UNIVERSITAT DE
BARCELONA

Exploring T-DNA gene mechanisms and optimizing centelloside production in *Centella asiatica* hairy root cultures

Miguel Angel Alcalde Alvites

UNIVERSITAT DE BARCELONA

Exploring T-DNA gene mechanisms and optimizing centelloside production in *Centella asiatica* hairy root cultures

Miguel Angel Alcalde Alvites

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

Exploring T-DNA gene mechanisms and optimizing centelloside production in *Centella asiatica* hairy root cultures

Miguel Angel Alcalde Alvites

2023

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

PROGRAMA DE DOCTORAT

Exploring T-DNA gene mechanisms and optimizing centelloside production in *Centella asiatica* hairy root cultures

Memòria presentada per Miguel Angel Alcalde Alvites per optar al títol de doctor per la Universitat de Barcelona

Mercedes Bonfill Baldrich

Director de tesis

Diego Alberto Hidalgo Martínez

Director de tesis

Javier Palazón Barandela

Tutor

Miguel Angel Alcalde Alvites

Doctorando

Acknowledgments

First and foremost, I would like to express my heartfelt gratitude to the organizations that financially supported the research presented in this thesis: the Ministry of Education and Science of Spain (BIO2017-82374-R) and the Generalitat of Catalunya (2017SGR242). I extend my appreciation to the Faculty of Pharmacy and Food Sciences at the University of Barcelona for granting me access to their facilities for conducting the experimental work in this thesis.

I also wish to convey my gratitude to the individuals who contributed directly or indirectly to this research; I will forever be indebted to them. To my advisors, Dr. Mercedes Bonfill and Dr. Diego Hidalgo, for their unwavering dedication, time, and support throughout these years of guiding this work. The experience, both within and beyond the realm of science, has been immensely valuable for my personal development.

To my mentor, Dr. Javier Palazón, I am thankful for guidance, ideas, and support, both on a scientific and personal level. He provided the support necessary for me to seamlessly integrate into the research group and adapt to life in this country since my arrival.

I want to express my appreciation to Dr. Rosa Cusido for her support and valuable advice during the initial years of this thesis. I would like to thank Dr. Roque Bru for allowing me to undertake a research stay in his laboratory at the University of Alicante, offering his knowledge and dedication for a chapter of this work. I would like to express my gratitude to Dr. Sergi Munné and Dr. Maren Muller for their valuable collaboration in the analysis of phytohormones in hairy roots. Additionally, I extend my thanks to Dr. Pedro Pablo Gallego for his insightful machine learning analysis, which played a pivotal role in one of our published articles.

Throughout these years, I am grateful to my colleagues in the research group with whom I shared numerous situations, conversations, journeys, experiences, and learned a great deal from their cultures and advice. To Edgar, Ainoa, Ester, Lucía, Chi, Ji Qi, and Luciana, I cherish the fond memories we created together.

Similarly, I want to express my gratitude to those who made research stays in the group and played a fundamental role in the experimental phase of this work. To Laia, Toni, Luz, and Paula, thank you for your support with your work and time.

To the friends I met during my doctoral journey in Barcelona, who made these years enjoyable and formed a beautiful group: Yelder, Ricardo, Verónica, Belén, and Alejandra. And to my lifelong friends in Peru, who always inquire about and care for my well-being abroad: Anderson, Christian, Jan, and Tamara.

In a special way, I want to thank Fabiola, my life partner during these last years of my doctoral studies, for her unwavering support and patience, which enabled me to complete this academic milestone.

Lastly, I want to express my gratitude to one of the driving forces in my life: my family. They provided me with unending support, love, and words of encouragement to pursue this achievement, despite the distance. To my parents, Cleofé and Anibal, and my brother David, who were there through the ups and downs of these years. To my uncles who always offered words of support: Pedro, Mariella, Javier, Juan, Miriam, Mirna, and Domi. To my grandparents: Simón, Miguelina, Modesta, and Pánfilo, especially the latter, who passed away this year. I will forever remember him for his words of encouragement, his joy, and the character that always defined him. This achievement is dedicated to him.

To all of you and many more, thank you sincerely for contributing to this phase of my life. Without your tremendous support, I would never have been able to complete this work.

List of abbreviations

A: Asiaticoside

AA: Asiatic acid

AACT: acetoacetyl-CoA thiolase

ABA: Abscisic acid

ADH: Alcohol Dehydrogenase

ADP: Adenosine Diphosphate

ANN: Artificial Neural Network

APX: Ascorbate Peroxidase

ASA: Acetyl salicylic acid

At-SQS: *Arabidopsis thaliana* Squalene Synthase

cAMP: Cyclic Adenosine Monophosphate

CAT: Catalase

CORO: Coronatine

CS: Chitosan

CYP: Cytochrome P450-dependent monooxygenase

CYP11: CYP716C11

CYS: Cycloartenol Synthase

DMAPP: Dimethylallyl Diphosphate

DW: Dry Weight

EF: Elongation Factor

FAK: Focal Adhesion Kinase

FPP: Farnesyl Pyrophosphate

FPS: Farnesyl diphosphate synthase

GTP: Guanosine Triphosphate

HMG-CoA: 3-hydroxy-3-methylglutaryl-CoA

HMGR: 3-hydroxy-3-methylglutaryl-CoA reductase

HMGS: 3-hydroxy-3-methylglutaryl-CoA synthase

HPLC: High Performance Liquid Chromatography

HPLC-MS/MS: HPLC coupled to Mass Spectrometry

HR: Hairy Root

HRGPs: Hydroxyproline-rich glycoproteins

IAA: Indole-3-acetic acid

IDI: Isopentenyl Diphosphate Isomerase

IDI: Isopentenyl phosphate

IPA: Isopentenyladenosine

IPK: Isopentenyl Phosphate Kinase

IPP: Isopentenyl Phosphate

JA: Jasmonic acid

LDA: Linear Discriminant Analysis

LS: Hairy root Lines transformed with At-*SQS* transgene

LST: Hairy root Lines transformed with At-*SQS* transgene and *TSAR2* transgene

LT: Hairy root Lines transformed with *TSAR2* transgene

M: Madecassoside

MA: Madecasic acid

MAPK: Mitogen-activated Protein Kinase

MeJa: Methyl jasmonate

MEP: Methylerythritol 4-phosphate/deoxyxylulose 5-phosphate

MVA: Mevalonate

MIC: Minimum Inhibitory Concentration

ML: Machine Learning

MLP: Major Latex Protein

MTT: 3-[4,5-dimethyl-2-thiazolyl]-2,5-diphenyl-2H-tetrazolium bromide

MVD: MVA-5-diphosphate decarboxylase

MVK: Mevalonate kinase

NADPH: Nicotinamide Adenine Dinucleotide Phosphate

OCD: Ornithine cyclodeaminase

OPLS-DA: Orthogonal partial least squares-discriminant analysis

ORF: Open Reading Frames

OSC: 2,3-oxidosqualene cyclase

PBPS: Plant Biotechnological Production System

PCA: Principal Component Analysis

PKA: Protein Kinase A

PMK: Phosphomevalonate kinase

PMV: Phosphomevalonate

POD: Peroxidase Dismutase

qRT-PCR: Quantitative Real Time Polymerase Chain Reaction

RF: Random Forest

RLK: Receptor-like protein kinase

RoI: Root loci

ROS: Reactive Oxygen Species

SA: Salicylic acid

SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

SOD: Superoxide Dismutase

SQE: Squalene Epoxidase

SQS: Squalene Synthase

SVM: Support Vector Machines

TC: Total Centelloside

T-DNA: Transfer DNA

TL-DNA: Left T-DNA

TR-DNA: Right T-DNA

TSAR2: Triterpene saponin biosynthesis activating regulator2.

UGT: UDP-glucosyltransferase

UV: Ultraviolet

WW: Wet Weight

YE: Yeast extract

αβAS: αβ-amyrin synthase

Abstract

Plant biofactories constitute an eco-friendly alternative to chemical synthesis as they have a lower environmental impact and therefore contribute to a more sustainable future. Hairy root cultures, generated by infecting plant material with *Rhizobium rhizogenes*, are efficient plant biofactories for producing high-value bioactive plant compounds such as centellosides, which are found in *Centella asiatica*. An important challenge in plant biotechnology is the sustainable production of centellosides, which have widespread use in dermopharmaceuticals and cosmeceuticals. This PhD thesis aimed to investigate the effects of randomly integrated genes, specifically the *rol* genes, following *R. rhizogenes* infection in *C. asiatica*. The focus was on understanding the significance of changes in morphology and the expression of triterpene biosynthetic genes in hairy root lines to enhance centelloside production. Machine learning models were used to analyze the impact of T-DNA gene expression levels on morphological traits, hormonal profiles, and centelloside contents. The results indicated that a high expression of these genes was associated with increased centelloside production and positively influenced the expression of centelloside biosynthetic genes. Furthermore, we established connections between elevated centelloside levels and isopentenyladenosine, as well as between reduced centelloside content and abscisic acid. Proteomics analysis identified key biomarkers of centelloside production, notably ornithine cyclodeaminase. This biomarker is codified by the *rolD* gene, previously considered of low importance in hairy root induction, which was found to effect hairy root growth and centelloside content. To improve centelloside yields in high-growth *C. asiatica* hairy roots, various biotic elicitors were employed. Elicitation with coronatine/methyl jasmonate resulted in the highest total centelloside content, obtained at day 14 of treatment. Elicitation also notably enhanced the expression of centelloside pathway genes, particularly those involved in the final stages, leading to a greater centelloside production. To optimize centelloside yields in *C. asiatica* hairy roots through metabolic engineering, two transgenes were inserted: a transcription factor *TSAR2* and squalene synthase from *Arabidopsis thaliana*. This resulted in the upregulation of genes involved in the initial steps of the centelloside metabolic pathway, thereby increasing the availability of precursors for an enhanced centelloside production. Some of these precursors were redirected toward phytosterol biosynthesis, leading to morphological changes in the hairy root phenotype and indicating a competitive interaction between the centelloside and phytosterol pathways.

Table of contents

Outline of the PhD thesis

Global context of the PhD thesis: The recent SARS-CoV-2 pandemic has highlighted the importance of access to vaccines and medicines for the entire population of the planet. Additionally, the overexploitation of plant resources and desertification caused by climate change means there is an urgent need to develop biosustainable production systems of phyto- and biopharmaceuticals. Both challenges fall within the 17 goals of the United Nations 2030 Agenda for Sustainable Development as shown below

(https://www.un.org/sustainabledevelopment/sustainable-development-goals/).

The 2030 Agenda for Sustainable Development:

Goal 3. "Ensure healthy lives and promote well-being for all at all ages." **Target 3.b**. "Support the research and development of vaccines and medicines for the communicable and non-communicable diseases that primarily affect developing countries. Provide access to affordable essential medicines and vaccines in accordance with the Doha Declaration on the TRIPS Agreement and Public Health, which affirms the right of developing countries to the fullest use of the provisions in the Agreement on Trade-Related Aspects of Intellectual Property Rights regarding flexibilities to protect public health and, in particular, provide access to medicines for all."

Goal 15. "Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss." **Target 15.5**. "Take urgent and significant action to reduce the degradation of natural habitats, halt the loss of biodiversity and, by 2020, protect and prevent the extinction of threatened species".

Hence, strategies are needed to ensure a global supply of essential medicines, including in developing countries where plants constitute an important source of therapeutic products and high demand can drive species to extinction.

Plant biofactories have been developed in the last decades as sustainable and ecofriendly production platforms of bio- and phytochemicals, offering a range of compelling advantages. Firstly, they play a pivotal role in reducing the environmental impact of specific industries by providing sustainable alternatives to petroleum-based products and chemical synthesis. This shift towards eco-friendly production methods is crucial for a more sustainable future. Secondly, plant biofactories contribute significantly to biodiversity conservation by mitigating the risk of overharvesting or depleting natural resources. By leveraging renewable plant sources, they help to preserve vital ecosystems. Thirdly, the innate ability of plants to grow rapidly and produce biomass makes plant biofactories highly scalable, enabling large-scale production of valuable compounds. This feature, together with customization and diversification opportunities, renders biofactories able to meet a broad range of market demands. The production of biopharmaceuticals and other specialized compounds through natural metabolic pathways further enhances the safety and acceptance of these products. Therefore, the implementation of plant biofactories not only boosts production efficiency but is also aligned with sustainable and environmentally conscious practices.

Plant biofactories: biotechnological platforms based on plant organ and cell cultures for the production of pharmaceuticals, nutraceuticals, and cosmeceuticals.

Hairy root cultures, which are obtained by infecting plant material with a pathogenic strain of *Rhizobium rhizogenes*, are used in plant biofactories to produce naturally scarce plant compounds with a high market price. Their advantages as a biotechnological tool include the ability to grow in hormone-free media, a high growth and branching capacity, a metabolic profile very similar to the plant roots, and high genetic stability.

Hairy roots: fine, hair-like structures with vigorous uncontrolled growth formed after the insertion of specific genes from *Rhizobium rhizogenes* into the plant genome. They can serve as a powerful tool for the sustainable production of bioactive substances.

A major target in plant biotechnology today is the biosustainable production of centellosides, bioactive compounds widely utilized in dermopharmaceuticals and

Centellosides: amyrin-derived triterpene saponins and sapogenins produced in *Centella asiatica*.

4

cosmeceuticals that are produced by *Centella asiatica*, a threatened plant species.

The main objective of this PhD thesis was to study and improve centelloside production in hairy root cultures and gain new insights into the effects of *rol* genes from *Rhizobium rhizogenes* on the morphology and productive capacity of the transformed roots.

The reader of the PhD thesis is participating in a game of "snakes and ladders", as we describe not only the ladders or strategies that allowed us to reach the proposed goals, but also the snakes or pitfalls encountered along the way. Thus, the rungs of the ladders are the techniques applied to gain new knowledge about the biotechnological production of centellosides and the function of *rol* genes, including *in vitro* cultures, genetic transformation, hormonal and metabolic analyses, artificial intelligence, proteomics and metabolic engineering. Hindering our progress were the morphological and productive variability of the different hairy root lines obtained, the metabolic complexity of plant systems, the interplay between primary and specialized metabolism, and the need to interpret large volumes of data generated in experiments.

We believe that these pitfalls, and the solutions subsequently found to overcome them, form an important part of the rational approach undertaken to achieve the goals of the thesis. See Figure

Rol genes: also known as root locus genes, they originate from *Rhizobium rhizogenes*, and when integrated into the plant genome induce the development of hairy roots.

The development of the PhD thesis was inspired by the excellent review by Farré et al. Annu. Rev. Plant Biol. 2014. 65: 187-223, which we recommend to all readers.

Snakes and ladders:

5
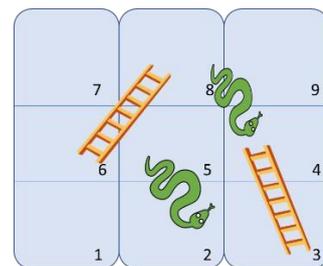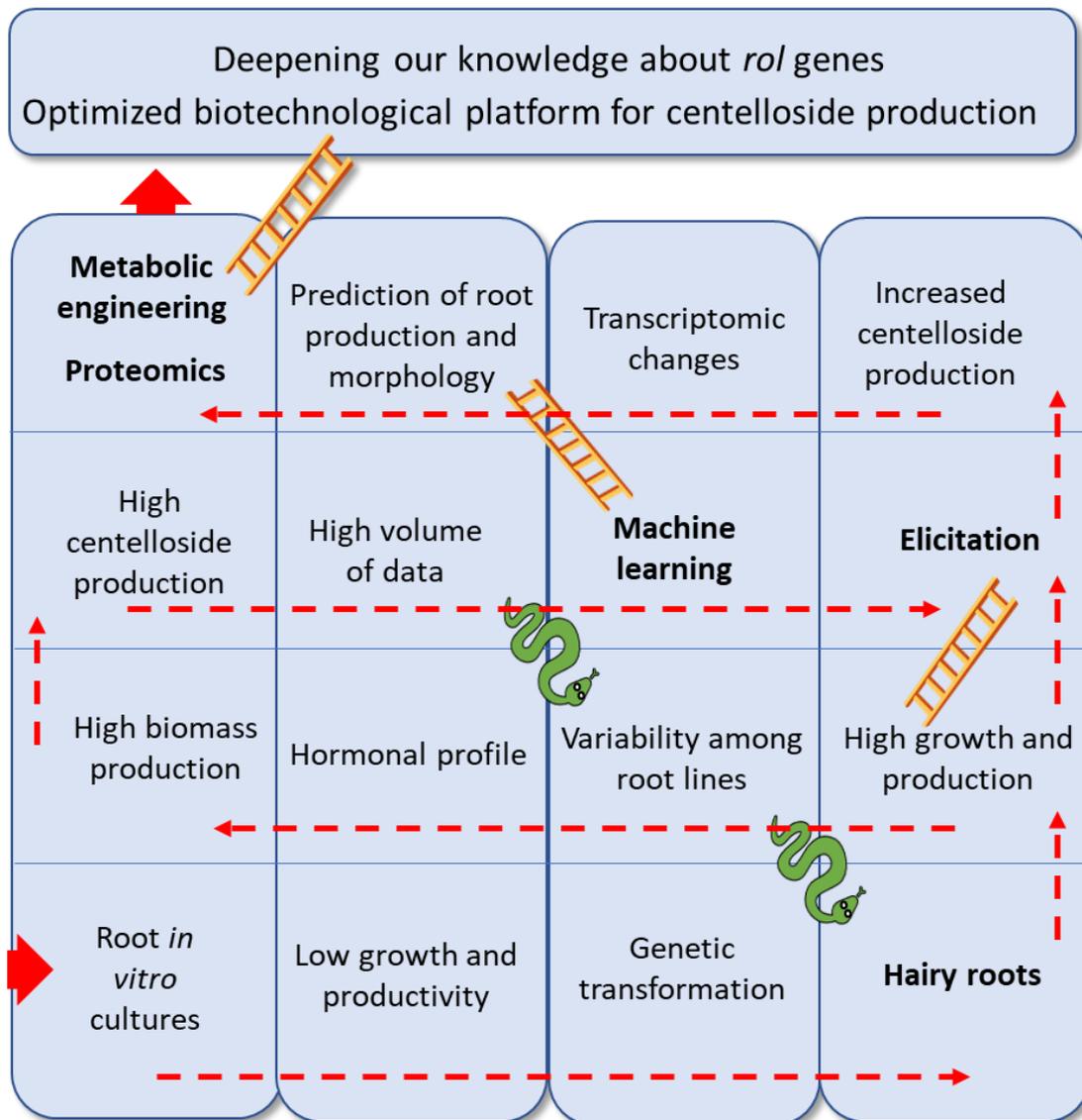
Blueprint of the approach followed in the PhD thesis (modified from Farré et al. Annu. Rev. Plant Biol. 2014. 65: 187-223). As players of snakes and ladders, we used a set of cutting-edge technologies (ladders) to achieve our main objectives, but several pitfalls (snakes) forced us to return to earlier stages of the research. The implementation of alternative techniques, some of them in collaboration with other research groups, allowed us to finally overcome the obstacles and reach our proposed goals.

The PhD thesis is presented as a compendium of four publications, all them authored by the PhD candidate as the first author. The first section is the **Introduction**, which summarizes the main topics dealt with in the thesis, some of them extensively described in two review articles authored by the PhD candidate [Alcalde et al. *Molecules* 2022, 27(16), 5253; Alcalde et al. Critical Reviews in Biotechnology 2023 &&&]. In this part, a series of open questions leads the reader to the **Objectives** of the thesis and the corresponding **Results**. The latter section is divided in four chapters, each one constituting a paper published by the PhD candidate in 1st quartile peer-review ISI

Journals within the Plant Science research field. Each chapter is preceded by an abstract in Spanish and a graphical abstract of the article.

Briefly:

Chapter 1. In the described research, machine learning techniques were applied to shed light on the connection between a small set of T-DNA genes (*rol* and *aux*) and the hormone profile in *C. asiatica* hairy roots, focusing on the development, morphology, and production of secondary metabolites, specifically centellosides [Alcalde, et al., *Frontiers in Plant Science* 2022, 13, 1001023].

Open Questions: The discovered differences among hairy root lines in morphology, centelloside production, hormonal levels, and the expression of the *rol* genes of *Rhizobium rhizogenes*, are due to the integration of different copies of *R. rhizogenes* T-DNA into the plant genome and proteomic alterations compared with untransformed roots?

Chapter 2. The research aim was to determine the gene copy number of some T-DNA genes and perform a comparative proteomic analysis of *C. asiatica* hairy roots with low, medium or high centelloside production as well as adventitious roots to assess the impact on centelloside production [Alcalde et al. *Frontiers in Plant Science* 2023].

Open question: Can strategies such as elicitation, that induce changes in specialized metabolism, generate new knowledge about the transcriptomic control of the centelloside biosynthetic pathway, that cannot be detected by analysis of the complete proteome?

Chapter 3. With the aim of overcoming the low production of *C. asiatica* plants, the effect of different elicitors on centelloside yields and the expression of centelloside biosynthetic genes in *C. asiatica* hairy root cultures was investigated [Alcalde, et al. *Industrial Crops and Products*, 2022, 184, 114988].

Open questions: In centelloside biosynthesis, elicitation affects the expression pattern of downstream genes. What if we unlock the production of precursors of the biosynthetic pathway through metabolic engineering to further enhance the production of these target compounds in hairy root cultures?

Chapter 4. This study explores the overexpression of the squalene synthase (*SQS*) and transcription factor *TSAR2* genes in *C. asiatica* hairy root lines to investigate their impact on the centelloside pathway, morphological traits, and content of squalene, phytosterol, and centellosides [Alcalde, et al. *Plants*, 2023, 12(19), 3363].

<u>Open questions and future perspectives</u>: The discovery that part of the precursor flow for centelloside biosynthesis derives towards early branches of the pathway or accumulates points to downstream blockages. A combined strategy with new stepwise or holistic engineering approaches (using transcription factors) together with elicitation, could unlock the last steps of centelloside biosynthesis and allow the design of new transformed root lines with improved productive capacities.

The main results reported in these chapters are integratively examined in the **Discussion** section, highlighting the most relevant findings. In the **Conclusions,** the main achievements of the PhD thesis are summarized. Finally, in the **References** section, a complete list of all the articles cited in the PhD thesis is provided.

Introduction

Centella asiatica: Unveiling the wonders of a time-honored herb, benefits, and key constituents.

Few plants have captivated the attention of herbalists, scientists, and skincare enthusiasts as profoundly as *Centella asiatica*. Renowned for its rich history in traditional medicine and its modern applications in skincare and wellness, *Centella asiatica*, also known as *Gotu Kola*, *Bua-bok*, *Tiger grass*, *Pegaga*, *Madukaparni* or *Indian Pennywort*, stands as a remarkable botanical wonder with a story that spans centuries. A brief timeline about *C. asiatica* was describe in Figure 1 (Arribas-López et al., 2022; Kunjumon et al., 2022).
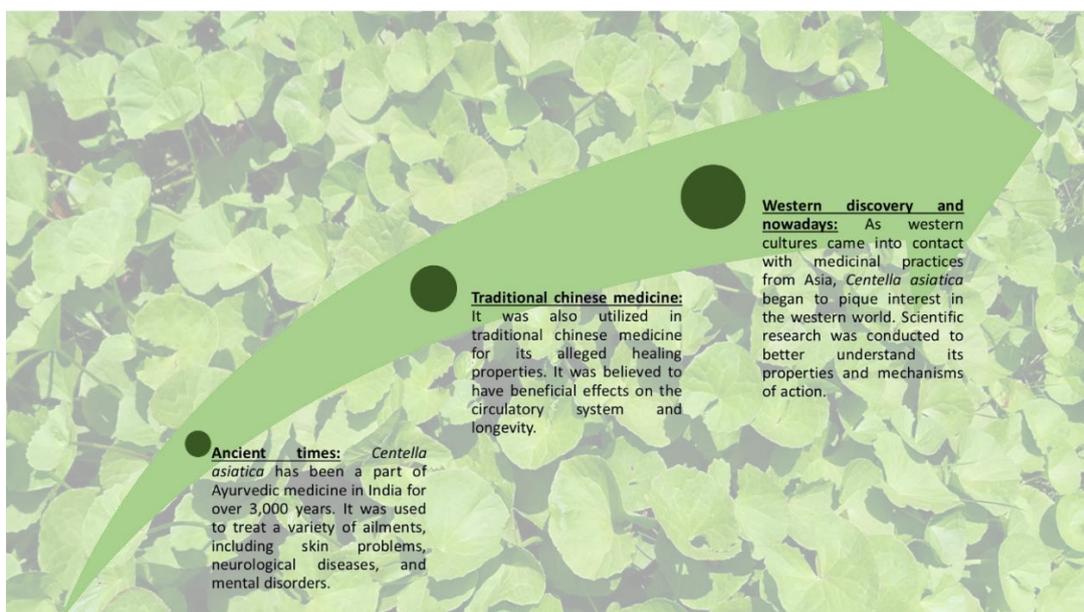
Figure 1. Timeline: Discovery and knowledge transfer journey of *Centella asiatica*.

C. asiatica is a perennial herbaceous plant belonging to the Apiaceae family typically thriving in tropical and subtropical environments, it commonly grows at altitudes exceeding 1800 meters, but its growth originating is in the wetlands of Asia, *C. asiatica* has been an integral part of traditional healing practices across various cultures, from ancient Ayurveda in India to Traditional Chinese Medicine and beyond (Sen and Chakraborty, 2017). Its distinctive fan-shaped leaves and delicate pink or white flowers are a testament to its natural beauty. However, it's not just its appearance that has captured the imagination; it's the potential benefits it offers (Chandrika and Prasad Kumara, 2015).

The comprehensive array of properties exhibited by this herbaceous plant is attributed to a rich repertoire of over 130 secondary metabolites. Among these, triterpenes are the

most important ones, the four principal centellosides stand out: asiatic acid and madecassic acid, classified as sapogenins, and asiaticoside and madecassoside, classified as saponins (Fig. 2) (Luthra et al., 2022).
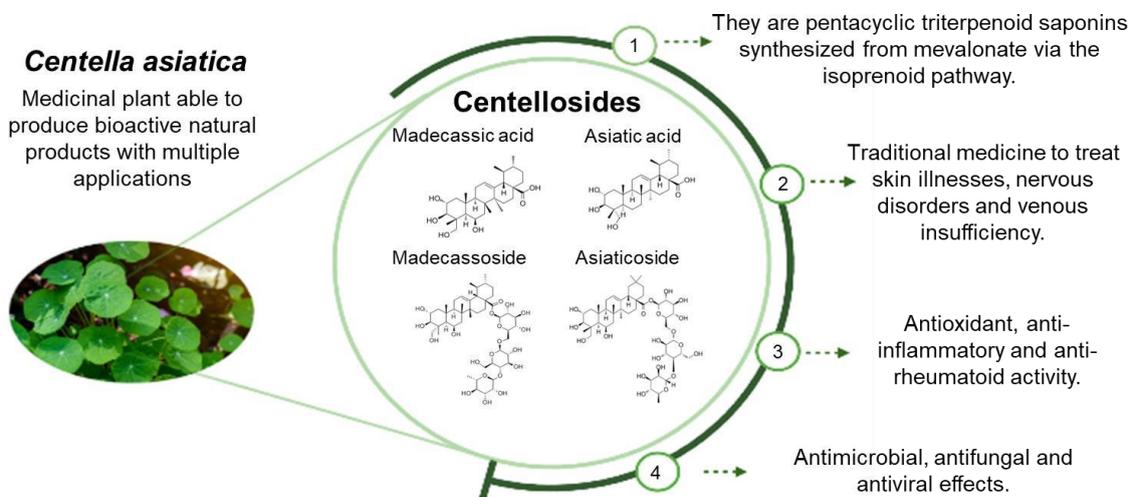
Figure 2. *Centella asiatica*: Key components and properties.

This herb has earned a reputation for its potential cognitive-enhancing properties, leading to its colloquial name *Gotu Kola*, which means fountain of youth in Sanskrit (Belwal et al., 2019). For generations, it has been regarded as a tonic for the mind, believed to promote mental clarity and vitality. Asiatic acid emerges as a pivotal player in the realm of cognitive enhancement, as evidenced by its ability to safeguard spatial working memory and counteract neurogenesis deficiencies in the hippocampal region triggered by 5-fluorouracil chemotherapy (Chaisawang et al., 2017). An aqueous extract derived from *C. asiatica* unveils its potential to amplify synaptic differentiation and dendritic arborization, particularly in response to Aβ—a protein notorious for its role in cognitive decline—thus offering a pathway to cognitive amelioration (Gray et al., 2017). In a notable study, the supplementation of *C. asiatica* extract for a defined duration yielded promising results, effectively addressing cognitive impairment following stroke. Furthermore, asiatic acid not only exhibits the potential to rectify the inhibitions in cell proliferation and spatial working memory prompted by valproic acid treatment, but also holds promise as a multifaceted agent for cognitive revitalization and restoration (Lv et al., 2018).

The neuroprotective facets of *C. asiatica* encompass a spectrum of actions, including enzyme inhibition, thwarting amyloid plaque formation linked to Alzheimer's disease, mitigating dopamine-induced neurotoxicity in Parkinson's disease, and curbing oxidative

stress (Rahman et al., 2012). This plant's contemporary medicinal contributions to neuroprotection spotlight constituents such as asiatic acid, madecassic acid, and brahmaside, collectively amplifying its potential as a versatile ally in safeguarding neurological well-being (Haleagrahara and Ponnusamy, 2010).

Moreover, *C. asiatica* has been revered for its potential to support skin health, thanks to compounds like triterpenoids and flavonoids that contribute to its antioxidant and anti-inflammatory properties (Theerawitaya et al., 2023). The extract of *Pegaga* significantly improved wound tensile strength in an incision model compared to controls, exhibiting accelerated wound contraction rates and enhanced epithelialization (Bylka et al., 2013). In Sprague-Dawley rats, both the extract and powder of *C. asiatica* were assessed for their potential to reduce oxidative stress, revealing a decrease in Reactive Oxygen Species (ROS) generation and oxidative stress, along with a noteworthy decline in Superoxide Dismutase (SOD) levels (Hussin et al., 2007; Pittella et al., 2009). The essential oil of *C. asiatica*, obtained through steam distillation, displayed impressive antioxidant capabilities for lipid-containing foods, comparable to the synthetic antioxidant butylhydroxyanisole (Shenoy et al., 2023). The herb boasts a rich profile of polyphenols, flavonoids, β-carotene, tannins, Vitamin C, and 2,2-diphenyl-1-picrylhydrazyl compounds, contributing to its robust antioxidant activity (Chandrika and Prasad Kumara, 2015). The amalgamation of pentacyclic triterpenoids plays a pivotal role in its therapeutic potential. Notably, *C. asiatica* exhibited inhibitory effects on hypotonicity-induced human red blood cell membrane breakdown, displaying membrane stabilization akin to diclofenac sodium and methanolic extracts, with a remarkable 94.97% stabilization noticed at a 2000 µg/ml dose of *C. asiatica* extracts (Prakash et al., 2017).

The allure of *C. asiatica* has expanded beyond traditional medicine into the realm of modern skincare and wellness. With the rise of natural and holistic approaches to self-care, *Bua-bok* has found its way into various cosmetic formulations, promising to soothe irritation, boost collagen production, and alleviate the effects of stress on the skin (Sun et al., 2020).

The antidiabetic potential of *Tiger grass* leaf extract was assessed in alloxan-induced rats at concentrations of 250, 500, and 1000 mg/kg. Post-ingestion, blood glucose levels decreased by 32.6%, 38.8%, and 29.9%, respectively (Emran et al., 2016). Furthermore, the effect of ethanol extract was investigated in streptozotocin-induced Wistar rats (50 mg/kg). The extract, administered at a minimum concentration of 300 mg/kg, exhibited antidiabetic activity as evidenced by alterations in serum glucose, urea cholesterol, lipid levels, liver glycogen levels, and body weight (Sasikala et al., 2015). In comparison to

13

acarbose and an anti-diabetic drug, the inhibitory activities of α-amylase from *C. asiatica* extract and rutin were relatively lower (Mehta et al., 2016).

The antibacterial efficacy of methanol extract from *C. asiatica* leaves was evaluated using the disc diffusion method, demonstrating zone of inhibition and a minimum inhibitory concentration (MIC) value of 2 µg/disc (Afzan et al., 2005). Notably, the extract exhibited in vitro antibacterial activity against *Staphylococcus aureus* ATCC 25923 and methicillin-resistant *S. aureus* (wild type), resulting in zone of inhibition measurements of 5 mm and 7 mm, respectively (Wong and Ramli, 2021). Furthermore, the essential oil extract displayed significant antibacterial properties against both Gram-positive bacteria (*Bacillus subtilis* and *S. aureus*) and Gram-negative bacteria (*Escherichia coli*, *Pseudomonas aeruginosa*, and *Shigella sonnei*), with MIC values ranging from 1.25 to 0.039 mg/ml (Oyedeji and Afolayan, 2008; Wong and Ramli, 2021).

In the evaluation of antifungal potential, diverse extracts of *Pegaga* were investigated for their efficacy against *Aspergillus niger* and *Candida albicans*. The petroleum ether, ethanol, chloroform, n-hexane, and aqueous extracts demonstrated varying levels of activity, producing inhibition zones ranging from 11 mm to 16 mm against the two fungi (Rupesh Kumar and Ramesh, 2021). Among these, the ethanolic extract stood out, displaying potent antifungal action against *Aspergillus flavus* and *Penicillium citrinum*, with a remarkable mycelial inhibition percentage of 26.3 mm (Shenoy et al., 2023). Further, the 100% ethanolic extract showcased substantial efficacy against *Aspergillus niger*, resulting in an impressive inhibition zone measuring 15.4 mm (Mohd Nadzir et al., 2017).

Indian Pennywort has demonstrated notable anticancer properties. A prime example lies in the assessment of asiatic acid's antiproliferative impact on lung cancer cells, carried out through the 3-[4,5-dimethyl-2-thiazolyl]-2,5-diphenyl-2H-tetrazolium bromide (MTT) assay. Additionally, oral administration of asiatic acid resulted in a substantial reduction in weight and tumor volume within a lung cancer xenograft model (Wu et al., 2017). Further reinforcing its potential, another study illuminated asiatic acid's ability to induce apoptosis and decrease viability in human melanoma SK-MEL-2 cells in a dose-dependent manner (Park et al., 2005). Notably, the antiproliferative effects of asiatic acid extended to RPMI 8226 cells, coinciding with a reduction in focal adhesion kinase (FAK) expression levels. This suggests a plausible link between asiatic acid and the inhibition of FAK-mediated signal transduction (Zhang et al., 2013). Moreover, the titrated extract of *C. asiatica* prominently featured asiatic acid, asiaticoside, and madecassic acid. Within this composition, asiaticoside emerged as a key player, effectively curtailing

melanogenesis in B16F10 mouse melanoma by regulating tyrosinase mRNA expression through microphthalmia-associated transcription factor (Kwon et al., 2014).

Centellosides: Exploring the metabolic and intricate pathway.

In the intricate tapestry of plant biochemistry, few pathways are as captivating as the metabolic journey of centellosides, prominent bioactive compounds found in *C. asiatica* (Figure 3), a plant revered for its traditional medicinal significance and modern therapeutic potential. As the botanical world's chemistry unfolds, the elucidation of centelloside metabolic pathway stands as a testament to the complexity and ingenuity of nature's processes, that is a pitfall that we need to overcome.
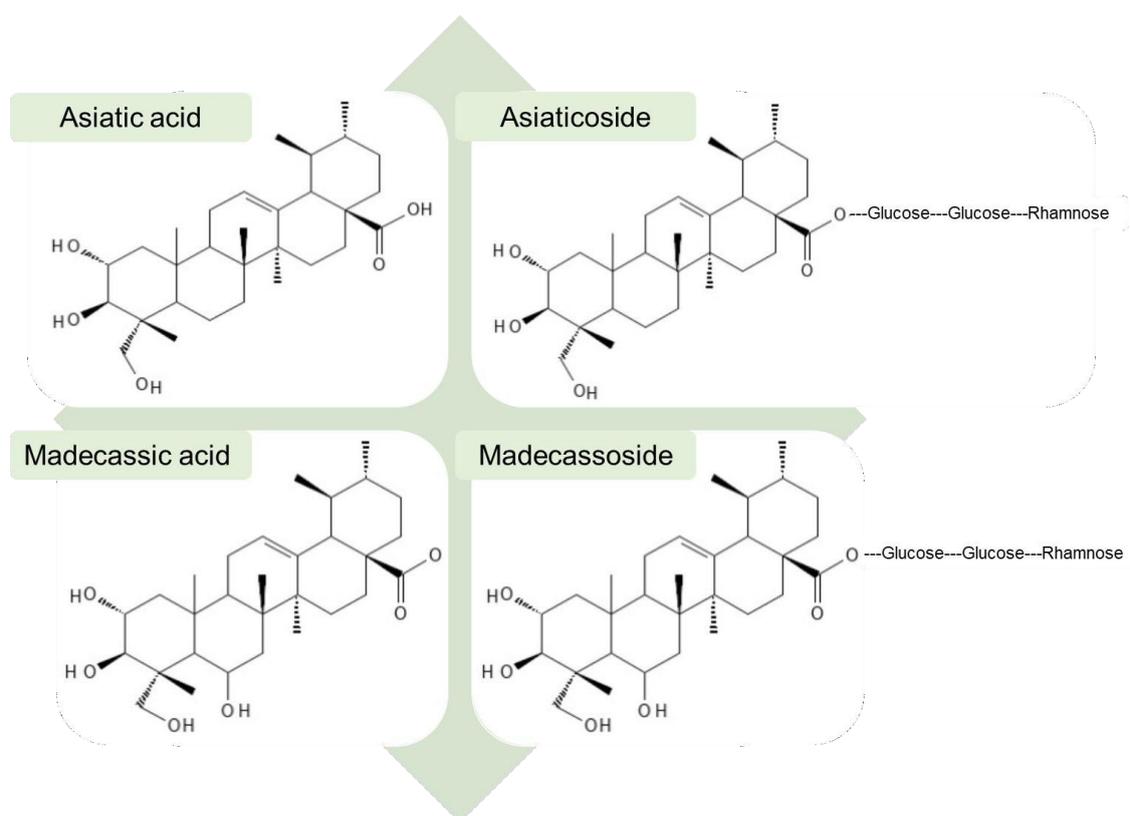
Figure 3. Main centellosides from *Centella asiatica*

Centellosides, characterized by their triterpenoid saponin and sapogenin structure with either a ursane or oleane skeleton, have captivated widespread interest due to their multifaceted pharmacological attributes, as previously highlighted (Prasad et al., 2019). The exploration of their metabolic pathway offers an intricate orchestration of enzymatic reactions, intricate molecular transformations, and precise regulatory mechanisms that underlie their synthesis, transformation, and integral role in shaping the exceptional qualities of *Centella asiatica* (Luthra et al., 2022).
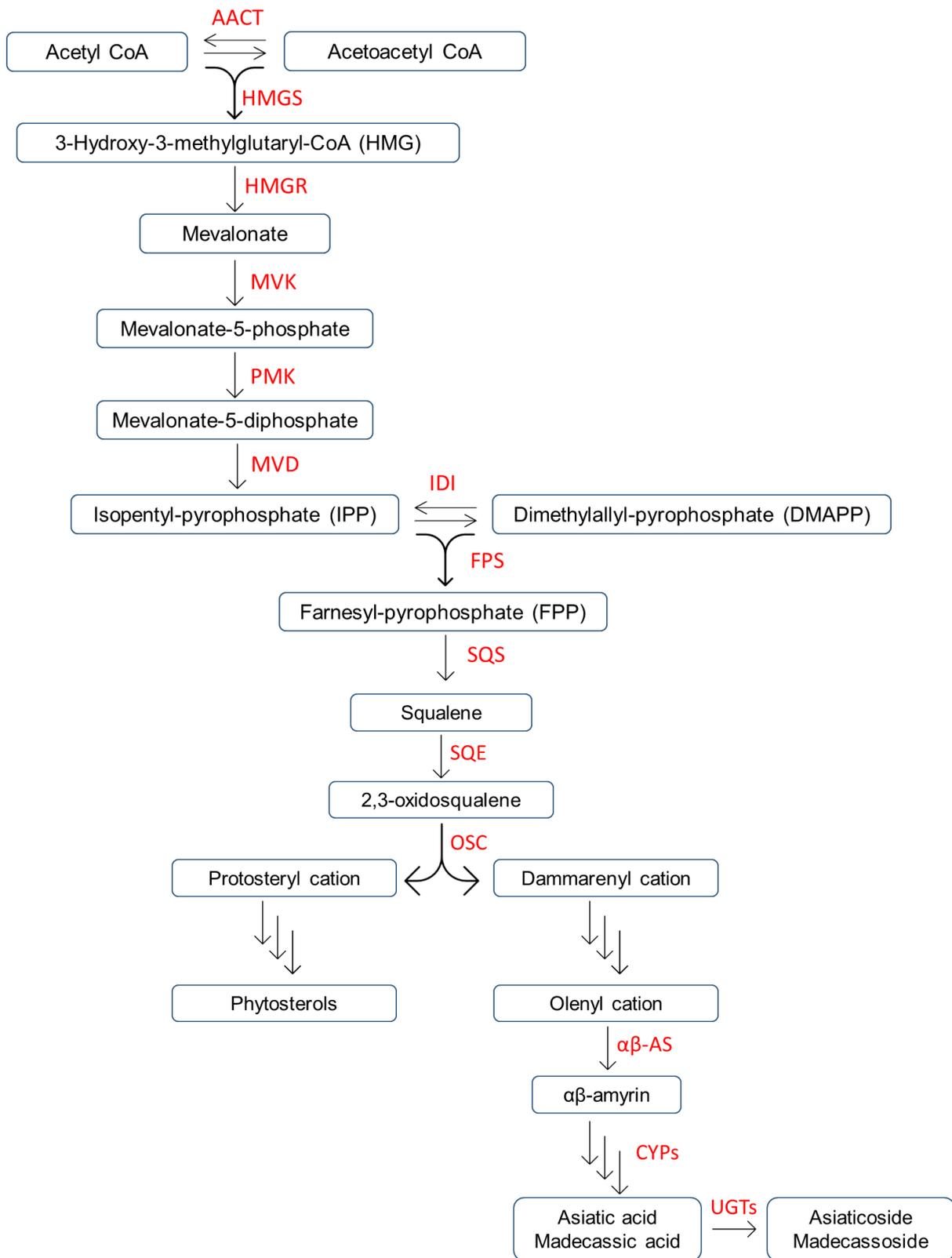
Figure 4. Overview of the centelloside biosynthetic mevalonate (*MVA*) pathway: Key precursors and enzymes. *AACT*, acetoacetyl-CoA thiolase; *HMGS*, 3-hydroxy-3-methylglutaryl-CoA synthase; *HMGR*, 3-hydroxy-3-methylglutaryl-CoA reductase; *MVK*,

mevalonate kinase; *PMK*, phosphomevalonate kinase; *MVD*, MVA-5-diphosphate decarboxylase; *IDI*, Isopentenyl phosphate; FPS, farnesyl diphosphate synthase; *SQS*, squalene synthase; *SQE*, squalene epoxidase; OSC, 2,3-oxidosqualene cyclase; *αβAS*, αβ-amyrin synthase; *CYPs*, cytochrome P450-dependent monooxygenases; and *UGTs*, UDP-glucosyltransferases. Figure generated based on research by Gallego et al. (2014), Aminfar et al. (2019) and Noushahi et al. (2022).

The journey into the biosynthesis of centellosides commences with the realm of terpenoid synthesis, two distinctive pathways are recognized: the methylerythritol 4-phosphate/deoxyxylulose 5-phosphate pathway (MEP) and the mevalonate (MVA) pathway (Noushahi et al., 2022). The latter has garnered particular attention due to its complexity (Figure 4), with enzymes distributed across various subcellular compartments. The MVA route begins with acetyl-CoA being catalyzed by acetoacetyl-CoA thiolase (*AACT*) to produce acetoacetyl-CoA and 3-hydroxy-3-methylglutaryl-CoA synthase (*HMGS*) to generate 3-hydroxy-3-methylglutaryl-CoA (*HMG-CoA*). Then, the pivotal enzyme, 3-hydroxy-3-methylglutaryl-CoA reductase (*HMGR*), stands as a cornerstone, located on the endoplasmic reticulum with its catalytic domain exposed towards the cytosol to produce *MVA*. Other pathway enzymes are found in the cytosol and peroxisomes, contributing to the synthesis of different isoprenoids leading up to Farnesyl pyrophosphate (*FPP*) via the *MVA* route (Vranová et al., 2013; Afroz et al., 2022).

MVA pathway activity is meticulously regulated at multiple levels. Plant cells finely tune *HMGR* activity through differential induction of gene family members at the mRNA level and post-translationally via enzyme modification. This orchestration extends to translation rates, determined by the cell's need for non-sterol isoprenoids, while degradation rates of *HMGR* respond to both sterol and non-sterol isoprenoid requirements. An essential enzyme, mevalonate kinase (*MVK*), guides the phosphorylation of mevalonic acid into phosphomevalonate (*PMV*) subsequent to *HMGR*. *MVK*'s activity is subject to feedback inhibition by farnesyl diphosphate and geranyl diphosphate, both integral intermediates in the isoprenoid pathway (Noushahi et al., 2022).

The journey further unfolds with phosphomevalonate kinase (*PMK*), which catalyzes the transformation of ATP and mevalonate 5-phosphate into ADP and mevalonate 5-diphosphate. Yet, the pathway's regulation encompasses realms yet undiscovered, governing the pathway's flow and ensuing metabolite yield. Amidst this complex

landscape, an additional key enzyme emerges— isopentenyl phosphate kinase (IPK). Found within plant genomes, *IPK* resides in the cytoplasm, converting isopentenyl phosphate (IPP) and possibly dimethylallyl phosphate (*DMAP*) into their phosphorylated forms, isopentenyl diphosphate (*IPP)* and dimethylallyl diphosphate (*DMAPP)*, through ATP-dependent phosphorylation. At once, MVA-5-diphosphate decarboxylase (MVD) produces IPP from mevalonate 5-diphosphate (Henry et al., 2018).

The architectural foundation of the centelloside pathway is crafted by the actions of isopentenyl diphosphate isomerase (*IDI*), which generates *DMAPP* from IPP. As we traverse this intricate mosaic of enzymatic interplay and regulatory control, we begin to fathom the depth of complexity that underpins the synthesis of centellosides, offering a glimpse into the dynamic dance of biochemistry within the realm of *C. asiatica* (Gallego et al., 2014).

Intriguingly, plants orchestrate this process through either the MVA pathway in the cytosol or the methylerythritol phosphate pathway in plastids, carving divergent routes to the synthesis of *IPP* and *DMAPP*. Farnesyl diphosphate synthase (*FPS*) then emerges as a pivotal catalyst, uniting *IPP* and *DMAPP* to yield *FPP*, the precursor to a plethora of sesquiterpenes produced by plants (Azerad, 2016).

The investigation delves further, uncovering the key players that sculpt centellosides. Squalene synthase (*SQS*), an influential microsomal polypeptide enzyme, guides the condensation of *FPP* molecules, yielding squalene—an essential milestone in the biosynthesis of steroids and triterpenoids (Kim, 2003; Moyano, 2007). The 2,3-oxidosqualene cyclases (*OSCs*) family steps onto the stage, wielding their transformative power to cyclize 2,3-oxidosqualene into protosteryl cation or dammarenyl cation, initiating a branching point pivotal to the biosynthesis of sterols and triterpenoid saponins. This step commences with the breaking of the oxirane ring in acidic media, and after various rearrangements, two distinct conformations are obtained. In plants, the protosteryl cation, which serves as the precursor for sterols, is transformed by the enzyme cycloartenol synthase (*CYS*) into cycloartenol (Haralampidis et al., 2002; Azerad, 2016).

As we navigate deeper, the centelloside saga continues with intricate transformations from dammarenyl cation involving enzymes such as dammarendiol synthase, lupeol synthase, and α/β-amyrin synthases (α/βASs). These sculptors of molecular structure guide the metamorphosis of cationic intermediates into diverse compounds, granting rise to the rich tapestry of centelloside precursors: α/β-amyrin (Kim et al., 2005).

Yet, the pathway's endgame remains shrouded in mystery, as the final stages of centelloside biosynthesis continue to elude our grasp. Enzymes like cytochrome P450-dependent monooxygenases (CYPs) and UDP-glucosyltransferases (UGTs) perform the final choreography, modifying products with oxidation, hydroxylation, glycosylation, and other elusive transformations (Augustin, 2011). Last studies identified 3 UGTs (*UGT73AD1*, *UGT73AH1* and *CaUGT*) that acts in the conversion of asiatic/madecassic acid to asiaticosid/madecassoside (de Costa et al., 2017; Kim et al., 2017; Han et al., 2022).

Despite the enigma that lingers, one truth remains: centelloside metabolic pathway is a testament to nature's symphony of complex biochemical interplay, orchestrating the creation of compounds that echo both tradition and innovation in the world of therapeutic possibilities (Mitra et al., 2021).

As we venture further, it becomes evident that the metabolic pathway of centellosides is influenced not only by genetic factors but also by environmental cues. Factors such as light, temperature, stress factors and soil composition are known to impact the expression of key biosynthetic genes and the activity of relevant enzymes, thereby shaping the production of centellosides within the plant (Müller et al., 2013; Nouri Nav et al., 2021; Pipatsitee et al., 2023).

Moreover, the unraveling of this pathway provides a window into the potential biotechnological applications that can be harnessed for optimized centelloside production.

Hairy roots: mechanism and influence of *rol* genes

In the fascinating realm of biotechnology, the concept of plant biotechnological production systems (PBPSs) emerges as a groundbreaking paradigm that harnesses the inherent potential of plants to serve as efficient and sustainable factories to produce valuable compounds. Just as industrial facilities churn out products on a mass scale, these living green production powerhouses enable us to cultivate a diverse array of bioactive molecules, pharmaceuticals, industrial chemicals, and other high-demand substances.

PBPSs, also known as plant biofactories, take advantage of the intricate cellular machinery within plants to synthesize complex molecules with precision, offering an eco-friendly and renewable alternative to conventional manufacturing methods. By leveraging the natural metabolic pathways and genetic potential of plants, scientists have unlocked the ability to engineer these living systems for enhanced production, scalability, and cost-effectiveness. The most frequently used plant cell and organ systems are cell suspension, immobilized cells, shoot and hairy root cultures (Hidalgo et al., 2018).

Hairy roots (HRs) are a captivating and enigmatic facet that has garnered considerable attention in recent years. Hairy roots, though visually distinct from conventional plant roots, hold immense promise as a powerful tool for scientific research (Gerszberg and Wiktorek-Smagur, 2022).

HRs are often known to grow much faster than plant cell cultures and do not require phytohormones in the medium for growth. They often possess a greater biosynthetic capacity to produce secondary metabolite in comparison to their mother plants. The hairy roots lack geotropism, show lateral branching and are genetically more stable. It can also serve the purpose of being a model system for the study of plant metabolism and physiology (Hussain et al., 2022).

The enigmatic realm of plant-microbe interactions reveals the intriguing genesis of HRs. These distinctive structures emerge as a consequence of plant infection by specific soil-dwelling pathogenic bacteria, known as *Rhizobium rhizogenes* strains. Classified based on the type of opine they produce, these strains encompass agropine-, cucumopine-, and mannopine-type categories. Within the mosaic of this bacterial diversity, Ri plasmids of certain strains, such as the agropine-type, encompass dual T-DNA sequences referred to as the left (TL-DNA) and right (TR-DNA) regions. A remarkable characteristic of these regions is their independent transfer potential (Amani et al., 2020).

Among these genetic regions, the TR-DNA segment houses the genes responsible for auxin biosynthesis (aux genes) and opine production. These factors act in concert to

provide essential support for the intricate development of hairy roots (Pavlova et al., 2014). Conversely, the TL-DNA region accommodates the influential rol genes, a quartet composed of *rolA*, *rolB*, *rolC*, and *rolD*. These genes wield not only a pivotal role in shaping the morphological symptoms of infections but also exert a profound impact on the emergence of novel phenotypic traits. This profound influence on both form and function underscores the intricate interplay between the plant and its bacterial companions (Gerszberg and Wiktorek-Smagur, 2022) (Figure 5).
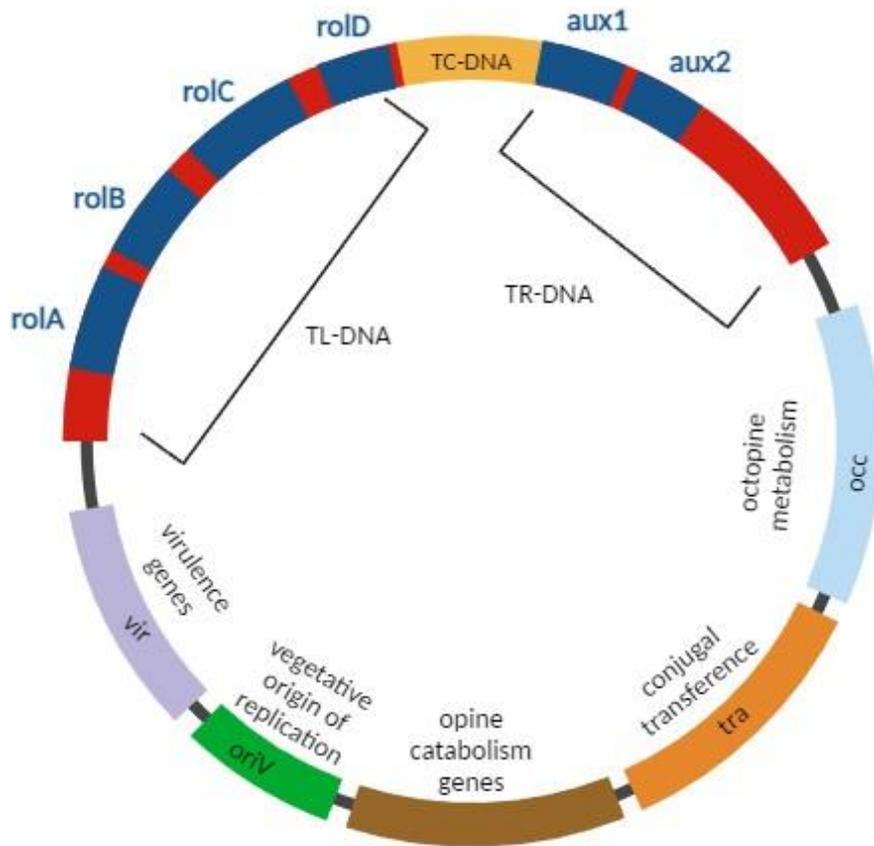
Figure 5. Structure agropine type Ri-plasmid resalting *rol* (*rolA*, *rolB*, *rolC* and *rolD*) and *aux* (*aux1* and *aux2*) genes from T-DNA.

Each *rol* gene possesses distinct characteristics and mechanisms that play a crucial role. The *rolA* gene, a constant presence in all known Ri-plasmids, exhibits sizes ranging from 281 to 302 bp (Pavlova et al., 2014). The resulting rolA protein, a compact 11.4 kDa molecule, is situated in the plasma membrane. Structurally akin to the DNA-binding domain of papillomavirus E2, it interacts with nucleic acids, orchestrating gene expression regulation in plants (Vilaine et al., 2007). The far-reaching influence of the *rolA* gene is evident through its capacity to amplify compound production in diverse plant

species, including artemisinin in *Artemisia dubia*, alkaloids in *Nicotiana tabacum*, and flavonoids in *Artemisia carvifolia* (Palazon et al., 1997; Amanullah et al., 2016; Khan and Dilshad, 2023).

The *rolB* open reading frames (ORFs), spanning 762 to 837 bp, code for a 259-amino acid protein of approximately 30 kDa (Dilshad et al., 2021). This protein catalyzes the hydrolysis of indole glycosides, releasing active auxins from their inactive conjugates (Estruch et al., 1991b). The versatile *rolB* gene takes center stage as a pivotal oncogene driving autonomous compound production in various plant species. It fosters anthraquinone production in *Rubia cardifolia* callus cultures (Shkryl et al., 2008), resveratrol synthesis in *Vitis amurensis* cells (Kiselev et al., 2007), and isoflavonoid generation in *Maackia amurensis* cells (Grishchenko et al., 2016).

In contrast, the *rolC* oncogene spans a sequence of 537 to 543 bp (Rangslang et al., 2018) and stands as one of the most conserved among the *rol* genes (Britton et al., 2008). The resulting *rolC* protein, comprising 178-180 amino acids and localized in the cytosol, exhibits beta-glucosidase activity (Dilshad et al., 2021). This activity, in turn, frees active cytokinins from their conjugates. The *rolC* oncogene exerts a substantial influence over the production of tropane alkaloids in transformed *Atropa belladonna* roots (Bonhomme et al., 2000) and indole alkaloids in *Catharanthus roseus* cell cultures (Palazon et al., 1998). Moreover, it augments phenol production, including ginsenosides in *Panax ginseng* hairy roots (Bulgakov et al., 1998) and resveratrol in *Vitis amurensis* cell cultures (Dubrovina et al., 2010).

Less explored yet equally intriguing, the *rolD* gene, present only in specific *R. rhizogenes* strains, encodes a 344-amino acid protein functioning as an ornithine cyclodeaminase (OCD), converting ornithine to proline (Trovato et al., 2001; Dilshad et al., 2021). Despite its relative novelty, the *rolD* gene's potential impact on specialized metabolite production remains largely unexplored.

The concurrent expression of multiple *rol* genes demonstrates enhanced effects on specialized metabolism. For instance, co-expression of *rolA*, *rolB*, and *rolC* genes led to a significant increase in anthraquinone production in *R. cardifolia* callus cultures (Shkryl et al., 2008). Similarly, rolABC-transgenic plants displayed heightened production of phenolics and flavonoids in *Lactuca sativa* (Ismail et al., 2016) and *Artemisia dubia* (Kiani et al., 2019). Notably, the acclaimed antimalarial compound artemisinin experienced elevated production in transgenic *Artemisia annua* and *Artemisia dubia* plants (Kiani et al., 2016). Another similar case is observed in HRs of *Panax quinquefolium* showing increased production of ginsenosides (Mathur et al., 2010). This underscores the

influential synergy achieved by harnessing the cooperative effects of multiple rol genes in advancing plant biotechnology.

To get more and detailed information about *rol* genes, see recommended Reviews in Box 1.

Box 1

- **Alcalde, MA. Palazon, J., Bonfill, M. & Hidalgo-Martinez, D. (2023). Exploring the key aspects of rol genes: An overview. Phytochemistry Reviews. Under review: PHYT-D-23-00286.**
- Sarkar, S., Ghosh, I., Roychowdhury, D., & Jha, S. (2018). The effects of rol genes of *Agrobacterium rhizogenes* on morphogenesis and secondary metabolite accumulation in medicinal plants. Biotechnological approaches for medicinal and aromatic plants: Conservation, genetic improvement and utilization, 27-51.
- Rangslang, R. K., Liu, Z., Lütken, H., & Favero, B. T. (2018). *Agrobacterium spp.* genes and ORFs: Mechanisms and applications in plant science. Ciência e Agrotecnologia, 42, 453-463.
- Mauro, M. L., Costantino, P., & Bettini, P. P. (2017). The never ending story of *rol* genes: a century after. Plant Cell, Tissue and Organ Culture (PCTOC), 131, 201-212.
- Khan, S., Saema, S., Banerjee, S., & ur Rahman, L. (2016). Role of *rol* Genes: Potential Route to Manipulate Plants for Genetic Improvement. Plant Tissue Culture: Propagation, Conservation and Crop Improvement, 419-446.
- Pavlova, O. A., Matveyeva, T. V., & Lutova, L. A. (2014). *rol*-Genes of *Agrobacterium rhizogenes*. Russian Journal of Genetics: Applied Research, 4, 137-145.
- Bulgakov, V. P. (2008). Functions of *rol* genes in plant secondary metabolism. Biotechnology advances, 26(4), 318-324.

In the intricate dance of plant-microbe interactions, *R. rhizogenes* takes center stage by orchestrating the transfer of select genes from its T-DNA into the host plant's genome during infection (Nemoto et al., 2009). In controlled laboratory settings, this microbial choreography gives rise to the emergence of roots at sites of bacterial invasion, which can be meticulously excised and nurtured in separate cultures without the need for external plant growth regulators (Bensaddek et al., 2008).

The profound ramifications of hairy roots traverse diverse domains. Within agriculture's realm, these enigmatic roots serve as a gateway to deciphering plant responses to stress, nutrient absorption dynamics, and interactions with pathogens. Beyond their investigative role, the robust growth attributes of hairy roots render them promising candidates for phytoremediation endeavors, wherein they wield the power to cleanse contaminated soils and waters. Venturing further, the realm of hairy roots unfolds a realm of possibilities for the sustainable production of prized entities encompassing pharmaceuticals, nutraceuticals, and bioactive compounds (Sharma et al., 2013; Roy, 2021; Gerszberg and Wiktorek-Smagur, 2022).

Intriguingly, hairy roots stand as prolific producers of an array of chemical constituents classified as secondary metabolites (Alcalde et al., 2022). This repertoire spans the likes of alkaloids, diterpenes, flavonoids, furan hormones, coumarins, naphthoquinones, and stilbenes (Tiwari and Rana, 2015). These compounds, instrumental in the growth and development of plants, also hold sway as pivotal factors shaping plant resilience in the face of both biotic and abiotic stressors (Guerriero et al., 2018). Moreover, the secondary metabolites' multifaceted utility extends into the realms of pharmaceuticals, agrochemicals, cosmetics, and food industries, rendering them valuable commodities with applications encompassing antibacterial, antifungal, antiviral, and anticancer effects, along with serving as fragrances, dyes, flavors, and more (Morey and Peebles, 2022; Sathasivam et al., 2022).

Delving into the specific case of *C. asiatica*, hairy root culture technique stands out as a remarkable solution to mitigate the overexploitation of this valuable resource. The optimal tissue for infection and the subsequent formation of transformed roots has been discerned to be the leaves (Gandi and Giri, 2012; Baek et al., 2022), culminating in a succinct depiction of this process in Figure 6. Notably, this knowledge has been harnessed to augment the production of asiaticoside (Ruslan et al., 2012; Zahanis et al., 2014). A significant breakthrough emerged through the overexpression of hairy roots, resulting in elevated levels of triterpenes, including asiaticoside and madecassoside, facilitated by *R. rhizogenes* R1000 in *C. asiatica* (Kim et al., 2010; Mandal et al., 2023).

2
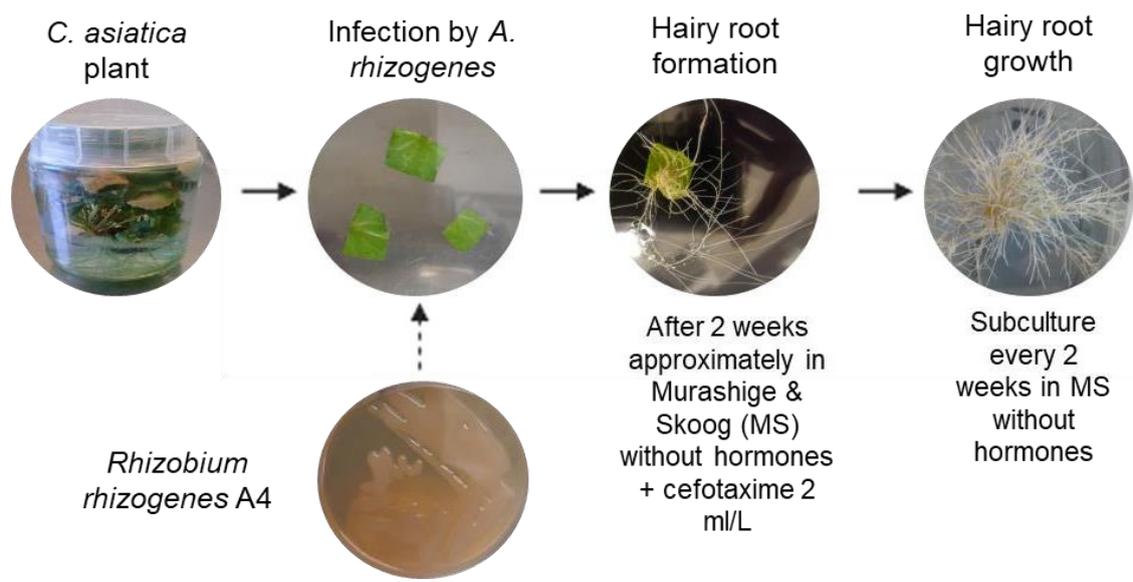
Figure 6. Overview of *Centella asiatica* hairy root establishment process via *Rhizobium rhizogenes* A4

As we delve into the intricacies of hairy roots, we unearth their potential to accelerate the discovery of novel genetic traits, elucidate complex biochemical pathways, and revolutionize the way we approach both basic plant science and industrial applications. Through a combination of new approaches, that can be compared as ladders in a game of "snakes and ladders", can improve secondary metabolite production. They are related to molecular biology, genetics, and innovative cultivation techniques, scientists are harnessing the potential of these extraordinary roots to drive innovation, address challenges, and pave the way for a greener, more sustainable future.

2

Exploring omics and machine learning in the context of hairy roots

In the dynamic realm of biological sciences, a revolution has taken place through the advent of "omics" technologies, offering a multifaceted lens to delve into the intricate web of molecular interactions that orchestrate the functions of living organisms. The prefix "omics" encompasses a range of disciplines such as genomics, transcriptomics, proteomics, metabolomics, and more, collectively facilitating a holistic understanding of cellular and organismal processes (Vailati-Riboni et al., 2017; Subramanian et al., 2020). By adopting high-throughput approaches and data-driven methodologies, omics technologies have transformed our ability to decipher the underlying mechanisms governing biological systems at various levels of complexity. This introduction embarks on an exploration of the diverse omics disciplines, their significance, and their interconnectedness, unveiling how they collectively unravel the intricate tapestry of life's molecular intricacies (Misra et al., 2019).

Proteomics illustrates the structures, functions and modifications of proteins, and the protein–protein interactions taking place under *in vitro* and *in vivo* conditions. The authentication of post-translational modifications, such as protein phosphorylation, protein acetylation, protein glycosylation and proteolysis, can also be performed (Pandita et al., 2021). Notably, within this framework, proteomics emerges as a potent and yet under-explored tool for elucidating secondary metabolites in hairy root cultures. Noteworthy strides have been taken in applying the proteomic approach to hairy root cultures, with initial strides observed in species like *Panax ginseng*, *Papaver somniferum, Salvia miltiorrhiza*, and *Alisma orientale* (Contreras et al., 2019; Rency et al., 2019; Rong et al., 2021).

The transcriptomic studies help in the characterization of key characteristics involved in the formation of secondary metabolites and in probing pharmaceutically important mechanisms at the molecular level. Respect to transcriptomic analysis of hairy root cultures has been done in several plants including *P. ginseng* (ginsenoside), *C. roseus* (indole alkaloids), *Medicago truncatula* (anthocyanin), *S. miltiorrhiza* (tanshinones), *Centella asiatica* (centellosides) (Sangwan et al., 2013; Prasad et al., 2019).

The main reason to study a lot of data by different omics is the possibility to determinate biomarkers (Figure 7). They are substances, structures, or processes that can be measured in any living being and can provide surrogate information about the presence of a disease, condition, or trait (Reel et al., 2021). It helps us to understand and analyze the underlying characteristics and complexities of biological systems.

2
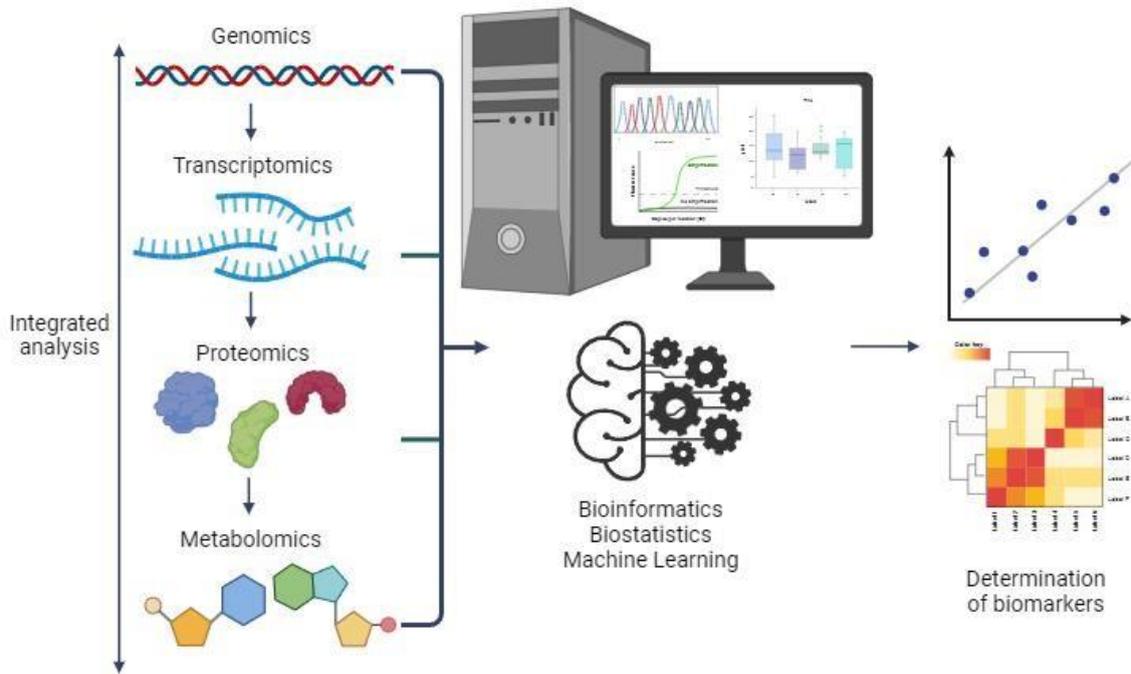
Figure 7. Exploring and validating biomarkers through a multi-omics approach

To determine the most important biomarkers we need to use new statistical analysis such as Machine Learning (ML). ML is a set of computational approaches that can find these predictive patterns in data which plays an increasingly important role in a particular condition. In various scientific and engineering domains.

ML encompasses two primary approaches, each with distinctive roles and applications. The first approach is known as supervised learning, which revolves around the utilization of labeled training data to train a model that can subsequently make predictions. This method is further categorized into two main problem types: regression, where the predicted variable assumes a numeric value, and classification, where the predicted variable is categorical in nature (Jiang et al., 2020).

The supervised learning process involves a sequence of three crucial steps. Firstly, the model is trained on a sample of input observations. Secondly, the model's performance is evaluated, followed by an iterative process of hyper-parameter tuning to refine its accuracy (Foster et al., 2014). Finally, the model is optimized for deployment in a production setting, enabling it to make predictions based on new input data.

In contrast, the second approach, known as unsupervised learning, operates on unlabelled data, focusing on extracting inherent patterns from input feature variables without a target/output variable. This technique serves various purposes, including

2

clustering, where data points are grouped based on similarities; anomaly detection, which identifies deviations from expected patterns; and dimensionality reduction, where complex data is simplified while retaining essential information, for example plant biology (Yan and Wang, 2022).

The different approaches use several models grouped in families (Table 1 summary).

Table 1. Diverse approaches in Machine Learning

Machine Learning approaches	Families	Main characteristics
Bayesian or probability-based	Bayesian network and naive beyond	Incorporation of prior knowledge or beliefs into the analysis process
Tree or information-based	Decision tree, random forest, and gradient boosting	Based on the concept of hierarchical structures and information gain
Linear or error-based	Linear, logistic and partial linear regression	Built upon the principles of linear relationships and error minimization
Instance or similarity-based	K nearest neighbour and Self-Organising Maps	It works comparing instances and measuring similarity
Support vector machines (SVM)	Linear and no-linear SVM	Depends on support vectors that define the optimal hyperplane
Neural network-based	Artificial neural network and deep learning	Ability to learn hierarchical representations

The Bayesian or probability-based family encompasses a wide range of techniques, including Bayesian networks and naive bayes methods, and probabilistic graphical models. A fundamental feature of Bayesian analysis is its incorporation of prior knowledge or beliefs into the analysis process. Prior probabilities are combined with observed data using Bayes' theorem to yield updated, or posterior, probabilities. This blending of prior information and current evidence empowers Bayesian methods to provide nuanced insights, especially when dealing with small datasets or in cases where prior expertise is available (Draper, 2012; van de Schoot et al., 2021).

The Tree or Information-Based family is rooted in the concept of hierarchical structures and information gain, this family of methods offers a structured way to organize and process complex data while uncovering meaningful insights. A key principle of these methods is their focus on information gain. When building a decision tree, algorithms seek to identify attributes that best separate data points into distinct classes or categories. The chosen attributes maximize the information gained at each step, allowing for efficient classification and prediction. As a result, decision trees not only provide actionable insights but also offer interpretability by showcasing the sequence of decisions leading to a particular outcome (Loh, 2008; Niazian and Niedbała, 2020).

The Tree or Information-Based Family encompasses various techniques, with one of the most prominent being the Classification and Regression Tree algorithm. In addition, Random Forests and Gradient Boosting are advanced methods that leverage the power of ensemble learning, combining multiple decision trees to enhance predictive accuracy (Kotsiantis, 2013; Hesami and Jones, 2020).

The Linear or Error-Based family is built upon the principles of linear relationships and error minimization. When constructing linear models, the goal is to find the line that best fits the observed data points, minimizing the discrepancies, or errors, between the predicted and actual values. This optimization process is typically achieved using techniques such as the method of least squares, which ensures that the model aligns as closely as possible with the underlying data distribution (Maulud and Mohsin Abdulazeez, 2020).

The Linear or Error-Based Family encompasses various techniques, including simple linear regression, multiple linear regression, and generalized linear models. Moreover, it serves as a foundation for more advanced methods like ridge regression and lasso regression, which address issues of multicollinearity and overfitting, respectively (Goswami et al., 2018).

The Instance or Similarity-Based is built upon the concept of comparing instances and measuring similarity, this family of methods offers a practical and adaptable framework for solving a wide array of problems. When dealing with new, unseen instances, these methods gauge the likeness between the instance of interest and those already present in the dataset. This often involves calculating distances or dissimilarities in a feature space. By identifying the most similar instances, these methods infer the likely class or category of the new instance, allowing for quick and context-aware decision-making (Zhang, 2016).

The Instance or Similarity-Based Family encompasses various techniques, including k-nearest neighbors, self-organizing maps, case-based reasoning, and collaborative filtering. These methods have applications ranging from recommendation systems and medical diagnosis to anomaly detection and image recognition (Moshou et al., 2006).

The appeal of these methods lies in their simplicity and adaptability. By basing decisions on local information and similarity metrics, instance-based approaches can capture intricate patterns that might be challenging for traditional model-based methods to discern. This makes them especially useful in scenarios where explicit model construction is difficult or where the data distribution is complex (Dong et al., 2019).

Support Vector Machines (SVMs) have emerged as a powerful and widely used technique for classification and regression tasks. Central to the concept of SVMs are the fundamental elements known as "support vectors". These vectors not only form the backbone of the SVM algorithm but also encapsulate the essence of the method's effectiveness. SVMs achieve this by identifying a hyperplane that maximizes the margin – the distance between the hyperplane and the nearest data points of each class. These crucial data points lying on the margins of the separation are referred to as support vectors (Somvanshi and Mishra, 2015).

Support vectors play a dual role in SVMs. First, they define the optimal hyperplane, as the margin is directly influenced by the distances from these vectors to the decision boundary. Second, they encapsulate the most informative and influential data points for the given classification or regression task. By focusing on these key instances, SVMs excel at generalizing and making accurate predictions even on unseen data (Hesami and Jones, 2021).

The concept of support vectors goes beyond binary classification scenarios. SVMs can be extended to handle multi-class classification and regression problems through techniques such as one-vs-one and one-vs-rest strategies. Additionally, SVMs can be adapted to work with non-linearly separable data by using kernel functions that transform the input space, effectively projecting the data into a higher-dimensional space where linear separation becomes feasible (Mayoraz and Alpaydm, 1999).

The Neural Network-Based is rooted in the inspiration drawn from the human brain's intricate network of neurons, this family of methods comprises artificial neural networks, deep learning architectures, and convolutional networks, collectively reshaping the boundaries of what machines can achieve (Shi et al., 2021). A defining characteristic of this family is its ability to learn hierarchical representations. Deep neural networks, which have multiple hidden layers, excel at capturing intricate features in data, progressively

abstracting information from raw inputs to high-level representations (Ma et al., 2014; Qi et al., 2019). This capacity is particularly advantageous for domains where complex structures or subtle patterns need to be uncovered. The Neural Network-Based Family encompasses various architectures, including feedforward neural networks, recurrent neural networks, and convolutional neural networks. Each architecture is tailored to excel in specific tasks, such as sequential data analysis, image recognition, and natural language processing (Kaya et al., 2019).

ML has driven a spur of recent innovations, and it is set to do the same in plant research. It is used to focus on the optimization for secondary metabolism production in different biotechnological platforms: growth modeling of *Medicago sativa* shoots, biomass estimation of *Daucus carota* cell cultures, model *in vitro* rhizogenesis and subsequent acclimatization of *Vitis vinifera* (Helmy et al., 2020; Dirk et al., 2021).

In the context of hairy root cultures, significant advancements have been made through the application of neural network models. Mehrotra et al. (2008) devised a feed-forward back-propagation neural network model aimed at predicting *in vitro* culture conditions conducive to robust hairy root growth. By incorporating key parameters such as inoculum size, fresh weight, density, culture temperature, pH, and time of inoculation as inputs, coupled with the final fresh weight of roots as the output parameter, their model exhibited the potential to forecast growth outcomes effectively.

Building upon this foundation, Prakash et al. (2010) introduced a comprehensive approach by developing both regression and feedforward neural network models. Their objective was to predict the optimal culture conditions for maximizing hairy root biomass yield. While both networks demonstrated promising predictive capabilities, the regression neural network emerged as notably more precise in its predictions.

Furthermore, neural networks have extended their influence into the modeling of bioreactors tailored for hairy root growth. In a study by Osama et al. (2013), a neural network model was formulated to encapsulate the complexities of hairy root growth within a nutrient mist reactor. Notable culture parameters, including factors such as inoculum size, initial packing density, media volume, initial sucrose concentration in the media, and culture duration, were harnessed as inputs. This model, constructed with a focus on predicting the final biomass of hairy roots in terms of dry weight, demonstrated its potential as a versatile tool for reactor modeling.

Expanding beyond traditional neural network approaches, a noteworthy fusion of techniques emerged in the realm of optimal culture condition prediction for maximum biomass yields in *Rauwolfia serpentina* hairy root cultures. This integration involved the

utilization of a hidden Markov model within a neural network-based combinatorial framework. This innovative approach showcased the multidisciplinary potential of neural networks in addressing intricate challenges, thereby contributing to the advancement of both hairy root cultures and predictive modeling techniques (Somvanshi and Mishra, 2015; Goswami et al., 2018).

As we delve deeper into this symbiotic relationship between machine learning and hairy root cultures, new frontiers emerge. From generating insights into growth patterns to predicting metabolite production and optimizing bioprocessing conditions, this dynamic pairing holds vast potential for accelerating biotechnological advancements and sustainable plant-based solutions. To get more and detailed information about machine learning in plant biotechnology, see recommended Reviews in Box 2.

Box 2

- ❖ Niazian, M., & Niedbała, G. (2020). Machine learning for plant breeding and biotechnology. Agriculture, 10(10), 436.
- ❖ Ma, C., Zhang, H. H., & Wang, X. (2014). Machine learning for big data analytics in plants. Trends in plant science, 19(12), 798-808.
- ❖ Osama, K., Mishra, B. N., & Somvanshi, P. (2015). Machine learning techniques in plant biology. PlantOmics: The Omics of plant science, 731-754.
- ❖ Kim, G. B., Kim, W. J., Kim, H. U., & Lee, S. Y. (2020). Machine learning applications in systems metabolic engineering. Current opinion in biotechnology, 64, 1-9.
- ❖ Reel, P. S., Reel, S., Pearson, E., Trucco, E., & Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: A review. Biotechnology Advances, 49, 107739.

Elicitation: influence on terpenoid metabolism in hairy roots as a biotechnological platform

Elicitation is a common method for inducing or boosting secondary metabolite production in plants as a survival mechanism (Figure 8 summary elicitation). This technique has been extensively employed in plant *in vitro* cultures to enhance their secondary metabolite production. Hence, we can aptly refer to it as a "ladder", guarding against "spinfall" like the naturally low production of specific bioactive components (Vasconsuelo and Boland, 2007; Halder et al., 2019).
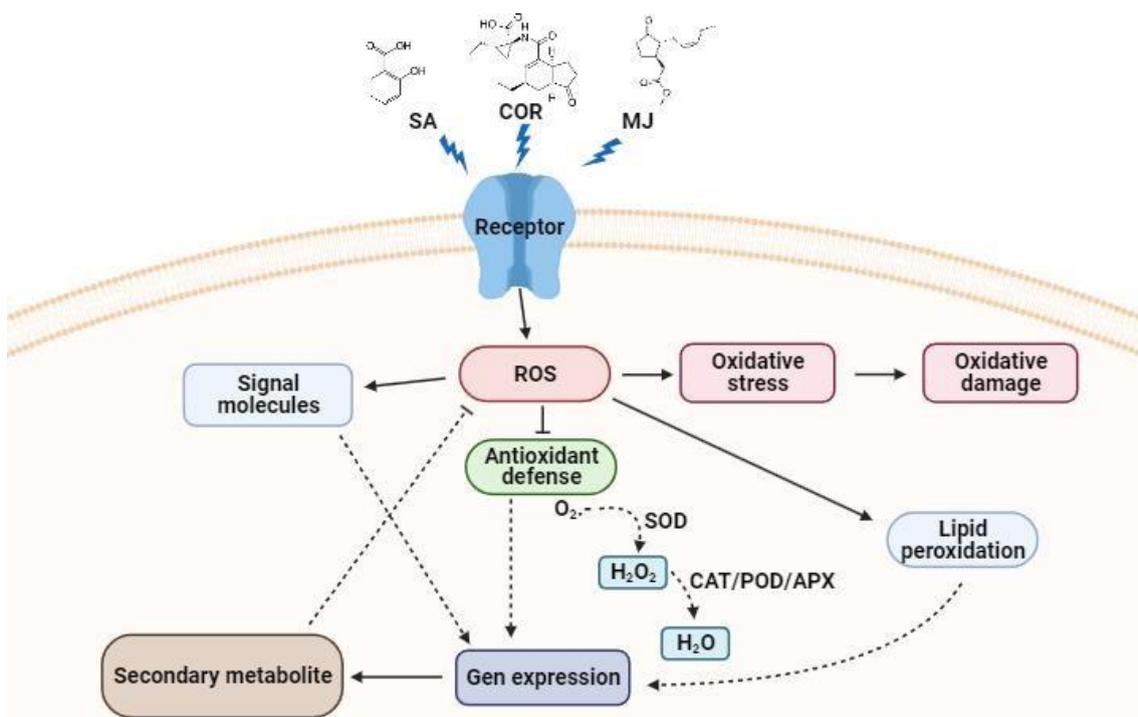
Figure 8. Overview of elicitation processes using various elicitors: Salicylic acid (SA), Coronatine (COR), and Methyl Jasmonate (MJ). ROS: Reactive Oxygen Species. SOD: Superoxide Dismutase. CAT: Catalase. POD: Peroxidase Dismutase. APX: Ascorbate Peroxidase.

The impact on elicited plant cells is delineated by the induction of defense or stress-related responses. A diverse array of compound classes, often lacking a shared chemical structure, participates in this process, with not all plants being equally responsive to all of them. Compounds that effectively trigger the production of specific metabolites in certain plants may remain inert in other species. Conversely, different plant species may exhibit responsiveness to the same eliciting agent. Furthermore, it is evident that elicitor specificity extends to the signal components activated during the elicitation process.

These observations suggest that plants possess the capacity to recognize various structurally distinct molecules as signals, potentially driven by the presence of dedicated receptors for each elicitor class (Vasconsuelo and Boland, 2007).

The inception of the elicitor signal transduction cascade heralds the stage of signal perception, a pivotal juncture in the orchestration of plant response. This response, as demonstrated through the activation of kinases, generation of reactive oxygen species, modulation of ion fluxes, and adjustments in cytoplasmic acidity, hinges on the discernment of diverse stimuli. The intricate web of these biochemical reactions underpins the plant's ability to sense and interpret a variety of environmental cues (Angelova et al., 2014).

In the elaborate choreography of this process, multiple sites for elicitor binding have been discerned. These potential receptor locales predominantly find their abode within the plant's plasma membranes, acting as sentinels poised to receive incoming signals. Amid the array of receptor classes explored, the transmembrane receptor-like protein kinases (RLKs) stand prominently, emerging as key contenders involved in pathogen perception. The wide diversity of plant RLKs, coupled with their remarkable prevalence within the Arabidopsis genome, positions them as potential participants in a wide spectrum of stimulus responses (Montesano et al., 2003).

Amid the RLK cohort, a particular standout is the extensively characterized flagellin receptor—a remarkable member of the leucine-rich repeat class. Through the meticulous dissection afforded by molecular techniques, the genetic intricacies of this receptor have been unveiled, contributing to our understanding of the delicate interplay between perception and response (Gómez-Gómez and Boller, 2000).

Another distinct category of elicitor receptors embraces plant R-proteins, notable for their capacity to recognize race-specific elicitors encoded by avirulence (*avr*) genes. This recognition mechanism serves as a sentinel for potential threats, spotlighting the dynamic interplay between plants and pathogens in their ongoing co-evolutionary struggle (Montesano et al., 2003).

Once the initial signal perception takes place, a fascinating array of intracellular signaling systems is set into motion. Among these, a diverse landscape of Guanosine Triphosphate (GTP)-binding proteins assumes a central role, orchestrating a multitude of cellular processes entwined with growth, hormonal signaling, development, and defense responses (Vasconsuelo and Boland, 2007).

Another crucial system emerges post-elicitation, marked by the activation of phospholipase C within plant cells. This event triggers a cascade leading to polyphosphoinositide turnover and the generation of vital second messengers, inositol trisphosphate and diacylglycerol. These messengers play a pivotal role in amplifying the elicitation-driven response (Pan et al., 2005).

Adding to the intricate web of intracellular interactions, the adenylyl cyclase/cyclic adenosine monophosphate (cAMP)/protein kinase A (PKA) pathway comes into play. Its involvement in the elicitation process has been implicated in both *Cupressus lusitanica* and *Arabidopsis* species, illuminating its significance across different plant contexts (Ebel and Mithöfer, 1998).

Calcium (Ca^{+2}) emerges as a ubiquitous signal in the plant world, serving as a conduit for the regulation of myriad cellular processes in response to various stimuli, including elicitation. This ion's role in translating external signals into intracellular responses is particularly noteworthy (Harmon et al., 2000).

Further into the intracellular landscape, the phosphatidylinositol 3-kinase type gains prominence, exerting its influence within the nucleus and nucleolus. This factor's interplay with mitogen-activated protein kinases (MAPKs) suggests its potential role in mediating specific aspects of the elicitation response. At the heart of these intricate interactions lie the MAPK cascades, pivotal components downstream of receptors and sensors. They serve as the bridge that translates external cues into intracellular responses across eukaryotes, often exerting their influence on targeted gene expression (Cardinale et al., 2000).

Once the mechanism of elicitation is understood, conducting this procedure requires consideration of several parameters. These include the concentration and selectivity of the inducer, treatment duration, growth stage of the culture, cell line, growth regulation, and nutrient composition. All these variables collectively influence the production of secondary metabolites (Singh et al., 2018).

Elicitors can be categorized according to their inherent characteristics as either abiotic or biotic elicitors (Table 1). Abiotic elicitors encompass a diverse array of chemical and physical stressors, ranging from light and UV radiation to the presence of heavy metal salts such as silver thiosulfate ($Ag_2S_2O_3$), silver nitrate ($AgNO_3$), cadmium chloride (CdCl2), cupric chloride ($CuCl_2$), cupric sulfate ($CuSO_4$), vanadyl sulfate ($VOSO_4$), nickelous sulfate ($NiSO_4$), and selenium. Additionally, abiotic elicitors include factors such as temperature shifts, osmotic stress induced by substances like mannitol, sorbitol,

sodium chloride, potassium chloride, cadmium chloride, and Polyvinylpyrrolidone, among others (Mishra et al., 2012).

Table 2. Classification of biotic and abiotic elicitors by their nature adapted from Rogowska and Szakiel (2021).

<u>**Biotic elicitors**</u>	<u>**Abiotic elicitors**</u>
Simple composition Polysaccharides, glycoproteins, chitosan, pectin, coronatine, etc.	**Salts of heavy metals** Silver nitrate, copper chloride, copper sulfate, etc.
Complex composition Crude extract of yeast, bacterial and fungal, etc.	**Osmotic stressors** Mannitol, sorbitol, potassium chloride, etc.
Intracellular signaling molecules Jasmonic acid, methyl jasmonate, salicylic acid, acetyl salicylic acid, etc.	**Gaseous substances** NO, ethylene, etc.
	Physical stressors Light and UV-radiation, temperature shift, drought and salinity.

Biotic elicitors encompass a spectrum of substances, which can either be crude extracts or partially purified products originating from pathogens (fungi, bacteria, yeast) or the host plant itself. These elicitors exhibit a diverse range of compositions. Some are meticulously defined, such as polysaccharides, glycoproteins, inactivated enzymes, purified chitosan (CS), pectin, chitin, alginate, curdlan, xanthan, and the notable coronatine (CORO). Others possess a more complex composition, such as yeast extract (YE) and fungal homogenate. Intriguingly, biotic elicitors also include key intracellular signaling molecules like jasmonic acid (JA), methyl jasmonate (MeJa), salicylic acid (SA), acetyl salicylic acid (ASA), each playing an essential role in orchestrating plant responses (Ramirez-Estrada et al., 2016). For further details regarding these elicitors, please refer to Box 3.

Box 3

- **Alcalde, M. A., Perez-Matas, E., Escrich, A., Cusido, R. M., Palazon, J., & Bonfill, M. (2022). Biotic elicitors in adventitious and hairy root cultures: A review from 2010 to 2022. Molecules, 27(16), 5253.**
- Bhaskar, R., Xavier, L. S. E., Udayakumaran, G., Kumar, D. S., Venkatesh, R., & Nagella, P. (2022). Biotic elicitors: A boon for the in-vitro production of plant secondary metabolites. Plant Cell, Tissue and Organ Culture (PCTOC), 149(1-2), 7-24.
- Halder, M., Sarkar, S., & Jha, S. (2019). Elicitation: A biotechnological tool for enhanced production of secondary metabolites in hairy root cultures. Engineering in life sciences, 19(12), 880-895.
- Ramirez-Estrada, K., Vidal-Limon, H., Hidalgo, D., Moyano, E., Golenioswki, M., Cusidó, R. M., & Palazon, J. (2016). Elicitation, an effective strategy for the biotechnological production of bioactive high-added value compounds in plant cell factories. Molecules, 21(2), 182.

The main biotic elicitors used in hairy roots are ASA, CS, CORO, JA, MeJa, Pectin, SA and YE (Alcalde et al., 2022).

SA is a small molecule with a crucial role in plant defense regulatory systems, often induced by numerous pathogens as part of the plant's defense response mechanism. During plant-pathogen interactions, there is a swift accumulation of SA at the infection site, triggering a hypersensitive response. Subsequently, this signal propagates to other parts of the plant, inducing various defense responses, including the production of secondary metabolites (Ramirez-Estrada et al., 2016).

MeJa is a naturally occurring compound in plants, commonly elicited in response to insect attacks, rendering it a hormone associated with biotic stress response (Singh et al., 2018). MeJa represents the methyl ester form of jasmonic acid, a pivotal compound in signal transduction that governs plant defense responses and culminates in heightened secondary metabolite production.

CORO is a phytotoxin produced by *Pseudomonas syringae*. Structurally, it is a polyketide derived from coronafacic acid and serves as a more stable analog of MeJa, known as coronamic acid (Singh et al., 2018).

Within hairy root cultures from 2010 to 2022, MeJa emerges as the preeminent elicitor for augmenting phenol production, while JA demonstrates superior efficacy in enhancing terpene yields, and ASA exhibits potency in stimulating alkaloid synthesis. Remarkably, the pinnacle of phenol production in hairy roots, yielding values of 123.6, 80, and 75.65

mg/g DW respectively, materialized with the application of 100 μM MeJa (Alcalde et al., 2022).

Notable instances are observed in *Glycyrrhiza inflata* hairy roots, which, when elicited by MeJa, achieve a production of 109 μg of glycyrrhizin per gram of dry weight on the fifth day of elicitation (Wongwicha et al., 2011). Similarly, in *Rehmannia glutinosa* hairy roots, elicitation with 200 μM SA led to an augmentation in the production of iridoids (catalpol, harpagide) and phenylethanoids (verbascoside and isoverbascoside) (Piątczak et al., 2016). Additionally, Vaccaro et al. (2017) demonstrated the significant enhancement of bioactive abietane diterpenes in *Salvia sclarea* hairy roots through transcriptional reprogramming induced by CORO.

Noteworthy advancements have been observed in the elicitation techniques applied to *C. asiatica*, notably leading to the augmented biosynthesis of triterpenoids, particularly centellosides in cell cultures and hairy root cultures (Gallego et al., 2014; Baek et al., 2020; Ganie et al., 2022).

For instance, the impact of MeJa (at a concentration of 50 μM) on the biosynthesis of asiaticoside in cell cultures of *C. asiatica* (Ruslan et al., 2012). This investigation highlighted the capability of methyl jasmonate to stimulate asiaticoside production, showcasing its potential as an elicitor for enhancing the bioactive compound's yield.

Similarly, the augmentation of asiaticoside content was also prompted by the introduction of 100 μM of SA at 10 days of post-elicitation in *C. asiatica* cell cultures (Loc and Giang, 2012).

Furthermore, in a separate study, MeJa was employed to elicit enhanced asiaticoside synthesis from hairy root cultures of *C. asiatica* (L.) Urban, leading to an amplified production of this valuable compound. The elicitation of hairy root cultures resulted in a significant increase in asiaticoside content, reaching an impressive level of 7.12 mg/g dry weight (Kim et al., 2007).

Metabolic engineering in hairy roots as plant biofactories

In the intricate world of plant biology, the concept of metabolic engineering emerges as a transformative endeavor that merges the elegance of nature's processes with the ingenuity of human innovation. This cutting-edge technology can work as a "ladder" in a game of "snakes and ladders". At its core, metabolic engineering aims to harness the inherent metabolic pathways within plants, directing them towards the synthesis of novel compounds or the enhancement of existing ones. This dynamic field lies at the intersection of biology, genetics, biochemistry, and engineering, offering a canvas upon which scientists can paint extraordinary possibilities (Nielsen, 2001).

Plants, endowed with an intricate web of biochemical reactions, produce a myriad of compounds essential for their growth, survival, and interaction with the environment. Through metabolic engineering, researchers seek to unveil the molecular intricacies underlying these processes, altering them with precision to yield desired outcomes. Whether it's fortifying nutritional content, amplifying the production of pharmaceutical compounds, or bolstering resistance against environmental stressors, metabolic engineering holds the key to unlocking the full potential of plants (Figure 9) (Woolston et al., 2013).
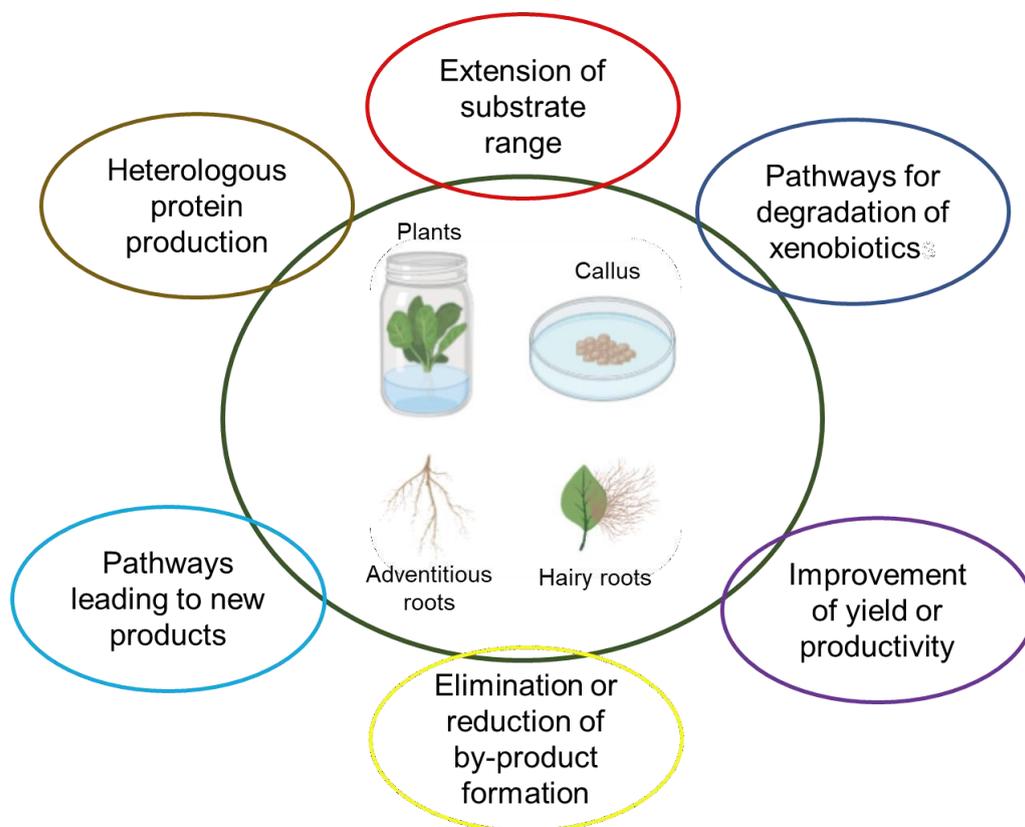
Figure 9. Primary approaches from principal plant biotechnological platforms for metabolic engineering.

The toolbox of a metabolic engineer encompasses a diverse array of tools and strategies. Genetic manipulation, molecular biology techniques, synthetic biology constructs, and computational modeling converge to rewire metabolic pathways, enabling plants to synthesize compounds that may not naturally be present in their repertoire. As technology evolves, the scope of metabolic engineering broadens, offering tantalizing prospects for sustainable agriculture, biopharmaceutical production, biofuel generation, and environmental remediation (Kim et al., 2020).

There are many categories or approaches in plant metabolic engineer. For instance, heterologous protein production that refers to the process of introducing and expressing a gene that encodes a protein from one organism (the donor or source organism) into a different host organism, which may belong to a different species or even a different biological kingdom. The host organism is chosen based on its capacity to efficiently produce and accumulate the desired protein, which may have various applications such as industrial enzymes, therapeutic proteins, vaccines, or research tools. The best plant platforms for metabolic engineer are *Arabidopsis thaliana*, *Nicotiana benthamiana*, *Physcomitrella patens*, *Brassica oleracea*, etc. (Ghag et al., 2021).

Additionally, other approach is how the pathways leading to new products. It refers about to the intricate series of biochemical reactions and enzymatic processes that enable the synthesis of novel compounds within plant cells. These pathways involve the conversion of precursor molecules into complex end products through a sequence of intermediate steps, often guided by specific enzymes and regulated by genetic and environmental factors (Verpoorte et al., 1999).

These pathways can give rise to a diverse array of compounds, including secondary metabolites, specialized chemicals, bioactive molecules, and even therapeutic compounds. Examples range from plant-derived pharmaceuticals and flavors to pigments and fragrances. The exploration and manipulation of these pathways offer the potential to harness nature's intricate biochemistry for applications ranging from biotechnology and agriculture to medicine and environmental management (Verpoorte et al., 2002).

Engineering of cellular physiology for process improvement in plants that deliberate modification and optimization of the metabolic and physiological functions within plant cells to enhance specific processes or outcomes. This involves the strategic manipulation of cellular pathways, gene expression patterns, and biochemical reactions to achieve improved performance, efficiency, or production of desired compounds. In this context, cellular physiology encompasses a wide range of biological processes, including

metabolism, growth, stress responses, and the synthesis of valuable compounds (Stephanopoulos, 1999).

Furthermore, the elimination or reduction of by-product formation in plants that is related to the intentional effort to mitigate or eliminate the production of unintended or undesired compounds that arise as side products during metabolic processes within plant cells. By employing biotechnological strategies, genetic manipulation, and metabolic engineering techniques, researchers aim to reconfigure biochemical pathways to minimize the synthesis of these by-products while enhancing the desired primary products (Nielsen, 2001).

Yet, realizing successful metabolic engineering within hairy root cultures demands a comprehensive grasp of the biosynthetic pathways of target molecules, the genes encoding associated enzymes, and their intricate regulatory mechanisms (Bagal et al., 2023).

This approach has proven highly effective in enhancing alkaloid biosynthesis in *Catharanthus roseus*, as well as in various HR cultures of solanaceous plants (Moyano et al., 2003; Palazón et al., 2003). It has also yielded impressive results in the production of significant biomolecules. Examples include the secretion of embryonic alkaline phosphatase in humans (Gaume et al., 2003), the fusion proteins of GFP and ricin toxin B from tobacco HR cultures (Medina-Bolivar et al., 2007), the accumulation of poly-3-hydroxybutyrate from HR cultures of sugar beet (Menzel et al., 2003), and the generation of solanoside glycoside from HR cultures of *Solanum khasianum* (Putalun et al., 2003) , enhanced production of terpenoid indole alkaloids in C*atharanthus roseus* HRs by overexpression of tryptophan decarboxylase, 1-deoxy-D-xylulose 5-phosphate synthase and octadecanoid-responsive *Catharanthus* AP2/ERF domain genes related from the terpenoid indole alkaloids biosynthesis pathway (Zhao et al., 2013).

Biotransformation, on the other hand, involves the conversion of naturally occurring phytochemicals into other biochemicals by inducing structural modifications in the original metabolite. This process harnesses the biological systems of plants to create novel and more valuable biochemicals that are better suited for human economic utilization. Biotransformation leads to the development of second-generation pharmaceuticals characterized by improved pharmacokinetics, enhanced solubility within biological systems, and reduced toxicity. As affirmed by the comprehensive review by Banerjee et al. (2012), HR cultures have emerged as highly advantageous platforms for facilitating this transformative process. These cultures offer advantages such as cost-effectiveness, genetic stability, and remarkable potential (Gantait and Mukherjee, 2021).

One prevailing strategy in the realm of terpenoid metabolic engineering is heterologous expression or the introduction of transcription factors that is involved in holistic approachs (Capell and Christou, 2004). Transcription factors play a pivotal role in orchestrating the intricate dance of gene expression within living organisms. These regulatory proteins possess the remarkable ability to modulate the activity of specific genes, dictating when and to what extent their genetic information is transcribed into functional RNA molecules. By binding to DNA sequences in the vicinity of target genes, transcription factors exert influence over a wide array of biological processes, ranging from development and growth to immune response and cellular adaptation (Yusuf et al., 2012).

Typically, plant transcription factors exhibit a characteristic composition, encompassing a DNA-binding region, an oligomerization site, a transcription regulation domain, and a nuclear localization signal. However, some transcription factors may lack either a transcription regulation domain or a specific DNA-binding region, adding to the intricate diversity of their structures (Schwechheimer and Bevan, 1998).

Transcription factors are categorized based on their structural attributes, often involving the division of families according to the number and arrangement of conserved residues within their most analogous domains. For instance, zinc finger-containing factors fall into distinct classes—C2H2, C3H, C2C2 (GATA finger), C3HC4 (RING finger), and C2HC5 (LIM finger)—dictated by the arrangement and abundance of cysteine and histidine residues. Alternatively, transcription factors belonging to the same family are grouped by referencing domains that extend beyond the most universally conserved region. As an illustrative case, homeodomain factors are clustered into five groups: homeodomain zipper, homeodomain finger, GLABRA2, ELK-homeodomain, and twin-homeodomain factors. Initially presumed to solely participate in sequence-specific DNA interactions, the conserved regions underpinning nomenclature have revealed unexpected versatility. Remarkably, excluding high mobility group finger, homeodomain finger, and *A. thaliana* HAT homeobox zipper factors, each plant transcription factor incorporates a sole variety of combined DNA-binding/oligomerization domain, albeit it might appear in multiple copies (Liu et al., 1999). Furthermore, there is databases about these genome sequences that stick out *Arabidopsis thaliana*, *Solanum tuberosum, Gossypium hirsutum*, *Helianthus annuus, Hordeum vulgare*, *Solanum tuberosum and Medicago truncatula* (Guo et al., 2008).

This approach aligns with the fact that the biosynthesis of specialized terpenoids is often tissue- and developmental stage-specific (Nagegowda and Gupta, 2020). The nuanced spatial and temporal formation of these specialized terpenoids is predominantly

governed at the transcriptional level through trans-acting factors, commonly known as transcription factors. These pivotal proteins intricately manage gene expression by binding to DNA regulatory sequences, referred to as enhancers and silencers, which respectively boost or inhibit the transcription of target genes (Bagal et al., 2023). Hairy root cultures have proven valuable platforms for engineering the production of a spectrum of specialized terpenoids. In the case of P. ginseng, the overexpression of distinct triterpenoid pathway genes like HMGR or SQS in adventitious root cultures yielded a remarkable 2–3-fold increase in total ginsenosides (Nagegowda and Gupta, 2020). By concurrently expressing HMGR and geranylgeranyl diphosphate synthase in *Salvia miltiorrhiza* hairy roots, the synthesis of tanshinone reached to impressive 2.727 mg/g dw (Kai et al., 2011). For *C. roseus* hairy roots the induction expression of a transcription factor ORCA3 and elicitation with JA increased terpenoid indol alkaloid metabolites and transcripts of pathway genes (Peebles et al., 2009). For further details regarding about metabolic engineering, please refer to Box 4.

> ### **Box 4**
>
> ❖ **Capell, T., & Christou, P. (2004). Progress in plant metabolic engineering. Current opinion in biotechnology, 15(2), 148-154.**
>
> ❖ Zhou, M. L., Zhu, X. M., Shao, J. R., Tang, Y. X., & Wu, Y. M. (2011). Production and metabolic engineering of bioactive substances in plant hairy root culture. Applied microbiology and biotechnology, 90, 1229-1239.
>
> ❖ Woolston, B. M., Edgar, S., & Stephanopoulos, G. (2013). Metabolic engineering: past and future. Annual review of chemical and biomolecular engineering, 4, 259-288.
>
> ❖ Volk, M. J., Tran, V. G., Tan, S. I., Mishra, S., Fatma, Z., Boob, A. & Zhao, H. (2022). Metabolic engineering: methodologies and applications. Chemical reviews, 123(9), 5521-5570.
>
> ❖ Bagal, D., Chowdhary, A. A., Mehrotra, S., Mishra, S., Rathore, S., & Srivastava, V. (2023). Metabolic engineering in hairy roots: An outlook on production of plant secondary metabolites. Plant Physiology and Biochemistry, 107847.

Objectives

In the current context of climate change, with the need to save water and preserve land for food production and improve the health of the world population regardless of their habitat of origin, a new challenge arises: the **growing demand** for products natural, mainly of plant origin, to cover the **growing demand** for these compounds to meet the existing needs for them, for the pharmaceutical, cosmetic and food industries in a biosustainable way. An example of this is the production of centellosides, triterpenes with important biological activities obtained from *Centella asiatica* plant extracts. In this pipeline, Plant Biofactories based on cell and organ cultures of *C. asiatica* may represent an eco-friendly alternative to meet the needs of these compounds at an industrial level.

As mentioned in the outline, this PhD thesis has as its main objective the study and improvement of centelloside production in hairy root cultures and deepen knowledge in functions of *rol* genes from *Rhizobium rhizogenes* on the morphology and productive capacity of the hairy roots.

Regarding that, we will try to answer 4 important open questions according to our specific goals that are schematized in Fig. 10:

Open question:	Objective:
How *Rhizobium rhizogenes* T-DNA genes modulates root growth, hormonal profiles and centelloside production using machine learning as a potent tool for classifying hairy roots lines, either by the level of gene expression or hormonal profile.	To understand the role of T-DNA genes in production of centellosides performing a combined analysis of the expression levels of *rol* and *aux* genes, morphological traits, centelloside content, and hormonal profiling in *Centella asiatica* hairy roots using machine learning; as well as it enables to determinate productive lines (Chapter 1).
The recent discovery that the morphological differences, the centellosides production, hormonal levels and the expression of the *rol genes* of *Rhizobium rhizogenes* among the hairy root lines, could be due to the integration of different copies of the T-DNA of *R. rhizogenes* into the plant genome and proteomic alterations generated by this	To investigate the molecular mechanisms underlying hairy root development and centelloside production in order to enhance *C. asiatica*'s biofactories through an analysis of hairy root protein profiles and number of *rol* gene copies (Chapter 2).

integration when compared with untransformed roots.	
Strategies such as elicitation that induces changes in specialized metabolism can be added to the new knowledge generated, which could be used to understand the transcriptomic control of the centellosides pathway, revealing changes that cannot be visualized through an analysis of the complete proteome	To assess the impact of various elicitors on the biotechnological production of major centellosides in a specific rapidly growing *C. asiatica* hairy root line, alongside analyzing the expression levels of genes responsible for crucial enzymes in the triterpene biosynthetic pathway (Chapter 3).
Elicitation changes the expression pattern of downstream genes related to centelloside biosynthesis, so unblocking the production of precursors of the biosynthetic pathway, through metabolic engineering techniques, could be an effective strategy to further increase the production of these target compounds in hairy root cultures.	To establish *C. asiatica* hairy roots that overexpress *At-SQS* and *TSAR2* genes and investigate how these genetic alterations influence the expression of centelloside pathway genes, the correlation between squalene/phytosterol contents, and morphological traits, thereby elucidating the regulatory mechanisms governing centelloside biosynthesis (Chapter 4).

Future perspectives: The discovery that part of the precursors flow for centellosides biosynthesis derives to previous branches of the pathway or is accumulate, reveals downstream blockages. A combined strategy with new stepwise or holistic engineering approaches (using transcription factors) together with elicitation, could unlock the last steps of centelloside biosynthesis and allow the design of new transformed root lines with improved productive capacities.
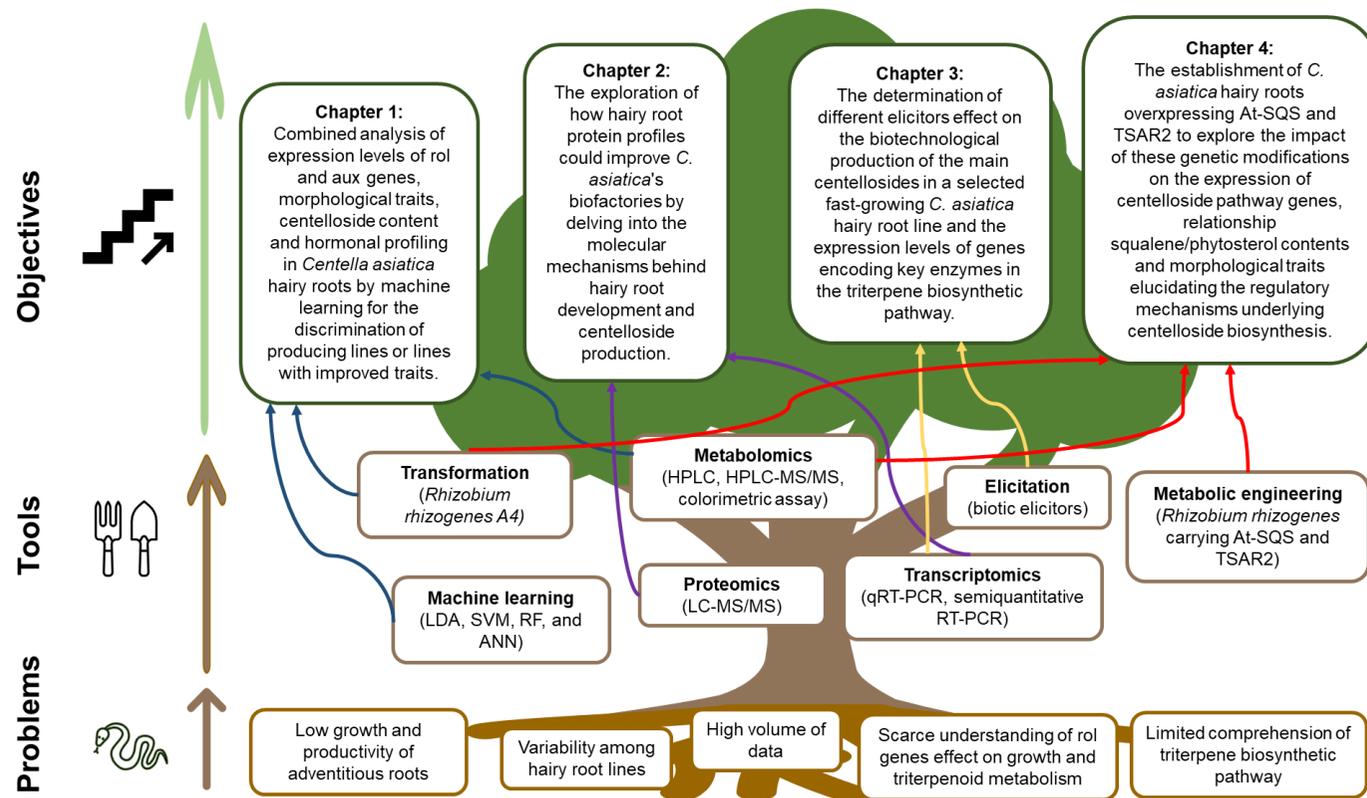
Figure 10. Objectives of PhD thesis: Identifying problems (spinfalls), tools and objectives (ladders). LDA, linear discriminant analysis; SVM, support vector machines; RF, random forest; ANN, artificial neural network; HPLC, high performance liquid chromatography; HPLC-MS/MS, HPLC coupled to mass spectrometry; qRT-PCR, quantitative real time polymerase chain reaction; *At-SQS*, *Arabidopsis thaliana* squalene synthase; *TSAR2*, triterpene saponin biosynthesis activating regulator2.

Supervisor`s report

Mercedes Bonfill, Professor of Plant Phisiology of the University of Barcelona and Diego Hidalgo, Maria Zambrano researcher of the University of Barcelona as Directors of the PhD thesis entitled "Exploring T-DNA gene mechanisms and optimizing centelloside production in *Centella asiatica* hairy root cultures" certify that the thesis present here is the result of the work carried out by Miguel Angel Alcalde Alvites under our guidance and supervision. The own contribution of the PhD candidate to each one of the manuscripts included in the thesis is detailed below.

Chapter 1. Using machine learning to link the influence of transferred *Agrobacterium rhizogenes* genes to the hormone profile and morphological traits in *Centella asiatica* hairy roots.

Alcalde MA, Müller M, Munné-Bosch S, Landín M, Gallego PP, Bonfill M, Palazon J and Hidalgo-Martinez D.

Frontiers in Plant Science (2022). 13:1001023. doi: 10.3389/fpls.2022.1001023

Impact Factor (2022): 5.6. Position 27/238, Q1 in Plant Sciences of JCR.

This article is a part of Miguel Alcalde´s PhD thesis.

MA: Miguel Alcalde performed tissue culture and metabolites determination, determined the hormone profile, and build the different machine learning models.

Chapter 2. Insights into enhancing *Centella asiatica* organ cell biofactories via hairy root protein profiling.

Alcalde MA, Hidalgo-Martinez D, Bru-Martínez R, Sellés-Marchart S, Mercedes Bonfill M, and Palazon J.

Frontiers in Plant Science (2023). XX:XXXXXX. doi: XXXX/XXX (Accepted-In press).

Impact Factor (2022): 5.6. Position 27/238, Q1 in Plant Sciences of JCR.

This article is a part of Miguel Alcalde´s PhD thesis.

MA: Miguel Alcalde designed the research, performed tissue culture, genomic and transcriptomics analysis, determined the proteomics profiles and build machine learning models.

Chapter 3. Metabolic gene expression and centelloside production in elicited *Centella asiatica* hairy root cultures.

Alcalde MA, Cusido RM, Moyano E, Palazon J, and Bonfill, M.

Industrial Crops and Products (2022). 184:114988. doi: 10.1016/j.indcrop.2022.114988.

Impact Factor (2022): 5.9. Position 3/14, Q1 in Agricultural Engineering of JCR.

This article is a part of Miguel Alcalde´s PhD thesis.

MA: Miguel Alcalde designed the research, carried out the experiments, analyzed all the data and wrote the draft manuscript.

Chapter 4. Enhancing centelloside production in *Centella asiatica* hairy root lines through metabolic engineering of triterpene biosynthetic pathway early genes.

Alcalde MA, Palazon J, Bonfill M, and Hidalgo-Martinez D.

Plants (2023). 12:19,3363. https://doi.org/ 10.3390/plants12193363

Impact Factor (2022): 4.5. Position 43/238, Q1 in Plant Sciences of JCR.

This article is a part of Miguel Alcalde´s PhD thesis.

MA: Miguel Alcalde designed the research, performed methodology, analyzed data and wrote the original draft.

For all the above, we consider that the work developed by the PhD candidate grants him the right to defend his thesis in front of a Scientific Committee.

Barcelona, 30th September of 2023

Fdo: Mercedes Bonfill   Diego Hidalgo

Chapter 1. Using machine learning to link the influence of transferred *Agrobacterium rhizogenes* genes to the hormone profile and morphological traits in *Centella asiatica* hairy roots.

Alcalde MA, Müller M, Munné-Bosch S, Landín M, Gallego PP, Bonfill M, Palazon J, & Hidalgo-Martinez D. *Front. Plant Sci.* (2022). 13:1001023. doi: 10.3389/fpls.2022.1001023

Centella asiatica hairy roots

Data collection
(gene expression, centelloside content and hormone profiles)

Machine learning analysis
(Estimation of model accuracy, specificity and sensitivity)
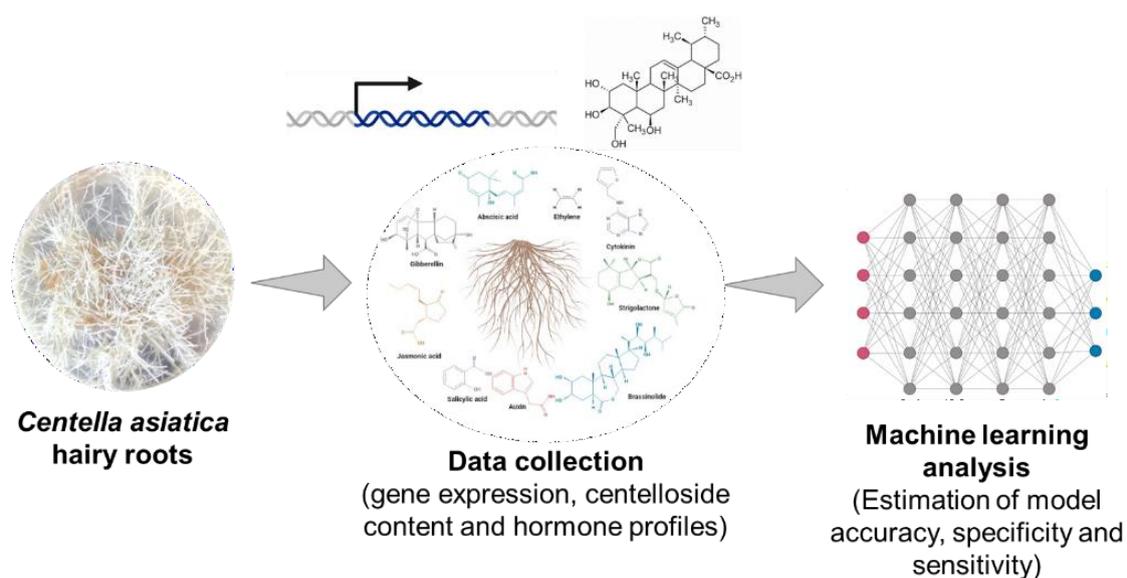
Spanish summary

Las raíces transformadas se forman después de la integración de los genes del T-DNA de *Rhizobium rhizogenes* en el genoma de la planta. Se sabe poco acerca de cómo este pequeño conjunto de genes está relacionado con el perfil hormonal, el cual determina el desarrollo, la morfología y los niveles de producción de metabolitos secundarios en esta plataforma biotecnológica. Para este trabajo empleamos diferentes cultivos de líneas de raíces transformadas de *Centella asiatica* para determinar los posibles vínculos entre la expresión de los genes *rol* y *aux* con los rasgos morfológicos, los perfiles hormonales y la producción de centellósidos. Los resultados obtenidos después de 14 y 28 días de crecimiento se procesaron mediante análisis multivariado y procesos de aprendizaje automático con los algoritmos de bosque aleatorio, máquinas de soporte vectorial, análisis discriminante lineal y redes neuronales. Esto nos permitió obtener modelos capaces de discriminar líneas de raíces altamente productivas según sus niveles de expresión genética (genes *rol* y *aux*) o su perfil hormonal. En total, se evaluaron 12 hormonas, de las cuales se detectaron satisfactoriamente 10. Dentro de este conjunto de hormonas, el ácido abscísico y la citoquinina isopentenil adenosina resultaron ser críticos para definir los rasgos morfológicos y el contenido de centellósidos. Los resultados mostraron que isopentenil adenosina aporta más beneficios a esta plataforma biotecnológica. Además, determinamos el grado de influencia de cada uno de los genes evaluados en el perfil hormonal individual, encontrando que *aux1* tiene una influencia significativa en el perfil de IPA, mientras que los genes *rol* están estrechamente vinculados al perfil de ABA. Finalmente, verificamos de manera efectiva la influencia de los genes en estas dos hormonas específicas a través de observar el efecto en la morfología de las raíces y el contenido de centellósidos.

Check for updates

Using machine learning to link the influence of transferred *Agrobacterium rhizogenes* genes to the hormone profile and morphological traits in *Centella asiatica* hairy roots

Miguel Angel Alcalde[1], Maren Müller[2], Sergi Munné-Bosch[2], Mariana Landín[3], Pedro Pablo Gallego[4], Mercedes Bonfill[1], Javier Palazon[1]*and Diego Hidalgo-Martinez[1,5]*

[1]Department of Biology, Healthcare and the Environment, Faculty of Pharmacy and Food Sciences, University of Barcelona, Barcelona, Spain, [2]Department of Evolutionary Biology, Ecology and Environmental Sciences, Faculty of Biology, University of Barcelona, Barcelona, Spain, [3]Department of Pharmacology, Pharmacy and Pharmaceutical Technology, Group I+D Farma (GI-1645), Faculty of Pharmacy, University of Santiago, Santiago de Compostela, Spain, [4]Agrobiotech for Health, Department of Plant Biology and Soil Science, Faculty of Biology, University of Vigo, Vigo, Spain, [5]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, United States

Hairy roots are made after the integration of a small set of genes from *Agrobacterium rhizogenes* in the plant genome. Little is known about how this small set is linked to their hormone profile, which determines development, morphology, and levels of secondary metabolite production. We used *C. asiatica* hairy root line cultures to determine the putative links between the *rol* and *aux* gene expressions with morphological traits, a hormone profile, and centelloside production. The results obtained after 14 and 28days of culture were processed *via* multivariate analysis and machine-learning processes such as random forest, supported vector machines, linear discriminant analysis, and neural networks. This allowed us to obtain models capable of discriminating highly productive root lines from their levels of genetic expression (*rol* and *aux* genes) or from their hormone profile. In total, 12 hormones were evaluated, resulting in 10 being satisfactorily detected. Within this set of hormones, abscisic acid (ABA) and cytokinin isopentenyl adenosine (IPA) were found to be critical in defining the morphological traits and centelloside content. The results showed that IPA brings more benefits to the biotechnological platform. Additionally, we determined the degree of influence of each of the evaluated genes on the individual hormone profile, finding that *aux1* has a significant influence on the IPA profile, while the *rol* genes are closely linked to the ABA profile. Finally, we effectively verified the gene influence on these two specific hormones through feeding experiments that aimed to reverse the effect on root morphology and centelloside content.

59

Introduction

Plant specialized metabolism is the source of a plethora of bioactive compounds, some of which are uncommon and with important pharmacological activities (Rai et al., 2017). The overuse of medicinal plants and the fact that some of their bioactive compounds are only found in trace amounts in plant tissues has prompted the search for alternative sources of these compounds (Chen et al., 2016).

This is the case with *Centella asiatica* (L.) Urban, which has been used in traditional medicine to treat several chronic diseases since ancient times (Prasad et al., 2019). *C. asiatica* extracts have antidepressant, antiepileptic, antidiabetic, anxiolytic, neuroprotective, antioxidant, antiulcer, antitumor, anti-inflammatory, and healing properties (Gallego et al., 2014; Sun et al., 2020; Arribas-López et al., 2022), about 139 metabolites has been isolated from this plant most of them extracted from leaves and roots (Kunjumon et al., 2022). *C. asiatica* is well known for the accumulation of triterpenoid centelloside, such as madecasosside and asiaticoside, as well as its relevant aglycones madecassic and asiatic acid (Joshi and Chaturvedi, 2013), which contribute to its clinical efficacy (Prakash et al., 2017). As a result, host plants have been over-exploited, and excessive uprooting has put *C. asiatica* in danger of extinction (Mangas et al., 2008). The chemical synthesis of centellosides is either impossible or economically unviable. Research has focused on the potential offered by plant cell culture technology, also known as Plant Biofactories, for efficient specialized metabolite production, such as centellosides (Dhillon et al., 2017).

One of these technologies is the generation of hairy roots, which has been used in several biotechnological approaches for phytochemical production (Donini and Marusic, 2018; Roy, 2020). Hairy root cultures are induced in most dicotyledonous plants by incorporating a segment of *Agrobacterium rhizogenes* DNA (T-DNA) from the plasmid Ri-DNA into the plant cell genome, where the expression of the genes carried out by the T-DNA promotes rooting at the site of infection. Hairy roots can grow rapidly even in the absence of exogenous growth regulators, which is why they are widely used as a transgenic tool to produce specialized metabolites, therapeutic proteins, etc. (Gutierrez-Valdes et al., 2020).

The plasmid Ri-DNA of *A. rhizogenes* strain A4 contains two regions: TL-DNA and TR-DNA. The first contains four *rol* genes (rooting locus): A, B, C, and D, which improve plant cell susceptibility to auxins and cytokinins and are responsible for the formation of these roots (Supplementary Figure 1; Mauro et al., 2017). However, little is known about the molecular changes induced in plant cells by the expression of *rol* genes into the plant genome. *rol*A is found on all Ri plasmids and encodes a small protein with a basic isoelectric point whose expression showed a dramatic reduction in several classes of hormones (Ozyigit et al., 2013). *rol*B may play a critical role in the early stages of hairy-root induction and is the most powerful inducer of secondary metabolism (Dilshad et al., 2021). *rol*C is considered the most conserved of all *rol* genes and has a minor impact on root formation (Makhzoum et al., 2013). T-DNA also contains genes that encode enzymes that direct the production of opines, which are synthesized and excreted by transformed cells and consumed as a source of carbon and nitrogen by *A. rhizogenes* (Matveeva and Otten, 2021). The second region of T-DNA is TR-DNA which contains genes related to auxin biosynthesis, known as *aux*1 and *aux*2. Both regions can be transferred to the nuclear genome of infected plant cells independently (Nemoto et al., 2009).

Most studies on the expression of *rol* and *aux* genes in hairy root cultures have focused on demonstrating their effect on specialized metabolism for increasing the production of phytochemicals, such as centelloside production in *C. asiatica* hairy root cultures (Dilshad et al., 2021), anthraquinones in *rol*A-transgenic tissues of *Rubia cordifolia* calli (Nemoto et al., 2009), or nicotine in tobacco hairy roots (Palazón et al., 1997). The *rol*B transformation was shown to induce resveratrol production in *Vitis amurensis* cells (Kiselev et al., 2007), and *rol*C gene expression was shown to be highly efficient for increasing the production of morphinan, tropane alkaloids (Cardillo et al., 2013; Hashemi and Naghavi, 2016), anthraquinones (Bulgakov et al., 2002).

On the contrary, the relationship between *rol* and *aux* genes expression and the hormonal profile, which is a determinant in root development and morphology (Wahby et al., 2012), has received little attention. As a result, we focus on analyzing the relationship between: (1) expression levels of different *rol* and *aux* genes, (2) morphological traits, (3) production of triterpene saponins and (4) hormonal profiling of various hairy root lines of *C. asiatica*. The combined analysis of these variables enabled us to generate machine learning models that allow for the discrimination of producing lines or lines with improved traits, either by the level of gene expression or hormonal profile. The extent to which *A. rhizogenes* genes influenced each of the hormones measured has been determined.

Materials and methods

Establishment of hairy root culture

The *A. rhizogenes* A4 strain was used in transformation experiments. Bacteria were grown for 48 h (OD 600 = 0.5–0.6)

in liquid YEB (Yeast Extract Beef) medium at 28°C on a rotary shaker at 130 rpm. Explants for co-cultivation and hairy root induction were leaf segments from a *in-vitro* 2-month-old seedling of *C. asiatica*. The explants were cut into 1.5–2 cm² disks with the tip of a scalpel containing a colony of *A. rhizogenes*, cultured at 25°C. All excised explants were then co-cultured in solid MS hormone-free media enriched with 3% sucrose and pH = 5.8. The explants were transferred to a fresh solid MS medium containing 500 mg/l cefotaxime after 48 h of cocultivation in the dark at 28°C. The emerging hairy roots were excised and transferred to a fresh solid MS medium containing 500 mg/l cefotaxime, where they were grown in darkness at 25°C on a rotary shaker. To eliminate the bacteria, this step was repeated every 2 weeks for about 2 months.

Semiquantitative RT-PCR detection and expression of transgene integrations

Semiquantitative RT-PCR was used to detect the integration and expression of *A. rhizogenes* T-DNA genes (*rolA*, *rolB, rolC*, and *aux1*) at the transcript level in the studied transgenic clones, this analysis was previously perform in *C. asiatica* by Mangas et al. (2008). *PureLink RNA Mini Kit (Invitrogen)* was used to isolate RNA from 200 mg of fresh hairy roots lines according to the manufacturer's instructions. The amount and quality of each RNA sample were determined using the *NanoDrop 2000 Spectrophotometer (Thermo Scientific)*. The integrity of the RNA was assessed using agarose gel electrophoresis. The total RNA at a fixed concentration (1.5 μg of RNA) was used as the template for the DNAse treatment. For this purpose, the required sample volume was calculated, taking into account the volume of DNAse I and buffer needed, and brought up to a final volume of 10 μl per sample with sterile H_2O. After adding the DNAse mix, the samples were heated at 37°C for 30 min. Then, for each sample, 1 μl of 50 mM EDTA was added and incubated at 65°C for 10 min to inactivate it. First-strand cDNA was synthesized using the *SuperScriptTM IV First-Strand Synthesis System (Invitrogen)* kit and 2 μl of RNA according to the manufacturer's instructions. Primer3Plus software[1] was used to design PCR primers with G/C content and the presence of introns (Supplementary Table 1). A volume of 1 μl of cDNA products were amplified with 12.5 μl of Green Taq polymerase, 1 μl of each specific primer, and 9.5 μl of H_2O milliQ. A 5-min cycle at 94°C was followed by 60 s at 94°C, 30 s at 60°C, and 1 min at 72°C for 35 cycles, and then another 5-min cycle at 72°C. A no-sample negative control was always included in each set of reactions. PCR products were loaded onto 1%

agarose gels in TAE buffer (1X), and pictures were taken using a Gel Logic 100 camera (KODAK). The bands were quantified using the Kodak Gel Logic 100 Digital Imaging System software (KODAK).

Evaluation of some morphology parameters of hairy root lines

An inoculum of 10 mg fresh weight (FW) from each hairy root line was placed in plates with MS medium solid, and the cultures were maintained for two subcultures every 2 weeks at 25°C in dark conditions, as we had done in previous studies (Alcalde et al., 2022), before evaluating the growth parameters considered using three replicas of each line. The branching rate was defined as the number of lateral roots per cm of initial stem root (number of lateral roots/cm); the growth rate as the average length of the lateral roots (mm/day); and the biomass productivity as the final FW minus initial FW divided by the number of growing days (mg/day).

Extraction and quantification of centellosides

Centellocide production was determined according to Hidalgo et al. (2016) and Alcalde et al. (2022) with slight modification. We weighted 0.5 g of freeze dry material (DW) of hairy roots and added 10 ml of methanol: H_2O (9:1) suspension, which was sonicated for 1 h at room temperature. The following step was to centrifuge at 20,000 rpm for 10 min. After separating the supernatant, the previous step was repeated. The supernatants of the various samples were placed on porcelain mortars and evaporated at 38°C for approximately 24 h before being redissolved in 1.5 ml of methanol. The methanolic extract was filtered through a 0.22 μm filter for HPLC quantification of centellosides. The HPLC system consisted of a Waters 600 Controller pump, a Waters 717 Autosampler automatic injector, a Jasco variable length (UV) 1570 detector, and Borwin data analysis software version 1.5. At room temperature, a Lichrospher 100 RP18 5 μm column (250 × 0.4 mm) was used for gradient chromatography, as described in Supplementary Table 2. The mobile phase consisted of acetonitrile and ammonium phosphate (10 mM), which had been acidified with ortho-phosphoric acid to a pH of 2.5. The acidification improved the definition of the compound peaks. The flow rate was 1 ml/min, and the injection volume was 10 μl. The detector wavelength was set to 214 nm, 1.00 au/v, and the run time was 45 min. To quantify the centellosides (asiatic acid, madecassic acid, asiaticoside, and madecasoside), standards of these 4 compounds were used to prepare calibration curves at concentrations of 10, 25, 50, 100, 250, and 500 ppm.

61

FIGURE 1
Average morphology of the different transformed hairy roots of
C. asiatica at different stages of development.
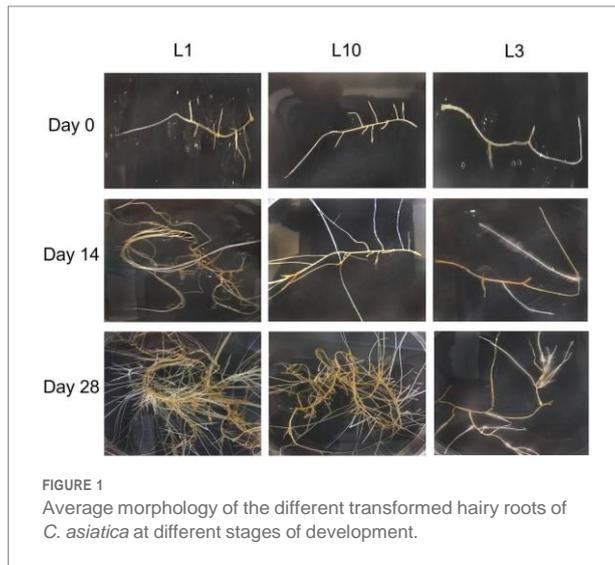
Hormonal profiling of hairy root lines

Samples were collected for hormonal profiling and immediately frozen in liquid nitrogen before being stored at −80°C for subsequent analyses. The endogenous concentrations of the compounds: abscisic acid (ABA), salicylic acid (SA), jasmonates (12-*oxo*-phytodienoic acid [OPDA], jasmonic acid [JA], and jasmonoyl isoleucine [JA-Ile]), cytokinins (2-isopentenenyladenine [2iP], IPA, *trans*-zeatin [*t*-Z], and *trans*-zeatin riboside [*t*-ZR]), the auxin indole 3-acetic acid (IAA), and gibberellins (GA_1, GA_4, and GA_7) were quantified using a protocol modified by Müller and Munné-Bosch (2011). For each hairy root line, 100 ± 5 mg samples were placed in liquid nitrogen in a 2 ml Eppendorf tube using the mixer mill MM400 (Retsch GmbH, Haan, Germany), and then extracted twice with extraction solvent (methanol:isopropanol:acetic acid in a proportion 50:49:1 [v/v/v] with 1% of glacial acetic acid) using ultrasonication (4–7°C). Deuterium-labeled compounds (Olchemim, Olomuc, Czech Republic) were used as internal standards for all phytohormones to estimate recovery rates for each sample. The quantifications were performed by preparing a calibration curve with each of the 13 compounds analyzed and calculating the compound/standard ratio using Analyst™ software (Applied Biosystems, Inc., Foster City, CA, United States). The results were expressed using the FWof the samples.

Statistical analysis and machine learning models

The statistical analysis was carried out using GraphPad Prism version 6.04 for Windows, GraphPad Software, La Jolla California,

United States.[2] Data are presented as the mean ± standard deviation. For statistical comparison, a multifactorial ANOVA analysis was performed, followed by Tukey's multiple comparison test. For morphologic traits, phytohormone concentration, and centelloside production, a p-value ≤ 0.05 was assumed to indicate a significant difference.

Machine learning models, principal component analysis (PCA), and Pearson correlation were performed using R Statistical Software (R Core Team, 2021). The caret package (Kuhn, 2021) was used to create LDA, SVM, RF, and ANN models, whereas randomForestSRC package (Ishwaran and Kogalur, 2022) was used to execute multivariate multiple regression models. A general scheme of the modeling process is presented in Supplementary Figure 9. For model validation, data split and repeated cross-validation methods were used, with 10-fold cross-validation repeated 5 times. Accuracy or coefficient of determination (R-squared) was used as metrics to evaluate the performance of each model, along with specificity and sensitivity. Factoextra (Kassambara and Mundt, 2020) package was used for the PCA, while corrplot package (Wei and Simko, 2021) was used for the calculation of the Pearson correlation coefficient (r). The multivariate multiple regression model's decreased accuracy values were used to perform a hierarchical clustering analysis and were displayed as a heatmap. Datasets for model development and PCA analysis were preprocess by autoscaling method.

Feeding experiment

To validate the model predictions and multivariate analysis results from the previous sections, we designed an experiment that consisted of supplying for 14 and 28 days ABA (13 and 1,300 ng/l) to the line L1 and IPA (1.5 and 150 ng/l) to the line L3. We used a single root to maximize visualization of the effect on growth and branching rates. ABA was bought to Sigma-Aldrich (Steinheim am Albuch, Germany) and IPA was bought to Cayman chemical (Ann Arbor, Michigan, USA). The stock solutions where prepared at 1 mg/ml in methanol followed by serial dilution prior to be added to culture media.

Results

Hairy root traits and centelloside content

The plant material used in this study consisted of 10 hairy root lines free of *Agrobacterium rhizogenes* (L1, L2, L3, L4, L6, L7, L8, L10, L12, and L14), which were cultivated for 2 and 4 weeks, before morphological traits and biomass production were registered, as described in the Material and methods section. Figure 1 shows the development of the root lines throughout the

---

2   www.graphpad.com (Accessed August 22, 2022).

experiment. Most of the lines had the typical morphology of hairy roots, but there were noticeable differences in branching rate, growth rate, and biomass accumulation. Statistical differences were found between lines in terms of branching and growth rate, but there were no significant changes between sampling times. Figure 2A shows the branching rate, with L1 standing out from the others, followed by L12. L3 had the lowest branching values at 2 and 4 weeks of growth. In terms of growth rate (Figure 2B), two lines predominated: L1 and L2. Line L14 has the lowest growth rate value at the end of 4 weeks. Finally, we observed significant differences in biomass productivity (Figure 2C) between weeks 2 and 4 for lines L1, L2, L8, L10, and L12, with increases ranging from 2 to 5 times.

The total productivity of centelloside was expressed in two ways: mg per g of dry weight (DW) and mg per liter of culture medium (Figure 3). These results showed a similar profile between hairy root lines as the traits mentioned previously, as well as significant differences between sampling times. Lines 1 and 2 had the highest production values, followed by lines 10, 12, and 14. Lines 3 to 8 had the lowest values. The productivity

range of centelloside values in mg/g DW oscillated between 0.14 ± 0.02 and 0.96 ± 0.03 after 14 days, and between 0.44 ± 0.27 and 5.49 ± 0.20 after 28 days. These results revealed the root lines' various capacities to accumulate this type of compound over time. Some of them, such as L2, showed a content that was 8 times higher at 28 days than it was at 14 days, whereas line L7's increase was only 0.7 times. In terms of centellosides profile, made cassoside was the main compound in almost all lines (Supplementary Figure 2). We can see from these results that the transformed root lines can be divided into at least 3 groups: L1 and L2, then L3, L4, L6, L7, and L8, and finally L10, L12, and L14.

Gene expression

PCR (data not shown) and semi-quantitative RT-PCR were used to confirm the integration and expression of the *rol* and *aux* genes. Supplementary Figure 3 shows the results of semi-quantitative RT-PCR in 10 *C. asiatica* hairy root lines grown in MS basal medium after 28 days, with the 5.8 s rRNA used as a housekeeping gene for normalization. A principal component analysis (PCA) was used to investigate how the expression of these genes was related to hairy root traits and centelloside content. Figure 4A summarizes the information about hairy root samples and their multiple gene expression by two components: PC1 = 63.5% and PC2 = 27.5%, which account for 91% of the model's total variance. L3, L4, L6, and L7 are the lines with the lowest expression of all genes, according to PC1. According to PC2, other subgroups can be seen within the lines with higher expression. The first (L10, L12, and L14) was associated with high expressions of the *rolC* and *rolB* genes, while the second (L1, L2, and L8) showed higher expressions of the *rolA* and *aux1* genes. A positive correlation was observed between centelloside content, branching, biomass productivity, *rol*, and *aux*1 genes (Figure 4B). The highest centelloside productions were strongly related to *rol*A ($r = 0.71$) and *aux*1 ($r = 0.70$) genes. The *rol*C was the least effective, with a slightly negative effect on elongation rates ($r = -0.29$). Similar behavior was observed for samples at 14 days (Supplementary Figure 4).

Prediction of production degree based on gene expression

The statistically significant correlations and differences established the concept of identifying production lines based on their gene expression profile. We assigned the following tags to the subgroups that represent the centelloside content: HIGH (L1 and L2), MID (L10, L12, and L14), and LOW (L3, L4, L6, L7, and L8). Four different classification machine learning algorithms were tested on a dataset (Supplementary Table 3) containing gene expression information from 10 lines cultivated for 14 and 28 days, and accuracy was used to track the models' performance (Table 1).

FIGURE 2
Morphological traits of hairy root lines of *C. asiatica*.
(A) Branching rate. **(B)** Growth rate. **(C)** Biomass productivity.
Data represent the mean±SD of three replicates. Different letters show significant differences between hairy root lines. ns=no differences (α =0.05).

63
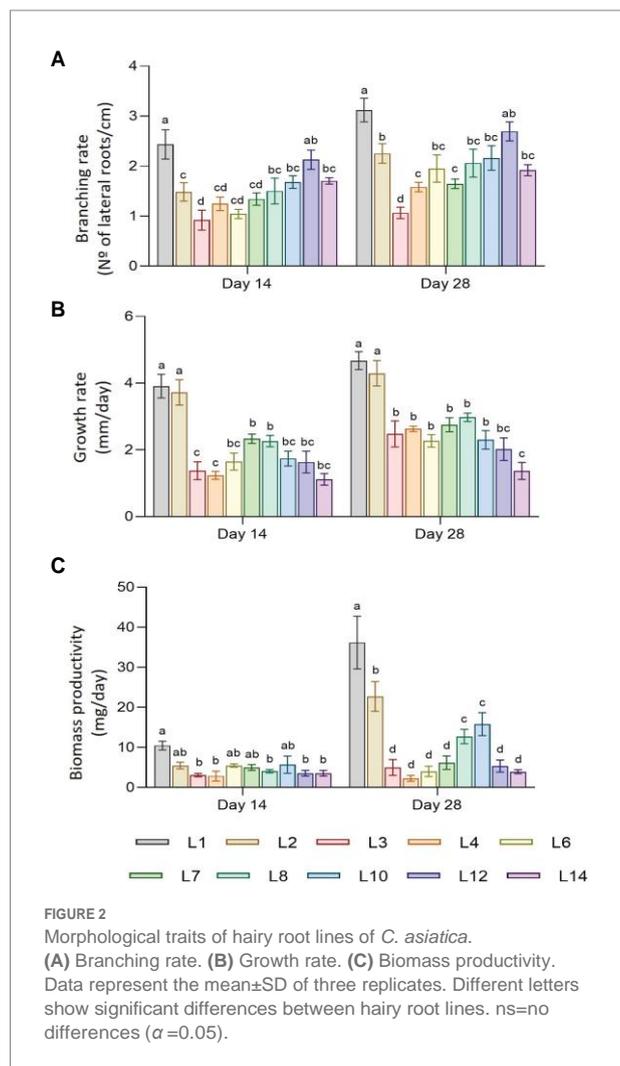
The most accurate model was random forest (RF) with 91.33% correct classification, and the least accurate was artificial neural network (ANN) with 89.5% correct classification. The sensitivity range was 62.5 to 100%, with RF having the highest values, followed by linear discriminant analysis (LDA) and ANN, and supported vector machine (SVM) having the lowest values. Model specificity ranged between 78.6 and 100%. The importance of each gene for correct classification was calculated and ranked by the mean decrease in Gini. The *rol*C is at the top of the list, followed by *rol*A, *rol*B, and finally *aux*1 (Supplementary Figure 5A).

FIGURE 3
Centelloside content in mg/g DW of hairy root lines of *C. asiatica* measured as the sum of asiaticoside + madecassoside + asiatic acid + madecassic acid. Data represent the mean±SD of three replicates. Different letters show significant differences between hairy root lines. ns=no differences (α =0.05).
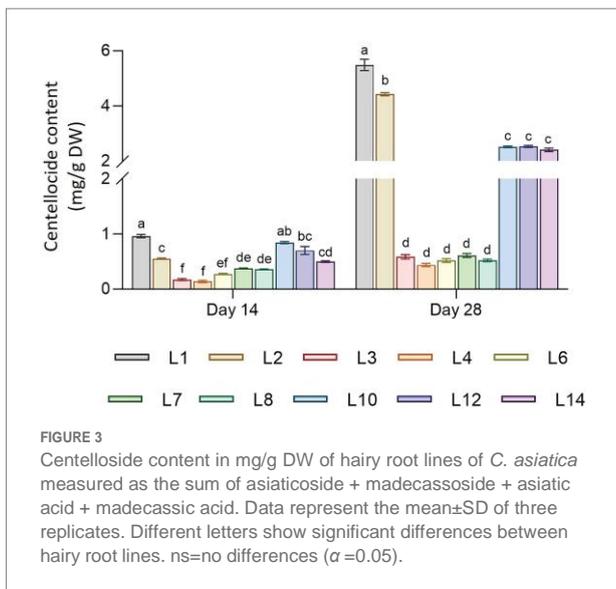
Hormone profiles

The hormone profile of the hairy root lines was evaluated at 14 and 28 days, with 13 compounds measured, including ABA, SA, OPDA, JA, JA-Ile, 2iP, IPA, *t*-Z, *t*-ZR, IAA, GA_1, GA_4, and GA_7. GA_4 was detected, but GA_1 and GA_7 values were below the detection threshold for all rhizoclones (Supplementary Figure 6). LOW centelloside producer lines had higher concentrations of ABA, SA, JA-Ile, and IAA, especially at 4 weeks of growth. MID centelloside producer lines had the highest concentration of 2-iP throughout the experiment. HIGH producer centelloside lines had the highest concentration value for IPA. The other hormones (JA, OPDA, *t*-Z, *t*-ZR, and GA_4) showed different values depending on the week of growth for each transformed root line.

The loadings plot in Figure 5 depicts the relationship between hairy root traits, centelloside contents, and hormone profile at various sampling times. The PCA was composed by nine components, where PC1 covered the 56% and PC2 = 22.7% which account for 78.9% of the model's total variance. At 14 days (Figure 5A; Supplementary Figure 7), the only hormone with a strong positive correlation ($r > 0.7$) for all traits and centelloside content was IPA hormone. OPDA showed a positive effect for branching and centelloside content, and *t*-Z only showed a positive correlation for centelloside content. IAA, on the other hand, had a strong negative effect ($r < 0.7$) on centelloside production, branching, and biomass production. ABA also had a negative effect on centelloside production and branching. IPA was strongly correlated ($r > 0.8$) with hairy root elongation and had a positive effect on centelloside content and biomass productivity at
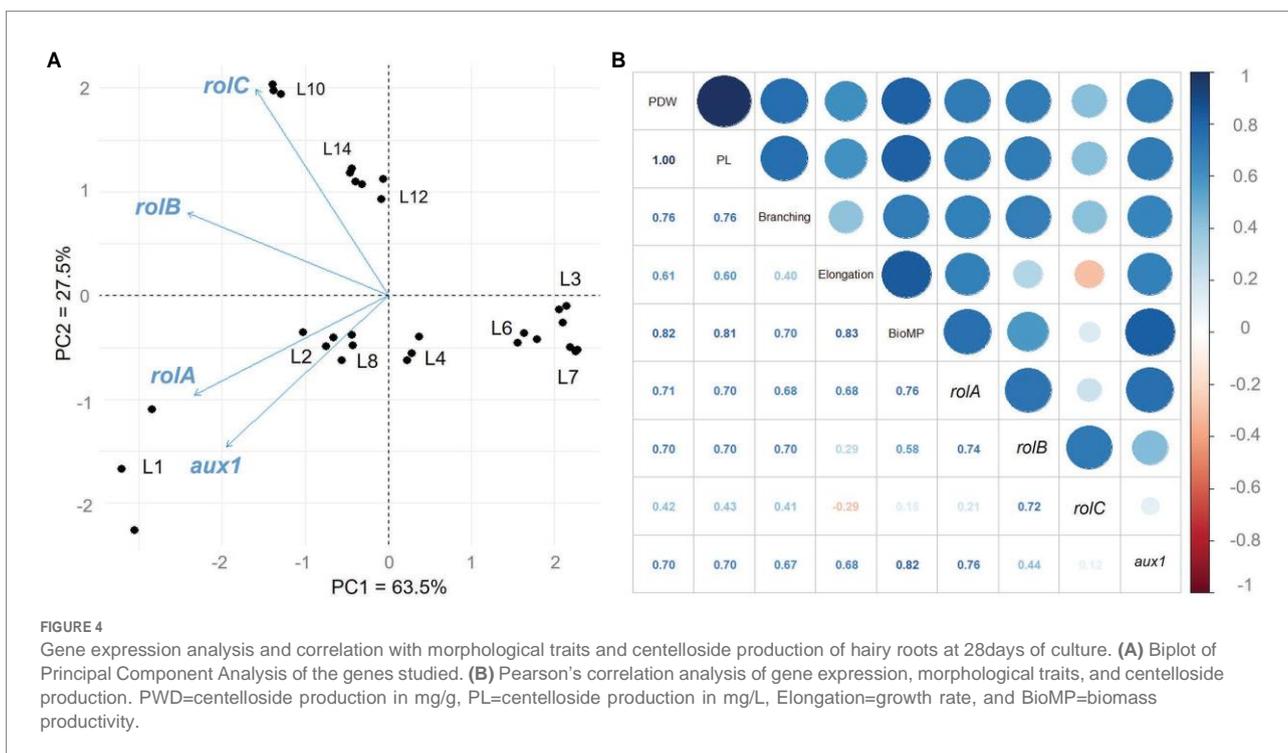
FIGURE 4
Gene expression analysis and correlation with morphological traits and centelloside production of hairy roots at 28days of culture. **(A)** Biplot of Principal Component Analysis of the genes studied. **(B)** Pearson's correlation analysis of gene expression, morphological traits, and centelloside production. PWD=centelloside production in mg/g, PL=centelloside production in mg/L, Elongation=growth rate, and BioMP=biomass productivity.

Model	Accuracy %	Sensitivity %			Specificity %		
		HIGH	MID	LOW	HIGH	MID	LOW
Prediction of production degree based on gene expression							
LDA	90.77	66.67	100	100	100	100	83.33
SVM	90.05	83.33	100	62.5	78.57	100	91.67
RF	91.33	83.33	100	87.5	92.9	100	91.7
ANN	89.5	66.67	100	100	100	100	83.33
Prediction of production degree based on hormone profiles							
LDA	82.57	100	100	50	75	100	100
SVM	100	100	100	100	100	100	100
RF	100	100	100	100	100	100	100
ANN	93.75	100	100	100	100	100	100

28 days (Figure 5B; Supplementary Figure 8). OPDA and GA4 showed a negative correlation ($r < 0.73$) for all traits and centelloside content. ABA, SA, and t-Z were also negatively correlated except for root elongation. When the hormone content is compared between the two sampling times, the high amount of IPA hormone is repeatedly associated with a high content of centelloside, while the high amount of ABA is associated with low content.

Prediction of production degree based on hormone profiles

Similarly, we investigated whether the hormonal profiles of transformed root lines could predict the degree of centelloside production. Four different classification machine learning algorithms were tested on a dataset (Supplementary Table 4) containing hormone content from the HIGH, MID, and LOW groups, and accuracy was used to track the models' performance (Table 1). The most accurate models were RF and SVM, with 100% correct classification, and the least accurate model was LDA, with 82.57%. The sensitivity ranged from 50 to 100%, with SVM, RF, and ANN having the highest values and LDA having the lowest. The specificity range of the models was between 75 and 100%. The SVM model performed the best in classifying different degrees of centelloside production. The importance of each hormone for correct classification was calculated and ranked by the mean decrease in the Gini coefficient. The top 5 on this list were IPA, 2iP, ABA, IAA, and GA4 (Supplementary Figure 5B).

Gene influence on hormone profiles

We investigated how genes influence hormone profiles after studying how traits and centelloside content can be predicted by their gene expression or hormone profile. A multivariate multiple regression model was created for this

purpose using the randomForestSRC R-package (see Materials and methods). The different hormones were treated as dependent variables in the model, while the genetic expression of *rol*A, *rol*B, *rol*C, and *aux*1 genes was treated as independent variables. The R-squared value for the model with all hormones was 0.788, which was used to measure the goodness of fit, the performance error of the model was 0.1998. The variable importance (VIMP) was used to compare the influence of genes on each hormone profile to aid in the interpretation of the multivariate regression model. Figure 6 shows a cluster analysis of hormones based on the influence of gene expression. ABA was most influenced by the behavior of *rol*B, followed by *rol*A and *rol*C, and *aux*1 had a low influence on its content. IPA was heavily influenced by *aux*1, followed by *rol*A. All genes had a greater than average influence on the 2 iP, with *rol*C having the greatest effect. The GA4 behavior was slightly affected by *rol*A and *aux*1. The genes *rol*A and *rol*C had a positive impact on IAA, while *aux1* and *rol*B had an average impact. Starting with SA, gene influence declined, followed by JA-ile, t-Z, OPDA, and JA as the hormone least influenced by gene expression. To improve the model's fitness, the less influenced hormones by the *rol* or *aux* genes were eliminated one by one. When JA, OPDA, t-Z, JA-ile, SA, and GA4 were excluded, the best fitness was 0.837. In contrast, the absence of ABA resulted in the greatest decrease in model fitness, with an R-squared value of less than 0.3, followed by IPA and 2 iP. This abrupt decrease in R-squared value confirms the connection between the expression of these genes and the hormone profile of ABA, IPA, and 2 iP.

Y-randomization was implemented to prove accurate prediction potential of the model. We selected the top two most influenced hormones ABA, IPA and JA which was the lowest influenced by the gene expression to do this test. Individual models were built for each of these three hormones, the R-squared value was calculated and compared against the population of R-squared values obtained after 1,000 permutation (Supplementary Figure 10). As a result of this test, ABA and IPA showed to be accurate predicted by the regression model. Additionally, JA showed the overlapping of the simulated values with the original value of R-squared. The above results matched with the multivariate multiple regression model previously developed.

Feeding experiments

Figure 7A shows the development of root lines under the exogenous hormonal influence of ABA or IPA. The controls behaved consistently with the previous experiments, whereas the effects of ABA were perceived after 14 days at concentrations of 13 ng/l, causing significant decreases in the measured traits (Figures 7B–D) on the HIGH line. After 28 days, the lower concentration reduced branching, growth rate, and biomass

FIGURE 5

Loading plot of hormonal profiling, morphological traits and centelloside production of hairy roots cultures **(A)** at 14days and **(B)** at 28days. PWD=centelloside production in mg/g, PL=centelloside production in mg/L, Elongation=growth rate, and BioMP=biomass productivity.

FIGURE 6

Global heatmap of the influence of *rol* and *aux1* genes on the hormone profile in hairy roots of *C. asiatica*.

productivity by 0.83, 0.54, and 0.45 times, respectively compared to the control. At 1,300 ng/l, the same traits were reduced 0.65, 0.24, and 0.28 times, respectively. When IPA was given to the LOW line, the opposite effect was seen; the effects on traits were visible at 28 days and concentrations of 150 ng/l. At the highest concentration, branching, growth rate, and biomass productivity were increased 2.60, 1.25, and 3.8 times after 28 days, respectively. Branching rate and biomass productivity were the most affected by ABA and IPA. Regarding centelloside content (Supplementary Figure 11), after 28 days on ABA treatment the HIGH line showed a decrease in 21% compared to the control when ABA was at 13 ng/l, while 74% of control when ABA was at

1300 ng/l. In contrast, the LOW line after 28 days on IPA treatment showed an increase in 1.76 times compared to the control when IPA was at 1.5 ng/l, while 3.21 times when IPA was 150 ng/l.

Discussion

The hairy root syndrome is a disease that affects many plants and is caused by the infection and subsequent insertion of a fragment of the *A. rhizogenes* plasmid known as T-DNA ("transfer" DNA; Mauro et al., 2017). The *rol* genes, which are found in the TR-DNA region, are primarily responsible for the

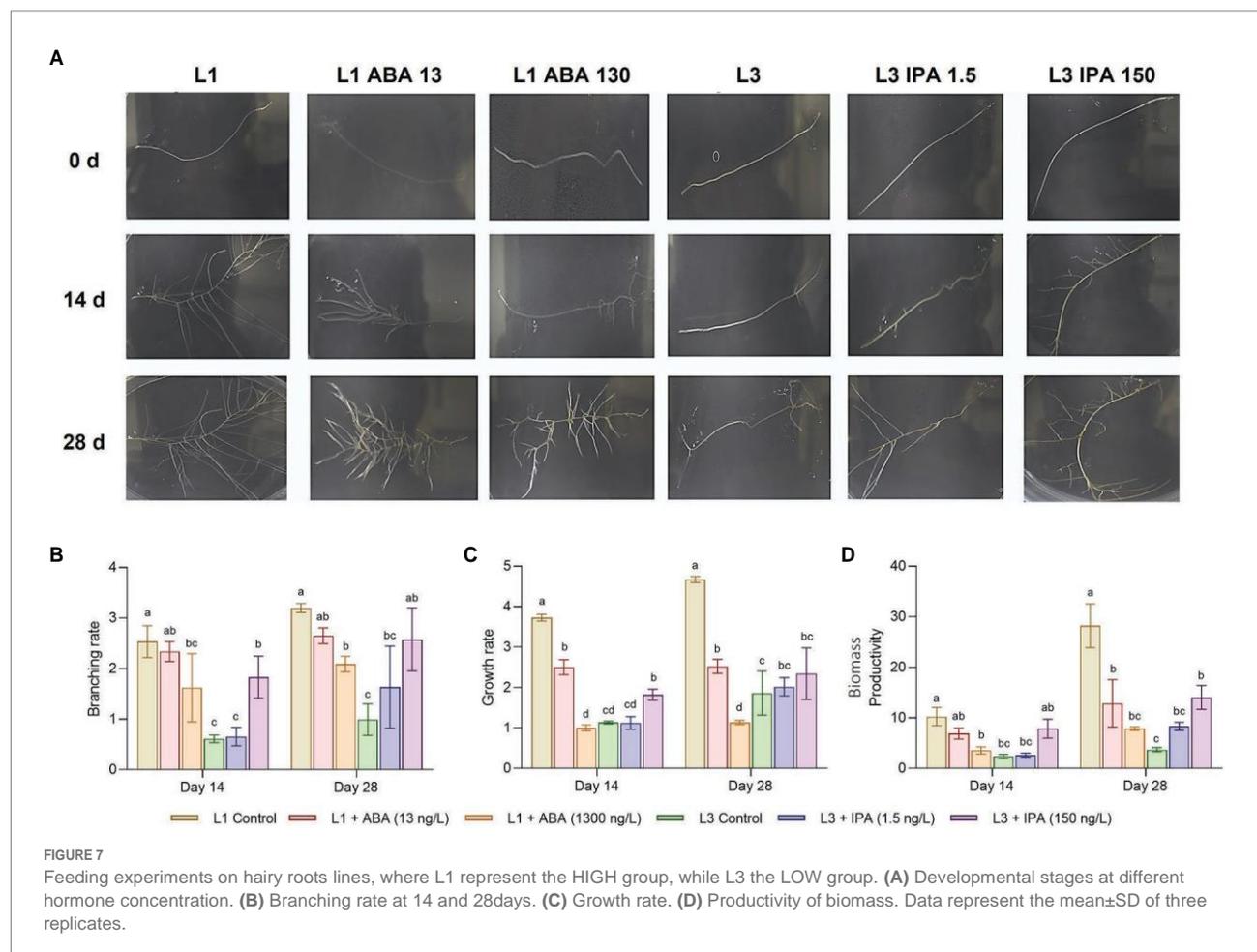morphology and formation of these hairy roots (Ozyigit et al., 2013).

This type of *in vitro* culture system is useful for secondary metabolites biosynthesis and biotechnological production (Bonfill et al., 2015) since these cultures grow much faster than other types of *in vitro* cultures (Kim et al., 2007) and produce a spectrum of secondary metabolites similar to plant roots (Ruslan et al., 2012).

We observed the presence and expression of all *rol* and *aux* genes in all hairy lines studied (Supplementary Figure 3), which is similar to the work of Komarovská et al. (2009) in *Hypericum tomentosum*. and *Hypericum tetrapterum*, but differs from what was observed by Alpizar et al. (2008) in *Coffea arabica*, where the presence of any *aux* gene was found in all transformed lines. We selected the A. rhizogenes TL-DNA genes rolA, B and C because they have been shown to play the most relevant role in hairy root development (Sarkar et al., 2018; Bahramnejad et al., 2019), and the aux1 gene as a representative of the proper integration of the TR-DNA region in the transformed root lines (Mano and Nemoto, 2012). In our study, the expression level of all *rol* genes was higher than that of the *aux* gene (Supplementary Figure 3), which may be since only the presence of TL-DNA genes is required for long-term hairy root growth

(Dessaux and Petit, 1994; Chriqui et al., 1996). The *rol*D was not analyzed because it was not detected in all *A. rhizogenes* strains (Pavlova et al., 2014).

When we compared all the hairy root lines or rhizoclones, we observed differences in growth and morphology that could be grouped into different categories, and it was thanks to the cultivation of these roots in a solid medium that the quantification of their traits was easily reproducible and simple to do. These differences between rhizoclones were attributed to variations in the nature, size, and number of T-DNA integrations into the host genome (Alpizar et al., 2008; Roychowdhury et al., 2015; Thwe et al., 2016). An evaluation with new methods to determine the number of copies of T-DNA integrated in the root lines could be an extension of the present study (Głowacka et al., 2016), which would allow identifying if the number of integrated copies is more relevant than the place of integration of the genes *rol* or *aux1*.

HIGH group lines had the highest expression value of the *rol*A gene, as well as the highest rooting rate, growth rate, biomass production, and centelloside content (Supplementary Figure 3). This gene is found in all Ri-plasmids and may be involved in the interaction with nucleic acids, which may be related to the regulation of gene expression in plants (Pavlova et al., 2014). The

FIGURE 7
Feeding experiments on hairy roots lines, where L1 represent the HIGH group, while L3 the LOW group. **(A)** Developmental stages at different hormone concentration. **(B)** Branching rate at 14 and 28days. **(C)** Growth rate. **(D)** Productivity of biomass. Data represent the mean±SD of three replicates.

67

importance of this gene in secondary metabolism production was observed in *Rubia cordifolia* callus for anthraquinone production (Shkryl et al., 2008).

The evaluation of the gene expression using a multivariable analysis such as PCA allowed us to visualize in a simplified way the correlation and behavior of each of the characteristics of the different root lines (centellosides content, branching, elongation, and biomass production) over time. This analysis identified the grouping of root lines based on the degree of specific and recurrent gene expression. In general, low expression of *rol* and *aux*1 genes was associated with poor traits and low centelloside production. The genes *aux*1 and *rol*A were found to be more closely linked to lines with higher centelloside and trait content values. The *aux*1 gene is responsible for differences in hairy root growth and morphology (Ozyigit et al., 2013). This could be related to the line's high rooting and growth rate in long-term culture (Figures 2, 4A), which was also observed by Lütken et al., 2017.

MID group lines had the highest levels of *rol*B and *rol*C gene expression (Figure 4A). The *rol*B is the most powerful inducer of secondary metabolism and the greatest suppressor of cell growth (Bulgakov, 2008). The MID group lines produced lower amount of centellosides than the HIGH group lines, which could be attributed to the high expression of *rol*C since it has previously been shown to have antagonistic effects between these two genes (Bulgakov et al., 2003). The *rol*C gene can stimulate the production of tropane alkaloids (Bonhomme et al., 2000), pyridine alkaloids (Palazon et al., 1998), ginsenosides (Bulgakov et al., 1998), and flavonoids (Ismail et al., 2017) in different *in vitro* culture systems, which differs somewhat from our studies.

The multivariable analysis (PCA) also exposed the behavior of the different hormones between traits and centelloside production, revealing a positive correlation with the IPA hormone content and a negative correlation with the ABA hormone. It is well known that ABA regulates numerous aspects of plant growth. Dicot plants deficient in this hormone have reduced seed dormancy and wilty phenotypes (Harris, 2015; Nambara, 2016). However, high levels of ABA have been shown to inhibit cell division in apical meristems and root elongation (Bai et al., 2009; Takatsuka and Umeda, 2014; Yang et al., 2014; Sun et al., 2018). This last scenario, in which high content is negatively correlated with elongation and branching, is consistent with our results, as is the low production of triterpene saponins. In terms of IPA, its high presence in hairy root lines was consistent with improved biomass production, elongation, and branching, and is supported by the abundance of lateral root meristems, which are one of the main sites of cytokinin synthesis (Nordström et al., 2004).

We focus on centelloside production for machine learning modeling since it is one of the most important parameters for biotechnological applications. The correlation analysis allowed us to consider the positive relationship between centelloside production and biomass production, elongation, and branching rate, allowing us to omit them from the models and simplify their execution. By discretizing the content of centellosides, it was possible to apply different supervised machine learning models, such as dimensionality reduction (LDA; Zhao et al., 2020), instance-based (SVM; Cristianini and Ricci, 2008), ensemble methods (RF; Liaw and Wiener, 2002), and artificial neural network (ANN; Venables and Ripley, 2002). The reason for selecting these models was due to their nature, as mentioned above. In general, RF and SVM performed the best with the data presented in this work, correctly classifying the samples into the classes proposed. The LDA, on the other hand, produced the worst results, which could be attributed to the fact that the data distribution was not normal for all variables. The data distribution was identified by Shapiro–Wilk test and the results are shown in Supplementary Table 5.

Multiple multivariate regression models were used to understand how certain hormones and gene expression levels (*rol* and *aux1* genes) interact to coordinate root growth and development. This allowed us to simultaneously evaluate the influence of each level of genetic expression on the profile of each hormone studied. The random forest method has the advantage of being able to work with data whose distribution may or may not be normal; it evaluated the importance of each variable within the model, allowing identifying the degree of influence of genes on each of the growth regulators. Feeding experiments validated the model by demonstrating that the analyzed phytohormones (IPA and ABA) were determinants in increasing the high producer line and decreasing the low producer line, as well as influencing centelloside production.

The use of this biotechnological platform together with machine learning techniques resulted in the implementation of models that allow us to discriminate root lines based on their level of production of secondary metabolites such as centelloside, with random forest outperforming all others. This discrimination was made possible by using gene expression levels of the *rol* and *aux1* genes, as well as hormone profiles. Furthermore, the degree of influence of each gene on the individual profile of each hormone studied was determined, with IPA and ABA being the most affected due to the action of the *rol* and *aux1* genes. Finally, the results of the gene influence analysis on these two specific hormones were successfully tested with feeding experiments aimed at reversing the effect on root morphology and centelloside content.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

MB, JP, and DH-M designed the research. MA, MM, and SM-B determined the hormone profile. MA, ML, and PG

build neural networks models. MA and DH-M build the others machine learning models. MA performed tissue culture and metabolites determination. All authors contributed to the article and approved the submitted version.

Funding

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1001023/full#supplementary-material

References

Alcalde, M. A., Cusido, R. M., Moyano, E., Palazon, J., and Bonfill, M. (2022). Metabolic gene expression and centelloside production in elicited Centella asiatica hairy root cultures. *Ind. Crop. Prod.* 184:114988. doi: 10.1016/J.INDCROP.2022.114988

Alpizar, E., Dechamp, E., Lapeyre-Montes, F., Guilhaumon, C., Bertrand, B., Jourdan, C., et al. (2008). Agrobacterium rhizogenes-transformed roots of coffee (Coffea arabica): conditions for Long-term proliferation, and morphological and molecular characterization. *Ann. Bot.* 101, 929–940. doi: 10.1093/AOB/MCN027

Arribas-López, E., Zand, N., Ojo, O., Snowden, M. J., and Kochhar, T. (2022). A systematic review of the effect of Centella asiatica on wound healing. *Int. J. Environ. Res. Public Health* 19, 3266. doi: 10.3390/ijerph19063266

Bahramnejad, B., Naji, M., Bose, R., and Jha, S. (2019). A critical review on use of agrobacterium rhizogenes and their associated binary vectors for plant transformation. *Biotechnol. Adv.* 37:107405. doi: 10.1016/J.BIOTECHADV.2019.06.004

Bai, L., Zhang, G., Zhou, Y., Zhang, Z., Wang, W., Du, Y., et al. (2009). Plasma membrane-associated proline-rich extensin-like receptor kinase 4, a novel regulator of Ca2+ signalling, is required for abscisic acid responses in Arabidopsis thaliana. *Plant J.* 60, 314–327. doi: 10.1111/J.1365-313X.2009.03956.X

Bonfill, M., Cusidó, R. M., Lalaleo, L., and Palazón, J. (2015). "Plant cell and organ cultures as a source of phytochemicals," in *Recent Advances in Pharmaceutical Sciences*. eds. D. Muñoz-Torrero, M. P. Vinardell and J. Palazón (Kerala, India: Research Signpost), 33–49.

Bonhomme, V., Laurain-Mattar, D., and Fliniaux, M. A. (2000). Effects of the rol C gene on hairy root: induction development and Tropane alkaloid production by Atropa belladonna. *J. Nat. Prod.* 63, 1249–1252. doi: 10.1021/NP990614L

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

Bulgakov, V. P., Khodakovskaya, M. V., Labetskaya, N. V., Chernoded, G. K., and Zhuravlev, Y. N. (1998). The impact of plant rolC oncogene on ginsenoside production by ginseng hairy root cultures. *Phytochemistry* 49, 1929–1934. doi: 10.1016/S0031-9422(98)00351-3

Bulgakov, V. P., Tchernoded, G. K., Mischenko, N. P., Khodakovskaya, M. V., Glazunov, V. P., Radchenko, S. V., et al. (2002). Effect of salicylic acid, methyl jasmonate, ethephon and cantharidin on anthraquinone production by Rubia cordifolia callus cultures transformed with the rolB and rolC genes. *J. Biotechnol.* 97, 213–221. doi: 10.1016/S0168-1656(02)00067-6

Bulgakov, V. P., Tchernoded, G. K., Mischenko, N. P., Shkryl, Y. N., Glazunov, V. P., Fedoreyev, S. A., et al. (2003). Effects of Ca2+ channel blockers and protein kinase/phosphatase inhibitors on growth and anthraquinone production in Rubia cordifolia callus cultures transformed by the rolB and rolC genes. *Planta* 217, 349–355. doi: 10.1007/S00425-003-0996-5/TABLES/2

Cardillo, A. B., Giulietti, A. M., Palazón, J., and Bonfill, M. (2013). Influence of hairy root ecotypes on production of tropane alkaloids in Brugmansia candida. *Plant Cell Tissue Organ Cult.* 114, 305–312. doi: 10.1007/S11240-013-0326-Y/FIGURES/5

Chen, S. L., Yu, H., Luo, H. M., Wu, Q., Li, C. F., and Steinmetz, A. (2016). Conservation and sustainable use of medicinal plants: problems, progress, and prospects. *Chinese Med. (United Kingdom)* 11, 1–10. doi: 10.1186/S13020-016-0108-7/TABLES/5

Chriqui, D., Guivarc'h, A., Dewitte, W., Prinsen, E., and Van Onkelen, H. (1996). Rol genes and root initiation and development. *Plant Soil* 187, 47–55. doi: 10.1007/BF00011656

Cristianini, N., and Ricci, E. (2008). "Support vector machines," in *Encyclopedia of Algorithms*. ed. M. Y. Kao (Boston, MA: Springer US), 928–932. doi: 10.1007/978-0-387-30162-4_415

Dessaux, Y., and Petit, A. (1994). "Opines as screenable markers for plant transformation," in *Plant Molecular Biology Manual*. eds. B. Gelvin and A. S. Robbert (Netherlands: Springer), 171–182.

Dhillon, P., Kumari, P., Kumari, C., and Singh, P. S. (2017). Phytochemical screening of selected medicinal plants for secondary metabolites. *J. Life. Sci. Sci. Res* 3, 1151–1157. doi: 10.21276/ijlssr.2017.3.4.9

Dilshad, E., Noor, H., Nosheen, N., Gilani, S. R., Ali, U., and Khan, M. A. (2021). "Influence of rol Genes for Enhanced Biosynthesis of Potent Natural Products," in *Chemistry of Biologically Potent Natural Products and Synthetic Compounds*. (USA: John Wiley & Sons, Ltd), 379–404.

Donini, M., and Marusic, C. (2018). "Hairy roots as bioreactors for the production of biopharmaceuticals," in *Hairy Roots: An Effective Tool of Plant Biotechnology*. eds. S. Vikas, M. Shakti and M. Sonal (Singapore: Springer), 213–225.

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

Głowacka, K., Kromdijk, J., Leonelli, L., Niyogi, K. K., Clemente, T. E., and Long, S. P. (2016). An evaluation of new and established methods to determine T-DNA copy number and homozygosity in transgenic plants. *Plant Cell Environ.* 39, 908–917. doi: 10.1111/pce.12693

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.* 11:33. doi: 10.3389/FPLS.2020.00033/BIBTEX

Harris, J. M. (2015). Abscisic acid: hidden architect of root system structure. *Plants* 4, 548–572. doi: 10.3390/PLANTS4030548

Hashemi, S. M., and Naghavi, M. R. (2016). Production and gene expression of morphinan alkaloids in hairy root culture of Papaver orientale L. using abiotic elicitors. *Plant Cell Tissue Organ Cult.* 125, 31–41. doi: 10.1007/S11240-015-0927-8/FIGURES/8

69

Hidalgo, D., Steinmetz, V., Brossat, M., Tournier-Couturier, L., Cusido, R. M., Corchete, P., et al. (2016). An optimized biotechnological system for the production of centellosides based on elicitation and bioconversion of Centella asiatica cell cultures. *Eng. Life Sci.* 17, 413–419. doi: 10.1002/ELSC.201600167

Ishwaran, H., and Kogalur, U. B. (2022). Fast unified random forests for survival, regression, and classification (RF-SRC). Available at: https://cran.r-project.org/package=randomForestSRC (Accessed August 22, 2022).

Ismail, H., Dilshad, E., Waheed, M. T., and Mirza, B. (2017). Transformation of lettuce with rol ABC genes: extracts show enhanced antioxidant, analgesic, anti-inflammatory, antidepressant, and anticoagulant activities in rats. *Appl. Biochem. Biotechnol.* 181, 1179–1198. doi: 10.1007/S12010-016-2277-3/FIGURES/7

Joshi, K., and Chaturvedi, P. (2013). Therapeutic efficiency of Centella asiatica (L.) Urb. An underutilized green leafy vegetable: an overview. *Int. J. Pharma Bio Sci.* 4, 135–149. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.448.1137&rep=rep1&type=pdf (Accessed August 12, 2022).

Kassambara, A., and Mundt, F. (2020). Factoextra: extract and visualize the results of multivariate data analyses. Available at: https://cran.r-project.org/package=factoextra (Accessed August 22, 2022).

Kim, O. T., Bang, K. H., Shin, Y. S., Lee, M. J., Jung, S. J., Hyun, D. Y., et al. (2007). Enhanced production of asiaticoside from hairy root cultures of Centella asiatica (L.) urban elicited by methyl jasmonate. *Plant Cell Rep.* 26, 1941–1949. doi: 10.1007/S00299-007-0400-1/FIGURES/6

Kiselev, K. V., Dubrovina, A. S., Veselova, M. V., Bulgakov, V. P., Fedoreyev, S. A., and Zhuravlev, Y. N. (2007). The rolB gene-induced overproduction of resveratrol in Vitis amurensis transformed cells. *J. Biotechnol.* 128, 681–692. doi: 10.1016/J.JBIOTEC.2006.11.008

Komarovská, H., Košnuth, J., Čnellárová, E., and Giovannini, A. (2009). Agrobacterium rhizogenes-mediated transformation of Hypericum tomentosum L. and Hypericum tetrapterum fries. *Zeitschrift fur Naturforsch. C* 64, 864–868. doi: 10.1515/ZNC-2009-11-1218/MACHINEREADABLECITATION/RIS

Kuhn, M. (2021). Caret: classification and regression training. Available at: https://cran.r-project.org/package=caret (Accessed August 22, 2022).

Kunjumon, R., Johnson, A. J., and Baby, S. (2022). Centella asiatica: secondary metabolites, biological activities and biomass sources. *Phytomedicine Plus* 2:100176. doi: 10.1016/j.phyplu.2021.100176

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22. Available at: https://cran.r-project.org/doc/Rnews/ (Accessed August 22, 2022).

Lütken, H., Meropi-Antypa, N., Kemp, O., Hegelund, J. N., and Müller, R. (2017). "Hairy root cultures of Rhodiola rosea to increase valuable bioactive compounds," in *Production of Plant Derived Natural Compounds Through Hairy Root Culture* ed. S. Malik (Switzerland: Springer), 65–88.

Makhzoum, A. B., Sharma, P., Bernards, M. A., and Trémouillaux-Guiller, J. (2013). Hairy roots: An ideal platform for transgenic plant production and other promising applications. *Phytochem. Plant Growth, Environ.* 42, 95–142. doi: 10.1007/978-1-4614-4066-6_6/COVER

Mangas, S., Moyano, E., Osuna, L., Cusido, R. M., Bonfill, M., and Palazón, J. (2008). Triterpenoid saponin content and the expression level of some related genes in calli of Centella asiatica. *Biotechnol. Lett.* 30, 1853–1859. doi: 10.1007/S10529-008-9766-6/FIGURES/5

Mano, Y., and Nemoto, K. (2012). The pathway of auxin biosynthesis in plants. *J. Exp. Bot.* 63, 2853–2872. doi: 10.1093/JXB/ERS091

Matveeva, T., and Otten, L. (2021). Opine biosynthesis in naturally transgenic plants: genes and products. *Phytochemistry* 189:112813. doi: 10.1016/J.PHYTOCHEM.2021.112813

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

Müller, M., and Munné-Bosch, S. (2011). Rapid and sensitive hormonal profiling of complex plant samples by liquid chromatography coupled to electrospray ionization tandem mass spectrometry. *Plant Methods* 7, 1–11. doi: 10.1186/1746-4811-7-37/FIGURES/4

Nambara, E. (2016). *Abscisic Acid*. 2nd ed. Amsterdam, Netherlands: Elsevier doi:10.1016/B978-0-12-394807-6.00098-8.

Nemoto, K., Hara, M., Suzuki, M., Seki, H., Oka, A., Muranaka, T., et al. (2009). Function of the aux and rol genes of the Ri plasmid in plant cell division in vitro. *Plant Signal. Behav.* 4, 1145–1147. doi: 10.4161/PSB.4.12.9904

Nordström, A., Tarkowski, P., Tarkowska, D., Norbaek, R., Åstot, C., Dolezal, K., et al. (2004). Auxin regulation of cytokinin biosynthesis in Arabidopsis thaliana: a factor of potential importance for auxin-cytokinin-regulated development. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8039–8044. doi: 10.1073/pnas.0402504101

Ozyigit, I. I., Dogan, I., and Artam Tarhan, E. (2013). "Agrobacterium rhizogenes-Mediated Transformation and Its Biotechnological Applications in Crops," in *Crop*

Improvement: New Approaches and Modern Techniques. eds. H. K. Rehman, A. Parvaiz and O. Munir (Boston, MA: Springer US), 1–48.

Palazon, J., Cusido, R. M., Gonzalo, J., Bonfill, M., Morales, C., and Pinol, M. T. (1998). Relation between the amount of rolC gene product and indole alkaloid accumulation in Catharanthus roseus transformed root cultures. *J. Plant Physiol.* 153, 712–718. doi: 10.1016/S0176-1617(98)80225-3

Palazón, J., Cusidó, R. M., Roig, C., and Pinol, M. T. (1997). Effect of rol genes from agrobacterium rhizogenes TL-DNA on nicotine production in tobacco root cultures. *Plant Physiol. Biochem.* 35, 155–162. Available at: https://agris.fao.org/agris-search/search.do?recordID=FR9702596 (Accessed April 18, 2022).

Pavlova, O. A., Matveyeva, T. V., and Lutova, L. A. (2014). rol-genes of agrobacterium rhizogenes. *Russ. J. Genet. Appl. Res* 4, 137–145. doi: 10.1134/S2079059714020063

Prakash, V., Jaiswal, N., and Srivastava, M. (2017). A review on medicinal properties of Centella asiatica characterization of Endophytic fungi of Mimusops elengi (Bakul) view project. *Artic. Asian J. Pharm. Clin. Res.* 10, 69–74. doi: 10.22159/ajpcr.2017.v10i10.20760

Prasad, A., Mathur, A. K., and Mathur, A. (2019). Advances and emerging research trends for modulation of centelloside biosynthesis in Centella asiatica (L.) urban: a review. *Ind. Crop. Prod.* 141:111768. doi: 10.1016/J.INDCROP.2019.111768

R Core Team (2021). R: a language and environment for statistical computing. Available at: https://www.r-project.org/ (Accessed August 22, 2022).

Rai, A., Saito, K., and Yamazaki, M. (2017). Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J.* 90, 764–787. doi: 10.1111/TPJ.13485

Roy, A. (2020). Hairy root culture an alternative for bioactive compound production from medicinal plants. *Curr. Pharm. Biotechnol.* 22, 136–149. doi: 10.2174/1389201021666201229110625

Roychowdhury, D., Basu, A., and Jha, S. (2015). Morphological and molecular variation in Ri-transformed root lines are stable in long term cultures of Tylophora indica. *Plant Growth Regul.* 75, 443–453. doi: 10.1007/s10725-014-0005-y

Ruslan, K., Selfitri, A. D., Bulan, S. A., Rukayadi, Y., and Elfahmi. (2012). Effect of agrobacterium rhizogenes and elicitation on the asiaticoside production in cell cultures of Centella asiatica. *Pharmacogn. Mag.* 8, 111–115. doi: 10.4103/0973-1296.96552

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 *Biotechnological Approaches for Medicinal and Aromatic Plants: Conservation, Genetic Improvement and Utilization*, ed. K. Nitish (Singapore: Springer), 27–51.

Shkryl, Y. N., Veremeichik, G. N., Bulgakov, V. P., Tchernoded, G. K., Mischenko, N. P., Fedoreyev, S. A., et al. (2008). Individual and combined effects of the rolA, B, and C genes on anthraquinone production in Rubia cordifolia transformed calli. *Biotechnol. Bioeng.* 100, 118–125. doi: 10.1002/BIT.21727

Sun, L. R., Wang, Y. B., He, S. B., and Hao, F. S. (2018). Mechanisms for abscisic acid inhibition of primary root growth. *Plant Signal. Behav.* 13, e1500064–e1500069. doi: 10.1080/15592324.2018.1500069

Sun, B., Wu, L., Wu, Y., Zhang, C., Qin, L., Hayashi, M., et al. (2020). Therapeutic potential of Centella asiatica and its Triterpenes: a review. *Front. Pharmacol.* 11:568032. doi: 10.3389/fphar.2020.568032

Takatsuka, H., and Umeda, M. (2014). Hormonal control of cell division and elongation along differentiation trajectories in roots. *J. Exp. Bot.* 65, 2633–2643. doi: 10.1093/jxb/ert485

Thwe, A., Arasu, M. V., Li, X., Park, C. H., Kim, S. J., Al-Dhabi, N. A., et al. (2016). Effect of different agrobacterium rhizogenes strains on hairy root induction and phenylpropanoid biosynthesis in tartary buckwheat (Fagopyrum tataricum Gaertn). *Front. Microbiol.* 7:318. doi: 10.3389/FMICB.2016.00318/BIBTEX

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics With S.*. New York: Springer. Available at: http://www.stats.ox.ac.uk/pub/MASS4 (Accessed August 22, 2022).

Wahby, I., Caba, J. M., and Ligero, F. (2012). Agrobacterium infection of hemp (Cannabis sativa L.): establishment of hairy root cultures. *J. Plant Interact.* 8, 312–320. doi: 10.1080/17429145.2012.746399

Wei, T., and Simko, V. (2021). R package "corrplot": visualization of a correlation matrix. Available at: https://github.com/taiyun/corrplot (Accessed August 22, 2022).

Yang, L., Zhang, J., He, J., Qin, Y., Hua, D., Duan, Y., et al. (2014). ABA-mediated ROS in mitochondria regulate root meristem activity by controlling PLETHORA expression in Arabidopsis. *PLoS Genet.* 10, e1004791. doi: 10.1371/journal.pgen.1004791

Zhao, H., Lai, Z., Leung, H., and Zhang, X. (2020). "Linear discriminant analysis," in *Feature Learning and Understanding: Algorithms and Applications* (Cham: Springer International Publishing), 71–85. doi: 10.1007/978-3-030-40794-0_5

70

Chapter 2. Insights into enhancing *Centella asiatica* organ cell biofactories via hairy root protein profiling.

Alcalde MA, Hidalgo-Martinez D, Bru-Martínez R, Sellés-Marchart S, Mercedes Bonfill M, & Palazon J. *Front. Plant Sci.* (2023). XX:XXXXXX. doi: XXXX/XXX (Accepted-In press).

Spanish summary

Las raíces transformadas son una plataforma biotecnológica prometedora debido a su potencial para la producción de metabolitos especializados y su rápido crecimiento. En este estudio, se utilizó el análisis de proteómica para obtener nuevos conocimientos sobre el desarrollo de las raíces transformadas, identificar biomarcadores y potencialmente optimizar la transformación genética para mejorar la producción de compuestos bioactivos en *Centella asiatica*. Este estudio involucró la categorización de líneas establecidas de raíces transformadas de *C. asiatica* en función de su capacidad de producción de centellósidos; clasificadas como ALTA, MEDIA o BAJA. Mediante un exhaustivo análisis proteómico sin marcaje, se identificaron un total de 213 proteínas cuantificables. La subsiguiente aplicación de análisis multivariados reveló diferencias y similitudes proteómicas distintas, especialmente entre la línea BAJA y las otras líneas de raíces transformadas. El análisis de red de proteínas reveló que estas macromoléculas se relacionan con procesos esenciales como la fotosíntesis y el metabolismo especializado. Se identificaron biomarcadores potenciales de diferentes rasgos, incluidas la proteína Tr-type G domain-containing y el alcohol deshidrogenasa en la línea ALTA. El hallazgo interesante se observó con la presencia de la ornitina ciclodeaminasa en las raíces transformadas. En base a lo anterior, se realizó la cuantificación de los números de copias de genes rol mediante qPCR, seguida de mediciones de expresión génica. Sin embargo, los resultados de este análisis cuantitativo contradicen una correlación uniforme con los niveles de expresión de los genes rol. Es particularmente destacable la expresión significativamente más elevada en la línea ALTA, específicamente en relación con el gen *rolD*. La asociación directa y la correlación de este gen con el biomarcador ornitina ciclodeaminasa afianzaron aún más esta conexión, evidenciando su posible importancia en la producción de centellósidos.

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

Miguel A. Alcalde Alvites[1, 2*], Diego A. Hidalgo Martinez[1], Roque Bru-Martinez[3], Susana Sellés-Marchart[3], Mercedes Bonfill[1], Javier Palazon[1*]

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

Scope Statement

The influence of rol genes in metabolism and physiology and the mechanism of how they act is still unclear. The manuscript is the continuation of a previous one titled as "Using machine learning to link the influence of transferred Agrobacterium rhizogenes genes to the hormone profile and morphological traits in Centella asiatica hairy roots" (https://www.frontiersin.org/articles /10.3389/fpls.2022.1001023/full) which we observed that there is great variability between the morphological parameters and centelloside production in different Centella asiatica hairy root lines. In this manuscript we determinate the alterations at the protein level and its impact in different hairy root lines, supplemented with the study of the expression and copy number of rol genes.

Conflict of interest statement

The authors declare a potential conflict of interest and state it below

The author(s) declared that they were not an editorial board member of Frontiers, at the time of submission.

CRedIT Author Statement

Diego Alberto Hidalgo Martinez: Conceptualization, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. **Javier Palazon**: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft. **Mercedes Bonfill**: Formal Analysis, Investigation, Supervision, Writing – original draft. **Miguel Angel Alcalde Alvites**: Conceptualization, Formal Analysis, Investigation, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing. **Roque Bru-Martinez**: Methodology, Resources, Writing – original draft. **Susana Sellés-Marchart**: Methodology, Resources, Writing – original draft.

Keywords

Rhizobium rhizogenes, biomarkers, Centelloside production, Molecular Farming, plant biotechnology

Abstract

Word count: 280

Recent advancements in plant biotechnology have led to the emergence of hairy roots as a promising biotechnological platform due to their potential for specialized metabolite production and rapid growth. In this study, proteomics analysis was used to obtain new insights into hairy root development, identify biomarkers, and potentially optimize genetic transformation for enhanced bioactive compound production in C.asiatica. This study involved the categorization of established C. asiatica hairy root lines based on their centelloside production capacity. These lines were classified as HIGH, MID, or LOW. Through meticulous label-free proteomic analysis, a total of 213 quantifiable proteins were successfully identified. The subsequent application of multivariate analysis elucidated distinct proteome differences and commonalities, particularly between the LOW line and other root lines. A protein network analysis revealed proteins related to essential processes such as photosynthesis and specialized metabolism. Potential biomarkers of different traits, were identified, including Tr-type G domain-containing protein and alcohol dehydrogenase in the HIGH group. Additionally, the control wild adventitious roots showcased a collection of proteins linked to photosynthesis. An intriguing finding arose from the presence of ornithine cyclodeaminase in the hairy roots. This protein serves as a traceable biomarker correlated with centelloside production capacity lines and serves as an indicator of the successful influence of Rhizobium-mediated genetic transformation. Based on the above, quantification of rol gene copy numbers through qPCR was undertaken, followed by gene expression measurements. However, the results of this quantitative analysis contradicted a uniform correlation with rol gene expression levels. Particularly noteworthy was the significantly higher expression observed in the HIGH line, specifically in relation to the rolD gene. The direct association and correlation of this gene with the ornithine cyclodeaminase biomarker further solidified this connection.

Funding information

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: No human studies are presented in the manuscript.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable images or data are presented in this study.

Data availability statement

Generated Statement: The authors acknowledge that the data presented in this study must be deposited and made publicly available in an acceptable repository, prior to publication. Frontiers cannot accept a manuscript that does not adhere to our open data policies.

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

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

6 [1]Department of Biology, Healthcare and the Environment. Faculty of Pharmacy and Food Sciences.
7 University of Barcelona. Spain

8 [2]Plant Proteomics and Functional Genomics Group, Department of Biochemistry and Molecular
9 Biology, Soil science and Agricultural Chemistry, Faculty of Science, University of Alicante,
10 Alicante, Spain

11 [3]Biotechnology, Health and Education Research Group, Posgraduate School, Cesar Vallejo
12 University, Trujillo, Peru.

14 * Correspondence:
15 Javier Palazon and Miguel Angel Alcalde
16 javierpalazon@ub.edu; miguel.psr.94@gmail.com

21 ABSTRACT

22 Recent advancements in plant biotechnology have led to the emergence of hairy roots as a promising
23 biotechnological platform due to their potential for specialized metabolite production and rapid
24 growth. In this study, proteomics analysis was used to obtain new insights into hairy root
25 development, identify biomarkers, and potentially optimize genetic transformation for enhanced
26 bioactive compound production in *C.asiatica*. This study involved the categorization of established
27 *C. asiatica* hairy root lines based on their centelloside production capacity. These lines were classified
28 as HIGH, MID, or LOW. Through meticulous label-free proteomic analysis, a total of 213
29 quantifiable proteins were successfully identified. The subsequent application of multivariate analysis
30 elucidated distinct proteome differences and commonalities, particularly between the LOW line and
31 other root lines. A protein network analysis revealed proteins related to essential processes such as
32 photosynthesis and specialized metabolism. Potential biomarkers of different traits, were identified,
33 including Tr-type G domain-containing protein and alcohol dehydrogenase in the HIGH group.
34 Additionally, the control wild adventitious roots showcased a collection of proteins linked to
35 photosynthesis. An intriguing finding arose from the presence of ornithine cyclodeaminase in the
36 hairy roots. This protein serves as a traceable biomarker correlated with centelloside production
37 capacity lines and serves as an indicator of the successful influence of *Rhizobium*-mediated genetic
38 transformation. Based on the above, quantification of *rol* gene copy numbers through qPCR was
39 undertaken, followed by gene expression measurements. However, the results of this quantitative

40 analysis contradicted a uniform correlation with *rol* gene expression levels. Particularly noteworthy
41 was the significantly higher expression observed in the HIGH line, specifically in relation to the *rolD*
42 gene. The direct association and correlation of this gene with the ornithine cyclodeaminase biomarker
43 further solidified this connection.

44

INTRODUCTION

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

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

58 The oleanyl cation, catalyzed by α/β-amyrine synthase, leads to the production of α or β-amyrin
59 (Azerad, 2016). After cyclization, further diversity in the resulting compounds is introduced through
60 diverse modifications which are facilitated by enzymes like cytochrome P450-dependent
61 monooxygenases and glycosyltransferases. UDP-glycosyltransferases (UGTs) play a key role in
62 glucosylating asiatic acid and madecassic acid to yield asiaticoside and madecassoside (Kim et al.,
63 2017).

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

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

77 To fully exploit the advantages offered by hairy roots, a comprehensive understanding of the intricate
78 molecular processes governing their development and metabolic capabilities is essential. One less-
79 utilized tool for achieving this understanding is proteomics, which entails studying the complete set
80 of proteins expressed by an organism and has revolutionized our comprehension of cellular processes
81 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.

108108

MATERIAL 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

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

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 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

167   standard curve. The average Ct values obtained from the four dilutions were utilized. These collected
168   values were then integrated into an equation, which was subsequently plotted. In the context of each
169   group of hairy root lines (HIGH, MID, and LOW), Cx and Cr represent the average Ct values
170   corresponding to the transgene and β-AS, respectively. Ix and Ir denote the intercepts associated with
171   the transgene and β-AS, while Sx and Sr signify the slopes for the transgene and β-AS, respectively.
172   To derive the copy number, the X/R value is multiplied by two.

173173

174   Gene expression

175   The *rol* genes expression in the transgenic lines was verified using quantitative real-time polymerase
176   chain reaction (qRT-PCR). Gene normalization was accomplished using the β-actin gene. Total RNA
177   was isolated from plant material utilizing TRIzol reagent (Invitrogen, Carlsbad, CA). For the qRT-
178   PCR, cDNA was synthesized from RNA treated with DNase I (Invitrogen, Carlsbad, CA) using
179   SuperScript IV reverse transcriptase (Invitrogen, Carlsbad, CA) according to the manufacturer's
180   instructions. The qRT-PCR assays were performed employing the iTAqTM Universal SYBR Green
181   Supermix (Bio-Rad, Hercules, CA, USA) in the QuantStudio3 System (Thermo Fisher). Each sample
182   was analyzed in triplicate under the following conditions: an initial step at 95°C for 60 s, followed by
183   40 cycles of denaturation at 95°C for 10 s, annealing at 60°C for 20 s, and extension at 72°C for 30
184   s. Subsequent to amplification, a melting curve analysis was conducted. To check reproducibility,
185   each assay was performed with technical triplicates for each of the three biological samples. Gene-
186   specific primers were designed using Primer-BLAST (Table 1).

187187

188   Label-free proteomic analysis

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

196   Peptide separation was performed at 50 °C using a 140-minute linear gradient of 3-40 % ACN in 0.1
197   % formic acid at a flow rate of 0.400 mL/min. Source parameters included a gas temperature of
198   250°C, drying gas at 14 L/min, nebulizer at 35 psi, sheath gas temperature at 250°C, sheath gas flow
199   at 11 L/min, capillary voltage at 3,500 V, and fragmentor at 360 V. Data were acquired in positive-
200   ion mode using Agilent MassHunter Workstation Software, LC/MS Data Acquisition B.08.00 (Build
201   8.00.8058.0). Operating in high sensitivity mode, MS and MS/MS data were collected in Auto
202   MS/MS mode. This involved selecting the 20 most intense parent ions (charge states from 2 to 5)
203   within the 300 to 1,700 m/z mass range, provided they exceeded a threshold of 1,000 counts, for
204   subsequent MS/MS analysis. MS/MS spectra spanning the 50-1,700 m/z range were gathered with
205   the quadrupole set to "narrow" resolution. Data acquisition was continued until either a total count of
206   25,000 was reached or a maximum accumulation time of 333 ms was achieved.

207   Each MS/MS spectrum was subjected to preprocessing using the extraction tool within the Spectrum
208   Mill Proteomics Workbench (Agilent). This step aimed to generate a peak list and enhance spectral
209   quality by merging MS/MS spectra sharing the same precursor (with a Δm/z < 1.4 Da and

210 chromatographic Δt < 15s). The resulting refined dataset was then subjected to a search against the
211 proteome database, encompassing primary species from the *Apiaceae* family and the
212 *Rhizobium*/*Agrobacterium* genus. Additionally, the analysis included identification of contaminant
213 proteins using the identity mode of the MS/MS search tool in the Spectrum Mill Proteomics
214 Workbench, configured as follows: trypsin enzyme specificity, allowance for up to 2 missed
215 cleavages, fixed modification of Cys by carbamidomethylation, variable modification of Met by
216 oxidation, and mass tolerance of 20 ppm for precursor ions and 50 ppm for product ions. The peptide
217 hits obtained were subjected to filtering, retaining those with a score of ≥ 6 and a percent scored peak
218 intensity (%SPI) of ≥ 60.

219 The LC-MS raw files were imported into Progenesis QI for Proteomics (Nonlinear Dynamics) version
220 4.0, a label-free analysis software. Quantification was based on MS1 intensity. The data file with the
221 highest number of features (peaks) served as a reference for aligning the retention times of all other
222 chromatographic runs and for normalizing MS feature signal intensity (peak area). To address
223 experimental variations, a robust distribution of all ratios (log(ratio)) was computed for correction
224 purposes. MS features were filtered to encompass only those with a charge state ranging from two to
225 five. Employing a "between subjects" experimental design mode, samples were clustered according
226 to their respective experimental groups (Adv, LOW, MID, and HIGH). Average intensity ratios of
227 matched features across experimental groups, along with p-values from one-way ANOVA, were
228 automatically computed.

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

236236

237 Statistical Analysis and protein network.

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

246 Statistical analyses to assess differences in protein levels were conducted using GraphPad Prism 7.0
247 software. Analysis of variance (ANOVA) was employed, followed by Tukey's post hoc test to
248 determine significant variations between protein levels. To discern pairwise differences between
249 means, the Tukey-Kramer multiple-comparison test was utilized, with a significance threshold set at
250 $p < 0.05$. To predict protein-protein interactions, a list of protein identifiers was submitted to the web
251 interface of the STRING. (2023). https://string-db.org/ [Accessed July 27, 2023].

252252

253 RESULTS

254 Proteomic profiles in hairy and adventitious roots

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

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

265 As a first approximation after this data filtering, an OPLS-DA was carried out with the 213 proteins
266 quantified with the *Apiaceae* database. The score scatter plot of this model (Fig. 1) displays the
267 variation between the groups, with components 1 (48.7%) and 2 (27.9%) representing the maximum
268 separation. The model provided a satisfactory explanation of the variation, showing a good fit with
269 an R2X (cum) value of 0.786. Further, the reliability of the model was confirmed by cross-validation,
270 resulting in a Q2 (cum) value of 0.861. Interestingly, the proteomes of HIGH and MID roots exhibited
271 minimal differences, whereas both differed significantly from the ADV and LOW proteomes.
272 Notably, the most pronounced difference was observed between the LOW group and the others.

273 The list of quantified plant proteins was analyzed using STRING. Out of 213 proteins, 184 were
274 found in the STRING database and utilized to construct a network, displaying significant interactions
275 (refer to Figure S1). The identified proteins were further classified based on their biological processes
276 (Gene Ontology) and KEGG Pathways (see Table S2). The classification revealed the recurrence of
277 proteins associated with essential processes such as photosynthesis and amino acid biosynthesis.
278 Additionally, proteins related to pathways of specialized metabolites, including phenylpropanoids,
279 were also prominent.

280 In the next step, ANOVA and Tukey tests were conducted to identify proteins with significant
281 differences among the different root lines. The comparison between the HIGH and LOW lines
282 revealed that only 44 proteins exhibited statistical variations (Table S3). Out of these, only 38 could
283 be utilized to construct a protein network using STRING (Figure 2). Surprisingly, the majority of the
284 differentially expressed proteins were found in higher abundance in the LOW hairy roots, and only a
285 few were found in greater quantities in the HIGH group. The differentially expressed proteins are as
286 follows: alcohol dehydrogenase, UDP-arabinopyranose mutase, Tr-type G domain-containing
287 protein, plant heme peroxidase, D-3-phosphoglycerate dehydrogenase, and ketol-acid
288 reductoisomerase.

289 The proteins from this network were further classified based on their biological processes and KEGG
290 pathways (Table S4). This classification revealed the absence of proteins associated with
291 photosynthesis, previously found when the Adv roots were included in the multivariable analysis.
292 Additionally, proteins related to pantothenate, CoA, and phenylpropanoid and amino acid
293 biosynthesis received good scores in the classification.

294294

295 **Biomarker discovery**

296 To identify significant markers among the Adv and hairy root groups, an OPLS-DA analysis was
297 conducted for each comparison using the 213-protein dataset mentioned above. To aid the
298 visualization of the discrimination model in terms of biomarkers, an S-plot (Wiklund et al., 2008) was
299 utilized to filter potential proteins. The S-plot of the Adv vs HIGH model (Figure 3A) illustrates the
300 magnitude (modeled covariation) and reliability (modeled correlation) of each protein. Putative
301 biomarkers were identified based on a small set of proteins exhibiting high magnitude ($\geq |0.1|$) and
302 reliability ($\geq |0.8|$). In this specific model, we identified only two biomarkers positively correlated
303 with the HIGH group: a Tr-type G domain-containing protein and an alcohol dehydrogenase protein.
304 In contrast, 13 proteins were correlated with the Adv group (Table S5), most of them related to
305 photosynthesis.

306 The S-plot analysis of the Adv vs MID model identified alcohol dehydrogenase protein as a putative
307 biomarker, showing a significant correlation with the MID group. Intriguingly, the clearest distinction
308 between two groups was observed in the Adv vs LOW model. At least six distinct proteins were found
309 to be positively correlated with the LOW roots, including the PCMH-type FAD-binding domain,
310 eukaryotic translation initiation factor, peroxidase, and cysteine protease.

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

316 Finally, we conducted a proteomic analysis by searching against the Uniprot databases for
317 *Rhizobium/Agrobacterium* taxonomy, as described in the Material and Methods section. A total of
318 100 quantifiable proteins were identified and carefully selected based on the specific criteria
319 described above. This dataset was then used for an OPLS-DA analysis comparing Adv and HIGH
320 lines, followed by an S-plot to identify potential biomarkers. The proteins showing high magnitude
321 ($\geq |0.1|$) and reliability ($\geq |0.8|$) were manually curated to ensure that only proteins potentially
322 originating from *Rhizobium* via T-DNA were included, filtering out any proteins that could be derived
323 from other sources, to ensure the selection of true *Rhizobium* biomarkers. Consequently, only the
324 ornithine cyclodeaminase protein (OCD) complied with the curation process.

325 In *R. rhizogenes*, OCD is located in the T-DNA region of the Ri plasmid and is referred to as RolD
326 by Trovato et al. (2001). The presence of OCD in the hairy roots is noteworthy, as this enzyme plays
327 a crucial role in synthesizing proline from ornithine in a single step. According to Trovato et al.
328 (2018), this metabolic capability could be functionally involved in the process of root elongation
329 and/or maturation. The production of higher amounts of proline, an important osmolyte and signaling
330 molecule, may contribute to stress tolerance and growth regulation in the hairy roots, making it a
331 significant biomarker of *Rhizobium*-mediated genetic transformation.

332

333 **Impact of transgene copy number on gene expression.**

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

85

338 The transgene copy number for all transgenic lines was determined using qPCR relative to the
339 endogenous reference gene β-amyrin synthase (β-AS), a gene involved in the biosynthesis of
340 centellosides (Kim et al., 2005). The quantifiable parameters collected are shown in Table S6. After
341 calculations, all the tested lines were estimated to have two copies of each transgene (see Table 2).
342 Nevertheless, upon scrutinizing the gene expression patterns, clear distinctions emerged between the
343 groups (Figure 4). In particular, the HIGH line stands out due to heightened expression levels of most
344 *rol* genes, especially the *rol*D gene, which was found to be 5 to 8 times more pronounced than in the
345 LOW and MID lines, respectively. These differences suggest that the presence of two copies of the
346 transgenes does not necessarily correlate with uniform expression levels. Furthermore, the greater
347 abundance of *rol*D gene transcripts appears to correlate with the OCD biomarker, suggesting its
348 potential significance.

349 349

350 DISCUSSION:

351 The literature contains numerous examples of plant biofactories based on hairy root cultures
352 designed to produce plant bioactive compounds scarcely synthesized in nature (Hasnain et al., 2022;
353 Bapat et al., 2023; Sonkar et al., 2023). Most of these studies adopt empirical approaches,
354 concentrating on establishing hairy root cultures to explore their production of specialized plant
355 compounds and optimizing the biotechnological production system. In contrast, only a limited body
356 of research has taken a rational approach, trying to understand how a specific set of genes from *R.*
357 *rhizogenes* (*rol* genes) can modify cell growth and metabolism (Mauro et al., 2017; Sarkar et al.,
358 2018).

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

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

373 Despite the care taken in cultivating the Adv roots, proteins related to photosynthesis were
374 surprisingly evident in their protein profiles, the levels being significantly higher than in the
375 transgenic roots. This discrepancy led us to investigate further and prompted a comparison between
376 the HIGH, MID, and LOW lines. The results of this analysis proved to be particularly relevant in our
377 quest to unravel the protein networks that significantly impact the performance of hairy root lines in
378 biotechnological applications. The comparisons revealed several key biomarkers, notably the Tr-type
379 G domain-containing protein, alcohol dehydrogenase protein, and Bet v I/Major latex protein.

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

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

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

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

404   Although numerous authors have suggested that the *rol*D gene of the *R. rhizogenes* T-DNA
405 does not play a significant role in hairy root induction, our results contradict this hypothesis (Trovato
406 et al., 1997; Bulgakov, 2008). The detection of the OCD biomarker indicates a potential connection
407 between *Rhizobium*-mediated genetic alteration and the proteome composition of hairy roots. The
408 enzymatic transformation of ornithine into proline through the OCD-like function of rolD may
409 provide a credible rationale for its involvement in the generation of hairy roots (Trovato et al., 2008,
410 2018). It is worth noting that prior studies have reported a significant escalation in proline levels
411 within the growth region of primary maize roots under conditions of limited water availability,
412 implying a crucial function of proline synthesis in maintaining root growth (Verslues and Sharp,
413 1999). An elevated concentration of proline has the potential to influence the synthesis of
414 hydroxyproline-rich glycoproteins (HRGPs), encompassing extensins and arabinogalactan proteins,
415 which serve as integral structural constituents of the plant cell wall (Okumoto et al., 2015). HRGPs
416 are thought to oversee essential processes such as cell division, the self-assembly of the cell wall, and
417 cell elongation, which could contribute to the noted impacts of RolD on root growth (Trovato et al.,
418 2001). Alternatively, the promotion of root growth by *rolD* could also be associated with the reduction
419 of ornithine, thus impacting the polyamine reservoir, where ornithine operates as a precursor. The
420 overexpression of arginine decarboxylase, another polyamine precursor, has been demonstrated to
421 increase putrescine levels and hinder root growth in tobacco plants (Masgrau et al., 1997). Further
422 research into the specific role of OCD protein in root development and stress responses could enhance
423 our knowledge of the mechanisms underlying hairy root formation. This would potentially open new
424 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.

460460

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary

material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

M.B., J.P., M.A. and D.H. designed the research. M.A., R.B., S.S. determined the proteomics profiles. M.A. and D.H. build machine learning models. M.A. Performed tissue culture, genomic and transcriptomics analysis. All the authors analyzed the results and contributed to the final version of the manuscript.

469469

478478

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, 1001023. 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 Science* 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. *International Journal of Molecular Sciences 2020, Vol. 21, Page 2873* 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

89

SmWRKY61 on tanshinone biosynthesis in *Salvia miltiorrhiza* and *Salvia castanea*. *Front Plant Sci* 12, 820582. 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. *Scientific Reports 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. *Applied Microbiology and Biotechnology 2020 104:20* 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. 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. *Phytochemistry Reviews 2017 17:1* 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, 504243. 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, 1009395. 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, 1208218. 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. 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 and Organ Culture (PCTOC) 2017 131:2* 131, 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, 544–544. 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. *Journal of Chromatography 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. *Biotechnological Approaches for Medicinal and Aromatic Plants: Conservation, Genetic Improvement and Utilization*, 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.

582 Shilpha, J., Largia, M. J. V., Kumar, R. R., Satish, L., Swamy, M. K., and Ramesh, M. (2023).
583   Hairy Root Cultures: A Novel Way to Mass Produce Plant Secondary Metabolites.
584   *Phytochemical Genomics: Plant Metabolomics and Medicinal Plant Genomics*, 417–445.
585   doi: 10.1007/978-981-19-5779-6_17.

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

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

591 Sonkar, N., Shukla, P. K., and Misra, P. (2023). Plant Hairy Roots as Biofactory for the
592   Production of Industrial Metabolites. *Plants as Bioreactors: An Overview*, 273–297. doi:
593   10.1002/9781119875116.CH11.

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

598 Strommer, J. (2011). The plant ADH gene family. *Plant J* 66, 128–142. doi: 10.1111/J.1365-
599   313X.2010.04458.X.

600 Triba, M. N., Le Moyec, L., Amathieu, R., Goossens, C., Bouchemal, N., Nahon, P., et al.
601   (2015). PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on
602   the K-fold cross-validation quality parameters. *Mol Biosyst* 11, 13–19. doi:
603   10.1039/C4MB00414K.

604 Trovato, M., Maras, B., Linhares, F., and Costantino, P. (2001). The plant oncogene *rolD*
605   encodes a functional ornithine cyclodeaminase. *Proc Natl Acad Sci U S A* 98, 13449–
606   13453. doi: 10.1073/pnas.231320398.

607 Trovato, M., Mattioli, R., and Costantino, P. (2008). Multiple roles of proline in plant stress
608   tolerance and development. *Rendiconti Lincei* 19, 325–346. doi: 10.1007/S12210-008-
609   0022-8.

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

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

615 Veena, V., and Taylor, C. G. (2007). *Agrobacterium rhizogenes*: Recent developments and
616   promising applications. *In Vitro Cellular and Developmental Biology - Plant* 43, 383–403.
617   doi: 10.1007/S11627-007-9096-8.

618 Veenstra, T. D. (2021). Omics in Systems Biology: Current Progress and Future Outlook.
619   *Proteomics* 21, 2000235. doi: 10.1002/PMIC.202000235.

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

623 Wang, W., Vignani, R., Scali, M., and Cresti, M. (2006). A universal and rapid protocol for
624   protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis*
625   27, 2782–2786. doi: 10.1002/ELPS.200500722.

626 Wiklund, S., Johansson, E., Sjöström, L., Mellerowicz, E. J., Edlund, U., Shockcor, J. P., et al.
627   (2008). Visualization of GC/TOF-MS-based metabolomics data for identification of
628   biochemically interesting compounds using OPLS class models. *Anal Chem* 80, 115–122.
629   doi: 10.1021/ac0713510.

630 Xu, B., Liu, L., and Song, G. (2022). Functions and Regulation of Translation Elongation
631   Factors. *Front Mol Biosci* 8. doi: 10.3389/FMOLB.2021.816398.

632 Yang, L., Ding, J., Zhang, C., Jia, J., Weng, H., Liu, W., et al. (2005). Estimating the copy
633   number of transgenes in transformed rice by real-time quantitative PCR. *Plant Cell Rep*
634   23, 759–763. doi: 10.1007/S00299-004-0881-0.

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

638

639 **FIGURE LEGENDS.**

640 Figure 1. Score scatter plot of the OPLS-DA model conducted using the quantified data of 213
641 proteins from the *Apiaceae* database for the various root lines.

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

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

652 Figure 4. Normalized gene expression values from the transgenic lines: HIGH, MID, and LOW.
653   Asterisks indicate statistical differences among the lines solely for the *rol*D gene ($\alpha = 0.05$). Data
654 represent the mean $\pm$ SD of three replicates.

655

656 **TABLES**

657 **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

658

659

660

661

662

663 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

664

Figure 1.JPEG

Figure 2.JPEG
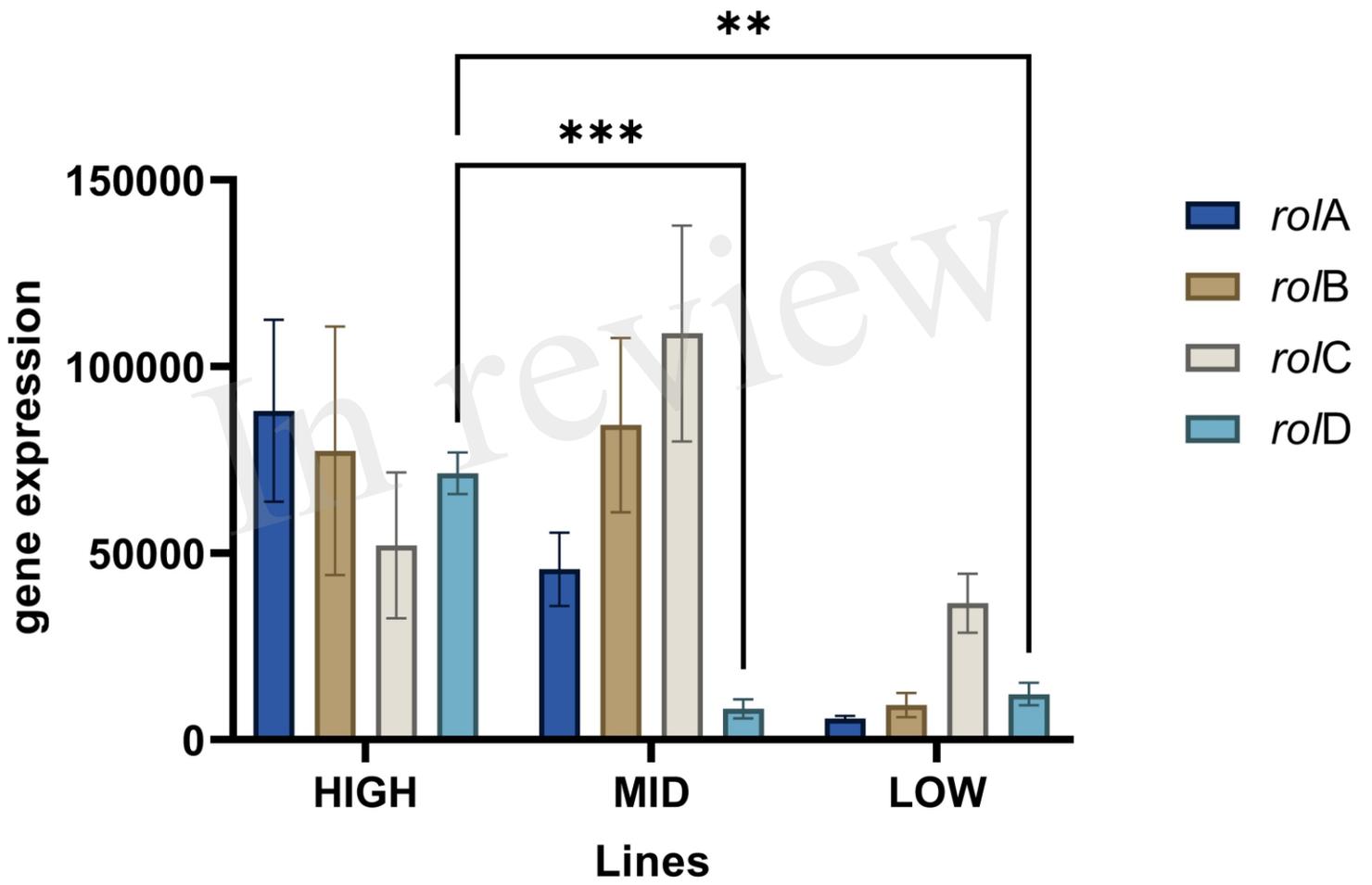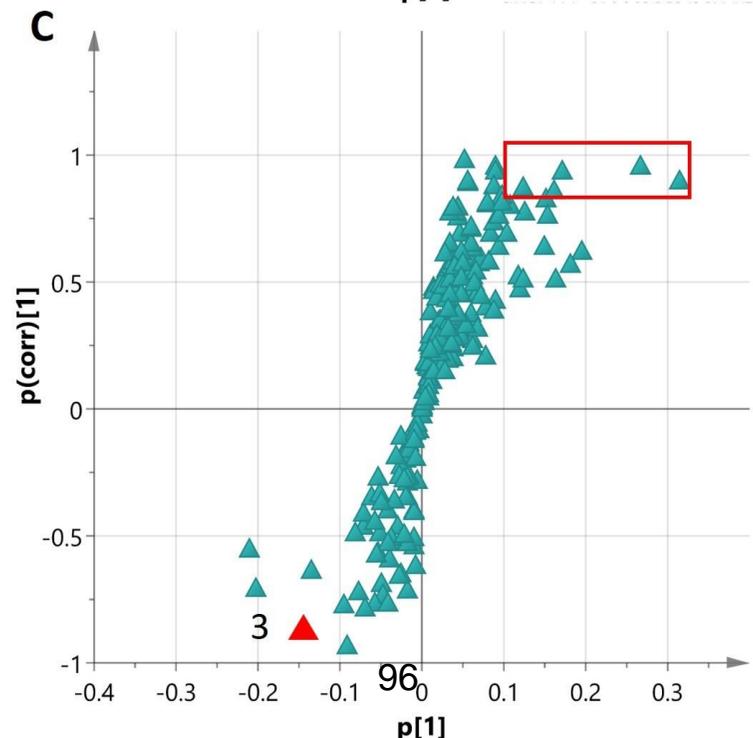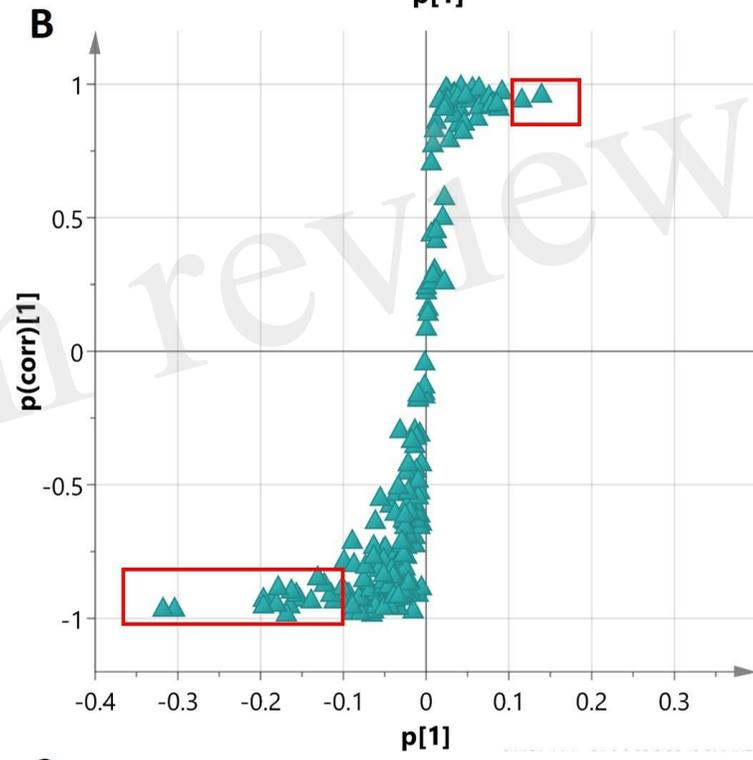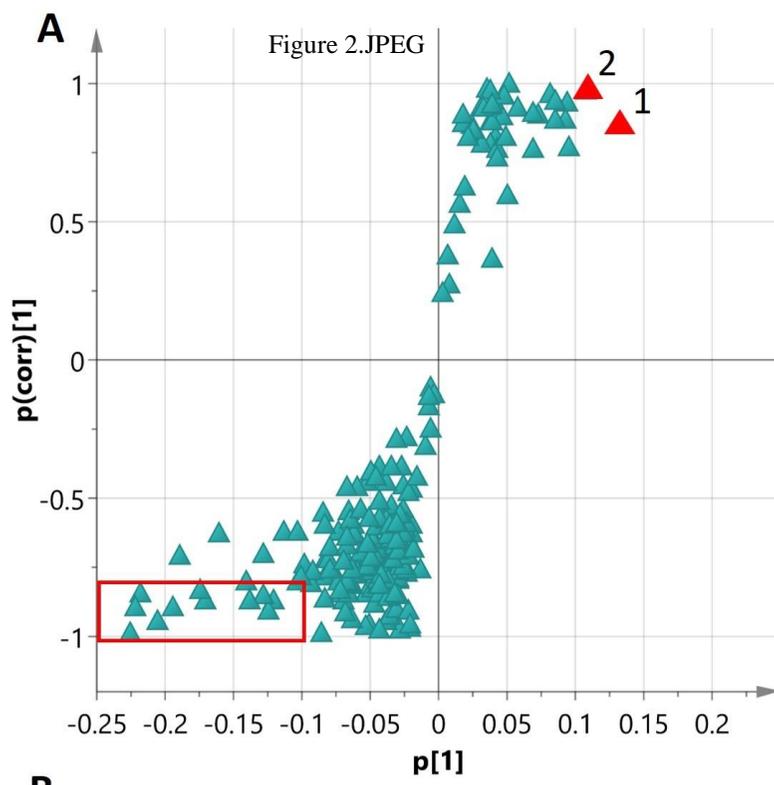
Figure 3.JPEG
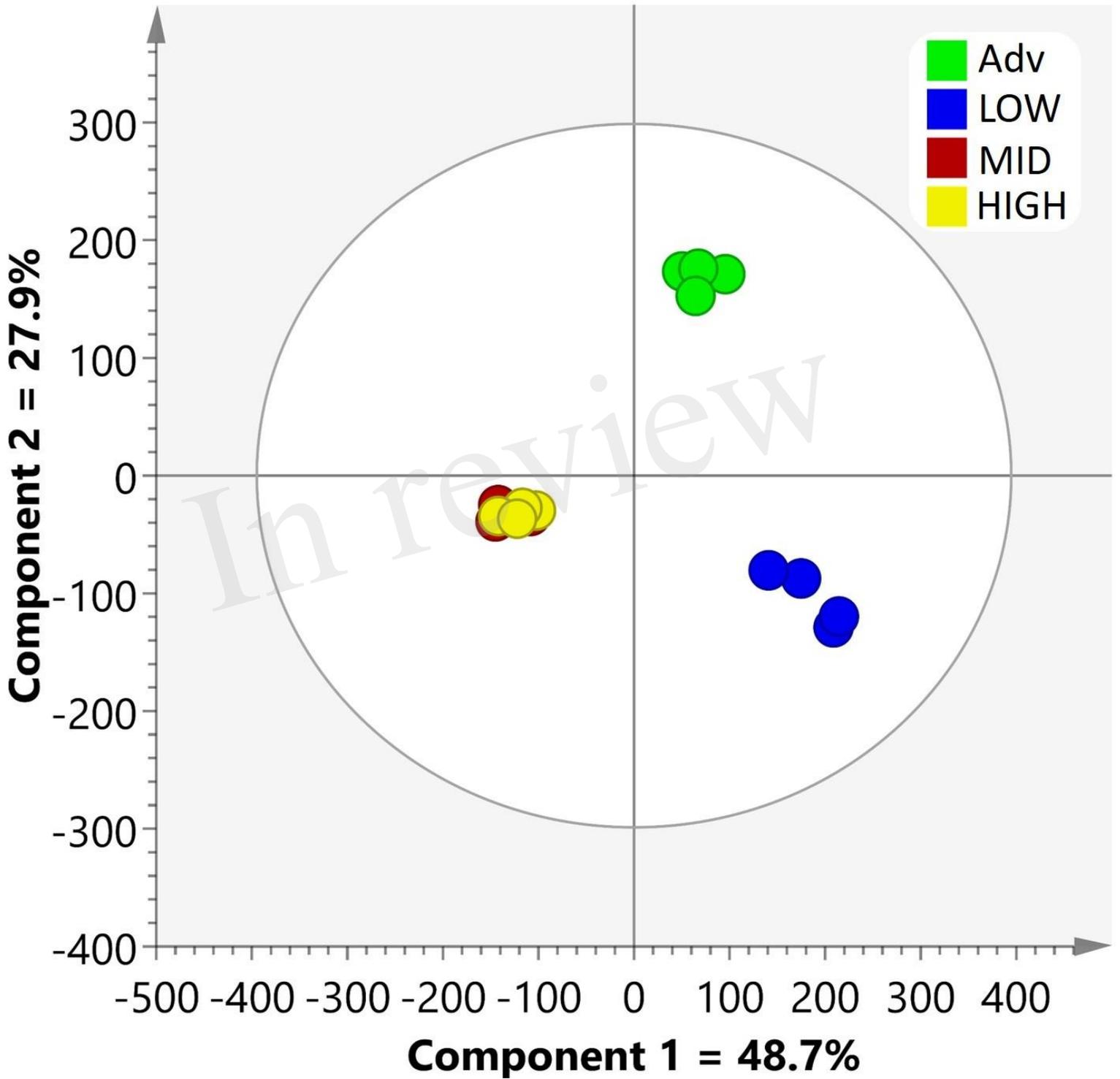
Figure 4.JPEG
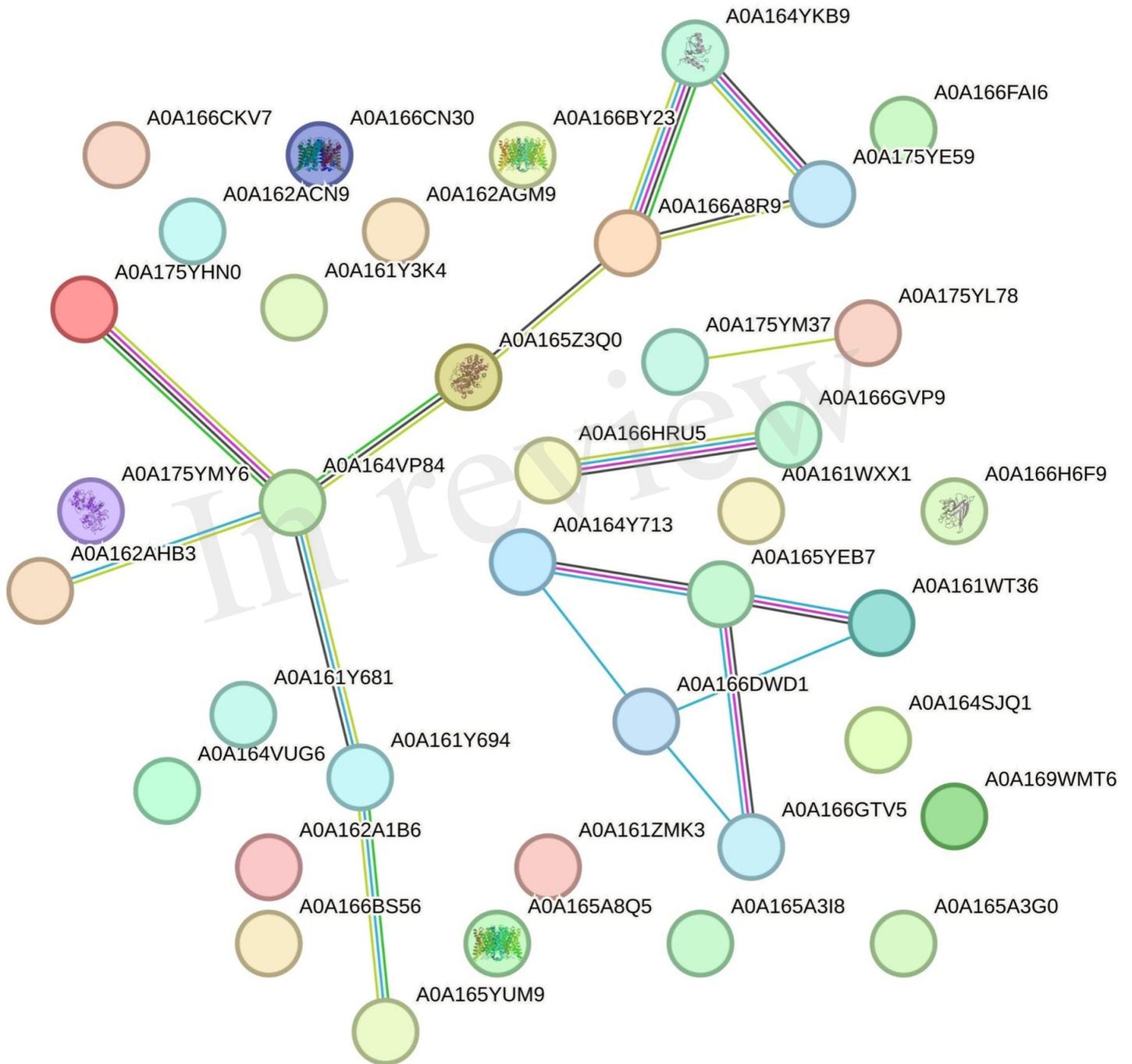
Chapter 3. Metabolic gene expression and centelloside production in elicited *Centella asiatica* hairy root cultures.

Centella asiatica hairy roots
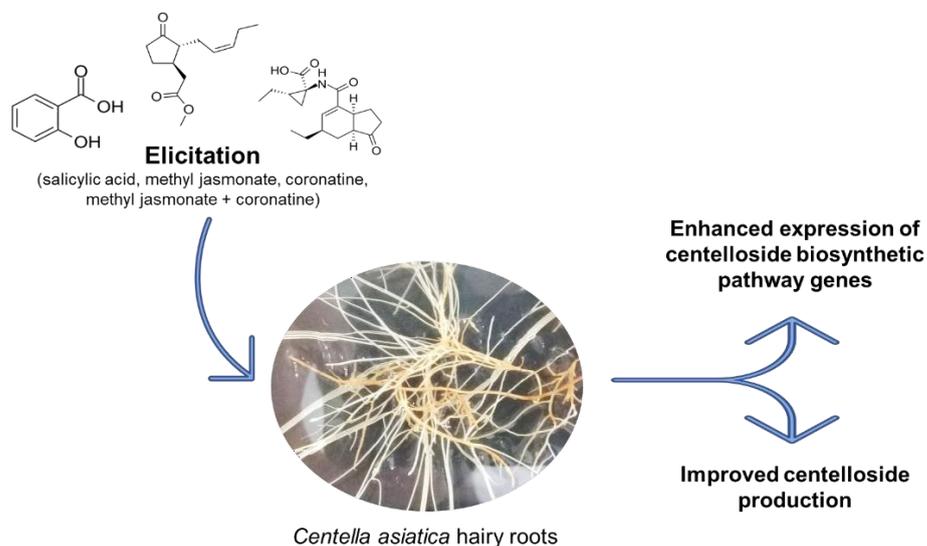
Spanish summary

Para la industria farmacológica y cosmética los centellósidos más importantes son el madecasósido y el asiaticósido, que se producen en cantidades muy bajas de manera natural en *Centella asiatica* (L.) Urban, llevando a la sobreexplotación de este recurso. Con el objetivo de superar esta limitación, en este trabajo comparamos el efecto de diferentes elicitores en la producción de centellósidos en cultivos de raíces transformadas establecidos mediante la infección de *Rhizobium rhizogenes*. Además, se estudiaron los cambios inducidos por los elicitores en la expresión de genes biosintéticos clave para evidenciar relaciones sobre la regulación de la vía metabólica de triterpenos. Los elicitores probados fueron la coronatina y el metil jasmonato, añadidos por separado o juntos, y el ácido salicílico. El contenido de los cuatro centellósidos principales (asiaticósido, madecasósido, ácido asiático y ácido madecásico) se determinó mediante cromatografía liquida de alta resolución, y el nivel de expresión de los genes después de la elicitación se analizó mediante reacción en cadena de polimerasa cuantitativa en tiempo real. El mayor aumento en la producción de centellósidos, especialmente madecasósido, se logró con la coronatina, aplicada sola o con el metil jasmonato. Este tratamiento también elevó la expresión de los genes de la ruta metabólica de los centellósidos, particularmente a las primeras horas posteriores a la elicitación. Los genes más altamente expresados estuvieron involucrados en las oxidaciones, lo que indica que los elicitores estudiados actuaron específicamente sobre los genes de los pasos finales de la vía biosintética de los centellósidos. Estos resultados respaldan que las raíces transformadas son una plataforma biotecnológica prometedora para mejorar la producción de centellósidos.
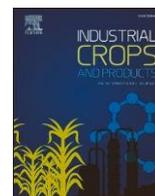
Metabolic gene expression and centelloside production in elicited *Centella asiatica* hairy root cultures

Miguel Angel Alcalde [a], Rosa Maria Cusido [a], Elisabeth Moyano [b], Javier Palazon [a], Mercedes Bonfill [a,*]

[a] *Laboratori de Fisiologia Vegetal, Facultat de Farmacia, Universitat de Barcelona, Av. Joan XXIII 27-31, 08028 Barcelona, Spain*
[b] *Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Avda. Dr. Aiguader 80, 08003 Barcelona, Spain*

A B S T R A C T

The most economically important centellosides, madecassoside and asiaticoside, are produced in very low amounts in *Centella asiatica* (L.) Urban roots. With the aim of overcoming this limitation, in this work we compared the effect of different elicitors on centelloside production in hairy root cultures established by *Agrobacterium rhizogenes* infection. Additionally, elicitor-induced changes in the expression of key biosynthetic genes were studied to shed light on the regulation of the triterpene metabolic pathway. The elicitors tested were coronatine and methyl jasmonate, added separately or together, and salicylic acid. The content of the four main centellosides (asiaticoside, madecassoside, asiatic acid and madecassic acid) was determined by HPLC/DAD, and the expression level of genes after elicitation was analyzed by real-time quantitative polymerase chain reaction. The greatest increase in the production of centellosides, especially madecassoside, was achieved with coronatine, applied alone or with methyl jasmonate. This treatment also enhanced the expression of the target genes, particularly at the start of elicitation. By far the most highly expressed were those involved in oxidations, indicating that the tested elicitors did not act specifically on key genes in the centelloside biosynthetic pathway. These results support that hairy roots are a promising biotechnological platform for improved centelloside production and that this approach warrants further research.

1. Introduction

Centella asiatica (L.) Urban is a herbaceous, perennial plant species of the Apiaceae (Umbelliferae) family. Requiring a humid climate, it grows up to 1800 m of altitude in tropical and subtropical regions such as southeastern Asia (Azerad, 2016). Reported to have neuroprotective, antioxidant, antidiabetic, antimicrobial, antitumor and antidepressant properties, among others, *C. asiatica* has been used in a wide range of medicinal and cosmetic applications.

The main bioactive compounds in *C. asiatica* are pentacyclic triterpenoid saponins called centellosides. The centelloside pathway is initiated *via* the mevalonate pathway to generate a sesquiterpene precursor, farnesyl diphosphate (FPP), for subsequent formation of an intermediate, squalene, by squalene synthase (James and Dubery, 2009). Squalene is then oxidized to 2,3-oxidosqualene, a branch point in the sterol and triterpenoid saponin biosynthesis, which cyclizes to a protosteryl or dammarenyl cation. Cyclization through a dammarenyl cation

generates, among others, an oleanyl cation that, finally catalyzed by α/β-amyrine synthase, will form α or β-amyrin (Azerad, 2016). The last steps of centelloside pathway have not yet been fully elucidated. Following cyclization, additional diversity is conferred by modification of the products by oxidation, hydroxylation, glycosylation, and other substitutions mediated by cytochrome P450-dependent mono-oxygenases, glycosyltransferases, and other enzymes. Asiatic acid and madecassic acid are glucosylated by UDP-glycosyltransferases (UGTs) to obtain asiaticoside and madecassoside (O.T. Kim et al., 2017). These highly specialized secondary metabolites form part of the plant defense system against pathogens and herbivores (Ramirez-Estrada et al., 2016). The biological activities of *C. asiatica* are mainly attributed to the pentacyclic triterpenoid saponins madecassoside and asiaticoside, and their corresponding sapogenins, madecassic acid and asiatic acid. Madecassoside has anti-inflammatory properties and significantly increases collagen III secretion, while asiaticoside stimulates wound healing and is useful in the treatment of leprosy (Singh et al., 2015).

In the last years, there has been growing interest in the therapeutic potential of *C. asiatica* extracts containing specific proportions of centellosides (Wen Su, 1995). Subsequent over-exploitation of wild-growing plants and excessive uprooting has brought *C. asiatica* to the brink of extinction (Mangas et al., 2008). As the chemical synthesis of centellosides is either extremely difficult or economically unviable, attention has turned to plant cell culture technology (plant biofactories) in the search for alternative systems of efficient secondary metabolite production (Verpoorte, 1999). In particular, hairy roots have been used as a biotechnological platform to produce plant bioactive compounds that are scarce in nature and whose complex structures make them difficult to synthesize. The advantages of hairy roots include genetic stability, hormonal independence, rapid growth and the capacity to produce the same metabolite spectrum as the roots of the whole plant (Gallego et al., 2014).

Endogenous levels of plant secondary metabolites are influenced by multiple environmental stresses and signals classified as biotic or abiotic (Gandhi et al., 2015). In plant *in vitro* cultures, it is possible to activate the secondary metabolism response to stress factors by the exogenous addition of biotic and abiotic elicitors (*i.e.*, phytohormones or plant pathogen particles) (Cusido et al., 2014), an approach commonly employed to boost metabolite production in plant-based production platforms (Thakur and Sohal, 2013). Furthermore, supplementing the media of a plant *in vitro* culture with elicitors under controlled conditions and observing the response in terms of gene expression can provide new insights into metabolite biosynthesis (Pauwels et al., 2009). Elicitors can be abiotic or biotic, the latter being the ones that have presented the best results to increase the production of secondary metabolites (Ramirez-Estrada et al., 2016). Among biotic elicitors, 100 mM methyl jasmonate (MeJa) has shown great results in achieving high terpene content in different cells and organ cultures *in vitro*. Thus, centelloside production in cell suspension cultures of *C. asiatica* was significantly increased using 100 µ M MeJa (Bonfill et al., 2011). On the other hand, 100 µ M MeJa was the best elicitor to increase the production of ginsenosides in hairy roots of *Panax ginseng* (Kim et al., 2013) and that od the diterpene tanshinone in hairy roots of *Salvia miltiorrhiza* (Liang et al., 2012; Kai et al., 2012). Coronatine (COR) is a polyketide phytotoxin produced by microbes. In some cases, it has shown greater activity as elicitor than Meja, such as in the production of sakuranetin and momilactone A in rice leaves (Tamogami and Kodama, 2000), or in cell suspension cultures of *Taxus media*, where 1 µM Coro had a activating effect on the production of taxanes (Onrubia et al., 2013). Vaccaro et al. (2017) obtained the best content of abietane diterpenes in *Salvia sclarea* hairy roots elicited with 100 µM MeJa or with 1 µM CORO. Salicylic acid (SA), a small molecule with a vital role in plant defense regulatory systems, has also shown interesting results, such as increased ginsenoside production after elicitation of *Panax ginseng* adventitious roots with 100 µM SA (Tewari and Paek, 2011). On the other hand, in *Cichorium intybus* hairy root cultures, sesquiterpene lactones production increased after elicitation with 100 µM SA (Malarz et al., 2007), similar results were observed on the diterpenoid andrographolide content in *Andrographis paniculate* hairy roots elicited with 100 mM SA (Sharmila and Subburathinam, 2013).

To improve centelloside production, more knowledge is required of their complex biosynthetic pathways (Fig. 1). Although many centelloside structures have been elucidated (Azerad, 2016), little is known about their metabolism, and only a few of the biosynthetic genes have been sequenced and cloned.

In this work it has been determined the effect of different elicitors on the biotechnological production of the main centellosides in a selected fast-growing *C. asiatica* hairy root line and on the expression levels of genes encoding key enzymes in the triterpene biosynthetic pathway, analyzed by real-time quantitative polymerase chain reaction (RT-qPCR) assays. The aim was to identify which genes are essential for the
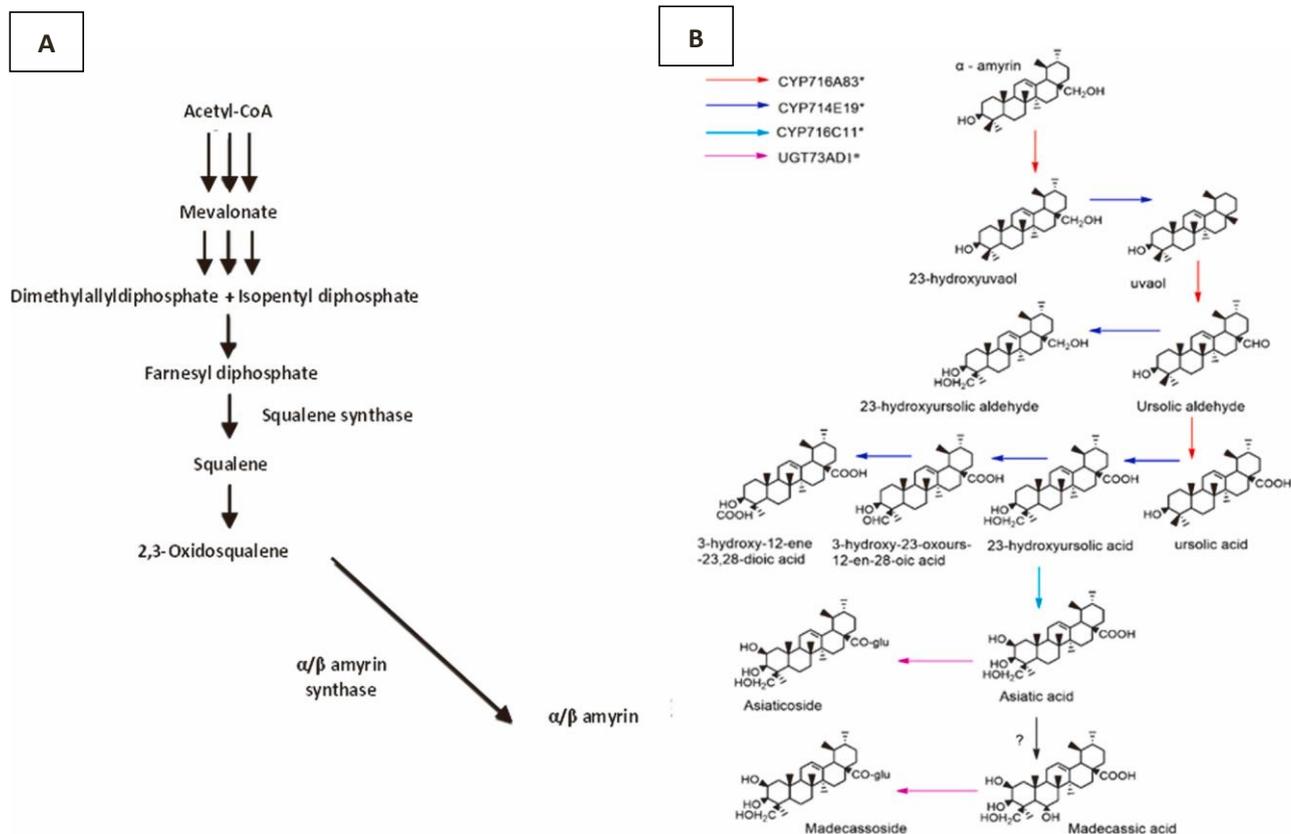

Fig. 1. A summarized scheme of centelloside biosynthesis. (A) Main steps in the biosynthesis of both α - and β -amyrin. (B) Last steps of oxidations and glycosylations. Modified figure from Mangas et al. (2008) and Kim et al. (2018).

Table 1
Primers used for confirmation of transformed roots.

Gene		Primer sequence (5´—3´)	Temperature melting (°C)	Amplicon size (pb)	Accession number
virD	Forward	ATGTCGCAAGGCAGTAAGCCA	56	700	KX986281.1
	Reverse	GGAGTCTTTCAGCATGGAGCAA			(*A. rhizogenes* A4)
rolA	Forward	TGGAATTAGCCGGACTAAAC	60	660	KX986281.1
	Reverse	GCGTACGTTGTAATGTGTTG			(*A. rhizogenes* A4)
rolB	Forward	AGTTCAAGTCGGCTTTAGGC	60	770	KX986281.1
	Reverse	TCCACGATTTCAACCAGTAG			(*A. rhizogenes* A4)
rolC	Forward	TAACATGGCTGAAGACGACC	60	534	KX986281.1
	Reverse	AAACTTGCACTCGCCATGCC			(*A. rhizogenes* A4)

control of centelloside production and ascertain which elicitor most effectively increases the expression of triterpene biosynthetic genes.

2. Material and methods

2.1. Establishment of hairy root cultures

In vitro plants of *C. asiatica* were cultured on MS medium including vitamins (Murashige and Skoog, 1962) and supplemented with 30% (w/v) sucrose, and 2.7% (w/v) gelrite. They were maintained in a controlled growth chamber at 25 °C and with a long-day photoperiod (16 h light/8 h dark).

Leaves from *in vitro* cultured *C. asiatica* plantlets were used to obtain explants. They were separated in a sterile Petri dish and cut into sections of approximately 1.5‑2 cm². A colony of *Agrobacterium rhizogenes* A4 strain, grown at 25 °C on solid YEB medium for 48 h and subsequently stored at − 70 °C, was detached with the tip of a scalpel and inoculated onto leaves from the underside. Leaves were placed on plates with solid MS medium, up to 4 leaf segments per plate, and incubated in the dark at 25 °C for 2 days. After 48 h, the explants were transferred to a new plate with solid MS medium + 500 mg/L cefotaxime and incubated at 25 °C for approximately 4 weeks.

After this time, the roots appearing on the leaf discs were separated and placed on new solid MS medium + 500 mg/L cefotaxime at 25 °C in darkness. This step was repeated every 2 weeks for approximately 2 months to remove any remaining *A. rhizogenes*. Transformed roots growing well were selected by cutting the top 5‑7 cm of the root apex

Table 2
Gradient used for HPLC separation of the centellosides.

Time (min)	Flow (mL/min)	Aqueous solvent (%)	Organic solvent (%)
0	1	80	20
15	1	62	38
30	1	30	70
35	1	30	70
37	1	80	20
45	1	80	20

with its lateral branches and cultured on new solid MS medium at 25 °C in darkness. Each root line was designated a number. Transformed root lines were sub-cultured in the same conditions every 2 weeks.

2.2. Confirmation of root transformation by PCR

Plant tissue (0.2 g) was pulverized in liquid nitrogen and placed in an Eppendorf tube, to which was added 0.75 mL of extraction buffer (50 mM EDTA Ph 8.0, 100 mM Tris pH 8.0 and 500 mM NaCl), 0.6 ⊮l of β-mercaptoethanol and 50 ⊮l of 20% sodium dodecyl sulphate (SDS). After incubation at 65 °C for 10 min, 250 ⊮l of 5 M potassium acetate was added, followed by incubation on ice for 20 min and then centrifugation at 4 °C for 20 min at 9000 rpm. After retention of the supernatant, 1 mL of isopropanol was added, followed by incubation at − 20 °C for 1 h. The pellet was centrifuged for 15 min at 8000 rpm, dried, and after the addition of 140 ⊮l of T10E1 (Tris 10 mM EDTA 1 mM), centrifuged again for 10 min at 14,000 rpm. After retention of the supernatant, 15 ⊮l of 3 M sodium acetate and 100 ⊮l of isopropanol were added. The supernatant was retained, centrifuged for 10 min at 15,000 rpm, and the pellet was kept. The Eppendorf tube was dried at 37 °C for 10 min, and 30 ⊮l of T10E1 buffer was added. Finally, 1 ⊮l of RNAse (10 mg/mL) was added and the purity of the DNA was measured using a *NanoDrop 2000 Spectrophotometer (Thermo Scientific)*.

Amplification of *rol* (*rol*A, *rol*B and *rol*C) and *vir* (*vir*D) genes was performed by PCR in 0.5 mL tubes. To each tube was added 12.5 ⊮l of Green taq polymerase, 1 ⊮l of forward primer, 1 ⊮l of reverse primer, 2 ⊮l of DNA and 8.5 ⊮l of milliQ water. The specific primers used are described in Table 1. The PCR conditions for *rol* and *vir* genes were as follows: one cycle at 94 °C for 5 min, followed by 35 cycles at 94 °C for 1 min, 60 °C for 30 s and 72 °C for 1 min, and a final extension step of 5 min at 72 °C. The size of the PCR products was determined by agarose gel electrophoresis (1%).

2.3. Selection and culture of the different hairy root lines

To choose the best hairy root line for the elicitation studies, 1 g of each line was sub-cultured in a glass flask with 20 mL of liquid MS

Table 3
Primers used for gene expression analysis.

Gene		Primer sequence (5´—3´)	Melting temperature (°C)	Amplicon size (pb)	Reference – Accession number
Squalene synthase (SQS)	Forward	TGAGGAGAATTCGGTCAAGG	60	126	AY787628.1
	Reverse	GCACAAAACCGGAAGATAGC			(*Centella asiatica*)
β-Amyrin synthase (β-AS)	Forward	TTACGTTTGCTGGGAGAAGG	60	134	AY520818.1
	Reverse	TATACCCCAAGAACCGCAAG			(*Centella asiatica*)
CYP716A83 (CYP83)	Forward	TAGCTCTGCATGCACTTTCG-	60	105	KU878849.1
	Reverse	CCAGGTCCCTTTGATTTCAC			(*Centella asiatica*)
CYP714E19 (CYP19)	Forward	AACCACACACACATCCTTGG	60	116	KT004520.1
	Reverse	TCCTTTGATGAGCCCAAGAC			(*Centella asiatica*)
CYP716C11 (CYP11)	Forward	CCCGATTCAACGACTCTTTC	60	104	KU878852.1
	Reverse	CCGGTTGTTTGATCGCACTTC			(*Centella asiatica*)
UGT73AD1 (UGT)	Forward	TCTGGAAGCAGTTTGTGAGG	59	101	KP195716.1
	Reverse	CGCCAATCTTCACTACATCG			(*Centella asiatica*)
β-actin	Forward	TGACAATGGAACTGGAATGG	58	80	Kim et al. (Kim et al., 2018) - KF699319.1
	Reverse	CAACAATACTGGGGAACACT			(*Panax ginseng*)

Fig. 2. Gel electrophoresis of PCR products from *C. asiatica* transformed roots after *A. rhizogenes* infection. (A) *A. rhizogenes* (positive control), (B) non-transformed roots (negative control), (C—L) hairy root lines (L1, L2, L3, L4, L6, L7, L8, L10, L12 and L14). MW: 100 bp molecular weight marker.
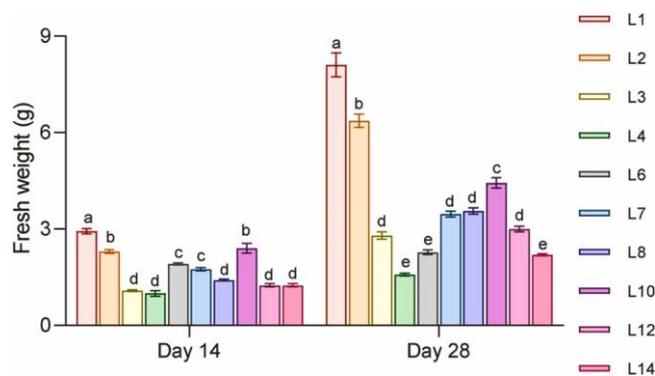
Fig. 3. Fresh weight of *C. asiatica* hairy roots. Data represent the mean ± SD of three replicates. Different letters show significant differences ($p < 0.05$) between hairy root lines at each time point.

medium for 4 weeks, maintained in a rotatory shaker at 115 rpm in darkness at 25 °C. The fresh weight (FW) was evaluated after 14 and 28 days. The growth of each hairy root line was established as the mean FW value of 3 replicates.

2.4. Elicitation studies

The elicitation experiment was performed using the transformed root line that produced the highest amount of biomass, applying the following elicitors at the indicated concentrations: 1 μM coronatine (Coro), 100 μM methyl jasmonate (Meja), 100 μM salicylic acid (SA) and 1 μM coronatine/100 μM methyl jasmonate (Coro Meja). These concentrations were the best concentrations studied up to now in different culture systems (Loc and An, 2010; Narayani and Srivastava, 2017; O.T. Kim et al., 2017; J.Y. Kim et al., 2017; Krishnan et al., 2019). As a pre-culture, inocula of 1 g FW of the selected hairy root line were placed in flasks containing 20 mL of liquid MS, and the cultures were maintained for 2 weeks in a rotatory shaker at 115 rpm in darkness at 25 °C.

After 2 weeks, the cultures were treated with the respective elicitors and maintained under the same conditions. After the treatment, samples were collected at 8 h and days 7 and 14 to determine the centelloside production, and at 0, 8, 12, 24, and 36 h and day 7 to study the expression of the targeted genes from the centelloside metabolic pathway. The samples were freeze-dried, and the dry weight (DW) was registered. The elicitation experiment was performed in triplicate.
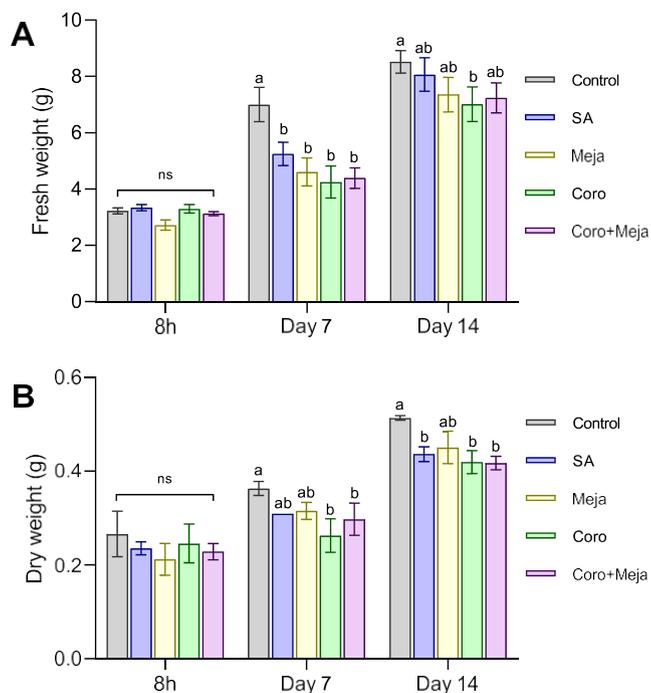
Fig. 5. Fresh (A) and dry (B) weight of elicited and non-elicited hairy roots. Data represents the mean ± SD of three replicates. Different letters show significant differences ($p < 0.05$) between elicited and non-elicited cultures at each time point.

2.5. Extraction and quantification of centellosides by HPLC/DAD

To determine the centelloside production, 0.5 g DW of elicited and non-elicited hairy roots were added to 10 mL of methanol: H$_2$O (9:1), and this suspension was sonicated for 1 h at room temperature, followed by centrifugation at 20,000 rpm for 10 min. The supernatant was separated, and the previous step was repeated. The supernatants of the different samples were placed in porcelain mortars to be evaporated at 38 °C for approximately 24 h and finally redissolved in 1.5 mL of methanol. The methanolic extract was filtered through a 0.22 μm filter for HPLC quantification of centellosides. The system consisted of a Waters 600 Controller Pump, a Waters 717 Autosampler Automatic Injector, a photodiode array (PDA) detector and Empower data analysis software version 1.5. Chromatographic analysis was performed by
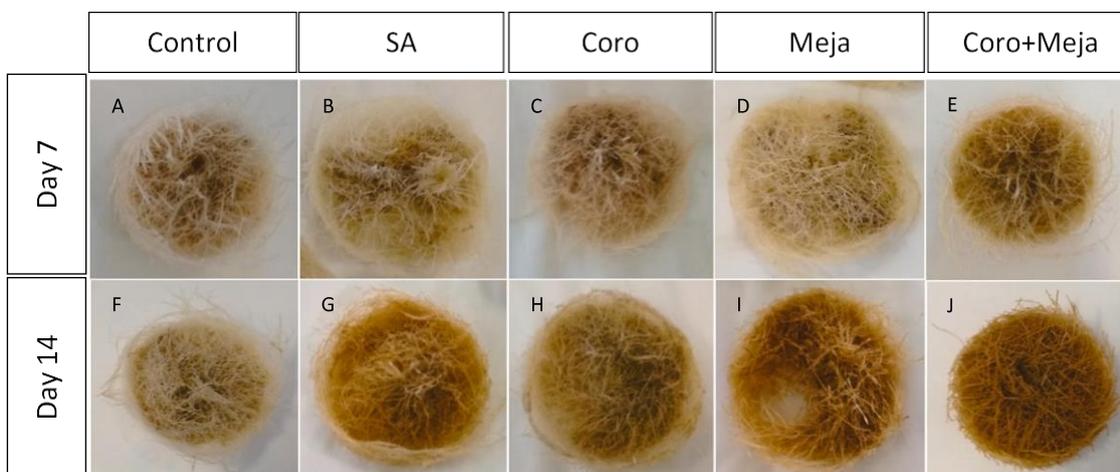
Fig. 4. Morphology and coloring of *C. asiatica* hairy roots (*Line 1*) at day 7 (photographs in the top row) and day 14 (photographs in the bottom row): non-elicited (A and F); elicited with salicylic acid (B and G); elicited with coronatine (C and H); elicited with methyl jasmonate (D and I); elicited with coronatine and methyl jasmonate (E and J).
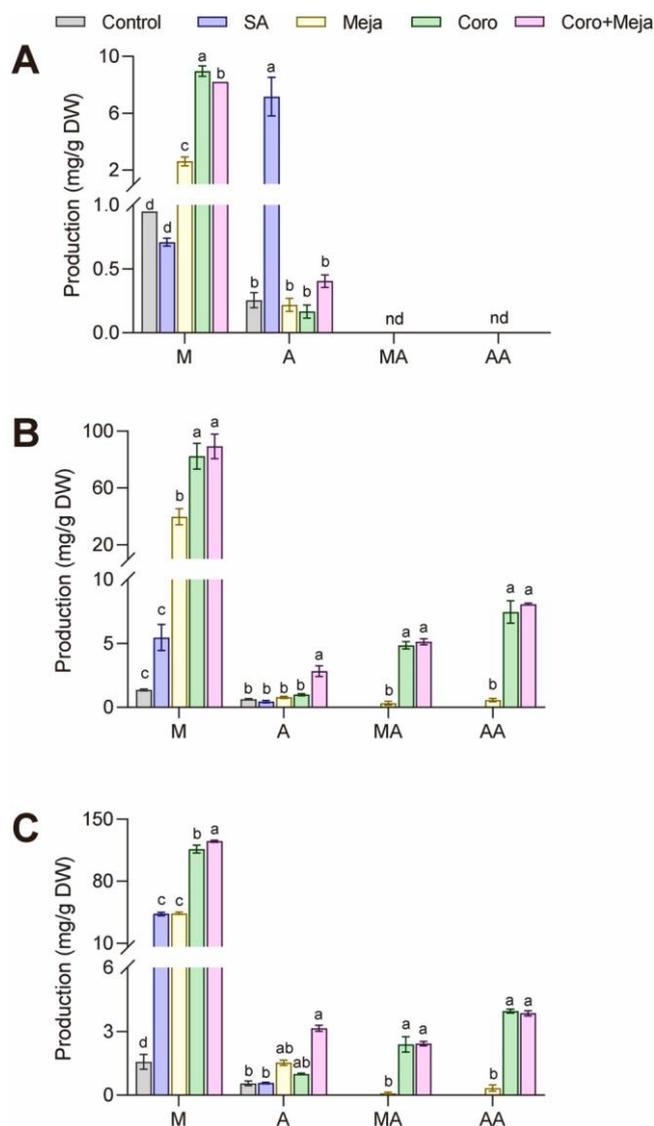
Fig. 6. Centelloside production (mg/g DW) in *C. asiatica* transformed roots. Madecassoside (M), asiaticoside (A), madecassic acid (MA) and asiatic acid (AA) production with different elicitation conditions: salicylic acid (SA), methyl jasmonate (Meja), coronatine (Coro), Coro + Meja, and control (no elicitation). (A) 8 h, (B) 7 days, and (C) 14 days after elicitation treatment. Data represent the mean ± SD of three replicates. Different letters show significant differences ($p < 0.05$) between elicited and non-elicited cultures for each centelloside. nd = not detected.

Fig. 7. Total centelloside production (mg/g DW) with elicitation treatments over time. Data represent the mean ± SD of three replicates. Different letters show significant differences ($p < 0.05$) between elicited and non-elicited cultures at each time point.

gradient (Table 2) at room temperature using a Lichrospher 100 RP18 5 μm column (250 × 0.4 mm). The mobile phase consisted of acetonitrile and ammonium phosphate (10 mM) acidified to pH 2.5 with orthophosphoric acid. The acidification improved the definition of the compound peaks. The flow rate was 1 mL/min, and the injection volume was 10 μl. The detector wavelength was set at 214 nm, 1.00 au/v, and the run time was 45 min. To quantify the centellosides (asiatic acid, madecassic acid, asiaticoside and madecasoside), standards of these four compounds were used at concentrations of 10, 25, 50, 100, 250 and 500 ppm to prepare the respective calibration curves.

2.6. Expression of centelloside pathway genes analyzed by RT-qPCR

The studied genes are representative of the centelloside biosynthetic pathway: squalene synthase (SQS), β-amyrin synthase (β-AS), CYP716A83 (CYP83), CYP714E19 (CYP19), CYP716C11 (CYP11) and

UGT73AD1 (UGT).

Total RNA extractions were obtained using elicited and non-elicited *C. asiatica* transformed roots. Elicited and control samples of approximately 5 g FW collected at each time point were frozen in liquid nitrogen and stored at − 70 °C. The *PureLink RNA Mini Kit (Invitrogen)* was used according to the manufacturer's instructions to extract the RNA. The amount and quality of each RNA sample was quantified using the *NanoDrop 2000 Spectrophotometer (Thermo Scientific)*. The integrity of RNA was also assessed by agarose gel electrophoresis.

DNAse treatment was performed using the total RNA with a fixed concentration as a template (1.5 μg of RNA). For this purpose, the required sample volume was calculated considering the volume of DNAse I and buffer needed and made up to a final volume of 10 μl per sample with sterile H_2O. After addition of the DNAse mix, the samples were heated at 37 °C for 30 min; then 1 μl of 50 mM EDTA was added per sample to inactivate DNAse by incubation at 65 °C for 10 min. First-strand cDNA was synthesized using the *SuperScriptTM IV First-Strand Synthesis System (Invitrogen)* and 2 μl of RNA according to the manufacturer's instructions.

After designing primers with Primer3Plus software (Table 3), the RT-qPCR assays were performed using the *LightCycler 480 System (Roche Diagnostics)*. Amplifications were carried out in 7.5 μl reaction solutions containing 2.5 μl first-strand cDNA (diluted 1:5), 1.7 μl of sterile H_2O milliQ, 2.5 μl of *iTaq™ Universal SYBR® Green Supermix (BIO-RAD)* and 0.4 μl of each specific primer, at a concentration of 10 μM. PCR conditions were 95 °C for 30 s followed by 45 cycles of 95 °C for 15 s and 60 °C for 30 s. The specificity of each pair of primers was checked by melting curve analysis (95 °C for 15 s, 65 - 95 °C, with an increment of 0.25 °C/s ramp rate followed by 95 °C for 15 s). To check reproducibility, each assay was performed with technical triplicates for each of the three biological samples. The $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001) was used to analyze the data obtained, using the constitutively expressed β-actin as the reference gene.

2.7. Statistical analysis

The statistical analysis was carried out using GraphPad PRISM software. Data are presented as the mean ± standard deviation. Multifactorial ANOVA analysis was performed, followed by Tukey's multiple comparison test for statistical comparison. A *p*-value of < 0.05 was assumed as a significant difference for growth parameters and centelloside production, and < 0.01 for gene expression analysis.

3. Results and discussion

3.1. Culture of hairy roots and confirmation of their transformed nature

The integration of *A. rhizogenes* plasmid T-DNA in the *C. asiatica*
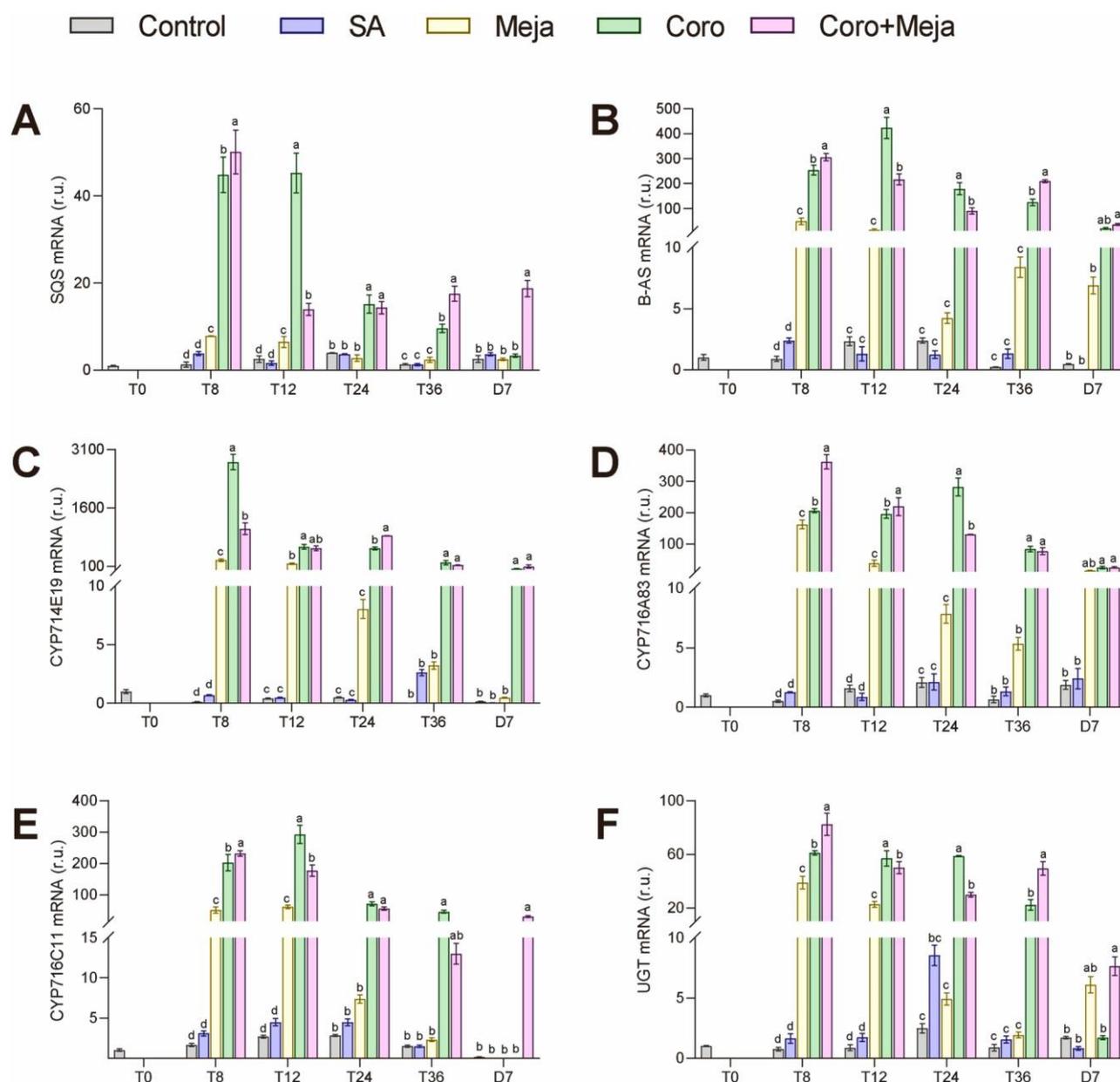
Fig. 8. RT-qPCR analysis of different genes from the centelloside biosynthetic pathway. (A) SQS, (B) β -AS, (C) CYP714E19, (D) CYP716A83, (E) CYP716C11 and (F) UGT gene expression in *C. asiatica* hairy roots after different elicitation treatments: salicylic acid (SA), coronatine (Coro), methyl jasmonate (Meja), Coro + Meja and control (no elicitation) at 0 (T0), 8 (T8), 12 (T12), 24 (T24), and 36 h (T36) and 7 days (D7). Data represent the mean ± SD of three replicates. Different letters show significant differences (*p < 0.01*) between elicited and non-elicited cultures at each time point.

Table 4
Correlation matrix between genes from the centelloside biosynthetic pathway. The asterisk indicates a significant difference (*p < 0.001*).

Gene expression	SQS	β -AS	CYP19	CYP83	CYP11	UGT
SQS	1					
β -AS	0.792*	1				
CYP19	0.669*	0.681*	1			
CYP83	0.706*	0.844*	0.729*	1		
CYP11	0.855*	0.890*	0.639*	0.730*		
UGT	0.656*	0.837*	0.712*	0.921*	0.702*	1

genome was used to confirm the transformed nature of the root lines. For this purpose, three *rol* genes (*rolA*, *rolB* and *rolC*) involved in hairy root formation (Ruslan et al., 2012) were analyzed by PCR. Different DNAs (from the hairy root lines, *A. rhizogenes* and untransformed *C. asiatica*

roots) were used as a template. Also, the *virD* gene was used as a negative control to check for residual *A. rhizogenes* in the transformed tissues; this gene is present in *A. rhizogenes* but not inside the T-DNA region, so does not integrate in plant cells (Sevón and Oksman-Caldentey, 2002). *Rol* genes (*rolA*, *rolB* and *rolC*) (660 pb, 770 pb and 534 pb, respectively) were found in *A. rhizogenes* (positive control) (Fig. 2A) and transformed *C. asiatica* root lines (Fig. 2C-L) but not in untransformed roots (Fig. 2B). The *virD* gene was only found in *A. rhizogenes* (Fig. 2A) and not in transformed and untransformed *C. asiatica* roots. The next step was to select the transformed root line with the highest biomass production measured as fresh weight (FW) after 3–4 weeks of growth.

3.2. Selection of the transformed root line with the highest biomass production

The growth course of the different transformed root lines is shown in

Table 5

Correlation matrix between gene expression and centelloside production. One to three asterisks indicate significant differences ($p < 0.05, 0.01, 0.001$ respectively).

Gene expression															
UGT					CYP11					CYP83					CYP19
D7	T36	T24	T16	T8	D7	T36	T24	T16	T8	D7	T36	T24	T16	T8	D7
0.067	0.574 *	0.740**	0.611*	0.551*	0.366	0.516*	0.715**	0.595 *	0.671 * *	0.531*	0.694**	0.685**	0.659**	0.521*	0.599*
0.526 *	0.813***	0.836***	0.947***	0.978***	0.548*	0.592 *	0.916***	0.802***	0.979***	0.908***	0.931***	0.856***	0.925***	0.930***	0.828***
0.440	0.811***	0.872***	0.898***	0.911***	0.578*	0.571*	0.913***	0.760**	0.958***	0.866***	0.923***	0.856***	0.913***	0.889***	0.836***

Fig. 3. As the growth capacity of *line 1* (*L1*) was significantly higher from the second week of culture, it was selected for the subsequent elicitation studies.

3.3. Effect of different elicitors on growth and centelloside production

The morphology and changes in color of hairy roots (*L1*) was observed at days 7 and 14 after elicitation (days 21 and 28 of the experiment). In general, elicitation caused hairy roots to darken with respect to the control, an effect that intensified over time (Fig. 4).

The growth course of the transformed roots (*L1*) under control or elicited conditions during a culture period of 14 days is shown in Fig. 5. The results indicate that elicitors generally inhibited the growth of the transformed roots, measured as FW and DW, compared to the non-elicited cultures.

The transformed root line producing the most biomass was used to analyze the production of the main centellosides in *C. asiatica*, madecassoside and asiaticoside, and their respective sapogenins, madecassic acid and asiatic acid. The yields were compared between elicited and non-elicited cultures and expressed in mg/g DW.

Madecassoside was the main centelloside detected in all cultures and at all time points (8 h, days 7 and 14). At 8 h of elicitation with Meja, Coro, or Coro+Meja, its production was significantly higher compared with the untreated samples, whereas SA did not have a significant effect. Conversely, SA induced a significant increase in asiaticoside, whose levels were not changed by the other treatments. Asiatic acid and madecassic acid were not detected at 8 h (Fig. 6 A). After 7 days of elicitation, madecassoside production was significantly higher in all the elicited samples compared with the control; the yields of most of the targeted centellosides were increased by elicitation with Coro or Coro + Meja (Fig. 6B).

Madecassoside production continued to increase significantly, in all treated cultures, at 14 days of elicitation. Coronatine or Coro + Meja significantly enhanced the production of all the centollosides, whereas SA and Meja did not affect madecassic and asiatic acid yields. The maximum production of madecassoside was achieved with Coro or Coro + MeJa (116 and 125 mg/g DW, respectively) after 14 days of elicitation (Fig. 6 C).

The total centelloside production (measured as the sum of madecassoside, asiaticoside, madecassic acid and asiatic acid) was higher in *C. asiatica* transformed roots treated with different elicitors (SA, Coro, Meja and Coro+Meja) than in the untreated samples (control) and their levels increased throughout the elicitation period. The treatments resulting in the highest increase were Coro and Coro + Meja (Fig. 7).

3.4. Gene expression analysis of control and elicited C. asiatica hairy root cultures

To understand how the elicitation treatments were affecting centelloside biosynthesis, the expression levels of six genes involved in different steps of the biosynthetic pathway were analyzed by RT-qPCR in elicited *C. asiatica* transformed roots sampled at 0, 8, 12, 24, and 36 h and 7 days after treatment (Table 3). The results are expressed in relative units, using β-actin as a reference gene, and non-elicitation as an internal control. For each treatment, expression levels were compared with the control sample (transformed but not elicited).

Expression levels of all targeted genes were significantly higher in samples treated with Coro or Coro+Meja compared with the control samples, at 8, 12, and 24 h post-treatment (Fig. 8). Both treatments had significantly upregulated all the genes at 36 h, except CYP716C11 in the case of Coro + Meja (Fig. 8E). Only elicitation with Coro+ Meja, at day 7, had a significant enhancing effect on the expression levels of all the genes, except CYP714E19 (Fig. 8 C). Treatment with Meja induced a significant upregulation of all the genes at 8 and 12 h, but at subsequent time points (24 h, 36 h and 7 days) no significant differences with control samples were found (Fig. 8). After SA elicitation, the expression levels of all the genes remained quite constant in all tested samples, the only significant increase being observed for UGT at 24 h (Fig. 8 F).

The different elicitors had variable effects on the gene expression patterns during the study period (from 8 h to day 7). Levels of SQS reached a maximum at 8 h with Coro + Meja and at 8 and 12 h with Coro, subsequently decreasing, although remaining significantly higher compared to the control (Fig. 8A). The expression of β-AS, CYP714E19, and CYP716A83 was consistently higher than control levels in samples treated with Coro or Coro + Meja, reaching peaks at 12 h with Coro, at 8 h with Coro, and at 8 h with Coro+Meja, respectively (Fig. 8B, C and D). CYP716C11 and UGT were also upregulated throughout the study period by Coro+Meja, but their expression level had fallen at day 7 in the Coro-treated samples (Fig. 8E and F). Finally, Meja increased the expression levels of SQS, CYP714E19 and CYP716C11 at 8 and 12 h, which decreased thereafter (Fig. 8A, C and E). The same effect was observed on β-AS, CYP716A83 and UGT, except their levels had increased again at day 7 (Fig. 8B, D and F).

A high correlation was observed between the expression levels of all the genes, especially between β-AS and the others, while the highest correlation value was found between UGT and CYP716A83 (Table 4).

A high level of correlation was also found between centelloside production and the expression levels of most genes, above all β-AS and CYP716A83, and mainly at 7 days post-elicitation. However, the expression of UGT at day 7 was not correlated with the production values at the earliest and last time points analyzed (Table 5).

3.5. Discussion

In this work it has been studied how different elicitors act at the transcriptomic level to increase centelloside production, with the aim of determining any correlation between improved yields and the upregulation of key biosynthetic genes. Hairy root cultures were used due to their numerous advantages as a biotechnological production platform for bioactive plant secondary compounds (Gutierrez-Valdes et al., 2020). These include: a rapid growth capacity, hormone-independent growth, genetic stability, and a capacity to biosynthesize the same type of compounds as the roots of the mother plant. The transformed *C. asiatica* roots had a high growth rate (Figs. 5 and 6) and were able to biosynthesize centellosides (Figs. 7 and 8).

Although the underlying mechanism has not been fully clarified, hairy roots can act as a self-sufficient production system of secondary metabolites without any special induction signal or metabolic

Gene expression														Time	
CYP19				β-AS					SQS						
T36	T24	T16	T8	D7	T36	T24	T16	T8	D7	T36	T24	T16	T8		
0.636*	0.694**	0.640*	0.634*	0.577*	0.685**	0.679**	0.667**	0.669**	0.412	0.574*	0.746**	0.566*	0.709**	8 h	Centelloside production (mg/g DW) after elicitation
0.756**	0.924***	0.956***	0.811***	0.870***	0.906***	0.840***	0.832***	0.961***	0.573*	0.842***	0.844***	0.724**	0.968***	Day 7	
0.767***	0.930***	0.915***	0.793***	0.878***	0.917***	0.839***	0.823***	0.951***	0.615***	0.801***	0.896***	0.703**	0.961***	Day 14	

engineering strategy (Hayat et al., 2010). However, this self-induction has not been observed in *C. asiatica*, and therefore strategies such as elicitation are required for centelloside production (Hidalgo et al., 2017). Elicitation is routinely used to manipulate secondary metabolite yields, imitating stress-induced production in nature. Considerable research has been dedicated to enhancing centelloside biosynthesis in *C. asiatica* by the application of biotic and abiotic elicitors, with different results achieved depending on the elicitor and/or type of culture used (Gallego et al., 2014). Hidalgo et al. (2017) and Mangas et al. (2006) found that all tested elicitors inhibited the growth of *C. asiatica* cell and plant cultures, measured as FW and DW.

One of the elicitors used in the present work is salicylic acid, a well-known inducer of systemic acquired resistance to many plant pathogens of secondary metabolites productions (Pieterse and Van Loon, 1999; Walker et al., 2002). It was observed that the application of SA improved the production of asiaticoside at 8 h post-elicitation and madecassoside at days 7 and 14. Demonstrating the effects of SA on the yields of a specific centelloside, Loc and Giang (2012) reported a production of 45.35 mg of asiaticoside/g of DW in non-elicited *C. asiatica* cell cultures, which they later improved five-fold by applying 100 μM of SA, reaching a value of 229.83 mg/g of DW (Loc and An, 2010).

The elicitor traditionally considered most effective in *C. asiatica* is Meja, which together with jasmonic acid, is produced widely in plants as a "stress hormone", in response to insect attacks (Kim et al., 2018). When exogenously applied to plant cell cultures of a variety of species, Meja (100-200 μM) enhances the workflow of secondary biosynthetic pathways, leading to an increased production of diverse compounds, including terpenoids, flavonoids, alkaloids and phenylpropanoids (Ramirez-Estrada et al., 2016). Recent studies in whole plant cultures demonstrated that the transcript levels of *C. asiatica* squalene synthase (CaSQS) and *C. asiatica* dammarenediol synthase (CaDDS), both genes associated with the triterpenoid pathway, are increased by the application of 100 μM Meja (Kim et al., 2018). In the present work, Meja elicitation mainly improved the production of madecassoside, which was higher than in the control at all time points (8 h, and days 7 and 14) (Fig. 6). Different studies have improved centelloside production using 100 μM Meja. Kim et al. (2007) obtained 7.12 mg/g DW of asiaticoside in transformed *C. asiatica* roots after 3 weeks of elicitation, a 5-fold increase compared to the control, and Ruslan et al. (2012) reported a very similar 4.9-fold increase in asiaticoside in cell suspension cultures. Yoo et al. (2011) described an enhanced production of both asiaticoside and madecassoside, but a reduction in madecassic acid levels, an effect we also observed, the amount decreasing from 0.32 mg/g to 0.09 mg/g of DW (Fig. 6).

The strongest elicitation effect on centelloside production reported to date was recently achieved by the exogenous application of the bacterial toxin coronatine (Coro) in cell suspension cultures (Onrubia et al., 2013). Due to a similar chemical structure, Coro mimics the function of the isoleucine-conjugated form of jasmonic acid and has a comparable mode of action to Meja. As an elicitor, Coro is of particular interest, as it enhances metabolite production more effectively than Meja and at lower concentrations (Ramirez-Estrada et al., 2016). This was confirmed in the present study, where the production of all centellosides was improved by Coro treatment, especially at day 14 after elicitation (Fig. 7). Similar results have been described by Hidalgo et al. (2017) in cell suspensions

of *C. asiatica*, and Onrubia et al. (2013) in cell cultures of *Taxus media*, in which Coro induced a higher taxane production than Meja. This superior performance could be due to the greater stability of Coro (Gallego et al., 2014), which would allow a more constant response in the plant tissue throughout the treatment period. The highest total centelloside production in *C. asiatica* hairy roots was achieved at day 14, the end of the experiment, when applying Coro, either alone or in combination with Meja, and the centelloside with the highest yield was madecassoside (Figs. 7 and 8). Possible explanations are that madecassoside (and asiaticoside) is more stable in intracellular spaces than its acid form, and, that the elicitors favored the glycosidation of these compounds.

Apart from improving metabolite production, elicitation of plant cultures provides a biological platform in which the stress response can be studied under controlled conditions (Lee et al., 2004). Changes in gene expression and proteins can be correlated with the accumulation of metabolites and messenger RNAs (mRNAs). Transcriptomes are highly dynamic, and mRNA levels provide information about gene expression within the signal response period. Therefore, to better understand the observed changes in centelloside production, key biosynthetic genes were analyzed by qRT-PCR, and the transcriptomic and metabolomic data derived from each elicitor treatment were integrated.

Among the targeted genes, SQS is involved in the formation of the universal precursor of triterpenoid saponins and phytosterols (Azerad, 2016). The elicitation treatments enhanced its expression, above all in cultures treated with Coro or Coro+MeJa at the beginning of the experiment (8 and 12 h, Fig. 8A). The subsequent decrease in expression is likely because SQS activity is not limited to centelloside biosynthesis. As SQS plays a key role in the upregulation of triterpene production (Lee et al., 2004), modulating its expression could avoid high accumulations of saponins and phytosterols.

β-amyrin synthase is a versatile gene involved in multiple steps in the centelloside metabolic pathway (Bonfill et al., 2011), catalyzing the formation of α and β-amyrin, the initial precursors of centellosides (Kim et al., 2007). The overexpression of β-AS reached a peak (an increase of up to 400-fold) at 12 h after elicitation with Coro, and remained consistently high throughout the study period in hairy roots treated with Coro or Coro+Meja, decreasingly slightly by day 7 (Fig. 8B). As CYP450s (CYP716A83, CYP714E19 and CYP716C11) are involved in the biosynthesis of different saponins derived from α and β-amyrin, an expression pattern similar to that of β-AS could be expected. Thus, maximum upregulation of CYP716A83 and CYP716C11 was observed at 8 h with Coro + Meja and at 12 h with Coro treatments, respectively (Fig. 8D and E). The expression of CYP714E19 underwent a highly significant increase, being almost 3000-fold higher compared to the control at 8 h in Coro-treated roots (Fig. 8C); this result could be explained by its involvement in more metabolic steps than the other CYP450s.

At the end of the biosynthetic pathway, UGT glycosylates sapogenins (acids) to yield the corresponding saponins (Kim et al., 2018). The expression of UGT increased in roots treated with Coro or Coro+Meja throughout the elicitation period, with maximum levels observed at 8 h of Coro+Meja treatment (Fig. 8F), like other targeted genes. The low yields (mg/g DW) of madecassic acid compared with madecassoside in these two treatments could be related with the elicitor-induced overexpression of UGT, which, as mentioned, is involved in the production of madecassoside from madecassic acid. Additionally, the complete

conversion of madecassic acid may have stimulated further production of this sapogenin from asiatic acid. The overexpression of UGT could also explain why madecassoside was the most produced centelloside in the other elicitor treatments. At day 7, the expression levels had mostly decreased, although they were maintained in Coro- and Coro + Meja-elicited cultures, probably due to transcript accumulation (Fig. 8).

Although the relative expression levels of genes varied throughout the period of elicitation (8–36 h), the most noticeable differences were observed in Coro and Coro + Meja treatments at 8 and 12 h (Fig. 8). Meja and SA also increased gene expression levels but not enough to have a clear impact on centelloside production, especially in the case of SA. A high degree of correlation was observed between the expression of all the studied genes, especially between β-AS and the others, which indicates that β-AS plays a major role in the regulation of the centelloside biosynthetic pathway(Kim et al., 2007).

The results demonstrate that the upregulation of biosynthetic genes can result in a higher centelloside production, especially long-term (7 days after elicitation), as shown by the values in Table 5. The key genes in the regulation of centelloside biosynthesis were found to be β-AS and CYP716A83 (Table 5); the former is responsible for modifying β-amyrin into oleanane-type saponins and the CYP gene family acts in multiple steps toward the end of the metabolic pathway (Kim et al., 2018). However, the most highly expressed genes after elicitation were mainly those involved in oxidations (CYP19), indicating that the tested elicitors did not act only on the key centelloside biosynthetic genes (b-AS, CYP83 and UGT).

Although gene expression was measured at specific hours and centelloside production on different days, any correlation between the two events is not necessarily evident immediately and, they can be separated by many intermediate steps. The lag between gene expression and secondary metabolite production is a common feature in *in vitro* cultures, as demonstrated in other studies. Ramirez-Estrada et al. (2015) observed that although taxane production was related with the expression of biosynthetic genes, other factors may also be involved, such as post-transcriptional and post-translational regulation.

4. Conclusions

New insights into metabolism and its regulation were obtained by studying centelloside production in elicited *C. asiatica* hairy roots together with the expression of key biosynthetic genes. Coro, especially in combination with Meja, proved to be the most effective elicitor in terms of enhanced gene expression and centelloside production, above all that of madecassoside. These findings may have useful application for increasing centelloside production in biotechnological platforms based on metabolically engineered *C. asiatica* hairy roots.

CRediT authorship contribution statement

Mercedes Bonfill supervised the project and wrote the draft manuscript, Javier Palazon designed and lead the Project, Elisabeth Moyano reviewed the manuscript, Rosa Maria Cusido analyzed all the data, Miguel Angel Alcalde carried out the experiments.

Funding

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

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

Bonfill, M., Mangas, S., Moyano, E., Cusido, R.M., Palazon, J., 2011. Production of centellosides and phytosterols in cell suspension cultures of *Centella asiatica*. Plant Cell Tissue Organ Cult. 104 (1), 61–67. https://doi.org/10.1007/s11240-010-9804-7.

Cusido, R.M., Onrubia, M., Sabater-Jara, A.B., Moyano, E., Bonfill, M., Goossens, A., Angeles, M., Palazon, J., 2014. A rational approach to improving the biotechnological production of taxanes in plant cell cultures of *Taxus* spp. Biotechnol. Adv. 32 (6), 1157–1167. https://doi.org/10.1016/j.biotechadv.2014.03.002.

Gallego, A., Ramirez-Estrada, K., Vidal-Limon, H.R., Hidalgo, D., Lalaleo, L., Khan Kayani, W., Cusido, R., Palazon, J., 2014. Biotechnological production of centellosides in cell cultures of *Centella asiatica* (L) Urban. Eng. Life Sci. 14 (6), 633–642. https://doi.org/10.1002/elsc.201300164.

Gandhi, S.G., Mahajan, V., Bedi, Y.S., 2015. Changing trends in biotechnology of secondary metabolism in medicinal and aromatic plants. Planta 241 (2), 303–317. https://doi.org/10.1007/s00425-014-2232-X.

Gutierrez-Valdes, N., Häkkinen, S.T., Lemasson, C., Guillet, M., Oksman-Caldentey, K.M., Ritala, A., Cardon, F., 2020. Hairy root cultures—a versatile tool with multiple applications. Front. Plant Sci. 11, 33. https://doi.org/10.3389/fpls.2020.00033.

Hayat, Q., Hayat, S., Irfan, M., Ahmad, A., 2010. Effect of exogenous salicylic acid under changing environment: a review. Environ. Exp. Bot. 68 (1), 14–25. https://doi.org/10.1016/j.envexpbot.2009.08.005.

Hidalgo, D., Steinmetz, V., Brossat, M., Tournier-Couturier, L., Cusido, R.M., Corchete, P., Palazon, J., 2017. An optimized biotechnological system for the production of centellosides based on elicitation and bioconversion of *Centella asiatica* cell cultures. Eng. Life Sci. 17 (4), 413–419. https://doi.org/10.1002/elsc.201600167.

James, J.T., Dubery, I.A., 2009. Pentacyclic triterpenoids from the medicinal herb, *Centella asiatica* (L.) Urban. Molecules 14 (10), 3922–3941. https://doi.org/10.3390/molecules14103922.

Kai, G., Liao, P., Xu, H., Wang, J., Zhou, C., Zhou, W., Qi, Y., Xiao, J., Wang, Y., Zhang, L., 2012. Molecular mechanism of elicitor-induced tanshinone accumulation in *Salvia miltiorrhiza* hairy root cultures. Acta Physiol. Plant. 34 (4), 1421–1433. https://doi.org/10.1007/s11738-012-0940-z.

Kim, J.Y., Kim, H.Y., Jeon, J.Y., Kim, D.M., Zhou, Y., Lee, J.S., Lee, H., Choi, H.K., 2017. Effects of coronatine elicitation on growth and metabolic profiles of *Lemna paucicostata* culture. PloS One 12 (11), e0187622. https://doi.org/10.1371/journal.pone.0187622.

Kim, O.T., Bang, K.H., Shin, Y.S., Lee, M.J., Jung, S.J., Hyun, D.Y., Kim, Y.G., Seong, N.S., Cha, S.W., Hwang, B., 2007. Enhanced production of asiaticoside from hairy root cultures of *Centella asiatica* (L.) Urban elicited by methyl jasmonate. Plant Cell Rep. 26 (11), 1941–1949. https://doi.org/10.1007/s00299-007-0400-1.

Kim, O.T., Yoo, N.H., Kim, G.S., Kim, Y.C., Bang, K.H., Hyun, D.Y., Kim, S.H., Kim, M., 2013. Stimulation of Rg3 ginsenoside biosynthesis in ginseng hairy roots elicited by methyl jasmonate. Plant Cell, Tissue Organ Cult. 112 (1), 87–93. https://doi.org/10.1007/s11240-012-0218-6.

Kim, O.T., Jin, M.L., Lee, D.Y., 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. https://doi.org/10.3390/ijms18122630.

Kim, O.T., Um, Y., Jin, M.L., Kim, J.U., Hegebarth, D., Busta, L., Racovita, R., Jetter, R., 2018. A novel multifunctional C-23 oxidase, CYP714E19, is involved in asiaticoside biosynthesis. Plant Cell Physiol. 59 (6), 1200–1213. https://doi.org/10.1093/pcp/pcy055.

Krishnan, M.L., Roy, A., Bharadvaja, N., 2019. Elicitation effect on the production of asiaticoside and asiatic acid in shoot, callus, and cell suspension culture of *Centella asiatica*. J. Appl. Pharm. Sci. 9 (06), 067–074. https://doi.org/10.7324/JAPS.2019.90609.

Lee, M.H., Jeong, J.H., Seo, J.W., Shin, C.G., Kim, Y.S., In, J.G., Yang, D.C., Yi, J.S., Choi, Y.E., 2004. Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene. Plant Cell Physiol. 45 (8), 976–984. https://doi.org/10.1093/pcp/pch126.

Liang, Z.S., Yang, D.F., Liang, X., Zhang, Y.J., Liu, Y., Liu, F.H., 2012. Roles of reactive oxygen species in methyl jasmonate and nitric oxide-induced tanshinone production in *Salvia miltiorrhiza* hairy roots. Plant Cell Rep. 31 (5), 873–883.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 25 (4), 402–408. https://doi.org/10.1006/meth.2001.1262.

Loc, N.H., An, N.T.T., 2010. Asiaticoside production from centella (*Centella asiatica* L. Urban) cell culture. Biotechnol. Bioprocess Eng. 15 (6), 1065–1070. https://doi.org/10.1007/s12257-010-0061-8.

Loc, N.H., Giang, N.T., 2012. Effects of elicitors on the enhancement of asiaticoside biosynthesis in cell cultures of centella (*Centella asiatica* L. Urban). Chem. Pap. 66 (7), 642–648. https://doi.org/10.2478/s11696-012-0168-9.

Mangas, S., Bonfill, M., Osuna, L., Moyano, E., Tortoriello, J., Cusido, R.M., Piñol, M.T., Palazon, J., 2006. The effect of methyl jasmonate on triterpene and sterol metabolisms of *Centella asiatica*, *Ruscus aculeatus* and *Galphimia glauca* cultured plants. Phytochemistry 67 (18), 2041–2049. https://doi.org/10.1016/j.phytochem.2006.06.025.

Mangas, S., Moyano, E., Osuna, L., Cusido, R.M., Bonfill, M., Palazon, J., 2008. Triterpenoid saponin content and the expression level of some related genes in calli

of *Centella asiatica*. Biotechnol. Lett. 30 (10), 1853-1859. https://doi.org/10.1007/s10529-008-9766-6.

Malarz, J., Stojakowska, A., Kisiel, W., 2007. Effect of methyl jasmonate and salicylic acid on sesquiterpene lactone accumulation in hairy roots of *Cichorium intybus*. Acta Physiol. Plant. 29 (2), 127-132. https://doi.org/10.1007/s11738-006-0016-z.

Murashige, T., Skoog, F., 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant. 15 (3), 473-497. https://doi.org/10.1111/j.1399-3054.1962.tb08052.X.

Narayani, M., Srivastava, S., 2017. Elicitation: a stimulation of stress in in vitro plant cell/tissue cultures for enhancement of secondary metabolite production. Phytochem. Rev. 16 (6), 1227-1252. https://doi.org/10.1007/s11101-017-9534-0.

Onrubia, M., Moyano, E., Bonfill, M., Cusidó, R.M., Goossens, A., Palazón, J., 2013. Coronatine, a more powerful elicitor for inducing taxane biosynthesis in *Taxus media* cell cultures than methyl jasmonate. J. Plant Physiol. 170 (2), 211-219. https://doi.org/10.1016/j.jplph.2012.09.004.

Pauwels, L., Inzé, D., Goossens, A., 2009. Jasmonate-inducible gene: what does it mean? Trends Plant Sci. 14 (2), 87-91. https://doi.org/10.1016/j.tplants.2008.11.005.

Pieterse, C.M., Van Loon, L.C., 1999. Salicylic acid-independent plant defense pathways. Trends Plant Sci. 4 (2), 52-58. https://doi.org/10.1016/S1360-1385(98)01364-8.

Ramirez-Estrada, K., Osuna, L., Moyano, E., Bonfill, M., Tapia, N., Cusido, R.M., Palazon, J., 2015. Changes in gene transcription and taxane production in elicited cell cultures of *Taxus× media* and *Taxus globosa*. Phytochemistry 117, 174-184. https://doi.org/10.1016/j.phytochem.2015.06.013.

Ramirez-Estrada, K., Vidal-Limon, H., Hidalgo, D., Moyano, E., Golenioswki, M., Cusidó, R.M., Palazon, J., 2016. Elicitation, an effective strategy for the biotechnological production of bioactive high-added value compounds in plant cell factories. Molecules 21 (2), 182. https://doi.org/10.3390/molecules21020182.

Ruslan, K., Selfitri, A.D., Bulan, S.A., Rukayadi, Y., 2012. Effect of *Agrobacterium rhizogenes* and elicitation on the asiaticoside production in cell cultures of *Centella asiatica*. Pharmacogn. Mag. 8 (30), 111. https://doi.org/10.4103/0973-1296.96552.

Sevón, N., Oksman-Caldentey, K.M., 2002. *Agrobacterium rhizogenes*-mediated transformation: root cultures as a source of alkaloids. Planta Med 68 (10), 859 – 868. https://doi.org/10.1055/S-2002-34924.

Sharmila, R., Subburathinam, K.M., 2013. Effect of signal compounds on andrographolide in the hairy root cultures of *Andrographis paniculata*. Int. J. Pharm. Sci. Res. 4 (5), 1773. https://doi.org/10.13040/IJPSR.0975-8232.4(5).1773-76.

Singh, J., Sabir, F., Sangwan, R.S., Narnoliya, L.K., Saxena, S., Sangwan, N.S., 2015. Enhanced secondary metabolite production and pathway gene expression by leaf explants-induced direct root morphotypes are regulated by combination of growth regulators and culture conditions in *Centella asiatica* (L.) urban. Plant Growth Regul. 75 (1), 55-66. https://doi.org/10.1007/s10725-014-9931-y.

Tamogami, S., Kodama, O., 2000. Coronatine elicits phytoalexin production in rice leaves (*Oryza sativa* L.) in the same manner as jasmonic acid. Phytochemistry 54 (7), 689-694. https://doi.org/10.1016/s0031-9422(00)00190-4.

Tewari, R.K., Paek, K.Y., 2011. Salicylic acid-induced nitric oXide and ROS generation stimulate ginsenoside accumulation in *Panax ginseng* roots. J. Plant Growth Regul. 30 (4), 396-404. https://doi.org/10.1007/s00344-011-9202-3.

Thakur, M., Sohal, B.S., 2013. Role of elicitors in inducing resistance in plants against pathogen infection: a review. Int. Sch. Res. Not. 1-10. https://doi.org/10.1155/2013/762412.

Vaccaro, M.C., Mariaevelina, A., Malafronte, N., De Tommasi, N., Leone, A., 2017. Increasing the synthesis of bioactive abietane diterpenes in *Salvia sclarea* hairy roots by elicited transcriptional reprogramming. Plant Cell Rep. 36 (2), 375-386. https://doi.org/10.1007/s00299-016-2076-X.

Verpoorte, R., 1999. Chemodiversity and the biological role of secondary metabolites, some thoughts for selecting plant material for drug development. In: Bohlin, L., Bruhn, J.G. (Eds.), Bioassay Methods in Natural Product Research and Drug Development. Proceedings of the Phytochemical Society of Europe, vol 43. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-4810-8_2.

Walker, T.S., Bais, H.P., Vivanco, J.M., 2002. Jasmonic acid-induced hypericin production in cell suspension cultures of *Hypericum perforatum* L.(St. John̓s wort). Phytochemistry 60 (3), 289-293. https://doi.org/10.1016/S0031-9422(02)00074-2.

Wen Su, W., 1995. Bioprocessing technology for plant cell suspension cultures. Appl. Biochem. Biotechnol. 50 (2), 189-230. https://doi.org/10.1007/BF02783455.

Yoo, N.H., Kim, O.T., Kim, J.B., Kim, S.H., Kim, Y.C., Bang, K.H., Hyun, D.Y., Cha, S.W., Kim, M.Y., Hwang, B., 2011. Enhancement of centelloside production from cultured plants of *Centella asiatica* by combination of thidiazuron and methyl jasmonate. Plant Biotechnol. Rep. 5 (3), 283-287. https://doi.org/10.1007/s11816-011-0173-4.

Chapter 4. Enhancing centelloside production in *Centella asiatica* hairy root lines through metabolic engineering of triterpene biosynthetic pathway early genes.
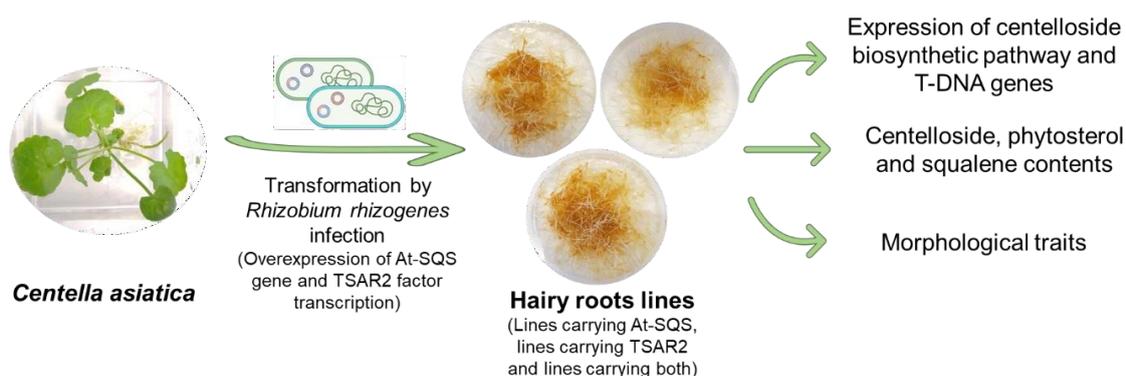
Spanish summary

Centella asiatica es una planta medicinal con una rica historia de uso tradicional debido a sus propiedades terapéuticas. Entre sus compuestos bioactivos, los centellosidos, han llamado la atención debido a sus potentes actividades farmacológicas. La ingeniería metabólica ha surgido como una poderosa herramienta biotecnológica para mejorar la producción de compuestos deseados. En este estudio, exploramos los efectos de la sobreexpresión del gen de la escualeno sintasa (*At-SQS*) de *Arabidopsis thaliana* y el factor de transcripción *TSAR2* en diversos aspectos: la expresión de genes en la ruta metabólica de los centellósidos, rasgos morfológicos, así como el contenido de escualeno, fitoesteroles y centellósidos en líneas de raíces transformadas de *C. asiatica*. Nuestros esfuerzos resultaron en tres categorías distintas de líneas transformadas: líneas LS, donde estaba presente *At-SQS*; líneas LT, con sobreexpresión de *TSAR2*; y líneas LST, con ambos transgenes expresados simultáneamente. Estas líneas mostraron alteraciones notables en los rasgos morfológicos, especialmente la tasa de ramificación y la producción de biomasa. Además, observamos que la expresión de genes clave en la biosíntesis de centellósidos experimentó modulación debido a la expresión de genes T-DNA, con un impacto notable de los genes *aux2* y *rolC*. Las líneas LS, que presentaron un contenido elevado de centellósidos pero al mismo tiempo mostraron un contenido reducido de fitoesteroles. Este hallazgo destacó una intrigante relación antagónica entre las vías de fitoesteroles y triterpenos. La correlación inversa inicialmente observada entre las líneas LS y los rasgos morfológicos es contrarrestada por la acción de *TSAR2* en las líneas LST y LT. Esta alteración podría atribuirse al aumento simultáneo del contenido de fitoesteroles, que está estrechamente relacionado con el desarrollo de las raíces. En conjunto, estos descubrimientos ofrecen información valiosa sobre las perspectivas biotecnológicas de las raíces transformadas de *C. asiatica* y su potencial para aumentar la producción de centellósidos.

Enhancing Centelloside Production in *Centella asiatica* Hairy Root Lines through Metabolic Engineering of Triterpene Biosynthetic Pathway Early Genes

Miguel Angel Alcalde [1,2] , **Javier Palazon** [1,*] , **Mercedes Bonfill** [1] and **Diego Hidalgo-Martinez** [1,*]

[1] Department of Biology, Healthcare and the Environment, Faculty of Pharmacy and Food Sciences, University of Barcelona, 08028 Barcelona, Spain; miguel.psr.94@gmail.com (M.A.A.); mbonfill@ub.edu (M.B.)

[2] Biotechnology, Health and Education Research Group, Posgraduate School, Cesar Vallejo University, Trujillo 13001, Peru

* Correspondence: javierpalazon@ub.edu (J.P.); dhidalgo@ub.edu (D.H.-M.)

Abstract: *Centella asiatica* is a medicinal plant with a rich tradition of use for its therapeutic properties. Among its bioactive compounds are centellosides, a group of triterpenoid secondary metabolites whose potent pharmacological activities have attracted significant attention. Metabolic engineering has emerged as a powerful biotechnological tool to enhance the production of target compounds. In this study, we explored the effects of overexpressing the squalene synthase (*SQS*) gene and transcription factor *TSAR2* on various aspects of *C. asiatica* hairy root lines: the expression level of centelloside biosynthetic genes, morphological traits, as well as squalene, phytosterol, and centelloside content. Three distinct categories of transformed lines were obtained: LS, harboring *At-SQS*; LT, overexpressing *TSAR2*; and LST, simultaneously carrying both transgenes. These lines displayed noticeable alterations in morphological traits, including changes in branching rate and biomass production. Furthermore, we observed that the expression of T-DNA genes, particularly *aux2* and *rolC* genes, significantly modulated the expression of pivotal genes involved in centelloside biosynthesis. Notably, the LS lines boasted an elevated centelloside content but concurrently displayed reduced phytosterol content, a finding that underscores the intriguing antagonistic relationship between phytosterol and triterpene pathways. Additionally, the inverse correlation between the centelloside content and morphological growth values observed in LS lines was countered by the action of *TSAR2* in the LST and LT lines. This difference could be attributed to the simultaneous increase in the phytosterol content in the *TSAR2*-expressing lines, as these compounds are closely linked to root development. Overall, these discoveries offer valuable information for the biotechnological application of *C. asiatica* hairy roots and their potential to increase centelloside production.

Keywords: metabolic engineering; centelloside; triterpene; synthetic biology; machine learning; plant biotechnology

1. Introduction

Centella asiatica, commonly known as *Gotu Kola*, is a renowned medicinal plant with a rich history of traditional use in various cultures for its therapeutic properties [1]. Among its numerous plant secondary bioactive compounds are centellosides, a group of triterpenes that have attracted considerable attention due to their potent pharmacological activities, including anti-inflammatory, antioxidant, and wound-healing effects. The high medicinal value of centellosides has prompted extensive research aimed at understanding their biosynthesis and developing strategies to enhance their production [2].

The biosynthetic pathways of secondary metabolites, including centellosides, are complex and tightly regulated processes. In plants, secondary metabolite production is influenced by various factors, such as gene expression, enzymatic activity, and metabolic

flux. Among the key regulatory components are phytosterols; these primary metabolites and membrane constituents serve as precursors for various hormones and play a crucial role in determining the production of specific triterpenes. It has been observed that when sufficient phytosterols have been synthesized to meet the plant's growth requirements, the biosynthesis of triterpenes takes precedence [3,4].

Metabolic engineering has emerged as a powerful biotechnological tool to manipulate the biosynthetic pathways of secondary metabolites in plants. By introducing specific genes into plant cells, researchers can modulate gene expression and metabolic pathways to enhance the production of desired compounds [5,6]. In the context of centellosides, studies have focused on manipulating key biosynthetic genes to augment their yields in *C. asiatica*.

One such gene of interest encodes squalene synthase (*SQS*), a key enzyme in the biosynthesis of triterpenes [4,7]. By overexpressing the *SQS* gene, it is possible to increase levels of the precursor squalene, potentially leading to higher centelloside production. The engineering of secondary metabolite production also involves transcription factors (TS) [8,9]. For example, *TSAR2* (triterpenoid saponin activating regulator 2) has been identified as having a direct effect on the expression of the *HMGR* gene, which encodes 3-hydroxy-3-methylglutaryl-coenzyme A reductase, an important enzyme in the triterpene biosynthetic pathway [10–12]. The overexpression of the *HMGR* gene has shown promising results in increasing phytosterol and triterpenoid content in different plant species [11,13,14].

Hence, the generation of *C. asiatica* hairy roots harboring *SQS* and *TSAR2* offers a potential opportunity to enhance centelloside yields. In this study, we explored the impact of these genetic modifications on the expression of centelloside pathway genes and morphological traits in different *C. asiatica* hairy root lines. Additionally, we examined the correlation between squalene, phytosterol, and centelloside contents in the transformed roots, aiming to elucidate the regulatory mechanisms underlying centelloside biosynthesis in this medicinal plant.

The findings from this study could offer valuable insights into the biotechnological potential of *C. asiatica* hairy root lines and pave the way for novel approaches to enhance centelloside production. Ultimately, these advancements could provide the pharmaceutical and cosmeceutical industries with a sustainable source of centellosides and help to unlock their therapeutic potential for various health applications.

2. Results

2.1. Integration of Transgenes by PCR

In the present study, we obtained four different types of hairy root lines: LS, carrying the exogenous *SQS* gene from *A. thaliana*; LT, which contains the transcription factor *TSAR2*; LST, with both of the aforementioned genes; and L1, utilized as the control. These hairy root lines harbored all the *rol* and *aux* genes from the pRi plasmid of *R. rhizogenes*.

Genomic DNAs from transformants and *C. asiatica* leaves were used for PCR analysis of multiple lines (Figure S1). Initially, the *rolC* gene served as a positive control to confirm the transformed nature of the obtained roots, given its involvement in hairy root formation. Following three rounds of selection with cefotaxime, the *VirD* gene was utilized to verify the absence of *R. rhizogenes* in the transformants, as this gene is present in the bacterial plasmid but not in the T-DNA region integrated in plant cells. The results demonstrated the presence of the *rolC* gene in *R. rhizogenes* (positive control) and the transformants but not in *C. asiatica* leaves (negative control). Conversely, the *VirD* gene was absent in all the lines. Finally, the presence of the target genes, *At-SQS* and *TSAR2*, was verified, and the following lines were selected for further analysis: LST(1-3), LS(1-3), and LT(1-3).

2.2. Influence of Transgenes on Endogenous Gene Expression

By conducting a genetic expression analysis, we explored the impact of the *At-SQS* and *TSAR2* transgenes on the centelloside biosynthetic genes. The results, illustrated in Figure 1, depict the expression profiles of the targeted endogenous genes.
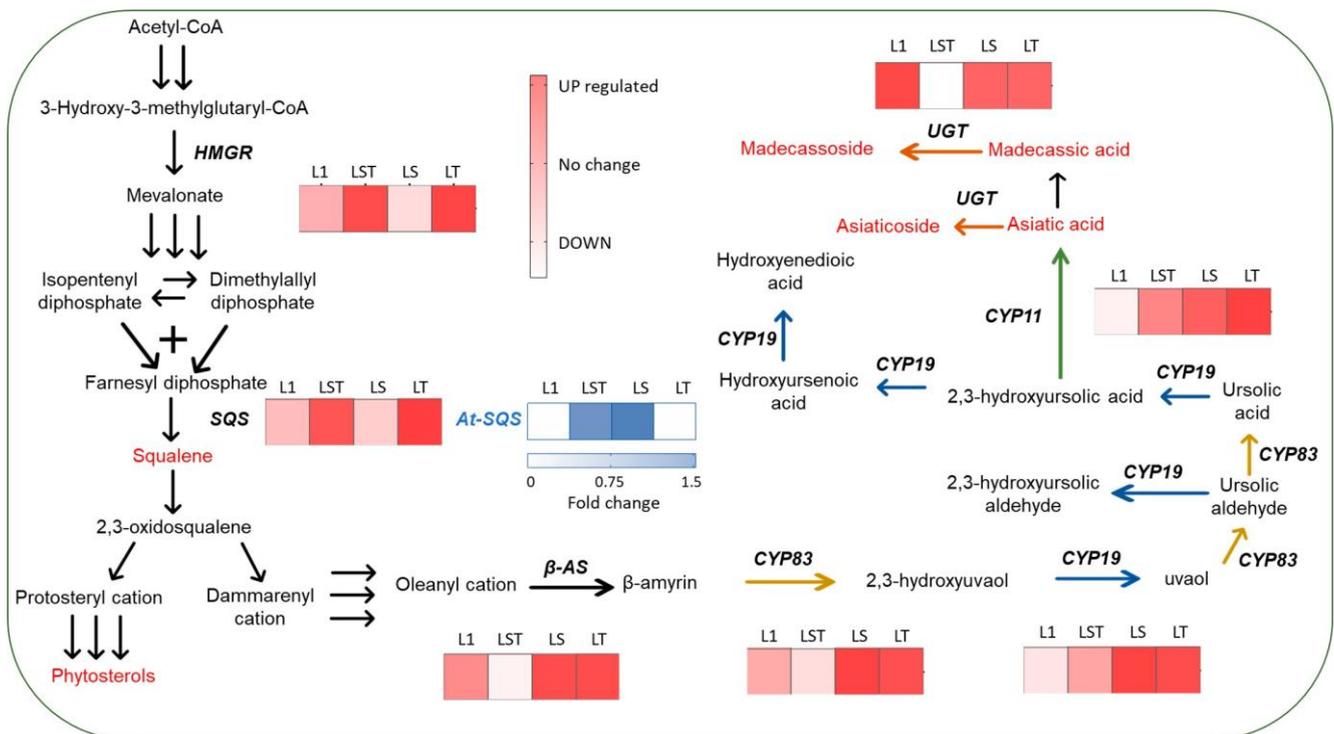
Figure 1. Overview of the centelloside biosynthetic pathway and gene expression profile. L1 denotes the control line, while LST, LS, and LT represent the transformed lines. The reddish heatmap indicates the upregulation (UP) or downregulation (DOWN) of gene expression for *HMGR* (encoding the 3-hydroxy-3-methylglutaryl coenzyme A reductase), *SQS* (encoding squalene synthase), *β-AS* (encoding beta amyrin synthase), *CYP83* (cytochrome *CYP716A83*), *CYP19* (cytochrome *CYP714E19*), *CYP11* (cytochrome *CYP716C11*), and *UGT* (*UGT73AD1*). The bluish heatmap reflects the fold change in gene expression of *At-SQS* (squalene synthase from *Arabidopsis thaliana*).

The expression of the HMGR and SQS genes differed notably among the examined lines. Specifically, lines overexpressing TSAR2 exhibited significantly elevated expression levels of these two endogenous genes compared with the control. A correlation analysis revealed a strong positive correlation (r = 0.98) between HMGR and SQS, suggesting that TSAR2 enhances the expression of these genes.

Surprisingly, this transcription factor exhibited no remarkable effects on or correlations with other genes involved in centelloside biosynthesis. In addition, the transformants exclusively overexpressing At-SQS had gene expression values on par with those of the control line. Consequently, no conspicuous augmentations or detrimental influences on the expression of the endogenous genes within the pathway were observable.

To examine the impact of the *rol* and *aux* transgenes derived from the pRi plasmid of *R. rhizogenes* on the centelloside biosynthetic genes, we generated multiple regression models using the randomForestSRC R-package (see Section 4). The models treated the endogenous genes *HMGR*, *SQS*, *β-AS*, *CYP83*, *CYP19*, *CYP11*, and *UGT* as dependent variables, whereas the expression levels of *rolA*, *rolB*, *rolC*, *aux1*, and *aux2* genes were considered independent variables. Among these models, the one pertaining to *CYP11* exhibited the highest R-squared value (0.7349), indicating the strongest level of fit. Conversely, the *SQS* gene model had the smallest R-squared value (0.1387). This suggests that the expression levels of the pRi plasmid transgenes are either unconnected with or exert a very limited influence on the expression behavior of the *SQS* gene.

The variable importance metric was utilized to compare the impact of pRi plasmid transgenes on the profile of each endogenous gene, aiding in the interpretation of the multivariate regression models. For a comprehensive understanding of the influence of transgene expression, Figure 2 presents a cluster analysis of the pathway genes.
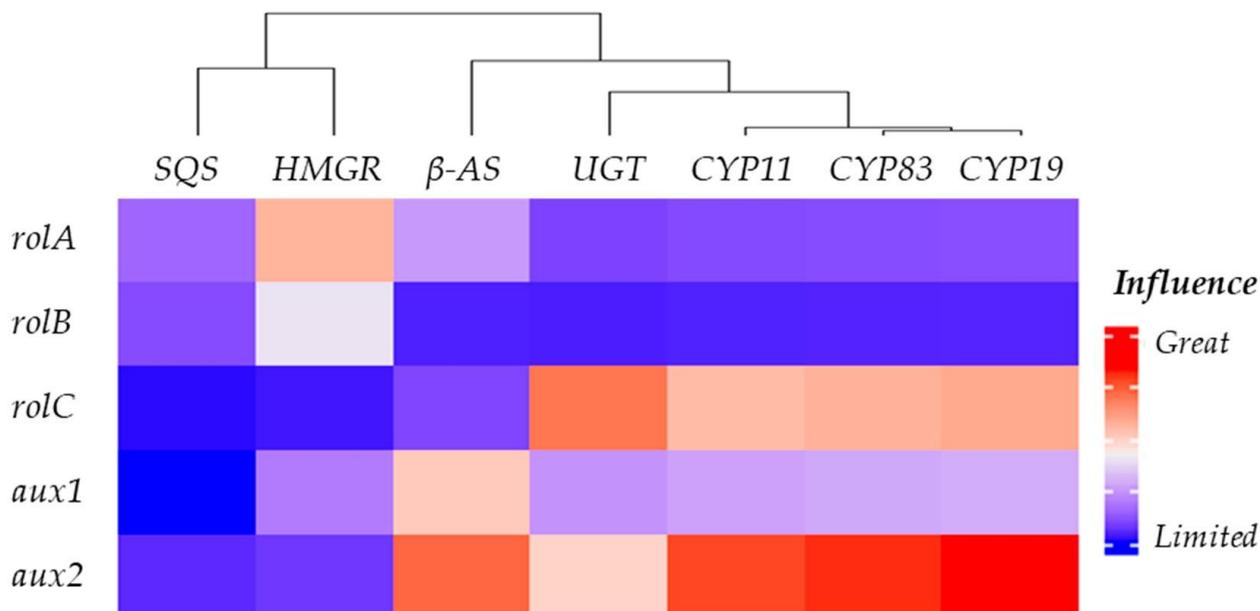
Figure 2. Comprehensive heatmap illustrating the impact of *rol* and *aux* genes on the endogenous gene expression profile in *Centella asiatica* hairy roots.

Remarkably, *aux2* exhibited a favorable impact on *β-AS*, *CYP83*, *CYP19*, and *CYP11*, with *UGT* displaying the least susceptibility and *CYP19* showing the most pronounced effect. Following closely, *rolC* emerged as the subsequent influential factor, specifically influencing *CYP83, CYP19, CYP11*, and *UGT* compared with *aux2*. The influence of *rolC* was particularly prominent on the *UGT* gene. The *aux1* gene, on the other hand, was notable solely for its impact on *β-AS*, while *rolA* exhibited only a marginal effect on *HMGR*. Among the endogenous genes, *SQS* was the least susceptible to the influence of the *rol* or *aux* genes, closely followed by *HMGR*.

2.3. Sterols, Squalene, and Centolloside Profiles in Response to Transgenes TSAR2 and At-SQS

Through a comprehensive phytochemical analysis, we examined the effects of the At-*SQS* and *TSAR2* transgenes on centelloside, squalene, and total sterol profiles. Sterols constitute key compounds within a competitive pathway toward centellosides, while squalene serves as a common precursor.

In Figure 3, the upper section illustrates the total content (measured in mg/g DW) of centellosides, sterols, and squalene in the four types of hairy root lines. Notably, the LS lines exhibited the highest centelloside content (6.56 ± 0.26 mg/g DW), differing significantly from the other lines (Figure S2). The direct influence of the *At-SQS* transgene on production levels was evident.

Greater variability was observed in the sterol content, which ranged from 188.31 to 56.88 mg/g DW among the root lines. Specifically, the LST lines displayed a significant increase compared with the control, whereas a pronounced decrease was observed in the LS lines, resulting in values lower than the control. Conversely, the values of the LT lines did not differ significantly from those of the LST lines. Noteworthy in this context was the clear positive effect of the transcription factor on sterol content, as well as on the *HMGR* gene, as previously mentioned.
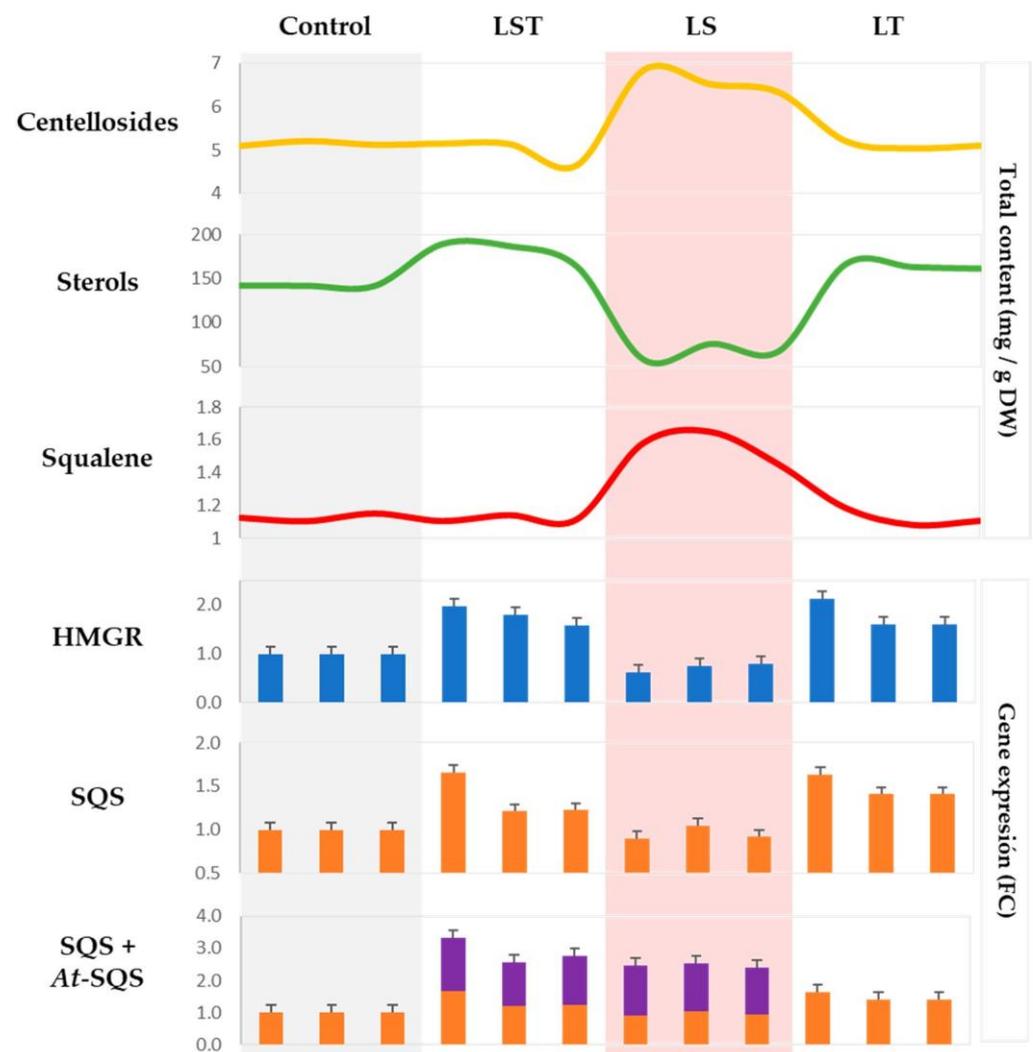
Figure 3. Comparative plot of various types of hairy root lines. The density plot illustrates the profiles of sterols, squalene, and centellosides, while the bar plots display the gene expression profiles of *HMGR, SQS,* and *At-SQS* in terms of fold change values. The stacked column highlighted in purple represents the contribution of *At-SQS*. See Table S1 for more details.

The profile of the squalene content was similar to that of centellosides, with a significant increase in the LS lines (1.56±0.10 mg/g DW), and nearly constant values in the other three lines. Accumulated expression values of native *SQS* and *At-SQS* genes in the LS line clearly demonstrated the positive impact of the *At-SQS* transgene on centelloside production. However, this positive effect was only observed in the absence of the transcription factor, suggesting that the activation of the sterol biosynthetic pathway effectively utilized the additional squalene generated by the heterologous enzyme.

2.4. Influence of Squalene and Sterol Content on Morphological Traits

In this study, we deemed the branching rate, growth rate, and biomass productivity to be pertinent morphological traits owing to their biotechnological implications. The branching rate denotes the number of lateral roots per centimeter, growth rate signifies the elongation of roots in millimeters per day, and biomass productivity quantifies the daily biomass yield for each line. Various morphologies of hairy root lines after 28 days of growth are illustrated in Figures 4 and S3.
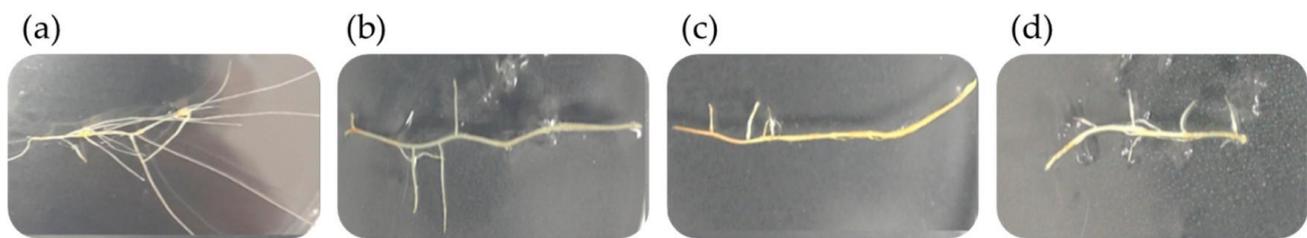
Figure 4. Average morphology of different hairy root lines of *Centella asiatica* at day 28 of growth: (**a**) control, (**b**) LST, (**c**) LS, (**d**) LT. The initial inoculum was a small section of 10 mg of fresh weight.

By conducting a thorough correlation analysis between the morphological traits and the contents of squalene and sterols, we have unveiled a compelling depiction of how the accrual or depletion of these two distinct compounds intricately interweaves with the measured characteristics (Figure 5). The control line maintained a robustly positive correlation between the variables. However, in the LS line, which yielded significantly diminished levels of sterols yet substantially elevated quantities of squalene and, therefore, centellosides, a different pattern was discernible. Here, it became distinctly apparent that the aggregation of squalene exerted a deleterious impact on the underlying morphological traits.
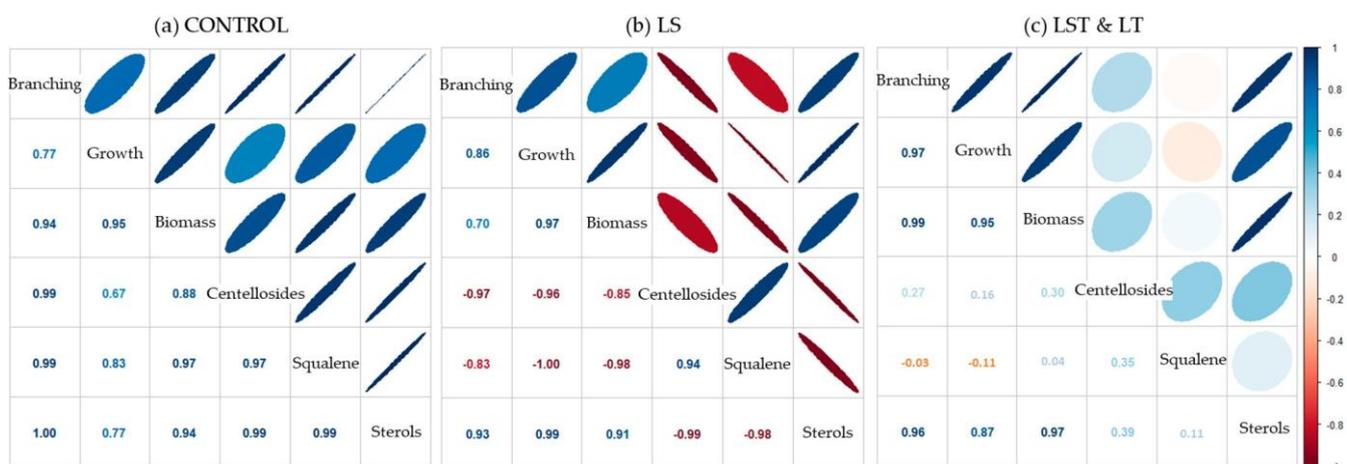
Figure 5. Pearson's correlation analysis of morphological traits and metabolite content. (**a**) Correlation plot for the control line; (**b**) correlation plot for the LS lines; (**c**) correlation plot for the lines overexpressing the transcription factor. Values indicate the Pearson correlation coefficient (r).

Intriguingly, in the LST line, this adverse effect appeared to be counterbalanced by the presence of the transcription factor. This regulatory entity seemed to orchestrate a harmonizing influence, fostering amplified sterol production while concurrently addressing the extra squalene availability. This bifaceted mechanism effectively overcame the negative impact observed in the LS line, thus indicating the intricate interplay between genetic modulation and metabolite accumulation.

3. Discussion

This study further elucidates the complex interplay between specific genes and the centelloside biosynthetic pathway. Notably, the *C. asiatica* hairy root lines harboring *TSAR2* exhibited a substantial upregulation of the endogenous genes *SQS* and *HMGR*, suggesting they play pivotal roles in the regulation of centelloside biosynthesis. These findings align with those of Xu et al. [12], who observed increased *HMGR* gene expression in *Ganoderma lucidum* overexpressing the *GlbHLH5* transcription factor under methyl jasmonate elicitation, and Mertens et al. [11], who reported higher transcript levels of the *HMGR* gene in *Medicago truncatula* hairy roots overexpressing *TSAR2*. Moreover, instances of

transcription-factor-mediated modulation of *SQS* gene homolog expression, leading to activated triterpenoid biosynthesis, have been observed in various species such as *Ziziphus jujuba* and *Torreya grand* [15,16].

A key factor in triterpenoid structural diversity is the cytochrome P450 family, particularly the CYP716 enzymes. These versatile enzymes, distributed widely across diverse plant species, have the capacity to catalyze triterpene oxidation at various positions [17]. Among the ensemble of genes, *CYP16C11* emerges as a significant player and is modulated by the insertion of T-DNA genes (Figure 2). Our findings are in parallel to those of Pandey et al. [18], who investigated the complex interactions of *CYP716C55* gene expression and metabolic pathways under Ri-T-DNA-mediated insertional mutagenesis, unraveling the intricacies of gene modulation within the dynamic landscape of *Ocimum* species hairy roots.

The *aux2* gene (1398 bp), a component of TR-DNA, plays a pivotal role in the induction and continuation of hairy root growth [19,20]. This oncogene orchestrates the conversion of naphthalene acetamide into the potent auxin naphthalene acetic acid, a crucial stage in hairy root development [21]. Notably, Srivastava et al. [22] revealed the significant influence of the *aux2* gene on 3-indol acetic acid (IAA) levels, morphological traits, and metabolite content in *Ocimum basilicum* hairy roots, with implications mirroring our own results, as its overexpression had a favorable impact on the expression of *β-AS, CYP,* and *UGT* genes.

By boosting beta-glucosidase activity, the *rolC* gene provides a gateway for the release of active cytokinins from their inactive counterparts [23]. Its versatile effects include roles in nicotine production in *Nicotiana tabacum* hairy roots [24], ginsenoside synthesis in *Panax ginseng* hairy roots [25], and resveratrol production in transgenic *Vitis amurensis* cell cultures [26]. In our study, we observed how the *rolC* gene positively influences *CYP* and *UGT* gene expression. Similarly, Inyushkina [27] noted that *rolC* gene expression led to increased transcript levels of *CYP98A3* subfamily members, crucial genes in rosmarinic acid biosynthesis.

The important role of the *rolA* gene in triterpenoid production has been reported in *Artemisia dubia* hairy roots, where it was found to directly enhance artemisinin production [28]. Interestingly, although the influence of the *rolA* gene in secondary metabolism remains less explored compared with other oncogenes, our work reveals its notable impact on the early triterpene pathway gene *HMGR*, which encodes an enzyme catalyzing vital steps in IPP biosynthesis [29].

The LST and LT lines stood out for their high phytosterol content, a direct effect of *TSAR2*. The influence of other transcription factors such as *WRKY* or *MYC2* (*WsWRKY1* and *WsMYC2*) on phytosterol biosynthesis has been demonstrated in studies on *Withania somnifera* [30,31].

The LS lines exhibited the highest squalene content, mirroring their centelloside content (Figure 3), which illustrates the effect of introducing the *SQS* gene from *A. thaliana* into the *C. asiatica* hairy root lines. This phenomenon has been previously studied in *Medicago truncatula* [32], *P. ginseng* [4], *W. somnifera* [33], *Taraxacum koksaghyz* [34], and *Eleutherococcus senticosus* [35], the results attesting to the intriguing connection between squalene synthase overexpression and enhanced triterpene levels.

As well as the highest centelloside content, the LS lines possessed the lowest phytosterol content (Figure 3), which reflects the fact that triterpene biosynthesis occurs once sufficient phytosterol has been synthesized, signaling the end of growth [3,4]. Supporting this observation, studies have shown that inhibiting the phytosterol biosynthetic pathway by antisense cycloartenol synthase led to a notable increase in ginsenosides (triterpenes) in *P. Ginseng* hairy roots [13]. A similar antagonistic interaction between the triterpene and sterol pathways has been observed in transgenic *Buplerum falcatum* roots [36] and in methyl-jasmonate-elicited *C. asiatica* plants, where free sterol decreased while centellosides increased [37].

Furthermore, it has been reported that triterpenic saponins can also contribute to root growth, as demonstrated in the case of a γ-pyronyl triterpenoid saponin, known as chromosaponin 1 in *Lactuca sativa* [38]. In *A. thaliana*, this compound interacts with the *aux1* protein, regulating the root gravitropic response [39].

Regarding the hairy root morphological traits, the strongest positive correlation between the different growth values with metabolite composition was observed in the control lines (Figure 5a). Conversely, the traits of LS lines exhibited an opposing trend (Figure 5b), which implies that the accumulation of squalene and subsequent elevation in centelloside content functions as a trigger for secondary metabolism, marked by the cessation of cell division and growth [40]. Curiously, in the LST and LT lines, the negative correlation observed in the LS lines disappears (Figure 5c). This remarkable fact finds its origins in the insertion of *TSAR2*, which directly affects the *HMGR* gene and, quite possibly, genes of the sterol biosynthetic pathway, resulting in a higher phytosterol content. These effects of the transcription factor contrast with the findings of Hey et al. [14], who observed a 2.44-fold increase in sterol production in transformed *A. thaliana* compared with the control.

Our findings further highlight the regulatory interplay between the phytosterol and triterpene production pathways, along with their impacts on hairy root morphological traits. Future research should focus on refining the modulation of the transgenes to achieve a harmonious equilibrium between production levels and morphological attributes.

4. Materials and Methods

4.1. Establishment and Selection of C. asiatica Hairy Roots

The protocol followed in this study was based on Alcalde et al. [41] with slight modifications. In vitro cultures of *C. asiatica* plants were maintained on Murashige and Skoog (MS) medium supplemented with vitamins, 30% (*w/v*) sucrose, and 2.7% (*w/v*) gelrite in a controlled growth chamber. The growth chamber provided a constant temperature of 25 °C and a long-day photoperiod of 16 h of light followed by 8 h of darkness.

To initiate the hairy root induction process, leaf segments from the plantlets were wounded and infected with two distinct strains of *R. rhizogenes* A4. The first strain carried the binary plasmid pBI121_*At-SQS*, which encodes squalene synthase 1 from *A. thaliana* [42]. The second strain carried the plasmid pK7WG2D,1_*TSAR2*, encoding a transcription factor [43]. To achieve a double mutation, a simultaneous infection involving both strains was performed. Prior to infection, these bacterial strains were cultured at 25 °C on solid YEB medium for 48 h, supplemented with 100 mg/L of kanamycin and 50 mg/L of rifampicin.

After the infection, the wounded leaf segments were placed on solid MS medium enriched with 50 mg/L of acetosyringone. Up to four leaf segments were placed per plate, and the plates were then incubated in darkness at 25 °C for 2 days. Subsequently, the explants were transferred to fresh plates with solid MS medium containing 500 mg/L of cefotaxime and incubated at 25 °C for 4 weeks.

Following this incubation period, the roots that developed on the leaf segments were carefully isolated and transferred to new solid MS medium supplemented with 500 mg/L of cefotaxime and 0.1 mg/L of kanamycin. This step was repeated every 2 weeks over a period of about 2 months to eliminate any residual presence of *R. rhizogenes*. Robustly growing transgenic roots were identified by excising the top 5–7 cm of the root apex along with its lateral branches. These selected root lines were then cultured on fresh solid MS medium at a temperature of 25 °C in darkness, and subculturing of the transgenic root lines was carried out under the same conditions every 2 weeks.

4.2. DNA Isolation and Transgenic Confirmation by PCR Analysis

The DNA isolation protocol, adapted from Alcalde et al. [41], involved several steps. Initially, 200 mg of hairy root tissue was pulverized with liquid nitrogen and combined with 0.75 mL of extraction buffer (containing 50 mM EDTA pH 8.0, 100 mM Tris pH 8.0, and 500 mM NaCl), 0.6 µL of β-mercaptoethanol, and 50 µL of 20% sodium dodecyl sulfate (SDS). This mixture was incubated at 65 °C for 10 min, followed by the addition of 250 µL

of 5 M potassium acetate. After an ice incubation for 20 min and centrifugation at 4 °C for 20 min at 9000 g, the supernatant was collected, mixed with 1 mL of isopropanol, and incubated at -20 °C for 1 h. The resulting pellet was centrifuged at 8000 g for 15 min and air-dried, then resuspended in 140 μL of T10E1 buffer (Tris 10 mM, EDTA 1 mM) and centrifuged for 10 min at 14,000 g. To the retained supernatant, 15 μL of 3 M sodium acetate and 100 μL of isopropanol were added. The resulting supernatant was again retained and centrifuged at 15,000 g for 10 min. The genomic DNA within the pellet was air-dried at 37 °C for 10 min and then resuspended in 30 μL of T10E1 buffer. Subsequently, 1 μL of RNAse (10 mg/mL) was added, and the purity of the DNA was assessed using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

For the purpose of transgenic confirmation, transgenes from *R. rhizogenes* strains, including *rolC*, *At-SQS*, *TSAR2*, and *virD* genes, were amplified by PCR within 0.5 mL tubes. To each tube, 12.5 μL of Green taq polymerase, 1 μL of forward primer, 1 μL of reverse primer, 2 μL of DNA, and 8.5 μL of Milli-Q water were added. The specific primers used are detailed in Table S1. The PCR protocol comprised an initial cycle at 94 °C for 5 min, followed by 35 cycles at 94 °C for 1 min, 60 °C for 30 s, and 72 °C for 1 min, with a final extension step at 72 °C for 5 min. The size of the PCR products was determined using agarose gel electrophoresis (1%).

4.3. Analysis of Morphological Traits

To initiate the cultures, an inoculum of 10 mg of fresh weight (FW) was introduced for each hairy root line onto solid MS medium plates. These cultures were consistently maintained through two consecutive subcultures, with intervals of 2 weeks, at a temperature of 25 °C in a dark environment. This approach is in accordance with our established methodology from previous investigations [44].

Following this cultivation period, the assessment of growth parameters was carried out, with three replicates conducted for each distinct line. The evaluated growth parameters included the branching rate, quantified as the number of lateral roots per centimeter of the initial stem root (number of lateral roots/cm). Additionally, the growth rate was calculated by measuring the average length of the lateral roots (mm/day). Finally, biomass productivity was determined by dividing the difference between the final FW and the initial FW by the number of days the growth occurred (mg/day).

4.4. Quantification of Centelloside Production

The determination of centelloside production was conducted in accordance with the method outlined by Alcalde et al. [41]. Initially, 0.5 g of freeze-dried hairy root material (DW) was carefully weighed, followed by the addition of a 10 mL suspension of methanol:H_2O (9:1). This suspension was then subjected to sonication for 1 h at room temperature to facilitate extraction. Subsequently, the suspension underwent centrifugation at 2000 g for 10 min to separate the solid and liquid phases. The supernatant obtained from this process was carefully separated, and the centrifugation step was repeated once more to ensure thorough extraction. The supernatants collected from the various samples were placed into porcelain mortars and evaporated at 38 °C for approximately 24 h. Following this, the resulting material was redissolved in 1 mL of methanol.

The methanolic extracts were subsequently filtered using a 0.22 μm filter for HPLC quantification of centellosides. The HPLC system consisted of a Waters 600 Controller pump, a Waters 717 Autosampler automatic injector, a Jasco variable-length (UV) 1570 detector, and Borwin data analysis software version 1.5. A Lichrospher 100 RP18 5 μm column (250 × 0.4 mm) was employed for gradient chromatography at room temperature following the method described by Alcalde et al. [41]. The mobile phase consisted of acetonitrile and ammonium phosphate (10 mM), acidified to a pH of 2.5 with orthophosphoric acid to enhance the definition of compound peaks. The flow rate was maintained at 1 mL/min, with an injection volume of 10 μL. The detector wavelength was set at 214 nm, 1.00 au/v, and the total run time was 45 min.

For the quantification of centellosides, namely asiatic acid, madecassic acid, asiaticoside, and madecassoside, calibration curves were established using standards of these four compounds. Calibration concentrations ranged from 10 to 500 ppm, with specific points at 10, 25, 50, 100, 250, and 500 ppm.

4.5. Expression Analysis of Centelloside Pathway Genes by RT-qPCR

Total RNA extraction was performed as in Alcalde et al. [41]. Trizol reagent was added to an Eppendorf tube with 200 mg of FW material and incubated for 20 min at room temperature. After shaking to mix the contents, the tubes were centrifuged at 7500 g for 10 min at 4 °C. The RNA pellet obtained was subjected to a single wash with 1 mL of 75% ethanol. After gentle shaking to mix the contents, the tubes were centrifuged at 7500 g for 5 min at 4 °C. Following removal of the supernatant, vacuum drying was employed for 10 min to desiccate the tubes. To dissolve the RNA samples, 20 µL of DEPC-treated water was utilized. Subsequent to extraction, RNA samples were assessed for purity and concentration using the NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

The treatment with DNAse was administered using the total RNA as a template, employing a fixed concentration (2 µg of RNA). Additionally, the RNA integrity was confirmed via agarose gel electrophoresis. The synthesis of the first-strand cDNA was performed using the SuperScriptTM IV First-Strand Synthesis System (Invitrogen from Thermo Fisher Scientific, Waltham, MA, USA) with 2 µg of RNA, precisely following the manufacturer's protocols.

After the design of primers via Primer3Plus software (version: 3.3.0) (Table S2), RT-qPCR assays were performed utilizing the QuantStudio3 System (Thermo Fisher Scientific, Waltham, MA, USA). These amplifications were conducted in 10 µL reaction volumes, consisting of 1 µL of first-strand cDNA (25 ng/µL), 2 µL of DEPC-treated water, 5 µL of iTaq Universal SYBR Green Supermix (Biorad, Hercules, CA, USA), and 1 µL of each specific primer at a concentration of 10 µM. The PCR conditions involved an initial phase at 95 °C for 30 s, followed by 40 cycles of 95 °C for 15 s and 60 °C for 30 s. To confirm the specificity of each primer pair, a melting curve analysis was performed, starting at 95 °C for 15 s, followed by a gradual increase from 60 to 95 °C at a rate of 0.1 °C/s, and ending with another step at 95 °C for 15 s.

4.6. Determination of Squalene Content

The quantification of squalene content was conducted following the methodologies of Shen et al. [45] and Rothblat et al. [46], with minor adjustments. In brief, the hairy roots were subjected to overnight drying in an oven set at 60 °C. For the subsequent steps, 200 mg of the dried sample was employed. Extraction was carried out within glass test tubes, utilizing 2 mL of n-hexane, and conducted in a supersonic water bath maintained at 40 °C for a duration of 30 min. The supernatants thus obtained were collected, and the extraction process was repeated once more to ensure thorough extraction. The pooled fractions were subjected to evaporation at 40 °C. Subsequently, 1 mL of concentrated sulfuric acid was added, and the mixture was subjected to a 5 min incubation in a water bath set at 70 °C. Then, 0.5 mL of 37% formaldehyde solution was gradually added, with the tubes being shaken for comprehensive mixing, and the mixture was placed in a water bath set at 95 °C for a 5 min period. Finally, to reach a final volume of 4 mL, 2.5 mL of glacial acetic acid was added promptly, and the mixture was meticulously mixed. After an incubation of 5 min, the absorbance at 400 nm was gauged using a UV2310 spectrophotometer (DINKO instruments, Barcelona Spain).

4.7. Determination of Phytosterol Content

The quantification of phytosterol content was conducted as in Araújo et al. [47], with minor modifications. For the preparation of the Liebermann–Burchard (LB) reagent, 50 mL of acetic anhydride was added to an amber glass bottle, which was maintained at a chilled temperature over ice. After a 30 min period, 5 mL of concentrated sulfuric acid was cautiously added to the bottle.

To establish the phytosterol reference curve, β-sitosterol was employed as the standard. Specifically, 10 mg of β-sitosterol was dissolved within 1 mL of chloroform, yielding a concentration of 10 mg/mL. Gradient concentrations spanning from 0.02 to 0.1 mg/mL of β-sitosterol were generated by diluting the stock solution with chloroform. Within the context of this reference curve, 0.5 mL of each concentration from the gradient was combined with 0.5 mL of LB reagent. Additionally, a blank solution comprising 0.5 mL of chloroform and 0.5 mL of LB reagent was prepared.

To determine the phytosterol content, 0.1 g of the dried sample material was mixed with 5 mL of chloroform within a glass tube. Subsequently, sonication was carried out for 30 min at room temperature. The supernatants thus obtained were collected, and the extraction process was repeated once more. The pooled fractions were subjected to evaporation at 40 °C, and the resulting residues were resuspended within 2 mL of chloroform. A mixture comprising 0.5 mL of the resuspension and 0.5 mL of LB reagent was prepared. The absorbance was measured at 625 nm, ensuring that this measurement was taken within a maximum of 5 min after the addition of LB reagent. For the purpose of comparison, a blank solution was generated, comprising 0.5 mL of chloroform and 0.5 mL of LB reagent.

4.8. Statistical Analysis and Machine Learning Models

Statistical analysis was performed using GraphPad Prism version 6.04 for Windows, developed by GraphPad Software in La Jolla, CA, USA. The data are presented as mean values with standard deviations. To conduct statistical comparisons, a multifactorial ANOVA analysis was employed, followed by Tukey's multiple comparison tests. A significance level of $p \leq 0.05$ was used to indicate statistical significance.

Machine learning models and Pearson correlation analysis were carried out using R Statistical Software (version 4.3.1) [48]. The randomForestSRC package [49] was employed to implement multivariate regression models. The coefficient of determination (R-squared) was used as a metric to evaluate the fitting of each model. The corrplot package [50] was utilized to compute the Pearson correlation coefficient (r). The reduced accuracy values of the multivariate multiple regression model were employed for hierarchical clustering analysis and were visually represented through a heatmap.

5. Conclusions

In this study, we explored the effects of *At-SQS* and *TSAR2* transgenes on the centelloside biosynthetic pathway, revealing distinct impacts on gene expression profiles. *TSAR2* overexpression notably elevated *HMGR* and endogenous *SQS* expression, demonstrating a strong positive correlation with these two genes. The transcription factor exhibited limited influence on other pathway genes, although its overexpression enhanced sterol production. Overexpression of *At-SQS* alone had minimal effects on endogenous genes but significantly increased the squalene content. Further investigation of pRi plasmid-derived *rol* and *aux* transgenes revealed a significant positive impact of *aux2* and *rolC* genes on centelloside pathway genes, whereas the influence of *aux1* and *rolA* was more restricted. The overarching theme from the findings of this study is the intricate interplay between genetic modulation and gene expression, shedding light on the complex mechanisms governing centelloside biosynthesis. Additionally, a comprehensive phytochemical analysis unveiled the interactions between *At-SQS* and *TSAR2* transgenes and centelloside, sterol, and squalene content. In general, the accumulation of squalene in the absence of *TSAR2* had a detrimental effect on the sterol content. Moreover, by measuring the branching rate, growth

rate, and biomass productivity of the hairy roots, we revealed that these morphological traits of biotechnological importance are negatively correlated with squalene, whereas their improvement was strongly positively associated with sterol.

References

1. Sabaragamuwa, R.; Perera, C.O.; Fedrizzi, B. *Centella asiatica* (*Gotu Kola*) as a neuroprotectant and its potential role in healthy ageing. *Trends Food Sci. Technol.* **2018**, *79*, 88–97. [CrossRef]
2. Kunjumon, R.; Johnson, A.J.; Baby, S. *Centella asiatica*: Secondary Metabolites, biological activities and biomass sources. *Phytomed. Plus* **2022**, *2*, 100176. [CrossRef]
3. Rogowska, A.; Szakiel, A. Enhancement of phytosterol and triterpenoid production in plant hairy root cultures-simultaneous stimulation or competition? *Plants* **2021**, *10*, 2028. [CrossRef] [PubMed]
4. Lee, M.H.; Jeong, J.H.; Seo, J.W.; Shin, C.G.; Kim, Y.S.; In, J.G.; Yang, D.C.; Yi, J.S.; Choi, Y.E. Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase Gene. *Plant Cell Physiol.* **2004**, *45*, 976–984. [CrossRef] [PubMed]
5. Mehrotra, S.; Rahman, L.U.; Kukreja, A.K. An Extensive Case Study of hairy-root cultures for enhanced secondary-metabolite production through metabolic-pathway engineering. *Biotechnol. Appl. Biochem.* **2010**, *56*, 161–172. [CrossRef] [PubMed]
6. Staniek, A.; Bouwmeester, H.; Fraser, P.D.; Kayser, O.; Martens, S.; Tissier, A.; van der Krol, S.; Wessjohann, L.; Warzecha, H. Natural Products–Modifying Metabolite Pathways in Plants. *Biotechnol. J.* **2013**, *8*, 1159–1171. [CrossRef]
7. Kim, O.T.; Kim, S.H.; Ohyama, K.; Muranaka, T.; Choi, Y.E.; Lee, H.Y.; Kim, M.Y.; Hwang, B. Upregulation of phytosterol and triterpene biosynthesis in *Centella asiatica* hairy roots overexpressed ginseng farnesyl diphosphate synthase. *Plant Cell Rep.* **2010**, *29*, 403–411. [CrossRef]
8. Yu, H.; Liu, M.; Yin, M.; Shan, T.; Peng, H.; Wang, J.; Chang, X.; Peng, D.; Zha, L.; Gui, S. Transcriptome analysis identifies putative genes involved in triterpenoid biosynthesis in *Platycodon grandiflorus*. *Planta* **2021**, *254*, 34. [CrossRef]
9. Broun, P. Transcription factors as tools for metabolic engineering in plants. *Curr. Opin. Plant Biol.* **2004**, *7*, 202–209. [CrossRef]
10. Chezem, W.R.; Clay, N.K. Regulation of plant secondary metabolism and associated specialized cell development by *MYBs* and *BHLHs*. *Phytochemistry* **2016**, *131*, 26–43. [CrossRef]
11. Mertens, J.; Pollier, J.; Bossche, R.V.; Lopez-Vidriero, I.; Franco-Zorrilla, J.M.; Goossens, A. The *BHLH* Transcription Factors *TSAR1* and *TSAR2* regulate triterpene saponin biosynthesis in *Medicago truncatula*. *Plant Physiol.* **2016**, *170*, 194–210. [CrossRef] [PubMed]
12. Xu, J.; Wang, Y.; Zhang, Y.; Xiong, K.; Yan, X.; Ruan, S.; Wu, X. Identification of a novel metabolic target for bioactive triterpenoids biosynthesis in *Ganoderma lucidum*. *Front. Microbiol.* **2022**, *13*, 878110. [CrossRef] [PubMed]
13. Liang, Y.; Zhao, S.; Zhang, X. Antisense suppression of cycloartenol synthase results in elevated ginsenoside levels in *Panax ginseng* hairy roots. *Plant Mol. Biol. Rep.* **2009**, *27*, 298–304. [CrossRef]
14. Hey, S.J.; Powers, S.J.; Beale, M.H.; Hawkins, N.D.; Ward, J.L.; Halford, N.G. Enhanced seed phytosterol accumulation through expression of a modified *HMG-CoA* Reductase. *Plant Biotechnol. J.* **2006**, *4*, 219–229. [CrossRef]
15. Wen, C.; Zhang, Z.; Shi, Q.; Niu, R.; Duan, X.; Shen, B.; Li, X. Transcription Factors *ZjMYB39* and *ZjMYB4* regulate farnesyl diphosphate synthase- and squalene synthase-mediated triterpenoid biosynthesis in Jujube. *J. Agric. Food Chem.* **2023**, *71*, 4599–4614. [CrossRef]

16. Zhang, F.; Kong, C.; Ma, Z.; Chen, W.; Li, Y.; Lou, H.; Wu, J. Molecular characterization and transcriptional regulation analysis of the *Torreya grandis* squalene synthase gene involved in sitosterol biosynthesis and drought response. *Front. Plant Sci.* **2023**, *14*, 1136643. [CrossRef]

17. Miettinen, K.; Pollier, J.; Buyst, D.; Arendt, P.; Csuk, R.; Sommerwerk, S.; Moses, T.; Mertens, J.; Sonawane, P.D.; Pauwels, L.; et al. The Ancient *CYP716* family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nat. Commun.* **2017**, *8*, 14153. [CrossRef]

18. Pandey, P.; Singh, S.; Negi, A.S.; Banerjee, S. Harnessing the versatility of diverse pentacyclic triterpenoid synthesis through hairy root cultures of various *Ocimum* species: An unprecedented account with molecular probing and up-scaling access. *Ind. Crops Prod.* **2022**, *177*, 114465. [CrossRef]

19. Gaudin, V.; Camilleri, C.; Jouanin, L. Multiple regions of a divergent promoter control the expression of the *Agrobacterium rhizogenes aux1* and *aux2* plant oncogenes. *Mol. Gen. Genet.* **1993**, *239*, 225–234. [CrossRef]

20. Camilleri, C.; Jouanin, L. The TR-DNA region carrying the auxin synthesis genes of the *Agrobacterium rhizogenes* agropine-type plasmid PRiA4: Nucleotide sequence analysis and introduction into tobacco plants. *Mol. Plant Microbe Interact.* **1991**, *4*, 155–162. [CrossRef]

21. Béclin, C.; Charlot, F.; Botton, E.; Jouanin, L.; Dore, C. Potential Use of the *aux2* gene from *Agrobacterium rhizogenes* as a conditional negative marker in transgenic cabbage. *Transgenic Res.* **1993**, *2*, 48–55. [CrossRef]

22. Srivastava, S.; Conlan, X.A.; Adholeya, A.; Cahill, D.M. Elite hairy roots of *Ocimum basilicum* as a new source of rosmarinic acid and antioxidants. *Plant Cell Tissue Organ Cult.* **2016**, *126*, 19–32. [CrossRef]

23. Dilshad, E.; Noor, H.; Nosheen, N.; Gilani, S.R.; Ali, U.; Khan, M.A. Influence of rol genes for enhanced biosynthesis of potent natural products. In *Chemistry of Biologically Potent Natural Products and Synthetic Compounds*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2021; pp. 379–404. ISBN 9781119640929. [CrossRef]

24. Palazon, J.; Cusido, R.M.; Gonzalo, J.; Bonfill, M.; Morales, C.; Pinol, M.T. Relation between the amount of *rolC* gene product and indole alkaloid accumulation in *Catharanthus roseus* transformed root cultures. *J. Plant Physiol.* **1998**, *153*, 712–718. [CrossRef]

25. Bulgakov, V.P.; Khodakovskaya, M.V.; Labetskaya, N.V.; Chernoded, G.K.; Zhuravlev, Y.N. The Impact of plant *rolC* oncogene on ginsenoside production by ginseng hairy root cultures. *Phytochemistry* **1998**, *49*, 1929–1934. [CrossRef]

26. Dubrovina, A.S.; Manyakhin, A.Y.; Zhuravlev, Y.N.; Kiselev, K.V. Resveratrol content and expression of phenylalanine ammonia-lyase and stilbene synthase genes in *rolC* transgenic cell cultures of *Vitis amurensis*. *Appl. Microbiol. Biotechnol.* **2010**, *88*, 727–736. [CrossRef]

27. Inyushkina, Y.V.; Kiselev, K.V.; Bulgakov, V.P.; Zhuravlev, Y.N. Specific genes of cytochrome p450 monooxygenases are implicated in biosynthesis of caffeic acid metabolites in *rolC*-transgenic culture of *Eritrichium sericeum*. *Biochemistry* **2009**, *74*, 917–924. [CrossRef]

28. Amanullah, M.; Mirza, B.; Zia, M. Production of artemisinin and its derivatives in hairy roots of *Artemisia dubia* induced by *rolA* gene transformation. *Pak. J. Bot.* **2016**, *48*, 699–706.

29. Kalita, R.; Patar, L.; Shasany, A.K.; Modi, M.K.; Sen, P. Molecular cloning, characterization and expression analysis of 3-Hydroxy- 3-Methylglutaryl Coenzyme A Reductase Gene from *Centella asiatica* L. *Mol. Biol. Rep.* **2015**, *42*, 1431–1439. [CrossRef]

30. Singh, A.K.; Kumar, S.R.; Dwivedi, V.; Rai, A.; Pal, S.; Shasany, A.K.; Nagegowda, D.A. A WRKY Transcription Factor from *Withania somnifera* regulates triterpenoid withanolide accumulation and biotic stress tolerance through modulation of phytosterol and defense pathways. *New Phytol.* **2017**, *215*, 1115–1131. [CrossRef]

31. Sharma, A.; Rather, G.A.; Misra, P.; Dhar, M.K.; Lattoo, S.K. Jasmonate responsive transcription factor *WsMYC2* regulates the biosynthesis of triterpenoid withanolides and phytosterol via key pathway genes in *Withania somnifera* (L.) Dunal. *Plant Mol. Biol.* **2019**, *100*, 543–560. [CrossRef]

32. Suzuki, H.; Achnine, L.; Xu, R.; Matsuda, S.P.T.; Dixon, R.A. A Genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant J.* **2002**, *32*, 1033–1048. [CrossRef]

33. Patel, N.; Patel, P.; Kendurkar, S.V.; Thulasiram, H.V.; Khan, B.M. Overexpression of squalene synthase in *Withania somnifera* leads to enhanced withanolide biosynthesis. *Plant Cell Tissue Organ Cult.* **2015**, *122*, 409–420. [CrossRef]

34. Unland, K.; Pütter, K.M.; Vorwerk, K.; van Deenen, N.; Twyman, R.M.; Prüfer, D.; Schulze Gronover, C. Functional characterization of squalene synthase and squalene epoxidase in *Taraxacum koksaghyz*. *Plant Direct* **2018**, *2*, e00063. [CrossRef]

35. Seo, J.W.; Jeong, J.H.; Shin, C.G.; Lo, S.C.; Han, S.S.; Yu, K.W.; Harada, E.; Han, J.Y.; Choi, Y.E. Overexpression of Squalene synthase in *Eleutherococcus senticosus* increases phytosterol and triterpene accumulation. *Phytochemistry* **2005**, *66*, 869–877. [CrossRef]

36. Kim, Y.S.; Cho, J.H.; Park, S.; Han, J.Y.; Back, K.; Choi, Y.E. Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in *Bupleurum falcatum*. *Planta* **2011**, *233*, 343–355. [CrossRef] [PubMed]

37. Mangas, S.; Bonfill, M.; Osuna, L.; Moyano, E.; Tortoriello, J.; Cusido, R.M.; Teresa Piñol, M.; Palazón, J. The Effect of Methyl jasmonate on triterpene and sterol metabolisms of *Centella asiatica*, *Ruscus aculeatus* and *Galphimia glauca* Cultured Plants. *Phytochemistry* **2006**, *67*, 2041–2049. [CrossRef] [PubMed]

38. Tsurumi, S.; Tsujino, Y. Chromasaponin I stimulates the growth of lettuce roots. *Physiol. Plant.* **1995**, *93*, 785–789. [CrossRef]

39. Tsurumi, S.; Ishizawa, K.; Rahman, A.; Soga, K.; Hoson, T.; Goto, N.; Kamisaka, S. Effects of chromosaponin i and brassinolide on the growth of roots in etiolated Arabidopsis seedlings. *J. Plant Physiol.* **2000**, *156*, 60–67. [CrossRef]

40. Li, Y.; Kong, D.; Fu, Y.; Sussman, M.R.; Wu, H. The Effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiol. Biochem.* **2020**, *148*, 80–89. [CrossRef]

41. Alcalde, M.A.; Cusido, R.M.; Moyano, E.; Palazon, J.; Bonfill, M. Metabolic *Gene Expression and Centelloside Production in Elicited Centella asiatica* hairy root cultures. *Ind. Crops Prod.* **2022**, *184*, 114988. [CrossRef]

42. Mirjalili, M.H.; Moyano, E.; Bonfill, M.; Cusido, R.M.; Palazón, J. Overexpression of the *Arabidopsis thaliana* squalene synthase gene in withania coagulans hairy root cultures. *Biol. Plant.* **2011**, *55*, 357–360. [CrossRef]

43. Sanchez-Muñoz, R.; Almagro, L.; Cusido, R.M.; Bonfill, M.; Palazon, J.; Moyano, E. Transfecting *Taxus Media* protoplasts to study transcription factors *BIS2* and *TSAR2* as activators of taxane-related genes. *Plant Cell Physiol.* **2020**, *61*, 576–583. [CrossRef] [PubMed]

44. Alcalde, M.A.; Müller, M.; Munné-Bosch, S.; Landín, M.; Gallego, P.P.; Bonfill, M.; Palazon, J.; Hidalgo-Martinez, D. 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.* **2022**, *13*, 1001023. [CrossRef] [PubMed]

45. Sheng, Y.Y.; Xiang, J.; Wang, K.R.; Li, Z.Y.; Li, K.; Lu, J.L.; Ye, J.H.; Liang, Y.R.; Zheng, X.Q. Extraction of squalene from tea leaves (*Camellia sinensis*) and its variations with leaf maturity and tea cultivar. *Front. Nutr.* **2022**, *9*, 755514. [CrossRef] [PubMed]

46. Rothblat, G.H.; Martak, D.S.; Kritchevsky, D. A quantitative colorimetric assay for squalene. *Anal. Biochem.* **1962**, *4*, 52–56. [CrossRef]

47. Araújo, L.B.D.C.; Silva, S.L.; Galvão, M.A.M.; Ferreira, M.R.A.; Araújo, E.L.; Randau, K.P.; Soares, L.A.L. Total phytosterol content in drug materials and extracts from roots of *Acanthospermum hispidum* by UV-VIS Spectrophotometry. *Rev. Bras. De Farmacogn.* **2013**, *23*, 736–742. [CrossRef]

48. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.

49. Ishwaran, H.; Kogalur, U.B. Fast unified random forests for survival, regression, and classification (RF-SRC) 2022. Available online: https://rdrr.io/cran/randomForestSRC/ (accessed on 18 July 2023).

50. Wei, T.; Simko, V. R Package "Corrplot": Visualization of a Correlation Matrix 2021. Available online: https://github.com/taiyun/corrplot (accessed on 18 July 2023).

General discussion

The primary aim of this thesis was to shed light on the effects of randomly integrated genes, especially the so-called *rol* genes, following infection of *Centella asiatica* with *Rhizobium rhizogenes*. In the studies carried out, we sought to determine the significance of morphological differences and changes in the expression levels of triterpene biosynthetic genes in the obtained hairy root lines in view of enhancing the production of centellosides, the most important bioactive compounds of *C. asiatica*. Gaining deeper understanding of how these plant biofactories function will facilitate the development of new approaches to the production of various compounds of great pharmacological and cosmetic interest.

Other goals of this work involve sustainable development, above all protecting terrestrial ecosystems and promoting their sustainable use. Given that the commercial importance of *C. asiatica* has led to its overexploitation in nature, hairy roots are envisaged as a viable long-term alternative production system that can protect this plant species from extinction. Additionally, achieving higher centelloside production levels will reduce their cost, rendering this valuable bioactive ingredient more accessible for the global population, and thereby promoting well-being and a healthier life.

To achieve our main goal of optimizing centelloside production in hairy roots of *C. asiatica*, we employed various biotechnological tools, namely genetic transformation, proteomics, transcriptomics, elicitation, machine learning, metabolomics, and metabolic engineering.

C. asiatica is known for its wide range of health and cosmetic benefits associated with its triterpenic components, centellosides, which are mainly concentrated in the leaves of the plant. Due to the scarcity of many geographical varieties of this species (Sabaragamuwa, 2018), adventitious roots from leaf-derived calli have been developed to produce asiaticoside, a predominant centelloside (Mercy et al., 2012). Although this approach proved viable for this component, its feasibility has not been demonstrated for the other key centellosides: madecassoside, madecassic acid, and asiatic acid.

Similarly, little success has been achieved in the production of these four triterpenic compounds in *C. asiatica* callus cultures, the values being lower than in the plant itself (Mangas et al., 2008). Nevertheless, the production of centellosides in cell cultures was shown to increase after elicitation with coronatine (Hidalgo et al., 2017), similar to the results obtained in the present study with hairy roots (Chapter 3).

Hairy roots have become a widely used biotechnological platform over the past decade in response to the need to create eco-sustainable long-term production systems of

compounds scarce in nature. The overexploitation of the producing plants is frequently driven by a low content and/or high medical-cosmetic interest of the target compound (Bensaddek et al., 2008). Hairy root cultures of various plants with these characteristics have been successfully established, including *Salvia miltiorrhiza* (diterpenoid tanshinones), *Ocimum basilicum* (rosmarinic acid), *Artemisia dubia* and *A. indica* (artemisinin), *Taxus cuspidata* (taxol), *Panax ginseng* (ginsenosides), among others (Pistelli et al., 2010; Ono and Tian, 2011; Trémouillaux-Guiller, 2013).

The widespread development of plant biofactories based on hairy roots is also due to their primary advantages, above all genetic stability, high biomass productivity, and efficient biosynthetic capacity (Gutierrez-Valdes et al., 2020). In this study, these benefits were evident in the acquisition and confirmation of various hairy root lines of *C. asiatica*, which were established to determine their morphological parameters and centelloside content (Chapter 1). These traits and production values remained constant over the following years when the hairy roots were studied for their proteomic profile and elicitation effects (Chapters 2 and 3). The experiments were generally developed over a 28-day period, throughout which growth and production rates remained constant.

Despite the aforementioned studies that have genetically transformed *C. asiatica* through infection with various strains of *R. rhizogenes*, the impact and underlying mechanisms of T-DNA genes, especially the *rol* and *aux* genes, once integrated in the plant genetic material are still not entirely clear. It has been suggested that the random insertion of these genes could lead to morphological and/or metabolic variations among different rhizoclones. This phenomenon was observed in our work with *C. asiatica* lines transformed by *R. rhizogenes* A4 (Chapter 1), *TSAR2*, and *At-SQS* (Chapter 4).

To study how these T-DNA genes affect and relate to morphological traits, hormonal profiles, and centelloside contents, various machine learning models were employed, both supervised and unsupervised. These computational and mathematical models are gaining traction in the field of plant biotechnology (Somvanshi and Mishra, 2015) and have also been utilized with hairy root cultures (Goswami et al., 2018). For example, an artificial neural network-based model demonstrated the effect of different reactor parameters on *Artemisia annua* hairy root biomass (Osama et al., 2013), and agent-based modeling approaches were employed to enhance total root length, branching point distribution, segment distribution, and secondary metabolite content in *Beta vulgaris* hairy root cultures (Lenk et al., 2014).

The use of principal component analysis (PCA), an unsupervised machine learning model, revealed that the high expression of genes integrated via *R. rhizogenes*,

specifically the *rol* and *aux* genes, was associated with *C. asiatica* hairy root lines producing higher levels of centellosides, labeled "HIGH" (Chapter 1). It was also observed that these T-DNA genes had a favorable impact on the expression of centelloside biosynthetic genes, especially *CYP* and *UGT* genes (Chapter 4). We can therefore infer that the *rol* and *aux* genes, directly or indirectly, generated a positive effect on the centelloside content in various transformed lines of this plant by increasing the expression levels of metabolic genes, especially those involved at the end of the pathway. The new insights and most interesting results about *rol* genes obtained in this PhD thesis are summarized in Table 3.

Table 3. Key insights and intriguing discoveries regarding *rol* genes obtained in this PhD thesis.

Chapter	New findings
Chapter 1	Enhanced expression of T-DNA genes, especially *rolA* and *aux1*, correlates with improved morphological traits and centelloside content.
Chapter 2	The *rolD* gene appears to impact secondary metabolite production of specialized metabolites, specifically centellosides. The copy number of *rol* genes appears to have no influence on morphological and metabolic differences.
Chapter 4	Across various lines of *R. rhizogenes*-transformed hairy roots (A4, At-*SQS*, and *TSAR2*): *rolA* gene expression influences *HMGR* gene expression. *rolC* gene expression influences *CYP* and *UGT* gene expression. *aux2* gene expression affects β-*AS*, CYP, and *UGT* gene expression. *rol* and *aux1* gene expression influences *CYP716C11* gene expression.

To determine if gene expression levels or hormonal profiles could be used to classify the different lines of hairy roots according to their centelloside production, four different machine learning classification algorithms were employed: support vector machine (SVM), random forest (RF), artificial neural network (ANN), and linear discriminant analysis (LDA) (Chapter 1). Our results classified the lines as HIGH, MEDIUM, and LOW

producers, showing that the best algorithm for categorization based on gene expression was random forest, whereas for hormonal profiles, RF and SVM algorithms were optimal. These algorithms have already been employed in biotechnological platforms; for example, in callus cultures of *Cannabis sativa*, the effect of various plant growth regulators was evaluated using RF and SVM in combination with a genetic algorithm (Hesami and Jones, 2021). To determine the segmentation of individual cells in *Arabidopsis thaliana*, the SVM classifier algorithm was used, based on a cell descriptor constructed from the shape and edge strength of the cell contours (Marcuzzo et al., 2009).

PCA also allowed us to determine which phytohormone was positively related to increased growth and centelloside production in the *C. asiatica* hairy root lines and which had the opposite effect (Chapter 1). The results showed that isopentenyladenosine (IPA) was positively associated with hairy root lines that produced more centellosides, while abscisic acid (ABA) was more closely related to those with lower production. In other words, high levels of IPA can lead to increased growth and centelloside contents, while high levels of ABA would produce the opposite effect. This result was confirmed through a feeding experiment for low- and high-producing lines, supplementing the media with IPA and ABA, respectively. The relationship between these two hormones has been observed previously, with higher levels of IPA reported in *rolC*-transgenic tobacco plants compared to wild plants (Nilsson et al., 1993), whereas ABA decreased in transgenic plants of the same species (Estruch et al., 1991a). In our study, ABA was most influenced by the *rol* genes, especially *rolB*.

Despite being able to classify the transformed lines based on various parameters using the "ladder" of machine learning, and avoiding the "snake" of rhizoclone variation, we still did not have a clear idea of how the *rol* and *aux* genes influence the centelloside production process, including at what molecular level the changes are initiated and what genes are involved.

To address this, and climb further toward our objectives, we decided to employ the tool of proteomics. Thus, we determined which differential proteins were associated with the different centelloside production levels in the three groups of transformed lines, in addition to the adventitious roots used as a control (Chapter 2).

Orthogonal partial least squares-discriminant analysis (OPLS-DA), a multivariate statistical method similar to PCA, was used to identify biomarkers that could differentiate the groups of hairy root lines. Standing out among them were the Tr-type G domain-containing protein, classified as a translation factor within the elongation factor (EF-

G/EF-2) subfamily related to the protein synthesis process, an alcohol dehydrogenase (ADH) protein, and Bet v I/Major latex protein (MLP). The most interesting findings regarding centelloside content obtained in this PhD thesis are summarized in Table 4.

Table 4. Comprehensive insights and intriguing findings about centelloside content obtained in this PhD Thesis.

Chapter	New findings
Chapter 1	In *R. rhizogenes* A4-transformed lines, a correlation between centelloside content and morphological parameters was observed. Machine learning models were used to group different hairy root lines based on their centelloside production levels: high, medium, and low. High centelloside content was associated with elevated levels of the phytohormone IPA, while low centelloside content was linked to elevated levels of the phytohormone ABA.
Chapter 2	High centelloside content was associated with various proteins, including the rolD protein with ornithin cyclase desaminase activity, as well as the Tr-type G domain-containing protein and an alcohol dehydrogenase protein.
Chapter 3	*C. asiatica* hairy roots elicited with CORO and CORO+MeJa showed the highest total centelloside production values after 14 days. The predominant centelloside was madecassoside, regardless of the elicitor used. Centelloside patterns varied depending on the elicitor and time of elicitation, but madecassoside and asiaticoside predominated.
Chapter 4	Hairy root lines harboring the At-*SQS* gene (LS) exhibited the highest centelloside contents (from the HIGH group). LS lines showed a negative correlation between centelloside content and morphological traits. In contrast, in LST and LT lines, this effect seemed to be balanced by the *TSAR2* transcription factor.

High levels of alcohol dehydrogenase (ADH) expression in *A. thaliana* have been associated with increased tolerance to anoxia and improved root growth (Shiao et al., 2002). Bet v I/Major latex proteins (MLPs) are linked to enhanced 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 IAA levels in hairy roots. This is significant due to the crucial role that this plant hormone plays in root induction and development, a fact that could be related to the high levels of *aux1* gene expression observed in the HIGH hairy root lines (Chapter 1). This gene is reported to cause differences in root growth and morphology (Ozyigit et al., 2013) because it mediates the transport/flow of the aforementioned phytohormone (Yu et al., 2015).

Lastly, the most intriguing finding yielded by OPLS-DA algorithm was the detection of the ornithin cyclase desaminase (OCD) biomarker, which suggests a potential link between *Rhizobium*-mediated genetic alteration and the proteome composition of hairy roots. This is of interest because the *rolD* gene induces an enzymatic transformation of ornithine into proline through an OCD-like function, which may offer a credible rationale for its involvement in the generation of hairy roots (Trovato et al., 2001, 2018).

It is worth noting that previous studies have reported a significant increase in proline levels in the growth region of primary maize roots under conditions of limited water availability, indicating that proline synthesis plays a crucial role 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), including extensins and arabinogalactan proteins, which serve as integral structural components of the plant cell wall (Kishor et al., 2015). HRGPs are believed to oversee essential processes such as cell division, the self-assembly of the cell wall, and cell elongation, all of which could contribute to the observed effects 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 serves as a precursor. The overexpression of arginine decarboxylase, another polyamine precursor, has been shown to increase putrescine levels and hinder root growth in tobacco plants (Masgrau et al., 1997).

This result contradicts the hypothesis that the *rolD* gene of *R. rhizogenes* T-DNA is of low importance in hairy root induction (Trovato et al., 1997; Bulgakov, 2008). In some cases, this gene has not even been found in *rol* gene detection analyses (Trovato et al., 2018).

Among the *rol* genes, as mentioned, the *rolB* gene is the most significant for the production and accumulation of various compounds and can promote root growth in different plant species on its own. For instance, in *Artemisia carvifolia* hairy roots, it is responsible for the production of terpenoids such as artemisinin and its derivatives (Dilshad et al., 2015). The same gene also plays a role in the production of alkaloids such as anthraquinones in *Rubia cordifolia* callus cultures (Shkryl et al., 2008). Additionally, this gene is involved in the production of phenolic compounds like resveratrol in *rolB*-transgenic *Vitis amurensis* cells (Kiselev et al., 2007) and isoflavonoids in transgenic *Maackia amurensis* cells (Grishchenko et al., 2016).

The *rolB* oncogene also affects hairy root growth because it increases the level of endogenous IAA through the hydrolysis of bound auxins. Auxins accumulate at the top of the root and stimulate the production of H^+ ions in the cells of the root elongation zone. This lowers the pH in the cell walls, which activates enzymes that loosen the cell walls and allow cell elongation. The result is rapid downward root growth (Grieneisen et al., 2007).

Similarly, the *rolC* gene has $\beta-$glucosidase activity, promoting the release of free active cytokinins from their active conjugates (Estruch et al., 1991a). It is responsible for the production of different specialized metabolites such as tropane alkaloids in rolC-transformed *Atropa belladonna* roots (Bonhomme et al., 2000). In rolC-transgenic *Catharanthus roseus* cell cultures, it produces indole alkaloids (Palazon et al., 1998). Another alkaloid modified by this oncogene is nicotine in *Nicotiana tabacum* (Palazon et al., 1997). Additionally, the gene is involved in increased production of phenolic compounds such as ginsenosides in *Panax ginseng* hairy roots (Bulgakov et al., 1998), and resveratrol in transgenic *Vitis amurensis* cell cultures (Dubrovina et al., 2010), compared to non-transformed roots and callus cultures, respectively. Other phenolic compounds increased by the *rolC* oncogene include caffeoylquinic acid in transformed *Cynara cardunculus* cells (Vereshchagina et al., 2014), and rutin, quercetin, isoquercetin, and caffeic acid in rolC-transgenic *Artemisia annua* plants (Dilshad et al., 2016).

The least studied T-DNA gene is *rolD*, so to assess its importance, as well as the impact of all the *R. rhizogenes* oncogenes on the proteomic profiles of the hairy root lines, we analyzed and compared the *rol* gene copy numbers and expression levels. The copy numbers were consistent (2 copies) for all the genes in all the transformed root lines. However, the expression level of the *rolD* gene increased significantly in the HIGH lines,

correlating with the proteomic observations, being approximately 5 to 8-fold higher than in the other hairy root groups.

In many plants, it is commonly observed that one or two integration events often result in a higher level of expression of the exogenous gene, but additional copies may lead to gene silencing (Yang et al., 2005). However, according to our findings, the expression and accumulation of the *rolD* gene appear to play a determinant role. Little is known about the *rolD* gene's effect on hairy roots, but this result led us to hypothesize that each *rol* gene is crucial for the growth and metabolism of hairy roots, the effect depending on the location of its insertion into the plant genetic material. As the insertion of the oncogenes occurs randomly, this hypothesis could be challenging to demonstrate, although *rol* genes can be targeted with techniques such as CRISPR-CAS9, as reported recently in *Soybean* hairy roots (Niazian et al., 2022).

Unexpectedly, in our study, proteomics analysis did not identify other proteins associated with the T-DNA. This could be attributed to the specific techniques and parameters used, which may not have captured the full spectrum of proteins, especially those present in low abundance.

Considering the above, other "ladders" or biotechnological tools were required to increase centelloside yields in the hairy root lines showing optimal growth and basal production levels. Therefore, based on the machine learning classification (Chapter 1), we selected the transformed *C. asiatica* line (L1) from the HIGH group with the best growth and the highest endogenous levels of these triterpenoid saponins. L1 was treated with various elicitors to study their impact on growth, the expression of centelloside metabolic genes, and the production of the target compounds (Chapter 3).

Elicitation is a mechanism that activates specialized metabolism by simulating stress conditions through the addition of various compounds known as elicitors, which can be either abiotic or biotic. In this study, biotic elicitors were employed because they often exhibit specificity, meaning they can activate specific defense responses that redirect resources towards particular threats such as pathogens and insects, inducing secondary metabolism (Yoshikawa et al., 1993; Thakur et al., 2019).

The elicitors used were salicylic acid (SA), methyl jasmonate (MeJa), coronatine (Coro), and MeJa/Coro. The initial effect of elicitation in all cases was a reduction in growth in the transformed root lines after 14 days of supplementation, a result previously observed in *Glycine max* hairy roots following treatment with MeJa (Theboral et al., 2014). Mangas et al. (2006) and Hidalgo et al. (2017) found that all tested elicitors inhibited the growth

of *C. asiatica* cell and plant cultures, measured as fresh weight (FW) and dry weight (DW). SA was the elicitor that caused the least reduction in growth, similar to observations in *Tripterygium wilfordii* and *G. max* hairy root cultures (Theboral et al., 2014; Zhu et al., 2014).

Regarding the total centelloside yields, SA-elicited hairy roots exhibited the lowest contents at day 7 post-elicitation (10 mg/g DW), possibly because this elicitor did not create sufficient stress, as indicated by its minimal impact on growth. CORO and CORO+MeJa resulted in the highest total centelloside content at day 14 post-elicitation (116 and 125 mg/g DW, respectively.)

In recent years, various studies on *C. asiatica* have improved centelloside production using 100 µM MeJa. For example, (Kim et al., 2007) achieved 7.12 mg/g DW of asiaticoside in *C. asiatica* hairy roots after 3 weeks of elicitation, a 5-fold increase compared to the control, and (Ruslan et al., 2012) reported a very similar 4.9-fold increase in asiaticoside in cell suspension cultures. Yoo et al. (2011) described an enhanced production of both asiaticoside and madecassoside but a reduction in madecassic acid, an effect also observed here, with the amount decreasing from 0.32 mg/g to 0.09 mg/g of DW. Similarly, 100 µM MeJa increased asiaticoside and asiatic acid content in shoot, callus, and cell suspension cultures of *C. asiatica* (Krishnan et al., 2019).

The other elicitor employed, CORO, is a bacterial toxin that has garnered interest in recent years as it mimics the activity of the isoleucine-conjugated form of jasmonic acid but is effective at lower concentrations (Ramirez-Estrada et al., 2016). Its effectivity was confirmed in the present study, the production of all centellosides being enhanced by CORO treatment, especially at day 14 after elicitation. Similar results have been described by Hidalgo et al. (2017) in cell suspensions of *C. asiatica* and Onrubia et al.)2013) in cell cultures of *Taxus media*, in which CORO induced higher taxane production than MeJa.

The synergistic effect between CORO and MeJa yielded the best results, although only a slight increase was observed compared to elicitation with CORO alone. We hypothesize that the latter is the primary driver of enhanced centelloside production.

Regarding the centelloside pattern, we observed slight variations depending on the duration of elicitation and the elicitor used, but madecassoside always predominated. This could be due to the fact that asiaticoside and madecassoside are more stable in intracellular spaces than their acid forms, and the elicitors favored the glycosidation of

these compounds. A comparison of centelloside contents between different plant biotechnological platforms is presented in Table 5.

Table 5. Centelloside content in elicited and non-elicited cell suspensions and hairy roots of *C. asiatica.*

Culture system	Elicitor	Triterpene content	Reference
Cell suspensions	MeJa (100 µM)	A + M: 1.1 mg/g DW	Bonfill et al. (2011)
Cell suspensions	SA (100 µM)	A: 102.6 mg/g DW	Loc and Giang, (2012)
Cell suspensions	MeJa (200 µM)	M: 10.6 µg/g FW A: 20.9 µg/g WW MA: 8 µg/g WW AA: 4.9 µg/g WW	James et al. (2013)
Cell suspensions	SA (100 µM)	M: 114.2 mg/g DW	Loc et al. (2017)
Hairy roots (wild type)	MeJa (100 µM)	A: 7.1 mg/g DW	Kim et al. (2007)
Hairy roots (*PgFPS* gene)	MeJa (100 µM)	M: 0.7% DW A: 0.38% DW	Kim et al. (2010)
Hairy root (wild type)	CORO (1 µM)	TC: 7.3 mg/g DW	Hidalgo et al. (2017)

Hairy root (wild type)	No elicitation CORO (1 µM) CORO (1 µM) +MeJa (100 µM)	TC: 5.2 mg/g DW TC: 116 mg/g DW TC: 125 mg/g DW	Chapter 3
Hairy root (*At-SQS*, *TSAR2* and *At-SQS*+*TSAR2*)	*At-SQS* (no elicitation) *TSAR2* (no elicitation) *At-SQS*+*TSAR2* (no elicitation)	TC: 6.5 mg/g DW TC: 5.4 mg/g DW TC: 5.7 mg/g DW	Chapter 4

WW:Wet Weight, DW:Dry Weight, M: Madecassoside, A:Asiaticoside, MA: Madecasic acid, AA: Asiatic acid and TC:Total Centelloside.

We also analyzed how elicitation affected the expression of genes in the centelloside metabolic pathway. We observed that after elicitation, gene expression increased in the first 8 to 12 hours and then gradually decreased over time. The defense response and activation of related genes are generally rapid after elicitation process (Angelova et al., 2014). However, it was shown that the effect of elicitation can persist in the medium to long term, as the values obtained with Coro and Coro+MeJa treatment after 7 days were similar to those at 24-36 hours after elicitation, especially for the cytochrome P450-dependent monooxygenase (*CYP*s) genes.

*CYP*s are a class of enzymes classified as oxidoreductases, constituting one of the largest enzyme families. *CYP*s play a significant role in generating diverse metabolites through a wide range of chemical transformations (Pandian et al., 2020). They are responsible for catalyzing hydroxylation reactions in primary and secondary metabolism across various organisms. They are involved in end pathway steps, modifying products with oxidation, hydroxylation, and glycosylation in the centelloside biosynthetic pathway (Augustin et al., 2011).

The new findings and most interesting results regarding centelloside biosynthetic gene expression obtained in this PhD thesis are summarized in Table 6.

Table 6. Novel insights and findings regarding gene expression in the biosynthetic pathway of centellosides: A PhD thesis overview.

Chapter	New findings
Chapter 3	Gene expression in the centelloside biosynthetic pathway increased in the initial hours following elicitation, gradually diminishing after 36 hours. *CYP* and β-*AS* genes were the most highly overexpressed after elicitation, particularly with CORO and CORO+MeJa treatments. The overexpression of the *CYPs* and *UGT* genes post-elicitation showed an enhancement of madecassoside and asiaticoside contents.
Chapter 4	The genes *HMGR* and *SQS* are influenced by the overexpression of the transcription factor *TSAR2*. Gene expression in the final steps of centelloside biosynthesis, involving *CYP* and *UGT* genes, were influenced by the *rol* and *aux* genes.

Similarly, the *CYP* genes were the most overexpressed genes compared to the non-elicited transformed root line (300 to 3000-fold higher), particularly *CYP714E19*, highlighting its important role in the elicitation process, above all in response to CORO.

The *UGT* gene, which plays a key role in glycosylating sapogenins (acids) to yield the corresponding saponins (Kim et al., 2017), showed an increase in expression after elicitation but not to the extent of the *CYP* genes (only 100-fold higher than the control values). Hence, the overexpression of these genes was correlated with the elevated contents of madecasoside and asiaticoside.

Considering the roles of *CYP* and *UGT* genes, the establishment of *C. asiatica* hairy roots elicited with biotic factors could achieve similar results to a metabolic engineering approach directed specifically at centelloside production. More information about the *CYP* and *UGT* genes responsible for the final steps of the metabolic pathway would be useful, as they are are favorably affected by the *rol* and *aux* genes (Chapter 4).

The β-*AS* gene was also overexpressed after elicitation. This gene is involved in a key step of triterpenoid biosynthesis, catalyzing the formation of α and β-amyrin, the initial precursors of centellosides (Kim et al., 2007). Finally, the *SQS* gene, which is involved in the formation of the universal precursor of triterpenoid saponins and phytosterols (Azerad, 2016), was overexpressed in all treatments. The subsequent decrease in expression is likely because *SQS* activity is not limited to centelloside biosynthesis. As it

is involved in the upregulation of triterpene production (Lee et al., 2004), modulating its expression could prevent high accumulations of saponins and phytosterols.

Based on the findings described in Chapter 3, it was determined that excessive flux in the centelloside metabolic pathway could be a pitfall that should be addressed to optimize the biotechnological system of *C. asiatica* hairy roots. Therefore, to further improve production in *C. asiatica* hairy roots and our understanding of the process, the "ladder" of metabolic engineering was employed to introduce one or more genes involved in the triterpenoid metabolic pathway. The engineered root cultures were studied for changes in the expression levels of triterpenoid pathway genes, morphological parameters, and the content of key bioactive compounds such as squalene, sterols, and centellosides.

The first transgene introduced into the plant genetic material was the transcription factor known as triterpenoid saponin activating regulator 2 (*TSAR2*), which promotes triterpenoid production by increasing the expression of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase (*HMGR*) gene. The second transgene was the *SQS* gene from *A. thaliana*, which is a key enzyme in triterpenoid biosynthesis. Three types of transformed lines were obtained, transformed with the *TSAR2* transgene (LT lines), the *At-SQS* transgene (LS lines), and both transgenes (LST lines). The highly productive "HIGH" transformed line, derived from wild-type *R. rhizogenes* A4 (Chapter 2), was used as the control.

In general, the LT and LST lines overexpressed the *HMGR* and *SQS* genes compared to the control line, demonstrating that the transcription factor directly affects these two key genes of the triterpenoid metabolic pathway. Similarly, overexpression of the transcription factor *TSAR2* in *Medicago truncatula* hairy roots increased the transcript levels of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase (*HMGR*) gene, which is responsible for catalyzing the conversion of 3-hydroxy-3-methylglutaryl-CoA (*HMG-CoA*) to MVA via two successive hydride transfers, using two molecules of NADPH (Mertens et al., 2015).

It is interesting to note that in the root lines transformed with the transcription factor *TSAR2*, a natural increase in the expression of two key genes in the pathway was observed. Moreover, in the LST lines, which also exhibited exogenous overexpression of At-*SQS*, we may conclude that these two effects together could address the issue of the relatively low level of *SQS* overexpression after elicitation in hairy roots transformed by wild-type *R. rhizogenes* A4 strain (Chapter 2). Thus, employing LST lines in combination with an elicitation process could constitute a holistic approach to increase

the expression levels of genes involved in the centelloside biosynthetic pathway and therefore the production of these compounds.

The utility of pairing elicitation with metabolic engineering was reinforced by the observation that the LT and LST lines (especially the latter) did not exhibit significant overexpression of the *CYP* and *UGT* genes compared to the control, a characteristic that could be reversed with elicitation, specifically with CORO and CORO+MeJa (Chapter 2).

We then applied a regression analysis model to analyze the impact of the *rol* and *aux* transgenes from the *R. rhizogenes* plasmid on various genes in the centelloside biosynthetic pathway. This model revealed that the *rol* and *aux1* genes affected the expression of the *CYP716C11* (*CYP11*) gene, precisely the *CYP* gene that maintained high medium- to long-term expression levels with CORO+MeJa elicitation (Chapter 3). This could suggest that the action of these elicitors somehow affects the *rol* and *aux1* genes with a consequent increase in the *CYP11* gene expression levels.

The *aux2* gene affected the expression of β-*AS*, *CYPs*, and *UGT* genes, whereas the *rolC* gene only impacted the *CYP* and *UGT* genes. This demonstrates that the *rol* and *aux* genes are primarily involved in the overexpression of genes responsible for the final steps in the centelloside biosynthetic pathway.

The *rolA* gene was the only T-DNA gene found to influence the expression of the *HGMR* gene. Considering that, according to the PCA, the *rolA* gene was among the genes with the greatest influence on transformed root growth and centelloside production (Chapter 1), we can infer that it has a strong impact on the metabolic pathway to ultimately increase triterpene production, unlike the other *rol* genes.

The importance of the *rolA* gene was observed in *Artemisia dubia* hairy roots, where it directly increased the production of terpenoids such as artemisinin (Amanullah et al., 2016). In transgenic roots of *Nicotiana tabacum*, the *rolA* gene was found to enhance alkaloid production, specifically nicotine (Palazon et al., 1997). Similarly, in transgenic *Rubia cordifolia* callus cultures, *rolA* has been demonstrated to increase the production of anthraquinones (Veremeichik et al., 2019).

Subsequently, we determined how the *TSAR2* and *At-SQS* transgenes affected the profiles of sterols, squalene, and centellosides. As mentioned, squalene serves as the universal precursor for triterpenoid saponins and phytosterols (Azerad, 2016), isoprenoid-derived compounds that play crucial roles in plant growth and development. Sterols are integral components of biological membranes and act as precursors for the

synthesis of essential steroid hormones such as brassinosteroids in plants (Boutté and Grebe, 2009).

The LST and LT lines exhibited significantly higher levels of phytosterols compared to the control line, suggesting a clear effect of the transcription factor on the production of these metabolites. This effect has already been observed with other transcription factors such as *WRKY* or *MYC2* (*WsWRKY1* and *WsMYC2*), which influenced phytosterol biosynthesis in *Withania somnifera* (Singh et al., 2017; Sharma et al., 2019).

The LS lines did not show higher levels of phytosterols compared to the LST and LT lines, but their contents of squalene and centellosides were higher, indicating that *At-SQS* affected centelloside accumulation. A possible explanation is that triterpene biosynthesis occurs once enough phytosterol has been synthesized, signaling the end of growth (Rogowska and Szakiel, 2021; Kunjumon et al., 2022). Supporting this observation, inhibiting the phytosterol biosynthetic pathway by antisense cycloartenol synthase led to a notable increase in ginsenosides (triterpenes) in *Panax Ginseng* hairy roots (Liang et al., 2009). A similar antagonistic interaction between the triterpene and sterol pathways has been observed in transgenic *Bupleurum falcatum* roots (Kim et al., 2011). Also, similar results were observed in *C. asiatica* plants under MeJa elicitation, in which free sterol decreased while centellosides increased (Mangas et al., 2006).

To complete the analysis, we determined how the content of squalene, sterols, and centellosides influences morphological traits in different hairy root lines. In the control lines, robust positive correlations were observed between morphological growth factors and the contents of squalene and phytosterol, a relationship previously observed with centellosides (Chapter 1).

In the LS lines, the morphological values were negatively correlated with centelloside and squalene contents. Conversely, in the LST and LT lines, this negative effect on morphological traits appeared to be counteracted by the presence of the transcription factor *TSAR2*, which directly affects the *HMGR* gene and quite possibly genes from the sterol biosynthetic pathway, resulting in a higher phytosterol content. These effects are akin to those induced by this transcription factor and contrast with the findings documented by (Hey et al., 2006), who observed a 2.44-fold increase in sterol production compared to the control in *A. thaliana*. These findings further emphasize the regulatory interplay between the phytosterol and triterpene production pathways, along with their impacts on the morphological traits of C*. asiatica* hairy roots.

Conclusions

Based on the results described in the chapters of this PhD thesis, whose primary goal was to determine the significance of randomly integrated genes following *Rhizobium rhizogenes* infection, particularly the so-called *rol* genes, and their impact on the formation of hairy roots in *Centella asiatica,* ultimately to increase centelloside production, we can conclude that:

1. Machine learning models demonstrated that changes in the expression levels of *aux1* and *rol* genes are associated with variations in the hormonal profile, morphological traits and centelloside production in *C. asiatica* hairy roots.

2. Determining the degree of influence of each target gene on the profile of individual hormones allowed us to establish connections between elevated centelloside levels and isopentenyladenosine (IPA), as well as between reduced centelloside content and acid abscisic (ABA). These correlations were confirmed in feeding experiments, which reversed the effects on the growth and metabolite production of the hairy root lines.

3. Alterations in morphology and centelloside production cannot be attributed to the copy number of T-DNA integrated in the plant genome. Proteomic analysis indicates that the primary factor behind these changes is the ornithine cyclodeaminase (OCD) protein encoded by the *rolD* gene.

4. After elicitation, a selected hairy root line exhibited increased expression of genes in the final stages of the centelloside biosynthetic pathway, heading towards a greater centelloside production.

5. The application of metabolic engineering, through the introduction of the transcription factor *TSAR2* and the *At-SQS* gene, resulted in enhanced expression of genes in the initial steps of the centelloside metabolic pathway, boosting precursor availability for centelloside production.

6. Some of the increased precursors were driven to the biosynthesis of phytosterols and caused morphological alterations in the hairy root phenotype, showing the competition between centelloside and phytosterol pathways.
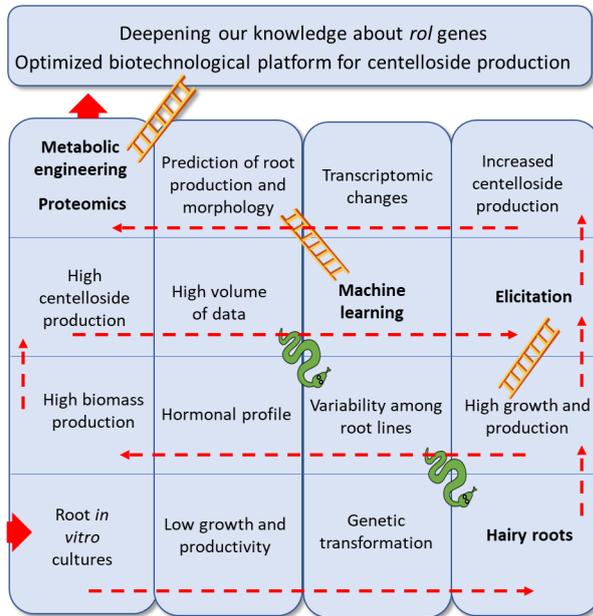
Finally, we can conclude that the game of snakes and ladders is not over, and it is necessary to continue climbing ladders by applying new strategies to explore the mechanisms of action of *R. rhizogenes* genes and developing new approaches to optimize the centelloside production in *C. asiatica* hairy root cultures.

General references

Afroz, S., Warsi, Z. I., Khatoon, K., Sangwan, N. S., Khan, F., and Rahman, L. U. (2022). Molecular cloning and characterization of Triterpenoid Biosynthetic Pathway Gene HMGS in *Centella asiatica* (Linn.). *Mol Biol Rep* 49, 4555–4563. doi: 10.1007/S11033-022-07300-9.

Afzan, A., Ahmad, N., Zaidan, M. R. S., Rain, N., Badrul, A. R., Adlin, A., et al. (2005). In vitro screening of five local medicinal plants for antibacterial activity using disc diffusion method Molecular characterization of *Corynebacterium diptheria* View project In vitro screening of five local medicinal plants for antibacterial activity using disc diffusion method. *Trop Biomed* 22, 165–170.

Alcalde, M. A., Perez-Matas, E., Escrich, A., Cusido, R. M., Palazon, J., and Bonfill, M. (2022). Biotic Elicitors in Adventitious and Hairy Root Cultures: A Review from 2010 to 2022. *Molecules* 27. doi: 10.3390/MOLECULES27165253.

Alpizar, E., Dechamp, E., Lapeyre-Montes, F., Guilhaumon, C., Bertrand, B., Jourdan, C., et al. (2008). *Agrobacterium rhizogenes*-transformed Roots of Coffee (*Coffea arabica*): Conditions for Long-term Proliferation, and Morphological and Molecular Characterization. *Ann. Bot.* 101, 929–940. doi: 10.1093/AOB/MCN027.

Altabella, T., Angel, E., Biondi, S., Palazón, J., Bagni, N., and Piñol, M. T. (1995). Effect of the rol genes from *Agrobacterium rhizogenes* on polyamine metabolism in tobacco roots. *Physiol. Plant.* 95, 479–485. doi: 10.1111/J.1399-3054.1995.TB00866.X

Amani, S., Mohebodini, M., Khademvatan, S., and Jafari, M. (2020). *Agrobacterium rhizogenes* mediated transformation of *Ficus carica* L. for the efficient production of secondary metabolites. *J Sci Food Agric* 100, 2185–2197. doi: 10.1002/JSFA.10243.

Amanullah, M., Mirza, B., and Zia, M. (2016). Production of artemisinin and its derivatives in hairy roots of *Artemisia dubia* induced by rolA gene transformation. *Pak J Bot* 48, 699–706.

Aminfar, Z., Rabiei, B., Tohidfar, M., and Mirjalili, M. H. (2019). Identification of key genes involved in the biosynthesis of triterpenic acids in the mint family. *Scientific Reports 2019 9:1* 9, 1–15. doi: 10.1038/s41598-019-52090-z.

Angelova, Z., Georgiev, S., and Roos, W. (2014). Elicitation of Plants. *Biotechnology & Biotechnological Equipment* 20, 72–83. doi: 10.1080/13102818.2006.10817345.

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.

Araújo, L.B.D.C.; Silva, S.L.; Galvão, M.A.M.; Ferreira, M.R.A.; Araújo, E.L.; Randau, K.P.; Soares, L.A.L. (2013). Total phytosterol content in drug materials and extracts from roots of *Acanthospermum hispidum* by UV-VIS Spectrophotometry. *Revista Brasileira de Farmacognosia* 23, 736–742, doi:10.1590/S0102-695X2013000500004.

Arribas-López, E., Zand, N., Ojo, O., Snowden, M. J., and Kochhar, T. (2022). A Systematic Review of the Effect of *Centella asiatica* on Wound Healing. *International Journal of Environmental Research and Public Health 2022, Vol. 19, Page 3266* 19, 3266. doi: 10.3390/IJERPH19063266.

Augustin, J. M., Kuzina, V., Andersen, S. B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* 72, 435–457. doi: 10.1016/J.PHYTOCHEM.2011.01.015.

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.

Baek, S., Han, J. E., Ho, T. T., and Park, S. Y. (2022). Development of Hairy Root Cultures for Biomass and Triterpenoid Production in *Centella asiatica*. *Plants* 11, 148. doi: 10.3390/PLANTS11020148/S1.

Baek, S., Ho, T. T., Lee, H., Jung, G., Kim, Y. E., Jeong, C. S., et al. (2020). Enhanced biosynthesis of triterpenoids in *Centella asiatica* hairy root culture by precursor feeding and elicitation. *Plant Biotechnol Rep* 14, 45–53. doi: 10.1007/S11816-019-00573-W.

Bagal, D., Chowdhary, A. A., Mehrotra, S., Mishra, S., Rathore, S., and Srivastava, V. (2023). Metabolic engineering in hairy roots: An outlook on production of plant secondary metabolites. *Plant Physiology and Biochemistry* 201, 107847. doi: 10.1016/J.PLAPHY.2023.107847.

Bai, L., Zhang, G., Zhou, Y., Zhang, Z., Wang, W., Du, Y., et al. (2009). Plasma membrane-associated proline-rich extensin-like receptor kinase 4, a novel regulator of Ca2+ signalling, is required for abscisic acid responses in *Arabidopsis thaliana*. *Plant J.* 60, 314–327. doi: 10.1111/J.1365-313X.2009.03956.X.

Banerjee, S., Singh, S., and Rahman, L. U. (2012). Biotransformation studies using hairy root cultures — A review. *Biotechnol Adv* 30, 461–468. doi: 10.1016/J.BIOTECHADV.2011.08.010.

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.

Batra, J., Dutta, A., Singh, D., Kumar, S., and Sen, J. (2004). Growth and terpenoid indole alkaloid production in *Catharanthus roseus* hairy root clones in relation to left- and right-termini-linked Ri T-DNA gene integration. *Plant Cell Rep.* 23, 148–154. doi: 10.1007/s00299-004-0815-x.

Belwal, T., Andola, H. C., Atanassova, M. S., Joshi, B., Suyal, R., Thakur, S., et al. (2019). *Gotu Kola* (*Centella asiatica*). *Nonvitamin and Nonmineral Nutritional Supplements*, 265–275. doi: 10.1016/B978-0-12-812491-8.00038-2.

Bensaddek, L., Villarreal, M. L., and Fliniaux, M.-A. (2008). Induction and growth of hairy roots for the production of medicinal compounds. *Electronic Journal of Integrative Biosciences* 3, 2–9.

Bonfill, M., Mangas, S., Moyano, E., Cusido, R. M., and Palazón, J. (2011). Production of centellosides and phytosterols in cell suspension cultures of *Centella asiatica*. *Plant Cell Tissue Organ Cult* 104, 61–67. doi: 10.1007/S11240-010-9804-7.

Bonhomme, V., Laurain-Mattar, D., and Fliniaux, M. A. (2000). Effects of the *rol C* gene on hairy root: induction development and tropane alkaloid production by *Atropa belladonna*. *J Nat Prod* 63, 1249–1252. doi: 10.1021/NP990614L.

Boutté, Y., and Grebe, M. (2009). Cellular processes relying on sterol function in plants. *Curr Opin Plant Biol* 12, 705–713. doi: 10.1016/J.PBI.2009.09.013.

Britton, M. T., Escobar, M. A., and Dandekar, A. M. (2008). The Oncogenes of *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *Agrobacterium: From Biology to Biotechnology*, 523–563. doi: 10.1007/978-0-387-72290-0_14.

Broun, P. (2004). Transcription factors as tools for metabolic engineering in plants. *Curr Opin Plant Biol* 7, 202–209, doi:10.1016/j.pbi.2004.01.013.

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.

Bulgakov, V. P., Khodakovskaya, M. V., Labetskaya, N. V., Chernoded, G. K., and Zhuravlev, Y. N. (1998). The impact of plant *rolC* oncogene on ginsenoside production by ginseng hairy root cultures. *Phytochemistry* 49, 1929–1934. doi: 10.1016/S0031-9422(98)00351-3.

Bulgakov, V. P., Tchernoded, G. K., Mischenko, N. P., Shkryl, Y. N., Glazunov, V. P., Fedoreyev, S. A., et al. (2003). Effects of Ca2+ channel blockers and protein kinase/phosphatase inhibitors on growth and anthraquinone production in *Rubia cordifolia* callus cultures transformed by the *rolB* and *rolC* genes. *Planta* 217, 349–355. doi: 10.1007/S00425-003-0996-5.

Bylka, W., Znajdek-Awiżeń, P., Studzińska-Sroka, E., and Brzezińska, M. (2013). Review paper *Centella asiatica* in cosmetology. *Advances in Dermatology and Allergology/Postępy Dermatologii i Alergologii* 30, 46–49. doi: 10.5114/PDIA.2013.33378.

Camilleri, C.; Jouanin, L. (1991) The TR-DNA region carrying the auxin synthesis genes of the *Agrobacterium rhizogenes* agropine-type plasmid PRiA4: nucleotide sequence analysis and introduction into tobacco plants. *Mol Plant Microbe Interact* 4, 155–162, doi:10.1094/MPMI-4-155.

Capell, T., and Christou, P. (2004). Progress in plant metabolic engineering. *Curr Opin Biotechnol* 15, 148–154. doi: 10.1016/J.COPBIO.2004.01.009.

Cardinale, F., Jonak, C., Ligterink, W., Niehaus, K., Boller, T., and Hirt, H. (2000). Differential activation of four specific MAPK pathways by distinct elicitors. *J Biol Chem* 275, 36734–36740. doi: 10.1074/JBC.M007418200.

Chaisawang, P., Sirichoat, A., Chaijaroonkhanarak, W., Pannangrong, W., Sripanidkulchai, B., Wigmore, P., et al. (2017). Asiatic acid protects against cognitive deficits and reductions in cell proliferation and survival in the rat hippocampus caused by 5-fluorouracil chemotherapy. *PLoS One* 12, e0180650. doi: 10.1371/JOURNAL.PONE.0180650.

Chandrika, U. G., and Prasad Kumara, P. A. A. S. (2015). Gotu Kola (Centella asiatica): Nutritional Properties and Plausible Health Benefits. *Adv Food Nutr Res* 76, 125–157. doi: 10.1016/BS.AFNR.2015.08.001.

Chen, C., Hou, J., Tanner, J. J., and Cheng, J. (2020). Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis. *International Journal of Molecular Sciences* 2020, Vol. 21, Page 2873 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, 820582. doi: 10.3389/FPLS.2021.820582.

Chriqui, D., Guivarc'h, A., Dewitte, W., Prinsen, E., and Van Onkelen, H. (1996). Rol genes and root initiation and development. *Plant Soil* 187, 47–55. doi: 10.1007/BF00011656.

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

Cusido, R. M., Onrubia, M., Sabater-Jara, A. B., Moyano, E., Bonfill, M., Goossens, A., Angeles, M., Palazon, J. (2014). A rational approach to improving the biotechnological production of taxanes in plant cell cultures of *Taxus* spp. *Biotechnology Advances* 32:6, 1157-1167. doi: biotechadv.2014.03.002.

de Costa, F., Barber, C. J. S., Kim, Y. bok, Reed, D. W., Zhang, H., Fett-Neto, A. G., et al. (2017). Molecular cloning of an ester-forming triterpenoid: UDP-glucose 28-O-glucosyltransferase involved in saponin biosynthesis from the medicinal plant Centella asiatica. *Plant Science* 262, 9–17. doi: 10.1016/J.PLANTSCI.2017.05.009.

Dessaux, Y., and Petit, A. (1994). Opines as screenable markers for plant transformation. *Plant Mol. Biol. Man.*, 171–182. doi: 10.1007/978-94-011-0511-8_11.

Dilshad, E., Cusido, R. M., Estrada, K. R., Bonfill, M., and Mirza, B. (2015). Genetic Transformation of *Artemisia carvifolia* Buch with rol Genes Enhances Artemisinin Accumulation. *PLoS One* 10, e0140266. doi: 10.1371/JOURNAL.PONE.0140266.

Dilshad, E., Noor, H., Nosheen, N., Gilani, S. R., Ali, U., and Khan, M. A. (2021). Influence of rol Genes for Enhanced Biosynthesis of Potent Natural Products. *Chemistry of Biologically Potent Natural Products and Synthetic Compounds*, 379–404. doi: 10.1002/9781119640929.CH13.

Dilshad, E., Zafar, S., Ismail, H., Waheed, M. T., Cusido, R. M., Palazon, J., et al. (2016). Effect of Rol Genes on Polyphenols Biosynthesis in *Artemisia annua* and Their Effect on Antioxidant and Cytotoxic Potential of the Plant. *Appl Biochem Biotechnol* 179, 1456–1468. doi: 10.1007/S12010-016-2077-9.

Dirk, A., Van Dijk, J., Kootstra, G., Kruijer, W., and De Ridder, D. (2021). iScience Machine learning in plant science and plant breeding. *iScience* 24, 101890. doi: 10.1016/j.isci.

Dong, X., Lin, L., Zhang, R., Zhao, Y., Christiani, D. C., Wei, Y., et al. (2019). TOBMI: trans-omics block missing data imputation using a k-nearest neighbor weighted approach. *Bioinformatics* 35, 1278–1283. doi: 10.1093/BIOINFORMATICS/BTY796.

Draper, D. (2012). Bayesian statistics. *Computational Complexity: Theory, Techniques, and Applications* 9781461418009, 254–274. doi: 10.1007/978-1-4614-1800-9_17.

Dubrovina, A. S., Manyakhin, A. Y., Zhuravlev, Y. N., and Kiselev, K. V. (2010). Resveratrol content and expression of phenylalanine ammonia-lyase and stilbene

synthase genes in *rolC* transgenic cell cultures of *Vitis amurensis*. *Appl Microbiol Biotechnol* 88, 727–736. doi: 10.1007/S00253-010-2792-Z.

Duca, D. R., and Glick, B. R. (2020). Indole-3-acetic acid biosynthesis and its regulation in plant-associated bacteria. *Applied Microbiology and Biotechnology 2020 104:20* 104, 8607–8619. doi: 10.1007/S00253-020-10869-5.

Ebel, J., and Mithöfer, A. (1998). Early events in the elicitation of plant defense. *Planta* 206, 335–348. doi: 10.1007/S004250050409.

Emran, T. Bin, Dutta, M., Muhammad, M., Uddin, N., Kumar, A., and Bgctub, N. (2016). Antidiabetic potential of the leaf extract of *Centella asiatica* in alloxan-induced diabetic rats Risk factors of cancer in the Bangladeshi population View project COVID-19-related projects View project. *Article in Jahangirnagar University Journal of Biological Sciences*. doi: 10.3329/jujbs.v4i1.27785.

Estruch, J. J., Chriqui, H., Grossmann, K., Schell, J., and Spena, A. (1991a). The plant oncogene *rolC* is responsible for the release of cytokinins from glucoside conjugates. *EMBO J* 10, 2889–2896. doi: 10.1002/J.1460-2075.1991.TB07838.X.

Estruch, J. J., Schell, J., and Spena, A. (1991b). The protein encoded by the rolB plant oncogene hydrolyses indole glucosides. *EMBO J* 10, 3125–3128. doi: 10.1002/J.1460-2075.1991.TB04873.X.

Foster, K. R., Koprowski, R., and Skufca, J. D. (2014). Machine learning, medical diagnosis, and biomedical engineering research - commentary. *Biomed Eng Online* 13, 1–9. doi: 10.1186/1475-925X-13-94.

Fujita, K., and Inui, H. (2021). Review: Biological functions of major latex-like proteins in plants. *Plant Sci* 306. 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.

Gandi, S., and Giri, A. (2012). Genetic transformation of *Centella asiatica* by Agrobacterium rhizogenes. *Journal of Pharmacognosy* 3, 82–84.

Gandhi, S. G., Mahajan, V., Bedi, Y. S. (2015). Changing trends in biotechnology of secondary metabolism in medicinal and aromatic plants. *Planta*. 241:2, 303-317. doi:10.1007/s00425-014-2232-x.

Ganie, I. B., Ahmad, Z., Shahzad, A., Zaushintsena, A., Neverova, O., Ivanova, S., et al. (2022). Biotechnological Intervention and Secondary Metabolite Production in *Centella asiatica* L. *Plants* 11, 2928. doi: 10.3390/PLANTS11212928.

Gantait, S., and Mukherjee, E. (2021). Hairy root culture technology: applications, constraints and prospect. *Appl Microbiol Biotechnol* 105, 35–53. doi: 10.1007/S00253-020-11017-9.

Gaume, A., Komarnytsky, S., Borisjuk, N., and Raskin, I. (2003). Rhizosecretion of recombinant proteins from plant hairy roots. *Plant Cell Rep* 21, 1188–1193. doi: 10.1007/S00299-003-0660-3.

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.

Gerszberg, A., and Wiktorek-Smagur, A. (2022). Hairy root cultures as a multitask platform for green biotechnology. *Plant Cell, Tissue and Organ Culture (PCTOC) 2022 150:3* 150, 493–509. doi: 10.1007/S11240-022-02316-2.

Ghag, S. B., Adki, V. S., Ganapathi, T. R., and Bapat, V. A. (2021). Plant Platforms for Efficient Heterologous Protein Production. *Biotechnology and Bioprocess Engineering 2021 26:4* 26, 546–567. doi: 10.1007/S12257-020-0374-1.

Gómez-Gómez, L., and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Mol Cell* 5, 1003–1011. doi: 10.1016/S1097-2765(00)80265-8.

Goswami, M., Akhtar, S., and Osama, K. (2018). Strategies for monitoring and modeling the growth of hairy root cultures: An in silico perspective. *Hairy Roots: An Effective Tool of Plant Biotechnology*, 311–327. doi: 10.1007/978-981-13-2562-5_14.

Gray, N. E., Zweig, J. A., Murchison, C., Caruso, M., Matthews, D. G., Kawamoto, C., et al. (2017). Centella asiatica attenuates Aβ-induced neurodegenerative spine loss and dendritic simplification. *Neurosci Lett* 646, 24–29. doi: 10.1016/J.NEULET.2017.02.072.

Grieneisen, V. A., Xu, J., Marée, A. F. M., Hogeweg, P., and Scheres, B. (2007). Auxin transport is sufficient to generate a maximum and gradient guiding root growth. *Nature* 449, 1008–1013. doi: 10.1038/nature06215.

Grishchenko, O. V., Kiselev, K. V., Tchernoded, G. K., Fedoreyev, S. A., Veselova, M. V., Bulgakov, V. P., et al. (2016). RolB gene-induced production of isoflavonoids in transformed *Maackia amurensis* cells. *Appl Microbiol Biotechnol* 100, 7479–7489. doi: 10.1007/S00253-016-7483-Y.

Guerriero, G., Berni, R., Muñoz-Sanchez, J. A., Apone, F., Abdel-Salam, E. M., Qahtan, A. A., et al. (2018). Production of Plant Secondary Metabolites: Examples, Tips and Suggestions for Biotechnologists. *Genes (Basel)* 9. doi: 10.3390/GENES9060309.

Guo, A. Y., Chen, X., Gao, G., Zhang, H., Zhu, Q. H., Liu, X. C., et al. (2008). PlantTFDB: a comprehensive plant transcription factor database. *Nucleic Acids Res* 36, D966–D969. doi: 10.1093/NAR/GKM841.

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, 504243. doi: 10.3389/FPLS.2020.00033.

Halder, M., Sarkar, S., and Jha, S. (2019). Elicitation: A biotechnological tool for enhanced production of secondary metabolites in hairy root cultures. *Eng Life Sci* 19, 880–895. doi: 10.1002/ELSC.201900058.

Haleagrahara, N., and Ponnusamy, K. (2010). Neuroprotective effect of Centella asiatica extract (CAE) on experimentally induced parkinsonism in aged Sprague-Dawley rats. *J Toxicol Sci* 35, 41–47. doi: 10.2131/JTS.35.41.

Han, X., Zhao, J., Chang, X., Li, Q., Deng, Z., and Yu, Y. (2022). Revisiting the transcriptome data of *Centella asiatica* identified an ester-forming triterpenoid:

UDP-glucose 28-O-glucosyltransferase. *Tetrahedron* 129, 133136. doi: 10.1016/J.TET.2022.133136.

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.

Harmon, A. C., Gribskov, M., and Harper, J. F. (2000). CDPKs - a kinase for every Ca2+ signal? *Trends Plant Sci* 5, 154–159. doi: 10.1016/S1360-1385(00)01577-6.

Harris, J. M. (2015). Abscisic Acid: Hidden Architect of Root System Structure. *Plants* 4, 548–572. doi: 10.3390/PLANTS4030548.

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, 1009395. doi: 10.3389/FPLS.2022.1009395.

Hayat, Q., Hayat, S., Irfan, M., Ahmad, A. (2010). Effect of exogenous salicylic acid under changing environment: a review. *Environmental Experiment*. Bot. 68(1), 14-25. doi:10.1016/j.envexpbot.2009.08.005.

Helmy, M., Smith, D., and Selvarajoo, K. (2020). Systems biology approaches integrated with artificial intelligence for optimized metabolic engineering. *Metab Eng Commun* 11. doi: 10.1016/J.MEC.2020.E00149.

Henry, L. K., Thomas, S. T., Widhalm, J. R., Lynch, J. H., Davis, T. C., Kessler, S. A., et al. (2018). Contribution of isopentenyl phosphate to plant terpenoid metabolism. *Nat Plants* 4, 721–729. doi: 10.1038/s41477-018-0220-z.

Hesami, M., and Jones, A. M. P. (2020). Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture. *Applied Microbiology and Biotechnology 2020 104:22* 104, 9449–9485. doi: 10.1007/S00253-020-10888-2.

Hesami, M., and Jones, A. M. P. (2021). Modeling and optimizing callus growth and development in Cannabis sativa using random forest and support vector machine in combination with a genetic algorithm. *Appl Microbiol Biotechnol* 105, 5201–5212. doi: 10.1007/S00253-021-11375-Y.

Hey, S. J., Powers, S. J., Beale, M. H., Hawkins, N. D., Ward, J. L., and Halford, N. G. (2006). Enhanced seed phytosterol accumulation through expression of a modified HMG-CoA reductase. *Plant Biotechnol J* 4, 219–229. doi: 10.1111/J.1467-7652.2005.00174.X.

Hidalgo, D., Sanchez, R., Lalaleo, L., Bonfill, M., Corchete, P., and Palazon, J. (2018). Biotechnological Production of Pharmaceuticals and Biopharmaceuticals in Plant Cell and Organ Cultures. *Curr Med Chem* 25, 3577–3596. doi: 10.2174/0929867325666180309124317.

Hidalgo, D., Steinmetz, V., Brossat, M., Tournier-Couturier, L., Cusido, R. M., Corchete, P., et al. (2017). An optimized biotechnological system for the production of centellosides based on elicitation and bioconversion of *Centella asiatica* cell cultures. *Eng Life Sci* 17, 413–419. doi: 10.1002/ELSC.201600167.

Hussain, M. J., Abbas, Y., Nazli, N., Fatima, S., Drouet, S., Hano, C., et al. (2022). Root Cultures, a Boon for the Production of Valuable Compounds: A Comparative Review. *Plants* 11. doi: 10.3390/PLANTS11030439.

Hussin, M., Abdul-Hamid, A., Mohamad, S., Saari, N., Ismail, M., and Bejo, M. H. (2007). Protective effect of Centella asiatica extract and powder on oxidative stress in rats. *Food Chem* 100, 535–541. doi: 10.1016/J.FOODCHEM.2005.10.022.

Ismail, H., Dilshad, E., Waheed, M. T., Sajid, M., Kayani, W. K., and Mirza, B. (2016). Transformation of *Lactuca sativa* L. with *rol C* gene results in increased antioxidant potential and enhanced analgesic, anti-inflammatory and antidepressant activities in vivo. *3 Biotech* 6, 1–11. doi: 10.1007/S13205-016-0533-4.

James, J. T., Tugizimana, F., Steenkamp, P. A., and Dubery, I. A. (2013). Metabolomic Analysis of Methyl Jasmonate-Induced Triterpenoid Production in the Medicinal Herb *Centella asiatica* (L.) Urban. *Molecules 2013, Vol. 18, Pages 4267-4281* 18, 4267–4281. doi: 10.3390/MOLECULES18044267.

Jiang, T., Gradus, J. L., and Rosellini, A. J. (2020). Supervised Machine Learning: A Brief Primer. *Behav Ther* 51, 675–687. doi: 10.1016/J.BETH.2020.05.002.

Kai, G., Xu, H., Zhou, C., Liao, P., Xiao, J., Luo, X., et al. (2011). Metabolic engineering tanshinone biosynthetic pathway in *Salvia miltiorrhiza* hairy root cultures. *Metab Eng* 13, 319–327. doi: 10.1016/J.YMBEN.2011.02.003.

Kalita, R.; Patar, L.; Shasany, A.K.; Modi, M.K.; Sen, P. (2015). Molecular cloning, characterization and expression analysis of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase Gene from *Centella asiatica* L. *Mol Biol Rep* 42, 1431–1439, doi:10.1007/S11033-015-3922-6.

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.

Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., and Tekinerdogan, B. (2019). Analysis of transfer learning for deep neural network based plant classification models. *Comput Electron Agric* 158, 20–29. doi: 10.1016/J.COMPAG.2019.01.041.

Khan, A. N., and Dilshad, E. (2023). Enhanced Antioxidant and Anticancer Potential of *Artemisia carvifolia* Buch Transformed with *rol A* Gene. *Metabolites* 13, 351. doi: 10.3390/METABO13030351.

Kiani, B. H., Suberu, J., and Mirza, B. (2016). Cellular engineering of *Artemisia annua* and *Artemisia dubia* with the rol ABC genes for enhanced production of potent anti-malarial drug artemisinin. *Malar J* 15, 1–17. doi: 10.1186/S12936-016-1312-8.

Kiani, B. H., Ullah, N., Haq, I. ul, and Mirza, B. (2019). Transgenic *Artemisia dubia* WALL showed altered phytochemistry and pharmacology. *Arabian Journal of Chemistry* 12, 2644–2654. doi: 10.1016/J.ARABJC.2015.04.020.

Kim, G. B., Kim, W. J., Kim, H. U., and Lee, S. Y. (2020). Machine learning applications in systems metabolic engineering. *Curr Opin Biotechnol* 64, 1–9. doi: 10.1016/J.COPBIO.2019.08.010.

Kim, O. T., Bang, K. H., Shin, Y. S., Lee, M. J., Jung, S. J., Hyun, D. Y., et al. (2007). Enhanced production of asiaticoside from hairy root cultures of *Centella asiatica* (L.) Urban elicited by methyl jasmonate. *Plant Cell Rep* 26, 1941–1949. doi: 10.1007/S00299-007-0400-1.

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. doi: 10.3390/IJMS18122630.

Kim, O. T., Kim, S. H., Ohyama, K., Muranaka, T., Choi, Y. E., Lee, H. Y., et al. (2010). Upregulation of phytosterol and triterpene biosynthesis in *Centella asiatica* hairy roots overexpressed ginseng farnesyl diphosphate synthase. *Plant Cell Rep* 29, 403–411. doi: 10.1007/S00299-010-0831-Y.

Kim, O. T., Seong, N. S., Kim, M. Y., and Hwang, B. (2005). Isolation and characterization of squalene synthase cDNA from *Centella asiatica* (L.) Urban. *Journal of Plant Biology* 48, 263–269. doi: 10.1007/BF03030521.

Kim, Y. S., Cho, J. H., Park, S., Han, J. Y., Back, K., and Choi, Y. E. (2011). Gene regulation patterns in triterpene biosynthetic pathway driven by overexpression of squalene synthase and methyl jasmonate elicitation in *Bupleurum falcatum*. *Planta* 233, 343–355. doi: 10.1007/S00425-010-1292-9.

Kiselev, K. V., Dubrovina, A. S., Veselova, M. V., Bulgakov, V. P., Fedoreyev, S. A., and Zhuravlev, Y. N. (2007). The *rolB* gene-induced overproduction of resveratrol in *Vitis amurensis* transformed cells. *J Biotechnol* 128, 681–692. doi: 10.1016/J.JBIOTEC.2006.11.008.

Kishor, P. B. K., Hima Kumari, P., Sunita, M. S. L., and Sreenivasulu, N. (2015). Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. *Front Plant Sci* 6, 1–17. doi: 10.3389/FPLS.2015.00544.

Komarovská, H., Košnuth, J., Čnellárová, E., and Giovannini, A. (2009). *Agrobacterium rhizogenes*-Mediated Transformation of *Hypericum tomentosum* L. and *Hypericum tetrapterum* Fries. Zeitschrift fur Naturforsch. - Sect. C J. Biosci. 64, 864–868. doi: 10.1515/ZNC-2009-11-1218.

Kotsiantis, S. B. (2013). Decision trees: A recent overview. *Artif Intell Rev* 39, 261–283. doi: 10.1007/S10462-011-9272-4.

Krishnan, M. L., Roy, A., and Bharadvaja, N. (2019). Elicitation effect on the production of asiaticoside and asiatic acid in shoot, callus, and cell suspension culture of *Centella asiatica*. *J Appl Pharm Sci* 9, 067–074. doi: 10.7324/JAPS.2019.90609.

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.

Kunjumon, R., Johnson, A. J., and Baby, S. (2022). *Centella asiatica*: Secondary metabolites, biological activities and biomass sources. *Phytomedicine Plus* 2, 100176. doi: 10.1016/J.PHYPLU.2021.100176.

Kwon, K. J., Bae, S., Kim, K., An, I. S., Ahn, K. J., An, S., et al. (2014). Asiaticoside, a component of *Centella asiatica*, inhibits melanogenesis in B16F10 mouse melanoma. *Mol Med Rep* 10, 503–507. doi: 10.3892/MMR.2014.2159.

Lee, M. H., Jeong, J. H., Seo, J. W., Shin, C. G., Kim, Y. S., In, J. G., et al. (2004). Enhanced Triterpene and Phytosterol Biosynthesis in *Panax ginseng* Overexpressing Squalene Synthase Gene. *Plant Cell Physiol* 45, 976–984. doi: 10.1093/PCP/PCH126.

Lenk, F., Sürmann, A., Oberthür, P., Schneider, M., Steingroewer, J., and Bley, T. (2014). Modeling hairy root tissue growth in in vitro environments using an agent-based, structured growth model. *Bioprocess Biosyst Eng* 37, 1173–1184. doi: 10.1007/S00449-013-1088-Y.

Liang, Y., Zhao, S., and Zhang, X. (2009). Antisense suppression of cycloartenol synthase results in elevated ginsenoside levels in *Panax ginseng* hairy roots. *Plant Mol Biol Report* 27, 298–304. doi: 10.1007/S11105-008-0087-7.

Liu, L., White, M. J., and MacRae, T. H. (1999). Transcription factors and their genes in higher plants. *Eur J Biochem* 262, 247–257. doi: 10.1046/J.1432-1327.1999.00349.X.

Livak, K. J., Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. Methods. 25(4), 402-408. doi: 10.1006/meth.2001.1262.

Loc, N. H., and Giang, N. T. (2012). Effects of elicitors on the enhancement of asiaticoside biosynthesis in cell cultures of centella (*Centella asiatica* L. Urban). *Chemical Papers* 66, 642–648. doi: 10.2478/S11696-012-0168-9.

Loc, N. H., Giang, N. T., Huy, N. D., and Lan, T. T. P. (2017). Accumulation of madecassoside - a major component of centelloside - in centella (*Centella asiatica* (L.) Urban) cells elicited by salicylic acid. *Period Biol* 119, 81–85. doi: 10.18054/PB.V119I1.4386.

Loh, W.-Y. (2008). Classification and Regression Tree Methods. *Encyclopedia of Statistics in Quality and Reliability*, 315–323.

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.

Luthra, R., Satija, G., and Roy, A. (2022). Centellosides: pharmaceutical applications and production enhancement strategies. *Plant Cell Tissue Organ Cult* 149, 25–39. doi: 10.1007/S11240-021-02187-Z.

Lv, J., Sharma, A., Zhang, T., Wu, Y., and Ding, X. (2018). Pharmacological Review on Asiatic Acid and Its Derivatives: A Potential Compound. *SLAS Technol* 23, 111–127. doi: 10.1177/2472630317751840.

Ma, C., Zhang, H. H., and Wang, X. (2014). Machine learning for Big Data analytics in plants. *Trends Plant Sci* 19, 798–808. doi: 10.1016/j.tplants.2014.08.004.

Mandal, S., Das, T., Nandy, S., Ghorai, M., Saha, S. C., Gopalakrishnan, A. V., et al. (2023). Biotechnological and endophytic-mediated production of centellosides in *Centella asiatica*. *Appl Microbiol Biotechnol* 107, 473–489. doi: 10.1007/S00253-022-12316-Z.

Mangas, S., Bonfill, M., Osuna, L., Moyano, E., Tortoriello, J., Cusido, R. M., et al. (2006). The effect of methyl jasmonate on triterpene and sterol metabolisms of *Centella asiatica*, *Ruscus aculeatus* and *Galphimia glauca* cultured plants. *Phytochemistry* 67, 2041–2049. doi: 10.1016/J.PHYTOCHEM.2006.06.025.

Mangas, S., Moyano, E., Osuna, L., Cusido, R. M., Bonfill, M., and Palazón, J. (2008). Triterpenoid saponin content and the expression level of some related genes in calli of *Centella asiatica*. *Biotechnol Lett* 30, 1853–1859. doi: 10.1007/S10529-008-9766-6.

Marcuzzo, M., Quelhas, P., Campilho, A., Maria Mendonça, A., and Campilho, A. (2009). Automated *Arabidopsis* plant root cell segmentation based on SVM classification and region merging. *Comput Biol Med* 39, 785–793. doi: 10.1016/J.COMPBIOMED.2009.06.008.

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. *The Plant Journal* 11, 465–473. doi: 10.1046/J.1365-313X.1997.11030465.X.

Mathur, A., Gangwar, A., Mathur, A. K., Verma, P., Uniyal, G. C., and Lal, R. K. (2010). Growth kinetics and ginsenosides production in transformed hairy roots of American ginseng-*Panax quinquefolium* L. *Biotechnol Lett* 32, 457–461. doi: 10.1007/S10529-009-0158-3.

Maulud, D. H., and Mohsin Abdulazeez, A. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends* 1, 140–147. doi: 10.38094/jastt1457.

Mauro, M. L., Costantino, P., and Bettini, P. P. (2017). The never ending story of rol genes: a century after. Plant Cell, Tissue Organ Cult. 2017 1312 131, 201–212. doi: 10.1007/S11240-017-1277-5.

Mayoraz, E., and Alpaydm, E. (1999). Support vector machines for multi-class classification. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 1607, 833–842. doi: 10.1007/BFB0100551.

Medina-Bolivar, F., Condori, J., Rimando, A. M., Hubstenberger, J., Shelton, K., O'Keefe, S. F., et al. (2007). Production and secretion of resveratrol in hairy root cultures of peanut. *Phytochemistry* 68, 1992–2003. doi: 10.1016/J.PHYTOCHEM.2007.04.039.

Mehrotra, S., Prakash, O., Mishra, B. N., and Dwevedi, B. (2008). Efficiency of neural networks for prediction of in vitro culture conditions and inoculum properties for optimum productivity. *Plant Cell Tissue Organ Cult* 95, 29–35. doi: 10.1007/S11240-008-9410-0.

Mehta, V., Sharma, A., Tanwar, S., and Malairaman, U. (2016). *In vitro* and *In silico* evaluation of the antidiabetic effect of hydroalcoholic leaf extrac of *Centella asiatica*. *Int J Pharm Pharm Sci*.

Menzel, G., Harloff, H. J., and Jung, C. (2003). Expression of bacterial poly(3-hydroxybutyrate) synthesis genes in hairy roots of sugar beet (*Beta vulgaris* L.). *Appl Microbiol Biotechnol* 60, 571–576. doi: 10.1007/S00253-002-1152-Z.

Mercy, S., Sangeetha, N., and Ganesh, D. (2012). In vitro production of adventitious roots containing asiaticoside from leaf tissues of *Centella asiatica* L. *In Vitro Cellular and Developmental Biology - Plant* 48, 200–207. doi: 10.1007/S11627-011-9416-X.

Mertens, J., Pollier, J., Bossche, R. Vanden, Lopez-Vidriero, I., Franco-Zorrilla, J. M., and Goossens, A. (2015). The bHLH Transcription Factors *TSAR1* and *TSAR2* Regulate Triterpene Saponin Biosynthesis in *Medicago truncatula*. *Plant Physiol* 170, 194–210. doi: 10.1104/PP.15.01645.

Miettinen, K.; Pollier, J.; Buyst, D.; Arendt, P.; Csuk, R.; Sommerwerk, S.; Moses, T.; Mertens, J.; Sonawane, P.D.; Pauwels, L.; et al. (2017). The Ancient *CYP716* family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. *Nature Communications* 8:1, 1–13, doi:10.1038/ncomms14153.

Mirjalili, M.H.; Moyano, E.; Bonfill, M.; Cusido, R.M.; Palazón, J. (2011). Overexpression of the *Arabidopsis thaliana* squalene synthase gene in *Withania coagulans* hairy root cultures. *Biol Plant* 55, 357–360, doi:10.1007/S10535-011-0054-2.

Mishra, A. K., Sharma, K., and Misra, R. S. (2012). Elicitor recognition, signal transduction and induced resistance in plants. *J Plant Interact* 7, 95–120. doi: 10.1080/17429145.2011.597517.

Misra, B. B., Langefeld, C., Olivier, M., and Cox, L. A. (2019). Integrated omics: tools, advances and future approaches. *J Mol Endocrinol* 62, R21–R45. doi: 10.1530/JME-18-0055.

Mitra, S. S., Mukherjee, A., Nongdam, P., Pandey, D. K., and Dey, A. (2021). Endophytes producing active constituents in *Centella asiatica* with a special emphasis on asiaticoside and madecassoside: a review update. *Volatiles and Metabolites of Microbes*, 409–427. doi: 10.1016/B978-0-12-824523-1.00013-4.

Mohd Nadzir, M., Jaromír Klemeš, J., Yen Liew, P., Shin Ho, W., Shiun Lim, J., and Nazira Idris, F. (2017). Antimicrobial Activity of *Centella Asiatica* on *Aspergillus Niger* and *Bacillus Subtilis*. *Chem Eng Trans* 56. doi: 10.3303/CET1756231.

Montesano, M., Brader, G., and Palva, E. T. (2003). Pathogen derived elicitors: searching for receptors in plants. *Mol Plant Pathol* 4, 73–79. doi: 10.1046/J.1364-3703.2003.00150.X.

Morey, K. J., and Peebles, C. A. M. (2022). Hairy roots: An untapped potential for production of plant products. *Front Plant Sci* 13, 937095. doi: 10.3389/FPLS.2022.937095.

Moshou, D., Bravo, C., Wahlen, S., West, J., McCartney, A., De Baerdemaeker, J., et al. (2006). Simultaneous identification of plant stresses and diseases in arable crops using proximal optical sensing and self-organising maps. *Precis Agric* 7, 149–164. doi: 10.1007/S11119-006-9002-0.

Moyano, E., Jouhikainen, K., Tammela, P., Palazón, J., Cusidó, R. M., Piñol, M. T., et al. (2003). Effect of pmt gene overexpression on tropane alkaloid production in

transformed root cultures of *Datura metel* and *Hyoscyamus muticus*. *J Exp Bot* 54, 203–211. doi: 10.1093/JXB/ERG014.

Müller, M., and Munné-Bosch, S. (2011). Rapid and sensitive hormonal profiling of complex plant samples by liquid chromatography coupled to electrospray ionization tandem mass spectrometry. *Plant Methods* 7, 1–11. doi: 10.1186/1746-4811-7-37.

Müller, V., Lankes, C., Zimmermann, B. F., Noga, G., and Hunsche, M. (2013). Centelloside accumulation in leaves of *Centella asiatica* is determined by resource partitioning between primary and secondary metabolism while influenced by supply levels of either nitrogen, phosphorus or potassium. *J Plant Physiol* 170, 1165–1175. doi: 10.1016/J.JPLPH.2013.03.010.

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.

Nagegowda, D. A., and Gupta, P. (2020). Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. *Plant Science* 294, 110457. doi: 10.1016/J.PLANTSCI.2020.110457.

Narayani, M., Srivastava, S. (2017). Elicitation: a stimulation of stress in *in vitro* plant cell/tissue cultures for enhancement of secondary metabolite production. Phytochemistry reviews. 16(6), 1227-1252. doi: 10.1007/s11101-017-9534-0

Nemoto, K., Hara, M., Suzuki, M., Seki, H., Oka, A., Muranaka, T., et al. (2009). Function of the *aux* and *rol* genes of the Ri plasmid in plant cell division in vitro. *Plant Signal Behav* 4. doi: 10.4161/PSB.4.12.9904.

Niazian, M., Belzile, F., and Torkamaneh, D. (2022). CRISPR/Cas9 in Planta Hairy Root Transformation: A Powerful Platform for Functional Analysis of Root Traits in Soybean. *Plants* 11, 1044. doi: 10.3390/PLANTS11081044.

Niazian, M., and Niedbała, G. (2020). Machine Learning for Plant Breeding and Biotechnology. *Agriculture 2020, Vol. 10, Page 436* 10, 436. doi: 10.3390/AGRICULTURE10100436.

Nielsen, J. (2001). Metabolic engineering. *Appl Microbiol Biotechnol* 55, 263–283. doi: 10.1007/S002530000511.

Nilsson, O., Moritz, T., Imbault, N., Sandberg, G., and Olsson, O. (1993). Hormonal Characterization of Transgenic Tobacco Plants Expressing the *rolC* Gene of *Agrobacterium rhizogenes* TL-DNA. *Plant Physiol* 102, 363–371. doi: 10.1104/PP.102.2.363.

Nouri Nav, S., Nejad Ebrahimi, S., Sonboli, A., and Mirjalili, M. H. (2021). Variability, association and path analysis of centellosides and agro-morphological characteristics in Iranian *Centella asiatica* (L.) Urban ecotypes. *South African Journal of Botany* 139, 254–266. doi: 10.1016/J.SAJB.2021.03.006.

Nordström, A., Tarkowski, P., Tarkowska, D., Norbaek, R., Åstot, C., Dolezal, K., et al. (2004). Auxin regulation of cytokinin biosynthesis in *Arabidopsis thaliana*: A factor of potential importance for auxin-cytokinin-regulated development. Proc*. Natl. Acad. Sci. U. S. A.* 101, 8039–8044. doi: 10.1073/pnas.0402504101.

Noushahi, H. A., Khan, A. H., Noushahi, U. F., Hussain, M., Javed, T., Zafar, M., et al. (2022). Biosynthetic pathways of triterpenoids and strategies to improve their Biosynthetic Efficiency. *Plant Growth Regul* 97, 439. doi: 10.1007/S10725-022-00818-9.

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, 544–544. doi: 10.3389/FPLS.2015.00544.

Ono, N. N., and Tian, L. (2011). The multiplicity of hairy root cultures: Prolific possibilities. *Plant Science* 180, 439–446. doi: 10.1016/J.PLANTSCI.2010.11.012.

Onrubia, M., Moyano, E., Bonfill, M., Cusidó, R. M., Goossens, A., and Palazón, J. (2013). Coronatine, a more powerful elicitor for inducing taxane biosynthesis in *Taxus media* cell cultures than methyl jasmonate. *J Plant Physiol* 170, 211–219. doi: 10.1016/J.JPLPH.2012.09.004.

Osama, K., Somvanshi, P., Pandey, A., and Mishra, B. (2013). Modelling of Nutrient Mist Reactor for Hairy Root Growth using Artificial Neural Network. *European Journal of Scientific Research* 97, 516–526.

Oyedeji, O. A., and Afolayan, A. J. (2008). Chemical Composition and Antibacterial Activity of the Essential Oil of *Centella asiatica*. Growing in South Africa. *Pharm Biol* 43, 249–252. doi: 10.1080/13880200590928843.

Ozyigit, I. I. ., Dogan, I. ., and Tarhan, E. A. (2008). Induction and growth of hairy roots for the production of medicinal compounds. Electron. J. Integr. Biosci. 3, 2–9.

Ozyigit, I. I., Dogan, I., and Artam Tarhan, E. (2013). *Agrobacterium rhizogenes*-mediated transformation and its biotechnological applications in crops. *Crop Improvement: New Approaches and Modern Techniques*, 1–48. doi: 10.1007/978-1-4614-7028-1_1.

Palazón, J., Cusidó, R. M., Bonfill, M., Mallol, A., Moyano, E., Morales, C., et al. (2003). Elicitation of different Panax ginseng transformed root phenotypes for an improved ginsenoside production. *Plant Physiology and Biochemistry* 41, 1019–1025. doi: 10.1016/J.PLAPHY.2003.09.002.

Palazon, J., Cusido, R. M., Gonzalo, J., Bonfill, M., Morales, C., and Pinol, M. T. (1998). Relation between the amount of *rolC* gene product and indole alkaloid accumulation in *Catharanthus roseus* transformed root cultures. *J Plant Physiol* 153, 712–718. doi: 10.1016/S0176-1617(98)80225-3.

Palazon, J., Cusido, R., Roig, C., and Pinol, M. T. (1997). Effect of *rol* genes from *Agrobacterium rhizogenes* TL-DNA on nicotine production in tobacco root cultures. *Plant Physiology and Biochemistry* 35, 155–162.

Pan, Y. Y., Wang, X., Ma, L. G., and Sun, D. Y. (2005). Characterization of phosphatidylinositol-specific phospholipase C (PI-PLC) from *Lilium daviddi* pollen. *Plant Cell Physiol* 46, 1657–1665. doi: 10.1093/PCP/PCI181.

Pandian, B. A., Sathishraj, R., Djanaguiraman, M., Prasad, P. V. V., and Jugulam, M. (2020). Role of Cytochrome P450 Enzymes in Plant Stress Response. *Antioxidants 2020, Vol. 9, Page 454* 9, 454. doi: 10.3390/ANTIOX9050454.

Pandita, D., Pandita, A., Wani, S. H., Abdelmohsen, S. A. M., Alyousef, H. A., Abdelbacki, A. M. M., et al. (2021). Crosstalk of Multi-Omics Platforms with Plants of Therapeutic Importance. *Cells* 10, 1296. doi: 10.3390/CELLS10061296.

Park, B. C., Bosire, K. O., Lee, E. S., Lee, Y. S., and Kim, J. A. (2005). Asiatic acid induces apoptosis in SK-MEL-2 human melanoma cells. *Cancer Lett* 218, 81–90. doi: 10.1016/J.CANLET.2004.06.039.

Pavlova, O. A., Matveyeva, T. V., and Lutova, L. A. (2014). rol-Genes of *Agrobacterium rhizogenes*. *Russ J Genet Appl Res* 4, 137–145. doi: 10.1134/S2079059714020063.

Peebles, C. A. M., Hughes, E. H., Shanks, J. V., and San, K. Y. (2009). Transcriptional response of the terpenoid indole alkaloid pathway to the overexpression of *ORCA3* along with jasmonic acid elicitation of *Catharanthus roseus* hairy roots over time. *Metab Eng* 11, 76–86. doi: 10.1016/J.YMBEN.2008.09.002.

Piątczak, E., Kuźma, Ł., and Wysokińska, H. (2016). The influence of methyl jasmonate and salicylic acid on secondary metabolite production in *Rehmannia glutinosa* Libosch. Hairy root culture. *Acta Biol Crac Ser Bot* 58, 57–65. doi: 10.1515/ABCSB- 2016-0004.

Pipatsitee, P., Praseartkul, P., Theerawitaya, C., Taota, K., Tisarum, R., Chungloo, D., et al. (2023). Impact of Temperature on Centelloside Content, Growth Characters, Physio-morphological Adaptations, and Biochemical Changes in Indian Pennywort (*Centella asiatica*). *J Plant Growth Regul*, 1–12. doi: 10.1007/S00344-023-11022-Y.

Pistelli, L., Giovannini, A., Ruffoni, B., Bertoli, A., and Pistelli, L. (2010). Hairy root cultures for secondary metabolites production. *Adv Exp Med Biol* 698, 167–184. doi: 10.1007/978-1-4419-7347-4_13.

Pittella, F., Dutra, R. C., Junior, D. D., Lopes, M. T. P., and Barbosa, N. R. (2009). Antioxidant and Cytotoxic Activities of *Centella asiatica* (L) Urb. *International Journal of Molecular Sciences 2009, Vol. 10, Pages 3713-3721* 10, 3713–3721. doi: 10.3390/IJMS10093713.

Prakash, O., Mehrotra, S., Krishna, A., and Mishra, B. N. (2010). A neural network approach for the prediction of in vitro culture parameters for maximum biomass yields in hairy root cultures. *J Theor Biol* 265, 579–585. doi: 10.1016/J.JTBI.2010.05.020.

Prakash, V., Jaiswal, N., and Srivastava, M. (2017). A review on medicinal properties of *Centella asiatica*. *Article in Asian Journal of Pharmaceutical and Clinical Research* 10. doi: 10.22159/ajpcr.2017.v10i10.20760.

Prasad, A., Mathur, A. K., and Mathur, A. (2019). Advances and emerging research trends for modulation of centelloside biosynthesis in *Centella asiatica* (L.) Urban- A review. *Ind Crops Prod* 141, 111768. doi: 10.1016/J.INDCROP.2019.111768.

Putalun, W., Taura, F., Qing, W., Matsushita, H., Tanaka, H., and Shoyama, Y. (2003). Anti-solasodine glycoside single-chain Fv antibody stimulates biosynthesis of solasodine glycoside in plants. *Plant Cell Rep* 22, 344–349. doi: 10.1007/S00299-003-0689-3.

Qi, X., Chen, G., Li, Y., Cheng, X., and Li, C. (2019). Applying Neural-Network-Based Machine Learning to Additive Manufacturing: Current Applications, Challenges, and Future Perspectives. *Engineering* 5, 721–729. doi: 10.1016/J.ENG.2019.04.012.

Rafamantanana, M. H., Rozet, E., Raoelison, G. E., Cheuk, K., Ratsimamanga, S. U., Hubert, P., et al. (2009). An improved HPLC-UV method for the simultaneous quantification of triterpenic glycosides and aglycones in leaves of *Centella asiatica* (L.) Urb (APIACEAE). *J. Chromatogr. B* 877, 2396–2402. doi: 10.1016/J.JCHROMB.2009.03.018.

Rahman, M. M., Shahdaat, M., Sayeed, B., Haque, A., Hassan, M., and Islam, S. M. A. (2012). Phytochemical screening, Antioxidant, Anti-Alzheimer and Anti-diabetic activities of *Centella asiatica*. *J. Nat. Prod. Plant Resour*, 504–511.

Rai, A., Saito, K., and Yamazaki, M. (2017). Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J.* 90, 764–787. doi: 10.1111/TPJ.13485.

Ramirez-Estrada, K., Osuna, L., Moyano, E., Bonfill, M., Tapia, N., Cusido, R. M., Palazon, J. (2015). Changes in gene transcription and taxane production in elicited cell cultures of *Taxus×media* and *Taxus globosa*. Phytochemistry. 117, 174-184. doi: 10.1016/j.phytochem.2015.06.013.

Ramirez-Estrada, K., Vidal-Limon, H., Hidalgo, D., Moyano, E., Golenioswki, M., Cusidó, R. M., et al. (2016). Elicitation, an Effective Strategy for the Biotechnological Production of Bioactive High-Added Value Compounds in Plant Cell Factories. *Molecules* 21, 182. doi: 10.3390/MOLECULES21020182.

Rangslang, R. K., Liu, Z., Lütken, H., and Favero, B. T. (2018). *Agrobacterium* spp. genes and ORFs: Mechanisms and applications in plant science. *Ciência e Agrotecnologia* 42, 453–463. doi: 10.1590/1413-70542018425000118.

Reel, P. S., Reel, S., Pearson, E., Trucco, E., and Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: A review. *Biotechnol Adv* 49, 107739. doi: 10.1016/J.BIOTECHADV.2021.107739.

Rency, A. S., Pandian, S., Kasinathan, R., Satish, L., Swamy, M. K., and Ramesh, M. (2019). Hairy root cultures as an alternative source for the production of high-value secondary metabolites. *Natural Bio-active Compounds: Volume 3: Biotechnology, Bioengineering, and Molecular Approaches*, 237–264. doi: 10.1007/978-981-13-7438-8_10.

Rogowska, A., and Szakiel, A. (2021). Enhancement of Phytosterol and Triterpenoid Production in Plant Hairy Root Cultures-Simultaneous Stimulation or Competition? *Plants* 10. doi: 10.3390/PLANTS10102028.

Rong, T., Chunchun, Z., Wei, G., Yuchen, G., Fei, X., Tao, L., et al. (2021). Proteomic insights into protostane triterpene biosynthesis regulatory mechanism after MeJA treatment in Alisma orientale (Sam.) Juz. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1869, 140671. doi: 10.1016/J.BBAPAP.2021.140671.

Rothblat, G.H.; Martak, D.S.; Kritchevsky, D. A quantitative colorimetric assay for squalene. (1962). *Anal Biochem* 4, 52–56, doi:10.1016/0003-2697(62)90019-2.

Roy, A. (2021). Hairy Root Culture an Alternative for Bioactive Compound Production from Medicinal Plants. *Curr Pharm Biotechnol* 22, 136–149. doi: 10.2174/1389201021666201229110625.

Roychowdhury, D., Basu, A., and Jha, S. (2015). Morphological and molecular variation in Ri-transformed root lines are stable in long term cultures of *Tylophora indica*. *Plant Growth Regul.* 75, 443–453. doi: 10.1007/s10725-014-0005-y.

Rupesh Kumar, M., and Ramesh, B. (2021). Ethnobotanical and Pharmacological Profile of *Centella asiatica*. *GIS Science Journal*..

Ruslan, K., Selfitri, A. D., Bulan, S. A., Rukayadi, Y., and Elfahmi (2012). Effect of Agrobacterium rhizogenes and elicitation on the asiaticoside production in cell cultures of *Centella asiatica*. *Pharmacogn Mag* 8, 111–115. doi: 10.4103/0973-1296.96552.

Sabaragamuwa, R.; Perera, C.O.; Fedrizzi, B. *Centella asiatica* (*Gotu Kola*) as a neuroprotectant and its potential role in healthy ageing. *Trends Food Sci Technol* 2018, 79, 88–97, doi:10.1016/J.TIFS.2018.07.024.

Sanchez-Muñoz, R.; Almagro, L.; Cusido, R.M.; Bonfill, M.; Palazon, J.; Moyano, E. (2020). Transfecting *Taxus × Media* protoplasts to study transcription factors *BIS2* and *TSAR2* as activators of taxane-related genes. *Plant Cell Physiol* 61, 576–583, doi:10.1093/PCP/PCZ225.

Sangwan, R. S., Tripathi, S., Singh, J., Narnoliya, L. K., and Sangwan, N. S. (2013). De novo sequencing and assembly of *Centella asiatica* leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism. *Gene* 525, 58–76. doi: 10.1016/J.GENE.2013.04.057.

Sasikala, S., Lakshminarasaiah, S., and Naidu, M. D. (2015). Antidiabetic activity of *Centella asiatica* on streptozotocin induced diabetic male albino rats. *World Journal of Pharmaceutical Sciences*, 1701–1705.

Sathasivam, R., Choi, M., Radhakrishnan, R., Kwon, H., Yoon, J., Yang, S. H., et al. (2022). Effects of various *Agrobacterium rhizogenes* strains on hairy root induction and analyses of primary and secondary metabolites in. *Front Plant Sci* 13, 983776. doi: 10.3389/FPLS.2022.983776.

Schwechheimer, C., and Bevan, M. (1998). The regulation of transcription factor activity in plants. *Trends Plant Sci* 3, 378–383. doi: 10.1016/S1360-1385(98)01302-8.

Sen, S., and Chakraborty, R. (2017). Revival, modernization and integration of Indian traditional herbal medicine in clinical practice: Importance, challenges and future. *J Tradit Complement Med* 7, 234–244. doi: 10.1016/J.JTCME.2016.05.006.

Sharma, A., Rather, G. A., Misra, P., Dhar, M. K., and Lattoo, S. K. (2019). Jasmonate responsive transcription factor WsMYC2 regulates the biosynthesis of triterpenoid withanolides and phytosterol via key pathway genes in *Withania somnifera* (L.) Dunal. *Plant Mol Biol* 100, 543–560. doi: 10.1007/S11103-019-00880-4.

Sharma, P., Padh, H., and Shrivastava, N. (2013). Hairy root cultures: A suitable biological system for studying secondary metabolic pathways in plants. *Eng Life Sci* 13, 62–75. doi: 10.1002/ELSC.201200030.

Shenoy, A., Dsouza, V. S., Shabaraya, R., and Prof, A. (2023). An update on the ayurvedic herb *Centella asiatica*. *Indian Journal of Pharmacy & Drugs Studies*, 98–103.

Shi, Q., Katuwal, R., Suganthan, P. N., and Tanveer, M. (2021). Random vector functional link neural network based ensemble deep learning. *Pattern Recognit* 117, 107978. doi: 10.1016/J.PATCOG.2021.107978.

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.

Shkryl, Y. N., Veremeichik, G. N., Bulgakov, V. P., Tchernoded, G. K., Mischenko, N. P., Fedoreyev, S. A., et al. (2008). Individual and combined effects of the rolA, B, and C genes on anthraquinone production in *Rubia cordifolia* transformed calli. *Biotechnol Bioeng* 100, 118–125. doi: 10.1002/BIT.21727.

Shilpha, J., Largia, M. J. V., Kumar, R. R., Satish, L., Swamy, M. K., and Ramesh, M. (2023). Hairy Root Cultures: A Novel Way to Mass Produce Plant Secondary Metabolites. *Phytochemical Genomics: Plant Metabolomics and Medicinal Plant Genomics*, 417–445. 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.

Singh, A., Dwivedi, P., and Padmanabh Dwivedi, C. (2018). Methyl-jasmonate and salicylic acid as potent elicitors for secondary metabolite production in medicinal plants: A review. *J Pharmacogn Phytochem* 7, 750–757.

Singh, A. K., Kumar, S. R., Dwivedi, V., Rai, A., Pal, S., Shasany, A. K., et al. (2017). A WRKY transcription factor from *Withania somnifera* regulates triterpenoid withanolide accumulation and biotic stress tolerance through modulation of phytosterol and defense pathways. *New Phytologist* 215, 1115–1131. doi: 10.1111/NPH.14663.

Somvanshi, P., and Mishra, B. N. (2015). Machine learning techniques in plant biology. *PlantOmics: The Omics of Plant Science*, 731–754. doi: 10.1007/978-81-322-2172-2_26.

Sonkar, N., Shukla, P. K., and Misra, P. (2023). Plant Hairy Roots as Biofactory for the Production of Industrial Metabolites. *Plants as Bioreactors: An Overview*, 273–297. doi: 10.1002/9781119875116.CH11.

Stephanopoulos, G. (1999). Metabolic Fluxes and Metabolic Engineering. *Metab Eng* 1, 1–11. doi: 10.1006/MBEN.1998.0101.

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights* 14. doi: 10.1177/1177932219899051.

Sun, B., Wu, L., Wu, Y., Zhang, C., Qin, L., Hayashi, M., et al. (2020). Therapeutic Potential of *Centella asiatica* and Its Triterpenes: A Review. *Front Pharmacol* 11, 568032. doi: 10.3389/FPHAR.2020.568032.

Sun, L. R., Wang, Y. Bin, He, S. Bin, and Hao, F. S. (2018). Mechanisms for Abscisic Acid Inhibition of Primary Root Growth. *Plant Signal. Behav.* 13, 1–4. doi: 10.1080/15592324.2018.1500069.

Takatsuka, H., and Umeda, M. (2014). Hormonal control of cell division and elongation along differentiation trajectories in roots. *J. Exp. Bot.* 65, 2633–2643. doi: 10.1093/jxb/ert485.

Thakur, M., Bhattacharya, S., Khosla, P. K., and Puri, S. (2019). Improving production of plant secondary metabolites through biotic and abiotic elicitation. *J Appl Res Med Aromat Plants* 12, 1–12. doi: 10.1016/J.JARMAP.2018.11.004.

Theboral, J., Sivanandhan, G., Subramanyam, K., Arun, M., Selvaraj, N., Manickavasagam, M., et al. (2014). Enhanced production of isoflavones by elicitation in hairy root cultures of Soybean. *Plant Cell Tissue Organ Cult* 117, 477–481. doi: 10.1007/S11240-014-0450-3.

Theerawitaya, C., Praseartkul, P., Taota, K., Tisarum, R., Samphumphuang, T., Singh, H. P., et al. (2023). Investigating high throughput phenotyping based morpho-physiological and biochemical adaptations of indian pennywort (*Centella asiatica* L. urban) in response to different irrigation regimes. *Plant Physiology and Biochemistry* 202, 107927. doi: 10.1016/J.PLAPHY.2023.107927.

Tiwari, R., and Rana, C. S. (2015). Plant secondary metabolites: a review. *International Journal of Engineering Research and General Science* 3.

Trémouillaux-Guiller, J. (2013). Hairy root culture: An alternative terpenoid expression platform. *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes*, 2941–2970. doi: 10.1007/978-3-642-22144-6_136.

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. (2018). From *A. rhizogenes rolD* to Plant P5CS: Exploiting Proline to Control Plant Development. *Plants* 7, 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.

Tsurumi, S.; Tsujino, Y. (1995). Chromosaponin I stimulates the growth of lettuce roots. *Physiol Plant* 93, 785–789, doi:10.1111/J.1399-3054.1995.TB05132.X.

Tsurumi, S.; Ishizawa, K.; Rahman, A.; Soga, K.; Hoson, T.; Goto, N.; Kamisaka, S. (2000). Effects of chromosaponin i and brassinolide on the growth of roots in etiolated *Arabidopsis* seedlings. *J Plant Physiol* 156, 60–67, doi:10.1016/S0176-1617(00)80273-4.

Unland, K.; Pütter, K.M.; Vorwerk, K.; van Deenen, N.; Twyman, R.M.; Prüfer, D.; Schulze Gronover, C. (2018). Functional characterization of squalene synthase and squalene epoxidase in *Taraxacum koksaghyz*. *Plant Direct* 2, e00063, doi:10.1002/PLD3.63.

Vaccaro, M. C., Mariaevelina, A., Malafronte, N., De Tommasi, N., and Leone, A. (2017). Increasing the synthesis of bioactive abietane diterpenes in *Salvia sclarea* hairy roots by elicited transcriptional reprogramming. *Plant Cell Rep* 36, 375–386. doi: 10.1007/S00299-016-2076-X.

Vailati-Riboni, M., Palombo, V., and Loor, J. J. (2017). What are omics sciences? *Periparturient Diseases of Dairy Cows: A Systems Biology Approach*, 1–7. doi: 10.1007/978-3-319-43033-1_1.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., et al. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers 2021 1:1* 1, 1–26. doi: 10.1038/s43586-020-00001-2.

Vasconsuelo, A., and Boland, R. (2007). Molecular aspects of the early stages of elicitation of secondary metabolites in plants. *Plant Science* 172, 861–875. doi: 10.1016/J.PLANTSCI.2007.01.006.

Venables, W. N., and Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth. *New York: Springer* Available at: http://www.stats.ox.ac.uk/pub/MASS4.

Veremeichik, G. N., Bulgakov, V. P., Shkryl, Y. N., Silantieva, S. A., Makhazen, D. S., Tchernoded, G. K., et al. (2019). Activation of anthraquinone biosynthesis in long-cultured callus culture of *Rubia cordifolia* transformed with the *rolA* plant oncogene. *J Biotechnol* 306, 38–46. doi: 10.1016/J.JBIOTEC.2019.09.007.

Vereshchagina, Y. V., Bulgakov, V. P., Grigorchuk, V. P., Rybin, V. G., Veremeichik, G. N., Tchernoded, G. K., et al. (2014). The *rolC* gene increases caffeoylquinic acid production in transformed artichoke cells. *Appl Microbiol Biotechnol* 98, 7773–7780. doi: 10.1007/S00253-014-5869-2.

Verpoorte, R., Contin, A., and Memelink, J. (2002). Biotechnology for the production of plant secondary metabolites. *Phytochemistry Reviews* 1, 13–25. doi: 10.1023/A:1015871916833.

Verpoorte, R., Van Der Hejden, R., Ten Hoopen, H. J. G., and Memelink, J. (1999). Metabolic engineering of plant secondary metabolite pathways for the production of fine chemicals. *Biotechnol Lett* 21, 467–479. doi: 10.1023/A:1005502632053.

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.

Vilaine, F., Rembur, J., Chriqui, D., and Tepfer, M. (2007). Modified Development in Transgenic Tobacco Plants Expressing a rolA::GUS Translational Fusion and Subcellular Localization of the Fusion Protein. *Mol Plant Microbe Interact* 11, 855–859. doi: 10.1094/MPMI.1998.11.9.855.

Vranová, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu Rev Plant Biol* 64, 665–700. doi: 10.1146/ANNUREV-ARPLANT-050312-120116.

Walker, T. S., Bais, H. P., Vivanco, J. M. (2002). Jasmonic acid-induced hypericin production in cell suspension cultures of *Hypericum perforatum* L.(St. John's wort). Phytochemistry. 60:3, 289-293. doi: 10.1016/S0031-9422(02)00074-2.

Wei, T., and Simko, V. (2021). R package "corrplot": Visualization of a Correlation Matrix. Available at: https://github.com/taiyun/corrplot.

Wong, J. X., and Ramli, S. (2021). Antimicrobial activity of different types of *Centella asiatica* extracts against foodborne pathogens and food spoilage microorganisms. *LWT* 142, 111026. doi: 10.1016/J.LWT.2021.111026.

Wongwicha, W., Tanaka, H., Shoyama, Y., and Putalun, W. (2011). Methyl jasmonate elicitation enhances glycyrrhizin production in *Glycyrrhiza inflata* hairy roots cultures. *Zeitschrift fur Naturforschung - Section C Journal of Biosciences* 66 C, 423–428. doi: 10.1515/ZNC-2011-7-815.

Woolston, B. M., Edgar, S., and Stephanopoulos, G. (2013). Metabolic Engineering: Past and Future. *Annu Rev Chem Biomol Eng* 4, 259–288. doi: 10.1146/ANNUREV-CHEMBIOENG-061312-103312.

Wu, T., Geng, J., Guo, W., Gao, J., and Zhu, X. (2017). Asiatic acid inhibits lung cancer cell growth in vitro and in vivo by destroying mitochondria. *Acta Pharm Sin B* 7, 65–72. doi: 10.1016/J.APSB.2016.04.003.

Xu, B., Liu, L., and Song, G. (2022). Functions and Regulation of Translation Elongation Factors. *Front Mol Biosci* 8. doi: 10.3389/FMOLB.2021.816398.

Yan, J., and Wang, X. (2022). Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology. *The Plant Journal* 111, 1527–1538. doi: 10.1111/TPJ.15905.

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.

Yang, L., Zhang, J., He, J., Qin, Y., Hua, D., Duan, Y., et al. (2014). ABA-Mediated ROS in Mitochondria Regulate Root Meristem Activity by Controlling PLETHORA Expression in *Arabidopsis. PLoS Genet.* 10. doi: 10.1371/journal.pgen.1004791.

Yoo, N. H., Kim, O. T., Kim, J. B., Kim, S. H., Kim, Y. C., Bang, K. H., et al. (2011). Enhancement of centelloside production from cultured plants of *Centella asiatica* by combination of thidiazuron and methyl jasmonate. *Plant Biotechnol Rep* 5, 283–287. doi: 10.1007/S11816-011-0173-4.

Yoshikawa, M., Yamaoka, N., and Takeuchi, Y. (1993). Elicitors: Their Significance and Primary Modes of Action in the Induction of Plant Defense Reactions. *Plant Cell Physiol* 34, 1163–1173. doi: 10.1093/OXFORDJOURNALS.PCP.A078537.

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.

Yu, C., Sun, C., Shen, C., Wang, S., Liu, F., Liu, Y., et al. (2015). The auxin transporter, OsAUX1, is involved in primary root and root hair elongation and in Cd stress responses in rice (*Oryza sativa* L.). *The Plant Journal* 83, 818–830. doi: 10.1111/TPJ.12929.

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. doi: 10.1016/J.GENE.2019.144275.

Yusuf, D., Butland, S. L., Swanson, M. I., Bolotin, E., Ticoll, A., Cheung, W. A., et al. (2012). The transcription factor encyclopedia. *Genome Biol* 13, 1–25. doi: 10.1186/GB-2012-13-3-R24.

Zahanis, Z., Mansyurdin, M., Noli, Z. A., and Bachtiar, A. (2014). Induction of Hairy roots of Pegagan (*Centella Asiatica* (l.) Urban using Several Explant Sources with Several *Agrobacterium rhizogenes* Strains in Vitro. *Int J Adv Sci Eng Inf Technol* 4, 286–289. doi: 10.18517/IJASEIT.4.4.418.

Zhang, J., Ai, L., Lv, T., Jiang, X., and Liu, F. (2013). Asiatic acid, a triterpene, inhibits cell proliferation through regulating the expression of focal adhesion kinase in multiple myeloma cells. *Oncol Lett* 6, 1762–1766. doi: 10.3892/OL.2013.1597.

Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Ann Transl Med* 4. doi: 10.21037/ATM.2016.03.37.

Zhao, H., Lai, Z., Leung, H., and Zhang, X. (2020). "Linear Discriminant Analysis," in Feature Learning and Understanding: Algorithms and Applications (*Cham: Springer International Publishing*), 71–85. doi: 10.1007/978-3-030-40794-0_5.

Zhao, L., Sander, G. W., and Shanks, J. V. (2013). Perspectives of the Metabolic Engineering of Terpenoid Indole Alkaloids in *Catharanthus roseus* Hairy Roots. *Adv Biochem Eng Biotechnol* 134, 23–54. doi: 10.1007/10_2013_182.

Zhu, C., Miao, G., Guo, J., Huo, Y., Zhang, X., Xie, J., et al. (2014). Establishment of *Tripterygium wilfordii* Hook. f. Hairy root culture and optimization of its culture conditions for the production of triptolide and wilforine. *J Microbiol Biotechnol* 24, 823–834. doi: 10.4014/JMB.1402.02045.

Annexes

Annex 1.

Supplementary material

for Chapter 1

Supplementary material

Using machine learning to link the influence of transferred *Agrobacterium rhizogenes* genes to the hormone profile and morphological traits in *Centella asiatica* hairy roots.

Miguel Angel Alcalde[1], Maren Müller[2], Sergi Munné-Bosch[2], Mariana Landín[3], Pedro Pablo Gallego[4], Mercedes Bonfill[1], Javier Palazon[1], Diego Hidalgo-Martinez[1,5]

[1]Department of Biology, Healthcare and the Environment. Faculty of Pharmacy and Food Sciences. University of Barcelona. Spain

[2]Department of Evolutionary Biology, Ecology and Environmental Sciences, Faculty of Biology. University of Barcelona. Spain

[3]Pharmacology, Pharmacy and Pharmaceutical Technology Department, Grupo I+D Farma (GI-1645), Faculty of Pharmacy, University of Santiago, E-15782 Santiago de Compostela, Spain

[4]Agrobiotech for Health, Plant Biology and Soil Science Department, Faculty of Biology, University of Vigo, E-36310 Vigo, Spain

[5]Department of Plant and Microbial Biology, University of California, Berkeley, California 94720-3102, United States.
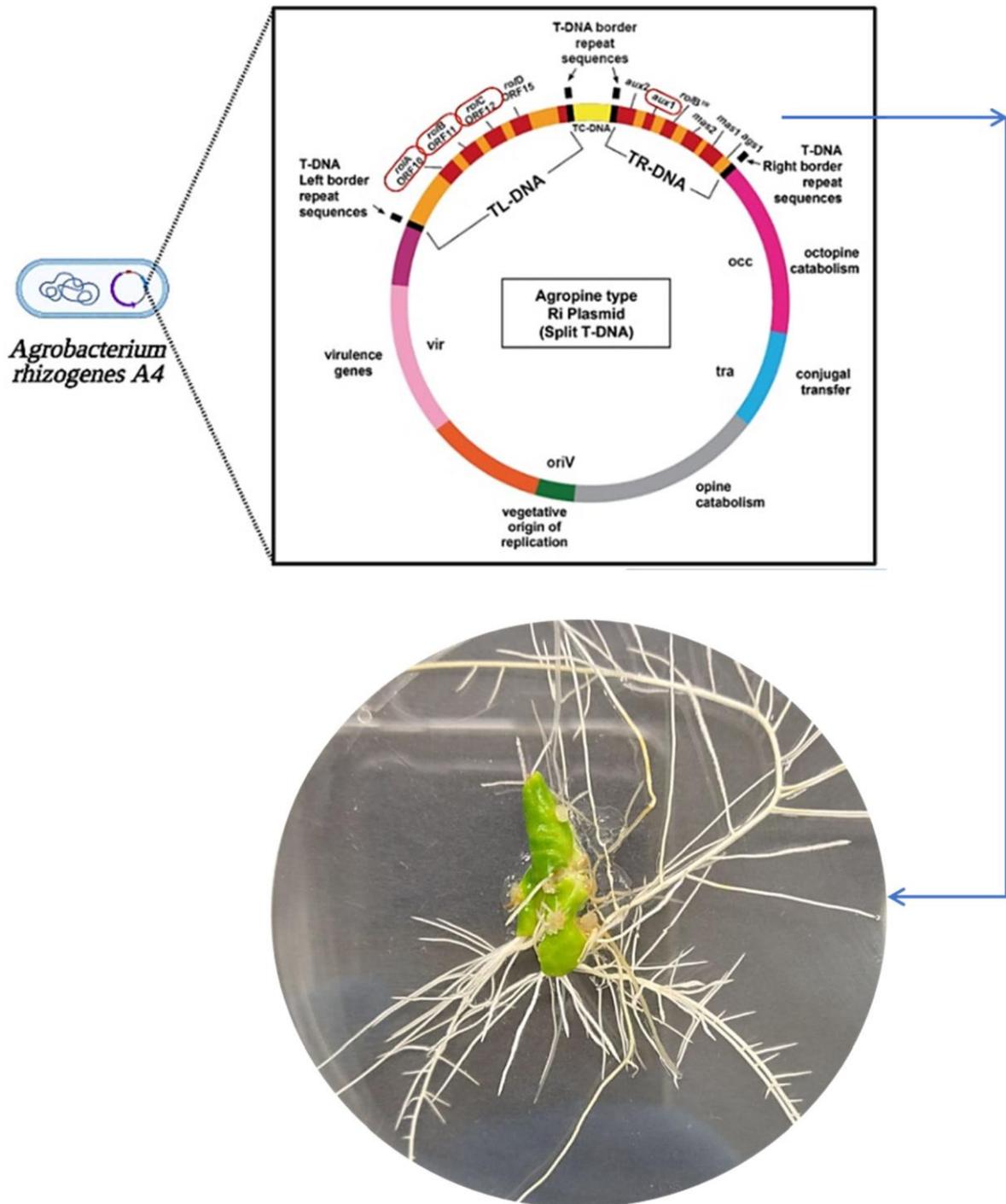
Figure 1. Ri plasmid from *Agrobacterium rhizogenes* A4 and photograph of hairy root induction from infected explants. Red circles are remarking genes in study. The Image of plasmid was modified from Ozyigit *et al.* (2013).
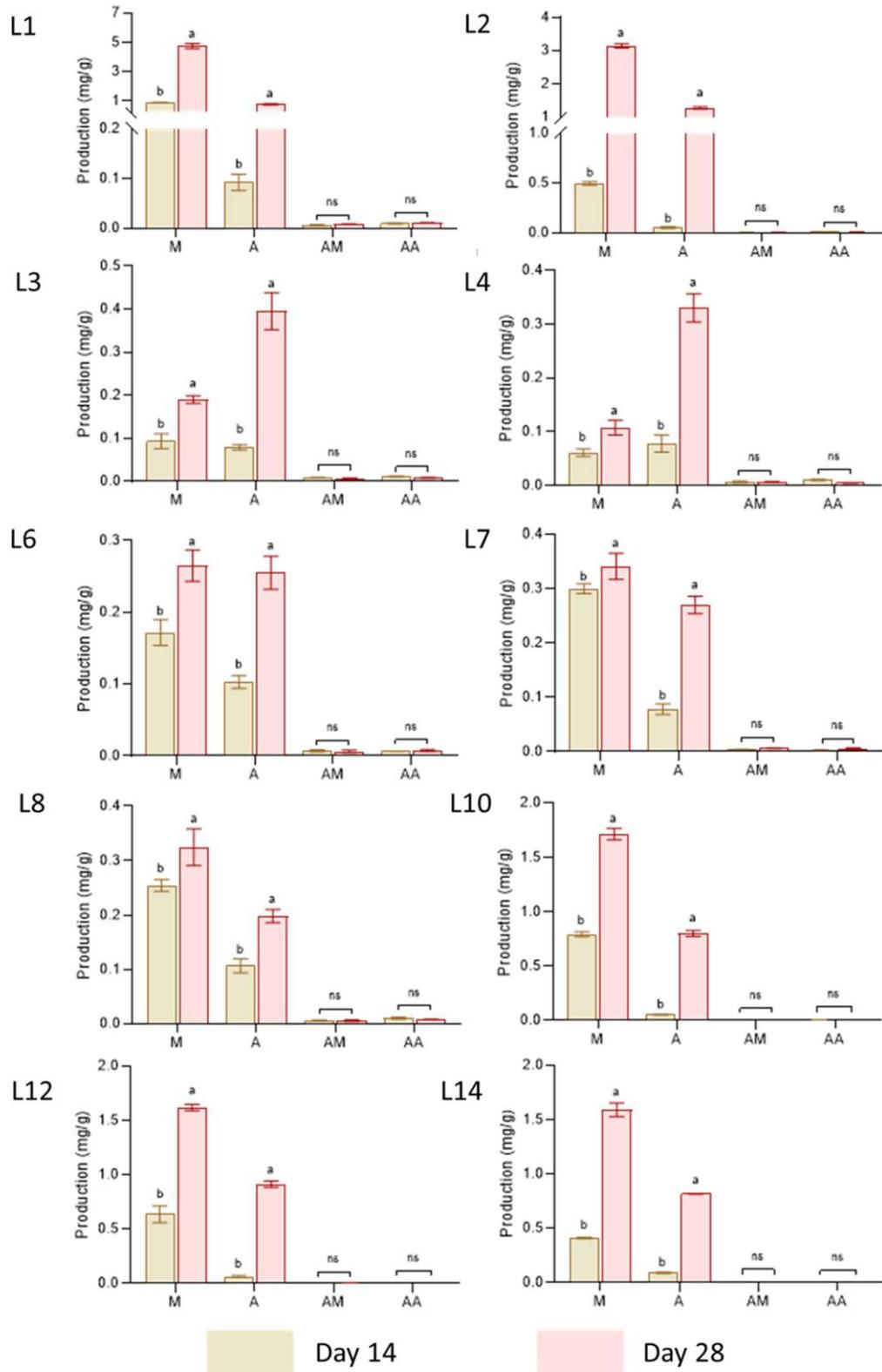
Figure 2. Specific centelloside production (mg/g DW). Madecassoside (M), Asiaticoside (A),

Madecassic acid (AM) and Asiatic acid (AA) of *C. asiatica* hairy root lines at day 14 and 28.
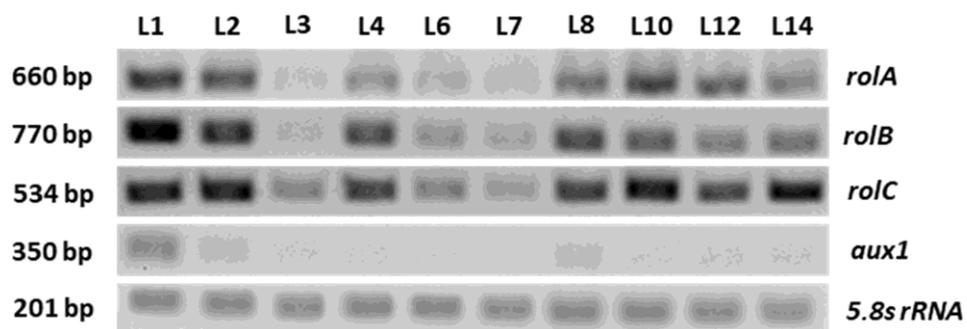
Figure 3. Semiquantitative reverse transcriptase PCR analysis. Expression of rol (rolA, rolB and rolC) and aux (aux1) genes in different *C. asiatica* hairy root lines in MS basal medium after 28 days of growth. 5.8s rRNA was used as housekeeping gene.
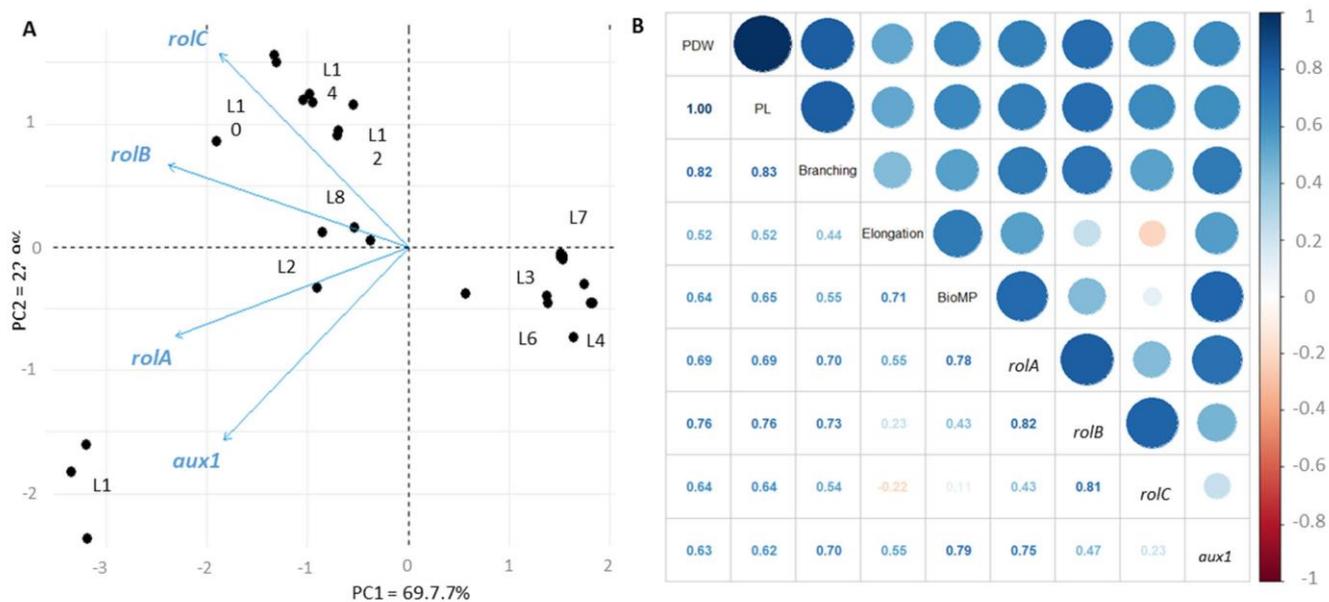
Figure 4. Gene expression analysis and correlation study with morphological parameters and centelloside production of *C. asiatica* hairy roots at day 14. (A) Principal Component Analysis of genes studied. (B) Correlation study of gene expression, morphological parameters and centelloside production. Elongation, refers to growth rate, while BioMP refers to productivity of biomass
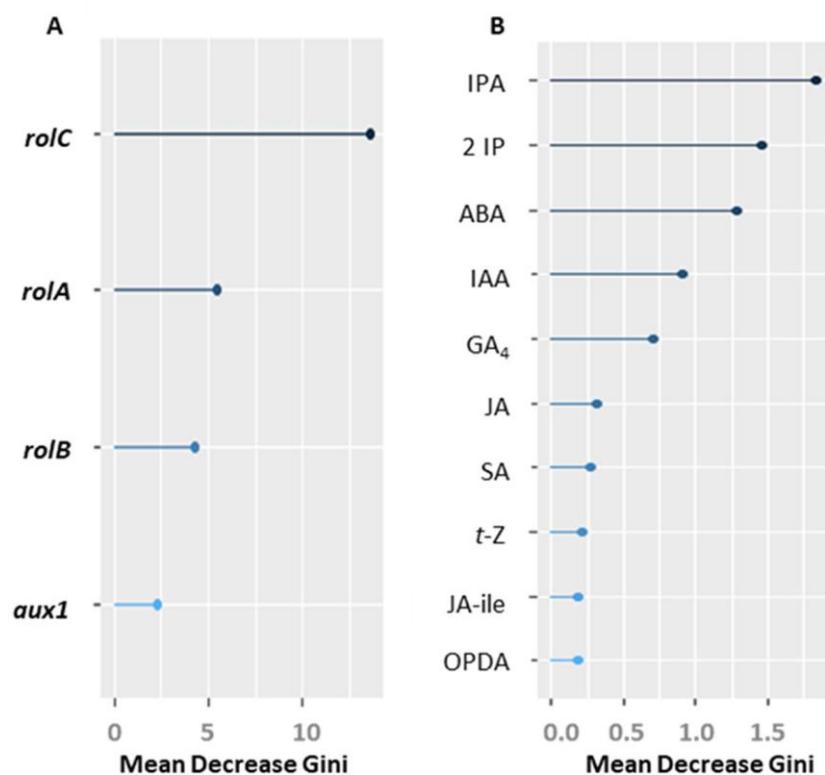
Figure 5. Effect of gene expression and hormone profiling in hairy root lines. (A) Importance of genes studied in hairy root lines. (B) Importance of plant hormones analyzed in hairy root lines.
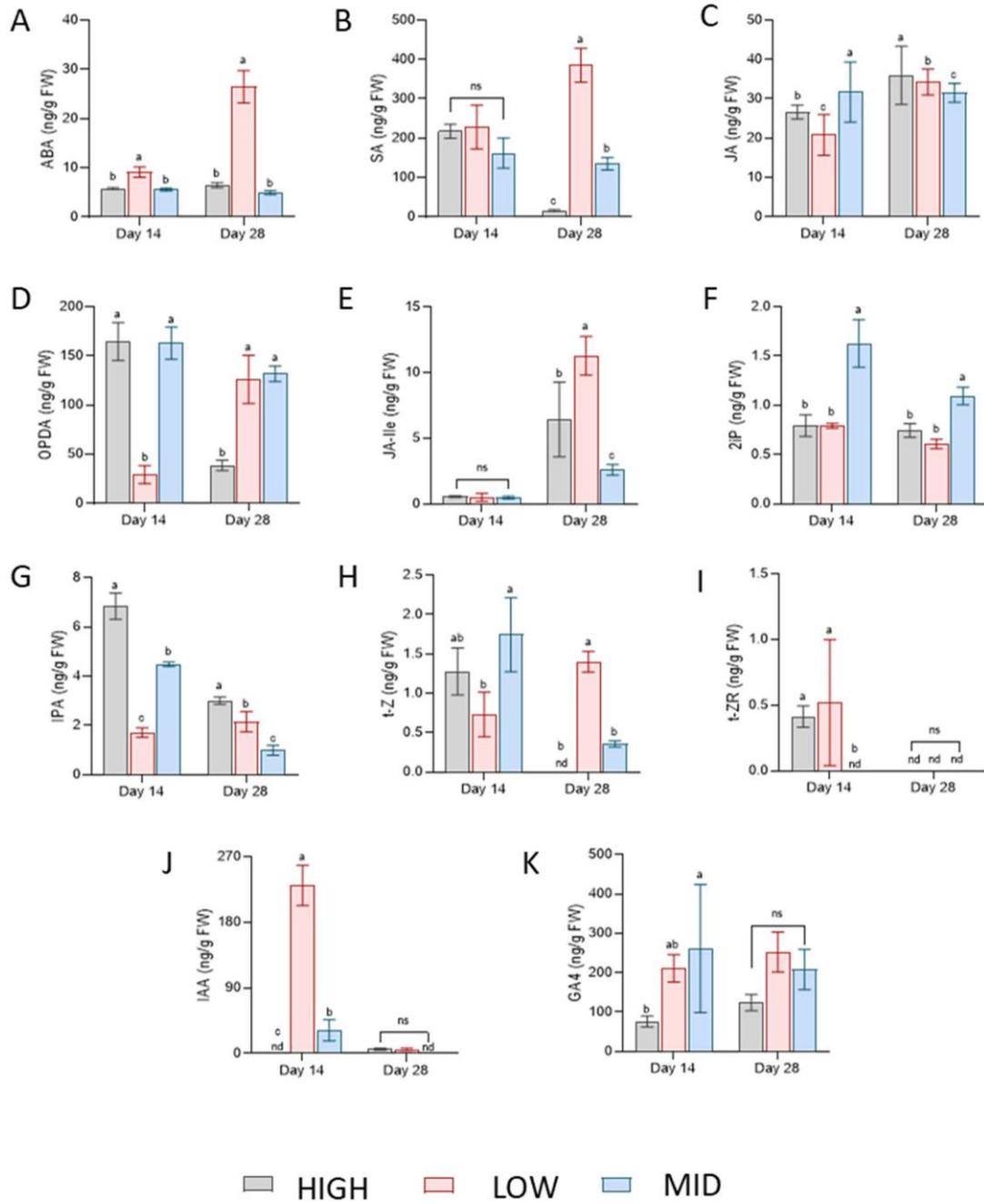
Figure 6. Phytohormone concentration (ng/g FW) in *C. asiatica* hairy root lines. (A) ABA, (B) SA, (C) JA, (D) OPDA, (E) JA-Ile, (F) 2iP, (G) IPA, (H) t-Z, (I) *t*-ZR, (J) IAA, and (K) GA$_4$. Data represent the mean ± SD of three replicates. Different letters show significant differences between hairy root lines at each time point. nd=no determinated. ns=no significance.

Figure 7. Correlation study of morphological parameters, production rates and plant hormone values at 14-day culture. Elon., refers to growth rate, while BioMP refers to biomass productivity.

Figure 8. Correlation study of morphological parameters, production rates and plant hormone values at 28-day culture. Elon., refers to growth rate, while BioMP refers to biomass productivity.

Figure 9. Machine Learning model workflow.
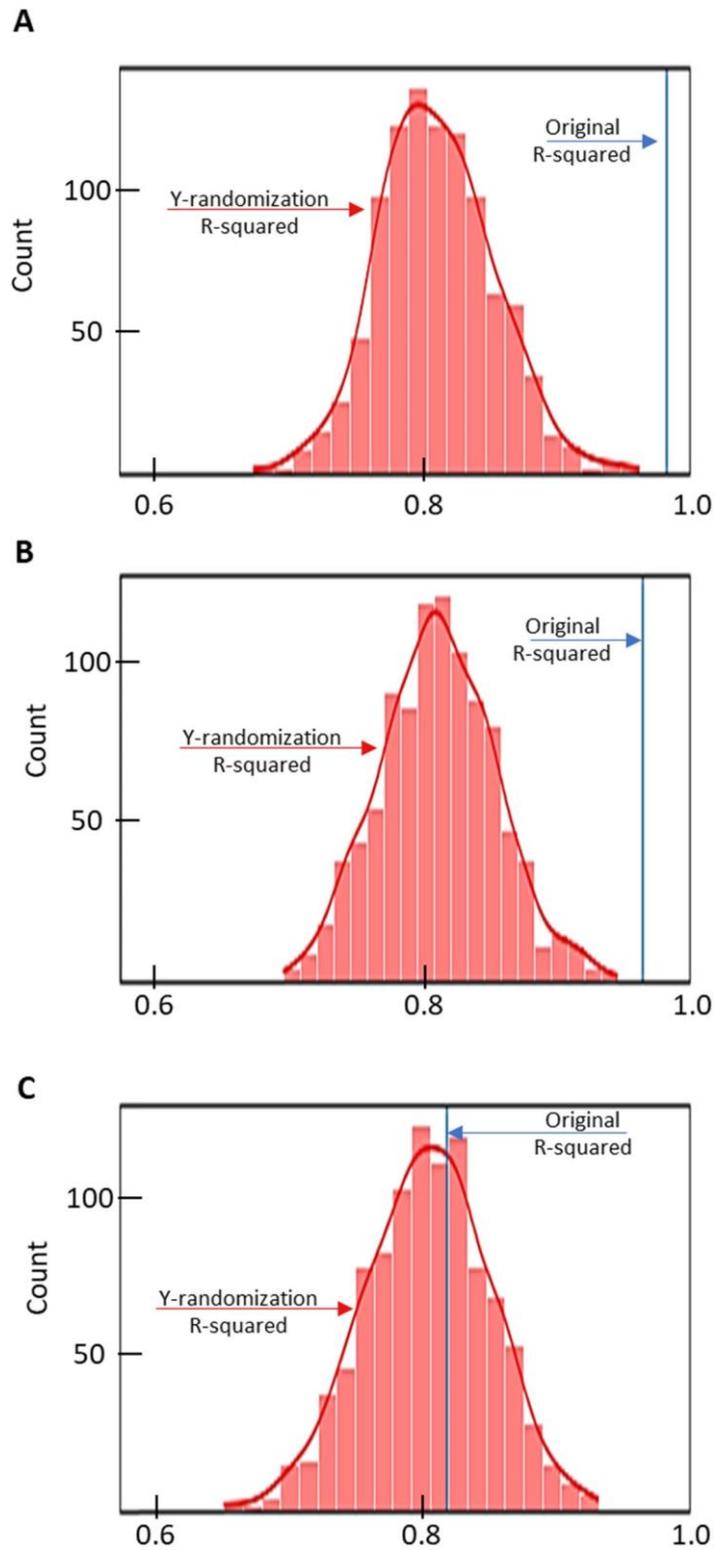
Figure 10. Y-randomization analysis for (A) ABA. (B) IPA. (C) JA. Blue line indicate the R-squared value with the origianal data. Red line and bars indicate the distribution of R-squered values after 1000 permutation.
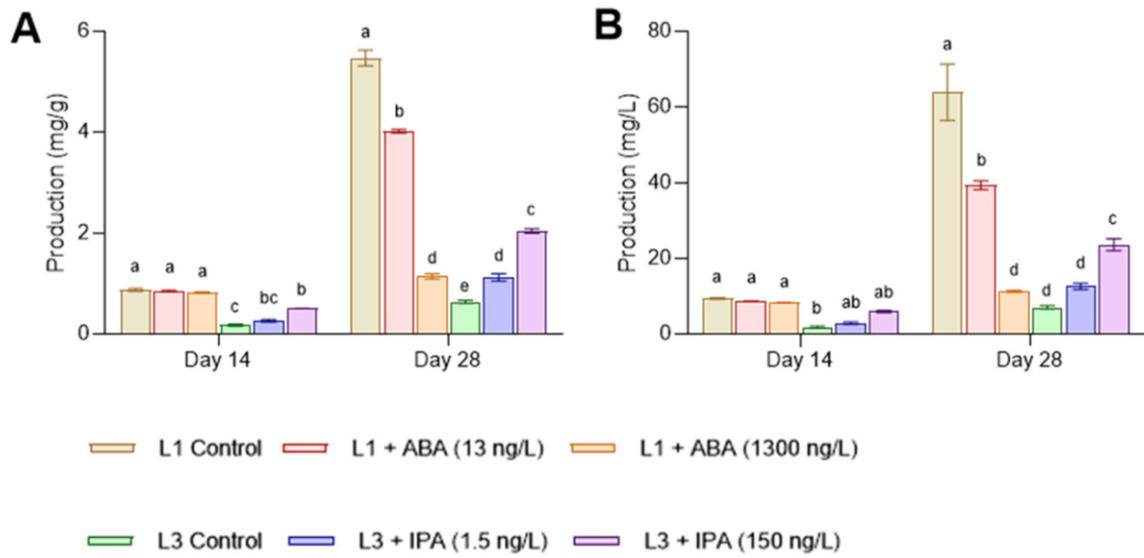
Figure 11. Centelloside content in feeding experiments on hairy roots lines, where L1 represent the HIGH group, while L3 the LOW group. (A) Production (mg/g DW) and (B) Production (mg/L of media) at 14 and 28 day. Data represent the mean ± SD of three replicates. Different letters show significant differences (α=0.05). between treatments.

Table 1. Primers used for confirmation of transformed roots

Gene		Primer sequence	Temperature melting (°C)	Amplicon size (pb)	Reference – Accession number
aux1	Forward	5´-TTCGAAGGAAGCTTGTCAGAA-3´	60	350	KX986281.1
	Reverse	5´-CTTAAATCCGTGTGACCATAG-3´			(*A. rhizogenes* A4)
rolA	Forward	5´-TGGAATTAGCCGGACTAAAC-3´	60	660	KX986281.1
	Reverse	5´-GCGTACGTTGTAATGTGTTG-3´			(*A. rhizogenes* A4)
rolB	Forward	5´-AGTTCAAGTCGGCTTTAGGC-3´	60	770	KX986281.1
	Reverse	5´-TCCACGATTTCAACCAGTAG-3´			(*A. rhizogenes* A4)
rolC	Forward	5´-TAACATGGCTGAAGACGACC-3´	60	534	KX986281.1
	Reverse	5´-AAACTTGCACTCGCCATGCC-3´			(*A. rhizogenes* A4)
5.8s rRNA	Forward	5´-CGGCAACGGATATCTCGGCTCT-3´	66	201	OK440405.1 – Mangas et al., 2008 [7] (*Arabidopsis thaliana*)
	Reverse	5´-TCCGCCCCGACCCCTTTC-3´			

Table 2. Gradient used for HPLC separation of the centellosides

Time (min)	Flow (mL/min)	Aqueous solvent (%)	Organic solvent (%)
0	1	80	20
15	1	62	38
30	1	30	70
35	1	30	70
37	1	80	20
45	1	80	20

Table 3. Dataset for prediction of production degree based on gene expression.

Class	Line	Day	PDW	PL	Branching	Growth rate	BioMP	RolA	RolB	RolC	Aux1
HIGH	L1	14	0.983	12	2.1	3.6	8.6	12373	10609	7093	9827
HIGH	L1	14	0.975	12.2	2.5	4.3	10	11208	10441	7639	9727
HIGH	L1	14	0.926	11.6	2.7	3.8	10.7	12374	8242	7099	11404
HIGH	L2	14	0.544	6.9	1.5	4.2	4.3	8618	9347	5112	288
HIGH	L2	14	0.564	7.1	1.3	3.5	5.7	1471	1938	2379	368
HIGH	L2	14	0.55	7	1.6	3.6	4.3	1472	1932	2370	368
LOW	L3	14	0.173	1.8	0.7	1.4	2.1	633	1930	4736	1156
LOW	L3	14	0.191	2.1	0.9	1.6	2.1	603	1939	4835	1107
LOW	L3	14	0.153	1.9	1.1	1.1	2.9	630	1923	4631	1138
LOW	L4	14	0.138	1.7	1.1	1.1	1.4	632	1837	4777	1096
LOW	L4	14	0.163	1.7	1.3	1.3	1.4	1472	1936	2379	368
LOW	L4	14	0.118	1.3	1.4	1.3	3.6	2354	2769	3184	818
LOW	L6	14	0.286	3.6	1.1	1.4	5	2748	2350	3199	792
LOW	L6	14	0.276	3.4	0.9	1.9	4.3	1071	1556	3506	732
LOW	L6	14	0.264	3.4	1.1	1.6	5	7081	7248	5441	678
LOW	L7	14	0.372	4.7	1.4	2.2	5	7603	8104	5375	294
LOW	L7	14	0.382	4.7	1.4	2.5	3.6	1518	1998	2391	377
LOW	L7	14	0.378	4.8	1.2	2.3	4.3	1472	1931	2379	368
HIGH	L8	14	0.36	4.5	1.8	2.4	3.6	7359	8259	5522	2733
HIGH	L8	14	0.358	4.3	1.2	2.3	2.9	1472	1933	2379	1638
HIGH	L8	14	0.368	4.6	1.5	2.1	3.6	3981	5077	3763	1463
MID	L10	14	0.853	10.6	1.8	1.6	2.9	4446	9504	11512	1520
MID	L10	14	0.822	10.5	1.6	2	7.1	8999	9059	10834	1940
MID	L10	14	0.857	10.5	1.6	1.7	5	4447	9506	11712	1441
MID	L12	14	0.78	9.8	2.2	1.8	2.1	3962	8266	9050	1515
MID	L12	14	0.681	8.6	2.2	1.3	2.9	3979	7690	9460	1857
MID	L12	14	0.639	8	1.9	1.8	3.6	3149	8242	9351	1156
MID	L14	14	0.5	6.4	1.7	1.1	2.9	4826	8245	10244	1257
MID	L14	14	0.486	6	1.7	1.3	2.1	4824	8243	10535	1196
MID	L14	14	0.511	6.4	1.6	1	3.6	4837	8291	10630	1520
HIGH	L1	28	5.25	64.7	3.4	4.7	41.4	12219	8441	7123	11404
HIGH	L1	28	5.64	70.8	3.1	4.9	37.5	10296	11686	7795	8656
HIGH	L1	28	5.57	70.7	2.9	4.4	28.6	12374	10831	7459	9720
HIGH	L2	28	4.42	54.2	2.1	4.7	23.6	9362	7378	5210	2224
HIGH	L2	28	4.37	54.4	2.2	4.3	25.4	8324	11175	4877	2611
HIGH	L2	28	4.48	57.4	2.5	3.9	18.2	8950	9034	5044	2289
LOW	L3	28	0.56	7	1	2.9	6.8	667	1169	4766	1267
LOW	L3	28	0.63	8.1	1	2.3	4.3	777	1144	4443	1641
LOW	L3	28	0.56	7	1.2	2.2	2.9	633	1634	4605	1472
LOW	L4	28	0.43	5.2	1.5	2.6	2.1	7633	5527	4471	1616
LOW	L4	28	0.47	5.5	1.6	2.6	1.1	5943	6171	4587	1775
LOW	L4	28	0.42	5.1	1.7	2.7	2.5	7038	5643	4529	1715
LOW	L6	28	0.54	6.9	1.6	2.5	6.4	2777	2375	3576	914
LOW	L6	28	0.54	6.7	2	2.2	8.9	3506	3262	3355	836
LOW	L6	28	0.48	6	2.2	2.1	7.1	2816	3349	3466	901
LOW	L7	28	0.65	8	1.5	2.9	7.1	1738	1769	2379	454
LOW	L7	28	0.6	7.3	1.7	2.8	6.4	2219	1718	2691	368
LOW	L7	28	0.58	7.3	1.6	2.5	3.9	1713	1594	2535	663
HIGH	L8	28	0.53	6.8	1.7	3	10.7	7506	7346	5663	2998
HIGH	L8	28	0.54	6.8	2.2	2.8	14.3	8098	8693	5458	2634
HIGH	L8	28	0.5	6.4	2.2	3.1	12.1	7293	7985	5561	2754
MID	L10	28	2.51	32.4	2.4	2.1	15.7	5980	10560	13379	1374
MID	L10	28	2.55	31.9	2	2.6	18.2	5550	12186	12665	1314
MID	L10	28	2.49	31.3	2.1	2.2	12.5	5710	11962	13022	1206
MID	L12	28	2.52	31.6	2.9	1.9	8.2	3916	7582	9312	1899
MID	L12	28	2.5	31.7	2.6	1.7	3.9	4560	8720	9483	1495
MID	L12	28	2.57	32.7	2.6	2.4	6.1	3543	8242	9397	1501
MID	L14	28	2.44	30.6	2	1.1	5.4	5117	6837	11080	2037
MID	L14	28	2.45	30.7	2	1.6	3.9	5211	8091	10573	1546
MID	L14	28	2.34	30	1.8	1.5	3.2	4724	7942	10827	1811

Table 4. Dataset for prediction of production degree based on hormone profiles.

Class	Day	ABA	SA	JA	OPDA	Ile-JA	2-iP	IPA	t-Z	t-ZR	IAA	GA4
HIGH	14	5.88	215.04	27.62	157.46	0.53	0.92	7.30	1.04	0.47	0.00	81.90
HIGH	14	5.91	236.97	27.76	186.40	0.52	0.70	6.27	1.61	0.45	0.00	86.58
HIGH	14	5.59	201.18	24.58	150.12	0.63	0.76	6.97	1.19	0.32	0.00	61.04
LOW	14	8.16	211.28	22.26	28.22	0.28	0.82	1.56	1.04	1.07	213.60	173.86
LOW	14	9.00	183.05	15.11	38.95	0.38	0.78	1.93	0.68	0.29	262.23	217.65
LOW	14	10.25	290.80	25.19	20.56	0.84	0.78	1.66	0.47	0.19	215.43	242.90
MID	14	5.26	197.92	25.11	178.33	0.52	1.86	4.38	1.50	0.00	43.46	290.38
MID	14	5.67	121.71	30.03	145.89	0.57	1.38	4.54	1.67	0.00	15.67	86.40
MID	14	5.90	165.78	40.04	164.91	0.38	1.65	4.54	2.06	0.00	37.05	407.67
HIGH	28	6.20	18.50	29.95	42.65	3.16	0.82	3.07	0.00	0.00	7.23	101.40
HIGH	28	7.06	14.34	33.72	40.42	7.83	0.69	3.11	0.00	0.00	5.06	140.81
HIGH	28	6.17	15.65	44.21	32.81	8.32	0.72	2.83	0.00	0.00	5.30	130.00
LOW	28	25.77	434.46	37.21	126.34	10.74	0.65	2.60	1.25	0.00	4.92	238.72
LOW	28	23.66	354.31	34.90	101.45	10.17	0.56	2.04	1.49	0.00	4.34	210.99
LOW	28	30.09	367.40	30.69	150.44	12.96	0.62	1.82	1.46	0.00	7.39	309.17
MID	28	5.24	152.46	28.88	122.97	2.16	1.19	0.94	0.40	0.00	0.00	262.27
MID	28	5.13	129.30	32.19	137.81	2.96	1.02	0.90	0.34	0.00	0.00	161.13
MID	28	4.48	122.92	33.49	135.20	2.70	1.08	1.16	0.34	0.00	0.00	200.94

Table 5. Data distribution.

Variable	14 days		28 days	
	Shapiro-Wilk (W)	p(normal)	Shapiro-Wilk (W)	p(normal)
ABA	0.7885	1.51E-02	0.7304	3.21E-03
SA	0.9694	8.90E-01	0.8535	8.15E-02
JA	0.9228	4.16E-01	0.8855	1.79E-01
OPDA	0.7884	1.51E-02	0.8308	4.56E-02
Ile-JA	0.9443	6.28E-01	0.8806	1.59E-01
2-iP	0.7896	1.55E-02	0.8997	2.50E-01
IPA	0.8883	1.92E-01	0.8873	1.87E-01
t-Z	0.9751	9.34E-01	0.802	2.15E-02
t-ZR	0.8436	6.33E-02	0.3898	3.22E-07
IAA	0.7595	6.99E-03	0.8323	4.73E-02
GA4	0.8979	2.40E-01	0.9734	9.22E-01
PDW	0.931	5.21E-02	0.8026	7.32E-05
PL	0.9324	5.69E-02	0.8065	8.63E-05
Branching	0.9653	4.19E-01	0.9743	6.61E-01
Growth rate	0.8583	9.31E-04	0.9081	1.34E-02
BioMP	0.8507	6.41E-04	0.8218	1.68E-04
RolA	0.8746	2.12E-03	0.9668	4.57E-01
RolB	0.7973	5.85E-05	0.9201	2.70E-02
RolC	0.8921	5.41E-03	0.8998	8.32E-03
Aux1	0.537	1.35E-08	0.5983	6.96E-08

References

Ozyigit, I. I., Dogan, I., and Artam Tarhan, E. (2013). Agrobacterium rhizogenes-Mediated Transformation and Its Biotechnological Applications in Crops. *Crop Improv. New Approaches Mod. Tech.*, 1–48. doi: 10.1007/978-1-4614-7028-1_1.

Annex 2.

Supplementary material

for Chapter 2

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

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

[1]Department of Biology, Healthcare and the Environment. Faculty of Pharmacy and Food Sciences. University of Barcelona. Spain

[2]Plant Proteomics and Functional Genomics Group, Department of Agrochemistry and Biochemistry, Faculty of Science, University of Alicante, Alicante, Spain

[3]Biotechnology, Health and Education Research Group, Posgraduate School, Cesar Vallejo University, Trujillo, Peru.

* Correspondence:
Javier Palazon and Miguel Angel Alcalde
javierpalazon@ub.edu; miguel.psr.94@gmail.com
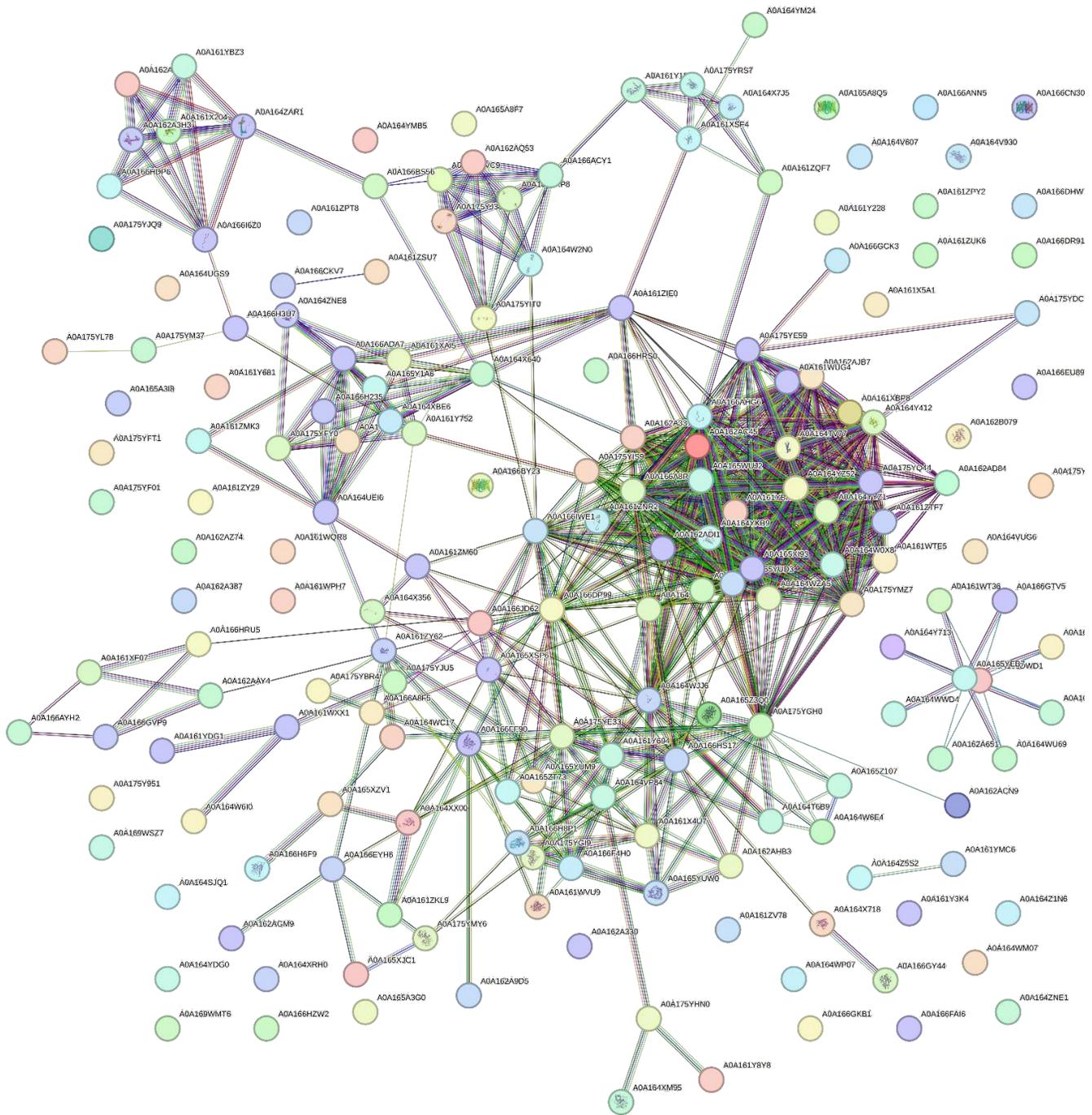
number of nodes:	**184**
number of edges:	689
average node degree:	7.49
avg. local clustering coefficient:	0.433
expected number of edges:	459
PPI enrichment p-value:	< 1.0e-16

Figure S1. Protein network constructed using the STRING web interface, illustrating the network formed by 184 proteins found in the STRING database. Network nodes depict proteins, and edges symbolize protein-protein associations. The accompanying legend displays the UniProt accession numbers.
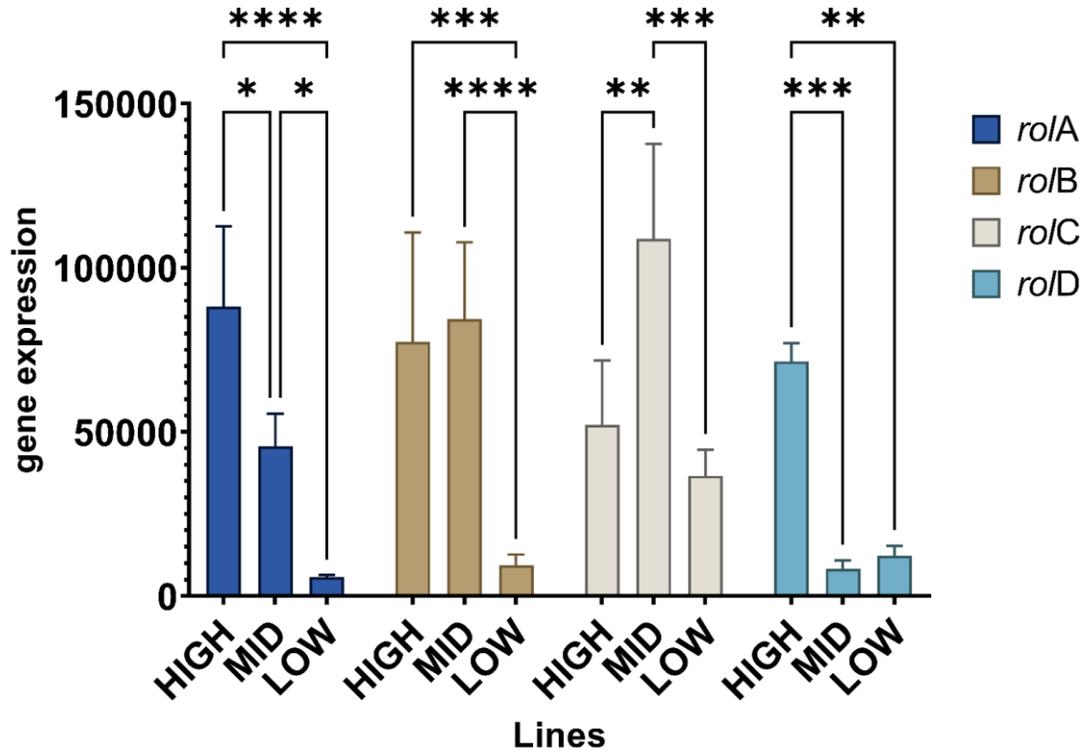
Figure S2. Normalized gene expression values from the transgenic lines: HIGH, MID, and LOW. Asterisks indicate statistical differences among the lines solely for the *rol*D gene (α = 0.05). Data represent the mean ± SD of three replicates.

Table S1

Biological process (Gene Ontology)					
GO-term	Description	observed gene count	background gene count	strength	false discovery rate
GO:1990547	Mitochondrial phosphate ion transmembrane transport	2	5	1,84	0,0335
GO:0097167	Circadian regulation of translation	3	10	1,71	0,0045
GO:0046294	Formaldehyde catabolic process	3	12	1,64	0,006
GO:0009768	Photosynthesis, light harvesting in photosystem I	4	23	1,48	0,0019
GO:0032922	Circadian regulation of gene expression	3	18	1,46	0,014
GO:0015977	Carbon fixation	4	32	1,33	0,0047
GO:0042744	Hydrogen peroxide catabolic process	10	100	1,24	2,40E-07
GO:0045037	Protein import into chloroplast stroma	3	31	1,22	0,0449
GO:0045454	Cell redox homeostasis	3	32	1,21	0,047
GO:0072593	Reactive oxygen species metabolic process	11	128	1,17	1,73E-07
GO:1990748	Cellular detoxification	18	219	1,15	1,34E-11
GO:0098869	Cellular oxidant detoxification	15	198	1,12	1,77E-09
GO:0006099	Tricarboxylic acid cycle	4	63	1,04	0,0314
GO:0042026	Protein refolding	5	84	1,01	0,0111
GO:0006096	Glycolytic process	5	86	1	0,0122
GO:1990542	Mitochondrial transmembrane transport	4	70	0,99	0,0413
GO:0009060	Aerobic respiration	7	127	0,98	0,0016
GO:0046034	ATP metabolic process	7	130	0,97	0,0018
GO:0015979	Photosynthesis	13	250	0,95	1,11E-06
GO:0019684	Photosynthesis, light reaction	5	126	0,84	0,0449
GO:0006979	Response to oxidative stress	15	389	0,82	2,92E-06
GO:0006412	Translation	23	613	0,81	2,07E-09
GO:0043604	Amide biosynthetic process	25	692	0,8	9,66E-10
GO:0006457	Protein folding	12	332	0,8	0,00013
GO:0006091	Generation of precursor metabolites and energy	17	478	0,79	1,11E-06
GO:0009150	Purine ribonucleotide metabolic process	9	269	0,76	0,0034
GO:0043603	Cellular amide metabolic process	26	859	0,72	5,00E-09
GO:0016052	Carbohydrate catabolic process	10	332	0,72	0,0031
GO:1901605	Alpha-amino acid metabolic process	9	299	0,72	0,006
GO:0098656	Anion transmembrane transport	7	249	0,69	0,035
GO:0009117	Nucleotide metabolic process	10	376	0,66	0,0062
GO:0044282	Small molecule catabolic process	8	305	0,66	0,0256
GO:0055086	Nucleobase-containing small molecule metabolic process	12	468	0,65	0,0026
GO:0006417	Regulation of translation	9	365	0,63	0,0185
GO:0044271	Cellular nitrogen compound biosynthetic process	32	1421	0,59	2,29E-08
GO:1901566	Organonitrogen compound biosynthetic process	34	1556	0,58	1,39E-08
GO:0019752	Carboxylic acid metabolic process	20	1058	0,51	0,00061
GO:0044281	Small molecule metabolic process	35	1928	0,5	4,92E-07
GO:0009056	Catabolic process	34	1957	0,48	2,19E-06
GO:0070887	Cellular response to chemical stimulus	25	1494	0,46	0,00034
GO:0044248	Cellular catabolic process	22	1349	0,45	0,0017
GO:1901575	Organic substance catabolic process	24	1742	0,38	0,006

204

GO:0005975	Carbohydrate metabolic process	17	1237	0,38	0,0449
GO:1901576	Organic substance biosynthetic process	44	3279	0,36	1,92E-05
GO:0044249	Cellular biosynthetic process	41	3093	0,36	6,90E-05
GO:0042221	Response to chemical	33	2512	0,36	0,0011
GO:0010467	Gene expression	25	2141	0,3	0,0336
GO:0034641	Cellular nitrogen compound metabolic process	41	3822	0,27	0,0058
GO:0044237	Cellular metabolic process	91	9779	0,21	1,92E-05
GO:1901564	Organonitrogen compound metabolic process	57	6024	0,21	0,0056
GO:0008152	Metabolic process	116	12879	0,19	2,01E-07
GO:0009987	Cellular process	142	17087	0,16	1,83E-08
GO:0071704	Organic substance metabolic process	91	11799	0,12	0,0237

KEGG Pathway					
Pathway	**Description**	**observed gene count**	**background gene count**	**strength**	**false discovery rate**
dcr00196	Photosynthesis - antenna proteins	3	19	1,44	0,0033
dcr00195	Photosynthesis	5	42	1,31	0,00014
dcr00220	Arginine biosynthesis	3	28	1,27	0,0072
dcr03010	Ribosome	20	227	1,18	1,62E-15
dcr00710	Carbon fixation in photosynthetic organisms	5	59	1,17	0,00052
dcr00620	Pyruvate metabolism	6	79	1,12	0,00017
dcr00940	Phenylpropanoid biosynthesis	11	155	1,09	9,95E-08
dcr00260	Glycine, serine and threonine metabolism	4	56	1,09	0,0045
dcr03050	Proteasome	3	42	1,09	0,0166
dcr00010	Glycolysis / Gluconeogenesis	7	107	1,05	0,0001
dcr00592	alpha-Linolenic acid metabolism	3	47	1,04	0,0213
dcr01200	Carbon metabolism	14	225	1,03	4,48E-09
dcr00630	Glyoxylate and dicarboxylate metabolism	4	67	1,01	0,0072
dcr04146	Peroxisome	4	80	0,94	0,0113
dcr00190	Oxidative phosphorylation	5	123	0,85	0,0075
dcr03018	RNA degradation	4	100	0,84	0,0213
dcr01230	Biosynthesis of amino acids	7	182	0,82	0,0016
dcr04141	Protein processing in endoplasmic reticulum	6	172	0,78	0,0062
dcr01110	Biosynthesis of secondary metabolites	30	1124	0,66	2,18E-10
dcr01100	Metabolic pathways	53	2108	0,64	6,86E-18

Table S2

ID	UNIPROT Accession	Protein description	Adjusted p-value	Significance	more abundant in:
		LOW vs HIGH			
2	A0A175YMY6	Alcohol dehydrogenase	0.0015	**	HIGH
6	A0A162AGM9	Alcohol dehydrogenase	0.0149	*	HIGH
48	A0A6M3W1G9	ATP synthase subunit beta, chloroplastic	0.0002	***	LOW
50	A0A166H6F9	Superoxide dismutase [Cu-Zn]	<0.0001	****	LOW
57	A0A175YL78	Cysteine protease	0.0184	*	LOW
60	A0A175YM37	Thiamine thiazole synthase, chloroplastic	<0.0001	****	LOW
66	A0A165A8Q5	Uncharacterized protein	0.0004	***	LOW
82	A0A161Y694	dihydrolipoyllysine-residue succinyltransferase	0.0031	**	LOW
83	A0A164VUG6	Beta-glucosidase	0.0308	*	LOW
87	A0A165A3I8	FAD-binding PCMH-type domain-containing protein	<0.0001	****	LOW
89	A0A166DWD1	FAD-binding PCMH-type domain-containing protein	<0.0001	****	LOW
90	A0A161Y681	FAD-binding PCMH-type domain-containing protein	<0.0001	****	LOW
102	A0A161WT36	Peroxidase	<0.0001	****	LOW
104	A0A164SJQ1	Fibronectin type III-like domain-containing protein	0.0329	*	LOW
105	A0A166CN30	Uncharacterized protein	<0.0001	****	LOW
107	A0A162AHB3	Glyceraldehyde 3-phosphate dehydrogenase NAD(P) binding domain-containing protein	0.0026	**	LOW
108	A0A165A3G0	FAD-binding PCMH-type domain-containing protein	<0.0001	****	LOW
110	A0A175YE59	Eukaryotic translation initiation factor 3 subunit B	<0.0001	****	LOW
116	A0A161Y3K4	FAD-binding PCMH-type domain-containing protein	<0.0001	****	LOW
126	A0A1B0YYH0	Peroxidase	<0.0001	****	LOW
129	A0A164Y713	Peroxidase	<0.0001	****	LOW
131	A0A165YEB7	FAD-binding PCMH-type domain-containing protein	<0.0001	****	LOW
136	A0A162ACN9	Uncharacterized protein	0.0389	*	LOW
137	A0A166BY23	Uncharacterized protein	0.0326	*	LOW

142	A0A166BS56	UDP-arabinopyranose mutase	0.0478	*	HIGH
151	A0A161ZMK3	Clp R domain-containing protein	<0.0001	****	LOW
152	A0A166A8R9	Tr-type G domain-containing protein	<0.0001	****	HIGH
159	Q6F6A3	Cysteine protease	<0.0001	****	LOW
160	Q6F6A5	Cysteine protease	<0.0001	****	LOW
170	A0A6L5B772	Plant heme peroxidase family profile domain-containing protein	0.0025	**	HIGH
178	A0A166CKV7	Tropinone reductase-like 3	<0.0001	****	LOW
180	A0A166FAI6	Uncharacterized protein	0.0329	*	LOW
182	A0A175YHN0	Cysteine synthase	<0.0001	****	LOW
189	A0A166HRU5	Uncharacterized protein	0.0111	*	LOW
200	A0A166GVP9	ADP/ATP translocase	0.0018	**	LOW
204	A0A165YUM9	FAD/NAD(P)-binding domain-containing protein	<0.0001	****	LOW
209	A0A166GTV5	Peroxidase	<0.0001	****	LOW
220	A0A164VP84	D-3-phosphoglycerate dehydrogenase	0.0268	*	HIGH
248	A0A162A1B6	Histone H2A	<0.0001	****	LOW
249	A0A165Z3Q0	Ketol-acid reductoisomerase	0.0045	**	HIGH
251	A0A164YKB9	30S ribosomal protein S8, chloroplastic	0.0240	*	LOW
259	A0A6L5BB43	Tr-type G domain-containing protein	<0.0001	****	HIGH
272	A0A169WMT6	Enoyl reductase (ER) domain-containing protein	0.0003	***	LOW
289	A0A161WXX1	ADP-ribosylation factor	0.0232	*	LOW

Table S3

Biological process (Gene Ontology)					
GO-term	**Description**	**observed gene count**	**background gene count**	**strength**	**false discovery rate**
GO:0046294	Formaldehyde catabolic process	2	12	2,14	0,0489
GO:1990748	Cellular detoxification	7	219	1,43	0,0000374
GO:0072593	Reactive oxygen species metabolic process	4	128	1,42	0,0089
GO:0098869	Cellular oxidant detoxification	5	198	1,32	0,0026
GO:0006979	Response to oxidative stress	6	389	1,11	0,0037
GO:0044281	Small molecule metabolic process	10	1928	0,64	0,0289
KEGG Pathway					
Pathway	**Description**	**observed gene count**	**background gene count**	**strength**	**false discovery rate**
dcr00770	Pantothenate and CoA biosynthesis	2	35	1,68	0,0302
dcr00940	Phenylpropanoid biosynthesis	5	155	1,43	0,0000742
dcr01230	Biosynthesis of amino acids	3	182	1,14	0,0377
dcr01110	Biosynthesis of secondary metabolites	10	1124	0,87	0,0000742
dcr01100	Metabolic pathways	13	2108	0,71	0,0000742

Table S4

Model	ID	UNIPROT Accession	Magnitude. p[1]	Reliability. p(corr)[1]	Protein description	Biomarker to:
Adv vs HIGH	259	A0A6L5BB43	0,133426	0,84764	Tr-type G domain-containing protein	HIGH
	2	A0A175YMY6	0,107198	0,972109	Alcohol dehydrogenase	HIGH
	31	A0A161Y1E5	-0,225166	-0,994278	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplastic	Advs
	48	A0A6M3W1G9	-0,222258	-0,902116	ATP synthase subunit beta, chloroplastic	Advs
	29	A0A164W2N0	-0,218023	-0,851047	Uncharacterized protein	Advs
	27	A0A6M3W3P4	-0,204972	-0,9502	ATP synthase subunit alpha, chloroplastic	Advs
	41	A0A6M3W1B1	-0,194244	-0,899031	Ribulose bisphosphate carboxylase large chain	Advs
	30	A0A162AQ53	-0,174538	-0,836359	Chlorophyll a-b binding protein, chloroplastic	Advs
	24	A0A6M3W1R0	-0,170415	-0,876275	Photosystem II CP47 reaction center protein	Advs

	54	A0A141CK 52	-0,140342	-0,805622	Photosystem II CP43 reaction center protein	Advs
	66	A0A165A8 Q5	-0,138187	-0,87275	Uncharacterized protein	Advs
	50	A0A166H6 F9	-0,128709	-0,856844	Superoxide dismutase [Cu-Zn]	Advs
	35	A0A166AC Y1	-0,12439	-0,912772	23 kDa subunit of oxygen evolving system of photosystem II	Advs
	23	A0A8A9IT E7	-0,120288	-0,87129	Photosystem I P700 chlorophyll a apoprotein A1	Advs
	43	A0A6M3W 3D2	-0,103038	-0,80787	Photosystem I P700 chlorophyll a apoprotein A2	Advs
Adv vs MID	2	A0A175YM Y6	0,0958309	0,935986	Alcohol dehydrogenase	MID
	31	A0A161Y1 E5	-0,216865	-0,994257	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplastic	Advs
	48	A0A6M3W 1G9	-0,212829	-0,899382	ATP synthase subunit beta, chloroplastic	Advs
	29	A0A164W2 N0	-0,210303	-0,849192	Uncharacterized protein	Advs
	27	A0A6M3W 3P4	-0,200677	-0,951815	ATP synthase subunit alpha, chloroplastic	Advs
	41	A0A6M3W 1B1	-0,187932	-0,898658	Ribulose bisphosphate carboxylase large chain	Advs
	30	A0A162AQ 53	-0,169269	-0,838643	Chlorophyll a-b binding protein, chloroplastic	Advs
	24	A0A6M3W 1R0	-0,164781	-0,876803	Photosystem II CP47 reaction center protein	Advs
	66	A0A165A8 Q5	-0,137168	-0,883628	Uncharacterized protein	Advs
	54	A0A141CK 52	-0,134585	-0,802704	Photosystem II CP43 reaction center protein	Advs
	19 2	A0A162A3 H3	-0,130667	-0,875075	Histone H2B	Advs
	50	A0A166H6 F9	-0,128992	-0,870864	Superoxide dismutase [Cu-Zn]	Advs
	35	A0A166AC Y1	-0,118609	-0,910098	23 kDa subunit of oxygen evolving system of photosystem II	Advs
	23	A0A8A9IT E7	-0,115888	-0,869613	Photosystem I P700 chlorophyll a apoprotein A1	Advs
Adv vs LOW	90	A0A161Y68 1	0,288943	0,943453	FAD-binding PCMH-type domain-containing protein	LOW
	11 6	A0A161Y3 K4	0,261717	0,841533	FAD-binding PCMH-type domain-containing protein	LOW
	11 0	A0A175YE 59	0,183437	0,902943	Eukaryotic translation initiation factor 3 subunit B	LOW
	13 1	A0A165YE B7	0,171391	0,850351	FAD-binding PCMH-type domain-containing protein	LOW
	60	A0A175YM 37	0,157644	0,92907	Thiamine thiazole synthase, chloroplastic	LOW
	20 9	A0A166GT V5	0,153533	0,874317	Peroxidase	LOW
	16 0	Q6F6A5	0,14936	0,814818	Cysteine protease	LOW

	#	ID	Value1	Value2	Protein	Category
	108	A0A165A3G0	0,148984	0,888081	FAD-binding PCMH-type domain-containing protein	LOW
	105	A0A166CN30	0,137822	0,93839	Uncharacterized protein	LOW
	159	Q6F6A3	0,10675	0,84086	Cysteine protease	LOW
	29	A0A164W2N0	-0,20701	-0,870135	Uncharacterized protein	Advs
	31	A0A161Y1E5	-0,205389	-0,982586	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplastic	Advs
	27	A0A6M3W3P4	-0,199608	-0,965169	ATP synthase subunit alpha, chloroplastic	Advs
	48	A0A6M3W1G9	-0,179818	-0,859137	ATP synthase subunit beta, chloroplastic	Advs
	41	A0A6M3W1B1	-0,172795	-0,8605	Ribulose bisphosphate carboxylase large chain	Advs
	30	A0A162AQ53	-0,167807	-0,860586	Chlorophyll a-b binding protein, chloroplastic	Advs
	24	A0A6M3W1R0	-0,163908	-0,899508	Photosystem II CP47 reaction center protein	Advs
	54	A0A141CK52	-0,134697	-0,82957	Photosystem II CP43 reaction center protein	Advs
	35	A0A166ACY1	-0,118701	-0,927738	23 kDa subunit of oxygen evolving system of photosystem II	Advs
	23	A0A8A9ITE7	-0,115221	-0,893216	Photosystem I P700 chlorophyll a apoprotein A1	Advs
	43	A0A6M3W3D2	-0,100028	-0,83764	Photosystem I P700 chlorophyll a apoprotein A2	Advs
LOW vs HIGH	259	A0A6L5BB43	0,139244	0,957466	Tr-type G domain-containing protein	HIGH
	152	A0A166A8R9	0,115929	0,938448	Tr-type G domain-containing protein	HIGH
	116	A0A161Y3K4	-0,318132	-0,96664	FAD-binding PCMH-type domain-containing protein	LOW
	90	A0A161Y681	-0,30325	-0,966754	FAD-binding PCMH-type domain-containing protein	LOW
	131	A0A165YEB7	-0,197024	-0,953451	FAD-binding PCMH-type domain-containing protein	LOW
	126	A0A1B0YYH0	-0,195764	-0,931565	Peroxidase	LOW
	110	A0A175YE59	-0,187379	-0,933272	Eukaryotic translation initiation factor 3 subunit B	LOW
	129	A0A164Y713	-0,180432	-0,944644	Peroxidase	LOW
	182	A0A175YHN0	-0,178691	-0,88692	Cysteine synthase	LOW
	108	A0A165A3G0	-0,169706	-0,987492	FAD-binding PCMH-type domain-containing protein	LOW
	60	A0A175YM37	-0,164818	-0,956669	Thiamine thiazole synthase, chloroplastic	LOW
	160	Q6F6A5	-0,16324	-0,897911	Cysteine protease	LOW
	89	A0A166DWD1	-0,157289	-0,909891	FAD-binding PCMH-type domain-containing protein	LOW

	50	A0A166H6F9	-0,152484	-0,928006	Superoxide dismutase [Cu-Zn]	LOW
	102	A0A161WT36	-0,13933	-0,93381	Peroxidase	LOW
	209	A0A166GTV5	-0,131599	-0,846206	Peroxidase	LOW
	178	A0A166CKV7	-0,123794	-0,876318	Tropinone reductase-like 3	LOW
	159	Q6F6A3	-0,114354	-0,911217	Cysteine protease	LOW
	105	A0A166CN30	-0,112004	-0,938319	Uncharacterized protein	LOW
	204	A0A165YUM9	-0,101801	-0,886908	FAD/NAD(P)-binding domain-containing protein	LOW
	87	A0A165A3I8	-0,101763	-0,9343	FAD-binding PCMH-type domain-containing protein	LOW

MID vs HIGH	192	A0A162A3H3	0,315458	0,895239	Histone H2B	HIGH
	105	A0A166CN30	0,266209	0,950556	Uncharacterized protein	HIGH
	63	A0A161WPH7	-0,147028	-0,881294	Bet v I/Major latex protein domain-containing protein	MID
	238	A0A164ZNE1	0,171347	0,934605	peroxidase	HIGH
	178	A0A166CKV7	0,161729	0,857005	Tropinone reductase-like 3	HIGH
	160	Q6F6A5	0,15231	0,824951	Cysteine protease	HIGH
	264	A0A162A651	0,124555	0,866691	peroxidase	HIGH

MID vs LOW	259	A0A6L5BB43	0,110044	0,899095	Tr-type G domain-containing protein	MID
	152	A0A166A8R9	0,106355	0,944438	Tr-type G domain-containing protein	MID
	116	A0A161Y3K4	-0,318903	-0,971591	FAD-binding PCMH-type domain-containing protein	LOW
	90	A0A161Y681	-0,301064	-0,970668	FAD-binding PCMH-type domain-containing protein	LOW
	126	A0A1B0YYH0	-0,207903	-0,954703	Peroxidase	LOW
	131	A0A165YEB7	-0,197045	-0,956845	FAD-binding PCMH-type domain-containing protein	LOW
	129	A0A164Y713	-0,184618	-0,951689	Peroxidase	LOW
	110	A0A175YE59	-0,174848	-0,926359	Eukaryotic translation initiation factor 3 subunit B	LOW
	182	A0A175YHN0	-0,171134	-0,883212	Cysteine synthase	LOW
	108	A0A165A3G0	-0,168863	-0,976467	FAD-binding PCMH-type domain-containing protein	LOW
	60	A0A175YM37	-0,16844	-0,959581	Thiamine thiazole synthase, chloroplastic	LOW
	160	Q6F6A5	-0,168281	-0,91361	FAD-binding PCMH-type domain-containing protein	LOW

89	A0A166DW D1	-0,158319	-0,922465	FAD-binding PCMH-type domain-containing protein	LOW
50	A0A166H6 F9	-0,151607	-0,932725	Superoxide dismutase [Cu-Zn]	LOW
10 2	A0A161WT 36	-0,143393	-0,945007	Peroxidase	LOW
10 5	A0A166CN 30	-0,139968	-0,974077	Uncharacterized protein	LOW
17 8	A0A166CK V7	-0,133421	-0,90631	Tropinone reductase-like 3	LOW
15 9	Q6F6A3	-0,114916	-0,919073	Cysteine protease	LOW
19 2	A0A162A3 H3	-0,109918	-0,850763	Histone H2B	LOW
15 1	A0A161ZM K3	-0,104047	-0,818885	Clp R domain-containing protein	LOW

Table S5

Line	Gene	Cx	Ix	Sx	Cr	Ir	Sr	X/R	2(X/R)
HIGH	*rolA*	33,415	41,068	-5,144	22,216	26,347	-2,777	1,000	2,001
	rolB	36,063	42,813	-5,158	22,216	26,347	-2,777	0,662	1,325
	rolC	33,148	36,422	-2,201	22,216	26,347	-2,777	1,000	2,000
	rolD	24,635	28,812	-2,808	22,216	26,347	-2,777	1,000	2,000
MID	*rolA*	30,717	35,184	-2,903	21,403	25,102	-2,486	1,124	2,248
	rolB	35,295	41,099	-3,901	21,403	25,102	-2,486	1,000	1,999
	rolC	30,798	33,694	-1,882	21,403	25,102	-2,486	1,124	2,248
	rolD	24,345	28,621	-2,875	21,403	25,102	-2,486	0,999	1,997
LOW	*rolA*	32,963	43,992	-7,641	21,345	26,847	-3,454	0,709	1,417
	rolB	36,92	43,803	-4,321	21,345	26,847	-3,454	1,000	2,000
	rolC	32,536	37,376	-2,948	21,345	26,847	-3,454	1,119	2,238
	rolD	24,265	29,565	-3,327	21,345	26,847	-3,454	1,000	2,000

Annex 3.

Supplementary material

for Chapter 4

Enhancing Centelloside Production in *Centella asiatica* Hairy Root Lines through Metabolic Engineering of Triterpene Biosynthetic Pathway Early Genes

Miguel Angel Alcalde[1,2], Javier Palazon[1*], Mercedes Bonfill[1], Diego Hidalgo-Martinez[1,*]

[1] Department of Biology, Healthcare and the Environment, Faculty of Pharmacy and Food Sciences, University of Barcelona, 08028 Barcelona, Spain; miguel.psr.94@gmail.com (M.A.); mbonfill@ub.edu (M.B.)

[2] Biotechnology, Health and Education Research Group, Posgraduate School, Cesar Vallejo University, Trujillo, Peru.

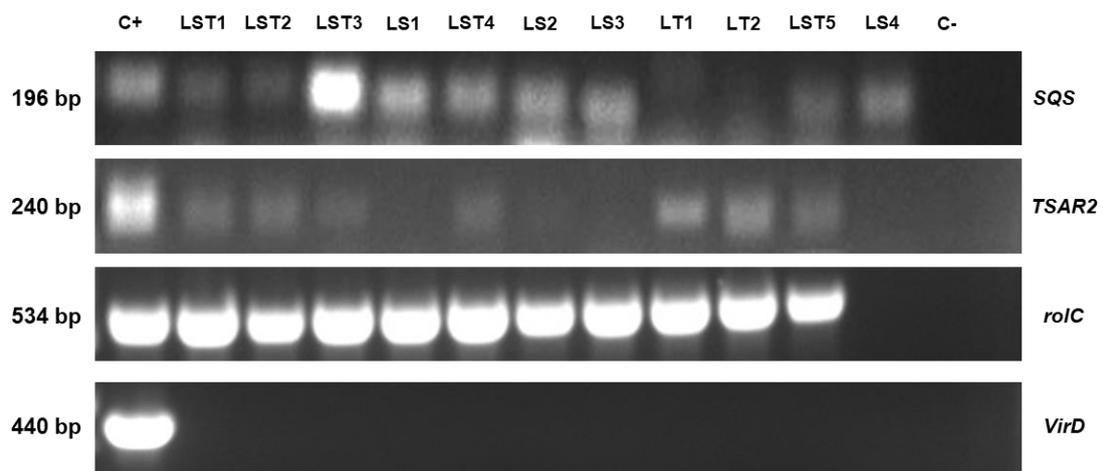* Correspondence: dhidalgo@ub.edu (D.H.); javierpalazon@ub.edu (J.P.)

Figure S1. PCR analysis: detection of *SQS*, *TSAR2*, *rolC* and *VirD* genes in different *C. asiatica* transgenic root lines.
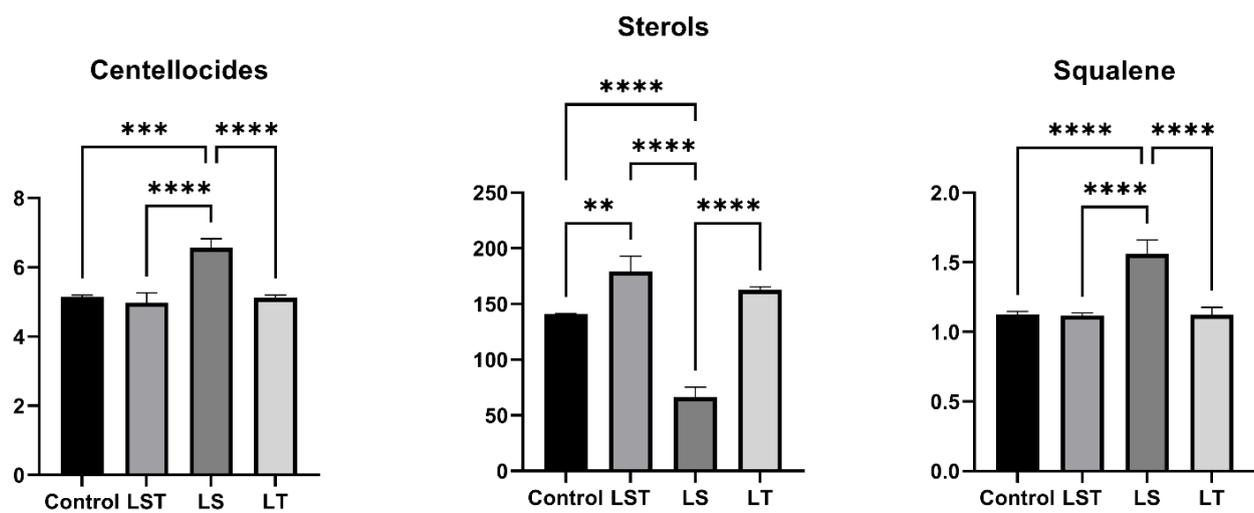

Figure S2. Statistical differences between different types of lines.
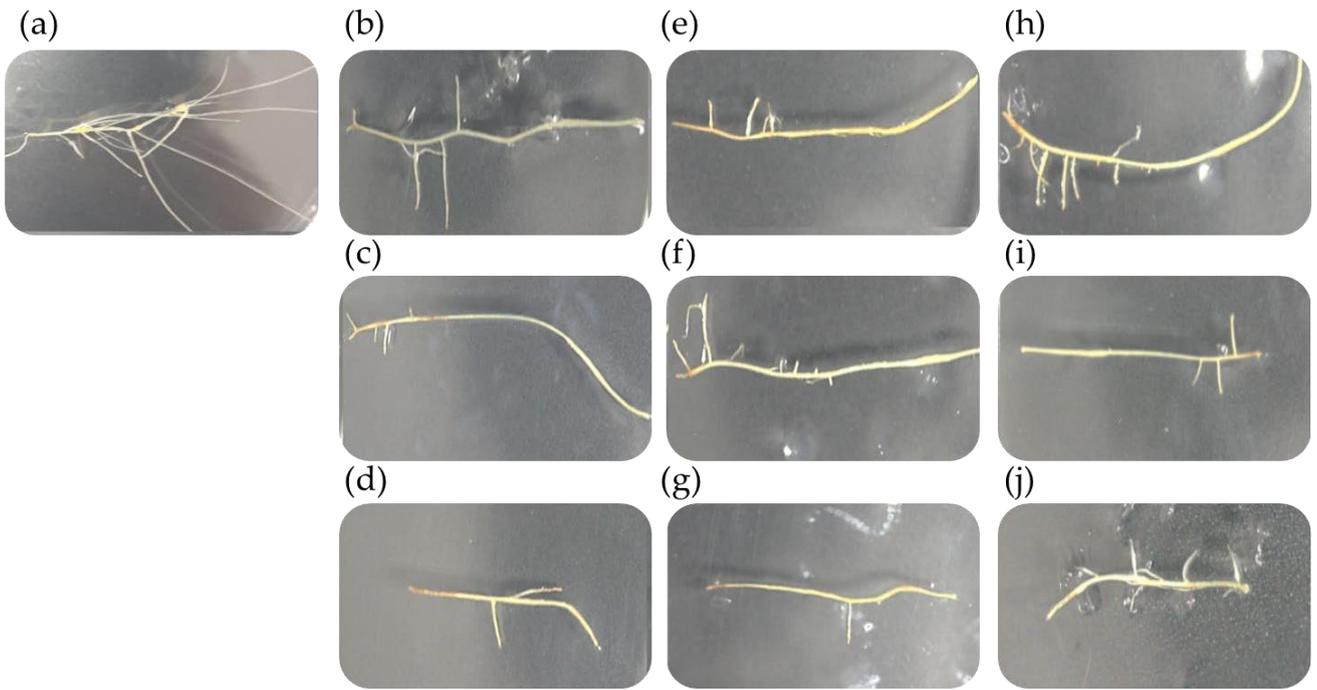
Figure S3. Morphology of different transformed root lines of *C. asiatica* on day 14 of their growth: (a) Control, (b) LST-1, (c) LST-2, (d) LST-3, (e) LS-1, (f) LS-2, (g) LS-3, (h) LT-1, (i) LT2 and (j) LT-3. The initial inoculum was a small section of 10 mg of fresh weight.

Table S1: Quantitative data for gene expression and metabolite profiling.

Type of Line	Gene expression (fold change)								Metabolites					
	HMGR		*SQS*		*At-SQS*		*SQS + At-SQS*		Centellosides content (mg/g DW)		Squalene content (mg/g dry weight)		Total Sterols (mg/g DW)	
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD
Control	1.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	5.14	0.06	1.12	0.02	141.12	0.23
LST	1.79	0.20	1.37	0.25	1.52	0.15	2.88	0.39	4.98	0.29	1.12	0.02	179.17	13.92
LST	0.73	0.09	0.96	0.08	1.50	0.04	2.46	0.07	6.56	0.26	1.56	0.10	66.20	9.12
LT	1.78	0.30	1.48	0.13	0.00	0.00	1.48	0.13	5.11	0.08	1.12	0.05	162.85	2.74

Table S2. Primers used for confirmation of transformed roots.

Gene	Primer sequence	Temperature melting (ºC)	Amplicon size (pb)	Accession number
rolC	F: TAACATGGCTGAAGACGACC R: AAACTTGCACTCGCCATGCC	60	534	KX986281.1 (*R. rhizogenes* A4)
At-SQS	F: TGGGGAGCTTGGGGACGATGC R: CGGCGTTACGGAGCTCGGTGTT	58	196	Mirjalili et al. [26] - NM_119630.4 (*Arabidopsis thaliana*)
TSAR2	F: CCCCAAATTCATCCCCTACT R: CAAGAGCTGCAAGAGCAATG	60	240	KR349466.1 (*Medicago truncatula*)
VirD	F: GGAGTCTTTCAGCATGGAGCAA R: GGAGTCTTTCAGCATGGAGCAA	56	440	KX986281.1 (*R. rhizogenes* A4)

Gene	Primer sequence	Temperature melting (ºC)	Amplicon size (pb)	Accession number
3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR)	F: CTCCTTCGTACTCCCTCGAGTC R: CCCGTCCAGTGACTTTCCAG	60	101	KJ939450.2 (*Centella asiatica*)
Squalene synthase (SQS)	F: TGAGGAGAATTCGGTCAAGG R: GCACAAAACCGGAAGATAGC	60	126	AY787628.1 (*Centella asiatica*)
beta-amyrin synthase (β-AS)	F: TTACGTTTGCTGGGAGAAGG R: TATACCCCAAGAACCGCAAG	60	134	AY520818.1 (*Centella asiatica*)
CYP716A83 (CYP83)	F: TAGCTCTGCATGCACTTTCG R: CCAGGTCCCTTTGATTTCAC	60	105	KU878849.1 (*Centella asiatica*)
CYP714E19 (CYP19)	F: AACCACACACACATCCTTGG R: TCCTTTGATGAGCCCAAGAC	60	116	KU878852.1 (*Centella asiatica*)
CYP716C11 (CYP11)	F: CCCGATTCAACGACTCTTTC R: CCGGTTGTTTGATCGACTTC	60	104	KU878852.1 (*Centella asiatica*)
UGT73AD1 (UGT)	F: TCTGGAAGCAGTTTGTGAGG R: CGCCAATCTTCACTACATCG	59	101	KP195716.1 (*Centella asiatica*)
rolA	F: GAATGGCCCAGACCTTTGGA R: TTGGTCAGGGAGGAAATCGC	60	116	EF433766.1 (*R. rhizogenes* A4)
rolB	F: CAACCGGATTTGGCCAGAGA R: ATAGGGTTGCATCGTGGTCG	60	135	EF433766.1 (*R. rhizogenes* A4)
rolC	F: CGCGCTCATCACCAATCTTC R: ACAGAAAGTGCGGCGAAGTA	60	135	EF433766.1 (*R. rhizogenes* A4)
aux1	F: TTCGAAGGAAGCTTGTCAGAA R: CTTAAATCCGTGTGACCATAG	60	350	M61151.1 (*R. rhizogenes* A4)
aux2	F: GAACCACAAACCCAAGACGC R: CTGTGGCGAAGTTGTTGCTC	60	125	M61151.1 (*R. rhizogenes* A4)
β-actin	F: TGACAATGGAACTGGAATGG R: CAACAATACTGGGGAACACT	58	80	KF699319.1 (*Panax ginseng*)

Appendixes

Other ways of disseminating the results:

- Garcia-Baeza, A. **Alcalde, M. A.**, Palazon, J. & Ramirez-Estrada, K. Induction of hairy roots in *Centella asiatica* (L.) Urban: a preliminary outcome towards the biotechnological production of anxiolytic natural compounds. 2nd International Congress of NanoBioengineering - Autonomous University of Nuevo León, Nuevo León (Mexico); 2020. *Poster*.

- Garcia-Baeza, A. **Alcalde, M. A.**, Ponce, A., Palazon, J. & Ramirez-Estrada, K. Obtaining and characterization of hairy roots of *Centella asiatica* (L.) Urban: a preliminary result towards the biotechnological production of natural anxiolytic compounds. VII National Symposium of Pharmaceutical Sciences and Biomedicine and the V National Symposium of Applied Microbiology, Nuevo León (Mexico); 2020. *Poster*.

- **Alcalde, M. A.**, Escrich, A., Bonfill, M. & Palazon, J. Plant biofactories: centelloside production in hairy root cultures of *Centella asiatica*. XVII Spanish-Portuguese Congress of Plant Biology, Vigo (Spain); 2021**.** *Poster*.

- **Alcalde, M. A.**, Escrich, A., Perez-Matas, E., Hidalgo-Martinez, D. & Bonfill, M. Effect of different elicitors on metabolic gene expression and centelloside production in *Centella asiatica* hairy root cultures. European Biotechnology Congress, Prague (Czech Republic); 2022. *Poster.*

- **Alcalde, M. A**. Use of Machine Learning to study the effects of the role genes of *Rhizobium rhizogenes* in hairy root cultures of *Centella asiatica*. XIII Research Conference of the Faculty of Pharmacy and Food Sciences, Barcelona (Spain); 2022. *Oral communication*.

- **Alcalde, M. A.**, Muller, M., Munné-Bosch, S., Landín, M., Gallego, P., Bonfill, M. & Hidalgo-Martínez, D. Enhancing insights into *Centella asiatica* Hairy Roots: Unraveling the influence of transferred *Agrobacterium rhizogenes* genes on hormone profiles and morphological traits through Machine Learning. XVIII Spanish-Portuese Congress of Plant Biology, Braga (Portugal); 2023. *Poster*.

- **Alcalde, M. A**. Perez-Matas, E., Palazon, J., Bonfill, M., Hidaldo, D. Metabolic engineering of triterpenes in transformed roots of *Centella asiatica* through overexpression of squalene synthase and the transcription factor *TSAR2*. XV Spanish Society Meeting of Plant Culture *In vitro* Culture, Lleida (Spain); 2023. *Oral communication-Poster.*

Other publications directly related with the PhD thesis:

- **Alcalde, M. A.**, Perez-Matas, E., Escrich, A., Cusido, R. M., Palazon, J., & Bonfill, M. (2022). (Review). Biotic elicitors in adventitious and hairy root cultures: A review from 2010 to 2022. Molecules, 27(16), 5253. This is a review elaborated to describe about elicitation in hairy roots, this research can amplify knowledge described in Chapter 3.

- **Alcalde, M. A**. Palazon, J., Bonfill, M. & Hidalgo-Martinez, D. (2023). Exploring the key aspects of rol genes: An overview. Phytochemistry Reviews. Under review: PHYT-D-23-00286. This is a review to describe mechanism and importance about *rol* genes, this research can increase knowledge depicted in Chapters 1, 2 and 4.

- Garcia-Baeza, A., **Alcalde, M. A**., Grovel, O., Balderas-Renteria, I., Villa-Ruano, N., Velázquez-Ponce, M., & Ramirez-Estrada, K. (2023).

Metabolic changes in hairy root cultures of *Centella asiatica* treated with methyl-jasmonate and coronatine: a 1H-NMR-based metabolomics approach. In Vitro Cellular & Developmental Biology-Plant, 1-18. This is an original research article amplying new knowledge of Chapter 3 by metabolomic analysis.

Other reviews co-authored by the PhD candidate related with the topic of the thesis:

- Khojasteh, A., Mirjalili, M. H., **Alcalde, M. A.**, Cusido, R. M., Eibl, R., & Palazon, J. (2020). Powerful Plant Antioxidants: A New Biosustainable Approach to the Production of Rosmarinic Acid. Antioxidants, 9(12), 1273.
- Perez-Matas, E., Hidalgo-Martinez, D., Escrich, A., **Alcalde, M. A.**, Moyano, E., Bonfill, M., & Palazon, J. (2023). Genetic approaches in improving biotechnological production of taxanes: An update. Frontiers in Plant Science, 14.