PNAS
www.pnas.org

Supplementary Information for
"A mycorrhiza-associated receptor-like kinase with an ancient origin in the green lineage".

Authors: Héctor Montero*, Tak Lee, Boas Pucker, Gabriel Ferreras, Giles Oldroyd, Samuel Brockington, Akio Miyao, Uta Paszkowski*.

*To whom correspondence may be addressed. Email: hm530@cam.ac.uk (H.M.), up220@cam.ac.uk (U.P.)

This PDF file includes:

Supplementary extended methods
Description of supplementary datasets S1 to S4
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S4

Supplementary extended methods

Molecular characterization of *ark2* mutant allele. All experiments were conducted using *Oryza sativa* ssp. Japonica cv. Nipponbare as a background cultivar. A mutant allele for *OsARK2* was identified from the *Tos17* insertional mutant collection (ID: NG0028, ref. 1). From a segregating population, homozygous plants were identified by PCR genotyping. Genomic DNA (gDNA) extraction for genotyping was carried out following a modified sucrose solution extraction protocol originally developed for *A. thaliana* tissues (2). Briefly, a 1 cm-piece of rice leaf was collected and frozen in liquid nitrogen. Leaf pieces were lysed in 2 mL safe-lock microcentrifuge tubes with glass beads using a Qiagen TissueLyser II (Hilden, Germany) and mixed with 200 µL sucrose gDNA extraction buffer containing 50 mM Tris-HCl, pH 7.5; 300 mM NaCl and 300 mM sucrose. Samples were incubated in a heat block for 10 min at 95°C and centrifuged at 14,000 rpm for 2 min. Supernatant containing extracted gDNA was used as a template for PCR genotyping with recombinant *Taq* DNA Polymerase (Invitrogen, Carlsbad, USA) in a PTC-225 Peltier thermal cycler (Watertown, USA). PCR conditions were: 94°C initial denaturation, 35 cycles of 3 min denaturation at 94°C, 40 s annealing at 56°C, 40 s elongation at 72°C followed by 2 min of final elongation at 72°C. 1 µg mL^{-1} ethidium bromide agarose gel electrophoresis (1% agarose) was carried out over PCR products. Gels were visualized under UV light and images were acquired using a GBox Chemi 16 Bio Imaging System (Syngene, Cambridge, United Kingdom). Oligonucleotides for genotyping *ark2* were: *ark2*-NG0028 Fwd, 5'-AACGGGTTATTTCTTGCCCT-3' and *ark2*-NG0028 Rev, 5'-CACGTTGCCTTCTCTGATGA-3'. To amplify the junction between the *OsARK2* gene and the *Tos17* element, the *ark2*-NG0028 Rev primer was used in conjunction with a *Tos17*-specific primer (5'-GACAACACCGGAGCTATACAAATCG-3'). To establish the exact location of the *Tos17* insertion, excised electrophoresis bands were purified using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany) following manufacturer's instructions and 5 µL of 10 ng $µL^{-1}$ of DNA and 3.2 pmol $µL^{-1}$ of gene-specific oligonucleotides were submitted for an external Sanger sequencing service (Source BioScience, Cambridge, United Kingdom). A previously characterized *Tos17* insertional mutant allele for *OsARK1* (*ark1-2*, NF4582, ref. 3, here referred simply as *ark1*) was employed in this study and crossed with the *ark2* allele to generate the *ark1/ark2* double knockout (*dKO*) line. Mutant plants had no evident alterations in growth and morphology.

Plant growth conditions, AM inoculation and AM fungal colonization assessments. The AM fungal model species *R. irregularis* (DAOM197198) was employed for all inoculation assays. Commercially available *R. irregularis* aseptic spores in suspension (Mycorise® ASP, Premier Tech Biotechnologies, Rivière-du-loup, Canada) were used as the source of spore inoculum for most assays and spores extracted from *Agrobacterium rhizogenes* transformed carrot hairy root cultures (4) were used for the RNA-seq experiment. Aliquoted inoculum was applied onto the substrate of half-filled cones (12 cm depth, 2.5 cm diameter), after which sand substrate was replenished and germinated seedlings transferred to the containers. Plants were watered every second day, the first two weeks with reverse osmosis water (R-O H_2O) followed by low Pi fertilization every other watering day with half-strength Hoagland solution containing 25 µM KH_2PO_4. Time of harvesting varied per experiment and it is specified in the main text. Different time points and inoculum pressures were assayed following the guidelines of (5). To stain AM fungal structures for microscopy quantification, root pieces were incubated at 95°C in 10% (w/v) KOH for 30 min. After rinsing with diH_2O, roots were incubated in 0.3 M HCl for 30 min at room temperature. HCl was removed and 0.1% (w/v) trypan blue (Sigma-Aldrich, St. Louis, USA) diluted in a 2:1:1 lactic acid/glycerol/diH_2O solution was added. Roots in trypan blue solution were incubated for 8 min at 95°C. Trypan blue was removed and roots were washed in 50% v/v acidic glycerol. Ten 2-cm root pieces per biological replicate were mounted in a glass microscope slide. Quantification of fungal colonization took place by recording the presence of AM fungal structures along 100 different visual fields under a 20× magnification objective using a GXM-L2800 microscope (GT Vision, Stanfield, United Kingdom) and expressed as percentage of the total root length scored.

To visualize AM fungal structures using confocal laser scanning microscopy (CLSM), roots were treated with Wheat Germ Agglutinin (WGA)-Alexa Fluor™ 633 (Invitrogen, Carlsbad, USA)

exhibiting far-red fluorescence. Roots were excised and incubated in 50% (v/v) ethanol for one hour. Ethanol was removed and 20% (w/v) KOH solution added. After two days, KOH was removed and roots were rinsed with diH_2O after which samples were incubated for two hours in 0.1 M HCl. Roots were rinsed with 1× phosphate-buffered saline (PBS, pH 7.4) solution. A 0.2 µg mL^{-1} WGA-Alexa Fluor™ 633 solution in 1× PBS was added and samples were incubated at 4°C in the dark for one week before imaging. Visualization of roots took place using a Leica TCS SP8 (Leica Microsystems, Wetzlar, Germany). WGA-Alexa Fluor™ 633 was detected using WLL with an excitation wavelength of 630 nm (5.6% laser power), and emitted wavelengths collected at 650-730 nm. Images of arbusculated cells were taken at the point of maximal arbuscule expansion at a line average of 2 and dimension of 2048 × 2048 pixels. Roots were observed using a 40× water immersion objective. Image processing was carried out using the Fiji package under the ImageJ software license (6). The source of roots was from the RNA-seq experiment. Three random plants per genotype were selected for imaging and one representative fully branched arbuscule per genotype is presented.

Gene expression assays. For RNA extraction and cDNA synthesis, root tissues of *O. sativa* were ground using mortar and pestle in liquid nitrogen and 100 mg were used for RNA extraction following a TRIzol reagent-based method as described in (3). After testing for RNA integrity under 1.5% agarose gel electrophoresis and quantifying RNA concentrations in a NanoDrop™ 2000 spectrophotometer (ThermoFisher, Waltham, USA), 1 µg of RNA was treated with DNase I (Invitrogen, Carlsbad, USA) to remove genomic contamination following manufacturer's guidelines. First-strand cDNA synthesis was carried out using SuperScript™ II reverse transcriptase (Invitrogen, Carlsbad, USA) according to manufacturer's instructions.
RT-PCR gene expression assays were carried out using recombinant *Taq* DNA Polymerase (Invitrogen, Carlsbad, USA) in a PTC-225 Peltier thermal cycler (Watertown, USA). PCR conditions were: 94°C initial denaturation, 35 cycles of 3 min denaturation at 94°C, 40 s annealing at 56°C, 40 s elongation at 72°C followed by 2 min of final elongation at 72°C. 1 µg mL^{-1} ethidium bromide agarose gel electrophoresis (1.5% agarose) was carried out over PCR products. Gels were visualized under UV light and images were acquired using a GBox Chemi 16 Bio Imaging System (Syngene, Cambridge, United Kingdom). Oligonucleotides used were *OsARK2* LP 5'-ATGGGCAGGAGGAATGTTT-3', *OsARK2* RP 5'-CCGACAATTTGGGTTCCATATC-3'. Gene expression of the housekeeping gene *OsGAPDH* (*LOC_Os08g03290*) was measured in parallel to *OsARK2* using the following oligonucleotides: *OsGAPDH* LP 5'-AGGTTCTTCCTGATTTGAATGG-3', *OsGAPDH* RP 5'-CAACTGCACTGGACGGCTTA-3'. For this assay, RNA extracted from ten WPI plants inoculated with 250 spores was employed.
qRT-PCR gene expression assays were carried out using GoTaq® G2 DNA Polymerase (Promega) in a CFX Connect™ Real-Time System (Singapore). qRT-PCR cycling conditions were: 96°C initial denaturation, 40 cycles of 30 s denaturation at 96°C, 60 s annealing at 59°C, 30 s elongation at 72°C followed by 1 min of final elongation at 95°C. The average relative transcript levels of three technical replicates per sample were normalized to the geometric mean of amplification of three housekeeping genes: *OsCYCLOPHILIN2* (*LOC_Os02g02890*), *OsPOLYUBIQUITIN* (*LOC_Os06g46770*) and *OsGAPDH*. Expression values are displayed as a function of *OsCYCLOPHILIN2*. Oligonucleotides employed were; *OsCYCLOPHILIN2*, Fwd 5'-GTGGTGTTAGTCTTTTTATGAGTTCGT-3', Rev 5'-ACCAAACCATGGGCGATCT-3'; *OsPOLYUBIQUITIN*, Fwd 5'-CATGGAGCTGCTGCTGTTCTAG-3', Rev 5'-CAGACAACCATAGCTCCATTGG-3'; *OsGAPDH*, Fwd 5'-CTGATGATATGGACCTGAGTCTACTTTT-3', Rev 5'-CAACTGCACTGGACGGCTTA-3'; *OsARK1* Fwd 5'-ACTTGTGCCAGTGACATCTAC-3', Rev 5'-CACATCCCTGGCCTTCTTT-3'; *OsARK2* Fwd 5'-TGGCCTTTGGTTCCTTCA-3', Rev 5'-CCGACAATTTGGGTTCCATATC-3'; *RiEF*, Fwd 5'-GCTATTTTGATCATTGCCGCC-3', Rev 5'-TCATTAAAACGTTCTTCCGACC-3'. For validation of RNA-seq experiment, root material employed corresponded to samples from colonization experiment shown in Fig. 1E (inoculated with 300 *R. irregularis* spores). qRT-PCR oligonucleotides used were: *LOC_Os03g19610* Fwd 5'-GGCGACAAGATTGCCTTCTA-3', Rev 5'-CGATGTTGCCCTGGATGTAA-3'; *LOC_Os06g11240*, Fwd 5'-TGTCACCGTTCATCGACTATATG-3', Rev 5'-CACCATGTGGCAGTAGAGAAA-3';

LOC_Os04g55420, Fwd 5'-GCAGGAGAAGACGGAGTTTATT-3', Rev 5'-CCATGGGACGTGGAGATATTG-3'; *LOC_Os11g36790*, Fwd 5'-GGCAAAGCTTCGACGAGATA-3', Rev 5'-CACCAGAAGAACGAGCAGATAA-3'.

RNA-seq assays. RNA was extracted using the TRIzol method as before. The experiment had 10-12 biological replicates per genotype and 10 final replicates per genotype were selected based on RNA yield. For both RNA-seq assays and subsequent independent gene expression validation experiment, 300 spores were used as inoculum and harvesting took place at six WPI. For RNA-seq library preparation and sequencing, RNA integrity was monitored on 1% agarose gels, using the NanoPhotometer® spectrophotometer (IMPLEN, Los Angeles, USA) and the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, USA). 1 µg RNA per sample was used for library preparation generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Hitchin, United Kingdom) following manufacturer's instructions. Library quality was monitored on the Agilent Bioanalyzer 2100 system. After cluster generation on a cBot Cluster Generation System using PE Cluster Kit cBot-HS (Illumina, San Diego, USA), the library preparations were sequenced on an Illumina NovaSeq 6000 System (Illumina, San Diego, USA) and paired-end reads were generated. RNA-seq library preparation and sequencing were performed by Novogene (Cambridge, United Kingdom).
Raw RNA-seq reads of FASTQ format were processed to remove low quality reads and reads containing adapter and poly-N sequences. Q20, Q30 and GC content of the clean data were calculated. Error rates between samples were homogeneous averaging 0.021% and with an average GC content of 51.7%. Paired-end clean reads were mapped to the *O. sativa* Nipponbare reference genome (Os-Nipponbare-Reference-IRGSP-1.0, ref. 7) using STAR v2.7.3a (8). The average mapping rate was of 81.9%. The expression levels of genes were calculated in raw counts and FPKM (Fragments Per Kilobase of exon model per Million reads mapped) with featureCounts in R package Rsubread (version 2.0.0) (9). TPM (Transcripts Per Million) values were calculated from raw counts using a custom Python script. Genes with CPM (Counts Per Million) values (calculated from R package edgeR 3.26.8, ref. 10) lower than 0.5 in more than 10 samples were filtered out to avoid confounding results in differential expression analysis due to small expression changes. The remaining 27,498 genes were subjected to pairwise differential expression analysis between three genotypes (*ark1*, *ark2*, *dKO*) in relation to expression levels in the wild-type using the R package DESeq2 v1.26.0 (11) employing the raw count numbers. Genes with log2FC greater than 1 or less than −1 that had the significance of Benjamini-Hochberg false discovery rate (FDR) corrected *p* value lower than 0.05 were classified as DEGs. A total of 142 genes were upregulated and a much higher number of 566 genes were downregulated in any of the three mutant genotypes compared to the wild-type, arguably due to differences in colonization levels. A total of 379 DEGs were identified in the *dKO* (97 upregulated and 282 downregulated). Correlation analyses were implemented for each gene comparing percentage of arbuscules and expression levels across replicates. The correlation between the expression level of genes (in TPM) in the different genotypes and percentage of arbuscule colonization was determined with Spearman's correlation coefficient using a *p* value of 0.05 as a significance cut-off (corresponding to Spearman's R ≥ 0.631, Dataset S2). Of the downregulated genes in the *dKO*, 30% still had positive correlation between gene expression and arbuscule levels and were therefore unlikely to directly depend on *OsARK1* and/or *OsARK2.* This prompted us to refine the dataset focusing on genotype effects as opposed to transcriptional responses solely associated with differential colonization levels. We identified those genes that had no-overlap in expression comparing all ten biological replicates per genotype in the wild-type-*dKO* comparison, resulting in 37 DEGs. The annotations of the genes were collected from three different sources: Gene Ontology (GO, ref. 12) terms used in (13), GO slim (14) terms from http://rice.plantbiology.msu.edu/ and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (15).

Phylogenetic analyses. Protein sequences from the different RLK subfamilies under investigation were initially retrieved from the iTAK database (http://itak.feilab.net/cgi-bin/itak/index.cgi, ref. 16). This database does not contain information of plant genomes released after 2018. Therefore, subsequent orthologue searches were carried out employing SymDB

(www.polebio.lrsv.ups-tlse.fr/symdb/, ref. 17) and Phytozome v13 (https://phytozome-next.jgi.doe.gov/). Additional orthologue searches among the less represented non-flowering plants and basal angiosperms were carried out using the One Thousand Plant Transcriptomes (1KP) database (18) employing the built-in tools BLASTP (https://db.cngb.org/blast/blast/blastp/) and TBLASTN (https://db.cngb.org/blast/blast/tblastn). When several genomes were available for a given genus in angiosperms, one species was selected to achieve a balanced representation. Best reciprocal BLAST hits were considered as orthologues. When sequences were incomplete or had low quality, the following resources were used to find better quality sequences; NCBI standard protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins), NCBI translated BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=tblastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and UniProt BLAST (https://www.uniprot.org/blast/).
For phylogenetic analyses, kinase domain amino acid sequences were aligned using MAFFT v7 (19).The alignment was trimmed to a minimal occupancy of 10% via pxclsq (20) and phylogenetic tree construction was performed using RAxML-NG v9 (21) applying the LG+G8+F model and 1000 bootstrapping rounds. All included sequences are listed in Dataset S3 following the same order they appear in the phylogenetic trees (Datasets S1 and S4). The outgroups of the phylogenetic tree of Figure 1A correspond to a collection of 210 kinase sequences from land plants, streptophyte algae and chlorophyte algae. The outgroups of the phylogenetic tree of Figure 2B consist of seven kinase sequences from chlorophyte algae. The mycorrhizal status of the different plant species was assigned based on the recommended criteria established in (22) and is depicted in the detailed phylogenetic tree of Dataset S1.

Identification of protein motifs and domains. The SPARK domain consists of a conserved arrangement of cysteines interspaced by otherwise variable sequences and it was first identified via protein alignments of SPARK-I subfamily members. The variable sequences of the domain render it not amenable for BLAST searches and conceivable hindered an earlier identification. In addition, the SPARK domain does not have homology to known domains making homology modelling attempts unfruitful. We preliminarily identified SPARK domain harbouring proteins outside the SPARK-I subfamily by performing a pattern search using the signature arrangement of cysteines, yielding a list of sequences. A multiple protein sequence alignment of these sequences was used to build a Hidden Markov Model (HMM)-based profile generated with HMMER3 (http://hmmer.org/) as described in (23) and a Pfam entry for the SPARK domain was created (https://pfam.xfam.org/family/PF19160). This allowed for the identification of all SPARK domain-containing proteins from UniProt. The alignment of SPARK domain-containing proteins from the SPARK-I subfamily was performed using the built-in MAFFT tool in Jalview v2.10.5 (24). All sequences employed for the alignment are listed in Dataset S3. The occurrence of the SPARK domain in different taxonomic groups as well as its domain architecture configurations were extracted from InterPro (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR043891/).
Signal peptides and transmembrane regions were predicted using Phobius (http://phobius.sbc.su.se/, ref. 25). Glycosylphosphatidylinositol (GPI)-anchored proteins were predicted using PredGPI (http://gpcr.biocomp.unibo.it/predgpi/, ref. 26). Kinase domains were predicted using ScanProsite (https://prosite.expasy.org/scanprosite/, ref. 27).

Statistical analysis. Statistically significant differences were assessed using the non-parametric Kruskal-Wallis test followed by Dunn's multiple comparison post hoc tests in *R* (v *4.0.0.*). We displayed all data points in every graph.

Description of supplementary datasets S1 to S4

Dataset S1: detailed phylogenetic tree of Fig. 1A. Colour coding is as in Fig. 1A. Sequence IDs are followed by plant species names and protein identities. Plants that do not form AM symbiosis or that engage in different types of mycorrhizal associations are indicated. EcM, ectomycorrhizal symbiosis; ErM, ericoid mycorrhiza; OM, orchid mycorrhiza; NM-AM, facultative arbuscular mycorrhizal; NM, non-mycorrhizal.
Dataset S2: RNA-seq data.

Dataset S3: list of sequences used for phylogenetic trees.
Dataset S4: detailed phylogenetic tree of Fig. 2B. Sequence IDs are followed by identity of plant species, proteins and RLK subfamilies.

SI References

(1) Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., ... & Hirochika, H. (2003). Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. The Plant Cell, 15(8), 1771-1780.

(2) Berendzen, K., Searle, I., Ravenscroft, D., Koncz, C., Batschauer, A., Coupland, G., ... & Ülker, B. (2005). A rapid and versatile combined DNA/RNA extraction protocol and its application to the analysis of a novel DNA marker set polymorphic between Arabidopsis thaliana ecotypes Col-0 and Landsberg erecta. *Plant Methods*, *1*(1), 4.

(3) Roth, R., Chiapello, M., Montero, H., Gehrig, P., Grossmann, J., O'Holleran, K., ... & Paszkowski, U. (2018). A rice Serine/Threonine receptor-like kinase regulates arbuscular mycorrhizal symbiosis at the peri-arbuscular membrane. Nature communications, 9(1), 1-12.

(4) Bécard, G., & Fortin, J. A. (1988). Early events of vesicular–arbuscular mycorrhiza formation on Ri T-DNA transformed roots. *New Phytologist*, *108*(2), 211-218.

(5) Montero, H., Choi, J., & Paszkowski, U. (2019). Arbuscular mycorrhizal phenotyping: the dos and don'ts. The New Phytologist, 221(3), 1182.

(6) Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., ... & Tinevez, J. Y. (2012). Fiji: an open-source platform for biological-image analysis. Nature methods, 9(7), 676-682.

(7) Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., ... & Childs, K. L. (2013). Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice*, *6*(1), 4.

(8) Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21.

(9) Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923-930.

(10) Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140.

(11) Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 550.

(12) Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... & Harris, M. A. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, *25*(1), 25-29.

(13) Choi, J., Lee, T., Cho, J., Servante, E. K., Pucker, B., Summers, W., ... & Paszkowski, U. (2020). The negative regulator SMAX1 controls mycorrhizal symbiosis and strigolactone biosynthesis in rice. *Nature Communications*, *11*(1), 1-13.

(14) Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic acids research*, *47*(D1), D419-D426.

(15) Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30.

(16) Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., ... & Zhao, P. X. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Molecular plant*, *9*(12), 1667-1670.

(17) Radhakrishnan, G. V., Keller, J., Rich, M. K., Vernié, T., Mbadinga, D. L. M., Vigneron, N., ... & Delaux, P. M. (2020). An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. *Nature Plants*, 1-10.

(18) Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., … & Wong, G. K. S. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, 574(7780), 679.

(19) Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772-780.

(20) Brown, J. W., Walker, J. F., & Smith, S. A. (2017). Phyx: phylogenetic tools for unix. *Bioinformatics*, *33*(12), 1886-1888.

(21) Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453-4455.

(22) Soudzilovskaia, N. A., Vaessen, S., Barcelo, M., He, J., Rahimlou, S., Abarenkov, K., ... & Tedersoo, L. (2020). FungalRoot: global online database of plant mycorrhizal associations. *New Phytologist*, *227*(3), 955-966.

(23) El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., ... & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic acids research*, *47*(D1), D427-D432.

(24) Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189-1191.

(25) Käll, L., Krogh, A., & Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology*, *338*(5), 1027-1036.

(26) Pierleoni, A., Martelli, P. L., & Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC bioinformatics*, *9*(1), 392.

(27) De Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., ... & Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research*, *34*(suppl_2), W362-W365.