Amino Acids (2018) 50:79–94
https://doi.org/10.1007/s00726-017-2491-5

ORIGINAL ARTICLE

Structural complexity and functional diversity of plant NADPH oxidases

Gurpreet Kaur[1,2,3,5] · Kunchur Guruprasad[2] · Brenda R. S. Temple[4] · David G. Shirvanyants[3] · Nikolay V. Dokholyan[3] · Pratap Kumar Pati[1]

Received: 29 June 2017 / Accepted: 11 September 2017 / Published online: 25 October 2017

Abstract Plant NADPH oxidases also known as respiratory burst oxidase homologs (Rbohs) are a family of membrane-bound enzymes that play diverse roles in the defense response and morphogenetic processes via regulated generation of reactive oxygen species. Rbohs are associated with a variety of functions, although the reason for this is not clear. To evaluate using bioinformatics, the possible mechanisms for the observed functional diversity within the plant kingdom, 127 Rboh protein sequences representing 26 plant species were analyzed. Multiple clusters were identified with gene duplications that were both dicot as well as monocot-specific. The N-terminal sequences were observed to be highly variable. The conserved cysteine (equivalent of Cys890) in C-terminal of AtRbohD suggested that the redox-based modification like *S*-nitrosylation may regulate the activity of other Rbohs. Three-dimensional models corresponding to the N-terminal domain for Rbohs from *Arabidopsis thaliana* and *Oryza sativa* were constructed and molecular dynamics studies were carried out to study the role of Ca^{2+} in the folding of Rboh proteins. Certain mutations indicated possibly affect the structure and function of the plant NADPH oxidases, thereby providing the rationale for further experimental validation.

Keywords Plant NADPH oxidase · Phylogenetic analysis · Mutation · Ortholog · Homology modeling · Discrete molecular dynamics

Introduction

Membrane-localized NADPH oxidases known as respiratory burst oxidase homologs (Rbohs) in plants are key players in the generation of superoxide radicals (O_2^-) (a type of reactive oxygen species; ROS). Reports on their role in the defense response and vital morphogenetic processes such as cell expansion, pollen tube growth, seed germination and after-ripening, root hair elongation, Ca^{2+}-dependent stomatal closure have been documented (Kaur et al. 2014). The first plant NADPH oxidase was identified in *Oryza sativa*, called as *OsRbohA* (Groom et al. 1996). Subsequently, more Rbohs were identified in other species. Rboh protein family displays a diverse portfolio and comprises 127 members from 26 plant species with isoforms ranging from A to J.

Rbohs encode homologs of the mammalian NADPH oxidase (Nox) catalytic subunit known as $gp91^{phox}$. The domain architecture comprises an extended N-terminal region containing two Ca^{2+}-binding EF-hand motifs, six α-helical transmembrane domains (TMD-I to TMD-VI) connected by five loops (loops A–E) and a C-terminal FNR

Handling Editor: J. D. Wade.

Electronic supplementary material The online version of this article (doi:10.1007/s00726-017-2491-5) contains supplementary material, which is available to authorized users.

✉ Pratap Kumar Pati
pkpati@yahoo.com

1 Department of Biotechnology, Guru Nanak Dev University, Amritsar, India

2 Bioinformatics, Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad, India

3 Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, NC, USA

4 R. L. Juliano Structural Bioinformatics Core Facility, School of Medicine, University of North Carolina, Chapel Hill, NC, USA

5 Present Address: Max Planck Institute for Developmental Biology, Tuebingen, Germany

(ferredoxin-NADP$^+$ reductase) domain containing FAD (flavin adenine dinucleotide) and NADPH-binding moieties (Kaur et al. 2014). Experimental evidence suggest that the N-terminal region plays a key role in the regulation of Rbohs via Ca^{2+}, Rac GTPase, protein kinases (CDPK; calcium-dependent protein kinase, CcaMK; Ca^{2+}/CaM-dependent protein kinase, MAPK; mitogen activated protein kinase, OST1; open stomata 1, receptor-like cytoplasmic kinase BIK1, histidine kinase), GIRAFFE heme oxygenase, extracellular ATP, phospholipase Dα1, and phosphatidic acid (Kaur et al. 2014; Baxter et al. 2014). The N-terminal region of RbohB from *O. sativa* (OsRbohB$^{138-313}$) comprises a homodimer, where each monomer contains two EF-hands and two EF-hand-like motifs with Ca^{2+}-binding sites (Oda et al. 2010). Ca^{2+} binding to the EF-hand and Ca^{2+}-dependent phosphorylation act together to control ROS production in OsRbohB (Wong et al. 2007) and *Arabidopsis thaliana* AtRbohC and AtRbohD (Kobayashi et al. 2007; Takeda et al. 2008). However, Ca^{2+}-independent intramolecular interaction was also observed between the N and C termini in OsRbohB, which requires the N-terminal region upstream of EF-hands (Oda et al. 2010). The possible lack of comprehensive phylogenetic information, three-dimensional structure, folding mechanism and physiological function for many Rbohs creates a gap in our understanding of Rbohs function. To bridge this gap, various bioinformatics approaches were employed (Lee et al. 2007; Kaur et al. 2015) to analyze the Rboh proteins from Arabidopsis and rice. A range of gene duplications were identified that were dicot as well as monocot-specific. Further, the sequence and disorder analyses indicated the N-terminal to be highly variable compared to the transmembrane and C-terminal domains. Three-dimensional models for the N-terminal domain for Rbohs from *A. thaliana* and *O. sativa* were generated and amino acid residues likely to affect functions were identified. Further, the role of Ca^{2+} in the folding of Rboh proteins was studied using discrete molecular dynamics. The present study provides a rational to evaluate the effects of certain amino acid residues on the structure and function of the plant NADPH oxidases in crop improvement programmes, for instance, crops with better stress tolerance.

Materials and methods

Phylogenetic analysis

A total of 127 Rboh protein sequences were retrieved from 26 plant species (Kaur et al. 2014). Multiple sequence alignments were generated using the Clustal Omega program (Sievers et al. 2011). Neighbor-joining and maximum likelihood based trees were generated using the ClustalX (Thompson et al. 1997) and PhyML 3.0 programs (Dereeper et al. 2008), respectively. Phylogenetic tree topology was drawn with the MEGA6 software (Tamura et al. 2013) and the evolutionary trace analysis (ETA) was carried out using the evolutionary trace server (Innis et al. 2000).

Ancestral sequence reconstruction

The sequence alignment was divided into five monophyletic groups, each containing monocot and dicot sequences as described here: (1) AtRbohD, AtRbohA, AtRbohG, AtRbohC, OsRbohI; (2) AtRbohB, OsRbohB, OsRbohH; (3) AtRbohF, OsRbohC, OsRbohA; (4) AtRbohE, OsRbohG, OsRbohF; and (5) AtRbohH, AtRbohJ, OsRbohD, OsRbohE. The four lower plant sequences CmRboh1, CmRboh2, PyRboh and CcNoxD were used as outgroups for ancestral reconstruction. The alignment for each monophyletic group with outgroup sequences were submitted to the FastML server (http://fastml.tau.ac.il/), and ancestral sequences were calculated using the JTT model of substitution with a Gamma distribution.

Multiple sequence alignment, identification of orthologs and structure analysis

Protein sequences of *A. thaliana* (AtRbohA-J) and *O. sativa* (OsRbohA-I) were aligned using Clustal Omega (Sievers et al. 2011). Secondary structures were predicted using PSIPRED (Jones 1999) through the Ali2D (Alva et al. 2016) server (http://toolkit.tuebingen.mpg.de/ali2d) and JPred3 (Cole et al. 2008) (http://www.compbio.dundee.ac.uk/jpred3/refs.html). Orthologous sequences were identified using InParanoid8 (Sonnhammer and Östlund 2015) (http://inparanoid.sbc.su.se/cgi-bin/faq.cgi), OMA Browser (Schneider et al. 2007) and OrthologID (Chiu et al. 2006).

Template searching was performed with PSI-BLAST (Altschul et al. 1997) and HHPred (Soding et al. 2005), and homology modeling was carried out using Modeller9v8 (Sali and Blundell 1993) for N-terminal EF-hand and C-terminal NADPH-binding regions. *Ab initio* modeling for N-terminal region upstream of EF-hands was carried out using I-TASSER (Roy et al. 2010) and Phyre software (Kelley and Sternberg 2009); and the I-TASSER models were used for subsequent studies. Structure optimization for models was carried out with Chiron (Ramachandran et al. 2011), Gaia (Kota et al. 2011) and KoBa^{MIN} (Rodrigues et al. 2012). The stereochemical and quality of models were assessed with structural analysis and verification server (http://nihserver.mbi.ucla.edu/SAVES/) and QMEAN server (Benkert et al. 2009) (http://swissmodel.expasy.org/qmean/cgi/index.cgi).

Simulations were conducted using replica exchange discrete molecular dynamics (RX-DMD) via πDMD software package (http://www.moleculesinaction.com/pdmd.html) (Ding and Dokholyan 2006; Ding et al. 2008; Dokholyan

et al. 1998; Proctor et al. 2011; Shirvanyants et al. 2012). RX-DMD simulations for the three OsRbohB forms were performed with 18 replicas at temperatures 0.480, 0.520, 0.540, 0.560, 0.580, 0.600, 0.620, 0.640, 0.650, 0.660, 0.670, 0.680, 0.690, 0.700, 0.710, 0.720, 0.740 and 0.760. The weighted histogram analysis method (WHAM) was used to analyze the thermodynamics of protein folding via rexwham program on replica exchange trajectories (Kumar et al. 1992). Electrostatic surface potentials were generated using Poisson–Boltzmann electrostatic calculation methods (PDB2PQR and ABPS input) (Dolinsky et al. 2004). Docking simulations were performed with ClusPro (Comeau et al. 2004) and ZDOCK (Chen et al. 2003). Structure visualization and the generation of figures were carried out using PyMol (DeLano 2002).

Disordered regions were predicted using metaPrDOS (Ishida and Kinoshita 2008) and GSmetaDisorder (Kozlowski and Bujnicki 2012). PrDOS (Ishida and Kinoshita 2007) was used for OsRbohF and OsRbohG (> 1000 aa) and the results were evaluated using DynaMine (Cilia et al. 2013).

Results

Phylogenetic analysis of Rboh family within plant kingdom

The phylogeny showing the closely and distantly related plant NADPH oxidases is presented in Fig. 1. It corresponds to 127 full-length protein sequences for experimentally characterized Rbohs, obtained from 26 plant species including dicots, monocots and lower plants (S1 File). We reconstructed phylogeny established using neighbor-joining method (Fig. 1) and a very similar phylogenetic tree was also generated using maximum likelihood method. As expected on the basis of the evolutionary separation existing between lower and higher plants, Rbohs from red algae formed a separate sub-cluster 1A. Bootstrap support for this division was as high as 98.2% (Fig. 1). In higher plants, we observed a more extensive expansion in the number of Rboh proteins due to gene duplications (Fig. 2). The Rboh proteins from higher plants were grouped into nine main clusters (2–10), with clusters 2–4, six being subdivided into two sub-clusters A and B. Clusters 2–6 contained both dicot and monocot Rbohs while cluster 7 was monocot-specific and clusters 8, 9 and 10 were dicot-specific. Cluster 1 showed members of red algae to be phylogenetically more proximal to Rbohs from Fabaceae (*Glycine max* and *Phaseolus vulgaris*). The distribution of Rbohs in red algae indicated the presence of one Rboh in *Chondrus crispus*, *Porphyra yezoensis,* whereas two in the case of *Cyanidioschyzon merolae*. No introgression of any dicot and monocot species was observed in this cluster. Cluster 2 contained representative members of both dicots and monocots. The legume species predominate the dicots, whereas members of *A. thaliana* and *Populus trichocarpa* have also been included in the cluster (sub-cluster 2A). In sub-cluster 2A, the presence of *A. thaliana* AtRbohH and AtRbohJ and in the same sub-cluster, an orthologous relationship of two *P. trichocarpa* Rbohs (PtRbohH and PtRbohJ) with AtRbohs was observed. Another group constituted five Rbohs from four legumes. A member of Vitaceae family (*Vitis vinifera* VvRbohH) with unknown function was also included in the sub-cluster. However, the other sub-cluster (2B) was predominated by cereal crops belonging to monocots. Sub-cluster 2B contained one protein from *Hordeum vulgare* and two proteins each from *O. sativa* and *Zea mays*. Cluster 3 included both monocot and dicot Rbohs within sub-cluster 3A and 3B, respectively, belonging to same plant families as cluster 2 but different Rboh subfamilies. Cluster 4 appears to be complex showing *A. thaliana* AtRbohI closer to monocots than dicots. Monocot sub-cluster 4A contained Rbohs from *Z. mays* (ZmRbohB-α and ZmRbohB-β), *O. sativa* (OsRbohA) and *H. vulgare* (HvRbohA and HvRbohF1). This sub-cluster was more close to that part of dicotic sub-cluster (cluster 4B) which also contained Rbohs involved in biotic stress from Solanaceae family; *Lycopersicon esculentum* (LeRboh1), *Solanum tuberosum* (StRbohA) and *Nicotiana benthamiana* (NbRbohA) and *Nicotiana tabacum* (NtRbohF). Members from Vitaceae and Salicaceae family were closer to Solanaceae than legumes. Cluster 5 contained different members belonging to family Brassicaceae (*Lepidium sativum* LesaRbohA, *Brassica oleracea* BoRbohD and BoRbohF), Cucurbitaceae (*Cucumis sativus* CsRboh) and monocot. The introgression of members of leguminous and Solanaceae have not been observed in this cluster. Cluster 6 was represented by monocots and dicots. The members of different families such as Poaceae, Brassicaceae, Fabaceae, Salicaceae, Vitaceae and Solanaceae were observed in this cluster. This cluster can be sub-divided into two major sub-clusters, one representing the members of monocots (6A) and the other of dicots (6B). Cluster 7 was exclusively a monocot-specific cluster with representation from rice and maize. On the other hand, clusters 8–10 were dicot-specific with majority of members from Brassicaceae. Cluster 9 was mostly dominated by members of family Fabaceae. Besides Fabaceae, isolated member of Vitaceae family (*V. vinifera* VvRbohA) of unknown function was observed in this cluster. Cluster 10 was predominated by members of family Solanaceae including members from *P. trichocarpa* (PtRbohA and PtRbohC).

The ancestral sequence reconstruction using the multiple sequence alignment of 127 Rbohs and the phylogenetic tree were used to identify amino acid substitutions that possibly account for the functional versatility of Rbohs within the plant kingdom. Red algae Rbohs (CmRboh1, CmRboh2,

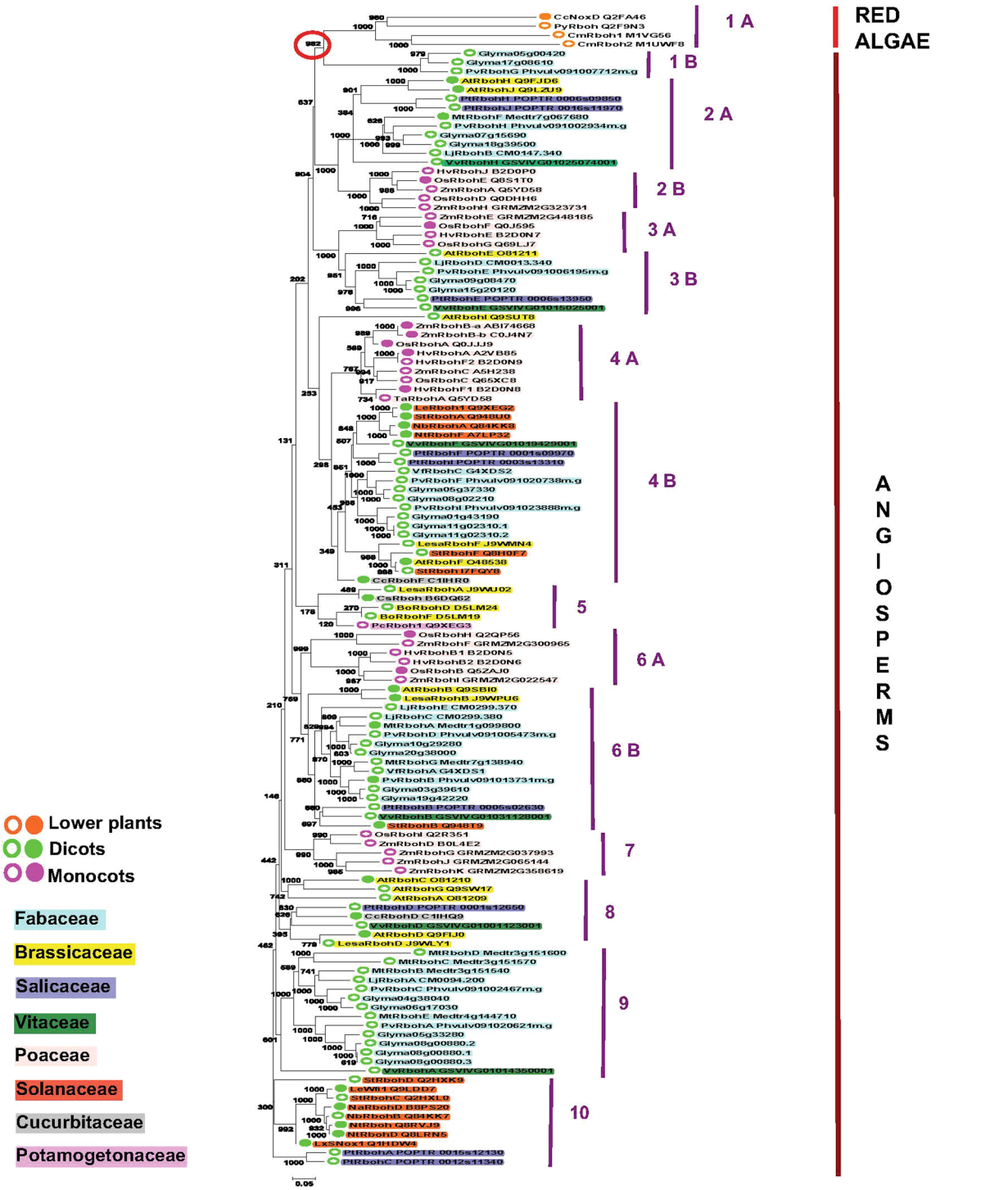


Fig. 1 Phylogenetic tree for127 Rbohs from 26 plant species. The unrooted tree was inferred by the neighbor-joining method with 1000 bootstraps after the alignment of the Rboh amino acid sequences listed in Supplementary File S1. Rbohs are labeled with their name and accession number. Green, pink and orange circles indicate dicot, monocot and lower plant Rbohs, respectively. Filled and empty circles denote experimentally validated and unknown functions respectively. At, *Arabidopsis thaliana*; Bo, *Brassica oleracea*; Cc, *Citrullus colocynthis*; Cs, *Cucumis sativus*; Glyma, *Glycine max*; Lesa, *Lepidium sativum*; Lj, *Lotus japonicus*; Le, *Lycopersicon esculentum*; Mt, *Medicago truncatula*; Na, *Nicotiana attenuata*; Nb, *Nicotiana benthamiana*; L × S, *Nicotiana langsdorffii* × *Nicotiana sanderae*; Nt, *Nicotiana tabacum*; Pv, *Phaseolus vulgaris*; Pt, *Populus trichocarpa*; St, *Solanum tuberosum*; Vf, *Vicia faba*; Vv, Vitis vinifera; Hv, *Hordeum vulgare*; Os, *Oryza sativa*; Pc, *Potamogeton crispus*; Ta, *Triticum aestivum*; Zm, *Zea mays*; Cc, *Chondrus crispus*; Cm, *Cyanidioschyzon merolae*; Py, *Porphyra yezoensis* (color figure online)

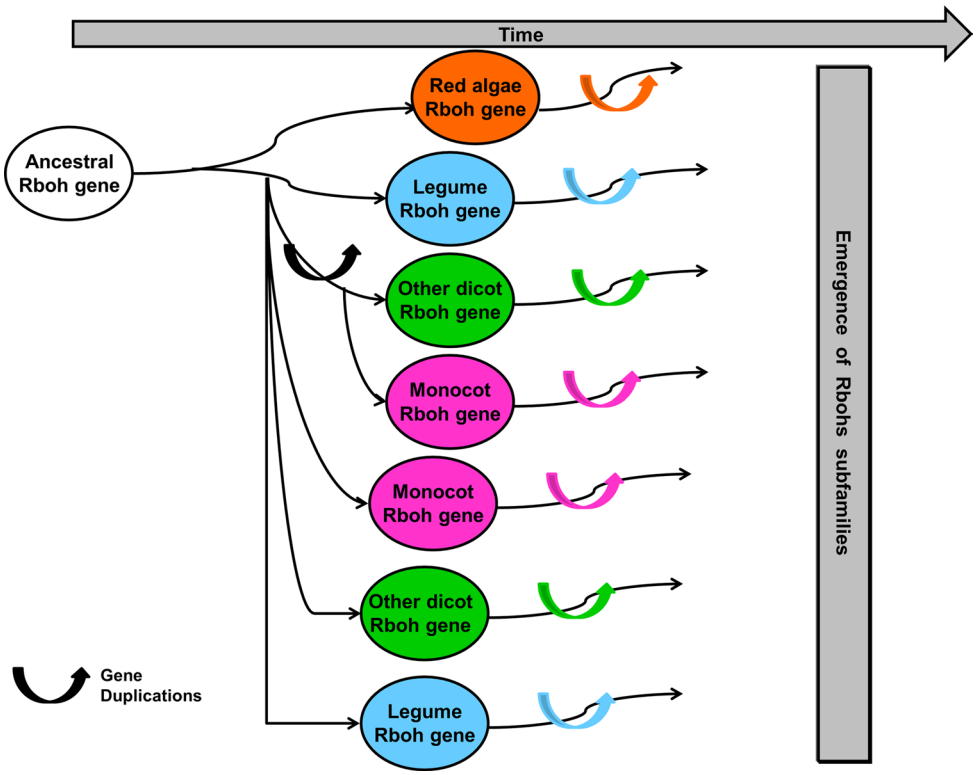


Fig. 2 Gene duplications in Rbohs. Model drawn in reference to Fig. 1, indicating evolutionary separation and duplication of Rbohs between lower (algae) and higher plants, and further extensive expansion in the number of Rbohs due to gene duplications in legume, other dicots and monocots

PyRboh and CcNoxD) were used as outgroups. First, ancestors were predicted for the nodes closest to the monocot and dicot clusters 2A, 2B, 3A, 3B, 4A, 4B, 6A, and 6B and the ancestors to the combined clusters 2A/2B, 3A/3B, 4A/4B and 6A/6B. Further, 2A/2B, 3A/3B, 4A/4B and 6A/6B clusters were analyzed for ancestral sequences (ANC), sites predicted with high confidence (INV) and discriminative residues among the clusters (DISCRIM) were determined from the alignments. The final output indicating ANC, INV and DISCRIM residues among AtRbohs and OsRbohs from the four clusters is shown in S2 File.

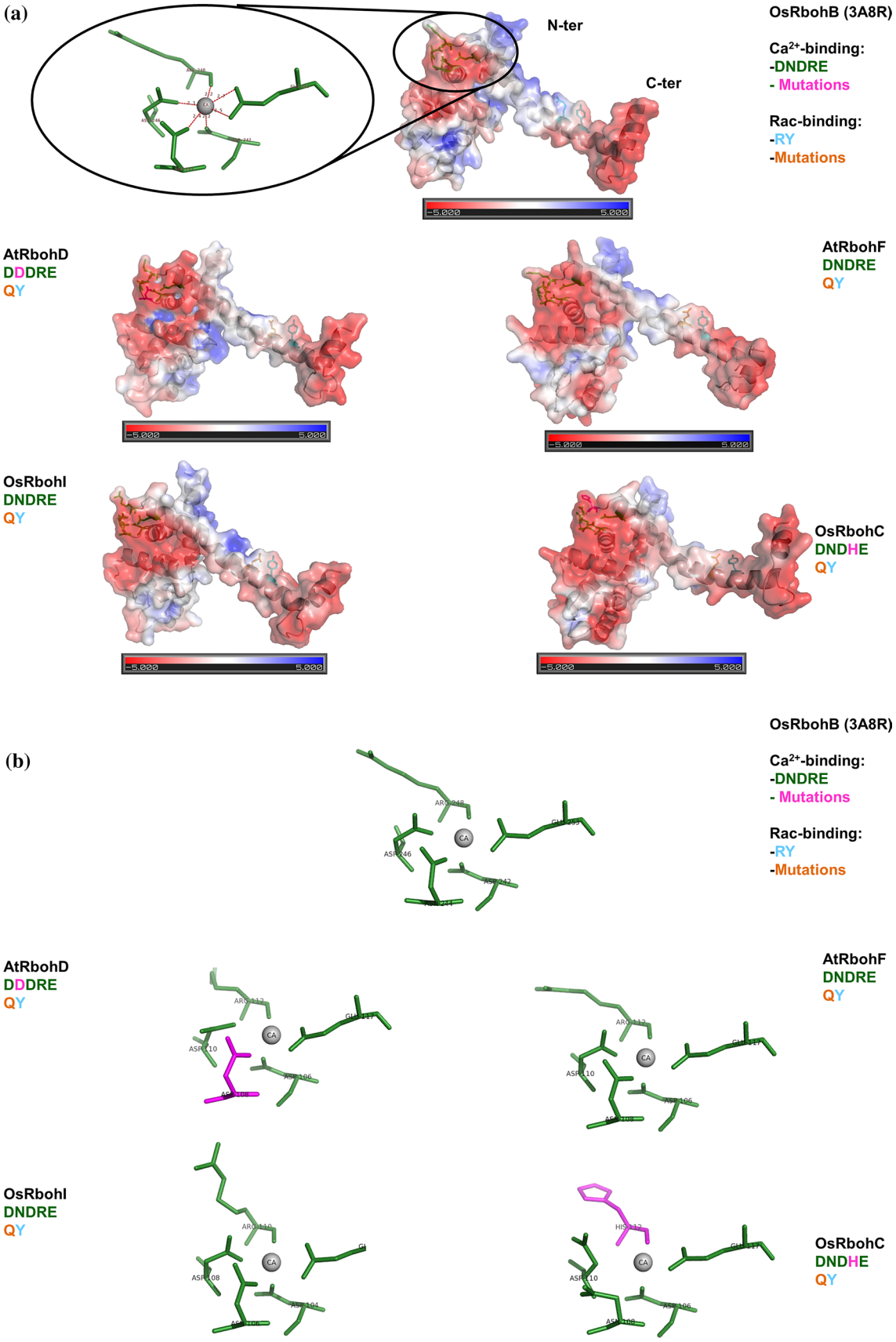
According to the evolutionary trace analysis (ETA), different partitions P01–P20 divide the phylogenetic tree into classes (S3 File). The ET classes identified all clusters obtained from phylogenetic analysis. Each partition included different numbers of classes, where each class contained cluster of similar sequences originating from a given node within that partition. In partition P01, all 127 sequences cluster into one class. In partition P02, these sequences divide into two classes 1 and 2. In the present work, we focused on ten Arabidopsis and nine rice Rbohs.

Sequence analysis and structural studies

The multiple sequence alignment for the 19 Rboh proteins (10 from Arabidopsis and 9 from rice), showed regions of conserved and non-conserved residues (S4 File). OsRbohI and OsRbohC were identified as putative orthologs of AtRbohD and AtRbohF, respectively (S5 File).

N-terminal EF-hands

The crystal structure corresponding to the N-terminal domain of OsRbohB (PDB: 3A8R) comprises a homodimer with two EF-hands and two EF-hand-like motifs. The Ca^{2+}-binding site EF-hand I comprises the amino acid residues; D242, N244, D246, R248 and E253 (Oda et al. 2010). The sequence identities for the 18 Rbohs corresponding to the N-terminal EF-hand ranged from 42 to 67%. The three-dimensional models comprised ~89 to ~95% residues in the 'most favoured regions' of the Ramachandran plot with deletions and substitutions associated in Ca-binding sites (S6 File). Except for five Rbohs (OsRbohA, OsRbohG, OsRbohI, AtRbohB and AtRbohF), the remaining were associated with substitutions corresponding to the equivalent Ca^{2+}-binding residues (S4 File). OsRbohF was associated with a deletion of five amino acid residues corresponding to the Ca^{2+}-binding ligands in EF-hand I, whereas AtRbohI was associated with deletion of D242. AtRbohA, AtRbohC, AtRbohD, AtRbohG and AtRbohJ were associated with N244D substitution, and AtRbohE, AtRbohH and AtRbohJ with R248K. In the case of OsRbohD and OsRbohH, the R248Q substitution was observed, whereas OsRbohC and OsRbohE were associated with R248H and R248M


(a)
N-ter
C-ter
OsRbohB (3A8R)
Ca²⁺-binding:
-DNDRE
- Mutations
Rac-binding:
-RY
-Mutations
AtRbohD
DDDRE
QY
AtRbohF
DNDRE
QY
OsRbohI
DNDRE
QY
OsRbohC
DNDHE
QY
(b)
OsRbohB (3A8R)
Ca²⁺-binding:
-DNDRE
- Mutations
Rac-binding:
-RY
-Mutations
AtRbohD
DDDRE
QY
AtRbohF
DNDRE
QY
OsRbohI
DNDRE
QY
OsRbohC
DNDHE
QY

◂**Fig. 3** **a** Electrostatic surfaces showing N-terminal EF-hands for (PDB code: 3A8R) and four Rbohs models. Ca^{2+}-and Rac-binding residues in 3A8R are indicated in green and cyan, respectively. Conserved residues are shown in green and cyan while respective substitutions are shown in magenta and orange. Gray sphere denotes calcium ion. The electrostatic potential displayed is between −5 (red) and +5 (blue) kT e-1. **b** Ca^{2+}-binding residues and respective substitutions are shown as sticks in four Rboh models with reference to 3A8R. Conserved residues are shown in green while respective substitutions are shown in magenta. Gray sphere denotes calcium ion (color figure online)

accepted substitution, respectively. In all 19 Rbohs, D246 was conserved.

Another important binding partner of Rbohs was Rac GTPase. R273 and Y277 are important for Rac-binding in OsRbohB (Oda et al. 2010). Nine Rbohs (OsRbohA, OsRbohC, OsRbohF, OsRbohG, OsRbohI, AtRbohA, AtRbohD, AtRbohE and AtRbohF) were associated with R273Q substitution. In the case of OsRbohD and OsRbohE, R273H substitution was observed, whereas AtRbohB, AtRbohH and AtRbohJ were associated with the R273N substitution and AtRbohG with the R273K substitution. In AtRbohC, AtRbohI and OsRbohH, R273 was conserved and Y277 was conserved among the 19 Rbohs.

Two pairs of putative orthologous Rbohs; AtRbohD–OsRbohI and AtRbohF–OsRbohC were selected for further studies. The percentage sequence identities for the Rbohs; AtRbohD, AtRbohF, OsRbohC and OsRbohI with the PDB code: 3A8R were 53, 55, 54 and 62%, respectively (S2 File). Electrostatic surfaces of the protein (PDB code: 3A8R) and four Rboh models are shown in Fig. 3a. Among four Rbohs, N244D and R248H substitution were observed for Ca-binding sites in AtRbohD and OsRbohC; and R273Q in all four Rbohs (Fig. 3b). Further, OsRac1 was modeled and used for docking with Rboh models (Fig. 4). Docking studies show that monomer models of two Rbohs interact in the region corresponding to R273 and Y277, although no interactions were observed in the dimer state (Fig. 5).

To study the role of Ca^{2+} in folding of Rboh proteins, we performed ab initio folding simulations of EF-hands from AtRbohD and OsRbohB. Further, OsRbohB was studied in more detail. Weighted histogram analysis method (WHAM) was used to calculate the folding thermodynamics from replica exchange simulations for three OsRbohB forms: apo form, containing one and two Ca^{2+} ions. The thermodynamic parameters such as specific heat were computed to analyze the unfolding transition temperature. The specific heat of wild-type apoOsRbohB dimer showed a major peak at $T = 320$ K (Fig. 6a), which may refer to the global unfolding or melting transition of the protein. However, at lower temperatures, the protein appeared to be folded or at least partially folded. A minor peak emerged at a lower temperature, $T = 300$ K, which corresponds to a partial unfolding transition. Above $T = 340$ K, the protein appeared to be in a random coil state. Similarly, we computed the specific heat of two other forms (Fig. 6a) that showed two peaks in the specific heat at $T = 320$, 342 K and $T = 340$, ~352 K. The presence of two peaks may indicate the existence of intermediate states in the folding pathway. Compared to the wild-type apoOsRbohB, single Ca^{2+} OsRbohB has higher melting temperature (342 K), and hence stronger thermal stability, suggesting that Ca^{2+} enhance the folding of apoOsRbohB. Furthermore, the peaks in specific heat indicate the correspondence with the transition between energetic states, where energy is required to increase the potential of the complex (Fig. 6b). The presence of three low-energy states in the potential energy distribution with Gaussian-like peaks in three forms was also observed (Fig. 6c).

N-terminal region upstream of EF-hands

The N-terminal region upstream of EF-hands was highly diverse within Rbohs as observed from multiple sequence alignment. Further, ETA confirmed these results. In addition, disordered regions were observed in N-terminal region as indicated by disorder prediction (Fig. 7) and DynaMine analysis (S7 File). Unlike EF-hands, ab initio structural models for two pair of putative orthologs Rbohs and OsRbohB appear more basic and have different conformations (Fig. 8).

Transmembrane region

The TMDs (I–VI) were conserved including a pair of His residues in TMD-III and TMD-V (S4 File). The alignment of the six TMDs in 19 Rbohs connected by five loops (A–E) was similar to human Nox2. A large insertion comprising ~ 57 amino acid residues was observed in TMD-III of OsRbohF compared to the other Rbohs. The amino acid residues in TMD-II were most conserved among the six TMDs.

C-terminal NADPH-binding regions

Our results indicated two FAD and four NADPH-binding domains (I–IV) in 19 Rbohs (S4 File). The three-dimensional models corresponding to the C-terminal NADPH-binding domains were constructed using the crystal structure of the NADPH-binding domain of $gp91^{phox}$ (PDB: 3A1F) as template (S8 File). The sequence identities corresponding to the individual Rbohs with the sequence of 3A1F vary from 36 to 43%. Certain large insertions were observed in OsRbohH. The presence of large insertions between NADPH-I and NADPH-II precluded the possibility of constructing reliable models for some of these Rbohs. The FAD-binding regions were also excluded from the model due to the lack of suitable templates. The NADPH-I domain comprises

Fig. 4 Model of OsRac1 showing structural alignment with template. Magenta and green color indicate Rboh and template (PDB code: 3A8R), respectively (color figure online)

the glycine-rich motif (GXGXG), which is conserved in all 19 Rbohs (S4 File). In addition, other conserved residues among Rbohs correspond to human Nox2 (Pro-415 and Asp-500) and AtRbohD (Cys890).

Discussion

In the present study, ten Rbohs from *A. thaliana* and nine from *O. sativa* were analyzed using phylogenetic, evolutionary trace and sequence-structure approaches.

Phylogenetic analysis of Rboh family within plant kingdom

In the present study, an extensive phylogenetic analysis of 127 Rboh proteins from 26 plant species involving members of dicots, monocots and lower plants was conducted. In an earlier phylogenetic study, 50 ferric reduction oxidases (FRO) and 77 Noxes from plants were reported (Chang et al. 2016). This paper also suggests that though FROs are closely related proteins but still differ from the plant Noxes, and therefore the comprehensive phylogenetic analysis of Noxes conducted in the present paper holds significance. In the present study, representative members from red algae, Fabaceae, Brassicaceae, Salicaceae, Vitaceae, Poaceae, Solanaceae, Cucurbitaceae and Potamogetonaceae were distributed in ten phylogenetic clusters. The phylogeny along with functional data available for few Rboh proteins indicated by filled circles in Fig. 1 were used for functional analysis of Rbohs. Cluster 1 indicated a close relationship of Rbohs from legumes with red algae, suggesting the possibility of evolution of legume Rbohs from lower plants (red algae). In the evolutionary context, it also implies that Rbohs of land plants arose from lower plants dwelled in water. Among four red algae Rbohs, the role of only *C. crispus* CcRbohD in pathogen-induced ROS production has been established (Herve et al. 2006). The analysis of clusters 2 and 6 indicated duplications of *M. truncatula* Rboh proteins relative to AtRbohs. These observations were supported by the genome duplications in legumes relative to *A. thaliana* (Yan et al. 2003; Cannon et al. 2006). *Arabidopsis thaliana* AtRbohH and AtRbohJ are implicated to have pollen-specific functions (Potocký et al. 2007), while the role of only *M. truncatula* Rboh (MtRbohF) in root hair development has been reported (Marino et al. 2011). From sub-cluster 2B, only *O. sativa* OsRbohE is known to involve in biotic stress (Yoshie et al. 2005). It is interesting to infer that cluster 2 involves development and biotic stress-related Rbohs. Further, majority of the Rboh isoforms in this cluster belong to H and J subfamily both from dicots and monocots. In cluster 3, except OsRbohF, known to play role in defense (Lin et al. 2009b), the functions of other proteins are still not known. Subfamily E seems to be dominant among monocots and dicots in this cluster. Further, we noticed the emergence of abiotic stress-related functions with biotic stress (pathogen/wound) and development-related Rbohs in cluster 4. For example, *Z. mays* ZmRbohB-α and ZmRbohB-β are involved in abiotic stress (Lin et al. 2009a, b), *O. sativa* OsRbohA in drought and growth regulation (Wang et al.

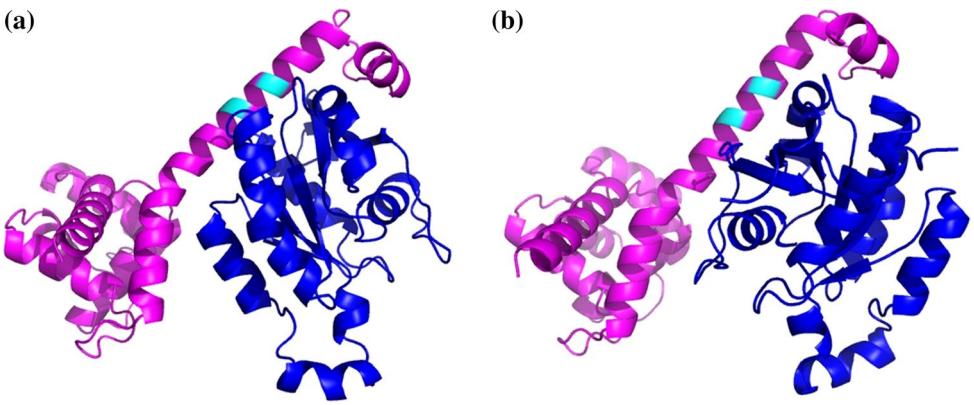


Fig. 5 Docked monomer Rboh models of **a** OsRbohC and **b** OsRbohI with OsRac1. Magenta and blue colors indicate Rboh and OsRac1, respectively. Cyan color indicates Rac-binding residues in Rbohs (color figure online)

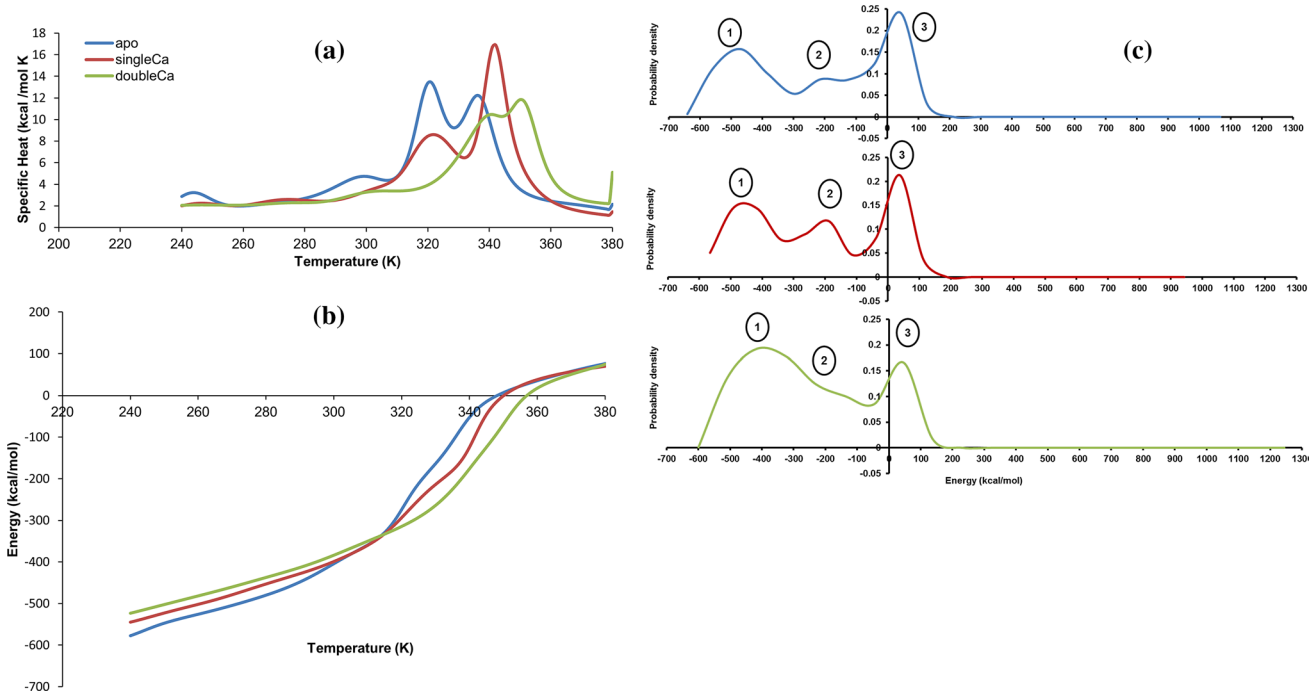


Fig. 6 Comparison of wild-type apoOsRbohB form with one and two Ca^{2+} forms **a** specific heat and **b** potential energy as a function of temperature, **c** potential energy distributions

2016), OsRbohC in drought (Kaur and Pati 2016), *H. vulgare* HvRbohA and HvRbohF1 in biotic stress from sub-cluster 4A (Yoshie et al. 2005; Trujillo et al. 2006; Lightfoot et al. 2008). The sub-cluster 4B which contains Rbohs involved in biotic stress from Solanaceae family; *L. esculentum* LeRboh1, *S. tuberosum* StRbohA and *N. benthamiana* NbRbohA (Sagi et al. 2004; Kumar et al. 2007; Yoshioka et al. 2003), while *N. tabacum* NtRbohF plays role in pollen tube growth (Potocký et al. 2007). Few Rbohs are known to be involved in abiotic stress (*A. thaliana* AtRbohF and *Citrullus colocynthis* CcRbohF) form a separate group in sub-cluster 4B. AtRbohF is the best example and the most studied Rboh possessing all the three kinds of function: abiotic stress, biotic stress and development (Marino et al. 2012). CcRbohF from Cucurbitaceae family is reported in drought tolerance (Si et al. 2010). The weak phylogenetic support at the start of cluster 4 further validates this observation and provides proof for this group's closeness with abiotic stress related monocot Rbohs. In cluster 5, the role of LesaRbohA is still unknown, whereas BoRbohD and BoRbohF in ethylene signaling and heavy metal stress have been reported (Jakubowicz et al. 2010). The critical role of CsRboh in brassinosteroid-induced stress tolerance has also been demonstrated (Xia et al. 2009). However, the role of only monocot Rboh (*Potamogeton crispus* PcRboh1) present in this cluster has not been identified yet. In the case of cluster 6, out of six proteins in sub-cluster 6A, only two (OsRbohB and OsRbohH) were reported to be drought-induced (Wang et al. 2013; Kaur and Pati 2016). We hypothesize that other monocot Rbohs (ZmRbohF, ZmRbohI, HvRbohB1 and HvRbohB2) may function in abiotic stress. The sub-cluster 6B which predominantly involves Rboh from B subfamily has wide range of functions. The role of *S. tuberosum* StRbohB in biotic stress (Kobayashi et al. 2006), *A. thaliana* AtRbohB in seed germination and after-ripening (Müller et al. 2009), *L. sativum* LesaRbohB in root development and auxin signaling (Müller et al. 2012), *M. truncatula* MtRbohA in nodule nitrogen fixation (Marino et al. 2011) and *P. vulgaris* PvRbohB in root growth and nodule nitrogen fixation (Montiel et al. 2012) were reported recently. Further, we observed that cluster 7 was represented by monocots, cluster 8 contained members of dicots, and cluster 9 was represented exclusively by members of legumes. Instead of three clusters as observed in the present study, they were placed in one cluster in an earlier study where only limited plant species were considered (Marino et al. 2011). The functions of most Rbohs from cluster 7 are still unknown. However, a recent analysis from our group has indicated the potential role of OsRbohI in abiotic stresses (Kaur and Pati 2016). On the other hand, members from clusters 8 are implicated in an array of biological activities. Most members are from Brassicaceae and belong to D subfamily. *Arabidopsis thaliana* AtRbohD is reported to play versatile roles in biotic and abiotic stress (Marino et al. 2012) while *C.*

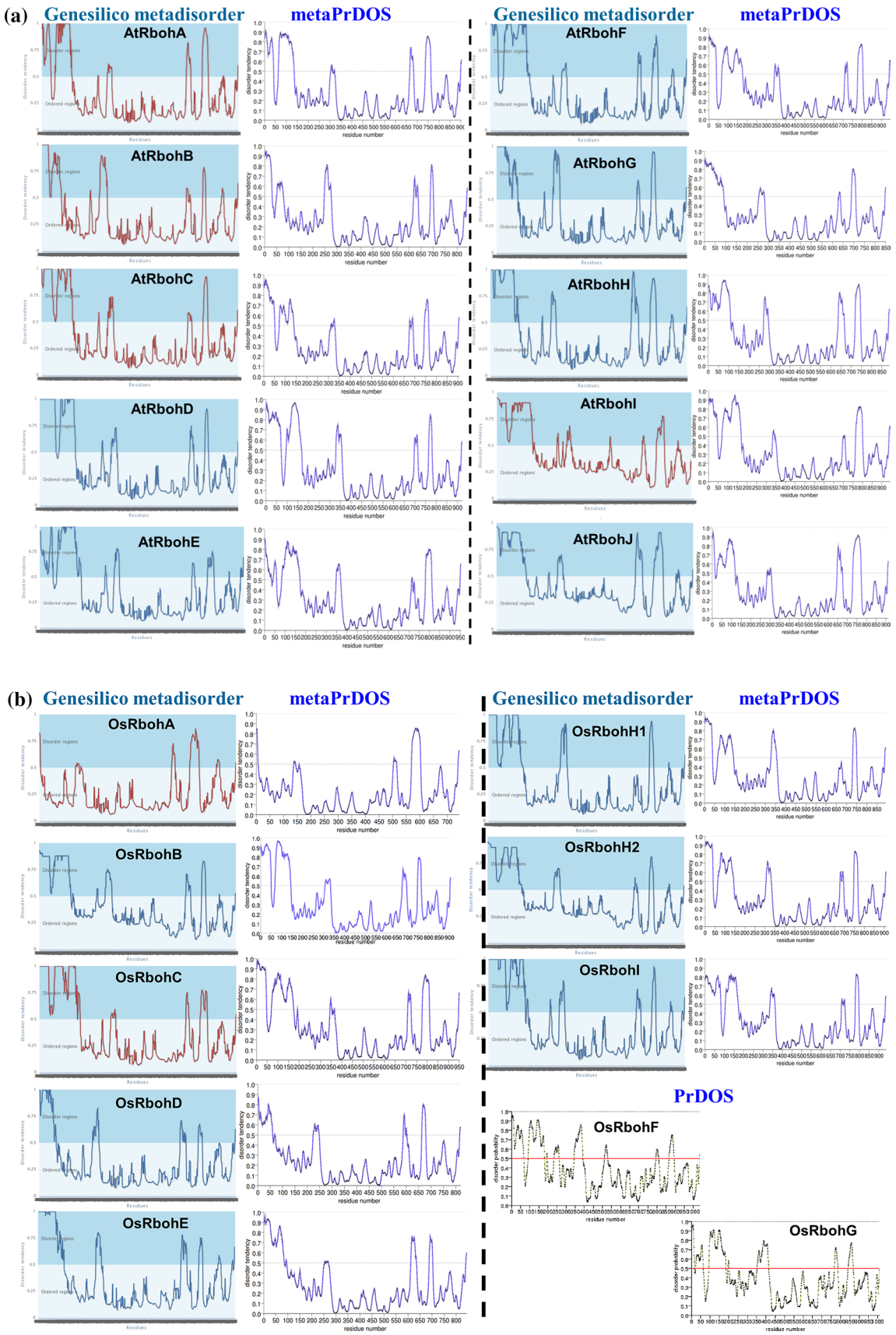


Fig. 7 Disorder prediction **a** AtRbohs and **b** OsRbohs using Genesilico metadisorder, metaPrDOS and PrDOS

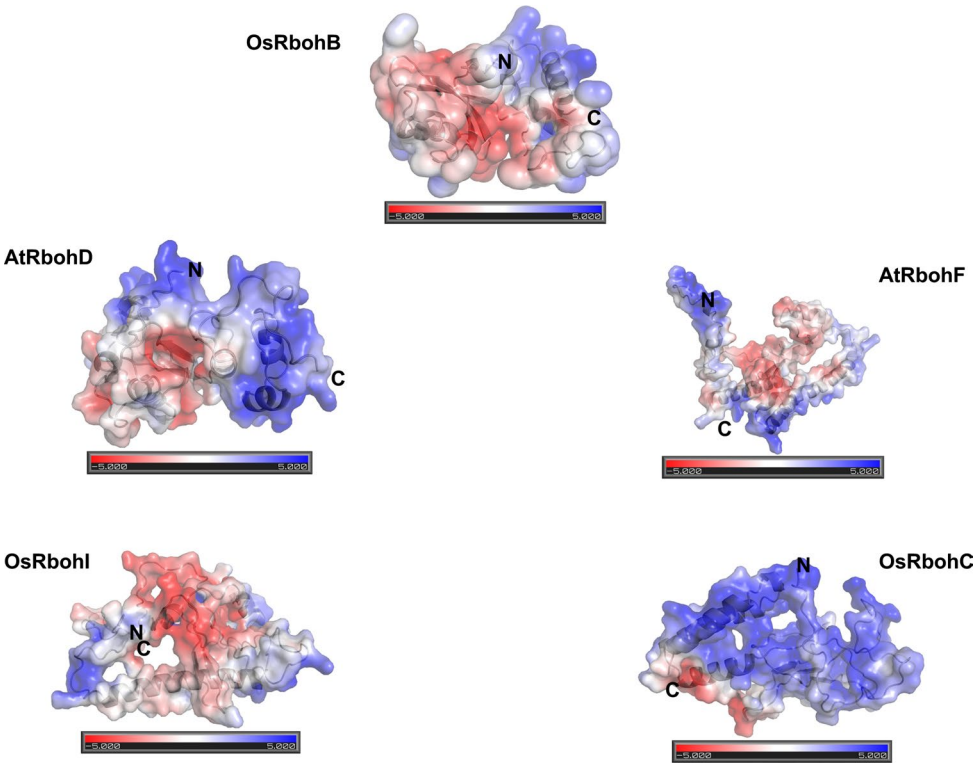


Fig. 8 Electrostatic surfaces showing N-terminal upstream of EF-hand-like motifs for (PDB code: 3A8R) and four Rbohs ab initio models. The electrostatic potential displayed is between −5 (red) and +5 (blue) kT e-1 (color figure online)

colocynthis CcRbohD in drought tolerance (Si et al. 2010). *Arabidopsis thaliana* AtRbohC plays role in root growth and cell wall integrity (Marino et al. 2012). The function of members in cluster 9 is still unknown. The divergence observed in the form of three clusters (cluster 7, 8 and 9) hints towards gene duplications among various members placed in three clusters. Such mechanisms of gene duplications and models; and their role in the evolution of novel gene functions have been highlighted earlier (Innan and Kondrashov 2010; Lawton-Rauh 2003). It has been observed that gene duplications may result due to nonfunctionalization, neofunctionalization (evolving novel functions) or subfunctionalization (partition of gene functions) of genes (Moore and Purugganan 2005; Rensing 2014). In our study, we also observed AtRbohH and AtRbohJ of cluster 2 are having pollen-specific function (Potocký et al. 2007) while MtRbohF of same cluster is linked to root hair development (Marino et al. 2011). The divergence of function of the members present in the same cluster also indicate that proteins encoded by duplicating Rboh genes may interact with different substrates leading to versatility in their functions (Huberts and van der Klei 2010). In addition, gene duplications also have implications in the evolution of regulatory networks (Babu et al. 2004). Therefore, gain or loss of function during gene duplications may result in new orloss of connections with the existing partners in the network. Further, the observed ancestral, invariant and discriminative residues among four dicot–monocot clusters from ancestral sequence reconstruction also hint towards a number of gene duplication events among Rbohs.

Sequence and structure analyses

In the present work, we have selected two model species, one representing dicots (Arabidopsis) and the other a member of monocots, i.e., rice. We identified the conserved and variable sequences and their link to structural features among their Rbohs. Using multiple sequence alignment (MSA), we observed many conserved and variable regions as well as secondary structural elements among 19 Rbohs, which may hint towards their functions. MSA shows N-terminal to be highly variable with no regular structure possibly contributing to the diverse functions of the Rbohs. The identification of orthologous genes is considered important in understanding the function of unknown genes, and hence is also explored in the present work (Kristensen et al. 2011). For this purpose, specific orthologous resources to identify orthologous sequences for each of ten AtRbohs and nine OsRbohs were investigated. This study provides important information not only on the orthologous proteins but on also diverse roles of Rbohs among members of dicots and monocots. The validation of these ortholog partners will further substantiate the relationship during the process of plant evolution.

Structural characterization based on the availability of crystal structure has been accepted means for the prediction of function of protein (Ramachandran and Dokholyan 2012). Rbohs are important transmembrane proteins involved in various biological functions that need proper characterization. We therefore analyzed 18 Rbohs to study the existing variations in the residues and predicted their structure using homology modeling. In the present work, emphasis was laid on N-terminal region of Rbohs involving Ca^{2+}- and Rac-binding sites critical for its activation and function.

In our study, we observed the following substitutions: N244D, R248K/Q/H/M in Ca^{2+}-binding sites with reference to OsRbohB. In earlier studies, it has been observed that the nature of amino acids in the six positions from the loop region of helix-loop-helix of Ca^{2+}-binding motifs affects the Ca^{2+}-affinity (Procyshyn and Reid 1994). N244D substitution at the loop region increased the Ca^{2+}-affinity in synthetic peptides (Procyshyn and Reid 1994) and lead to conformational changes in troponin C-based calcium biosensors (Mank et al. 2006, 2008). In addition, shift from monodentate to bidentate or vice versa in carboxylate-binding mode of Asp (D) also affect the Ca^{2+}-binding affinity and the catalytic activity and function of proteins (Dudev and Lim 2007). Apart from Ca^{2+}-binding proteins, N244D substitution is also known in effecting substrate binding of aspartate aminotransferase (Oue et al. 1999), conformational changes in phenylalanine hydroxylase and NAD-dependent D-lactate dehydrogenase (Carvalho et al. 2003; Shinoda et al. 2005). Further, R248Q/H substitutions can affect the helix stability in centrin, EF-hand containing centrosome-associated protein (Taillon and Jarvik 1995). A recent study revealed that R248Q substitution in EF-hand influence the structural flexibility and salt bridges in neuronal calcium sensor-1 protein (Zhu et al. 2014). R248M substitution involving charged to uncharged amino acid, is known to inhibit ability of eukaryotic translation initiation factor 5 (eIF5) in stimulating GTP hydrolysis (Paulin et al. 2001), affecting substrate binding by zinc-endopeptidases (Marie-Claire et al. 1997) and lactate monooxygenase (Sanders et al. 1999), and Rho-GAP activity of myosin myr 5 (Müller et al. 1997). Thermodynamics studies involving assessment of free energy changes will be helpful in understanding the effect of the above mutations on Ca^{2+}-binding as well as the stability of the mutants in the respective Rbohs.

The other component in the N terminus is the Rac-binding region, which is reported to play a very important role in the regulation of Rboh (Oda et al. 2010; Wong et al. 2007). The present study revealed R273N/Q/K/H substitutions in the Rac-binding sites in 18 Rbohs. Three Rbohs (AtRbohB, AtRbohH and AtRbohJ) showed similar kind of substitution (R273N), although they are implicated in different functions (Kaur et al. 2014). R273Q is known to prevent Rac-binding to N-terminal tetratricopeptide repeat motifs from $p67^{phox}$ (one of the components of human Nox) while R273K has weak interaction (Koga et al. 1999). In another experiment, two rice Rboh models (OsRbohI and OsRbohC) were docked with deduced structure of OsRac1. The rice Rbohs (OsRbohI and OsRbohC) are putative orthologs of AtRbohD and AtRbohF, respectively, linked to plant stress. The monomer models of two Rbohs show interaction in the region corresponding to R273 and Y277, although no interactions were observed in the dimer state. This might be due to the unstable nature of monomer, which needs to bind with some substrate as compared to dimer.

While analyzing the available information regarding various Rbohs, the involvement of AtRbohD in a range of biological functions such as plant development, biotic and abiotic stress was observed (Kaur et al. 2014). Hence, it is important for us to further study ab initio folding of its EF-hands, and hence the role of Ca^{2+} using replica exchange discrete molecular dynamics. In the current study, the critical role of Ca^{2+} in the folding of EF-hands was found. It is observed that EF-hands feature three helices instead of four in the absence of Ca^{2+}. It indicates the involvement of Ca^{2+} in specific folding of EF-hands. In WHAM analysis from replica exchange, trajectories of OsRbohB using its three forms (apo form, containing one and two Ca^{2+} ions) might suggest it as a two-state protein. It further provides evidence of strong link of number of Ca^{2+} ions with the folding event and its probable involvement in the regulation of Rbohs.

We observed disordered regions upstream of the EF-hand-like motifs in the N terminus of 19 Rbohs. This hints towards the presence of different binding partners. To get insights about their structural composition, ab initio structure prediction and electrostatic studies were conducted for two pair of putative orthologous Rbohs as well as for OsRbohB. We observed that the N-terminal region was rich in basic amino acids, which may participate in electrostatic interactions with acidic phospholipids as well as subcellular localization and function.

In the present study, six TMDs downstream of N-terminal region were observed from sequence analysis and in agreement with previous studies on Rbohs (Lin et al. 2009b; Trujillo et al. 2006). However, it has not been possible to predict the structure of TMDs due to non-availability of suitable templates for modeling. A pair of His residues involved in heme binding in the TMD-III and V was conserved among 19 Rbohs. Six TMDs involve five loops; two intracellular loops—B and D; and three extracellular loops—A, C and E. The B-loop has been recognized as an important structural element for human Nox2 function and plays a critical role in ROS production. Arg73, Arg80, Arg91, Arg92, Leu94 and Asp95 residues are involved in electron flow from NADPH across membrane in Nox2 (von Löhneysen et al. 2010). Among 19 Rbohs, Arg73 and Arg80 are conserved. Leu94 which is conserved in

human Noxes is replaced by a more bulky group, i.e., phenylalanine in all Rbohs. Previous study on $L \rightarrow F$ substitution has documented its effect on the binding properties of platelet glycoprotein Ibα for von Willebrand factor (Miller et al. 1992). However, the equivalent of Asp95 is conserved in all Rbohs, except OsRbohB and OsRbohH that are substituted by Asn. The $D \rightarrow N$ substitution in the loop region between TMDs have increased the signaling activity in Kaposi's sarcoma-associated herpesvirus G protein-coupled receptors (Ho et al. 2001). Further, the cysteine residues in E-loop of Nox4 that form disulphide bridges to maintain stability of the E-loop (Takac et al. 2011), are absent in Rbohs.

Apart from N-terminal and TMD, other component of Rbohs is C-terminal, which shows interaction with N-terminal from Rboh (Oda et al. 2010). C-terminal consists of FAD and NADPH-binding regions where NADPH-binding region is involved in electron transfer from the cytosolic NADPH (electron donor), through FAD and then across the membrane through hemes (in TMDs) to molecular oxygen (electron acceptor) leading to superoxide production (Kaur et al. 2014). In the present study, two FADs and four NADPH-binding regions in the C-terminal region of 19 Rbohs were observed and in agreement with earlier reports on Rbohs (Lin et al. 2009b; Trujillo et al. 2006). Structure analysis of the predicted C-terminal models shows a Rossmann fold for binding with NADPH. It consists of alternating β-strands and α-helical regions where all strands forms a central β-sheet. The glycine-rich motif (GXGXG) in NADPH-I domain was conserved among 19 Rbohs. Only the highly conserved leucine (L) at the first X position in the above motif in Rbohs was substituted by Ala and Gly in human Noxes. It has been proposed that the glycine-rich motif is involved in substrate binding, where substrate is ATP in bacterial sensor histidine kinases and their eukaryotic counterparts (Egger et al. 1997) and *S*-adenosyl-L-methionine (SAM) in SAM-dependent methyltransferases (Kozbial and Mushegian 2005). Another type of glycine-rich motif (GXXXG) has been observed in transmembrane α-helices and known to stabilize the oligomerization of several membrane proteins (Kleiger and Eisenberg 2002). Amino acid residues that are required for human Nox2 function (Pro-415 and Asp-500) were conserved in C-terminal region of Rbohs. A recent study has proposed the regulatory role of Cys890 from AtRbohD during plant defense response and *S*-nitrosylation of this residue inhibits the activity of the enzyme (Yun et al. 2011). The equivalent cysteine residue was conserved in the other 18 Rbohs and human Nox2 in NADPH-IV. This data suggest that the redox-based modification like *S*-nitrosylation at the conserved cysteine might regulate the activity of other 18 Rbohs and govern a feedback loop in the synthesis of reactive oxygen species.

Conclusion

The phylogenetic analysis corresponding to 127 Rbohs representing 26 plant species suggests a number of both ancient and recent gene duplication events that may contribute to functional versatility of plant NADPH oxidases. Comprehensive sequence analysis accompanied with evolutionary trace, orthologous identification and disorder studies identified the conserved and variable residues among Rbohs. These residues are important candidates for mutation and functional studies. The N-terminal domain is associated with variability relative to the transmembrane and C-terminal domains. The conserved cysteine residue (Cys890) in C-terminal domain in AtRbohD may regulate the activity of other Rbohs via redox-based modification like *S*-nitrosylation. Three-dimensional models corresponding to the N-terminal indicated deletions and substitutions associated with the Ca^{2+}- and Rac-binding sites. The role of Ca^{2+} in folding of EF-hands from Rboh proteins was also explored using discrete molecular dynamics. The present study provides crucial insights to design functional genomics experiments to validate the function of the plant NADPH oxidases.

Acknowledgements The authors acknowledge the financial support received under Innovation in Science Pursuit for Inspired Research (INSPIRE) Programme, Department of Science and Technology (DST), Government of India (Grant no. DST/INSPIRE Fellowship/2010[79]) and Fulbright-Nehru Doctoral and Professional Research Fellowship, United States-India Educational Foundation (USIEF), New Delhi, India (Grant no. 1663/DPR/2012-2013). The support from Department of Biotechnology (DBT), Government of India (Grant no. BT/PR13965/BRB/10/883/2010) is also acknowledged. The authors are thankful to Onur Dagliyan for his helpful suggestions during simulation sessions.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest.

Research involving human participants and/or animals This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This article does not contain any studies involving human participants.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402

Alva V, Nam S-Z, Söding J, Lupas AN (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Res 44(W1):W410–W415

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14(3):283–291

Baxter A, Mittler R, Suzuki N (2014) ROS as key players in plant stress signalling. J Exp Bot 65(5):1229–1240

Benkert P, Kunzli M, Schwede T (2009) QMEAN server for protein model quality estimation. Nucleic Acids Res 37(Web Server issue):W510–W514. doi:10.1093/nar/gkp322

Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, Scheix T, Spannagl M, Monaghan E, Nicholson C, Humphray SJ, Schoof H, Mayer KFX, Rogers J, Quétier F, Oldroyd GE, Debellé F, Cook DR, Retzel EF, Roe BA, Town CD, Tabata S, Van de Peer Y, Young ND (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proc Natl Acad Sci USA 103(40):14959–14964

Carvalho RN, Solstad T, Bjørgo E, Barroso JF, Flatmark T (2003) Deamidations in recombinant human phenylalanine hydroxylase. Identification of labile asparagine residues and functional characterization of Asn→Asp mutant forms. J Biol Chem 278(17):15142–15152

Chang Y-L, Li W-Y, Miao H, Yang S-Q, Li R, Wang X, Li W-Q, Chen K-M (2016) Comprehensive genomic analysis and expression profiling of the NOX gene families under abiotic stresses and hormones in plants. Genome Biol Evol 8(3):791–810

Chen R, Li L, Weng Z (2003) ZDOCK: an initial-stage protein-docking algorithm. Proteins 52(1):80–87. doi:10.1002/prot.10389

Chiu JC, Lee EK, Egan MG, Sarkar IN, Coruzzi GM, DeSalle R (2006) OrthologID: automation of genome-scale ortholog identification within a parsimony framework. Bioinformatics 22(6):699–707

Cilia E, Pancsa R, Tompa P, Lenaerts T, Vranken WF (2013) From protein sequence to dynamics and disorder with DynaMine. Nat Commun 4:2741. doi:10.1038/ncomms3741

Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res 36(Web Server issue):W197–W201. doi:10.1093/nar/gkn238

Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res 32(Web Server issue):W96–W99. doi:10.1093/nar/gkh354

DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA, USA. http://www.pymol.org

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465–W469. doi:10.1093/nar/gkn180

Ding F, Dokholyan NV (2006) Emergence of protein fold families through rational design. PLoS Comput Biol 2(7):e85

Ding F, Tsao D, Nie H, Dokholyan NV (2008) Ab initio folding of proteins with all-atom discrete molecular dynamics. Structure 16(7):1010–1018. doi:10.1016/j.str.2008.03.013

Dokholyan NV, Buldyrev SV, Stanley HE, Shakhnovich EI (1998) Discrete molecular dynamics studies of the folding of a protein-like model. Fold Des 3(6):577–587. doi:10.1016/S1359-0278(98)00072-8

Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res 32(Web Server issue):W665–W667. doi:10.1093/nar/gkh381

Dudev T, Lim C (2007) Effect of carboxylate-binding mode on metal binding/selectivity and function in proteins. Acc Chem Res 40(1):85–93

Egger LA, Park H, Inouye M (1997) Signal transduction via the histidyl-aspartyl phosphorelay. Genes Cells 2(3):167–184

Groom QJ, Torres MA, Fordham-Skelton AP, Hammond-Kosack KE, Robinson NJ, Jones JD (1996) rbohA, a rice homologue of the mammalian gp91phox respiratory burst oxidase gene. Plant J 10(3):515–522

Herve C, Tonon T, Collen J, Corre E, Boyen C (2006) NADPH oxidases in Eukaryotes: red algae provide new hints! Curr Genet 49(3):190–204. doi:10.1007/s00294-005-0044-z

Ho HH, Ganeshalingam N, Rosenhouse-Dantsker A, Osman R, Gershengorn MC (2001) Charged residues at the intracellular boundary of transmembrane helices 2 and 3 independently affect constitutive activity of Kaposi's sarcoma-associated herpesvirus G protein-coupled receptor. J Biol Chem 276(2):1376–1382

Huberts DHEW, van der Klei IJ (2010) Moonlighting proteins: an intriguing mode of multitasking. Biochim et Biophys Acta 1803(4):520–525

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11(2):97–108

Innis CA, Shi J, Blundell TL (2000) Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis. Protein Eng 13(12):839–847

Ishida T, Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res 35(Web Server issue):W460–W464. doi:10.1093/nar/gkm363

Ishida T, Kinoshita K (2008) Prediction of disordered regions in proteins based on the meta approach. Bioinformatics 24(11):1344–1348. doi:10.1093/bioinformatics/btn195

Jakubowicz M, Galganska H, Nowak W, Sadowski J (2010) Exogenously induced expression of ethylene biosynthesis, ethylene perception, phospholipase D, and Rboh-oxidase genes in broccoli seedlings. J Exp Bot 61(12):3475–3491. doi:10.1093/jxb/erq177

Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292(2):195–202. doi:10.1006/jmbi.1999.3091

Kaur G, Pati PK (2016) Analysis of cis-acting regulatory elements of respiratory burst oxidase homolog (Rboh) gene families in Arabidopsis and rice provides clues for their diverse functions. Comput Biol Chem 62:104–118

Kaur G, Sharma A, Guruprasad K, Pati PK (2014) Versatile roles of plant NADPH oxidases and emerging concepts. Biotechnol Adv 32(3):551–563

Kaur G, Singh P, Pati PK (2015) Integrating knowledge of bioinformatics in medicinal plant research. In: Keshavachandran R, Raji Radhakrishnan S (eds) Agriculture bioinformatics. NIPA Publishers, New Delhi, pp 91–142

Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4(3):363–371

Kleiger G, Eisenberg D (2002) GXXXG and GXXXA motifs stabilize FAD and NAD(P)-binding Rossmann folds through C(alpha)-H···O hydrogen bonds and van der Waals interactions. J Mol Biol 323(1):69–76

Kobayashi M, Kawakita K, Maeshima M, Doke N, Yoshioka H (2006) Subcellular localization of Strboh proteins and NADPH-dependent O2(−)-generating activity in potato tuber tissues. J Exp Bot 57(6):1373–1379

Kobayashi M, Ohura I, Kawakita K, Yokota N, Fujiwara M, Shimamoto K, Doke N, Yoshioka H (2007) Calcium-dependent protein kinases regulate the production of reactive oxygen species by potato NADPH oxidase. Plant Cell 19(3):1065–1080. doi:10.1105/tpc.106.048884

Koga H, Terasawa H, Nunoi H, Takeshige K, Inagaki F, Sumimoto H (1999) Tetratricopeptide repeat (TPR) motifs of p67(phox) participate in interaction with the small GTPase Rac and activation of the phagocyte NADPH oxidase. J Biol Chem 274(35):25051–25060

Kota P, Ding F, Ramachandran S, Dokholyan NV (2011) Gaia: automated quality assessment of protein structure models. Bioinformatics 27(16):2209–2215. doi:10.1093/bioinformatics/btr374

Kozbial PZ, Mushegian AR (2005) Natural history of *S*-adenosylmethionine-binding proteins. BMC Struct Biol 5:19

Kozlowski LP, Bujnicki JM (2012) MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinform 13:111. doi:10.1186/1471-2105-13-111

Kristensen DM, Wolf YI, Mushegian AR, Koonin EV (2011) Computational methods for gene orthology inference. Brief Bioinform 12(5):379–391

Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13(8):1011–1021

Kumar GNM, Iyer S, Knowles NR (2007) Strboh A homologue of NADPH oxidase regulates wound-induced oxidative burst and facilitates wound-healing in potato tubers. Planta 227(1):25–36

Lawton-Rauh A (2003) Evolutionary dynamics of duplicated genes in plants. Mol Phylogenet Evol 29(3):396–409

Lee D, Redfern O, Orengo C (2007) Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol 8(12):995

Lightfoot DJ, Boettcher A, Little A, Shirley N, Able AJ (2008) Identification and characterisation of barley (*Hordeum vulgare*) respiratory burst oxidase homologue family members. Funct Plant Biol 35(5):347–359. doi:10.1071/FP08109

Lin F, Ding H, Wang J, Zhang H, Zhang A, Zhang Y, Tan M, Dong W, Jiang M (2009a) Positive feedback regulation of maize NADPH oxidase by mitogen-activated protein kinase cascade in abscisic acid signalling. J Exp Bot 60(11):3221–3238

Lin F, Zhang Y, Jiang M-Y (2009b) Alternative splicing and differential expression of two transcripts of nicotine adenine dinucleotide phosphate oxidase B gene from *Zea mays*. J Integr Plant Biol 51(3):287–298

Mank M, Reiff DF, Heim N, Friedrich MW, Borst A, Griesbeck O (2006) A FRET-based calcium biosensor with fast signal kinetics and high fluorescence change. Biophys J 90(5):1790–1796

Mank M, Santos AF, Direnberger S, Mrsic-Flogel TD, Hofer SB, Stein V, Hendel T, Reiff DF, Levelt C, Borst A (2008) A genetically encoded calcium indicator for chronic in vivo two-photon imaging. Nat Methods 5(9):805–811

Marie-Claire C, Ruffet E, Antonczak S, Beaumont A, O'Donohue M, Roques BP, Fournié-Zaluski MC (1997) Evidence by site-directed mutagenesis that arginine 203 of thermolysin and arginine 717 of neprilysin (neutral endopeptidase) play equivalent critical roles in substrate hydrolysis and inhibitor binding. Biochemistry 36(45):13938–13945

Marino D, Andrio E, Danchin EGJ, Oger E, Gucciardo S, Lambert A, Puppo A, Pauly N (2011) A *Medicago truncatula* NADPH oxidase is involved in symbiotic nodule functioning. New Phytol 189(2):580–592

Marino D, Dunand C, Puppo A, Pauly N (2012) A burst of plant NADPH oxidases. Trends Plant Sci 17(1):9–15

Miller JL, Lyle VA, Cunningham D (1992) Mutation of leucine-57 to phenylalanine in a platelet glycoprotein Ib alpha leucine tandem repeat occurring in patients with an autosomal dominant variant of Bernard–Soulier disease. Blood 79(2):439–446

Montiel J, Nava N, Cárdenas L, Sánchez-López R, Arthikala M-K, Santana O, Sánchez F, Quinto C (2012) A *Phaseolus vulgaris* NADPH oxidase gene is required for root infection by Rhizobia. Plant Cell Physiol 53(10):1751–1767

Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol 8(2):122–128

Müller RT, Honnert U, Reinhard J, Bähler M (1997) The rat myosin myr 5 is a GTPase-activating protein for Rho in vivo: essential role of arginine 1695. Mol Biol Cell 8:2039–2053

Müller K, Carstens AC, Linkies A, Torres MA, Leubner-Metzger G (2009) The NADPH-oxidase AtrbohB plays a role in Arabidopsis seed after-ripening. New Phytol 184(4):885–897. doi:10.1111/j.1469-8137.2009.03005.x

Müller K, Linkies A, Leubner-Metzger G, Kermode AR (2012) Role of a respiratory burst oxidase of *Lepidium sativum* (cress) seedlings in root development and auxin signalling. J Exp Bot 63(18):6325–6334. doi:10.1093/jxb/ers284

Oda T, Hashimoto H, Kuwabara N, Akashi S, Hayashi K, Kojima C, Wong HL, Kawasaki T, Shimamoto K, Sato M, Shimizu T (2010) Structure of the N-terminal regulatory domain of a plant NADPH oxidase and its functional implications. J Biol Chem 285(2):1435–1445. doi:10.1074/jbc.M109.058909

Oue S, Okamoto A, Yano T, Kagamiyama H (1999) Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. J Biol Chem 274(4):2344–2349

Paulin FE, Campbell LE, O'Brien K, Loughlin J, Proud CG (2001) Eukaryotic translation initiation factor 5 (eIF5) acts as a classical GTPase-activator protein. Curr Biol 11(1):55–59

Potocký M, Jones MA, Bezvoda R, Smirnoff N, Zarsky V (2007) Reactive oxygen species produced by NADPH oxidase are involved in pollen tube growth. New Phytol 174(4):742–751. doi:10.1111/j.1469-8137.2007.02042.x

Proctor EA, Ding F, Dokholyan NV (2011) Discrete molecular dynamics. Wiley Interdiscip Rev Comput Mol Sci 1(1):80–92

Procyshyn RM, Reid RE (1994) A structure/activity study of calcium affinity and selectivity using a synthetic peptide model of the helix-loop-helix calcium-binding motif. J Biol Chem 269(3):1641–1647

Ramachandran S, Dokholyan NV (2012) Homology modeling: generating structural models to understand protein function and mechanism. In: Dokholyan NV (ed) Computational modeling of biological systems. Springer, New York, pp 97–116

Ramachandran S, Kota P, Ding F, Dokholyan NV (2011) Automated minimization of steric clashes in protein structures. Proteins 79(1):261–270. doi:10.1002/prot.22879

Rensing SA (2014) Gene duplication as a driver of plant morphogenetic evolution. Curr Opin Plant Biol 17:43–48

Rodrigues JP, Levitt M, Chopra G (2012) KoBaMIN: a knowledge-based minimization web server for protein structure refinement. Nucleic Acids Res 40(Web Server issue):W323–W328. doi:10.1093/nar/gks376

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738. doi:10.1038/nprot.2010.5

Sagi M, Davydov O, Orazova S, Yesbergenova Z, Ophir R, Stratmann JW, Fluhr R (2004) Plant respiratory burst oxidase homologs impinge on wound responsiveness and development in *Lycopersicon esculentum*. Plant Cell 16(3):616–628

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779–815. doi:10.1006/jmbi.1993.1626

Sanders SA, Williams CH, Massey V (1999) The roles of two amino acid residues in the active site of L-lactate monooxygenase. Mutation of arginine 187 to methionine and histidine 240 to glutamine. J Biol Chem 274(32):22289–22295

Schneider A, Dessimoz C, Gonnet GH (2007) OMA Browser—exploring orthologous relations across 352 complete genomes. Bioinformatics 23(16):2180–2182

Shinoda T, Arai K, Shigematsu-Iida M, Ishikura Y, Tanaka S, Yamada T, Kimber MS, Pai EF, Fushinobu S, Taguchi H (2005) Distinct conformation-mediated functions of an active site loop in the catalytic reactions of NAD-dependent D-lactate dehydrogenase and formate dehydrogenase. J Biol Chem 280(17):17068–17075

Shirvanyants D, Ding F, Tsao D, Ramachandran S, Dokholyan NV (2012) Discrete molecular dynamics: an efficient and versatile simulation method for fine protein characterization. J Phys Chem B 116(29):8375–8382. doi:10.1021/jp2114576

Si Y, Dane F, Rashotte A, Kang K, Singh NK (2010) Cloning and expression analysis of the Ccrboh gene encoding respiratory burst oxidase in *Citrullus colocynthis* and grafting onto *Citrullus lanatus* (watermelon). J Exp Bot 61(6):1635–1642

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:Artn 539. doi:10.1038/Msb.2011.75

Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Web Server issue):W244–W248. doi:10.1093/nar/gki408

Sonnhammer EL, Östlund G (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res 43(D1):D234–D239

Taillon BE, Jarvik JW (1995) Helix mutations in the centrosome-associated EF-hand protein centrin. Protoplasma 189:203–215

Takac I, Schröder K, Zhang L, Lardy B, Anilkumar N, Lambeth JD, Shah AM, Morel F, Brandes RP (2011) The E-loop is involved in hydrogen peroxide formation by the NADPH oxidase Nox4. J Biol Chem 286(15):13304–13313

Takeda S, Gapper C, Kaya H, Bell E, Kuchitsu K, Dolan L (2008) Local positive feedback regulation determines cell shape in root hair cells. Science 319(5867):1241–1244. doi:10.1126/science.1152505

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725–2729. doi:10.1093/molbev/mst197

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25(24):4876–4882. doi:10.1093/nar/25.24.4876

Trujillo M, Altschmied L, Schweizer P, Kogel K-H, Hückelhoven R (2006) Respiratory burst oxidase homologue A of barley contributes to penetration by the powdery mildew fungus *Blumeria graminis* f. sp. hordei. J Exp Bot 57(14):3781–3791

von Löhneysen K, Noack D, Wood MR, Friedman JS, Knaus UG (2010) Structural insights into Nox4 and Nox2: motifs involved in function and cellular localization. Mol Cell Biol 30(4):961–975

Wang G-F, Li W-Q, Li W-Y, Wu G-L, Zhou C-Y, Chen K-M (2013) Characterization of Rice NADPH oxidase genes and their expression under various environmental conditions. Int J Mol Sci 14(5):9440–9458

Wang X, Zhang MM, Wang YJ, Gao YT, Li R, Wang GF, Li WQ, Liu WT, Chen KM (2016) The plasma membrane NADPH oxidase OsRbohA plays a crucial role in developmental regulation and drought-stress response in rice. Physiol Plant 156:421–443

Wong HL, Pinontoan R, Hayashi K, Tabata R, Yaeno T, Hasegawa K, Kojima C, Yoshioka H, Iba K, Kawasaki T, Shimamoto K (2007) Regulation of rice NADPH oxidase by binding of Rac GTPase to its N-terminal extension. Plant Cell 19(12):4022–4034

Xia X-J, Wang Y-J, Zhou Y-H, Tao Y, Mao W-H, Shi K, Asami T, Chen Z, Yu J-Q (2009) Reactive oxygen species are involved in brassinosteroid-induced stress tolerance in cucumber. Plant Physiol 150(2):801–814

Yan HH, Mudge J, Kim DJ, Larsen D, Shoemaker RC, Cook DR, Young ND (2003) Estimates of conserved microsynteny among the genomes of Glycine max, *Medicago truncatula* and *Arabidopsis thaliana*. Theor Appl Genet 106(7):1256–1265

Yoshie Y, Goto K, Takai R, Iwano M, Takayama S (2005) Function of the rice gp91phox homologs OsrbohA and OsrbohE genes in ROS-dependent plant immune responses. Plant Biotechnol 22:127–135 **(SRC—GoogleScholar)**

Yoshioka H, Numata N, Nakajima K, Katou S, Kawakita K, Rowland O, Jones JDG, Doke N (2003) Nicotiana benthamiana gp91phox homologs NbrbohA and NbrbohB participate in H_2O_2 accumulation and resistance to *Phytophthora infestans*. Plant Cell 15(3):706–718

Yun B-W, Feechan A, Yin M, Saidi NBB, Le Bihan T, Yu M, Moore JW, Kang J-G, Kwon E, Spoel SH, Pallas JA, Loake GJ (2011) *S*-nitrosylation of NADPH oxidase regulates cell death in plant immunity. Nature 478(7368):264–268

Zhu Y, Wu Y, Luo Y, Zou Y, Ma B, Zhang Q (2014) R102Q mutation shifts the salt-bridge network and reduces the structural flexibility of human neuronal calcium sensor-1 protein. J Phys Chem B 118(46):13112–13122