Australian Journal of Crop Science
3(6):322-328 (2009)
ISSN: 1835-2707

Southern Cross Journals©2009
www.cropj.com

A computational and experimental approach for developing jute ESTs from genomic clones

Salim Ahmed[1], Md. Zinnatun Nabi [2], Md. Maksudul Alam[3], Md. Sazzadul Islam[3], Rozalynne Samira[3], Mahdi M Moosa[1], Haseena Khan[3,*]

[1]*Department of Genetic Engineering and Biotechnology, University of Dhaka, Bangladesh*
[2]*Department of Biochemistry and Medical Genetics, University of Manitoba, Canada*
[3]*Department of Biochemistry and Molecular Biology, University of Dhaka, Bangladesh*

* *Correspondending author: pmblab@gmail.com*

Abstract

Gene identification is a leading and essential step in understanding the genome of an organism or for manipulation of genes necessary for incorporating any novel characteristics. However, current technologies and resources are far from shedding light on the genome of organisms like jute for which the whole genome sequence is yet to be determined. An alternative approach for finding genes in such genomes is to find and characterize specific genes, especially those which are differentially expressed under certain conditions. Though expressed sequence tags (ESTs) are widely used for this purpose, generation and characterization of ESTs are not straight forward in many cases. In this situation the available genomic clones can be used to identify and characterize ESTs with the combination of accessible bioinformatics tools such as gene prediction software, NCBI web portal etc and experimental approaches such as RT-PCR followed by sequencing. A number of ESTs for jute have been developed from a previously constructed jute genomic library enriched for simple sequence repeats and also by using the technique of differential display. Some of these ESTs were found to be interesting in terms of their agronomic properties.

***Keyword**:* Bioinformatics, RT-PCR, DNA sequencing

Introduction

Jute is one of the most important and useful fiber crops but efforts on studies on jute at the molecular level are surprisingly limited. As of January 2009, only about 1210 DNA sequences have been deposited in GenBank. There are few research works at the molecular level on jute like DNA markers (Belayat et al., 2002; Gupta et al.,1996; Gupta and Varshney 2000; Basu et al., 2004; Jesmin et al., 2008; Keka et al., 2008; Mir et al., 2008), tissue culture (Saha et al., 1999; Huda et al., 2007), somatic hybridization (Khatun, 2007), genetic transformation (Ghosh et al., 2002; Sajib et al., 2008) and gene identification. But there are no studies on bioinformatics analysis of jute sequences. The main hindrance to conduct bioinformatics study on jute is the unavailability of genome sequence information of this fiber crop. But fortunately some genomic clones of jute have been constructed and sequences of such clones are available. This study analyzed these clones by chronological application of online available gene finding tools, database resources and common laboratory techniques to develop ESTs for jute to identify agronomically important genes along with a comparison with EST prediction using non-radioactive differential display. Computational gene identification in eukaryotic genomes remains a challenging task (Zhang, 2002). It can be divided into two major disciplines: homology based gene identification and *ab initio* gene identification (Mathe et al., 2002). Homology based gene identification includes database searches for sequence similarity and *ab initio* gene finding approaches include gene prediction software and tools which employs statistical parameters. *Ab initio* Gene finding tools like GenmarkHMM, Genscan and GlimmerHMM are already available in the internet, trained by the data set of *Arabidopsis thaliana* and can be used as a standard to identify plant genes (Partea et al., 2002).

For homology based searching, the databases (NCBI, EBI etc.) are a trove of sequence information. To prove, whether a DNA sequence is gene/part of gene or not, RT-PCR (Reverse Transcription Polymerase Chain Reaction) can be said as gold standard. RT-PCR is a very useful and sensitive technique in molecular biology which is used to produce the DNA copy of RNA. Since its introduction the method is constantly finding new applications, especially in projects aimed at detecting and cloning differentially expressed genes (Kusec, 1998).

Presently this technique followed by sequencing is widely used for gene annotation with low cost and high accuracy (Tenney et al., 2004*).* Since whole genome sequence of jute is not available, no computational analyses are perfect or near perfection. All the *ab initio* searches in this study are done against *Arabidopsis thaliana*. In this study, RT-PCR was the ultimate validation of the gene finding tools for identification of jute ESTs. Gene finding tools are found with both false positive and false negative results which has given an opportunity to justify the accuracy of these tools to identify ESTs from random jute genomic clones. Apart from the techniques mentioned above, differential display is used worldwide as a method to identify changes in gene expression and to discover novel genes and this leads to the development of ESTs.

Screening for differentially expressed genes is a straightforward approach to study the molecular basis of a biological system. In the last two decades, a rapid evolution has taken place in the field of differential screening technology and now-a-days high-throughput tools for genome-wide transcript profiling, such as expressed sequence tags and microarray analysis, are becoming readily available (Leivens et al., 2001). The main focus of this study was to devise a method for identification and confirmation of jute ESTs from an available jute genomic library. Starting with a random selection of twenty clones from the library, this study has identified and confirmed sixteen as ESTs which justified the use of this approach.

Materials and methods

Jute genomic clones

The jute genomic clones used in this study were selected randomly and the clones are named as J 6, J 7, J 8, J 14, J 17, J 23, J 26, J 58, J 61, J 62, J 93, J 120, J 122, J 124, J 126, J 128, J 132, J 137, J 140, and J 161

VecScreen Analysis

Genomic clones were first analyzed by the NCBI VecScreen,tool http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html) to determine the vector contamination in the clones. Most of the clones were found contaminated by vector sequences in their 5’ and 3’ end. For successive analyses vector sequences were removed.

NCBI Blastx

When the genomic clones were analyzed by the gene finding tools, they were further analyzed by the NCBI Blastx (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to search protein database using translated nucleotide query of the given query sequences of randomly selected jute genomic clones.

NCBI Nucleotide Blast

From the nucleotide Blast option of NCBI Blast page (http://blast.ncbi.nlm.nih.gov/Blast.cgi), Blastn was performed for each of the sequences to find the similarity between the query sequences with any EST (Expressed Sequence Tags) listed in the database. Non-human, Non-mouse EST database was selected as references.

Open Reading Frame (ORF) finding

To explore the ORF in the query sequences, ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) was used from NCBI web portal. To conduct this analysis standard genetic code was selected.

Gene prediction tools

Sequences of genomic clones after removal of vector contamination were analyzed by three online available gene prediction tools. These are GENSCAN (http://genes.mit.edu/GENSCAN.html),GenMark.HMM(http://exon.gatech.edu/genemark/eukhmm.cgi) and Glimmer HMM (http://nbc11.biologie.unikl.de/ framed/left/menu/auto/right/glimmerhmm/). For every analysis *Arabidopsis* was selected as reference data set against which the query sequences were compared.

Primer design

Online available Primer3 tool (http://frodo.wi.mit.edu/) was used to design gene specific primers for each of the genomic clones. For designing primers Tm was chosen between 55^0C – 65^0C. Among the

Table 1. Blastx results for query sequences with top three matched sequence, their gi number, species name in bracket with E-value and bit score. Sequences marked in green did not amplify in RT-PCR, while others gave amplification products.

Sequence Name	Matched with	E-value	Bit score
J6	GI:7267526 [*Arabidopsis thaliana*]	5.00E-15	85.9
	GI:218186460 [*Oryza sativa* Indica Group]	8.00E-15	85.1
	GI:147859821 [*Vitis vinifera*]	5e-15	73.9
J 7	GI:147783561[*Vitis vinifera*]	2.00E-10	70.5
	GI:224106718 [*Populus trichocarpa*]	8.00E-10	68.6
	GI:15233296 [*Arabidopsis thaliana*]	6.00E-09	65.5
J8	GI:157357130 [*Vitis vinifera*]	1.00E-05	54.3
	GI:223549553 [*Ricinus communis*]	2e-05	53.9
	GI:4586103 [*Arabidopsis thaliana*]	2.00E-05	53.90
J14	GI:224057990 [*Populus trichocarpa*]	3.00E-22	109.00
	GI:223530527 [*Ricinus communis*]	2e-21	107
	GI:18026956 [*Arabidopsis thaliana*]	2e-12	74.7
J17	GI:1545805 [*Nicotiana tabacum*]	6.00E-88	293
	GI:49532956[*Citrullus lanatus*]	3.00E-28	100
	GI:110740129[*Arabidopsis thaliana*]	4.00E-22	108
J23	GI:51534819 [uncultured bacterium]	3.9	35.4
	GI:189183803 [*Orientia tsutsugamushi* str. Ikeda]	5.2	35
	GI:110832726 [*Medicago sativa* subsp. x varia]	8.00E-17	91.7
J26	GI:125555678 [*Oryza sativa* (indica cultivar-group)]	1.00E-11	74.7
	GI:115468452 [*Oryza sativa* (japonica cultivar-group)]	1.00E-11	74.7
	GI:115468452 [*Oryza sativa* (japonica cultivar-group)]	1.00E-11	74.7
J58	GI:108864262 [*Oryza sativa* (japonica cultivar-group)]	2.00E-21	107
	GI:77549442 [*Oryza sativa* (japonica cultivar-group)]	2.00E-21	107
	GI:148546044 [*Pseudomonas putida*]	1.00E-52	210
J61	GI:11094371 [*Medicago sativa* subsp. x varia]	1.00E-16	91.3
	GI:157340050 [*Vitis vinifera*]	1.00E-16	90.9
	GI:157340050 [*Vitis vinifera*]	1.00E-16	90.9
J62	GI:220983459 [*Welwitschia mirabilis*]	6.00E-25	118
	GI:11465941 [*Nicotiana tabacum*]	7.00E-24	115
	GI:7525020 [*Arabidopsis thaliana*]	9.00E-24	114
J93	GI:149789412 [*Ricinus communis*]	1.00E-18	98.6
	GI:30697740(*Arabidopsis thaliana)*	4.00E-15	86.7
	GI:224101661[*Populus trichocarpa*]	3e-20	103
	GI:87116466 [*Ipomoea batatas*]	3.00E-22	82
J120	GI:87116459 [*Ipomoea batatas*]	3.00E-22	82
	GI:149277062 [*Pedobacter* sp. BAL39]	1.00E-19	100
J122	GI:119368534[*Gossypium barbadense*]	1.00E-65	253
	GI:7525068[*Arabidopsis thaliana*]	2.00E-59	233
J124	GI:157345191 [*Vitis vinifera*]	1.00E-52	211
	GI:110006646 [*Populus trichocarpa*]	8.00E-50	201
	GI:2865437[*Arabidopsis arenosa*]	3e-31	139
J126	GI:147798271 [*Vitis vinifera*]	1.00E-15	87.8
	GI:157331387 [*Vitis vinifera*]	1.00E-15	87.4
	GI:150025305 [*Flavobacterium psychrophilum* JIP02/86]	2.00E-85	319
J128	GI:109676342[*Populus trichocarpa*]	1.00E-32	144
	GI:157331950 [*Vitis vinifera*]	1.00E-31	140
	GI:147790729 [*Vitis vinifera*]	1.00E-31	140
J132	GI:147802484 [*Vitis vinifera*]	6.00E-79	298
	GI:15240528 [*Arabidopsis thaliana*]	1.00E-77	286
	GI:149279338 [*Pedobacter* sp. BAL39]	1.00E-54	217
J137	GI:110006646 [*Populus trichocarpa*]	1.00E-41	174
	GI:147817139 [*Vitis vinifera*]	4.00E-41	172
	GI:15224592 [*Arabidopsis thaliana*]	3.00E-12	76.6
J140	GI:222641022 [*Oryza sativa* (Japonica Group)]	0.025	42.7
	GI:224118590[*Populus trichocarpa*]	0.043	42
	GI:223535970[*Ricinus communis*]	0.025	42.7
J161	GI:157359191 [*Vitis vinifera*]	1.00E-15	87.8
	GI:224122746 [*Populus trichocarpa*]	4.00E-15	86.3
	GI:218196773 [*Oryza sativa* (japonica cultivar group)]	2.00E-13	80.5

proposed primers from primer3 output primer with least self complimentarily were chosen.

Plant materials, seed germination, collection and storage

Leaf and seed from jute variety O-9897 was used for this study. Seeds from 0-9897 were incubated on water soaked filter paper in separate Petri dishes at room temperature in the absence of light for 4 days. All developing seedlings were collected after 4 days of their germination. Fresh and healthy seedlings were collected for the maximum yield of high quality total RNA, isolated from these plant tissues. After harvesting, samples were stored at -80°C until the extraction.

Total RNA and Genomic DNA isolation

Total RNA from O-9897 leaf and seedling was isolated using TRIZOL Reagent (GibcoBRL). After isolating RNA, its quality was confirmed by running RNA on a 1.3% denaturing agarose gel. Purity of RNA was determined by spectrophotometric analyses

Table 2. Blastx results for the sequences obtained by non radioactive differential display

Sequence Name	Matched with	E-value	Bit score
EXGT	GI:21553421 [*Arabidopsis thaliana*]	7e-44	179
	GI:179743760 [*Brassica rapa* subsp. Pekinensis]	3e-43	177
	GI:217074188 [Medicago truncatula]	8e-43	176
CT2	GI:222102895 [*Agrobacterium vitis* S4]	5e-22	97.8
	GI:17546264 [*Ralstonia solanacearum* GMI1000]	7e-22	93.2
	GI:207724015 [*Ralstonia solanacearum* MolK2]	1e-19	87.8

using the ratio of absorbance at 260 nm and 280 nm (*A*260/*A*280). Quantity of RNA was determined using Nanodrop (nanodrop 100, Spectophotometer). Genomic DNA was isolated from O-9897 using CTAB method and it's purity was confirmed after running the genomic DNA on a 2% agarose gel.

RT-PCR using RNA and PCR from genomic DNA

For the construction of cDNA from the total RNA using the primers used in this study, GeNei™ one step M-MuLV RT-PCR kit was used. Mixture of enzyme, template and primer was prepared according to the manufacturer's protocol. Thermal cycler profile followed for the one step RT-PCR was - RT incubation: 50^0C for 30 minutes, Initial denaturation: 95ºC for 5minutes, Denaturation 95ºC for 30 seconds, Annealing: according to primer Tm, Elongation 72ºC for 45 seconds, Final elongation: 72ºC for 5.0 minutes. Denaturation, Annealing, Elongation steps were continued for 35 cycles and rest of the steps were of single cycle. PCR from genomic DNA using the same primer pairs was carried out under identical conditions like that of RT-PCR except for the RT incubation step. Every time, a control PCR was conducted using RNA as template to find out whether the isolated RNA is contaminated with DNA or not. Each time PCR products were run on a 2% agarose gel.

DNA extraction from agarose gel for sequencing

For sequencing purposes DNA bands from agarose gel were extracted using QIAGEN MinElute Gel Extraction Kit followed by gel run to check the presence of band and for measuring concentration. Each of the gel extracted products were then sequenced using ABI-prism DNA sequencer.

RT-PCR, cloning and sequencing for non-radioactive differential display

Total RNA was isolated from different varieties of jute using the same method stated above. RT-PCR was carried using anchored oligo dT-12 as a 3'-end primer, 5X RT-Buffer, 0.1 M dTT, 2.0 mM dNTPs, 40 units of RNaseOUT (Invitrogen), reverse transcriptase. RT-products were then used for PCR reaction using a thermal cycling profiling-94^0C for 5 mins; followed by 35 cycles of 40s at 94^0C, 1.20s at 42^0C and 40s at 72^0C; and the final extension at 72^0C for 7 mins. PCR products were analyzed by separating them in 6% polyacrylamide gel followed by silver staining. The amplified PCR products were cloned into T-tailed pCR2.1 vector (Invitrogen) and sequenced.

Results

Result of homology based study was based on the outcome of Blastx, Blastn and ORF finder results. Blastx align the query nucleotide sequences with the protein database (Altschul et al., 1990). This search is more sensitive than nucleotide blast since the comparison is performed at the protein level (NCBI Blast program selection guide). Results of Blastx analysis are given in Table 1. When the sequences of the clones of differential display were obtained (unpublished data), they were aligned using the Blastx program from NCBI. The result of this Blastx analysis is given in Table 2.

Traditionally, gene prediction programs that rely only on the statistical qualities of exons have been referred to as performing *ab initio* prediction (Semple, 2000) which upon statistical analysis of the raw genomic sequence can detect upto 98% genes present. In this study, Genmark.HMM, Genscan and GlimmerHMM were used to predict ESTs from jute genomic clones. However there are differences in the algorithms among these software like GenMarkHMM and Genscan are both built on Hidden Markov Model (HMM) whereas Glimmer is built on Interpolated Markov Model (IMM). IMMs are a generalization of fixed-order Markov chains (Pertea et al., 2002). The main distinction lies on deciding in advance how many bases to consider for each prediction. These models therefore vary in the number of bases considered for making the prediction (Mahony et al., 2004). Results from three gene finding tools are given

Table 3. Results of three gene prediction tools (Genmark HMM, Genscan, GlimmerHMM). 'Y' represents the sequence predicted as gene by software and 'N' represents sequence not predicted as gene. The left most column represents the jute genomic clones. The clones marked in red are those which after software prediction were proven to be part of a gene by RT-PCR. The clones in green are those which were predicted as part of a gene but failed to amplify in RT-PCR Clones in blue were not predicted as part of a gene by any software but amplified to give band(s) in RT-PCR and clones in pink are those which were not predicted by software as part of a gene and failed to amplify in RT-PCR.

Sequence	GenMark.hmm	Genscan	GlimmerHMM
J 6	Y	N	Y
J 7	N	Y	Y
J 8	Y	Y	Y
J 14	Y	Y	N
J 17	N	N	Y
J 23	Y	N	N
J 26	Y	Y	Y
J 58	Y	N	N
J 61	Y	Y	Y
J 62	Y	Y	Y
J 93	Y	Y	Y
J 120	Y	Y	Y
J 122	N	Y	N
J 124	Y	Y	Y
J 126	Y	Y	Y
J 128	Y	Y	Y
J 132	Y	Y	Y
J 137	Y	Y	Y
J 140	Y	N	N
J 161	Y	Y	Y
EXGT	Y	Y	Y
CT 2	Y	Y	Y

in Table 2. ESTs (EXGT and CT 2) developed by differential display were also analyzed by these three software. To verify the results of gene prediction by the software RT-PCR was performed using GeNei™ one step M-MuLV RT-PCR kit. Results of RT-PCR are given in Figure 1 (a and b). While RT-PCR positive clones were sequenced after extracting from the gel, each of the sequences matched with the sequences of initial clones which represented that RT-PCR positive clones are jute ESTs. Gene prediction accuracy is calculated from the results of the gene finding software and RT-PCR analyses. While each software predicted some genes correctly, wrong prediction was not uncommon. Accuracy is the ratio of number of correct predictions and total number of

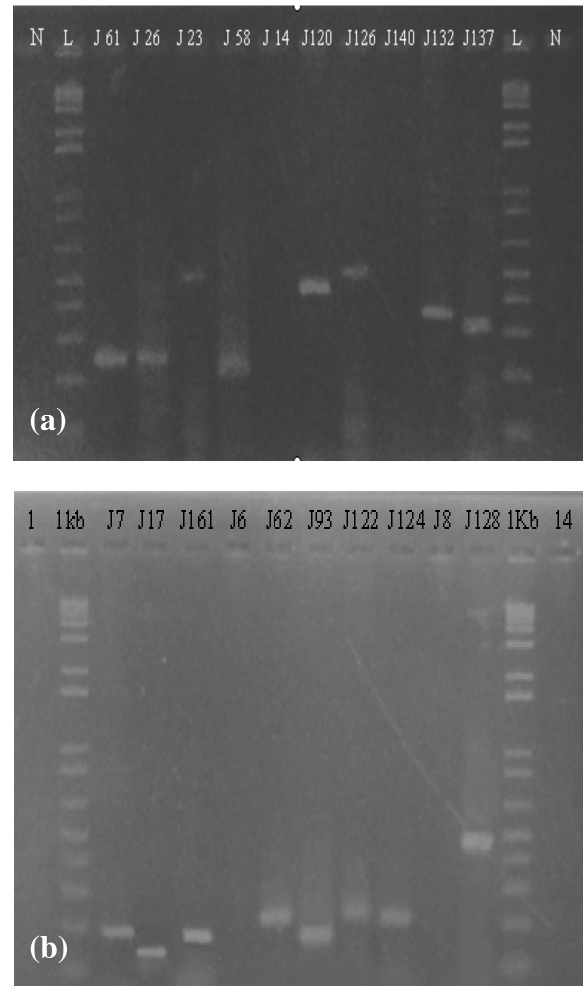

Fig 1a and b : The bands of RT-PCR positive clones. The right and left most lanes of both pictures represent PCR using RNA only.

predictions (Wong, 2004). Number of correct prediction is the sum of total positive (TP) and total negative (TN) where TP stands for correctly predicted as coding and TN stands for correctly predicted as non-coding. Number of total prediction includes two additional counts, false positive (FP) and false negative (FN) along with TP and TN. So accuracy in this study was calculated using the formula $\{(TP+TN)/ (TP+TN+FP+FN)\}\times 100\%$. The accuracy calculation so obtained for the three software mentioned above are given in table 4. However it should be noted that this accuracy does not reflect the actual precision of the software for predicting the genes of other plant species. It can only be valid for random jute genomic clones. It is mentionable that concomitant to this study the whole sequence of the putative leucine rich repeat transmembrane protein kinase gene (gi|209168629|) and ribosomal S8 protein gene (GQ325661) in jute were identified using the

techniques like degenerate primer, gene walking etc. starting from the genomic clone J 132 (unpublished data) and J 122 respectively.

Discussion

In recent years, researchers are getting interested in improving agronomic traits of jute through molecular approaches, which has however failed to gain momentum since the whole genome sequence is yet to be determined. Moreover, most of the small number of jute sequences deposited in GenBank is still uncharacterized. Thus the huge number of genes, responsible for various characteristics of the jute fiber, their functions as well as the analyses and characterization of the respected protein products are of fundamental importance in understanding and manipulating the different properties of jute. In this context, the present study was under taken to predict genes from the jute genomic clones and to attest them as expressed sequence tags of jute. Nowadays advancement of bioinformatics research at the molecular level has generated a tremendous pace. However without the complete genome sequence it is not easy to achieve speed and attain quality of such research. In this daunting situation bioinformatics analyses were initiated in the present study with the easy and existing tools available on the internet and to confirm these analyses, appropriate laboratory techniques were used. Jute genomic clones were analyzed by three gene prediction software: Gen MarkHMM, Genscan and GlimmerHMM. Although these software have never been trained for jute, they are well trained for *Arabidopsis* sequences. Therefore for every analysis *Arabidopsis* was used as a standard. In 2005 GeneMark was improved by addition of parameters allowing detection of genes in eukaryotic genomes with the predictors for splice sites, translation initiation signals and exons and introns (Besemer et al., 2005). Genscan and GlimmerHMM were also trained with *Arabidopsis* data and were tested on *Arabidopsis* genes confirmed by the presence of cDNAs (Burge et al., 1997, Majoros et al., 2004). GenmarkHMM, Genscan and GlimmerHMM recognized majority of the clones as gene sequences but also gave false positive and false negative results. After experimental verification, the accuracy of these software to predict jute ESTs were calculated, however it should be noted that this accuracy is based on the small population of jute genomic clones. Extensive analysis with large data set is required to identify their accuracy more appropriately but the approach will remain the same. Blastx was performed to match the clones with the existing protein coding genes in the database. Finally, the RT-PCR positive clones were those which were found to be similar to plant genes on Blastx analysis and were positive in *ab initio* (except for the few false negatives) prediction. These results highlight the finding that only the *ab initio* approach of gene finding is not satisfactory in identifying ESTs. A combination of *ab initio* procedures, homology based approach and experimental techniques like RT-PCR can be used to predict jute ESTs with optimum decisiveness. In contrast to this combination of computational and experimental approach, different-tial display has some limitations in establishing ESTs although differential display is commonly used to study differentially expressed genes at the mRNA level. The high percentage of false positive results (Zegzouti et al., 1997), labor intensive downstream verification process and necessity of high quantity of RNA (Leivens et al., 2001) have limited use in EST development when genomic clones are available. Moreover the cost of this technology remains prohibitive for most laboratories and does not give rise to many differential ESTs. To conclude, it may be said that the protocol standardized in this study can be the beginning of analyzing jute at the molecular level with the assistance of bioinformatics along with the available molecular biology laboratory techni-ques. Identifying the genes of an organism, using traditional molecular biology techniques for which little or no information is available in the global database, is painstaking, time consuming and requires extensive labor. This problem can be alleviated by blending available computational tools and laboratory techniques. This research has successfully achieved this goal. Although there is still a long way to go but the jute EST collection is already starting to have an impact on research and this procedure is not only applicable for jute but also can be applied for other species whose genome is yet to be sequenced and have diminutive information in the database.

Table 4. Accuracy determination of the gene the three finding tools (GenscanHMM, Genscan and Glimmer-HMM) for random jute genomic clones.

Software	TP	TN	FP	FN	Accuracy (%)
GenMarkHMM	16	0	4	2	72.72%
Genscan	15	3	2	2	80.81%
GlimmerHMM	15	3	2	2	80.81%

Acknowledgement

This work was financially supported by US Department of Agriculture (USDA). The authors are grateful to Bangladesh Jute Research Institute (BJRI) for supplying jute seeds.

References

Akter J, Islam MS, Sajib AA., Ashraf N, Haque S, Khan H (2008) Microsatellite markers for determining genetic identities and genetic diversity among jute cultivars. *Aust J* Crop *Sci* 1(3): 97-107

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

Basu A, Ghosh M, Meyer R, Powell W, Basak SL, Sen SK (2004) Analysis of Genetic Diversity in Cultivated Jute Determined by Means of SSR Markers and AFLP Profiling. *Crop Science* 44: 678–685

Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* 33: 451-454

Burge C, Karlin S (1997) Prdiction of complete gene structures in human genomic DNA. *J Mol Biol* 268: 78-94

Colin S (2000) Gene prediction: the end of the beginning. *Genome Biology* 1(2):4012.1 – 4012.3

Ghosh M, Saha T, Nayak P, Sen SK (2002) Genetic transformation by particle, bombardment of cultivated jute, *Corchorus capsularis* L. *Plant Cell Rep* 20: 936–942

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113:163–185

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. *Curr Sci* 70: 45–54

Hossain MB, Haque, Khan H (2002) DNA Fingerprinting of Jute Germplasm by RAPD. *Journal of Biochemistry and Molecular Biology* 35 (4): 414-419

Keka SI, Shamsuzzaman M, Pahloan MU, Pervin S, Rahman MM, Khan H (2008) Identifying Simple Sequence Repeat (SSR) Marker Linked To Mite Tolerance In Jute Species. *Bangladesh J Bot* 37(2): 161-171

Khatun A (2007) Recent agricultural developments in Jute, Kenaf and Mesta through Traditional and Biotechnological approaches. Paper presented at a seminar in Myanmar co-organized by IJSG (www. jute.org/news.htm).

Kusec R (1998) Clinical applications of PCR. *Met in Mol Med* 16: 149-158 Lievens S, Goormachtig S, Holsters M (2001) A critical evaluation of differential display ,as a tool to identify genes involved in legume nodulation: looking back and looking forward. *Nucleic Acid Res* 29 (17): 3459-3468

Mahony S, McInerney JO, Smith TJ Golden A (2004) Gene prediction using the Self- Organizing Map: automatic generation of multiple gene models, *BMC Bioinformatics* 5:23

Majoros WH, Pertea M, Salzberg SL (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20: 2878-9

Mathe C, Sagot MF, Schiex T, Rouze P (2002) Current methods of gene prediction, their strengths and weaknesses. *Nucleic Acid Res* 30: 4103-17

Mir RR, Rustgi S, Sharma S, Singh R, Goyal A, Kumar J, Gaur A, Tyagi AK, Khan H, Sinha MK, Balyan HS and Gupta PK (2008) A preliminary genetic analysis of fibre traits and the use of new genomic SSRs for genetic diversity in jute. *Euphytica* 161: 413-427

Pertea M, Salzberg SL (2002) Computational gene finding in plants. *Plant Molecular Biology* 48: 39-48

Saha T, Ghosh M, Sen SK (1999) Plant regeneration from cotyledonary explants of jute *Corchorus capsularis* L. *Plant Cell Rep* 18:544–548

Sajib AA, Islam MS, Reza MS, Bhowmik A, Fatema L, Khan H (2008) Tissue culture independent transformation for *Corchorus olitorius*. *plant cell tiss and organ cult* 95 (3): 333-340

Singh VK, Kumar A (2001) PCR Primer Design. *Mol. Biol Today* 2: 27-32 Tenney AE, Brown RH, Vaske C, Lodge JK, Doering TL, Brent MR (2004) Gene, prediction and verification in a compact genome with numerous small introns. *Genome Res* 14: 2330-2335

Wong L (2004) The Practical Bioinformatician, Chapter 6, p 128, World Scientific, New Jersey

Zegzouti H, Marty C, Jones B, Bouquin T, Latche A, Pech JC, Bouzayen M (1997) Improved Screening of cDNAs Generated by mRNA Differential Display Enables the Selection of True Positives and the Isolation of Weakly Expressed Messages. Plant *Mol Biol Rep*. 15: 236–245

Zhang MQ (2002) Computational prediction of eukaryotic protein-coding genes. *Nature Rev Genet* 3: 698–710