Journal Homepage: -www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI: 10.21474/IJAR01/2336
DOI URL: http://dx.doi.org/10.21474/IJAR01/2336

RESEARCH ARTICLE

IN-SILICO CHARACTERIZATION OF EST SEQUENCES FOR CELLULOSE SYNTHASE IN SUGARCANE.

Jyotika Bhati[1], Neetesh Pandey[1], *Monendra Grover[1], NiharRanjan Nayak[2], Nitesh Kumar Mund[2]

1. Centre for Agricultural Bio-Informatics, ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, PUSA, New Delhi, India.
2. Regional Plant Resource Centre, Nayapalli, Bhubaneswar, Odisha, 751 015 India.

Manuscript Info

Manuscript History

Received: 30 September 2016
Final Accepted: 30 October 2016
Published: November 2016

Key words:-
Sugarcane, EST sequences, Cellulose synthase, Molecular function, Biological process and Cellular component.

Abstract

Sugarcane is an economically important perennial grass of genus *Sacchrum.* Cellulose is important product of sugarcane and is synthesized by cellulose synthase. Since the complete genome sequence of sugarcane is not available, studies on ESTs can be a valuable source of information for the genes coding for important traits. We aim to contribute to more complete understanding of economically important process of cellulose production which has multifarious uses in industry. In this study we have downloaded ESTs for cellulose synthase in sugarcane. These have been assembled and annotated.438 GO terms were retrieved which means on an average 8 GO terms per contigs were obtained. A maximum of 15 GO terms were found for two contigs. Furthermore EST-contigs sequences were categorized according to the GO vocabularies i.e.Molecular Function, Cellular Component and Biological Process with 120, 118 and 200 GO terms obtained for each category respectively. All the EST-contigs showed involvement in starch and sucrose metabolism.

Introduction:-

Sugarcane is one of the numerous species of tall perennial true grasses of the genus *Saccharum*, community *Andropogoneae*, family *Poaceae* and is most prominently used for production of sugar. It has determined jointed fibrous stalks that are rich in the sugar sucrose. The sugarcane plant is in between two to six meters tall. The sugarcane is also used as a fuel in sugar mills, for paper production and as a component of fibreboard. The Sugar, extracted in specialized mill factories, is used as important material in human food industries. The maximum amount of sugar is used in the production of soft drinks, ice-creams, alcoholic beverages, chocolates and biscuits etc.

Plant cellulose synthases belong to the family of glycosyltransferases, which are proteins essential in the biosynthesis and hydrolysis of the majority of earth's biomass(Campbell, et al., 1997). Cellulose is synthesized by large cellulose synthase complexes (CSCs), which consist of synthase protein isoforms (CesA) that are arranged into a unique hexagonal structure known as a particle rosette (Bowling and Brown Jr, 2008; Giddings, et al., 1980; Yin, et al., 2009). There are more than 20 of these full-length integral membrane proteins, each of which is around 1000 amino acid long (Olek, et al., 2014; Richmond, 2000). Cellulose is a total of unbranched polymers of -1,4-linked glucose residues, is the main constituent of wood and accordingly paper, and is synthesized by plants, some

Corresponding Author: -Monendra Grover
Address: - Centre for Agricultural Bio-Informatics, ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, PUSA, New Delhi, India.

animal, fungi , bacteria and mostly algae. There are no known crystal structures for cellulose synthase proteins, and precise enzymatic mechanism is unidentified. A number of mutations have been identified in cellulose synthase genes in the model organism *Arabidopsis thaliana*. Due to lack of a properly developed cell wall, the mutations show the altered morphology (Richmond, 2000). Cellulose is the richest constituent of plant cell walls and its significance is well defined. In the primary cell wall it generates part of the load-behaviour system that both maintains and controls cell shape, while permitting regulated cell expansion that is important during growth. Cellulose is also the chief component of secondary cell walls where it plays an important role for mechanical strength for the plant (Cosgrove, 2005; Taylor, 2008).

Expressed sequence tag assemblies may be valuable as shortest sources of genes, but their main benefit is in permitting comparisons of expression pattern to be made between target tissues. Those evaluations have been made highly well-organized by the application of array screening machineries and these have been profitably employed in sugarcane, using macroarray and microarray systems (Carson and Botha, 2002; Casu, et al., 2003). ESTs development has countless benefits for facilitating crop development for stress resistance. It is becoming useful, for example, to combine large quantities of DNA sequence data with high-throughput molecular biology approaches to identify genes that are differentially expressed under specific environmental conditions (Hamilton and Robin Buell, 2012; Henry, 2012). The studies on Sugarcane ESTs are of special significance since complete genome of sugarcane is not available. In this study the sugarcane EST sequences have been assembled and annotated for more complete understanding of cellulose synthase genes and consequently cellulose production.

Materials and Methods:-

Data Source and Assembly:-

In this study 69 ESTs corresponding to cellulose synthase in sugarcane were downloaded from GenomeNet from EGENOME T20007. In order to improve the efficiency of the similarity based clustering, ESTs were masked to eliminate sequence parts that would cause incorrect clustering (Nagaraj, et al., 2007). The processed ESTs sequences were grouped into clusters based on sequences similarity to have stable clusters using CAP3 software, with threshold set at default. Each cluster of ESTs were assembled separately with overlap 80 percent and minimum number of nucleotides 65 (Huang and Madan, 1999) of EGAssembler software (Masoudi-Nejad, et al., 2006). The sequences which cannot be grouped due to their low similarity to other ESTs results in singletons. These singletons may represent genes where only single mRNA has been collected for the expressed gene or may be a result of contamination, were not considered for further analysis in this study.

Gene prediction (partial / full length):-

These sequences were used to predict the structure of gene with TSS (Transcription Start Site), PolyA tails at the extremes and CDS (Coding Sequence) in between by a gene prediction program FGENESH (Conesa, et al., 2005). *Triticumaestivum* based gene finding algorithm was selected for finding the gene structure.

Functional analysis of EST-contigs:-

Functional analysis of the EST-contigs was performed using Blast2GO v 2.5 (Conesa and Götz, 2008). Blast2GO is Gene Ontology based annotation tool and found to be effective in the functional characterization of plant sequence data (Salamov and Solovyev, 2000). The EST-contigs homologous with annotated proteins in nr database were selected for functional characterisation based on maximum E-value ($1E^{-3}$) and the minimum alignment size (HSP length 33) using BLASTX. The EST-contigs sequences were then categorized according to the GO vocabularies into three categories i.e. molecular function, biological process and cellular component. The distribution of GO terms was analysed at level 2 of the Directed Acyclic Graphs.

Results and Discussion:-

Assembling of ESTs into Contigs:-

A total of 69 EST sequences related to cellulose synthase of sugarcane were downloaded from GenomeNet. The average length of theses ESTs is 952 base pairs. The ESTs sequences for *Saccharum* were assembled into 57 EST-contigs. These assembled EST account for 82.6% of the size of total ESTs. Less abundant or lowly expressed transcripts could not be assembled into larger contigs and remained as 12 singletons.

Gene Structural Prediction:-

All 57 EST-contigs were submitted to FGENESH program for structural prediction of genes, comprising TSS, CDS and PolA as shown in Table 1. Almost all the EST-contigs had partial gene structure, mainly comprising of CDS region and PolA, except three EST-contigs which showed TSS and CDS regions. Only one EST-contig had the complete gene structure with TSS, CDS and PolA regions.

Table 1: Coding position of EST-contigs

ID	Length	TSS	CDS	CDS Start	CDS End	PolA
458	784		CDSi	108	213	410
			CDSl	341	390	
2512	1775		CDSo	67	1266	1432
2694	1350		CDSo	11	1267	1306
4498	830	741	CDSo	169	660	149
10675	3777		CDSi	1	3115	3397
			CDSl	3229	3317	
12205	921		CDSo	204	692	816
12324	977		CDSo	13	576	719
13032	1144		CDSi	249	676	
			CDSi	741	1116	
14543	634		CDSi	1	603	
17321	1059		CDSf	57	450	
			CDSi	504	1006	
19135	767		CDSo	38	679	730
21175	2887		CDSo	47	2551	2857
27502	643		CDSl	39	439	623
33648	948	113	CDSf	151	944	
36308	901		CDSl	189	791	22
40327	965		CDSo	147	773	912
42091	686		CDSo	46	651	671
42473	699		CDSi	175	448	
42487	505		CDSi	1	496	
48353	669		CDSl	96	358	637
48412	744		CDSl	4	523	716
48751	879		CDSi	1	458	
49293	555		CDSl	84	440	
51150	604		CDSf	85	580	
52004	815	354	CDSf	22	343	
52014	718		CDSo	208	582	654
52380	557		CDSi	57	342	
53104	661		CDSf	59	297	623
			CDSl	525	597	
53434	684		CDSi	1	622	
53735	841		CDSl	38	611	722
54137	608		CDSo	122	550	588
54162	729		CDSf	30	263	
			CDSl	317	685	
55281	624		CDSl	4	404	540
55469	475		CDSl	30	349	464
56522	524		CDSo	107	505	
56627	623		CDSo	95	598	
56799	684		CDSf	68	542	
56823	644		CDSi	35	247	605
			CDSl	360	514	
7082	2191		CDSf	126	2188	
8385	1516		CDSl	57	1348	1362

9753	4054		CDSl	71	3406	3862
56842	702		CDSo	42	698	
58073	752		CDSl	78	493	602
58721	699		CDSi	1	444	
59804	749		CDSo	8	613	645
61251	729		CDSi	75	77	
			CDSi	277	560	
63327	583		CDSo	61	549	
67711	673		CDSi	70	125	386
			CDSl	206	292	
70198	538		CDSi	1	505	
70897	1154		CDSi	45	175	663
			CDSl	348	648	
71721	692	659	CDSo	113	616	
Contig1	3815		CDSf	113	388	3369
			CDSi	438	529	
			CDSl	651	3354	
Contig2	1488		CDSi	1	859	
			CDSl	945	1460	
Contig3	957		CDSi	1	907	
Contig4	738		CDSo	147	617	
Contig5	1041		CDSi	1	1008	

Functional Annotation of EST-contigs:-

For functional characterization 57 assembled and translated EST-contigs were compared against NCBI *nr* database. Out of 57 EST contigs, 54 contigs were selected based on homology search, which were further subjected to GO functional classification. The GO terms were available for all 54 EST contigs. It has been noticed that overall 438 GO terms were retrieved which means on an average 8 GO terms per contigs were obtained. A maximum of 15 GO terms were found for two contigs. Furthermore EST-contigs sequences were categorized according to the GO vocabularies i.e.Molecular Function (Fig 1), Cellular Component (Fig 2) and Biological Process (Fig 3) with 120, 118 and 200 GO terms obtained for each category respectively. All the EST-contigs showed involvement in starch and sucrose metabolism (Fig 4).

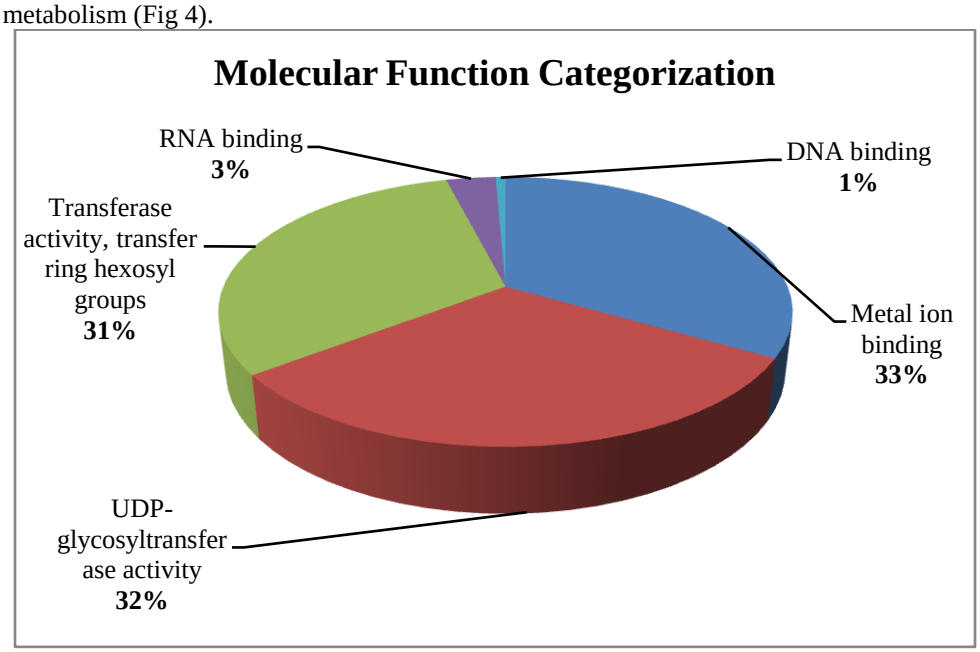

Fig 1: Distribution of GO terms in the Molecular Function category

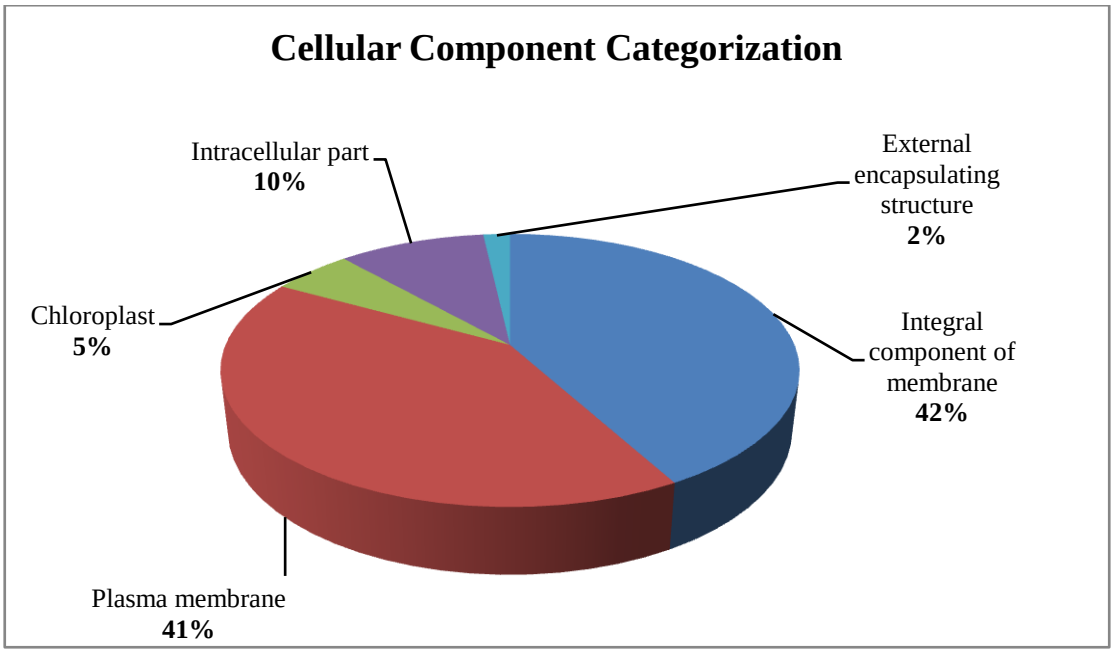

Fig 2: Distribution of GO terms in the Cellular Component category

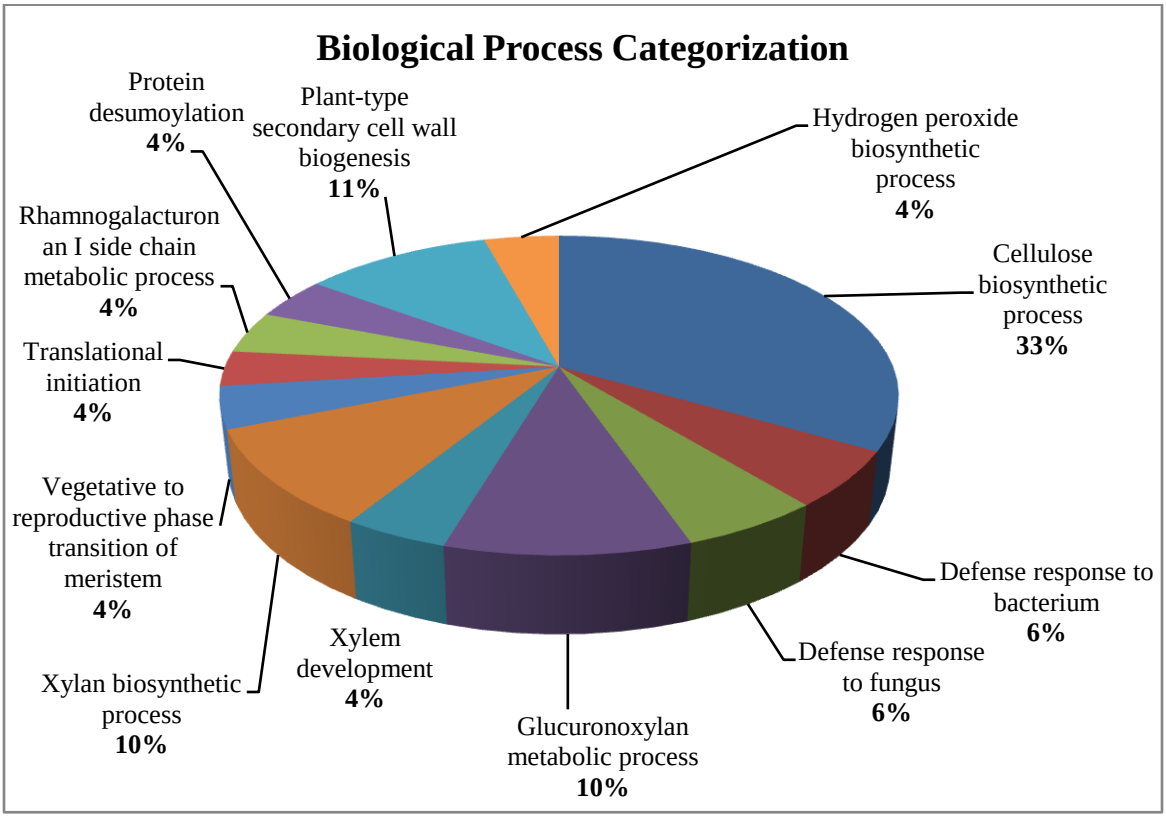

Fig 3: Distribution of GO terms in the Biological Process category

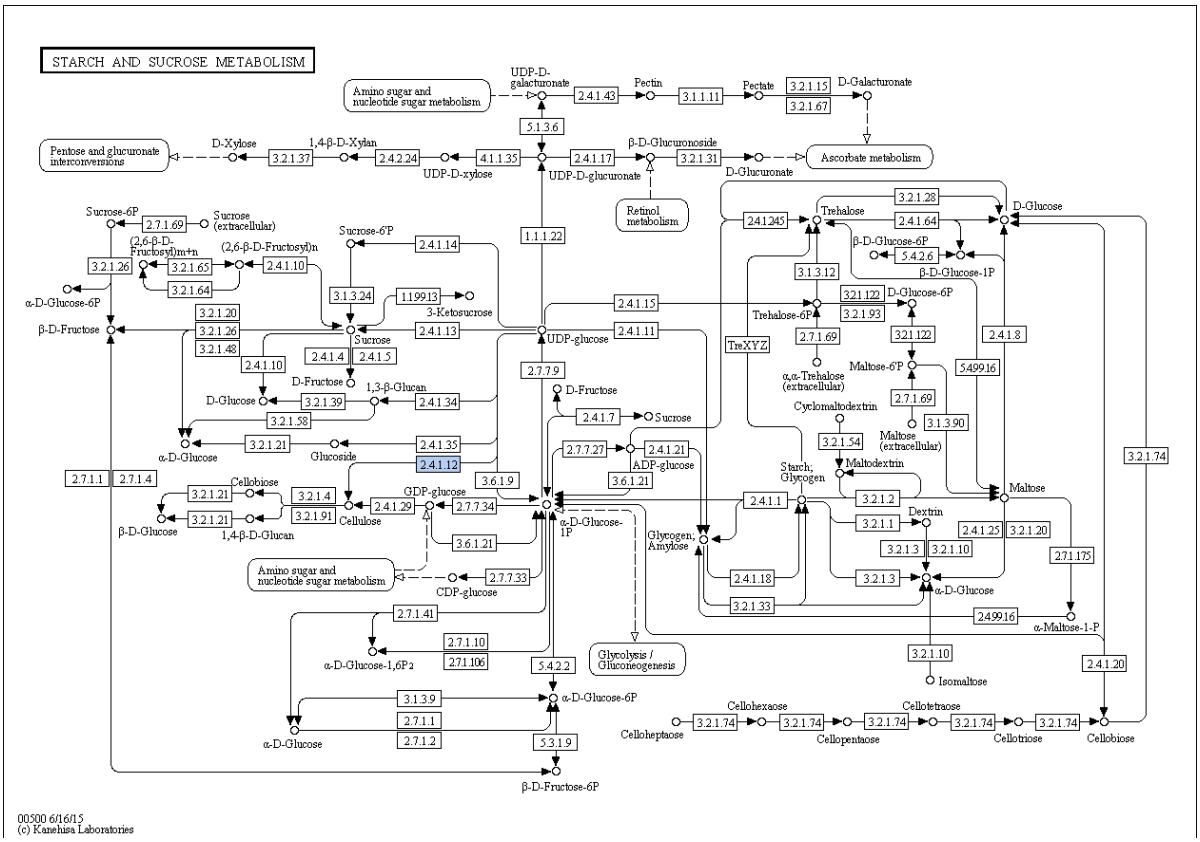

Figure 4: Sucrose and Starch Metabolism Pathway

The EST-contigs classified with respect to different molecular functional activities showed almost equal match (~31%) to metal ion binding, UDP-glycosyltransferase activity and transferase activity (hexosyl groups). Association of metal ion binding activity with cellulose synthase genes is also displayed in *Gossypiumhirsutum*(Jacob-Wilk, et al., 2006). Many cellulose synthase genes showed UDP-forming activity in crops viz., *Arabidopsis*(Arioli, et al., 1998) and Rice (Kikuchi, et al., 2003).

Sugarcane EST-contigs showed cellular localization to integral component of membrane (42%), plasma membrane (41%) and rest few to intracellular part (10%), chloroplast (5%) and external encapsulating structure (2%). Cellulose represents an essential component of plant's integral cell membrane. Cellulose synthesis and transport across the inner membrane is mediated by a complex of the membrane-integrated catalytic cellulose synthase subunits (Morgan, et al., 2013). Fujii*et. al* suggested that the plasma membrane-associated rosette anchors the catalytic unit of cellulose synthesis to form the functional cellulose synthesis machinery (Fujii, et al., 2010).

Out of GO terms pertaining to biological process 33% belong to cellulose biosynthetic process (33%), which clearly showed the association with cellulose synthase. Other related biological functions characterized were secondary cell wall biogenesis, glucuronoxylan metabolic process, xylan biosynthetic process, defense response to fungus and bacterium. Similar functions for cellulose synthase are also shown in *Arabidopsis* for secondary cell wall processes. A complex form of asymmetrical cellular differentiation occurs in *Arabidopsis* seed coat epidermal cells, where it was recently showed that two secondary cell wall processes occur that utilize different cellulose synthase (CESA) proteins (Mendu, et al., 2011; Stork, et al., 2010). Glucuronoxylan metabolic process is associated with cellulose synthase in sugarcane and same association is also observed in *Medicagotruncatula*(Li, et al., 2012).

Conclusion:-

Cellulose synthase enzymes (CESAs) synthesize cellulose and are regarded as a main source for atmospheric carbon in plants because it is the main component of the plant cell wall. The characterization of ESTs for CesA proteins

might provide insights about regulatory processes involved in the specific expression patterns of CesAs genes and consequently cellulose production.

Important abbreviations:-
EST: Expressed Sequence Tag, **GO:** Gene Ontology, **CSCs:** Cellulose Synthase Complexes, **CDS**: Coding Sequence.

Conflict of interest:-
The authors declare that they have no conflict of interest.

Acknowledgement:-

We are thankful to the Indian Council of Agricultural Research (ICAR) for financial assistance under the network project of Centre for Agricultural Bioinformatics Scheme (CABin project code 1004936). We are also thankful to our colleagues from ICAR-Indian Agricultural Statistics Research Institute, New Delhi who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations of this paper.

References:-

1. Arioli, T., *et al.* (1998). Molecular analysis of cellulose biosynthesis in Arabidopsis, *Science*, **279**, 717-720.
2. Bowling, A. and Brown Jr, R. (2008). The cytoplasmic domain of the cellulose-synthesizing complex in vascular plants, *Protoplasma*, **233**, 115-127.
3. Campbell, J.A., *et al.* (1997). A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities, *Biochemical Journal*, **326**, 929.
4. Carson, D. and Botha, F. (2002). Genes expressed in sugarcane maturing internodal tissue, *Plant cell reports*, **20**, 1075-1081.
5. Casu, R.E., *et al.* (2003). Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis, *Plant molecular biology*, **52**, 371-386.
6. Conesa, A. and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics, *International journal of plant genomics*, **2008**.
7. Conesa, A., *et al.* (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics*, **21**, 3674-3676.
8. Cosgrove, D.J. (2005). Growth of the plant cell wall, *Nature reviews molecular cell biology*, **6**, 850-861.
9. Fujii, S., Hayashi, T. and Mizuno, K. (2010). Sucrose synthase is an integral component of the cellulose synthesis machinery, *Plant and cell physiology*, **51**, 294-301.
10. Giddings, T.H., Brower, D.L. and Staehelin, L.A. (1980). Visualization of particle complexes in the plasma membrane of Micrasterias denticulata associated with the formation of cellulose fibrils in primary and secondary cell walls, *The Journal of cell biology*, **84**, 327-339.
11. Hamilton, J.P. and Robin Buell, C. (2012). Advances in plant genome sequencing, *The Plant Journal*, **70**, 177-190.
12. Henry, R.J. (2012). Next-generation sequencing for understanding and accelerating crop domestication, *Briefings in functional genomics*, **11**, 51-56.
13. Huang, X. and Madan, A. (1999). CAP3: A DNA sequence assembly program, *Genome research*, **9**, 868-877.
14. Jacob-Wilk, D., *et al.* (2006). The cotton fiber zinc-binding domain of cellulose synthase A1 from Gossypium hirsutum displays rapid turnover in vitro and in vivo, *Proceedings of the National Academy of Sciences*, **103**, 12191-12196.
15. Kikuchi, S., *et al.* (2003). Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice, *Science*, **301**, 376-379.
16. Li, J., *et al.* (2012). LegumeIP: an integrative database for comparative genomics and transcriptomics of model legumes, *Nucleic acids research*, **40**, D1221-D1229.
17. Masoudi-Nejad, A., *et al.* (2006). EGassembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments, *Nucleic acids research*, **34**, W459-W462.
18. Mendu, V., *et al.* (2011). Cellulose synthesis in two secondary cell wall processes in a single cell type, *Plant signaling & behavior*, **6**, 1638-1643.
19. Morgan, J.L., Strumillo, J. and Zimmer, J. (2013). Crystallographic snapshot of cellulose synthesis and membrane translocation, *Nature*, **493**, 181-186.

20. Nagaraj, S.H., Gasser, R.B. and Ranganathan, S. (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis, *Briefings in bioinformatics*, **8**, 6-21.
21. Olek, A.T., *et al.* (2014). The structure of the catalytic domain of a plant cellulose synthase and its assembly into dimers, *The Plant Cell*, **26**, 2996-3009.
22. Richmond, T. (2000). Higher plant cellulose synthases, *Genome Biol*, **1**, 3001.3001-3001.3006.
23. Salamov, A.A. and Solovyev, V.V. (2000). Ab initio gene finding in Drosophila genomic DNA, *Genome research*, **10**, 516-522.
24. Stork, J., *et al.* (2010). CELLULOSE SYNTHASE9 serves a nonredundant role in secondary cell wall synthesis in Arabidopsis epidermal testa cells, *Plant physiology*, **153**, 580-589.
25. Taylor, N.G. (2008). Cellulose biosynthesis and deposition in higher plants, *New Phytologist*, **178**, 239-252.
26. Yin, Y., Huang, J. and Xu, Y. (2009). The cellulose synthase superfamily in fully sequenced plants and algae, *BMC plant biology*, **9**, 1.