RESEARCH ARTICLE

STRUCTURE BASED COMPARATIVE IN SILICO ANALYSES OF INVERTASE GENE FAMILIES IN SUGARCANE

P.T. Prathima* and S. Arun Meena

Abstract

Sugarcane is an important commercial crop of the tropics which supplies raw material to produce much of the sugar requirement of the world. As the highest biomass producing crop, it has gained focus in the current biofuel research programs across the world. Sucrose is accumulated in sugarcane to a level approaching molar concentrations (0.6-1.0 M) and studies on sucrose metabolism identified several important genes controlling the sugar content. In a crop like sugarcane wherein large amounts of sucrose are accumulated, the genes and their regulation are presumed to be different from those of other crops that do not accumulate high levels of sucrose. Invertases are a group of enzymes that play a key role in the maintenance of sugar levels inside the plant cell. The present study was conducted to identify, characterize and differentiate sugarcane invertases from those of different crops using molecular tools. Invertase nucleotide and protein sequences were retrieved from NCBI-Genbank database and characterized for homology search, multiple sequence alignment for conserved domains, distribution of conserved motifs and phylogenetic relatedness with crop species like sorghum, rice and arabidopsis. Phylogenetic analyses indicated the systematic position of sugarcane invertases in comparison with other invertases of closely related crops. The findings also confirm that the sugarcane invertases are different from those of other crops that are highly conserved and may prove to be useful in planning further studies like genetic transformation or gene manipulations.

Introduction

In sugarcane, sucrose is synthesized in the cytosol and transported and accumulated in the vacuoles to a concentration of about 0.6 - 1.0 M. Sucrose is a disaccharide (alpha-D-Glucopyranosyl beta-D-fructofuronaside) consisting of a molecule of glucose and a molecule of fructose joined through a $\alpha 1–\beta 2$-glycosidic bond. Before sucrose can be used, this bond has to be broken by the action of specific enzymes, which are present in all tissues that require sucrose. Both glucose and fructose are messenger molecules, implicated in several signalling pathways alerting the plant cell to nutritional requirements or environmental constraints. In plants, two enzymes, namely invertase and sucrose synthase are known to cleave sucrose. One process is catalyzed by invertase (sucrose + $H_2O \rightarrow$ glucose and fructose) and the other by sucrose synthase (sucrose + UDP $\rightarrow$ fructose + UDP-glucose). Though both enzymes use sucrose as substrate, their reaction products are quite different (Winter and Huber 2000). While invertases produce free glucose, sucrose synthase produces UDP-glucose. Invertase produces twice as many hexoses as sucrose synthase and thus invertase action amplifies the sucrose signal by producing two "messenger" molecules, which makes invertase to be a bifunctional enzyme, both catalysing

P.T. Prathima* and S. Arun Meena
ICAR-Sugarcane Breeding Institute, Coimbatore 641007, Tamil Nadu, India
*Email : prathimasambandam@gmail.com

sucrose breakdown and amplifying information on carbon status. On the other hand, the reversible cleavage of sucrose via sucrose synthase preserves much of the energy stored in the chemical bond of glucose and fructose.

In general, depending on their optimum pH, solubility and sub-cellular locations, invertases are classified as vacuolar, apoplasmic/cell wall and cytoplasmic isoforms having similar enzymatic and biochemical properties and sharing a high degree of overall sequence homology (Table 1). They are classisfied under β-fructofuranosidases, which have an acidic pH optimum and are also able to accept other fructofuranoside substrates like stachyose (a tetrasaccharide) and raffinose (a trisaccharide) (Fotopoulos 2005). Vacuolar acid invertase, also known as soluble acid invertase, determines the level of sucrose stored in the vacuole and remobilises it for various metabolic processes of the cell. Cell wall invertases, also referred to as extracellular, apoplasmic, periplasmic or free-space invertases, are characterized by a low pH-optimum (pH 3.5–5.0), a high isoelectric point and are ionically bound to the cell wall. The acid invertases are inhibited by heavy metals and are said to have evolved from non-photosynthetic, respiratory, eukaryotic organisms (Chen et al. 2015).

Unlike acid invertases, neutral invertase does not have the characteristics of a typical plant β-fructofuranosidase and sucrose is reported to be its sole substrate; they are not inhibited by heavy metals. The deduced amino acid sequence shares no similarity with sequences of acid invertases, suggesting a prokaryotic (cyanobacteria) origin (Chen et al. 2015). The polypeptide is cysteine-rich and homologous sequences were only detected in the genomes of plants and photosynthetic bacteria. Hence, this protein must have evolved independently of other sucrose-cleaving enzymes. These elusive enzymes have been barely studied in the past because of their low and unstable activity. They were described as cytosolic proteins, and no definite role/ physiological functions were assigned in carbon metabolism and plant development. Recent findings about the presence of functional alkaline/neutral invertase forms inside plant organelles like mitochondria and chloroplast as well as the description of novel physiological roles suggested novel cytosol-organelle metabolic connections and previously unforeseen roles in plant development (Sturm and Tang 1999). The fact that alkaline/neutral invertases are proteins with multiple locations in the plant cell (cytosol, chloroplasts, mitochondria and nuclei) makes them excellent candidates for the coordination of metabolic processes that take place in the different compartments. Sucrose hydrolysis by an alkaline/neutral invertase inside the chloroplast probably participates in controlling chloroplast-cytosolic carbon partitioning (Chen et al. 2015) The molecular mechanisms and signalling pathways by which sucrose hydrolysis inside the chloroplast mediates the regulation of carbon exchange and starch accumulation are still unclear.

In sugarcane, which is the only plant that accumulates large amounts of sucrose, invertases must be occupying a pivotal role. Several studies have proven that any manipulation of invertase activity disrupts the normal functioning and physiology of crop growth. The availability of multiple copies of invertase genes together with various isoforms localized to different compartments makes it one of the complex metabolic networks in sugarcane. This study was formulated to have a basic understanding on the gene structure at the protein level of the invertase family in order to devise strategies for regulating the invertase activity and, in turn, maintain better sucrose levels in the crop.

Table 1. Classification of invertases and their characteristic features

Feature	Alkaline/neutral invertase	Acid invertase	
Localization	Cytoplasm, chloroplast, mitochondria, nuclei	Vacuole	Cell wall
pH	7.0–8.0	~4.5–5.5	~3.5–5.5
pI	Neutral	Neutral	Basic
Solubility	Soluble	Soluble	Insoluble, ionically bound
Sequence motif	>12 predicted conserved motifs	Val residue in the WECVD, Ile residue in the case of sugarcane and other monocots	Pro residue in the WECPD motif
Function	Involved in sugar/ABA signalling	Sugar composition, osmoregulation and cell enlargement, drought stress response	Apoplastic cleavage of sucrose, response to wounding and pathogen infection, regulation of seed and pollen development
Signal peptide	Lacks the N-terminal signal peptide	Often exhibit extra ~20–50 and ~3–10 amino acid residues in their N- and C-terminus	100 amino acid residues comprising a signal peptide and an N-terminal propeptide that may be related to protein folding, targeting, and the regulation of activity
Substrate	Sucrose specific	Also hydrolyse other β-fructose-containing oligosaccharides including raffinose and stachyose other than sucrose	
Classification	Glycoside hydrolase family 100	They are also called β-fructofuranosidases, classified under glycoside hydrolase family 32	
Evolutionary origin	Evolved from prokaryotic cyanobacterial photosynthetic organisms	Evolved from non-photosynthetic, respiratory, aerobic eukaryotic organisms	
Glycosylation	Not glycosylated	Glycosylated	
Inhibitors	Inhibited by tris glucose and fructose	Inhibited by heavy metal ions	

Materials and methods

Database search and sequence retrieval

Invertase gene sequences were obtained from NCBI-GenBank nucleotide and protein databases. Bioinformatics analyses, such as composition, physical and chemical characterisation, and conserved functional domains of the invertase gene family were performed using the Expert Protein Analysis System (ExPASy) and BioEdit tools. Structure and modelling information of sugarcane invertase was used from a previous study (P.T. Prathima unpubl. data) in Deepview Swiss-Pdb Viewer v.4.1 for finding the residues of the conserved WECP(V,I)D domain.

Multiple sequence alignment, motif prediction and phylogenetic analysis

The details of nucleotide sequences of invertases from different crop species and their accession numbers are furnished in Table 2. The protein sequences obtained were grouped together and aligned to compare equivalent residues using the program ClustalW. A global dynamic programming algorithm was used to construct an alignment for full length of the sequences. Bioedit was used to produce the graphical representation of the alignment. The phylogenetic trees were inferred by bootstrap (1000 reiterations), maximum likelihood phylogenic inference and UPGMA methods using Mega6 (Tamura et al. 2013). The conserved motifs were predicted using MEME (multiple em for motif elicitation) web server (Bailey et al. 2015) taking 10 as the number of motifs to be predicted with default parameters.

Results and discussion

Multiple sequence alignment and prediction of conserved motifs

In the present study, computational analysis of available sequences of the three different invertase enzymes of sugarcane was used to identify functionally important residues. Comparison of the sequences of homologous proteins and analysis of large multiple sequence alignments can help to identify sequence, structural conservation and conserved interactions that are crucial for protein stability and function. The multiple sequence alignment of various invertases along with available sugarcane invertases revealed both consensus and unique motifs and sites (Fig. 1). When the acid invertase sequences were aligned, based on the conserved motif WECP(V)D, a clear distinction of cell wall and vacuolar invertases could be made. A search for acid invertases in NCBI brings up both cell wall and vacuolar forms, or a search for cell wall invertases also brings up vacuolar forms. Only upon alignment, the WECPD motif could differentiate them into cell wall and vacuolar isoforms. Within the conserved WECP/VD amino acid motif, all vacuolar invertases posses a valine (Val) residue whereas all cell wall invertases are characterized by a proline (Pro) residue. This conserved single amino acid difference determines the acidic pH-optimum of cell wall invertases and their substrate specificity.

In maize and rice, both forms, i.e. vacuolar acid invertase with Isoleucine (Ile) and Val are found whereas in sugarcane and other monocots, only Ile occurs in place of Val. In the same conserved motif, Ile forms a similar structure as that of Val residue but the significance of this substitution is not yet reported in crops. However, the Val to Ile substitution is a very conservative substitution in that the two amino acids differ by only a methylene group. The Ile side chain is very non-reactive and thus rarely directly involved in protein functions like catalysis, though it can play a role in substrate recognition. In the structure of WECPD, the Pro residue forms a cyclic structure whereas Val residue forms an open structure maintaining a similar structure orientation

Table 2. Details of the various invertase sequences and their accession numbers taken up for the study

Neutral/alkaline invertases		Cell wall invertases		Soluble acid Invertases	
Source	**Accession no.**	**Source**	**Accession no.**	**Source**	**Accession no.**
Saccharum hybrid cultivar FN-41	gb|AGG41113.1	*Saccharum* hybrid cultivar Pindar	gb|AAP59437.1	*Saccharum* hybrid cultivar GT28	gb|AFV94412.1
Saccharum hybrid cultivar GT28	gb|AFV94466.1	*Saccharum* hybrid cultivar GT28	gb|AFV09274.1	*Saccharum* hybrid cultivar GT28	gb|AFV94413.1
Arabidopsis thaliana	ref|NP_197643.1	*Sorghum bicolor*	gb|ABM65156.1	*Saccharum* hybrid cultivar GT28	gb| AFV94414.1
Zea mays	ref|NP_001142296.1	*Zea mays*	ref|NP_001105369.1	*Saccharum* hybrid cultivar Pindar	gb| AAP59436.1
Zea mays	gb|ACG27641.1	*Zea mays*	gb|AAD02511.1	*Saccharum* hybrid cultivar GT28	gb| AFV94467.1
Oryza sativa Japonica Group	dbj|BAD53496.1	*Zea mays*	gb|AAD02263.1	*Saccharum* hybrid cultivar GT28	gb| AFV94472.1
Nicotiana tomentosiformis	ref|XP_009602730.1	*Zea mays*	gb|AAD02510.1	*Saccharum officinarum*	gb| AAC16655.1
Nicotiana tomentosiformis	ref|XP_009595687.1	*Zea mays*	tpg|DAA35504.1	*Saccharum robustum*	gb| AAC16654.1

Table 2. Details of the various invertase sequences and their accession numbers taken up for the study (Cont'd)

Neutral/alkaline invertases		Cell wall invertases		Soluble acid Invertases	
Source	**Accession no.**	**Source**	**Accession no.**	**Source**	**Accession no.**
Medicago truncatula	ref|XP_003630134.1	*Zea mays*	gb|AAF06992.1	*Saccharum* hybrid cultivar GT28	gb| AFV94468.1
Populus trichocarpa	gb|EEE80691.2	*Arabidopsis thaliana*	gb|AAA63802.1	*Saccharum* hybrid cultivar GT28	gb| AFV94469.1
Beta vulgaris	emb|CAD19320.1	*Arabidopsis thaliana*	gb|AED95779.1	*Saccharum* hybrid cultivar GT28	gb| AFV94470.1
		Oryza sativa Japonica Group	ref|NP_001054172.2	*Saccharum* hybrid cultivar GT28	gb| AFV94475.1
		Oryza sativa Japonica Group	dbj|BAF16086.2	*Saccharum* hybrid cultivar GT28	gb| AFV94473.1
		Oryza sativa Japonica Group	dbj|BAF14661.1	*Saccharum* hybrid cultivar GT28	gb| AFV94474.1
		Beta vulgaris	emb|CAC81921.1	*Saccharum* hybrid cultivar GT28	gb| AFV94471.1
		Triticum aestivum	gb|AAC96065.1	*Sorghum bicolor*	gb| AGO03820.1
		Solanum lycopersicum	emb|CAD30649.1	*Zea mays*	ref| NP_001292789.1
		Manihot esculenta	gb|AFR69123.1	*Oryza sativa* Japonica Group	gb| AAK72492.2
				Beta vulgaris	emb|CAD19321.1
				Arabidopsis thaliana	gb| AAA63802.1

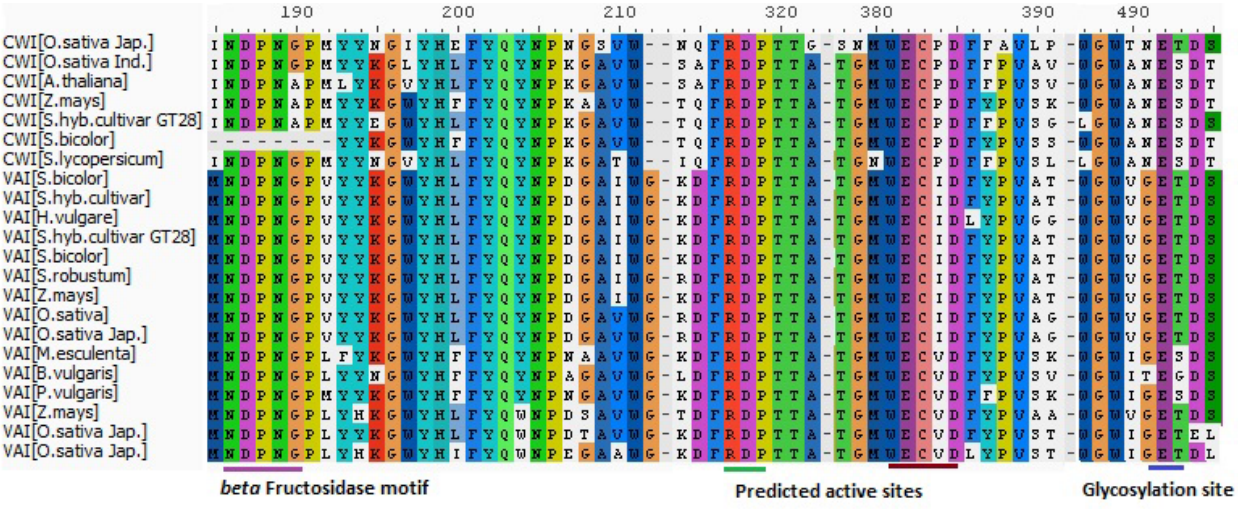

Fig. 1. Protein alignment and domain structure of the various cell wall and vacuolar acid invertase sequences. Three conserved sequence domains-NDPNG (β-fructosidase motif), RDP and WECP(V, I)D-which contain the predicted active sites of the enzyme, are shown in solid lines. Blue line indicates the predicted glycosylation site NES in the case of cell wall invertases, GET in the case of vacuolar acid invertases.

(Fig. 2). The role of this distinct difference was studied by substituting the Pro residue of extracellular invertase CIN1 from *Chenopodium rubrum* with a Val residue by site-directed mutagenesis. The mutated gene was heterologously expressed in an invertase-deficient *Saccharomyces cerevisiae* strain. The single amino-acid difference was shown to be the molecular basis for two enzymatic properties of invertase isoenzymes, for both pH optimum and substrate specificity. A Pro in the WEC-P/V-D box of extracellular invertases determines its more acidic pH optimum and higher cleavage rate of raffinose, compared to vacuolar invertases that have a Val residue at this position (Chen et al. 2015).

All the conserved positions are not related to function but some amino acids tend to have structural roles when conserved (e.g. Trp, Leu, Gly and Cys) while others (mainly polar amino acids or specific types such as Asp, Ser, Cys and His) tend to be part of binding and active sites (Pazos and Bang 2006). Here, the role of conserved residue Ile occurring in WECP/I/D motif and if it is responsible for the structural role of the protein are not known. Also, it would be interesting to examine if these predicted conserved positions can be important in the functional diversity of the proteins.

The three conserved sequence domains (NDPNG, RDP and WECP(V)D) that contain catalytic residues in this module in the cell wall invertases ensure that the catalytic centre is for sucrose degradation. In the conserved motif NDPNG, the DPN is a functional exon, the smallest known in the plant kingdom (Kim et al. 2000), which encodes three amino acids (DPN), which are part of the highly conserved -fructosidase motif NDPN (Goetz and Roitsch 2000). This motif, together with the well-conserved cysteine catalytic site (WECV/ P/I DF), might have an important function in the enzyme. The focus of the present study was to map conserved positions to a representative structure and orthologous sequences of the three different invertases occurring in sugarcane.

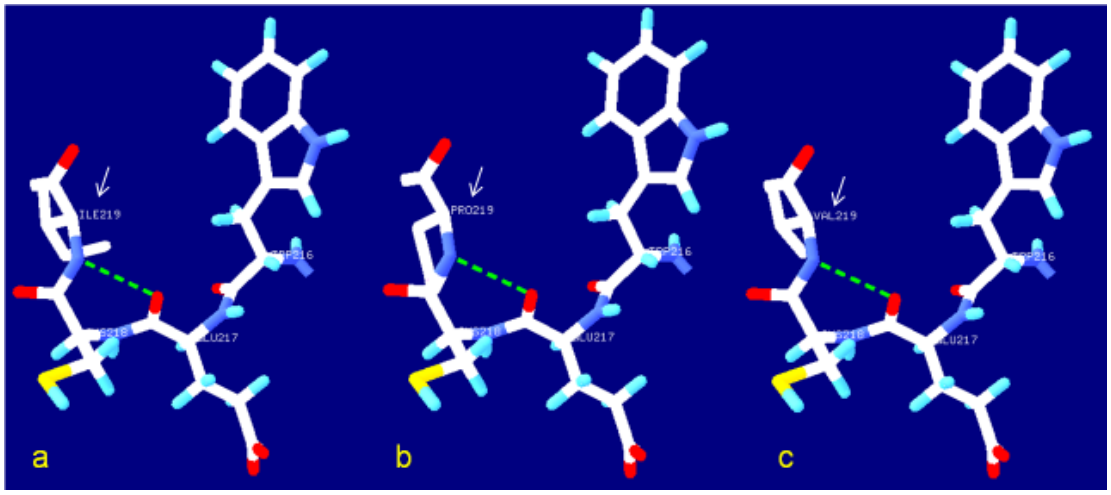

Fig. 2. Close-up view of comparison of orientation of predicted structure of Ile (a), Pro (b) and Val (c) residues in WECP(V,I)D domain of invertase protein. The invertase (vacuolar) sequence from *Saccharum officinarum* was used for modelling (data not shown here). The residues were characterized using DeePView-Swiss-Pdb Viewer V.4.1

The asparagine-X-serine/threonine (NXS/T) motif, where X is any amino acid except Pro, is the consensus motif for N-linked glycosylation. N-linked glycosylation occurs as a post-translational modification and a co-translational process through which carbohydrates (glycans) are added to an asparagine (N) at the consensus motif asparagine-X-serine/threonine (NXS/T), in which X is any amino acid except Pro (Lam et al. 2013). The glycosylation site is NES in the case of cell wall invertase, including sugarcane, and GET in the case of vacuolar acid invertase. Asparagine, when occurring in a particular motif (Asn-X-Ser/Thr) can be *N*-glycosylated (Gavel and von Heijne 1990). Thus, in this context it is impossible to substitute it with any amino acid at all. The uniqueness of glycine also means that it can play a distinct functional role, such as using its backbone (without a side chain) to bind to phosphates (Schulze-Gahmen et al. 1996). This means that if one sees a conserved glycine changing to any other amino acid, the change could have a drastic impact on the function. Only glycines can function to bind to the phosphates of the ATP molecule using their main chains. In this case, the change in NES in cell wall invertase to GET in vacuolar acid invertase remains to be elucidated.

In the case of alkaline invertase, a previous study reported 12 conserved motifs in neutral/alkaline invertases in green plants (Ji et al. 2005). In this study, all 12 of the neutral/alkaline invertases, including sugarcane, were predicted to contain all of the conserved motifs (Fig. 3, numbered lines).

Prediction of conserved motifs and distribution in the protein secondary structure

Identification of phylogenetic motifs helps to understand known functional features, including catalytic site, substrate binding epitopes and protein-protein interfaces. Furthermore, they are strongly conserved among orthologs, indicating their evolutionary importance. A group of evolutionarily related ortholog sequences of invertases were identified and aligned to find the phylogenetically conserved regions of proteins. There are no striking differences found among the conserved motifs in the invertase sequences taken up for the study. However, there are variations found in the motif

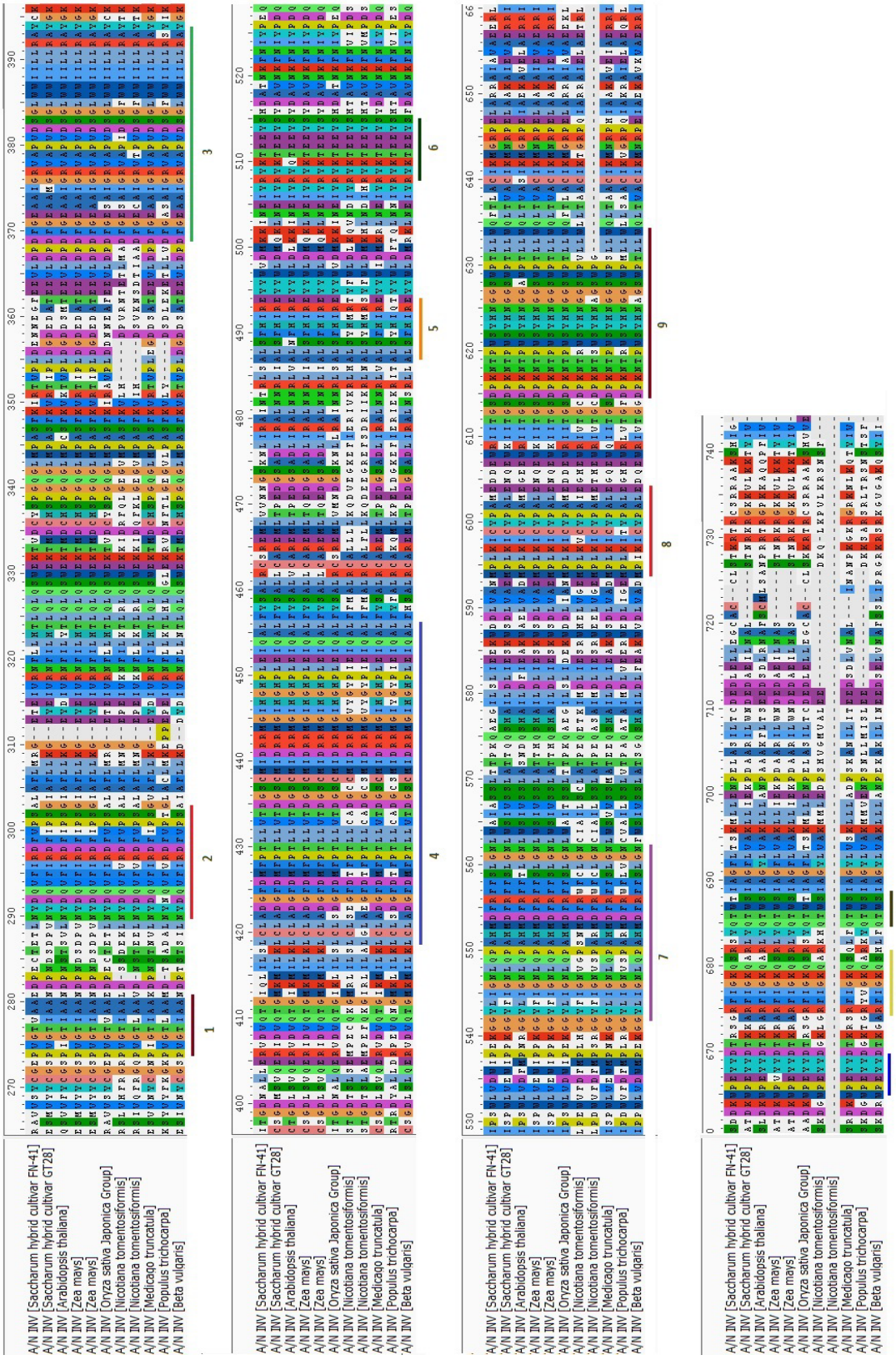

Fig. 3. Alignment of the conserved regions from neutral/alkaline invertases. The numbered lines indicate the 12 well-conserved regions from known neutral/alkaline invertases of selected crop species.

distribution between monocots and dicots crop species studied (Fig. 4). The phylogenetic trees showed closer relationship of sorghum and *Zea* with sugarcane for all the invertases. The vacuolar acid invertase of rice was found to lack three motifs, *Beta vulgaris* two motifs and arabidopsis two motifs when compared to sugarcane. Similarly, in the case of cell wall invertases, one or two motifs were not found in some of the *Zea* sequences, arabidopsis and rice. The only available sequence from sorghum was a partial sequence, which lacked the first motif when compared to sugarcane. The sequences of alkaline invertases were found to be conserved across the monocot and dicot crops chosen and all predicted motifs were found on all sequences.

The evolution of proteins involves mutations of single residues, insertions or deletions, gene duplication or fusion, and exon duplication, deletion or shuffling. Such changes accumulate over time and result in structural differences between the two proteins. These changes affect the proteins, except for the functional sites, which tend to be conserved if the protein retains the same molecular function. Knowledge of the distribution of a specific motif in the secondary structure elements can be useful in predicting the functional relevance. The presence or absence of certain motifs gains significance, if genes are to be used from related/unrelated species for improving traits like sucrose levels or any trait in sugarcane. Also, there are certain invertase inhibiting proteins, wherein they interact with very specific domains and residues of the protein. The associated function of the motifs is to be known before such transgenesis is planned in sugarcane. Most of the identified conserved residues were expected to be critically related to the function of the protein. Further investigation on these functional sites can be used to prioritize targets for manipulating invertases and thereby regulating sucrose accumulation in sugarcane.

References

Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite. Nucleic Acids Res. 1;43(W1):W39-49. doi: 10.1093/nar/gkv416. Epub 2015 May 7.

Chen Z, Gao K, Su X, Rao P, An X (2015) Genome-Wide Identification of the Invertase Gene Family in *Populus*. PLoS ONE 10(9): e0138540. doi:10.1371/journal.pone.0138540.

Fotopoulos V (2005) Plant invertases: structure, function and regulation of a diverse enzyme family. J Biol Res 4: 127 - 137.

Gavel Y, von Heijne G (1990) Cleavage-site motifs in mitochondrial targeting peptides. Protein Eng 4(1):33-7. Pub Med PMID : 2290832.

Goetz M, Roitsch T (2000) Identification of amino acids essential for enzymatic activity of plant invertases. J Plant Physiol 157: 581-585.

Ji X, Van den Ende W, Van Laere A, Cheng S, Bennett J (2005) Structure, evolution, and expression of the two invertase gene families of rice. J Mol Evol 60: 615-634.

Kim J-Y, Mahé A, Guy S, Brangeon J, Roche O, Chourey PS, Prioul J-L (2000) Characterization of two members of the maize gene family, Incw3 and Incw4, encoding cell-wall invertases. Gene 245: 89-102.

Pazos F, Bang JW (2006) Computational prediction of functionally important regions in proteins. Curr Bioinform 1:15-23.

Lam PVN, Goldman R, Karagiannis K, Narsule T, Simonyan V, Soika V, Mazumder R (2013) Structure-based Comparative Analysis and Prediction of N-linked Glycosylation Sites in Evolutionarily Distant Eukaryotes. Genomics Proteomics Bioinform 11(2): 96-104.

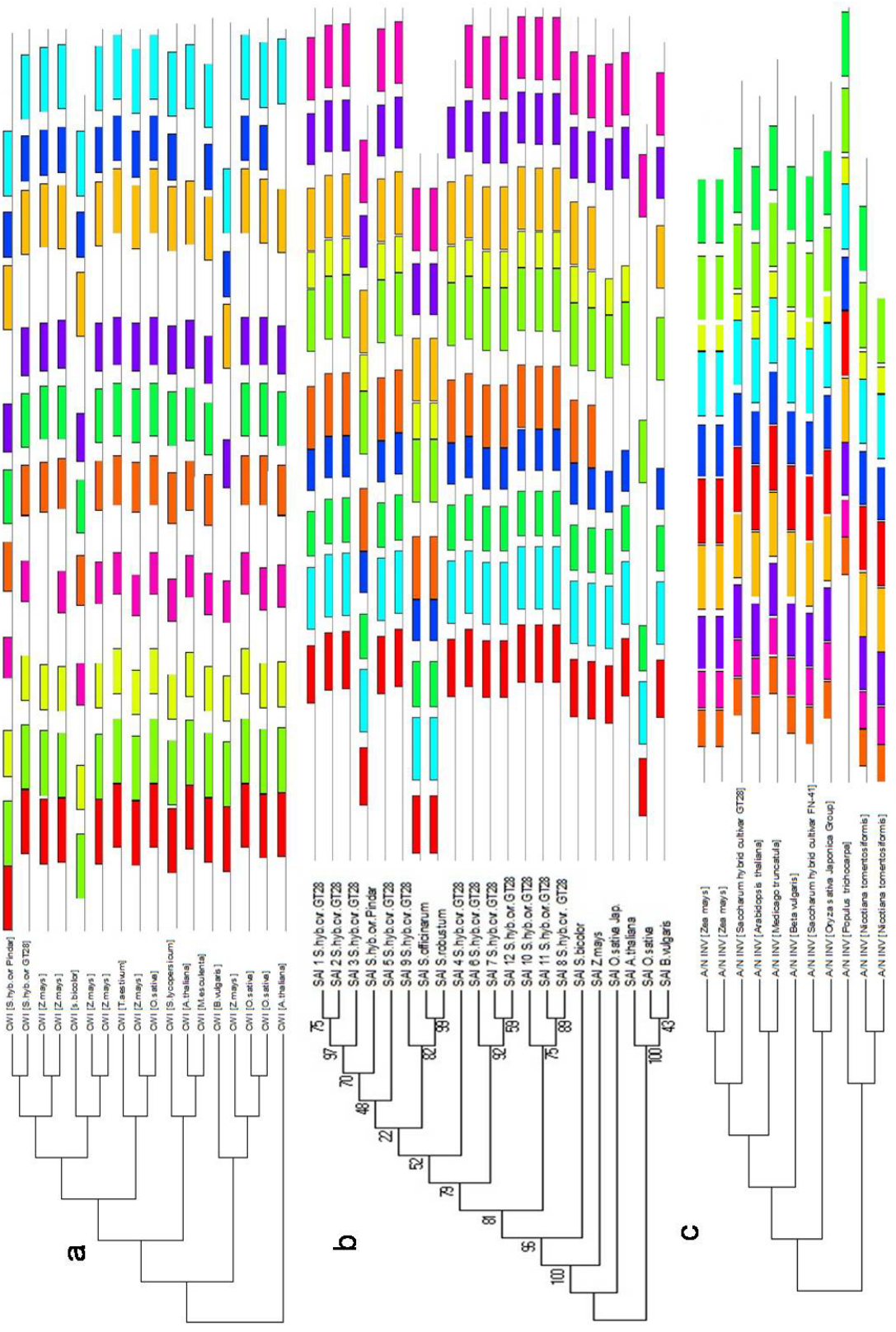

Fig. 4. Phylogenetic tree and motif distribution in sugarcane invertases. Motifs of acid invertases (cell wall (a) vacuolar (b)) and neutral/alkaline invertases (c) were investigated using the MEME web server. Phylogenetic trees were constructed using Mega 6.0. Bootstrap values were generated as a percentage of 1000 repetitions by neighbour joining statistical method.

Schulze-Gahmen U, De Bondt HL, Kim SH (1996) High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design. J Med Chem 8;39(23):4540-6.

Sturm A, Tang G-Q (1999) The sucrose-cleaving enzymes of plants are crucial for development, growth and carbon partitioning. Trends Plant Sci, 4: 401-407.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30: 2725-2729.

Winter H, Huber SC (2000) Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. Crit Rev Biochem Mol Biol 35(4):253-89.