ibc Interdisciplinary Bio Central

IBC 2012;4:1, 1-7 • DOI: 10.4051/ibc.2012.4.1.0001

Open Access

FULL REPORT

Members of *Ectocarpus siliculosus* F-box Family Are Subjected to Differential Selective Forces

Niaz Mahmood[1], Mahdi Muhammad Moosa[1,†], S. Abdul Matin[2] and Haseena Khan[1,*]

[1]Molecular Biology Laboratory, Department of Biochemistry and Molecular Biology, University of Dhaka, Ramna, Dhaka, Bangladesh
[2]DataSoft Systems Bangladesh Limited, Monipuripara, Dhaka, Bangladesh
[†]Current address: Graduate Program in Biological Sciences, the Scripps Research Institute, La Jolla, California, USA

Subject areas; Bioinformatics/Computational biology/Molecular modeling

Author contribution; N.M. conceptualized the study; N.M. and M.M.M. acquired, analyzed, interpreted the data and drafted the initial version of the manuscript; S.A.M. wrote the perl script; H.K. critically reviewed and revised the manuscript and contributed to the preparation of the final draft; All authors approved the final version of the manuscript.

***Correspondence** and requests for materials should be addressed to H.K. (haseena@univdhaka.edu).

Editor; Hong Gil Nam, POSTECH, Korea

Received January 26, 2012
Accepted February 06, 2012
Published February 07, 2012

Citation; Mahmood, N., et al. Members of *Ectocarpus siliculosus* F-box Family Are Subjected to Differential Selective Forces. IBC 2012, 4:1, 1-7. doi: 10.4051/ibc.2012.4.1.0001

Funding; The authors wish to thank Data Soft Systems Bangladesh Limited for providing computational support during this study.

Competing interest; All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

SYNOPSIS

Background: The F-box proteins represent one of the largest families of proteins in eukaryotes. Apart from being a component of the ubiquitin (Ub)/26 S proteasome pathways, their regulatory roles in other cellular and developmental pathways have also been reported. One interesting feature of the genes encoding the proteins of this particular family is their variable selection patterns across different lineages. This resulted in the presence of lineage specific F-box proteins across different species.

Findings: In this study, 48 non-redundant F-box proteins in *E. siliculosus* have been identified by a homology based approach and classified into three classes based on their variable C-terminal domains. A greater number of the F-box proteins have domains similar to the ones identified in other species. On the other hand, when the proteins having unknown or no C-terminal domain (as predicted by InterProScan) were analyzed, it was found that some of them have the polyglutamine repeats. To gain evolutionary insights on the genes encoding the F-box proteins, their selection patterns were analyzed and a strong positive selection was observed which indicated the adaptation potential of the members of this family. Moreover, four lineage specific F-box genes were found in *E. siliculosus* with no identified homolog in any other species.

Conclusions: This study describes a genome wide *in silico* analysis of the F-box proteins in *E. siliculosus* which sheds light on their evolutionary patterns. The results presented in this study provide a strong foundation to select candidate sequences for future functional analysis.

Subfamily	Total members	Example	Graphical representation	C-terminal domain
FBXL	17	CBN75250		LRR
FBXW	12	CBJ29278		WD40
FBXO	19	CBN75763		None
		CBJ29000		Zn
		CBJ25753		SPRY
		CBN80232		Jmj C

Key Words: *Ectocarpus siliculosus*; heterokont lineage; brown algae; F-box; ubiquitinylation; evolutionary analyses

INTRODUCTION

In order to carry out multifarious cellular functions, living systems produce different proteins; and degradation of the damaged or unneeded proteins is necessary for maintaining a controlled protein turnover rate. In eukaryotes, such degradation of most of the intracellular proteins happens via the ubiquitin (Ub)/26S proteasome pathway[1]. This pathway consists of three steps[2]. First, the activation of ubiquitously present ubiquitin molecule by ubiquitin-activating enzyme (E1); followed by the transfer of the activated ubiquitin to ubiquitin-conjugating enzyme (E2). The final step of substrate recognition and ubiquitinylation is catalyzed by cullin-based E3 ubiquitin ligases, an enzyme complex composed of four subunits: cullin1, Rbx1/Roc1, Skp1, and a member of the F-box family of proteins[3,4]. The F-box proteins are the substrate-recognition components of E3 ubiquitin-protein ligases[5]. They bind to the catalytic core of the enzyme complex by means of their respective F-box motifs interacting with Skp1, and bind substrates through their variable protein–protein interaction domains present at the C-terminus[5,6].

Members of the F-box protein family have a well-conserved motif of around 40-50 amino acids (AAs) at their N-terminal region and this motif was named 'F-box' after its identification in human cyclin F by Bai and co-workers[7]. Apart from functioning as a component of the eukaryotic E3 enzyme complexes, F-box proteins have also been found to play roles in several other cellular contexts like male sex determination in *Caenorhabditis elegans*[8], recycling of the v-SNARE Snc1p in *Saccharomyces cerevisiae*[9], regulation of cell differentiation in mammals[10] and regulation of various developmental processes in plants[11,12].

Brown algae along with other Heterokont have diverged from other phyla comprising multicellular organisms, such as Opisthokonta (metazoan and fungi), Viridiplantae, and the red algal lineage, more than 1,000 million years ago[13]. Consequently they have an independent evolutionary pattern which is quite different from other multicellular eukaryotes[14-17]. This makes them an ideal candidate to study various biological processes from an evolutionary perspective.

In this study, F-box domain containing sequences in the brown algae *Ectocarpus siliculosus* was identified through InterProScan. To explore the evolutionary and selective forces acting on the identified sequences, a phylogenetic tree was constructed and non-synonymous (dS) and synonymous divergence (dN) ratio (dN: dS) was calculated.

RESULTS AND DISSCUTION

Identification and classification of F-box proteins

In eukaryotes, the F-box proteins represent a large protein family with huge expansion in their numbers in nematodes and plant species[18,19]. We identified forty eight F-box proteins in *E. siliculosus* through a homology based approach (Additional File 1 and 2). This accounts for 0.3% of the total number of proteins in *E. siliculosus*, a percentage similar to that of other eukaryotes like animals and fungus but much less than nematode and land plants (Table 1).

By following the nomenclature system described by Jin and colleagues[20], the identified proteins were classified into three groups depending on the domains present in their C-termini: FBXW containing WD40 repeats, FBXL containing LRR domains, and FBXO with other or no predictable domains (Table 2). A greater percentage of the F-box proteins (60%) in *E. siliculosus* belong to either FBXL or FBXW subfamily. Similar predominance of the members of these two subfamilies is also observed in yeast and mammalian F-box proteins[7,21,22]. As many as seventeen F-box proteins in *E. siliculosus* had the C-terminal LRR domain for protein-protein interactions. One protein (GenBank Accession number CBJ27973) had both the LRR and Zn finger domains at the C-terminus. This protein was categorized as FBXL in this study.

The leucine-rich repeat (LRR) is typically composed of around 20-30 amino acids with a characteristic repetitive sequence pattern rich in leucine and it is widely distributed in the primary structure of thousands of protein sequences in all life forms,

Table 2. Classification of *E. siliculosus* F-box proteins

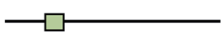
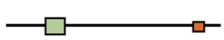
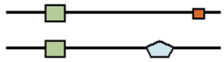
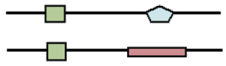
Subfamily	Total members	Example	Graphical representation	C-terminal domain
FBXL	17	CBN75250		LRR
FBXW	12	CBJ29278		WD40
FBXO	19	CBN75763		None
		CBJ29000		Zn
		CBJ25753		SPRY
		CBN80232		Jmj C

Table 1. Distribution of F-box proteins across different species

Species	*E. siliculosus*	Human[a)]	Mouse[a)]	Drosophila[a)]	Yeast[a)]	*C. elegans*[b)]	Arabidopsis[b)]	Rice[c)]
F-box protein	48	68	74	22	11	520	701	687
Total proteins	16,256	25,000	25,000	13,600	6,000	25,010	25,498	37,544
Percentage of F-box	0.3%	0.28%	0.3%	0.16%	0.18%	2.07%	2.74%	1.82%

Data source[a)]: [20], Data source[b)]: [18], Data source[c)]: [23].

from viruses to eukaryotes[24]. Multiple LRRs assemble to form an arched docking structure[25]. Apart from their presence in the F-box proteins, the LRRs are widely distributed in *E. siliculosus* with presence in more than a hundred other proteins having a wide range of functions (data not shown). In *E. siliculosus*, twelve out of the forty eight F-box proteins had c-terminal WD-40 repeats. These WD-40 repeats are short structural motif of approximately 40 amino acids that often terminate in a tryptophan-aspartic acid (W-D) dipeptide[26]. Typically 4-16 units of WD-40 repeats form circular beta-propeller structures which serve as rigid scaffolds for protein interactions[27].

Seven out of the nineteen FBXO proteins had shown the presence of known functional domain at their C-termini. One (CBN80232) of them had the Jumonji C (Jmj C) domain, predicted to be a metal-binding site[28]. Studies on various plant and metazoan lineages revealed the presence of a conserved Jmj C domain in a large group of histone demethylase enzymes that play role in demethylation of damaged DNA[29,30]. DNA-binding domains such as zinc finger and C2H2 zinc finger were also found in *E. siliculosus* F-box proteins (CBJ29000, CBJ49180, CBJ27973, and CBN78274), which may be directly or indirectly involved in transcriptional regulation. These domains were also observed in rice F-box proteins[23]. Two proteins (CBJ25753, CBJ29317) had the 'SPla/RYanodine receptor (SPRY)' domain which typically functions as Ca(2+)-release channel[31]. In addition, two *E. Siliculosus* FBXO proteins (CBJ26383, CBJ33554) contained domains of unknown function (DUF525). The presence of such domains with no known function is also common in F-box proteins from other species[23]. Interestingly, none of the *E. Siliculosus* FBXO proteins had kelch repeat in their C-terminus. Although absent in prokaryotes, kelch repeat containing F-box proteins are widespread in plants and occur rarely in non-plant eukaryotes[32]. The absence of kelch repeat in this heterokont further verifies the fact that such repeats are unique for plants.

A large percentage (21%) of the *E. siliculosus* F-box proteins do not harbor any known functional domain other than the F box. In order to identify the unknown putative conserved motifs within these type of FBXO proteins (without considering the F-box region), MEME Suite[33,34] was used and the consensus sequences were graphically visualized by the WebLogo tool[35]. Four statistically significant motifs were identified by this process (shown in Supplementary Figure 1, Additional File 3). Each of these motifs was more than five amino acids (AAs) in length and was conserved in at least three of the predicted FBXO proteins. Motif 1 with polyglutamine repeat (polyQ) tract was present in four proteins - CBJ48573, CBN74590, CBN77748 and CBN74582. This motif was also found in four other F-box family members (CBN78274 with c-terminal Zinc finger domain, CBN79186 with WD repeats, CBJ48439 with LRRs and CBJ29317 with SPla/RYanodine receptor SPRY) of *E. siliculosus*. The presence of the polyQ tract in proteins from all the three subfamilies indicates that they may have specific biological role in *E. siliculosus*. To our knowledge the role of the polyQ tract in *E. siliculosus* has not yet been defined experimentally. Nevertheless, such repeats are found in various types of proteins including transcription factors and co-factors; and the products of triplet-repeat disease genes in humans[36]. PolyQ tracts have also been found to play significant roles in circadian regulation in *Arabidopsis thaliana* and *Drosophila melanogaster*[37,38]. We found that the polyQ repeats in *E. siliculosus* are polymorphic (data not shown). One previous study revealed that polymorphism in such repeats may have role in morphological evolution which ultimately resulted in incipient speciation[39]. When Motif 2 (present in CBN74590, CBN 74582, and CBN79375), motif 3 (in CBN74590, CBN74582, and CBN75763) and motif 4 (in CBN74590, CBN74582, and CBJ48573) were screened, they were neither found in other members in *E. siliculosus* nor in any proteins with known functions deposited in the publicly available protein databases.

For the remaining four proteins (CBJ34162, CBJ25451, CBJ29976, and CBN75917) in the 'unknown' group, no consensus C-terminal domains were detected, suggesting they are either improperly annotated, are pseudogenes, or use novel domains to interact with their respective targets.

The diversity in C-terminal interaction domains in F-box proteins suggests that F-box proteins use a broad palette of mechanisms for target recognition. Some exploit protein-interaction domains common in F-box proteins from yeast and animals (e.g., LRR, WD-40), some are DNA binding domains, some are common in plant F-box proteins (e.g., Zinc finger domain, Jmj C domain); whereas others appear to employ domains unique to *E. siliculosus* F-box proteins, such as the SPRY domain. Such diversity indicates that the recognition sites within the targets are heterogeneous.

Sub-cellular localization

E. siliculosus possesses a complex chloroplast acquired through a secondary endosymbiosis event[40]. So they have different localization signals which cannot be accurately predicted by the conventional algorithms. To address this issue, heterokont specific sub-cellular localization prediction tool known as HECTAR was used during this study. Out of the 48 proteins, 46 had no N-terminal target peptide as predicted by HECTAR. Only one protein (CBJ26383) with a signal peptide and one (CBN79186) with a mitochondrial transit peptide was predicted.

Evolutionary analyses

Phylogenetic tree, constructed by MEGA 4.1 using the alignments of full-length F-box protein sequences, clustered the

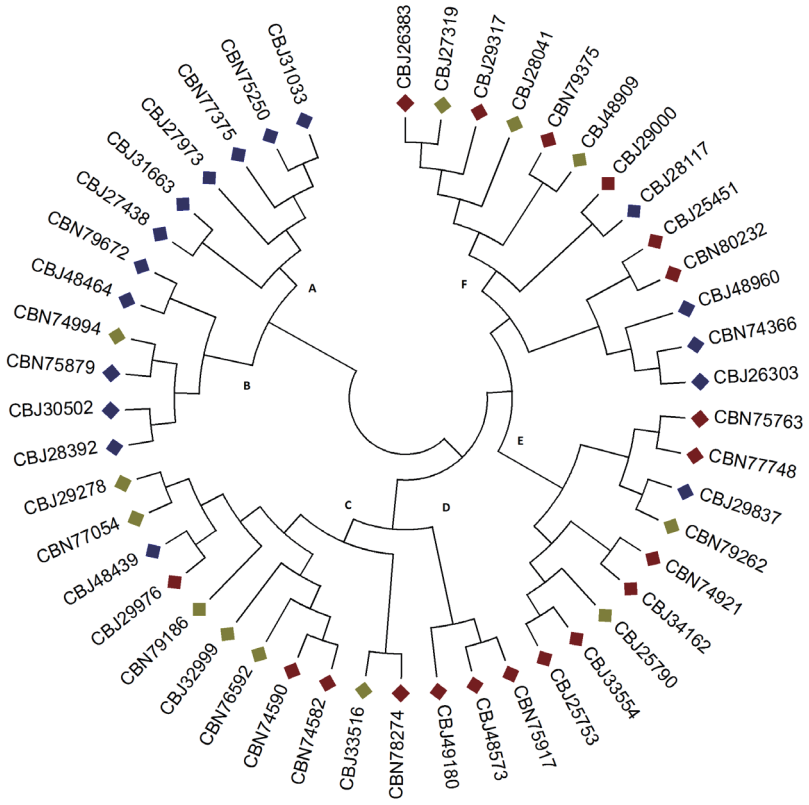


Figure 1. Phylogenetic tree of the *E. siliculosus* F-box proteins. Multiple alignments of the full length F-box protein sequences were done by Clustal X version 2.0[41] and phylogenetic tree was generated by MEGA 4.1 using neighbor-joining method. The proteins from FBXW, FBXL and FBXO families are shown in yellow, blue and red, respectively.

members into six different groups (A-F) as shown in Figure 1. Like other model species, the phylogenetic tree of F-box sequences reflected a striking clustering of the C-terminal domains. F-box proteins of the same subfamily were clustered together in three out of the six phylogenetic groups. Groups A and B mainly contained the members of FBXL subfamily, while group D contained the members of FBXO subfamily. This correlation further suggests about co-evolution of the F-box motif with the target-interaction domain. However, in the other three groups, there was no clear division of the subfamilies; members of different subfamily clustered together in the same group. This can be explained by the fact that all these proteins have similar F-box domains; and therefore clustered in a group irrespective of their C-terminal domains. Phylogenetic trees of F-box proteins of other eukaryotes also showed similar phenomena[23]. Such clustering with similar F-box domains but dissimilar C-terminal domains further supports the idea that domain shuffling has contributed to the expansion of F-box protein diversity[42]. In addition, we also constructed a comparative phylogenetic tree (shown in Supplementary Figure 2, Additional File 3) using the alignments of full-length *E. siliculosus* F-box protein sequences along with several F-box proteins with known function in *A. thaliana* and *Oryza sativa* and found that most members of the same subfamily were clustered together. For example, FBXL protein of *A. thaliana* (NP_569047) clustered with the members of same subfamily in *E. siliculosus*. Similarly, FBXO proteins were clustered with the members of the same subfamily in *O. sativa* (AAP52051) and *A. thaliana* (NP_569047). However, the clustering of F-box proteins with dissimilar C-terminal domains was also found here.

To analyze the synonymous and non-synonymous substitution pattern, nucleotide sequences of the F-box proteins were used. The observed dN:dS ratio suggests positive selection of the F-box family of *E. siliculosus* (Table 3).

The codon based Z test indicated positive selection (data not shown) for most of the pairwise comparisons of the F-box genes.

Based on evolutionary stability, the F-box genes were further divided into two categories: stable genes with clear well-conserved homologs and unstable genes without clear homologs

Table 3. Synonymous divergence (dS), non-synonymous divergence (dN) ratios of the F-box genes

N	S	dN (mean)	dS (mean)	dN/dS
486.628	210.242	2.151	1.232	1.746

that is undergoing rapid birth–death evolution[18]. In this study, we termed a gene as unstable if no specific/significant hits were found in protein blast with NCBI non redundant (nr) protein database (result in Supplementary Table 1, Additional File 3) using their protein sequences. Genes with significant protein homologue in other species are termed as stable genes.

Out of the 48 F-box proteins, four (CBN75763, CBJ25451, CBN78274 and CBN77748) were found to be unstable as they do not have any clear homologue (Supplementary Table 1, Additional File 3). Interestingly, proteins encoded by the unstable genes either have no domains or have 'domains of unknown function' at their C-terminal end.

Le Bail and colleagues[17] stated that brown algae share several obvious features with photosynthetic organisms and some features with the metazoans. Similar pattern was also observed in F-box family members of *E. siliculosus*; 36% of the stable F-box genes had top blast hits with non-photosynthetic organisms, while 64% had top homology with photosynthetic organisms. In non-photosynthetic organisms, the homologs of the *E. siliculosus* F-box genes were distributed across diverse taxa ranging from bacteria to mammals. On the other hand, the photosynthetic organisms with *E. siliculosus* F-box homologs were mainly the other members of heterokont family. Among land plants, top blast hits with each of the two members of the moss lineage -*Selaginella moellendorffi* and *Physcomitrella patens* were found; while each of *A. thaliana*, *Vitis vinifera* and *Ricinus communis* had a single top hit with *E. siliculosus* F-box.

However, the dN: dS ratio was 1.8 and 1.5 for the stable F-box genes that matched with photosynthetic and non-photosynthetic organisms respectively. The result indicated that Darwinian selection pressure was higher on the genes having homolog in photosynthetic organisms than those with non-photosynthetic homologs.

The distant phylogenetic relationship between brown algae and other eukaryotes raises the possibility that they have developed distinct cellular mechanisms to achieve multicellular development. Till now, the cellular mechanisms that govern the developmental patterning in *E. siliculosus* are poorly characterized. Studies on photosynthetic land plants revealed the involvement of F-box proteins in regulating various developmental processes like photomorphogenesis and circadian clock regulation[23]. Being a photosynthetic heterokont, there is a strong possibility that the *E. siliculosus* F-box proteins might have similar roles in its developmental processes.

CONCLUSION AND PROSPECTS

To our knowledge, none of the *E. siliculosus* F-box proteins have been experimentally characterized. Apart from revealing patterns of evolution of F-box family, this manuscript provides a solid basis for selecting promising candidates for future reverse genetic and functional characterization of this family of proteins in *E. siliculosus*.

MATERIALS AND METHODS

Retrieval of F-box proteins

The *E. siliculosus* protein sequences were downloaded from the publicly available database of Bioinformatics Gent (https://bioinformatics.psb.ugent.be/gdb/ectocarpus/).Then domain annotation of all the 16,256 proteins of *E. siliculosus* was done by the standalone version of InterProScan[43], and F-box proteins were screened by searching for the domains IPR001810, PTHR14289, PTHR22844, PTHR23123, PTHR23125 and SSF81383 using an in-house perl script (Additional File : 'protein_search.pl'). The identified sequences were compiled together and made free of redundancy manually. Then the GenBank identification number for each of the proteins was retrieved from the NCBI database. The subfamilies within the F-box protein families were classified according to the domains present in the C-terminus which generally consists of one or more highly variable protein–protein interaction domains like leucine rich repeat (LRR), kelch repeat, WD40 repeat and many other domains that are known to interact with specific targets[20]. Putative unknown conserved motifs were identified using MEME (Multiple Expectation Maximization for Motif Elicitation)[33,34]. A limit of 20 motifs with other options set to default values was specified. Sub-cellular localization was predicted by HECTAR[44].

Phylogenetic tree construction

Phylogenetic tree was constructed using the neighbor-joining method[45] with bootstrap multiple alignment resampling set at 10,000 using Molecular Evolutionary Genetics Analysis (MEGA) software version 4.1[46]. Uniform rates among sites and pairwise deletion of gaps was assumed for the analysis.

Substitution pattern analysis

Synonymous and non-synonymous substitution pattern was determined using the modified Nei-Gojobori[47] method with the Jukes-Cantor correction as implemented in MEGA4.1[46].

ACKNOWLEDGEMENTS

The authors wish to thank DataSoft Systems Bangladesh Limited for providing computational support during this study.

REFERENCES

1. Smalle, J., and Vierstra, R.D. (2004). The ubiquitin 26S proteasome proteolytic pathway. *Annu Rev Plant Biol* 55, 555-590.

2. Stone, S.L., and Callis, J. (2007). Ubiquitin ligases mediate growth and development by promoting protein death. *Curr Opin Plant Biol* 10, 624-632.
3. Feldman, R., Correll, C.C., Kaplan, K.B., and Deshaies, R.J. (1997). A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* 91, 221-230.
4. Cardozo, T., and Pagano, M. (2004). The SCF ubiquitin ligase: insights into a molecular machine. *Nat Rev Mol Cell Biol* 5, 739-751.
5. Schulman, B.A., Carrano, A.C., Jeffrey, P.D., Bowen, Z., Kinnucan, E.R.E., Finnin, M.S., Elledge, S.J., Harper, J.W., Pagano, M., and Pavletich, N.P. (2000). Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex. *Nature* 408, 381-386.
6. Cao, P.R., Kim, H.J., and Lecker, S.H. (2005). Ubiquitin-protein ligases in muscle wasting. *The Int J Biochem Cell Biol* 37, 2088-2097.
7. Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W., and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* 86, 263-274.
8. Clifford, R., Lee, M.H., Nayak, S., Ohmachi, M., Giorgini, F., and Schedl, T. (2000). FOG-2, a novel F-box containing protein, associates with the GLD-1 RNA binding protein and directs male sex determination in the *C. elegans* hermaphrodite germline. *Development* 127, 5265-5276.
9. Galan, J.M., Wiederkehr, A., Seol, J.H., Haguenauer-Tsapis, R., Deshaies, R.J., Riezman, H., and Peter, M. (2001). Skp1p and the F-box protein Rcy1p form a non-SCF complex involved in recycling of the SNARE Snc1p in yeast. *Mol Cell Biol* 21, 3105-3117.
10. Smaldone, S., Laub, F., Else, C., Dragomir, C., and Ramirez, F. (2004). Identification of MoKA, a novel F-box protein that modulates Kruppel-like transcription factor 7 activity. *Mol Cell Biol* 24, 1058-1069.
11. Levin, J.Z., and Meyerowitz, E.M. (1995). UFO: an *Arabidopsis* gene involved in both floral meristem and floral organ development. *Plant Cell* 7, 529-548.
12. Samach, A., Klenz, J.E., Kohalmi, S.E., Risseeuw, E., Haughn, G.W., and Crosby, W.L. (1999). The UNUSUAL FLORAL ORGANS gene of *Arabidopsis thaliana* is an F box protein required for normal patterning and growth in the floral meristem. *Plant J* 20, 433-445.
13. Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol* 21, 809-818.
14. Lau, S., Jurgens, G., and De Smet, I. (2008). The evolving complexity of the auxin pathway. *Plant Cell* 20, 1738-1746.
15. Baldauf, S.L. (2008). An overview of the phylogeny and diversity of eukaryotes. *J Syst Evol* 46, 263–273.
16. Simon, D., and Sylvie, R. (2011). Microarray estimation of genomic inter-strain variability in the genus Ectocarpus (Phaeophyceae). *BMC Mol Biol* 12, 1-12.
17. Le Bail, A., Billoud, B., Kowalczyk, N., Kowalczyk, M., Gicquel, M., Le Panse, S., Stewart, S., Scornet, D., Cock, J.M., and Ljung, K. (2010). Auxin metabolism and function in the multicellular brown alga *Ectocarpus siliculosus*. *Plant Physiol* 153, 128-144.
18. Thomas, J.H. (2006). Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. *Genome Res* 16, 1017-1030.
19. Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., and Hinds, D.A. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317, 338-342.
20. Jin, J., Cardozo, T., Lovering, R.C., Elledge, S.J., Pagano, M., and Harper, J.W. (2004). Systematic analysis and nomenclature of mammalian F-box proteins. *Gene Dev* 18, 2573-2580.
21. Cenciarelli, C., Chiaur, D., Guardavaccaro, D., Parks, W., Vidal, M., and Pagano, M. (1999). Identification of a family of human F-box proteins. *Curr Biol* 9, 1177-1179, S1171-S1173.
22. Winston, J.T., Koepp, D.M., Zhu, C., Elledge, S.J., and Harper, J.W. (1999). A family of mammalian F-box proteins. *Curr Biol* 9, 1180-1182.
23. Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., Kapoor, S., Tyagi, A., and Khurana, J. (2007). F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol* 143, 1467-1483.
24. Bella, J., Hindle, K., McEwan, P., and Lovell, S. (2008). The leucine-rich repeat structure. *Cell Mol Life Sci* 65, 2307-2333.
25. Andrade, M.A., Perez-Iratxeta, C., and Ponting, C.P. (2001). Protein repeats: structures, functions, and evolution. *J Struct Biol* 134, 117-131.
26. Neer, E.J., Schmidt, C.J., Nambudripad, R., and Smith, T.F. (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature* 371, 297-300.
27. Smith, T.F., Gaitatzes, C., Saxena, K., and Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* 24, 181-185.
28. Clissold, P.M., and Ponting, C.P. (2001). JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2 [beta]. *Trends Biochem Sci* 26, 7-9.
29. Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 439, 811-816.
30. Lu, F., Li, G., Cui, X., Liu, C., Wang, X. J., and Cao, X. (2008). Comparative Analysis of JmjC Domain-containing Proteins Reveals the Potential Histone Demethylases in *Arabidopsis* and Rice. *J Integr Plant Biol* 50, 886-896.
31. Ponting, C., Schultz, J., and Bork, P. (1997). SPRY domains in ryanodine receptors (Ca2+-release channels). *Trends Biochem Sci* 22, 193-194.
32. Schumann, N., Navarro-Quezada, A., Ullrich, K., Kuhl, C., and Quint, M. (2011). Molecular Evolution and Selection Patterns of Plant F-Box Proteins with C-Terminal Kelch Repeats. *Plant Physiol* 155, 835-850.
33. Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2, 28-36.
34. Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34, W369-373.
35. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res* 14, 1188-1190.
36. Imafuku, I., Waragai, M., Takeuchi, S., Kanazawa, I., Kawabata, M., Mouradian, M.M., and Okazawa, H. (1998). Polar amino acid-rich sequences bind to polyglutamine tracts. *Biochem Biophys Res Comm* 253, 16-20.
37. Tajima, T., Oda, A., Nakagawa, M., Kamada, H., and Mizoguchi, T. (2007). Natural variation of polyglutamine repeats of a circadian clock gene

ELF3 in Arabidopsis. *Plant Biotechnol* 2, 237-240.
38. Saleem, Q., Anand, A., Jain, S., and Brahmachari, S.K. (2001). The polyglutamine motif is highly conserved at the Clock locus in various organisms and is not polymorphic in humans. *Hum Genet* 109, 136-142.
39. Lindqvist, C., Laakkonen, L., and Albert, V. (2007). Polyglutamine variation in a flowering time protein correlates with island age in a Hawaiian plant radiation. *BMC Evol Biol* 7, 105.
40. Cavalier-Smith, T. (2003). Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Phil Trans R Soc B* 358, 109.
41. Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
42. Morgenstern, B., and Atchley, W.R. (1999). Evolution of bHLH transcription factors: modular evolution by domain shuffling? *Mol Biol Evol* 16, 1654.
43. Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res* 33, W116-120.
44. Bernhard, G., Yann, G., and Mark, C.J. (2008). HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* 9, 393.
45. Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425.
46. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24, 1596-1599.
47. Zhang, J., Rosenberg, H.F., and Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *PNAS* 95, 3708-3713.