Comparative transcriptomics and network reconstruction with applications to auxin signaling

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät III
Agrar- und Ernährungswissenschaften,
Geowissenschaften und Informatik

der Martin-Luther-Universität Halle–Wittenberg,

vorgelegt von

Frau Pöschl, Yvonne

Geb. am 21.01.1981 in Wolfen

Gutachter:
1. Prof. Dr. Ivo Große
2. Dr. Dirk Walther

Datum der Verteidigung: 29.06.2016

Contents

List of abbreviations

A. lyrata	*Arabidopsis lyrata*
ANOVA	Analysis of variance
ARF	AUXIN RESPONSE FACTOR
AS	Amino acid sequence
A. thaliana	*Arabidopsis thaliana*
AUX/IAA	AUXIN/INDOLE-3-ACETIC ACID
AuxRE	Auxin-responsive element
DNA	Deoxyribonucleic acid
HCLUST	Hierarchical CLUSTer method
HOPACH	Hierarchical Ordered Partitioning And Collapsing Hybrid
IAA	INDOLE-3-ACETIC ACID
LCF	Local Context Finder
Mio.	Million
mRNA	messenger Ribonucleic acid
PIF	Profile Interaction Finder
PMP	Probe Masking Pipeline
RMA	Robust Multi-array Average (normalization)
RNA	Ribonucleic acid
RT-qPCR	Real-Time quantitative Polymerase Chain Reaction
ssp.	*subspecies*
TIR1/AFB	TRANSPORT INHIBITOR RESPONSE1/AUXIN SIGNALING F-BOX1-5

1. Summary

1.1. English version

Genes code the blue prints for proteins and need to undergo the molecular processes of transcription and subsequent translation to result in the proteins they code for. The amount of protein is mainly determined by the amount of available transcript. Hence, inspecting the amount of transcript of a gene gives information about its expression level. The more transcript is present for a gene the higher the gene is expressed and the more protein can be synthesized. Proteins have different functions and are involved in different processes, like enzymes that change the activity of proteins or transcription factors that regulate the transcriptional process. In summary, the expression of a gene depends on other genes or their corresponding proteins, which are regulated themselves. Hence, the regulatory interaction of genes can be described as a network where the nodes represent the genes and the edges represent regulatory relationships. If the expression level of one gene is changed, this change affects other genes and thus triggers a cascade that propagates the change through the network.

The expression level of genes can be altered in response to a signal. A signal is perceived and transduced by the corresponding signaling network. This signaling network translates the signal into gene responses by affecting the transcription of genes that lead to an increase or decrease in the amount of the respective transcripts.

We have developed algorithms for inspecting the responses of thousands of genes. We applied these algorithms to study expression responses of genes from the plant species *Arabidopsis thaliana* and its closely related sister species *Arabidopsis lyrata* to treatment with the signal molecule auxin. Although both species are closely related they show differences in their genomic sequences that have to be considered.
While for the well studied species *A. thaliana*, the infrastructure for measuring the expression of thousands of genes by microarrays is available and well established, it is not available for *A. lyrata*. Due to the fact that no microarray is available for *A. lyrata* we chose the microarray that was specifically designed to target transcripts of *A. thaliana*. We have developed the PMP (Probe Masking Pipeline) algorithm that makes use of transcript sequences and therefore can deal with the problems that arise due to differences in the genomic sequence of *A. lyrata* and *A. thaliana* and provides reliable and comparable expression values for *A. thaliana* and *A. lyrata*. The PMP is designed in a modular fashion and can be applied to different use cases. It is capable of providing reliable expression values not only for a single species but also for two or more species by taking their transcript sequences into account simultaneously which is necessary for comparing gene responses of closely related species.

To inspect the expression responses of genes along a set of different experiments or samples (expression profiles) or rather to inspect potential regulatory relationship of genes, we have developed the PIF (Profile Interaction Finder) algorithm employing a linear model. The PIF algorithm inspects the relationship of the expression profiles of genes by reconstructing the expression profile of a gene as a linear combination of the expression profiles of other genes. We used the inferred relationships between the genes to reconstruct networks in which the nodes represent the genes and the edges represent the relations. To distinguish between relationships that are inferred because of similar or opposite expression responses of genes, we incorporated an additional set of parameters which is directly attached to the weights of the linear model. We refer to a positive relationship if two genes show a similar expression response and to a negative relationship if two genes show an opposite expression response. Therefore the set of edges comprises two subsets of edges, the first representing the positive relationships and the second representing the negative relationships.

We inferred the positive relationships of genes from different *A. thaliana* ecotypes, to inspect the expression response of genes to auxin within the *A. thaliana* species. In performing this intra-species comparison, we statistically evaluated the amplitudes of gene responses to auxin and the topology of the reconstructed networks. We found evidence for the existence of natural variation in the gene responses, especially for the genes coding for the components of the auxin signaling network. This finding lead to a model of how responses of genes in the auxin signaling network affect each other and downstream responding genes.
We expanded the analysis of auxin gene responses to an inter-species comparison of *A. thaliana* and *A. lyrata*. We applied the PMP to obtain reliable estimates for gene responses of *A. lyrata*. We inferred networks from gene expression profiles of both species using the PIF algorithm and subsequently evaluated positive and negative relationships between genes. We observed that a set of genes shows very conserved responses to auxin and concluded that this set of genes comprises genes that might be essential for auxin response. However, we also spotted genes showing a very different auxin response in both species and concluded that these genes might be responsible for different downstream responses in *A. thaliana* and *A. lyrata* as proposed in the model derived from the intra-species comparison.

We also found evidence for naturally occurring variation in the expression of reproductive traits of different ecotypes of *A. thaliana* in response to ambient temperature changes. We obtained these findings from inspecting traits measured along entire life cycles of different *A. thaliana* ecotypes at different ambient temperatures. Hence, for each trait we had measurements at different temperatures for different ecotypes of *A. thaliana*. To analyze the impact of ambient temperature change on the expression of each trait in each ecotype, we fitted a linear model. The inspection of the absolute value and the sign of slope parameter of the fitted linear model allowed us to distinguish between traits that have always the same sign for all ecotypes or have different signs. The second group possibly constitutes traits that show variation due to natural variation. But to dissect the effect of the ecotype and the effect of temperature, we presented a measure based on the intra-class correlation coefficient. To this end, we analyzed the decomposed total variance for each of the traits in two ways: (i) for the impact of the ecotype and (ii) for the impact of temperature. By evaluating both measures for all traits capturing an entire life cycle, we identified the reproductive traits as highly affected by ecotype

and temperature and thus as worthwhile candidate traits for further scientific investigation and breeding.

We showed that specific biological questions lead to new bioinformatics algorithms whose application in turn provides new insights into biological systems.

1.2. German version

Gene kodieren die Baupläne für Proteine. Durch die Transkription der Gene und die anschließende Translation des Transkriptes werden Proteine synthetisiert, wobei die Menge des synthetisieren Proteins hauptsächlich von der Menge an verfügbarem Transkript abhängt. Eine Analyse der zur Verfügung stehenden Transkriptmenge eines Genes gibt also Hinweise auf dessen Expressionszustand. Je mehr Transkript eines Genes verfügbar ist, desto stärker ist das Gen exprimiert und desto mehr Protein kann synthetisiert werden. Proteine haben verschiedene Funktionen und sind in unterschiedliche Prozesse involviert, wie z.B. Enzyme, die die Aktivität von Proteinen verändern oder Transkriptionsfaktoren, die die Transkription der Gene regulieren. Im Allgemeinen hängt die Expression eines Genes von anderen Genen bzw. deren korrespondierenden Proteinen ab, die aber wiederum auch der Regulation unterliegen. Die regulatorischen Zusammenhänge zwischen Genen lassen sich durch Netzwerke beschreiben, in welchen die Knoten die Gene und die Kanten mögliche regulatorische Beziehungen zwischen Genen repräsentieren. Ändert sich die Expression eines Genes, wirkt sich dies auch auf die Expression anderer Gene aus. Es wird eine Kaskade in Gang gesetzt, welche die Änderung durch das Netzwerk propagiert.

Die Expression eines Genes kann auf ein Signal hin verändert werden. Signale werden durch das entsprechende Signalnetzwerk wahrgenommen und weitergeleitet. Das Signalnetzwerk überführt das Signal in Genreaktionen, indem die Transkription der Gene beeinflusst wird. Dies hat eine Verringerung oder Erhöhung der zur Verfügung stehenden Transkriptmenge zur Folge.

Wir haben Algorithmen entwickelt, die der Analyse der Reaktion tausender Gene dienen. Diese haben wir eingesetzt um die Genreaktion der nah verwandten Pflanzenspezies *Arabidopsis thaliana* und *Arabidopsis lyrata* auf Behandlung mit Auxin zu studieren. Obwohl beide Spezies nah verwandt sind, existieren nicht zu vernachlässigende Unterschiede in ihren genomischen Sequenzen.
Für die gut erforschte Pflanzenspezies *A. thaliana* steht sowohl ein Microarray zum Messen der Genexpression als auch die zugehörige etablierte Infrastruktur zur Verfügung. Allerdings ist das für *A. lyrata* nicht der Fall. Aus diesem Grund haben wir auch für *A. lyrata* auf das Microarray, welches spezifisch zum Messen von *A. thaliana*-Gen-Transkripten geschaffen wurde, zurückgegriffen. Wir haben den PMP-(Probe Masking Pipeline)-Algorithmus entwickelt um Probleme zu kompensieren, die durch die genomischen Unterschiede von *A. thaliana* und *A. lyrata* hervorgerufen werden. Hierführ bezieht der PMP-Algorithmus die Sequenzen der Transkripte mit ein und liefert am Ende verlässliche und vergleichbare Genexpressionswerte für *A. thaliana* und *A. lyrata*. Der PMP-Algorithmus hat durch seinen modularen Aufbau vielfältige Anwendungsbereiche. Er liefert nicht nur verlässliche Genexpressionswerte für eine Spezies sondern auch für mehrere, indem er die Sequenzen der Transkripte aller Spezies gleichzeitig berücksichtigt. Letzteres ist dann erforderlich, wenn die Genexpression mehrerer nah verwandter Spezies miteinander verglichen werden soll.

Um das Expressionsverhalten von Genen über mehrere Experimente (Expressionsprofile) hinweg bzw. mögliche regulatorischen Beziehungen zwischen Genen untersuchen zukönnen, haben

wir den PIF-(Profile Interaction Finder)-Algorithmus entwickelt. Dieser beinhaltet als Kernstück ein lineares Modell, das verwendet wird, um das Expressionsprofil eines Genes durch Linearkombination der Expressionsprofile anderer Gene zu rekonstruieren. Die so ermittelten Beziehungen zwischen den Genen haben wir in Netzwerken dargestellt, in denen die Knoten die Gene und die Kanten die ermittelten Beziehungen zwischen den Gene repräsentieren.
Das Expressionsverhalten von Genen, die unter dem gleichen regulatorischen Einfluss stehen, kann gleich oder entgegengesetzt sein. Um zwischen diesen beiden Fällen unterscheiden zu können, haben wir zusätzliche Parameter, die in direkter Beziehung zu den Gewichten des linearen Modells stehen, eingeführt. Ist im Netzwerk eine Kante durch ein sehr ähnliches Expressionsverhalten zweier Gene zustande gekommen, bezeichnen wir diese Beziehung als positive Beziehung. Ist hingegen die Kante durch ein entgegengesetztes Expressionsverhalten zweier Gene zustande gekommen, bezeichnen wir die Beziehung als negative Beziehung. Im ersten Fall hat ein potentiell gemeinsamer regulatorischer Einfluss den gleichen Effekt im Expressionsverhalten beider Gene ausgelöst oder eines der Gene wirkt positiv regulierend auf die Expression des anderen Genes. Wohingegen im zweiten Fall durch einen potentiell gemeinsamen oder direkten regulatorischen Einfluss ein entgegengesetzter Effekt im Expressionsverhalten hervorgerufen wurde.

Für die vergleichende Analyse des Expressionsverhalten von Genen verschiedener *A. thaliana*-Ökotypen unter Auxinbehandlung haben wir Genexpressionsnetzwerke für Ökotypen basierend auf den positiven Beziehungen rekonstruiert. Dieser Intra-Spezies-Vergleich beinhaltete die statistische Analyse der Stärke der Genexpression sowie die statistische Analyse der Topologie der rekonstruierten Netzwerke. Wir fanden Anhaltspunkte für die Existenz einer natürlichen Variation im Expressionsverhalten der Gene, insbesondere bei Genen, welche die Komponenten des Auxin-Signal-Netzwerkes kodieren. Diese Erkenntnis führte zu einem Modell, das den Einfluss des Expressionsverhalten der Gene des Auxin-Signal-Netzwerks untereinander und auf das Expressionsverhalten nachfolgender Gene zeigt.

Nach dem Intra-Spezies-Vergleich von *A. thaliana* erweiterten wir die vergleichende Analyse auf einen Inter-Spezies-Vergleich von *A. thaliana* und *A. lyrata*. Wir wendeten den PMP-Algorithmus an, um auch für *A. lyrata* verläßliche Expressionswerte für diesen Vergleich zur Verfügung zu haben. Unter Verwendung des PIF-Algorithmus rekonstruierten wir Expressionsnetzwerke beider Spezies und werteten sowohl die positiven als auch die negativen Beziehungen aus. Wir ermittelten eine Gruppen von Genen, die ein sehr ähnliches Expressionverhalten in Bezug auf die Auxinbehandlung zeigt und folgerten, dass diese Gengruppe essenziell für die Auxinantwort sein könnte. Wir ermittelten eine weitere Gruppe von Genen, die ein unterschiedliches Expressionsverhalten in beiden Spezies zeigten. Wir folgerten, dass diese Gene für unterschiedliche nachfolgende Auxinantworten verantwortlich sein könnten. Dies steht in Übereinstimmung mit dem Modell, das aus dem Intra-Spezies-Vergleich abgeleitet wurde.

Wir fanden auch Hinweise auf natürliche Variation in der Ausbildung von Merkmalen der reproduktiven Phase verschiedener *A. thaliana*-Ökotypen als Reaktion auf veränderte Umgebungstemperaturen. Wir erlangten diese Erkenntnisse durch die Analyse von Merkmalen, die über vollständige Lebenszyklen verschiedener *A. thaliana*-Ökotypen bei verschiedenen Umgebungstemperaturen gemessen wurden. Für jedes dieser Merkmale hatten wir Messungen für die verschiedenen *A. thaliana*-Ökotypen zu den verschiedenen Umgebungstemperaturen zur

Verfügung. Um den Einfluss der Umgebungstemperatur auf die Expression eines Merkmales zu untersuchen, haben wir für jeden Ökotypen ein lineares Modell gefittet. Durch die Analyse der Stärke und des Vorzeichens des Steigungsparameters des gefitteten linearen Modells, konnten wir die Merkmale unterscheiden in solche, die in allen Ökotypen das gleiche Vorzeichen hatten und in solche die unterschiedliche Vorzeichen hatten. Die unterschiedlichen Vorzeichen in der letzteren Gruppe könnten auf natürlicher (genetischer) Variation in den Ökotypen beruhen. Um aber den Einfluss der Ökotypen und den Einfluss der Umgebungstemperatur zu untersuchen, haben wir ein Maß basierend auf dem Intra-Klassen-Korrelationskoeffizienten entwickelt. Unter Verwendung dieses Maßes wird die Gesamtvarianz eines jeden Merkmales zerlegt und analysiert auf (i) den Einfluss durch die Ökotypen und (ii) den Einfluss durch die Temperatur. Durch die Bewertung beider Einflussfaktoren aller Merkmale des gesamten Lebenszyklusses, konnten wir die Merkmale, die die reproduktive Phase beschreiben als diejenigen identifizieren, die am stärksten durch die Ökotypen und die Umgebungstemperatur beinflusst wurden. Diese Merkmale wären vielversprechende Kandidaten für nachfolgende wissenschaftliche Untersuchungen oder für die Pflanzenzucht.

Wir haben gezeigt, dass gezielte biologische Fragen zur Entwicklung neuer bioinformatischer Algorithmen führen, deren Anwendung wiederum zu neuen Einblicken in biologische Systeme führt.

2. Introduction

Organisms are organized in organs, tissues, and cells, where the cells are the smallest unit that contains the genetic information. The genetic information is stored in form of genes in the DNA (Deoxyribonucleic acid). Genes code blueprints for proteins that control processes in the organism. If the information stored by a specific gene is needed then a working copy of the respective gene is generated by transcription. Subsequently, the working copy of a gene is translated into a protein with a specific function. Some proteins regulate the transcription of genes, but proteins can also regulate other proteins by changing their activity. Combinations of different genes, more precisely of proteins produced from different genes, control different processes. Such processes could, for example, be important for the survival of the organism or its appearance.

It was observed that organisms with nearly identical genetic information show differences in their appearance, although they were exposed to the same environmental conditions. This leads to the conclusion that somehow the processes and more precisely their regulatory mechanisms have changed. In particular, we aim at identifying the processes and understanding the regulatory mechanisms that are behind these processes. We also aim at comparing regulatory mechanisms of processes, to find and understand similarities and differences and their impact on the appearance of an organism. To achieve these goals we developed various bioinformatics algorithms that are presented in this thesis. We have designed algorithms to compute the amount of working copies of genes from measurements and also to uncover regulatory mechanisms, which is to uncover the relationships of genes that determine specific processes, e.g., different enzymatic processes.

In this context, we developed bioinformatics algorithms to facilitate the analysis of measurements from the plant genus *Arabidopsis* exposed to an auxin stimulus.

2.1. Biological background

This section introduces the reader into gene expression and its regulation. The introduction also includes a general description of how signals are transduced in the plants by means of gene expression and additionally it includes a more detailed description of this process for the signal molecule auxin.

2.1.1. Gene expression

Whenever a protein having a specific function (e.g., an enzyme) is needed then the corresponding gene needs to be expressed. The process of gene expression comprises two main processes, transcription and translation, and related post-processing steps (Figure 2.1). The expression of a gene starts with the process of transcription, where the DNA sequence of the gene is transcribed into the corresponding RNA (Ribonucleic acid) sequence. This process is driven by the binding of transcription-regulating proteins (transcription factors) to regulatory elements (specific short sequences) in the promoter region (upstream) of the genes. The binding of a transcription factor to its corresponding regulatory element can either activate or repress the transcription of a gene. Besides transcription factors, several additional proteins play a role in the transcription process. After transcription-related post-processing steps the transcription of the gene results in the messenger RNA (mRNA). The mRNA serves as input of the translation process where the mRNA sequence is translated into the corresponding amino acid sequence (AS). After several post-processing and folding steps this sequence of amino acids results in a mature protein. Proteins are also often referred to as gene products. Proteins have special functions, e.g., they are enzymes and catalyze enzymatic reactions or they are transcription factors and regulate the transcription of genes.

The whole gene expression process is regulated by different specific mechanisms on the transcriptional and translational level.

For simplicity, we assume a gene to be expressed whenever mRNA of this gene is present and do not take into account whether the corresponding protein is synthesized or not. Measuring the amount of mRNA that is available in different experimental setups in a high-throughput manner (e.g. using expression microarrays) is more convenient than measuring complex proteins. All available transcripts (mRNAs) taken together are denoted as the transcriptome. Depending on the amount of available mRNA of a gene we can assume how strong a gene is expressed. A gene is highly expressed if there is a high amount of its mRNA available, whereas it is lowly expressed if there is only a low amount of its mRNA available. The set of measurements of the amount of available mRNA of a gene in different samples (e.g., tissues or experiments) is therefore denoted as its respective expression profile.

2.1.2. Gene expression regulation

The expression of genes is regulated by genes having transcription factor activity. The activity of a transcription factor depends on its corresponding mRNA and protein level. The protein level directly depends on the mRNA level and the mRNA level is controlled by transcription-regulating proteins. The activity of transcription-regulating proteins is again regulated by other proteins. And, additionally, proteins with specific functions are needed to produce mRNA of the gene coding for these transcription factors. To summarize, different genes especially their corresponding proteins and their relationships to each other have an influence on the expression of other genes, e.g., transcription factors, and the activity of other proteins. Hence, the regulation of gene expression constitutes a network of genes.

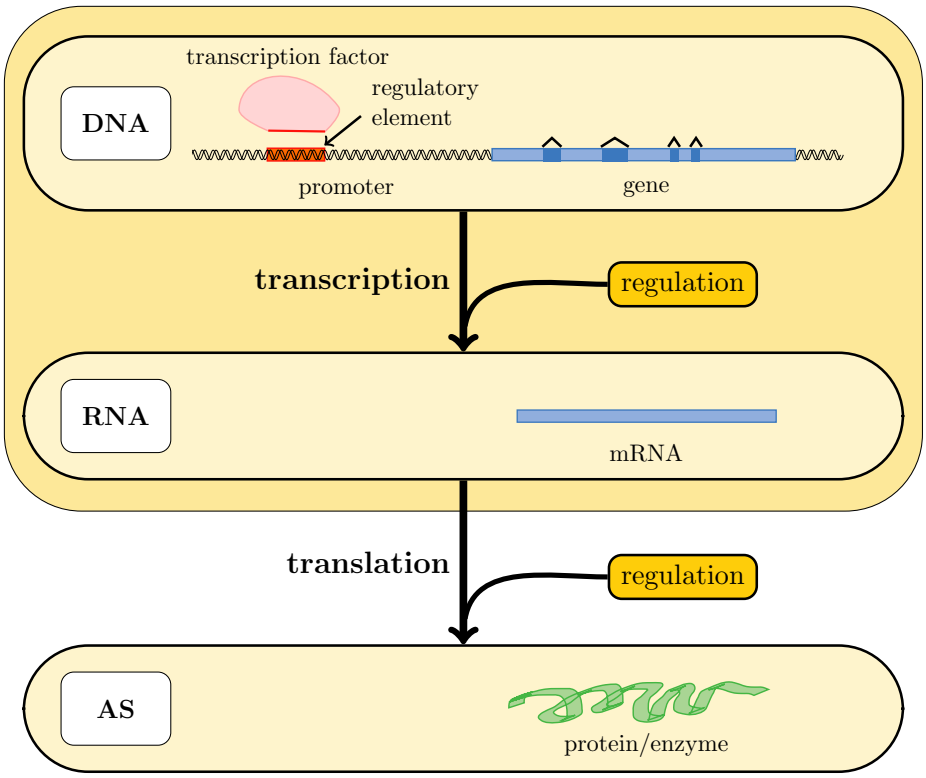


Figure 2.1.: Flowchart showing the expression of a gene. Genes are regions on the DNA that code information about proteins. The transcription of genes is regulated by transcription factors which bind to regulatory elements in the promoter regions of genes. The binding to regulatory elements affects the regulation and thus the transcription of genes. Transcription is the process by which the DNA sequence of the gene is transcribed into the RNA sequence. After several post-processing steps the transcription yields the messenger RNA (mRNA). By the translation process the mRNA is translated into a sequence of amino acids (AS) which results in mature protein after several steps of post-processing. Both on the transcriptional and on the translational level are mechanisms that regulate both processes. The transcription and the translation together with their respective post-processing steps comprise the processes of gene expression.

The expression of genes can be changed as a response to a stimulus or a signal (Figure 2.2). A stimulus can be a signal from outside or inside an organism. From the outside it can, e.g., be a change in the ambient temperature or a change in the availability of water. From the inside it can, e.g., be a change in the concentration of a hormone. A signal is recognized by a signal-specific receptor. The receptor is one of the main components of the corresponding signaling network which recognizes and processes the signal. A signaling network transduces the signal by activating or repressing other components of the signaling network which directly regulate the activity of other proteins or directly affect the transcription of genes. Each component of the signaling network has a specific function and the interplay of the different components directly determines the primary responses triggered by the signal. The primary responses lead to additional downstream responses e.g., changes in the phenotype (physiological aspects). In summary, the signal triggers a cascade of gene-regulatory events (signaling network) that lead to a signal-specific response, which might be visible at the physiological (phenotypical) level.

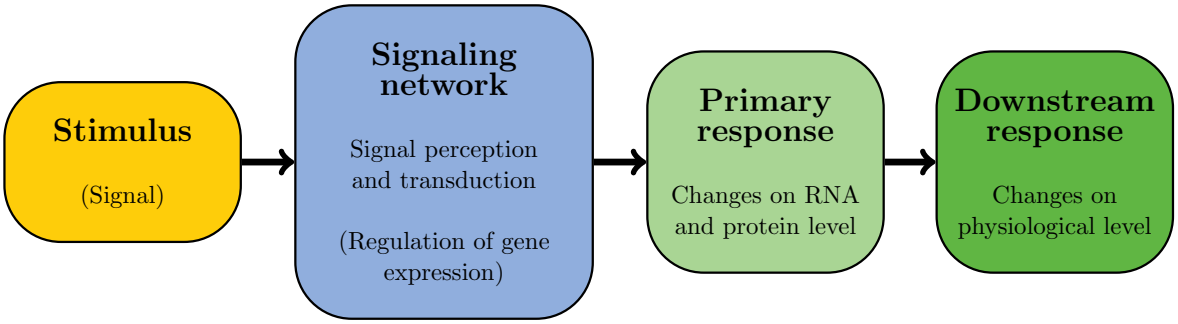


Figure 2.2.: Flowchart showing stimulus perception and transduction. The stimulus (signal) is recognized by the corresponding receptor. The receptor is one component of the respective signaling network. The signaling network plays a key role in the perception and transduction of the signal via regulation of gene expression. After signal recognition the activated receptors activate or repress other components of the signaling network that regulate the expression of primary response genes. As a consequence the signal is transduced by changing the expression of primary response genes. This causes changes in the overall RNA and protein level (Figure 2.1). This in turn causes downstream responses, which are i.e., changes on the physiological level.

2.1.3. Auxin signaling network

A very important stimulus a plant reacts to is a change in the auxin concentration in the cell. Auxin is a very powerful plant hormone that controls processes such as cell division, cell differentiation, and cell elongation: essential cellular processes necessary for plant developmental events and reactions in response to environmental challenges. At the cellular level, the auxin signal is recognized and transduced by the auxin signaling pathway (Figure 2.3). The auxin signaling pathway is a network that is formed by three main components: (i) the TRANSPORT INHIBITOR RESPONSE1/AUXIN SIGNALING F-BOX1-5 (TIR1/AFBs) auxin receptors, (ii) AUXIN/INDOLE-3-ACEDIC ACID (AUX/IAA) family of auxin co-receptors/transcriptional repressors, and (iii) the AUXIN RESPONSE FACTOR (ARF) family of transcription factors (Quint et al., 2006).

ARFs regulate the transcription of auxin-responsive genes by binding to auxin-responsive elements (AuxRE) located in their promoters (Guilfoyle et al., 1998; Ulmasov et al., 1999). The central function of the auxin signaling network is to regulate the transcription of ARF-controlled auxin-responsive genes. An AUX/IAA is bound to the ARFs as long as the auxin concentration in the cell is low. This binding prevents the ARF to act as a transcription factor and thus represses the transcription of the respective genes. An increase of auxin concentration in the cell is recognized by the auxin receptors (TIR1/AFBs), which are part of an E3-ligase complex. The TIR1/AFBs and the AUX/IAAs form co-receptor complexes and together bind auxin molecules. To form this co-receptor complex the binding of the AUX/IAAs to the ARFs is released and the AUX/IAAs are marked by the E3-ligase complex for degradation. The marked AUX/IAAs are subsequently degraded which results in a reduced AUX/IAA concentration in the cell. As a consequence of the released ARF-to-AUX/IAA binding, the transcription factor activity of the ARFs is no longer repressed and the respective auxin-responsive genes are transcribed. This set of genes contains transcription factors, enzymes and also genes of the AUX/IAA family. As long as the auxin level in the cell is high enough

auxin together with the AUX/IAAs is bound to the receptors (TIR1/AFBs). If the auxin level decreases, the newly synthesized AUX/IAAs bind to the ARFs and repress the transcription of auxin-responsive genes. The interaction of these three main components of the auxin signaling network causes an auxin-specific reaction. The three main components of receptors, co-receptors/transcriptional repressors and transcription factors are encoded by gene families of six, 29, and 23 known members, respectively (Chapman et al., 2009). This allows 4002 theoretically possible specific interaction scenarios of these three components that trigger different gene regulation events (primary responses) which result in different downstream responses (Calderón Villalobos et al., 2012; Salehin et al., 2015). Hence, the auxin signal processed by the auxin signaling network can trigger a wide variety of downstream responses (Ramos et al., 2001; Zenser et al., 2001; Guilfoyle et al., 1998; Ulmasov et al., 1999) with some of them leading to visible changes in the physiological phenotype of the plant.

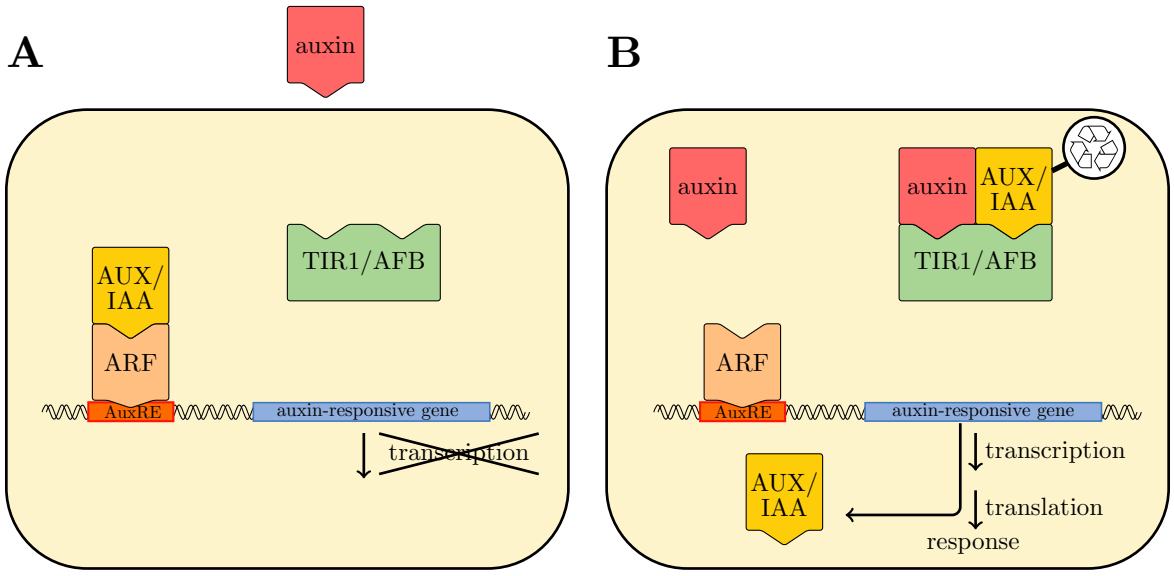


Figure 2.3.: The auxin signaling network. The auxin receptors (TIR1/AFBs), Auxin Response Factors (ARFs), and auxin co-receptors/repressors (AUX/IAAs) together form the auxin signaling network. (A) Low auxin concentration: ARFs bind to auxin responsive-elements (AuxRE) in promoters of auxin-responsive genes. In case of low auxin concentration in the cell, the AUX/IAAs repress the transcription factor activity of the ARFs by directly binding them. (B) High auxin concentration: An increase in cellular auxin levels is perceived by the auxin co-receptor complex that consists of a TIR1/AFBs and an AUX/IAA protein. The AUX/IAAs release the binding to the ARFs and bind together with the auxin to the TIR1/AFBs. Simultaneously the AUX/IAAs are tagged for degradation and their concentration in the cell is reduced. The ARFs recover their transcription factor activity and initiate downstream auxin responses. As a direct consequence, the ARFs could initiate the transcription of auxin-responsive genes like AUX/IAAs. The transcription factor activity of the ARFs is repressed again by the newly synthesized AUX/IAAs when the auxin level decreases.

2.2. Objectives and outline

Driven by one major question in auxin biology: “How can this small auxin signaling network that consists of only three main components trigger a wide variety of downstream responses?”,

we were interested in developing biological and bioinformatics methods to study the reactions of the model plant species *Arabidopsis thaliana* on application of an auxin stimulus. *A. thaliana* as a model organism is well established and easy to cultivate and to handle. Additionally, it is completely sequenced and well annotated.
We analyzed the expression levels or changes of the expression levels of genes of *A. thaliana* exposed to an auxin stimulus to get insights into regulatory relationships and interactions between genes that are involved in the auxin signaling network and genes that show primary or downstream responses.

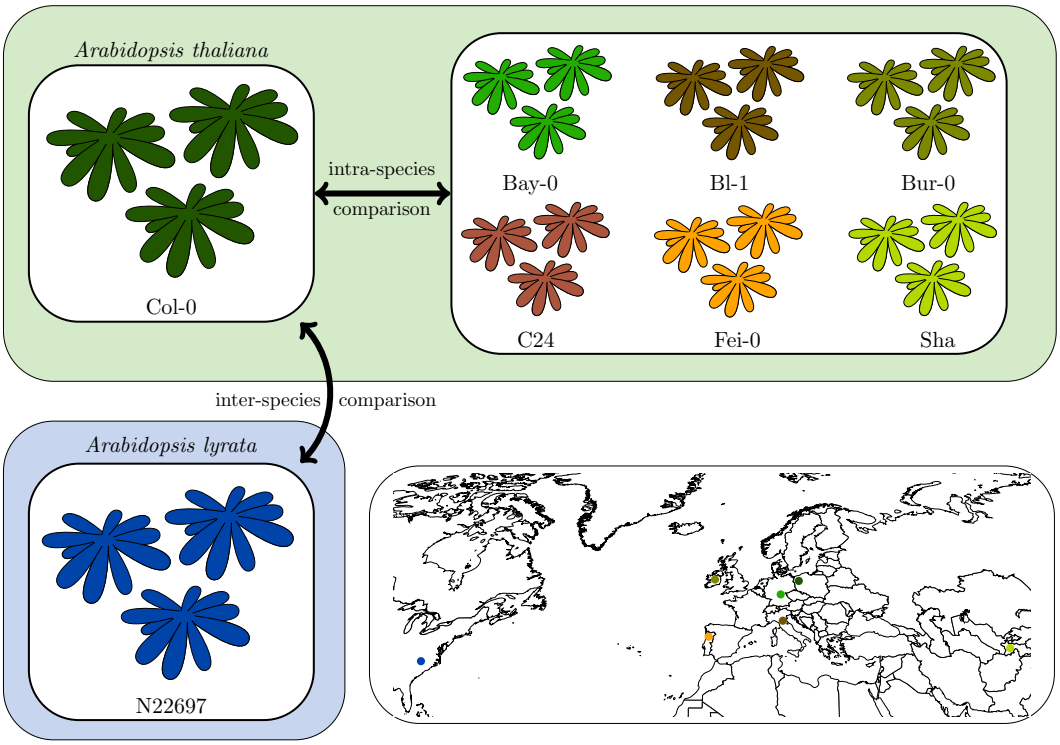


Figure 2.4.: Comparisons performed with *A. thaliana* and *A. lyrata* and their distribution over the world. The green and the blue rectangle contain representative ecotypes of *A. thaliana* and *A. lyrata*. Each ecotype is shown in three copies. Hence, each ecotype is analyzed by its three biological replicates. We performed an intra-species comparison by comparing ecotypes of *A. thaliana* plants; we compared the reference ecotype Col-0 to six other *A. thaliana* ecotypes. We additionally performed an inter-species comparison by comparing *A. thaliana* Col-0 to *A. lyrata ssp. lyrata* N22697. The map shows the distribution of the analyzed *A. thaliana* ecotypes and *A. lyrata ssp. lyrata* over the planet.

We considered different types of analyses (Figure 2.4). First, we analyzed plants of the well studied *A. thaliana* reference ecotype Col-0. The analysis of gene expression levels of a single Col-0 plant provides a snapshot of the reactions, changes in the genes expression levels, and gene-to-gene interactions. To get information of the variation and reliability of the observed gene interactions and thus relationships, we took multiple Col-0 plants with an identical genetic background into account. Although these plants originate from the same seeds and were exposed to the same conditions, they will react as individuals and will possibly show differences in their reactions. Second, we extended this kind of analysis to six other ecotypes that are available for *A. thaliana*. The reference ecotype and the other ecotypes are very similar

in their genomic sequences but originate from different geographic locations with different environmental factors. Analyzing these ecotypes provides information on how strong slight differences in the genomic sequences and the adaption to different environmental factors affect the reaction to an auxin stimulus. We did not only analyze the reference ecotype and ecotypes separately, but also compared them based on gene expression levels to find similarities in gene expression among the ecotypes and also differences which might be present due to the former adaption processes to different environmental factors. This intra-species comparison gives deep insight into the naturally occurring variation in response to an auxin stimulus. Third, for our analyses we did not only take one species but also a second species into account *Arabidopsis lyrata*. *A. lyrata* is a close relative of *A. thaliana* (Hu et al., 2011). Both species diverged 5 Mio. years ago and show more genetic variation compared to the variation of the ecotypes. Again, we considered to analyze the representatives of both species first separately and second by comparing them. The inter-species comparison allows us to identify genes that show similar auxin responses and are therefore either essential for auxin response or are conserved primary or downstream responses.

To perform the considered analyses of the gene expression levels of several *Arabidopsis* plants exposed to auxin, we selected ecotypes based on their physiological response to an auxin stimulus. Root growth is known to be affected by auxin, therefore we selected ecotypes that cover a wide range of different auxin-related root growth responses.

We already have published or will publish the performed analyses. We visualize their relationships which define the outline of this thesis in Figure 2.5. A detailed analysis of auxin responses within and between species, which includes transcriptomic, genomic, and physiological data was not performed before. The performed analyses and publications highly depend on bioinformatics knowledge and algorithms. Algorithms are needed to integrate these three levels of data for analysis and, e.g., to inspect if the expression response of a gene is related to its promoter sequence. We will present algorithms and measures to fulfill this task.
Whereas sequence for both the model species *A. thaliana* and the non-model species *A. lyrata* were available, the computation of reliable microarray expression values for *A. lyrata* is still an open task. The problem results from the fact that there is no microarray available for *A. lyrata* and using the microarray designed for *A. thaliana* causes problems. Indeed there are algorithms available to solve these problems by probe masking, but neither the number of remaining genes nor the quality of the expression values are satisfying (Khaitovich et al., 2004; Broadley et al., 2008; Graham et al., 2007; Hammond et al., 2005; Poeschl et al., 2013). We will present an algorithm that fills this gap and yields a satisfying number of genes and additionally reliable expression values.
Reliable expression values are the basis for further analyses like comparing expression profiles using clustering algorithms or inferring gene-to-gene relationships from co-expression networks. For clustering genes using hierarchical clustering algorithms various distance measures are available (Yona et al., 2006), but none of these addresses that the clustering might be biased by noise. Whereas hierarchical clustering algorithms allow for studying co-expressions of genes on a global level, algorithms, like the Local Context Finder (LCF, Katagiri et al., 2003), are available to perform a local and more detailed analysis of gene co-expression and thus potential regulatory relationships. The LCF is capable of inferring gene-to-gene relationships from gene expression profiles that are due to positive regulation events, but neglects existing

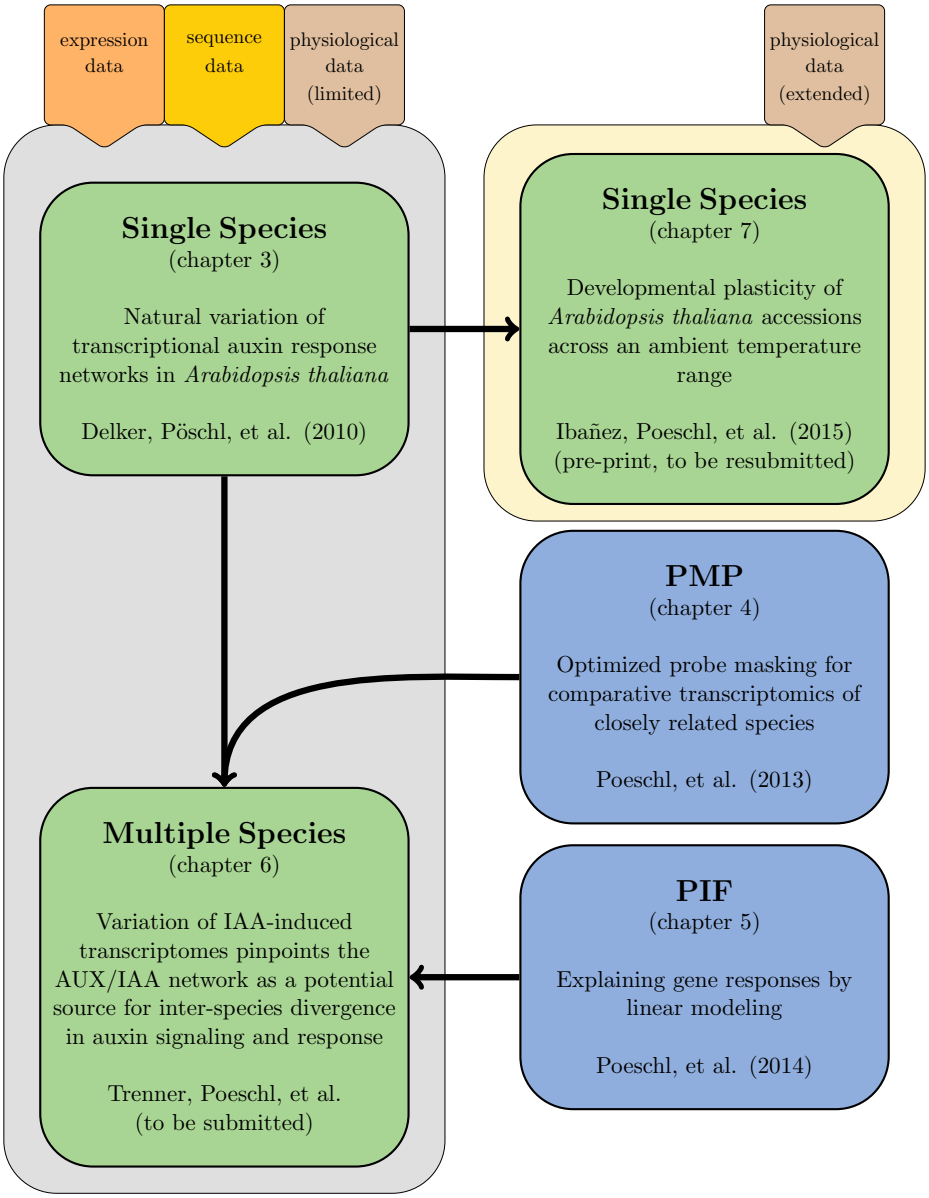


Figure 2.5.: Flowchart showing the relationship of the presented publications. Boxes either in green or blue contain information about the respective publication and chapter it is presented in. The green color shows publications that have their main focus on biology and are attempted for readers with biological background, but highly depend on bioinformatics knowledge. The blue color shows publications that have their main focus on bioinformatics, but depend on the biological input data/biological question. The small puzzle-like pieces show the type of data that is analyzed in the respective publications. Publications are linked by black arrows to show their dependencies.

negative regulation events. We will present an algorithm that can handle also negative regulation events. Gene-to-gene relationships are transferred into a network for a more intelligible representation. We will present a measure to compare two networks based on their topology.

In "Natural variation of transcriptional auxin response networks in *Arabidopsis thaliana*" (Delker et al., 2010), the first mainly biology-focused work, we performed intra-species comparisons of *A. thaliana* representatives to get a basic understanding on how the components

of auxin signaling network interact. We included expression, sequence, and physiological data to analyse how the auxin treatment affects the interaction of these components and and the remaining genes. We describe the bioinformatics algorithms used and analyses performed in more detail in chapter 3 and give a short introduction in section 2.2.1.

To enable the inter-species comparison of *A. thaliana* and *A. lyrata* representatives by integrative analyses of expression and sequence data, we developed and published two new bioinformatics algorithms, the PMP and the PIF. We developed the "Probe Masking Pipeline" (PMP) to address and overcome the problem of computing reliable expression values for a sufficient number of genes of the non-model species *A. lyrata*. We published the PMP in "Optimized probe masking for comparative transcriptomics of closely related species" (Poeschl et al., 2013). We give a short introduction into this publication in section 2.2.2 and present the full article in chapter 4. For a more comprehensive analysis of gene expression profiles, which also includes negative regulation events besides positive regulation events, we introduced the "Profile Interaction Finder" (PIF) in "Explaining gene responses by linear modeling" (Poeschl et al., 2014). We give a short introduction into this publication in section 2.2.3 and present the full article in chapter 5.
To make the inter-species comparison more accurate we additionally introduce two measures for quantifying the diversity of expression and promoter sequences of genes in both species. We will publish the inter-species comparison together with selected and new introduced bioinformatics algorithms in "Variation of IAA-induced transcriptomes pinpoints the AUX/IAA network as a potential source for inter-species divergence in auxin signaling and response" (Trenner et al., in prep.). We give a short introduction into this work in section 2.2.4 and present the full article in chapter 6.

Previous analyses (Balasubramanian et al., 2006; Delker et al., 2010) proved that ecotypes of *A. thaliana* show variations in, e.g., root growth, hypocotyl elongation or flowering time in response to auxin treatment. In the last work (Ibañez et al., 2015) presented in this thesis, we address the question if this natural variation can also be observed in the development of other traits of *A. thaliana*. We inspected the physiological responses of ten *A. thaliana* ecotypes exposed to different ambient temperatures. To perform the analyses, we measured 34 traits including hypocotyl length and flowering time. We addressed the question of natural variation by inspecting how strong the slight differences in the genomic sequence affect the temperature-related response (observable in the traits) of the individual ecotypes. This study provides a deeper insight into which phenotypes are affected at different ambient temperatures, which phenotypes show the same temperature-related differences for all ecotypes and which phenotypes show temperature-related differences in a subset of ecotypes. The last group might be determined by the genome of the ecotypes and thus be worthwhile candidates for existing natural variation. We addressed the task on quantifying the variance of phenotype expression due to a change in ambient temperature and proposed a measure that fulfills this task in "Developmental plasticity of *Arabidopsis thaliana* accessions across an ambient temperature range" (Ibañez et al., 2015), which is a pre-print that will be re-submitted soon. We give a short introduction into this work in section 2.2.5 and present the full article in chapter 7.

The reader will be introduced into the publications comprising this work (Figure 2.5) by the following subsections giving a more detailed overview on the objectives and methods addressed

in these works. The full articles describing the complete work are presented in chapters 3 to 7.

2.2.1. Natural variation of transcriptional auxin response networks in Arabidopsis thaliana

The main objective of this first biology-focused work (Delker et al., 2010) is to determine whether natural intra-species variation of physiological and molecular auxin responses occurs in *A. thaliana*. Furthermore, we intend to analyze at which molecular levels within the hierarchical signaling network variation can occur and which signaling components might contribute to natural variation visible on the physiological level.
These analyses are supposed to provide an overall view and a basic understanding of auxin responses, especially of the components of the auxin signaling network. This knowledge will be the basis for further studies on natural variation of auxin responses.

The question of potential natural variation is initially addressed by classic physiological auxin response assays which are followed by extensive transcriptional profiling of auxin-induced changes of transcriptomes in different ecotypes of *A. thaliana* at different time points (control and 0.5, 1 and 3 h post induction) in three biological replicates each.

In the following we will outline the bioinformatics methods that were used to address these objectives in this intra-species comparison.

Bioinformatics methods

To answer the main question of whether natural intra-species variation of physiological and molecular auxin responses occurs in *A. thaliana*, we decided to cluster on the one hand ecotypes and on the other hand genes based on their auxin response.

By literature research we found an bioinformatics algorithm proposed as the Local Context Finder (LCF) by Katagiri et al. (2003) that fulfills our needs and seems promising in assisting to answer our questions. The Local Context Finder (LCF) algorithm is generally used to generate co-expression networks. In contrast to other co-expression algorithms where the co-expression of ecotypes or genes is studied on a global level using conventional clustering methods like HCLUST (Murtagh et al., 2011) or HOPACH (Laan et al., 2003), the LCF algorithm performs a local, more precise analysis of potential ecotype or gene regulation relationships. An important advantage of the LCF algorithm is the translation of multidimensional relationships between expression profiles into a network that makes complex interactions more intelligible. In these networks ecotypes or genes are the nodes and edges represent mathematical relations between nodes. Whenever two nodes are connected in a network we can hypothesize that there might be some biological reason or process which relates these two nodes to each other.

To reduce the effect of possible noise and to filter for robust co-expressions relations of ecotypes or genes, we implemented the LCF algorithm and extended the LCF algorithm by the suggested sampling-with-replacement (bootstrapping) step (Katagiri et al., 2003). We additionally linked

the LCF algorithm to the scriptable network visualization program Graphviz (Gansner et al., 2000) to allow a co-expression analysis and visualization in a high-throughput manner.

The comparison of gene expression profiles is valuable to detect similarities and differences between the ecotypes but does not consider the actual level of gene expression. Hence, we needed a second measure to assess quantitative differences of gene expression levels among ecotypes. To asses and detect differentially expressed genes we used well established statistical testing procedures, ANOVA for two-way testing and a Student's t test for small sample sizes (Opgen-Rhein et al., 2007) for one-way testing.

To answer the second, more specialized question on which molecular levels within the hierarchical signaling network variation can occur and which signaling components might contribute to variability on the physiological level, we focused on a subset of genes coding the components of the auxin signaling network (section 2.1.3). To analyse the expression profiles of the selected genes by means of co-expression networks, we applied the LCF algorithm. We studied the resulting networks, where the nodes represent the genes and the edges represent the inferred interactions, and compared their topology among the ecotypes. To analyse the gene interactions we introduced an hypergeometric test to asses how likely the number of common edges occurs by chance. For analysis of the gene responses we additionally introduced a modified Student's t test to identify differently responding genes of two ecotypes.

Results, discussion, and conclusions

We could answer the main research questions in a combination of applying existing and established algorithms and measures, and of applying modified or extended versions of existing algorithms and measures. From applying the LCF algorithm on the ecotypes, we found that the ecotypes form subgroups, where different subgroups show different behaviors on the transcriptional level. This might indicate that there is intra-species natural variation which occur due to differences at the transcriptional level. We additionally found by applying the LCF algorithm, and known and newly introduced statistical testing procedures that transcriptional differences already occur in the auxin signaling network which is the beginning of auxin response. Hence, we proposed that due to differences in the expression of genes contained in the auxin signaling network, the auxin signal transmission differs between ecotypes causing clearly distinguishable physiological phenotypes. With these findings we proposed a model showing that the expression levels of the auxin co-receptors/transcription repressors (AUX/IAAs) and transcription factors (ARFs), and consequently their interaction, affect the regulation of the transcription of downstream genes that cause physiological responses.

2.2.2. Optimized probe masking for comparative transcriptomics of closely related species

The key question of this part of the project (Poeschl et al., 2013) was, "How to compare gene expression values of different species when a microarray is available only for one species?".

We provided a solution and demonstrated its utility for the well known model plant *A. thaliana* and its closely related sister species *Arabidopsis lyrata* (both treated with auxin, and samples taken at three time points 0 h, 1h and 3h in three replicates each). *A. thaliana* and *A. lyrata* both diverged about 5 Mio. years ago. While still closely related, *A. thaliana* and *A. lyrata* show considerable differences in numerous physiological and morphological traits. Furthermore, the genome size of *A. lyrata* is considerably larger but the genomes still show a high level synteny (i.e., co-localization of genes). Using sequence information we determined orthologous gene pairs between both species which are the basis of the proposed algorithm. Orthologous genes are genes in different species that originated from a common gene in their last common ancestor. Orthologs often, but not always, have the same function (Fang et al., 2010).

The cheapest way to analyze samples taken from a non-model species is not to design a new microarray but to use an existing microarray of a closely related (model) species and to perform hybridization of control and auxin treated samples from both species on the same microarray architecture. In case of the non-model species *A. lyrata* this is the ATH1 microarray from Affymetrix (Redman et al., 2004) specifically designed for the model species *A. thaliana*. This microarray contains probe sets of small oligonucleotide sequences that specifically target the transcript of a unique gene or the transcripts of a gene family of *A. thaliana*. But species-specific differences in the sequences of the genes or more precisely in the transcripts of genes can cause problems, such as the following: (i) lower hybridization accuracy of probes due to mismatches or deletions, (ii) probes binding multiple transcripts of different genes, and (iii) probes binding transcripts of non-orthologous genes. All three aspects can have considerable impact on the accuracy of transcript level detection and need to be addressed in cross-species microarray analyses.

Bioinformatics methods

The key question of this work evolved into a more specific question of how to allow for the direct comparison of expression values of genes from closely related species measured on the same microarray. There are bioinformatics algorithms available that compute expression values for the mRNAs of genes of non-model species measured on microarrays that are not designed for them. However they mostly concentrate on the problem of lower hybridization accuracy and neglect the other two aspects mentioned before. We were faced with the challenge to develop a bioinformatics algorithm that addresses all three problems and yields reliable gene expression values.

One of the available algorithms is a sequence-based approach proposed by Khaitovich et al. (2004). This algorithm uses three sets of sequences, the sequences of the microarray probes, the sequences of the transcripts of *A. thaliana* and the sequences of the transcripts of *A. lyrata* to determine which probe likely binds to which transcripts. Inspired by this sequence-based approach, we based our new probe masking algorithm, the probe masking pipeline (PMP), on sequences of probes and transcripts, too. But we solved the task of determining which probe binds to which transcript in a different way.
Khaitovich et al. (2004) went for comparing the sequences of the transcripts of two species

first, to determine orthologous genes and to identify and keep identical regions. Subsequently, Khaitovich et al. (2004) determined the probe-to-ortholog sequence relation by comparing the probe sequences and the kept identical regions. In contrast, we decided to first compare the sequences of probes and transcripts to use as many sequence information as possible and post-process the results in the PMP.

We designed the PMP in a modular fashion that allows us to specifically address and solve all three mentioned problems. First, we aligned the sequences of the probes to the sequences of the transcripts of both species allowing at most one mismatch. Second, we removed probes that do not show any similarity to a transcript. We processed the remaining probes that show high similarity to at least one transcript according to a decision tree presented in Poeschl et al. (2013) to determine if they provide reliable or unreliable hybridization intensities. Finally, the PMP retained only probes that are orthologous gene pair-specific and can be used for the comparative gene expression analysis. The mismatch that we allowed in the comparison of the probes and the transcripts could cause probes to show an artificially decreased hybridization intensity, because the hybridization was not perfect. This causes no problems if fold changes are used for comparing genes by their responses. But problems arise, if actual expression values are used in the comparison. Therefore, we proposed a correction of the hybridization intensities on the probe level based on a fit of a fourth-degree polynomial. We included the correction of the intensities of the probes as an additional step in the RMA-normalization procedure (Irizarry et al., 2003). The correction was necessary for direct comparison of the expression values of *A. thaliana* and *A. lyrata* in chapters 5 and 6.

We compared our algorithm with the sequence-based approach proposed by Khaitovich et al. (2004) and a genomic DNA hybridization-based approach proposed by Hammond et al. (2005). The sequence-based approach addresses the first and the last problem and has very stringed settings for the sequence comparisons. The hybridization-based approach addresses only the first problem. It requires the user to set a hybridization intensity threshold. Intensity values below this threshold are discarded.

We were also faced with the challenge to validate and to compare the output of the three algorithms. We compared the resulting number of genes and the expression responses of 40 randomly selected genes. We also compared the computed expression responses with independent wet-lab (RT-qPCR) produced expression values to assess the validity of the computed microarray expression values.

Results, discussion, and conclusions

By comparing our algorithm with the two previously published algorithms, we found that both sequence-based algorithms yield fewer genes than the hybridization-based algorithm. Our sequence-based algorithm including the relaxed sequence comparison results in significantly more genes retained for the analysis than the sequence-based algorithm by Khaitovich et al. (2004). We could also show that both sequence-based algorithms yield comparable and more reliable expression response values than the hybridization-based algorithm. Our new algorithm yields as many genes as possible that also have reliable expression responses. By

using this algorithm for probe masking and additional probe intensity normalization, comparative transcriptomics of two or more closely related species via classic microarray approaches becomes feasible.

2.2.3. Explaining gene responses by linear modeling

Co-expression on the simplest level addresses genes that show the same expression response over time or to treatment. The expression of these genes is putatively triggered by the same biological process or stimulus and can indicate a function of genes in the same signaling or response pathway. The relationship of co-expressed genes can be studied on a global level using conventional clustering methods like HCLUST (Murtagh et al., 2011) or HOPACH (Laan et al., 2003). But for a more precise analysis of potential gene regulation relationships, a study on the local level is needed as provided by the Local Context Finder (LCF) algorithm proposed by Katagiri et al. (2003).

In more detail, the LCF algorithm reconstructs the high dimensional expression profile of a gene as a linear combination of the expression profiles of other genes. These relations can be translated into graphical representations, where the nodes represent the genes and the edges represent the mathematically inferred relations. In a network representation, genes that contribute to the reconstruction of a specific gene would have a directed edge pointing to the specific gene. Genes that are connected in a network have similar expression profiles and therefore show similar expression responses. The biological assumption is that genes showing similar expression profiles and thus responses, are either regulated by the same regulatory acting gene or regulate each other.

Bioinformatics methods

For a more comprehensive analysis of gene expression responses we wanted to include the knowledge that gene regulation networks often function in both up- and down-regulation to initiate response, which the LCF cannot do.
We proposed a new bioinformatics algorithm, the Profile Interaction Finder (PIF, Poeschl et al. (2014)) that now incorporates both directions of gene responses. We based the reconstruction of a gene expression profile on the same mathematical model using linear combinations as proposed by Katagiri et al. (2003). In more detail, we used a linear model and incorporated the constraints that the weights have to be positive and have to sum up to one. To model the possible opposite direction of responses, we extended the model by an additional set of parameters directly coupled to the weights. This extended linear model is still a convex linear combination which can be solved analytically.

In contrast to the LCF algorithm, the PIF algorithm comes in two variants.
We make use of the assumption that genes that are closely connected in biological pathways, and thus have a biological relationship, will also tend to have similar expression patterns in the first variant of PIF algorithm. We therefore feed the PIF algorithm with the expression profiles of all genes to compute gene-to-gene co-expression networks that reflect this assumption. In the gene-to-gene co-expression networks, gene expression profiles are reconstructed

using the expression profiles of other genes. These reconstructed networks consist of genes and edges connecting genes that show a response either in the same or in the opposite direction. These networks could serve as starting point for elucidating possible functions of unknown genes by incorporating their (co-expression) network relation to known and thus annotated genes. The networks could also give hints to possible regulatory relations between connected and not connected genes.
We were also faced with the challenge to identify genes that respond due to a specific experimental condition, which in other words means, genes that show a very condition-, treatment-, or stimulus-specific expression profile. Therefore, in the second variant, the input of the PIF algorithm comprises not only the expression profiles of the genes but also pre-defined, artificially created, condition-specific prototype profiles. In this variant the gene expression profiles are reconstructed from these condition-specific prototype profiles. Hence, we used the PIF algorithm to generate gene-to-treatment networks that represent gene-to-treatment relationships. Based on the inferred treatment relationships from the gene-to-treatment networks we assigned genes to clusters.

Results, discussion, and conclusions

We showed that the PIF algorithm is capable of producing biologically relevant results when applied to reconstructing gene-to-gene networks and clustering genes according to their response to experimental conditions. We applied the PIF algorithm, in both variants, to the *A. thaliana* and *A. lyrata* data set described in section 2.2.2. By applying the PIF algorithm with very stringent parameters to reduce the inference of false positive relations in the first variant, we generated gene-to-gene co-expression networks of all genes in the data set. For 15 % of the genes we found strong evidence for possible regulatory connections to other genes. For the *A. thaliana* and *A. lyrata* data set, we found that 36 % of these genes are "regulated" by genes showing an opposite expression response. We would have missed these relations when applying only the LCF algorithm instead of the PIF algorithm to this data set. From a biological point of view, we identified a reasonable number of genes that are potentially up or down regulated by the presence or absence of other genes or their gene products.

For application of the second variant of the PIF algorithm we created prototype profiles according to the time point of post treatment with auxin. Using these time point-dependent prototype profiles we were able to cluster genes according to the time they needed for their response to auxin. Response can result in increased or reduced gene expression. Besides identifying genes that showed the same direction as the prototype profiles, we additionally identified relations of gene expression and prototype profiles showing opposite directions. We found a reasonable number of genes that are only or additionally down regulated at a specific time point.

We additionally demonstrated the applicability and the utility of the PIF algorithm on a second data, which is a synthesis data set comprising samples of different tissues of *Apis mellifera* treated with different pathogens (*The Trans-Bee workshop* 2014).

Hence, we concluded that the PIF algorithm in its two variants is applicable for a more comprehensive and complex analysis of gene-to-gene and gene-to-treatment relationships, and produces biologically relevant results.

2.2.4. Variation of IAA-induced transcriptomes pinpoints the AUX/IAA network as a potential source for inter-species divergence in auxin signaling and response

A key question in auxin biology is still: "Do auxin signaling and response contribute to adaptive processes to local environmental changes?". By studying the auxin gene responses in ecotypes of the model plant *A. thaliana*, Delker et al. (2010) detected remarkably high variation in the auxin response among the *A. thaliana* ecotypes for early auxin signaling components. These findings lead to a model which illustrates the hypothesis that the expression levels of the auxin co-receptors/transcription repressors (AUX/IAAs) and auxin response factors (ARFs) and consequently their interactions contribute to variations observed on the gene expression level and on the physiological level, e.g., reduced root growth.
In this work (Trenner et al., in prep.), we went from the ecotype level to the genus level and compared auxin responses of the closely related sister species *A. thaliana* and *A. lyrata*. We combined physiological, transcriptomic and genomic information to inspect variations of auxin responses in both species. The increased genetic variation between the two *Arabidopsis* species allowed (i) the identification of genes with different or similar auxin response in both species. Genes with a similar response in both species might constitute essential or conserved auxin response genes. We furthermore aimed (ii) at exploiting the genetic variation in the promoter sequences to identify regulatory elements that might contribute to similar or differential auxin responses. And finally (iii) we aimed at testing whether the previously proposed model which identifies the expression level of the early auxin signaling components as the source of downstream variation could be verified on the species level that has higher genetic variation.

We addressed and assessed these tasks on the *A. thaliana* and *A. lyrata* expression data set already used in section 2.2.2.

The basis to face these three tasks are reliable expression values not only for the model species *A. thaliana* but also for the non-model species *A. lyrata*. We addressed and solved the problem of getting reliable expression values particularly for the non-model species *A. lyrata* in Poeschl et al. (2013) presented in section 2.2.2 and chapter 4.

Bioinformatics methods

To study the patterns of expression response of genes on a global level and to identify genes with different or similar auxin responses (i) we choose HCLUST (Murtagh et al., 2011) a hierarchical clustering approach. To cluster together genes whose expression values change across treatments/over time in a similar fashion, the Pearson correlation coefficient is a valid measure (Yona et al., 2006). However, to avoid the clustering of the expression profiles to be biased by noise, we proposed a modified version of the Pearson correlation coefficient. For

this modified correlation coefficient we proposed to compute the co-variances of the replicate means of each time point and to use all individual measurements to compute the variances.

The expression of a gene is substantially influenced by the binding of transcription factors to regulatory elements in the promoters of genes. Alterations in the promoter region of a gene especially at regulatory sites, might affect the transcription of the respective gene. Changes occurring at regulatory sites can alter the transcriptional process by preventing or altering binding of transcription factors and thus affect transcription.
To identify potential sources for the distinct transcriptional behavior and to exploit the genetic variation in the promoter sequences (ii) we analyzed the presence of known regulatory elements within the promoter regions of the genes. To this end, we extracted the motif sequences of auxin related elements from a database (Yilmaz et al., 2011) and literature. Motifs can have unspecific positions, where various nucleotides are tolerated for binding, thus we represented the sequences of the motifs by regular expressions to perform inexact pattern matching on the promoter sequence. We found no clear pattern of motif occurrence that could explain the auxin-related expression responses of the genes in the clusters resulting from the HCLUST approach. To complete the analysis on regulatory elements we applied a statistical model-based *de-novo* motif discovery approach (Grau et al., 2013).
To exploit the genetic variation in the promoter sequences (ii) on a more general level and to assess the relationship between the diversity of the promoter sequence and the diversity in expression response of orthologous genes we introduced measures for both diversities. To measure the diversity of two promoter sequences of an orthologous gene pair we selected a k-mer-based correlation coefficient proposed by Vinga et al. (2003). To asses the diversity on the gene expression level we used the newly introduced modified Pearson correlation coefficient.

To inspect if early auxin signaling components could be the source of downstream variation (iii), we again used the modified Pearson correlation coefficient to cluster auxin co-receptors/transcription repressors (AUX/IAAs) genes. The AUX/IAA genes are classical and conserved auxin response genes (Paponov et al., 2008) and were identified in Delker et al. (2010) to be highly variable in their gene expression and thus be a potential source for downstream variations on the expression and physiological level.
To identify genes with expression profiles that are either positively or negatively correlated to individual AUX/IAA gene clusters, we used the Profile Interaction Finder (PIF) algorithm which we introduced in section 2.2.3 and present in chapter 5.

Results, discussion, and conclusions

The global clustering approach revealed groups of genes with similar and distinct expression response patterns. Among the group with similar and thus conserved auxin response we could identify members of prominent auxin-response gene families and also several genes with so far unknown function.

The analysis of the presence of known auxin-related regulatory elements in the promoters of the genes revealed no clear pattern that could explain the expression responses of the genes in the specific clusters. This indicates either a highly complex regulation involving multiple regulatory elements and/or the presence of additional, so far, unidentified regulatory elements.

From *de-novo* motif discovery we identified variants of known regulatory elements that are annotated to be auxin-related, which might indicate a high variability in certain nucleotides putatively accounting for differential binding affinities of distinct auxin-related transcription factors (ARF) (Boer et al., 2014). While we identified modified motifs of known regulatory elements, we also discovered potentially new motifs that might contribute to auxin signaling. Hence, these newly discovered motifs need to be validated experimentally to verify their roles and effects in auxin signaling transduction.

By addressing question (iii) and clustering the expression responses of the AUX/IAA gene family we identified three main types of clusters: clusters containing AUX/IAA genes specifically induced in *A. lyrata*, clusters containing genes primarily induced in *A. thaliana*, and one large cluster containing AUX/IAA genes showing conserved and significant induction in both species. By inspecting the expression profiles of the AUX/IAA genes and the remaining genes using the PIF algorithm, we identified known auxin-related genes that show the same classical auxin response profile like the conserved AUX/IAA genes in the large cluster. This group of genes seems to be part of a conserved auxin response in both species. AUX/IAA gene clusters with species-specific gene induction showed correlations to auxin-relevant genes involved in biosynthesis, signaling, transport, and response. Positive and negative correlations of downstream-responding genes to AUX/IAA genes indicate that variations in the beginning of auxin signaling may reach downstream genes and thus may contribute to differences observed at the physiological level.

Hence, our analyses, which integrated information from genomic, transcriptomic and physiological level, identified new possibly auxin response-related regulatory elements and the gene families of the auxin signaling network as potential source for adaptation.

2.2.5. Developmental plasticity of Arabidopsis thaliana accessions across an ambient temperature range

Changes in ambient temperature can affect plant growth and development, and flowering processes (CaraDonna et al., 2014; Fitter et al., 2002). Hence, it becomes increasingly important to get a deeper understanding of developmental temperature responses.

Most of our present knowledge about molecular responses to ambient temperature signaling has been gained from studies in *A. thaliana*. Model temperature-related phenotypes such as hypocotyl elongation (Gray et al., 1998) and alterations in flowering time have been studied to identify components of the molecular signal transduction that are involved in triggering temperature-related responses. However, considerable naturally occurring variation in temperature-related traits like hypocotyl elongation and flowering time has been demonstrated (Balasubramanian et al., 2006; Delker et al., 2010). Higher temperatures result in higher levels of endogenous auxin that mediate parts of the temperature-related response e.g. dramatic hypocotyl elongation (Franklin et al., 2011). Natural variation in temperature-related traits might be due to local adaptation processes of diverse *A. thaliana* ecotypes to their environments and indicates a high degree of freedom in the development of traits.

Future challenges in sciences and plant breeding will, however, require a systematic assessment of temperature-related variations of developmental phenotypes across a complete life cycle. As such, on the biological side, we aim at (i) identifying phenotypes that are sensitive to ambient temperature changes. We also aim at investigating (ii) which phenotypes are robustly affected by temperature within all ecoytpes and (iii) which phenotypes are affected in only one or few ecotypes. Changes in robustly affected phenotypes in all ecotypes could be mainly due to general thermodynamic effects, but changes of phenotypes in few ecotypes could be due to natural variation in temperature-related responses. These natural variations might be consequences of adaptation processes of the affected ecotypes to cope with local climate or general environmental conditions.

In this work (Ibañez et al., 2015), we addressed these questions by profiling 34 developmental and morphological-associated traits (phenotypes) in the vegetative and reproductive growth phases of ten *A. thaliana* ecotypes which were grown at 16, 20, 24, and 28 °C.

Bioinformatics methods

To address the three main biological questions and to perform a systematic assessment of the variations of the 34 phenotypes measured for ten *A. thaliana* ecotypes at four different temperatures we focused on the application of descriptive statistical methods.
The main challenge was to select or extend descriptive statistical methods that on the one hand can be used to address the three questions and on the other hand are intuitively and easy to interpret and to visualize. Therefore, we conducted linear regression analyses to address question (i). We fitted linear models to measure the trend of phenotype variation of each ecotype across the four ambient temperatures. We used the slope, provided by the fitted linear model, to analyse the direction and strength of the phenotype variation. As an example, considering the number of days a plant needs to start flowering, a change in ambient temperature could cause a prematurely (positive slope) or delayed (negative slope) flowering of the plants. Both parameters of the linear model, the intercept and the slope, are perfectly suited for visualization and to give an impression of the trend of change in a phenotype across the four temperatures.
To identify phenotypes that show significant variation in their response to temperature changes, we additionally assessed the variances using one-way ANOVAs. However, a change of a phenotype can be due to the genome of an ecotype (genotype), or the temperature, or a mixture of both. In order to quantify the distinct influences of genotype and temperature on a given phenotype and to answer questions (ii) and (iii), we determined modified version of the intraclass correlation coefficients λ_{gen} and λ_{temp} using squared differences similar to the ANOVA framework used in Donner et al. (1980). For instance, to compute λ_{gen} for a given phenotype and temperature, we estimated the total variance of the measurements for all ten ecotypes. We decomposed this total variance into two fractions, the between variance which measures the variance between the ten ecotypes, and the within variance which measures the variance within the ten ecotypes. To yield λ_{gen}, we computed the ratio of the between variance and the total variance. Hence, λ_{gen} measures the proportion of the between variance contained in the total variance. Therefore, λ_{gen} ranges from 0 to 1. While a λ_{gen} value $= 1$ indicates a strong genotype effect on the observed variability of the phenotype, no effect of natural variation on

the phenotypic differences can be assumed for $\lambda_{\text{gen}} = 0$. The same scheme holds for λ_{temp}, which measures the effect of temperature for a given phenotype and a given ecotype. Both measures have the advantage that they are in the range of 0 to 1 and that a pair of λ_{temp} and λ_{gen} is related to one phenotype. We made use of this advantage by visualizing both measures together in a 2D-scatter plot to inspect and identify possible relationships of genotype and temperature effect.

Results, discussion, and conclusions

The selection of descriptive statistical methods that are intuitive to interpret and easy to visualize, made the interpretation and the inspection of their outcomes straightforward. By inspecting slope values and λ_{gen} and λ_{temp} values, we found temperature-related variations for almost all of the 34 phenotypes. This shows the fundamental impact of ambient temperature on plant physiology. In more detail, for phenotypes measuring the leaf production phase we found a high temperature but small genotype effect. This could indicate either a highly conserved regulation within *A. thaliana* ecotypes or a regulation due to general thermodynamic effects on metabolic rates and enzyme activities. In contrast, we found a higher effect of the genotype than of the temperature for phenotypes measuring senescence. For phenotypes representing the reproductive phase such as flowering time we found both a high genotype and a high temperature effect. Phenotypes that show a high degree of genotype and temperature effects might rather be influenced by one or more specific genes that contribute to trait expression in a quantitative manner. It has been shown by Koini et al. (2009) and Kumar et al. (2012) that there are central signaling elements which are involved in the induction of flowering time. Natural variation in temperature-related responses could be caused by different polymorphisms of signaling or response genes ranging from alteration in gene sequence to expression level polymorphism (Delker et al., 2011) due to adaption to local environmental conditions. As they provide keys to altered temperature responses that could be utilized in specific breeding approaches, these genes would thus be of high interest.

In conclusion linear models and the quantification of variances between genotypes and temperatures (λ_{gen} and λ_{temp} values) have been shown to be useful approaches to perform a systematic assessment of a phenotypic data set covering the whole life cycle of different *A. thaliana* ecotypes. The application of these methods allowed us to identify phenotypes whose temperature-related changes are possibly due to natural variation and thus driven by the genotypes.

2.2.6. Applications beyond Arabidopsis and auxin

The methods and bioinformatics algorithms we introduced are neither restricted to *Arabidopsis* and nor to auxin treatment.
For instance, the PMP (Poeschl et al., 2013) is composed of modules. These modules are replaceable by other modules as long as these modules fulfill the recommended interface requirements. Also, the modules themselves can solve various different tasks. The first module can be used to verify selected primers for RT-qPCR. Input primer sequences are compared to

available transcriptomes and thus be checked for cross-hybridization. Another specific application of the first and the second module is that they can be used to re-annotate the probes on microarrays if new genome annotations are available. For example, we used the first and second module of the PMP to re-annotate a tiling microarray used for honey bee to measure gene expression (Dussaubat et al., 2012). This data set was part of the honey bee synthesis data set analysed for the project Trans-Bee (*The Trans-Bee workshop* 2014). This synthesis data set besides the *A. thaliana* and *A. lyrata* data set was used in Poeschl et al. (2014) to show that the introduced PIF algorithm provides biological relevant information. The synthesis data set included data sets measured on different platforms for different tissues of bees that were exposed to different pathogens and viruses. To make the data sets comparable, we transformed the provided fold changes, which represent the change between treatment and control sample, into relative ranks. In Poeschl et al. (2014) we showed that the PIF algorithm is also applicable to data matrices containing relative ranks and provides biological information that can be used for further gene analyses.

Hence, we presented bioinformatics algorithms that are not only designed for one special purpose but can be applied to address various tasks.

2.3. References

Balasubramanian, S., Sureshkumar, S., Lempe, J., and Weigel, D. (2006). Potent Induction of *Arabidopsis* thaliana Flowering by Elevated Growth Temperature. *PLoS Genetics*, 2 (7), e106.

Boer, D. R., Freire-Rios, A., Berg, W. A. M. van den, Saaki, T., Manfield, I. W., Kepinski, S., López-Vidrieo, I., Franco-Zorrilla, J. M., Vries, S. C. de, Solano, R., Weijers, D., and Coll, M. (2014). Structural Basis for DNA Binding Specificity by the Auxin-Dependent ARF Transcription Factors. *Cell*, 156 (3), pp. 577–589.

Broadley, M. R., White, P. J., Hammond, J. P., Graham, N. S., Bowen, H. C., Emmerson, Z. F., Fray, R. G., Iannetta, P. P. M., McNicol, J. W., and May, S. T. (2008). Evidence of neutral transcriptome evolution in plants. *New Phytologist*, 180 (3), pp. 587–593.

Calderón Villalobos, L. I. A., Lee, S., De Oliveira, C., Ivetac, A., Brandt, W., Armitage, L., Sheard, L. B., Tan, X., Parry, G., Mao, H., Zheng, N., Napier, R., Kepinski, S., and Estelle, M. (2012). A combinatorial TIR1/AFB–Aux/IAA co-receptor system for differential sensing of auxin. *Nature Chemical Biology*, 8 (5), pp. 477–485.

CaraDonna, P. J., Iler, A. M., and Inouye, D. W. (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences*, 111 (13), pp. 4916–4921.

Chapman, E. J. and Estelle, M. (2009). Mechanism of Auxin-Regulated Gene Expression in Plants. *Annual Review of Genetics*, 43 (1). PMID: 19686081, pp. 265–285.

Delker, C., Pöschl, Y., Raschke, A., Ullrich, K., Ettingshausen, S., Hauptmann, V., Grosse, I., and Quint, M. (2010). Natural Variation of Transcriptional Auxin Response Networks in *Arabidopsis thaliana*. *The Plant Cell*, 22 (7), pp. 2184–2200.

Delker, C. and Quint, M. (2011). Expression level polymorphisms: heritable traits shaping natural variation. *Trends in Plant Science*, 16 (9), pp. 481–488.

Donner, A. and Koval, J. J. (1980). The Estimation of Intraclass Correlation in the Analysis of Family Data. *Biometrics*, 36 (1), pp. 19–25.

Dussaubat, C., Brunet, J.-L., Higes, M., Colbourne, J. K., Lopez, J., Choi, J.-H., Martín-Hernández, R., Botías, C., Cousin, M., McDonnell, C., Bonnet, M., Belzunces, L. P., Moritz, R. F. A., Le Conte, Y., and Alaux, C. (2012). Gut Pathology and Responses to the Microsporidium *Nosema ceranae* in the Honey Bee *Apis mellifera*. *PLoS ONE*, 7 (5), e37017.

Fang, G., Bhardwaj, N., Robilotto, R., and Gerstein, M. B. (2010). Getting Started in Gene Orthology and Functional Analysis. *PLoS Computational Biology*, 6 (3), e1000703.

Fitter, A. H. and Fitter, R. S. R. (2002). Rapid Changes in Flowering Time in British Plants. *Science*, 296 (5573), pp. 1689–1691.

Franklin, K. A., Lee, S. H., Patel, D., Kumar, S. V., Spartz, A. K., Gu, C., Ye, S., Yu, P., Breen, G., Cohen, J. D., Wigge, P. A., and Gray, W. M. (2011). PHYTOCHROME-INTERACTING FACTOR 4 (PIF4) regulates auxin biosynthesis at high temperature. *Proceedings of the National Academy of Sciences*, 108 (50), pp. 20231–20235.

Gansner, E. R. and North, S. C. (2000). An open graph visualization system and its applications to software engineering. *Software - Practice and Experience*, 30 (11), pp. 1203–1233.

Graham, N., Broadley, M., Hammond, J., White, P., and May, S. (2007). Optimising the analysis of transcript data using high density oligonucleotide arrays and genomic DNA-based probe selection. *BMC Genomics*, 8 (1), p. 344.

Grau, J., Posch, S., Grosse, I., and Keilwagen, J. (2013). A general approach for discriminative de novo motif discovery from high-throughput data. *Nucleic Acids Research*, 41 (21), e197.

Gray, W. M., Östin, A., Sandberg, G., Romano, C. P., and Estelle, M. (1998). High temperature promotes auxin-mediated hypocotyl elongation in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 95 (12), pp. 7197–7202.

Guilfoyle, T. J., Ulmasov, T., and Hagen, G. (1998). The ARF family of transcription factors and their role in plant hormone-responsive transcription. *Cellular and Molecular Life Sciences*, 54 (7), pp. 619–627.

Hammond, J., Broadley, M., Craigon, D., Higgins, J., Emmerson, Z., Townsend, H., White, P., and May, S. (2005). Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. *Plant Methods*, 1 (1), p. 10.

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., Fahlgren, N., Fawcett, J. A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J. D., Ossowski, S., Ottilar, R. P., Salamov, A. A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M. E., Bergelson, J., Carrington, J. C., Gaut, B. S., Schmutz, J., Mayer, K. F. X., Van de Peer, Y., Grigoriev, I. V., Nordborg, M., Weigel, D., and Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, 43 (5), pp. 476–481.

Ibañez, C., Poeschl, Y., Peterson, T., Bellstädt, J., Denk, K., Gogol-Döring, A., Quint, M., and Delker, C. (2015). Developmental plasticity of *Arabidopsis thaliana* accessions across an ambient temperature range. *bioRxiv*, pre-print, doi: 10.1101/017285.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4 (2), pp. 249–264.

Katagiri, F. and Glazebrook, J. (2003). Local Context Finder (LCF) reveals multidimensional relationships among mRNA expression profiles of *Arabidopsis* responding to pathogen infection. *Proceedings of the National Academy of Sciences*, 100 (19), pp. 10842–10847.

Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigele, S., Do, H.-H., Weiss, G., Enard, W., Heissig, F., Arendt, T., Nieselt-Struwe, K., Eichler, E. E., and Pääbo, S. (2004). Regional Patterns of Gene Expression in Human and Chimpanzee Brains. *Genome Research*, 14 (8), pp. 1462–1473.

Koini, M. A., Alvey, L., Allen, T., Tilley, C. A., Harberd, N. P., Whitelam, G. C., and Franklin, K. A. (2009). High Temperatur-Mediated Adaptations in Plant Architecture Require the bHLH Transcription Factor PIF4. *Current Biology*, 19 (5), pp. 408–413.

Kumar, S. V., Lucyshyn, D., Jaeger, K. E., Alos, E., Alvey, E., Harberd, N. P., and Wigge, P. A. (2012). Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature*, 484 (7393), pp. 242–245.

Laan, M. J. van der and Pollard, K. S. (2003). A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap. *Journal of Statistical Planning and Inference*, 117 (2), pp. 275–303.

Murtagh, F. and Contreras, P. (2011). Methods of Hierarchical Clustering. *CoRR*, abs/1105.0121.

Opgen-Rhein, R. and Strimmer, K. (2007). Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach. *Statistical Applications in Genetics and Molecular Biology*, 6 (1), pp. 477+.

Paponov, I. A., Paponov, M., Teale, W., Menges, M., Chakrabortee, S., Murray, J. A. H., and Palme, K. (2008). Comprehensive Transcriptome Analysis of Auxin Responses in *Arabidopsis*. *Molecular Plant*, 1 (2), pp. 321–337.

Poeschl, Y., Delker, C., Trenner, J., Ullrich, K. K., Quint, M., and Grosse, I. (2013). Optimized Probe Masking for Comparative Transcriptomics of Closely Related Species. *PLoS ONE*, 8 (11), e78497.

Poeschl, Y., Grosse, I., and Gogol-Döring, A. (2014). Explaining gene responses by linear modeling. *German Conference on Bioinformatics*, Volume P-235 of Lecture Notes in Informatics (LNI) - Proceedings, pp. 27–35.

Quint, M. and Gray, W. M. (2006). Auxin signaling. *Current Opinion in Plant Biology*, 9 (5), pp. 448–453.

Ramos, J. A., Zenser, N., Leyser, O., and Callis, J. (2001). Rapid Degradation of Auxin/Indoleacetic Acid Proteins Requires Conserved Amino Acids of Domain II and Is Proteasome Dependent. *The Plant Cell*, 13 (10), pp. 2349–2360.

Redman, J. C., Haas, B. J., Tanimoto, G., and Town, C. D. (2004). Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *The Plant Journal*, 38 (3), pp. 545–561.

Salehin, M., Bagchi, R., and Estelle, M. (2015). SCFTIR1/AFB-Based Auxin Perception: Mechanism and Role in Plant Growth and Development. *The Plant Cell*, 27 (1), pp. 9–19.

The Trans-Bee workshop (2014). http://www.idiv-biodiversity.de/sdiv/workshops/workshops-2013/stransbee (accessed 2014/07/23).

Trenner, J., Poeschl, Y., Grau, J., Gogol-Döring, A., Quint, M., and Delker, C. (in prep.). Variation of IAA-induced transcriptomes pinpoints the AUX/IAA network as a potential source for inter-species divergence in auxin signaling and response. *not announced*.

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. (1999). Activation and repression of transcription by auxin-response factors. *Proceedings of the National Academy of Sciences*, 96 (10), pp. 5844–5849.

Vinga, S. and Almeida, J. (2003). Alignment-free sequence comparison—a review. *Bioinformatics*, 19 (4), pp. 513–523.

Yilmaz, A., Mejia-Guerra, M. K., Kurz, K., Liang, X., Welch, L., and Grotewold, E. (2011). AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. *Nucleic Acids Research*, 39 (suppl 1), pp. D1118–D1122.

Yona, G., Dirks, W., Rahman, S., and Lin, D. M. (2006). Effective similarity measures for expression profiles. *Bioinformatics*, 22 (13), pp. 1616–1622.

Zenser, N., Ellsmore, A., Leasure, C., and Callis, J. (2001). Auxin modulates the degradation rate of Aux/IAA proteins. *Proceedings of the National Academy of Sciences*, 98 (20), pp. 11795–11800.

The Plant Cell, Volume 22, Pages 2184-2200, July 2010

3. Natural variation of transcriptional auxin response networks in Arabidopsis thaliana

Carolin Delker[1,*], Yvonne Poeschl[2,*], Anja Raschke[1], Kristian Ullrich[1], Stefan Ettinghausen[1], Valeska Hauptmann[1], Ivo Grosse[2], and Marcel Quint[1]

[1] Leibniz Institute of Plant Biochemistry, Independent Junior Research Group, 06120 Halle (Saale), Germany

[2] Institute of Computer Science, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany

[*] These authors contributed equally to this work.

3.1. Abstract

Natural variation has been observed for various traits in Arabidopsis thaliana. Here, we investigated natural variation in the context of physiological and transcriptional responses to the phytohormone auxin, a key regulator of plant development. A survey of the general extent of natural variation to auxin stimuli revealed significant physiological variation among 20 genetically diverse natural accessions. Moreover, we observed dramatic variation on the global transcriptome level after induction of auxin responses in seven accessions. Although we detect isolated cases of major-effect polymorphisms, sequencing of signaling genes revealed sequence conservation, making selective pressures that favor functionally different protein variants among accessions unlikely. However, coexpression analyses of a priori defined auxin signaling networks identified variations in the transcriptional equilibrium of signaling components. In agreement with this, cluster analyses of genome-wide expression profiles followed by analyses of a posteriori defined gene networks revealed accession-specific auxin responses. We hypothesize that quantitative distortions in the ratios of interacting signaling components contribute to the detected transcriptional variation, resulting in physiological variation of auxin responses among accessions.

3.2. Introduction

Naturally occurring genetic variation has been reported for numerous phenotypes in *Arabidopsis thaliana*. In addition to various developmental traits, response phenotypes that are primarily correlated with adaptations to natural environments have been under investigation. The stimuli triggering the respective responses ranged from pathogens or effectors to different

light conditions, abiotic stress, and a variety of other environmental perturbations (reviewed in Alonso-Blanco et al., 2009).

The translation of a stimulus into cellular responses is often mediated by plant hormones. Auxin in particular is known to be a potent regulator of various aspects of plant development (Delker et al., 2008). At the cellular level, auxin responses are initiated by altering the expression of a multitude of genes, which requires the proteolytic degradation of transcriptional repressors by the 26S proteasome (Quint et al., 2006). In the absence of auxin, AUXIN/INDOLE-3-ACETIC ACID (Aux/IAA) proteins repress auxin signaling by heterodimerization with transcription factors of the AUXIN RESPONSE FACTOR (ARF) family (Tiwari et al., 2003). With increasing auxin levels, the Aux/IAA proteins bind to the auxin receptors. These consist of a small family of F-box proteins (TRANSPORT INHIBITOR RESPONSE1/AUXIN SIGNALING F-BOX PROTEIN [TIR1/AFB]) that integrate into functional S-phase kinase-associated protein, Cullin, F-box (SCF)TIR1/AFB complexes (Dharmasiri et al., 2005b; Dharmasiri et al., 2005a; Kepinski et al., 2005; Parry et al., 2006) and confer substrate specificity to the complex. Aux/IAA proteins are recruited for polyubiquitination and are subsequently degraded by the proteasome (Ramos et al., 2001; Zenser et al., 2001). This allows the ARF transcription factors to initiate downstream auxin responses by regulating the expression of auxin-responsive genes (Guilfoyle et al., 1998; Ulmasov et al., 1999). Such auxin responses can be summarized as cell division, cell differentiation, and cell elongation: essential cellular processes that can translate into an array of different physiological phenotypes.

Many plant developmental events and reactions in response to environmental cues are tightly regulated by auxin and other phytohormones. Natural variation in hormone responses, however, has not been studied in detail as yet (Maloof et al., 2001; Delker et al., 2008). Phytohormones usually act via extensive reprogramming of expression patterns for a unique cassette of genes (Nemhauser et al., 2006). Up to now, the intraspecific variation in phytohormone-induced transcriptional responses has only been assessed for salicylic acid (SA; Leeuwen et al., 2007). For other phytohormones (e.g., auxins), the impact or even presence of natural variation has hardly been approached experimentally at all. While it is obvious that natural variation should exist for pathways that specifically regulate adaptation to certain natural environment perturbations, it is uncertain whether this is also true for essential conserved messenger systems that transduce multiple environmental or developmental signals into specific responses. As such, the auxin signaling pathway is an ideal model to study naturally occurring genetic variation of essential messenger systems.

We have investigated the natural variation in auxin responses and signaling at the physiological, population genetic, and transcriptional levels. First, classic physiological auxin response assays were used to assess the general extent of natural variation. Second, nucleotide diversities were estimated for early auxin signaling elements to determine potential differences in the signaling ability of natural accessions. Third, network analyses of ATH1-based transcriptional profiles were used to investigate the variation and outcomes in global transcriptome changes of seven accessions in response to an auxin stimulus. Finally, based on our data, we present a model to explain the observed variation in various response levels.

3.3. Results

3.3.1. Natural variation of physiological auxin responses

The high degree of natural variation observed for numerous physiological traits prompted us to study the physiological responses to auxin in 20 different accessions, which represent a maximal degree of genetic diversity (Clark et al., 2007). We performed standard bioassays to quantify root inhibition and hypocotyl elongation in response to auxin and found significant differences between accessions with respect to absolute root length and growth responses (Figure 3.1; see Supplemental Figures A.1-A.3 online). Phenotypic variation in root growth was higher in response to the synthetic auxins naphthylacetic acid (NAA) and 2,4-D than to the natural auxin IAA (Figures 1A-1C). This phenomenon is likely attributable to a slower removal via catabolization of the synthetic auxins, whereas a large excess of IAA is usually rapidly removed by conjugation to amino acids, sugars, or direct oxidation (Delker et al., 2008). High temperatures promote auxin-mediated hypocotyl elongation by increasing endogenous auxin contents (Gray et al., 1998; Stavang et al., 2009). To analyze potential variations in the response to resulting increased endogenous auxin levels, plants were grown at elevated temperatures (29°C) and the increase in hypocotyl elongation was quantified for each accession and found to differ significantly in many pair-wise comparisons (Figure 3.1D; see Supplemental Figure A.4 online). Remarkably, individual accessions varied in their responses depending on the specific auxin and type of assay (root versus hypocotyl assays). One can assume, therefore, that the mechanisms underlying the variations in response to different auxins are not uniformly regulated but rather result from complex mechanisms in a tissue-specific manner.

Additional evidence for intraspecific variation in auxin responses was obtained by analysis of the activation of the synthetic auxin reporter construct *DR5:GUS* in three accessions that differed significantly in their response to IAA-induced root growth inhibition (see Supplemental Figure A.3 online). The analysis of several independent and homozygous T3 lines revealed considerable differences among Fei-0, Sha, and Col-0 in histochemical β-glucuronidase (GUS) assays (Figure 3.1E; see Supplemental Figure A.5 online). The extent of DR5 promoter activation was determined by quantitative (q)RT-PCR of *GUS* expression after mock treatment or treatment with three different IAA concentrations. Col-0 showed the strongest response in auxin-induced expression changes, whereas the levels in Sha were significantly lower. Fei-0 exhibited *GUS* expression responses intermediate to Col-0 and Sha (Figure 3.1F). In addition, we analyzed two known endogenous auxin-responsive genes, *GH3.1* and *IAA2*, in the transgenic *DR5:GUS* lines. The expression response of *GH3.1* showed similar results to those already detected for the *GUS* gene. Even although the accession-specific differences in the expression response of *IAA2* were not quite as distinct, the general trend in expression responses was confirmed. Here, too, significant differences between Col-0 and the other two accessions were detectable (Figure 3.1F).

Alterations in the expression responses could be the result of differences in endogenous auxin concentrations causing hypersensitive/hyposensitive reactions to an additional exogenous auxin stimulus. Therefore, we quantified free IAA levels in 7-d-old seedlings and were unable to identify significant differences among the seven accessions that were further analyzed in this study

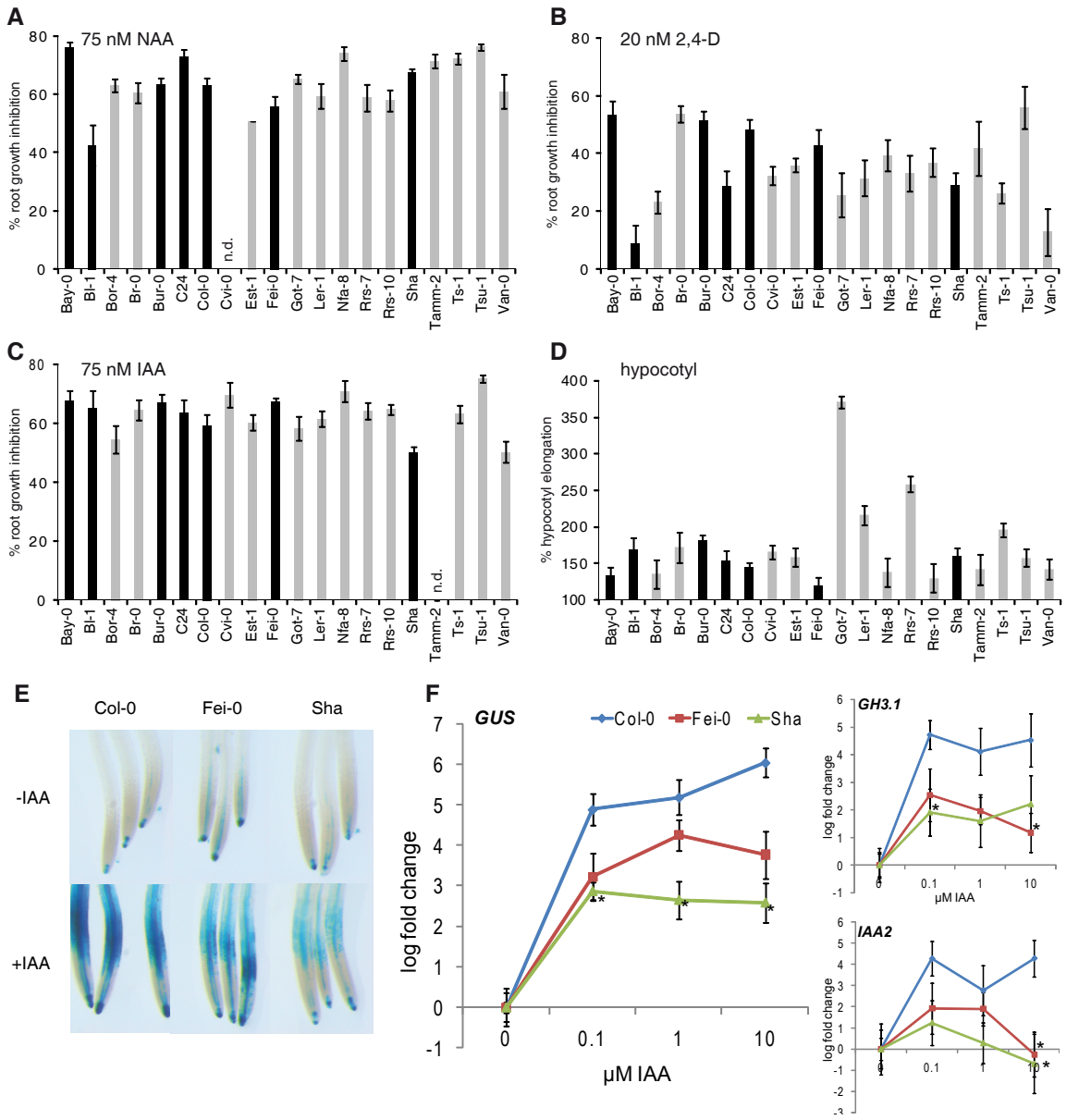


Figure 3.1.: Natural Variation in Physiological Auxin Responses. **(A)** to **(D)** Physiological auxin responses of 20 *Arabidopsis* accessions were determined in root growth inhibition and hypocotyl elongation assays of 8- and 10-d-old seedlings (n = 12), respectively. Black bars highlight accessions that were subsequently analyzed for whole genome transcriptome changes. Error bars show SD. Experiments were repeated twice with similar results. Data of absolute root and hypocotyl lengths and statistical analyses are shown in Supplemental Figures A.1 to A.4 online. **(A)** to **(C)** Bars represent mean root length of treated roots as a percentage of untreated roots. **(D)** Hypocotyl elongation of seedlings grown at 29°C is given in % relative to seedlings grown at 20°C. **(E)** Histochemical detection of GUS activity after 3 h of mock treatment (-IAA) or treatment with 1 μM IAA (+IAA). Three seedlings of a single representative T3 line are shown for each accession. All independent T3 lines for each accession are shown in Supplemental Figure A.5 online. **(F)** Quantification of *GUS*, *IAA2* and *GH3.1* expression by qRT-PCR 1 h post induction (p.i.) with 0.1, 1 and 10 μM IAA, respectively. Mean log fold changes (treatment versus mock) in expression were determined by analysis of eight, six, and seven independent T3 lines for Fei-0, Sha, and Col-0, respectively. Error bars denote SE. Significant differences from Col-0 expression responses were assessed by two-way ANOVA and are marked by asterisks.

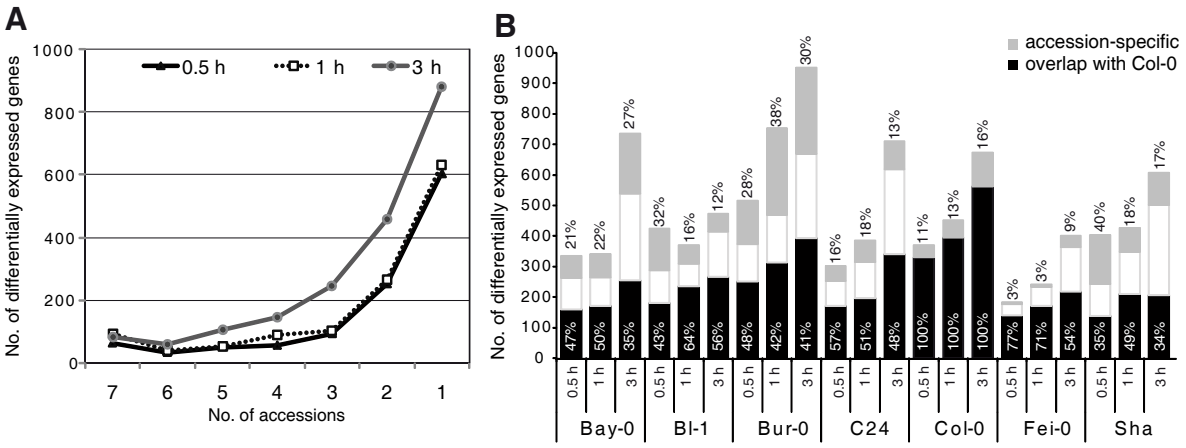


Figure 3.2.: Accession-Specific Differences in Auxin-Induced Transcriptional Changes. (A) Differentially expressed genes with a significant ($P < 0.05$; Benjamini-Hochberg corrected) expression change of at least twofold (i.e., $\Delta \log_2 > 1$) compared with untreated plants were categorized by the number of accessions in which they were differentially expressed. **(B)** Bar plots show the number of differentially expressed genes in individual expression profiles. The fraction of genes that is specifically regulated in an individual accession is indicated in gray (black numbers), whereas the fraction of genes also differentially expressed in Col-0 is marked in black (white numbers).

(see Supplemental Figure A.6 online). Thus, auxin responsiveness is most likely not affected by endogenous IAA levels in these accessions.

3.3.2. Arabidopsis accessions differ in auxin-induced transcriptional changes

The expression data of the *DR5:GUS* transgenic lines suggested that differences in auxin sensitivity and expression responses might contribute to the observed variation. To gain a more global insight into the differential auxin responses on a transcriptional level, we performed ATH1-based expression profiling of auxin responses with a set of 7 of the 20 accessions that differed in their phenotypic auxin response (Figure 3.1). To avoid potential secondary effects, we performed a time-course analysis that focused on the early transcriptional changes induced by auxin. Seven-day-old *Arabidopsis* seedlings were treated with 1 μM IAA, and samples were taken before induction (0 h) and at 0.5, 1, and 3 h post induction (hpi). Auxin-induced transcriptional changes were detectable in all seven accessions, with an average of 651 genes that showed a significant (Benjamini-Hochberg-corrected $P < 0.05$) auxin response of at least twofold change in expression levels at 3 hpi. Surprisingly, many of these genes are differentially expressed in three or fewer accessions, whereas only ~100 genes showed a twofold or higher expression change in all seven accessions (Figure 3.2A). Auxin-induced transcriptional responses of 17 arbitrary genes of the latter group were independently reexamined across all time points by qRT-PCR. The relatively high correlation coefficient of $r_s = 0.8$ (Spearman correlation coefficient) between both data sets offered further validation of the microarray data (see Supplemental Figure A.7 online) and indicated the robustness of the expression levels detected by microarray analysis (Czechowski et al., 2004).

While the total number of genes with an auxin-induced transcriptional response was similar for 0.5 and 1 hpi, the numbers of differentially expressed genes increased notably 3 h after the

auxin stimulus. This is in agreement with previously published data (Goda et al., 2008) and most likely denotes the establishment of secondary responses following an auxin treatment. To obtain further insight into the apparent diversity of the transcriptome, we compared the differentially expressed genes between all analyzed accessions. The overlap of individual accessions with Col-0 ranged from only 34% (Sha; 3 hpi) up to 77% (Fei-0; 0.5 hpi). Hence, a relatively large proportion of genes showed an auxin-induced expression change in one or more accessions other than Col-0 or were specifically induced in a single accession (Figure 3.2B).

The variation in differentially expressed genes could be indicative of hypersensitive and hyposensitive auxin responses on the expression level. To address this hypothesis, we compared the number of differentially expressed genes as well as the respective amplitudes of expression changes between all accessions. In both cases, significant differences were observed (see Supplemental Figure A.8 online). However, no clear correlation between the number of differentially induced genes and median fold changes in expression was observed; thus, based on this criterion, we could not justify the classification in truly hyperresponsive or hyporesponsive accessions. As such, the variation in the total number of genes as well as the different degrees of accession specificity and Col-0 overlap can serve only as general indicators for a high variability in auxin-induced transcriptional changes in different *Arabidopsis* accessions.

3.3.3. Intraspecific variation of whole genome responses

Whole genome expression profiles of all accessions at individual time points were compared to further assess the degree of natural variation. Identification of common patterns in such complex data sets is usually complicated by the multidimensional nature of the data. Thus, we used the Local Context Finder (LCF; Katagiri et al., 2003), a nonlinear dimensionality reduction method for pattern recognition. In contrast to other coexpression algorithms, an important advantage of the LCF is the translation of multidimensional relationships between expression profiles into a two-dimensional network that makes complex interactions more intelligible. To reduce the effect of possible noise and to filter for robust coexpressions, we applied a bootstrapping procedure as suggested by Katagiri et al. (2003). Expression profiles are presented as nodes within the LCF-generated networks, and interconnections between them are presented as directed edges.

LCF analysis of whole genome transcriptome profiles separated the seven analyzed accessions into three groups (Figure 3.3A). Bay-0 and Sha represent one isolated group, and C24 and Fei-0 constitute another. The third group is formed by Col-0 and Bur-0. Bl-1 shows no clear affiliation with a specific group, and Bl-1 nodes share edges with all accessions except Fei-0 (Figure 3.3). While edges within each group were quite frequent, considerably fewer edges connect nodes of one group with nodes of another. In general, all nodes of an accession are tightly linked to each other regardless of the time point. Edges between nodes of different accessions can only be detected for identical time points (Figure 3.3). This illustrates a tight temporal regulation of auxin responses and argues against delays or shifts in the kinetics of auxin responses as the cause for the observed variation. In summary, global auxin-induced expression changes among *Arabidopsis* accessions differ considerably in comparison with each

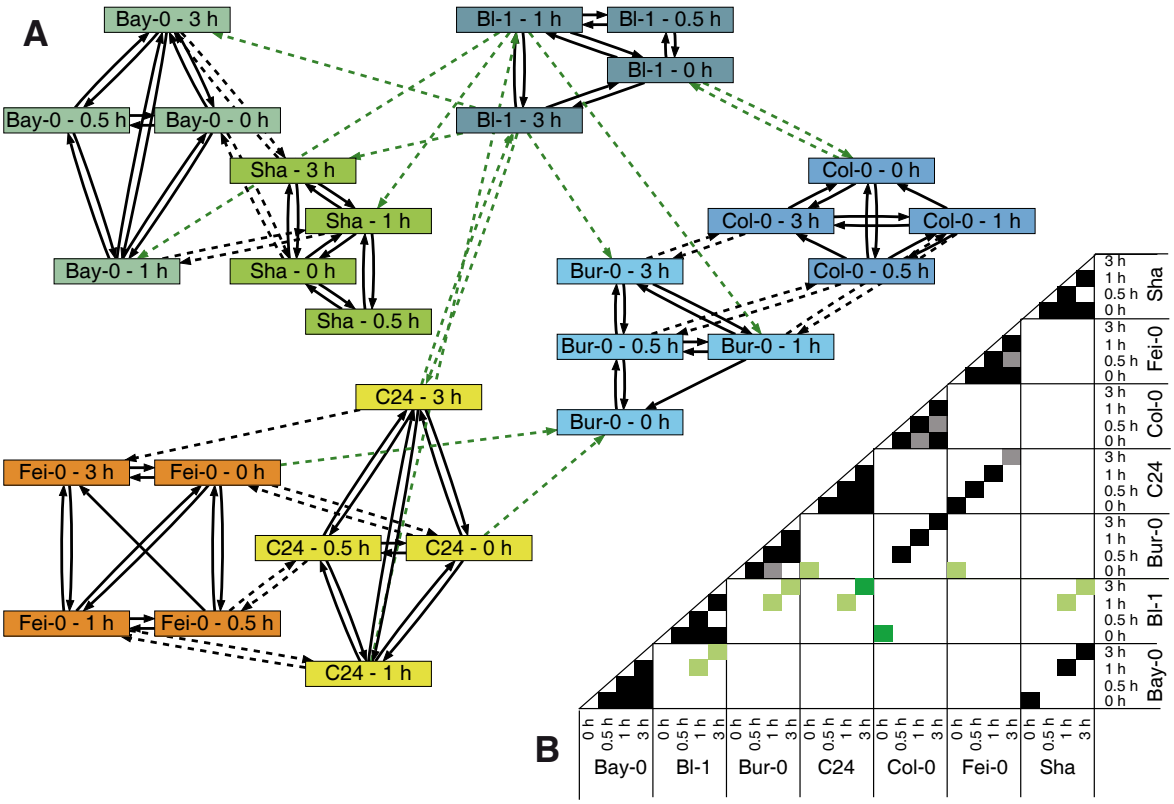


Figure 3.3.: Intraspecific Variation in Whole Genome Transcriptome Profiles. (A) Profiles for individual accessions were compared by LCF. Similarities in profiles of an individual accession at different time points post induction are indicated by solid lines; dashed lines represent similarities between different accessions at similar time points. Edge colors specify similarities between accessions within a subgroup (black) or between accessions of another subgroup (green). **(B)** Tabular presentation of edges detected within the LCF network. Black/dark green and gray/light green squares denote the presence of two edges and one edge between nodes, respectively; black/gray squares represent edges within the same subgroup; green squares represent edges between different subgroups.

other as well as with the reference accession Col-0, illustrating the large potential for variation in the regulation of diverse auxin-regulated processes.

3.3.4. Sequence diversity of auxin signaling genes

The $SCF^{TIR1/AFB}$-dependent signaling pathway regulates the expression of auxin response genes (Quint et al., 2006). A possible cause for the above-described natural variation in the transcriptional and subsequent physiological auxin responses, therefore, may be variations at the level of early signaling events. It is likely that slight changes in the function or the equilibrium of signaling components would contribute to the dramatic differences we observed in the transcriptional response downstream of the initial signaling events. Genes encoding signaling elements may display accession-specific differences at the sequence level, possibly resulting in signaling components with altered biochemical properties between accessions. To test this hypothesis, we analyzed the sequence diversity of auxin signaling genes for 19 of the accessions used in this study. Three gene families were considered: (1) the *TIR1/AFB*

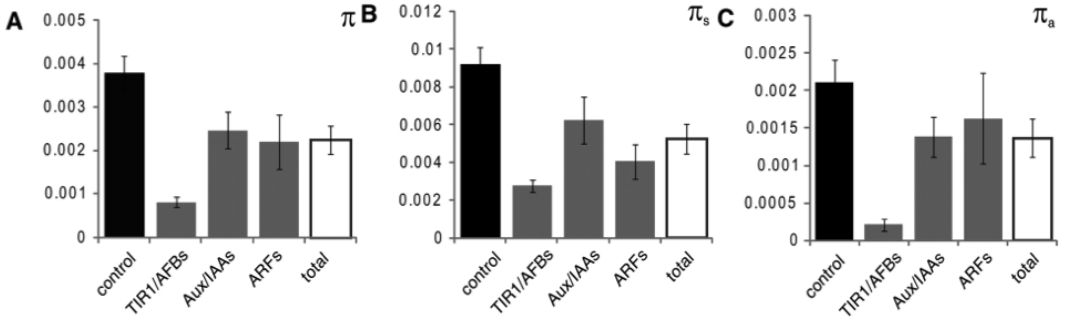


Figure 3.4.: Nucleotide Diversity of Auxin Signaling Genes. Nucleotide diversity (p) was determined for all sites **(A)**, synonymous sites **(B)**, and nonsynonymous sites **(C)** by comparing ∼1-kb fragments in 19 accessions. Results were summarized for control genes (n = 236) as well as for the three gene families of receptors (n = 4), *Aux/IAA*s (n = 29), and *ARF*s (n = 16) separately and combined (total, n = 49). Error bars denote SE. Summary statistics for individual genes are presented in Supplemental Data Set 1 online.

auxin receptors, (2) the *Aux/IAA* repressors, and (3) the *ARF* transcription factors. We sequenced ∼1-kb fragments and identified two major-effect changes with potential functional consequences. First, *IAA11* contained a splice site specific to Col-0 and Ler-1, which results in a different splice variant than in the other 17 accessions. Second, *ARF13* contains premature stop codons or alternative splice variants in Sha, Tsu-1, Tamm-2, and Bay-0. However, *ARF13* generally does not have the ARF–Aux/IAA interaction domains III and IV (Okushima et al., 2005); therefore, it is unlikely that these mutations are of functional significance.

Taking a molecular population genetic approach, we analyzed whether patterns of selection resulting in sequence diversification would favor the possibility that functional differences in signaling genes contribute to the transcriptional and physiological variation in response to an auxin signal. On the other hand, selective constraints resulting in sequence conservation might argue against such a hypothesis. As a control data set for the population genetic approach, we used an empirical distribution of genome-wide polymorphisms as suggested previously by Kreitman (2000) and Nordborg et al. (2005). For the analysis of the empirical distribution, we took advantage of 876 equally spaced fragments sequenced from a panel of 96 accessions (which include all our analyzed accessions except for Bl-1) generated by Nordborg et al. (2005). We then calculated nucleotide diversities for the coding sequences of the auxin signaling genes and compared them with the empirical distribution (i.e., control genes). The nucleotide diversity π can be used to measure the degree of polymorphism within a population (Nei et al., 1979). We measured π for all sites and for synonymous (π_s) and nonsynonymous (π_a) sites separately. Figure 3.4 depicts the nucleotide diversities for the auxin signaling gene families and the control genes in a bar plot. The underlying gene-wise summary statistics are shown in Supplemental Data Set 1 online.

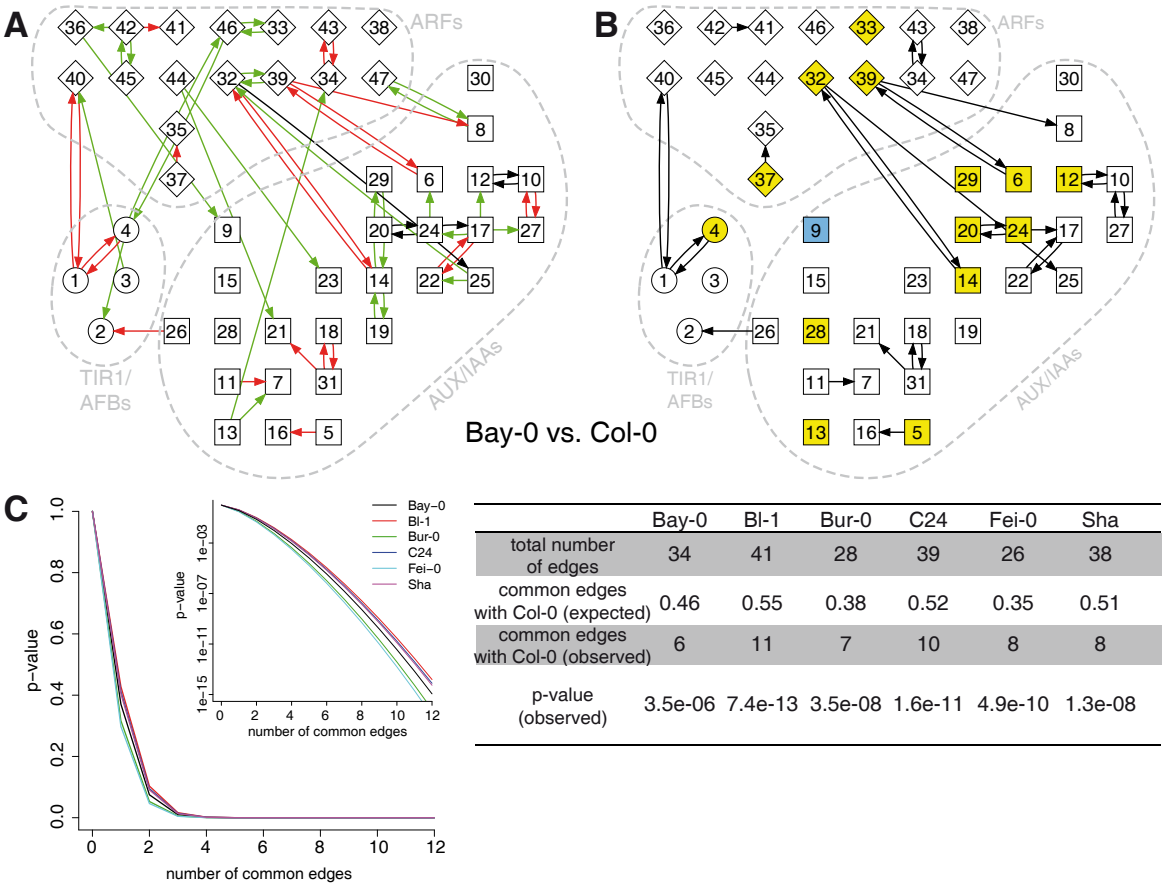
The summary statistics showed no significant deviations from the control genes. For the auxin receptors, however, we found evidence for lower values for π (P = 0.15) and π_a (P = 0.06). Auxin receptors are members of the superfamily of F-box genes that belong, together with nucleotide-binding-Leu-rich repeat genes, to the most diverse and rapidly evolving gene families in the *Arabidopsis* genome (Clark et al., 2007). Hence, if these data were compared with those of the F-box gene superfamily instead of the empirical distribution, significant

differences indicating some degree of purifying selection could be expected. Likewise, we detected lower values for the assayed nucleotide diversities for *Aux/IAA*s and *ARF*s (Figure 3.4). However, although we identified genes with no amino acid substitutions in each gene family (see Supplemental Table A.1 online), it has to be kept in mind that, theoretically, a single nonsynonymous mutation at a functional residue might result in functional variation at the protein level. In summary, although not statistically significant, these results demonstrate that sequence diversity among the accessions tested was rather low for auxin signaling genes.

3.3.5. Coexpression networks of auxin signaling genes

While the conservation of auxin signaling genes at the sequence level indicates a possible conservation of functional protein properties, differences in the transcriptional regulation of signaling genes may directly influence auxin responses by causing changes in specific TIR1/AFB-Aux/IAA and/or Aux/IAA-ARF interactions. This would in turn contribute to the diversity in downstream transcriptional responses of different accessions. To address this hypothesis, we used the LCF to inspect the transcriptional coregulation of signaling genes. This a priori defined network represented the same three gene families as above: (1) *TIR1/AFB*s, (2) *Aux/IAA*s, and (3) *ARF*s (see Supplemental Table A.2 online for gene list). After LCF analysis of individual accessions, including bootstrapping of the signaling gene expression profiles, we performed pair-wise comparisons between networks of Col-0 and the other accessions. This revealed remarkable differences between the individual network structures (Figure 3.5A; see Supplemental Figure A.10 online). As in every coexpression analysis, the edges detected by LCF do not necessarily reflect a true functional or physical interaction between linked neighbors. Nevertheless, the results bear functional significance for the comparison of different accessions, as the LCF provides characteristic pattern information or fingerprints for each individual complex data set. The overlap of edges detected in the networks of Col-0 and the respective other accessions ranged from 18% for Bay-0 to 31% for Fei-0 (Figures 3.5A and 3.5C; see Supplemental Figure A.10 online). The majority of edges seemed to be specific for the network of an individual accession, which indicates that the individual expression profiles of genes differ considerably between accessions. Since *Aux/IAA* genes are known to be auxin-inducible themselves, it is not surprising that connections in individual networks were most prevalent among *Aux/IAA* gene family members, which confirms the validity of this approach. Most of these connections, however, seem to be specific for the expression set of a single accession, whereas only a few appear to be more conserved and can be detected in several of the analyzed comparisons, such as the edges connecting genes 20 (*IAA2*) and 24 (*IAA1/AXR5*).

To ensure that the detected common edges between Col-0 and the respective edges in other accessions represent robust congruencies and did not result by chance, we computed P values based on the hypergeometric distribution of the number of expected common edges, which ranged from 0.35 to 0.55 (Figure 3.5C). For each accession, the number of common edges is significantly higher than expected (Figure 3.5C). The probability of the six common edges between Bay-0 and Col-0 occurring by chance was found to be 3.5×10^{-6}. This probability was even lower for the 11 common edges detected between Bl-1 and Col-0 (7.4×10^{-13}; Figure 3.5C). Hence, the detected common edges in our pair-wise comparisons are highly significant and represent true overlaps in the coexpression networks of the respective accessions.



	Bay-0	Bl-1	Bur-0	C24	Fei-0	Sha
total number of edges	34	41	28	39	26	38
common edges with Col-0 (expected)	0.46	0.55	0.38	0.52	0.35	0.51
common edges with Col-0 (observed)	6	11	7	10	8	8
p-value (observed)	3.5e-06	7.4e-13	3.5e-08	1.6e-11	4.9e-10	1.3e-08

Figure 3.5.: Coexpression Analyses of Auxin-Induced Transcriptional Changes in Signaling Genes. **(A)** Coregulation across time points of genes encoding TIR1/AFB auxin receptors (1-4, circles), AUX/IAA proteins (5-31, squares), and ARFs (32-47, diamonds) was analyzed by LCF. Red edges indicate connections detected specifically in the network of Col-0, green edges are specific for Bay-0, and black edges represent connections detected in both networks. **(B)** Differences in auxin-induced changes for signaling genes at 1 hpi are highlighted in blue and yellow for significantly lower or higher transcriptional responses, respectively, in Bay-0 compared with Col-0 ($P < 0.05$, Benjamini-Hochberg corrected). A complete summary of pair-wise comparisons of LCF networks and time point-specific expression changes of all accessions is available as Supplemental Figures A.9 to A.11 online. **(C)** Probability of the number of common edges between Col-0 (29 edges) and all other accessions. The plot shows probabilities of common edges, the inset shows identical data on a logarithmic scale, and the tabulated data are a summary of the results.

The overall low numbers of common edges with Col-0 indicate considerable deviations in the transcriptional equilibrium of signaling components. Since the LCF procedure only considers the shapes of expression profiles but fails to provide information about the respective amplitudes, we performed a modified t test for small sample sizes (Opgen-Rhein et al., 2007) to compare the degree of expression changes induced by the auxin stimulus. For all three time points, significant differences were detected for various genes in pair-wise comparisons with Col-0 (Figure 3.5B; see Supplemental Figure A.11 online). Highest variations in expression levels were detected for genes of the *Aux/IAA* family, whereas expression changes in *ARF* genes varied considerably less. Almost no variation was detectable in the transcriptional changes of the auxin receptor family (see Supplemental Figure A.11 online).

3.3.6. Cluster analysis

The significant differences in the coexpression networks and expression levels of signaling genes suggested a more detailed analysis of the global transcriptional responses. To reduce the complexity of the expression data sets, we first performed cluster analysis using the Col-0 expression data. This resulted in the identification of 112 clusters, many of which contained only a single gene. We chose a minimum cutoff of six genes per cluster, which reduced the cluster number to 51. These were further reduced to 46 by only considering clusters that showed a significant transcriptional response in at least one of the analyzed accessions at least at one time point.

Inspection of heat maps of the mean expression changes seen at 0.5, 1, and 3 hpi illustrates that some clusters (e.g., clusters 91, 97, and 100) showed relatively strong differences among accessions, whereas others (e.g., clusters 99 and 101) showed a relatively uniform response, with more subtle differences between accessions (Figure 3.6A). A dendrogram based on hierarchical clustering of all three time points separates the accessions into different groups (Figure 3.6B). Bay-0 and Sha as well as Fei-0 and C24 form distinct groups from Bl-1, Col-0, and Bur-0, confirming the pattern in expression variation between individual accessions detected by LCF analysis of whole genome expression data (Figure 3.3A).

For a more detailed analysis of the expression differences, we (1) compared the coexpression of clusters by LCF and (2) inspected the mean expression levels of clusters at individual time points post induction. In both cases, pair-wise comparisons between Col-0 and the other accessions were performed.

For the coexpression analysis, the mean expression response profiles of genes within a cluster were generated and subjected to LCF analysis, resulting in accession-specific coexpression networks. These were subjected to pair-wise comparisons of individual accessions and Col-0, which again revealed a high degree of diversity between the individual LCF networks, in agreement with the high degree of diversity observed for the signaling genes (Figure 3.5A; see Supplemental Figures A.10 and A.12 online).

To assess the extent to which the mean cluster expression changes differ among accessions, we inspected auxin-induced expression changes at individual time points after auxin induction. The network structure and connections detected by LCF for the Col-0 data were selected for

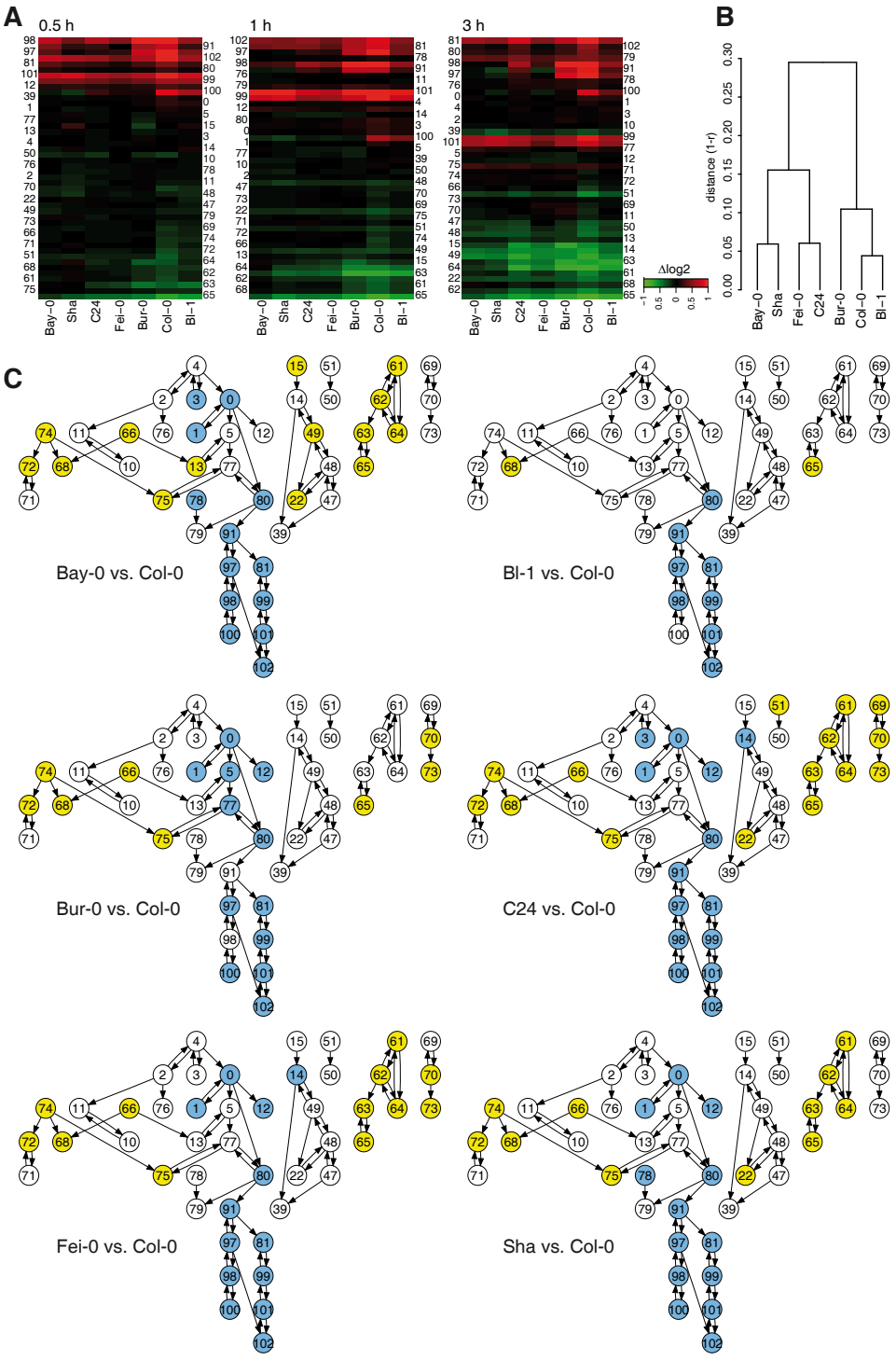


Figure 3.6.: Accession-Specific Differences in the Expression Response of Gene Clusters. **(A)** Mean expression changes of all genes within individual clusters were calculated for each time point post induction and are presented as heat maps. **(B)** Dendrogram based on the hierarchical clustering of expression data of all four time points. 1 - r(Pearson) was used as a distance measure of the agglomerative hierarchical clustering with complete linkage. **(C)** Pair-wise comparison of cluster expression levels at 1 hpi ($P < 0.05$, Benjamini-Hochberg corrected). Deviations from Col-0 expression changes are highlighted in blue and yellow for significantly lower and higher expression changes, respectively. Network structure was obtained by LCF analysis of Col-0 cluster expression data (see Supplemental Figure A.12 online).

convenient visualization. The significant differences in the mean expression changes between an accession and Col-0 are indicated by colored nodes (Figure 3.6C). Numerous alterations from the Col-0 expression levels were detected for several clusters, and some of them were accession-specific. Interestingly, some general patterns of congruency can be observed as well, since many clusters with significantly altered expression levels were detectable as such in almost all pair-wise comparisons (Figure 3.6C). In this respect, the group formed by clusters 81 to 102 is of special interest. In most cases, the expression changes of these clusters are significantly lower than in Col-0 with only a few exceptions (e.g., cluster 100, 1 hpi; Figure 3.6C; see Supplemental Figure A.13 online). This group of clusters is also highlighted by the LCF coexpression analysis (see Supplemental Figure A.12 online). First, the expression profiles of clusters within this group showed strong responses to the auxin treatment (see Supplemental Data Set 2 online). Second, since common edges between Col-0 and other accessions appeared in many of the pair-wise comparisons, auxin responsiveness seemed to be rather conserved within this group. Although accession-specific edges also occurred among this group of clusters, the number of edges that overlapped with Col-0 was exceptionally high (see Supplemental Figure A.12 online). Last, many genes that have been shown to be auxin-responsive in a broad-scale microarray data-mining approach (Paponov et al., 2008) were found within these clusters (see Supplemental Data Set 3 online). Among them were several *GH3* and *Aux/IAA* genes, transport-associated genes (e.g., *PIN3*), and representatives of the LOB domain (LBD) transcription factor family (e.g., *LBD16*).

Similarly, expression changes of several clusters seemed to be uniformly higher in the six accessions than in Col-0 (e.g., 61-65; Figure 3.6C), indicating extensive deviations from the transcriptional response observed in Col-0.

3.3.7. Accession-specific expression differences in selected clusters

To demonstrate an example of accession-specific differences, we further investigated the transcriptional responses of the 100 individual genes of cluster 100. A close-up view of the expression changes of individual genes 1 hpi within this cluster demonstrated that the majority of genes indeed showed almost contrary transcriptional responses between several accessions and Col-0 (Figure 3.7; see Supplemental Figure A.14 online). Expression changes in Bl-1 and Col-0 genes were similar, and Bl-1 and Col-0 were thus grouped into a clade separated from the five other accessions. This grouping was caused mainly by a large block of ∼60 genes that were upregulated in Col-0 and Bl-1 and downregulated in the other five accessions (Figure 3.7). To verify the differential response detected in the microarray data, we reexamined the transcriptional responses for a subset of 10 of the total 100 genes by qRT-PCR. For 8 out of 10 genes, auxin-induced transcriptional responses detected by microarray analyses were reproducible by qRT-PCR (Figure 3.7), which confirms that accession-specific regulation of gene clusters does occur in response to auxin stimuli.

The high degree of accession specificity observed on the level of gene clusters complicates a direct identification of a single cluster or a few clusters of genes, which are potentially responsible for the observed natural variation of physiological auxin responses. Therefore, we performed a second analysis to correlate the expression and physiological IAA responses on

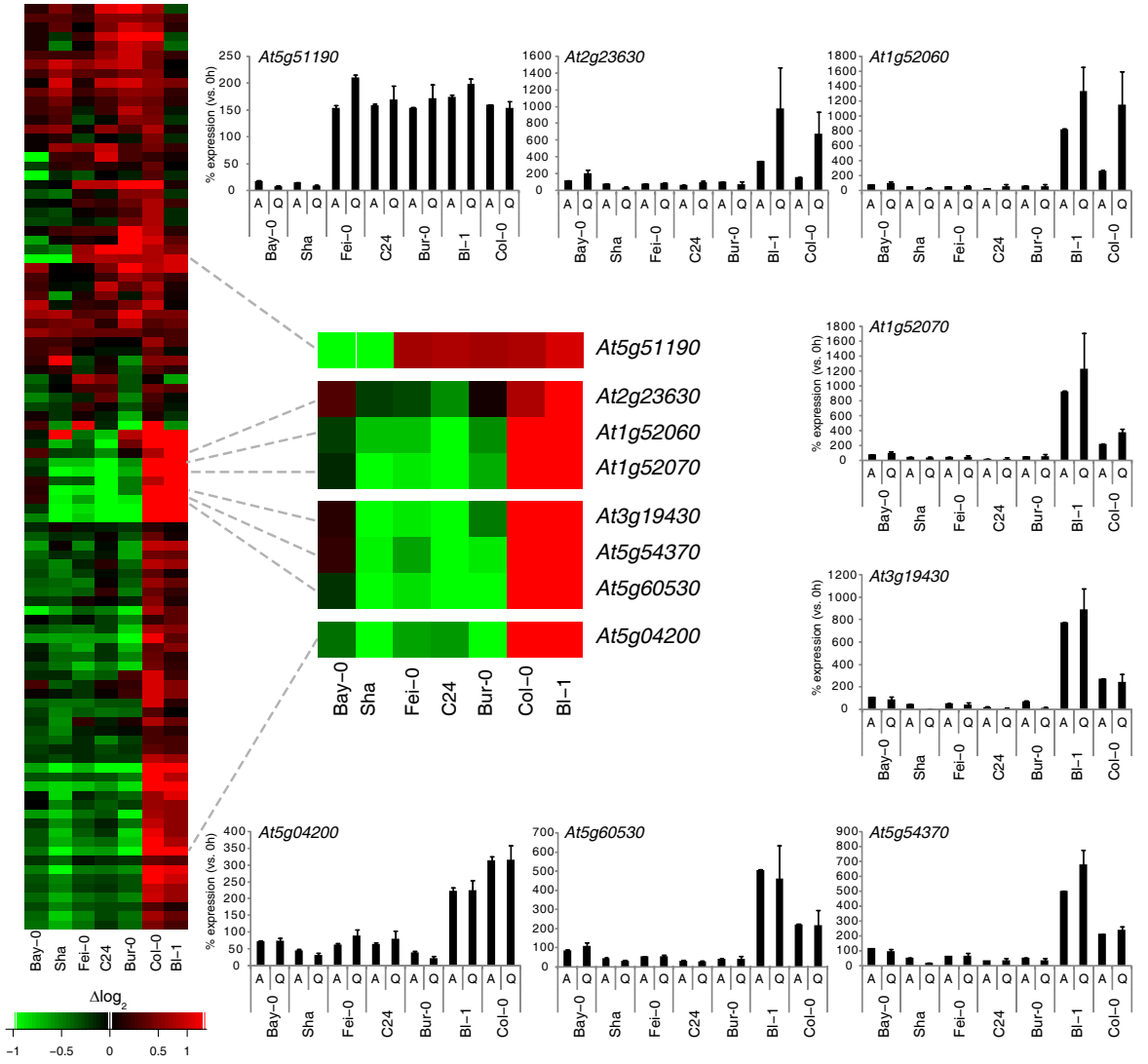


Figure 3.7.: Accession-Specific Expression Differences at 1 hpi within Cluster 100. Microarray data (A) were validated with qRT-PCR (Q) for eight selected genes. Error bars represent SE (n = 3). For a higher resolution of the cluster 100 heat map, see Supplemental Figure A.14 online.

the level of individual genes. The expression profiles of the genes with highest and lowest Pearson r values mirror almost perfectly the variation observed in the physiological assay (see Supplemental Figures A.15A and A.15C online). Among the 230 genes with $r > 0.8$, a significant enrichment of Gene Ontology terms related to hormone responses and transcription factor activity was detected (see Supplemental Figure A.15B online). Notably, one member of the *Aux/IAA* gene family, *IAA5* (*AT1G15580*), was highly correlated, with an r value of 0.94. However, determination of the functional relevance of these genes for the variation in downstream auxin responses needs to be further investigated, and these results can only be seen as indicators of potential factors that contribute to the extensive variation found in auxin responses.

3.4. Discussion

In plants, many traits exhibit a high degree of interspecific as well as intraspecific variation, and many of these traits have been extensively analyzed in the reference plant species *A. thaliana*. The vast majority of traits investigated from the perspective of natural variation are studied for their proposed roles in adaptations to natural environments. Less is known about integrative signaling pathways essential for the coordination of a multitude of specific environmental and/or developmental stimuli. Therefore, we investigated the genetic variation of auxin responses in natural accessions of *Arabidopsis* on the physiological, molecular population genetic, and transcriptional levels.

3.4.1. Natural variation of physiological and transcriptional auxin responses

We first demonstrated for a set of 20 genetically diverse accessions that variations in physiological auxin responses were evident (Figure 3.1). Auxin responses were, at least in part, tissue-specific, and they depended on the auxin compound applied. Differences observed on the physiological level were also reflected in variations in a *DR5:GUS* reporter assay (Figures 3.1E and 3.1F). In agreement with this, we detected extensive variation in auxin-induced transcriptional responses in seven of the phenotyped accessions (Figure 3.2). The actual degree of accession specificity measured was surprisingly high. However, similar or even higher accession-specific differences were previously described for transcriptional changes induced by SA (Kliebenstein et al., 2006; Leeuwen et al., 2007) or between parental strains of recombinant inbred line populations used in expression quantitative trait locus (eQTL) studies (Keurentjes et al., 2007). Likewise, the presence of intraspecific variation for pathogen responses in *Arabidopsis* has been demonstrated for various pathogens (Poecke et al., 2007; Rowe et al., 2008; Narusaka et al., 2009). In this respect, natural variation in SA responses might likely aid adaptation processes related to plant defense mechanisms. While the known SA functions are restricted, one would assume a much tighter regulation of responses to a signaling molecule that is known to translate a multitude of stimuli into diverse responses, such as auxin. Therefore, the degree of variation we detected might be considered unexpected.

3.4.2. Global auxin response networks

LCF analysis enabled us to detect and visualize patterns of congruency and variation in global auxin-induced transcriptomes between accessions (Figure 3.3). The results divided the accessions into different subgroups (Figure 3.3A). This classification was largely confirmed by hierarchical clustering of the expression data, which indicates robust differences between accessions (Figure 3.6B). Congruencies were detected most frequently for the transcriptional profiles of an individual accession at different time points (Figure 3.3B). This can be expected, since only ∼1 to 3% of the genes were differentially regulated by auxin. The strong similarities of transcriptomes, therefore, are most likely due to the remaining ∼97 to 99% nonresponsive genes whose expression levels are constant within the same accession. However, the finding that transcriptional profiles of different accessions are only coregulated at similar time points

(Figure 3.3) indicates a tight temporal regulation of auxin responses. This argues against accession-specific time shifts in transcriptional regulation as a possible cause for the observed variation.

3.4.3. Sequence conservation of auxin signaling genes

Mechanisms known to be involved in the manifestation of natural variation must be encoded genetically or epigenetically. Evidence for extensive sequence variation among Arabidopsis accessions has been obtained in several studies (Borevitz et al., 2007; Clark et al., 2007; Zhang et al., 2008). We considered genetic variation in the *TIR1/AFB*, *Aux/IAA*, and *ARF* gene families. All signaling genes known from the Col-0 reference genome were present in the other tested accessions. Therefore, we can exclude the absence of signaling genes in accessions other than Col-0 as a cause for variation, although we cannot rule out the presence of additional family members in these accessions that are absent in Col-0. We identified two isolated events of major-effect polymorphisms, but their functional significance is questionable and restricted to the affected accessions. On a population level auxin signaling genes seemed more conserved than the control genes (Figure 3.4; see Supplemental Table A.1 online). This argues against selective pressures that favor the existence of multiple protein variants with different functions among accessions as the primary cause for the downstream transcriptional variation.

3.4.4. Transcriptional networks of auxin signaling and response Genes

Alternatively, sequence polymorphisms in *cis*-regulatory regions or *trans*-acting factors can affect the regulation of gene expression, which will influence the spatiotemporal conditions and amounts in which the resulting gene product is present to exert its function. The regulation of auxin responses requires direct physical interactions between signaling components, including (1) TIR1/AFBs recruiting Aux/IAAs for proteasomal degradation, (2) heterodimerization of Aux/IAAs and ARFs, and (3) heterodimerization and homodimerization of ARFs. All three signaling components are encoded by gene families in *Arabidopsis*. It has been hypothesized that a major key to auxin's ability to regulate such a wealth of diverse processes is constituted by the multitude of putatively different interactions of this tripartite signaling ensemble (Lokerse et al., 2009). In agreement with this hypothesis, individual members of each gene family are at least partially expressed in a temporal- and tissue-specific manner (Dharmasiri et al., 2005b; Teale et al., 2006). Analyses of loss- and gain-of-function mutants have provided further evidence for functional specificities as well as redundancies of TIR1/AFBs, ARFS, and Aux/IAAs, respectively (Chapman et al., 2009; Parry et al., 2009). While several *Aux/IAA* loss-of-function mutants do not show obvious phenotypes in the Col-0 background (Overvoorde et al., 2005), others do exhibit auxin-related defects, as in the case of *SHORT HYPOCOTYL2* (*SHY2/IAA3*; Tian et al., 1999). Interestingly, *SHY2/IAA3*, among others, shows accession-specific differences in coexpression patterns and expression levels in several pair-wise comparisons with Col-0 (Figure 3.6C; see Supplemental Figures A.12 and A.13 online). This does not necessarily prove a relevant role for *SHY2/IAA3* in the observed variation in downstream auxin responses. However, it shows that we find alterations for signaling components with known specificity. Taken together, selected *Aux/IAA*s may be sensitive

to transcriptional downregulation (such as *SHY2/IAA3*), while others most likely are not. Furthermore, there are several examples of phenotypes resulting from overexpression alleles of individual *Aux/IAA*s (Nagpal et al., 2000; Ploense et al., 2009). Therefore, *Aux/IAA*s may be more prone to cause downstream natural variation by transcriptional upregulation in certain accessions.

In agreement with these findings, it has been shown for Col-0 that modifications in expression patterns that ultimately change the protein abundance of signaling components can have a profound effect on auxin signaling. Changes in perception of the auxin stimulus to more or less receptive can be modulated by increasing or decreasing the levels of auxin receptors (Ruegger et al., 1998; Navarro et al., 2006; Pérez-Torres et al., 2008). If such differences in receptor levels were genetically fixed between accessions, auxin responses would likewise differ. However, we have detected only minimal variation in expression profiles of receptor-encoding genes. This argues against a general hyperresponse or hyporesponse of an accession, even though this might be the case in specific tissues or could be mediated by posttranscriptional processes. Alterations in transcriptional responses and coexpression networks were more frequent among *ARF*s and *Aux/IAA*s than between receptor and ARF- or Aux/IAA-coding genes, respectively (Figures 3.5A and 3.5B; see Supplemental Figures A.10 and A.11 online), which suggests that the accession-specific equilibria of *Aux/IAA*s and *ARF*s show considerable variation, at least in comparison with the reference strain Col-0. Here, posttranscriptional processes such as modifications or turnover rates might further affect the equilibrium of signaling components. For instance, both Aux/IAAs and ARFs are subjected to proteasomal degradation. In contrast to Aux/IAAs, the degradation of ARFs seems to be auxin- and TIR1-independent, as shown in the case of ARF1 (Salmon et al., 2008). In this respect, auxin-independent factors could also contribute to the observed natural variation in expression profiles. Based on the assumption that specific compositions of available signaling components trigger specific auxin responses, we hypothesize that the accession-specific alterations in the equilibria of available signaling components contribute considerably to the variation in downstream transcriptional responses. As a consequence, accession-specific clusters of coregulated response genes could be expected. These were indeed observed and validated by qRT-PCR (Figures 3.6 and 3.7). Ultimately, this may then translate into quantitative differences in physiological responses to the auxin stimulus. In a highly simplified scheme, the basal auxin response is caused by a balanced equilibrium between Aux/IAA and ARF proteins that function together in the regulation of a particular trait, such as root growth inhibition (Figure 3.8). In response to an auxin stimulus, a shift in the composition and relative amount of Aux/IAA and ARF proteins enables downstream responses. The strength of the triggered response in different accessions depends on the equilibria of the activating and repressing signaling components involved in the specific process. Since our data set is based on whole seedlings, functional analysis and dissection of putative changes in Aux/IAA-ARF compositions and subsequent correlation to specific physiological processes such as root growth are not easily achieved. This would require tissue-specific or even cell type-specific analysis of transcriptional auxin responses, which would be essential to circumvent dilution effects caused by the mixture of different tissue types and/or the simultaneous exhibition of different auxin responses at the transcriptional level (Teale et al., 2006; Paponov et al., 2008).

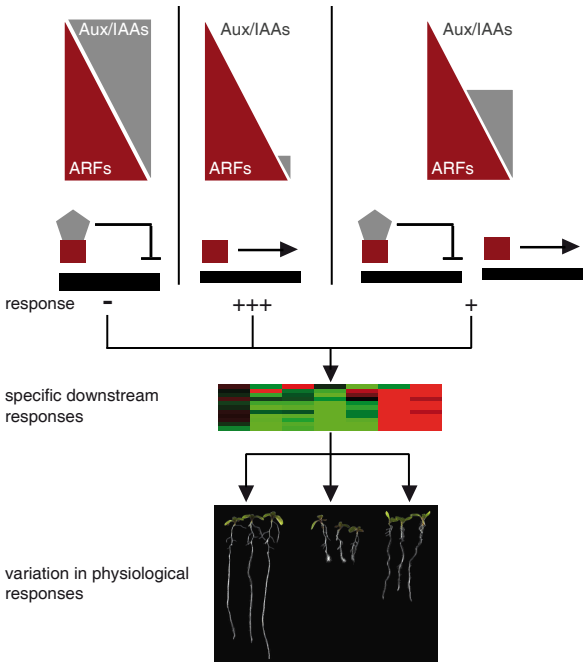


Figure 3.8.: Model of the Putative Impact of Different Aux/IAA-ARF Equilibria on Downstream Transcriptional and Physiological Responses. ARFs (red square) are known to regulate the activity of primary auxin response genes. The direct interaction with Aux/IAA proteins (gray pentagon), however, inhibits ARF function. Therefore, differences in the equilibria of interacting Aux/IAAs and ARFs are likely to trigger alterations in downstream transcriptional responses that might ultimately contribute to variations at the physiological level.

3.4.5. Identification of specific factors involved in the natural variation of auxin responses

Our model suggests that quantitative distortions in signaling element compositions contribute to the downstream variation in auxin responses. This raises two major questions: (1) what influences/regulates the variation at the level of signaling genes? and (2) what are the relevant downstream factors that are actually involved in the regulation of auxin responses at the physiological level? Auxin biology is influenced by several major processes, such as biosynthesis, metabolism, transport, and signaling. Natural variation in signaling gene expression and subsequent responses could, in principle, be caused by variation in all of these processes or by different sensitivities to stimuli that have been shown to influence them, such as other phytohormones or circadian rhythms (Covington et al., 2007). While extensive accession-specific differences in expression patterns have been observed in several studies, it is also evident that most likely only a fraction of the detectable differences will actually contribute to the variation seen on a phenotypic level. This effect of genetic buffering has also been assessed in *Arabidopsis*, and it was shown that 16% of the transcriptional variation detected between Col-0 and Ler can be attributed to as few as six QTL "hot spot" regions (Fu et al., 2009). Based on our data, we can make no estimation on how many factors are actually involved in the regulation of the observed expression level polymorphisms (ELPs). Some ongoing experiments in our laboratory based on accession intercrosses and QTL analysis favor quantitative genetics versus Mendelian inheritance. Hence, we assume that the causative factors for the observed natural variation in auxin responses are regulated and inherited in a quantitative genetic manner. Future studies,

such as eQTL analysis, need to be employed to unravel the cause of the detected ELPs. Even though we were able to identify genes whose variations in expression profiles correlate with the physiological responses (see Supplemental Figure A.15 online), their actual relevance in the observed variation in auxin responses needs to be verified in tissue-specific approaches.

In summary, we found that natural variation does not only exist for traits and pathways that display obvious ecological relevance in terms of adaptive advantages/specializations, but natural variation is similarly present in a pathway that is essential for the integration of numerous developmental and environmental stimuli. The extensive accession-specific variations in auxin responses add yet another fundamental set of data to the startlingly complex picture of intraspecific variation. The data obtained in this study suggest that the temporal response to auxin stimuli is tightly regulated and conserved across accessions. Furthermore, natural *Arabidopsis* accessions generally possess the same set of signaling genes. Although isolated polymorphisms resulting in functional variation on the protein level cannot be ruled out, auxin signaling genes seem to be highly conserved at the population level. However, ELPs within these gene families seem to contribute considerably to the variation in downstream responses. This may ultimately impact the observed natural variation in physiological responses and, thereby, potentially contribute to adaptation processes. Future approaches will need to focus on specific physiological responses or tissue-specific assays in order to facilitate the association of phenotypes to specific genes or to relevant expression profiles in causative tissues.

3.5. Methods

3.5.1. Plant material and growth conditions

Arabidopsis thaliana accessions were obtained from the Nottingham Arabidopsis Stock Centre and the Arabidopsis Biological Resource Center (see Supplemental Table A.3 online for stock numbers). Seeds were surface-sterilized and imbibed in deionized H_2O for 3 d at 4°C before sowing. Seedlings were germinated and grown under sterile conditions on Arabidopsis thaliana solution (ATS) nutrient medium (Lincoln et al., 1990).

For root growth assays, seedlings were cultivated vertically on unsupplemented ATS for 3 d (IAA) or 5 d (2,4-D and NAA) before transfer to plates supplemented with IAA, 2,4-D, or NAA at the indicated concentrations. Root lengths were quantified after an additional 5 d (IAA) or 3 d (2,4-D and NAA). Hypocotyl growth was quantified in seedlings cultivated for 10 d under long-day lighting conditions at 29°C. Root and hypocotyl lengths of hormone- and heat-treated seedlings, respectively, were determined in relation to seedlings grown on unsupplemented ATS medium at 20°C.

For expression studies, seedlings of the seven accessions were germinated and cultivated in liquid ATS under continuous illumination to minimize potential circadian effects. After 7 d, ATS was supplemented with 1 μM IAA for 0, 0.5, 1, and 3 h. Yellow long-pass filters were applied in all IAA treatment experiments to prevent photodegradation of IAA.

3.5.2. Statistical analysis of physiological data

After $\log_2$ transformation of the four physiological growth response data sets (IAA, 2,4-D, NAA, and hypocotyl), the following two variants of analysis of variance (ANOVA) were performed. (1) Treated and untreated samples were separated for each of the data sets. For both sets of 20 treated and untreated samples, a one-way ANOVA with 20 groups corresponding to the 20 accessions studied was performed. Subsequently, the Tukey post hoc test was conducted to identify the pairs of accessions that are significantly different among the 20 treated samples and among the 20 untreated samples. (2) A two-way ANOVA with 2 $\times$ 2 groups was conducted for each of the 190 pairs of the 20 accessions, testing which pairs of accessions show a significantly different response to auxin. The resulting P values were corrected for multiple testing using the Benjamini-Hochberg correction method.

3.5.3. Microarray experiments and qRT-PCR analyses

RNA was extracted from the homogenized plant material of whole seedlings (7 d) of seven accessions grown in liquid culture in three biological replicates for each time point (with the exception of only two replicates for the 3-hpi Fei-0 sample) using the Qiagen RNeasy Plant Mini Kit with an on-column DNase treatment. Further processing of purified RNA and hybridization to whole genome Affymetrix ATH1 GeneChip microarrays was performed by the Nottingham Arabidopsis Stock Centre's International Affymetrix Service (`http://affymetrix.arabidopsis.info/`). Processing of plant material and RNA purification for qRT-PCR were performed similarly. One microgram of total RNA was subjected to reverse transcription by SuperScript III reverse transcriptase (Invitrogen). Power SYBR Green PCR Master Mix (Applied Biosystems) was used for subsequent quantitative real-time PCR analyses. The *PP2A* catalytic subunit gene *At1g13320* served as the constitutively expressed reference gene (Czechowski et al., 2005). Comparative expression levels for the respective genes of interest were calculated as $2^{(\text{Ct reference gene - Ct gene of interest})}$. For oligonucleotide sequences and a complete list of analyzed genes, see Supplemental Table A.3 online.

3.5.4. DR5:GUS cloning, plant transformation, and histochemical Glucuronidase staining

The *DR5:GUS* construct (Ulmasov et al., 1997) was transferred into the binary vector pGWB1 (Nakagawa et al., 2007) by Gateway cloning via the entry vector pDONR221 (Invitrogen). *Agrobacterium tumefaciens*-mediated (GV3101) transformation of *Arabidopsis* accessions was performed by floral dip (Clough et al., 1998), and transgenic seedlings were selected as described previously (Quint et al., 2009). For the histochemical glucuronidase assays, seedlings of eight (Fei-0), six (Sha), and seven (Col-0) independent and homozygous T3 lines were mock treated (0.1% ethanol in liquid ATS) or treated with 1 μM IAA in liquid ATS medium for 3 h and stained overnight at 37°C, as described previously (Stomp, 1991). *GUS* expression in seedlings of the same transgenic lines mock treated or treated with 0.1, 1, or 10 μM IAA for 1 h was quantified by qRT-PCR as described above. For statistical analysis of the *GUS*

expression response, a two-way ANOVA was performed on the log-transformed comparative expression level data as described for the analysis of the physiological data (see above).

3.5.5. Quantitation of free IAA

Five hundred milligrams of plant material of 7-d-old seedlings was homogenized with 10 mL of methanol and 100 ng of [$^{13}C_6$]IAA as internal standard. The homogenate was filtered and placed on a 3-mL DEAE-Sephadex A25 column (Amersham Pharmacia Biotech) followed by three subsequent wash steps with 0.1, 1, and 1.5 N acetic acid in methanol. Elution with 3 mL of 3 N acetic acid in methanol was repeated three times. The solvent of the eluate was evaporated, and the residue was resuspended in 110 μL of 50% methanol followed by HPLC using a Eurospher 100-C18 column. Appropriate fractions were evaporated, resuspended in 100 μL of methanol, and incubated with 400 μL of diazomethane for 10 min. After evaporation of the solvent, samples were resuspended in 60 μL of acetonitrile and analyzed by gas chromatographymass spectrometry (PolarisQ; Thermo-Finnigan).

3.5.6. Statistical analyses

Statistical analyses were performed using the software R. The hopach package (2.4.0), multtest package (2.0.0), and simpleaffy package (2.20.0) were obtained from `www.bioconductor.org`. The fdrtool package (1.2.5), gplots package (2.6.0), st package (1.1.1), and stats package (2.9.0) were downloaded from `www.r-project.org`.

3.5.7. Processing of microarray data

The simpleaffy package was used to obtain robust multi-chip average-normalized $\log_2$ expression levels using default settings. A quality control analysis was conducted using the simpleaffy package to verify that the arrays had similar hybridization efficiencies and background intensities for all accessions.

3.5.8. Defining gene clusters

Genes were classified according to their expression profile patterns in Col-0 using a hierarchical clustering algorithm named HOPACH (Laan et al., 2003) from the hopach package with Pearson correlation coefficient (1 - r) as a distance measure. In total, 112 clusters were identified. The number of genes per cluster ranged from 1 to 1301, with an average of 203.1 genes. Only clusters with at least six genes were retained, which resulted in 51 clusters. The gene-to-cluster mapping of the Col-0 clustering was also used for the other six accessions. The mean expression profile of each of the clusters was computed for each of the seven accessions as the arithmetic mean of the expression values of the genes contained in the cluster. Each mean expression profile has 12 values corresponding to the four time points and the three biological replicates (two replicates for Fei-0, 3 hpi) for each time point. Complete lists of cluster expression profiles and gene cluster identities are presented in Supplemental Data Sets

2 and 3 online, respectively. Each mean expression profile was tested for significant changes in expression levels to retain only the relevant clusters that respond significantly to auxin treatment in at least one accession and at least one time point. To this end, we conducted a one-way ANOVA with four groups (corresponding to the four time points) for each of the seven accessions using the stats package. The obtained P values were adjusted for multiple testing using the Benjamini-Hochberg procedure from the multtest package. Five of the 51 clusters did not reach a Benjamini-Hochbergcorrected P value below 0.05 and were eliminated, yielding a final set of 46 clusters. The initial cluster numbers according to the HOPACH clustering result were maintained ranging from 0 to 111.

3.5.9. Coexpression network analysis by LCF

Network analyses were conducted on the expression data using LCF following the procedure described by Katagiri et al. (2003), with a maximum number of seven possible neighbors (k = 7). The visualization of the graphs was done using Graphviz version 2.20.3 (`http://www.graphviz.org`) with neato as layout algorithm.

Three data sets were subjected to LCF analysis: (1) whole transcriptome data sets of all accessions, (2) expression profiles of 47 auxin signaling genes analyzed separately for each accession, and (3) mean expression profiles of 46 predefined clusters also analyzed separately for each accession. To filter for robust edges in the LCF networks, we used the bootstrapping approach (Katagiri et al., 2003) by keeping genes and drawing with replacement experiment vectors. For each of the 1000 surrogate data sets generated by the bootstrapping approach, one LCF network was computed.

For analysis of variations in whole genome transcriptomes, the averaged $\log_2$ expression values of the three replicates of each of the four time points for each of the seven accessions were analyzed by LCF. All of the 28 analyzed transcriptome profiles consisted of 22,746 genes. For analysis of auxin signaling elements, all genes coding for TIR1/AFBs, Aux/IAAs, or ARFs were selected that are represented by a single, unique probe on the ATH1 microarray, resulting in 47 genes (see Supplemental Table A.2 online). LCF and bootstrapping were performed for each accession individually based on the expression profiles consisting of all replicates at all time points. Subsequently, the Col-0 network was compared with the networks of the other six accessions by determining common edges of both networks. To determine the robustness and significance of the number of common edges, 1000 surrogate data sets were generated by bootstrapping for each of the accessions.

LCF analysis and bootstrapping of cluster coexpressions was also performed separately for each accession based on the averaged expression profiles, each consisting of 12 values of the 46 predefined clusters. Individual networks were subsequently compared with the Col-0 network.

3.5.10. Expression level analysis

Variations in auxin-induced expression changes of a gene or cluster at individual time points post induction were analyzed by using the modified t test of the st package from Opgen-Rhein et al. (2007). The mean $\Delta\log_2$ expression values of a gene between each treatment time point and the 0-h control were computed for each accession. The resulting P values were Benjamini-Hochberg corrected for multiple testing. The results were projected on the Col-0 LCF network structures defined above. Significant differences ($P < 0.05$) in expression changes of a gene or a cluster are denoted by colored nodes for each accession. Nodes were colored blue and yellow if the corresponding t value was lower or higher than 0, respectively.

3.5.11. Heat maps

Heat maps were generated using the heatmap.2 function of the gplots package. The Pearson correlation coefficient (1 - r) was used as a distance measure of the agglomerative hierarchical clustering with complete linkage. Heat maps of mean $\Delta\log_2$ values were generated for clusters and signaling genes (see Supplemental Figure A.9 online). Data sets were separated into three parts based on the time point post induction. The dendrogram of the accessions was computed based on the mean $\Delta\log_2$ cluster data of all three time points also using agglomerative hierarchical clustering with complete linkage and 1 - r as a distance measure.

3.5.12. Correlation analysis of physiological and expression data

The $\Delta\log_2$ levels of the physiological IAA responses were computed for each of the seven accessions used for the expression studies, resulting in a profile consisting of seven values. For each of the 22,810 genes in the expression data set, the $\Delta\log_2$ values of the treated (mean of expression values of all replicates for 0.5 and 1 h) and untreated (0 h) samples for each ecotype were computed, resulting in 22,810 profiles consisting of seven values each. The Pearson r value between each of the $\Delta\log_2$ profiles of the 22,810 genes and the $\Delta\log_2$ profile of the physiological IAA responses was computed. The $\Delta\log_2$ profiles of the 10 genes with highest and lowest correlation coefficients are presented in Supplemental Figure A.15 online. Genes with $r > 0.8$ and $r < -0.8$ were considered to be positively and negatively correlated, respectively. Both groups were subjected to Gene Ontology term enrichment analysis using the AmiGO tool (`http://amigo.geneontology.org/cgi-bin/amigo/term_enrichment`).

3.5.13. Sequence analysis of signaling genes

For sequence analysis of auxin signaling genes, the respective gene fragments from the 18 accessions used in the physiological growth assays were sequenced (all except Van-0). The available Col-0 reference sequence was also included in all subsequent analyses. DNA was extracted from leaf tissue. Primers were designed on the basis of the Col-0 reference genome (see Supplemental Table A.1 online) for *TIR1/AFB*s, *Aux/IAA*s, or *ARF*s, resulting in 49

genes (see Supplemental Table A.1 online). Sequences of ∼1-kb PCR products were generated on an ABI 3730 XL (Applied Biosystems) automated sequencer in collaboration with the Leibniz Institute of Plant Genetics and Crop Research in Gatersleben, Germany. Fragments were sequenced in both directions. All sequences and polymorphisms were validated by visual inspection of the chromatograms and edited where appropriate. Alignments were performed with BioEdit version 7.0.5 software (Hall, 1999). Nucleotide diversity for all sites (π), synonymous sites (π_s), and nonsynonymous sites (π_a) was calculated in DnaSP 5.1 (Librado et al., 2009) after Nei (1987). Heterozygous sites were treated as missing data. The 19 accessions included in this study are part of a set of 96 accessions that have been extensively characterized for polymorphism in 876 genomic fragments (Nordborg et al., 2005). The sequences of 236 fragments were extracted with a minimum of 400-bp coding sequence from this data set for the 19 accessions. This empirical distribution was aligned and analyzed in exactly the same manner as the auxin signaling genes. Distributions of nucleotide diversity summary statistics calculated for the auxin signaling genes were then compared with the empirical distributions of the control genes by performing Mann-Whitney U tests in R.

3.5.14. Accession numbers

All microarray data from this article are publicly available at the Gene Expression Omnibus under accession number GSE18975. Sequence data from this article have been submitted to GenBank (accession numbers GU348425-GU348653 and HM487319-HM487971). Arabidopsis Genome Initiative locus identifiers of individual genes analyzed in this article are listed in Supplemental Tables A.1,A.2, and A.3 online.

3.6. Acknowledgments

We are grateful to Tom Guilfoyle for providing the *DR5:GUS* construct, Renate Schmidt for sequencing support, Kathrin Denk for technical assistance, and Jan Grau, Jens Keilwagen, and Steffen Neumann for valuable discussions. We thank Claus Wasternack and Jerry Cohen for critical reading of the manuscript and anonymous reviewers for constructive comments. Furthermore, M.Q. thanks Bill Gray for support in the initial phase of the project. Our work was supported by a grant from the Exzellenznetzwerk Biowissenschaften "Structures and Mechanisms of Biological Information Processing" funded by the Federal State of Sachsen-Anhalt to M.Q.

3.7. References

Alonso-Blanco, C., Aarts, M. G., Bentsink, L., Keurentjes, J. J., Reymond, M., Vreugdenhil, D., and Koornneef, M. (2009). What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? *The Plant Cell*, 21 (7), pp. 1877–1896.

Borevitz, J. O., Hazen, S. P., Michael, T. P., Morris, G. P., Baxter, I. R., Hu, T. T., Chen, H., Werner, J. D., Nordborg, M., Salt, D. E., Kay, S. A., Chory, J., Weigel, D., Jones, J. D. G., and Ecker, J. R. (2007). Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 104 (29), pp. 12057–12062.

Chapman, E. J. and Estelle, M. (2009). Mechanism of Auxin-Regulated Gene Expression in Plants. *Annual Review of Genetics*, 43 (1). PMID: 19686081, pp. 265–285.

Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T. T., Fu, G., Hinds, D. A., Chen, H., Frazer, K. A., Huson, D. H., Schölkopf, B., Nordborg, M., Rätsch, G., Ecker, J. R., and Weigel, D. (2007). Common Sequence Polymorphisms Shaping Genetic Diversity in *Arabidopsis thaliana*. *Science*, 317 (5836), pp. 338–342.

Clough, S. J. and Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal*, 16 (6), pp. 735–743.

Covington, M. F. and Harmer, S. L. (2007). The Circadian Clock Regulates Auxin Signaling and Responses in *Arabidopsis*. *PLoS Biology*, 5 (8), e222.

Czechowski, T., Bari, R. P., Stitt, M., Scheible, W.-R., and Udvardi, M. K. (2004). Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *The Plant Journal*, 38 (2), pp. 366–379.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W.-R. (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in *Arabidopsis*. *Plant Physiology*, 139 (1), pp. 5–17.

Delker, C., Raschke, A., and Quint, M. (2008). Auxin dynamics: the dazzling complexity of a small molecule's message. *Planta*, 227 (5), pp. 929–941.

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005a). The F-box protein TIR1 is an auxin receptor. *Nature*, 435 (7041), pp. 441–445.

Dharmasiri, N., Dharmasiri, S., Weijers, D., Lechner, E., Yamada, M., Hobbie, L., Ehrismann, J. S., Jürgens, G., and Estelle, M. (2005b). Plant Development Is Regulated by a Family of Auxin Receptor F-Box Proteins. *Developmental Cell*, 9 (1), pp. 109–119.

Fu, J., Keurentjes, J. J. B., Bouwmeester, H., America, T., Verstappen, F. W. A., Ward, J. L., Beale, M. H., Vos, R. C. H. de, Dijkstra, M., Scheltema, R. A., Johannes, F., Koornneef, M., Vreugdenhil, D., Breitling, R., and Jansen, R. C. (2009). System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nature Genetics*, 41 (2), pp. 166–167.

Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., Ogawa, M., Yamauchi, Y., Preston, J., Aoki, K., Kiba, T., Takatsuto, S., Fujioka, S., Asami, T., Nakano, T., Kato, H., Mizuno, T., Sakakibara, H., Yamaguchi, S., Nambara, E., Kamiya, Y., Takahashi, H., Hirai, M. Y., Sakurai, T., Shinozaki, K., Saito, K., Yoshida, S., and Shimada, Y. (2008). The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *The Plant Journal*, 55 (3), pp. 526–542.

Gray, W. M., Östin, A., Sandberg, G., Romano, C. P., and Estelle, M. (1998). High temperature promotes auxin-mediated hypocotyl elongation in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 95 (12), pp. 7197–7202.

Guilfoyle, T. J., Ulmasov, T., and Hagen, G. (1998). The ARF family of transcription factors and their role in plant hormone-responsive transcription. *Cellular and Molecular Life Sciences*, 54 (7), pp. 619–627.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, pp. 95–98.

Katagiri, F. and Glazebrook, J. (2003). Local Context Finder (LCF) reveals multidimensional relationships among mRNA expression profiles of *Arabidopsis* responding to pathogen infection. *Proceedings of the National Academy of Sciences*, 100 (19), pp. 10842–10847.

Kepinski, S. and Leyser, O. (2005). The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature*, 435 (7041), pp. 446–451.

Keurentjes, J. J. B., Fu, J., Terpstra, I. R., Garcia, J. M., Ackerveken, G. van den, Snoek, L. B., Peeters, A. J. M., Vreugdenhil, D., Koornneef, M., and Jansen, R. C. (2007). Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proceedings of the National Academy of Sciences*, 104 (5), pp. 1708–1713.

Kliebenstein, D. J., West, M. A. L., Leeuwen, H. van, Kim, K., Doerge, R. W., Michelmore, R. W., and St. Clair, D. A. (2006). Genomic Survey of Gene Expression Diversity in *Arabidopsis thaliana*. *Genetics*, 172 (2), pp. 1179–1189.

Kreitman, M. (2000). Methods to detect selection in populations with application to the human. *Annual Review of Genomics and Human Genetics*, 1 (1). PMID: 11701640, pp. 539–559.

Laan, M. J. van der and Pollard, K. S. (2003). A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap. *Journal of Statistical Planning and Inference*, 117 (2), pp. 275–303.

Leeuwen, H. van, Kliebenstein, D. J., West, M. A., Kim, K., Poecke, R. van, Katagiri, F., Michelmore, R. W., Doerge, R. W., and St.Clair, D. A. (2007). Natural Variation among *Arabidopsis thaliana* Accessions for Transcriptome Response to Exogenous Salicylic Acid. *The Plant Cell*, 19 (7), pp. 2099–2110.

Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25 (11), pp. 1451–1452.

Lincoln, C., Britton, J. H., and Estelle, M. (1990). Growth and development of the axr1 mutants of *Arabidopsis*. *The Plant Cell*, 2 (11), pp. 1071–80.

Lokerse, A. S. and Weijers, D. (2009). Auxin enters the matrix–assembly of response machineries for specific outputs. *Current Opinion in Plant Biology*, 12 (5), pp. 520–526.

Maloof, J. N., Borevitz, J. O., Dabi, T., Lutes, J., Nehring, R. B., Redfern, J. L., Trainer, G. T., Wilson, J. M., Asami, T., and Berry, C. C. (2001). Natural variation in light sensitivity of *Arabidopsis*. *Nature Genetics*, (4), pp. 441–446.

Nagpal, P., Walker, L. M., Young, J. C., Sonawala, A., Timpte, C., Estelle, M., and Reed, J. W. (2000). AXR2 Encodes a Member of the Aux/IAA Protein Family. *Plant Physiology*, 123 (2), pp. 563–574.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering*, 104 (1), pp. 34–41.

Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraishi, T., Iwabuchi, M., and Narusaka, Y. (2009). RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *The Plant Journal*, 60 (2), pp. 218–226.

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., Voinnet, O., and Jones, J. D. G. (2006). A Plant miRNA Contributes to Antibacterial Resistance by Repressing Auxin Signaling. *Science*, 312 (5772), pp. 436–439.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M. and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76 (10), pp. 5269–5273.

Nemhauser, J. L., Hong, F., and Chory, J. (2006). Different Plant Hormones Regulate Similar Processes through Largely Nonoverlapping Transcriptional Responses. *Cell*, 126 (3), pp. 467–475.

Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N. A., Shah, C., Wall, J. D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M., and Bergelson, J. (2005). The Pattern of Polymorphism in *Arabidopsis thaliana*. *PLoS Biology*, 3 (7), e196.

Okushima, Y., Overvoorde, P. J., Arima, K., Alonso, J. M., Chan, A., Chang, C., Ecker, J. R., Hughes, B., Lui, A., Nguyen, D., Onodera, C., Quach, H., Smith, A., Yu, G., and Theologis, A. (2005). Functional Genomic Analysis of the AUXIN RESPONSE FACTOR Gene Family Members in *Arabidopsis thaliana*: unique and Overlapping Functions of ARF7 and ARF19. *The Plant Cell*, 17 (2), pp. 444–463.

Opgen-Rhein, R. and Strimmer, K. (2007). Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach. *Statistical Applications in Genetics and Molecular Biology*, 6 (1), pp. 477+.

Overvoorde, P. J., Okushima, Y., Alonso, J. M., Chan, A., Chang, C., Ecker, J. R., Hughes, B., Liu, A., Onodera, C., Quach, H., Smith, A., Yu, G., and Theologis, A. (2005). Functional Genomic Analysis of the AUXIN/INDOLE-3-ACETIC ACID Gene Family Members in *Arabidopsis thaliana*. *The Plant Cell*, 17 (12), pp. 3282–3300.

Paponov, I. A., Paponov, M., Teale, W., Menges, M., Chakrabortee, S., Murray, J. A. H., and Palme, K. (2008). Comprehensive Transcriptome Analysis of Auxin Responses in *Arabidopsis*. *Molecular Plant*, 1 (2), pp. 321–337.

Parry, G., Calderon-Villalobos, L. I., Prigge, M., Peret, B., Dharmasiri, S., Itoh, H., Lechner, E., Gray, W. M., Bennett, M., and Estelle, M. (2009). Complex regulation of the TIR1/AFB family of auxin receptors. *Proceedings of the National Academy of Sciences*, 106 (52), pp. 22540–22545.

Parry, G. and Estelle, M. (2006). Auxin receptors: a new role for F-box proteins. *Current Opinion in Cell Biology*, 18 (2), pp. 152–156.

Pérez-Torres, C.-A., López-Bucio, J., Cruz-Ramírez, A., Ibarra-Laclette, E., Dharmasiri, S., Estelle, M., and Herrera-Estrella, L. (2008). Phosphate Availability Alters Lateral Root Development in *Arabidopsis* by Modulating Auxin Sensitivity via a Mechanism Involving the TIR1 Auxin Receptor. *The Plant Cell*, 20 (12), pp. 3258–3272.

Ploense, S. E., Wu, M.-F., Nagpal, P., and Reed, J. W. (2009). A gain-of-function mutation in IAA18 alters *Arabidopsis* embryonic apical patterning. *Development*, 136 (9), pp. 1509–1517.

Poecke, R. M. van, Sato, M., Lenarz-Wyatt, L., Weisberg, S., and Katagiri, F. (2007). Natural Variation in RPS2-Mediated Resistance among *Arabidopsis* Accessions: correlation between Gene Expression Profiles and Phenotypic Responses. *The Plant Cell*, 19 (12), pp. 4046–4060.

Quint, M., Barkawi, L. S., Fan, K.-T., Cohen, J. D., and Gray, W. M. (2009). *Arabidopsis* IAR4 Modulates Auxin Response by Regulating Auxin Homeostasis. *Plant Physiology*, 150 (2), pp. 748–758.

Quint, M. and Gray, W. M. (2006). Auxin signaling. *Current Opinion in Plant Biology*, 9 (5), pp. 448–453.

Ramos, J. A., Zenser, N., Leyser, O., and Callis, J. (2001). Rapid Degradation of Auxin/Indoleacetic Acid Proteins Requires Conserved Amino Acids of Domain II and Is Proteasome Dependent. *The Plant Cell*, 13 (10), pp. 2349–2360.

Rowe, H. C. and Kliebenstein, D. J. (2008). Complex Genetics Control Natural Variation in *Arabidopsis thaliana* Resistance to *Botrytis cinerea*. *Genetics*, 180 (4), pp. 2237–2250.

Ruegger, M., Dewey, E., Gray, W. M., Hobbie, L., Turner, J., and Estelle, M. (1998). The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast Grr1p. *Genes & Development*, 12 (2), pp. 198–207.

Salmon, J., Ramos, J., and Callis, J. (2008). Degradation of the auxin response factor ARF1. *The Plant Journal*, 54 (1), pp. 118–128.

Stavang, J. A., Gallego-Bartolomé, J., Gómez, M. D., Yoshida, S., Asami, T., Olsen, J. E., García-Martínez, J. L., Alabadí, D., and Blázquez, M. A. (2009). Hormonal regulation of temperature-induced growth in *Arabidopsis*. *The Plant Journal*, 60 (4), pp. 589–601.

Stomp, A.-M. (1991). Histochemical localization of β-glucuronidase. In: *GUS Protocols*. Ed. by S. Gallagher. London: Academic Press, pp. 103–113.

Teale, W. D., Paponov, I. A., and Palme, K. (2006). Auxin in action: signalling, transport and the control of plant growth and development. *Nature Reviews Molecular Cell Biology*, 7 (11), pp. 847–859.

Tian, Q. and Reed, J. (1999). Control of auxin-regulated root development by the *Arabidopsis thaliana* SHY2/IAA3 gene. *Development*, 126 (4), pp. 711–721.

Tiwari, S. B., Hagen, G., and Guilfoyle, T. (2003). The Roles of Auxin Response Factor Domains in Auxin-Responsive Transcription. *The Plant Cell*, 15 (2), pp. 533–543.

Ulmasov, T., Murfett, J., Hagen, G., and Guilfoyle, T. J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell*, 9 (11), pp. 1963–71.

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. (1999). Activation and repression of transcription by auxin-response factors. *Proceedings of the National Academy of Sciences*, 96 (10), pp. 5844–5849.

Zenser, N., Ellsmore, A., Leasure, C., and Callis, J. (2001). Auxin modulates the degradation rate of Aux/IAA proteins. *Proceedings of the National Academy of Sciences*, 98 (20), pp. 11795–11800.

Zhang, X., Shiu, S., Cal, A., and Borevitz, J. O. (2008). Global Analysis of Genetic, Epigenetic and Transcriptional Polymorphisms in *Arabidopsis thaliana* Using Whole Genome Tiling Arrays. *PLoS Genetics*, 4 (3), e1000032.

PLoS ONE, Volume 8, Issue 11, November 2013

4. Optimized probe masking for comparative transcriptomics of closely related species

Yvonne Poeschl[1], Carolin Delker[2], Jana Trenner[2], Kristian Karsten Ullrich[2], Marcel Quint[2], Ivo Grosse[1,3]

[1] Martin Luther University Halle–Wittenberg, Institute of Computer Science, 06099 Halle (Saale), Germany
[2] Leibniz Institute of Plant Biochemistry, Department of Molecular Signal Processing, 06120 Halle (Saale), Germany
[3] German Center of Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, 04103 Leipzig, Germany

4.1. Abstract

Microarrays are commonly applied to study the transcriptome of specific species. However, many available microarrays are restricted to model organisms, and the design of custom microarrays for other species is often not feasible. Hence, transcriptomics approaches of non-model organisms as well as comparative transcriptomics studies among two or more species often make use of cost-intensive RNAseq studies or alternatively, by hybridizing transcripts of a query species to a microarray of a closely related species. When analyzing these cross-species microarray expression data, differences in the transcriptome of the query species can cause problems, such as: (i) lower hybridization accuracy of probes due to mismatches or deletions, (ii) probes binding multiple transcripts of different genes, and (iii) probes binding transcripts of non-orthologous genes. So far, methods for (i) exist, but these neglect (ii) and (iii).

Here, we propose an approach for comparative transcriptomics addressing problems (i) to (iii), which retains only transcript-specific probes binding transcripts of orthologous genes. We apply this approach to an *Arabidopsis lyrata* expression data set measured on a microarray designed for *Arabidopsis thaliana*, and compare it to two alternative approaches, a *sequence-based* approach and a genomic DNA *hybridization-based* approach. We investigate the number of retained probe sets, and we validate the resulting expression responses by qRT-PCR. We find that the proposed approach combines the benefit of sequence-based stringency and accuracy while allowing the expression analysis of much more genes than the alternative sequence-based approach. As an added benefit, the proposed approach requires probes to detect transcripts of orthologous genes only, which provides a superior base for biological interpretation of the measured expression responses.

4.2. Introduction

While RNAseq approaches gained increased popularity for transcriptome analyses, microarrays are still in use due to their simplicity and easier data processing. In addition, a plethora of tools, comparable data sets, and experiences make microarrays attractive, also in transcriptomics studies of non-model organisms or in comparative transcriptomics studies among different species. However, one problem in cross-species analyses is that microarrays are usually designed for a specific *reference species*. If no specific microarray is available for a *query species*, the microarray of a closely related species can be utilized, as sequences tend to be more similar among closely related species.

Sequence differences in target genes of the query species can cause three problems. First (i), the hybridization signal can be reduced due to mismatches or deletion of the target. Second (ii), the hybridization signal can be increased due to cross hybridization, where probes do not only detect the transcript of the intended target gene but hybridize also to transcripts of other genes (Gilad et al., 2006; Bar-Or et al., 2007; Orlov et al., 2007). And third (iii), probes can target transcripts that are highly similar in the targeted region, but are not products of orthologous genes in the reference and the query species.

Two popular approaches addressing problem (i) are the *sequence-based* approach by Khaitovich et al. (2004) and the genomic DNA *hybridization-based* approach by Hammond et al. (2005), Graham et al. (2007), and Broadley et al. (2008). The sequence-based approach by Khaitovich et al. (2004) uses the transcript of the annotated target gene of the reference species to determine the transcript of the target gene of the query species prior. Afterwards, the sequences of the target transcripts of the query species are compared to the probe sequences of the microarray of the reference species to identify and mask probes that are affected by the mentioned problem (i). The sequence-based approach retains probes that perfectly match the sequences of the target transcripts of the query species.

The hybridization-based approach by Hammond et al. (2005), Graham et al. (2007), and Broadley et al. (2008) hybridizes genomic DNA (gDNA) of the query species to the microarray of the reference species to detect probes affected by problem (i). The hybridization-based approach by Broadley et al. (2008) address problem (i) by masking probes below a given gDNA hybridization intensity value. However, it retains probes that possibly match regions on the genomic DNA outside transcribed regions.

Common to both approaches is that they fail to provide solutions for problem (ii) and (iii), the problems of cross hybridization and transcripts of non-orthologous genes, respectively.

Here, we propose a sequence-based approach similar to Khaitovich et al. (2004), but in contrast to Khaitovich et al. (2004), we account for a slight sequence divergence of the query species to the reference species by allowing probes to match the target transcript with at most one mismatch. We additionally address problems (ii) and (iii) to facilitate reliable comparative transcriptomics analyses.

We apply the proposed approach (*1mm*) to an auxin expression data set of *A. lyrata* measured on the Affymetrix ATH1-121501 microarray designed for *A. thaliana* (Redman et al., 2004)

and compare it to the sequence-based approach (*0mm*) by Khaitovich et al. (2004), the gDNA hybridization-based approach (*gDNA*) by Broadley et al. (2008), and a *naive* approach that uses all probes on the microarray.

We investigate the effect of using probes matching with a single mismatch in contrast to using only perfectly matching probes, and we additionally study the effect of addressing problems (ii) and (iii). Therefore, we compare the number of transcript-specific probe sets retained by each of the three masking approaches and the naive approach. We validate the accuracy of the resulting expression responses for the auxin-treated query species *A. lyrata* by qRT-PCR.

4.3. Methods

4.3.1. 1mm approach

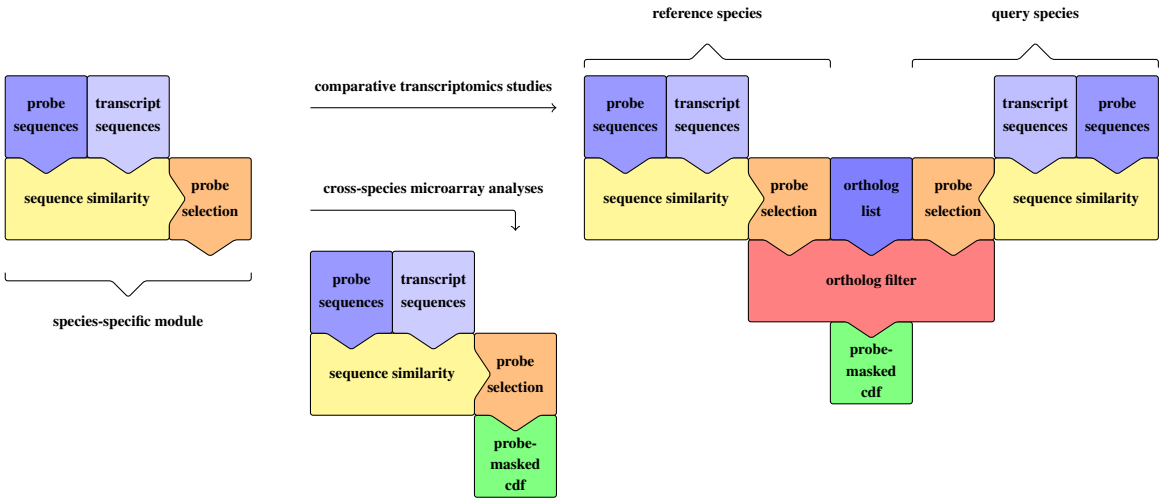


Figure 4.1.: The two possible workflows of the *1mm* approach. The *1mm* approach can be used in two different ways: For cross-species microarray analyses or for comparative transcriptomics studies. Each of the two workflows results in a probe-masked cdf colored in green. The tips of the colored pieces show the flow of information. The blue colored pieces show the input data provided by the user, whereas the yellow, orange, and red pieces show the two or three steps of the *1mm* approach leading to a probe mask. The *species-specific module* consists of the sequence similarity step with the microarray-specific probe sequences and the species-specific transcript sequences as input, and the probe selection step that results in a list of probe sets containing only reliable probes. The *species-specific module* can be used for generating a probe-masked cdf for cross-species microarray analyses of non-model species. Two different *species-specific modules* can be used with an orthologous gene list for generating a probe-masked cdf for comparative transcriptomics studies.

To facilitate a reliable comparative transcriptomics analysis based on microarray hybridization of a reference species and a closely related query species, transcript-specific probes must be separated from probes affected by at least one of the problems (i) to (iii) mentioned in the introduction. We declare a probe to be transcript-unspecific if it is affected by cross hybridization or if it targets transcripts of non-orthologous genes. Our goal is to detect and mask transcript-unspecific probes and probes that target no transcript from subsequent analyses.

We generate a probe mask for comparative transcriptomics in four steps (Figure 4.1): first, we align the sequences of the probes to all known transcripts, including all known *isoforms*, of the reference and the query species to find regions of high similarity. Here and in the following, the term *isoform* refers to *one possible transcript of a gene*. Second, we filter the alignments and mask probes that match no transcripts or are affected by cross-hybridization. Third, we verify if the transcripts targeted by a probe set are transcripts of orthologous genes of the reference and the query species. Fourth, we generate the final probe mask based on the outcome of the previous three steps.

Sequence similarity

To find regions of high similarity, we use *BLASTN-short* of the BLAST+ package (Camacho et al., 2009) for computing an alignment of the sequences of the perfectly matching probes present on the Affymetrix microarray of the reference species to the protein coding transcripts of the reference and query species, respectively. We require BLASTN-short to align all 25 bases of the probe sequences and allow at most one mismatch (perc_identity = 96 and ungapped = 1) to account for sequence variation in the transcriptome of the query species. We record for each probe all matching transcripts. Subsequently, we process the results of the alignment for the reference species and the query species separately but follow the same *probe selection* steps.

Probe selection

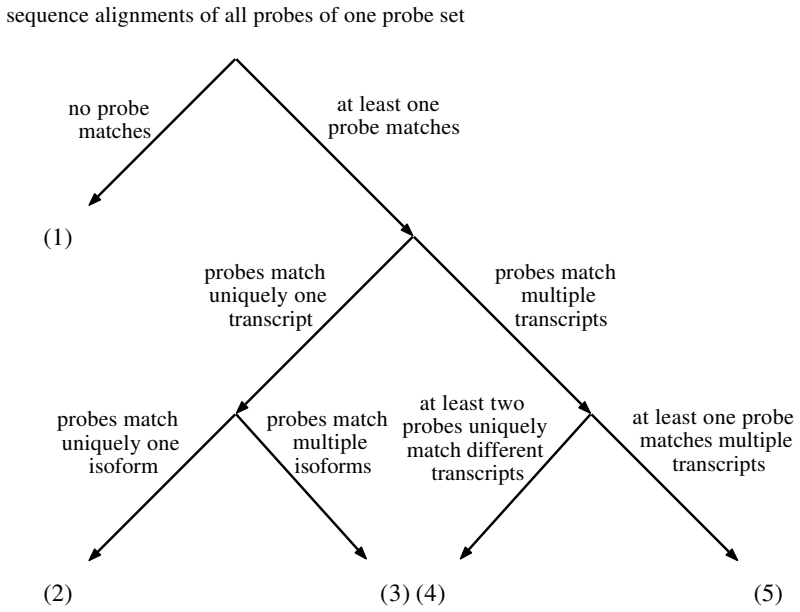


Figure 4.2.: Assignment of a probe set to one of five groups. The assignment of a probe set to a specific group depends on the characteristics of the matching probes in the probe sets. The term *isoform* refers to one possible transcript of a gene.

We subdivide the probe sets into five disjoint groups (Figure 4.2) consisting of probe sets in which: (1) none of the probes match any transcript, (2) all matching probes uniquely match one isoform of one gene, (3) all matching probes uniquely match several isoforms of one gene, (4) at least two of the matching probes uniquely match different transcripts, and (5) at least one of the matching probes matches multiple transcripts. We process the probe sets of the five groups as follows:

(1): Mask all probes.

(2): Mask probes that do not match.

(3): Mask probes that do not match.

(4): Mask probes that do not match any transcript.

(5): Mask probes that do not match any transcript and
mask probes that match multiple transcripts.

We mask probes in probe sets of groups (1) to (5) matching no transcript because they are expected to generate no gene-specific hybridization signals. We compute for each probe set of group (3) the union of probes matching any known isoform of a specific gene. Probe sets of group (4) contain probes that uniquely match different transcripts. For each matching transcript, we process the respective probes according to the rules of groups (2) or (3). Probe sets of group (5) contain at least one probe that matches multiple transcripts and thus is affected by cross hybridization. We mask such probes and process the remaining probes according to the rules of the four previous groups.

As a result of the masking, the number of unmasked probes within a probe set can vary from zero to 11. We mask probe sets containing less than three probes, because we consider these unreliable (Fujimoto et al., 2011). We perform the processing step for the alignment of the reference species and the query species, which results in two species-specific modules, which return two lists of *mappings* of probe sets to genes, one for the reference and one for the query species, respectively (Figure 4.1).

Filtering of orthologous genes

We join both lists of mappings by the probe set names to obtain gene-pair-matchings of the corresponding probe sets. Multiple gene pairs are possible for probe sets belonging to groups (4) or (5), because they can target multiple transcripts. We consider a gene pair orthologous if it is contained in the list of orthologs (Methods List of orthologous genes) of the reference species and the query species.

If a gene pair is orthologous, we compute the intersection of the probes. We mask probes that are not contained in the intersection and thus do not target both transcripts. For probe sets of groups (4) and (5) we keep the orthologous gene pair with the largest number of probes. We again mask a probe set if it contains less than three probes after the intersection.

This filtering step leads to a *final list* of probe sets targeting only transcripts of orthologous genes using the same probes.

Probe masking

We modify the original chip definition file (cdf), which contains the locations of the probes on the respective microarray, based on the final list of probe sets. We create the probe-masked cdf using the function make.cdf.package of the makecdfenv package (Irizarry et al., 2006) to allow using the probe-masked cdf with R (R Development Core Team, 2010) for further analyses. The probe-masked cdf can be used by the function read.affybatch of the affy package (Gautier et al., 2004).

4.3.2. Data sets

Transcript sequences

We obtain the transcript data sets of *A. thaliana* (Swarbreck et al., 2008) and *A. lyrata* (Hu et al., 2011) from Phytozome v7.0 (`http://www.phytozome.com`). These data sets contain sequences of transcripts of 35386 and 32670 protein-coding genes for *A. thaliana* and *A. lyrata*, respectively.

Probe sequences

We obtain the sequences of the probes of the ATH1-121501 microarray from Affymetrix (`http://www.affymetrix.com`). The data set comprises 251078 sequences including 975 sequences of control probes.

Target sequences

We obtain the sequences of the targets of the ATH1-121501 microarray from Affymetrix (`http://www.affymetrix.com`) and proceed according to Khaitovich et al. (2004). The data set comprises 22814 sequences.

List of orthologous genes

We obtain the protein sequences of *A. thaliana* and *A. lyrata* from Phytozome v7.0 (`http://www.phytozome.com`). These data sets contain protein sequences of 35386 and 32670 protein-coding genes for *A. thaliana* and *A. lyrata*, respectively. We generate the list containing orthologous genes of *A. thaliana* and *A. lyrata* using BLASTP (Altschul et al., 1990) setting the maximal E-value to 1e-05 and retaining only the best BLASTP hit.

gDNA hybridization data set

We obtain the .cel file containing the hybridization intensities of the gDNA of *A. lyrata* from `http://affy.arabidopsis.info/xspecies/` and proceed it according to Broadley et al. (2008).

Chip definition file

We obtain the chip definition file (cdf) for the ATH1-121501 microarray from Affymetrix (`http://www.affymetrix.com`). The cdf contains the locations of the PM and the MM probes of the ATH1-121501 microarray which target the 3'-end of *A. thaliana* transcripts.

Expression data set

We obtain a cross-species hybridization data set of *A. thaliana* and *A. lyrata* using the ATH1 microarray from NASC (`http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid=579`). The data set assesses the variation of auxin responses in seven days old *A. thaliana* and *A. lyrata* seedlings. Information on the experimental procedures are provided at
`http://affymetrix.arabidopsis.info/narrays/experimentpage.pl?experimentid=579`.

We load the .cel files into R (R Development Core Team, 2010) using the masked cdfs resulting from the 1mm approach (Methods 1mm approach), the 0mm approach, the gDNA approach, and the non-masked cdf of the naive approach, and the affy package. We perform background correction, quantile normalization, and summary of the expression data using RMA (Irizarry et al., 2003) of the simpleaffy package (Wilson et al., 2005), which returns $\log_2$-transformed expression values.

4.3.3. qRT-PCR analysis

We perform a verification of transcription levels by qRT-PCR, to assess the accuracy of the expression responses resulting from the four studied approaches. Plant material was subjected to the same experimental conditions as described in Methods 4.3.2. $3\,\mu g$ of total RNA was subjected to reverse transcription using the RevertAid First Strand cDNA Synthesis Kit by Fermentas according to the manufacturers description. Power SYBR Green PCR Master Mix (Applied Biosystems) was used for subsequent quantitative real-time PCR analyses. Expression of the PP2A catalytic subunit gene AT1G13320 (array element: 259407_at) served as the constitutively expressed reference gene (Czechowski et al., 2005). Comparative expression levels (CELs) for the respective genes of interest were calculated as
$\Delta\mathrm{Ct} := \mathrm{Ct}^{\text{reference gene}} - \mathrm{Ct}^{\text{gene of interest}}$. Oligonucleotide sequences and a complete list of analyzed genes are presented in Supplementary Table B.3.

Table 4.1.: A table of verified candidate genes.

ae name	locus At	locus Al	probes 1mm	probes 0mm	probes gDNA	ΔΔCt	*1mm	*0mm	*gDNA	*naive	category
245245_at	AT1G44318	314128	110----10-0	--0-----0-0	11-xxxx-0--	-1.64	-1.46	-2.04	-0.22	-0.35	A,B,C,D
245696_at	AT5G04190	939816	0-0-1-001--	0-0---00---	0x0x--00-x-	3.03	1.68	2.53	0.98	0.55	A,B,C,D
246270_at	AT4G36500	490986	10110---110	-0--0-----0	1---0-x-11-	-2.27	-2.18	-2.80	-1.42	-1.69	A,B,C,D
248676_at	AT5G48850	494948	00-11001-10	00---00---0	-0-1-001x10	3.30	1.86	2.77	2.07	1.55	A,B,C,D
251705_at	AT3G56400	486080	00-10--1--0	00--0-----0	0-x10--1-x-	-2.90	-2.46	-2.90	-1.81	-1.36	A,B,C,D
252205_at	AT3G50350	485386	0---0-0-0-1	0---0-0-0--	0x-x0x0x0x1	1.82	1.36	1.65	0.31	0.47	A,B,C,D
252626_at	AT3G44940	484892	-11010-0---	---0-0-0---	--10-0-0---	-1.14	-0.85	-0.93	-1.22	-0.42	A,B,C,D
253287_at	AT4G34270	491240	01100001-00	0--0000--00	0-100-01-00	-0.10	-0.05	-0.10	-0.07	-0.08	A,B,C,D
253908_at	AT4G27260	492072	101-0---011	-0--0---0--	101x0xx-011	3.11	2.73	2.70	2.46	2.42	A,B,C,D
254175_at	AT4G24050	492457	---000100--	---000-00--	-x-00--00xx	-1.14	-1.01	-0.95	-0.60	-0.64	A,B,C,D
255788_at	AT2G33310	482270	1101110-00-	--0---0-00-	11011-0-00x	2.46	2.09	2.22	2.01	1.94	A,B,C,D
256131_at	AT1G13600	920239	1--101-0100	----0--0-00	1--1-1-0-00	-1.21	-0.83	-0.82	-1.03	-0.61	A,B,C,D
257153_at	AT3G27220	936451	11--00010--	----000-0--	-1xx-00-0-x	-3.95	-3.80	-4.21	-3.65	-3.24	A,B,C,D
259407_at	AT1G13320	920212	-000000-000	-000000-000	-000000-000	-0.35	0.10	0.09	0.10	0.10	A,B,C,D
260904_at	AT1G02450	470205	-0--0011110	-0--00----0	x-xx-0111--	-2.13	-1.29	-1.05	-0.83	-0.61	A,B,C,D
261892_at	AT1G80840	477161	--0000-1100	--0000---00	x-0000x--0-	-2.40	-2.27	-2.55	-2.05	-1.56	A,B,C,D
263970_at	AT2G42850	346095	0100001-1-1	0-0000-----	0-00-01---1	-1.67	-1.11	-0.82	-1.16	-0.97	A,B,C,D
264867_at	AT1G24150	313260	010001-1-0-	0-000----0-	-1--01---0x	-1.37	-1.63	-1.85	-0.92	-1.27	A,B,C,D
265452_at	AT2G46510	483808	001--00----	00-x-00----	00---0-xx--	-1.89	-1.43	-1.35	-0.26	-0.20	A,B,C,D
265856_at	AT2G42430	935111	000--010-10	000--0-0--0	-00xx0-0-10	1.98	1.65	2.03	0.99	0.97	A,B,C,D
245336_at	AT4G16515	493225	-111--1--11	----------	-1-1-x1xx11	2.76	2.16		0.97	0.70	C,D
245369_at	AT4G15975	329916	-0--1----1-	----------	--xx-x-----	-2.48	-1.64		-0.13	-0.15	C,D
245397_at	AT4G14560	946923	-1----110--	----------	x1-x-x11-x-	2.74	2.37		1.44	1.36	C,D
246993_at	AT5G67450	496850	0---01----1	----------	0-x-0---xx-	-2.57	-0.80		-1.10	-0.29	C,D
247524_at	AT5G61440	496303	--01--01--1	----------	--01--01xx1	-3.48	-2.49		-1.22	-0.77	C,D
248858_at	AT5G46630	948276	1--010-----	----------	1--010xx-xx	-0.10	0.28		0.25	0.14	C,D
250937_at	AT5G03230	939701	10--110---1	----------	10--1-0xxx1	-2.64	-2.74		-2.05	-1.94	C,D
251910_at	AT3G53810	485775	-1---0---0-	----------	x1xxx--xx0-	-1.31	-1.10		-0.28	-0.24	C,D
253400_at	AT4G32860	491410	--11-010--1	----------	x-11x0-0xx1	-1.42	-0.91		-0.09	-0.12	C,D
253959_at	AT4G26410	945436	1011----01-	----------	--11-xxx0--	-0.27	0.07		0.08	0.00	C,D
261766_at	AT1G15580	471758	--1--0-111-	----------	x-1xx0x---x	4.46	1.54		0.38	0.61	C,D
262085_at	AT1G56060	474673	---1----100	----------	xxx-x-xx100	1.70	1.67		0.38	0.15	C,D
265256_at	AT2G28390	481666	----1-11100	----------	--xx--11-0-	0.10	0.11		-0.01	0.14	C,D
266649_at	AT2G25810	932757	11-01----1-	----------	11--1---x-x	-0.68	-0.44		-1.05	-0.72	C,D
266820_at	AT2G44940	483623	1-0-111-11-	----------	--0-11-x-1-	-2.24	-1.44		-1.73	-0.87	C,D
266974_at	AT2G39370	482956	--11-1-1011	----------	--1-x--1011	4.02	0.85		1.56	0.67	C,D
254761_at	AT4G13195	333009	--0----1--0	----------	----------	2.22	1.70			0.33	D
265806_at	AT2G18010	931672	1111---100-	----------	----------	3.96	0.62			0.53	D
247215_at	AT5G64905	951330	--000------	--000------	----------	-4.60	-1.98	-1.85		-0.03	B, D
248539_at	AT5G50130	495070	--01--00---	--0---00---	----------	2.03	1.03	1.61		0.19	B, D

*:$\Delta\mu$, ae: array element, At: *Arabidopsis thaliana*, Al: *Arabidopsis lyrata*
For each gene the corresponding array element name and the orthologous gene pair (locus *A. thaliana* by the TAIR id and locus *A. lyrata* by the Phytozome gene id) are listed. Additionally, the composition of the probe set in the 1mm mask, the 0mm mask, and the gDNA mask are shown. Originally, each probe set consists of 11 probes. The "-" represents a masked probe, "0" a perfectly matching probe, "1" a probe matching with one mismatch, and "x" represents a transcript-unspecific probe. The ΔΔCt labeled column contains the Δ expression values of the 1h treatment versus no treatment experiments derived from qRT-PCR. The next four columns contain the $\Delta\mu$ expression values of the 1h treatment versus no treatment experiments derived from the three probe masking approaches and the non-masking approach. The last column contains the category used for computation of the Pearson correlation coefficient.

4.3.4. Candidate selection

We choose 40 genes as candidate genes for verification by qRT-PCR based on the response to auxin treatment and the composition of the corresponding probe sets (Table 4.1 and Supplementary Table B.4). The number of probes per probe set ranges from three to ten, and the number of imperfectly matching probes with a single mismatch ranges from zero to all. We choose these 40 candidate genes from four categories:

(A): 20 candidate genes present in all four approaches.

(B): 22 candidate genes: 20 genes of category (A) and 2 candidate genes present in the naive, 0mm, and 1mm approaches.

(C): 36 candidate genes: 20 genes of category (A) and 16 candidate genes present in the naive, gDNA, and 1mm approaches.

(D): 40 candidate genes: 20 genes of category (A) and 20 candidate genes present in the naive and 1mm approaches.

We choose genes of categories (A) and (B) for studying the impact of using probes with a single mismatch and removing probes affected by cross hybridization. We use genes of category (C) for studying the effect of using probes with a single mismatch and removing probes affected by cross hybridization on a larger set of genes, which contains 16 genes that are not retained by the 0mm approach. We use genes of category (D) for studying the overall performance of the 1mm approach.

4.3.5. Correlation analysis

We compute the mean $\log_2$ expression values of the 1 hour post auxin treatment samples and control samples (n=3 biological replicates) for each of the 40 candidate genes, for the $\log_2$ expression values resulting from the three masking approaches and the non-masking approach. We compute the log-fold changes (responses), which are the differences of the mean $\log_2$ expression values of treated ($\mu_{\text{treatment}}$) and control (μ_{control}) samples as $\Delta\mu := \mu_{\text{treatment}} - \mu_{\text{control}}$. Similarly, we calculate the $\Delta\Delta\text{Ct} := \Delta\text{Ct}_{\text{treatment}} - \Delta\text{Ct}_{\text{control}}$ values of the comparative expression levels produced by qRT-PCR (Methods qRT-PCR analysis) of all candidate genes. We compute Pearson, Spearman, and Kendall correlation coefficients for all four candidate gene categories between the log-fold changes $\Delta\mu$ resulting from each approach and the $\Delta\Delta$Ct values resulting from qRT-PCR.

4.3.6. Source code

Source code is available at `http://sourceforge.net/projects/probemaskingpipeline` online.

4.4. Results and discussion

For a reliable comparative transcriptomics analysis of a reference species and a closely related query species based on microarray hybridization, transcript-specific probes must be separated from (i) probes matching no transcripts in at least one of the species, and transcript-unspecific probes that (ii) are affected by cross hybridization when they target multiple transcripts or (iii) target transcripts of non-orthologous genes.

Current approaches address these problems only partially. While hybridization-based techniques fail to address any of the problems (i) to (iii) in a specific manner, they have the benefit of usually allowing the analysis of a large set of transcripts. Sequence-based approaches, so far, offer relatively high stringency and specificity since only perfectly matching probes are retained in the analyses. This usually results in a high loss of genes for subsequent analyses since minor changes in sequences are frequent even among closely related species. Furthermore, the issue of gene orthology has been neglected in the masking approaches, so far.

Orthologous genes are relevant in comparative transcriptomics analyses, because they are derived from a common ancestor. Keeping the focus of the analysis on orthologous genes provides a solid base for biological interpretation of the expression data.

The goal of the 1mm approach (Methods 1mm approach) is to mask transcript-unspecific probes and probes matching no transcripts, and to keep only transcript-specific probes that target transcripts of orthologous genes. We permit probes to match transcripts with at most one mismatch in order to account for a possible sequence divergence between the query species and the reference species.

We apply the 1mm approach, the 0mm approach described by Khaitovich et al. (2004), the gDNA approach described by Broadley et al. (2008), and the naive approach to *A. thaliana* as reference species and its closely related sister species *A. lyrata* as query species based on the Affymetrix ATH1-121501 microarray designed for *A. thaliana* (Redman et al., 2004). The naive approach uses all probes of all probe sets as originally designed by Redman et al. (2004).

We require the probe sets of all three masking approaches to contain at least three probes to enhance the reliability of the expression values for a gene as previously proposed by Fujimoto et al. (2011).

To assess the performance of the different masking approaches and the non-masking approach, we analyze gene expression data in response to an auxin stimulus for the query species *A. lyrata*, determined by hybridization to an ATH1-121501 microarray.

Auxin is a plant hormone that is involved in virtually all aspects of plant development and is known to induce rapid transcriptome changes as part of its primary signaling response (Kleine-Vehn et al., 2006; Delker et al., 2008).

First, we compare the four approaches with respect to the number of retained probe sets. Second, we perform qRT-PCR experiments for 40 genes, and we compare the four approaches with respect to the Pearson correlation coefficients of the resulting microarray data with

the qRT-PCR data. While mismatches can affect hybridization intensities, we show that the tolerance of one mismatch per probe in the proposed approach accurately detects gene expression changes in response to an external treatment of the query species.

4.4.1. Number and composition of probe sets

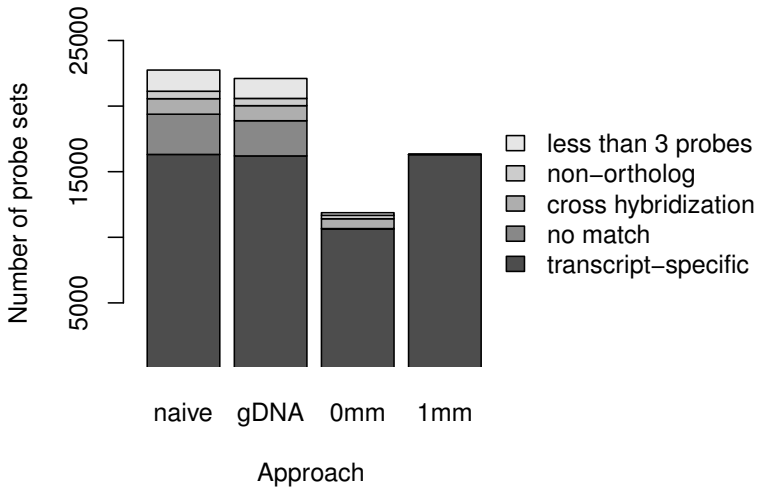


Figure 4.3.: Number of probe sets obtained by the three masking approaches and the naive approach. The height of each bar shows the number of probe sets falling in one of the following categories: *transcript-specific*: retained probe sets targeting orthologs, not affected by cross hybridization, and containing at least 3 probes; *no match*: probe sets matching no transcript in *A. thaliana* or *A. lyrata*; *cross hybridization*: probe sets affected by cross hybridization; *non-ortholog*: probe sets targeting non-orthologs, and *less than 3 probes*: probe sets containing less than 3 matching probes in the 1mm approach but at least 3 probes in the other approach. The naive approach, the gDNA approach, and the 1mm approach retain approximately 16000 transcript-specific probe sets, and the 0mm approach retains approximately 10500 transcript-specific probe sets.

The ATH1-121501 microarray represents approximately 24000 *A. thaliana* genes by 22746 probe sets, which are all contained in the naive approach. 22105 probe sets are retained by the gDNA approach, while 11873 and 16315 probe sets are retained by the 0mm and the 1mm approach, respectively (Supplementary Figure B.7). Depending on the respective masking approach, retained probe sets can be transcript-specific, transcript-unspecific, can match no transcript or contain less than three probes (Figure 4.3 and Supplementary Table B.1).

First, we consider transcript-specific probe sets, which contain at least three probes that uniquely target transcripts of orthologs as these would represent the genes most relevant in any comparative transcriptomics approach. We find 16315 transcript-specific probe sets retained by the naive approach, 16202 retained by the gDNA approach, 10629 retained by the 0mm approach, and 16315 retained by the 1mm approach. The naive, the gDNA, and the 1mm approach yield approximately the same number of transcript-specific probe sets. These three approaches retain approximately 5500 more transcript-specific probe sets than the 0mm approach, because the 0mm approach only retains perfectly matching probes.

Second, we consider transcript-unspecific probe sets, which contain probes that target multiple transcripts or transcripts of non-orthologs. These probe sets would likely result in biased expression values or artifacts and would be undesired in any transcriptomics analysis. Approximately 1700 of the retained probe sets of the naive and the gDNA approach, and approximately 1000 of the 0mm approach are transcript-unspecific, which comprise approximately 8 % of the retained probe sets, respectively. Furthermore, two thirds of the transcript-unspecific probe sets are affected by cross hybridization and one third of the transcript-unspecific probe sets target transcripts of non-orthologous genes in each of the three approaches.

Third, we consider the probe sets that match no transcript in any of the two species with the 1mm approach. We find that approximately 3000 of the retained probe sets of the naive approach, approximately 2700 of the gDNA approach, and approximately 30 of the 0mm approach match no transcript. This indicates that approximately 12 % of the retained probe sets of the naive and the gDNA approach, and that 0.3 % of the retained probe sets of the 0mm approach match no transcript. In case of the gDNA approach, this may be caused by the possibility that probes target regions on the genomic DNA outside transcribed regions. The 0mm approach retains only a few probe sets whose probes match no transcript in the 1mm approach, because in the 0mm approach probes are checked to be similar to *A. lyrata* but not to *A. thaliana*. Thus, these probes are unspecific for *A. thaliana* and would be uninformative in comparative transcriptomics analysis.

Fourth, we consider those probe sets that contain less than three probes in the 1mm approach after masking of probes matching no transcripts or multiple transcripts, but contain at least three probes in the other approaches. We find that approximately 1600 of the retained probe sets of the naive and the gDNA approach, and that approximately 200 of the retained probe sets of the 0mm approach contain at least three probes. This states that approximately 7 % of the retained probe sets by the naive and the gDNA approach, and that approximately 2 % of the retained probe sets of the 0mm approach contain at least three probes, whereas they contain less than three probes in the 1mm approach. Again, for the gDNA approach probes of these probe sets possibly target regions on the genomic DNA outside transcribed regions. And again, these probe sets could result in biased expression values that are undesired in any transcriptomics analysis.

The 1mm approach efficiently masks probes matching no or multiple transcripts, and probes matching transcripts of non-orthologs. Due to the tolerance of probes with a single mismatch, the number of transcript-specific probe sets retained by the 1mm approach is similar to that of the gDNA approach and increases from 10629 to 16315 compared to the 0mm approach (Figure 4.3 and Supplementary Table B.1).

4.4.2. qRT-PCR verification

To evaluate the quality of the three masking approaches and the naive approach, we perform qRT-PCR experiments for 40 *A. lyrata* genes (Table 4.2 and Supplementary Figure B.6). We apply the four respective approaches and subsequently compute the Pearson correlation coefficients c (Arikawa et al., 2008) of the auxin induced log-fold changes ($\Delta\mu$) and the $\Delta\Delta$Ct values obtained by qRT-PCR of an independent experiment.

Table 4.2.: qRT-PCR verification of masked and non-masked microarray data.

category	naive	gDNA	0mm	1mm
(A)	0.91	0.93	0.98	0.98
(B)	0.82		0.95	0.96
(C)	0.83	0.87		0.94
(D)	0.78			0.92

Pearson correlation coefficients of (i) the $\Delta\mu$ expression responses resulting from the three masking approaches and the naive approach, and (ii) the $\Delta\Delta$Ct expression responses resulting from qRT-PCR of the genes of category A, B, C, and D (Methods Candidate selection). We find that the 1mm approach and the 0mm approach yield similar Pearson correlation coefficients that are higher than those of the gDNA approach and the naive approach.

First, we consider category (A), which contains 20 genes that are present in all three masking approaches and the naive approach. We find Pearson correlation coefficients of $c = 0.91$ for the naive approach, $c = 0.93$ for the gDNA approach, $c = 0.98$ for the 0mm approach, and $c = 0.98$ for the 1mm approach (Table 4.2 and Supplementary Table B.2). Hence, the sequence-based approaches (0mm and 1mm) yield more accurate expression response values than the naive and the gDNA approach for this category. Although the 1mm approach permits single mismatches and the more stringent 0mm approach does not, both approaches yield similarly high Pearson correlation coefficients.

Second, we consider category (B), which contains 22 genes that are present in the naive, the 0mm, and the 1mm approach, and we find Pearson correlation coefficients of $c = 0.82$ for the naive approach, $c = 0.95$ for the 0mm approach, and $c = 0.96$ for the 1mm approach. Again, both sequence-based approaches yield similar, but higher Pearson correlation coefficients than the naive approach.

The similar Pearson correlation coefficients indicate that, despite probes matching with one mismatch can have a reduced hybridization efficacy (Supplementary Figures B.1 and B.2) (Gilad et al., 2006; Naiser et al., 2008; Dannemann et al., 2009), the accuracy of the log-fold changes ($\Delta\mu$) is not reduced by using probes matching with a single mismatch (Supplementary Figure B.3). To account for the reduced hybridization efficacy of probes matching with one mismatch, we suggest a correction approach using a fourth-degree polynomial, which corrects the nominal expression values according to the positional effect of the respective mismatch but does not have a significant effect on the log-fold changes (Supplementary Figures B.4 and B.5).

Third, we consider category (C), which contains 36 genes that are present in the naive, the gDNA, and the 1mm approach, and we find Pearson correlation coefficients of $c = 0.83$ for the naive approach, $c = 0.87$ for the gDNA approach, and $c = 0.94$ for the 1mm approach, stating that also for the genes of category (C) the 1mm approach yields higher Pearson correlation coefficients than the naive and the gDNA approach. This difference might be explained by the fact that, even though the 1mm and the gDNA approaches retain the probe sets of the 36 genes of category (C), the probe sets contain different probes. The probe sets of the gDNA

approach lack approximately 30 % of the probes matching with at most one mismatch, but approximately 35 % of probes possibly match regions on the DNA outside transcribed regions, or match multiple targets (Table 4.1).

Fourth, we consider category (D), which contains 40 genes that are present in the naive and the 1mm approach, and we find Pearson correlation coefficients of $c = 0.78$ for the naive approach and $c = 0.92$ for the 1mm approach, stating that also for genes of category (D) the 1mm approach yields a higher Pearson correlation coefficient than the naive approach.

For all four categories, we find similar results for Spearman and Kendall correlation as for the Pearson correlation (Supplementary Table B.2).

In summary, we find that both sequence-based approaches yield more accurate expression responses than the naive and the gDNA approach. This finding is interesting, because the 1mm approach retains approximately 5500 additional transcript-specific probe sets than the more stringent 0mm approach, which allows a more comprehensive yet still accurate analysis of transcriptome changes/responses.

4.5. Conclusions

We address the problem of obtaining reliable expression response data for microarray-based comparative transcriptomics studies of a reference species and a closely related query species. We propose an approach that can be used if whole-transcriptome sequence information is available for the query species and that addresses the problems of (i) probes targeting no transcript, (ii) probes affected by cross hybridization, and (iii) probes targeting transcripts of non-orthologous genes.

We find that the 1mm and the 0mm approach yield a similar accuracy in qRT-PCR verification of the expression response values and outperform the naive and the gDNA approach, indicating that imperfectly matching probes with a single mismatch do not reduce the quality of the recorded $\Delta\mu$ log fold-changes.

However, using imperfectly matching probes with a single mismatch increases the number of transcript-specific probes per probe set and the number of transcript-specific probe sets of orthologous genes from 10629 for the 0mm approach to 16315 for the 1mm approach. In addition, the 1mm approach reduces the number of probe sets that are potentially affected by cross hybridization or that target transcripts of non-orthologous genes, and we conjecture that the proposed 1mm approach will considerably improve future comparative transcriptomics studies.

4.6. Acknowledgments

We thank Jan Grau for valuable discussions.

4.7. References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215 (3), pp. 403–410.

Arikawa, E., Sun, Y., Wang, J., Zhou, Q., Ning, B., Dial, S., Guo, L., and Yang, J. (2008). Cross-platform comparison of SYBR(R) Green real-time PCR with TaqMan PCR, microarrays and other gene expression measurement technologies evaluated in the MicroArray Quality Control (MAQC) study. *BMC Genomics*, 9 (1), p. 328.

Bar-Or, C., Czosnek, H., and Koltai, H. (2007). Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends in Genetics*, 23 (4), pp. 200–207.

Broadley, M. R., White, P. J., Hammond, J. P., Graham, N. S., Bowen, H. C., Emmerson, Z. F., Fray, R. G., Iannetta, P. P. M., McNicol, J. W., and May, S. T. (2008). Evidence of neutral transcriptome evolution in plants. *New Phytologist*, 180 (3), pp. 587–593.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10 (1), p. 421.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W.-R. (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in *Arabidopsis*. *Plant Physiology*, 139 (1), pp. 5–17.

Dannemann, M., Lorenc, A., Hellmann, I., Khaitovich, P., and Lachmann, M. (2009). The effects of probe binding affinity differences on gene expression measurements and how to deal with them. *Bioinformatics*, 25 (21), pp. 2772–2779.

Delker, C., Raschke, A., and Quint, M. (2008). Auxin dynamics: the dazzling complexity of a small molecule's message. *Planta*, 227 (5), pp. 929–941.

Fujimoto, R., Taylor, J. M., Sasaki, T., Kawanabe, T., and Dennis, E. S. (2011). Genome wide gene expression in artificially synthesized amphidiploids of *Arabidopsis*. *Plant Molecular Biology*, 77 (4), pp. 419–431.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20 (3), pp. 307–315.

Gilad, Y. and Borevitz, J. (2006). Using DNA microarrays to study natural variation. *Current Opinion in Genetics and Development*, 16 (6), pp. 553–558.

Graham, N., Broadley, M., Hammond, J., White, P., and May, S. (2007). Optimising the analysis of transcript data using high density oligonucleotide arrays and genomic DNA-based probe selection. *BMC Genomics*, 8 (1), p. 344.

Hammond, J., Broadley, M., Craigon, D., Higgins, J., Emmerson, Z., Townsend, H., White, P., and May, S. (2005). Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. *Plant Methods*, 1 (1), p. 10.

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., Fahlgren, N., Fawcett, J. A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J. D., Ossowski, S., Ottilar, R. P., Salamov, A. A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M. E., Bergelson, J., Carrington, J. C., Gaut, B. S., Schmutz, J., Mayer, K. F. X., Van de Peer, Y., Grigoriev, I. V., Nordborg, M., Weigel, D., and Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, 43 (5), pp. 476–481.

Irizarry, R. A., Gautier, L., Huber, W., and Bolstad, B. (2006). *makecdfenv: CDF Environment Maker*. R package version 1.30.0.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4 (2), pp. 249–264.

Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigele, S., Do, H.-H., Weiss, G., Enard, W., Heissig, F., Arendt, T., Nieselt-Struwe, K., Eichler, E. E., and Pääbo, S. (2004). Regional Patterns of Gene Expression in Human and Chimpanzee Brains. *Genome Research*, 14 (8), pp. 1462–1473.

Kleine-Vehn, J., Dhonukshe, P., Swarup, R., Bennett, M., and Friml, J. (2006). Subcellular Trafficking of the *Arabidopsis* Auxin Influx Carrier AUX1 Uses a Novel Pathway Distinct from PIN1. *The Plant Cell Online*, 18 (11), pp. 3171–3181.

Naiser, T., Kayser, J., Mai, T., Michel, W., and Ott, A. (2008). Position dependent mismatch discrimination on DNA microarrays - experiments and model. *BMC Bioinformatics*, 9 (1), p. 509.

Orlov, Y., Zhou, J., Lipovich, L., Shahab, A., and Kuznetsov, V. (2007). Quality assessment of the Affymetrix U133A&B probesets by target sequence mapping and expression data analysis. *In Silico Biol*, 7 (3), pp. 241–60.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. URL: R-project website. Available: http://www.R-project.org/. Accessed 2013 October 8.

Redman, J. C., Haas, B. J., Tanimoto, G., and Town, C. D. (2004). Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *The Plant Journal*, 38 (3), pp. 545–561.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., and Huala, E. (2008). The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research*, 36 (suppl 1), pp. D1009–D1014.

Wilson, C. L. and Miller, C. J. (2005). Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*, 21 (18), pp. 3683–3685.

German Conference on Bioinformatics, Volume P-235 of Lecture Notes in Informatics (LNI) - Proceedings, Pages 27-35, Bielefeld, September 2014

5. Explaining gene responses by linear modeling

Yvonne Poeschl, Ivo Grosse, and Andreas Gogol-Döring

German Center of Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany
Institute of Computer Science, Martin Luther University Halle–Wittenberg, Germany

5.1. Abstract

Increasing our knowledge about molecular processes in response to a certain treatment or infection in plants, insects, or other organisms requires the identification of the genes involved in this response. In this paper, we propose the *Profile Interaction Finder* (PIF) to identify such genes from gene expression data which is based on a convex linear model, and we investigate its efficacy for two applications related to stimulus response. First, we seek to identify sets of putative regulatory genes that explain the expression levels of a gene under different stimuli best. Second, we aim at identifying genes that show a specific response to a stimulus or a combination of stimuli. For both applications, we study the expression response of two *Arabidopsis* species to treatment with the plant hormone auxin and of *Apis mellifera* to pathogen infection. The proposed approach may be of general utility for analyzing expression data with a focus on genes and gene sets that explain specific stimulus response.

5.2. Introduction

Genes in a living cell form a complex network in which the expression level of each gene, i.e., the concentration of messenger RNA molecules, depends on the expression level of other genes. For instance, the expression of a gene encoding a transcription factor (TF) could rise because of an external stimulus, which consequently influences – directly or indirectly – the transcription of tens or hundreds of other genes.

The investigation of the causal effects between one TF and its target genes is a difficult task requiring complex laboratory experiments. Fortunately, it is possible to get indications of potential regulatory relationships between genes by comparing their expression levels under different conditions, e.g., before and after stimulation. A variety of methods have been developed for this purpose, for a recent review see Wang et al. (2014). The higher the number of

the involved data sets, and the more different conditions (treatments, time points, cell types, pathogens, etc.) are covered, the more detailed and accurate a prediction of the regulatory network could be. The underlying assumption is that genes that are closely connected in the regulatory network will also tend to have similar expression patterns under varying conditions. Mathematically speaking, gene expression levels obtained from M experiments can be represented by an M-dimensional vector, and if two genes are neighbors within this M-dimensional space, they presumably have a tight relation to each other in terms of their regulation.

A conventional clustering method like HCLUST (Murtagh et al., 2011) relying only on the relation between pairs of genes sometimes fails to model cases in which one gene is jointly regulated by several other genes while it is only loosely correlated with each individual regulator. The Local Context Finder (LCF) introduced by Katagiri et al. (2003) addresses this problem by reconstructing the M-dimensional expression profile of a gene as a linear combination of the expression profiles of other (neighboring) genes. One limitation of this approach is that it does not regard anti-correlated expression profiles. Although it is well known that, e.g., TFs could either increase or suppress the transcription of target genes, the latter case is not considered by the LCF.

In this paper we propose a new approach, the *Profile Interaction Finder* (PIF), that uses a distance metric that takes into account both positive and negative correlations. The approach selects for each gene a set of neighboring reference profiles that together explain the expression values of the gene best. Reference profiles are either expression profiles of other genes, which possibly have a regulatory influence on the current gene, or prototype profiles that reflect in which data set a certain experimental condition was present or absent. The proposed approach extends the LCF in two aspects, namely by considering both positive and negative interactions, and by using the flexible and generalizing notion of *reference profiles*. These extensions are instrumental in answering two central questions when analyzing expression data: (i) Which genes might have a positive or negative influence on the expression pattern of other genes?, And (ii) which genes respond positively or negatively to certain experimental conditions?

5.3. Methods

Supposed that we measure the expression of genes under varying conditions in M different experiments. To each gene we assign an *expression profile* $\underline{x} = (x_1, \ldots, x_M)$ containing the expression values of this gene. All expression profiles are normalized using a linear transformation such that the length $\|\underline{x}\| = 1$ and the mean $\bar{\underline{x}} = 0$. This normalization does not affect the Pearson correlation coefficient between two profiles $\underline{x}$ and $\underline{y}$, but it simplifies its calculation as the dot product $\underline{x} \cdot \underline{y}^T$ which can be interpreted as the cosine of the angle between the two vectors.

The goal of the proposed algorithm is to approximate a given expression profile by a linear model of *reference profiles* that could be either expression profiles of other genes or artificially created prototype profiles describing experimental conditions. Supposed for example that we set $n_m = 1$ if the m-th experiment is measured under a certain condition c, and $n_m = 0$ otherwise, then $\underline{n} = (n_1, \ldots, n_M)$ is after normalization a prototype profile for the condition

'measured on condition c'. More detailed examples for prototype profiles will be given in Section 5.4.2.

PIF returns for each gene $\underline{x}$ a set of *neighboring profiles* which are most informative for predicting $\underline{x}$. The proposed approach consists of three steps: (i) PIF first selects candidate reference profiles $\underline{n}_1, \ldots, \underline{n}_K$ related to $\underline{x}$ (Section 5.3.1), which (ii) are used to reconstruct $\underline{x}$ by a linear model (Section 5.3.2), and finally (iii) the results are filtered using bootstrapping (Section 5.3.3).

If gene expression profiles are used as references, the output could be interpreted as a gene regulatory network in which every gene is linked to all genes in its neighborhood. In case of prototype profiles, the genes could be sorted into clusters according to their neighborhoods. Examples for both applications are discussed in Section 5.4.

5.3.1. Selection of reference profiles

Fitting a linear model to a given input profile $\underline{x}$ could be computationally demanding, especially if the number of reference profiles is large. We therefore restrict the calculation to the subset of reference profiles that are most appropriate for reconstructing the input profile. This filtering process reduces computational costs and also improves the quality of the reconstruction by reducing noise.

For scoring the predictive power of a reference profile $\underline{n}$ relative to $\underline{x}$, we first compute the Pearson correlation coefficient between the two profiles. If this value is either close to 1 (positive correlation) or close to -1 (anti-correlation), then the two profiles are strongly connected, and in both cases the reference profile would be appropriate for reconstructing the input profile. A correlation coefficient of 0 on the other hand means that both vectors are orthogonal and no information about the input profile could be derived from the reference profile. The absolute value of the correlation coefficient $s = |\underline{x} \cdot \underline{n}^T|$ is a good indicator for the applicability of $\underline{n}$ for reconstructing $\underline{x}$. In contrast to the LCF given in Katagiri et al. (2003), which only chooses reference profiles with maximum *positive* dot product, PIF also takes highly informative reference profiles with *negative* dot product into account.

We select at most K profiles with maximal score $s \geq t$, where t is a user-defined threshold. A high value of t ensures that only reference profiles in close proximity to the input profile are used, whereas with $t = 0$ the filtering step would be omitted completely. In this paper we use $K = 10$ and $t = 0.25$.

5.3.2. Linear model reconstruction

In the main step of our approach, we reconstruct the input profile $\underline{x}$ as a linear combination of the reference profiles $\underline{n}_1, \ldots, \underline{n}_K$ selected in step (i) (Section 5.3.1). We calculate non-negative

weights $w_1, \ldots, w_K$ by a constrained linear fit such that the squared error function $f(\underline{w})$ is minimized,

$$f(\underline{w}) \quad = \quad \left\| \underline{x} - \sum_{k=1}^{K} w_k \mu_k \underline{n}_k \right\|^2 \quad \text{and} \quad 1 = \sum_{k=1}^{K} w_k, \tag{5.1}$$

where $\underline{\mu} = (\mu_1, \ldots, \mu_K) \in \{-1, 1\}^K$ denotes the signs of the dot products $\underline{x} \cdot \underline{n}_k^T$, i.e., $\mu_k = 1$ if $\underline{x} \cdot \underline{n}_k^T \geq 0$, and $\mu_k = -1$ if $\underline{x} \cdot \underline{n}_k^T < 0$. For reference profiles $\underline{n}_k$ that are anti-correlated to $\underline{x}$ the factor $\mu_k = -1$ reverts the direction of the reference profile such that the resulting profile $\underline{v}_k = \mu_k \underline{n}_k$ and $\underline{x}$ are *positively* correlated. This reduces the reconstruction to a convex linear combination, where all weights w_k are non-negative and sum to one.

We reformulate the optimization problem by including the constraint on the weights by introducing the Lagrangian multiplier λ:

$$L(\underline{w}, \lambda) \quad = \quad \left\| \underline{x} - \sum_{k=1}^{K} w_k \underline{v}_k \right\|^2 + \lambda \left(1 - \left(\sum_{k=1}^{K} w_k \right) \right) \tag{5.2}$$

We minimize $L(\underline{w}, \lambda)$ in eq. 5.2 by computing the derivatives for all w_k and then use the constraint in eq. 5.2 to compute λ, yielding

$$w_k = \sum_{j=1}^{K} s_{k,j}^{-1} \left(\frac{\lambda}{2} + \underline{v}_j \underline{x}^T \right) \qquad \lambda = 2 \cdot \frac{1 - \left(\sum_{j=1}^{K} \underline{v}_j \underline{x}^T \left(\sum_{k=1}^{K} s_{k,j}^{-1} \right) \right)}{\sum_{j=1}^{K} \sum_{k=1}^{K} s_{jk}^{-1}}, \tag{5.3}$$

where $\underline{v}_j = \mu_j \underline{n}_j$, and $\underline{\underline{s}}^{-1}$ is the inverse of $\underline{\underline{s}} = \underline{\underline{v}} \cdot \underline{\underline{v}}^T$ with $\underline{\underline{v}} = (\underline{v}_1, \ldots, \underline{v}_K)^T$. If $\underline{\underline{s}}$ becomes singular due to the linear dependence of some reference profiles, we compute the pseudo-inverse as suggested by Roweis et al. (2000).

The intended reconstruction of $\underline{x}$ is then given by the linear combination $\underline{r} = \sum_{k=1}^{K} w_k \mu_k \underline{n}_k$.

5.3.3. Determining robust neighborhoods

The weights $w_1, \ldots, w_K$ calculated in the previous section can be interpreted as degrees of relative importance of the reference profiles $\underline{n}_1, \ldots, \underline{n}_K$ for the explanation of an expression profile $\underline{x}$. Reference profiles $\underline{n}_k$ with a low weight w_k are likely expendable. Given a user-defined threshold r, we call the set $\{\underline{n}_k | w_k \geq r\}$ of all reference profiles with weights of at least r the *neighborhood* of $\underline{x}$. In this paper, we set $r = 0.1$.

The approach comprised of step (i) and (ii) (Section 5.3.1 and 5.3.2) described so far could be affected by noise in the gene expression data. Hence, we use bootstrapping in order to increase the reliability of the results. Given a data set with M samples, bootstrapping samples M out of these M samples with replacement, and we apply PIF to this sampled data set. We perform this bootstrapping step $L = 1000$ times and keep only reference profiles in the neighborhood of a gene which occurred in this neighborhood for at least p percent of the L repeats. In this

paper, we use a thresholds of $p = 50\%$ for gene expression reference profiles (Section 5.4.1) and $p = 75\%$ for prototype profiles (Section 5.4.2).

5.4. Results

We will now investigate if PIF is capable of producing biologically relevant results when applied to reconstructing gene regulatory networks (Section 5.4.1) and to clustering genes according to experimental conditions (Section 5.4.2).

5.4.1. Reconstruction of regulatory networks

Auxin is one of the key phytohormones that controls plant development and growth. So far, only parts of auxin signaling are understood (Delker et al., 2008). For the identification of novel candidate genes that might be involved in auxin signaling network, we applied PIF on a time-series of gene expression data of the two closely related plant species *Arabidopsis thaliana* and *Arabidopsis lyrata*, measured using expression microarrays at 0, 1, and 3 hours after auxin treatment. Each measurement was repeated three times, yielding $M = 2 \times 3 \times 3 = 18$ data sets.

We processed and normalized the raw data as described in Poeschl et al. (2013). 9091 genes with a coefficient of variation above 0.05 were selected for further analysis. Each of these genes could be regulated either enhanced or repressed by any of the other genes, so we used the expression profiles of all 9091 genes as possible reference profiles.

Figure 5.1 shows a part of the reconstructed gene network connected to the well known auxin responsive gene *AT5G54510* that is up-regulated upon auxin stimulation. According to the PIF analysis, AT5G54510 is part of the neighborhoods of four other genes. The correlation coefficients between AT5G54510 and the two genes AT3G58190 and AT4G37295 are positive, so AT5G54510 might have an enhancing effect on their expression. In contrast to that, the correlation coefficients to the other two target genes AT4G10270 and AT3G10040 are negative, suggesting that AT5G54510 possibly suppresses their transcription.

None of the four genes related to AT5G54510 had been identified to be involved in the auxin signaling pathway. Nevertheless, especially AT3G58190 seems to be very likely involved in hormone signaling, since this gene is also connected to two more factors AT4G14560 and AT4G27260 both related to the hormone metabolism.

5.4.2. Prototype analysis

In addition to the reconstruction of gene regulatory networks we can use the *Arabidopsis* data from the previous section to address various further questions. Examples are: 'Which genes respond quickly, or with a delay to auxin stimulation?' or 'Which genes are regulated differently in the two species?'. PIF is capable of answering these questions by using prototype profiles that reflect the different time points and species of the data sets (Figure 5.2A). Figure 5.2B-D

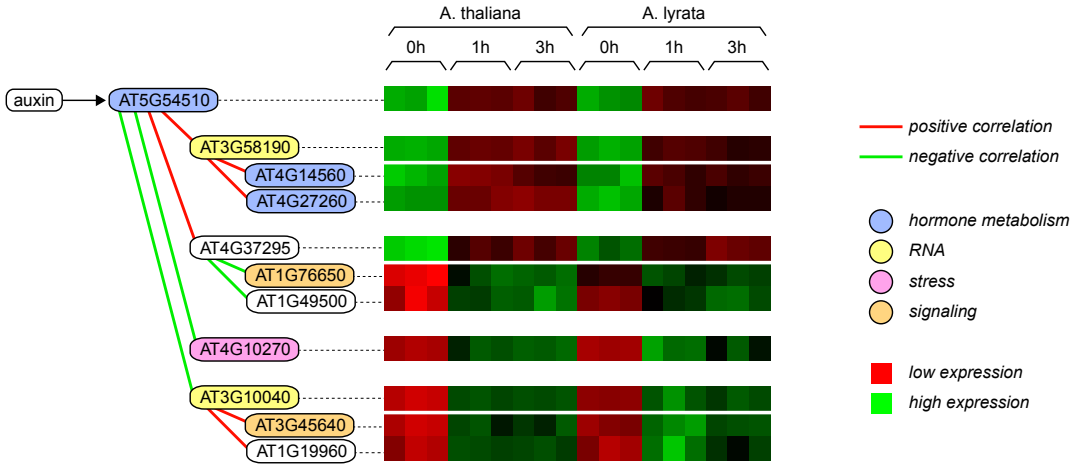


Figure 5.1.: A part of the regulatory network for *Arabidopsis* reconstructed by PIF showing genes related to the auxin responsive gene *AT5G54510*. Genes connected with red edges are positively correlated; green edges mean negative correlation. The gene colors correspond to the GO-terms they are annotated with using MapMan (Thimm et al., 2004). The heat map shows the expression levels of the 11 genes for each of the 18 data set. Red fields mean that the gene is highly expressed due to auxin treatment, while green fields mean low expression.

shows an example of the results of this analysis, a cluster of 16 genes initially highly expressed in both species and later down-regulated, but more strongly in *Arabidopsis thaliana* than in *Arabidopsis lyrata*.

This expression pattern is described by a combination of three prototype profiles (Figure 5.2B). Each single prototype profile differs strongly from the expression profiles of the genes in this cluster (Figure 5.2C and D), so the cluster could only be found because PIF reconstructs expression profiles by combining several reference profiles (Section 5.3.2).

Statistical analysis reveals that for the GO-term (Thimm et al., 2004) 'RNA' the number of annotated genes in this cluster is significantly higher than expected (p-value > 0.05, Fisher's exact test). This indicates that PIF possibly sorted the genes into biological meaningful clusters.

To investigate if PIF could also handle more diverse input data, we applied it to multiple data sets collected for a metastudy (*The Trans-Bee workshop* 2014) concerning the impact of different pathogens on gene expression in honeybee (*Apis mellifera*), see Table 5.1. The expression data were collected from different sources, measured for different tissues and on different platforms, and preprocessed with different methods, so they have very different dynamic ranges. Hence, we decided to use relative ranks (Breitling et al., 2004) instead of raw gene expressions as input for PIF.

We group 6242 genes present in all 9 data sets according to their response pattern to different experimental conditions, namely pathogens and tissues, see Figure 5.2A. Figure 5.2B-D show the example of a gene cluster containing 15 genes that respond positively to nosema infection in the fat body but negatively in the gut. Gut and fat body are distinct parts of the honeybee abdomen; genes in this group may be related to the immune response activated due to the infection. Although the individual genes within the clusters are more diverse than in the

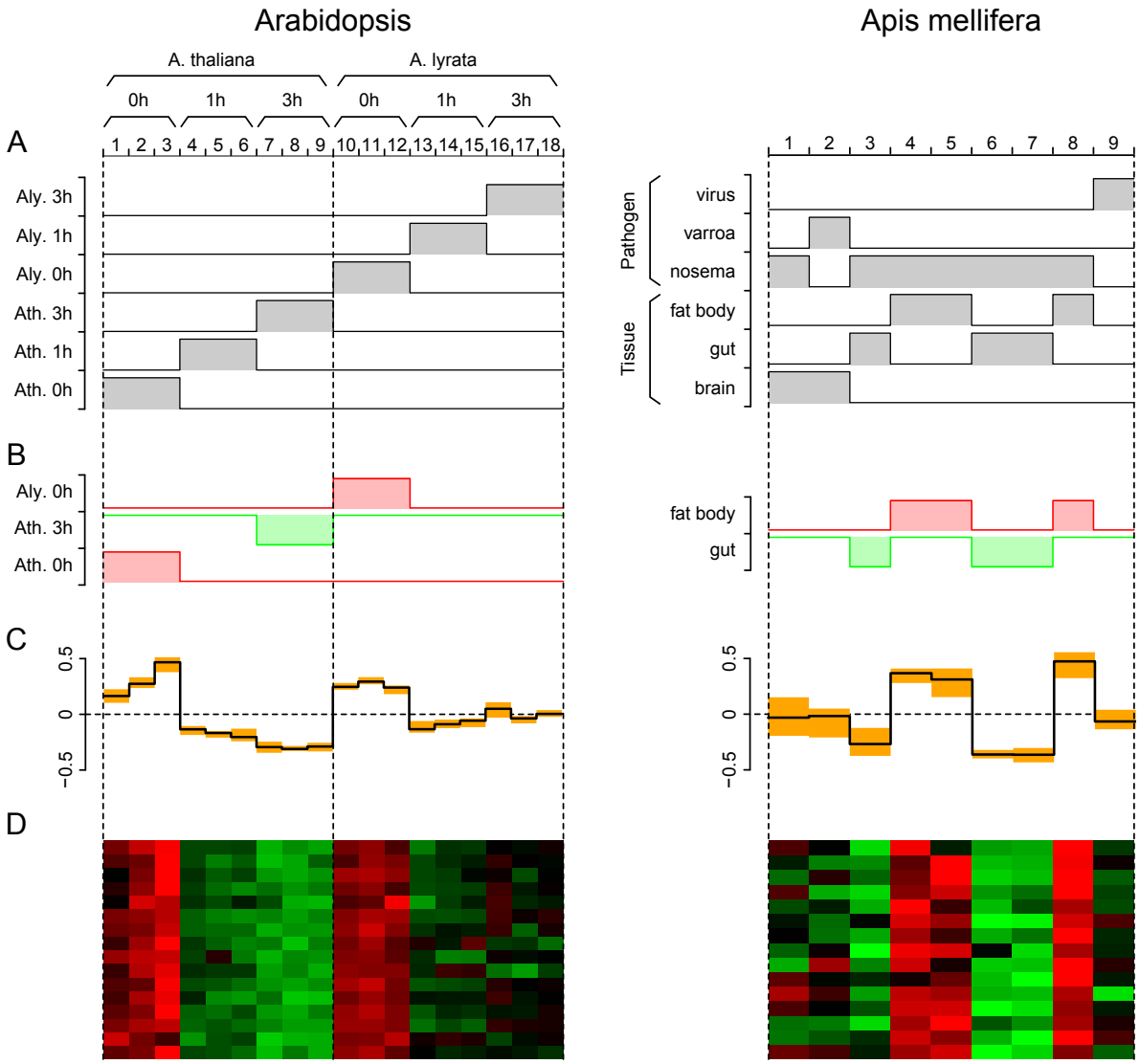


Figure 5.2.: PIF analysis using prototype profiles as reference. The left panel shows results of the *Arabidopsis* data analysis, and the right panel shows results of the *Apis mellifera* metastudy. A: The complete set of prototype profiles (before normalization) used in the analysis. B: Neighboring prototype profiles for one selected gene cluster. Prototype profiles which correlate positively to the genes in the cluster are shown in red; anti-correlations are shown in green. C: Averaged expression profiles of the genes in the cluster. The orange boxes show the area between the first and the third quartile. D: Heat maps showing the expression profiles of genes in the cluster. Each line represents one gene. Red boxes show highly expressed/up-regulated genes, and green boxes show low expressed/down-regulated genes.

	Pathogen	Tissue	Platform	Source
1	Nosema	brain	RNA-seq	McDonnell et al., 2013
2	Varroa	brain	RNA-seq	McDonnell et al., 2013
3	Nosema	gut	tiling microarray	Dussaubat et al., 2012
4	Nosema	fat body	expression microarray	Holt et al., 2013
5	Nosema	fat body	expression microarray	Holt et al., 2013
6	Nosema	gut	expression microarray	Holt et al., 2013
7	Nosema	gut	expression microarray	Holt et al., 2013
8	Nosema	fat body	expression microarray	Holt et al., 2013
9	Virus	whole bee	expression microarray	Flenniken et al., 2013

Table 5.1.: List of data sets used in the metastudy of *A. mellifera*.

data set for *Arabidopsis*, their expression profiles broadly follow the pattern defined by the prototypes.

5.5. Conclusions

The identification of genes acting as regulators of other genes or responding specifically to certain experimental conditions is an important aspect of gaining knowledge about gene regulatory processes in response to a treatment or infection. In this paper, we propose PIF, the Profile Interaction Finder, a novel approach that can be applied to expression data sets in order to tackle these questions.

Studying data sets of *A. thaliana* and *A. lyrata* after auxin treatment, and of *A. mellifera* after infection with different pathogens, PIF successfully identified genes related to the cell responses for the respective stimulus. In addition to that, PIF determined novel putative regulators that might affect several other genes in the downstream response. The detected targets of the *Arabidopsis* gene AT5G54510 for example had not yet been identified to be involved in the auxin signaling pathway. This shows that PIF is capable to discover previously unknown relationships between genes. The obtained results are highly relevant, as shown by linking them to already existing biological knowledge, represented for example in the gene ontology. Being capable to identify not only enhancing but also suppressing regulators is another advantageous feature of PIF. For example, with our method we were able to find two genes which are possibly down-regulated by AT5G54510.

Hence we conclude that PIF is a valuable tool for getting deeper insights into biological processes by analyzing gene expression data under varying experimental conditions.

5.6. Acknowledgements

We thank Carolin Delker, Jan Grau, Marcel Quint, Jana Trenner, and all participants of the Trans-Bee workshop for valuable discussions.

The honeybee transcriptome data used in Section 5.4.2 were collected and analyzed for the project Trans-Bee (*The Trans-Bee workshop* 2014), which was kindly supported by sDiv, the Synthesis Centre for Biodiversity Sciences – a unit of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118).

5.7. References

Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Letters*, 573 (1–3), pp. 83–92.

Delker, C., Raschke, A., and Quint, M. (2008). Auxin dynamics: the dazzling complexity of a small molecule's message. *Planta*, 227 (5), pp. 929–941.

Dussaubat, C., Brunet, J.-L., Higes, M., Colbourne, J. K., Lopez, J., Choi, J.-H., Martín-Hernández, R., Botías, C., Cousin, M., McDonnell, C., Bonnet, M., Belzunces, L. P., Moritz, R. F. A., Le Conte, Y., and Alaux, C. (2012). Gut Pathology and Responses to the Microsporidium *Nosema ceranae* in the Honey Bee *Apis mellifera*. *PLoS ONE*, 7 (5), e37017.

Flenniken, M. L. and Andino, R. (2013). Non-Specific dsRNA-Mediated Antiviral Response in the Honey Bee. *PLoS ONE*, 8 (10), e77263.

Holt, H., Aronstein, K., and Grozinger, C. (2013). Chronic parasitization by *Nosema* microsporidia causes global expression changes in core nutritional, metabolic and behavioral pathways in honey bee workers (*Apis mellifera*). *BMC Genomics*, 14 (1), p. 799.

Katagiri, F. and Glazebrook, J. (2003). Local Context Finder (LCF) reveals multidimensional relationships among mRNA expression profiles of *Arabidopsis* responding to pathogen infection. *Proceedings of the National Academy of Sciences*, 100 (19), pp. 10842–10847.

McDonnell, C., Alaux, C., Parrinello, H., Desvignes, J.-P., Crauser, D., Durbesson, E., Beslay, D., and Le Conte, Y. (2013). Ecto- and endoparasite induce similar chemical and brain neurogenomic responses in the honey bee (*Apis mellifera*). *BMC Ecology*, 13 (1), pp. 1–15.

Murtagh, F. and Contreras, P. (2011). Methods of Hierarchical Clustering. *CoRR*, abs/1105.0121.

Poeschl, Y., Delker, C., Trenner, J., Ullrich, K. K., Quint, M., and Grosse, I. (2013). Optimized Probe Masking for Comparative Transcriptomics of Closely Related Species. *PLoS ONE*, 8 (11), e78497.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE*, 290, pp. 2323–2326.

The Trans-Bee workshop (2014). http://www.idiv-biodiversity.de/sdiv/workshops/workshops-2013/stransbee (accessed 2014/07/23).

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L. A., Rhee, S. Y., and Stitt, M. (2004). MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal*, 37 (6), pp. 914–939.

Wang, Y. R. and Huang, H. (2014). Review on statistical methods for gene network reconstruction using expression data. *Journal of Theoretical Biology*.

6. Variation of IAA-induced transcriptomes pinpoints the AUX/IAA network as a potential source for inter-species divergence in auxin signaling and response

Jana Trenner[1,*], Yvonne Poeschl[2,3,*], Jan Grau[3], Andreas Gogol-Döring[2,3], Marcel Quint[1,4], and Carolin Delker[1,4]

[1] Department of Molecular Signal Processing, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale), Germany
[2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany
[3] Institute of Computer Science, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany
[4] Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle-Wittenberg, Betty-Heimann-Str. 5 , 06099 Halle (Saale), Germany
[*] These authors contributed equally to this work.

6.1. Abstract

Auxin is an essential regulator of virtually all aspects of plant growth and development and components of the auxin signaling pathway are conserved among land plants. Yet, a remarkable degree of natural variation in physiological and transcriptional auxin responses has been described among *Arabidopsis thaliana* accessions. Such variations might be caused by divergence in promoter or coding sequences of signaling and/or response genes that ultimately result in altered protein levels or functions. As intra-species comparisons offer only limited sequence variation, we here combined physiological, transcriptomic and genomic information to inspect the variation of auxin responses of *A. thaliana* and *A. lyrata*. This approach allowed the identification of genes with conserved auxin responses in both species and provided novel genes with potential relevance for auxin biology. Furthermore, gene expression and promoter sequence divergence were exploited to assess potential sources of variation. *De novo* motif discovery identified variants of known as well as novel promoter elements with potential relevance for transcriptional auxin responses. Furthermore, expression of *AUXIN/INDOLE-3-ACETIC ACID (AUX/IAA)* signaling genes was highly diverse between *A. thaliana* and *A. lyrata*. Network analysis revealed positive and negative correlations of inter-species differences in

the expression of *AUX/IAA* gene clusters and classic auxin-related genes. We conclude that variation in general transcriptional and physiological auxin responses may originate substantially from functional or transcriptional variations in early auxin signaling components. Hence, transcriptional and/or functional variations within the network of *TRANSPORT INHIBITOR RESPONSE 1/AUXIN SIGNALING F-BOX*, *AUX/IAA* and *AUXIN RESPONSE FACTOR* gene family members may be targets for adaptation processes that contribute to phenotypic plasticity within and between species.

6.2. Introduction

Auxin's capacity to regulate the essential cellular processes of division, elongation and differentiation integrates it in the regulation of virtually all developmental and physiological plant processes. On a molecular level, auxin responses involve extensive and rapid changes in the transcriptome (Paponov et al., 2008). This response depends on a signaling pathway which is constituted by three main signaling components: (i) TRANSPORT INHIBITOR RESPONSE1/AUXIN SIGNALING F-BOX1-5 (TIR1/AFBs) auxin-co-receptors, (ii) AUXIN/INDOLE-3-ACEDIC ACID (AUX/IAA) family of auxin co-receptors/transcriptional repressors, and (iii) the AUXIN RESPONSE FACTOR (ARF) family of transcription factors (Quint et al., 2006).

ARFs induce or repress the expression of genes by binding to auxin-responsive elements (AuxRE) in the respective promoter regions (Guilfoyle et al., 1998; Ulmasov et al., 1999). When auxin levels are low, AUX/IAAs in concert with additional repressors such as TOPLESS heterodimerize with ARFs which prevents ARF regulatory action on auxin-responsive genes (Weijers et al., 2005; Szemenyei et al., 2008). The presence of auxin is sensed by a co-receptor complex formed by the cooperative binding of auxin by the TIR1/AFB F-box subunit of an SCF-type E3 ligase and an AUX/IAA protein (Dharmasiri et al., 2005a; Kepinski et al., 2005; Calderón Villalobos et al., 2012). This binding results in the polyubiquitylation of the AUX/IAAs by the $SCF^{TIR1/AFB}$ complex (Dos Santos Maraschin et al., 2009). The subsequent proteasomal degradation of the tagged AUX/IAAs causes a de-repression of ARF transcription factors, which are then released to initiate transcriptional changes (Ramos et al., 2001; Zenser et al., 2001). The three key signaling elements of TIR1/AFBs, AUX/IAAs, and ARFs are encoded by gene families of six, 29 and 23 members, respectively (Chapman et al., 2009). The virtually infinite possibilities of combinations among the individual gene family members with putatively different signaling capacities could ultimately be responsible for the wide range of auxin signaling outputs observed throughout plant growth and development (Calderón Villalobos et al., 2012; Salehin et al., 2015).

The auxin signaling pathway seems to be conserved among land plants as individual core components are present already in the liverwort *Marchantia polymorpha* (Kato et al., 2015). With the universal impact of auxin on plant growth and development, an open question in auxin biology remains whether auxin signaling and response contribute to adaptive processes to local environmental conditions and challenges. First data indicating that the read-out of an auxin stimulus can be highly variable were obtained by the analysis of natural variation

of physiological and transcriptional auxin responses among different accessions of *A. thaliana* (Delker et al., 2010). Apart from a striking diversity in auxin-induced transcriptome changes, a remarkably high variation among accessions was detectable for co-expression networks of early auxin signaling components. These variations gave rise to the hypothesis that altered equilibria of specific signaling components might contribute to the variation observed on the general transcriptome and ultimately on the physiological level (Delker et al., 2010).
Here, we performed a cross-species analysis of auxin responses in the closely related sister species *A. thaliana* and *A. lyrata* in a comparative transcriptomics approach. The increased genetic variation between the two *Arabidopsis* species compared to the variation among different accessions allowed (i) the identification of genes with similar auxin responses in both species that might constitute essential or conserved auxin response genes. We furthermore aimed (ii) to exploit the genetic variation in promoter sequences to identify *cis*-regulatory elements that might contribute to similar or differential auxin responses, and (iii) we aimed to test whether the previously hypothesized variation in early auxin signaling gene expression as a source for downstream variation could be verified in a system with higher genetic variation.

6.3. Materials and methods

6.3.1. Plant material and growth conditions

A. thaliana accession Col-0 (N1092) and *A. lyrata* accession (N22696) were obtained from the Nottingham Arabidopsis Stock Centre. Seeds were surface-sterilized and imbibed in deionized H_2O for 3 d at 4 °C before sowing. Seedlings were germinated and grown under sterile conditions on solid or in liquid *Arabidopsis thaliana* solution (ATS) nutrient medium (Lincoln et al., 1990). For growth assays, seedlings were cultivated on vertical un-supplemented ATS for 3 d (IAA), 4 d (TIHE) or 5 d (2,4-D and NAA) before transfer to plates supplemented with IAA, 2,4-D, or NAA at the indicated concentrations or before transfer of plates to 28 °C (TIHE). Root lengths were quantified after an additional 5 d (IAA) or 3 d (2,4-D and NAA), hypocotyl growth was quantified after additional 4 d at 28 °C. All experiments were performed in long-day conditions (16 h light/8 h dark) and a fluence rate of $\sim 230\,\mu\text{mol}\,\text{m}^{-2}\,\text{sec}^{-1}$ (root growth assays) or $30\,\mu\text{mol}\,\text{m}^{-2}\,\text{sec}^{-1}$ (TIHE). To visualize auxin and temperature responses, relative root and hypocotyl lengths of hormone- and temperature-treated seedlings, respectively, were determined as percent in relation to the median value of 20 °C grown plants. Statistical analyses (1- and 2-way ANOVAs) were performed on the untransformed raw data. For expression studies and [^{3}H]-IAA uptake assays, seeds were germinated and cultivated in liquid ATS under continuous illumination to minimize potential circadian effects. For expression analyses, ATS was supplemented with 1 μM IAA for 0 h, 1 h, and 3 h after seven days. Yellow long-pass filters were applied in all IAA treatment experiments to prevent photodegradation of IAA.

6.3.2. [^{3}H]-IAA uptake assay

Three biological replicates of seven days-old seedlings were treated with 2 nM of [^{3}H]-IAA (Hartmann Analytic, Germany) per mg seedling fresh weight in liquid ATS for 1 h. Samples

were subsequently washed with liquid ATS ten times before quantification via scintillation count.

6.3.3. RNA extraction and microarray hybridization

RNA was extracted from three biological samples of seven days-old whole seedlings using the RNeasy Plant Mini Kit (Qiagen) including the on-column Dnase treatment according to the manufacturers description. After assessment of RNA integrity the samples were sent to the Nottingham Arabidopsis Stock Centre's microarray hybridization service for further processing and hybridization to the ATH1-121501 microarray.

6.3.4. Probe masking, data normalization and data processing

The raw data generated by NASC was pre-processed and corrected according to (Poeschl et al., 2013) including the proposed polynomial correction of probe intensities. The data matrix contained the expression values for 16315 genes at three time points (with three biological replicates each) for both species.
Significant changes in auxin-induced expression were determined by a modified t-test (Opgen-Rhein et al., 2007). P-values were Benjamini-Hochberg-corrected for multiple testing and genes significantly (fdr $< 5\,\%$) changed by a factor of two or more ($|\log_2$ fold change$| > 1$) where considered to be differentially expressed.

6.3.5. Modified Pearson correlation

To incorporate the information on variation among the biological replicate measurements at each time point in the correlation analyses, a modified Pearson correlation coefficient ($mod.r$) was introduced. $mod.r(\underline{x}_A, \underline{x}_B)$ of the expression profiles for two genes A and B was computed by dividing the covariance of the mean expression profiles $cov(\bar{\underline{x}}_A, \bar{\underline{x}}_B)$ by the product of the standard deviations of the expression profiles $sd(\underline{x}_A) \cdot sd(\underline{x}_B)$, which is given by the formula:

$$mod.r(\underline{x}_A, \underline{x}_B) = \frac{cov(\bar{\underline{x}}_A, \bar{\underline{x}}_B)}{sd(\underline{x}_A) \cdot sd(\underline{x}_B)}$$

The mean expression profiles ($\bar{\underline{x}}_A$ and $\bar{\underline{x}}_B$)) consist of one value per time point, which represent the means of the respective replicates.

6.3.6. Cluster analysis

A total of N = 9091 genes were selected based on a coefficient of variation (cv) in expression profiles of cv > 0.05. A hierarchical clustering with average linkage was performed on N expression profiles using $1 mod.r$ as distance measure. Each expression profile consists of 18 measurements representing the three biological replicates of three time points and two species. The resulting dendrogramm was cut at level 0.1 ($mod.r = 0.9$) and resulting clusters were subsequently filtered by the following parameters: Clusters needed to contain at least 5 genes

of which 70% showed a significant dierence in species, time point and interaction as assessed by two-way ANOVAs which resulted in 14 clusters containing 337 genes in total.

6.3.7. Promoter analysis

Promoter sequences for *A. thaliana* and *A. lyrata* were extracted using the annotation provided by Phytozome v7.0 (`http://www.phytozome.com`). A promoter sequence was defined as 500 bp upstream the transcription start site to 100 bp downstream the transcription start site, or to the start codon, whichever came first.

6.3.8. Extraction and assignment of known cis-elements

Extracted promoter sequences were analyzed for the presence of a set of annotated cis-elements and their reverse complements from `http://arabidopsis.med.ohio-state.edu/AtcisDB/bindingsites.html` (last accessed 2014/02/03) extended by a set of 10 *cis*-elements described in literature to be involved in auxin response/signaling (Tab. C.2). Motifs shorter than six bases were excluded from the analysis. The sequences of the motifs were used as regular expressions to compute their occurrences in the promoter sequences.

6.3.9. Determination of promoter and expression divergence

Similarities of promoter sequences of an orthologous gene pair was assessed by determining the occurrence of each possible 8-mer in each of the two promoter sequences and computing the Pearson correlation coefficient of the two vectors of k-mer counts ($kmer.r$) as proposed in (Vinga et al., 2003). Promoter divergence was assessed as 1-$kmer.r$ and expression divergence was determined as 1-$mod.r$.

6.3.10. De-novo identification of putative cis-elements

Dimont (Grau et al., 2013) was used for identification of putative novel *cis*-elements with slight modifications from the published procedure which are comprehensively described in the Supplemental Methods Section.

6.3.11. Co-expression analysis using Profile Interaction Finder (PIF)

The Profile Interaction Finder algorithm (PIF, Poeschl et al., 2014) was applied in its second mode using eight input profiles of the individual mean expression profiles of the eight AUX/IAA gene clusters (Figure 6.5B). We applied the PIF to the set of genes showing a cv > 0.05 to prevent false-positive correlation based on noise. Parameters and thresholds for the identification of positively or negatively correlated genes were set to a |PIF-correlation| of > 0.7, neighbor number k = 1 and a 75 % bootstrap occurrence (n = 1000).

6.3.12. GO-term analysis

GO-terms for *A. thaliana* genes were provided by MapMan (Thimm et al., 2004). Over- or under-representation of GO-terms was assessed by a two-sided Fisher's exact test using the *stats* package. Resulting p-values were Benjamini-Hochberg corrected for multiple testing using *multtest* package (Pollard et al., 2005).

6.3.13. Statistical and computational analyses

Analyses were performed using the software R (R Core Team, 2012) with implementation of the following packages: beeswarm (Eklund, 2015), gplots (Warnes et al., 2014), st (Opgen-Rhein et al., 2007), multtest (Pollard et al., 2005).

6.3.14. Accession numbers

The cross-species hybridization microarray data analyzed in this article are publicly available at `http://data.iplantcollaborative.org/quickshare/8e9b2f0212c8a1bc/Exp579.zip`.

6.4. Results and discussion

We inspected inter-species variation of auxin responses between *A. thaliana* and *A. lyrata* with the aim to assess whether auxin signaling and responses differentially contribute to adaptive variation and phenotypic plasticity. We took advantage of the close relation of the two *Arabidopsis* species providing extensive synteny despite considerable genetic variation, for example in total genome size (Hu et al., 2011). The aim was to combine physiological, transcriptomic and genomic information to assess the extent of inter-species variation in auxin responses on several levels and to identify genes with conserved transcriptional responses. Furthermore, we wanted to exploit the genetic variation among the two sister species to gain further insights into the molecular mechanisms that contribute to naturally occurring variation in auxin responses which might ultimately reflect consequences of adaptation processes.

6.4.1. Physiological auxin responses

To assess whether *A. thaliana* and *A. lyrata* show differences in physiological auxin responses we used classic auxin response assays that focus on the quantitative reaction of seedling growth to exogenously applied auxin or to a temperature-induced increase of endogenous auxin levels. We performed several of these assays, testing the response to the naturally prevalent auxin indole-3-acetic acid (IAA) as well as several synthetic auxins, to assess the extent of natural inter-species variation between *A. thaliana* and *A. lyrata*.

In terms of relative growth effects, a high diversity in responses to natural and synthetic auxins was observed (Fig. 6.1A-D). While *A. thaliana* is less sensitive with respect to IAA-induced

root growth inhibition (Fig. 6.1A), a higher sensitivity in temperature-induced hypocotyl elongation (TIHE) was observed (Fig. 6.1D). *A. thaliana*'s response to the synthetic auxin 2,4-Dichlorophenoxyacetic acid (2,4-D) was significantly stronger than the response of *A. lyrata* (Fig. 6.1B). In contrast, 1-Naphthaleneacetic acid (NAA)-induced root growth inhibition was almost similar in both species (Fig. 6.1C). Overall, the extent of variation in auxin responses between *A. thaliana* and *A. lyrata* seems to be highly dependent on the specific auxinic compound and the analyzed organ. The compound- and tissue-specificity might indicate differential sources for the observed response differences putatively involving any or all aspects of auxin biology ranging from biosynthesis (in case of TIHE) to transport, sensing, signal transduction and/or metabolism.

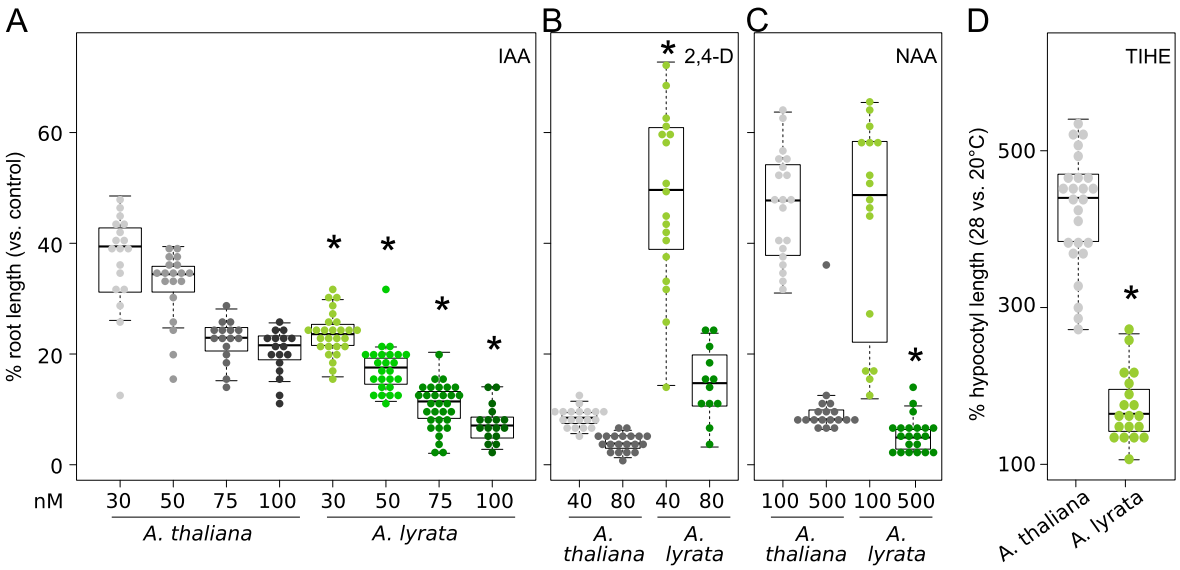


Figure 6.1.: Physiological auxin responses of *A. thaliana* and *A. lyrata*. Relative root length (treated vs. control) of seedlings grown on different concentrations of (A) IAA, (B) 2,4-D, or (C) NAA. 3 (A) or 5 (B,C) days-old seedlings were transferred to hormone-containing medium and grown for additional 5 (A) or 3 (B,C) days. (D) Relative hypocotyl length (28 °C/20 °C) of 8 days-old seedlings. Box plots show medians (horizontal bar), interquartile ranges (IQR, boxes), and data ranges (whiskers) excluding outliers (defined as $> 1.5 \times$ IQR). Individual data points are superimposed as beeswarm plots. Asterisks denote significant differences between treatment responses of *A. thaliana* and *A. lyrata* as assessed by two-way ANOVA (i.e. genotype x treatment effect, $P < 0.05$) of the absolute data presented in Fig. C.1.

6.4.2. Microarray-based transcriptional profiling of auxin responses

For *A. thaliana*, natural variation among different accessions was observed on physiological as well as on transcriptional levels (Delker et al., 2010). We thus conducted a similar microarray-based analysis of transcriptional auxin responses comparing *A. thaliana* and *A. lyrata* using a cross-species hybridization approach. The experimental set up was similar to that reported previously (Delker et al., 2010). In brief, seven days-old seedlings grown in liquid culture were treated with 1 μM IAA for one and three hours, respectively. Isolated RNAs from treated and control (untreated) seedlings were subsequently processed and hybridized to the Affymetrix ATH1 microarray. To exclude potential effects of differential auxin uptake on the transcriptional read-out, we quantified the amount of radio-labeled auxin in seven days-old seedlings

exposed to [^{3}H]-IAA for one hour (Fig. 6.2A). The lack of statistically significant differences in [^{3}H]-IAA levels indicated similar IAA uptake capacities in *A. thaliana* and *A. lyrata*.

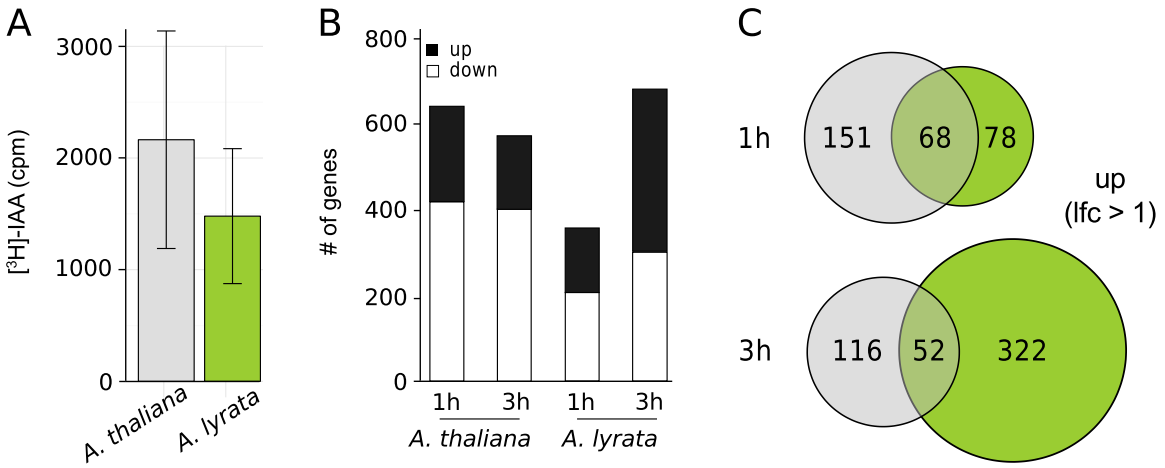


Figure 6.2.: Quantification of [^{3}H]-IAA uptake and ATH1-based assessment of auxin-induced transcriptome changes. (A) 7 days-old seedlings were treated with 2ng [^{3}H]-IAA per mg seedling fresh weight for 1h in liquid ATS medium. Scintillation counts were recorded after removal of radiolabeled IAA and ten subsequent wash steps with liquid ATS. Bar plots show mean [3H]-IAA levels of three biological replicates and error bars denote SEM. No significant differences were detected by a two-sided t-test ($P < 0.05$). (B) Stacked bars show the number of up- and down-regulated genes with an auxin-induced significant (fdr ≤ 0.05) change in expression level in black and white, respectively. (C) Venn diagrams illustrate the number of genes commonly or specifically up-regulated in *A. thaliana* (gray) and *A. lyrata* (green) after 1 h and 3 h of auxin treatment (lfc = $\log_2$ fold change).

The hybridization of a non-intended species to a species-specific microarray requires a probe-masking procedure in the processing of the expression data to avoid false-positive and false-negative results caused by mis-hybridization of probes due to sequence variations between the two species. Here, a sequence-based masking approach was applied that allows for one mismatch per probe and retained only those genes that are represented by at least three probes per probe set and uniquely hybridize to orthologous genes in *A. thaliana* and *A. lyrata* (Poeschl et al., 2013). As a result of the masking procedure, 16315 genes were retained for expression comparisons between *A. thaliana* and *A. lyrata*. To correct for putative effects of one tolerated mismatch per probe on the expression level we implemented a fourth-degree polynomial correction option in the RMA-normalizing procedure as suggested by Poeschl et al. (2013). After normalization we inspected the expression levels of various constitutively expressed genes designated as superior expression reference genes in *A. thaliana* (Czechowski et al., 2005). This subset of genes showed similar transcription profiles as well as largely similar expression levels in both *Arabidopsis* species indicating the comparability of the two data sets (Fig. C.2).

To analyze auxin-induced transcriptome changes, differentially expressed genes in both species were identified based on a significant (fdr < 0.05) change in expression with a $|\log_2$ fold change$| > 1$. Several hundred genes were differentially regulated in response to auxin in both species (Fig. 6.2B). Considerably more genes were differentially regulated in *A. thaliana* in response to one hour of auxin treatment than in *A. lyrata*, whereas after three hours more genes were responsive in *A. lyrata*. Overall, the number of down-regulated genes

was relatively high in comparison to other auxin-response transcriptome analyses (Paponov et al., 2008; Delker et al., 2010). In accordance with previous studies, we focused primarily on differentially up-regulated genes in the subsequent analyses.

6.4.3. Identification of conserved response genes

Several gene families are known to be up-regulated by elevated auxin levels in *A. thaliana* (Paponov et al., 2008). The cross-species approach might provide further insights into the identity of genes that are conserved in their response to auxin and might thus be of particular importance for auxin signaling, metabolism and/or response. The intersection of up-regulated genes among the two *Arabidopsis* species was moderate at both time points (Fig. 6.2C). Among the commonly up-regulated genes were individual members of prominent auxin-response gene families such as the *ASYMMETRIC LEAVES/LATERAL ORGAN BOUNDARIES DOMAIN (ASL/LBD)*, *GRETCHEN HAGEN 3 (GH3)*, *AUX/IAA* and *SMALL AUXIN UPREGULATED (SAUR)* families (Tab. 6.1 and Tab. C.1), validating the successful auxin induction. In addition, numerous other genes were induced by auxin treatment in both species. This included known auxin-responsive genes (e.g. *ARABIDOPSIS THALIANA HOMEOBOX 2 (HAT2)/AT5G47370)*, genes associated with other phytohormones (e.g. *1-AMINOCYCLOPROPANE-1-CARBOXYLATE SYNTHASE 11 (ACS11)/ AT4G08040*, *BRASSINOSTEROID INSENSITIVE LIKE 3 (BRL3)/AT3G13380*, *GIBBERELLIN 2-OXIDASE 8 (GA2ox8)/AT4G21200)* as well as several genes with so far unknown function (e.g. *AT1G29195*, *AT1G64405*, etc). The latter group in particular might be of interest as the conserved response to the auxin stimulus in both species might indicate potential new candidate genes relevant for auxin responses.

6.4.4. Inter-species expression responses in auxin-relevant gene families

To further investigate similarities and specificities of transcriptional auxin responses in *A. thaliana* and *A. lyrata*, we performed a cluster analysis of genes that showed a change in expression in at least one species at any of the analyzed time points with a coefficient of variation (cv) > 0.05. A modified Pearson correlation ($mod.r$) was used as a distance measure in the hierarchical clustering to incorporate information on the variation among the three biological replicates at each analyzed time point. To filter for correlations among genes with potential biological relevance, we further applied a minimum cut-off in correlation of $mod.r = 0.7$. The resulting 14 gene clusters fall into two clearly distinct groups (Fig. 6.3). Clusters 1 - 8 and clusters 9 - 14 are predominantly characterized by genes that show a higher expression level and/or response in *A. lyrata* or *A. thaliana*, respectively. Only very few clusters show high similarities among the expression profiles of both species (e.g. cluster 2 and 9). The majority of cluster profiles show small to striking differences between the two species in either expression levels (e.g. cluster 8) or expression response in terms of induction/repression profiles (e.g. cluster 3) or both (e.g. cluster 11). We next inspected whether the presence and frequency of known *cis*-regulatory elements in the promoters of clustered genes could explain the observed patterns of similarities or differences in the expression profiles of individual

Table 6.1.: Conserved auxin up-regulated genes. Genes significantly up-regulated ($|\log_2$ fold change$| > 1$) in *A. thaliana* and *A. lyrata* after 1h (1) and/or 3h (3) of auxin treatment in 7 days-old seedlings. Detailed information on *A. lyrata* locus identifiers, corresponding ATH1 array elements and expression levels are shown in Tab. C.1.

AUX/IAA	
AT1G04240	IAA3 [1]
AT1G15580	IAA5 [1]
AT2G33310	IAA13 [13]
AT3G15540	IAA19 [13]
AT3G23030	IAA2 [13]
AT3G62100	IAA30 [1]
AT4G14560	IAA1 [13]
AT4G28640	IAA11 [13]
AT4G32280	IAA29 [13]
AT5G43700	IAA4 [1]
auxin transport	
AT1G23080	PIN7 [1]
AT1G70940	PIN3 [13]
AT1G73590	PIN1 [1]
AT2G17500	PILS5 [3]
AT2G21050	LAX2 [1]
ASL/LBD	
AT2G42430	ASL18/LBD16 [13]
AT2G42440	ASL15/LBD17 [13]
AT3G58190	ASL16/LBD29 [13]
expansins	
AT3G45970	EXLA1 [1]
AT4G17030	EXLB1 [3]
AT4G38400	EXLA2 [1]
GH3	
AT2G14960	GH3.1 [1]
AT2G23170	GH3.3 [13]
AT4G27260	GH3.5 [13]
AT5G54510	GH3.6 [13]
SAUR	
AT2G18010	SAUR10 [1]
AT4G34760	SAUR50 [1]
AT4G34770	SAUR1 [1]
AT4G38850	SAUR15 [13]
AT4G38860	SAUR16 [13]

others	
AT1G02850 [3]	AT3G28740 [3]
AT1G05560 [3]	AT3G30180 [3]
AT1G05680 [13]	AT3G42800 [1]
AT1G10380 [3]	AT3G43270 [3]
AT1G14280 [1]	AT3G44540 [3]
AT1G17170 [3]	AT3G50340 [13]
AT1G17180 [3]	AT3G51410 [1]
AT1G21980 [1]	AT3G54950 [1]
AT1G23340 [1]	AT4G15550 [3]
AT1G23730 [3]	AT4G16515 [1]
AT1G29195 [13]	AT4G16515 [3]
AT1G30100 [1]	AT4G17350 [13]
AT1G30760 [3]	AT4G21200 [1]
AT1G32870 [3]	AT4G30140 [3]
AT1G57560 [1]	AT4G37295 [13]
AT1G59740 [1]	AT5G02760 [13]
AT1G64405 [13]	AT5G04190 [1]
AT1G70270 [13]	AT5G06860 [3]
AT2G03760 [1]	AT5G12050 [13]
AT2G26710 [1]	AT5G18560 [1]
AT2G29490 [1]	AT5G26930 [1]
AT2G39370 [13]	AT5G47370 [13]
AT2G41820 [1]	AT5G50130 [1]
AT2G47550 [3]	AT5G51440 [3]
AT3G03660 [1]	AT5G52900 [13]
AT3G09270 [3]	AT5G53290 [1]
AT3G13380 [3]	AT5G57760 [1]
AT3G22370 [3]	AT5G61820 [3]
AT3G26760 [1]	AT5G62280 [1]
AT3G26960 [1]	AT5G65320 [3]
AT3G28420 [1]	AT5G66940 [1]

clusters. We limited the size of the putative promoter region to 500 bp upstream of the transcription start site. While eukaryotic promoters can arguably be much larger, the majority of *cis*-regulatory sequences should be present within this 500 bp interval (Franco-Zorrilla et al., 2014). We analyzed the presence of 99 known *cis*-regulatory elements taken from the *Arabidopsis cis*-regulatory element database (`http://arabidopsis.med.ohio-state.edu/AtcisDB/`) and additional literature (Tab. C.2). Of the total number of motifs (n = 109) 35 known *cis*-elements were detected in at least one of the promoter sequences of clustered genes with significantly altered expression (Tab. C.3). To assess whether the presence of certain regulatory sequences explains the distinct expression profiles, we initially focused on *cis*-elements known or predicted to be involved in auxin responses such as different varieties of the auxin responsive element (AuxRE), the E-box/hormone up at dawn (HUD) element and the TGA2 binding site motif (Keilwagen et al., 2011; Liu et al., 1994; Nemhauser et al., 2004; Vert et al.,

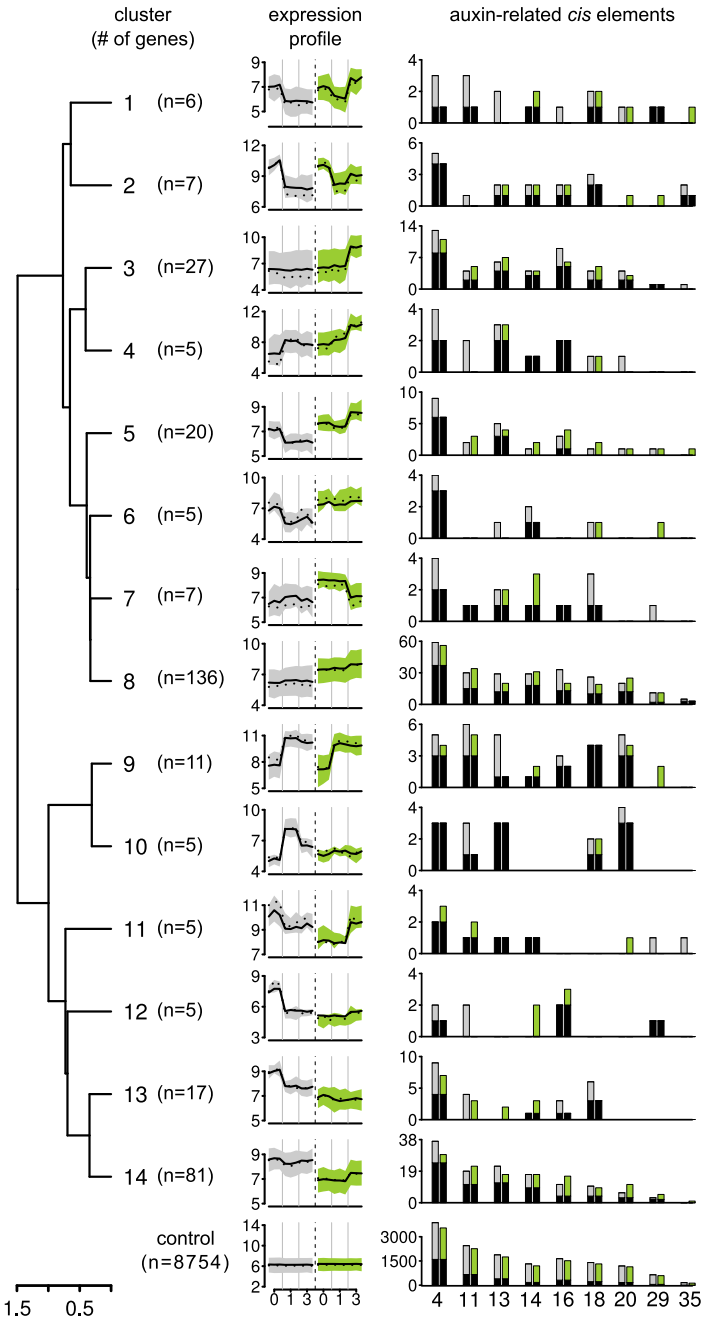


Figure 6.3.: Cluster analysis of auxin-regulated genes and allocation of known *cis*-regulatory elements. Hierarchical clustering of genes that showed an auxin-induced expression response (coefficient of variation (cv) > 0.5) in at least one species at one time point of auxin treatment using a modified Pearson correlation (1-*mod.r*) among expression profiles as distance measure. A threshold of 1-*mod.r* $= 0.3$ provided 14 clusters. Expression profiles show mean (solid lines) and median (dotted lines) expression levels of genes in one cluster. Areas shaded in grey and green denote interquartile ranges for *A. thaliana* and *A. lyrata*, respectively. Bar plots illustrate the presence of known *cis*-element sequences with functional relevance in auxin biology. "4": AATAAG, "11": TGTCTC, "13": CACATG, "13": CGTG[TC]G, "16": CACCAT, "18": TGTCTG, "20": TGT[CG]T[CG][CGT]C, "29": TGTATATAT, and "35": ATACGTGT. A full description of *cis*-elements is shown in Tab. C.2 and C.3. A comprehensive analysis of the presence of known regulatory sequences is depicted in Fig. C.3).

2008).

Auxin-related *cis*-regulatory elements were detected in all of the clusters. There was a certain degree of redundancy in the analysis due to sequence overlaps among differently labeled or modified sequences of elements, e.g., in various versions of the AuxRE (Nos. 11, 18, and 20, Fig. 6.3, Tab. C.3). Yet, neither the frequency of AuxREs nor any other *cis*-element seemed to explain the similarities or differences in the expression behavior (i.e., auxin response pattern) of the gene clusters (Fig. 6.3, Fig. C.3). Even for cluster 9, which shows clearly up-regulated profiles in both species and includes several prominent auxin-responsive genes, only roughly 50% of the genes contained a version of the AuxREs. This observation is in accordance with several previous studies in *A. thaliana* which showed a lack of AuxREs in a substantial number of auxin-regulated genes (Nemhauser et al., 2004). Furthermore, expression differences among *A. thaliana* and *A. lyrata* did not show a clear pattern of correlation to the species-specific presence of individual regulatory elements in the promoters of *A. thaliana* (gray) or *A. lyrata* (green). However, these observations remain subjective as statistical tests for over- or under-representation of elements are hindered by the low number of genes present in several of the clusters identified here.

6.4.5. Expression divergence vs. promoter divergence

The lack of any obvious correlation of known *cis*-elements and auxin-induced expression patterns prompted a *de novo* search for putatively new regulatory sequences. The data set seemed ideal as the two *Arabidopsis* species are distant enough to provide considerable sequence variation in promoter regions while providing sufficient similarities to allow for local alignments of the sequences (Hu et al., 2011).

However, a prerequisite for this approach would be a general correlation between the diversity in the promoter sequence and the differences detected on the expression level. To evaluate this assumption, we compared promoters of three groups of genes: (i) the set of conserved genes with a significant induction in expression in response to 1 h of auxin treatment in both species (n = 68), (ii) promoters of genes that are up-regulated in at least one of the analyzed species (n = 297) which include also the 68 genes of group (i) that met the threshold of auxin-induction in both species. We retained this gene set in group (ii) as the kinetics of expression profiles might still show differences among the two species. Group (iii) included neutral genes that did not show a significant alteration in expression as a control set (n = 11195). We then calculated the expression divergence of expression profiles between each orthologous gene pair using *mod.r*. Similarities of promoter sequences were assessed by a sliding window approach to compute the correlation of the occurrence of all possible 8-mers across the promoters of orthologous genes (*kmer.r*, Vinga et al., 2003).

As expected, expression divergence for genes with a conserved up-regulation in both species is rather low and seems to be independent of promoter divergences (Fig. 6.4A). Similarly, no correlation among expression and promoter divergence was observed for neutral genes that did not show expression changes in response to auxin. However, for group (ii) including all genes with a differential response in at least one of the two analyzed species, a wide range in expression divergence as well as promoter divergence was observed which showed a

considerably higher correlation compared to the other two gene sets (Fig. 6.4A). Hence, both auxin-responsive gene sets showed the expected pattern of relationships between expression and promoter divergence, which made them suitable candidate sets for *de novo* identification of regulatory promoter elements.

6.4.6. De novo identification of putative cis-regulatory elements

Based on the promoter divergence analysis we selected two gene sets for motif discovery (positive data sets). The first set comprised an extended set of genes that were induced in both species after 1 h of auxin treatment. As we did not limit the selection by filtering via corrected p-values, this set extended the previously shown set of genes of up-regulated in both species to a total of 81 orthologous gene pairs. Data set 2 comprises promoters of an extended set of genes that were up-regulated in at least one species. For this second data set we only included the promoter sequence of the species that showed a significant up-regulation of a gene in response to auxin (n = 845 promoter sequences). The corresponding promoter sequence of the other species of an orthologous gene pair was included in the control data set 2 following the rationale that regulatory elements required for the auxin response are absent in this case.

Applying the discriminative motif discovery tool *Dimont* (Grau et al., 2013), we identified motifs with significant over-representation in each of the two data sets of auxin-induced genes in comparison to their respective control data sets (see Methods for details). Among the motifs identified in both data sets were sequences with high to medium similarities to TATA box elements (Fig. 6.4B, motifs A-C). TATA boxes are present in approximately 28 % of all *Arabidopsis* genes with a predominance of non-housekeeping genes (Molina et al., 2005). Interestingly, yeast genes containing a TATA box showed increased inter-species variation in expression responses to a variety of environmental stresses (Tirosh et al., 2006). It was hypothesized that core promoters including a TATA box might be more sensitive to genetic perturbations and could be a driving factor in expression divergence (Tirosh et al., 2006). As TATA-like elements were enriched in both analyzed data sets they might rather reflect the general rapid and partially strong induction of these genes in response to an external stimulus. In yeast, TATA-containing promoters showed a slightly higher tendency for higher expression after a heat shock (Kim et al., 2004). The identification of novel variants of AuxRE- and HUD-like motifs (Fig. 6.4B, motifs D - F) corresponds with their previously demonstrated function in auxin-mediated expression induction (Walcher et al., 2012). The identification of these putatively novel variations of known elements may indicate a higher tolerance for sequence variation in the *cis*-regulatory motif that only becomes evident with a higher degree of genetic variation among genome sequences included in this analysis. Recent advances in understanding the mode of ARF transcription factor binding to target promoter sequences substantiates this assumption. Structure-function analysis indicated that different ARF proteins seem to have altered affinities for different variations of AuxREs (Boer et al., 2014). These specificities could account at least partially for functional specifications of individual ARFs and might also be a contributing factor in natural variation of transcriptional auxin responses.

In addition, other putatively novel *cis*-regulatory sequences were found to be significantly enriched in genes that were induced by auxin in both species (Fig. 6.4 and Fig. C.4, motifs

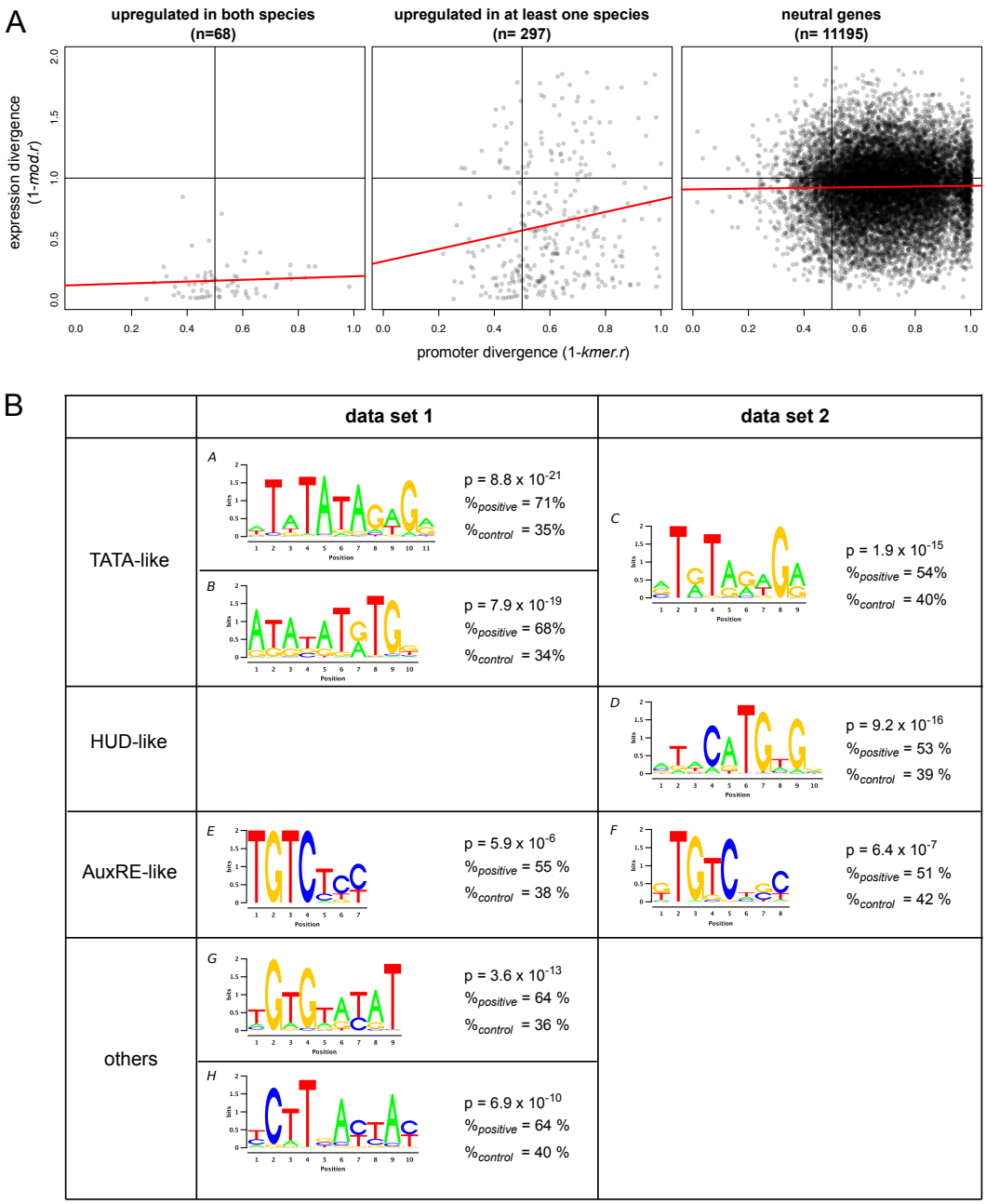


Figure 6.4.: *De novo* identification of promoter elements. (A) Analysis of promoter and expression diversity in genes that are significantly up-regulated in both species, up-regulated in either *A. thaliana* or *A. lyrata* or non-responsive (neutral) to 1 hour of auxin treatment. Divergence among expression profiles and promoter sequences was assessed by *mod.r* correlation of expression profiles and 8-mer sliding window correlation (*kmer.r*) results of promoter sequences, respectively. (B) *De novo* identification of putative *cis*-regulatory elements significantly overrepresented in auxin-induced genes identified using *Dimont*. Motifs shown were significantly enriched in genes up-regulated in both species (data set 1) or in at least one species (data set 2). Motifs were additionally tested for enrichment in an independent auxin-induced expression data set of *A. thaliana* (see p' values in Fig. C.4). Frequency of occurrence [%] in the positive and control data sets are denoted by $\%_{\text{positive}}$ and $\%_{\text{control}}$, respectively.

G-L). To the best of our knowledge, these sequences have not been described previously. To assess the potential significance of these elements with respect to auxin responses, we tested whether they were also enriched in auxin-induced genes in an independent auxin-response tran-

scriptome data set generated for *A. thaliana* seedlings (Nemhauser et al., 2006). Two of the sequences (Fig. 6.4B, motifs G+H) were indeed found to be enriched significantly ($p < 0.05$) in differentially expressed genes in this additional data set (Fig. C.4), highlighting their potential relevance for auxin-induced transcriptional regulation. We then inspected whether the presence/absence of any of the *de novo*-identified promoter sequences can account for the differential expression responses or levels of distinct gene clusters (Fig. C.3). However, similarly to the analysis of already known *cis*-elements taken from literature or databases, no coincidence pattern of *de novo* promoter elements and expression response could be identified despite the enrichment of these sequences in auxin-regulated genes. While we cannot exclude that the newly identified promoter sequences may be of minor functional relevance, the analysis as a whole rather points towards a highly complex orchestration of auxin-induced expression responses involving multiple *cis*-element variations. The identification of sequences with homology to AuxREs and HUD indicates a general success in the analytical approach. The diversity in auxin-induced expression responses via combinations of multiple different transcription factors and their individual target promoter sequences has been shown previously in case of the AuxRE and HUD elements (Nemhauser et al., 2006). Unraveling the combinatorial code of regulatory elements will require highly sophisticated bioinformatics approaches, a higher number of transcription profile data sets from diverse genetic backgrounds for in-depth phylogenetic footprinting analyses and ultimately extensive functional validation.

While the complex promoter code of auxin-induced transcriptional variation remains somewhat elusive, the general hierarchy of the auxin signal transduction pathway is well known. Transcriptional responses to auxin are primarily mediated via the TIR1/AFB-AUX/IAA-ARF signaling pathway. All three components are encoded by gene families. Individual members of these families seem to have partial redundancies in their spatio-temporal expression patterns and have at least partially distinct biochemical properties (Okushima et al., 2005; Paponov et al., 2008; Rademacher et al., 2011; Parry et al., 2009; Calderón Villalobos et al., 2012). As quantitative alterations in the equilibrium of these signaling components may significantly affect downstream responses, we next focused on this particular group of genes specifically.

6.4.7. Divergence of AUX/IAA gene expression is reflected in downstream responses

Highly diverse co-expression profiles of signaling components have been shown previously in a comparison among seven different accessions of *A. thaliana* (Delker et al., 2010). Variation in gene expression levels and co-expression patterns are indicative of altered levels of individual signaling proteins that might contribute to the differential responses observed initially on transcriptional and ultimately also a on physiological levels (Delker et al., 2010). Differential expression was predominantly evident for *AUX/IAA* genes which are generally more responsive to auxin treatment than *ARF*s or *TIR1/AFB*s (Paponov et al., 2008). Alterations in AUX/IAA protein levels will likely impact on auxin sensing by affecting the availability of individual auxin co-receptor complexes with potentially specific auxin sensitivities. Furthermore, preferential formation of specific ARF-AUX/IAA heteromerizations may affect transcriptional regulation. As such, the intra-specific comparison of auxin-regulated expression responses in *A. thaliana* accessions highlighted the early auxin signaling network as a potential source for

the observed variation in downstream responses (Delker et al., 2010). In this study, we challenged this hypothesis by inspecting the expression responses of the core auxin signaling gene families in the cross-species comparison of auxin responses.

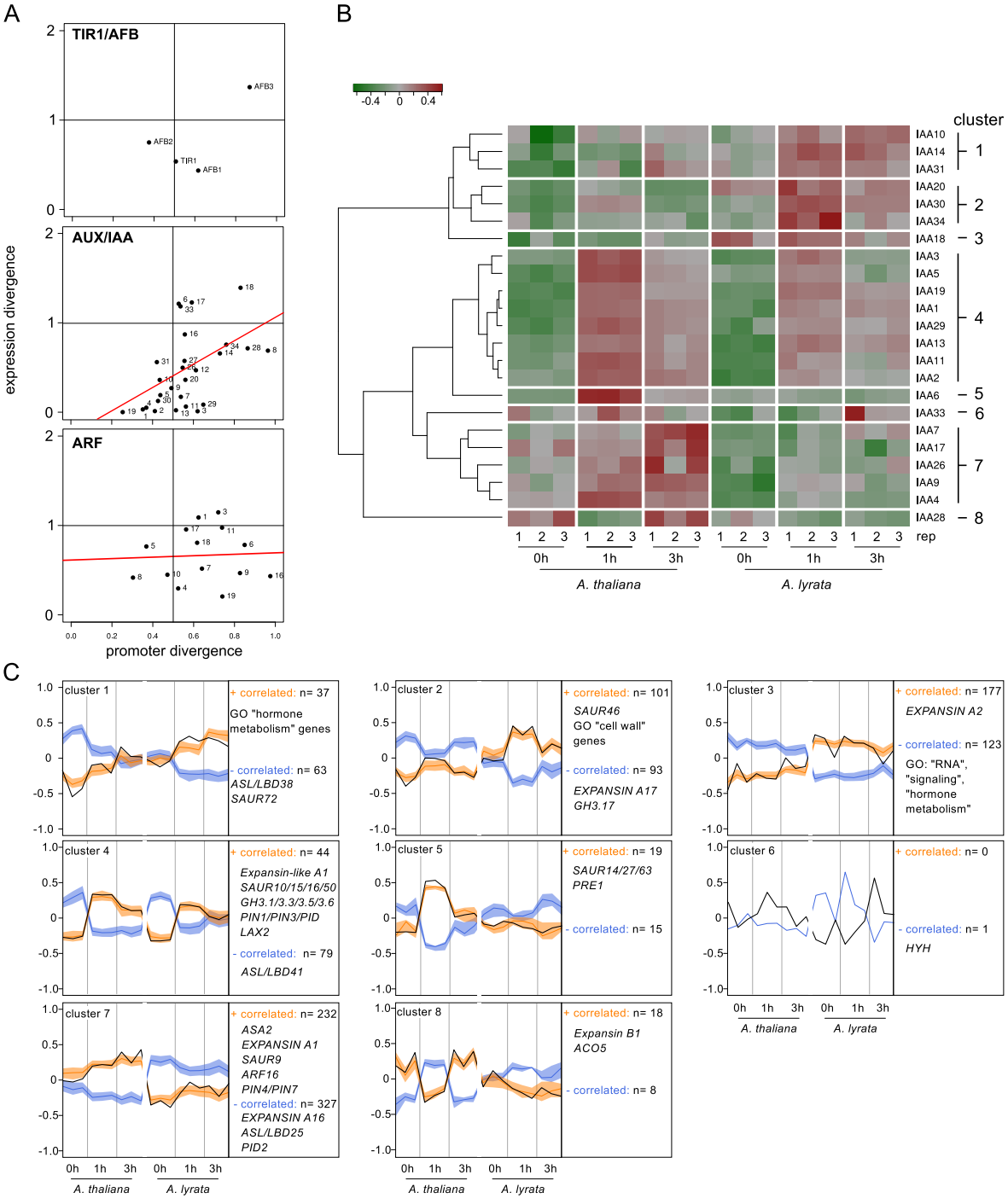


Figure 6.5.: *AUX/IAA* expression divergence correlates with downstream expression profiles. (A) Promoter divergence for core auxin signaling genes was determined as described in Fig. 6.4A. (B) Hierarchical clustering of PIF-normalized (mean-centered) *AUX/IAA* expression profiles using 1-*mod.r* as distance measure. (C) Selected genes with expression profiles that are positively (+) or negatively (-) correlated to the mean expression profiles (solid black lines) of *AUX/IAA* clusters shown in B as determined by the profile interaction finder (PIF) algorithm. A complete list of identified genes is presented as Data file C.3.

Members of all three gene families showed differential expression responses between the two species. Analysis of expression and promoter divergences showed a considerably stronger correlation for the highly auxin-responsive *AUX/IAA* gene family (Fig. 6.5A). This might be similar for the *TIR/AFB* family but the total number of only four genes retained in this analysis is generally low and effects by individual outliers might be high. While promoter divergences of *ARF* family members are also quite high, expression divergence is only low to medium (1-*mod.r* values in expression divergence from 0-1, Fig. 6.5A). AUX/IAAs have a unique role among the signaling components. Apart from their dual function in signaling as repressors of ARF transcription factors and co-receptors of auxin, they also constitute a group of classic and conserved auxin response genes which provide a readout for auxin responsiveness (Tab. 6.1, Paponov et al., 2008). Due to this prominent role, we inspected the expression responses of the *AUX/IAA* gene family in more detail. Hierarchical clustering allowed the identification of *AUX/IAA* subgroups based on the correlation (1-*mod.r*) in expression profiles (Fig. 6.5B). While cluster 1 - 3 contained *AUX/IAA* genes that were specifically induced by auxin in *A. lyrata*, cluster 5 - 8 *AUX/IAA* genes responded primarily in *A. thaliana*. In contrast, cluster 4 contained *AUX/IAA* genes that showed significantly changed expression levels in response to auxin treatment in both species. These genes are part of the conserved auxin-response gene set (Tab. 6.1) and form the largest cluster among the *AUX/IAA* genes (Fig. 6.5B). Consequently, *AUX/IAA* genes with similar expression profiles in *A. thaliana* and *A. lyrata* are indicative for similar upstream transcriptional activation/signaling events and their corresponding gene products can be speculated to have similar downstream signaling effects. In contrast to that, gene clusters with species-specific auxin responses could be indicative for the sources of natural variation seen in downstream auxin responses.

To identify genes with expression profiles that are either positively or negatively correlated to individual *AUX/IAA* gene clusters (Data file C.3), we used the recently introduced Profile Interaction Finder (PIF) algorithm (Poeschl et al., 2014). As expected, members of several of the classic and conserved auxin response gene families showed positively correlated expression profiles to cluster 4 (Fig. 6.5C). This cluster shows a classic response profile of transient expression induction in both species. The respective *AUX/IAA* and co-regulated genes of known auxin-related genes seem to be part of a conserved auxin response in both species.

Clusters with more species-specific expression responses also showed correlations with genes relevant for auxin biology. For example, the expression profile of cluster 7 shows a higher expression and gradual auxin induction in *A. thaliana*, while the expression levels in *A. lyrata* are generally lower. A similar, positively correlated pattern in expression was observed for several auxin-relevant genes ranging from biosynthesis (*ASA2*), to signaling (*ARF16*), transport (*PIN4, PIN7*), and response (*EXPANSIN A1*). In addition, genes with negatively correlated expression profiles were also identified (e.g. *ASL/LBD25*).

The positive and negative correlation of numerous auxin-associated genes with *AUX/IAA* gene clusters indicates that variation in early auxin signaling may penetrate to downstream response levels. Ultimately, these differences could quantitatively contribute to the variation observed on physiological levels. Whether the major source of variation is actually caused by differential expression or rather by altered biochemical properties due to non-synonymous mutations of signaling genes remains to be elucidated. For example, the genome-wide variation

in auxin-induced gene expression may originate in the differential gene regulation and subsequent protein levels of AUX/IAAs themselves. Alternatively and/or in addition, differential upstream events such as auxin sensing or initial gene activation may be the actual source of initial variation which then results in differential activation of *AUX/IAA*s and other genes.

6.5. Summary and conclusions

We studied natural inter-species variation of physiological and transcriptional auxin responses to assess whether the highly conserved auxin signaling and response pathway might contribute to adaptive processes in growth and development. Transcriptome analysis allowed the identification of genes with a highly conserved response to the auxin treatment which included members of known auxin-responsive gene families and so far uncharacterized genes alike. However, the majority of differentially expressed genes in response to auxin showed significant variation in expression levels and/or response pattern between the two *Arabidopsis* species. Neither similar nor species-specific expression patterns of auxin-regulated gene clusters could be explained by the presence of individual known or *de novo*-identified promoter elements. Thus, it remains likely that a sophisticated code of element combinations accounts for the diversity in transcriptional auxin responses. Breaking this particular code will require extensive efforts by bioinformaticians and far more available expression data from genetically diverse backgrounds.

A significant source for variation in auxin-induced transcriptome changes likely originates within the initial auxin signal transduction pathway itself. Distinct patterns of *AUX/IAA* gene cluster expressions were found to penetrate to the level of numerous response genes, many of which with a known functional relevance for auxin biology. Our analysis has spotlighted the triumvirate of TIR1/AFBs, AUX/IAAs, and ARFs as substantial initiators for variation in downstream auxin signaling and response, highlighting this group of gene families as potential targets of adaptation.

6.6. Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (Qu 141/3-1 to MQ).

6.7. References

Boer, D. R., Freire-Rios, A., Berg, W. A. M. van den, Saaki, T., Manfield, I. W., Kepinski, S., López-Vidrieo, I., Franco-Zorrilla, J. M., Vries, S. C. de, Solano, R., Weijers, D., and Coll, M. (2014). Structural Basis for DNA Binding Specificity by the Auxin-Dependent ARF Transcription Factors. *Cell*, 156 (3), pp. 577–589.

Calderón Villalobos, L. I. A., Lee, S., De Oliveira, C., Ivetac, A., Brandt, W., Armitage, L., Sheard, L. B., Tan, X., Parry, G., Mao, H., Zheng, N., Napier, R., Kepinski, S., and Estelle, M. (2012). A combinatorial TIR1/AFB–Aux/IAA co-receptor system for differential sensing of auxin. *Nature Chemical Biology*, 8 (5), pp. 477–485.

Chapman, E. J. and Estelle, M. (2009). Mechanism of Auxin-Regulated Gene Expression in Plants. *Annual Review of Genetics*, 43 (1). PMID: 19686081, pp. 265–285.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W.-R. (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in *Arabidopsis*. *Plant Physiology*, 139 (1), pp. 5–17.

Delker, C., Pöschl, Y., Raschke, A., Ullrich, K., Ettingshausen, S., Hauptmann, V., Grosse, I., and Quint, M. (2010). Natural Variation of Transcriptional Auxin Response Networks in *Arabidopsis thaliana*. *The Plant Cell*, 22 (7), pp. 2184–2200.

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005a). The F-box protein TIR1 is an auxin receptor. *Nature*, 435 (7041), pp. 441–445.

Dos Santos Maraschin, F., Memelink, J., and Offringa, R. (2009). Auxin-induced, SCFTIR1-mediated poly-ubiquitination marks AUX/IAA proteins for degradation. *The Plant Journal*, 59 (1), pp. 100–109.

Eklund, A. (2015). *beeswarm: The Bee Swarm Plot, an Alternative to Stripchart*. R package version 0.2.0.

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences*, 111 (6), pp. 2367–2372.

Grau, J., Posch, S., Grosse, I., and Keilwagen, J. (2013). A general approach for discriminative de novo motif discovery from high-throughput data. *Nucleic Acids Research*, 41 (21), e197.

Guilfoyle, T. J., Ulmasov, T., and Hagen, G. (1998). The ARF family of transcription factors and their role in plant hormone-responsive transcription. *Cellular and Molecular Life Sciences*, 54 (7), pp. 619–627.

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., Fahlgren, N., Fawcett, J. A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J. D., Ossowski, S., Ottilar, R. P., Salamov, A. A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M. E., Bergelson, J., Carrington, J. C., Gaut, B. S., Schmutz, J., Mayer, K. F. X., Van de Peer, Y., Grigoriev, I. V., Nordborg, M., Weigel, D., and Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, 43 (5), pp. 476–481.

Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J. L., Nishihama, R., and Kohchi, T. (2015). Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in *Marchantia polymorpha*. *PLoS Genet*, 11 (5), e1005084.

Keilwagen, J., Grau, J., Paponov, I. A., Posch, S., Strickert, M., and Grosse, I. (2011). De-Novo Discovery of Differentially Abundant Transcription Factor Binding Sites Including Their Positional Preference. *PLoS Computational Biology*, 7 (2), e1001070.

Kepinski, S. and Leyser, O. (2005). The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature*, 435 (7041), pp. 446–451.

Kim, J. and Iyer, V. R. (2004). Global Role of TATA Box-Binding Protein Recruitment to Promoters in Mediating Gene Expression Profiles. *Molecular and Cellular Biology*, 24 (18), pp. 8104–8112.

Lincoln, C., Britton, J. H., and Estelle, M. (1990). Growth and development of the axr1 mutants of *Arabidopsis*. *The Plant Cell*, 2 (11), pp. 1071–80.

Liu, Z. B., Ulmasov, T., Shi, X., Hagen, G., and Guilfoyle, T. J. (1994). Soybean GH3 promoter contains multiple auxin-inducible elements. *The Plant Cell*, 6 (5), pp. 645–57.

Molina, C. and Grotewold, E. (2005). Genome wide analysis of *Arabidopsis* core promoters. *BMC Genomics*, 6, pp. 25–25.

Nemhauser, J. L., Hong, F., and Chory, J. (2006). Different Plant Hormones Regulate Similar Processes through Largely Nonoverlapping Transcriptional Responses. *Cell*, 126 (3), pp. 467–475.

Nemhauser, J. L., Mockler, T. C., and Chory, J. (2004). Interdependency of Brassinosteroid and Auxin Signaling in *Arabidopsis*. *PLoS Biol*, 2 (9), e258.

Okushima, Y., Overvoorde, P. J., Arima, K., Alonso, J. M., Chan, A., Chang, C., Ecker, J. R., Hughes, B., Lui, A., Nguyen, D., Onodera, C., Quach, H., Smith, A., Yu, G., and Theologis, A. (2005). Functional Genomic Analysis of the AUXIN RESPONSE FACTOR Gene Family Members in *Arabidopsis thaliana*: unique and Overlapping Functions of ARF7 and ARF19. *The Plant Cell*, 17 (2), pp. 444–463.

Opgen-Rhein, R. and Strimmer, K. (2007). Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach. *Statistical Applications in Genetics and Molecular Biology*, 6 (1), pp. 477+.

Paponov, I. A., Paponov, M., Teale, W., Menges, M., Chakrabortee, S., Murray, J. A. H., and Palme, K. (2008). Comprehensive Transcriptome Analysis of Auxin Responses in *Arabidopsis*. *Molecular Plant*, 1 (2), pp. 321–337.

Parry, G., Calderon-Villalobos, L. I., Prigge, M., Peret, B., Dharmasiri, S., Itoh, H., Lechner, E., Gray, W. M., Bennett, M., and Estelle, M. (2009). Complex regulation of the TIR1/AFB family of auxin receptors. *Proceedings of the National Academy of Sciences*, 106 (52), pp. 22540–22545.

Poeschl, Y., Delker, C., Trenner, J., Ullrich, K. K., Quint, M., and Grosse, I. (2013). Optimized Probe Masking for Comparative Transcriptomics of Closely Related Species. *PLoS ONE*, 8 (11), e78497.

Poeschl, Y., Grosse, I., and Gogol-Döring, A. (2014). Explaining gene responses by linear modeling. *German Conference on Bioinformatics*, Volume P-235 of Lecture Notes in Informatics (LNI) - Proceedings, pp. 27–35.

Pollard, K. S., Dudoit, S., and Laan, M. J. van der (2005). *Multiple Testing Procedures: R multtest Package and Applications to Genomics, in Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer.

Quint, M. and Gray, W. M. (2006). Auxin signaling. *Current Opinion in Plant Biology*, 9 (5), pp. 448–453.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vol. ISBN 3-900051-07-0. Vienna, Austria.

Rademacher, E. H., Möller, B., Lokerse, A. S., Llavata-Peris, C. I., Berg, W. van den, and Weijers, D. (2011). A cellular expression map of the *Arabidopsis* AUXIN RESPONSE FACTOR gene family. *The Plant Journal*, 68 (4), pp. 597–606.

Ramos, J. A., Zenser, N., Leyser, O., and Callis, J. (2001). Rapid Degradation of Auxin/Indoleacetic Acid Proteins Requires Conserved Amino Acids of Domain II and Is Proteasome Dependent. *The Plant Cell*, 13 (10), pp. 2349–2360.

Salehin, M., Bagchi, R., and Estelle, M. (2015). SCFTIR1/AFB-Based Auxin Perception: Mechanism and Role in Plant Growth and Development. *The Plant Cell*, 27 (1), pp. 9–19.

Szemenyei, H., Hannon, M., and Long, J. A. (2008). TOPLESS Mediates Auxin-Dependent Transcriptional Repression During *Arabidopsis* Embryogenesis. *Science*, 319 (5868), pp. 1384–1386.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L. A., Rhee, S. Y., and Stitt, M. (2004). MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal*, 37 (6), pp. 914–939.

Tirosh, I., Weinberger, A., Carmi, M., and Barkai, N. (2006). A genetic signature of interspecies variations in gene expression. *Nature Genetics*, 38 (7), pp. 830–834.

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. (1999). Activation and repression of transcription by auxin-response factors. *Proceedings of the National Academy of Sciences*, 96 (10), pp. 5844–5849.

Vert, G., Walcher, C. L., Chory, J., and Nemhauser, J. L. (2008). Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proceedings of the National Academy of Sciences*, 105 (28), pp. 9829–9834.

Vinga, S. and Almeida, J. (2003). Alignment-free sequence comparison—a review. *Bioinformatics*, 19 (4), pp. 513–523.

Walcher, C. L. and Nemhauser, J. L. (2012). Bipartite Promoter Element Required for Auxin Response. *Plant Physiology*, 158 (1), pp. 273–282.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. (2014). *gplots: Various R programming tools for plotting data*. R package version 2.15.0.

Weijers, D., Benkova, E., Jäger, K. E., Schlereth, A., Hamann, T., Kientz, M., Wilmoth, J. C., Reed, J. W., and Jürgens, G. (2005). Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators. *The EMBO Journal*, 24 (10), pp. 1874–1885.

Zenser, N., Ellsmore, A., Leasure, C., and Callis, J. (2001). Auxin modulates the degradation rate of Aux/IAA proteins. *Proceedings of the National Academy of Sciences*, 98 (20), pp. 11795–11800.

7. Developmental plasticity of Arabidopsis thaliana accessions across an ambient temperature range

Carla Ibañez[1,*], Yvonne Poeschl[2,3,*], Tom Peterson[1], Julia Bellstädt[1], Kathrin Denk[1], Andreas Gogol-Döring[2,3], Marcel Quint[1], and Carolin Delker[1]

[1] Department of Molecular Signal Processing, Leibniz Institute of Plant Biochemistry, Weinberg 3, 06120 Halle (Saale), Germany
[2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany
[3] Institute of Computer Science, Martin Luther University Halle-Wittenberg, 06120 Halle (Saale), Germany
[*] These authors contributed equally to this work.

7.1. Abstract

The global increase in ambient temperature constitutes a significant challenge to wild and cultivated plant species. Yet, a comprehensive knowledge on morphological responses and molecular mechanisms involved is scarce. Studies published to date have largely focused on a few, isolated temperature-relevant phenotypes such as flowering time or hypocotyl elongation. To systematically describe thermomorphogenesis, we profiled more than 30 phenotypic traits throughout an entire life cycle in ten distinct accessions of *Arabidopsis thaliana* grown in four different ambient temperatures. We observed a uniform acceleration of developmental timing in the vegetative growth phase with a low contribution of genotype effects on variation indicating a passive effect of temperature. In contrast, reproduction-associated phenotypes and several quantitative growth traits were sensitive to both, genotype and temperature effects or could be attributed primarily to either factor. Therefore, the results argue against a general mechanism of passive temperature effects by thermodynamic processes. Temperature responses of several phenotypes rather implicate differential function of specific signaling components that might be targets of adaptation to specific environmental conditions.

7.2. Introduction

Recurrent changes in ambient temperature provide plants with essential information about time of day and seasons. Yet, even small changes in mean ambient temperature can profoundly affect plant growth and development which collectively can be summarized as thermomorphogenesis. In crops like rice, a season-specific increase in the mean minimum temperature of 1 °C results in approximately a 10% reduction in grain yield (Peng et al., 2004). Similarly, up to 10% of the yield stagnation of wheat and barley in Europe over the past two decades can be attributed to climate trends (Moore et al., 2015). Current projections indicate that mean global air temperatures will increase up to 4.8 °C by the end of the century (*IPCC Climate change 2013: The physical science basis. Fifth assessment report.* 2013; Lobell et al., 2012). Global climate change will thus have significant implications on biodiversity and future food security.

Naturally, increased ambient temperatures also affect wild species and natural habitats. Long-term phenology studies of diverse plant populations have revealed an advance in first and peak flowering and alterations in the total length of flowering times (CaraDonna et al., 2014; Fitter et al., 2002). Furthermore, estimates project that temperature effects alone will account for the extinction of up to one-third of all European plant species (Thuiller et al., 2005). As the impact of changes in ambient temperature on crop plants and natural habitats emerge, a comprehensive understanding of thermomorphogenesis and developmental temperature responses becomes paramount.

Our present knowledge on molecular responses to ambient temperature signaling has largely been gained from studies in *Arabidopsis thaliana*. Model thermomorphogenesis phenotypes such as hypocotyl elongation (Gray et al., 1998), hyponastic leaf movement (Zanten et al., 2009), and alterations in flowering time have served in forward or reverse genetic approaches to identify some of the molecular signal transduction components involved in triggering thermomorphogenic responses. So far, the main molecular players identified seem to function in response to both temperature and light stimuli and form a highly interconnected network of signaling elements. Prominent members of this network are PHYTOCHROME INTERACTING FACTOR 4 (PIF4, Franklin et al., 2011; Koini et al., 2009; Proveniers et al., 2013), the DE-ETIOLATED1-CONSTITUTIVELY PHOTOMORPHOGENIC1- ELONGATED HYPOCOTYL 5 (DET1-COP1-HY5) cascade (Delker et al., 2014; Toledo-Ortiz et al., 2014) and EARLY FLOWERING 3 (ELF3) as a component of the circadian clock (Box et al., 2015; Raschke et al., 2015). In addition, considerable naturally occurring variation in thermomophogenic traits like hypocotyl elongation and flowering time has been demonstrated (Balasubramanian et al., 2006; Delker et al., 2010). This variation might be attributed to local adaptation processes of diverse *A. thaliana* accessions and indicates a high variability regarding temperature-induced phenotypic plasticity.

The use of thermomophogenic model phenotypes has undoubtedly been useful for the identification of several molecular signaling components. Meeting future challenges in plant breeding will, however, require more extensive knowledge about temperature effects on plant development and morphology beyond commonly described traits. As such, it would be vital to determine (i) which phenotypes are sensitive to ambient temperature effects, (ii) which of

these traits are robustly affected by temperature within a gene pool, and (iii) which phenotypic traits show natural variation in temperature responses and thus might be consequences of adaptation processes to cope with local climate or general environmental conditions. Robustly affected temperature response might indicate passive consequences of general thermodynamic effects. According to basic principles of thermodynamics, temperature-induced changes in free energy will affect the rates of biological reactions. As these effects should occur more generally and non-selective, phenotypic responses can be expected to occur robustly and rather independently of genetic variation. However, natural variation in thermomorphogenesis could implicate the relevance of specific signaling elements showing natural genetic variation as a consequence of adaptation. Such genes would represent attractive candidates for targeted breeding approaches.

Here, we aim to address these questions by profiling of more than 30 developmental and morphological traits of ten *A. thaliana* accessions which were grown at 16, 20, 24, and 28 °C. In addition, we provide accession-specific developmental reference maps of temperature responses that can serve as resources for future experimental approaches in the analysis of ambient temperature responses in *A. thaliana*.

7.3. Materials and methods

7.3.1. Plant material and growth conditions

A. thaliana accessions were obtained from the Nottingham Arabidopsis Stock Centre (Scholl et al., 2000). Detailed information on stock numbers and geographic origin are listed in Supplementary Tab. D.1. For seedling stage analyses, surface-sterilized seeds were stratified for 3 days in deionized water at 4 °C and subsequently placed on *A. thaliana* solution (ATS) nutrient medium (Lincoln et al., 1990). Seeds were germinated and cultivated in growth chambers (Percival) at constant temperatures of 16, 20, 24 or 28 °C under long day photoperiods (16h light/8h dark) and a fluence rate of 90 μmol·m^{-2}·sec^{-1}. We refrained from including a vernalization step because the primary focus of this study was to record morphology and development in response to different constant ambient temperature conditions.

Germination rates were assessed daily and hypocotyl, root length, and petiole angles were measured in 7 days old seedlings with ImageJ (`http://imagej.nih.gov/ij/`) and Root Detection (`http://www.labutils.de/rd.html`).
All other analyses were performed on soil-grown plants at a fluence rate of 140 μmol·m^{-2}·sec^{-1}. After imbibition for 3 days at 4 °C, seeds were grown in individual 5 × 5 cm pots, which were randomized twice a week to minimize position effects. Relative humidity of growth chambers was maintained at 70% and plants were watered by subirrigation. Plants were photographed daily for subsequent determination of phenotypic parameters using Image J (`http://imagej.nih.gov/ij/`). At transition to the reproductive growth phase, the number of leaves was determined by manual counting in addition to recording the days after germination.

Spectrophotometric determination of chlorophyll content was performed as described in Porra et al. (1989). Rates of germination and seedling establishment were determined from ∼100 individual seeds. Two different seed pools were generated by proportional merging of four different seed batches from individuals from one accession (1:1:1:1). Both sample pools were used in the actual experiments. Sterilized and stratified seeds were germinated on ATS medium without sucrose. Germination was determined in the first three days and seedling establishment data was recorded at day six. Morphological markers for germination and seedling establishment are described in Table 7.1. Data were recorded from three independent germination experiments of which one representative set is shown.

7.3.2. Data analysis

Data visualization and statistical analyses of the data were performed using the software R (R Core Team, 2012). For visualization of the data set, box plots were generated using the *boxplot* function contained in the graphics package. For visualization of the statistical measures, heat maps were generated using the *heatmap.2* function contained in the gplots package, which is available on `http://cran.r-project.org`.

7.3.3. ANOVA for single factors

ANOVAs for a single factor (either accession or temperature) were done using the *anova* function contained in the R stats package. In case of temperature, the factor had four levels. In case of accession, the factor had ten levels. As post hoc test Tukey's 'Honest Significant Difference' test was used to determine the pairs of factor levels that are significantly different. To perform this test, the function *TukeyHSD* contained in the stats package was used.

7.3.4. Calculation of intraclass correlation coefficients λ

In order to quantify the distinct influences of genotype and temperature on a given phenotype, we determined intraclass correlation coefficients λ_{gen} and λ_{temp} using the ANOVA framework similar to (Donner et al., 1980). This involved the calculation of sum of squared differences *SSD* values, which are defined for a set of data points $M = \{x_1, x_2, \ldots, x_m\}$ as $SSD(M) = \sum(x_i - \bar{x})^2$, $\bar{x} = \frac{1}{m}\sum x_i$ is the mean of all values in M. In the case of λ_{temp} we split all data points M corresponding to a given phenotype and genotype into four groups M_{16}, M_{20}, M_{24}, and M_{28} according to the temperatures. The total variation of the data given by the $SSD_{\text{total}} = SSD(M)$ is the composition of two components, namely the variation between the groups SSD_{between} representing the effect of the temperature, and the variation inside of the groups SSD_{within} representing the accession-specific biological variability. The latter component can be calculated by adding up the *SSD* values computed separately for each groups, i.e., $SSD_{\text{within}} = SSD(M_{16}) + SSD(M_{20}) + SSD(M_{24}) + SSD(M_{28})$, while the former is given by $SSD_{\text{between}} = SSD_{\text{total}} - SSD_{\text{within}}$. We defined the value λ_{temp} to be the fraction of variation due to the temperature, i.e., $\lambda_{\text{temp}} = SSD_{\text{between}}/SSD_{\text{total}}$. Accordingly, the fraction of variation due to the genotype λ_{gen} was calculated by splitting the set of data points

M corresponding to a given phenotype and temperature into ten groups according to the accessions.

7.3.5. Regression analysis

Linear regression analyses were conducted using the *lm* function contained in the stats package to get a trend of the temperature effect. The slope of the resulting regression line was used to determine the direction (and strength) of the effect caused by temperature (for a specific phenotype).

7.4. Results

To assess phenotypic plasticity in a range of ambient temperatures, *A. thaliana* plants were cultivated throughout an entire life cycle at four different temperatures (16, 20, 24 and 28 °C) under otherwise similar growth conditions (see Materials and methods for further details). More than 30 morphological and development-associated traits were recorded in the vegetative and reproductive growth phases (Tab. 7.1).

7.4.1. Temperature responses in the A. thaliana reference accession Col-0

In Col-0, almost all phenotypes analyzed in this study were affected by the cultivation in different ambient temperatures. Only seed weight and maximum height remained constant regardless of the growth temperature (Fig. 7.1A, Supplementary Fig. D.1). Among the temperature sensitive traits were several growth-associated phenotypes in early vegetative stages. Primary root length, hypocotyl and petiole elongation all increased with elevated temperatures which concurs with previously published results (Gray et al., 1998; Zanten et al., 2009). As a further example, yield-related traits, such as the number of siliques per plant and the number of seeds per silique decreased with an increase in ambient temperature (Fig. 7.1A).

As reported previously, Col-0 plants showed a decrease in developmental time until flowering with increasing ambient temperatures (Balasubramanian et al., 2006). The transition from vegetative to reproductive phase occurred about 25 days earlier at 28 °C than at 16 °C (Fig. 7.1B). Similarly, the number of rosette leaves developed at time of bolting differed by 26 leaves between 28 °C and 16 °C (Fig. 7.1A).

The fact that only a very limited number of phenotypes was insensitive to cultivation in different temperatures clearly illustrates the fundamental impact of ambient temperature on plant growth and development.

Table 7.1.: Growth and development phenotypes analyzed for temperature sensitivity

Trait	Morphological marker/time point	Units	#Traits
Developmental data			
Germination			
Germination time	radicle emergence	days	1
Seedling establishment	cotyledons opened fully	days	2
Leaf production			
2 rosette leaves	rosette leaves > 1 mm in length	days	3
3 rosette leaves	rosette leaves > 1 mm in length	days	4
4 rosette leaves	rosette leaves > 1 mm in length	days	5
5 rosette leaves	rosette leaves > 1 mm in length	days	6
6 rosette leaves	rosette leaves > 1 mm in length	days	7
7 rosette leaves	rosette leaves > 1 mm in length	days	8
8 rosette leaves	rosette leaves > 1 mm in length	days	9
9 rosette leaves	rosette leaves > 1 mm in length	days	10
10 rosette leaves	rosette leaves > 1 mm in length	days	11
11 rosette leaves	rosette leaves > 1 mm in length	days	12
12 rosette leaves	rosette leaves > 1 mm in length	days	13
13 rosette leaves	rosette leaves > 1 mm in length	days	14
14 rosette leaves	rosette leaves > 1 mm in length	days	15
Reproductive development			
Inflorescence emergence	First flower buds visible	days	16
Flowering time_days	Bolt >1 cm	days	17
Flowering time_n leaves	Bolt >1 cm	n° leaves	18
Flowering time_first flower	First flower full opened	days	19
Siliques production	First silique appear	days	20
Quantitative/morphometric phenotypes			
Vegetative stage			
Hypocotyl length	7 days old seedlings	pixels	21
Petiole angle	7 days old seedlings	°	22
Length of primary root	7 days old seedlings	pixels	23
Petiole length	20 days old seedlings	pixels	24
Chlorophyll content	14 days old seedlings	μg/mg leave	25
Foliar surface	Bolt >1 cm	mm^2	26
Senescence			
Total no. of siliques per plant	First silique shattered	Count	27
Max. plant height	First silique shattered	cm	28
Seed phenotype			
Seed area		pixels	29
Seed length		pixels	30
Seed weight		μgr.	31
Total no. of seeds per plant		Count	32
Total no. of seeds per silique		Count	33
Silique length		mm	34

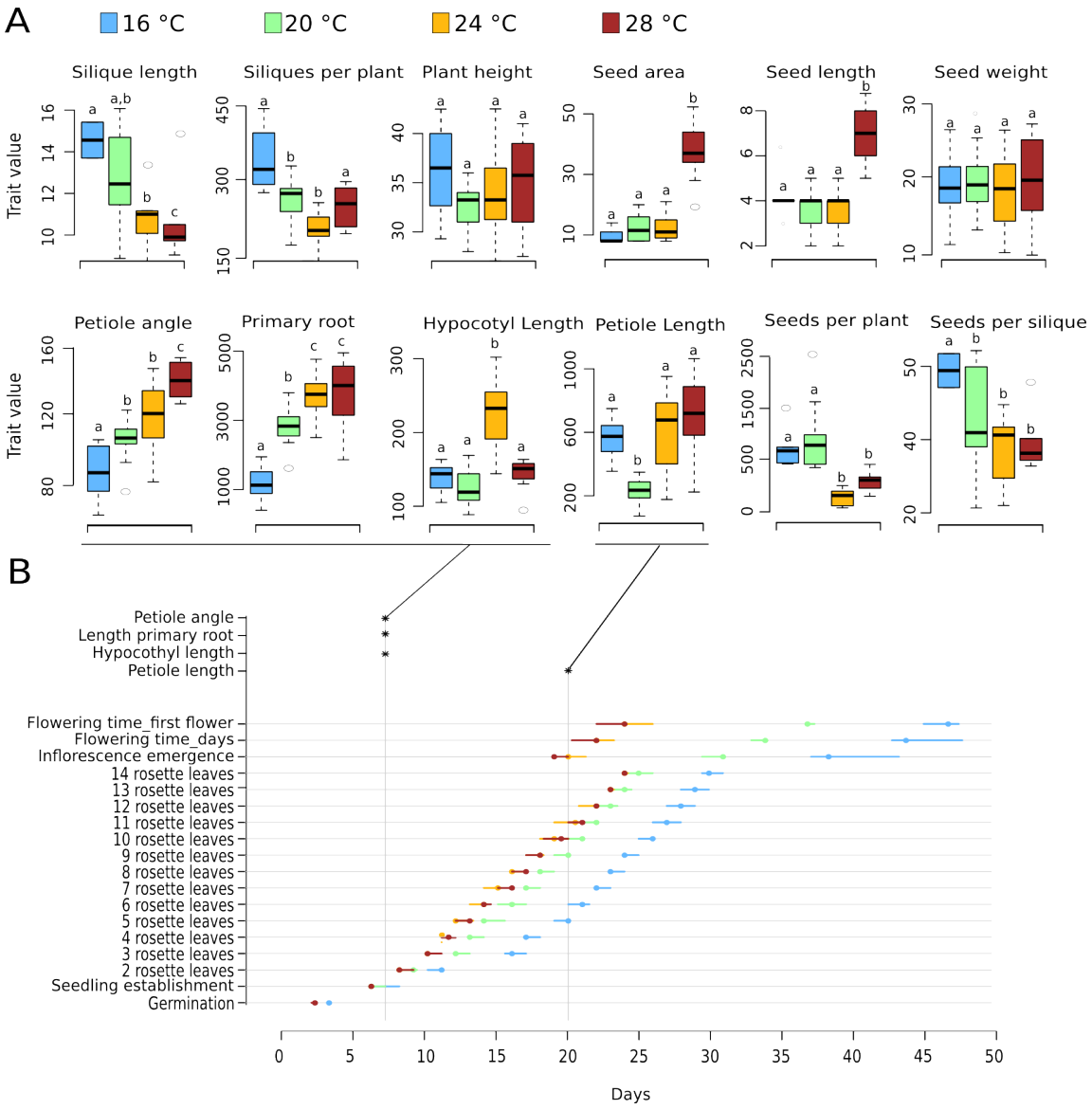


Figure 7.1.: Col-0 growth and development in response to different ambient temperatures. (A) Quantification of phenotypic traits recorded at different growth temperatures. Box plots show median and interquartile ranges (IQR), outliers (> 1.5 times IQR) are shown as circles. Units for each trait are specified in Table 7.1. Different letters denote statistical differences ($P > 0.05$) among samples as assessed by one-factorial ANOVA and Tukey HSD. (B) Summary of temperature effects on developmental timing. Circles denote medians, bars denote IQRs ($n > 15$). Time of phenotypic assessment for selected traits in (A) is indicated by asterisks.

7.4.2. Natural variation of temperature responses

To assess whether the observed temperature responses in Col-0 are robust throughout the *A. thaliana* population or which of the responses are affected by natural variation, phenotypic profiling was performed in nine other *A. thaliana* accessions parallel to the analysis in Col-0 (Supplementary Tab. D.1, Fig. D.1-D.10). Although a panel of ten accession does of course not represent the world-wide *A. thaliana* gene pool in its entity, it is certainly sufficient to address the aim of this study, i.e. to identify and distinguish between traits that may be

targets for adaptation and those that are genetically fixed.

To approximate and to compare temperature sensitivity of traits among different accessions, we transformed individual trait values into temperature responses by linear regression of values across all four ambient temperature regimes (Fig. 7.2A). The slope values were then normalized to the respective trait median of all temperatures combined to allow comparison and cluster analysis of phenotypes with different dimensions of units (Fig. 7.2A).

Fig. 7.2B shows that hierarchical clustering of temperature responses (slopes) clearly separated seedling growth traits and chlorophyll content from all other phenotypes due to the strong increase of trait values with increasing temperatures. An additional cluster was constituted by phenotypes associated with the transition to reproductive development. Here, most of the accessions showed a temperature-induced reduction in time/development to flowering as indicated by negative slope values. However, in accordance with previously published results on natural variation of temperature-induced flowering (Balasubramanian et al., 2006) the strength of the response differed. Most striking in this respect was the temperature response of Rrs-7 and Got-7. In contrast to the other accessions, they showed a delay in flowering time with increasing temperature (Fig. 7.2B). Got-7 did not flower within the first 90 days of cultivation when grown in 24 or 28 °C likely caused by the lack of vernalization (Supplementary Fig. D.5). Thus, initiated leaf senescence at bolting stage prevented accurate determination of leaf number at the onset of flowering.

A third cluster is formed by traits associated with the timing of vegetative development. Negative slope values for germination and induction of rosette leaves indicate accelerated development in response to higher temperatures, which was uniformly observed in all analyzed accessions.

A direct comparison of leaf number and time of development corroborates a sudden increase in variation at the transition to flowering. However, at 16 °C and 20 °C several accessions contribute to the overall variability in the graph, whereas at 24 °C and 28 °C, C24 and Rrs-7 are the main determinants of variation due to their massive number of leaves corresponding to an extension of the vegetative growth phase (Supplementary Fig. D.11). This finding harbors several interesting aspects. First, natural variation in the transition to flowering is already observed at lower temperatures. As the flowering time differences of Rrs-7 and Got-7 (Fig. 7.2B) become pronounced primarily at temperatures above 24 °C, the general variation in flowering time seems to be largely, independent of vernalization requirements. Furthermore, C24 contributes considerably to the variability of the reproductive traits, even though the general C24 temperature response follows the common pattern of earlier transition to flowering at higher temperatures (Fig. 7.2B, Supplementary Fig. D.3).

To further substantiate this analysis and to identify specific traits with adaptive potential, we aimed to dissect and quantify the individual effects of temperature and genotype on the observed variability of each trait/phenotype in the following.

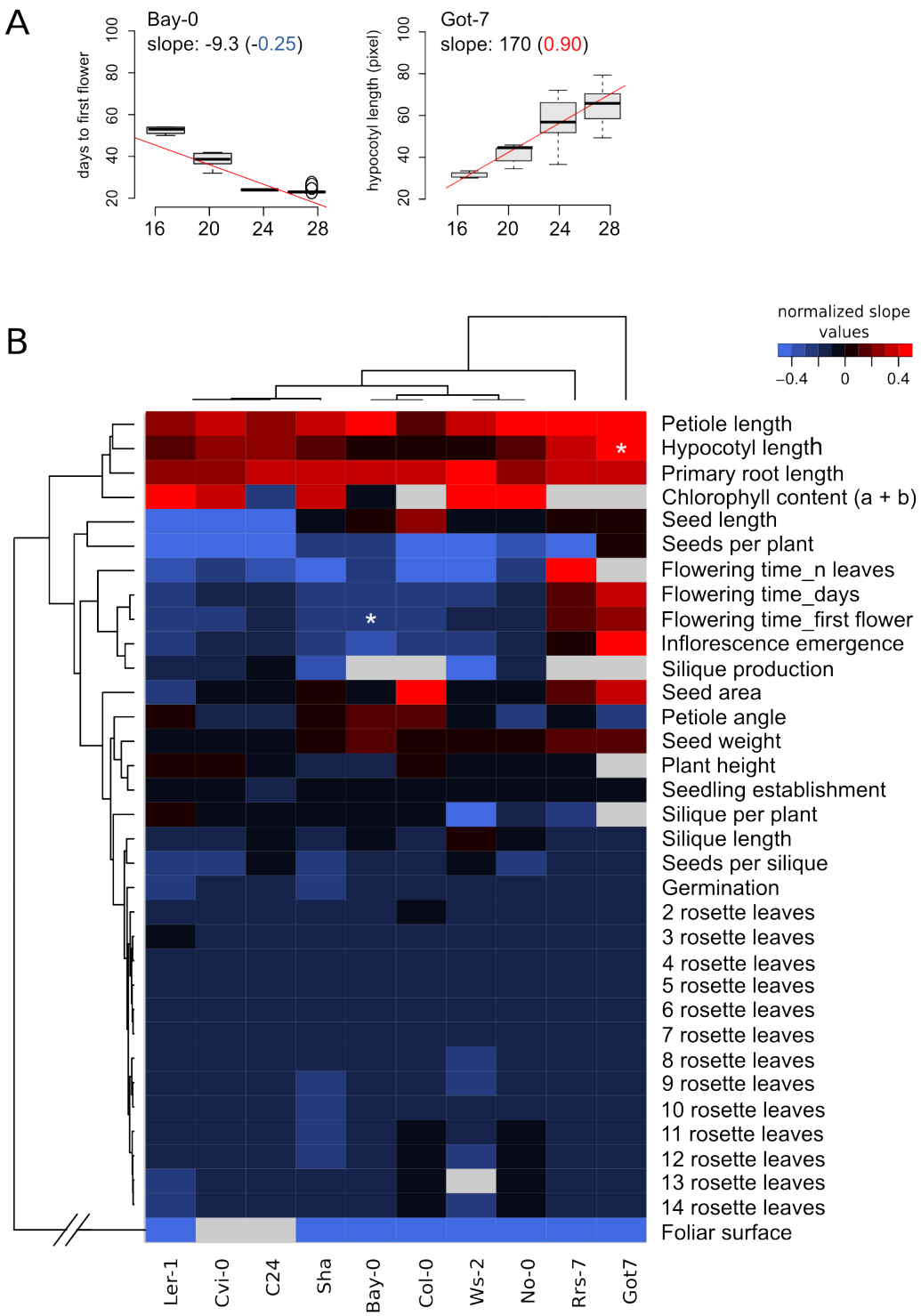


Figure 7.2.: Natural variation in temperature sensitivity of phenotypic traits. (A) Example graphs illustrating the origin of slope values (in black) for each phenotype and genotype combination. Median-normalized slope values are shown in red and blue for increasing and decreasing values, respectively and are highlighted by asterisks in (B). Corresponding figures for all other available combinations of phenotypes and genotypes are shown in Supplementary Fig. D.1-D.10. (B) Heatmap and hierarchical clustering of normalized slope values derived for each phenotype/genotype combination as indicated in (A).

7.4.3. Genotype contributions to phenotypic variation

For genotype effects, we assessed the variation that occurs within each individual accession and compared it to the total variation occurring among all accessions for each phenotypic trait at each given temperature. As a measure for variabilit we made use of the sum of squared differences (*SSD*). While the SSD_{within} represents the biological variation within an individual accession (e.g. Ler-1 or Got-7, Fig. 7.3A), SSD_{between} describes the range of variability that is observed among the mean values across the ten analyzed accessions. Values of SSD_{within} and SSD_{between} were subsequently used to obtain a unit-free measure of genotype effects on variation (λ_{gen}). While a λ_{gen} value = 1 indicates a strong genotype effect on the observed variability, no effect of natural variation on the phenotypic differences can be assumed for $\lambda = 0$ (Fig. 7.3A).

Assessing the degree of genotype effects on the overall range of phenotypic variation observed at each temperature showed highly variable patterns. Regardless of the individual temperature, genotype effects on the developmental timing throughout the vegetative phase was generally very low. This objectively supports the above described initial impression of low natural variation observed in the general temperature sensitivities of traits (Fig. 7.2B). Similarly, strong genotype effects were observed for many reproductive traits. Other phenotypes show more differential or even gradual genotype effects at different temperatures. For example, effects of natural variation on plant height, silique production and silique length decreased with an increase in temperature, whereas opposite effects are observed for hypocotyl and petiole length as well as flowering time (number of leaves). Although in some cases, such as flowering time, a strong genotype effect seems to correlate also with a strong general temperature sensitivity (Fig. 7.3B and Fig. 7.2B), this differs in case of root length. Here, only low genotype effects were observed (Fig. 7.3B), even though the phenotype was highly sensitive to a change in ambient temperature (Fig. 7.2B).

7.4.4. Temperature contributions to phenotypic variation

To further dissect and differentiate genotype and temperature effects, we also computed the degree of temperature effects (λ_{temp}) on the total variation for each of the ten accessions (Supplementary Fig. D.12A). The heatmap representation of λ_{temp} partially mirrors the λ_{gen} results, for instance in the strong temperature effect on the timing of vegetative development (Supplementary Fig. D.12A). However, many traits exhibit highly differential temperature responses among accessions. This is particularly obvious for yield-related traits such as total number of seeds per plant and silique as well as silique length. Here, temperature effects on total phenotype variation were low for Col-0, C24 and Bay-0, whereas higher λ_{temp} values were determined for the other accessions. Importantly, the latter could be of relevance for future breeding approaches. Similar distinct patterns of temperature effects were observed for a number of traits indicating a highly diverse and complex interplay of temperature and genotype effects on phenotypic plasticity.

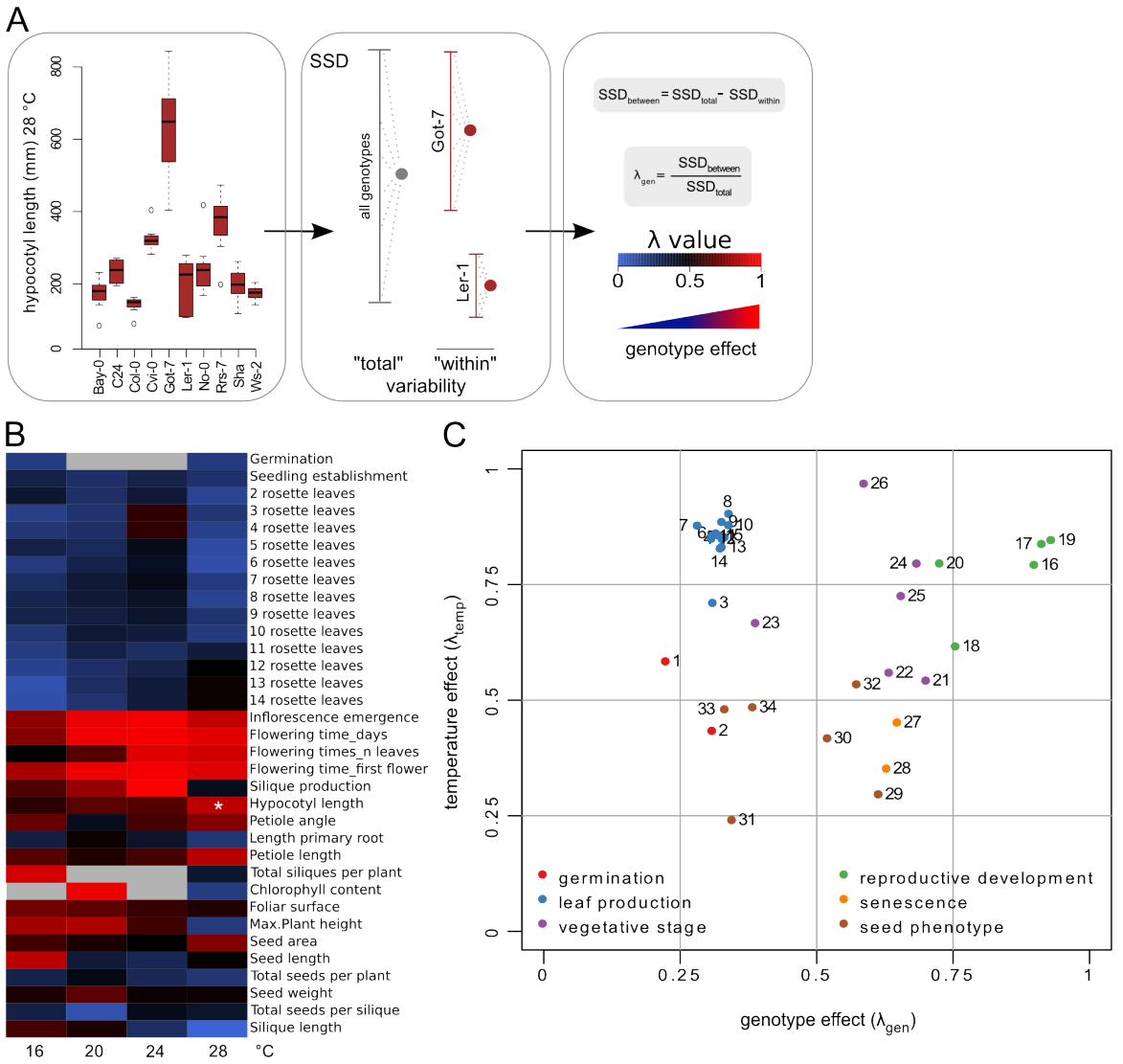


Figure 7.3.: Genotype and temperature effects on phenotypic variation. (A) Illustration of the concept of "within" and "between" variability and the calculation of genotype effects (λ_{gen}) taking hypocotyl elongation at 28 °C as an example. Variation "within" a genotype was calculated as the sum of squared differences (SSD) between individual data points of one accession to the respective accession mean (SSD_{within}) as shown for Ler-1 and Got-7 as an example. Total variation between genotypes was calculated by assessing the SSD of all data to the global mean of values of all accessions combined (SSD_{total}). Both values were used to calculate λ_{gen} which provides a measure of genotype effects on the variation observed for individual phenotypes. (B) Heat map representation of the intraclass correlation coefficient λ_{gen} of all recorded phenotypes. Missing data is shown in grey. The asterisk highlights data shown in (A). (C) Scatter plot of mean λ_{gen} and λ_{temp} values over all temperatures and accessions, respectively. Phenotypes are color-coded according to developmental stage. Heat maps of individual λ_{temp}, mean λ values and standard deviations are shown in Supplementary Fig. D.12A-C.

7.4.5. Comparison of temperature and genotype effects

To identify global effects of both contributing factors, we computed mean values for λ_{gen} across all temperatures and λ_{temp} across all accessions (Supplementary Fig. D.12B). A direct comparison of mean λ_{gen} and λ_{temp} pinpoints the predominant temperature effect on changes in

the timing of leaf development (Fig. 7.3C Supplementary Fig. D.12C). In contrast, the variation in quantitative growth phenotypes in the vegetative growth phase displayed considerably higher degrees of genotype effects with similarly high temperature effects. This combination of factorial effects is most prominent for phenotypes associated with shifts to reproductive development. Phenotypes associated with late developmental stages or senescence as well as seed phenotypes were generally less affected by both factors with a general tendency of slightly higher genotype than temperature effects (Fig. 7.3C, Supplementary Fig. D.12C).

Several yield-associated phenotypes such as total number of seeds, seed size and seed weight showed varying degrees of temperature sensitivity, likely caused by the partially distinct temperature effects on individual accessions (Fig. 7.3B, Supplementary Fig. D.11A). A comparison of total seed numbers harvested from plants grown at 28 °C or 16 °C clearly illustrates that for most accessions higher temperatures cause a strong decrease in total yield (Fig. 7.4A, Supplementary Fig. D.13). However, Got-7 showed an opposite trend even though the overall yield was severely reduced at both temperatures (Supplemental Fig. D.13). This illustrates that the extension of the vegetative growth phase might positively affect yield (it has to be noted that in the case of Got-7 this observation might be affected by the vernalization requirement). This would require further inspection using accessions, ideally those with less pronounced vernalization requirements.

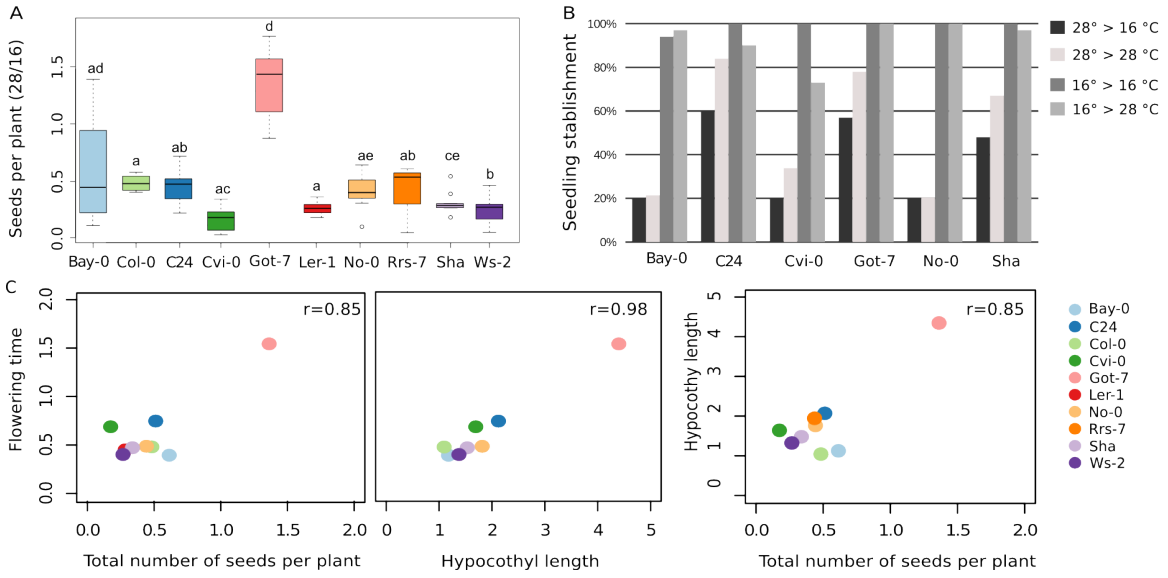


Figure 7.4.: Yield, trans-generational effects and phenotypic correlations. (A) Comparison of temperature sensitivities of accession yield. Box plots show relative seed numbers (28 °C vs. 16 °C median). Different letters denote significant differences ($P < 0.05$) as assessed by two-factorial ANOVA of absolute data shown in Supplementary Fig. D.13. Got-7 was significantly less affected by high temperature but showed lower absolute yield values at all analyzed temperatures (Supplementary Fig. D.13). (B) Rates of seedling establishment of 6 days old seedlings. Seeds were collected from plants grown at 16 or 28 °C for an entire life cycle and were germinated either the same (16 > 16 °C and 28 > 28 °C) or the respective other growth temperature (16 > 28 °C and 28 > 16 °C). The experiment was performed three times with similar results of which one representative is shown. (C) Scatter plot of temperature response ratios (28 vs. 16 °C) of selected phenotypes. Pearson correlation coefficients (r) of trait temperature response ratios (28 vs. 16 °C) are shown in the upper right corners. See Supplementary Fig. D.14 for complete set of pair-wise comparisons among traits.

The observed differences in yield and some of the seed size parameters prompted us to inspect potential trans-generational effects of ambient growth temperatures on the following generation. Therefore, we tested the rates of germination and seedling establishment of seeds collected from plants grown at 16 °C and 28 °C when cultivated again at the same or the respective other temperature. Germination rates ranged between 97 to 100% and were similar among all analyzed samples. Seedling establishment (= fully opened cotyledons) after 6 days, however, showed reproducible differences among the different samples. Seeds collected from plants grown at 16 °C showed almost no differences in seedling establishment when germinated at 16 or 28 °C (Fig. 7.4B). However, seeds collected from plants grown at 28 °C seem to show higher seedling establishment rates when grown under the same temperature (28 °C) compared to seeds germinated at 16 °C (Fig. 7.4B). This improved development might indicate trans-generational priming of seeds for development at higher temperatures, putatively involving epigenetic processes. While these effects were repeatedly observed for individual seed pools, extensive analysis of seeds collected from independently cultivated parental lines need to be analyzed to substantiate these observations.

7.4.6. Correlation of phenotypic temperature responses

Finally, we analyzed putative correlations in temperature responses (28 vs. 16 °C) among different phenotypes. We used Pearson correlation coefficients for pairwise comparisons of trait ratios (28 vs. 16 °C) among all accessions. As to be expected from the varying degrees of genotype and temperature effects on different traits, correlations among phenotypes covered a wide range (Supplementary Figure D.14). Particularly high correlation values were observed among flowering time, hypocotyl length and seed production (Fig. 7.4C), indicating that traits with strong adaptive potential seem to be affected similarly. Moreover, these data reveal that model phenotypes used in classic forward genetic approaches (such as hypocotyl elongation) are at least partially indicative for general temperature responses in plants.

7.5. Discussion

Increased ambient temperatures have been shown to affect thermomorphogenesis for selected phenotypes (Gray et al., 1998; Zanten et al., 2009). A systematic assessment of developmental plasticity across a complete life cycle has, to the best of our knowledge, been lacking so far. This study provides a solid base of temperature effects on plants by consecutive profiling of plant growth and development throughout a life cycle of *A. thaliana* grown in four different ambient temperatures. Furthermore, including several distinct *A. thaliana* accessions reduced potential genotype-specific biases in the data and allowed the analysis of temperature and genotype effects on the different phenotypic traits.

Of the 34 phenotypes analyzed, almost all were affected by different growth temperatures illustrating the fundamental impact of ambient temperature on plant physiology (Fig. 7.1, Supplementary Fig. D.1- D.10).

Temperature-sensitive traits can be divided into two distinct groups. First, phenotypes that were similarly affected in all analyzed accessions. Second, phenotypes that showed natural variation in temperature responses. The induction of leaf development throughout the vegetative growth phase was uniformly accelerated by increasing temperatures in all analyzed genotypes. This could indicate either a highly conserved regulation within *A. thaliana* or a regulation due to passive temperature effects. Indeed, thermomorphogenic responses are often speculated to be primarily caused by the effect of free energy changes on biological reactions (e.g. enzyme activities). The validity of the early proposed temperature coefficient (Q10) for plant development was demonstrated for germination rates and plant respiration (Atkin et al., 2003; Hegarty, 1973). The strong temperature effect on the acceleration of developmental timing throughout the vegetative phase, which was only weakly affected by genotypes would certainly fit to this theory. When adopting the terms of "passive" and "active" temperature effects as proposed by Penfield and MacGregor (Penfield et al., 2014), timing of vegetative development would represent a passive temperature response that might be caused by thermodynamic effects on metabolic rates and enzyme activities.

On the other hand, phenotypes that show a high degree of genotype and temperature effects might rather be influenced by one or more specific genes that contribute to trait expression in a quantitative manner. As such, these phenotypes would represent "active" temperature effects (Penfield et al., 2014). Natural variation in thermomorphogenic responses could be caused by different polymorphisms of signaling or response genes ranging from alteration in gene sequence to expression level polymorphism (Delker et al., 2011) due to adaptation to local environmental conditions. As they provide keys to altered temperature responses that could be utilized in specific breeding approaches, these genes would thus be of high interest. Several phenotypes analyzed here have the potential to contribute to adaptation to environmental conditions. Particularly hypocotyl and petiole elongation as well as hyponastic leaf movement (increased petiole angles) have previously been shown to improve leaf cooling by increased transpiration rates (Bridge et al., 2013; Crawford et al., 2012). As such, variation in any of these traits could significantly impact on photosynthesis rates and affect further growth and development. In fact, the ratio of hypocotyl elongation showed a high correlation with the ratio of flowering induction and yield (28 vs. 16 °C, Fig. 7.4C). This could indicate that early seedling development significantly affects the timing of further development. Alternatively, these processes might involve similar signaling elements. In fact, PIF4 and ELF3 as central signaling elements that integrate multiple environmental stimuli have been shown to be involved in both, temperature induced hypocotyl elongation and the induction of flowering (Koini et al., 2009; Kumar et al., 2012).

In addition, natural allelic variation in the circadian clock components *ELF3* and in the regulation of *GIGANTEA* have recently been shown to directly affect PIF4-mediated hypocotyl elongation in response to elevated temperatures (Box et al., 2015; Montaigu et al., 2015; Raschke et al., 2015). Therefore, PIF4 and PIF4-regulating components could be important targets of adaptation.

The increasing number of identified genes and allelic variations that contribute to specific phenotypic changes in response to elevated ambient temperatures argue against a general explanation of morphological and developmental changes due to passive effects by thermodynamic

processes.

Exploiting natural genetic variation to identify genes that are involved in the regulation of temperature effects on specific traits (e.g., *ELF3* and *PIF4*) can provide new avenues in breeding. Specific approaches will depend on the focus on either yield- or biomass-associated traits. In addition, initial evidence for trans-generational effects require further analysis to account for potential epigenetic transduction of temperature cues on growth and development.

In conclusion, our work provides a data resource that allows the dissection of thermomorphogenesis in phenotypic traits that are either robustly affected by temperature or traits that are differentially affected by temperature among different accessions; the latter might be a consequence of adaptive processes. While robust temperature-sensitive phenotypes might indeed be caused by thermodynamic acceleration of metabolism, natural genetic variation of temperature responses implicate the relevance of specific regulatory cascades that might be targets of adaptation to local environmental conditions.

7.6. Acknowledgements

This study was supported by the Leibniz association and a grant from the Deutsche Forschungsgemeinschaft to M.Q. (Qu 141/3-1).

7.7. References

Atkin, O. K. and Tjoelker, M. G. (2003). Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends in Plant Science*, 8 (7), pp. 343–351.

Balasubramanian, S., Sureshkumar, S., Lempe, J., and Weigel, D. (2006). Potent Induction of *Arabidopsis* thaliana Flowering by Elevated Growth Temperature. *PLoS Genetics*, 2 (7), e106.

Box, M. S., Huang, B. E., Domijan, M., Jaeger, K. E., Khattak, A. K., Yoo, S. J., Sedivy, E. L., Jones, D. M., Hearn, T. J., Webb, A. A., Grant, A., Locke, J. C., and Wigge, P. A. (2015). ELF3 Controls Thermoresponsive Growth in *Arabidopsis*. *Current Biology*, 25 (2), pp. 194–199.

Bridge, L. J., Franklin, K. A., and Homer, M. E. (2013). Impact of plant shoot architecture on leaf cooling: a coupled heat and mass transfer model. *Journal of The Royal Society Interface*, 10 (85).

CaraDonna, P. J., Iler, A. M., and Inouye, D. W. (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences*, 111 (13), pp. 4916–4921.

Crawford, A. J., McLachlan, D. H., Hetherington, A. M., and Franklin, K. A. (2012). High temperature exposure increases plant cooling capacity. *Current Biology*, 22 (10), R396–R397.

Delker, C., Pöschl, Y., Raschke, A., Ullrich, K., Ettingshausen, S., Hauptmann, V., Grosse, I., and Quint, M. (2010). Natural Variation of Transcriptional Auxin Response Networks in *Arabidopsis thaliana*. *The Plant Cell*, 22 (7), pp. 2184–2200.

Delker, C. and Quint, M. (2011). Expression level polymorphisms: heritable traits shaping natural variation. *Trends in Plant Science*, 16 (9), pp. 481–488.

Delker, C., Sonntag, L., James, G. V., Janitza, P., Ibañez, C., Ziermann, H., Peterson, T., Denk, K., Mull, S., Ziegler, J., Davis, S. J., Schneeberger, K., and Quint, M. (2014). The DET1-COP1-HY5 Pathway Constitutes a Multipurpose Signaling Module Regulating Plant Photomorphogenesis and Thermomorphogenesis. *Cell Reports*, 9 (6), pp. 1983–1989.

Donner, A. and Koval, J. J. (1980). The Estimation of Intraclass Correlation in the Analysis of Family Data. *Biometrics*, 36 (1), pp. 19–25.

Fitter, A. H. and Fitter, R. S. R. (2002). Rapid Changes in Flowering Time in British Plants. *Science*, 296 (5573), pp. 1689–1691.

Franklin, K. A., Lee, S. H., Patel, D., Kumar, S. V., Spartz, A. K., Gu, C., Ye, S., Yu, P., Breen, G., Cohen, J. D., Wigge, P. A., and Gray, W. M. (2011). PHYTOCHROME-INTERACTING FACTOR 4 (PIF4) regulates auxin biosynthesis at high temperature. *Proceedings of the National Academy of Sciences*, 108 (50), pp. 20231–20235.

Gray, W. M., Östin, A., Sandberg, G., Romano, C. P., and Estelle, M. (1998). High temperature promotes auxin-mediated hypocotyl elongation in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 95 (12), pp. 7197–7202.

Hegarty, T. W. (1973). Temperature Coefficient (Q10), Seed Germination and Other Biological Processes. *Nature*, 243 (5405), pp. 305–306.

IPCC Climate change 2013: The physical science basis. Fifth assessment report. (2013). `http://www.ipcc.ch/report/ar5/wg1/`.

Koini, M. A., Alvey, L., Allen, T., Tilley, C. A., Harberd, N. P., Whitelam, G. C., and Franklin, K. A. (2009). High Temperatur-Mediated Adaptations in Plant Architecture Require the bHLH Transcription Factor PIF4. *Current Biology*, 19 (5), pp. 408–413.

Kumar, S. V., Lucyshyn, D., Jaeger, K. E., Alos, E., Alvey, E., Harberd, N. P., and Wigge, P. A. (2012). Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature*, 484 (7393), pp. 242–245.

Lincoln, C., Britton, J. H., and Estelle, M. (1990). Growth and development of the axr1 mutants of *Arabidopsis*. *The Plant Cell*, 2 (11), pp. 1071–80.

Lobell, D. B. and Gourdji, S. M. (2012). The Influence of Climate Change on Global Crop Productivity. *Plant Physiology*, 160 (4), pp. 1686–1697.

Montaigu, A. de, Giakountis, A., Rubin, M., Tóth, R., Cremer, F., Sokolova, V., Porri, A., Reymond, M., Weinig, C., and Coupland, G. (2015). Natural diversity in daily rhythms of gene expression contributes to phenotypic variation. *Proceedings of the National Academy of Sciences*, 112 (3), pp. 905–910.

Moore, F. C. and Lobell, D. B. (2015). The fingerprint of climate trends on European crop yields. *Proceedings of the National Academy of Sciences*, 112 (9), pp. 2670–2675.

Penfield, S. and MacGregor, D. (2014). "Temperature sensing in plants". In: *Temperature and Plant Development*. John Wiley & Sons, Inc, pp. 1–18.

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., Centeno, G. S., Khush, G. S., and Cassman, K. G. (2004). Rice yields decline with higher night tempera-

ture from global warming. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (27), pp. 9971–9975.

Porra, R., Thompson, W., and Kriedemann, P. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 975 (3), pp. 384–394.

Proveniers, M. C. and Zanten, M. van (2013). High temperature acclimation through PIF4 signaling. *Trends in Plant Science*, 18 (2), pp. 59–64.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vol. ISBN 3-900051-07-0. Vienna, Austria.

Raschke, A., Ibañez, C., Ullrich, K. K., Anwer, M. U., Becker, S., Glöckner, A., Trenner, J., Denk, K., Saal, B., Sun, X., Ni, M., Davis, S. J., Delker, C., and Quint, M. (2015). Natural Variants of ELF3 Affect Thermomorphogenesis by Transcriptionally Modulating PIF4-Dependent Auxin Response Genes. *bioRxiv*.

Scholl, R. L., May, S. T., and Ware, D. H. (2000). Seed and Molecular Resources for Arabidopsis. *Plant Physiology*, 124 (4), pp. 1477–1480.

Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., and Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (23), pp. 8245–8250.

Toledo-Ortiz, G., Johansson, H., Lee, K. P., Bou-Torrent, J., Stewart, K., Steel, G., Rodríguez-Concepción, M., and Halliday, K. J. (2014). The HY5-PIF Regulatory Module Coordinates Light and Temperature Control of Photosynthetic Gene Transcription. *PLoS Genetics*, 10 (6), e1004416.

Zanten, M. van, Voesenek, L. A., Peeters, A. J., and Millenaar, F. F. (2009). Hormone- and Light-Mediated Regulation of Heat-Induced Differential Petiole Growth in *Arabidopsis*. *Plant Physiology*, 151 (3), pp. 1446–1458.

Bibliography

Alonso-Blanco, C., Aarts, M. G., Bentsink, L., Keurentjes, J. J., Reymond, M., Vreugdenhil, D., and Koornneef, M. (2009). What Has Natural Variation Taught Us about Plant Development, Physiology, and Adaptation? *The Plant Cell*, 21 (7), pp. 1877–1896.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215 (3), pp. 403–410.

Arikawa, E., Sun, Y., Wang, J., Zhou, Q., Ning, B., Dial, S., Guo, L., and Yang, J. (2008). Cross-platform comparison of SYBR(R) Green real-time PCR with TaqMan PCR, microarrays and other gene expression measurement technologies evaluated in the MicroArray Quality Control (MAQC) study. *BMC Genomics*, 9 (1), p. 328.

Atkin, O. K. and Tjoelker, M. G. (2003). Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends in Plant Science*, 8 (7), pp. 343–351.

Balasubramanian, S., Sureshkumar, S., Lempe, J., and Weigel, D. (2006). Potent Induction of *Arabidopsis* thaliana Flowering by Elevated Growth Temperature. *PLoS Genetics*, 2 (7), e106.

Bar-Or, C., Czosnek, H., and Koltai, H. (2007). Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends in Genetics*, 23 (4), pp. 200–207.

Boer, D. R., Freire-Rios, A., Berg, W. A. M. van den, Saaki, T., Manfield, I. W., Kepinski, S., López-Vidrieo, I., Franco-Zorrilla, J. M., Vries, S. C. de, Solano, R., Weijers, D., and Coll, M. (2014). Structural Basis for DNA Binding Specificity by the Auxin-Dependent ARF Transcription Factors. *Cell*, 156 (3), pp. 577–589.

Borevitz, J. O., Hazen, S. P., Michael, T. P., Morris, G. P., Baxter, I. R., Hu, T. T., Chen, H., Werner, J. D., Nordborg, M., Salt, D. E., Kay, S. A., Chory, J., Weigel, D., Jones, J. D. G., and Ecker, J. R. (2007). Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, 104 (29), pp. 12057–12062.

Box, M. S., Huang, B. E., Domijan, M., Jaeger, K. E., Khattak, A. K., Yoo, S. J., Sedivy, E. L., Jones, D. M., Hearn, T. J., Webb, A. A., Grant, A., Locke, J. C., and Wigge, P. A. (2015). ELF3 Controls Thermoresponsive Growth in *Arabidopsis*. *Current Biology*, 25 (2), pp. 194–199.

Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Letters*, 573 (1–3), pp. 83–92.

Bridge, L. J., Franklin, K. A., and Homer, M. E. (2013). Impact of plant shoot architecture on leaf cooling: a coupled heat and mass transfer model. *Journal of The Royal Society Interface*, 10 (85).

Broadley, M. R., White, P. J., Hammond, J. P., Graham, N. S., Bowen, H. C., Emmerson, Z. F., Fray, R. G., Iannetta, P. P. M., McNicol, J. W., and May, S. T. (2008). Evidence of neutral transcriptome evolution in plants. *New Phytologist*, 180 (3), pp. 587–593.

Calderón Villalobos, L. I. A., Lee, S., De Oliveira, C., Ivetac, A., Brandt, W., Armitage, L., Sheard, L. B., Tan, X., Parry, G., Mao, H., Zheng, N., Napier, R., Kepinski, S., and Estelle, M. (2012). A combinatorial TIR1/AFB–Aux/IAA co-receptor system for differential sensing of auxin. *Nature Chemical Biology*, 8 (5), pp. 477–485.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10 (1), p. 421.

CaraDonna, P. J., Iler, A. M., and Inouye, D. W. (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences*, 111 (13), pp. 4916–4921.

Chapman, E. J. and Estelle, M. (2009). Mechanism of Auxin-Regulated Gene Expression in Plants. *Annual Review of Genetics*, 43 (1). PMID: 19686081, pp. 265–285.

Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T. T., Fu, G., Hinds, D. A., Chen, H., Frazer, K. A., Huson, D. H., Schölkopf, B., Nordborg, M., Rätsch, G., Ecker, J. R., and Weigel, D. (2007). Common Sequence Polymorphisms Shaping Genetic Diversity in *Arabidopsis thaliana*. *Science*, 317 (5836), pp. 338–342.

Clough, S. J. and Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal*, 16 (6), pp. 735–743.

Covington, M. F. and Harmer, S. L. (2007). The Circadian Clock Regulates Auxin Signaling and Responses in *Arabidopsis*. *PLoS Biology*, 5 (8), e222.

Crawford, A. J., McLachlan, D. H., Hetherington, A. M., and Franklin, K. A. (2012). High temperature exposure increases plant cooling capacity. *Current Biology*, 22 (10), R396–R397.

Czechowski, T., Bari, R. P., Stitt, M., Scheible, W.-R., and Udvardi, M. K. (2004). Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *The Plant Journal*, 38 (2), pp. 366–379.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W.-R. (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in *Arabidopsis*. *Plant Physiology*, 139 (1), pp. 5–17.

Dannemann, M., Lorenc, A., Hellmann, I., Khaitovich, P., and Lachmann, M. (2009). The effects of probe binding affinity differences on gene expression measurements and how to deal with them. *Bioinformatics*, 25 (21), pp. 2772–2779.

Delker, C., Pöschl, Y., Raschke, A., Ullrich, K., Ettingshausen, S., Hauptmann, V., Grosse, I., and Quint, M. (2010). Natural Variation of Transcriptional Auxin Response Networks in *Arabidopsis thaliana*. *The Plant Cell*, 22 (7), pp. 2184–2200.

Delker, C. and Quint, M. (2011). Expression level polymorphisms: heritable traits shaping natural variation. *Trends in Plant Science*, 16 (9), pp. 481–488.

Delker, C., Raschke, A., and Quint, M. (2008). Auxin dynamics: the dazzling complexity of a small molecule's message. *Planta*, 227 (5), pp. 929–941.

Delker, C., Sonntag, L., James, G. V., Janitza, P., Ibañez, C., Ziermann, H., Peterson, T., Denk, K., Mull, S., Ziegler, J., Davis, S. J., Schneeberger, K., and Quint, M. (2014).

The DET1-COP1-HY5 Pathway Constitutes a Multipurpose Signaling Module Regulating Plant Photomorphogenesis and Thermomorphogenesis. *Cell Reports*, 9 (6), pp. 1983–1989.

Dharmasiri, N., Dharmasiri, S., and Estelle, M. (2005a). The F-box protein TIR1 is an auxin receptor. *Nature*, 435 (7041), pp. 441–445.

Dharmasiri, N., Dharmasiri, S., Weijers, D., Lechner, E., Yamada, M., Hobbie, L., Ehrismann, J. S., Jürgens, G., and Estelle, M. (2005b). Plant Development Is Regulated by a Family of Auxin Receptor F-Box Proteins. *Developmental Cell*, 9 (1), pp. 109–119.

Donner, A. and Koval, J. J. (1980). The Estimation of Intraclass Correlation in the Analysis of Family Data. *Biometrics*, 36 (1), pp. 19–25.

Dos Santos Maraschin, F., Memelink, J., and Offringa, R. (2009). Auxin-induced, SCFTIR1-mediated poly-ubiquitination marks AUX/IAA proteins for degradation. *The Plant Journal*, 59 (1), pp. 100–109.

Dussaubat, C., Brunet, J.-L., Higes, M., Colbourne, J. K., Lopez, J., Choi, J.-H., Martín-Hernández, R., Botías, C., Cousin, M., McDonnell, C., Bonnet, M., Belzunces, L. P., Moritz, R. F. A., Le Conte, Y., and Alaux, C. (2012). Gut Pathology and Responses to the Microsporidium *Nosema ceranae* in the Honey Bee *Apis mellifera*. *PLoS ONE*, 7 (5), e37017.

Eklund, A. (2015). *beeswarm: The Bee Swarm Plot, an Alternative to Stripchart*. R package version 0.2.0.

Fang, G., Bhardwaj, N., Robilotto, R., and Gerstein, M. B. (2010). Getting Started in Gene Orthology and Functional Analysis. *PLoS Computational Biology*, 6 (3), e1000703.

Fitter, A. H. and Fitter, R. S. R. (2002). Rapid Changes in Flowering Time in British Plants. *Science*, 296 (5573), pp. 1689–1691.

Flenniken, M. L. and Andino, R. (2013). Non-Specific dsRNA-Mediated Antiviral Response in the Honey Bee. *PLoS ONE*, 8 (10), e77263.

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences*, 111 (6), pp. 2367–2372.

Franklin, K. A., Lee, S. H., Patel, D., Kumar, S. V., Spartz, A. K., Gu, C., Ye, S., Yu, P., Breen, G., Cohen, J. D., Wigge, P. A., and Gray, W. M. (2011). PHYTOCHROME-INTERACTING FACTOR 4 (PIF4) regulates auxin biosynthesis at high temperature. *Proceedings of the National Academy of Sciences*, 108 (50), pp. 20231–20235.

Fu, J., Keurentjes, J. J. B., Bouwmeester, H., America, T., Verstappen, F. W. A., Ward, J. L., Beale, M. H., Vos, R. C. H. de, Dijkstra, M., Scheltema, R. A., Johannes, F., Koornneef, M., Vreugdenhil, D., Breitling, R., and Jansen, R. C. (2009). System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nature Genetics*, 41 (2), pp. 166–167.

Fujimoto, R., Taylor, J. M., Sasaki, T., Kawanabe, T., and Dennis, E. S. (2011). Genome wide gene expression in artificially synthesized amphidiploids of *Arabidopsis*. *Plant Molecular Biology*, 77 (4), pp. 419–431.

Gansner, E. R. and North, S. C. (2000). An open graph visualization system and its applications to software engineering. *Software - Practice and Experience*, 30 (11), pp. 1203–1233.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20 (3), pp. 307–315.

Gilad, Y. and Borevitz, J. (2006). Using DNA microarrays to study natural variation. *Current Opinion in Genetics and Development*, 16 (6), pp. 553–558.

Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., Ogawa, M., Yamauchi, Y., Preston, J., Aoki, K., Kiba, T., Takatsuto, S., Fujioka, S., Asami, T., Nakano, T., Kato, H., Mizuno, T., Sakakibara, H., Yamaguchi, S., Nambara, E., Kamiya, Y., Takahashi, H., Hirai, M. Y., Sakurai, T., Shinozaki, K., Saito, K., Yoshida, S., and Shimada, Y. (2008). The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *The Plant Journal*, 55 (3), pp. 526–542.

Graham, N., Broadley, M., Hammond, J., White, P., and May, S. (2007). Optimising the analysis of transcript data using high density oligonucleotide arrays and genomic DNA-based probe selection. *BMC Genomics*, 8 (1), p. 344.

Grau, J., Posch, S., Grosse, I., and Keilwagen, J. (2013). A general approach for discriminative de novo motif discovery from high-throughput data. *Nucleic Acids Research*, 41 (21), e197.

Gray, W. M., Östin, A., Sandberg, G., Romano, C. P., and Estelle, M. (1998). High temperature promotes auxin-mediated hypocotyl elongation in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 95 (12), pp. 7197–7202.

Guilfoyle, T. J., Ulmasov, T., and Hagen, G. (1998). The ARF family of transcription factors and their role in plant hormone-responsive transcription. *Cellular and Molecular Life Sciences*, 54 (7), pp. 619–627.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, pp. 95–98.

Hammond, J., Broadley, M., Craigon, D., Higgins, J., Emmerson, Z., Townsend, H., White, P., and May, S. (2005). Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. *Plant Methods*, 1 (1), p. 10.

Hegarty, T. W. (1973). Temperature Coefficient (Q10), Seed Germination and Other Biological Processes. *Nature*, 243 (5405), pp. 305–306.

Holt, H., Aronstein, K., and Grozinger, C. (2013). Chronic parasitization by *Nosema* microsporidia causes global expression changes in core nutritional, metabolic and behavioral pathways in honey bee workers (*Apis mellifera*). *BMC Genomics*, 14 (1), p. 799.

Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., Fahlgren, N., Fawcett, J. A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J. D., Ossowski, S., Ottilar, R. P., Salamov, A. A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M. E., Bergelson, J., Carrington, J. C., Gaut, B. S., Schmutz, J., Mayer, K. F. X., Van de Peer, Y., Grigoriev, I. V., Nordborg, M., Weigel, D., and Guo, Y.-L. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nature Genetics*, 43 (5), pp. 476–481.

Ibañez, C., Poeschl, Y., Peterson, T., Bellstädt, J., Denk, K., Gogol-Döring, A., Quint, M., and Delker, C. (2015). Developmental plasticity of *Arabidopsis thaliana* accessions across an ambient temperature range. *bioRxiv*, pre-print, doi: 10.1101/017285.

IPCC Climate change 2013: The physical science basis. Fifth assessment report. (2013). `http://www.ipcc.ch/report/ar5/wg1/`.

Irizarry, R. A., Gautier, L., Huber, W., and Bolstad, B. (2006). *makecdfenv: CDF Environment Maker*. R package version 1.30.0.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4 (2), pp. 249–264.

Katagiri, F. and Glazebrook, J. (2003). Local Context Finder (LCF) reveals multidimensional relationships among mRNA expression profiles of *Arabidopsis* responding to pathogen infection. *Proceedings of the National Academy of Sciences*, 100 (19), pp. 10842–10847.

Kato, H., Ishizaki, K., Kouno, M., Shirakawa, M., Bowman, J. L., Nishihama, R., and Kohchi, T. (2015). Auxin-Mediated Transcriptional System with a Minimal Set of Components Is Critical for Morphogenesis through the Life Cycle in *Marchantia polymorpha*. *PLoS Genet*, 11 (5), e1005084.

Keilwagen, J., Grau, J., Paponov, I. A., Posch, S., Strickert, M., and Grosse, I. (2011). De-Novo Discovery of Differentially Abundant Transcription Factor Binding Sites Including Their Positional Preference. *PLoS Computational Biology*, 7 (2), e1001070.

Kepinski, S. and Leyser, O. (2005). The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature*, 435 (7041), pp. 446–451.

Keurentjes, J. J. B., Fu, J., Terpstra, I. R., Garcia, J. M., Ackerveken, G. van den, Snoek, L. B., Peeters, A. J. M., Vreugdenhil, D., Koornneef, M., and Jansen, R. C. (2007). Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proceedings of the National Academy of Sciences*, 104 (5), pp. 1708–1713.

Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigele, S., Do, H.-H., Weiss, G., Enard, W., Heissig, F., Arendt, T., Nieselt-Struwe, K., Eichler, E. E., and Pääbo, S. (2004). Regional Patterns of Gene Expression in Human and Chimpanzee Brains. *Genome Research*, 14 (8), pp. 1462–1473.

Kim, J. and Iyer, V. R. (2004). Global Role of TATA Box-Binding Protein Recruitment to Promoters in Mediating Gene Expression Profiles. *Molecular and Cellular Biology*, 24 (18), pp. 8104–8112.

Kleine-Vehn, J., Dhonukshe, P., Swarup, R., Bennett, M., and Friml, J. (2006). Subcellular Trafficking of the *Arabidopsis* Auxin Influx Carrier AUX1 Uses a Novel Pathway Distinct from PIN1. *The Plant Cell Online*, 18 (11), pp. 3171–3181.

Kliebenstein, D. J., West, M. A. L., Leeuwen, H. van, Kim, K., Doerge, R. W., Michelmore, R. W., and St. Clair, D. A. (2006). Genomic Survey of Gene Expression Diversity in *Arabidopsis thaliana*. *Genetics*, 172 (2), pp. 1179–1189.

Koini, M. A., Alvey, L., Allen, T., Tilley, C. A., Harberd, N. P., Whitelam, G. C., and Franklin, K. A. (2009). High Temperatur-Mediated Adaptations in Plant Architecture Require the bHLH Transcription Factor PIF4. *Current Biology*, 19 (5), pp. 408–413.

Kreitman, M. (2000). Methods to detect selection in populations with application to the human. *Annual Review of Genomics and Human Genetics*, 1 (1). PMID: 11701640, pp. 539–559.

Kumar, S. V., Lucyshyn, D., Jaeger, K. E., Alos, E., Alvey, E., Harberd, N. P., and Wigge, P. A. (2012). Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature*, 484 (7393), pp. 242–245.

Laan, M. J. van der and Pollard, K. S. (2003). A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap. *Journal of Statistical Planning and Inference*, 117 (2), pp. 275–303.

Leeuwen, H. van, Kliebenstein, D. J., West, M. A., Kim, K., Poecke, R. van, Katagiri, F., Michelmore, R. W., Doerge, R. W., and St.Clair, D. A. (2007). Natural Variation among *Arabidopsis thaliana* Accessions for Transcriptome Response to Exogenous Salicylic Acid. *The Plant Cell*, 19 (7), pp. 2099–2110.

Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25 (11), pp. 1451–1452.

Lincoln, C., Britton, J. H., and Estelle, M. (1990). Growth and development of the axr1 mutants of *Arabidopsis*. *The Plant Cell*, 2 (11), pp. 1071–80.

Liu, Z. B., Ulmasov, T., Shi, X., Hagen, G., and Guilfoyle, T. J. (1994). Soybean GH3 promoter contains multiple auxin-inducible elements. *The Plant Cell*, 6 (5), pp. 645–57.

Lobell, D. B. and Gourdji, S. M. (2012). The Influence of Climate Change on Global Crop Productivity. *Plant Physiology*, 160 (4), pp. 1686–1697.

Lokerse, A. S. and Weijers, D. (2009). Auxin enters the matrix–assembly of response machineries for specific outputs. *Current Opinion in Plant Biology*, 12 (5), pp. 520–526.

Maloof, J. N., Borevitz, J. O., Dabi, T., Lutes, J., Nehring, R. B., Redfern, J. L., Trainer, G. T., Wilson, J. M., Asami, T., and Berry, C. C. (2001). Natural variation in light sensitivity of *Arabidopsis*. *Nature Genetics*, (4), pp. 441–446.

McDonnell, C., Alaux, C., Parrinello, H., Desvignes, J.-P., Crauser, D., Durbesson, E., Beslay, D., and Le Conte, Y. (2013). Ecto- and endoparasite induce similar chemical and brain neurogenomic responses in the honey bee (*Apis mellifera*). *BMC Ecology*, 13 (1), pp. 1–15.

Molina, C. and Grotewold, E. (2005). Genome wide analysis of *Arabidopsis* core promoters. *BMC Genomics*, 6, pp. 25–25.

Montaigu, A. de, Giakountis, A., Rubin, M., Tóth, R., Cremer, F., Sokolova, V., Porri, A., Reymond, M., Weinig, C., and Coupland, G. (2015). Natural diversity in daily rhythms of gene expression contributes to phenotypic variation. *Proceedings of the National Academy of Sciences*, 112 (3), pp. 905–910.

Moore, F. C. and Lobell, D. B. (2015). The fingerprint of climate trends on European crop yields. *Proceedings of the National Academy of Sciences*, 112 (9), pp. 2670–2675.

Murtagh, F. and Contreras, P. (2011). Methods of Hierarchical Clustering. *CoRR*, abs/1105.0121.

Nagpal, P., Walker, L. M., Young, J. C., Sonawala, A., Timpte, C., Estelle, M., and Reed, J. W. (2000). AXR2 Encodes a Member of the Aux/IAA Protein Family. *Plant Physiology*, 123 (2), pp. 563–574.

Naiser, T., Kayser, J., Mai, T., Michel, W., and Ott, A. (2008). Position dependent mismatch discrimination on DNA microarrays - experiments and model. *BMC Bioinformatics*, 9 (1), p. 509.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *Journal of Bioscience and Bioengineering*, 104 (1), pp. 34–41.

Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraishi, T., Iwabuchi, M., and Narusaka, Y. (2009). RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *The Plant Journal*, 60 (2), pp. 218–226.

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., Voinnet, O., and Jones, J. D. G. (2006). A Plant miRNA Contributes to Antibacterial Resistance by Repressing Auxin Signaling. *Science*, 312 (5772), pp. 436–439.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M. and Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76 (10), pp. 5269–5273.

Nemhauser, J. L., Hong, F., and Chory, J. (2006). Different Plant Hormones Regulate Similar Processes through Largely Nonoverlapping Transcriptional Responses. *Cell*, 126 (3), pp. 467–475.

Nemhauser, J. L., Mockler, T. C., and Chory, J. (2004). Interdependency of Brassinosteroid and Auxin Signaling in *Arabidopsis*. *PLoS Biol*, 2 (9), e258.

Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N. A., Shah, C., Wall, J. D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M., and Bergelson, J. (2005). The Pattern of Polymorphism in *Arabidopsis thaliana*. *PLoS Biology*, 3 (7), e196.

Okushima, Y., Overvoorde, P. J., Arima, K., Alonso, J. M., Chan, A., Chang, C., Ecker, J. R., Hughes, B., Lui, A., Nguyen, D., Onodera, C., Quach, H., Smith, A., Yu, G., and Theologis, A. (2005). Functional Genomic Analysis of the AUXIN RESPONSE FACTOR Gene Family Members in *Arabidopsis thaliana*: unique and Overlapping Functions of ARF7 and ARF19. *The Plant Cell*, 17 (2), pp. 444–463.

Opgen-Rhein, R. and Strimmer, K. (2007). Accurate Ranking of Differentially Expressed Genes by a Distribution-Free Shrinkage Approach. *Statistical Applications in Genetics and Molecular Biology*, 6 (1), pp. 477+.

Orlov, Y., Zhou, J., Lipovich, L., Shahab, A., and Kuznetsov, V. (2007). Quality assessment of the Affymetrix U133A&B probesets by target sequence mapping and expression data analysis. *In Silico Biol*, 7 (3), pp. 241–60.

Overvoorde, P. J., Okushima, Y., Alonso, J. M., Chan, A., Chang, C., Ecker, J. R., Hughes, B., Liu, A., Onodera, C., Quach, H., Smith, A., Yu, G., and Theologis, A. (2005). Functional Genomic Analysis of the AUXIN/INDOLE-3-ACETIC ACID Gene Family Members in *Arabidopsis thaliana*. *The Plant Cell*, 17 (12), pp. 3282–3300.

Paponov, I. A., Paponov, M., Teale, W., Menges, M., Chakrabortee, S., Murray, J. A. H., and Palme, K. (2008). Comprehensive Transcriptome Analysis of Auxin Responses in *Arabidopsis*. *Molecular Plant*, 1 (2), pp. 321–337.

Parry, G., Calderon-Villalobos, L. I., Prigge, M., Peret, B., Dharmasiri, S., Itoh, H., Lechner, E., Gray, W. M., Bennett, M., and Estelle, M. (2009). Complex regulation of the TIR1/AFB family of auxin receptors. *Proceedings of the National Academy of Sciences*, 106 (52), pp. 22540–22545.

Parry, G. and Estelle, M. (2006). Auxin receptors: a new role for F-box proteins. *Current Opinion in Cell Biology*, 18 (2), pp. 152–156.

Penfield, S. and MacGregor, D. (2014). "Temperature sensing in plants". In: *Temperature and Plant Development*. John Wiley & Sons, Inc, pp. 1–18.

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., Centeno, G. S., Khush, G. S., and Cassman, K. G. (2004). Rice yields decline with higher night temperature from global warming. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (27), pp. 9971–9975.

Pérez-Torres, C.-A., López-Bucio, J., Cruz-Ramírez, A., Ibarra-Laclette, E., Dharmasiri, S., Estelle, M., and Herrera-Estrella, L. (2008). Phosphate Availability Alters Lateral Root Development in *Arabidopsis* by Modulating Auxin Sensitivity via a Mechanism Involving the TIR1 Auxin Receptor. *The Plant Cell*, 20 (12), pp. 3258–3272.

Ploense, S. E., Wu, M.-F., Nagpal, P., and Reed, J. W. (2009). A gain-of-function mutation in IAA18 alters *Arabidopsis* embryonic apical patterning. *Development*, 136 (9), pp. 1509–1517.

Poecke, R. M. van, Sato, M., Lenarz-Wyatt, L., Weisberg, S., and Katagiri, F. (2007). Natural Variation in RPS2-Mediated Resistance among *Arabidopsis* Accessions: correlation between Gene Expression Profiles and Phenotypic Responses. *The Plant Cell*, 19 (12), pp. 4046–4060.

Poeschl, Y., Delker, C., Trenner, J., Ullrich, K. K., Quint, M., and Grosse, I. (2013). Optimized Probe Masking for Comparative Transcriptomics of Closely Related Species. *PLoS ONE*, 8 (11), e78497.

Poeschl, Y., Grosse, I., and Gogol-Döring, A. (2014). Explaining gene responses by linear modeling. *German Conference on Bioinformatics*, Volume P-235 of Lecture Notes in Informatics (LNI) - Proceedings, pp. 27–35.

Pollard, K. S., Dudoit, S., and Laan, M. J. van der (2005). *Multiple Testing Procedures: R multtest Package and Applications to Genomics, in Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer.

Porra, R., Thompson, W., and Kriedemann, P. (1989). Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 975 (3), pp. 384–394.

Proveniers, M. C. and Zanten, M. van (2013). High temperature acclimation through PIF4 signaling. *Trends in Plant Science*, 18 (2), pp. 59–64.

Quint, M., Barkawi, L. S., Fan, K.-T., Cohen, J. D., and Gray, W. M. (2009). *Arabidopsis* IAR4 Modulates Auxin Response by Regulating Auxin Homeostasis. *Plant Physiology*, 150 (2), pp. 748–758.

Quint, M. and Gray, W. M. (2006). Auxin signaling. *Current Opinion in Plant Biology*, 9 (5), pp. 448–453.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vol. ISBN 3-900051-07-0. Vienna, Austria.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. URL: R-project website. Available: http://www.R-project.org/. Accessed 2013 October 8.

Rademacher, E. H., Möller, B., Lokerse, A. S., Llavata-Peris, C. I., Berg, W. van den, and Weijers, D. (2011). A cellular expression map of the *Arabidopsis* AUXIN RESPONSE FACTOR gene family. *The Plant Journal*, 68 (4), pp. 597–606.

Ramos, J. A., Zenser, N., Leyser, O., and Callis, J. (2001). Rapid Degradation of Auxin/Indoleacetic Acid Proteins Requires Conserved Amino Acids of Domain II and Is Proteasome Dependent. *The Plant Cell*, 13 (10), pp. 2349–2360.

Raschke, A., Ibañez, C., Ullrich, K. K., Anwer, M. U., Becker, S., Glöckner, A., Trenner, J., Denk, K., Saal, B., Sun, X., Ni, M., Davis, S. J., Delker, C., and Quint, M. (2015).

Natural Variants of ELF3 Affect Thermomorphogenesis by Transcriptionally Modulating PIF4-Dependent Auxin Response Genes. *bioRxiv*.

Redman, J. C., Haas, B. J., Tanimoto, G., and Town, C. D. (2004). Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *The Plant Journal*, 38 (3), pp. 545–561.

Rowe, H. C. and Kliebenstein, D. J. (2008). Complex Genetics Control Natural Variation in *Arabidopsis thaliana* Resistance to *Botrytis cinerea*. *Genetics*, 180 (4), pp. 2237–2250.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *SCIENCE*, 290, pp. 2323–2326.

Ruegger, M., Dewey, E., Gray, W. M., Hobbie, L., Turner, J., and Estelle, M. (1998). The TIR1 protein of *Arabidopsis* functions in auxin response and is related to human SKP2 and yeast Grr1p. *Genes & Development*, 12 (2), pp. 198–207.

Salehin, M., Bagchi, R., and Estelle, M. (2015). SCFTIR1/AFB-Based Auxin Perception: Mechanism and Role in Plant Growth and Development. *The Plant Cell*, 27 (1), pp. 9–19.

Salmon, J., Ramos, J., and Callis, J. (2008). Degradation of the auxin response factor ARF1. *The Plant Journal*, 54 (1), pp. 118–128.

Scholl, R. L., May, S. T., and Ware, D. H. (2000). Seed and Molecular Resources for Arabidopsis. *Plant Physiology*, 124 (4), pp. 1477–1480.

Stavang, J. A., Gallego-Bartolomé, J., Gómez, M. D., Yoshida, S., Asami, T., Olsen, J. E., García-Martínez, J. L., Alabadí, D., and Blázquez, M. A. (2009). Hormonal regulation of temperature-induced growth in *Arabidopsis*. *The Plant Journal*, 60 (4), pp. 589–601.

Stomp, A.-M. (1991). Histochemical localization of β-glucuronidase. In: *GUS Protocols*. Ed. by S. Gallagher. London: Academic Press, pp. 103–113.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., and Huala, E. (2008). The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research*, 36 (suppl 1), pp. D1009–D1014.

Szemenyei, H., Hannon, M., and Long, J. A. (2008). TOPLESS Mediates Auxin-Dependent Transcriptional Repression During *Arabidopsis* Embryogenesis. *Science*, 319 (5868), pp. 1384–1386.

Teale, W. D., Paponov, I. A., and Palme, K. (2006). Auxin in action: signalling, transport and the control of plant growth and development. *Nature Reviews Molecular Cell Biology*, 7 (11), pp. 847–859.

The Trans-Bee workshop (2014). http://www.idiv-biodiversity.de/sdiv/workshops/workshops-2013/stransbee (accessed 2014/07/23).

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L. A., Rhee, S. Y., and Stitt, M. (2004). MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal*, 37 (6), pp. 914–939.

Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., and Prentice, I. C. (2005). Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (23), pp. 8245–8250.

Tian, Q. and Reed, J. (1999). Control of auxin-regulated root development by the *Arabidopsis thaliana* SHY2/IAA3 gene. *Development*, 126 (4), pp. 711–721.

Tirosh, I., Weinberger, A., Carmi, M., and Barkai, N. (2006). A genetic signature of interspecies variations in gene expression. *Nature Genetics*, 38 (7), pp. 830–834.

Tiwari, S. B., Hagen, G., and Guilfoyle, T. (2003). The Roles of Auxin Response Factor Domains in Auxin-Responsive Transcription. *The Plant Cell*, 15 (2), pp. 533–543.

Toledo-Ortiz, G., Johansson, H., Lee, K. P., Bou-Torrent, J., Stewart, K., Steel, G., Rodríguez-Concepción, M., and Halliday, K. J. (2014). The HY5-PIF Regulatory Module Coordinates Light and Temperature Control of Photosynthetic Gene Transcription. *PLoS Genetics*, 10 (6), e1004416.

Trenner, J., Poeschl, Y., Grau, J., Gogol-Döring, A., Quint, M., and Delker, C. (in prep.). Variation of IAA-induced transcriptomes pinpoints the AUX/IAA network as a potential source for inter-species divergence in auxin signaling and response. *not announced*.

Ulmasov, T., Murfett, J., Hagen, G., and Guilfoyle, T. J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell*, 9 (11), pp. 1963–71.

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. (1999). Activation and repression of transcription by auxin-response factors. *Proceedings of the National Academy of Sciences*, 96 (10), pp. 5844–5849.

Vert, G., Walcher, C. L., Chory, J., and Nemhauser, J. L. (2008). Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proceedings of the National Academy of Sciences*, 105 (28), pp. 9829–9834.

Vinga, S. and Almeida, J. (2003). Alignment-free sequence comparison—a review. *Bioinformatics*, 19 (4), pp. 513–523.

Walcher, C. L. and Nemhauser, J. L. (2012). Bipartite Promoter Element Required for Auxin Response. *Plant Physiology*, 158 (1), pp. 273–282.

Wang, Y. R. and Huang, H. (2014). Review on statistical methods for gene network reconstruction using expression data. *Journal of Theoretical Biology*.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W. H. A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, B. (2014). *gplots: Various R programming tools for plotting data*. R package version 2.15.0.

Weijers, D., Benkova, E., Jäger, K. E., Schlereth, A., Hamann, T., Kientz, M., Wilmoth, J. C., Reed, J. W., and Jürgens, G. (2005). Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators. *The EMBO Journal*, 24 (10), pp. 1874–1885.

Wilson, C. L. and Miller, C. J. (2005). Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*, 21 (18), pp. 3683–3685.

Yilmaz, A., Mejia-Guerra, M. K., Kurz, K., Liang, X., Welch, L., and Grotewold, E. (2011). AGRIS: the Arabidopsis Gene Regulatory Information Server, an update. *Nucleic Acids Research*, 39 (suppl 1), pp. D1118–D1122.

Yona, G., Dirks, W., Rahman, S., and Lin, D. M. (2006). Effective similarity measures for expression profiles. *Bioinformatics*, 22 (13), pp. 1616–1622.

Zanten, M. van, Voesenek, L. A., Peeters, A. J., and Millenaar, F. F. (2009). Hormone- and Light-Mediated Regulation of Heat-Induced Differential Petiole Growth in *Arabidopsis*. *Plant Physiology*, 151 (3), pp. 1446–1458.

Zenser, N., Ellsmore, A., Leasure, C., and Callis, J. (2001). Auxin modulates the degradation rate of Aux/IAA proteins. *Proceedings of the National Academy of Sciences*, 98 (20), pp. 11795–11800.

Zhang, X., Shiu, S., Cal, A., and Borevitz, J. O. (2008). Global Analysis of Genetic, Epigenetic and Transcriptional Polymorphisms in *Arabidopsis thaliana* Using Whole Genome Tiling Arrays. *PLoS Genetics*, 4 (3), e1000032.

Appendix

A. Supporting Information: Natural variation of transcriptional auxin response networks in Arabidopsis thaliana

A.1. Figures

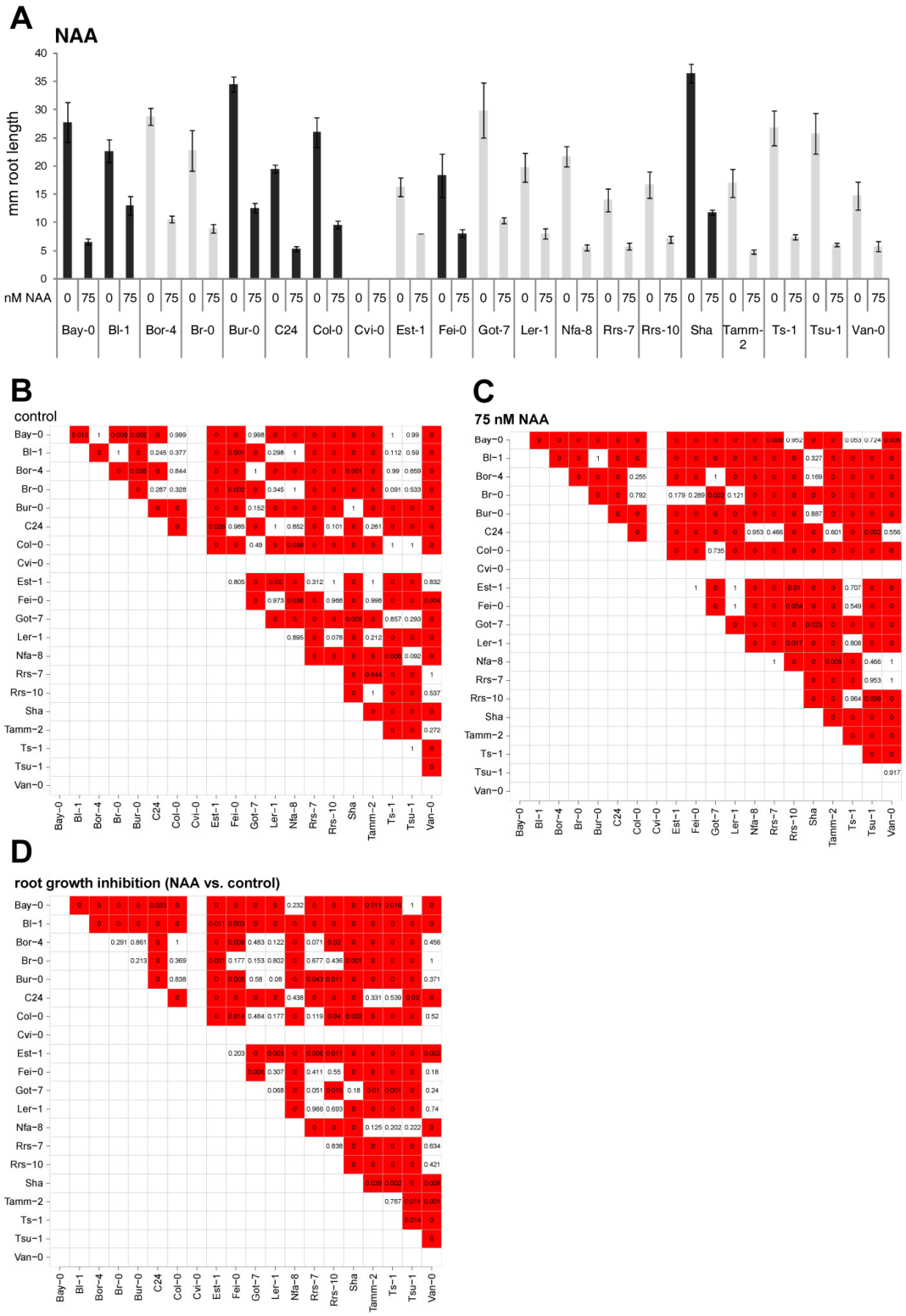


Figure A.1.: NAA root growth assay (data and statistics). (A) Seedling root lengths grown on unsupplemented (control) or 75 nM NAA containing medium, respectively (n=12). Error bars denote standard deviations in percent. **(B+C)** Statistical differences between root lengths of accessions were analyzed by one-way ANOVA and Tukey test. **(D)** Differences in growth responses were assessed by performing two-way ANOVA on individual accession pairs. All analyses were performed on log-transformed data. Tables show p-values. Significant differences ($p<0.05$) are highlited in red. 0 indicates p-values <0.001.

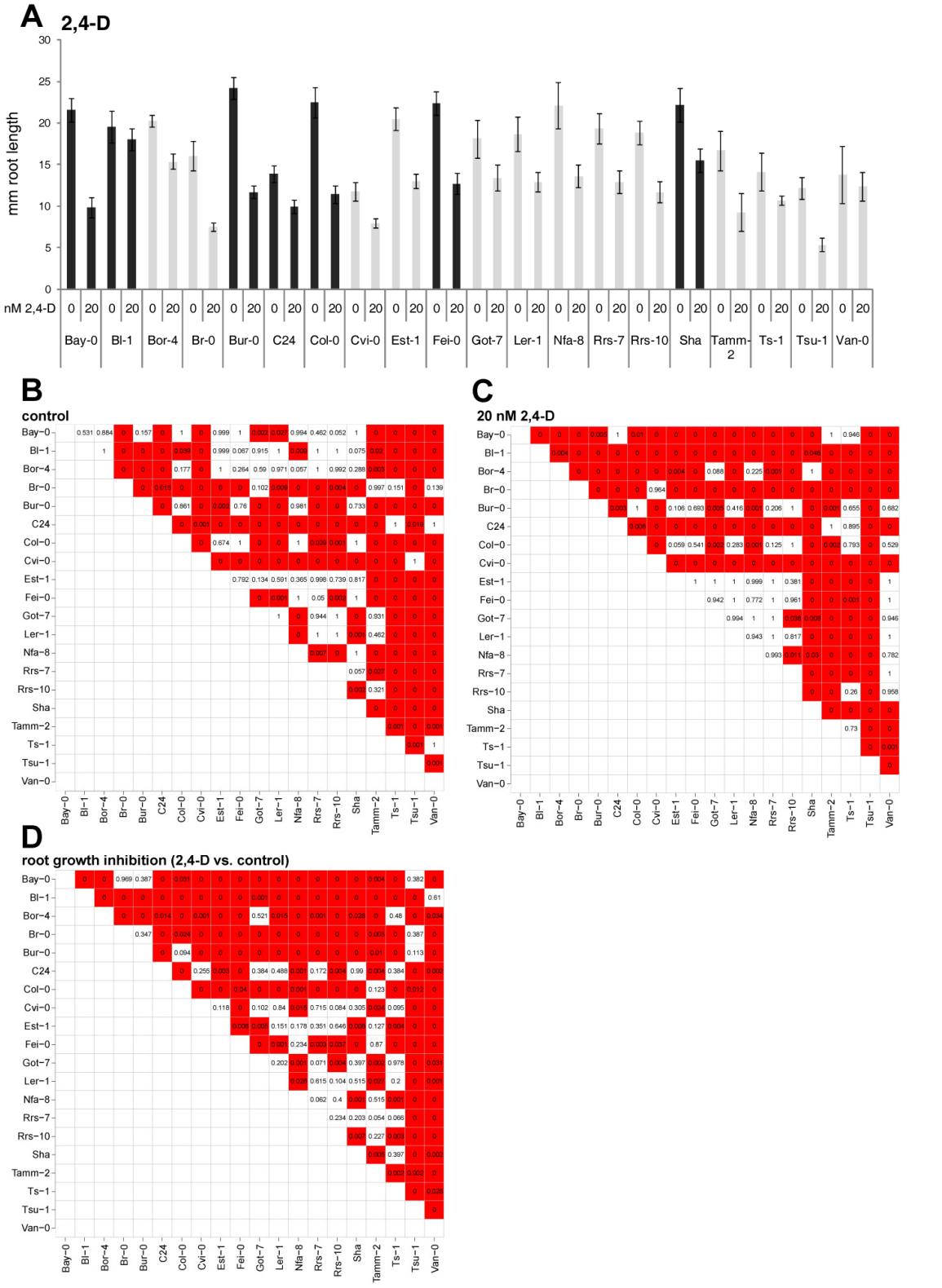


Figure A.2.: 2,4-D root growth assay (data and statistics). **(A)** Seedling root lengths grown on unsupplemented (control) or 20 nM 2,4-D containing medium, respectively (n=12). Error bars denote standard deviations in percent. **(B+C)** Statistical differences between root lengths of accessions were analyzed by one-way ANOVA and Tukey test. **(D)** Differences in growth responses were assessed by performing two-way ANOVA on individual accession pairs. All analyses were performed on log-transformed data. Tables show p-values. Significant differences ($p<0.05$) are highlited in red. 0 indicates p-values <0.001.

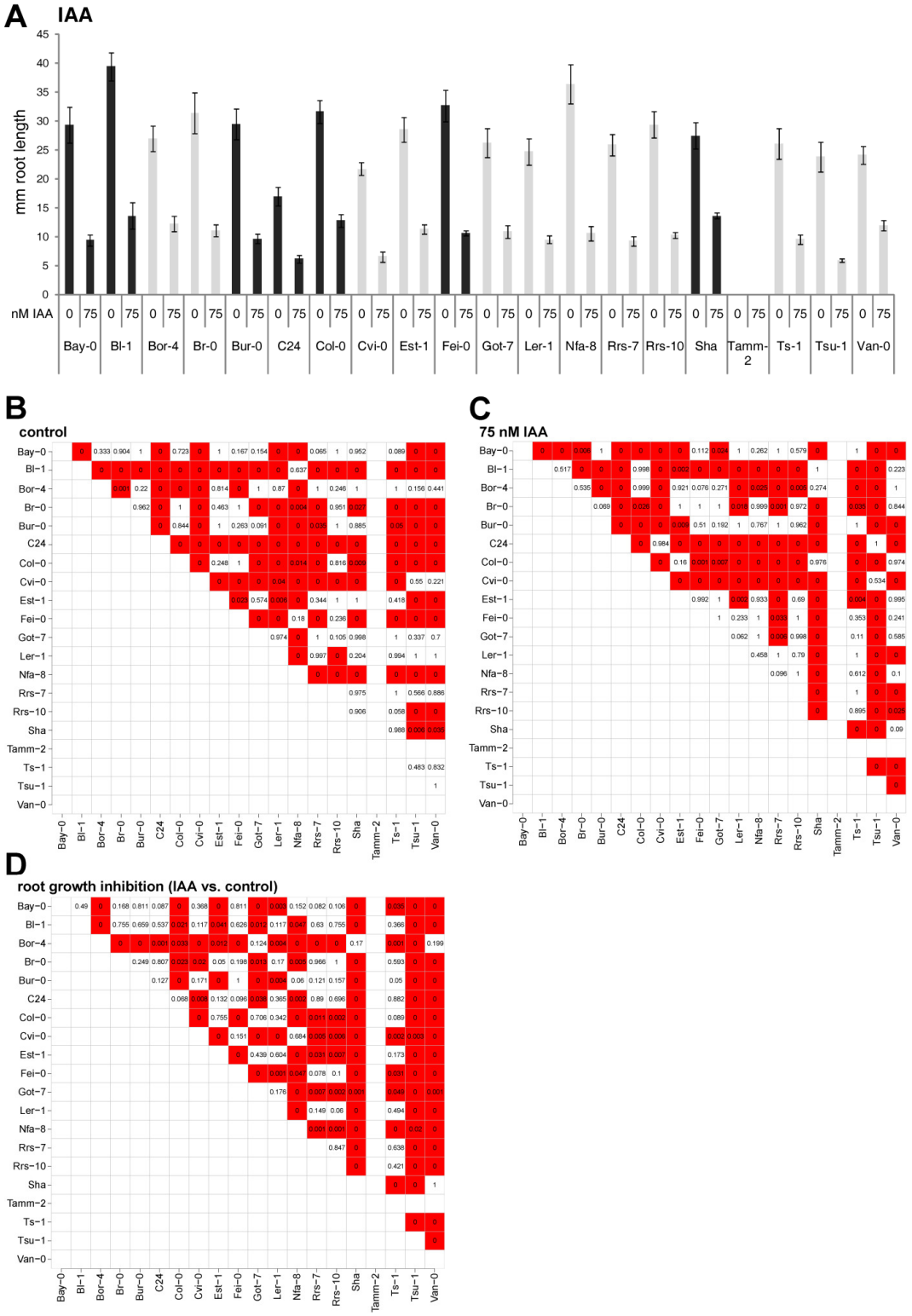


Figure A.3.: IAA root growth assay (data and statistics). **(A)** Seedling root lengths grown on unsupplemented (control) or 75 nM IAA containing medium, respectively (n=12). Error bars denote standard deviations in percent. **(B+C)** Statistical differences between root lengths of accessions were analyzed by one-way ANOVA and Tukey test. **(D)** Differences in growth responses were assessed by performing two-way ANOVA on individual accession pairs. All analyses were performed on log-transformed data. Tables show p-values. Significant differences ($p<0.05$) are highlited in red. 0 indicates p-values <0.001.

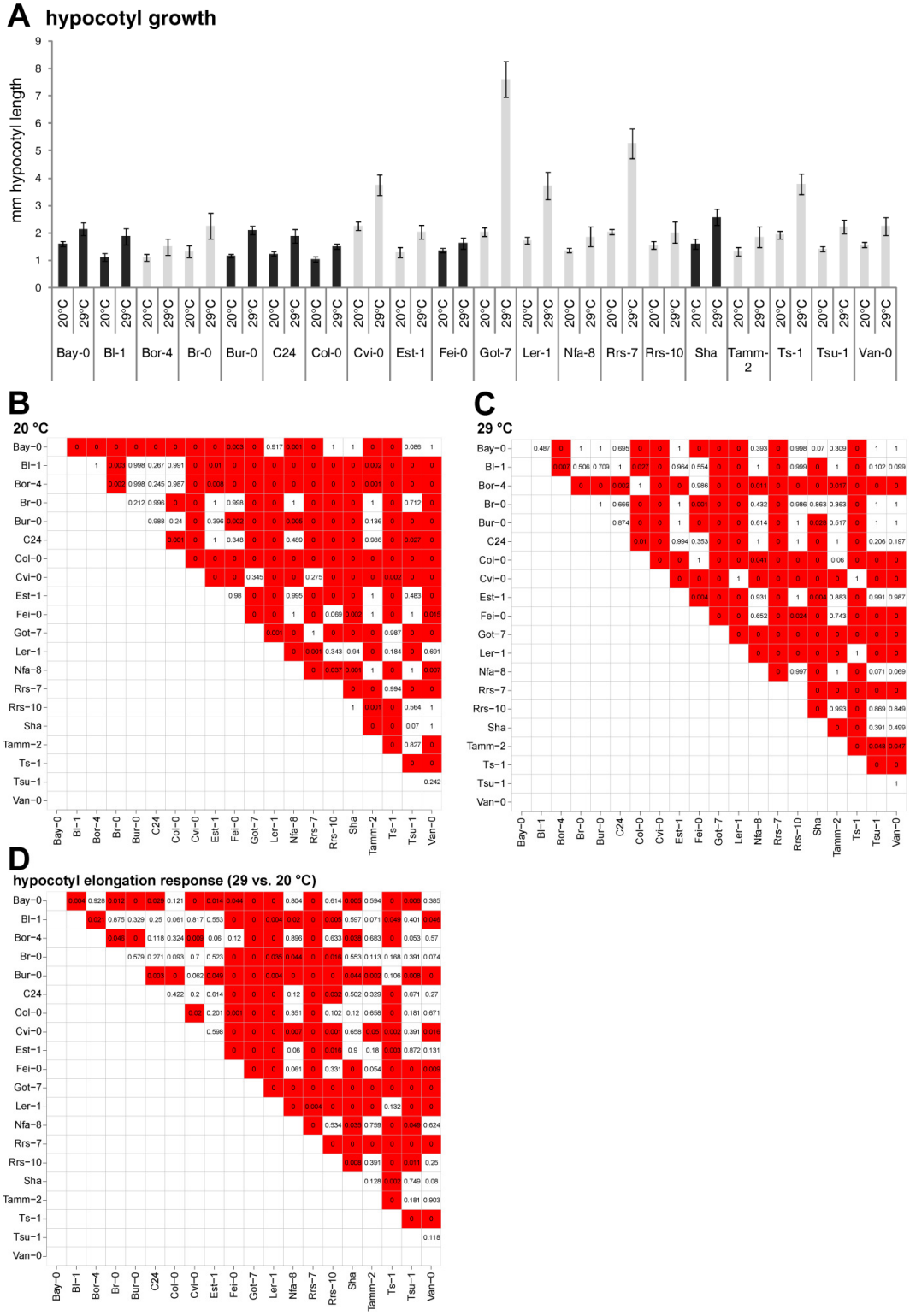


Figure A.4.: Temperature-induced hypocotyl elongation assay (data and statistics). (A) Seedling hypocotyl lengths grown at 20 and 29∘C, respectively (n=12). Error bars denote standard deviations in percent. **(B+C)** Statistical differences between hypocotyl lengths of accessions were analysed by one-way ANOVA and Tukey test. **(D)** Differences in elongation responses were assessed by performing two-way ANOVA on individual accession pairs. All analyses were performed on log-transformed data. Tables show p-values. Significant differences ($p<0.05$) are highlited in red. 0 indicates p-values <0.001.

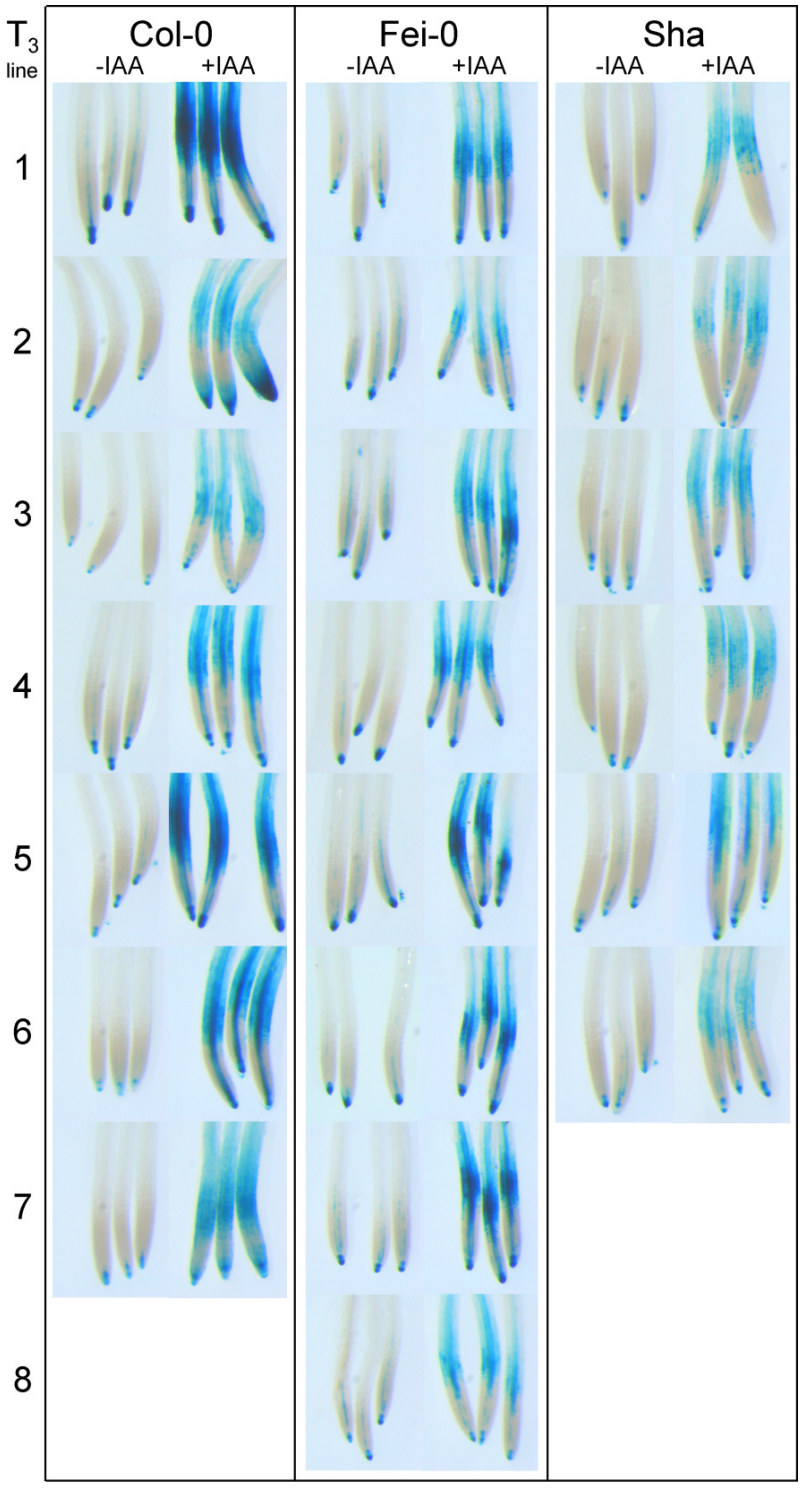


Figure A.5.: Histochemical detection of DR5:GUS activity (individual lines). DR5:GUS activity was detected in individual T_3 lines after 3 h treatment with mock (-IAA) or 1 μM IAA (+IAA).

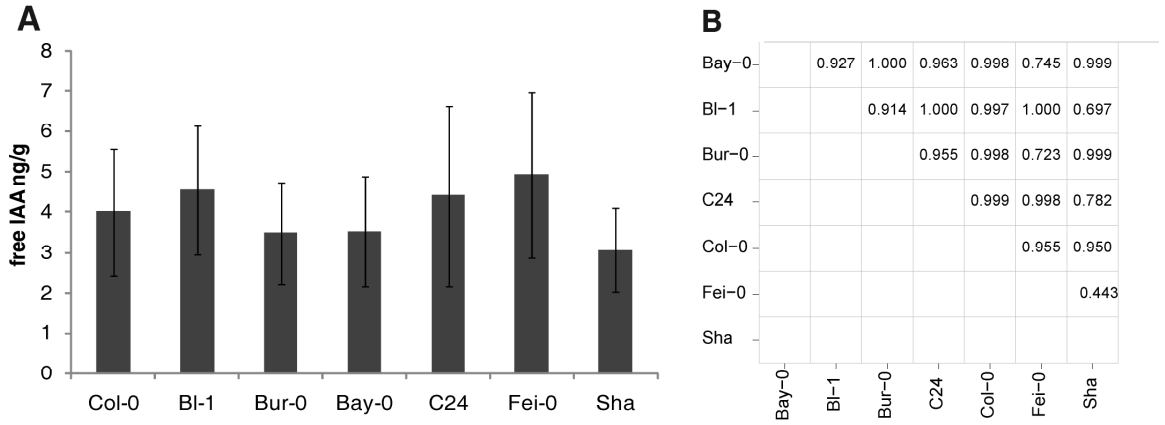


Figure A.6.: IAA quantitation. (A) Mean levels of free IAA in pooled plant material (∼500 mg) of 7-d-old seedlings (n= 6). Error bars denote standard deviations. **(B)** Table of p-values obtained by one-way ANOVA and Tukey test. No significant differences between accessions were identified.

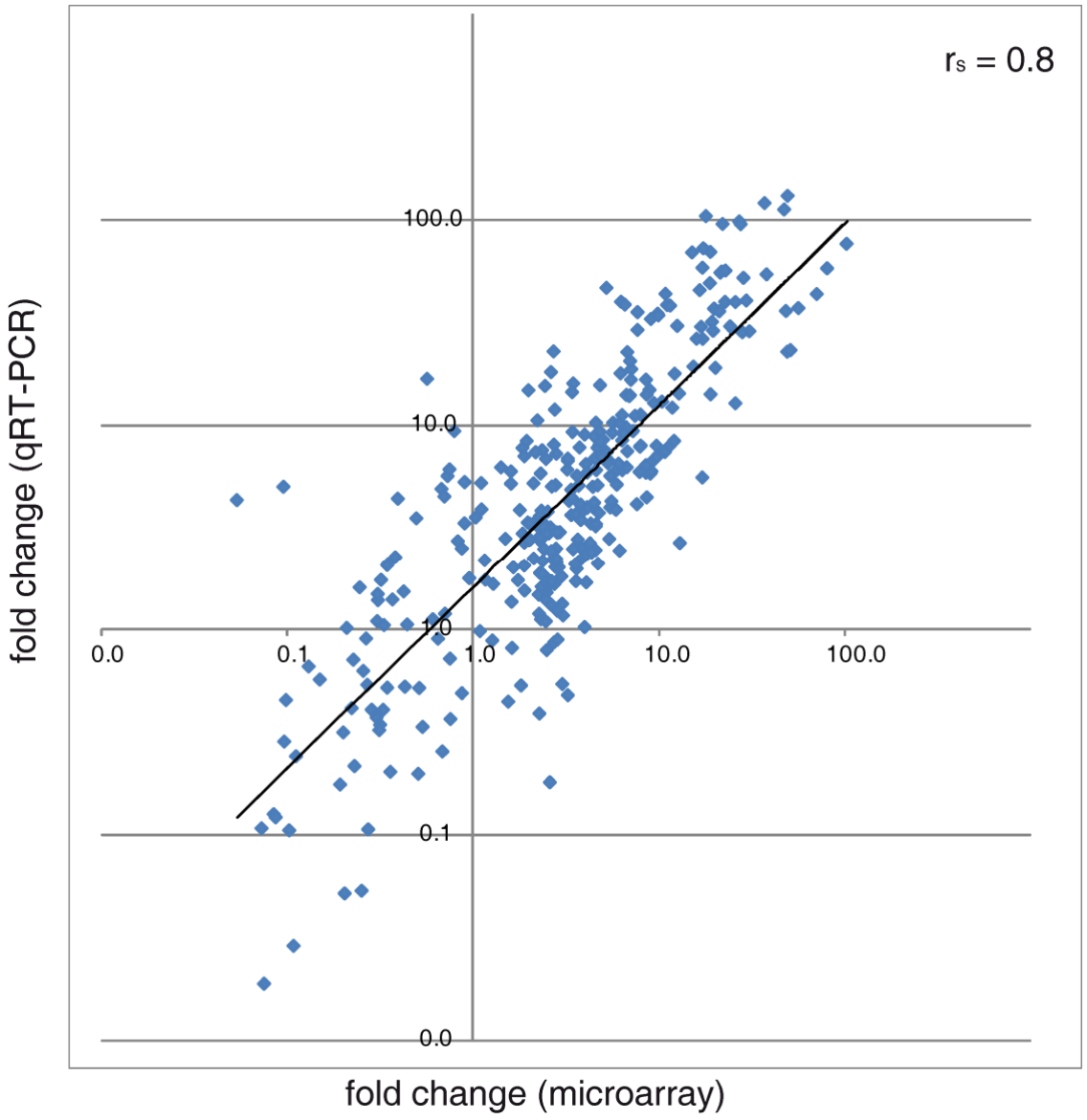


Figure A.7.: Correlation of qRT-PCR and microarray data. Expression of 18 arbitrary genes was independently re-examined by qRT-PCR for all seven accessions (see Supplemental Table A.3 for primer sequences). For qRT-PCR, comparative expression levels (CELs) were determined relative to the constitutively expressed gene *At1g13320*. Fold changes in the expression of individual genes were calculated using mean CEL values of three biological replicates for each time point (0.5 h, 1 h, and 3 h) relative to the mean CEL of untreated seedlings (0 h). Fold changes in expression detected by qRT-PCR and microarray were plotted on a log scale for each gene at individual time points. Spearman's correlation coefficient between both data sets was calculated in R using default settings.

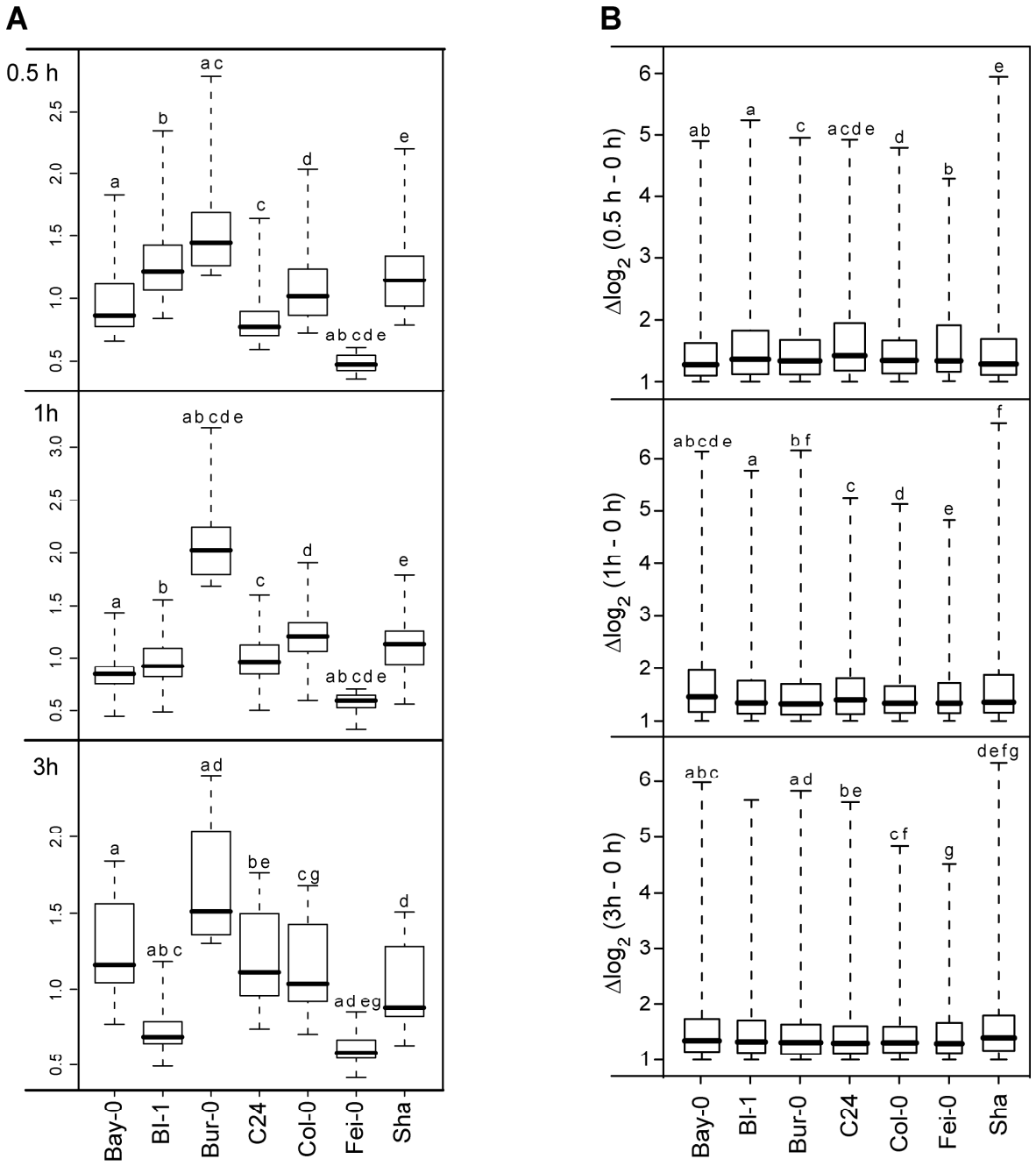


Figure A.8.: Identification of putative hyper- and hypo-responsive accessions. **(A)** Ratios of numbers of differentially expressed genes ($\Delta \log_2 > 1$) and **(B)** the $\Delta \log_2$ expression changes of a given accession and every other accessions were calculated and subjected to statistical analysis. Boxplots show interquartile ranges and medians (black bar), whiskers comprise min - max range of values. Samples marked with identical letters differ significantly ($p < 0.05$) in pair-wise comparisons (U-test).

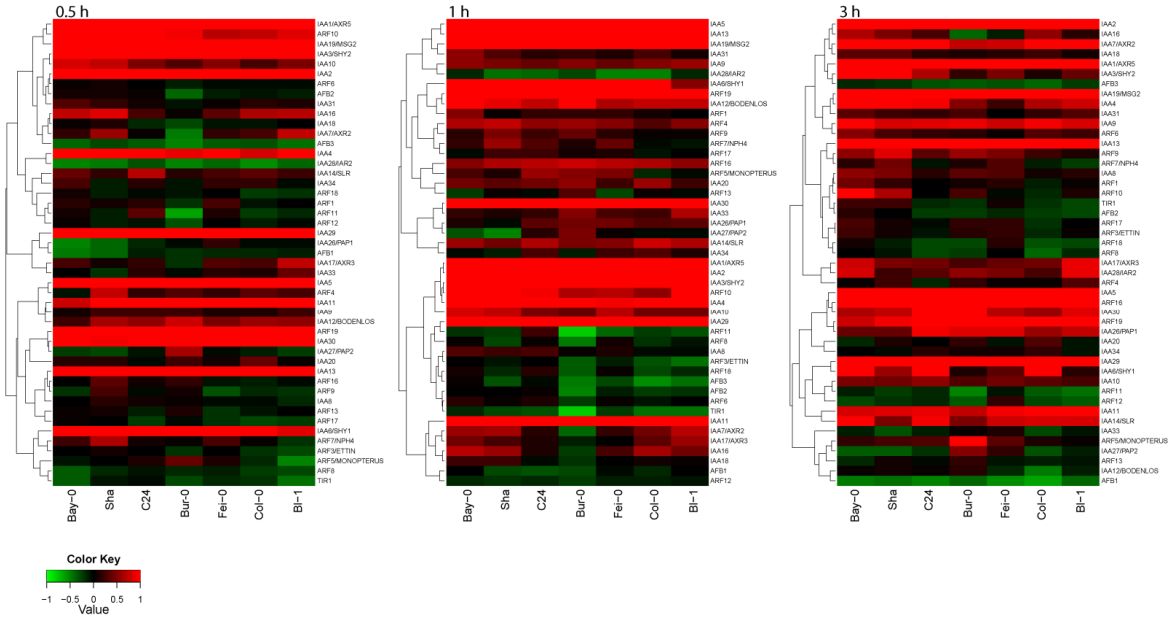


Figure A.9.: Heat maps of signaling gene expression. Heat map representation of expression changes ($\Delta \log_2$) of signaling genes for individual time points p.i. (cf. Supplemental Table A.1 online for complete list of genes).

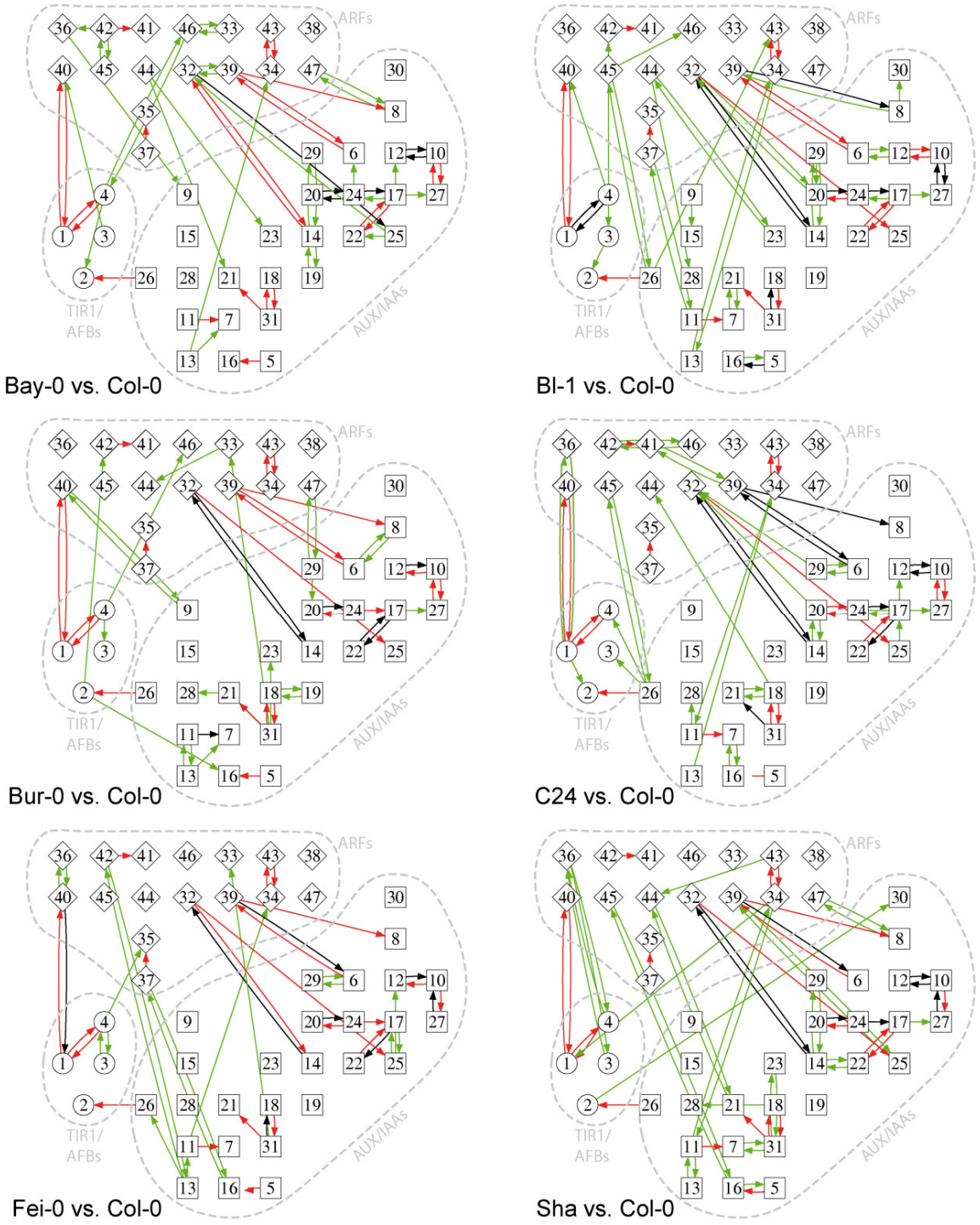
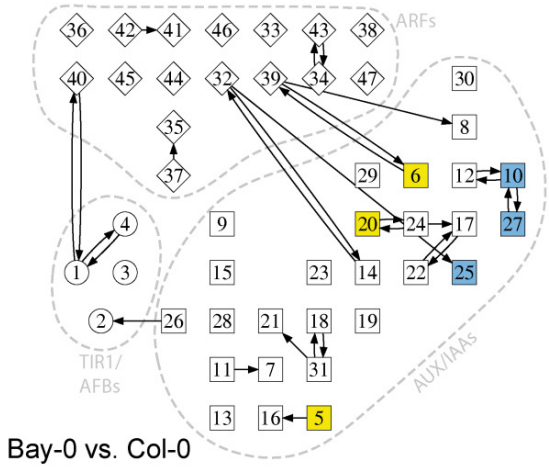
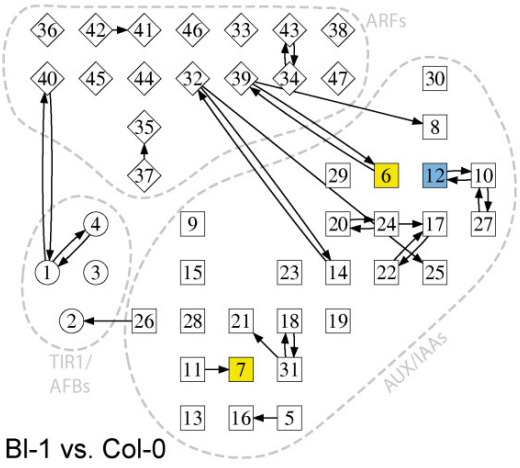


Figure A.10.: Pair-wise comparisons of LCF networks for signaling genes. Expression changes of genes encoding TIR1/AFB auxin receptors (1-4, circles), AUX/IAA proteins (5-31, squares) and ARFs (32-47, diamonds) were analyzed by LCF to identify patterns of co-regulation in each accession (cf. Supplemental Table A.1 online for complete list of genes). Individual networks were compared to the Col-0 reference network exhibiting 29 edges that connect individual node pairs and indicate co-regulation of the respective genes. Edges with a bootstrapping value of > 0.75 are presented. Red edges indicate connections detected specifically in the network of Col-0, green edges are specific for the connections in the respective other accessions and black edges represent connections detected in both networks.

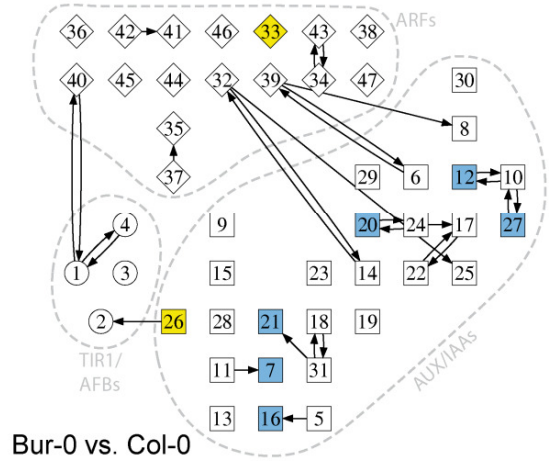
A



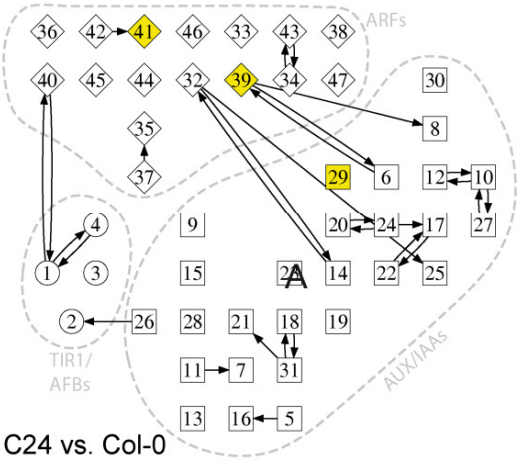
Bay-0 vs. Col-0



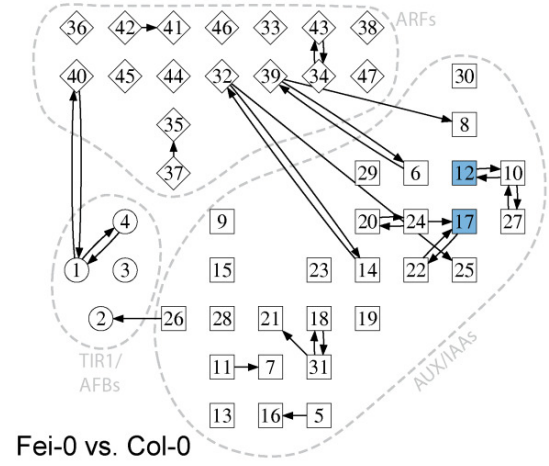
Bl-1 vs. Col-0



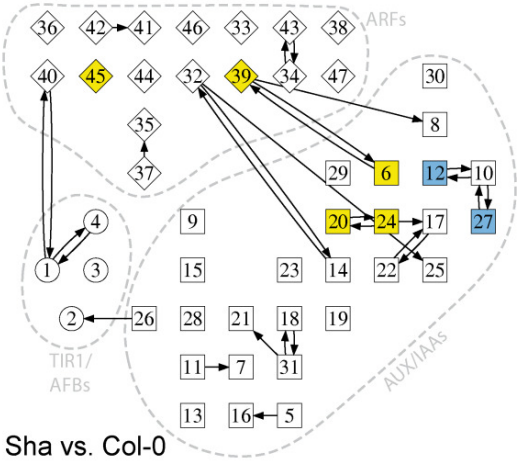
Bur-0 vs. Col-0



C24 vs. Col-0

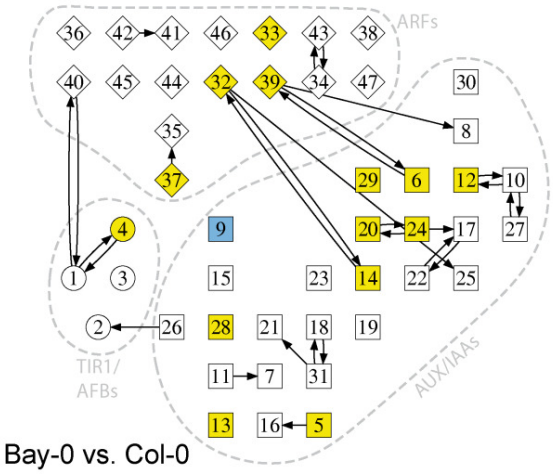


Fei-0 vs. Col-0

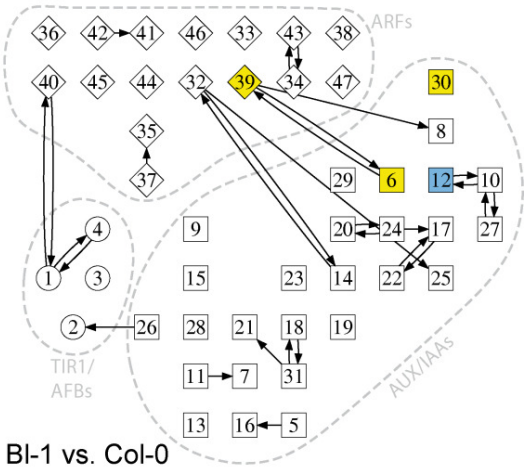


Sha vs. Col-0

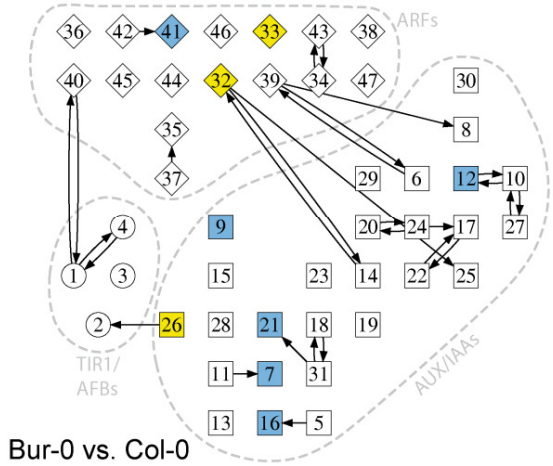
B



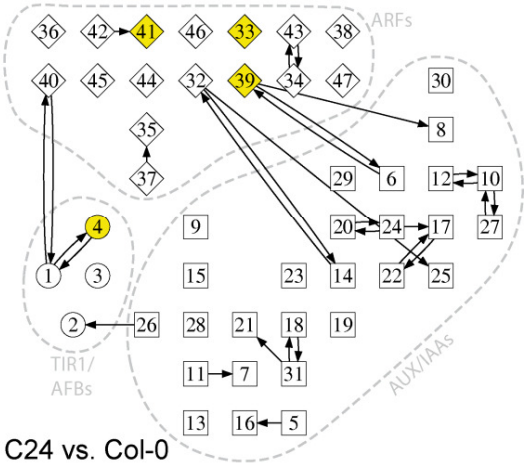
Bay-0 vs. Col-0



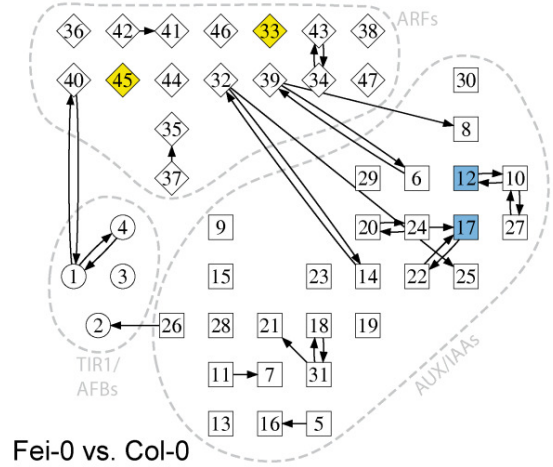
Bl-1 vs. Col-0



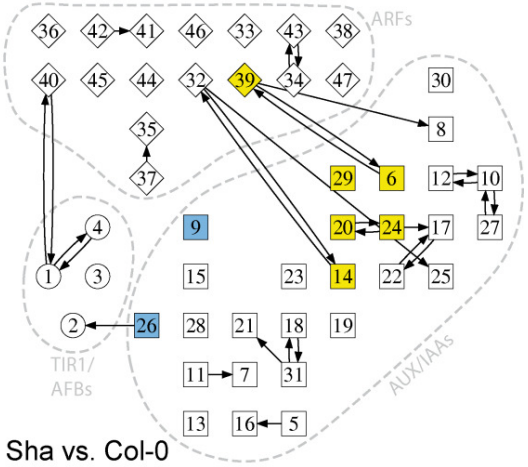
Bur-0 vs. Col-0



C24 vs. Col-0



Fei-0 vs. Col-0



Sha vs. Col-0

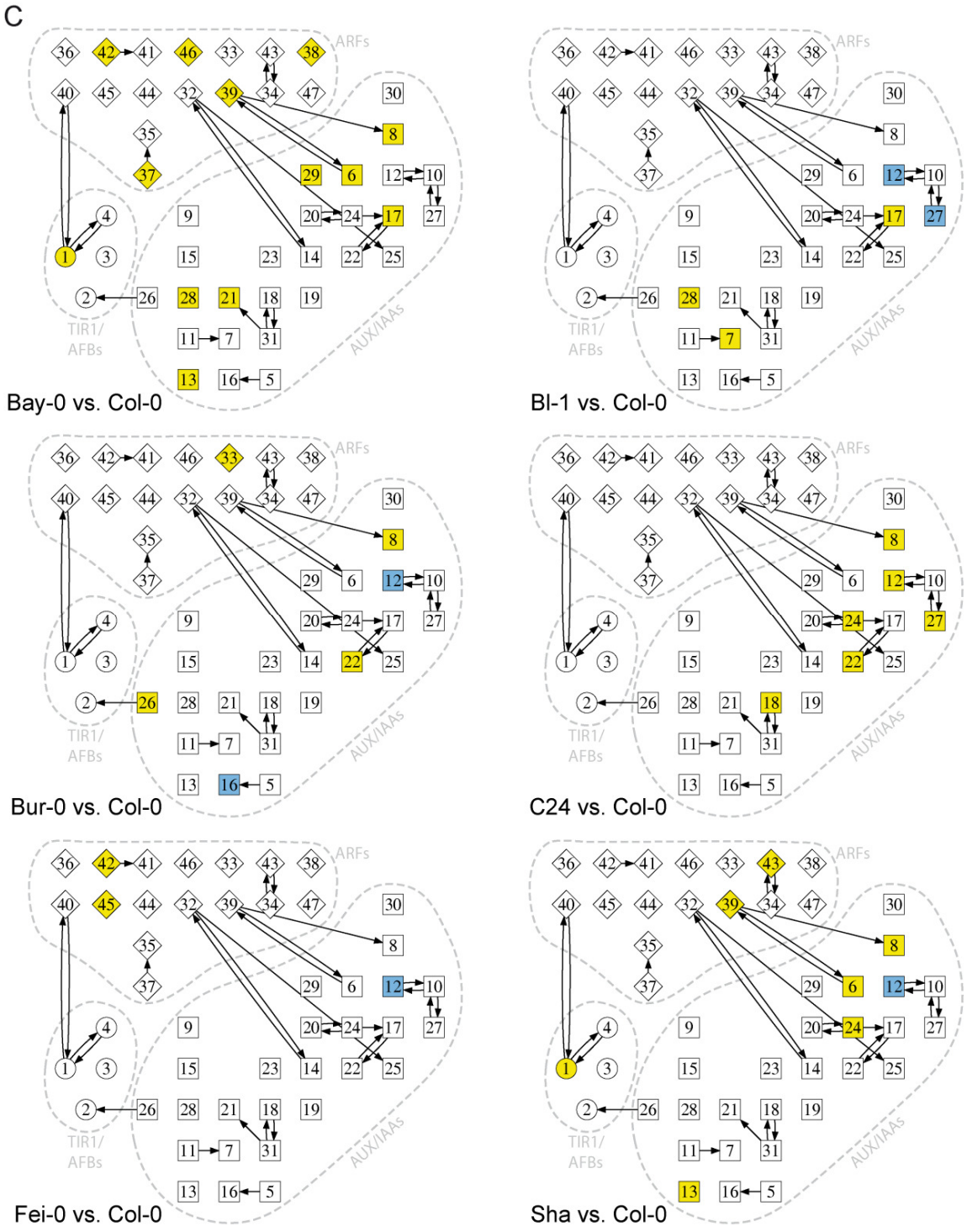


Figure A.11.: Pair-wise comparison of signaling gene expression. Auxin-induced expression changes in auxin signaling genes represented by a single specific probe on the ATH1 chip (see Supplemental Table 1 online). Significant differences in auxin-induced expression changes for signaling genes **(A)** 0.5 h, **(B)** 1 h and **(C)** 3 h p.i. are highlighted in blue and yellow for significantly lower or higher expression responses, respectively, detected in the indicated accession in comparison to Col-0 ($p < 0.05$, modified t-statistics and Benjamini-Hochberg correction). The network structure was obtained by LCF analysis of Col-0 data. Edges presented had a bootstrapping value of > 0.5.

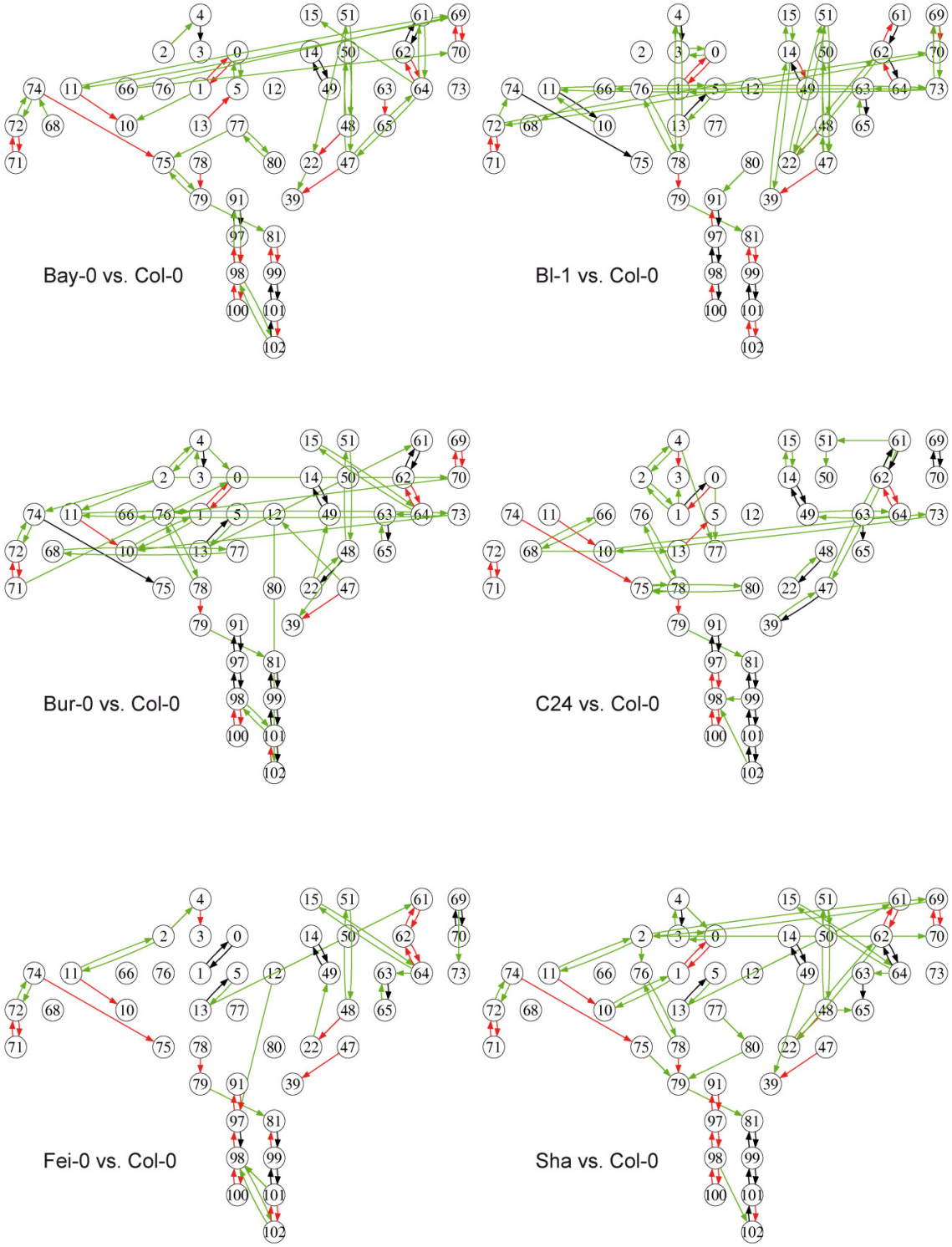
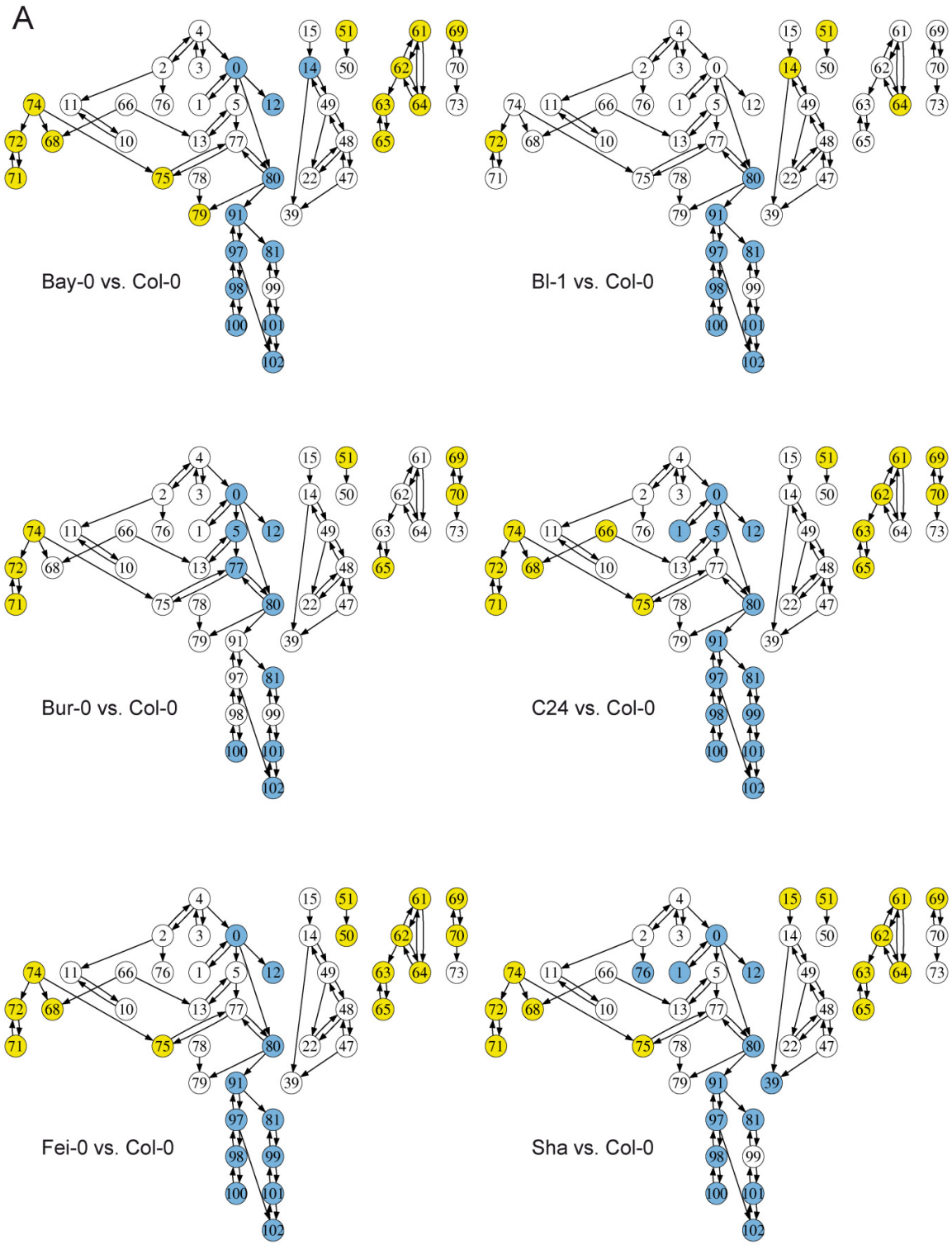


Figure A.12.: Pair-wise comparison of LCF networks of cluster genes. Mean expression profiles of clusters were analyzed by LCF to identify patterns of co-regulation in each accession. Clusters were determined based on the Col-0 dataset by hierarchical clustering. Networks obtained for individual accessions were compared to the Col-0 reference network. Red edges indicate connections detected specifically in the network of Col-0, edges in green are specific for the respective connections in other accessions and black edges are common in both networks.


A
Bay-0 vs. Col-0
Bl-1 vs. Col-0
Bur-0 vs. Col-0
C24 vs. Col-0
Fei-0 vs. Col-0
Sha vs. Col-0

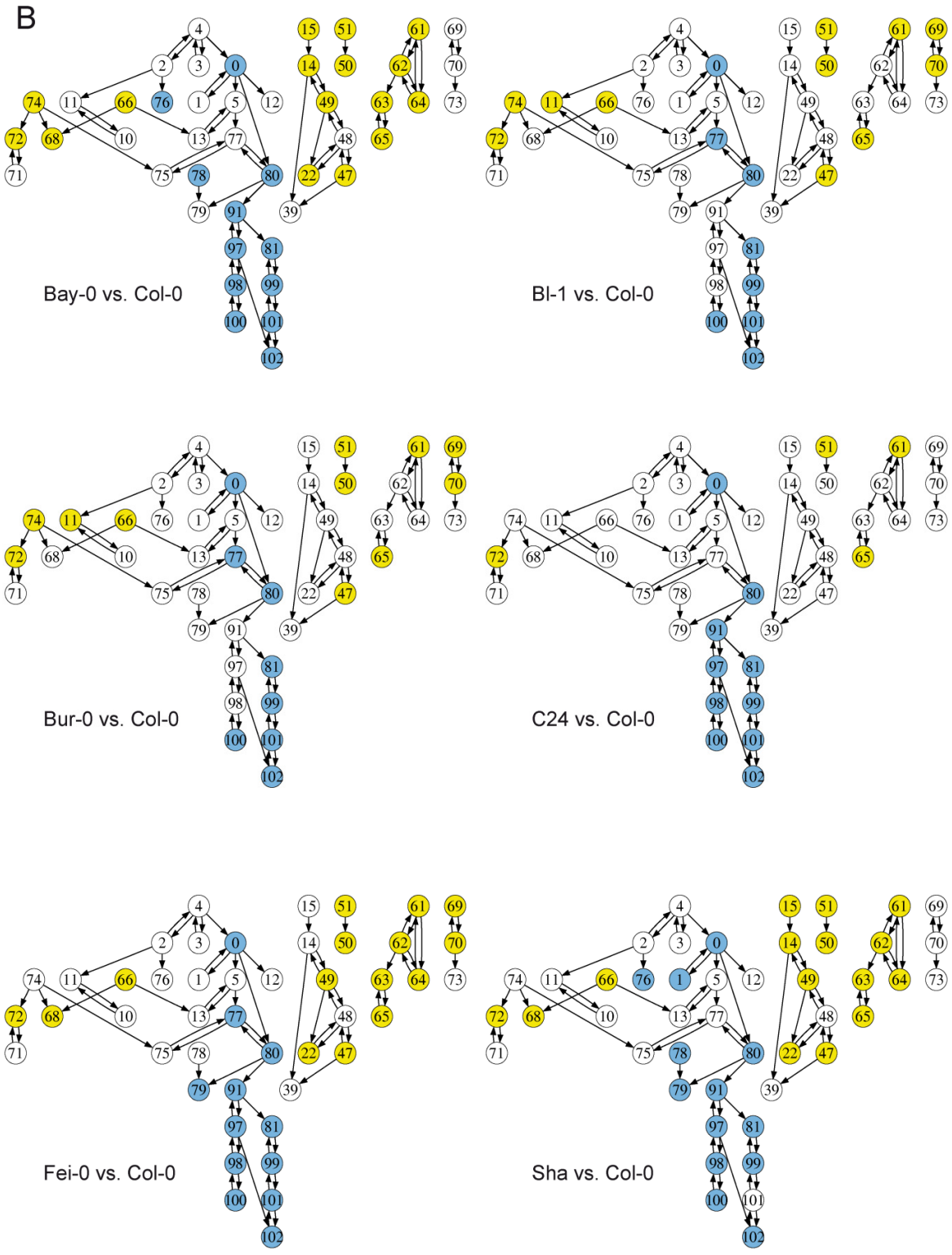


Figure A.13.: Pair-wise comparison of cluster gene expression. Significant alterations in mean expression changes **(A)** 0.5 h, and **(B)** 3 h p.i. are highlighted in blue and yellow for significantly lower or higher expression responses, respectively, detected in the indicated accession in comparison to Col-0 ($p < 0.05$, modified t-test and Benjamini-Hochberg correction). The network structure was obtained by LCF analysis of Col-0 data (Supplemental Figure A.12 online).

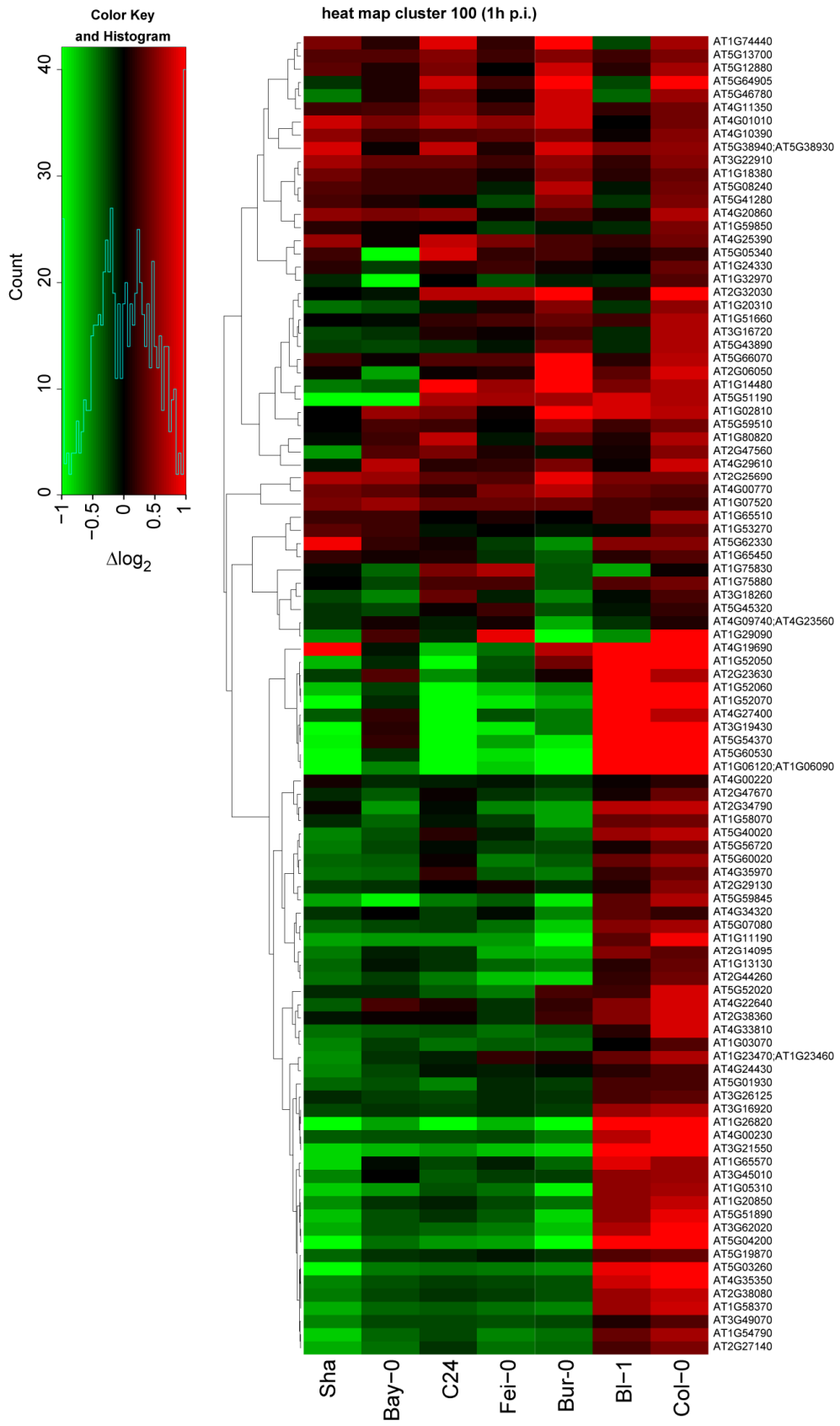
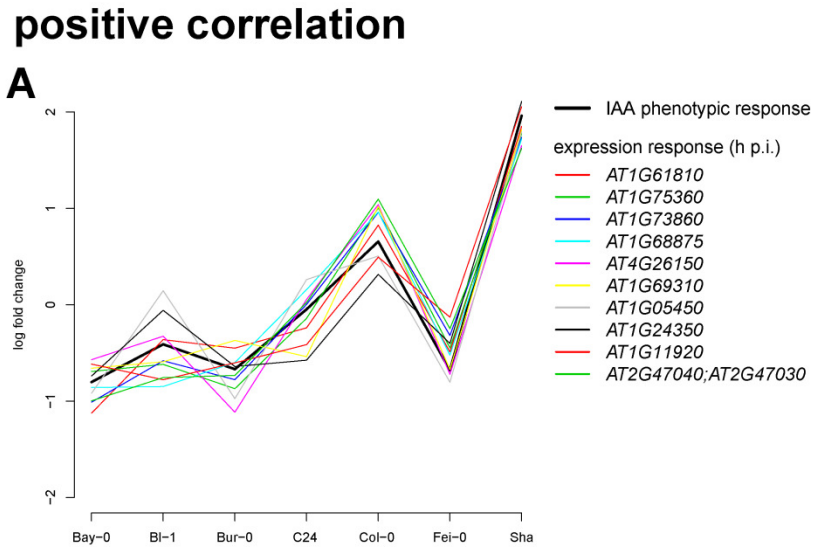


Figure A.14.: Detailed heat map presentation of mean expression changes ($\Delta \log_2$) of all genes within cluster 100 (1 h p.i.). A histogram of $\Delta \log_2$ values is shown on the left as a turquoise line within the color key.



B

Biological Process	p-value
GO:0050896 response to stimulus	1.08E-04
GO:0009725 response to hormone stimulus	9.50E-03
Molecular Function	
GO:0003700 transcription factor activity	1.19E-04
GO:0030528 transcription regulator activity	1.30E-04
GO:0003677 DNA binding	4.45E-03

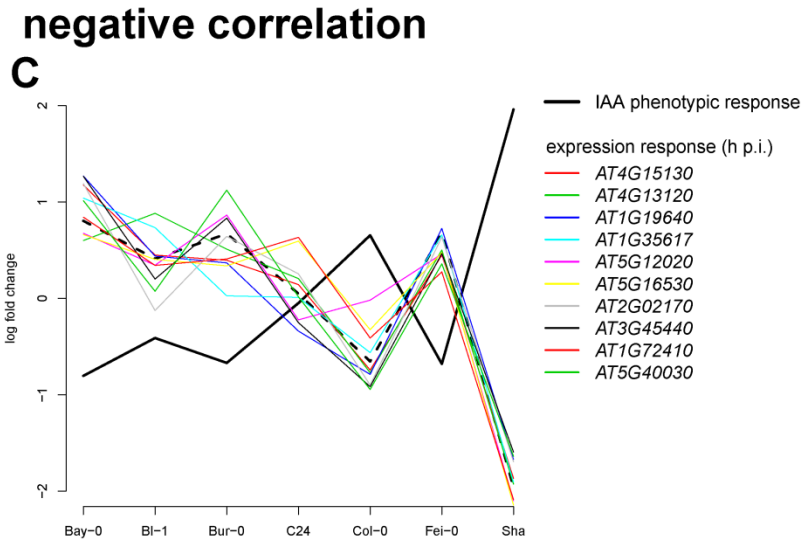


Figure A.15.: Correlation between phenotypic and expression responses to IAA. Pearson correlation coefficients (r) were generated by comparison of $\Delta \log_2$ values of physiological IAA responses and mean expression responses of individual genes (0,5 h and 1 h combined). Genes with r > 0.8 and < -0.8 were considered to be positively and negatively correlated, respectively. **(A)** $\Delta \log_2$ profiles of the 10 genes with highest r-values (coloured lines) and of physiological IAA responses (solid black line). **(B)** Over-represented GO terms in the 230 genes with r > 0.8. **(C)** $\Delta \log_2$ profiles of 10 genes with lowest r-values (coloured lines). The $\Delta \log_2$ profile of physiological IAA responses is shown as observed (solid black line) and mirrored (broken black line) to visualize a perfect negative correlation.

A.2. Tables

Table A.1.: Signaling genes selected for LCF analyses.

No.	AGI	gene
1	*AT3G62980*	*TIR1*
2	*AT4G03190*	*AFB1*
3	*AT3G26810*	*AFB2*
4	*AT1G12820*	*AFB3*
5	*AT1G04100*	*IAA10*
6	*AT1G04240*	*IAA3/SHY2*
7	*AT1G04250*	*IAA17/AXR3*
8	*AT1G04550*	*IAA12/BODENLOS*
9	*AT1G15050*	*IAA34*
10	*AT1G15580*	*IAA5*
11	*AT1G51950*	*IAA18*
12	*AT1G52830*	*IAA6/SHY1*
13	*AT2G22670*	*IAA8*
14	*AT2G33310*	*IAA13*
15	*AT2G46990*	*IAA20*
16	*AT3G04730*	*IAA16*
17	*AT3G15540*	*IAA19/MSG2*
18	*AT3G16500*	*IAA26/PAP1*
19	*AT3G17600*	*IAA31*
20	*AT3G23030*	*IAA2*
21	*AT3G23050*	*IAA7/AXR2*
22	*AT3G62100*	*IAA30*
23	*AT4G14550*	*IAA14/SLR*
24	*AT4G14560*	*IAA1/AXR5*
25	*AT4G28640*	*IAA11*
26	*AT4G29080*	*IAA27/PAP2*
27	*AT4G32280*	*IAA29*
28	*AT5G25890*	*IAA28/IAR2*
29	*AT5G43700*	*IAA4*
30	*AT5G57420*	*IAA33*
31	*AT5G65670*	*IAA9*
32	*AT1G19220*	*ARF19*
33	*AT1G19850*	*ARF5/MONOPTERUS*
34	*AT1G30330*	*ARF6*
35	*AT1G34170*	*ARF13*
36	*AT1G34310*	*ARF12*
37	*AT1G59750*	*ARF1*
38	*AT1G77850*	*ARF17*
39	*AT2G28350*	*ARF10*
40	*AT2G33860*	*ARF3/ETTIN*
41	*AT2G46530*	*ARF11*
42	*AT3G61830*	*ARF18*
43	*AT4G23980*	*ARF9*
44	*AT4G30080*	*ARF16*
45	*AT5G20730*	*ARF7/NPH4*
46	*AT5G37020*	*ARF8*
47	*AT5G60450*	*ARF4*

Table A.2.: *Arabidopsis* **accession numbers.**

Accession	NASC/ABRC stock no.
Bay-0	N57923
Bl-1	N968
Bor-4	N22677
Br-0	N22678
Bur-0	N1028
C24	N906
Col-0	N1092
Cvi-0	N22682
Est-1	N22683
Fei-0	CS28250
Got-7	N22685
L*er*-1	N22686
Lov-5	N22695
Nfa-8	N22687
Rrs-7	N22688
Rrs-10	N22689
Shakdara	N929
Tamm-2	N22691
Ts-1	N22692
Tsu-1	N22693
Van-0	N22694

Table A.3.: qRT-PCR primers.

General validation of array data			
AGI	**gene / description**	**forward primer**	**reverse primer**
AT1G15580	*IAA5*	GCTTCCGCTCTGCAAATTCT	CTTGCACGATCCAAGGAACA
AT1G27730	*STZ*	CGCCGTGACTACTGGAAGTG	AGGGCTCATGACTTCGTCGT
AT1G01060	*LHY*	GAACATCTTGAACCGCGTTG	GCCTGGGAACAACGGTACAT
AT1G68520	zinc finger family protein	CGAGGCCTTTTCTCTGCATT	CCGGGTGCCATCTGAAATAG
AT1G29510	*IAA6*	ATCTCCGACGAGCATCCAGT	ATCACTGCCGGTTGTGAAGA
AT2G14960	*GH3.1*	ATGGCTTCGTTGGGACTTGT	TCGTCGCCAGCTTCTTTACA
AT2G23170	*GH3.3*	CGTGACGCATAAGGAGACCA	ATGGACCGACGTCAGCTTTT
AT2G33310	*IAA13*	TTGCGCCAAATTCTCGTAAG	TCTGCTTCTCATGCTGGTTCA
AT2G34650	*PID*	AAGATTCAACGGCTGCGATT	TTGCCGACTCTTTACGCTGA
AT3G06490	*BOS1/ AtMYB108*	AATGGAGAAGGTCGCTGGAA	CCGCGTCCTCCAGTAGTTCT
AT3G10040	transcription factor	TTGCGCCAAATTCTCGTAAG	TCTGCTTCTCATGCTGGTTCA
AT3G15540	*IAA19*	AGTGAGCATGGATGGTGTGC	CCGGTAGCATCCGATCTTTT
AT3G58190	*ASL16*	ATCATGCTTTGTGCTGCTTG	TTCACACTTTGCAGCCATTC
AT4G12410	auxin response protein	AAGTAAATCACCGGCACCAC	CGGTGGAAGCAAGAAGAATC
AT5G47370	*HAT2*	GCTTCTCTACGCACCGTTTC	AACGTCGAGGAAGAAGCTCA
AT4G14560	*IAA1*	GATTACCCGGAGCACAAGAA	GAGATATGGAGCTCCGTCCA
AT3G23030	*IAA2*	TGGATTACCCGGAAGAACAG	AGGAGCTCCGTCCATACTCA

Cluster 100			
AGI	**gene / description**	**forward primer**	**reverse primer**
AT5G51190	ERF/AP2 transcription factor	GCCTCCACCATCCACCACTGC	GGGTCTCTGATCTCAGCCGCA
AT2G23630	*SKU5*	GACGGCGAATGCAGCAAGGC	TGGAGGGTCCTGGTGTGGGAATG
AT1G52060	unknown protein	ATGGATCCGCGGGGAAAGCC	CAGTAGTCGCCGCCCCAACC
AT1G52070	jacalin lectin family protein	TGGCTTCCATGGATCTGCTGGGA	TCCCAAGGTTCGCCTCCGGT
AT3G19430	LEA protein-related	TCCGCCAGTTTCACCACCGC	CGTCTGATGGCGGGGGAACG
AT5G54370	LEA protein-related	CAGACCCGCTGGTCGTGCC	TGAGGCGGTTTTCGCGGCTT
AT5G60530	LEA protein-related	AGTTGCCGCAGGATGACGCT	TGCACAGTCTGCAAGTCGGCG
AT5G04200	*ATMC9*	GCTGACGTCGGCGTTGGGAA	GCGTCTGCGTTCTGGTCGCT

DR5:GUS			
AGI	**gene**	**forward primer**	**reverse primer**
-	*uidA*	TCAGCAAGCGCACTTACAGG	GAGTTTACGCGTTGCTTCCG

Constitutively expressed control gene			
AGI	**gene**	**forward primer**	**reverse primer**
AT1G13320	*PP2AA3*	AGACAAGGTTCACTCAATCCGTG	CATTCAGGACCAAACTCTTCAGC

B. Supporting Information: Optimized probe masking for comparative transcriptomics of closely related species

B.1. Figures

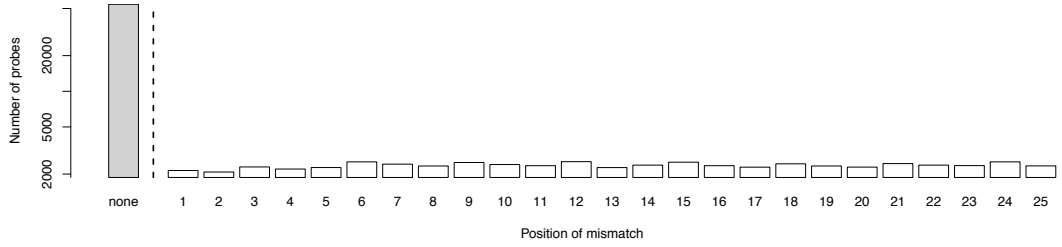


Figure B.1.: Number of probes of the 1mm mask that match the transcripts of *A. lyrata* without any mismatch (none) or with one mismatch at a specific position. The number of probes with a mismatch is similar for all mismatch positions.

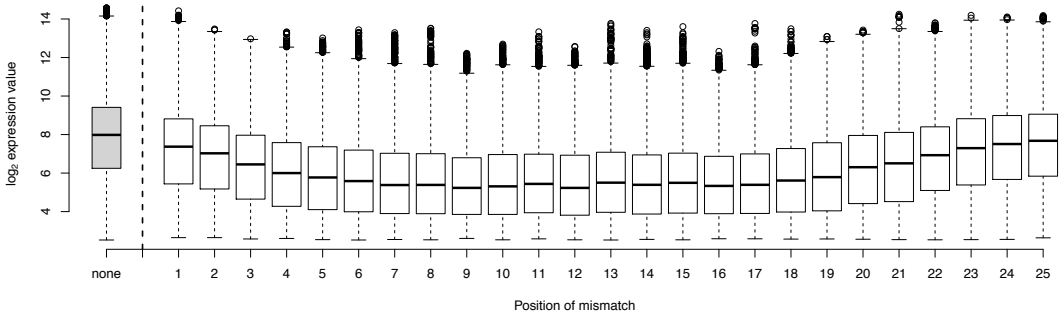


Figure B.2.: $\log_2$ expression values of probes of the 1mm mask that match the transcripts of *A. lyrata* without any mismatch (none) or with one mismatch at a specific position. The expression values measured depend on the occurrence of a mismatch and its position within the probe sequence. Hence, a correction for this positional bias would be required to compare expression values between *A. thaliana* and *A. lyrata*.

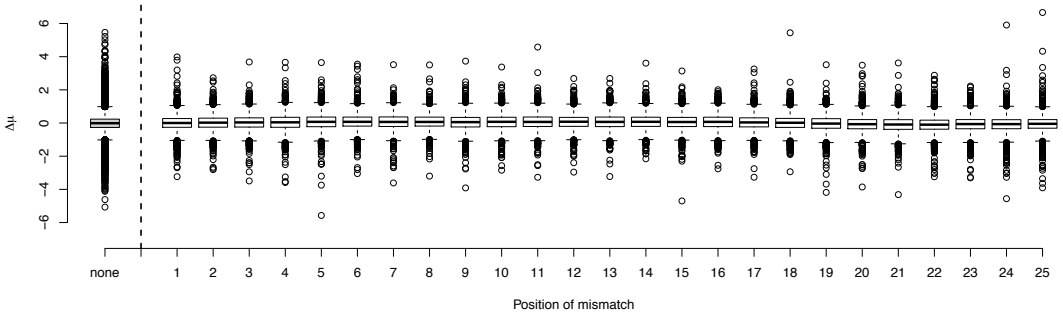


Figure B.3.: $\Delta\mu$ expression responses of probes of the 1mm mask that match the transcripts of *A. lyrata* without any mismatch (none) or with one mismatch at a specific position. The expression responses of probes are similar for all mismatch position as well as for the perfectly matching probes. In contrast to the $\log_2$ expression values, the comparison of $\Delta\mu$ expression responses between *A. thaliana* and *A. lyrata* does not require a correction.

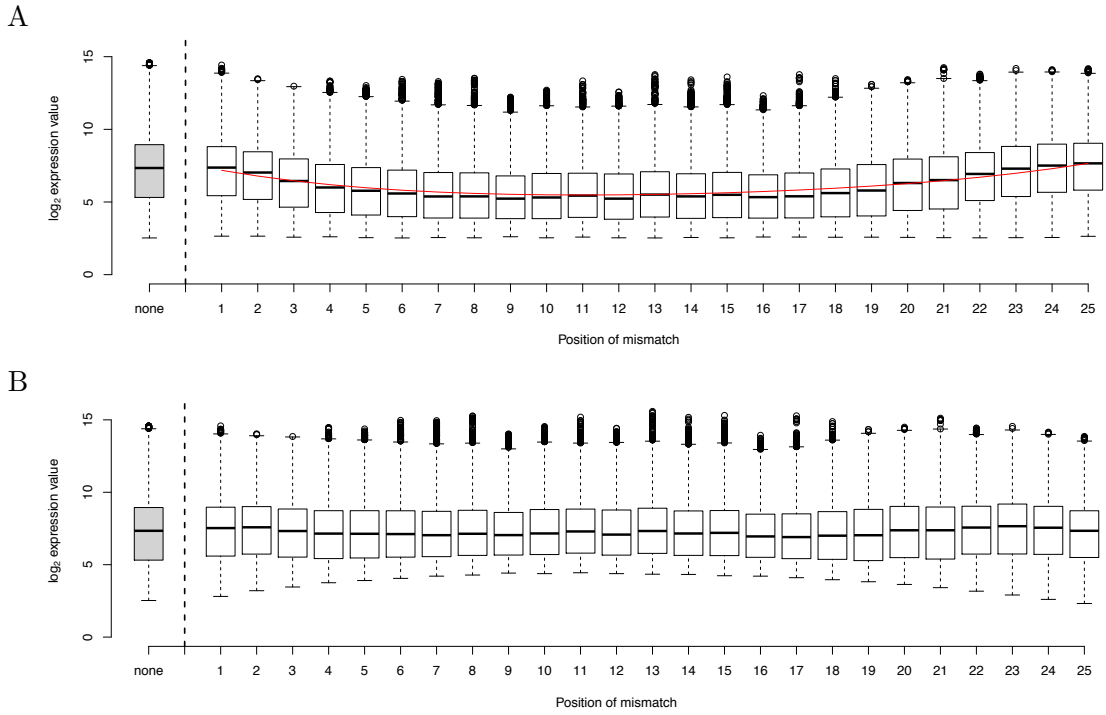


Figure B.4.: $\log_2$ expression values of probes of the 1mm mask. (A) $\log_2$ expression values of probes of the 1mm mask that match the transcripts of *A. thaliana* and *A. lyrata* without any mismatch (none) or with one mismatch at a specific position. The expression values measured depend on the occurrence of a mismatch and its position within the probe sequence. To correct for this positional bias we fit a fourth-degree polynomial to the data (red curve). (B) Corrected $\log_2$ expression values based on the polynomial fit. The corrected expression values are not affected by the occurrence of a mismatch and its position within the probe sequence any more.

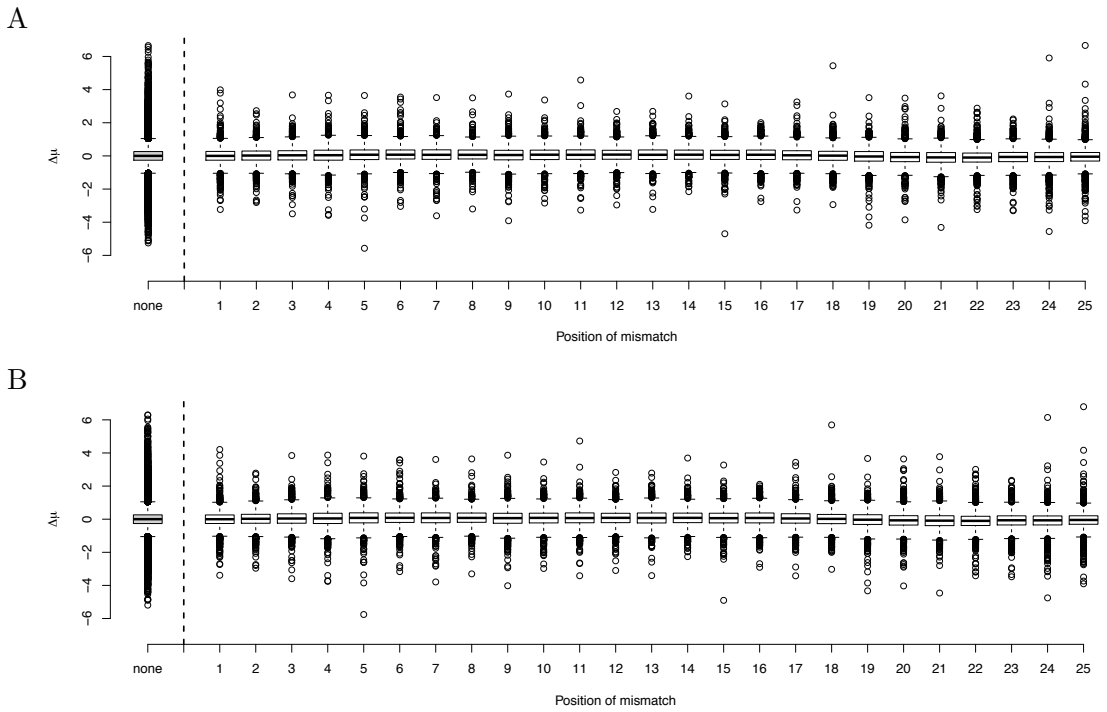


Figure B.5.: $\Delta\mu$ expression responses of probes of the 1mm mask. (A) $\Delta\mu$ expression responses of probes of the 1mm mask that match the transcripts of *A. thaliana* and *A. lyrata* without any mismatch (none) or with one mismatch at a specific position. The expression responses of probes are similar for all mismatch position as well as for the perfectly matching probes. (B) Corrected (Figure B.4) $\Delta\mu$ expression responses of probes of the 1mm mask. The expression responses of probes are similar to the uncorrected expression responses in (A). The suggested correction does not have a significant effect on the expression responses.

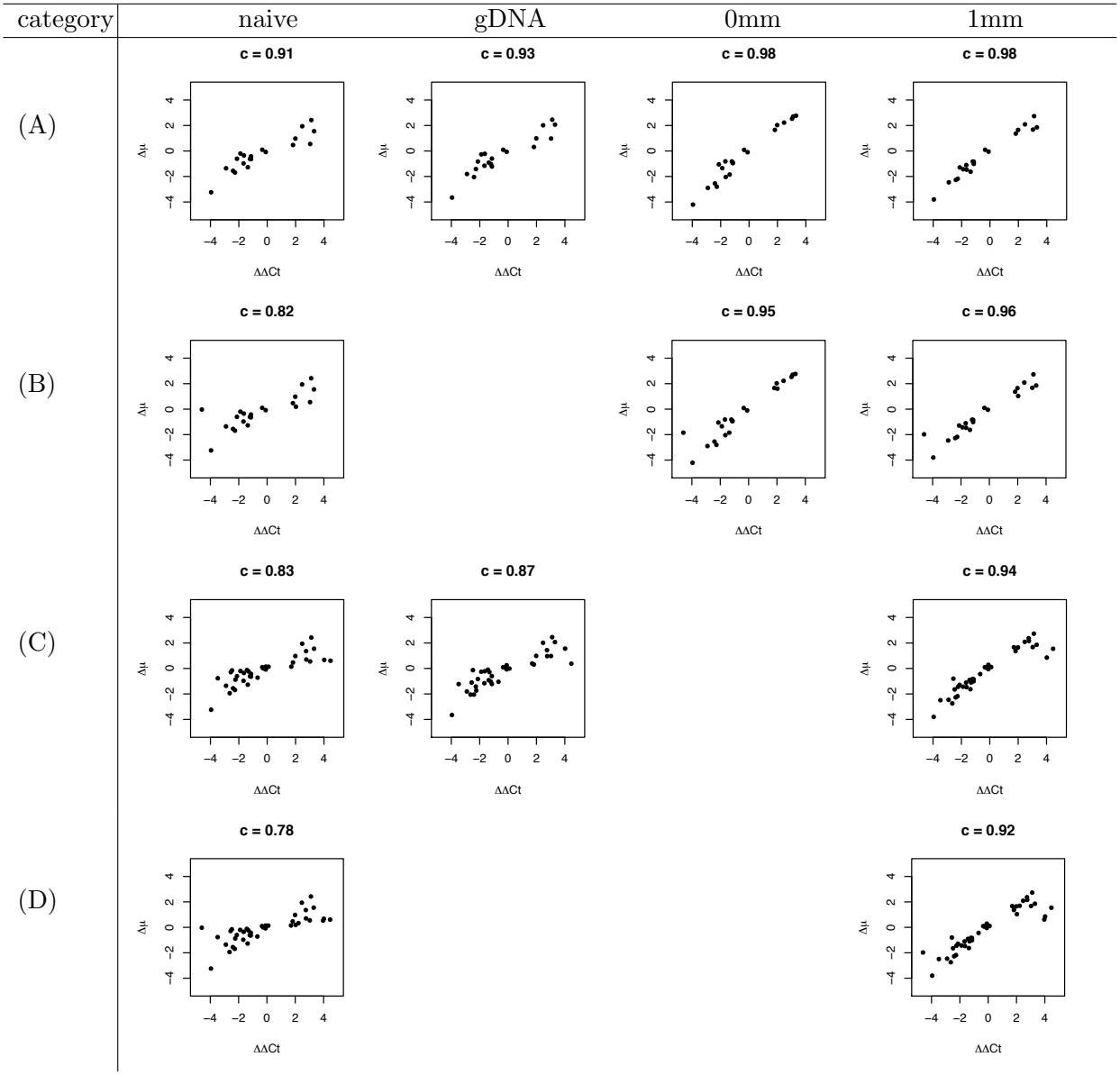


Figure B.6.: Scatterplots and Pearson correlation coefficients. Correlation coefficients of (i) the $\Delta\mu$ expression responses of *A. lyrata* resulting from the three masking approaches and the naive approach, and (ii) the $\Delta\Delta$Ct expression responses resulting from qRT-PCR of the genes of category A, B, C, and D (Methods Candidate selection).

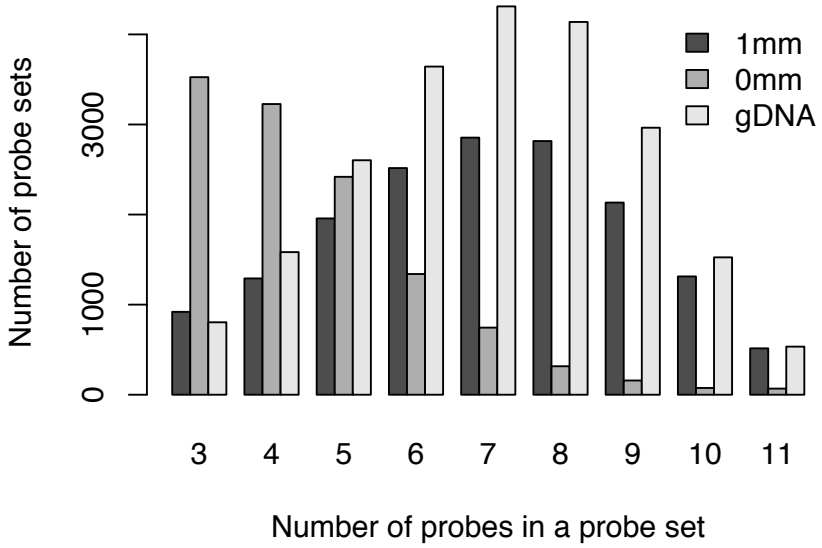


Figure B.7.: Frequency of probes per probe set. The height of the bars represents the absolute frequency of probe sets containing a defined number of probes. For each number of probes three bars are shown, one for each of the three probe masking approaches. Total number of probe sets: 16315 (1mm approach), 11873 (0mm approach), and 22105 (gDNA approach).

B.2. Tables

Table B.1.: Number of probe sets and probes, and average number of probes per probe set for all three masking approaches and the naive approach. Values are rounded to the first or second position after decimal point. Additionally, the number of probe sets falling in one of the following categories are listed: *transcript-specific*: probe sets targeting orthologs, not affected by cross hybridization, and containing at least 3 probes; *transcript-unspecific* probe sets that can be separated in *cross hybridization*: probe sets affected by cross hybridization and *non-ortholog*: probe sets targeting non-orthologs; *no match*: probe sets matching no transcript in *A. thaliana* or *A. lyrata*; and *less than 3 probes*: probe sets containing less than 3 matching probes in the 1mm approach, but at least 3 probes in the other approach. The 1mm approach retains a similar number of transcript-specific probe sets as the gDNA and the naive approach, but retains more transcript-specific probe sets than the 0mm approach.

	naive	**gDNA**	**0mm**	**1mm**
transcript-specific	16315 (71.7 %)	16202 (73.3 %)	10629 (89.5 %)	16315 (100 %)
transcript-unspecific:	1749 (7.7 %)	1701 (7.7 %)	1012 (8.5 %)	0
cross hybridization	1183	1149	749	0
non-ortholog	566	552	263	0
no match	3067 (13.5 %)	2682 (12.1 %)	33* (0.3 %)	0
less than 3 probes	1615 (7.1 %)	1520 (6.9 %)	199 (1.7 %)	0
total number of probe sets	22746	22105	11873	16315
total number of probes	250103	154698	54281	113303
average number of probes per probe set	11.00	7.00	4.57	6.95

*The 0mm approach uses the targets for *A. thaliana* annotated by Affymetrix to determine the target sequences of *A. lyrata*, by aligning the transcript sequences of *A. lyrata* to the annotated target sequences of *A. thaliana*. The sequences of the probes are aligned to the sequences of target transcripts of *A. lyrata*. Only probes are retained that perfectly match the target transcripts of *A. lyrata*.

The 1mm approach aligns the sequences of the probes to the sequences of protein-coding transcripts of *A. thaliana* and *A. lyrata*. A probe set has *no match* if the probes do not match any transcript in *A. thaliana* or *A. lyrata*.

Table B.2.: Pearson, Spearman and Kendall correlation coefficients. Correlation coefficients of (i) the $\Delta\mu$ expression values resulting from the three masking approaches and the naive approach, and (ii) the $\Delta\Delta$Ct expression values resulting from qRT-PCR of the genes of category A, B, C, and D (Methods Candidate selection). The 1mm approach and the 0mm approaches yield similar correlation coefficients that are higher than those of the gDNA and the naive approaches.

category	mask	Pearson	Spearman	Kendall
A	1mm	0.980	0.961	0.853
	gDNA	0.929	0.902	0.758
	0mm	0.978	0.950	0.853
	naive	0.913	0.905	0.747
B	1mm	0.961	0.956	0.827
	0mm	0.952	0.939	0.818
	naive	0.818	0.820	0.662
C	1mm	0.938	0.931	0.797
	gDNA	0.867	0.898	0.717
	naive	0.830	0.877	0.692
D	1mm	0.920	0.920	0.774
	naive	0.777	0.856	0.669

Table B.3.: Primer sequences for *A. lyrata* of the 40 candidate genes used for verification by qRT-PCR. The locus identifier for *A. thaliana* is given by the TAIR id and that for *A. lyrata* by the Phytozome gene id.

ae name	locus At	locus Al	forward primer	reverse primer
245245_at	AT1G44318	314128	AACTGGGCACGGTGGGATCG	CGCGCCTACACGACCATCCA
245336_at	AT4G16515	493225	GCTGCCGCTCGTCGGTTAGG	CGACCCCACCCAACACTCGC
245369_at	AT4G15975	329916	TAGCCGCTTCAACCGCACCG	GGTTGCGGCGGAGAACGTGA
245397_at	AT4G14560	946923	ACCGAGCTTCGTTTGGGATTACCTG	GGAGGCCATCCCACGATTTGTGTT
245696_at	AT5G04190	939816	GCTCGTCCATGGGCTCCACC	CCGGCTCGGCGGTCATAACG
246270_at	AT4G36500	490986	GGTGCTGGTGGTGTTCCGACC	CGGGTGGCTAAATTTGCCTGTTGG
246993_at	AT5G67450	496850	ACGGAAGTAGCAGCAACAGCGT	GGCCACCAATAGCACTTTCTTCCGA
247215_at	AT5G64905	951330	GGCGATTTTTCGTCATCTCACAGCG	GTCTTGGTCTTCCCTCGCGCTT
247524_at	AT5G61440	496303	ACGATGCAGCCTCGGGCCA	TTCCCCAACCGATGCCAAAGCC
248539_at	AT5G50130	495070	GCCAGGGCGCAGCTACAACA	TGGGTGCATAGCTTGAAAGCCACA
248676_at	AT5G48850	494948	ACCCACCAAGACCGCTCGCT	TGTATACCGCCTCTGCCGACAAGT
248858_at	AT5G46630	948276	CGAAGATGCCGGTGGCTGCT	CGACGTCATCACGGTAGGTGCG
250937_at	AT5G03230	939701	CGCACGAGTATTTAGCGCGGC	TTCGCCGTTCCACCGATTCCTTC
251705_at	AT3G56400	486080	GGGTGCAAGGCAACAAAGCAAGT	TGCGTTGGTGTTACACGTGTGGT
251910_at	AT3G53810	485775	GGCCGGGACGGTTTTAGCGG	ATCAAACATCACCTCCCACCGGAGA
252205_at	AT3G50350	485386	CCGGGGGTAGTGCGTCGTCT	GCTCTTTCCACGGCGGCGAT
252626_at	AT3G44940	484892	CGGTCCTCTGCCTCAATGGCG	AGTGGTGGGATGGTGACAGGAGG
253287_at	AT4G34270	491240	GTGAAAACTGTTGGAGAGAAGCAA	TCAACTGGATACCCTTTCGCA
253400_at	AT4G32860	491410	AGCGCCAACGCCAAGCGTAT	ACGTCGGGAAACACTGCCGC
253908_at	AT4G27260	492072	TGGATGGAAACACGCCGATCCC	AAGCGGGCCTATGGACTTGTCACT
253959_at	AT4G26410	945436	ACTGCTGCTGATGCGACGGTG	GAGCAGAGCGCATGGCGGAA
254175_at	AT4G24050	492457	ACCGTCCCTCAAGCAGCAGCA	TTGGATCCGAGTCCAGACGGCG
254761_at	AT4G13195	333009	ACTCACACTAGACGGAGTCGCAGT	TGAGCTTCAGCTCCGAACTCTCTCC
255788_at	AT2G33310	482270	ACCAAGCTACGAAATCTGCGAGGG	ACCCAACTGGCACTTTCCCATTCAC
256131_at	AT1G13600	920239	GGGAATCTGCTCGTAGGTCA	TCAGATACGCGGTTCAGCTT
257153_at	AT3G27220	936451	GGGAGGCTTCATGTGATGGGTGG	GCCCTGTGTGGTCCACCTCG
259407_at	AT1G13320	920212	AGACAAGGTTCACTCAATCCGTG	CATTCAGGACCAAACTCTTCAGC
260904_at	AT1G02450	470205	TAGCGGCGGAGAAAGATCCGGT	GCGGCTTCAAATCCGTACGACACT
261766_at	AT1G15580	471758	GGCCTCTCCGGAAGTGGAGAGTAA	AACCGGTGGCCAACCCACAA
261892_at	AT1G80840	477161	AGGACCAGTCCGTGTTGGTTGC	GCTGCTGCGGGTGTTGAAGC
262085_at	AT1G56060	474673	CGTATGTGACAGCTCCGCCACC	AGCAGCAGCACATTGCAGCCA
263970_at	AT2G42850	346095	AGGAGGGCGCTGAGAAGCCA	TGGCCATGGCGTAAGAGGTTGTG
264867_at	AT1G24150	313260	TCAGAGGGGAAGCGATGTGTGCT	TCGAACTGCTGCTGCTACGGC
265256_at	AT2G28390	481666	AACTCTATGCAGCATTTGATCCACT	TGATTGCATATCTTTATCGCCATC
265452_at	AT2G46510	483808	GGCGGTCCGGGGAGGTGTTA	TCGTTCTGATTCCCGCAGATTTCGC
265806_at	AT2G18010	931672	CCGTTTACGTGGGACCGAACCG	CCTCGGCTAGTCGGAGCAACG
265856_at	AT2G42430	935111	GCTGTCGTCACCATCGCCTACG	GGCCGGCGATCTGTGCCTTC
266649_at	AT2G25810	932757	CGCCATGGCCACCGACAGTT	GTGACCGCGGGGTTGAGGTG
266820_at	AT2G44940	483623	CGGCGAGCTTCCTCGTCCAG	GCTCGACTCGGCTCGGCTTC
266974_at	AT2G39370	482956	CATGCGCGTCTTCCGCTGC	ACGATCTCCTATGGCTCCCGGAAA

ae: array element, At: *Arabidopsis thaliana*, Al: *Arabidopsis lyrata*

Table B.4.: Probe sets of the 40 candidate genes containing the position of the mismatch. A mismatch can occur at position 1 to 25. A "0" indicates that the probe matches perfectly without any mismatch and a "–" indicates that the probe is masked. Originally, each of the 40 probe sets consists of 11 probes.

ae name	probes matching transcripts of At											probes matching transcripts of Al										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
245245_at	0	0	0	–	–	–	–	0	0	–	0	10	9	0	–	–	–	–	24	0	–	0
245696_at	0	–	0	–	0	–	0	0	0	–	–	0	–	0	–	6	–	0	0	3	–	–
246270_at	0	0	0	0	0	–	–	–	0	0	0	25	0	1	14	0	–	–	–	9	10	0
248676_at	0	0	–	0	0	0	0	0	–	0	0	0	0	–	25	4	0	0	17	–	8	0
251705_at	0	0	–	0	0	–	–	0	–	–	0	0	0	–	22	0	–	–	1	–	–	0
252205_at	0	–	–	–	0	–	0	–	0	–	0	0	–	–	–	0	–	0	–	0	–	24
252626_at	–	0	0	0	0	0	–	0	–	–	–	–	13	13	0	23	0	–	0	–	–	–
253287_at	0	0	0	0	0	0	0	0	–	0	0	0	15	10	0	0	0	0	8	–	0	0
253908_at	0	0	0	–	0	–	–	–	0	0	0	8	0	9	–	0	–	–	–	0	6	14
254175_at	–	–	–	0	0	0	0	0	0	–	–	–	–	–	0	0	0	4	0	0	–	–
255788_at	0	0	0	0	0	0	0	–	0	0	–	3	17	0	18	1	17	0	–	0	0	–
256131_at	0	–	–	0	0	0	–	0	0	0	0	24	–	–	22	0	20	–	0	18	0	0
257153_at	0	0	–	–	0	0	0	0	0	–	–	11	15	–	–	0	0	0	12	0	–	–
259407_at	–	0	0	0	0	0	0	–	0	0	0	–	0	0	0	0	0	0	–	0	0	0
260904_at	–	0	–	–	0	0	0	0	0	0	0	–	0	–	–	0	0	21	10	17	5	0
261892_at	–	–	0	0	0	0	–	0	0	0	0	–	–	0	0	0	0	–	15	15	0	0
263970_at	0	0	0	0	0	0	0	–	0	–	0	0	14	0	0	0	0	25	–	3	–	9
264867_at	0	0	0	0	0	0	–	0	–	0	–	0	8	0	0	0	12	–	12	–	0	–
265452_at	0	0	0	–	–	0	0	–	–	–	–	0	0	12	–	–	0	0	–	–	–	–
265856_at	0	0	0	–	–	0	0	0	–	0	0	0	0	0	–	–	0	15	0	–	6	0
245336_at	–	0	0	0	–	–	0	–	–	0	0	–	17	3	20	–	–	23	–	–	1	8
245369_at	–	0	–	–	16	–	–	–	–	0	–	–	0	–	–	24	–	–	–	–	8	–
245397_at	–	0	–	–	–	–	0	0	0	–	–	–	9	–	–	–	–	10	22	0	–	–
246993_at	0	–	–	–	0	0	–	–	–	–	0	0	–	–	–	0	15	–	–	–	–	14
247524_at	–	–	0	0	–	–	0	0	–	–	0	–	–	0	16	–	–	0	23	–	–	19
248858_at	0	–	–	0	0	0	–	–	–	–	–	8	–	–	0	12	0	–	–	–	–	–
250937_at	0	0	–	–	0	0	0	–	–	–	0	7	0	–	–	25	11	0	–	–	–	19
251910_at	–	0	–	–	–	0	–	–	–	0	–	–	5	–	–	–	0	–	–	–	0	–
253400_at	–	–	0	0	–	0	0	0	–	–	0	–	–	21	22	–	0	10	0	–	–	5
253959_at	0	0	0	0	–	–	–	–	0	0	–	13	0	22	9	–	–	–	–	0	17	–
261766_at	–	–	0	–	–	0	–	0	0	0	–	–	–	21	–	–	0	–	15	11	18	–
262085_at	–	–	–	0	–	–	–	–	0	0	0	–	–	–	7	–	–	–	–	2	0	0
265256_at	–	–	–	–	0	–	0	0	0	0	0	–	–	–	–	11	–	6	7	23	0	0
266649_at	0	0	–	0	0	–	–	–	–	0	–	18	16	–	0	24	–	–	–	–	9	–
266820_at	0	–	0	–	0	0	0	–	0	0	–	19	–	0	–	17	23	22	–	17	22	–
266974_at	–	–	0	0	–	0	–	0	0	0	0	–	–	13	13	–	14	–	20	0	7	4
254761_at	–	–	0	–	–	–	–	0	–	–	0	–	–	0	–	–	–	–	10	–	–	0
265806_at	0	0	0	0	–	–	–	0	0	0	–	14	18	1	25	–	–	–	9	0	0	–
247215_at	–	–	0	0	0	–	–	–	–	–	–	–	–	0	0	0	–	–	–	–	–	–
248539_at	–	–	0	0	–	–	0	0	–	–	–	–	–	0	7	–	–	0	0	–	–	–

ae: array element, At: *Arabidopsis thaliana*, Al: *Arabidopsis lyrata*

C. Supporting Information: Variation of IAA-induced transcriptomes pinpoints the AUX/IAA network as a potential source for inter-species divergence in auxin signaling and response

C.1. Figures

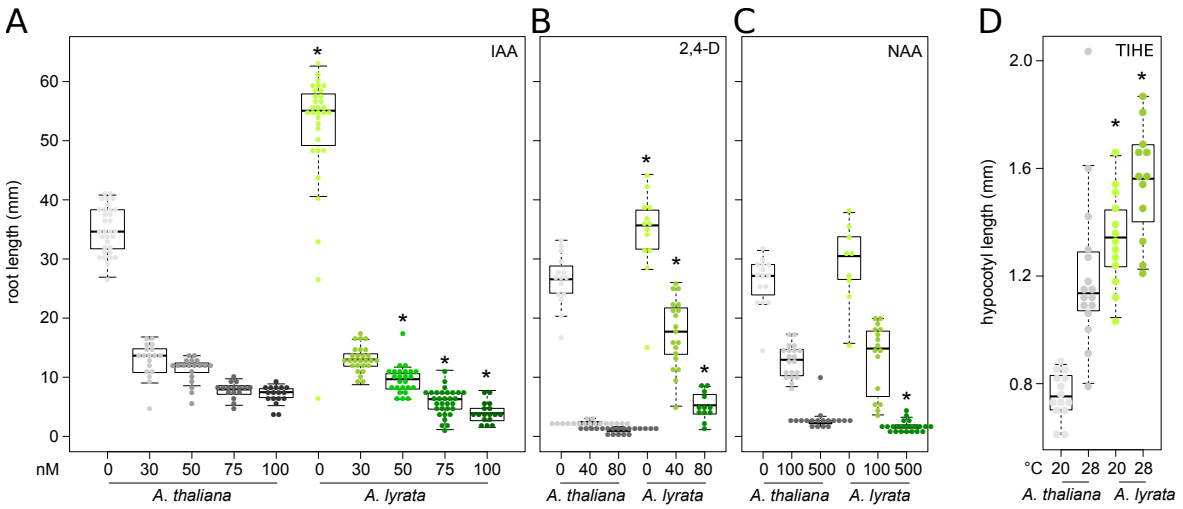


Figure C.1.: Absolute lengths in physiological auxin responses. Absolute root length of seedlings grown on different concentrations of (A) IAA, (B) 2,4-D, or (C) NAA. 3 (A) or 5 (B,C) days-old seedlings were transferred to hormone-containing medium and grown for additional 5 (A) or 3 (B,C) days. (D) Hypocotyl length of 8 days-old seedlings grown either at 20C for 8 days or for 4 days at 20 °C and additional 4 days at 28 °C. (A-D) Box plots show medians (horizontal bar), interquartile ranges (IQR, boxes), and data ranges (whiskers) excluding outliers (defined as $> 1.5 \times$ IQR). Individual data points are superimposed as beeswarm plots. Asterisks denote significant differences between *A. thaliana* and *A. lyrata* values at similar treatments as assessed by 1-factorial ANOVA and Tukey HSD ($p < 0.05$).

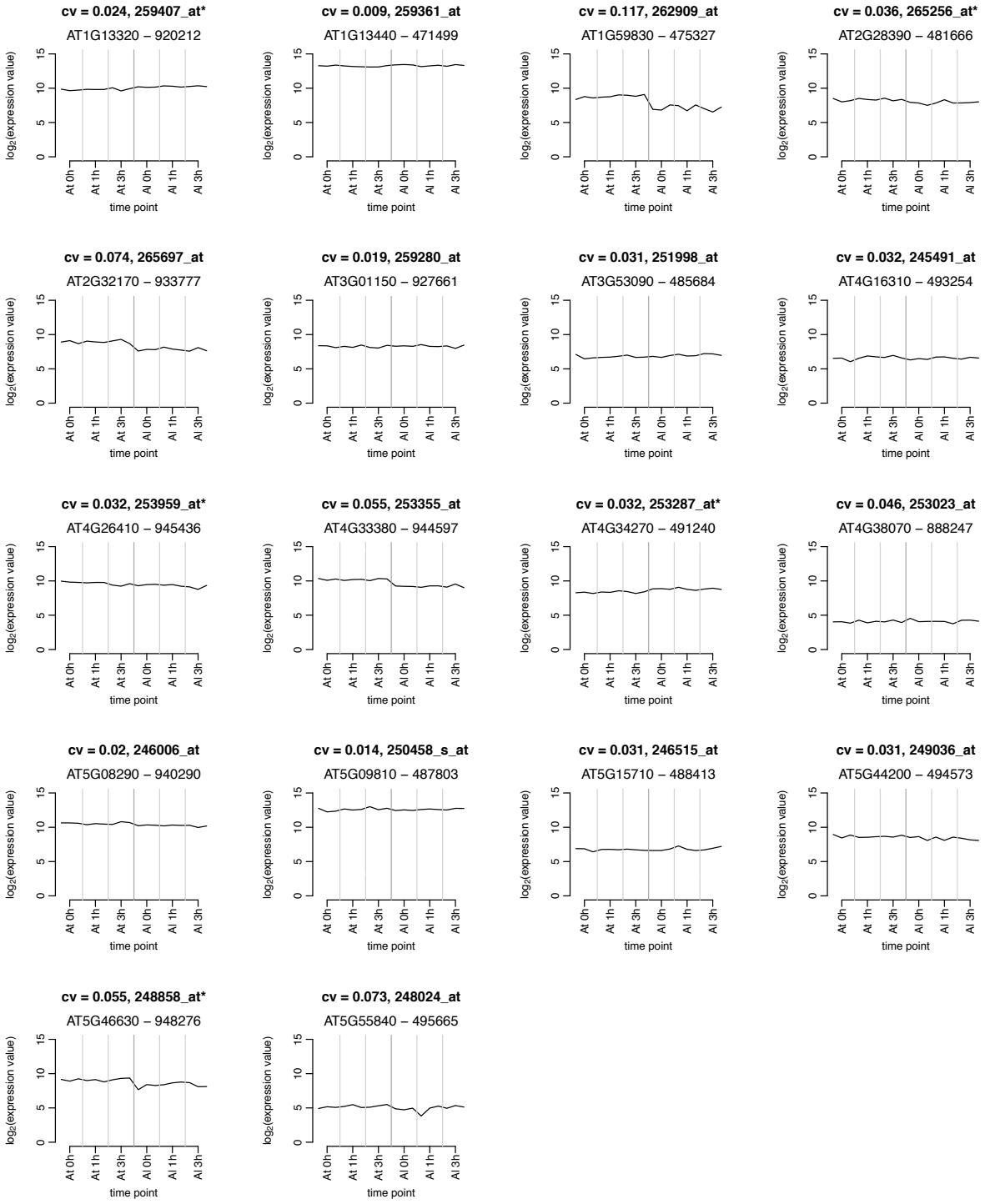


Figure C.2.: Expression levels of non-responsive reference genes confirm successful normalization of the cross-species microarray data. $\log_2$ expression values of a set of reference genes (Czechowski et al., 2005) after probe masking and array normalization including the correction by a fourth-degree polynome. Asterisks mark genes that have been previously verified independently by qRT-PCR in (Poeschl et al., 2013).

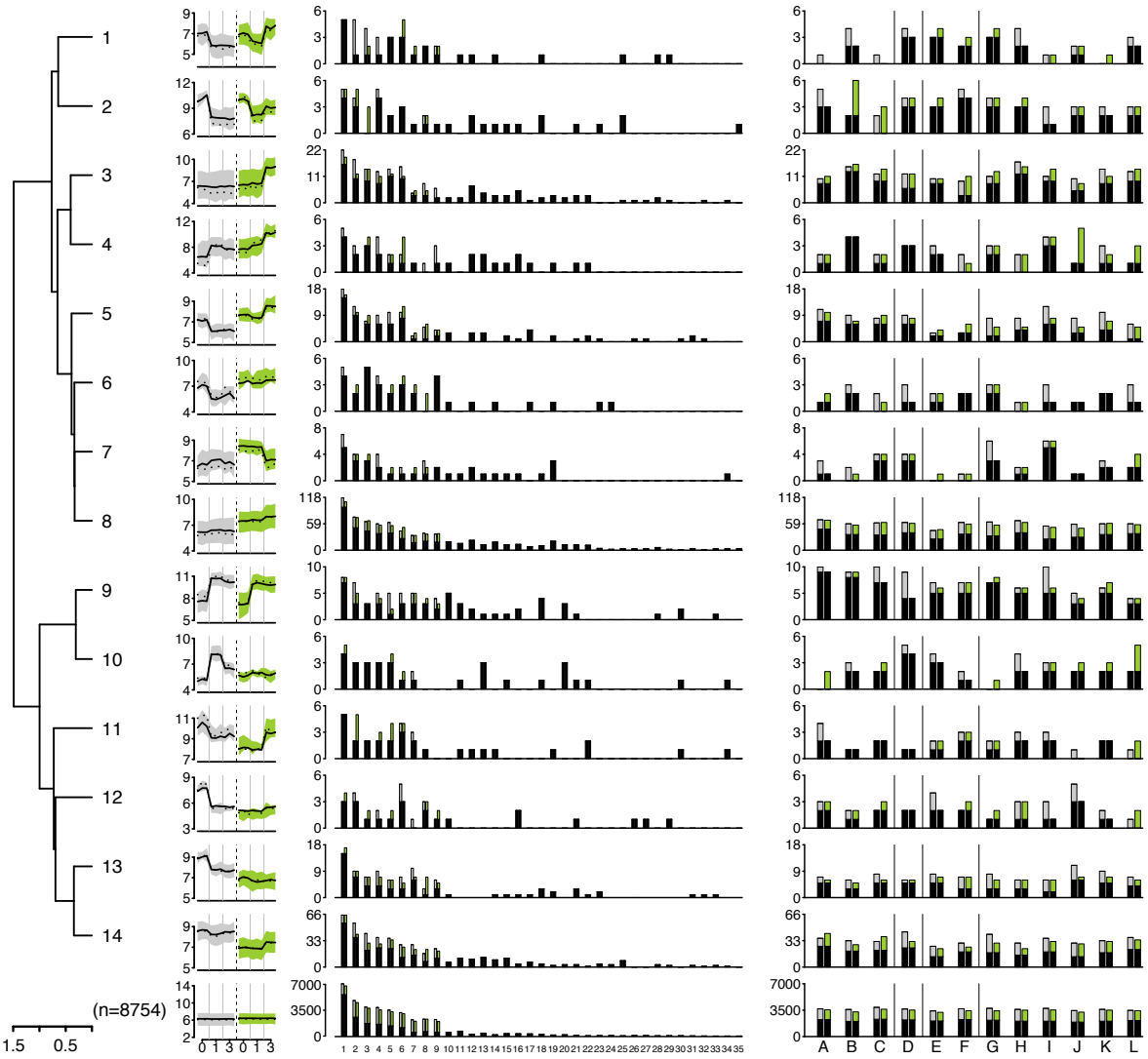


Figure C.3.: Assignment of 35 known and 8 *de novo*-identified *cis*-elements to auxin-regulated gene clusters. Bar plots show the occurrence of known (1-35, Table C.1) or *de novo*-identified (Fig. 6.4 and C.4) *cis*-elements in promoters of auxin-regulated gene clusters. Hierarchical clustering and expression profiles of gene clusters are the same as in Fig. 6.3.

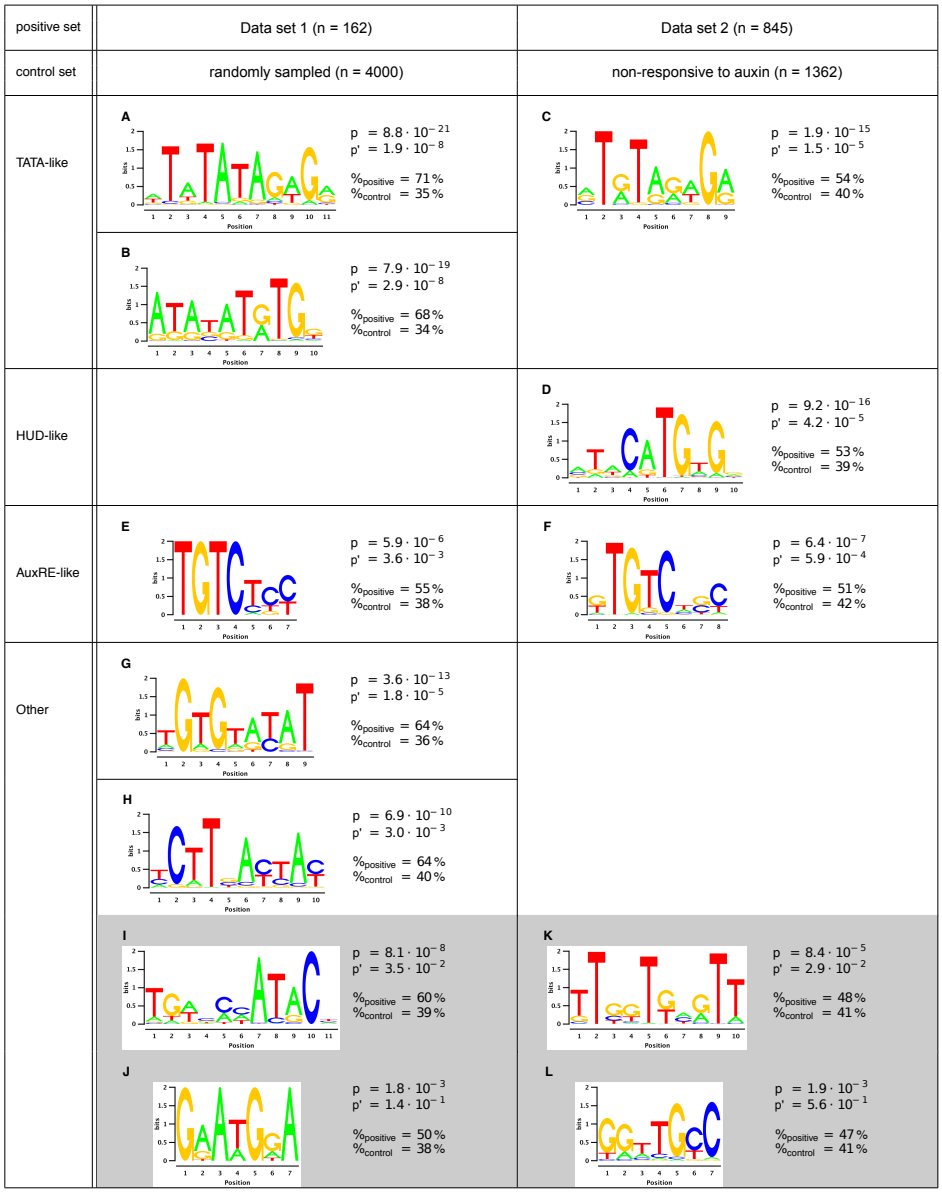


Figure C.4.: *De novo*-identified *cis*-elements. *De novo* identification of putative *cis*-regulatory elements significantly overrepresented in auxin-induced genes using *Dimont*. Data set1, includes promoters of genes that were up-regulated in both species after 1 h and/or 3 h of auxin treatment (n = 162) which were compared to a set of randomly sampled, non-responsive promoter sequences (n = 4000). Data set 2 comprises promoter sequences of genes up-regulated in at last one species. Importantly, data set 2 included only the promoter sequences of a species if the corresponding gene showed an actual response to the auxin stimulus (n = 845), whereas the corresponding promoter sequence of the other species was added to the control set (see text and Supplemental Methods for detailed descriptions). Motifs with significant over-representation in either data set (p) were tested for over-representation in an independent auxin-induced expression data set of *A. thaliana* ($p^{'}$). Motifs that were not significantly enriched in the independent data set are shaded in gray. Frequency of occurrence in the positive and control data sets are denoted by $\%_{\text{positive}}$ and $\%_{\text{control}}$, respectively.

C.2. Tables

Table C.1.: Expression response of conserved auxin up-regulated genes in *A. thaliana* and *A. lyrata*

gene ID		array element		log_2 fold change 1h vs. 0h		log_2 fold change 3h vs. 0h	
A. thaliana	*A. lyrata*			*A. thaliana*	*A. lyrata*	*A. thaliana*	*A. lyrata*
AUX/IAA							
AT1G04240	919100	263656_at	IAA3	1.799	1.818	0.742	0.653
AT1G15580	471758	261766_at	IAA5	5.472	2.225	3.158	0.197
AT2G33310	482270	255788_at	IAA13	1.422	1.892	1.073	1.593
AT3G15540	479003	258399_at	IAA19	3.563	2.804	2.380	2.005
AT3G23030	930309	257766_at	IAA2	1.991	2.183	1.469	1.649
AT3G62100	486723	251246_at	IAA30	1.627	1.656	0.357	1.061
AT4G14560	946923	245397_at	IAA1	3.331	2.357	2.671	2.275
AT4G28640	491899	253791_at	IAA11	1.999	1.825	1.230	1.294
AT4G32280	328503	253423_at	IAA29	3.359	2.018	2.403	1.632
AT5G43700	494639	249109_at	IAA4	1.356	1.117	0.620	0.631
auxin transport							
AT1G23080	472559	264900_at	PIN7	1.055	1.004	1.263	0.537
AT1G70940	908693	262263_at	PIN3	1.696	1.233	1.170	1.037
AT1G73590	476492	259845_at	PIN1	1.757	1.764	0.569	1.074
AT2G17500	480634	263073_at	PILS5	0.179	-0.230	1.142	2.399
AT2G21050	932026	264025_at	LAX2	1.589	1.138	0.988	1.625
ASL/LBD							
AT2G42430	935111	265856_at	ASL18/LBD16	2.161	1.682	1.612	1.147
AT2G42440	346041	257386_at	ASL15LBD17	2.565	2.256	1.379	1.997
AT3G58190	486271	251565_at	ASL16/LBD29	4.783	4.221	4.963	3.756
expansins							
AT3G45970	484976	252563_at	EXLA1	1.617	1.118	0.558	0.262
AT4G17030	946531	245463_at	EXLB1	0.050	-0.259	1.003	1.055
AT4G38400	490602	252997_at	EXLA2	2.165	1.173	0.685	0.624
GH3							
AT2G14960	480379	266611_at	GH3.1	2.601	1.547	3.221	0.916
AT2G23170	481188	245076_at	GH3.3	5.122	5.865	4.838	6.794
AT4G27260	492072	253908_at	GH3.5	2.826	2.504	2.986	2.209
AT5G54510	950094	248163_at	GH3.6	1.779	1.533	1.736	1.463
SAUR							
AT2G18010	931672	265806_at	SAUR10	2.482	1.072	1.304	0.650
AT4G34760	491179	253255_at	SAUR50	1.558	1.683	1.188	0.737
AT4G34770	491178	253207_at	SAUR1	1.724	1.477	0.353	0.968
AT4G38850	943848	252970_at	SAUR15	3.491	2.369	1.851	1.413
AT4G38860	943849	252965_at	SAUR16	1.258	2.240	1.115	1.327
others							
AT1G02850	909946	262118_at	BGLU11	1.242	0.516	1.846	1.434
AT1G05560	919257	263184_at	UDP-glucose transferase	1.518	0.507	1.285	2.892
AT1G05680	470572	263231_at	UDP-glucosyltransferase	3.271	1.698	2.467	4.518
AT1G10380	919813	264466_at	Putative membrane lipoprotein	0.250	0.371	1.058	1.209
AT1G14280	920316	261480_at	phytochrome kinase substrate 2	1.132	1.983	0.472	1.557
AT1G17170	471920	262518_at	glutathione transferase	1.291	0.631	1.399	2.019
AT1G17180	911524	262517_at	glutathione transferase	1.943	0.398	1.065	3.657
AT1G21980	472433	255959_at	type I phosphatidylinositol-4-phosphate 5-kinase	1.267	1.052	0.924	0.587
AT1G23340	472593	263042_at	unknown function	1.287	1.005	0.351	0.065
AT1G23730	912233	265170_at	BCA3	-0.728	0.056	1.190	1.005
AT1G29195	921952	260841_at	unknown protein	1.167	1.124	1.763	2.308
AT1G30100	473228	256190_at	9-cis-epoxycarotenoid dioxygenase	1.804	1.043	0.395	0.338
AT1G30760	473307	264527_at	BBE-like enzyme	0.937	0.295	2.284	1.501
AT1G32870	473471	261192_at	NAC13	1.311	0.854	1.226	1.865
AT1G57560	314948	246401_at	member of the R2R3 factor gene family	1.386	1.413	1.278	0.062
AT1G59740	475340	262912_at	major facilitator superfamily protein	1.224	1.182	1.325	-0.489
AT1G64405	474908	259735_at	unknown protein	2.241	2.174	2.785	2.766
AT1G70270	476150	264341_at	unknown protein	1.237	1.103	1.130	1.704
AT2G03760	484285	264042_at	brassinosteroid sulfotransferase	1.280	1.028	0.982	3.201
AT2G26710	932970	267614_at	cytochrome p450	1.956	2.788	0.688	2.381
AT2G29490	481864	266290_at	glutathione transferase	1.266	1.179	0.601	3.258
AT2G39370	482956	266974_at	member of the MAKR gene family	3.059	2.156	3.303	1.112
AT2G41820	483263	260494_at	LRR protein kinase family member	1.607	1.287	0.784	0.936
AT2G47550	483936	245151_at	plant invertase/pectin methylesterase inhibitor superfamily	1.461	0.718	1.184	1.835
AT3G03660	477611	259223_at	WUSCHEL-related homeobox gene family member	1.445	1.447	1.059	0.421
AT3G09270	903793	259040_at	glutathione transferase	0.489	1.128	1.304	2.121
AT3G13380	478719	256981_at	similar to BRI	0.627	0.019	1.392	1.012
AT3G22370	479749	258452_at	AOX1a	1.434	0.312	1.482	1.478
AT3G26760	484451	258253_at	NAD(P)-binding Rossmann-fold superfamily protein	1.841	1.516	1.005	0.752
AT3G26960	896580	257793_at	Pollen Ole e 1 allergen and extensin family protein	1.848	1.107	1.137	0.061
AT3G28420	484617	257900_at	putative membrane lipoprotein	3.003	1.635	0.916	0.727
AT3G28740	484649	256589_at	cytochrome p450	2.069	0.357	1.517	3.285
AT3G30180	484717	256598_at	cytochrome p450	-0.244	-0.049	1.070	1.299
AT3G42800	484775	252765_at	unknown protein	1.798	1.898	0.926	0.655
AT3G43270	936938	252740_at	plant invertase/pectin methylesterase inhibitor superfamily	0.361	0.280	1.061	1.268
AT3G44540	484866	252638_at	alcohol-forming fatty acyl-CoA reductase	-0.109	0.324	1.391	1.738
AT3G50340	485385	252204_at	unknown protein	1.531	2.052	1.161	1.553
AT3G51410	485493	252103_at	unknown protein	2.248	1.906	2.649	0.632
AT3G54950	485912	251839_at	pPLAIIIbeta	1.138	1.036	0.638	1.097
AT4G08040	327073	255177_at	aminotransferase	1.459	1.894	0.956	-0.028
AT4G15550	493327	245277_at	indole-3-acetate beta-D-glucosyltransferase	1.944	0.737	1.072	1.564
AT4G16515	493225	245336_at	RGF family member	1.592	1.953	1.680	1.579
AT4G17350	890488	245416_at	unknown protein	1.174	1.819	1.793	1.185
AT4G21200	890071	254459_at	gibberellin 2-oxidase	1.929	2.475	0.478	1.063
AT4G30140	491739	253660_at	GDSL lipase/esterase family member	-0.013	-0.031	1.026	1.596
AT4G37295	490890	253047_at	unknown protein	2.202	1.316	2.458	1.699
AT5G02760	939651	251017_at	protein phosphatase 2C family protein	3.072	3.908	2.221	3.411
AT5G04190	939816	245696_at	phytochrome kinase substrate 4	1.553	2.197	0.327	1.638
AT5G06860	487501	250670_at	polygalacturonase inhibiting protein	1.441	0.748	1.810	2.165
AT5G12050	488048	250327_at	unknown protein	1.849	2.410	1.207	1.837
AT5G18560	488739	249992_at	PUCHI	1.215	1.671	0.982	1.168
AT5G26930	894620	246798_at	GATA factor family member	1.223	1.826	0.768	0.296
AT5G47370	494245	248801_at	homeobox-leucine zipper	3.627	2.848	2.457	2.522
AT5G50130	495070	248539_at	NAD(P)-binding Rossmann-fold superfamily protein	1.790	1.321	0.413	0.190
AT5G51440	495215	248434_at	HSP20-like chaperone	1.915	0.589	1.358	1.714
AT5G52900	495355	248282_at	member of the MAKR gene family	3.030	2.595	2.100	2.372
AT5G53290	949919	248253_at	ERF family member	1.008	1.178	0.250	0.622
AT5G57760	950468	247878_at	unknown protein	1.758	1.673	0.884	1.022
AT5G61820	496351	247488_at	unknown protein	0.787	0.229	1.075	2.146
AT5G62280	332462	247474_at	unknown protein	1.785	1.290	1.259	0.741
AT5G65320	358733	247179_at	bHLH family member	1.188	0.358	1.986	1.180
AT5G66940	951535	247066_at	DOF-domain binding transcription factor	1.149	1.419	-0.075	0.159

Table C.2.: Collection of known *cis*-regulatory elements. Promoter element sequences were taken from `http://arabidopsis.med.ohio-state.edu/AtcisDB/bindingsites.html` (last accessed 2014/02/03). Asterisks denote additional auxin-relevant promoter elements taken from literature.

Name	Consensus motif	Reference
ABFs binding site motif	CACGTGGC	Guiltinan, 1990
ABRE binding site motif	(C/T)ACGTGGC	Choi, 2000
ABRE-like binding site motif	(C/G/T)ACGTG(G/T)(A/C)	Shinozaki, 2000
ACE promoter motif	GACACGTAGA	Hartmann, 1998
AG binding site motif	TT(A/G/T)CC(A/T)(A/T)(A/T)(A/T)(A/T)GG(A/C/T)	Shiraishi, 1993
AG BS in AP3	CCATTTTTAGT	Tilly, 1998
AG BS in SUP	CCATTTTTGG	Riechmann, 1996
AGL1 binding site motif	NTT(A/G/T)CC(A/T)(A/T)(A/T)(A/T)NNGG(A/T)AAN	Huang, 1996
AGL2 binding site motif	NN(A/T)NCCA(A/T)(A/T)(A/T)(A/T)T(A/G)G(A/T)(A/T)AN	Huang, 1996
AGL3 binding site motif	TT(A/T)C(C/T)A(A/T)(A/T)(A/T)(A/T)T(A/G)G(A/T)AA	Huang, 1995
AP1 BS in AP3	CCATTTTTAG	Tilly, 1998
AP1 BS in SUP	CCATTTTTGG	Riechmann, 1996
ATHB1 binding site motif	CAAT(A/T)ATTG	Sessa, 1993
ATHB2 binding site motif	CAAT(C/G)ATTG	Sessa, 1993
ATHB5 binding site motif	CAATNATTG	Johannesson, 2001
ATHB6 binding site motif	CAATTATTA	Himmelbach, 2002
AtMYB2 BS in RD22	CTAACCA	Shinozaki, 1997
AtMYC2 BS in RD22	CACATG	Abe, 1997
Box II promoter motif	GGTTAA	Le Gourrierec, 1999
CArG promoter motif	CC(A/T)(A/T)(A/T)(A/T)(A/T)(A/T)GG	Hepworth, 2002
CArG1 motif in AP3	GTTTACATAAATGGAAAA	Tilly, 1998
CArG2 motif in AP3	CTTACCTTTCATGGATTA	Tilly, 1998
CArG3 motif in AP3	CTTTCCATTTTTAGTAAC	Tilly, 1998
CBF1 BS in cor15a	TGGCCGAC	Hao, 2002
CBF2 binding site motif	CCACGTGG	Pla, 1993

Name	Consensus motif	Reference
CCA1 binding site motif	AA(A/C)AATCT	Wang, 1997
CCA1 motif1 BS in CAB1	AAACAATCTA	Wang, 1997
CCA1 motif2 BS in CAB1	AAAAAAAATCTATGA	Wang, 1997
DPBF1&2 binding site motif	ACACNNG	Kim, 1997
DRE promoter motif	TACCGACAT	Kasuga, 1999
DREB1&2 BS in rd29a	TACCGACAT	Kasuga, 1999
DRE-like promoter motif	(A/G/T)(A/G)CCGACN(A/T)	Chen, 2002
E2F binding site motif	TTTCCCGC	Chaboute, 2000
E2F/DP BS in AtCDC6	TTTCCCGC	de Jager, 2001
E2F-varient binding site motif	TCTCCCGCC	Ramirez-Parra, 2003
EIL1 BS in ERF1	TTCAAGGGGGCATGTATCTTGAA	Solano, 1998
EIL2 BS in ERF1	TTCAAGGGGGCATGTATCTTGAA	Solano, 1998
EIL3 BS in ERF1	TTCAAGGGGGCATGTATCTTGAA	Solano, 1998
EIN3 BS in ERF1	GGATTCAAGGGGGCATGTATCTTGAATCC	Solano, 1998
ERE promoter motif	TAAGAGCCGCC	Shinshi, 1995
ERF1 BS in AtCHI-B	GCCGCC	Solano, 1998
EveningElement promoter motif	AAAATATCT	Harmer, 2000
GATA promoter motif	(A/T)GATA(G/A)	Teakle, 2002
GBF1/2/3 BS in ADH1	CCACGTGG	de Vetten, 1995
G-box promoter motif	CACGTG	Menkens, 1994
GCC-box promoter motif	GCCGCC	Shinozaki, 2000
GT promoter motif	TGTGTGGTTAATATG	Green, 1987
Hexamer promoter motif	CCGTCG	Chaubet, 1996
HSEs binding site motif	AGAANNTTCT	Nover, 2001
Ibox promoter motif	GATAAG	Giuliano, 1988
JASE1 motif in OPR1	CGTCAATGAA	He, 2001
JASE2 motif in OPR2	CATACGTCGTCAA	He, 2001
L1-box promoter motif	TAAATG(C/T)A	Abe, 2001
LS5 promoter motif	ACGTCATAGA	Despres, 2000
LS7 promoter motif	TCTACGTCAC	Despres, 2000

Name	Consensus motif	Reference
LTRE promoter motif	ACCGACA	Nordin, 1993
MRE motif in CHS	TCTAACCTACCA	Hartmann, 1998
MYB binding site promoter	(A/C)ACC(A/T)A(A/C)C	Sablowski, 1994
MYB1 binding site motif	(A/C)TCC(A/T)ACC	Menkens, 1994
MYB2 binding site motif	TAACT(G/C)GTT	Martin, 1997
MYB3 binding site motif	TAACTAAC	Chen, 2002
MYB4 binding site motif	A(A/C)C(A/T)A(A/C)C	Chen, 2002
Nonamer promoter motif	AGATCGACG	Chaubet, 1996
OBF4,5 BS in GST6	ATCTTATGTCATTGATGACGACCTCC	Chen, 1996
OBP-1,4,5 BS in GST6	TACACTTTTGG	Chen, 1996
OCS promoter motif	TGACG(C/T)AAG(C/G)(A/G)(A/C)T(G/T)ACG(C/T)(A/C)(A/C)	Bouchez, 1989
octamer promoter motif	CGCGGATC	Chaubet, 1986
PI promoter motif	GTGATCAC	Chan, 2001
PII promoter motif	TTGGTTTTGATCAAAACCAA	Chan, 2001
PRHA BS in PAL1	TAATTGACTCAATTA	Plesch, 1997
RAV1-A binding site motif	CAACA	Kagaya, 1999
RAV1-B binding site motif	CACCTG	Kagaya, 1999
RY-repeat promoter motif	CATGCATG	Dickinson, 1988
SBP-box promoter motif	TNCGTACAA	Cardon, 1999
T-box promoter motif	ACTTTG	Chan, 2001
TEF-box promoter motif	AGGGGCATAATGGTAA	Tremousayque, 1999
TELO-box promoter motif	AAACCCTAA	Tremousayque, 1999
W-box promoter motif	TTGAC	Yu, 2001
AG BS in SPL/NOZ	AAAACAGAATAGGAAA	Ito, 2004
Bellringer/replumless/pennywise BS1 IN AG	AAATTAAA	Bao, 2004
Bellringer/replumless/pennywise BS2 IN AG	AAATTAGT	Bao, 2004
Bellringer/replumless/pennywise BS3 IN AG	ACTAATTT	Bao, 2004
AGL15 BS in AtGA2ox6	CCAATTTAATGG	Wang, 2004
ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	ACTCAT	Satoh, 2004
LFY BS in AP3	CTTAAACCCTAGGGGTAAT	Lamb, 2002

Name	Consensus motif	Reference
SORLREP1	TT[AT]TACTAGT	Hudson, 2003
SORLREP2	ATAAAACGT	Hudson, 2003
SORLREP3	TGTATATAT	Hudson, 2003
SORLREP4	CTCCTAATT	Hudson, 2003
SORLREP5	TTGCATGACT	Hudson, 2003
SORLIP1	AGCCAC	Hudson, 2003
SORLIP2	GGGCC	Hudson, 2003
SORLIP3	CTCAAGTGA	Hudson, 2003
SORLIP4	GTATGATGG	Hudson, 2003
SORLIP5	GAGTGAG	Hudson, 2003
LFY consensus	CCANTG	
ARF binding site motif	TGTCTC	Ulmasov, 1995
TGA1 binding site motif	TGACGTGG	Schindler, 1992
Z-box promoter motif	ATACGTGT	Ha, 1988
*AATAAG	AATAAG	Liu, 1994 [5]
*B-box promoter motif	CACCAT	Hagen, 1991 [2]
*E-box / HUD	CACATG	Walcher, 2012 [6]
*BRRE	CGTG[TC]G	Walcher, 2012 [6]
*AuxRE variant	TGTGCTC	Walcher, 2012 [6]
*AuxRE SAUR	TGTCTG	Walcher, 2012 [6]
*AuxRE SAUR II	[TG]GTCCCAT	Li, 1994 [4]
*TGA2 binding site motif	TGACGTAA	Liu, 1994 [5]
*AuxRE Dispom	TGT[CG]T[CG][CGT]C	Keilwagen, 2011 [3]
*Z-element	GCACATACGT	An, 1990 [1]

References

[1] G. An, M. A. Costa, and S. B. Ha. Nopaline synthase promoter is wound inducible and auxin inducible. *The Plant Cell*, 2(3):225–233, 03 1990.

[2] G. Hagen, G. Martin, Y. Li, and T. Guilfoyle. Auxin-induced expression of the soybean gh3 promoter in transgenic tobacco plants. *Plant Mol Biol*, 17(3):567–579, 1991.

[3] J. Keilwagen, J. Grau, I. A. Paponov, S. Posch, M. Strickert, and I. Grosse. De-novo discovery of differentially abundant transcription factor binding sites including their positional preference. *PLoS Comput Biol*, 7(2):e1001070, 02 2011.

[4] Y. Li, Z. B. Liu, X. Shi, G. Hagen, and T. J. Guilfoyle. An auxin-inducible element in soybean saur promoters. *Plant Physiology*, 106(1):37–43, 1994.

[5] Z. B. Liu, T. Ulmasov, X. Shi, G. Hagen, and T. J. Guilfoyle. Soybean gh3 promoter contains multiple auxin-inducible elements. *The Plant Cell*, 6(5):645–57, 1994.

[6] C. L. Walcher and J. L. Nemhauser. Bipartite promoter element required for auxin response. *Plant Physiology*, 158(1):273–282, 2012.

References for elements taken from the Atcis data base are listed at `http://arabidopsis.med.ohio-state.edu/AtcisDB/bindingsites.html`.

Table C.3.: *cis*-regulatory elements identified in significantly up-regulated genes. Subset of *cis*-elements presented in Tab. C.2 that were identified in at least one promoter of clustered genes presented in Fig. 6.3 and supplemental Fig. C.3. Numbers of elements correspond to numbers in the respective figures. Elements with previously demonstrated function in auxin biology are highlighted in gray.

No.	Name of *cis*-element	Consensus motif
1	GATA promoter motif	(A/T)GATA(G/A)
2	MYB4 binding site motif	A(A/C)C(A/T)A(A/C)C
3	DPBF1&2 binding site motif	ACACNNG
4	AATAAG	AATAAG
5	LFY consensus	CCANTG
6	T-box promoter motif	ACTTTG
7	Ibox promoter motif	GATAAG
8	Box II promoter motif	GGTTAA
9	ATB2/AtbZIP53/AtbZIP44/GBF5 BS in ProDH	ACTCAT
10	Bellringer/replumless/pennywise BS1 IN AG	AAATTAAA
11	ARF binding site motif	TGTCTC
12	ABRE-like binding site motif	(C/G/T)ACGTG(G/T)(A/C)
13	E-box (HUD) / AtMYC2 BS in RD22	CACATG
14	BRRE	CGTG[TC]G
15	MYB binding site promoter	(A/C)ACC(A/T)A(A/C)C
16	B-box promoter motif	CACCAT
17	CCA1 binding site motif	AA(A/C)AATCT
18	AuxRE SAUR	TGTCTG
19	G-box promoter motif	CACGTG
20	AuxRE Dispom	TGT[CG]T[CG][CGT]C
21	SORLIP1	AGCCAC
22	DRE-like promoter motif	(A/G/T)(A/G)CCGACN(A/T)
23	Hexamer promoter motif	CCGTCG
24	L1-box promoter motif	TAAATG(C/T)A
25	GCC-box promoter motif / ERF1	GCCGCC
26	Bellringer/replumless/pennywise BS2 IN AG	AAATTAGT
27	Bellringer/replumless/pennywise BS3 IN AG	ACTAATTT
28	RAV1-B binding site motif	CACCTG
29	SORLREP3	TGTATATAT
30	AtMYB2 BS in RD22	CTAACCA
31	EveningElement promoter motif	AAAATATCT
32	CArG promoter motif	CC(A/T)(A/T)(A/T)(A/T)(A/T)(A/T)GG
33	ATHB6 binding site motif	CAATTATTA
34	LTRE promoter motif	ACCGACA
35	Z-box promoter motif	ATACGTGT

C.3. Data file

Profile interaction finder results for positively and negatively correlated genes of AUX/IAA gene cluster. (A file containing the names of the genes comprising the individual clusters will be available online.)

C.4. Methods - Comprehensive description of de novo identification of cis-elements

De novo motif discovery was performed using promoter sequences of auxin responsive genes of *A. thaliana* and *A. lyrata*. Promoter sequences of 500 bp upstream and 100 pb downstream of the transcriptional start site or ATG (whichever came first) were used in all analyses.

C.4.1. Selection of data sets

We selected two different sets of genes to identify promoter sequences with potential regulatory function in auxin response. Data set 1 contained promoters of genes that were up-regulated in both species with a $\log_2$ fold change (lfc) > 1 after 1 h of auxin treatment in either species and after 1 or 3 h in the respective other species. A total of 81 orthologous promoter pairs of both species were included in the analysis. As a control gene set we randomly sampled 2000 promoter sequences of *A. thaliana* and *A. lyrata*, respectively, which did not show an auxin response in either species.

For data set 2 we selected the promoters of 474 *A. thaliana* genes and 371 *A. lyrata* genes that showed an expression response of lfc > 1 and/or genes with expression levels higher than the median expression level of the microarray and an expression response of lfc $> \log_2(1.5)$ after 1 h and/or 3 h of auxin treatment. Consequently, data set 2 includes commonly and species-specifically up-regulated genes alike. In case of species-specific up-regulation, we included the promoter sequence of the responsive species in data set 2 whereas the promoter of the non-responsive species was included into the control data set 2. Thus, 202 and 304 promoter sequences of species-specific non-responsive genes for *A. thaliana* and *A. lyrata*, respectively were included in control data set 2 which also included 856 genes that showed no response after 1 h of auxin treatment.

C.4.2. Motif discovery

Dimont, a discriminative motif discovery tool especially suited for large data sets was used for *de novo* motif discovery as described previously (Grau et al., 2013) with the following minor adaptations. First, the option for weighted input of data was omitted and promoters from data sets and control sets were hard-labeled with 1 and 0, respectively. In addition, no specific assumptions about motif localization were made but a uniform distribution across all promoter positions was used. Finally, we did not use the speed-up strategy based on subsets

of data but rather used the complete data sets for all optimization runs. For comparisons of data set 1 and 2 with their respective control sets, the modified Dimont was started with 40 initial seeds and the remaining parameters set to default values.

C.4.3. Prediction, assessment and validation

We assessed the significance of the discovered motifs by predicting motif occurrences in the corresponding data and control sets. The prediction threshold for each motif was determined independently so that 0.001 of all positions of all promoters in the control set were predicted as motif occurrences. For each motif, we independently determined the proportion of promoter sequences that contained the motif at least once for the data set ($\%_{\text{positive}}$) and control set ($\%_{\text{control}}$). A one-sided Fischer test was used to assess whether motif occurrence was significantly enriched in the data set in comparison to its respective control set (see p-values in Fig. 6.4 and C.4). To assess a more general validity and relevance of the identified motifs, we analyzed their occurrence in auxin-induced genes of an independent, previously published data set (Keilwagen et al., 2011) using a one-sided Fischer test (see $p^{'}$ values in Fig. C.4).

C.5. References

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W.-R. (2005). Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in *Arabidopsis*. *Plant Physiology*, 139 (1), pp. 5–17.

Grau, J., Posch, S., Grosse, I., and Keilwagen, J. (2013). A general approach for discriminative de novo motif discovery from high-throughput data. *Nucleic Acids Research*, 41 (21), e197.

Keilwagen, J., Grau, J., Paponov, I. A., Posch, S., Strickert, M., and Grosse, I. (2011). De-Novo Discovery of Differentially Abundant Transcription Factor Binding Sites Including Their Positional Preference. *PLoS Computational Biology*, 7 (2), e1001070.

Poeschl, Y., Delker, C., Trenner, J., Ullrich, K. K., Quint, M., and Grosse, I. (2013). Optimized Probe Masking for Comparative Transcriptomics of Closely Related Species. *PLoS ONE*, 8 (11), e78497.

D. Supporting Information: Developmental plasticity of Arabidopsis thaliana accessions across an ambient temperature range

D.1. Figures

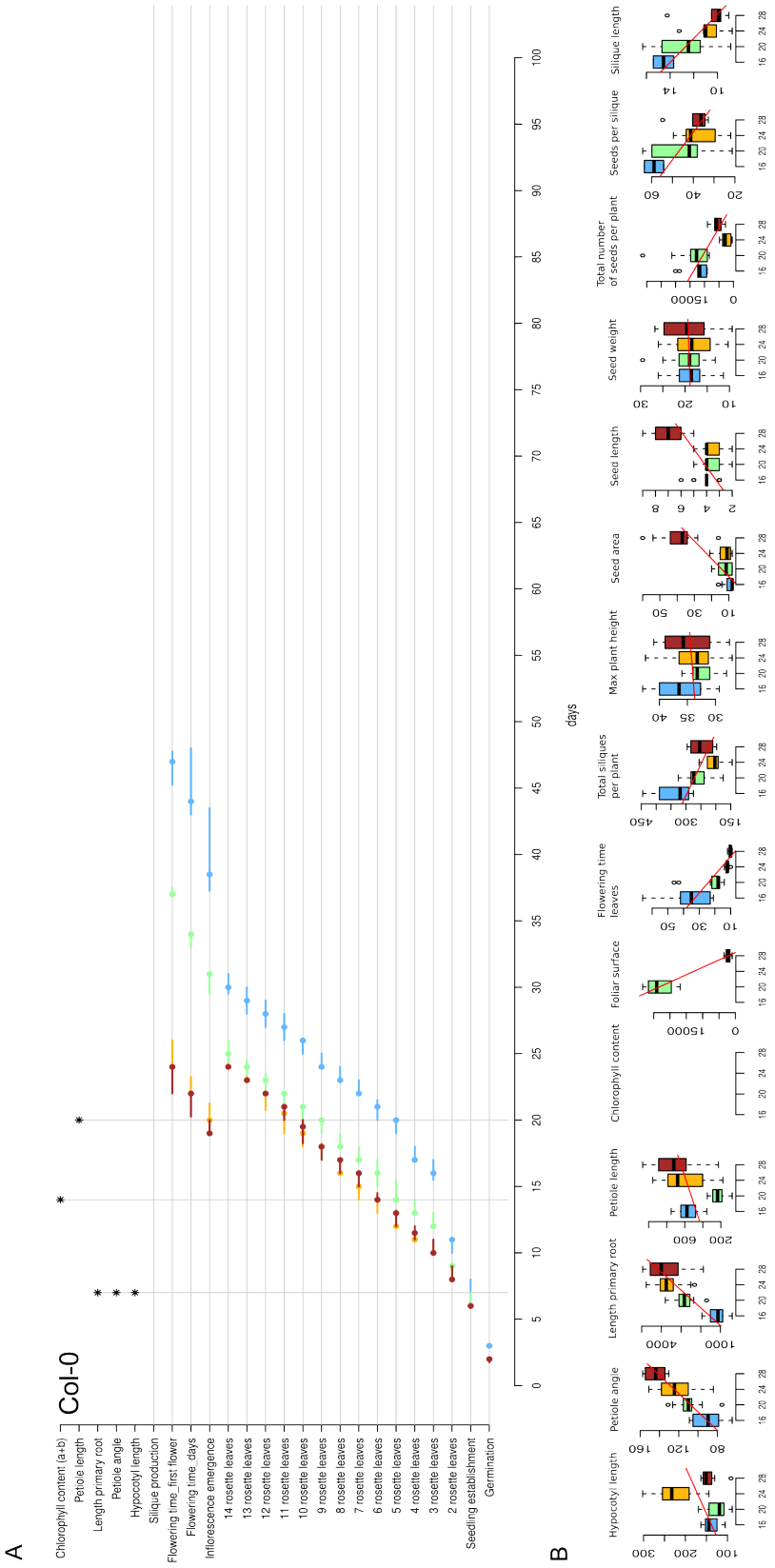


Figure D.1.: Summary of Col-0 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of Col-0 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Data presented in Fig. 7.1 is shown again to allow comparison among accession maps. Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

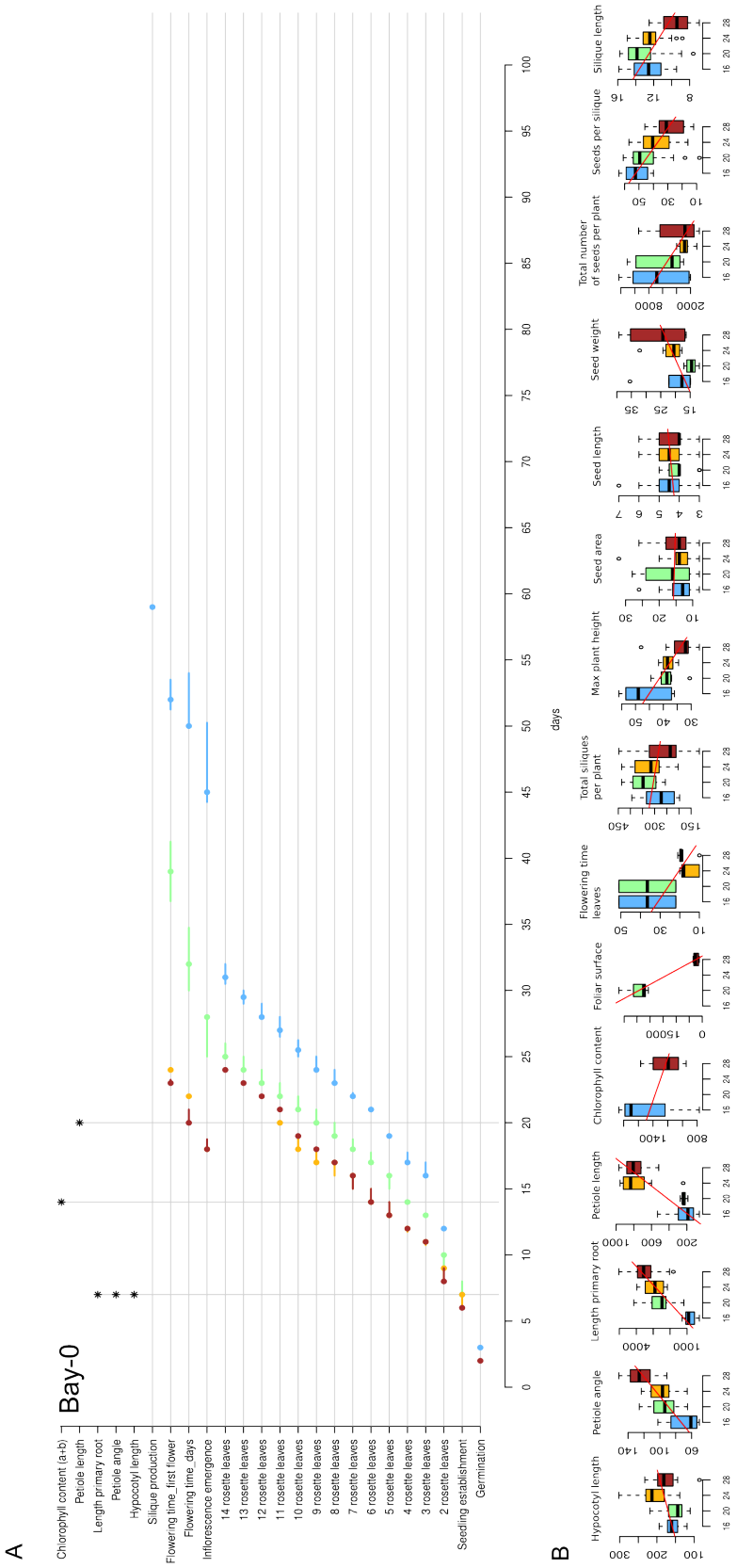


Figure D.2.: Summary of Bay-0 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of Bay-0 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

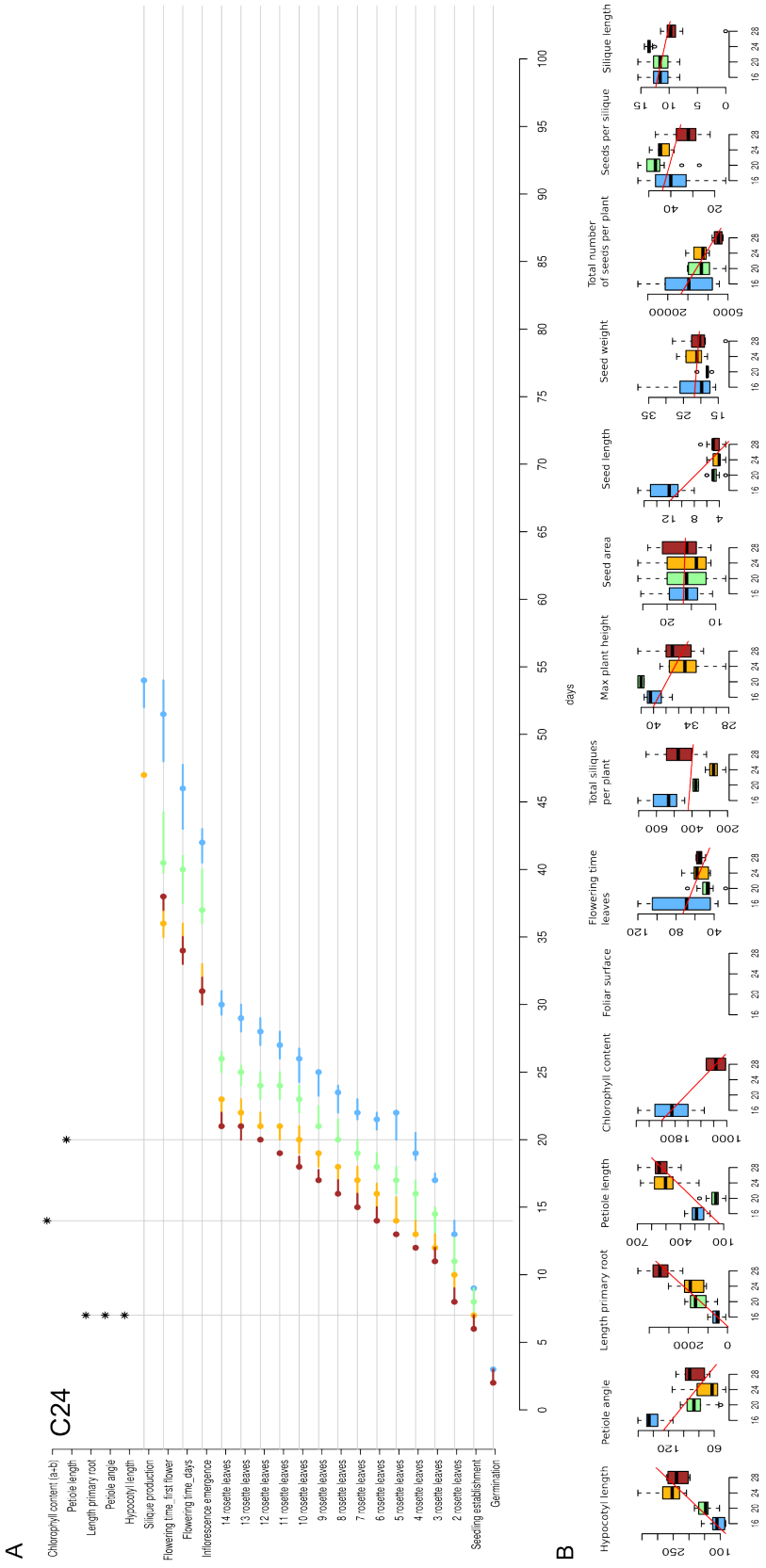


Figure D.3.: Summary of C24 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of C24 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

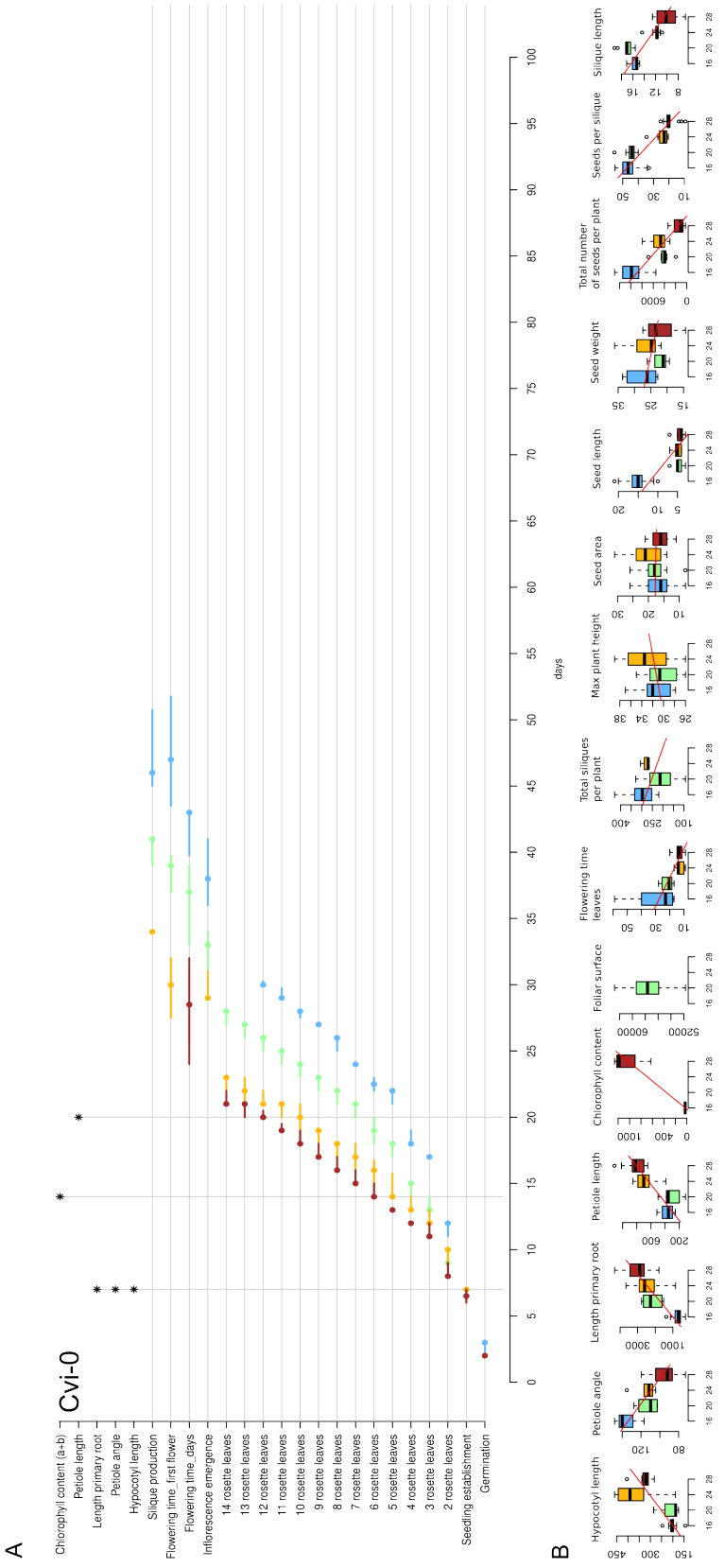


Figure D.4.: Summary of Cvi-0 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of Cvi-0 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

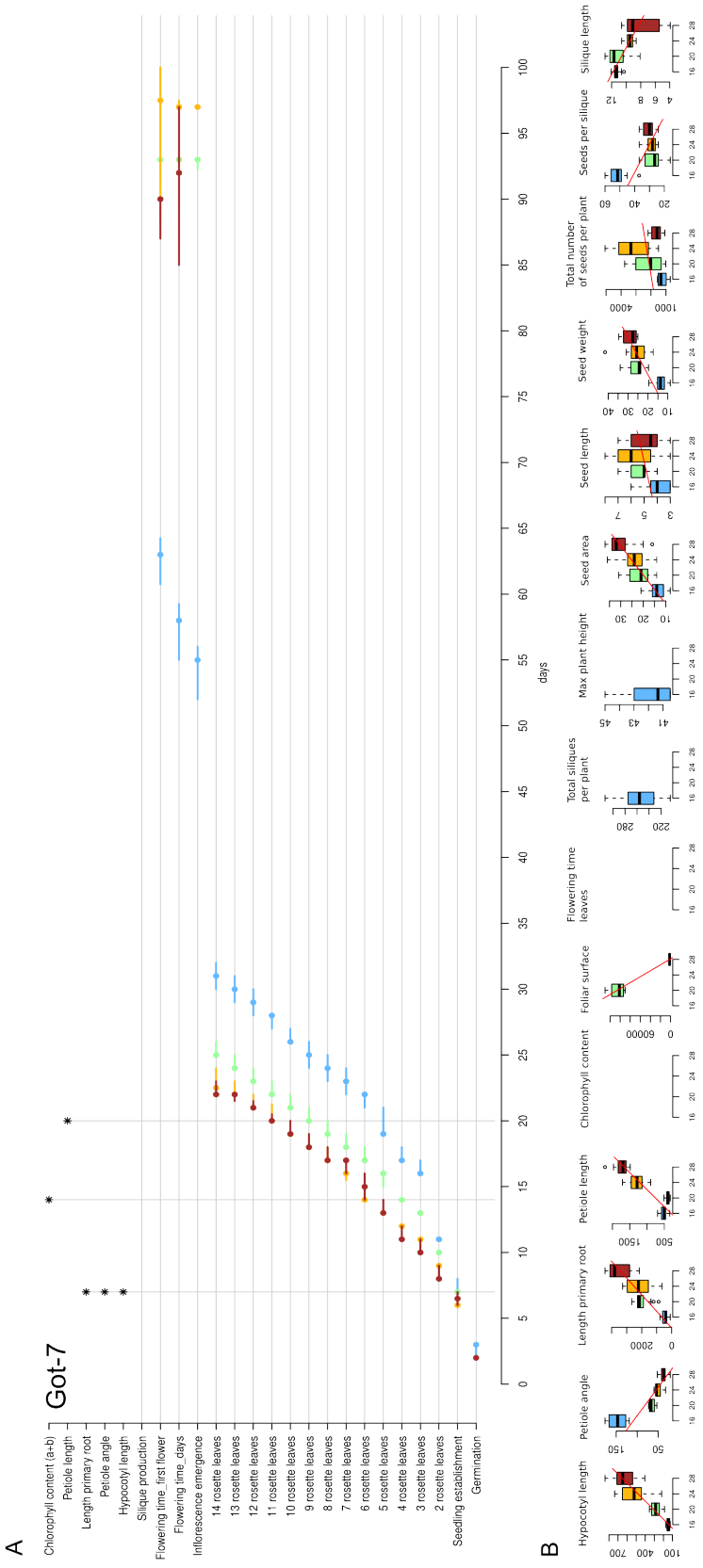


Figure D.5.: Summary of Got-7 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of Got-7 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

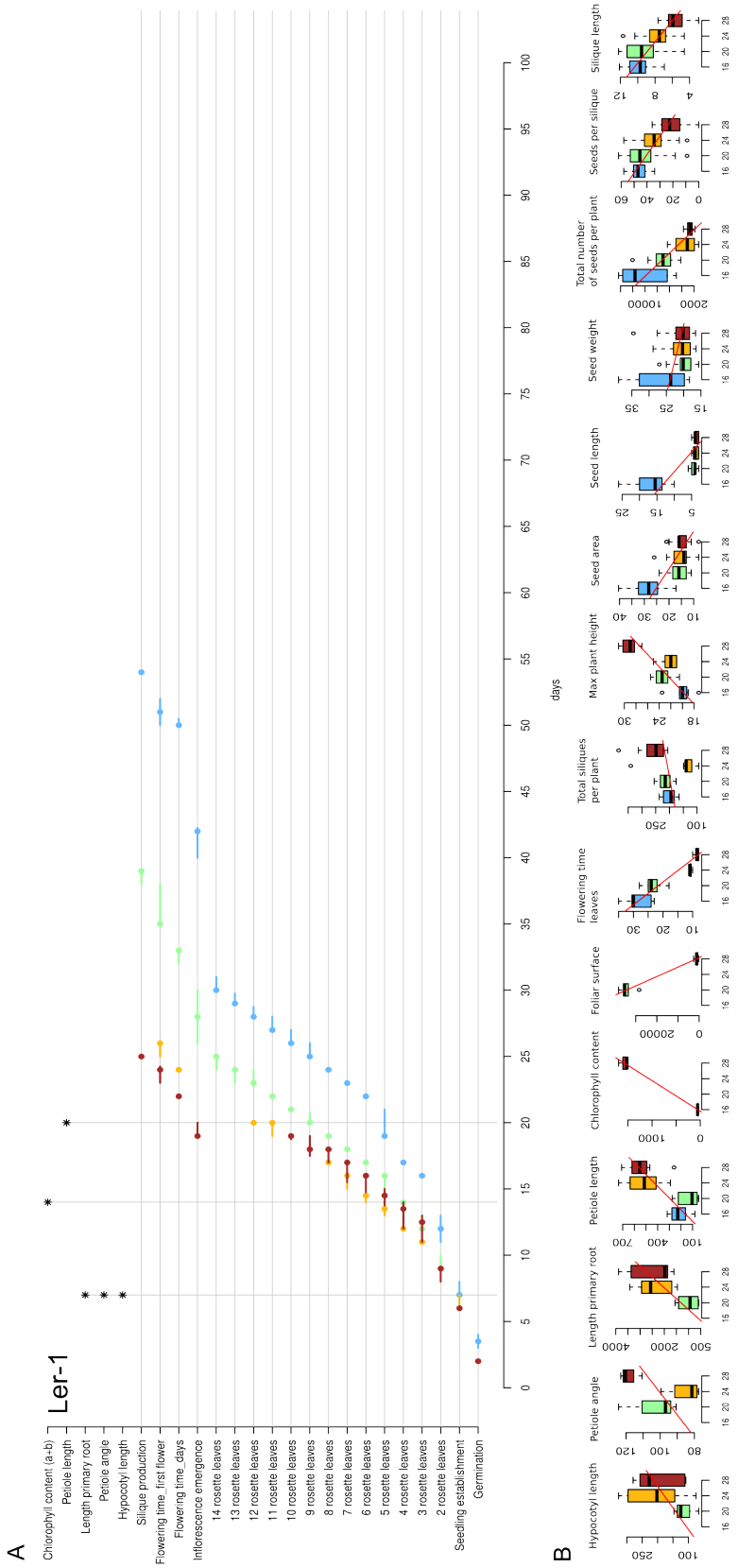


Figure D.6.: Summary of Ler-1 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of Ler-1 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

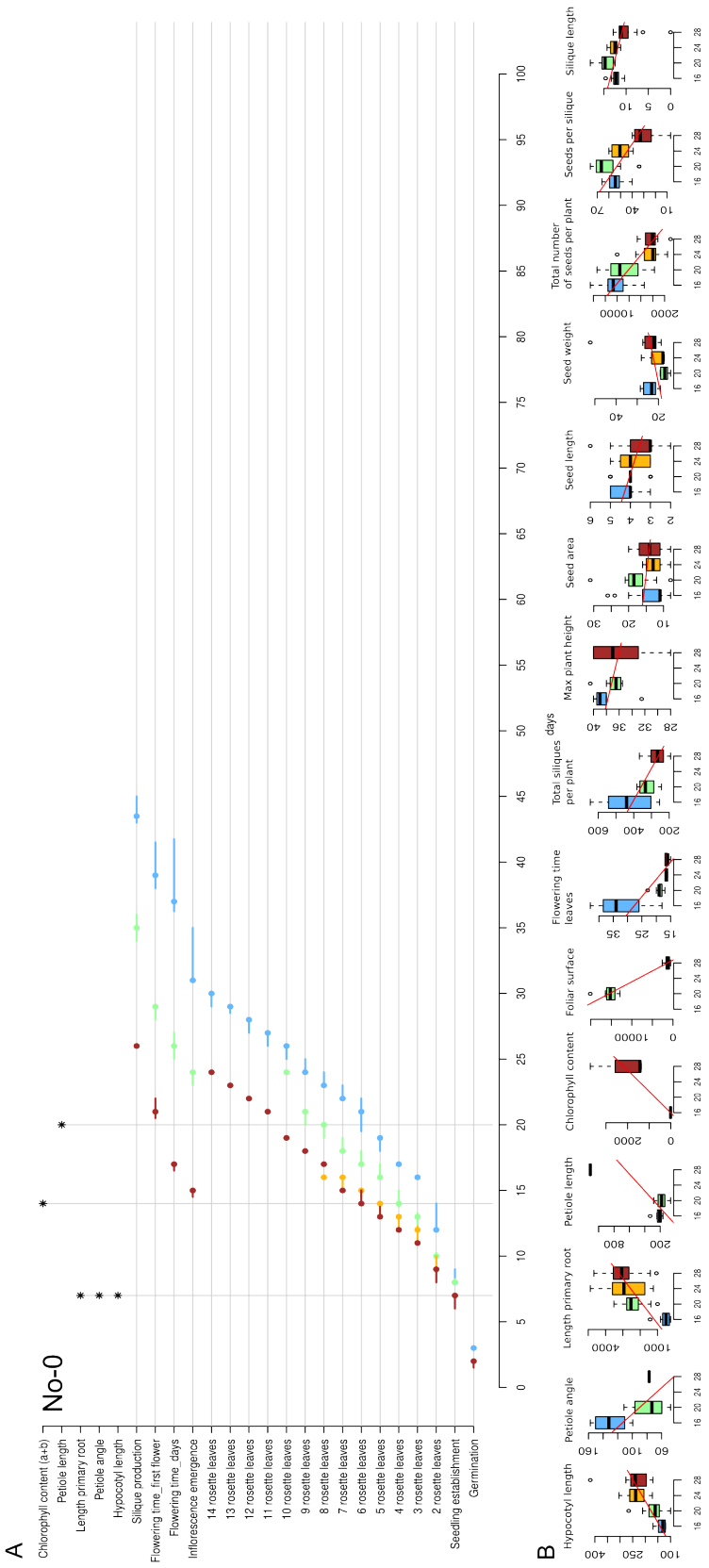


Figure D.7.: Summary of No-0 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of No-0 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

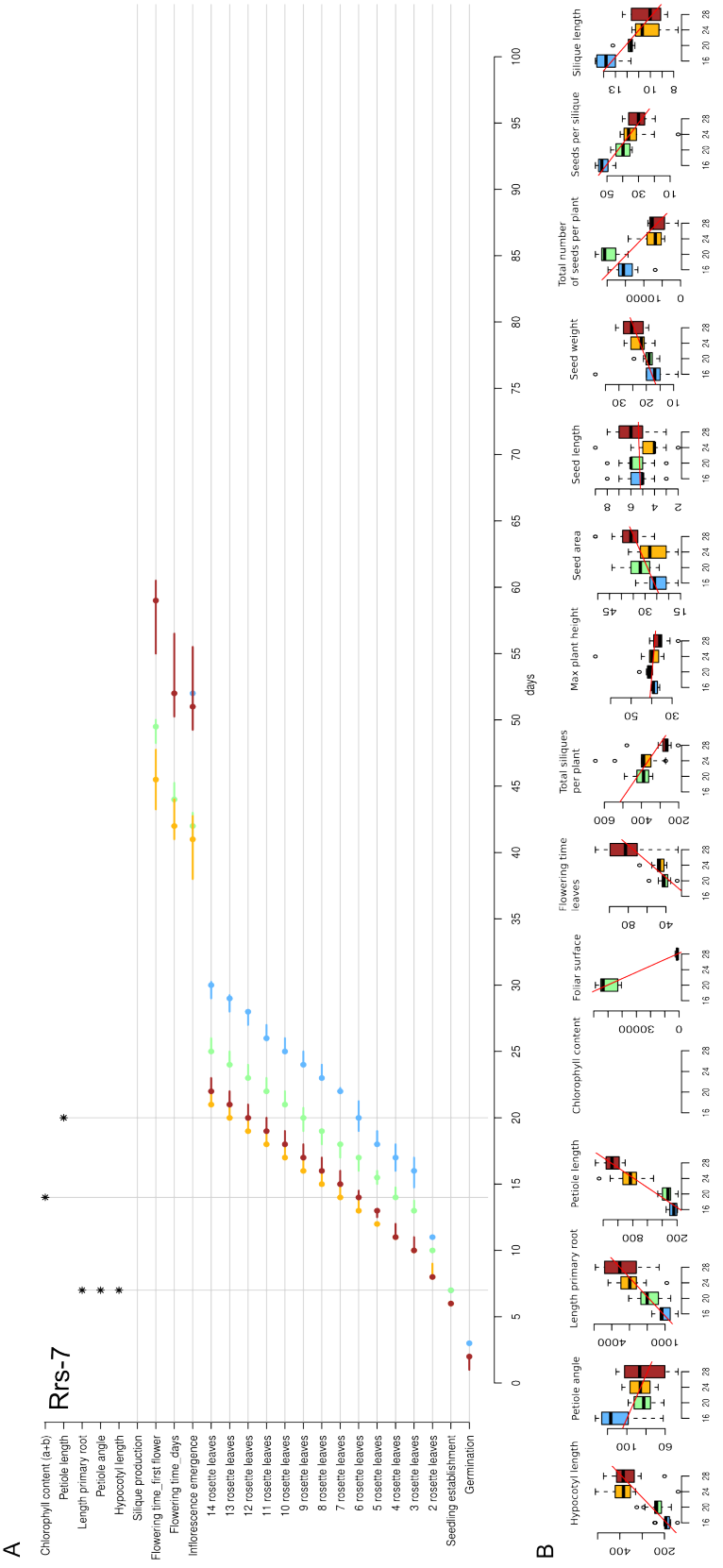


Figure D.8.: Summary of Rrs-7 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of Rrs-7 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

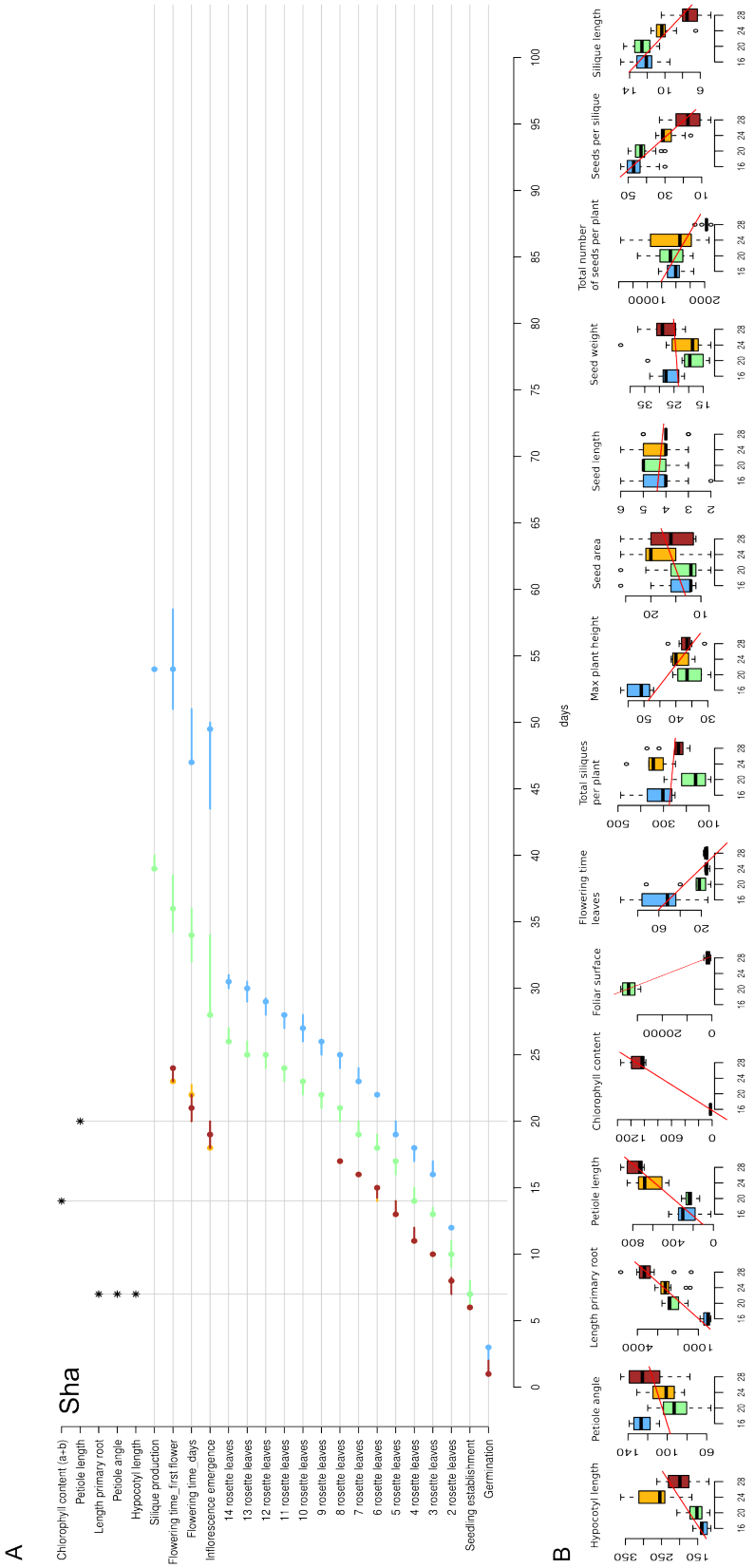


Figure D.9.: Summary of Sha thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of Sha grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

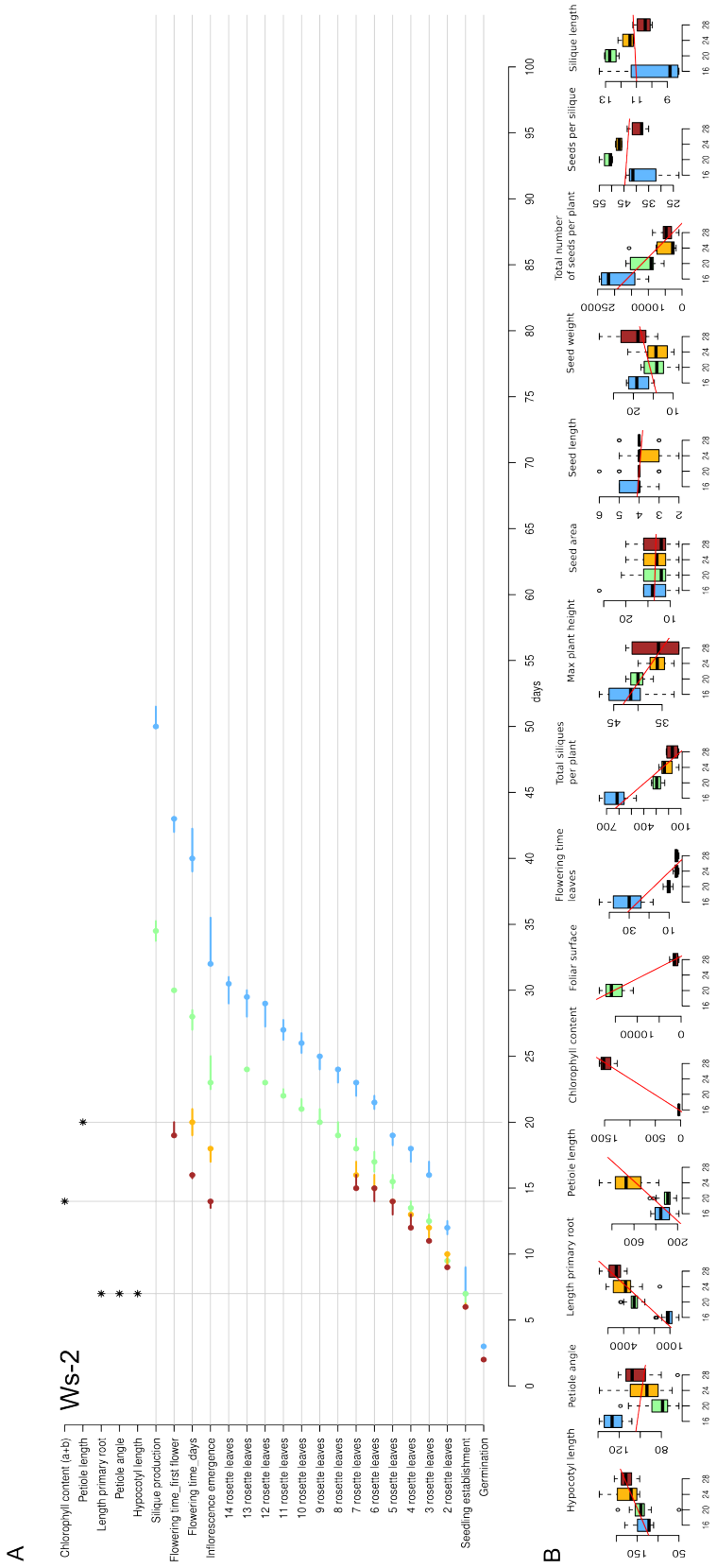


Figure D.10.: Summary of Ws-2 thermomorphogenesis (A) Developmental timing and (B) quantitative phenotypes of Ws-2 grown at 16 °C (blue), 20 °C (green), 24 °C (yellow), or 28 °C (red). Solid red lines in box plots show slopes derived from linear regression of data. Trait units (x-axis) are noted in Table 7.1. Asterisks denote time of phenotypic assessments as described in Figure 7.1.

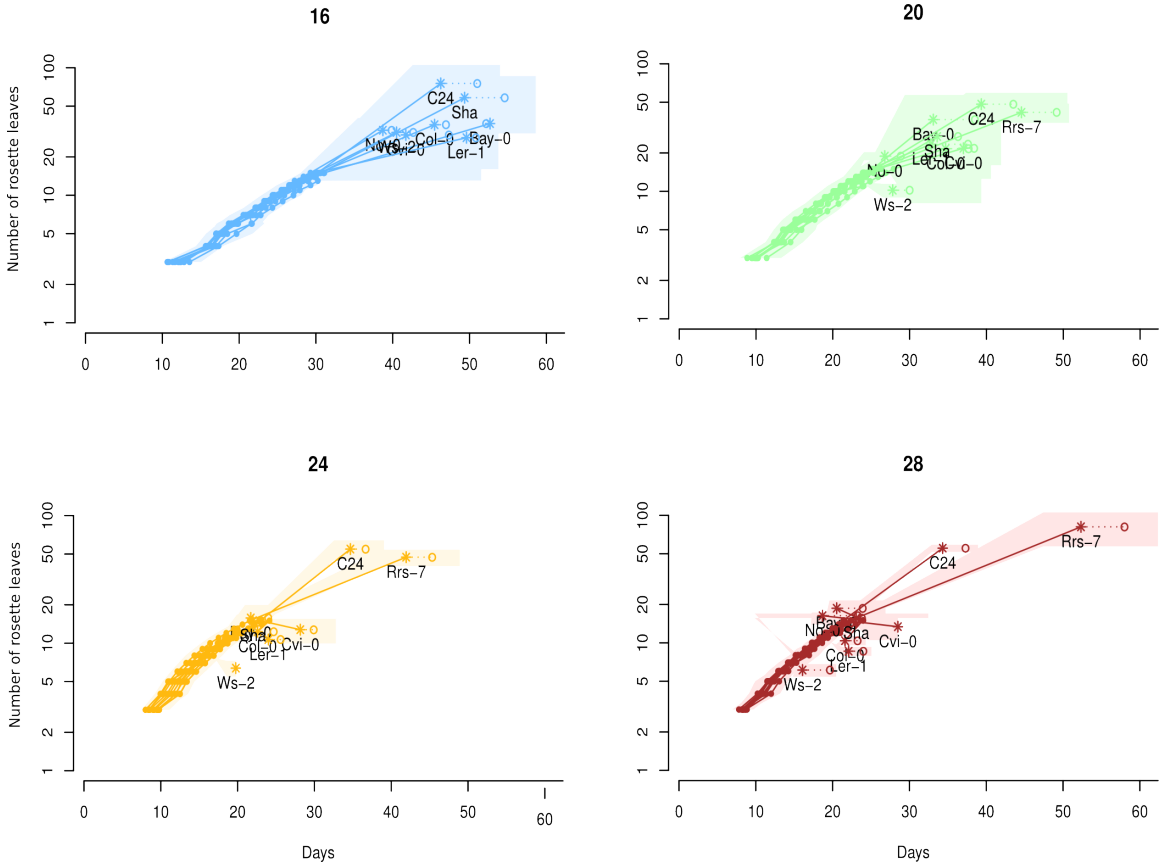


Figure D.11.: Natural variation in developmental timing (leaves vs. days) Plot of the relationship of leaf development over time in developmental timing for each ambient temperature profile. Data points show mean values with filled circles representing vegetative development, asterisks show flowering time (bolting) and open circles the time of first flower opening. Shaded areas denote ranges of standard deviations.

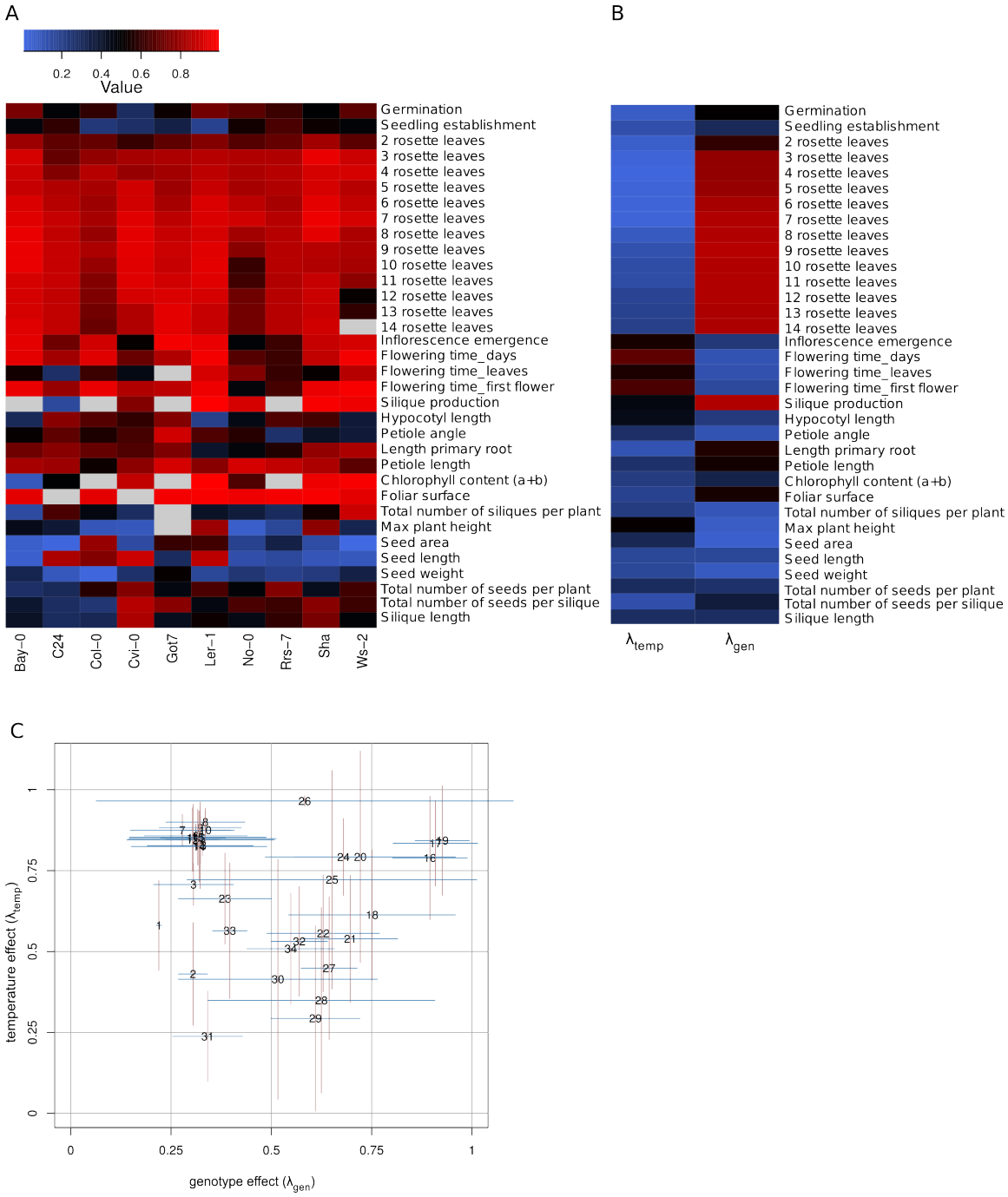


Figure D.12.: Temperature effects on phentoypic variation ($\lambda_{\mathbf{temp}}$), mean and standard deviation of $\lambda_{\mathbf{temp}}$ and $\lambda_{\mathbf{gen}}$ values. (A) Similarly to λ_{gen}, λ_{temp} values were calculated for all accessions to assess temperature effects on phentotypic variation. (B) To assess global patterns, mean λ values were determined across all temperatures (mean λ_{gen}) or across all accessions (mean λ_{temp}). (C) Scatter plot of mean λ_{gen} and mean λ_{temp} values including standard deviation, corresponding to data presented in Figure 7.3C. Missing data is shown in grey.

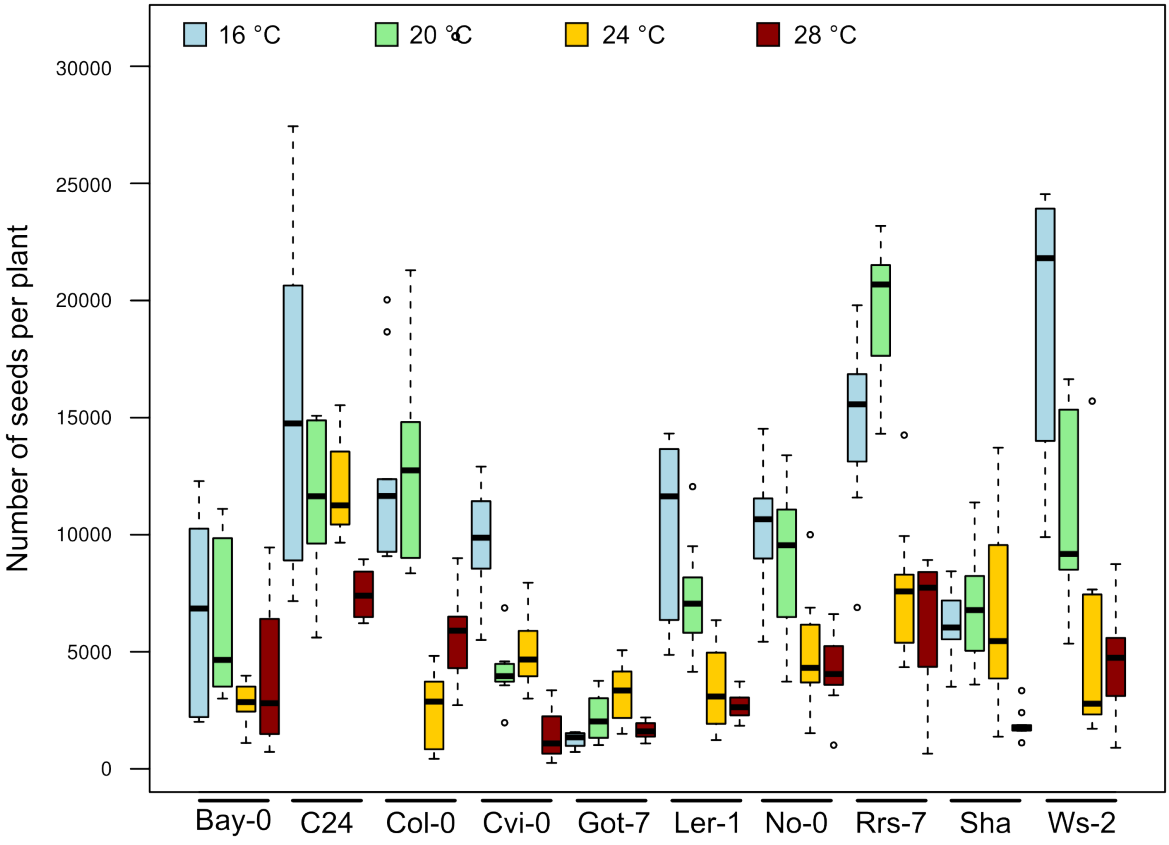


Figure D.13.: Temperature effect on yield (absolute values) Box plots of total number of seeds per plant corresponding to the relative data presented in Figure 7.4A.

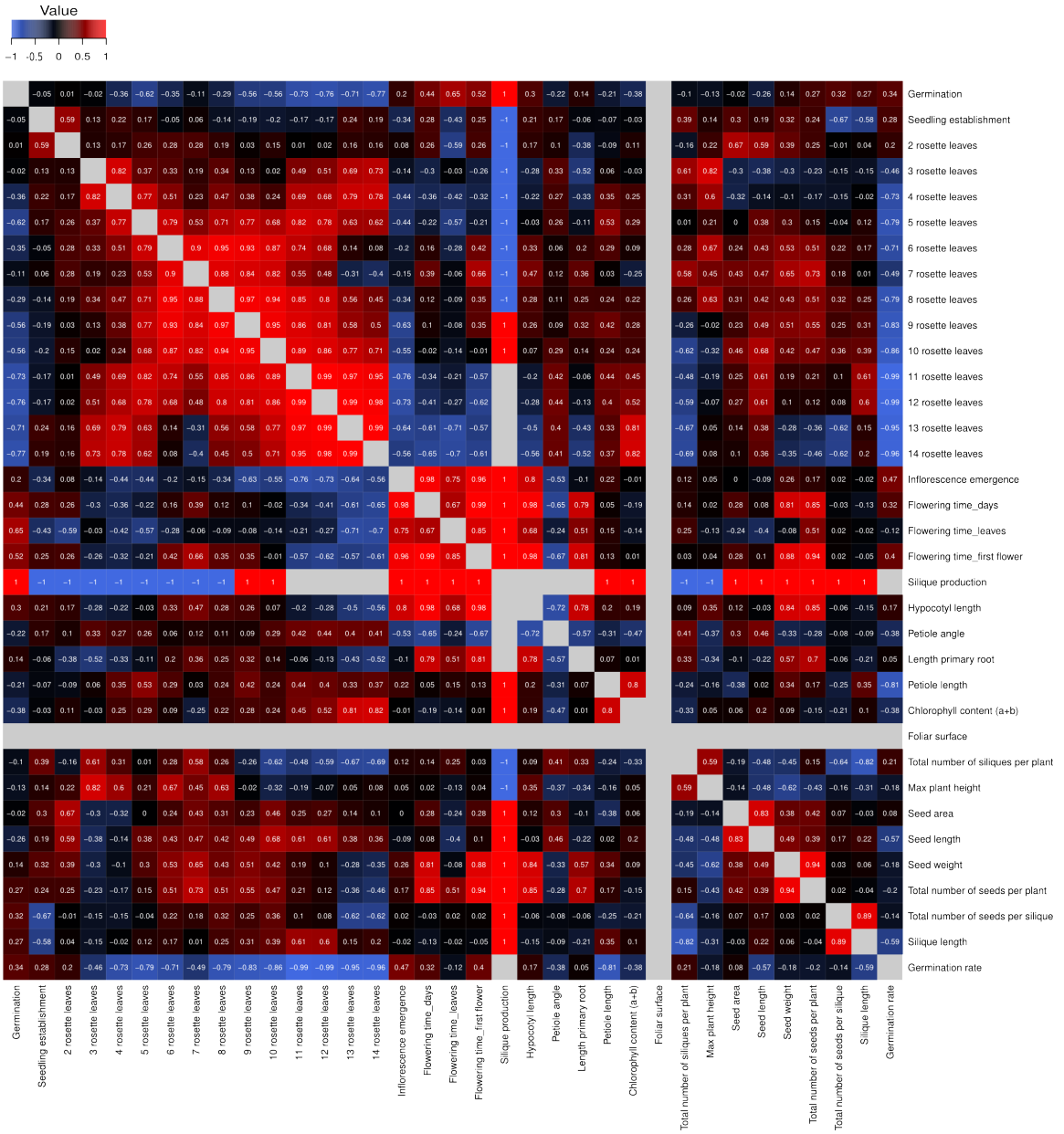


Figure D.14.: Correlations among temperature response ratios (28 vs. 16 °C) Heatmap of Pearson correlation values of temperature responses (28 vs. 16 °C) among all phenotype pairs. Data corresponds to example data presented in Figure 7.4C. Missing data is shown in grey.

D.2. Tables

Table D.1.: Identity and geographic origin of analyzed *A. thaliana* accessions

Abbreviation	Name	Stock ID	Country	Latitude	Longitude
Bay-0	Bayreuth	N954	Germany	N49	E11
Col-0	Columbia	N1092	USA	N38	W92
C24	C24	N906	Portugal	N40	W8
Cvi-0	Cape Verde Islands	N22682	Cape Verde	N17	W23
Got7	Goettingen-7	N22685	Germany	N51	E10
Ler-1	Landsberg (er)	N22686	Poland	N51	E19
No-0	Nossen	N28564	Germany	N51	E10
Rrs-7	RRS	N22688	USA	N41	W86
Sha	Shakdara	N929	Tadjikistan	N38	E68
Ws-2	Wassilewskija	N2360	Russia	N52	E30

Eidesstattliche Erklärung / *Declaration under Oath*

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.
I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

_____ _____

Datum / Date Unterschrift des Antragstellers / *Signature of the applicant*

Lebenslauf

Persönliche Daten

Geburtsdatum:	21. Januar 1981
Geburtsort:	Wolfen (Bitterfeld)
Familienstand:	ledig, ein Kind

Schulbildung

1987–1991	Oberschule Otto Pawliki, Wolfen
1991–1999	Gymnasium Wolfen Nord, Wolfen
09.07.1999	Abitur

Universitäre Bildung

ab 10/1999	Studium der Bioinformatik (Diplom) an der Martin-Luther-Universität Halle–Wittenberg
05.01.2005	Diplom der Bioinformatik

Tätigkeiten

03/2005-08/2008	wissenschaftliche Mitarbeiterin in der Arbeitsgruppe "Bioinformatik und Massenspektrometrie", Abteilung Stress- und Entwicklungsbiologie, Leibniz-Institut für Pflanzenbiochemie
09/2008-08/2013	wissenschaftliche Mitarbeiterin in der Arbeitsgruppe "Bioinformatik", Institut für Informatik, Naturwissenschaftliche Fakultät III, Martin-Luther-Universität Halle–Wittenberg
11/2013	wissenschaftliche Mitarbeiterin in der "Bioinformatics Unit" am Deutschen Zentrum für integrative Biodiversitätsforschung (iDiv) Halle-Jena-Leipzig und in der Arbeitsgruppe "Bioinformatik", Institut für Informatik, Naturwissenschaftliche Fakultät III, Martin-Luther-Universität Halle–Wittenberg

Halle, den

_____

Yvonne Pöschl