Comparative In Silico Profiling of Anthocyanin Biosynthetic Enzymes in Oryza Species: Evolutionary Conservation versus Functional Diversification

Shravani R[1], Swathi B[1], Dhanalakshmi G[2*]

1, 2 (Corresponding author) Department of Biochemistry, Padmashree Institute of Management and Sciences, Bangalore 560060, Karnataka, India.*

dhanu.bio@gmail.com

Abstract: Anthocyanins, a class of flavonoid pigments, underlie plant coloration, environmental stress resilience, and dietary value. Whether the enzymes governing anthocyanin formation are equally constrained across cultivated rice and its wild relatives, or diverge unevenly along the pathway, remains poorly resolved at the sequence level. We addressed this through a comparative bioinformatic evaluation of three enzymes central to the late stages of anthocyanin biosynthesis—leucoanthocyanidin dioxygenase 1 (LOD1/ANS), glycosyl transferase, and malonyl transferase—across four Oryza taxa: cultivated O. sativa aus N22 and the wild relatives O. nivara, O. rufipogon, and O. punctata. Multiple sequence alignment coupled with phylogenetic reconstruction placed glycosyl transferase as the most conserved of the three loci (92.08–100% identity) and showed its topology tracking genome-level relatedness (AA- versus BB-genome accessions) closely, while malonyl transferase displayed the widest divergence (56.48–95.25%) and, notably, departed from that same genome-relatedness structure in the O. rufipogon lineage. Physicochemical profiling indicated that LOD1 and glycosyl transferase are consistently stable, hydrophilic, acidic polypeptides, whereas malonyl transferase showed more variable stability indices and an isoelectric point range that crosses from acidic to basic across taxa, implying a shift in net charge at physiological pH not seen in the other two enzymes. Subcellular localization predictions placed LOD1 and malonyl transferase in the cytoplasm and glycosyl transferase in the chloroplast, a pattern conserved across all four taxa despite chloroplast targeting being atypical for this enzyme family. Mapping these enzymes onto the KEGG anthocyanin pathway situated this conservation gradient within a broader, chemically plausible logic: catalytically constrained, substrate-specific reactions (LOD1, glycosylation) are tightly conserved, while the terminal, acyl-donor-flexible modification step (malonylation) is comparatively free to diversify. Collectively, these results point to strong purifying constraint on the core biosynthetic machinery alongside more permissive, lineage-specific evolution of the terminal modification step, a pattern with implications for understanding pigment diversity across the genus Oryza.

Keywords: *Oryza sativa*, anthocyanin biosynthesis, LOD1, glycosyl transferase, malonyl transferase, phylogenetic analysis.

1. Introduction

Rice (*Oryza sativa* L.) ranks among the world's most consequential cereal crops, underpinning caloric intake for a substantial share of the global population. Beyond its role as a staple carbohydrate source, the grain accumulates an array of secondary metabolites that contribute to defense, environmental adaptation, and nutritional quality. Anthocyanins, a prominent class of flavonoid pigments within this metabolic repertoire, are responsible for the characteristic coloration of rice grains, leaves, and other tissues. Interest in anthocyanin-pigmented rice varieties has grown considerably owing to their

© The Author(s) 2026
P. Bagchi et al. (eds.), *Proceedings of the International Conference on Information Systems Design And Intelligent Applications (ICISDA 2026)*, Advances in Intelligent Systems Research 215,
https://doi.org/10.2991/978-94-6239-821-4_4

capacity to scavenge free radicals, which is thought to confer protection against oxidative stress and associated chronic disease risk [1, 2].

The biosynthesis of anthocyanins forms a well-defined branch of the broader flavonoid pathway, proceeding through a coordinated sequence of enzymatic transformations under the control of structural and regulatory genes [3]. Its terminal stages involve several enzymes responsible for pigment formation and subsequent chemical modification. Leucoanthocyanidin dioxygenase (LOD1), also known as anthocyanidin synthase (ANS; EC 1.14.20.4), oxidizes colourless leucoanthocyanidin intermediates into the coloured anthocyanidins pelargonidin, cyanidin, and delphinidin—a step that effectively sets the final pigment profile of the plant [4].

Anthocyanidins generated by LOD1 are chemically unstable and require further processing. UDP-dependent glycosyl transferases attach sugar moieties that confer stability, water solubility, and intracellular mobility, shielding the pigments from degradation and enabling their accumulation within vacuoles. A subsequent round of acylation, carried out by malonyl transferases, appends malonyl groups to these anthocyanin glucosides. Together, these modifications reinforce pigment stability, modulate colour expression, and generate the structural diversity characteristic of plant anthocyanins [5].

Wild relatives of cultivated rice—*O. nivara*, *O. rufipogon*, and *O. punctata*—offer a valuable comparative resource for tracing rice evolution and identifying alleles linked to stress tolerance, pigmentation, and metabolic variation. Critically, these taxa are not equally related to cultivated rice at the genome level: *Oryza* encompasses species with markedly different genomic architectures, spanning both AA-genome accessions (which include cultivated rice together with *O. nivara* and *O. rufipogon*) and the more distantly related BB-genome *O. punctata*—contrasting cultivated and wild lineages across this genomic divide provides a window into how genes are conserved, diverge, and acquire new functions over evolutionary time, and offers a basis for asking whether individual anthocyanin biosynthetic loci track this genome-level structure uniformly or diverge from it [6, 7].

Advances in bioinformatics now make it possible to characterize genes and their protein products in considerable depth—through sequence comparison, phylogenetic reconstruction, prediction of physicochemical behaviour, and inference of subcellular targeting—without recourse to extensive wet-laboratory experimentation. These computational approaches offer an efficient route to inferring evolutionary relationships, structural attributes, and probable biological roles [7].

Despite the well-established enzymology of the anthocyanin pathway, it remains unclear whether the three terminal biosynthetic enzymes are subject to comparable evolutionary constraint, or whether their divergence patterns diverge from one another—and from the broader genomic relationships among taxa—in ways that reflect their distinct catalytic roles. Building on this rationale, the present study set out to comparatively characterize three anthocyanin biosynthetic enzymes—LOD1, glycosyl transferase, and malonyl transferase—across four *Oryza* taxa (*O. sativa* aus N22, *O. nivara*, *O. rufipogon*, and *O. punctata*) using an *in silico* framework. Specifically, the analysis aimed to

(i) assess sequence similarity and infer evolutionary relationships through multiple sequence alignment (MSA) and phylogenetic reconstruction;

(ii) characterize the physicochemical properties of the corresponding proteins;

(iii) predict their subcellular localization; and

(iv) interpret the functional contribution of each enzyme to anthocyanin biosynthesis.

Taken together, these objectives were designed to yield a comparative framework for understanding how the anthocyanin pathway has been conserved and diversified across cultivated rice and its wild relatives.

2. Materials and Methodology

This study drew on a series of publicly accessible bioinformatics databases and computational tools to identify, characterize, and functionally annotate anthocyanin biosynthesis-related enzymes across the four selected *Oryza* taxa (Fig. 1).

2.1 Retrieval of mRNA and Protein Sequences

Genomic data for the four *Oryza* species examined here were sourced from the Rice Genome Hub (https://rice-genome-hub.southgreen.fr/organisms), an open-access repository of plant genome assemblies, gene annotations, and comparative genomic resources. Three enzymes central to anthocyanin biosynthesis—leucoanthocyanidin dioxygenase 1 (LOD1), glycosyl transferase, and malonyl transferase—were selected as the focus of this analysis, and their corresponding mRNA and amino acid sequences were retrieved from this database for subsequent computational analysis.

2.2 Multiple Sequence Alignment (MSA) and Phylogenetic Analysis

mRNA and protein sequences for the three target enzymes were obtained from the Rice Genome Hub across the cultivated and wild *Oryza* accessions considered here. Sequence alignment was carried out using Clustal Omega (EMBL-EBI; https://www.ebi.ac.uk/jdispatcher/msa/clustalo/), an open-source tool for nucleotide and protein alignment. Each gene was aligned independently across the four species, and the resulting alignments were used to construct phylogenetic trees (phylograms) so as to gauge sequence conservation and evolutionary relatedness.

2.3 Protein Physicochemical Characterization

Amino acid sequences of the three enzymes were submitted to ExPASy ProtParam (https://web.expasy.org/protparam/), a tool maintained by the Swiss Institute of Bioinformatics for predicting protein physicochemical parameters. Molecular weight, theoretical isoelectric point, amino acid and atomic composition, extinction coefficient, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) were computed for each sequence, providing a basis for comparing the stability, hydrophobicity, and structural character of LOD1, glycosyl transferase, and malonyl transferase.

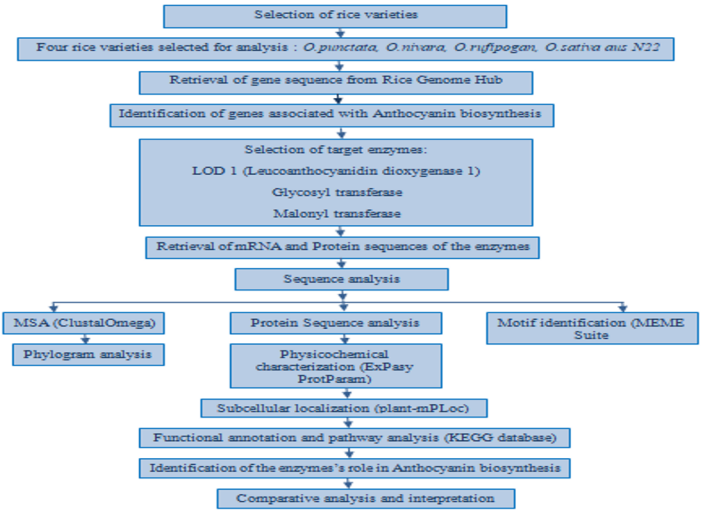


Fig. 1. Workflow of the study

2.4 Protein Motif Identification

Conserved motifs within the three protein families were identified using the MEME Suite (https://meme-suite.org/meme/), a platform for motif discovery in DNA, RNA, and protein sequences. Sequences from all four *Oryza* species were screened under the following settings:

- Number of motifs searched: 5
- Motif distribution: zero or one occurrence per sequence (ZOOPS)
- Motif width: 6–600 amino acids
- Number of motif sites: 2–600

The resulting motifs were compared across species to identify conserved functional regions and to corroborate the sequence similarity patterns obtained from alignment.

2.5 Subcellular Localization Prediction

Subcellular localization of the three target proteins was predicted using Plant-mPLoc (Cell-PLoc; http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/), a platform developed specifically for predicting the cellular compartmentalization of plant proteins from amino acid sequence. These predictions were used to infer the likely intracellular sites at which LOD1, glycosyl transferase, and malonyl transferase operate during anthocyanin biosynthesis.

2.6 Functional Annotation and Pathway Analysis

The metabolic roles and pathway context of the three enzymes were established using the KEGG database (https://www.genome.jp/kegg/). Their involvement in the flavonoid, flavanone, flavonol, and anthocyanin biosynthetic pathways was examined to clarify how these enzymes act sequentially in pigment formation, glycosylation, modification, and stabilization across the four *Oryza* species.

3. Results and Discussion

3.1 Multiple Sequence Alignment (MSA) and Phylogenetic Analysis

Sequence alignment of the LOD1, glycosyl transferase, and malonyl transferase genes across *O. sativa* aus N22, *O. nivara*, *O. rufipogon*, and *O. punctata* was performed with Clustal Omega to characterize sequence conservation and infer evolutionary relationships among the four taxa.

3.1.1 LOD1

Pairwise identity for LOD1 spanned 70.21–97.87%. The closest pairings were *O. punctata–O. nivara* and *O. sativa* aus N22–*O. rufipogon*, both at 97.87% identity, while *O. rufipogon* and *O. nivara* formed the most divergent pair (70.21%) (Table 1). In the corresponding phylogenetic tree, *O. sativa* aus N22 and *O. nivara* clustered together, with *O. rufipogon* joining this group, while *O. punctata* formed a separate branch (Fig. 2). This topology is consistent with established relationships within *Oryza*, in which AA-genome species cluster more closely with one another than with BB-genome taxa such as *O. punctata* [7, 8]. Notably, the pairwise values themselves do not track genome grouping in a simple way: *O. rufipogon* and *O. nivara*, both AA-genome accessions, form the single most divergent pair in the LOD1 dataset—more divergent from one another than either is from the BB-genome *O. punctata* in several comparisons. This indicates that within-AA-genome variation at this locus is not negligible, and that broad genome relatedness sets the overall topology without fully predicting locus-specific divergence.

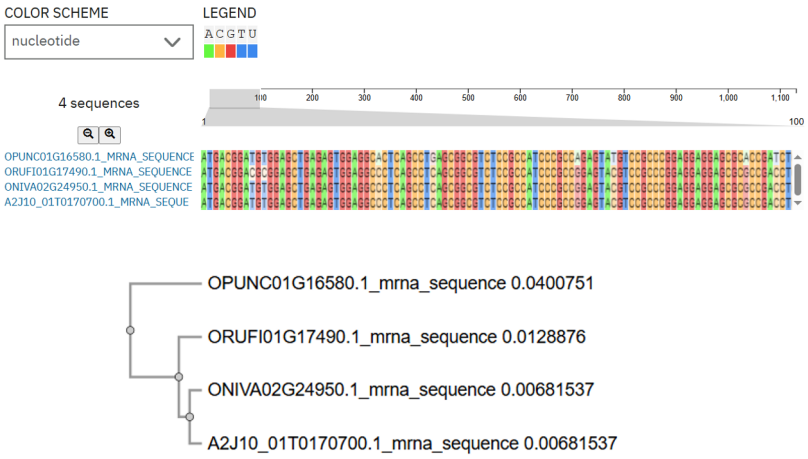


Fig. 2. Multiple sequence alignment and phylogenetic relationship of LOD1 mRNA sequences among O. punctata, O. rufipogon, O. nivara, and O. sativa

3.1.2 Glycosyl Transferase

Of the three genes examined, glycosyl transferase was the most conserved, with pairwise identities ranging from 92.08% to 100%. *O. sativa* aus N22 and *O. nivara* were identical at this locus, and *O. rufipogon* shared 99.93% identity with both (Table 1). Phylogenetic reconstruction grouped these three AA-genome accessions in a single cluster, leaving *O. punctata* as an independent lineage (Fig. 3). Conservation of this magnitude across four taxa spanning two genome types is difficult to attribute to genome relatedness alone, since the AA/BB split still permits appreciable divergence elsewhere in the dataset (see 3.1.1 and 3.1.3). A more parsimonious explanation is functional: glycosyl transferase

catalyzes a chemically constrained reaction—UDP-glucose-dependent glycosylation—using a substrate-binding architecture broadly conserved across plant secondary metabolism, leaving little sequence space free to vary without compromising catalysis [9]. The conservation observed here is therefore consistent with strong purifying selection acting on a functionally constrained enzyme, rather than being a rice-specific outcome.

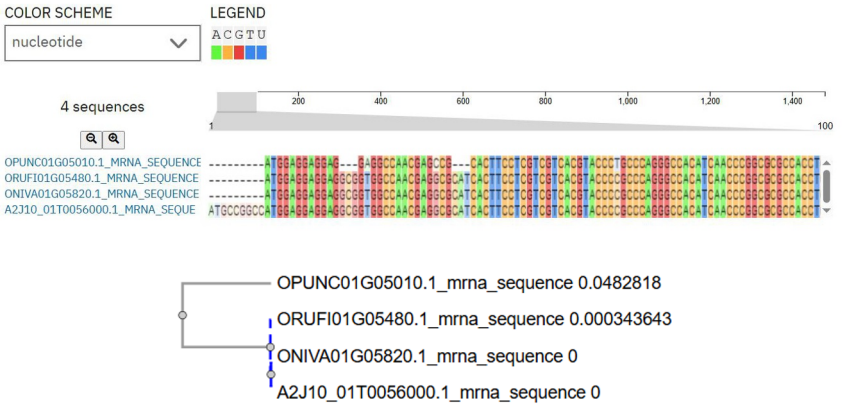


Fig. 3. Multiple sequence alignment and phylogenetic relationship of glycosyl transferase mRNA sequences among O. punctata, O. rufipogon, O. nivara, and O. sativa

3.1.3 Malonyl Transferase

Malonyl transferase displayed the greatest divergence of the three loci, with pairwise identities ranging from 56.48% to 95.25%. *O. sativa* aus N22 and *O. nivara* were most similar (95.25%), while *O. rufipogon* diverged markedly from the remaining taxa (56.48–58.31%) (Table 1). As with the other two genes, the phylogenetic tree placed *O. sativa* aus N22 and *O. nivara* together, while *O. punctata* and *O. rufipogon* fell into separate branches (Fig. 4). This is the clearest case in the dataset where genome relatedness and locus-specific divergence decouple: *O. rufipogon* clusters with the other AA-genome accessions for both LOD1 and glycosyl transferase, yet diverges sharply from them at this locus—including from its fellow AA-genome relatives. Whatever evolutionary process has acted on malonyl transferase in the *O. rufipogon* lineage has therefore done so independently of the broader genomic relationships that otherwise structure this dataset. This is in line with evidence that plant anthocyanin malonyltransferases diversify rapidly and often lineage-specifically, reflecting variation in acyl-donor specificity and downstream modification chemistry rather than tracking overall genome phylogeny [10]; a comparable pattern of lineage-specific malonyltransferase evolution, decoupled from broader relatedness, has also been reported outside *Oryza*, in the inheritance and enzymatic basis of anthocyanin malonylation in cultivated strawberry [17].

Table 1. Pairwise sequence alignment scores (%) of LOD1, glycosyl transferase, and malonyl transferase genes among Oryza species

Pairwise Comparison	LOD1 (%)	Glycosyl Transferase (%)	Malonyl Transferase (%)

Pairwise Comparison	LOD1 (%)	Glycosyl Transferase (%)	Malonyl Transferase (%)
O. punctata vs. O. nivara	97.87	92.08	82.13
O. rufipogon vs. O. nivara	70.21	99.93	56.48
O. punctata vs. O. rufipogon	93.21	92.15	58.31
O. sativa aus N22 vs. O. nivara	71.01	100.00	95.25
O. sativa aus N22 vs. O. punctata	93.29	92.08	85.90
O. sativa aus N22 vs. O. rufipogon	97.87	99.93	56.65

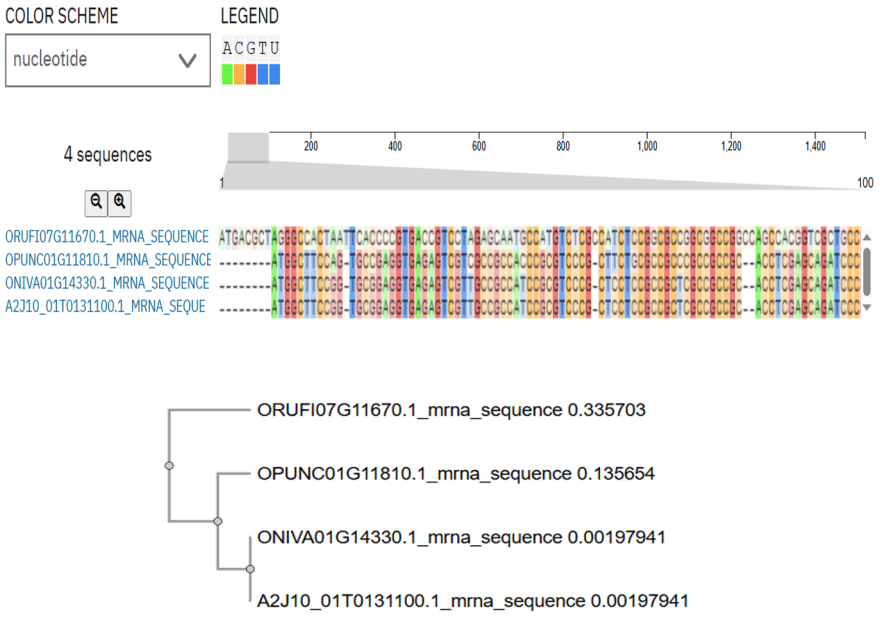


Fig. 4. Multiple sequence alignment and phylogenetic relationship of malonyl transferase mRNA sequences among O. punctata, O. rufipogon, O. nivara, and O. sativa

Read together, the three phylograms indicate that AA/BB-genome relatedness is the dominant signal for two of the three loci examined—LOD1 and glycosyl transferase—but that malonyl transferase evolves substantially outside that structure, at least in the *O. rufipogon* lineage. This positions malonyl transferase not merely as "the least conserved of the three" but as a locus whose evolutionary

trajectory is decoupled from the phylogenetic backbone that otherwise organizes this dataset, a distinction with a plausible mechanistic basis discussed further in Section 3.4.

3.2 Protein Physicochemical Properties

ProtParam analysis uncovered marked differences in the physicochemical behaviour of the three enzyme families (Table 2a–c).

Theoretical isoelectric points for LOD1 (5.20–5.39) and glycosyl transferase (4.99–5.74) fell consistently in the acidic range, whereas malonyl transferase spanned a considerably wider interval (5.97–8.42)—not merely a wider range, but one that crosses from acidic to basic. Because a protein's net charge at physiological pH is set by its pI relative to cytosolic pH (~7.0–7.5), this is not a cosmetic difference: the *O. rufipogon* malonyl transferase (pI 8.42) is predicted to carry a net positive charge under conditions where its counterparts in the other three taxa would be net negative or near-neutral. Charge state at physiological pH influences substrate-binding electrostatics and protein interaction surfaces, so this shift plausibly reflects genuine functional divergence in the *O. rufipogon* enzyme rather than neutral sequence drift, consistent with this taxon also standing out as the sequence-level outlier for this locus (Section 3.1.3). More broadly, this variation in pI reflects underlying differences in amino acid composition, which can shape protein folding, solubility, and catalytic behaviour [11].

Molecular weight was tightly conserved for glycosyl transferase (~51 kDa across all four species), whereas LOD1 showed greater variability—most notably in *O. nivara*—and malonyl transferase ranged from 44.8 to 50.7 kDa, in keeping with its comparatively higher sequence divergence.

Instability index values placed LOD1 and glycosyl transferase within the stable range (<40) across all species, while malonyl transferase exceeded this threshold in three of the four taxa (>40). This distinction implies that LOD1 and glycosyl transferase retain more conserved structural features, whereas malonyl transferase may tolerate greater conformational flexibility [12]. Here again, *O. rufipogon* is the exception rather than the rule: its instability index (39.56) falls just under the stability threshold despite this taxon being the most sequence-divergent at this locus, indicating that sequence divergence and predicted structural stability are not simply co-varying in this enzyme.

All three enzymes exhibited high aliphatic indices (>86), pointing to enhanced thermostability, since elevated aliphatic index values are generally associated with greater thermal resilience in globular proteins. The correspondingly negative or near-zero GRAVY scores indicate that these enzymes are predominantly hydrophilic and likely to be readily soluble in the aqueous cytosolic environment, consistent with their function as soluble catalytic proteins within plant cells [13].

Table 2a. Physicochemical properties of the proteins

Gene	Species	Theoretical pI	Molecular Weight (Da)	Instability Index	Aliphatic Index	GRAVY
LOD1	O. sativa aus N22	5.38	40,708.01	35.15	96.56	-0.195
LOD1	O. nivara	5.39	29,116.77	36.19	86.38	-0.297
LOD1	O. punctata	5.20	40,663.86	39.61	94.40	-0.239
LOD1	O. rufipogon	5.39	40,644.03	35.78	96.32	-0.184
Glycosyl Transferase	O. sativa aus N22	5.74	51,405.12	37.45	89.43	0.066
Glycosyl Transferase	O. nivara	5.74	51,105.73	36.32	89.77	0.062

Gene	Species	Theoretical pI	Molecular Weight (Da)	Instability Index	Aliphatic Index	GRAVY
Glycosyl Transferase	O. punctata	4.99	51,708.00	37.49	86.92	-0.051
Glycosyl Transferase	O. rufipogon	5.74	51,139.75	36.32	88.97	0.060
Malonyl Transferase	O. sativa aus N22	6.64	46,984.64	54.03	91.68	0.063
Malonyl Transferase	O. nivara	6.87	44,831.09	52.61	89.52	0.037
Malonyl Transferase	O. punctata	5.97	48,746.56	53.05	89.67	0.025
Malonyl Transferase	O. rufipogon	8.42	50,698.81	39.56	90.30	-0.071

Patterns in charged amino acid content mirrored the pI results, and the amino acid and atomic composition data track with overall protein size. Extinction coefficient values were highest for glycosyl transferase, followed by malonyl transferase in *O. punctata* and then LOD1, pointing to differences in aromatic amino acid content among the three enzyme families.

Table 2b. Physicochemical properties of the proteins

Gene	Species	Positively Charged Residues	Negatively Charged Residues	Atomic Composition	Extinction Coefficient (M-1 cm-1)
LOD1	O. punctata	38	55	C1807 H2842 N514 O546 S5	40910
LOD1	O. nivara	29	39	C1291 H2016 N372 O388 S5	30940
LOD1	O. rufipogon	39	54	C1808 H2850 N518 O538 S6	39420
LOD1	O. sativa aus N22	38	54	C1811 H2850 N520 O540 S5	39420
Glycosyl Transferase	O. punctata	46	65	C2279 H3554 N658 O697 S12	67380
Glycosyl Transferase	O. nivara	47	55	C2254 H3544 N660 O675 S13	68870
Glycosyl Transferase	O. rufipogon	47	55	C2257 H3542 N660 O675 S13	68870
Glycosyl Transferase	O. sativa aus N22	47	55	C2267 H3565 N663 O678 S14	68870
Malonyl Transferase	O. punctata	47	54	C2165 H3409 N629 O628 S14	45950
Malonyl Transferase	O. nivara	46	47	C1998 H3136 N590 O568 S10	44460
Malonyl Transferase	O. rufipogon	57	54	C2259 H3559 N669 O643 S10	44920
Malonyl Transferase	O. sativa aus N22	47	49	C2097 H3301 N615 O596 S10	44460

	LOD 1				Glycosyl transferase				Malonyl transferse			
	O.punctata	O.nivara	O.rufipogon	O.sativa aus N22	O.punctata	O.nivara	O.rufipogon	O.sativa aus N22	O.punctata	O.nivara	O.rufipogon	O.sativa aus N22
Ala (A)	10.70%	14.20%	13.10%	12.50%	15.30%	16.90%	16.90%	17.00%	15.50%	16.20%	12.60%	16.30%
Arg (R)	7.00%	7.10%	7.20%	7.20%	7.90%	8.30%	8.30%	8.20%	9.50%	9.80%	10.30%	9.50%
Asn (N)	2.10%	3.00%	2.40%	2.70%	2.50%	2.10%	2.10%	2.10%	0.90%	0.70%	0.40%	0.70%
Asp (D)	7.80%	7.50%	7.50%	7.20%	7.90%	7.00%	7.00%	7.00%	6.60%	7.10%	7.10%	6.80%
Cys (C)	0.80%	0.70%	0.80%	0.80%	1.20%	1.20%	1.20%	1.20%	1.30%	1.20%	1.70%	1.10%
Gln (Q)	3.20%	2.60%	2.10%	2.10%	2.30%	2.10%	2.10%	2.10%	1.80%	1.90%	1.70%	2.00%
Glu (E)	7.60%	7.10%	6.90%	7.20%	5.60%	4.30%	4.30%	4.30%	5.30%	4.00%	4.50%	4.30%
Gly (G)	8.30%	7.80%	8.00%	8.00%	9.10%	9.70%	9.70%	9.70%	6.40%	8.10%	9.20%	7.70%
His (H)	3.50%	3.40%	3.70%	4.00%	2.10%	2.10%	2.10%	2.10%	2.60%	2.90%	3.40%	2.70%
Ile (I)	3.50%	2.60%	2.70%	2.90%	2.70%	2.70%	2.70%	2.70%	2.00%	2.90%	2.60%	2.70%
Leu (L)	11.60%	10.10%	12.50%	12.30%	7.20%	7.20%	7.20%	7.20%	9.50%	8.30%	10.70%	8.80%
Lys (K)	3.20%	3.70%	3.20%	2.90%	1.70%	1.40%	1.40%	1.40%	0.90%	1.20%	1.90%	1.10%
Met (M)	0.50%	1.10%	0.80%	0.50%	1.20%	1.40%	1.40%	1.60%	1.80%	1.20%	0.40%	1.10%
Phe (F)	2.70%	3.40%	2.70%	2.70%	3.50%	2.90%	2.90%	2.90%	4.40%	4.80%	4.70%	4.50%
Pro (P)	6.70%	6.70%	6.90%	6.90%	4.50%	4.50%	4.50%	4.70%	7.90%	7.90%	6.60%	8.40%
Ser (S)	5.40%	5.20%	5.10%	5.10%	5.40%	5.20%	5.20%	5.10%	6.80%	6.20%	5.60%	6.10%
Thr (T)	3.50%	2.20%	2.70%	2.90%	4.30%	4.50%	4.50%	4.50%	4.00%	2.90%	4.70%	2.90%
Trp (W)	1.30%	1.50%	1.30%	1.30%	1.90%	1.90%	1.90%	1.80%	1.50%	1.70%	1.30%	1.60%
Tyr (Y)	2.40%	2.20%	2.10%	2.10%	2.50%	2.70%	2.70%	2.70%	1.10%	1.00%	1.70%	0.90%
Val (V)	9.40%	7.80%	8.30%	8.50%	11.40%	11.80%	11.80%	11.70%	10.20%	10.20%	9.00%	10.40%

Table 2c. Physicochemical properties of the proteins (amino acid composition)

3.3 Subcellular Localization

Predicted subcellular localization was fully conserved across all four *Oryza* species. LOD1 and malonyl transferase were both predicted to reside in the cytoplasm, while glycosyl transferase was assigned to the chloroplast [14].

The cytoplasmic placement of LOD1 is consistent with its position in the late stages of flavonoid biosynthesis, where anthocyanidins are generated prior to further chemical modification. Malonyl transferase likewise appears to act in the cytoplasm, transferring malonyl groups onto anthocyanin glucosides ahead of their eventual sequestration in the vacuole.

The predicted chloroplast localization of glycosyl transferase raises the possibility that glycosylation occurs in close association with plastid-derived metabolites, or that this enzyme participates in a more specialized branch of flavonoid metabolism. Although cytosolic localization is typical of many plant glycosyltransferases, several members of this enzyme family have also been localized to plastids and other organelles, depending on substrate specificity and the presence of targeting peptides [15].

This assignment nonetheless warrants a methodological caveat. Tools such as Plant-mPLoc infer localization from sequence-derived features—targeting-peptide motifs and amino acid composition—rather than direct experimental evidence, and single-location calls for proteins with weak or ambiguous targeting signals carry non-trivial uncertainty. The fact that the chloroplast assignment recurs identically across all four independently retrieved and independently analyzed sequences argues against it being a stochastic artifact of any single sequence, and lends the prediction more weight than a single-species result would. It remains, however, a computational inference rather than

a confirmed localization, and is best read as a hypothesis for experimental follow-up—for example, fluorescent fusion-protein localization or compartment-resolved proteomic profiling—rather than an established finding.

The uniformity of this localization pattern across all four species suggests that intracellular targeting of these three enzymes has remained stable over the course of *Oryza* evolution.

3.4 Contribution of the Three Enzymes to Anthocyanin Biosynthesis

Together, LOD1, glycosyl transferase, and malonyl transferase catalyze the terminal steps of anthocyanin biosynthesis (Table 3).

Leucoanthocyanidin dioxygenase (LOD1/ANS; EC 1.14.20.4) oxidizes leucoanthocyanidins into pelargonidin, cyanidin, and delphinidin, a step that effectively determines the pigment profile of the plant [16].

Anthocyanidin 3-O-glucoside 5-O-glucosyltransferase (EC 2.4.1.298) subsequently transfers glucose onto anthocyanidin 3-O-glucosides to yield anthocyanidin 3,5-di-O-glucosides. This glycosylation step markedly improves pigment stability, water solubility, and vacuolar accumulation while protecting the pigment from enzymatic breakdown [9].

Malonyl transferases (EC 2.3.1.171, EC 2.3.1.172, and EC 2.3.1.214) then malonylate these anthocyanin glucosides, a modification that further stabilizes the pigment, intensifies colour expression, and supports efficient vacuolar transport and long-term storage. The pronounced sequence divergence observed for malonyl transferase in this study is consistent with a role in generating species-specific anthocyanin diversity and in mediating environmental adaptation [10].

The evolutionary pattern established above maps onto this reaction sequence in a way that is chemically sensible rather than coincidental. LOD1 and glycosyl transferase—the two loci showing tight sequence conservation, narrow physicochemical parameter ranges, and taxon-invariant localization—catalyze reactions with limited substrate flexibility: LOD1 performs a defined oxidation of a small set of leucoanthocyanidin substrates, and the glycosyltransferase examined here transfers a single defined donor (UDP-glucose) to a specific acceptor position. Malonyl transferase, by contrast, belongs to a family known to use variable acyl-CoA donors and to act on positionally distinct hydroxyl groups on the anthocyanin glucoside, producing structurally diverse malonylated products across plant taxa [10]. A locus with more chemically flexible catalytic requirements has correspondingly more sequence space over which mutations can be tolerated without loss of core function—consistent with, and offering a mechanistic account for, the sequence divergence, physicochemical variability, and phylogenetic decoupling from genome relatedness observed for malonyl transferase in this study. A comparable pattern of comparatively unconstrained, lineage-specific malonyltransferase evolution generating chemotype diversity has also been documented in strawberry (*Fragaria* × *ananassa*) [17], suggesting that the terminal-modification-enzyme flexibility observed here in *Oryza* may reflect a broader property of plant anthocyanin malonyltransferases rather than a genus-specific idiosyncrasy—a comparison that would benefit from direct cross-genus sequence analysis in future work.

Table 3. Function of selected enzymes in anthocyanin biosynthesis

Enzyme	EC Number	Reaction Catalyzed	Role in Anthocyanin Biosynthesis
Leucoanthocyanidin dioxygenase (LOD1/ANS)	EC 1.14.20.4	Leucoanthocyanidins → Anthocyanidins	Catalyzes the formation of coloured anthocyanidins

Enzyme	EC Number	Reaction Catalyzed	Role in Anthocyanin Biosynthesis
			from colourless precursors.
Anthocyanidin 3-O-glucoside 5-O-glucosyltransferase (5GT)	EC 2.4.1.298	Anthocyanidin 3-O-glucoside + UDP-glucose → Anthocyanidin 3,5-di-O-glucoside + UDP	Glycosylates anthocyanins, increasing their stability, solubility, and accumulation.
Anthocyanin malonyl transferase	EC 2.3.1.171, EC 2.3.1.172, EC 2.3.1.214	Malonylation of anthocyanin glucosides	Stabilizes anthocyanins, enhances pigmentation, and facilitates vacuolar storage.

Considered as a whole, these three enzymes function in a coordinated sequence: LOD1 establishes the anthocyanidin backbone, glycosyl transferase stabilizes the resulting pigment through glycosylation, and malonyl transferase carries out the final acylation step that reinforces stability and colour expression. The strong conservation observed for LOD1 and glycosyl transferase, set against the comparatively greater divergence of malonyl transferase, points to an anthocyanin pathway in which the core biosynthetic machinery has been evolutionarily conserved while the terminal modification enzyme has diversified to generate structural and functional variation among *Oryza* species.

3.5 KEGG Pathway Context of Anthocyanin Biosynthesis

The three enzymes examined here catalyze the terminal reactions of the anthocyanin biosynthesis pathway (KEGG: ko00942). Anthocyanin production begins within the broader flavonoid pathway, where phenylalanine is progressively converted into flavonoid precursors by PAL, CHS, CHI, and F3H. Formation of the anthocyanidins themselves is governed by leucoanthocyanidin dioxygenase (LOD1/ANS; EC 1.14.20.4), which oxidizes leuco-anthocyanidins into pelargonidin, cyanidin, and delphinidin [17].

The resulting anthocyanidins are inherently unstable and are subsequently modified by glycosyl transferases (EC 2.4.1.298), which install sugar moieties using UDP-glucose as the donor substrate. This glycosylation step enhances pigment stability, solubility, and tissue-level accumulation. Malonyl transferases (EC 2.3.1.171/172/214) act next, appending malonyl groups to the resulting anthocyanin glucosides to yield stable, acylated derivatives with improved pigmentation and storage characteristics [17].

The conservation of these three enzymes across the *Oryza* species examined here reinforces their essential role in anthocyanin biosynthesis. High sequence conservation for LOD1 and glycosyl transferase points to strong functional constraint, while the greater divergence of malonyl transferase suggests species-specific variation in how anthocyanins are chemically modified. Overall, the KEGG pathway mapping confirms that these three enzymes act in a defined sequence to govern anthocyanin formation, stabilization, and diversification across rice species. The conservation gradient identified here—tight constraint on LOD1 and glycosyl transferase, comparatively loose constraint on malonyl transferase—is consistent with a broader tendency in specialized plant metabolism for constraint to relax toward the periphery of a pathway: core, non-substitutable steps tend to be conserved, while terminal modification steps, which diversify chemical output without altering the core scaffold, are freer to vary and often underlie chemotype and species-level diversity. The present dataset is consistent with, rather than establishing, that general pattern, since only three loci and four taxa were examined; extending this comparison to additional pathway-adjacent genes (PAL, CHS, CHI, F3H) across the same four accessions would be a natural next step to test whether the conservation gradient holds across the full pathway or is specific to the three terminal enzymes examined here.

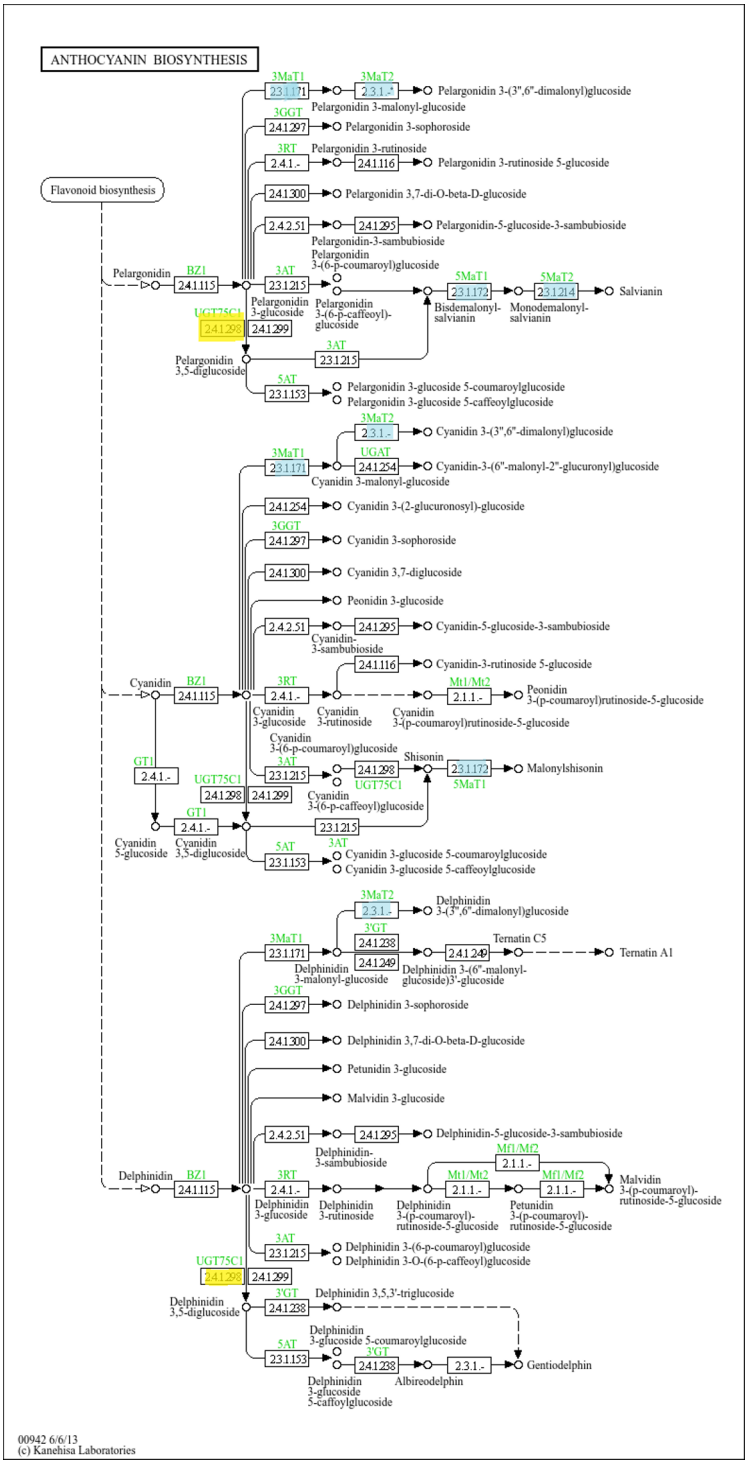


Fig. 5. KEGG pathway of anthocyanin biosynthesis

3.6 Limitations

This study is based on single-accession, sequence-derived computational prediction rather than experimental validation. Physicochemical parameters (ProtParam) and subcellular localization (Plant-mPLoc) are inferred from primary sequence features and have not been confirmed biochemically or through in vivo localization assays; the chloroplast assignment for glycosyl transferase in particular should be treated as a hypothesis for experimental follow-up rather than an established finding. Phylogenetic inference was based on four taxa without an outgroup beyond the genus-internal AA/BB-genome comparison, which constrains the resolution of the evolutionary conclusions drawn. Extending this analysis to additional *Oryza* accessions, incorporating experimental validation of localization and expression, and comparing malonyltransferase evolution across a broader set of plant genera would strengthen confidence in the patterns identified here.

4. Conclusion

This study presents a comprehensive *in silico* characterization of three anthocyanin biosynthetic enzymes—LOD1, glycosyl transferase, and malonyl transferase—across four *Oryza* species. Comparative sequence and phylogenetic analysis identified glycosyl transferase as the most conserved of the three loci, while malonyl transferase showed substantially greater evolutionary divergence. Physicochemical characterization indicated that all three enzymes possess favourable structural properties, including high thermostability and hydrophilic character, consistent with their functional roles in anthocyanin metabolism. Subcellular localization predictions revealed a conserved distribution pattern, with LOD1 and malonyl transferase localized to the cytoplasm and glycosyl transferase associated with the chloroplast. KEGG pathway analysis confirmed the sequential involvement of these enzymes in anthocyanin biosynthesis, with LOD1 driving anthocyanidin formation, glycosyl transferase enhancing pigment stability, and malonyl transferase contributing to further chemical modification. Collectively, these findings underscore the evolutionary conservation and functional significance of the anthocyanin biosynthetic pathway in *Oryza* species and offer a comparative foundation for future work on pigment enhancement, nutritional improvement, and stress-responsive traits in rice.

Acknowledgments. The authors thank the Management and Principal (Dr. Anuradha M) of Padmashree Institute of Management and Sciences, Bangalore, Karnataka, India, for their support of this research.

Disclosure of Interests. The authors declare no competing interests relevant to the content of this article.

References

1. Bao, J. (Ed.). (2019). Rice: Chemistry and Technology (3rd ed.). AACC International Press. https://doi.org/10.1016/C2016-0-04339-3.

2. Deng, G. F., Xu, X. R., Zhang, Y., Li, D., Gan, R. Y., & Li, H. B. (2013). Phenolic compounds and bioactivities of pigmented rice. Critical Reviews in Food Science and Nutrition, 53(3), 296–306. https://doi.org/10.1080/10408398.2010.529624.

3. Winkel-Shirley, B. (2001). Flavonoid biosynthesis: A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology, 126(2), 485–493. https://doi.org/10.1104/pp.126.2.485.

4. Petroni, K., & Tonelli, C. (2011). Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant Science, 181(3), 219–229. https://doi.org/10.1016/j.plantsci.2011.05.009.

5. Yoshida, K., Mori, M., & Kondo, T. (2009). Blue flower color development by anthocyanins: From chemical structure to cell physiology. Natural Product Reports, 26(7), 884–915. https://doi.org/10.1039/B800165K.

6. Jacquemin, J., Bhatia, D., Singh, K., & Wing, R. A. (2013). The International Oryza Map Alignment Project: Development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Current Opinion in Plant Biology, 16(2), 147–156. https://doi.org/10.1016/j.pbi.2013.02.014

7. Stein, J. C., Yu, Y., Copetti, D., Zhang, L., et al. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover, and innovation across the genus Oryza. Nature Genetics, 50(2), 285–296. https://doi.org/10.1038/s41588-018-0040-0

8. Vaughan, D. A., Lu, B. R., & Tomooka, N. (2008). The evolving story of rice evolution. Plant Science, 174, 394–408.

9. Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., & Martens, S. (2012). A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases. DNA Research, 19, 103–114.

10. Suzuki, H., Nakayama, T., & Nishino, T. (2004). Proposed mechanism and functional diversity of plant anthocyanin malonyltransferases. Phytochemistry, 65, 517–524.

11. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), The Proteomics Protocols Handbook (pp. 571–607). Humana Press. https://doi.org/10.1385/1-59259-890-0:571.

12. Guruprasad, K., Reddy, B. V. B., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition. Protein Engineering, 4, 155–161.

13. Khosravi, F., Fard, E. M., Hosseininezhad, M., & Shoorideh, H. (2023). Identification and characterization of inulinases by bioinformatics analysis of bacterial glycoside hydrolases family 32 (GH32). Engineering in Life Sciences, 23(8), e2300003. https://doi.org/10.1002/elsc.202300003

14. Chou, K.-C., & Shen, H.-B. (2010). Plant-mPLoc: A top-down strategy to augment the power for predicting plant protein subcellular localization. PLoS ONE, 5(6), e11335. https://doi.org/10.1371/journal.pone.0011335

15. Bowles, D., Lim, E. K., Poppenberger, B., & Vaistij, F. E. (2006). Glycosyltransferases of lipophilic small molecules. Annual Review of Plant Biology, 57, 567–597.

16. Holton, T. A., & Cornish, E. C. (1995). Genetics and biochemistry of anthocyanin biosynthesis. The Plant Cell, 7, 1071–1083.

17. Wang, X., Trinkl, J., Wulanjati, M., et al. (2025). Inheritance and enzymatic basis of anthocyanin malonylation in Fragaria × ananassa. The Plant Journal, 124(2), e70485. https://doi.org/10.1111/tpj.70485.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

