

POGED: A Plant Orthologous Group Evolution Database

Yuanyuan Wang

Adelaide University

Guang Chen

Guangdong Academy of Sciences

Shouli Feng

Xianghu Laboratory

Cong Tan

BGI Research

Lulu Wang

Adelaide University

Hangjin Jiang

Zhejiang University

Shengchun Xu

Xianghu Laboratory

Shengguan Cai

Zhejiang University

Feibo Wu

Zhejiang University

Guoping Zhang

Zhejiang University

Zhong-Hua Chen

`zhong-hua.chen@adelaide.edu.au`

Adelaide University

Research Article

Keywords: comparative genomics, gene orthology, green plants, molecular evolution, phylogenetic analysis

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10838442/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Comparative analyses of orthologous gene families are central to understanding the evolution of gene function and regulatory pathways across species. However, accurate and efficient orthology assignment across diverse plant and algal genomes remains challenging for many researchers with limited bioinformatics background. We designed a Plant Orthologous Group Evolution Database (POGED, <http://poged.xhlab.ac.cn:30001/#/>), which integrates orthology inference and domain annotation across 57 species. POGED includes the major evolutionary trajectory of land plants and algae, providing a comprehensive framework for robust evolutionary analyses. Specifically, POGED provides (i) an OrthoFinder-based Search module for orthogroup queries by protein identifiers with downloadable tables and phylogenetic trees; (ii) a PfamScan-based Search module for identifying proteins that share identical Pfam domain architectures and for querying by both protein and Pfam identifiers; (iii) a BLAST module for similarity searches against all integrated species; and (iv) a Submit module that enables users to upload custom protein datasets for orthogroup or domain architecture analyses. In addition, analysis of MAPK orthogroups and the protein kinase domain PF00069 (Pkinase) demonstrated that POGED is ideal for orthogroup-based phylogenetic analyses and cross-lineage comparisons of domain architectures. We also highlighted significant applications of POGED in gene family evolution analysis and gene discovery for function verification in plants.

Highlight

By integrating precomputed orthogroups with protein domain architecture searches across a phylogenetically comprehensive set of 57 plant and algal species, POGED allows researchers to conduct comparative genomics across green plants by visualizing and exporting for downstream analyses, enabling intuitive exploration of how gene families have evolved across different lineages.

Introduction

Accurate assignment of genes to orthologous groups is a prerequisite for genome-scale phylogenetics and comparative genomic analyses (Dewar et al., 2025). However, orthology inference is often complicated by the need to distinguish orthologs (genes that diverged through speciation from a single gene in the last common ancestor) from paralogs (genes that diverged through duplication) (Sonnhammer & Koonin, 2002). Growing interest in gene orthology has driven the development of a wide range of computational approaches, including phylogeny-based methods (Goodman et al., 1979; Hernández-Plaza et al., 2026), pairwise similarity-based methods (Barton, 1990), and profile-based methods (Henikoff, 1996; Milner et al., 2019; Taylor, 1986).

Phylogeny-based methods can be highly accurate, but they are computationally demanding when applied to large gene sets across many species. Excellent databases such as OrthologID and eggNOG (Chiu et al., 2006; Hernández-Plaza et al., 2026) and PhylomeDB (Huerta-Cepas et al., 2014) rely on curated phylogenetic pipelines. In contrast, pairwise similarity-based approaches were developed to reduce

computational cost. Basic Local Alignment Search Tool (BLAST) remains one of the most widely used tools for DNA and protein sequence similarity searches (Altschul et al., 1997; Johnson et al., 2008; Priyam et al., 2019). Building on sequence similarity graphs, OrthoFinder is a fast and scalable orthology inference method that constructs orthogroups from BLAST-derived similarity scores (Emms & Kelly, 2019). By leveraging these scores, OrthoFinder reduces biases associated with gene length variation and phylogenetic distance, and benchmarking analyses have shown that it performs competitively compared with widely used approaches such as TreeFam, OrthoMCL and OMA (Altenhoff et al., 2011; Altenhoff et al., 2021; Emms & Kelly, 2019; Li et al., 2006; Roth et al., 2008).

Profile-based approaches provide an additional layer of sensitivity for detecting remote homology. A position specific scoring matrix (PSSM; also called a profile) represents position-wise residue probabilities derived from a multiple sequence alignment and allows conserved and variable positions to be weighted differently (Baldi et al., 1994; Mohammadi et al., 2022). Profile hidden Markov models (profile HMMs) further improve on traditional profiles, which are widely used for detecting distant homologs (Eddy, 1998; Karplus & Hu, 2001). Large, curated libraries of profile HMMs are available, including HMMER, Pfam, PfamScan, KofamKOALA and SMART (Letunic et al., 2021; Schultz et al., 1998). PfamScan is commonly used to annotate protein sequences by searching them against Pfam HMM libraries.

Many plant proteins contain multiple functional domains, which is challenging to assess whether candidate homologs share comparable Pfam domain compositions (domain architectures), especially in large scale multi-species datasets. Domain architecture is a key determinant of protein function, and systematic comparisons of domain architectures can therefore provide insight into functional diversification and evolutionary innovation (Liang et al., 2025; Yu et al., 2019). Multidomain proteins are pervasive across the tree of life. Both the number of unique domains and the fraction of multidomain proteins tend to increase with organismal complexity (Scaiewicz & Levitt, 2018). This long-term trend of domain accretion represents an important evolutionary strategy to enable new combinations of modular domains and the reuse of common catalytic cores in different functional contexts (Sonnhammer & Koonin, 2002). Despite the widespread adoption of Pfam-based annotations, dedicated plant resources that systematically aggregate, classify and enable comparative analysis of Pfam domain architectures across large protein datasets remain limited.

To fill this gap, Plant Orthologous Group Evolution Database (POGED) supports Pfam identifier-based search, allowing users to retrieve all proteins and domain architecture combinations that contain a queried Pfam domain across species. The rapid accumulation of genome assemblies and the growing number of users without extensive bioinformatics training have created an increasing demand for user-friendly tools that enable reliable homolog identification and evolutionary analyses in green plants. In addition, researchers without a strong background in plant systematics often encounter difficulties selecting representative species that capture key evolutionary lineages with sufficiently high-quality genome assemblies. Here, we present POGED, a genome-wide ortholog identification database for green

plants, built on precomputed OrthoFinder orthology inferences, PfamScan domain annotations and user defined BLAST searches across 57 representative plant and algal species.

Materials and Methods

Species selection and phylogenetic framework

With more than 1,800 plant genomes now sequenced (Schwacke et al., 2025), it remains challenging for researchers to select an appropriate set of species for phylogenetic analyses, particularly when aiming to include well studied taxa with high quality genome assemblies. To provide a broad evolutionary framework, POGED incorporates 57 publicly available plant genomes integrate a total of 2,008,443 protein sequences spanning major lineages of green plants and model species (Table 1). Species were selected to represent key evolutionary branches while avoiding unnecessary redundancy. The dataset includes representative chlorophyte algae (e.g. *Chlamydomonas reinhardtii* (Merchant et al., 2007), *Volvox carteri* (Prochnik et al., 2010)), early-diverging streptophyte algae (*Klebsormidium nitens* (Hori et al., 2014), *Chara braunii* (Nishiyama et al., 2018)), bryophytes such as liverworts (*Marchantia polymorpha* (Bowman et al., 2017)), mosses (*Physcomitrella patens* (Lang et al., 2018), *Sphagnum fallax* (Healey et al., 2023)), and hornworts (*Anthoceros* spp. (Li et al., 2020)), as well as vascular plants including lycophytes (*Selaginella moellendorffii* (Banks et al., 2011)), ferns (*Ceratopteris richardii* (Marchant et al., 2022), *Adiantum capillus-veneris* (Fang et al., 2022), *Azolla filiculoides* (Li et al., 2018), *Salvinia cucullata* (Li et al., 2018)) and angiosperms.

Table 1

Phylogenetic coverage of species included in Plant Orthologous Group Evolution Database (POGED).

Clade	Species	Reference
Core eudicots; Rosids	<i>Arabidopsis thaliana</i>	(de, 2000)
Core eudicots; Rosids	<i>Brassica napus</i>	(Chalhoub et al., 2014)
Core eudicots; Rosids	<i>Gossypium raimondii</i>	(Paterson et al., 2012)
Core eudicots; Rosids	<i>Citrus clementina</i>	(Wu et al., 2014)
Core eudicots; Rosids	<i>Populus trichocarpa</i>	(Tuskan et al., 2006)
Core eudicots; Rosids	<i>Manihot esculenta</i>	(Prochnik et al., 2012)
Core eudicots; Rosids	<i>Glycine max</i>	(Schmutz et al., 2010)
Core eudicots; Rosids	<i>Medicago truncatula</i>	(Young et al., 2011)
Core eudicots; Rosids	<i>Citrullus lanatus</i>	(Guo et al., 2013)
Core eudicots; Rosids	<i>Cucumis sativus</i>	(Huang et al., 2009)
Core eudicots; Rosids	<i>Eucalyptus grandis</i>	(Myburg et al., 2014)
Core eudicots; Rosids	<i>Vitis vinifera</i>	(French, 2007)
Core eudicots; Asterids	<i>Helianthus annuus</i>	(Badouin et al., 2017)
Core eudicots; Asterids	<i>Nicotiana attenuata</i>	(Xu et al., 2017)
Core eudicots; Asterids	<i>Solanum tuberosum</i>	(Guzman, 2011)
Core eudicots; Asterids	<i>Solanum lycopersicum</i>	(Tomato Genome Consortium, 2012)
Core eudicots; Asterids	<i>Capsicum annuum</i>	(Qin et al., 2014)
Core eudicots; Asterids	<i>Daucus carota</i>	(Iorizzo et al., 2016)
Core eudicots; Caryophyllales	<i>Beta vulgaris</i>	(Dohm et al., 2014)
Basal eudicots	<i>Nelumbo nucifera</i>	(Ming et al., 2013)
Basal eudicots	<i>Aquilegia coerulea</i>	(Filiault et al., 2018)
Monocots; Poaceae	<i>Hordeum vulgare</i>	(Mascher et al., 2021; Mayer et al., 2012)
Monocots; Poaceae	<i>Triticum aestivum</i>	(Consortium et al., 2018)
Monocots; Poaceae	<i>Brachypodium distachyon</i>	(Vogel et al., 2010)
Monocots; Poaceae	<i>Oryza sativa</i>	(Yu et al., 2002)
Monocots; Poaceae	<i>Sorghum bicolor</i>	(Paterson et al., 2009)

Clade	Species	Reference
Monocots; Poaceae	<i>Zea mays</i>	(Gore et al., 2009)
Monocots; Commelinids	<i>Ananas comosus</i>	(Chen et al., 2019; Ming et al., 2015)
Monocots; Commelinids	<i>Musa acuminata</i>	(D'hont et al., 2012)
Monocots; Asparagales	<i>Asparagus officinalis</i>	(Harkess et al., 2017)
Monocots; Asparagales	<i>Phalaenopsis equestris</i>	(Cai et al., 2015)
Basal angiosperms	<i>Nymphaea colorata</i>	(Zhang et al., 2020)
Basal angiosperms	<i>Amborella trichopoda</i>	(Project et al., 2013)
Gymnosperms	<i>Ginkgo biloba</i>	(Guan et al., 2016)
Gymnosperms	<i>Pinus lambertiana</i>	(Stevens et al., 2016)
Gymnosperms	<i>Gnetum montanum</i>	(Wan et al., 2018)
Ferns (Polypodiopsida)	<i>Ceratopteris richardii</i>	(Marchant et al., 2022)
Ferns (Polypodiopsida)	<i>Adiantum capillus-veneris</i>	(Fang et al., 2022)
Ferns (Polypodiopsida)	<i>Azolla filiculoides</i>	(Li et al., 2018)
Ferns (Polypodiopsida)	<i>Salvinia cucullata</i>	(Li et al., 2018)
Lycophytes	<i>Selaginella moellendorffii</i>	(Banks et al., 2011)
Bryophytes; Hornworts	<i>Anthoceros agrestis</i>	(Li et al., 2020)
Bryophytes; Hornworts	<i>Anthoceros punctatus</i>	(Li et al., 2020)
Bryophytes; Hornworts	<i>Anthoceros angustus</i>	(Zhang et al., 2020)
Bryophytes; Mosses	<i>Sphagnum fallax</i>	(Healey et al., 2023)
Bryophytes; Mosses	<i>Physcomitrium patens</i>	(Lang et al., 2018)
Bryophytes; Liverworts	<i>Marchantia polymorpha</i>	(Bowman et al., 2017)
Streptophyte algae; Zygnematophyceae	<i>Spirogloea muscicola</i>	(Cheng et al., 2019)
Streptophyte algae; Zygnematophyceae	<i>Mesotaenium endlicherianum</i>	(Cheng et al., 2019)
Streptophyte algae; Charophyceae	<i>Chara braunii</i>	(Nishiyama et al., 2018)
Streptophyte algae; Klebsormidiophyceae	<i>Klebsormidium nitens</i>	(Hori et al., 2014)

Clade	Species	Reference
Chlorophytes	<i>Chlamydomonas reinhardtii</i>	(Merchant et al., 2007)
Chlorophytes	<i>Volvox carteri</i>	(Prochnik et al., 2010)
Prasinodermophyta	<i>Prasinoderma coloniale</i>	(Li et al., 2020)
Rhodophytes	<i>Porphyra umbilicalis</i>	(Brawley et al., 2017)
Rhodophytes	<i>Chondrus crispus</i>	(Janouskovec et al., 2013)

Within angiosperms, we selected phylogenetically reference species spanning major early and derived lineages. These include the basal angiosperms *Amborella trichopoda* (Project et al., 2013) and *Nymphaea colorata* (water lily) (Zhang et al., 2020); the basal eudicot *Aquilegia coerulea* (columbine) (Filiault et al., 2018) and *Nelumbo nucifera* (lotus) (Ming et al., 2013); the monocot *Phalaenopsis equestris* (orchid) (Cai et al., 2015); together with the model core eudicot *Arabidopsis thaliana* (de, 2000), these species provide broad phylogenetic representation across major angiosperm clades. We also included major cereal crops (*Triticum aestivum* (wheat) (Consortium et al., 2018), *Oryza sativa* (rice) (Yu et al., 2002), *Hordeum vulgare* (barley) (Mascher et al., 2021; Mayer et al., 2012), *Zea mays* (maize) (Gore et al., 2009), *Sorghum bicolor* (sorghum) (Paterson et al., 2009) and the modern grass *Brachypodium distachyon* (Vogel et al., 2010)), legumes (*Glycine max* (soybean) (Schmutz et al., 2010), *Medicago truncatula* (Young et al., 2011)), solanaceous crops (*Solanum lycopersicum* (tomato) (Tomato Genome Consortium, 2012), *Capsicum annuum* (pepper) (Qin et al., 2014), *Solanum tuberosum* (potato) (Guzman, 2011)), oilseed and industrial crops (*Brassica napus* (rapeseed) (Chalhoub et al., 2014), *Gossypium raimondii* (cotton) (Paterson et al., 2012), *Helianthus annuus* (sunflower) (Badouin et al., 2017), *Manihot esculenta* (cassava) (Prochnik et al., 2012), and *Beta vulgaris* (sugar beet) (Dohm et al., 2014)), fruit and horticultural species (*Citrus clementina* (clementine) (Wu et al., 2014), *Vitis vinifera* (grapevine) (French, 2007), *Musa acuminata* (banana) (D'hont et al., 2012), *Ananas comosus* (pineapple) (Chen et al., 2019; Ming et al., 2015), *Citrullus lanatus* (watermelon) (Guo et al., 2013), *Cucumis sativus* (cucumber) (Huang et al., 2009), *Daucus carota* (carrot) (Iorizzo et al., 2016), and *Asparagus officinalis* (asparagus) (Harkess et al., 2017)), as well as woody perennials (*Populus trichocarpa* (poplar) (Tuskan et al., 2006) and *Eucalyptus grandis* (eucalyptus) (Myburg et al., 2014)). Detailed species information is accessible through scrollable panels on both the home page and the search interface of the database (Fig. 1). Species taxonomy was curated primarily based on NCBI Taxonomy (Schoch et al., 2020), iPlant (<http://www.iplant.cn/>), World Flora Online (<https://wfoplantlist.org/plant-list>), AlgaeBase (<https://www.algaebase.org/>), and the APG IV classification system (The Angiosperm Phylogeny, 2016). This selection captures major transitions in plant evolution, from aquatic algae to land plants, the emergence of vascular tissues and seeds, and the diversification of flowering plants. By integrating broad phylogenetic coverage with key agricultural and horticultural species, the database facilitates evolutionary analyses and gene family investigations in crops.

Precomputed orthology datasets

Orthologous relationships among all protein sequences were inferred using OrthoFinder (Emms & Kelly, 2019). Sequence similarity searches within OrthoFinder were performed using BLAST (Cock et al., 2015), while all other parameters were kept at default settings to ensure reproducibility. Pfam domain annotation was performed using PfamScan (Potter et al., 2018), with default Pfam seed alignment defined E-value thresholds applied for each hidden Markov model (HMM) seed. Following PfamScan annotation, tabular summaries were generated for each species. Within each species, Pfam domains associated with individual protein sequences were first consolidated. Protein sequences from all species were then classified based on their complete Pfam domain combinations (domain architectures), enabling cross-species searches for proteins with shared identical Pfam HMM compositions.

Utility and web interface

POGED comprises four main functional modules: Home, Search, BLAST, and Submit, which together support comprehensive orthology exploration and comparative analyses (Fig. 2). The Home module provides an overview of the database content, supported species, and navigation to all major functions.

Users can select alternative search strategies and species datasets, with detailed species information and corresponding identifiers displayed below the Search interface to guide query selection. Search results are typically returned within seconds and are presented as interactive visualizations alongside downloadable tables (Fig. 3). For OrthoFinder based queries, POGED reports orthogroup membership across species and provides downloadable ortholog lists, FASTA sequences, and precomputed OrthoFinder phylogenetic trees, which can be directly reused for downstream phylogenetic analyses (Fig. 3C). For PfamScan based queries, POGED summarizes proteins sharing identical Pfam domain architectures across species and links each Pfam entry to the Pfam database for functional annotation (Fig. 3D). In addition, searches can be initiated using Pfam identifiers, returning all domain architectures that contain the queried Pfam domain across species (Fig. 3E).

To further facilitate user defined and exploratory analyses, POGED also incorporates BLAST and Submit modules. The BLAST module enables similarity searches against protein datasets from all species integrated in POGED, allowing users to perform BLAST searches with customizable parameters to rapidly identify candidate homologs across lineages integrated. The Submit module extends POGED beyond the pre-integrated species set by allowing users to upload their own protein sequence datasets. To perform customized analyses using the same OrthoFinder and PfamScan workflows, users can upload three to five protein datasets, typically representing different species. This function supports flexible investigation of orthologous and paralogous relationships in user defined datasets.

Results and Discussion

POGED is designed to enable rapid cross-species analyses of orthology and protein domain architectures across the major green plant lineages. Although the online platform is newly released, the underlying comparative workflow has been applied in many previous studies, including orthogroup inference across representative plant lineages coupled with phylogenetic analysis and functional verification for light responses, abiotic stress tolerance, miRNA evolution, stomatal evolution, and protein target prediction (Cai et al., 2021; Cai et al., 2017a; Chen et al., 2017b; Deng et al., 2022; Feng et al., 2020; Li et al., 2025; Singh et al., 2024; Zhao et al., 2019). Analyses of published papers using the core algorithm of POGED demonstrated that the database is a powerful tool in the discovery of origin, expansion, and contraction of gene families and functional characterization of genes in combination with phylogenetic, molecular biological, biochemical, and physiological tools.

We first examined mitogen activated protein kinases (MAPKs) as a case study for orthogroup based queries in POGED. MAPK is a highly conserved family of serine threonine kinases that function as central signalling modules downstream of cell surface receptors and intracellular sensors. MAPK regulate diverse aspects of plants such phytohormones, growth and developmental signals, abiotic and biotic stresses, and stress adaptation (Manna et al., 2023; Meng & Zhang, 2013; Singh et al., 2024). In plants, MAPKs are commonly classified according to their activation loop motifs, TEY (Thr-Glu-Tyr; typically found in MAPK groups A, B, and C) and TDY (Thr-Asp-Tyr; Characteristic of Group D MAPKs), which define two major evolutionary clades with distinct expansion histories and functional specialization (Tena et al., 2001). Consistent with this classification, precomputed OrthoFinder results in POGED assigned MAPK sequences to two major orthogroups. Conserved domain detection (Wang et al., 2023) further indicated that these orthogroups correspond to the STKc_Tey_MAPK and STKc_Tdy_MAPK motif classes, respectively (Fig. 4). This partition strongly aligned with previous phylogenetic analyses, showing that TEY and TDY of MAPKs represent deeply conserved lineages that have undergone lineage specific expansion and functional diversification across land plants (Cristina et al., 2010; Manna et al., 2023).

As a complementary demonstration of domain architecture-based analyses in POGED, we used the protein kinase core domain PF00069 (Pkinase) in MAPKs as an entry point to summarise PF00069 containing domain architectures across species and phylogenetic clades. While the Pkinase domain alone represents the most prevalent architecture across green plants (Fig. 5A), many abundant PF00069 containing architectures incorporate additional extracellular or regulatory domains. The fusion of PF00069 with diverse ligand binding or regulatory domains provides a powerful route to functional evolutionary innovation by expanding signal perception capabilities. PF00069 combined with leucine rich repeat domains (LRRNT and LRR) represents the classical receptor like kinase (RLK) architecture. The mediate perception of diverse extracellular peptide and hormone signals plays central roles in plant development and immunity, gene family evolution and crop domestication (Liu et al., 2017; Morris & Walker, 2003; Yu et al., 2025). PF00069 and LysM domain combinations are also widely observed and are characteristic of kinases involved in microbial recognition, symbiosis and innate immune signaling (Tan et al., 2025; Tanaka et al., 2013). In addition, PF00069 architectures incorporating lectin or malectin like domains such as Wall-Associated Kinases (WAK) domain is associated with carbohydrate and cell wall

related sensing, linking extracellular matrix status to intracellular phosphorylation dependent responses (Herold et al., 2025). Search results retrieved from POGED enable the construction of a PF00069 centred domain architecture network (Fig. 3E), providing an intuitive view of the diversification and evolutionary relationships of PF00069 containing domain combinations across plant lineages (Fig. 5B). This network representation further highlights simple gain and loss relationships among related domain architectures, illustrating how stepwise domain additions around the conserved kinase core have driven functional diversification during plant evolution. Taken together, the MAPK and PF00069 case studies illustrate how POGED enables comparative evolutionary analyses across green plant lineages.

Based on orthologous identification through the POGED, it was demonstrated evolutionarily conserved function of some key gene families in ABA signaling pathway is linked to stomatal closure among many land plant lineages (Cai et al., 2017a). Furthermore, it was found that ABA reception and signalling existed in stomata-bearing terrestrial plants including ferns, which indicated the ABA induced stomatal closure is present in seed plants and in early divergent extant ferns (Cai et al., 2017a). It was hypothesized that the genes involved in stomatal development are likely evolutionarily conserved among the land plants, since many key genes have been conserved in the first stomatal-bearing bryophytes (Cai et al., 2017b). Subsequent studies implemented the evolutionary analysis of relevant gene families among 26 key plant and algae species confirmed the hypothesis (Cai et al., 2017a; Chater et al., 2017). Using the POGED method, we discovered an early evolution of the nucleotide phosphatase (SAL1) and 3'-phosphoadenosine-5'-phosphate (PAP), a key element of chloroplast retrograded signalling, evolved in streptophyte algae, the algal sister group of land plants. It indicated SAL1-PAP retrograde signalling pathway predates the evolutionary of stomata and the evolutionary coordination of the SAL1-PAP signalling is crucial for green plants adaptation to land and abiotic stress tolerance (Calderon & Strand, 2021; Jiang et al., 2024; Zhao et al., 2019).

Furthermore, results derived from the workflow of POGED also provided new insight into complex evolutionary scenarios such as gene family expansion and contraction of green plants (Chen et al., 2024; Singh et al., 2024). For instance, comparative genomics analysis of the blue-light signaling pathway indicated that the multiple duplications of cryptochrome (CRY) may have contributed to the evolution of rapid stomatal opening to blue light in the Polypodiales ferns. The four gene duplication events of CRYs could have benefited the colonization of the low light forest understory habitat and diversification of Polypodiales to become the largest clade of ferns with more than 80% of extant fern species (Cai et al., 2021). In addition to the evolutionary analysis, the workflow of POGED facilitated many explorations or supplements for cloning and verification of genes involved in abiotic stress responses such as drought (Feng et al., 2020), salinity (Kuang et al., 2025; Shen et al., 2020), heat (Chen et al., 2021), and heavy metal stress (Chen et al., 2021; Liu et al., 2021). POGED also is practical for many other applications encompassing design of molecular markers, multi-omics analysis, and literature review for future research projects (Deng et al., 2021; Li et al., 2025; Yang et al., 2025). For example, the coevolution of miRNAs and their target genes demonstrated the key roles of miRNAs in the diversification, adaptation, and evolution of green plants and the potential for gene editing of miRNAs in plants (Deng et al., 2022), which was recently validated for improving salt tolerance in barley (Kuang et al., 2025).

Conclusion

Integrative comparative genomics across diverse species creates strong opportunities for interdisciplinary collaboration and new discoveries in life sciences. We propose that POGED will provide plant, agricultural, horticultural and environmental scientists a convenient entrance into evolutionary analysis of genes and gene families of their interests. The consequent results facilitate many preliminary explorations for cloning and functional validation of target genes and cross-species, multi-omics studies. Therefore, the release of the POGED tool is timely and important for research endeavours in plant biology, evolutionary biology, molecular biology and biochemistry, agriculture, and environmental science.

Declarations

Acknowledgements

We thank Gang Xu and Jianbo Li at BGI-Shenzhen who provided technical supports to this project. We thank Dr. Yuanqing Zheng and Dr. Fei Xu for their support and guidance in database design.

Author contribution

Z-HC, YW, GC, SX, YZ and HJ conceived and designed the database. YW and GC conducted data analyses and integration. YW and Z-HC wrote the manuscript with contribution from all authors.

Funding

This work is funded by the Australian Research Council (FT210100366), the National Natural Science Foundation of China (31620103912 and 31571578), China Agriculture Research System (CARS-05) and Adelaide University Top Talent Program.

Data Availability Statement

All protein sequence datasets used in POGED were downloaded from publicly available databases (<https://www.ncbi.nlm.nih.gov>; <https://phytozome-next.jgi.doe.gov>; <https://plants.ensembl.org/index.html>; <https://gigadb.org>; <https://www.fernbase.org>; <https://www.hornworts.uzh.ch/en.html>). For transparency and traceability, the original sources and associated publications are cited and linked for each species within the database.

Conflict of interest

None declared

References

1. Altenhoff, A. M., Schneider, A., Gonnet, G. H., & Dessimoz, C. (2011). OMA 2011: orthology inference among 1000 complete genomes. *Nucleic Acids Res.*, 39, D289-D294.
<https://doi.org/10.1093/nar/gkq1238>
2. Altenhoff, A. M., Train, C. M., Gilbert, K. J., Mediratta, I., de Farias, T. M., Moi, D., Nevers, Y., Radoykova, H. S., Rossier, V., Vesztochy, A. W., Glover, N. M., & Dessimoz, C. (2021). OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res.*, 49(D1), D373-D379. <https://doi.org/10.1093/nar/gkaa1007>
3. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25(17), 3389–3402. <https://doi.org/DOI 10.1093/nar/25.17.3389>
4. Badouin, H., Gouzy, J., Grassa, C.J., Murat, F., Staton, S.E., Cottret, L., Lelandais-Brière, C., Owens, G.L., Carrère, S., Mayjonade, B. and Legrand, L., (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, 546(7656), 148–152.
5. Baldi, P., Chauvin, Y., Hunkapiller, T., & McClure, M. A. (1994). Hidden Markov-Models of Biological Primary Sequence Information. *Proc. Natl. Acad. Sci. U.S.A.* 91(3), 1059–1063. <https://doi.org/DOI 10.1073/pnas.91.3.1059>
6. Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., DePamphilis, C., Albert, V. A., Aono, N., Aoyama, T., & Ambrose, B. A. (2011). The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science*, 332(6032), 960–963.
7. Barton, G. J. (1990). Protein Multiple Sequence Alignment and Flexible Pattern-Matching. *Methods in Enzymology*, 183, 403–428.
8. Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., & Berger, F. d. r. (2017). Insights into land plant evolution garnered from the Marchantia polymorpha genome. *Cell*, 171(2), 287–304. e215.
9. Brawley, S. H., Blouin, N. A., Ficko-Blean, E., Wheeler, G. L., Lohr, M., Goodson, H. V., Jenkins, J. W., Blaby-Haas, C. E., Helliwell, K. E., Chan, C. X., Marriage, T. N., Bhattacharya, D., Klein, A. S., Badis, Y., Brodie, J., Cao, Y., Collā©n, J., Dittami, S. M., Gachon, C. M. M.,...Prochnik, S. E. (2017). Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). *Proc. Natl. Acad. Sci. U.S.A.*, 114(31), E6361-E6370.
<https://doi.org/10.1073/pnas.1703088114>
10. Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.-C., Liu, K.-W., Chen, L.-J., He, Y., Xu, Q., & Bian, C. (2015). The genome sequence of the orchid Phalaenopsis equestris. *Nat. Genet.*, 47(1), 65–72.
11. Cai, S., Chen, G., Wang, Y., Huang, Y., Marchant, D. B., Wang, Y., Yang, Q., Dai, F., Hills, A., Franks, P. J., Nevo, E., Soltis, D. E., Soltis, P. S., Sessa, E., Wolf, P. G., Xue, D., Zhang, G., Pogson, B. J., Blatt, M. R., & Chen, Z. H. (2017a). Evolutionary Conservation of ABA Signaling for Stomatal Closure. *Plant Physiol*, 174(2), 732–747. <https://doi.org/10.1104/pp.16.01848>
12. Cai, S., Huang, Y., Chen, F., Zhang, X., Sessa, E., Zhao, C., Marchant, D. B., Xue, D., Chen, G., Dai, F., Leebens-Mack, J. H., Zhang, G., Shabala, S., Christie, J. M., Blatt, M. R., Nevo, E., Soltis, P. S., Soltis, D.

- E., Franks, P. J.,...Chen, Z. H. (2021). Evolution of rapid blue-light response linked to explosive diversification of ferns in angiosperm forests. *New Phytol*, 230(3), 1201–1213.
<https://doi.org/10.1111/nph.17135>
13. Cai, S. G., Papanatsiou, M., Blatt, M. R., & Chen, Z. H. (2017b). Speedy Grass Stomata: Emerging Molecular and Evolutionary Features. *Molecular Plant*, 10(7), 912–914.
<https://doi.org/10.1016/j.molp.2017.06.002>
 14. Calderon, R. H., & Strand, Å. s. (2021). How retrograde signaling is intertwined with the evolution of photosynthetic eukaryotes. *Curr. Opin. Plant Biol.*, 63, 102093.
 15. Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., & Samans, B. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, 345(6199), 950–953.
 16. Chater, C. C. C., Caine, R. S., Fleming, A. J., & Gray, J. E. (2017). Origins and evolution of stomatal development [Article]. *Plant Physiol.*, 174(2), 624–638. <https://doi.org/10.1104/pp.17.00183>
 17. Chen, F., Dong, G., Wang, F., Shi, Y., Zhu, J., Zhang, Y., Ruan, B., Wu, Y., Feng, X., & Zhao, C. (2021). A β -ketoacyl carrier protein reductase confers heat tolerance via the regulation of fatty acid biosynthesis and stress signaling in rice. *New Phytol.*, 232(2), 655–672.
 18. Chen, G., Qin, Y., Wang, J., Li, S., Zeng, F., Deng, F., Chater, C., Xu, S., & Chen, Z. H. (2024). Stomatal evolution and plant adaptation to future climate. *Plant, Cell & Environ.*, 47(9), 3299–3315.
 19. Chen, L.-Y., VanBuren, R., Paris, M., Zhou, H., Zhang, X., Wai, C. M., Yan, H., Chen, S., Alonge, M., & Ramakrishnan, S. (2019). The bracteatus pineapple genome and domestication of clonally propagated crops. *Nat. Genet.*, 51(10), 1549–1558.
 20. Chen, Z. H., Chen, G., Dai, F., Wang, Y. Z., Hills, A., Ruan, Y. L., Zhang, G. P., Franks, P. J., Nevo, E., & Blatt, M. R. (2017). Molecular Evolution of Grass Stomata. *Trends in Plant Sci.*, 22(2), 124–139.
<https://doi.org/10.1016/j.tplants.2016.09.005>
 21. Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T., Sun, W., Li, X., Xu, Y., & Zhang, Y. (2019). Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell*, 179(5), 1057–1067. e1014.
 22. Chiu, J. C., Lee, E. K., Egan, M. G., Sarkar, I. N., Coruzzi, G. M., & DeSalle, R. (2006). OrthologID: automation of genome-scale ortholog identification within a parsimony framework. *Bioinformatics*, 22(6), 699–707. <https://doi.org/10.1093/bioinformatics/btk040>
 23. Cock, P. J. A., Chilton, J. M., Gruning, B., Johnson, J. E., & Soranzo, N. (2015). NCBI BLAST plus integrated into Galaxy. *GigaScience*, 4. <https://doi.org/ARTN 3910.1186/s13742-015-0080-7>
 24. Consortium, I. W. G. S., Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C. J., Choulet, F., & Distelfeld, A. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403), eaar7191.
 25. Cristina, M. S., Petersen, M., & Mundy, J. (2010). Mitogen-activated protein kinase signaling in plants. *Annu. Rev. plant Biol.*, 61(1), 621–649.

26. D'hont, A. I., Denoeud, F., Aury, J.-M., Baurens, F.-C., Carreel, F. o., Garsmeur, O., Noel, B., Bocs, S. p., Droc, G. t., & Rouard, M. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, 488(7410), 213–217.
27. de, A. G. I. g. t. o. g. g. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), 796–815.
28. Deng, F., Zeng, F., Chen, G., Feng, X., Riaz, A., Wu, X., Gao, W., Wu, F., Holford, P., & Chen, Z. H. (2021). Metalloid hazards: From plant molecular evolution to mitigation strategies. *J Hazard Mater*, 409, 124495. <https://doi.org/10.1016/j.jhazmat.2020.124495>
29. Deng, F., Zeng, F., Shen, Q., Abbas, A., Cheng, J., Jiang, W., Chen, G., Shah, A. N., Holford, P., Tanveer, M., Zhang, D., & Chen, Z. H. (2022). Molecular evolution and functional modification of plant miRNAs with CRISPR. *Trends Plant Sci*, 27(9), 890–907. <https://doi.org/10.1016/j.tplants.2022.01.009>
30. Dewar, A. E., Belcher, L. J., & West, S. A. (2025). A phylogenetic approach to comparative genomics. *Nat. Rev. Genet.*, 26(6), 395–405.
31. Dohm, J.C., Minoche, A.E., Holtgräwe, D., Capella-Gutiérrez, S., Zakrzewski, F., Tafer, H., Rupp, O., Sörensen, T.R., Stracke, R., Reinhardt, R. and Goesmann, A., (2014). The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*, 505(7484), 546–549.
32. Eddy, S. R. (1998). Hidden Markov models and genome sequence analysis. *FASEB Journal*, 12(8), A1327-A1327.
33. Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*, 20(1), 238. <https://doi.org/10.1186/s13059-019-1832-y>
34. Fang, Y., Qin, X., Liao, Q., Du, R., Luo, X., Zhou, Q., Li, Z., Chen, H., Jin, W., & Yuan, Y. (2022). The genome of homosporous maidenhair fern sheds light on the euphyllophyte evolution and defences. *Nature Plants*, 8(9), 1024–1037.
35. Feng, X., Liu, W. X., Qiu, C. W., Zeng, F. R., Wang, Y. Z., Zhang, G. P., Chen, Z. H., & Wu, F. B. (2020). HvAKT2 and HvHAK1 confer drought tolerance in barley through enhanced leaf mesophyll H⁺ homeostasis. *Plant Biotechnol. J.*, 18(8), 1683–1696. <https://doi.org/10.1111/pbi.13332>
36. Filiault, D.L., Ballerini, E.S., Mandakova, T., Aköz, G., Derieg, N.J., Schmutz, J., Jenkins, J., Grimwood, J., Shu, S., Hayes, R.D. and Hellsten, U., (2018). The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *eLife*, 7, e36426.
37. French. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. (2007). *Nature*, 449(7161), 463–467.
38. Goodman, M., Czelusniak, J., Moore, G. W., Romeroherrera, A. E., & Matsuda, G. (1979). Fitting the Gene Lineage into Its Species Lineage, a Parsimony Strategy Illustrated by Cladograms Constructed from Globin Sequences. *Systematic Zoology*, 28(2), 132–163. <https://doi.org/Doi 10.2307/2412519>
39. Gore, M. A., Chia, J.-M., Elshire, R. J., Sun, Q., Ersoz, E. S., Hurwitz, B. L., Peiffer, J. A., McMullen, M. D., Grills, G. S., & Ross-Ibarra, J. (2009). A first-generation haplotype map of maize. *Science*, 326(5956), 1115–1117.

40. Guan, R., Zhao, Y., Zhang, H., Fan, G., Liu, X., Zhou, W., Shi, C., Wang, J., Liu, W., & Liang, X. (2016). Draft genome of the living fossil *Ginkgo biloba*. *GigaScience*, 5(1), s13742-13016-10154-13741.
41. Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., Zheng, Y., Mao, L., Ren, Y., & Wang, Z. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.*, 45(1), 51–58.
42. Guzman, F. (2011). Genome sequence and analysis of the tuber crop potato. *Nature*.
43. Harkess, A., Zhou, J., Xu, C., Bowers, J. E., Van der Hulst, R., Ayyampalayam, S., Mercati, F., Riccardi, P., McKain, M. R., & Kakrana, A. (2017). The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.*, 8(1), 1279.
44. Healey, A. L., Piatkowski, B., Lovell, J. T., Sreedasyam, A., Carey, S. B., Mamidi, S., Shu, S., Plott, C., Jenkins, J., & Lawrence, T. (2023). Newly identified sex chromosomes in the Sphagnum (peat moss) genome alter carbon sequestration and ecosystem dynamics. *Nature Plants*, 9(2), 238–254.
45. Henikoff, S. (1996). Scores for sequence searches and alignments. *Curr. Opin. Struct. Biol.*, 6(3), 353–360. [https://doi.org/Doi 10.1016/S0959-440x\(96\)80055-8](https://doi.org/Doi%2010.1016/S0959-440x(96)80055-8)
46. Hernández-Plaza, A., Deng, Z., Robledo-Yagüe, F., Szklarczyk, D., von Mering, C., Bork, P., & Huerta-Cepas, J. (2026). eggNOG v7: phylogeny-based orthology predictions and functional annotations. *Nucleic Acids Research*, 54(D1), D402-D408.
47. Herold, L., Ordon, J., Hua, C., Kohorn, B. D., NÄ¼rnberger, T., DeFalco, T. A., & Zipfel, C. (2025). Arabidopsis WALL-ASSOCIATED KINASES are not required for oligogalacturonide-induced signaling and immunity. *The Plant Cell*, 37(1), koae317.
48. Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., & Tajima, N. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.*, 5(1), 3978.
49. Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W. J., Wang, X., Xie, B., & Ni, P. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.*, 41(12), 1275–1281.
50. Huerta-Cepas, J., Capella-Gutierrez, S., Pryszcz, L. P., Marcet-Houben, M., & Gabaldon, T. (2014). PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Research*, 42(D1), D897-D902. <https://doi.org/10.1093/nar/gkt1177>
51. Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., Bowman, M., Iovene, M., Sanseverino, W., & Cavagnaro, P. (2016). A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nature Genetics*, 48(6), 657–666.
52. Janouskovec, J., Liu, S. L., Martone, P. T., Carré, W., Leblanc, C., Collén, J., & Keeling, P. J. (2013). Evolution of Red Algal Plastid Genomes: Ancient Architectures, Introns, Horizontal Gene Transfer, and Taxonomic Utility of Plastid Markers. *PLoS ONE*, 8(3). [https://doi.org/ARTN e5900110.1371/journal.pone.0059001](https://doi.org/ARTN%20e5900110.1371/journal.pone.0059001)
53. Jiang, W., Deng, F., Babla, M., Chen, C., Yang, D., Tong, T., Qin, Y., Chen, G., Marchant, B., & Soltis, P. (2024). Efficient gene editing of a model fern species through gametophyte-based transformation.

Plant Physiol., 196(4), 2346–2361.

54. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res*, 36(Web Server issue), W5-9. <https://doi.org/10.1093/nar/gkn201>
55. Karplus, K., & Hu, B. R. (2001). Evaluation of protein multiple alignments by SAM-T99 using the BALiBASE multiple alignment test set. *Bioinformatics*, 17(8), 713–720. <https://doi.org/DOI10.1093/bioinformatics/17.8.713>
56. Kuang, L., Zhou, H., Zhang, T., Gao, F., Yan, T., Chen, Z.-H., Shen, Q., Zhang, G., Li, L., & Wu, D. (2025). MicroRNA164d suppresses the HvNAC92-HvHKT1; 5 module to enhance salinity tolerance in barley. *Proc. Natl. Acad. Sci. U.S.A.*, 122(52), e2514555122.
57. Lang, D., Ullrich, K. K., Murat, F., Fuchs, J. r., Jenkins, J., Haas, F. B., Piednoel, M., Gundlach, H., Van Bel, M., & Meyberg, R. (2018). The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. *The Plant Journal*, 93(3), 515–533.
58. Letunic, I., Khedkar, S., & Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.*, 49(D1), D458-D460. <https://doi.org/10.1093/nar/gkaa937>
59. Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., De Vries, J., Delaux, P.-M., Eily, A., Koppers, N., Kuo, L.-Y., & Li, Z. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, 4(7), 460–472.
60. Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., Fernandez-Pozo, N., Barker, M. S., Bennett, T., & Blázquez, M. A. (2020). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. *Nature Plants*, 6(3), 259–272.
61. Li, H., Coghlan, A., Ruan, J., Coin, L. J., Heriche, J. K., Osmotherly, L., Li, R. Q., Liu, T., Zhang, Z., Bolund, L., Wong, G. K. S., Zheng, W. M., Dehal, P., Wang, J., & Durbin, R. (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.*, 34, D572-D580. <https://doi.org/10.1093/nar/gkj118>
62. Li, L., Wang, S., Wang, H., Sahu, S.K., Marin, B., Li, H., Xu, Y., Liang, H., Li, Z., Cheng, S. and Reder, T. (2020). The genome of Prasinoderma coloniale unveils the existence of a third phylum within green plants. *Nature Ecology & Evolution*, 4(9), 1220–1231. <https://doi.org/10.1038/s41559-020-1221-7>
63. Li, T., Li, B., Wang, Y., Xu, J., Li, W., Chen, Z. H., Mou, W., & Xue, D. (2025). WRKY transcription factors in rice: key regulators orchestrating development and stress resilience. *Plant, Cell & Environment*, 48(11), 8388–8406.
64. Liang, Y., Gao, Q., Li, F., Du, Y., Wu, J., Pan, W., Wang, S., Zhang, X., Zhang, M., & Song, X. (2025). The giant genome of lily provides insights into the hybridization of cultivated lilies. *Nature Commun.*, 16(1), 45.
65. Liu, P.-L., Du, L., Huang, Y., Gao, S.-M., & Yu, M. (2017). Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evolutionary Biology*, 17(1), 47.
66. Liu, W., Feng, X., Cao, F., Wu, D., Zhang, G., Vincze, E., Wang, Y., Chen, Z. H., & Wu, F. (2021). An ATP binding cassette transporter HvABCB25 confers aluminum detoxification in wild barley. *J Hazard*

67. Manna, M., Rengasamy, B., & Sinha, A. K. (2023). Revisiting the role of MAPK signalling pathway in plants and its manipulation for crop improvement. *Plant, Cell & Environment*, 46(8), 2277–2295.
68. Marchant, D. B., Chen, G., Cai, S., Chen, F., Schafran, P., Jenkins, J., Shu, S., Plott, C., Webber, J., & Lovell, J. T. (2022). Dynamic genome evolution in a model fern. *Nature Plants*, 8(9), 1038–1051.
69. Mascher, M., Wicker, T., Jenkins, J., Plott, C., Lux, T., Koh, C. S., Ens, J., Gundlach, H., Boston, L. B., & Tulpová, Z. (2021). Long-read sequence assembly: a technical evaluation in barley. *The Plant Cell*, 33(6), 1888–1906.
70. Mayer, K., Waugh, R., Langridge, P., Close, T., Wise, R., Graner, A., Matsumoto, T., Sato, K., Schulman, A., & Muehlbauer, G. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, 491, 711–716.
71. Meng, X., & Zhang, S. (2013). MAPK cascades in plant disease resistance signaling. *Annu. Rev. phytopathol*, 51(1), 245–266.
72. Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., & Maréchal-Drouard, L. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science*, 318(5848), 245–250.
73. Milner, S. G., Jost, M., Taketa, S., Mazon, E. R., Himmelbach, A., Oppermann, M., Weise, S., Knupffer, H., Basterrechea, M., Konig, P., Schuler, D., Sharma, R., Pasam, R. K., Rutten, T., Guo, G. G., Xu, D. D., Zhang, J., Herren, G., Muller, T.,...Stein, N. (2019). Genebank genomics highlights the diversity of a global barley collection. *Nature Genet.*, 51(2), 319–+. <https://doi.org/10.1038/s41588-018-0266-x>
74. Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L.-T., Zhang, Q., Kim, M.-J., Schatz, M. C., & Campbell, M. (2013). Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biology*, 14(5), R41.
75. Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., Lyons, E., Wang, M.-L., Chen, J., & Biggers, E. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.*, 47(12), 1435–1442.
76. Mohammadi, A., Zahiri, J., Mohammadi, S., Khodarahmi, M., & Arab, S. S. (2022). PSSMCOOL: a comprehensive R package for generating evolutionary-based descriptors of protein sequences from PSSM profiles. *Biol Methods Protoc*, 7(1), bpac008. <https://doi.org/10.1093/biomethods/bpac008>
77. Morris, E. R., & Walker, J. C. (2003). Receptor-like protein kinases: the keys to response. *Current Opinion in Plant Biology*, 6(4), 339–342.
78. Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., & Bauer, D. (2014). The genome of *Eucalyptus grandis*. *Nature*, 510(7505), 356–362.
79. Nishiyama, T., Sakayama, H., De Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K. K., Haas, F. B., Vanderstraeten, L., Becker, D., & Lang, D. (2018). The Chara genome: secondary complexity and implications for plant terrestrialization. *Cell*, 174(2), 448–464. e424.

80. Paterson, A. H., Bowers, J. E., Bruggmann, R. m., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., & Poliakov, A. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature*, 457(7229), 551–556.
81. Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K. C., Shu, S., & Udall, J. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature*, 492(7429), 423–427.
82. Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., & Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res*, 46(W1), W200–W204. <https://doi.org/10.1093/nar/gky448>
83. Priyam, A., Woodcroft, B. J., Rai, V., Moghul, I., Munagala, A., Ter, F., Chowdhary, H., Pieniak, I., Maynard, L. J., Gibbins, M. A., Moon, H., Davis-Richardson, A., Uludag, M., Watson-Haigh, N. S., Challis, R., Nakamura, H., Favreau, E., Gomez, E. A., Pluskal, T.,...Wurm, Y. (2019). Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. *Mol Biol Evol*, 36(12), 2922–2924. <https://doi.org/10.1093/molbev/msz185>
84. Prochnik, S., Marri, P. R., Desany, B., Rabinowicz, P. D., Kodira, C., Mohiuddin, M., Rodriguez, F., Fauquet, C., Tohme, J., & Harkins, T. (2012). The cassava genome: current progress, future directions. *Tropical Plant Biology*, 5(1), 88–94.
85. Prochnik, S. E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L. K., Hellsten, U., Chapman, J., Simakov, O., Rensing, S. A., Terry, A., Pangilinan, J., Kapitonov, V., Jurka, J., Salamov, A.,...Rokhsar, D. S. (2010). Genomic Analysis of Organismal Complexity in the Multicellular Green Alga. *Science*, 329(5988), 223–226. <https://doi.org/10.1126/science.1188800>
86. Project, A. G., Albert, V. A., Barbazuk, W. B., dePamphilis, C. W., Der, J. P., Leebens-Mack, J., Ma, H., Palmer, J. D., Rounsley, S., & Sankoff, D. (2013). The Amborella genome and the evolution of flowering plants. *Science*, 342(6165), 1241089.
87. Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., & Luo, Y. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. *Proc. Natl. Acad. Sci. U.S.A.*, 111(14), 5135–5140.
88. Roth, A. C., Gonnet, G. H., & Dessimoz, C. (2008). Algorithm of OMA for large-scale orthology inference. *BMC Bioinformatics*, 9, 518. <https://doi.org/10.1186/1471-2105-9-518>
89. Scaiewicz, A., & Levitt, M. (2018). Unique function words characterize genomic proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 115(26), 6703–6708.
90. Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., & Cheng, J. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, 463(7278), 178–183.
91. Schoch, C. L., Ciufo, S., Domrachev, M., Hottton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., & Robbertse, B. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database*, 2020, baaa062.

92. Schultz, J., Milpetz, F., Bork, P., & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A*, 95(11), 5857–5864. <https://doi.org/10.1073/pnas.95.11.5857>
93. Schwacke, R., Bolger, M. E., & Usadel, B. r. (2025). PubPlant-a continuously updated online resource for sequenced and published plant genomes. *Frontiers in Plant Science*, 16, 1603547.
94. Shen, Q. F., Fu, L. B., Su, T. T., Ye, L. Z., Huang, L., Kuang, L. H., Wu, L. Y., Wu, D. Z., Chen, Z. H., & Zhang, G. P. (2020). Calmodulin HvCaM1 Negatively Regulates Salt Tolerance via Modulation of HvHKT1s and HvCAMTA4. *Plant Physiology*, 183(4), 1650–1662. <https://doi.org/10.1104/pp.20.00196>
95. Singh, V. P., Jaiswal, S., Wang, Y., Feng, S., Tripathi, D. K., Singh, S., Gupta, R., Xue, D., Xu, S., & Chen, Z.-H. (2024). Evolution of reactive oxygen species cellular targets for plant development. *Trends in Plant Science*, 29(8), 865–877. <https://doi.org/10.1016/j.tplants.2024.03.005>
96. Sonnhammer, E. L., & Koonin, E. V. (2002). Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet*, 18(12), 619–620. [https://doi.org/10.1016/s0168-9525\(02\)02793-2](https://doi.org/10.1016/s0168-9525(02)02793-2)
97. Stevens, K. A., Wegrzyn, J. L., Zimin, A., Puiu, D., Crepeau, M., Cardeno, C., Paul, R., Gonzalez-Ibeas, D., Koriabine, M., & Holtz-Morris, A. E. (2016). Sequence of the sugar pine megagenome. *Genetics*, 204(4), 1613–1626.
98. Tan, X., Wang, D., Zhang, X., Zheng, S., Jia, X., Liu, H., Liu, Z., Yang, H., Dai, H., & Chen, X. (2025). A pair of LysM receptors mediates symbiosis and immunity discrimination in *Marchantia*. *Cell*, 188(5), 1330–1348. e1327.
99. Tanaka, K., Nguyen, C. T., Liang, Y., Cao, Y., & Stacey, G. (2013). Role of LysM receptors in chitin-triggered plant innate immunity. *Plant signaling & behavior*, 8(1), e22598.
100. Taylor, W. R. (1986). Identification of protein sequence homology by consensus template alignment. *J Mol Biol*, 188(2), 233–258. [https://doi.org/10.1016/0022-2836\(86\)90308-6](https://doi.org/10.1016/0022-2836(86)90308-6)
101. Tena, G., Asai, T., Chiu, W.-L., & Sheen, J. (2001). Plant mitogen-activated protein kinase signaling cascades. *Curr. Opin. Plant Biol.*, 4(5), 392–400.
102. The Angiosperm Phylogeny, G. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181(1), 1–20. <https://doi.org/https://doi.org/10.1111/boj.12385>
103. Tomato Genome Consortium, x. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485(7400), 635.
104. Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J. r., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., & Salamov, A. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313(5793), 1596–1604.
105. Vogel, J. P., Garvin, D. F., Mockler, T. C., Schmutz, J., Rokhsar, D., Bevan, M. W., Barry, K., Lucas, S., Harmon-Smith, M., Lail, K., Tice, H., Schmutz, J., Grimwood, J., McKenzie, N., Bevan, M. W., Huo, N., Gu, Y. Q., Lazo, G. R., Anderson, O. D.,...gene family, a. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, 463(7282), 763–768.

106. Wan, T., Liu, Z.-M., Li, L.-F., Leitch, A. R., Leitch, I. J., Lohaus, R., Liu, Z.-J., Xin, H.-P., Gong, Y.-B., & Liu, Y. (2018). A genome for gnetophytes and early evolution of seed plants. *Nature Plants*, 4(2), 82–89.
107. Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., & Yamashita, R. A. (2023). The conserved domain database in 2023. *Nucleic acids Res.*, 51(D1), D384-D388.
108. Wu, G. A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., Perrier, X., Ruiz, M., Scalabrin, S., & Terol, J. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology*, 32(7), 656–662.
109. Xu, S., Brockmöller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., Zhou, W., Kreitzer, C., Stanke, M., & Tang, H. (2017). Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.*, 114(23), 6133–6138.
110. Yang, D., Chen, G., Jiang, W., Marchant, D. B., Cai, S., Zhu, X., Soltis, P. S., Soltis, D. E., & Chen, Z. H. (2025). Harnessing fern stress adaptations: From evolution and ecophysiology to molecular biology. *The Plant Journal*, 124(3), e70572.
111. Young, N.D., Debelle, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H. and Van de Peer, Y., (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature*, 480(7378), 520–524.
112. Yu, J., Chen, Q., Yuan, L., Feng, S., Huang, M., Zheng, P., Chen, G., Tao, X., Edwards, D., & Chen, Z. H. (2025). Phylogenomic and super-pangenome analyses unveil the genetic landscape of tomato evolution and domestication. *Plant Biotechnology Journal*, 23(9), 3783–3797.
113. Yu, J., Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., & Zhang, X. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science*, 296(5565), 79–92.
114. Yu, L., Tanwar, D. K., Penha, E. D. S., Wolf, Y. I., Koonin, E. V., & Basu, M. K. (2019). Grammar of protein domain architectures. *Proc. Natl. Acad. Sci. U.S.A.*, 116(9), 3636–3645.
115. Zhang, J., Fu, X.-X., Li, R.-Q., Zhao, X., Liu, Y., Li, M.-H., Zwaenepoel, A., Ma, H., Goffinet, B., & Guan, Y.-L. (2020). The hornwort genome and early land plant evolution. *Nature Plants*, 6(2), 107–118.
116. Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho, S. Y., & Liu, X. (2020). The water lily genome and the early evolution of flowering plants. *Nature*, 577(7788), 79–84.
117. Zhao, C., Wang, Y., Chan, K. X., Marchant, D. B., Franks, P. J., Randall, D., Tee, E. E., Chen, G., Ramesh, S., Phua, S. Y., Zhang, B., Hills, A., Dai, F., Xue, D., Gilliam, M., Tyerman, S., Nevo, E., Wu, F., Zhang, G.,...Chen, Z. H. (2019). Evolution of chloroplast retrograde signaling facilitates green plant adaptation to land. *Proc. Natl. Acad. Sci. U.S.A.*, 116(11), 5015–5020.

Figures

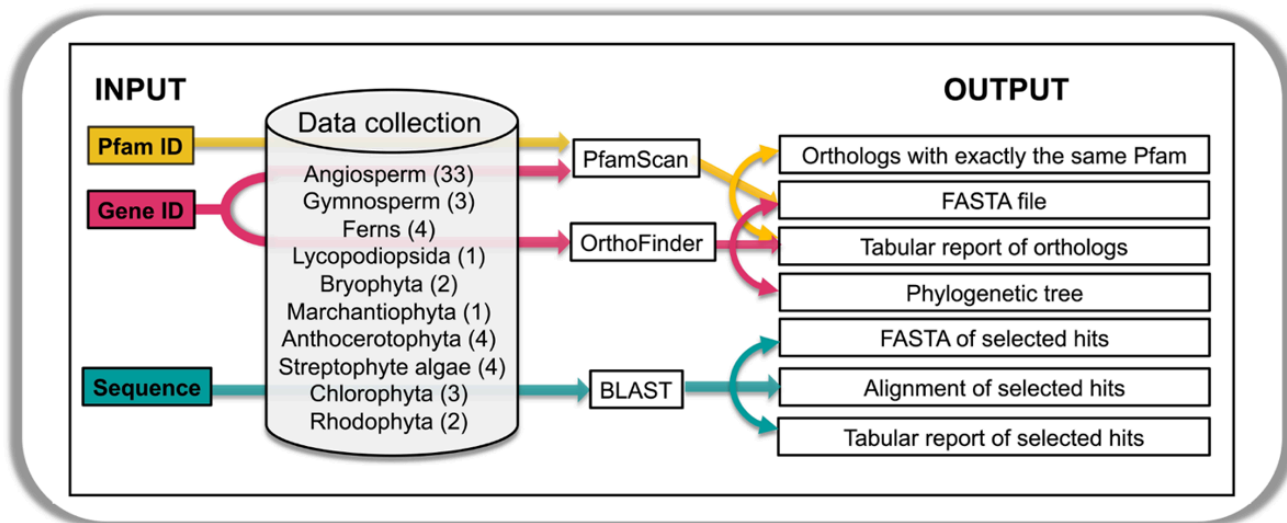


Figure 1

Schematic overview of the Plant Orthologous Group Evolution Database (POGED) data integration and analysis workflow. POGED supports two types of user input, including protein identifiers and protein sequences. Curated protein datasets from representative plant and algal lineages are collected and integrated into a unified reference framework, covering angiosperms, gymnosperms, ferns, lycophytes, bryophytes, streptophyte algae, chlorophytes, and rhodophytes. Domain based queries are analyzed using PfamScan to identify orthologous proteins sharing identical HMM domain compositions, with outputs including ortholog lists and downloadable FASTA files. Blast based queries are processed using OrthoFinder to infer orthologous groups, generating tabular orthology reports, FASTA sequences, and phylogenetic trees. Protein sequence-based queries are analyzed using BLAST to identify homologous sequences within POGED, enabling users with custom or unannotated sequences to retrieve database entries for downstream comparative and evolutionary analyses.

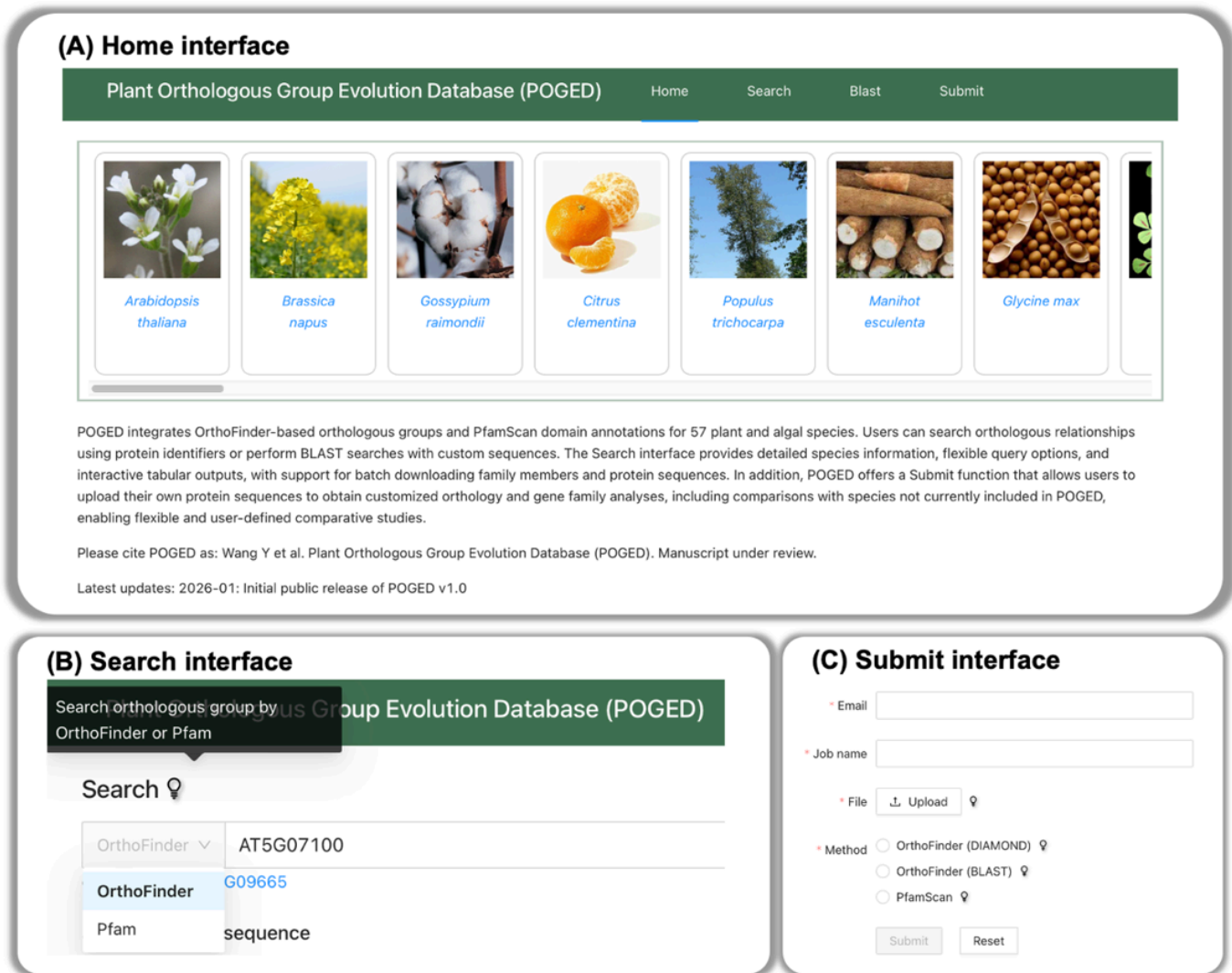


Figure 2

Overview of the user interfaces of the Plant Orthologous Group Evolution Database (POGED). (A) The home interface of POGED, showing an overview of supported plant and algal species and providing direct access to the main functional modules, including Search, BLAST, and Submit. (B) The Search interface, which allows users to query orthologous groups using OrthoFinder based orthogroup identifiers or Pfam domain annotations, as well as to retrieve detailed species information and orthologous relationships through flexible query options. (C) The Submit interface allows users to upload custom protein sequences for user defined orthology or domain based analyses, extending POGED analyses to species not currently included in the database.



Figure 3

Search interfaces and outputs of the Plant Orthologous Group Evolution Database (POGED). (A) OrthoFinder based search interface. (B) PfamScan based search interface for identifying proteins with identical Pfam domain combinations. (C) Tabular OrthoFinder results showing orthologous genes across species. (D) PfamScan search results summarising proteins sharing the same Pfam combinations. (E) Querying by Pfam identifier retrieves all proteins and domain architecture combinations containing the specified Pfam domain across species.

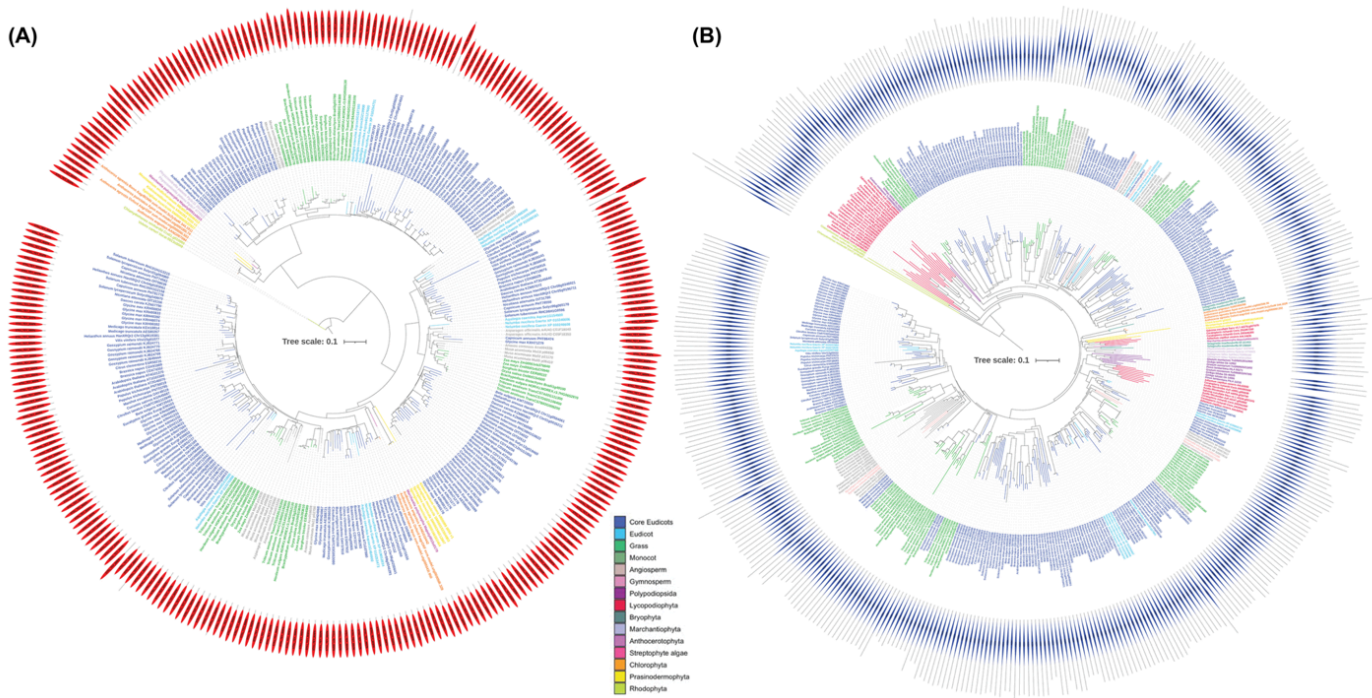


Figure 4

Phylogenetic trees of two MAPK orthogroups identified by OrthoFinder in POGED. (A) The left panel shows the orthogroup enriched for the STKc-TEY-MAPK activation-loop motif identified by CDD search (Wang et al., 2023), highlighted by red ellipses. (B) The right panel shows the orthogroup enriched for the STKc-TDY-MAPK motif (highlighted by blue diamonds). Tip colors denote species clades as indicated in the legend. Sequences lacking canonical TEY/TDY motifs were removed prior to visualization. Trees were annotated in iTOL.

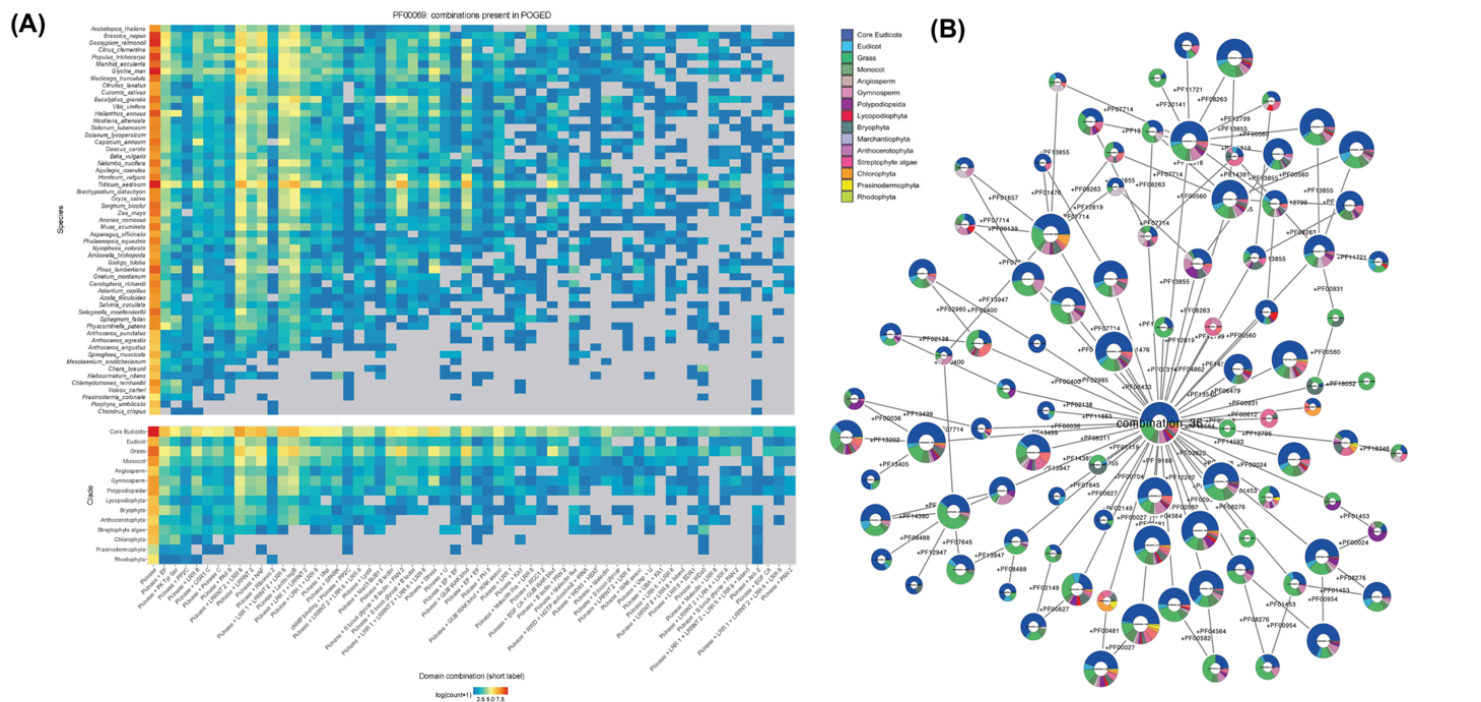


Figure 5

Pfam domain architecture of the MAPK core domain PF00069 across green plant lineages. (A) Heatmaps showing the distribution of PF00069 containing domain architecture combinations across individual species (upper panel) and major phylogenetic clades (lower panel). Each column represents a distinct domain architecture combination containing PF00069, and colours indicate the log transformed protein counts for each species or clade, only Pfam domain names are displayed in the heatmap labels. (B) Network representation of PF00069 based domain architecture diversification. The central node represents the core PF00069 domain, and edges indicate the gain of additional Pfam domains relative to the core architecture. Nodes correspond to distinct PF00069 containing domain combinations. Pie charts illustrate the proportional contribution of different phylogenetic clades to each domain architecture, and node size reflects the total number of species in which the corresponding combination is observed.