Biochemical and Biophysical Research Communications 438 (2013) 526–532

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Characterization of bud emergence 46 (BEM46) protein: Sequence, structural, phylogenetic and subcellular localization analyses

Abhishek Kumar, Krisztina Kollath-Leiß, Frank Kempken *

Abteilung für Botanik mit Schwerpunkt Genetik und Molekularbiologie, Botanisches Institut und Botanischer Garten, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

ARTICLE INFO

Article history:
Received 17 July 2013
Available online 31 July 2013

Keywords:
Bud emergence 46 (*bem46*)
Evolution
Fungi
Indels
Endoplasmic retention signal
eGFP-fusion
α/β-hydrolase

ABSTRACT

The bud emergence 46 (BEM46) protein from *Neurospora crassa* belongs to the α/β-hydrolase superfamily. Recently, we have reported that the BEM46 protein is localized in the perinuclear ER and also forms spots close by the plasma membrane. The protein appears to be required for cell type-specific polarity formation in *N. crassa*. Furthermore, initial studies suggested that the BEM46 amino acid sequence is conserved in eukaryotes and is considered to be one of the widespread conserved "known unknown" eukaryotic genes. This warrants for a comprehensive phylogenetic analysis of this superfamily to unravel origin and molecular evolution of these genes in different eukaryotes. Herein, we observe that all eukaryotes have at least a single copy of a *bem46* ortholog. Upon scanning of these proteins in various genomes, we find that there are expansions leading into several paralogs in vertebrates. Usingcomparative genomic analyses, we identified insertion/deletions (indels) in the conserved domain of BEM46 protein, which allow to differentiate fungal classes such as ascomycetes from basidiomycetes. We also find that exonic indels are able to differentiate BEM46 homologs of different eukaryotic lineage. Furthermore, we unravel that BEM46 protein from *N. crassa* possess a novel endoplasmic-retention signal (PEKK) using GFP-fusion tagging experiments. We propose that three residues namely a serine 188S, a histidine 292H and an aspartic acid 262D are most critical residues, forming a catalytic triad in BEM46 protein from *N. crassa*. We carried out a comprehensive study on *bem46* genes from a molecular evolution perspective with combination of functional analyses. The evolutionary history of BEM46 proteins is characterized by exonic indels in lineage specific manner.

1. Introduction

Bud emergence 46 (*bem46*) genes encode proteins belonging to the BEM46 family within the α/β-hydrolase superfamily. This α/β-hydrolase superfamily possesses a typical α/β-hydrolase domain, which is characterized by a β-sheet core of five to eight strands connected by α-helices to form a α/β/α sandwich [1,2]. Several hydrolytic enzymes share the common α/β-hydrolase domain with a wide array of catalytic functions with different phylogenetic history [2,3]. In fact, there are over 30,000 manually annotated members of this domain in the ESTHER database [1]. However, the functional characterization of BEM46 family members remainslargely unknown with no information of their hydrolase activity and substrate binding. The *S. pombe bem46* gene was originally described as a temperature-sensitive suppressor of *S. cerevisiae bem1* and *bud5* double mutant [4]. Mutations in *bem1* and *bud5* show defects in cell polarization and establishment of non-random budding patterns [5,6]. BUD5 is a GTP–GDP exchange factor (GEF) for a BUD1 Ras-like small G protein and is necessary for bud site selection [7]. BEM1 is a scaffold protein binding to BUD1, the Cdc24p GEF for the Cdc42p Rho-like small G protein, GTP-CDC42p, and to actin [5,8,9]. The *bem46* homologous gene-YNL320W in *S. cerevisiae* is not an essential single-copy gene [10]. Likewise, a *bem46* homolog *WAV2* in *Arabidopsis thaliana* was identified and it has no essential function [11]. Global analysis of protein interactions in *D. melanogaster* indicate that there is a low-confidence interaction between *Drosophila* BEM46 and RAPSYNOID [12], which is a putative GEF for Gα protein and involved in the control of asymmetric cell divisions [13]. In *N. crassa*, BEM46 may play a role in a signal transduction pathway involved in determining or maintaining cell type-specific polarity [14]. Hence members of the BEM46 family may modulate the function of some morphogenic determinants, like GEF or actin, on the cell surface. However, *bem46* remains as one of the widespread conserved "known unknown" eukaryotic genes, whose function remains elusive [15,16]. Further studies are clearly required for a better understanding of the function(s) of members belonging to the BEM46 family.

In the current study, we analyzed the molecular evolution of BEM46 family and we compiled BEM46 repository from selected eukaryotic genomes. We demonstrate that BEM46 from different

* Corresponding author.
E-mail address: fkempken@bot.uni-kiel.de (F. Kempken).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.07.103

eukaryotic lineages show lineage-specific indels that can be used as molecular marker for differentiating BEM46 in different lineages. Using GFP-fusion tagging, we provide evidence that the BEM46 protein (from *N. crassa*) has a novel and previously uncharacterized endoplasmic-retention signal (PEKK).

2. Materials and methods

We describ details of strains, DNA isolation, gel electrophoresis and PCR amplification in supplementary section S1. We also describ details of vector construction, transformation, transformant analysis and microscopic methods in supplementary section S2. We extracted sequences using full-length *N. crassa* BEM46 via NCBI, JGI and Ensembl [17] using BLAST suite [18]. We generated protein alignments of BEM46 sequences with MUSCLE [19] and visualized using GENEDOC [20]. We constructed Bayesian phylogenetic tree in the MrBayes 3.2.1 [21]. We performed divergence analysis for fungal BEM46 protein as described previously [22,23]. We created the homology model of BEM46 protein from *N. crassa* using the I-TASSER [24] and visualized it using YASARA [25]. We predicted active site residues of BEM46 from *N. crassa* using COFACTOR [26]. We provide details of data collection, sequence-structural and phylogenetic analyses in supplementary section S3.

3. Results and discussion

3.1. All eukaryotic organisms possess at least one copy of bem46 gene

We have compiled a comprehensive repository of BEM46 from representative eukaryotic organisms, as summarized by the Bayesian phylogenetic tree (Fig. 1). The majority of eukaryotic organisms have a single copy of BEM46 homolog, however, vertebrates have several paralogs of BEM46 (originated by duplication events), which separated in three different branches in this Bayesian phylogenetic tree (Fig. 1). Fungal BEM46 proteins are grouped into four sub-branches namely ascomycetes, basidiomycetes, *saccharomycetes* and methylotrophic yeasts. Interestingly, the parasitic fungus *Batrachochytrium dendrobatidis* (infecting frogs) has two copies of BEM46 and one copy is a close homolog of vertebrate BEM46 homologs (marked by arrow in Fig. 1). Possibly, this bem46 gene originated from horizontal gene transfer.

We have not detected *bem46* genes in bacterial genomes, however they possess other α/β-hydrolase superfamily members such as lysophospholipase in *Thioalkalivibrio* sp. ALJ9 (Genbank accession id: WP_018173930.1). The members of α/β-hydrolase superfamily are one of the largest and most diverse protein families known and these are comprised of proteases, lipases, esterases, dehalogenases, peroxidases, and epoxide hydrolases with different phylogenetic origin and catalytic function [2,3]. These members have been extensively studied in last two decades after the first report on α/β-hydrolases in 1992 [3]. Hence, we limited our study to the BEM46 protein family. To evaluate the sequence features of BEM46, we examined protein sequences in different eukaryotic lineages and were able to illustrate several insertions/deletions (indels) that might have played important roles in enhancing local diversities within overall conserved BEM46 core domain as described below. The existence of lineage-specific indels may also explain the differences between Bem46 homologs within different eukaryotic lineages.

3.2. Fungal BEM46 proteins have indels specific for fungal phyla

Using protein sequence analyses, we illustrate that there are distinct differences in terms of indels for ascomycetes and basidiomycetes (Fig. 2). A large indel is found in the region between 87P-128I amino acids (according to BEM46 protein numbering from *N. crassa*) as an insertion in basidomycetes but a deletion in ascomycetes. Another two amino acid indels, located between positions 205G-206D, are present in ascomycetes except yeast, but absent in basidiomycetes except rusts. These indels were mapped to BEM46 protein model structure to highlight their respective locations. The large indel is localized in between the β-sheet β2 to the helix α3, just before the α/β-hydrolase domain, while the two amino acid indel between positions 205G-206D is localized in the loop L10 in the α/β-hydrolase domain.

3.3. BEM46 protein from N. crassa possesses novel endoplasmic reticulum (ER)-retention signal

We considered BEM46 protein from *N. crassa* as a standard fungal BEM46, which is characterized by the presence of a signal peptide at the N-terminus with hydrophobic residues and an endoplasmic reticulum (ER)-retention signal PEKK at the C-terminal end (Fig. 3A). The ER-retention signal is designated by a prosite pattern as [KRHQSA]-[DENQ]-E-L. The ER-retention signal for *N. crassa* BEM46 does not fit into this typical pattern, however, it fits into recently reported 35 newly defined ER-retention signals [27] by using various permutation and combinations at these four positions. To examine if this putative motif serves as an authentic ER retention signal, we performed a localization study with (Fig. 3B, C) and without PEKK (Fig. 3D) at the C-terminal end using eGFP as a tag. The native protein carrying the putative ER retention signal coupled with eGFP showed the previously described localization to the perinuclear ER and in additional small spots near to the plasma membrane [14]. In contrast, the eGFP tagged protein without the PEKK sequence was not located to the ER any more, but forms several small cytoplasmic spots (Fig. 3C). Our experiment indicates that the PEKK motif is indeed an ER retention signal.

3.4. Structural analysis of BEM46 protein reveals catalytic triad

We modeled the BEM46 protein from *N. crassa* using I-TASSER [24]. It illustrates that BEM46 has a core α/β-hydrolase domain with eight β-sheets (Fig. 4A). In addition, BEM46 has an extra region with two β-sheets and two α-helices in the N-terminal region. To evaluate diversity within secondary structural elements, we carried out divergence analysis of individual secondary structural elements of different BEM46 proteins during fungal evolution. Fig. 4B illustrates the divergence pattern of these elements among different BEM46 proteins. We found that four loops L5, L11, L13 and L15 are particularly conserved. Normally, loops are conserved in a typical α/β-hydrolase and they harbor active site for ligand bindings. We computed putative binding site residues for three different ligands (Table 1) with C-score higher than −1.5, which reflects correct fold prediction whereas TM-score illustrates structural similarity between the template and the predicted model, [24]. We found that loop L2 possess an important residue, a glycine 117G, for two ligands namely acetic acid and glycerol. Furthermore, amino acid positions between 186G-189L spanning over β-sheet β6 and the loop L9 appear to be critical positions for putative binding sites of fungal BEM46 protein as summarized in Table 1. Among β-sheets, the β-sheet β7 is highly conserved and it harbors a putative binding site residue (isoleucine 211I). The loop L11 harbors an important residue, a phenyl alanine (216F). Among α-helices, the helix α6 appears to be a critical one, as it possesses an important residue, leucine 222L, while the next loop L12 possesses another critical position (proline 228P). Additionally, a potential binding site residue, valine 265 V, is also located near the loop L15, and also a histidine (292H) is located in loop L17. By combining these data, we propose that positions 117G, 188S,

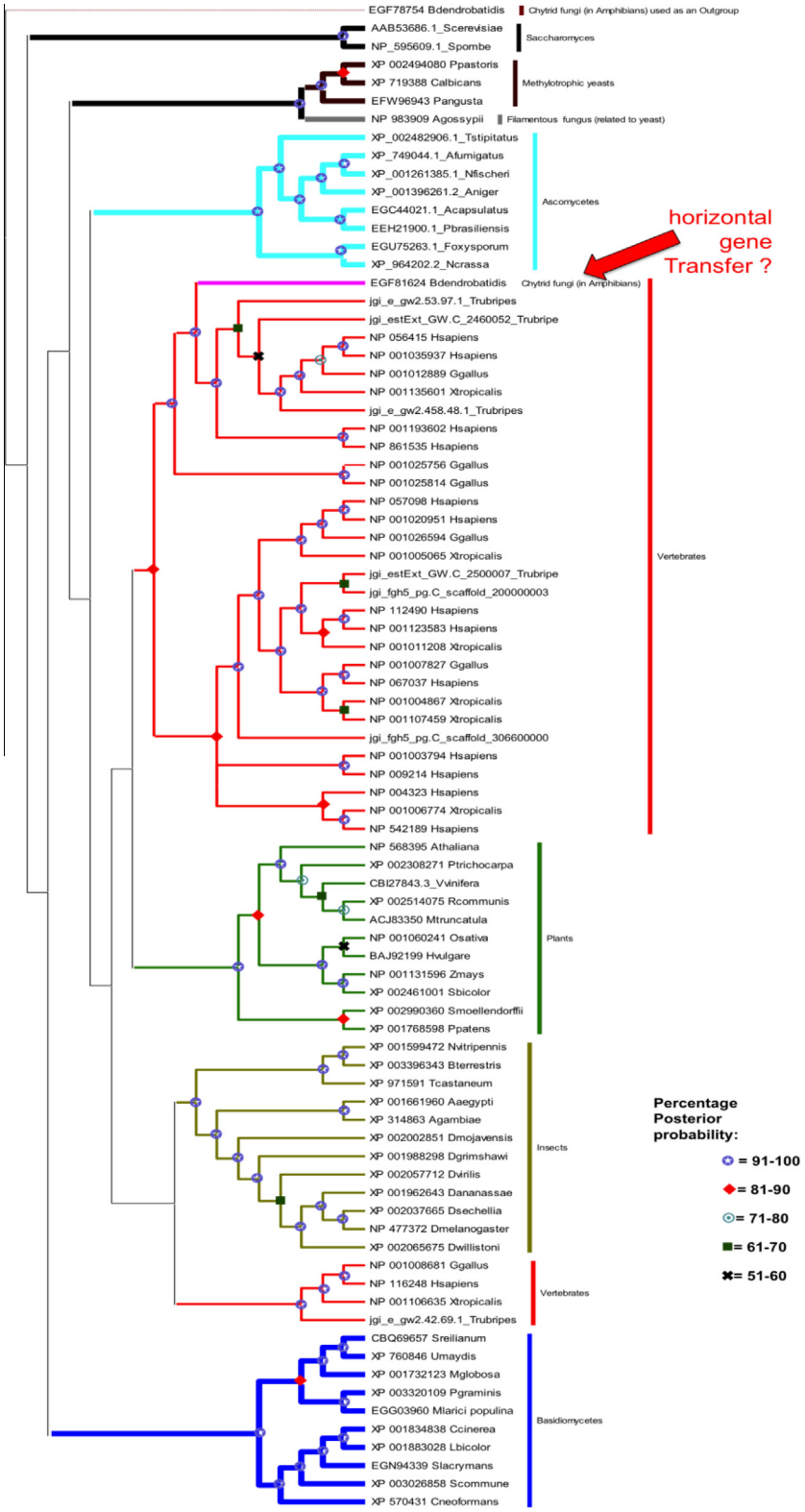

Fig. 1. Bayesian phylogenetic analysis depicts BEM46-like proteins are conserved across different lineages of eukaryotic organisms. Fungal branches are shown by thicker lines.

189L are very critical as more than two ligands are binding to these residues. On closer inspection, region 186G-189L harbors critical residues that are important for ligand binding (Table 1).

To further define the most essential residues, we predicted active site residues for BEM46 protein of *N. crassa,* and found that residues 188S, 262D and 292H are most influential residues for putative active site of BEM46 (marked by bold letters in Table 2). An E-score (higher than 1.0) indicate an excellent structural/functional prediction [24]. Serine 188S appears to be highly important residue in the BEM46 active site along with 262D and 292H. These two independent methods hint for four most critical residue for ligand binding and for activity of BEM46 protein as a serine 188S,

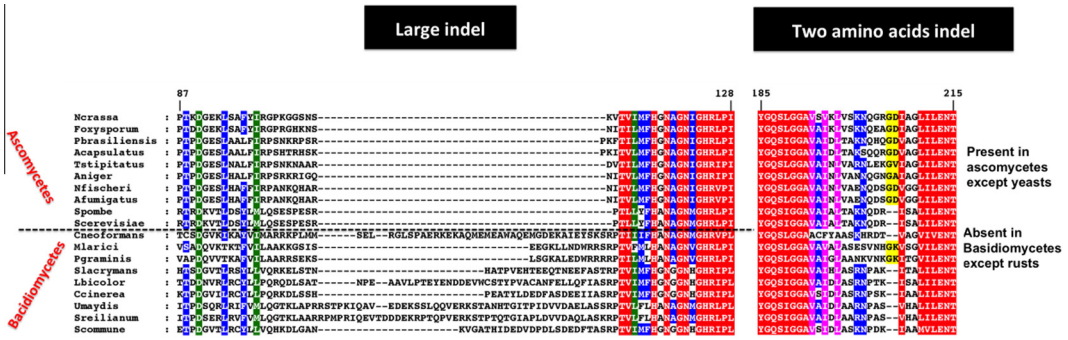

Fig. 2. Major insertions/deletions (indels) segments in BEM46-like proteins from different fungi. Numbering designates amino acid numbering of BEM46 from *N. crassa*. 100, 90 and 70% (with allowed substitutions) conserved residues is marked in red, green and blue, respectively. (For interpretation of the reference to color in this figure legend, the reader is referred to the web version of this article.)

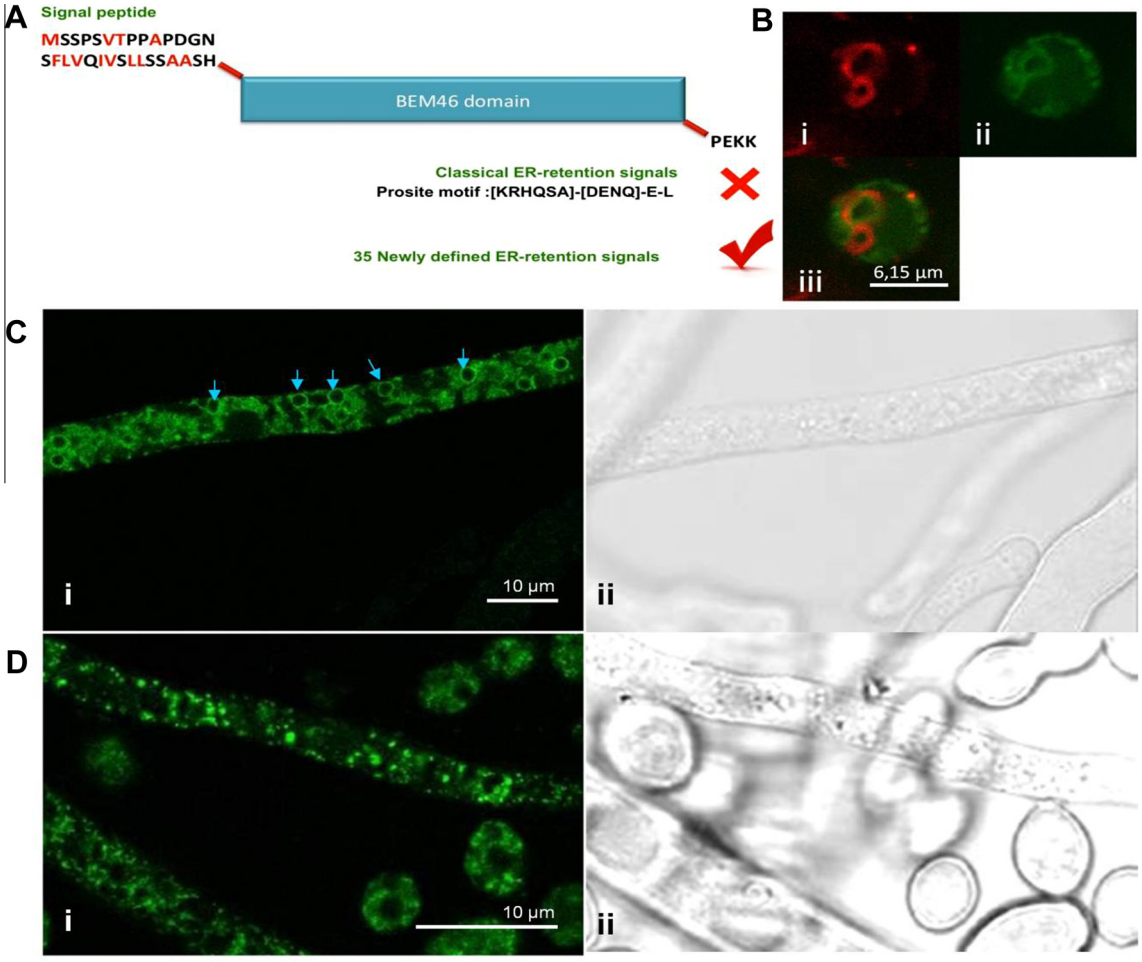

Fig. 3. Cellular localization analysis of BEM46 protein. The BEM46 protein of *N. crassa* carries a PEKK motif at the C-terminal end, which acts as an ER retention signal. (A) BEM46 protein from *N. crassa* is characterized by a signal peptide with patches of hydrophobic residues (red color) at N-terminus and an exceptional endoplasmic-reticulum (ER) retention signal at the C-terminal end. Raykhel et al. reported 35 newly defined ER-retention signals [27]. (B) Intracellular localization of Bem46::eGFP in a macroconidospore stained with the ER specific fluorescent marker ER Tracker™ Red. As previously reported [13] the Bem46::eGFP fusion construct colocalizes with the perinuclear ER stained with the fluorescent marker ER Tracker™ Red. Confocal laser scanning microscopical images of ER Tracker™ Red (i) Bem46::eGFP (ii) and overlay of both (iii). (C) Intracellular localization of the eGFP-tagged BEM46 protein with the PEKK motif. (D) Intracellular localization of the eGFP-tagged BEM46 protein without the PEKK motif. The perinuclear localization (green circles marked by blue arrows) of the BEM46:eGFP protein disappears, when the PEKK motif at the C-terminal end is missing. Confocal laser scanning microscopical images of young hyphae and macroconidia exhibiting (i) eGFP fluorescence, (ii) calculated bright field. (For interpretation of the reference to color in this figure legend, the reader is referred to the web version of this article.)

followed by a leucine 189L, a histidine 292H and an aspartic acid 262D. Upon comparing known canonical α/β-hydrolases, it is prominent that catalytic triads of these α/β-hydrolases are composed of a nucleophile (serine/cysteine/aspartic acid), an acid residue and a conserved histidine and they must be localized in the loops [2]. Three out of four proposed residues 188S, 262D and 292H fulfill these requirements for forming a catalytic triad of BEM46 protein. Furthermore, these residues are structurally con-

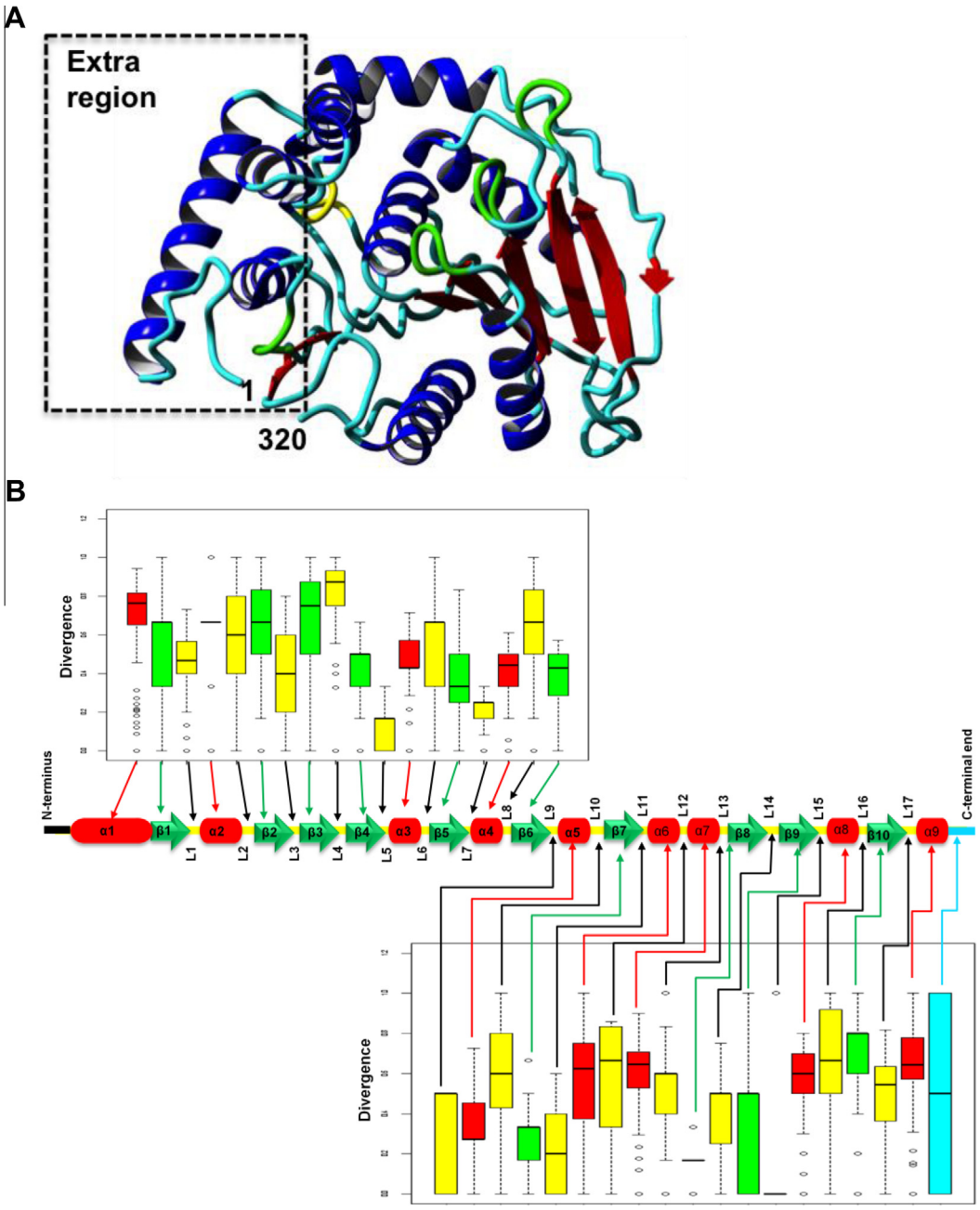

Fig. 4. Structural analysis of fungal BEM46. (A) Homology model of BEM46 from *N. crassa* suggests that BEM46 is composed of an extra region with two β-sheets and two α-helices in N-terminal region other than a typical α/β hydrolase with eight β-sheets. (B) Divergence analysis of different secondary structural elements in fungal BEM46 homologs. All values are significant ($P < 2.2e\text{-}16$, Kuskal–Wallis test).

Table 1
Prediction of putative binding sites of BEM46 from *N. crassa* based on known crystal structures as template with similar binding site. Predicted with I-TASSER software. Putative active site residues conserved in more than one ranking are marked in bold.

Rank	Cscore	PDB Hit	TM-score	RMSD	IDEN	Cov.	BS-score	Ligands	Predicted binding site residues in the model
1	0.45	3fyuC	0.786	2.34	0.181	0.866	1.30	Acetic acid (ACY)	**117G, 188S, 189L**, 216F, 292H
2	0.35	2z3wA	0.677	3.33	0.117	0.784	0.88	Glycerol (GOL)	**117G, 188S, 189L**, 222L, 228, 265 V
3	0.06	3dduA	0.692	3.49	0.130	0.819	0.85	Glycerol (GOL)	186G, 187Q, **189L**, 211I, 256F

Cscore is the confidence score of predicted binding site. Cscore values range in between [0–1]; where a higher score indicates a more reliable ligand-binding site prediction. BS-score is a measure of local similarity (sequence and structure) between template binding site and predicted binding site in the query structure. Based on large scale benchmarking analysis, we have observed that a BS-score > 1.1 reflects a very good local match between the predicted and template binding site.
TM-score is a measure of global structural similarity between query and template protein.
RMSD is the RMSD between residues that are structurally aligned by TM-align.
IDEN is the percentage sequence identity in the structurally aligned region.
Cov. represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

Table 2
Prediction of putative active site residues of BEM46 from *N. crassa* in comparison of top 5 enzyme homologs in protein databank (PDB) using COFACTOR software [35]. Putative active site residues conserved in more than one ranking are marked in bold.

Ranking	EC-score	PDB Hit	TM-score	RMSD	IDEN	Cov.	EC Number	Predicted active site residues
1	1.365	2jbwD	0.597	2.34	0.141	0.750	3.7.1.-	**188S, 292H**
2	1.210	1yr2A	0.591	3.32	0.106	0.800	3.4.21.26	191G, **262D**
3	1.188	1thtA	0.539	3.08	0.143	0.722	2.3.1.-	**188S, 292H**
4	1.109	1l7aA	0.692	2.35	0.161	0.872	3.1.1.72 3.1.1.41	125R
5	1.087	3ga7A	0.550	3.21	0.111	0.734	3.1.1.-	**262D, 292H**

EC-score is a confidence score for the Enzyme Classification (EC) Number prediction and it is used for ranking purpose. E-score (higher than 1.0) is indicatively of excellent structural/functional prediction [33].
PDB Hit is known enzyme structure in Protein databank
TM-score is a measure of global structural similarity between query and template protein.
RMSD is the RMSD between models and the PDB structure in the structurally aligned regions by TM-align.
IDEN is percentage sequence identity in the structurally aligned region.
Cov. represents the coverage of the alignment and is equal to the number of structurally aligned residues divided by length of model.

served in same positions as of the residues in the catalytic triad among several canonical α/β-hydrolases. A catalytic triad is defined by residues which share surface in the three dimensional space. To illustrate that we mapped surface area of these three residues (188S, 262D and 292H) as illustrated in Fig. S1 and it reveals that their combined surface area is interacting as required for any other catalytic triad. This triad is conserved across all known alpha/beta-hydrolases, which is also true with BEM46 as illustrated for insect BEM46 proteins aligned with BEM46 from *N. crassa* (Fig. S2).

3.5. BEM46 homologs share indels in other eukaryotic lineages

BEM46-like proteins from insects have only one amino acid indel at the position 80 (Fig. S2). Furthermore, these BEM46 homologs have an extension at the C-terminal end.

BEM46-like proteins from plant genomes have five indels (Fig. S3) namely one amino acid indel at the position 80, two amino acids indel between positions 205–206, eight to twelve amino acids indel between positions 232–233, two to three amino acids indel between positions 282–283 and two amino acids indel between positions 300–301.

Vertebrates possess several paralogs of fungal BEM46 protein, which are evolved by gene duplication events. These proteins have been termed as abhydrolase domain containing (ABHD) proteins, due to vertebrate-specific duplication events. Most prominent and close members to these ABHD proteins to fungal BEM46 are ABHD12 and ABHD13. These proteins are characterized by indels in comparison to BEM46 from *N. crassa* as summarized in Fig. S4. It is noteworthy that loss of function mutations in ABHD12 cause polyneuropathy, hearing loss, ataxia, retinitis pigmentosa, and cataract (PHARC) [28]. The ABHD12 has unique insertions in its core domain that are marked by red bars in Fig. S4. Mammalian ABHD proteins have important roles in lipid metabolism, lipid signal transduction, and metabolic disease, characterized via global profiling of dynamic protein palmitoylation [29,30]. However, annotations and computational characterizations of ABHD proteins are urgently required [29]. Our analysis shed first light into this direction.

3.6. Evolutionary history of BEM46 proteins are characterized by exonic indels

The findings here reveal a clear phylogenetic history of bem46 genes. BEM46 homologs are conserved in all eukaryotic organisms from fungi to human (Fig. 1). We demonstrated that the conserved BEM46 proteins are characterized by indels that discriminate these homologs in different lineages. Indels are generally considered as rare genomic characters and they serve a phylogenetic marker during evolution [31]. The phylogeny of eukaryotes is always under state of flux [32]. An influx of mutations governs the evolution of genes/genomes and these mutations undergo a stochastic process of neutral fixation via application of multiple selective pressures that can control the neutral fixation dynamics [32,33]. Characterization of both the mutational and fixation events assists in understanding the evolutionary processes. Indels are structural genomic changes, which play instrumental roles in applications that try to reveal genomic loci that are evolving under selection by virtue of either slowly or rapidly evolving sequences. In these analyses, it is essential that the mutational input at the genomic regions under study is not abnormally high or low [33–36]. Otherwise, the artifacts of the mutational dynamics predominate over inferred selection and it will be not a true indication for a functional constraint on the sequence. Genetic changes are introduced into genomes through three methods such as point mutations, insertions and deletions. The dynamics of each of these mechanisms are influenced by the genomic fragments under consideration and also by the presence of various factors acting in *trans*.

Prior to availability of genome sequences, evolutionary studies were primarily focused on either replacements of entire genes or fragments of genomic regions or point mutations. Loss or gains of genetic contents have direct immediate functional implications and can be scanned over long evolutionary times. Small structural changes such as 1–50 bps indels are intermediate in scale and are less frequent than single base substitutions. However, they can account for comparable base pairs of change. For instance, previously it was reported that 3.2% and 0.8% of the base pair changes between the fly species and in the primate species, respectively, were affected by indels. In contrast, only 1.8% and 1.5% of the base pairs were affected by point mutations in flies and primates, respectively [37]. Therefore, short indels are as sensitive as primary sequences and serve as significant factor in the mutational input that feeds into the evolutionary process, a fact that underlines the importance of characterizing the mechanisms that induce or suppress their activity. Earlier works were focused on human indels at disease loci [38–41] or on indels detected between relatively distant species [42–44]. These led to the suggestion that such events are correlated with specific sequence contexts. More recent works characterized extensive collections of indels in the human-chimp lineages [45,46], or human-mice lineages [37] and vertebrate-specific contexts of gene superfamily evolution [44]. These data further motivate a comprehensive approach to the description of their sequence contexts. Likewise, indels were also described in various studies in basal organisms such as for phylogenetic position of trichomonads [47] and also fungal genome analyses such

as during P-type ATPase gene family evolution [48] and during understanding of the anamorphic fungus *Rhizoctonia* species complex [49].

Generally, exonic indels require more complex mutational mechanisms and are more constrained than single base substitutions. In this work, we illustrate how indels play important roles in bringing diversity in homologous protein coding regions of *bem46* genes in various eukaryotic lineages, which could be implicated on functional level. However, attention is needed for further functional characterization in different eukaryotic lineages, which is yet in its infancy.

To the best of our knowledge, this is the first comprehensive study on *bem46* genes demonstrating BEM46 proteins being conserved from fungi to human, and BEM46 homologues being characterized by exonic indels that differentiate lineage of a particular BEM46 homolog. We defined the active site of fungal BEM46 using sequence and structural comparisons.

Acknowledgements

We thank the central microscopy plat form in the Biozentrum of our university (http://www.uni-kiel.de/biologie/zm/) for use of their equipment. We also thank Chandan Goswami for editing final version of this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2013.07.103.

References

[1] N. Lenfant, T. Hotelier, E. Velluet, et al., ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions, Nucleic Acids Res. 41 (2013) D423–429.
[2] M. Nardini, B.W. Dijkstra, Alpha/beta hydrolase fold enzymes: the family keeps growing, Curr. Opin. Struct. Biol. 9 (1999) 732–737.
[3] D.L. Ollis, E. Cheah, M. Cygler, et al., The alpha/beta hydrolase fold, Protein Eng. 5 (1992) 197–211.
[4] M.L. Valencik, J.R. Pringle, *Schizosaccharomyces pombe* bem1/bud5 suppressor (bem46) mRNA, EMBL database, accession number U29892. Website: <http://www.ebi.ac.uk/Tools/dbfetch/dbfetch?db=embl&id=U29892&style=raw>), 1995 (Accessed on 20 October 2011).
[5] E. Cabib, J. Drgonova, T. Drgon, Role of small G proteins in yeast cell polarization and wall biosynthesis, Annu. Rev. Biochem. 67 (1998) 307–333.
[6] K. Madden, M. Snyder, Cell polarity and morphogenesis in budding yeast, Annu. Rev. Microbiol. 52 (1998) 687–744.
[7] J. Chant, K. Corrado, J.R. Pringle, et al., Yeast BUD5, encoding a putative GDP-GTP exchange factor, is necessary for bud site selection and interacts with bud formation gene BEM1, Cell 65 (1991) 1213–1224.
[8] H.O. Park, E. Bi, J.R. Pringle, et al., Two active states of the Ras-related Bud1/Rsr1 protein bind to different effectors to determine yeast cell polarity, Proc. Natl. Acad. Sci. USA. 94 (1997) 4463–4468.
[9] J.E. Irazoqui, A.S. Gladfelter, D.J. Lew, Scaffold-mediated symmetry breaking by Cdc42p, Nat. Cell. Biol. 5 (2003) 1062–1070.
[10] G. Giaever, A.M. Chu, L. Ni, et al., Functional profiling of the *Saccharomyces cerevisiae* genome, Nature 418 (2002) 387–391.
[11] S. Mochizuki, A. Harada, S. Inada, et al., The *Arabidopsis* WAVY GROWTH 2 protein modulates root bending in response to environmental stimuli, Plant Cell 17 (2005) 537–547.
[12] L. Giot, J.S. Bader, C. Brouwer, et al., A protein interaction map of *Drosophila melanogaster*, Science 302 (2003) 1727–1736.
[13] M.L. Parmentier, D. Woods, S. Greig, et al., Rapsynoid/partner of inscuteable controls asymmetric division of larval neuroblasts in *Drosophila*, J. Neurosci. 20 (2000) RC84.
[14] M. Mercker, K. Kollath-Leiss, S. Allgaier, et al., The BEM46-like protein appears to be essential for hyphal development upon ascospore germination in *Neurospora crassa* and is targeted to the endoplasmic reticulum, Curr. Genet. 55 (2009) 151–161.
[15] M.Y. Galperin, E.V. Koonin, From complete genome sequence to 'complete' understanding?, Trends Biotechnol 28 (2010) 398–406.
[16] M.Y. Galperin, E.V. Koonin, 'Conserved hypothetical' proteins: prioritization of targets for experimental study, Nucleic Acids Res. 32 (2004) 5452–5463.
[17] P. Flicek, I. Ahmed, M.R. Amode, et al., Ensembl, Nucleic Acids Res. 41 (2013) (2013) D48–55.
[18] S.F. Altschul, T.L. Madden, A.A. Schaffer, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[19] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[20] K.B. Nicholas, H.B. Nicholas Jr., D.W.I. Deerfield, GeneDoc: analysis and Visualization of Genetic Variation, EMBNEW NEWS 4 (1997) 14.
[21] F. Ronquist, J.P. Huelsenbeck, MrBayes 3: bayesian phylogenetic inference under mixed models, Bioinformatics 19 (2003) 1572–1574.
[22] P. Sardar, A. Kumar, A. Bhandari, et al., Conservation of Tubulin-binding sequences in TRPV1 throughout evolution, PLoS ONE 7 (2012) e31448.
[23] A. Kumar, A. Bhandari, R. Sinha, et al., Molecular phylogeny of OVOL genes illustrates a conserved C2H2 zinc finger domain coupled by hyper variable unstructured regions, PLoS ONE 7 (2012) e39399.
[24] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat. Protoc. 5 (2010) 725–738.
[25] E. Krieger, G. Koraimann, G. Vriend, Increasing the precision of comparative models with YASARA NOVA – a self-parameterizing force field, Proteins 47 (2002) 393–402.
[26] A. Roy, J. Yang, Y. Zhang, COFACTOR: an accurate comparative algorithm for structure-based protein function annotation, Nucleic Acids Res. 40 (2012) W471–W477.
[27] I. Raykhel, H. Alanen, K. Salo, et al., A molecular specificity code for the three mammalian KDEL receptors, J. Cell Biol. 179 (2007) 1193–1204.
[28] T. Fiskerstrand, D. H'Mida-Ben Brahim, S. Johansson, et al., Mutations in ABHD12 cause the neurodegenerative disease PHARC: an inborn error of endocannabinoid metabolism, Am. J. Hum. Genet. 87 (2010) 410–417.
[29] C.C. Lord, G. Thomas, J.M. Brown, Mammalian alpha beta hydrolase domain (ABHD) proteins: lipid metabolizing enzymes at the interface of cell signaling and energy metabolism, Biochim. Biophys. Acta. 1831 (2013) 792–802.
[30] B.R. Martin, C. Wang, A. Adibekian, et al., Global profiling of dynamic protein palmitoylation, Nat. Methods 9 (2012) 84–U205.
[31] A. Rokas, P.W. Holland, Rare genomic changes as a tool for phylogenetics, Trends Ecol. Evol. 15 (2000) 454–459.
[32] H. Philippe, P. Lopez, H. Brinkmann, et al., Early-branching or fast-evolving eukaryotes? an answer based on slowly evolving positions, Proc. Biol. Sci. 267 (2000) 1213–1221.
[33] G. Lunter, C.P. Ponting, J. Hein, Genome-wide identification of human functional DNA using a neutral indel model, PLoS Comput. Biol. 2 (2006) e5.
[34] A.G. Clark, The search for meaning in noncoding DNA, Genome Res. 11 (2001) 1319–1320.
[35] D.J. Gaffney, P.D. Keightley, The scale of mutational variation in the murid genome, Genome Res. 15 (2005) 1086–1094.
[36] G. Lunter, Probabilistic whole-genome alignments reveal high indel rates in the human and mouse genomes, Bioinformatics 23 (2007) i289–296.
[37] A. Tanay, E.D. Siggia, Sequence context affects the rate of short insertions and deletions in flies and primates, Genome Biol. 9 (2008) R37.
[38] N.A. Chuzhanova, E.J. Anassis, E.V. Ball, et al., Meta-analysis of indels causing human genetic disease: mechanisms of mutagenesis and the role of local DNA sequence complexity, Hum. Mutat. 21 (2003) 28–44.
[39] A.S. Kondrashov, I.B. Rogozin, Context of deletions and insertions in human coding sequences, Hum. Mutat. 23 (2004) 177–185.
[40] E.V. Ball, P.D. Stenson, S.S. Abeysinghe, et al., Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity, Hum. Mutat. 26 (2005) 205–213.
[41] J.M. Chen, N. Chuzhanova, P.D. Stenson, et al., Complex gene rearrangements caused by serial replication slippage, Hum. Mutat. 26 (2005) 125–134.
[42] E.E. Thomas, N. Srebro, J. Sebat, et al., Distribution of short paired duplications in mammalian genomes, Proc. Natl. Acad. Sci. USA 101 (2004) 10349–10354.
[43] S. Sinha, E.D. Siggia, Sequence turnover and tandem repeats in cis-regulatory modules in *Drosophila*, Mol. Biol. Evol. 22 (2005) 874–885.
[44] A. Kumar, H. Ragg, Ancestry and evolution of a secretory pathway serpin, BMC Evol. Biol. 8 (2008) 250.
[45] F.C. Chen, C.J. Chen, W.H. Li, et al., Human-specific insertions and deletions inferred from mammalian genome sequences, Genome Res. 17 (2007) 16–22.
[46] P.W. Messer, P.F. Arndt, The majority of recent short DNA insertions in the human genome are tandem duplications, Mol. Biol. Evol. 24 (2007) 1190–1197.
[47] E. Bapteste, H. Philippe, The potential value of indels as phylogenetic markers: position of trichomonads as a case study, Mol. Biol. Evol. 19 (2002) 972–977.
[48] N. Corradi, I.R. Sanders, Evolution of the P-type II ATPase gene family in the fungi and presence of structural genomic changes among isolates of *Glomus intraradices*, BMC Evol. Biol. 6 (2006) 21.
[49] D. Gonzalez, M.A. Cubeta, R. Vilgalys, Phylogenetic utility of indels within ribosomal DNA and beta-tubulin sequences from fungi in the *Rhizoctonia solani* species complex, Mol. Phylogenet. Evol. 40 (2006) 459–470.