Department of Plant Physiology, 2026

Comparative co-expression network analysis of abiotic stress response in boreal conifers

Elena van Zalen

UMEÅ UNIVERSITY

ISBN: 978-91-8070-882-1 (print)
ISBN: 978-91-8070-883-8 (pdf)
Cover by Elena van Zalen
Electronic version available at: http://umu.diva-portal.org/
Printed by: Scandinavian Print Group, Skarpnäck, 2026

“Would you tell me, please, which way I ought to go from here?”

“That depends a good deal on where you want to get to,” said the Cat.

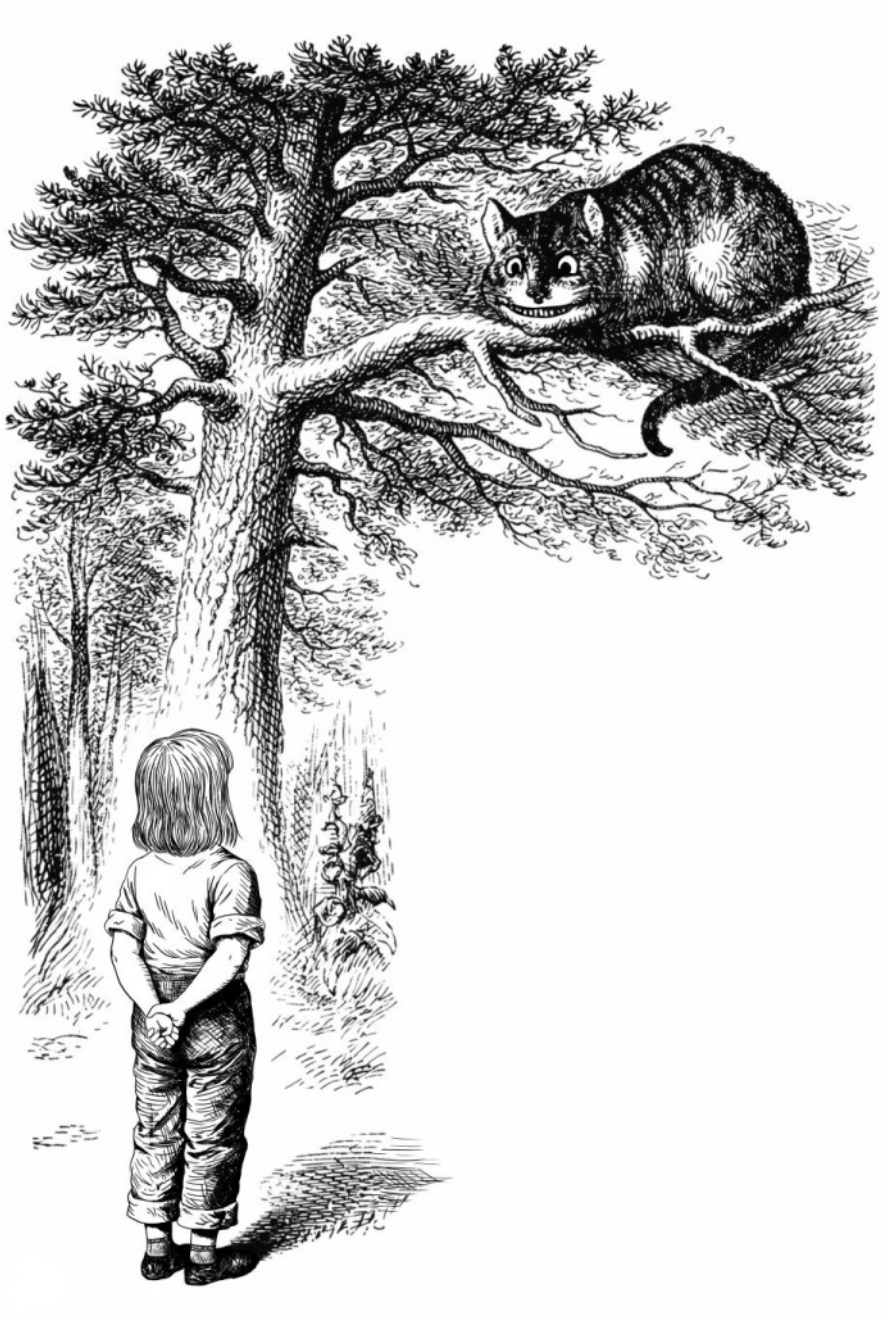
“I don’t much care where—” said Alice.

“Then it doesn’t matter which way you go,” said the Cat.

“—so long as I get somewhere,” Alice added as an explanation.

“Oh, you’re sure to do that,” said the Cat, “if you only walk long enough.”

Lewis Carroll, Alice’s Adventures in Wonderland, Chapter VI: Pig and Pepper, 1865.



Adapted from the original illustration by John Tenniel 1865

Table of Contents

Abstract

Gene co-expression networks (GCNs) are a powerful approach for exploring transcriptional regulation by identifying functionally related genes through their expression patterns across various conditions. The inference of GCNs can be achieved by various computational algorithms, each with distinct merits and limitations. The choice of algorithm can influence the network structure and the biological interpretation derived from it. By using a combination of different methods, biases can be minimised providing more robust and complementary insights. These methodologies are particularly valuable for non-model species, a challenge exemplified by Norway spruce and Scots pine. With ongoing climate change, drought and cold stresses are becoming increasingly important factors shaping the survival of these boreal conifers. Boreal regions are experiencing more frequent and prolonged drought periods, alongside greater variability in early spring, including sudden freeze-thaw events and episodes of extreme cold. Understanding the genetic regulation through which species such as Norway spruce and Scots pine, perceive, respond to, and potentially recover from drought and cold is therefore of high importance.

In this thesis I have used an extensive collection of transcriptomic data generated from boreal tree species under abiotic stress conditions to infer GCNs to reveal coordinated patterns of gene expression responses to environmental challenges. In addition, comparative analyses of GCNs enabled the systematic assessment of conservation and divergence of co-expression among these species, identifying both shared regulatory circuits and species-specific adaptations. Analyses uncovered down-regulated modules of developmental processes, up-regulated modules of abiotic stress response, and several candidate transcription factors directly connected to these stress-responsive pathways. Comparison with boreal angiosperms revealed divergent responses in core cold-regulatory processes, most notably in the regulation and representation of C- repeat Binding Factor (CBF) transcription factors. The abiotic stress response patterns of both cold and drought were largely shared between the two conifer species, indicating a high degree of conservation in their transcriptional responses. This conservation extended to the organisation of topologically associated domains, where a subset of highly conserved co-expressed orthologs were found at the same location in the genomes of these conifers.

Together, these analyses demonstrated the utility of comparative co-expression networks as a tool for understanding both conserved and diverged regulatory mechanisms, while offering new perspectives on the resilience of conifers in the context of environmental change.

Sammanfattning

Samuttrycksnätverk (GCNs) är ett kraftfullt verktyg för att utforska transkriptionsreglering genom att identifiera funktionellt relaterade gener via deras uttrycksmönster under olika förhållanden. GCNs kan konstrueras med hjälp av olika beräkningsalgoritmer, som var och en har styrkor och begränsningar. Valet av algoritm kan påverka nätverkets struktur och den biologiska tolkning som görs utifrån det. Genom att använda en kombination av olika metoder kan bias minimeras, vilket ger mer robusta och kompletterande insikter. Dessa metoder är särskilt värdefulla för icke-modellorganismer, vilket illustreras av gran och tall. I takt med det pågående klimatförändringarna blir torka och kyla allt viktigare faktorer som påverkar överlevnaden hos dessa boreala barrträd. Boreala regioner upplever nu mer frekventa och långvariga torkperioder, tillsammans med större variationer under tidig vår, inklusive plötsliga frysnings-tiningscykler och perioder av extrem kyla. Att förstå den genetiska regleringen genom vilken arter som gran och tall uppfattar, reagerar på och potentiellt återhämtar sig från torka och kyla är därför av stor betydelse.

I denna avhandling har jag använt en omfattande samling transkriptomdata genererad från boreala trädslag under abiotiska stressförhållanden för att konstruera GCNs och därigenom avslöja koordinerade uttrycksmönster som svar på miljöutmaningar. Dessutom möjliggjorde jämförande analyser av GCNs en systematisk bedömning av bevarande och divergens i samuttryck mellan arterna, vilket identifierade både delade regulatoriska nätverk och artspecifika anpassningar. Analyserna avslöjade nedreglerade moduler kopplade till utvecklingsprocesser, uppreglerade moduler involverade i abiotisk stressrespons, och flera kandidattranskriptionsfaktorer som var direkt kopplade till reglering av dessa stressrespons. Jämförelser mellan boreala angiospermer visade divergerande response i centrala processer reglerade av kyla, mest noterbart i regleringen och representationen av C-repeat Binding Factor (CBF)-transkriptionsfaktorer. Responsmönstren för både kyla och torka var i stor utsträckning delade mellan de två barrträdsarterna, vilket tyder på en hög grad av bevarande i deras transkriptionella svar. Detta bevarande sträckte sig även till organisationen av topologiskt associerade domäner, där en undergrupp av starkt bevarade samuttryckta ortologer hittades på samma position i båda arternas genom.

Sammantaget visar dessa analyser nyttan av att jämföra samuttrycksnätverk som ett verktyg för att förstå både bevarade och divergerade regulatoriska mekanismer, och ger samtidigt nya perspektiv på barrträdens resiliens i en föränderlig miljö.

List of publications

This thesis discusses the work contained in the following papers, referred to by Roman numerals in the text:

Paper I: Bastian Schiffthaler, **Elena van Zalen**, Alonso R. Serrano, Nathaniel R. Street, Nicholas Delhomme, (2023). Seiðr: Efficient calculation of robust ensemble gene networks. Heliyon, 9(6). https://doi.org/10.1016/j.heliyon.2023.e16811

Paper II: Ove Nilsson*, Teitur Ahlgren Kalman*, Nicolas Delhomme*, Mimmi Eriksson*, Jason Hill*, Vikash Kumar*, Tomas Larsson*, Niklas Mähler*, Soumyadeep Nandi*, Per Unneberg*, Tom van der Valk*, Andrea Zuccolo, Zhi-Qiang Chen, Maximiliano Estravis Barcala, Eduardo Rodriguez, Tomáš Funda, Rajiv Chaudhary, Siri Birkeland, Remi-Andre Olsen, Ignas Bunikis, Jamie Mccann, Hannele Tuominen, Camilla Canovi, Edoardo Piombo, Zulema Carracedo Lorenzo, **Elena van Zalen**, Mari Suontama, Henrik Hallingbäck, Mai-Britt Mosbech, Helena Vassilieff, Bastian Schiffthaler, Manfred Grabherr, Linda Bakker, Elio Schijlen, Ruben Benstein, Jens Sundström, Karl Johan Westrin, Olof Emanuelsson, Olga Pettersson, Torgeir Hvidsten, Ellen Sherwood, Pär Ingvarsson, Harry Wu, Ulf Gyllensten, Ove Nilsson, Björn Nystedt, Nathaniel Street, (In review). 1000 conifer genomes: Genome innovation, organisation and diversity. PREPRINT (Version 1) available at Research Square: https://doi.org/10.21203/rs.3.rs-6502828/v1

Paper III: Tuuli Aro*, **Elena van Zalen***, Alexander Vergara, Camilla Canovi, Vikash Kumar, Ellen Dimmen Chapple, Torgeir R. Hvidsten, Nathaniel R. Street, Vaughan Hurry, (In preparation). Comparative genomics of cold temperature responses in boreal tree roots. (Manuscript)

Paper IV: **Elena van Zalen**, Camilla Canovi, Vikash Kumar, Ellen Dimmen Chapple, David Castro, Sonja Viljamaa, Torgeir R. Hvidsten, Vaughan Hurry, Nathaniel R. Street, (In preparation). Comparative genomics of abiotic stress response in Norway spruce and Scots pine. (Manuscript)

*These authors contributed equally to this work

Authorship contributions

Paper I. Schiffthaler *et al.* [Published]: I performed the clustering calculations and analysis, enrichment and transcription factor analysis, and generated the curated list of genes putatively involved in drought response.

Paper II. Kalman *et al.* [In review]: I generated co-expression networks for multiple datasets involved in the study. I performed gene expression analysis of the FTL-genes. I identified conserved co-expression of orthologs between Norway spruce and Scots pine of the abiotic stress datasets, and assessed their expression patterns in conserved TADs, including calculation and analysis of network centrality metrics. Additionally, I contributed to the Scots pine genome by generating the drought stress data.

Paper III. Aro & Van Zalen *et al.* [In preparation]: I ran the RNA-seq processing-pipeline and biological quality assessment on the Norway spruce and Scots pine data. I identified conserved co-expression of ortholog between all six plants, performed the comparative analysis and contributed to the discussion of results, and the writing and revision of the manuscript.

Paper IV. Van Zalen *et al.* [In preparation]: I generated the drought stress data in Scots pine. I performed the preprocessing and quality assessment of all datasets involved. I performed all analyses and wrote the manuscript.

Abbreviations

ABA	Abscisic Acid
ARACNE	Algorithm for the Reconstruction of Accurate Cellular Networks
AtBG1	*Arabidopsis thaliana* Beta-Glucosidase1
AUCR	Area Under the Curve Ratio
bZIP	basic leucine Zipper
CaM	Calmodulin
Cas	CRISPR-associated proteins
CBF	C-repeat Binding Factor
CBL	Calcineurin B-like protein
CCA1	Circadian Clock Associated1
CDPK	Calcium-Dependent Protein Kinase
ChIP-Seq	Chromatin Immunoprecipitation-Sequencing
CML	CaM-Like proteins
ComPlEx	Comparative analysis of Plant co-Expression networks
CLR	Context Likelihood of Relatedness
COR	Cold-Regulated gene
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRT	Cold-Responsive element or C-repeat
DAP-Seq	Deoxyribonucleic acid Affinity Purification-Sequencing
DNA	Deoxyribonucleic Acid
DREB	Dehydration-Responsive Element Binding protein
ELNET	Elastic NET
ERF	Ethylene Response Factor
FAD	Fatty Acid Desaturate
FC	Field Capacity
FDR	False Discovery Rate
GCN	Gene Co-expression Network
GENIE3	GEne Network Inference with Ensemble of trees
GO	Gene Ontology
GRN	Gene Regulatory Network
HSP	Heat Shock Protein
HSF	Heat Shock Factor
ICE1	Inducer of *CBF* Expression 1
IRP	Inverse Rank Product
LASSO	Least Absolute Shrinkage and Selection Operator
LBD	Lateral organ Boundaries Domain
LEA	Late Embryogenesis Abundant
LHY	Late elongated Hypocotyl
lncRNA	long non-coding Ribonucleic Acid

MAPK	Mitogen-Activated Protein Kinase
MI	Mutual Information
MR	Mutual Ranking
mRNA	messenger-Ribonucleic Acid
MYB	Myeloblastosis
MYC	Myelocytomatosis
NVN	Node Vicinity Network
PCA	Principal Component Analysis
PCor	Partial Correlation
PlantGenIE	Plant Genome Integrative Explorer
PLSNET	Partial Least Squares-based gene NETwork inference method
PYL	PYR1-Like
PYR	Pyrabactin Resistance 1
RCAR	Regulatory Component of ABA Receptors
Ribo-Seq	Ribosome-Sequencing or Ribosome Profiling
RNA-Seq	Ribonucleic Acid-Sequencing
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
TAD	Topologically Associating Domain
TE	Transposable Element
TF	Transcription Factor
TIGRESS	Trustful Inference of Gene REgulation using Stability Selection
TN	True Negative
TOC1	Timing Of Cab expression1
TP	True Positive
TSS	Transcription Starting Site
VST	Variance-Stabilizing Transformation

Introduction

Over the past few decades, advances in data science and bioinformatics have revolutionised biological research. The rapid growth in computational power, coupled with the development of efficient and more sophisticated algorithms and high-throughput sequencing technologies, has made it possible to process and interpret vast and complex biological datasets. As a result, genome assembly and annotation have reached levels of quality and completeness that were previously unattainable, even for species with exceptionally large and challenging genomes. These developments have opened new opportunities to explore the genetic and molecular bases of development, adaptation, resilience, and response to environmental change.

In the context of anthropogenic climate change, such tools are particularly valuable for studying long-lived and ecologically important species like conifers. Their large genomes, slow growth, and environmental sensitivity make them both challenging and critical subjects for understanding plant stress responses and adaptive mechanisms.

This PhD thesis brings together computational and experimental approaches to investigate how conifers respond to abiotic stress conditions such as drought and cold. Using transcriptomics data and network-based analyses, the work aims to uncover conserved and species-specific regulatory patterns underlying stress tolerance. Together, these studies contribute to a broader understanding of conifer abiotic stress response and provide resources that can support future forest management and conservation strategies under a changing climate.

Norway spruce and Scots pine as a study system

In biology, model species are organisms selected for their experimental tractability, typically having short generation times, well-characterised genetics, high quality annotated genomes, and ease of cultivation and genetic manipulation. These features, combined with extensive data resources and well-established research communities, make model species powerful systems for investigating fundamental biological mechanisms (Leonelli & Ankeny, 2013; Müller & Grossniklaus, 2010). In plant biology, *Arabidopsis thaliana* (L.) Heynh (Arabidopsis) is widely recognised as the primary model system for plants (Meinke et al., 1998; Yaschenko et al., 2024). Its relatively small genome of around 135 Mbp (Lian et al., 2024), short generation time of 4 to 8 weeks, and high degree of genetic conservation with other plant species allows insights gained from Arabidopsis to be broadly translated across the plant kingdom.

In contrast, *Picea abies* (L.) H. Karst. (Norway spruce) and *Pinus sylvestris* L. (Scots pine) are long-lived gymnosperms with large, repetitive genomes of 17,7 Gbp and 20,3 Gbp, respectively (**Paper II**). Their generation times range from 25 to 60 years and they do not reach sexual maturity for about 25 years. Despite this, these trees have been researched relatively extensively as they are directly linked to our well-being: They are the most abundant tree species in the boreal forest, the largest terrestrial biome of our planet, where they provide essential ecosystem services, sustain the forestry industry, support indigenous lifestyles, hold high cultural importance, and represent the largest terrestrial carbon storage on earth (Bradshaw & Warkentin, 2015; Garnett et al., 2018; Johnson & Miyanishi, 2012; Snyder et al., 2004). Economically, these trees are of high importance for several countries around the world for their use in timber and pulpwood production. In commercial Swedish forests alone, Norway spruce and Scots pine represent 42% and 39% of the total forest standing volume (Kumar et al., 2021). In all of Europe, Scots pine represents 20% of the forest area (Durrant et al., 2016).

Increased exposure to environmental changes makes boreal conifers even more relevant for extensive research into gene regulation. Norway spruce has been identified as particularly vulnerable to climate-related stressors due to its relatively shallow root system and sensitivity to water deficits, making it a key species for research on stress resilience and gene regulation under changing environmental conditions (Caudullo et al., 2016; Lévesque et al., 2013). In contrast, Scots pine is generally considered more drought-tolerant and exhibits a wider ecological range across Eurasia (Durrant et al., 2016). Despite these physiological differences, the two species are phylogenetically related gymnosperms and share a high degree of genomic and functional similarity (**Paper II**), making them well-suited for comparative studies. Analysing Norway spruce and Scots pine in parallel enables the identification of both shared and species-specific gene regulatory mechanisms in response to abiotic stress.

Gene regulation at a core level

Gene regulation is a fundamental component of the "central dogma" of molecular biology, where DNA is transcribed into RNA, which in turn is translated into protein (Crick, 1970). Transcriptional regulation controls when, where, and how actively genes are expressed, and enables rapid responses to environmental changes through appropriate signalling pathways. This process is driven by transcription factors (TFs), proteins that bind to specific DNA sequences such as promoters and enhancers. Promoters are located proximal to the transcription start site (TSS) of a gene, whereas enhancers are often positioned far from the TSS and can increase transcription of one or several genes (Zabidi & Stark, 2016). While promoter and enhancer elements are often thought of as transcriptional activators, they can

equally act as repressors or silencers that decrease or block transcription (Duttke et al., 2024). Epigenetic modifications, including chemical changes to DNA or histone proteins, can also influence transcription by altering how tightly DNA is packed, thereby affecting the accessibility of genes to the transcriptional machinery. Together, these mechanisms form a complex system that finetunes regulation of gene activity and that enables flexible responses to internal signals and external environmental cues.

Transcriptomics, the study of RNA expression patterns, provides a valuable perspective into this transcriptional regulation. It enables the identification of genes that are activated or repressed under specific conditions, offering insights into cellular and physiological responses. While transcriptomics has some inherent limitations (e.g., not capturing protein abundance or post-translational regulation), it represents the level of the central dogma where we have the best current technologies to perform accurate, quantitative and high throughput analyses, and therefore remains central to systems biology approaches, particularly in species where experimental tools are limited.

Why transcriptomics?

Transcriptomics captures gene expression at the messenger-RNA (mRNA) transcript level, offering a snapshot of transcript abundance under specific environmental or physiological conditions. While transcriptomic data does not encompass regulatory events occurring at the translational or post-translational levels, it does provide a direct means of examining the product of transcriptional regulation. Given the size and complexity of conifer genomes, transcriptomics offers a more practical alternative to whole-genome approaches. By focusing on just expressed protein coding genes, it bypasses many of the challenges posed by large, repeat-rich genomes.

Technological advances in sequencing have revolutionised our ability to study gene expression. Early methods such as Sanger sequencing (Sanger et al., 1977) enabled targeted analysis of individual transcripts. Later, high-throughput approaches such as microarrays allowed genome wide surveys of transcript abundance, but were limited in their range, sensitivity, and detection of lowly expressed or novel transcripts (Bumgarner, 2013). RNA-Sequencing (RNA-Seq) has overcome many of these limitations, expanding coverage of the transcriptome and increasing the range of detectable expression (Giorgi et al., 2013; Ozsolak & Milos, 2011). Transcript quantification using bias-aware tools such as Salmon (Patro et al., 2017), in combination with the now fully assembled and annotated reference genomes for Norway spruce and Scots pine (**Paper II**), the use of RNA-Seq enables robust investigation of gene expression dynamics. This facilitates comprehensive

downstream analyses, including differential expression profiling, gene co-expression network construction, and ortholog-based cross species comparative genomics.

In conifers, which exhibit strong local adaptation with limited genomic divergence (Wang et al., 2020), transcriptomics can reveal important regulatory differences that are not evident from the DNA alone (Krieger et al., 2020; McManus et al., 2010). This is critical for uncovering the molecular basis of phenotypic plasticity and response across environmental gradients (Yeaman et al., 2016). This provides a foundation for investigating how conifers respond to environmental change and supports efforts to enhance forest resilience through molecular insights.

Transcriptional networks

Transcriptional networks are a way of studying the expression dynamics of genes across a range of conditions: they are systems-level representations of genes that show similar expression patterns across multiple conditions, time points, or tissue types. Within a transcriptional network, nodes represent genes, and edges (the connection between nodes) represent the nodes with similarity in their expression patterns. Edge weights reflect the strength of the expression similarity between nodes.

Transcriptional networks can either be directional or non-directional. Directional networks, commonly referred to as Gene Regulation Networks (GRNs), represent known or predicted regulatory relationships between genes. They are typically inferred with the use of bait genes, which are well-characterised genes, often TFs, part of specific pathways with known targets. This directed approach has been widely used for the identification of key genes in specific pathways (Schlitt & Brazma, 2007). Another approach to generating transcriptional networks is the construction of an undirected Gene Co-expression Network (GCN), which reflects potential functional associations rather than known regulatory direction. In this approach, no bait genes are used, but instead the network is constructed from pairwise relationships among all genes across a given sample set (Aoki et al., 2007; Liu et al., 2019), or from multivariate methods that estimate conditional dependencies between genes (Magwene & Kim, 2004; Margolin et al., 2010). This strategy provides a global overview of co-expressed genes, though its application to large datasets can be highly computationally intensive. Genes showing similar expression patterns across multiple samples, could suggest shared biological function rather than coincidental patterns (Jiménez-Gómez, 2014). Using the “Guilt by Association” principle, gene function of unknown genes can be predicted by evaluating highly connected neighbouring genes that have known biological functions (Liu et al., 2023; Wolfe et al., 2005). Particular attention is often given to

so-called hub genes, which are characterised by high connectivity within the network, as they may act as key regulators of biological processes of the given sample set (Dai et al., 2018; Johnson et al., 2025; Yang et al., 2014). The expression dynamics of these hub genes across conditions of the sample set can reveal involvement in stress response, developmental timing, or tissues specific specialisation, depending on the experimental setup.

Application of the Guilt by Association principle has highlighted transcriptional network analysis as a powerful tool for exploring gene function and regulatory mechanisms in plants, including for genes of unknown function (Proost & Mutwil, 2017). The approach enables systematic investigation of complex biological systems and can broadly be divided into three categories (Rao & Dixon, 2019): First, they facilitate the identification of regulatory relationships between TFs or other gene regulators and their downstream targets, providing direct insight into gene regulation (Cortijo et al., 2020; van Dam et al., 2017); second, transcriptional networks can be used to link genotype with biochemical function by integrating metabolic data and a network to predict involvement of genes in metabolic pathways (Liu, 2015; Rydman et al., 2025; Wisecaver et al., 2017); third, by performing comparative network analysis across multiple plant species, functional annotations from well-characterised model species can be transferred to unannotated orthologs (Vercruysse et al., 2020). Collectively, these applications highlight the versatility of transcriptional network analysis. This thesis focusses on the use of the undirected approach, both exploring species specific GCNs (**Paper I**) and performing comparative analyses across multiple species (**Paper II, III & IV**). The use of an undirected approach overcomes the limitations posed by the absence of prior regulatory knowledge, which would enable the application of a GRN approach.

A variety of gene inference approaches

The concept of gene networks was first introduced as “relevance networks” (Butte & Kohane, 1999), where Pearson correlation was applied to identify relationships among results from medical tests. This approach was later extended to gene expression data with Mutual Information (MI) as a complementary method, marking the origins of gene co-expression networks (Butte AJ, 2000). Subsequent advancements in high-throughput transcriptomic technologies such as microarrays and RNA sequencing required the application of more sophisticated inference algorithms. These include methods based on different mathematical approaches such as regression, MI, Bayesian network models or hybrid strategies that integrate multiple methodologies, with each method having its own merits and limitations.

Correlation methods were the earliest methods used to infer expression similarity, and they are still widely used today. They remain popular because they are

computationally efficient, easy to interpret, and have therefore been used as the basis for many widely used plant co-expression databases (Obayashi et al., 2009). Pearson correlation calculates the linear relationship of gene expression by calculating the correlation between each gene pair and scoring the genes in the network accordingly. Even though expression profiles are expected to mostly follow a linear relationship (Daub et al., 2004), biological data is often noisy and linear correlation methods might miss outliers or biological nuances. While linear correlation works well for simple regulatory relationships such as activation and repression, it will miss more complex patterns where a gene is expressed only when multiple regulators are inactive, or when it responds differently to combinations of regulators, or when regulatory effects occur with temporal delay. Spearman's rank correlation captures the monotonic relationships across the data, finding relationships between genes based on the rank correlation coefficient (ρ) rather than the magnitude of expression change (de Siqueira Santos et al., 2013). Spearman is therefore less sensitive to outliers or a lack of strict linearity, but misses relationships that are strictly linear.

Regression-based methods provide another versatile approach beyond regular correlation methods for gene network inference. Regression methods allow the identification of direct associations while accounting for the influence of other genes. While regression methods have primarily been applied in the context of GRNs to identify directed regulator-target relationships, they can also be adapted for the construction of undirected GCNs (Lin et al., 2024). General linear regression models how the expression of a single gene can be explained by the combined influence of other genes, each being weighted based on the strength of its association. One example of this is partial correlation, which uses the Gaussian Graphical Model (GGM) to calculate the pairwise conditional dependencies between nodes, forming an edge only when the nodes are conditionally dependent compared to other nodes in the network (Lauritzen, 1996; Shutta et al., 2022). While straightforward, this approach becomes unreliable in high-dimensional expression datasets, where the number of genes far exceeds the number of samples. These issues are overcome by the regularisation of regression models. Regularisation adds penalties to regression models to shrink the data and produce more stable and interpretable results (Schäfer & Strimmer). Least Absolute Shrinkage and Selection Operator (LASSO), or L1 regularisation, applies a penalty that forces many regression coefficients to zero, removing weaker connections and producing sparse networks that emphasize the most important associations (Friedman et al., 2010; Tibshirani, 1996). L2 regularisation, also known as ridge regression, adds the sum of squared regression coefficients as a penalty in order to minimise overfitting. Elastic NET (ELNET) combines the L1 penalty with the L2 penalty to shrink the size of all coefficients, effectively normalising the results (Zou & Hastie, 2005). This makes the model more stable when predictors are correlated, which occurs often in gene expression data. In addition to these established traditional regression methods, several more

sophisticated approaches have been developed to better handle the complexity of gene expression data. Gene Network Inference with Ensemble of trees (GENIE3) uses an ensemble of regression trees with random forests to predict the expression of each gene in relation to all others (Huynh-Thu et al., 2010). The importance of each predictor gene is then used to rank edges, producing a network that can be interpreted as a directed GRN or symmetrised into an undirected GCN. Trustful Inference of Gene Regulation using Stability Selection (TIGRESS) also builds on regression but incorporates stability selection by repeatedly resampling the data and identifying only those edges that are consistently recovered, thereby increasing the robustness of the inferred network (Haury et al., 2012). Partial Least Squares-based gene NETwork inference method (PLSNET) takes yet another approach, using Partial Least Squares regression to reduce dimensionality of the predictors before estimating gene to gene connections (S. Guo et al., 2016), which again is particularly useful when working with thousands of genes but fewer samples.

A third approach for inferring expression similarity is based on information theory, represented by MI methods. Here the concept of entropy (Shannon, 1948) is used to measure the uncertainty or predictability of the expression of one gene across an experiment. Entropy is low when gene expression remains relatively constant, and entropy is high when gene expression fluctuates. MI quantifies how much information is shared between two genes by measuring the reduction in uncertainty. If two genes are highly co-expressed, knowing the expression of one gene makes the other much more predictable, resulting in high MI. Several MI methods exist to infer edges between nodes of a GCN. Context Likelihood of Relatedness (CLR) normalises MI values by calculating the statistical likelihood in context of the entire network, removing indirect influences and keeping interactions that are significant (Faith et al., 2007). Another prominent method is the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE), which uses Data Processing Inequality (DPI) by looking at triplets of genes to systematically remove indirect connections and retaining only the most likely direct associations between two genes (Margolin et al., 2006). Additionally, partial correlation also has a mathematical link to information theory and can be interpretable as MI when assuming a Gaussian distribution, as is done by PCor (Schäfer & Strimmer).

Bayesian models use probabilistic reasoning to infer gene relationships by combining observed expression data with prior knowledge, such as assumptions about network sparsity, modularity, or preliminary correlation estimates from the data itself. These methods estimate the posterior probability of connections, providing a measure of confidence rather than a fixed threshold. Bayesian network models, such as BANJO (Yu et al., 2004) and BNFinder (Wilczyński & Dojer, 2009), represent conditional dependencies between genes as directed edges, while Bayesian Gaussian graphical models such as DBgraph (Mohammadi & Wit, 2015), infer undirected networks based on posterior distributions of partial correlations.

These approaches are particularly useful when data are noisy or sparse, as they explicitly account for uncertainty in network inference (Ni et al., 2022).

In short, there are many methods available to infer expression similarity, and the optimal choice(s) depends on the characteristics of the data, such as distribution, sample size, and noise level, as well as the biological question being addressed. Independent studies have shown that no single algorithm outperforms others, but that combining results from multiple methods often yields improved results by mitigating the biases inherent to individual approaches (Marbach et al., 2012). This principle is referred to as the "Wisdom of the Crowd" (**Figure 1**), which suggests that consensus networks derived from diverse inference strategies are more likely to capture underlying biological relationships. In conclusion, the key challenge is not simply selecting a single inference method, but determining how to effectively integrate, benchmark, and interpret the results in a biologically meaningful and reproducible manner.

Figure 1 Wisdom of the Crowd concept, or "ask the audience". Multiple answers to the same question sketch the bigger and more complete picture closer to the truth.

Exploration of the network

After building a GCN, the next challenge is to make sense of the resulting "hairball" of interconnected genes (**Figure 2A**). This involves statistically analysing how nodes and edges are organised, identifying key regulators based on these statistics, and grouping genes into modules that may represent shared biological functions. A range of network metrics and clustering approaches can be used to characterise these features.

Network statistics

Network topology measures provide valuable insights into the relative importance of genes within the network. Metrics such as degree, PageRank, eigenvector, Katz, betweenness, closeness, Laplacian centrality, and strength are commonly used, and most are integrated in Seidr (**Paper I**). The majority of these statistical formulas were initially developed for social network studies, but can as easily be applied to GCNs.

Degree centrality is the simplest measure of the connectivity of a given node, counting the number of edges directly linked to a gene (Liao et al., 2017). In weighted networks, degree centrality can be extended to the strength centrality metric, calculated by the sum of edge weights (Candeloro et al., 2016). This reflects the overall level of co-expression, providing a simple yet informative measure of how strongly a gene correlates with other genes in the dataset.

PageRank can evaluate the importance of a given gene based on the connectivity of its neighbours, effectively identifying genes that are influential within densely connected network regions (Brin & Page, 1998). Eigenvector centrality captures similar information but places more emphasis on nodes connected to other highly connected genes, identifying globally influential nodes (Iacobucci et al., 2017). Katz centrality extends this further by incorporating direct and indirect paths, allowing detection of genes that have network-wide influence even if they are not locally dense hubs (Katz, 1953).

Betweenness centrality highlights genes that frequently lie on the shortest paths between other nodes, identifying potential regulatory intermediates that link otherwise distinct modules or pathways (Bavelas, 1948). Closeness centrality instead measures how effectively a node can reach all other nodes in the network, with high-closeness genes potentially acting as information distributors that can rapidly influence multiple targets (Newman, 2010). Laplacian centrality quantifies how much a gene contributes to the overall connectivity by evaluating the effect on the structure of the network after removal of the gene. Genes with high Laplacian

centrality cause a measurable disruption when removed, highlighting their importance for network integrity (Qi et al., 2012).

Genes that score highly across these metrics are often referred to as hub genes. In GCNs, hub genes can represent TFs, signalling proteins, or genes coding for metabolic enzymes that coordinate multiple downstream targets. Within modules, intramodular hubs maintain the coherence of co-expressed gene sets and often play key regulatory roles.

Module detection

While node-level statistics highlight hub genes, module detection reveals the collective organisation of the network (**Figure 2C**). GCNs typically exhibit a modular structure, composed of groups of genes with highly correlated expression patterns. These modules can be identified using community detection algorithms such as Louvain (Blondel et al., 2008), Leiden (Traag et al., 2019), and InfoMap (Rosvall et al., 2009).

The Louvain algorithm partitions the network by iteratively grouping nodes to maximise modularity. Conceptually, it can be viewed as a random walker approach, where nodes are assigned to communities that a hypothetical walker is more likely to remain within, capturing regions of dense intra-community connectivity. The Leiden algorithm improves upon Louvain by providing more stable and well-defined partitions, while also being computationally faster. These attributes make Leiden particularly suitable for large scale transcriptomic networks. Infomap, in contrast, uses information theory to identify modules by compressing the paths of random walks on the network, effectively finding partitions that minimise the expected description length of the path of the walker and highlighting tightly connected subnetworks.

Modules often correspond to sets of co-regulated genes involved in related biological processes or pathways (Liu et al., 2019; Usadel et al., 2009), embodying the Guilt by Association principle. Once defined, these modules provide a basis for downstream analyses, such as identifying module-specific hub genes or testing the enrichment of known biological functions using functional annotations such as Gene Ontology (GO) or the presence of protein domains or genes assigned to biochemical pathways.

Visualisation, functional analyses and candidate gene identification

Visualisation is a crucial step in interpreting GCNs, as it translates complex mathematical relationships into interpretable patterns. Network layouts can reveal the distribution of hubs, the boundaries between modules, and the interactions of different modules and genes. Integrating differential expression values such as $\log_2$

fold changes and adjusted *P*-values, adds an additional layer of information by highlighting modules or subnetworks associated with change in conditions or treatment (**Figure 2B**). Even in densely connected or "hairball-like" networks, such overlays can uncover regions of biological interest.

Significant co-expression does not necessarily mean biological relevance (Usadel et al., 2009), it is a combination of all components discussed before that strengthens the interpretation.

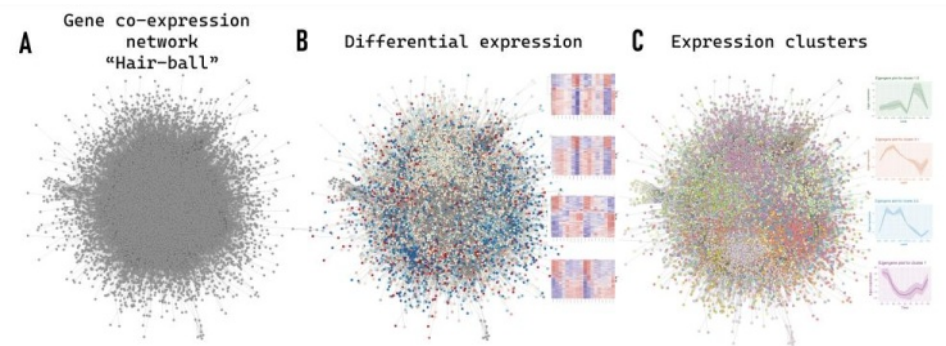


Figure 2 Layers of gene co-expression network analyses. A) Un-edited "hair-ball" of a gene co-expression network (GCN) with nodes as genes and their connection as edges. B) GCN coloured by differential expression values, with blue as significantly down-regulated and red as significantly up-regulated. Heatmaps showing extracted modules of co-expressed genes. C) GCN coloured by Leiden-cluster identification. Plots showing eigenvector expression of four different modules.

Comparative network analysis

An individual GCN captures gene-gene relationships as they appear in a defined experimental context, such as a single tissue type under specific conditions. Comparative network analysis expands this approach by examining how co-expression patterns differ across conditions, tissues, or species (van Dam et al., 2017). Through these comparisons, it becomes possible to identify conserved regulatory processes, context-specific rewiring, and candidate regulators that are shared or unique to specific tissue types, conditional responses or phylogenetic lineages (Jiang et al., 2016; Ovens, Eames, et al., 2021).

The most basic approach in comparative GCN analysis is to infer separate networks for each condition or species and simply compare their global topological metrics (Eidsaa et al., 2017). As an example, one might construct networks under two distinct treatments and then examine summary statistics such as connectivity, clustering coefficient, edge density or degree distributions. Although conceptually

straightforward, these approaches are limited in resolution, since they do not encompass changes to specific functional modules or genes, and are highly sensitive to network inference and thresholding procedures (Elo et al., 2007).

To gain greater biological insight, module-based comparative methods such as Differential Co-Expression (DiffCoEx, (Tesson et al., 2010), Module Differential Analysis (MODA, (Li et al., 2016) and related frameworks were developed. These approaches build upon weighted co-expression methods such as Weighted Gene Co-expression Network Analysis (WGCNA, (Langfelder & Horvath, 2008) to identify modules based on connectivity measurements and subsequently quantify how these modules differ across conditions. Differential module analysis can reveal whether a co-expression module becomes more or less cohesive under specific conditions, whether modules split or merge between them, or whether eigengene correlations between modules are rewired (Erola et al., 2019). These types of comparisons yield directly interpretable results since modules often link to biological pathways. However, mapping modules between networks remains challenging as their composition can drift and correspondences are not always one-to-one, especially across species.

Node-level and edge-level differential co-expression methods take a complementary perspective by focusing on individual genes or gene-gene pairs whose correlation structure changes significantly across different networks. Rather than comparing modules or global topology, these methods assess the statistical difference in correlation or MI for each gene pair across networks. Approaches such as Gene Sets Net Correlations Analysis (GSNCA) (Rahmatallah et al., 2013) evaluate entire gene sets for differential co-expression, while others test specific edges for gain, loss, or sign reversal of correlation (Yu et al., 2011). This fine-grained resolution makes it possible to identify key regulatory genes that change connectivity between the different networks suggesting altered functional roles or responsiveness (Savino et al., 2020). Nevertheless, the large number of possible gene pairs poses statistical and computational challenges, and their interpretation often requires subsequent aggregation into functional modules or pathways (McKenzie et al., 2016; Yu et al., 2011).

More recently, comparative network research has moved toward network alignment and neighbourhood-based methods that operate at an intermediate structural level. Instead of comparing global network properties, these methods focus on the local environments of each gene. The concept of Node Vicinity Networks (NVNs) (Mutwil et al., 2010) with NetworkComparer and related embedding-based approaches, such as Juxtapose (Ovens, Maleki, et al., 2021), represent each gene by the pattern of its surrounding connectivity, enabling direct comparison of these local structures across conditions or species. In cross-species contexts, neighbourhood conservation can be quantified by assessing the similarity of orthologous gene

neighbourhoods, as implemented in comparative frameworks such as Comparative analysis of Plant co-Expression networks (ComPlEx, Netotea et al., 2014) and OrthoClust (Yan et al., 2014). These methods capture high-order structural changes that module- or edge-based approaches may miss and are particularly valuable for identifying conserved regulatory architectures in species where orthology is imperfect or network topology is only partially conserved (Ryan et al., 2012).

Comparing networks across species, can yield deeper biological insight into the regulation of systems in species that are less well understood and that lack functional annotations of genes or information on regulatory mechanisms (e.g. knowledge of which genes are regulated by a TF). For example, the model plant Arabidopsis has a large proportion of its genes experimentally verified in laboratory conditions, providing a rich functional reference. Mappings networks from non-model species to Arabidopsis via best hit orthology (for example, using DIAMOND BLAST) allows the transfer of functional annotation, module context or additional information such as TF-target gene interaction or protein-protein interaction. Orthologous genes pairs can then be placed within comparative networks and analysed for conservation of neighbourhood structure or rewiring of regulatory context.

One must be cautious, however, when interpreting such sequence similarity based orthology mappings. Orthology assignments can be uncertain, particularly in gene families that have undergone lineage-specific expansion or duplication, and such evolutionary events may lead to functional divergence. In roughly half of Arabidopsis-poplar ortholog comparisons, (Netotea et al., 2014) reported that the most sequence-similar ortholog is not necessarily the functional ortholog when co-expression conservation is additionally considered. To identify the correct functional orthologs, it is therefore essential to consider a combination of co-expression values and orthology information.

Regardless of the chosen methodology, valid comparative analyses rely critically on consistent data processing and network inference. Differences in normalisation, filtering, or parameterisation can easily produce artificial discrepancies between networks that are unrelated to biology (Johnson & Krishnan, 2022). Therefore, datasets must be processed using harmonised pipelines, typically including comparable gene sets, normalisation strategies, and comparable and preferably identical inference methods and filtering thresholds.

The importance of data quality

Constructing biologically meaningful GCNs begin not with the algorithms, but with the data itself. Biological data is inherently noisy and heterogenous, where outliers are expected and both linear and non-linear dependencies shape expression relationships. Sample heterogeneity is both a strength and a challenge in co-expression analyses. Diverse samples allow the network to capture broader transcriptional programs, but excessive heterogeneity, such as mixing tissues, or several conditions, can obscure condition-specific patterns and inflate noise (Ballouz et al., 2015). Thoughtful experimental design that balances within-group consistency and between-group contrast yields more interpretable co-expression modules. In practise, richer more structured experimental setups, encompassing multiple time points, treatments, or genotypes, tend to produce networks with modules that correspond more clearly to biological processes. Variation can also arise from technical influences including sequencing depth, preprocessing steps, mapping accuracy, and normalisation. Low coverage or uneven sampling can lead to spurious co-expression, particularly among lowly expressed genes. Filtering out low-abundance transcripts and applying appropriate normalisation steps can help minimise such biases.

The quality of underlying genomic resources also shapes network reliability. Incomplete or fragmented reference genomes, annotation gaps, or transcriptome assemblies with missing isoforms all constrain the detection of true co-expression. Likewise, uneven sampling for instance, underrepresentation of certain tissues or conditions, limits the interpretability of identified modules. High-quality, well-annotated reference genomes and comprehensive sampling thus provide the foundation for accurate inference.

Ultimately, the adage “garbage in, garbage out” is exceptionally relevant in co-expression analysis. The interpretability and biological relevance of inferred networks depend fundamentally on data quality, consistency, and experimental design. Investing effort in careful data preprocessing, adequate sampling, and rigorous quality control not only improves network accuracy, but also ensures that downstream analyses yield insights that truly reflect the underlying biology.

The data used in this thesis consist of gene expression profiles from Norway spruce and Scots pine subjected to cold and drought stress. All expression data were mapped to the latest genome annotations provided by **Paper II**, providing updated gene models that improve the accuracy of transcript quantification and ensure reliable construction of GCN for subsequent comparative analyses.

Abiotic stress response in Norway spruce and Scots pine

In temperate and boreal regions, plants increasingly face cold and drought stress (Seidl et al., 2017). Both impose osmotic and oxidative stress, triggering overlapping transcriptional and physiological responses (Kim et al., 2024). In contrast, stresses like heat and salinity are less representative of the environmental conditions faced by conifer species in boreal regions. While heat is still increasing due to ongoing climate change, periods with extraordinarily high-temperatures remain rare and are usually of short duration. Conifers generally withstand warmer temperatures with water availability, though in combination with drought this might become a problem (D'Orangeville et al., 2018).

Plants have evolved several mechanisms to deal with abiotic stress. When first perceived, they initiate a complex cascade of signal transduction pathways that covert the external stimulus into an intracellular response. This cascade includes the production of Reactive Oxygen Species (ROS), the accumulation of Abscisic Acid (ABA), the activation of secondary messengers such as calcium ions (Ca^{2+}), and phosphorylation relays mediated by protein kinases (Xu et al., 2022). These signalling components collectively induce changes in gene expression, physiological and developmental responses that enhance the ability of the plant to cope with stress. Key regulators of these gene expression changes, include several families of TFs. In the context of abiotic stress, the main TF families include or C-repeat Binding Factors or Dehydration-Responsive Element Binding factors (CBF/DREBs), NAM/ATAF/CUC (NAC), WRKY, Myeloblastosis (MYB), APETALA2/Ethylene Response Factor (AP2/ERF), and basic leucine Zipper (bZIP). Genes within these families are excellent targets for breeding programs and forestry (Kim et al., 2024). However, most knowledge is based on model herbaceous, annual angiosperms and it is not known how well these results translate to conifers, which have been separated, and independently evolving for hundreds of millions of years.

Cold stress response and regulation

Upon exposure to colder temperatures, plants limit growth-related processes in order to prioritise a cold acclimation response. Depending on what species of plant, the degree of cold tolerance varies (Rihan et al., 2017). Plants can grow in environments where cold occurs either on a diurnal or seasonal basis, and species have adapted to these climatic patterns (Preston & Sandve, 2013). Although certain genetic responses to cold are conserved across species, variation in timing and duration of cold exposure have driven the evolution of different adaptive strategies. Even among temperate species that experience seasonal cold (like the ones discussed in this thesis), different evolutionary outcomes have emerged that induce different mechanisms to cope with prolonged low temperature conditions.

Temperate annual species have evolved to align their growth and reproductive cycles with the more favourable conditions in spring and summer (King & Heide, 2009). Spring annuals such as the *A. thaliana* Col-0 accession have evolved so that germination reproduction and senescence occur during the warm seasons. In contrast, winter annuals such as the *A. thaliana* Ost-0 accession, germinate during autumn, are dormant during winter, and flower in spring (Hemming & Trevaskis, 2011). Species that undergo any type of dormancy during winter can overcome these repressive mechanisms by a process known as vernalisation. This process re-initiates growth or flowering in many temperate species after prolonged cold exposure (Bouché et al., 2017; Chouard, 1960). For winter annuals, vernalisation is an important process as they are subjected to longer periods of cold and it is important that resumption of growth does not occur prematurely. Temperate perennials show similar seasonal strategies. Herbaceous perennials experience vernalisation, become cold and frost tolerant and resume growth from dormant meristems after winter. Woody perennials can delay flowering for years until sufficient carbon reserves have accumulated. Additionally, many temperate and boreal trees enter endodormancy to prevent new growth during winter (Lloret et al., 2018; Maurya & Bhalerao, 2017).

To prepare for longer periods of cold exposure, such as during winter, plants go through the physiological and molecular process of cold acclimation. Lower non-freezing temperatures (<18 °C), induce changes in the fluidity of cellular membranes, subsequently changing water potential within cells and disturbing ionic balances (Sardoei & Fazeli-Nasab, 2025). Lipid remodelling by Fatty Acid Desaturates (FADs) during cold stress alters membrane properties making them more rigid (Murata & Los, 1997). This chemically induced membrane rigidity enhances electron transfer to oxygen increasing the production of ROS, in addition to elevated cytosolic Ca^{2+} levels and regulation of hormone signalling (Örvar et al., 2000; Xu et al., 2022). While ROS can be harmful to cellular structures and functions, plants have evolved sophisticated antioxidant systems to moderate their accumulation and mitigate oxidative damage (Baxter et al., 2013; Huang et al., 2019). ROS and calcium-binding proteins enhance cold tolerance by regulating specific Mitogen-Activated Protein Kinases (MAPKs) through phosphorylation of threonine and tyrosine residues, interacting with Inducer of CBF Expression 1 (ICE1) to initiate the ICE1-CBF (C-repeat binding factor) transcriptional cascade (**Figure 3**) (Jalmi & Sinha, 2015; Yuan et al., 2018).

Additionally, Ca^{2+} is perceived by proteins such as Calmodulin (CaM), CaM-like proteins (CMLs), Calcium-Dependent Protein Kinases (CDPKs), and Calcineurin B-Like protein (CBLs), which directly contribute to the regulation of CBF and downstream cold-regulated gene (*COR*) expression (Atif et al., 2019; Satyakam et al., 2022). In Arabidopsis, the circadian clock also plays an important role in cold stress response regulation. Circadian Clock Associated1 (CCA1) and Late elongated

Hypocotyl (LHY) proteins act as positive regulators of *CBF* expression during the morning-loop, promoting cold tolerance at dawn (Eriksson & Webb, 2011). In poplar, LHY has also been shown to be a promoter of cold hardiness, indicating that this regulatory mechanism could be conserved across plant lineages (Ibáñez et al., 2010). During prolonged cold exposure, the evening-phased clock component Timing Of Cab expression1 (*TOC1*) is down-regulated, increasing levels of LHY, subsequently resulting in higher expression of *CBF* (Bieniawska et al., 2008; Ramos et al., 2005).

In Arabidopsis, transcriptional changes to promote cold tolerance are initiated by ICE1, a MYC-type bHLH transcriptional activator located upstream of the CBF genes (Chinnusamy et al., 2003). The CBF/DREBs regulons then coordinate the activation of numerous genes linked to metabolic pathways, stress, signalling and protective cellular functions through Dehydration-Responsive Element/Cold-Responsive element or C-repeat (DRE/CRT) (Chinnusamy et al., 2007). Homologs of both ICE1 and CBF genes are present across many plant lineages, including gymnosperms, where CBF-like genes in conifers show strong cold-induced expression suggesting that they contribute to freezing tolerance (Vergara et al., 2022).

Regulation of drought stress response

Water availability is one of the key factors for plant growth, productivity and survival. Water is taken up from the soil through the roots and transported to the leaves via xylem, where transpiration creates a negative tension that pulls water upward. This movement of water is maintained by a negative gradient in water potential, from higher potential in soil and roots to lower potential in the leaves, and is explained by a process described as cohesion-tension theory (Dixon, 1914).

Under drought conditions, water potential declines throughout the plant, increasing tension in the water column continuum. If the tension within the xylem surpasses a critical threshold, air can be drawn into the water column, causing cavitation. The resulting blockage of water flow leads to hydraulic and symplastic failure that may result in plant desiccation and even mortality (Choat et al., 2018; Sperry & Tyree, 1988). Structural traits such as organ volume and water storage capacity strongly influence hydraulic functionality to withstand such stress (Brodribb & Cochard, 2009; Hubbard et al., 2001; Powers et al., 2020; Sperry et al., 2008). In trees, the long transport pathway between roots and canopy increases the risk of hydraulic failure, as greater height is associated with more negative water potential required to maintain water flow during drought (McDowell et al., 2008).

To prevent cavitation, plants actively regulate stomatal aperture on the surface of the leaves. Stomata sense both leaf water potential and environmental cues such as light, temperature and humidity, and respond dynamically by adjusting their

aperture accordingly. During drought conditions, closure of the stomata reduces transpiration, protecting the xylem from cavitation, but this simultaneously limits CO_2 uptake, reducing photosynthetic carbon assimilation and growth and potentially leading to carbon starvation (Anderegg et al., 2019; Cowan & Farquhar, 1977; Hubbard et al., 2001). The timing and degree of stomatal closure vary across species, reflecting different drought survival strategies that likely stem from evolutionary trade-offs (Brodribb & McAdam, 2013; Brodribb et al., 2014). In anisohydric species, stomatal closure occurs after a prolonged decline in water potential and such species are, therefore, more susceptible to hydraulic failure (Sade et al., 2012). Their greater degree in stomatal plasticity allows anisohydric species to occupy drought-prone areas as they can maintain a higher productivity (Almeida-Rodriguez et al., 2010). In contrast, stomatal closure in isohydric species occurs early and thereby maintains relatively stable leaf water potentials, avoiding critical xylem tensions and cavitation during drought conditions. Scots pine is a typical example of this isohydric hydraulic behaviour, maintaining stable needle water potential even during soil drying (Irvine et al., 1998; Salmon et al., 2015). Similarly, Norway spruce displays strongly regulated stomatal conductance and limited hydraulic plasticity, consistent with its high vulnerability to xylem embolism (Arend et al., 2021; Kurjak et al., 2012).

Stomatal behaviour is tightly coupled to ABA-signalling, a central hormonal pathway in drought response (Yamaguchi-Shinozaki & Shinozaki, 2006). Under declining water availability, ABA accumulates in guard cells, where it modulates ion fluxes through plasma membrane channels, leading to turgor loss and stomatal closure (Murata et al., 2001; Pei et al., 2000; Schroeder & Hagiwara, 1990). The exact mechanism that causes drought-induced ABA accumulation is still unclear, but several models in different plant species have been proposed. These include the reduction of ABA catabolism (Z. Wang et al., 2018), increased transport of ABA through the shoot (Long et al., 2019), and the hydrolysis of conjugated ABA-GE (ABA-glucose ester) by beta-glucosidase AtBG1 (Lee et al., 2006). After sufficient accumulation, ABA binds to the Regulatory Component of ABA Receptors (RCARs)/PYR/PYL (Pyrabactin Resistance 1/ PYR1-like), inhibiting group A 2C-type protein phosphatases (PP2Cs). This inhibition allows activation of Subclass III SNF1-Related protein Kinases 2 (SnRK2s), which in turn phosphorylate and activate ABA-Responsive Element Binding proteins AREB/ABF (ABRE binding factor). These AREB/ABFs bind to ABA-Responsive Elements (ARBEs) in promoters of stress-inducible genes, initiating the transcription of downstream drought responsive genes (**Figure 3**) (Joshi et al., 2016; Nakashima et al., 2014; Singh & Laxmi, 2015).

In parallel to this ABA-dependent pathway, ABA-independent pathways induce other Cis-Regulatory Elements, such as DRE/CRT, to promote drought-responsive genes (Kasuga et al., 1999; Shinozaki & Yamaguchi-Shinozaki, 2000). DRE/CRT is specifically activated by the cold-response DREB/CBF regulons, DREB2A and

DREB2B, which act under osmotic stress (Agarwal et al., 2006; Nakashima et al., 2000). These interactions highlight the cross-talk between pathways involved in drought, salinity and cold stress responses (Roychoudhury et al., 2013; Seki et al., 2002). Additional TFs, including MYB/MYC, WRKY, NAC, and NF-Y families, act either downstream of ABA signalling cascades or in coordination with independent osmotic stress pathways. They are involved in the synthesis of osmolytes, detoxification enzymes, and structural proteins such as Heat Shock Proteins (HSPs) and Late Embryogenesis Abundant (LEA) proteins that stabilize cellular components under dehydration stress (Singh & Laxmi, 2015). Collectively, these transcriptional networks form an integrated regulatory system that enables plants to perceive, signal and mitigate drought induced damage.

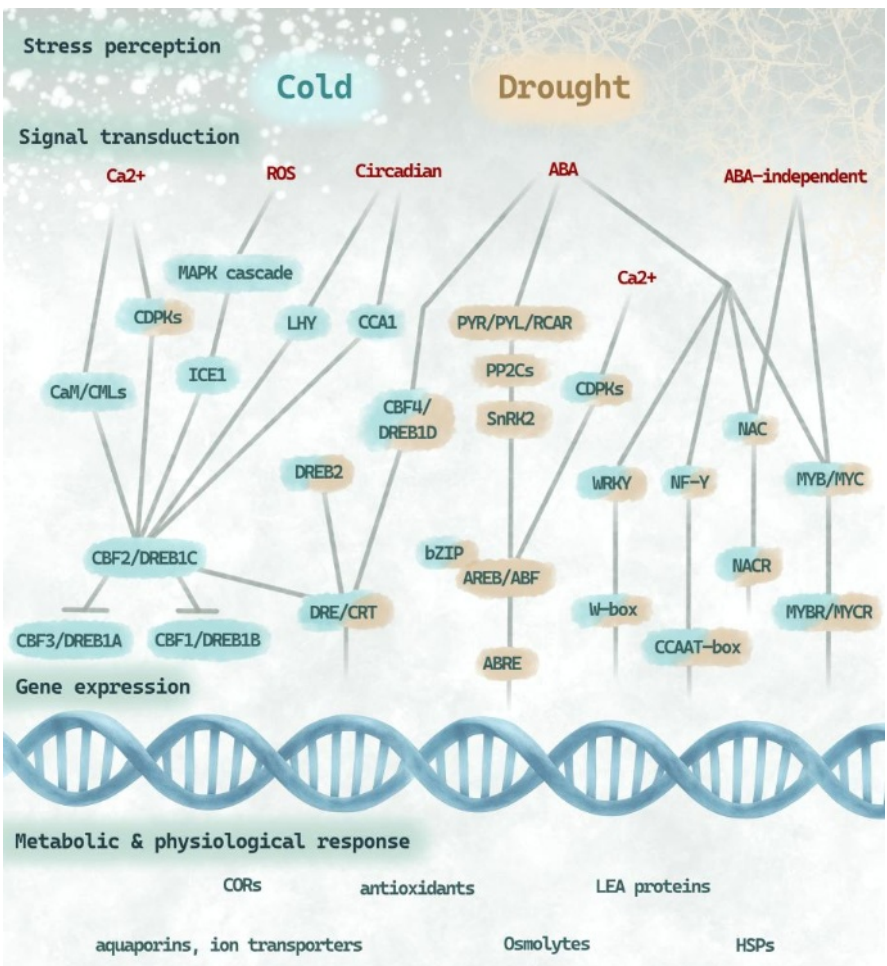


Figure 3 Summary of key regulators activated during cold and drought stress in plants. Based on Hirayama & Shinozaki, 2010; Raza et al., 2020; Singh & Laxmi, 2015.

In summary, abiotic stress responses arise from the coordinated activity of numerous regulatory components, many of which act in both cold and drought conditions while others function in a more stress-specific and also species-specific manner. Understanding how these components interact, and how they differ or are conserved across different tissues or species, requires approaches that move beyond single-gene analyses. GCNs offer a powerful framework for identifying groups of genes that act together under specific conditions and for highlighting potential key regulators that are unique or in common. When combined with comparative network analyses, GCNs also provide a way to evaluate how expression patterns are conserved across species, allowing for the potential identification of a conserved response to abiotic stress. Applying these methods to conifers is particularly compelling given their large and newly annotated genomes, long life cycles, and remarkable resilience to environmental stressors. By integrating GCN-based approaches, this thesis aims to advance our understanding of the regulatory architecture of abiotic stress responses in Norway spruce and Scots pine, and provide a foundation for interpreting how these regulatory systems contribute to conifer adaptation and stress tolerance.

Materials and Methods

This study involved multiple species, however for this thesis the primary focus was on the conifers *Picea abies* (L.) H. Karst (**Paper I – IV**) and *Pinus sylvestris* (L.) (**Paper II – IV**)**.** Other species used during the thesis include *Arabidopsis thaliana* of the Col-0 (Columbia) accession, *Arabidopsis thaliana* of the Ost-0 (Östhammar) accession, *Populus tremula*, and *Betula pendula*, which are all part of **Paper III**.

The abiotic stress conditions the plants were exposed to were replicated for each species in **Papers I** to **IV**. In **Paper III**, data were derived from the first four cold stages of the experimental design in **Figure 4** and replicated for each species, focusing exclusively on root tissue.

Detailed descriptions of the experimental design, RNA extraction, -sequencing and -processing, quality reassurance, and paper-specific analyses are found in the respective papers. This thesis chapter only discusses the core methods to provide sufficient context.

Exposure to abiotic stress conditions

In the drought experiment, Norway spruce and Scots pine seedlings were grown under long-day conditions (16 h light, 20 °C), with soil moisture monitored as field capacity (FC), with 80% FC representing well-watered control conditions. Drought stress was induced by withholding water, reducing FC to 60%, 40%, and 30% (**Figure 4A**). The 30% FC condition, considered a mild, non-lethal drought, was maintained for seven days (30%7d) with limited watering. Water was then withheld entirely to induce severe drought, characterised by visible photosynthetic dysfunction (Collapsed) and maintained for two days (C2d). After experiencing the severe drought conditions, seedlings were rehydrated back to 80% FC. Needle and root samples were collected from the most recent shoot and from fine roots of the seedlings at each of these drought stages.

In the cold stress experiment, control samples were taken after a two-week growth period at 18 °C/ 15 °C light/dark long-day conditions (16 h light, 20 °C). Seedling were gradually acclimated from 18 C ° to 5 °C still maintaining a 16-h long-day period (**Figure 4B**). Needle and root samples were then collected after 6 h, 24 h, 3 days, and 10 days of exposure to 5 °C. Following acclimation, seedlings were transferred to -5 °C, and another set of samples were taken after 6 h, 24 h, 3 days, and 10 days.

In both experiments, samples were frozen with liquid nitrogen immediately after collection, and stored at -80 °C before RNA-extraction.

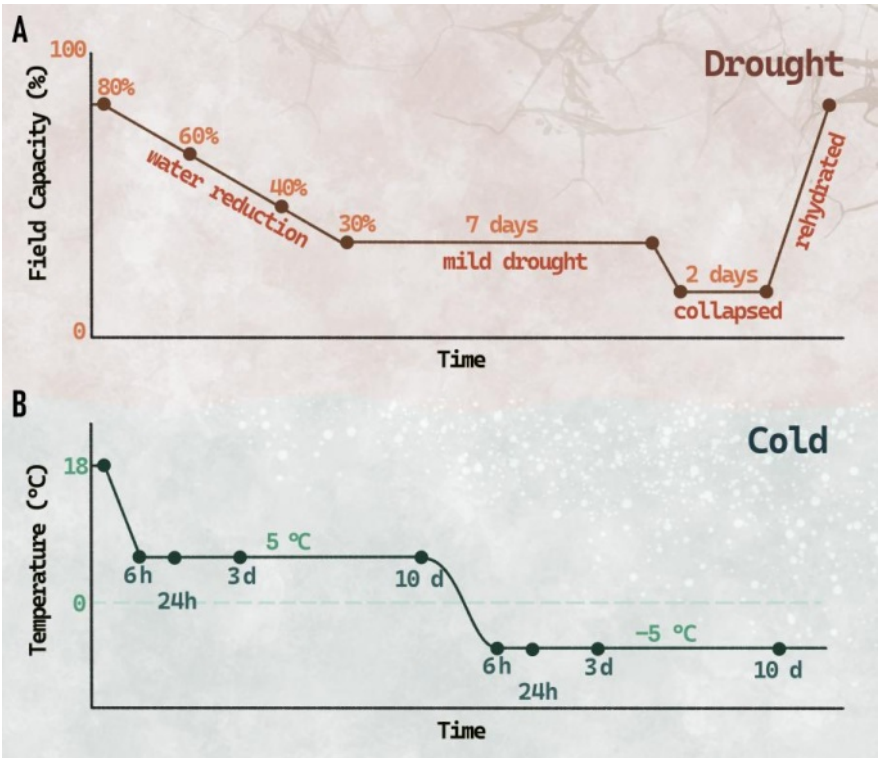


Figure 4 Experimental design of drought and cold experiments in Norway spruce and Scots pine. Replicated samples were taken at each dot on the line-plot. A) Experimental design of drought experiment. B) Experimental design of cold experiment.

From stress to sequences: Transcriptome construction

RNA was extracted from the abiotic stress samples using the CTAB method (Chang et al., 1993) with the following modifications: 1 mg/mL of proteinase K was added to the lysis buffer, ground samples were incubated for 10 min at 65 °C before adding the Chloroform:Isoamyl Alcohol (Ch:IAA). Samples were precipitated for 1 h by addition of 2X volume of EtOH at 4 °C, after which RNA was harvested by centrifugating at 14.000 rpm for 20 min at 4 °C. Purification of the samples was performed using the RNeasy mini kit (Qiagen, Hilden, Germany). Additionally, DNase Digestion was performed using the RNase-free DNase set (Qiagen) with a modified incubation time of 30 min. After extraction all samples were assessed for quality before being sent for sequencing at Novogene.

Paired-end (2 x 126 bp) sequencing was performed using an Illumina HiSeq 2500 platform. RNA-Seq data was pre-processed and analysed as follows. The quality of the raw sequencing reads was first assessed using FastQC (Andrews, 2010).

Ribosomal RNA (rRNA) contamination was then removed using SortMeRNA v1.9 (Kopylova et al., 2012) with the default rRNA sequence databases provided with the software. Sequencing adapters were removed, and low-quality bases trimmed using Trimmomatic v0.32 (Bolger et al., 2014). A second quality assessment with FastQC confirmed no technical artefacts were introduced during pre-processing.

Transcript quantification was performed using Salmon v0.11.3 (Patro et al., 2017), performing pseudo-alignment the pre-processed reads to the updated Norway spruce and newly generated Scots pine reference genomes (**Paper II**). The resulting output was subsequently processed to generate count data, giving quantitative information on transcript abundance summed at the gene level. These count data were assessed by using Principal Component Analyses (PCAs) to visualise sample relationships. Finally, read counts were normalised using the Variance-Stabilizing Transformation (VST) implemented in the DESeq2 package v1.16.1 (Love et al., 2014), correcting for the mean-variance relationship between samples.

The VST count data facilitated downstream analyses such as differential expression analysis of the genes per experiment, including the visualisation of expression in heatmaps. Differential expression analyses were conducted for all experiments using a standardised workflow to ensure comparability between treatments and species. The default thresholds set for the selection of differentially expressed genes was at a *P*-adjusted value below 0.01 and an absolute $\log_2$ fold change of 2. VST values were also used for co-expression network analyses, with additional filtering steps applied where required for specific methods, as detailed in respective chapters.

Seiðr: A wisdom of the crowd approach to network inference

The Seidr toolkit efficiently generates crowd (aggregate) co-expression networks using a collection of well-established inference algorithms (**Paper I**). A VST count matrix filtered to exclude genes with zero-variance, was used as input. A number of different inference algorithms were selected for network inference. The outcome of each algorithm can vary, as some might produce networks that have edges with edge weights of 0, whereas some methods might produce a fully dense network containing an edge weight for each possible edge in the network.

Each algorithm produces an individual network, which can be aggregated into a consensus (crowd) network using the Inverse-Rank Product (IRP) method (Zhong et al., 2014) (**Figure 5**). This non-parametric approach mitigates differences in edge-weight distributions among networks by ranking edges within each network independently (with the strongest edge ranked as 1) and calculating a combined IRP-score for every edge across all methods. This results in a fully dense aggregated

network, meaning each possible connection in the network has a score, therefore it also includes a high degree of technical noise.

To reduce this noise, Seidr implements the backbone method (Coscia & Neffke, 2017), which assumes a binomial distribution of edge weights and applies arbitrary cut-offs (typically 1 to 10%) to retain only the top-ranking edges. Networks generated at each cut-off are usually evaluated for quality by using a golden standard, comparing known True Positive (TP) and True Negative (TN) edges to those retained in the backbone network, and summarising performance with an Area Under the Curve Ratio (AUCR) score. However, since no such gold standards were available for Norway spruce and Scots pine, network quality was instead assessed using Seidr's internal threshold method, which evaluates network properties such as the number of edges and nodes, the R^2 fit of the degree distribution to a scale-free model, and the average clustering coefficient. This method should however be used with caution since scale free networks are rare, or at least their prevalence is debated (Broido & Clauset, 2019). Additionally, Seidr can calculate various centrality metrics on the filtered network to give a measure of similarity between nodes or edges.

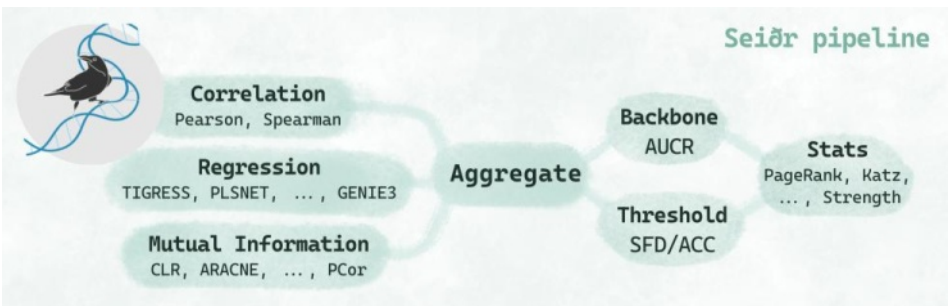


Figure 5 Seiðr pipeline. Starting with the various inference algorithms, one can pick and choose which ones to use. Each individual network is then aggregated with seidr aggregate and one of two threshold method can be applied. Seidr stats will then calculate centrality metrics on the filtered network.

Comparing species with ComPlEx

Comparative co-expression network analysis to identify functionally conserved gene relationships across species, was performed using ComPlEx (Comparative analysis of Plant co-Expression networks; Netotea et al., 2014). For this method, VST matrices were filtered to exclude genes with zero-variance and genes below an arbitrary density threshold. Individual co-expression networks were first generated for each species by calculating pairwise gene expression correlations using Pearson correlation coefficients. In these networks, genes were represented as nodes, and edges indicated the strength of their co-expression. To enhance correlation robustness and minimise the effect of outlier values, Mutual Ranking (MR) was applied (Obayashi et al., 2009), following the same pipeline as used by PlaNet (Proost & Mutwil, 2017). Networks were subsequently pruned by applying a density threshold of 0.03, retaining only the top 3% of the strongest connections.

Orthologous gene pairs were identified using OrthoFinder v2.5.2 results (Emms & Kelly, 2019) from **Paper II**, where genes were assigned to their orthogroup based on protein sequence similarity. To assess the degree of regulatory conservation, each orthologous gene pair (A^{PA} and A^{PS}), the co-expression neighbourhood of gene A^{PA} in Norway spruce was mapped to its corresponding orthologous neighbours of gene A^{PS} in the Scots pine network (**Figure 6A**). Overlap between the orthologous neighbourhoods of gene A^{PA} and A^{PS} was evaluated statistically using a bidirectional hypergeometric test, quantifying the likelihood that the observed overlap occurs by chance. This produced two *P*-values, one for each species background network. Following a False Discovery Rate (FDR) correction, gene pairs were classified as co-expressologs only when both adjusted *P*-values were below 0.1, indicating significant reciprocal conservation of co-expression patterns and neighbourhoods. In cases where no orthologous neighbours were found, the test was omitted, and a default *P*-value of 1 was assigned.

The resulting ortholog comparisons were compiled into a summary file listing gene identifiers, number of neighbours, shared orthologous neighbours, and associated FDR corrected *P*-values. In the case of more than two species comparisons (**Paper III**), co-expressologs consistently showing significant co-expression overlap across all pairwise species comparisons were grouped into cliques (**Figure 6B**), reflecting conserved regulatory modules. As multiple cliques could be associated with a given orthogroup, one representative clique was selected for further analysis. Selection was based on the lowest *P*-adjusted value mean of a clique per orthogroup, prioritising the strongest overall co-expression.

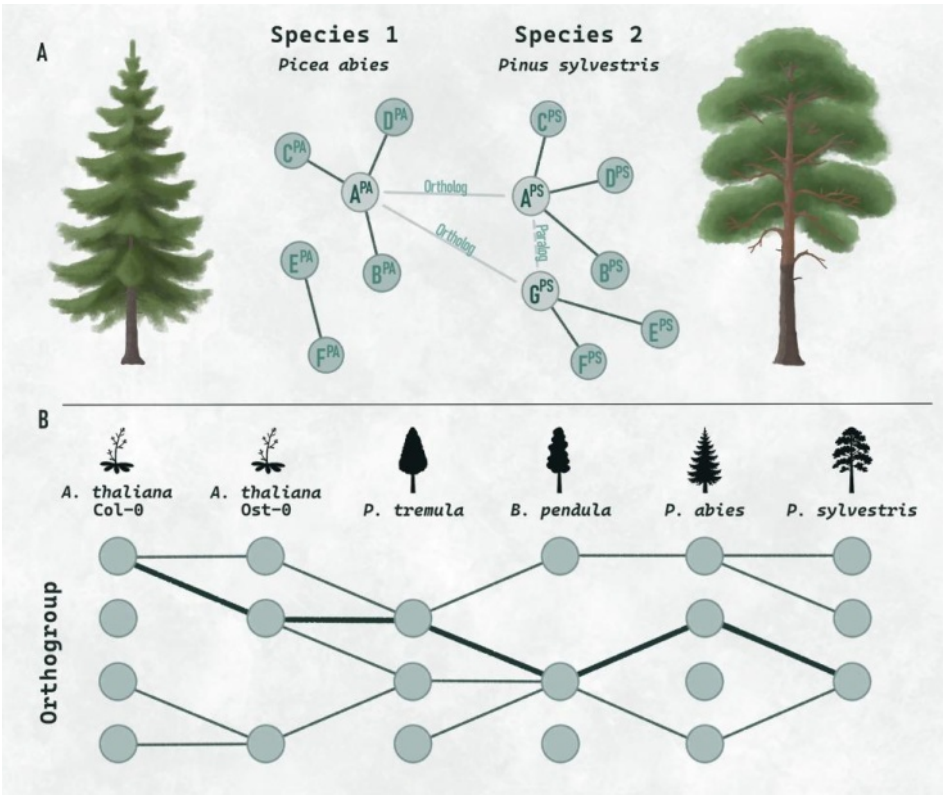


Figure 6 Conceptual representation of the Comparative analysis of Plant Co-Expression networks (ComPlEx) algorithm. A) Representation of a pairwise comparison where gene A^{PA} forms a co-expressolog with A^{PS} and G^{PS} in the other species. This makes A^{PS} and G^{PS} paralogs of each other. B) For a given orthogroup co-expressologs can form cliques when genes connect across all species. The strongest connection based on the lowest mean *P*-adjusted value is then kept to represent this given orthogroup, representing the thickest connecting line here.

Results and Paper discussion

The aim of this thesis was to provide novel insights into the conservation of abiotic stress response regulation in boreal tree species, with a particular focus on Norway spruce and Scots pine. Knowing the extent of this conservation is increasingly important as boreal forests face rapid environmental change, including more frequents droughts, increasingly early ice formation and extreme weather events (Rantanen et al., 2022). Identifying conserved regulators, hubs, and gene modules responsive to abiotic stress therefore provides a foundation for future breeding and biotechnology strategies and provides valuable insight as to whether model species, such as Arabidopsis, can be used to translate knowledge to boreal tree species.

The results presented in the work included in this thesis address these objectives by examining transcriptional responses across species, identifying conserved and divergent regulatory components, and characterising network structures associated with abiotic stress response.

A demonstration of the importance of data quality

In **Paper I** we were working with the old genome annotation of Norway spruce (v1, (Nystedt et al., 2013). **Paper II** resulted in the new and updated version of the Norway spruce genome. To measure the influence of the annotation difference on network inference, Seidr was used to infer two GCNs from drought exposed root data, selecting Pearson correlation, CLR, and PLSNET to get a wide coverage of methods. Centrality metrics were calculated for both network and non-TFs and TFs from each network were compared (**Figure 7**).

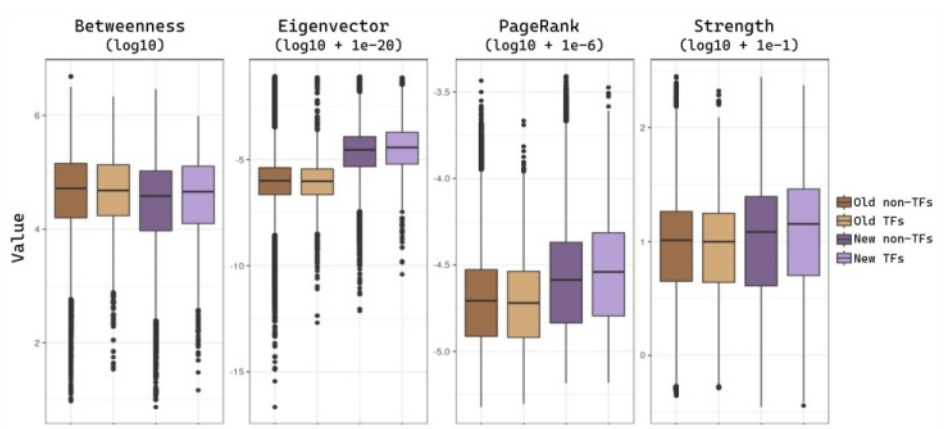


Figure 7 Boxplots of centrality metrics between network inferred from data aligned to the old version of the Norway spruce genome (v1, 2013) and the new version (v2, **Paper II**).

Centrality metrics of both non-TFs and TFs in the network based on the new annotation exhibited higher values of centrality for eigenvector, PageRank, and strength metrics. This might indicate that reads previously mis-aligned under the old annotation were correctly mapped to the new annotation, yielding more accurate expression estimates for affected genes. As a result, the GCN generated from the new annotation more effectively captured the highly interconnected nature of genuine co-expression relationships. Betweenness values were slightly lower in the new annotation as compared to the old annotation, which reflects a reduction in artificial gene-bridges between different network modules and indicates a more cohesive network structure.

The complexity of ComPlEx

Constant quality assurance is an important factor to obtain biologically informative results throughout any project, but even then, some noise slips through the cracks.

Genes with flat expression profiles ended up being integrated in the set of co-expressologs after initial runs of ComPlEx on data from **Paper III**. Presumably, flat expression still maintained high correlation across data sets in addition to having highly overlapping orthologous neighbourhoods. Unfortunately, these initial results contained a considerable amount of technical noise, which obscured the underlying biological signal. To address this issue and ensure that co-expression patterns reflected genuine biological relationships rather than artifacts, additional filtering steps were applied.

A density distribution filter was used to remove genes with sparsely distributed expression values, which likely represented noise or unreliable measurements. In addition, a standard deviation filter was applied to exclude genes with minimal expression variability across samples. After applying these filters and rerunning ComPlEx, the resulting co-expressolog expression patterns displayed clearer and more biologically coherent patterns and genes with zero-variance and flat expression were removed (**Figure 8**).

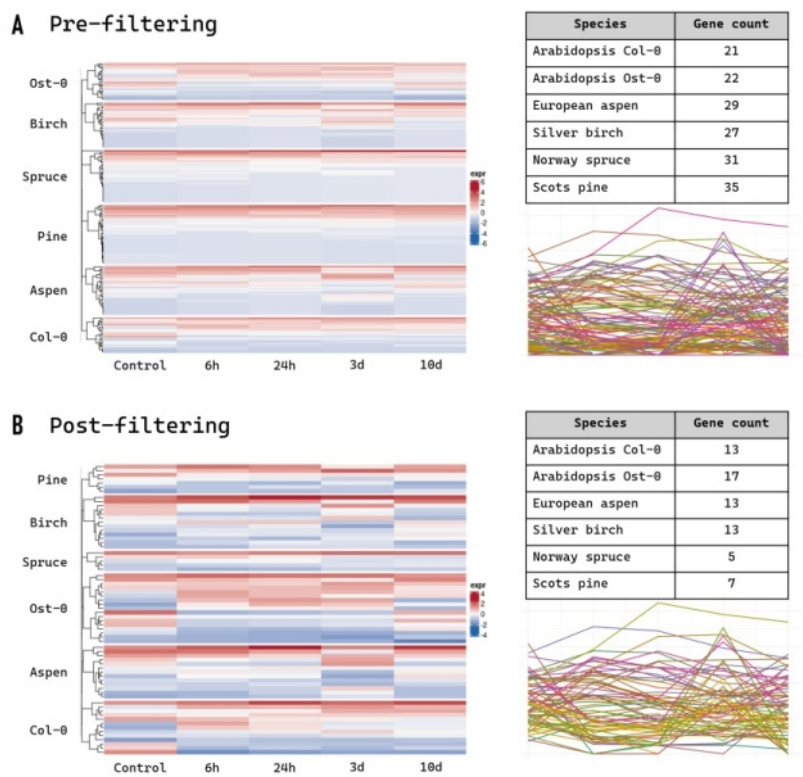


Figure 8 Filtering effect on *C-repeat Binding Factor* (*CBF*) co-expressologs. Additional filtering steps show change in gene-composition of orthogroup OG0000037. A) Heatmap, count-table, and median expression line plot of co-expressologs before filtering. B) Post-filtering effect on heatmap, count-table, and line-plot.

Even with additional filtering and noise reduction, deriving biological meaning from the data remained a challenge due to the variety of species and their differences in expression dynamics. While co-expressologs might vary extensively in expression patterns between species, their gene-neighbourhoods in terms of gene-families or orthogroup annotation was highly similar. Functional enrichment of the co-expressologs using Gene Ontology revealed that many carried similar, and in some cases identical, annotations across species. The clique set from **Paper III** was another proof of this concept, as enrichment results of the genes across species overlap.

Adding extra layers to the co-expressologs, such as identifying which ones were TFs and DEGs, also help to derive biological meaning from the data. However, not all of the DEGs could be included in the comparative network analyses, as some did not have a shared orthogroup annotation and were species-specific. Unfortunately, this was one of the limitations of the comparative analysis.

Paper I – Ensemble network inference revealed key regulators of drought response in Norway spruce

Inferring networks from expression data is a powerful way to study complex biological processes, but individual network inference methods often produce biased or incomplete results due to their specific assumptions and limitations. In **Paper I**, Seidr was developed to integrate the results of several inference algorithms into one consensus network, obtaining the wisdom of the crowd. This outperformed individual methods by reducing methodological bias, improving sensitivity, specificity, and precision (**Paper I, Figure 1**). We demonstrated the use of this wisdom of the crowd principle to decipher the transcriptional regulation of drought response in Norway spruce as an example of how the method is equally applicable to non-model species.

The performance of Seidr was assessed by generating a drought-specific and an unstressed network. Centrality analyses revealed that genes associated with drought stress were highlighted only in the drought network, emphasising the value of experiment-specific networks. Partitioning the network with InfoMap identified 16 modules (**Paper I, Figure 3**), three of which were significantly enriched for stress response related GO terms. One module in particular contained six TFs with an expression peak at FC30, two of which displayed a second expression peak at the Collapsed state (**Paper I, Figure 4**). These two TFs were both NAC-related TFs, with MA_103341g0010 annotated as a NAC-NAM, ATAF1/2 and CUC2-domain-containing 35-like, and MA_10426365g0010 as a NAC domain-containing 86-like TF. Homologs of both genes have been shown to have a link to abiotic stress response in various plant species (Cai et al., 2025; Fang et al., 2008; Guérin et al., 2019; Le et al., 2011; Ooka et al., 2003; Su et al., 2013). However, none of the six TFs were among the genes in a previously identified list of drought response genes. This was not surprising as these annotations were lifted from Arabidopsis, and thus may miss spruce-specific regulators. Rather than being a limitation, this highlights the potential for uncovering novel spruce-specific stress responses. In addition to these TFs, we used network neighbourhood analyses and centrality metrics to infer possible functions for previously unannotated genes. The first high-centrality gene, MA_10432675g0010, was annotated as a cellulase, and its neighbours were predominantly enzymes involved in cell wall modification, suggesting a role for cell wall reorganisation under drought stress. The second unannotated gene, MA_69177g0010, had a neighbour annotated as MAPK, indicating it may function within stress signalling pathways mediated by MAPKs. The third, MA_10192193g0010, was positioned near a photolyase and a CONSTANS-like gene, implying a potential role in ROS detoxification and maintenance of photosynthesis under drought conditions, paralleling functions reported for CONSTANS-like genes in photoperiod regulation and stress resistance.

Paper II – Co-expression networks provide biological insight into genomic features

Conifer genomes are among the largest in the plant kingdom, being mostly composed of repetitive non-coding DNA. Genes can be separated by large intergenic regions and can even contain very long intron sequences (>15Kbp) within the gene itself. One naturally starts to wonder how conifer genomes are organised in terms of expression, considering these genomic characteristics.

The three-dimensional organisation of the genome plays a crucial role in regulating gene expression, and two key structural features underpin this architecture: Topologically Associating Domains (TADs) and chromatin loops (van Steensel & Furlong, 2019). TADs are contiguous genomic regions in which sequences interact more frequently with each other than with neighbouring regions, creating isolated domains of regulatory activity (Bonev & Cavalli, 2016). These domains help insulate enhancers and promoters, ensuring that regulatory interactions occur within defined boundaries. Within TADs, chromatin loops bring distant genomic elements into close contact, allowing enhancers, silencers, and promoters to interact despite being far apart in linear sequence (Tourdot & Grob, 2023). Loop formation is often mediated by architectural proteins, which facilitate loop extrusion and stabilise long-range interactions. Together, TADs and chromatin loops create a hierarchical and dynamic chromatin landscape regulating gene expression. While these features have been established in some plant species (Tian et al., 2021; M. Wang et al., 2018; Xie et al., 2019), **Paper II** provides the first insights into this structural organisation in conifers.

In **Paper II** we showed that TADs were encompassed by one or two genes on either side of the TAD (TAD borders). Interestingly, genes at TAD borders generally had longer intron sequences, and each TAD typically contained around 3 to 10 genes. To see if genes at TAD borders, genes within a TAD or gene-gene pairs acting as chromatin loop anchors had any co-expression signal, a GCN was inferred from seasonal needle data in Norway spruce, which contains expression values from growing needles at different seasonal timepoints. Co-expression correlation values of different classes of gene-pairs were compared to see if there was any biological pattern. Gene-pairs classed as loop anchors showed the highest co-expression correlation, whereas gene-pairs at TAD borders, inside a TAD, or gene-pairs from the TAD border to inside the TAD, showed weak co-expression correlation values (**Paper II, Figure 2f**). This suggests that the physical interaction of loop anchors helps facilitate a high correlation of co-expression. To benchmark these results, co-expression correlation values of a random set of gene-pairs proportional to the investigated set were compared.

To investigate whether TADs across Norway spruce and Scots pine followed similar regulation, conserved TADs were identified by taking the orthologuos gene pairs at each TAD border and comparing them across species. Combinations of orthogroups of the genes at these TAD borders that matched between species were considered conserved, with an additional sanity check confirming that these TADs occupied corresponding syntenic genomic regions in each genome. Using ComPlEx, co-expressologs were identified between Norway spruce and Scots pine abiotic stress datasets. The most significant pairings of co-expressologs per orthogroup were retained and subsequently assessed to determine whether they were located within equivalent conserved TADs. It was found that when the co-expressolog of one species was located within a conserved TAD, the corresponding co-expressolog in the other species was located in the same conserved TAD in 87% of cases. Co-expressologs located within conserved TADs were assessed for network centrality and compared to co-expressologs not within conserved TADs (**Figure 9**).

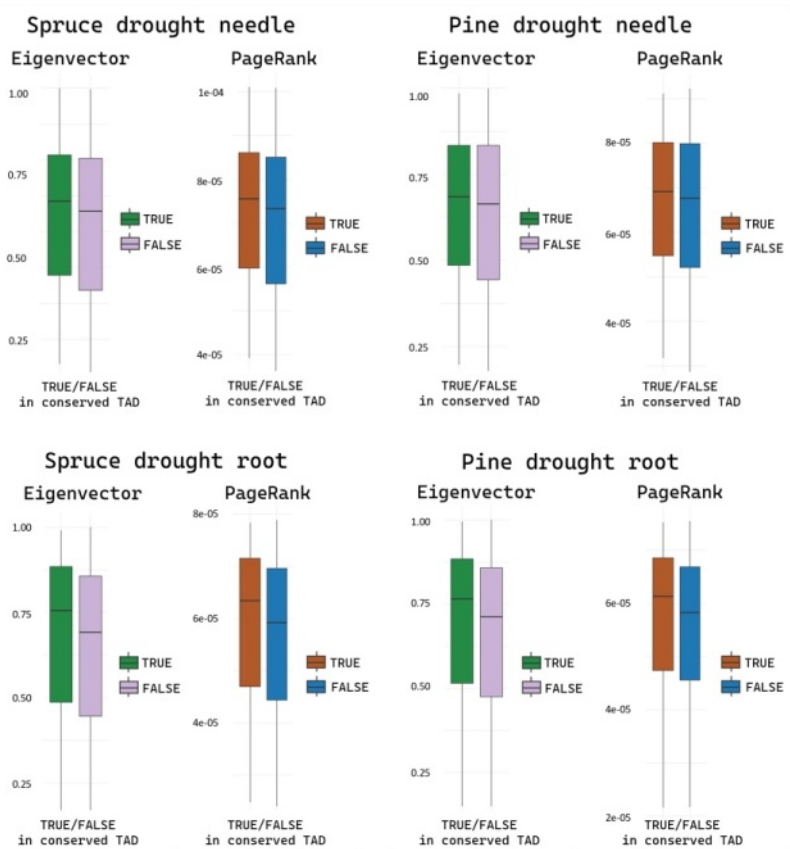


Figure 9 Eigenvector and PageRank boxplots of co-expressologs found in conserved and non-conserved Topologically Associating Domains (TADs). Generated from drought stress data in Norway spruce and Scots pine.

Across the different abiotic stress GCNs, co-expressologs within conserved TADs consistently displayed a higher degree of centrality, suggesting that these genes could act as main regulators (**Figure 9**). Centrality boxplots from co-expressologs generated from cold stress data displayed a slightly weaker signal in Scots pine needles and Norway spruce root tissue. Considering the centrality calculation was restricted to genes that were part of orthogroups present in both Norway spruce and Scots pine and needed to meet the expression threshold requirements, these results indicate that genes of high importance within conserved TADs have conserved co-expression and are highly connected within abiotic stress co-expression networks, suggesting that they are of evolutionary importance.

The expression profiles of the co-expressologs in two of these conserved TADs were visualised to get an indication of how these genes were regulated (**Figure 10**). Genes in **Figure 10A** very clearly demonstrate that co-expressolog genes within a conserved TAD do not necessarily follow the same expression profile between species despite having conserved co-expression neighbourhoods, in line with the results reported in **Paper III**. The same was true for genes within one TAD of the same species, suggesting that they are not co-regulated. The up-regulation of several genes during more severe drought conditions in **Figure 10A**, suggests that genes within this TAD may play an active role in drought response. While no significant enrichment of GO terms was found for the genes within this TAD, the genes were annotated for processes involved in “cytokinesis" and other metabolic and developmental processes. Additionally, several genes were annotated to GO terms for stimulus response. Similarly, no significant enrichments were found for the genes down-regulated during cold stress in the conserved TAD depicted in **Figure 10B**. However, these genes were annotated to several abiotic stress related GO terms, such as “response to cold” and “response to abiotic stimulus”. Additionally, genes within this TAD were also annotated to GO terms associated with development, consistent with genes within the conserved TAD represented in **Figure 10A**. All data and results from **Paper II** were integrated within the Plant Genome Integrative Explorer (PlantGenIE; (Sundell et al., 2015) to ensure compliance with the Findable, Accessible, Interoperable, and Reuseable (FAIR) principles (Wilkinson et al., 2016). The results presented here have not yet been integrated in the preprint version of **Paper II**, but will be part of the published work.

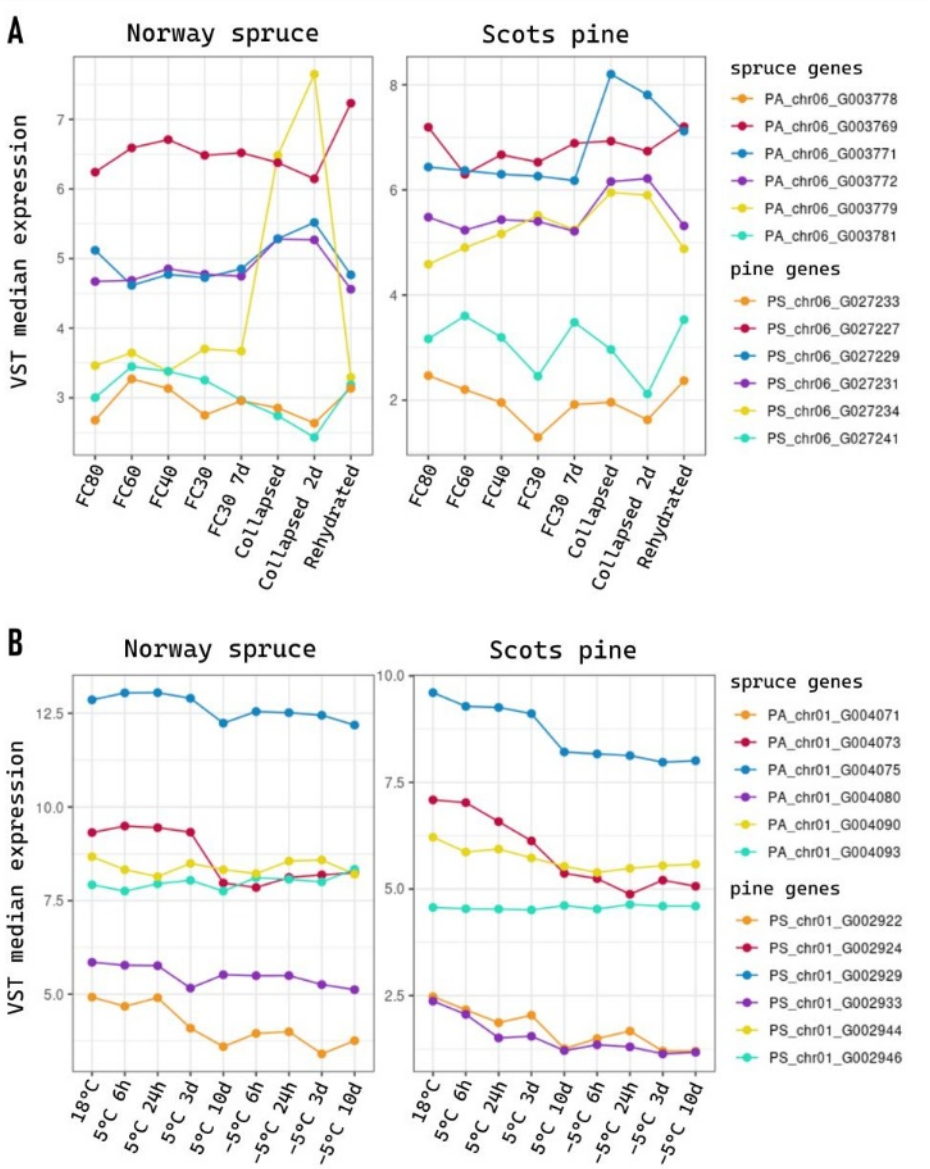


Figure 10 Co-expressolog expression of conserved Topologically Associating Domains (TADs). Equivalent co-expressolog pairings share the same colour across Norway spruce and Scots pine plots. A) Variance Stabilizing Transformation (VST) median expression of co-expressologs in needles during drought stress exposure in conserved TAD OG0015698_OG0015709. B) VST median expression of co-expressologs in needles during cold stress exposure in conserved TAD OG0009457_OG0009477.

Paper III – Conserved core rewired to regulate different set of genes as response to cold stress in roots

Regulation of abiotic stress responses in plants differ extensively between species, therefore very little is known about the conserved gene regulation of these responses. Plant research has historically focussed less on root tissue as roots are generally harder to access, manipulate and image, as compared to above-ground tissues. However, with increasing temperatures due to climate change, roots are more often exposed to ice formation as the insulating snowpack becomes thinner during spring (Richardson et al., 2024). In **Paper III** we looked at the cold response regulation across four boreal tree species, two angiosperms (aspen and birch) and two gymnosperms (Norway spruce and Scots pine). Additionally, we included the herbaceous model species Arabidopsis Col-0 along with a more northern ecotype, Arabidopsis Ost-0, to enable functional annotation based on the guilt by association principle and as a reference point to the extensive body of literature on cold response of, primarily, aboveground tissue responses of Arabidopsis.

Temporal expression differences were observed between the DEGs of the five species, which appeared to be lineage-specific. The two Arabidopsis ecotypes generally displayed a quicker reaction with up- or down-regulation of genes already after 6 hours of cold exposure. The two broadleaf trees, aspen and birch, generally displayed an up- or down-regulation spike after 3 days of exposure, and the conifers, Norway spruce and Scots pine, generally showed a late response at 10 days.

To analyse whether this response involved the same gene families, we translated all expressed genes and all DEGs into their respective orthogroups. A high overlap of orthogroups across all expressed genes was observed (**Paper III, Figure 1A**), representing approximately half of the expressed genes per species (**Paper III, Figure 1B**). Low overlap among DEG orthogroups representing 13-18% of the total DEGs and ~6% of all expressed genes (**Paper III, Figure 1C & 1D**), indicated that species regulate largely distinct gene sets in response to cold exposure.

By integrating comparative network analysis with ComPlEx, orthogroup and expression information was used to identify co-expressologs between every pairwise species comparison. The orthogroup overlap between all pairwise comparisons revealed the highest overlapping sets to be between the Arabidopsis ecotypes (1579 orthogroups), the conifer trees (612 orthogroups), and the broadleaf trees (293 orthogroups) (**Paper III, Figure S4**). Genes within each of these pairwise sets were independently enriched for stress-responsive GO terms, suggesting each lineage has their own way of dealing with abiotic stress. Across all pairwise comparisons an overlap of 181 orthogroups was observed. Enriching the genes within this set also revealed stress-responsive GO terms, in addition to many developmental processes. From this common set we identified a core set of genes (a

subset of 116 orthogroups) forming cliques with conserved cold-responsive co-expression across all species. While these genes were identified as having conserved co-expression neighbourhoods, their expression profiles still appeared lineage-specific (**Paper III, Figure 5**). This highlights the power of the comparative co-expression approach to identify conserved regulatory relationships despite lineage, or species-specific differences in expression response timing. Eight of these cliques were composed entirely of TFs, four of which were part of an expression-profile-based cluster that was functionally enriched for "response to cold stress" (**Figure 11**).

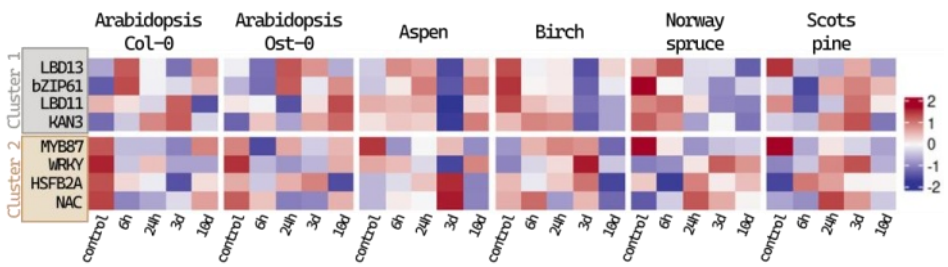


Figure 11 Heatmap of Transcription Factor (TF)-cliques subset. This heatmap represents the TF specific subset of cliques from **Paper III**, with z-score scaled median Variance Stabilizing Transformation (VST) expression values. Cluster 1 contains TFs annotated for tissue development and growth processes. Cluster 2 contains the set of TFs annotated for 'response to stress'.

TFs in Cluster 2 had direct involvement in stress response based on literature evidence. A MYB, specifically *MYB87*, was differentially down-regulated in the two Arabidopsis ecotypes and its ortholog in Norway spruce. MYB87 has been shown to promote longitudinal growth in several plant organs, including roots (Fujiwara et al., 2014). This suggests that in Arabidopsis growth was initially supressed and later reactivated, whereas in Norway spruce, the *MYB87* co-expressolog remained down-regulated from 24 hours onward, potentially indicating that reduced growth is an important and sustained adaptive mechanism for cold tolerance in Norway spruce. Two distinct WRKY TFs were identified in the Arabidopsis ecotypes, *WRKY48* in Col-0 and *WRKY23* in Ost-0, of which only *WRKY48* was differentially expressed in Col-0, and its aspen and Scots pine co-expressologs. WRKY48 has been shown to play a role in biotic stress response (Xing et al., 2008), and in response to combined high light and heat stress (Balfagón et al., 2024), where it functions as a repressor of genes that promote acclimation to these stressors. While both previous studies were conducted in leaf tissues, WRKYs act broadly across plant organs. The expression pattern observed in **Figure 11** may be consistent with these findings, potentially indicating a similar, and conserved, role for *WRKY48* in root tissue in response to cold stress. An initial down-regulation could facilitate activation of early stress-response pathways, followed by a later up-regulation as Arabidopsis Col-0 and

aspen transition towards acclimation. However, further functional evidence would be required to confirm this interpretation, especially since the opposite expression pattern was observed in Scots pine. Heat Shock Factor (HSF) proteins are known to be key regulators involved in abiotic stress response (M. Guo et al., 2016), although not much is known about HSF2BA specifically. *HSFB2A* and its co-expressologs were differentially expressed across all species, generally showing down-regulation at 6 hours and an up-regulation later on, being down-regulating again at 10 days. *HSFB2A* was annotated for "transcription cis-regulatory region binding", potentially suggesting that it has a role in either promoting or repressing genes in response to abiotic stress. Finally, two *NAC* genes were identified in the Arabidopsis ecotypes, *ANAC018/ATNAM* in Col-0 and *ANAC002/ATAF1* in Ost-0, and were differentially expressed in both Arabidopsis ecotypes, and the aspen and Norway spruce co-expressologs. Although no direct link to cold stress has been reported for *ANAC018/ATNAM*, *ANAC002/ATAF1* has an identified association with cold stress response in birch (Chen et al., 2019). In addition, *ANAC002/ATAF1* appears to function as a broader regulator of general stress responses (Wu et al., 2009), but particularly in drought stress responses where it promotes perception of ABA (Lu et al., 2007; Puranik et al., 2012).

Cluster 1 was most notably annotated for growth-related and developmental processes, with some TFs in this cluster having interesting roles in roots specifically. *LBD13* promotes lateral root development in Arabidopsis (Cho et al., 2019). Only its co-expressologs in aspen and Norway spruce were differentially expressed, and these were primarily down-regulated, suggesting that growth was mainly repressed in response to cold stress. *LBD11* is related to ROS detoxification (Dang et al., 2023), but its involvement in cold response has not yet been identified (Rong et al., 2024). *KANADI3* (*KAN3*), a GOLDEN2-like (GL2-like) TF, has no direct association with root-specific processes, however, as a member of the GARP TF family, it could be involved in cold stress regulation by altering gene expression under freezing conditions (Hernández-Verdeja & Lundgren, 2024). *KAN3* was up-regulated in the two Arabidopsis ecotypes and in Norway spruce, but down-regulated in the other tree species. This suggests a diverged role of this gene among these species.

While none of the TFs in the cliques were *CBFs*, we were still interested to see whether the Arabidopsis *CBFs* had any co-expressologs in root tissue of other species, as they are considered to be key regulators during cold stress in leaf tissue (Cook et al., 2004; Fowler & Thomashow, 2002; Medina et al., 1999). Interestingly, several co-expressologs were found in aspen and birch, while the two conifers only had a few co-expressologs with divergent expression patterns (**Figure 12**).

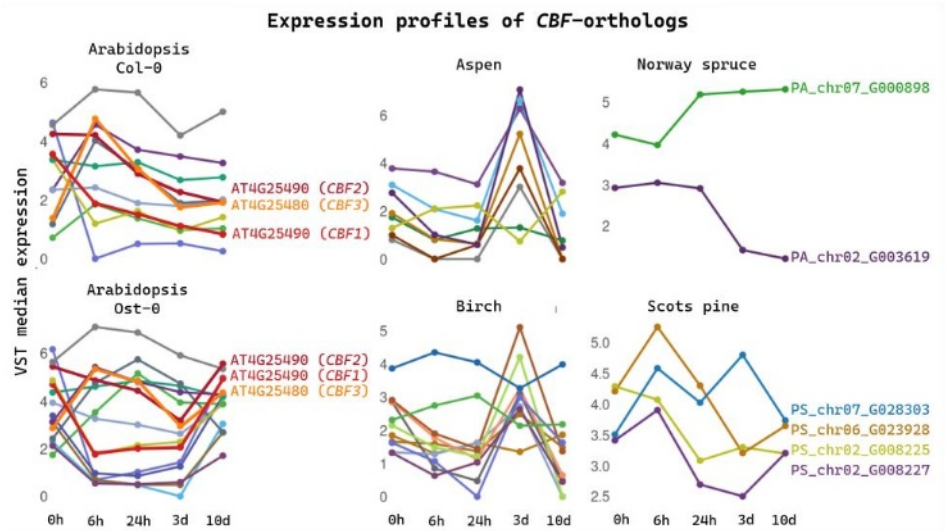


Figure 12 Expression profiles of co-expressologs (*P*-value < 0.05) linked to the Arabidopsis *C-repeat Binding Factors* (*CBFs*); *CBF1*: AT4G25490, *CBF2*: AT4G25470 and *CBF3*: AT4G25480.

In Arabidopsis Col-0, *CBF1* and *CBF2* were both down-regulated throughout the experiment, while *CBF3* was initially up-regulated at 6 hours and thereafter down-regulated. In the Ost-0 ecotype all three *CBFs* followed the same expression pattern as their Col-0 counterpart, but prolonged cold triggered a second up-regulation of *CBF*-related genes, suggesting that certain genotypes may retain or re-initiate CBF-mediated signalling under extended stress conditions. In aspen and birch the *CBF* orthologs were mostly up-regulated at 3 days after cold exposure, showing a delayed response compared to Arabidopsis. A limited number of *CBF* orthologs were identified in conifers, with two in Norway spruce and four in Scots pine. In Norway spruce, both genes were *CBF3* orthologs, with one (PA_chr07_G00898) up-regulated after 24 hours and the other (PA_chr02_G003619) was down-regulated after 3 days. In Scots pine, four potential *CBF* orthologs were identified, two of which (PS_chr07_G028303 and PS_chr06_G023928) were up-regulated after 6 hours, while the remaining two (PS_chr02_G008227 and PS_chr02_G008225) were down-regulated after 24 hours. Together, these results suggest that although the CBF-dependent cold response pathway actively operates in the root tissue of the angiosperms, it appears to play a comparatively limited role in gymnosperms and may be complemented by an alternative regulatory mechanism.

Overall, root cold responses were shown to involve a relatively small but deeply conserved set of regulatory modules with lineage-specific abiotic stress-response dynamics. This demonstrates substantial regulatory diversification among angiosperms and gymnosperms, confirming that abiotic stress responses in plants indeed differ extensively.

Paper IV – Comparative abiotic stress transcriptomics reveal common regulatory modules in conifers

Most research performed on abiotic stress response in plants has been focused on crop species, such as rice and wheat, and model herbaceous angiosperms, such as Arabidopsis, and not on unconventional model species, such as aspen or conifers. In **Paper IV**, we took one of the first looks into the conservation of stress response regulation in conifers enabled by availability of the newly annotated genomes of Norway spruce and Scots pine presented in **Paper II**.

Our experimental system included RNA-Seq profiling of the transcriptional response to cold and drought stress in needles and roots of Norway spruce and Scots pine. Analysis of gene expression from time series samples for both stresses revealed a noticeable difference between needle and root tissues of the abiotic stress exposed seedlings (**Figure 13A-D**). Approximately 30% of the variance between samples was explained by the difference in tissue type, whereas 10% was explained by the stress conditions. Drought stress caused an extensive transcriptional response in root tissue, where genes enriched for developmental processes were significantly down-regulated during the most severe conditions of the drought experiment (**Paper IV, Figure 1e**). Needles however, had far fewer differentially expressed genes across the entire drought experiment. The opposite was observed in the cold-stressed tissues, with needles having a higher increase in differential expression as compared to the root tissue. Analysis of GO enrichment indicated that down-regulated genes were largely linked to development and growth (**Paper IV, Supplementary Table 1**), whereas up-regulated genes were primarily associated with abiotic stress-responsive terms in both cold and drought conditions.

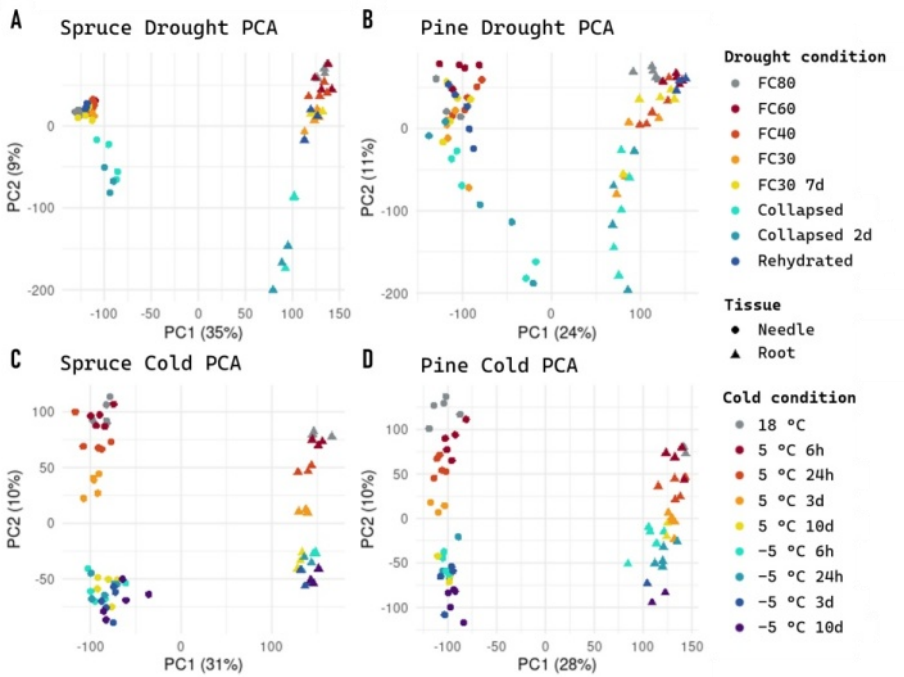


Figure 13 Principal Component Analyses (PCA) of abiotic stress datasets. Needle samples are shown as circles and root samples as triangles. Conditions range from grey as control to red at the start, to yellow, and to blue at the end of the experiment. A) PCA of Norway spruce drought experiment. B) PCA of Scots pine drought experiment. C) PCA of Norway spruce cold experiment. D) PCA of Scots pine cold experiment.

In addition to the similarities in tissue-specific responses between species revealed from the PCA plots, a large number of highly significant co-expressolog pairs were found directly related to abiotic stress response in Norway spruce and Scots pine. In drought stress data we identified 7313 orthogroups containing co-expressologs (*P*-value < 0.01), and in cold stress data we retain 7322 orthogroups of equal significance. At first glance, this large overlap might be expected given that these species are both gymnosperms and thus share a relatively recent common ancestry compared to more distantly related plant lineages. However, the evolutionary separation of Norway spruce and Scots pine is still quite distant, as they diverged ~130 million years ago and have since undergone independent evolutionary trajectories (Kumar et al., 2022). While they are sympatric across much of their boreal regions, they also occupy distinct ecological niches (Dering et al., 2021). These ecological and developmental differences could possibly have driven divergence in abiotic stress response pathways. Therefore, it was not guaranteed that co-expressed gene architecture would show the level of conservation we observed. The findings in **Paper IV** provide some of the first evidence that, despite their long independent evolutionary histories since their divergence, these two

conifer species retain an extensive degree of co-expression conservation in their response to abiotic stress. This overlap could also suggest that certain gene families are broadly important across species, possibly maintained by selective pressures.

Selecting for the co-expressologs where both orthologous genes were additionally differentially expressed, retained a highly conserved set of core stress responsive genes (**Paper IV, Figure 4**). Both stresses displayed distinct patterns of expression in accordance with the stress conditions. During drought stress, both conifers showed almost the exact same pattern of expression with either an up or down-regulation peak at the point of photosynthetic collapse. Testing for enrichment of GO terms within these sets of up and down-regulated genes, revealed enrichment of terms associated with a down-regulation of developmental processes and enrichment of terms associated with osmotic stress response within up-regulated genes (**Paper IV, Table 1**). Unfortunately, no significant GO enrichments were found for the cold-stressed set of co-expressologs.

From this highly conserved set of co-expressologs, 23 TFs in drought and 34 TFs in cold were extracted and examined. Across both stress conditions, TF expression profiles formed distinct clusters of up- and down-regulated genes (**Figure 14**). Enrichment analysis of the up- and down-regulated sets of TFs revealed significantly enriched terms associated with gene regulation in all TF sets, with the exception of down-regulated TFs under drought stress. These GO enrichment terms suggest that these TFs were up- or down-regulated specifically to activate appropriate stress responsive genes and repress the activation of processes unrelated to, or that could negatively impact, stress-response. There were seven TFs in common among the up-regulated TFs, comprising of two NACs, two WRKYs, one LBD, one ERF, and one MYB-related. Common down-regulated TFs included one LBD and one G2-like. These shared TFs represent particularly interesting targets for future functional validation as they potentially represent universal stress response activators.

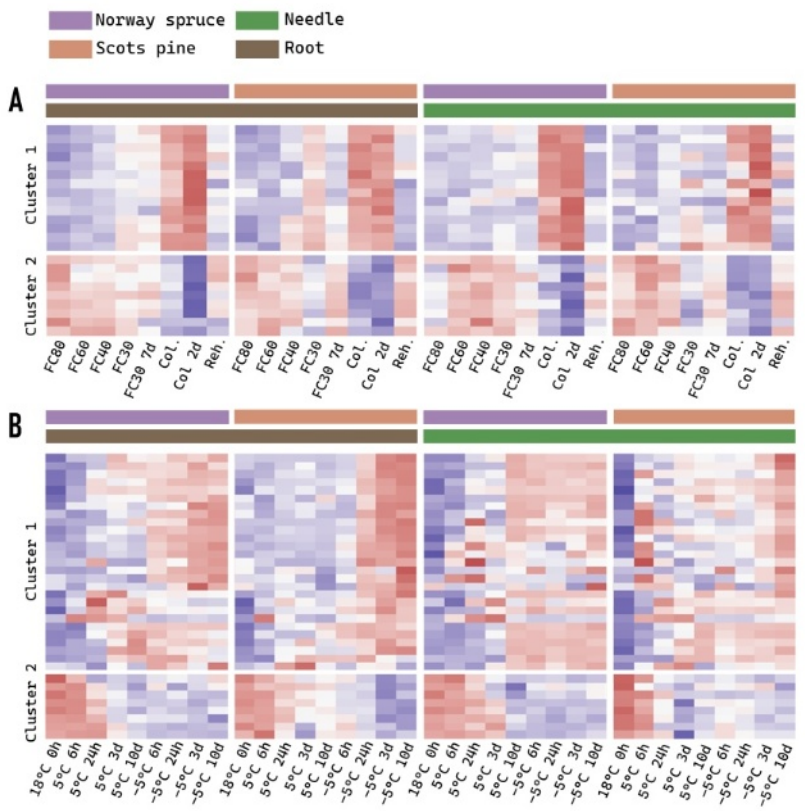


Figure 14 Heatmap of Transcription Factors (TFs) among highly conserved co-expressologs (*P*-value < 0.01). A) Heatmap of Variance Stabilising Transformation (VST) normalised median TF expression in conserved drought stress response co-expressologs. Col. = Collapsed, Col 2d = Collapsed for 2 days, Reh. = Rehydrated. b) Heatmap of VST median TF expression in conserved cold stress response co-expressologs. Ribbon colouring: needles in green, roots in brown, Norway spruce in purple, Scots pine in light orange.

Conclusions and future perspectives

This thesis set out to advance our understanding of abiotic stress response gene regulation in the closely related boreal conifers, Norway spruce and Scots pine. Through comparative gene expression analyses and network-based approaches, conserved regulatory components, and abiotic stress-activated candidate genes were identified. These findings contribute to a broader and more integrated view of how conifers respond to environmental challenges.

Consensus networks inferred using Seidr provided a more reliable and biologically informative view of gene regulation than any single network inference method alone (**Paper I**). Drought-response related genes of Norway spruce were identified by looking at high ranking genes within the network based on centrality measurements and comparison to an unstressed network. This highlighted the importance of an appropriate experimental setup for the generation of meaningful GCNs, as this response was not found in the unstressed network.

Paper II demonstrated that there is a clear relationship between three-dimensional genome structure and gene-co-expression in Norway spruce, as genes acting as loop anchors showed higher co-expression correlation. Cross-species analyses further revealed that a portion of TAD organisation was conserved across these conifer genomes. We found that when a co-expressolog in Norway spruce was located in a conserved TAD, its counterpart in Scots pine was highly likely to be found in a syntenically conserved equivalent TAD of Scots pine. Gene expression from these co-expressologs revealed that the genes located within each of these TADs are not co-expressed as the expression profiles of genes within the same TADs differed. However, co-expressologs within conserved TADs displayed a generally higher degree of network centrality, indicating that these genes could serve as highly conserved regulatory hubs.

Paper III revealed that while cold-responsive regulation in roots was temporally divergent across plant lineages there was a distinct subset of genes with functionally conserved co-expression, indicating conserved regulation. There was a relatively large set of orthogroups containing DEGs that were only differentially expressed in single species, indicating that distinct gene sets were transcriptionally regulated in response to cold stress in roots. Eight core TFs with conserved co-expression across all species were found, with several of these having documented roles in abiotic stress response. While their gene-neighbourhoods were highly similar, their expression still displayed temporal differences between species. Although they were not represented within the conserved core response, Arabidopsis *CBFs* had several significant co-expressologs in the other species, and predominantly in the other angiosperms, aspen and birch. This is in agreement with previous research

performed on CBFs in angiosperms, showing that CBFs also potentially have a cold regulation function in root tissue (Sang et al., 2025; Zhao et al., 2023). In the conifers however, only a few co-expressologs were found mainly for *CBF3*. The low copy number in conifers reflects the known expansion of CBFs in the angiosperm lineage with subsequent sub/neo-functionalisation related to stress response (Nie et al., 2022). All together the results from **Paper III** suggest that abiotic stress response in root tissue of the different species is rewired, which is in concurrence with the developmental system drift (True & Haag, 2001).

In **Paper IV** I show that Norway spruce and Scots pine possess a high degree of conservation for abiotic stress-specific co-expression. Nonetheless, the transcriptional response seems to remain strongly tissue- and stress-specific as shown in **Paper IV Figure 2**, where the largest sets of unique DEGs and orthogroups showed little overlap with other conditions. This suggests that conserved abiotic stress-response regulation is deployed in a context-dependent manner rather than as a common species-wide response. The conserved core responses shown in **Paper IV Figure 4** were potentially established prior to the divergence of Norway spruce and Scots pine in their common ancestor. As these conifers are sympatric but inhabit contrasting ecological niches, divergence of regulatory pathways was likely required.

All together, common conserved abiotic responses in boreal conifers and angiosperms (**Figure 15**) have been revealed and could potentially be used as targets for breeding programs in the future. However, even with everything that was accomplished in these works, relatively speaking, the surface was barely scraped.

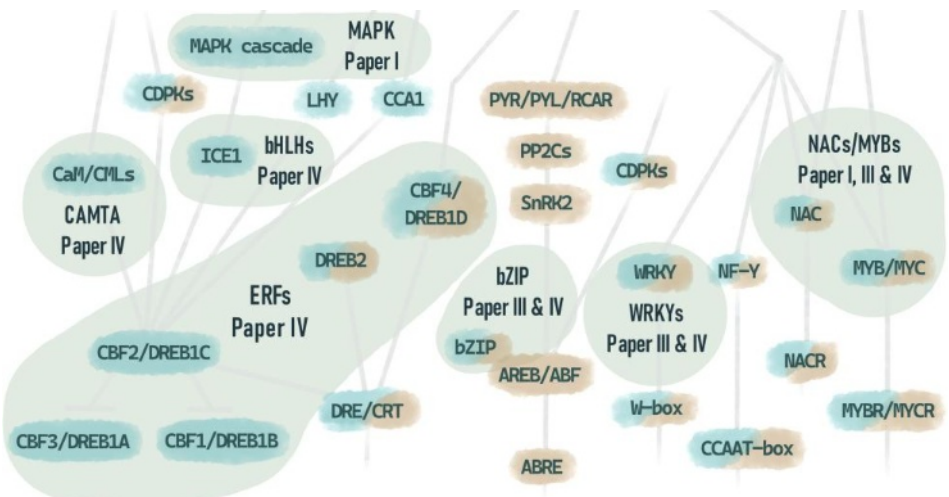


Figure 15 Summary picture of transcription factors found related to regulating abiotic stress in boreal species from Paper I, III and IV.

The increasing accuracy of network inference

As time goes on, high-throughput data will continue to accumulate and network inference approaches will become increasingly more sophisticated and accurate. The availability of larger and more diverse datasets will improve statistical power, better noise reduction, and the predictive performance of machine learning models. The integration of multiple data types in addition to transcriptomics, including epigenomics, proteomics, and chromatin accessibility data, will further strengthen network predictions by capturing complementary layers of information (Montesinos-López et al., 2025).

Other components of transcriptional regulation

While the main focus in this thesis was on protein-coding genes, there are many more directions that can be explored with the methods used in this thesis. Transcriptional components such as Transposable Elements (TEs), long non-coding RNAs (lncRNAs), and pseudogenes might all play a role in the regulation of abiotic stress responses.

Conifer genomes are replete with TEs, mobile genetic elements that can affect gene expression. Comparative TE-expression analysis across species, tissues, and stress conditions could reveal whether particular TEs become transcriptionally activated or repressed under abiotic stress. TEs may contribute to abiotic stress regulation, as they can act as alternative promoters, reshape chromatin accessibility and donate regulatory motifs (Hassan et al., 2024; Sigman & Slotkin, 2016). Integrating TE annotations into co-expression networks could identify TE-gene interactions, while long-read transcriptomics may help detect TE-derived transcripts and TE-associated protein isoforms (Zhang et al., 2025). A cross-species TE comparison would also shed light on how genome size expansion and TE retention influence transcriptional plasticity in long-lived trees.

LncRNAs remain one of the least investigated components of conifer transcriptomes, despite emerging evidence in angiosperms that they play essential roles in stress signalling, chromatin modulation, and fine-tuning of TF activity (Song et al., 2021). Steps have been made to identify and annotate lncRNAs in Norway spruce (Canovi et al., 2024). Future work could implement lncRNAs in co-expression network to investigate their expression across different abiotic stress conditions and assess their interaction with protein-coding genes.

Aside from functional gene duplicates, frequent duplications in conifers also led to the existence of pseudogenes. These are genes copies that show limited or no coding potential, yet might still play an active role in gene regulation. Rather than being functionless remnants of their source gene, pseudogenes can influence gene

regulation by acting as decoys for miRNAs, producing regulatory RNAs, or participating in epigenetic control (Sasidharan & Gerstein, 2008). Many pseudogenes were identified in Norway spruce and Scots pine in **Paper II** (131,696 and 155,941 pseudogenes, respectively). Incorporating these pseudogenes into GCNs would provide a way for uncovering potential pseudogene-gene regulatory interactions. Additionally, such analyses could reveal patterns of neo- or sub-functionalisation among pseudogenes. A preliminary assessment of the expression of TEs and pseudogenes in Norway spruce and Scots pine identified a subset that was differentially expressed, indicating that these components may indeed contribute to functional processes (data not shown).

Finally, transcriptional profiling provides only a snapshot of gene activity at the time of sampling, and does not necessarily reflect protein abundance or activity. Post-transcriptional regulation, such as messenger-RNA (mRNA) stability, alternative splicing, and other translational processes, influence whether a transcript is ultimately translated into a protein. Therefore, functional validation remains essential. Techniques such as DNA Affinity Purification-Sequencing (DAP-Seq) for TFs would allow direct identification of TF binding sites and provide confirmation of network-derived predictions (Bartlett et al., 2017). Complementary approaches include Chromatin Immunoprecipitation-Sequencing (ChIP-Seq) to TF binding in vivo (Song et al., 2016), Ribosome Profiling (Ribo-Seq) to assess which transcripts are actively being translated (Mahboubi et al., 2021), and targeted proteomics or mass spectrometry to measure protein abundance directly. While not currently achievable in Norway spruce or Scots pine, reporter assays can test regulatory activity of promoters or TFs in specific tissues, while gene knockout or knockdown strategies (e.g. Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins (CRISPR-Cas) or RNA-interference (RNAi)) would provide functional evidence for gene involvement in stress responses (Touzdjian Pinheiro Kohlrausch Távora et al., 2022). Together, these avenues highlight that while progress has been made, conifer abiotic stress biology still remains a vast unexplored landscape.

Acknowledgements

It has been a long and winding journey, marked by uncertainty about direction, yet ultimately an enriching one. I'm recalling the reminder that the value of the path lies not only in knowing where it leads, but in choosing to move forward and explore. Along the way, I was fortunate to be surrounded by people who offered guidance, encouragement, and good company. Their support and presence made the journey much easier to navigate and without them I would not have gotten where I am now.

I would like to sincerely thank my supervisor, **Nat**, for giving me the opportunity to pursue this PhD. I truly could not have wished for a better supervisor. Your enthusiasm and dedication to research, commitment to the group, and constant support made this journey both motivating and enjoyable, and I am deeply grateful for the guidance and trust you showed throughout these years.

Thank you, **Nico**, **Bastian**, **Alonso** and everyone else in the **Street-lab**, for giving me a solid foundation to build the rest of the PhD upon. Without your knowledge and many tricks and tips in coding, repositories and even game recommendations, this journey would have been a lot harder and less fun.

Thanks to **Vaughan**, **Tuuli**, and **Alex**, for the enjoyable collaboration on the root-project. I appreciate all of the good discussions we had and the fun anecdotes that were shared.

Thank you, **Torgeir**, for allowing me to have the opportunity to come visit NMBU in Ås, for teaching me about the complexities of ComPlEx, and always being up for a conversation and discussion of results. Thank you, **Ellen**, for always making the time to go through the various scripts with me and explain how you approached the vast conservation of wood development.

Thanks, **Sonja** and **Vikash** for all your help with the RNA-extractions, and to **David** for your support with the physiological measurements in the drought experiments.

Thank you, **Maria** and **Judith** for being part of my reference group and giving your expertise advice and encouragement over the years.

Thank you, to the people at **UPSC**, for the many lunch chats, company and support. To all the **PhD** students, I see you, and you got this!

Thanks to **Sara** and **Camy**, for going on this journey alongside me; shared load is half the load, they say. Thanks for the valuable discussions about this process, the advice and the many many laughs.

Thank you, **Teitur**, for checking up on me occasionally. Conversations with you were always very inspiring and gave me the motivation needed to go that extra mile.

A special thanks to you, **Amanda**, for taking me up on that fika and for being a listening ear. I hope I was able to return the favour. It's so rare to find someone with such similar interests. I enjoyed every conversation, Baldur's Gate evening, climbing session and moment shared. Thanks also for translating the abstract!

Thank you, **Edo**, for being a great friend who's always game for a game. Thank you for your insurmountable wit and endlessly hilarious puns. Working and hanging out with you has always kept my spirits a-pun-dantly high. **Giulia**, thank you for your warmth, your kindness, and the many rides home that were more than just a lift.

The Bakers Guild, dear friends who I met during the pandemic (online of course). Thank you for all the time spend together in a time where loneliness was the norm. Special thanks to **Cup**, for bringing everyone together.

Thank you to the current and past members of the **Bakers D&D group**. I have always treasured our adventures in the Warring Continents and Faerûn, and will continue doing so in the future with more adventures to come.

Lieve **mamma**, bedankt voor alles! Door jouw sterkte ben ik waar ik vandaag ben. Jij en **Saskia** hebben mij laten zien wat echte liefde is en wat het betekent om echt gelukkig te zijn. Dankjulliewel voor de warme open armen en gastvrijheid waardoor Zefrin en ik ons altijd zo thuis hebben gevoeld. Jullie hebben ontzettend veel voor ons betekent in de afgelopen jaren en zullen dat ook altijd blijven doen.

Dankjewel, **Jarl**, voor alle goeie gesprekken en onvoorwaardelijke liefde. Jouw en **Brittney's** steun kwam uit een plek van begrip en jullie advies hielp mij het nodige perspectief te krijgen. Bedankt dat jullie altijd voor me klaar stonden en weet dat ik ook altijd klaar sta voor jullie.

Ook wil ik graag **oma Adri** bedanken voor alle liefde, wijsheid, steun en betrokkenheid die je mij mijn hele leven gegeven hebt.

My cat **Grietje**, thank you for the cuddles and company, especially when it was just us in the apartment. Thank you also for making sure I go to bed on time and often doubling as an extra alarm-clock. Prrt!

And last, but definitely not least, **Zefrin**, my partner. Thank you for always being there when I needed someone to talk to, no matter what hour of the day or night it was. Thank you also, for bearing with me these last few stressful months. Het is tijd voor een nieuw hoofdstuk.

'All right,' said the Cat; and this time it vanished quite slowly,

beginning with the end of the tail, and ending with the grin,

which remained some time after the rest of it had gone.

Lewis Carroll, Alice's Adventures in Wonderland, Chapter VI: Pig and Pepper, 1865.

References

Agarwal, P. K., Agarwal, P., Reddy, M. K., & Sopory, S. K. (2006). Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Reports*, *25*(12), 1263-1274. https://doi.org/10.1007/s00299-006-0204-8

Almeida-Rodriguez, A. M., Cooke, J. E., Yeh, F., & Zwiazek, J. J. (2010). Functional characterization of drought-responsive aquaporins in Populus balsamifera and Populus simonii×balsamifera clones with different drought resistance strategies. *Physiol Plant*, *140*(4), 321-333. https://doi.org/10.1111/j.1399-3054.2010.01405.x

Anderegg, W. R. L., Trugman, A. T., Bowling, D. R., Salvucci, G., & Tuttle, S. E. (2019). Plant functional traits and climate influence drought intensification and land–atmosphere feedbacks. *Proceedings of the National Academy of Sciences*, *116*(28), 14071-14076. https://doi.org/doi:10.1073/pnas.1904747116

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data.

Aoki, K., Ogata, Y., & Shibata, D. (2007). Approaches for Extracting Practical Information from Gene Co-expression Networks in Plant Biology. *Plant and Cell Physiology*, *48*(3), 381-390. https://doi.org/10.1093/pcp/pcm013

Arend, M., Link, R. M., Patthey, R., Hoch, G., Schuldt, B., & Kahmen, A. (2021). Rapid hydraulic collapse as cause of drought-induced mortality in conifers. *Proceedings of the National Academy of Sciences*, *118*(16), e2025251118. https://doi.org/doi:10.1073/pnas.2025251118

Atif, R. M., Shahid, L., Waqas, M., Ali, B., Rashid, M. A. R., Azeem, F., Nawaz, M. A., Wani, S. H., & Chung, G. (2019). Insights on Calcium-Dependent Protein Kinases (CPKs) Signaling for Abiotic Stress Tolerance in Plants. *International Journal of Molecular Sciences*, *20*(21), 5298. https://www.mdpi.com/1422-0067/20/21/5298

Balfagón, D., Pascual, L. S., Sengupta, S., Halliday, K. J., Gómez-Cadenas, A., Peláez-Vico, M., Sinha, R., Mittler, R., & Zandalinas, S. I. (2024). WRKY48 negatively regulates plant acclimation to a combination of high light and heat stress. *Plant J*, *117*(6), 1642-1655. https://doi.org/10.1111/tpj.16658

Ballouz, S., Verleyen, W., & Gillis, J. (2015). Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics*, *31*(13), 2123-2130. https://doi.org/10.1093/bioinformatics/btv118

Bartlett, A., O'Malley, R. C., Huang, S.-s. C., Galli, M., Nery, J. R., Gallavotti, A., & Ecker, J. R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature Protocols*, *12*(8), 1659-1672. https://doi.org/10.1038/nprot.2017.055

Bavelas, A. (1948). A Mathematical Model for Group Structures. *Human Organization*, *7*(3), 16-30. https://doi.org/10.17730/humo.7.3.f4033344851gl053

Baxter, A., Mittler, R., & Suzuki, N. (2013). ROS as key players in plant stress signalling. *Journal of Experimental Botany*, *65*(5), 1229-1240. https://doi.org/10.1093/jxb/ert375

Bieniawska, Z., Espinoza, C., Schlereth, A., Sulpice, R., Hincha, D. K., & Hannah, M. A. (2008). Disruption of the Arabidopsis Circadian Clock Is Responsible for Extensive Variation in the Cold-Responsive Transcriptome *Plant Physiology*, *147*(1), 263-279. https://doi.org/10.1104/pp.108.118059

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

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

Bonev, B., & Cavalli, G. (2016). Organization and function of the 3D genome. *Nature Reviews Genetics*, *17*(11), 661-678. https://doi.org/10.1038/nrg.2016.112

Bouché, F., Woods, D. P., & Amasino, R. M. (2017). Winter Memory throughout the Plant Kingdom: Different Paths to Flowering. *Plant Physiol*, *173*(1), 27-35. https://doi.org/10.1104/pp.16.01322

Bradshaw, C. J. A., & Warkentin, I. G. (2015). Global estimates of boreal forest carbon stocks and flux. *Global and Planetary Change*, *128*, 24-30. https://doi.org/https://doi.org/10.1016/j.gloplacha.2015.02.004

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, *30*(1-7), 107-117.

Brodribb, T. J., & Cochard, H. (2009). Hydraulic failure defines the recovery and point of death in water-stressed conifers. *Plant Physiol*, *149*(1), 575-584. https://doi.org/10.1104/pp.108.129783

Brodribb, T. J., & McAdam, S. A. M. (2013). Abscisic Acid Mediates a Divergence in the Drought Response of Two Conifers *Plant Physiology*, *162*(3), 1370-1377. https://doi.org/10.1104/pp.113.217877

Brodribb, T. J., McAdam, S. A. M., Jordan, G. J., & Martins, S. C. V. (2014). Conifer species adapt to low-rainfall climates by following one of two divergent pathways. *Proceedings of the National Academy of Sciences*, *111*(40), 14489-14493. https://doi.org/doi:10.1073/pnas.1407930111

Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, *10*(1), 1017. https://doi.org/10.1038/s41467-019-08746-5

Bumgarner, R. (2013). Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol*, *Chapter 22*, Unit 22.21. https://doi.org/10.1002/0471142727.mb2201s101

Butte AJ, K. I. (2000). Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. In *Biocomputing* (pp. 418-429). https://doi.org/10.1142/9789814447331_0040

Butte, A. J., & Kohane, I. S. (1999). Unsupervised knowledge discovery in medical databases using relevance networks. *Proc AMIA Symp*, 711-715.

Cai, M., Sun, C., Ou, J., & Zeng, T. (2025). Genome identification of NAC gene family and its gene expression patterns in responding to salt and drought stresses in Rhododendron delavayi. *BMC Plant Biology*, *25*(1), 924. https://doi.org/10.1186/s12870-025-06965-1

Candeloro, L., Savini, L., & Conte, A. (2016). A New Weighted Degree Centrality Measure: The Application in an Animal Disease Epidemic. *PLOS ONE*, *11*(11), e0165781. https://doi.org/10.1371/journal.pone.0165781

Canovi, C., Stojkovič, K., Benítez, A. A., Delhomme, N., Egertsdotter, U., & Street, N. R. (2024). A resource of identified and annotated lincRNAs expressed during somatic embryogenesis development in Norway spruce. *Physiologia Plantarum*, *176*(5), e14537. https://doi.org/https://doi.org/10.1111/ppl.14537

Caudullo, G., Tinner, W., & de Rigo, D. (2016). Picea abies in Europe: distribution, habitat, usage and threats. In.

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. https://doi.org/10.1007/BF02670468

Chen, S., Lin, X., Zhang, D., Li, Q., Zhao, X., & Chen, S. (2019). Genome-Wide Analysis of NAC Gene Family in Betula pendula. *Forests*, *10*(9), 741. https://www.mdpi.com/1999-4907/10/9/741

Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B. H., Hong, X., Agarwal, M., & Zhu, J. K. (2003). ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes Dev*, *17*(8), 1043-1054. https://doi.org/10.1101/gad.1077503

Chinnusamy, V., Zhu, J., & Zhu, J.-K. (2007). Cold stress regulation of gene expression in plants. *Trends in Plant Science*, *12*(10), 444-451. https://doi.org/https://doi.org/10.1016/j.tplants.2007.07.002

Cho, C., Jeon, E., Pandey, S. K., Ha, S. H., & Kim, J. (2019). LBD13 positively regulates lateral root formation in Arabidopsis. *Planta*, *249*(4), 1251-1258. https://doi.org/10.1007/s00425-018-03087-x

Choat, B., Brodribb, T. J., Brodersen, C. R., Duursma, R. A., López, R., & Medlyn, B. E. (2018). Triggers of tree mortality under drought. *Nature*, *558*(7711), 531-539. https://doi.org/10.1038/s41586-018-0240-x

Chouard, P. (1960). Vernalization and its relations to dormancy. *Annual review of plant physiology*, *11*(1), 191-238.

Cook, D., Fowler, S., Fiehn, O., & Thomashow, M. F. (2004). A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. *Proceedings of the National Academy of Sciences*, *101*(42), 15243-15248. https://doi.org/doi:10.1073/pnas.0406069101

Cortijo, S., Bhattarai, M., Locke, J. C. W., & Ahnert, S. E. (2020). Co-expression Networks From Gene Expression Variability Between Genetically Identical Seedlings Can Reveal Novel Regulatory Relationships. *Front Plant Sci*, *11*, 599464. https://doi.org/10.3389/fpls.2020.599464

Coscia, M., & Neffke, F. M. H. (2017, 19-22 April 2017). Network Backboning with Noisy Data. 2017 IEEE 33rd International Conference on Data Engineering (ICDE),

Cowan, I. R., & Farquhar, G. D. (1977). Stomatal function in relation to leaf metabolism and environment. *Symp Soc Exp Biol*, *31*, 471-505.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561-563. https://doi.org/10.1038/227561a0

D'Orangeville, L., Houle, D., Duchesne, L., Phillips, R. P., Bergeron, Y., & Kneeshaw, D. (2018). Beneficial effects of climate warming on boreal tree growth may be transitory. *Nature Communications*, *9*(1), 3213. https://doi.org/10.1038/s41467-018-05705-4

Dai, H., Zhou, J., & Zhu, B. (2018). Gene co-expression network analysis identifies the hub genes associated with immune functions for nocturnal hemodialysis in patients with end-stage renal disease. *Medicine (Baltimore)*, *97*(37), e12018. https://doi.org/10.1097/md.0000000000012018

Dang, T. V. T., Lee, S., Cho, H., Choi, K., & Hwang, I. (2023). The LBD11-ROS feedback regulatory loop modulates vascular cambium proliferation and secondary growth in Arabidopsis. *Mol Plant*, *16*(7), 1131-1145. https://doi.org/10.1016/j.molp.2023.05.010

Daub, C. O., Steuer, R., Selbig, J., & Kloska, S. (2004). Estimating mutual information using B-spline functions--an improved similarity measure for analysing gene expression data. *BMC Bioinformatics*, *5*, 118. https://doi.org/10.1186/1471-2105-5-118

de Siqueira Santos, S., Takahashi, D. Y., Nakata, A., & Fujita, A. (2013). A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in Bioinformatics*, *15*(6), 906-918. https://doi.org/10.1093/bib/bbt051

Dering, M., Baranowska, M., Beridze, B., Chybicki, I. J., Danelia, I., Iszkuło, G., Kvartskhava, G., Kosiński, P., Rączka, G., Thomas, P. A., Tomaszewski, D., Walas, Ł., & Sękiewicz, K. (2021). The evolutionary heritage and ecological uniqueness of Scots pine in the Caucasus ecoregion is at risk of climate changes. *Scientific Reports*, *11*(1), 22845. https://doi.org/10.1038/s41598-021-02098-1

Dixon, H. H. (1914). *Transpiration and the Ascent of Sap in Plants*. Macmillan and co.

Durrant, T., de Rigo, D., & Caudullo, G. (2016). Pinus sylvestris in Europe: distribution, habitat, usage and threats. In.

Duttke, S. H., Guzman, C., Chang, M., Delos Santos, N. P., McDonald, B. R., Xie, J., Carlin, A. F., Heinz, S., & Benner, C. (2024). Position-dependent function of human sequence-specific transcription factors. *Nature*, *631*(8022), 891-898. https://doi.org/10.1038/s41586-024-07662-z

Eidsaa, M., Stubbs, L., & Almaas, E. (2017). Comparative analysis of weighted gene co-expression networks in human and mouse. *PLOS ONE*, *12*(11), e0187611. https://doi.org/10.1371/journal.pone.0187611

Elo, L. L., Järvenpää, H., Orešič, M., Lahesmaa, R., & Aittokallio, T. (2007). Systematic construction of gene coexpression networks with applications to human T

helper cell differentiation process. *Bioinformatics*, *23*(16), 2096-2103. https://doi.org/10.1093/bioinformatics/btm309
Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y
Eriksson, M. E., & Webb, A. A. R. (2011). Plant cell responses to cold are all about timing. *Current opinion in plant biology*, *14*(6), 731-737. https://doi.org/https://doi.org/10.1016/j.pbi.2011.08.005
Erola, P., Bonnet, E., & Michoel, T. (2019). Learning Differential Module Networks Across Multiple Experimental Conditions. *Methods Mol Biol*, *1883*, 303-321. https://doi.org/10.1007/978-1-4939-8882-2_13
Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J. J., & Gardner, T. S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol*, *5*(1), e8. https://doi.org/10.1371/journal.pbio.0050008
Fang, Y., You, J., Xie, K., Xie, W., & Xiong, L. (2008). Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Mol Genet Genomics*, *280*(6), 547-563. https://doi.org/10.1007/s00438-008-0386-6
Fowler, S., & Thomashow, M. F. (2002). Arabidopsis transcriptome profiling indicates that multiple regulatory pathways are activated during cold acclimation in addition to the CBF cold response pathway. *Plant Cell*, *14*(8), 1675-1690. https://doi.org/10.1105/tpc.003483
Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1 - 22. https://doi.org/10.18637/jss.v033.i01
Fujiwara, S., Mitsuda, N., Nakai, Y., Kigoshi, K., Suzuki, K., & Ohme-Takagi, M. (2014). Chimeric repressor analysis identifies MYB87 as a possible regulator of morphogenesis via cell wall organization and remodeling in Arabidopsis. *Biotechnology Letters*, *36*(5), 1049-1057. https://doi.org/10.1007/s10529-013-1451-8
Garnett, S. T., Burgess, N. D., Fa, J. E., Fernández-Llamazares, A., Molnár, Z., Robinson, C. J., Watson, J. E. M., Zander, K. K., Austin, B., Brondizio, E. S., Collier, N. F., Duncan, T., Ellis, E., Geyle, H., Jackson, M. V., Jonas, H., Malmer, P., McGowan, B., Sivongxay, A., & Leiper, I. (2018). A spatial overview of the global importance of Indigenous lands for conservation. *Nature Sustainability*, *1*(7), 369-374. https://doi.org/10.1038/s41893-018-0100-6
Giorgi, F. M., Del Fabbro, C., & Licausi, F. (2013). Comparative study of RNA-seq- and Microarray-derived coexpression networks in Arabidopsis thaliana. *Bioinformatics*, *29*(6), 717-724. https://doi.org/10.1093/bioinformatics/btt053
Guérin, C., Roche, J., Allard, V., Ravel, C., Mouzeyar, S., & Bouzidi, M. F. (2019). Genome-wide analysis, expansion and expression of the NAC family under

drought and heat stresses in bread wheat (T. aestivum L.). *PLOS ONE*, *14*(3), e0213390. https://doi.org/10.1371/journal.pone.0213390

Guo, M., Liu, J.-H., Ma, X., Luo, D.-X., Gong, Z.-H., & Lu, M.-H. (2016). The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses [Review]. *Frontiers in Plant Science*, *Volume 7 - 2016*. https://doi.org/10.3389/fpls.2016.00114

Guo, S., Jiang, Q., Chen, L., & Guo, D. (2016). Gene regulatory network inference using PLS-based methods. *BMC Bioinformatics*, *17*(1), 545. https://doi.org/10.1186/s12859-016-1398-6

Hassan, A. H., Mokhtar, M. M., & El Allali, A. (2024). Transposable elements: multifunctional players in the plant genome [Review]. *Frontiers in Plant Science*, *Volume 14 - 2023*. https://doi.org/10.3389/fpls.2023.1330127

Haury, A.-C., Mordelet, F., Vera-Licona, P., & Vert, J.-P. (2012). TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. *BMC Systems Biology*, *6*(1), 145. https://doi.org/10.1186/1752-0509-6-145

Hemming, M. N., & Trevaskis, B. (2011). Make hay when the sun shines: the role of MADS-box genes in temperature-dependant seasonal flowering responses. *Plant Sci*, *180*(3), 447-453. https://doi.org/10.1016/j.plantsci.2010.12.001

Hernández-Verdeja, T., & Lundgren, M. R. (2024). GOLDEN2-LIKE transcription factors: A golden ticket to improve crops? *PLANTS, PEOPLE, PLANET*, *6*(1), 79-93. https://doi.org/https://doi.org/10.1002/ppp3.10412

Hirayama, T., & Shinozaki, K. (2010). Research on plant abiotic stress responses in the post-genome era: Past, present and future. *The Plant journal : for cell and molecular biology*, *61*, 1041-1052. https://doi.org/10.1111/j.1365-313X.2010.04124.x

Huang, H., Ullah, F., Zhou, D.-X., Yi, M., & Zhao, Y. (2019). Mechanisms of ROS Regulation of Plant Development and Stress Responses [Review]. *Frontiers in Plant Science*, *Volume 10 - 2019*. https://doi.org/10.3389/fpls.2019.00800

Hubbard, R. M., Ryan, M. G., Stiller, V., & Sperry, J. S. (2001). Stomatal conductance and photosynthesis vary linearly with plant hydraulic conductance in ponderosa pine. *Plant, Cell & Environment*, *24*(1), 113-121. https://doi.org/https://doi.org/10.1046/j.1365-3040.2001.00660.x

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLOS ONE*, *5*(9), e12776. https://doi.org/10.1371/journal.pone.0012776

Iacobucci, D., McBride, R., & Popovich, D. (2017). Eigenvector Centrality: Illustrations Supporting the Utility of Extracting More Than One Eigenvector to Obtain Additional Insights into Networks and Interdependent Structures. *Journal of Social Structure*, *18*, 1-21. https://doi.org/10.21307/joss-2018-003

Ibáñez, C., Kozarewa, I., Johansson, M., Ögren, E., Rohde, A., & Eriksson, M. E. (2010). Circadian Clock Components Regulate Entry and Affect Exit of Seasonal Dormancy as Well as Winter Hardiness in Populus Trees *Plant Physiology*, *153*(4), 1823-1833. https://doi.org/10.1104/pp.110.158220

Irvine, J., Perks, M. P., Magnani, F., & Grace, J. (1998). The response of Pinus sylvestris to drought: stomatal control of transpiration and hydraulic conductance. *Tree Physiology*, *18*(6), 393-402. https://doi.org/10.1093/treephys/18.6.393

Jalmi, S. K., & Sinha, A. K. (2015). ROS mediated MAPK signaling in abiotic and biotic stress- striking similarities and differences [Review]. *Frontiers in Plant Science*, *Volume 6 - 2015*. https://doi.org/10.3389/fpls.2015.00769

Jiang, Z., Dong, X., Li, Z.-G., He, F., & Zhang, Z. (2016). Differential Coexpression Analysis Reveals Extensive Rewiring of Arabidopsis Gene Coexpression in Response to Pseudomonas syringae Infection. *Scientific Reports*, *6*(1), 35064. https://doi.org/10.1038/srep35064

Jiménez-Gómez, J. M. (2014). Network types and their application in natural variation studies in plants. *Curr Opin Plant Biol*, *18*, 80-86. https://doi.org/10.1016/j.pbi.2014.02.010

Johnson, E. A., & Miyanishi, K. (2012). The boreal forest as a cultural landscape. *Annals of the New York Academy of Sciences*, *1249*(1), 151-165.

Johnson, K. A., & Krishnan, A. (2022). Robust normalization and transformation techniques for constructing gene coexpression networks from RNA-seq data. *Genome Biology*, *23*(1), 1. https://doi.org/10.1186/s13059-021-02568-9

Johnson, Z., Anderson, D., Cheung, M. S., & Bohutskyi, P. (2025). Gene network centrality analysis identifies key regulators coordinating day-night metabolic transitions in Synechococcus elongatus PCC 7942 despite limited accuracy in predicting direct regulator-gene interactions [Original Research]. *Frontiers in Microbiology*, *Volume 16 - 2025*. https://doi.org/10.3389/fmicb.2025.1569559

Joshi, R., Wani, S. H., Singh, B., Bohra, A., Dar, Z. A., Lone, A. A., Pareek, A., & Singla-Pareek, S. L. (2016). Transcription Factors and Plants Response to Drought Stress: Current Understanding and Future Directions [Review]. *Frontiers in Plant Science*, *Volume 7 - 2016*. https://doi.org/10.3389/fpls.2016.01029

Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K., & Shinozaki, K. (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nature Biotechnology*, *17*(3), 287-291. https://doi.org/10.1038/7036

Katz, L. (1953). A New Status Index Derived from Sociometric Analysis. *Psychometrika*, *18*(1), 39-43. https://doi.org/10.1007/BF02289026

Kim, J.-S., Kidokoro, S., Yamaguchi-Shinozaki, K., & Shinozaki, K. (2024). Regulatory networks in plant responses to drought and cold stress. *Plant Physiology*, *195*(1), 170-189. https://doi.org/10.1093/plphys/kiae105

King, R. W., & Heide, O. M. (2009). Seasonal flowering and evolution: the heritage from Charles Darwin. *Functional Plant Biology*, *36*(12), 1027-1036. https://doi.org/10.1071/fp09170

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

Krieger, G., Lupo, O., Levy, A. A., & Barkai, N. (2020). Independent evolution of transcript abundance and gene regulatory dynamics. *Genome Res*, *30*(7), 1000-1011. https://doi.org/10.1101/gr.261537.120

Kumar, A., Adamopoulos, S., Jones, D., & Amiandamhen, S. O. (2021). Forest Biomass Availability and Utilization Potential in Sweden: A Review. *Waste and Biomass Valorization*, *12*(1), 65-80. https://doi.org/10.1007/s12649-020-00947-0

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution*, *39*(8). https://doi.org/10.1093/molbev/msac174

Kurjak, D., Střelcová, K., Ditmarová, Ľ., Priwitzer, T., Kmeť, J., Homolák, M., & Pichler, V. (2012). Physiological response of irrigated and non-irrigated Norway spruce trees as a consequence of drought in field conditions. *European Journal of Forest Research*, *131*(6), 1737-1746. https://doi.org/10.1007/s10342-012-0611-z

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559. https://doi.org/10.1186/1471-2105-9-559

Lauritzen, S. L. (1996). *Graphical models* (Vol. 17). Clarendon Press.

Le, D. T., Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K., & Tran, L. S. (2011). Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. *DNA Res*, *18*(4), 263-276. https://doi.org/10.1093/dnares/dsr015

Lee, K. H., Piao, H. L., Kim, H. Y., Choi, S. M., Jiang, F., Hartung, W., Hwang, I., Kwak, J. M., Lee, I. J., & Hwang, I. (2006). Activation of glucosidase via stress-induced polymerization rapidly increases active pools of abscisic acid. *Cell*, *126*(6), 1109-1120. https://doi.org/10.1016/j.cell.2006.07.034

Leonelli, S., & Ankeny, R. A. (2013). What makes a model organism? *Endeavour*, *37*(4), 209-212. https://doi.org/10.1016/j.endeavour.2013.06.001

Lévesque, M., Saurer, M., Siegwolf, R., Eilmann, B., Brang, P., Bugmann, H., & Rigling, A. (2013). Drought response of five conifer species under contrasting water availability suggests high vulnerability of Norway spruce and European larch. *Global Change Biology*, *19*(10), 3184-3199. https://doi.org/https://doi.org/10.1111/gcb.12268

Li, D., Brown, J., Orsini, L., Pan, Z., Hu, G., & He, S. (2016). MODA: MOdule Differential Analysis for weighted gene co-expression network. *bioRxiv*. https://doi.org/10.1101/053496

Lian, Q., Huettel, B., Walkemeier, B., Mayjonade, B., Lopez-Roques, C., Gil, L., Roux, F., Schneeberger, K., & Mercier, R. (2024). A pan-genome of 69 Arabidopsis thaliana accessions reveals a conserved genome structure throughout the global species range. *Nature Genetics*, *56*(5), 982-991. https://doi.org/10.1038/s41588-024-01715-9

Liao, H., Mariani, M. S., Medo, M., Zhang, Y.-C., & Zhou, M.-Y. (2017). Ranking in evolving complex networks. *Physics Reports*, *689*, 1-54. https://doi.org/https://doi.org/10.1016/j.physrep.2017.05.001

Lin, Z., Pan, I., & Pan, W. (2024). On network deconvolution for undirected graphs. *Biometrics*, *80*(4). https://doi.org/10.1093/biomtc/ujae112

Liu, L., Heidecker, M., Depuydt, T., Manosalva Perez, N., Crespi, M., Blein, T., & Vandepoele, K. (2023). Transcription factors KANADI 1, MYB DOMAIN PROTEIN 44, and PHYTOCHROME INTERACTING FACTOR 4 regulate long intergenic noncoding RNAs expressed in Arabidopsis roots. *Plant Physiology*, *193*(3), 1933-1953. https://doi.org/10.1093/plphys/kiad360

Liu, W., Lin, L., Zhang, Z., Liu, S., Gao, K., Lv, Y., Tao, H., & He, H. (2019). Gene co-expression network analysis identifies trait-related modules in Arabidopsis thaliana. *Planta*, *249*(5), 1487-1501. https://doi.org/10.1007/s00425-019-03102-9

Liu, Z. P. (2015). Reverse Engineering of Genome-wide Gene Regulatory Networks from Gene Expression Data. *Curr Genomics*, *16*(1), 3-22. https://doi.org/10.2174/1389202915666141110210634

Lloret, A., Badenes, M. L., & Ríos, G. (2018). Modulation of Dormancy and Growth Responses in Reproductive Buds of Temperate Trees. *Front Plant Sci*, *9*, 1368. https://doi.org/10.3389/fpls.2018.01368

Long, H., Zheng, Z., Zhang, Y., Xing, P., Wan, X., Zheng, Y., & Li, L. (2019). An abscisic acid (ABA) homeostasis regulated by its production, catabolism and transport in peanut leaves in response to drought stress. *PLOS ONE*, *14*(6), e0213963. https://doi.org/10.1371/journal.pone.0213963

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

Lu, P.-L., Chen, N.-Z., An, R., Su, Z., Qi, B.-S., Ren, F., Chen, J., & Wang, X.-C. (2007). A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. *Plant Molecular Biology*, *63*(2), 289-305. https://doi.org/10.1007/s11103-006-9089-8

Magwene, P. M., & Kim, J. (2004). Estimating genomic coexpression networks using first-order conditional independence. *Genome Biology*, *5*(12), R100. https://doi.org/10.1186/gb-2004-5-12-r100

Mahboubi, A., Delhomme, N., Häggström, S., & Hanson, J. (2021). Small-scale sequencing enables quality assessment of Ribo-Seq data: an example from Arabidopsis cell culture. *Plant Methods*, *17*(1), 92. https://doi.org/10.1186/s13007-021-00791-w

Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Kellis, M., Collins, J. J., & Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. *Nat Methods*, *9*(8), 796-804. https://doi.org/10.1038/nmeth.2016

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., & Califano, A. (2006). ARACNE: An Algorithm for the Reconstruction of Gene

Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics*, *7*(1), S7. https://doi.org/10.1186/1471-2105-7-S1-S7
Margolin, A. A., Wang, K., Califano, A., & Nemenman, I. (2010). Multivariate dependence and genetic networks inference. *IET Syst Biol*, *4*(6), 428-440. https://doi.org/10.1049/iet-syb.2010.0009
Maurya, J. P., & Bhalerao, R. P. (2017). Photoperiod- and temperature-mediated control of growth cessation and dormancy in trees: a molecular perspective. *Ann Bot*, *120*(3), 351-360. https://doi.org/10.1093/aob/mcx061
McDowell, N., Pockman, W. T., Allen, C. D., Breshears, D. D., Cobb, N., Kolb, T., Plaut, J., Sperry, J., West, A., Williams, D. G., & Yepez, E. A. (2008). Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? *New Phytologist*, *178*(4), 719-739. https://doi.org/https://doi.org/10.1111/j.1469-8137.2008.02436.x
McKenzie, A. T., Katsyv, I., Song, W. M., Wang, M., & Zhang, B. (2016). DGCA: A comprehensive R package for Differential Gene Correlation Analysis. *BMC Syst Biol*, *10*(1), 106. https://doi.org/10.1186/s12918-016-0349-1
McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res*, *20*(6), 816-825. https://doi.org/10.1101/gr.102491.109
Medina, J., Bargues, M., Terol, J., Pérez-Alonso, M., & Salinas, J. (1999). The Arabidopsis CBF gene family is composed of three genes encoding AP2 domain-containing proteins whose expression Is regulated by low temperature but not by abscisic acid or dehydration. *Plant Physiol*, *119*(2), 463-470. https://doi.org/10.1104/pp.119.2.463
Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., & Koornneef, M. (1998). Arabidopsis thaliana: a model plant for genome analysis. *Science*, *282*(5389), 662, 679-682. https://doi.org/10.1126/science.282.5389.662
Mohammadi, A., & Wit, E. C. (2015). Bayesian Structure Learning in Sparse Gaussian Graphical Models. *Bayesian Analysis*, *10*(1), 109-138, 130. https://doi.org/10.1214/14-BA889
Montesinos-López, O. A., Montesinos-López, A., Mosqueda-González, B. A., Delgado-Enciso, I., Chavira-Flores, M., Crossa, J., Dreisigacker, S., Sun, J., & Ortiz, R. (2025). Genomic prediction powered by multi-omics data. *Front Genet*, *16*, 1636438. https://doi.org/10.3389/fgene.2025.1636438
Müller, B., & Grossniklaus, U. (2010). Model organisms--A historical perspective. *J Proteomics*, *73*(11), 2054-2063. https://doi.org/10.1016/j.jprot.2010.08.002
Murata, N., & Los, D. A. (1997). Membrane Fluidity and Temperature Perception. *Plant Physiol*, *115*(3), 875-879. https://doi.org/10.1104/pp.115.3.875
Murata, Y., Pei, Z. M., Mori, I. C., & Schroeder, J. (2001). Abscisic acid activation of plasma membrane Ca(2+) channels in guard cells requires cytosolic NAD(P)H and is differentially disrupted upstream and downstream of reactive oxygen species production in abi1-1 and abi2-1 protein phosphatase 2C mutants. *Plant Cell*, *13*(11), 2513-2523. https://doi.org/10.1105/tpc.010210

Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhöh, O., & Persson, S. (2010). Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiol*, *152*(1), 29-43. https://doi.org/10.1104/pp.109.145318

Nakashima, K., Shinwari, Z. K., Sakuma, Y., Seki, M., Miura, S., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2000). Organization and expression of two Arabidopsis DREB2 genes encoding DRE-binding proteins involved in dehydration- and high-salinity-responsive gene expression. *Plant Molecular Biology*, *42*(4), 657-665. https://doi.org/10.1023/A:1006321900483

Nakashima, K., Yamaguchi-Shinozaki, K., & Shinozaki, K. (2014). The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat [Review]. *Frontiers in Plant Science*, *Volume 5 - 2014*. https://doi.org/10.3389/fpls.2014.00170

Netotea, S., Sundell, D., Street, N. R., & Hvidsten, T. R. (2014). ComPlEx: conservation and divergence of co-expression networks in A. thaliana, Populus and O. sativa. *BMC Genomics*, *15*(1), 106. https://doi.org/10.1186/1471-2164-15-106

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

Ni, Y., Baladandayuthapani, V., Vannucci, M., & Stingo, F. C. (2022). Bayesian graphical models for modern biological applications. *Stat Methods Appt*, *31*(2), 197-225. https://doi.org/10.1007/s10260-021-00572-8

Nie, Y., Guo, L., Cui, F., Shen, Y., Ye, X., Deng, D., Wang, S., Zhu, J., & Wu, W. (2022). Innovations and stepwise evolution of CBFs/DREB1s and their regulatory networks in angiosperms. *J Integr Plant Biol*, *64*(11), 2111-2125. https://doi.org/10.1111/jipb.13357

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L.,…Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, *497*(7451), 579-584. https://doi.org/10.1038/nature12211

Obayashi, T., Hayashi, S., Saeki, M., Ohta, H., & Kinoshita, K. (2009). ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Res*, *37*(Database issue), D987-991. https://doi.org/10.1093/nar/gkn807

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., Matsubara, K., Osato, N., Kawai, J., Carninci, P., Hayashizaki, Y., Suzuki, K., Kojima, K., Takahara, Y., Yamamoto, K., & Kikuchi, S. (2003). Comprehensive analysis of NAC family genes in Oryza sativa and Arabidopsis thaliana. *DNA Res*, *10*(6), 239-247. https://doi.org/10.1093/dnares/10.6.239

Örvar, B. L., Sangwan, V., Omann, F., & Dhindsa, R. S. (2000). Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity. *The Plant Journal*, *23*(6), 785-794. https://doi.org/https://doi.org/10.1046/j.1365-313x.2000.00845.x

Ovens, K., Eames, B. F., & McQuillan, I. (2021). Comparative Analyses of Gene Co-expression Networks: Implementations and Applications in the Study of Evolution [Review]. *Frontiers in Genetics*, *Volume 12 - 2021*. https://doi.org/10.3389/fgene.2021.695399

Ovens, K., Maleki, F., Eames, B. F., & McQuillan, I. (2021). Juxtapose: a gene-embedding approach for comparing co-expression networks. *BMC Bioinformatics*, *22*(1), 125. https://doi.org/10.1186/s12859-021-04055-1

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, *12*(2), 87-98. https://doi.org/10.1038/nrg2934

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

Pei, Z.-M., Murata, Y., Benning, G., Thomine, S., Klüsener, B., Allen, G. J., Grill, E., & Schroeder, J. I. (2000). Calcium channels activated by hydrogen peroxide mediate abscisic acid signalling in guard cells. *Nature*, *406*(6797), 731-734. https://doi.org/10.1038/35021067

Powers, J. S., Vargas, G. G., Brodribb, T. J., Schwartz, N. B., Pérez-Aviles, D., Smith-Martin, C. M., Becknell, J. M., Aureli, F., Blanco, R., Calderón-Morales, E., Calvo-Alvarado, J. C., Calvo-Obando, A. J., Chavarría, M. M., Carvajal-Vanegas, D., Jiménez-Rodríguez, C. D., Murillo Chacon, E., Schaffner, C. M., Werden, L. K., Xu, X., & Medvigy, D. (2020). A catastrophic tropical drought kills hydraulically vulnerable tree species. *Glob Chang Biol*, *26*(5), 3122-3133. https://doi.org/10.1111/gcb.15037

Preston, J. C., & Sandve, S. R. (2013). Adaptation to seasonality and the winter freeze. *Front Plant Sci*, *4*, 167. https://doi.org/10.3389/fpls.2013.00167

Proost, S., & Mutwil, M. (2017). PlaNet: Comparative Co-Expression Network Analyses for Plants. In A. D. J. van Dijk (Ed.), *Plant Genomics Databases: Methods and Protocols* (pp. 213-227). Springer New York. https://doi.org/10.1007/978-1-4939-6658-5_12

Puranik, S., Sahu, P. P., Srivastava, P. S., & Prasad, M. (2012). NAC proteins: regulation and role in stress tolerance. *Trends in Plant Science*, *17*(6), 369-381. https://doi.org/https://doi.org/10.1016/j.tplants.2012.02.004

Qi, X., Fuller, E., Wu, Q., Wu, Y., & Zhang, C.-Q. (2012). Laplacian centrality: A new centrality measure for weighted networks. *Information Sciences*, *194*, 240-253. https://doi.org/https://doi.org/10.1016/j.ins.2011.12.027

Rahmatallah, Y., Emmert-Streib, F., & Glazko, G. (2013). Gene Sets Net Correlations Analysis (GSNCA): a multivariate differential coexpression test for gene sets. *Bioinformatics*, *30*(3), 360-368. https://doi.org/10.1093/bioinformatics/btt687

Ramos, A., Pérez-Solís, E., Ibáñez, C., Casado, R., Collada, C., Gómez, L., Aragoncillo, C., & Allona, I. (2005). Winter disruption of the circadian clock in chestnut. *Proceedings of the National Academy of Sciences*, *102*(19), 7037-7042. https://doi.org/doi:10.1073/pnas.0408549102

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., & Laaksonen, A. (2022). The Arctic has warmed

nearly four times faster than the globe since 1979. *Communications Earth & Environment*, *3*(1), 168. https://doi.org/10.1038/s43247-022-00498-3

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

Raza, A., Charagh, S., Sadaqat, N., & Jin, W. (2020). Arabidopsis thaliana: Model Plant for the Study of Abiotic Stress Responses. In M. Hasanuzzaman (Ed.), *The Plant Family Brassicaceae: Biology and Physiological Responses to Environmental Stresses* (pp. 129-180). Springer Singapore. https://doi.org/10.1007/978-981-15-6345-4_3

Richardson, A. D., Schädel, C., Westergaard-Nielsen, A., Novick, K. A., Basler, D., Phillips, J. R., Krassovski, M. B., Warren, J. M., Sebestyen, S. D., & Hanson, P. J. (2024). Experimental Whole-Ecosystem Warming Enables Novel Estimation of Snow Cover and Depth Sensitivities to Temperature, and Quantification of the Snow-Albedo Feedback Effect. *Journal of Geophysical Research: Biogeosciences*, *129*(3), e2023JG007833. https://doi.org/https://doi.org/10.1029/2023JG007833

Rihan, H. Z., Al-Issawi, M., & Fuller, M. P. (2017). Advances in physiological and molecular aspects of plant cold tolerance. *Journal of Plant Interactions*, *12*(1), 143-157. https://doi.org/10.1080/17429145.2017.1308568

Rong, M., Gao, S. X., Wen, D., Xu, Y. H., & Wei, J. H. (2024). The LOB domain protein, a novel transcription factor with multiple functions: A review. *Plant Physiol Biochem*, *214*, 108922. https://doi.org/10.1016/j.plaphy.2024.108922

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *The European Physical Journal Special Topics*, *178*(1), 13-23. https://doi.org/10.1140/epjst/e2010-01179-1

Roychoudhury, A., Paul, S., & Basu, S. (2013). Cross-talk between abscisic acid-dependent and abscisic acid-independent pathways during abiotic stress. *Plant Cell Reports*, *32*(7), 985-1006. https://doi.org/10.1007/s00299-013-1414-5

Ryan, C. J., Roguev, A., Patrick, K., Xu, J. W., Jahari, H., Tong, Z. T., Beltrao, P., Shales, M., Qu, H., Collins, S. R., Kliegman, J. I., Jiang, L. L., Kuo, D., Tosti, E., Kim, H. S., Edelmann, W., Keogh, M. C., Greene, D., Tang, C.,...Krogan, N. J. (2012). Hierarchical Modularity and the Evolution of Genetic Interactomes across Species. *Molecular Cell*, *46*(5), 691-704. https://doi.org/10.1016/j.molcel.2012.05.028

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. *Physiol Plant*, *177*(5), e70567. https://doi.org/10.1111/ppl.70567

Sade, N., Gebremedhin, A., & Moshelion, M. (2012). Risk-taking plants: anisohydric behavior as a stress-resistance trait. *Plant Signal Behav*, *7*(7), 767-770. https://doi.org/10.4161/psb.20505

Salmon, Y., Torres-Ruiz, J. M., Poyatos, R., Martinez-Vilalta, J., Meir, P., Cochard, H., & Mencuccini, M. (2015). Balancing the risks of hydraulic failure and carbon starvation: a twig scale analysis in declining Scots pine. *Plant, Cell & Environment*, *38*(12), 2575-2588. https://doi.org/https://doi.org/10.1111/pce.12572

Sang, S., Zhou, Y., & Liu, Y. (2025). Genome-wide analysis of the CBF gene family and their transcriptional response to cold stress in Hibiscus mutabilis. *Scientific Reports*, *15*(1), 23808. https://doi.org/10.1038/s41598-025-05040-x

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, *74*(12), 5463-5467. https://doi.org/10.1073/pnas.74.12.5463

Sardoei, A. S., & Fazeli-Nasab, B. (2025). Physiological and Molecular Mechanisms of Freezing in Plants. *Phyton-International Journal of Experimental Botany*, *94*(6), 1601-1630. https://doi.org/https://doi.org/10.32604/phyton.2025.064729

Sasidharan, R., & Gerstein, M. (2008). Protein fossils live on as RNA. *Nature*, *453*(7196), 729-731. https://doi.org/10.1038/453729a

Satyakam, Zinta, G., Singh, R. K., & Kumar, R. (2022). Cold adaptation strategies in plants—An emerging role of epigenetics and antifreeze proteins to engineer cold resilient plants [Review]. *Frontiers in Genetics*, *Volume 13 - 2022*. https://doi.org/10.3389/fgene.2022.909007

Savino, A., Provero, P., & Poli, V. (2020). Differential Co-Expression Analyses Allow the Identification of Critical Signalling Pathways Altered during Tumour Transformation and Progression. *International Journal of Molecular Sciences*, *21*(24), 9461. https://www.mdpi.com/1422-0067/21/24/9461

Schäfer, J., & Strimmer, K. A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. *Statistical Applications in Genetics and Molecular Biology*, *4*(1). https://doi.org/doi:10.2202/1544-6115.1175

Schlitt, T., & Brazma, A. (2007). Current approaches to gene regulatory network modelling. *BMC Bioinformatics*, *8*(6), S9. https://doi.org/10.1186/1471-2105-8-S6-S9

Schroeder, J. I., & Hagiwara, S. (1990). Repetitive Increases in Cytoslic Ca²⁺ of Guard Cells by Abscisic Acid Activation of Nonselective Ca²⁺ Permeable Channels. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(23), 9305-9309. http://www.jstor.org/stable/2355810

Seidl, R., Thom, D., Kautz, M., Martin-Benito, D., Peltoniemi, M., Vacchiano, G., Wild, J., Ascoli, D., Petr, M., Honkaniemi, J., Lexer, M. J., Trotsiuk, V., Mairota, P., Svoboda, M., Fabrika, M., Nagel, T. A., & Reyer, C. P. O. (2017). Forest disturbances under climate change. *Nature Climate Change*, *7*(6), 395-402. https://doi.org/10.1038/nclimate3303

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y., & Shinozaki, K. (2002).

Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant Journal*, *31*(3), 279-292. https://doi.org/https://doi.org/10.1046/j.1365-313X.2002.01359.x

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379-423. https://doi.org/https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shinozaki, K., & Yamaguchi-Shinozaki, K. (2000). Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Current opinion in plant biology*, *3*(3), 217-223.

Shutta, K. H., De Vito, R., Scholtens, D. M., & Balasubramanian, R. (2022). Gaussian graphical models with applications to omics analyses. *Stat Med*, *41*(25), 5150-5187. https://doi.org/10.1002/sim.9546

Sigman, M. J., & Slotkin, R. K. (2016). The First Rule of Plant Transposable Element Silencing: Location, Location, Location. *Plant Cell*, *28*(2), 304-313. https://doi.org/10.1105/tpc.15.00869

Singh, D., & Laxmi, A. (2015). Transcriptional regulation of drought response: a tortuous network of transcriptional factors. *Front Plant Sci*, *6*, 895. https://doi.org/10.3389/fpls.2015.00895

Snyder, P. K., Delire, C., & Foley, J. A. (2004). Evaluating the influence of different vegetation biomes on the global climate. *Climate Dynamics*, *23*(3), 279-302. https://doi.org/10.1007/s00382-004-0430-0

Song, L., Koga, Y., & Ecker, J. R. (2016). Profiling of transcription factor binding events by chromatin immunoprecipitation sequencing (ChIP-seq). *Curr Protoc Plant Biol*, *1*(2), 293-306. https://doi.org/10.1002/cppb.20014

Song, X., Hu, J., Wu, T., Yang, Q., Feng, X., Lin, H., Feng, S., Cui, C., Yu, Y., Zhou, R., Gong, K., Yu, T., Pei, Q., & Li, N. (2021). Comparative analysis of long noncoding RNAs in angiosperms and characterization of long noncoding RNAs in response to heat stress in Chinese cabbage. *Horticulture Research*, *8*(1), 48. https://doi.org/10.1038/s41438-021-00484-4

Sperry, J. S., Meinzer, F. C., & McCulloh, K. A. (2008). Safety and efficiency conflicts in hydraulic architecture: scaling from tissues to trees. *Plant, Cell & Environment*, *31*(5), 632-645. https://doi.org/https://doi.org/10.1111/j.1365-3040.2007.01765.x

Sperry, J. S., & Tyree, M. T. (1988). Mechanism of water stress-induced xylem embolism. *Plant Physiol*, *88*(3), 581-587. https://doi.org/10.1104/pp.88.3.581

Su, H., Zhang, S., Yuan, X., Chen, C., Wang, X. F., & Hao, Y. J. (2013). Genome-wide analysis and identification of stress-responsive genes of the NAM-ATAF1,2-CUC2 transcription factor family in apple. *Plant Physiol Biochem*, *71*, 11-21. https://doi.org/10.1016/j.plaphy.2013.06.022

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y. C., Sjödin, A., Van de Peer, Y., Jansson, S., Hvidsten, T. R., & Street, N. R. (2015). The Plant Genome Integrative Explorer Resource: PlantGenIE.org. *New Phytol*, *208*(4), 1149-1156. https://doi.org/10.1111/nph.13557

Tesson, B. M., Breitling, R., & Jansen, R. C. (2010). DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics*, *11*(1), 497. https://doi.org/10.1186/1471-2105-11-497

Tian, L., Ku, L., Yuan, Z., Wang, C., Su, H., Wang, S., Song, X., Dou, D., Ren, Z., Lai, J., Liu, T., Du, C., & Chen, Y. (2021). Large-scale reconstruction of chromatin structures of maize temperate and tropical inbred lines. *Journal of Experimental Botany*, *72*(10), 3582-3596. https://doi.org/10.1093/jxb/erab087

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267-288. http://www.jstor.org/stable/2346178

Tourdot, E., & Grob, S. (2023). Three-dimensional chromatin architecture in plants – General features and novelties. *European Journal of Cell Biology*, *102*(4), 151344. https://doi.org/https://doi.org/10.1016/j.ejcb.2023.151344

Touzdjian Pinheiro Kohlrausch Távora, F., de Assis dos Santos Diniz, F., de Moraes Rêgo-Machado, C., Chagas Freitas, N., Barbosa Monteiro Arraes, F., Chumbinho de Andrade, E., Furtado, L. L., Osiro, K. O., Lima de Sousa, N., Cardoso, T. B., Márcia Mertz Henning, L., Abrão de Oliveira Molinari, P., Feingold, S. E., Hunter, W. B., Fátima Grossi de Sá, M., Kobayashi, A. K., Lima Nepomuceno, A., Santiago, T. R., & Correa Molinari, H. B. (2022). CRISPR/Cas- and Topical RNAi-Based Technologies for Crop Management and Improvement: Reviewing the Risk Assessment and Challenges Towards a More Sustainable Agriculture [Review]. *Frontiers in Bioengineering and Biotechnology*, *Volume 10 - 2022*. https://doi.org/10.3389/fbioe.2022.913728

Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 5233. https://doi.org/10.1038/s41598-019-41695-z

True, J. R., & Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evolution & Development*, *3*(2), 109-119. https://doi.org/https://doi.org/10.1046/j.1525-142x.2001.003002109.x

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F. M., Bassel, G. W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S., & Provart, N. J. (2009). Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant, Cell & Environment*, *32*(12), 1633-1651. https://doi.org/https://doi.org/10.1111/j.1365-3040.2009.02040.x

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., & de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in Bioinformatics*, *19*(4), 575-592. https://doi.org/10.1093/bib/bbw139

van Steensel, B., & Furlong, E. E. M. (2019). The role of transcription in shaping the spatial organization of the genome. *Nature Reviews Molecular Cell Biology*, *20*(6), 327-337. https://doi.org/10.1038/s41580-019-0114-6

Vercruysse, J., Van Bel, M., Osuna-Cruz, C. M., Kulkarni, S. R., Storme, V., Nelissen, H., Gonzalez, N., Inzé, D., & Vandepoele, K. (2020). Comparative transcriptomics enables the identification of functional orthologous genes

involved in early leaf growth. *Plant Biotechnology Journal*, *18*(2), 553-567. https://doi.org/https://doi.org/10.1111/pbi.13223

Vergara, A., Haas, J. C., Aro, T., Stachula, P., Street, N. R., & Hurry, V. (2022). Norway spruce deploys tissue-specific responses during acclimation to cold. *Plant, Cell & Environment*, *45*(2), 427-445. https://doi.org/https://doi.org/10.1111/pce.14241

Wang, M., Wang, P., Lin, M., Ye, Z., Li, G., Tu, L., Shen, C., Li, J., Yang, Q., & Zhang, X. (2018). Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nature Plants*, *4*(2), 90-97. https://doi.org/10.1038/s41477-017-0096-3

Wang, X., Bernhardsson, C., & Ingvarsson, P. K. (2020). Demography and Natural Selection Have Shaped Genetic Variation in the Widely Distributed Conifer Norway Spruce (Picea abies). *Genome Biol Evol*, *12*(2), 3803-3817. https://doi.org/10.1093/gbe/evaa005

Wang, Z., Wang, F., Hong, Y., Yao, J., Ren, Z., Shi, H., & Zhu, J.-K. (2018). The Flowering Repressor SVP Confers Drought Resistance in Arabidopsis by Regulating Abscisic Acid Catabolism. *Molecular Plant*, *11*(9), 1184-1197. https://doi.org/https://doi.org/10.1016/j.molp.2018.06.009

Wilczyński, B., & Dojer, N. (2009). BNFinder: exact and efficient method for learning Bayesian networks. *Bioinformatics*, *25*(2), 286-287. https://doi.org/10.1093/bioinformatics/btn505

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R.,…Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wisecaver, J. H., Borowsky, A. T., Tzin, V., Jander, G., Kliebenstein, D. J., & Rokas, A. (2017). A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. *The Plant Cell*, *29*(5), 944-959. https://doi.org/10.1105/tpc.17.00009

Wolfe, C. J., Kohane, I. S., & Butte, A. J. (2005). Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics*, *6*, 227. https://doi.org/10.1186/1471-2105-6-227

Wu, Y., Deng, Z., Lai, J., Zhang, Y., Yang, C., Yin, B., Zhao, Q., Zhang, L., Li, Y., Yang, C., & Xie, Q. (2009). Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. *Cell Res*, *19*(11), 1279-1290. https://doi.org/10.1038/cr.2009.108

Xie, T., Zhang, F.-G., Zhang, H.-Y., Wang, X.-T., Hu, J.-H., & Wu, X.-M. (2019). Biased gene retention during diploidization in Brassica linked to three-dimensional genome organization. *Nature Plants*, *5*(8), 822-832. https://doi.org/10.1038/s41477-019-0479-8

Xing, D.-H., Lai, Z.-B., Zheng, Z.-Y., Vinod, K. M., Fan, B.-F., & Chen, Z.-X. (2008). Stress- and Pathogen-Induced Arabidopsis WRKY48 is a Transcriptional Activator that Represses Plant Basal Defense. *Molecular Plant*, *1*(3), 459-470. https://doi.org/https://doi.org/10.1093/mp/ssn020

Xu, T., Niu, J., & Jiang, Z. (2022). Sensing Mechanisms: Calcium Signaling Mediated Abiotic Stress in Plants. *Frontiers in Plant Science*, *13*, 925863. https://doi.org/10.3389/fpls.2022.925863

Yamaguchi-Shinozaki, K., & Shinozaki, K. (2006). Transcriptional Regulatory Networks in Cellular Responses and Tolerance to Dehydration and Cold Stresses. *Annual Review of Plant Biology*, *57*(Volume 57, 2006), 781-803. https://doi.org/https://doi.org/10.1146/annurev.arplant.57.032905.105444

Yan, K.-K., Wang, D., Rozowsky, J., Zheng, H., Cheng, C., & Gerstein, M. (2014). OrthoClust: an orthology-based network framework for clustering data across multiple species. *Genome Biology*, *15*(8), R100. https://doi.org/10.1186/gb-2014-15-8-r100

Yang, Y., Han, L., Yuan, Y., Li, J., Hei, N., & Liang, H. (2014). Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nature Communications*, *5*(1), 3231. https://doi.org/10.1038/ncomms4231

Yaschenko, A. E., Alonso, J. M., & Stepanova, A. N. (2024). Arabidopsis as a model for translational research. *The Plant Cell*, *37*(5). https://doi.org/10.1093/plcell/koae065

Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., Nurkowski, K. A., Smets, P., Wang, T., Gray, L. K., Liepe, K. J., Hamann, A., Holliday, J. A., Whitlock, M. C., Rieseberg, L. H., & Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, *353*(6306), 1431-1433. https://doi.org/10.1126/science.aaf7812

Yu, H., Liu, B. H., Ye, Z. Q., Li, C., Li, Y. X., & Li, Y. Y. (2011). Link-based quantitative methods to identify differentially coexpressed genes and gene pairs. *BMC Bioinformatics*, *12*, 315. https://doi.org/10.1186/1471-2105-12-315

Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J., & Jarvis, E. D. (2004). Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics*, *20*(18), 3594-3603. https://doi.org/10.1093/bioinformatics/bth448

Yuan, P., Yang, T., & Poovaiah, B. W. (2018). Calcium Signaling-Mediated Plant Response to Cold Stress. *International Journal of Molecular Sciences*, *19*(12), 3896. https://www.mdpi.com/1422-0067/19/12/3896

Zabidi, M. A., & Stark, A. (2016). Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet*, *32*(12), 801-814. https://doi.org/10.1016/j.tig.2016.10.003

Zhang, Y., She, J., Hu, X., Jin, Y., Zhao, J., Hou, S., Tao, C., Du, M., & Yang, E. (2025). Transposable elements drive species-specific and tissue-specific transcriptomes in human development. *Genome Biol*, *26*(1), 379. https://doi.org/10.1186/s13059-025-03855-5

Zhao, Q., Han, R., Cai, K., Yan, H., Li, Y., Qu, G., Liu, L., & Zhao, X. (2023). Identification and Analysis of the CBF Gene Family in Three Species of Acer under Cold Stress. *Int J Mol Sci*, *24*(3). https://doi.org/10.3390/ijms24032088

Zhong, R., Allen, J. D., Xiao, G., & Xie, Y. (2014). Ensemble-based network aggregation improves the accuracy of gene network reconstruction. *PLOS ONE*, *9*(11), e106319. https://doi.org/10.1371/journal.pone.0106319

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, *67*(2), 301-320. https://EconPapers.repec.org/RePEc:bla:jorssb:v:67:y:2005:i:2:p:301-320