

Comprehensive in silico analysis of the Δ^1 -Pyrroline-5-Carboxylate Synthase (P5CS) gene elucidates structure regulation and stress responsive roles in *Vigna unguiculata*

Soumyadipa Das¹, Nirjhar Dasgupta^{1#}

¹Department of Biotechnology, Techno India University, EM-4, EM-4/1, EM Block, Sector V, Bidhannagar, Kolkata, West Bengal 700091

Abstract - Proline biosynthesis plays a pivotal role in plant responses to abiotic stresses, particularly drought and salinity. The key enzyme Δ^1 -pyrroline-5-carboxylate synthase (P5CS) is central to this process, catalysing the rate-limiting step in the conversion of glutamate to proline. Despite its significance, limited data are available on the structural, functional, and regulatory features of P5CS in *Vigna unguiculata* (cowpea), an important legume crop valued for its drought tolerance and nutritional content. This study employed a comprehensive *in silico* approach to characterize the P5CS gene in *V. unguiculata*, including sequence analysis, motif prediction, phylogenetic comparison, structural modelling, regulatory element identification, subcellular localization, miRNA interaction, synteny mapping, and functional enrichment. The gene was found to encode a 751-amino-acid protein localized predominantly to the plastid, with conserved catalytic domains relevant to glutamate kinase and P5C synthase activity. Promoter analysis revealed the presence of stress-responsive cis-elements (e.g., ABRE, MYB, DRE), while miRNA analysis suggested cleavage-based regulation by miR856-3p. Comparative synteny analysis using CoGe SynMap confirmed conserved collinearity with orthologous genes in rice, soybean, Arabidopsis, and common bean. The findings provide molecular insight into the adaptive role of P5CS and lay a foundation for its functional exploitation in stress-resilient crop development.

Keywords: P5CS gene, proline biosynthesis, abiotic stress, cowpea, in silico analysis, gene regulation.

Abbreviations:

P5CS: Δ^1 -Pyrroline-5-Carboxylate Synthase
Vu: *Vigna unguiculata*
WRKY: WRKY transcription factor
PPI: Protein-Protein Interaction
GO: Gene Ontology
KEGG: Kyoto Encyclopedia of Genes and Genomes
CDS: Coding DNA Sequence
ORF: Open Reading Frame
MW: Molecular Weight
pI: Isoelectric Point
GRAVY: Grand Average of Hydropathicity
AFLP: Amplified Fragment Length Polymorphism
qRT-PCR: Quantitative Real-Time Polymerase Chain Reaction
BLAST: Basic Local Alignment Search Tool

1. INTRODUCTION

Abiotic stresses such as drought, salinity, extreme temperatures, and nutrient deficiency severely disrupt plant growth, development, and productivity by affecting water balance, ion homeostasis, and metabolic activities (Mittler, 2006; Hasanuzzaman et al., 2013). These stresses lead to oxidative damage, membrane instability, protein denaturation, and impaired photosynthesis, ultimately reducing crop yield and survival (Wang et al., 2003; Zhu, 2002). As climate change intensifies the frequency and severity of such stressors, understanding plant adaptive responses has become crucial for ensuring agricultural sustainability (Boyer, 1982; Suzuki et al., 2014).

Plants frequently encounter abiotic stresses such as drought, salinity, and temperature extremes that adversely affect growth and yield (Zhu, 2002; Dasgupta et al., 2010). One of the well-documented adaptive mechanisms

involves the accumulation of osmo-protectants like proline, which stabilizes proteins, membranes, and cellular structures, and acts as a redox buffer (Szabados & Savouré, 2010). The biosynthesis of proline is primarily governed by the enzyme Δ^1 -pyrroline-5-carboxylate synthase (P5CS), which catalyses the conversion of glutamate to Δ^1 -pyrroline-5-carboxylate (P5C), a precursor of proline (Kishor et al., 2005; Verbruggen & Hermans, 2008). In higher plants, P5CS is encoded by small multigene families, often stress-inducible and functionally specialized (Yoshida et al., 1999).

Vigna unguiculata (cowpea) is a legume crop widely cultivated in semi-arid regions and is known for its resilience to drought (Ehlers & Hall, 1997). However, the molecular mechanisms underlying its stress tolerance, particularly the role of P5CS in stress-induced proline accumulation, remain underexplored (Inal et al., 2024).

The advent of genome sequencing and the availability of transcriptomic and proteomic datasets have enabled functional annotation of stress-related genes through computational methods (Kersey et al., 2018). This study presents a detailed *in silico* characterization of the P5CS gene in *V. unguiculata*, utilizing a variety of bioinformatics tools to investigate its sequence features, structural domains, evolutionary relationships, regulatory control, and interaction with other stress-responsive proteins (Que et al., 2021). The findings contribute to a broader understanding of stress adaptation mechanisms and offer potential genetic targets for crop improvement.

2. MATERIALS AND METHODS

The Δ^1 -pyrroline-5-carboxylate synthase (P5CS) gene sequence from *V. unguiculata* was obtained from the Phytozome database (Goodstein et al., 2012). The genomic locus and coding sequence were retrieved and translated to obtain the amino acid sequence. Homology searches were performed using BLASTP (Altschul et al., 1990) against the NCBI non-redundant protein database to confirm gene identity and find related P5CS proteins in other species. Top BLAST hits were retained based on E-value cutoff ($1e-5$) and alignment score, ensuring that the cowpea P5CS sequence corresponded to known P5CS homologs.

InterProScan (Jones et al., 2014) was used to annotate conserved domains and functional sites in the P5CS protein. Amino acid sequence analysis was performed with ExPASy ProtParam (Gasteiger et al., 2005) to compute physicochemical parameters such as molecular weight, isoelectric point, and stability index. Gene structure (exon–intron organization) was visualized by uploading the P5CS gene model to the GSDS 2.0 server (Hu et al., 2015), which graphically mapped CDS regions onto the genomic sequence. Conserved motifs within the P5CS protein (and its homologs) were identified using the MEME Suite (Bailey et al., 2015). MEME was run with default settings (maximum motif width, and a specified number of motifs), and the resulting motifs were compared to known protein signatures.

A phylogenetic analysis was conducted to compare P5CS proteins from cowpea and other legumes. Protein sequences from *V. unguiculata* and selected species were aligned using Clustal Omega (Sievers et al., 2011). The multiple sequence alignment was then used to construct a phylogenetic tree in MEGA11 (Tamura et al., 2021) by the Maximum Likelihood method with 1,000 bootstrap replicates. Tree topology and bootstrap support values were examined to infer evolutionary relationships of the P5CS gene family.

Tertiary structure prediction of the cowpea P5CS protein was performed with the Phyre2 web portal (Kelley et al., 2015). Phyre2 builds homology-based models by threading the query sequence onto known structures. In parallel, the AlphaFold algorithm (Jumper et al., 2021) was used to generate an ab initio structural model. Predicted models were evaluated by confidence scores (e.g. Phyre2 confidence, AlphaFold pLDDT) and visualized in PyMOL to inspect active sites and binding pockets.

3. RESULTS

3.1 Retrieval and Annotation of the P5CS Gene

The cowpea P5CS coding sequence (Vigun10g052600.1) was retrieved from Phytozome (version 1.1) and found to span 2,256 bp, encoding a 751-residue protein. Domain annotation (InterProScan) identified three conserved regions characteristic of P5CS enzymes: a glutamate 5-kinase domain (IPR005715), a P5C synthase domain (IPR005766), and a glutamate/acetylglutamate kinase domain (IPR001057), confirming the bifunctional nature of the protein (Jones et al., 2014).

3.2 Physicochemical Properties of the Protein

ProtParam (Gasteiger et al., 2005) analysis yielded the following properties for the P5CS protein (Table I). It has a molecular weight of 81.3 kDa, a theoretical pI of 6.49, and a stability index of 34.3, classifying it as stable. The

high aliphatic index (106.1) suggests thermostability, and a GRAVY score of -0.035 indicates slight hydrophilicity.

These parameters indicated a stable, cytosolic enzyme with good thermal tolerance, consistent with a role in stress-response metabolism.

Feature	Value	Interpretation
Length	751 amino acids	Full-length P5CS protein
Molecular Weight (MW)	81.3 kDa	Typical size for a bifunctional enzyme like P5CS
Theoretical pI	6.49	Slightly acidic isoelectric point
Charged residues	Negative: 89, Positive: 84	Near balance — stable in physiological pH
Instability Index	34.30	Stable protein (<40 = stable)
Aliphatic Index	106.10	High → thermostable, good structural rigidity
GRAVY	-0.035	Slightly hydrophilic, expected for cytoplasmic enzymes
Half-life	>30h (mammals), >10h (E. coli)	Reasonable expression stability
Extinction Coefficient	~37,360 M ⁻¹ cm ⁻¹	Useful for UV-based protein quantification
Amino Acid Richness	High Leu, Ala, Gly	Typical of metabolically active proteins

Table I. Physicochemical parameters of *V. unguiculata* P5CS protein (ProtParam).

3.3 Conserved Motif and Gene Structure Analysis

Conserved motif analysis (MEME Suite; Bailey et al., 2009) identified 10 motifs of 6–10 amino acids in the P5CS protein (Fig. 2). Motif 1 (6 aa) had the lowest E-value ($\sim 1.1 \times 10^{-1}$), suggesting it represents a critical catalytic core, while motifs 2–10 exhibited E-values from $\sim 1.6 \times 10^{-1}$ to 8.9×10^{-2} (Table II). Gene structure mapping (GSDS) showed that the cowpea P5CS gene contains 20 exons (Fig. 2), mirroring the 20-exon organization observed in *Arabidopsis thaliana* P5CS (Araport11). This exon–intron architecture is consistent with other plant P5CS orthologs, indicating a conserved gene organization.

Motif	Width	E-value	Biological Interpretation
Motif 1	6 aa	1.1e+001	Possibly catalytic core region
Motif 2–10	6–10 aa	1.6e+001 to 8.9e+002	May correspond to binding sites, structural regions, or flexible loops

Table II. Conserved motifs of P5CS identified by MEME (motif width and E-value).

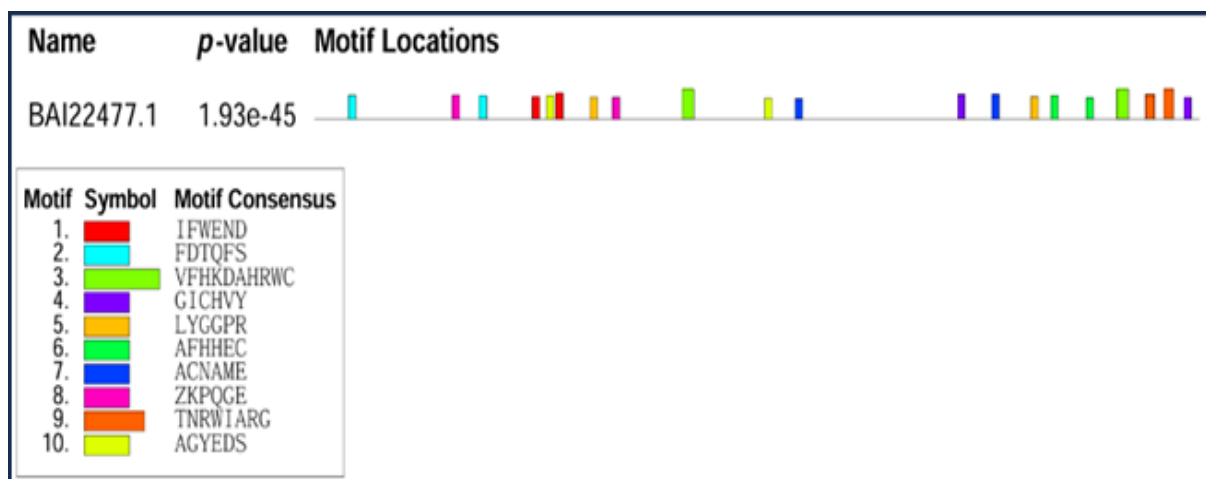


Fig. 2. Conserved motif map and exon–intron structure of cowpea P5CS.

3.4 Phylogenetic and Similarity Analysis

Pairwise sequence comparison revealed high amino acid identity between cowpea P5CS and orthologs in related species: 85.6% identity with *Phaseolus vulgaris*, 85.5% with *Glycine max* (soybean), and 77.0% with *Oryza sativa* (rice). An unrooted phylogenetic tree (Fig. 1) placed *V. unguiculata* P5CS in a clade with other legume P5CS sequences, reflecting the close relationship among dicot homologs (Tamura et al., 2021). This clustering supports evolutionary conservation of the P5CS gene within dicots and highlights divergence from monocots.

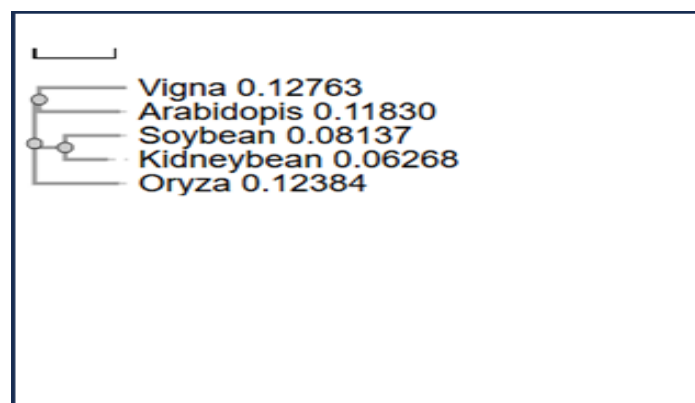


Fig. 1. Phylogenetic tree of P5CS proteins from cowpea and related plants (NJ method).

3.5 3D Structural Modelling

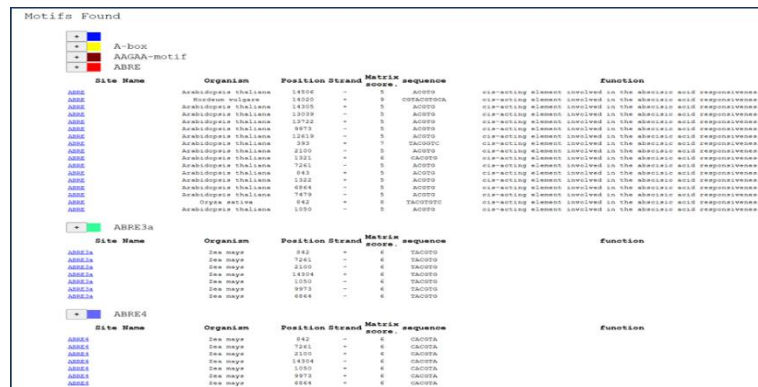
Tertiary structures of the P5CS protein were predicted by both Phyre2 (Kelley et al., 2015) and AlphaFold2 (Jumper et al., 2021). The AlphaFold2 model had a high confidence (predicted LDDT = 87.4). Structural superposition of the Phyre2 and AlphaFold2 models in PyMOL produced a root-mean-square deviation (RMSD) of 0.747 Å, indicating strong concordance between models. Both models exhibited the expected fold of the bifunctional enzyme, with stacked α -helices and β -sheets forming the ATP- and glutamate-binding domains.

3.6 Promoter and Cis-Regulatory Element Analysis

The 2,000 bp upstream promoter was scanned using PlantCARE (Lescot et al., 2002), revealing numerous abiotic stress-responsive elements:

- **ABRE** – Absciscic acid-responsive element
- **MYB/MYC** sites – Drought and salt stress regulatory elements

- These cis-elements (Fig. 3) indicate that P5CS transcription is potentially regulated by drought, salinity, cold, and ABA signaling (Lescot et al., 2002). Their presence suggests a role for P5CS in the transcriptional stress response of cowpea.



3.7 Subcellular Localization Prediction

No.	Protein ID	Localization	Distance	Identity (%)	Description
1	At2g36910.1	Plastid	230.41	14.54	Arabidopsis [Arath]
2	At2g38170.1	Vacuole	234.06	15.85	Arabidopsis [Arath]
3	PMAA_ARATH	Plastid	234.48	14.77	Integral membrane protein
4	GR11_ARATH	Plastid	256.53	13.37	Integral membrane protein
5	ACAC_ARATH	Plastid	260.03	14.71	Integral membrane protein
6	AHM4_ARATH	Plastid	264.61	12.78	Integral membrane protein
7	PMA5_ARATH	Plastid	282.67	13.61	Integral membrane protein
8	ATC1_DUNBI	Plastid	287.64	15.53	Integral membrane protein
9	HMD2_CAPAN	E.R.	290.11	14.23	ER membrane-anchored
10	NCPR_PHA AU	E.R.	291.18	13.72	ER-anchored
11	PMA1_LYCES	Plastid	292.23	14.44	Plasma membrane protein
12	GIGA_ARATH	Nucleus	292.26	14.92	Nuclear protein
13	PMAB_ARATH	Plastid	293.51	15.57	Integral membrane protein

No.	Protein ID	Localization	Distance	Identity (%)	Description
14	PT2B_ARATH	Plastid	296.20	13.45	Integral membrane protein

Table III. Wolf PSORT output, which explain P5CS protein is predicted to be localized in the plastid (chloroplast).

3.8 miRNA Target Prediction

Using psRNATarget (Dai et al., 2018), one cowpea miRNA was predicted to target the P5CS mRNA. The candidate *aly-miR856-3p* (from *Arabidopsis lyrata*) was found with an expectation score of 3.0, suggesting moderate confidence in a cleavage-mode interaction on the P5CS coding sequence. This implies possible post-transcriptional regulation of P5CS by miRNA under stress conditions (Table IV).

Field	Value
miRNA	aly-miR856-3p
Target Gene	Vigun10g052600.1 (P5CS CDS)
Expectation	3.0 (<i>threshold cutoff, borderline match</i>)
Inhibition Type	Cleavage
Alignment	Imperfect pairing with typical miRNA pattern
Multiplicity	1 (only one predicted target site)

Table IV. Predicted miRNA target for cowpea P5CS (psRNATarget output).

3.9 Synteny and Comparative Genomics

Comparative synteny analysis (CoGe SynMap, Lyons et al., 2008) identified conserved collinear genomic regions surrounding P5CS across diverse plants. The cowpea P5CS locus aligned with orthologous regions in *Oryza sativa* (Fig. 4). GEvo visualization confirmed that local gene order and content are conserved in these blocks, indicating evolutionary preservation of the P5CS chromosomal context and supporting its functional importance.

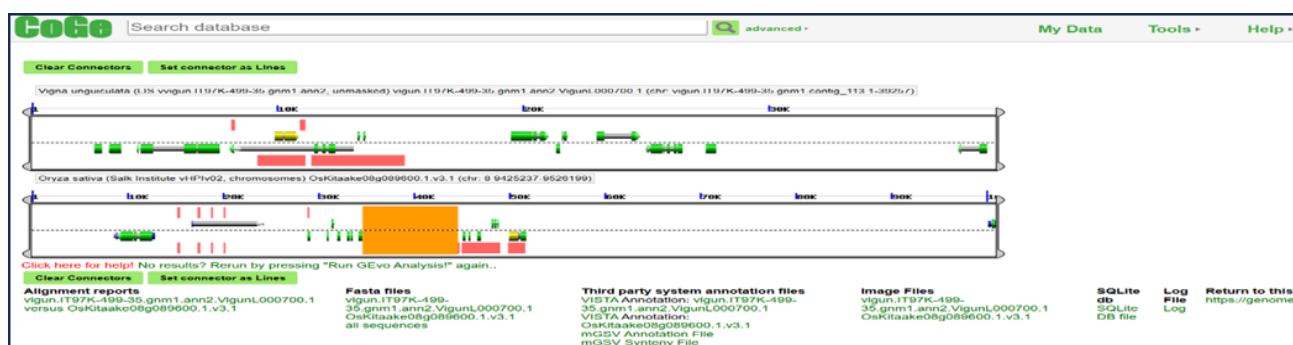


Fig. 4. The connected red lines or boxes show where genes (or non-coding regions) from cowpea are conserved or matched in rice. The orange rectangle shows a high-confidence syntenic region between the two species. This means the P5CS gene or its neighbouring genes in cowpea are conserved in rice.

3.10 Protein-Protein Interaction and Functional Enrichment

A STRING v11.5 network (Szklarczyk et al., 2021) was constructed for the Arabidopsis P5CS ortholog (AT2G39800) to infer P5CS interactions. The network revealed links to enzymes in proline and glutamate metabolism:

- **Proline biosynthesis** (e.g., ALDH12A1, PROT3)
- **Glutamate metabolism** (e.g., glutamate decarboxylases (GADs), glutamate dehydrogenases (GDHs), GLT1)
- **Nitrogen assimilation/redox regulation** (e.g., GS, GOGAT)

KEGG pathway enrichment of these interactors highlighted strong significance for amino acid pathways. Notably, *glutamate metabolism* was enriched ($FDR = 1.41 \times 10^{-22}$), along with arginine/proline metabolism, nitrogen metabolism, and the urea cycle (Table V). These results underscore the integration of P5CS into core metabolic and stress-response networks (Kanehisa & Goto, 2000; Szklarczyk et al., 2021).

Pathway	Description	Significance (FDR)	Relevance
ath00250	Alanine, aspartate, and glutamate metabolism	1.41e-22	Glutamate is the direct precursor of proline (P5CS substrate).
ath00330	Arginine and proline metabolism	8.54e-11	Core pathway for P5CS—this confirms its role in proline biosynthesis.
ath00220	Arginine biosynthesis	1.40e-11	Tied to nitrogen utilization; overlaps with proline pathway.
ath00910	Nitrogen metabolism	2.46e-11	Reflects involvement in amino acid balance.
ath01230	Biosynthesis of amino acids	3.24e-07	Confirms its role in primary metabolism.

Table V. KEGG pathway enrichment for P5CS-associated proteins (FDR values).

4. DISCUSSION

Sequence analysis confirmed that the cowpea P5CS gene encoded a 751-amino-acid enzyme (~81.3 kDa) with a near-neutral isoelectric point and high aliphatic index, consistent with a stable, soluble protein. Domain annotation identified the canonical bifunctional P5CS architecture – a γ -glutamyl kinase domain fused to a Δ^1 -pyrroline-5-carboxylate synthetase domain – which is characteristic of plant P5CS enzymes (Verbruggen and Hermans, 2008). Ten conserved motifs were detected, several of which aligned with these catalytic core regions, indicating that structural constraints have preserved essential active-site features. These structural insights support the expected dual role of *V. unguiculata* P5CS in phosphorylating glutamate and reducing γ -glutamyl phosphate during proline biosynthesis.

Phylogenetic analysis showed that cowpea P5CS clustered closely with homologs from other legumes, sharing ~85% identity with *Phaseolus* and *Glycine* P5CS, and ~77% identity with the rice ortholog. Comparative synteny mapping revealed conserved colinear blocks around the P5CS locus in soybean, *Arabidopsis*, rice, and common bean genomes, indicating that the gene's genomic context has been maintained over evolution. This strong conservation across both dicots and monocots implies that P5CS has been under purifying selection, reflecting its essential function in stress adaptation.

Structural modeling yielded highly similar tertiary structures from Phyre2 and AlphaFold2 (RMSD ~0.75 Å); the AlphaFold2 model had a high predicted LDDT (87.4), indicating reliable folding. The model displayed the expected α/β folds around the ATP- and substrate-binding sites, supporting catalytic competence. Subcellular localization prediction indicated plastid (chloroplast) targeting, consistent with known plastidic proline biosynthesis in plants. Promoter analysis identified multiple abiotic stress-responsive cis-elements (e.g. ABRE, MYB/MYC, DRE and LTR motifs) upstream of the gene, suggesting transcriptional activation by drought, salinity, and cold signals. These regulatory motifs (Nakashima *et al.*, 2009) are consistent with the observed induction of P5CS under stress and underscore proline's role as an osmoprotectant and redox buffer in stress tolerance (Szabados and Savouré, 2010).

A putative microRNA regulator (aly-miR856-3p) was predicted to target the P5CS mRNA, indicating an additional layer of post-transcriptional control. In plants, such miRNA-guided cleavage of stress-related transcripts is well documented (Jones-Rhoades *et al.*, 2006), and the identified target site suggests that cowpea P5CS expression could be fine-tuned by RNA silencing during stress responses.

Interaction network analysis placed P5CS at the center of a metabolic cluster. In the Arabidopsis ortholog network, P5CS interacted with enzymes of glutamate metabolism (glutamate decarboxylases, glutamate synthase) and proline biosynthesis (aldehyde dehydrogenases, PROT3). KEGG enrichment highlighted glutamate metabolism, arginine/proline metabolism, and nitrogen assimilation pathways with very low FDRs. These links imply that P5CS channels glutamate-derived nitrogen toward proline under stress, consistent with proline's known involvement in nitrogen homeostasis during osmotic stress (Szabados and Saviouré, 2010). Overall, the network and pathway results reinforce the concept that P5CS acts as a metabolic hub, integrating amino acid and nitrogen metabolism with osmotic stress response.

Together, these findings underscore P5CS as a promising target for enhancing stress tolerance in cowpea. Proline accumulation is correlated with drought and salinity resilience, so increasing P5CS activity could improve osmotic adjustment and redox buffering. Indeed, overexpression of P5CS from other species has elevated proline levels and stress tolerance in transgenic plants. In legumes, breeding or biotechnological approaches to up-regulate P5CS (e.g. marker-assisted selection or genome editing) may thus be effective. Notably, the successful application of CRISPR/Cas9 in cowpea (Ji *et al.*, 2019) demonstrates that targeted editing of the P5CS gene is feasible, paving the way for developing stress-resilient cultivars by precise gene modification.

Future work should validate these computational predictions. Quantitative expression profiling of cowpea P5CS under drought and salt treatments would confirm stress responsiveness. Functional studies – for example, generating overexpression lines, RNAi knockdowns, or CRISPR knockouts – are needed to test the gene's role in vivo. Transcriptomic and metabolomic analyses of such genotypes would clarify downstream effects and pathway integration. By combining in silico insights with wet-lab experimentation, researchers can fully elucidate P5CS's function in cowpea stress adaptation and accelerate its deployment in crop improvement programs.

5. CONCLUSION

In this study, comprehensive in silico analysis reaffirmed that the *Vigna unguiculata* P5CS gene encodes the bifunctional Δ^1 -pyrroline-5-carboxylate synthase responsible for the rate-limiting step of proline biosynthesis (Kishor *et al.*, 2005). This enzyme's central role in converting glutamate to proline underscores its importance in osmoprotection and ROS scavenging during abiotic stress (Verbruggen and Hermans, 2008; Szabados and Saviouré, 2010). Sequence analysis confirmed the conserved glutamate kinase and glutamylphosphate reductase domains characteristic of P5CS proteins. Promoter examination revealed multiple stress-responsive cis-elements (e.g. ABRE, DRE, MYB/MYC motifs) indicative of regulation by drought, salinity and ABA signaling (Verbruggen and Hermans, 2008). Subcellular localization predictors placed the cowpea P5CS protein predominantly in chloroplasts, consistent with experimental evidence that plant P5CS isoforms localize to plastids under stress (Székely *et al.*, 2008). Together, these structural and regulatory features established that VuP5CS was poised for strong induction under environmental stress.

Protein–protein interaction and pathway enrichment analyses further supported VuP5CS's integration into stress-response networks. The STRING-based interaction map connected P5CS with key enzymes of glutamate, proline and nitrogen metabolism, forming a tightly linked amino acid–metabolism hub. Enrichment tests identified highly significant associations of P5CS with glutamate and arginine/proline metabolic pathways (FDR < 1e-10). These systems-level findings corroborated that P5CS channelled nitrogen flux toward proline synthesis during stress, in agreement with its role in coordinating nitrogen balance and redox homeostasis (Szabados and Saviouré, 2010; Szklarczyk *et al.*, 2023). In addition, a putative stress-induced miRNA (miR856) target site on P5CS suggested post-transcriptional control under abiotic challenge. Importantly, the in-silico results aligned with empirical evidence that elevating P5CS activity enhances stress tolerance: for example, constitutive overexpression of P5CS in cowpea and other crops significantly improved drought resilience and root growth under water deficit (Okeyo-Ikawa *et al.*, 2016; Kishor *et al.*, 2005). In summary, this integrative analysis confirmed that cowpea P5CS possesses the structural, regulatory and interaction characteristics of a key stress-responsive gene. These insights underscore the strong potential of VuP5CS as a candidate for genetic engineering or marker-assisted breeding to develop stress-resilient cowpea cultivars.

6. SUPPLEMENTARY MATERIALS

Gene Report:

Vigun10g052600

Genome: [Vigna unguiculata v1.1](#) (Phytozome genome ID: 469 • NCBI taxonomy ID: [3917](#))

Gene Identifier: Vigun10g052600 (PAC:39033479)

Transcript: Vigun10g052600.1

Location: Vu10:8458269..8469678 reverse

Auto Defline: (1 of 2) 1.2.1.41//2.7.2.11 - Glutamate-5-semialdehyde dehydrogenase / Glutamyl-gamma-semialdehyde dehydrogenase // Glutamate 5-kinase / Gamma-glutamyl kinase

PFAM: [PF00696](#) AA_kinase

PFAM: [PF00171](#) Aldedh [Q9AYM4](#) VIGUN (Q9AYM4)

Description: Delta-1-pyrroline-5-carboxylate synthase [Includes: Glutamate 5-kinase (EC 2.7.2.11) (GK) (Gamma-glutamyl kinase); Gamma-glutamyl phosphate reductase (EC 1.2.1.41) (GPR) (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase)]

Organism: *Vigna unguiculata* (Cowpea)

Bioinformatic tools & links

Tool Name	Purpose	Link
Phytozome	Gene sequence retrieval	https://phytozome-next.jgi.doe.gov/
ProtParam (ExPASy)	Physicochemical properties of protein	https://web.expasy.org/protparam/
InterProScan	Protein domain and functional annotation	https://www.ebi.ac.uk/interpro/search/sequence/
MEME Suite	Conserved motif discovery	https://meme-suite.org/meme/tools/meme
GSDS (Gene Structure Display Server)	Exon-intron structure visualization	http://gsds.gao-lab.org/
Phyre2	3D structure prediction	http://www.sbg.bio.ic.ac.uk/phyre2/
AlphaFold Protein Structure Database	High-accuracy protein structure prediction	https://alphafold.ebi.ac.uk/
PlantCARE	Cis-regulatory element analysis	http://bioinformatics.psb.ugent.be/webtools/plantcare/html/
WoLF PSORT	Subcellular localization prediction	https://wolfsort.hgc.jp/
psRNATarget	miRNA target prediction	https://www.zhaolab.org/psRNATarget/
CoGe SynMap & GEvo	Synteny and collinearity analysis	https://genomevolution.org/coge/SynMap.pl
STRING	Protein-protein interaction (PPI) network	https://string-db.org/
KEGG Enrichment Tools	Pathway enrichment analysis (via STRING or DAVID)	https://www.genome.jp/kegg/

3. RESULTS

3.1 Retrieval and annotation of the P5CS Gene

Blastp output Sequence:

```
>BAI22477.1 delta 1-pyrroline-5-carboxylate synthase [Vigna unguiculata]
MELLQNGTKAKPSEIPLVNGTALTHLNSLFQTQFSGNIDPSRAVSKVKRLIVKVGTA VVTRSDGRLALG
RIGALCEQLKELSSQGFEVILVTSGAVGLGRQRLRYRKLANSFSDLQKPQGELDGKACA AVGQSSLMA
LYDTMFSQLDVTSSQLLVNDGFFRDAGFRKQLSDTVNALLDLRVIPIFNENDAVSTRKAPYEDSSGIFWD
NDSLAGLLALELKADLLVLLSDVEGLYSGPPSDPN SKLIHTYVKEKHQGEITFGDKSRLGRGGM TAKVN
AAVCAAHAGIPVIITSGYATNNIIRVLQGERIGTVFHKDAHLWSNIKEESAREMAVSARESSRLQILKSE
ERSKILLAIADALEKNASAI SLENEADVAVAVLAGYENSLISRLTLKHEKISKLAKSVRMLAAMEEPIGQI
LKRTEIADKLILEKISCPLGVLLVIFESRPDALVQIAALAIRSGNGLLLKGGKEAQRSNAILHKVITSVIPNT
VGDKLIGLVNSRDEIPYLLKLDDVIDLVVPRGSNKLVSQIKESTKIPVLGHADGICHVYVDKSAKIDMAK
QIIRDAKTDYPACNAMETLLVHKDLSNNGGLNELVVELQREGVKLYGGPRASGILNIVETSAFHHEYS
SLACTVEIVEDVFDAINHIHEHGS AHTECIVTEDREVAETFLSQVDSA AVFHNASTRFCDGARFGLGA EV
GISTSRIHARGPVGVEGLLTNRWILRGSGHV VNGDQGINYTYKELPLEA
```

InterPro Scan output:

Sequence Length: [751 amino acids](#); Protein family membership: [Glutamate/acetylglutamate kinase](#) (IPR001057)
[Glutamate 5-kinase/delta-1-pyrroline-5-carboxylate synthase](#) (IPR005715)
[Delta 1-pyrroline-5-carboxylate synthetase](#) (IPR005766).

Domain Name	InterPro ID	Meaning
Glutamate/acetylglutamate kinase	IPR001057	Matches PF01180 → Glutamate 5-kinase → first domain of P5CS
Glutamate 5-kinase/delta-1-pyrroline-5-carboxylate synthase	IPR005715	Combined domain annotation related to P5CS multi-domain enzymes
Delta-1-pyrroline-5-carboxylate synthetase	IPR005766	General family classification → confirms P5CS family membership

Table I. This InterPro Scan results clarifies the presence of a full, functional P5CS homolog in Vigna Unguiculata.

3.3 Conserved motif and gene structure analysis

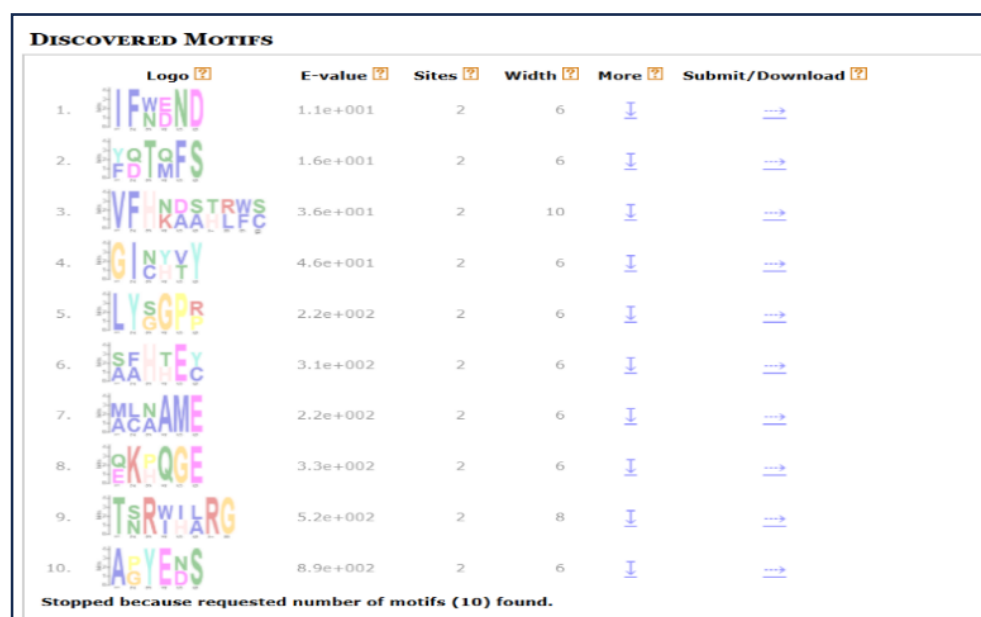


Fig 01:

Conserved motifs in *Vigna unguiculata* P5CS protein identified by MEME Suite.

Table II. Amino Acid Composition in Cowpea P5CS Protein (751 aa)

Amino Acid	Code	Frequency
Alanine	A	0.0919
Cysteine	C	0.0120
Aspartic acid	D	0.0533
Glutamic acid	E	0.0652
Phenylalanine	F	0.0266
Glycine	G	0.0786
Histidine	H	0.0266
Isoleucine	I	0.0706
Lysine	K	0.0586
Leucine	L	0.1190
Methionine	M	0.0120
Asparagine	N	0.0439
Proline	P	0.0280
Glutamine	Q	0.0266
Arginine	R	0.0533
Serine	S	0.0866
Threonine	T	0.0493
Valine	V	0.0799
Tryptophan	W	0.00399
Tyrosine	Y	0.0186

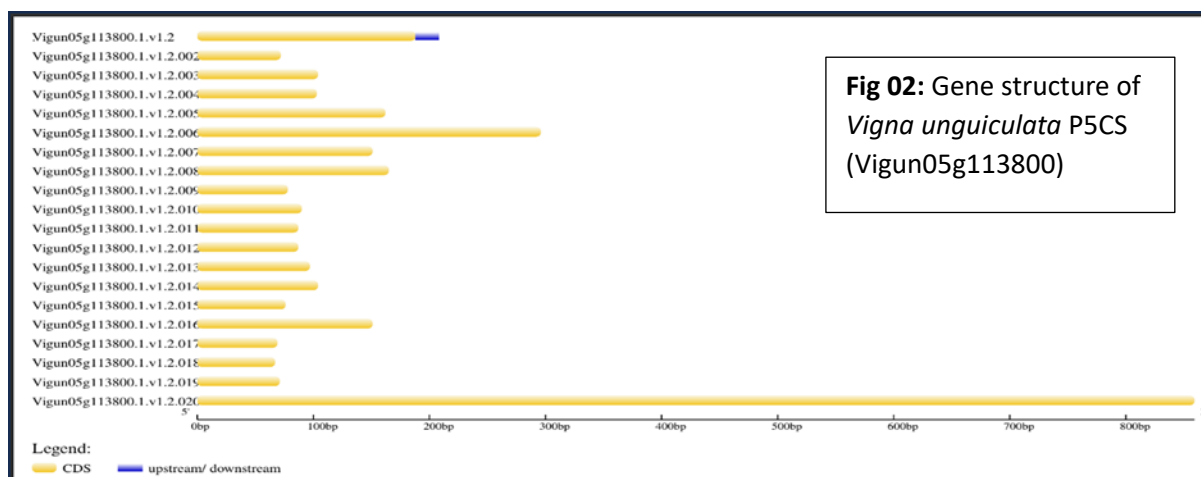


Fig 02: Gene structure of *Vigna unguiculata* P5CS (Vigun05g113800)

Key features of GSDS2.0 output:

- **20 exons** are consistent with known P5CS genes in other species (e.g., *Arabidopsis* P5CS1 also has 20 exons).
- This confirms that **Vigun05g113800** is a **full-length, properly structured P5CS gene** in cowpea.
- The **intron-exon structure** is typical of stress-responsive metabolic enzymes, which often have multiple regulatory segments.

3.4 Phylogenetic and similarity analysis

A **Percent Identity Matrix (PIM)** showing how similar the **P5CS protein sequences** are between five plant species.

Table III. PIM of P5CS gene in different organisms

	Vigna (Cowpea)	Arabidopsis	Oryza (Rice)	Soybean	Kidney Bean
Vigna	100.00%	75.41%	74.58%	73.71%	75.28%
Arabidopsis	75.41%	100.00%	74.71%	75.33%	77.12%
Oryza	74.58%	74.71%	100.00%	74.83%	77.06%
Soybean	73.71%	75.33%	74.83%	100.00%	85.59%
Kidney Bean	75.28%	77.12%	77.06%	85.59%	100.00%

Table IV. Branch lengths (evolutionary distance)

Species Pair	Distance	Meaning
Vigna ↔ Arabidopsis	0.12763 vs 0.11830	Moderately close
Soybean ↔ Kidney bean	0.08137 vs 0.06268	Very close (short branches)
Oryza	0.12384	Intermediate, more distant to all

The shorter the branch, the closer the evolutionary relationship.

3.5 3D structural modelling

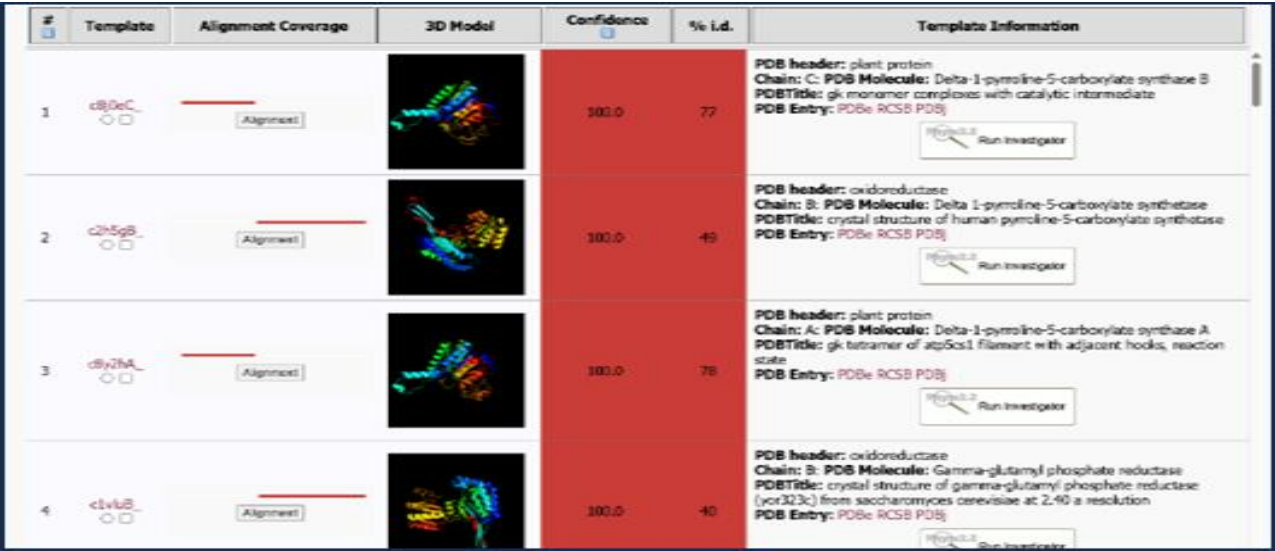


Fig 03: P5CS Protein 3D Structure Prediction template via Phyre2.2

Rank	Template	% Identity	Confidence	Function
1	c8j0eC_	77%	100%	Plant P5CS B
2	c2h5gB_	49%	100%	Human P5CS
3	c8y2hA_	78%	100%	P5CS A (plant)
4	c1vluB_	40%	100%	Gamma-glutamyl phosphate reductase

Table V. All top hits are P5CS or related proline biosynthetic enzymes.

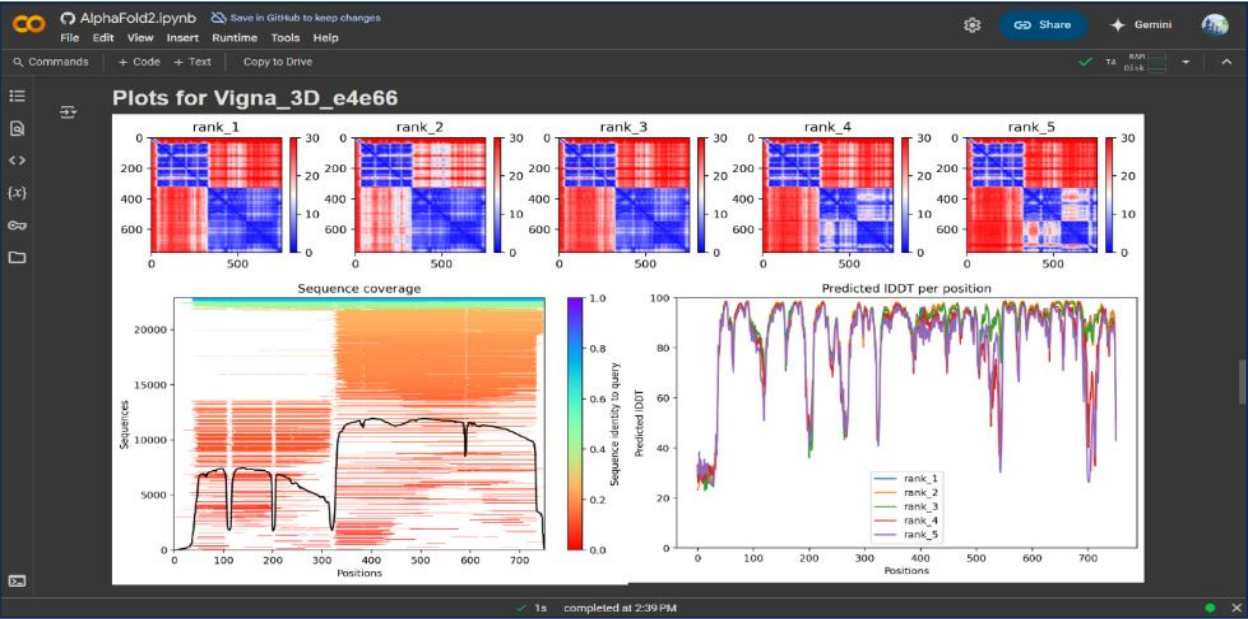


Fig 04: AlphaFold2 prediction using ColabFold; full-length 3D structure of cowpea P5CS protein.

Rank	Model ID	pLDDT (Confidence)	pTM (Global Fold Accuracy)
1	rank_001_alphafold2_ptm_model_3_seed_000	87.4	0.616
2	rank_002_alphafold2_ptm_model_5_seed_000	87.3	0.667
3	rank_003_alphafold2_ptm_model_4_seed_000	86.8	0.603
4	rank_004_alphafold2_ptm_model_2_seed_000	83.2	0.574
5	rank_005_alphafold2_ptm_model_1_seed_000	82.2	0.529

Table VI. AlphaFold2 results.



Fig 05: The core catalytic domain of P5CS predicted by Phyre2.2 is very well aligned with the same region in the AlphaFold model. (RMSD Value: 0.747)

3.7 Subcellular localization prediction

Fig 06: The top 14 sequence-based neighbours of cowpea P5CS show predominant plastid localization, with over 70% matching plastid-associated proteins.

3.9 Synteny and Comparative Genomics

Page 15

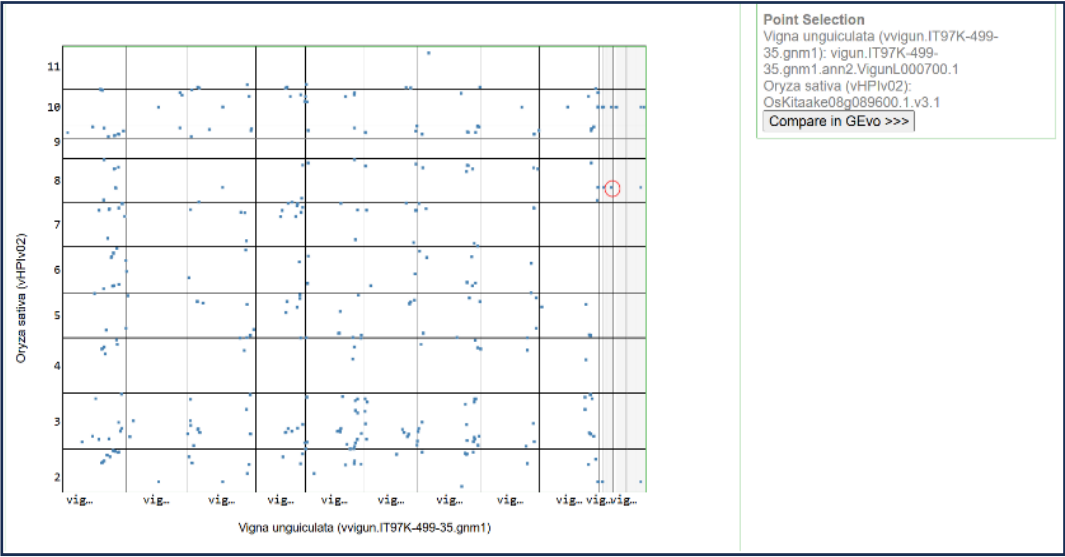


Fig 07: A synteny analysis between *Vigna unguiculata* and *Oryza sativa* using CoGe SynMap

3.10 Protein-Protein Interaction and Functional Enrichment

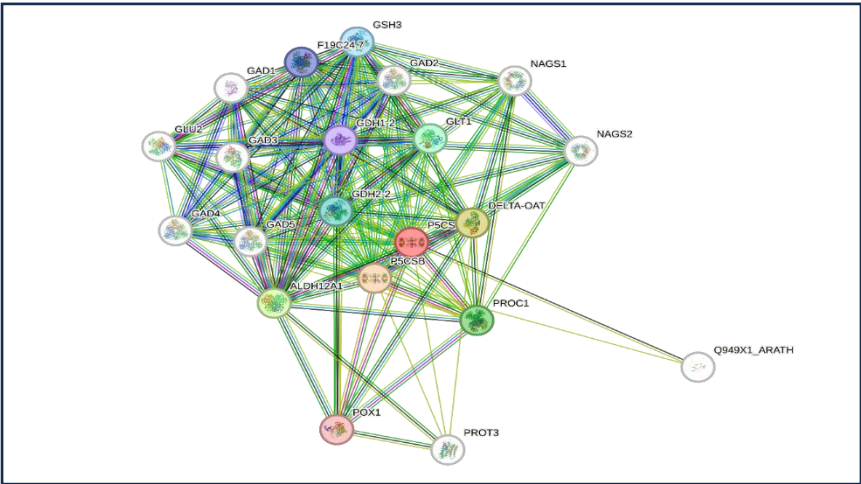


Fig 08: This network shows that P5CS is highly integrated into the glutamate–proline metabolism cycle, a key pathway in: Osmotic stress tolerance (drought, salt), Reactive oxygen species (ROS) regulation, Energy balance during stress. P5CS works upstream of proline synthesis, and the presence of ALDH12A1 and PROT3 shows a complete stress-tolerant proline biosynthesis module is being regulated.

Protein	Full Name / Function	Role in Metabolism / Stress
P5CS	Δ1-Pyrroline-5-Carboxylate Synthase	Central enzyme in proline biosynthesis, stress response regulator
P5CSB	P5CS isoform	Redundant or complementary to P5CS
ALDH12A1	Aldehyde Dehydrogenase	Converts P5C to proline; downstream of P5CS
POX1	Proline oxidase	Proline degradation (reverse pathway), modulated under stress

Protein	Full Name / Function	Role in Metabolism / Stress
PROT3	Proline transporter	Moves proline across membranes; key in osmotic adjustment
GAD family (GAD1–GAD5)	Glutamate decarboxylase	Competes for glutamate (P5CS substrate)
GDH1/2	Glutamate dehydrogenase	Interconversion between glutamate and α -KG
GLT1	Glutamate synthase	Regenerates glutamate
NAGS1/2	N-acetylglutamate synthase	Glutamate-linked metabolism
PROC1	Proline-rich protein	Often upregulated in stress conditions

Table VIII. Interacting proteins which shows P5CS is highly integrated into the glutamate–proline metabolism cycle, a key pathway in: Osmotic stress tolerance (drought, salt), Reactive oxygen species (ROS) regulation and Energy balance during stress.

7. DECLARATIONS

Ethics approval

Not applicable. This study does not involve human participants, human data, or animal subjects.

Consent to participate and Consent to publish

Not applicable. This manuscript does not contain data from any individual person.

Availability of data and materials

The physicochemical parameters of P5CS protein are available in the Supplementary Materials for Paper.

Conserved motifs of P5CS identified by MEME can be found in the Supplementary Materials for Paper.

The phylogenetic tree of P5CS proteins is provided in the Supplementary Materials for Paper.

3D structural models of P5CS protein are included in the Supplementary Materials for Paper.

Cis-regulatory element distribution in P5CS promoter is detailed in the Supplementary Materials for Paper.

miRNA target prediction for cowpea P5CS is available in the Supplementary Materials for Paper.

Synten analysis of P5CS gene is presented in the Supplementary Materials for Paper.

Protein-protein interaction network for P5CS is available in the Supplementary Materials for Paper.

Reused Data

The P5CS gene sequence from *Vigna unguiculata* is accessible at the URL of dataset [<https://phytozome.jgi.doe.gov/pz/portal.html>].

Clinical Trial Details

This present work contains no clinical trials.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Authors' contributions

SD collected the data and prepared the manuscript. NDG planned the work, provided critical revisions. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge the Chancellor and Vice -Chancellor of Techno India University, West Bengal for providing continuous support to carry out the work.

Conflict of Interests

The authors declare that they have no conflict of Interests.

8. REFERENCES

- [1] P. B. K. Kishor *et al.*, "Regulation of proline biosynthesis, degradation, uptake and transport in higher plants: Its implications in plant growth and abiotic stress tolerance," *Curr. Sci.*, vol. 88, no. 3, pp. 424–438, 2005.
- [2] N. Verbruggen and C. Hermans, "Proline accumulation in plants: a review," *Amino Acids*, vol. 35, no. 4, pp. 753–759, 2008.
- [3] L. Szabados and A. Savouré, "Proline: a multifunctional amino acid," *Trends Plant Sci.*, vol. 15, no. 2, pp. 89–97, 2010.
- [4] G. Székely *et al.*, "Duplicated P5CS genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis," *Plant J.*, vol. 53, no. 1, pp. 11–28, 2008.
- [5] R. Okeyo-Ikawa *et al.*, "In planta seed transformation of Kenyan cowpeas (*Vigna unguiculata*) with P5CS gene via *Agrobacterium tumefaciens*," *J. Agric. Biotechnol. Sustainable Dev.*, vol. 8, no. 5, pp. 32–45, 2016.
- [6] D. Szklarczyk *et al.*, "The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest," *Nucleic Acids Res.*, vol. 51, no. D1, pp. D638–D646, 2023.
- [7] K. Nakashima, Y. Ito, and K. Yamaguchi-Shinozaki, "Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses," *Plant Physiol.*, vol. 149, no. 1, pp. 88–95, 2009.
- [8] M. Jones-Rhoades, D. Bartel, and B. Bartel, "MicroRNAs and their regulatory roles in plants," *Annu. Rev. Plant Biol.*, vol. 57, pp. 19–53, 2006.
- [9] J. Ji *et al.*, "Genome editing in cowpea (*Vigna unguiculata*) using CRISPR-Cas9," *Genes*, vol. 10, no. 3, art. 131, 2019.
- [10] P. Jones *et al.*, "InterProScan 5: genome-scale protein function classification," *Bioinformatics*, vol. 30, no. 9, pp. 1236–1240, 2014.
- [11] E. Gasteiger *et al.*, "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, J. M. Walker, Ed., Humana Press, 2005, pp. 571–607.
- [12] B. Hu *et al.*, "GSDS 2.0: an upgraded gene feature visualization server," *Bioinformatics*, vol. 31, no. 8, pp. 1296–1297, 2015.
- [13] T. L. Bailey *et al.*, "The MEME Suite," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W39–W49, 2015.
- [14] F. Sievers *et al.*, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Mol. Syst. Biol.*, vol. 7, art. 539, 2011.
- [15] K. Tamura, G. Stecher, and S. Kumar, "MEGA11: Molecular evolutionary genetics analysis version 11," *Mol. Biol. Evol.*, vol. 38, no. 7, pp. 3022–3027, 2021.
- [16] L. A. Kelley *et al.*, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nat. Protoc.*, vol. 10, no. 6, pp. 845–858, 2015.
- [17] J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, 2021.
- [18] M. Lescot *et al.*, "PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences," *Nucleic Acids Res.*, vol. 30, no. 1, pp. 325–327, 2002.
- [19] P. Horton *et al.*, "WoLF PSORT: protein localization predictor," *Nucleic Acids Res.*, vol. 35, suppl. 2, pp. W585–W587, 2007.
- [20] X. Dai, Z. Zhuang, and P. X. Zhao, "psRNATarget: a plant small RNA target analysis server (2017 release)," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W49–W54, 2018.
- [21] E. Lyons and M. Freeling, "How to usefully compare homologous plant genes and chromosomes as DNA sequences," *Plant J.*, vol. 53, no. 4, pp. 661–673, 2008.
- [22] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, 2000.
- [23] B. Jassal *et al.*, "The Reactome pathway knowledgebase," *Nucleic Acids Res.*, vol. 48, no. D1, pp. D498–D503, 2020.
- [24] D. M. Goodstein *et al.*, "Phytozome: a comparative platform for green plant genomics," *Nucleic Acids Res.*, vol. 40, no. D1, pp. D1178–D1186, 2012.
- [25] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, 1990.
- [26] Dasgupta, N., Nandy, P., Tiwari, C. and Das, S., 2010. Salinity-imposed changes of some isozymes and total leaf protein expression in five mangroves from two different habitats. *Journal of Plant Interactions*, 5(3), pp.211–221.