Insights into enhancing *Centella asiatica* organ cell biofactories via hairy root protein profiling

Miguel Angel Alcalde[1,2]*, Diego Hidalgo-Martinez[2], Roque Bru Martínez[3], Susana Sellés-Marchart[3], Mercedes Bonfill[2] and Javier Palazon[2]*

[1]Biotechnology, Health and Education Research Group, Posgraduate School, Cesar Vallejo University, Trujillo, Peru, [2]Department of Biology, Healthcare and the Environment, Faculty of Pharmacy and Food Sciences, University of Barcelona, Barcelona, Spain, [3]Plant Proteomics and Functional Genomics Group, Department of Biochemistry and Molecular Biology, Soil Science and Agricultural Chemistry, Faculty of Science, University of Alicante, Alicante, Spain

Recent advancements in plant biotechnology have highlighted the potential of hairy roots as a biotechnological platform, primarily due to their rapid growth and ability to produce specialized metabolites. This study aimed to delve deeper into hairy root development in *C. asiatica* and explore the optimization of genetic transformation for enhanced bioactive compound production. Previously established hairy root lines of *C. asiatica* were categorized based on their centelloside production capacity into HIGH, MID, or LOW groups. These lines were then subjected to a meticulous label-free proteomic analysis to identify and quantify proteins. Subsequent multivariate and protein network analyses were conducted to discern proteome differences and commonalities. Additionally, the quantification of rol gene copy numbers was undertaken using qPCR, followed by gene expression measurements. From the proteomic analysis, 213 proteins were identified. Distinct proteome differences, especially between the LOW line and other lines, were observed. Key proteins related to essential processes like photosynthesis and specialized metabolism were identified. Notably, potential biomarkers, such as the Tr-type G domain-containing protein and alcohol dehydrogenase, were found in the HIGH group. The presence of ornithine cyclodeaminase in the hairy roots emerged as a significant biomarker linked with centelloside production capacity lines, indicating successful *Rhizobium*-mediated genetic transformation. However, qPCR results showed an inconsistency with rol gene expression levels, with the HIGH line displaying notably higher expression, particularly of the *rolD* gene. The study unveiled the importance of ornithine cyclodeaminase as a traceable biomarker for centelloside production capacity. The strong correlation between this biomarker and the *rolD* gene emphasizes its potential role in optimizing genetic transformation processes in *C. asiatica*.

Introduction

Centella asiatica is a perennial plant, native to parts of Asia, and has garnered significant attention in modern times due to its potential health benefits and diverse applications in medicine and cosmetics (Gray et al., 2017).

The primary bioactive compounds identified in *C. asiatica* are centellosides, which are categorized as pentacyclic triterpenoid saponins. These compounds are utilized to treat a variety of conditions including skin ailments, nervous disorders, and venous insufficiency. The centelloside biosynthesis pathway originates from the mevalonate pathway, ultimately yielding farnesyl diphosphate (FPP) as a sesquiterpene precursor. Squalene synthase further converts FPP into squalene, serving as an intermediate (Gallego et al., 2014). It undergoes oxidation to form 2,3-oxidosqualene, a pivotal branching point in both sterol and triterpenoid saponin biosynthesis. This compound cyclizes into a protosteryl or dammarenyl cation, which subsequently generates various products, including the oleanyl cation (Haralampidis et al., 2002).

The oleanyl cation, catalyzed by α/β-amyrine synthase, leads to the production of α or β-amyrin (Azerad, 2016). After cyclization, further diversity in the resulting compounds is introduced through diverse modifications which are facilitated by enzymes like cytochrome P450-dependent monooxygenases and glycosyltransferases. UDP-glycosyltransferases (UGTs) play a key role in glucosylating asiatic acid and madecassic acid to yield asiaticoside and madecassoside (Kim et al., 2017).

The field of plant biotechnology has experienced significant advancements in recent years, marked by a growing interest in leveraging genetic transformation for diverse applications. In this context, hairy roots obtained by *Rhizobium*-mediated genetic transformation constitute a promising biotechnological platform, owing to their remarkable potential for specialized metabolite production and rapid growth (Gutierrez-Valdes et al., 2020).

Hairy root cultures are initiated through the random integration of a segment of *Rhizobium rhizogenes* DNA (T-DNA), mainly *rol* and *aux* genes, derived from the Ri-DNA plasmid into the plant cell genome, where the expression of the genes carried out by the T-DNA promotes rooting at the site of infection (Veena and Taylor, 2007). Concurrently with this stochastic integration process, the quantity of integrated heterologous genes presents a pertinent yet unexplored aspect. Due to the variable copy numbers of introduced transgenes, specifically T-DNA genes in this instance, have the potential to exert an influence on the collective expression levels of target genes, consequently affecting protein composition and metabolic pathways within hairy roots (Bhat and Srinivasan, 2002).

To fully exploit the advantages offered by hairy roots, a comprehensive understanding of the intricate molecular processes governing their development and metabolic capabilities is essential. One less-utilized tool for achieving this understanding is proteomics, which entails studying the complete set of proteins expressed by an organism and has revolutionized our comprehension of cellular processes and their intricate regulation. Within plant biology, proteomics has emerged as a powerful tool for unraveling the molecular mechanisms underpinning various physiological phenomena (Chen et al., 2020). Proteomics analysis may therefore shed new light on the genes associated with centelloside biosynthesis in *Centella asiatica* hairy roots, similar to observations reported in other plant species (Kim et al., 2003; Contreras et al., 2019; Chen et al., 2022).

Besides offering insights into the dynamic metabolic processes that drive hairy root development, advanced protein profiling techniques may uncover biomarkers of desirable traits. This approach therefore opens the way to achieving the production levels and developmental capacities in hairy roots necessary for their sustainable application as a biotechnological platform (Padilla et al., 2021). Furthermore, the use of omics techniques to study differentially expressed genes in *C. asiatica* hairy roots lays the groundwork for further investigation into the transcriptional regulation of centelloside content (Khan et al., 2023; Shilpha et al., 2023).

The objectives of this study were to conduct a comprehensive investigation into the potential role of protein profiling in hairy roots for the optimization and enhancement of *C. asiatica* organ cell biofactories. Understanding the relationship between transgene copy numbers and gene expression is crucial for developing strategies to improve the stability and performance of transformed lines in various biotechnological applications. Our work significantly extends the findings reported by Alcalde et al. (2022), where distinctive morphological and metabolic variations were observed among different *C. asiatica* hairy root lines, likely due to the random insertion of a limited number of genes from the T-DNA, particularly the *rol* and *aux* genes.

This study aims to provide new insights into the intricate molecular mechanisms governing hairy root development and their impact on the production of specialized metabolites, especially centelloside biosynthesis. By applying advanced protein profiling techniques, our research seeks to identify key proteins and biomarkers associated with enhanced organ cell biofactory performance. Ultimately, our goal is to contribute to the advancement of biotechnological applications by unveiling novel strategies to optimize the production of bioactive compounds through the manipulation of hairy root protein profiles.

Materials and methods

Plant material

The hairy root lines utilized in this study were established and morphologically characterized by Alcalde et al. (2022). To achieve this, we utilized the *Rhizobium rhizogenes* A4 strain and employed leaf segments from 2-month-old *in-vitro C. asiatica* seedlings as explants. These leaf segments, measuring 1.5–2 cm², were cut and exposed to *R. rhizogenes* colonies, then cultured at 25°C. Following this, the explants were co-cultivated in solid MS hormone-free medium with 3% sucrose and pH set at 5.8. After 48 hours of cocultivation in the dark at 28°C, the explants were transferred to fresh solid MS medium containing 500 mg/l cefotaxime. The emerging hairy roots were subsequently excised and placed on a fresh solid MS medium with 500 mg/l cefotaxime in darkness at 25°

C. This process was repeated every 2 weeks for approximately 2 months to eliminate bacteria from the culture.

Transformation confirmation was conducted using a semi-quantitative RT-PCR approach. This method allowed us to detect both the integration and expression of *R. rhizogenes* T-DNA genes (*rolA, rolB, rolC,* and *aux1*) at the transcript level across the different hairy root lines. The validated lines were categorized as HIGH (4.96 ± 0.75), MID (2.48 ± 0.07), or LOW (0.54 ± 0.067), each corresponding to their respective centelloside production levels, expressed in milligrams per gram of dry weight. As a comparative control, wild adventitious (Adv) roots were excised from *in vitro C. asiatica* seedlings and cultivated on solid MS medium at 25°C in darkness.

Five samples, one gram as the initial fresh weight, from the HIGH (formerly designated as L1 by Alcalde et al. (2022), MID (L10), LOW (L3) hairy root, and adventitious root were grown on solid MS medium at 25°C in darkness and subcultured every two weeks. Sampling for protein extraction was conducted two weeks after the last subculture.

Genomic DNA isolation

Hairy root tissue (200 mg) was pulverized in liquid nitrogen and transferred to a 1.5 mL tube. To this was added 0.75 mL of extraction buffer (50 mM EDTA, pH 8.0; 100 mM Tris, pH 8.0; and 500 mM NaCl), along with 0.6 µl of β-mercaptoethanol and 50 µl of 20% SDS. The mixture was incubated at 65°C for 10 minutes. Subsequently, 250 µl of 5 M potassium acetate was introduced, followed by an ice incubation for 20 minutes. The sample was then centrifuged at 4°C for 20 minutes at 10000 g. After recovering the supernatant, 1 mL of isopropanol was added, and the solution was kept at -20°C for 1 hour. The resulting pellet was subjected to centrifugation for 15 minutes at 10000 rpm, followed by drying.

To the dried pellet, 140 µl of T10E1 buffer (Tris 10 mM, EDTA 1 mM) was added. This mixture was then centrifuged for 10 minutes at 14000 g, the supernatant was retained, and 15 µl of 3 M sodium acetate and 100 µl of isopropanol were incorporated into the sample. After mixing, the supernatant was again recovered, and centrifugation was carried out for 10 minutes at 14000 g. The resulting pellet was dried at 37°C for 10 minutes, followed by the addition of 30 µl of T10E1 buffer. Finally, the purity of the DNA was assessed using a NanoDrop 2000 Spectrophotometer (Thermo Scientific) and 1 µl of RNAse (10 mg/mL) was introduced to remove residual RNA.

Determination of gene copy number by qPCR

The genomic DNA (gDNA) from each sample was subjected to various dilutions, spanning concentrations from 100 ng/µl to 5 µg/µl. The dilutions were quantified utilizing a NanoDrop 2000 Spectrophotometer (Thermo Scientific). For the quantification of copy numbers of transgenes (*rol*A, *rol*B, *rol*C, and *rol*D), primer sequences were designed using Primer-BLAST (Table 1). As a

TABLE 1 List of primers used for gene copy number estimation and gene expression.

Gene	Primer sequence from 5′ to 3′
rolA	FW: GAATGGCCCAGACCTTTGGA
	RV: TTGGTCAGGGAGGAAATCGC
rolB	FW: CAACCGGATTTGGCCAGAGA
	RV: ATAGGGTTGCATCGTGGTCG
rolC	FW: CGCGCTCATCACCAATCTTC
	RV: ACAGAAAGTGCGGCGAAGTA
rolD	FW : GCGTCGTTCCTCCCTATCAG
	RV: TCTGGCAAGATCGCCACAAA
β-AS	FW: CGGAGATTTCCCTCAGCAGG
	RV: CACAAGCGTTTGCGGTACTC
β-actin	FW: TGACAATGGAACTGGAATGG
	RV: CAACAATACTGGGGAACACT

reference gene (internal control), β-amirin synthase (β-AS) was employed, given its single-copy nature within the genome of *C. asiatica* (Kim et al., 2005).

The quantitative polymerase chain reaction (qPCR) assays were conducted using the QuantStudio3 System (Thermo Fisher). Amplifications were carried out in 10 µl reaction solutions, comprising 1 µl of gDNA from each dilution sample, 2 µl of sterile milliQ H₂O, 5 µl of iTaq™ Universal SYBR® Green Supermix (BIO-RAD), and 1 µl of each specific primer at a concentration of 10 µM. The PCR conditions consisted of an initial step at 95°C for 60 seconds, followed by 40 cycles of denaturation at 95°C for 10 s, annealing at 60°C for 20 s and extension at 72°C for 30 s. The specificity of each primer pair was validated by melting curve analysis (95°C for 15 s, a temperature range of 60–95°C with a ramp rate of 0.1°C/s, followed by 95°C for 15 s). To ensure reproducibility, each assay was performed with three technical replicates for each of the three biological samples.

To calculate the transgene copy number, we adopted the formula outlined by Kanwar et al. (2022) $X/R= 10^{(((Cx-Ix)/Sx-(Cr-Ir)/Sr))}$, incorporating the slope and intercept values obtained from the standard curve. The average Ct values obtained from the four dilutions were utilized. These collected values were then integrated into an equation, which was subsequently plotted. In the context of each group of hairy root lines (HIGH, MID, and LOW), Cx and Cr represent the average Ct values corresponding to the transgene and β-AS, respectively. Ix and Ir denote the intercepts associated with the transgene and β-AS, while Sx and Sr signify the slopes for the transgene and β-AS, respectively. To derive the copy number, the X/R value is multiplied by two.

Gene expression

The *rol* genes expression in the transgenic lines was verified using quantitative real-time polymerase chain reaction (qRT-PCR).

Gene normalization was accomplished using the β-actin gene. Total RNA was isolated from plant material utilizing TRIzol reagent (Invitrogen, Carlsbad, CA). For the qRT-PCR, cDNA was synthesized from RNA treated with DNase I (Invitrogen, Carlsbad, CA) using SuperScript IV reverse transcriptase (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The qRT-PCR assays were performed employing the iTAqTM Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) in the QuantStudio3 System (Thermo Fisher). Each sample was analyzed in triplicate under the following conditions: an initial step at 95°C for 60 s, followed by 40 cycles of denaturation at 95°C for 10 s, annealing at 60°C for 20 s, and extension at 72°C for 30 s. Subsequent to amplification, a melting curve analysis was conducted. To check reproducibility, each assay was performed with technical triplicates for each of the three biological samples. Gene-specific primers were designed using Primer-BLAST (Table 1).

Label-free proteomic analysis

A time series proteomic experiment was conducted using quadruplicates of whole cell extracts from each transgenic line and adventitious roots. Trypsin protein digestion and peptide cleanup were carried out following the procedure described by Wang et al. (2006). For analysis, 30 mg of desalted peptide digests were directly injected onto a reverse phase Agilent AdvanceBio Peptide mapping column (2.1 mm x 250 mm, 2.7 µm particle size) attached to an Agilent 1290 Infinity UHPLC, coupled through an Agilent Jet Stream® interface to an Agilent 6550 iFunnel Q-TOF mass spectrometer (Agilent Technologies) system.

Peptide separation was performed at 50°C using a 140-minute linear gradient of 3-40% ACN in 0.1% formic acid at a flow rate of 0.400 mL/min. Source parameters included a gas temperature of 250°C, drying gas at 14 L/min, nebulizer at 35 psi, sheath gas temperature at 250°C, sheath gas flow at 11 L/min, capillary voltage at 3,500 V, and fragmentor at 360 V. Data were acquired in positive-ion mode using Agilent MassHunter Workstation Software, LC/MS Data Acquisition B.08.00 (Build 8.00.8058.0). Operating in high sensitivity mode, MS and MS/MS data were collected in Auto MS/MS mode. This involved selecting the 20 most intense parent ions (charge states from 2 to 5) within the 300 to 1,700 m/z mass range, provided they exceeded a threshold of 1,000 counts, for subsequent MS/MS analysis. MS/MS spectra spanning the 50-1,700 m/z range were gathered with the quadrupole set to "narrow" resolution. Data acquisition was continued until either a total count of 25,000 was reached or a maximum accumulation time of 333 ms was achieved.

Each MS/MS spectrum was subjected to preprocessing using the extraction tool within the Spectrum Mill Proteomics Workbench (Agilent). This step aimed to generate a peak list and enhance spectral quality by merging MS/MS spectra sharing the same precursor (with a Δm/z < 1.4 Da and chromatographic Δt < 15s). The resulting refined dataset was then subjected to a search against the proteome database, encompassing primary species from the *Apiaceae* family and the *Rhizobium/Agrobacterium* genus. Additionally, the analysis included identification of contaminant

proteins using the identity mode of the MS/MS search tool in the Spectrum Mill Proteomics Workbench, configured as follows: trypsin enzyme specificity, allowance for up to 2 missed cleavages, fixed modification of Cys by carbamidomethylation, variable modification of Met by oxidation, and mass tolerance of 20 ppm for precursor ions and 50 ppm for product ions. The peptide hits obtained were subjected to filtering, retaining those with a score of ≥ 6 and a percent scored peak intensity (%SPI) of ≥ 60.

The LC-MS raw files were imported into Progenesis QI for Proteomics (Nonlinear Dynamics) version 4.0, a label-free analysis software. Quantification was based on MS1 intensity. The data file with the highest number of features (peaks) served as a reference for aligning the retention times of all other chromatographic runs and for normalizing MS feature signal intensity (peak area). To address experimental variations, a robust distribution of all ratios (log (ratio)) was computed for correction purposes. MS features were filtered to encompass only those with a charge state ranging from two to five. Employing a "between subjects" experimental design mode, samples were clustered according to their respective experimental groups (Adv, LOW, MID, and HIGH). Average intensity ratios of matched features across experimental groups, along with p-values from one-way ANOVA, were automatically computed.

For protein identification, the filtered SpectrumMill peptide hits files were introduced into Progenesis QIp. Here, conflicts in peptide assignments were resolved, either by selecting the highest score as the winner or by retaining unresolved conflicts in cases of equal scores and sequences. The inferred protein list was then filtered to include entries with a score of ≥15. To determine protein abundance, the Hi-3 method described by Silva et al. (2006) and implemented in Progenesis QI for Proteomics was employed. Differential protein abundance across experimental groups was evaluated using advanced statistical tools available in Progenesis QIp, including ANOVA and Power analysis.

Statistical analysis and protein network

The protein abundance datasets were imported into SIMCA-P software, version 14.1 (Umetrics, Umea, Sweden) for further analysis. To ensure comparability, all variables underwent pareto scaling prior to multivariate analysis. The pareto-scaled variables underwent orthogonal partial least squares discriminant analysis (OPLS-DA) to elucidate distinctive components within the sample set. The OPLS-DA predictive component loading was visualized using S-plots, a technique proven effective for enhancing model interpretation and biomarker discovery (Wiklund et al., 2008). Model quality assessment was performed using R2X (cumulative) and Q2 values, following the criteria established by Triba et al. (2015).

Statistical analyses to assess differences in protein levels were conducted using GraphPad Prism 7.0 software. Analysis of variance (ANOVA) was employed, followed by Tukey's *post hoc* test to determine significant variations between protein levels. To discern pairwise differences between means, the Tukey-Kramer multiple-comparison test was utilized, with a significance threshold set at p <

0.05. To predict protein-protein interactions, a list of protein identifiers was submitted to the web interface of the STRING. (2023). https://string-db.org/ [Accessed July 27, 2023].

Results

Proteomic profiles in hairy and adventitious roots

The experimental design proposed in this study aimed to investigate and compare the complete set of soluble proteins expressed in both transformed and adventitious roots. The Adv roots were collected from *in vitro*-grown *C. asiatica* seedlings. The transformed roots, which were morphologically and phytochemically characterized by (Alcalde et al., 2022), were categorized as HIGH, MID, or LOW based on their respective capacities for the production of centellosides – the key bioactive compounds of *C. asiatica* plants.

Samples of each root group were subjected in quadruplicate to a label-free proteomic analysis, which was conducted by searching against the Uniprot databases for *Apiaceae* taxonomy (see Material and Methods). A total of 213 quantifiable proteins were identified and selected based on specific criteria: they were required to have a SCORE ≥15, p-value ≥ 0.05, and a fold change (FC) ≥2.

As a first approximation after this data filtering, an OPLS-DA was carried out with the 213 proteins quantified with the *Apiaceae* database. The score scatter plot of this model (Figure 1) displays the variation between the groups, with components 1 (48.7%) and 2 (27.9%) representing the maximum separation. The model provided a satisfactory explanation of the variation, showing a good fit with an R2X (cum) value of 0.786. Further, the reliability of the model was confirmed by cross-validation, resulting in a Q2 (cum) value of 0.861.

FIGURE 1
Score scatter plot of the OPLS-DA model conducted using the quantified data of 213 proteins from the Apiaceae database for the various root lines.
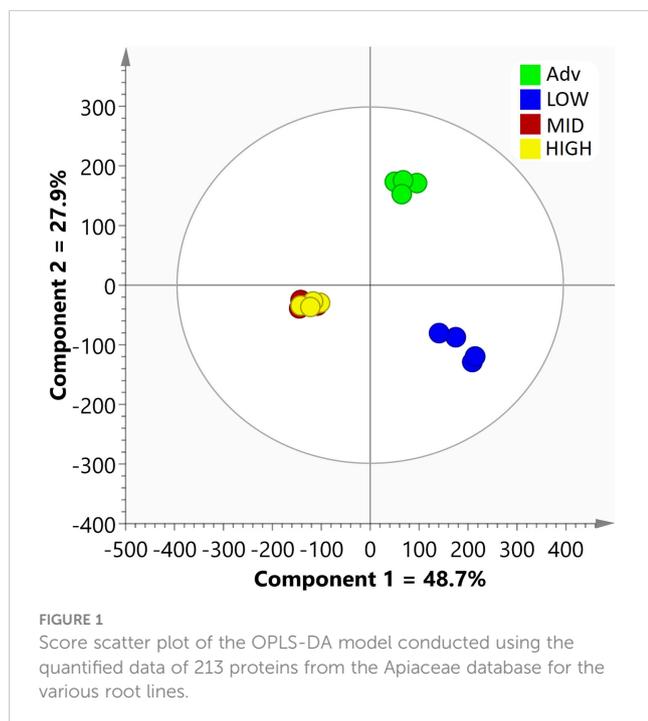
Interestingly, the proteomes of HIGH and MID roots exhibited minimal differences, whereas both differed significantly from the ADV and LOW proteomes. Notably, the most pronounced difference was observed between the LOW group and the others.

The list of quantified plant proteins was analyzed using STRING. Out of 213 proteins, 184 were found in the STRING database and utilized to construct a network, displaying significant interactions (refer to Supplementary Figure 1). The identified proteins were further classified based on their biological processes (Gene Ontology) and KEGG Pathways (see Supplementary Table 1). The classification revealed the recurrence of proteins associated with essential processes such as photosynthesis and amino acid biosynthesis. Additionally, proteins related to pathways of specialized metabolites, including phenylpropanoids, were also prominent.

In the next step, ANOVA and Tukey tests were conducted to identify proteins with significant differences among the different root lines. The comparison between the HIGH and LOW lines revealed that only 44 proteins exhibited statistical variations (Supplementary Table 2). Out of these, only 38 could be utilized to construct a protein network using STRING (Figure 2). Surprisingly, the majority of the differentially expressed proteins were found in higher abundance in the LOW hairy roots, and only a few were found in greater quantities in the HIGH group. The differentially expressed proteins are as follows: alcohol dehydrogenase, UDP-arabinopyranose mutase, Tr-type G domain-containing protein, plant heme peroxidase, D-3-phosphoglycerate dehydrogenase, and ketol-acid reductoisomerase.

The proteins from this network were further classified based on their biological processes and KEGG pathways (Supplementary Table 3). This classification revealed the absence of proteins associated with photosynthesis, previously found when the Adv roots were included in the multivariable analysis. Additionally, proteins related to pantothenate, CoA, and phenylpropanoid and amino acid biosynthesis received good scores in the classification.

Biomarker discovery

To identify significant markers among the Adv and hairy root groups, an OPLS-DA analysis was conducted for each comparison using the 213-protein dataset mentioned above. To aid the visualization of the discrimination model in terms of biomarkers, an S-plot (Wiklund et al., 2008) was utilized to filter potential proteins. The S-plot of the Adv vs HIGH model (Figure 3A) illustrates the magnitude (modeled covariation) and reliability (modeled correlation) of each protein. Putative biomarkers were identified based on a small set of proteins exhibiting high magnitude (≥ |0.1|) and reliability (≥ |0.8|). In this specific model, we identified only two biomarkers positively correlated with the HIGH group: a Tr-type G domain-containing protein and an alcohol dehydrogenase protein. In contrast, 13 proteins were correlated with the Adv group (Supplementary Table 4), most of them related to photosynthesis.

The S-plot analysis of the Adv vs MID model identified alcohol dehydrogenase protein as a putative biomarker, showing a

FIGURE 2
Protein network constructed using the STRING web interface, illustrating the network formed by 38 proteins found in the STRING database. Network nodes depict proteins, and edges symbolize protein-protein associations. The accompanying legend displays the UniProt accession numbers.

FIGURE 3
The S-plot depicts the magnitude (modeled covariation) and reliability (modeled correlation) of each protein, represented by triangles. The red rectangle highlights the region with high magnitude ($\geq |0.1|$) and reliability ($\geq |0.8|$). **(A)** S-plot analysis of the Adv vs HIGH OPLS-DA model: Number 1 corresponds to Tr-type G domain-containing protein, and number 2 corresponds to alcohol dehydrogenase protein. **(B)** S-plot analysis of the LOW vs HIGH OPLS-DA model. **(C)** S-plot analysis of the MID vs HIGH OPLS-DA model: Number 3 represents Bet v I/Major latex protein.

significant correlation with the MID group. Intriguingly, the clearest distinction between two groups was observed in the Adv vs LOW model. At least six distinct proteins were found to be positively correlated with the LOW roots, including the PCMH-type FAD-binding domain, eukaryotic translation initiation factor, peroxidase, and cysteine protease.

Additional comparisons between the transgenic lines revealed that Tr-type G domain-containing protein was the only one that could potentially serve as a reliable indicator to distinguish between the HIGH/MID and the LOW lines (Figure 3B). A detachable potential biomarker differentiating between HIGH and MID was the Bet v I/Major latex protein, which was more strongly correlated with the MID group (Figure 3C).

Finally, we conducted a proteomic analysis by searching against the Uniprot databases for *Rhizobium/Agrobacterium* taxonomy, as described in the Material and Methods section. A total of 100 quantifiable proteins were identified and carefully selected based on the specific criteria described above. This dataset was then used for an OPLS-DA analysis comparing Adv and HIGH lines, followed by an S-plot to identify potential biomarkers. The proteins showing high magnitude ($\geq |0.1|$) and reliability ($\geq |0.8|$) were manually curated to ensure that only proteins potentially originating from *Rhizobium* via T-DNA were included, filtering out any proteins that could be derived from other sources, to ensure the selection of true *Rhizobium* biomarkers. Consequently, only the ornithine cyclodeaminase protein (OCD) complied with the curation process.

In *R. rhizogenes*, OCD is located in the T-DNA region of the Ri plasmid and is referred to as RolD by Trovato et al. (2001). The presence of OCD in the hairy roots is noteworthy, as this enzyme plays a crucial role in synthesizing proline from ornithine in a single step. According to Trovato et al. (2018), this metabolic capability could be
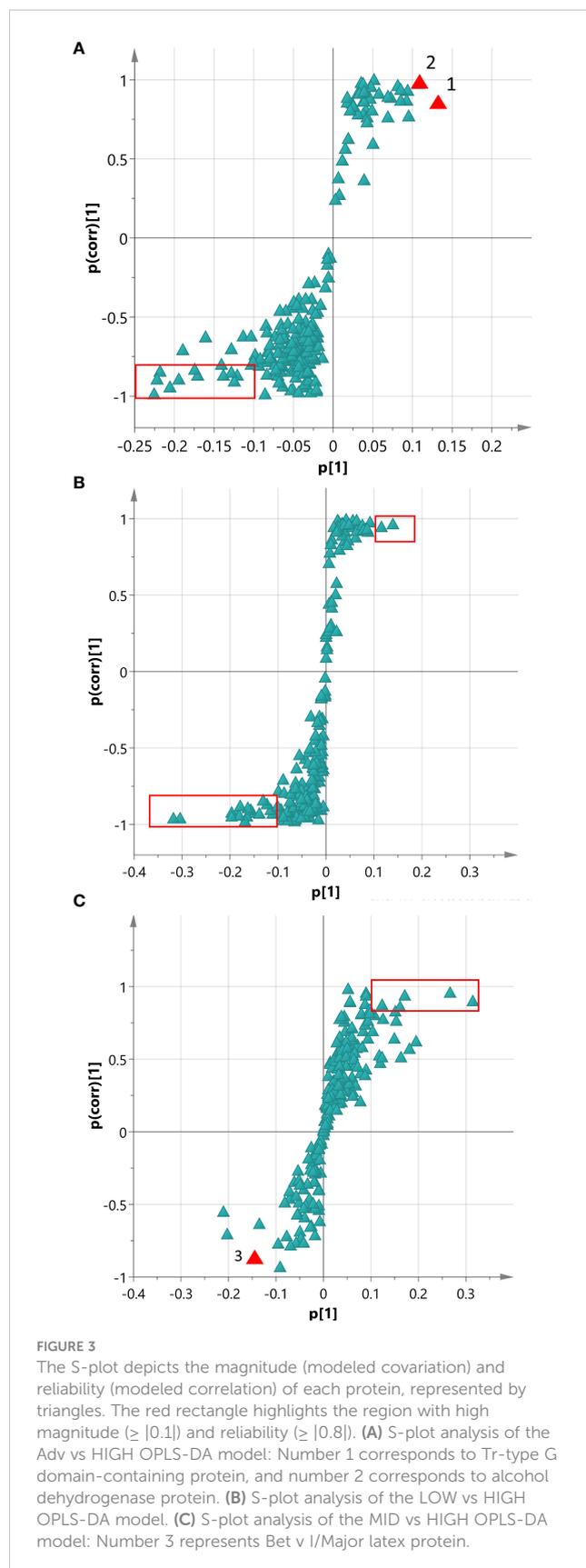
functionally involved in the process of root elongation and/or maturation. The production of higher amounts of proline, an important osmolyte and signaling molecule, may contribute to stress

tolerance and growth regulation in the hairy roots, making it a significant biomarker of *Rhizobium*-mediated genetic transformation.

Impact of transgene copy number on gene expression

Based on the discovery of the OCD biomarker, and with the objective of investigating the potential impact of the transgenes from the T-DNA of *R. rhizogenes* on the hairy root proteome profiles, we proceeded to quantify the number of copies and the level of expression of the *rol* genes in the transgenic lines.

The transgene copy number for all transgenic lines was determined using qPCR relative to the endogenous reference gene β-amyrin synthase (β-AS), a gene involved in the biosynthesis of centellosides (Kim et al., 2005). The quantifiable parameters collected are shown in Supplementary Table 5. After calculations, all the tested lines were estimated to have two copies of each transgene (see Table 2). Nevertheless, upon scrutinizing the gene expression patterns, clear distinctions emerged between the groups (Figure 4). In particular, the HIGH line stands out due to heightened expression levels of most *rol* genes, especially the *rol*D gene, which was found to be 5 to 8 times more pronounced than in the LOW and MID lines, respectively. These differences suggest that the presence of two copies of the transgenes does not necessarily correlate with uniform expression levels. Furthermore, the greater abundance of *rol*D gene transcripts appears to correlate with the OCD biomarker, suggesting its potential significance.

Discussion

The literature contains numerous examples of plant biofactories based on hairy root cultures designed to produce plant bioactive

FIGURE 4
Normalized gene expression values from the transgenic lines: HIGH, MID, and LOW. Asterisks indicate statistical differences among the lines solely for the rolD gene (a = 0.05). Data represent the mean ± SD of three replicates.

compounds scarcely synthesized in nature (Hasnain et al., 2022; Bapat et al., 2023; Sonkar et al., 2023). Most of these studies adopt empirical approaches, concentrating on establishing hairy root cultures to explore their production of specialized plant compounds and optimizing the biotechnological production system. In contrast, only a limited body of research has taken a rational approach, trying to understand how a specific set of genes from *R. rhizogenes* (*rol* genes) can modify cell growth and metabolism (Mauro et al., 2017; Sarkar et al., 2018).

With the aim of increasing our understanding of the effects of *rol* genes on plant metabolism, in the present study, an experimental design was meticulously crafted to determine a complete representative repertoire of soluble proteins expressed within two distinct root types: transformed (carrying the *rol* genes) and adventitious roots (excised directly from the plants and cultured separately). The protein profiling approach employed in this study not only provides insights into the dynamic metabolic processes underpinning hairy root development but also furnishes a lens through which we can discern biomarkers associated with traits that bestow the coveted production and developmental capacities required for a sustainable biotechnological platform.

The Adv roots were sampled from *C. asiatica* seedlings cultivated within a carefully controlled *in vitro* environment. Moreover, strong precautions were exercised to regulate the growth conditions of these roots, including light shielding, to ensure a robust basis for comparison with the conventionally grown transformed roots. The specific transformed lines were selected according to their previous characterization by (Alcalde et al., 2022), which benchmarked their performance for biotechnological production application.

Despite the care taken in cultivating the Adv roots, proteins related to photosynthesis were surprisingly evident in their protein profiles, the levels being significantly higher than in the transgenic roots. This discrepancy led us to investigate further and prompted a comparison between the HIGH, MID, and LOW lines. The results of this analysis proved to be particularly relevant in our quest to

TABLE 2   Estimated number of copies of each transgene.

gene	Line	2(X/R)	Estimated number of copies
rolA	HIGH	2.001	2
	MID	2.248	2
	LOW	1.417	1-2
rolB	HIGH	1.325	1-2
	MID	1.999	2
	LOW	2.000	2
rolC	HIGH	2.000	2
	MID	2.248	2
	LOW	2.238	2
rolD	HIGH	2.000	2
	MID	1.997	2
	LOW	2.000	2

unravel the protein networks that significantly impact the performance of hairy root lines in biotechnological applications. The comparisons revealed several key biomarkers, notably the Tr-type G domain-containing protein, alcohol dehydrogenase protein, and Bet v I/Major latex protein.

The Tr-type G domain-containing protein is categorized within the GTPase family of classical translation factors under the EF-G/EF-2 subfamily. These elongation factors play a fundamental role in the process of translation, which constitutes a fundamental step in the intricate process of protein synthesis (Xu et al., 2022). This observation aligns well with the morphological traits characterizing the HIGH line, as reported by (Alcalde et al., 2022), which demonstrated superior elongation, branching, and biomass production.

Alcohol dehydrogenases (ADH) in plants, pioneering subjects in early molecular research (App and Meiss, 1958), play an important role in orchestrating the conversion of ethanol to acetaldehyde, as described by Strommer (2011). Notably, increased ADH expression in *Arabidopsis* has been associated with enhanced tolerance to anoxia and improved root growth, as demonstrated by Shiao et al. (2002). Interestingly, these traits align with the morphological characteristics of the hairy roots in our study.

The major latex protein (MLP) subfamily has pivotal functions in defense and stress responses, and forms the second-largest category of the birch pollen allergen Bet v 1 superfamily, as elucidated by Yuan et al. (2020). MLPs are frequently sequestered within laticifers – latex-filled tubular structures – distributed throughout the plant. Such compartments serve as optimal reservoirs for defense metabolites, functioning as a frontline defense mechanism, an aspect highlighted by Musidlak et al. (2020).

MLPs also play a role in enhancing stress tolerance through intricate plant hormone signaling pathways. It is plausible that MLPs interact with indole-3-acetic acid (IAA) and participate in the auxin signaling pathway, influencing the IAA levels in hairy roots. This is significant due to the relevant role that this plant hormone plays in root induction and development (Patten and Glick, 2002; Duca and Glick, 2020). This intriguing connection could potentially explain their function as biomarkers for MID lines, as suggested by Fujita and Inui (2021).

Although numerous authors have suggested that the *rol*D gene of the *R. rhizogenes* T-DNA does not play a significant role in hairy root induction, our results contradict this hypothesis (Trovato et al., 1997; Bulgakov, 2008). The detection of the OCD biomarker indicates a potential connection between *Rhizobium*-mediated genetic alteration and the proteome composition of hairy roots. The enzymatic transformation of ornithine into proline through the OCD-like function of rolD may provide a credible rationale for its involvement in the generation of hairy roots (Trovato et al., 2008; Trovato et al., 2018). It is worth noting that prior studies have reported a significant escalation in proline levels within the growth region of primary maize roots under conditions of limited water availability, implying a crucial function of proline synthesis in maintaining root growth (Verslues and Sharp, 1999). An elevated concentration of proline has the potential to influence the synthesis of hydroxyproline-rich glycoproteins (HRGPs), encompassing extensins and arabinogalactan proteins, which serve as integral structural constituents of the plant cell wall (Okumoto et al., 2015). HRGPs are thought to oversee essential processes such as cell division, the self-assembly of the cell wall, and cell elongation, which could contribute to the noted impacts of RolD on root growth (Trovato et al., 2001). Alternatively, the promotion of root growth by *rolD* could also be associated with the reduction of ornithine, thus impacting the polyamine reservoir, where ornithine operates as a precursor. The overexpression of arginine decarboxylase, another polyamine precursor, has been demonstrated to increase putrescine levels and hinder root growth in tobacco plants (Masgrau et al., 1997). Further research into the specific role of OCD protein in root development and stress responses could enhance our knowledge of the mechanisms underlying hairy root formation. This would potentially open new avenues for biotechnological applications in agriculture and plant biotechnology, specifically in the development of new plant biofactories for the production of high added value compounds synthesized in plant roots.

The non-identification of proteins associated with the T-DNA is not unexpected and can be attributed to the particular technique and parameters employed (see Material and Methods), which might not capture the full range of proteins, especially those of lower abundance. Moreover, the observed variations in gene expression across the different lines could potentially arise from a multitude of factors, including the exact insertion site of the transgene within the genome, given that the insertion of the T-DNA is a random process (Gelvin, 2017; Singer, 2018). Our findings regarding the quantity of copies originating from the T-DNA support the notion that it is the insertion site, rather than the number of transgene copies, which exerts a more pronounced influence on elevated expression levels and subsequent protein translation. This stands in contrast to the conventional belief that comprehending the impact of transgene copy numbers is paramount for optimizing genetic transformation, thereby ensuring consistent and predictable outcomes in hairy root growth and the production of secondary metabolites (Ludwig-Müller et al., 2014). Furthermore, it should be noted that high-copy number transgenes may be more susceptible to instability, potentially resulting in the loss of the inserted genes. Therefore, gaining a comprehensive understanding of copy number dynamics remains critical for maintaining stable genetic modifications (Yang et al., 2005).

Furthermore, the intricate network of molecular mechanisms orchestrating gene expression further contributes to this variability. It is noteworthy that the interplay between transcriptional regulation, post-transcriptional modifications, and protein turnover can often lead to discrepancies between gene expression levels and actual protein abundances. This divergence reflects the complexity inherent in translating genetic information into functional proteins and shows that the underlying processes can only be comprehensively understood through a holistic approach

encompassing both transcriptomic and proteomic analyses (Kumar et al., 2016; Stenton et al., 2020; Veenstra, 2021).

The findings in this work shed light on the proteomic differences among the Adv, MID, and LOW root lines, contributing to a deeper understanding of the molecular basis underlying their diverse characteristics. The comprehensive proteomic analysis performed provides valuable insights into the proteins associated with *Rhizobium* infection. The identified biomarkers hold great promise for further investigations into the mechanisms of *Rhizobium*-mediated genetic transformation and their implications in biotechnology and plant genetic engineering. Our findings underscore the importance of not only quantifying transgene copy numbers but also assessing their impact on gene expression and protein accumulation. Understanding these effects will be crucial for optimizing genetic transformation strategies and ensuring consistent and predictable outcomes in biotechnological applications.

Data availability statement

The mass spectrometry proteomics data presented in the study have deposited to the ProteomeXchange Consortium via the PRIDE partner repository, accession number PXD045705.

Author contributions

MA: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. DH-M: Conceptualization, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. RB-M: Methodology, Resources, Writing – original draft. SS-M: Methodology, Resources, Writing – original draft. MB: Formal Analysis, Investigation, Supervision, Writing – original draft. JP: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1274767/full#supplementary-material

References

Alcalde, M. A., Müller, M., Munné-Bosch, S., Landín, M., Gallego, P. P., Bonfill, M., et al. (2022). Using machine learning to link the influence of transferred *Agrobacterium rhizogenes* genes to the hormone profile and morphological traits in *Centella asiatica* hairy roots. *Front. Plant Sci.* 13. doi: 10.3389/FPLS.2022.1001023

App, A. A., and Meiss, A. N. (1958). Effect of aeration on rice alcohol dehydrogenase. *Arch. Biochem. Biophys.* 77, 181–190. doi: 10.1016/0003-9861(58)90054-7

Azerad, R. (2016). Chemical structures, production and enzymatic transformations of sapogenins and saponins from *Centella asiatica* (L.) Urban. *Fitoterapia* 114, 168–187. doi: 10.1016/J.FITOTE.2016.07.011

Bapat, V. A., Kavi Kishor, P. B., Jalaja, N., Jain, S. M., and Penna, S. (2023). Plant Cell Cultures: Biofactories for the production of bioactive compounds. *Agronomy* 13, 858. doi: 10.3390/AGRONOMY13030858

Bhat, S. R., and Srinivasan, R. (2002). Molecular and genetic analyses of transgenic plants: Considerations and approaches. *Plant Sci.* 163, 673–681. doi: 10.1016/S0168-9452(02)00152-8

Bulgakov, V. P. (2008). Functions of *rol* genes in plant secondary metabolism. *Biotechnol. Adv.* 26, 318–324. doi: 10.1016/J.BIOTECHADV.2008.03.001

Chen, C., Hou, J., Tanner, J. J., and Cheng, J. (2020). Bioinformatics methods for mass spectrometry-based proteomics data analysis. *Int. J. Mol. Sci.* 21, 2873. doi: 10.3390/IJMS21082873

Chen, Y., Wang, Y., Guo, J., Yang, J., Zhang, X., Wang, Z., et al. (2022). Integrated transcriptomics and proteomics to reveal regulation mechanism and evolution of SmWRKY61 on tanshinone biosynthesis in *Salvia miltiorrhiza* and *Salvia castanea*. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.820582

Contreras, A., Leroy, B., Mariage, P. A., and Wattiez, R. (2019). Proteomic analysis reveals novel insights into tanshinones biosynthesis in *Salvia miltiorrhiza* hairy roots. *Sci. Rep. 2019 9:1* 9, 1–13. doi: 10.1038/s41598-019-42164-3

Duca, D. R., and Glick, B. R. (2020). Indole-3-acetic acid biosynthesis and its regulation in plant-associated bacteria. *Appl. Microbiol. Biotechnol.* 104, 8607–8619. doi: 10.1007/S00253-020-10869-5

Fujita, K., and Inui, H. (2021). Review: Biological functions of major latex-like proteins in plants. *Plant Sci.* 306, 110856. doi: 10.1016/J.PLANTSCI.2021.110856

Gallego, A., Ramirez-Estrada, K., Vidal-Limon, H. R., Hidalgo, D., Lalaleo, L., Khan Kayani, W., et al. (2014). Biotechnological production of centellosides in cell cultures of *Centella asiatica* (L) Urban. *Eng. Life Sci.* 14, 633–642. doi: 10.1002/ELSC.201300164

Gelvin, S. B. (2017). Integration of *Agrobacterium* T-DNA into the plant genome. *Annu. Rev. Genet.* 51, 195–217. doi: 10.1146/ANNUREV-GENET-120215-035320

Gray, N. E., Alcazar Magana, A., Lak, P., Wright, K. M., Quinn, J., Stevens, J. F., et al. (2017). *Centella asiatica*: phytochemistry and mechanisms of neuroprotection and cognitive enhancement. *Phytochem. Rev. 2017* 17, 161–194. doi: 10.1007/S11101-017-9528-Y

Gutierrez-Valdes, N., Häkkinen, S. T., Lemasson, C., Guillet, M., Oksman-Caldentey, K. M., Ritala, A., et al. (2020). Hairy root cultures—A versatile tool with multiple applications. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.00033

Haralampidis, K., Trojanowska, M., and Osbourn, A. E. (2002). Biosynthesis of triterpenoid saponins in plants. *Adv. Biochem. Eng. Biotechnol.* 75, 31–49. doi: 10.1007/3-540-44604-4_2

Hasnain, A., Naqvi, S. A. H., Ayesha, S. I., Khalid, F., Ellahi, M., Iqbal, S., et al. (2022). Plants in *vitro* propagation with its applications in food, pharmaceuticals and cosmetic industries; current scenario and future approaches. *Front. Plant Sci.* 13. doi: 10.3389/FPLS.2022.1009395

Kanwar, P., Ghosh, S., Sanyal, S. K., and Pandey, G. K. (2022). Identification of gene copy number in the transgenic plants by Quantitative Polymerase Chain Reaction (qPCR). *Methods Mol. Biol.* 2392, 161–171. doi: 10.1007/978-1-0716-1799-1_12

Khan, M. N., Kamal, A. H. M., and Yin, X. (2023). Editorial: Advancements in plant omics for tackling biotic and abiotic stresses. *Front. Plant Sci.* 14. doi: 10.3389/FPLS.2023.1208218

Kim, O. T., Jin, M. L., Lee, D. Y., and Jetter, R. (2017). Characterization of the asiatic acid glucosyltransferase, UGT73AH1, involved in asiaticoside biosynthesis in *Centella asiatica* (L.) Urban. *Int. J. Mol. Sci.* 18 (12), 2630. doi: 10.3390/IJMS18122630

Kim, O. T., Kim, M. Y., Huh, S. M., Bai, D. G., Ahn, J. C., and Hwang, B. (2005). Cloning of a cDNA probably encoding oxidosqualene cyclase associated with asiaticoside biosynthesis from *Centella asiatica* (L.) Urban. *Plant Cell Rep.* 24, 304–311. doi: 10.1007/S00299-005-0927-Y

Kim, S., Kim, J. Y., Kim, E. A., Kwon, K. H., Kim, K. W., Cho, K., et al. (2003). Proteome analysis of hairy root from *Panax ginseng* C. A. Meyer using peptide fingerprinting, internal sequencing and expressed sequence tag data. *Proteomics* 3, 2379–2392. doi: 10.1002/PMIC.200300619

Kumar, D., Bansal, G., Narang, A., Basak, T., Abbas, T., and Dash, D. (2016). Integrating transcriptome and proteome profiling: Strategies and applications. *Proteomics* 16, 2533–2544. doi: 10.1002/PMIC.201600140

Ludwig-Müller, J., Jahn, L., Lippert, A., Püschel, J., and Walter, A. (2014). Improvement of hairy root cultures and plants by changing biosynthetic pathways leading to pharmaceutical metabolites: Strategies and applications. *Biotechnol. Adv.* 32, 1168–1179. doi: 10.1016/J.BIOTECHADV.2014.03.007

Masgrau, C., Altabella, T., Farrás, R., Flores, D., Thompson, A. J., Besford, R. T., et al. (1997). Inducible overexpression of oat arginine decarboxylase in transgenic tobacco plants. *Plant J.* 11, 465–473. doi: 10.1046/J.1365-313X.1997.11030465.X

Mauro, M. L., Costantino, P., and Bettini, P. P. (2017). The never ending story of *rol* genes: a century after. *Plant Cell Tissue Organ Culture (PCTOC)* 131 (2), 201–212. doi: 10.1007/S11240-017-1277-5

Musidlak, O., Bałdysz, S., Krakowiak, M., and Nawrot, R. (2020). Plant latex proteins and their functions. *Adv. Bot. Res.* 93, 55–97. doi: 10.1016/BS.ABR.2019.11.001

Okumoto, S., Anjum, N. A., Signorelli, S., Kishor, P. B. K., Kavi Kishor, P. B., Kumari, P. H., et al. (2015). Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. *Front. Plant Sci.* 6. doi: 10.3389/FPLS.2015.00544

Padilla, R., Gaillard, V., Le, T. N., Bellvert, F., Chapulliot, D., Nesme, X., et al. (2021). Development and validation of a UHPLC-ESI-QTOF mass spectrometry method to analyze opines, plant biomarkers of crown gall or hairy root diseases. *J. Chromatogr. B* 1162, 122458. doi: 10.1016/J.JCHROMB.2020.122458

Patten, C. L., and Glick, B. R. (2002). Role of Pseudomonas putida indoleacetic acid in development of the host plant root system. *Appl. Environ. Microbiol.* 68, 3795–3801. doi: 10.1128/AEM.68.8.3795-3801.2002

Sarkar, S., Ghosh, I., Roychowdhury, D., and Jha, S. (2018). The effects of rol genes of *Agrobacterium rhizogenes* on morphogenesis and secondary metabolite accumulation in medicinal plants. In: Kumar, N (ed) *Biotechnological approaches for medicinal and aromatic plants. Conservation, genetic improvement and utilization.* Springer, Singapore,. pp 27–51. doi: 10.1007/978-981-13-0535-1_2

Shiao, T. L., Ellis, M. H., Dolferus, R., Dennis, E. S., and Doran, P. M. (2002). Overexpression of alcohol dehydrogenase or pyruvate decarboxylase improves growth of hairy roots at reduced oxygen concentrations. *Biotechnol. Bioeng* 77, 455–461. doi: 10.1002/bit.10147

Shilpha, J., Largia, M. J. V., Kumar, R. R., Satish, L., Swamy, M. K., and Ramesh, M. (2022). Hairy root cultures: a novel way to mass produce plant secondary metabolites. In: Swamy, M. K., and Kumar, A.(eds) *Phytochemical Genomics*. Springer, Singapore. doi: 10.1007/978-981-19-5779-6_17

Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. C., and Geromanos, S. J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell Proteomics* 5, 144–156. doi: 10.1074/MCP.M500230-MCP200

Singer, K. (2018). The mechanism of T-DNA integration: Some major unresolved questions. *Curr. Top. Microbiol. Immunol.* 418, 287–317. doi: 10.1007/82_2018_98

Sonkar, N., Shukla, P. K., and Misra, P. (2023). Plant hairy roots as biofactory for the production of industrial metabolites. *Plants as Bioreactors: Overview*, 273–297. doi: 10.1002/9781119875116.CH11

Stenton, S. L., Kremer, L. S., Kopajtich, R., Ludwig, C., and Prokisch, H. (2020). The diagnosis of inborn errors of metabolism by an integrative "multi-omics" approach: A perspective encompassing genomics, transcriptomics, and proteomics. *J. Inherit Metab. Dis.* 43, 25–35. doi: 10.1002/JIMD.12130

Strommer, J. (2011). The plant ADH gene family. *Plant J.* 66, 128–142. doi: 10.1111/J.1365-313X.2010.04458.X

Triba, M. N., Le Moyec, L., Amathieu, R., Goossens, C., Bouchemal, N., Nahon, P., et al. (2015). PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol. Biosyst.* 11, 13–19. doi: 10.1039/C4MB00414K

Trovato, M., Maras, B., Linhares, F., and Costantino, P. (2001). The plant oncogene *rolD* encodes a functional ornithine cyclodeaminase. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13449–13453. doi: 10.1073/pnas.231320398

Trovato, M., Mattioli, R., and Costantino, P. (2008). Multiple roles of proline in plant stress tolerance and development. *Rendiconti Lincei* 19, 325–346. doi: 10.1007/S12210-008-0022-8

Trovato, M., Mattioli, R., and Costantino, P. (2018). From *A. Rhizogenes* rold to plant P5Cs: Exploiting proline to control plant development. *Plants* 7 (4), 108. doi: 10.3390/plants7040108

Trovato, M., Mauro, M. L., Costantino, P., and Altamura, M. M. (1997). The *rolD* gene from *Agrobacterium rhizogenes* is developmentally regulated in transgenic tobacco. *Protoplasma* 197, 111–120. doi: 10.1007/BF01279889

Veena, V., and Taylor, C. G. (2007). *Agrobacterium rhizogenes*: Recent developments and promising applications. *In Vitro Cell. Dev. Biol. - Plant* 43, 383–403. doi: 10.1007/S11627-007-9096-8

Veenstra, T. D. (2021). Omics in systems biology: current progress and future outlook. *Proteomics* 21, 2000235. doi: 10.1002/PMIC.202000235

Verslues, P. E., and Sharp, R. E. (1999). Proline accumulation in maize (*Zea mays* L.) primary roots at low water potentials. II. Metabolic Source of Increased Proline Deposition in the Elongation Zone. *Plant Physiol.* 119, 1349–1360. doi: 10.1104/PP.119.4.1349

Wang, W., Vignani, R., Scali, M., and Cresti, M. (2006). A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis* 27, 2782–2786. doi: 10.1002/ELPS.200500722

Wiklund, S., Johansson, E., Sjöström, L., Mellerowicz, E. J., Edlund, U., Shockcor, J. P., et al. (2008). Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Anal. Chem.* 80, 115–122. doi: 10.1021/ac0713510

Xu, J., Liu, L., and Song, G. (2022). Functions and regulation of translation elongation factors. *Front. Mol. Biosci.* 8. doi: 10.3389/FMOLB.2021.816398

Yang, L., Ding, J., Zhang, C., Jia, J., Weng, H., Liu, W., et al. (2005). Estimating the copy number of transgenes in transformed rice by real-time quantitative PCR. *Plant Cell Rep.* 23, 759–763. doi: 10.1007/S00299-004-0881-0

Yuan, G., He, S., Bian, S., Han, X., Liu, K., Cong, P., et al. (2020). Genome-wide identification and expression analysis of major latex protein (MLP) family genes in the apple (*Malus domestica* Borkh.) genome. *Gene* 733, 144275. doi: 10.1016/J.GENE.2019.144275