Department of Plant Physiology, 2025

A systems genetics approach to identify candidate genes driving salicinoid diversity in *Populus tremula*

Sara Rydman

UMEÅ UNIVERSITY

This work is protected by the Swedish Copyright Act (Act 1960:729)
ISBN: 978-91-8070-866-1 (print)
ISBN: 978-91-8070-867-8 (pdf)
Cover image generated by Microsoft Copilot
Electronic version available at: http://umu.diva-portal.org/
Printed by: Scandinavian Print Group, Skarpnäck, 2025

I vart såm I vart, a I jär såm I jär.
a sä hav I ållti vöre.
Hä går såm hä går, a hä bär dit hä bär,
men än hav int håppe före.
Hä val såm hä val, da man djär hä man djär,
men än järe eingen fåra.
Fast I vart såm I vart, a I jär såm I jär,
a sä ska I ållti våra.

– Erik Grahn (2014)

Table of Contents

Abstract .... **iii**

Sammanfattning .... **v**

Acronyms .... **vii**

List of Publications .... **x**

Authorship Contributions .... **xi**

Objectives .... **xii**

Introduction **1**

Background **4**
- Specialised metabolites in defense and medicine 4
  - *Terpenes and terpenoids* *6*
  - *Phenylpropanoids* *7*
  - *Alkaloids* *9*
  - *Human utilisation of specialised metabolites* *10*
- Chemical defences in Salicaceae 12
  - *Diversity and evidence of bioactivity in salicinoid phenolic glycosides* *14*
- Systems biology 26
  - *The study of 'omics* *27*
  - *Capturing the interplay between 'omics* *37*

Material and Methods **47**
- Plant material: Swedish Aspen Collection 48
- Transcriptomic profiling 50
- Metabolomic profiling 50
- Association studies 52
- Network analyses 52
- No-choice insect experiment 53
- Image analysis of foliar damage 54

Results and Discussion **56**
- An Improved Chromosome-scale Genome Assembly and Population Genetics resource for *Populus tremula* (Paper I) 56

An integrated multi-omic resource to identify candidate genes in the salicinoid phenolic glycosides pathway (Paper II) 58
A systems genetics study reveals novel candidate genes for salicinoid biosynthesis in aspen (Paper III) 61
Herbivory by *Orgyia antiqua* is unaffected by salicinoid phenolic glycoside chemotypes in *Populus tremula* 63

Conclusions and Perspectives 69

Acknowledgements 74

References 77

Abstract

Trees have evolved an impressive array of strategies to cope with the challenges of having a long and sessile life. Not only must they withstand a fluctuating climate, but they also face instantaneous pressures from herbivores and other attackers. To protect themselves, plants can produce defence compounds, many of which are highly specialised and taxon specific. Within the Salicaceae family, a key group of such defence compounds are the salicinoid phenolic glycosides (SPGs). Many structural variants of SPGs have been identified, in which acyl groups (e.g., cinnamoyl, benzoyl, and acetyl) are common. Some of these SPGs can have toxic and deterrent effects against attackers, and a few are known for their medicinal properties in humans. However, the biological function of most SPGs *in planta* remains unclear, and the causal enzymes for the majority of SPGs are yet to be identified.

The aim of this thesis was to uncover the genetic basis of SPG biosynthesis in European aspen (*Populus tremula* L.) and to determine the extent of ontogenic and organ-specific variation among individuals. To achieve this, SPG variation within a collection of natural aspen, the Swedish aspen (SwAsp) collection, was investigated using an integrative multi-omic approach. By analysing the metabolome and transcriptome of multiple leaf ages from aspen individuals with varying levels of cinnamoyl and acetylated SPGs, a set of candidate transferases and novel putative SPGs were identified. These analyses further suggested that young leaf tissue is a highly active site of SPG biosynthesis, compared with mature leaves.

To extend this analysis, we performed genome-wide association studies on transcriptomic and metabolomic data from leaf buds to identify genomic regions associated with variation in SPG abundance and gene expression. These data were integrated into a systems genetics network, visualising the intricate relationship between candidate genes and the diversity of SPGs. Among the candidates, an acyltransferase was highly associated with both acetyl- and cinnamoyl-SPGs. Heterologous expression assays in *Escherichia coli* (*E. coli*) confirmed its acetylation activity. In line with these findings, overexpression of the gene *in planta* led to increased levels of acetyl-SPGs, suggesting acetylation activity of the enzyme.

In summary, these results have enhanced our understanding of SPG biosynthesis and provide a foundation for future studies aimed at elucidating the *in planta* function of the remaining candidate genes.

Keywords: aspen | *Populus tremula* | systems genetics | GWAS | eQTL | metabolomics | specialised metabolites | salicinoid phenolic glycosides | chemotype | liquid chromatography-mass spectrometry | transcriptomics | RNA-Seq

Sammanfattning

Träd har utvecklat en imponerande uppsättning strategier för att hantera utmaningarna som följer med ett långt och stationärt liv. De måste inte bara tåla ett fluktuerande klimat, utan utsätts också för plötsliga angrepp från växtätare och andra angripare. För att skydda sig kan växter producera olika försvarsämnen, av vilka många är taxonspecifika. Inom familjen Salicaceae utgör sekundär fenol glykosider (SPGs) en viktig grupp av sådana försvarsämnen. Många strukturella varianter av SPGs har identifierats, där acylgrupper (ex., cinnamoyl, benzoyl, samt acetyl) är vanligt förekommande. Vissa av dessa SPGs kan ha toxiska och avskräckande effekter på angriparen, och några har läkande egenskaper på människor. Trots detta är den biologiska funktionen för de flesta SPGs i växten okänd, och de synteserande enzymen för majoriteten av SPGs är ännu inte identifierade.

Syftet med denna avhandling var att kartlägga den genetiska grunden för SPG biosyntes i asp (*Populus tremula* L.) samt fastställa omfattningen av ontogenetisk och organ-specifik variation bland individerna. För att åstadkomma detta analyserades variationen av SPGs inom en naturlig aspkollektion, den svenska aspkollektionen (SwAsp), med hjälp av en integrativ multi-omic strategi. Genom att analysera metabolomet och transkriptomet från löv av olika åldrar hos aspindivider med olika nivåer av cinnamoyl och acetylerade SPGs identifierade vi ett antal kandidattransferaser samt nya potentiella SPGs. Analyserna tyder även på att unga löv är högst aktiva SPG-producerande vävnader i jämförelse med äldre löv.

För att vidareutveckla denna analys, genomförde vi även helgenomstudier på transkriptomet och metabolomet från lövknoppar för att identifiera genomiska regioner associerade med variation i SPG-nivå samt genuttryck. Dessa data integrerades till ett systemgenetiskt nätverk som visualiserade den intrikata relationen mellan kandidatgener och diversiteten av SPGs. Bland kandidaterna fanns en acyltransferas som var starkt associerad med både acetyl- och cinnamoyl-SPGs. Heterologt uttryck av genen i *Escherichia coli* (*E. coli*) bekräftade enzymets acetylerande egenskaper. I linje med dessa upptäckter ledde överuttryck av genen i växten till ökade nivåer av acetyl-SPGs, vilket tyder på acetyleringsaktivitet hos enzymet.

Sammantaget har dessa resultat ökat vår förståelse av SPG-biosyntesen i asp samt utgör en grund för framtida studier ämnade att kartlägga funktionen av de återstående kandidatgenerna i växten.

Acronyms

ATAC-Seq Assay for Transposable Accessible Chromatin Sequencing

BLUP Best Linear Unbiased Prediction

cDNA Complementary DNA

CNV Copy Number Variation

CT Condensed Tannins

DAM Differentially Abundant Metabolites

DEG Differentially Expressed Genes

DGE Differential Gene Expression

DNA Deoxyribonucleic Acid

ddNTP Dideoxyribonucleoside Triphosphates

DPI Dots Per Inch

eQTL Expression Quantitative Trait Locus

ESI Electrospray Ionisation

GC-MS Gas Chromatography-Mass Spectrometry

GWA Genome-Wide Association

GWAS Genome-Wide Association Study

HGP Human Genome Project

HPLC High-Performance Liquid Chromatography

LC-MS Liquid Chromatography-Mass Spectrometry

lncRNA Long Non-Coding RNA

miRNA MicroRNA

mQTL Metabolomic Quantitative Trait Locus

mRNA Messenger RNA

MS/MS Tandem Mass Spectrometry

ncRNA Non-Coding RNA

NGS Next-Generation Sequencing

pQTL Protein Quantitative Trait Locus

QqQ Triple Quadrupole Mass Spectrometer

Q-TOF Quadrupole Time-of-Flight Spectrometer

QTL Quantitative Trait Locus

RNA Ribonucleic Acid

RNA-Seq RNA Sequencing

rRNA Ribosomal RNA

SNP Single Nucleotide Polymorphism

SPG Salicinoid Phenolic Glycoside

TPS Terpene Synthases

tRNA Transfer RNA

UHPLC Ultra-High Performance Liquid Chromatography

UPSC Umeå Plant Science Center

VOC Volatile Organic Compound

VSN Variance Stabilizing Normalization

VST Variance Stabilizing Transformation

WGCNA Weighted Gene Co-expression Network Analysis

List of Publications

1. Robinson, K. M.*, Schiffthaler, B.*, Liu, H., **Rydman, S. M**., Rendón-Anaya, M., Kalman, T. A., Kumar, V., Canovi, C., Bernhardsson, C., Delhomme, N., Jenkins, J., Wang, J., Mähler, N., Richau, K. H., Stokes, V., A'Hara, S., Cottrell, J., Coeck, K., Diels, T., Vandepoele, K., Mannapperuma, C., Park, E-J, Plaisance, S., Jansson, S., Ingvarsson, P. K. & Street, N. R. (2024). An Improved Chromosome-scale Genome Assembly and Population Genetics resource for *Populus tremula*. *Physiologia Plantarum, 176*(5), e14511.

2. **Rydman, S. M**., Lihavainen, J., Robinson, K. M., Jansson, S., Albrectsen, B. R. & Street, N. R. (2025). A Metabolomics and Transcriptomics Resource for Identifying Candidate Genes in the Biosynthesis of Specialised Metabolites in *Populus tremula*. *Physiologia Plantarum,* 177(5), e70567.

3. Robinson, K. M.*, **Rydman, S. M.*,** Kumar, V., Lihavainen-Bag, J., Cottrell, J., Albrectsen, B. R., Jansson, S. & Street, N. R. A systems genetics study reveals novel candidate genes for salicinoid biosynthesis in aspen [Unpublished manuscript]

* these authors contributed equally to this work.

Authorship Contributions

1. Robinson et al., I designed and implemented the best linear unbiased prediction (BLUP) pipeline used to preprocess the traits for genome-wide association studies (GWAS). I also mapped expression quantitative trait loci (eQTL) using fastJT in R.

2. Rydman et al., I generated, preprocessed, and explored the metabolite and transcriptomic data, and co-wrote the manuscript in collaboration with the other authors.

3. Robinson et al., I created the BLUP and Qst pipeline, performed GWAS on the Swedish Aspen (SwAsp) metabolomic data from leaf buds, and conducted the network analysis. I also performed the integrative analysis to create a systems genetics network.

Objectives

The primary objective of this thesis was to elucidate the genetic basis of salicinoid phenolic glycoside (SPG) biosynthesis by exploiting natural variation within a collection of wild European aspen (*Populus tremula* L.; SV asp) known as the Swedish Aspen (SwAsp) collection. More specifically:

In Paper I, we provide an updated genome version of *P. tremula* generated using long-read sequencing, complemented with optical and high-density genetic maps. We demonstrate the utility of the new genome version by performing genome-wide association studies (GWAS) to find single nucleotide polymorphisms (SNPs) associated with leaf physiognomy phenotypes and gene expression in buds. Moreover, we also identified candidate long non-coding RNAs and accessible chromatin regions using Assay for Transposable Accessible Chromatin sequencing (ATAC-Seq).

In Paper II, we present a metabolomic and transcriptomic resource for identifying candidate genes in the SPG biosynthesis pathway. The 'omics data were derived from various organs, ranging from shoot tips to roots, of SwAsp individuals with different SPG profiles, allowing us to measure the extent of ontogenic and organ-specific variation in SPG abundance and diversity. Exploration of the resource led to the identification of three novel putative candidate genes in the SPG pathway.

In Paper III, we present a framework for identifying candidate genes in the SPG biosynthesis pathway using a systems genetics approach. Specifically, whole-genome resequencing, transcriptomic and metabolomic data from the SwAsp collection were integrated into a systems genetic network, allowing us to identify genes associated with

SPG abundance. An acyltransferase emerged as a strong candidate, being associated with both cinnamoyl- and acetyl-SPGs. Consistent with this, the enzyme acetylated a set of SPGs in a heterologous expression assay in *Escherichia coli* (*E. coli*), an activity further supported by overexpression of the gene *in planta*.

Within this thesis, I also present results from an herbivore experiment designed to assess differences in feeding damage by a generalist larva (*Orgyia antiqua*; SV aprikostofsspinnare) among SwAsp individuals with different SPG profiles.

Introduction

[T]he trees in whose shade we wander, as well as the plants which we meet time and again, are chemically unknown entities. We know that salicin occurs in the willow, populin in the poplar; we know amygdalin of the almond and the volatile oils of chamomile and sage, but we do not have a picture of the exact chemical composition of any of these plants. One who knows the chemical composition of valerian oil, nevertheless, knows little about the composition of the plant. Frequently one finds that more is known about the cinchona tree or other foreign plants than is known about the linden or the nut trees of our own yards.

Rochleder (1858) as cited in Burrell (1937)

Although Rochleder wrote the original quote over a century ago, the text remains relevant today: most plants are still chemically uncharacterised. This is not surprising, given that researchers estimate there are approximately 500,000 land plant species globally – many of which have yet to be discovered (Corlett, 2016). These plants are, in turn, estimated to produce about 1,000,000 different compounds (Afendi et al., 2012), although the true value is likely to be significantly higher (Perez de Souza et al., 2021). Thus, much of the vast chemical diversity within the plant kingdom remains to be explored.

Rather than focusing on all species within the plant kingdom, the scientific community generally concentrates on a few selected species referred to as model plants. These model species are generally not chosen for our interest in their chemical properties but for attributes such as short generation time, small genomes, or ease of breeding or genetic manipulation. Among the established plant model species, we find

Arabidopsis (*Arabidopsis thaliana*; SV backtrav), purple false brome (*Brachypodium distachyon*; SV grusskafting), barrel clover (*Medicago truncatula*; SV tornlusern) (Cesarino et al., 2020), and European aspen (*P. tremula*) (Jansson & Douglas, 2007). While research on these species has significantly advanced our knowledge of plant biology, we do not know their exact chemical composition, emphasising the complexity of plant metabolomes and the immense, ongoing effort to characterise even a single species. In addition, biosynthesis can vary between tissues and developmental stages (Balandrin et al., 1985), adding yet another layer of complexity to metabolomic studies.

Beyond chemically characterising plants, a major challenge is knowing the function of a compound within the plant: Does the compound protect the plant against insects?; provide resistance to drought?; or is it just a by-product of metabolic processes? To answer these questions, researchers are increasingly turning to holistic approaches. Rather than viewing metabolism as a linear cascade from DNA to RNA to protein to metabolite, these systems-level approaches use so-called omic techniques (e.g., genomics, transcriptomics, proteomics, and metabolomics) to capture the dynamic and interconnected nature of biological information flow. By integrating multiple layers of biological data, researchers can potentially unravel the molecular mechanisms underlying a compound; identifying genes responsible for their biosynthesis and uncovering how perturbations at one level affect the entire system, providing valuable insight into the role of the compound within the plant and its extended effects on other organisms interacting with the plant.

Even if much appears to remain the same since Rochleder reflected on the chemical mystery of the plant kingdom over a century ago, the scientific landscape has fundamentally changed. Today, we have powerful tools, like

‘omic technologies, to tackle many of the previously intractable questions. Moreover, understanding of one biological system, such as a model plant, can give useful clues for unravelling the biology of other plant species. Since all plants share a common ancestor, the same genes can be conserved across species, especially those within the same family of plants. Therefore, uncovering the function of a single gene in one species can be beneficial for researchers studying another species, as that gene might perform a similar role. In this way, research can have far-reaching implications, potentially shedding light on the linden or nut tree growing in our own yard.

Background

Specialised metabolites in defense and medicine

Plants are sessile and cannot escape harsh conditions by moving and must adopt other methods to avoid being consumed. Walters (2010) is reviewing evolutionary strategies that might explain plants' ability to avoid being eaten. One of these hypotheses suggests that plants use their metabolic machinery to produce deterrent and toxic compounds, called specialised (or secondary) metabolites, which, through evolution, have been advantageous. Specialised metabolites are more or less taxon-specific and often defined as non-essential for plant development or growth, instead serving adaptive and ecological functions, including resistance to biotic and abiotic stress, communication between individuals, attracting pollinators, and facilitating beneficial microbial interactions (Hartman, 2007; Hartmann, 2008). Based on this terminology, specialised metabolites are thus distinct from general (or primary) metabolites, which are compounds needed for growth and development (Salam et al., 2023). However, the distinction between general and specialised metabolism has been questioned, as several studies have shown that specialised metabolites can influence plant development (Durán-Medina et al., 2021), for example, in roots (Wasson et al., 2006) and pollen (Muhlemann et al., 2018). Moreover, it is increasingly accepted that specialised metabolism is derived from and deeply interconnected with general metabolism through processes such as gene duplication, mutation, precursor utilisation (or competition), and enzyme promiscuity (Ji et al., 2024). Thus, adopting an integrated perspective on general and specialised metabolism may better reflect plant biology (Durán-Medina et al., 2021).

While the total number of plant metabolites is estimated to exceed one million, a large proportion, possibly ranging from 200,000 to 1,000,000, are thought to be specialised compounds (Wu et al., 2023). While phytochemists and plant physiologists initially regarded them as "metabolic waste" in the 1950s, improved knowledge of their biochemistry eventually changed the attitude towards specialised metabolites, recognising them as dynamic constituents of metabolism (Hartmann, 2007). Specialised metabolites can directly affect herbivores by tasting bad, affecting nutrient uptake, or being toxic, for example, by interacting with enzymes or interfering with metabolic processes. The effect of compounds varies among attackers, and some specialist insects are highly adapted to feed on these metabolites by excreting or detoxifying them, with some using them for their own defence. Plants can also use indirect defences by releasing volatile organic compounds (VOCs) to attract enemies of the plant attacker. As these specialised metabolites require energy and resources that could otherwise be used for growth and development, they can be costly for the plant to produce. These costs vary among metabolites, with rarer elements being more precious and difficult to obtain. For example, nitrogen-containing metabolites are thought to be more energetically demanding to synthesise than those without nitrogen (Mithöfer & Boland, 2012).

Initially, the study of specialised metabolites focused mainly on setting up chemical classes (e.g., terpenoids, alkaloids, phenylpropanoids, and polyketides) and on their distribution within and across plant species (Hartmann, 2007). Some of these are present in most plant species, whereas others are confined to only a few plant families or individual species (Walters, 2010). Thanks to the availability of radioactive labelling

techniques in the 1950s, the subsequent decades involved the discovery of biosynthetic routes for many classes of specialised metabolites (Hartmann, 2007). As these metabolites can be harmful to the plant itself, they are often neutralised by adding a sugar moiety and are often stored in the vacuole or apoplast, minimising interference with metabolic processes and cell structures (Mithöfer & Boland, 2012; Walters, 2010). Among the major classes of specialised metabolites, we find terpenes, phenolics, and alkaloids (Nawrot-Chorabik et al., 2022), which will briefly be reviewed in terms of their chemical structures, biosynthetic pathways, and links to general metabolism. Examples of arthropod defence properties and human usage of each class will also be provided.

Terpenes and terpenoids

Terpenes are a large and diverse class of lipophilic specialised metabolites (Salam et al., 2023) with more than 65,000 identified structures (Wu et al., 2023). The building blocks of terpenes are isoprenoid units (C_5H_8), which are also used for classification, ranging from hemiterpenes (C_5) to polyterpenes ($>C_{40}$) (Singh & Sharma, 2015). Terpenes that have been oxidised or that have had their carbon skeleton rearranged are called terpenoids (Abdel-Aziz et al., 2016). Nevertheless, terpenes and terpenoids are often referred to interchangeably in the literature (Boncan et al., 2020).

Terpenes are biosynthesised from two precursors: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). These precursors originate from either the mevalonate pathway (MVA) or the 2C-methyl-D-erythritol-4-phosphate (MEP) pathway, both of which utilise glucose derived from glycolysis in plants (Câmara et al., 2024). MVA is present in animals, archaea, plants, and some bacteria, whereas

MEP is restricted to plants, bacteria, and green algae. IPP and DMAPP undergo a number of enzymatic reactions (such as structural rearrangements, repetition, and cyclisation) to produce terpenes. The vast diversity of terpenes, however, stems from the activity of the terpene synthases superfamily (TPS) (Ninkuu et al., 2021) and P450 enzymes (Wu et al., 2023).

Terpenes are known to play a role in plant defences against microbes and herbivores (Ninkuu et al., 2021) and are important for plant-insect interactions (Singh & Sharma, 2015). Terpenoids constitute a considerable portion of VOC, which can be used to attract or repel insects. A study on plant pollination showed that *Silene latifolia* (SV vitblära) releases monoterpenoids to attract *Hadena bicruris* (SV större vitblärefly) (Dötterl et al., 2006). Another study illustrated the toxic effects of terpenoid constituents on the granary weevil (*Sitophilus granarius* L.; SV kornvivel), presenting an alternative pest-control method (Plata-Rueda et al., 2018). Among the highly successful terpenoid-based biopesticides, we find azadirachtin, extracted from the seeds of the Indian neem tree (*Azadirachta indica;* SV nimträd), which is an efficient insect antifeedant and growth disruptor (Kilani-Morakchi et al., 2021). Other applications of volatile terpenes include perfumes and flavours, including ginger, saffron, wine, basil, citrus fruits, cannabis, and tea (Pichersky & Raguso, 2018; Elshafie et al., 2023). Moreover, several terpenes and terpenoids have shown antimicrobial, anticancer, and anti-inflammatory effects (Masyita et al., 2022).

Phenylpropanoids

Phenylpropanoids are a diverse set of compounds that can be divided into several subgroups based on chemical structure: phenolic acids,

flavonoids, monolignols, stilbenes, and coumarins. The three former groups are found in most land plants, whereas stilbenes and coumarins have a more restricted distribution in land plants (Deng & Lu, 2017). A common feature of phenylpropanoids is their structure, which includes a phenyl group and a three-carbon side chain (C6-C3) (Wu et al., 2023).

In most plants, phenylpropanoids are synthesised from the amino acid phenylalanine, which is formed via a general metabolic pathway called the shikimate pathway (Dong & Lin, 2021). The core phenylpropanoid pathway includes three key enzymes; phenylalanine is first converted into cinnamic acid by phenylalanine ammonia-lyase (PAL), which is hydroxylated to *p*-coumaric acid by cinnamate 4-hydroxylase (C4H). The next enzymatic reaction involves the transformation of *p*-coumaric acid into *p*-coumaroyl-CoA, which branches into different phenylpropanoid subgroups. These compounds (e.g., lignins, flavonoids, and stilbenes) are synthesised through the action of enzymes such as transferases, reductases, lyases, cytochrome P450 membrane-bound monooxygenases (P450), and others (Deng & Lu, 2017).

Phenylpropanoids are involved in a myriad of processes in plants, including plant resistance (Rahim et al., 2023), photoprotection, structure and growth (Weng & Chapple, 2010). Monolignols, for example, are the building blocks of lignin, an aromatic polymer that provides mechanical support and allows water transport from roots to shoots (Dixon & Barros, 2019). Flavonoids also play a vital role in plants, involved in, among other things, pollination, seed development, abiotic stress, and plant defence. For example, the application of flavonoids on the abaxial leaf surface negatively affected reproduction, feeding, and honeydew production in the cotton aphid (*Aphis gossypii*; SV gurkbladlus) (Zhang et al., 2025). Another study observed higher mortality rate and reduced body weight

in corn earworm (*Helicoverpa zea*) feeding on maize lines with high flavonoid levels (Chatterjee et al., 2025). Moreover, flavonoids have been exploited for their pharmacological activities, including antioxidant, anti-inflammatory, and antimicrobial activities (Chen et al., 2023).

Alkaloids

Alkaloids are nitrogen-containing compounds, comprising about 20,000 molecules. These compounds are commonly derived from amino acids such as lysine, tyrosine, tryptophan, phenylalanine, and ornithine. Although alkaloids have been identified in microorganisms, toads, and insects, they are mainly found in plants (Faisal et al., 2023), present in about 20% of plant species (Ziegler & Facchini, 2008).

Alkaloids are biosynthesised via diverse metabolic pathways and exhibit many structural forms (Faisal et al., 2023; Ziegler & Facchini, 2008). Most alkaloids have a heterocyclic ring system, which, together with their biosynthetic precursors, is commonly used to classify them (Wu et al., 2023). The largest alkaloid classes are indole alkaloids and isoquinoline alkaloids, which together comprise about 8,000 compounds (Faisal et al., 2023).

Alkaloids can serve as nitrogen reservoirs, growth regulators, and protectants against abiotic and biotic stresses (Bhambhani et al., 2021). For example, in a feeding experiment with eri silk moth (*Philosamia ricini*), larval growth and development were negatively affected by pyrrolizidine alkaloids (PA) applied to a non-PA-producing plant (*Ailanthus altissima*; SV gudaträd) (Narberhaus et al., 2005). In another study, the defence role of the alkaloid nicotine was assessed in a trial with *Manduca sexta* (SV tobakssvärmare) larvae feeding on tobacco (*Nicotiana attenuata*) mutants exhibiting severely reduced nicotine

levels. The larvae grew and developed better on the mutant compared to the wild type. Similar results were observed for field-grown mutants, which lost threefold more leaf biomass to native herbivores than the wild type (Steppuhn et al., 2004).

The bioactivity of alkaloids can affect a wide range of metabolic systems in animals, and have been widely used as analgesics (e.g., morphine and codeine), anticancer drugs (e.g., vinblastine and taxol), antimalarial drugs (quinine), nicotine, caffeine, etc (Facchini, 2001; Abdel-Aziz et al., 2016).

Human utilisation of specialised metabolites

Humans have been exploiting the biological activities of specialised metabolites for millennia. Prehistoric humans used, for example, toxic plants such as *Strychnos toxifera* (SV kurarebuske) by soaking arrow tips in the plant's toxic juice, efficiently paralysing their prey by blocking respiration (Nepovimova & Kuca, 2019). Plant toxins were also used as an execution method, as seen in the case of the Greek philosopher Socrates, who was condemned to death in 399 B.C. through the consumption of extracts from the hemlock plant (*Conium maculatum* L.; SV odört) (Dayan, 2024).

Although poisonous when ingested at high doses, some specialised metabolites can be exploited as medicine at lower concentrations (Wink, 1998). One of the earliest documented examples of plant-derived medicines is depicted in a carving on a Sumerian clay slab from around 3000 B.C., which reveals 12 medicinal recipes derived from 250 plants, including myrrh and opium (Tembo et al., 2021; Abdel-Aziz et al., 2016). In addition, the Eber Papyrus, dating back to approximately 1500 B.C., lists over 850 plant medicines, including aloe, garlic, and cannabis (Bryan, 1930; Abdel-Aziz et al., 2016). Other famous records include the Chinese

Materia Medica (1100 B.C.), the Shennong Herbal (ca 100 B.C.) (Cragg & Newman, 2013), and publications by Hippocrates (ca 400 B.C.) (Sinha et al., 2023).

For most of human history, we relied on crude plant extracts. The Swedish-German pharmacist Scheele (1742-1786) isolated citric, malic, oxalic, and tartaric acids from plant-derived materials, thus moving the field from water or alcohol extracts to the pure isolation of plant compounds (Burrell, 1937). The isolation of morphine from *Papaver somniferum* (SV opievallmo) by Sertürner in 1803 became an important milestone, initiating the era of plant-derived drug discovery (Nasim et al., 2022; Hartmann, 2007).

Today, plants continue to be an essential source of medicines, with 80% of the global population dependent on ethnobotanical remedies (Chakraborty, 2018). However, using plants for medicines is challenging due to certain characteristics of specialised metabolite production (Chakraborty, 2018). Plants often make specialised metabolites in specific cell types and developmental stages, in low amounts and with somewhat unpredictable yields/concentrations, thereby constraining isolation and purification (Balandrin et al., 1985; Fu et al., 2018), as exemplified by the production of the antimalarial compound artemisinin in only the trichomes of artemisia plants (SV malörter) (Ikram & Simonsen, 2017). As a result, most uses of single-defined-component medicines have shifted to synthetic production in industrialised countries rather than the isolation of the natural product (De Luca et al., 2012).

Nature remains an active research field for inspiring drug discovery. For example, specialised metabolites in the phenylpropanoid pathway, called salicinoid phenolic glycosides (SPGs), have attracted attention for their potential applications in anti-obesity therapy (Kim et al., 2022), cancer

treatment (Kwon et al., 2014; Shah et al., 2016), and neuroprotection (Kim et al., 2015). Historically, SPGs have been used by humans for their pain-relieving properties, with records traced back to ancient Egypt and the Hippocratic era. During the 18th century, one of these active metabolites, salicin, was isolated from Willows, belonging to the Salicaceae family (Rainsford, 2004). Salicin later had a significant impact on the treatment of acute rheumatism (MacLagan, 1876) and, eventually, on the synthesis of the well-known painkiller Aspirin (Rainsford, 2004).

SPGs are the focus of this thesis and are discussed in depth in the following chapters.

Chemical defences in Salicaceae

Salicaceae is a plant family with a broad geographical distribution, commonly growing in cooler regions of the Northern and Southern hemispheres (Julkunen-Tiitto & Virjamo, 2017). The family currently includes about 54 genera with more than 1,400 species. Notable members of this family include *Populus* and *Salix* species, including willows, aspens, poplars, and cottonwoods. These plants hold significant ecological and economic value as they provide habitats and food for various species and serve as a source of wood products (Ogutcen et al., 2024). Furthermore, *Populus* trees possess certain characteristics that make them attractive as study organisms, including fast growth, clonal propagation, and small genomes. As a result, *Populus* has been designated as a model organism for trees (Jansson & Douglas, 2007; Ogutcen et al., 2024).

There are many research areas where *Populus* serves as an attractive study system (Jansson & Douglas, 2007). Since these species are rich in

phenolic specialised metabolites derived from the phenylpropanoid pathway, they provide a particularly suitable system for studying phenylpropanoid biosynthesis. Phenylpropanoids represent a significant fitness trait for long-lived plants growing in fluctuating environments (Tsai et al., 2006), with the majority of specialised metabolites produced by *Populus* species originating from the phenylpropanoid pathway (Philippe & Bohlmann, 2007). Among these specialised metabolites, two groups are relatively well studied regarding defence: condensed tannins (CTs, proanthocyanidins) and salicinoid phenolic glycosides (SPGs) (Philippe & Bohlmann, 2007).

CTs are a subgroup of flavonoids. The basic structure of CTs is constituted of flavan-3-ol units (commonly catechin or epicatechin) linked together through C-C bonds, forming oligomer and polymer structures (Deng & Lu, 2017; Barbehenn & Constabel, 2011). Stevens and Lindroth (2005) observed a rapid induction (within a week) of CTs in herbivore-damaged leaves of *Populus tremuloides* (SV amerikansk asp), a species that can accumulate large amounts of CT – up to 20% of leaf dry weight (Donaldson et al., 2006). In aspen, high CT concentrations have been associated with reduced insect richness and density (Barker et al., 2018) and larval growth of *Chrysomela scripta* F. (Donaldson & Lindroth, 2004). Studies on other *Populus* species have also reported a negative correlation between CT abundance and beaver browsing (Bailey et al., 2004). This, however, is not universal, as some aspen-adapted species are positively associated with CTs. A study on, for example, the gypsy moth (*Lymantria dispar;* SV lövskogsnunna) revealed a positive correlation between CTs and improved performance (developmental time and pupal weight) (Hemming & Lindroth, 1995) and increased food consumption (Osier et al., 2000). Other ecological effects of CTs have been observed in

litter and roots, influencing microbial composition and nitrogen availability (Barbehenn & Constabel, 2011).

Diversity and evidence of bioactivity in salicinoid phenolic glycosides

Structure and occurrence

SPGs are common in Salicaceae members, such as *Populus*, *Salix* (Boeckler et al., 2011), and *Homalium* (Ishikawa et al., 2004). They are also common in other plants, such as meadowsweet (*Filipendula ulmaria*; SV älggräs eller älgört) (Blazics et al., 2010). SPGs have been referred to as both a constitutive defence (i.e., always present) (Philippe & Bohlmann, 2007) and an intermediate-delayed induced resistance, with increased levels observed eight weeks after damage (Stevens & Lindroth, 2005). However, SPG induction in willow has been shown to be highly dependent on the developmental stage of the sampled tissue, individual SPG compounds measured, and genotype (Fields & Orians, 2006; Ruuhola et al., 2001).

The basic structure of SPGs is composed of a beta-D-glucopyranose bonded to a salicyl alcohol through an ether linkage. In its most basic form, this structure is called salicin, and phenolic glycosides are often referred to as salicinoids (Figure 1). There are more than 20 structures, all with a salicin core and the addition of moieties (e.g., acetyl, cinnamoyl, coumaroyl, hydroxy-6-oxo-2-cyclohexene (HCH), and benzoyl) (Boeckler et al., 2011; Abreu et al., 2020; Abreu et al., 2011; Keefover-Ring et al., 2014; Paper II-III). As these defences are mainly composed of carbon, their biosynthesis is thought to incur low energy investment (Ruuhola & Julkunen-Tiitto, 2003).

The occurrence of SPGs varies among species and tissues (Boeckler et al., 2011). For example, Palo (1984) observed a generally higher phenolic glycoside content in bark than in leaves in both *Salix* and *Populus* species, while younger leaves have been shown to have a higher SPG concentration than older leaves (Gould et al., 2007; Massad et al., 2014; Ruuhola & Julkunen-Tiitto, 2000; Bingman & Hart, 1993; Coleman, 1986; Yepes-Vivas et al., 2025; Paper II). SPGs have been found in other plant parts as well, including buds (Julkunen-Tiitto & Virjamo, 2017), roots (Massad et al., 2014), and phloem sap (Gould et al., 2007). However, there are a few *Salix* species with low or nondetectable levels of SPGs in leaves, twigs (e.g., *S. myrtilloides, S. triandra,* and *S. viminatis*), or roots (e.g., *S. lapponum*, *S. myrsinifolia,* and *S. pentandra*) (Julkunen-Tiitto & Virjamo, 2017; Palo, 1984). The concentration of SPGs is also highly variable, even within the same species, ranging as much as tenfold among genetically distinct aspens (Hemming & Lindroth, 1995). SPG levels in *Populus* leaves are typically 1-8% (Fellenberg et al., 2020), but levels as high as 30% have been reported (Donaldson et al., 2006). Many studies point to a genetic basis for the amount and composition of SPGs in Salicaceae (Hwang & Lindroth, 1997; Osier & Lindroth, 2006; Donaldson & Lindroth, 2007; Keefover-Ring et al., 2014; Barker et al., 2018; Paper III). In a study by Osier and Lindroth (2001), genotype explained 93% of the SPG variation in *P. tremuloides*. Although genotype strongly influence SPG production, it can also be affected by nutrient availability (Donaldson & Lindroth, 2007), age (Donaldson et al., 2006), and season (Wimp et al., 2007; Massad et al., 2014). As such, SPG abundance and composition appear to be influenced by both genotype and environment. Adding to the complexity, recent studies also suggest trade-offs between SPGs and growth, and between SPGs and CTs (Cole et al., 2021). A biological function for the majority of SPGs remains to be defined.

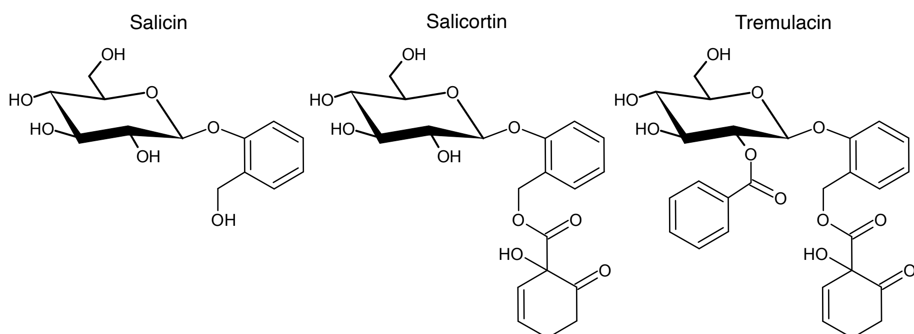

Figure 1. Structural overview of three salicinoid phenolic glycosides (SPGs) with documented defence activity. The figure was made using ACD/ChemSketch Freeware.

Evidence for salicinoid phenolic glycosides as defense compounds and their biosynthesis

Generalists

Studies of SPGs and herbivores have suggested that these compounds function as a defence resulting from their toxicity and bitterness (Mason et al., 2021). However, a lack of SPG mutants has forced researchers to rely on correlative studies, such as feeding experiments with pure extracts or using plants with naturally varying levels of specialised metabolites. In a feeding trial with free-ranging mountain hares (*Lepus timidus*; SV skogshare), the hares preferred feeding on willow with low SPG concentrations (Tahvanainen et al., 1985). In an additional experiment, the hares were offered oat grains supplemented with purified SPG fractions isolated from bark, indicating that SPGs have a negative effect on hare preference (Tahvanainen et al., 1985). Similar patterns have been observed in feeding trials in which elk were offered bundles of aspen shoots from different clones with varying compositions of SPGs. The results indicated that SPGs, such as salicortin and tremulacin (Figure 1), may reduce elk consumption (Wooley et al., 2008). There is some evidence suggesting that SPGs may act as a selective force. In a survey of elk (*Cervus elaphus*; SV kronhjort) browsing on *P. tremuloides,* an

induction of the SPGs salicortin and tremulacin was observed in response to browsing (Bailey et al., 2007). Moreover, trees with higher tremulacin concentrations in their terminal twigs experienced lower mortality than those with lower levels.

SPGs have been shown to negatively affect several life-history traits of gypsy moth (*L. dispar*) including developmental time, fecundity, pupal weight (Osier et al., 2000), growth (Boeckler et al., 2016), and growth rate (Hwang & Lindroth, 1997). Another species, the southern armyworm (*Spodoptera eridania*), showed reduced growth rate, digestibility, and consumption rate when fed salicortin and tremulacin. Examination of the midgut revealed that these metabolites act as gut toxins, causing degenerative lesions (Lindroth & Peterson, 1988). The toxicity of salicortin has been related to 6-HCH. This moiety is thought to be converted to the toxic compound catechol under basic conditions (e.g., the gut of many herbivores) and phenol under acidic conditions (e.g., the mammalian gut) (Clausen et al., 1989; Martinsen et al., 1998). The gypsy moth (*L. dispar*) has evolved mechanisms to detoxify SPG derivatives by conjugation with glucose, amino acid, or phosphate (Boeckler et al., 2016). In a follow-up experiment, the first two strategies were also identified in ten additional lepidopteran species and one coleopteran species, of which only the rusty tussock moth (*O. antiqua*) could conjugate with phosphate, suggesting the evolution of both general and specific detoxification mechanisms. However, further studies on SPG detoxification and the underlying enzymes are still needed (Boeckler et al., 2016).

Most feeding experiments have focused on a few SPGs characteristic of *Populus* (such as salicortin and tremulacin). However, the diversity of SPGs is far greater, partly due to the addition of cinnamoyl and acetyl functional groups, and the effects of these modifications on SPG toxicity

and digestibility remain largely unstudied. In our own study (see Results and Discussion), which assessed feeding damage by *O. antiqua* on aspen individuals that naturally vary in their acetylated SPG abundance, we did not detect a significant difference in feeding damage among individuals, further highlighting the need to investigate the functional role of these compounds.

Generalist herbivores vary in their adaptation to SPGs. One study observed that forest tent caterpillars (*Malacosoma disstria*; SV skogstältlarv), which commonly feed on *P. tremuloides*, are better adapted to SPGs than gypsy moth (*L. dispar*), which only recently started to feed on *Populus* (Hemming & Lindroth, 1995). A similar pattern has been documented for two species of tiger swallowtail (*Papilio glaucus*; SV östlig tigersvalstjärt) (Lindroth et al., 1988). An in-depth investigation of gypsy moth showed that the gut microbiota changed when ingesting leaves with different SPG compositions (Mason et al., 2015) — presenting an opportunity to study how specialised metabolites can shape the evolution of symbioses (Woolbright et al., 2018), and serving as a compelling example of the extended phenotype (Dawkins, 1982). However, the level of detoxification varies between species, and in an experiment with the Asian long-horned beetle (*Anoplophora glabripennis;* SV asiatisk långhorning) feeding on red maple (*Acer rubrum;* SV rödlönn) twigs supplemented with SPGs extracts, minimal detoxification occurred (Mason et al., 2021). Furthermore, the beetle was affected by SPGs in a dose-dependent manner, with higher doses significantly reducing consumption and egg production (Mason et al., 2021). Heiska et al. (2007) reported a critical SPG concentration, below which there was no effect on vole (*Microtus* spp.; SV sork) feeding.

Gordon et al. (2022) conducted a feeding trial in which white-marked tussock moth (*Orgyia leucostigma*) larvae were exposed to a SPG mutant plant lacking a functional UGT71L1 gene, a glycosyl transferase (UGT) hypothesised to glycosylate a SPG precursor. The mutant exhibits a >90% reduction in tremulacin and salicortin levels compared with the control plants, and in a choice assay, the larvae preferred the mutant. However, because the mutant's overall metabolite profile and physiology were altered, the reduced larval feeding cannot be conclusively attributed to SPG content alone.

Although not intended as a SPG mutant, overexpressing the MYB134 tannin regulatory gene in *P. tremula x tremuloides* resulted in drastically increased levels of CT and about a 50% reduction in salicortin and tremulacin concentration (Boeckler et al., 2014). When performing a two-choice feeding assay, two generalist species (*L. dispar* and *M. disstra*) preferred feeding on leaf discs from the MYB134 overexpressing mutant compared to the control. The authors also reported higher pupal weight and higher survival and pupation rates for *L. dispar* feeding on the MYB134 mutant. The authors suggest that reduced leaf palatability is most likely due to reduced SPG levels rather than CTs (Boeckler et al., 2014).

No clear effect of SPGs at the community level has been established to date. This includes surveys of endophyte (Albrectsen et al., 2018) and arthropod communities in *P. tremula* (Robinson et al., 2012), and insect community in *P. tremuloides* (Barker et al., 2018; Wimp et al., 2007). Neither of these studies excludes the possibility of a community effect, which may be driven by seasonal variation in SPGs (Wimp et al., 2007; Barker et al., 2018).

Specialist herbivores

There are examples of specialist species that are highly adapted to feed on Salicaceae and can utilise SPGs for their own benefit. The specialist sawfly (*Euura anerinae* L.) can, for example, differentiate between willow trees based on SPG concentration (Kolehmainen et al., 1994). Notably, a study by Nyman and Julkunen-Tiitto (2000) suggests that gall-inducing sawflies manipulate the synthesis of phenolic compounds in willows to reduce the level of defence compounds in the gall interior.

Among the specialist beetles, we find *Chrysomeloidea* beetles, such as *Phratora vitellinae* L. (SV bronsvidebagge) and *Chrysomela populi* (SV aspglansbagge). These beetles are known for incorporating salicin in their eggs, providing efficient protection against predators. The eggs of, for example, *Chrysomela* contain ten times as much salicin as those of *Ph. Vitellinae* – at levels high enough to kill an ant (Pasteels et al., 1986). Moreover, a major fraction of the larvae's secretions contains salicylaldehyde (derived from salicin), a potent deterrent against predators (Pasteels et al., 1983). In a study by Martinsen et al. (1998), they observed that *Chrysomela* prefers to feed on cottonwood tissue rich in SPGs. In turn, the adult beetles were larger, developed faster and were better protected against ants. However, this is not efficient against all predators, as some use the secretions to locate the larvae (Köpf et al., 1997). As such, specialist herbivores have not only evolved mechanisms to negate the toxic or adverse effects of SPGs but also ways to use these compounds for their own defence.

Biosynthesis

The SPG biosynthesis pathway remains largely unresolved, although many discoveries have provided insight into its origin. Zenk (1967) tracked the incorporation of isotope-labelled precursors (cinnamic acid,

benzoic acid, salicyl alcohol, and benzyl alcohol) in *Salix purpurea* L. (SV rödvide) to show that salicin is derived from the phenylpropanoid pathway. Another study confirmed a connection to the phenylpropanoid pathway by inhibiting phenylalanine ammonia-lyase (PAL) activity, showing a clear reduction of SPGs in the shoot tips of *Salix pentandra* (SV jolster) (Ruuhola & Julkunen-Tiitto, 2003). Later studies demonstrated that both the HCH and the salicyl moiety of salicortin are derived from cinnamic acid, with benzoates acting as intermediates, by exposing young leaf discs of *Populus nigra* (SV svartpoppel) to a broad range of isotope-labelled phenylpropanoid compounds (Babst et al., 2010). They also observed that some precursors (salicylaldehyde and salicyl alcohol) were only incorporated into salicin. Given these results, the authors proposed that SPG biosynthesis should be considered as a metabolic grid rather than a strictly hierarchical pathway.

Although isotope-labelled carbon provided insight into the connection between SPGs and the phenylpropanoid pathway, the enzymes active in the salicinoid pathway remained unknown. Chedgy et al. (2015) were the first to identify two enzymes in *Populus trichocarpa* (SV jättepoppel) in the previously proposed phenylpropanoid pathway. Using a mutant line overexpressing the MYB134 tannin regulator, which exhibited reduced SPG levels (Mellway et al., 2009), Chedgy et al. targeted genes downregulated in the mutant. Additionally, the authors considered only genes belonging to the BAHD-type acyl transferase family, which have frequently been implicated in specialised metabolite biosynthesis. This approach led to the identification of two candidate acyltransferases: salicyl alcohol O-benzoyltransferase (SABT) and benzyl alcohol O-benzoyltransferase (BEBT). *In vitro* experiments showed that the enzyme SABT catalyses the reaction between salicyl alcohol and benzoyl-CoA to form salicyl benzoate. The other enzyme, BEBT, catalyses the formation

of benzyl benzoate from benzyl alcohol and benzoyl-CoA. Both benzyl benzoate and salicyl benzoate are hypothesised to be required for the formation of salicortin and tremualcin. However, both enzymes were promiscuous, accepting several substrates to produce acetates and benzoates.

When studying gene co-expression among the genes encoding these acyltransferases, Fellenberg et al. (2020) discovered two new candidate enzymes. These enzymes were both UDP-glucose-dependent glycosyltransferases, capable of glycosylating salicyl benzoate, salicylaldehyde, and 2-hydroxycinnamic acid, with salicyl benzoate as the preferred substrate *in vitro*. Although the knockout mutants of one of the genes (UGT78M1) in *P. tremula* x *Populus alba* hairy root culture were futile due to errors in construct design, the knockout of the *UGT71L1* gene was successful. Inactivation of the *UGT71L1* gene resulted in a loss of salicortin, tremulacin, and tremuloidin production and a reduction of salicin production, again indicating multiple routes for salicin biosynthesis (Figure 2).

Gordon et al. (2022) investigated the role of UGT71L1 in greenhouse-grown plants. They found that its absence led to a 90% reduction in the concentrations of tremulacin and salicortin compared with control plants. Notably, significant effects were also observed on non-SPG compounds, with over 3,000 metabolite features showing altered abundance, including flavonoids, apigenin, kaempferol, and benzoic acid-derived metabolites, indicating systemic effects. The physiology of UGT71L1-deficient plants was also severely affected, exhibiting reduced growth, shorter internodes, and altered leaf size and shape. The authors suggest that these physiological changes stem from elevated levels of plant hormones, specifically salicylic acid and jasmonic acid. In the absence of

UGT71L1, the substrate salicyl salicylate may accumulate and become unstable, resulting in the production of salicylic acid as a by-product (Gordon et al., 2022) – a metabolite considered unlikely to serve as an SPG precursor (Zenk, 1967).

The UGT71L1 mutant is not the first phenylpropanoid mutant with non-target effects. For example, Coleman et al. (2008) generated a hybrid poplar mutant with suppressed expression of *p*-coumaroyl-CoA 3'-hydroxylase (C3'H), a gene in lignin biosynthesis. This mutant showed a significant reduction in lignin content accompanied by a dramatic accumulation of phenolic glycosides. Similarly, Boeckler et al. (2014) constructed a MYB134 tannin regulatory mutant with markedly increased CT and decreased SPG levels. The authors suggest that these unforeseen changes in metabolite profiles highlight the inherent plasticity of the phenylpropanoid pathway, making it especially challenging to generate mutants without non-target effects.

Lackus et al. (2020) reported the existence of two sulphated SPGs (salirepin-7-sulfate and salicin-7-sulfate) in *P. tricocharpa,* where they were found to be highly abundant in leaves. To identify the enzyme responsible for this sulfation, the authors searched for Arabidopsis sulfotransferase (*AtSOT12*) homologs that showed increased expression in *P. trichocarpa* leaves. Two genes met these criteria, but only one (*SOT1*) produced sulphated SPGs in a heterologous expression experiment. Consistent with this finding, *Populus* X *canescens* mutants with suppressed *SOT1* expression exhibited reduced levels of sulphated SPGs compared with controls. To evaluate possible plant defence-related functions of these sulphated SPGs, the authors performed a feeding preference assay with *L. dispar* larvae, offering them leaf discs from *SOT1*-suppressed and wild-type plants. No significant feeding preference was

detected, leading the authors to postulate that sulphated SPGs may serve as sulphur storage compounds.

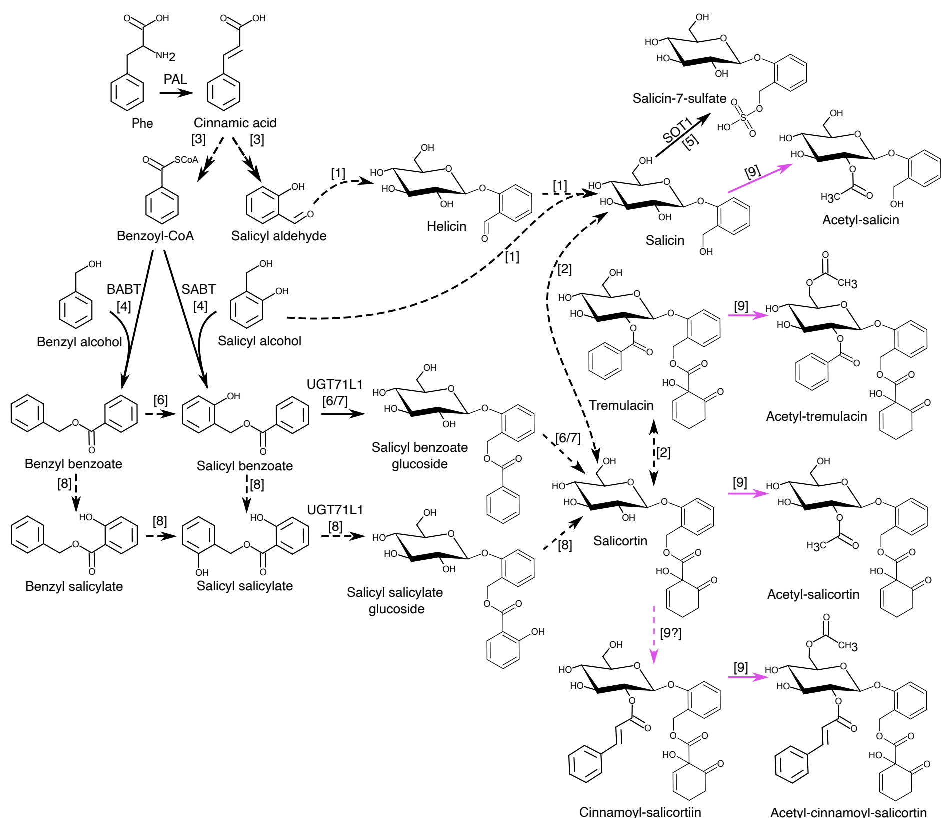

Figure 2. An overview of salicinoid phenolic glycoside (SPG) biosynthesis showing both confirmed (solid arrows) and hypothetical pathways (black dashed arrows) with their associated genes and/or references. Purple solid arrows denote proposed biosynthetic reactions based on the heterologous expression in *Escherichia coli* and the overexpression experiment in *Populus tremula* (Paper III). Purple dashed arrows denote hypothetical biosynthetic reactions in Paper III. [1] Zenk, (1967); [2] Ruuhola & Julkunen-Tiitto (2003); [3] Babst et al. (2010); [4] Chedgy et al. (2015); [5] Lackus et al. (2020); [6] Fellenberg et al. (2020); [7] Kulasekaran et al., (2021); [8] Gordon et al. (2022); [9] Paper III (Gene ID: Potra2n4c9093). The figure was made using ACD/ChemSketch Freeware.

As discussed above, the plasticity of the phenylpropanoid pathway can lead to non-target effects in mutants, making it difficult to assign precise

gene functions. However, a more nuanced view of specialised metabolism is to regard it as a complex trait, in which the biosynthesis of a metabolite depends on multistep metabolic pathways catalysed by numerous enzymes (Weng et al., 2014), some of which potentially exhibit catalytic promiscuity and influence multiple traits (Weng et al., 2012). In addition, the expression of the enzyme might be modulated by complex transcriptional regulation, for example, by MYB (Liu et al, 2015) and WRKY (Li et al., 2025) transcription factors, which coordinate pathway activity. As such, studying plant specialised metabolism as a multi-level system – integrating, among others, genetic variation, transcriptomic regulation, and protein and metabolite abundance – provides a more complete mechanistic picture of how biosynthetic pathways are controlled. This also means that perturbations at one level can propagate through the system and influence other levels. Natural genetic variation within a population provides a source of such perturbations, enabling the dissection of how changes in genotype affect the transcriptome and other phenotypic layers.

Such multi-level studies, or systems biology approaches (introduced below), should be particularly suitable for studying SPGs, since they have previously been shown to be highly heritable (Hwang & Lindroth, 1997; Osier & Lindroth, 2006; Donaldson & Lindroth, 2007; Keefover-Ring et al., 2014; Barker et al., 2018; Osier & Lindroth, 2001), suggesting a genetic basis for trait variation. Furthermore, extensive SPG diversity has been reported in the SwAsp collection, which has been grouped into chemical phenotypes (or chemotypes) based on the abundance of SPGs adorned with cinnamoyl, acetyl, or benzoyl (Keefover-Ring et al., 2014; Abreu et al., 2011). The SwAsp collection comprises 112 genetically unrelated individuals (Wang et al., 2018), sampled in 2004 from twelve locations spanning from Arjeplog in northern Sweden to Ronneby in the south

(Luquez et al., 2008). Moreover, whole-genome resequencing of the SwAsp individuals is available – first reported by Wang et al. (2018) and has since been refined through alignment to the latest *P. tremula* reference genome (Paper I) – making this population a suitable system for integrative analyses.

In this thesis, candidate genes associated with leaf physiognomy (Paper I) and SPGs (Paper III) are identified by exploiting the natural variation within the SwAsp collection. The following chapter describes the concept of systems genetics, key discoveries and technological advancements that have contributed to its emergence.

Systems biology

For much of the 20th century, the dominant scientific philosophy was reductionism. According to reductionism, the world can be viewed as a clockwork mechanism; by breaking it down into smaller components, analysing each individually, and reassembling the parts, we can ultimately understand how it works (Beresford, 2010). This philosophy is applicable to organisms as well, which can be viewed as collections of smaller units, including molecules and atoms, that can be further reduced and explained by physics and chemistry. However, for most traits, it is rarely sufficient to list individual components to go from phenotype to genotype – we also need to understand how the components dynamically interact (Kitano, 2002). Since genes and gene products do not act in isolation, but rather form an intricate network of pathways, the role of individual genes becomes apparent only when considered within this higher level of organisation. Multiple genes in the genome may also perform redundant roles, such that the loss of one gene would not be apparent as long as the others are functional. Knockout experiments in these scenarios often fail

to reveal gene function, yielding either null or unforeseen phenotypes. As such, alternative approaches acknowledging the dynamic interactions across biological levels are needed (Mazzocchi, 2012).

Systems biology represents one such approach, being a highly interdisciplinary field that builds on biology, chemistry, physics, mathematics, and computer science (Pazhamala et al., 2021; Voit, 2022). Although the exact definition varies among researchers, systems biology generally involves integrating multiple types of large-scale data, typically created using high-throughput approaches (i.e., 'omics data). The most common data types include genomics, transcriptomics, proteomics and metabolomics, which are used to construct models of biological systems to predict how these systems will respond to perturbations (Veenstra, 2021; Pazhamala et al., 2021; Civelek & Lusis, 2014). The main driver enabling systems biology has been technological advancements, particularly the completion of the Human Genome Project (HGP) in 2003. With enhanced computational power and the availability of high-throughput methods, we can now analyse the entire genome, epigenome, transcriptome, proteome, metabolome, and microbiome (Delker & Mann, 2017; Voit, 2022; Veenstra, 2021).

The study of 'omics

Genomics and sequencing technologies

The field of genomics entails studying the structure, function, and expression of all genes in a genome (Horgan & Kenny, 2011). A vital step in genomics is therefore accessing the complete genome sequence, that is, determining the exact order of each and every nucleotide base in the DNA (Archibald, 2018). This would not have been possible without the elucidation of DNA's double-helix structure by Watson and Crick (Watson

& Crick, 1953) and the development of DNA sequencing techniques by Sanger and co-workers (Sanger et al., 1977).

The first organism to have its genome sequenced was the bacteriophage phiX174 (5,368 bp, base pairs) in 1977, utilising a method known as the "plus/minus" system (or Sanger/Coulson method). This event can be regarded as the "birth of first-generation" DNA sequencing, which could sequence about 50 base pairs in a single experiment (Heather & Chain, 2016; Giani et al., 2020). The same year, sequencing technology took another leap with Sanger's "chain-termination" technology. This improved method involved synthesising DNA strands using a modified version of radiolabelled nucleotides, called dideoxynucleotides (ddNTP) (Heather & Chain, 2016). With Sanger sequencing, it became possible to sequence about 400 bases per day (Giani et al., 2020).

Since 1977, three generations of sequencing technologies have entered the market (Ambardar et al., 2016). The later generations enable parallelisation (i.e., sequencing millions of DNA fragments at the same time), thereby reducing both time and cost for sequencing. However, these technologies cannot sequence an entire genome in one go and rely on sequencing fragmented DNA strings (the maximum reported length being more than 4 Mbp) (Herbert et al., 2025). Assembling a genome is very much like solving a puzzle – requiring matching millions of DNA fragments (called reads) to one another *in silico* to identify overlapping sequences and using these overlaps to recreate the original sequence. Genome assembly thus represents a major analytical challenge, especially for genomes with high frequencies of repetitive DNA sequences. The latest (3rd) generation of sequencing technology, therefore, aims to optimise read length, sequence quality, and parallelisation (Archibald, 2018), and

as of 2024, the genomes of 1,482 plant species had been sequenced (Bernal-Gallardo & de Folter, 2024).

Illumina (second-generation sequencing)

Illumina belongs to the second-generation sequencing technologies and is the most popular sequencing platform, producing most of the next-generation sequencing (NGS) data today. Illumina uses "sequencing-by-synthesis", exploiting nature's way of replicating DNA using a DNA polymerase. However, unlike normal DNA replication, the DNA polymerase is given modified nucleotides that carry a fluorescent label and a reversible terminator. Each nucleotide is labelled with a nucleotide-specific fluorescent label (dATP, dCTP, dGTP, and dTTP), allowing it to be distinguished from the others. Thus, as DNA polymerase extends the DNA strand, a fluorescent light specific for each nucleotide base is emitted. The reverse terminator prevents DNA polymerase from continuing to replicate the DNA strand, allowing optical detection of the fluorescent light. The terminator and fluorescent label are attached to the nucleotide with a single chemical bond, allowing the DNA polymerase to continue extending the DNA strand once the light is emitted and the bond is broken, releasing the terminator (Wang, 2016).

To make the fluorescent signal detectable, each DNA fragment needs to be amplified before sequencing. Amplification takes time and can introduce biases for regions with high guanine-cytosine content (GC%) due to inefficient amplification. The error rate of Illumina is, however, generally low (<1%), and the throughput is high. One of the major disadvantages of Illumina is the read length (typically 150 bp) (Espinosa et al., 2024), since short reads make genome assembly challenging. Short reads are especially problematic for mapping repeated sequences, often leading to gaps and mismapping (Scarano et al., 2024; Wang, 2016).

However, the very high number of sequencing reads produced by Illumina (billions per machine run) offers a huge dynamic range, which is highly advantageous for the application of the technology to quantifying gene expression by RNA sequencing (RNA-Seq).

Pacific Biosciences (third-generation sequencing)

Pacific Biosciences (PacBio) uses third-generation sequencing technology that is sensitive enough to sequence a single DNA molecule, eliminating the need to amplify the DNA. PacBio relies on sequencing-by-synthesis, but uses flow cells with tiny wells, each containing a single DNA polymerase attached to the bottom of the well. Since DNA polymerase is stationary, the camera can detect the signal from the bottom of the well, allowing it to register a single nucleotide (Archibald, 2018). In contrast to Illumina, the nucleotides used for replication lack a terminator group. Instead, the fluorescent label is attached to the end phosphate group, which is released upon incorporation into the DNA strand, thus allowing for real-time sequencing (Wang, 2016). Compared to second-generation sequencing, PacBio technology is faster and can generate long reads, with maximum read lengths exceeding 250 Kbp and 99.9% accuracy in PacBio high-fidelity (HiFi) sequencing (Wang et al., 2025).

Genome re-sequencing and genetic variants

Once the genome of a species is sequenced and assembled, it can be used as a reference to map future read data generated from other individuals of the same species. This can be done either by re-sequencing the entire genome and mapping the sequence reads to the reference genome, or by targeting specific segments of the genome, such as the exons (Archibald, 2018). Whole-genome re-sequencing allows researchers to identify sequence variations at every nucleotide position in the genome, also known as single-nucleotide polymorphisms (SNPs). To some extent, other

types of genomic variations can also be detected, such as large structural rearrangements, copy number variations (CNVs), and small insertions and deletions (indels) (Archibald, 2018; Martin et al., 2018). Identification of the set of genetic variants (or alleles) present within a population is essential for understanding the genetic basis of trait variation. These DNA variants can serve as genetic markers, helping to pinpoint the genomic locations associated with trait variation in a population (Jiang, 2017).

Transcriptomics

Just as crucial as identifying which genes are present in a genome is knowing when they are active (i.e., turned on or off). Gene activity often varies both spatially (in different locations) and temporally (over time), leading to cellular and tissue differentiation. Indeed, all cells of an individual contain the same genome, and thus the same set of genes; therefore, the diversity of cell types can only be achieved by the differential expression of this set of genes. Thus, by knowing where and when a gene is active, we can potentially gain insight into the function of these genes (Archibald, 2018). Transcriptomics entails studying all RNA transcripts within a cell or group of cells (Alpay Savasan et al., 2021), a term coined by Auffray in 1996 (McGettigan, 2013).

There are multiple types of RNAs, which can be classified into protein-coding RNAs (messenger RNAs, mRNAs) and non-coding RNAs (ncRNAs). The latter includes ribosomal RNA (rRNA), transfer RNA (tRNA), microRNA (miRNA), long non-coding RNA (lncRNA), and others. The ncRNAs can serve a regulatory or housekeeping role (Dong & Chen, 2013; Park & Kim, 2023; Santosh et al., 2015), constituting a significant portion of the transcribed RNA. In humans, only 2% of the

genome is transcribed to mRNA, whereas 98% is transcribed into ncRNAs (Vicentini et al., 2019)

RNA sequencing techniques actually preceded DNA sequencing, mainly because it is easier to obtain higher RNA yields than DNA yields (Archibald, 2018). The first published sequence of a nucleic acid was in 1965, when Holley and co-workers sequenced yeast alanine tRNA, which required about 140 kg of yeast (Holley et al., 1965; Holley, 1968). The Holley sequencing approach involved fragmenting RNA with enzymes and isolating each fragment by ion-exchange chromatography (Holley, 1968), requiring substantial laboratory effort.

Today, next-generation sequencing technologies are widely used for transcriptome sequencing, including Illumina, PacBio, and others. In contrast to Holley sequencing, RNA must first be reverse transcribed into complementary DNA (cDNA) (Deshpande et al., 2023). The RNA selected for cDNA synthesis and cDNA library may differ depending on the RNA targeted subclass (mRNA, miRNA, or total RNA), but the sequencing method for each of these classes is identical (Guo et al., 2015). Once the reads are sequenced, they are mapped to the reference genome to identify the genomic region from which they were transcribed, with the number of reads in the data reflecting the transcript or gene activity (Wang, 2016).

Standard downstream analyses of RNA-Seq data include normalisation and differential gene expression (DGE) analysis. Normalisation of RNA-Seq data is crucial to control for factors that influence read counts across samples, such as differences in sequencing depth (Evans et al., 2018). After normalisation, DGE analysis is typically performed to test for differences in expression level (genes, transcripts, and/or transcripts/exons) between groups, for example, differences between male and female, tissues, treated and untreated (Deshpande et al., 2023;

Upton et al., 2023). Identifying genes that are up- or downregulated across test groups provides valuable insight into regulatory mechanisms and helps identify key genes involved in biological processes (Upton et al., 2023).

Metabolomics

The metabolome encompasses all metabolites within a biological system (e.g., a cell or tissue), where a single metabolite can be defined as an organic molecule synthesised by enzymes (Tebani et al., 2018). Among these molecules are amino acids, carbohydrates, vitamins, organic acids, lipids, and many other compounds (Ovbude et al., 2024). As part of the "omic" technologies, metabolomics entails analysing the complete biochemical profile of a sample (Tebani et al., 2018), including the identification and quantification of molecules with molecular masses up to approximately 1,500 Da (Idle & Gonzalez, 2007; Qiu et al., 2023). Unlike other 'omic levels, such as the genome, transcriptome and proteome, metabolites exhibit greater chemical diversity and serve as a phenotypic fingerprint, influenced by biological and environmental stimuli (Schrimpe-Rutledge et al., 2016; Aderemi et al., 2021). As such, metabolomic analyses gives a snapshot of metabolite dynamics and offer valuable insights into the effects of multiple perturbations within a biological system, including genetic alterations and environmental influences (Emwas et al., 2021). In contrast to transcripts and proteins, which are sequence-based, metabolites lack a direct link to the genome sequence, posing unique analytical challenges.

There are currently two main approaches to metabolomic profiling: targeted and untargeted. Targeted metabolomics typically focuses on a small number of known metabolites, usually ranging from tens to hundreds, that are specifically targeted for quantification. In contrast,

untargeted metabolomics aims to identify all measurable metabolites in a sample, whether they are known (i.e., can be assigned a meaningful identifier) or not (Schrimpe-Rutledge et al., 2016).

Chromatographic analysis and mass spectrometry techniques

There are several analytical platforms for metabolomics, including liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS). LC-MS can identify and quantify a large set of metabolites with high sensitivity and precision, including amino acids, carbohydrates, sugars, fatty acids, and nucleotides. In contrast, GC-MS is particularly suited for detecting volatile metabolites. Since LC-MS can detect a broader range of metabolites, it is more commonly used than GC-MS (Doğan, 2024). However, one single platform is not sufficient to obtain complete coverage (Perez de Souza et al., 2021).

As the name implies, LC-MS is a liquid chromatograph coupled with a mass spectrometer (Figure 3). The liquid chromatograph (LC) physically separates individual metabolites from the rest of the sample. Specifically, the sample is injected into a mobile phase, which is transported into a column via a high-pressure pump (referred to as HPLC). The column is packed with a stationary phase (e.g., C18 alkyl group bound to silica) that will interact with the metabolites in the sample. The stationary phase will separate metabolites based on characteristics such as size, affinity, and electrostatic forces (Rusli et al., 2022). The metabolite will interact with both the stationary and mobile phase, travelling through the stationary phase in response to sorption and desorption. Most LC separations are conducted with reversed-phase LC, in which the stationary phase is nonpolar while the mobile phase is polar (Ardrey, 2003).

The time it takes for a metabolite to elute from the column is called the retention time, which varies with column length and the mobile phase flow rate. The retention time can be used for metabolite identification by comparing it with that of a reference material analysed under the same experimental conditions. However, given the vast number of metabolites, identification based on retention time alone is generally insufficient. This is where the mass spectrometer comes in. Mass spectra are often specific enough for identification, providing molecular weight and structural information. The mass spectrometer can also quantify the metabolite with high accuracy (Ardrey, 2003).

The mass spectrometer consists of three parts: an ionisation source, a mass analyser, and a detector (Figure 3). The role of the ionisation source is to charge the metabolite. There are many techniques for ionising a metabolite, with electrospray ionisation (ESI) being the most popular for biological molecules. With ESI, the liquid sample is nebulised into charged droplets that evaporate when heated and dried with nitrogen (Pitt, 2009). The ionised metabolite travels into the mass analyser, which sorts ions by their mass-to-charge ratio (m/z) and eventually hits a detector that quantifies the number of ions. The quadrupole mass analyser is most commonly used for metabolite sorting due to a broad mass range coverage and ease of use (Parasuraman et al., 2014). The quadrupole consists of four parallel metallic rods that apply a voltage, allowing ions to reach the detector. By varying the voltage, ions with different m/z values can be scanned, yielding a mass spectrum. For targeted metabolomic studies, specific m/z values can be monitored, whereas the remaining ions collide with the rods (Pitt, 2009; Parasuraman et al., 2014). When ions reach the detector, a mass spectrum is generated, plotting the number of ions as a function of the m/z ratio (Ardrey, 2003).

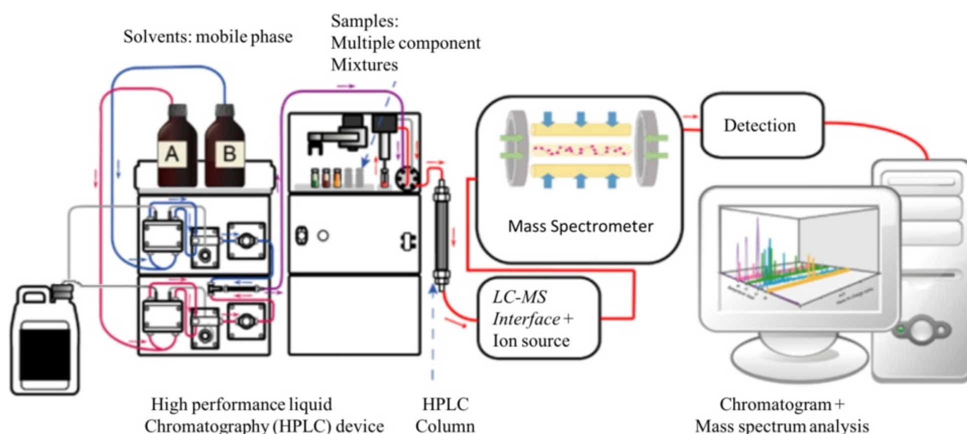

Figure 3. A schematic overview of liquid chromatography-mass spectrometry.
Source: Wikimedia Commons
[https://commons.wikimedia.org/wiki/File:Liquid_Chromatography_Mass_Spectrometer.png].

Tandem mass spectrometry (MS/MS)

Registering the molecular weight of a metabolite does not convey structural information. For this, tandem mass spectrometry (MS/MS) is used, an umbrella term for the coupling of two or more mass analysers (Pitt, 2009). Triple quadrupoles (QqQ) are commonly used, which is the coupling of three quadrupoles. The second quadrupole will, however, be used as a collision cell containing an inert gas, which will fragment ions (called product ions) before they travel to the third quadrupole, where a mass spectrum of the product ions is generated (Ardrey, 2003; Pitt, 2009). Other tandem mass spectrometry includes Quadrupole Time-of-Flight (Q-TOF) and Orbitrap, which are commonly used for untargeted metabolomics studies (Doğan, 2024).

Capturing the interplay between 'omics

Linkage analysis

Familial aggregation (similarity among relatives) may indicate a genetic contribution to a specific trait. As such, comparing the genetic makeup of family members can help identify causal variants (Bailey-Wilson & Wilson, 2011). These family-based studies are known as linkage analyses, which track the co-inheritance of a variant and a trait within pedigrees (March, 1999).

Linkage analyses involve genotyping family members to identify sequence variants that serve as genetic markers with a known genomic location (Bailey-Wilson & Wilson, 2011). Markers located close to a causal variant tend to be inherited together more frequently than expected by chance because of the low recombination rate at these positions. This non-random co-inheritance indicates that the genetic marker and trait locus (i.e., genomic region influencing the trait) are "linked", providing an approximate genomic location of the causal variant (Pulst, 1999).

The foundation of linkage was laid in 1905, when Bateson and co-workers observed that certain sweet pea (SV luktärt) traits violated Mendel's law of independent assortment by appearing "coupled" (Bateson et al., 1905). A few years later, Thomas H. Morgan made several crucial discoveries by studying inheritance patterns in fruit flies (*Drosophila melanogaster*; SV bananfluga), becoming the first to report that genes are carried on chromosomes (Morgan, 1910). When Morgan later studied multiple traits simultaneously, he realised that their inheritance patterns could only be explained by crossover events between homologous chromosomes (Morgan, 1911a). Some traits, however, remained strongly associated with one another. Morgan postulated that this was due to their physical

location on the chromosome, with genes located close together being less likely to be separated by a crossover event (Morgan, 1911b). One of Morgan's students, Alfred Sturtevant, advanced these discoveries by using linkage to estimate relative gene distances, thereby creating the first genetic map (Sturtevant, 1913).

Linkage analysis has been successful in uncovering the genetic basis of several human diseases, including Huntington's disease (Gusella et al., 1983), Alzheimer's disease (Corder et al., 1993), cystic fibrosis (Tsui et al., 1986), and psychiatric disorders such as bipolar disorder (Stine et al., 1995) and schizophrenia (Schwab et al., 1998). With the ease of crossing inbred lines, linkage mapping is also widely used to study plant traits, with studies in cotton (Wang et al., 2006), soybean (Wang et al., 2019), maize (Khanal et al., 2015), and rice (Shanmugavadivel et al., 2013). It has also been an important approach for investigating model species, including Arabidopsis (Koornneef & Meinke, 2010), Drosophila (Zwick et al., 2000; Mackay, 2004), mouse (Dietrich et al., 1995) and Populus (Woolbright et al., 2018).

There are, however, several notable drawbacks to linkage analysis that have become evident over time. First, since linkage analyses are family-based, they are limited by the number of recombination events that occur within a pedigree, making it challenging to achieve fine-scale resolution of causal loci without generating unrealistically large family sizes (De Resende et al., 2014). For some organisms, performing controlled crosses can be difficult, time-consuming, or infeasible. Second, findings from linkage studies have often been challenging to reproduce, because the genetic variation within a single family does not capture the diversity of the broader population (Myles et al., 2009). Finally, linkage analyses have limited power to detect loci with modest influence (effect size), a

characteristic of most complex traits, which are typically influenced by many loci, each having a small effect (Hirschhorn & Daly, 2005). For these reasons, a shift from family-based to population-based studies has been highly beneficial for the study of complex traits. This transition became feasible with the adoption of SNPs as genetic markers, which are not only cheap to genotype but also represent the most abundant of polymorphisms in the genome, enabling much finer mapping resolution (De Resende et al., 2014).

Linkage disequilibrium

Linkage disequilibrium analysis (also known as association analysis) exploits the historical linkage between markers and causative trait loci in a natural population of unrelated individuals (March, 1999). This approach leverages the high number of recombination events that accumulate over many generations to reduce the number of markers still linked to the causal variant. In principle, when a new variant is introduced in a population, either through mutation or migration, it is initially associated with nearby genetic markers. Over time, recombination events break down these associations, leaving only markers in close physical proximity in linkage with the variant. Consequently, for a marker and a causal variant to remain linked in such populations, they must be located reasonably close to each other, thereby making it easier to identify the causal variant, which in some cases may even be the marker itself (De Resende et al., 2014).

A linkage disequilibrium analysis that genotypes markers across the entire genome is known as a genome-wide association study (GWAS) (Myles et al., 2009). SNPs are commonly used as markers, but other genomic features are also suitable. To pinpoint markers significantly associated with the trait, an association test is performed for each marker. In large

and highly heterozygous species, such as many forest trees, this may require performing millions of association tests (Jiang, 2017; Sul et al., 2018). Given these attributes, GWAS is generally more accessible and cost-efficient than linkage analysis and resolves many of the challenges encountered with linkage analysis. However, GWAS has limited power to detect rare variants, as they contribute little to the population-level signal. This limitation does not apply to linkage analyses (Myles et al., 2009), and therefore many studies combine both linkage analysis and GWAS (e.g., Xiao et al., 2019; Zhang et al., 2019).

Since the first GWAS publications in 2002 of myocardial infarction (Ozaki et al., 2002; Ikegawa, 2012), human-related GWAS have been continuously published, with over 7,000 GWAS publications available in the human GWAS catalogue database (https://www.ebi.ac.uk/gwas, 2025-04-01). GWAS has also been highly successful in identifying SNPs associated with plant traits. For example, Atwell et al. (2010) conducted a GWAS of over 100 traits in *A. thaliana*, identifying many variants with large effects. There are likewise many publications applying GWAS to crop traits, including flavour in tomato (Zhang et al., 2015), agronomic traits such as yield and quality in rice (Huang et al., 2010) and adaptive traits such as the timing of autumn bud set in *P. tremula* (Wang et al., 2018).

Despite the success of population-based association studies in dissecting the genetic basis of complex traits, GWAS has notable shortcomings. A common finding is that most heritability remains unexplained by significant associations, and the variants that are identified typically account for only a small fraction of the total heritable variation. This so-called “missing heritability” has been a heated topic. Potential factors include the low frequency of rare variants in the population, allelic heterogeneity, epigenetic effects, gene-by-environment interaction, gene-

by-gene interaction, and technical measurement errors (Manolio et al., 2009; Brachi et al., 2011; Civelek & Lusis, 2014; Ingvarsson & Street, 2011). Another limitation, relevant for both GWAS and linkage analyses, is that historically, many studies focused on identifying SNPs predicted to change protein sequences. Such an approach may be highly flawed, as most variants lie outside protein-coding genes, suggesting that regulatory mechanisms play a crucial role in trait variation (Ingvarsson & Street, 2011). Shifting the focus toward genetic variation that influences gene expression can therefore provide valuable insights into trait variation. Moreover, it offers an opportunity to "functionally characterise" individual SNPs located outside coding regions (van der Sijde et al., 2014).

Genetical genomics

Jansen and Nap (2001) were the first to propose treating genome-wide expression profiles as quantitative traits that could be mapped to the genome. They introduced the field "genetical genomics", which exploits genetic variation within a family to identify genomic regions influencing gene expression. When gene expression is mapped to the genome, the approach is commonly referred to as expression quantitative trait locus (eQTL) mapping. Mapping of other 'omic levels, such as metabolomics (mQTL) and proteomics (pQTL), is equally applicable and informative (Jansen et al., 2009). Unfortunately, the terminology can be confusing, as the term eQTL is sometimes used to describe GWAS of gene expression, while some studies prefer the term expression GWAS (eGWAS).

eQTL mapping provides important information by (I) identifying SNPs that influence gene expression among individuals and (II) providing insight into regulatory mechanisms based on the genomic location of the eQTL (Kadarmideen et al., 2006). Depending on their genomic location relative to the target gene, eQTLs are commonly classified as local or

distant. A local eQTL maps close to the target gene, whereas a distant eQTL maps elsewhere in the genome. These eQTLs can also act in *cis* or *trans*. If an eQTL only alters the expression of the allele it maps to, it is called a *cis*-eQTL and is therefore typically local, while a *trans*-eQTL influences the expression of both alleles, with most *trans*-eQTL being distant (Mähler et al., 2017). By examining the mapping locations and *cis*/*trans* actions, eQTL studies can provide important insight into the regulatory architecture of gene expression.

Several studies have used eQTL mapping to explore gene expression regulation in a variety of species, including yeast (Brem et al., 2002), humans (Morley et al., 2004), *Eucalyptus* (Kirst et al., 2005), *Populus* (Drost et al., 2010; Mähler et al., 2017) and *A. thaliana* (Keurentjes et al., 2007), highlighting the importance of both *cis*- and *trans*-regulation for expression variation. eQTL mapping has also provided insight into gene-by-gene (Brem et al., 2005; Becker et al., 2012) and gene-by-environment effects on gene expression variation (Romanoski et al., 2010; Smith & Kruglyak, 2008). In addition to transcriptomic variation, other ‘omic levels have been investigated using similar approaches, including metabolomics (mQTL) (Alseekh et al., 2020; Knoch et al., 2017) and proteomics (pQTL) (Rodziewicz et al., 2019) in plants.

Systems genetics

With the availability of linkage analysis, GWAS, and genetical genomics, studies combining multiple approaches to uncover trait variation have become increasingly common (e.g., Lisec et al., 2008; Xiao et al., 2013; Drost et al., 2015). Integrating mapping results from ‘omic and phenotypic traits offers several advantages. For example, after mapping a phenotypic trait to the genome, eQTL mapping can help identify which GWAS associations are likely causal through *cis*-eQTLs and can also reveal

potential downstream target genes via *trans*-eQTLs (van der Sijde et al., 2014). However, genetic variants that affect protein structure or activity are not detectable at the eQTL level (Jansen et al., 2009), highlighting the need to survey additional molecular levels. Therefore, a comprehensive view of factors influencing trait variation requires a systems genetics approach.

Systems genetics can be regarded as a subfield of systems biology in which the perturbations arise from naturally occurring genetic variation (i.e., mutations) within a population. The goal of systems genetics is to identify genes that regulate a trait and to detect significant interactions within and across biological levels that contribute to trait variation, thereby encompassing the central dogma of molecular biology (Figure 4). To connect traits with underlying biological levels, systems genetics identifies genomic regions that co-map to both phenotypic and molecular traits (Nadeau & Dudley, 2011; Voy, 2011; Civelek & Lusis, 2014). In this way, it renders a "clearer path" from genome to phenotype, offering mechanistic insight into how trait variation arises (Myburg et al., 2019). However, such relationships are only detectable if the molecular traits influence the final phenotype at the time of sampling, which may also vary among tissues. This highlights the importance of carefully designed studies (Ingvarsson et al., 2016). Study designs in systems genetics can vary in complexity and methodology depending on the number of traits and molecular traits included, but they generally combine GWAS or linkage analysis, genetical genomics, correlation, and network analyses (van der Sijde et al., 2014).

As previously described, GWAS or linkage analysis identify potential connections between the genome and phenotype through association or linkage, whereas genetical genomics establishes links between the genome and molecular traits. Genomic regions that map to both

phenotypic and molecular traits may indicate a causal relationship between them, making them strong candidates for further investigation (Civelek & Lusis, 2014; van der Sijde et al., 2014). Systems genetics, therefore, leverages the collective information from multiple analyses to identify potential candidates for trait variation.

Another key component of systems genetics analyses is correlation analysis, which can help establish potential relationships between molecular and phenotypic traits or different layers of the central dogma. A significant correlation between a molecular trait and a phenotype may suggest a causal relationship in which a molecular trait influences the phenotype. However, the reverse is also possible, with the phenotype affecting the molecular trait, or a confounding factor influencing both (Civelek & Lusis, 2014). Therefore, further validation, such as the experiments using knockout mutants, can help determine the direction and nature of the relationship (van der Sijde et al., 2014).

Network analyses can further enhance our understanding of molecular interactions and provide an effective way to visualise massive 'omics-scale datasets. Networks are typically classified based on the type of biological entity and interaction, such as gene co-expression networks, metabolic networks, and protein-protein interaction networks. For example, gene co-expression networks are constructed based on similarities in pairwise gene expression patterns. In systems genetics studies, this co-expression reflects similarity in expression among individuals within the study population, whereas classical systems biology typically measures variation within a reference genotype during development or in response to a specific stress. Expression similarity is quantified as a similarity score, calculated by comparing expression profiles for each gene pair. Gene pairs exceeding a predefined threshold are used to construct the network, with

genes represented as nodes and their interactions as edges. Following the "guilt-by-association" principle, genes with similar expression patterns are assumed to be involved in the same biological processes, forming clusters or modules within the network. These clusters offer an opportunity to identify genes involved in similar biological processes and to assign putative functions to unannotated genes based on the known functions of other genes within the same cluster (Serin et al., 2016). Network analyses can also help identify key regulators of gene expression (Rao & Dixon, 2019), such as "hub" genes that are highly connected to other genes in the network (Upton et al., 2023).

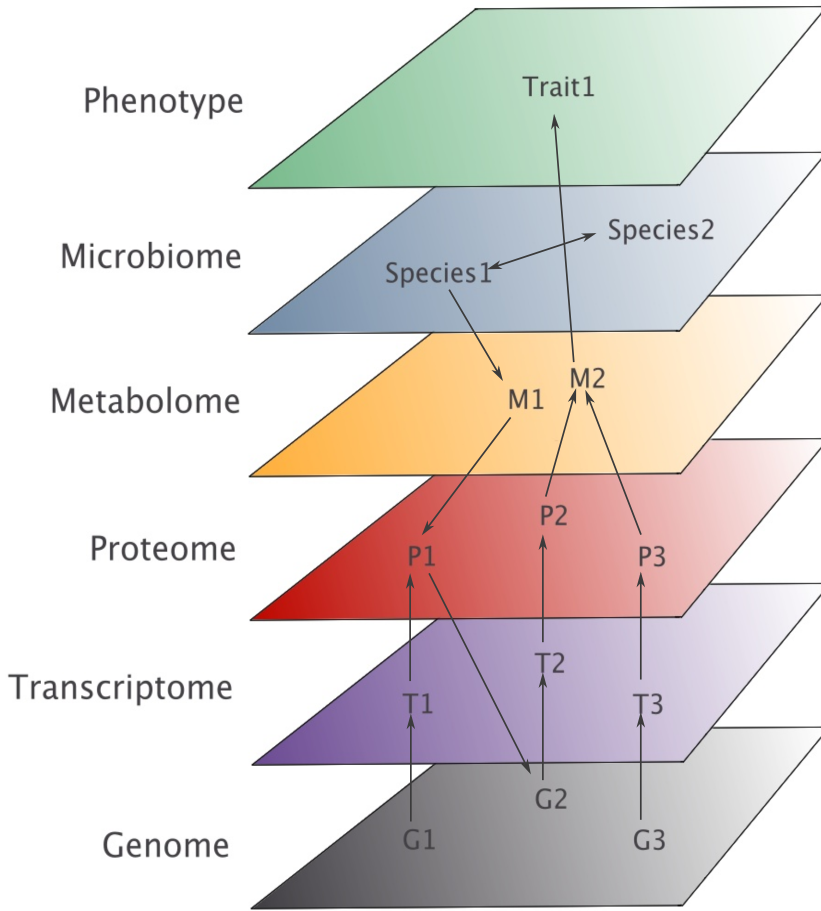

Figure 4. An overview of a systems genetics approach, illustrating how individual 'omic levels are exploited to connect the genome to a phenotypic trait. Figure adapted from Civelek & Lusis (2014).

Material and Methods

This thesis builds on three papers (Figure 5) that collectively address two central research aims:

1. To explore whether natural variation in salicinoid phenolic glycosides (SPGs) composition among *P. tremula* individuals can be used to uncover genes and regulatory mechanisms underlying their biosynthesis.
2. To gain an overview of ontogenic and organ-specific variation in SPG profiles in aspen.

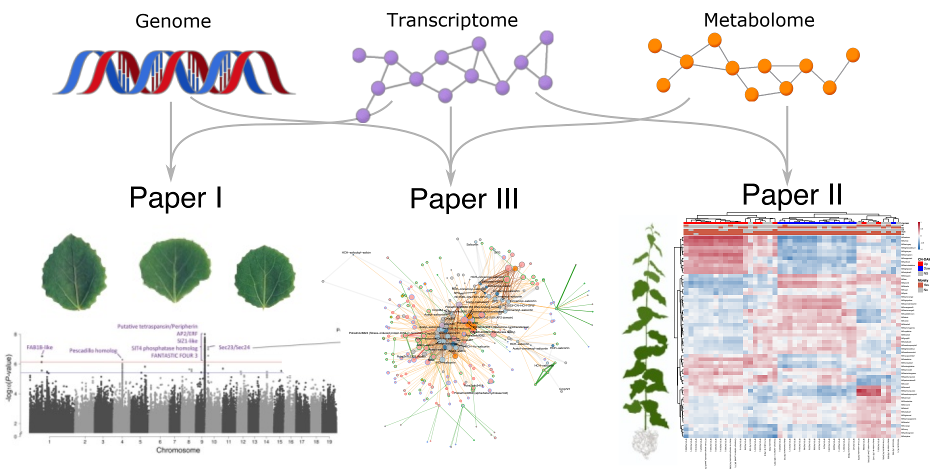

Figure 5. An overview of the 'omic datasets explored in this thesis.

Paper I provides a resource for conducting genome-wide association studies (GWAS) in *P. tremula*, including an improved genome assembly, single nucleotide polymorphism (SNP) datasets from three populations, a best linear unbiased prediction (BLUP) pipeline, expression quantitative trait loci (eQTL) from leaf buds, long non-coding RNAs (lncRNAs) in

leaves, and open chromatin regions (ATAC-Seq). These resources enable genome-wide studies in aspen.

Paper II provides an integrated multi-omic resource, combining transcriptomic and metabolomic data, to explore ontogenic and organ-specific variation in *P. tremula* and to identify candidate genes in the SPG pathway.

Paper III leverages the resources from Papers I and II to construct a systems genetics network, integrating whole-genome resequencing, transcriptomic, and metabolomic data to identify candidate genes in SPG biosynthesis.

Plant material: Swedish Aspen Collection

In Papers I and III, branches from 8-year-old trees in the Swedish Aspen Collection (SwAsp; 110 genotypes) were collected from the Sävar common garden on May 27, 2012, and placed in the greenhouse at the Umeå Plant Science Centre (UPSC) to induce bud break (Mähler et al., 2017). Once bud breaking was initiated, the terminal leaf bud was collected and flash frozen. Buds originating from the same branch and sampled on the same day were used for RNA-Seq (Paper I) and metabolomic analysis (LC-MS) (Paper III).

In Paper II, a subset of the genotypes in the SwAsp collection, belonging to distinct chemotype groups of cinnamoyl (CN, genotypes 14, 34, and 76), acetyl (AC, genotypes 47, 56, and 60), and benzoyl (AC-/CN-, genotypes 41, 45, and 51) were cultivated under *in vitro* conditions (day/night: 18/6h light and 22/18°C). On October 1, 2020, the plants were transferred to soil

(1 L pots) and maintained under greenhouse conditions (day/night 18/6h, 20/15°C). After ten weeks, four biological replicates per genotype were sampled over two days, except genotype 76, which had only three replicates. The following sample types were collected: shoot tips (including the apical meristem and youngest emerging leaves), youngest leaf (3rd from the top), expanding leaf (5th-6th leaf), mature leaf (12th-13th leaf), bark (including cortex, phloem, periderm, and cambium), axillary buds (ca 25-30 buds per sample), and young roots (5-6 white lateral roots). The samples were frozen in liquid nitrogen upon collection and split into two batches – one for LC-MS analysis and one for RNA-Seq.

For the herbivore experiment (see Results and Discussion), aspen (*P. tremula* L.) root cuttings were collected on May 24, 2022, at the Skogforsk station in Sävar (63.4°N, Umeå district). The plants were grown under controlled greenhouse conditions: 20/15°C day/night temperatures, and an 18/6 h photoperiod (LED light FL300 Sunlight, Senmatic, at 80% intensity), and 60% relative humidity. No fungicides or pesticides were applied throughout the experiment. On July 5 and 8, 2022, the trees were re-potted into 1 L pots. The number of replicates obtained from the root cuttings varied among genotypes. In total, 165 plants were included in the experiment: 75 individuals from the AC chemotype (genotypes 47, 56, 60, 61, 67, 86, and 89), 29 individuals from the CN chemotype (genotypes 12 and 22), 32 plants from the AC-/CN- chemotype (genotypes 37, 41, 70, and 71) and 29 plants from the cinnamoyl-intermediate (CN-intermediate) chemotype (genotypes 48 and 51, exhibiting higher CN levels than AC-/CN, but not enough to belong to the CN chemotype). The plants were moved outdoors at the start of the insect experiment (August 3, 2022), at which point tree heights were measured.

Transcriptomic profiling

Frozen plant material was ground with tungsten carbide beads (3 mm) using a Retsch MM 400 Mixer Mill. RNA from leaf material was extracted using the RNeasy Mini Kit (74104; Qiagen), while non-leaf RNA was extracted following a modified protocol version of Chang et al. (1993). In Paper I, RNA sequencing was performed by the Science for Life Laboratory using the Illumina HiSeq 2000 platform to generate paired-end 2 x 100 bp reads. In Paper II, library construction and RNA sequencing were performed by Novogene using the Illumina NovaSeq platform, producing 2 x 150 bp paired-end reads.

The raw read files (FASTQ) were pre-processed using the UPSC bioinformatics (UPSCb) pipeline. Briefly, rRNA was removed with SortMeRna (Kopylova et al., 2012), followed by adapter trimming (Trimmomatic) and removal (Bolger et al., 2014). The processed reads were then mapped to the *P. tremula* v. 2 transcriptome (Paper I), quantified with Salmon (Patro et al., 2017), and aggregated to the gene level with tximport in R (Soneson et al., 2015). The data were normalised using a variance stabilizing transformation (VST) in DESeq2 (Love et al., 2014).

Metabolomic profiling

In Paper II, freeze-dried material (~3 mg) was mixed in a bead mill for 3 min (30 Hz, Retsch GmbH) with 1 mL of cold extraction buffer containing methanol, water, and chloroform (3:1:1). In Paper III, the extraction buffer volume was adjusted to match extracting a 10 mg dry weight bud in 1 mL. The samples were subsequently centrifuged (at ~20,000×*g*) for 10 min (4°C). The upper, aqueous phase was collected, out of which an

aliquot (100 µl in Paper II and 200 µl in Paper III) was dried using a vacuum. The samples were stored at -80°C until further analyses.

Methanol and water (10 µl of each in Paper II and 20 µl of each in Paper III) were added to the samples prior to analysis using an Agilent 1290 Infinity UHPLC system coupled to an Agilent 6546 Q-TOF mass spectrometer (Agilent Technologies). The samples were analysed in both negative and positive modes. In Paper II, quality control (QC) samples containing aliquots of all sampled genotypes were analysed at multiple collision energies (10V, 20V, and 40V) to generate MS/MS spectra used for tentative annotation.

In Paper II, the negative ion mode data were processed in R using XCMS (Smith et al., 2006) to extract feature intensities. Subsequently, RAMClustR (Broeckling et al., 2014) was used to group XCMS features (e.g., adducts, fragments, and isotopes) derived from the same compound, generating 3,620 feature clusters (excluding, among others, ungrouped features). The feature intensities were normalised against sample dry weight (mg) and variance stabilizing normalization (vsn) using MetaboDiff in R (Mock et al., 2018). The MS/MS data were used to annotate targeted compounds tentatively. Furthermore, putative SPGs were detected by identifying differentially abundant metabolites (DAMs) between SwAsp chemotypes using MetaboDiff (Mock et al., 2018).

In Paper III, the peak areas of previously reported SPGs (Abreu et al., 2011; Keefover-Ring et al., 2014; Abreu et al., 2020; Paper II) were calculated using Agilent Masshunter Profinder (Agilent Technologies Inc., Santa Clara, CA, USA) and normalised by sample concentration.

Association studies

Each trait (leaf shape, Paper I) and metabolomic features (Paper III) were pre-processed using the BLUP pipeline to evaluate whether residuals conformed to the normality assumptions of linear regression models. Traits with non-normal residuals were normalised using the BestNormalize package in R (Peterson, 2021) and outlier samples were identified and removed using the outlierTest function from the car package (Fox & Weisberg, 2019). GWAS studies were conducted using linear mixed models implemented in GEMMA (Zhou & Stephens, 2012) or fastJT (Lin et al., 2019) to identify SNPs significantly associated (q-value < 0.05) with the traits (Papers I and III).

Expression quantitative trait loci (eQTL) analyses were performed using Matrix eQTL (Shabalin, 2012) and fastJT to identify SNPs significantly associated with gene expression (FDR < 0.05) and classify genes as acting in *cis* or *trans* (Paper I).

Network analyses

In Paper II, a gene co-expression network was generated for each leaf developmental stage using Weighted Gene Co-expression Network Analysis (WGCNA; Langfelder & Horvath, 2008). In Paper III, both the metabolomic and transcriptomic datasets were included in the WGCNA analysis, yielding a combined metabolite–gene expression network.

In Paper III, the GWAS and eQTL results were further integrated to create an association network visualising metabolite–gene and gene–gene relations based on shared SNP associations. Specifically, a gene and a metabolite were linked when both were significantly associated with the

same SNP ("Exact SNP (mGWAS/eQTL)"). Similarly, if both were associated with SNPs within the same genomic region (±2 Kb), an edge classified as "SNP window (mGWAS/eQTL)" was used. eQTLs and mGWAS hits located within or near a gene (upstream/downstream) were also represented in the network, classified as eQTL and mGWAS, respectively. The weight of the relationship was generated by scaling the -log(*P*-value) to range from 0.1 to 1. The association network and the metabolite–gene WGCNA network were subsequently integrated to generate a systems genetics network.

No-choice insect experiment

In the herbivore experiment (see Results and Discussion), eggs of *O. antiqua* were purchased from Worldwide Butterflies (https://www.wwb.co.uk/, DT9 4QN, Sherborne, United Kingdom), and the emerging larvae were reared exclusively on rowan (SV rönn) leaves. At the start of the experiment, three second-instar larvae were placed on each tree. To track stem growth during the experiment, a piece of cotton yarn was loosely tied around the youngest leaf. Each plant was covered with a garden fleece bag (115 x 40 cm) and secured with a drawstring to confine the larvae. A small piece of garden fleece was placed between the stem and the bag opening to prevent any gaps. If the bag did not enclose the entire plant, a piece of cotton yarn was loosely tied around the leaf closest to the bag opening to mark the boundary between insect-exposed and non-exposed foliage. The plants were grown outdoors throughout the experiment (Figure 6a), which ended after 2 weeks (on August 17). Two days after the insects were collected, the shoot apex was removed to prevent further growth, and insect damage was assessed. Damaged leaves with holes larger than 0.5-1 mm were collected and scanned (300 DPI). A

plant label was included in one of the scans as a reference of known size. The plant height was later measured to assess tree growth. The number of leaves developed during the experiment (i.e., leaves above the top cotton string), the total number of leaves that were enclosed by the fleece bag (i.e., above the bottom cotton string), and those not included in the experiment (i.e., below the bottom cotton string) were also documented.

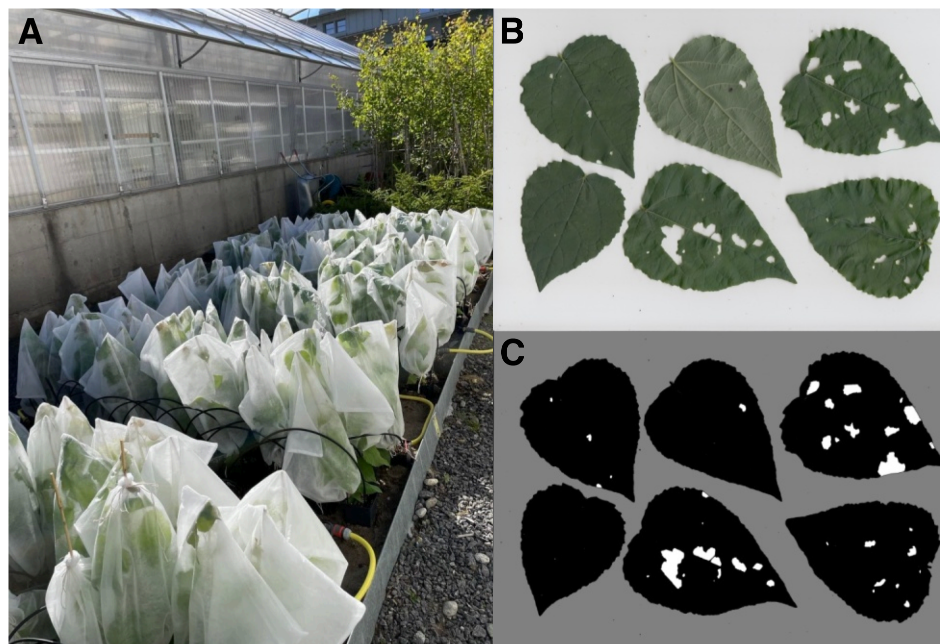

Figure 6. Feeding experiment with rusty tussock moth (*Orgyia antiqua*). (a) Overview of the experimental site in the greenhouse courtyard. Example scan of larval feeding damage on leaves (b) pre-image and (c) post-image analysis.

Image analysis of foliar damage

In the herbivore experiment (see Results and Discussion), foliar damage was analysed using ImageJ version 1.52 (Schneider et al., 2012). To reconstruct the leaf area missing along the margins, a line was manually drawn between the edges of the eaten sections (Figure 6b). Pre-processed

images were then analysed automatically (see https://github.com/sarawestman/Herbivore/) to quantify foliage consumed by larvae. First, the global scale was calibrated using a plant label as a reference (pixel distance = 1169.6438, known distance = 9 cm). The colour threshold was set to: Hue = 0-255, Saturation = 0-255, Brightness = 156-255. The image was converted to binary (black and white), then re-converted to an 8-bit image and the background was set to grey. The number of white, grey, and black pixels was quantified using the Histogram tool. White pixels were used as a proxy for damaged leaf area (Figure 6c).

Results and Discussion

An Improved Chromosome-scale Genome Assembly and Population Genetics resource for *Populus tremula* (Paper I)

High-quality, contiguous genome assemblies and gene annotations are foundational resources for performing systems genetics studies. In this paper, we presented an updated genome assembly of *P. tremula* generated from long-read sequencing, complemented by optical and high-density genetic maps. This high-quality reference genome was then annotated using an extensive collection of RNA sequencing evidence and homology-based approaches. We subsequently performed SNP calling and GWAS on three aspen populations to identify genes associated with leaf physiognomy traits, including leaf circularity, leaf indentations, and 23 additional leaf shape and size metrics.

Previous association studies of leaf shape variation in aspen suggested that it is a highly complex trait influenced by many small-effect size variants (Mähler et al., 2020). This, together with the stringent multiple-testing correction required in GWAS, can result in insufficient statistical power to detect causal variants (Benstock et al., 2024). Moreover, linear regression models, such as GWAS, are sensitive to both outliers and the trait's underlying distribution (assuming normal residuals). If ignored, the type I error rate and statistical power can be affected, with the strongest effects observed on rare variants (Chien, 2020). To increase statistical power, we conducted GWAS in three aspen populations of different sizes, analysing the complete set of SNPs and a subset comprising SNPs only found in open chromatin regions. Furthermore, to

comply with the underlying GWAS assumptions, we created a best linear unbiased prediction (BLUP) pipeline (Figure 7a) that removes outlier samples, tests for normality and normalises non-normally distributed traits. This pipeline had a major effect on the number of significantly associated SNPs from the GWAS, as illustrated in Figure 7b, which compares the *P*-values obtained from GWAS based on BLUPs with or without outlier removal and normalisation. Consequently, all physiognomy traits were processed using the BLUP pipeline before performing the GWAS.

No significant associations were identified in the SwAsp collection (n = 99) when including all SNPs (6,806,717 SNPs). However, performing GWAS on SNPs located in open chromatin regions (185,616 SNPs) resulted in four significant associations (q-value < 0.05) for two traits. When we examined the same leaf phenotypes in two larger *P. tremula* populations from northern Sweden (UmAsp, n = 227) and Scotland (ScotAsp, n = 105), all SNP-GWAS identified significant associations for one leaf shape trait in UmAsp and two traits in ScotAsp. Significant associations for two additional traits in the ScotAsp collection and one additional trait in the UmAsp collection were identified when performing GWAS on SNPs located in open chromatin regions.

Taken together, we identified 25 genes with SNPs associated with seven leaf physiognomy traits. A subset of these genes was annotated as being involved in plant development (including Potra2n9c199975, Potra2n9c199981, and Potra2n9c199982), representing compelling candidates for functional validation.

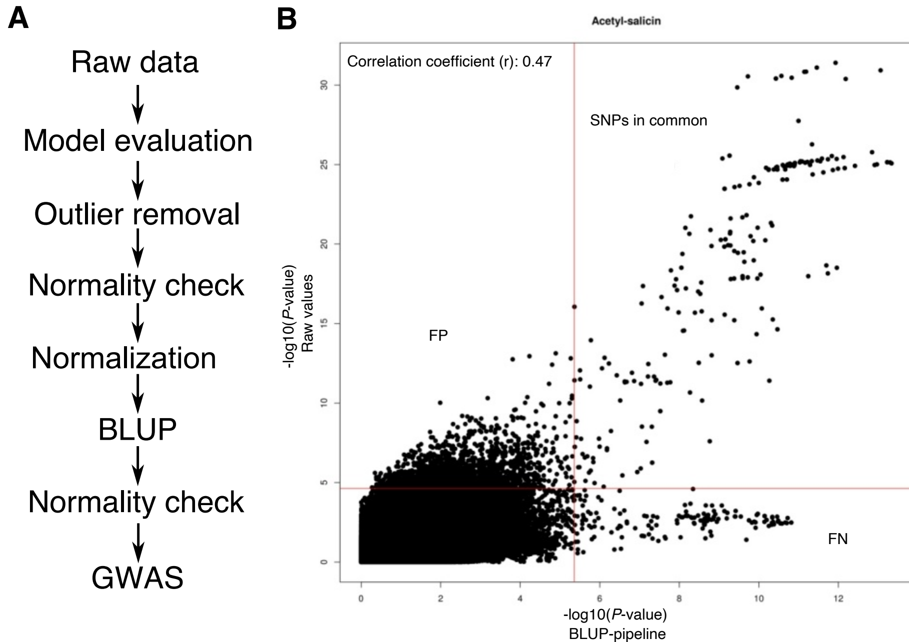

Figure 7. Overview of the best linear unbiased prediction (BLUP) pipeline, showing (a) the processing steps and (b) an example comparing genome-wide association (GWA) results (-log10(*P*-value)) obtained from BLUPs without outlier removal and trait normalisation (x-axis) against BLUPs with outlier removal and normalisation (y-axis). The red lines represent the significance threshold (adjusted *P*-value < 0.05). FP = False Positive; FN = False Negative; SNPs = single nucleotide polymorphisms.

An integrated multi-omic resource to identify candidate genes in the salicinoid phenolic glycosides pathway (Paper II)

In this study, we aimed to identify candidate genes involved in SPG biosynthesis by integrating transcriptomic (mRNA-Seq) and metabolomic (LC-MS) data from aspen (*P. tremula*) individuals from the SwAsp collection that exhibit contrasting SPG profiles, or chemotypes. We compared individuals with a high abundance of cinnamoyl-SPGs (CN chemotype) to those with higher abundance of benzoyl (AC-/CN-

chemotype). Moreover, data were collected from various plant organs, ranging from shoot tips to roots, to determine the extent of ontogenic and organ-specific variation in SPG abundance and diversity – crucial information for designing future experiments linking gene expression to SPG biosynthesis. For this paper, we focused particularly on SPG variation across four developmental leaf stages: shoot tip, young leaf (3rd leaf), emerging leaf (5th-6th leaf), and mature leaf (12th-13th leaf). All data are publicly available at PlantGenie.org, providing a resource for exploring the full range of SPG chemotypes and plant organs.

Differential expression analysis comparing the CN chemotype to the AC-/CN- chemotype revealed a higher number of differentially expressed genes (DEGs) in younger leaves compared to older leaves. More specifically, 306 and 178 genes for shoot tips and young leaves and 106 and 115 genes for emerging and mature leaves, respectively (Table S5 in Paper II). Eleven DEGs were shared across all leaf stages, three of which have annotated Protein family (Pfam) domains potentially related to disease resistance (Table S6, Paper II). In addition, DAM analysis on the LC-MS data identified 27 putative SPGs (Table S2 in Paper II).

Gene co-expression networks were constructed for each leaf stage to identify modules (or clusters) potentially involved in SPG biosynthesis. Each network was correlated with the abundance of known and putative SPGs, and the module showing the most statistically significant correlation with HCH-cinnamoyl-salicortin (the most influential CN-SPG, Figure 3 in Paper II) and/or containing the highest proportion of DEGs was identified for each leaf stage. Once again, younger leaves stood out, with the shoot tip and younger leaf module containing a higher proportion of DEGs (28% and 18%, respectively) compared to emerging and mature leaves (5% and 6%, respectively) (Table S7, Paper II). Notably, the shoot

tip module was enriched for genes associated with phenylpropanoid biosynthesis (Figure 5 in Paper II), including several genes previously linked to this pathway by Tsai et al. (2006). Thus, candidate modules from younger leaf stages are particularly conducive to identifying candidate genes involved in SPG biosynthesis.

Previous studies have suggested that immature leaf tissue is an active site of SPG biosynthesis. For example, Massad et al. (2014) found that SPGs are likely to be synthesised in immature leaves by studying SPG turnover in *P. trichocarpa*. They observed faster ^{13}C incorporation into SPGs in immature leaves than in mature leaves. Moreover, damaged plants growing early in the season exhibited a more rapid SPG turnover and a greater accumulation of glucose, the core sugar molecule of SPGs, than mature leaves, suggesting that immature leaves are active sites of SPG production. In another study, inhibition of Phe ammonia-lyase (*PAL*) activity in *Salix myrsinifolia* led to a reduction of SPGs and an increase of phenylalanine in the shoot tips (Ruuhola & Julkunen-Tiitto, 2000). Mature leaves, however, contained the lowest phenylalanine level with no significant SPG turnover detected, indicating that the plant prioritises protecting the shoot tip. In addition, Subramaniam et al. (1993) found that mature leaves and older stems exhibited the lowest *PAL1/2* expression in *P. trichocarpa* x *Populus deltoides*. In contrast, young leaves, young stems and apical buds showed the highest expression levels. Together, these studies support our findings that younger leaf stages are active sites of SPG biosynthesis.

Hypothesising that genes in the SPG pathway are more active in younger than in older leaves, we identified two HXXXD-type acyltransferases and one UDP-glycosyltransferase in the candidate shoot tip network module. Such enzymes are common in the phenylpropanoid pathway, in which

acyltransferases diversify phenylpropanoids (Bontpart et al., 2015) and UDP-glycosyltransferases perform sugar transfers (Le Roy et al., 2016). The *in vitro* function of these genes needs further validation.

A systems genetics study reveals novel candidate genes for salicinoid biosynthesis in aspen (Paper III)

In this paper, we aimed to utilise the natural variance among individuals of the SwAsp collection to unravel the genetic basis of SPG biosynthesis. We performed metabolite GWAS (mGWAS) on SwAsp leaf bud, including twenty-one known SPGs. Significant mGWAS associations (ranging from 2 to 602 per metabolite) were identified for ten SPGs adorned with cinnamoyl and/or acetyl moieties. A gene hotspot on chromosome 4 emerged from these analyses, encompassing six genes (Potra2n4c9081, Potra2n4c9088, Potra2n4c9089, Potra2n4c9091, Potra2n4c9092, and Potra2n4c9093) shared by at least two SPGs with an acetyl group and two SPGs with a cinnamoyl/salicyloyl group. Within this hotspot, Potra2n4c9093 had the greatest number of significant SNPs across all SPGs, with more than two associations, and was the only gene positively correlated with the abundance of all cinnamoyl- and acetyl-SPGs (Paper III).

We also performed mGWAS on two larger aspen populations, UmAsp (n = 227) and ScotAsp (n = 105), for the majority of SPGs tested in SwAsp buds. Six genes were common to all populations: Potra2n4c9072, Potra2n4c9081, Potra2n4c9089, Potra2n4c9091, Potra2n4c9092, and Potra2n4c9093. Notably, salicin, salicortin, and tremulacin, among other metabolites, showed a significant association with Potra2n4c9093 in

UmAsp, while in ScotAsp, only the acetylated SPGs were associated with the gene.

A systems genetics network was constructed using metabolite and gene expression data from the SwAsp collection to identify candidate genes and regulatory pathways involved in SPG biosynthesis. Briefly, we integrated eQTLs (Paper I) and mGWAS (Paper III) from leaf buds to generate an "association network". This association network was then combined with a leaf bud gene–metabolite co-expression network, together forming the systems genetics network. When extracting the first-degree neighbours (FDNs) of the SPGs with significant associations from the systems genetics network, we observed that acetylated SPGs and cinnamoyl/salicyloyl-SPGs formed two distinct clusters. Both clusters were significantly associated with ten genes (Potra2n9080, Potra2n4c9081, Potra2n4c9090, Potra2n4c9091, Potra2n4c9092, Potra2n4c9093, Potra2n6c13578, Potra2n8c17226, Potra2n14c26839, and Potra2n17c31159). Among these genes, Potra2n3c9093 emerged as the strongest candidate – showing the highest eigenvector centrality within the systems genetics network, the strongest association weight with acetylated SPGs, and the greatest number of SNPs associated with these metabolites. Moreover, the gene was annotated as an HXXXD-type acyltransferase, a family known to diversify phenylpropanoids through acyl-CoA-dependent reactions (D'Auria, 2006).

To functionally characterise Potra2n4c9093, we performed cDNA cloning and heterologous expression in *E. coli*. The purified enzyme was then used in an acetylation assay with SPGs (salicin, salicortin, cinnamoyl-salicortin, and tremulacin, respectively) and acetyl-CoA, confirming its acetyltransferase activity. Consistent with these results, overexpressing the gene *in planta* in a genotype with naturally low abundance of

acetylated-SPGs increased the levels of acetylated-SPGs indicating that the enzyme can acetylate SPGs. The mutants also showed elevated cinnamoyl-salicortin levels, suggesting that the enzyme might catalyse additional acyl group transfers, although this needs further verification.

Herbivory by *Orgyia antiqua* is unaffected by salicinoid phenolic glycoside chemotypes in *Populus tremula*

In this study, we aimed to determine whether there were differences in rusty tussock moth (*O. antiqua*) feeding on *P. tremula* individuals with distinct SPG profiles, or chemotypes. The rusty tussock moth (or vapourer moth) is a generalist insect native to the Palaearctic region of Europe but can also be found in North America. The larvae feed on coniferous and broadleaf species, as well as some groundcover plants, including *Plantago major* (SV groblad) and *Vaccinium myrtillus* (SV blåbär). Since the adult moth does not feed, it must survive on the energy stored as a larva (Weir & Boyes, 2024), making the larvae an attractive study system in feeding experiments.

Four-week-old plants (n = 165) were used in the experiment, representing contrasting SPG chemotypes (AC, CN, AC-/CN- and CN-intermediate). Each plant was exposed to three second-instar *O. antiqua* larvae in a no-choice field experiment, but to mimic natural conditions, the larvae were allowed to feed on the whole plant, with the occasional exception of basal leaves. Moreover, they were exposed to the environmental fluctuations in the greenhouse courtyard. The experiment ended after two weeks of feeding, during which the larvae were collected, and feeding was assessed by scanning damaged leaves, using pixel loss as a proxy for feeding damage.

The feeding damage varied among genotypes, with high variation observed across most replicates of the same genotype (Figure 8). This intragenotype variation may partly be due to differences in the number of surviving larvae at the end of the experiment. To account for larval counts, we calculated BLUPs using genotype and the number of surviving larvae as random and fixed effects, respectively. To capture both the herbivore and plant perspectives, BLUPs were calculated based on the leaf area consumed (herbivore perspective) and the proportion of leaf area loss (plant perspective). Since the number of genotypes for CN and CN-intermediate chemotypes was few (only two genotypes per chemotype), they were not further assessed.

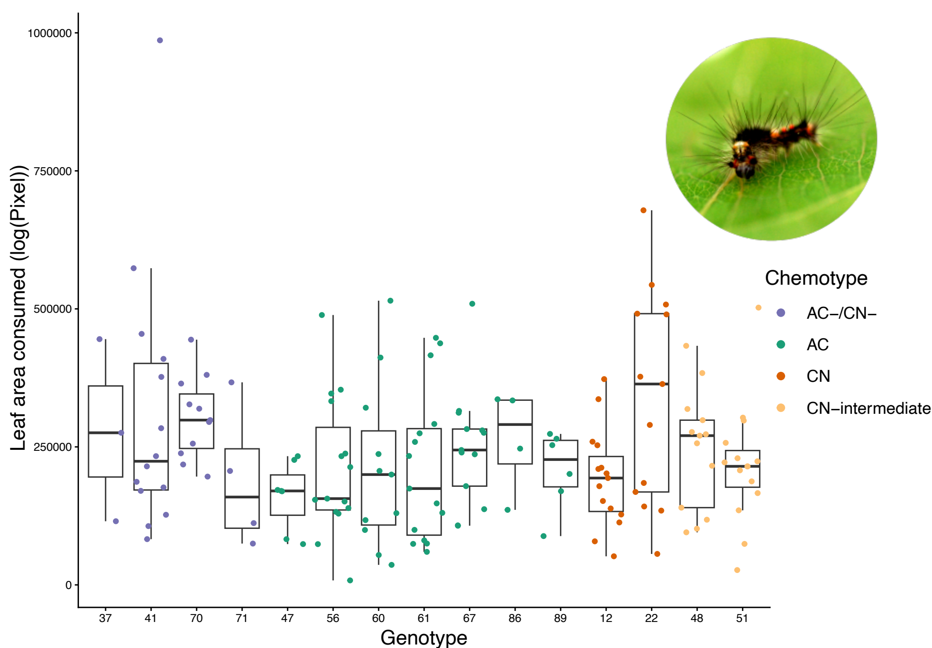

Figure 8. Consumption of Swedish Aspen (SwAsp) genotype leaves by rusty tussock moth (*Orgyia antiqua)* in a no-choice assay. The x-axis shows the genotype identifiers (IDs) of the SwAsp individuals, while the y-axis shows the leaf area consumed by the larvae. Log-transformed pixel count was used as a proxy of consumed leaf area. Dot colour reflects the genotype's chemotype classification.

There was no significant difference between AC and AC-/CN- chemotypes, whether BLUPs were based on leaf area consumed (Figure 9a; Wilcoxon test: P = 0.16) or proportion of leaf area loss (Figure 9b; Wilcoxon test: P = 0.53). Moreover, there was no significant correlation between the leaf area consumed (Pixel) and height of the trees at the start of the experiment (Pearson's correlation coefficient r = -0.12, P = 0.14) nor the total number of leaves on the tree at the end of the experiment (Pearson's correlation coefficient r = -0.15, P = 0.07). There was also no significant difference in larval survival between chemotypes (Kruskal-Wallis test: P = 0.51) or genotypes (Kruskal-Wallis test: P = 0.71).

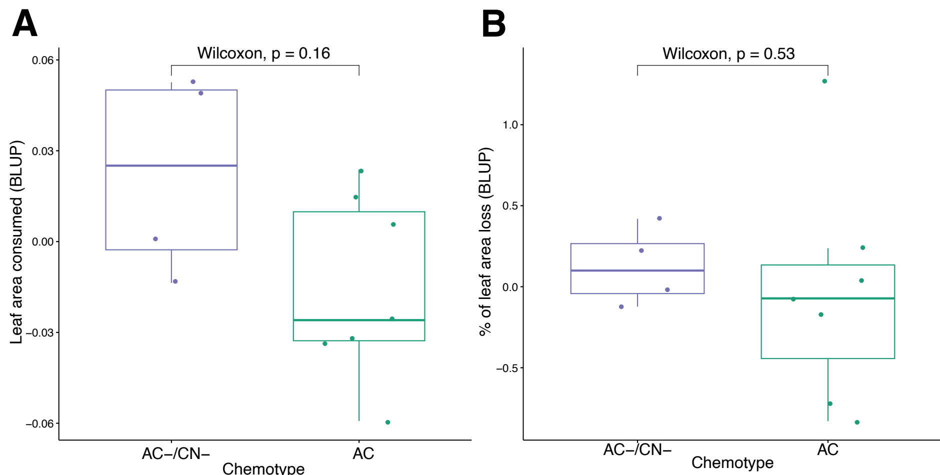

Figure 9. Consumption of chemotype leaves by rusty tussock moth (*Orgyia antiqua*) in a no-choice assay. The x-axis shows the chemotype groups, with each dot representing a genotype. The y-axis reports the leaf area consumed by the larvae, accounting for the number of surviving larvae by calculating best linear unbiased prediction (BLUP). The BLUPs were calculated based on (a) the leaf area consumed, and (b) the percentage (%) of leaf area loss.

Previous studies investigating the effects of naturally varying SPG-producing genotypes on herbivore performance have typically only focused on salicortin and tremulacin levels, finding that high SPG concentration led to reduced feeding and/or had a negative impact on

herbivore performance (e.g., Wooley et al., 2008; Osier & Lindroth, 2001; Osier & Lindroth, 2006; Rubert-Nason et al., 2015; Villalba et al., 2014). However, few studies have explored how other SPGs, such as cinnamoylated or acetylated SPGs, affect herbivore performance.

Ikonen et al. (2002) performed a two-choice laboratory feeding trial with the leaf beetle *Agelastica alni* (SV allövbagge), offering the beetle leaves from five different plant species that varied in their content of specialised metabolites. The beetle consistently rejected willow species with a high concentration of salicortin or acetyl-salicortin. Additionally, Kolehmainen et al. (1994) examined the effects of SPGs (including acetyl-salicortin, salicortin, and salicin) on the oviposition preference of the specialist shoot galling sawfly *Euura amerinae* L. They observed that the fly could discriminate among willow plants, preferring to lay eggs on those with high levels of acetyl-salicortin. The effect of salicortin, however, was significantly weaker, and salicin showed no detectable impact. Moreover, Fields and Orians (2006) found a negative correlation between cinnamoyl-salicortin concentration and leaf area consumed by *Plagiodera versicolora* (SV bred videbagge) larvae, but not by *P. versicolora* and *Calligrapha multipunctata bigsbyana* adults. Four days after feeding, systemic induction of cinnamoyl-salicortin was detected in undamaged leaves above the feeding site for all insect species tested. However, no systemic effects of salicortin were observed. Taken together, these results highlight the ecological importance of SPG diversity in Salicaceae.

This study aimed to assess variation in feeding by *O. antiqua* among different SPG chemotypes in the SwAsp population. Although the genotype replication was relatively high (on average, 10.6 trees per genotype), the chemotype representation was suboptimal for CN and CN-

intermediate chemotypes; therefore, these chemotypes were not assessed in this study. To summarise, we did not detect a difference in feeding intensity by *O. antiqua* when comparing AC and AC-/CN- chemotypes. Although many studies consider SPGs important plant defences, there are studies where it was shown that other factors than SPGs had a greater impact on feeding preference. For example, Holeski et al. (2016) did not detect a significant effect of SPGs on white-tailed deer (*Odocoileus virginianus;* SV vitsvanshjort) feeding preference. Instead, nutritional factors, such as sugar, mineral, and protein concentrations, appeared to influence their food intake. In agreement, Villalba et al. (2014) observed that sheep (*Ovis aries;* SV tamfår) increased their intake of aspen when the nutritional quality of other vegetation was poor, suggesting that nutrition rather than SPG concentration was the decisive factor. The sheep, however, could discriminate between high- and low-SPG-producing aspens. In a follow-up study, Heroy et al. (2018) also observed that sheep consumed more aspen when the leaves had lower SPG concentration. However, when the sheep could choose between high- and low-SPG-producing aspens, factors such as nutrition and chemical composition were more important than SPG concentration. The authors suggest that herbivore preference is most likely multidimensional, resulting from the herbivore's nutritional state and plant characteristics. However, it is important to note that very few SPGs (tremulacin and salicortin) were used as proxies for SPGs.

Although SPG chemotypes are generally robust in both field and greenhouse experiments (Keefover-Ring et al., 2014), SPG abundance and diversity were not assessed in this experiment. Consequently, we did not directly determine the relationship between individual SPGs and insect feeding. Future assessment of *O. antiqua* feeding preferences should ideally be conducted as a multi-choice experiment, which may provide a

better understanding of the herbivore's chemotype preferences. In conclusion, additional experiments are needed to determine whether SPG chemotypes affect *O. antiqua* feeding.

Conclusions and Perspectives

This work provides new insights into the biosynthesis of SPGs and the genes involved.

Paper I provides scientists with a resource for identifying genes contributing to trait variation in *P. tremula* by analysing genomic, transcriptomic (RNA-Seq and lncRNAs), and epigenetic (ATAC-Seq) data. The utility of the resource was demonstrated through GWAS of leaf shape, yielding a set of candidate genes requiring further validation. The value of the resource was further demonstrated in the following papers, which used the updated assembly for read mapping (Paper II), and for eQTL and GWAS analyses (Paper III).

Paper II provides an additional resource for identifying candidate genes in the SPG pathway, offering exploration of metabolomic and transcriptomic data. Moreover, because the resource included data from multiple organs (including buds, roots, bark, and leaves), it enabled exploration of ontogenic and organ-specific variation in metabolite and gene expression. Two HXXXD-type acyltransferases and one UDP-glycosyltransferase were identified as compelling candidates in the SPG pathway by analysing leaves at four developmental stages. In agreement with previous studies on SPG turnover (Massad et al., 2014), immature leaf tissues emerged as a site of active SPG biosynthesis. For all identified candidates, future studies assessing the *in vitro* function of these genes are needed.

Paper III integrates a gene–metabolite co-expression network, mGWAS and eQTL results, together forming a systems genetics network. Within this integrated network, an HXXXD-type acyltransferase showed a strong association with both acetyl- and cinnamoyl-SPGs. We performed heterologous expression of the gene in *E. coli,* purified the protein, and performed an enzyme acetylation assay using four different SPGs and acetyl-CoA. The enzyme acetylated all tested SPGs, confirming its acetyltransferase activity under *in vitro* conditions. Overexpression of the gene *in planta* resulted in increased levels of acetylated SPGs, further confirming its acetylation activity. Notably, the mutant also accumulated higher levels of cinnamoyl-salicortin, suggesting that the enzyme may transfer additional acyl groups, although this requires further verification. The systems genetics network offers a framework for further exploration of candidate genes in the SPG pathway.

Taken together, these studies highlight the main challenges biologists face when unravelling the genetic basis of complex traits. First, the trait variation within a population, along with the number of individuals, sets a limit on what can be detected. This likely explains the differences in the number of significant associations observed among aspen populations in **Papers I** and **III**. Second, the tissue selected, the number of 'omic levels examined, and the timing of sampling will influence the extent to which we can interpret the underlying biology. For example, as shown in **Paper II**, linking gene expression to metabolite abundance requires the gene to be active at the right time and in the right tissue. Moreover, such links are further complicated by the transport and accumulation of metabolites across tissues, which introduces additional complexity. Third, the interpretation depends on the sensitivity and accuracy of the instruments measuring the 'omic levels and, subsequently, on the downstream analytical tools used. This was clearly illustrated by the BLUP pipeline in

Papers I and **III**, showing that the underlying model assumptions should be carefully considered to reduce the number of spurious associations identified by GWAS, which would introduce noise to the systems genetics network, hampering interpretation.

Once candidate genes are identified, researchers face an additional challenge: confirming gene function. In this thesis, we have identified a set of candidate genes involved in SPG biosynthesis and leaf shape, the latter of which is expected to have a more complex genetic architecture. The *in vitro* and *in vivo* experiments conducted in **Paper III** confirmed the acetylation activity of the enzyme. However, overexpression of the gene additionally resulted in elevated levels of cinnamoyl-SPGs, suggesting potential substrate promiscuity. An alternative explanation is that increased levels of acetylated SPGs trigger a metabolic feedback loop that influences the activity of other genes in the SPG pathway. To improve our understanding of the gene function, a broader set of acyl donors (including cinnamoyl-CoA) and kinetic assays would help determine substrate preference and enzyme promiscuity. Moreover, comparing the metabolomic and transcriptomic profiles of the overexpression line with those of the wild type could reveal effects on upstream phenylpropanoid metabolism and on candidate genes identified in the systems genetics network. A knockout would further confirm gene function and uncover potential genetic redundancy.

Despite numerous publications on SPGs and herbivory, we did not detect a link between SPGs and *O. antiqua* feeding preference in our trial, highlighting the need to broaden the scope of future work. Expanding the representation of chemotypes, particularly those enriched in cinnamoyl and benzoyl moieties, could reveal patterns that were not apparent in the current design. Moving beyond no-choice assays to experiments that

allow insects to select among multiple chemotypes may uncover preference dynamics that better reflect natural conditions. Integrating transcriptomic and metabolomic profiling during herbivory would enable capturing induced responses and identification of regulatory networks that govern plant defence. Functional validation using overexpression lines and knockout mutants of candidate genes identified in this thesis will be critical for establishing causal links between SPG biosynthesis and herbivore performance. Finally, extending studies to field conditions and incorporating multiple herbivore species, including browsing mammals, will provide a more realistic perspective on how SPG diversity shapes community-level interactions. This experiment underscored the complexity of plant defence and opens exciting opportunities for future research. By leveraging the genomic and multi-omic resources developed in this thesis, subsequent studies can move toward a mechanistic understanding of SPG function and its ecological significance, ultimately advancing our knowledge of how chemical diversity contributes to plant resilience.

In summary, 'omic integration represents a powerful tool for studying biological systems, but it is still in its infancy. We should, however, not forget that we have come a long way since the elucidation of the DNA double helix (Watson & Crick, 1953). The pace of technological advancement is perhaps best illustrated by the dramatic increase in throughput and reduction in sequencing costs – from approximately $2.7 billion for the first human genome project (HGP) (Giani et al., 2020) to roughly $200 per genome today (Schwartz, 2023). Over time, many of the challenges emphasised in the previous paragraphs will likely diminish. However, we will probably continue to face another crucial obstacle, one that remains relevant even with the comparatively "small" datasets we are handling today: That is, how to disentangle these 'omics to extract

meaningful biological signals and understand how each of them interacts to shape a biological system. As stated by Trewavas (2006):

> *"Understanding the complexity of biological systems represents the greatest intellectual and experimental challenge yet faced by any biologist"*

This will undoubtedly remain a major bottleneck in biology for the foreseeable future, although the application of new machine learning and artificial intelligence approaches holds much potential. Ultimately, the only true limitation we may face is our own imagination.

Acknowledgements

Nat

After many years of studying biology, I finally know what I want to do in life – and I have you to thank for that! I am so thankful for your support and guidance. You have always been there to listen, offer support, and give advice. I also want to thank you for giving me the freedom to explore this project in a way that fitted my interest – it made the journey incredibly exciting. I thought bioinformatics would initially freak me out, given that I had made a solemn vow not to pursue it after my master's, but you made sure that it was manageable and (eventually) fun!

Benedicte

The topic of this thesis has been a perfect match for my scientific interests – a blend of ecology and molecular biology – a field you first introduced me to during my master's. Thank you for your support and guidance throughout this project. It has been a delight to discuss insects and the ecological aspects of this work with you, and it has been comforting to know that I could always turn to you for advice and support.

Tiggy

Thank you for your endless support. With you by my side, it felt manageable and fun to dive into a completely new topic. I have spent countless hours discussing the project with you, and your enthusiasm has truly inspired me. It has been especially helpful during those times when I have been stuck or lost in a knee-high pile of data, not knowing which direction to take. So, thank you for being my partner in crime!

UPSC mates

Among the many things life has taught me, it is only as fun as the people you share it with. Thank you to the Street lab – Edo, Elena, Amanda, Nat, Nico, Jamie, Teitur, Fai, Vikash, Ashish, Sagar, and Eduardo – and to my former co-workers – Kristina, Camy, Chanaka, Niklas, Bastian, Alonso, Aarón, and Andreas. You have made this journey highly entertaining with movie nights, decorating the Christmas sofa, board game evenings, BBQs, and more. A special thanks to Niklas, who was my GWAS guru, and to Vikash for generating the mutant lines and performing the heterologous expression experiment for this thesis.

Thank you to my fellow PhD students and UPSC coworkers – especially Sanchali, who has been a fantastic friend and PhD companion.

Thank you, Hasse, Jenna, and Krister, for being my metabolomic support group. Without you, I would have been completely lost in the metabolomic realm – where my sanity sure would have been lost as well.

I also want to thank my reference group, Markus and Maria, for your valuable input.

To my family

Thank you to the loves of my life – my husband and daughter – for your unconditional love and support. You two kept me sane, even though I am sure that there were moments when you must have questioned my sanity – especially that time I brought home our new "summer pets" (i.e., thousands of rusty tussock moth larvae). Thank you, Jocke, for being my rock throughout this PhD – not only when I needed a pair of extra hands for making leaf discs for my insects, but also for listening to me ramble about my experiments and making sure that my work–life balance was

(somewhat) balanced. And thank you, Thea, for making it so easy to focus on something other than my PhD. One smile or hug from you can brighten up the gloomiest days. Thank you to my mother, Karin, and my sister, Ida, for your unwavering love and support. I cannot find the right words to describe how much the four of you mean to me. To put it simply: If I were a table, you would be the legs that keep me standing.

Dad – losing you has been the biggest challenge life has sent my way so far. Your motto, “nothing is impossible”, has carried me through many of the rough patches this PhD offered. Still, I wanted your presence to be reflected in another way – so why not through a poem in Norsjömål, the place where you grew up?

References

Abdel-Aziz, S. M., Aeron, A. & Kahil, T. A. (2016). Health Benefits and Possible Risks of Herbal Medicine. In Garg, N., Abdel-Aziz, S. M. & Aeron, A. (eds). *Microbes in Food and Health*. Springer International.

Abreu, I. N., Ahnlund, M., Moritz, T. & Albrectsen, B. R. (2011). UHPLC-ESI/TOFMS Determination of Salicylate-like Phenolic Gycosides in *Populus tremula* Leaves. *Journal of Chemical Ecology*, *37*(8), 857–870. https://doi.org/10.1007/s10886-011-9991-7

Abreu, I. N., Johansson, A. I., Sokołowska, K., Niittylä, T., Sundberg, B., Hvidsten, T. R., Street, N. R. & Moritz, T. (2020). A metabolite roadmap of the wood-forming tissue in *Populus tremula*. *New Phytologist*, *228*(5), 1559–1572. https://doi.org/10.1111/nph.16799

Aderemi, A. V., Ayeleso, A. O., Oyedapo, O. O. & Mukwevho, E. (2021). Metabolomics: A Scoping Review of Its Role as a Tool for Disease Biomarker Discovery in Selected Non-Communicable Diseases. *Metabolites*, *11*(7), 418. https://doi.org/10.3390/metabo11070418

Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L. K., Saito, K. & Kanaya, S. (2012). KNApSAcK family databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology*, *53*(2), 1–12. https://doi.org/10.1093/pcp/pcr165

Albrectsen, B. R., Siddique, A. B., Decker, V. H. G., Unterseher, M. & Robinson, K. M. (2018). Both plant genotype and herbivory shape aspen endophyte communities. *Oecologia*, *187*(2), 535–545. https://doi.org/10.1007/s00442-018-4097-3

Alpay Savasan, Z., Kim, S. K., Oh, K. J. & Graham, S. F. (2021). Advances in cerebral palsy biomarkers. *Advances in Clinical Chemistry*, *100*, 139-169. https://doi.org/10.1016/bs.acc.2020.04.006

Alseekh, S., Ofner, I., Liu, Z., Osorio, S., Vallarino, J., Last, R. L., Zamir, D., Tohge, T. & Fernie, A. R. (2020). Quantitative trait loci analysis of seed-specialized metabolites reveals seed-specific flavonols and differential

regulation of glycoalkaloid content in tomato. *The Plant Journal*, *103*(6), 2007–2024. https://doi.org/10.1111/tpj.14879

Ambardar, S., Gupta, R., Trakroo, D., Lal, R. & Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology*, *56*(4), 394–404. https://doi.org/10.1007/s12088-016-0606-4

Archibald, J. (2018). *Genomics: A Very short Introduction*. Oxford University Press.

Ardrey, R. E. (2003). *Liquid Chromatography - Mass Spectrometry: An Introduction*. John Wiley & Sons, Incorporated.

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., Hu, T. T., Jiang, R., Muliyati, N. W., Zhang, X., Amer, M. A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., de Meaux, J., Ecker, J. R., Faure, N., Kniskern, J. M., Jones, J. D. G., Michael, T., Nemri, A., Roux, F., Salt, D. E., Tang, C., Todesco, M., Traw, M. B., Weigel, D., Marjoram, P., Borevitz, J. O., Bergelson, J. & Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature*, *465*(7298), 627–631. https://doi.org/10.1038/nature08800

Babst, B. A., Harding, S. A. & Tsai, C-J. (2010). Biosynthesis of Phenolic Glycosides from Phenylpropanoid and Benzenoid Precursors in *Populus*. *Journal of Chemical Ecology*, *36*(3), 286–297. https://doi.org/10.1007/s10886-010-9757-7

Bailey, J. K., Schweitzer, J. A., Rehill, B. J., Irschick, D. J., Whitham, T. G. & Lindroth, R. L. (2007). Rapid shifts in the chemical composition of aspen forests: an introduced herbivore as an agent of natural selection. *Biological Invasions*, *9*(6), 715–722. https://doi.org/10.1007/s10530-006-9071-z

Bailey, J. K., Schweitzer, J. A., Rehill, B. J., Lindroth, R. L., Martinsen, G. D. & Whitham, T. G. (2004). Beavers as molecular geneticists: A genetic basis to the foraging of an ecosystem engineer. *Ecology*, *85*(3), 603–608. https://doi.org/10.1890/03-3049

Bailey-Wilson, J. E. & Wilson, A. F. (2011). Linkage Analysis in the Next-Generation Sequencing Era. *Human Heredity*, *72*(4), 228–236. https://doi.org/10.1159/000334381

Balandrin, M. F., Klocke, J. A., Wurtele, E. S. & Bollinger, W. H. (1985). Natural Plant Chemicals: Sources of Industrial and Medicinal Materials. *Science*, *228*(4704), 1154–1160. https://doi.org/10.1126/science.3890182

Barbehenn, R. V. & Constabel, C. P. (2011). Tannins in plant-herbivore interactions. *Phytochemistry*, *72*(13), 1551–1565. https://doi.org/10.1016/j.phytochem.2011.01.040

Barker, H. L., Holeski, L. M. & Lindroth, R. L. (2018). Genotypic variation in plant traits shapes herbivorous insect and ant communities on a foundation tree species. *PLoS ONE*, *13*(7), e0200954. https://doi.org/10.1371/journal.pone.0200954

Bateson, W., Saunders, E. R. & Punnett, R. C. (1905). Experimental studies in the physiology of heredity. *Reports to the Evolution Committee of the Royal Society*, *2*(1), 80–99. https://doi.org/10.1007/bf01975751

Becker, J., Wendland, J. R., Haenisch, B., Nöthen, M. M. & Schumacher, J. (2012). A systematic eQTL study of *cis-trans* epistasis in 210 HapMap individuals. *European Journal of Human Genetics*, *20*(1), 97–101. https://doi.org/10.1038/ejhg.2011.156

Benstock, S. E., Weaver, K., Hettema, J. M. & Verhulst, B. (2024). Using Alternative Definitions of Controls to Increase Statistical Power in GWAS. *Behavior Genetics*, *54*(4), 353–366. https://doi.org/10.1007/s10519-024-10187-w

Beresford, M. J. (2010). Medical reductionism: lessons from the great philosophers. *QJM: An International Journal of Medicine*, *103*(9), 721–724. https://doi.org/10.1093/qjmed/hcq057

Bernal-Gallardo, J. J. & de Folter, S. (2024). Plant genome information facilitates plant functional genomics. *Planta*, *259*(5), 117. https://doi.org/10.1007/s00425-024-04397-z

Bhambhani, S., Kondhare, K. R., & Giri, A. P. (2021). Diversity in Chemical Structures and Biological Properties of Plant Alkaloids. *Molecules*, *26*(11), 3374. https://doi.org/10.3390/molecules26113374

Bingman, B. R. & Hart, E. R. (1993). Clonal and Leaf Age Variation in *Populus* Phenolic Glycosides: Implications for Host Selection by *Chrysomela*

scripta (Coleoptera: Chrysomelidae). *Environmental Entomology*, *22*(2), 397–403. https://doi.org/10.1093/ee/22.2.397

Blazics, B., Papp, I. & Kéry, Á. (2010). LC-MS Qualitative Analysis and Simultaneous Determination of Six *Filipendula* Salicylates with Two Standards. *Chromatographia*, *71*(1), 61–67. https://doi.org/10.1365/s10337-010-1502-4

Boeckler, G. A., Gershenzon, J. & Unsicker, S. B. (2011). Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. *Phytochemistry*, *72*(13), 1497–1509. https://doi.org/10.1016/j.phytochem.2011.01.038

Boeckler, G. A., Paetz, C., Feibicke, P., Gershenzon, J. & Unsicker, S. B. (2016). Metabolism of poplar salicinoids by the generalist herbivore *Lymantria dispar* (Lepidoptera). *Insect Biochemistry and Molecular Biology*, *78*, 39-49. https://doi.org/10.1016/j.ibmb.2016.08.001

Boeckler, G. A., Towns, M., Unsicker, S. B., Mellway, R. D., Yip, L., Hilke, I., Gershenzon, J. & Constabel, C. P. (2014). Transgenic upregulation of the condensed tannin pathway in poplar leads to a dramatic shift in leaf palatability for two tree-feeding Lepidoptera. *Journal of Chemical Ecology*, *40*(2), 150–158. https://doi.org/10.1007/s10886-014-0383-7

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boncan, D. A. T., Tsang, S. S. K., Li, C., Lee, I. H. T., Lam, H-M., Chan, T-F. & Hui, J. H. L. (2020). Terpenes and Terpenoids in Plants: Interactions with Environment and Insects. *International Journal of Molecular Sciences*, *21*(19), 7382. https://doi.org/10.3390/ijms21197382

Bontpart, T., Cheynier, V., Ageorges, A. & Terrier, N. (2015). BAHD or SCPL acyltransferase? What a dilemma for acylation in the world of plant phenolic compounds. *New Phytologist*, *208*(3), 695–707. https://doi.org/10.1111/nph.13498

Brachi, B., Morris, G. P. & Borevitz, J. O. (2011). Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biology*, *12*(10), 232. https://doi.org/10.1186/gb-2011-12-10-232

Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. (2005). Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature*, *436*(7051), 701–703. https://doi.org/10.1038/nature03865

Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science*, *296*(5568), 752–755. https://doi.org/10.1126/science.1069516

Broeckling, C. D., Afsar, F. A., Neumann, S., Ben-Hur, A. & Prenni, J. E. (2014). RAMClust: A Novel Feature Clustering Method Enables Spectral-Matching-Based Annotation for Metabolomics Data. *Analytical Chemistry*, *86*(14), 6812–6817. https://doi.org/10.1021/ac501530d

Bryan, C. P. (1930). *The Papyrus Ebers* (Translated from the German version). Geoffrey Bles.

Burrell, R. C. (1937). Phytochemistry. What it is and how it has developed. *Journal of Chemical Education*, *14*(11), *520*–527. https://doi.org/10.1021/ed014p520

Câmara, J. S., Perestrelo, R., Ferreira, R., Berenguer, C. V., Pereira, J. A. M. & Castilho, P. C. (2024). Plant-Derived Terpenoids: A Plethora of Bioactive Compounds with Several Health Functions and Industrial Applications—A Comprehensive Overview. *Molecules*, *29*(16), 3861. https://doi.org/10.3390/molecules29163861

Cesarino, I., Dello Ioio, R., Kirschner, G. K., Ogden, M. S., Picard, K. L., Rast-Somssich, M. I. & Somssich, M. (2020). Plant science's next top models. *Annals of Botany*, *126*(1), 1–23. https://doi.org/10.1093/aob/mcaa063

Chakraborty, P. (2018). Herbal genomics as tools for dissecting new metabolic pathways of unexplored medicinal plants and drug discovery. *Biochimie Open*, *6*, 9–16. https://doi.org/10.1016/j.biopen.2017.12.003

Chang, S., Puryear, J. & Cairney, J. (1993). A Simple and Efficient Method for Isolating RNA From Pine Trees. *Plant Molecular Biology Reporter, 11*(2), 113–116.

Chatterjee, D., Colvin, C., Lesko, T., Peiffer, M., Felton, G. W. & Chopra, S. (2025). Plant defense against insect herbivory: Flavonoid-mediated growth inhibition of *Helicoverpa zea*. *Plant Stress*, *15*, 100738. https://doi.org/10.1016/j.stress.2025.100738

Chedgy, R. J., Köllner, T. G. & Constabel, C. P. (2015). Functional characterization of two acyltransferases from *Populus trichocarpa* capable of synthesizing benzyl benzoate and salicyl benzoate, potential intermediates in salicinoid phenolic glycoside biosynthesis. *Phytochemistry*, *113*, 149–159. https://doi.org/10.1016/j.phytochem.2014.10.018

Chen, S., Wang, X., Cheng, Y., Gao, H. & Chen, X. (2023). A Review of Classification, Biosynthesis, Biological Activities and Potential Applications of Flavonoids. *Molecules*, *28*(13), 4982. https://doi.org/10.3390/molecules28134982

Chien, L-C. (2020). A rank-based normalization method with the fully adjusted full-stage procedure in genetic association studies. *PLoS ONE, 15*(6), e0233847. https://doi.org/10.1371/journal.pone.0233847

Civelek, M. & Lusis, A. J. (2014). Systems genetics approaches to understand complex traits. *Nature Reviews Genetics*, *15*(1), 34–48. https://doi.org/10.1038/nrg3575

Clausen, T. P., Reichardt, P. B., Bryant, J. P., Werner, R. A., Post, K. & Frisby, K. (1989). Chemical model for short-term induction in quaking aspen (*Populus tremuloides*) foliage against herbivores. *Journal of Chemical Ecology, 15*(9), 2335-2346. https://doi.org/10.1007/BF01012085

Cole, C. T., Morrow, C. J., Barker, H. L., Rubert-Nason, K. F., Riehl, J. F. L., Köllner, T. G., Lackus, N. D. & Lindroth, R. L. (2021). Growing up aspen: ontogeny and trade-offs shape growth, defence and reproduction in a foundation species. *Annals of Botany*, *127*(4), 505–517. https://doi.org/10.1093/aob/mcaa070

Coleman, H. D., Park, J-Y., Nair, R., Chapple, C. & Mansfield, S. D. (2008). RNAi-mediated suppression of *p*-coumaroyl-CoA 3′-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4501–4506. https://doi.org/10.1073/pnas.0706537105

Coleman, J. S. (1986). Leaf development and leaf stress: increased susceptibility associated with sink-source transition. *Tree Physiology*, *2*, 289–299. https://doi.org/10.1093/treephys/2.1-2-3.289

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L. & Pericak-Vance, M. A. (1993). Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset Families. *Science*, *261*(5123), 921–923. https://doi.org/10.1126/science.8346443

Corlett, R. T. (2016). Plant diversity in a changing world: Status, trends, and conservation needs. *Plant Diversity*, *38*(1), 10–16. https://doi.org/10.1016/j.pld.2016.01.001

Cragg, G. M. & Newman, D. J. (2013). Natural products: A continuing source of novel drug leads. *Biochimica et Biophysica Acta*, *1830*(6), 3670–3695. https://doi.org/10.1016/j.bbagen.2013.02.008

D'Auria, J. C. (2006). Acyltransferases in plants: a good time to be BAHD. *Current Opinion in Plant Biology*, *9*(3), 331–340. https://doi.org/10.1016/j.pbi.2006.03.016

Dayan, A. D. (2024). Death of Socrates: a likely case of poison hemlock (*Conium maculatum*) poisoning. *Clinical Toxicology*, *62*(1), 56-60. https://doi.org/10.1080/15563650.2024.2309328

Dawkins, R. (1982). *The Extended Phenotype: The Long Reach of the Gene*. Oxford University Press.

De Luca, V., Salim, V., Atsumi, M. S & Yu, F. (2012). Mining the Biodiversity of Plants: A Revolution in the Making. *Science*, *336*(6089), 1658–1661. https://doi.org/10.1126/science.1217410

De Resende, M. D. V., Silva, F. F., Resende, M. F. R. & Azevedo, C. F. (2014). Genome-Wide Association Studies (GWAS). In Borém, A. & Fritsche-Neto, R. (eds.). *Biotechnology and Plant Breeding: Applications and Approaches for Developing Improved Cultivars*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-418672-9.00004-0

Delker, R.K. & Mann, R.S. (2017). From Reductionism to Holism: Toward a More Complete View of Development Through Genome Engineering. In Tsang, S.H. (eds.). *Precision Medicine, CRISPR and Genome Engineering: Moving from Association to Biology and Therapeutics*. Springer International Publishing.

Deng, Y. & Lu, S. (2017). Biosynthesis and Regulation of Phenylpropanoids in Plants. *Critical Reviews in Plant Sciences*, *36*(4), 257–290. https://doi.org/10.1080/07352689.2017.1402852

Deshpande, D., Chhugani, K., Chang, Y., Karlsberg, A., Loeffler, C., Zhang, J., Muszyńska, A., Munteanu, V., Yang, H., Rotman, J., Tao, L., Balliu, B., Tseng, E., Eskin, E., Zhao, F., Mohammadi, P., Łabaj, P. P. & Mangul, S. (2023). RNA-seq data science: From raw data to effective interpretation. *Frontiers in Genetics*, *14*, 997383. https://doi.org/10.3389/fgene.2023.997383

Dietrich, W. F., Copeland, N. G., Gilbert, D. J., Miller, J. C., Jenkins, N. A. & Lander, E. S. (1995). Mapping the mouse genome: current status and future prospects. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(24), 10849–10853. https://doi.org/10.1073/pnas.92.24.10849

Dixon, R. A. & Barros, J. (2019). Lignin biosynthesis: old roads revisited and new roads explored. *Open Biology*, *9*(12), 190215. https://doi.org/10.1098/rsob.190215

Doğan, H. O. (2024). Metabolomics: a review of liquid chromatography mass spectrometry-based methods and clinical applications. *Turkish Journal of Biochemistry*, *49*(1), 1–14. https://doi.org/10.1515/tjb-2023-0095

Donaldson, J. R. & Lindroth, R. L. (2004). Cottonwood Leaf Beetle (Coleoptera: Chrysomelidae) Performance in Relation to Variable Phytochemistry in Juvenile Aspen (*Populus tremuloides* Michx.). *Environmental Entomology*, *33*(5), 1505–1511. https://doi.org/10.1603/0046-225X-33.5.1505

Donaldson, J. R. & Lindroth, R. L. (2007). Genetics, Environment, and Their Interaction Determine Efficacy of Chemical Defense in Trembling Aspen. *Ecology, 88*(3), 729-739. https://doi.org/10.1890/06-0064

Donaldson, J. R., Stevens, M. T., Barnhill, H. R. & Lindroth, R. L (2006). Age-Related Shifts in Leaf Chemistry of Clonal Aspen (*Populus tremuloides*). *Journal of Chemical Ecology*, *32*(7), 1415–1429. https://doi.org/10.1007/s10886-006-9059-2

Dong, N. Q. & Lin, H-X. (2021). Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *Journal of*

Integrative Plant Biology, *63*(1), 180–209. https://doi.org/10.1111/jipb.13054

Dong, Z. & Chen, Y. (2013). Transcriptomics: Advances and approaches. *Science China Life Sciences*, *56*(10), 960–967. https://doi.org/10.1007/s11427-013-4557-2

Dötterl, S., Jürgens, A., Seifert, K., Laube, T., Weißbecker, B. & Schütz, S. (2006). Nursery pollination by a moth in *Silene latifolia*: The role of odours in eliciting antennal and behavioural responses. *New Phytologist*, *169*(4), 707–718. https://doi.org/10.1111/j.1469-8137.2005.01509.x

Drost, D. R., Benedict, C. I., Berg, A., Novaes, E., Novaes, C. R. D. B., Yu, Q., Dervinis, C., Maia, J. M., Yap, J., Miles, B. & Kirst, M. (2010). Diversification in the genetic architecture of gene expression and transcriptional networks in organ differentiation of *Populus*. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(18), 8492–8497. https://doi.org/10.1073/pnas.0914709107

Drost, D. R., Puranik, S., Novaes, E., Novaes, C. R. D. B., Dervinis, C., Gailing, O. & Kirst, M. (2015). Genetical genomics of *Populus* leaf shape variation. *BMC Plant Biology*, *15*, 166. https://doi.org/10.1186/s12870-015-0557-7

Durán-Medina, Y., Ruiz-Cortés, B. E., Guerrero-Largo, H. & Marsch-Martínez, N. (2021). Specialized metabolism and development: An unexpected friendship. *Current Opinion in Plant Biology*, *64*, 102142. https://doi.org/10.1016/j.pbi.2021.102142

Elshafie, H. S., Camele, I. & Mohamed, A. A. (2023). A Comprehensive Review on the Biological, Agricultural and Pharmaceutical Properties of Secondary Metabolites Based-Plant Origin. *International Journal of Molecular Sciences*, *24*(4), 3266. https://doi.org/10.3390/ijms24043266

Emwas, A-H., Szczepski, K., McKay, R. T., Asfour, H., Chang, C-k., Lachowicz, J. & Jaremko, M. (2021). Pharmacometabolomics: A New Horizon in Personalized Medicine. In Zhan, X. (eds.). *Metabolomics: Methodology and Applications in Medical Sciences and Life Sciences*. IntechOpen, pp. 31-65.

Espinosa, E., Bautista, R., Larrosa, R. & Plata, O. (2024). Advancements in long-read genome sequencing technologies and algorithms. *Genomics*, *116*(3), 110842. https://doi.org/10.1016/j.ygeno.2024.110842

Evans, C., Hardin, J. & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, *19*(5), 776–792. https://doi.org/10.1093/bib/bbx008

Facchini, P. J. (2001). Alkaloid biosynthesis in plants: Biochemistry, cell biology, molecular regulation, and metabolic engineering applications. *Annual Review of Plant Biology*, *52*, 29–66. https://doi.org/10.1146/annurev.arplant.52.1.29

Faisal, S., Badshah, S. L., Kubra, B., Emwas, A-H. & Jaremko, M. (2023). Alkaloids as potential antivirals. A comprehensive review. *Natural Products and Bioprospecting*, *13*, 4. https://doi.org/10.1007/s13659-022-00366-9

Fellenberg, C., Corea, O., Yan, L-H., Archinuk, F., Piirtola, E-M., Gordon, H., Reichelt, M., Brandt, W., Wulff, J., Ehlting, J. & Constabel, C. P. (2020). Discovery of salicyl benzoate UDP-glycosyltransferase, a central enzyme in poplar salicinoid phenolic glycoside biosynthesis. *The Plant Journal*, *102*(1), 99-115. https://doi.org/10.1111/tpj.14615

Fields, M. J. & Orians, C. M. (2006). Specificity of Phenolic Glycoside Induction in Willow Seedlings (*Salix sericea*) in Response to Herbivory. *Journal of Chemical Ecology*, *32*(12), 2647–2656. https://doi.org/10.1007/s10886-006-9188-7

Fox, J. & Weisberg, S. (2019). An R Companion to Applied Regression, Third edition. Sage, Thousand Oaks CA. https://www.john-fox.ca/Companion/

Fu, R., Martin, C. & Zhang, Y. (2018). Next-Generation Plant Metabolic Engineering, Inspired by an Ancient Chinese Irrigation System. *Molecular Plant*, *11*(1), 47–57. https://doi.org/10.1016/j.molp.2017.09.002

Giani, A. M., Gallo, G. R., Gianfranceschi, L. & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, *18*, 9–19. https://doi.org/10.1016/j.csbj.2019.11.002

Gordon, H., Fellenberg, C., Lackus, N. D., Archinuk, F., Sproule, A., Nakamura, Y., Köllner, T. G., Gershenzon, J., Overy, D. P. & Constabel, C. P. (2022). CRISPR/Cas9 disruption of *UGT71L1* in poplar connects salicinoid and salicylic acid metabolism and alters growth and morphology. *The Plant Cell*, *34*(8), 2925-2947. https://doi.org/10.1093/plcell/koac135

Gould, G. G., Jones, C. G., Rifleman, P., Perez, A. & Coleman, J. S. (2007). Variation in Eastern Cottonwood (*Populus deltoides* Bartr.) Phloem Sap Content Caused by Leaf Development May Affect Feeding Site Selection Behavior of the Aphid, *Chaitophorous populicola* Thomas (Homoptera: Aphididae). *Environmental Entomology*, *36*(5), 1212–1225. https://doi.org/10.1603/0046-225X(2007)36[1212:VIECPD]2.0.CO;2

Grahn, Erik. (2014*). 'n Erik deekt oppa boonska*. Ord & Visor förslag.

Guo, Y., Zhao, S., Sheng, Q., Guo, M., Lehmann, B., Pietenpol, J., Samuels, D. C. & Shyr, Y. (2015). RNAseq by Total RNA Library Identifies Additional RNAs Compared to Poly(A) RNA Library. *BioMed Research International*, *2015*, 862130. https://doi.org/10.1155/2015/862130

Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y., Young, A. B., Shoulson, I., Bonilla, E. & Martin, J. B. (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, *306*(5940), 234–238. https://doi.org/10.1038/306234a0

Hartmann, T. (2007). From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry*, *68*(22–24), 2831–2846. https://doi.org/10.1016/j.phytochem.2007.09.017

Hartmann, T. (2008). The lost origin of chemical ecology in the late 19th century. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4541–4546. https://doi.org/10.1073/pnas.0709231105

Heather, J. M. & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Heiska, S., Tikkanen, O-P., Rousi, M. & Julkunen-Tiitto, R. (2007). Bark salicylates and condensed tannins reduce vole browsing amongst

cultivated dark-leaved willows (*Salix myrsinifolia*). *Chemoecology*, *17*(4), 245–253. https://doi.org/10.1007/s00049-007-0385-9

Hemming, J. D. C. & Lindroth, R. L. (1995). Intraspecific variation in aspen phytochemistry: effects on performance of gypsy moths and forest tent caterpillars. *Oecologia*, *103*(1), 79–88. https://doi.org/10.1007/BF00328428

Herbert, J., Thompson, S., Beckett, A. H. & Robson, S. C. (2025). Impact of microbiological molecular methodologies on adaptive sampling using nanopore sequencing in metagenomic studies. *Environmental Microbiome*, *20*, 47. https://doi.org/10.1186/s40793-025-00704-7

Heroy, K. Y., St. Clair, S. B. S., Burritt, E. A., Durham, S. L. & Villalba, J. J. (2018). Nutritional state and secondary compounds influence trembling aspen (*Populus tremuloides*) intake by lambs. *Applied Animal Behaviour Science*, *206*, 39–51. https://doi.org/10.1016/j.applanim.2018.05.031

Hirschhorn, J. N. & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, *6*(2), 95–108. https://doi.org/10.1038/nrg1521

Holeski, L. M., McKenzie, S. C., Kruger, E. L., Couture, J. J., Rubert-Nason, K. & Lindroth, R. L. (2016). Phytochemical traits underlie genotypic variation in susceptibility of quaking aspen (*Populus tremuloides*) to browsing by a keystone forest ungulate. *Journal of Ecology*, *104*(3), 850–863. https://doi.org/10.1111/1365-2745.12559

Holley, R.W. (1968). *Alanine transfer RNA*. Nobel Lecture, 12 December 1968. https://www.nobelprize.org/prizes/medicine/1968/holley/lecture/ [Accessed 2025-10-8].

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. & Zamir, A. (1965). Structure of a Ribonucleic Acid. *Science*, *147*(3664), 1462–1465. https://doi.org/10.1126/science.147.3664.1462

Horgan, R. P. & Kenny, L. C. (2011). 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, *13*(3), 189–195. https://doi.org/10.1576/toag.13.3.189.27672

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E. S., Qian, Q., Zhang, Q-F., Li, J. & Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, *42*(11), 961-967. https://doi.org/10.1038/ng.695

Hwang, S-Y. & Lindroth, R. L. (1997). Clonal variation in foliar chemistry of aspen: effects on gypsy moths and forest tent caterpillars. *Oecologia*, *111*(1), 99–108. https://doi.org/10.1007/s004420050213

Idle, J. R. & Gonzalez, F. J. (2007). Metabolomics. *Cell Metabolism*, *6*(5), 348–351. https://doi.org/10.1016/j.cmet.2007.10.005

Ikegawa, S. (2012). A Short History of the Genome-Wide Association Study: Where We Were and Where We Are Going. *Genomics & Informatics*, *10*(4), 220–225. https://doi.org/10.5808/GI.2012.10.4.220

Ikonen, A., Tahvanainen, J. & Roininen, H. (2002). Phenolic secondary compounds as determinants of the host plant preferences of the leaf beetle, *Agelastica alni*. *Chemoecology*, *12*(3), 125–131. https://doi.org/10.1007/s00012-002-8337-2

Ikram, N. K. B. K. & Simonsen, H. T. (2017). A Review of Biotechnological Artemisinin Production in Plants. *Frontiers in Plant Science*, *8*, 1966. https://doi.org/10.3389/fpls.2017.01966

Ingvarsson, P. K., Hvidsten, T. R. & Street, N. R. (2016). Towards integration of population and comparative genomics in forest trees. *New Phytologist*, *212*(2), 338–344. https://doi.org/10.1111/nph.14153

Ingvarsson, P. K. & Street, N. R. (2011). Association genetics of complex traits in plants. *New Phytologist*, *189*(4), 909-922. https://doi.org/10.1111/j.1469-8137.2010.03593.x

Ishikawa, T., Nishigaya, K., Takami, K., Uchikoshi, H., Chen, I-S. & Tsai, I-L. (2004). Isolation of Salicin Derivatives from *Homalium cochinchinensis* and Their Antiviral Activities. *Journal of Natural Products*, *67*(4), 659–663. https://doi.org/10.1021/np0340520

Jansen, R. C. & Nap, J-P. (2001). Genetical genomics: the added value from segregation. *Trends in Genetics*, *17*(7), 388–391. https://doi.org/10.1016/S0168-9525(01)02310-1

Jansen, R. C., Tesson, B. M., Fu, J., Yang, Y. & McIntyre, L. M. (2009). Defining gene and QTL networks. *Current Opinion in Plant Biology*, *12*(2), 241–246. https://doi.org/10.1016/j.pbi.2009.01.003

Jansson, S. & Douglas, C. J. (2007). *Populus*: A Model System for Plant Biology. *Annual Review of Plant Biology*, *58*, 435–458. https://doi.org/10.1146/annurev.arplant.58.032806.103956

Ji, W., Osbourn, A. & Liu, Z. (2024). Understanding metabolic diversification in plants: branchpoints in the evolution of specialized metabolism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *379*(1914), 20230359. https://doi.org/10.1098/rstb.2023.0359

Jiang, G. (2017). Molecular Markers. In Thomas, B., Murray, B. G. & Murphy, D. J. (eds.). *Encyclopedia of Applied Plant Sciences*. 2nd ed. Elsevier Inc. pp. 207-214. https://doi.org/10.1016/B978-0-12-394807-6.00193-3

Julkunen-Tiitto, R. & Virjamo, V. (2017). *Biosynthesis and roles of Salicaceae salicylates*. In Arimura, G-I. & Maffei, M. (eds.). *Plant Specialized Metabolism: Genomics, Biochemistry, and Biological Functions*. CRC Press, pp. 65-93.

Kadarmideen, H. N., Von Rohr, P. & Janss, L. L. G. (2006). From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. *Mammalian Genome*, *17*(6), 548–564. https://doi.org/10.1007/s00335-005-0169-x

Keefover-Ring, K., Ahnlund, M., Abreu, I. N., Jansson, S., Moritz, T. & Albrectsen, B. R. (2014). No evidence of geographical structure of salicinoid chemotypes within *Populus tremula*. *PLoS ONE*, *9*(10), e107189. https://doi.org/10.1371/journal.pone.0107189

Keurentjes, J. J. B., Fu, J., Terpstra, I. R., Garcia, J. M., Van Den Ackerveken, G., Snoek, L. B., Peeters, A. J. M., Vreugdenhil, D., Koornneef, M. & Jansen, R. C. (2007). Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(5), 1708–1713. https://doi.org/10.1073/pnas.0610429104

Khanal, R., Navabi, A. & Lukens, L. (2015). Linkage map construction and quantitative trait loci (QTL) mapping using intermated vs. selfed recombinant inbred maize line (*Zea mays* L.). *Canadian Journal of Plant Science*, *95*(6), 1133–1144. https://doi.org/10.4141/CJPS-2015-091

Kilani-Morakchi, S., Morakchi-Goudjil, H. & Sifi, K. (2021). Azadirachtin-Based Insecticide: Overview, Risk Assessments, and Future Directions. *Frontiers in Agronomy*, *3*, 676208. https://doi.org/10.3389/fagro.2021.676208

Kim, C. S., Subedi, L., Park, K. J., Kim, S. Y., Choi, S. U., Kim, K. H. & Lee, K. R. (2015). Salicin derivatives from *Salix glandulosa* and their biological activities. *Fitoterapia*, *106*, 147-152. https://doi.org/10.1016/j.fitote.2015.08.013

Kim, H. J., Jang, Y. S., Ha, J. W., Ra, M-J., Jung, S-M., Yu, J-N., Kim, K., Kim, K. H. & Um, S. H. (2022). Bioactive Phytochemicals from *Salix pseudolasiogyne* Twigs: Anti-Adipogenic Effect of 2′-O-Acetylsalicortin in 3T3-L1 Cells. *International Journal of Molecular Sciences*, *23*(19), 12006. https://doi.org/10.3390/ijms231912006

Kirst, M., Basten, C. J., Myburg, A. A., Zeng, Z-B. & Sederoff, R. R. (2005). Genetic Architecture of Transcript-Level Variation in Differentiating Xylem of a Eucalyptus Hybrid. *Genetics, 169*(4), 2295-2303. https://doi.org/10.1534/genetics.104.039198

Kitano, H. (2002). Systems Biology: A Brief Overview. *Science*, *295*(5560), 1662–1664. https://doi.org/10.1126/science.1069492

Knoch, D., Riewe, D., Meyer, R. C., Boudichevskaia, A., Schmidt, R. & Altmann, T. (2017). Genetic dissection of metabolite variation in Arabidopsis seeds: evidence for mQTL hotspots and a master regulatory locus of seed metabolism. *Journal of Experimental Botany*, *68*(7), 1655–1667. https://doi.org/10.1093/jxb/erx049

Kolehmainen, J., Roininen, H., Julkunen-Tiitto, R. & Tahvanainen, J. (1994). Importance of Phenolic Glucosides in Host Selection of Shoot Galling Sawfly. *Journal of Chemical Ecology*, *20*(9), 2455–2466. https://doi.org/10.1007/BF02033213

Koornneef, M. & Meinke, D. (2010). The development of Arabidopsis as a model plant. *Plant Journal*, *61*(6), 909–921. https://doi.org/10.1111/j.1365-313X.2009.04086.x

Köpf, A., Rank, N. E., Roininen, H. & Tahvanainen, J. (1997). Defensive larval secretions of leaf beetles attract a specialist predator *Parasyrphus nigritarsis*. *Ecological Entomology*, *22*(2), 176–183. https://doi.org/10.1046/j.1365-2311.1997.t01-1-00061.x

Kopylova, E., Noé, L. & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Kulasekaran, S., Cerezo-Medina, S., Harflett, C., Lomax, C., De Jong, F., Rendour, A., Ruvo, G., Hanley, S. J., Beale, M. H. & Ward, J. L. (2021). A willow UDP-glycosyltransferase involved in salicinoid biosynthesis. *Journal of Experimental Botany*, *72*(5), 1634–1648. https://doi.org/10.1093/jxb/eraa562

Kwon, D-J., Bae, Y-S., Ju, S. M., Youn, G. S., Choi, S. Y. & Park, J. (2014). Salicortin suppresses lipopolysaccharide-stimulated inflammatory responses via blockade of NF-κB and JNK activation in RAW 264.7 macrophages. *BMB Reports*, *47*(6), 318–323. https://doi.org/10.5483/BMBRep.2014.47.6.200

Lackus, N. D., Müller, A., Kröber, T. D. U., Reichelt, M., Schmidt, A., Nakamura, Y., Paetz, C., Luck, K., Lindroth, R. L., Constabel, C. P., Unsicker, S. B., Gershenzon, J. & Köllnera, T. G. (2020). The Occurrence of Sulfated Salicinoids in Poplar and Their Formation by Sulfotransferase1. *Plant Physiology*, *183*(1), 137–151. https://doi.org/10.1104/pp.19.01447

Langfelder, P. & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559. https://doi.org/10.1186/1471-2105-9-559

Le Roy, J., Huss, B., Creach, A., Hawkins, S. & Neutelings, G. (2016). Glycosylation Is a Major Regulator of Phenylpropanoid Availability and Biological Activity in Plants. *Frontiers in Plant Science*, *7*, 735. https://doi.org/10.3389/fpls.2016.00735

Li, Y., Li, Z., Xu, C. & Wang, Q. (2025). WRKYs as regulatory hubs of secondary metabolic networks: Diverse inducers and distinct responses. *Plant Communications*, *6*(9), 101438. https://doi.org/10.1016/j.xplc.2025.101438

Lin, J., Sibley, A., Shterev, I., Nixon, A., Innocenti, F. Chan, C. & Owzar, K. (2019). FastJT: An R package for robust and efficient feature selection for machine learning and genome-wide association studies. *BMC Bioinformatics*, *20*, 333. https://doi.org/10.1186/s12859-019-2869-3

Lindroth, R. L. & Peterson, S. S. (1988). Effects of plant phenols on performance of southern armyworm larvae. *Oecologia, 75*(2), 185-189. https://doi.org/10.1007/BF00378595

Lindroth, R. L., Scriber, J. M. & Hsia, M. T. S. (1988). Chemical Ecology of the Tiger Swallowtail: Mediation of Host Use by Phenolic Glycosides. *Ecology*, *69*(3), 814–822. https://doi.org/10.2307/1941031

Lisec, J., Meyer, R. C., Steinfath, M., Redestig, H., Becher, M., Witucka-Wall, H., Fiehn, O., Törjék, O., Selbig, J., Altmann, T. & Willmitzer, L. (2008). Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant Journal*, *53*(6), 960–972. https://doi.org/10.1111/j.1365-313X.2007.03383.x

Liu, J., Osbourn, A. & Ma, P. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Molecular Plant*, *8*(5), 689–708. https://doi.org/10.1016/j.molp.2015.03.012

Love, M. I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Luquez, V., Hall, D., Albrectsen, B. R., Karlsson, J., Ingvarsson, P., & Jansson, S. (2008). Natural phenological variation in aspen (*Populus tremula*): The SwAsp collection. *Tree Genetics and Genomes*, *4*(2), 279–292. https://doi.org/10.1007/s11295-007-0108-y

Mackay, T. F. C. (2004). The genetic architecture of quantitative traits: lessons from *Drosophila*. *Current Opinion in Genetics & Development*, *14*(3), 253–257. https://doi.org/10.1016/j.gde.2004.04.003

MacLagan, T. (1876). The Treatment of Rheumatism by Salicin and Salicylic Acid. *The British Medical Journal*, *1*(803), 627. https://doi.org/10.1136/bmj.1.803.627

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnek, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A. & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature, 461*(7265), 747-753. https://doi.org/10.1038/nature08494

March, R. E. (1999). Gene mapping by linkage and association analysis. *Molecular Biotechnology*, *13*(2), 113–122. https://doi.org/10.1385/MB:13:2:113

Martin, J., Schackwitz, W. & Lipzen, A. (2018). Genomic Sequence Variation Analysis by Resequencing. In de Vries R. P., Tsang, A. & Grigoriev, I. (eds.). *Fungal Genomics: Methods and Protocols*. 2nd ed. Humana Press. https://doi.org/10.1007/978-1-4939-7804-5_18

Martinsen, G. D., Driebe, E. M. & Whitham, T. G. (1998). Indirect interactions mediated by changing plant chemistry: Beaver browsing benefits beetles. *Ecology*, *79*(1), 192–200. https://doi.org/10.1890/0012-9658(1998)079[0192:IIMBCP]2.0.CO;2

Mason, C. J., Rubert-Nason, K. F., Lindroth, R. L. & Raffa, K. F. (2015). Aspen Defense Chemicals Influence Midgut Bacterial Community Composition of Gypsy Moth. *Journal of Chemical Ecology*, *41*(1), 75–84. https://doi.org/10.1007/s10886-014-0530-1

Mason, C. J., Rubert-Nason, K., Lindroth, R. L., Shi, J. & Hoover, K. (2021). Salicinoid phenolics reduce adult *Anoplophora glabripennis* (Cerambicidae: Lamiinae) feeding and egg production. *Arthropod-Plant Interactions*, *15*(1), 127–136. https://doi.org/10.1007/s11829-020-09802-4

Massad, T. J., Trumbore, S. E., Ganbat, G., Reichelt, M., Unsicker, S., Boeckler, A., Gleixner, G., Gershenzon, J. & Ruehlow, S. (2014). An optimal defense strategy for phenolic glycoside production in *Populus trichocarpa* – isotope labeling demonstrates secondary metabolite production in growing leaves. *New Phytologist*, *203*(2), 607–619. https://doi.org/10.1111/nph.12811

Masyita, A., Mustika Sari, R., Dwi Astuti, A., Yasir, B., Rahma Rumata, N., Emran, T. B., Nainu, F. & Simal-Gandara, J. (2022). Terpenes and terpenoids as main bioactive compounds of essential oils, their roles in human health and potential application as natural food preservatives. *Food Chemistry: X*, *13*, 100217. https://doi.org/10.1016/j.fochx.2022.100217

Mazzocchi, F. (2012). Complexity and the reductionism-holism debate in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *4*(5), 413–427. https://doi.org/10.1002/wsbm.1181

McGettigan, P. A. (2013). Transcriptomics in the RNA-seq era. *Current Opinion in Chemical Biology*, *17*(1), 4–11. https://doi.org/10.1016/j.cbpa.2012.12.008

Mellway, R. D., Tran, L. T., Prouse, M. B., Campbell, M. M. & Constabel, C. P. (2009). The Wound-, Pathogen-, and Ultraviolet B-Responsive *MYB134* Gene Encodes an R2R3 MYB Transcription Factor That Regulates Proanthocyanidin Synthesis in Poplar. *Plant Physiology*, *150*(2), 924–941. https://doi.org/10.1104/pp.109.139071

Mithöfer, A. & Boland, W. (2012). Plant Defense Against Herbivores: Chemical Aspects. *Annual Review of Plant Biology*, *63*, 431–450. https://doi.org/10.1146/annurev-arplant-042110-103854

Mock, A., Warta, R., Dettling, S., Brors, B., Jäger, D. & Herold-Mende, C. (2018). MetaboDiff: an R package for differential metabolomic analysis. *Bioinformatics*, *34*(19), 3417–3418. https://doi.org/10.1093/bioinformatics/bty344

Morgan, T. H. (1910). Sex Limited Inheritance in *Drosophila*. *Science*, *32*(812), 120–122. https://doi.org/10.1126/science.32.812.120

Morgan, T. H. (1911a). The Application of the Conception of Pure Lines to Sex-Limited Inheritance and to Sexual Dimorphism. *The American Naturalist*, *45*(530), 65–78. https://doi.org/10.1086/279195

Morgan, T. H. (1911b). Random Segregation Versus Coupling in Mendelian Inheritance. *Science*, *34*(873), 384. https://doi.org/10.1126/science.34.873.384

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S. & Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature*, *430*(7001), 743–747. https://doi.org/10.1038/nature02797

Muhlemann, J. K., Younts, T. L. B. & Muday, G. K. (2018). Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(47), E11188–E11197. https://doi.org/10.1073/pnas.1811492115

Myburg, A. A., Hussey, S. G., Wang, J. P., Street, N. R. & Mizrachi, E. (2019). Systems and Synthetic Biology of Forest Trees: A Bioengineering Paradigm for Woody Biomass Feedstocks. *Frontiers in Plant Science*, *10,* 775. https://doi.org/10.3389/fpls.2019.00775

Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E. & Buckler, E. S. (2009). Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *The Plant Cell*, *21*(8), 2194–2202. https://doi.org/10.1105/tpc.109.068437

Mähler, N., Schiffthaler, B., Robinson, K. M., Terebieniec, B., Vučak, M., Mannapperuma, C., Bailey, M. E. S., Jansson, S., Hvidsten, T. R. & Street, N. R. (2020). Leaf shape in *Populus tremula* is a complex, omnigenic trait. *Ecology and Evolution,* 10(21), 11922-11940. https://doi.org/10.1002/ece3.6691

Mähler, N., Wang, J., Terebieniec, B. K., Ingvarsson, P. K., Street, N. R. & Hvidsten, T. R. (2017). Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genetics*, 13(4), e1006402. https://doi.org/10.1371/journal.pgen.1006402

Nadeau, J. H. & Dudley, A. M. (2011). Systems Genetics. *Science*, *331*(6020), 1015-1016. https://doi.org/10.1126/science.1203869

Narberhaus, I., Zintgraf, V. & Dobler, S. (2005). Pyrrolizidine alkaloids on three trophic levels – evidence for toxic and deterrent effects on phytophages and predators. *Chemoecology*, *15*, 121–125. https://doi.org/10.1007/s00049-005-0302-z

Nasim, N., Sandeep, I. S. & Mohanty, S. (2022). Plant-derived natural products for drug discovery: current approaches and prospects. *The Nucleus*, *65*(3), 399–411. https://doi.org/10.1007/s13237-022-00405-3

Nawrot-Chorabik, K., Sułkowska, M. & Gumulak, N. (2022). Secondary Metabolites Produced by Trees and Fungi: Achievements So Far and Challenges Remaining. *Forests*, *13*(8), 1338. https://doi.org/10.3390/f13081338

Nepovimova, E. & Kuca, K. (2019). The history of poisoning: from ancient times until modern ERA. *Archives of Toxicology*, *93*(1), 11–24. https://doi.org/10.1007/s00204-018-2290-0

Ninkuu, V., Zhang, L., Yan, J., Fu, Z., Yang, T. & Zeng, H. (2021). Biochemistry of Terpenes and Recent Advances in Plant Protection. *International Journal of Molecular Sciences*, *22*(11), 5710. https://doi.org/10.3390/ijms22115710

Nyman, T. & Julkunen-Tiitto, R. (2000). Manipulation of the phenolic chemistry of willows by gall-inducing sawflies. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(24), 13184–13187. https://doi.org/10.1073/pnas.230294097

Ogutcen, E., de Lima Ferreira, P., Wagner, N. D., Marinček, P., Vir Leong, J., Aubona, G., Cavender-Bares, J., Michálek, J., Schroeder, L., Sedio, B. E., Vašut, R. J. & Volf, M. (2024). Phylogenetic insights into the Salicaceae: The evolution of willows and beyond. *Molecular Phylogenetics and Evolution*, *199*, 108161. https://doi.org/10.1016/j.ympev.2024.108161

Osier, T. L., Hwang, S-Y. & Lindroth, R. L. (2000). Effects of phytochemical variation in quaking aspen *Populus tremuloides* clones on gypsy moth *Lymantria dispar* performance in the field and laboratory. *Ecological Entomology*, *25*(2), 197–207. https://doi.org/10.1046/j.1365-2311.2000.00245.x

Osier, T. L. & Lindroth, R. L. (2001). Effects of Genotype, Nutrient Availability, and Defoliation on Aspen Phytochemistry and Insect Performance. *Journal of Chemical Ecology*, *27*(7), 1289–1313. https://doi.org/10.1023/A:1010352307301

Osier, T. L. & Lindroth, R. L. (2006). Genotype and environment determine allocation to and costs of resistance in quaking aspen. *Oecologia*, *148*(2), 293–303. https://doi.org/10.1007/s00442-006-0373-8

Ovbude, S. T., Sharmeen, S., Kyei, I., Olupathage, H., Jones, J., Bell, R. J., Powers, R. & Hage, D. S. (2024). Applications of chromatographic methods in metabolomics: A review. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *1239*, 124124. https://doi.org/10.1016/j.jchromb.2024.124124

Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Sato, H., Hori, M., Nakamura, Y. & Tanaka, T. (2002). Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genetics, 32*(4), 650-654. https://doi.org/10.1038/ng1047

Palo, R. T. (1984). Distribution of birch (*Betula* SPP.), willow (*Salix* SPP.), and poplar (*Populus* SPP.) secondary metabolites and their potential role as chemical defense against herbivores. *Journal of Chemical Ecology*, *10*(3), 499–520. https://doi.org/10.1007/BF00988096

Parasuraman, S., Anish, R., Balamurugan, S., Muralidharan, S., Jayaraj Kumar, K. & Vijayan, V. (2014). An Overview of Liquid Chromatography-Mass Spectroscopy Instrumentation. *Pharmaceutical Methods*, *5*(2), 47–55. https://doi.org/10.5530/phm.2014.2.2

Park, Y-B. & Kim, J-M. (2023). Identification of long non-coding RNA-mRNA interactions and genome-wide lncRNA annotation in animal transcriptome profiling. *Journal of Animal Science and Technology*, *65*(2), 293–310. https://doi.org/10.5187/jast.2023.e17

Pasteels, J. M., Daloze, D. & Rowell-Rahier, M. (1986). Chemical defence in chrysomelid eggs and neonate larvae. *Physiological Entomology*, *11*(1), 29–37. https://doi.org/10.1111/j.1365-3032.1986.tb00388.x

Pasteels, J. M., Rowell-Rahier, M., Breakman, J. C. & Dupont, A. (1983). Salicin from host plant as precursor of salicylaldehyde in defensive secretion of Chrysomeline larvae. *Physiological Entomology*, *8*(3), 307–314. https://doi.org/10.1111/j.1365-3032.1983.tb00362.x

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Pazhamala, L. T., Kudapa, H., Weckwerth, W., Millar, A. H. & Varshney, R. K. (2021). Systems biology for crop improvement. *Plant Genome*, *14*(2), e20098. https://doi.org/10.1002/tpg2.20098

Perez de Souza, L., Alseekh, S., Scossa, F. & Fernie, A. R. (2021). Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. *Nature Methods*, *18*(7), 733–746. https://doi.org/10.1038/s41592-021-01116-4

Peterson, R. A. (2021). Finding Optimal Normalizing Transformations via bestNormalize. *The R Journal*, **13**(1), 310–329. https://doi.org/10.32614/RJ-2021-041

Philippe, R. N. & Bohlmann, J. (2007). Poplar defense against insect herbivores. *Canadian Journal of Botany*, *85*(12), 1111–1126. https://doi.org/10.1139/B07-109

Pichersky, E. & Raguso, R. A. (2018). Why do plants produce so many terpenoid compounds? *New Phytologist*, *220*(3), 692–702. https://doi.org/10.1111/nph.14178

Pitt, J. J. (2009). Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *Clinical Biochemist Reviews*, *30*(1), 19–34.

Plata-Rueda, A., Campos, J. M., da Silva Rolim, G., Martínez, L. C., Dos Santos, M. H., Fernandes, F. L., Serrão, J. E. & Zanuncio, J. C. (2018). Terpenoid constituents of cinnamon and clove essential oils cause toxic effects and behavior repellency response on granary weevil, *Sitophilus granarius*. *Ecotoxicology and Environmental Safety*, *156*, 263–270. https://doi.org/10.1016/j.ecoenv.2018.03.033

Pulst, S. M. (1999). Genetic Linkage Analysis. *Archives of Neurology*, *56*(6), 667–672. https://doi.org/10.1001/archneur.56.6.667

Qiu, S., Cai, Y., Yao, H., Lin, C., Xie, Y., Tang, S. & Zhang, A. (2023). Small molecule metabolites: discovery of biomarkers and therapeutic targets. *Signal Transduction and Targeted Therapy*, *8*(1), 132. https://doi.org/10.1038/s41392-023-01399-3

Rahim, M. A., Zhang, X. & Busatto, N. (2023). Editorial: Phenylpropanoid biosynthesis in plants. *Frontiers in Plant Science*, *14*, 1230664. https://doi.org/10.3389/fpls.2023.1230664

Rainsford, K. D. (2004). *Aspirin and Related Drugs*. CRC Press.

Rao, X. & Dixon, R. A. (2019). Co-expression networks for plant biology: why and how. *Acta Biochimica et Biophysica Sinica*, *51*(10), 981–988. https://doi.org/10.1093/abbs/gmz080

Robinson, K. M., Ingvarsson, P. K., Jansson, S. & Albrectsen, B. R. (2012). Genetic Variation in Functional Traits Influences Arthropod Community Composition in Aspen (*Populus tremula* L.). *PLoS ONE*, *7*(5), e37679. https://doi.org/10.1371/journal.pone.0037679

Rodziewicz, P., Chmielewska, K., Sawikowska, A., Marczak, Ł., Łuczak, M., Bednarek, P., Mikołajczak, K., Ogrodowicz, P., Kuczyńska, A., Krajewski, P. & Stobiecki, M. (2019). Identification of drought responsive proteins and related proteomic QTLs in barley. *Journal of Experimental Botany*, *70*(10), 2823–2837. https://doi.org/10.1093/jxb/erz075

Romanoski, C. E., Lee, S., Kim, M. J., Ingram-Drake, L., Plaisier, C. L., Yordanova, R., Tilford, C., Guan, B., He, A., Gargalovic, P. S., Kirchgessner, T. G., Berliner, J. A. & Lusis, A. J. (2010). Systems Genetics Analysis of Gene-by-Environment Interactions in Human Cells. *The American Journal of Human Genetics*, *86*(3), 399–410. https://doi.org/10.1016/j.ajhg.2010.02.002

Rubert-Nason, K. F., Couture, J. J., Major, I. T., Constabel, C. P. & Lindroth, R. L. (2015). Influence of Genotype, Environment, and Gypsy Moth Herbivory on Local and Systemic Chemical Defenses in Trembling Aspen (*Populus tremuloides*). *Journal of Chemical Ecology*, *41*(7), 651–661. https://doi.org/10.1007/s10886-015-0600-z

Rusli, H., Putri, R. M. & Alni, A. (2022). Recent Developments of Liquid Chromatography Stationary Phases for Compound Separation: From Proteins to Small Organic Compounds. *Molecules*, *27*(3), 907. https://doi.org/10.3390/molecules27030907

Ruuhola, T. & Julkunen-Tiitto, R. (2003). Trade-Off Between Synthesis of Salicylates and Growth of Micropropagated *Salix pentandra*. *Journal of Chemical Ecology*, *29*(7), 1565-1588. https://doi.org/10.1023/a:1024266612585

Ruuhola, T. M. & Julkunen-Tiitto, M-R. K. (2000). Salicylates of Intact *Salix myrsinifolia* Plantlets Do Not Undergo Rapid Metabolic Turnover. *Plant Physiology*, *122*(3), 895–906. https://doi.org/10.1104/pp.122.3.895

Ruuhola, T. M., Sipura, M., Nousiainen, O. & Tahvanainen, J. (2001). Systemic Induction of Salicylates in *Salix myrsinifolia* (Salisb.). *Annals of Botany*, *88*(3), 483–497. https://doi.org/10.1006/anbo.2001.1491

Salam, U., Ullah, S., Tang, Z-H., Elateeq, A. A., Khan, Y., Khan, J., Khan, A. & Ali, S. (2023). Plant Metabolomics: An Overview of the Role of Primary and Secondary Metabolites against Different Environmental Stress Factors. *Life*, *13*(3), 706. https://doi.org/10.3390/life13030706

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Santosh, B., Varshney, A. & Yadava, P. K. (2015). Non-coding RNAs: biological functions and applications. *Cell Biochemistry and Function*, *33*(1), 14–22. https://doi.org/10.1002/cbf.3079

Scarano, C., Veneruso, I., De Simone, R. R., Di Bonito, G., Secondino, A. & D'Argenio, V. (2024). The Third-Generation Sequencing Challenge: Novel Insights for the Omic Sciences. *Biomolecules*, *14*(5), 568. https://doi.org/10.3390/biom14050568

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods,* 9(7), 671-675. https://doi.org/10.1038/nmeth.2089

Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & McLean, J. A. (2016). Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*, *27*(12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y

Schwab, S. G., Hallmayer, J., Lerer, B., Albus, M., Borrmann, M., Hönig, S., Strauß, M., Segman, R., Lichtermann, D., Knapp, M., Trixler, M., Maier, W. & Wildenauer, D. B. (1998). Support for a Chromosome 18p Locus Conferring Susceptibility to Functional Psychoses in Families with Schizophrenia, by Association and Linkage Analysis. *American Journal of Human Genetics*, *63*(4), 1139–1152. https://doi.org/10.1086/302046

Schwartz, S. 2023. Illumina's revolutionary NovaSeq X exceeds 200th order milestone in first quarter 2023. https://emea.illumina.com/company/news-center/press-releases/2023/4b48580b-42d4-419b-8512-5adcbb069836.html [Accessed 2025-11-04]

Serin, E. A. R., Nijveen, H., Hilhorst, H. W. M. & Ligterink, W. (2016). Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Plant Science*, *7*, 444. https://doi.org/10.3389/fpls.2016.00444

Shabalin, A. A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, *28*(10), 1353–1358. https://doi.org/10.1093/bioinformatics/bts163

Shah, Z. A., Hameed, A., Ahmed, A., Simjee, S. U., Jabeen, A., Ullah, A. & Shaheen, F. (2016). Cytotoxic and anti-inflammatory salicin glycosides from leaves of *Salix acmophylla*. *Phytochemistry Letters, 17,* 107-113. https://doi.org/10.1016/j.phytol.2016.07.013

Shanmugavadivel, P. S., Mithra, S. V. A., Dokku, P., Kumar, K. A. R., Rao, G. J. N., Singh, V. P., Singh, A. K., Singh, N. K. & Mohapatra, T. (2013). Mapping quantitative trait loci (QTL) for grain size in rice using a RIL population from Basmati × *indica* cross showing high segregation distortion. *Euphytica*, *194*(3), 401–416. https://doi.org/10.1007/s10681-013-0964-5

Singh, B. & Sharma, R. A. (2015). Plant terpenes: defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech*, *5*(2), 129–151. https://doi.org/10.1007/s13205-014-0220-2

Sinha, D., Odoh, U. E., Ganguly, S., Muhammad, M., Chatterjee, M., Chikeokwu, I. & Egbuna, C. (2023). Phytochemistry, history, and progress in drug discovery. In Egbuna, C., Rudrapal, M. & Tijjani, H. (eds.). *Phytochemistry, Computational Tools, and Databases in Drug Discovery*. Elsevier Inc, pp. 1-26. https://doi.org/10.1016/B978-0-323-90593-0.00001-0

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. (2006). XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Analytical Chemistry*, *78*(3), 779–787. https://doi.org/10.1021/ac051437y

Smith, E. N. & Kruglyak, L. (2008). Gene-Environment Interaction in Yeast Gene Expression. *PLoS Biology*, *6*(4), e83. https://doi.org/10.1371/journal.pbio.0060083

Soneson, C., Love, M. I. & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.1

Steppuhn, A., Gase, K., Krock, B., Halitschke, R. & Baldwin, I. T. (2004). Nicotine's Defensive Function in Nature. *PLoS Biology*, *2*(8), e217. https://doi.org/10.1371/journal.pbio.0020217

Stevens, M. T. & Lindroth, R. L. (2005). Induced resistance in the indeterminate growth of aspen (*Populus tremuloides*). *Oecologia*, *145*(2), 298–306. https://doi.org/10.1007/s00442-005-0128-y

Stine, O. C., Xu, J., Koskela, R., McMahon, F. J., Gschwend, M., Friddle, C., Clark, C. D., McInnis, M. G., Simpson, S. G., Breschel, T. S., Vishio, E., Riskin, K., Feilotter, H., Chen, E., Shen, S., Folstein, S., Meyers, D. A., Botstein, D., Marr, T. G. & DePaulo, R. (1995). Evidence for Linkage of Bipolar Disorder to Chromosome 18 with a Parent-of-Origin Effect. *American Journal of Human Genetics*, *57*(6), 1384–1394.

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *The Journal of Experimental Zoology*, *14*(1), 43–59. https://doi.org/10.1002/jez.1400140104

Subramaniam, R., Reinold, S., Molitor, E. K. & Douglas, C. J. (1993). Structure, Inheritance, and Expression of Hybrid Poplar (*Populus trichocarpa* × *Populus deltoides*) Phenylalanine Ammonia-Lyase Genes. *Plant Physiology*, *102*(1), 71–83. https://doi.org/10.1104/pp.102.1.71

Sul, J. H., Martin, L. S. & Eskin, E. (2018). Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genetics, 14*(12), e1007309. https://doi.org/10.1371/journal.pgen.1007309

Tahvanainen, J., Helle, E., Julkunen-Tiitto, R. & Lavola, A. (1985). Phenolic compounds of willow bark as deterrents against feeding by mountain hare. *Oecologia*, *65*(3), 319–323. https://doi.org/10.1007/BF00378905

Tebani, A., Afonso, C. & Bekri, S. (2018). Advances in metabolome information retrieval: turning chemistry into biology. Part I: analytical chemistry of the metabolome. *Journal of Inherited Metabolic Disease*, *41*, 379–391. https://doi.org/10.1007/s10545-017-0074-y

Tembo, N., Lampiao, F., Mwakikunga, A. & Chikowe, I. (2021). Ethnobotanical survey of medicinal plants used for cervical cancer management in Zomba District, Malawi. *Scientific African*, *13*(2), e00941. https://doi.org/10.1016/j.sciaf.2021.e00941

Trewavas, A. (2006). A Brief History of Systems Biology. *The Plant Cell*, *18*(10), 2420–2430. https://doi.org/10.1105/tpc.106.042267

Tsai, C-J., Harding, S. A., Tschaplinski, T. J., Lindroth, R. L. & Yuan, Y. (2006). Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. *New Phytologist*, *172*(1), 47–62. https://doi.org/10.1111/j.1469-8137.2006.01798.x

Tsui, L. C., Buetow, K. & Buchwald, M. (1986). Genetic analysis of cystic fibrosis using linked DNA markers. *American Journal of Human Genetics*, *39*(6), 720–728.

Upton, R. N., Correr, F. H., Lile, J., Reynolds, G. L., Falaschi, K., Cook, J. P. & Lachowiec, J. (2023). Design, execution, and interpretation of plant RNA-seq analyses. *Frontiers in Plant Science*, *14*, 1135455. https://doi.org/10.3389/fpls.2023.1135455

van der Sijde, M. R., Ng, A. & Fu, J. (2014). Systems genetics: From GWAS to disease pathways. *Biochimica et Biophysica Acta*, *1842*(10), 1903–1909. https://doi.org/10.1016/j.bbadis.2014.04.025

Veenstra, T. D. (2021). Omics in Systems Biology: Current Progress and Future Outlook. *Proteomics*, *21*(3–4), e2000235. https://doi.org/10.1002/pmic.202000235

Vicentini, C., Galuppini, F., Corbo, V. & Fassan, M. (2019). Current role of non-coding RNAs in the clinical setting. *Non-Coding RNA Research*, *4*(3), 82–85. https://doi.org/10.1016/j.ncrna.2019.09.001

Villalba, J. J., Burritt, E. A. & Clair, S. B. S. (2014). Aspen (*Populus tremuloides* Michx.) Intake and Preference by Mammalian Herbivores: The Role of Plant Secondary Compounds and Nutritional Context. *Journal of Chemical Ecology*, *40*(10), 1135–1145. https://doi.org/10.1007/s10886-014-0507-0

Voit, E. O. (2022). Perspective: Systems biology beyond biology. *Frontiers in Systems Biology*, *2*, 987135. https://doi.org/10.3389/fsysb.2022.987135

Voy, B. H. (2011). Systems Genetics: a Powerful Approach for Gene-Environment Interactions. *The Journal of Nutrition*, *141*(3), 515–519. https://doi.org/10.3945/jn.110.130401

Walters, D. R. (2010). *Plant Defense: Warding off Attack by Pathogens, Herbivores and Parasitic Plants*. Wiley-Blackwell.

Wang, B-H., Wu, Y-T., Huang, N-T., Zhu, X-F., Guo, W-Z. & Zhang, T-Z. (2006). QTL Mapping for Plant Architecture Traits in Upland Cotton Using RILs and SSR Markers. *Acta Genetica Sinica*, *33*(2), 161–170. https://doi.org/10.1016/S0379-4172(06)60035-8

Wang, B., Jia, P., Gao, S., Zhao, H., Zheng, G., Xu, L. & Ye, K. (2025). Long and Accurate: How HiFi Sequencing is Transforming Genomics. *Genomics, Proteomics and Bioinformatics*, *23*(1). qzaf003. https://doi.org/10.1093/gpbjnl/qzaf003

Wang, J., Ding, J., Tan, B., Robinson, K. M., Michelson, I. H., Johansson, A., Nystedt, B., Scofield, D. G., Nilsson, O., Jansson, S., Street, N. R. & Ingvarsson, P. K. (2018). A major locus controls local adaptation and adaptive life history variation in a perennial plant. *Genome Biology*, *19*, 72. https://doi.org/10.1186/s13059-018-1444-y

Wang, J., Wang, J., Tian, B., Li, Q., Zhu, J., Liu, X., Ma, C., Li, C., Qi, Z., Zhu, R., Shi, Y., Zou, J., Wen, Y., Sun, Z., Liu, H., Jiang, H., Yin, Z., Hu, Z., Chen, Q., Xin, D. & Liu, C. (2019). QTL mapping and gene mining to identify genes on soybean (*Glycine max*) associated with NopB of *Sinorhizobium fredii* HH103. *Plant Breeding*, *138*(6), 677–685. https://doi.org/10.1111/pbr.12714

Wang, X. (2016). *Next-Generation Sequencing Data Analysis*. CRC Press.

Wasson, A. P., Pellerone, F. I. & Mathesius, U. (2006). Silencing the Flavonoid Pathway in *Medicago truncatula* Inhibits Root Nodule Formation and Prevents Auxin Transport Regulation by Rhizobia. *The Plant Cell*, *18*(7), 1617–1629. https://doi.org/10.1105/tpc.105.038232

Watson, J. D. & Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, *171*(4356), 737–738. https://doi.org/10.1038/171737a0

Weir, J. C. & Boyes, D. (2024). The genome sequence of the Vapourer moth, *Orgyia antiqua* (Linnaeus, 1758). *Wellcome Open Research, 8*, 314. http://doi.org/10.12688/wellcomeopenres.19480.2

Weng, J-K. (2014). The evolutionary paths towards complexity: a metabolic perspective. *New Phytologist*, *201*(4), 1141–1149. https://doi.org/10.1111/nph.12416

Weng, J-K. & Chapple, C. (2010). The origin and evolution of lignin biosynthesis. *New Phytologist*, *187*(2), 273–285. https://doi.org/10.1111/j.1469-8137.2010.03327.x

Weng, J-K., Philippe, R. N. & Noel, J. P. (2012). The Rise of Chemodiversity in Plants. *Science*, *336*(6089), 1667–1670. https://doi.org/10.1126/science.1217411

Wimp, G. M., Wooley, S., Bangert, R. K., Young, W. P., Martinsen, G. D., Keim, P., Rehill, B., Lindroth, R. L. & Whitham, T. G. (2007). Plant genetics predicts intra-annual variation in phytochemistry and arthropod community structure. *Molecular Ecology*, *16*(23), 5057–5069. https://doi.org/10.1111/j.1365-294X.2007.03544.x

Wink, M. (1998). A Short History of Alkaloids. In Roberts, M. F. & Wink, M. (eds.). *Alkaloids: Biochemistry, Ecology, and Medicinal Applications*. Springer Science+Business Media New York, pp. 11-44.

Woolbright, S. A., Rehill, B. J., Lindroth, R. L., DiFazio, S. P., Martinsen, G. D., Zinkgraf, M. S., Allan, G. J., Keim, P. & Whitham, T. G. (2018). Large effect quantitative trait loci for salicinoid phenolic glycosides in *Populus*: Implications for gene discovery. *Ecology and Evolution*, *8*(7), 3726–3737. https://doi.org/10.1002/ece3.3932

Wooley, S. C., Scott, W., Vernon, J. & Lindroth, R. L. (2008). Aspen Decline, Aspen Chemistry, and Elk Herbivory: Are They Linked? *Rangelands*, *30*(1), 17–21. https://doi.org/10.2458/azu_rangelands_v30i1_wooley

Wu, M., Northen, T. R. & Ding, Y. (2023). Stressing the importance of plant specialized metabolites: omics-based approaches for discovering specialized metabolism in plant stress responses. *Frontiers in Plant Science*, *14*, 1272363. https://doi.org/10.3389/fpls.2023.1272363

Xiao, D., Zhao, J. J., Hou, X. L., Basnet, R. K., Carpio, D. P. D., Zhang, N. W., Bucher, J., Lin, K., Cheng, F., Wang, X. W. & Bonnema, G. (2013). The *Brassica rapa FLC* homologue *FLC2* is a key regulator of flowering time, identified through transcriptional co-expression networks. *Journal of Experimental Botany*, *64*(14), 4503–4516. https://doi.org/10.1093/jxb/ert264

Xiao, S., Wang, B., Liu, Y., Miao, T., Zhang, H., Wen, P., He, J., Huang, J., Liu, D., Qiu, Z., Liu, L., Liu, S., Jiang, L., Cheng, X., Wang, C., Xu, D. & Wan, J. (2019). Genome-wide association study and linkage analysis on resistance to rice black-streaked dwarf virus disease. *Molecular Breeding*, *39*, 73. https://doi.org/10.1007/s11032-019-0980-9

Yepes-Vivas, S., Popp, M., Reichelt, M., Gershenzon, J., Schnitzler, J-P. & Unsicker, S. B. (2025). Deciphering organ-specific chemical changes following insect herbivory in *Populus nigra* using comparative metabolomics. *Plant Biology*, *27*(5), 834–846. https://doi.org/10.1111/plb.70032

Zenk, M. H. (1967). Pathways of salicyl alcohol and salicin formation in *Salix purpurea* L. *Phytochemistry*, *6*(2), 245–252.

Zhang, J., Zhao, J., Xu, Y., Liang, J., Chang, P., Yan, F., Li, M., Liang, Y. & Zou, Z. (2015). Genome-Wide Association Mapping for Tomato Volatiles Positively Contributing to Tomato Flavor. *Frontiers in Plant Science*, *6*, 1042. https://doi.org/10.3389/fpls.2015.01042

Zhang, T., Wu, T., Wang, L., Jiang, B., Zhen, C., Yuan, S., Hou, W., Wu, C., Han, T. & Sun, S. (2019). A Combined Linkage and GWAS Analysis Identifies QTLs Linked to Soybean Seed Protein and Oil Content. *International Journal of Molecular Sciences*, *20*(23), 5915. https://doi.org/10.3390/ijms20235915

Zhang, Y., Wang, X., Bian, Z., Se, C., Yang, G. & Lu, Y. (2025). Application of flavonoid compounds suppresses the cotton aphid, *Aphis gossypii*. *Frontiers in Plant Science*, *16*, 1545499. https://doi.org/10.3389/fpls.2025.1545499

Zhou, X. & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics, 44*(7), 821-824. https://doi.org/10.1038/ng.2310

Ziegler, J. & Facchini, P. J. (2008). Alkaloid Biosynthesis: Metabolism and Trafficking. *Annual Review of Plant Biology*, *59*, 735–769. https://doi.org/10.1146/annurev.arplant.59.032607.092730

Zwick, M. E., Cutler, D. J. & Chakravarti, A. (2000). Patterns of Genetic Variation in Mendelian and Complex Traits. *Annual Review of Genomics and Human Genetics*, *1,* 387-407. https://doi.org/10.1146/annurev.genom.1.1.387