12-2013

DEVELOPMENT OF Liriodendron EST-SSR MARKERS AND GENETIC COMPOSITION OF TWO Liriodendron tulipifera L. ORCHARDS

Xinfu Zhang
Clemson University, zhangxinfuralph@gmail.com

DEVELOPMENT OF *Liriodendron* EST-SSR MARKERS AND GENETIC COMPOSITION OF TWO *Liriodendron tulipifera* L. ORCHARDS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Genetics

---

by
Xinfu Zhang
December 2013

Accepted by:
Dr. Haiying Liang, Committee Chair
Dr. Ksenija Gasic
Dr. James Morris
Dr. Alex Feltus

ABSTRACT

Liriodendron tulipifera L., commonly known as yellow-poplar, is a fast-growing hardwood tree species with great ecological and economic value and is native to eastern North America. *Liriodendron* occupies an important phylogenetic position as a basal angiosperm and has been used in studies of the evolution of flowering plants. Genomic resources, such as Expressed Sequence Taq (EST) databases and Bacterial Artificial Chromosome (BAC) libraries, have been developed for this species. However, no genetic map is available for *Liriodendron*, and very few molecular markers have been developed.

In this study, a total of 119 informative genomic SSR markers suitable were identified for genetic linkage map construction with an F1 progeny from #UT108A × #UT23 cross, that have been developed. The full-sibship of 213 seedlings were validated. These informative SSR markers and full-sib seedlings are essential in construction of linkage maps. Linkage map will enable molecular breeding and quantitative trait locus (QTL) mapping, and provide framework for sequencing the *Liriodendron* genome. In addition we characterized 20 EST-SSR markers with 174 trees from two yellow-poplar seed orchards (residing in Knoxville, Tennesse, and Clemson, South Carolina, respectively), and the US National Arboretum, and provided a first look at the genetic diversity and allele richness among selections of this unique native species. Analysis revealed only one locus significantly deviating from Hardy-Weinberg proportions in the Clemson population, and 10 loci in Knoxville population ($p>0.05$). In addition, the Clemson orchard exhibited higher values of observed and effective number of alleles,

observed heterozygosity, and Nei's expected heterozygosity than the Knoxville orchard, revealing larger genetic diversity in the Clemson seed orchard.

DEDICATION

I dedicate this thesis to my family: my father, Hongyin Zhang, and my Mother, Jing Liu. They gave me positive encouragement during my graduate study. My family will always be the most important part of my life.

ACKNOWLEDGMENTS

Firstly, I would like to give my sincere thanks to my advisor, Dr. Haiying Liang, for her patient guidance, lasting assistance, and positive encouragement to me. Secondly, I would like to thank to Dr. Ksenija Gasic, Dr. James Morris, and Dr. Alex Feltus for serving on my committee, and for their sparkling wisdom, insightful feedback, and professional suggestions. I thank Jeanice Troutman, Christopher Saski, Xiaoxia Xia, and Margaret Staton in Clemson University Genome Institute (CUGI) for technical support in fragment analysis. I also thank Dr. Nannan Yang, a former Ph.D. student in Dr. Gasic's lab, Dr. Jeanne Romero-Severson, and Tim McCleary in University of Notre Dame, who gave me professional suggestions about the whole work. I thank Dr. Scott Schlarbaum's group in The University of Tennessee, who provided us plant materials. I also thank the former and current lab members in our lab, Dr. Haiying Liang, Dr. Shivegowda Thammannagowda, Dr. Yi Xu, Tao Xu, Dr. Tieguang He, Dr. Zhenkun Tian, and Ying Zhao. I thank the promising undergraduate students who worked with me, including Alanna Carlson, Christopher Saindon, and Jasmine Ruizyi in Clemson University, and Melany Moore from The University of Tennessee. Finally, I would like to thank all the faculty, staff, students, and visiting scholars in the Department of Genetics and Biochemistry. I am proud of being a member in this wonderful department.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

LITERATURE REVIEW

Liriodendron tulipifera L., a potential tree model research system for comparative genomics

Liriodendron tulipifera L., commonly called yellow-poplar, is an attractive large deciduous tree species, which is widely distributed in the eastern United States, and is the state tree of Tennessee. The only other species in *Liriodendron*, *Liriodendron chinense* (Hemsl.) sarg, inhabits the mountains in southern China and Northern Vietnam (Hao et al., 1995). Although these two species were deduced to have separated 10-16 million years ago from molecular divergence (Parks and Wendel, 1990), and exist on opposite sides of the globe, their hybrid have demonstrated healthy growth and aesthetic appearance (Wang, 2005), making *Liriodendron* available for mating system, systematic evolution and population genetics studies (Xu et al., 2006).

Even though yellow-poplar has wood structure and density similarity to *Populus* species, it does not belong to *Malpighiales*, the core eudicot order of *Populus*; instead, it is a member of *Magnoliaceae* in the order *Magnoliales*, which occupy an early branching of "basal angiosperm" lineages (Soltis et al., 2000). *Magnoliales* have been proved to be intermediate sister group to a large clade of angiosperms of monocots and eudicots by

recent molecular phylogenetic analysis of genome sequence datasets (Qiu et al., 2005; Jansen et al., 2007; Moore et al., 2007). Phylogenetic analysis of GIGANTEA amino acid sequences from yellow-poplar, monocots, and eudicots also classified yellow-poplar closer to eudicots than to monocots (Liang et al., 2010). In addition, its unusual flower structures also place it into an important phylogenetic position. Distinguished from mainstream angiosperms with whorled floral organs, yellow-poplar always arranges its stamens and pistils in a spiral pattern, which is probably an ancestral trait of flowering plants (Soltis et al., 2004). Thus, *Liriodendron tulipifera* has been deeply studied as a candidate for comparative studies and genome evolution of angiosperms (De Craene et al., 2003; Zahn et al., 2005; Liang et al., 2011).

Yellow-poplar has great economic and ecological values. It can attain a height of 150 feet with a trunk of 5 feet in diameter (Burns and Honkala, 1990), with fast growing rate, and outstanding resistance to insects, diseases, and damaging metals (Klugh and Cumming, 2007; Chen et al., 2012). Its ability to grow in barren soil and highland areas and carbon absorption capacity has been reported (Gwak et al., 2009; Kim et al., 2012b). Yellow-poplar is a productive source of industrial raw material and wild life food (Moody et al., 1993; Hernandez et al., 1997; Williams and Feist, 2004). As a result, it is cultivated worldwide for wood production and waste landfill remediation, and as urban avenue trees (Hunt, 1998; Kim and Lee, 2005). Recently, 21 compounds have been isolated and studied from yellow-poplar (Chen et al., 2012), including four alkaloids (Chi et al., 2006; Chen et al., 2008; Sawasdee et al., 2010), three lignans (Xu et al., 2001; Kim et al., 2010; Yang et al., 2010), four steroids (Gaspar and Dasneves, 1993; Liu et al.,

2010), and 10 benzenoids (Rojas et al., 2000; Antolovich et al., 2004; Voitl and von Rohr, 2008; Liu et al., 2009; Piao et al., 2009; Shi et al., 2009; Wang et al., 2009; Zhao et al., 2009; Lin et al., 2010). Chemical extracts from yellow-poplar have anti-tumor effects and herbivore anti-feeding activity (Moon et al., 2007). Lastly, its biomass can be converted into biofuels (Xiang et al., 2004; Berlin et al., 2006; Celen et al., 2008; Hwang et al., 2008; Koo et al., 2008; Koo et al., 2009).

Due to its ecological and economical importance and unique phylogenetic position, genomic resources of *L. tulipifera* have been developed rapidly in recent years. Yellow-poplar's chromosome number, 2n=2x=38, is lowest in *Magnoliaceae* family. It is estimated that its genome size is 1802 Mbp per haploid genome (SD=16; Liang et al., 2007). One Bacterial Artificial Chromosome (BAC) library for yellow-poplar has been constructed, which consists of 73,728 clones with an average insert size as 117 kb (Liang et al., 2007). This library, containing 4.8 haploid genome equivalents, has a 99.2% probability of recovering any interested specific sequence of interest. In addition, a small shotgun library containing 3,072 clones with an average insert size of 3 kb is available for purchase at http://genome.arizona.edu/orders/.

The first complementary DNA (cDNA) library was constructed from young floral buds (Albert et al., 2005; Liang et al., 2008). From 9531 high-quality ESTs, 6520 unigenes were yielded, including 5251 singletons and 1269 contigs (unigene build number 4, 2004-12-4, http://pgn.cornell.edu). Among the unigenes, 16% contained full-length coding regions, of which 90% had simple sequence repeats (SSRs). From this

dataset primers for a total of 176 SSR markers were designed and characterized on 8% polyacrylamide gels (Xu et al., 2006; Xu et al., 2010). Sixty-six of these markers produced polymorphic products in *L. tulipifera* (Table 1). Another Thirty-nine polymorphic EST-SSR loci were identified, among which 32 showed interspecific transferability and polymorphism in the related species, *L. chinense* (Yang et al., 2012). Since then a more comprehensive EST dataset, which consisted of 132,905 contigs and 4599 singletons, was constructed from 10 different *L. tulipifera* tissues: (1) premeiotic flower buds, (2) postmeiotic flower buds, (3) open flowers, (4) developing fruit, (5) terminal buds, (6) leaves, (7) cambium, (8) xylem, (9) roots, and (10) seedlings (Liang et al., 2011). *In silico* mining identified 1244 SSRs, for which candidate primers were designed by Clemson University Genomics Institute (CUGI). Another cDNA library was also constructed from secondary xylem during the early stages of tension wood formation, which contained 1,733 unigenes (Jin et al., 2011).

Table 1. Sixty-six polymorphic yellow-poplar EST-SSR markers that identified by Xu et al. (2006; 2010)

No.	Locus	Forward primer sequence(5'--3')	Reverse primer sequence(3'--5')	Repeat motif	Unigene	Expected size
1	LT002	CCTACCACCAGCAATACCTA	TCTCGTCGCTGAAGATATG	(GCA)8	247326	189
2	LT005	GTTTCTCATTTCCACCTCTG	TCCTTACACGAACCTGATCT	(AAGGA)4	247498	258
3	LT009	GAGGAGAGCCAATACACC	CAATGTAGTAGGGGATATGA	(CT)11	247603	173
4	LT013	CATGTCTGGTGGAAGAGAAT	CCATGAGAAGAGGATGAAAC	(GA)17	247671	271
5	LT015	TCCGTTATCTCTCTCAAAA	CTAGACAGGTGCTCGGATAC	(CCGAAC)5	247687	110
6	LT017	AAAGAAATGCCCATCCAC	CCTCGAAATATCCACTAACG	(CTT)8	247745	247
7	LT018	GGACCTATTCCGTCACTACA	GAAACAAAGACGTTCACCAT	(TCTT)5	247752	208
8	LT020	TCCCTTGACTGAGAGAGAGA	CTGTCTAGCCTCCTTGCTTA	(GA)15	247804	240
9	LT021	CAAATACCATTGCACCTTGT	ACGCATCCTCTTCCACTAC	(TTC)8	247804	180
10	LT022	AAACGTCTTCATGTGGAACT	CCTCACCTCAAATCCATTC	(AG)17	247805	139
11	LT023	TGATAGATATGGAGGGTGGA	TGAAGACGAGTTCCCAGTAT	(TTCGTC)5	247852	161
12	LT026	ACCCTGTGTGAGGTTGATAA	TTTTTGTGAGAGCTAGTGTCC	(ATG)7	247927	232
13	LT028	GACAGACCACACTCCATTTT	GATGTTTCCTTTCCCCTATC	(CT)17	248012	105
14	LT031	TGAAGAACCCAACAACTCTC	GTCGTAGCAGGTAGGTATGC	(GA)18	248063	203
15	LT037	CCCTAAATTCTCATCACACC	CCAGATCGTCTTGTTCTCAT	(GCA)12	248227	268
16	LT040	CCTGTGGATAAACTAGCTGAA	CTCTCCTTCCTCTTCCTCTC	(GAA)6	248316	180
17	LT045	TACTCTTCGCAAGCTCTTTT	CACAAGATTCCCATCAGTTT	(CTC)7	248480	271
18	LT048	CCTCTCCCACTCTTGAAA	TTGAGTTTGGATCTTTGACC	(AG)13	248535	249
19	LT051	GGTGAACTCCTTCAACACTC	CCTAACAGGGGATTTTATCA	(AG)24	248661	262
20	LT055	CTCTCTACCGATCCCTCTCT	GCTCATTCTCTTGTTTCCAC	(GAA)8	248713	281
21	LT056	CTTGGGTTCTTTATGCAGTC	TCTGTAGCATCTCCTTGACC	(CT)26	248760	223
22	LT057	CATGGTGGACATCACATAAA	CTTACAGGCAAATCTTACAGG	(CA)11	248773	299
23	LT058	TTAAAATGGAGGAACGAGAG	GTAGAGGCTTCGAGTTTGTG	(GA)9	248788	208
24	LT059	GGCAAAGAGATGTGATTTTG	ACAATCTTCACCAGTGTCCT	(TC)15	248853	254

25	LT061	CTTCGATCCTGAAATCGTAT	GAGCGAGAGAGAGAGAAGAA	(TCTG)5	248879	210
26	LT066	TACTGAGAGAGGGAGAGAGG	ACTGCTCATTTAGACGATCC	(AG)9	249185	269
27	LT067	TGCATTTGGTTCTCTCTTCT	GAGGGGGTTTTATTTTTCTG	(GCA)8	249243	286
28	LT070	CAAGCAAAGGTGTCTGTCTC	AATGCGACTGTTGGTTTTAC	(TC)18	249523	149
29	LT071	GCGCTTCCTCTAAAATCTCT	CCAAATCCATGCAACATC	(TC)17	249542	154
30	LT073	ACTTTTTCTCCACCGACTG	ATTGGATGGCTAGAGTGAAA	(TC)18	249565	235
31	LT075	GGTCGTTCTCTCTGTCTCTC	ACCAAATCAGTCATGCTCTT	(CTT)6	249574	226
32	LT076	TGTCCAACAATCCAAAAGTC	AGTACAGTTGTCGCAATTCA	(ATT)8	249607	112
33	LT077	TACCAGCCATTGAAGAGTTT	CGATTACAGAAGCAACAACA	(GA)19	249636	293
34	LT079	AGAGAGAGGGAGGGAGAAG	GCCTGTATGTTGGGTAAAAA	(AG)20	249840	294
35	LT081	GCAAGGCTAGTGAAAGACTG	GAGTCACCGAAGACAAAGAG	(TCA)6	249946	181
36	LT082	CGTTTTCTTGCTAGGGTTTA	CTAACGTAGAGGGGCTTGAT	(AGC)6	250004	228
37	LT086	AAGACAGGACTTTCCACTGA	GAACGAACCTAACCAAATGA	(CTT)10	250300	274
38	LT090	TGCTTTACCTGAGCATCTCT	GACGAGAACCTGTAGCACAC	(AT)9	250477	210
39	LT091	ATTTTCGTGTGCTACAGGTT	GGAAGGATGTTGGTTAGACA	(TC)19	250477	193
40	LT092	GGGGTTTTGCTTAATGTGA	CATTCCCTACCTCCTTCTCT	(GGAGCC)4	250526	229
41	LT096	TGCAACCTAACAAGATGTGT	TGAAAAGCAACCAAGTTACC	(CT)20	250709	272
42	LT101	CCACAGGTTTTCTTCATTTC	CGCATTGGATCTTCATCTTA	(CT)10	250850	404
43	LT102	GGAAACCAAACACAATCACT	TCCGTCACCACTAATCTCTC	(GA)9	250871	163
44	LT103	CCTCTCCCTCTCTCATTTCT	CGATGGTATCCAAACACAA	(CT)15	250979	232
45	LT105	TCCGAGACATCTAATCAACA	AAACTCCCAGGAACAAATCT	(TC)19	251003	108
46	LT111	ACGACCAGATGGCTATAATG	AGTCTACACAGGGAGAGAGC	(TCT)7	251113	229
47	LT113	CCAAGTGAAAATCAACTCCT	ATCTCGACGGTGTTCTGAT	(CT)18	251149	252
48	LT115	CTCTCATTCCGACCTTCATA	ACTTTTCCTGCAACTACTGC	(TCA)9	251177	129
49	LT117	GGGTACATGAGTTGGGTACT	GGGAGTTCCTTAGCCTTATC	(CT)27	251237	196
50	LT120	CCTTTTCTCAATGTCCTGAA	CACAGACTCCCAAACCTTAC	(AG)14	251410	167

51	LT121	GCATGAAATCCAAAGAAGAG	CTGCGAAAGAAGAGAAGAAG	(TC)23	251431	152
52	LT124	TTAAAACTGGGATCTGCACT	AACCCACAAACATCAGACAT	(TAT)11	251477	127
53	LT125	GTCCAAGATCAAGGGTAGTG	TAGATGGATTGACCCACTTG	(TC)15	251589	274
54	LT127	GTTGGGTTCATGTTTATGGT	GGAGGAAAATCACAGTATCG	(TC)11	251695	199
55	LT131	GCAGCATCTCCTCATATTCT	TTGCAGTTGAGCTATTGTTG	(AC)22	251877	240
56	LT135	CCCTCCAGAGAGAGACTTTT	CTCTTCCCTTTCACCATCTT	(AG)13	251931	101
57	LT137	ATACCTTCACCCAACCTGAT	GGATTGACCAACACTCAAAT	(ATT)7	251951	257
58	LT138	AAACCCATCTTTCTCCTTTC	AGCCCATATTTCTTCACCTT	(GA)16	251970	245
59	LT139	CTAGAAGGTGGATTGTGTACG	ACTGCTATAAGGGCATATCA	(TTTC)5	251972	226
60	LT141	CCCTGTAAATAACCCAATCA	CCGTTCTCTCCTTCTTCTCT	(CT)14	252005	143
61	LT150	TGGGTAGGGTCTAAGTTGTG	CCTTGTCTCAAAATGGTTGT	(TC)22	252115	297
62	LT152	GCTGCTTCTTCTTTCATCTT	GGAACTTGTTGCTGGTGTAG	(AG)9	252137	194
63	LT157	AGTTGCCCTTTAGCTTCTTT	GCCACAGAGTTTTGGAAGTA	(TTC)6	252634	222
64	LT158	ACTGTTCGATGAAATGTTCC	TATCGGAGGAGTTTCTCTTG	(GCG)8	252745	167
65	LT161	AGCCTTCTTCTCCATCTCTT	TCGGATTATGGTGTTTATGG	(CCATCT)6	252990	122
66	LT170	GACGATGTTGTTCTTGGAGT	CAGACAGAAGCGAGTAGAGG	(CAG)6	253417	235

We are interested in developing yellow-poplar as a new tree model research system for comparative genomics of secondary cell wall formation. Primary cell walls, which exist inside the intercellular layer, mainly contain cellulose and a small amount of hemicelluloses and pectin. Primary cell walls are usually as thin as 1-3 micrometer, with soft texture and great plasticity, which allow cell expansion and division in young plants. All plant cells have primary cell walls. Secondary cell walls appear at maturity, and exist inside primary cell walls. Secondary cell walls are usually 5-10 micrometer, and their structures are cellulose microfibrils matrix, commonly filled with hemicelluloses and lignin. Not all the plant cells have secondary cell walls, and most of the plant cells with secondary cell wall will experience protoplast death. The remaining secondary cell walls provide physical support and protection to the whole plant. Their cross-linked structures make secondary cell walls rigid and water resistant; as a result, they can function as water and nutrition transportation corridors. Secondary cell walls also function in herbivore and microbe defense. As mentioned above, yellow-poplar has great economic value as a productive source of industrial raw material (Hernandez et al., 1997; Moody et al., 1993; Williams and Feist, 2004). Its lignocellulosic biomass, which is largely stored in its secondary cell walls, has been used in paper pulp production. In this traditional fossil energy shortage age, alternative energy source development is urgently needed for human sustainable development. Lignocellulosic biomass is an ideal source for biofuel

production, because it is not related to food production, unlike crops, the use of which in biofuel production has raised concerns about food/energy production balance.

Although lignin is important to plant structure support and protection, it is undesired in many industrial procedures. Lignin is highly resistant to both mechanical disruption and enzymatic degradation, and makes lignocellulosic biomass recalcitrant to industrial treatment. As a result, a better understanding of lignin biosynthesis and secondary cell wall formation would help us improve lignocellulosic biomass as a better resource for industrial production. Yellow-poplar is a potential tree model research system for comparative genomics of secondary cell wall formation because of its fast growing ability, strong resilience and survival capabilities, lignocellulosic biomass richness, and its phylogenetic position as a basal angiosperm species.

Applications of molecular markers in forest tree species

Genetic linkage map represents a linear map of genes and other DNA markers along a chromosome. It is based on chromosomal exchanges and recombination frequency, which reflects the mapped distances. A detailed genetic linkage map can be an efficient

tool for molecular evolution studies, QTL positioning, genetic factor identification, and breeding (Kim et al., 2012a). It is the first prerequisite for map-based gene cloning and a cornerstone of genomics. Consequently, many linkage maps of different plant species have been developed, including tomato (Tanksley et al., 1992), potato (Tanksley et al., 1992), *Arabidopsis thaliana* (Cho et al., 1999), peach (Dettori et al., 2001), perennial ryegrass (Jones et al., 2002), olive (de la Rosa et al., 2003), hexaploid wheat (Akbari et al., 2006), diploid *Fragaria* (Sargent et al., 2006), *Rubus idaeus* (Sargent et al., 2007; Fernández-Fernández et al., 2011), cultivated grapevine (Vezzulli et al., 2008), *Prunus* (Illa et al., 2009), domesticated apple (Velasco et al., 2010), and sweet cherry (Klagges et al., 2013).

Traditianal genetic linkage mapping with phenotypes has conspicuous defects, including a low quantity of available markers and severe dependence on gene expression. Forest tree species usually have a long life cycle, including a long juvenile stage, requiring much time to wait for phenotype development to allow for phenotype-based selection. Molecular markers can overcome this problem, and have proven valuable in plant breeding, as well as in studies of genetic diversity, genome mapping, gene tagging, phylogeny, and evolution (Reddy et al., 2002).

In the 1970s, restriction fragment length polymorphisms (RFLPs) began to be recognized and utilized as valuable genetic mapping markers (Botstein et al., 1980). Digestion of genomic DNA with a restriction enzyme generates DNA fragments of different lengths, which are then separated electropheretically and transferred to a membrane. Hybridization with randomly chosen DNA fragments, probes, produces codomenant banding pattern on membrane. Fragment number and length depend on the distribution of restriction endonuclease cleavage sites. As a result, differences induced by evolutional mutation on those sites can result in length changes of the DNA fragments, and can be used as DNA molecular markers. RFLPs are suitable for species maps, because the same RFLP hybridization probes are restriction enzymes that can be used in related species (Devey et al., 1994). As a result, RFLPs have been used in linkage map construction for some tree species, including poplar (Bahrman and Damerval, 1989) and eucalyptus (Byrne et al., 1995). However, the application of RFLPs is labor intensive and requires development of specific probes, and radioactive or fluorescent labeling.

In the last 20 years, many Polymerase Chain Reaction (PCR)-based DNA molecular markers have emerged and been widely used, such as randomly amplified polymorphic DNA (RAPD) analysis (van Heusden and Bachmann, 1992; Lynch and Milligan, 1994), amplified fragment length polymorphism (AFLP) analysis (Vos et al., 1995), simple

sequence repeat (SSR, also called microsatellite) analysis (Akkaya et al., 1992; Gupta and Varshney, 2000), and single nucleotide polymorphism (SNP) analysis (Jordan and Humphries, 1994). RAPDs are PCR amplified DNA fragments using arbitrary 10-base oligonucleotides as primers. The detected polymorphisms result from individual sequence changes in the primer binding sites (Williams et al., 1990; Welsh et al., 1992). RAPDs require small amounts of DNA and no prior knowledge of sequence, and have the advantages of rapid polymorphism screening and efficient generation of large number of marker, making it a very powerful tool to construct linkage maps. RAPDs have been used for genetic map construction of white spruce (Tulsieram et al., 1992), slash pine (Nelson et al., 1993), longleaf pine (Nelson et al., 1994), norway spruce (Binelli and Bucci, 1994), and maritime pine (Plomion et al., 1995a; Plomion et al., 1995b). RAPDs have also been used as molecular markers in gene identification and genetic analysis. For example, the resistance gene to white pine blister rust in sugar pine was linked to six RAPD markers (Devey et al., 1995).

AFLP technology is a combination of PCR amplification and RFLP technology. AFLPs are selective PCR amplifications of restriction endonuclease digested genomic DNA fragments, which show reproducible fingerprint after electrophoretic separation and hybridization with probe (Mueller and Wolfenbarger, 1999). They have been widely used

in linkage map development (Vaneck et al., 1995; vanderVoort et al., 1997; Saliba-Colombani et al., 2000), full sib exclusion (Gerber et al., 2000), genetic diversity analysis (Mariette et al., 2002; Nybom, 2004), and many other applications.

An SNP is a single nucleotide variation in DNA sequence. Almost all SNPs have only two alleles (Vignal et al., 2002), and they occur more frequently in non-coding regions than in coding regions (Nachman, 2001). SNP frequency in forest tree species is generally as high as 1 per 100 bp (Neale, 2007; Savolainen and Pyhajarvi, 2007; Neale and Ingvarsson, 2008;). SNPs have been developed by Sanger sequencing for candidate genes in tree species, and application of Next Generation Sequencing (NGS) will surely make SNP development much faster and easier (Neale and Kremer, 2011). They are primarily used in development of highly saturated linkage maps and genome-wide association studies (GWAS) in tree species because of their high frequency and availability of more efficient, cost effective and rapid sequencing methods. SNPs are promising molecular markers in forest tree species studies, even given their bi-allelic and the uneven distribution characterizations, which make them less informative than SSRs (Slate et al., 2009; Ball et al., 2010).

Simple sequence repeats (SSRs), also called microsatellites, variable number of tandem repeats (VNTRs), or short tandem repeat polymorphism (STRP), are simple

duplication of 2-6 base pairs of DNA. They are widely distributed throughout plant and animal genomes (Tautz and Renz, 1984; Turnpenny and Ellard, 2005). Different repeat numbers result in extremely rich inter- and intra- polymorphisms (Queller et al., 1993), making SSRs valuable molecular markers for plant and animal genetic mapping. Polymorphism between different individuals is varieties derived from the difference in the length of their repeating region lengths, and not from variability in upstream or downstream DNA sequences. The upstream or downstream DNA sequences are highly conserved (Jarne and Lagoda, 1996; Queller et al., 1993). (AT) n and (TA)n repeats are the most abundant, while (AG)n, (GA)n, (CT)n, (TC)n, (AC)n, and (CA)n repeats usually show higher level of polymorphism than other primers (Reddy et al., 2002). Compared with other molecular markers, SSRs are highly reproducible and polymorphic, and exhibit high degree of heterozygosity, co-dominant inheritance, and even distribution along chromosomes (Cuadrado and Schwarzacher, 1998; Xu et al., 2010). Primers developed for SSRs are usually 16-25 mers long, permitting higher annealing temperature (ranges from 45 to 65 ^{o}C), which subsequently leads to higher stringency (Reddy et al., 2002).

Genomic fingerprinting, also called DNA profiling, represents individual specific DNA polymorphism, which has been successfully used in germplasm characterization

and varieties/hybrids/parental identification (Reddy et al., 2002). Molecular markers are efficient tools in genetic diversity and phylogenic analysis in forest tree species. RFLPs and RAPDs have also been used in genomic fingerprinting, but SSRs are perhaps more useful because of the multi-allelic nature for a given SSR and the high reproducibility (Rauscher and Simko, 2013). RFLPs were successfully used to detect cultivar variation in apple (*Malus×domestica* Borkh., Watillon et al., 1991), but failed to differentiate between different sports of the apple "Red Delicious" (Nybom, 1990). Thirty-four *Prunus persica* L. cultivars were successfully identified with nine RFLP fragment probes (Rajapakse et al., 1995), and fifty-two *Prunus armeniaca* L. were identified with thirty-one probes (de Vicente et al., 1998).

Because of its simplicity, short experimental time, and low cost, RAPDs are still used in genomic fingerprinting and genetic diversity analysis as well (Wunsch and Hormaza, 2002). They have been used in fingerprinting and genetic similarity analysis for apple (Koller et al., 1993; Mulcahy et al., 1993; Landry et al., 1994;; Tancred et al., 1994; Autio et al., 1998; Oraguzie et al., 2001), pear (Oliveira et al., 1999; Monte-Corvo et al., 2000), peach (Lu et al., 1996; Warburton and Bliss, 1996; Casas et al., 1999), plum (Ortiz et al., 1997; Shimada et al., 1999), sweet cherry (Gerlach and Stosser, 1997), apricot (Shimada et al., 1994; Takeda et al., 1998), olive (Fabbri et al., 1995; Wiesman et

al., 1998; Mekuria et al., 1999; Claros et al., 2000; Belaj et al., 2001; Besnard et al., 2001; Sanz-Cortes et al., 2001), walnut (Nicese et al., 1998) and chestnut (Galderisi et al., 1998; Oraguzie et al., 1998). SSRs have been used to fingerprint a cocoa collection (Charters and Wilkinson, 2000), distinguish various chrysanthemum cultivars (Wolff et al., 1995), and discriminate fourteen rice varieties cultivated in India (Sarao et al., 2010), and eight closely related wheat cultivars (Zhu et al., 2011). SSRs have been successfully used both to examine genetic diversity and to investigate phylogeny for finger millet (Salimath et al., 1995), wheat (Nagaoka and Ogihara, 1997), rice (Joshi et al., 2000), Vigna (Ajibade et al., 2000) and *Diplotaxis* (Martin and Sanchez-Yelamo, 2000). Both AFLPs and SSRs were used in genetic diversity analysis in the mangrove species *Avicennia marina* (Maguire et al., 2002).

DNA markers that are closely linked to valuable trait QTLs, can contribute greatly to the tree species improvement. Some of the most intensively studied valuable traits include growth and biomass, biotic or abiotic stress resistance, and wood properties (Neale and Kremer, 2011). In peach, thirty-four AFLPs and three SSRs markers were used to identify sequence tagged sites (STSs) for *br* gene, responsible for architectural properties of peach canopy, where *brbr* homologous peaches exhibit pillar growth habit (Sajer et al., 2012). In chickpea, two SSR markers, UBC 855_{500} and UBC 825_{1200}, were

tagged to the *Fusarium* wilt race 4 resistance gene (Ratnaparkhe et al., 1998). In perennial ryegrass (*Lolium perenne*), four QTLs are associated with crown rust (*Puccinia coronata* f. sp. *lolii*) resistance, and were found to closely link to several different AFLPs, SSRs, RFLPs and STSs markers (Muylle et al., 2005). QTL mapping has been applied in forest tree studies for more than 20 years; however, positional cloning of genes residing in QTLs has not been undertaken because of the large size of tree genomes and expensive cost. Trait associated DNA markers can help in early breeding selection (Neale and Kremer, 2011).

Project objectives

Liriodendron presents a suitable model for mating system, systemic evolution and population genetics studies, and has been deeply studied as a candidate for comparative studies and evolution of angiosperms. In addition, yellow-poplar has great economic and ecological values. However, the genome of yellow-poplar has not been sequenced, and less than 200 SSR markers have been characterized in *Liriodendron*, indicating that more

informative markers are needed for applications, such as linkage map construction, molecular breeding, trait improvement, and other studies.

The specific objectives of this project include:

1. To develop informative SSR markers for construction of the first genetic linkage map for yellow-poplar. Such linkage maps are essential for future molecular breeding and QTL mapping, and as a framework for sequencing the *Liriodendron* genome in the future.

2. To investigate the genetic composition of two yellow-poplar breeding orchards; one in Clemson University, South Carolina, and the other one in University of Tennessee, Tennessee. This would provide a first look at the genetic diversity and allele richness among selections of this unique native species.

REFERENCES

Ajibade, S.R., Weeden, N.F., and Chite, S.M. (2000). Inter simple sequence repeat analysis of genetic relationships in the genus Vigna. Euphytica 111, 47-55.

Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H., et al. (2006). Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. TAG 113, 1409-1420.

Akkaya, M.S., Bhagwat, A.A., and Cregan, P.B. (1992). Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132, 1131-1139.

Albert, V.A., Soltis, D.E., Carlson, J.E., Farmerie, W.G., Wall, P.K., Ilut, D.C., Solow, T.M., Mueller, L.A., Landherr, L.L., Hu, Y., et al. (2005). Floral gene resources from basal angiosperms for comparative genomics research. BMC Plant Biol 5.

Antolovich, M., Bedgood, D.R., Bishop, A.G., Jardine, D., Prenzler, P.D., and Robards, K. (2004). LC-MS investigation of oxidation products of phenolic antioxidants. J Agr Food Chem 52, 962-971.

Arús P, Yamamoto T, Dirlewanger E, Abbott AG (2005). Synteny in the Rosaceae. Plant Breeding Reviews 27, 175-211.

Autio, W.R., Schupp, J.R., Ferree, D.C., Glavin, R., and Mulcahy, D.L. (1998). Application of RAPDs to DNA extracted from apple rootstocks. Hortscience 33, 333-335.

Bahrman, N., and Damerval, C. (1989). Linkage Relationships Of Loci Controlling Protein Amounts In Maritime Pine (*Pinus pinaster* Ait). Heredity 63, 267-274.

Ball, A.D., Stapley, J., Dawson, D.A., Birkhead, T.R., Burke, T., and Slate, J. (2010). A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (*Taeniopygia guttata*). BMC genomics 11, 218.

Belaj, A., Trujillo, I., de la Rosa, R., Rallo, L., and Gimenez, M.J. (2001). Polymorphism and discrimination capacity of randomly amplified polymorphic markers in an olive germplasm bank. J Am Soc Hortic Sci 126, 64-71.

Berlin, A., Maximenko, V., Bura, R., Kang, K.Y., Gilkes, N., and Saddler, J. (2006). A rapid microassay to evaluate enzymatic hydrolysis of lignocellulosic substrates. Biotechnol Bioeng 93, 880-886.

Besnard, G., Baradat, P., and Berville, A. (2001). Genetic relationships in the olive (*Olea europaea* L.) reflect multilocal selection of cultivars. TAG 102, 251-258.

Binelli, G., and Bucci, G. (1994). A Genetic-Linkage Map Of Picea-Abies Karst, Based on Rapd Markers, as a Tool In Population-Genetics. TAG 88, 283-288.

Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics 32, 314-331.

Burns, R., and Honkala, B. (1990). Silvics of North America, Vol 2 (United States Department of Agriculture (USDA) Forest Service).

Byrne, M., Murrell, J.C., Allen, B., and Moran, G.F. (1995). An Integrated Genetic-Linkage Map for Eucalypts Using Rflp, Rapd And Isozyme Markers. TAG 91, 869-875.

Casas, A.M., Igartua, E., Balaguer, G., and Moreno, M.A. (1999). Genetic diversity of Prunus rootstocks analyzed by RAPD markers. Euphytica 110, 139-149.

Celen, I., Harper, D., and Labbe, N. (2008). A multivariate approach to the acetylated poplar wood samples by near infrared spectroscopy. Holzforschung 62, 189-196.

Charters, Y.M., and Wilkinson, M.J. (2000). The use of self-pollinated progenies as 'in-groups' for the genetic characterization of cocoa germplasm. TAG 100, 160-166.

Chen, C., Liu, T., Tseng, W., Lu, F., Hung, R., Chen, C., and Chen, C. (2008). (-)-anonaine induces apoptosis through Bax- and caspase-dependent pathways in human cervical cancer (HeLa) cells. Food Chem Toxicol 46, 2694-2702.

Chen, C., Wang, Y., Juan, S., and Huang, J. (2012). Chemical Constituents From the Stems Of *Liriodendron Tulipifera*. Chem Nat Compd 47, 1035-1037.

Chi, T.C., Lee, S.S., and Su, M.J. (2006). Antihyperglycemic effect of aporphines and their derivatives in normal and diabetic rats. Planta Med 72, 1175-1180.

Cho, R.J., Mindrinos, M., Richards, D.R., Sapolsky, R.J., Anderson, M., Drenkard, E., Dewdney, J., Reuber, T.L., Stammers, M., Federspiel, N., et al. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nature genetics 23, 203-207.

Claros, M.G., Crespillo, R., Aguilar, M.L., and Canovas, F.M. (2000). DNA fingerprinting and classification of geographically related genotypes of olive-tree (*Olea europaea* L.). Euphytica 116, 131-142.

Cuadrado, A., and Schwarzacher, T. (1998). The chromosomal organization of simple sequence repeats in wheat and rye genomes. Chromosoma 107, 587-594.

De Craene, L.P.R., Soltis, P.S., and Soltis, D.E. (2003). Evolution of floral structures in basal angiosperms. Int J Plant Sci 164, S329-S363.

de la Rosa, R., Angiolillo, A., Guerrero, C., Pellegrini, M., Rallo, L., Besnard, G., Berville, A., Martin, A., and Baldoni, L. (2003). A first linkage map of olive (*Olea europaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. TAG 106, 1273-1282.

de Vicente, M.C., Truco, M.J., Egea, J., Burgos, L., and Arus, P. (1998). RFLP variability in apricot (*Prunus armeniaca* L.). Plant Breeding 117, 153-158.

Dettori, M.T., Quarta, R., and Verde, I. (2001). A peach linkage map integrating RFLPs, SSRs, RAPDs, and morphological markers. Genome 44, 783-790.

Devey, M.E., Delfinomix, A., Kinloch, B.B., and Neale, D.B. (1995). Random amplified polymorphic DNA markers tightly linked to a gene for resistance to white-pine blister rust in sugar pine. P Natl Acad Sci USA 92, 2066-2070.

Devey, M.E., Groover, A.T., Jermstad, K.D., Neale, D.B., and Ahuja, M.R. (1994). mapped DNA probes from loblolly-pine can be used for

restriction-fragment-length-polymorphism mapping in other conifers. TAG 88, 279-282.

Fabbri, A., Hormaza, J.I., and Polito, V.S. (1995). Random amplified polymorphic DNA analysis of olive (*Olea Europaea* L) Cultivars. J Am Soc Hortic Sci 120, 538-542.

Fernández-Fernández, F., Antanaviciute, L., Govan, C., and Sargent, D. (2011). Development of a multiplexed microsatellite set for fingerprinting red raspberry (*Rubus idaeus*) germplasm and its transferability to other *Rubus* species. Journal of Berry Research 1, 177-187.

Galderisi, U., Cipollaro, M., Di Bernardo, C., De Masi, L., Galano, G., and Cascino, A. (1998). Molecular typing of Italian sweet chestnut cultivars by random amplified polymorphic DNA analysis. J Hortic Sci Biotech 73, 259-263.

Gaspar, E.M.M., and Dasneves, H.J.C. (1993). Steroidal constituents from mature wheat-straw. Phytochemistry 34, 523-527.

Gerber, S., Mariette, S., Streiff, R., Bodenes, C., and Kremer, A. (2000). Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. Molecular ecology 9, 1037-1048.

Gerlach, H.K., and Stosser, R. (1997). Patterns of random amplified polymorphic DNAs for sweet cherry (*Prunus avium* L.) cultivar identification. J Appl Bot-Angew Bot 71, 212-218.

Gupta, P., and Varshney, R. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113, 163-185.

Gwak, K., Kim, H., Ryu, K., Choi, H., Cho, D., Kim, P., and Choi, I. (2009). Growth improvement of *Liriodendron tulipifera* through SCB manure treatment. Forest Bioenergy 28, 7-14.

Hao, R., He, S., Tang, S., and Wu, S. (1995). Geographical distribution of *Liriodederon chinense* in China and its significance. J Plant Resour Environ 4, 1-6.

Hernandez, R., Davalos, J., Sonti, S., Kim, Y., and Moody, R. (1997). Strength and stiffness of reinforced yellow-poplar glued-laminated beams. Forest Service FPL, 554.

Hunt, D. (1998). Magnolias and Their Allies.

Hwang, S.S., Lee, S.J., Kim, H.K., Ka, J.O., Kim, K.J., and Song, H.G. (2008). Biodegradation and saccharification of wood chips of *Pinus strobus* and

Liriodendron tulipifera by white rot fungi. Journal of microbiology and biotechnology 18, 1819-1826.

Illa, E., Lambert, P., Quilot, B., Audergon, J.M., Dirlewanger, E., Howad, W., Dondini, L., Tartarini, S., Lain, O., Testolin, R., et al. (2009). Linkage map saturation, construction, and comparison in four populations of *Prunus*. J Hortic Sci Biotech, 168-175.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Muller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. PNAS 104, 19369-19374.

Jarne, P., and Lagoda, P.J. (1996). Microsatellites, from molecules to populations and back. Trends in ecology & evolution 11, 424-429.

Jin, H., Do, J., Moon, D., Noh, E.W., Kim, W., and Kwon, M. (2011). EST analysis of functional genes associated with cell wall biosynthesis and modification in the secondary xylem of the yellow poplar (*Liriodendron tulipifera*) stem during early stage of tension wood formation. Planta 234, 959-977.

Jones, E.S., Dupal, M.P., Dumsday, J.L., Hughes, L.J., and Forster, J.W. (2002). An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). TAG 105, 577-584.

Jordan, S.A., and Humphries, P. (1994). Single nucleotide polymorphism in exon 2 of the BCP gene on 7q31-q35. Human molecular genetics 3, 1915.

Joshi, S.P., Gupta, V.S., Aggarwal, R.K., Ranjekar, P.K., and Brar, D.S. (2000). Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. TAG 100, 1311-1320.

Kim, C., Zhang, D., Auckland, S.A., Rainville, L.K., Jakob, K., Kronmiller, B., Sacks, E.J., Deuter, M., and Paterson, A.H. (2012a). SSR-based genetic maps of *Miscanthus sinensis* and *M. sacchariflorus*, and their comparison to sorghum. TAG 124, 1325-1338.

Kim, K.D., and Lee, E.J. (2005). Potential tree species for use in the restoration of unsanitary landfills. Environ Manage 36, 1-14.

Kim, K.H., Moon, E., Kim, S.Y., and Leet, K.R. (2010). Lignans from the tuber-barks of *Colocasia antiquorum* var. esculenta and their antimelanogenic activity. J Agr Food Chem 58, 4779-4785.

Kim, Y.H., Lee, S.M., Lee, H.W., and Lee, J.W. (2012b). Physical and chemical characteristics of products from the torrefaction of yellow poplar (*Liriodendron tulipifera*). Bioresource technology 116, 120-125.

Klagges, C., Campoy, J.A., Quero-Garcia, J., Guzman, A., Mansur, L., Gratacos, E., Silva, H., Rosyara, U.R., Iezzoni, A., Meisel, L.A., et al. (2013). Construction and Comparative Analyses of Highly Dense Linkage Maps of Two Sweet Cherry Intra-Specific Progenies of Commercial Cultivars. PloS one 8.

Klugh, K.R., and Cumming, J.R. (2007). Variations in organic acid exudation and aluminum resistance among arbuscular mycorrhizal species colonizing *Liriodendron tulipifera*. Tree physiology 27, 1103-1112.

Koller, B., Lehmann, A., Mcdermott, J.M., and Gessler, C. (1993). Identification of apple cultivars using rapd markers. TAG 85, 901-904.

Koo, B., Min, B., Park, N., Eom, C., Yeo, H., Ryu, K., and Choi, I. (2008). Chemical and physical characterizations of *Liriodendron tulipifera* on growth periods. Proceedings of the 30th Symposium on Biotechnology for Fuels and Chemicals.

Koo, B., Park, N., Yeo, H., Lee, S., Kim, H., Kim, H., and Choi, I. (2009). Organosolv pretreatment of *Liriodendron tulipifera* with acid and alkali catalysts. Proceedings of the 30th Symposium on Biotechnology for Fuels and Chemicals.

Landry, B.S., Li, R.Q., Cheung, W.Y., and Granger, R.L. (1994). Phylogeny analysis of 25 apple rootstocks using rapd markers and tactical gene tagging. TAG 89, 847-852.

Liang, H., Barakat, A., Schlarbaum, S.E., Mandoli, D.F., and Carlson, J.E. (2010). Comparison of gene order of GIGANTEA loci in yellow-poplar, monocots, and eudicots. Genome 53, 533-544.

Liang, H., Ayyampalayam, S., Wickett, N., Barakat, A., Xu, Y., Landherr, L., Ralph, P.E., Jiao, Y., Xu, T., Schlarbaum, S.E., et al. (2011). Generation of a large-scale genomic resource for functional and comparative genomics in *Liriodendron tulipifera* L. Tree Genet Genomes 7, 941-954.

Liang, H., Carlson, J.E., Leebens-Mack, J.H., Wall, P.K., Mueller, L.A., Buzgo, M., Landherr, L.L., Hu, Y., DiLoreto, D.S., Ilut, D.C., et al. (2008). An EST database for *Liriodendron tulipifera* L. floral buds: the first EST resource for functional and comparative genomics in *Liriodendron*. Tree Genet Genomes 4, 419-433.

Liang, H., Fang, E., Tomkins, J., Luo, M., Kudrna, D., Kim, H.R., Arumuganathan, K., Zhao, S.Y., Leebens-Mack, J., Schlarbaum, S.E., et al. (2007). Development of a BAC library for yellow-poplar (*Liriodendron tulipifera*) and the identification of genes associated with flower development and lignin biosynthesis. Tree Genet Genomes 3, 215-225.

Lin, I.J, Lo, W.L., Chia, Y.C., Huang, L.Y., Cham, T.M., Tseng, W.S., Yeh, Y.T., Yeh, H.C., Wang, Y.D., and Chen, C.Y. (2010). Isolation of new esters from the stems of *Cinnamomum reticulatum* Hay. Nat Prod Res 24, 775-780.

Liu, B., Zhang, T., Zhang, X., Ye, W., and Y, L. (2010). Chemical constituents of Laggera pterodonta. Jounal of Traditional Chinese Medicine (Zhongguo Zhong Yao Za Zhi) 35, 602-606.

Liu, R., Mei, C., Shao, F., Ren, G., Huang, H., Chen, S., and Yang, W. (2009). Studies on the chemical constituents from *Daphne tangutica*. Chinese herbal medicine (Zhong Yao Cai) 32, 1846-1847.

Lu, Z., Reighard, G.L., Baird, W.V., Abbott, A.G., and Rajapakse, S. (1996). Identification of peach rootstock cultivars by RAPD markers. Hortscience 31, 127-129.

Lynch, M., and Milligan, B.G. (1994). Analysis of population genetic structure with RAPD markers. Molecular ecology 3, 91-99.

Maguire, T.L., Peakall, R., and Saenger, P. (2002). Comparative analysis of genetic diversity in the mangrove species *Avicennia marina* (Forsk.) Vierh. (*Avicenniaceae*) detected by AFLPs and SSRs. TAG 104, 388-398.

Mariette, S., Le Corre, V., Austerlitz, F., and Kremer, A. (2002). Sampling within the genome for measuring within-population diversity: trade-offs between markers. Molecular ecology 11, 1145-1156.

Martin, J.P., and Sanchez-Yelamo, M.D. (2000). Genetic relationships among species of the genus *Diplotaxis* (*Brassicaceae*) using inter-simple sequence repeat markers. TAG 101, 1234-1241.

Mekuria, G.T., Collins, G.G., and Sedgley, M. (1999). Genetic variability between different accessions of some common commercial olive cultivars. J Hortic Sci Biotech 74, 309-314.

Monte-Corvo, L., Cabrita, L., Oliveira, C., and Leitao, J. (2000). Assessment of genetic relationships among Pyrus species and cultivars using AFLP and RAPD markers. Genet Resour Crop Ev 47, 257-265.

Moody, R., Hernandez, R., Davalos, J., and Sonti, S. (1993). Yellow poplar glulam timber beam performance. Forest Service FPL, 520.

Moon, M.K., Oh, H.M., Kwon, B.M., Baek, N.I., Kim, S.H., Kim, J.S., and Kim, D.K. (2007). Farnesyl protein transferase and tumor cell growth inhibitory activities of lipiferolide isolated from *Liriodendron tulipifera*. Archives of pharmacal research 30, 299-302.

Moore, M.J., Bell, C.D., Soltis, P.S., and Soltis, D.E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. PNAS 104, 19363-19368.

Mueller, U.G., and Wolfenbarger, L.L. (1999). AFLP genotyping and fingerprinting. Trends in ecology & evolution 14, 389-394.

Mulcahy, D.L., Cresti, M., Sansavini, S., Douglas, G.C., Linskens, H.F., Mulcahy, G.B., Vignani, R., and Pancaldi, M. (1993). The use of random amplified polymorphic DNAs to fingerprint apple genotypes. Sci Hortic-Amsterdam 54, 89-96.

Muylle, H., Baert, J., Van Bockstaele, E., Pertijs, J., and Roldan-Ruiz, I. (2005). Four QTLs determine crown rust (*Puccinia coronata* f. sp lolii) resistance in a perennial ryegrass (*Lolium perenne*) population. Heredity 95, 348-357.

Nachman, M.W. (2001). Single nucleotide polymorphisms and recombination rate in humans. TIG 17, 481-485.

Nagaoka, T., and Ogihara, Y. (1997). Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. TAG 94, 597-602.

Neale, D.B. (2007). Genomics to tree breeding and forest health. Current opinion in genetics & development 17, 539-544.

Neale, D.B., and Ingvarsson, P.K. (2008). Population, quantitative and comparative genomics of adaptation in forest trees. Current opinion in plant biology 11, 149-155.

Neale, D.B., and Kremer, A. (2011). Forest tree genomics: growing resources and applications. Nature reviews Genetics 12, 111-122.

Nelson, C.D., Kubisiak, T.L., Stine, M., and Nance, W.L. (1994). A genetic-linkage map of longleaf pine (*Pinus palustris* Mill) based on random amplified polymorphic DNAs. J Hered 85, 433-439.

Nelson, C.D., Nance, W.L., and Doudrick, R.L. (1993). A partial genetic-linkage map of slash pine (*Pinus elliottii* Engelm. var. *elliottii*) based on random amplified polymorphic DNAs. TAG 87, 145-151.

Nicese, F.P., Hormaza, J.I., and McGranahan, G.H. (1998). Molecular characterization and genetic relatedness among walnut (*Juglans regia* L.) genotypes based on RAPD markers. Euphytica 101, 199-206.

Nybom, H. (1990). DNA fingerprints in sports of red delicious apples. Hortscience 25, 1641-1642.

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molecular ecology 13, 1143-1155.

Oliveira, C.M., Mota, M., Monte-Corvo, L., Goulao, L., and Silva, D.M. (1999). Molecular typing of *Pyrus* based on RAPD markers. Sci Hortic-Amsterdam 79, 163-174.

Oraguzie, N.C., Gardiner, S.E., Basset, H.C.M., Stefanati, M., Ball, R.D., Bus, V.G.M., and White, A.G. (2001). Genetic diversity and relationships in *Malus* sp. germplasm collections as determined by randomly amplified polymorphic DNA. J Am Soc Hortic Sci 126, 318-328.

Oraguzie, N.C., McNeil, D.L., Paterson, A.M., and Chapman, H. (1998). Comparison of RAPD and morpho-nut markers for revealing genetic relationships between chestnut

species (*Castanea* spp.) and New Zealand chestnut selections. New Zeal J Crop Hort 26, 109-115.

Ortiz, A., Renaud, R., Calzada, I., and Ritter, E. (1997). Analysis of plum cultivars with RAPD markers. J Hortic Sci 72, 1-9.

Parks, C.R., and Wendel, J.F. (1990). Molecular divergence between Asian And North-American species of *Liriodendron* (Magnoliaceae) with Implications for Interpretation Of Fossil Floras. American journal of botany 77, 1243-1256.

Piao, Y.Z., Kim, Y.J., Kim, Y.A., Lee, H.S., Hammock, B.D., and Lee, Y.T. (2009). Development of ELISAs for the class-specific determination of organophosphorus pesticides. J Agr Food Chem 57, 10004-10013.

Plomion, C., Bahrman, N., Durel, C.E., and Omalley, D.M. (1995a). Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. Heredity 74, 661-668.

Plomion, C., Omalley, D.M., and Durel, C.E. (1995b). Genomic analysis in maritime pine (*Pinus pinaster*) - comparison of 2 RAPD maps using selfed and open-pollinated seeds of the same individual. TAG 90, 1028-1034.

Qiu, Y.L., Dombrovska, O., Lee, J., Li, L.B., Whitlock, B.A., Bernasconi-Quadroni, F., Rest, J.S., Davis, C.C., Borsch, T., Hilu, K.W., et al. (2005). Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. Int J Plant Sci 166, 815-842.

Queller, D.C., Strassmann, J.E., and Hughes, C.R. (1993). Microsatellites and kinship. Trends in ecology & evolution 8, 285-288.

Rajapakse, S., Belthoff, L.E., He, G., Estager, A.E., Scorza, R., Verde, I., Ballard, R.E., Baird, W.V., Callahan, A., Monet, R., et al. (1995). Genetic-linkage mapping in peach using morphological, RFLP and RAPD markers. TAG 90, 503-510.

Ratnaparkhe, M.B., Santra, D.K., Tullu, A., and Muehlbauer, F.J. (1998). Inheritance of inter-simple-sequence-repeat polymorphisms and linkage with a fusarium wilt resistance gene in chickpea. TAG 96, 348-353.

Rauscher, G., and Simko, I. (2013). Development of genomic SSR markers for fingerprinting lettuce (*Lactuca sativa* L.) cultivars and mapping genes. Bmc Plant Biol 13, 11.

Reddy, M.P., Sarla, N., and Siddiq, E.A. (2002). Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica 128, 9-17.

Rojas, I.S., Lotina-Hennsen, B., and Mata, R. (2000). Effect of lichen metabolites on thylakoid electron transport and photophosphorylation in isolated spinach chloroplasts. J Nat Prod 63, 1396-1399.

Sajer, O., Scorza, R., Dardick, C., Zhebentyayeva, T., Abbott, A.G., and Horn, R. (2012). Development of sequence-tagged site markers linked to the pillar growth type in peach (*Prunus persica*). Plant Breeding 131, 186-192.

Saliba-Colombani, V., Causse, M., Gervais, L., and Philouze, J. (2000). Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. Genome 43, 29-40.

Salimath, S.S., Deoliveira, A.C., Godwin, I.D., and Bennetzen, J.L. (1995). Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. Genome 38, 757-763.

Sanz-Cortes, F., Badenes, M.L., Paz, S., Iniguez, A., and Llacer, G. (2001). Molecular characterization of olive cultivars using RAPD markers. J Am Soc Hortic Sci 126, 7-12.

Sarao, N.K., Vikal, Y., Singh, K., Joshi, M.A., and Sharma, R.C. (2010). SSR marker-based DNA fingerprinting and cultivar identification of rice (*Oryza sativa* L.) in Punjab state of India. Plant Genet Resour-C 8, 42-44.

Sargent, D.J., Clarke, J., Simpson, D.W., Tobutt, K.R., Arus, P., Monfort, A., Vilanova, S., Denoyes-Rothan, B., Rousseau, M., Folta, K.M., et al. (2006). An enhanced microsatellite map of diploid Fragaria. TAG 112, 1349-1359.

Sargent, D.J., Fernandez-Fernandez, F., Rys, A., Knight, V.H., Simpson, D.W., and Tobutt, K.R. (2007). Mapping of A1 conferring resistance to the aphid Amphorophora idaei and dw (dwarfing habit) in red raspberry (*Rubus idaeus* L.) using AFLP and microsatellite markers. BMC Plant Biol 7, 15.

Savolainen, O., and Pyhajarvi, T. (2007). Genomic diversity in forest trees. Current opinion in plant biology 10, 162-167.

Sawasdee, K., Chaowasku, T., and Likhitwitayawuid, K. (2010). New neolignans and a phenylpropanoid glycoside from twigs of *Miliusa mollis*. Molecules 15, 639-648.

Shi, H., Wang, H., Wang, M., and Li, X. (2009). Antioxidant activity and chemical composition of *Torreya grandis* cv. *Merrillii* Seed. Natural product communications 4, 1565-1570.

Shimada, T., Haji, T., Yamaguchi, M., Takeda, T., Nomura, K., and Yoshida, M. (1994). Classification of mume (*Prunus mume* Sieb. et Zucc.) by RAPD assay. J Jpn Soc Hortic Sci 63, 543-551.

Shimada, T., Hayama, H., Haji, T., Yamaguchi, M., and Yoshida, M. (1999). Genetic diversity of plums characterized by random amplified polymorphic DNA (RAPD) analysis. Euphytica 109, 143-147.

Slate, J., Gratten, J., Beraldi, D., Stapley, J., Hale, M., and Pemberton, J.M. (2009). Gene mapping in the wild with SNPs: guidelines and future directions. Genetica 136, 97-107.

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., et al. (2000). Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. Bot J Linn Soc 133, 381-461.

Soltis, P.S., Soltis, D.E., Chase, M.W., Endress, P.K., and Crane, P.R. (2004). The diversification of flowering plants. Assembling the Tree Of Life, 154-167.

Takeda, T., Shimada, T., Nomura, K., Ozaki, T., Haji, T., Yamaguchi, M., and Yoshida, M. (1998). Classification of apricot varieties by RAPD analysis. J Jpn Soc Hortic Sci 67, 21-27.

Tancred, S.J., Zeppa, A.G., and Graham, G.C. (1994). The use of the PCR-RAPD technique in improving the plant variety rights description of a new Queensland Apple (*Malus domestica*) Cultivar. Aust J Exp Agr 34, 665-667.

Tanksley, S.D., Ganal, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., et al. (1992). High density molecular linkage maps of the tomato and potato genomes. Genetics 132, 1141-1160.

Tautz, D., and Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic acids research 12, 4127-4138.

Tulsieram, L.K., Glaubitz, J.C., Kiss, G., and Carlson, J.E. (1992). Single tree genetic-linkage mapping in conifers using haploid DNA from Megagametophytes. Bio-Technol 10, 686-690.

Turnpenny, P., and Ellard, S. (2005). Emerys Elements of Medical Genetics 12th ed.

van Heusden, A., and Bachmann, K. (1992). Genotype relationships in *Microseris elegans* (*Asteraceae*, *Lactuceae*) revealed by DNA amplification from arbitrary primers (RAPDs). Plant Syst Evol 179, 221-233.

vanderVoort, J.N.A.M.R., vanZandvoort, P., vanEck, H.J., Folkertsma, R.T., Hutten, R.C.B., Draaistra, J., Gommers, F.J., Jacobsen, E., Helder, J., and Bakker, J. (1997). Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol Gen Genet 255, 438-447.

Vaneck, H.J., Vandervoort, J.R., Draaistra, J., Vanzandvoort, P., Vanenckevort, E., Segers, B., Peleman, J., Jacobsen, E., Helder, J., and Bakker, J. (1995). The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. Mol Breeding 1, 397-410.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., et al. (2010). The genome of the domesticated apple (*Malus* x *domestica* Borkh.). Nature genetics 42, 833-+.

Vezzulli, S., Troggio, M., Coppola, G., Jermakow, A., Cartwright, D., Zharkikh, A., Stefanini, M., Grando, M.S., Viola, R., Adam-Blondon, A.F., et al. (2008). A reference integrated map for cultivated grapevine (*Vitis vinifera* L.) from three crosses, based on 283 SSR and 501 SNP-based markers. TAG 117, 499-511.

Vignal, A., Milan, D., SanCristobal, M., and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. GSE 34, 275-305.

Voitl, T., and von Rohr, P.R. (2008). Oxidation of lignin using aqueous polyoxometalates in the presence of alcohols. Chemsuschem 1, 763-769.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. Nucleic acids research 23, 4407-4414.

Wang, L., Zhang, C., Wang, Z., Zhang, M., and Xu, L. (2009). Five new compounds from *Dendrobium crystallinum*. J Asian Nat Prod Res 11, 903-911.

Wang, Z. (2005). Utilization and species hybridization in *Liriodendron*. Chinese Forestry Press.

Warburton, M.L., and Bliss, F.A. (1996). Genetic diversity in peach (*Prunus persica* L. batch) revealed by randomly amplified polymorphic DNA (RAPD) markers and compared to inbreeding coefficients. J Am Soc Hortic Sci 121, 1012-1019.

Watillon, B., Druart, P., Dujardin, P., Kettmann, R., Boxus, P., and Burny, A. (1991). Use of random cDNA probes to detect restriction-fragment-length-polymorphisms among apple clones. Sci Hortic-Amsterdam 46, 235-243.

Welsh, J., Chada, K., Dalal, S.S., Cheng, R., Ralph, D., and McClelland, M. (1992). Arbitrarily primed PCR fingerprinting of RNA. Nucleic acids research 20, 4965-4970.

Wiesman, Z., Avidan, N., Lavee, S., and Quebedeaux, B. (1998). Molecular characterization of common olive varieties in Israel and the West Bank using randomly amplified polymorphic DNA (RAPD) markers. J Am Soc Hortic Sci 123, 837-841.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. Nucleic acids research 18, 6531-6535.

Williams, R., and Feist, W. (2004). Durability of yellow-poplar and sweetgum and service life of finishes after long-term exposure. Forest Prod J 54, 96-101.

Wolff, K., Zietkiewicz, E., and Hofstra, H. (1995). Identification of chrysanthemum cultivars and stability of DNA fingerprint patterns. TAG 91, 439-447.

Wunsch, A., and Hormaza, J.I. (2002). Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. Euphytica 125, 59-67.

Xiang, Q., Lee, Y.Y., and Torget, R.W. (2004). Kinetics of glucose decomposition during dilute-acid hydrolysis of lignocellulosic biomass. Applied biochemistry and biotechnology 113, 1127-1138.

Xu, M., Li, H., and Zhang, B. (2006). Fifteen polymorphic simple sequence repeat markers from expressed sequence tags of *Liriodendron tulipifera*. Mol Ecol Notes 6, 728-730.

Xu, M., Sun, Y.G., and Li, H.G. (2010). EST-SSRs development and paternity analysis for *Liriodendron* spp. New Forest 40, 361-382.

Xu, Z., Qin, G., Li, X., and Xu, R. (2001). New biflavanones and bioactive compounds from Stellera chamaejasme L. Pharmaceutical Journal (Yao Xue Xue Bao) 36, 669-671.

Yang, A., Zhang, J., Tian, H., and Yao, X. (2012). Characterization of 39 novel EST-SSR markers for *Liriodendron tulipifera* and cross-species amplification in *L. chinense* (Magnoliaceae). American journal of botany 99, e460-464.

Yang, B., Chen, G., Song, X., Chen, Z., Song, X., and Wang, J. (2010). Chemical constituents and antimicrobial activities of Canthium horridum. Natural product communications 5, 913-914.

Zahn, L.M., Leebens-Mack, J., DePamphilis, C.W., Ma, H., and Theissen, G. (2005). To B or Not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. J Hered 96, 225-240.

Zhao, C., Shao, J., and Li, X. (2009). Chemical constituents from fruits of *Ailanthus altissima*. Jounal of Traditional Chinese Medicine (Zhongguo Zhong Yao Za Zhi) 34, 2197-2199.

Zhu, Y., Hu, J., Han, R., Wang, Y., and Zhu, S. (2011). Fingerprinting and identification of closely related wheat (*Triticum aestivum* L.) cultivars using ISSR and fluorescence-labeled TP-M13-SSR markers. Aust J Crop Sci 5, 846-850.

CHAPTER TWO

DEVELOPMENT OF FULL-SIB PROGENY AND INFORMATIVE SSR MARKERS FOR YELLOW-POPLAR GENETIC LINKAGE MAP CONSTRUCTION

Abstract

Liriodendron tulipifera L., commonly known as yellow-poplar, is a member of the Magnoliaceae family. It is a fast-growing hardwood tree species with great ecological and economic value. *Liriodendron* occupies an important phylogenetic position as a basal angiosperm and has been used in studies of the evolution of flowering plants. Genomic resources, such as EST databases and BAC libraries, have been developed for this species. However, genetic map is not available for *Liriodendron*, and very few well developed molecular markers have been available. A total of 119 informative SSR markers were identified in this project for genetic linkage map construction with an F1 progeny from #UT108A × #UT23 cross. In addition, the full-sibship for 213 seedlings was validated. These informative SSR markers and full-sib seedlings are essential in construction of

linkage maps, which would enable molecular breeding and quantitative trait locus (QTL) mapping, as well as provide a framework for sequencing of the *Liriodendron* genome.

Introduction

Forest trees amount to more than 80 percent of continental biomass (Roy et al., 2001), and provide materials for building, paper production, and biofuel production (Neale and Kremer, 2011). Yellow-poplar (*Liriodendron tulipifera* L.), which has great economic and ecological values, is a fast growing deciduous hardwood tree species found in the eastern United States, that has been a productive source of industrial raw material (Hernandez et al., 1997; Moody et al., 1993; Williams and Feist, 2004). Its lignocellulosic biomass in the secondary cell walls has been used in paper pulp production. In this fossil energy shortage age, lignocellulosic biomass of yellow-poplar and other hardwood tree species could provide an alternative energy source. Yellow-poplar has been cultivated worldwide for wood production and waste landfill remediation because of its beautiful outward appearance (especially its unique flower shape), outstanding resistance to insects, diseases, and damaging metals (Chen et al., 2012; Klugh and Cumming, 2007), barren

soil tolerance and highland growth suitability and carbon absorption capacity (Gwak et al., 2009; Kim et al., 2012). Although yellow-poplar has similar properties to poplar, such as wood structure, rapid growth, and biomass accumulation, it does not belong to *Malpighiales*, or any other eudicot or monocot; instead, it is a member of the *Magnoliaceae* family and occupies an early branch on phylogenetic tree as a basal angiosperm, with unusual flower structure (De Craene et al., 2003; Jansen et al., 2007; Liang et al., 2010; Moore et al., 2007; Qiu et al., 2005; Soltis et al., 2004; Zahn et al., 2005). These characteristics make it a candidate research model for plant evolutionary and floral structure studies.

Genomic research on hardwood tree species, including yellow-poplar, is motivated by their improvement, conservation, restoration, and population management programs (Neale and Kremer, 2011). However, unlike classical research model plants, such as *Arabidopsis thaliana* and *Nicotiana benthamiana*, or annual crops, such as maize, rice and wheat, genomic research of hardwood tree species is hindered by their long generation time and large genomes, and is restricted to highly domesticated species, including *Pinus*, *Populus*, *Eucalyptus*, and *Quercus* (Neale and Kremer, 2011). Fortunately, yellow-poplar's chromosome number, 2n=2x=38, is lowest in Magnoliaceae family, and its genome size, about 1802 Mbp per haploid genome, is also relatively small

in forest tree species (Liang et al., 2007). In recent years, several large-insert genomic libraries and expressed sequence tags (EST) datasets have become available for yellow-poplar. The bacterial artificial chromosome (BAC) library, containing 4.8× haploid genome equivalents, provides 99.2% high probability of recovering any specific sequence of interest. In addition, a small shotgun library containing 3,072 clones with an average insert size of 3 kb is also available for order at http://genome.arizona.edu/orders/. Three EST datasets consisting of 6520 unigenes, 137,504 unigenes and 1,733 unigenes, respectively, have been constructed in succession (Albert et al., 2005; Liang et al., 2008; Jin et al., 2011; Liang et al., 2011).

Besides genomic and EST libraries, molecular markers and reference genetic maps are important genomic resources. While several thousand putative EST-SSR markers being mined from the EST datasets (Liang et al., 2011; Liang et al., 2008), only 176 yellow-poplar EST-SSR markers have been characterized, and only 66 of them produced polymorphic amplification in *L. tulipifera* (Table 1; Xu et al., 2006; Xu et al., 2010). Yellow-poplar is in need of informative markers to enable linkage and QTL mapping as well as to provide the framework for sequencing of the *Liriodendron* genome. The specific objectives of the present study were to develop (1) a set of discriminating

microsatellite markers for validation of full-sibship; (2) and a set of informative markers that can be used in map construction.

Materials and Methods

1. Plant Materials and DNA isolation

Plant material, consisted of 500 seedlings, from controlled pollination between #UT108A and #UT108B as the mother trees, and #UT23 as the pollen donor, and parental trees grown at The University of Tennessee were provided by Dr. Scott Schlarbaum at The University of Tennessee. Total genomic DNA was isolated from leaf tissues using a Cetyl Trimethyl Ammonium Bromide (CTAB) protocol (Kobayashi et al., 1998). The quality and concentrations of genomic DNA from individual plants were determined with a NanoDrop 3300 (Thermo Scientific, Wilmington, DE, USA) and by electrophoresis on 0.8% agarose gels.

2. Characterization of EST-SSR markers

A total of 604 primer pairs of EST-SSR markers were synthesized by Integrated DNA Technologies. Forward primers were tailed with an M13 forward (5'-CACGACGTTGTAAAACGAC-3') at 5'-end of the forward primer. This allowed for three primer PCR reaction, with M13-tailed marker specific forward primer, marker specific reverse primer, and fluorescently labeled M13 forward primer (6-FAM, VIC, NED, or PET; Applied Biosystems, Foster City, CA, USA) (Oetting et al. 1995). These 604 markers include the sixty six EST-SSRs that have previously been characterized by electrophoresis on 8% polyacrylamide gels (Table 1; Xu et al., 2010), and 538 putative markers that were chosen from a comprehensive EST dataset (Table 2; Liang et al., 2011). The 538 markers were selected based on the following criteria: PCR amplicon size ranging from 150-350 nt, repeat number ranging from 8-30, with the sequence length at least 30 nt away from the ends of contigs, and having a melting temperature $\geq 50\ ^{0}C$.

Polymerase chain reactions (PCR) were carried out in total volume of 12.5μL, with DNA from two parents (clones 108A and 23) as templates. The reactions were set up as follows: 112.5 ng DNA template, 0.052 U/μL Promega Taq DNA polymerase, 0.16 nM forward-tailed primer, 0.4 nM reverse primer, 0.4 nM fluorescently labeled M13-forward primer, 0.24 mM each dNTPs, and 1.2×Promega PCR buffer. The PCR conditions used were 3 minutes of an initial denaturation at 94℃, 1 minute at 94℃, 1 minute at annealing

temperature (Ta, Table 2), and 1 minute and 15 seconds at 72℃ , for 10 cycles; and then 1 minute at 94℃ , 1 minute at 58℃ , and 1 minute at 72°C, for 35 cycles, with a final extension of 5 minutes at 72°C. An aliquot of 1.5 µl of PCR was cleaned with ten-fold-dilution of ExoSAP-IT (Affymetrix Inc. Cleveland, OH, USA) according to manufacturer's specifications. ExoSAP-IT was applied to remove excess of primers. After dilution to 100 ng/µl, fragments were separated on an ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and scored with GeneMapper (4.0) (Applied Biosystems, Foster City, CA, USA). Gradient PCR was conducted for optimization of annealing temperature for the makers that produced multiple bands at the initial screening.

Table 2. Putative EST-SSR markers from a comprehensive EST dataset (Liang et al., 2011)

No.	Locus	Motif	Start Position	Stop Position	Forward primer sequence(5'--3')	Reverse primer sequence(5'--3')	Product Size	Arabidopsis Hit	E value	Score
1	isotig17398	(tg)8	552	563	ATTAATTCGTGTGTGCGCGT	CGTCAAACCGAGTGCCTAAT	152			
2	isotig17677	(ctg)8	106	120	CGACTGGTCTGCGAATTCAT	ATCCGTCTATCCCCCAAGAc	199			
3	Contig08105	(ac)9	1511	1528	CGTTTGGTTGCATCAAATATCA	CCAATGGCCACAGTAGAGGT	171			
4	isotig06476/77	(tc)8	2184	2199	GACGACGACCTTGTCTCCAC	TTCGCACTCGTTTTAGGAGC	201	AT3G61600.1 | Symbols: ATPOB1, POB1 | POZ/BTB containin G-protein 1 |	9E-32	139
5	isotig13347	(ag)16	321	352	CCGCAACTAACACGACATCA	AATTCCGGCCCCATAATAAG	164	AT1G02500.1 | Symbols: SAM1, SAM-1, MAT1, AtSAM1 |	0	708
6	isotig14672	(ac)13	136	161	GGGATGTCCTGGGAGAAATC	TGTGGACATTAAGTTATGGTTTTG A	153			
7	isotig12564	(tc)8	1857	1872	TGCCTCTGTAAGGAAACCCA	AAACATGGCACACCTGTTCA	178			
8	isotig03894/95	(ttg)9	2546	2572	CCGCATTTGGTGGTTGTTAT	CCAAGCCTCTTTTCCTTTCC	209	AT4G32640.1 | Symbols: | Sec23/Sec24 protein transport family	6E-30	133
9	isotig31399	(aat)9	131	157	TGAGAGCATTTCAGCATGTCA	AGTGGGGCACTCCTACAATG	171			
10	isotig19678	(tc)17	593	626	TAGTATAGGCTCCCCCTGCC	GGGACCCTAAAGCTTCATCC	158			
11	isotig07006	(ct)10	250	269	TAGCAGGAAAGGACAAACGC	AGGTTCTCGCCTTCCAATTT	164			
12	isotig07448/49	(ag)15	273	302	TCATGGAATCCACACACTGG	GTAGGGCCATGCTTCGTAAA	158			
13	isotig21806	(ct)10	1	30	CGGAAGGAAAAACAGAAGCA	CATCATCGCATTTCATTTGC	174			
14	isotig02980	(tc)10	442	461	GGAGGAAGCACTGATCTGGA	TCCTCTGGCAGATTGACCTT	195			
15	isotig14235	(ttc)8	218	241	TCTGATTCGATCTCTTCGGG	TCATCTCTGCTCCAATCGTG	150	AT2G43360.1 | Symbols: BIO2, BIOB | Radical SAM superfamily protein |	7E-50	198
16	isotig26672	(ct)10	101	120	GGGAGAGGATCGGAGAGAAG	ACCGACTCCACACCATTAGC	170			

17	isotig30818	(ct)8	118	133	CGTAATCAGAGCTGGACCGT	AAGAATGAGGAGAGGGAGGc	194			
18	isotig03613/15	(ag)25	940	989	ACATCTGCAACGAGTCCTCC	CGCTTTTAGGCTGGATTCTG	242			
19	isotig13819	(ag)10	1309	1320	GTGGATTGCAAAGGCAGAGT	AAAACAAAAGCAAGCAAGCC	183	AT2G37620.1 | Symbols: ACT1, AAc1 | actin 1 | chr2:15779761-15781241	0	813
20	isotig05571	(tc)10	444	463	GGTGGTGGTGAGTTTGATCC	TTTCAGAAtGCGTTACGACG	162			
21	isotig33307	(ag)10	126	145	CCCCACTTACCTTCCACCTT	TCTCTGGGATTTCAGCAGC	180			
22	isotig11892	(ct)9	2456	2473	TCAGCCTTATTGAAGTGGGC	CCTGAAGTGGGTCTCCTAGtG	216	AT3G02050.1 | Symbols: KUP3, ATKUP3, ATKT4 | K+ uptake transporter 3	2E-17	91.7
23	isotig11814	(tc)8	2519	2534	CCTTCGTACCCAAAACCCTT	TAGAAAAgGacgggggagTT	154	AT1G34130.1 | Symbols: STT3B | staurosporin and temperature sensitive	4E-62	240
24	isotig22892	(tc)8	126	141	CCAGCTTGCATTGGATttCT	CCCACTACCATTTGCCTGAC	198			
25	isotig13816	(ttc)8	1481	1504	TTGAATCGTCCGTGATCTGA	TCTGagGGATCCCATTTCTG	154	AT5G66280.1 | Symbols: GMD1 | GDP-D-mannose 4,6-dehydratase 1 |	2E-56	220
26	isotig03209	(tc)16	253	284	TGCCCGTGATACCGATTATT	tgaagcctttcgattgctct	153			
27	isotig23428	(tgt)9	450	476	CGGTGGAAGTGGTGTAGCTT	TTCTCATTTATTCTTCAAAGCCAA	150			
28	isotig11568	(tat)10	3781	3801	CCCTTGTAACAGCTCGTGGT	CCGTTCAAGGAAAGATGGAA	167			
29	isotig06261/62	(ca)10	3524	3535	GTTtggatgtgttgtgctgc	TGGTGGTAgGGCaAAGaaag	170	AT5G62000.1 | Symbols: ARF2, ARF1-BP, HSS, ORE14 | auxin response	2E-52	208
30	isotig18778	(ttg)12	70	84	TCTCTTGGCGCCAGACTACT	TGACAAgaaGAAGAAGAAGAAGAA GA	171			
31	isotig26603	(ag)17	242	275	AGCGCAAGAGCATAGCATTT	tGGGtTTTTCTCGtTTCTCC	150			
32	isotig21507	(ggt)12	150	185	GTTGGAGGATGCACAGGAGT	ACCTGCCTACTTCTCTCCCC	178			
33	isotig05048	(tc)12	709	732	CCAGCTTGGAACAAGGTCAT	TCACCAACATCACATGAGCA	172	AT4G02520.1 | Symbols: ATGSTF2, ATPM24.1, ATPM24, GST2, GSTF2 |	3E-14	79.8
34	isotig07042/43	(gtt)11	1303	1335	GAACCAAACACCCAAACACC	ATAACCCCATTCGAAATCCC	229			

35	isotig16499	(ac)14	59	86	TCAATTGACAAaTCACAGGCA	TTCCACGTGTGTCACTTTGG	156	AT4G31330.1 | Symbols: | Protein of unknown function, DUF599 |	4E-20	99.6
36	isotig28623	(ag)8	231	246	GCAGCCCAGAGAGAGATTTG	TGGGGGTCTTCTTCTTGTTG	179			
37	isotig18722	(tc)8	109	118	CAATCTCCCTCAGCTTGCTC	ACAACAAATTGAgGGACCCA	158			
38	isotig12559	(gca)8	150	173	TTACCATCACACCCACCCTT	GTGGTAGGCCATGAAATGCT	193	AT1G49600.1 | Symbols: ATRBP47A, RBP47A | RNA-binding protein 47A |	7E-20	99.6
39	isotig03958/59	(ttc)12	407	442	TTGCGGTTCCATAGGAGTTC	tTTTTACCAAAATCCCTAGAAGG	222	AT1G78240.1 | Symbols: TSD2, QUA2, OSU1 |	4E-28	127
40	isotig19384	(atg)8	159	182	TTGCGTAAATGCATCCAAAA	GAAGCCtaTGCAAGATGCAA	181			
41	isotig09130	(tct)8	641	650	TCCTACATTTCCGACAAGGC	CTGCTGCTGCTGAAGATGAG	152			
42	isotig12747	(tc)13	110	135	aCCCCcAAATCTCTCTGCTT	TTCCGCCCAGACAAAGATAc	199			
43	isotig04395	(ga)12	389	412	TTCGGTGGAATTAGCTTTGG	GGCTGAGCCTAATGAGATCG	185	AT1G76860.1 | Symbols: | Small nuclear ribonucleoprotein family	1E-22	107
44	isotig04397	(ga)12	389	412	TTCGGTGGAATTAGCTTTGG	CAAAAGATGCAGaAGGGGaa	198			
45	isotig03014/15/16	(ct)15	4315	4344	TACGAAACCTTCGAGGATGG	TGGCTGAAAATGCACTGCTA	153	AT1G59890.2 | Symbols: SNL5 | SIN3-like 5 | chr1:22044326-22050670	4E-23	111
46	isotig03780/81	(tc)12	1295	1309	CCACCTCTCAACGATCCCTA	CCAGAGACtTCCCCATCaAA	174	AT5G51570.1 | Symbols: | SPFH/Band 7/PHB domain-containing	3E-62	240
47	isotig15644	(ct)16	205	236	TCCCGCTATAGCCACAAATC	TGTGGCGAGAGAATTTAGGG	213			
48	isotig12944	(ggt)9	321	335	GGGGTGTAACTGGATGATGG	ATGTGGCCACCTGCAGAC	156			
49	isotig14718	(tc)18	1203	1238	GAAAGGAGAAGGGTTGGGAG	TCAGCAAgGCACAACAGTTC	176			
50	isotig17436	(ct)15	586	615	ATCATTGGGCTTCAATCAGC	aACGGTTCATTCACGATTGG	168			
51	isotig23696	(ct)18	311	346	ATCACCATCTTCCTCATCGC	AAACCATTCCAACCATCCAA	198			
52	isotig23903	(tgc)9	266	292	CTCGGGACCTATCGATTTCA	AAGACGCCACAGaagTCCAG	159			
53	isotig12995	(tg)14	146	173	CGGATCTTTCTCTTtCCATCC	AAGAAGATTGCAGAGGCAGAA	223	AT3G53570.1 | Symbols: AFC1, AME2, FC1 | FUS3-complementing gene 1 |	1E-15 7	555

54	isotig16653	(ctt)13	776	814	CAAACGTCTCTGCAACTGGA	GCAAAACCCATCTCCTGAAA	158	AT3G21510.1 | Symbols: AHP1 | histidine-containing	5E-13	75.8
55	isotig04682	(ct)18	1930	1965	CAATGCTCACTGCATTGCTT	TTGGGCaAAACCAGGTTAAT	168	AT1G52150.2 | Symbols: ATHB-15, ATHB15, CNA, ICU4 | Homeobox-leucine	4E-40	167
56	isotig23360	(ttc)10	172	201	GGGATTTATGTCGGAGGGTT	CTCCGCCTGTAAACAGAAGC	159			
57	isotig03362/63	(ga)10	893	912	TCTCGTAGCTTGCCTGGTTT	CTTCTCATTTGTTCCCCACC	174			
58	isotig03665/66/67/68	(gcc)8	645	668	GCCACAACGTTTTTCACCTT	GGCTTTGGTTCCACTTCTCA	164			
59	isotig15508	(ct)8	827	842	GGATCCAAATCTCAAGCCAA	AAAGGCAGCTAAGGCAAACA	176	AT3G62870.1 | Symbols: | Ribosomal protein L7Ae/L30e/S12e/Gadd45	1E-75	283
60	isotig13566	(tc)23	631	651	CACCTtATGCGCTCTCAACA	GCCCTTTCTCTTTTTAAtTGGA	170			
61	isotig17264	(tcc)12	817	828	GACTGGACGAACCACCTGTT	GGCACACGAAGAGGAAGATT	168	AT2G23980.1 | Symbols: ATCNGC6, CNGC6 | cyclic nucleotide-gated	3E-14	79.8
62	isotig17519	(tc)9	117	134	CGTCTGCTCGTTTTtCcTTC	CAGTCATCATCCACAGCCAT	178	AT1G01750.1 | Symbols: ADF11 | actin depolymerizing factor 11 |	5E-22	105
63	isotig07238/39	(acc)10	1600	1629	CGTTATCAACATGGGCACTG	GGACCaTGCTCATCCAAAAT	173	AT3G51860.1 | Symbols: CAX3, ATHCX1, CAX1-LIKE, ATCAX3 | cation	7E-17	89.7
64	isotig11944	(ag)8	2338	2353	GAAAGCAGTAAATGCGCTCC	TCTCCCGATCTTCAATTTGc	188	AT2G38120.1 | Symbols: AUX1, WAV5, PIR1, MAP1 | Transmembrane amino	1E-71	272
65	isotig13272	(ag)24	1540	1587	CAAGCTTTCTCAGGACCAGG	GGGCAAATTTCCTCCATTTA	190	AT4G13940.1 | Symbols: HOG1, EMB1395, SAHH1, MEE58, ATSAHH1 |	1E-15 8	557
66	isotig12340	(ct)17	2041	2074	GCTAAGCCAGAGCAAAATGG	TGTGGCTTGTTCTCCATTCA	169	AT4G37740.1 | Symbols: AtGRF2, GRF2 | growth-regulating factor 2 |	1E-18	95.6

67	isotig05553	(ct)17	18	35	GAAGGCGAGAGTATTGCTGG	TCATCACAACATAATTCCATTGC	156			
68	isotig10814	(tc)13	555	580	AaTAGAGCTCCcAGCACGAa	GACCTGCATCAGCCCATTAT	158	AT1G64740.1 | Symbols: TUA1 | alpha-1 tubulin |	8E-36	151
69	isotig05551/52	(ct)15	18	35	AAGGCGAGAGTATTGCTGGA	CTGACCGATGTGGATCGAG	174	AT4G14960.2 | Symbols: TUA6 | Tubulin/FtsZ family protein |	0	1057
70	isotig21845	(tc)25	475	524	GCAGCCTATCGTTTCTCAGG	CAAACTTCTCACGCGCAAAT	183			
71	isotig11972	(ct)10	124	143	GTTTTGTACCCCAGCCAAGA	TATCTTGCGTTTTCCGAACC	162	AT3G21650.1 | Symbols: | Protein phosphatase 2A regulatory B subunit	3E-35	151
72	isotig05098	(ga)15	171	200	CTTTGTGCCCATGGAGTTTT	GGGGCGAGAACTGGAATAAC	156			
73	isotig09069	(tgt)12	115	126	CAACCGTCCATTCTCCAGTT	ACCCAAATAAAAATGCGTGC	196			
74	isotig15002	(ag)14	368	382	ATGTCCTGAAGTGGAAACCG	ATCTGCCAAAAAGGCATACG	177	AT4G21450.3 | Symbols: | PapD-like superfamily protein |	3E-27	123
75	isotig03107/10	(tc)8	1430	1445	tcAACGGTGGATGAGGTGTA	GAAATTTTCTGGATTTTCCAACTT	177			
76	isotig26946	(tc)9	252	269	GATTTTtCGAGCGTTtCGAG	CAAGTAGACAAAACGCCGGT	152			
77	contig08221	(ct)8	548	563	CAATGCCCCATTACTCGTCT	AAAGCCCaAAGCAAACCATA	153			
78	isotig14650	(gt)23	1143	1188	AGCCCATTtATCACGTCCAG	ACACAACCAGAGGACCCAAG	151	AT2G39900.1 | Symbols: WLIM2a | GATA type zinc finger transcription	3E-30	133
79	isotig12460	(ttg)8	1904	1927	ATTTGGTGAGCTCGGAGAGA	ACaGCGaCGaAGaCGaAAAT	180			
80	isotig05643/45	(ag)8	1367	1382	GAGCAGGAGAGATTTCGTGG	TCTCTCTATCCGAAAGCCGA	162	AT3G11320.1 | Symbols: | Nucleotide-sugar transporter family	8E-59	228
81	isotig05545/46	(gct)10	1773	1788	TGGAAGAAAAACaCCGGTTC	CGAGTGTTGGAGGATTTGGT	185			
82	isotig05547	(gct)10	1399	1414	TGGAAGAAAAACaCCGGTTC	CGAGTGTTGGAGGATTTGGT	185			
83	isotig13163	(tc)15	1652	1681	TTTCCCGTAAGAACAATGCC	GTCGGtGGCAAGAAAACATC	215			
84	isotig14849	(ag)12	459	482	AATCCCAACTCAGTGATGGC	AAAGACAAGTGCTGCTCCCT	153			
85	isotig06190	(cag)8	100	123	TAAAGCCCACATCCTTCCAC	GGAGGGGTAAGGGTGGTAAA	194			
86	isotig09553	(tgc)8	428	451	gggtttgaatacctgctgga	ACCTATTTGGGAGTGGGGAG	171			

87	isotig02981	(tc)10	324	343	GgtAGAGAAGGGGGtGGAAG	TCCTCTGGCAGATTGACCTT	207			
88	isotig04396	(ga)19	305	322	TGAAAGTTAGTCCCCGGATG	CAAAAGATGCAGaAGGGGaa	156			
89	isotig15138	(ct)16	1107	1138	GGGCTTTAACCGAGGGATAG	TCCTGCCTCACATAGCTTAACA	226	AT1G69530.3 | Symbols: ATEXPA1, EXP1, AT-EXP1, ATEXP1, ATHEXP ALPHA	5E-29	129
90	isotig10813	(ag)13	161	186	GACCTGCATCAGCCCATTAT	AATAGAGCTCCCAGCACGAA	158	AT5G19770.1 | Symbols: TUA3 | tubulin alpha-3 | chr5:6682761-6684474	0	638
91	isotig16887	(tc)18	871	906	TTATGAGGACTGTGGGGGAG	ACAATCACTGCATTTTGGCA	188			
92	isotig16962	(tca)10	697	708	GCATCCTCCTCATCTTCTGC	AGGTTTTACCCTCCAGCGAC	208			
93	isotig23746	(cag)8	285	296	TTGACCACTTGAGCGACAAc	TCTTCttcTTCGTGGCCATT	178			
94	isotig13870	(ag)8	1462	1477	TTTtGGAGCATTTCCCAATC	AAGACTCAAGAGCAGAGCgG	172			
95	isotig30473	(cca)10	192	206	GCAGCAGTACCTTCCTTTGG	ATGGTAGTGGTGGtGGTGGT	156			
96	isotig12394	(tct)12	359	376	ACAaCGACGATTCTGGCTCT	TCATCATCATCAAGGGACGA	187			
97	isotig28546	(gtt)8	215	238	TTCTGTACATTTGCTTGCGG	TCGACAAGCTTTTCCATGCT	167			
98	isotig05387/89	(ag)18	1563	1598	GACGGGGTACTGAGAAGCTG	CCATTCCTGCACGTTATCCT	163			
99	isotig13874	(at)16	819	830	TGTCCAGAGTTCCAGTCGTG	GCAACCAcCCaaAAAGAAAA	162			
100	isotig11381/82	(ag)17	364	397	CGTTCAAATCTACACCCCGT	AGGCAAATAGCAACAGCAGG	174			
101	isotig07208/09	(ctt)10	57	86	TCAAATGGCGATTACAGCAG	TGGGAGGAAGAGGAAGATGA	165	AT5G48300.1 | Symbols: ADG1, APS1 | ADP glucose pyrophosphorylase 1	1E-15 9	563
102	isotig14101	(cag)8	615	638	CTAGAAGGAAGCATCCCACG	TTTCGGAGGtTgGATCTTGT	181			
103	isotig14899	(ct)19	357	394	GGCGGTTAGTTATGGTCCAA	AAgGGGAGAAGCCAGTGAAT	161			
104	isotig13846	(gaa)12	110	145	GTGTTTTGGGGTTTTGGAGA	GGTGGTGATCCTCAATCCAT	201			
105	isotig16173	(ct)15	177	206	CCTTCCCCTCAGACTCCTCT	CGCCAGATCTGAATGTGTTG	180	AT1G08465.1 | Symbols: YAB2 | Plant-specific transcription factor	1E-16	87.7
106	isotig11289	(ag)10	1415	1434	CAACGGAATTTCCACTCCAT	ATGTGCCTCGTTCCAATCTC	180			
107	isotig15694	(ct)8	84	99	GCGCAATCATCATTTTCTCA	GGATTCCGATGAGGTTGTTG	159	AT2G23810.1 | Symbols: TET8 |	6E-13	75.8

								tetraspanin8 | chr2:10135859-10137352		
								AT1G43850.1 | Symbols: SEU | SEUSS		
108	isotig11889	(tgt)12	2321	2356	AGCTGGAGAAGTCTGCCTTG	CCACAACAACTCCCGTCTTT	162	transcriptional co-regulator |	3E-44	180
109	isotig15655	(ttc)12	897	906	GTGTGCGCCTTTTGTGATAA	CGCAGCAACACATTCAAATAG	202			
								AT1G06430.1 | Symbols: FTSH8 | FTSH		
110	isotig11898	(tc)12	2465	2488	CCCGTTCACTGACTTTAGCC	GGGAATTTCATCCCGAATTT	200	protease 8 | chr1:1960214-1962525	0	682
								AT2G03510.1 | Symbols: | SPFH/Band		
111	isotig04691	(ac)18	192	227	tcAagCccTtAatgCcAact	ATTCCCACTCGGTTGAACAC	163	7/PHB domain-containing	1E-66	254
						ACACATCTTTGATATTAGAAATTC				
112	isotig18778	(ctt)9	70	84	TCGAGGAAGGACCAAGTGTT	CAT	153			
113	isotig20478	(ag)9	176	193	TCCTCTTTGGCTCTTTCGAG	ACTGCAGGTAAAACGATGCC	165			
								AT1G15690.1 | Symbols: AVP1,		
								ATAVP3, AVP-3, AtVHP1;1 | Inorganic		
114	isotig11820	(ag)12	218	241	AAGCTCAAAACCCATCTCCA	CCAAAGCAAACACAATCCCT	168	H	3E-87	323
								AT4G09160.1 | Symbols: | SEC14		
115	isotig12312	(ga)12	254	277	CGCAGAAATCCAACAAATCA	AGTTGGGTTTTCCATTTTGG	158	cytosolic factor family protein /	3E-13	77.8
								AT1G20330.1 | Symbols: SMT2, CVP1,		
116	isotig13520	(ga)8	447	460	GACCCGAGTACACACACGAA	ATaAATCCCCATCGACTCCC	165	FRL1 | sterol methyltransferase 2	7E-32	139
117	isotig26389	(cac)8	84	107	GGTTAGGGTTTTTCGCCTCT	TTGGAGCAAATCCGTAGCTT	151			
118	isotig15788	(tc)19	118	155	TGCAGCTGTTGGATCTGACT	CCGGAACGGAATTTCAGATA	168			
					TTTCAGCATTCATTCAGAATACA			AT4G26610.1 | Symbols: D6PKL1,		
119	isotig05673	(cca)10	149	178	AC	ATTGGGGAAAGAAGAGGTGG	197	AGC1-2 | D6 protein kinase like 1 |	1E-45	184
								AT3G25560.3 | Symbols: NIK2 |		
120	isotig02154/55/57/58/59/60	(ata)9	2159	2185	TCGCTGGATGCTAGAACAAA	GGAGATgGGCAAACAACACT	185	NSP-interacting kinase 2 |	1E-27	125
								AT3G53610.1 | Symbols: ATRAB8,	1E-10	
121	isotig16102	(tc)16	65	96	GAaaTCTTCAACGACCGACC	TCTTACCAACACCGCTGTCA	174	AtRab8B, AtRABE1a, RAB8 | RAB	0	365

								GTPase		
122	isotig25924	(aac)8	31	54	TCCATCTCCAACAACTACAaCAA	CCTCTCAGGCAGATGAAAGC	150			
								AT5G66730.1 | Symbols: | C2H2-like		
123	isotig06764	(gca)8	1547	1570	AGCAACATCAATCCTCCGAT	TGTCAATCCCAACCAGATGA	178	zinc finger protein |	9E-16	85.7
								AT1G01090.1 | Symbols: PDH-E1		
					TCTAGCAACTTCTTGATAATGCA			ALPHA | pyruvate dehydrogenase E1		
124	isotig15731	(ga)9	275	292	AA	CGCTTTCACATGGTTAGTTGG	167	alpha	4E-17	89.7
								AT5G19090.1 | Symbols: | Heavy		
125	isotig13485	(tc)8	1018	1029	CGAAAGACATTCCCATCACA	CCATTACAATCCACAGCCAA	205	metal transport/detoxification	6E-29	129
126	isotig03364/66	(ga)13	36	61	tgagACAGAGCAGAGAGAGATCA	CTTCTCATTTGTTCCCCACC	156			
127	isotig19526	(aag)8	84	107	TGGGTGCTATGTTGGTTTTG	TAATTGTATGCTGCTGCCCA	180			
								AT4G39350.1 | Symbols: CESA2,		
								ATH-A, ATCESA2 | cellulose synthase		
128	isotig11541	(tc)8	106	120	AGCCCCCAACTAATCAAACA	GAAATGGGAGAGGGTCAACA	156	A2	3E-61	238
129	isotig03236/38	(ttc)9	989	1015	TTCTTTCAATGGCAGCAAAA	AGACAATTCAGCTTGCCTCC	160			
					CCTCTTCTCTCTCTATCTCTTTCTT					
130	isotig12099	(ct)9	2118	2132	CA	GTCCCTCAAAAGGGTTTCCT	168			
								AT1G20050.1 | Symbols: HYD1 | C-8,7		
131	isotig16848	(tc)10	115	134	TTCTCGGAGGAAACGAGAAC	CGTAGTCGGGGAGATTGAGA	213	sterol isomerase |	3E-14	79.8
								AT5G37600.1 | Symbols: ATGSR1,		
132	isotig12454	(ga)12	1321	1344	TCCATATGTATTCGGCGATG	AGtAGGCACGTTCCTTGCAC	189	GLN1;1, GSR 1, ATGLN1;1 | glutamine	4E-55	216
								AT3G55940.1 | Symbols: |		
								Phosphoinositide-specific phospholipase		
133	isotig12356	(ga)8	161	176	GCTCAAAATAAAAAGCCCGA	GAAGGTGGAGAAAACACGGA	178	C	2E-14	81.8
								AT2G05990.1 | Symbols: MOD1, ENR1		
134	isotig13703	(tc)15	1415	1444	CGACAGTCCGATATTTGCAG	CCGGAACAAAATCCCCTATT	165	| NAD(P)-binding Rossmann-fold	5E-88	325

135	isotig08403	(ga)9	598	615	gagaggtagaacgagcggag	CCTCTTCaAACCCtTCTccC	189			
136	isotig15780	(tct)8	1049	1072	GCAGTAGGAAGGAAAGATCCC	AGCATTATCCGTTCCCTTCC	156			
137	isotig31377	(ta)8	105	120	TGCCTTTTTCTTTAATGTGtGG	TCTGTTCAAGGCCCTTCTTT	172			
138	isotig29434	(tg)21	57	98	CTCCAGCTTGaggggtGTAT	ATGGTTGGACGCTTGAGATT	179			
139	isotig31749	(tct)10	198	227	GAATAACcGCtCTTTTGGGA	AAGCCAAGTGGCAAAGAAGA	164			
						GGaaGATCTCTCTCTCTCTATCTCCT				
140	isotig33292	(ga)16	24	37	TtGCATCTTCTTTTTCATCCC	T	167			
								AT1G20330.1 | Symbols: SMT2, CVP1,		
141	isotig13520	(ata)9	447	460	GCGCTATCCCATCTTCAAAT	AAGGAAGTTGCAGGCAGAGA	182	FRL1 | sterol methyltransferase 2	7E-32	139
142	isotig14887	(aat)8	387	410	TGGTGCATATGGGCTTAGAA	TATTCCCCCAGCTTCTCCTT	171			
					AAAAATGCtAAtCCAATAACTTTC					
143	isotig11620	(tg)13	3286	3311	G	TATCCAACCGATCACCCATT	160			
								AT2G01190.1 | Symbols: |		
								Octicosapeptide/Phox/Bem1p family		
144	isotig12076	(tct)10	1475	1489	TGCTTTGCATTTTCTTCTGTG	CCAAACACAGCATTTTCCAA	187	protein	1E-15	85.7
145	isotig08911	(ga)18	187	222	TTGAAGTCCAGATTGATTGATTG	GCCTAGGGaGATGtTTTTGG	157			
146	isotig06349	(ga)12	96	119	caaccactcaccaaaattgc	AGCTCGAtTTGAGAGCGaAG	151			
147	isotig07967/68	(ct)10	1213	1232	TCAGTTCGAAGGTCTTGTGC	AGAATCCGCTAGGTGGGAGT	159			
								AT2G14910.1 | Symbols: | unknown		
148	isotig07108	(ag)13	1658	1675	CCTCAGGGGTCAATTCCTTA	GAAGAAGGATCAGAGCGTGG	151	protein; LOCATED IN: chloroplast;	6E-14	79.8
149	isotig14934	(at)14	130	157	GAAATGGACGACTAACCCAAA	TACGGCTGCGATTGTATTGA	175			
								AT5G13420.1 | Symbols: |		
150	isotig11603	(tc)10	1567	1588	TCTTCAAACCAAGGCTGTTG	GCACTACATCCCTTTTcCCA	167	Aldolase-type TIM barrel family protein |	1E-34	149
151	isotig22220	(tc)11	482	503	TGAGGTGACTTTGGCTTTTG	GACCCgaGCTGTAAAATGGA	189			
								AT2G03510.1 | Symbols: | SPFH/Band		
152	isotig04692	(ca)17	171	204	CATCCAAATGCAGCAGAAAT	ATTCCCACTCGGTTGAACAC	177	7/PHB domain-containing	1E-66	254

153	isotig15826	(ga)8	197	208	CAAGATTTCGTGCTAAACAGAGT	TTGGGAAAATGGAGAAATGG	184			
154	isotig09599	(ct)10	72	91	gatgaaggagaattctataTTTTCTGA	CCAGCCAAGAAAGAAAATGG	156			
155	isotig06536/37	(tc)9	123	140	AAAAaGGCTCCTCTCCCATC	CGTCTGATCTGCCTGCATAA	147	AT2G24050.1 | Symbols: eIFiso4G2 | MIF4G domain-containing protein /	9E-14	79.8
156	isotig15738	(ctg)7	941	961	AAaACCATGATGAAGGCGAC	ACAACCAAACAGCCACAACA	149			
157	isotig04927/28	(ag)6	1058	1069	aaaaggggagatgggtatgc	TTTCtTGGCTCTCCCTCTCA	177			
158	isotig17591	(tctt)3	165	176	AAAAGGGTGCTATCAGACCG	AgTGTGAGGGGTTCATCGAG	174			
159	isotig17626	(ct)6	83	94	AAACGCCACAGTTctGAAGG	GATCCACCTTCGTGAACACC	160			
160	isotig14895	(ttc)4	1189	1200	AAACGGTGCAATCTAATGGG	CCTCCTCTCTCTGCACATCA	171	AT2G42320.1 | Symbols: | nucleolar protein gar2-related |	7E-22	105
161	isotig12555	(gca)6	200	217	AAAGGAGCGAGATTTCCGTT	TCTTCTTCCTCGTCCTTCCA	173	AT1G58440.1 | Symbols: XF1, SQE1 | FAD/NAD(P)-binding oxidoreductase	2E-17	91.7
162	isotig22733	(aaag)3	217	228	AAAGTCCATGTCTGGATCGC	GCAGGCATGGTAAGAGAAGG	146	AT1G67785.1 | Symbols: | unknown protein; Has 30 Blast hits to 30	2E-11	69.9
163	isotig20648	(tct)6	273	290	AAAGTTAGTGCGGTTCCAGG	AACAGAGCAGGCTTGTCGAT	182			
164	isotig08475/76	(cggtt)3	638	652	AAAGTTTTGgATTGAGGCCC	TGAGGGAAAATATCCAACCG	150			
165	isotig06326	(tc)6	3030	3041	AAATCTGAAATCTTGCGGgA	CAGCCCTTCCTCtTCTTCCT	151	AT5G57210.1 | Symbols: | Ypt/Rab-GAP domain of gyp1p superfamily	3E-29	131
166	isotig14308	(tttc)4	352	367	AACAAAATGCAAACAAATGGG	AAAAGGTGAGAGGCAACGAA	167			
167	isotig07339/40-a	(gac)5	692	706	AACAACAACAAGAAAGCGCC	CCTCTTCCTCCTCATCCTCC	208			
168	isotig07269	(aaag)3	91	102	AACACAACATTGCAAGCCAA	AACTTTGAGCCTCTTATGGGAA	149			
169	isotig06269	(ag)8	432	447	AACAGCTTGTACCTGTCCGAA	GAACGTAGGATCGGAGTCCA	145	AT2G29140.1 | Symbols: APUM3, PUM3 | pumilio 3 |	1E-65	252
170	isotig29851	(aac)5	173	187	AACCACCgtGTTGTgTTTGA	CCTAAGCCAACGGAAGAAGA	152			
171	isotig13800	(aca)4	129	140	AACCCCACAACAACAAGAGC	GAGGCAGATCTTTCTGCGG	179			

172	isotig11723	(ctcc)4	128	143	AACCCTAGAAAGAAAGGGCG	TCTTCgTTCCTTCCCTGAAA	150	AT4G39680.1 | Symbols: | SAP domain-containing protein |	7E-21	103
173	isotig13127	(ca)5	213	222	AACGCAGGTGATAGGGTTTG	TATGCACACATCATGGAGGG	150	AT2G35210.1 | Symbols: RPA, AGD10, MEE28 | root and pollen arfgap |	1E-39	165
174	isotig12051	(ga)6	2321	2332	AACGCTCTCTGTGTCGGATT	CCTCCTTTGGTGATTTTCCA	147	AT5G51060.1 | Symbols: RHD2, ATRBOHC, RBOHC | NADPH/respiratory burst	2E-14	81.8
175	isotig11587	(cag)7 (aaaag)	183	203	AACTGAGCTTCCCAAAGCAG	CCTCGACCACTTTCTCCTTG	162	AT1G48410.2 | Symbols: AGO1 | Stabilizer of iron transporter SufD /	2E-18	95.6
176	isotig12785	3	1448	1462	AAGAAGCCACCATCGTTCTC	GGTTTTCGTTGTGGTTGTGTT	145			
177	isotig07066/67	(tg)8	1818	1833	AaGACGTACGGATCGTCAGG	GTCACCCACGTTTGGAATTG	148			
178	isotig04394/96	(aag)6	305	322	AAGAGCCCCAAAATGACCTT	TTTCGTCCATTTCAATGCTG	164	AT1G76860.1 | Symbols: | Small nuclear ribonucleoprotein family	2E-22	107
179	isotig12197	(ga)17	1898	1931	AAGATTGGCGAGATACCAGC	TCGATGCAAGTACACCGAAC	201	AT4G34200.1 | Symbols: EDA9 | D-3-phosphoglycerate dehydrogenase |	5E-24	113
180	isotig11187	(aag)5	387	401	AAGCCAGAGGAGAAGAAGGC	GATGGACCTGCTTCAAGACC	175	AT1G07790.1 | Symbols: HTB1 | Histone superfamily protein |	1E-37	157
181	isotig06460/61	(cag)5	1262	1276	aaGCcGAGAAaTGgAGTTtG	tgcacattgttttcttccca	203	AT5G12370.1 | Symbols: SEC10 | exocyst complex component sec10 |	4E-43	176
182	isotig13031	(tca)4	140	151	AAGCCTTTACCCCTACGCAT	GATGATTAGCAATGGCTCGC	147	AT5G01590.1 | Symbols: | unknown protein; FUNCTIONS IN:	4E-21	103
183	isotig07949/50	(gaa)9	165	191	AAGCTCCACCCCATCTCTCT	CCAATTGTTGGCTCGTTCTT	191			
184	isotig20661	(aag)4	138	149	AAGGCAGAGAAGAGGCTTCC	CTGGAGATCCCAATGTCAGG	146	AT1G07790.1 | Symbols: HTB1 | Histone superfamily protein |	2E-45	182
185	isotig22945	(gca)4	379	390	AAGGCTGCTGGATATCGTTG	GTTGCTTGTTACCCTGCGAT	176			

186	isotig00731-38	(tg)7	97	110	AAGGGCATTTGAGTGAGTGC	TTGGAGGTCAGGTTCTtTGC	180			
187	isotig16780	(cag)5	89	103	AAGGGTGGAAAAaGaGGGAA	GCTGGAACTGTGGATCTGGT	154			
188	isotig12986	(ca)7	1619	1632	AAGGTACACCTAGGAGGGTGG	CACTCAATTCTCATTCCCATGT	175			
189	isotig11896	(tc)5 (aaaag)	745	754	AAGTTTCATGAATGCCAGCC	AGAAGGCGCTTGAGGTAACA	171			
190	isotig17502	3	499	513	AATCAGGCTGTCCGCTAGAA	CTTCCTGTTTAGGCCCCTTC	178			
191	isotig16813	(tc)7	418	431	AATCCAACATCCAACACCGT	TGCTATGCGAAATGATCTGG	170			
192	isotig19191	(ttttc)3 (aaaac)	251	265	AATCTCGAAGGGGAACCTGT	CGCGTTAGCCTTGAAGAAGA	148			
193	isotig05078	4 (aaaac)	701	720	AATCTGGgCCCTTGGATTAC	CGTGGTTGCCATCAAGTTTT	162			
194	isotig05080	4	285	304	AATCTGGgCCCTTGGATTAC	CGTGGTTGCCATCAAGTTTT	162			
195	isotig13026	(gag)4	1409	1420	AATGAAAATCGTTGGGGAGG	GACCCGCTTGTCATATTCGT	154	AT1G60940.1 | Symbols: SNRK2-10, SNRK2.10, SRK2B | SNF1-related	8E-81	301
196	isotig00591/601	(tct)4	110	121	aatGATTGGAgCCCCcTTt	CGGCTTGGATTCAAAGAAAA	157			
197	isotig01596/98	(aggg)3	421	432	AATGCTATTGTCAAaGgCGg	AATCACATCCAACGGCTCTc	152			
198	isotig14510	(ct)7	1318	1331	AATTGCACTTGCTGCTGTTG	TGACACCCACGGAAGAGAAT	174	AT2G20060.1 | Symbols: | Ribosomal protein L4/L1 family |	2E-22	107
199	isotig07339/40-b	(gac)5	692	706	ACAAAAAGCCCATCAGCAAC	CTTCTTCCTCATCCTCCTCG	208			
200	isotig08103/04	(aga)5	221	235	ACAAAGAACCCCAATCTCCC	GCCCCAAAGCCATAGAGAAT	187			
201	isotig16705	(ag)5	739	748	ACAACAACAGCATGAGGCAG	TACGGCAGAAAATTTGGGAG	180	AT5G55190.1 | Symbols: RAN3, ATRAN3 | RAN GTPase 3 |	1E-12 4	446
202	isotig12491	(cag)7	435	455	ACACCCTCCTTGTAAACCCC	GATACGGCACGAAATGCTG	194	AT1G49600.1 | Symbols: ATRBP47A, RBP47A | RNA-binding protein 47A |	2E-14	81.8
203	isotig14397	(ta)6	156	167	ACACGTCCCAACTTCGAGTC	GGATGCTGTGAAACGGAGAT	153			
204	isotig02409/10	(gaaa)3	2439	2450	ACAGATCGGCACAGAACAAA	GTTGGGTATTGATCCGTTCG	148	ATCG00020.1 | Symbols: PSBA |	0	1590

								photosystem II reaction center protein		
								AT3G18480.1 | Symbols: AtCASP,		
205	isotig15677	(gat)4	1045	1056	ACAGATGCAATGGGAGAAGG	TGAAAGAAGGGAAGGAAGGAA	147	CASP | CCAAT-displacement protein	9E-46	184
206	isotig20355	(cag)4	324	335	ACATCcaTTCTTCcgTCGTC	AAGAAGAAGAAGCAGCCACG	165			
207	isotig02995/97	(aag)5	399	413	ACATGGCCGTGTGTCTCTAA	AGATTGAAGCGATTCCGAAG	145			
								AT4G10710.1 | Symbols: SPT16 | global		
208	isotig11575	(ag)5	2714	2723	ACATTGTGCAGACATTGGGA	CAAGTCCTTTAAACTGCGGC	165	transcription factor C |	1E-28	129
209	isotig22080	(gagg)5	486	505	ACCCCTACCTCTCCCTTCCT	CCCTCTTCTTTTTGGCCTTC	147			
								AT4G11420.1 | Symbols: EIF3A,		
								ATEIF3A-1, EIF3A-1, ATTIF3A1,		
210	isotig04591-93	(gaa)7	113	133	ACCCGTTTCTCTTGCTGAAA	TTTCtGGCTTCGCAAAAGTT	170	TIF3A1 |	3E-29	131
								AT4G39350.1 | Symbols: CESA2,		
								ATH-A, ATCESA2 | cellulose synthase		
211	isotig11541	(cac)5	106	120	ACCCTCCAGAGCAAACACAG	AGGACTTGCAACTGTTTGCC	201	A2	3E-61	238
212	isotig12102	(aag)4	2130	2141	ACCTGCTGACCAAGAAGTGG	ATTACTGCCAGGTCCCACAG	147			
213	isotig14317	(tct)4	1128	1139	ACCTGGTTCGTCTGGATCTG	CCAATTAAAGAGGCCCAACA	149			
214	isotig02516/17/20	(ct)5	954	963	ACCTTCaTtGAGCACCTtGg	GAcGtTcagCCCCTcTaATG	157			
215	isotig02521	(ct)5	655	664	ACCTTCaTtGAGCACCTtGg	GAcGtTcagCCCCTcTaATG	157			
216	isotig19540	(caa)4	120	131	ACGACAACAGCAaCCATCAT	CGTTGTTGTTGCCACCATAG	149			
								AT4G29390.1 | Symbols: | Ribosomal		
217	isotig24972	(aag)4	152	163	ACGCCATGGGTAAGGTACAC	CCCCTCTTCTTTCCAAATCC	172	protein S30 family protein |	2E-26	119
218	isotig14889	(tct)4	285	296	ACGTGGTGAGGAAATTCCAA	ATCAAGGGGAGGAGGAAAGA	165			
219	isotig06390	(cca)4	1582	1593	ACTACAATTCACCACCCCCA	AAGCATGAGTGGGGAGAAGA	165			
220	isotig10096	(tcg)6	282	299	ACTCGAGCGGGATTTTCTCT	AAAATCCCAGATCCTTCGCT	153			
								AT4G30210.1 | Symbols: ATR2, AR2 |		
221	isotig11786	(ttc)5	2486	2500	ACTCTGTTGTGCTTTTGCCC	GCCCACCACCTACCAAACTA	171	P450 reductase 2 |	3E-38	161

222	isotig02438/39/41	(tacaa)3	2148	2162	ACTCTTGCTCCAGTGGCTGT	CCTTCTTCAAGCGAGCAGAT	199	AT5G08450.1 | Symbols: | CONTAINS InterPro DOMAIN/s: Histone	4E-53	210
223	isotig02436	(tacaa)3	2148	2162	ACTCTTGCTCCAGTGGCTGT	CAGACGTGTAAAGGGGCTTC	184	AT5G08450.1 | Symbols: | CONTAINS InterPro DOMAIN/s: Histone	6E-49	196
224	isotig02437/40	(tacaa)3	2058	2072	ACTCTTGCTCCAGTGGCTGT	CAGACGTGTAAAGGGGCTTC	184	AT5G08450.1 | Symbols: | CONTAINS InterPro DOMAIN/s: Histone	6E-49	196
225	isotig15448	(cctc)3	1072	1083	ACTTAGCATTCTGCGCCAGT	TCCAATTGCTTGCTCTCTGA	154	AT3G48420.1 | Symbols: | Haloacid dehalogenase-like hydrolase	6E-19	95.6
226	isotig11386	(ac)5	59	68	AGAAaCAAGTGATATTACACACT GCT	cgggtgcctaagctAAGATG	147			
227	isotig02369	(taca)3	1522	1533	AGAAGACGAGCTCCAACCAA	tagtggatgcaacaagcagc	170			
228	isotig02871	(cat)4	163	174	AGAATCCATGGCTCTCCTCC	TCATCCACCTTCGAGACCAC	195	AT5G11770.1 | Symbols: | NADH-ubiquinone oxidoreductase 20 kDa	9E-58	224
229	isotig12156	(aga)7	1809	1829	AGAATCCCTGGCTTCCTCAT	CTGAAGGCCCATCTGTTTTG	171	AT3G15990.1 | Symbols: SULTR3;4 | sulfate transporter 3;4 |	1E-12	75.8
230	isotig16670	(ag)5	965	974	AGAGAGTtGCCGGAATCTGA	TTCACCTTTTTCACTAAACCACAA	164			
231	isotig03761	(attt)3	1349	1360	AGAGCaaaaaggTGGGgAAT	TCATGCATTGTTCTGCCATT	146			
232	isotig11529	(ccct)3	5337	5348	AGAGCTTGAACCTGATGGGA	CTCCCCATCAACCCTTCATA	190	AT3G08530.1 | Symbols: | Clathrin, heavy chain |	0	1316
233	isotig03228	(tgta)3	1667	1678	aGCACTCCCCCTTTCATttt	GTCAACGGAGtcGTAGGAGC	189			
234	isotig30166	(aata)5	197	216	AGCATTCTGGTCCTGGAAGA	CAACTCGtCtAACAGGCAGG	195			
235	isotig11945	(ga)13	64	89	AGCATTTTCAGCTGATCGAAG	CAcGGATCTCGTCCGTACAT	145	AT5G49720.1 | Symbols: ATGH9A1,	1E-31	139

								TSD1, DEC, KOR, RSW2, IRX2, KOR1,		
								AT3G16785.1 | Symbols: PLDP1,		
236	isotig14545	(tcaa)3	207	218	AGCCATTTCTCCTCCTCCAT	TAGGCCCCACCTGATCTACA	194	PLDZ1, PLDZETA1, PLD ZETA 1 |	2E-16	87.7
237	isotig19496	(cta)4	437	448	AGCCCATAAACCAGTTTCCC	TCAAATGCGAAAGCATTGAC	160			
								AT3G59360.1 | Symbols: UTR6,	1E-10	
238	isotig12271	(ct)7	2118	2131	AGCGAGAAGGAGAGGAGAGG	CGAAAGCTCAGAGGGAATTG	150	ATUTR6 | UDP-galactose transporter 6 |	1	371
								AT1G02890.1 | Symbols: | AAA-type		
239	isotig11544	(at)6	358	369	AGCtGAGCGGGAACACTTT	CAAAAGCCCAAGTAAGCTgC	197	ATPase family protein |	8E-80	299
240	isotig32418	(ta)6	138	149	AGCTGCTGTGATGGGAAAAT	TCCCAGACAGAGAAGAGGTTG	191			
								AT1G07790.1 | Symbols: HTB1 |		
241	isotig08756	(aag)4	212	223	AGGAGAAGAAGGCCGAAAAG	GGATGGACCTGCTTCAAGAC	166	Histone superfamily protein |	3E-44	178
242	isotig11092	(aga)6	224	241	aggatcccgaactccctaga	Gacgagagctgtaaccagcc	157			
		(agagg)						AT3G44110.1 | Symbols: ATJ3, ATJ |		
243	isotig13231	3	540	554	AGGCCATGACCCATTTGATA	CCCTTGCACTTGGAACAGAT	198	DNAJ homologue 3 |	5E-48	192
								AT5G62090.1 | Symbols: SLK2 |		
244	isotig12216	(cag)7	342	362	AGGCCTTAATTCAGCAGCAG	CTTCATACCGGATGTTGGCT	185	SEUSS-like 2 | chr5:24935221-24938540	3E-19	97.6
								AT2G35610.1 | Symbols: XEG113 |		
245	isotig04669-71	(catc)4	56	71	aggcggaagaaagaccttgt	aAcgGtgGagaAgatgatgc	145	xyloglucanase 113 |	7E-51	202
246	isotig22958	(cat)5	125	139	AGGGCGATATCACAAATTCC	ATGGTGGATGATTATGCGTG	171			
								AT5G19350.1 | Symbols: |		
								RNA-binding (RRM/RBD/RNP motifs)		
247	isotig13052	(ctg)4	1758	1769	AGGGTTCTTATCTCGTCGGC	AAATGGAAATGGAGTCGCAG	154	family	3E-25	117
248	isotig11012/13	(aga)4	1019	1030	AGGTGCAGCAAAGACAGGTT	GGTGtCGGTCTATGGGTGTT	184			
249	isotig32954	(aga)5	53	67	AGTAaTGGCGGCAGAGAGAG	TGATTTGGGGAGAAACCCTA	160			
								AT4G11240.1 | Symbols: TOPP7 |	1E-10	
250	isotig06830	(ag)6	108	119	AGTAGCCCACAACGCTCCTA	AGAAGGGGTAGACGTCCGAT	159	Calcineurin-like	4	379

								AT1G27530.1 | Symbols: |		
251	isotig12525	(ct)6	971	982	AGTCGATTTCGTGTTCGGAT	TGGCGTTTTGATTCCTAACC	172	CONTAINS InterPro DOMAIN/s:	3E-25	117
252	isotig15822	(tttgg)4	967	986	AGTGCCTAAAGGCCCATTTT	TGATCCCAAAACAAAGCAAA	151			
253	isotig18301	(catg)3	454	465	AGTTGCACCATGCAATTCAG	AGTAGAACCGGTcCACCTCA	195			
								AT5G21090.1 | Symbols: |		
254	isotig11030	(gt)6	904	915	AGTTTTGAAAAAGGGTGGGG	GCAGCAATGCTACCGAAAGT	150	Leucine-rich repeat (LRR) family protein	4E-23	109
								AT3G27090.1 | Symbols: | DCD		
255	isotig06928/29	(aac)4	517	528	ATAACACGGTTGCCCACATT	CAAGAACTTCATTGCGTGGA	183	(Development and Cell Death) domain	1E-51	204
								AT5G15490.1 | Symbols: |		
								UDP-glucose 6-dehydrogenase family		
256	isotig12659	(ctcc)3	1257	1268	ATAAGATCCGATCGTCAGGC	TAGCAAGGGCATCAACTTCC	185	protein	2E-63	244
								AT4G39800.1 | Symbols: MI-1-P		
		(caaac)						SYNTHASE, MIPS1, ATMIPS1,	1E-10	
257	isotig12715	3	49	63	ATAGGCCACAACAGGCGTAG	TGTAATCTTTCGCTTTGGGC	149	ATIPS1 |	6	385
								AT2G32830.1 | Symbols: PHT5, PHT1;5		
258	isotig12267	(cac)4	1027	1038	ATAGTGGCGGAGCAAGAGaa	ATCCACCCGATAGCAGAGAA	182	| phosphate transporter 1;5 |	1E-24	115
								AT5G20240.1 | Symbols: PI | K-box		
259	isotig18288	(ag)6	597	608	ATATCCTGCATCAGCAGCAA	TGGCATCCGTTcCTTCTTAC	165	region and MADS-box transcription	2E-15	83.8
								AT4G36480.1 | Symbols: ATLCB1,		
260	isotig04690	(aata)4	237	252	ATCAAACCCTTAGCCAACCG	AGGACCaTTACTGGCAGGTG	196	LCB1, EMB2779, FBR11 | long-chain	1E-30	135
261	isotig25647	(tc)5	232	241	ATCACGAGGACGTAGAAGCG	TCCCATTGAGTTGtTTgCAc	156			
262	isotig19934	(ctt)5	246	260	ATCGAGGCATCCAATATCCA	AACTGAAATTTaACAAAAACCCAA	152			
								AT1G17370.1 | Symbols: UBP1B |		
263	isotig10831	(gca)7	174	194	ATCTTGGGAAAATGCAGCAG	CGGCATGTACTCGAATCAAA	187	oligouridylate binding protein 1B |	7E-75	281
264	isotig11568	(ttc)7	3781	3801	ATGACGGCTCTCAGAAATGG	CCTCTTTCGTTCCCTCTCCT	172			
265	isotig06526	(tgc)7	1546	1566	ATGAGAGGGTCAGCTGGCTA	CCGGTCCAAGTGCATTACAT	160	AT3G21175.1 | Symbols: ZML1,	7E-14	79.8

								TIFY2B, GATA24 | ZIM-like 1 |		
								AT3G21175.1 | Symbols: ZML1,		
266	isotig06527	(tgc)7	1433	1453	ATGAGAGGGTCAGCTGGCTA	CCGGTCCAAGTGCATTACAT	160	TIFY2B, GATA24 | ZIM-like 1 |	6E-14	79.8
								AT5G62650.1 | Symbols: | Tic22-like		
267	isotig12520	(ttttc)3	1924	1938	ATGAGGATCTGCCATTCTGG	TTCCAgAAGCCATCATTTCC	151	family protein |	2E-20	101
								AT1G79550.1 | Symbols: PGK |		
268	isotig06195/96	(ga)5	1342	1351	ATGATGCGTAGCCTGCTTCT	GGTAACTCCCGAGCAACTCA	188	phosphoglycerate kinase |	6E-97	355
269	isotig19767	(ag)7	465	478	ATGATGCTTGCAAaGGAGGT	CCATCGATgCATCTTGGTAA	147			
270	isotig15821	(tga)4	427	438	ATGATGTTGAAGAGGTCCCG	TACTCCGTTCCCATGTCCTC	184			
								AT5G25220.1 | Symbols: KNAT3 |		
271	isotig11579	(caatt)3	563	577	ATGATTGCAATTTGGGGCTA	GTGCATAGCCAATGTTGTGG	177	KNOTTED1-like homeobox gene 3 |	7E-49	196
272	isotig09794	(ttc)4	186	197	AtgCCACTgTGTCTCcATCa	TCCCAGGTGCCATTCTTATC	151			
273	isotig05796-98	(tc)5	483	492	atggaatttggtgtgaagcc	ggTTTGGTCGGTgaTGAGTT	170			
								AT2G23350.1 | Symbols: PAB4, PABP4		
274	isotig10701	(cag)4	2251	2262	ATGGACCAGACTGAGGTGCT	CTAACCTGCCTCAGGAAACG	193	| poly(A) binding protein 4 |	1E-28	129
								ATCG01310.1 | Symbols: RPL2.2 |	1E-17	
275	isotig02401-08	(atgt)3	2488	2499	ATGGCAAAGTGGAGAAGGTG	TGTTTCCTCATAGGAACGCC	159	ribosomal protein L2 |	9	630
276	isotig26680	(tgc)6	304	321	ATGGCAACGGAAGAAGAAGA	ATGCAGAGCTTTTCAGCCAT	209			
277	isotig20314	(tgg)4	147	158	ATGGCAACGGTTGAAGTAGG	AAGCACTGGGGAAATGACAC	153			
								AT3G61710.1 | Symbols: ATATG6,		
								ATG6, BECLIN1, AtBECLIN1 |		
278	isotig12962	(ggag)5	83	102	ATGGCAATCTCTCTCCCCTT	ACAGTTCTGGCAAACCCATC	175	AUTOPHAGY 6	2E-47	190
279	isotig11172	(cta)4	173	184	ATGGGAACAAGACCAAGTGC	CAAAAGAACGACAACCAAGACA	171			
280	isotig13542	(gaa)6	989	1006	ATGGGATGTTTCGGAGACTG	AGGGCAACTTTCCTCTCCTC	207			
281	isotig26286	(tttta)3	233	247	ATGGTGGGAACAGCACTACC	TGGATGTGGACGTCTGAACT	147			
282	isotig11090	(tc)5	298	307	ATGTTATTGTCCTGCGGAGC	GCGCTGAGATTTCGAGAGAG	160	AT2G02760.1 | Symbols: ATUBC2,	9E-41	167

								UBC2 | ubiquiting-conjugating enzyme		
283	isotig01741-44	(gag)5	625	639	ATGTTCCAAaGGACCCATCA	CaATgGACGAGtTgGGttTT	176			
								AT3G14940.1 | Symbols: ATPPC3,		
284	isotig11625	(ac)7	227	240	ATTACCAGTCCGAACGGTTG	CATCTTCCGAAACCTTTCCA	168	PPC3 | phosphoenolpyruvate carboxylase	0	644
285	isotig13566	(gat)7	631	651	ATTCTCAGCTTTTGGACCGA	CCGACTTTTTGAAGGGAACA	182			
286	isotig17666	(cca)4	282	293	ATTGAGGTCTCGCTCTCCC	CTCCTTTCCGCTGTTGTTGT	187			
287	isotig03271	(ag)7	278	291	ATTtCAGtgCcgAaatgagC	TCCATCGACTGCTACGAGTG	184			
						TGCTaAATGGATGATTGATATGtTT				
288	isotig01761/65	(atcgt)3	187	201	ATTTCCTCGATGAGGGATCA	T	173			
289	isotig11171	(tag)4	558	569	CAAAAGAACGACAACCAAGACA	ATGGGAACAAGACCAAGTGC	171			
								AT4G18800.1 | Symbols: ATHSGBP,		
								ATRAB11B, ATRABA1D, RABA1d |		
290	isotig05980/81	(ag)7	163	176	CAAAAGCAAGAACAGAGGTGG	CTCCCGAATCACCGATTAGA	153	RAB	4E-60	232
291	isotig13971	(gaa)5	217	231	CAAACGCACCCACAATACCT	GGTGTGTGCATGAGAGAGGA	155			
292	isotig11738	(cttct)3	179	193	CAAAGGCTCCTAAACAAGCA	ACATACTGCTGCGCTTGAGA	148			
293	isotig11914	(ttc)5	2494	2508	CAAATCCAGGAAGAATCCGA	CTTCGAACTCGAATCAAGGC	177			
								AT4G17900.1 | Symbols: | PLATZ		
294	isotig16414	(aag)6	919	936	CAACACACGCGCTTTATACG	TTTCTACCCCTCACATTTCCA	185	transcription factor family protein	9E-18	91.7
295	isotig17906	(ttttc)4	428	447	CAACAGAAAACCATCATTTCACA	GCGTGAGAAATTGCATCAGA	186			
								AT3G02750.3 | Symbols: | Protein		
296	isotig12666	(aat)5	252	266	CAACTTGCGGTAGGTAGGGA	GAAGAGGCCTGAAATTCAAAA	148	phosphatase 2C family protein |	7E-26	119
								AT1G08190.1 | Symbols: ATVPS41,		
								ZIP2, VPS41, ATVAM2, VAM2 |		
297	isotig12767	(ac)14	208	235	CAAGCAATAATGCAAAGGGG	TAATGTTGGGTCTGGTGGGT	149	vacuolar	1E-33	145
298	isotig10711/12	(gct)6	1125	1142	CAAGCCCCATCATCATCTTC	CAGAATCCGAACCTGCAAAT	145			
299	isotig02866-70	(aag)7	207	227	CAAGCTCGAGCCTAAACCC	CATCCTCGTCTTCGTCCACT	146			

300	isotig13292	(aaac)3 (cgagc)	1626	1637	CAAGGaTCGCTTCCAGGTAA	TCCCTCCACTTCCCTTTCTT	180	AT5G11110.1 | Symbols: SPS1, ATSPS2F, KNS2, SPS2F | sucrose phosphate	7E-38	159
301	isotig02476-77/79-80	3	1015	1029	CAAGTGGATCACACCACAGG	TTCAAACCAAACCAAGCCTc	167			
302	isotig20082	(tgt)13	255	293	CAATAAAACACAAGGGGGCA	GCAAGGGCTCTTCTAAGCAA	149			
303	isotig04681	(ct)18	1930	1965	CAATGCTCACTGCATTGCTT	TTGTATTGAAATGGGCGGAT	148	AT1G52150.2 | Symbols: ATHB-15, ATHB15, CNA, ICU4 | Homeobox-leucine	4E-40	167
304	isotig13711	(aag)11	89	121	CAATGGCCAAAAaGAGGAAA	TTCTACCTCTTCAgGaTCGGA	146			
305	isotig14844	(ctca)3	1268	1279	CACACTTGAGGGACATGGAA	GtGAGctGCaAAGGGAAGAg	164			
306	isotig07277/78	(aag)4	1133	1144	CACAGGAGAGAAGGGAATGG	aCATTCGAATTCCGGTTTTG	158			
307	isotig07753/54	(tc)7	1237	1250	CACAGTTTCTTGGGCTGGTT	CAAACGTGGAAAGACGCTAA	197	AT4G02590.1 | Symbols: UNE12 | basic helix-loop-helix (bHLH)	3E-27	123
308	isotig01746/48	(gt)5	1342	1351	CACGAAGCCTTCTAGGTTGG	AGAGGGTCTAGTGGACTGTAAAA A	184			
309	isotig07874/75	(aag)5	735	749	CACGGACCGATTTCAAACTT	CGGTATGGAAGTAAGCCAGC	188	AT1G30540.1 | Symbols: | Actin-like ATPase superfamily protein |	3E-36	153
310	isotig14037	(cca)4	161	172	CACTCACACTCCCTTGCTTG	GCACGGGCTTGACTTATAGC	172			
311	isotig10804	(tca)4 (aaaag)	406	417	CAGAAGCCGAACCTGAACTC	gtgaagacctggttgggaaa	145			
312	isotig02096/76	3	209	223	CAGAAGCGCTGGAATTtCTT	CAATTGCATCGAGACTCTGAA	149	AT1G34750.1 | Symbols: | Protein phosphatase 2C family protein |	7E-20	99.6
313	isotig01066	(atc)6	1403	1420	CAGCATTCAGATCCTCAGCA	AGAATGGGACGGGATTTCAT	153	AT5G48240.3 | Symbols: | unknown protein; FUNCTIONS IN:	1E-18	95.6
314	isotig01067/70	(atc)6	1280	1297	CAGCATTCAGATCCTCAGCA	AGAATGGGACGGGATTTCAT	153	AT5G48240.3 | Symbols: | unknown protein; FUNCTIONS IN:	9E-19	95.6

								AT3G06130.1 | Symbols: | Heavy		
315	isotig12446	(cat)4	1284	1295	CAGCTCCTCATGCTCATCAA	GCTACCCATCAAGGACCAGA	148	metal transport/detoxification	2E-20	101
316	isotig11070	(ctttt)3	886	900	CAGCTCTTCAAGCAAGCAAA	CCTAACAAGCCAAACCCAAA	170			
								AT5G41060.1 | Symbols: | DHHC-type		
317	isotig06638	(ct)6	1376	1387	CAGGCCTGGTAAATCCTCG	TGGTTCTCCTTCTGCTCCAT	180	zinc finger family protein |	4E-21	103
318	isotig29174	(tct)4	256	267	CAGGTGATATCTCgCCCATC	AAgggATGTTTCAATGAAGAAAA	169			
319	isotig14236	(ttc)4	1229	1240	CAGGTTCCCATTCGAATCAT	GTCGTCGTCATCATCGTCTC	169			
320	isotig13281	(gga)4	1451	1462	CAGTCTTCTGCTGCATGGAA	CCAAGATCCACGACCACAG	156			
					CAGTGACACATAGTTAGGAAAAT			AT3G12710.1 | Symbols: | DNA		
321	isotig13505	(ctt)4	154	165	CA	CCTCCCACtaCCCAACTGTG	149	glycosylase superfamily protein |	1E-17	91.7
322	isotig17677	(aatac)3	106	120	CATCCAATAAGTGGGCCATC	ACTTCTGTGAGGCGCTTTGT	178			
323	isotig12438	(ag)6	81	92	CATCCCAGTGTCACTCCCTC	GCGATTTAGGGTTTTTGGAA	169			
324	isotig17935	(cttc)3	96	107	CATCTCCCTTCCCTTCCTTC	CCCTtttTCTGGAAGTGTGC	183			
								AT5G53300.1 | Symbols: UBC10 |		
325	isotig18404	(tc)7	649	662	caTGGAGGGCTCCTTATTGT	TGCATGAACACTTCTCCAGC	180	ubiquitin-conjugating enzyme 10 |	4E-78	291
								AT5G56670.1 | Symbols: | Ribosomal		
326	isotig21619	(aag)4	132	143	CATGGGAAAGGTACACGGAT	TTCCCAAATCCAACAACAGC	159	protein S30 family protein |	6E-27	121
327	isotig06376/77	(catc)3	1171	1182	CATGTCATGAGTCGAATCGC	ACTCATTAGCTcccGAAGCA	163			
328	isotig17733	(cca)4	340	351	CATGTCCAACTAAACTCCCTTG	CTGCATGAACGAGAAAGCAA	177			
329	isotig01349/57	(tgat)3	998	1009	caTTcAAGACACAcAAGcGt	TcCCagAAACTCaTGAGATGaA	178			
330	isotig20031	(ga)5	86	95	CATTCTCGAAATGGGGCTAA	CAGTGGACGGTGACATTCTG	147			
331	isotig10706	(tgt)4	1068	1079	CATTGAGATCACATCACCCG	aaatgttgggtctccaatgc	179			
332	isotig13717	(ttc)4	1094	1105	CATTGCATGGTTTTCACCAG	GGAACCAGTATCACAGGGGA	172			
								AT5G03760.1 | Symbols: ATCSLA09,		
333	isotig11921	(tctt)4	320	335	CATTTCATGCTTATGGCTCTGA	CTTCCATGCTTCTTGCTGTG	146	CSLA09, ATCSLA9, CSLA9, RAT4 |	3E-50	200
334	isotig12161	(gaat)3	189	200	CATTTCCTTGAAAGGGAGCA	AACTCTGCATTCATCCACCC	189	AT1G12240.1 | Symbols:	2E-11	71.9

								ATBETAFRUCT4, VAC-INV | Glycosyl hydrolases		
335	isotig12696	(ag)9	112	129	CATTTCTGTCCTCCCTTCCA	AAACCCTAGCAGACAAGcGA	149	AT1G58440.1 | Symbols: XF1, SQE1 | FAD/NAD(P)-binding oxidoreductase	4E-24	113
336	isotig06770/71	(ct)8	89	104	CCAAAAaCGGAAACGAACTT	GATGGAAAGCACGAGCAGAT	146	AT4G05020.2 | Symbols: NDB2 | NAD(P)H dehydrogenase B2 |	5E-21	103
337	isotig22468	(aaaat)3	78	92	CCAAAAGTTAAATCTTCCAGACG	TTGGCAGTAccaGATTGCTG	185			
338	isotig15185	(aaaat)3	1004	1018	CCAAACAGGCCCTAAAAcAa	TGAATGTTGTAGCGTTTGGc	175			
339	isotig08014/15	(aag)5	210	224	CCAAAGAACCATGTCCATCA	GCTAGTTCATTTTCGCACCC	147			
340	isotig12771	(aag)5	1280	1294	CCAAAGGAGGACATGCCTAA	ATCGTGGTGTTCCCAGACTC	167			
341	isotig13512	(tgc)4	303	314	CCAAAGTCTGCCAAAACCAT	GcCTGttTTTAtTcCgGTCA	147	AT2G33630.1 | Symbols: | NAD(P)-binding Rossmann-fold superfamily	8E-16	85.7
342	isotig05054/54	(aaat)3	650	661	CcAAGCTTGAATTTAGGCCA	TGCAATGACTCTATGGGAGG	160			
343	isotig18564	(ttc)5	597	611	CCAATCGAATCGTGTGTAAGA	AAACCTCGATCGAAAGGGAC	180			
344	isotig15895	(aataa)4	964	983	CCAATGAAGAAGAGCTTGACC	CATTGGGAGCAAAACAACCt	150	AT5G09230.7 | Symbols: SRT2 | sirtuin 2 | chr5:2871559-2873613	4E-63	242
345	isotig19538	(aag)5	241	255	CCAATGCCATGTAAGCCTTT	GGGGGAGATTGGATCTGTTT	164	AT4G27230.1 | Symbols: HTA2 | histone H2A 2 |	1E-22	107
346	isotig17471	(ct)7	316	329	CCaaTTGGTGTTAGAAACTCGG	GATgcTTTACTGCCCAtTgC	183			
347	isotig21667	(cac)5	198	212	CCACCACCTCCATACCACTA	TGGTGGTGGAGGAGATTTGTA	186			
348	isotig05943	(ttct)3	1239	1250	CCACCATTGGATTTTGGTTC	GCCACCCATTCAACAAGAGT	156	AT1G76170.1 | Symbols: | 2-thiocytidine tRNA biosynthesis protein,	2E-65	250
349	isotig01144-48	(aagag)3	2209	2223	CCACCTTCACtATGGGTCGT	TGAAACgGAGCTTATTTgGC	162			

350	isotig01456-61	(cca)5	618	632	ccaccttgtccatcatCCTC	GAGTATGGGAACCCGATTCA	164			
351	contig00750	(tct)4	191	202	CCAGCTTAACACgACAaCcA	GGGACTAGCtTATCCCCAGc	191			
352	isotig04258/59	(tc)6	462	473	ccaggaattcgtcactgctc	gGagGagGAGAGAGAgGGAG	169	AT5G53800.1 | Symbols: | unknown protein; Has 30201 Blast hits to	4E-21	103
353	isotig00534-44/46-48/50-55/57-59 /62	(caga)6	2166	2189	CCAGGGTCATtGctCaATTT	GCGGtcGTGATTGGTTAGAT	185			
354	isotig23863	(ctt)7	369	389	CCAGTATCACTGCATGTGGG	TCGTGGTGTTTGTGTCCCTA	188			
355	isotig06544/45	(cca)7	1598	1618	CCATAACGATCCCCATTACG	GTTCTAGCGGTGGAGAGTGC	184	AT3G26420.1 | Symbols: ATRZ-1A | RNA-binding (RRM/RBD/RNP motifs)	4E-12	73.8
356	isotig17733	(cca)4	340	351	CCATCAAGCCCCTCATAAAA	CCTACCAATAGAGGGTGCCA	177			
357	isotig10991/92	(gaa)5	236	250	CCATTTGAAGAGGACTTATGCC	GAGCGATGCTAGGAAATTCG	190			
358	contig00488	(attt)4	727	742	CCCAAAGGAATCACCAAAGA	GCCATCAGGAGTTGTCCATT	165			
359	isotig12143	(agaaa) 3	2188	2202	CCCAACAACAGCAGCAGATA	AGATTTGGATGagCTGGGTG	193	AT2G41170.1 | Symbols: | F-box family protein |	7E-11	69.9
360	isotig13421	(ta)5	1464	1473	CCCAACCTAAAATCTGCCAA	CCAAGGAGATCGGTCACAGT	145	AT3G56440.1 | Symbols: ATATG18D, ATG18D | homolog of yeast autophagy	6E-26	119
361	isotig12558	(ag)6	1962	1973	CCCACGATATCCCTCCTATG	CTTCGCAGAAAATCCCAAAC	150			
362	isotig15561	(tc)7	211	224	CCCATCCACCCTTAAATCCT	GGAGACCGAAAAGCAACTGA	156			
363	isotig13132	(tgc)4	131	142	CCCATGCAGAAAACAAaGAAA	AAGTCCCCAAAGGGAAGAGA	176	AT4G32570.1 | Symbols: TIFY8 | TIFY domain protein 8 |	6E-14	79.8
364	isotig03228/29	(tgta)3	1667	1678	CCCATTCATTGGTGACTTGA	CAAAGCACAAATGTAATCaAACG	189			
365	isotig11095	(aac)4	213	224	CCCCCTACTGATGACAAGGA	CAGACGAGGGTCAAGCTTCT	159	AT1G17880.1 | Symbols: BTF3, ATBTF3 | basic transcription factor 3	1E-31	137
366	isotig29623	(ga)7	143	156	CCCGTCAATTCAAGATCGAG	TGCAACGGTTTGAAAATGAG	170			
367	isotig21380	(ga)5	160	169	CCCGTTTTTCTTCTTCTCTCC	TCTGACGCCTACATCCAACA	148			
368	isotig06234/35	(gttt)5	48	67	CCcTTAGAAAAGACACCCAAA	CAGTTTCCACCACAGCAAGA	147			

369	isotig10205	(ttc)4	95	106	CCcTTTAAaTCAACGGCAGA	GACTTCGAAGACGAACAGCC	148			
370	isotig06498/99	(ct)5	155	164	CCCTTTTCTGCCTCTAGGGT	CATCGTATGTGGACGGTGAC	168	AT2G17190.1 | Symbols: | ubiquitin family protein |	2E-23	111
371	isotig05652	(ga)9	1301	1318	CCGAGGCTTCTGAGATTGTC	AACAGCTCACGCTACCCATT	147	AT5G54770.1 | Symbols: THI1, TZ, THI4 | thiazole biosynthetic enzyme,	3E-43	176
372	isotig12298	(catc)5	314	333	CCGATCATCTCTCTCCACCT	CTAGCGTAACGGACTTCGGA	145	AT2G24270.4 | Symbols: ALDH11A3 | aldehyde dehydrogenase 11A3 |	9E-69	262
373	isotig14740	(tc)5	148	157	CCGCACACAAATCAGGTAAA	GGGTGTGGGAGTTTTGTTTG	154	AT3G12130.1 | Symbols: | KH domain-containing protein / zinc finger	7E-16	85.7
374	isotig17824	(aaaga)3	230	244	CCGCATATCATCCATCACAC	TTACTGTCAATGGTGGGTGC	177			
375	isotig12801	(gat)4	219	230	CCGGAtCTTTTCCtCCAAGT	GCTGCTGCTCCTCTCTGTTT	165	AT5G22400.1 | Symbols: | Rho GTPase activating protein with	3E-25	117
376	isotig12623	(aca)9	403	429	CCGTTATCATCATCCCCATC	GGGAATTTATAGGCGAAGGC	148	AT5G07120.1 | Symbols: SNX2b | sorting nexin 2B |	5E-24	113
377	isotig02786/88/90/92	(attg)3	902	913	CCGTTCCCCATTTACCTTCT	TGACCTTGGTCTCAAAACCC	155			
378	isotig14866	(at)5	1229	1238	CCTACAAGCTTGGTTTGGATG	CGCCCCAAAGTCTTGAAAT	152	AT5G56280.1 | Symbols: CSN6A | COP9 signalosome subunit 6A |	7E-90	331
379	isotig11353/54	(ccta)3	389	400	CCTCAAGtAACCTCACCCCA	CCGAAAGATGAGGCAGAGAT	159			
380	isotig03156	(cct)4	761	772	CCTCCTCCACCACCTTACAA	GGGTGGATACTCTTTGTGGG	186			
381	isotig15931	(atcc)4	869	884	CCTGGACCATATTCGCCTAA	AGCCGATCCAATCTTCAATG	172	AT2G39830.1 | Symbols: DAR2 | DA1-related protein 2 |	1E-20	101
382	isotig10885	(aaaag)3	1344	1358	CCTGGGTTCAGTTGGACACT	CCAAACATGTGAATGAAAGACC	182			
383	isotig31913	(ct)6	151	162	CCTTAGTGGGGAGATTATTCACTT	GCACCATATCATTTCCCaaAA	193			
384	isotig21193	(ctg)5	498	512	CGAAtCCAAtGCAACACATC	CCGCCATGAAAGAAAGaAAA	148	AT2G29020.1 | Symbols: |	2E-14	79.8

								Rab5-interacting family protein |		
								AT5G11680.1 | Symbols: |		
								FUNCTIONS IN: molecular_function		
385	isotig16754	(ag)5	58	67	CGAATtGGGAATTGGAGAGA	CAAGACGAACATTTCGCTGA	153	unknown;	6E-22	105
								AT1G18070.3 | Symbols: | Translation		
386	isotig06552/53	(cag)5	65	79	CGACAGAAACCCTCGAAATC	CTCCAATCGTACGGCTCAC	154	elongation factor	3E-65	250
387	isotig15322	(ag)6	1164	1175	CGACATCTTTTGGAATGCCT	TCtGtTCtCCTCTTCTTCTTCTCTG	148			
388	isotig25039	(ga)7	113	126	CGAGAGAAGGGATTACGACG	GCCGGATCTGTGAGCATAAT	174			
389	isotig17060	(gct)7	780	800	CGATCACTTTGTGGCCTCTT	AGCGGAAACCCTAGCTCTCT	188			
390	isotig14284	(ct)7	1356	1369	CGATGGGAGCCCTAGAATTt	AGCTCGGAAGGGAGAGAGAG	150			
391	isotig03186	(ctc)5	1469	1483	cgcaatcaattcAaGCACAT	AAAACGTCGCAGAGAAAGGA	154			
								AT2G04030.1 | Symbols: CR88,		
								EMB1956, HSP90.5, Hsp88.1,		
392	isotig11802	(tct)6	1728	1745	CGCATCCAAATAGGCTTTGT	ATCAAGGACTGTGGAGGTGG	150	AtHsp90.5 |	6E-98	359
								AT3G13224.2 | Symbols: |		
								RNA-binding (RRM/RBD/RNP motifs)		
393	isotig07460/61	(ct)5	69	78	CGCCAACCCTAACCCTAGAT	GCCTGAGAGAGGCAGAGAGA	162	family	2E-22	107
								AT5G64260.1 | Symbols: EXL2 |		
394	isotig15330	(ta)5	1110	1119	CGGACAGAGATCAAGAGGTG	GGTACAcAGTCCACAACAGCA	149	EXORDIUM like 2 |	5E-29	129
395	isotig13569	(aaga)3	99	110	CGGACCACTGCCCTATCTAC	ATCCATCAAGCTTCCACCAC	170			
396	isotig13569	(aaga)3	99	110	CGGACCACTGCCCTATCTAC	TCCACCCAtTTTTCAAGAGG	170			
								AT2G28910.1 | Symbols: CXIP4 | CAX		
397	isotig06808/09	(aga)4	1071	1082	CGGATGATGAGGATGATGG	ATCCGAATCATCAGAATGCC	155	interacting protein 4 |	3E-12	73.8
398	isotig22967	(aag)6	154	171	CGGATTCCGATACAGAcACA	CCACATGACAATTTCGCATC	166			
399	isotig17943	(ctcc)3	398	409	CGGCACCACCTACTTCATCT	GACTGCGATTCTGTTCATGG	146			
400	isotig10832	(tgc)7	1030	1050	cggcatgtactcgaatcaaa	ATCTTGGGAAAATGCAGCAG	187	AT1G17370.1 | Symbols: UBP1B |	3E-70	266

								oligouridylate binding protein 1B |		
401	isotig23882	(atttt)3	396	410	CGGCGTCGATaTATTGCTTT	GCACCTGAGGTCCAGAACTT	193			
								AT1G76810.1 | Symbols: | eukaryotic		
402	isotig05873/74	(aag)5	1423	1437	CGGGATGACTTGTCtATCGAA	CACAGCTGAACAAGGGTTGa	169	translation initiation factor 2	8E-74	280
403	isotig11385	(tg)5	436	445	CGGGTGCCTAAGCTAAGATG	cCAACAGACGAGATTTCAACA	181			
404	isotig20564	(ag)7	569	582	cGGTGCAACAATCTTGAAGTAA	TTGCGCACtTAAAATTCGAT	204			
								AT4G21450.3 | Symbols: | PapD-like		
405	isotig15002	(cag)5	368	382	CGTATGCCTTTTTGGCAGAT	AGGGATCAAGACGCAACCT	172	superfamily protein |	3E-27	123
406	isotig29148	(cga)4	272	283	CGTCAGAGGAGTGGTGCTTT	GCCTCATCTCCGTCTTCATC	163			
407	isotig08477	(tct)4	368	379	CGTCCTCATCATTCACCTCA	GCCAAGCAAGAGAAGAGGAA	148			
408	isotig05545-47	(ga)8	1773	1788	CGTCGTCTAAATCCTCTGCC	ACTTCtTTTGCTGGGACCCt	147			
409	isotig12679	(aaat)4	606	621	CGTTCCCAGAGACGACATTT	TTTCAGGTTCCAGATTTGCC	145			
410	isotig13422	(gaaa)3	1264	1275	CTACTTGCCACAGGTCAGCA	TTGCATTCGTTCAAGACGAG	192			
411	isotig07252	(tgt)5	1218	1232	CTATCGATAAGGGCAGCGAG	AGCAGCAGCAGTACGTTGAA	196			
412	isotig09108	(taaa)3	84	95	CTCATCTTGAACACTGCCCA	GTCAAATTCCAAGTGTGCCC	156			
413	isotig17060	(gct)7	780	800	CTCCAAGGACAGGAGGCAT	GCCAGGAGGATGAATGAGAG	188			
414	isotig13641	(tttg)6	1403	1426	CTCCGAAGATGGACAATGTG	TGCAGCAACAACTGATGaAA	171			
415	contig00968	(cca)4	48	59	CTCCTCCACCTCCTAAGAAATC	GTGATGGTGGTGGTGGTGAT	159			
416	isotig01319/22	(gga)4	1204	1215	cTCCTCCTCCTCCTCCAAAG	GCATCTTCACTAGCGCCTTC	150			
								AT4G30190.2 | Symbols: HA2 |		
								H(+)-ATPase 2 |	1E-15	
417	isotig11618	(tct)7	329	349	CTCCTGCCAGAAAATGAAGC	AAACAAGGCAAGAAAGCGAA	162	chr4:14770820-14775920	7	557
418	isotig21278	(ta)7	382	395	CTCCTGTGTGCACCTGTACG	TAGCATTGGCGTGTTATCCA	206			
								AT3G07440.1 | Symbols: | unknown		
419	isotig17095	(ctg)5	816	830	CTCGTCTCCCTTCTTTGCTG	TCATCCATTGACGATGCTGT	160	protein; BEST Arabidopsis	3E-42	172
420	isotig07394	(ggc)7	462	482	CTGCAATGGATGGAAAGGAT	AGAACCACCACCATAGCTGC	179			

421	isotig12209	(ttat)3	1983	1994	CTGCTACTGCTGCCAAGATG	TCTCCCACAAACCAACAACA	165			
422	isotig08067	(gtttt)3	563	577	CTGCTGCTGATCGAGATTTG	TTCTGCATATGCTGCTGACC	173	AT2G45000.1 | Symbols: EMB2766 | structural constituent of nuclear	1E-19	97.6
423	isotig12836	(gaa)4	1562	1573	CTGTTAACATGGCATGGACG	CAAGTGAACCACACATGCAA	155			
424	isotig13589	(aaaag)4	1331	1350	CTTGAAGCCTTAAATGCGGT	CCGGTTGAGAAGGATCAAAA	194	AT1G05350.1 | Symbols: | NAD(P)-binding Rossmann-fold superfamily	8E-47	188
425	isotig18763	(ta)5	753	762	CTTGCTTCAGAGGACCAAATG	AGCTGATGGGAATGCTGACT	156	AT5G27720.1 | Symbols: emb1644 | Small nuclear ribonucleoprotein	1E-65	250
426	isotig02617/18	(ag)5	749	758	GAAAACCTGGGCAAGACAAG	AGAGACGTGACAGAGACCCAA	183	AT3G56800.1 | Symbols: CAM3, acam-3 | calmodulin 3 |	9E-95	347
427	isotig04793/95	(agc)5	1740	1754	GAAAATAGTGAGAGAATCTTCCG TG	CCTCATTTATTGGTGGTGGG	195	AT3G11700.1 | Symbols: FLA18 | FASCICLIN-like arabinogalactan	4E-15	83.8
428	isotig07578/79	(gag)5	1021	1035	GAAAATGGTGGGTGGAATTG	ATCCGTCTTTTAGGCCTGGT	192	AT3G07030.1 | Symbols: | Alba DNA/RNA-binding protein |	1E-14	81.8
429	isotig05280	(gact)3	76	87	GAAAaTTGGAAATGgCctGT	AActGAACAGTCAGccAGTCC	169			
430	isotig07670	(ga)7	491	504	GAAaTGCcTAaTtGCcGAGA	CACCAtTCATCGTCATCCTT	176			
431	isotig11553	(ctttt)4	294	313	GAAATTTTCTGGTGGGAGCA	GCCTGCTATGGAACATGAGG	146	AT5G64740.1 | Symbols: CESA6, IXR2, E112, PRC1 | cellulose synthase 6	3E-36	155
432	isotig24566	(gct)7	295	315	GAACGAAGGGAATAAAGGGC	GCTCGTCTCCACAAGGTTTC	173			
433	isotig19544	(at)5	256	265	GAACGGCCAAGATCACCTAA	TTAACGGCCGTGATTTGATT	153			
434	isotig03272	(ag)5	554	563	gaagaacatggatgCcACCT	TCCATCGACTGCTACGAGTG	162			
435	isotig06243	(gaa)5	5694	5708	GAAGAACTGGGCCATGAAAA	ATCAGCGGCTTCTGTGATCT	147			
436	isotig12267	(cac)4	1027	1038	GAAGAACTGGGGACGATTGA	CCCCACAGTGACAAACACAT	182	AT2G32830.1 | Symbols: PHT5, PHT1;5 | phosphate transporter 1;5 |	1E-24	115
437	isotig03356/58	(gat)4	1316	1327	GAAGAAGTACGGATCTGGCG	GGAGTTTCAGTTTGGCAGGA	161	AT3G08580.1 | Symbols: AAC1 |	1E-12	438

								ADP/ATP carrier 1 |	2	
								AT5G08300.1 | Symbols: |	1E-13	
438	isotig14531	(ct)11	1261	1282	GAAGCTATGGAACCCAGCAG	ACCCTCTCTTTCTCGCCATT	145	Succinyl-CoA ligase, alpha subunit |	7	488
439	isotig20342	(gcg)4	607	618	GAAGGTGAaGGCaGAAGCAC	CTTTGCAGCGACCTTTCTTC	157			
								AT3G15060.1 | Symbols: AtRABA1g,	1E-10	
440	isotig20606	(tc)6	524	535	GAATCGCCGATCAATACCAC	AAGAGGAAGAAGGAAGGCCC	147	RABA1g | RAB GTPase homolog A1G |	2	373
								AT1G79040.1 | Symbols: PSBR |		
441	isotig20085	(ctt)4	115	126	GAATGGCAAGCAGTGTTTTG	TGCCCCCATTAATTCCATAA	171	photosystem II subunit R |	5E-12	71.9
442	isotig18099	(aag)6	325	342	GAATTCGACGGAGAAGACGA	GCTTTTCCCTTTCCCTTGAC	179			
								AT5G53800.1 | Symbols: | unknown		
443	isotig04258	(tc)6	462	473	gaattcgtcactgctccctt	aAgGagGagGAGAGAGAgGG	169	protein; Has 30201 Blast hits to	4E-21	103
								AT5G53800.1 | Symbols: | unknown		
444	isotig04259	(tc)6	462	473	gaattcgtcactgctccctt	aAgGagGagGAGAGAGAgGG	169	protein; Has 30201 Blast hits to	2E-21	103
		(aagaa)						AT2G42590.3 | Symbols: GRF9, GF14		
445	isotig14823	3	185	199	GACAAACACGGTGAAGCTGA	GAGTCCTATGAATGCGGGAA	167	MU | general regulatory factor 9	2E-25	117
								AT5G28840.1 | Symbols: GME |		
446	isotig14243	(tttc)3	1269	1280	GACAACCGTTGAGCCCTAGA	TCACTCCTAACCTCgAACAGAA	170	GDP-D-mannose 3',5'-epimerase |	8E-19	95.6
								AT5G63110.1 | Symbols: HDA6, AXE1,		
447	isotig15122	(ctt)4	414	425	GACCAGAATCCAGCTCGAAG	CATTCTATGGGGAACCCTGA	196	ATHDA6, RTS1, RPD3B, SIL1 |	2E-19	97.6
448	isotig02042/43/46/47	(gat)6	391	408	gaccccaacaatctttctgc	CAtCTCCACCATCCCCAtag	148			
449	isotig00494	(aag)4	250	261	GACTGATCcaGGATGGATGG	TTgGtTGcCTCCTCTcTCTC	162			
								AT2G38710.1 | Symbols: | AMMECR1		
450	isotig10936	(ct)5	137	146	GACTTTCTCCGATCGCATTC	CGAAACAGTAGACCGCCATT	149	family | chr2:16184517-16186764	6E-13	75.8
451	isotig17586	(ccg)4	422	433	GAGAGAGCATTTGGCTGTCC	TGGACGGAACATTCGTGTTA	168			
								AT1G18660.4 | Symbols: | zinc finger		
452	isotig13041	(aag)4	1642	1653	GAGAGCAGCAGAACAAGGCT	TCGACATTTTCAATTGGCTG	173	(C3HC4-type RING finger) family	6E-11	69.9

453	isotig11803	(ttc)6	2525	2542	GAGAGGTAGAGTGGGGTGTGA	AAGCCAATCAAACAGCATCC	164			
454	isotig18972	(ctg)5	543	557	GAGATTGCTTTGTGGTTCCC	GGCACAAAGAGACATAGGGC	191	AT4G09320.1 | Symbols: NDPK1 | Nucleoside diphosphate kinase family	7E-24	111
455	isotig13875	(tctt)3	1430	1441	GAGCAGAACGCTCCTCGTTA	GAAACAACAACAACAACGGC	157	AT5G48230.2 | Symbols: EMB1276, ACAT2 | acetoacetyl-CoA thiolase 2 |	5E-48	192
456	isotig22320	(agg)7	323	343	GAGcGATGACATGGAGgTCT	GATCTTGGGGGTGGTTTTct	158			
457	isotig12426	(tttta)3	1900	1914	GAGGATCCAGACGTGGATGT	CAGAACCTAGCAGCAGATCAAA	162	AT1G13320.1 | Symbols: PP2AA3 | protein phosphatase 2A subunit A3 |	1E-17 6	618
458	isotig11541	(cac)5	106	120	GAGGCGCAAAAGAAAGAAGA	ATCCCTCTTCAGATGCTCCA	201	AT4G39350.1 | Symbols: CESA2, ATH-A, ATCESA2 | cellulose synthase A2	3E-61	238
459	isotig19717	(cagct) 3	266	280	GAGGGATGCATCAGGCTTAC	TTGAAATCGCCTGTATTCCC	168			
460	isotig01788/91	(gaa)4	428	439	GAGGGTTTtGTTGGgTTGAA	AAcCatCCcTTCCAATCACA	192			
461	isotig17350	(tttc)4	844	859	GAGTGCTTGGAATGGgAAGA	CAGACAGCAAGTTGGCaAAA	186	AT1G25560.1 | Symbols: TEM1, EDF1 | AP2/B3 transcription factor	8E-18	91.7
462	isotig11367	(gat)5	254	268	GATAGAGTGGAATGGcGgAA	GCGAAACACCCTTGTCCTTA	161			
463	isotig26426	(gggt)3	174	185	GATCCGAGCCATTGATTCAT	CGATCCTGGATTCCAATCTG	186			
464	isotig25991	(aaaat)3	165	179	GATCGCTTTTCAAGCTACGG	GAATGGCCACTTAGCTCTGG	189			
465	isotig10950	(gc)5	722	731	GATCTTAGTCGGCAGCCTTG	AGCGAGATCTCCATGGCTAA	192	AT3G16780.1 | Symbols: | Ribosomal protein L19e family protein |	1E-75	283
466	isotig16956	(atc)6	655	672	GATGCCTTAACAGCTGCCTC	TGCAAGGTTCCCTGGTAAAG	192			
467	isotig11188	(ctt)5	269	283	gatggacctgcttcaagACC	AAGCCAGAGGAGAAGAAGGC	175	AT1G07790.1 | Symbols: HTB1 | Histone superfamily protein |	2E-36	153
468	isotig12917	(ttc)5	213	227	GATGGAGTGGTGCTCTGGAT	ATCAGGCGAGCCATTATCAC	151	AT1G66430.1 | Symbols: | pfkB-like carbohydrate kinase family	3E-65	250

469	isotig14657	(aag)4	643	654	GATTCGATGATCGCTAGCCT	CCGTCACTGTCCTCCAATTT	161	AT3G57150.1 | Symbols: NAP57, AtNAP57, CBF5, AtCBF5 | homologue of	1E-35	151
470	isotig25126	(tc)6	332	343	GATTTTCCTGGCTTCCTTCC	GCGaAAGATAGAAAAGCCGA	158	AT2G45200.2 | Symbols: GOS12 | golgi snare 12 |	2E-20	99.6
471	isotig07297/98	(agc)4	637	648	GATTTTGAAGtCCGCCACTC	GGGTGGTTTCATCATCCTTG	166			
472	isotig02411/14/17	(ctcg)3	123	134	gaTTTTtGCGTATGCCCAct	GAGTCgAGTGACCCaATGGT	179			
473	isotig12970	(ga)7	559	572	GCAAAACCCAACACCTCATT	GGGTGGTGtaGTTGGATTGG	183			
474	isotig05919/20	(cctc)3	868	879	GCAAGAGAGAACGAGGAGGA	GAAGGCATGGTCAAAGGAAA	172	AT5G47730.1 | Symbols: | Sec14p-like phosphatidylinositol transfer	3E-37	157
475	isotig03296	(tggtg)3	371	385	GCACCTGTGGCAATTAAGGT	AGCCACATAGGAAAGCCAGA	165			
476	isotig10821	(ag)6	204	215	GCACGATTGCAGGATCTACC	GCCACCGAGAATAAGTTCCA	160			
477	isotig20174	(gct)6	619	636	GCAGCAGATGGTGTGAAAGA	ACATTGGCATTTGTCGTGAA	177	AT1G07990.1 | Symbols: | SIT4 phosphatase-associated family	4E-65	248
478	isotig11781	(ag)6	258	269	GCAGCAGCTCAGAAAGAGAGA	GGGTTTGAAAACGGATGAGA	171	AT3G63450.3 | Symbols: | RNA-binding (RRM/RBD/RNP motifs) family	2E-21	105
479	isotig13819	(gctt)3	1309	1320	GCAGGAGGCTTGATGAAGAG	ATTCAAAAGGCCAACACGAC	164	AT2G37620.1 | Symbols: ACT1, AAc1 | actin 1 | chr2:15779761-15781241	0	813
480	isotig12446	(cat)4	1284	1295	GCATGTACGCCACCTGTGTA	AAACCCCTTCTACCAGCAGC	148	AT3G06130.1 | Symbols: | Heavy metal transport/detoxification	2E-20	101
481	isotig21605	(ttgc)3	208	219	GCATTGGGTGgTGAATTTCT	GACCCGTGCAACTCTACTCC	203			
482	isotig07010/11	(gca)6	1071	1088	GCCAAAGACGTGACGAGAAG	TCTAAGCATAAAGGGGCGAA	172	AT1G07570.3 | Symbols: APK1A | Protein kinase superfamily protein |	4E-39	163
483	isotig11321	(aag)5	260	274	Gccacaagacgaaggagaag	TACATGATCGTATGCCACCC	199			
484	isotig10822	(ct)6	1195	1206	gccaccGAGAATAAGTTCCA	GCACGATTGCAGGATCTACC	160			

485	isotig31045	(aag)7	109	129	GCCACcTGGCAGTTACATTT	CTCCTCATCAGCCACCTGTT	151			
486	isotig06612/13	(ctg)7	487	507	GCcATGAACCCAGacAGATT	GATTCCAAATGAATGGGGTG	166			
487	isotig12618	(aaat)3	1826	1837	GCCCGTTTGGATCATTGTAT	ATTTAACGGTCTGCCATCCA	152	AT4G13430.1 | Symbols: IIL1, ATLEUC1 | isopropyl malate isomerase	2E-51	204
488	isotig23421	(tcttt)3	122	136	GCCCTTAATTAGTGGGATGG	AATGGTGtgTGTTGGAGCAA	172			
489	isotig22933	(ag)6	47	58	GCCGAGGAGAGTAGAAAAGG	ACGGGTCTCCCCCTTATAGA	149	AT2G43810.1 | Symbols: | Small nuclear ribonucleoprotein family	2E-17	89.7
490	isotig14348	(gca)4	1051	1062	GCCTTCATGTCGAAGGAGTC	CCAAACACAACCTCGTTCGT	191			
491	isotig13685	(gaa)7	1482	1502	GCCTTGGTTTACCCATTTGA	CACaAAACACTTCCTTGCCA	161			
492	isotig14600	(agaa)3	1037	1048	GCGAAGGAGTACCATCCAGA	TTGATCCACTTTGCTTGCTG	181	AT4G07990.1 | Symbols: | Chaperone DnaJ-domain superfamily protein	3E-18	93.7
493	isotig17184	(cga)4	616	627	GCGACAGCTCCTCTTTGTCT	CTGTCAAGCTGAGAAGTGCG	187			
494	isotig17495	(aag)10	115	144	GCGAGGCAACGAGAAaTTAT	GTCTTTGTCAAACGGACGGT	147			
495	isotig11089	(ga)5	203	212	GCGCTGAGATTTCGAGAGAG	ATGTTATTGTCCTGCGGAGC	160	AT2G02760.1 | Symbols: ATUBC2, UBC2 | ubiquiting-conjugating enzyme	8E-58	224
496	isotig24438	(aat)7	291	311	GCGgCCcgTACTActAAgc	tgcttttagcttgtccggtt	190			
497	isotig19842	(cttt)3	599	610	GCGGGAGATGCAAGTAGTGT	CCACAACCAACAGTCCAATCT	177	AT5G12190.1 | Symbols: | RNA-binding (RRM/RBD/RNP motifs) family	1E-34	147
498	isotig16351	(ac)6	92	103	GCGTAAATCGGACCTCTGAA	TTTTACTGGAAGCCTGGTCG	157			
499	isotig20832	(aga)4 (aacca)	586	597	GCGTCACCAACGGTATAGGT	GAGAAGAAGAAAGGGCGAGC	164			
500	isotig09673	4	112	131	GCtcaaaaaGACGaagaCCg	TTctttgggTGGTCTTTTGG	205			
501	isotig11147/48	(cca)4	656	667	GCTCTCCTCCTGCACCGTAG	GGAGAGCATGGTGGTGGATA	145			
502	isotig31218	(tgt)4	181	192	GCTGAAGTGCACCCCATTAT	CTCCACCAAACCTCCATTGT	167			
503	contig00545	(atttt)3	921	935	GCTGCAGTAAACGAACGAGA	TGCCACTCTGCTCTGTTTTG	200			

504	isotig04064	(gtt)6	1544	1561	GCTGTAGCTGCTGcttgTTG	CCTCCGCCTATACACCAAAA	149			
505	isotig13709	(gcc)6	1115	1132	GCTTGTTCTCGACTGTGGGT	GGAGTATTCCCCAACGGATT	187			
506	isotig12929	(ag)7	361	374	GCTTTTCTCAAACTGCAGGG	TACTCGGCCGCTTCATAACT	146			
507	isotig10701	(cag)4	2251	2262	GgAAGCTTTAAAGGCGAAGG	CTAACCTGCCTCAGGAAACG	193	AT2G23350.1 | Symbols: PAB4, PABP4 | poly(A) binding protein 4 |	1E-28	129
508	isotig02647/51	(tttc)3	1269	1280	GgAaGTtCcTGtCTTtCCca	CAGTGGGCATTTTTCAGCTT	165	AT5G07440.1 | Symbols: GDH2 | glutamate dehydrogenase 2 |	1E-29	131
509	isotig12223	(caa)5	394	408	GGAATCCGAACACAAACCCT	ATTGCTGCTGTTGTTGTTGC	165	AT4G00660.2 | Symbols: RH8, ATRH8 | RNAhelicase-like 8 |	1E-15 4	547
510	isotig13490	(caggc) 3	1520	1534	GGAATGTATTGTACATGATTGGC	GTGGGAACACACTTCACCCT	161	AT3G27925.1 | Symbols: DEGP1, Deg1 | DegP protease 1 |	1E-10 7	389
511	isotig01745/46	(gag)5	201	215	ggaGgAGCATGTTTCAAAGG	CaATgGACGAGtTgGGttTT	184			
512	isotig16614	(tttc)6	394	417	GGAGTGGACATCCAAGGAGA	CATTCGGGAGGCAGTTACAT	146			
513	isotig01895-906	(gag)4	442	453	GGATCTGCCTGGCATTAAAA	CTTCACTTCCTCCACCTTCG	190	AT3G46230.1 | Symbols: ATHSP17.4, HSP17.4 | heat shock protein 17.4	2E-24	113
514	isotig15832	(gagg)3	1122	1133	GGATGGTCATGTTTGGCAAT	CACATTTCCACCTATTTTCTAGCTT	147	AT4G25820.1 | Symbols: XTR9, XTH14, ATXTH14 | xyloglucan	2E-12	73.8
515	isotig13935	(ttct)3	1324	1335	GGCAAGGGAACATACAGGAA	ACACATGGGACCCTTTTCAC	152	AT1G09340.1 | Symbols: CRB, CSP41B, HIP1.3 | chloroplast RNA	2E-22	107
516	isotig07878	(tgg)4	992	1003	GGCAATGGTGCTTCTTTGTT	TCATGAGATCTTGGAAGTGGC	166	AT3G13810.2 | Symbols: IDD11 | indeterminate(ID)-domain 11 |	1E-27	125
517	isotig03484	(ag)5	174	183	GGCCGTTAATAACTCCTCCC	AGATAGCGTGGCAAGGTGAC	163			
518	isotig06072/73	(ag)5	304	313	GGCCTTGGAGACTTGGAACT	AAAAAGGCAtTCATAccCCC	154			
519	isotig10705	(tgt)4	2477	2488	GGGAAAAGGGTTGCTAGGTC	GGCTTGCCTGTGATAACAAAA	190	AT4G32551.2 | Symbols: LUG | LisH dimerisation motif;WD40/YVTN	8E-14	79.8
520	isotig16422	(tgaa)3	683	694	GGGCGTGGATGATACAATTT	CTTCCTCCACaCCCACATTC	165	AT5G27720.1 | Symbols: emb1644 |	2E-40	167

								Small nuclear ribonucleoprotein		
521	isotig06996	(ga)5	1587	1596	GGGTGGTCTATGGCGATCTA	GGGTGTGCACATcGAAAAG	199			
522	isotig0047577/81-83	()			GGgTtGTAGAgAATGaGGgAAa	AAACCTTTGGTTGCCTCCTC				
523	isotig02571-74	(ac)5	71	80	ggtcaatgcctacaacagga	CGGCTCCTAGCTATCCTTCC	156			
524	isotig07659/60	(aaga)3	715	726	GGTCAGTGAAAAGTGCCGAT	CCGTTGCCTTGAGATGGTTA	153			
525	isotig05467-70	(aga)5	509	523	GGTCTCTGCTCCTGAGATGG	GCGTTGCAACCTTCTTCTTC	206			
526	isotig23047	(tc)5	140	149	GGTGACCTCAAGCAAGGTTT	ccCagtttTacCatagccCa	200			
								AT2G03890.1 | Symbols: ATPI4K		
								GAMMA 7, UBDK GAMMA 7, PI4K		
527	isotig11658	(agc)6	2265	2282	GGTGAGAAGAACCAGCGGTA	AGCTTCACAAAGCTCGCACT	149	GAMMA 7 |	1E-25	119
528	isotig19481	(tatt)3	130	141	GGTGTCGTCCACACAATGAG	TGAGAAGATCAGTGCAAGGC	153			
								AT2G25910.2 | Symbols: | 3'-5'		
529	isotig15130	(ct)7	1173	1186	GGTTCTcTGCCTCCTGACTG	GGAAAATGCAATTCCAGAGC	168	exonuclease domain-containing protein	7E-22	105
								AT5G11670.1 | Symbols:		
								ATNADP-ME2, NADP-ME2 |	1E-10	
530	isotig12036	(tttg)5	2247	2266	GGTtGGTACACGGAGTTTGG	CACCACAAGATTCAAAACGG	148	NADP-malic enzyme 2 |	2	373
531	isotig01788-91	(gaa)4	428	439	GgTTTTGGGGAAAAacGAAT	tcTCtcCTTCtCCATTCAAgG	192			
532	isotig01792-93	(gaa)4	428	439	GgTTTTGGGGAAAAacGAAT	GGtgGaTTCCtcCAACAACT	182			
533	isotig03156	(cct)4	761	772	GTACAAGTCCCCACCACCAC	ATTCGGCTTTTCACCATCTT	186			
534	isotig14693	(tc)6	1170	1181	GTACATCGTCCGATCCACCT	GCCAGTTTCTTTCTGCAACC	147			
								AT5G62930.1 | Symbols: | SGNH		
535	isotig16913	(tctt)3	939	950	GtAGGTGTCGGCAaGAGCAG	tGCACTTGCATTTTGCTTTC	177	hydrolase-type esterase superfamily	2E-12	73.8
536	isotig11375/76	(tc)7	491	504	GTAtCCACACCTCTCCCTCG	CTAAGGCCGTGGAGATTTTG	201			
								AT3G02630.1 | Symbols: | Plant		
537	isotig14282	(aag)7	184	204	GTATCTACAAACCCGCCACC	ATCCCCTTCAGCTGGATTTT	169	stearoyl-acyl-carrier-protein	3E-40	167

538	isotig16717	(tct)6	273	290	GTCACTACCGAAGCCGAGTC	TTTGACGTTGTCGATGGTGT	190	AT1G31340.1 | Symbols: RUB1, NEDD8, ATRUB1 | related to ubiquitin 1	5E-90	331

3. Validation of full-sibship

A total of 9 EST-SSR makers (Table 3) with the M13 tail added were used to amply DNA from seedlings and parent trees. PCR reactions were conducted as described above, with an annealing temperature specific for each marker (Table 3). PCR products were prepared as described above before being separated on an ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and scored with GeneMapper (4.0) (Applied Biosystems, Foster City, CA, USA). Functional annotation of EST-SSRs was performed by applying a homology search of reassembled ESTs against the non-redundant (nr) NCBI database using the BLASTx algorithm (Altschul et al., 1997). Full-sibship was validated by the Mendelian segregation of parent allele in the progeny.

Table 3. Characteristics of 9 EST-SSR markers used for validation of full-sibship

Marker name	Repeat motif	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')	T_a (°C)	Range (bp)
isotig17398	(TG)8	ATTAATTCGTGTGTGCGCGT	CGTCAAACCGAGTGCCTAAT	55	152
isotig13347	(AG)16	CCGCAACTAACACGACATCA	AATTCCGGCCCCATAATAAG	61	164
isotig03894/95	(TTG)9	CCGCATTTGGTGGTTGTTAT	CCAAGCCTCTTTTCCTTTCC	59	209
isotig07448/49	(AG)15	TCATGGAATCCACACACTGG	GTAGGGCCATGCTTCGTAAA	61	158
isotig03209	(TC)16	TGCCCGTGATACCGATTATT	TGAAGCCTTTCGATTGCTCT	55	153
isotig19384	(ATG)8	TTGCGTAAATGCATCCAAAA	GAAGCCTATGCAAGATGCAA	55	181
LT102	(GA)9	GGAAACCAAACACAATCACT	TCCGTCACCACTAATCTCTC	56	163
LT141	(CT)14	CCCTGTAAATAACCCAATCA	CCGTTCTCTCCTTCTTCTCT	56	143
LT170	(CAG)6	GACGATGTTGTTCTTGGAGT	CAGACAGAAGCGAGTAGAGG	59	253

Results and Discussion

1. Development of informative EST-SSR markers

A total of 604 loci have been amplified with an initial annealing temperature of 55 °C, using the DNA of #UT108A and #UT23 as templates. Among the 66 previously tested loci, 64 were successfully amplified with the annealing temperature of 55 °C, and the other two, LT026 and LT061, were amplified at lower annealing temperatures, which was consistent with the previous study. A total of 112 new EST-SSR markers were chosen for estimation of PCR amplification success rate and polymorphism rate. Among them, 80 loci were successfully amplified with the annealing temperature of 55°C, giving rise to a PCR amplification success rate of 71.43% at this annealing temperature (Figure 1A). For instance, Among the six marker loci shown in Figure 1A, five of them were successful with the annealing temperature, while isotig07006 did not after many repeats. It is noteworthy that although the primer pair of marker isotig03894/95 amplified an unwanted band, the band was much shorter than the expected length, which can be separated easily. Thus, it was not regarded as a multiple bands producer, or excluded. Six of the 80 markers were excluded either because their PCR products were shorter than 150 nt (isotig07006, isotig02980, and isotig23746), or the corresponding primer pairs

amplified multiple bands within the expected length range (isotig17398, isotig19678, and isotig22892). The ideal annealing temperature of a certain primer pair was defined as the highest one with which a clear band was amplified in both parents. The ideal annealing temperatures of the remaining 74 primer pairs were determined by gradient PCR. For instance, the ideal annealing temperature of isotig07448/49 was identified as 61 ^{0}C (Figure 1B). As shown in Figure 1C, 12 markers had 55 °C as the optimized annealing temperature, eight 57 °C, twenty-two 59 °C, twenty-three 61 °C, eight 64 °C, and one 66 °C. Among the 74 markers that were analyzed on an ABI 3730 Genetic Analyzer, 63 were polymorphic in the two parent trees, giving an 85.14% of polymorphic loci. Xu et al. (2010) developed 176 primer pairs from ESTs of *L. tulipifera* and yielded 132 EST-SSR markers that amplified clear SSR bands with genomic DNA, giving rise to a PCR amplification success rate of 75%. Among the 132 markers, 66 were polymorphic in 5 *L. tulipifera* provenances, resulting 50% of polymorphic loci (Xu et al. 2010). Compared with the previously developed EST-SSR markers, the new markers had a similar PCR amplification success rate and a much higher percentage of polymorphism.

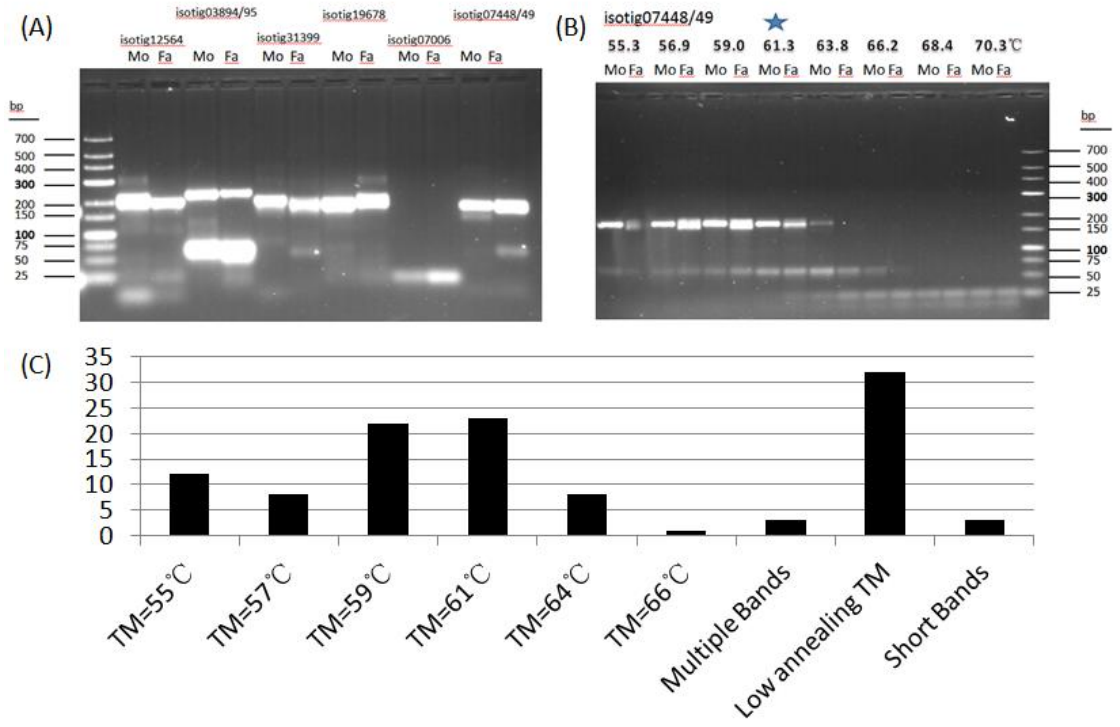

Figure 1. PCR amplification success rate study with 112 chosen SSR markers. The 112 marker loci were amplified with an initial annealing temperature of 55 ^{0}C, using the DNA of #UT108A (Mo) and #UT23 (Fa) as templates. Figure (A): an example of the amplification results of six marker loci. Figure (A): an example of the amplification results of six marker loci. (B) The ideal annealing temperature of isotig07448/49 was identified as 61 ^{0}C by gradient PCR. (C) Distribution of EST-SSR markers based on amplification results.

Ultimately, we will use F1 progeny from a controlled cross to construct the first genetic linkage map for yellow-poplar. Based on the requirement of JoinMap 4 (Kyazma B.V., Wageningen, Netherlands), a software for the calculation of genetic linkage maps in experimental populations, five groups of SSR markers will be informative for mapping: (1) both parent trees are heterozygous at a locus and do not share any same allele (AB×CD); (2) both two parent trees are heterozygous at a locus and share two same alleles (HK×HK); (3) and (4) one of the parent tree is homozygous at a locus, while the other one is heterozygous, and they share a same allele (LM×LL and NN×NP); (5) both of the two parent trees are heterozygous at a locus and share one and only one same allele (EF×EG). Among the 604 SSR markers that have been tested with #UT108A and #UT23, 20 of them belonged to group 1, 17 of them belonged to group 2, 29 of them belonged to group 3, 25 of them belonged to group 4, and 28 of them belonged to group 5 (Table 4). Since yellow-poplar has 19 chromosomes in one haploid genome, a relatively dense genetic linkage map with at least ten SSR loci in every chromosome of yellow-poplar requires at least 190 independent informative EST-SSR markers, which suggest that more informative SSR markers are still in need for constructing a framework genetic linkage map.

Table 4. Distribution of the 119 informative EST-SSR markers in the five groups

Female × male				
AB×CD	HK×HK	LM×LL	NN×NP	EF×EG
LT013	LT045	LT076	LT015	LT022
LT028	LT121	LT091	LT021	LT023
LT056	isotig02409/10	isotig12995	LT073	LT115
LT071	isotig08756	isotig26946	LT127	LT125
LT075	isotig22958	isotig20478	LT161	isotig31399
LT102	isotig18301	isotig06764	isotig13819	isotig11892
LT158	isotig11579	isotig11541	isotig11568	isotig22892
isotig03665/66/67/68	isotig09794	isotig12356	isotig14650	isotig13816
isotig21845	isotig11172	isotig22220	isotig05545/46	isotig23428
isotig15002	isotig11914	isotig29851	isotig13703	isotig03014/15/16
contig08221	isotig13292	isotig15677	isotig15780	isotig04682
isotig13874	isotig12446	isotig10096	isotig11587	isotig05098
isotig13485	isotig06638	isotig15448	isotig04394/96	isotig15788
isotig31377	isotig12161	isotig02871	isotig16780	isotig05673
isotig31749	isotig01144-48	isotig11012/13	isotig07339/40-b	isotig12099
isotig11620	isotig24566	isotig06928/29	isotig11386	isotig14887
isotig09599	isotig18099	isotig11090	isotig32954	isotig08911
isotig13800		isotig13566	isotig11030	isotig11603
isotig20355		isotig13971	isotig12520	isotig04692
isotig11945		isotig15185	isotig10701	isotig06536/37
		isotig06544/45	isotig14236	isotig06830
		isotig12801	isotig18763	isotig07277/78
		isotig15002	isotig11553	isotig01746/48
		isotig29148	isotig19842	isotig13505
		isotig17060	isotig12929	isotig18564
		isotig13641		isotig19538
		isotig01319/22		isotig05943
		isotig13589		isotig06552/53
		isotig01895-906		

2. Validation of full-sib seedlings from a controlled cross (108A and 23)

Progeny of 500 one-year-old yellow-poplars from controlled pollination was generated by Dr. Schlarbaum at The University of Tennessee, using #UT108A and #UT108B as mother trees, and #UT23 as the father tree, with an understanding that #UT108A and #UT108B were clones. However, our analysis with nine SSR markers, isotig17398, isotig13347, isotig03894/95, isotig07448/49, isotig03209, isotig19384, LT102, LT141, and LT170, revealed that less than 10% of seedlings were identified as full-sibs. When checking the genotypes of #UT108A and #UT108B with the same nine markers, the observed differences suggested two separate genotypes (Table 5). Subsequently, 213 seedlings were identified as full-sibs of #UT108A and #UT23. Since yellow-polar, as a hardwood tree species, has a large genome, more full-sibs are needed to construct a relatively dense genetic linkage map.

Table 5. Alleles at the nine SSR loci of the three parents used in genetype checking of #UT108A and #UT108B

	LT102	LT141	LT170	isotig17398	isotig13347	isotig03894/95	isotig07448/49	isotig03209	isotig19384
#UT108A	196, 204	160, 160	323, 333	172, 172	176, 178	228, 228	173, 173	217, 219	197, 197
#UT108B	178, 178	158, 158	329, 329	174, 174	180, 180	229, 232	171, 171	196, 203	200, 203
#UT23	178, 196	152, 158	331, 357	174, 174	172, 172	232, 232	161, 169	213, 213	203, 203

In summary, a total of 538 new EST-SSR markers had been available for yellow-poplar genome study, which were speculated to have relatively high PCR amplification success rate and percentage of polymorphism. Among the 538 new markers and 66 previously characterized polymorphic markers, a total of 119 informative SSR markers were identified for genetic linkage map construction with an F1 progeny with #UT108A and #UT23 as parents. The full-sibship for 213 seedlings were validated. These informative SSR markers and full-sib seedlings are essential in construction of linkage maps, which are valuable for future molecular breeding and quantitative trait locus (QTL) mapping, and as a framework for sequencing the *Liriodendron* genome. However, because yellow-poplar has a large genome with 19 chromosomes in one haploid genome, more informative SSR markers and full-sib seedlings are needed to construct a relatively dense genetic linkage map.

REFERENCES

Albert, V.A., Soltis, D.E., Carlson, J.E., Farmerie, W.G., Wall, P.K., Ilut, D.C., Solow, T.M., Mueller, L.A., Landherr, L.L., Hu, Y., et al. (2005). Floral gene resources from basal angiosperms for comparative genomics research. Bmc Plant Biol 5.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25, 3389-3402.

Chen, C., Wang, Y., Juan, S., and Huang, J. (2012). Chemical constituents from the stems of *Liriodendron tulipifera*. Chem Nat Compd 47, 1035-1037.

De Craene, L.P.R., Soltis, P.S., and Soltis, D.E. (2003). Evolution of floral structures in basal angiosperms. Int J Plant Sci 164, S329-S363.

Gwak, K., Kim, H., Ryu, K., Choi, H., Cho, D., Kim, P., and Choi, I. (2009). Growth improvement of *Liriodendron tulipifera* through SCB manure treatment. Forest Bioenergy 28, 7-14.

Hernandez, R., Davalos, J., Sonti, S., Kim, Y., and Moody, R. (1997). Strength and stiffness of reinforced yellow-poplar glued-laminated beams. Forest Service FPL, 554.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Depamphilis, C.W., Leebens-Mack, J., Muller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. PNAS 104, 19369-19374.

Jin, H., Do, J., Moon, D., Noh, E.W., Kim, W., and Kwon, M. (2011). EST analysis of functional genes associated with cell wall biosynthesis and modification in the secondary xylem of the yellow poplar (*Liriodendron tulipifera*) stem during early stage of tension wood formation. Planta 234, 959-977.

Kim, Y.H., Lee, S.M., Lee, H.W., and Lee, J.W. (2012). Physical and chemical characteristics of products from the torrefaction of yellow poplar (*Liriodendron tulipifera*). Bioresource technology 116, 120-125.

Klugh, K.R., and Cumming, J.R. (2007). Variations in organic acid exudation and aluminum resistance among arbuscular mycorrhizal species colonizing *Liriodendron tulipifera*. Tree physiology 27, 1103-1112.

Kobayashi, N., Horikishi, T., Katsuyama, H., Handa, T., and Takayanagi, K. (1998). A simple and efficient DNA extraction method for plants, especially woody plants. Plant Tissue Culture Biotechnol 4, 76-80.

Liang, H., Barakat, A., Schlarbaum, S.E., Mandoli, D.F., and Carlson, J.E. (2010). Comparison of gene order of GIGANTEA loci in yellow-poplar, monocots, and eudicots. Genome 53, 533-544.

Liang, H., Ayyampalayam, S., Wickett, N., Barakat, A., Xu, Y., Landherr, L., Ralph, P.E., Jiao, Y.N., Xu, T., Schlarbaum, S.E., et al. (2011). Generation of a large-scale genomic resource for functional and comparative genomics in *Liriodendron tulipifera* L. Tree Genet Genomes 7, 941-954.

Liang, H., Carlson, J.E., Leebens-Mack, J.H., Wall, P.K., Mueller, L.A., Buzgo, M., Landherr, L.L., Hu, Y., DiLoreto, D.S., Ilut, D.C., et al. (2008). An EST database for *Liriodendron tulipifera* L. floral buds: the first EST resource for functional and comparative genomics in *Liriodendron*. Tree Genet Genomes 4, 419-433.

Liang, H., Fang, E.G., Tomkins, J.P., Luo, M., Kudrna, D., Kim, H.R., Arumuganathan, K., Zhao, S., Leebens-Mack, J., Schlarbaum, S.E., et al. (2007). Development of a BAC library for yellow-poplar (*Liriodendron tulipifera*) and the identification of

genes associated with flower development and lignin biosynthesis. Tree Genet Genomes 3, 215-225.

Moody, R., Hernandez, R., Davalos, J., and Sonti, S. (1993). Yellow poplar glulam timber beam performance. Forest Service FPL, 520.

Moore, M.J., Bell, C.D., Soltis, P.S., and Soltis, D.E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. PNAS 104, 19363-19368.

Neale, D.B., and Kremer, A. (2011). Forest tree genomics: growing resources and applications. Nature reviews Genetics 12, 111-122.

Oetting, W.S., Lee, K.H., Flanders, D.J., Wiesner, G.L., Sellers, T.A. and King R.A. (1995). Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. Genomics 30, 450-458

Qiu, Y., Dombrovska, O., Lee, J., Li, L.B., Whitlock, B.A., Bernasconi-Quadroni, F., Rest, J.S., Davis, C.C., Borsch, T., Hilu, K.W., et al. (2005). Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. Int J Plant Sci 166, 815-842.

Roy, J., Saugier, B., and Mooney, H. (2001). Terrestrial Global Productivity (Academic Press, London).

Soltis, P.S., Soltis, D.E., Chase, M.W., Endress, P.K., and Crane, P.R. (2004). The diversification of flowering plants. Assembling the Tree Of Life, 154-167.

Williams, R., and Feist, W. (2004). Durability of yellow-poplar and sweetgum and service life of finishes after long-term exposure. Forest Prod J 54, 96-101.

Xu, M., Li, H., and Zhang, B. (2006). Fifteen polymorphic simple sequence repeat markers from expressed sequence tags of *Liriodendron tulipifera*. Mol Ecol Notes 6, 728-730.

Xu, M., Sun, Y., and Li, H. (2010). EST-SSRs development and paternity analysis for *Liriodendron* spp. New Forest 40, 361-382.

Zahn, L., Leebens-Mack, J., DePamphilis, C.W., Ma, H., and Theissen, G. (2005). To B or Not to B a flower: the role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. J Hered 96, 225-240.

CHAPTER THREE

DETECTING GENETIC CONSTITUTION OF TWO *Liriodendron* SEED ORCHARDS WITH EST-SSR MARKERS

Abstract

Because of its ecological and economic value and phylogenetic position as a basal angiosperm, genomic resources, such as EST databases and BAC libraries, have been developed for *Liriodendron tulipifera* L., a tree species native to eastern North America. However, molecular marker resources are under developed with only a few hundred SSRs. One hundred seventy six SSR markers have been previously tested for amplification success and polymorphism rate. The lack of molecular markers has hindered the construction of genetic maps, molecular breeding, and study of population dynamics and adaptive variation in *Liriodendron*. In this study, we characterized 20 EST-SSR markers with 174 trees from two yellow-poplar seed orchards and the US National Arboretum, and provided a first look at the genetic diversity and allele richness among selections of this unique native species. The two yellow-poplar seed orchards, residing in Knoxville, Tennessee, and Clemson, South Carolina, were established in 1966 and 1976, respectively, and have provided seeds for distribution. Analysis revealed only

one locus significantly deviating from Hardy-Weinberg proportions in Knoxville population ($p>0.05$). In addition, the Clemson orchard exhibited higher values of observed and effective number of alleles, observed heterozygosity, and Nei's expected heterozygosity than the Knoxville orchard. Therefore, revealing larger genetic diversity in the Clemson seed orchards.

Introduction

The genus *Liriodendron* consists of two species, one native to China and Vietnam [*Liriodendron chinense* (Hemsl.) Sarg], and another to eastern North America (*Liriodendron tulipifera* L.). They are quite similar morphologically, although *L. chinense* is smaller in stature than *L. tulipifera*. Analysis of fossil evidence, allozyme polymorphisms, and chloroplast DNA variation suggested that these two species separated 10~16 million years ago. However, their hybrids are highly vigorous in growth (Wang 2005). *L. chinense* pupulates small isolated areas and is now an endangered species due to its limited seed production (He and Hao 1999). In contrast, *L. tulipifera*, commonly known as yellow-poplar, tulip tree, or tulip-poplar, is a fast-growing and one of the most important hardwood species used for the production of pulpwood and timber. It is distributed from 28° to 43° north latitude and predominantly east of the Mississippi

River (Sewell et al. 1996). As a member of the *Magnoliaceae* family, *L. tulipifera* occupies an important phylogenetic position as a basal angiosperm species and 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). The recent assembly of the mitochondrial genome of *L. tulipifera* confirmed its exceptionally slow rate of evolution with the lowest known genome-wide absolute silent substitution rate (Richardson et al. 2013).

Because of its ecological and economic value and phylogenetic position as a basal angiosperm, genomic resources, such as expressed sequence tag (EST) databases and genomic DNA libraries, have been developed for *L. tulipifera*. The deep transcriptome sequence resource reported by Liang et al. (2011) contained 568.5 Mb bases and were developed from ten different tissue types (premeiotic flower buds, postmeiotic flower buds, open flowers, developing fruit, terminal buds, leaves, cambium, xylem, roots, and seedlings). The EST dataset for early stage of tension wood formation contained 5,982 high-quality ESTs, which were clustered into 1,733 unigenes (Jin et al. 2011). Recently, approximate 4.2 Gb of new EST data from ozone treatment leaves have been obtained with Illumina sequencing (personal communication with Dr. John Carlson at The Pennsylvanian State University, USA). The genomic DNA libraries includes a 5X BAC library with 73,728 large-insert clones and a shotgun library containing 3,072 clones,

with an average insert size of 117 kb and 3 kb, respectively (Liang et al. 2007). In addition, the chloroplast genome of *L. tulipifera* has been assembled (Cai et al. 2006). These resources have generated several thousand of putative simple sequence repeat (SSR) markers (also called microsatellites). Once being characterized, these markers will be valuable in studies of genetic diversity and functional diversity related to adaptive variation, as well as in molecular breeding and construction of genetic maps (e.g. Tomlinson et al. 2000). SSR markers are co-dominant, easily reproduced and scored, highly polymorphic, abundant through the genome, and have higher information content than isoenzyme and dominant markers (Zane et al. 2002).

Compared to other forest tree species, such as *Populus*, *Eucalyptus*, and loblolly pines, SSR markers are under developed in *Liriodendron*. So far only 171 EST-SSR markers have been developed from *L. tulipifera* sequences and tested for their transferability to *L. chinense* (Xu et al. 2010; Yang et al. 2012), in addition to 14 *L. chinense* genomic microsatellites and 11 *L. tulipifera* chloroplast SSRs. This explains why no molecular tools have been used in breeding programs and no linkage maps have been constructed in *Liriodendron*. In contrast to *L. chinense* in China, information of genetic variation of *L. tulipifera* in the United States is limited. The very few reports include surveys of restriction site variation in chloroplast DNA and allozymes, which suggested two distinct haplotypes (northern and southern) with an intermediate group that

was putatively formed from recent hybridizations between these entities (Parks et al. 1994; Sewell et al. 1996). Recently, assessment of genetic variation of *L. tulipifera* populations in unmanaged forests of the Southeast United States was reported by Kovach (1992) utilizing amplified fragment length polymorphism with five primers.

In this study, we characterized 20 EST-SSR markers with a total of 174 trees from two yellow-poplar seed orchards and the US National Arboretum. The two yellow-poplar seed orchards are residing in Knoxville, Tennessee, and Clemson, South Carolina, and have produced seeds for distribution. The Clemson orchard contains 165 trees and was established in 1976 by grafting scions of big trees from the horseback riding trails in South Carolina by Dr. Roland E. Schoenike (Figure 2A). Established in 1966, the Tennessee orchard contains 100 grafted clones, representing 31 genotypes (Figure 2B). Genetic diversity of yellow-poplar seed orchards has not received any considerable attention. Our study not only has discovered highly polymorphic and multiallelic loci that will be useful in the study of population dynamics and adaptive variation in *Liriodendron*, but also provided a first look at the genetic diversity and allele richness among selections of this unique native species.

Figure 2. *Liriodendron* orchards in Clemson, SC (A), and Knoxville, TN (B).

Materials and Methods

1. Plant materials and DNA isolation

Fresh leaves of *Liriodendron* trees from two seed orchards in Clemson University, South Carolina, and The University of Tennessee, Tennessee, USA, were collected in the spring of 2013 and stored in plastic bags at -80°C prior to DNA isolation. Leaves from a *Liriodendron tulipifera* tree (accession number 70921 H) and *Liriodendron Chinense* tree (accession number 62539.H) from the US National Arboretum (collected by Kevin Conrad) were also included in the study. Total genomic DNA was isolated from leaves using a CTAB protocol as described in Kobayashi et al. (1998) and suspended in TE buffer (Tris base 6.1g/L, EDTA 0.37 g/L, pH 8). The quality and concentrations of genomic DNA from individual plants were determined with a NanoDrop 3300 (Thermo Scientific, Wilmington, Delaware, USA) and by electrophoresis on 0.8% agarose gels.

2. Distinguishing two *Liriodendron* species based on maturase K sequence and leaf morphology

There is a mix of *L. tulipifera* and *L. chinense* in the Clemson orchard according to the records provided by Mr. Knight Cox, manager of the Clemson University Experimental Forest. Due to missing labels on the surviving trees and death of trees, records for the trees on the site could not be matched. In order to distinguish these two species, a chloroplast gene, matK (maturase K), was amplified with forward (5'-CGATCTATTCATTCAATATTTC-3') and reverse primers (5'-TCTAGCACACGAAAGTCGAAGT-3') in a 12.5-µl reaction containing 6.875 uL ddH_2O, 1 uL $MgCl_2$ (25 mM), 0.5 uL forward primer (10uM), 0.5 uL reverse primer (10uM), 0.25 uL dNTPs (10 mM each), 0.25 uL BSA (0.8ug/uL), 0.125 uL Taq Pololymerase (5u/uL), 0.5 uL DNA (in uL) (~20ng/ul), 2.50 uL 5X PCR buffer (-Mg). The conditions for polymerase chain reactions (PCR) were as follows: 5 minutes of initial denaturation at 94°C, 35 cycles of touch-down PCR with 30 seconds of denaturation at 94°C, 30 seconds of annealing at 60-50°C (first cycle 60°, then each subsequent cycle 1°C lower than the previous until 51°C annealing temperature. Then 25 cycles each with a 50°C annealing temperature), and 3 minutes of extension at 72°C, and a final extension at 72°C for 10 minutes. Before being sequenced with 1 ul of 10 uM forward or reverse

primer, PCR products were cleaned with ExoAP mix (89 uL H_2O + 10 uL 5000U/mL Antarctic Phosphatase + 1 uL 20000U/mL Exonuclease I) for 30 minutes in a reaction containing 1uL of PCR product and 1uL of ExoAP mix, followed by a heat inactivation step at 80^0C for 15min. An 834 bp-segment of maturase K gene from each tree was used for alignment with MUSCLE, curated with Gblocks, and a phylogenetic tree was built with maximum likelihood (PhyML) (http://www.phylogeny.fr/) (Dereeper et al. 2008). The maturase K gene sequence of *L. tulipifera* (GI: 5731451), *L. chinense* (GI: 7239759), and a hybrid (GI: 389955358) available in GenBank were included in the analysis.

3. *L. tulipifera* EST-SSR markers, PCR amplification, and allele sizing

Seven EST-SSR markers (LT002, LT015, LT021, LT086, LT096, LT131, LT157) previously characterized by electrophoresis on 8% polyacrylamide gels (Xu et al. 2011) as well as thirteen markers (isotig13819, isotig19384, isotig23696, isotig12995, isotig13485, isotig31749, isotig14887, isotig11620, isotig08911, isotig11603, isotig22220, isotig04692, isotig09599) mined from a comprehensive EST dataset (Table 6; Liang et al. 2011) were used for amplification with genomic DNA of *Liriodendron* trees from both the Clemson University and The University of Tennessee seed orchards as well as the US National Arboretum. For a more cost-effective primer screening, a M13 tail (5'-CACGACGTTGTAAAACGAC-3') was added to the 5'-end of the forward

primer of each marker pair in order to amplify the fragments using a complementary adapter with a fluorescent dye (6-FAM, VIC, NED, or PET) at its 5'-end (Applied Biosystems, Foster City, California, USA). Polymerase chain reactions were carried out in a 12.5-µl solution comprising: approximate 75 ng DNA template, 0.052 U/µL Promega Taq DNA polymerase, 0.16 nM forward primer, 0.4 nM reverse primer, 0.4 nM fluorescent M13 primer, 0.24 mM each dNTPs, and 1.2×Promega PCR buffer. The PCR profile consisted of an initial denaturation at 94℃ for 3 minutes followed by 10 cycles of 1 minute at 94℃, 1 minute at annealing temperature (Ta, Appendix I), and 1 minute 15 seconds at 72℃, and then 35 cycles of 1 minute at 94℃, 1 minute at 58℃, and 1 minute at 72°C, with a final extension of amplified DNA at 72℃ for 5 minutes.

An aliquot of 1.5 µl PCR products were treated with 1.5 ul of 10-fold diluted ExoSAP-IT (Affymetrix Inc. Cleveland, OH, USA) to remove single stranded primers which might influence fragment analysis at 37°C for 30 minutes and then at 80°C for 15 minutes. After being diluted to 100 ng/µl, 1ul of each sample was mixed with 0.1 ul of LIZ600 and 8.9 ul of Hi-Di Formamide, denatured at 95 C for 5 minutes, and then put on ice for 10 minutes before being separated on an ABI 3730 Genetic Analyzer. The Dye set was DS-33 (6-FAM, VIC, NED, PET and LIZ). Allele sizes were scored with GeneMapper (4.0) (Applied Biosystems, Foster City, California, USA). Functional annotation of EST-SSRs was performed by applying a homology search of reassembled

ESTs against the non-redundant (nr) NCBI database using the BLASTx algorithm (Altschul et al. 1997).

Table 6. Characteristics of 20 EST-SSR loci.

Marker name	Repeat motif	Expected size	Stuttering	Presence of null alleles	>50% of alleles are of one allele size class	Annealing temperature ^{0}C
LT002	(GCA)8	189	N	N	N	59
LT015	(CCGAAC)5	110	N	N	N	59
LT021	(TTC)8	180	N	N	N	57
LT086	(CTT)10	274	N	Y	Y	55
LT096	(CT)20	272	N	Y	N	55
LT131	(AC)22	240	Y	Y	N	55
LT157	(TTC)6	222	Y	Y	Y	55
isotig13819	(AG)10	183	N	Y	Y	57
isotig19384	(ATG)8	181	Y	Y	Y	55
isotig23696	(CT)18	198	N	Y	N	55
isotig12995	(TG)14	223	N	Y	N	57
isotig13485	(TC)8	205	N	N	N	55
isotig31749	(TCT)10	164	Y	Y	N	55
isotig14887	(AAT)8	171	Y	Y	N	55
isotig11620	(TG)13	160	N	N	N	55
isotig08911	(GA)18	157	N	N	N	55
isotig11603	(TC)10	167	N	Y	Y	55
isotig22220	(TC)11	189	N	Y	N	55
isotig04692	(CA)17	177	N	Y	N	55
isotig09599	(CT)10	156	N	Y	N	55

4. Data analysis

Overrall homozygote excess and high observed homozygote frequencies can sometimes cause null alleles, large allele drop-out, and stuttering. MICRO-CHECKER (van Oosterhout et al. 2004) was employed to check for potential genotyping errors arising from null alleles, large allele drop-out, and stuttering. Observed and expected heterozygosities and polymorphic information content (PIC) were performed using Cervus 2.0 (Marshall et al. 1998). Deviations from Hardy–Weinberg equilibrium and the Shannon's Information index were calculated with GENEPOP (http://genepop.curtin.edu.au/, Raymond M, Rousset 1995) and POPGENE version 1.32 (Yeh et al 2000).

Results and Discussion

1. Distinguishing two *Liriodendron* species with maturase K primer

The two *Liriodendron* species are thought to have separated 10–16 million years ago (Parks and Wendel 1990), but hybridize readily (Merkle et al. 1993). They are quite similar morphologically, except that *L. chinense* is smaller in stature and has larger leaves more deeply lobed and smaller flowers. However, our attempt to tell these two species

apart by leaf shape failed since it varied depending on age (Figure 3). Due to the high location of flowers, we were not able to obtain any flowers. In order to distinguish these two species, the maturase K gene sequence was amplified from each tree. Among the 165 *Liriodendron* trees in Clemson orchard, PCR amplification was successful for 143 trees (Figure 4). When the amplicons were pair-end sequenced, an 834-bp segment of high quality was obtained for each tree, representing 55% of the full-length gene. There were only eight different nucleotides between the two *Liriodendron* species within the 834-bp segment. As shown in Figure 5, only Tree#CU24 was not *L. tulipifera*, grouping with the hybrid. Tree#CU24 was more close to *L. Chinense* in heredity than to *L. tulipifera* in the matK sequence since there were seven different nucleotides between Tree#CU24 and *L. tulipifera* and there were only two between Tree#CU24 and *L. Chinense*. Because PCR amplification failed for 22 trees, these trees were excluded from the remaining analyses. Our study indicates that *L. tulipifera*, *L. Chinense*, and their hybrids contain unique nucleotide compositions in the maturase K gene sequence that can be utilized in distinguishing the genotypes. The *matK* gene locates within the intron of the *trnK* and codes for maturase like protein involved in Group II intron splicing (Turmel et al., 2006). The *trnKUUU-matK* region, ranging from approximately 2.2 kb (liverworts) to 2.6 kb (seed plants) in size, is universally present in land plants and only few exceptions of a secondary loss or reorganisations are known to date (Wicke and Quandt 2009 and

reference therein). Because the *matK* gene evolves more rapidly, compared to other plastid genes, it has become a valuable marker for systematic and evolutionary studies.

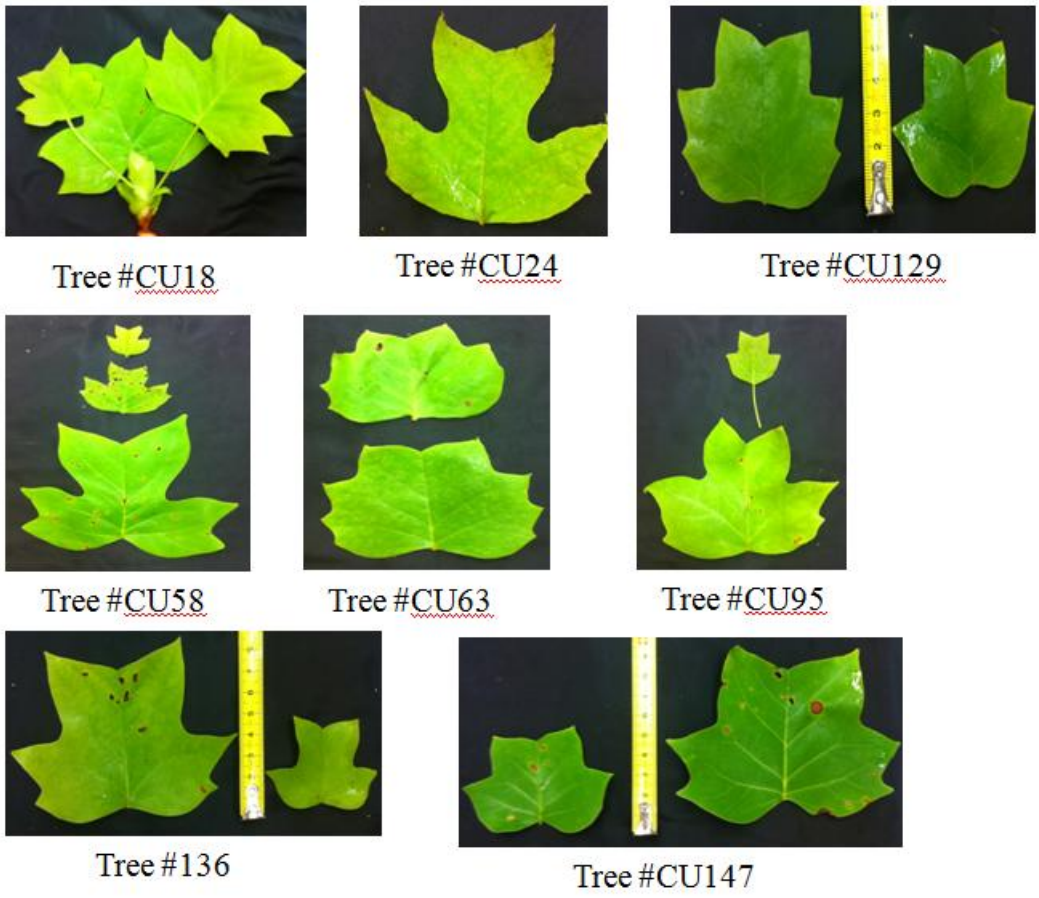

Figure 3. Examples of leaf shape of *Liriodendron* trees in Clemson orchard.

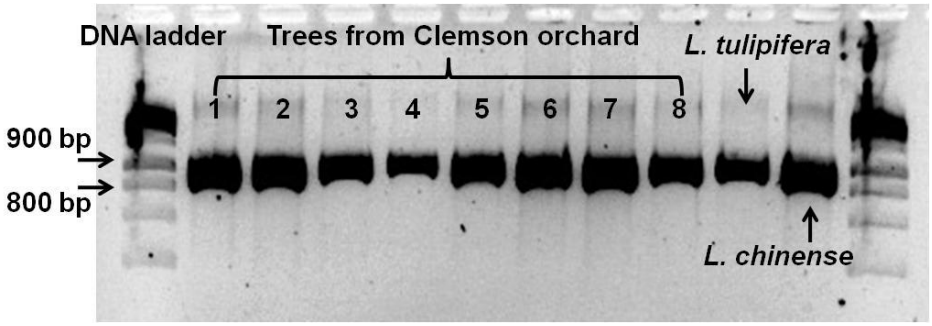

Figure 4. PCR amplification of maturase K gene in *Liriodendron*. The *L. tulipifera* and *L. chinense* samples were from the US National Arboretum provided by Kevin Conrad.

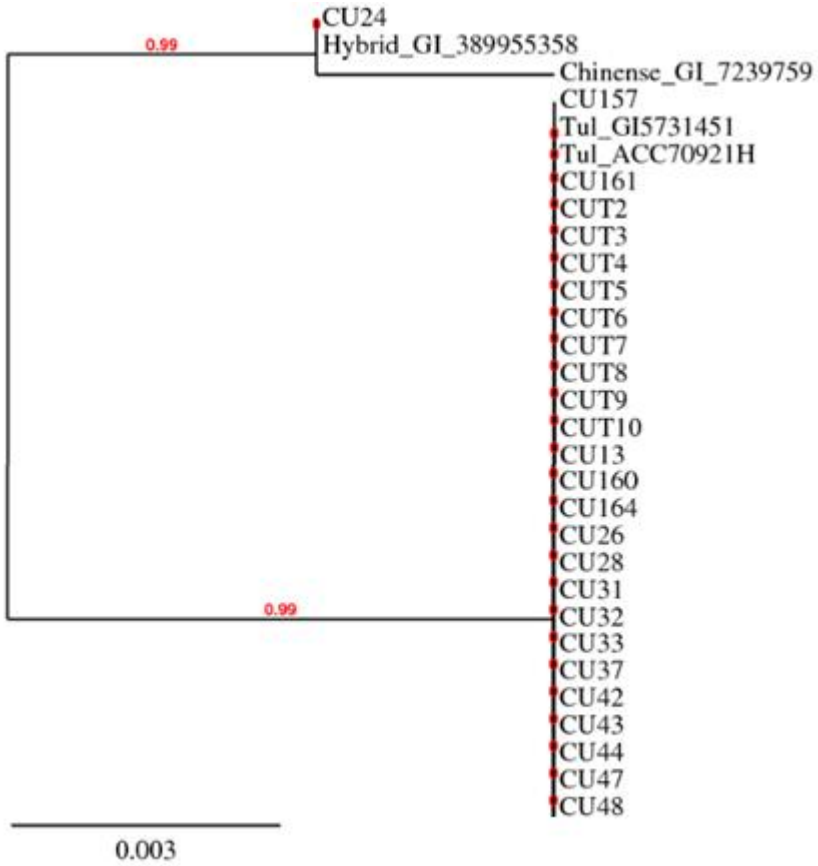

Figure 5. Comparison of maturase K gene sequences.

2. Characterization of 20 EST-SSR loci with *L. tulipifera*

All *L. tulipifera* trees from both orchards and the US National Arboretum were used for the estimation of genetic variation based on 20 EST-SSR loci. No evidence for large allele dropout was found for any markers. Stuttering might have resulted in scoring errors in five markers (LT131, LT157, isotig19384, isotig31749, and isotig14887) (Table 6). Since PCR reactions needed to be optimized for these markers to avoid this issue, they were excluded in further analyses. Null alleles might be present at nine of the remaining 15 loci, as suggested by the general excess homozygotes for most allele size classes, with three of them having more than 50% of alleles at each locus being one allele size class. The number of alleles per locus ranged from 3 to 26 (mean=13.0) (Table 7). The observed and expected heterozygosities (H_0 and H_e) ranged from 0.17 to 0.89 and from 0.19 to 0.93, with averages of 0.62 and 0.74, respectively. The polymorphic information content (PIC) ranged from 0.17 to 0.92, with an average of 0.71. Two loci (LT015 and isotig08911) were found to be out of Hardy-Weinberg equilibrium (Hardy Weinberg Exact Tests, $p>0.05$) (Appendix I). Overall, we obtained three markers (LT002, isotig13485, and isotig11620) with a PIC ≥ 0.5, without the presence of null alleles, and within Hardy-Weinberg equilibrium. These highly polymorphic, multiallelic loci will be useful in the study of population dynamics and adaptive variation in *Liriodendron*. Lastly, two markers, isotig19384 and isotig23696, amplified alleles 191/191 and 188/190,

respectively, and were at least 9 nucleotides smaller in *L. Chinense* than in *L. tulipifera*. Because only one *L. Chinense* tree was included in the study, further tests with more genotypes are need to confirm if these markers are truly informative in distinguishing the two *Liriodendron* species.

Table 7. Statistics of the 15 markers analyzed by Cervus.

Locus	**K**	**N**	**H_O**	**H_E**	**PIC**
LT002	6	174	0.718	0.69	0.646
LT015	6	172	0.547	0.598	0.537
LT021	3	173	0.173	0.185	0.173
LT086	7	170	0.359	0.532	0.499
LT096	15	165	0.648	0.759	0.733
isotig13819	12	174	0.598	0.709	0.69
isotig23696	18	172	0.738	0.868	0.851
isotig12995	13	167	0.629	0.84	0.819
isotig13485	18	164	0.878	0.89	0.877
isotig11620	14	164	0.738	0.795	0.762
isotig08911	11	172	0.837	0.858	0.839
isotig11603	15	172	0.512	0.737	0.718
isotig22220	11	158	0.513	0.757	0.725
isotig04692	19	143	0.65	0.925	0.916
isotig09599	26	160	0.738	0.931	0.923
Average	13	167	0.618	0.738	0.714

K: number of alleles; N: number of individuals; H_0: observed heterozygosity; He: expected heterozygosity; PIC: polymorphic information content.

3. Transferability of *L. tulipifera* EST-SSR markers

PCR amplification of 20 surveyed markers was successful in *L. Chinense*, although sizing in an ABI 3730 Genetic Analyzer failed for LT157 and isotig14887, due to stuttering observed in *L. tulipifera*. This indicates a high frequency of transferability of *L. tulipifera* EST-SSR markers in *L. Chinense*, supporting the previous findings of 72.4% success rate by Xu et al. (2011) and 82.1% by Yang et al. (2012). This is expected because EST-SSRs have generally demonstrated a high frequency of cross-species transferability despite less polymorphism compared to genomic SSRs (Yu et al. 2004; Elli and Burke et al. 2007; Han et al. 2009).

4. Comparison of genetic composition of two *L. tulipifera* orchards

All of the 15 markers were polymorphic in both orchards. The 15 markers were also tested on one *L. tulipifera* tree from the US National Arboretum and 10 of the loci were heterozygous (data not shown). While there was only one locus (LT015) significantly deviating from Hardy-Weinberg proportions in the Clemson population, 10 loci were observed in Knoxville population ($p>0.05$) (Appendix II and III). In terms of observed number of alleles, effective number of alleles, observed heterozygosity, and Nei's expected heterozygosity, the Clemson orchard exhibited higher values than the Knoxville orchard (Table 8 and 9). Kovach (2012) utilized amplified fragment length polymorphism

with five primers to determine the level of genetic diversity of *Liriodendron tulipifera* samples collected from six unmanaged populations in the Mountains and Coastal Plain of the Southeastern U.S. Observed overall genetic diversity was higher (He: 0.289) than within the Mountain populations (He: 0.281) or the Coastal Plain populations (He: 0.271). Thus, both Clemson and Knoxville orchards had a much higher level of expected heterozygosity (He), 0.71 and 0.70, respectively, than the unmanaged populations in the Mountains and Coastal Plain, and similar to a cultivated population of *L. tulipifera* in China (Yang et al. 2012).

Nei's genetic distance between the two orchards was 0.38, which was the lowest among all comparisons (Table 10). The *L. chinense* and *L. tulipifera* trees from the National Arboretum (0.97) exhibited the largest genetic distance (1.17). The two orchards and the *L. tulipifera* sample from the US National Arboretum grouped together in the UPGMA dendrogram. In accordance with the results of matK alignments, Tree#CU24 from the Clemson orchards did not group with *L. tulipifera*. The genetic distance of the Tree#CU24 was closest to the Clemson orchard (0.60), followed by the Knoxville orchard and *L. chinense* from the National Arboretum (0.97), and then by the *L. tulipifera* from the National Arboretum (1.02) (Figure 6).

In summary, 20 EST-SSR markers have been characterized with trees from two *Liriodendron* orchards and the US National Arboretum. Our study indicated high

frequency of transferability of *L. tulipifera* EST-SSR markers in *L. Chinense*. The multiallelic loci (LT002, isotig13485, and isotig11620) having a PIC ≥0.5, without presence of null alleles, and within Hardy-Weinberg equilibrium will be useful in the study of population dynamics and adaptive variation in *Liriodendron*. Genetic diversity of the Knoxville and Clemson orchards is higher than the unmanaged populations and similar to a cultivated population in China. The information obtained from this study provides a foundation for further genetic and breeding exploration with this economically important tree species.

Table 8. Genetic variation at six EST-SSR loci characterized in Clemson orchard.

Locus	Clemson orchard (142 trees)								
	Sample Size	Na	Ne	Obs_Hom	Obs_Het	Exp_Hom	Exp_Het	Nei's	I
LT1	284	6	3.1	0.27	0.73	0.32	0.68	0.68	1.33
LT5	284	6	2.29	0.46	0.54	0.43	0.57	0.56	1.07
LT9	282	3	1.26	0.82	0.18	0.79	0.21	0.21	0.41
LT37	276	6	1.8	0.7	0.3	0.55	0.45	0.44	0.95
LT41	268	14	3.71	0.36	0.64	0.27	0.73	0.73	1.74
isotig13819	284	12	3.33	0.38	0.62	0.3	0.7	0.7	1.71
isotig23696	282	17	7.71	0.26	0.74	0.13	0.87	0.87	2.26
isotig12995	272	12	5.05	0.29	0.71	0.19	0.81	0.8	1.9
isotig13485	266	17	8.62	0.08	0.92	0.11	0.89	0.88	2.4
isotig11620	266	12	4.55	0.25	0.75	0.22	0.78	0.78	1.7
isotig08911	282	11	6.99	0.17	0.83	0.14	0.86	0.86	2.9
isotig11603	280	13	2.8	0.45	0.55	0.36	0.64	0.64	1.57
isotig22220	258	9	3.06	0.52	0.48	0.32	0.68	0.67	1.4
isotig04692	220	15	11.32	0.34	0.66	0.08	0.92	0.91	2.53
isotig09599	260	22	11.37	0.27	0.73	0.08	0.92	0.91	2.66
Mean	271	11.7±5.1	5.13±3.31	0.37±0.20	0.63±0.20	0.29±0.19	0.71±0.19	0.71±0.19	1.71±0.62

Table 9. Genetic variation at six EST-SSR loci characterized in Knoxville orchard.

	Knoxville orchard (31 trees)								
Locus	Sample Size	Na	Ne	Obs_Hom	Obs_Het	Exp_Hom	Exp_het	Nei's	I
LT1	62	5	3.65	0.32	0.68	0.26	0.74	0.73	1.43
LT5	60	5	3.38	0.43	0.57	0.28	0.72	0.7	1.36
LT9	62	2	1.17	0.84	0.16	0.85	0.15	0.15	0.28
LT37	62	6	2.81	0.35	0.65	0.35	0.65	0.64	1.22
LT41	62	10	4.75	0.32	0.68	0.2	0.8	0.79	1.85
isotig13819	62	9	3.59	0.52	0.48	0.27	0.73	0.72	1.68
isotig23696	60	12	5.84	0.27	0.73	0.16	0.84	0.83	2
isotig12995	62	6	2.77	0.74	0.26	0.35	0.65	0.64	1.32
isotig13485	60	15	6.14	0.3	0.7	0.15	0.85	0.84	2.17
isotig11620	60	8	5.26	0.33	0.67	0.18	0.82	0.81	1.8
isotig08911	60	8	4.64	0.1	0.9	0.2	0.8	0.78	1.71
isotig11603	62	5	3.29	0.68	0.32	0.29	0.71	0.7	1.33
isotig22220	56	3	2.26	0.39	0.61	0.43	0.57	0.56	0.89
isotig04692	62	9	5.88	0.42	0.58	0.16	0.84	0.83	1.93
isotig09599	58	8	3.83	0.24	0.76	0.25	0.75	0.74	1.62
Mean	61	7.4±3.4	3.95±1.44	0.42±0.20	0.58±0.20	0.29±0.17	0.7±0.17	0.58±0.16	1.51±0.48

Na = Observed number of alleles. Ne = Effective number of alleles (Kimura and Crow 1964). Obs_Hom/Obs_Het: Observed homozygosity/heterozygosity. Ext_Het/Exp_Het: expected homozygosity/heterozygosity (Levene 1949). Nei's (1973) expected heterozygosity. I = Shannon's Information index (Lewontin 1972).

Table 10. Nei's (1978) unbiased identity (above diagonal) and distance (below diagonal).

pop ID	Clemson	Knoxville	NA *L. tulipifera*	Clemson Tree# CU24 (hybrid)	NA *L. chinense*
Clemson	—	0.6856	0.6238	0.5507	0.4035
Knoxville	0.3775	—	0.7662	0.4188	0.3495
NA *L. tulipifera*	0.4719	0.4648	—	0.3600	0.3097
Clemson Tree# 24 (hybrid)	0.5965	0.9714	1.0217	—	0.9714
NA *L. chinense*	0.9076	1.0513	1.1721	0.3785	—

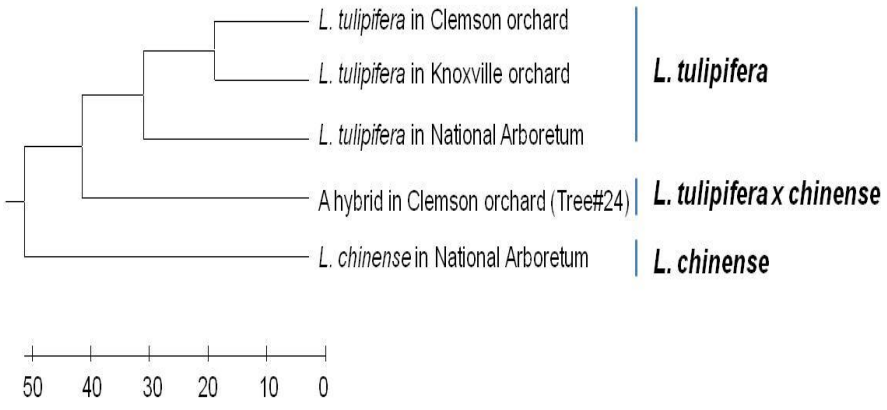

Figure 6. The UPGMA dendrogram based on Nei's (1978) genetic distance. Bootstrap replicates = 1,000.

References

Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson JE, de Pamphilis 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

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufa-yard J-F, Guindon S, Lefort V, Lescot M, Claverie J-M, Gascuel O (2008) Phylogeny.fr: robust phylogeneticanalysisforthenon-specialist,Nucleic Acids Res 36:W465–469.

Ellis JR and Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125–132.

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

Han Y, Chagné D, Gasic K, Rikkerink EH, Beever JE, Gardiner SE, Korban SS (2009) BAC-end sequence-based SNPs and Bin mapping for rapid integration of physical and genetic maps in apple. Genemics 93(3): 282-8

He S, Hao R (1999) Study on natural population dynamics and the endangering habitat of *Liriodendron chinense* in China. Acta Phytoecologica Sinica 23:87-95

Jin H, Do J, Moon D, Noh EW, Kim W, Kwon M (2011) EST analysis of functional genes associated with cell wall biosynthesis and modification in the secondary xylem of the yellow poplar (*Liriodendron tulipifera*) stem during early stage of tension wood formation. Planta 234:959-977

Kovach KE (2012) Assessment of genetic variation of *Acer rubrum* L. and *Liriodendron tulipifera* L. populations in unmanaged forests of the Southeast United States. Master Thesis, Virginia Polytechnic Institute and State University.

Kimura M, Crow JF (1964) The measurement of effective population number. Evolution 17:279-288.

Kobayashi N, Horikishi T, Katsuyama H, Handa T, Takayanagi K (1998) A simple and efficient DNA extraction method for plants, especially woody plants. Plant Tissue Culture Biotechnol 4:76-80

Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:381–398.

Liang H, Ayyampalayam S, Wickett N, Barakat A, Xu Y, Landherr L, Ralph P, Xu T, Schlarbaum SE, Leebens-Mack JH, dePamphilis CW (2011) Generation of a large-scale genomic resource for functional and comparative genomics in Liriodendron. Tree Gen Genom 7:941-954

Marshall TC, Slate J, Kruuk LEB and Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639-655.

Merkle SA, Hoey MT, Watson-Pauley BA, Schlarbaum SE (1993) Propagation of Liriodendron hybrids via somatic embryogenesis. Plant Cell Tissue Organ Cult 34:191–198

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.

Parks CR, Wendel JF, Sewell MM, Qiu YL (1994) The significance of allozyme variation and introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). Am J Bot 81:878-889

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Heredity, 86:248-249

Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD (2013) The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. BMC Biol 11:29

Sewell MM, Parks CR, Chase MW (1996) Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L. (Magnoliaceae). Evolution 50:1147–1154

Tomlinson PT, Jensen RJ, Hancock JF (2002) Do whole tree silvic characters indicate hybridization in red oak (*Quercus Section* Lobatae)? Am Midl Nat 143:154-168

Turmel M, Otis C, Lemieux C (2006) The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants. Mol Biol Evol 23: 1324-1338

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535– 538.

Wang Z (2005) Utilization and Species Hybridization in Liriodendron. Chinese Forestry Press, Beijing (in Chinese).

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

Wicke S and Quandt D (2009) Universal primers for the amplification of the plastid trnK/matK region in land plants. Anales Jard Bot Madrid 66: 285-288.

Xu M, Sun Y, Li H (2010) EST-SSRs development and paternity analysis for *Liriodendron* spp. New Forests 40:361-382

Yeh F, Rongcai Y, Boyle T (2000) POPGENE-1.32: A free program for the analysis of genetic variation among and within populations using co-dominant and dominant markers. Department of Renewable Resources at the University of Alberta, Canada. At http://www.ualberta.ca/~fyeh/index.htm.

Yang A, Zhang J, Tian H, Yao X (2012) Characterization of 39 novel EST-SSR markers for *liriodendron tulipifera* and cross-species amplification in *L. chinense* (magnoliaceae). Am J Bot 99:e460–e464

Yang A, Zhang J, Yao X, Huang H (2011) Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and crossspecies amplification in *L. chinense*. Am J Bot 98 e123 – e126 .

Yao, X, Zhang J, Ye Q, Huang H (2008) Characterization of 14 novel microsatellite loci in the endangered *Liriodendron chinense* (Magnoliaceae) and cross-species amplification in closely related taxa. Conservation Gen 9:483–485.

Yu JK, Rota ML, Kantety RV, Sorrells ME (2004) EST derived SSR markers for comparative mapping in wheat and rice. Mol Gen Genomics 271:742–277.

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.

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1-16

CONCLUSION AND FUTURE DIRECTION

Conclusion

Liriodendron is a suitable model for mating system, systemic evolution and population genetics studies, and has been deeply studied as a candidate for comparative studies and evolution of angiosperms. Besides, yellow-poplar has great economic and ecological values. We are interested in developing yellow-poplar as a new tree model research system for comparative genomics of secondary cell wall formation.However, the genome of yellow-poplar has not been sequenced, and less than 200 SSR markers have been characterized in *Liriodendron*. The specific objectives of this project were to develop informative SSR markers for construction of the first genetic linkage map for yellow-poplar; and to investigate the genetic composition of two yellow-poplar breeding orchards.

A total of 538 new EST-SSR markers had been available for yellow-poplar genome study, which were speculated to have relatively high PCR amplification success rate and percentage of polymorphic loci. Among the 538 new markers and 66 previously

characterized polymorphic markers, a total of 119 informative SSR markers were identified for genetic linkage map construction with an F1 progeny with #UT108A and #UT23 as parents. The full-sibship for 213 seedlings were validated. These informative SSR markers and full-sib seedlings are essential in construction of linkage maps, which are valuable for future molecular breeding and quantitative trait locus (QTL) mapping, and as a framework for sequencing the *Liriodendron* genome.

Twenty EST-SSR markers have been characterized with trees from two *Liriodendron* orchards and the US National Arboretum, and provided a first look at the genetic diversity and allele richness among selections of this unique native species. Our study indicated high frequency of transferability of *L. tulipifera* EST-SSR markers in *L. Chinense*. The multiallelic loci (LT002, isotig13485, and isotig11620) having a PIC ≥0.5, without presence of null alleles, and within Hardy-Weinberg equilibrium will be useful in the study of population dynamics and adaptive variation in *Liriodendron*. Genetic diversity of the Knoxville and Clemson orchards is higher than the unmanaged populations and similar to a cultivated population in China. The information obtained from this study provides a foundation for further genetic and breeding exploration with this economically important tree species.

Future direction

Yellow-poplar has a large genome with 19 chromosomes in one haploid genome. In order to construct a dense genetic linkage map, which is essential for future molecular breeding and quantitative trait locus (QTL) mapping and *Liriodendron* genome sequencing, more informative SSR markers and full-sib seedlings are needed.

APPENDICES

Appendix I. Genepop version 4.2: Hardy-Weinberg test with all *L. tulipifera* trees

```
Number of populations detected:    1
Number of loci detected:          15


Estimation of exact P-Values by the Markov chain method.
--------------------------------------------
Markov chain parameters for all tests:
Dememorization:            1000
Batches:                   100
Iterations per batch:       1000
Hardy Weinberg: Probability test
       ************************



=========================================
    Results by population
=========================================


 Pop : tulip
-----------------------------------------
                        Fis estimates
                       ---------------
locus      P-val   S.E.    W&C     R&H     Steps
---------- ------- ------- ------- ------- ------
```

```
LT002      0.0393  0.0045  -0.0468 -0.0087  30593 switches
LT015      0.1035  0.0111   0.0855  0.0336  20318 switches
LT021      0.0241  0.0017   0.1095  0.1877  27047 switches
LT086      0.0000  0.0000   0.3268  0.3563  14194 switches
LT096      0.0009  0.0009   0.1421  0.1001   5805 switches
Ltu19      0.0000  0.0000   0.1575  0.1673  14234 switches
Ltu51      0.0090  0.0057   0.1444  0.1119   5538 switches
Ltu53      0.0000  0.0000   0.2494  0.2219  12804 switches
Ltu125     0.0000  0.0000   0.0127  0.0331   8488 switches
Ltu143     0.0000  0.0000   0.0787  0.1519   4842 switches
Ltu145     0.2648  0.0219   0.0179  0.0172  23786 switches
Ltu150     0.0000  0.0000   0.3034  0.1833   7557 switches
Ltu151     0.0000  0.0000   0.3304  0.2863   8290 switches
Ltu152     0.0000  0.0000   0.2998  0.2791  13651 switches
Ltu154     0.0000  0.0000   0.2141  0.1248   4584 switches

All (Fisher's method):
 Chi2 :    Infinity
 Df   :   30.0000
 Prob :    High. sign.
```

Appendix II. Genepop version 4.2: Hardy-Weinberg test with all *L. tulipifera* trees in Clemson orchard.

- Number of populations detected: 1
- Number of loci detected: 15
-
-
- Estimation of exact P-Values by the Markov chain method.
- ---
- Markov chain parameters for all tests:
- Dememorization: 1000
- Batches: 100
- Iterations per batch: 1000
- Hardy Weinberg: Probability test
- ********
-
-
-
- ==========
- Results by population
- ==========
-
-
- Pop : CU165
- ---
- Fis estimates
- ---
- locus P-val S.E. W&C R&H Steps
- --- --- --- --- --- ---
- LT002 0.0413 0.0049 -0.0781 -0.0370 26992 switches
- LT015 0.2254 0.0143 0.0415 0.0054 14803 switches
- LT021 0.0181 0.0019 0.1391 0.1999 27410 switches
- LT086 0.0000 0.0000 0.3337 0.4793 17503 switches
- LT096 0.0000 0.0000 0.1249 0.0746 4918 switches
- Ltu19 0.0361 0.0134 0.1177 0.1148 10265 switches
- Ltu51 0.0077 0.0032 0.1479 0.1283 6113 switches
- Ltu53 0.0000 0.0000 0.1145 0.1633 10646 switches

- Ltu125 0.0000 0.0000 -0.0423 -0.0098 8250 switches
- Ltu143 0.0027 0.0011 0.0401 0.0227 5525 switches
- Ltu145 0.0303 0.0064 0.0350 0.0263 22390 switches
- Ltu150 0.0000 0.0000 0.1475 0.1297 5328 switches
- Ltu151 0.0000 0.0000 0.2894 0.3310 7471 switches
- Ltu152 0.0000 0.0000 0.2763 0.2625 16550 switches
- Ltu154 0.0000 0.0000 0.2024 0.1378 5139 switches
-
- All (Fisher's method):
- Chi2 : Infinity
- Df : 30.0000
- Prob : High. sign.

Appendix III. Genepop version 4.2: Hardy-Weinberg test with all *L. tulipifera* trees in Knoxville orchard.

```
Number of populations detected:    1
Number of loci detected:           15


Estimation of exact P-Values by the Markov chain method.
-----------------------------------------
Markov chain parameters for all tests:
Dememorization:              1000
Batches:                     100
Iterations per batch:        1000
Hardy Weinberg: Probability test
        ************************



=========================================
    Results by population
=========================================


 Pop : tulip
-----------------------------------------
                        Fis estimates
                       ---------------
locus      P-val   S.E.    W&C    R&H    Steps
---------- ------- ------- ------- ------- ------
LT002      0.0393  0.0045  -0.0468 -0.0087  30593 switches
LT015      0.1035  0.0111   0.0855  0.0336  20318 switches
LT021      0.0241  0.0017   0.1095  0.1877  27047 switches
LT086      0.0000  0.0000   0.3268  0.3563  14194 switches
LT096      0.0009  0.0009   0.1421  0.1001   5805 switches
Ltu19      0.0000  0.0000   0.1575  0.1673  14234 switches
Ltu51      0.0090  0.0057   0.1444  0.1119   5538 switches
Ltu53      0.0000  0.0000   0.2494  0.2219  12804 switches
```

```
Ltu125     0.0000  0.0000   0.0127  0.0331   8488 switches
Ltu143     0.0000  0.0000   0.0787  0.1519   4842 switches
Ltu145     0.2648  0.0219   0.0179  0.0172  23786 switches
Ltu150     0.0000  0.0000   0.3034  0.1833   7557 switches
Ltu151     0.0000  0.0000   0.3304  0.2863   8290 switches
Ltu152     0.0000  0.0000   0.2998  0.2791  13651 switches
Ltu154     0.0000  0.0000   0.2141  0.1248   4584 switches
```