Study of Molecular Mechanism of Gall Development in Medicinally Important Plants Using Omics Approach

A THESIS SUBMITTED TO

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

FOR THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY

SANTHOSHA N HEGDE

UNDER THE GUIDANCE OF

DR. SUBRAHMANYA KUMAR K (GUIDE)

AND

DR. PAVITHRA N (CO-GUIDE)

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND
TECHNOLOGY, BENGALURU - 560064

APRIL, 2024

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064

DECLARATION BY THE CANDIDATE

I declare that this thesis entitled " **Study of Molecular Mechanism of Gall Development in Medicinally Important Plants Using Omics Approach**" submitted for the award of Doctor of Philosophy to THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY, Bengaluru, is my original work, conducted under the supervision of my guide **Dr. Subrahmanya Kumar K** and co-guide, **Dr. Pavithra N**. I also wish to inform that no part of the research has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged

I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

Place: Bengaluru **Signature of the Candidate**

Date: 24-04-2024

Name of candidate: SANTHOSHA N HEGDE
Reg. No.: 20219020002

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064

CERTIFICATE

This is to certify that the work incorporated in this thesis **"Study of Molecular Mechanism of Gall Development in Medicinally Important Plants Using Omics Approach"** submitted by **SANTHOSHA N HEGDE** was carried out under my supervision. No part of this thesis has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged. I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

K. Subrahmanya

Research Supervisor

Dr. Subrahmanya Kumar K

Associate professor

The University of Trans-Disciplinary Health Sciences and Technology,

Bengaluru, India

Date: 24-04-2024

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY

Private University Established in Karnataka by ACT 35 of 2013
BENGALURU - 560064

CERTIFICATE

This is to certify that the work incorporated in this thesis **“Study of Molecular Mechanism of Gall Development in Medicinally Important Plants Using Omics Approach”** submitted by **SANTHOSHA N HEGDE** was carried out under my supervision. No part of this thesis has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged. I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

Pavithra. N

Co-Supervisor
Dr. Pavithra N

Assistant professor

The University of Trans-Disciplinary Health Sciences and Technology,

Bengaluru, India

Date: 24-04-2024

Acknowledgement

I would like to express my sincere gratitude to The University of Trans-Disciplinary Health Sciences and Technology for providing me with the invaluable opportunity to pursue my research and complete this thesis.

I am grateful for the opportunity to work on this project under the guidance of Dr. Subrahmanya Kumar K. His unwavering support and mentorship have been invaluable throughout my PhD journey, and I consider it a privilege to express my heartfelt appreciation for his guidance.

I extend my sincere gratitude to Dr. Pavithra N for her invaluable guidance and unwavering support throughout the course of this project. I consider it a privilege to convey my deep appreciation for her guidance, the resources she made available, and her constant support during the project. I've been able to complete my thesis successfully due to her expert guidance.

I would also like to express my gratitude to the members of the Doctoral Advisory Committee members, namely Dr. Noorunnisa Begum, Dr. Sheetal Ambardar and Dr. Preetha Nair. I appreciate their accessibility and the enlightening discussions we've had, as well as the support they've provided over the past years.

I extend my deepest gratitude to Prof. Malali Gowda for being a pivotal role in my academic journey and for his valuable support during the initial stages of my PhD work.

I would like to record my deep sense of gratitude to lab members Dr. Lavanya DK, Dr. Prasanna Koti, Ms. Namitha, Ms. Manasa Ms. Mahalakshmi from Centre for Functional Genomics & Bioinformatics, The University of TransDisciplinary Health Sciences and Technology, Bengaluru for their moral support and kind encouragement. I want to thank you for your excellent cooperation and for all of the opportunities I was given to conduct my dissertation.

I would like to extend my heartfelt appreciation to Dr. Keshava Prasad from the Center for Systems Biology and Molecular Medicine, Yenepoya Research Centre, for his invaluable assistance in conducting the metabolomics work. His expertise and support have been instrumental in the successful execution of this research project.

I would like to extend my appreciation to Dabur India Pvt Ltd. for their generous funding support, which has been instrumental in facilitating and carrying out this work. I also want to acknowledge Mr. Pattu Raj, Dr. Amith Bhatt for their invaluable assistance in collecting samples and capturing images.

I am deeply grateful to Dr. Gurmeet Singh and Rural India Supporting Trust (RIST) for his generous financial support during the final stages of my research.

I take this opportunity to express my sincere thanks to my friends who have lent their support in various ways throughout the duration of this project. Moreover, I want to show my appreciation to every individual who, whether knowingly or unknowingly, may have contributed to the project's advancement.

Lastly, but of equal importance, I am profoundly grateful to all my family members for their continuous support and encouragement.

Dedication

I dedicate this thesis to my source of inspiration, Dr. V K Shreepad Rao - my grandfather; my mentors, Dr. Pavithra N & Dr. Subrahmanya Kumar K; and my parents, Nagesh Hegde & Kathyayini.

Contents

Chapters

List of Tables

List of Figures

Abbreviations

ABA	Abscisic acid
ASVs	Amplicon sequence variants
BUSCO	Benchmarking universal single-copy orthologs
CAMs	Cell adhesion molecules
CoA	Coenzyme A
CTAB	Cetyltrimethylammonium bromide
CTM	Chinese traditional medicine
DEGs	Differentially expressed genes
EG	Ethyl gallate
EMS	Enhanced mass spectra
EPI	Enhanced product ion
EST	Expressed sequence tags
GO	Gene ontology
HMDB	Human metabolome database
IAA	Indole 3 acetic acid
IDA	Information dependent acquisition
ITS	Internal transcribed spacer
KEGG	Kyoto encyclopedia of genes and genomes
LCMS	Liquid chromatography–mass spectrometry
NGS	Next generation sequencing
NR	Non redundant
OTUs	Operational taxonomic unit
PCR	Polymerase chain reaction
PFAM	Protein families
QC	Quality check
ROS	Reactive oxygen species
SSR	Simple sequence repeat
TDF	Triadimefon
TEs	Transposable elements
TLC	Thin layer chromatography
TM	Traditional medicine

Abstract

Pistacia chinensis subsp. integerrima (J. Stewart) Rech. f., known as *P. integerrima*, is a medicinal deciduous tree native to Asia. One interesting aspect of *P. integerrima* is the formation of galls on its apical meristem. These galls are rich in bioactives and used to treat various diseases. The exact molecular mechanism of gall formation in this species is still unknown. This study was carried out to explore the molecular mechanisms underlying gall formation on its apical meristem, a distinctive feature caused by insect infestation. A multi-omics approach encompassing genomics, transcriptomics, metabolomics, and microbiome was employed to unravel the relationship between the host plant and invading organisms.

Genomic analysis resulted in a 462 Mb draft assembly with 39,452 genes, providing insights into secondary metabolite biosynthesis. Transcriptome analysis identified differentially expressed genes, emphasizing the higher expression of phytohormone signaling pathways in galls. Also the transcripts involved in plant-aphid intearctions were highly expressed in gall. The microbiome study uncovered beneficial bacterial species such as *Aeromonas, Pseudomonas* and *Bacillus* and highlighted *Glomerellaceae* as the dominant fungal family in galls.

Metabolite profiling of *P. integerrima* and *Terminalia chebula*-one of the market substitute of *P. integerrima* galls showcased unique and shared metabolites. Enriched pathways in *P. integerrima* gall included linoleic acid metabolism and sesquiterpenoid biosynthesis. Correlation analysis linked specific bacterial genera to metabolites, covering 13 identified compounds, including serotonin and glyoxylic acid.

This comprehensive study highlights the complex interactions between microorganisms, metabolites, and host plants during gall formation. The findings offer insights into the roles of beneficial microbial communities and metabolic pathways, with potential applications in traditional medicine, plant health management, and the discovery of bioactive compounds derived from gall-associated interactions.

Publications

1. Hegde, S.N., Begum, N., Bhatt, A. *et al*. De novo genome assembly and annotation of gall-forming medicinal plant Pistacia chinensis subsp. integerrima (J. L. Stewart ex Brandis) Rech. f.. J Genet 101, 51 (2022). https://doi.org/10.1007/s12041-022-01391-w

Overview of Thesis

The study was carried out using a multi omics approach to understand the molecular mechanism of gall development in *P. integerrima*; one of the medicinally important plant species.

Chapter I: Introduction provides brief introduction to entire thesis

Chapter II: Review of literature explains the importance of medicinal plant *P. integerrima* and its galls, summarizes studied gall formation in different species and use of different omics approaches to study the formation of galls.

Chapter III: Whole genome and transcriptome sequencing of *P. integerrima* provides an in-depth exploration of various aspects related to the genome of *P. integerrima* such as assembly, annotation, metabolic pathway and phylogeny. It compiles transcriptome assembly, annotation, gene expression and pathway analysis of *P. integerrima* gall and leaf.

Chapter IV: Microbiome of *P. integerrima* gall and leaf describes the detected microbes along with their respective abundance, diversity, and corresponding microbial genes and metabolic pathways associated with gall and leaf of *P. integerrima*.

Chapter V: Global metabolite profiling of gall and leaf of *P. integerrima* and *T. chebula* explains metabolites associated with *P. integerrima* gall and as well as specific pathways and their involvement in gall formation.

Chapter VI: Correlation of data from different omics approaches involves the correlation analysis of different omics approaches and statistical analysis.

Chapter VII: Conclusion concludes the thesis with scope for further research on the development of *P. integerrima* galls.

CHAPTER I

INTRODUCTION

This chapter provides a general introduction/background of entire research work in this thesis

1. Introduction

1.1. Medicinal Plants

The importance of natural products in the treatment of diseases is well known. Amongst the various natural sources, plants are an important source of bioactive constituents, including anticancer, antifungal and antimicrobial drugs (Bibi *et al*, 2012). Since at least 60,000 years, plants are utilized as medicines, demonstrating their capacity to create mixtures of secondary metabolites with a variety of medicinal uses (Buyel, 2018). The earliest known health care items utilized by humans are medicinal plants, which are also important ingredients in many of the formulations used in indigenous systems of medicine (Tandon *et al*, 2017). These indigenous systems of medicines also known as Traditional Medicine (TM) include medical procedures, methods, theories, and practises that use drugs derived from plants, animals, and minerals as well as spiritual therapies, manual therapies, and exercises. These methods may be used separately or in combination to treat, diagnose, prevent, or maintain health (Alves *et al*, 2007; Fokunang *et al*, 2011). Medicinal plants are used in about 90% of traditional medicine formulations and treatments (Sofowora *et al*, 2013).

Around the world, hundreds of higher plants are cultivated to provide valuable ingredients for medicine and pharmacy. The therapeutic properties of plants have led to the development of medicinal drugs (Salmerón-Manzano *et al*, 2020). The global market for herbal medicines (medicine derived from medicinal plants) was estimated by Fortune Business Insights to be at USD 151.91 billion in 2021. According to projections, the market for herbal medicines would increase from $165.66 billion in 2022 to $347.50 billion in 2029 (https://www.fortunebusinessinsights.com/herbal-medicine-market-106320).

Medicinal properties derived from plants can come from many different parts of a plant including leaves, roots, bark, fruit, seeds, flowers, galls (Prasad *et al*, 2010). The different parts of plants may contain different active ingredients within one plant.

1.2. The Plant Galls

Galls, the abnormal growths in plants, induced by viruses, bacteria, fungi, nematodes and arthropods. However, each type of gall-producer is specific to a particular kind of plant (Patel *et al*, 2018). Galls exhibit special geometrical shapes that are ordinarily absent in the typical plant system. Galls are the greatest instances of naturally changed formations that only appear as a result of communications from foreign organisms (Raman, 2011). Nematodes, insects, and mites are the typical causes for galls, whereas bacteria, fungi, and viruses account for a very minor portion of all cases. There are thousands of species of insects in the world that induce gall formation on the roots, stems, leaves, buds, flowers and fruits of plants in a wide variety of plant families (Peck *et al*, 2008; Figure 1.1). Galls may appear as balls, knobs, lumps, or warts, each being characteristic of the causal organism and the host plant. In addition to the unusual structure of galls, they draw attention due to their range of colors: red, green, yellow, or black and structures: oval, curved, round, horn shaped, *etc*. Factors such as weather, plant susceptibility, and pest populations affect the occurrence of galls on plants from year to year or region to region.

The galls, occur are phytochemically-distinct from the normal plant tissues. The plant family characteristics and biochemistry of gall has shown, galls are rich in bioactive phytochemicals (Figure 1.1). Triterpene, gallic acid, ethyl gallate, catechin, epicatechin, tannic acid, resin, and a variety of other phytochemicals are among the substances generally found in most of the galls (Patel *et al*, 2018). Mankind has learnt to employ the galls as medicine during the lengthy evolutionary history, just like other plant components. For example, *Quercus infectoria* galls are frequently utilized in Malay traditional medicine for managing post-childbirth wound infections. In India, they have also been traditionally employed for dental purposes, including the treatment of toothaches and gingivitis (Basri *et al*, 2012). In Ayurveda, the powdered form of *P. integerrima* galls is employed either on its own or in conjunction with other herbal ingredients for medicinal purposes. This gall powder is commercially available as Shringyadi Churna and is used in the management of asthma. Furthermore, Chyawanprash containing *P. integerrima* galls powder is utilized for its digestive, bronchodilator, and antitussive properties (Kadam *et al*, 2023; Alhat and Kulkarni, 2023). Common plant galls used for medicinal purposes include *Terminalia, Quercus*, *Pistacia*, and *Rhus,* among others. (Patel *et al*, 2018).

1.3. *Pistacia chinensis* subsp. *integerrima* (J. Stewart) Rech. f.

Pistacia chinensis subsp. *integerrima* (J. Stewart) Rech. f. (hereafter called as *P. integerrima),* commonly called as zebrawood is one of the most important ethnomedicinal plant species; traditionally used in the treatment of coughs, phthisis, jaundice, antiseptic, chronic wounds, asthma and dysentery (Chopra *et al*, 1986; Ahmad *et al*, 2020). *P. integerrima* belongs to the *Anacardiaceae* family indigenous to India and is found in the outer ranges of the North-Western Himalayas at an altitude of 500 to 2500 m (Kaur and Singh, 2015). It is a moderate, branched tree that lives in dry slopes with shallow soils and may reach heights of up to 25 meters (Ahmad *et al*, 2020; Ismail *et al*, 2011; Figure 1.2).

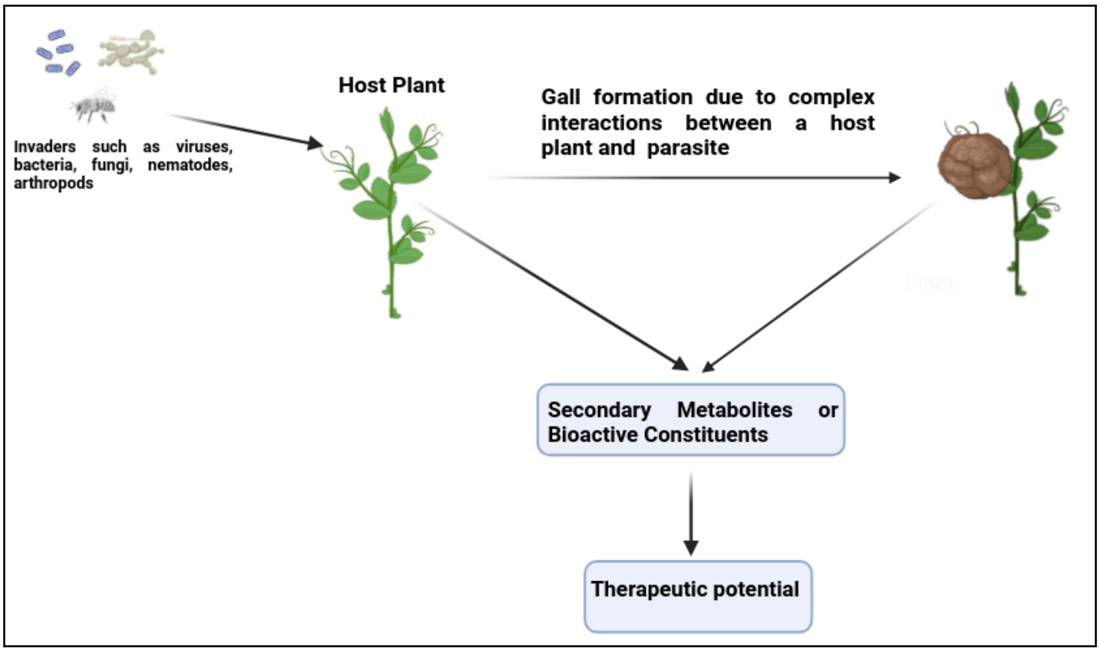


Figure 1.1: Plant galls, fascinating structures induced by a range of organisms, have developed impressive adaptations, such as distinctive metabolic compositions and potential therapeutic uses. Notably, galls are abundant sources of secondary metabolites, rendering them valuable for addressing human health issues.

Figure 1.2: ***P. integerrima*** **tree**

1.4. *P. integerrima* Galls

P. integerrima serves as an illustrative example of a plant renowned for its diverse array of properties, which researchers have been diligently investigating and characterizing. The activities vary from antibacterial to anticancer (Bibi *et al*, 2015). *P. integerrima* is distinguished by the rough, horn-shaped, rugose, hollow galls that grow on the plant and store most of its metabolites. These galls grow on the plant's apical buds of *P. integerrima* (Figure 1.3) as a result of aphid *Baizongia pistaciae* L. infection. These gall forming insects follow a two year life cycle. First instar nymphs (fundatrices), which emerge from overwintering eggs in the spring, cause galls. Two to three generations are parthenogenetically generated within each gall. Flying aphids leave the galls in the winter, and their young grow on the roots of the secondary hosts. Another flying morph (sexuparae) migrates back to the main host the following spring to give birth to males and females. One year after mating, the fertilized eggs are left on the host, where the fundatrices emerge (two-year life cycle) (Inbar *et al*, 2004; Wool, 2012; Wool, 2013).

These horn-shaped galls grow as a result of an insect infestation (Zahoor *et al*, 2018). The galls of *P. integerrima* are highly appreciated in Ayurvedic system of medicine and called as “Karkatshringi.” The galls of this plant included in a variety of Ayurvedic preparations, including "Dasamularista," "Chayavanaprasa," "Shringyadi churna," "Balachaturbhadra churna," *etc* for treatment of different diseases such as asthma, phthisis, and other respiratory tract illnesses, dysentery, chronic bronchitis, hiccough, child-inflicted vomiting, skin conditions, psoriasis, fever, snake bite, *etc*. (Uddin *et al*, 2011; Kaur and Singh, 2015).

Figure 1.3: *P. integerrima* gall. A - Green gall; B - Dry gall

The galls have been shown to contain secondary metabolites such as alkaloids, tannins, terpenoids, and flavonoids. Minor components such as crystalline hydrocarbon, gum mastic, resin, and crystalline acids are also present in addition to these. Terpenoids and flavonoids are present in the bark. The galls are enriched with pistagremic acid, serving as a natural terpene inhibitor targeting β-secretase activity (Kaur and Singh, 2015; Kadam *et al*, 2023).

Hence, this plant contains various important phyto-constituent of commercial value and therapeutic potential and these galls are used in traditional medicines in India with different reported activities as antispasmodic, carminative, antiamoebic and anthelmintic potential (Warrier *et al*, 1995; Rauf *et al*, 2015).

1.5. Plant Gall Formation

An abnormal growth of tissue and altered external morphologies may result in the development of tumor-like galls in plants which have been infected by viruses, bacteria, or insects. The type of induced gall morphologies depends on the feeding preferences of the various insect species (Nabity *et al*, 2013). The phenomenon of certain bacteria changing hosts is relatively well-documented, whereas our understanding of how insect gall formers establish and maintain genetic control over plant growth remains less explored. However, research has revealed that gall formation involves more complex interactions than just a parasitic or defensive connection between a gall-inducing insect and its host plant (Gatjens-Boniche, 2019; Chen *et al*, 2020).

Complex insect galls produce a specific "nutritive tissue" on which the insect feeds on and are very divergent from the plant tissues on which they form (Schultz *et al*, 2019). Gall forming insects manage their host to get a better source of nutrients as well as chemical and physical defense from abiotic and natural enemies (Kurzfeld-Zexer and Inbar, 2021). However, according to certain studies, the interaction between gall inducers and host plants is not only a straightforward parasitic/defensive relationship (Chen *et al*, 2020).

Gall initiation and development are manifested by the cell hypertrophy and tissue hyperplasia, as well as by changes in the chemical composition (Zorić *et al*, 2019). The development of galls is mediated by the following mechanisms: (1) release of signalling molecules from insects or microbes; (2) plant processing of the signals; (3) proliferation and division of plant cells; and (4) structuring of gall tissue. In these procedures, insects must prevent the plant's defense responses (Giron *et al*, 2016). Insects that cause gall production initiate triggers that start the growth and maintenance of gall tissue. Gall-inducing insects trigger specific developmental pathways in their host plants, exerting significant influence over these pathways. This is evident from the consistent observation that each bug species tends to produce a distinct gall form on its respective host plant. (Takeda *et al* 2021). Different hypotheses for the adaptive significance of gall induction have been proposed by shape and cell status of galls: nutrient hypothesis, microenvironment hypothesis, and enemy hypothesis (Stone *et al*, 2003; Takeda *et al* 2021; Gatjens-Boniche *et al*, 2019). According to the nutrition hypothesis, galls offer superior nutrition to alternative feeding methods, and improved nutrition is generally acknowledged as a benefit of gall induction. According to the microenvironment theory, the galler is protected from harmful abiotic conditions, notably desiccation, by the gall tissues. According to the enemy concept, galls shield gallers from attack by their own natural foes. Galls do offer some defense against diseases and non-specialist predators (Stone *et al*, 2003).

1.6. Micro-organisms in Gall Formation

Bacteria, fungi, and viruses only cause a very tiny fraction of galls, and *Agrobacterium tumefaciens* is widely known for producing crown galls (Mapes, 2008). Different studies suggest that phytohormones and growth regulators such auxins, cytokinins, indole-3-acetic acid (IAA), and others are produced by the gall inducer and serve as the catalyst for gall formation (Gatjens-Boniche, 2019). Faist *et al*, described the microbiota of *Vitis vinifera*'s crown gall and reported the bacteria that produce IAA, including *Pseudomonas* sp. and *Enterobacteriaceae* sp (Faist *et al*, 2016). The capacity of many phytopathogenic bacteria to generate galls is linked to the creation of cytokinins; these substances, when present with auxin, cause plant tissue to divide and multiply, resulting in the development of galls or tumors (Klimov *et al*, 2022). *Agrobacterium tumefaciens*, which is referred to as "promiscuous" in the world of bacteria since it infects plant species belonging to at least 60 families and 140 genera (Harris and Pitzschke, 2020).

Previous studies have given some useful information about how galls form in plants, which is important in plant biology. However, it is crucial to recognize that this data serves as an initial foundation. There is a lack of substantial knowledge persists when it comes to conducting an in-depth examination of this process of gall formation. It is essential to use advanced scientific methods like genomics, transcriptomics, proteomics, and metabolomics to gain an in-depth understanding of this complex process. This kind of in-depth investigation is crucial to understand how galls develop in medicinal plants. Furthermore, limited information is available regarding the specific genes and chemicals responsible for regulating gall formation in these plants.

1.7. Need for the Study

The study of plant galls is very important due its high medicinal value, high demand and low supply, need to conserve the galls by understanding the mechanism behind the gall formation, and there is not much information about molecular mechanism of gall formation. Since the plant galls have the high medicinal properties it is very essential to conserve the plant galls. The formation of galls in *P. integerrima* has decreased dramatically in recent decades, potentially as a result of climate change or human influence, making it an essential species for conservation. The demand for gall is high, but supply has decreased adequately. To conserve the galls of this plant, it is necessary to understand the molecular mechanism of gall formation. Comprehending the molecular mechanisms of gall formation is essential for conservation as it paves the way for targeted strategies aimed at boosting gall production and mitigating the supply decline. This knowledge can inform genetic interventions, cultivation practices, and habitat management efforts to create ideal conditions for host plants and gall-inducing insects, ensuring the sustainable availability of gall resources.

Since gall development is a complex process, the molecular and cellular processes of gall formation are still poorly understood. The overall process by which the insect might control and change plant growth and physiology is still not fully known. Similarly, we have yet to completely understand how various chemical compounds or plant hormones work in this process. The creation of innovative, effective management approaches is hampered by the dearth of knowledge on the molecular processes underlying interactions between plants and galling insects. The advent in recent technologies have enabled the study of plant galls through different perspectives.

Different parts of *Terminalia chebula* Retz. (*Combretaceae*) possesses numerous medicinal properties due to its abundant variety of phytoconstituents. *T. chebula* galls (Figure 1.4) have a wide range of ethno-botanical and diverse pharmacological applications (Eshwarappa *et al*, 2016). Ayurvedic literature, “Yogratnakar” points out that *T. chebula* possesses similar therapeutic attributes to *P. integerrima* (Saini *et al*, 2023). Galls from *T. chebula* are used as a substitute for *P. integerrima* in South Indian markets (Ravi Shankara *et al*, 2012). Therefore, given the common therapeutic characteristics and the utilization of *T. chebula* galls as a replacement for *P. integerrima* in South Indian markets, it is essential to undertake a thorough metabolite profiling analysis of both plant species. This analysis will shed light on the chemical constituents responsible for their respective medicinal properties and can offer valuable insights into their potential applications in traditional medicine formulations.

Over the past decade, the use of omics technologies has been a powerful tool for enhancing our understanding of interactions between plant-galling insects. Multiple omics technologies, such as genomics, transcriptomics, proteomics, and metabolomics, are very important to study the formation of plant galls.

Although studies have revealed disparities in the expression of genes involved in hormone synthesis and response, and differences in hormone levels, the regulatory roles of hormones in numerous interactions between plants and insects remain unclear. The integration of multi-omics data will play a crucial role in uncovering the molecular mechanisms involved.

Figure 1.4: ***T. chebula* galls**

Additionally, there is scarcity of studies examining the role of microbial association in the life cycles of gall-inducing insects. The existence of symbiotic relationships between these insects and microbes with regards to gall formation, plant utilization, and the evolution of galling behavior has been proposed. The impact of microbes on plant-galling insect interactions is undeniable, but the underlying mechanisms in this tripartite relationship remain largely unknown (Oates *et al*, 2016). Multi-omics approaches provide a comprehensive view of the molecular changes that occur during gall formation, from the genome to the proteome and metabolome. For example, transcriptomics can be used to study changes in gene expression, proteomics can be used to study changes in protein abundance and function, and metabolomics can be used to study changes in the plant's metabolic pathways.

By combining these omics technologies, a deeper understanding of the complex biological processes involved in plant gall formation can be obtained. They can identify the genes and metabolic pathways that are affected by the gall-inducing organisms thus contributing to the gall formation. This information can be used to develop new strategies for controlling gall-inducing organisms and improving plant health. Multi-omics approaches in plant gall formation allow researchers to study the complex interactions between plants and gall-inducing organisms at the molecular level, providing new insights into the biology of plant gall formation.

This study aimed to investigate the molecular mechanism of gall formation in *P. integerrima*, a significant medicinal plant species. The approach employed integrated omics techniques, encompassing genomics, transcriptomics, metabolomics, and microbiome analyses. It is essential to use an integrative strategy that incorporates multi-omics data to emphasize the inter relationships of the implicated biomolecules and their activities in order to understand complicated biological processes holistically. These omics techniques will help in understanding the gall formation at molecular level such as the expression of genes in gall, the enriched biochemical pathways in gall, the community of microbes associated with gall formation, and different chemical compounds (metabolites) which are involved in gall formation. Due to the quick production of massive volumes of *de novo* systems biology data, next-generation omics techniques make it easier to analyze non-model species, which makes them appealing choices for researching poorly characterized interactions (Oates *et al*, 2016). Multi-omics techniques have greater predictive power when identifying important targets for functional testing because of their integrated character.

In this study a combined approach of four different multi-omics approaches such as genomics, transcriptomics, metabolomics and microbiome are integrated and correlated the data from each omics approach to understand the molecular basis of gall formation (Figure 1.5).

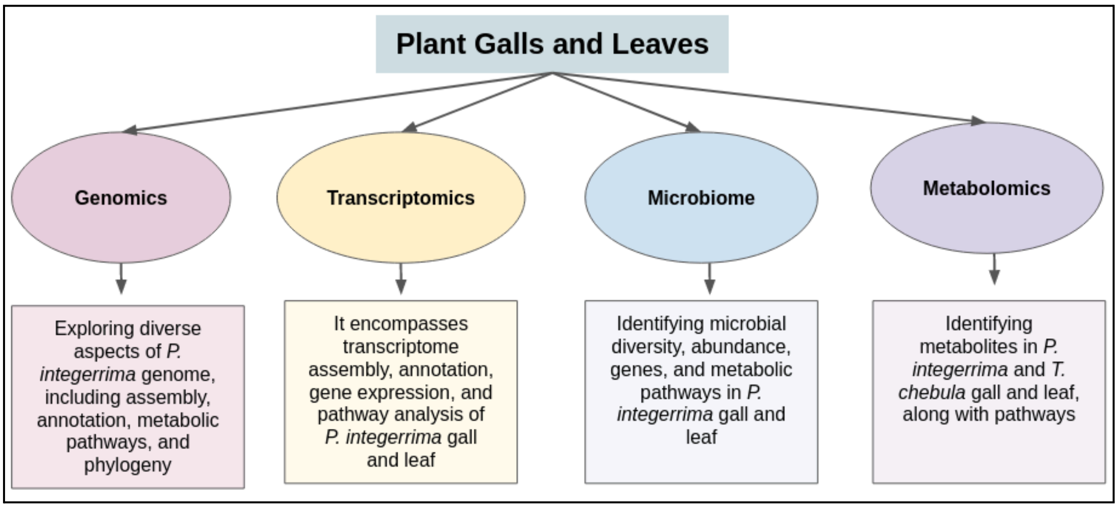


Figure 1.5: Omics approach to study formation of medicinally important plant galls

1.8. Objectives

The overall aim of the study is to understand the molecular mechanism behind the formation of galls in *P. integerrima*. The objectives of the study are;

1. Determination the function of genes and the elements that regulates the gall formation in *P. integerrima* using whole genome and transcriptome sequencing

2. Identification the potential microbes associated with gall in comparison to leaves of *P. integerrima*

3. Study the metabolite diversity in galls and leaves of *P. integerrima* and *T. chebula*- one of the market substitutes of *P. integerrima*

CHAPTER II
REVIEW OF LITERATURE

This chapter summarizes the importance of medicinal plant *P. integerrima* and its galls and the use of different omics approaches to study the formation of galls in plants.

2. Review of Literature

2.1. Plant Galls

Galls are distinctive plant formations that are created when parasitic or pathogenic organisms cause cell size growth and/or cell multiplication. Galls are frequently brought on by insects, and they can appear on plant leaves, stems, floral buds, flowers, fruits, or roots (Takeda *et al*, 2021). Insects that cause gall production create triggers that start the growth and development of gall tissue. The gall-inducing bugs produce more phytohormones, such as indole 3 acetic acid (IAA) and cytokinins, than host plant galls (Tooker *et al*, 2011; Yamaguchi *et al*, 2012; Andreas *et al*, 2020). It has been found that several phytohormones have a role in gall formation, host modification, and plant response to galling insects. Auxins and cytokinins are well-known growth regulators that have been frequently linked to the formation of insect galls. It is commonly believed that these hormones play a role in the interactions between plants and galling insects (Oates *et al*, 2016).

Plant galls, which are triggered by arthropods and other organisms, share a close relationship with their host plants. The organisms inducing these galls typically have limited mobility. Galls, besides conducting their own photosynthesis, act as nutrient-rich resource sinks, often relying on nearby plant organs as external sources of photosynthates. However, if galls lack effective defenses, they can become appealing food sources for herbivores (Yamazaki, 2016). Insects that induce the formation of galls manipulate their host plants to secure improved access to nutrients and to gain physical and chemical protection from natural predators and environmental factors (Kurzfeld-Zexer and Inbar, 2021). Arthropod-induced galling prompts the host plant to produce a variety of phytochemicals that it wouldn't typically generate under normal circumstances. However, this metabolic shift comes with a cost. These secondary metabolites are intended to serve as a defense mechanism against the invaders. The agents causing the galls disrupt the plant's auxin and cytokinin pathways (Patel *et al*, 2018).

Patel *et al*, 2018 listed out the gall inducers as well as respective host plants and is given in Table 2.1.

Table 2.1: A compilation of organisms causing galls and the plants they impact (Patel *et al*, 2018)

Galling Agent	**Scientific Names**	**Affected Plants**
Bacteria	*Agrobacterium tumefaciens*	Spindle tree, alfalfa, grapevine, blueberry, mum
Rust Fungi	*Gymnosporangium juniperi-virginianae*	Pine, cedar, cypress
Exobasidium	*Exobasidium rhododendri, Exobasidium camelliae*	Rhododendron, camellia
Ascomycota	*Taphrina pruni, Taphrina alni, Taphrina betulina, Endoconidiophora polonica*	Peach, blackthorn, alder, spruce
Ustilago	*Ustilagosp.*	Corn, rice, barley, oat, wheat, sorghum, sugarcane
Wasp	*Diplolepis rosae, Diplolepis polita, Diastrophus nebulosus, Diastrophus rubi, Callirhytis seminator, Andricus flocci, Andricus gigas, Neuroterus numismalis, Loxaulus maculipennis, Leptocybe invasa*	Pistachio, rose, oak, raspberry, blackberry, currant, rosinweed, quercusbaccarum, maple, lantana, eucalyptus, sweet chestnut, rosinweed
Mite	*Phytoptus padi, Eriophyes tiliae, Eriophyes fraxinivorus, Colomerus vitis, Cecidophyopsis betulae, Aceria aceris, Aceria lantanae*	Lime, linden, ash, grapevine, birch, maple, lantana
Midge	*Taxodiomyiacupressiananassa, Dasineura oxycoccana, Asphondyliasp., Mayetiola destructor, Lasioptera arundinis, Schizomyia galiorum, Kieffer*	Cypress, blueberry, creosote bush, common reed, hedge bedstraw, fern, goldenrod, thistle
Thrips	*Kladothrips maslini, Acacia*	Acacia

Fly	*Chirosia grossicauda, Eurosta solidaginis, Urophora stylata*	Fern, goldenrod, thistle
Scale Insect	*Apiomorpha munitatereticornuta*	Eucalyptus
Psyllid	*Pachypsylla celtidismamma, Calophya schini*	Hackberry, pepper tree
Aphid (Phylloxera)	*Hormaphis hamamelidis, Tamalia coweni, Tamalia dicksoni, Melaphis rhois, Colopha ulmicola, Slavum wertheimae, Daktulosphaira vitifoliae, Phylloxera caryaecaulis*	Witch-hazel, manzanita, sumac, elm, wild pistachio, grapevines, hickory
Sawflies	*Phyllocolpasp., Pontaniasp., Euurasp.*	Willow
Fly	*Chirosiabetuleti*	Ferns
Adelgid	*Adelges cooleyi*	Spruce
Bark Beetle	*Ips typographus*	Spruce
Weevil	*Bruchus pisorum*	Pea
Leafhopper	*Cicadulina bipunctata*	Maize
Parasite	*Pedisapis aceris*	Maple

2.1.1. Types of Galls

Galls can be classified based on different factors such as mechanism of formation, size, shape, gall inducers *etc*. Galls vary in complexity, ranging from relatively simple depressions or folds to formations where the galler is completely enclosed by plant tissues. Enclosed galls can vary widely in complexity. Some are relatively simple, with minimal variation within the same galler group, as seen in the galls induced by fig wasps and yucca moths. On the other hand, there are more intricate structures composed of numerous distinct layers of plant tissue, exhibiting significant diversity even within a specific galler group (Stone and Schönrogge, 2003).

Galls are classified based on their morphology, and they can be categorized into three main types. First, there are Filz galls, which consist of extensive epidermal hairy outgrowths and are often caused by external gall makers like eriophyid mites. Second, Fold and roll galls involve the folding or rolling of the leaf blade margin, accompanied by swelling of the affected area, typically induced by gall-inducing thrips. Lastly, Pouch galls result from an intense out-arching of the leaf blade, forming a bulge on one side and an invagination on the other. These galls are hollow and contain the gall-inducing organisms, with the externally situated organisms inducing the growth of pathological tissue around them, eventually enclosing them within the gall (Sahu *et al*, 2020).

Galls can be categorized based on the site of infection into various types. Blister galls are characterized by blister-like swellings on leaves, while Bud galls cause deformities in buds. Bullet galls are nearly solid, resembling bullets. *Cecidomyia* is a general term applied to any species related to gall midges. Erineum galls appear hairy or pile-like due to plant mites. Flower galls result in deformed flowers or masses of flowers, and Fruit galls cause deformities on fruits and seeds. Leaf galls lead to deformations in leaves, while leaf spots are marked discolorations rather than swellings or deformations. Oak apples refer to large galls on oaks of the genus *Amphibolips*. Pouch galls are simple, pouch-like deformities, and Roly-poly galls are loose, usually oval cells with a large cavity. Root galls appear on plant roots, Rosette galls form on bud tips, often consisting of a central cell surrounded by a rosette of partly developed leaves. Stem or twig galls cause deformations on twigs and stems, and Subcortical galls develop just under the bark, usually on one side of the stem or twig and have irregular shapes (Sahu *et al*, 2020; Ananthakrishnan, 2007).

2.2. Mechanism of Plant Gall Formation

Takeda *et al* studied the molecular mechanism of gall formation in four different host plants namely; *Glochidion obovatum, Eurya japonica, Artemisia montana, Rhus javanica* using comparative transcriptome analysis. The study indicates that in order to create certain morphology, galls are compelled to activate the genes that were initially involved in other many cellular mechanisms. Further authors identified 38 frequently up-regulated genes that may contribute to the emergence of more leaf galls. From this study, authors reported differentially expressed genes which involved in different biological, cellular as well as molecular processes/functions such as response to stress, cell division and cytokinesis, lignification and reactive oxygen species (ROS) generation, phytohormone signaling and cell regeneration, metabolic processes, *etc* (Takeda *et al*, 2021).

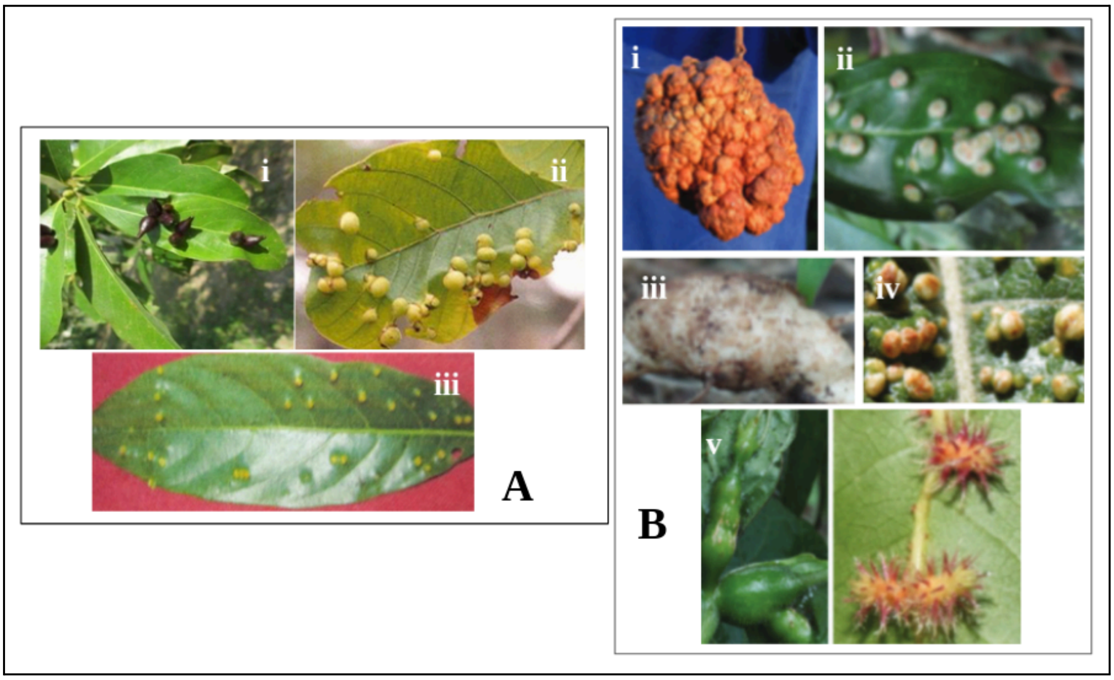


Figure 2.1: Different types of plant galls. A - Galls from different plants; i: *Persea bombycina*; ii: *Litsea monopetala* Roxb.; iii: *Litsea salicifolia* (Adopted from Sahu *et al*, 2020). B - Galls formed due to infestation of different invaders; i: Bacteria; ii: Fungus; iii: Nematodes; iv: Mites; v: Insects (Adopted from Gatjens-Boniche, 2019)

Shih *et al* conducted the transcriptome study of cup-shaped galls and its host leaf of *Litsea acuminata* to understand the mechanism of genetic regulation involved in the gall formation. The key metabolism-related genes, such as photosynthesis, cell wall reorganization, and sugar breakdown, expressed differentially in galls and leaves. A possible function for steroid hormones in controlling gall growth has been suggested by the analysis of gene expression, which revealed the involvement of genes in brassinosteroid production and responses showed a noteworthy modulation in cup-shaped galls from the study. The genetic responses of galls produced by a *Dipteran* insect, including those involved in reallocation of photosynthetically active parts of a plant and the areas of active growth and areas of storage (source-sin) and phytohormone metabolism, were disclosed in this work (Shih *et al*, 2018).

Li *et al* revealed the symbiotic mechanism of *Ustilago esculenta* induced gall formation of *Zizania latifolia* using triadimefon (TDF) treatment and RNA sequencing technology. A total of 17,541 differentially expressed genes (DEGs) were identified, based on the transcriptome in this study. Further, authors reported the significant enrichment of pathways such as plant hormone signal transduction, and cell wall–loosening factors. The relative expression levels of genes associated with hormones and downregulation of DEGs associated with indole 3-acetic acid (IAA) were discovered. The LCMS study revealed the IAA, zeatin+trans zeatin riboside, and gibberellin 3 were increased under *U. esculenta* infection (Li *et al*, 2021).

Nabity *et al* studied the reconfiguration of host metabolism and morphology of host plant *Vitis* spp. infected by *phylloxera*. Authors reported the morphological and functional characterization of insect-induced stomata. Further, they demonstrated that *phylloxera* may cause plants to develop functional stomata and that this insect alters the transcription of galler-induced tissue to include downstream secondary processes in addition to primary metabolism using gall transcriptome analysis. Around 2,750 genes in 15 functional categories; including photosynthesis, fermentation, and secondary metabolism were differentially expressed (DE) in galled leaf tissue. In comparison with ungalled leaf tissue, the expression levels involved in sucrose mobilization, glycolysis, and fermentation enhanced whereas genes involved in light harvesting and photosynthetic carbon absorption showed a marked drop. Additionally, the number of transcripts for water, oligopeptide, and amino acid transporters were enhanced (Nabity *et al,* 2013).

Schultz *et al* investigated how plant productive mechanisms are activated during gall development in *Vitis riparia*. They discovered that the expression of numerous potential genes associated in floral development was dramatically elevated, especially in later phases of gall formation, and that reproductive gene ontology groups were substantially enriched in growing galls. The transcriptome sequencing was carried out at four stages of galls as they developed. They discovered 11,049 transcripts with differential expression at least once in galls compared with ungalled leaves. As the gall grew, there were significantly more transcripts expressed differently (DEGs) in the galls compared to the leaves. They also noticed that when leaves and galls evolve, their transcriptomes vary more and more, starting at the initial stages of gall formation (Schultz *et al,* 2019).

Hirano *et al* studied the modifications of the developmental process of *Rhus javanica* at the early stages of gall induction by *Schlechtendalia chinensis*. They noticed that several functional domains, including plant hormone metabolism and signaling, stress-response pathways, secondary metabolic pathways, photosynthetic responses, and floral organ development, had drastically different expression patterns. Master transcription factors that control the growth of meristems, flowers, and fruits as well as genes that respond to biotic and abiotic stress were all significantly upregulated, while the expression of genes involved in photosynthesis was significantly downregulated in the early stages of gall development (Hirano *et al,* 2020) .

Gatjens-Boniche listed four basic explanations/hypotheses that may account for the development of plant galls. The primary of these theories proposes that gall induction could be mediated by the insertion of an insect secretion during the egg laying process. According to a second theory, the placement of a foreign substance on the plant tissue causes mechanical stimulation that leads to gall development. According to the third theory, the insect's saliva contains active ingredients that are secreted, which causes the galls to develop. A fourth theory contends that the discharge of biochemical byproducts from the insect is a mediator in the production of galls. Further, he reported that phytohormones such as auxins, cytokinins, indole-3-acetic acid (IAA), and other types of compounds are involved in the formation of gall. Also he mentioned it is yet unknown how these chemicals work and the overall method by which the insect might regulate and alter plant physiology and growth (Gatjens-Boniche, 2019).

Chen *et al* reported genes that are involved in plant-insect interactions and the molecular processes of tannin buildup in Rhus galls. In this study, *S. chinensis* caused horned galls were collected once per fifteen days from each embryonic stage. The concentration of tannins at different stages of gall development was measured and transcriptome sequencing was carried to identify the expressed genes involved in tannin biosynthesis or tannin accumulation inside the gall tissue. They observed that the amounts of hydrolyzable tannin in galls grew from gall initiation through gall maturity, and then progressively dropped after that. They also discovered 22 genes (TS1-22) for condensed tannin biosynthesis and a total of 81 genes (GTS1-81) with probable functions in gallotannin production (Chen *et al,* 2018).

In an assault on *Rhuschinensis Mill* (member of *Anacardiaceae* family) by the aphid *Schlechtendalia chinensis*, Wang *et al* showed the molecular reaction of horned gall induction. They sequenced the transcriptomes of four tissues, including galls, leaves that grew on the same branch as the gall that was removed, leaves from a branch without any galls, and leaves from a tree without any galls, using Illumina deep sequencing and digital gene expression. It was observed that In contrast to leaves (leaves that grew on the same branch of the gall and leaves from a branch without any gall), differentially expressed genes that were substantially enriched in the manufacture of bioactive molecules, interactions between plants and aphids, and signal transduction for phytohormones were strongly expressed in galls. Further it was found that, After an aphid assault, primary and secondary metabolism are coordinated by phytohormone signal transduction, which is controlled by Indole-3-Acetic Acid (IAA) and Abscisic Acid (ABA), and this causes gall formation and development (Wang *et al*, 2017).

2.3. Medicinally Important Plant Galls

Galls, found in a wide array of shapes, exhibit distinct phytochemical compositions compared to regular plant tissues. Over the course of evolution, humans have discovered the therapeutic properties of galls, much like they have with other parts of plants. Across various cultures, there is abundant evidence of the traditional use of galls. Galls from plant genera such as *Rhus, Pistacia, Quercus, Terminalia*, and more are well-known in the field of ethnomedicine. Evidence from folklore and pharmacopeias attests to their utilization in traditional remedies across various cultures, including Chinese, Greek, and Indian traditions. Plant galls serve as both household ingredients and medicinal substances (Patel *et al*, 2018).

Below is a compilation of important gall-producing plants and the traditional medicinal purposes of their galls.

*Quercus infectori*a (*Q. infectoria*), a member of the *Fagaceae* family, is a small shrub that is typically located in Greece, Asia Minor, and Iran. It is commonly referred to as gall oak. Galls produced by *Q. infectoria* have been recognized for their medicinal potential, exhibiting properties including astringency, anti-inflammatory, antiviral, antidiabetic, larvicidal, antibacterial, antiulcerogenic, and gastroprotective activities (Shrestha *et al*, 2014). *Q. infectoria* galls have been used by Malay women for post-partum and health benefits, while they are also employed in Thailand for stomach issues and in India for oral care and toothache relief, due to their tannin (50–70%; Mahboubi, 2020) content with astringent properties. These galls have a versatile history of traditional medicinal uses (Zin *et al*, 2019)

Rhus chinensis, a species within the *Rhus* genus, is known for its medicinal properties, with a focus on the galls found on its leaves, which have been traditionally used in Asia for preventing and treating various ailments, including diarrhea, dysentery, cancer, diabetes, sepsis, and oral inflammation (Djakpo and Yaro, 2010). Yang *et al*, reported the antidiarrheal property of galls of *Rhus chinensis* (Yang *et al*, 2017).

Terminalia chebula, belonging to the *Combretaceae* family, is another plant known for producing galls and is extensively utilized in traditional Indian and Iranian medicine to address conditions such as dementia, constipation, and diabetes (Jokar *et al*, 2016). The leaf gall extract of *T. chebula* was tested for its cytotoxicity on various cell lines, with the highest effectiveness observed against A-549 (Human lung cancer) cells, displaying an IC50 range of 208 to 643 µg/mL (Ravi Shankara *et al*, 2016).

Pistacia integerrima, a significant gall-producing plant species, holds great importance in traditional medicinal systems such as Ayurveda, Unani, Sidda, and Chinese Traditional Medicine (CTM). The galls of this plant are renowned for their storage of various secondary metabolites, including steroids, flavonoids, tannins, saponins, and phenols (Kanade and Awalaskar, 2021; Kaur and Singh, 2015).

2.4. *Pistacia chinensis* subsp. *integerrima* (J.L. Stewart) Rech. f.

Pistacia chinensis subsp. *integerrima* (J.L. Stewart) Rech. f. (Rechinger, 1969) belongs to the *Anacardiaceae* family with about 70 genera and over 600 species. This is one of the species among the economically important genus *Pistacia* (AL-Saghir and Porter, 2012). The natural habitat of this species includes the S. Transcaucasus, NW India, and Eastern Afghanistan up to the Himalayas. It mostly inhabits the temperate biome. This plant has a wide range of remedial and preventive medicinal capabilities, according to many indigenous medicinal and traditional databases of diverse pharmacological approaches. Various plant components have been used in the creation of herbal remedies since olden days (Jeet and Baldi, 2019). This plant is also called by different names such as Karkatashringi, Shringi, Karkatashringikaa, Karkata, Karkataakhya, Kulirashringaaya, Kuliravishaanikaa, Vakraa, Vishaani. Ajashringi in Ayurveda; Kaakraasingi, Kakar. In Unani; Karkatagasingi in Siddha/Tamil (Khare, 2007; Table 2.2).

Zohary, 1952 grouped the 11 *Pistacia* species and classified them into four categories namely *Lentiscella* Zoh, *Eu Lentiscus* Zoh, *Butmela* Zoh and *Eu Terebinthus*. Also Zohary reported that *P. integerrima* Stewart was considered to be a variety of *Pistacia chinensis* (Zohary, 1952; Parfitt and Badenes, 1997).

Parfitt and Badenes, 1997 showed the phylogenetic relationships of 10 *Pistacia* species including *Pistacia atlantica, Pistacia chinensis, Pistacia khinjuk, Pistacia lentiscus, Pistacia mexicana, Pistacia terebinthus, Pistacia texana, Pistacia vera, Pistacia weinmannifolia* and *P. integerrima* using variability in the DNA level in the chloroplast genome. They demonstrated how *P. integerrima* differed significantly from the more developed *Pistacia chinensis*. The first of the 10 species to blossom when cultivated in the same site in California is *P. integerrima*, and the last is *Pistacia chinensis* (Parfitt and Badenes, 1997).

Table 2.2: Vernacular names of *P. integerrima* (Grover, 2021; Alhat *et al*, 2023; Kadam *et al*, 2023)

Language	**Vernacular Name**
Latin	*Pistacia integerrima*
English	Crabs Claw, Zebrawood
Hindi	Kakdashingi, Kakarsingi, Kakra, Kakkatasingi
Urdu	Kakrasinghi, Kakra
Punjabi	Kakar, Kakarshingi, Drek, Gurgu, Kakkeran, Kakkrangehe, Kakala, Kangar Masna, Sumak, Tungu, Tanbari, Shne
Bengali	Kakra, Kakrashingi, Kandashringi
Gujarati	Kakadasingi, Kakra, Kakarshingi
Marathi	Karkadasringi, Kakra, Kakarsingi, Kakadshingi
Tamil	Karkata, Singi, Kakkatashingi
Telugu	Kakarashingi, Kakatakashrungi, Kakarasimga
Assamese	Kakiasrngi
Malayalam	Karkatasringi, Karkktakasingi
Oriya	Kakadashrungi, Kakadashringi
Sanskrit	Karkatashringi, Sringi, Kulirvishanika, Ajsringi, Karkatakhya
Kannada	Kakadshingi, Karkatashringi

Taxonomic classification of *P. integerrima* (Grover, 2021; Alhat *et al*, 2023; Kadam *et al*, 2023)

- Kingdom - Plantae
- Phylum - Tracheophytes
- Division - Angiosperm
- Subdivision - Eudicots
- Class - Rosides
- Order - Sapindales
- Family - *Anacardiaceae*
- Genus - *Pistacia*
- Scientific Name - *P. integerrima*

P. integerrima is a deciduous tree with a single, sturdy stem that can reach heights of up to 25 meters. It has a dense crown at the base and deep roots. The tree features large pinnate leaves, which are often paripinnate and consist of 2-6 pairs of elongated leaflets. Notably, the terminal leaflet is considerably smaller than the lateral leaflets and can even be reduced to a tiny point (Orwa *et al*, 2009).

This plant is renowned for its aromatic, astringent, and potent medicinal properties in the Ayurvedic system of medicine. It is known for its ability to balance the Kapha and Vata doshas within the body and is employed in the treatment of various ailments, including chronic respiratory disorders (kshayahara), fever (jwarahara), asthma and bronchitis (shwasa), cough (kasa), hiccup (hikka), vomiting (vamana), anorexia (aruchi), excessive thirst (trut), diarrhea (atisara), and bleeding disorders (asrapitta) (Grover, 2021).

Rasapanchak (properties) of *P. integerrima* (Grover, 2021; Alhat *et al*, 2023; Kadam *et al*, 2023)

- Taste - Bitter and astringent
- Physical properties - Dry
- Qualities - Light to digest
- Potency - Hot
- Metabolic property (after digestion) - Pungent
- Pharmacological action - Astringent, Expectorant, Stimulant

The plant is indigenous to India (Orwa *et al*, 2009) China (Grover, 2021) and is also found in the eastern region of the Indian Himalayan area, spanning from the Indus River to Kumaon. It can thrive at altitudes ranging from 350 to 400 meters in the sub-alpine regions of the Himalayas and is also cultivated in low-lying areas. This plant is distributed in various countries, including Nepal, China, Afghanistan, Pakistan, Armenia, as well as in the northwestern and western Himalayan regions worldwide (Grover, 2021). The plant can also be found in England. In India, it is cultivated in the North-West Himalayas, ranging from the Indus region to Kumaon, at altitudes varying from 350 to 2500 meters, as well as in the plains of Punjab (Alhat *et al*, 2023; Figure 2.2).

2.4.1. *P. integerrima* Galls

The petioles and leaves of *P. integerrima* develop tough, horn-shaped, rugose, hollow galls that contain the majority of the plant's metabolites (Patel *et al*, 2018). *P. integerrima* galls are characterized by their fragrant bitterness and find use as both a tonic and expectorant. In northern Pakistan, people consume roasted galls with honey to alleviate conditions like cough, asthma, and diarrhea. Moreover, in Pakistan, galls are employed for the treatment of hepatitis, liver diseases, and are often combined with other remedies to address issues like snake bites and scorpion stings (Kadam *et al*, 2023). In North India, galls are employed for managing inflammatory conditions, diabetes, liver infections, as well as for pain and fever relief. Additionally, both the gall and leaves of the plant are utilized in various regions to address ailments like cough, asthma, common fever, jaundice, diarrhea, and snakebites (Nadkarni and Nadkarni, 1976; Pant and Samant, 2010).

The galls are typically hard, hollow, thin-walled, and generally cylindrical in shape, narrowing at both ends. When the galls are broken open, one can observe a reddish inner surface that may seem dusty but is actually composed of insect debris and their excretory substances (Rauf, 2019).

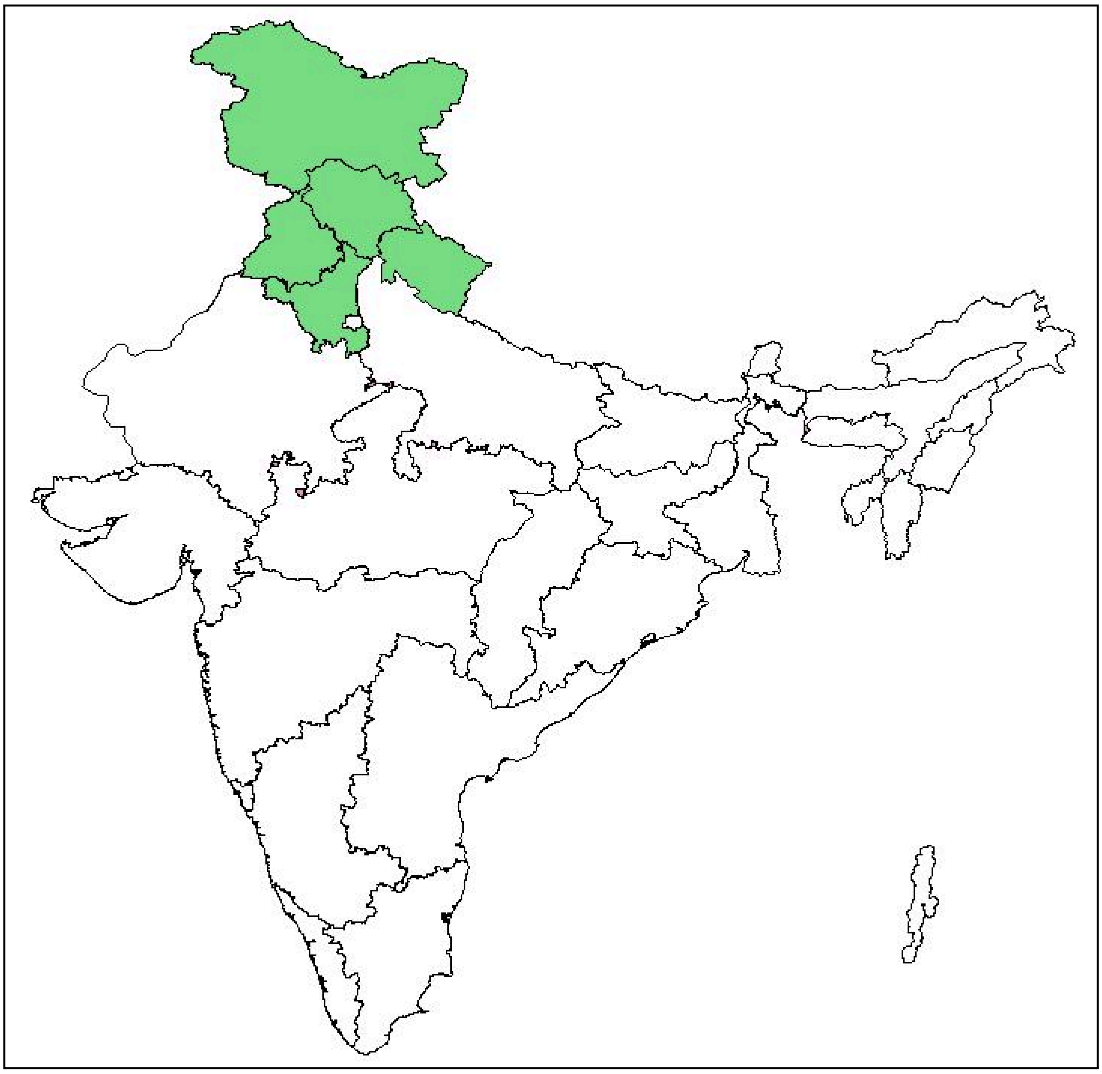


Figure 2.2: Geo distribution map of *P. integerrima* in India (Source: Ved *et al*, 2016; envis.frlht.org)

2.5. Phytochemicals and Pharmacological Uses of of *P. integerrima* Galls

Due to the wide variety of medicinal usage and the special feature of this plant i.e, gall formation; this plant got the attention of many researchers. Various phytochemical properties of this plant as well as corresponding activities have been reported.

P. integerrima, particularly in its galls, harbors a variety of chemical compounds. These compounds encompass tannins at a concentration of 60% and volatile oil at 1.2%. Furthermore, *P. integerrima* is known to contain tetracyclic triterpenes, resin, pistacieonic acids A and B, as well as essential oils comprising camphene, caprylic acids, cineol, α-pinene, and various other constituents (Kadam *et al*, 2023; Table 2.3).

Table 2.3: Phytochemical reported in *P. integerrima* from different studies

Phytochemicals	Reference
δ3-carene, β-pinene, α-phellandrene, α-pinene	Ansari and Qadri, 1993
β-phellandrene, γ-terpenene, α- and β-terpineol, α- and β-ocemene	Ansari and Qadri, 1993
Alkaloids, tannins, flavonoids, and others	Ismail *et al*, 2011
n-decan-3′-ol-yl-n-eicosanoate, n-octadecan-9,11-diol-7-one, 3-oxo-9β-lanost-1,20(22)-dien-26-oic acid, β-sitosterol	Ahmad *et al*, 2010
Flavonoids	Ahmed *et al*, 2008
Ethyl gallate	Mehla *et al*, 2011
Pistagremic acid	Uddin *et al*, 2011
	Uddin *et al*, 2012
	Rauf *et al*, 2014 Rauf et al, 2017

Ansari and Qadri showed that essential oil from *P. integerrima* has CNS-depressant properties. Authors showed the presence of hydroxylated hydrocarbons through IR analysis in essential oil extracted from *P. integerrima* galls (Ansari and Qadri, 1993).

Ismail *et al* investigated the pharmacognostic profile, phytochemical, and physical-chemical characteristics of the bark of *P. integerrima*. Authors also studied the cross section of bark and both microscopic and macroscopic features. The authors reported phytochemical content and physical-chemical characteristics, such as ash qualities, moisture contents, and extractive values, of the crude methanolic extract and its subsequent solvent fractions were assessed. From this study, they reported a preliminary phytochemical profile of *P. integerrima* bark (Ismail *et al*, 2011).

Ahmed *et al* studied the effect of *P. integerrima* leaf extracts on hyperuricemia and gout. From this study, it was found that the flavonoids in *P. integerrima* leaf extracts show potent anti-radical and xanthine oxidase (XO) inhibitory effects. In an animal study, the ethyl acetate fraction significantly lowered Uric Acid (UA) levels. They discovered that the XO inhibitory activity of the insoluble fraction of *P. integerrima* extracts in n-BuOH was extremely low and that extracts in chloroform exhibited intermediate XO inhibition activity (Ahmed *et al*, 2008).

Ahmed *et al* (b) examined the potential analgesic and anti-inflammatory properties of *P. integerrima* extracts in mice models. Using acetic acid to cause belly rigidity and formalin to cause paw licking in mice, analgesia was identified. The thermally generated algesia in mice led to the observation of an antinociceptive effect. They reported the extensive reaction towards chemically prompted ache with the aid of the usage of *P. integerrima* leaves extracts while galls extracts had tremendously extensive safety in a dose based manner. Further, they observed that the extracts of *P. integerrima* showed a significant response in reducing thermal pain, but less than pentazocine and diclofenac. The results of the investigation showed that *P. integerrima* extracts exhibit analgesic and antinociceptive properties, as well as no obvious acute toxicity when administered orally (Ahmed *et al*, 2010).

Mehla *et al* reported the inhibition of cell adhesion molecules (CAMs) using ethyl gallate isolated from *P. integerrima* Linn. In this study, activity-guided extraction from *Pistacia* galls was performed utilizing cell-based ELISA for LPS-induced CAMs expression in human vein endothelial cells (HUVECs), proceeded by analysis of structures of the active ingredients employing IR, MS, and NMR spectroscopy. The substance in *P. integerrima* Linn that is responsible for mediating its anti-inflammatory effect has been identified as ethyl gallate (EG) from the study. Additionally, they showed that EG prevented the production of CAMs that LPS produced by inhibiting the AP-1 transcription factor, revealing its mode of action (Mehla *et al*, 2011).

Uddin *et al* embarked on a comprehensive analysis of the entire *P. integerrima* plant to uncover the pharmaceutical basis for its traditional medicinal use in treating infections. The authors reported the leishmanicidal activity of different phytochemicals isolated from *P. integerrima* (Uddin *et al*, 2011).

Further, By employing α-glucosidase as a therapeutic target in molecular docking simulations, pistagremic acid's powerful α-glucosidase inhibitory capability was predicted. The computational predictions were validated by significant experimental α-glucosidase inhibitory action of pistagremic acid. It was reported that inhibitory action of pistagremic acid against yeast and rat intestinal α-glucosidases was strong (Uddin *et al*, 2012).

Rauf *et al* investigated the in-vivo antinociceptive, anti-inflammatory and antipyretic activity of pistagremic acid. They noted that pistagremic acid had substantial antinociceptive properties both peripherally and centrally that were independent of opioidergic effects and were enhanced by its anti-inflammatory and antipyretic properties (Rauf *et al*, 2014). Also Rauf *et al* evaluated the effect of pistagremic acid (PA) isolated from the galls of *P. integerima* in acute toxicity and in-vivo gastrointestinal (GIT) motility tests using charcoal screening model (Rauf *et al*, 2017).

2.6. *Terminalia chebula* - one of the market substitutes of *P. integerrima*

Terminalia chebula Retz., belonging to the *Combretaceae* family, consistently occupies the foremost position in the 'Ayurvedic Materia Medica' due to its remarkable healing capabilities. This entire plant is endowed with significant medicinal properties and has been a longstanding traditional remedy for treating a wide range of human ailments. *T. chebula* is distributed along the Sub-Himalayan region, stretching from Ravi eastwards to West Bengal and Assam. It can be found at elevations of up to 1,500 meters in the Himalayas. (Bag *et al*, 2013).

Taxonomic classification of *T. chebula* (Bag *et al*, 2013)

- Kingdom - Plantae
- Subkingdom - Tracheobionta
- Superdivision - Spermatophyta
- Division - Magnoliophyta
- Class - Magnoliopsida
- Subclass - Rosidae
- Order - Myrtales
- Family - *Combretaceae*
- Genus - *Terminalia* L
- Species - *T. chebula* (Gaertn) Retz.

The plant has been scientifically shown to exhibit numerous pharmacological and medicinal properties, including but not limited to antioxidant (Chang and Lin, 2010), anticracinogenic (Saleem *et al*, 2002), antimicrobial (Bag *et al*, 2011), antidiabetic (Kannan *et al*, 2012), anti-inflammatory (Pratibha *et al*, 2004), antimutagenic (Grover and Bala, 1992), antiviral (Jeong *et al*, 2002), cardioprotective (Reddy, 1990), gastrointestinal motility enhancement (Tamhane *et al*, 1997), and wound healing (Li *et al*, 2011) activities.

Ravi Shankara *et al* conducted a pharmacognostic examination on leaf galls derived from *T. chebula*, which are utilized as a substitute for *P. integerrima*. This comprehensive study encompassed the analysis of macro- and micromorphological characteristics of the galls, examination of powder fluorescence, physicochemical assessments, and preliminary phytochemical screening (Ravi Shankar *et al*, 2012).

Saini *et al* conducted a study involving the preparation of hydroalcoholic and aqueous extracts from *P. integerrima, Q. infectoria* and *T. chebula*, followed by TLC analysis and high performance liquid chromatography (HPLC) quantification of gallic acid and ellagic acid in the extracts. The thin layer chromatography (TLC) analysis revealed similarities in the metabolite patterns among the three plants, and a precise HPLC method demonstrated gallic acid content ranging from 3.74% to 10.16% w/w and ellagic acid content ranging from 0.10% to 1.24% w/w in these plants (Saini *et al*, 2023).

2.7. Use of Omics Approach to Study the Mechanism of Gall Formation

Omics technologies are becoming crucial tools for understanding the mechanisms behind interactions because they enable the analysis of both plant hosts and insect effectors, even when there is limited or no prior knowledge available. Over the past decade, omics technologies have emerged as a powerful new approach for advancing our understanding of interactions between plants and galling insects (Oates et al, 2016). The integration of multi-omics data will have a significant role in pinpointing the molecular mechanisms responsible for gall formation. Omics technologies are increasingly employed to study plant-insect interactions (Barah and Bones, 2015).

Metabolomic and transcriptomic methods are frequently applied to study alterations in phytohormone levels, as well as the genes involved in hormone synthesis and response, when plants are infested by insects. Zhang *et al*, investigated the changes in phytohormone levels during the interaction between apple and the leaf-mining lepidopteran *Phyllonorycter blancardella* using liquid chromatography and mass spectrometry (LC-MS) and microarray expression profiling. They found that the mined tissue exhibited elevated levels of cytokinin, jasmonic acid, and salicylic acid, along with a reduction in abscisic acid (Zhang *et al*, 2016).

Yamaguchi *et al*, conducted a study on the interaction between the gall-inducing sawfly and its host plant *Salix japonica*. They employed a combination of LC-MS and RT-qPCR to explore the host's signaling pathways. Their research provided the first evidence of a galling insect's capability to produce indole acetic acid (IAA) from tryptophan (Yamaguchi *et al*, 2012). Mukhtar *et al*, utilized proteomic data from *Arabidopsis thaliana* and two pathogens to illustrate that pathogen effectors have a tendency to focus on central components in the plant immune response network (Mukhtar *et al*, 2011)

CHAPTER III

WHOLE GENOME AND TRANSCRIPTOME SEQUENCING OF *P. INTEGERRIMA*

This chapter provides an in-depth exploration of various aspects related to the genome of *P. integerrima* such as assembly, annotation, metabolic pathway and phylogeny. It compiles transcriptome assembly, annotation, gene expression and pathway analysis of *P. integerrima* gall and leaf

3. Whole Genome and Transcriptome Sequencing of *P. integerrima*

3.1. Introduction

The sustainability of all higher animals on Earth relies on plants. Instead of focusing on individual genes, current molecular and computational methods have made it possible to analyze the structure, evolution, and function of whole plant genomes (Gebhardt *et al*, 2005). *Arabidopsis thaliana*, a flowering plant, is an ideal model system for finding genes and figuring out their activities. The plant genome sequence offers a way to comprehend the genetic underpinnings of variations between plants and other eukaryotes. This serves as the basis for in-depth functional characterisation of plant genes (The *Arabidopsis* genome initiative, 2000). Research on plant biology has dramatically changed as a result of the development in genomics technologies. To fully understand the vast genetic diversity of plants at the molecular level, plant researchers today have easy access to huge genomic data. The field of plant genomics research has recently experienced significant growth and prominence, primarily due to the substantial increase in the number of plant species with sequenced genomes. (Ong *et al*, 2016). The localisation of genetic elements that regulate mendelian or complex phenotypic features on molecular maps is a step in the functional study of plant genomes (Gebhardt *et al*, 2005).

The majority of DNA sequencing in the twentieth century used sanger sequencing, which had limitations on the reliability (precision) and volume of data that could be produced in comparison to next-generation sequencing (NGS) that is available today. Plant genome sequencing has advanced more quickly in recent years, notably due to the rapid development of NGS (Henry, 2022). The plant genomic era officially began in December 2000 with the complete genome sequencing of *Arabidopsis thaliana*, the first plant reference genome. Utilizing this reference genome, numerous plant species, ranging from nonvascular to flowering plants, have been generated over the last two decades (Sun *et al*, 2022). Plant systematics and evolutionary biologists will soon find constructing entire genomes from non-model plants to be normal thanks to the quick advancement of sequencing technology and the decline in cost (Li and Harkness, 2018).

Genome biology would benefit greatly from having a better understanding of the biochemical processes of important bioactive compounds from a genomics perspective in order to create and produce innovative chemicals on a big scale. Utilizing cutting-edge techniques like NGS, it may be possible to identify genes involved in certain metabolic pathways of several medicinal plants (Zhang *et al*, 2019). A genome analysis provides details on the genome makeup of a plant species, such as estimations of its size, levels of heterozygosity, and repeat information (Ziya *et al* 2016).

3.1.2. Expression of Gall Forming Genes

An important characteristic of insect-induced galls is that, in comparison to microbial galls, the majority are architecturally more structured and sophisticated, with unique tissue differentiation linked to activator feeding and defense (Hearn *et al*, 2019). Studies established the viability of researching gall development by transcriptome techniques and revealed the genetic adaptation mechanism of plant cells following gall induction (Shih *et al*, 2018).

Different studies have been carried out to understand altered cellular architecture of galls. A transcriptome analysis of flooded gum (*Eucalyptus grandis*) identified alterations in genotype and defensive mechanisms the plant established to fend off the gall wasp's egg production (Oates *et al*, 2015). The genes that react to auxin (a phytohormone) were highly expressed in galled *Myrtaceae* leaves, according to a comparative transcriptomic and genotypic study among galled and ungalled leaves (Bailey *et al,* 2015).

Investigation using RNA-seq and qPCR revealed substantial variations in the expression levels of several genes engaged in various metabolic and developmental processes, and analysis using gene ontology (GO) suggested multiple biological processes likely involved in the formation of the radish crown gall (Tkachenko *et al,* 2021). RNA-seq is a strong and effective approach for whole-genome level differentially expressed gene (DEG) identification, and it is particularly useful for screening the transcriptome of non-sequenced species (Wang *et al*, 2009).

For the first time, in the present work, we sequenced the transcriptome of P. integerrima gall and leaf for comparing the transcription of both. Gall and leaf transcripts were assembled and annotated entirely from scratch. To find distinctive genes, their expression patterns, and pathways, the transcriptomes of the gall and the leaf were compared and analyzed. The understanding of the genes and pathways involved in gall formation in *P. integerrima* will be aided by these molecular resources. For *P. integerrima*, no genetic data are available. The goal of this work was to characterize the *P. integerrima* genome in order to better understand the *Pistacia* genes and pathways involved in therapeutic application.

3.2. Materials and Methods

3.2.1. Sample Collection

For this investigation, the *P. integerrima* trees which were located in their natural habitat and identity were confirmed by field botanists and specialist taxonomists. Under the supervision of botanists, the leaf and gall samples were obtained from the Indian state of Uttarakhand (Figure 3.1). To prevent RNA degradation, the plant sample used for RNA isolation was immediately washed with normal saline and preserved in RNALater. The samples were shipped to the lab and stored at -80°C. Gall and leaf voucher specimens were deposited to the FRLH Herbarium in Bengaluru.

3.2.2. DNA Isolation, Genomic Library Preparation and Genome Sequencing

The cetyltrimethylammonium bromide (CTAB) technique was used to extract genomic DNA with slight modifications (Novaes *et al,* 2009). With the use of nanodrop and gel electrophoresis, the genomic DNA's quality was verified. Enzymatically, 250 bp fragments of the genomic DNA were produced. The NEBNext Ultra TM II DNA FS Library Prep kit for Illumina was used to create a sequencing library from about 50 ng of fragmented DNA. Qubit DNA High Sensitivity Quantitation Assay was used to quantify the library, and the Bioanalyzer 2100 was used to assess the library's quality. HiSeq 2500 was used for the DNA sequencing process.

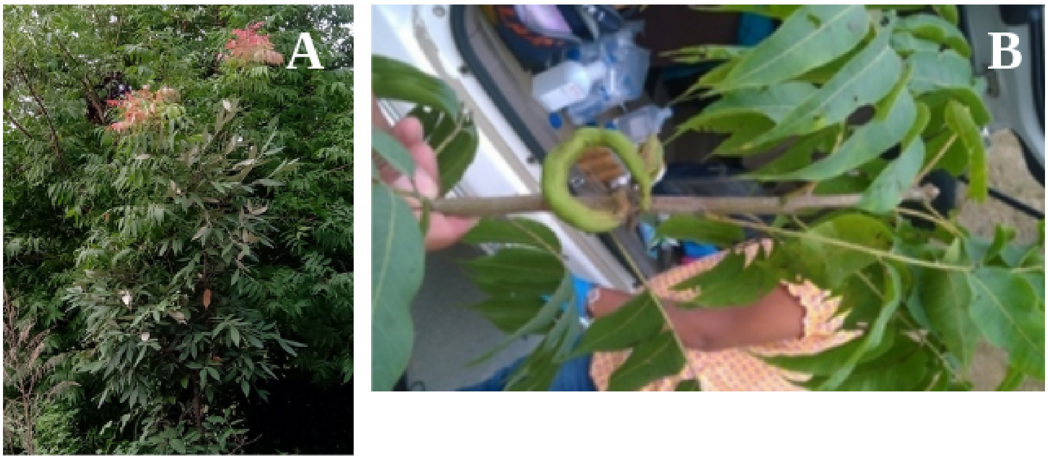


Figure 3.1: *P. integerrima* sample collection. A - *P. integerrim*a plant identified in its natural habitat; B - Collection of *P. integerrima* gall and leaf samples

3.2.3. RNA Library Preparation and Sequencing

Total RNA was extracted from gall and leaf samples. The quality of RNA was checked on Bioanalyzer and quantified using QUBIT dsRNA HS kit. The library was prepared using “NEBNext® Ultra™” RNA Library Prep Kit for Illumina NextSeq500 with standardized protocol.

The final enriched libraries were further validated for quality on Agilent Bioanalyzer using DNA High Sensitivity chip and for quantification on real time PCR (KAPA Library Quantification kit). The quality and quantity of the prepared library met the Illumina standards required for further sequencing and hence the library was normalized. The library was denatured using NaOH followed by neutralizing the pH conditions by adding 0.2N Tris, pH 7 and was taken further for cluster generation and sequencing. About 40 million paired-end reads were generated on Illumina NextSeq500 platform.

3.2.4. Data QC and Genome Assembly

FastQC v0.11.6 (Andrews 2010) was used to do the first quality assessment of the raw reads from the *P. integerrima* sequencing data from the Illumina library, and trim galore v0.4.4 dev (Krueger 2015) was used with the default settings to eliminate low-quality reads. In order to determine the genome size, Jellyfish v2.2.10 was used to further process the filtered reads (Marçais and Kingsford 2011). A significant peak at 69x was visible in the 21-kmer distribution. Using the equation genome size = k-mer number/peak depth, the total number of k-mers and the matching k-mer depth were used to determine the size of the *P. integerrima* genome.

KmerGenie v1.7048 was used to estimate the best k-mer from the high quality data (Chikhi and Medvedev 2014). Various k-mers, ranging from 21 to 121, were used to forecast the assembly size in diploid mode. For de novo assembly, the best k-mer that produced the best assembly size was chosen. gGenome assembly was performed using two assemblers; SOAPdenovo v2.04 (Luo *et al,* 2012) and SPAdes v3.11.1 (Bankevich *et al,* 2012).

SSPACE v5.26.2 (Boetzer *et al*, 2011) and GapClosure v1.12 (Luo *et al*, 2012) modules from SOAPdenovo2 were also utilized to scaffold the assembly. The sequences for the mitochondria, chloroplasts, and vectors were acquired from the NCBI database. Using Blastn v2.6.0+ (Camacho *et al*, 2009) (blastn parameter used for mitochondria and chloroplast; e-value is 1e-10 and for vector sequences; reward: 1, penalty: -5, gapopen: 3, gapextend: 3, dust: yes, soft masking: true, searchsp: 1750000000000), the chloroplast, mitochondrial, and vector sequences were recognized and eliminated from the assembly. The genomic assembly of *P. integerrima* was submitted to Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 (Simão *et al*, 2015) to evaluate the genome completion after the deletion of chloroplast, mitochondrial, and vector regions.

3.2.5. RNA Seq Assembly and Annotation

Utilizing FastQC v0.11.613, quality of RNA seq data from both gall and leaf was examined (Andrews 2010). Using Trimgalore v0.4.4 dev14, reads with a Phred quality score of less than 30 were eliminated (Krueger 2015). Using Trinity v2.4.031, the gall and leaf reads that had been clipped and were assembled into transcripts (Grabherr *et al*, 2011). The transcript sequences were examined using TransDecoder-v5.0.2 (https://transdecoder.github.io/), which was utilized to find potential coding regions. The Viridiplantae sequences from the UniProtKB database was used for the annotation of assembled transcripts.

3.2.6. Genome Annotation

3.2.6.1. Repeats Identification

Repeat modeler v1.0.11 (https://www.repeatmasker.org/RepeatModeler/) was used to generate the repetition library and identify de novo repeats. Repbase v23 8 (Bao *et al*, 2015) was used for repeat annotation, and *Arabidopsis thaliana* served as the reference. The assembly's recognised and annotated repetitions were masked up using the RepeatMasker (available at https://www.repeatmasker.org). Using the MIcroSAtellite (MISA) identification tool, the simple sequence repeats (SSRs) were identified (Thiel *et al*, 2003).

3.2.6.2. Genome Annotation

The repeat-masked assembly was then subjected to annotation using a variety of techniques, including predictions made using ab initio approaches, homology-based methods, and evidence-based methods. *A. thaliana* parameters were employed with the ab initio based prediction models Augustus v3.3.2 (Stanke and Waack 2003), GlimmerHMM v3.0.4 (Majoros *et al*, 2004), GeneID v1.4.4 (Blanco *et al*, 2007), and SNAP (Korf 2004). From the NCBI refseq database, predicted homologous protein sequences from the following species were obtained and utilized for homology-based gene prediction: *Pistacia vera* (GCF 008641045.1), *Arabidopsis thaliana* (GCF 000001735.4), *Theobroma cacao* (GCF 000208745.1), *Citrus sinensis* (GCA 000695605.1), *Vitis vinifera* (GCF 000003745.3), *Beta vulgaris sub* (Keilwagen *et al*, 2016). The RNA-Seq reads were combined into de novo contigs using Trinity v2.4.031 (Grabherr *et al*, 2011), and the resulting unigenes were aligned using BLAT (Kent, 2002) to the repeat-masked assemblies. The gene structures of the BLAT alignment results were then modeled using a program to assemble spliced alignments (PASA v2.4.1; (Haas *et al*, 2003). With a maximum and minimum intron length of 500,000 and 50 bp, correspondingly, the RNA seq reads were also matched to a repeat-masked genome assembly using Tophat2 (Kim *et al*, 2013). The cufflinks were put together using cufflinks v2.2.1 (Trapnell *et al*, 2010), and they were used as proof in PASA. To create a consensus gene collection, predictions from all three approaches were integrated using EVidenceModeler v1.1.1 (Haas *et al*, 2008) (score > 1000). (Bi *et al*, 2019). By matching protein sequences to the Swissprot database, protein-coding genes' functional annotation was completed (Boeckmann *et al*, 2003). By doing a search against the Pfam (Mistry *et al*, 2021) database, the protein motifs and domains were annotated. The Kyoto Encyclopaedia of Genes and Genomes (KEGG) Automatic Annotation Server (KASS) was used to map gene pathways (Moriya *et al*, 2007).

3.2.6.3. Noncoding RNAs Identification

To predict tRNA genes, the tRNAscan-SE v.2.0.7 (Chan and Lowe 2019) method was used with default settings. Using the Rfam database and the INFERNAL v1.1.2 programme (Nawrocki and Eddy 2013), the genes for small nucleolar RNA (SnRNA) and miRNA were predicted (Griffiths-Jones *et al*, 2003).

3.2.7. Ortholog Detection, Gene Family Construction and Phylogenetic Analysis of *P. integerrima*

Protein sequences from the six species, including *Arabidopsis thaliana*, *Citrus sinensis*, *Pistacia vera, Solanum tuberosum, Vitis vinifera*, and *P. integerrima*, were clustered using OrthoVenn2 (Xu *et al*, 2019). A list of putative orthologs, co-orthologs, and paralogs was created using comparative data from the OrthoVenn2 analysis. By grouping the gene-pairs, the software has classified the proteins as well as organized them into distinct categories.

Using the ProteinOrtho v6.0.27 (Lechner *et al*, 2011) programme was used to search for homologues and unique genes in the proteomes of 13 sequenced plant species, including *P. integerrima*, A*rabidopsis thaliana, Oryza sativa, Vitis vinifera, Glycine max, Theobrama cocao, Beta vulgaris, Citrus sinesis, Ricinus communis, Pistacia vera, Solanum tuberosum, Solanum lycopersicum.* The ProteinOrtho programme was used to build the phylogenetic tree.

3.2.8. Transcriptome Annotation

Using BLASTX (Camacho *et al*, 2009; Altschul *et al*, 1990) and BLASTP (Camacho *et al*, 2009; Altschul *et al*, 1990) with the SwissProt [Boeckmann *et al*, 2003] database, the transcripts were utilized for functional annotation. Protein domains were identified using HMMER (Eddy, 1996) with the PFAM (Mistry *et al*, 2021) database.

The uniprot Retrieve/ID mapping tool (https://www.uniprot.org/uploadlists/) was used to find the Gene Ontology (GO) keywords.

3.2.9. Differential Gene Expression

Non-redundant concatenated assembled transcript sequences (at 90% sequence similarity) from CD-hit EST (Li and Godzik, 2006) of gall and leaf were used as a benchmark to calculate expression. Using bowtie2 (Langmead and Salzberg, 2012; Langmead *et al*, 2019), the RNA seq data from both samples were aligned to the non-redundant combined transcriptome assembly. TMM values were used to measure the level of RNA expression (Li *et al*, 2015; Robinson *et al,* 2010), and NOISeq v2.14.1 was used to analyze differential expression (Tarazona *et al*, 2015). The input for NOISeq-sim comprised raw counts, which were normalized using TMM values. The NOISeq-sim approach was used to perform differential expression analysis after the low abundant transcripts were filtered (CPM 5 in any of the datasets). With the sequencing depth (total number of reads) of the unique available sample given, the size of the simulated samples constitutes 10% of this sequencing depth, with a slight allowable variability of 0.05. Five replicates are to be simulated.

TMM values are used to express the gall and leaf transcript expressions. In order for RNA scaffolds or transcripts to be considered highly significantly differentially abundant, the fold-change between the two samples has to be at least 2. Additionally, a significance threshold (q-value) of 0.95 was utilized to ensure a 95% probability of true differential expression, as recommended in NOISeq for NOISeq-sim analysis. To identify potential functions, the differentially expressed genes were BLASTX (Camacho *et al*, 2009; Altschul *et al*, 1990) searched against the NR database. The Kyoto Encyclopedia of Genes and Genomes (KEGG) was utilized to describe related pathways, and KAAS (KEGG Automatic Annotation Server) (Moriya *et al*, 2007) to further examine the functional associated pathways for DEGs. Using ClusterProfiler, pathway enrichment was performed (Wu *et al*, 2021; Yu *et al*, 2012). The significantly enriched pathways were identified using p-value (< 0.05).

3.3. Results

3.3.1. *De novo* Genome Assembly of *P. integerrima*

Using the Illumina HiSeq 2500, the paired end (2 x 150 bp) DNA libraries for *P. integerrima* were created. 551 million readings (83 Gb) in total were produced (Table 3.1). With k-mer size 21, the high-quality reads (phred score >30) were utilized to predict genome size. The peak was projected to be 69 for k-mer depth vs k-mer coverage. K-mer coverage and K-mer depth were used to compute the genome size, which resulted in a haploid genome size of 452,531,485 base pairs (452 Mbp), where we anticipated a 55.4% (251,140,882 bp) single copy regions were predicted..

All Pistachio cultivars are diploid, with chromosomal numbers 2n = 24, 28, and 30, according to earlier research (Ghaffari and Shabazaz, 2005). The nearest species to *P. integerrima* is *Pistacia vera* L., whose genome and chromosomal size were determined to be 2n=30 and 513 Mb, respectively (Basr *et al*, 2003; Ziya *et al*, 2016).

The high-quality sequencing reads were then assembled using SOAPdenovo2 and SPAdes softwares. KmerGenie was used to verify the optimal k-mer prior to assembly. Additionally, the assembly size was projected using several k-mers with intervals of two, ranging from 21 to 121. Both stated that 115 was the ideal k-mer size. SPAdes has provided the best assembly with a k-mer size of 115 out of SOAPdenovo2 and SPAdes. With a total assembly size of 463,342,136 bp (463.3 Mb) and a highest contig size of 238,122 bp, we were able to acquire 205,922 contigs (>200 bp). The average scaffold size is 2250.1 bp, while the N50 value was discovered to be 16,145 bp.

Using the same set of reads as for the assembly, we attempted to construct a contig assembly using SSPACE (Boetzer *et al*, 2011) and GapClosure (Luo *et al*, 2012). We were unable to distinguish between the main contig assembly and the scaffold assembly since the identical set of data was employed. The assembly was also utilized to remove the vector, mitochondria, and chloroplast sequences. After removing these sequences, with an assembly of 462,064,448 base pairs (462 Mb), with the average contig size being 2243.9 base pairs and the maximum contig size being 238,122 base pairs was obtained (Table 3.2).

Table 3.1: ***P. integerrima*** **genome sequencing statistics**

	Before Trimming			**After Trimming**		
	Forward	**Reverse**	**Total**	**Forward**	**Reverse**	**Total**
Reads	27,55,54,024	27,55,54,024	55,11,08,04	26,68,55,293	26,68,55,293	53,37,10,586
Total bases (bp)	41,60,86,57,624	41,60,86,57,624	83,21,73,15,248	34,76,10,72,002	34,76,10,72,002	69,522,144,004

Table 3.2: ***P. integerrima*** **genome assembly statistics**

Attributes	**Assembly statistics**
Total assembly size (>200 bp)	462,064,448
Mean Contig/scaffold size (bp)	2,243.9
Largest contig/scaffolds (bp)	238,122
Total contigs/scaffolds (> 200 bp)	205,044
N50 (bp)	16,109
L50	6,843

The assembled sequences were verified using the eudicots odb10 lineage and benchmarking universal single-copy orthologs (BUSCO) for assembly completeness. A total of 2,121 BUSCO groups were examined, and we found 1,938 complete BUSCOs (1907 single copies and 31 duplicates), 108 fragmented BUSCOs, and 75 missing BUSCOs. As a result, the projected 91.4% of complete BUSCOs demonstrates good assembly (Table 3.3).

Table 3.3: Statistics of the completeness of the assembled *P. integerrima* genome predicted using BUSCO

Contents	BUSCO numbers	BUSCO percentage
Complete BUSCOs	1,938	91.4
Complete and single-copy BUSCOs	1,907	89.9
Complete and duplicated BUSCOs	31	1.5
Fragmented BUSCOs	108	5.1
Missing BUSCOs	75	3.5
Total BUSCO groups searched	2,121	-

3.3.2. *P. integerrima* Genome Annotation

The repeating sequences, including simple sequence repeats (SSRs) and complex repeats, were found using the final assembly. Total, 206,323 SSRs and 27,214 scaffolds containing multiple SSRs have been found. SSRs were divided into mono, di, tri, tetra, penta, and hexa categories. The assembly has more mono repetitions than di (152,221), tri (14,282), tetra (2184), penta (577) and hexa (493) repeats (Figure 3.2).

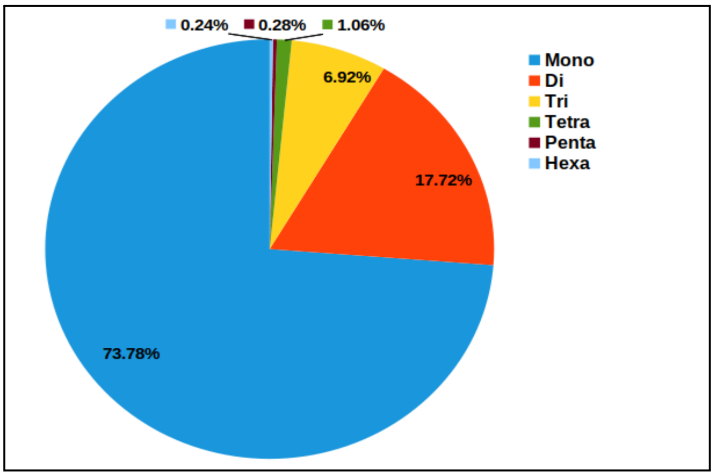


Figure 3.2: SSRs identified from *P. integerrima* genome

An examination of complex repeats revealed that the *P. integerrima* genome contained 57.31% transposable elements (TEs), the bulk of which were long terminal repeat (LTR) retroelements. A modest number of SINEs and DNA transposons were obtained (Table 3.4). The repetitions that were found were marked for more examination.

Table 3.4: *P. integerrima* repeat identification statistics

Type of repeat elements	**Number of elements**	**Length occupied (bp)**	**Percentage of sequence (%)**
LINEs	4,228	1,738,422	0.38
LINE1	4,228	1,738,422	0.38
LTR elements:	119,054	90,883,679	19.67
DNA elements:	25,707	15,895,578	3.44
Unclassified:	523,324	150,007,185	32.46
Total interspersed repeats		258,524,864	55.95
Simple repeats	130,472	6,089,413	1.32
Low complexity	21,701	1,180,083	0.26

The repeat-masked assembly was used to predict protein-coding genes as well. We predicted 39,425 gene models, of which 20,933 genes were discovered by *de novo* and 18,492 genes contain evidence in the form of RNA or proteins. Additionally, we discovered genes with mean CDS lengths of 1,113.5 nts and 2,107.6 nts, respectively, and mean gene lengths of 2,107.6 nts. We found three introns on average per gene, with an average length of 327.9 nts (Table 3.5). The predicted genes were checked against various databases, including Pfam and Swissprot. Total, 38,466 of the 39,425 genes share similarities to those in the Swissprot database. Gene ontology concepts (GO) were used to categorize the genes. According to the GO annotation, genes are equally divided across cellular components, biological process, and molecular function (Figure 3.3).

Table 3.5: *P. integerrima* genome annotation information

Annotation	**Protein coding genes**
Number	39,425
Transcript/Protein evidence	18,492
Ab initio	20,933
Single-exon gene count	13,612
Mean gene length (bp)	2,107.6
Mean CDS length (bp)	1,113.5
Mean introns per gene (bp)	3
Mean intron length (bp)	327.9
Mean intergenic region length (bp)	1,826.3
SwissProt hits	3,8426
Pfam hits	2,90400

The majority of genes under "molecular function" are engaged in binding, catalysis, molecular function regulation and transporter activity. Similar to this, more genes are involved in the protein-containing complexes referred to as cellular components and anatomical entities of the cell. More genes are engaged in biological activities such as cellular activity, metabolism, biological control, stimulus response, localisation, development, growth and reproduction (Figure 3.4). Gene models from *P. integerrima* were also checked against the Pfam database. Total, 34,850 of the 39,425 genes have been found to contain Pfam domains.

In the genome of *P. integerrima*, there are 488 tRNAs with an average length of 75 nts. Additionally, 25 tRNAs containing introns were found. *P. integerrima*'s genome included 1,602 small nucleolar RNA (snRNA), of which 1372 were unique to plants, 191 microRNA (miRNA), 99 spliceosomal RNA, 108 eukaryotic large subunit ribosomal RNA (rRNA), and 35 eukaryotic small subunit ribosomal RNA (rRNA) (Table 3.6).

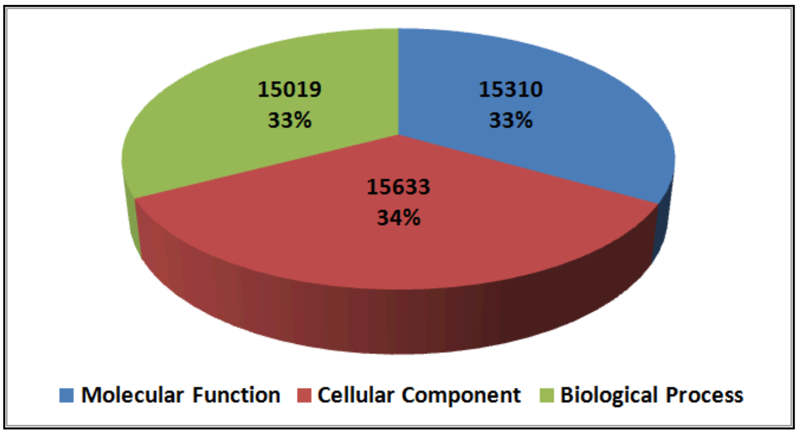


Figure 3.3: Gene ontology classification of *P. integerrima* genes

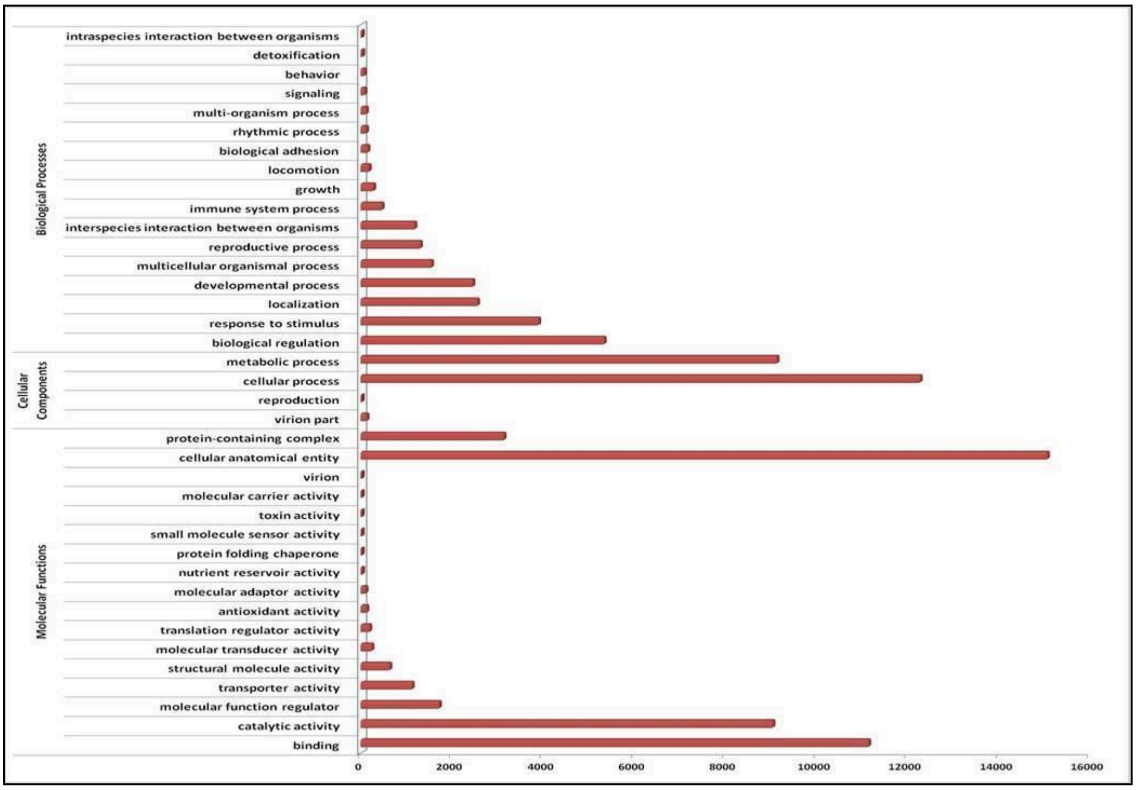


Figure 3.4: Gene Ontology (GO) subcategories of *P. integerrima*

Table 3.6: Statistics of Non-coding RNA (NC RNA) in *P. integerrima* genome

RNA type	Numbers
tRNA	488
tRNAs with introns	25
small nucleolar RNA (snRNA)	1602
microRNA (miRNA)	191
Spliceosomal RNA	99
Eukaryotic large subunit ribosomal RNA (rRNA)	108
Eukaryotic small subunit ribosomal RNA (rRNA)	35

3.3.3. *P. integerrima* Pathway Analysis

The protein sequences of predicted gene models in P. integerrima were queried against the KEGG database using KEGG-KAAS server to obtain KEGG Orthology (KO) IDs. A total of 3895 KO ids and 7893 gene models were assigned, resulting in 398 pathways. Three levels of pathways have been assigned to the genes. Cellular processes, human illnesses, metabolism, genetic information processing, environmental information processing, and organismal systems are all included in Level I. *P. integerrima* genes are more prevalent and more metabolically active (Figure 3.5).

Level II global and overview maps included 1997 genes, then 713 genes for neurological disorders, 438 genes for signal transduction, 334 genes for carbohydrate metabolism, etc. At level III, the top pathways with the most genes include metabolic pathways, production of secondary metabolites, microbial metabolism in various conditions, *etc*.

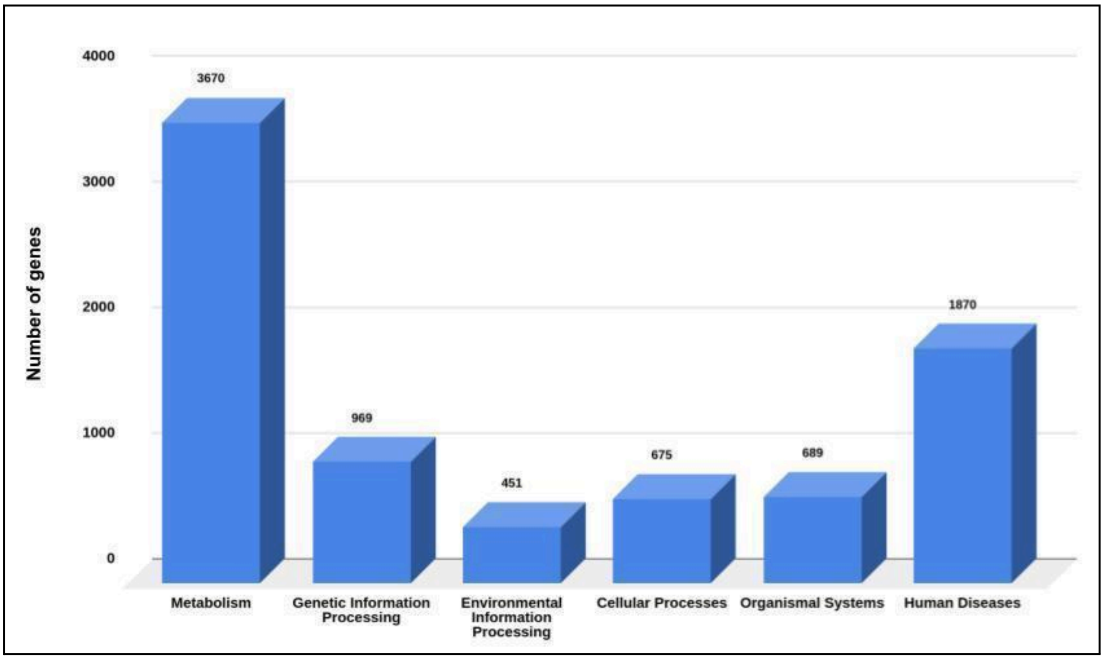


Figure 3.5: Predicted pathways of *P. integerrima* through predicted genes from the genome. More number of genes are identified in the metabolism pathway which is followed by human diseases, Genetic information processing, etc.

Upon discovery, it was found that 103 genes play a role in the synthesis of various secondary metabolites, and additional genes are implicated in the pathways for flavonoid and phenylpropanoid production. Additionally, 92 genes were identified that were involved in the metabolism of terpenoids and polyketides, of which 21 and 31 genes involved in “carotenoid biosynthesis” and “terpenoid backbone biosynthesis” respectively (Figure 3.6A and 3.6B). Mevalonic acid (MVA) in the cytoplasm and methylerythritol 4-phosphate (MEP) in the plastid are two distinct biochemical routes that are involved in the synthesis of terpenoids. The biosynthetic process for terpenoids mainly consists of three stages: the first stage involves the formation of intermediates, which are frequently precursors of terpenoids, such as isopentenyl-P (IPP) and dimethylallyl-PP (DMAPP); the second stage involves the synthesis of three direct precursors, geranyl-PP (GPP), farnesyl-PP (FPP), and geranyl geranyl-PP (GGPP); The MVA and MEP terpenoid backbone biosynthesis pathways' associated genes have all been successfully assigned.

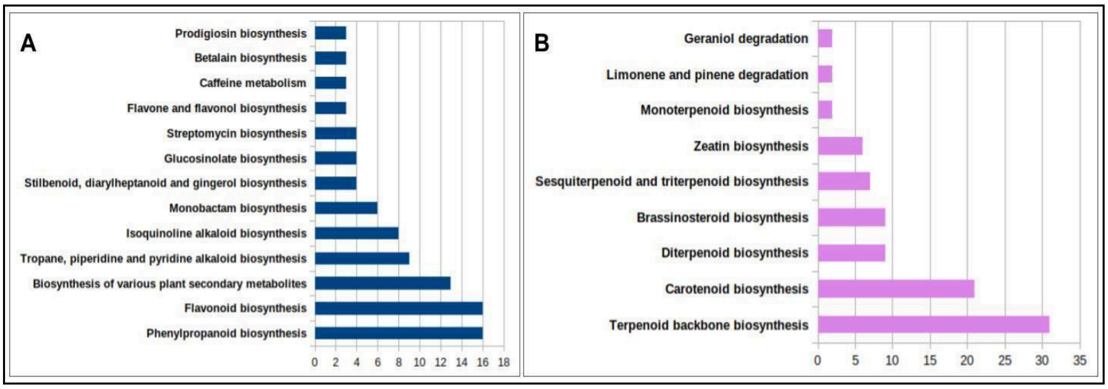


Figure 3.6: Pathways of biosynthesis of different secondary metabolites identified in *P. integerrima*. (A) Biosynthesis of different secondary metabolites pathway (phenylpropanoids, terpenoids, etc). (B) Biosynthesis of terpenoids. X-axis represents the number of genes.

3.3.4. Orthology detection and phylogenetic analysis of *P. integerrima*

The proteomes of 5 sequenced plant species (*Arabidopsis thaliana, Citrus sinensis, Pistacia vera, Solanum tuberosum,* and *Vitis vinifera*) were compared with the protein sequences for 39,425 genes in *P. integerrima*. The six species combine to generate 22,230 single-copy gene clusters, 23,944 orthologous clusters (at least two species), and 26,174 clusters (Table 3.7).

P. integerrima produces 10,673 singletons and 19,932 (76%) clusters (proteins are not in any cluster). For pathway analysis, the singletons from *P. integerrima* were aligned to the KEGG database. 93 KO ids were assigned to the 10,673 proteins. In metabolic pathways and the manufacture of secondary metabolites, more genes have been found. All six species shared 9065 core orthologous groups (COGs) gene families, of which 10,142 (13.7%) genes are from *P. integerrima*.

The similarity matrix revealed that *P. integerrima* and Pistacia vera share more clusters (18,325). (Figure 3.7). In addition to Orthovenn2, the proteomes of 13 different plant species were examined, and proteinortho software was used to construct a phylogenetic tree based on how similar the proteomes of the different plant species were. 13 species were analyzed using phylogenetic analysis, and *P. integerrima* was shown to be most closely related to *Pistacia vera* (Figure 3.8).

Table 3.7: Summary of proteins of *P. integerrima* shared with different species

Species	Proteins	Clusters	Singletons
Arabidopsis thaliana	27,413	14,425	5,211
Solanum tuberosum	37,475	13,990	9,593
Vitis vinifera	26,556	14,709	7,426
P. integerrima	39,425	19,932	10,673
Pistacia vera	41,299	19,935	7,033
Citrus sinensis	39,056	17,267	2,548

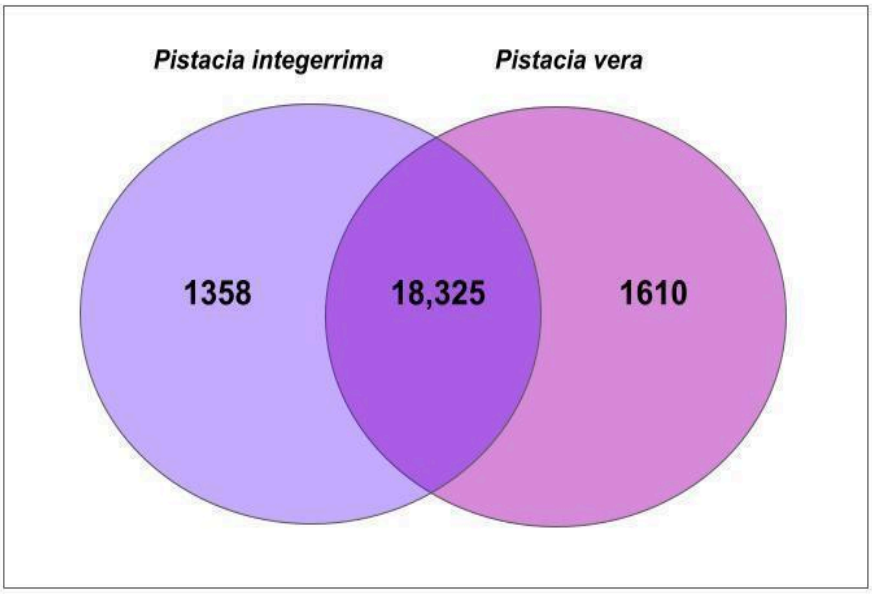


Figure 3.7: Clusters shared by *P. integerrima* and *Pistacia vera* based on number of genes shared between the two species

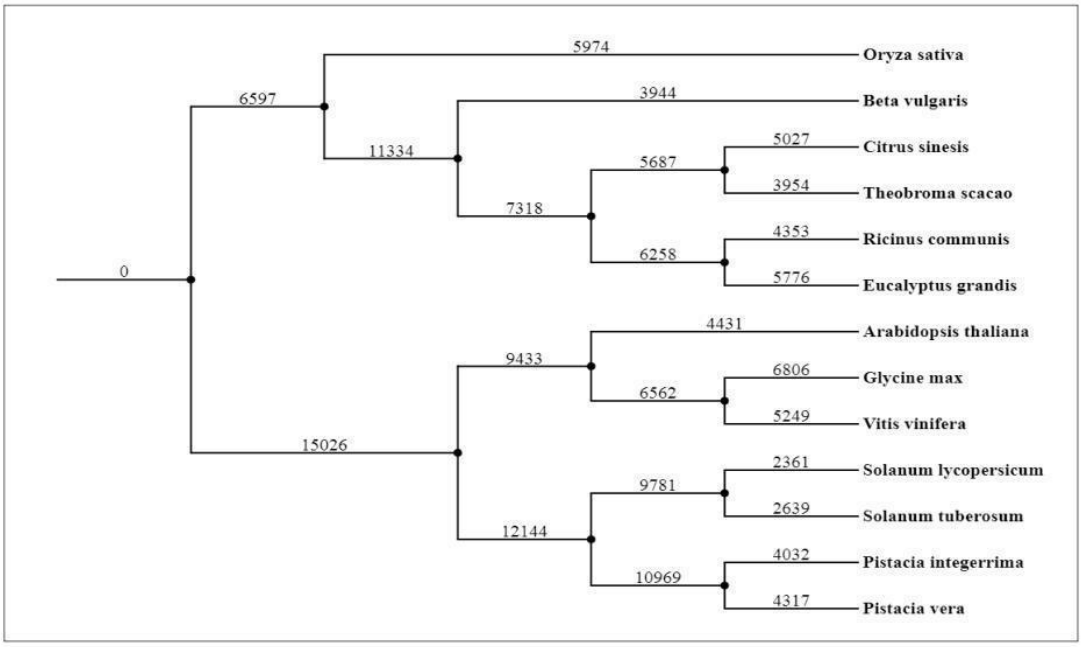


Figure 3.8: Phylogenetic tree of *P. integerrima* with other species. *P. integerrima* is closer to *Pistacia vera* which is another species of the same genus *Pistacia*.

3.3.5. Transcriptome Assembly and Gene Expression Analysis

Using the Illumina NextSeq500 platform, RNA sequencing data for leaf and gall samples were produced. For leaf and gall, 3.7 and 3.4 million reads forming 5.6 Gb and 5.1 Gb of data were produced respectively . Low quality data was removed, and the assembly consumed 4.6 and 4.1 Gb of data (Table 3.8).

Trinity was used to construct independent as well as combined *de novo* transcriptome assembly for *P. integerrima* leaf and gall samples. For leaf and gall, the optimal assembly produced 46,571 and 73,572 transcripts, respectively. 1,01,619 transcripts with a N50 of 511 bp and an average transcript length of 524 bp were produced by the combined assembly (Table 3.9).

Table 3.8: ***P. integerrima*** **transcriptome sequencing statistics**

Before Trimming						
	Leaf			**Gall**		
	Forward	**Reverse**	**Total**	**Forward**	**Reverse**	**Total**
Reads	1,88,66,673	1,88,66,673	3,77,33,346	1,70,32,124	1,70,32,124	3,40,64,248
Total bases (Gbp)	2.84	2.84	5.69	2.57	2.57	5.14
After Trimming						
	Leaf			**Gall**		
	Forward	**Reverse**	**Total**	**Forward**	**Reverse**	**Total**
Reads	1,79,25,334	1,79,25,334	3,58,50,668	1,64,32,362	1,64,32,362	3,28,64,724
Total bases (Gbp)	2.32	2.32	4.64	2.08	2.08	4.16

Using blastn, the transcripts of both gall and leaf were compared to find their shared and distinctive transcripts. We found 30,512 transcripts that were shared by both gall and leaf, and 40,060 and 16,059 transcripts that were exclusive to each (Figure 3.9).

3.3.6. Functional Annotation of Gall and Leaf Transcripts

The Swissprot database was used to functionally annotate the transcripts from both the gall and the leaf using blastx and blastp. Using blastx, a total of 22,683 and 11,793 transcripts from gall and leaf, respectively, were annotated. Total, 15,608 and 11,967 transcripts for gall and leaf, respectively, were annotated using the blastp homology search. For the gall and leaf, respectively, it was found 20,485 and 15833 transcripts were linked to 2,924 and 2,676 different pfam domains. The outputs of blastp and blastx were both used to find GO keywords. With the greatest fractions in catalytic activity and binding category, we discovered more transcripts linked to molecular function. The most often occurring categories for cellular components were cellular anatomical activity and biological processes were metabolic process and cellular process (Figure 3.10).

Table 3.9: RNA seq assembly statistics of gall, leaf and gall-leaf combined

Contents	Leaf	Gall	Gall and leaf combined
Transcripts	46,571	73,572	101,619
Unigenes	38,098	61,390	95,229
N50 (bp)	467	486	511
Mean transcript length (bp)	492.75	499.17	524.18
Assembled bases (bp)	22,947,973	36,724,725	53,266,502

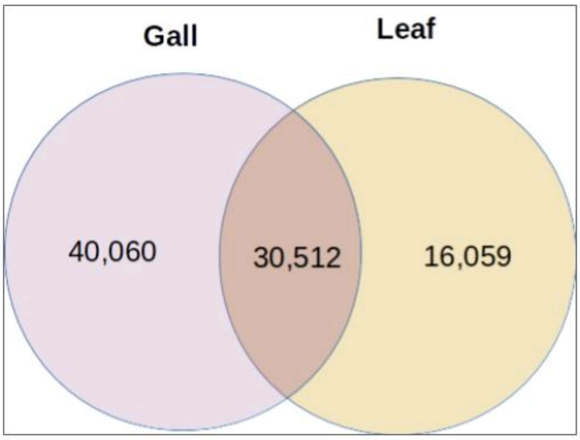


Figure 3.9: Comparison of transcripts in gall and leaf; tissue Specific and common transcripts identified in gall and leaf

Additionally, the transcripts from both gall and leaf were divided into various enzyme groups. In both the gall and the leaf, more transcripts were assigned to the family of enzymes known as transfereases, then oxidoreductases, and finally hydrolases (Figure 3.11).

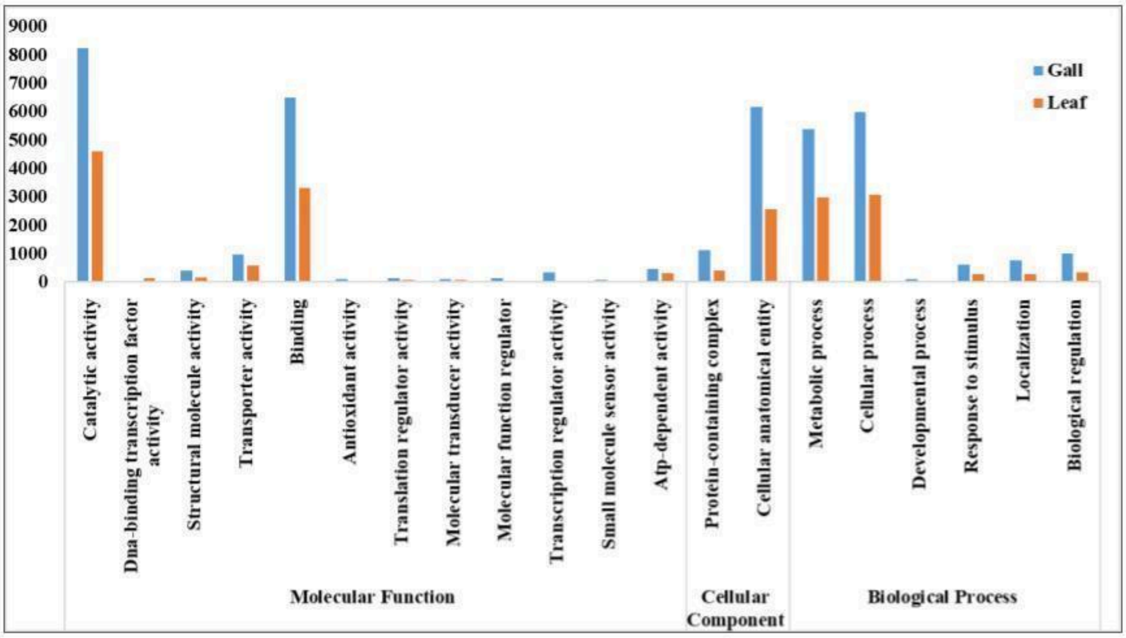


Figure 3.10: Gene Ontology (GO) distribution of *P. integerrima* gall and leaf transcripts. Y-axis represents the number of transcripts

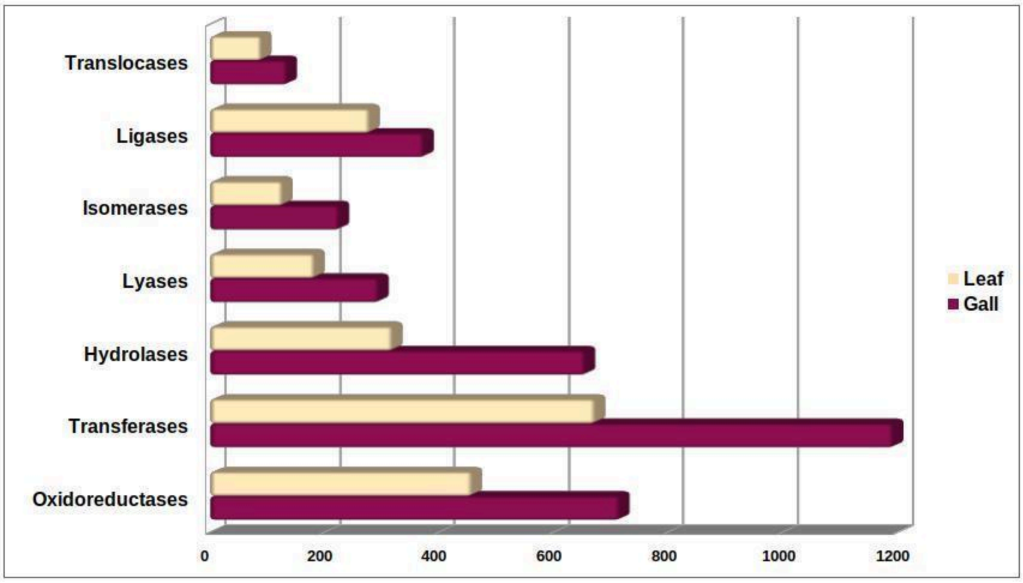


Figure 3.11: Transcripts of gall and leaf categorized into different enzyme classes. X-axis represents the number of transcripts involved in each enzyme class

3.3.7. Differential Expression of Transcripts Between Gall and Leaf

The differentially expressed genes were found using simulation after the leaf and gall data were mapped to combined assembly. Total, 17,875 genes with differential expression were found. There are several genes that are both up-regulated and down-regulated in the gall and leaf. In comparison to leaves, more genes are upregulated in galls. Using the NR database, differentially expressed genes were annotated. A total of 15,013 genes were annotated, including 9,596 and 5,547 upregulated genes from gall and leaf. Terpene synthase 10 like, lipid transferase, beta amyrin 28 oxidase, acetyl CoA, etc are highly expressed in gall compared to leaf. Stress-response A/B barrel domain-containing proteins in plants are involved in defense against biotic stress and are highly expressed in gall. In case of leaf, N-acetyl muramic acid 6-phosphate etherase, integrase, glucose/quinate/shikimate family membrane bond, *etc* genes are highly expressed in gall compared to leaf (Figure 3.12).

Using uniprot retrieve ID/mapping, the GO terms were assigned to the genes with differential expression. A higher percentage of genes, 44% in leaves and 40% in galls, are assigned to molecular functions. For the highly expressed genes in galls, 30% are associated with biological processes, and 25% are linked to cellular components. In leaves, the predominant allocation for highly expressed genes is to cellular components (32%), followed by biological processes (29%). In comparison to the leaf, the gall has a greater number of up-regulated genes associated with "response to stimuli," "developmental process," "biological control," and "biological process engaged in interspecies contact." (Figure 3.13)

3.3.8. Pathway Analysis of Differentially Expressed Genes Between Gall and Leaf

Using clusterProfiler, the pathway enrichment of genes with differential expression was performed. The highly enriched pathways include the production of amino acids, cofactors, carbon metabolism, *etc*. The investigations revealed that plant growth hormones such auxins (especially Indole-3-acetic acid), cytokinins, gibberellic acid, and brassinosteroid play a significant role in gall development. Additionally, the pathway enrichment revealed that the MAPK and signal transduction pathways for plant hormones were strongly enriched (Figure 3.14).

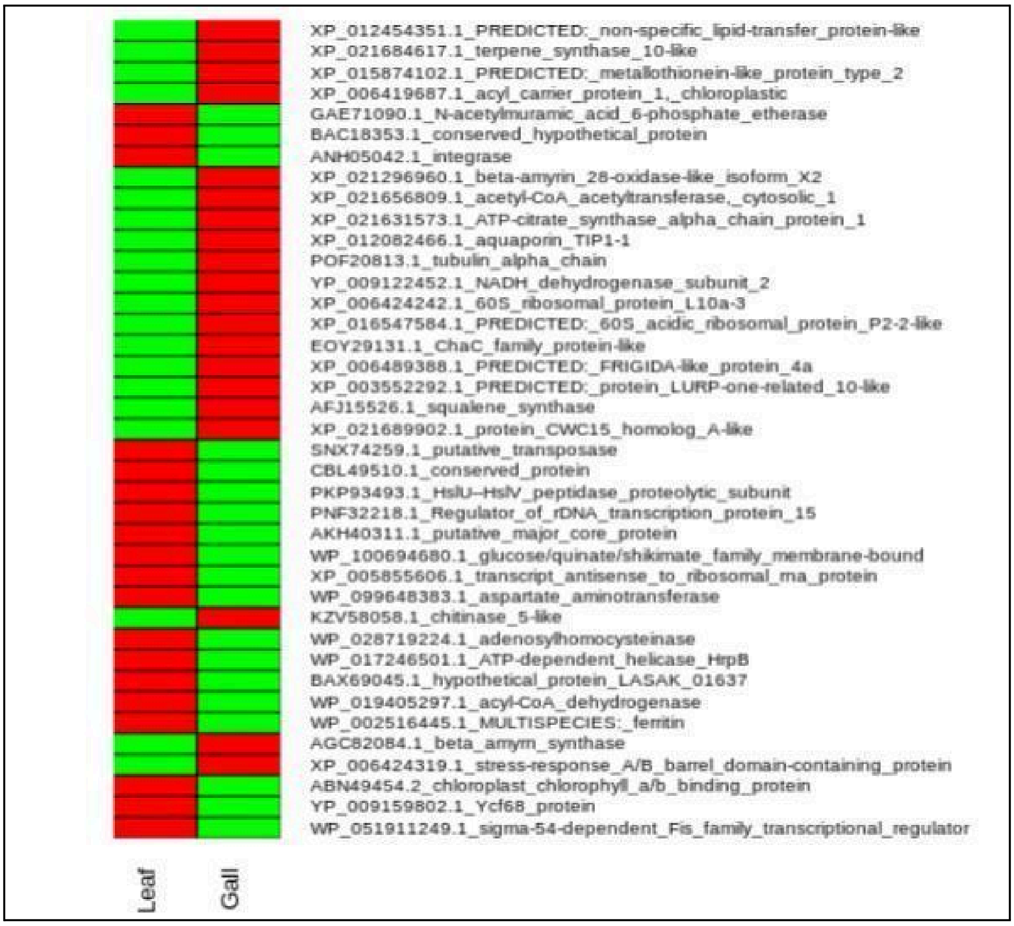


Figure 3.12: Heatmap showing top 40 highly expressed genes between gall and leaf. Red color represent high expression while green color represents low expression

Using the KEGG pathway database, differentially expressed genes from both gall and leaf were further classified into several pathways. To determine the differences between the pathways in the gall and the leaf, the pathway analysis was carried out independently for each gene that was strongly expressed in both tissues. For the gall, 399 pathways and 4,386 highly expressed genes were assigned to KO ids. For the leaf, 229 pathways and 852 highly expressed genes were assigned to KO ids. Different levels of pathways are classified by KEGG.

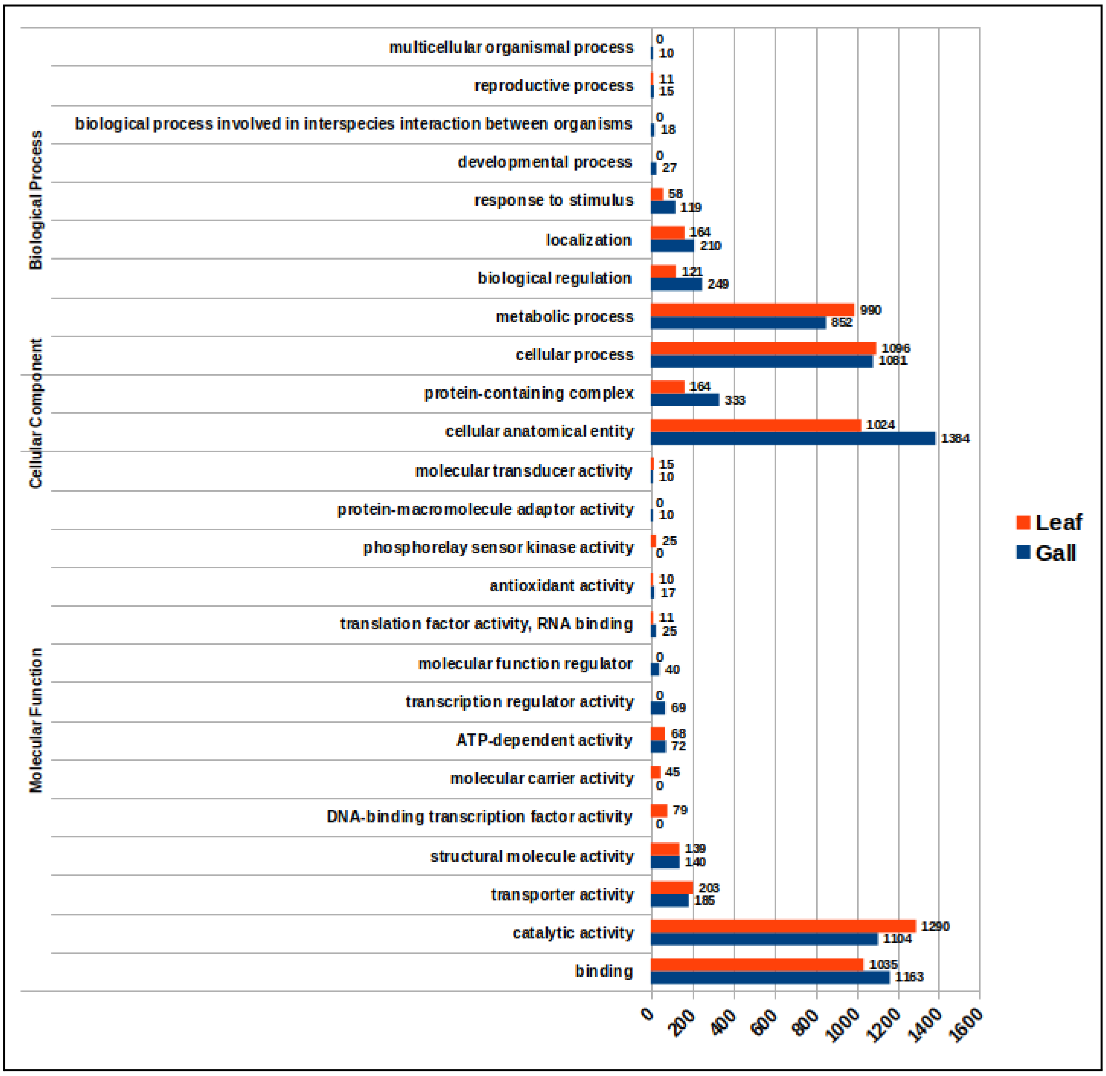
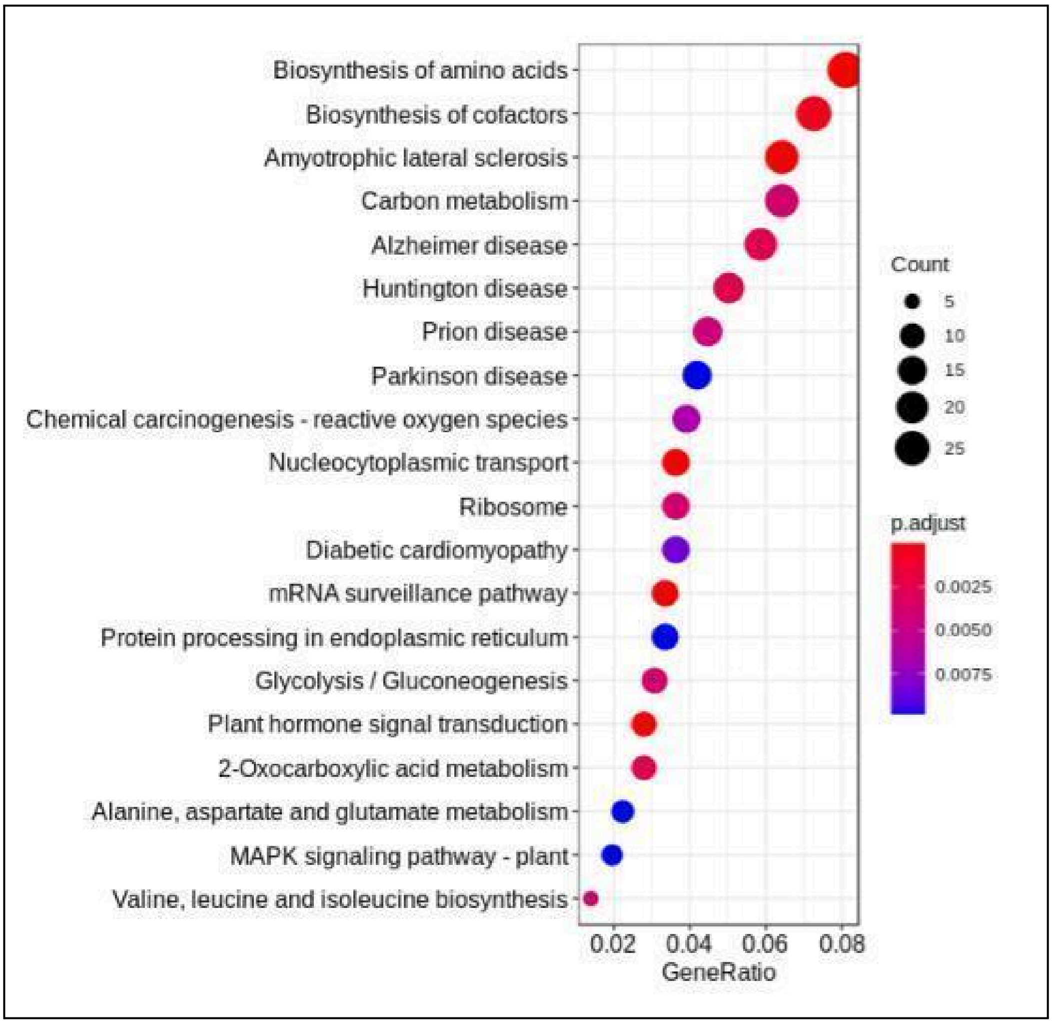


Figure 3.13: Gene Ontology (Go) terms associated with differentially expressed genes between gall and leaf. X-axis represents number of DEGs

Level 1 gene classification placed the genes in categories such metabolism, genetic information processing, cellular activities, organismal systems, and environmental information processing. The metabolism-related genes were more prevalent in both gall and leaf tissues (2,812 and 1,242 genes in gall and leaf respectively). While 82 genes were upregulated in the genetic information processing pathway from leaf and the genetic information processing pathway in gall contains 674 upregulated genes (Figure 3.15). At level 2, the global and overview maps revealed 1511 more highly expressed genes, followed by 337 for signal transduction, 298 for carbohydrate metabolism, and 266 for translation in the case of gall. Global and overarching maps (679), followed by the metabolism of carbohydrates (135), amino acids (131), and energy (75), etc. Additionally, we discovered 77 genes that are more highly expressed in gall than leaf and are involved in the synthesis of secondary metabolites (24 genes)



3.14: KEGG pathway enrichment analysis. Pathway enrichment of differentially differentially expressed genes. y-Axis indicates the pathway name, x-axis indicates the gene ratio in each of the pathways. The bubble size indicates the number of genes. The color bar indicates the corrected p-value, the blue color represents higher value, the red color represents lower value.

3.3.8.1. Differentially Expressed Genes Involved in Plant Hormone Signal Transduction Pathway

The role that plant hormones play in the growth and development of plants is well established. They have the capacity to regulate a wide range of processes, including cell division, stomata closure, shoot initiation, stem growth, and cell expansion.

According to this study, the highly expressed genes for plant signal transduction are more prevalent in the gall than the leaf (Table 3.10).

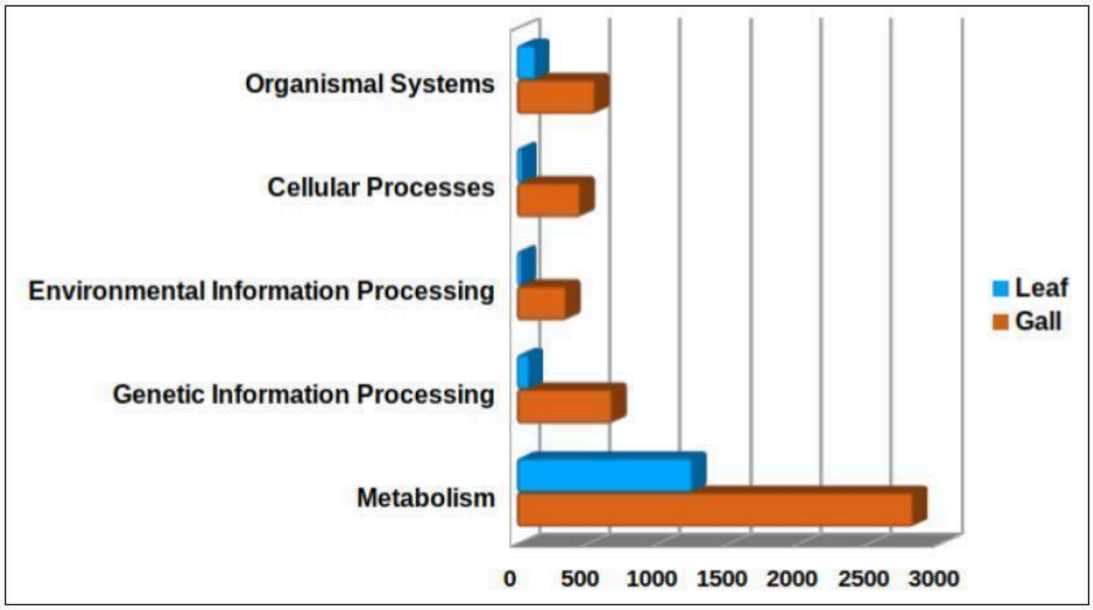


Figure 3.15: KEGG pathways of highly expressed genes from gall and leaf. X-axis represents the number of highly expressed genes among gall and leaf involved in each pathway.

Table 3.10: KEGG pathways identified from up regulated genes both gall and leaf

KEGG Pathway		**Number of Genes**	
Level I	**Level II**	**Gall**	**Leaf**
Metabolism	Global and overview maps	1,511	679
Environmental Information Processing	Signal transduction	337	23
Metabolism	Carbohydrate metabolism	298	135
Genetic Information Processing	Translation	266	40
Metabolism	Amino acid metabolism	235	131
Genetic Information	Folding, sorting and degradation	220	19

Processing			
Cellular Processes	Transport and catabolism	209	16
Cellular Processes	Cell growth and death	177	9
Metabolism	Energy metabolism	168	75
Metabolism	Lipid metabolism	139	41
Genetic Information Processing	Transcription	114	7
Organismal Systems	Environmental adaptation	113	3
Organismal Systems	Endocrine system	109	5
Organismal Systems	Nervous system	100	1
Metabolism	Metabolism of cofactors and vitamins	97	58
Organismal Systems	Immune system	93	21
Metabolism	Biosynthesis of other secondary metabolites	77	24
Genetic Information Processing	Replication and repair	74	16
Metabolism	Glycan biosynthesis and metabolism	72	13
Metabolism	Nucleotide metabolism	62	31
Metabolism	Metabolism of other amino acids	56	30
Metabolism	Metabolism of terpenoids and polyketides	55	9
Metabolism	Xenobiotics biodegradation and metabolism	42	16

Organismal Systems	Aging	33	37
Organismal Systems	Digestive system	29	2
Organismal Systems	Circulatory system	25	7
Organismal Systems	Excretory system	24	16
Organismal Systems	Development and regeneration	19	25
Cellular Processes	Cellular community - prokaryotes	15	7
Cellular Processes	Cell motility	14	4
Environmental Information Processing	Membrane transport	10	8
Organismal Systems	Sensory system	8	11

Due to their capacity to regulate cell proliferation, growth, and differentiation, phytohormones have long been the subject of speculation regarding their potential role in gall formation. In this investigation, we discovered that the auxin-responsive GH3 gene family, the SAUR family protein, the auxin influx carrier protein, the auxin-responsive protein (IAA), the transport inhibitor response 1 (TIR1), and the auxin response factor (ARF) are all substantially expressed in gall tissue. Four genes, including the cytokinin receptor (CRE1), the histidine-containing phosphotransfer protein (AHP), and the two-component response regulator ARR family (B-ARR and A-ARR) were substantially expressed in gall when it came to cytokinin signal transduction. In addition, it was discovered that the gall tissue had high expression levels of two genes involved in gibberellin signal transduction: the gibberellin receptor (GID1) and one DELLA protein. Four genes were discovered to be elevated in gall in the ABA signal transduction category, including the abscisic acid receptor PYR/PYL family (PYL), protein phosphatase 2C (PP2C), serine/threonine-protein kinase (SRK2), and ABA response element binding factor (ABF). Additionally, gall tissue had elevated expression levels of several genes connected to other plant hormone signal transduction pathways, including ethylene, brassinosteroid, jasmonate, and salicylic acid signal transduction (Table 3.11).

Table 3.11: The expression level of transcripts identified in plant hormone signal transduction

KEGG ID	KEGG Enzyme	Gall FPKM	Leaf FPKM	log2FC
K13416	BAK1	92.54	9.12	3.34
K13422	MYC2	161.38	0.50	8.33
		88.03	0.50	7.46
		81.26	0.50	7.34
K13463	COI1	1,061.96	0.50	11.05
K13464	JAZ	150.10	0.50	8.23
		90.28	0.50	7.50
		88.03	0.50	7.46
		0.50	62.66	-6.97
		10.16	54.69	-2.43
K13946	AUX1	56.43	0.50	6.82
		12.41	64.72	-2.38
K14431	TGA	3378.88	8.20	8.69
		91.41	0.50	7.51
		0.50	45.58	-6.51
K14432	ABF	1,204.16	4.56	8.05
		101.57	0.50	7.67
K14484	AUX/IAA	389.35	0.50	9.60
		360.01	0.50	9.49

		255.05	0.50	8.99
		254.77	0.50	8.99
		255.05	0.50	8.99
		134.30	0.50	8.07
		134.30	0.50	8.07
		134.30	0.50	8.07
		102.70	0.50	7.68
		102.70	0.50	7.68
		92.54	0.50	7.53
		89.16	0.50	7.48
		76.74	0.50	7.26
		67.71	0.50	7.08
		4.51	51.05	-3.50
K14485	TIR1	399.51	0.50	9.64
K14486	ARF	194.11	0.50	8.60
		185.08	0.50	8.53
		135.43	0.50	8.08
		56.78	0.50	6.83
		11.29	72.93	-2.69
K14487	GH3	4,443.10	0.50	13.12
		2,33.61	0.50	8.87

		2.26	49.23	-4.45
K14488	SAUR	375.81	0.50	9.55
		2.26	36.46	-4.01
		69.97	17.32	2.01
K14489	CRE1	209.91	0.50	8.71
		203.14	0.50	8.67
		97.06	0.50	7.60
		76.74	0.50	7.26
		72.23	0.50	7.17
		71.10	0.50	7.15
K14490	AHP	92.54	0.50	7.53
		0.50	32.82	-6.04
K14491	B-ARR	1,443.39	27.35	5.72
		217.20	0.50	8.76
		121.88	0.50	7.93
		0.50	33.07	-6.05
K14492	A-ARR	396.12	0.50	9.63
K14494	DELLA	1,994.15	35.55	5.81
		0.50	37.37	-6.22
K14495	GID2	59.81	0.50	6.90
K14496	PYR/PYL	959.27	0.50	10.91

		787.73	0.50	10.62
		390.48	0.50	9.61
		69.97	0.50	7.13
		62.07	5.47	3.50
K14497	PP2C	869.61	0.50	10.76
		202.01	0.50	8.66
		139.94	0.50	8.13
		139.94	0.50	8.13
		139.94	0.50	8.13
		139.94	0.50	8.13
		0.50	59.25	-6.89
K14498	SnRK2	371.29	0.50	9.54
		106.08	0.50	7.73
		99.31	0.50	7.63
		99.31	0.50	7.63
		0.50	69.28	-7.11
		12.41	102.10	-3.04
K14500	BSK	135.43	26.44	2.36
K14502	BIN2	0.50	124.89	-7.96
		104.96	0.50	7.71
		104.96	0.50	7.71

K14504	TCH4	158.00	0.50	8.30
		158.00	0.50	8.30
K14505	CYCD3	720.01	0.50	10.49
		119.63	0.50	7.90
		117.37	0.50	7.87
K14506	JAR1	373.59	0.50	9.55
K14509	ETR	3,718.57	0.50	12.86
		714.37	0.50	10.48
		227.97	0.50	8.83
		64.33	0.50	7.01
K14512	MPK6	79.00	0.50	7.30
K14513	EIN2	409.66	0.50	9.68
K14514	EIN3	77.20	8.20	3.23
K14515	EBF1/2	85.77	0.50	7.42

3.3.9. Differentially Expressed Genes Involved in Plant-Aphid Interactions

Gall formation is a physiological process that is incredibly complex and is connected to aphid active assault, passive defense of trees, conflict between the host and the aphid, and symbiotic association.

Massive transcripts associated with plant-aphid interactions are predicted to be concentrated in galls. Protein phosphorylation has been discovered as one of the most significant steps in the disease resistance pathway. Serine/threonine-protein kinases phosphorylate the OH group of serine or threonine residues, changing the target protein's functionality (Cao *et al*, 2011). The network of serine threonine-protein kinases in plant cells appears to function as a "central processing unit," receiving input data from receptors that detect environmental conditions, phytohormones, and other external factors, and translating that data into useful outputs, like adjustments in metabolism, gene expression, and cell growth and division. Total, 62 transcripts encoding serine/threonine protein kinases were found in the current study, demonstrating their critical function in gall development and involvement in a variety of signaling pathways (Figure 3.16). We also identified 26 upregulated genes in gall and involved in plant-pathogen interaction whereas only 2 genes were upregulated in leaf. Two genes (CDPK and Rboh) are involved in the production of Reactive Oxygen Species (ROS) which is involved in hypersensitive response (HR) to biotic stress and cell wall reinforcement.

3.3.10. Differentially Expressed Genes Involved in Reactive Oxygen Species (ROS)

Reactive oxygen species (ROS), including superoxide and hydrogen peroxide, are produced by plants in response to a variety of biotic and abiotic stressors (Goggin *et al*, 2022). ROS levels are a mediator of the accumulation of (poly)phenols and phytohormones at gall sites, which supports new cell growth pathways and metabolic adjustments that result in gall formation (Oliveira *et al*, 2015). Reactive oxygen species modulator 1, a gene that stimulates the formation of ROS essential for cell proliferation was found to be substantially expressed in gall tissue in this study. By directly controlling the transcription of the “Orbitally manifested gene 1”, R2R3-MYB, a transcription factor, contributes to the metabolism of reactive oxygen species (ROS).

3.3.11. Gene Shift in Primary Metabolism in Galls

Galls have unequal changes in the genes involved in glycolysis. In comparison to the expression level in leaf tissue, pyruvate/orthophosphate dikinase was significantly down-regulated in gall. In comparison to the leaf, the genes for alcohol dehydrogenase, fructose-1,6-bisphosphatase, phosphoglucomutase, aldehyde dehydrogenase, and others are substantially expressed in the gall. Additionally, we discovered that the genes that control the breakdown of glycans into monosaccharides or disaccharides-benefiting aphids and other glycans were more strongly expressed in the gall.

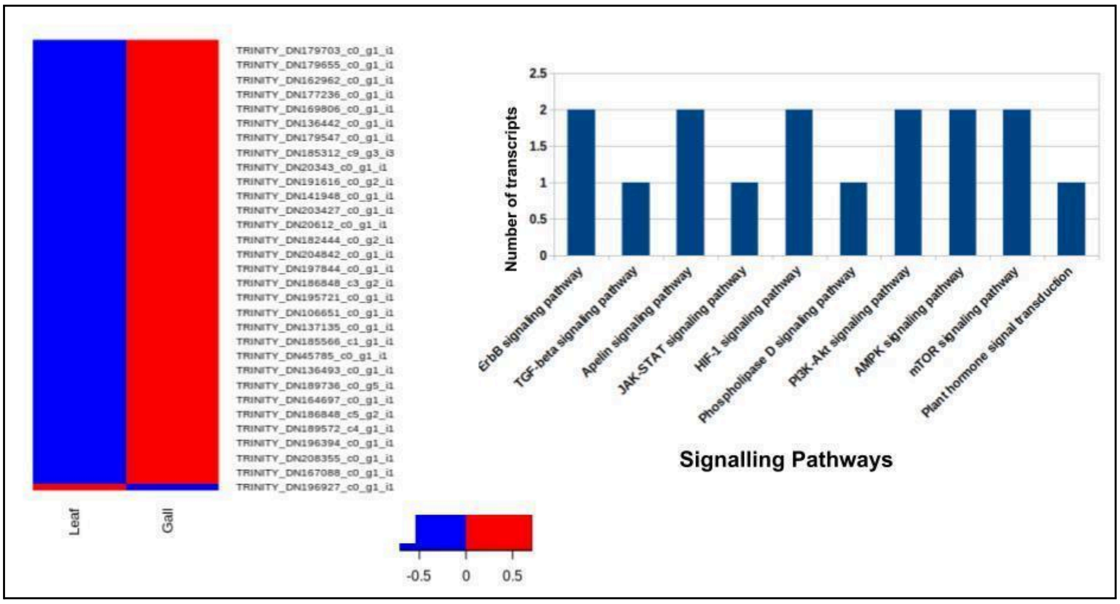


Figure 3.16: Differentially expressed genes between gall and leaf involved in plant aphid interaction and involved signaling pathways

It is interesting to note that gall has downregulated alpha-glucosidase, the enzyme that turns sucrose into fructose-D, leading to reduced amounts of fructose metabolism. Galls usually have up-regulated gene expression for important amino acid biosynthesis pathways, including those for histidine, leucine, lysine, methionine, tryptophan, and valine.

In contrast to leaf tissues, we discovered multiple distinct changes of the metabolic pathways in galls. Unbalanced alterations in the gene expression of the glycolysis-related genes were one of the modifications. The main metabolic process for proteins, lipids, and carbohydrates is known as glycolysis. Reduced simple sugar metabolism and increased NAD+/NADH buildup may be caused by the imbalanced shift of glycolysis. Simple sugar and NAD+/NADH are readily used by aphids as nutrition. This may be because aphids are likely to consume these compounds, although their digestive systems make it challenging for them to consume starch directly. Although aphids feed on plant phloem, their need for amino acids cannot be met since only a small portion of the amino acids in plant phloem can be collected. Our transcriptome investigations revealed that the genes involved in the manufacture of essential amino acids for aphids were more highly expressed in the gall than the leaf.

3.3.12. Gene Shift in Secondary Metabolism in Galls

The secondary metabolic pathways of the genes involved in phenylpropanoid pathways, which are involved in plant defense, differed most between galls and leaf tissues. Only two genes were down-regulated in galls, whereas 12 genes involved in the synthesis of phenylpropanoids were up-regulated. There were 19 genes that were up-regulated in galls, while there were none that were down-regulated, and these genes were involved in the metabolism of different phenylpropanoids, such as flavonol, isoflavonoids, and anthocyanins. Phenylpropanoids may be used by aphids as a chemical defense against possible secondary microbial infection, despite the fact that aphids have evolved to tolerate these poisonous substances through a lengthy period of coevolution. Similar to the majority of other defensive genes, such as the jasmonic acid and ethylene pathways, which were significantly expressed in gall, phenylpropanoid genes are up-regulated.

Figure 3.17 depicts a model for the mutualism between aphids and host plants. Aphids get photoassimilates, minerals, and other nutrients from host plants. The gall lessens the direct confrontation by generating unbalanced gene alterations and encouraging nutrient synthesis and metabolism, both of which the aphids may exploit for their own advantage. Aphids generate a lot of carbon dioxide, which might increase the photosynthesis of nearby leaf blades. Additionally, aphids may benefit host plants by repairing nitrogen and offering other benefits.

4. Discussion

Many scientists have been able to do in-depth research, mostly in nonmodel species, because of advancements in NGS (Lu *et al*, 2016). Understanding *P. integerrima*'s genetic makeup and doing a functional investigation of the pathways involved in the generation of the secondary metabolites that give the plant its therapeutic capabilities are both made possible by the genome and transcriptome sequencing of the plant.

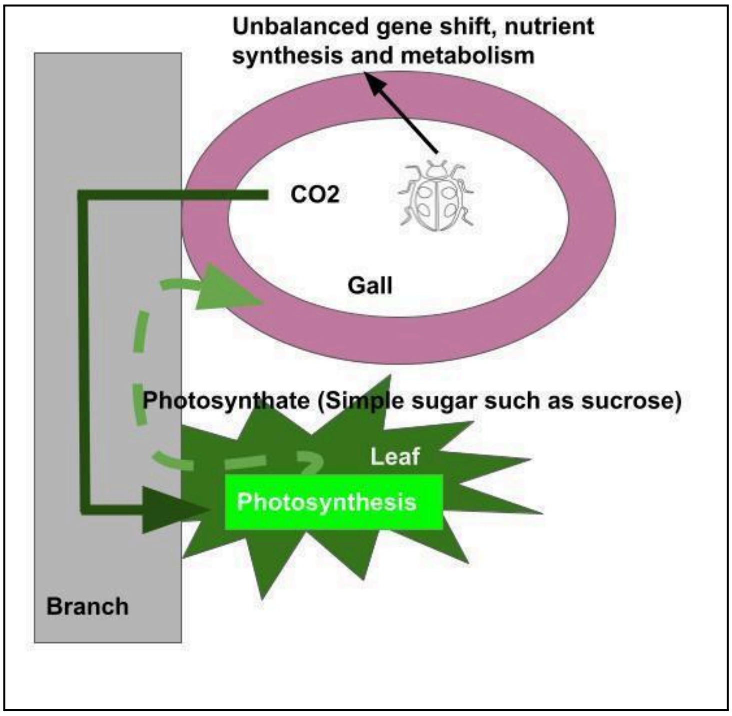


Figure 3.17: A model for mutualism between aphids and host plant

The size of the k-mers, which are substrings of length k that are created by several genome assemblers using the *de bruijn* network architecture, is the most important parameter (Chikhi and Medvedev, 2014). Here, we evaluated the assembly size for each of the different k-mers (ranging from 21 to 121, with a two-fold gap) to determine the optimal k-mer size. In order to acquire the optimum assembly utilizing short reads, we also utilized two assemblers, namely SPAdes and SOAPdenovo2. Additionally, the k-mer method has been utilized to accurately estimate genome size using NGS data without any prior knowledge of genome size (Lu *et al*, 2016).

Here, we disclose the *P. integerrima* genome's size for the first time. The genome size of *P. integerrima* was projected using the k-mer technique to be around 450 Mb, which is comparable to the genome size of *Pistacia vera* L (513 Mb) (Basr *et al*, 2003; Ziya *et al*, 2016). The high level of heterozygosity in *P. integerrima* was also discovered by the k-mer analysis, which is probably related to the dicotyledonous mating system of the genus.

It is essential to adopt a thorough strategy that incorporates both extrinsic data like proteins and transcripts as well as information intrinsic to the genome sequence (*ab* *initio*) for the prediction of more accurate gene models (Haas *et al*, 2008). Here, we used three strategies to discover the highly accurate gene models, including *ab initio*, homology-based, and evidence-based techniques. Using the Swissprot database, the genes were functionally annotated, and 97% of the predicted genes were functionally annotated. Phytochemical analysis of *P. integerrima* reveals the presence of alkaloids, flavonoids, tannins, saponins, sterols, and essential oils. Specifically, a comprehensive examination of *P. integerrima* leaves identified the existence of carotenoids, triterpenoids, catechins, and flavonoids. (Bibi *et al*, 2015). An examination of *P. integerrima*'s metabolic pathways revealed that a greater number of genes were involved in the sinthesis of secondary metabolites. Among these, it was found that there were more genes implicated in the pathways for phenylpropanoid production and flavonoids. All components of plant responses to biotic and abiotic stimuli are impacted by phenylpropanoids, and the metabolism of phenylpropanoids results in the production of a vast variety of secondary metabolites, such as lignin or flavonoid (Vogt 2010). Terpenoids are a class of organic compounds with a diverse variety of biological functions (tocopherol, brassinolide, and gibberellin, for instance, are involved in cell formation and defense) (Wang *et al*, 2019). The majority of the genes predicted to encode terpenoid-backbone biosynthesis enzymes were discovered using KEGG analysis among the many important secondary metabolites in *P. integerrima*. Genes predicted to produce farnesyl pyrophosphate (FPP) synthesizing enzymes have been successfully mined since the FPP is the major precursor for both primary (sterols, dolichols, ubiquinone, brassinosteroids, and protein prenylation) and secondary (triterpenes) terpene metabolism. Genes encoding geranylgeranyl pyrophosphate synthase (GGPS), a precursor to geranylgeranyl pyrophosphate (GGPP), a crucial component of the production of both main and specialized isoprenoid compounds (carotenoids, gibberellic acid, abscisic acid, tocopherol, diterpenes), were also discovered.

Noncoding RNAs (ncRNAs), which include long noncoding RNAs (lncRNAs), small interfering RNAs (siRNAs), and microRNAs, serve crucial regulatory roles in plant development and secondary metabolism as well as responses to biotic and abiotic stress. It has been hypothesized that five lncRNAs and more than 30 miRNAs govern the synthesis of bioactive compounds through a variety of regulatory pathways and modules (Li *et al*, 2021). We discovered miRNAs from the miR397 family, which primarily targets the laccase (LAC) genes involved in the synthesis of lignin (Huang *et al*, 2021) and miR828, which targets coding sequences of particular helix motifs in the mRNA sequences of MYB proteins and leads to the decay of MYB RNA and the production of a cascade of secondary siRNAs that depend on RNA-dependent RNA polymerase (Tirumalai *et al*, 2019).

In this work, we sequenced the mRNAs from the gall and leaf tissues in an effort to discover the genes that were differentially expressed in the two tissues of *P. integerrima*. Total 11,342 more genes were highly expressed in gall tissue than in leaves, which indicates that gall tissue has a substantial number of highly expressed genes. The fact that more genes are upregulated in gall than in leaf shows that metabolic activity is greater in gall than in leaf. When compared to leaves, it was discovered that gall contains a wide variety of enzymes. The identification of significant enzyme codes may give us information on crucial functional pathways and metabolic activities between the gall and leaf tissue of *P. integerrima* since enzymes are linked to a variety of biological processes (Schultz *et al*, 2019).

GO keywords were assigned to the highly expressed genes in both tissues. The fact that more genes from gall were assigned to GO keywords like cellular process and metabolic process shows that when gall develops, the host transcriptome and metabolome undergo extensive re-programming (Schultz *et al*, 2019). Additionally, the fact that more highly expressed genes in the gall were classified under the GO term cellular anatomical entity implies that a greater number of these genes were engaged in different aspects of gall formation.

4.1. DEGs Involved in Plant-aphid Interactions

Gall-forming insects frequently have easy access to food and protection from the elements and other natural predators. Brassinosteroid (BR) response genes are highly expressed in gall, which may indicate that steroid hormones have a role in controlling gall formation. By activating transcription and interacting with other phytohormones, BR are known to perform important roles in plant growth and stress responses (Shih *et al*, 2018). A highly complicated physiological mechanism called gall formation plays a role in aggressive aphid invasion, passive tree defense, host-versus-aphid conflict, and mutual communication between the two (Schultz *et al*, 2019). It is anticipated that gall will have a significant expression of transcripts linked to interactions between host plants and aphids. The presence of 62 serine/threonine protein kinase transcripts, which are highly expressed in gall, shows significant modifications in the disease resistance pathway that are connected to plant-aphid interactions. These transcripts were engaged in a number of signaling pathways, including the ErbB, TGF beta, and Apeline pathways.

The results of the GO analysis revealed that the gall had more up-regulated genes linked to the GO keyword "response to stimulation" than the leaf. Plants respond to a range of biotic and abiotic stressors by producing reactive oxygen species (ROS), including hydrogen peroxide and superoxide (Goggin and Fischer, 2022). The creation of new cell developmental pathways and metabolic alterations that result in gall formation are mediated by the accumulation of (poly)phenols and phytohormones at gall tissue (Oliveira *et al*, 2016). “Reactive oxygen species modulator 1”, a gene that stimulates the formation of ROS, which is essential for cell proliferation, was found to be abundantly expressed in gall tissue from this study. This gene may contribute to replicative senescence and the induction of oxidative DNA damage. This implies that ROS were involved in the formation of the gall.

According to the study conducted by Wang *et al*, 2017, which demonstrated the enrichment of plant hormone signal transduction pathways, the current analysis found a considerable enrichment of signal transduction pathways involving seven distinct hormones (Wang *et al*, 2017). Phytohormones play a vital role in both plant growth and development regulation and environmental adaptability. Auxin, cytokinin, abscisic acid (ABA), ethylene, jasmonic acid, salicylic acid, and other differentially up-regulated genes in gall tissue that are involved in hormone signal transduction and are involved in growth, development, cell division, elongation, stress response, and other processes suggest that phytohormones may be involved in gall formation.

Numerous studies (Yamaguchi *et al*, 2012; Wang *et al*, 2017; Takeda *et al*, 2021) have examined the role of phytohormones and their signal transduction in gall development. We discovered 96 transcripts that were differently expressed in gall and leaf tissue and connected to genes in the signal transduction pathways for plant hormones. In leaf, 15 transcripts were highly expressed, whereas 81 transcripts were up-regulated in gall. Gall tissue displayed an up-regulation of all transcripts involved in ethylene signal transduction pathway. This work establishes a foundation for understanding the genes that are expressed differently in gall compared to leaf, as well as the genes necessary for gall induction and development.

P. integerrima's transcriptome data are not yet available. In this study, the gene expression profiles of gall and leaf tissue were compared. Genes involved in plant-aphid interactions and signal transduction for plant hormones were shown to be overexpressed in galls compared to leaf tissue. These results open the door for more genetic research on *P. integerrima*. The absence of biological replicates poses a challenge in accurately assessing differential gene expression. In our study, we addressed this limitation by leveraging technical replicates simulated through the NOISeq-sim tool. Although computing differential expression from technical replicates deviates from the gold standard of biological replicates, we deemed it essential to utilize the available data for initial insights into gene expression dynamics. Despite these limitations, our study offers indicative insights into gene expression dynamics, especially considering that the genome under investigation was sequenced for the first time. These initial findings lay the groundwork for future studies incorporating biological replicates to validate and refine our observations.

The findings outlined in this chapter have laid the groundwork for the next chapter which describes exploration of the microbiome residing within both gall and leaf tissues of *P. integerrima*. This microbial investigation holds the promise of shedding light on the intricate relationships and ecological dynamics that underlie the unique interaction between microbes and host plants. Moreover, this study of the microbial community is expected to contribute significantly to our understanding of gall formation, a crucial aspect of this intricate ecological relationship.

CHAPTER IV

MICROBIOME OF *P. INTEGERRIMA* GALL AND LEAF

This chapter describes microbial diversity and abundance along with corresponding microbial genes and metabolic pathways associated with gall and leaf of *P. integerrima*

4. Microbiome of *P. integerrima* Gall and Leaf

4.1. Introduction

The varied population of microorganisms that coexist with plants, such as bacteria, fungus, viruses, and other microbes, is referred to as the microbiome of plants. Recent research has demonstrated that the plant microbiome is essential for plant health and growth, influencing factors including nutrient uptake, stress tolerance, and disease resistance (Berendsen *et al*, 2012; Bulgarelli *et al*, 2015). Metagenomics, transcriptomics, and metabolomics are just a few of the tools that have been used to explore the interactions between plants and microbes (Schlaeppi *et al*, 2014). To comprehend the underlying principles of plant-microbe interactions, several research have been carried out on model plant species including *Arabidopsis thaliana* and *Medicago truncatula* (Berendsen *et al*, 2012; Bulgarelli *et al*, 2015).

This growing body of research has provided new insights into the complex relationship between plants and their microbiomes. For example, a study by Schulze-Lefert *et al* revealed that the plant microbiome can influence the plant's immune response to pathogens, while a study by Bulgarelli *et al* showed that the plant microbiome can have a significant impact on plant growth and productivity (Schulze-Lefert and Panstrunga, 2011; Bulgarelli *et al*, 2015).

Overall, the research on the microbiome of plants is ongoing and continues to reveal new insights into the complex interactions between plants and microorganisms. This knowledge has the potential to be used for the development of new crop varieties and sustainable agricultural practices that can promote plant health and productivity.

Unique structures known as plant galls arise in response to a parasite such a virus, fungus, bacterium, worm, or arthropod (Redfern, 2011). The relationship between gall inducers and host plants is intricate and extends beyond parasitic/defensive relationships (Chen *et al*, 2020). Research has shown that the composition of the microbiome within galls is distinct from the surrounding plant tissue and certain bacterial taxa are enriched within galls. Faist *et al* used high-throughput sequencing to analyze the microbial communities within galls and surrounding plant tissue. The researchers found that the microbiome within galls is diverse and distinct from the surrounding plant tissue. They discovered that certain bacterial taxa, such as members of the genus *Pseudomonas*, were enriched within galls (Faist *et al*, 2016). They described the microbiota of *Vitis vinifera*'s crown gall and named the bacteria that produce IAA, including *Pseudomonas sp*. and *Enterobacteriaceae sp*. This genus is known to produce plant growth-promoting compounds, such as phytohormones, which are essential for cell division and elongation (Egamberdieva *et al*, 2017). This suggests that the microbiome may play a role in promoting gall formation.

Bacteria, fungi, and viruses only cause a very tiny fraction of galls, and *Agrobacterium tumefaciens* is widely known for producing crown galls. *Agrobacterium tumefaciens* is known to cause crown gall disease in a wide range of plants by transferring a portion of its own DNA, known as the T-DNA, into the plant genome. This transferred DNA causes the plant cells to divide and differentiate in an uncontrolled manner, leading to the formation of galls. (Peck *et al*, 2008). The galls provide a protected environment for the bacteria to grow and reproduce, and the plant cells in the galls also produce compounds that serve as a food source for the bacteria. According to the theory, the gall inducer releases phytohormones and growth regulators such auxins, cytokinins, and indole-3-acetic acid (IAA), among others, to start the development of galls (Gatjens-Boniche, 2019). The role of fungi in the formation of gall in some species of oak, *Quercus* spp has been reported. Fungi in the genus *Taphrina* causes galls in oak leaves, twigs, and fruits. The fungal spores infect the oak tissue and stimulate the growth of gall tissue, which provides a protected environment for the fungus to grow and reproduce. The fungus also benefits from the plant's nutrients and compounds that are produced in the gall tissue.

In addition to these studies, other research has also shown that the microbiome can influence plant growth and development. For example, studies have shown that plant-associated microbes can promote plant growth by providing nutrients, protecting against pathogens, and regulating plant hormone levels. These findings indicate that the microbiome plays a critical role in plant health and development.

Understanding how galls arise requires a thorough understanding of microbes. Metagenomics has emerged as a powerful tool for assessing the potential of microbes and their ecological significance in host-microbe interactions (Bashir *et al*, 2014). The limits of microbe culture have been solved by the development of the metagenomic approach, a DNA-based NGS technique. This genomic approach will focus on the full range of microorganisms in the environmental samples, including both culturable and nonculturable (William *et al*, 2010). Genomic-based monitoring offers the most accurate, cost-effective, and timely insights into the composition, quantity, and characteristics of microorganisms within galls. Results are also much more exact for the evaluation and much more accurate for taxonomic identification (Knief, 2014).

Through the use of a 16S rRNA metagenomic method, the microbiota of *P. integerrima* gall was examined and compared to leaf. To our knowledge, no research has been done on the microbiota of *P. integerrima* galls. The bioinformatics method was used in order to forecast the functional profile of microbial communities within the gall and leaf tissue.

The study of the microbiome in plant gall formation is still in the early stages, but the findings so far are promising. This research has the potential to provide new insights into plant-microbe interactions and the mechanisms that underlie plant gall formation. It also highlights the importance of considering the microbiome when studying plant health and development.

4.2. Materials and Methods

4.2.1. Sample Collection

For the purpose of this study, *P. integerrima* trees were located in their natural habitat and verified by skilled taxonomists and field botanists. In the Indian state of Uttarakhand, botanists oversaw the collection of the matured gall and young leaf samples. These samples were brought to the laboratory and stored there at -80°C. The gall and leaf voucher specimens were delivered to the FRLHT Herbarium in Bengaluru.

4.2.2. DNA Extraction and PCR Amplification 16s rRNA and ITS

Using commercially available kits from companies like QIAGEN, ZYMO RESEARCH, and ThermoFisher in accordance with the manufacturer's instructions, DNA is extracted from gall and leaf samples. Nanodrop was used to verify the DNA QC. Using V13F - 5' AGAGTTTGATCCTGGCTCAG 3' and V13R - 5' TTACCGCGGCMGCSGGCAC 3' forward and reverse primers, respectively, the V3-V4 region of 16S rRNA was amplified. In a nutshell, TAQ Master mix (High-Fidelity DNA polymerase, 0.5mM dNTPs, 3.2mM MgCl2, and PCR enzyme buffer) and 10pM of each primer are used to amplify 40 ng of extracted DNA. Denaturation at 95°C for 15 seconds, annealing at 60°C for 15 seconds, and elongation at 72°C for 2 minutes for 25 cycles of PCR. The last extension was kept at 4°C for 10 minutes at 72°C.

The amplification of the ITS gene involved a PCR reaction mixture containing High-Fidelity DNA Polymerase, 0.5 mM dNTPs, 3.2 mM MgCl2, and PCR Enzyme Buffer. Each reaction utilized 40 ng of extracted DNA and 10 pM of each primer. The PCR cycling conditions comprised 25 cycles, with denaturation occurring at 95°C for 15 seconds, annealing at 60°C for 15 seconds, and elongation at 72°C for 2 minutes. Subsequently, a final extension step was conducted at 72°C for 10 minutes, followed by holding the reaction at 4°C. This protocol was employed for the amplification of the ITS gene.

4.2.3. Library Preparation and Sequencing

Ampure beads were employed to eliminate any unnecessary primers from the amplicons.Quality control for the amplified 16S rRNA and ITS sequences was performed through gel electrophoresis and Nanodrop analysis. To prepare the sequencing libraries, an additional 8 cycles of PCR utilizing Illumina barcoded adapters were carried out. Libraries were quantitated using the Qubit dsDNA high sensitivity test kit and purified using Ampure beads. Illumina Miseq with 2x300 PE v3 sequencing kit was used for the sequencing.

4.2.4. Bioinformatics Analysis

4.2.4.1. Data QC and Preprocessing

Quantitative Insights Into Microbial Ecology 2 (QIIME2) v2020.6.0 was used to analyze the ITS and 16S rRNA sequencing data (Bolyen *et al*, 2019). Using the demux summarize plugin in Qiime2, the sequencing summary statistics were created, and Quality Check (QC) was carried out. The DADA2 (Callahan *et al*, 2016) plugin was used to dereplicate sequences, merge denoised paired-end reads, fix errors in marginal sequences, denoise paired-end reads, and cluster reads in precise amplicon sequence variations (ASVs; Callahan *et al*, 2017). Feature tables were used to depict the grouped sequences. The two techniques listed below were employed to eliminate ASVs: ASVs associated with mitochondria or chloroplasts were first removed, and then low-abundance ASVs with a frequency of less than 0.0005% of reads throughout the whole dataset were removed.

4.2.4.2. Taxonomy Classification

The Naive Bayes classifier q2-feature-classifier (Bokulich *et al*, 2018) in qiime2 was used to classify the taxonomy by comparing clustered sequences (i.e., feature table) to a reference database of sequences with known taxonomic composition. The SILVA (Quast *et al*, 2013) reference database for 16S rRNA data and UNITE database (version 8.2; Abarenkov *et al*, 2020) for ITS data were used to train the Naive Bayes classifier, which then identified sample sequences from our dataset. Both bacterial and fungi species were taxonomically classified using the trained classifier. Using the feature-table group approach in qiime2, the gall and leaf samples were both sequenced as duplicates and combined into a single group.

4.2.4.3. Diversity Analysis

The align-to-tree-mafft-fasttree process from the q2-phylogeny plugin was used to create the phylogenetic tree. The pipeline employs FastTree (Price *et al*, 2010) to generate phylogenetic trees and maft (Katoh *et al*, 2019; Kuraku *et al*, 2013) for multiple sequence alignment of the sequences in the feature table. Alpha diversity and beta diversity were employed to quantify the species variations within a sample and the differences in sample composition. The various matrices were utilized for beta diversity analysis (Bray-Curtis, Jaccard's, unweighted and weighted UniFrac distances), as well as alpha diversity analysis (Pielou's evenness and Shannon diversity). The beta diversity metrics were used to generate principal coordinates analysis (PCoA) diagrams using Emperor (Vázquez-Baeza *et al*, 2013). Using the qiime2 diversity plugin's beta-group-significance and beta-alpha-significance methods, the alpha and beta diversity indices were assessed for group significance.

4.2.4.4. Prediction of Metagenomic Functions of Bacterial Community Using PICRUSt2

A bioinformatic approach is utilized to estimate functional content: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt; Douglas *et al*, 2020). The sequences used in the PICRUSt investigation were initially organised into ASVs in QIIME2, utilising the Greengenes database (version 13.5) as a clustering reference and a similarity criterion of 0.99.

4.3. Results

4.3.1. Compositions of the Microbial Communities

A combined total of 763,182 raw 16S rRNA reads were obtained from both gall and leaf tissues, inclusive of replicates for each tissue type, with an average of 190,795 reads per sample. In case of ITS sequencing, a total of 1,123,760 ITS reads were obtained with an average 280,940 reads per sample (Table 4.1).

In the 16S rRNA and ITS data, 439 (with a total frequency of 171,969) and 130 (with a total frequency of 241,090) distinct ASVs, respectively, were found (Table 4.2). Filters were used to remove the ASVs with low abundance and chloroplast or mitochondrial affinities. After filtering, a total of 341 ASVs were observed for the 16S rRNA data, while 75 ASVs were identified for the ITS data. Further downstream analysis was conducted using these ASVs.

Table 4.1: ***P. integerrima*** **gall and leaf 16S rRNA and ITS raw data summary**

16S rRNA raw data summary			
sample ID	**Forward read count**	**Reverse read count**	**Total reads**
Leaf-1	125,824	125,824	251,648
Gall-2	116,166	116,166	232,332
Leaf-2	88,895	88,895	177,790
Gall-1	50,706	50,706	101,412
ITS raw data summary			
Sample ID	**Forward read count**	**Reverse read count**	**Total reads**
Leaf-1	251,117	251,117	502,234
Gall-1	143,591	14,3591	287,182
Leaf-2	93,505	93,505	187,010
Gall-2	73,667	73,667	147,334

4.3.2. Bacterial Community Structure in Gall and Leaf of *P. integerrima*

Most ASVs from 16S rRNA amplicon sequences were classified into the bacterial kingdom in both gall (99.8%) and leaf (96.8%), whereas 3.2 % (leaf) and 0.2% (gall) were unassigned (Table 4.3).

Proteobacteria, Firmicutes, Bacteroidota, Actinobacteriota, Chloroflexi, and Fusobacteriota were the leading phyla in the gall, accounting for over 100% of the total amplicon sequence variants (ASVs). *Firmicutes* (16.2%) was the next most prevalent phylum. The identified phyla in leaf *Proteobacteria, Firmicutes, Bacteroidota, Actinobacteria, Chloroflexi, Desulfobacter Ota, Cyanobacteria, Patescibacteria,* and *Planctomycetes* account for more than 95% of the total ASVs. *Proteobacteria* was the dominant phylum, making up about 60% of the total, followed by *Firmicutes* (21.8%; Figure 4.1).

Table 4.2: Summary of frequency of ASVs identified in 16S rRNA and ITS data

Feature Count		
Sample	**16S rRNA**	**ITS**
Gall-2	62,819	112,970
Leaf-1	43,342	75,594
Leaf-2	42,352	42,540
Gall-1	23,456	9,986
Frequency of ASVs per sample		
Statistic	**16S rRNA**	**ITS**
Minimum frequency	23,456.00	9,986.00
1st quartile	37,628.00	34,401.50
Median frequency	42,847.00	59,067.00
3rd quartile	48,211.25	84,938.00
Maximum frequency	62,819.00	112,970.00
Mean frequency	42,992.25	60,272.50

Table 4.3: Kingdom level abundance of bacteria identified in gall and leaf of *P. integerrima*

	Abundance (%)	
Kingdom	**Gall**	**Leaf**
Bacteria	99.82	96.8
Unassigned	0.18	3.15
Archaea	0	0.04

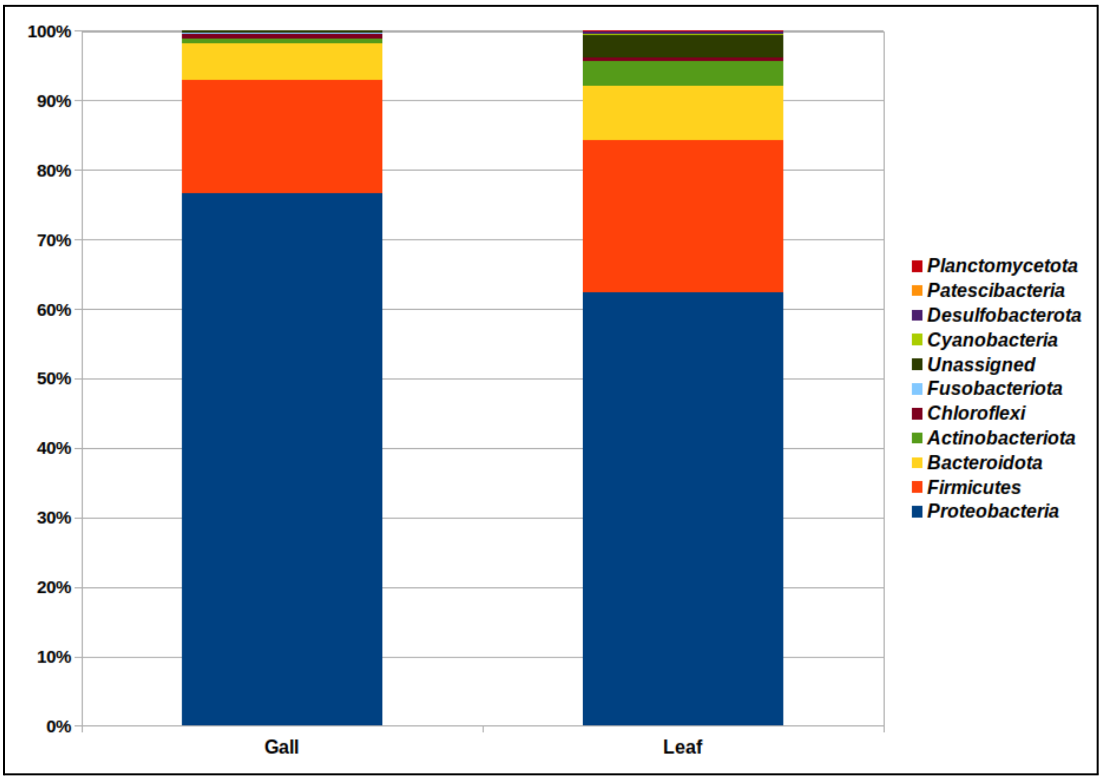


Figure 4.1: Phylum level abundance of bacteria in gall and leaf

The most common phylum in both gall and leaf is *Proteobacteria*, followed by *Firmicutes* and *Bacteroidota*, accounts for the second and third biggest phyla. *Actinobacteria* did, however, differ between the gall and the leaf, with a 3.5% abundance in the leaf and just 0.7% in the gall.

The most prevalent bacterial family in gall is *Erwiniacea*e (31.5%), followed by *Aeromonadaceae* (28.9%), *Clostridiaceae* (5.9%), *Prevotellaceae* (3.6%), *Bacillaceae* (3.2%), and other families. The major family in leaves is *Aeromonadaceae* (20.3%), which is followed by *Bacillaceae* (8.5%), *Xanthomonadaceae* (8.1%), *Enterobacteriaceae* (6.9%), *Clostridiaceae* (5.3%), etc (Figure 4.2).

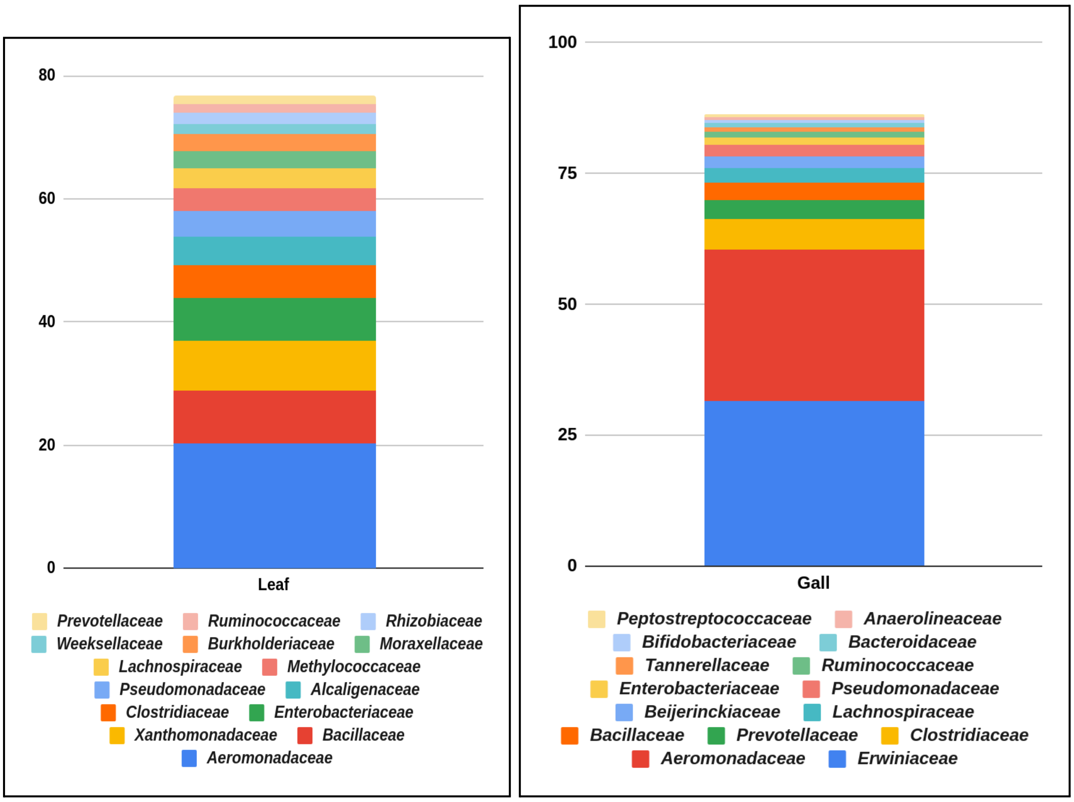


Figure 4.2: Top 15 Family level abundance of bacteria in gall and leaf. Y axis represents abundance (%)

Gall and leaf in the *Erwiniaceae* family differ significantly, with a presence in gall (31.5%) and absence in leaf. Similar large variations exist between the abundance of *Xanthomonadaceae* (absent in gall whereas present in leaf with an abundance of 8.1%), *Alcaligenaceae* (absent in gall whereas present in leaf with an abundance of 4.7%), *Methylococcaceae* (absent in gall whereas present in leaf with an abundance of 3.6%).

In leaf samples, only 51% of the ASVs were classified at the genus level, whereas 61% of the ASVs were categorized in gall samples. The predominant genus in both gall (28%) and leaves (20%) is *Aeromonas*. *Bacillus* ranks as the third most prevalent genus in gall (3.2%), while it holds the position of the second most common genus in leaves (8.5%). In comparison to gall, where it is the second most prevalent genus, *Clostridium sensu stricto* 13 is the third most numerous genus in leaf. The fourth most prevalent bacterial genus in both gall (2.1%) and leaf (4%) is *Pseudomonas* (Figure 4.3).

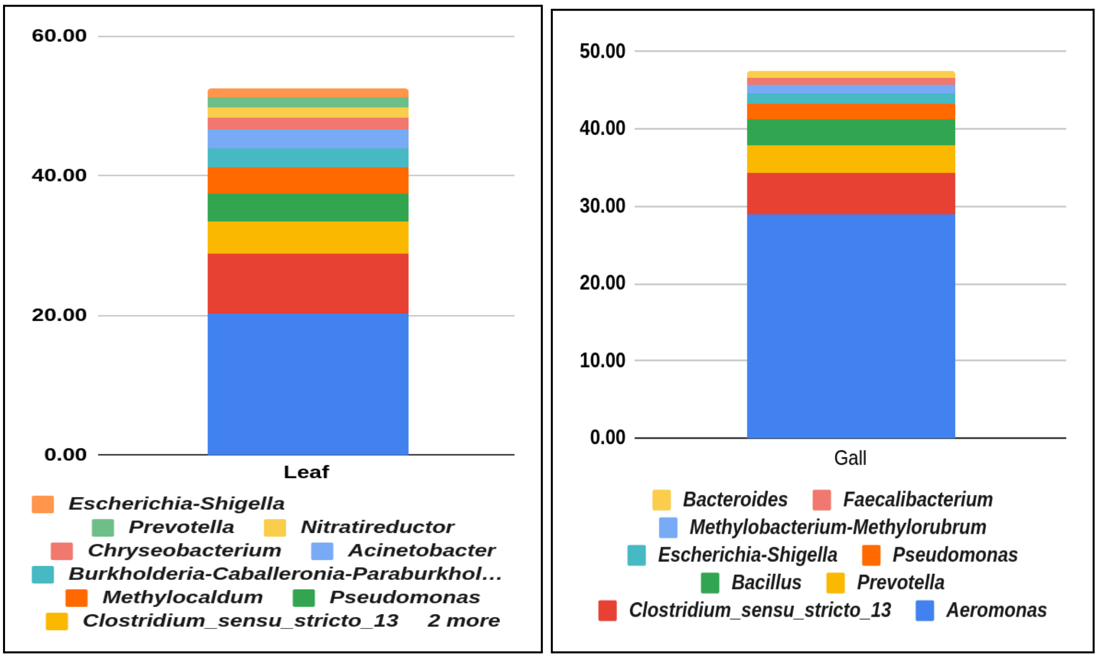


Figure 4.3: Genus level abundance of bacteria in gall and leaf. Y-axis represents abundance (%)

At the species level, 25% and 15% of ASVs in leaf and gall, respectively, were classified. With a 6.8% abundance, *Bacteroides xylanisolvens* is the main bacterial species found in gall. With a 3.9% abundance, *Acinetobacter towneri* is the most abundant bacterial species in leaf, while it is the second most abundant bacterial species in gall (Table 4.4).

Table 4.4: Top 10 species level abundance of bacteria in gall and leaf

Species	**Gall**	**Leaf**
Acinetobacter towneri	2.12	3.96
Trichuris trichiura	0.18	3.15
Bacteroides xylanisolvens	6.84	1.49
Candidatus obscuribacter	0.49	0.71
Dysgonomonas mossii	0.36	0.61
Clostridium tepidum	0.00	0.59

4.3.3. Fungal Community Structure in Gall and Leaf of *P. integerrima*

In both gall and leaf, all ASVs from ITS sequences were categorized into the fungal kingdom. *Ascomycota* comprises 97.7% of the gall and 99.8% of the leaf samples, while *Basidiomycota* accounts for 1.9% of the gall and 0.1% of the leaf samples. These two phyla, together, represent nearly 100% of the ASVs. (Figure 4.4).

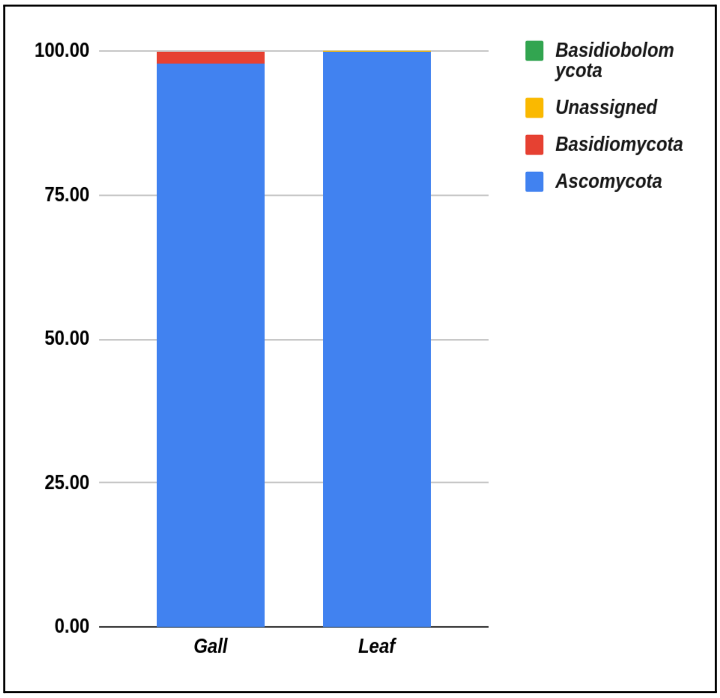


Figure 4.4: Phylum level abundance of fungus in gall and leaf. Y axis represents abundance (%)

The *Glomerellacea*e fungus family has a 96% prevalence in galls compared to a 0.5% abundance in leaves, making it the most prevalent fungal family in gall. In terms of leaf, *Didymellaceae* (23.9%), *Mycosphaerellaceae* (6.8%), *Aspergillaceae* (2.6%), and *Rutstroemiaceae* (56.5%) are the next most prevalent families (Figure 4.5).

In comparison to leaves, where more than 75% of the fungal ASVs were categorized at the genus level, very few fungal ASVs overall were found in galls. The most prevalent fungus genus in gall is *Curvibasidium* (1.7%), followed by *Aureobasidium* (0.48%). The most common fungus found in leaves is *Lambertella* (56.5%), which is followed by *Nothophoma* (17.6%), *Septoria* (5.1%), *Lophiostoma* (1.9%), *Penicillium* (1.9%), etc (Figure 4.6).

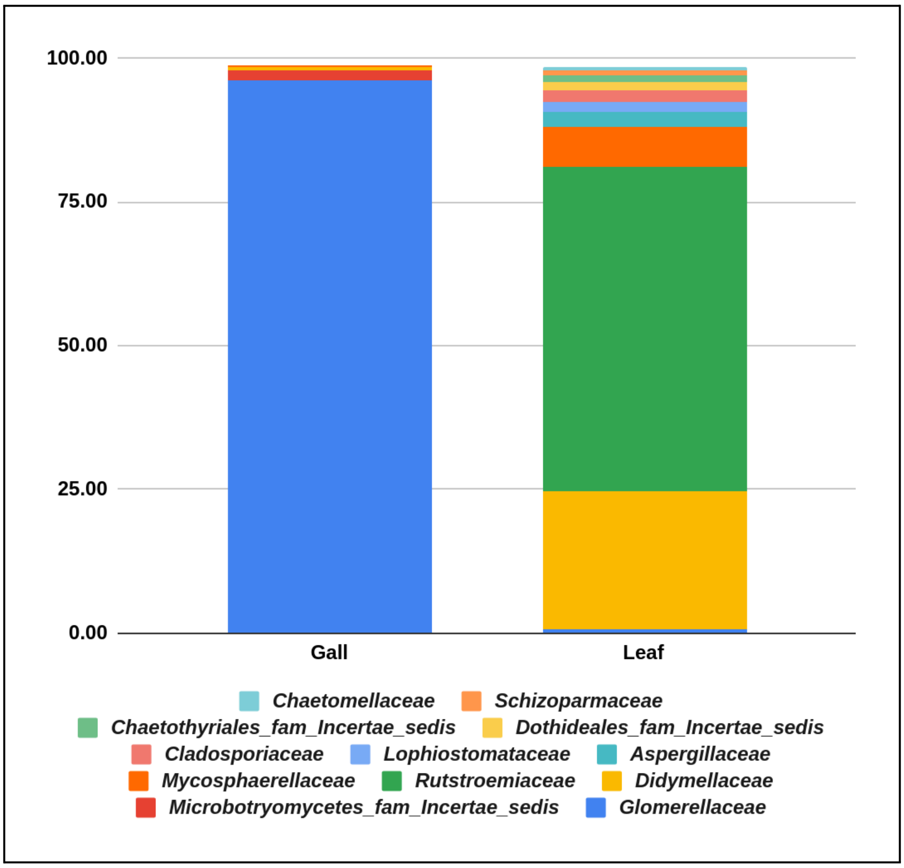


Figure 4.5: Family level abundance of fungus in gall and leaf. Y axis represents abundance (%)

Every major genus of fungus found in leaves has a very low prevalence in gall. Only a small number of ASVs were classified at the species level of taxonomy. However, in leaves, 30% of the ASVs were classified at the species level. With a 17.6% abundance, *Nothophoma variabilis* is the most common fungus, followed by *Septoria cretae* (5.1%), *Lophiostoma fuckelii* (1.9%), *Selenophoma mahoniae* (1.4%), etc (Figure 4.7).

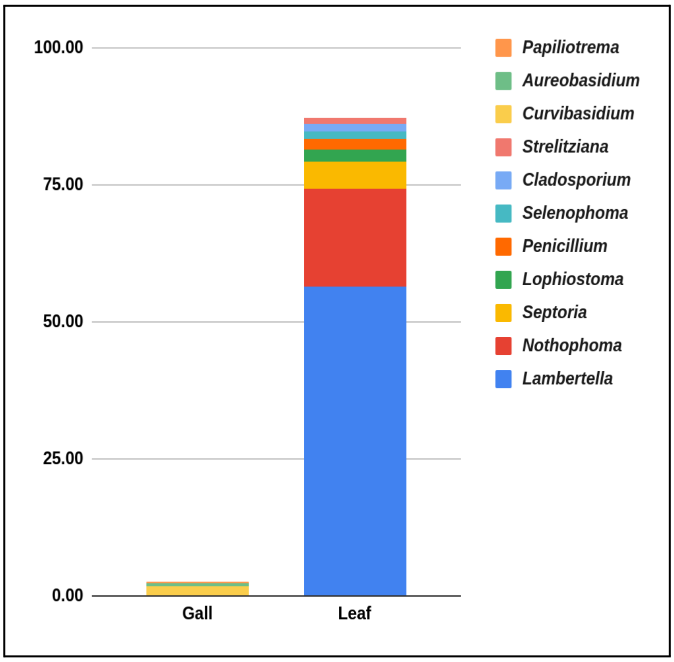


Figure 4.6: Genus level abundance of fungus in gall and leaf. Y axis represents abundance (%)

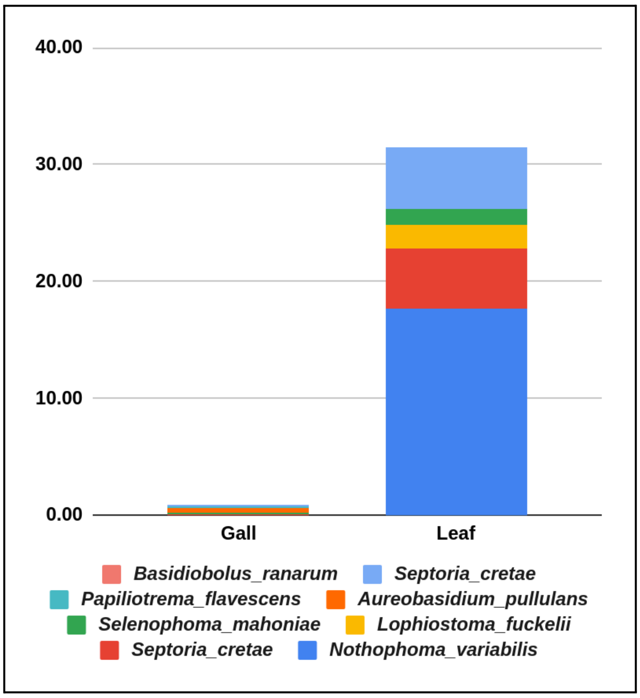


Figure 4.7: Species level abundance of fungus in gall and leaf. Y axis represents abundance (%)

4.3.4. Diversity Analysis

Alpha and beta diversities were calculated to better understand the bacterial and fungal community structures in samples. Alpha diversity provides insights into the diversity of species within a single sample or habitat. It quantifies the richness and evenness of species in a specific environment. Beta diversity examines the differences in species composition between different samples or habitats. It measures the turnover or change in species composition across spatial or temporal gradients. Common metrics used to assess alpha diversity include species richness, which counts the number of different species present, and Shannon diversity index, which takes into account both species richness and evenness. The alpha rarefaction curve used to determine if the richness of the samples has been fully observed or sequenced. At some sample depth along the x-axis, the lines in the figure appear to "level out" (i.e., approach zero slope), suggesting that collecting more sequences beyond that sampling depth is unlikely to provide new characteristics signaling the adequate sampling depth (Figure 4.8 and 4.9).

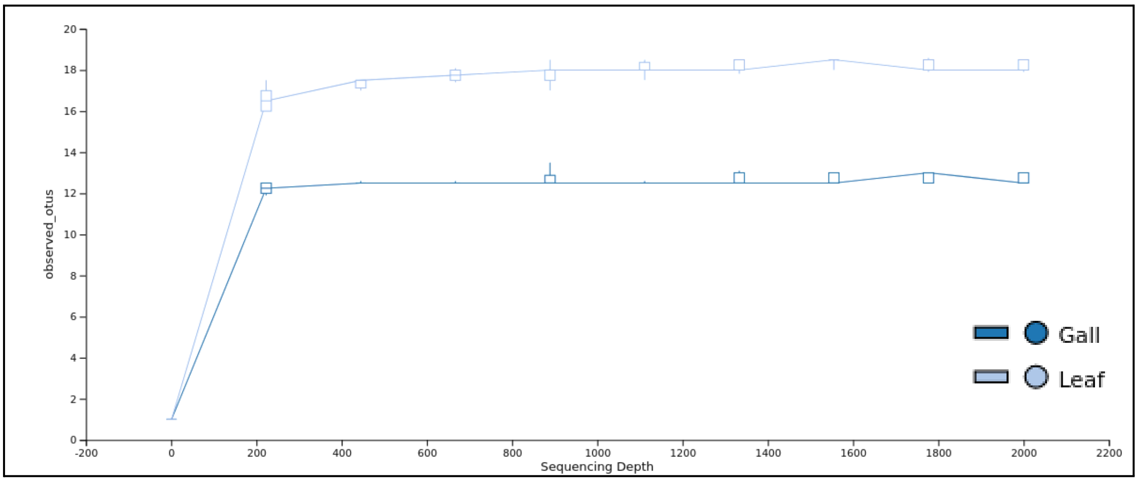


Figure 4.8: Rarefaction curves of gall and leaf bacterial species (16S rRNA). The number of various Operational Taxonomic Units (OTUs)/ ASVs as a function of the quantity of sequences examined is displayed.

Microbial diversity was further evaluated employing two metrics, namely Shannon and Pielou indices. Shannon diversity indices were utilized to quantify the richness and evenness of microbial populations. The results indicated higher bacterial diversity in the leaf sample compared to the gall for the 16S analysis. In contrast, for the ITS analysis, no significant difference in fungal diversity was observed between the gall and leaf samples (Figure 4.10). Further investigation into microbial distribution evenness was conducted using Pielou's evenness index. Notably, the leaf sample exhibited greater evenness in bacterial communities compared to the gall, while Pielou's index for fungal communities showed a similar level of evenness between the two samples (Figure 4.11). These findings underscore the importance of considering both bacterial and fungal components when evaluating microbial diversity in gall and leaf samples, revealing nuanced patterns that highlight differential influences on bacterial and fungal communities within these ecological niches.

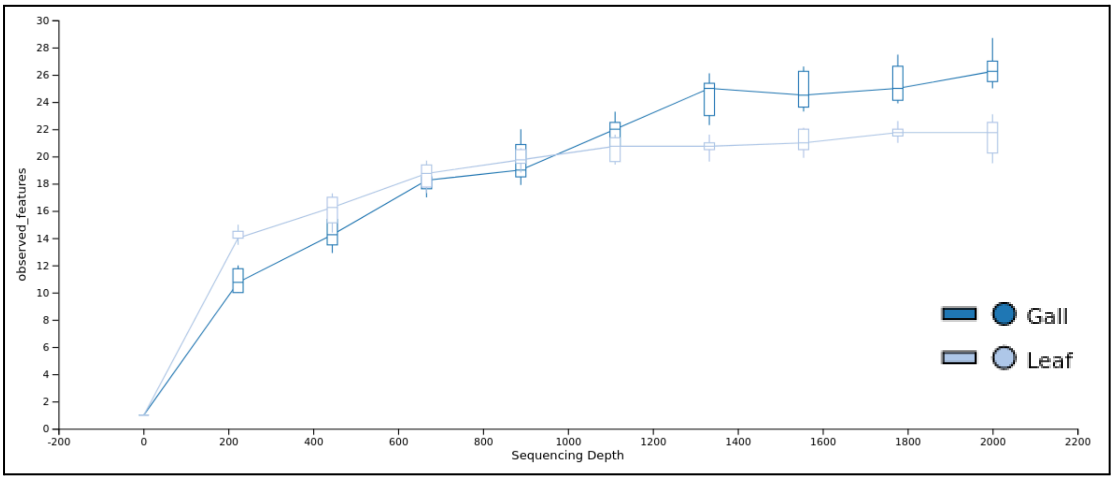


Figure 4.9: Rarefaction curves of gall and leaf fungal species (ITS). The number of various Operational Taxonomic Units (OTUs)/ ASVs as a function of the quantity of sequences examined is displayed.

For beta diversity weighted and unweighted UniFrac metrics to compare the similarity or dissimilarity of microbial communities across different samples or groups (Gall and Leaf). The main difference between these two metrics is how they take into account the evolutionary distance between different taxa. The unweighted UniFrac metric is based on the presence or absence of specific taxa in different samples. It calculates the proportion of shared taxa between two samples and the proportion of unique taxa in each sample. The weighted UniFrac metric, on the other hand, takes into account the evolutionary distance between different taxa and the relative abundance of those taxa in each sample. It calculates the proportion of the total evolutionary distance between taxa that is shared between two samples and the proportion that is unique to each sample.

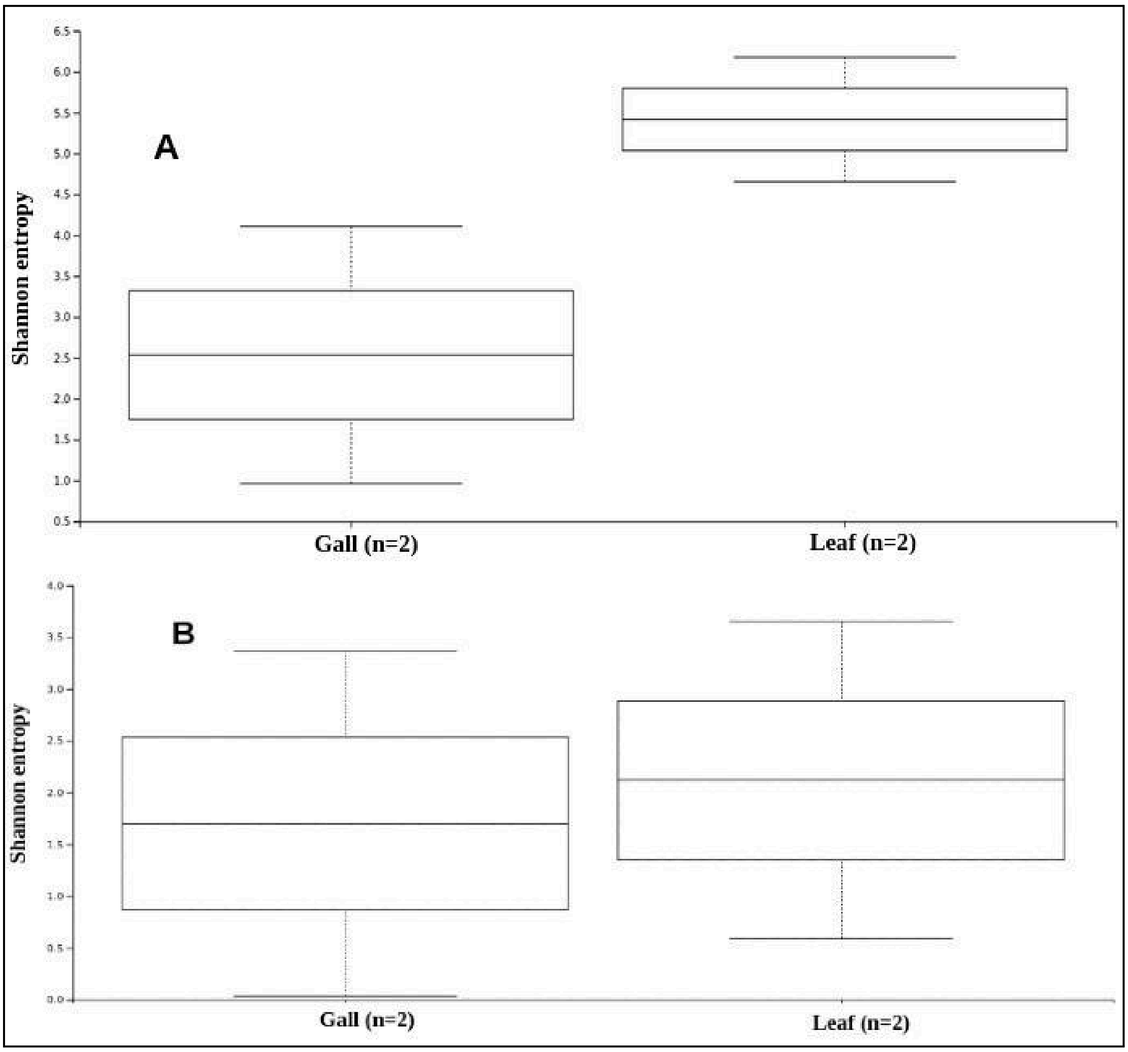


Figure 4.10: Alpha diversity of gall and leaf samples. Alpha-diversity, measured by Shannon diversity index, is plotted for gall and leaf samples. The line inside the box represents the median, while the whiskers represent the lowest and highest values within the 1.5 interquartile range (IQR). A - Bacterial; B - Fungal

4.3.4. Predictive Functionality from Identified Bacteria

PICRUSt2 (Douglas *et al*, 2020) was used to understand the functional (predictive) features of bacterial communities between gall and leaf. The KEGG database was used to annotate the metabolic pathways predicted from 16S rRNA gene sequences. Functional profiles were categorized into three levels: category (level 1), super pathway (level 2) and sub pathway (level 3). We identified 6942 KEGG orthologs (genes) involved in 362 pathways from both gall and leaf. According to mean abundance, the most dominant level 1 predictive functionality identified from bacterial sequences was metabolism (32.36 %) followed by genetic information processing (22.8%), environmental information processing (20.23%), cellular processes (5.76 %) and unclassified (18.78 %). At level 2, 22 super-families from different categories were identified (Table 4.5).

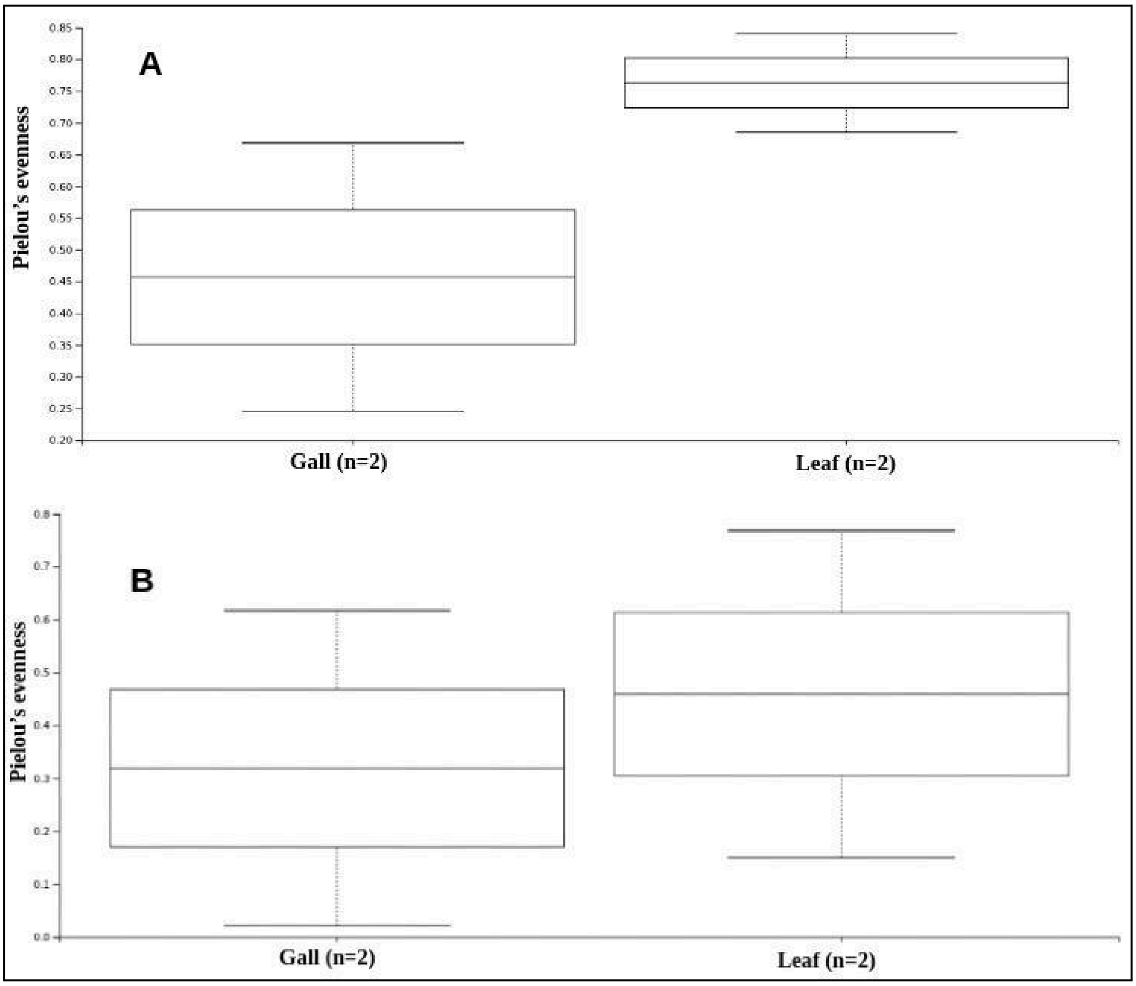


Figure 4.11: Alpha diversity of gall and leaf samples. Alpha-diversity, measured by pielou's evenness, is plotted for gall and leaf samples. The line inside the box represents the median, while the whiskers represent the lowest and highest values within the 1.5 interquartile range (IQR). A - Bacterial; B - Fungal

Table 4.5: Pathways and their respective abundance predicted from identified bacterial species from gall and leaf of *P. integerrima*

Category	**Superfamily**	**SubPathway**	**Pathway abundance (%)**	
			Gall	**Leaf**
Unclassified	Poorly characterized	Function unknown	6.23	6.15
Metabolism	Metabolism of cofactors and vitamins	Porphyrin and chlorophyll metabolism	3.04	1.57
	Enzyme families	Peptidases	2.99	2.33
	Energy metabolism	Oxidative phosphorylation	2.37	1.85
	Nucleotide metabolism	Purine metabolism	2.03	2.07
Genetic Information Processing	Translation	Ribosome	4.20	3.33
		Transfer RNA biogenesis	2.92	2.69
		Ribosome biogenesis	2.78	2.89
	Replication and repair	DNA repair and recombination proteins	2.74	2.58
	Transcription	Transcription factors	2.33	3.58
Environmental Information Processing	Membrane transport	Transporters	12.69	14.97
		ABC transporters	2.82	2.99
		Secretion system	1.76	2.19

The identified KEGG genes were further used to identify differentially abundant genes between gall and leaf. Four hundred bacterial genes were predicted and were more abundant in gall whereas 397 bacterial genes were less abundant in leaf.

Metabolic pathways has more number of highly abundant genes in leaf followed by biosynthesis of secondary metabolites, ABC transporters, biosynthesis of cofactors, etc. whereas microbial metabolism in diverse environments has a greater number of highly abundant genes in gall followed by two-component systems, carbon metabolism, methane metabolism, *etc* (Figure 4.12.).

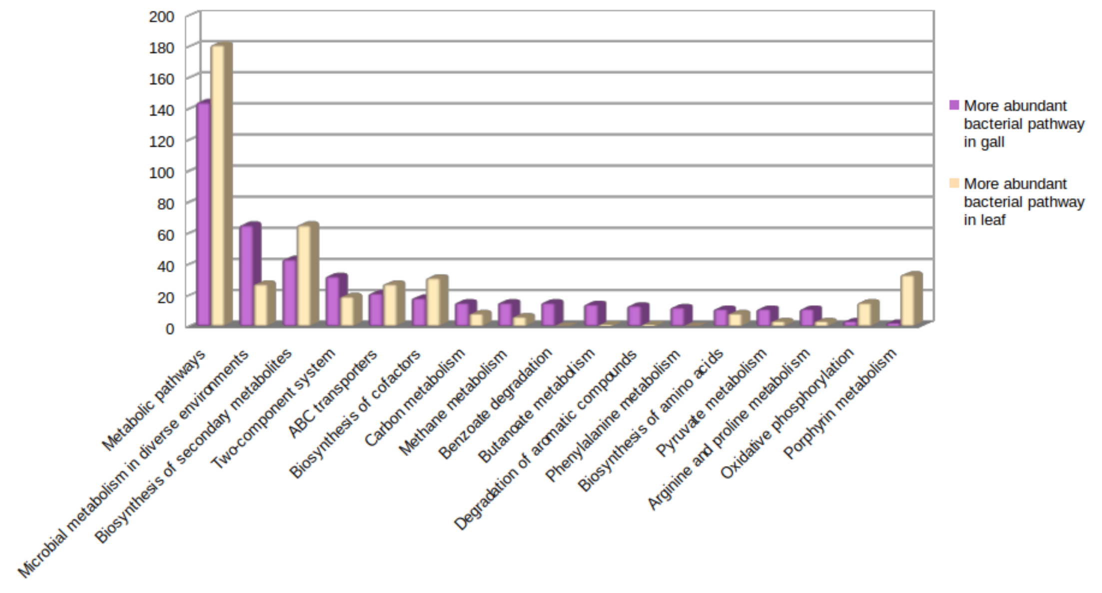


Figure 4. 12: Pathways identified with involving high abundant bacterial genes in gall and leaf of *P. integerrima*

4.5. Discussion

Gall formation in plants is often associated with microbial infection, and a number of studies have investigated the role of different types of microbes in gall formation. Bacteria are known to cause gall formation in a wide range of plant species. One of the most well-studied examples is the bacterium *Agrobacterium tumefaciens*, which causes crown gall disease in plants. *A. tumefaciens* injects a portion of its DNA into the plant cell, which leads to the activation of genes that cause the host plant cell to divide uncontrollably, resulting in the formation of a gall (Hwang and Gelvin, 2004). Other bacteria, such as *Rhodococcus fascians*, have also been shown to cause gall formation in plants. Fungi are also known to cause gall formation in plants. The fungus *Taphrina deformans* is known to cause peach leaf curl disease, which leads to the formation of galls on the leaves of peach trees (Cissé *et al*, 2013).

The role of microbes in gall formation in plants can be complex, and the exact mechanisms by which different microbes induce gall formation can vary. However, understanding the interactions between plants and microbes in the context of gall formation can help to understand the gall formation.

In this study, the microbes associated with the gall and leaf tissues were identified through 16S rRNA and ITS target gene sequencing. From this study, we couldn’t identify any microbe that is known to have a direct role in plant gall formation. However, it is important to note that there may be other microbes that have not been identified or characterized yet, and that could potentially have a role in gall formation. Nevertheless, some studies have shown that certain bacterial taxa, such as *Pseudomonas* and *Bacillus*, can have beneficial effects on plant growth and development by promoting nutrient acquisition, stimulating plant defenses, and suppressing pathogenic microbes. These beneficial bacteria are often referred to as plant growth-promoting bacteria (PGPB) and can potentially contribute to gall formation in plants.

From this study, bacterial species which are reported in plant growth and development have been identified. *Bacillus* species are known to produce a range of plant growth-promoting metabolites, such as auxins, cytokinins, gibberellins, and abscisic acid, which can stimulate plant growth and development (Kumar *et al*, 2012). *Bacillus* species have been shown to promote plant growth and increase crop yields in a variety of plant species, including wheat, maize, and tomato (Kanchiswamy *et al*, 2015). From this study *Bacillus* was identified with an abundance of 3.2% and 8.5 % in gall and leaf respectively.

Another Bacteria which is more abundant in gall and leaf is *Pseudomonas*. *Pseudomonas* species are known to produce a range of plant growth-promoting metabolites, such as indole-3-acetic acid, siderophores, and volatiles, which can stimulate plant growth and development. *Pseudomonas* species can also produce enzymes that promote nutrient uptake by plants, such as nitrogen-fixing enzymes, which can help plants access more nitrogen from the air. *Pseudomonas* is one of the known bacterial genus which synthesize IAA (Indole-3-acetic acid); an auxin.

Bacterial production of IAA can interfere with plant development by disturbing the auxin balance in plants. Production of IAA by phytopathogenic bacteria via the indole-3-acetamide pathway was first characterized in the gall-forming pathogens *Agrobacterium tumefaciens* and *Pseudomonas savastanoi* (Aragón *et al*, 2014)

One more bacterial genus which is significantly abundant in both gall and leaf was *Aeromonas*. There are limited studies on the role of *Aeromonas* specifically in plants. However, there are some studies that suggest that some species of *Aeromonas* have plant growth-promoting properties, while others may have pathogenic effects. One study investigated the effect of *Aeromonas hydrophila* on the growth of wheat plants and found that the bacterium had a positive effect on plant growth, including an increase in root length and dry weight of shoots and roots (Nawaz *et al*, 2020). The role of *Aeromonas* in plant growth appears to be context-dependent and may vary depending on the specific species and strain of the bacterium and the environmental conditions. In a study by Etminani *et al*, 61 endophytic bacteria were isolated from *Pistacia atlantica* L. and grouped based on their phenotypic properties. Further identification of 10 isolates from each group was performed by partial sequencing of the 16S rRNA gene, which showed that the isolates belonged to *Pseudomonas, Stenotrophomonas, Bacillus, Pantoea* and *Serratia* genera. The ability of these isolates for phytohormone production, such as auxin and gibberellin, siderophore production, etc was evaluated. The results showed that all strains were capable of producing the plant growth hormones auxin and gibberellin in varying amounts (Etminani and Harighi, 2018).

In this study the *Glomerellaceae* family was identified as the most abundant fungal family in gall. *Glomerellaceae* is a family of fungi that includes many plant pathogenic and endophytic species. Some members of this family are known to play important roles in plant growth and development. A study investigated the interaction between the plant *Arabidopsis thaliana* and the *Glomerellaceae* fungus *Colletotrichum tofieldiae*. The researchers found that this fungus could promote plant growth and increase the plant's resistance to drought stress. They further demonstrated that the fungus-induced growth promotion was due to the production of the plant hormone auxin by the fungus, which enhanced the plant's root growth and branching

(Díaz *et al*, 2020). This suggests that *Glomerellaceae* fungi can have a positive role in promoting plant growth and development.

Further functional profiling of identified bacterial communities was predicted using PICRUSt2. The result showed the different bacterial pathways from gall and leaf samples. The formation of galls has been associated with changes in the metabolism of vitamins and cofactors, which provide essential precursors for the biosynthesis of phytohormones and other secondary metabolites that mediate the interactions between the plant and the inducing agent.

A study on gall formation induced by the insect *Diaphorina citri* on citrus plants, it was found that the levels of vitamin B1 and its active form thiamine pyrophosphate were significantly increased in galled tissues compared to healthy tissues. This increase in thiamine levels was associated with the upregulation of genes involved in thiamine biosynthesis and transport, as well as the downregulation of genes involved in thiamine degradation. The authors suggested that the increase in thiamine levels may be involved in the regulation of plant hormone balance and the biosynthesis of secondary metabolites that promote gall formation (Nunes *et al*, 2018). Some bacteria have been found to produce phytohormones such as auxins, cytokinins, and gibberellins, or found to involve in the pathways of phytohormone biosynthesis which can cause cell division and expansion in the plant host and result in the formation of galls.

In addition, microbes can also produce enzymes that manipulate plant cell walls, resulting in the release of nutrients that support microbial growth and the formation of galls. Microbial enzymes such as pectinases, cellulases, and proteases have been shown to be involved in the breakdown of plant cell walls and the formation of galls. Moreover, some bacteria have been found to produce exopolysaccharides that play a critical role in the formation of the gall structure and help protect the microbial communities residing within the gall from host defenses.

Overall, the role of microbes in gall formation in plants can be complex, and the exact mechanisms by which different microbes induce gall formation can vary. However, understanding the interactions between plants and microbes in the context of gall formation can help to inform strategies for studying plant galls. This understanding of plant-microbe interactions in gall formation establishes the groundwork for our next chapter's investigation into global metabolite profiling. It helps to explore how these interactions affect the plant's chemistry, connecting genetic factors to its metabolic responses in *P. integerrima*.

CHAPTER V

GLOBAL METABOLITE PROFILING OF *P. INTEGERRIMA* AND *T. CHEBULA*

This chapter involves metabolites associated with gall and leaf of *P. integerrima* and *T. chebula* and their pathways specifying their involvement in gall formation

5. Global Metabolite Profiling of Gall and Leaf of *P. integerrima* and *T. chebula*

5.1. Introduction

The formation of galls is a complex and multistep process that involves significant changes in the metabolic profile of the affected plant tissue. This alteration results in the production of various secondary metabolites, such as alkaloids, flavonoids, tannins, terpenoids, and essential oils, which play a key role in the defense mechanism of the plant. The study of the metabolites involved in the formation of plant galls has received considerable attention in recent years due to their potential pharmacological activities, such as anti-inflammatory, anti-cancer, anticonvulsant, cardioprotective, and analgesic effects (Patel *et al*, 2018). These metabolites have been found to play a crucial role in the protection of the plant against environmental stress and herbivory, as well as in the regulation of gall formation (Qi *et al*, 2020).

A detailed analysis of the metabolic profile of the gall and leaf of *P. integerrima* will help uncover their biochemical composition, leading to the discovery of new phytochemicals. Metabolomics is the study of the complete set of metabolites present in a biological sample (Manchester and Anand, 2017). In recent years, it has gained significant attention as a powerful tool to study the metabolic changes that occur in plants in response to biotic and abiotic stress. One area of research that has received considerable attention in recent years is the role of metabolites in the formation of plant galls. The study of the metabolites involved in the formation of plant galls has been facilitated by the development of high-throughput analytical techniques such as liquid chromatography-mass spectrometry (LC-MS). This has allowed for the rapid identification of metabolites and the characterization of their metabolic pathways.

In this study, the metabolic constituents of *P. integerrima*'s gall and leaf were determined using a mass spectrometry-based global metabolomics approach. The metabolites identified in this study will aid in comprehending the mode of action of the gall and leaf in various Ayurvedic formulations, as well as identifying new phytochemicals for disease treatment.

5.2. Materials and Methods

5.2.1. Sample Preparation

The extraction of metabolites from 50 mg of gall and leaf of *P. integerrima* and *T. chebula* was conducted separately. The extraction process used a solvent mixture of water, methanol, and acetonitrile in a 1:2:2 ratio. The samples were crushed using a mortar and pestle and the extraction solvent was added. The mixture was agitated using vortexing for 5 minutes at room temperature, followed by sonication and centrifugation at 12,000 g for 15 minutes at 4°C. The supernatant was removed and dried using a SpeedVac, and the dried pellet was reconstituted in 0.1% formic acid in LC-MS grade water for LC-MS/MS analysis (Subbannayya *et al*, 2018) .

5.2.2. LC-MS/MS Analysis

The analysis of the metabolites was performed using a QTRAP 6500 mass spectrometer from ABSciex that was linked to an Agilent 1290 infinity II liquid chromatography system. The column used was a C18 RRHD Zorbax with a size of 20 x 150 mm and 1.8 µm particle size. The software used for the analysis was Analyst version 1.6.3 with the Analyst Device Driver, which was used to set the parameters for the analysis. The separation of the metabolites was achieved through a 20-minute liquid chromatography method, with solvent A being 0.1% formic acid in MilliQ water and solvent B being 0.1% formic acid in 90% acetonitrile at a flow rate of 0.3 mL/min. The mass spectrometry data was acquired using an Information dependent acquisition (IDA) method in low mass mode. The IDA method was built using the Enhanced mass spectra (EMS) to Enhanced product ion (EPI) modes, with the top five spectra from the EMS mode being used for analysis in the EPI (MS/MS) mode through the use of high-energy Collisionally induced dissociation (CID). The metabolite data was acquired in both positive and negative polarities at 4500 V and -4500 V respectively, with a probe temperature of 450°C. The compound parameters were set at a declustering potential (DP) of 75 V and a collision energy (CE) of 45 V, and the data was acquired in triplicates (Subbannayya *et al*, 2018).

5.2.3. Bioinformatics Data Analysis

5.2.3.1. Data Processing and Metabolite Identification

The raw data format .wiff files from the QTRAP-6500 were converted into .mzML using Proteowizard's MSConvert tool. The data was analyzed with XCMS online, using centwave parameters for feature detection with a peak width range of 10-30 seconds. The retention time was corrected with a non-linear algorithm using loess. The alignment of replicates was performed with a retention deviation limit of 12 seconds. The isotopes and adducts were annotated and identified using the METLIN database. The m/z values from the XCMS output file were searched in PlanCyc, HMDB, and KEGG databases, with [M+H]+ and [M-H]- being used as adducts for positive and negative modes. The adduct mass was modified, and the monoisotopic mass was used for searching in PlantCyc.

5.2.3.2. Pathway Analysis

Metaboanalyst 5.0 (Pang *et al*, 2021) was utilized to analyze the pathways of the identified metabolites using the *Arabidopsis thaliana* (thale cress) pathway database. A hypergeometric test was conducted to perform over-representation analysis and relative-betweenness centrality was utilized to analyze the topology of the pathways. The hypergeometric test assesses the probability of observing a specific set of metabolites within predefined pathways compared to what would be expected by chance. The metabolites were classified using classyFire (Djoumbou *et al*, 2016).

5.2.3.3. Target Analysis

The protein targets were predicted separately for exclusive metabolites of gall and leaf of *P. integerrima* using "Find my Compound's Targets" tool (http://bindingdb.org/bind/chemsearch/marvin/FMCT.jsp) in BindingDB (Nicola *et al*, 2015). The SMILES IDs for all the detected compounds were obtained through the PubChem Identifier Exchange Service (http://pubchem.ncbi.nlm.nih.gov/idexchange) and used in the study. The prediction of targets was carried out using a similarity criterion of 0.85. Accessions for all the identified target proteins were obtained and analyzed for protein-protein interaction networks using STRING (https://string-db.org/; Szklarczyk *et al*, 2019). The sources of interactions from experiments, databases, co-expression, neighborhood, gene fusion, co-occurrence, and text mining were all considered in the STRING analysis. High-confidence interactions with a minimum interaction score of 0.6 were evaluated, and the network was clustered into five clusters using the k-means clustering. GO analysis was performed for human targets using DAVID (Sherman *et al*, 2021; Huang *et al*, 2009). Pathway enrichment was done using gProfiler (Raudvere *et al*, 2019; Reimand *et al*, 2007) and the KEGG database with hypergeometric test and significantly enriched pathways were identified using p-value cutoff of 0.05.

5.3. Results

5.3.1. Identification of Metabolites from Gall and Leaf of *P. integerrima* and *T. chebula*

The global metabolite analysis of gall and leaf tissues of the medicinal plant *P. integerrima* and *T. chebula* was carried out using a mass spectrometry-based technique. Total of 4,414 non-redundant metabolites were identified from both gall and leaf of *P. integerrima* whereas 3,790 non-redundant metabolites were identified from both gall and leaf of *T. chebula*. The gall and leaf metabolite data from both the plants were collected in both positive and negative ion modes. The comparison between these datasets was given in Figure 5.1 (A & B) and 5.2 (A & B).

A total of 981 metabolites were exclusively identified in gall whereas 1,535 metabolites were exclusive for leaf samples of *P. integerrima*. The remaining 1898 metabolites were identified in both gall and leaf. When data from positive and negative ion modes were compared, it was discovered that 2,184 metabolites were identified in positive mode whereas 1,874 metabolites were identified in negative mode. Remaining 356 metabolites were identified in both positive and negative mode (Figure 5.1 A & B). In the case of *T. chebula,* 1,252 metabolites were unique to gall whereas 1,040 metabolites were unique to leaf samples. The remaining 1,498 metabolites were identified in both gall and leaf (Figure 5.2 A & B).

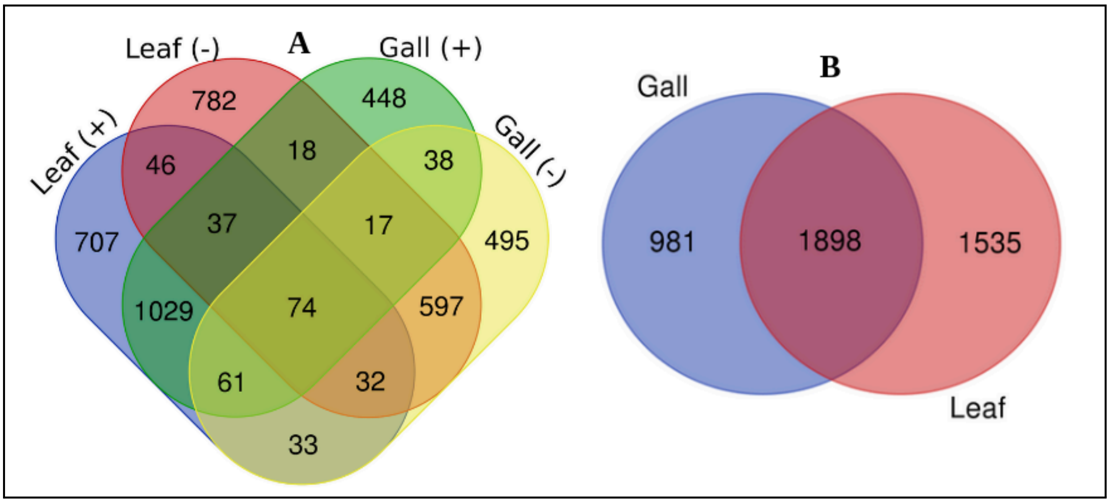


Figure 5.1: Venn diagram showing the overlapping and unique metabolites across gall and leaf of *P. integerrima.* A - Samples in positive and negative ion modes; B - Irrespective of ion modes

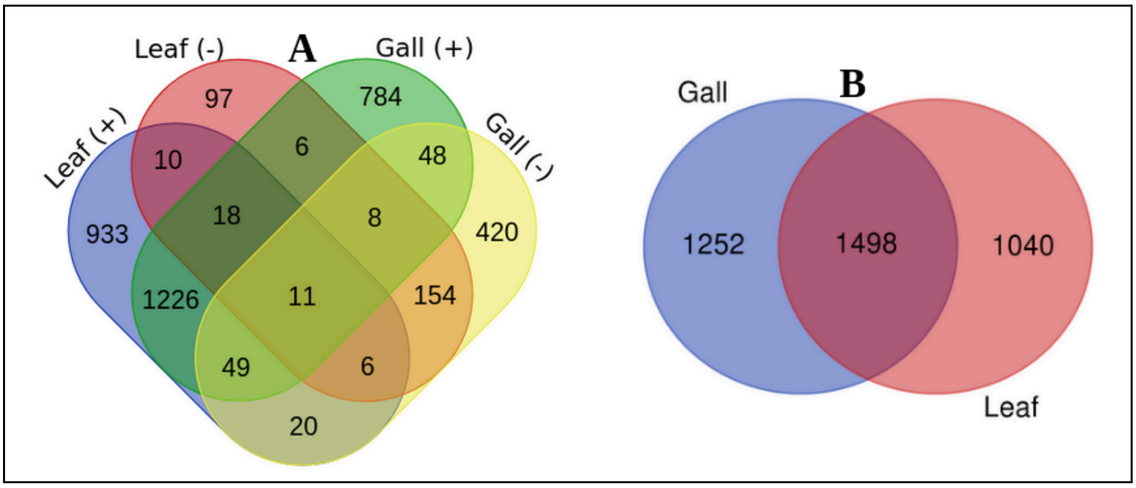


Figure 5.2: Venn diagram showing the overlapping and unique metabolites across gall and leaf of *T. chebula*. A - Samples in positive and negative ion modes; B - Irrespective of ion modes

5.3.2. Comparison of Metabolites from *P. integerrima* and *T. chebula*

The metabolites from both gall and leaf of *P. integerrima* and *T. chebula* were compared separately. There were 1,593 metabolites common between *P. integerrima* and *T. chebula* galls whereas 1,157 metabolites were unique to *T. chebula* gall and 1286 metabolites were unique to *P. integerrima* gall (Figure 5.3A). Similarly, 1,577 metabolites were common between *P. integerrima* and *T. chebula* leaves whereas 961 metabolites were unique to *T. chebula* leaf and 1,856 metabolites were unique to *P. integerrima* leaf (Figure 5.3B).

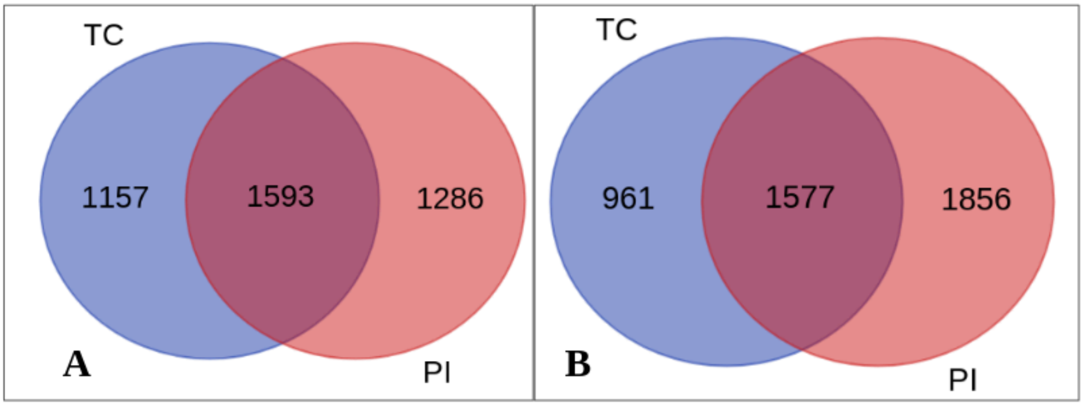


Figure 5.3: Venn diagram showing common and unique metabolites between *T. chebula* and *P. integerrima* galls and leaves; A-Gall; B-Leaf

Further metabolites from both gall and leaf from *T. chebula* and *P. integerrima* were compared together. There were 203 metabolites such as quercetin, arabidiol, octadecanol, palmitate, agmatine, *etc* which were identified between *P. integerrima* as well as *T. chebula* and unique to gall tissue. Total 243 metabolites including farrerol, rutinose, (+)-catechin, butalbital, *etc* which were identified between *P. integerrima* as well as *T. chebula* and unique to leaf tissue. Also 516 and 541 metabolites were unique to *P. integerrima* as well as *T. chebula* gall respectively. Total 511 and 870 metabolites were unique to *P. integerrima* as well as *T. chebula* leaf respectively (Figure 5.4).

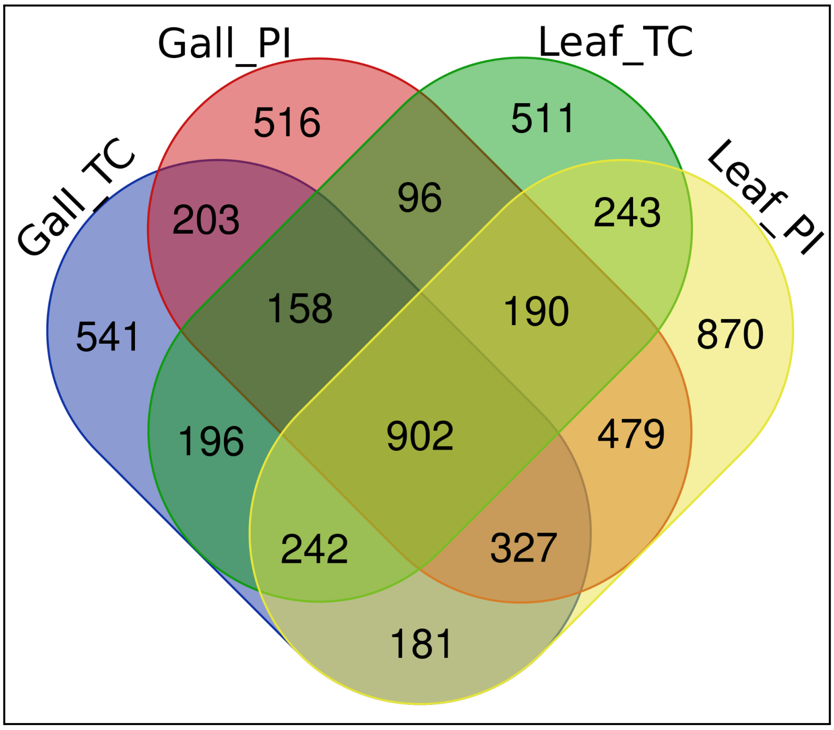


Figure 5.4: Comparison of metabolites between gall and leaf of *P. integerrima* and *T. chebula*

5.3.3. Pathway Analysis of Identified Metabolites in Gall and Leaf

In the current study, many novel metabolites were identified in *P. integerrima* gall and leaf. Various metabolite categories with putative therapeutic benefits were identified. Using Metaboanalyst 5.0, we performed pathway analysis for the metabolites found in both gall and leaf of *P. integerrima* and *T. chebula*. The top enriched pathways in gall of *P. integerrima* are linoleic acid metabolism, taurine and hypotaurine metabolism, sesquiterpenoid and triterpenoid biosynthesis, Sphingolipid metabolism etc (Figure 5.5A) whereas in leaf, taurine and hypotaurine metabolism, linoleic acid metabolism, indole alkaloid biosynthesis, sulfur metabolism, etc are enriched (Figure 5.5B). Many metabolites belong to different classes such as isoprenoids, alkaloids, phenols *etc* in *P. integerrima* gall and leaf.

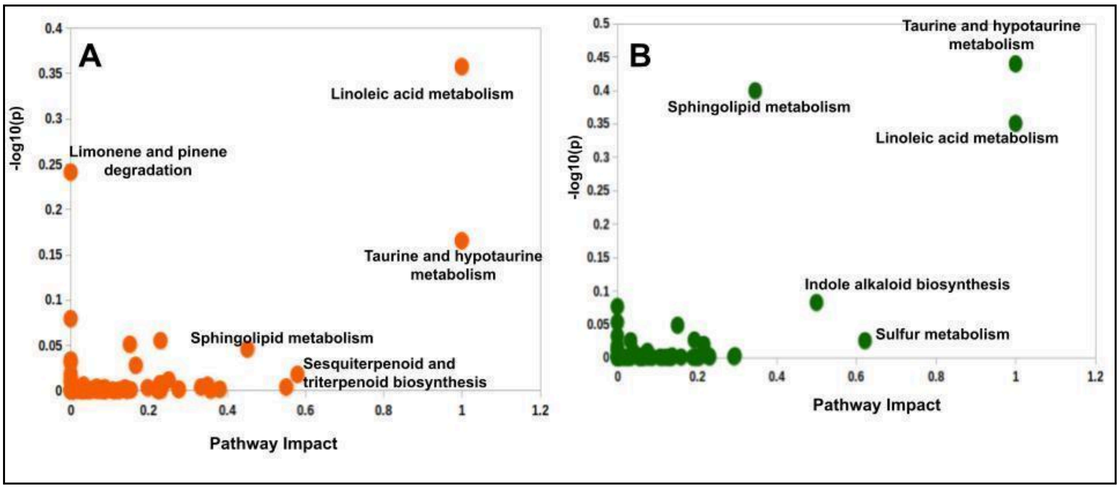


Figure 5.5: Pathway impact of metabolites identified in gall and leaf of *P. integerrima*; A-Gall, B-Leaf. Pathway impact represents the statistical significance or enrichment of the pathway

The top enriched pathways in gall of *T. chebula* are brassinosteroid metabolism, taurine and hypotaurine metabolism, sesquiterpenoid and triterpenoid biosynthesis, sphingolipid metabolism, etc (Figure 5.6A) whereas in leaf, sphingolipid metabolism, carotenoid metabolism, glycerophospholipid metabolism, sulfur metabolism, etc are enriched (Figure 5.6B). However, brassinosteroid biosynthesis and diterpenoid biosynthesis had less pathway impact. Many metabolites belong to different classes such as isoprenoids, alkaloids, phenols *etc* in *T. chebula* gall and leaf.

Further metabolites which were identified from both *P. integerrima* and *T. chebula* and were unique to gall and leaf tissue were analyzed for their role in gall formation. *I.e.,* 203 and 243 metabolites unique to gall and leaf tissues respectively and were commonly identified from both the plants. The pathway analysis of gall specific metabolites from both the plants showed enrichment of pathways such as flavone and flavonol biosynthesis, brassinosteroid biosynthesis, flavonoid biosynthesis, *etc* (Figure 5.7). In case of metabolites specific to leaf from both plants, pathways such as linoleic acid metabolism, arachidonic acid metabolism, sphingolipid metabolism, *etc* are enriched (Figure 5.8).

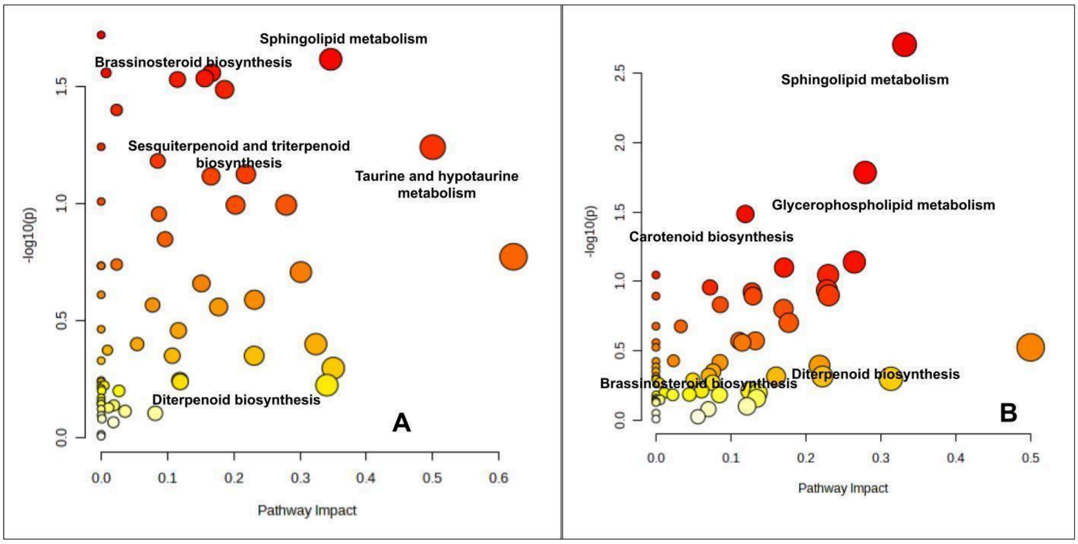


Figure 5.6: Pathway impact of metabolites identified in gall and leaf of *T. chebula*; A-Gall, B-Leaf

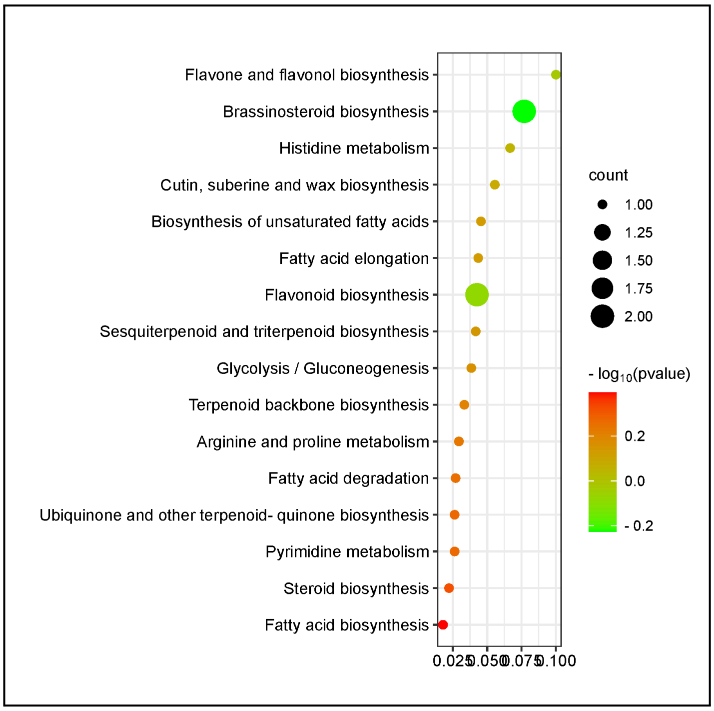


Figure 5.7: Pathways enrichment of gall specific metabolites from both *P. integerrima* and *T. chebula*

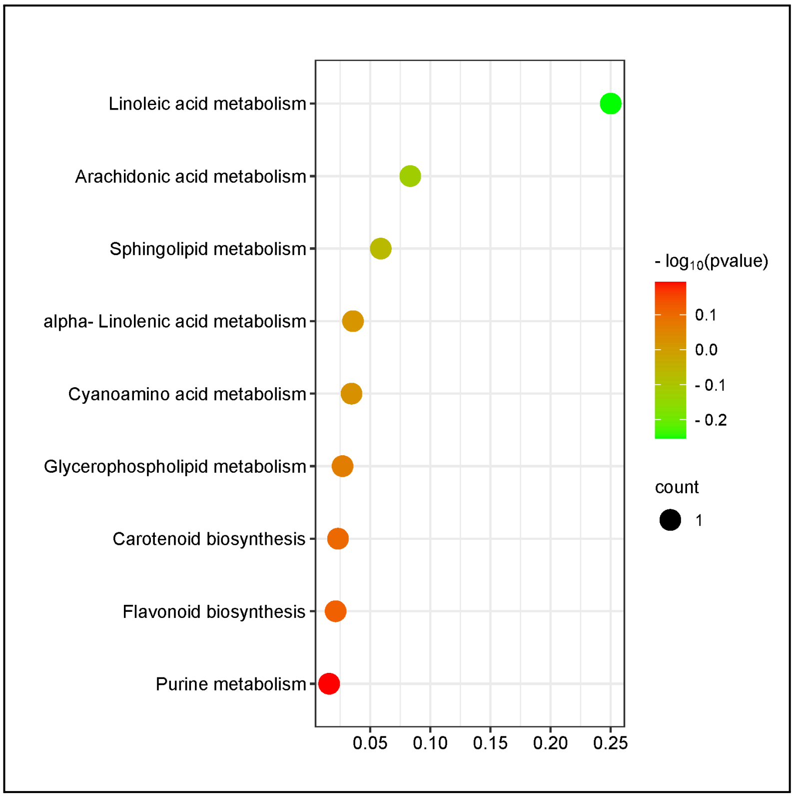


Figure 5.8: Pathways enrichment of leaf specific metabolites from both *P. integerrima* and *T. chebula*

5.3.4. Identification of Previously Known Metabolites of *P. integerrima*

Many studies have shown the activity of different metabolites from different classes extracted from *P. integerrima*. Table 5.1 summarizes the findings of these studies. We matched our findings to earlier research and discovered compounds in *P. integerrima* including quercetin, gamma terpinene, gallic acid, β-Sitosterol, β-Stigmasterol, *etc*.

Table 1: List of previously reported metabolites of *P. integerrima*

Class	Compound name	Reference
Triterpenes	Pistagremic acid	Uddin *et al*, 2011; Rauf *et al*, 2014; Rauf *et al*, 2015; Uddin *et al*, 2012; Rauf *et al*, 2016; Chauhan, Singh, & Agarwal, 2002; Joshi & Mishra, 2010; Ansari *et al*, 1993; Ullahi *et al*, 2012; Rauf *et al*, 2013; Alhumaydhi *et al*, 2021; Joshi & Mishra, 2010; Rauf *et al*, 2017; Zahoor *et al*, 2018; Mehla *et al*, 2011; Ahmad *et al*, 2010; Ulbricht, 2016
	Pistacienoic acid A	
	Pistacienoic acid B	
Flavonoids	Pistacide A	
	Quercetin	

	Luteolin	
Volatile oils	4-Carvomenthenol/1-Terpinen-4-ol	
	Acetic acid	
	β-Farnesene	
	α-Terpinolene	
	Limonene	
	α-Thujene/Origanene	
	α-Pinene	
	β-Pinene	
	α-Phellandrene/Menthadiene	
	β-Phellandrene	
	α-Terpineol	
	β-Terpineol	
	δ3-carene	
	γ-Terpenene	
	α-Ocemene	
	β-Ocemene	
Fatty acids	Palmitic acid	
	Palmitoleic acid	
	Caproic acid	
	Lauric acid	
	Myristic acid	

	Margaric acid	
	Capric acid	
	Caprylic acid	
	Oleic acid	
	Elaidic acid	
	Linoleic acid	
	Linolenic acid	
	Stearic acid	
	Arachidic acid	
	Behenic acid	
Phenolics	Pistaciaphenyl ether	
	Pistiphloro-glucinyl ester	
	Pisticiphloro-glucinyl ether	
	Ethyl gallate	
	Gallic acid	
Phytosterol	β-Sitosterol	
	β-Stigmasterol	

5.3.5. Classification of Metabolites

In the metabolite profile of gall samples from *P. integerrima*, the predominant category consists of isoprenoids (64), followed by amino acids and peptides (40), and benzenes (21). These classes represent the major components in the gall. Other classes, including flavonoids, fatty acids, and sterols, also contribute to the overall metabolite composition, with lower counts (Figure 5.9A).

Conversely, the metabolite composition in gall samples from *T. chebula* is characterized by a significant presence of isoprenoids (56) and amino acids and peptides (40). benzenes (18) are also notable components, albeit with a slightly lower count compared to *P. integerrima*. The distribution of other metabolite classes, such as fatty acids, flavonoids, and purines, varies between the two gall types, indicating potential differences in their chemical profiles (Figure 5.9B).

In the metabolite profile of *P. integerrima* leaf, the most prominent category is isoprenoids with 54 metabolites, followed by amino acids and peptides with 42 metabolites, and benzenes with 25 metabolites. These classes represent the major constituents within the leaf. Additionally, other classes like ceramides, sterols, pyrimidines, and fatty acids contribute to the overall metabolite composition, with varying counts (Figure 5.10A).

The metabolite composition in *T. chebula* leaf is characterized by a significant presence of isoprenoids with 50 metabolites and amino acids and peptides with 33 metabolites. Benzenes are also noteworthy with 15 metabolites, although slightly fewer than in *P. integerrima* leaf. The distribution of other metabolite classes, such as fatty acids, pyrimidines, and ceramides, varies between the two leaf types, indicating potential differences in their chemical profiles (Figure 5.10B).

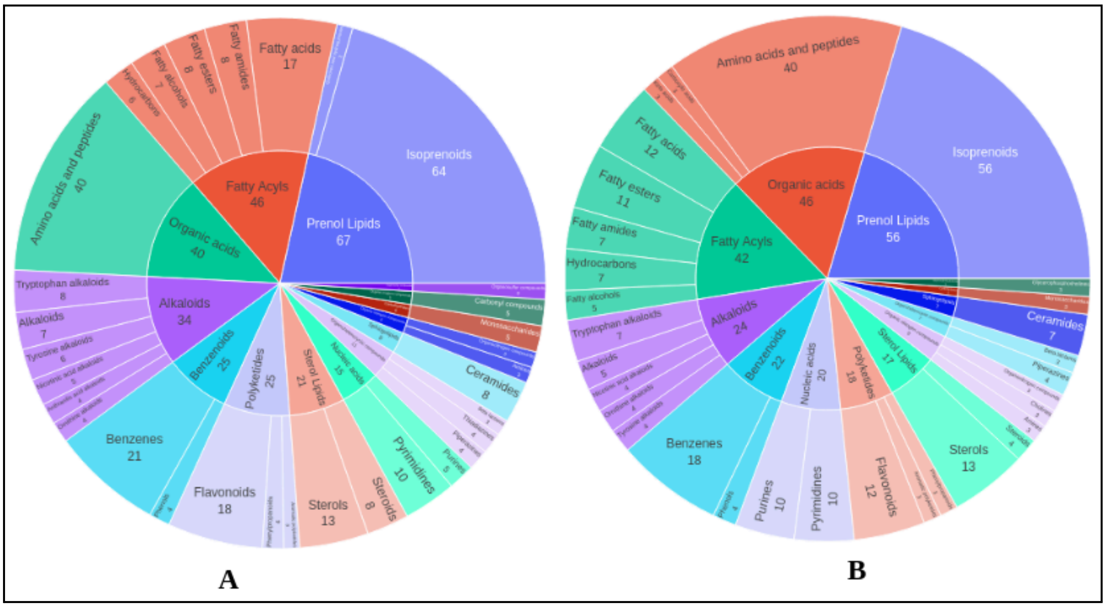


Figure 5.9: Classification of metabolites from glall samples. A - *P. integerrima*; B - *T. chebula*

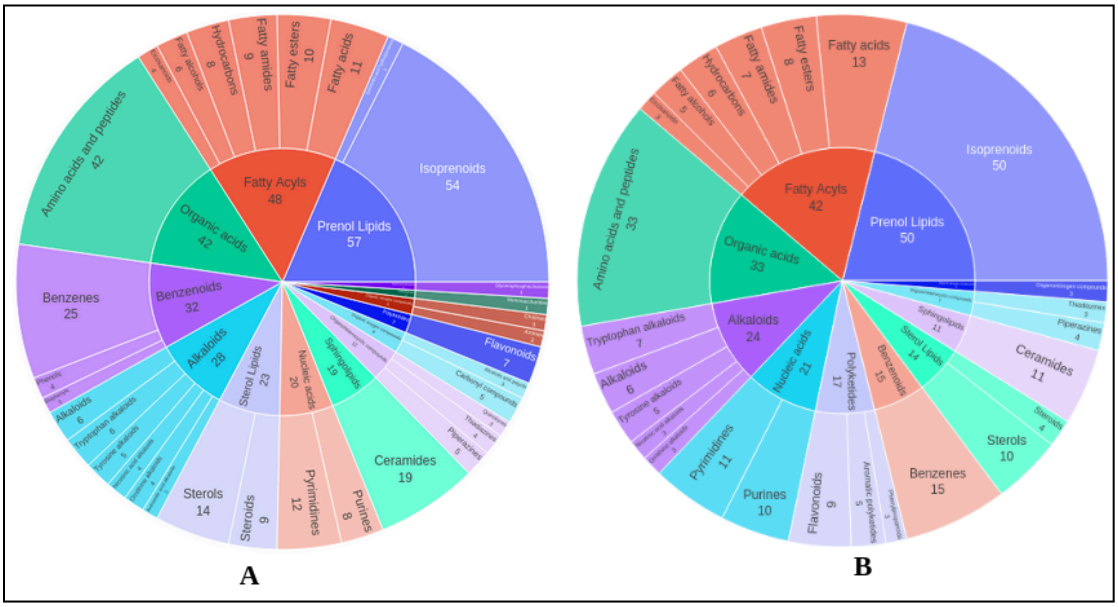


Figure 5.10: Classification of metabolites from leaf samples. A - *P. integerrima*; B - *T. chebula*

5.3.6. Prediction of Protein Targets of Metabolites Identified from *P. integerrima* Gall and Leaf

The protein targets of *P. integerrima* gall and leaf metabolites were identified using the BindingDB tool, resulting in 167 targets. Of these targets, 106 were found to be human proteins. The identified targets related to airway disease (e.g. Interleukin 13), proteins involved in the MAPK signaling pathway (such as MAPK1, ERBB2, TEK, and EGFR), and various protein kinases (e.g. Serine/threonine-protein kinase, Aurora kinase B, ULK3 kinase, and Myotonin-protein kinase).

STRING was used to assess the proteins that mapped to human proteins in order to find networks that were altered by *P. integerrima* metabolites (Figure 5.11). We also carried out gene ontology and pathway analysis for protein targets of *P. integerrima*. Several target pathways of *P. integerrima* were found to be enriched including Gap junction, PI3K-Akt signaling pathway, MAPK signaling pathway, pathways in cancer, *etc* (Figure 5.12). Gene ontology analysis showed that the protein targets of P. integerrima were involved in cellular processes, cellular anatomical entity, catalytic activity, *etc* (Figure 5.13).

5.4. Discussion

This study employed LC-MS/MS-based metabolomics to identify tissue-specific and plant-specific metabolites in the gall and leaf tissues of *P. integerrima* and *T. chebula*. Pathway analysis was conducted to elucidate pathways involved in plant gall formation. These metabolites exhibit diverse bioactivities, including anti-bacterial, anti-oxidant, anti-inflammatory, cardioprotective, anti-cancer, anti-diarrhoeal, anticonvulsant, and analgesic effects. Additionally, potential protein targets of these metabolites were explored.

Metabolites were identified from different classes in both gall and leaf. More metabolites were identified from metabolite class prenol lipids, fatty acyls, organic acids, steroids and steroid derivatives, flavonoids, alkaloids, *etc* classes from both gall and leaf tissues of *P. integerrima* and *T. chebula*. Metabolites such as prenol lipids, fatty acyls, carboxylic acids and derivatives, benzene and substituted derivatives, organonitrogen compounds, steroids and steroid derivatives, organo oxygen compounds, and flavonoids have been shown to play a role in gall formation.

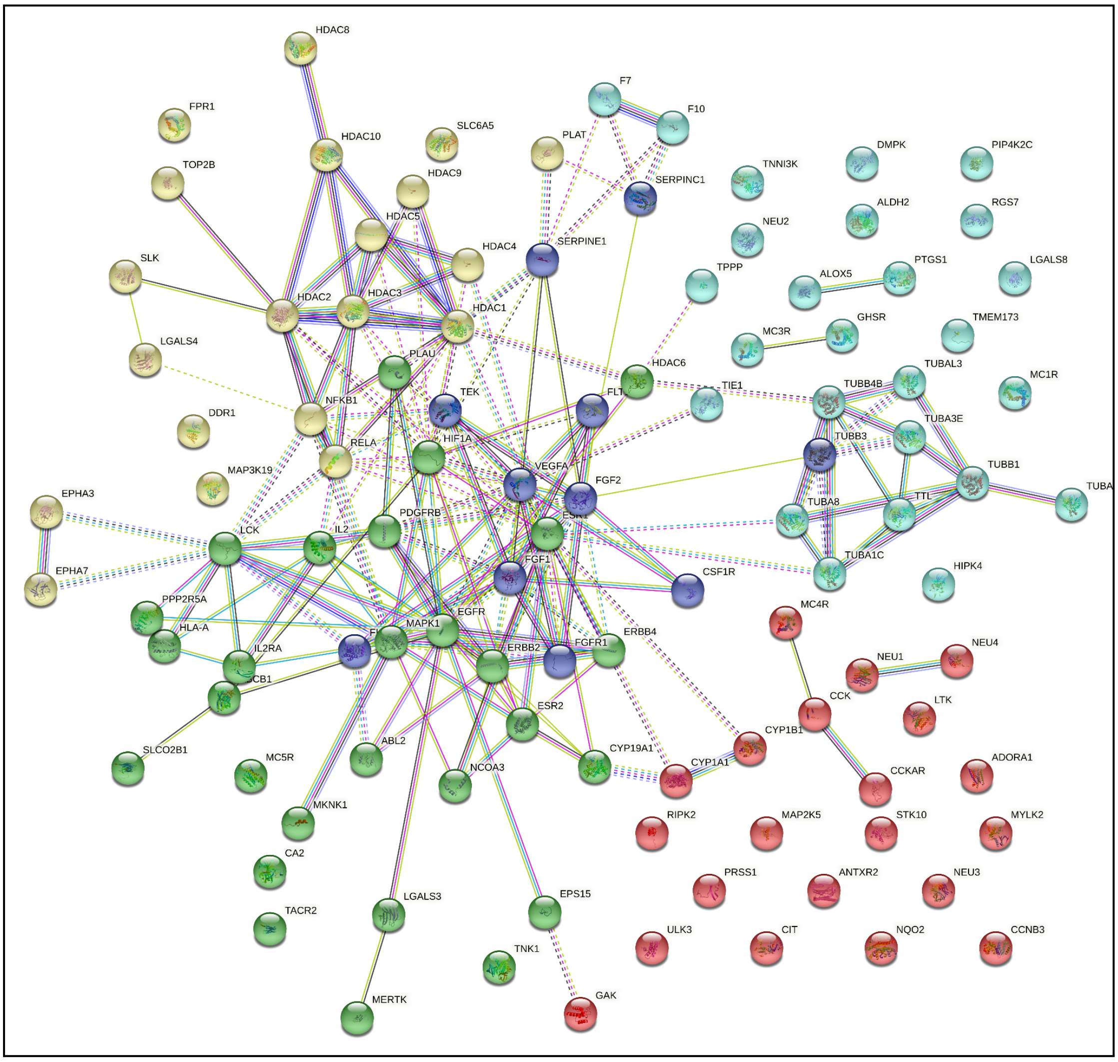


Figure 5.10: Metabolite target network in *P. integerrima*. BindingDB was used to predict the protein targets of *P. integerrima* metabolites, and STRING was used to do network analysis for proteins of human origin. STRING clusters proteins in different ways, as indicated by the different colors.

Prenol lipids are a group of lipids that are synthesized from isoprenoid precursors. These lipids have been implicated in the formation of plant galls. For example, a study by Giron *et al*, (2013) found that prenol lipids play a role in the induction of gall formation in response to herbivore feeding in tomato plants. The authors found that prenol lipids accumulate in the gall tissues and may act as signaling molecules to trigger the formation of the gall. Prenol lipids are a group of metabolites that include compounds such as plastoquinones and tocopherols. These lipids are thought to play a role in plant defense, as they have been shown to possess anti-oxidant and anti-inflammatory properties (Shorthouse *et al*, 1992; Giron *et al*, 2013).

ID	Source	Term ID	Term Name	p_{adj} (query_1)
1	KEGG	KEGG:04540	Gap junction	2.396×10^{-8}
2	KEGG	KEGG:04151	PI3K-Akt signaling pathway	1.323×10^{-6}
3	KEGG	KEGG:04066	HIF-1 signaling pathway	1.311×10^{-6}
4	KEGG	KEGG:04010	MAPK signaling pathway	1.070×10^{-6}
5	KEGG	KEGG:04014	Ras signaling pathway	4.706×10^{-6}
6	KEGG	KEGG:04613	Neutrophil extracellular trap formation	3.799×10^{-6}
7	KEGG	KEGG:05203	Viral carcinogenesis	5.923×10^{-6}
8	KEGG	KEGG:05200	Pathways in cancer	1.386×10^{-5}
9	KEGG	KEGG:05215	Prostate cancer	1.933×10^{-5}
10	KEGG	KEGG:05130	Pathogenic Escherichia coli infection	1.486×10^{-4}

Figure 5.11: Top 10 enriched pathways of targets of *P. integerrima* metabolites identified using gProfiler

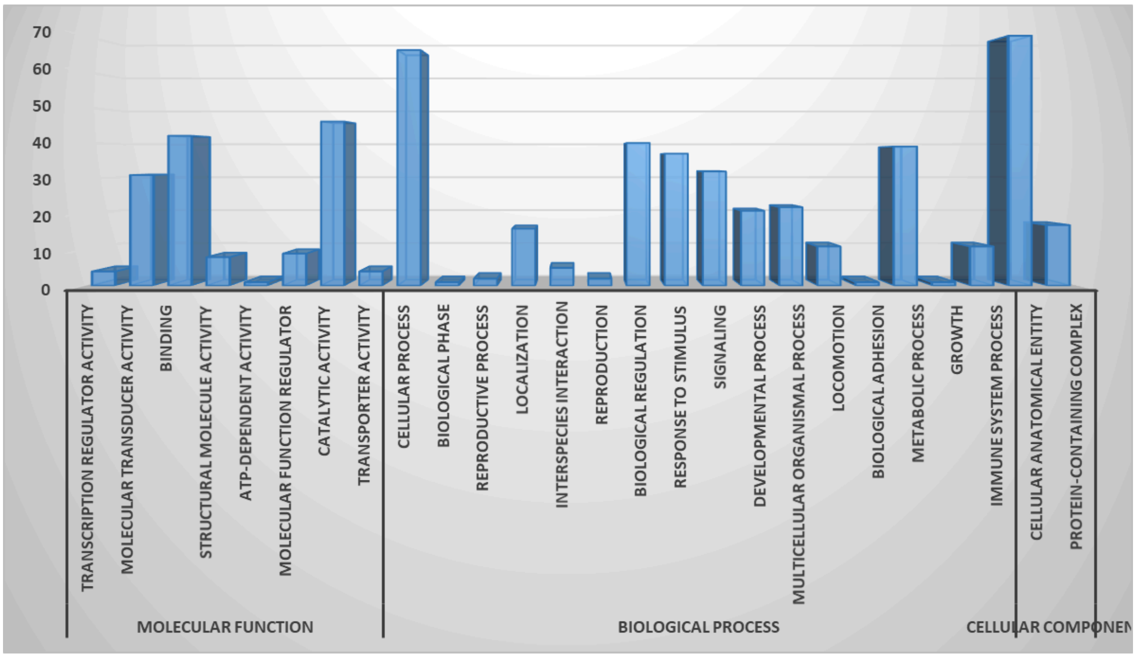


Figure 5.12: Gene ontology categories human protein targets identified from *P. integerrima* metabolite

Fatty acyls are a class of lipids that have been implicated in gall formation. A study by Wu *et al*, (2018) found that fatty acyls accumulate in the gall tissues of poplar trees in response to infestation by aphids. The authors suggest that these lipids may play a role in the defense response of the plant to the aphid infestation (Wu *et al*, 2018). Fatty acyls, such as jasmonic acid, are important signaling molecules in plant defense response against herbivores and pathogens. Studies have shown that jasmonic acid and its derivatives are synthesized in galls induced by aphids and mites, and they play a crucial role in regulating the plant's response to the gall-inducing stimulus (Howe and Jander, 2008).

Carboxylic acids and derivatives are a class of organic compounds that have been shown to play a role in gall formation. For example, a study by Yamamura et al. (2000) found that jasmonic acid, a carboxylic acid derivative, is involved in the formation of galls induced by the wasp *Diplolepis japonica* in wild rose plants. The authors suggest that jasmonic acid may act as a signaling molecule to trigger the formation of the gall (Yamamura *et al*, 2000).

Benzene and substituted derivatives are a class of organic compounds that have been implicated in gall formation. A study by Kuroda *et al*, (2005) found that gall formation induced by the aphid *Schlechtendalia chinensis* in Japanese elm trees is associated with the accumulation of benzyl glucosinolate, a derivative of benzene. The authors suggest that this compound may play a role in the defense response of the plant to the aphid infestation (Kuroda *et al*, 2005).

Organonitrogen compounds are a class of organic compounds that contain nitrogen atoms. These compounds have been shown to play a role in gall formation. Yamamoto *et al*, (2016) found that gall formation induced by the aphid *Tetraneura nigriabdominalis* in Japanese beech trees is associated with the accumulation of phenylalanine, an amino acid that contains a nitrogen atom. The authors suggest that this compound may play a role in the defense response of the plant to the aphid infestation (Yamamoto *et al*, 2016).

Steroids and steroid derivatives are a class of organic compounds that have been implicated in gall formation. Tamogami *et al*, (2015) found that the formation of galls induced by the wasp *Trichilogaster signiventris* in oak trees is associated with the accumulation of brassinosteroids, a class of steroid hormones. which suggest that these compounds may act as signaling molecules to trigger the formation of the gall (Tamogami *et al*, 2015).

Brassinosteroid biosynthesis and flavonoid biosynthesis pathways are enriched in gall from both *P. integerrima* and *T. chebula.* brassinosteroids (BRs) and flavonoids, have been shown to play a significant role in plant gall formation. BRs are a class of steroid hormones that regulate various physiological processes in plants, including cell elongation, differentiation, and division. Studies have shown that BRs play a crucial role in the formation of galls in plants. Galls are abnormal growths on plant tissues induced by various biotic and abiotic factors. These factors induce the production of BRs, which in turn promote cell proliferation and differentiation, leading to gall formation. In a study by Zhang *et al*. (2015), it was demonstrated that the BR signaling pathway is involved in the formation of gall tissue in Chinese bayberries. The authors showed that treatment with exogenous BRs increased the expression of genes involved in cell division and differentiation, resulting in the formation of larger galls. Similarly, a study by Jin *et al*. (2020) showed that the application of BRs to tea plants induced the formation of galls, which were associated with increased expression of genes involved in cell proliferation and expansion (Zhang *et al*, 2015; Jin *et al*, 2020).

Flavonoids, on the other hand, are a class of secondary metabolites that are widely distributed in plants and are involved in a variety of physiological processes, including UV protection, pigmentation, and defense against pathogens and herbivores. Several studies have shown that flavonoids also play a significant role in the formation of plant galls.

In a study by Tohge *et al*, (2013), it was shown that the expression of genes involved in flavonoid biosynthesis was upregulated in galls induced by the parasitic plant *Cuscuta pentagona*. The authors proposed that the accumulation of flavonoids in the gall tissue may play a role in protecting the plant against oxidative stress caused by the parasitic infection. Similarly, a study by Raja *et al,* (2018) showed that the accumulation of flavonoids in the gall tissue of sandalwood induced by the mite *Aceria santalae* may play a role in defense against the mite Tohge *et al*, 2013; Raja *et al,* 2018).

The analysis revealed a considerable number of metabolites in both the gall and leaf of *P. integerrima*, with a greater number of metabolites identified in the leaf. Pathway analysis of the identified metabolites showed similarities between the pathways in both the gall and leaf. We compared the metabolites of *P. integerrima* with those previously reported and found that Rauf *et al* had reported the presence of β-sitosterol and β-stigmasterol in this plant (Rauf *et al*, 2015). β-sitosterol has been found to have multiple biological effects, including antioxidant, anticancer, anti-diabetic, antibacterial, and immunomodulatory properties (Babu and Jayaraman, 2020). Zahoor *et al* extracted quercetin from *P. integerrima* gall and showed its anticholinesterase activity (Zahoor *et al*, 2018). Quercetin has been explored extensively in recent years for its anti-cancer activities, including cell signalling, apoptosis promotion, anti-proliferation and -oxidation, and growth inhibition (Yang *et al*, 2021). Hashemzaei *et al* showed the significant anticancer potential of quercetin both in vitro and in vivo; whereas, in vitro by MMT assay in 9 tumor cell lines and in in vivo experiments, mice bearing MCF‑7 and CT‑26 tumors (Hashemzaei *et al*, 2017).

A wide range of metabolites of the different classes mentioned already reported (Ahmad *et al*, 2020). In vivo and insilco study of different metabolites from the class of flavonoids were carried out in order to assess the antidiarrheal effect of the *P. integerrima* extracts (Alhumaydhi *et al* 2021). Eshwarappa *et al* reported phenolic compounds, tannins, phytosterols, triterpenoids, saponins, flavonoids, and reducing sugars in both aaqueous and ethanolic leaf gall extracts and reported their antioxidant activity (Eshwarappa *et al*, 2015). Ullah *et al* reported the flavonoid constituents of *P. integerrima* and isolated flavonoid glycosides Pistacides A and B (Ullah *et al*, 2012).

Further *P. integerrima* metabolites were used to identify potential targets. Human protein kinases are a complex system with complex internal and external interactions, and it has been demonstrated that they play a role in a number of disease conditions (Shchemelinin *et al* 2006). We identified different human protein kinases as the potential targets for *P. integerrima* metabolites. *P. integerrima* has been already reported to cure asthma. We identified Interleukin 13 (IL-13) as a potential human target which is involved in airway disease and provides strong evidence for the usage of *P. integerrima* in the treatment of asthma. Also, different human targets were identified to be involved in anticancer and neuromodulatory pathways suggesting these are areas where *P. integerrima* may show its activity. A proof-of-concept study using the metabolite profile of *P. integerrima* can be used to compare, characterize, and standardize other conventional formulations.

CHAPTER VI

CORRELATION OF DATA FROM DIFFERENT OMICS APPROACH

This chapter includes analysis of different omics approaches and statistical analysis of the results obtained

6. Correlation of Data from Different Omics Approach

6.1 Introduction

To understand the molecular mechanisms underlying gall formation, researchers often use an integrated multi-omics approach, which combines multiple omics data sets, such as genomics, transcriptomics, proteomics, and metabolomics. Advanced omics methodologies in the next generation enable the exploration of non-model organisms due to their ability to rapidly generate extensive de novo systems biology data. This makes them appealing choices for investigating interactions in less understood contexts (Oates *et al*, 2016).

Genomics can be used to identify genetic variations between resistant and susceptible plants, as well as to identify genes and pathways involved in gall formation. Transcriptomics can be used to determine which genes are differentially expressed in response to gall-inducing organisms. Proteomics can be used to identify which proteins are involved in gall formation and how they interact with each other. Metabolomics can be used to identify the metabolites that are involved in the formation of galls and how they are regulated (Oates *et al*, 2016).

By combining all these data sets, researchers can gain a comprehensive understanding of the molecular mechanisms underlying gall formation. For example, a study using an integrated multi-omics approach to investigate gall formation in *Arabidopsis thaliana* found that gall formation is associated with changes in gene expression, protein abundance, and metabolite levels. The study also identified several genes and pathways involved in gall formation, including those involved in cell wall modification, hormone signaling, and defense responses (Akhtar, 2014).

This study integrated/correlated the data from different omics approaches in order to understand the gall formation comprehensive and structurally. The data from RNASeq, microbiome as well metabolomics were used to correlate each other.

6.2. Material and Methods

6.2.1. Analysis of Unmapped (data not mapped to *P. integerrima* genome) RNA-Seq Reads

RNASeq reads from both gall and leaf samples were mapped to the assembled *P. integerrima* genome. The unmapped RNA-Seq reads (not mapped to the *P. integerrima* genome) were further mapped to the RefSeq microbial database using Kraken (Wood and Salzberg, 2014). To identify potential microbial genes and pathways, the unmapped reads were further classified using HUMAnN (Beghini *et al*, 2021). Spearman correlation analysis was conducted to assess the relationship between RNASeq expression values of genes in Pistacia integerrima and the abundance of associated microbes. Specifically, for each gene-microbe pair, the Spearman correlation coefficient was calculated to determine the strength and direction of their association.

6.2.2. Correlation Between Microbes and Metabolites in Gall Tissue

The microbes identified from gall tissue were used to identify potential metabolites already identified and reported using OmicsNet 2.0 (Zhou *et al*, 2022). The correlated metabolites with gall microbes were searched in gall metabolite datasets. Further microbe associated metabolites were used for pathway enrichment analysis using the Metaboanalyst 5.0 (Pang *et al*, 2021) package.

6.3. Results

6.3.3. Analysis of Unmapped RNA-Seq Reads

The unmapped reads (reads which are not mapped to *P. integerrima*) were checked for microbial composition by aligning them to the RefSeq microbial database using Kraken (Table 6.1). To identify potential microbial genes and pathways, the unmapped reads were further classified using HUMAnN.

Table 6.1: Percentage of reads classified into microbes using Kraken

Attribute	Gall	Leaf
Percentage of reads classified into microbes	47.8 %	52.4 %
Percentage of reads un-classified into microbes	52.1 %	47.5 %

Pseudomonas, Paracoccus, Aureimonas, Moraxella and *Acinetobacter* are the most abundant bacterial genus in both gall and leaf of *P. integerrima* (Figure 6.1). *Paracoccus haeundaensis, Paracoccus yeei, Moraxella osloensis, Cutibacterium acnes* and *Pseudomonas stutzeri* are the most abundant bacterial species in both gall and leaf of *P. integerrima* (Figure 6.2). There is not much difference observed in both gall and leaf bacterial diversity and is correlating with the 16S rRNA data.

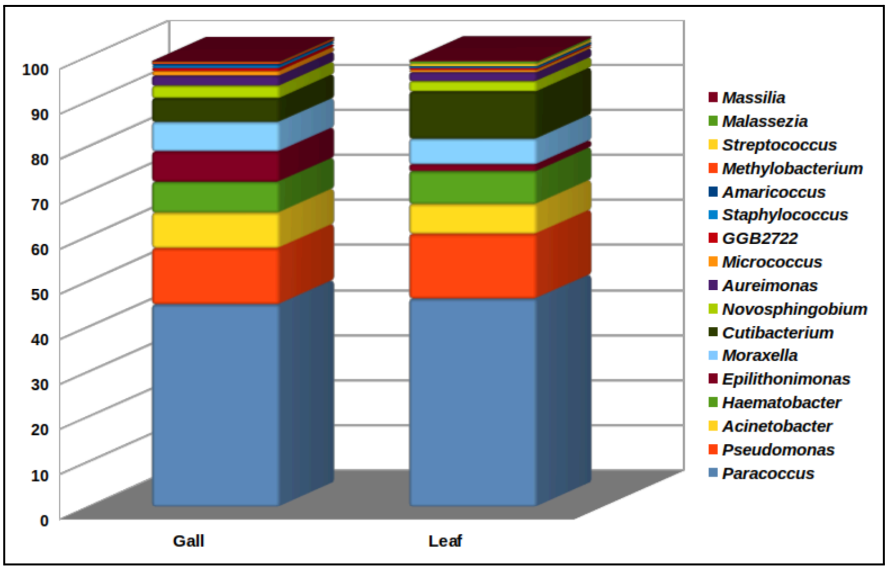


Figure 6.1: Genus level bacterial abundance of microbes from unmapped reads. Y-axis represents abundance (%)

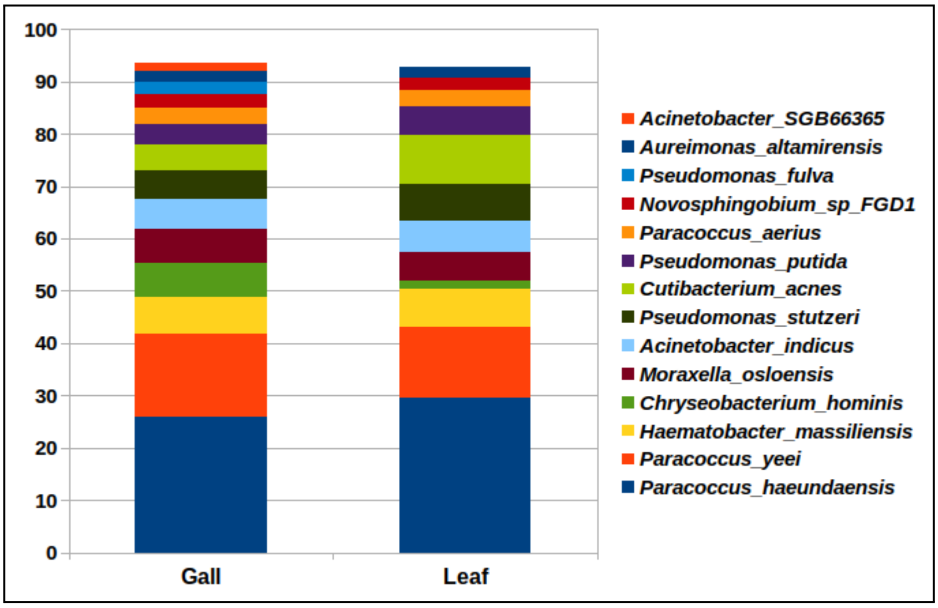


Figure 6.2: Species level bacterial abundance of microbes from unmapped reads. Y-axis represents abundance (%)

The abundance of different bacterial metabolic reactions involved in different biochemical pathways were identified from unmapped RNA Seq data.

The high abundant bacterial metabolic reactions in gall involves,

- Methane metabolism
- METHENYLTHFCYCLOHYDRO (MethenylTHF Cyclohydrolase)
- METHYLENETHFDEHYDROG-NADP (MethyleneTHF Dehydrogenase)
- Purine de novo biosynthesis pathway
- Biosynthesis of coenzyme A, an essential cofactor that is involved in many reactions (Figure 6.3).

Also microbial Pathways were identified from unmapped RNA-Seq reads. The two microbial metabolic pathways, Gondoate biosynthesis and Molybdopterin biosynthesis were suppressed in gall. Gondoate is monounsaturated fatty acid synthesized in anaerobic pathway. Aldehyde Oxidases is one of the MO (AO,EC1.2.3.1) catalyzing the final oxidation in the phytohormone biosynthesis of indole acetic acid and abscisic acid.

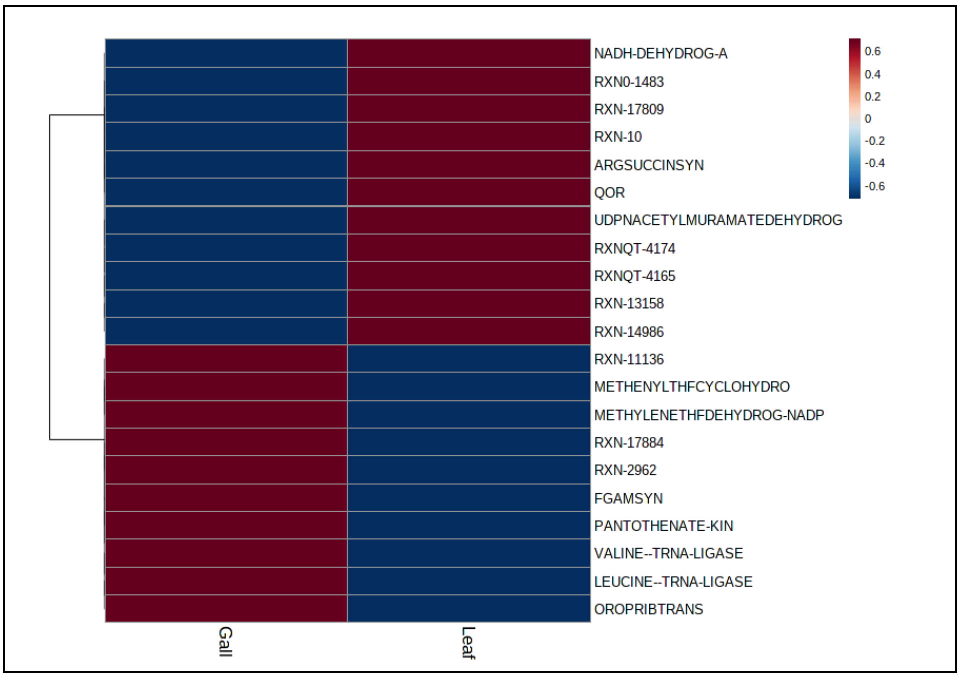


Figure 6.3: Abundance of different microbial metabolic reactions in gall and leaf of P. integerrima

To find the correlation between the gene expression and microbes, we have taken 2,785 differentially expressed genes between gall and leaf of *P. integerrima* as well as 40 bacterial genus (more abundant in gall tissue). Spearman correlation coefficient was calculated for each gene-microbe pair. *Pseudomonas* showed positive correlation between 1,525 gene expression and negative correlation with 1,060 gene expression. *Aureimonas* showed positive correlation with 980 gene expression whereas negative correlation with 1805 gene expression.

6.3.4. Correlation Between Microbes and Metabolites in Gall Tissue

The identification of association between the microbes and metabolites was carried out. Out of 40 bacterial genus (with high abundance in gall), 18 genus showed the association with 286 metabolites and 77 metabolic pathways (Figure 6.4). *Pseudomonas* is one of the known bacterial genus which synthesize IAA (Indole-3-acetic acid); an auxin. Bacterial production of IAA can interfere with plant development by disturbing the auxin balance in plants. Production of IAA by phytopathogenic bacteria via the indole-3-acetamide pathway was first characterized in the gall-forming pathogens *Agrobacterium tumefaciens* and *Pseudomonas savastanoi*. IAA is synthesized through the Tryptophan metabolism pathway and is identified as one of the associated pathways with microbes (Figure 6.5).

Out of 286 metabolites associated with microbes, 13 metabolites were identified in gall from LC_MS/MS metabolite profiling. The 13 metabolites are Glyoxylic acid, Oxidized glutathione, Agmatine, Glyceric acid, N-Carbamoylputrescine, Betaine aldehyde, Serotonin, 7,8-Dihydropteroic acid, Fucose 1-phosphate, Coproporphyrinogen III, Dihydroneopterin triphosphate, 3-Mercaptolactic acid, P1,P4-Bis(5'-uridyl) tetraphosphate. Tryptophan metabolism is the pathway where IAA is synthesized. Serotonin is the metabolite involved in the tryptophan metabolism pathway. This compound is synthesized from tryptophan and shares structural similarity with IAA acts as natural IAA inhibitor (Figure 6.6)

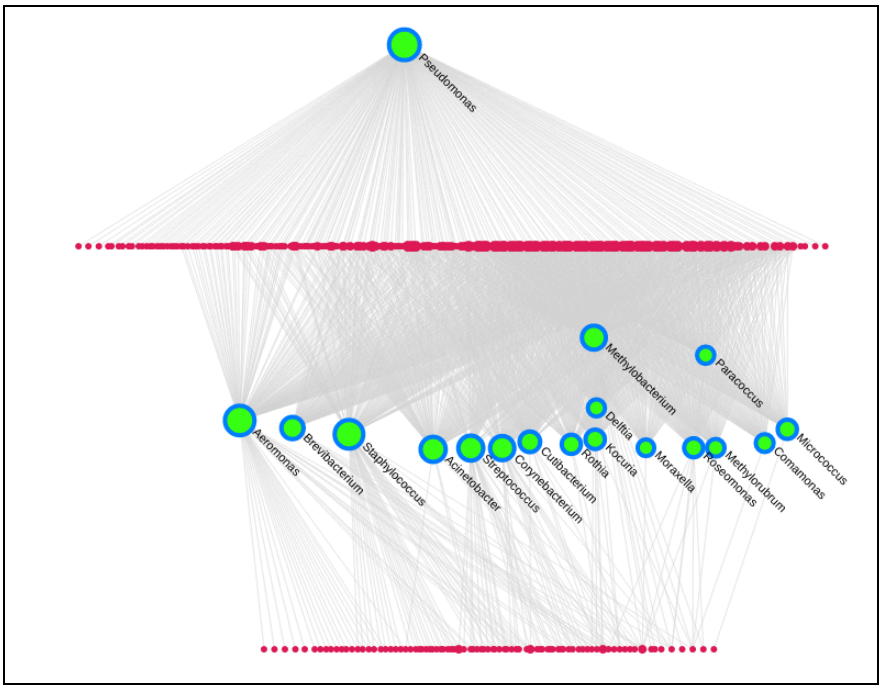


Figure 6.4: Correlation Between Microbes and Metabolites in Gall Tissue. Red color dots represents the different metabolic pathways

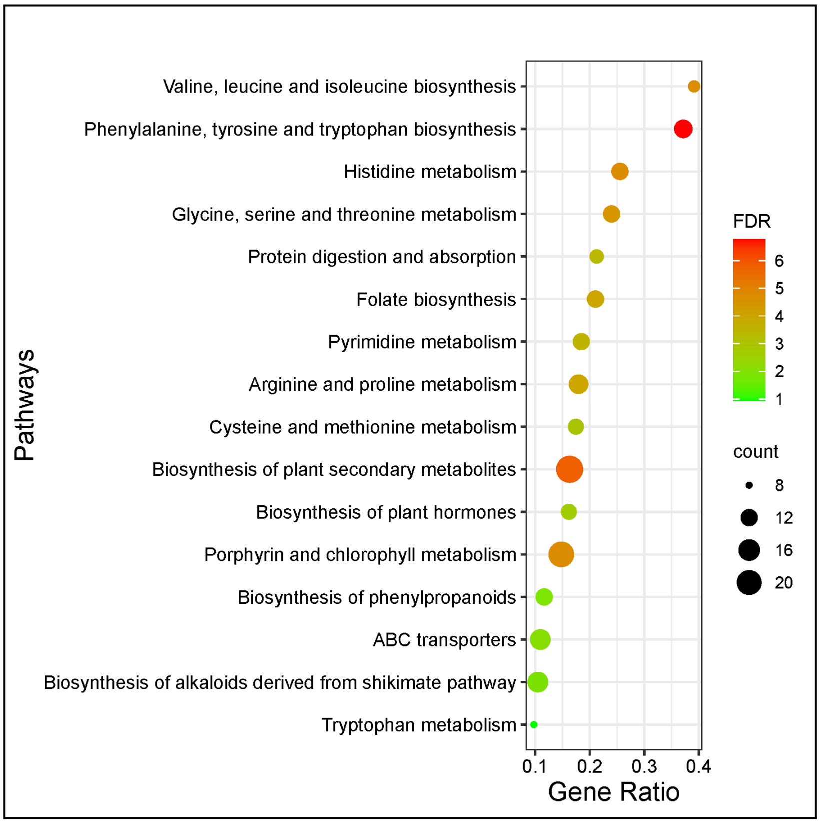


Figure 6.5: Top 15 metabolic pathways associated with microbes from *P. integerrima*

6.4. Discussion

Gall formation, a complex and intriguing phenomenon in plant biology, is influenced by a multitude of factors, including microbial communities inhabiting the gall tissue. In this study, an extensive analysis was conducted to identify associations between microbes and metabolites in gall tissue, shedding light on the intricate web of interactions occurring within this specialized environment. The identification of associations between microbes and metabolites in gall is a significant development in our understanding of the role of microbes in gall formation and development. The findings of this study suggest that microbes can influence plant development and health through a variety of mechanisms, including the production of plant hormones and other metabolites.

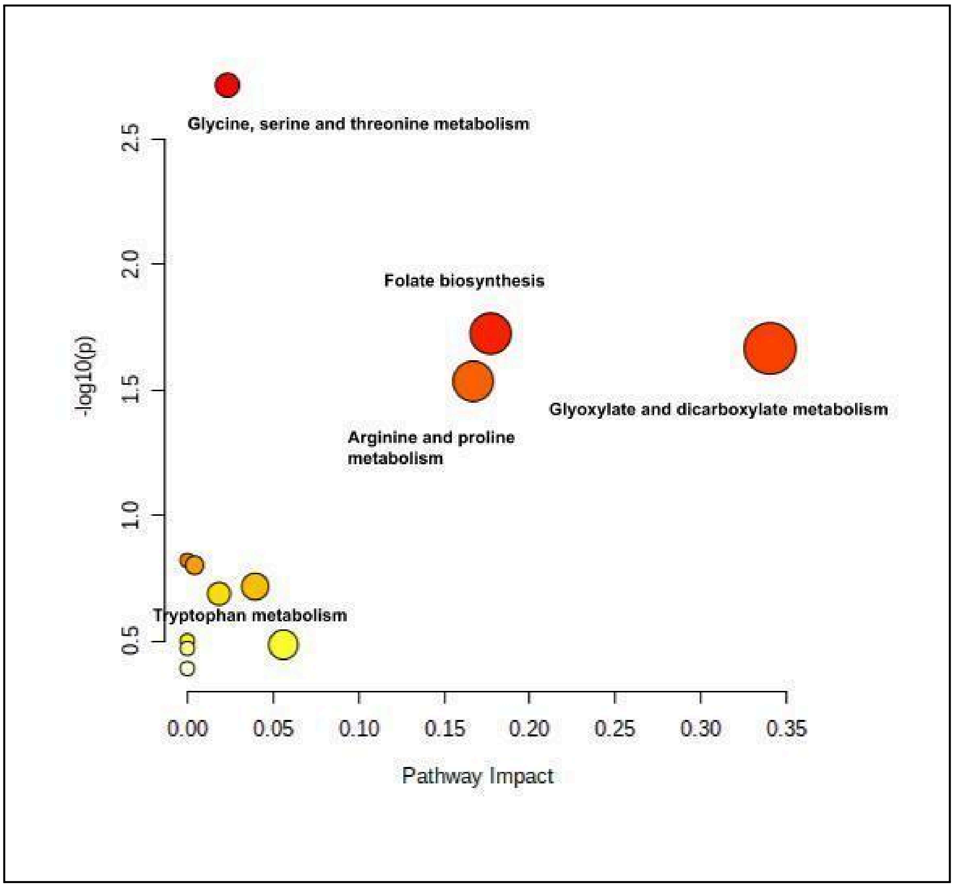


Figure 6.6: Pathway enrichment analysis of 13 metabolites associated microbes in the gall tissue

The discovery of a substantial presence of *Pseudomonas* and *Aureimonas* in the RNA sequencing data obtained from gall tissue aligns with the findings from the microbiome analysis. This congruence between the two datasets strengthens the evidence of these bacterial genera's prevalence within the gall environment, highlighting their potential significance in gall formation and development.

This analysis has unearthed a subset of bacterial metabolic reactions within gall tissue that hold significant implications for understanding gall formation. The presence of methane metabolism in gall tissue indicates the involvement of specific bacterial species capable of methane production or utilization, hinting at potential anaerobic conditions and localized oxygen depletion within the gall microenvironment. Enzymes such as MethenylTHF Cyclohydrolase and MethyleneTHF Dehydrogenase, integral to one-carbon metabolism, are highly abundant, emphasizing the importance of one-carbon units in gall tissue biochemistry and suggesting roles in nucleotide biosynthesis or other vital pathways.

In the study by Schönrogge *et al*, (2000), the detection of a biotin carboxyl carrier protein (BCCP), a component of acetyl CoA-carboxylase (ACCase), in galls is associated with the synthesis of triacylglycerol lipids, which act as an energy source. The upregulation of BCCP in response to gall formation indicates an increased lipid synthesis to meet the nutritional demands of growing larvae. The authors suggest that the identified proteins, including those related to respiratory stress and seed-specific proteins, play roles in the intricate processes associated with gall development (Schönrogge *et al*, 2000).

From this study, the high abundance of Coenzyme A within gall tissue was observed. Coenzyme A is known to play a crucial role in facilitating a diverse range of metabolic processes essential for gall formation. This includes its involvement in various biochemical reactions, such as the metabolism of lipids, carbohydrates, and amino acids. The enrichment of Coenzyme A in gall tissue underscores its importance in supporting the metabolic demands associated with the complex process of gall development.

The analysis also reveals intriguing instances of pathway suppression within gall tissue. The suppression of Gondoate Biosynthesis suggests an unfavorable anaerobic microenvironment, potentially linked to oxygen levels inhibiting anaerobic metabolic pathways. Furthermore, the suppressed Molybdopterin Biosynthesis pathway raises questions about reduced activity in molybdoenzymes within the gall, particularly Aldehyde Oxidase, with implications for the final oxidation step of phytohormone biosynthesis. These findings provide valuable insights into the complex biochemical processes and microbial interactions shaping gall development and highlight potential connections between metabolic pathways and the gall microenvironment.

One of the most notable findings of this study is the association between the bacterium *Pseudomonas* and the production of the plant hormone IAA. IAA is a key regulator of plant growth and development, and its production by *Pseudomonas* can have a significant impact on gall development. For example, *Pseudomonas savastanoi*, a gall-forming pathogen, produces IAA to induce gall formation in plants

(Glass *et al*, 1988; Spaepen and Vanderleyden, 2011).

The production of IAA by *Pseudomonas* can interfere with plant development by disturbing the auxin balance in plants. Auxins are plant hormones that play a role in a variety of processes, including cell division, cell elongation, and root growth (Tampakaki, 2009). When the auxin balance is disrupted, it can contribute to various anomalies in plants, potentially influencing phenomena such as gall formation, stunted growth, root distortion, and leaf deformation (Kunkel and Harper, 2018). In addition to IAA, the study also identified a number of other metabolites that are associated with microbes in gall. These metabolites include glyoxylic acid, oxidized glutathione, agmatine, glyceric acid, N-carbamoylputrescine, betaine aldehyde, serotonin, 7,8-dihydropteroic acid, fucose 1-phosphate, coproporphyrinogen III, dihydroneopterin triphosphate, 3-mercaptolactic acid, and P1,P4-bis(5'-uridyl) tetraphosphate. The precise role of these metabolites in gall formation and plant development is not fully understood, but they may play a role in a variety of processes, including signaling, transport, and metabolism.

One particularly interesting metabolite that was identified in this study is serotonin. In plants, serotonin has been shown to act as a natural IAA inhibitor (Pelagio-Flores *et al*, 2011). This suggests that the production of serotonin by microbes in gall may help to offset the effects of IAA produced by *Pseudomonas*. However, more research is needed to understand the precise role of serotonin in gall formation and plant development. Serotonin shares structural similarity with IAA and has been identified as a natural inhibitor of IAA. This discovery suggests a sophisticated mechanism by which plants may regulate the levels of IAA in response to microbial presence. Serotonin's role as an IAA inhibitor highlights the dynamic nature of plant-microbe interactions and their ability to fine-tune hormonal balance

The presence of the Tryptophan metabolism pathway associated with microbes in gall tissue, suggesting a direct link between microbial metabolism and IAA production. This connection underscores the significance of IAA in gall formation and prompts further investigation into the mechanistic details of this process

CHAPTER VII

CONCLUSION

This chapter concludes the thesis with the scope for further research on the development of *P. integerrima* galls

7. Conclusion

In conclusion, the comprehensive analysis of *P. integerrima* in this study has provided valuable insights into its genome, transcriptome, microbial communities, and metabolite profiles associated with gall formation.

The complete genetic code of *P. integerrima*, along with its leaf transcriptome, was analyzed through Illumina sequencing technology. The genome analysis resulted in a draft genome assembly spanning 462 Mb, featuring an N50 of 16,145 bp. Within this assembly, an impressive total of 39,452 genes were identified, with 18,492 of them having RNA or protein evidence. Notably, genes responsible for the synthesis of various plant secondary metabolites, including flavonoids and terpenoids, were meticulously characterized. Additionally, noncoding RNA miR397 and miR828 families were identified, primarily targeting the laccase (LCA) and MYB protein, respectively. Phylogenetic analysis unveiled a closer genetic relationship between *P. integerrima* and *Pistacia vera*. In the transcriptome analysis, 73,572 transcripts were identified for gall tissue, while 46,571 transcripts were found for leaf tissue. Notably, 17,257 genes displayed differential expression in both gall and leaf tissues, with genes related to phytohormone signaling pathways showing higher expression levels in gall tissue. Moreover, we uncovered 62 transcripts encoding serine/threonine protein kinases associated with plant-aphid interactions.

The draft genome assembly and gene discovery shed light on the genetic makeup of this plant species, revealing the presence of key genes involved in the biosynthesis of secondary metabolites, including flavonoids and terpenoids. The transcriptome analysis identified differentially expressed genes between gall and leaf tissues, with a particular emphasis on phytohormone signaling pathways. These findings suggest that phytohormones play a crucial role in gall formation and provide avenues for further investigation into the regulatory mechanisms involved in this process.

Although the specific microbes responsible for gall formation could not be identified in this study, the presence of certain bacterial taxa, such as *Aeromonas, Pseudomonas*, and *Bacillus*, implies their potential involvement in promoting plant growth, nutrient acquisition, and suppressing pathogenic microbes. These beneficial bacteria, known as plant growth-promoting bacteria (PGPB), represent promising candidates for future research exploring their direct role in gall formation. Moreover, the dominance of the *Glomerellaceae* fungal family in gall highlights the significance of fungal communities in the gall microenvironment.

A comprehensive metabolite analysis was carried out on both *P. integerrima* and *T. chebula* to explore their shared metabolite composition, considering that *T. chebula* galls are sometimes used as a substitute for *P. integerrima* in South Indian markets. A total of 4,414 non-redundant metabolites were identified from both gall and leaf samples of *P. integerrima*. Among these, 981 metabolites were exclusively found in gall tissue, while 1,591 metabolites were unique to leaf tissue, and 1,898 metabolites were common to both. In the case of *T. chebula*, 1,252 metabolites were specific to gall tissue, 1,040 were unique to leaf tissue, and 1,498 were shared between both. Interestingly, 1,593 metabolites were found to be shared between *P. integerrima* and *T. chebula* galls, with 1,157 metabolites unique to *T. chebula* gall and 1,286 exclusive to *P. integerrima* gall.

Similarly, in the leaves of these plants, 1,577 metabolites were common to both *P. integerrima* and *T. chebula*, while 961 metabolites were specific to *T. chebula* leaf and 1,856 were unique to *P. integerrima* leaf tissue. In terms of metabolic pathways, *P. integerrima* gall tissue exhibited enrichment in pathways like linoleic acid metabolism, taurine and hypotaurine metabolism, sesquiterpenoid and triterpenoid biosynthesis, and sphingolipid metabolism, among others. Conversely, *P. integerrima* leaf tissue displayed enrichment in pathways such as taurine and hypotaurine metabolism, linoleic acid metabolism, indole alkaloid biosynthesis, and sulfur metabolism. Both gall and leaf tissues yielded a diverse array of metabolites from various categories, including prenol lipids, organooxygen compounds, fatty acyls, carboxylic acids and derivatives, steroids and steroid derivatives, flavonoids, phenol ethers, tannins, and more.

Furthermore, a correlation analysis was conducted using data obtained from various omics studies. Notably, among the 40 bacterial genera that exhibited high abundance in gall tissue, 18 of them were found to be associated with 286 metabolites and 77 metabolic pathways. Within this group of 286 metabolites linked to microbes, 13 metabolites were specifically identified in gall tissue through LC-MS/MS metabolite profiling. These 13 metabolites included Glyoxylic acid, Oxidized glutathione, Agmatine, Glyceric acid, N-Carbamoylputrescine, Betaine aldehyde, Serotonin, 7,8-Dihydropteroic acid, Fucose 1-phosphate, Coproporphyrinogen III, Dihydroneopterin triphosphate, 3-Mercaptolactic acid, and P1,P4-Bis(5'-uridyl) tetraphosphate.

Understanding the intricate dynamics between host plants, invading organisms, and associated metabolites could unveil novel insights into the mechanisms underlying gall formation and plant defense responses. For instance, elucidating the specific roles of differentially expressed genes and enriched metabolic pathways within the galls could shed light on the regulatory networks governing gall development and the synthesis of bioactive compounds. Moreover, exploring the functional implications of the identified microbial communities within the galls, including their potential roles in gall formation and host defense mechanisms, may provide insights into strategies for modulating microbial populations to bolster plant resilience and maximize the production of bioactive compounds.

While this study sheds light on the molecular mechanisms underlying gall formation in Pistacia integerrima, several limitations warrant consideration and suggest avenues for future investigation. Functional validation of differentially expressed genes and enriched metabolic pathways is crucial to confirm their roles in gall development and plant defense responses. Additionally, identifying the specific microbial species responsible for gall formation and elucidating their functional roles using metagenomic approaches would enhance our understanding of host-microbe interactions. Further research focusing on the biological activities of specific metabolites, the impact of environmental factors on gall formation, and the regulatory networks governing host-microbe interactions will contribute to a more comprehensive understanding of gall formation

Overall, this study serves as a foundation for future investigations aiming to unravel the complex interactions between microbes, metabolites, and host plants during gall formation. The findings contribute to our understanding of the potential roles of beneficial bacteria, microbial communities, and metabolic pathways in the development and functioning of galls in *Pistacia integerrima*. Further research in this field holds promising prospects for applications in agriculture, plant health management, and the exploration of novel bioactive compounds derived from gall-associated microbial and plant interactions.

REFERENCES

References

Abarenkov, K., Zirk, A., Piirmann, T., Pöhönen, R., Ivanov, F., Nilsson, R. H., & Kõljalg, U. (2020). UNITE QIIME release for fungi [Data set]. UNITE Community.

Ahmad, N. S., Farman, M., Najmi, M. H., Mian, K. B., & Hasan, A. (2008). Pharmacological basis for use of Pistacia integerrima leaves in hyperuricemia and gout. Journal of Ethnopharmacology, 117(3), 478–482. https://doi.org/10.1016/j.jep.2008.02.031

Ahmad, N. S., Waheed, A., Farman, M., & Qayyum, A. (2010). Analgesic and anti-inflammatory effects of Pistacia integerrima extracts in mice. Journal of Ethnopharmacology, 129(2), 250–253. https://doi.org/10.1016/j.jep.2010.03.017

Ahmad, R., Almubayedh, H., Ahmad, N., Naqvi, A. A., & Riaz, M. (2020). Ethnobotany, ethnopharmacology, phytochemistry, biological activities and toxicity of Pistacia chinensis subsp. integerrima: A comprehensive review. Phytotherapy Research: PTR, 34(11), 2793–2819. https://doi.org/10.1002/ptr.6720

Ahmad, S., Ali, M., Ansari, S. H., & Ahmed, F. (2010). Phytoconstituents from the galls of Pistacia integerrima Stewart. Journal of Saudi Chemical Society, 14(4), 409–412. https://doi.org/10.1016/j.jscs.2010.05.003

Akhtar, S. A. (2014). The Molecular Mechanisms Underlying Gall Formation in Plasmodiophora brassicae Infected Arabidopsis thaliana [The University of Sheffield]. https://etheses.whiterose.ac.uk/6785/

AL-Saghir, M. G., & Porter, D. M. (2012). Taxonomic revision of the genus Pistacia L. (Anacardiaceae). American Journal of Plant Sciences, 03(01), 12–32. https://doi.org/10.4236/ajps.2012.31002

Alhat, M. S., & Kulkarni, S. (2023). Karkatshringi (Pistacia integerrima): A review on functional and pharmacological properties. International Journal of Research Publication and Reviews, 4(6), 1545–1551

Alhumaydhi, F. A., Rauf, A., Rashid, U., Bawazeer, S., Khan, K., Mubarak, M. S., Aljohani, A. S. M., Khan, H., El-Saber Batiha, G., El-Esawi, M. A., & Mishra, A. P. (2021). In vivo and in silico studies of flavonoids isolated from Pistacia integerrima as potential antidiarrheal agents. ACS Omega, 6(24), 15617–15624. https://doi.org/10.1021/acsomega.1c00298

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Alves, R. R. N., & Rosa, I. M. L. (2007). Biodiversity, traditional medicine and public health: where do they meet? Journal of Ethnobiology and Ethnomedicine, 3(1), 14. https://doi.org/10.1186/1746-4269-3-14

Ananthakrishnan, T. N. (2007). Distribution, adaptive dynamics, and phylogeny of gall-inducing thrips (Insecta: Thysanoptera) of the Indian subcontinent. Oriental Insects, 41(1), 27–33. https://doi.org/10.1080/00305316.2007.10417497

Andreas, P., Kisiala, A., Emery, R. J. N., De Clerck-Floate, R., Tooker, J. F., Price, P. W., Miller, D. G., Iii, Chen, M.-S., & Connor, E. F. (2020). Cytokinins are abundant and widespread among insect species. Plants, 9(2), 208. https://doi.org/10.3390/plants9020208

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ankita, S., Preeti, D., & Verma, M. B. (2021). Pistachiaintegerrima: An important source of Nutraceutical and Phytomedicine. Phytomedicine. World Journal of Pharmacy and Pharmaceutical Sciences, 10(4), 1542–1563

Ansari, S. H., Ali, M., & Qadry, J. S. (1993). Essential oils of Pistacia integerrima galls and their effect on the central nervous system. International Journal of Pharmacognosy, 31(2), 89–95. https://doi.org/10.3109/13880209309082924

Anuradha, S. (2010). Pharmacognostic and phytochemical evaluation of leaf galls of Kakadshringi used in the Indian system of medicine. Ournal of Scientific & Industrial Research, 69, 700–704

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, 408(6814), 796–815. https://doi.org/10.1038/35048692

Aragón, I. M., Pérez-Martínez, I., Moreno-Pérez, A., Cerezo, M., & Ramos, C. (2014). New insights into the role of indole-3-acetic acid in the virulence of Pseudomonas savastanoi pv. savastanoi. FEMS Microbiology Letters, 356(2), 184–192. https://doi.org/10.1111/1574-6968.12413

Babu, S., & Jayaraman, S. (2020). An update on β-sitosterol: A potential herbal nutraceutical for diabetic management. Biomedicine & Pharmacotherapy, 131. https://doi.org/10.1016/j.biopha.2020.110

Bag, A., Bhattacharyya, S. K., Pal, N. K., & Chattopadhyay, R. R. (2011). Synergistic effect of Terminalia chebula against multidrug-resistant uropathogenic Escherichia coli. Med Aromatic Plant Sci Biotech, 5(1), 70–73

Bag, Anwesa, Bhattacharyya, S. K., & Chattopadhyay, R. R. (2013). The development of Terminalia chebula Retz. (Combretaceae) in clinical research. Asian Pacific Journal of Tropical Biomedicine, 3(3), 244–252. https://doi.org/10.1016/s2221-1691(13)60059-3

Bailey, S., Percy, D. M., Hefer, C. A., & Cronk, Q. C. B. (2015). The transcriptional landscape of insect galls: psyllid (Hemiptera) gall formation in Hawaiian Metrosideros polymorpha (Myrtaceae). BMC Genomics, 16(1), 943. https://doi.org/10.1186/s12864-015-2109-9

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology: A Journal of Computational Molecular Cell Biology, 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA, 6(1). https://doi.org/10.1186/s13100-015-0041-9

Barah, P., & Bones, A. M. (2015). Multidimensional approaches for studying plant defence against insects: from ecology to omics and synthetic biology. Journal of Experimental Botany, 66(2), 479–493. https://doi.org/10.1093/jxb/eru489

Barinder, K., & Saurabh, S. (2015). A review on gall karkar shringi. J. Med. Plants Res, 9, 636–640

Bashir, Y., Singh, S., & Kumar Konwar, B. (2014). Metagenomics: An Application Based Perspective. Chinese Journal of Biology, 1–7. https://doi.org/10.1155/2014/14603

Basr Ila, H., Kafkas, S., & Topaktas, M. (2003). Chromosome numbers of four Pistacia(Anacardiaceae)species. The Journal of Horticultural Science & Biotechnology, 78(1), 35–38. https://doi.org/10.1080/14620316.2003.11511583

Basri, D., Fredalina, Tan, L., Si, & Zaleha Zin, N. (2012). In Vitro Antibacterial Activity of Galls of Quercus infectoria Olivier against Oral Pathogens. Evidence-Based Complementary and Alternative Medicine, 1–6. https://doi.org/10.1155/2012/632796

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife, 10. https://doi.org/10.7554/elife.65088

Berendsen, R. L., Pieterse, C. M. J., & Bakker, P. A. H. M. (2012). The rhizosphere microbiome and plant health. Trends in Plant Science, 17(8), 478–486. https://doi.org/10.1016/j.tplants.2012.04.001

Bi, Q., Zhao, Y., Du, W., Lu, Y., Gui, L., Zheng, Z., Yu, H., Cui, Y., Liu, Z., Cui, T., Cui, D., Liu, X., Li, Y., Fan, S., Hu, X., Fu, G., Ding, J., Ruan, C., & Wang, L. (2019). Pseudomolecule-level assembly of the Chinese oil tree yellowhorn (Xanthoceras sorbifolium) genome. GigaScience, 8(6). https://doi.org/10.1093/gigascience/giz070

Bibi, Y., Nisa, S., Zia, M., Waheed, A., Ahmed, S., & Chaudhary, M. F. (2012). The Study of Anticancer and Antifungal Activities of Pistacia integerrima Extract in vitro. Indian Journal of Pharmaceutical Sciences, 74(4), 375–379. https://doi.org/10.4103/0250-474X.107085

Bibi, Y., Zia, M., & Qayyum, A. (2015). An overview of Pistacia integerrima, a medicinal plant species: Ethnobotany, biological activities and phytochemistry, Pak. Pak. J. Pharm. Sci, 28, 1009–1013

Blanco, E., Parra, G., & Guigó, R. (2007). Using geneid to identify genes. Et al [Current Protocols in Bioinformatics], Chapter 4(1), Unit 4.3. https://doi.org/10.1002/0471250953.bi0403s18

Boeckmann, B. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Research, 31(1), 365–370. https://doi.org/10.1093/nar/gkg095

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics, 27(4), 578–579. https://doi.org/10.1093/bioinformatics/btq683

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, *6*(1). https://doi.org/10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Ver Loren Van Themaat, E., Schulze-Lefert, P., & Panstruga, R. (2015). Structure and function of the bacterial root microbiome in wild and domesticated barley. Nature, 517(7535), 609–613

Buyel, J. F. (2018). Plants as sources of natural and recombinant anti-cancer agents. Biotechnology Advances, 36(2), 506–520. https://doi.org/10.1016/j.biotechadv.2018.02.002

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal, 11(12), 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. BMC Bioinformatics, 10(1), 421. https://doi.org/10.1186/1471-2105-10-421

Chan, P. P., & Lowe, T. M. (2019). TRNAscan-SE: Searching for tRNA genes in genomic sequences. In Methods in Molecular Biology (pp. 1–14). Springer New York

Chang, C. L., & Lin, C. S. (2010). Development of antioxidant activity and pattern recognition of Terminalia chebula Retzius extracts and its fermented products. HungKuang J, 61, 115–129

Chen, H., Liu, J., Cui, K., Lu, Q., Wang, C., Wu, H., Yang, Z., Ding, W., Shao, S., Wang, H., Ling, X., King-Jones, K., & Chen, X. (2018). Molecular mechanisms of tannin accumulation in Rhus galls and genes involved in plant-insect interactions. Scientific Reports, 8(1). https://doi.org/10.1038/s41598-018-28153-y

Chen, X., Yang, Z., Chen, H., Qi, Q., Liu, J., Wang, C., Shao, S., Lu, Q., Li, Y., Wu, H., King-Jones, K., & Chen, M.-S. (2020). A complex nutrient exchange between a gall-forming aphid and its plant host. Frontiers in Plant Science, 11, 811. https://doi.org/10.3389/fpls.2020.00811

Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. Bioinformatics, 30(1), 31–37. https://doi.org/10.1093/bioinformatics/btt310

Chopra, R. N., Nayar, S. L., & Chopra, I. C. (1986). Glossary of Indian Medicinal Plants (Including the Supplement). Council of Scientific and Indust Res, 79–81

Choudhari, D., Kulkarni, Y., & Visave, P. (2021). Abhava pratinidhi dravya: A Comparative Phytochemical Study of Bharangi and Kantakari. International Journal of Ayurvedic Medicine, 12(3), 588–592. https://doi.org/10.47552/ijam.v12i3.2070

Cissé, O. H., Almeida, J. M. G. C. F., Fonseca, A., Kumar, A. A., Salojärvi, J., Overmyer, K., Hauser, P. M., & Pagni, M. (2013). Genome sequencing of the plant pathogen Taphrina deformans, the causal agent of peach leaf curl. MBio, 4(3), e00055-13. https://doi.org/10.1128/mBio.00055-13

D K Ved (Late), Suma Tagadur Sureshchandra, Vijay Barve, Vijay Srinivas, Sathya Sangeetha, K. Ravikumar, Kartikeyan R., Vaibhav Kulkarni, Ajith S. Kumar, S.N. Venugopal, B. S. Somashekhar, M.V. Sumanth, Noorunissa Begum, Sugandhi Rani, Surekha K.V., and Nikhil Desale. (2016). Encyclopedia on Indian Medicinal Plants. ENVIS Center on Medicinal Plants. envis.frlht.org

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology, 72(7), 5069–5072. https://doi.org/10.1128/AEM.03006-05

Díaz-González, S., Marín, P., Sánchez, R., Arribas, C., Kruse, J., González-Melendi, P., Brunner, F., & Sacristán, S. (2020). Mutualistic fungal endophyte Colletotrichum tofieldiae Ct0861 colonizes and increases growth and yield of maize and tomato plants. Agronomy (Basel, Switzerland), 10(10), 1493. https://doi.org/10.3390/agronomy10101493

Djakpo, O., & Yao, W. (2010). Rhus chinensisandGalla Chinensis- folklore to modern evidence: review: <I>RHUS CHINENSIS</i>ANDGALLA CHINENSIS. Phytotherapy Research: PTR, 24(12), 1739–1747. https://doi.org/10.1002/ptr.3215

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., & Wishart, D. S. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. Journal of Cheminformatics, 8(1), 61. https://doi.org/10.1186/s13321-016-0174-y

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6

Egamberdieva, D., Wirth, S. J., Alqarawi, A. A., Abd Allah, E. F., & Hashem, A. (2017). Phytohormones and beneficial microbes: Essential components for plants to balance stress and fitness. Frontiers in Microbiology, 8, 2104. https://doi.org/10.3389/fmicb.2017.02104

Eshwarappa, R. B., Dhananjaya, B., Ramachandra, Y. L., Subaramaihha, S., Subbaiah, S. P., & Austin, R. (2016). In vivo toxicity studies on gall extracts of Terminalia chebula (Gaertn.) Retz. (combretaceae). Pharmacognosy Research, 8(3), 199. https://doi.org/10.4103/0974-8490.182914

Eshwarappa, R. S. B., Lakshmikantha, R. Y., Subaramaihha, S. R., Subbaiah, S. G. P., Surendranath, A. R., Dhananjaya, B. L. (2015). Antioxidant activity of insect gall extracts of Pistacia integerrima. Acta Scientiarum Polonorum. Technologia Alimentaria, 14(4), 367–374. https://doi.org/10.17306/j.afs.2015.4.36

Etminani, F., & Harighi, B. (2018). Isolation and identification of endophytic bacteria with plant growth promoting activity and biocontrol potential from wild pistachio trees. The Plant Pathology Journal, 34(3), 208–217. https://doi.org/10.5423/PPJ.OA.07.2017.0158

Faist, H., Keller, A., Hentschel, U., & Deeken, R. (2016). Grapevine (Vitis vinifera) Crown Galls Host Distinct Microbiota. Applied and Environmental Microbiology, 82(18), 5542–5552. https://doi.org/10.1128/aem.01131-16

Fokunang, C. N., Ndikum, V., Tabi, O. Y., Jiofack, R. B., Ngameni, B., Guedje, N. M., Tembe-Fokunang, E. A., Tomkins, P., Barkwan, S., Kechia, F., Asongalem, E., Ngoupayou, J., Torimiro, N. J., Gonsu, K. H., Sielinou, V., Ngadjui, B. T., Angwafor, F., 3rd, Nkongmeneck, A., Abena, O. M., … Kamsu-Kom. (2011). Traditional medicine: past, present and future research and development prospects and integration in the National Health System of Cameroon. African Journal of Traditional, Complementary, and Alternative Medicines, 8(3), 284–295. https://doi.org/10.4314/ajtcam.v8i3.65276

Gatjens-Boniche, O. (2019). The mechanism of plant gall induction by insects: revealing clues, facts, and consequences in a cross-kingdom complex interaction. Revista de Biologia Tropical, 67(6). https://doi.org/10.15517/rbt.v67i6.33984

Gebhardt, C., Schmidt, R., & Schneider, K. (2005). Plant genome analysis: the state of the art. International Review of Cytology, 247, 223–284. https://doi.org/10.1016/S0074-7696(05)47005-9

Ghaffari, S., & Shabazaz, M. (2005.). Chromosome variation in Pistacia genus

Giron, D., Frago, E., Glevarec, G., Pieterse, C. M. J., & Dicke, M. (2013). Cytokinins as key regulators in plant-microbe-insect interactions: connecting plant growth and defense. Frontiers in Plant Science, 4, 1–13

Giron, D., Huguet, E., Stone, G. N., & Body, M. (2016). Insect-induced effects on plants and possible effectors used by galling and leaf-mining insects to manipulate their host-plant. Journal of Insect Physiology, 84, 70–89. https://doi.org/10.1016/j.jinsphys.2015.12.009

Glass, N. L., & Kosuge, T. (1988). Role of indoleacetic acid-lysine synthetase in regulation of indoleacetic acid pool size and virulence of Pseudomonas syringae subsp. savastanoi. Journal of Bacteriology, 170(5), 2367–2373. https://doi.org/10.1128/jb.170.5.2367-2373.1988

Goggin, F. L., & Fischer, H. D. (2021). Reactive oxygen species in plant interactions with aphids. Frontiers in Plant Science, 12, 811105. https://doi.org/10.3389/fpls.2021.811105

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644–652. https://doi.org/10.1038/nbt.1883

Griffiths-Jones, S. (2003). Rfam: an RNA family database. Nucleic Acids Research, 31(1), 439–441. https://doi.org/10.1093/nar/gkg006

Grover, I. S., & Bala, S. (1992). Antimutagenic activity of Terminalia chebula (myroblan) in Salmonella typhimurium. Indian Journal of Experimental Biology, 30(4), 339–341

Grover, M. (2021). Pistacia integerrima (Shringi)- A Plant With Significant Pharmacological Activities. The Journal of Phytopharmacology, 10(5), 323–330. https://doi.org/10.31254/phyto.2021.10508

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Jr, Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research, 31(19), 5654–5666. https://doi.org/10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biology, 9(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Harris, M. O., & Pitzschke, A. (2020). Plants make galls to accommodate foreigners: some are friends, most are foes. The New Phytologist, 225(5), 1852–1872. https://doi.org/10.1111/nph.16340

Hashemzaei, M., Far, A. D., Yari, A., Heravi, R. E., Tabrizian, K., Taghdisi, S. M., Sadegh, S. E., Tsarouhas, K., Kouretas, D., Tzanakakis, G., Nikitovic, D., Anisimov, N. Y., Spandidos, D. A., Tsatsakis, A. M., & Rezaee, R. (2017). Anticancer and apoptosis-inducing effects of quercetin in vitro and in vivo. Oncology Reports, 38(2), 819–828. https://doi.org/10.3892/or.2017.5766

Hearn, J., Blaxter, M., Schönrogge, K., Nieves-Aldrey, J.-L., Pujade-Villar, J., Huguet, E., Drezen, J.-M., Shorthouse, J. D., & Stone, G. N. (2019). Genomic dissection of an extended phenotype: Oak galling by a cynipid gall wasp. PLoS Genetics, 15(11), e1008398. https://doi.org/10.1371/journal.pgen.1008398

Henry, R. J. (2022). Progress in plant genome sequencing. Applied Biosciences, 1(2), 113–128. https://doi.org/10.3390/applbiosci1020008

Eddy, S. R. (1996). Hidden Markov Models. Current Opinion in Structural Biology, 6:361-365

Hirano, T., Kimura, S., Sakamoto, T., Okamoto, A., Nakayama, T., Matsuura, T., Ikeda, Y., Takeda, S., Suzuki, Y., Ohshima, I., & Sato, M. H. (2020). Reprogramming of the Developmental Program of Rhus javanica During Initial Stage of Gall Induction by Schlechtendalia chinensis. Frontiers in Plant Science, 11, 471. https://doi.org/10.3389/fpls.2020.00471

Howe, G. A., & Jander, G. (2008). Plant immunity to insect herbivores. *Annual Review of Plant Biology*, *59*(1), 41–66. https://doi.org/10.1146/annurev.arplant.59.032607.092825

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, S., Zhou, J., Gao, L., & Tang, Y. (2021). Plant miR397 and its functions. Functional Plant Biology: FPB, 48(4), 361–370. https://doi.org/10.1071/FP20342

Hwang, H.-H., & Gelvin, S. B. (2004). Plant proteins that interact with VirB2, the Agrobacterium tumefaciens pilin protein, mediate plant transformation. The Plant Cell, 16(11), 3148–3167. https://doi.org/10.1105/tpc.104.026476

Inbar, M., Wink, M., & Wool, D. (2004). The evolution of host plant manipulation by insects: molecular and ecological evidence from gall-forming aphids on Pistacia. Molecular Phylogenetics and Evolution, 32(2), 504–511. https://doi.org/10.1016/j.ympev.2004.01.006

Irfan, A., Imran, M., Sumrra, S. H., Qaisar, M. N., Khalid, N., Basra, M. A. R., Shah, A. T., Hussien, M., Assiri, M. A., & Al-Sehemi, A. G. (2021). Exploration of carbonic anhydrase inhibition of bioactive metabolites from Pistacia integerrima by molecular docking and first-principles investigations. Journal of Saudi Chemical Society, 25(10), 101324. https://doi.org/10.1016/j.jscs.2021.101324

Ismail, M., Shafiq-Ur-Rahman, M. N., Mohani, N., Khan, M., Barkatullah, A., & Hussain, J. (2011). Pharmacognostic and phy-tochemical investigation of the stem bark of Pistacia integerrima Stew ex Brandis. Journal of Medicinal Plants Research, 5(16), 3891–3895

Jeet, K., & Baldi, A. (2019). A comprehensive review on Pistacia integerrima. Current Traditional Medicine, 5(2), 98–113. https://doi.org/10.2174/2215083805666181213154344

Jena, A. K., Karan, M., & Vasisht, K. (2017). Plant parts substitution based approach as a viable conservation strategy for medicinal plants: A case study of Premna latifolia Roxb. Journal of Ayurveda and Integrative Medicine, 8(2), 68–72. https://doi.org/10.1016/j.jaim.2016.11.003

Jeong, A., Kim, C. Y., Lee, J. S., Kim, T. G., Kim, S. H., & Lee, C. K. (2002). Inhibition of HIV-1 integrase by galloyl glucoses from Terminalia chebula and flovonol glycoside gallates from Euphorbia pekinensis. Plant Med, 68, 457–459

Jin, X., Yao, M., Ma, Y., Ma, J., Chen, L., & Ji, R. (2020). Brassinosteroids regulate gall formation induced by Meloidogyne incognita in tea plants. Phytopathology, 110(2), 405–412

Jokar, A., Masoomi, F., Sadeghpour, O., Nassiri-Toosi, M., & Hamedi, S. (2016). Potential therapeutic applications for Terminalia chebula in Iranian traditional medicine. Chung i Tsa Chih Ying Wen Pan [Journal of Traditional Chinese Medicine], 36(2), 250–254. https://doi.org/10.1016/s0254-6272(16)30035-8

Kadam PV, Yadav KN, Karanje AS, Mukadam RJ, Giram DK and Patil MJ. (2023). Karkatshringi (Pistacia integerrima): phytopharmacological review. Int J Pharm Sci & Res, 14(3), 1063–1070

Kanade, M. L., & Awalaskar, A. (2021). Kanade ML and Awalaskar A:The comparative antimicrobial activity of karkatshringi. National Journal of Research in Ayurved, 9(3), 1–12

Kanchiswamy, C. N., Malnoy, M., & Maffei, M. E. (2015). Bioprospecting bacterial and fungal volatiles for sustainable agriculture. Trends in Plant Science, 20(4), 206–211. https://doi.org/10.1016/j.tplants.2015.01.004

Kannan, V. R., Rajasekar, G. S., Rajesh, P., Balasubram, V., Ramesh, N., Solomon, E. K., Nivas, D., & Chandru, S. (2012). Anti-diabetic activity on ethanolic extracts of fruits of Terminalia chebula Retz. Alloxan induced diabetic rats. American Journal of Drug Discovery and Development, 2(3), 135–142. https://doi.org/10.3923/ajdd.2012.135.142

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics, 20(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Kaur, B., & Singh, S. (2015). A review on gall karkatshringi. Journal of Medicinal Plant Research, 9(21), 636–640. https://doi.org/10.5897/jmpr2014.5404

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Research, 44(9), e89–e89. https://doi.org/10.1093/nar/gkw092

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. Genome Research, 12(4), 656–664. https://doi.org/10.1101/gr.229202

Khare, C. P. (2007). Pistacia integerrima Stewart ex Brandis. In Indian Medicinal Plants (pp. 1–1). Springer New York

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology, 14(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36

Klimov, P. B., Chetverikov, P. E., Dodueva, I. E., Vishnyakov, A. E., Bolton, S. J., Paponova, S. S., Lutova, L. A., & Tolstikov, A. V. (2022). Symbiotic bacteria of the gall-inducing mite Fragariocoptes setiger (Eriophyoidea) and phylogenomic resolution of the eriophyoid position among Acari. Scientific Reports, 12(1), 3811. https://doi.org/10.1038/s41598-022-07535-3

Knief, C. (2014). Analysis of plant microbe interactions in the era of next generation sequencing technologies. Frontiers in Plant Science, 5, 216. https://doi.org/10.3389/fpls.2014.00216

Korf, I. (2004). BMC Bioinformatics, 5(1), 59. https://doi.org/10.1186/1471-2105-5-59

Krueger, F. (2015). Trim galore, A Wrapper Tool around Cutadapt FastQC to Consistently Apply Qual. Adapt. Trimming to FastQ Files. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Kumar, P., Dubey, R. C., & Maheshwari, D. K. (2012). Bacillus strains isolated from the rhizosphere showed plant growth promoting and antagonistic activity against phytopathogens. Microbiological Research, 167(8), 493–499. https://doi.org/10.1016/j.micres.2012.05.002

Kunkel, B. N., & Harper, C. P. (2018). The roles of auxin during interactions between bacterial plant pathogens and their hosts. Journal of Experimental Botany, 69(2), 245–254. https://doi.org/10.1093/jxb/erx447

Kuraku, Z., & Nishimura, K. (2013). Katoh 2013 (Nucleic Acids Research 41:W22-W28) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Research, 41

Kuroda, H., Kuroda, K., & Fujii, H. (2005). Accumulation of benzyl glucosinolate and its neoglucobrassicin derivative in galls induced by Schlechtendalia chinensis (Hemiptera: Aphididae) on Japanese elm. Biotechnology, and Biochemistry, 69(10), 1832–1836

Kurzfeld-Zexer, L., & Inbar, M. (2021). Gall-forming aphids are protected (and benefit) from defoliating caterpillars: the role of plant-mediated mechanisms.

BMC Ecology and Evolution, 21(1), 124. https://doi.org/10.1186/s12862-021-01861-2

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Wilks, C., Antonescu, V., & Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. Bioinformatics (Oxford, England), 35(3), 421–432. https://doi.org/10.1093/bioinformatics/bty648

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics, 12(1). https://doi.org/10.1186/1471-2105-12-124

Li, C., Wang, M., Qiu, X., Zhou, H., & Lu, S. (2021). Noncoding RNAs in medicinal plants and their regulatory roles in bioactive compound production. Current Pharmaceutical Biotechnology, 22(3), 341–359. https://doi.org/10.2174/1389201021666200529101942

Li, F.-W., & Harkess, A. (2018). A guide to sequence your favorite plant genomes. Applications in Plant Sciences, 6(3), e1030. https://doi.org/10.1002/aps3.1030

Li, J., Lu, Z., Yang, Y., Hou, J., Yuan, L., Chen, G., Wang, C., Jia, S., Feng, X., & Zhu, S. (2021). Transcriptome analysis reveals the symbiotic mechanism of Ustilago esculenta-induced gall formation of Zizania latifolia. Molecular Plant-Microbe Interactions: MPMI, 34(2), 168–185. https://doi.org/10.1094/MPMI-05-20-0126-R

Li, K., Diao, Y., Zhang, H., Wang, S., Zhang, Z., Yu, B., Huang, S., & Yang, H. (2011). Tannin extracts from immature fruits of Terminalia chebula Fructus Retz. promote cutaneous wound healing in rats. BMC Complementary and Alternative Medicine, 11(1), 86. https://doi.org/10.1186/1472-6882-11-86

Li, P., Piao, Y., Shon, H. S., & Ryu, K. H. (2015). Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. BMC Bioinformatics, 16(1). https://doi.org/10.1186/s12859-015-0778-7

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics (Oxford, England), 22(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li, Y., Liu, Y., & Zhang, H. (2013). The role of auxins in plant-insect interactions. Frontiers in Plant Science, 4

Liu, Y., Wang, L., & Zhang, H. (2015). The role of glycosides in plant-insect interactions. Frontiers in Plant Science, 6

Lu, M., An, H., & Li, L. (2016). Genome survey sequencing for the characterization of the genetic background of Rosa roxburghii Tratt and leaf ascorbate metabolism genes. PloS One, 11(2), e0147530. https://doi.org/10.1371/journal.pone.0147530

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., … Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience, 1(1). https://doi.org/10.1186/2047-217x-1-18

Mahboubi, M. (2020). Quercus infectoria fruit hulls and galls and female genital disorders. Clinical Phytoscience, 6(1). https://doi.org/10.1186/s40816-020-00194-9

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics, 20(16), 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Manchester, M., & Anand, A. (2017). Metabolomics. In Advances in Virus Research (pp. 57–81). Elsevier

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics, 27(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R., & Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME Journal, 6(3), 610–618. https://doi.org/10.1038/ismej.2011.139

Mehla, K., Balwani, S., Kulshreshtha, A., Nandi, D., Jaisankar, P., & Ghosh, B. (2011). Ethyl gallate isolated from Pistacia integerrima Linn. inhibits cell adhesion molecules by blocking AP-1 transcription factor. Journal of Ethnopharmacology, 137(3), 1345–1352. https://doi.org/10.1016/j.jep.2011.07.068

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. Nucleic Acids Research, 49(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research, 35(Web Server issue), W182-5. https://doi.org/10.1093/nar/gkm321

Mukhtar, M. S., Carvunis, A.-R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M. T., Pevzner, S. J., Donovan, S. E., Ghamsari, L., Santhanam, B., Romero, V., Poulin, M. M., Gebreab, F., Gutierrez, B. J., Tam, S., … Dangl, J. L. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science (New York, N.Y.), 333(6042), 596–601. https://doi.org/10.1126/science.1203659

Munir, M., Khan, M., Ahmed, M., Bano, A., Ahmed, S., Tariq, K., Tabassum, S., Mukhtar, T., Ambreen, M., & Bashir, S. (2011). Foliar epidermal anatomy of some ethnobotanically important species of wild edible fruits of northern Pakistan. J. Med. Plants Res, 5, 5873–5880

Nabity, P. D., Haus, M. J., Berenbaum, M. R., & DeLucia, E. H. (2013). Leaf-galling phylloxera on grapes reprograms host metabolism and morphology. Proceedings of the National Academy of Sciences of the United States of America, 110(41), 16663–16668. https://doi.org/10.1073/pnas.1220219110

Nadkarni, K., & Nadkarni, A. K. (1976). Indian Materia Medica, Popular Prakashan Pvt. Popular Prakashan Pvt. Ltd, 1, 799–752

Nawaz, M. S., Arshad, A., Rajput, L., Fatima, K., Ullah, S., Ahmad, M., & Imran, A. (2020). Growth-stimulatory effect of quorum sensing signal molecule N-acyl-homoserine lactone-producing multi-trait Aeromonas spp. On wheat genotypes under salt stress. Frontiers in Microbiology, 11, 553621. https://doi.org/10.3389/fmicb.2020.553621

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics (Oxford, England), 29(22), 2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Nicola, G., Berthold, M. R., Hedrick, M. P., & Gilson, M. K. (2015). Connecting proteins with drug-like compounds: Open source drug discovery workflows with BindingDB and KNIME. Database: The Journal of Biological Databases and Curation, 2015, bav087. https://doi.org/10.1093/database/bav087

Nik Mat Zin, N. N. I., School of Health Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia, Wan Mohd Rahimi, W. N. A., Abu Bakar, N., School of Health Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia, & School of Health Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia. (2019). A review of Quercus infectoria (Olivier) galls as a resource for anti-parasitic agents: In vitro and in vivo studies. The Malaysian Journal of Medical Sciences: MJMS, 26(6), 19–34. https://doi.org/10.21315/mjms2019.26.6.3

Novaes, R. M. L., Rodrigues, J. G., & Lovato, M. B. (2009). An efficient protocol for tissue sampling and DNA isolation from the stem bark of Leguminosae trees. Genetics and Molecular Research: GMR, 8(1), 86–96. https://doi.org/10.4238/vol8-1gmr542

Nunes, C., Coelho, R. R., Cruz, A. C. M., De Souza, A. A., & Machado, M. A. (2018). Proteomic analysis of citrus fruit reveals roles of thiamine in the response to Huanglongbing. Journal of Proteomics, 173, 26–39

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research, 44(D1), D733-45. https://doi.org/10.1093/nar/gkv1189

Oates, C. N., Denby, K. J., Myburg, A. A., Slippers, B., & Naidoo, S. (2016). Insect gallers and their plant hosts: From omics data to systems biology. International Journal of Molecular Sciences, 17(11), 1891. https://doi.org/10.3390/ijms17111891

Oates, C. N., Külheim, C., Myburg, A. A., Slippers, B., & Naidoo, S. (2015). The transcriptome and terpene profile of Eucalyptus grandis reveals mechanisms of defense against the insect pest, Leptocybe invasa. Plant & Cell Physiology, 56(7), 1418–1428. https://doi.org/10.1093/pcp/pcv064

Oliveira, D. C., Isaias, R. M. S., Fernandes, G. W., Ferreira, B. G., Carneiro, R. G. S., & Fuzaro, L. (2016). Manipulation of host plant cells and tissues by gall-inducing insects and adaptive strategies used by different feeding guilds. Journal of Insect Physiology, 84, 103–113. https://doi.org/10.1016/j.jinsphys.2015.11.012

Ong, Q., Nguyen, P., Phuong Thao, N., & Le, L. (2016). Bioinformatics Approach in Plant Genomic Research. Current Genomics, 17(4), 368–378. https://doi.org/10.2174/1389202917666160331202956

Orwa, C., & Simons, A. (2009). Agroforestree Database: A Tree Reference and Selection Guide, version 4.0. World Agroforestry Centre ICRAF

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D. A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.-É., Li, S., & Xia, J. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Research, 49(W1), W388–W396. https://doi.org/10.1093/nar/gkab382

Pant, S., & Samant, S. S. (2010). Ethnobotanical observations in the Mornaula reserve forest of Komoun, West Himalaya, India. India. Ethnobotanical Leaflets, 2

Parfitt, D. E., & Badenes, M. L. (1997). Phylogeny of the genus Pistacia as determined from analysis of the chloroplast genome. Proceedings of the National Academy of Sciences of the United States of America, 94(15), 7987–7992. https://doi.org/10.1073/pnas.94.15.7987

Park, S. Y., Kim, J. W., & Lee, S. Y. (2014). The role of terpenoids in plant-insect interactions. Frontiers in Plant Science, 5.

Patel, S., Rauf, A., & Khan, H. (2018). The relevance of folkloric usage of plant galls as medicines: Finding the scientific rationale. Biomedecine & Pharmacotherapie [Biomedicine & Pharmacotherapy], 97, 240–247. https://doi.org/10.1016/j.biopha.2017.10.111

Peck, S. B., Mapes, C. C., Dorchin, N., Heppner, J. B., Buss, E. A., Moya-Raygoza, G., Hoy, M. A., Franz, G., Handler, A. M., Hoy, M. A., Capinera, J. L., Heppner, J. B., Striganova, B. R., Heppner, J. B., Heppner, J. B., Heppner, J. B., Heppner, J. B., Goula, M., Northfield, T. D., … Brewer, W. (2008). Gall Formation. In Encyclopedia of Entomology (pp. 1572–1576). Springer Netherlands

Pelagio-Flores, R., Ortíz-Castro, R., Méndez-Bravo, A., Macías-Rodríguez, L., & López-Bucio, J. (2011). Serotonin, a tryptophan-derived signal conserved in plants and animals, regulates root system architecture, probably acting as a natural auxin inhibitor in Arabidopsis thaliana. Plant & Cell Physiology, 52(3), 490–508. https://doi.org/10.1093/pcp/pcr006

Pratibha, N., Saxena, V. S., Amit, A., D'Souza, P., Bagchi, M., & Bagchi, D. (2004). Anti-inflammatory activities of Aller-7, a novel polyherbal formulation for allergic rhinitis. International Journal of Tissue Reactions, 26(1–2), 43–51

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE, 5(3). https://doi.org/10.1371/journal.pone.00094

Qi, Y., Duan, C., Ren, L., & Wu, H. (2020). Growth dynamics of galls and chemical defence response of Pinus thunbergii Parl. to the pine needle gall midge, Thecodiplosis japonensis Uchida & Inouye (Diptera: Cecidomyiidae). Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-69231-4

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590-6. https://doi.org/10.1093/nar/gks1219

Raja, A., Santha, R., & Kuriakose, G. (2018). Comparative transcriptomic analysis of gall tissue of Sandalwood induced by the mite, Aceria santalae and its effect on host plant defense response. Scientific Reports, 8(1)

Rajopadhye, A. A., Upadhye, A. S., Dandge, C. N., & Naik, D. G. (2016). Essential oil from leaf galls on Pistacia chinensis ssp. integerrima: Chemical composition, in vitro and in vivo antioxidant and hepatoprotective activity. Journal of Essential Oil-Bearing Plants, 19(7), 1648–1659. https://doi.org/10.1080/0972060x.2016.1174079

Raman, A. (2011). Morphogenesis of insect-induced plant galls: facts and questions. Flora, 206(6), 517–533. https://doi.org/10.1016/j.flora.2010.08.004

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Research, 47(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Rauf, A. (2019). A Mini Review on a Pistacia Integerrima Well-Known Medicinal Plant: its Active Phytochemicals with Exciting Pharmacological Profile”. Acta Scientific Nutritional Health, 3, 45–48

Rauf, A., Abu-Izneid, T., Maalik, A., Bawazeer, S., Khan, A., Hadda, T. B., Khan, H., Ramadan, M. F., Khan, I., Mubarak, M. S., Uddin, G., Bahadar, A., & Farooq, U. (2017). Gastrointestinal motility and acute toxicity of pistagremic acid isolated from the galls of Pistacia integerrima. Medicinal Chemistry (Shariqah (United Arab Emirates)), 13(3), 292–294. https://doi.org/10.2174/1573406412666161007145247

Rauf, A., Saleem, M., Uddin, G., Siddiqui, B. S., Khan, H., Raza, M., Hamid, S. Z., Khan, A., Maione, F., Mascolo, N., & De Feo, V. (2015). Phosphodiesterase-1 inhibitory activity of two flavonoids isolated from Pistacia integerrima J. l. stewart galls. Evidence-Based Complementary and Alternative Medicine: ECAM, 2015, 506564. https://doi.org/10.1155/2015/506564

Rauf, A., Uddin, G., Raza, M., Ahmad, B., Jehan, N., Siddiqui, B. S., Molnar, J., Csonka, A., & Szabo, D. (2016). Reversal of multidrug resistance in mouse lymphoma cells by extracts and flavonoids from Pistacia integerrima. Asian Pacific Journal of Cancer Prevention: APJCP, 17(1), 51–55. https://doi.org/10.7314/apjcp.2016.17.1.51

Rauf, A., Uddin, G., Siddiqui, B. S., Khan, A., Khan, H., Arfan, M., Muhammad, N., & Wadood, A. (2014). In-vivo antinociceptive, anti-inflammatory and antipyretic activity of pistagremic acid isolated from Pistacia integerrima. Phytomedicine: International Journal of Phytotherapy and Phytopharmacology, 21(12), 1509–1515. https://doi.org/10.1016/j.phymed.2014.07.015

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Ravi Shankara, B. E., Dhananjaya, B., Ramachandra, Y. L., Rajan, S., Sujan Ganapathy, P. S., Yarla, N., & Richard, S. A. (2016). Evaluating the anticancer potential of ethanolic gall extract of Terminalia chebula (Gaertn.) Retz. (combretaceae). Pharmacognosy Research, 8(3), 209. https://doi.org/10.4103/0974-8490.182919

Ravi Shankara, B. E., Ramachandra, Y. L., Sundara Rajan, S., Kaushik, V. S., & Sujan Ganapathy, P. S. (2012). Pharmacognostic and Phytochemical Studies of Leaf Gall of Terminalia chebula Retz. Used as Karkatashringi in South Indian Markets. The Asian and Australasian Journal of Plant Science and Biotechnology, 6(1), 75–77

Rechinger, K. H. (1969). Flora Iranica, Volume 63: Anacaridiaceae. Flora Iranica Project

Reddy, V. (1990). Cardioprotective activity of the fruit of Terminalia chebula. Fitoterapia, 61, 517–525

Redfern, M. P. (2011). The new naturalist library (Vol. 562). Harper Collins

Reimand, J., Kull, M., Peterson, H., Hansen, J., & Vilo, J. (2007). g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Research, 35(Web Server issue), W193-200. https://doi.org/10.1093/nar/gkm226

RepeatMasker Home Page. (n.d.). Org. Retrieved January 4, 2023, from http://www.repeatmasker.org

Robert Hubley, Arian Smit, Jullien M. Flynn. (n.d.). RepeatModeler. Institute for Systems Biology Is a Nonprofit Scientific Research Organization. https://www.repeatmasker.org/RepeatModeler/

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology, 11(3), R25. https://doi.org/10.1186/gb-2010-11-3-r25

Sahu, K. H., Samal, I., & Boruah, J. (2020). Gall diversity, causal agents, their adaptive significance and gall infestation in sericultural host plan. Journal of Entomology and Zoology Studie, 8(2), 1288–1292

Sahu, K.H, Samal, I, Boruah, J. (2020). Gall diversity, causal agents, their adaptive significance and gall infestation in sericultural host plan. Journal of Entomology and Zoology Studie, 8(2), 1288–1292

Saini, A. K., Sawant, L., & Zahiruddin, S. (2023). LC-MS/MS-based Targeted Metabolomic Profiling of Aqueous and Hydro-alcoholic Extracts of Pistacia integerrima Linn., Quercus infectoria Olivier and Terminalia chebula Retz. Pharmacognosy Magazine, 0. https://doi.org/10.1177/09731296221144809

Saini, A. K., Zahiruddin, S., Newton, K. G., Sawant, L., Mitra, R., Kumar Rai, R., & Ahmad, S. (2023). Validation and standardization of Gallic acid and ellagic acid in Quercus infectoria, Terminalia chebula, and Pistacia integerrima. Journal of AOAC International, 106(5), 1305–1312. https://doi.org/10.1093/jaoacint/qsad070

Saleem, M., Hushum, P., & Harkonen, K. (2002). Pihlaja Inhibition of cancer cell growth by crude extract and phenolics of Terminalia chebula fruit. J Ethnopharmacol, 81, 327–336

Salmerón-Manzano, E., Garrido-Cardenas, J. A., & Manzano-Agugliaro, F. (2020). Worldwide research trends on medicinal plants. International Journal of Environmental Research and Public Health, 17(10), 3376. https://doi.org/10.3390/ijerph17103376

Schlaeppi, K., Bulgarelli, D., Luginbuehl, C., Schulze-Lefert, P., & Panstruga, R. (2014). New insights into the microbiome of the phytosphere. Nature Reviews Microbiology, 12(11), 769–780

Schönrogge, K., Harper, L. J., & Lichtenstein, C. P. (2000). The protein content of tissues in cynipid galls (Hymenoptera: Cynipidae): Similarities between cynipid galls and seeds. Plant, Cell & Environment, 23(2), 215–222. https://doi.org/10.1046/j.1365-3040.2000.00543.x

Schultz, J. C., Edger, P. P., Body, M. J. A., & Appel, H. M. (2019). A galling insect activates plant reproductive programs during gall development. Scientific Reports, 9(1), 1833. https://doi.org/10.1038/s41598-018-38475-6

Schulze-Lefert, P., & Panstruga, R. (2011). A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends in Plant Science, 16(3), 117–125. https://doi.org/10.1016/j.tplants.2011.01.001

Shchemelinin, I., Sefc, L., & Necas, E. (2006). Protein kinases, their function and implication in cancer and other diseases. Folia Biologica, 52(3), 81–100

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Research, 50(W1), W216–W221. https://doi.org/10.1093/nar/gkac194

Shih, T.-H., Lin, S.-H., Huang, M.-Y., Sun, C.-W., & Yang, C.-M. (2018). Transcriptome profile of cup-shaped galls in Litsea acuminata leaves. PloS One, 13(10), e0205265. https://doi.org/10.1371/journal.pone.0205265

Shorthouse, J. D., & Rohfritsch, O. (1992). Biology of insect-induced galls. Oxford University Press

Shrestha, S., Kaushik, V. S., Eshwarappa, R. S. B., Subaramaihha, S. R., Ramanna, L. M., & Lakkappa, D. B. (2014). Pharmacognostic studies of insect gall of Quercus infectoria Olivier (Fagaceae). Asian Pacific Journal of Tropical Biomedicine, 4(1), 35–39. https://doi.org/10.1016/s2221-1691(14)60205-7

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics (Oxford, England), 31(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Sofowora, A., Ogunbodede, E., & Onayade, A. (2013). The role and place of medicinal plants in the strategies for disease prevention. African Journal of

Traditional, Complementary, and Alternative Medicines, 10(5), 210–229. https://doi.org/10.4314/ajtcam.v10i5.2

Spaepen, S., & Vanderleyden, J. (2011). Auxin and plant-microbe interactions. Cold Spring Harbor Perspectives in Biology, 3(4), a001438–a001438. https://doi.org/10.1101/cshperspect.a001438

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics, 19(Suppl 2), ii215–ii225. https://doi.org/10.1093/bioinformatics/btg1080

Stone, G. N., & Schönrogge, K. (2003). The adaptive significance of insect gall morphology. Trends in Ecology & Evolution, 18(10), 512–522. https://doi.org/10.1016/s0169-5347(03)00247-7

Subbannayya, Y., Karthikkeyan, G., Pinto, S. M., Kapoor, S., Tyagi, A., Pervaje, S. K., Pervaje, R., & Prasad, T. S. K. (2018). Global metabolite profiling and network pharmacology of Triphala identifies neuromodulatory receptor proteins as potential targets. J. Protein Proteonomics, 9(2), 101–114

Sun, Y., Shang, L., Zhu, Q.-H., Fan, L., & Guo, L. (2022). Twenty years of plant genome sequencing: achievements and challenges. Trends in Plant Science, 27(4), 391–401. https://doi.org/10.1016/j.tplants.2021.10.006

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research, 47(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Takeda, S., Hirano, T., Ohshima, I., & Sato, M. H. (2021). Recent progress regarding the molecular aspects of insect gall formation. International Journal of Molecular Sciences, 22(17), 9424. https://doi.org/10.3390/ijms22179424

Tamhane, M. D., Thorat, S. P., Rege, N. N., & Dahanukar, S. A. (1997). Effect of oral administration of Terminalia chebula on gastric emptying: an experimental study. Journal of Postgraduate Medicine, 43(1), 12–13

Tamogami, S., Rakwal, R., Kodama, O., & Komatsu, S. (2015). Proteomic analysis of the formation of insect galls in pedunculate oak (Quercus robur). Proteomics, 15(23–24), 4007–4015

Tampakaki, A. P. (2009). Encyclopedia of Microbiology || Plant Pathogens. Bacterial, 655–677. https://doi.org/10.1016/B978-012373944-5.00346-1

Tandon, N., & Yadav, S. S. (2017). Contributions of Indian Council of Medical Research (ICMR) in the area of Medicinal plants/Traditional medicine. Journal of Ethnopharmacology, 197, 39–45. https://doi.org/10.1016/j.jep.2016.07.064

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A. D., Nueda, M. J., Ferrer, A., & Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Research, 43(21), e140. https://doi.org/10.1093/nar/gkv711

Thiel, T., Michalek, W., Varshney, R., & Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theoretical and Applied Genetics, 106(3), 411–422. https://doi.org/10.1007/s00122-002-1031-0

Tirumalai, V., Swetha, C., Nair, A., Pandit, A., & Shivaprasad, P. V. (2019). miR828 and miR858 regulate VvMYB114 to promote anthocyanin and flavonol accumulation in grapes. Journal of Experimental Botany, 70(18), 4775–4792. https://doi.org/10.1093/jxb/erz264

Tkachenko;Maria, A. A., Gancheva, S., Varvara, E., Tvorogova;Lavrenty, G., Danilov;Alexander, V., Predeus;Irina, E., & Dodueva;Ludmila, L. (2021). Transcriptome analysis of crown gall in radish (Raphanus sativus L.) inbred lines. Annals of Applied Biology. https://doi.org/10.1111/aab.12669

Tohge, T., Watanabe, M., Hoefgen, R., & Fernie, A. R. (2013). The evolution of phenylpropanoid metabolism in the green lineage. Critical Reviews in Biochemistry and Molecular Biology, 48(2), 123–152. https://doi.org/10.3109/10409238.2012.758083

Tooker, J. F., & De Moraes, C. M. (2011). Feeding by a gall-inducing caterpillar species alters levels of indole-3-acetic and abscisic acid in Solidago altissima (Asteraceae) stems. Arthropod-Plant Interactions, 5(2), 115–124. https://doi.org/10.1007/s11829-010-9120-5

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology, 28(5), 511–515. https://doi.org/10.1038/nbt.1621

Uddin, G., Rauf, A., & Qaisar, M. (2011). Phytochemical screening of Pistacia chinensis var. integerrima. Integerrima. Middle-East J. Sci. Res, 7, 707–711

Uddin, G., Rauf, A., Arfan, M., Waliullah, Khan, I., Ali, M., Taimur, M., ur-Rehman, I., & Samiullah. (2012). Pistagremic acid a new leishmanicidal triterpene isolated from Pistacia integerrima Stewart. Journal of Enzyme Inhibition and Medicinal Chemistry, 27(5), 646–648. https://doi.org/10.3109/14756366.2011.604853

Uddin, Ghias, Rauf, A., Al-Othman, A. M., Collina, S., Arfan, M., Ali, G., & Khan, I. (2012). Pistagremic acid, a glucosidase inhibitor from Pistacia integerrima. Fitoterapia, 83(8), 1648–1652. https://doi.org/10.1016/j.fitote.2012.09.017

Ullah, Z., Mehmood, R., Imran, M., Malikb, A., & Afzal, R. A. (2012). Flavonoid constituents of Pistacia integerrima. Natural Product Communications, 7(8), 1011–1014. https://doi.org/10.1177/1934578x1200700813

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., & Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. GigaScience, 2(1), 16. https://doi.org/10.1186/2047-217X-2-16

Vogt, T. (2010). Phenylpropanoid biosynthesis. Molecular Plant, 3(1), 2–20. https://doi.org/10.1093/mp/ssp106

Wang, H., Cui, K., Shao, S., Liu, J., Chen, H., Wang, C., Wu, H., Yang, Z., Lu, Q., King-Jones, K., & Chen, X. (2017). Molecular response of gall induction by aphid Schlechtendalia chinensis (Bell) attack on Rhus chinensis Mill. Journal of Plant Interactions, 12(1), 465–479. https://doi.org/10.1080/17429145.2017.1392627

Wang, L., Liu, Y., & Zhang, H. (2013). The role of phenolic compounds in plant-insect interactions. Frontiers in Plant Science, 4

Wang, Q., Quan, S., & Xiao, H. (2019). Towards efficient terpenoid biosynthesis: manipulating IPP and DMAPP supply. Bioresources and Bioprocessing, 6(1). https://doi.org/10.1186/s40643-019-0242-z

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews. Genetics, 10(1), 57–63. https://doi.org/10.1038/nrg2484

Warrier, P. K., Nambiar, V., & Ramankutty, C. (1995). Indian Medicinal Plants, A Compendium of 500 species

William, J. J. (2010). High-throughput sequencing and metagenomics. Estuar Coast, 33, 944–952

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology, 15(3). https://doi.org/10.1186/gb-2014-15-3-r46

Wool, D. (2012). Autecology of Baizongia pistaciae (L.): a monographical study of a galling aphid. https://www.semanticscholar.org/paper/b0d2f5c5715af4898f7ab1d4da2084bd3af074ba

Wool, D. (2013). Aphid-induced galls on Pistacia in the natural Mediterranean forest of Israel: Which, where, and how many? Israel Journal of Zoology, 41

Wu, H., Liu, Q., Liu, L., Wu, X., Wu, Y., & Huang, W. (2018). Transcriptome profiling and identification of genes involved in gall formation of Populus simonii × Populus nigra. BMC Genomics, 19(1), 1–16

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Cambridge (Mass.)), 2(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141

Wu, Y., Zhang, H., & Liu, Y. (2013). The role of cytokinins in plant-insect interactions. Frontiers in Plant Science, 4

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q., & Wang, Y. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Research, 47(W1), W52–W58. https://doi.org/10.1093/nar/gkz333

Yamaguchi, H., Tanaka, H., Hasegawa, M., Tokuda, M., Asami, T., & Suzuki, Y. (2012). Phytohormones and willow gall induction by a gall-inducing sawfly. The New Phytologist, 196(2), 586–595. https://doi.org/10.1111/j.1469-8137.2012.04264.x

Yamamoto, T., Yokoyama, J., & Ozaki, K. (2016). Tetraneura nigriabdominalis (Homoptera: Pemphigidae) induces galls on Carpinus tschonoskii (Betulaceae) by accumulating excess amounts of free amino acids. Applied Entomology and Zoology, 51(3), 385–392

Yamamura, Y., Shimada, M., Sugimoto, K., & Fujimoto, H. (2000). Jasmonic acid and salicylic acid are involved in the regulation of the wound-induced transcript accumulation of diplolepis japonica-specific polygalacturonase in the wild rose. Plant Science, 153(1), 63–69

Yamazaki, K. (2016). Caterpillar mimicry by plant galls as a visual defense against herbivores. Journal of Theoretical Biology, 404, 10–14. https://doi.org/10.1016/j.jtbi.2016.05.028

Yang L., Wang Y.-Q., Huang X., & Yang Z.-M. (2021). Quercetin for the treatment of prostate cancer: Progress in studies. Zhonghua nan ke xue [National journal of andrology], 27(7), 654–658

Yang, Y., Huihui, Xu, Li, Xie, J., Jiajie, Renyong, Juchun, Yuanfeng, Lizi, Changliang, Xiaoxia, Guizhou, & Yin, Z. (2017). Preparation of Galla Chinensis Oral Solution as well as Its Stability, Safety, and Antidiarrheal Activity Evaluation. Evidence-Based Complementary and Alternative Medicine, 1–8. https://doi.org/10.1155/2017/1851459

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. Omics: A Journal of Integrative Biology, 16(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Zahoor, M., Zafar, R., & Rahman, N. U. (2018). Isolation and identification of phenolic antioxidants from Pistacia integerrima gall and their anticholine esterase activities. Heliyon, 4(12), e01007. https://doi.org/10.1016/j.heliyon.2018.e01007

Zhang, H., Dugé de Bernonville, T., Body, M., Glevarec, G., Reichelt, M., Unsicker, S., Bruneau, M., Renou, J.-P., Huguet, E., Dubreuil, G., & Giron, D. (2016). Leaf-mining by Phyllonorycter blancardella reprograms the host-leaf transcriptome to modulate phytohormones associated with nutrient mobilization and plant defense. Journal of Insect Physiology, 84, 114–127. https://doi.org/10.1016/j.jinsphys.2015.06.003

Zhang, H., Wang, J., Zhang, Z., Huang, H., Liu, B., & Xie, Z. (2015). Brassinosteroid regulates cell elongation by modulating gibberellin metabolism in Chinese Bayberry. Journal of Plant Growth Regulation, 34(2), 207–218.

Zhang, Y., Zheng, L., Zheng, Y., Zhou, C., Huang, P., Xiao, X., Zhao, Y., Hao, X., Hu, Z., Chen, Q., Li, H., Wang, X., Fukushima, K., Wang, G., & Li, C. (2019). Assembly and annotation of a draft genome of the medicinal plant Polygonum cuspidatum. Frontiers in Plant Science, 10, 1274. https://doi.org/10.3389/fpls.2019.01274

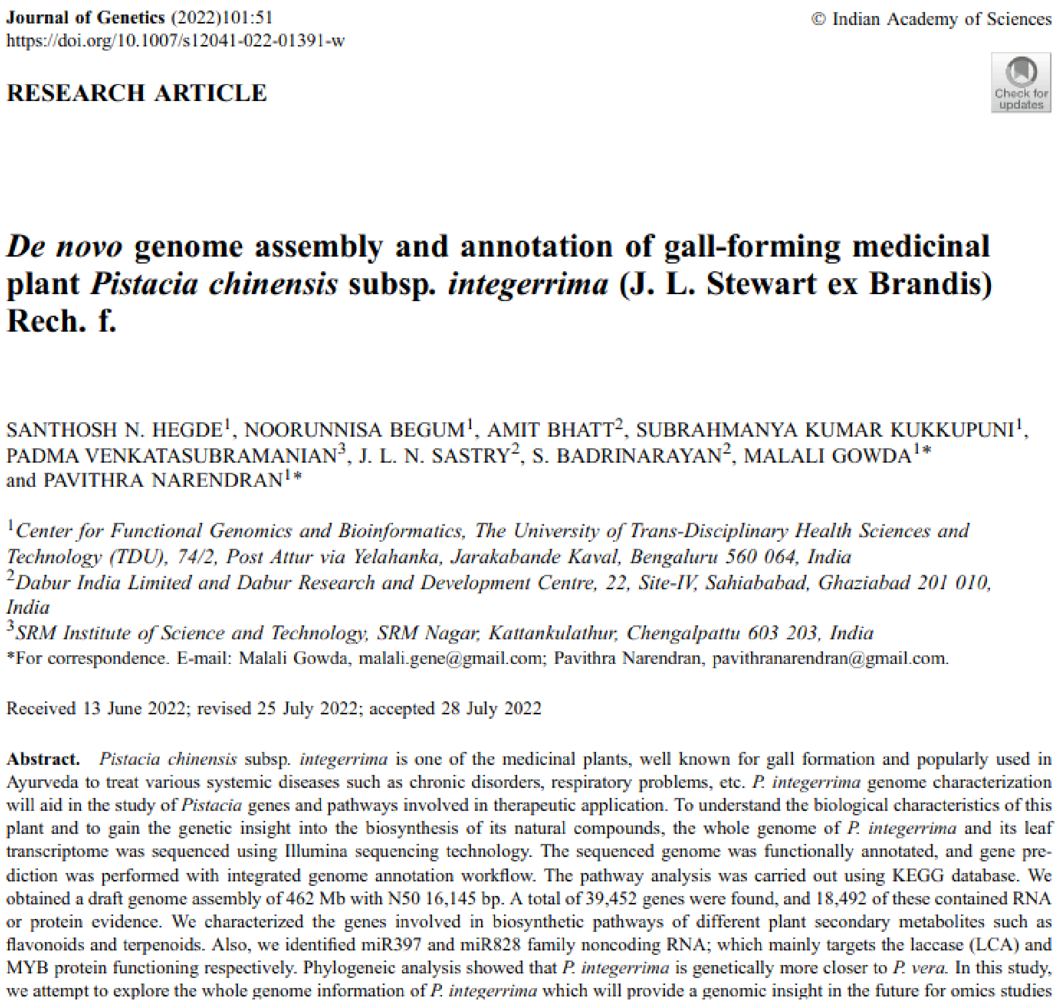
Zhou, G., Pang, Z., Lu, Y., Ewald, J., & Xia, J. (2022). OmicsNet 2.0: a web-based platform for multi-omics integration and network visual analytics. Nucleic Acids Research, 50(W1), W527–W533. https://doi.org/10.1093/nar/gkac376

Ziya, M. E., Kafkas, S., Khodaeiaminjan, M., C¸ Oban, N., & Go¨zel, H. (2016). Genome survey of pistachio (Pistacia vera L.) by next generation sequencing: development of novel SSR markers and genetic diversity in Pistacia species. BMCGenomics, 17

Zohary, M. (1952). A monographical study of the genus Pistacia. Palestine Journal of Botany, Jerusalem Series, 5, 5, 187–228

Zorić, A. S., Morina, F., Toševski, I., Tosti, T., Jović, J., Krstić, O., & Veljović-Jovanović, S. (2019). Resource allocation in response to herbivory and gall formation in Linaria vulgaris. Plant Physiology and Biochemistry, 135, 224–232. https://doi.org/10.1016/j.plaphy.2018.11.032

Appendix



Journal of Genetics (2022)101:51
https://doi.org/10.1007/s12041-022-01391-w

RESEARCH ARTICLE

De novo genome assembly and annotation of gall-forming medicinal plant *Pistacia chinensis* subsp. *integerrima* (J. L. Stewart ex Brandis) Rech. f.

SANTHOSH N. HEGDE[1], NOORUNNISA BEGUM[1], AMIT BHATT[2], SUBRAHMANYA KUMAR KUKKUPUNI[1], PADMA VENKATASUBRAMANIAN[3], J. L. N. SASTRY[2], S. BADRINARAYAN[2], MALALI GOWDA[1]* and PAVITHRA NARENDRAN[1]*

[1]*Center for Functional Genomics and Bioinformatics, The University of Trans-Disciplinary Health Sciences and Technology (TDU), 74/2, Post Attur via Yelahanka, Jarakabande Kaval, Bengaluru 560 064, India*
[2]*Dabur India Limited and Dabur Research and Development Centre, 22, Site-IV, Sahiababad, Ghaziabad 201 010, India*
[3]*SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Chengalpattu 603 203, India*
*For correspondence. E-mail: Malali Gowda, malali.gene@gmail.com; Pavithra Narendran, pavithranarendran@gmail.com.

Received 13 June 2022; revised 25 July 2022; accepted 28 July 2022

Abstract. *Pistacia chinensis* subsp. *integerrima* is one of the medicinal plants, well known for gall formation and popularly used in Ayurveda to treat various systemic diseases such as chronic disorders, respiratory problems, etc. *P. integerrima* genome characterization will aid in the study of *Pistacia* genes and pathways involved in therapeutic application. To understand the biological characteristics of this plant and to gain the genetic insight into the biosynthesis of its natural compounds, the whole genome of *P. integerrima* and its leaf transcriptome was sequenced using Illumina sequencing technology. The sequenced genome was functionally annotated, and gene prediction was performed with integrated genome annotation workflow. The pathway analysis was carried out using KEGG database. We obtained a draft genome assembly of 462 Mb with N50 16,145 bp. A total of 39,452 genes were found, and 18,492 of these contained RNA or protein evidence. We characterized the genes involved in biosynthetic pathways of different plant secondary metabolites such as flavonoids and terpenoids. Also, we identified miR397 and miR828 family noncoding RNA; which mainly targets the laccase (LCA) and MYB protein functioning respectively. Phylogeneic analysis showed that *P. integerrima* is genetically more closer to *P. vera*. In this study, we attempt to explore the whole genome information of *P. integerrima* which will provide a genomic insight in the future for omics studies