e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Volume 10, Issue 2, **440-449.**
https://e-journal.hamzanwadi.ac.id/index.php/edumatic
DOI: 10.29408/edumatic.v10i2.36252

Orthology-Based Reconstruction of Parental Gene Inheritance in the Chromosome-Scale PR25 Perennial Rice Genome

Martha Britany Napitupulu [1,*] , Chandro Pardede [1] , Pandapotan Siagian [1]

[1] Institut Teknologi Del, Indonesia
[*] Corresponding author: Martha Britany Napitupulu, Institut Teknologi Del, Indonesia
✉ iss21051@students.del.ac.id

Received: 19 July 2026 | Revised: 24 July 2026 | Accepted: 19 August 2026 | Published: 26 August 2026

Abstract
Perennial rice integrates a cultivated rice background with genetic elements from wild species; however, the role of parental gene repertoires in hybrid genomes is not fully understood. This study examined the gene repertoires linked to parents and their functional characteristics within the chromosome-scale PR25 perennial rice genome, which originated from Oryza sativa ssp. indica cv. RD23 and *Oryza longistaminata*. Structural annotations were transferred using Liftoff, followed by phylogenetic orthology analysis using OrthoFinder, and functional annotation using eggNOG-mapper v2 and InterProScan. Of the 37,581 structurally annotated PR25 loci, 33,305 were aligned using the parental orthology framework. Most of these genes (20,204; 60.7%) had identifiable orthologs in both parents (BOTH), whereas 12,808 (38.5%) were categorized as RD23_only and 293 (0.9%) as OL_only. Among the 25,828 genes with consistent functional annotations, RD23_only genes exhibited a higher presence of secretory and membrane-associated features, such as SignalP-TM (4.42%), whereas OL_only genes had a higher presence of defense-related features, including NB-ARC (6.22%) and Ser/Thr kinase domains (2.90%). Functional annotation was unresolved for 5,746 RD23_only genes (44.9%). These results provide a detailed catalog of parent-associated gene repertoires and functions in PR25. The small proportion of OL_only should be interpreted with caution because of the fragmented nature of the O. longistaminata assembly. This catalog serves as a foundation for functional validation and molecular breeding research.

Keywords: artificial intelligence; comparative genomics; functional annotation; orthology inference; perennial rice

To cite this article: Napitupulu, M. B., Pardede, C., & Siagian, P. (2026). Orthology-Based Reconstruction of Parental Gene Inheritance in the Chromosome-Scale PR25 Perennial Rice Genome. *Edumatic: Jurnal Pendidikan Informatika*, *10*(2), 440–449. https://doi.org/10.29408/edumatic.v10i2.36252

INTRODUCTION

Perennial rice cultivation has become a viable approach for sustainable rice cultivation, allowing multiple harvests from a single planting. This method can decrease the need for land preparation and replanting while sustaining yield levels (Guo et al., 2023; Zhang et al., 2023). The development of perennial rice has been facilitated by interspecific hybridization between cultivated rice and its perennial wild relatives, which integrates genetic traits for perennial growth and adaptation into cultivated varieties. PR25, which originated from a cross *between Oryza sativa* ssp. indica cv. RD23 and the perennial wild species *Oryza longistaminata* serves as a genomic framework for examining the contributions of cultivated and wild parents to a perennial rice hybrid (Zhang et al., 2023). To comprehend this hybrid genome, it is essential to

identify how the gene repertoires associated with each parent are distributed throughout the genome and to understand their functional roles.

The advent of chromosome-scale and telomere-to-telomere genome assemblies has significantly enhanced the ability of comparative genomics to explore the connections between cultivated, wild, and hybrid species. A telomere-to-telomere reference genome for *Oryza longistaminata* has revealed structural and segmental variations, offering a valuable resource for studying the wild perennial rice lineage (Guang et al., 2025). Concurrently, spatiotemporal transcriptomic analysis of *Oryza longistaminata* has pinpointed transitions linked to rhizome formation, shedding light on the processes that support perennial growth (Lian et al., 2024). These investigations illustrate how genome structure and gene activity provide insights into the biology of perennial rice. However, none of these studies individually clarified the organization of parent-associated gene repertoires within the PR25 hybrid genome.

Comparative genomics and phylogenetic orthology inferences provide a framework for resolving gene correspondence across genomes. Because homologous sequences may arise through speciation, duplication, or gene loss, sequence similarity alone is insufficient; orthology inference therefore helps characterize gene repertoire relationships (Emms & Kelly, 2019). Efficient protein alignment is essential for large-scale analyses, with DIAMOND providing sensitive and computationally efficient genome-wide alignments (Buchfink et al., 2021). However, evolutionary complexity, sequence divergence, and duplication can introduce inference errors (Natsidis et al., 2021). Thus, parent-associated categories should be interpreted as detectable orthologous relationships rather than definitive evidence of gene inheritance or origin. Functional annotation further complements this framework through protein domains and motifs, particularly defense- and signaling-related signatures such as NB-ARC and protein kinase domains in rice (Pan et al., 2022; Jin et al., 2025; Shi et al., 2025). eggNOG-mapper and InterPro provide complementary evidence for functional characterization and domain identification (Cantalapiedra et al., 2021; Hernández-Plaza et al., 2023; Blum et al., 2025).

Progress in artificial intelligence (AI) and machine learning has broadened the scope of genome-scale analyses. AI-enhanced genomics depends on organized biological datasets to discern patterns, aid in functional predictions, and rank genes. Within this framework, genome annotation and comparative genomics offer features such as orthology relationships, functional domains, parent-associated categories, and annotation confidence data to facilitate machine learning-driven prediction and prioritization. Advances in rice functional genomics have highlighted the importance of combining gene annotations, loci, expression data, and genomic variation in genomic analysis and molecular breeding (Kawahara et al., 2026). While comparative genomics focuses on establishing genomic and evolutionary connections, AI extracts predictive patterns from structured genomic data. Consequently, high-quality, biologically structured genomic datasets are crucial for AI-supported functional genomics.

Recent studies have provided supporting evidence for this framework. The enhanced reference genome of *Oryza longistaminata* bolsters the investigation of the parental lineage; however, it fails to reconstruct gene repertoires linked to parents within PR25 (Guang et al., 2025). Although rhizome transcriptomics has shed light on the mechanisms of perennial growth, it does not provide insight into the parental genomic context (Lian et al., 2024). Resources for rice annotation elucidate gene functions and loci, but are not intended to define parent-associated repertoires in interspecific hybrids (Kawahara et al., 2026). Orthological studies have demonstrated reliability; however, they do not connect parental classification with functional characterization in PR25 (Natsidis et al., 2021). Consequently, these elements remain distinct rather than unified for the genomic characterization associated with the parents.

This division results in a gap between the literature and science. There is still a scarcity of genome-wide data identifying which PR25 genes have orthologous connections with the cultivated RD23 parent, wild *O. longistaminata* parent, or both species. Classification linked

to parents without functional analysis offers genomic alignment, but lacks biological insight, whereas functional annotation without parental classification fails to provide a parental context for gene collection. The incomplete assembly of *O. longistaminata* further restricted the detection of orthologs in the wild parent. This study aimed to bridge this gap by combining structural annotation, orthology-based classification, and functional analysis in a chromosome-scale perennial rice hybrid, paving the way for future AI-assisted prediction and candidate selection.

This study outlines the genome-wide distribution of gene repertoires associated with the parents in PR25 and pinpoints functional traits in genes that have orthologous links to RD23, *Oryza longistaminata*, or both parental genomes. This study employed an analytical framework that merged structural annotation transfer using Liftoff, phylogenetic orthology inference using OrthoFinder, and functional annotation using eggNOG-mapper v2 and InterProScan. The novelty of this study does not reside in the individual tools, but in the integration of these components to create a structured catalog of parent-associated genes for a chromosome-scale perennial rice hybrid. Methodologically, it connects genome annotation, orthology, and functional characterization. Biologically, it uncovers asymmetric parent-associated repertoires and annotation uncertainties. Computationally, it provides a resource for AI and machine-learning applications in the breeding of perennial rice.

METHOD

This study employed a computational comparative genomics framework to characterize the parent-associated gene repertoires and functional characteristics of the perennial rice hybrid PR25 (*Oryza sativa* ssp. *indica* cv. RD23 × *Oryza longistaminata*). Genome assemblies of PR25 and RD23 were obtained during an internship at the BGI under restricted-access conditions and were therefore not redistributed. The reference genomes of *O. longistaminata* (GCA_000465655.1) and *O. sativa* ssp. *indica* cv. ZS97 (GCA_002151415.1) was retrieved from Ensembl Plants Release 61 and accessed in January 2024. ZS97 was selected as the reference for annotation transfer because of its high-continuity indica genomic background. The processed datasets, scripts, parameters, and metadata generated in this study were deposited in the project repository to support reproducibility.

The analytical workflow consisted of seven sequential stages: (1) genome quality assessment, (2) structural annotation transfer, (3) gene model standardization and integration, (4) protein sequence extraction, (5) sequence standardization, (6) parental orthology and functional analysis, and (7) integration of orthology and functional evidence for parent-associated classification and functional annotation (Figure 1). Figure shows a seven-stage analytical workflow for inference of parent-associated gene repertoires and functional characteristics in PR25. The workflow comprises genome quality assessment, structural annotation transfer, gene-model standardization, protein extraction, sequence standardization, orthology and functional analysis, and the final integration of parental classification with functional annotation.

Genome completeness was evaluated using BUSCO v5.2.2, with the Poales odb10 lineage in genome mode, while assembly contiguity was assessed using QUAST v5.2.0, in reference-free mode (Manni et al., 2021; Mikheenko et al., 2023). The *Oryza longistaminata* assembly was retained as the wild-parent reference despite lower completeness and fragmentation, which were considered when interpreting *Oryza longistaminata-associated* orthology assignments. Structural gene annotations were transferred from ZS97 to PR25 and RD23 using Liftoff v1.6.2, with parameters -sc 0.95 and -a 0.6. Unique *Oryza longistaminata* gene models were integrated into the PR25 backbone using AGAT v1.5.1 (Dainat et al., 2025). Protein sequences were extracted using GffRead v0.12.7, and processed with AWK scripts before downstream analyses.

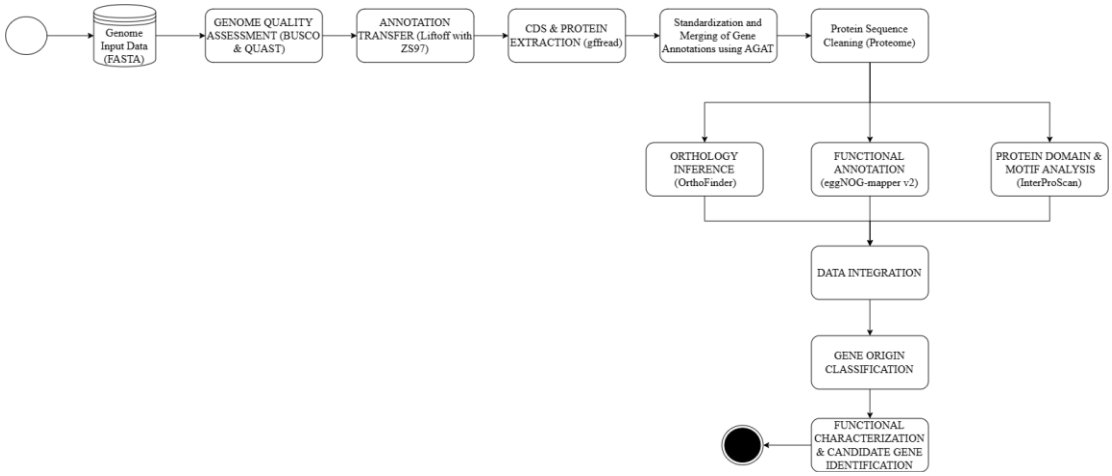

Figure 1. Seven-stage analytical workflow for inference of parental gene origins

Orthologous relationships among PR25, RD23, and *Oryza longistaminata* protein sequences were inferred using OrthoFinder v2.5.5 with DIAMOND for all-versus-all protein alignment and MCL clustering with an inflation parameter of 1.5. PR25 genes with detectable orthologs in both parental genomes were classified as BOTH, genes with detectable orthologs only in RD23 as RD23_only, and genes with detectable orthologs only in *Oryza longistaminata* as OL_only. Genes for which the parental orthology could not be resolved were excluded from these three categories. Ambiguous multicopy relationships were examined using OrthoFinder-generated gene and species trees. These categories represent operational classifications based on detectable orthologous relationships rather than direct evidence of physical gene inheritance.

Functional characterization was performed using eggNOG-mapper v2 and InterProScan as complementary sources of functional evidence. eggNOG-mapper provides orthology-based functional mapping, whereas InterProScan was used to characterize protein domains and functional signatures (Hernández-Plaza et al., 2023). A gene was classified as KNOWN when concordant functional evidence was obtained from both annotation sources; otherwise, it was classified as UNKNOWN.

Functional signatures were summarized by parent-associated categories among KNOWN genes, with differences interpreted as descriptive patterns because no formal enrichment analysis was performed. Orthology and functional annotations were integrated into a structured PR25 gene catalog. All analyses were conducted using Ubuntu 22.04 LTS through WSL2, with workflow scripts, parameters, and metadata deposited in the project repository. The resulting dataset provides a resource for functional validation and potential AI-assisted gene prioritization.

RESULTS AND DISCUSSION

Results

Genome assembly assessment showed high completeness and contiguity for PR25 and RD23, whereas the *Oryza longistaminata* assembly was substantially more fragmented (Table 1). PR25 and RD23 both achieved 99.8% BUSCO completeness, with N50 values of 31.96 Mbp and 32.12 Mbp, respectively. PR25 consisted of 13 contigs, whereas RD23 contained 190 contigs. In contrast, *O. longistaminata* showed 90.2% BUSCO completeness, an N50 of 97.75 kbp, and 22,477 contigs. Its gap density was 10,155.31 N's per 100 kbp, compared with 32.71 and 4.68 N's per 100 kbp for PR25 and RD23, respectively. The assembly metrics therefore indicated substantially greater continuity and completeness for the PR25 and RD23 genomes than for the available *Oryza longistaminata* reference.

Table 1. Genome assembly quality metrics (busco poales_odb10 and quast summary)

Genome	Complete (%)	Complete & Single-Copy (%)	Complete & Duplicated (%)	Fragmented (%)	Missing (%)	N50 (bp)	Contigs	Gap Density (N/100 kbp)
PR25	99.8	98.1	1.6	0.1	0.1	31,960,027	13	32.71
RD23	99.8	98.0	1.7	0.1	0.1	32,118,244	190	4.68
Oryza longistaminata	90.2	87.5	2.6	5.0	4.9	97,754	22,477	10,155.31

Structural annotation transfer from the ZS97 reference mapped 34,800 genes to PR25, corresponding to a mapping rate of 97.52 %, with 885 genes remaining unmapped. For RD23, 35,028 genes were mapped, corresponding to a 98.16% mapping rate, whereas 657 genes remained unmapped. Integration of unique *Oryza longistaminata* gene models using AGAT increased the PR25 structural annotation set to 37,581 loci. Subsequent orthology analysis resolved parental relationships for 33,305 PR25 genes, whereas 4,276 annotated loci could not be assigned to the BOTH, RD23_only, or OL_only categories and were excluded from the downstream parental classification.

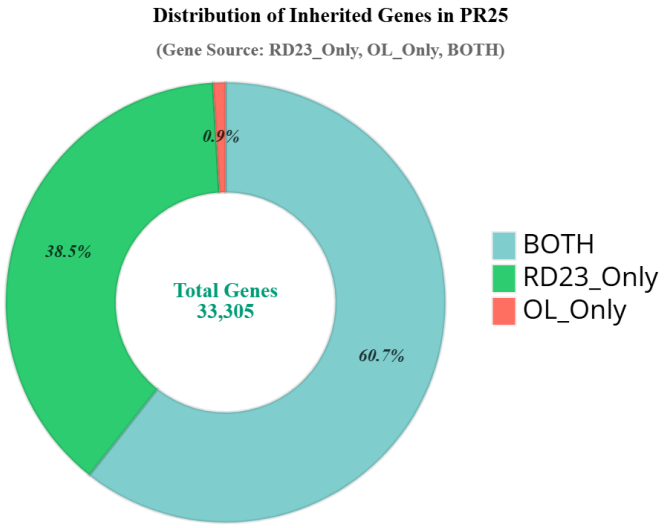

Figure 2. Distribution of pr25 genes across parental orthology

Among the 33,305 PR25 genes with resolved parental orthology, 20,204 (60.7%) were classified as BOTH, 12,808 (38.5%) as RD23_only, and 293 (0.9%) as OL_only (Figure 2). Therefore, the gene repertoire was dominated by shared orthologous relationships, followed by RD23_only genes, whereas OL_only genes represented the smallest category. These classifications reflect detectable orthologous relationships within the analytical framework and should not be interpreted as direct evidence of individual gene inheritance.

Functional annotation identified 25,828 PR25 genes with concordant functional evidence and classified them as KNOWN. Therefore, functional coverage represented 77.5% of the 33,305 genes assigned to the parental orthology framework. The remaining genes were classified as UNKNOWN based on the absence of concordant functional support from the annotation resources. The RD23_only group contained 7,062 KNOWN genes, whereas 5,746 genes (44.9% of the RD23_only category) remained UNKNOWN. Among the OL_only genes, 241 were classified as KNOWN. The distribution indicated that functional annotation coverage differed substantially among parent-associated categories, with the largest unresolved fraction occurring within RD23_only genes.

Figure 3. Top 10 dominant functions across inheritance categories

Functional signature analysis characterized the frequency and relative proportion of the most frequently detected annotations across the parent-associated categories (Figure 3). Across all KNOWN PR25 genes (n = 25,828), the most frequent signatures were SignalP-TM in 400 genes (1.55%), SignalP-TM|SignalP-noTM in 365 genes (1.41%), and SignalP-noTM in 267 genes (1.03%). Among RD23_only KNOWN genes (n = 7,062), SignalP-TM occurred in 312 genes (4.42%), followed by SignalP-TM|SignalP-noTM in 220 genes (3.13%), and SignalP-noTM in 197 genes (2.79%).

Among OL_only KNOWN genes (n = 241), the most frequent signatures included the NB-ARC domain in 15 genes (6.22%) and the Ser/Thr protein kinase family in seven genes (2.90%). Within the BOTH category (n = 18,525), protein kinase domains were detected in 247 genes (1.33%), NB-ARC domains in 159 genes (0.86%), and Cytochrome P450 in 152 genes (0.82%). Additional frequently detected signatures across the PR25 KNOWN gene set included Cytochrome P450 (216 genes; 0.84%), UDP-glycosyltransferases (134 genes; 0.52%), Myb-like DNA-binding domains (134 genes; 0.52%), Ser/Thr protein kinases (125 genes; 0.48%), and PPR repeats (119 genes; 0.46%). These values represent the observed frequencies and proportions of functional signatures; no formal statistical enrichment test was performed. Figure 3 compares the observed distribution of the most frequent functional signatures among the BOTH, RD23_only, and OL_only categories, based on genes classified as KNOWN.

Overall, 37,581 PR25 loci were structurally annotated, with 33,305 loci subsequently assigned to resolved categories of parent-associated orthology. Among these, 20,204 genes (60.7%) were classified as BOTH, 12,808 (38.5%) as RD23_only, and 293 (0.9%) as OL_only.

Concordant functional annotations were obtained for 25,828 genes, representing 77.5% of the classified gene set, whereas 5,746 RD23_only genes (44.9%) remained uncharacterized. The distribution of functional signatures varied across parent-associated categories, with membrane-related signatures more frequently observed in RD23_only genes and defense-related signatures more prominent among OL_only genes.

Discussion

This study provides genome-scale characterization of parent-associated gene repertoires in the chromosome-scale PR25 perennial rice genome by integrating structural annotation, orthology inference, and functional annotation. The analysis revealed a predominantly shared orthologous repertoire alongside a substantial RD23-associated component and a comparatively small *Oryza longistaminata-associated* component. Functional characterization further demonstrated that these categories differed in their observed distributions of functional signatures, whereas a substantial proportion of RD23_only genes remained functionally unresolved. Collectively, these findings indicate that PR25 contains a heterogeneous genomic repertoire that can be systematically characterized through the integration of evolutionary correspondence and functional evidence, rather than through sequence similarity or functional annotation alone.

These findings support a comparative genomics framework in which genomic correspondence and functional characterization provide complementary evidence. The predominance of the BOTH category indicates extensive shared orthology between PR25 and its parents, whereas RD23_only and OL_only reflect asymmetric parent-associated representation. RD23_only genes showed more frequent membrane- and secretion-associated signatures, whereas OL_only genes showed more frequent defense-related signatures. Protein kinase signatures are also biologically relevant, as Lin et al. (2021) demonstrated that kinases play important roles in plant signaling and stress responses, including the regulation of salt stress in rice. However, these patterns remain descriptive rather than constituting evidence of causal effects or functional enrichment. Moreover, orthology categories should be interpreted as detectable evolutionary relationships rather than definitive indicators of gene origin. As emphasized by Emms and Kelly (2019), orthology inference depends on the accuracy of evolutionary relationship reconstruction, while genome completeness and annotation quality may further influence comparative genomic interpretations (Natsidis et al., 2021).

These findings complement recent advances in perennial rice genomics while addressing a distinct analytical question. Zhang et al. (2023) demonstrated the sustained productivity and agronomic potential of perennial rice, while Guo et al. (2023) characterized its physiological performance under repeated harvesting; however, these studies focused primarily on agronomic and physiological characteristics rather than the genome-wide organization of parent-associated gene repertoires. Similarly, Guang et al. (2025) improved the genomic basis for comparative analysis through a telomere-to-telomere *Oryza longistaminata* reference genome, whereas Lian et al. (2024) provided transcriptomic insights into rhizome formation and perennial growth. Although these studies provide complementary evidence on agronomic performance, genome structure, and perennial growth biology, they do not specifically resolve parent-associated gene repertoires within PR25. Therefore, the present study extends previous research by establishing a gene-level orthology framework that connects the PR25 hybrid genome with its cultivated and wild parental backgrounds.

The functional annotation results were also consistent with the broader transition toward integrated functional genomics. eggNOG-mapper and InterProScan provide complementary information on functional relationships and protein domains, enabling the characterization of gene repertoires beyond simple sequence correspondence (Cantalapiedra et al., 2021; Hernández-Plaza et al., 2023; Paysan-Lafosse et al., 2023). The integration of these resources

with parental orthology provides a more structured representation of the PR25 genome than either functional annotation or orthology analysis considered independently. This is particularly relevant in the context of recent rice genomic resources, which increasingly integrate gene annotation, genome variation, expression information, and agronomically relevant loci to support functional genomics and molecular breeding (Kawahara et al., 2026).

One notable finding was the relatively small proportion of OL_only genes compared to RD23_only genes, which should not be interpreted as evidence that the wild parent contributed very few genes to PR25. The lower completeness and greater fragmentation of the *Oryza longistaminata* reference assembly may have reduced the detectability of orthologous relationships, making the observed OL_only proportion a conservative estimate under the current analytical framework (Natsidis et al., 2021). Nevertheless, the presence of NB-ARC-associated signatures within the OL_only category is noteworthy because wild *Oryza* spp. can harbor functionally relevant resistance-associated genes, including NB-ARC-containing candidates identified through genetic and sequence-based analyses (Sinha et al., 2023). This observation supports the value of wild parent-associated genes as candidates for further investigation, although the present analysis does not establish a direct role in disease resistance or other phenotypes.

The principal scientific contribution of this study lies in integrating structural annotation, phylogenetic orthology, and functional annotation to characterize the PR25 perennial rice hybrid genome. This framework distinguishes parent-associated orthological relationships from definitive gene inheritance while revealing conserved and asymmetric components of the hybrid gene repertoire. The study also generated a structured catalog linking PR25 genes with parental orthology and functional annotation, providing a foundation for future AI-assisted genomics. Recent advances have shown that machine learning and protein language models can support genome-wide identification of plant disease-resistance genes (Liu et al., 2026). However, the present study should be considered an AI-ready genomic framework rather than an AI model, as no machine- or deep-learning model was trained or evaluated.

These findings have practical implications for perennial rice genomics and molecular breeding by providing a systematic catalog for prioritizing gene sets for further study. RD23_only genes with unresolved annotations and OL_only genes with defense-associated signatures may serve as candidates for transcriptomic profiling and experimental validation. However, they should not be considered validated breeding targets because no functional assays or statistical enrichment analyses were conducted. Instead, the catalog provides a preliminary evidence layer for integrating genomic, expression, phenotypic, and functional data while improving wild rice reference genomes. Reproducible computational workflows and structured genomic metadata may further strengthen comparative analyses and support future artificial intelligence-based genomic research.

The main limitations include the fragmented *Oryza longistaminata* assembly, which may reduce wild-parent orthology detection; the inability of orthology-based classification to establish direct gene inheritance; the absence of formal enrichment testing; and substantial unresolved functional annotation, particularly among RD23_only genes. In addition, no experimental validation or AI/ML prediction was performed. Future studies should use improved *Oryza longistaminata* assemblies, formal enrichment analyses, and transcriptomic, phenotypic, and experimental validation. The PR25 gene catalog could also support the AI-assisted functional prediction of unresolved genes using well-annotated training datasets and independent validation.

CONCLUSION

This study established an integrated computational framework for characterizing parent-associated gene repertoires and functional features in the chromosome-scale PR25 perennial

rice genome. By combining structural annotation transfer, phylogenetic orthology inference, and complementary functional annotation, 33,305 PR25 genes were resolved into BOTH, RD23_only, and OL_only categories, revealing a predominantly shared gene repertoire along with asymmetric parent-associated components. Functional profiling further identified distinct patterns of membrane- and defense-associated signatures, while substantial functional uncertainty remained among RD23_only genes. These findings provide a structured genomic resource for interpreting parent-associated gene repertoires and prioritizing candidates for subsequent functional validation and molecular breeding. However, the fragmented *Oryza longistaminata* assembly, absence of formal enrichment testing, and lack of experimental or AI/ML validation limit the strength of biological inference. Future research integrating improved wild rice genome assemblies, multi-omics evidence, experimental validation, and AI-assisted functional prediction will be important for advancing the functional interpretation and breeding application of the PR25 genomic resource.

REFERENCES

Blum, M., Andreeva, A., Florentino, L. C., Chuguransky, S. R., Grego, T., Hobbs, E., Pinto, B. L., Orr, A., Paysan-Lafosse, T., Ponamareva, I., Salazar, G. A., Bordin, N., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunic, I., Llinares-López, F., … Bateman, A. (2025). InterPro: The protein sequence classification resource in 2025. *Nucleic Acids Research*, *53*(D1), D444–D456. https://doi.org/10.1093/nar/gkae1082

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Dainat, J., Cannoodt, R., Soares, A., Ruano, D. G., Hereñú, D., Murray, D. K. D., Davis, E., Ugrin, I., Crouch, K., Soler, L., Agostinho, N., pascal-git, Zollman, Z., & tayyrov. (2025). *NBISweden/AGAT: AGATv1.5.1* (Version v1.5.1) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.16317950

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Guang, X., Yang, J., Zhang, S., Guo, F., Li, L., Lian, X., Zeng, T., Cai, C., Liu, F., Li, Z., Hu, Y., Fang, D., He, W., Sahu, S. K., Li, W., Lu, H., Li, Y., Liu, H., Xu, X., … Wei, T. (2025). Telomere-to-telomere African wild rice (Oryza longistaminata) reference genome reveals segmental and structural variation. *GigaScience*, *14*, giaf074. https://doi.org/10.1093/gigascience/giaf074

Guo, C., Lin, W., Gao, W., Lan, C., Xu, H., Zou, J., Fallah, N., Wang, W., Lin, W., Chen, T., & Lin, W. (2023). Physiological Properties of Perennial Rice Regenerating Cultivation in Two Years with Four Harvests. *Plants*, *12*(22). https://doi.org/10.3390/plants12223910

Hernández-Plaza, A., Szklarczyk, D., Botas, J., Cantalapiedra, C. P., Giner-Lamia, J., Mende, D. R., Kirsch, R., Rattei, T., Letunic, I., Jensen, L. J., Bork, P., von Mering, C., & Huerta-Cepas, J. (2023). eggNOG 6.0: Enabling comparative genomics across 12 535 organisms. *Nucleic Acids Research*, *51*(D1), D389–D394. https://doi.org/10.1093/nar/gkac1022

Jin, B., Dong, J., Hu, X., Li, N., Li, X., Long, D., & Wu, X. (2025). RiceReceptor: The cell-surface and intracellular immune receptors of the *Oryza* genus. *Genes, 16*(5), 597. https://doi.org/10.3390/genes16050597

Kawahara, Y., Hirozane-Kishikawa, T., Hirata, R., Wang, X., Tamagaki, Y., Teramoto, Y., Tabei, N., Kumagai, M., Sakai, H., & Itoh, T. (2026). Rice Annotation Project Database (RAP-DB): Literature-Curated Gene Annotation and Integrated Omics Resources for Rice Functional Genomics and Molecular Breeding. *Rice*, *19*(1), 51. https://doi.org/10.1186/s12284-026-00924-6

Lian, X., Zhong, L., Bai, Y., Guang, X., Tang, S., Guo, X., Wei, T., Yang, F., Zhang, Y., Huang, G., Zhang, J., Shao, L., Lei, G., Li, Z., Sahu, S. K., Zhang, S., Liu, H., & Hu, F. (2024). Spatiotemporal transcriptomic atlas of rhizome formation in *Oryza longistaminata*. *Plant Biotechnology Journal*, *22*(6), 1652–1668. https://doi.org/10.1111/pbi.14294

Lin, W., Wang, Y., Liu, X., Shang, J.-X., & Zhao, L. (2021). OsWAK112, A Wall-Associated Kinase, Negatively Regulates Salt Stress Responses by Inhibiting Ethylene Production. *Frontiers in Plant Science*, *12*, 751965. https://doi.org/10.3389/fpls.2021.751965

Liu, Z., Wang, X., Cao, S., Lei, T., Chen, Z., Zhang, M., et al. (2026). Machine learning empowers precise discovery of disease-resistance genes in plants. *Plant Physiology, 201*(3), kiag276. https://doi.org/10.1093/plphys/kiag276

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Mikheenko, A., Saveliev, V., Hirsch, P., & Gurevich, A. (2023). WebQUAST: Online evaluation of genome assemblies. *Nucleic Acids Research*, *51*(W1), W601–W606. https://doi.org/10.1093/nar/gkad406

Natsidis, P., Kapli, P., Schiffer, P. H., & Telford, M. J. (2021). Systematic errors in orthology inference and their effects on evolutionary analyses. *iScience*, *24*(2), 102110. https://doi.org/10.1016/j.isci.2021.102110

Pan, Y.-H., Chen, L., Guo, H.-F., Feng, R., Lou, Q.-J., Rashid, M. A. R., Zhu, X.-Y., Qing, D.-J., Liang, H.-F., Gao, L.-J., Huang, C.-C., Zhao, Y., & Deng, G.-F. (2022). Systematic Analysis of NB-ARC Gene Family in Rice and Functional Characterization of GNP12. *Frontiers in Genetics*, *13*, 887217. https://doi.org/10.3389/fgene.2022.887217

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., … Bateman, A. (2023). Interpro in 2022. *Nucleic Acids Research*, *51*(D1), D418–D427. https://doi.org/10.1093/nar/gkac993

Shi, F., Li, L., Chen, M., Chang, J., Tu, M., He, G., Li, Y., & Yang, G. (2025). Genus-wide pan-genome analysis of *Oryza* calcium-dependent protein kinase genes and their related kinases highlights the complexity of protein kinase evolution. *Plants, 14*(10), 1542. https://doi.org/10.3390/plants14101542

Sinha, P., Kumar T, D., Sk, H., Solanki, M., Gokulan, C. G., Das, A., ... & Sundaram, R. M. (2023). Fine mapping and sequence analysis reveal a promising candidate gene encoding a novel NB-ARC domain derived from wild rice (Oryza officinalis) that confers bacterial blight resistance. *Frontiers in Plant Science*, *14*, 1173063.

Zhang, S., Huang, G., Zhang, Y., Lv, X., Wan, K., Liang, J., Feng, Y., Dao, J., Wu, S., Zhang, L., Yang, X., Lian, X., Huang, L., Shao, L., Zhang, J., Qin, S., Tao, D., Crews, T. E., Sacks, E. J., … Hu, F. (2023). Sustained productivity and agronomic potential of perennial rice. *Nature Sustainability*, *6*(1), 28–38. https://doi.org/10.1038/s41893-022-00997-3