Molecular Characterization and Phylogenetic Analysis of BZIP Protein in Plants

Dhivya Selvaraj[1], Arul Loganathan[2] and Sathishkumar Ramalingam[1]*

[1]Molecular Biology Laboratory, Department of Biotechnology, Bharathiar University, Coimbatore -46, India
[2]Centre for Plant Molecular Biology, Tamilnadu Agriculture University, Coimbatore-46, India

Abstract

BZIP are a class of dimeric sequence specific DNA-binding proteins, is bipartite in structure containing region of enriched basic amino acids which is adjacent to leucine zippers. It is characterized by several leucine residues regularly spaced at seven amino acid intervals, basic region directly contacts with DNA. The leucine zipper mediates heterodimerization and homodimerization of protein monomers through parallel interactions which is unique to eukaryotes. The plant *Arabidopsis thaliana* genome shows 67 BZIP proteins. We have predicted dimeric properties of alpha helical leucine zipper and coiled coil structure of BZIP proteins in plants. In this analysis the length of leucine zippers, placement of asparagines in the hydrophobic interface and presence of interhelical electrostatic interactions were focused. Phylogenetic tree was also constructed by studying evolutionary relationship of BZIB existing among the plants.

Keywords: BZIP; Transcription factor; Dimerization; Leucine zippers; Biophysical properties; Phylogenetic relationship

Introduction

Growth and development of all organisms depends on respective gene expression which is mainly controlled by transcription factors. Transcriptional regulators can be grouped into families of related proteins (Michel et al., 2001). The basic leucine zipper (BZIP) is one among the transcriptional regulatory factors that have been conserved in all eukaryotes. BZIP protein has DNA binding domain consisting of rich regions of basic amino acids that binds to DNA and so called leucine zippers. It consists of several heptad repeats of hydrophobic residues which cause dimerization. BZIP basic region shows a high degree of sequence similarity with *Homo sapiens* and *Arabidopsis thaliana* and contain two invariant residues of Asparagine and Arginine.

BZIP–DNA complex consists of two α helices lying perpendicular to the DNA, associated in a coiled coil structure with basic region contacting a half site in the DNA major groove. The previous study reveals that BZIP structures shows functional variability of conserved residues in DNA recognition (Maria et al., 2003). The genome of *Arabidopsis thaliana* have been sequenced and annotated. Findings suggest that BZIP proteins are important for pathogen defense, light- induced signaling, seed maturation and flower development in plants (Christopher et al., 2004). BZIP proteins form homodimers and heterodimers depending on the amino acid sequence of the leucine zipper (O'Shea et al., 1992). In the previous study of basic region of EmBp-1, eight or ten conserved residues were found in other leucine zipper proteins (Guiltinan et al., 1990).

Leucine zippers of monomeric BZIP have structural repeats of two α helical turns and the repeat is termed as heptad, with each seven positions assigned as *a,b,c,d,e,f* and *g*. The positions *d*, *e* and *g* are near to leucine zipper interface and shows dimerization specificity. Amino acids in *a* and *d* positions are hydrophobic that lies on same side of the helix. These hydrophobic amino acids interact interhelically with hydrophobic amino acids in the same *a* and *d* positions of the second α-helix of the leucine zipper that stabilize the dimerization property of the protein. The Amino acid leucine has better stabilizing property than other amino acids (Michel et

al., 2001) Dimerization specificity is regulated by amino acids in the *a, e* and *g* positions. The charged amino acids present in g and e positions actively involve in the formation of attractive electrostatic interhelical interactions. These interactions are denoted as *g↔e'* where the prime (') indicates a residue on the second α helix of the dimeric leucine zipper. Oppositely charged amino acid interactions promote dimerization specificity where as similarly charged amino acids shows repulsion and thereby inhibiting homodimerization. BZIP plays an important role in Abscisic Acid (ABA) signaling pathways in *Arabidopsis*. Through quantitative RT-PCR, it is analyzed that most of *OsbZIPs* were induced by ABA, ACC and abiotic stress. The RT-PCR reveals that rice BZIP has a positive role in drought tolerance (Lu et al., 2009). Phylogenetic analysis of BZIP protein was done in algae, mosses, ferns, gymnosperms and angiosperms. The result suggests that the ancestor of green plants possess four bZIP genes that actively involved in oxidative stress and also in light-dependent regulations. (Luiz et al., 2008).

In this study, 109 motifs of BZIP in the genome of *O. sativa japonica* were analyzed by comparing with the model plant *Arabidopsis thaliana* and other plants. None of the proteins are homologous to animal BZIP proteins but they have similar amino acids to regulate dimerization specificity. BZIP protein sequences from different plant source like *Phaseolus vulgaris, Capsicum annuum, Nicotiana tabacum, Antirrhinum majus, Hyacinthus orientalis, Malus x domestica, Phaseolus vulgaris, Triticum aestivum, Vitis vinifera, Glycine max, Lycopersicon esculentum, Catharanthus roseus, Spinacia oleracea and Psophocarpus tetragonolobus* have been annotated. This analysis

ACC.NO	BZIP IN PLANTS	NO OF LEUCINE REPEATS	SEQUENCE LENGTH	SECONDARY STRUCTURE		
				HELIX(H) %	EXTENDED STRAND (E) %	RANDOM COIL (C) %
AAK25822.1	*Phaseolus vulgaris*	1	193	53	6	42
AAX20030.1	*Capsicum annuum*	2	286	50	13	33
AAY82589.1	*Nicotiana tabacum*	2	400	39	14	41
AAL27150.1	*Nicotiana tabacum*	2	450	44	6.5	48
AAF06696.1	*Nicotiana tabacum*	2	325	68	7.4	22
CAA74023.1	*Antirrhinum majus*	3	140	73	3	30
CAA74022.1	*Antirrhinum majus*	3	133	65	3	31
AAX20038.1	*Capsicum annuum*	3	170	51	15	29
AAS21020.1	*Hyacinthus orientalis*	3	44	66	0	32
AAX11392.1	*Malus x domestica*	3	322	50	12	35
AAK92215.1	*Nicotiana tabacum*	3	138	78	3	20
AAK92214.1	*Nicotiana tabacum*	3	130	77	2	18
CAA41453.1	*Nicotiana tabacum*	3	401	50	6	43
AAK01953.1	*Phaseolus acutifolius*	3	193	54	7	37
AAK39132.1	*Phaseolus vulgaris*	3	415	40	60	38
AAK39131.1	*Phaseolus vulgaris*	3	397	23	6	70
AAK39130.1	*Phaseolus vulgaris*	3	417	22	7	70
BAD97366.1	*Triticum aestivum*	3	354	42	9	46
BAD97365.1	*Triticum aestivum*	3	150	60	1.5	37
CAB85632.1	*Vitis vinifera*	3	447	34	16	41
AAN03468.1	*Glycine max*	4	166	72	7	19
AAD55394.1	*Lycopersicon esculentum*	4	144	67	2	30
AAK92213.1	*Nicotiana tabacum*	4	170	62	11	23
BAD42432.1	*Psophocarpus tetragonolobus*	4	424	30	5	63
AAK14790.1	*Catharanthus roseus*	5	316	30	5	63
AAT08717.1	*Hyacinthus orientalis*	5	141	46	3.6	50
CAA11499.1	*Spinacia oleracea*	7	422	32	7	60

Table 1: Various plant BZIP proteins of heptad repeats, amino acid length and distribution of secondary structure elements.

Figure 1: Pie chart representing the frequency in all the *a,b,c,e,f* and *g* positions of the leucine zipper for *O.sativa* BZIP proteins. 100% of leucine was observed at *d* position.
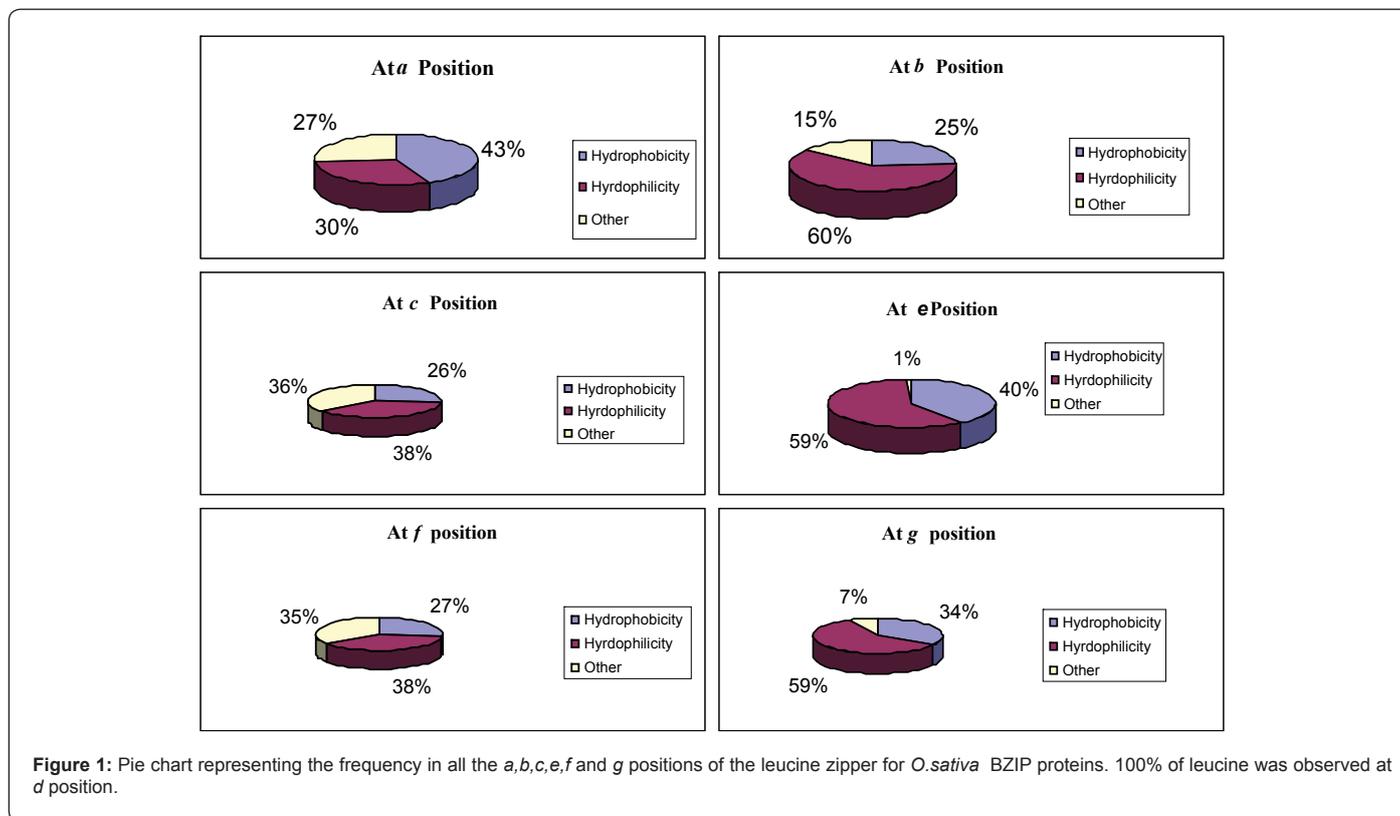
reveals that many O.sativa BZIP proteins have longer leucine zippers like *A.thaliana* and also shows similar dimerization property which is specified by attractive and repulsive g↔e' interactions. Finally evolutionary relationship were analysed in plants by considering the BZIP protein.

Materials and Methods

Data collection

The data set of *Indica Oryza sativa* of BZIP factors were obtained from the database of rice transcription factors (DART), http://drtf.

cbi.pku.edu.cn/. The different plant BZIP factors were obtained from protein sequence database of National centre for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/.

Pattern matching

The Pattern matching program Motifscan 3Dinsight is an integrated database and search tool for structure, function and sequence patterns of biomolecules which is used to identify the BZIP patterns existing between different plant sources. Two types of query of regular expressions were used in this database. The basic region expression [KR]-x (1, 3)-[RKSAQ]-N-{VL}-x-[SAQ] (2)-{L}-[RKTAENQ]-x-R-{S}-[RK] was found to be same in different plants. The numbers of motifs observed in the above plants were presented in Table1.

Secondary structure predictions

Protein secondary structural elements were predicted using a new method called self optimized prediction method (SOPMA), which accurately predicts 69.5% of amino acid for the three state describing the secondary structure (α-helix, b-beta sheet and coil). This tool works on the basis of neural network method (PHD) (Geourjon and Deleage, 1995).

Multiple alignment and phylogenetic tree analysis
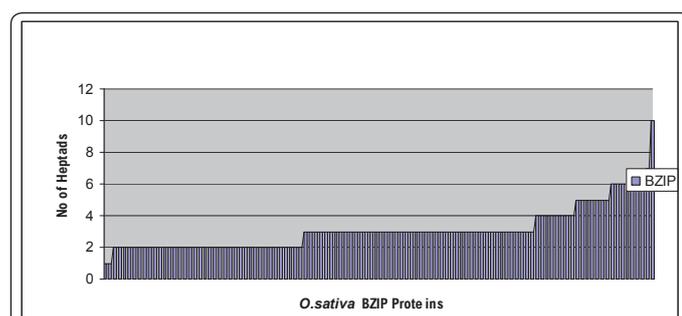
Protein sequences were aligned with Clustal X Program

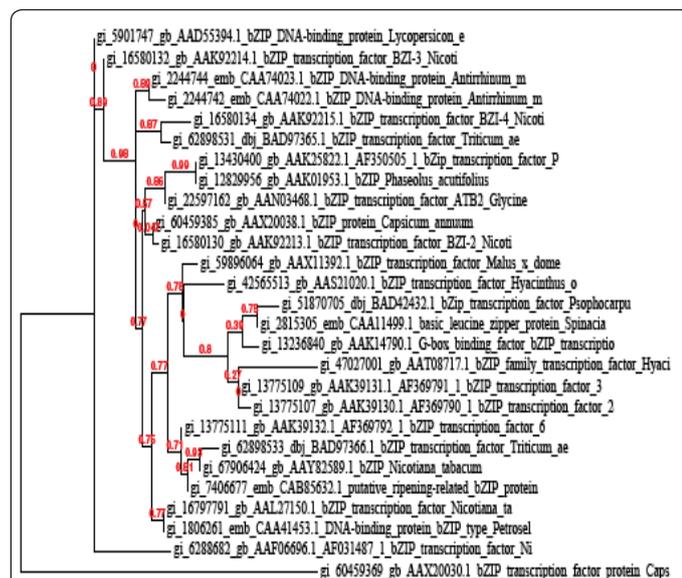**Figure 2:** BZIP distribution and number of heptad repeats in O.sativa.

Figure 3: Phylogenetic relations for BZIP proteins in plants.

(Thombson et al., 1997). Phylogenetic relationship of different Plant BZIP proteins were analyzed by the neighbour – joining method (Saitou and Nei, 1987) using Molecular evolutionary Genetic Analyis tool (MEGA).

Results and Discussion

Amino acid content and leucine zipper length

The interaction of g↔e position is characterized by charged amino acids like Arginine, serine, lysine, Proline and glycine. Pair of proline and glycine indicates the C- terminals. The plants like *Antirrhinum majus, Capsicum annuum, Hyacinthus orientalis, Malus x domestica, Nicotiana tabacum, Phaseolus acutifolius, Phaseolus vulgaris, Triticum aestivum, Vitis vinifera* shows triheptad repeats; *Glycine max, Lycopersicon esculentum Nicotiana tabacum, Psophocarpus tetragonolobus* contains tetra repeats; Catharanthus roseus and Hyacinthus orientalis contains penta heptads; *Spinacia oleracea* shows hepta heptads these were shown in the Table 1. The kind of amino acids found in the *a, d, e* and *g* regions of *O.sativa* are found to be coiled coil arrangements which are known to regulate dimerization stability and specificity , as shown in Figure 1. The number of leucine repeat distribution of BZIP in O.sativa was shown in the Figure 2. Allocation of Amino acid sequences of the Plants BZIP Domains were shown in Figure 4.

Biophysical analysis shows that how the positions of amino acids contribute to dimerization specificity. From the present data, serine(S) in the second position of the heptad interacts with I, N, K and S at the *'a'* position. These data indicates that serine contributes less to dimerization specificity than aliphatic amino acids, polar asparagines or charged lysine residues.

Phylogenetic analysis

The evolutionary relationships between the plants were evaluated by phylogenetic analysis of the aligned amino acids sequence of their BZIP domain. From the analysis BZIP factor 4 of *Nicotiana tabacum* is closely related with BZIP of *Antirrhinum majus* and *Lycopersicon esculentum*. The factor ATB2 BZIP of *Glycine max* is highly similar with BZIP-2 of *Nicotiana tabacum* and *Capsicum annum*. BZIP factor 2 and 3 is found to be same in *Phaseolus vulgaris*. Factor 6 of *Phaseolus vulgaris* is closely related with taxon *Triticum aestivum* and *Vitis vinifera* putative ripening-related BZIP and *Nicotiana tabacum*. The plant *Catharanthus roseus* is not related with the above plants it has BZIP of G BoX Binding protein, shown in the tree Figure 3. Evolutionary relationships of 23 taxa were inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 10.69915308 is shown. Phylogenetic analyses were conducted in MEGA4.

Future Perspectives

Experimental work should be done through *in vivo* and *in vitro* methods which show different binding activities of bHLH and bZIP protein motifs. Yeast one Hybrid provides a satisfactory technique for *in vivo* testing of Protein – DNA interactions like bHLHZ targets with E-box. Through *in vitro* fluorescence anisotropy titrations protein homodimer are to be measured: E-box dissociation constants and circular dichorisim can be used to demonstrate the leucine zipper significance.

Conclusion

In this analysis dimerization partners of different plant BZIP

Acc.No	TNF
AAY82589.1	Nicotiana tabacum
AAK01953.1	Phaseolus acutifolius
CAA11499.1	Spinacia oleracea
CAA74023.1	Antirrhinum majus
CAA74022.1	Antirrhinum majus
AAX20038.1	Capsicum annuum
AAX20030.1	Capsicum annuum
CAA41453.1	Petroselinum crispum
AAS21020.1	Hyacinthus orientalis
AAT08717.1	Hyacinthus orientalis
AAN03468.1	Glycine max
AAL27150.1	Nicotiana tabacum
AAK92215.1	Nicotiana tabacum
AAK92214.1	Nicotiana tabacum
AAK92213.1	Nicotiana tabacum
AAK39132.1	Phaseolus vulgaris
AAK39131.1	Phaseolus vulgaris
AAK39130.1	Phaseolus vulgaris
AAD55394.1	Lycopersicon esculentum
AAX11392.1	Malus x domestica
BAD97366.1	Triticum aestivum
BAD97365.1	Triticum aestivum
CAB85632.1	Vitis vinifera
CAA74022.1	Antirrhinum majus
BAD42432.1	Psophocarpus tetragonolobus
AAK25822.1	Phaseolus vulgaris
AAK14790.1	Catharanthus roseus
AAF06696.1	Nicotiana tabacum

Figure 4: Amino acid sequence of various plants BZIP domains. The leucine zipper region is divided into heptads (*a, b, c, d, e, f, g*) to help visualize the *g↔e*' pairs. Amino acids predicted to regulate dimerization specificity are color coded. If the *g* and *e* positions contain charged amino acids, the heptads from g to the following e were colored. Four colors were used to represent *g↔e*' pairs. Green is used for the attractive basic-acidic pairs (R↔ E and K↔ E), orange is for the attaractive acidic-basic pairs (K↔R, E↔K, D↔K), red is for repulsive acidic pairs (E↔E and E↔D), and blue is for repulsive basic pairs (K↔K and R↔K). The blue color represents the basic and red for acidic. The prolines and glycines are colored red to indicate potential break in α helical structure. The amino acid leucine is represented in yellow at *d* position and serine is represented in blue color in the second position of the heptad which interacts with I, N, K and S at the *e* position. These data indicates that serine contributes less to dimerization specificity than an aliphatic amino acid, polar asparagines or charged lysine residues.

proteins were predicted and also observed for leucine zippers. The result reveals that plants *Phaseolus vulgaris, Capsicum annuum, Lycopersicon esculentum* and *Hyacinthus orientalis*, have three repeats where as *Spinacia oleracea* has seven repeats of leucine. BZIP proteins were identified based on the presence of α – helix breakers, proline, pair of glycines, presence of leucines in the *d* position, presence of charged amino acids in the *g* and *e*. Very few histidine residues are distributed in the plant source and suggest that such signaling system is absent. Phylogenetic analysis reveals that all BZIP proteins use the same amino acids to regulate dimerization specificity. Further experimental studies can be done to prove the dimerization property.

References

1. Christopher DD, Asha A, Vikas R, Barry W, Sjef S, et al. (2004)Dimerization specificity of all 67 B-ZIP motifs in *Arabidopsis thaliana*: a comparison to *Homo sapiens* BZIP motifs. Nucleic Acids Res 32: 3435-3445.

2. Geourjon C, Deleage G (1995) SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11: 681-684.

3. Guiltinan MJ, Marcotte WR, Quatrano RS (1990) A plant leucine zipper protein that recognizes an abscisic acid response element. Science 250: 267-271.

4. Lu G, Gao C, Zheng X, Han B(2009) Identification of OsbZIP72 as a positive regulatorof ABA response and drought tolerance in rice. Planta 229: 605-615.

5. Luiz GC, Diego Mauricio RP, Carlos GS, Renato VS, Bernd MR, et al. (2008) The Role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS One 3: e2944

6. Maria M, Shuman JD, Subastian T, Dauter Z, Johnson PF (2003) Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer – binding protein alpha. J Biol Chem 278: 15178-15184.

7. Michel V, Paulo S, Luis G, Correa G, Fabiana K, Adilson L(2001) Phylogenetic relationships between Arabidopsis and sugarcane bZIP transcriptional regulatory factors. Genet Mol Biol 24: 55-60.

8. O'Shea EK, Rutkowski R, Kim PS (1992) Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. Cell 68: 699-708.

9. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425

10. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higguns DG (1997) The CLUSTAL_X Windows Interface: Flexible stratgies for Multiple Sequene Alignment Aided by Quality Analysis Tools. Nucleic Acids Res 24: 4876-4882.