kossack DS, Kinlaw CS (1999) IFG, a *gypsy*-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. Plant Mol Biol 39:417–426

Kumar A, Bennetzen JL (1999) Plant retrotranspons. Annu Rev Genet 33:479–532

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

Leebens-Mack J, Raubeson LA, Cui L, Kuehl J, Fourcade M, Chumley T, Boore JL, Jansen RK, dePamphilis CW (2005) Identifying the basal angiosperms in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol Biol Evol 22:1948–1963

Leeton PR, Smyth DR (1993) An abundant LINE-like element amplified in the genome of *Lilium speciosum*. Mol Gen Genet 237:97–104

Little EL Jr (1979) Checklist of United States trees (native and naturalized). U.S. Department of Agriculture, Agriculture Handbook 541. Washington, DC, pp 375

Luo M, Wang YH, Frisch D, Joobeur T, Wing RA, Dean RA (2001) Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (Fom-2). Genome 44:154–162

Maniatas T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, NY

McCarthy EF (1933) Yellow-poplar characteristics, growth, and management. U.S. Department of Agriculture, Technical Bulletin 356. Washington, DC, pp 58

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, WI

Moore G, Lucas H, Batty N, Flavell R (1991) A family of retrotransposons and associated genomic variation in wheat. Genomics 10:461–468

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

Noma K, Nakajima R, Ohtsubo H, Ohtsubo E (1997) RIRE1, a retrotransposon from wild rice (Oryza australiensis). Genes Genet Syst 72:131–140

Park D, Somers DE, Kim YS, Choy YH, Lim HK, Soh MS, Kim HJ, Kay SA, Nam HG (1999) Control of circadian rhythms and photoperiodic flowering by the *Arabidopsis GIGANTEA* gene. Science 285:1579–1582

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

Pearce SR, Harrison G, Li D, Heslop-Harrison JS, Kumar A, Flavell AJ (1996a) The Ty1-*copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250:305–315

Pearce SR, Pich U, Harrison G, Flavell AJ, Heslop-Harrison JS, Schubert I, Kumar A (1996b) The Ty1-*copia* group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. Chromosome Res 4:357–364

Qiu Y-L, Dombrovska 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 analysis of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. Int J Plant Sci 166:815–842

Ronse de Craene L, Soltis DE, Soltis PS (2003) Evolution of floral structures in basal angiosperms. Int J Plant Sci 164:S329–S363

San Miguel P, Tikhonov A, Jin YK, Motchoulskaian N, Zakharov D, Melake-Berhan A, Springer PS, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Shizuya H, Birren B, Kim U, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-

kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci U S A 89:8794–8797

Simpson GG (2003) Evolution of flowering in response to day length: flipping the *CONSTANS* switch. Bioessays 25:829–832

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

Tomkins JP, Mahalingham R, Miller-Smith H, Goicoechea JL, Knapp HT, Wing RA (1999) A soybean bacterial artificial chromosome library for PI 437654 and the identification of clones associated with cyst nematode resistance. Plant Mol Biol 41:25–32

Tomkins JP, Peterson DG, Yang TJ, Main D, Wilkins TA, Paterson AH, Wing RA (2001) Development of genomic resources for cotton (*Gosypium hirsutum*): BAC library development, preliminary STC analysis, and identification of clones associated with fiber development. Mol Breed 8:255–261

Tomkins JP, Davis G, Main D, Duru N, Musket T, Goicoechea JL, Frisch DA, Coe EH Jr, Wing RA (2002) Construction and characterization of a deep-coverage bacterial artificial chromosome library for maize. Crop Sci 42:928–933

Wada M, Cao QF, Kotoda N, Soejima J, Masuda T (2002) Apple has two orthologues of *FLORICAULA/LEAFY* involved in flowering. Plant Mol Biol 49:567–577

Wei ZX, Wu ZY (1993) Pollen ultrastructure of Liriodendron and its systematic significance. Acta Bot Yunnanica 15:163–166

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

Williams RS, Feist WC (2004) Durability of yellow-poplar and sweetgum and service life of finishes after long-term exposure. Forest Products J 54:96–101

Woo SS, Jiang J, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Res 22:4922–4931

Yoshioka Y, Matsumoto S, Kjima S, Ohshima K, Okada N, Machida Y (1993) Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc Natl Acad Sci USA 90: 6562–6566

Yu Y, Tomkins JP, Waugh R, Frisch DA, Kudrna D, Kleinhofs A, Brueggeman RS, Muehlbauer GJ, Wise RP, Wing RA (2000) A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. Theor Appl Genet 101:1093–1099

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