academicJournals

Vol. 14(23), pp. 1928-1935, 10 June, 2015
DOI: 10.5897/AJB2015.14619
Article Number: CF272B553541
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Construction of a full-length cDNA library and analysis of expressed sequence tags in white jute (*Corchorus capsularis* L.)

Aifen Tao[1], Xiaozhen Li[2], Jianmin Qi[1]*, Pingping Fang[1], Lihui Lin[1], Jiantang Xu[1], Liwu Zhang[1], Jianmei Wu[1] and Peiqing Lin[1]

[1]Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, Fujian Provincial Key Laboratory of Crops by Design(Fujian Agriculture and Forestry University), Fuzhou 350002, P.R. China.
[2]Datong Township Government of Changting County, Longyan 366300, China.

Received 3 April, 2015; Accepted 2 June, 2015

White jute (*Corchorus capsularis* L.) is recognized as an important industrial raw material fibre crop owing to its elite characters. However, little information is known about its molecular basis and genomics. In this study, a complementary DNA library of white jute was constructed and expressed sequence tags (ESTs) were characterized. The titers of original and amplified libraries were 2.32×10^7 and 1.07×10^9 pfu/mL, respectively. The recombinant frequency was 98.3% in the library. Most of the sequences ranged from 500 to 1500 bp with an average length of 750 bp. Results show 203 (73%) ESTs exhibited significant similarity with known or putative functional nucleotide sequences in the GenBank databases. Cluster analysis allowed the identification of 61 unique sequences. These genes were classified into six types by Gene Ontology (GO) annotation. The results also indicated that unigenes of *C. capsularis* have higher homology to *Populus trichocarpa*, *Ricinus communis* and *Corchorus olitorius*. This report will provide a valuable resource for the further investigations in the gene cloning, transcription or expression for white jute.

Key words: White jute, *Corchorus capsularis*, cDNA library, construction, ESTs, analysis.

INTRODUCTION

Jute (*Corchorus capsularis* L.) is an important bast fibre crop extensively grown in Southeast Asian countries. Jute fibres exhibit a characteristically high luster, good moisture absorption performance, rapid water loss capacity and easy degradation (Zhang et al., 2013). The textile and paper industry are interested in its potential as

*Corresponding author. E-mail: Qijm863@163.com.

Abbreviations: ESTs, Expressed sequence tags; **GO,** gene ontology; **cDNA,** complementary DNA; **SMART,** switching mechanism at 5' end of the RNA transcript; mRNA, messenger RNA; **SSR,** simple sequence repeat; **BGI,** Beijing genomics institution; **NCBI,** national center for biotechnology; **PVPP,** polyvinylpyrrolidone.

an important ingredient for producing paper, fine textiles as well as a renewable source for biofuel (Wazni et al., 2007). In recent years, with increasing uses of jute in diversified industries, there is a growing demand for high-yielding new cultivars. This has posed a serious challenge because no new breeding approaches have been developed for jute over the past seven decades (Kundu et al., 2015). Instead of using traditional breeding methods, jute breeders and biologists are now turning their attention to use molecular tools to improve agronomic traits (Khatun, 2007). However, many potentially fruitful research avenues, especially large-scale gene expression surveys and development of molecular genetic markers have been limited by a lack of sequence information in public databases (Zhang et al., 2011a). For a non-model species such as jute, with no prior genomic knowledge and resources (only 2,036 nucleotide and 54 EST sequences in NCBI database) (Kundu et al., 2015), it is critical to isolate, identify and understand the expressed sequences, with a long term goal of genetically modifying jute.

Generation and characterization of gene libraries is the priority for plant breeders because it is a fundamental platform in diverse aspects as genetic and physical mapping, molecular marker, new genes isolation and identification, and comparative genomics research (Talon and Gmitter, 2008; Zhang et al., 2011b; Zheng et al., 2011; Zhang et al, 2014). Expressed sequence tag (EST) projects provide a very useful and quick means of accessing gene sequence and expression information (Manickavelu et al., 2012). Some reports have proven that projects based on ESTs are powerful tools for both the analysis of gene expression patterns in a given tissue or at specific developmental stages (Chen et al., 2012; Tran et al., 2011; Wang et al., 2011b; Yang et al., 2009b; Zhang et al., 2011a). Moreover, ESTs from complementary DNA (cDNA) clones are inexpensive and efficient gene discovery tools (Wang et al., 2011a; Xiao et al., 2011; Yamagishi et al., 2011). As a molecular basis of information on whole genomes, the accumulation of ESTs is a promising strategy for studies in plant molecular biology (Rudd, 2003). These technologies are particularly important for plants lacking genomic sequence information such as *C. capsularis*.

Significant progress has been made in the last decade to understand the genome sequences of *Corchorus* species by ESTs analysis. So far, partial cDNA sequences of putative phosphate transport ATP-binding protein gene of C. *capsularis* var. CVL-1 were submitted (Islam et al., 2005). Amherst published partial cDNA base sequences of NADH dehydrogenase (ndhF) gene of C. *capsularis* (Whitlock et al., 2003). Determined complete cDNA sequence of caffeoyl-CoA-O-methyltransferase and cinnamyl alcohol dehydrogenase which are two of the three genes involved in lignin biosynthesis of *capsularis*. Basu et al., 2003a; Basu et al., 2003b Samanta et al. (2015) reported that WRKY transcript factor was a most important transcript in fibre development process of jute by EST analysis. In fact, there are some other major challenges for jute except poor fibre quality, which include susceptibility of the jute crop to fungal diseases, photoperiod sensitivity, and low yield under unfavorable growth conditions. The long term goal of this project is to better understand the jute genome and to produce transgenic jute varieties that will have higher fibre yield and quality than that of current cultivars without compromising their other important agronomic traits such as disease resistance, photoperiod insensitivity, strong and lustrous fibres, etc.

In this study, we aimed to construct a full-length cDNA library, to conduct EST analyses, and to analyze and classify gene functions from the leaves of '179' to lay foundations for the further utilization of the gene resources from '179' for the improvement of white jute by genetic transformation.

MATERIALS AND METHODS

Plant materials

The plant material used in this study was *C. capsularis* cultivar "179". "179" which has a fibre yield about 10% higher than that of Yue-yuan No. 5 (used as check) and also a high resistance to anthracnose, which is a potential material for molecular and biology study (Lu et al., 1983). The seedlings of "179" were grown in a greenhouse under natural conditions, then transplanted and grown under normal field management in Fujian Agriculture and Forestry University. Tender leaves were separated from the plants and immediately frozen in liquid nitrogen, then stored at -80°C until use.

Construction of the normalized cDNA library

Total RNA was extracted according to the protocol of RP3301 RNA extraction kit (Bioteke Corporation) with modified. Messenger RNA (mRNA) was isolated from total RNA using Oligotex (Qiagen, The Netherlands) mRNA extraction kit. The total RNA and mRNA quantities were determined spectrophotometrically at wavelengths of 230, 260, and 280 nm. The integrity of the total RNA and mRNA was verified by running samples on 1.1% agarose gels.

cDNA library construction and characterization

First and double-stranded cDNAs were synthesized as described in the manual of the SMART cDNA library construction kit (Clontech, USA). 50% of the double-stranded cDNAs were digested with proteinase K and fractionated by Creator^{TM} SMART^{TM} cDNA Library Construction Kit. The cDNA fragments were selected and purified and cloned into a pMD-18T vector; the recombinant DNA was put in *vitro* packaged using competent cell and LB liquid medium at 37°C for 1 h. *Escherichia coli* DH5α were infected with the phage from cDNA library to determine the titer of original and amplified library. The percentage of recombinant clones was determined by screening for blue/white plaques on medium containing X-gal and isopropyl β-D-1- thiogalactopyranoside. Finally, the titer and the recombination frequency of the library were calculated by the number of blue and white plaques: library titer P (pfu/mL) = number of plaques × dilution factor × 10^3 / volume of phage plates (μL), and recombination frequency (%) = number of white plaques / total

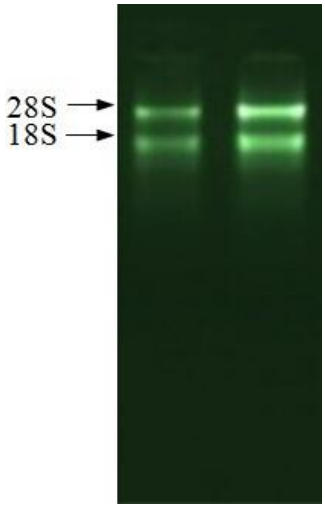

Figure 1. 1.1% agarose gel electrophoresis of total RNA from tender leaves of white jute cultivar '179'.

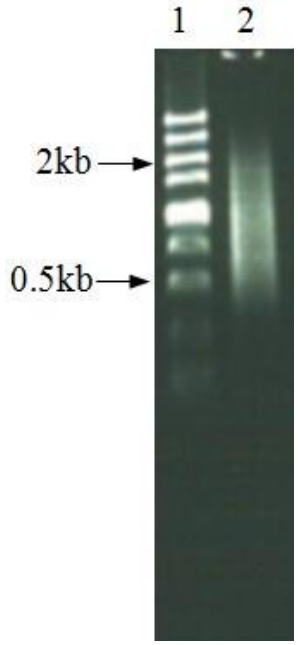

Figure 2. 1.1% agarose gel electrophoresis of double-strand cDNA. **1** represent DL5000 Marker; **2** represent the double-strand cDNA.

number of plaques. According to the sequences of two ends of the λ phage vector, forward primer was designed as 5'- CGCCA GGGTT TTCCC AGTCA CGAC-3', and the corresponding reverse primer was designed as 5'- GAGCG GATAA CAATT TCACA CAGG -3'. PCR procedures were carried out as follows: initial denaturation at 94°C for 5 min; 94°C for 30 s, 46°C for 45 s and 72°C for 2 min (35 recycles in total), and a primer extension at 72°C for 7 min. After amplification by plasmid PCR and identification by positive clone screening, suspected strains were combined with the same amount of 30% glycerin and then sent to Beijing Genomics institution (BGI, China) for sequencing.

Homology comparisons and analysis of ESTs

ESTs obtained were evaluated using software DNAstar and Chromas, and were spliced by software Phrad. Each edited EST was translated in reading frames and compared with the non-redundant database at the National Center for Biotechnology (NCBI) using the BLASTX program, which compares translated nucleotide sequences with protein sequences. ESTs longer than 100 bp and containing no more than 4% ambiguity were considered useful for data analysis. Using the BLAST service, at NCBI (http:// www.ncbi.nlm.nih.gov), sequences were searched against the protein and nucleic acid databases. Sequence similarities identified by the BLAST programs were considered to be statistically significant at an E-value of $\leq 10^{5}$. Molecular function of annotated genes was classified by Gene Ontology.

RESULTS

Construction of full-length cDNA library

Isolation of high quality total RNA is a critical step for constructing a cDNA library. In this study, the total RNA electrophoresis on 1.1% agarose gels showed distinct 18S and 28S (Figure 1), indicating good quality of total RNA. The optical density ratios at A260/280 and A260/230 were 1.99 and 2.26, respectively, suggesting little contamination of polysaccharides and proteins. The density of isolated total RNA was 418 ng/μL. The titer of the original library was approximately 2.32×10^{7} pfu/mL, much higher than 1.0×10^{6} pfu/mL, the estimated criteria for an available cDNA library. The amplified library was up to 1.07×10^{9} pfu/mL. The percentage of recombinants calculated through blue/white plaques was 98.3%. Size fractionated double-strand cDNA was visualized as a smear on the agarose gel (1.1%) with a size ranging from 0.6 to 1.5 kb (Figure 2). Further confirmation was obtained with electrophoresis PCR products of 44 randomly selected clones. The size of the cDNAs ranged from 0.5 to 1.5 kb, with an average size of 0.75 kb (Figure 3). Based on the above observation, the constructed cDNA library reached the criteria (library content and cDNA integrity) for isolating full-length expressed genes in *C. capsularis*.

Obtaining and splicing effective EST sequences

The sequences were trimmed off by their vector, adaptor, poly(A) tail, and low-quality sequences and filtered for minimum length (100 bp), resulting in a total of 219 high quality ESTs (Table 1), which percentage was 77.42%. The length of all the ESTs was from 106 to 1386 bp with an average size of 531 bp. Using Phrap software to

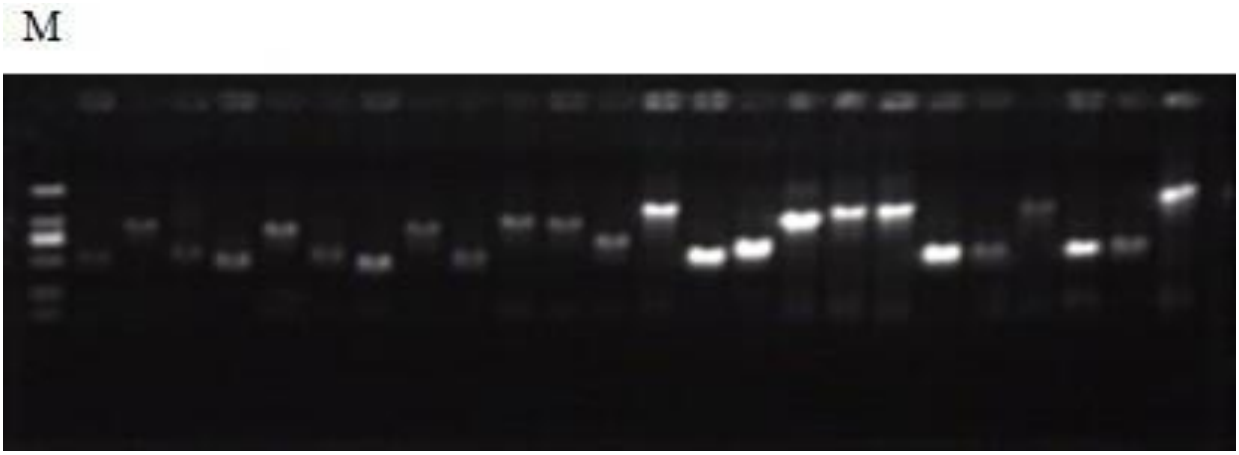

Figure 3. Twenty-four (24) clones in the library were selected randomly to evaluate their insert sizes. The size of the cDNAs ranged from 0.5 to 1.5 kb, with an average size of 0.75 kb. M represents DL2000 Marker.

Table 1. cDNA library, ESTs and cluster statistics of white jute cultivar '179'.

cDNA Library characteristics	Eigen value
Titer of cDNA library (pfu mL^{-1})	2.32×10^{7}
Recombination rate	98.3%
Average cDNA insert size	0.75 kb
Sequences passing quality check	219 (77.42%)
Average length	531bp
Singletons	37
Contigs	24

conduct cluster-analysis and post-splicing, a total of 61 non-repetitive sequences (unigenes) were found including 24 contigs and 37 singleton ESTs.

Functional analysis and classification based on gene ontology

To provide a deeper understanding of the gene expression in jute, 41 unigenes including 203 ESTs bearing known functions or assumed functions based on BLASTX were analyzed (Table 2). The results reveal putative functions of 41 unigenes include cold-inducible and light-inducible protein, chloroplast precursor, 50S ribosomal protein, oxygen-evolving enhancer protein and other hypothetical protein. There were other six unigenes which could be matched and annotated but which function could not be known. Moreover, a total number of 14 unigenes could not be matched and annotated, indicating represent genes are likely new genes with new function. GO has been used widely to predict gene functions and classification (Ashburner et al., 2000; Wang et al., 2007). In our study, there were 19 genes be classified successfully according to their function, which were clustered into 6 functional categories involved in the cellular processes of energy production and conversion, metabolism, translation, ribosomal structure and biogenesis (Figure 4). It revealed that the encoded energy production and conversion-related genes present the largest number of ESTs, with a total of 7 occupying 36.84% of the functionally described genes. Carbohydrate transport and metabolism-related genes also present a larger number of ESTs, with a total of 6 occupying 31.58% of the functionally described genes. Amino acid transport and metabolism genes occupied 10.53% of all described genes, with the same percentage of translation, ribosomal structure and biogenesis genes. Two types of genes had the least percentage of 5.26%, which were inorganic ion transport and metabolism genes and lipid transport and metabolism genes.

Homologous analysis

Sixty (61) unigenes were analyzed by BlastX and DNAman software, among which 47 unigenes were homologous with

Table 2. Annotations and functional classification of unigenes.

Putative function	Species matched	Matches no	E-value
Repeat-containing protein	*Glycine max*	ref|XP_003519707.1|	2.00E-61
Predicted protein	*Populus trichocarpa*	ref|XP_002312449.1|	2.00E-62
Predicted protein	*Populus trichocarpa*	ref|XP_002308765.1|	1.00E-57
Putative cold-inducible protein	*Camellia sinensis*	gb|ACB20694.1|	8.00E-43
Predicted protein	*Populus trichocarpa*	ref|XP_002303722.1|	2.00E-47
50S ribosomal protein L34, chloroplast precursor, putative	*Ricinus communis*	ref|XP_002522558.1|	2.30E-30
Ycf3	*Ranunculus macranthus*	gb|AAZ03964.1|	2.00E-16
Chalcone isomerase	*Gossypium hirsutum*	gb|ADG27840.1|	1.00E-88
Similar to Os01g0104000	*Vitis vinifera*	ref|XP_002263812.1|	4.00E-58
Ascorbate peroxidase	*Zea mays*	ref|NP_001170482.1|	6.00E-07
Oxygen-evolving enhancer protein3,chloroplast precursor, putative	*Ricinus communis*	ref|XP_002518571.1|	3.00E-25
Hypothetical protein	*Vitis vinifera*	ref|XP_002273533.1|	3.00E-31
Conserved hypothetical hypothetical protein	*Ricinus communis*	ref|XP_002517183.1|	5.00E-64
Light-inducible protein ATLS1	*Elaeis guineensis*	gb|ACF06473.1|	4.00E-54
Hypothetical protein osi_08842	*Oryza sativa* Indica Group	gb|EEC73954.1|	4.00E-28
Hypothetical protein	*Vitis vinifera*	ref|XP_002284937.1|	9.00E-46
Cytoplasmic ribosomal protein S13	*Solanum lycopersicum*	ref|NP_001234162.1|	9.00E-59
Hypothetical protein VITISV_001773	*Vitis vinifera*	emb|CAN64991.1|	6.00E-50
Predicted protein	*Populus trichocarpa*	ref|XP_002313119.1|	4.00E-37
Predicted protein	*Populus trichocarpa*	ref|XP_002306249.1|	4.00E-10
Hypothetical protein	*Vitis vinifera*	ref|XP_002263049.1|	7.00E-83
Serine hydroxymethyl transferase	*Corchorus olitorius*	gb|ABS72195.1|	8.00E-20
NADH dehydrogenase subunit 4	*Nicotiana tabacum*	dbj|BAD83480.2|	4.00E-10
Ribulose-1, 5-bisphosphate carboxylase/oxygenase small subunit 1A	*Corchorus olitorius*	gb|ABS72189.1|	3.00E-24
Tonoplast intrinsic protein putative	*Ricinus communis*	ref|XP_002531978.1|	1.00E-123
Unknown protein	*Nicotiana tabacum*	gb|AAA84679.1|	3.00E-06
Predicted protein	*Populus trichocarpa*	ref|XP_002327750.1|	1.00E-168
Conserved hypothetical protein	*Ricinus communis*	ref|XP_002525607.1|	4.00E-16
Acyl-coa-binding protein	*Jatropha curcas*	gb|ADB85092.1|	2.00E-40
Hypothetical protein osi_13874	*Oryza sativa* Indica Group	gb|EAY92161.1|	2.00E-22
Fructose -1, 6-bisphosphate aldolase class 1	*Carica papaya*	gb|AER26531.1|	2.00E-45
Ribulose-1, 5-bisphosphate carboxylase/oxygenase Small subunit 1a	*Corchorus olitorius*	gb|ABS72189.1|	1.00E-87
Hypothetical protein MTR_5g050970	*Medicago truncatula*	ref|XP_003614382.1|	2.00E-92
Predicted protein	*Populus trichocarpa*	ref|XP_002306693.1|	8.00E-88
Ribulose-1, 5-bisphosphate carboxylase/oxygenase Small subunit 1a	*Corchorus olitorius*	gb|ABS72189.1|	6.00E-76
Chloroplast precursor putative	*Ricinus communis*	ref|XP_002510603.1|	4.00E-60
Uncharacterized protein LOC100499745	*Glycine max*	ref|NP_001235442.1|	7.00E-54
Predicted protein	*Populus trichocarpa*	ref|XP_002316235.1|	8.00E-88
Hypothetical protein	*Vitis vinifera*	ref|XP_002263786.1|	1.00E-108
Ribulose-phosphate 3-epimerase, chloroplastic-like	*Glycine max*	ref|XP_003524706.1|	1.00E-114
Ribulose-1, 5-bisphosphate carboxylase/oxygenase Small subunit 1a	*Corchorus olitorius*	gb|ABS72189.1|	6.00E-90
Pantothenate kinase putative	*Ricinus communis*	ref|XP_002514412.1|	1.00E-139
3-oxo-5-alpha-steroid 4-dehydrogenase family protein	*Arabidopsis thaliana*	ref|NP_197105.1|	1.00E-23
Ribulose-1, 5-bisphosphate carboxylase/oxygenase Small subunit 1a	*Corchorus olitorius*	gb|ABS72189.1|	4.00E-92
NADH dehydrogenase subunit K	*Eucalyptus globulus* subsp. globulus	ref|YP_636303.1|	1.00E-140
Orf174	*Beta vulgaris* subsp vulgaris	dbj|BAD66815.1|	1.00E-08
Hypothetical protein POPTRDRAFT_712066	*Populus trichocarpa*	ref|XP_002304206.1|	7.00E-44

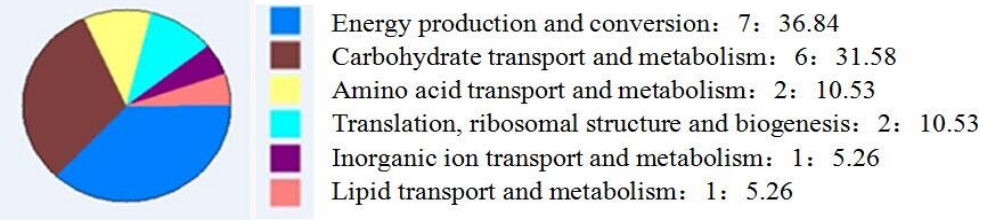

Figure 4. The GO classification of genes with function or putative function annotation.

Table 3. Numbers of homologous unigenes and matched species.

Species name	Homologous unigenes number	Percentage (%)
Populus trichocarpa	9	19.14
Ricinus communis	7	14.89
Corchorus olitorius	6	12.77
Vitis vinifera	6	12.77
Glycine max	3	6.38
Nicotiana tabacum	2	4.26
Oryza sativa Indica Group	2	4.26
Camellia sinensis	1	2.13
Arabidopsis thaliana	1	2.13
Ranunculus macranthus	1	2.13
Gossypium hirsutum	1	2.13
Zea mays	1	2.13
Elaeis guineensis	1	2.13
Solanum lycopersicum	1	2.13
Jatropha curcas	1	2.13
Carica papaya	1	2.13
Medicago truncatula	1	2.13
Beta vulgaris subsp. *vulgaris*	1	2.13
Eucalyptus globulus subsp. *globulus*	1	2.13

19 species (Table 3). The results show that unigenes of *C. capsularis* have highest homology to *Populus trichocarpa*, with a total of 9 occupying 19.14% of the analyzed unigenes. *Capsularis* unigenes also have higher homology to *Ricinus communis*, with a percentage 14.89% of the analyzed unigenes. The percentage of homologous unigenes to *Corchorus olitorius* of analyzed unigenes is 12.77%, with the same percentage of *Vitis vinifera.* Moreover, the *capsularis* unigenes are also homologous with other species, such as *Glycine max* (6.38%), *Nicotiana tabacum* (4.26%), *Oryza sativa* (Indica Group, 4.26%). While, the *capsularis* unigenes have lower homology percentage to *Camellia sinensis*, *Arabidopsis thaliana* and other 12 species, the percentage of analyzed unigenes is only 2.13%.

DISCUSSION

Quality of mRNA plays an important role in the construction of a full-length cDNA library, and high-quality mRNA is critical to the creation of full-length cDNA (Chen et al., 2012). Jute is a fairly primitive plant species somewhat different from those of other plant species (Samanta et al., 2011). It is difficult to isolate RNA from jute tissues because of the nature that much mucilage and phenolics are present in jute. The important steps taken into consideration in this study were the application of CsCl isopycnic centrifugation to remove insoluble poly-saccharides, and using polyvinylpyrrolidone (PVPP) to prevent oxidation of phenolics, as reported by Samanta et al (Samanta et al., 2011). The ratios of the optical density A260/230 and A260/280 were suggesting little contamination of polysaccharides and phenolics in isolated RNA, which was found to be appropriate for further downstream applications. A successful establish-ment of a cDNA library should contain almost all the expressed information possibly and should be examined with some quality index, such as abundance, integrity, and capacity (Zhang et al., 2012). Moreover, the titer of a cDNA library could be used as an evaluation criterion of the representativeness of the library (Yang et al., 2009a).

In general, it has been suggested that the titer of cDNA library should be above 1 × 10^6 pfu/mL. In the present study, the titer of the primary cDNA library was 2.32×10^7 pfu/mL, and the amplified library was up to 1.07×10^9 pfu/mL with a recombinant frequency of 98.3%, and the average insert size was 500 to 1500 bp. At the same time, the number of ESTs matching a particular gene should reflect the abundance of their corresponding cDNAs in the non-normalized library (Ewing et al., 1999). In our study, a total of 279 EST sequences were obtained after excluding the incomplete sequences, which provides the first nucleotide sequence data for white jute cultivar '179'. By functional speculation of the sequences of randomly chosen clones, 73% ESTs were supposed to be known putative function or have significant matches with hypothetical proteins, putative proteins, and 9% ESTs are assumed to be unknown proteins, 18% ESTs had no significant similarity to sequences in the public databases. Samanta et al. 2015 found 81% of the ESTs resulted by *C. capsularis* were similar with genes of known function, 2% showed homology with the putative sequences and 17% were similar to genes of unknown function (Samanta et al., 2015). The results indicated that the construction of the cDNA library of white jute was successful, which could serve as an important resource for the isolation of genes to be utilized in the genetic improvement of jute using genetic engineering.

The cDNA library of C. *capsularis* constructed by Islam et al. contained 106 primary clones comprising about 90% recombinants, and the average insert size as determined was 100 to 500 bp (Islam et al., 2005). Taliaferro et al. constructed a cDNA library of C. *capsularis*, which contained a sufficient number of primary clones comprising about 65% recombinants. The average insert size was 150 to 500 bp (Taliaferro et al., 2006). The library constructed by Samanta et al., 2015 using suppression subtractive hybridization resulted in 2,685 expressed sequence tags, which were assembled and clustered into 225 contigs and 231 singletons. It seems that the suppression subtractive hybridization library can result in more ESTs than cDNA library.

In our study, the functions of the encoded proteins from ESTs sequences were classified into six categories based on molecular function, and the most abundant GO terms are energy production and conversion-related genes, carbohydrate transport and metabolism-related genes. Additionally, amino acid transport and metabolism genes are also one important GO term, as well as the translation, ribosomal structure and biogenesis genes. While the other two types of genes, inorganic ion and lipid transport and metabolism genes, are the least abundant GO terms. Our results are similar with Islam's. According to the report of Islam, based on similar homology resulted the partial cDNAs encode proteins, including ribosomal protein, transport protein and chloroplast inner membrane protein. Taliaferro et al. also found several significant sequences of jute ESTs by analyzing the library, including those of the 60S acidic ribosomal protein and the Class I chitinase (Taliaferro et al., 2006). Moreover, our study indicated unigenes of *C. capsularis* have highest homology to *P. trichocarpa*, and which also have higher homology to *R. communis*, *C. olitorius* and *V. vinifera.* The results are different with Wazni et al., who reported the homology was maximum between black jute (*olitorius*) and cotton followed by citrus, grapevine, tobacco and arabidopsis (Wazni et al., 2007), which illustrates the differences between black and white jute. Above results would be a potential resource for comprehensive genomic studies in *Corchurus* species. The collection of ESTs presented in this study should prove useful tools for identifying *C. capsularis* homologues to important genes from other organisms. Nevertheless, the EST data presented here were limited. High and medium abundance expressed genes have a lower proportion of 39.35% relatively, while which often represent the specific characteristics or functions of cell or tissue. Whatever, the initial data of *C. capsularis* sequences will undoubtedly provide a foundation for future research.

In conclusion, this study presents the construction of cDNA library and analysis of ESTs from leaves of *C. capsularis*. The ultimate objective of the present study was to establish a database on white jute genome analysis and repository of ESTs and cDNA, which will be freely available to jute researchers all over the world to identify new genes and supply an effective alternative strategy for functional genomics. Further analysis will involve screening of the current library with probes of these expressed genes. Moreover, many of the sequences need to be isolated by cloning the cDNA ends. Future study aims at collecting suitable material from other parts of the plant such as young stem, bark tissue to clone genes controlling many important traits of agronomic importance including those for regulating lignin biosynthesis.

Conflict of interests

The author(s) did not declare any conflict of interest.

ACKNOWLEDGEMENTS

This project was sponsored by funds from Fujian Provincial Science Foundation (2013J01082); National Natural Science Foundation of China (31471549); National Agri-Industry Technology Research System for Crops of Bast and Leaf Fiber, China (CARS-19-E06);Construction of Germplasm Resources Platform for Bast Fiber Crops in Fujian, China (2010N2002).

REFERENCES

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M,

Rubin GM, Sherlock G (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25:25-9.
Basu A, Maiti MK, Sen SK (2003a). Nucleotide sequence of cinnamyl alcohol cDNA isolated from Corchorus capsularis. Direct submission to GenBank [Accesssion No.AY504425]
Basu A, Maiti MK, Sen SK (2003b). Cloning and sequencing of mRNA for caffeoyl- CoA-o-methyltransferase from Corchorus capsularis. Direct submission to GenBank. [Accession No. AY500159]
Chen G, Wang H, Gai JY, Zhu YL, Yang LF, Liu QQ, Zhang GC, Chen GH (2012). Construction and characterization of a full-length cDNA library and identification of genes involved in salinity stress in wild eggplant (Solanum torvum Swartz). Hortic. Environ. Biotechnol. 53: 158-166.
Ewing RM, Ben Kahla A, Poirot O, Lopez F, Audic S, Claverie JM, (1999). Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res. 9:950-9.
Islam AS, Taliaferro M, Lee CT, Ingram C, Montalvo RJ, Ende GVD, Alam S, Siddiqui J, Sathasivan K (2005). Preliminary progress in jute (Corchorus species) genome analysis. Plant Tissue Cult. Biotechnol. 15:145-156.
Khatun A (2007). Recent agricultural developments in jute, kenaf and mesta through traditional and biotechnological approaches. a seminar talk held in Myanmar in February 2007 under the auspices of International Jute Study Group 1, 1-5.
Kundu A, Chakraborty A, Mandal NA, Das D, Karmakar PG, Singh NK, Sarkar D (2015). A restriction-site-associated DNA (RAD) linkage map, comparative genomics and identification of QTL for histological fibre content coincident with those for retted bast fibre yield and its major components in jute (Corchorus olitorius L., Malvaceae s. l.). Mol. Breed. 35:19.
Lu HR, Zheng YY, Wang YJ, Qi JM (1983). The breeding of new jute cultivar "179" and its extension cultivation. Journal of Fujian Agricultural College 1, 1-6(in Chinese with English Abstract).
Manickavelu A, Kawaura K, Oishi K, Shin IT, Kohara Y, Yahiaoui N, Keller B, Abe R, Suzuki A, Nagayama T, Yano K, Ogihara Y(2012). Comprehensive functional analyses of expressed sequence tags in common wheat (Triticum aestivum). DNA Res. 19:165-77.
Rudd S (2003). Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci. 8:321-9.
Samanta P, Sadhukhan S, Das S, Joshi A, Sen SK, Basu A (2011). Isolation of RNA from field-grown jute (Corchorus capsularis) plant in different developmental stages for effective downstream molecular analysis. Mol. Biotechnol .49:109-115.
Samanta P, Sadhukhan S, Basu A (2015). Identification of differentially expressed transcripts associated with bast fibre development in Corchorus capsularis by suppression subtractive hybridization. Planta 241:371-385.
Taliaferro JM, Islam AS, Sathasivan K (2006). Expressed sequence tags (ESTs) from a jute (Corchorus olitorius) cDNA library. Plant Tissue Cult. Biotechnol. 16:95-104.
Talon M, Gmitter FG Jr. (2008). Citrus genomics. Int. J. Plant Genomics 2008:528361.
Tran T, Vu TQC, Abdullah MP, Omar H, Noroozi M, Ky H, Napis S (2011). Construction of cDNA library and preliminary analysis of expressed sequence tags from green microalga Ankistrodesmus convolutus Corda. Mol. Biol. Rep. 38:177-182.
Wang LN, Wu D, Yu SX, Fan SL, Song MZ, Pang CY, Liu JJ (2011a). Construction of a full-Length cDNA library of Gossypium hirsutum L. and identification of two MADS-Box genes. Agric. Sci. China 10:28-40.
Wang WK, Liu CC, Chiang TY, Chen MT, Chou CH, Yeh CH (2011b). Characterization of Expressed Sequence Tags from Flower Buds of Alpine Lilium formosanum using a Subtractive cDNA Library. Plant Mol. Biol. Rep. 29:88-97.
Wang YC, Chu YG, Liu GF, Wang MH, Jiang J, Hou YJ, Qu GZ, Yang CP (2007). Identification of expressed sequence tags in an alkali grass (Puccinellia tenuiflora) cDNA library. J. Plant Physiol. 164:78-89.
Wazni MW, Islam AS, Taliaferro JM, Anwar N, Sathasivan K (2007). Novel ESTs from a jute (Corchorus olitorius L.) cDNA library. Plant Tissue Cult. Biotechnol. 17:173-182.
Whitlock BA, Karol KG, Alverson WS (2003). Chloroplast DNA sequences confirm the placement of the enigmatic Oceanopapaver within Corchorus (Grewioideae: Malvaceae sensu lato., formerly Tiliaceae). Int. J. Plant Sci. 164 (1): 35-41.
Xiao D, Yin C, Zhang Q, Li JH, Gong PT, Li SH, Zhang GC, Gao YJ, Zhang XC (2011). Selection and identification of a new adhesion protein of Cryptosporidium parvum from a cDNA library by ribosome display. Exp. Parasitol .129:183-9.
Yamagishi J, Wakaguri H, Sugano S, Kawano S, Fujisaki K, Sugimoto C, Watanabe J, Suzuki Y, Kimata I, Xuan X (2011). Construction and analysis of full-length cDNA library of Cryptosporidium parvum. Parasitol. Int. 60:199-202.
Yang CJ, Wang J, Yang CP, Liu GF (2009a). Construction and analysis of a cDNA library from yellow-fruit ginseng (Panax ginseng CAMeyer.) leaf tissue. Afr. J. Biotechnol. 8:6082-6088.
Yang D, Tang ZH, Zhang LP, Zhao CP, Zheng YL (2009b). Construction, Characterization, and Expressed Sequence Tag (EST) Analysis of Normalized cDNA Library of Thermo-Photoperiod-Sensitive Genic Male Sterile (TPGMS) Wheat from Spike Developmental Stages. Plant Mol. Biol. Rep. 27:117-125.
Zhang D, Hu C, Ouyang Y, Yao J (2011a). Construction of a full-Length cDNA library and analysis of expressed sequence tags from inflorescence of apomictic Sabaigrass (Eulaliopsis binata). Plant Mol. Biol. Rep. 30:46-54.
Zhang GY, Qi JM, Xu JT, Niu XP, Zhang YJ, Tao AF, Zhang LW, Fang PP, Lin LH (2013). Overexpression of UDP-glucose pyrophosphorylase gene could increase cellulose content in jute (Corchorus capsularis L.). Biochem. Biophys. Res. Commun. 442: 153-158.
Zhang LW, Yuan MH, He XW, Liu X,Fang PP, Lin LH, Tao AF, Xu JT, a. Qi JM (2014). Development and universality evaluation of EST-SSR markers from GenBank in Jute. Acta Agromonica Sin. 40:1028-1034(in Chinese with English Abstract).
Zhang T, Li LJ, Hu FY, Zhao XQ, Fu BY, Yang DC (2012). Analysis of ESTs from a Normalized cDNA Library of the Rhizome Tip of Oryza longistaminata. J. Plant Biol. 55:33-42.
Zhang Z, Xiao G, Liu R, Tan T, Guan C, Wang G (2011b). Efficient construction of a normalized cDNA library of the high oleic acid rapeseed seed. Afr. J. Agric. Res. 6:4288-4292.
Zheng TG, Qiu WM, Fan GE, Zheng BB, Guo WW (2011). Construction and characterization of a cDNA library from floral organs and fruitlets of Citrus reticulata. Biol. Plant. 55:431-436.