Plant Omics Journal POJ

POJ 6(4):311-317 (2013) ISSN:1836-3644

Analysis of expressed sequence tags and alternative splicing genes in sacred lotus (*Nelumbo nucifera* Gaertn.)

Robert VanBuren[1], Braden Walters[2], Ray Ming[1], Xiang Jia Min[2*]

[1]Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[2]Center for Applied Chemical Biology, Department of Biological Sciences, Youngstown State University, Youngstown, OH 44555, USA

***Corresponding author: xmin@ysu.edu**

Abstract

Sacred lotus, *Nelumbo nucifera* (Gaertn.), a basal eudicot of agricultural, medicinal, cultural, and religious importance, is an under-studied species at the molecular level. Expressed sequence tags (ESTs) are particularly useful for gene identification, genome annotation and alternative splicing (AS) study. We generated 626,137 reads by 454 pyrosequencing of a cDNA library, which was constructed from mRNA collected from 7 different tissues of sacred lotus. The reads were assembled into 16,349 non-redundant EST contigs with an average length of 715 bp. Aligning ESTs to sacred lotus genome revealed the internal exons and introns had an average size of 140 and 2,005 bp, respectively. The average intron length in the sacred lotus is much larger than most sequenced plant genomes. We also identified 174 AS events involving 161 genes. Classification of the AS events showed that 62.6% of the AS events were retained introns, 8.0% were alternative acceptor sites, 7.5% were alternative donor sites, 4.0% were skipped exons, and 17.8% had "complex events" which had more than one basic event. Comparison of AS genes in sacred lotus, rice and Arabidopsis identified 7 commonly conserved genes between lotus and rice and 9 between sacred lotus and Arabidopsis. A gene encoding a 40S ribosomal protein was commonly conserved in all three species. These AS events and annotated ESTs are available through the Plant Alternative Splicing Database website (http://proteomics.ysu.edu/altsplice/plant/).

Keywords: alternative splicing; expressed sequence tags; exon, intron, sacred lotus.
Abbreviations: AS_alternative splicing; BLAST_basic local alignment search tool; CDS_coding DNA sequence; EST_expressed sequence tag.

Introduction

Expressed sequence tags (ESTs) are widely used for identification of genes and alternatively spliced transcripts. Alternative splicing (AS) is a process of generating more than one mRNA transcript or isoform from a single protein coding gene. Relative to the predominant transcript, four basic AS types have been observed, including exon skipping, alternative donor or acceptor site, and intron retention (Wang and Brendel, 2006). Some other basic types may also be counted as AS events such as alternative transcription initiation, alternative transcription termination, and mutually exclusive exons (Roberts and Smith, 2002; Sablok et al., 2011). However, various combinations of the above basic AS types can produce a transcript having more than one AS event. AS may generate functional transcripts that encode distinct functional proteins, or it may generate nonfunctional transcripts with premature stop codons. Nonfunctional isoforms can be degraded by the nonsense mediated mRNA decay (NMD) surveillance machinery. mRNA transcript abundance may be regulated using AS and NMD through a process known as regulated unproductive splicing and translation (RUST) (Lareau et al., 2007). Although AS appears to occur less frequently in plants than in animal systems, it is clearly significant, with 20-30% of genes in *Arabidopsis thaliana* and *Oryza sativa* (rice) showing AS (Wang and Brendel, 2006). Moreover, recently Filichkin et al. (2010) reported at least 42% of intron-containing genes in Arabidopsis are alternatively spliced using a deep transcriptome sequencing technique; this is significantly higher than previous estimates based on cDNA/expressed sequence tag sequencing. AS can be mediated mainly by three pathways: intron retention in the mature mRNA, accounting for 30–50% of AS in Arabidopsis and rice; alternative exon usage, resulting in exon skipping; and the use of cryptic splice sites that may elongate or shorten an exon (Wang and Brendel, 2006). Approximately 60 - 75% of AS events occur within the translated regions of mRNAs, and this can have dramatic effects on binding properties, intracellular localization, protein stability, enzymatic, and signaling activities (Stamm et al., 2005). ESTs represent the transcripts derived from genes expressed in different tissues and developmental processes. EST based analyses would be a suitable approach to discover and delineate these alternative-splicing variants through homology mapping (Wang and Brendel, 2006). In plants, the functional relevance of certain AS-derived isoforms has been observed in responses to biotic and abiotic stresses (Reddy, 2007; Mastrangelo et al., 2012). The roles of AS in the expression of various genes involved in the processes of photosynthesis, plant disease resistance, flowering, and grain quality in rice have been well illustrated (Reddy, 2007). The progress in plant AS research was

Table 1. Summary of sacred lotus EST assembly statistics.

Parameter	Value
Total number of reads	626,137
Mean read length (bp)	379
Inferred error rate (%)	0.87
Number of reads assembled	531649
Percentage of reads assembled (%)	84.9
Number of contigs	16394
Average contig length (bp)	715
Unassembled singletons	94488

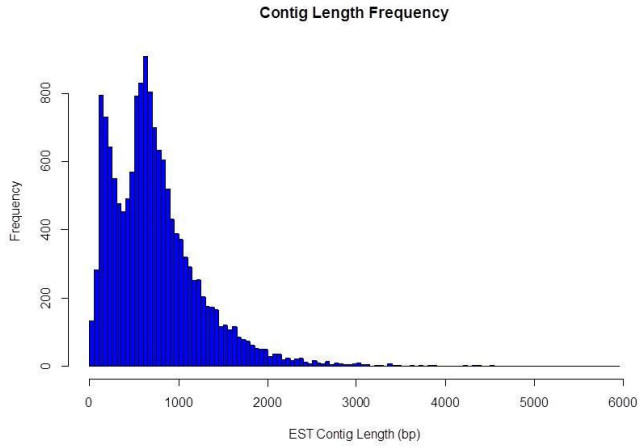

Fig 1. Overview of assembled EST contig length distribution. EST contig length is plotted by frequency, with a mean length of 715 bp.

recently comprehensively reviewed (Reddy et al., 2012). Sacred lotus is a basal eudicot, in the Proteales family with agricultural, medicinal, cultural, and religious importance. The genome of sacred lotus variety 'China Antique' was recently sequenced (Ming et al., 2013). It lacks the lambda triploidization event seen in all core eudicots, making sacred lotus an excellent model for reconstructing the pan-eudicot genome (Ming et al., 2013). It is also expected to serve as a reference for comparative genomics due to its remarkably low mutation rate and a higher retention of duplicated genes compared with most other angiosperm genomes (Ming et al., 2013). EST annotation and mapping ESTs to the genome will help to refine gene models predicted from genome sequence. The identification of AS genes and AS events is the first step for our understanding of how these genes are regulated post-transcriptionally, and may provide insights into the exceptional seed longevity, thermogenesis, water repellency, and other biological phenomenon seen in sacred lotus. In this report we present a set of 16,943 ESTs annotated and mapped to the genomic sequences and from them 161 AS genes identified. The availability of the data through the Plant Alternative Splicing Database website (http://proteomics.ysu.edu/altsplice/plant/) provides the community a starting point to perform additional detailed experiments for identifying the functional relevance of protein coding genes in this species.

Results and discussion

EST annotation and analysis

To capture a representative transcriptome of sacred lotus, RNA was collected from floating and aerial leaves, leaf petioles, rhizomes, roots, mixed flowers and germinating seeds. Equimolar concentrations of RNA from each tissue type were used for 454 library preparation. After trimming and removal of low quality reads, a total of 16,394 non-redundant EST contigs were assembled from 626,137 reads generated by 454 pyrosequencing (Table 1). The average trimmed read length was 379 bp, and the inferred error rate after assembly was low, at 0.87%. Approximately 85% of the reads were assembled into contigs, with the remaining 94,488 classified as singletons. The average EST length is 715 bp with sizes ranging from 100-5,500 bp (Fig. 1). Using OrfPredictor, a webserver for predicting open reading frames (ORFs) of ESTs (Min et al., 2005), a total of 16,350 ESTs (99.7%) were predicted to contain an ORF region for peptide translations, and among them 10,031 (61.2%) have an ortholog in the UniProtKB/Swiss-Prot database and 8,138 (49.6%) could be assigned to a Pfam domain with a cutoff E-value of 1e-10. A BLASTN search against the predicted coding DNA sequences (CDS) from the lotus draft genome identified 12,298 ESTs with annotated CDS having greater than 95% identity. We estimated the category distributions of ESTs using Gene Ontology (GO) classification. There are three main categories of GO terms including biological processes, cellular components, and molecular functions (http://www.geneontology.org/). Among 10,031 ESTs having a BLASTX hit against UniProtKB/Swiss-Prot database, 6,013 of them have one or more GO annotation. A total of 30,038 GO identifiers were extracted and further grouped using GO SlimViewer (McCarthy et al., 2006). The results are shown in Fig. 2 and Supplementary Fig. 1. Molecular function classification showed ~46% of predicted proteins coded by the sampled genes had enzyme or catalytic activity, including 12.9% hydrolase activity, 12.6% transferase activity, 5.4% kinase activity. About 40% of the proteins had molecular binding activity, including 15.8% nucleotide binding, 10.9% protein binding, 5% DNA binding and 4.2% RNA binding.

Table 2. Conserved alternative splicing genes in sacred lotus, rice (subspecies japonica) and Arabidopsis.

Lotus	Rice (gi)	Arabidopsis (gi)	Function
contig00537		42472437	CASD1 CAS1 domain-containing protein 1
contig00634	428757		Aspartic proteinase
contig00962		937199	RH15 DEAD-box ATP-dependent RNA helicase 15
contig00976	287081, 427172, 286977		TUBA Tubulin alpha chain
contig02039		19799440	Hypothetical protein
contig02899		16590	Pyrophosphate-energized vacuolar membrane proton pump
contig03074		5843939	PPD1 PsbP domain-containing protein 1, chloroplastic
contig03150	428212		pepN Aminopeptidase N
contig03511, contig08690	425900	1053681	RPS3C 40S ribosomal protein S3-3
contig03973	703053, 426964		RIC1 Ras-related protein RIC1
contig04075, contig12238		29468597	PISD Phosphatidylserine decarboxylase proenzyme
contig04258		17069	CAB8 Chlorophyll a-b binding protein 8, chloroplastic
contig04973	286284		Triosephosphate isomerase, cytosolic
contig05783, contig11093		19741797	lsm1 U6 snRNA-associated Sm-like protein LSm1
contig06207, contig11408	287248		Ras-related protein Rab7

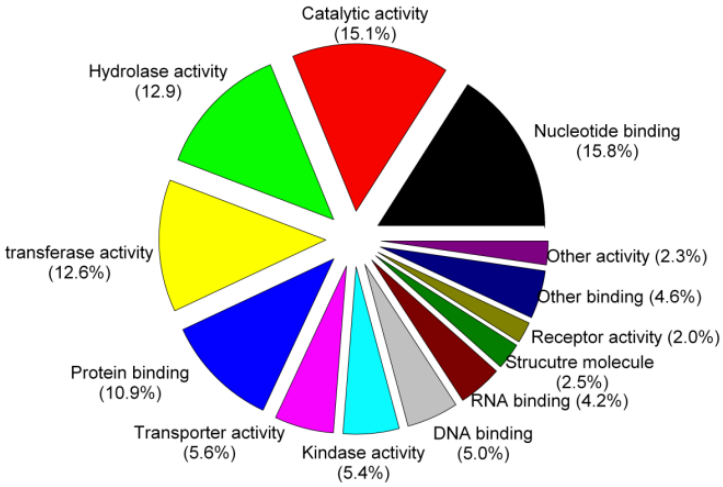

Fig 2. Distribution of molecular functions of ESTs based on Gene Ontology (GO) classification. The distribution of cellular components and biological processes of GO classification of ESTs was shown in Supplementary Fig 1.

Cellular component analyses showed about 24% membrane proteins, 19.4% intracellular, 17.4% cytoplasm, and variable distributions in other subcellular components. These proteins are involved in various biological processes including metabolic, biosynthesis, transport, responses to stress, etc. (Supplementary Fig. 1).

Features of exons and introns in protein coding genes

After mapping the ESTs to the corresponding draft genome, we obtained 21,220 internal exons and 29,315 introns. Internal exons range in size from 19 – 4,563 bp with an average size of 140 bp and a standard deviation of 124 bp. Approximately 94% of exons were less than 300 bp in length and 65% were between 40 to 140 bp (Fig. 3). The observed exon length is in line with mean internal exon length observed in Arabidopsis (172 bp), rice (subspecies japonica) (193 bp), and *Brachypoidum distachyon* (130 bp) (Wang and Brendel, 2006; Walters et al., 2013). The distribution of intron sizes was more variable than exon sizes with a range of 11 bp to >100 kb. Twenty three of the introns (0.08%) are unusually large with a size >50 kb. These exceptionally large introns may be indicative of possible errors in EST assembly and/or genome assembly, and warrant further examination. Excluding those long introns with >50 kb in size, the remaining introns had an average size of 2,005 bp with a standard deviation of 4,419 bp. Furthermore, 68% of the introns are shorter than 1 kb and 47% are in the range of 60 to 140 bp (Fig. 4). The observed intron length in sacred lotus was much longer than the average intron size observed in Arabidopsis (173 bp), rice (subspecies japonica) (433 bp), and *B. distachyon* (420 bp), calculated using a similar method (Wang and Brendel, 2006; Walters et al., 2013).

Detection and classification of alternative splicing events

Aligning the ESTs to the draft genome of sacred lotus using 97% minimum identity and >85% alignment of the EST to the genome, resulted in a total of 12,344 anchored ESTs aligned to 11,801 genomic loci. We further identified a total of 174 alternative splicing events from 161 genes involving 329 ESTs (Fig.5, Supplementary Table 1). Alternatively spliced transcripts represent 1.5% of the total ESTs in sacred lotus, significantly less than reported in Arabidopsis or rice, which was ~ 20% (Wang and Brendel, 2006). The low frequency of AS events may be due to the small number of ESTs used in this study as shown in some other plant species with relatively limited number of ESTs including *B. distachyon* (~6.3%) (Walters et al., 2013), *Medicago truncatula* (~10%) and *Lotus japonicus* (~3%) (Wang et al., 2008). However, intron retention was the predominant form of AS in sacred lotus, representing 62.6% of total AS events,

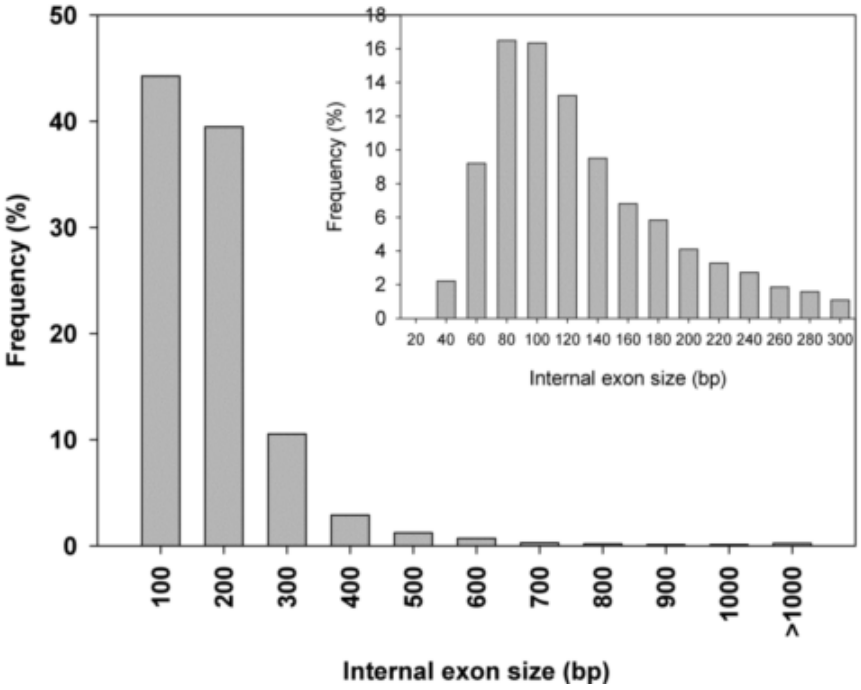

Fig 3. Distribution of internal exon sizes. Bin sizes in x axis are right inclusive (e.g., bin 100 comprises sequences of lengths 1 –100 nucleotides). Insets show a detailed distribution of smaller internal exons.

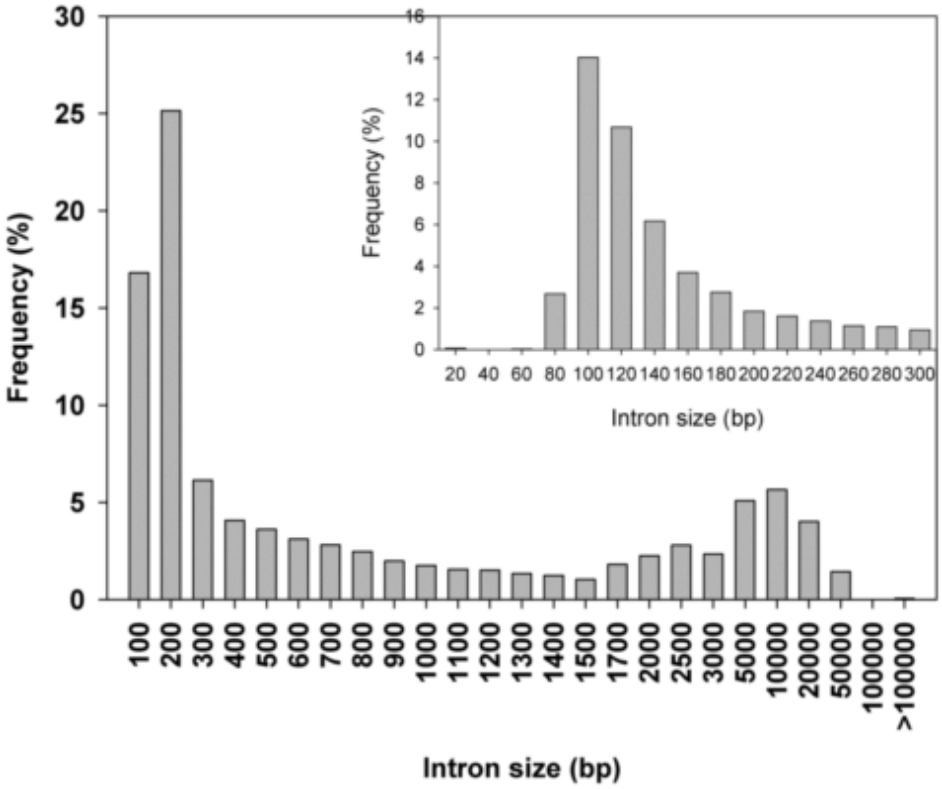

Fig 4. Distribution of intron sizes. Bin sizes in x axis are right inclusive (e.g., bin 100 comprises sequences of lengths 1 –100 nucleotides). Insets show a detailed distribution of smaller introns.

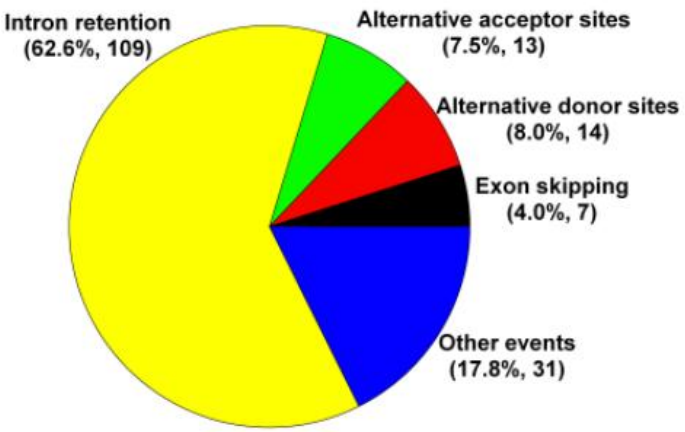

Fig 5. Frequency distribution of different types of alternative splicing events in sacred lotus.

in consistency with those reported previously for Medicago (39%), Populus (34%), Arabidopsis (56%), rice (54%) and *B. distachyon* (53 – 55.5%) (Wang and Brendel, 2006; Baek et al., 2008; Sablok et al., 2011; Walters et al., 2013). Alternative donor sites and alternative accepter sites each accounted for ~7- 8%, and exon skipping was the least prevalent event type, consistent with AS event distributions reported in other plant species (Wang and Brendel, 2006; Walters et al., 2013). The products of these AS genes had diverse functionalities including enzymes, transporters, and ribosomal proteins (Supplementary Table 1). The prevalence of intron retention in lotus and in other sequenced plants supports the intron definition model. That is, introns are identified by the splicing machinery splicesomes during pre-mRNA processing in plants. This contrasts the exon definition model seen in animals where exon skipping events occur much more frequently than intron retention events (McGuire et al., 2008). We also searched the unmapped ESTs that failed to align to the draft genome for AS events. ESTs were searched against themselves using the BLASTN method of ASFinder with 97% identity, a minimum aligned length of 60 bp and minimum unaligned gap length of 6 bp between two aligned segments (Min, 2013). In total, 21 additional AS genes were identified from 4,005 unmapped ESTs. However, AS events in these genes can not be identified due to the lack of genomic information.

Conserved alternatively spliced genes

Homologous AS genes conserved among multiple evolutionarily distant plant species are most likely functionally significant. It was reported that 41.7% of AS genes in Arabidopsis were found to have close homologs in rice that were also identified to be alternatively spliced (Wang and Brendel, 2006). Conserved AS genes and events were also reported between *M. truncatula* and *L. japonicus* (Wang et al., 2008), and among Arabidopsis, rice and *B. distachyon* (Walters et al., 2013). Using the AS data set identified in rice (subspecies japonica) and Arabidopsis (Wang and Brendel, 2006), we identified 7 homologous pairs of AS genes between sacred lotus and rice, and 9 homologous pairs of AS genes between sacred lotus and Arabidopsis (Table 2). Among them, one AS gene encoding a 40S ribosomal protein (UniProtKB accession number: Q9FJA6 in *A. thaliana*) was commonly conserved in all three species. These commonly conserved AS genes certainly are potential marker genes for examining the biological significance of AS in plant growth and development during evolution.

Materials and methods

Plant materials

Greenhouse grown sacred lotus cultivar 'China Antique' tissue was used for 454-pyrosequencing. Samples of floating and aerial leaves, leaf petioles, rizomes, roots, and mixed flowers (ranging from <10mm buds to full bloom) were collected from mature plants, frozen in liquid nitrogen and stored at -80 ^{o}C until processing. Newly germinated sacred lotus seedlings were grown in the greenhouse for 10 days post germination and harvested as described above.

RNA isolation and 454 cDNA library construction and sequencing

Total RNA was isolated using methods previously described in Yu et al. (2005). RNA quality was verified on a 2% agarose gel and quantified using a nanodrop spectrophotometer. RNA from the 7 tissue types were pooled in equimolar concentrations prior to cDNA synthesis. cDNA library construction and 454 pyrosequencing were carried out at the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign. Briefly, messenger RNA was enriched from 10 µg total pooled RNA using the Oligotex kit (Qiagen, Valencia, CA). The enriched mRNA was used to construct and normalize the 454 cDNA library following the procedure described in Lambert et al. (2010). The prepared library was quantified using a Quibit fluorometer (Invitrogen, CA) and average fragment size was measured with a Bioanalyzer (Agilent, CA). Emulsion-based amplification and sequencing was performed on the 454 Genome Sequencer FLX+ system according to manufacturer instructions (454 Life Sciences, Branford, CT). Raw sequences generated from a full 454 titanium plate were analyzed with the GS FLX software

v2.0.01.12 package (454 Life Sciences, Roche). Weak signals and low quality sequences were removed from the project and 454 adapters were trimmed, leaving 626,137 single end reads for assembly. Reads were assembled using GS De Novo Assembler from the GS FLX software package. Default parameters were used with the exception of a stringent overlap of 40bp and minimum identity of 95%. Large contigs were surveyed using BLASTX to check for chimeric gene assemblies.

EST annotation and analysis

The assembled EST contigs, hereafter referred as ESTs, were annotated for the following features: (1) open reading frame (ORF) and peptide prediction using OrfPredictor (http://proteomics.ysu.edu/tools/OrfPredictor.html) based on BLASTX search against UniProtKB/Swiss-Prot data or the intrinsic signals of the ESTs for sequences without a BLASTX hit (Min et al., 2005); (2) BLATN against predicted coding DNA sequences (CDS) of sacred lotus (Ming et al., 2013); (3) functional annotation including BLASTX search against UniProtKB/Swiss-Prot database and domain search against Pfam; (4) EST mapping to genomic scaffolds (see below). All BLAST searches were performed locally using a stand-alone BLAST package with a cut off E-value of 1e-10. To estimate different categories of genes sampled in the EST data set, we extracted the Gene Ontology (GO) annotation from the identifier mapping table which was downloaded from UniProt (http://www.uniprot.org/downloads) for ESTs having a BLASTX hit against UniProt-SP dataset. The retrieved GO IDs were then mapped to top categories using GO SlimViewer (McCarthy et al., 2006).

Homology mapping and identification of AS isoforms

ASFinder was used to map the lotus ESTs to the corresponding genomic megascaffold sequences and identify AS isoforms (http://proteomics.ysu.edu/tools/ASFinder.html/) (Min, 2013). ASFinder uses the SIM4 program to map ESTs to the genome (Florea et al., 1998). After mapping, ESTs that are mapped to the same genomic location but have variable exon-intron boundaries are flagged as AS isoforms. For genome mapping we used the following thresholds: a minimum of 97% identity of aligned ESTs with genomic sequences, a minimum of 80 bp of aligned length, and >85% of EST sequence aligned to the genome. The output of ASFinder was subsequently analyzed and AS events were identified using AStalavista server (http://genome.crg.es/astalavista/) (Foissac et al., 2007). Additional AS isoforms from ESTs that were not mapped the genome were identified using ASFinder BLASTN method to detect EST pairs having segmented alignments. A similar method using BLAT was used to identify AS from ESTs in plants previously (Ner-Gaon et al., 2007).

Conserved alternatively spliced genes in lotus, rice and Arabidopsis

To identify conserved alternatively spliced genes between sacred lotus and rice, and between sacred lotus and Arabidopsis, we downloaded mRNA sequences of AS genes in rice and Arabidopsis (http://www.plantgdb.org/ASIP/Download/) (Wang and Brendel, 2006). Rice and Arabidopsis mRNAs with AS were used as a query against sacred lotus AS ESTs using BLASTN with E-value of 1e-5. Reciprocal BLASTN hits with an identity of > 70% and a minimum aligned length of >100 bp were identified as conserved AS pairs.

Data access and visualization of AS

We developed a website for searching annotation information of all ESTs and AS events (http://proteomics.ysu.edu/altsplice/plant/), using EST ID, keywords, or AS event types. The AS events were displayed using GBrowse (http://proteomics.ysu.edu/cgi-bin/gb2/gbrowse/). All ESTs and AS data can also be downloaded and searched using BLASTN through links at the above website.

Conclusions

We generated and annotated 16,364 non-redundant EST contigs from sacred lotus. Using 97% minimum identity and >85% alignment of the EST to the genome, a total of 12,344 ESTs were mapped to 11,801 genomic loci. We identified 174 AS events involving 161 genes. Comparison of AS genes in sacred lotus, rice and Arabidopsis identified 7 commonly conserved genes between sacred lotus and rice and 9 between sacred lotus and Arabidopsis. The annotated ESTs and AS events are available through the Plant Alternative Splicing Database website.

Acknowledgements

The work was supported by the University of Illinois at Urbana-Champaign to RM and Youngstown State University (YSU) Research Council and the Ohio Plant Biotechnology Consortium to XJM.

References

Baek JM, Han P, Iandolino A, Cook DR (2008) Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, Arabidopsis and rice. Plant Mol Biol. 674:99–510

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res. 20:45-58

Florea L, Hartzell G, Zhang Z, Rubin GM, Miller W (1998) A computer program for aligning a cDNA sequence with a genomic DNA sequence. Genome Res. 8:967-974

Foissac S, Sammeth M (2007) ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. Nucleic Acids Res 35:W297–99.

Lambert JD, Chan XY, Spiecker B, Sweet HC (2010) Characterizing the embryonic transcriptome of the snail ilyanassa. Integr Comp Biol. 50:768-777

Lareau LF, Brooks AN, Soergel DAW, Meng Q, Brenner SE (2007) The coupling of alternative splicing and nonsense mediated mRNA decay. In: Blencowe BJ and Graveley BR (ed) Alternative splicing in the postgenomic era. Landes Biosciences, Austin, TX, p 191-212

Mastrangelo AM, Marone D, Laidò G, De Leonardis AM, De Vita P (2012) Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. Plant Sci. 185-186: 40-49.

McCarthy FM, Wang N, Magee GB, Williams WP, Luthe DS, Burgess SC (2006) AgBase: a functional genomics resource for agriculture. BMC Genomics. 7:229

McGuire AM, Pearson MD, Neafsey DE, Galagan JE (2005) Cross-kingdom patterns of alternative splicing and splice recognition. Genome Biol. 9:R50

Min XJ (2013) ASFinder: a tool for genome-wide identification of alternatively spliced transcripts from EST-derived sequences. Int J Bioinform Res Appl. 9:221-226

Min, XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res 33:W677-W680

Ming R, Vanburen R, Liu Y, Yang M, Han Y, Li LT, Zhang Q, Kim MJ, Schatz MC,Campbell M, Li J, Bowers JE, Tang H, Lyons E, Ferguson AA, Narzisi G, Nelson DR, Blaby-Haas CE, Gschwend AR, Jiao Y, Der JP, Zeng F, Han J, Min XJ, Hudson KA,Singh R, Grennan AK, Karpowicz SJ, Watling JR, Ito K, Robinson SA, Hudson ME, Yu Q, Mockler TC, Carroll A, Zheng Y, Sunkar R, Jia R, Chen N, Arro J, Wai CM, Wafula E, Spence A, Han Y, Xu L, Zhang J, Peery R, Haus MJ, Xiong W, Walsh JA, Wu J, Wang ML, Zhu YJ, Paull RE, Britt AB, Du C, Downie SR, Schuler MA, Michael TP, Long SP, Ort DR, Schopf JW, Gang DR, Jiang N, Yandell M, Depamphilis CW, Merchant SS, Paterson AH, Buchanan BB, Li S, Shen-Miller J. (2013) Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). Genome Biol. 14: R41

Ner-Gaon H, Leviatan N, Rubin E, Fluhr R (2007) Comparative cross-species alternative splicing in plants. Plant Physiol. 144:1632-1641

Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol. 58:267-294

Reddy ASN, Rogers MF, Richardson DN, Hamilton M and Ben-Hur A (2012) Deciphering the plant splicing code: experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. Front Plant Sci. 3:18.

Roberts GC, Smith CW (2002) Alternative splicing: combinatorial output from the genome. Curr Opin Chem Biol. 6:375-383

Sablok G, Gupta PK, Baek JM, Vazquez F, Min XJ (2011) Genome-wide survey of alternative splicing in the grass *Brachypodium distachyon*: an emerging model biosystem for plant functional genomics. Biotechnol Lett. 33:629-636

Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H (2005) Function of alternative splicing. Gene 344:1-20

Walters B, Lum G, Sablok G, Min XJ (2013) Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. DNA Res. 20:163-171

Wang B, Brendel V (2006) Genome wide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci USA 103:7175-7180

Wang BB, O'Toole M, Brendel V, Young ND. (2008) Cross-species EST alignments reveal novel and conserved alternative splicing events in legumes. BMC Plant Biol. 8:17

Yu Q, Moore PH, Albert HH, Roader AH, Ming R (2005) Cloning and characterization of a FLORICAULA/LEAFY ortholog, PFL, in polygamous papaya. Cell Res. 15:576-584