

Population and quantitative genetic
analysis of auxin response pathways in
Arabidopsis thaliana

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I – Biowissenschaften –

der Martin-Luther-Universität

Halle-Wittenberg,

vorgelegt

von Herrn **Kristian Karsten Ullrich**

geb. am 15.10.1980 in Kiel

Thesis referees

Prof. Dr. Dierk Scheel (1st referee)
Director of the department of Stress and Developmental Biology
Leibniz Institute of Plant Biochemistry, Halle (Saale)

Prof. Dr. Ivo Grosse (2nd referee)
Institute of Computer Science
Martin Luther University, Halle-Wittenberg

PD Dr. Ivan Paponov (3rd referee)
Institute of Biology II / Botany
Albert-Ludwigs-University, Freiburg

Thesis committee director

Prof. Dr. Gunter Reuter
Institute of Genetics / Developmental Genetics
Martin Luther University, Halle-Wittenberg

Thesis committee members

Prof. Dr. Dierk Scheel (Thesis supervisor)
Director of the department of Stress and Developmental Biology
Leibniz Institute of Plant Biochemistry, Halle (Saale)

Dr. Marcel Quint (Thesis co-supervisor)
Group leader at the department of Molecular Signal Processing
Leibniz Institute of Plant Biochemistry, Halle (Saale)

Prof. Dr. Ivo Grosse
Institute of Computer Science
Martin Luther University, Halle-Wittenberg

Prof. Dr. Björn Junker
Institute of Pharmacy / Biosynthesis of Active Substances
Martin Luther University, Halle-Wittenberg

PD Dr. Matthias Hoffmann
Institute of Biology, Geobotany and Botanical Garden
Martin Luther University, Halle-Wittenberg

Day of disputation: 09.10.2013

Abstract

Plant hormones are primary regulators of plant growth and development. The phytohormone auxin is related to almost all of these growth-related processes. In this thesis, I studied naturally occurring variation of growth-related traits in young *Arabidopsis thaliana* seedlings upon auxin treatments. In the context of adaptive selection, different quantitative genetic approaches were used and combined with the results of a population genetic study to identify genes, which might contribute to the observed phenotypic variation. The population genetic analysis included a total of 151 genes known to regulate auxin biosynthesis, metabolism, transport and signaling. While especially auxin transport genes seemed to be highly conserved, the auxin metabolism genes were identified as the most variable gene group among the auxin network genes. Consistent with these findings, the auxin metabolism genes also showed the highest transcript diversity for a set of six diverse *A. thaliana* accessions, based on a previously published gene expression data set (Delker et al., 2010). In addition to auxin metabolism genes, the results of the population genetic analysis indicated that the *AUXIN/INDOLE-3-ACETIC ACID* gene family is most variable among the gene families involved in auxin signaling. In contrast to that, like the auxin transport genes the co-receptor *TRANSPORT INHIBITOR RESISTANT 1/AUXIN F-BOX PROTEIN* gene family showed signatures of negative selection. To analyze auxin response traits on a functional level, quantitative genetic analyses like quantitative trait loci (QTL) mapping and genome-wide association (GWA) mapping were conducted. The successful development of the software RootDetection, which is capable of measuring primary root length in a high-throughput manner, assisted the phenotypic data acquisition. In general, the genetic architecture regulating the phenotypic variation in the investigated populations seems to be very complex and dominated by small effect loci. However, the QTL, with the strongest genotype-phenotype correlation harbours three auxin metabolism genes, supporting the findings of the population genetic analysis. Hence, auxin metabolism genes potentially contribute to the observed phenotypic variation. The results of a GWA study for auxin response traits identified *AUXIN UP-REGULATED F-BOX PROTEIN 2* as a candidate gene to influence phenotypic variation between the analyzed *A. thaliana* accessions. In addition, the results of the GWA study indicated many genotype-phenotype correlations of moderate significance, which once again highlights the complex nature of the investigated traits. Taken together, while the complex architecture of the auxin network probably prevented the identification of large effect loci, some promising candidate genes and genomic regions were identified which require future functional validation.

Index

Abstract	I
Index	II
1 General Introduction	1
1.1 Auxin and its impact on plant growth and development	1
1.2 Molecular sources of natural variation	3
1.3 Quantitative genetic methods to identify complex traits in <i>Arabidopsis thaliana</i>	6
1.4 Objectives and outline of this thesis	9
2 Materials and methods	10
2.1 Materials and methods for a population genetic analyses of auxin network genes in <i>A. thaliana</i>	10
2.2 Materials and methods for quantitative genetic analysis for two RIL populations	20
2.3 Materials and methods for quantitative genetic analysis for 80 <i>A. thaliana</i> accessions	24
3 Results	29
3.1 A population genetic analysis of genes regulating the auxin network	29
3.1.1 Major SNPs detected in the re-sequenced data and MPICao2010 data for auxin network genes	31
3.1.2 Nucleotide variation highlighted by different population genetic parameters	33
3.1.3 Variation of population genetic parameters inside the auxin signaling gene families	41
3.1.4 Interspecific nucleotide divergence	43
3.1.5 Extreme auxin network genes	45
3.1.6 Comparing sequence and transcript diversity	47
3.2 Analysis of auxin response traits in two RIL populations	49
3.2.1 A stable and user-friendly image processing software for high-throughput root length quantification and more	49
3.2.2 QTL analysis for auxin responses in two <i>Arabidopsis thaliana</i> recombinant inbred line populations	52
3.2.3 Validation of QTLs for auxin related traits by using near isogenic inbred line (NIL) populations	68
3.3 Natural genetic variation of genes involved in auxin biology evaluated by association mapping	72
3.3.1 80 <i>Arabidopsis thaliana</i> accessions screened for phenotypic and genotypic relatedness	72
3.3.2 Unravelling the correlation of phenotype and genotype for auxin response traits by the use of genome-wide association (GWA) mapping	76
3.3.3 A closer look at <i>a priori</i> candidate genes (CGs)	80

4 Discussion	86
4.1 Major SNPs, conserved pathways and amino acid changes	86
4.2 QTL analysis of auxin response traits	93
4.3 GWA analyses of auxin response traits	95
4.4 Concluding remarks	97
5 References	99
Appendix	111
List of figures	111
List of tables	113
Supplementary figures	115
Supplementary tables	144
Glossary	161
German abstract	IV
Acknowledgements	V
Curriculum vitae	VII
List of publications	VIII
Statutory declaration	IX

General Introduction

1.1 Auxin and its impact on plant growth and development

The phytohormone auxin is one of the major regulators of plant growth and development. The signaling molecule indole-3-acetic acid (IAA) itself was identified by Went and Thimann (1937), but the existence of such a signaling molecule had been predicted far earlier dating back to the days of Charles Darwin (Darwin and Darwin, 1880).

Figure F1.1 illustrates four important levels of auxin biology: synthesis, signaling, metabolism and transport.

A tryptophan (TRP)-independent and a TRP-dependent pathway have been proposed for the biosynthesis of auxin. Among the TRP-dependent pathway, four distinct routes have been proposed, named after their intermediate products (Indole-3-pyruvic acid (IPA), indole-3-acetamide (IAM), tryptamine (TAM) and indole-3-acetaldoxime (IAOx)). Recently, the major TRP-dependent auxin biosynthesis pathway was defined for *A. thaliana* to be the IPA pathway including the YUCCA genes which can directly convert IPA to IAA (Mashiguchi et al. (2011), Won et al. (2011) and Stepanova et al. (2011)). For a detailed description of the individual auxin synthesis pathways please see the review of Mano and Nemoto (2012).

As already noted by Thimann (1938), auxin stimulates or inhibits cell elongation, depending on concentration and plant tissue. As outlined in the review of De Smet et al. (2010), the distribution of auxin and its gradients determine the appropriate organ specifications in the *A. thaliana* embryo. Vanneste and Friml (2009) provide several examples on how important auxin gradients are to regulate plant growth in response to environmental stimuli and organogenesis. Auxin itself can be distributed through different types of auxin transporters. IAA influx into cells is mediated by auxin-uptake carriers encoded by the *AUXIN RESISTANT 1 (AUX1)* and *LIKE AUX1 (LAX)* gene family. Auxin moves out of plant cells through efflux proteins, including *PIN-FORMED (PIN)* and *ATP-BINDING CASSETTE FAMILY B (ABCB)* proteins. Recently, a new auxin transport family was detected by Barbez et al. (2012), the *PIN LIKES (PILS)* gene family. PILSs proteins regulate the intracellular auxin accumulation at the endoplasmatic reticulum.

Ludwig-Müller (2011) reviewed the process of auxin conjugation in the context of auxin homeostasis. The *GRETCHEN HAGEN 3 (GH3)* gene family, which also belongs to a group of genes rapidly induced by auxin stimuli, facilitates the conjugation of amides to IAA. Whereas some amide-conjugates are proposed to be storage forms of IAA, which can be hydrolysed into free IAA, other amide-conjugates are proposed to be involved in auxin degradation pathways. The amide-conjugate TRP-IAA is thought to be an inhibitor of auxin action. In addition to amide-conjugates, there exist ester-, protein- and glucose-conjugates. The corresponding enzymes, which facilitate the conjugation of IAA, can also con-

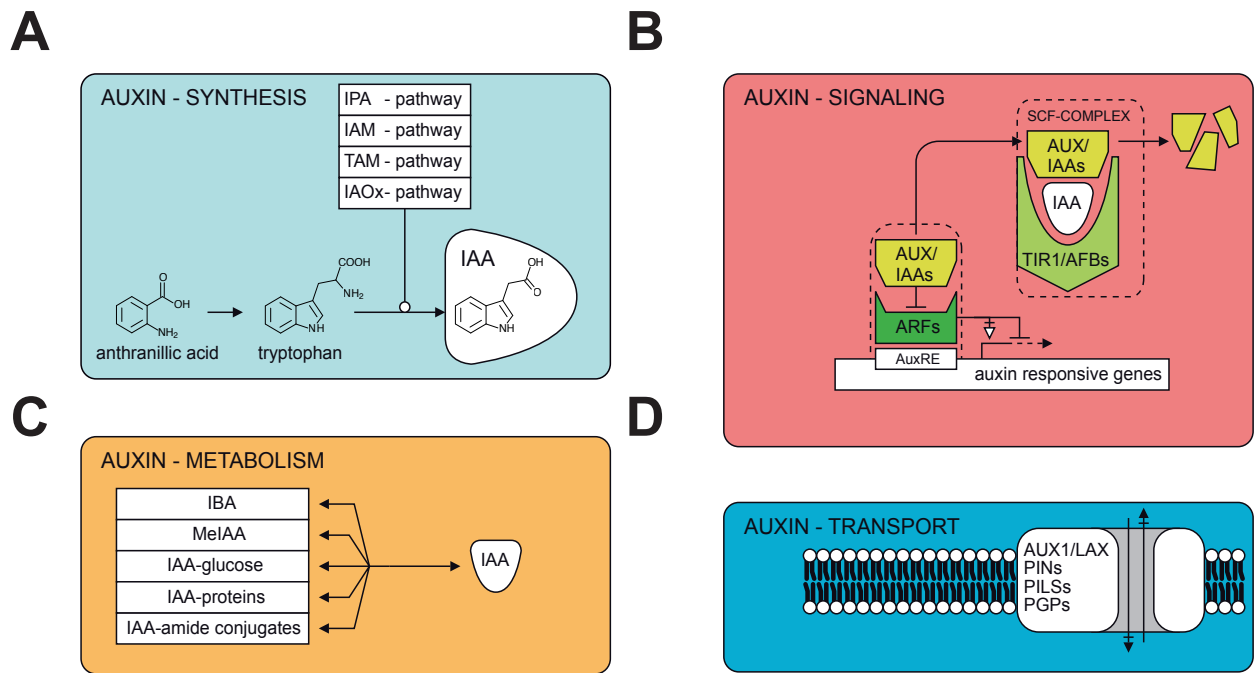


Figure F1.1: Overview of auxin biology. (A) Four postulated Trp-dependent IAA biosynthesis pathways. Indole-3-pyruvic acid (IPA) pathway, indole-3-acetamide (IAM) pathway, tryptamine (TAM) pathway and indole-3-acetaldoxime (IAOx) pathway. (B) Auxin signaling, with the *TIR1/AFB* co-receptor gene family, which recruit *AUX/IAA* repressor proteins to the *SCF*-complex. Together *TIR1/AFBs* and *AUX/IAAs* serve as a co-receptor complex (Calderón et al., 2012). The *AUX/IAAs* are flagged with ubiquitin for degradation by the proteasome. The auxin response factors (*ARFs*) are then no longer blocked by the *AUX/IAAs* and can induce or reduce gene expression of auxin responsive genes. (C) Auxin metabolism. IAA can be conjugated to glucose or even proteins and can be converted to a physiological inactive auxin form. (D) Auxin transport is facilitated by auxin importers like *AUX1/LAX* and auxin exporters like *PINs* and *PGPs* and intracellular auxin transporters, the *PILSs*.

jugate other molecules which show auxin activity, like indole-3-butyric acid (IBA), 4-chloride indole-3-acetic acid (4-Cl-IAA) and indole-3-propionic acid (IPA). As outlined by Vanneste and Friml (2009), the cellular concentration of auxin can be controlled at multiple levels, such as biosynthesis, conjugation, deconjugation, degradation and intercellular transport. The auxin metabolism pathways might contribute to form distinct gradients of active auxin molecules and in such a way fine-tune the plant reactions to auxin.

This leads to the question of how different auxin concentrations can trigger specific responses?

As outlined by Vanneste and Friml (2009), at low concentrations of auxin *AUXIN/INDOLE-3-ACETIC ACID (AUX/IAA)* proteins act as transcriptional repressors. They dimerize through their domains III and IV with *AUXIN RESPONSE FACTORS (ARFs)* and in such a way block auxin responsive gene expression (see Figure F1.1 C). Domains III and IV can also lead to homodimerization within the *AUX/IAAs*. Through their domain I, the co-repressor *TOPELESS* can also bind to the *AUX/IAAs* (Szemenyei et al., 2008).

At high concentrations of auxin, the *AUX/IAAs* are recruited to the F-Box proteins *TRANSPORT INHIBITOR RESISTANT 1 (TIR1)* and *AUXIN F-BOX PROTEINs (AFBs)* and form a co-receptor

complex. Here, the domain II of the *AUX/IAAs* plays a leading role in the interaction with the *SCF*-complex. The *AUX/IAAs* are then ubiquitinated and subsequently degraded by the 26S proteasome. *ARFs* are nuclear proteins which also contain domains III and IV. An additional domain is the amino-terminal DNA-binding domain (DBD), which lies into a plant-specific B3-type transcription factor domain. This DBD domain binds to so-called auxin response elements (AuxREs), which can be found in promoter regions of auxin responsive genes. The middle region of *ARFs* can be glutamine rich or proline rich, which determines whether the *ARFs* act as transcriptional activators or transcriptional repressors. After *AUX/IAA* degradation, the *ARFs* are free to induce or repress gene expression of auxin responsive genes. As reviewed by Quint and Gray (2006), *ARFs* are targets of microRNAs (miRNAs) and trans-acting small interfering RNAs (tasiRNAs), which reduce *ARF* gene expression and can influence plant development.

As one of the main regulators of plant growth and development, auxin is of course also involved in seedling development, the process in the focus of this thesis.

Due to its involvement in many aspects of plant development, only its key functions were highlighted here. For a very detailed view of auxin and its role during a plants life cycle, a number of detailed reviews can be consulted (Abel and Theologis (2010), Delker et al. (2008), Teale et al. (2008), Teale et al. (2006) and Weijers and Jürgens (2005)), for auxin-regulated gene expression (Chapman and Estelle (2009) and Quint and Gray (2006)), for auxin distribution (Vanneste and Friml (2009) and Benjamins and Scheres (2008)), for auxin biosynthesis (Mano and Nemoto, 2012), for auxin metabolism (Ludwig-Müller, 2011) and for auxin crosstalk to other phytohormones (Vanstraelen and Benkova (2012), Muday et al. (2012) and Depuydt and Hardtke (2011)).

1.2 Molecular sources of natural variation

With the tremendous decrease of sequencing costs, more and more plant species are and will be sequenced. The sequence information is used to analyze plant genomic variation not only between species but also within species. For *Arabidopsis thaliana* an initial step was made by Nordborg et al. (2005). In this study, 96 *A. thaliana* accessions were partially sequenced by Sanger-sequencing. A study of Clark et al. (2007), in which 20 diverse accessions were sequenced by an array-approach, increased the amount of single nucleotide polymorphism (SNP) data. Finally, in 2009 a project was initiated to sequence 1001 accessions of *A. thaliana* (Weigel and Mott, 2009). 80 out of the proposed 1001 genomes of *A. thaliana* have recently been published (Cao et al., 2011). Furthermore, Horton et al. (2012) provided SNP data for 1307 *A. thaliana* accessions based on a 250k SNP chip. These genotyping efforts provide a rich resource for GWAS for the *A. thaliana* community.

All these large-scale analyses provide genomic variation data. Based on these data, variation on polyploidization, transposable element insertions, gene duplication status and SNPs can be identified and prediction on protein function might lead to prediction on phenotypic effects (Günther and Schmid, 2010). With the help of all these resources, the molecular evolutionary scientist can answer questions concerning natural occurring phenotypic variation.

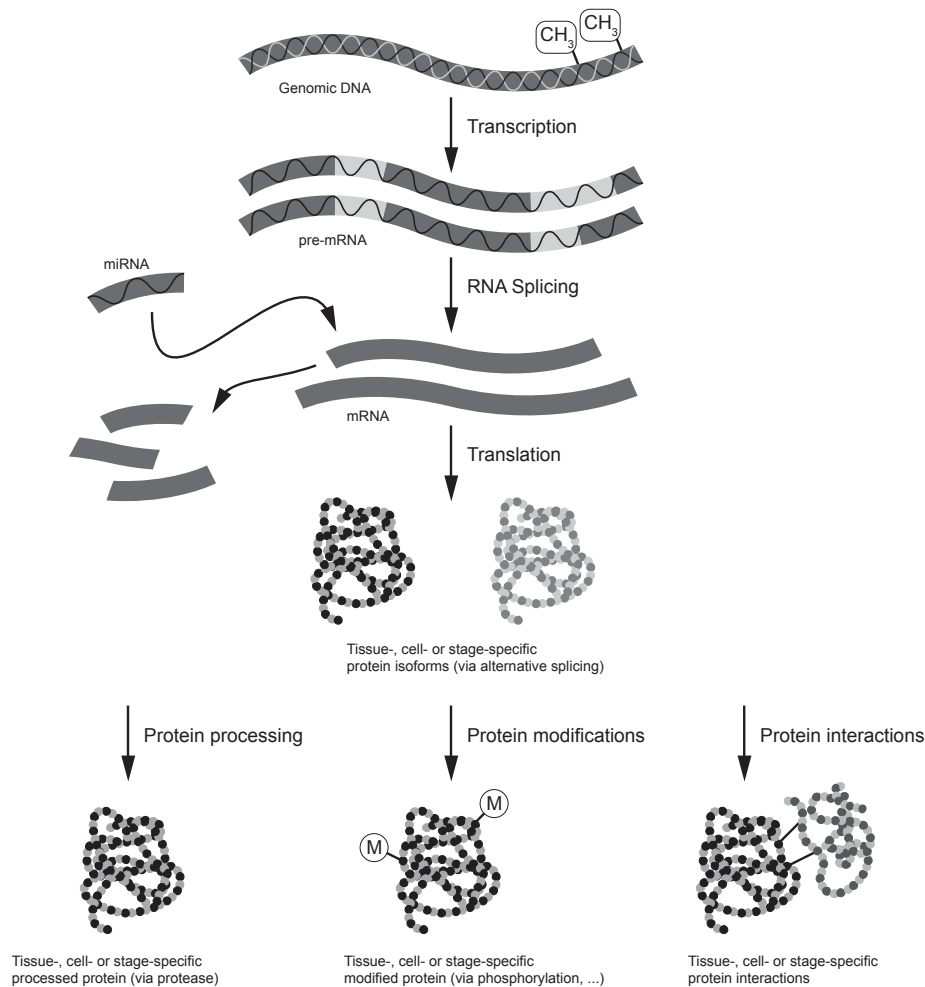


Figure F1.2: Molecular sources of phenotypic variation. Molecular sources of phenotypic variation illustrated as part of the protein synthesis pathway. Starting with genomic variation between species as divergence and within one species as diversity, different genes and/or different splice variants from the same gene can be present leading to different mRNAs. Depending on promoter sequences or methylation status, the transcription level might be altered, which can also be due to differences in transcription machinery. Diversity on the level of miRNA can alter mRNA levels and downstream abundance of translated proteins. Proteins might be processed in different ways via altered protease activity or altered modification machinery. These post-translational modifications might be different in a tissue-, cell- or stage-specific manner. This kind of variation might also be true for protein-protein interactions. All possible combinations of these different layers of molecular variation can be imagined, which might lead to phenotypic variation between organisms in an investigated population.

Genetic variation between species and within a species can be observed on the genomic level. Considering the genetic features, which are encoded on the DNA level, different layers can be affected and can lead to natural phenotypic variation (see Figure F1.2). One driving force, which can alter allele frequencies on the genomic level is natural selection. Organisms compete for resources, and those with genes that better adapt them to their environment have a greater probability that these genes are passed on to the next generation and will increase in frequency in the population (Page and Holmes, 1998). Kimura (1968) proposed the neutral theory of molecular evolution. In addition, two other theories exist: the nearly neutral model of molecular evolution (reviewed in Ohta (1992)) and the selcionist kind of view (outlined in Gillespie (1991)). All these theories have in common that the

majority of mutations are deleterious and will be removed by negative selection. Taking into account functional constraint of genes, differences will be expected. In a functionally constraint gene, most of the mutations will be deleterious and will be removed by negative selection. For genes with less functional constraint or even for non-coding DNA, the rate of neutral substitutions is higher so that the rate of substitutions increases (Page and Holmes, 1998). The theories diverge with regard to the substitutions which will be fixed in the population. For neutralists, the majority of these loci will be neutral and will be fixed just by random sampling, also called genetic drift. For selectionists, the majority of these fixed substitutions would be not neutral but advantageous and would be fixed by positive selection. Follower of the nearly neutral theory would expect that the group of fixed substitutions is splitted into one part fixed by genetic drift and another part fixed by slightly advantageous substitutions. The conclusion from all these theories is, that both natural selection and genetic drift determine the evolutionary fate of mutations (Page and Holmes, 1998).

In addition to natural selection and genetic drift, migration and demographic histories (e.g. recent bottlenecks) can influence allele frequencies. Schmid et al. (2005) and Nordborg et al. (2005) showed that population genetic parameters calculated for *A. thaliana* do not follow the expectations of the neutral model of molecular evolution. They also showed that population structure and a recent change in population size played a major role in shaping the genetic variation in *A. thaliana*, most likely due to the life history of the species influenced by glacial refugia and postglacial recolonization events (Sharbel et al., 2000).

One of the first systematic population genetic studies on large plant gene families has been conducted with resistance (R-)genes (Bergelson et al., 2001). In this study, amino acid substitution rates (K_{NONSYN}) have been compared to synonymous substitution rates (K_{SYN}) between *A. thaliana* and *A. lyrata* to infer selection forces acting on the R-genes. Bakker et al. (2006) conducted also a study on plant R-genes, in this case within 96 *A. thaliana* accessions. They could show that there exists weak balancing selection acting on the R-genes by comparing them to the data of Nordborg et al. (2005). This comparison was necessary to overcome the fact that the population genetic parameters do not follow the neutral model of molecular evolution. Hence, the data of Nordborg et al. (2005) were used in different studies to infer selection forces acting on gene families. Bakker et al. (2008) conducted a population genetic study on defense related genes and identified signatures of purifying selection. Sterken et al. (2009) conducted a similar population genetic approach and investigated core cell cycle genes in 30 *A. thaliana* accessions. Again, the observed data was compared to the data of Nordborg et al. (2005) and signatures of purifying selection were identified. However, the genetic robustness of cell cycle-related processes is supposed to be due to functional redundancy rather than high selective constraint (Sterken et al., 2009).

Vaughn et al. (2007) showed the impact of epigenetic natural variation in *A. thaliana* as an additional layer which can cause phenotypic variation. Alonso-Blanco et al. (2009) reviewed the findings based on natural variation and its impact on plant development and physiology. The majority of traits, which shows natural variation, is determined by multiple loci and genes and is therefore of multigenic nature. As outlined by Alonso-Blanco et al. (2009), the vast majority of genes, which causes natural variation, was identified by quantitative trait loci (QTL) analysis. The allelic variants of these genes contribute to adaptation to different environments. The genetic variation ranged from nonsense mutations and

indels to splice site mutations not only in coding but also in promoter regions. Some of the identified genetic variants represent natural null alleles of proteins. Other genetic variants affect the expression level, showing again the broad range of molecular sources of phenotypic variation (see Figure F1.2). The different quantitative genetic methods, which can identify and unravel complex traits are outlined in the next subsection.

1.3 Quantitative genetic methods to identify complex traits in *Arabidopsis thaliana*

A. thaliana is distributed in a wide range throughout the northern hemisphere and occur in different environments and different climates (Hoffmann, 2002). Figure F1.3 shows the distribution of 1307 *A. thaliana* accessions, which have been genotyped with a 250k SNP chip and provide a great resource for genome-wide association (GWA) analysis (Horton et al., 2012). In recent studies, local adaption to climate could be linked to allelic variants by conducting GWA analysis in *A. thaliana* (Fournier-Level et al. (2011) and Hancock et al. (2011)). The genetic variants between *A. thaliana* accessions can be used in forward genetic approaches to identify genes, which contribute to phenotypic variation. Phenotypic screening is used to identify individuals with altered phenotypes in the trait of interest, and then various genetic and molecular approaches are used to identify the relevant genetic alterations (Assmann, 2012).

Alonso-Blanco and Koornneef (2000) pointed the naturally occurring variation out as a great resource to detect the causative genes for phenotypic variation by the use of QTL analysis. In the following years until today recombinant inbred lines (RILs) were used to conduct QTL analysis in *A. thaliana*, (reviewed in Weigel (2012)).

Kowalski et al. (1994) identified QTLs for naturally occurring variation in flowering time and many other QTL studies followed for flowering time in *A. thaliana* as reviewed in Koornneef et al. (2004). Bentsink et al. (2006) identified *DELAY OF GERMINATION 1 (DOG1)* as a gene influencing seed dormancy, followed by additional genes influencing the same trait (Bentsink et al., 2010). Hence, QTL mapping has proven to be a successful approach to detect genes influencing natural variation.

In a classical linkage mapping approach, two diverse *A. thaliana* accessions are crossed and the F2 population is analyzed for the phenotypic trait of interest. To overcome the problem that an F2 population can only be screened once, RIL populations have been developed. They are produced by subsequent selfing of the F2 population of usually six rounds of inbreeding. The resulting RIL populations are almost completely homozygous and have the advantage that after genotyping, they can be used multiple times in different experiments and also different environments representing a theoretically immortal F2 population. As outlined by Bergelson and Roux (2010) and Weigel (2012), around 60 of such RIL populations are available for *A. thaliana*. The RIL population can be screened for the phenotypic trait of interest and correlation analysis can be performed (QTL mapping) to find genomic regions which are linked to the phenotypic variation. The resolution of the conducted QTL mapping can narrow down the genomic region of up to a few megabases depending on first instance of the population size and second on marker density, covering thousands of genes.

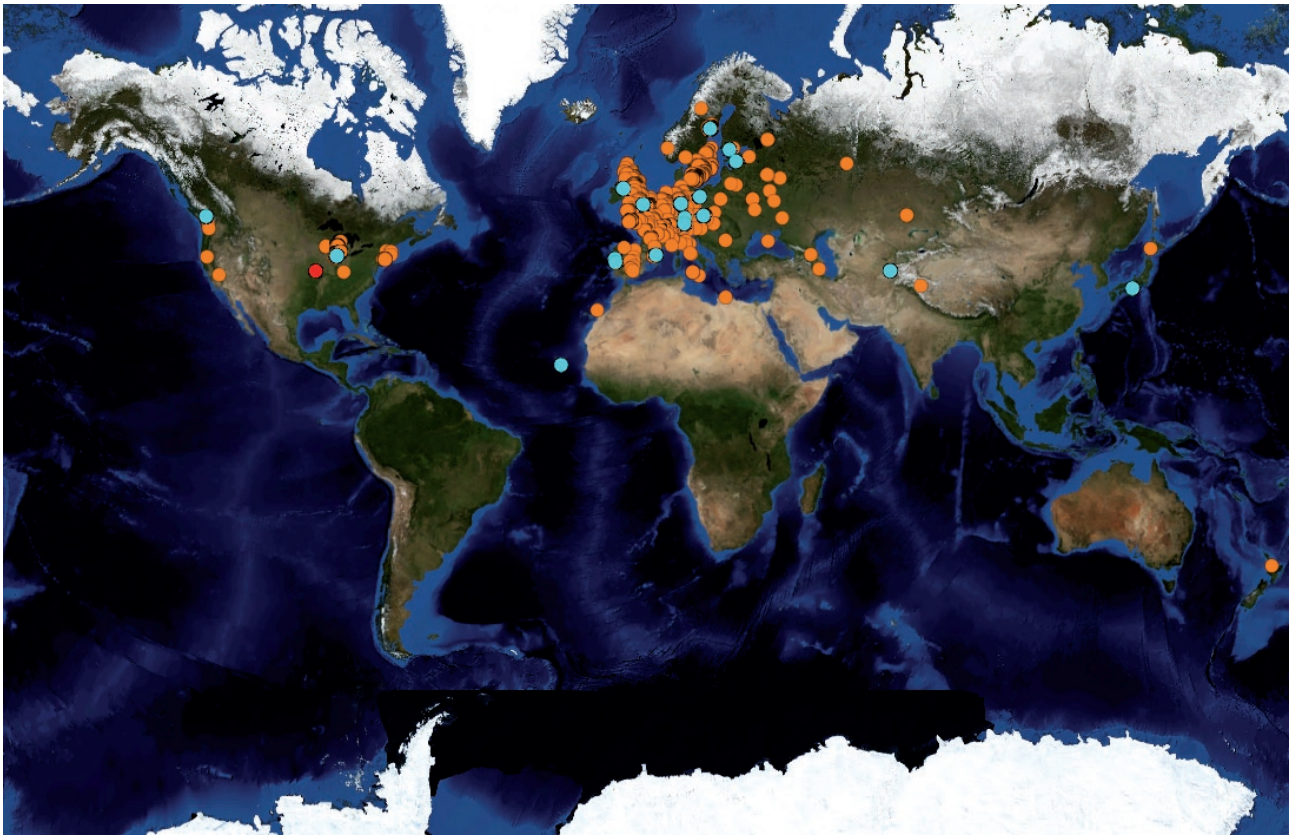


Figure F1.3: Distribution of 1307 genotyped *A. thaliana* accessions. Worldwide distribution of 1307 *A. thaliana* accessions according to Horton et al. (2012). Each orange dot represents the location of one accession. Lightblue dots represents the locations of 20 *A. thaliana* accessions used in a study by Clark et al. (2007). Red dot represents the location for Col-0 according to Anastasio et al. (2011). Locations were used as input for WORLDWIND (<http://worldwind.arc.nasa.gov/java/>).

As outlined by Weigel (2012), advanced mapping populations as introduced by Davarsi and Soller (1995) and Balsubramanian et al. (2009) can increase the power of mapping populations. In advanced intercross RILs individuals from the F₂ and later generations are intermated before inbred lines are derived.

Another approach to increase the power of mapping populations and to even incorporate genetic variation from more than two accessions is facilitated in the multiple advanced generation intercross, the so called MAGIC lines, and the *Arabidopsis* multiparent RIL (AMPRIL) design (Kover et al. (2009) and Huang et al. (2011)).

After a genomic region is detected for which a correlation with the phenotypic variation is apparent, near isogenic lines (NILs) or heterogenous inbred families (HIFs) are used to validate and further narrow down the mapping interval. NILs consist of small genomic regions introgressed from one parental line in the background of the other parental line. HIFs are generated from RILs carrying residual heterozygosity at genomic regions of interest. Basically, both NILs and HIFs provide material that is genetically identical with the exception of the region of interest, where they carry either one of the two parental alleles. Any phenotypic difference between these lines can thus be attributed to the region of interest, resulting in mendalization of a previously multigenic phenotype. After phenotyping for the trait of interest, the mentioned NILs and HIFs can be used to further fine map the genomic

region of interest to finally identify the candidate genes causing the phenotypic variation between the parental lines.

As outlined by Assmann (2012), also other strategies exist to identify the candidate gene in such a correlated genomic region. One approach is bulk segregant analysis. With this approach, individuals are grouped according to their phenotypic variation and genotyped in pools. It is assumed that the evaluated pools of plants segregate for many loci but are homozygous for the loci that cause the phenotypic variation they were grouped for. Another approach is to look into gene expression patterns of the RILs and try to find the responsible gene in the correlated genomic region under the assumption that variation in expression is linked to the phenotypic variation (DeRose-Wilson and Gaut (2011) and Assmann (2012)). One disadvantage of traditional linkage mapping is that the genetic regions, that are identified by QTL mapping, are specific to the parental lines of the experimental segregating populations and may not be representative of the genetic variation on which natural selection acts (Bergelson and Roux, 2010).

Here, GWA mapping can overcome the restrictions of QTL mapping. GWA mapping uses natural linkage disequilibrium (LD) to identify polymorphisms that are associated with phenotypic variation (Bergelson and Roux, 2010). One advantage of using SNP information between diverse accessions is that all recombinations events which happened between these individual accessions throughout the life history of the species can be used for fine mapping. The resolution in GWA mapping is therefore much higher than in traditional QTL mapping.

However, one disadvantage of GWA mapping is that it has less power to detect rare alleles. Another disadvantage of GWA mapping is the difficulty to correct for population structure, which clearly exists within *A. thaliana* (Nordborg et al., 2005). Here, advanced GWA mapping methods have been developed to overcome this problem and to correct for population structure (Yu et al. (2006) and Kang et al. (2008)) and new methods are developed to even conduct multiple trait GWA analysis (Korte et al., 2012).

Brachi et al. (2010) could show that the combination of traditional QTL mapping and GWA mapping leads to the identification of true positive associations for flowering time in *A. thaliana*. Atwell et al. (2010) could demonstrate the successful usage of GWA analysis within *A. thaliana* by elucidating associations for 107 phenotypes. The authors showed that *a priori* candidate genes are over-represented among the strongest associated SNPs. However, also the difficulties of evaluating complex traits have been shown by this study indicating the importance of functional validation of each true association, which can be detected by quantitative genetic methods as supposed by Weigel and Nordborg (2005).

In *A. thaliana*, natural variation was identified for many complex traits, like plant growth and morphology, seed dormancy, flowering time, primary and secondary metabolism, responses to jasmonate, etc. (reviewed by Alonso-Blanco et al. (2009) and Weigel (2012)). In this thesis, auxin response traits will be systematically analyzed for natural variation by quantitative genetic approaches.

1.4 Objectives and outline of this thesis

Both auxin signaling and transport proteins appeared early in the evolution of land plants (De Smet et al. (2011), Paponov et al. (2009) and Viane et al. (2012)). As sessile organisms, plants evolved an amazing degree of developmental plasticity, which enables them to respond to an ever changing environment. Usually, developmental plasticity is achieved by adjusting the plant's growth to environmental stimuli. Considering the fact that almost all growth-related processes are at least in part related to auxin, it becomes inevitable to assess the auxin pathway from an adaptive perspective. Does natural variation for auxin genes exist? Can this variation be used to learn about such a possible adaptive function?

First, a population genetic approach is used in this thesis to detect selection signatures among genes, which are related to auxin biology with a special focus on the auxin signaling gene families. Here, the question was, whether mutations lead to protein coding changes within the conserved auxin signaling genes and which gene family contributes at most to the phenotypic variation.

Second, to gain knowledge about auxin responses during the establishment of the root system of young seedlings and effects on hypocotyl growth in dark-grown seedlings, quantitative genetic approaches are applied to identify loci and evaluate their impact on natural variation for auxin response traits.

A QTL analysis of two RIL populations was conducted to identify genomic regions, which cause natural variation upon auxin response traits in young *A. thaliana* seedlings between the parental accessions. The detected QTLs were investigated for underlying SNPs between the parental accessions of the RIL population. In addition, a validation for the identified QTLs was conducted with suitable NILs.

To identify genes and their allelic variants which contribute to natural variation upon auxin response traits in the global population of *A. thaliana*, a GWA analysis was conducted in 80 accessions. In addition to a GWA method, which can identify differences in phenotypic mean values, also a recently published GWA method was applied, which can identify differences in phenotypic variance values (Shen et al., 2012). Here the question was, if the QTLs detected in a cross between two *A. thaliana* accessions are also involved in natural variation throughout the global population of *A. thaliana*. Another question was, if known candidate genes (CGs) for auxin response traits associate with the phenotypic variation in the GWA experiment.

Materials and methods

2.1 Materials and methods for a population genetic analyses of auxin network genes in *A. thaliana*

DNA samples

In this thesis 18 *A. thaliana* accessions (Bay-0 , Bor-4 , Br-0, Bur-0 , C24, Cvi-0, Est-1, Fei-0, Got-7, Ler-1, Lov-5, Nfa-8, Rrs-7, Rrs-10, Shadara, Tamm-2, Ts-1, Tsu-1) were used to get genomic sequence information by sequencing. These genotypically diverse accessions are part of 20 accessions, which were previously used in the study of Clark et al. (2007) (all except Van-0). The available Col-0 reference sequence was also included in all subsequent analysis, resulting in 19 *A. thaliana* accessions. Seeds of these accessions were obtained from the Nottingham Arabidopsis Stock Center and genomic DNA was extracted from young leaf tissue using the DNeasy Plant Mini Kit (Quiagen).

Primer design and sequencing

Primers were designed with FastPCR (Primer Digital Ltd) on the basis of the Col-0 reference genome for *TIR1/AFBs*, *AUX/IAAs* or *ARFs* (see Supplementary Table ST1), resulting in 49 genes for the auxin signaling group. Primers were selected to maximize coding sequence (CDS) regions within ~1-kb fragments by sequencing if possible the same domain regions between individual gene family members. 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.

Population genetic sequence data

Re-sequenced data

All fragment 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). In addition to the auxin signaling group (49 genes), sequencing data for the auxin synthesis (21 genes), auxin metabolism (12 genes) and auxin transport (15 genes) group were obtained from Stefan Eттingshausen (personal communication) and re-validated as described above. In total, 97 auxin biology genes in 19 *A. thaliana* accessions were analyzed further.

Accession numbers

Sequence data from this thesis for 49 auxin signaling genes have been submitted to GenBank (accession numbers GU348425-GU348653 and HM487319-HM487971).

Nordborg2005 data access

1214 partial genomic sequence alignments from 96 *A. thaliana* accessions (Nordborg et al., 2005) were downloaded from the Gregor-Mendel Institute web site (Nordborg2005 data; see URL subsection 2.1).

MPICao2010 data access

The sequence data for 80 *A. thaliana* accessions (Cao et al., 2011) was obtained from the 1001 Genomes Project web site (Weigel and Mott, 2009) (1001 Genomes Project; see URL subsection 2.1). For each of the 80 accessions, the filtered_variant.txt.gz containing the positions and annotations of single nucleotide polymorphisms (SNPs) and 1-3 bp deletions for the TAIR10 release, was downloaded (MPICao2010 data; see URL subsection 2.1) and the genomic sequence of the published genome of *Arabidopsis thaliana* (TAIR10 chromosome sequences; see URL subsection 2.1) was replaced by the filtered_variant.txt.gz entries with a custom R script and the Biostrings R package resulting in 80 pseudo-chromosome sequences.

Data preparation

Re-sequenced data

Only sequence alignments were processed, for which the total width of concatenated "A", "T", "G" and "C" characters was higher than 400 bp in 17 out of 19 accessions. If an accessions had fewer sites, it was excluded from the corresponding sequence alignment.

Each of the 97 multiple sequence alignments containing 17 to 19 *A. thaliana* accessions was used as a query and searched against the genomic sequence of the published genome of *Arabidopsis thaliana* (TAIR10 chromosome sequences; see URL subsection 2.1) with BLAT (Kent, 2002) to validate their chromosomal position.

The TAIR10 annotated positions for each of the sequence alignments were then used to parse the TAIR10 Genetic feature format version 3 (GFF3) file (TAIR10 GFF3; see URL subsection 2.1) with a custom R script. Genetic features like CDS, five-prime, three-prime or other features were extracted from the alignments with the Biostrings R package to conduct further a population genetic analyses.

If more than one gene model falls within one sequence alignment, the sequence alignment was split into the corresponding, possibly overlapping gene models. The resulting sequence alignments were then analyzed as independent sequence alignments.

By parsing the TAIR10 GFF3 file, for each sequence alignment an individually GFF3 file was constructed. This information could be used to assign the correct reading frame for CDS and to extract gene feature information, like intron start and end positions for a certain fragment.

To account for major or large-effect SNPs, that cause an altered start codon, an altered or preliminary stop codon, a change in the splice donor or acceptor site or lead to an overall sequence width of not multiple of three via deletions, the sequence alignments were checked with a custom R script and excluded from further analysis. Here, splice donor sites different from "GT" and splice acceptor sites different from "AG" were marked as major SNPs and were excluded from further analysis.

In addition to these filters, three variants SNP positions were processed and the corresponding codons were excluded from further analysis, because most population software only can handle biallelic SNPs. Under this criteria the following auxin network genes were excluded: *AT1G34170*, *AT1G68100*, *AT3G62100*, *AT4G13260*, *AT4G39950* and *AT5G54810*.

Four additional representative gene models had to be excluded from the auxin network genes due to major SNPs, which are listed in Table ST3.

For further analysis, additional filtering steps were applied and are indicated, if necessary.

Nordborg2005 data

First, each of the 1214 partial genomic sequence alignments containing 96 accessions was reduced to the following set of 19 accessions: Bay-0, Bor-4, Br-0, Bur-0, C24, Col-0, Cvi-0, Est-1, Fei-0, Got-7, Ler-1, Lov-5, Nfa-8, Rrs-7, Rrs-10, Shadara, Tamm-2, Ts-1, Tsu-1.

The next processing steps for each sequence alignment were performed as described for the re-sequenced data.

In total, ninety sequence alignments were processed and codons removed, because they contained three variants SNP positions.

Thirty-seven representative gene models were affected by deletions, so that the open reading frame (ORF) within the analyzed sequence alignment was shifted by a number of bp, which is not a multiple of three. These sequences were excluded from further analysis.

Nine representative gene models were affected by SNPs that cause a premature STOP codon and were excluded from further analysis.

Twenty-four representative gene models were affected by an altered splice donor or acceptor site and were excluded from further analysis.

MPICao2010 data

First, the TAIR10 GFF3 file (TAIR10 GFF3; see URL subsection 2.1) was parsed with a custom GFF3 R parser for "protein coding genes", resulting in 27206 hits for the five *Arabidopsis thaliana* chromosomes, without the chloroplast and mitochondrion entries.

For each of the 27206 hits the complete gene positions (five-prime to three-prime, introns, splice variants) were obtained from the TAIR10 GFF3 file. These positions were then used to extract the corresponding complementary DNA (cDNA) sequences from the 80 pseudo-chromosomes. The extracted cDNA sequences for each of the 80 accessions and for each "protein coding gene" were then concatenated to 27206 multiple sequence alignments containing 35176 annotated splice variants.

To account for major or large-effect SNPs, that cause an altered start codon, an altered or preliminary stop codon, a change in the splice donor or acceptor site or lead to an overall coding sequence width of not multiple of three via deletions. The 35176 splice variants were checked with a custom R script.

If a start codon deviated from "ATG", it was marked as a major SNP in the corresponding accessions for this particular splice variant. Any reverted stop codon was marked as a major SNP in the corresponding accessions. If a premature stop codon was found, it was marked in the corresponding accessions for this particular splice variant.

In addition to the common splice donor nucleotide dimer "GT" and splice acceptor dimer "AG", other dimers exist, which can be recognized by the spliceosome (Hiller et al. (2006) and Wang and Brendel (2006)). To account for alternative splice sites, first, the splice donor and splice acceptor sites of introns were counted for 35176 annotated splice variants of Col-0 (TAIR10). To further account for major SNPs affecting splice sites of introns of the 35176 multiple sequence alignments, only nucleotide dimers were kept, which exceed 0.05% occurrence of total splice sites (see Supplementary Table ST2).

For splice donor sites, "AT", "GC" and "GT" fall into this category. An accession was marked to contain a major SNP, if a splice donor site deviates from the three mentioned categories. If the splice acceptor site deviated from "AC" or "AG", the corresponding accession was marked containing a major SNP.

If a deletion would lead to an overall length of a coding sequence, which is not a multiple of three, it was marked as a major SNP in the corresponding accession for this particular splice variant.

Only splice variants defined as TAIR10 representative gene models (TAIR10 representative gene models; see URL subsection 2.1) were considered for further analysis.

Splice variants mentioned in the TAIR9_sequence_edits.txt and TAIR10_sequence_edits.txt (TAIR9 sequence edits, TAIR10 sequence edits; see URL subsection 2.1) were removed from further analysis, because their GFF3 information were not correctly assigned, which would mean manual assignment of exon-intron borders.

All gene models, which coding sequence width is not a multiple of three according to TAIR10 GFF3 file, were removed from further analysis.

Several gene models were excluded from further analysis, because the coding sequence of Col-0 reference contained ambiguous characters (*AT2G01120.2*, *AT2G07981.1* and *AT2G08986.1*) or the start codon of the corresponding Col-0 reference sequence deviated from "ATG" (*AT1G27565.1*, *AT3G09922.1*, *AT3G29255.1*, *AT3G63540.1* and *AT4G18960.1*).

By applying these filters, 18042 out of 27206 representative gene models showed no major SNP in any of the 80 accessions and could be directly used for further analysis.

In addition to these genes, representative gene models were retained, which show a major SNP only in one or two out of 80 accessions. In these cases, the affected accessions were removed from the complete multiple sequence alignment of 80 accessions, resulting in multiple sequence alignments of either 79 or 78 accessions. For 2446 representative gene models affected in one accession, and 1109 representative gene models affected in two accessions, the nucleotide sequences of these affected accessions were excluded from the multiple sequence alignment. After this removing step, 3555 additional representative gene models were used for further analysis. The resulting 21597 representative gene models were further reduced by 272 gene models, which showed no SNP at all in the corresponding multiple sequence alignment. This could be due to strong conservation of these splice variants for the 80 evaluated accessions or might just be due to SNP calling failure in the data of Cao et al. (2011). To be conservative and not to over-interpret these splice variants, they were initially reduced. Among the evaluated auxin network genes, four fall into this category (*AT1G24807.1*, *AT1G24909.1*, *AT1G25083.1* and *AT1G25155.1*) and were excluded from further analysis.

In total, 21325 out of 27206 representative gene models (five-prime to three-prime) were retained and could be used for the within species population genetic analyses. For each of the 21325 multiple sequence alignments containing sequence information of 80 accessions, the CDS or intron sequences were extracted and analyzed separately.

For promoter analysis, multiple sequence alignments were extracted from the pseudo-chromosomes, ranging from the annotated gene start position for all 27206 representative gene models to either 500 bp, 1000 bp or 3000 bp in five-prime direction. Promoter analyses were restricted to promoters corresponding to the 21325 representative gene models.

Due to the generated pseudo-chromosome data out of the data from Cao et al. (2011), 151 auxin network genes could be evaluated on full length CDS. However, 23 had to be excluded because of major SNPs affecting more than two accessions, additional four representative gene models were excluded because they contained no SNPs at all, resulting in 124 representative gene models for the auxin network genes, which were analyzed further. For the auxin signaling group 46 representative

gene models were analyzed, 38 for the auxin synthesis, 17 for the auxin metabolism and 23 for the auxin transport group.

Within species population genetic analyses (Re-sequenced data and Nordborg2005 data)

Nucleotide diversity and neutrality tests

The software DnaSP 5.10.01 (Librado and Rozas, 2009) was used to calculate the vast majority of population genetic parameters for multiple sequence alignments of 19 *A. thaliana* accessions.

The individual gene feature positions were used as input for a multiple domain analysis in DnaSP. By this, different domains were assigned to the sequence alignment and specific regions could be analyzed separately, like the complete sequence alignment (at least 200 sites analyzed), intron sites (at least 100 intron sites analyzed) and CDS (at least 120 codons (360 bp) analyzed).

For the assigned domains the **MultiDomain Analysis** option of DnaSP estimated the number of segregating sites (noS), the nucleotide diversity (π_T , π_I and π_{CDS}) according to equation 10.5 of Nei (1987), the Watterson estimator of θ_{total} , θ_{intron} and θ_{coding} based on noS_T , noS_I and noS_{CDS} and the haplotype diversity (H_d) according to Nei (1987), Tajima's D statistic (D_T , D_I and D_{CDS}) according to Tajima (1989a), Fu and Li's D* statistic (FLD_T , FLD_I and FLD_{CDS}), Fu and Li's F* statistic (FLF_T , FLF_I and FLF_{CDS}) and Fu's Fs statistic (FS_T , FS_I and FS_{CDS}) according to Fu and Li (1993).

The **Recombination** option of DnaSP estimated the minimum number of recombination events (R_{min}) according to Hudson and Kaplan (1985). This parameter indicates the minimum number of recombination events in the history of the sample.

The **Synonymous and Nonsynonymous Substitutions** option of DnaSP estimated the synonymous and nonsynonymous nucleotide diversity (π_{SYN} and π_{NONSYN}) according to Nei (1987) based on the number of synonymous and nonsynonymous segregating sites (noS_{SYN} and noS_{NONSYN}) and the total number of synonymous and total number of nonsynonymous sites ($TnoS_{SYN}$ and $TnoS_{NONSYN}$). From all possible pairwise combinations, the maximum synonymous nucleotide diversity ($\pi_{SYN_{MAX}}$) was extracted for each CDS.

noS_{CDS} , $TnoS_{CDS}$, noS_{SYN} , $TnoS_{SYN}$, noS_{NONSYN} and $TnoS_{NONSYN}$ were used to standardize the number of segregating sites by the corresponding length of total analyzed sites ($\frac{noS}{TnoS}$), resulting in S_{CDS} , S_{SYN} and S_{NONSYN} .

The **Tajima's Test** option of DnaSP was used to estimate Tajima's D test statistic also for synonymous and nonsynonymous sites (D_{SYN} and D_{NONSYN}) according to Tajima (1989a) (equation 38) for testing the hypothesis that all mutations are selectively neutral (Kimura, 1983).

The number of protein variants per codon was estimated with the Biostrings and seqinr R package based on CDS.

The fragments of the Nordborg2005 data were exactly analyzed like the re-sequenced data with DnaSP and due to the random distribution of the Nordborg2005 fragments across the *A. thaliana* genome, they could be used as an empirical null distribution.

The above mentioned population genetic parameters of the re-sequenced auxin network genes were compared with the empirical null distribution data of Nordborg2005 and significant differences were estimated based on the nonparametric Wilcoxon rank-sum test after Benjamini-Hochberg correction for multiple testing (Benjamini and Hochberg, 1995).

If genes from the auxin network genes fall into extreme parts of the empirical null distribution, they were highlighted in the specified Tables.

Site frequency spectrum

To calculate the site frequency spectrum of synonymous and nonsynonymous biallelic SNPs of multiple sequence alignments for the re-sequenced data and Nordborg2005 data, first, the **Synonymous and Nonsynonymous Substitutions** option of DnaSP was used. Using this option, each segregating site in the CDS of each sequence alignment could be assigned to either belonging to the synonymous or the nonsynonymous category. In case of a complex codon, DnaSP assumes the most conservative pathway to assign synonymous and nonsynonymous sites for this complex codon. In a second step, a custom R script was used to calculate the minor allele frequency (MAF), which is defined as the frequency of the rarest allele at each site, for each segregating site.

Difference between the empirical null distribution in site frequency spectra (Nordborg2005 data fragments with more than 120 codons; 336 fragments: 1242 synonymous SNPs, 972 nonsynonymous SNPs) and the site frequency spectra of the auxin network groups (re-sequenced data; 93 fragments: 391 synonymous SNPs, 274 nonsynonymous SNPs) were estimated based on the non-parametric Wilcoxon rank-sum test after Benjamini-Hochberg correction for multiple testing (Benjamini and Hochberg, 1995).

Within species population genetic analyses (MPICao2010 data)

Nucleotide diversity and neutrality tests

The software suite analysis-0.8.1 (in particular POLYDNDS and COMPUTE) using the C++ libsequence (Thornton, 2003) was used to calculate the vast majority of population genetic parameters for multiple sequence alignments of 80 *A. thaliana* accessions. Here, DnaSP was not used to calculate population genetic parameters due to its problem of handling large amounts of sequence information in a high-throughput manner. However, the same parameters as for the re-sequenced data could be calculated with the mentioned software.

The COMPUTE software with the options "-n" (use number of segregating sites), "-b" (only consider biallelic sites) and "-p" (calculate the probabilities of haplotype number and other summary statistics using coalescent simulations) was used to estimate the following population genetic parameters for each CDS of the 21325 representative gene models. π_{CDS} was estimated according to Tajima (1983), θ_{CDS} based on noS_{CDS} was estimated according to Watterson (1975), H_d according to Depaulis and Veuille (1998), D_{CDS} according to Tajima (1989a), FLD_{CDS} and FLF_{CDS} according to Fu and Li (1993) and R_{min} according to Hudson and Kaplan (1985).

In addition, COMPUTE was used to calculate the same population genetic parameters for the three different promoter data sets (500 bp, 1000 bp and 3000 bp) for 21325 representative gene models.

The individual gene feature positions were used as an input for the software POLYDNDS to estimate π_{SYN} , π_{NONSYN} , π_T and π_I according to Tajima (1983) based on noS_{SYN} , noS_{NONSYN} , noS_T and noS_I and $TnoS_{SYN}$, $TnoS_{NONSYN}$, $TnoS_T$ and $TnoS_I$ with the "-A" option. In case of a complex codon, the "-A" option of POLYDNDS provides an approximate treatment of codons and assumes the most conservative pathway to assign synonymous and nonsynonymous sites for this complex codon. From all possible pairwise combinations, $\pi_{SYN_{MAX}}$ was extracted for each CDS.

noS_{CDS} , $TnoS_{CDS}$, noS_{SYN} , $TnoS_{SYN}$, noS_{NONSYN} and $TnoS_{NONSYN}$ were used to standardize the number of segregating sites by the corresponding length of total analyzed sites ($\frac{noS}{TnoS}$), resulting in S_{CDS} , S_{SYN} and S_{NONSYN} .

The number of protein variants per codon were estimated with the Biostrings and seqinr R package based on CDS.

The above mentioned population genetic parameters of 124 auxin network genes were compared with 21201 representative gene models as the empirical null distribution and significant differences were estimated based on the nonparametric Wilcoxon rank-sum test after Benjamini-Hochberg correction for multiple testing (Benjamini and Hochberg, 1995).

If genes from the auxin network genes fell into extreme parts of the empirical null distribution, they were highlighted in the specified Tables.

Site frequency spectrum

To calculate the site frequency spectrum of synonymous, non-synonymous and introns_flanking (introns, five-prime, three-prime) biallelic SNPs of multiple sequence alignments for the MPICao2010 data, first the POLYDNDS software using the C++ libsequence (Thornton, 2003) was used with the option "-P -A" to generate polymorphism tables for each category.

In a second step, the obtained polymorphism tables were parsed with a custom R script, and the MAF was calculated using a sample size of 78 individuals. For multiple sequence alignments, which contained more than 78 accessions, sites were randomly pruned to 78 sites. If a site was turned by this pruning step into a fixed site, it was discarded from further analysis.

This can happen for rare alleles, which are completely culled from a given site by the pruning step, resulting in no SNPs for this site. For the MPICao2010 data consisting of 21201 representative gene models, 2105 synonymous segregating sites, 2561 segregating nonsynonymous sites and 6157 introns_flanking segregating sites were lost by this pruning step. In addition, seven representative gene models had lost all their synonymous segregating sites after this pruning step. Twelve lost all their nonsynonymous segregating sites and one all its introns_flanking segregating sites. Fivehundredfifty-five representative gene models had no synonymous segregating sites at all, 604 had no nonsynonymous segregating sites at all and 1564 had no introns_flanking segregating sites.

For the auxin network genes, 15 synonymous segregating sites, twelve nonsynonymous segregating sites and 51 nonsynonymous segregating sites were lost by the described pruning step. One auxin network gene had lost all its nonsynonymous segregating sites by the pruning step. Two of the 124 representative gene models had no nonsynonymous segregating sites at all.

In total, 20763 (20639 + 124) representative gene models could be analyzed for the synonymous site frequency spectrum, 20706 (20585 + 121) for the nonsynonymous site frequency spectrum and 19760 (19636 + 124) for the introns_flanking site frequency spectrum.

The theoretical expected folded spectrum was calculated with equation (51) from Tajima (1989a) as $G_n(i) = \frac{S(\frac{1}{i} + \frac{1}{n-1})}{\sum_{j=1}^{n-1} \frac{1}{j}}$, where n is the sample size, S is the number of segregating sites and i is the MAF category.

Differences between the empirical null distribution in site frequency spectra (MPICao2010 data; 20639, 20585, 19636 representative gene models: 292692 synonymous SNPs, 267796 nonsynonymous SNPs, 766610 introns_flanking SNPs) and the site frequency spectra of the auxin network groups (MPICao2010 data; 124, 121, 124 representative gene models: 2121 synonymous SNPs, 1524 nonsynonymous SNPs, 5241 introns_flanking SNPs) were estimated based on the non-parametric Wilcoxon rank-sum test after Benjamini-Hochberg correction for multiple testing (Benjamini and Hochberg, 1995).

Between species population genetic analyses (Re-sequenced data and Nordborg2005 data)

Nucleotide divergence analysis

For each of the 93 auxin network genes, which contained no major SNP, the representative protein sequence according to TAIR10 from *A. thaliana* Col-0 accession was downloaded (<http://www.arabidopsis.org>) and a blastp search (Altschul et al., 1990) against all annotated *A. lyrata* sequences (www.phytozome.net) was conducted.

The best blastp hits between *A. thaliana* and *A. lyrata* were retained and used for codon alignment preparation.

Amino acid sequence alignments of the best blastp hits were generated with MAFFT (Katoh et al., 2005) using the L-INS-i option and were then used to produce codon alignments with PAL2NAL (Suyama et al., 2006) and the CDS from *A. thaliana* (<http://www.arabidopsis.org>) and *A. lyrata* (www.phytozome.net).

The codon alignments were used to generate multiple sequence codon alignments of the 19 *A. thaliana* accessions containing the 19 partial coding sequences and the partial coding *A. lyrata* sequence as the outgroup sequence.

The Polymorphism and Divergence option of DnaSP was used to calculate the synonymous nucleotide divergence (K_{SNY}), the nonsynonymous nucleotide divergence (K_{NONSNY}) and the $\frac{K_{NONSNY}}{K_{SNY}}$ ratio (ω) according to Nei (1987).

McDonald-Kreitman test

The McDonald and Kreitman's Test option of DnaSP was used to perform the McDonald-Kreitman test (McDonald and Kreitman, 1991) of neutral evolution on each of the 97 partial codon alignments for the comparison between *A. thaliana* and *A. lyrata*.

If the McDonald-Kreitman test was significant, the corresponding auxin network gene was highlighted in the specified Table.

Between species population genetic analyses (MPICao2010 data)

Nucleotide divergence analysis

The TAIR10 annotated representative protein sequences from the *A. thaliana* Col-0 accession were downloaded (TAIR10 representative protein models; see URL subsection 2.1) and a blastp search against all annotated *A. lyrata*, *Brassica rapa* and *Thellungiella parvula* proteins was conducted (*A. lyrata* protein sequences, *B. rapa* protein sequences, *T. parvula* protein sequences; see URL subsection 2.1).

Only best blastp hits between *A. thaliana* and the other Brassicaceae species with E-values smaller than 1e-10 were retained and further analysed.

Amino acid sequence alignments of the best blastp hits (Col-0 sequence against other Brassicaceae sequence) were generated with MAFFT (Katoh et al., 2005) using the L-INS-i option and were then used to produce codon alignments with PAL2NAL (Suyama et al., 2006) and the CDS from *A. thaliana*, *A. lyrata*, *B. rapa* or *T. parvula* (*A. lyrata* CDS sequences, *A. lyrata* CDS sequences, *B. rapa* CDS sequences, *T. parvula* CDS sequences; see URL subsection 2.1).

The codon alignments were used to generate multiple sequence codon alignments of the 80 *A. thaliana* accessions containing the 80 sequences and one outgroup sequence from either *A. lyrata*, *B. rapa* or *T. parvula*. The multiple sequence alignments were then processed with a custom R script to retain at least 78 *A. thaliana* accession sequences based on the major SNP assignment described above.

For each codon alignment and each *A. thaliana* accession sequence within each codon alignment, the nucleotide divergence against the corresponding Brassicaceae sequence (*A. lyrata*, *B. rapa* and *T. parvula*) was calculated with the software KaKs-Calculator (Zhang et al., 2006) according to Yang and Nielsen 2000. For each codon alignment the average K_{SYN} , K_{NONSYN} and ω were calculated. Codon alignments with $K_{NONSYN} > 0.5$ and $K_{SYN} > 5$ were excluded from further analysis. In addition, K_{NONSYN} values which could not be calculated by the KaKs-Calculator software because of no nonsynonymous SNPs were manually set to 0.

Differences between the empirical null distribution in K_{SYN} , K_{NONSYN} and ω and the auxin network groups were estimated based on the non-parametric Wilcoxon rank-sum test after Benjamini-Hochberg correction for multiple testing (Benjamini and Hochberg, 1995).

If genes from the auxin network genes fell into extreme parts of the empirical null distribution, they were highlighted in the specified Tables.

McDonald-Kreitman test

The MKTEST software using the C++ libsequence (Thornton, 2003) with the "-A" option was used to perform the McDonald-Kreitman test (McDonald and Kreitman, 1991) of neutral evolution on each codon alignment for all three Brassicaceae comparisons. In case of a complex codon, the "-A" option of MKTEST provides an approximate treatment of codons and assumes the most conservative pathway to assign synonymous and nonsynonymous sites for this complex codon. The p-value of the test was computed by a Fisher's exact test and the g-test. All calculated p-values were corrected for multiple testing (Benjamini and Hochberg, 1995).

If genes from the auxin network genes fell into extreme parts of the empirical null distribution, they were highlighted in the specified Tables.

Transcript diversity analysis

The expression data of Delker et al. (2010) were used to calculate the transcript diversity between six *A. thaliana* accessions (Bay-0, Bur-0, C24, Col-0, Fei-0 and Shadara), which were also part of the 19 *A. thaliana* accessions of the re-sequenced data. For the expression data of Delker et al. (2010), each accession was treated in three biological replicates with exogenously applied indole-3-acetic acid. At four different time points (t_0 : 0 minutes, t_{30} : 30 minutes, t_{60} : 60 minutes, t_{180} : 180 minutes) RNA was extracted, further processed and hybridized to the whole genome Affymetrix ATH1-12150 GeneChip. For a detailed description of how the data was generated see Delker et al. (2010). Robust multi-chip average-normalized \log_2 expression levels were obtained for each accession and each array element (Delker et al., 2010).

For each of the 22810 ATH1 array elements, an average expression value was calculated for each accession at a certain auxin treatment time point ($\bar{e}_{(acc,t,ae)}$) according to the formula

$$\bar{e}_{(acc,t,ae)} = \frac{\sum_{r=1}^R e_{(acc,t,ae,r)}}{R} \quad (2.1)$$

, where t is the time point at a certain auxin treatment, R is the number of biological replicates for each accession, ae is the array element, and $e_{(acc,t,r,ae)}$ is the expression for one accession at a certain

auxin treatment time point for a certain biological replicate r , and a certain array element ae .

For each accession the log fold expression change $\Delta\bar{e}_{(acc,t,ae)}$ at a treatment time point t was calculated by subtracting $\bar{e}_{(acc,t_0,ae)}$ from $\bar{e}_{(acc,t_{30},ae)}$, $\bar{e}_{(acc,t_{60},ae)}$ and $\bar{e}_{(acc,t_{180},ae)}$. These three \bar{e} values constitute the expression profile $eprof_{acc,ae}$.

The transcript diversity $TRCDIV_{(acc_i,acc_j,ae)}$ between an accession i and an accession j for an array element ae was calculated as the euclidean distance between the expression profiles $eprof_{(acc_i,ae)}$ and $eprof_{(acc_j,ae)}$ with the formula, where for two accessions

$$TRCDIV_{(acc_i,acc_j,ae)} = \sqrt{\sum_1^t (\bar{e}_{(acc_i,t,ae)} - \bar{e}_{(acc_j,t,ae)})^2} \quad (2.2)$$

, the sum of the quadratic differences for each time point were calculated and the square root is taken.

The transcript diversity for an array element $TRCDIV_{ae}$ was then calculated as the average of all $m = \frac{n*(n-1)}{2}$ (n : number of accessions) possible pairwise transcript diversities $TRCDIV_{(acc_i,acc_j,ae)}$

$$TRCDIV_{ae} = \frac{\sum_{i=2}^n \sum_{j=1}^{i-1} TRCDIV_{(acc_i,acc_j,ae)}}{m} \quad (2.3)$$

, where m is the number of possible pairwise comparisons.

Transcript diversity values $TRCDIV_{ae}$ were only retained, when an array element mapped exclusively to a single annotated *A. thaliana* gene (21192 array elements).

In total, 107 $TRCDIV_{ae}$ values for the auxin network genes could be compared to 21085 $TRCDIV_{ae}$ values as the empirical null distribution. For one auxin network gene, two array elements are on the ATH1 GeneChip ("AT5G62000"). This holds true for 155 genes for the empirical null distribution, for three genes three array elements are on the ATH1-12150 GeneChip, for one gene six array elements are on the ATH1-12150 GeneChip and for one gene seven array elements are on the ATH1-12150 GeneChip.

Differences between the empirical null distribution and the auxin network groups were estimated based on the non-parametric Wilcoxon rank-sum test after Benjamini-Hochberg correction for multiple testing (Benjamini and Hochberg, 1995).

URLs

1001 Genome Project	http://www.1001genomes.org
<i>A. lyrata</i> CDS sequences	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v8.0/Alyrata/annotation/Alyrata_107_cds.fa.gz
<i>A. lyrata</i> protein sequences	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v8.0/Alyrata/annotation/Alyrata_107_peptide.fa.gz
<i>A. thaliana</i> CDS sequences	ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR10_blastsets/TAIR10_cds_20110103_representative_gene_model_updated
analysis	http://molpopgen.org/software/lseqsoftware.html
<i>B. rapa</i> CDS sequences	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v8.0/Brapa/annotation/Brapa_197_cds.fa.gz
<i>B. rapa</i> protein sequences	ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v8.0/Brapa/annotation/Brapa_197_peptide.fa.gz
DnaSP	http://www.ub.edu/dnasp/

KaKs-Calculator	http://code.google.com/p/kaks-calculator
libsequence	http://molpopgen.org/software/libsequence.html
MPICao2010 data	http://1001genomes.org/data/MPI/MPICao2010/releases/current/strains/
Nordborg2005 data	https://cynin.gmi.oeaw.ac.at/home/resources/atpolydb/miscellaneous-data/alignments_042006.tar.z
<i>T. parvula</i> CDS sequences	http://thelungiella.org/blast/TpV840RF.CDS
<i>T. parvula</i> protein sequences	http://thelungiella.org/blast/TpV840RF.protein
TAIR9 sequence edits	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR9_genome_release/TAIR9_sequence_edits.txt
TAIR10 chromosome sequences	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_chromosome_files/TAIR10_chr_all.fas
TAIR10 GFF3	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_gff3/TAIR10_GFF3_genes.gff
TAIR10 representative gene models	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_gene_lists/TAIR10_representative_gene_models
TAIR10 representative protein models	ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR10_blastsets/TAIR10_pep_20110103_representative_gene_model_updated
TAIR10 sequence edits	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_sequence_edits.txt

2.2 Materials and methods for quantitative genetic analysis for two RIL populations

Plant material and growth conditions

A subset of plant lines from two *A. thaliana* RIL populations derived by a reciprocal cross between the two *A. thaliana* accessions Col-0 and C24 (Törjék et al., 2006) was used to map QTLs. For RIL population $Q_{Col-0 \times C24}$, the number of lines used ranged from 167 to 211. For RIL population $R_{C24 \times Col-0}$, the number of lines used ranged from 133 to 193. A core subset of both populations was generated with GGT2 (van Berloo, 2008) and was used for the dose response analyses. The corresponding genetic map and the genotype information are described in Törjék et al. (2006) and Meyer et al. (2010). The subset of NILs, Col-0 NILs in C24 background and C24 NILs in Col-0 background, used to verify QTLs is described in Törjék et al. (2008).

For root growth assays, seeds were surface-sterilized, germinated and cultivated as described in Delker et al. (2010). In brief, seedlings were cultivated vertically on petri-dishes containing control medium for three days (IAA) or five days (2,4-D and NAA). Subsequently, the seedlings were transferred either to control medium or to phytohormone containing medium supplemented with the indicated concentrations. After additional five (IAA) or three days (2,4-D and NAA), the plates were photographed with a Canon EOS 550D digital single lens reflex (DSLR) camera or a Nikon D90 DSLR camera.

For hypocotyl growth assays, seeds were surface-sterilized and germinated vertically on control medium or phytohormone containing medium with the indicated concentrations. To synchronize the germination process, an initial light treatment of 5 hours ($175 \mu Em^{-2} sec^{-1}$) was applied to each plate. After wrapping the petri-dishes with aluminum foil, seedlings were cultivated at 20°C for 5 days and then photographed with a Canon EOS 550D DSLR camera or a Nikon D90 DSLR camera.

The experimental design of the RGI and HGI traits involves the separation of the control and phytohormone treated plants. To keep standard deviation caused by micro-environmental differences low,

so called "plate effects", the petri-dishes containing either the control plants or treated plants were placed directly next to each other in the growth chambers. The ten longest plants per RIL line or *A. thaliana* accessions were used for calculating RGI. For the HGI assays, likewise the ten longest hypocotyls were used for calculating HGI.

Phenotypical data acquisition and statistical analysis

Primary root and hypocotyl lengths were measured with the image processing software RootDetection (described in the results chapter). Growth inhibition GI (either RGI or HGI) for a certain phytohormone concentration j and the n longest roots or hypocotyls was calculated according to the formula

$$GI_j = 1 - \left(\frac{\frac{1}{n} \sum_1^n l_j}{\frac{1}{n} \sum_1^n l_0} \right) \times 100 \quad (2.4)$$

, where l_j are the root or hypocotyl lengths for concentration j and l_0 are the sorted root or hypocotyl lengths for control plant seedlings.

The measured phenotypic data for each growth assay and each RIL population was analyzed in R2.15 to describe the data distribution and to test for non-normality by applying a Shapiro-Wilk test. To account for outliers, the phenotypic mean values of each RIL were first scaled to z-scores with the `scale` function in R. RILs outside the z-score threshold of 3 were excluded from further analysis (Table T3.3). Pearson correlation coefficients between different auxin response traits within each RIL population and the corresponding pairwise p-values were calculated with the `cor.test` function in R.

For each growth assay, three different comparisons were performed on \log_2 transformed phenotypic data sets. First, for each data set an equal amount of plants per line was grouped into treated and untreated groups. (1) To compare the treated groups, one-way analysis of variance (ANOVA, R2.15) was performed. Next, the Tukey post hoc test was conducted to find pairs of accessions or pairs of plant lines among all compared treated groups which are significantly different from each other. (2) The same statistical analysis was performed to find pairs of accessions or pairs of plant lines which are significantly different from each other among untreated groups. (3) To identify the pairs of accessions or pairs of plant lines which significantly differ in their response to a treatment, a two-way ANOVA (R2.15) was conducted in a 2×2 design for all possible comparisons. Subsequently, the obtained P values were corrected for multiple testing using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

Estimation of the broad sense heritability in the RIL populations

The broad sense heritability (H^2) for root length, hypocotyl length, RGI and HGI for both RIL populations was estimated according to the formula

$$H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2) \quad (2.5)$$

, where σ_g^2 is the between genotype variance component and σ_e^2 is the within residual (error) variance component. Using a random-effects model (model 2) one-way ANOVA (R2.15) according to the model $\gamma_{ij} = \mu + g_i + e_{ij}$, the total phenotypic variance for each trait was partitioned and the mean square of between (MSB) groups and the mean square of within (MSW) groups were obtained. With these estimates σ_g^2 was then calculated as $\sigma_g^2 = (MSB - MSW)/n$, where n is the number of replicates and σ_e^2 was defined as $\sigma_e^2 = MSW$.

For the traits root length and hypocotyl length, each measured plant of each line in the RIL populations was considered as a replicate to estimate the broad sense heritability. For the traits RGI and HGI,

which are both calculated as ratios of mean values of several plants between a control and a treatment condition, three replicates were obtained by randomly choosing a subset of plants (3 out of 10) for the control and the treatment condition for three times and calculating the mean values for the sampled subsets (Table T3.3).

QTL analysis in the RIL populations

The two RIL populations $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ were phenotypically analyzed for RGI and HGI after treatment with the endogenous auxin IAA (40nM; 500nM) and the synthetic auxins 2,4-D (20nM; 375nM) and NAA (75nM; 500nM) as indicated above. RGI and HGI were evaluated on the 10 longest roots or hypocotyls and calculated as described above. The so defined phenotypic RIL mean values were used as traits for the QTL mapping in both RIL populations. Prior to the QTL analysis, the phenotypic data were checked for normal distribution and outliers were excluded from further analysis.

The QTL mapping software R/qtl (Broman et al., 2003) implemented as the qtl package in R was used for the QTL analyses of both RIL populations. The following steps were applied to identify QTLs:

1. *Main effect QTLs*: To detect *main effect* (single) QTLs, which contribute to the variation in the phenotypic data, interval mapping and composite interval mapping was used with the Haley-Knott regression (Haley and Knott, 1992). When the phenotypic data was not normally distributed for one trait, the non-parametric model of the `scanone` function was also applied according to Kruglyak and Lander (1995). The conditional genotype probabilities were calculated using the `calc.genoprob` function with a step size of 1cM and an assumed genotyping error probability of 0.001 using the Kosambi map function (Kosambi, 1943). For composite interval mapping (Zeng, 1994), the `cim` function was used with three covariates selected by a forward approach and a window size of 10cM. For each trait and mapping method a genome-wide logarithm of odds (LOD) score threshold was estimated by 1000 permutations to correct for type I error rates of $\alpha = 0.05$ as suggested by Churchill and Doerge (1994). If a QTL LOD score, which under a single QTL model is defined as the difference between the likelihood that there is a QTL and there is no QTL (see Broman and Sen (2009) pages 75-78), crosses the LOD score threshold it was declared as significant (see Supplementary Figure S16, Supplementary Figure S17 and Supplementary Table ST10).
2. *Two-QTL genome scans*: To find potential interactions among QTLs the `scantwo` function with the Haley-Knott regression (Haley and Knott, 1992) method was applied on each trait and RIL population by applying the same parameters as used with the `scanone` function. Genome-wide significant LOD score thresholds for the six different LOD scores, which are calculated between two chromosomal positions j and k by the `scantwo` function ($M_{full}(j, k)$, $M_{fv1}(j, k)$, $M_{int}(j, k)$, $M_{add}(j, k)$, $M_{av1}(j, k)$ and $M_{one}(j, k)$), were estimated for each trait by 10000 permutations (Supplementary Table ST11). A pair of chromosomal positions (j, k) is reported as interesting when either of the following holds true as described with equation 8.3 in Broman and Sen (2009):

$$\begin{aligned} M_{full}(j, k) \geq T_{full} \text{ and } [M_{fv1}(j, k) \geq T_{fv1} \text{ or } M_{int}(j, k) \geq T_{int}] \\ M_{add}(j, k) \geq T_{add} \text{ and } [M_{av1}(j, k) \geq T_{av1}] \end{aligned} \quad (2.6)$$

Next, for evaluating additive or interactive effects the estimated LOD score thresholds were further used to calculate "heavy" and "light" penalties with the `calc.penalties` function described by Manichaikul et al. (2009), which are necessary for the `stepwiseqtl` function to identify multiple QTL models (see Broman and Sen (2009) pages 273-280).

3. *Multiple QTL models*: For each trait and RIL population the `stepwiseqtl` function was used to identify multiple QTL models in a forward/backward selection approach. In brief, in the forward approach the algorithm creates (or takes) an initial QTL model and uses the precalculated

penalties to subsequently add new QTLs and QTL interactions up to a predefined maximum number of QTL. After each step the QTL positions are refined and in the backward approach *main effect* QTLs are dropped from the current model until the null model is reached (for a detailed description see Broman and Sen (2009) pages 273-280 and documentation of the `stepwiseqtl` function in R). Here, the LOD peak values found with the interval mapping and Haley-Knott regression (Haley and Knott, 1992) method (see Supplementary Table ST10) of the *main effect* QTL analysis was used as an initial model and multiple QTL models were analyzed up to a maximum of 10 QTLs (Table T3.5).

4. *QTL effect and LOD support interval estimation*: The multiple QTL models for each trait and RIL population, defined with the `stepwiseqtl` function, were finally used with the `makeqtl` and `fitqtl` functions to explore the estimated QTL effects and % variance explained by a QTL. The LOD support interval for each QTL in the multiple QTL model was evaluated as the 95% Bayes credible intervals with the `bayesint` function and the lowest and highest value indicated in Table T3.5.

Genome-wide interactions as defined by the two-QTL genome scans were combined with the final multiple QTL models for each trait and plotted with Circos (Krzywinski et al., 2009). Most significant interactions were evaluated with the `effectplot` and `plot.png` function of the R/qtl software (Broman et al., 2003).

Analysis of *a priori* auxin related candidate genes

95% CIs of *main effect QTLs* detected by the CIM method and their corresponding flanking markers were used as borders to search for auxin related candidate genes (CGs) in a CG list prepared as described in the following materials and methods section. The detected auxin related CGs were subsequently screened for synonymous or nonsynonymous SNPs between the *A. thaliana* accessions Col-0 and C24 in the data of Clark et al. (2007).

Root length assignment in RootDetection

The **RootDetection** software is described in the results chapter. Here, the equation, how **RootDetection** defines root length, is described in more detail.

The root length $l(G)$ is determined by calculating the distances between the weighted mean positions of each iteration step. The weighed mean positions $P(g)$ of the particles $(x_1, y_1), \dots, (x_{n_g}, y_{n_g})$ in each generation (iteration step) $g = 1, \dots, G$ are defined as:

$$P(g) = \frac{\sum_{i=1}^{n_g} \omega_i (x_i, y_i)}{\sum_{i=1}^{n_g} \omega_i} \quad (2.7)$$

The euclidean norm for $v = (v_1, v_2) \in \mathbb{R}^2$ is defined as:

$$\|v\|_2 = \sqrt{v_1^2 + v_2^2} \quad (2.8)$$

With this notation the root length $l(G)$ is defined as:

$$l(G) = \sum_{g=1}^{G-1} \|P(g) - P(g+1)\|_2 \quad (2.9)$$

Accession numbers

A. thaliana accessions were obtained from the Arabidopsis Nottingham Stock Center. The individual stock identities are indicated in table T3.6. *A. thaliana* seeds of the RILs and NILs were obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany, kindly provided by Professor Dr. Thomas Altmann.

URLs

R	http://www.r-project.org
R/QTL	http://www.rqt1.org
RootDetection	http://www.labutils.de

2.3 Materials and methods for quantitative genetic analysis for 80 *A. thaliana* accessions

Plant material and growth conditions

To explore the diversity within the *A. thaliana* species, eighty *A. thaliana* accessions out of 1307 genotyped accessions (Horton et al., 2012) were used to conduct genome-wide association studies (GWAS).

Root and hypocotyl growth assays were performed like described in section 2.2.

Phenotypical data acquisition and statistical analysis

Primary root and hypocotyl lengths were measured with the image processing software **RootDetection** described earlier. Phenotypic data analysis was performed like described in the previous section.

Population structure analysis

The following steps were applied to infer population structure for a subselection of 80 out of 1307 *A. thaliana* accessions (Horton et al., 2012):

1. *Marker data*: Marker data were downloaded from the Gregor Mendel Institute web site (1307 *A. thaliana* accessions marker data; see URL subsection 2.3).
2. *Data conversion and subsetting*: The data was converted into PLINK (Purcell et al., 2007) compatible files containing 214051 SNPs by a custom *python* script. Further work was performed either on the complete data (**1307 SNP data**) for population assignment or on a reduced subset (**80 SNP data**) for the association mapping.
3. *Minor allele frequency*: To correct for spurious low p-value enrichment at rare alleles when conducting an efficient mixed-model association (EMMA) and to keep only non-singelton SNPs, using the `--maf` option of PLINK, only SNPs were retained with a minor allele frequency (MAF)

above 0.1 as suggested by Kang et al. (2008) and Atwell et al. (2010). Applying these filter results in either 173220 SNPs (**1307 MAF01 SNP data**) for the **1307 SNP data** or 175655 SNPs (**80 MAF01 SNP data**) for the **80 SNP data**.

4. *SNP pruning*: To calculate the **kinship matrix (K-matrix)** the **1307 MAF01 SNP data** was further pruned with the `--indep-pairwise` option of PLINK with a sliding window size of 50, a sliding window step of 5 and an r^2 threshold of 0.2 to exclude SNPs in high linkage disequilibrium. By applying this pruning step 35733 SNPs were retained to calculate the K-matrix. For the **PCA matrix (P-matrix)** the r^2 threshold of 0.8 as used in Horton et al. (2012) was applied on the **1307 SNP data** retaining 165984 SNPs.
5. *Kinship matrix (K-matrix)*: The **K-matrix** describes the pairwise relatedness of individuals based on their genotypes. It is an $n \times n$ square matrix where n is the number of compared *A. thaliana* accessions and can be used to correct for population structure. It was calculated with the identity by state (IBS) method as described in Kang et al. (2008) on the pruned **1307 MAF01 SNP data** containing 35733 SNPs with the `snpGdsIBS` function of the SNPRelate R package (Zheng et al., 2012) according to equation 2.10:

$$K_{i,j} = K_{j,i} = \frac{1}{m} \left(\sum_1^m ((X_i) \times (X_j)) + ((1 - X_i) \times (1 - X_j)) \right) \quad (2.10)$$

For each pairwise comparison between individual i and j the **kinship** $K_{i,j} = K_{j,i}$ is calculated on an $n \times m$ SNP matrix, X , where n is the number of individuals and m is the number of SNPs. X_i is a vector of SNPs m for an individual i at a position n with SNPs decoded as 0 or 1. The pairwise **kinship** is then defined as the sum of all shared SNPs between individual i and j divided by the sum of all SNPs evaluated.

The resulting **K-matrix** was further reduced to the 80 screened *A. thaliana* accessions and converted to a distance matrix by simply subtracting each pairwise relatedness value from 1 to get a pairwise distance matrix. The pairwise distance matrix was used to conduct a UPGMA clustering of the 80 genotypes with the `hclust` function in R (see Figure F3.23).

6. *PCA matrix (P-matrix)*: The `snpGdsPCA` function of the SNPRelate R package (Zheng et al., 2012) was used to perform a principle component analysis (PCA) on the pruned **1307 SNP data** to obtain PCA loadings for the 1307 *A. thaliana* accessions. The PCA loadings were then reduced to the 80 phenotyped accessions and different numbers of top PCA loadings were further used with the general linear model (GLM) and mixed linear model (MLM) to correct for population structure as described in Price et al. (2006); Patterson et al. (2006).

Wilcoxon rank-sum test (WILCOXON)

As a non-parametric genome-wide association mapping method, which does not account for population structure, the Wilcoxon rank-sum test was applied for each trait on each marker. After excluding *A. thaliana* accessions for which no phenotypic data was available for the evaluated trait, the remaining accessions were grouped according to their alleles and the non-parametric test was conducted only on biallelic sites of the reduced **80 SNP data**. The expected p-value distribution for the non-parametric test was estimated using a permutation test as outlined in Atwell et al. (2010).

General linear model (GLM) and mixed linear model (MLM)

To take the population structure into account, the general linear model (GLM) and mixed linear model (MLM) implemented in the software TASSEL (Bradbury et al., 2007) were used for association testing. Population structure can be incorporated in both models with a matrix which describes the population structure of the analyzed individuals. Here, the population structure matrix was generated via PCA analysis as in DeRose-Wilson and Gaut (2011) (see also **P-matrix**).

Association between SNPs and phenotypes is tested with the GLM as a fixed effects linear model and population structure is taken into account in this model as main effects as covariates (for a detailed description see Buckler et al. (2011) pages 30-34).

The MLM approach implemented in TASSEL (Bradbury et al., 2007) includes both fixed and random effects and is described in detail in the online methods of Zhang et al. (2010). In brief, as described in Buckler et al. (2011) on pages 30-34, a genetic marker based kinship matrix (see also **K-matrix**) can be combined with a population structure matrix (see also **P-matrix**), which was previously shown to improve statistical power compared to a model only considering the population structure matrix (Yu et al., 2006). In addition to the standard MLM approach further improvements for statistical power and computing time reduction like the EMMA algorithm (Kang et al., 2008) and the population parameters previously determined (P3D) option (Zhang et al., 2010) are implemented in the TASSEL software (Bradbury et al., 2007). Zhang et al. (2010) also introduced a method which reduces the dimensionality of the **K-matrix** and should improve model fitting, called "compression" and can be used via the `compression level` option of the MLM approach (a detailed description can be found in the online methods of Zhang et al. (2010)).

To determine an appropriate model for the analyzed auxin response traits, different top principle components (P3, P6, P9) representing the population structure were combined with either GLM or MLM approach. If the MLM approach, for which a **K-matrix** is mandatory (see also **K-matrix**), was applied on the phenotypic data of the different auxin response traits, also combinations of the **P-matrix** (P3, P6, P9) and the P3D option and combinations of the **P-matrix** (P3, P6, P9) and the `compression level` option were used. The observed and expected genome-wide p-values were then used to generate quantile-quantile plots (QQ-plots) to choose appropriate models for association testing.

Variance-heterogeneity GWAS (vGWAS)

For each trait a genome-wide test for variance controlling genes was conducted as described in Shen et al. (2012) by using the **80 MAF01 SNP data** as the genome matrix with the R package vGWAS.

Manhattan plots and QQ-plots were generated with a custom R script in the style of the GAPIT R package (Lipka et al., 2012). SNP structure maps and LD plots were made with Genome Variation Server 137 (Genome Variation Server 137; see URL subsection 2.3).

Analysis of *a priori* candidate genes

A list of *a priori* candidate genes (CG)s was built by searching for genes with annotations related to auxin as follows:

1. *Data retrieval*: Gene aliases data, functional description of genes and gene ontology (GO) annotations were obtained from The Arabidopsis Information Resource (TAIR) webpage (<http://www.arabidopsis.org/>) (TAIR10 gene aliases data, TAIR10 functional description of genes, TAIR GO annotations; see URL subsection 2.3).

2. *Search for auxin related genes*: To build a list of auxin related *a priori* CGs the gene aliases data and functional description data was parsed for the keywords "Auxin", "auxin", "IAA" and "indole-3-acetic acid". The GO data was screened for auxin related GO terms ("GO:0010315", "GO:0060919", "GO:0010011", "GO:0060918", "GO:0010252", "GO:0009926", "GO:0009850", "GO:0009852", "GO:0010540", "GO:0010541", "GO:0009733", "GO:0009851", "GO:0009721", "GO:0009921", "GO:0080162", "GO:0009734", "GO:0009672", "GO:0010249", "GO:0071365", "GO:2000012", "GO:0090354", "GO:0010600", "GO:0080161", "GO:0010928", "GO:0090355", "GO:0090356", "GO:0010328", "GO:0010329", "GO:0010601", "GO:1901703", "GO:0010929", "GO:0010930", "GO:0060774", "GO:0090015", "GO:0044032", "GO:0010013", "GO:0080024", "GO:0010178"). Genes identified for the three different data sets were combined to one *a priori* candidate list containing 642 genes.
3. *Gene positions*: For each of the 642 genes the start and end positions were retrieved from the TAIR10 GFF3 file (TAIR10 GFF3; see URL subsection 2.3).

For all auxin response traits and the conducted GWA mapping methods (MLM, WILCOXON and vGWAS) the highest 50 ranked associated SNPs were checked if they were located within 10 kb (MLM and WILCOXON) or 20 kb (vGWAS) of one of the CGs or fall into 95% CI of the found QTL regions as described in Brachi et al. (2010). Clark et al. (2007) showed that applying a 20 kb window is conservative given that linkage disequilibrium in *A. thaliana* decays per 10 kb on average (Brachi et al., 2010). All *a priori* genes fulfilling these conditions were listed and SNP rank positions are indicated in Table T3.8 (MLM + WILCOXON) and Table T3.9 (vGWAS). To test whether "true" associations were detected by GWA mapping, assuming overrepresentation of top-ranked SNPs (Brachi et al., 2010), enrichment ratios of predefined top-ranked SNP categories were calculated. N defines the population size of SNPs tested for association (175655 for MLM and 214051 for WILCOXON) and $CSNP_d$ is the set of candidate SNPs at maximum distance d away from CGs. X is a $N \times 2$ matrix with analyzed SNPs (SNP_i) and their corresponding p-values (P_i), ordered by increasing p-values. The enrichment ratio of *a priori* candidate SNPs ($E_{CSNP_d,n}$) at a defined threshold of top ranked SNPs n is then defined by the formula:

$$E_{CSNP_d,n} = \left(\frac{\sum_{i=1}^n \mathbf{1}_{\mathbf{x}_i}}{|CSNP|} \right) \times \frac{N}{n} \quad (2.11)$$

$$\mathbf{1}_{\mathbf{x}_i} := \begin{cases} 1, & \text{if } SNP_i \in CSNP_d \\ 0, & \text{if } SNP_i \notin CSNP_d \end{cases}$$

Enrichment ratios according to equation 2.11 were calculated for different top ranked SNP thresholds (50, 100, 200, 300, 400, 500, 1000, 2000 and 3000) and are given in Supplementary Table ST16.

Geographical distance analysis

To assess potential correlations between phenotypic trait values and geographical origin of the analyzed *A. thaliana* accessions, geographical distance analyses was performed as described in Zuther et al. (2012). For each possible accession pair the geographic distance was calculated with the `distance` function of the R package `SDMTools` according to Vincenty (1975). Pearson correlations between phenotypic trait values, pairwise *A. thaliana* relatedness and geographic distance were calculated with R. The pairwise genetic relatedness was calculated as described earlier (see **K-matrix** equation 2.10). For all geographic distance analyses all *A. thaliana* accessions were excluded which fall into the "RED" category or are missing according to Anastasio et al. (2011). These accessions have most likely an uncertain origin and are indicated in table T3.6.

Accession numbers

Arabidopsis thaliana accessions were obtained from the Arabidopsis Nottingham Stock Center. The individual stock identities are indicated in table T3.6. *A. thaliana* seeds of the RILs and NILs were obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany, kindly provided by Professor Dr. Thomas Altmann.

URLs

1307 <i>A. thaliana</i> accessions marker data	https://cynin.gmi.oeaw.ac.at/home/resources/atpolydb/250k-snp-data/call_method_75.tar.gz
Genome Variation Server 137	http://gvs.gs.washington.edu/GVS137
TAIR10 functional description of genes	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_functional_descriptions
TAIR10 gene aliases data	ftp://ftp.arabidopsis.org/home/tair/Genes/gene_aliases.20120207.txt
TAIR10 GFF3	ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_gff3/TAIR10_GFF3_genes.gff
TAIR10 GO annotations	ftp://ftp.arabidopsis.org/home/tair/Ontologies/Gene_Ontology/ATH_GO_GOSLIM.txt
TASSEL and GAPIT	http://www.maizegenetics.net

Results

3.1 A population genetic analysis of genes regulating the auxin network

To uncover patterns of selection, which might contribute to natural variation observed for auxin response traits, a population genetic analysis of auxin network genes was conducted in this thesis.

In the beginning of the project, partial genomic fragments of 1 kb length for 49 auxin signaling genes in 18 *A. thaliana* accessions were resequenced. The 49 auxin signaling genes were restricted to the *TIR1/AFB*, *AUX/IAA* and *ARF* gene families. The 18 accessions used, are part of a set of 96 accessions, which have been analysed by Nordborg et al. (2005). Nordborg et al. (2005) sequenced 1214 randomly distributed genomic fragments in 96 accessions and showed a tremendous degree of population structure within *A. thaliana*. Despite the fact that *A. thaliana* is a highly selfing organism ($\sim 97\%$, Platt et al. (2010)), the pattern of polymorphism indicated values expected for a widely distributed, sexually reproducing species (Nordborg et al., 2005). Another very important finding was, that the analyzed population genetic parameters did not follow a distribution expected under the neutral model of molecular evolution. Even simple demographic models could not be applied to correct for the found patterns, suggesting a strong effect of population structure and demographic effects, which have influenced the pattern of polymorphisms within *A. thaliana* during evolution.

Bottlenecks, extinction/recolonization events and population structure, which can be found as demographic footprints among *A. thaliana* accessions (Nordborg et al. (2005), Schmid et al. (2006), Platt et al. (2010) and Cao et al. (2011)) can complicate the identification of adaptive footprints at the DNA level (Wright and Gaut (2005), Thornton et al. (2007), Pavlidis et al. (2008), Pavlidis et al. (2010) and Li et al. (2012)) as outlined by Puerma and Aguadé (2013).

To overcome this problem, the data produced by Nordborg et al. (2005) can be used as an empirical null distribution for *A. thaliana* and genomic sequences of interest, when the same set of accessions are used in a population genetic study, as already demonstrated by Bakker et al. (2006), Bakker et al. (2008), Ramos-Onsins et al. (2008), Sterken et al. (2009) and Puerma and Aguadé (2013).

The 18 accessions are also part of 20 diverse *A. thaliana* accessions, which have been intensively studied by Clark et al. (2007). In this study, a microarray approach was used to resequence the accessions on a genome-wide scale to examine sequence variation in *A. thaliana*. It could be shown that $\sim 10\%$ of all annotated protein coding genes were affected by large-effect SNPs, also called major SNPs, which have an effect on gene integrity (Clark et al., 2007). The analysis of individual gene families revealed an over-representation of major SNPs in nucleotide-binding leucine-rich repeat (*NB-LRR*) genes and the F-box superfamily. Families involved in transcriptional regulation harbored only few major SNPs (Clark et al., 2007). However, due to the microarray approach, very closely linked SNPs could not be called. Here, I decided to re-sequence auxin network genes in 18 *A. thaliana* accessions by Sanger-sequencing to overcome this restriction of missing SNP calls.

After sequencing of partial genomic fragments, multiple sequence alignments of partial coding sequences in 19 accessions (including Col-0 reference) were used to infer sequence diversity within the auxin signaling gene families, published in Delker et al. (2010). Although we only sequenced partial coding sequences, we were able to assess variation in the auxin signaling gene families.

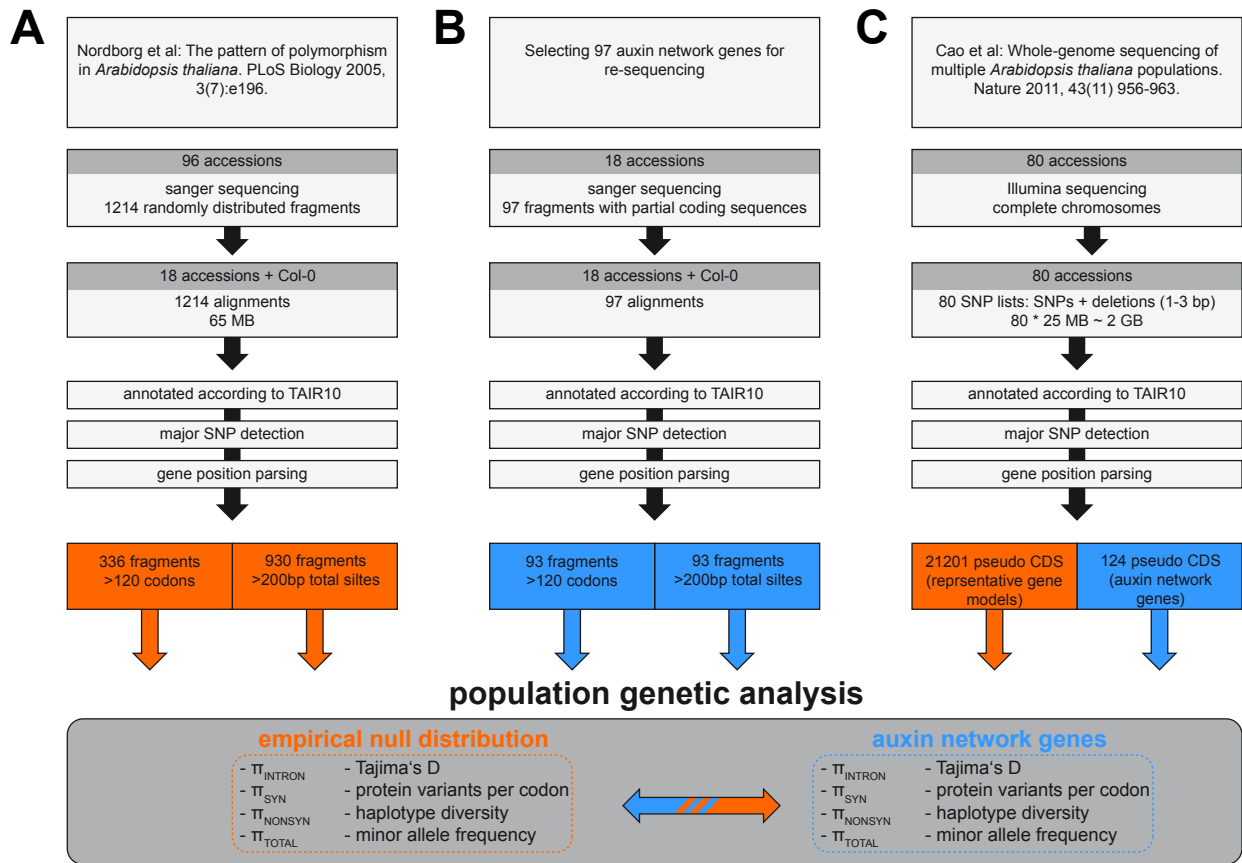


Figure F3.1: Overview of the population genetic analyses workflow. (A) Nordborg2005 data, (B) re-sequenced data and (C) MPICao2010 data.

To address further the question of sequence diversity among auxin network genes, additional partial genomic sequences were resequenced in the same 18 accessions for 21 auxin synthesis genes, 12 auxin metabolism genes and 15 auxin transport genes. The sequence data for these additional 48 auxin network genes were obtained from Ettingshausen (2010) and were processed like the 49 auxin signaling genes (see Figure F3.1 B).

In total, 97 auxin network genes with an average of ~ 824 -bp length containing 612-bp partial coding sites (further denoted as re-sequenced data) were compared to the publicly available data set (Nordborg et al., 2005). The 1214 partial genomic sequences (further denoted as Nordborg2005 data) were first reduced to the same 19 accessions and were then, as well as the 97 auxin network genes, annotated to TAIR10 and further processed to calculate population genetic parameters on coding as well as non-coding sites (see Figure F3.1 A + B).

To support and broaden the findings based on the partial sequence set in 19 accessions, I used a publicly available data set, which was published during the project, and which consists of nearly complete genome sequences of 80 *A. thaliana* accessions (Cao et al., 2011). This data set gives me the opportunity to calculate population genetic parameters on all TAIR10 annotated protein coding sequences in these 80 accessions. In this data set (further denoted as MPICao2010 data), due to next-generation sequencing method, SNP information for closely linked SNPs was not lost (see Figure F3.1 C). Not all 18 *A. thaliana* accessions used for the partial genomic resequencing, were represented among the 80 accessions. As a consequence, in the following result sections, the population genetic analysis for the 18 accessions plus Col-0 partial genomic fragments and the population genetic meta-analysis for the 80 *A. thaliana* accessions are presented.

3.1.1 Major SNPs detected in the re-sequenced data and MPICao2010 data for auxin network genes

Before calculating population genetic parameters for the re-sequenced data and MPICao2010 data, both data sets were pre-processed.

For the re-sequenced data, only genomic fragments were included, which had more than 400 total sites in at least 17 out of 19 accessions, and no major SNP as defined in the corresponding materials and methods section. To obtain a comparable empirical null distribution data set, the same filtering rules were also applied for the Nordborg2005 data.

For the MPICao2010 data, only the filtered variant SNP tables of this data set, containing SNPs and small deletions of up to three bp, were used and mapped on the TAIR10 Col-0 reference genome sequence. The resulting pseudochromosomes and 27206 representative gene models (further denoted as RGMs), which code for protein coding genes, were filtered according to major SNPs. All those RGMs (five-prime to three-prime) were excluded, for which major SNPs occurred in more than 2 accessions. All multiple sequence alignments, which showed no SNPs at all, most likely due to high repetitive regions in the Col-0 reference sequence, for which no base call was made among the filtered variant single nucleotide polymorphisms table, were subtracted from the RGMs. In total the primary 27206 available RGMs were reduced to 21325 RGMs. For each of the 21325 retained RGMs also 500 bp, 1000 bp and 3000 bp upstream of the TAIR10 annotated gene start position were extracted from the pseudochromosomes and analyzed for basic population genetic parameters.

According to the available literature, 151 genes were assigned to one of the following auxin network groups: auxin synthesis, auxin signaling, auxin metabolism and auxin transport. Not restricted to certain genomic fragments like for the re-sequenced data, even the recently detected auxin transport family, the PILSs (Barbez et al., 2012), could be integrated into the analysis of the MPICao2010 data.

After filtering for major SNPs, 93 auxin network genes were retained for the re-sequenced data and 124 for the MPICao2010 data. For all auxin network genes, which contained major SNPs, all annotated splice variants were further tested for major SNPs to see, if functional splice variants might exist in addition to the assigned representative gene model splice variant in the affected accessions.

The results of the major SNP analysis for the re-sequenced data and the MPICao2010 data are given in Supplementary Tables ST3 and ST4. Because there were many genes affected among the auxin network genes considering the MPICao2010 data, not every single gene can be discussed in this thesis. Among the auxin signaling components, the *AUXIN/INDOLE-3-ACETIC ACID INDUCIBLEs* (*AUX/IAAs*) (*IAA2*, *IAA6*, *IAA8*, *IAA11*, *IAA18*, *IAA26* and *IAA34*) and *AUXIN RESPONSE FACTORs* (*ARFs*) (*ARF1*, *ARF7*, *ARF12*, *ARF13*, *ARF14*, *ARF15*, *ARF16*, *ARF20*, *ARF21*, *ARF22* and *ARF23*) were affected.

Consistent within both data sets, *IAA11* was affected by major SNPs in the majority of accessions in the representative gene model *AT4G28640.2*. Seventeen accessions in the re-sequenced data and 72 accessions in the MPICao2010 data were affected. However, other annotated splice variants were not affected at all and might take over the function of the mis-functional annotated representative gene model in these accessions. Here it might be questionable, if *AT4G28640.2* should be considered as the representative gene model. Due to the data analysis workflow, *IAA11* was removed from further analyses.

IAA8 was also removed from further analyses, but again other splice variants were not affected. As six accessions were affected by a major SNP *IAA34* was also removed. For all the other *AUX/IAAs* only a minor fraction of accessions was affected by major SNPs and could be retained for further analyses.

Consistent within both data sets, *ARF13* was affected by major SNPs in the representative gene model *AT1G34170.3*. Three accessions in the re-sequenced data among all annotated splice variants

of *ARF13* and almost one third of all accessions in the MPICao2010 data among all annotated splice variants of *ARF13* were affected; it was therefore omitted from further analyses.

Except *ARF12*, all other members of the subgroup of ARFs originated from a recent tandem duplication (Class I' ARFs by Remington et al. (2004), *ARF12*, *ARF13*, *ARF14*, *ARF15*, *ARF20*, *ARF21*, *ARF22* and *ARF23*) were affected in more than two accessions and were removed from further analyses. Rademacher et al. (2011) showed, that Class I' ARFs are exclusively expressed in the endosperm. Hence, functional redundancy might compensate the loss of one functional ARF in the endosperm.

Apart from Class I' ARFs, *ARF1*, *ARF7* and *ARF16* were affected by major SNPs. *ARF1* was affected not in the representative gene model and could be retained for further analyses.

In addition to the auxin signaling group, also genes from the auxin synthesis group like *ASA1*, *ASB1*, *YUC5*, *YUC10* and *UGTB74B1* were excluded from further analyses, because more than two accession were effected by major SNPs. Also other YUCCAs, like *YUC4* and *YUC7* were affected by major SNPs, but only in a minor fraction of the accessions. With *IGS2*, another gene involved in the tryptophan dependent IAA synthesis was affected, again only in minor fractions of the investigated accessions.

For the auxin metabolism group, *GH3.6*, *ILL5*, *ILR2* and *UGTB84B1* were excluded from further analyses due to the presence of major SNPs. *GH3.2*, *GH3.4* and *ILL2* could be retained because only a minor fraction of accessions were affected. For the auxin transport group, *PIN4* and *PGP2* were excluded from further analyses. *LAX2*, *PILS1*, *PILS2*, *PIN7* and *PIN8* could be retained because only a minor fraction of accessions was affected. For *PIN4* and the representative gene model, three accessions were affected. However, other splice variants were not affected, so that they might take over the function of the representative gene model in these accessions. If one considers the important role of this auxin transporter in root patterning (Friml et al., 2002), this is the most likely scenario. *PGP2* was affected in 58 accessions showing an altered stop codon. However, the deletion in these 58 accessions at the last bp of the stop codon "TAA" resulted due to a following "A" in no alteration of the gene model. Due to the analysis pipeline, it was reduced from the population genetic analysis. However, a loss of gene function is not expected for these 58 accessions.

Considering major SNPs influencing gene integrity of auxin network genes, it seems that for any auxin network group, genes were affected. As Cao et al. (2011) showed previously, genes that were affected in more than one accession often had sustained independent drastic mutations. A compensation effect of other splice variants for effected accessions can not be ruled out as well as compensational effects of redundant family members. However, it is likewise possible, that the identified major SNPs might have an influence on the phenotypic variation.

The remaining representative gene models could be used to calculate population genetic parameters to infer possible hints for selection patterns among the remaining auxin network genes. All pre-processing steps were necessary to reject wrong CDS assignments, which would have a strong influence on the calculated parameters and would lead to false estimates for the auxin network genes as well as the empirical null distribution.

In the following subsections, individual population genetic parameters will be discussed and the results presented, always comparing auxin network gene groups against the empirical null distribution. The empirical null distribution for the re-sequenced data consists of the Nordborg2005 data. For the MPICao2010 data, 21201 representative gene models, which were not assigned to one of the auxin network groups, were used as the empirical null distribution. It is assumed that the majority of loci of the empirical null distribution evolved under nonadaptive forces, with only a small fraction of loci having evolved under adaptive selection as outlined by Puerma and Aguadé (2013). The latter loci would show differential levels and patterns of variation as compared to those in the rest of the genome, and statistics summarizing different aspects of variation at these loci would be found in the empirical distribution tails (Li et al. (2012) and Puerma and Aguadé (2013)).

3.1.2 Nucleotide variation highlighted by different population genetic parameters

Due to the sheer number of genes, it is not possible to present here the output as a table for all 21325 representative gene models for the MPICao2010 data, nor the output as one large table for the re-sequenced data for each individual representative gene model. This output includes suitable information, not only regarding the auxin network genes, but any kind of gene family one might be interested in. These output tables will be available on the attached media and can be obtained upon request.

In this thesis, I will focus on the auxin network gene groups and will discuss individually extreme outliers in an extra subsection. All four mentioned auxin network gene groups, auxin synthesis, auxin signaling, auxin metabolism and auxin transport were evaluated. However, in this thesis a focus was made on the auxin signaling.

As already mentioned, the MPICao2010 data consists of complete coding sequence information, whereas the re-sequenced data only considers partial coding sequence information. Here, the focus was made on the MPICao2010 data, which in addition also has more accessions included. Generally, the MPICao2010 data gives more reliable results. However, the results, which will be highlighted for the MPICao2010 data, are also illustrated for the re-sequenced data in the Supplementary Figures S7, S8, S9, S10 and S11. One exception is made for the first global population genetic parameter, nucleotide diversity, for which all auxin network genes were treated as one group. Here, the results of both data sets will be illustrated directly next to each other.

Nucleotide diversity between auxin network genes as one group and the empirical null distribution

First, the nucleotide diversity π , which is calculated based on nucleotide differences at certain nucleotide categories, is illustrated. Each category, intron segregating sites, coding segregating sites, synonymous segregating sites, nonsynonymous segregating sites and total segregating sites, is compared between the auxin network genes, analyzed as one group, and the corresponding empirical null distribution. As mentioned above, here the results for the re-sequenced data and the MPICao2010 data are illustrated directly next to each other.

While synonymous mutations do not result in an amino acid change, a nonsynonymous mutation causes an amino acid change and can therefore have direct influence on protein structure and function.

For the re-sequenced data and the corresponding empirical null distribution, which was calculated based on the Nordborg2005 data, not all partial genomic fragments harbored all different site classes. This results in different numbers of fragments, which could be analyzed for a certain nucleotide category and are given in Supplementary Table ST5.

As given in Figure F3.2 A, there was no significant difference between the partial 93 auxin network fragments and the corresponding empirical null distribution, if the auxin network genes were treated as one group.

For the MPICao2010 data, π_{NONSYN} is significantly different from the empirical null distribution. The median π_{NONSYN} of the auxin network genes, treated as one group, is 0.00087 and differs from the median of 0.00104 for the empirical null distribution (see Figure F3.2 B). The other evaluated site classes were not different from the empirical null distribution. Also no differences were detected for the promoter regions, defined as 500 kb, 1000 kb and 3000 kb from the RGM gene start position.

Consistent with previous findings (Sterken et al., 2009), lower median values were observed for π_{CDS} compared to π_I , indicating a greater constraint on coding sites.

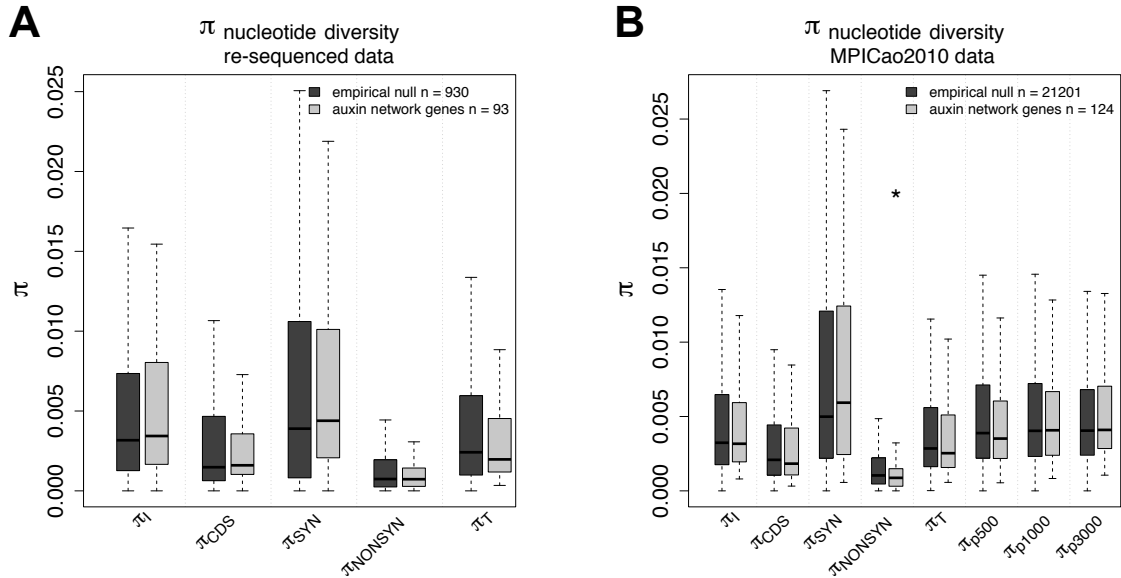


Figure F3.2: Nucleotide diversity comparing the empirical null distribution with combined auxin network groups as one group. (A) π_I , π_{CDS} , π_{SYN} , π_{NONSYN} and π_T comparison of 93 re-sequenced genomic fragments (58, 93, 93, 93 and 93 per category) to 930 fragments of the Nordborg2005 data as the empirical null distribution (396, 336, 336, 336 and 930 per category). (B) π_I , π_{CDS} , π_{SYN} , π_{NONSYN} and π_T comparison of 124 auxin network genes to 21201 genes of the MPICao2010 data as the empirical null distribution. Error bars denote the 95% confidence interval. Asterisk indicates a significant difference between empirical null distribution and combined auxin network groups (Wilcox Test, $p < 0.05 < 0.01$; Benjamini-Hochberg corrected).

That there are no clear differences between the empirical null distribution and the auxin network genes, analyzed as one group, was expected, since introns flanking sites and synonymous sites should evolve neutral. The lower median π_{NONSYN} value in the MPICao2010 data indicates that fewer nonsynonymous changes occur in this group, which would argue for negative selection acting on the auxin network genes.

To take a closer look at the individual auxin network gene group levels, the basic population genetic parameter, nucleotide diversity, and other parameters were compared to the empirical null distribution for each group separately.

Nucleotide variation of auxin network gene groups considering coding sequences

All median values for each auxin network group for different population genetic parameters, which will be discussed in detail in this subsection, are given in Supplementary Table ST6 for the MPICao2010 data. The corresponding median values for the re-sequenced data are given in Supplementary Table ST7.

Figure F3.3 shows the nucleotide diversity separated for each auxin network gene group for the MPICao2010 data. For π_{SYN} no significant differences comparing the individual auxin network gene groups and the empirical null distribution could be detected. However, for the auxin metabolism group the median value of 0.013 is shifted to higher values compared to the median value of 0.00499 of the empirical null distribution (see Figure F3.3 A). It might be, that auxin metabolism genes map to genomic regions with higher substitution rates. For π_{NONSYN} significant differences between the auxin network genes treated as one group, the auxin signaling group and the auxin transport group and the empirical null distribution could be detected (see Figure F3.3 B). Also for π_{NONSYN} the median value of 0.00157 of the auxin metabolism group is shifted to higher values compared to the empirical null distribution median value of 0.00104, like observed for π_{SYN} .

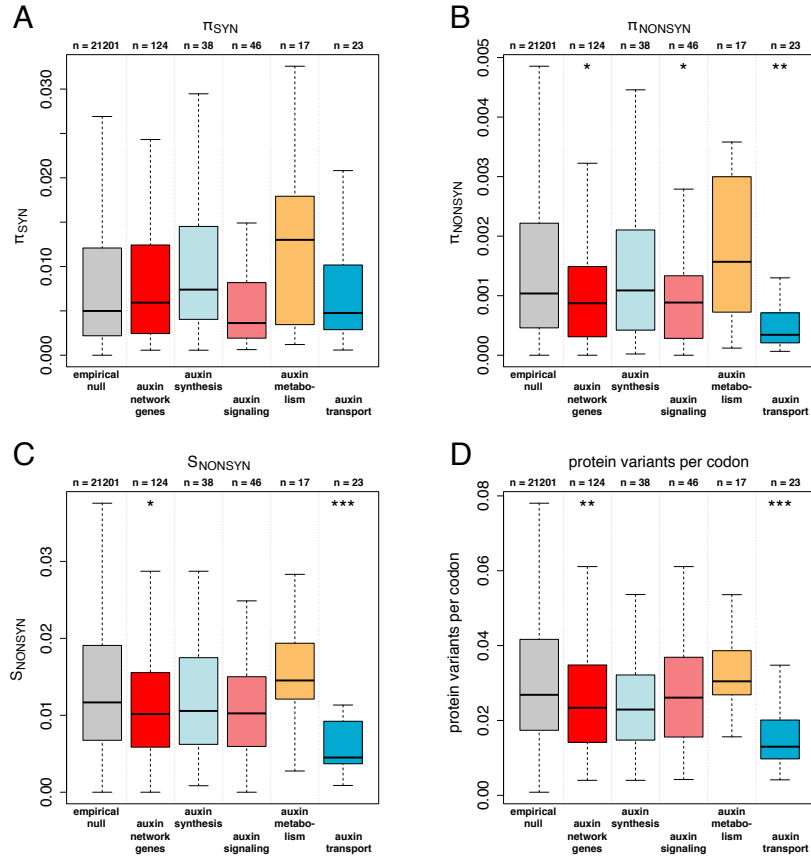


Figure F3.3: Nucleotide diversity of auxin network gene groups considering coding sequences. (A) π_{SYN} , (B) π_{NONSYN} , (C) S_{NONSYN} , (D) protein variants per codon of auxin network gene groups in the MPICao2010 data compared to the corresponding empirical null distribution. Grey: empirical null distribution; red: auxin network genes as one group; light blue: auxin synthesis group; light red: auxin signaling group; orange: auxin metabolism group; blue: auxin transport group. Asterisk indicates a significant difference between empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).

S_{NONSYN} is the total amount of nonsynonymous segregating sites divided by the total amount of nonsynonymous sites and indicates an over- or under-representation of nonsynonymous segregating sites, which lead to amino acid changes. The auxin network genes treated as one group and the auxin transport group are significant different from the empirical null distribution in this parameter (see Figure F3.3 C).

Consistent with S_{NONSYN} and π_{NONSYN} , also the population genetic parameter protein variants per codon is significantly different for the auxin network genes treated as one group and the auxin transport group in the MPICao2010 data (see Figure F3.3 D).

Since the auxin transport genes show for all these population genetic parameters significant lower values than the empirical null distribution, negative selection is acting on the auxin transport genes.

Like for π_{SYN} and π_{NONSYN} , also for the parameters S_{NONSYN} and protein variants per codon, the median values of the auxin metabolism group are shifted to higher values. However, again these observations are not significant, but indicate that the auxin metabolism genes have higher mutation rates than the other auxin network groups. Possibly, the investigated auxin metabolism genes map to genomic regions, with globally higher mutation rates than other genomic regions. In support of this, not only π_{NONSYN} but also π_{SYN} is affected in the same magnitude and direction. If only π_{NONSYN} would be affected, this would be an indication for positive selection, which was previously shown by Bakker et al. (2006) to exist for disease resistance genes.

The ratio of nonsynonymous nucleotide diversity and synonymous nucleotide diversity as an indicator of selection

Mutations in coding sequences that change amino acids are in most cases deleterious and will be removed from the population over time. As described by Hartl and Clark (2007), they may by chance persist in a population for a few generations, but eventually they are eliminated by selection, which is called in this context purifying selection. Another source of losing mutations is genetic drift. Here, random sampling will alter the frequency of alleles in a population. Since random sampling can only remove alleles from a population, over time depending on the population size, genetic drift will lead to the loss of genetic variance.

To identify, which kind of selection might be acting on a gene, the ratio of π_{NONSYN} and π_{SYN} , also called ω , can give an estimation for that. If there are no selective constraints, as in a pseudogene, then the expected value for ω would be 1. But if there is purifying selection, then the expectation is that $\omega < 1$ and $\omega > 1$ for positive selection (Hartl and Clark, 2007).

To illustrate ω , each π_{NONSYN} value was plotted against the corresponding π_{SYN} value for each representative gene model for the auxin network genes and the empirical null distribution. To better visualize the auxin network groups, trend lines were plotted as regression lines for each group, which were calculated with the `lm` function in R. The results for the MPICao2010 data are given in Figure F3.4.

All regression lines lie under the neutrality line, indicating a bias towards negative selection for the majority of the representative gene models. To allow conclusions here again the auxin network groups were compared to the empirical null distribution and not to the neutrality line.

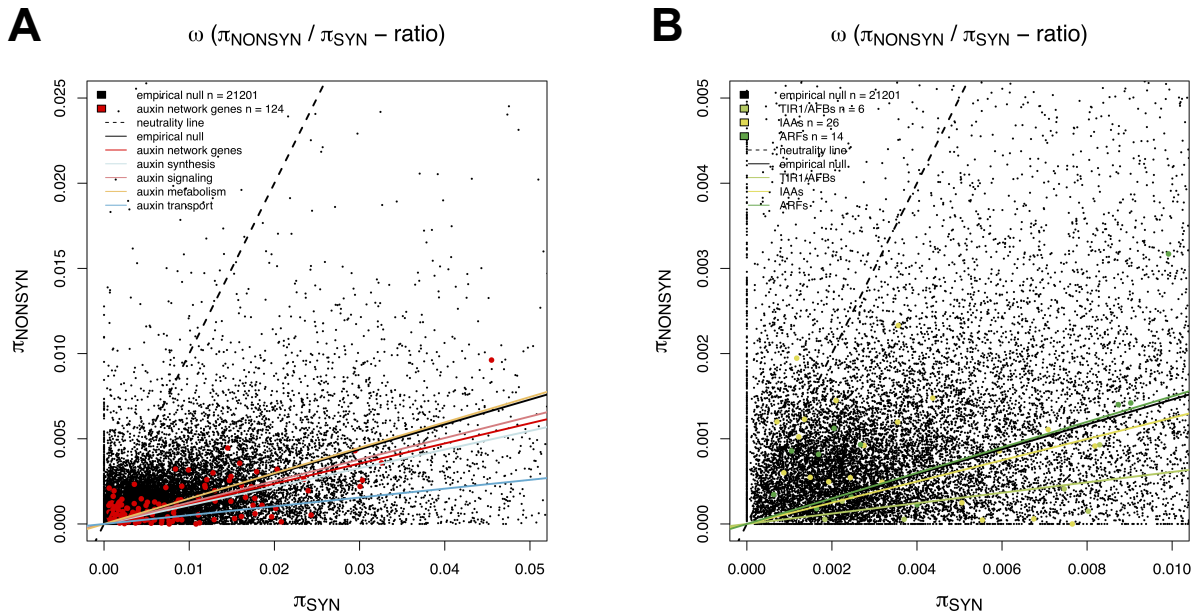


Figure F3.4: Ratio plot of π_{NONSYN} and π_{SYN} for auxin network gene groups. (A) π_{NONSYN} / π_{SYN} - ratio for auxin network gene groups in the MPICao2010 data and the corresponding empirical null distribution, (B) π_{NONSYN} / π_{SYN} - ratio for the auxin signaling group in the MPICao2010 data and the corresponding empirical null distribution. Each black dot represents one representative gene model of the empirical null distribution, each red dot represents one of the auxin network genes representative gene models. Dashed lines represent the neutrality line, if synonymous and nonsynonymous mutations would occur at same frequencies during evolution. Black lines represent regression lines through zero of the empirical null distribution. Colors given in the legend explain assignment of regression lines for the individual auxin network groups. Red: auxin network genes as one group; light blue: auxin synthesis group; light red: auxin signaling group; orange: auxin metabolism group; blue: auxin transport group.

As already expected from the low π_{NONSYN} value, the auxin transport group has the lowest regression line of all auxin network groups (see Figure F3.4 A). This indicates that negative selection acts on the auxin transport genes.

The regression line for the auxin metabolism group falls above the empirical null distribution regression line, indicating a faster evolution than the average *A. thaliana* gene. The regression line for the auxin signaling group falls below the empirical distribution regression line, indicating a slower evolution than the average *A. thaliana* gene. The regression line for the auxin synthesis group and the auxin network genes treated as one group, fall in below the empirical null distribution regression line, indicating a slight negative selection acting on these groups.

Nucleotide variation based on other population genetic parameters

In addition to population genetic parameters, which in principle draw conclusions based on possible amino acid changes, here, three additional population genetic parameters are discussed.

The haplotype diversity (H_d) is a measure for allele abundance on the coding sequence level in this case. A value of 1 would mean, that each accession has its own haplotype and with that its own allele for a given representative gene model. In contrast, low values would mean, that the majority of accessions share one common sequence and that this allele is over-represented throughout the accessions.

π_{SYNMAX} is a measure of the maximum number of synonymous substitutions between any pair of accessions at a representative gene model. It provides a gauge of the time to the common ancestry of the sample as outlined by Bakker et al. (2006). Large values indicate the presence of long-lived alleles at a locus, whereas small values imply a recent common ancestry of alleles.

Tajima's D, here conducted on coding sites (D_{CDS}), is a test statistic proposed by Tajima (1983). It is basically a test for neutrality with the assumption that all mutations are selectively neutral and the investigated population has been in mutation-drift balance for a long period of evolutionary time as outlined by (Kimura (1983) and Nei and Kumar (2000)). $D_{CDS} > 0$ would indicate an excess of polymorphisms in intermediate frequency alleles indicating balancing selection. $D_{CDS} < 0$ would indicate an excess of rare polymorphisms most likely due to singeltons, which would indicate purifying selection. However, if these assumptions are not true for a population, for example if a population has gone recently through a bottleneck with an expansion of population size after these bottleneck, D_{CDS} may become significantly positive or negative depending on the population history (Tajima, 1989b). Fu and Li's D* and F* statistics (Fu and Li, 1993), which relies on the comparison of singeltons to the diversity estimators θ_w and θ_π , can be interpreted like Tajima's D statistics with the same sensitivity to demographic events.

Here again applies what was mentioned above, that in *A. thaliana* this population genetic parameter did not follow a distribution expected under the neutral model of molecular evolution (Nordborg et al., 2005). The shift to lower values most likely reflect the postglacial recolonization events in the life history of *A. thaliana* (Sharbel et al., 2000). To overcome once again this restriction, the calculated values were compared to an empirical null distribution to detect outliers in the extreme tails of this empirical null distribution.

In Figure F3.5 the results for these three population genetic parameters in the auxin network gene groups are shown.

Auxin network genes treated as one group and the auxin synthesis group have almost equal median value of H_d (see Figure F3.5 A and Supplementary Table ST6). The auxin signaling group has the lowest median H_d value. The auxin metabolism has the broadest range and the auxin transport group the shortest range.

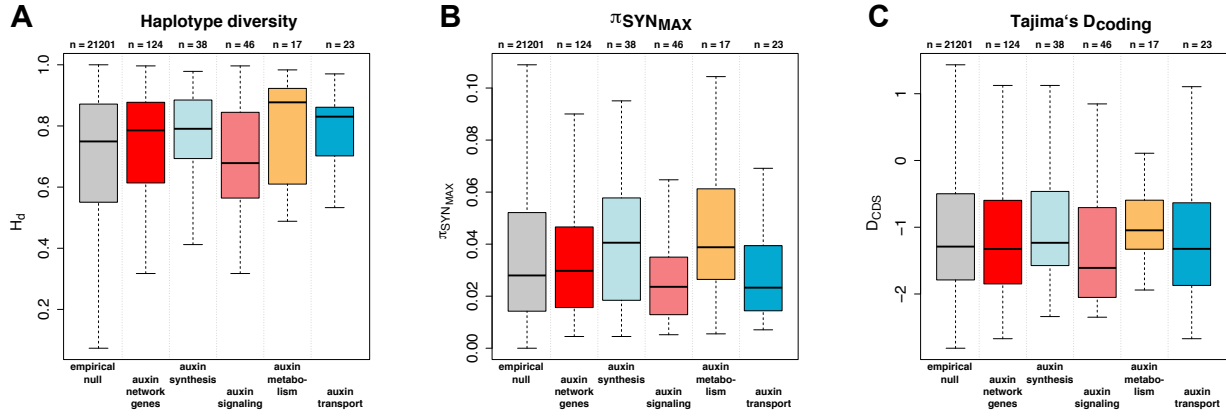


Figure F3.5: Nucleotide variation patterns of auxin network gene groups considering coding sequences. (A) Haplotype diversity (H_d), (B) $\pi_{SYN_{MAX}}$, (C) Tajima's D for coding sites (D_{CDS}) of auxin network gene groups in the MPICao2010 data compared to the corresponding empirical null distribution. Grey: empirical null distribution; red: auxin network genes as one group; light blue: auxin synthesis group; light red: auxin signaling group; orange: auxin metabolism group; blue: auxin transport group. Asterisk indicate significant differences between the empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$).

Consistent with the above mentioned higher π_{SYN} and π_{NONSYN} values of the auxin metabolism group, here the median H_d value was also higher compared to the empirical null distribution median H_d value, indicating that the majority of auxin metabolism genes evaluated map to genomic regions of faster evolution.

The considerably high median H_d value for the auxin transport group might just be observed because of the nature of their long coding sequences.

For the $\pi_{SYN_{MAX}}$ parameter, no auxin network group was significantly different from the empirical null distribution (see Figure F3.5 B). However, the median $\pi_{SYN_{MAX}}$ value for the auxin synthesis group and the auxin metabolism group was shifted to higher values, indicating long-lived alleles for these two groups. It is worth to mention, that among these two groups, also three outliers were present, which fall into the extreme upper tail of the empirical null distribution, indicating extremely old alleles. Two of them are in the group of auxin metabolism, *GRETCHEN HAGEN3.4* (*GH3.4*) and *IAA-ALANINE RESISTANT 1* (*IAR1*), and one of them is in the group of auxin synthesis, *PHOSPHORIBOSYLANTHRANILATE TRANSFERASE 1* (*PAT1*).

But also representative gene models, which fall into the lower tail of the empirical null distribution, indicating extremely young alleles, were identified. One for the auxin metabolism group, *GRETCHEN HAGEN3.9* (*GH3.9*), one for the auxin synthesis group, *TSK-ASSOCIATING PROTEIN 1* (*TSA1*), and three for the auxin signaling group, *AUXIN RESPONSE FACTOR 4* (*ARF4*), *AUXIN RESPONSE FACTOR 18* (*ARF18*) and *AUXIN RESPONSE FACTOR 19* (*ARF19*).

A D_{CDS} value of zero should indicate neutral evolution, whereas positive values indicate positive evolving loci and negative values indicate negative evolving loci. However, as already mentioned, the null hypothesis is not true for the majority of loci. No auxin network gene group is significantly different from the empirical null distribution. It seems, that the auxin signaling group is shifted to more negative values compared to the empirical null distribution, whereas the auxin metabolism group has slightly more positive values. Six outliers were detected within the MPICao2010 data set. Three in the upper tail of the empirical null distribution: *IAA19* ($D_{CDS} = 1.49246$), *AT2G28880* (AnthranilateSynthase) ($D_{CDS} = 1.773$) and *P-GLYCOPROTEIN 1* (*PGP1*) ($D_{CDS} = 1.50694$); and three in the lower tail of the empirical null distribution: *IAA14* ($D_{CDS} = -2.34793$), *ARF18* ($D_{CDS} = -2.34281$) and *PILS2* ($D_{CDS} = -2.66986$).

This analysis revealed that within the auxin network gene groups, the auxin signaling group has the

lowest median H_d value, the youngest alleles together with the auxin transport group by the $\pi_{SYN_{MAX}}$ parameter and the lowest median D_{CDS} value. In the next subsection, I investigate the auxin signaling genes in more detail.

Site frequency spectra of auxin network gene groups

As mentioned by Hartl and Clark (2007), natural selection is expected to distort the spectrum of frequencies of variants within genes. Purifying selection drives variants to lower frequency, and positive selection produces an excess of high-frequency derived alleles.

To identify footprints of selection, the site frequency spectra of the representative gene models were calculated for synonymous, nonsynonymous and introns_flanking sites in the MPICao2010 data set. Here, the minor allele frequency (MAF) was calculated, which can at most reach values of 0.5 or 39 out of 78 accessions. How the MAF was calculated is described in detail in the corresponding materials and methods section.

In addition, an individual site frequency plot was made for each auxin network gene and the resulting frequency classes for nonsynonymous and synonymous sites were divided, resulting in a quick overview of auxin network genes, which might be driven by positive selection (see Supplementary Figure S1, S2, S3, S4, S5 and S6).

As already published by Nordborg et al. (2005), the MAF for synonymous and nonsynonymous sites is shifted to rare alleles compared to standard neutral expectations. The MAF for nonsynonymous sites was even more biased to lower frequency values than the synonymous sites, suggesting, that selective factors must be involved (Nordborg et al., 2005). Figure F3.6 illustrates the MAF for synonymous, nonsynonymous and introns_flanking sites within the MPICao2010 data. For the empirical null distribution, the fraction of loci for nonsynonymous sites is even more skewed towards lower frequency values than the synonymous sites, consistent with the findings of Nordborg et al. (2005).

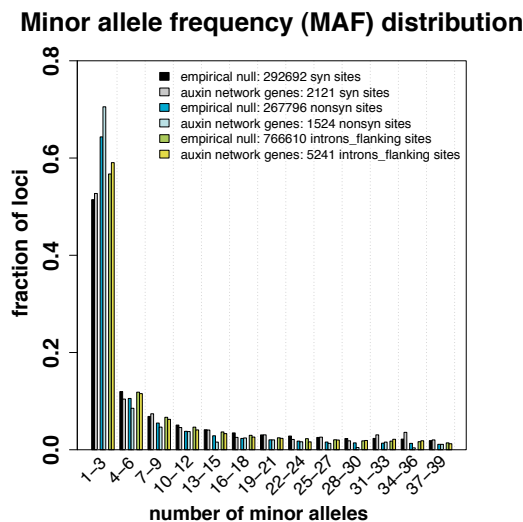


Figure F3.6: Minor allele frequency distribution of auxin network genes treated as one group. MAF for three different categories: synonymous, nonsynonymous and introns_flanking sites. The fraction of loci is plotted for minor allele classes. Black: synonymous sites evaluated for the empirical null distribution; grey: synonymous sites for the auxin network gene group; blue: nonsynonymous sites for the empirical null distribution; light blue: nonsynonymous sites for the auxin network gene group; green: introns_flanking sites for the empirical null distribution; light green: introns_flanking sites for the auxin network gene group. Number of sites evaluated for each site class is given in the legend.

For the auxin network genes treated as one group, the MAF for synonymous sites is in the same range like the empirical null distribution, but the MAF for nonsynonymous sites is even more skewed towards rare alleles. This indicates that negative or purifying selective forces acting on the auxin network genes are stronger than selective forces acting on the average *A. thaliana* gene.

To take a closer look at the individual auxin network groups, they were analyzed separately and results are given in Figure F3.7. The synonymous MAF for the auxin signaling group is significantly different from the empirical null distribution and skewed towards lower frequency values. The synonymous MAF of the auxin metabolism group is significantly different from the empirical null distribution and skewed towards higher frequency values (see Figure F3.7 A).

For the nonsynonymous MAF, all but the auxin metabolism group are significantly different from the empirical null distribution, with values shifted to rare alleles (see Figure F3.7 B). Together, this indicates purifying selective forces acting on these groups. Like identified for other population genetic parameters, here also negative selection acts on the auxin transport group.

In addition, for the auxin signaling group also the introns_flanking MAF is skewed towards lower MAF frequency values. Interestingly, the MAF values for the auxin signaling group is skewed towards rare alleles in all three MAF categories, this indicates negative selection even on non-coding sites (see Figure F3.7 C). This kind of MAF pattern is also seen for genes in close vicinity to a gene, which is under strong natural selection. Neighboring loci will show reduced variation in that genomic region. Alternatively, this observed pattern might be due to the evolution of auxin signaling genes, which exist in tandem duplicates. In this respect they might also show the low MAF frequency in introns_flanking sites.

In total, the site frequency spectra analysis revealed again hints for negative selection forces acting on the auxin transport group. But also hints for negative selection could be detected for the auxin signaling group and even the auxin synthesis group. For the auxin metabolism, no differences from the empirical null distribution could be detected for nonsynonymous MAF category.

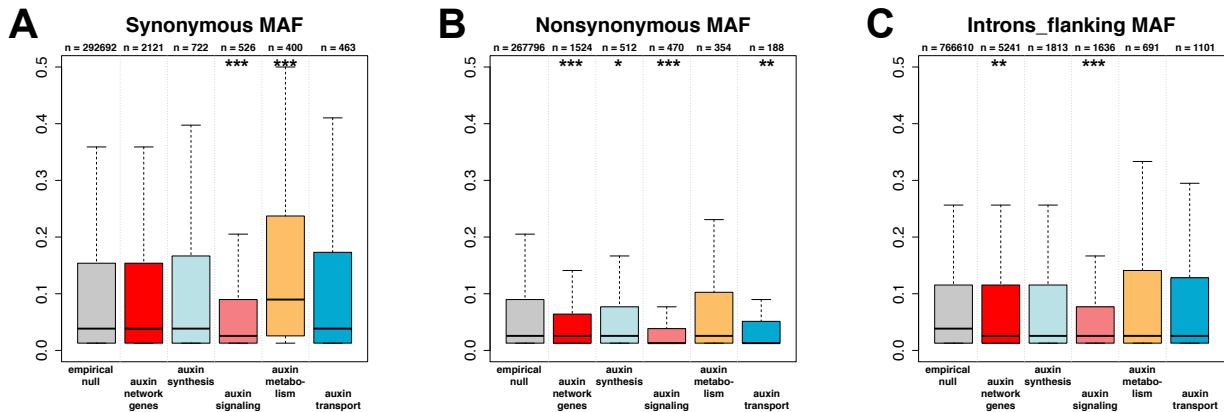


Figure F3.7: Minor allele frequency distribution of auxin network groups presented as boxplots. (A) Synonymous MAF (B) Nonsynonymous MAF (C) Introns_flanking MAF of individual auxin network groups. Grey: empirical null distribution; red: auxin network genes grouped as one group; light blue: auxin synthesis group; light red: auxin signaling group; orange: auxin metabolism group; blue: auxin transport group. The number of analyzed sites is given under each boxplot. Asterisk indicate significant differences between the empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).

3.1.3 Variation of population genetic parameters inside the auxin signaling gene families

To unravel selective forces inside the auxin signaling group, here the auxin signaling is divided into its three auxin signaling families: the *TIR1/AFB* co-receptor gene family, the *AUX/IAA* repressor gene family and the *ARF* transcription factor gene family.

Co-receptors, repressors and transcription factors - Is there a hierarchy?

An interesting question is, whether the three distinct and hierarchical signaling components show different types of selection, or if the signaling hierarchy is also detected based on population genetic parameters.

First, in Figure F3.8 the nucleotide variation between the individual auxin signaling gene families is shown.

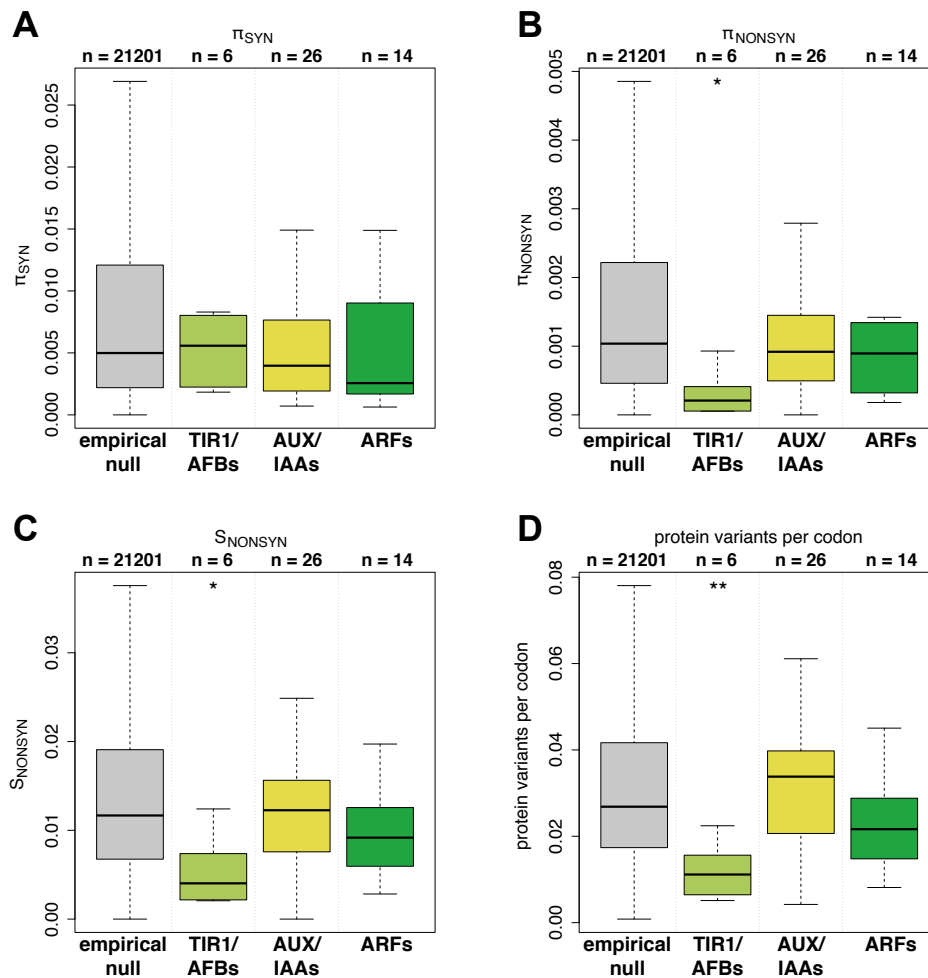


Figure F3.8: Nucleotide diversity of auxin signaling gene families considering coding sequences. (A) π_{SYN} , (B) π_{NONSYN} , (C) S_{NONSYN} , (D) protein variants per codon of auxin signaling gene families in the MPICao2010 data compared to the corresponding empirical null distribution. Grey: empirical null distribution; green: *TIR1/AFBs*; light green: *AUX/IAAs*; dark green: *ARFs*. Asterisk indicate significant differences between the empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).

For π_{SYN} no significant differences to the empirical null distribution can be detected, which is fine, because synonymous segregating sites should not be under selection. For π_{NONSYN} , S_{NONSYN} and protein variants per codon, the *TIR1/AFB* co-receptor gene family is significantly different from the empirical null distribution. Here, the lower values indicate negative selection acting on the *TIR1/AFB* co-receptor gene family. The *AUX/IAAs* and the *ARFs* are not significantly different from the empirical null distribution, but the *AUX/IAAs* show higher values compared to the *ARFs*. This suggests, that on coding sequence level the *AUX/IAAs* show more variation than the *ARFs*.

In Figure F3.9, the nucleotide pattern of the auxin signaling gene families is shown for the population genetic parameters, H_d , $\pi_{SYN_{MAX}}$ and D_{CDS} . The *IAAs* show a significantly reduced median H_d value, whereas the *ARFs* show a significantly increased median H_d value.

The significantly reduced median $\pi_{SYN_{MAX}}$ for the *ARFs* indicates a recent common ancestry allele. The *TIR1/AFBs* and the *IAAs* do not show this pattern.

The D_{CDS} values for the auxin signaling gene families are following the same trend as observed for the above mentioned nucleotide variation parameters. The *TIR1/AFBs* are on the bottom, the *IAAs* on the top and the *ARFs* in between. Indicating negative selection forces acting strongest on the *TIR1/AFBs*, followed by the *ARFs* and more relaxed for the *AUX/IAAs*.

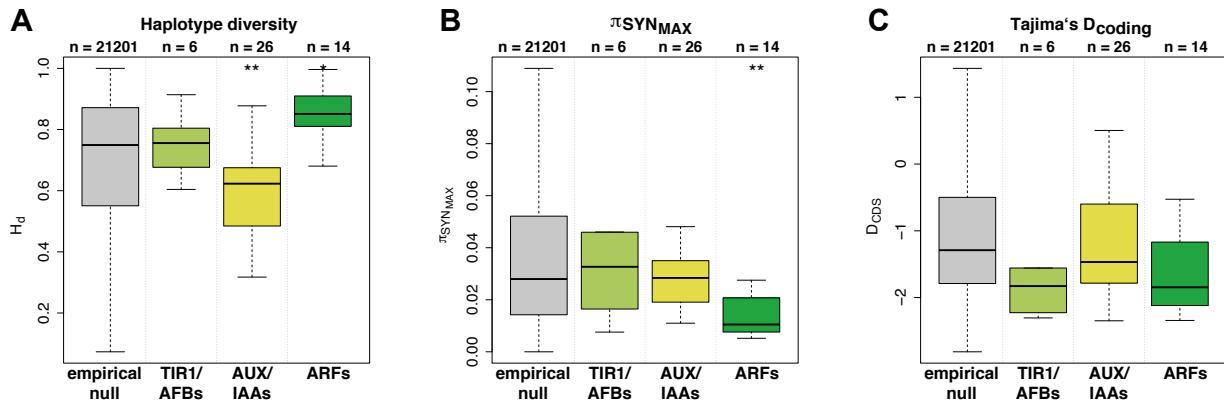


Figure F3.9: Nucleotide variation patterns of auxin signaling gene families considering coding sequences. (A) Haplotype diversity (H_d), (B) $\pi_{SYN_{MAX}}$, (C) Tajima's D for coding sites (D_{CDS}) of auxin signaling gene families in the re-sequenced data compared to the corresponding empirical null distribution. (D) Haplotype diversity (H_d), (E) $\pi_{SYN_{MAX}}$, (F) Tajima's D for coding sites (D_{CDS}) of auxin signaling gene families in the MPICao2010 data compared to the corresponding empirical null distribution. Grey: empirical null distribution; green: *TIR1/AFBs*; light green: *AUX/IAAs*; dark green: *ARFs*. Asterisk indicate significant difference between the empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).

Considering population genetic parameters, the *TIR1/AFBs* are under the strongest negative selection forces, which fit to their hierarchy in the signaling process as the stable part of the co-receptor complex. The *ARFs* as the transcription factors are more variable as a group, followed by the unstable part of the co-receptor complex, the *AUX/IAAs*. The *AUX/IAAs* are bound to the *ARFs* and repress their transcription activity. Upon an auxin stimulus, they are recruited to the *SCF-TIR1* complex and are marked for degradation. Here, for the *AUX/IAAs* the most relaxed negative selection forces were detected among the auxin signaling groups. Hence, the question of the beginning of this subsection has to be neglected in that way, that the signaling hierarchy *TIR1/AFBs* \rightarrow *AUX/IAAs* \rightarrow *ARFs* is not represented by population genetic parameters. Here, the hierarchy starts from the strongest negative selection forces acting on the *TIR1/AFBs* \rightarrow *ARFs* \rightarrow *IAAs*.

3.1.4 Interspecific nucleotide divergence

The nucleotide diversity can not only be calculated within one species, but also between species, and is then called nucleotide divergence. Here, the closely related Brassicaceae *A. lyrata*, *B. rapa* and *T. parvula* were analyzed. For this purpose, a BLASTP was conducted to find orthologous gene sequences between the 21325 representative gene models of the MPICao2010 data and the mentioned Brassicaceae. The resulting codon alignments were used to calculate the synonymous nucleotide divergence K_{SYN} , the nonsynonymous nucleotide divergence K_{NONSYN} and the $\frac{K_{NONSYN}}{K_{SYN}}$ - ratio, between *A. thaliana* and *A. lyrata*, *B. rapa* or *T. parvula*.

The nucleotide divergence can highlight selection forces, which might be active between two species. For the $\frac{K_{NONSYN}}{K_{SYN}}$ - ratio, like for the $\frac{\pi_{NONSYN}}{\pi_{SYN}}$ - ratio, values < 1 indicate negative selection, whereas values > 1 indicate positive selection.

Nucleotide divergence between *A. thaliana* and closely related Brassicaceae species

Table T3.1 shows the nucleotide divergence results for the different species comparisons. Between *A. thaliana* and *A. lyrata*, the auxin network genes treated as one group is significant different from the empirical null distribution for K_{NONSYN} and the $\frac{K_{NONSYN}}{K_{SYN}}$ - ratio. This indicates that negative selection is the preferred selection force acting on this gene group.

Within the auxin network gene groups, the auxin signaling group and the auxin transport group show significant differences to the empirical null distribution. This indicates that these two groups lead to the observed significant difference for the complete auxin network gene group and highlight purifying selection acting on these groups between *A. thaliana* and *A. lyrata*. However, if the auxin signaling group is further splitted into its auxin signaling gene families, the *TIR1/AFB* gene family, the *AUX/IAA* gene family and the *ARF* gene family, no significant differences from the empirical null distribution are detected anymore. This indicates that within each gene family, also genes exists, that are not strongly conserved between the evaluated species.

The pattern that the auxin network genes treated as one group and the auxin transport group is significantly different from the empirical null distribution for K_{NONSYN} and the $\frac{K_{NONSYN}}{K_{SYN}}$ - ratio holds true throughout all species comparisons, which have been calculated.

This suggests, that the auxin transport genes are very conserved, even between *A. thaliana* and its closely relatives, the three Brassicaceae *A. lyrata*, *B. rapa* and *T. parvula*.

The fact, that the other groups did not differ significantly from the empirical null distribution, does on the other hand not imply, that positive selection is acting on these groups. The observed $\frac{K_{NONSYN}}{K_{SYN}}$ - ratio for the empirical null distribution falls for all comparisons clearly under 1, which would argue for negative selection for the vast majority of genes evaluated here. The auxin transport genes just fall into the lower tail of this distribution, which highlight their extreme conservation between these species.

McDonald and Kreitman Test

The McDonald and Kreitman test examines, whether the ratio of nonsynonymous to synonymous substitutions differs within and between species.

Segregating sites are classified in this test to be either fixed or polymorphic. For example one site, which only consists of 'As' in one species and of a 'G' in the outgroup species, would be considered as a fixed site. If, for example one site in one species consists of 'As' and 'Gs' and of a 'G' in the outgroup species, it would be considered to be polymorphic.

Table T3.1: Nucleotide divergence between *A. thaliana* and three other Brassicaceae.

<i>A. thaliana</i> vs. <i>A. lyrata</i>		K_{NONSYN}^1		K_{SYN}^1		$K_{\text{NONSYN}}/K_{\text{SYN}} - \text{ratio}^1$
group	#rep. gene models ²					
auxin network genes	124	0.01745	***	0.13917		0.13063
synthesis	38	0.02048		0.13496		0.13063
signaling	46	0.01808	*	0.14557		0.13334
metabolism	17	0.01757		0.13089		0.14135
transport	23	0.00883	***	0.14906		0.06227
TIR1/AFBs	6	0.01437		0.14761		0.11244
AUX/IAAs	26	0.01841		0.14477		0.13693
ARFs	14	0.01825		0.14419		0.14119
PINs	7	0.00883		0.14906		0.06227
PILSs	7	0.01324		0.09851		0.13426
empirical null distribution	19785	0.02472		0.14128		0.17732

<i>A. thaliana</i> vs. <i>B. rapa</i>		K_{NONSYN}^1		K_{SYN}^1		$K_{\text{NONSYN}}/K_{\text{SYN}} - \text{ratio}^1$
group	#rep. gene models ²					
auxin network genes	124	0.06433	***	0.45276	*	0.13471
synthesis	38	0.06822		0.45987		0.14381
signaling	46	0.06714		0.44341	*	0.16143
metabolism	17	0.06530		0.45421		0.12971
transport	23	0.03646	***	0.48788		0.06299
TIR1/AFBs	6	0.05198		0.65089		0.08640
AUX/IAAs	26	0.07286		0.42850	**	0.18597
ARFs	14	0.06494		0.40772		0.15088
PINs	7	0.05069		0.34878		0.08337
PILSs	7	0.05495		0.41987		0.13089
empirical null distribution	19785	0.08018		0.47937		0.15989

<i>A. thaliana</i> vs. <i>T. parvula</i>		K_{NONSYN}^1		K_{SYN}^1		$K_{\text{NONSYN}}/K_{\text{SYN}} - \text{ratio}^1$
group	#rep. gene models ²					
auxin network genes	124	0.05099	***	0.34456		0.13871
synthesis	38	0.05998		0.32794		0.14479
signaling	46	0.05279		0.32617		0.14528
metabolism	17	0.04887		0.32310		0.16601
transport	23	0.03250	***	0.35941		0.07116
TIR1/AFBs	6	0.05665		0.45269		0.09390
AUX/IAAs	26	0.05634		0.30189		0.16937
ARFs	14	0.04353		0.30855		0.13511
PINs	7	0.03178		0.31671		0.05748
PILSs	7	0.05796		0.35681		0.16559
empirical null distribution	19711	0.06589		0.35470		0.17986

¹ Median average pairwise nucleotide divergence between 78 to 80 accession sequences and the best blast hit from different brassicaceae species calculated with KaKs-Calculator according to Yang and Nielson (2000).

² Number of representative gene models after excluding values which could not be calculated by KaKs-Calculator.

Asterisks denote significant differences between one group and the empirical null distribution (* P < 0.05; ** P < 0.01; *** P < 0.001) assessed by a two-sided Wilcoxon rank sum test Benjamini Hochberg corrected.

Another classification is made based on the effect of the site, if it is a synonymous or a nonsynonymous site. In total four different types are compared, nonsynonymous fixed sites N_F , synonymous fixed sites S_F , nonsynonymous polymorphic sites N_P and synonymous polymorphic sites S_P .

As described by Nei and Kumar (2000), if the ratio $\frac{N_F}{S_F}$ is significantly larger than the ratio $\frac{N_P}{S_P}$, it is assumed, that some of the nonsynonymous substitutions between the two species are caused by positive selection. In contrast, if the ratio $\frac{N_F}{S_F}$ is significantly smaller than the ratio $\frac{N_P}{S_P}$, it is assumed, that purifying selection has reduced the number of nonsynonymous substitutions between the two species. Significant differences were calculated based on the fisher test and the obtained p-values were corrected for multiple testing.

Here, the same codon alignments as described for the nucleotide divergence analysis above, were used to perform the McDonald and Kreitman test of *A. thaliana* sequences and the best BLASTP hit in the three mentioned Brassicaceae *A. lyrata*, *B. rapa* and *T. parvula*.

The following auxin network genes had a significant p-value after correcting for multiple testing, with the $\frac{N_F}{S_F}$ ratio significantly smaller than the $\frac{N_P}{S_P}$ ratio. This argues for purifying selection, which reduces the number of nonsynonymous substitutions between *A. thaliana* and *A. lyrata*: *ARF11*, *GH3.3*, *IGS1*, *PGP4*, *PILS1*, *PILS3*, *SUR1* and *TAR1*.

The following auxin network genes had a significant p-value after correcting for multiple testing, with the $\frac{N_F}{S_F}$ ratio significantly higher than the $\frac{N_P}{S_P}$ ratio. This argues for positive selection, which leads to nonsynonymous substitutions between *A. thaliana* and *A. lyrata*: *CYP71B15* and *NIT1*.

The following auxin network genes had a significant p-value after correcting for multiple testing, with the $\frac{N_F}{S_F}$ ratio significantly smaller than the $\frac{N_P}{S_P}$ ratio. This argues for purifying selection, which reduces the number of nonsynonymous substitutions between *A. thaliana* and *B. rapa*: Anthranilate Synthase, *AT3G55870*, *ARF1*, *ARF11*, *GH3.3*, *GH3.4* and *TSB2*.

The following auxin network gene had a significant p-value after correcting for multiple testing, with the $\frac{N_F}{S_F}$ ratio significantly higher than the $\frac{N_P}{S_P}$ ratio. This argues for positive selection, which leads to nonsynonymous substitutions between *A. thaliana* and *B. rapa*: *NIT1*.

The following auxin network genes had a significant p-value after correcting for multiple testing, with the $\frac{N_F}{S_F}$ ratio significantly smaller than the $\frac{N_P}{S_P}$ ratio. This argues for purifying selection, which reduces the number of nonsynonymous substitutions between *A. thaliana* and *T. parvula*: Anthranilate Synthase; *AT3G55870*, *ARF11*, *GH3.3*, *GH3.4*, *ILR3*, *TAR1* and *TSB2*.

The following auxin network genes had a significant p-value after correcting for multiple testing, with the $\frac{N_F}{S_F}$ ratio significantly higher than the $\frac{N_P}{S_P}$ ratio. This argues for positive selection, which leads to nonsynonymous substitutions between *A. thaliana* and *T. parvula*: *AFB5* and *NIT1*.

Consistent to the fact, that all comparisons have been made between *A. thaliana* and closely related Brassicaceae, overlapping genes are found to be under negative and positive selection. *ARF11* and *GH3.3* were identified in all three comparisons to be under negative selection, whereas *NIT1* was identified in all comparisons to be under positive selection between *A. thaliana* and the other Brassicaceae.

3.1.5 Extreme auxin network genes

Here, auxin network genes are highlighted, which fall into the extreme lower or upper tail of the empirical null distribution for the calculated population genetic parameters. Extreme outliers can not be recognized in functional groups, since many genes will average the complete group picture. However, the individual contribution of such outliers to the observed natural phenotypic variation might be of great concern.

In Table T3.2 these outliers are listed and the population genetic parameter is highlighted for which the specified outlier falls into an extreme tail of the empirical null distribution. If H_d , R_{min} , π_{SYN} , $\pi_{introns_flanking}$ and π_{SYN_MAX} were identified, these parameters are not indicating some selective forces to be present for these outliers. They just indicate extreme values for the specified outliers for that particular population genetic parameter.

More important are π_{NONSYN} , protein variants per codon, S_{NONSYN} , D_{CDS} , FLD_{CDS} and FLF_{CDS} , since they can be directly interpreted as clear indicators for selective forces, for negative selection for the lower tail and for positive selection for the upper tail.

The majority of outlier genes of the auxin signaling group falls into the lower tail of the empirical distribution. The lower tail of the empirical distribution indicates negative selection. However, *ARF12* showed up in the upper tail for several population genetic parameters. Hence, for *ARF12* positive selective forces seem to act on this auxin signaling gene.

The majority of outlier genes of the auxin transport group falls into the lower tail of the empirical null distribution. This findings are consistent with the previous findings for the auxin transport group, that negative selection is acting on the auxin transport group. However, *PGP1* and *PGP19* showed up in the upper tail of the empirical null distribution. Here, high H_d values do not indicate positive selection, so that only *PGP1* shows with the high D_{CDS} value hints for positive selection.

Table T3.2: Extreme auxin network genes in the tails of the empirical null distribution.

AGI	gene name	auxin network gene group	2.5 per. ^a	97.5 per. ^b
AT3G62980.1	TIR1	TIR1/AFBs	π_{NONSYN}	
AT4G03190.1	AFB1	TIR1/AFBs	π_{NONSYN} ; protein variants per codon; FLD_{CDS} ; FLF_{CDS}	
AT1G04240.1	IAA3	AUX/IAAs	$\pi_{introns_flanking}$	
AT1G04250.1	IAA17	AUX/IAAs	π_{NONSYN}	
AT1G80390.1	IAA15	AUX/IAAs		protein variants per codon
AT3G04730.1	IAA16	AUX/IAAs	S_{NONSYN} ; protein variants per codon	
AT3G15540.1	IAA19	AUX/IAAs		D_{CDS}
AT3G62100.1	IAA30	AUX/IAAs	S_{NONSYN} ; protein variants per codon	
AT4G14550.1	IAA14	AUX/IAAs	D_{CDS} ; FLD_{CDS} ; FLF_{CDS}	
AT4G14560.1	IAA1	AUX/IAAs		FLD_{CDS}
AT5G25890.1	IAA28	AUX/IAAs	π_{NONSYN}	
AT1G19220.1	ARF19	ARFs	π_{SYN_max}	
AT1G34310.1	ARF12	ARFs		S_{NONSYN} ; H_d ; R_{min} ; protein variants per codon
AT3G61830.1	ARF18	ARFs	D_{CDS} ; π_{SYN} ; π_{SYN_max}	
AT5G60450.1	ARF4	ARFs	π_{SYN_max}	
AT1G08980.1	AMI1	auxin synthesis	FLD_{CDS} ; FLF_{CDS}	
AT1G23320.1	TAR1	auxin synthesis	π_{SYN}	
AT1G25220.2	ASB1	auxin synthesis	π_{SYN}	$\pi_{introns_flanking}$
AT1G29410.2	PAI3	auxin synthesis	FLF_{CDS}	
AT2G28880.1	AnthranilateSynthase	auxin synthesis		D_{CDS}
AT3G54640.1	TSA1	auxin synthesis	π_{SYN_max}	
AT4G27070.1	TSB2	auxin synthesis		$\pi_{introns_flanking}$
AT4G31500.1	CYP83B1/SUR2	auxin synthesis	π_{NONSYN} ; protein variants per codon; S_{NONSYN}	
AT5G17990.1	PAT1	auxin synthesis		$\pi_{introns_flanking}$; π_{SYN_max}
AT5G38530.1	TrpSynBetaChain2	auxin synthesis		H_d
AT1G59500.1	GH3.4	auxin metabolism		$\pi_{introns_flanking}$; π_{CDS} ; π_{NONSYN} ; π_{SYN} ; H_d ; protein variants per codon; S_{NONSYN} ; π_{SYN_max}
AT1G68100.1	IAR1	auxin metabolism		FLD_{CDS} ; FLF_{CDS} ; π_{SYN_max}
AT2G47750.1	GH3.9	auxin metabolism	π_{SYN_max}	
AT1G20925.1	PILS1	auxin transport	FLD_{CDS} ; FLF_{CDS}	
AT1G71090.1	PILS2	auxin transport	D_{CDS} ; FLD_{CDS} ; FLF_{CDS} ; H_d	
AT2G21050.1	LAX2	auxin transport	S_{NONSYN} ; protein variants per codon	
AT2G34650.1	PID	auxin transport	π_{SYN} ; FLD_{CDS}	
AT2G36910.1	PGP1	auxin transport		D_{CDS}
AT2G38120.1	AUX1	auxin transport	π_{NONSYN}	
AT2G47000.1	PGP4	auxin transport	π_{CDS}	
AT3G28860.1	PGP19	auxin transport		H_d

^a Population genetic parameter, which falls into the lower tail (2.5 percentil) of the empirical null distribution.

^b Population genetic parameter, which falls into the upper tail (97.5 percentil) of the empirical null distribution.

For the auxin synthesis outlier genes, the pattern is very variable, both kind of outliers are represented in an equal amount.

For the auxin metabolism outlier genes, clearly *GH3.4* stands out, since it falls in the upper tail of the empirical null distribution in many population genetic parameters. Hence, *GH3.4* might map to a genomic region with a high mutation rate influencing nonsynonymous and synonymous changes.

Evaluating extreme auxin network genes had shown, that the majority of auxin signaling genes falls into the lower tail of population genetic parameters, which postulate negative selection. The same is true for the auxin transport genes.

3.1.6 Comparing sequence and transcript diversity

In the last subsection of the first results section, the sequence diversity within 80 *A. thaliana* accessions is compared to the transcript diversity of six *A. thaliana* accessions after exogenous auxin treatment. Here, it can be evaluated, if the sequence diversity between individual auxin network gene groups is reflected by transcriptional read out of six diverse *A. thaliana* accessions. A detailed description of how the transcript diversity was calculated can be found in the materials and methods section.

Figure F3.10 A illustrates the transcript diversity between the auxin network gene groups. In Figure F3.10 B the auxin signaling gene group is splitted into its gene families to illustrate their individual contribution to the transcript diversity among the six *A. thaliana* accessions.

The auxin network genes treated as one group are significantly different from the empirical null distribution. This indicates high variability of the transcriptional response to exogenous auxin treatment among the investigated six *A. thaliana* accessions. Looking at individual auxin network gene groups, the auxin synthesis, auxin signaling and auxin metabolism show a significant difference to the empirical null distribution. Here, the auxin transport genes do not show a difference to the empirical null distribution.

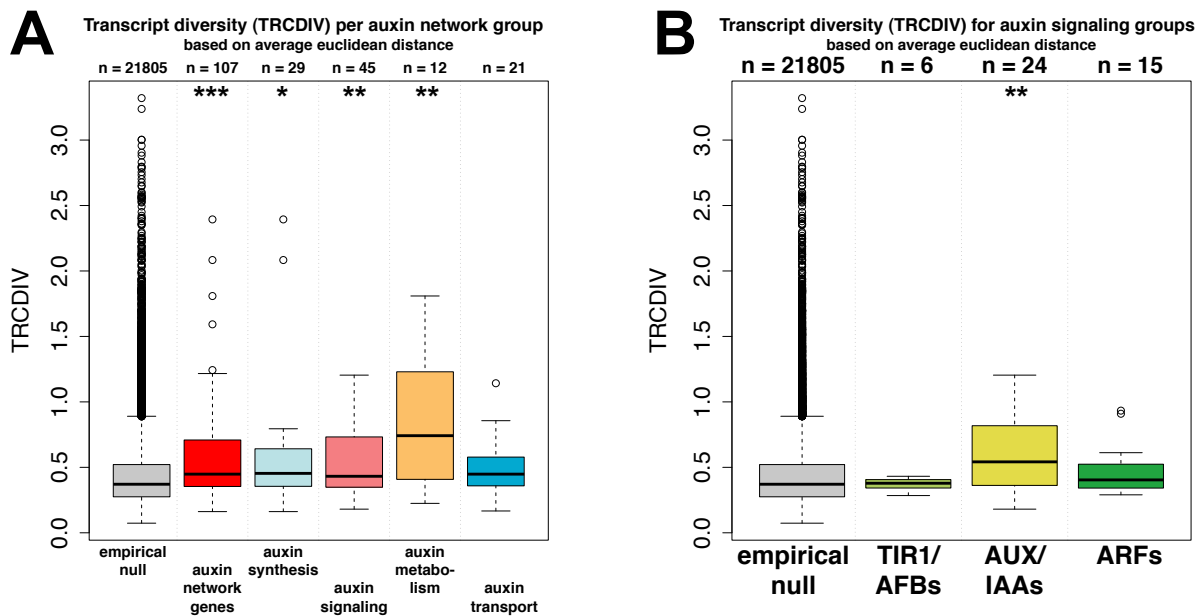


Figure F3.10: Transcript diversity among six *A. thaliana* accessions comparing auxin network gene groups to the empirical null distribution. (A) Transcript diversity between auxin network gene groups. (B) Transcript diversity within the auxin signaling group. Asterisks indicates a significant difference between empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).

The highest value for transcript diversity was identified for the auxin metabolism group. In addition, they showed also the highest values for π_{SYN} , π_{NONSYN} , protein variants per codon (see Figure F3.3) and D_{CDS} (see Figure F3.5). In total, the auxin metabolism group is the most variable among the evaluated auxin network gene groups.

The auxin signaling group is also significantly different from the empirical null distribution for transcript diversity. If one looks into the individual gene families within the auxin signaling group, the *AUX/IAAs* turn out to be the significant factor.

Here, it seems that the group with the highest π_{NONSYN} , the highest protein variants per codon (see Figure F3.8) and the highest D_{CDS} (see Figure F3.9) value contributes the most to the transcript diversity. The same kind of relation was seen for the auxin metabolism group, when the individual auxin network groups were compared to each other.

To evaluate, if this kind of relation is not only true for complete groups, but maybe also on the individual gene level, chromosome-wide plots of π_{SYN} , π_{NONSYN} , $\pi_{introns_flanking}$ have been generated and plotted together with the transcript diversity. The results are given in Supplementary Figure S13. For some genomic regions there seem to be a correlation of nucleotide diversity and transcript diversity, like for a region on chromosome one including *IAA18* and *IAA6*. However, for other regions the relation of nucleotide diversity and transcript diversity is not visible. This might be also due to the fact, that only six *A. thaliana* accessions could be analyzed for transcript diversity upon exogenous auxin treatment.

In the first part of the results, a population genetic analysis of auxin network genes was conducted, in a self-prepared re-sequenced data set and within a large sample of accessions via a meta-analysis. In this thesis the description of the results was restricted to the MPICao2010 data, but in most cases both data sets revealed consistent results. According to the results, strong negative selection is acting on the auxin transport group. Hence, most likely the nucleotide diversity of the auxin transport genes is not contributing to the observed natural phenotypic variation upon exogenous auxin stimulus.

The auxin metabolism group turned out to be the most variable group between the auxin network gene groups, which findings were also underlined in a transcript diversity analysis between six diverse *A. thaliana* accessions. Here, the auxin metabolism group showed the highest transcript diversity values of all conducted group comparisons.

Among the auxin signaling group, which treated as one group showed signatures of negative selection, the *AUX/IAA* gene family was identified as the most variable part. This was also reflected by a transcript diversity analysis between six *A. thaliana* accessions.

The signaling hierarchy *TIR1/AFBs* \rightarrow *AUX/IAAs* \rightarrow *ARFs* is not recapitulated within the population genetic parameters determined. However, the selection force hierarchy *TIR1/AFBs* \rightarrow *ARFs* \rightarrow *AUX/IAAs* was visible within the species *A. thaliana*. Between *A. thaliana* and closely related Brassicaceae this hierarchy of conservation was not so dominant.

3.2 Analysis of auxin response traits in two RIL populations

Existing and rapidly upcoming genetic resources allow for quantitative genetic approaches such as quantitative trait locus (QTL) analysis (described in this section) or genome-wide association studies (GWAS) (described in the next section). Such approaches demand high-throughput phenotyping. Image based efforts that have been undertaken to overcome this bottleneck capture the shape of whole organisms (Karaletsos et al., 2012) or concentrate on single organs like the green parts of plants (Granier et al., 2006; Hartmann et al., 2011). In addition to the shoot system, the plant root system plays an important role in terms of plant development and yield. The root system architecture (RSA) responds very sensitively to a multitude of stimuli, which is the reason why RSA has become a major target for phenotyping efforts (Armengaud et al., 2009; French et al., 2009; Clark et al., 2011; Lobet et al., 2011; Burton et al., 2012), reviewed in Zhu et al. (2011). While existing RSA software packages require a significant amount of manual curation, we introduce RootDetection which allows fully automatic high-throughput phenotyping of root lengths and a suit of post-processing tools which additionally enable the quantification of hypocotyl lengths.

As already outlined in the general introduction of this thesis, RSA and plant development in general are strongly influenced by the plant hormone auxin (Vanstraelen and Benkova, 2012). It has been shown previously (Delker et al., 2010) that exogenous plant treatment with the endogenous auxin indole-3-acetic acid (IAA), or the synthetic auxins 2,4-dichlorophenoxyacetic acid (2,4-D) and naphthaleneacetic acid (NAA) results in natural variation in root growth inhibition (RGI) in different *A. thaliana* accessions.

On basis of these results, in this study endogenous and synthetic auxins were used in a quantitative genetic approach to unravel the complexity of auxin response pathways in *A. thaliana*. I assessed the natural variation of auxin response traits in two RIL populations for RGI. In addition to RGI, experiments were conducted in this thesis on dark-grown seedlings treated with exogenous auxins, which has been shown to result in hypocotyl growth inhibition (HGI) (Boerjan et al., 1995; King et al., 1995; Thomine et al., 1997). Subsequently, QTL analysis for RGI and HGI identified several genomic regions explaining the measured phenotypic variation and promising QTLs were selected for validation with suitable NILs.

3.2.1 A stable and user-friendly image processing software for high-throughput root length quantification and more

In the first instance of all conducted quantitative genetic experiments, phenotyping of the traits of interest, which implicates screening of hundreds to thousands of plants and their trait characteristics, is mandatory. In this case, RGI as a root system architecture (RSA) trait and HGI were used as a readout for auxin responses in the investigated *A. thaliana* seedlings. To enable reliable, effective and automated quantification of root lengths, we developed the RootDetection software package. In what follows, the main features of RootDetection, including the possibility of RootDetection to measure hypocotyl lengths and a quality assessment of RootDetection, proofing the usability in a high-throughput manner, is described.

Implementation of the software

We have implemented RootDetection in Java (v1.6), providing platform-independence. RootDetection generated data is stored in an SQLite database, which can also be accessed by our additional R package for data visualization and easy data processing in the R environment.

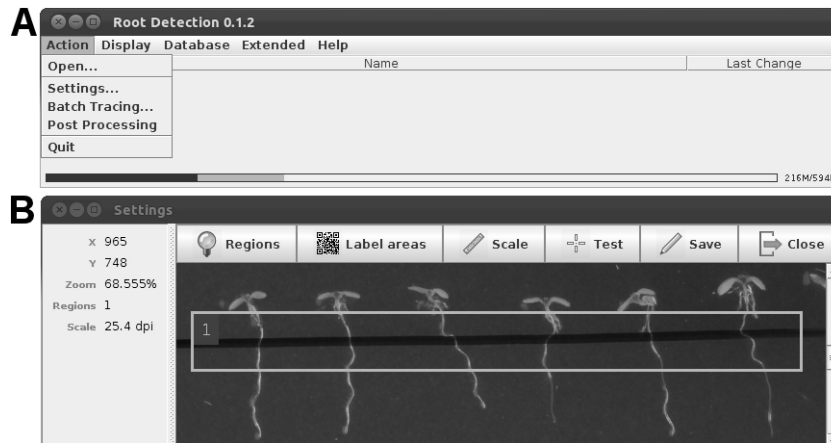


Figure F3.11: Screenshots from RootDetection. (A) Overview. (B) Settings window.

Software features of RootDetection

RootDetection (see Figure F3.11) was built to measure root length of young seedlings grown on agar plates in a fully automatic manner. In addition to automatic root length measurements, standard manual measurement procedures were also implemented, which can be used to measure root or hypocotyl lengths. In addition to classic small-scale experiments, RootDetection provides essential features with respect to high-throughput phenotyping:

- minimal, user-friendly configuration
- fast, fully automatic batch mode for large image sets
- optional manual post-processing of detected roots
- flexible data access

RootDetection is optimized for a commonly used plant transfer assay, in which young seedlings grow on control plates and then are transferred to plates containing screening substances (as conducted for the RGI experiments).

Exemplary workflow for high-throughput root length measurements with RootDetection

All steps described here can be looked up in detail in video tutorials or the online manual of RootDetection (www.labutils.de). An overview of the single steps of the workflow is given in Supplementary Figure S14.

In brief, the experimentator performs the following consecutive steps. First, seeds are germinated and grown on control plates and then seedlings are transferred to plates containing the screening substance of interest. To allow for automatic sample recognition, these plates are labeled with QR-codes generated and readable by RootDetection. Next, the plates are marked with a black start line to enable root start point detection. Subsequent to cultivation of the seedlings for the amount of time given by the experiment, the plates are photographed and the image files are saved in a target folder. After defining barcode label regions and image regions to be measured for the first picture, RootDetection automatically performs root length measurements for all images in the target folder. Finally, the results are exported in an MS Excel compatible format and/or visualized with the RootDetection R package (RRootDetection).

The most powerful feature, which distinguishes RootDetection from available software is the fully automatic batch trace mode that enables the user to analyze an indefinite number of plates with

minimal hands-on time. Here, the user initially specifies three basic settings: (i) region containing roots to be measured, (ii) image scale, and optionally (iii) region setting for barcode recognition (see Figure F3.11 B and Supplementary Figure S14 (5)). After these basic settings have been defined for the first image, the RootDetection algorithm adopts these settings for every single image in the target folder and sequentially performs five steps:

(1) Detection of the start line: A list of start line candidates is formed by evaluating the intensities of all pixels along the left and right vertical boundaries of a region containing roots. Candidates from the left and right boundaries, which have small intensities, are connected to form candidate lines. For each candidate line the sum of pixel intensities is computed and the line with smallest sum is selected as the start line.

(2) Detection of root start points: Intensity gradients are approximated along the detected start line using the Sobel operator. The mid point position of two detected edges (left root border: low to high intensity, right root border: high to low intensity) is then considered as a root start point.

(3) Tracing of root paths: French et al. (2009) first used the condensation algorithm of Isard and Blake (1998) to trace root paths. The particle weight is determined by comparing the pixel intensity at the particle position to a simple "foreground" model. The particle position of each iteration is determined by a model that remembers the previous path orientation and allows for a certain deviation while preferring a straight movement.

(4) Detection of root end points: Tracing of the root path stops as soon as one of several "root end tests" is fulfilled. For most roots an evaluation of the average particle weight at the current iteration step provides the most robust test. Other tests analyze whether total root length, particle spread, or other values are within expected limits. Final root length is determined by calculating the distances between the weighted mean positions of each iteration step (see equation 2.7, 2.8 and 2.9).

(5) Re-evaluation: Lastly, all roots are evaluated to detect outliers or mis-scans. Roots with a length below or above a certain range are marked as failure candidates (FC). The interval of valid lengths is calculated from the mean length of all roots and twice the standard deviation. To avoid negative and large interval boundaries, these values are clamped between an absolute minimum and an absolute maximum length. Note that these boundaries are only used to mark the respective roots as failure candidates, the actual length value and root path is left unchanged.

In addition, RootDetection offers an optional, extensive post-processing mode to easily correct barcode information or to re-evaluate automatically measured roots (see Supplementary Figure S15). In case of incorrect scan results, several tools, including a fast drag-and-drop tool, can be applied to correct or adjust the root path (see Supplementary Figure S15 B+C). A manual measuring tool can be used to measure any kind of line segments like hypocotyl length. The generated data can be exported from the embedded SQLite database into an MS Excel compatible file format or can be accessed via an included R package. The R package also provides the user with basic plot functions to directly analyze the obtained data.

Quality assessment of RootDetection in standard bioassays

RootDetection has been developed to build a fully automatic RSA software and is, at present, capable of quantifying hypocotyl length in a manual mode and the most widely used RSA trait root length in a fully automatic mode. We have tested RootDetection by performing standard root growth assays in response to the phytohormone auxin in a high-throughput manner. As a proof-of-concept, the data obtained by screening 80 *A. thaliana* accessions in response to NAA were processed in detail (see Supplementary Table ST9). Comparing mean root length values of manually curated data (using the RootDetection post processing mode) with automatically measured raw data (only correcting undetected barcode labels and removing FCs) results in an adjusted R^2 value of 0.94427 (see Figure

F3.12). In this data set the 95% confidence interval for errors of measurement ranges from -0.32 to 0.02 mm (see Supplementary Table ST9). RootDetection’s fully automatic batch trace mode returns values comparable to manually curated values, demonstrating its potential for fully automated high-throughput phenotyping. However, for the conducted RGI and HGI experiments for QTL and GWA mapping on top of each automatic root length measurement and the integrated FC removal, a manual post-processing was applied to take care of measurement errors and to reduce standard deviation.

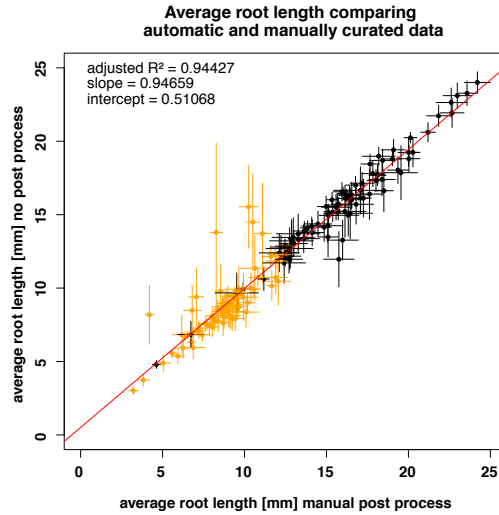


Figure F3.12: Comparing mean root length values of automatically measured raw data with manually curated data. 80 *A. thaliana* accessions have been analyzed by transferring them to plates containing 75nM of the phytohormone naphthylacetic acid (NAA) or unsupplemented control media (orange dots: +NAA, black dots: unsupplemented). Each dot represents mean root length values of 160 analyzed images ($n_{\text{automatic}} = 6 - 16$ roots per image, $n_{\text{manual}} = 9 - 13$ roots per image). The legend denotes values obtained by fitting the observed data by a linear model. X-axis and Y-axis are set to same scale. Error bars denote standard error of the mean (X-direction: manually curated, Y-direction: automatic mode).

RootDetection’s fully automatic batch trace mode proofed to return values comparable to manually curated values (see Figure F3.12) demonstrating its potential for fully automated high-throughput phenotyping. Taken advantage of this, phenotyping for the following QTL analyses and GWAS were performed with the software RootDetection.

3.2.2 QTL analysis for auxin responses in two *Arabidopsis thaliana* recombinant inbred line populations

For the two RIL populations $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$, which have been introduced by Törjék et al. (2006), QTL mapping can unravel the genetic architecture of complex traits (Lapin et al., 2012). To identify QTLs which contribute to RGI and HGI after exogenous auxin treatment, QTL analyses in both populations were performed for the above mentioned three different auxins 2,4-D, IAA and NAA.

Dose response analysis of auxin response traits in two RIL populations

Ideally, QTL analysis with a RIL population is performed under conditions that clearly discriminate the phenotype of the two parental lines. To identify such suitable auxin concentrations for QTL mapping for RGI and HGI in both RIL populations, a core subset was generated. These core subsets represent RILs which maximize the genotypic diversity of the population (van Berloo, 2008). The

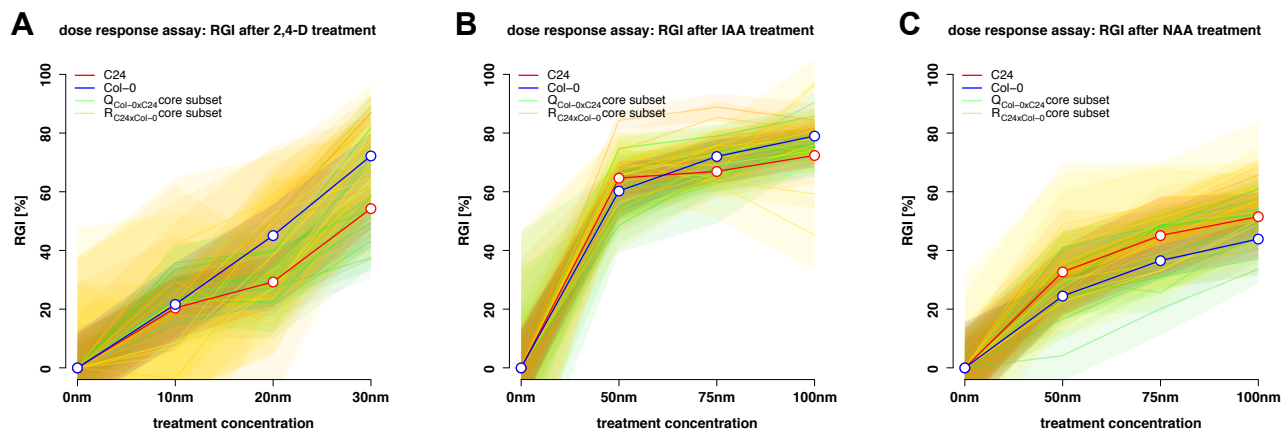


Figure F3.13: Root growth inhibition (RGI) on different auxin concentrations in the core subsets of the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations and the parental lines Col-0 and C24. Phenotypic mean values for each investigated RIL line and the parental *A. thaliana* accessions are shown as solid lines. The different core subset RIL lines are colored greenish for the $Q_{Col-0 \times C24}$ subset and yellowish for the $R_{C24 \times Col-0}$ subset, the parental lines are colored in blue (Col-0) and red (C24). The transparent area highlights the % standard deviation for each evaluated plant line. (A) 2,4-D treatment, (B) IAA treatment, (C) NAA treatment.

generated core subsets (data not shown) were tested for RGI and HGI on different phytohormone concentrations. To be able to combine the results for these RIL populations with another RIL population (*Bay-0* \times *Shadara*), the range of the tested phytohormone concentrations needed to be also suitable for this RIL population (personal communication with Anja Raschke, unpublished data).

Before starting the proper QTL mapping experiment, the experimentator has to consider two things. First of all, the measured phenotypic values and their distribution has to be known, which can be binary, ordinal or continuous metric values. In the case of growth inhibition, continuous metric values can be obtained within a segregating population. Growth inhibition can reach values up to 100%. Second, due to the fact that in a segregating population parental alleles are recombined during the propagation process, new allele combinations can lead to a phenomenon called "transgression" in which offsprings show a broader phenotypic variation than their parents. Considering transgression, the continuous metric mode of growth inhibition and the upper limit of 100% growth inhibition, the tested phytohormone concentrations for parental lines and RIL lines should be well below complete growth inhibition. They should also allow a broad spectrum of phenotypic values to maximize the detectable phenotypic variation in the RIL population.

The RGI dose response analyses for both RIL populations and the three different auxin treatments show that transgression can be observed even in the core subsets (see Figure F3.13). To optimize the comparison between the two analyzed RIL populations, a single concentration for each auxin treatment was determined to be used for QTL mapping in both RIL populations. For the 2,4-D treatment a concentration of 20 nM was determined. Although 30 nM shows a better phenotypic spectrum for the $Q_{Col-0 \times C24}$ RIL population, for the $R_{C24 \times Col-0}$ RIL population the phenotypic values for the extreme plant lines of the core subset are close to 100% RGI (see Figure F3.13 A). To avoid losing even more extreme plant lines in the full set of the $R_{C24 \times Col-0}$ RIL population, the decision for 20 nM was made. For IAA and NAA treatment the core subset of the $Q_{Col-0 \times C24}$ RIL population shows a shift towards lower RGI values, whereas the core subset of the $R_{C24 \times Col-0}$ RIL population shows a shift towards higher RGI values. For further IAA root growth assays an even smaller concentration of 40 nM than shown in figure F3.13 B was defined. For NAA root growth assays a concentration of 75 nM was chosen for the QTL analysis in both RIL populations.

For HGI the % standard deviations for each plant line show higher values than for the RGI dose response assays (see Figure F3.14). Similar to the RGI results, here transgression is also prevalent. For the 2,4-D treatment 375 nM and 500 nM seem to be suitable for further QTL mapping, also for IAA

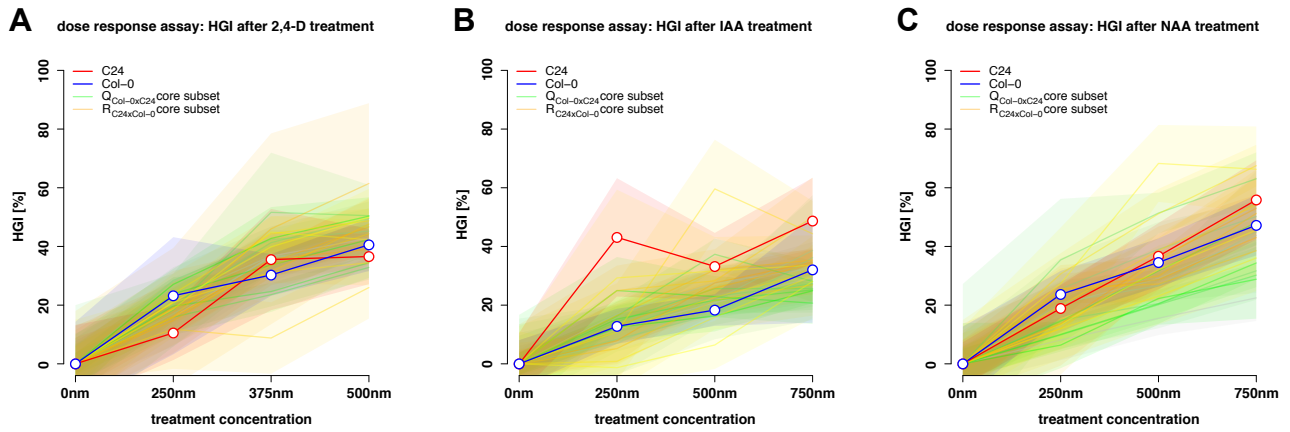


Figure F3.14: Hypocotyl growth inhibition (HGI) on different auxin concentrations in the core subsets of the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations and the parental lines Col-0 and C24. Phenotypic mean values for each investigated RIL line and the parental *A. thaliana* accessions are shown as solid lines. The different core subset RIL lines are colored greenish for the $Q_{Col-0 \times C24}$ subset and yellowish for the $R_{C24 \times Col-0}$ subset, the parental lines are colored in blue (Col-0) and red (C24). The transparent area highlights the % standard deviation for each evaluated plant line. (A) 2,4-D treatment, (B) IAA treatment, (C) NAA treatment.

and NAA 500 nM and 750 nM seem suitable for QTL mapping according to the observed phenotypic spectrum. Here, for a better comparison of the HGI results of the RIL populations $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ with the RIL population $Bay-0 \times Shadara$, we decided to use 375 nM for 2,4-D and 500 nM for IAA and NAA (personal communication with Anja Raschke).

By evaluating RGI and HGI in reduced subsets of both RIL populations, which still have high genotypic diversity, suitable auxin concentrations were defined to be further used in the proper QTL experiments including the vast majority of all RILs.

Descriptive statistics for root and hypocotyl growth phenotypes in two RIL populations

The $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations were subsequently screened for root growth inhibition (RGI) (20 nM 2,4-D, 40 nM IAA, 75 nM NAA) and hypocotyl growth inhibition (HGI) (375 nM 2,4-D, 500 nM IAA, 500 nM NAA) after treatment with the different auxins. For each trait and RIL population the number of analyzed RILs varied from 133 to 211 plant lines, depending on how many plant lines had to be discarded because of insufficient germination rates or bacterial or fungal contamination prior to the proper statistical analysis. The remaining plant lines were further reduced to remove extreme outliers which fall above a predefined z-score threshold of 3. A summary of all conducted growth inhibition assays is given in Table T3.3.

As already seen for the core subsets, transgression is observed for all RGI traits in both RIL populations (see Figure F3.15). In the 2,4-D treated $Q_{Col-0 \times C24}$ RIL population, the RIL phenotypic mean value of $28.69 \pm 11.33\%$ RGI is slightly lower than for the $R_{C24 \times Col-0}$ RIL population with a value of $35.04 \pm 15.91\%$ RGI. The phenotypic mean values of the parental lines were calculated based on four experimental replicates including the results of the genome-wide association (GWA) experiment described in the next results section. For the 2,4-D treatment the parental phenotypic mean values fall into different RGI classes but show an overlap when considering the % standard deviations, $30.88 \pm 11.48\%$ for Col-0 and $21.13 \pm 12.49\%$ for C24 (see Figure F3.15 A and Table T3.3). For IAA and NAA treatments both parental lines fall into the same phenotypic RGI classes (approximately 60% RGI for IAA treatment and 45% RGI for NAA treatment). RGI classes have been defined in the histograms by 10% growth inhibition classes (see Figure F3.15 B + C and Table T3.3). After IAA

Table T3.3: Descriptive statistics of auxin response traits in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations.

Trait	Treatment	Parental lines		Recombinant inbred lines		Mean \pm SD ^e	Range (Min – Max) ^e	SW test ^f	Skewness ^e	Kurtosis ^e	BSH \pm SD ^g
		Col-0 ^c	C24 ^c	Population	Line number ^d						
RG1 ^a	2,4-D 20nM	30.88 \pm 11.48	21.13 \pm 12.49	Q _{Col-0xC24}	167 (0)	28.69 \pm 11.33	0.92 – 57.28	0.58205	0.057	-0.267	0.714 \pm 0.021
		63.44 \pm 6.44	60.09 \pm 4.08	R _{C24xC0l-0}	169 (0)	35.04 \pm 15.91	0.97 – 72.27	0.13564	-0.051	-0.583	0.858 \pm 0.011
		43.26 \pm 2.93	46.82 \pm 3.06	Q _{Col-0xC24}	195 (0)	42.68 \pm 11.66	14.26 – 73.80	0.24097	0.258	-0.094	0.804 \pm 0.016
HGI ^b	2,4-D 375nM	43.26 \pm 2.93	46.82 \pm 3.06	R _{C24xC0l-0}	189 (2)	54.87 \pm 10.19	23.39 – 77.56	0.00069	-0.656	0.586	0.837 \pm 0.012
		27.31 \pm 6.46	25.23 \pm 3.78	Q _{Col-0xC24}	211 (2)	45.55 \pm 7.00	24.16 – 61.08	0.08098	-0.253	-0.278	0.690 \pm 0.019
		17.67 \pm 2.18	34.34 \pm 9.01	R _{C24xC0l-0}	193 (2)	45.37 \pm 7.32	26.99 – 60.62	0.07289	-0.242	-0.274	0.660 \pm 0.023
NAA 500nM	2,4-D 500nM	33.00 \pm 8.93	45.25 \pm 10.44	Q _{Col-0xC24}	203 (1)	34.47 \pm 8.16	14.77 – 56.39	0.73988	0.016	-0.344	0.738 \pm 0.018
		17.67 \pm 2.18	34.34 \pm 9.01	R _{C24xC0l-0}	189 (2)	26.88 \pm 8.29	4.03 – 49.42	0.24995	0.210	-0.299	0.669 \pm 0.023
		30.88 \pm 11.48	21.13 \pm 12.49	Q _{Col-0xC24}	176 (0)	31.89 \pm 13.60	5.40 – 70.08	0.02341	0.414	-0.232	0.706 \pm 0.021
NAA 75nM	2,4-D 75nM	30.88 \pm 11.48	21.13 \pm 12.49	R _{C24xC0l-0}	133 (0)	30.63 \pm 13.11	5.52 – 61.38	0.01575	0.162	-0.838	0.727 \pm 0.022
		63.44 \pm 6.44	60.09 \pm 4.08	Q _{Col-0xC24}	201 (0)	34.68 \pm 11.09	4.87 – 61.21	0.27148	0.136	-0.489	0.790 \pm 0.016
		43.26 \pm 2.93	46.82 \pm 3.06	R _{C24xC0l-0}	189 (1)	27.94 \pm 11.80	4.37 – 61.35	0.00024	0.609	0.027	0.774 \pm 0.017

^a Root growth inhibition. Phenotypic values are indicated in % inhibition.

^b Hypocotyl growth inhibition. Phenotypic values are indicated in % inhibition.

^c Mean and standard deviation for parental lines of 4 biological replicates.

^d RILs used for QTL analysis. Numbers in brackets indicate RILs outside z-score threshold of 3.

^e Values determined after removing RILs outside z-score threshold of 3. For details how phenotypic values are evaluated see methods.

^f Shapiro-Wilk test calculated after removing RILs outside z-score threshold of 3.

^g Broad-sense heritability calculated after removing RILs outside z-score threshold of 3. For details how BSH is evaluated see methods.

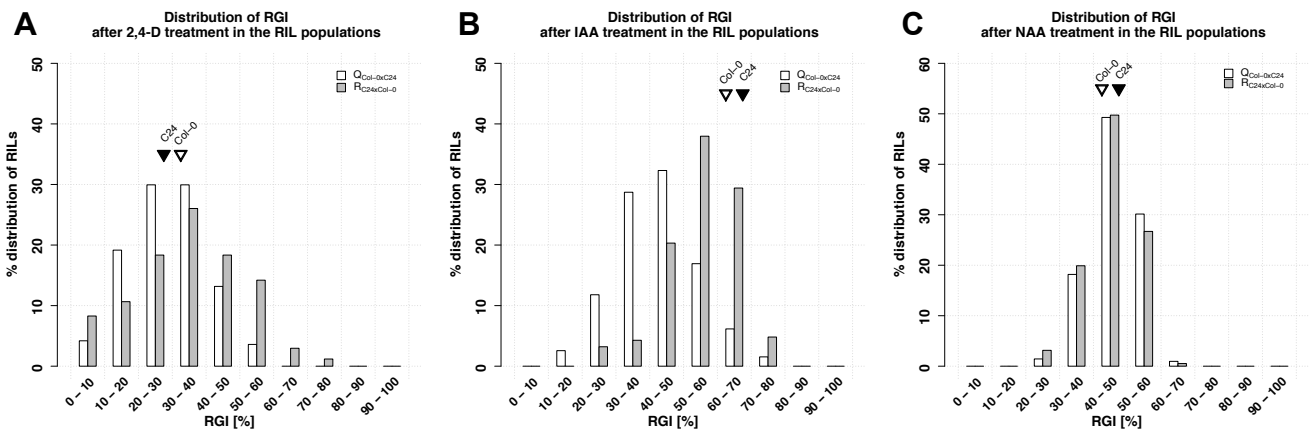


Figure F3.15: Percent frequency distribution of RGI for the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations after different auxin treatments. Histograms show the percent frequency distribution of RGI in two RIL populations after treatment with (A) 20 nM 2,4-D, (B) 40 nM IAA and (C) 75 nM NAA. Filled black triangle (C24) and open triangle (Col-0) indicate RGI category into which parental phenotypic mean values fall.

treatment, the $Q_{Col-0 \times C24}$ RIL population also shows a lower RIL phenotypic mean value of $42.68 \pm 11.66\%$ versus $54.87 \pm 10.19\%$ of the $R_{C24 \times Col-0}$ RIL population, indicating a possible maternal or paternal effect. After NAA treatment, both RIL populations show equal RIL phenotypic mean RGI values of $\sim 45\%$ and also share the same phenotypic range of about 25 to 60% RGI (see Table T3.3). For 2,4-D and IAA, the phenotypic ranges of the RIL populations are broader compared to the NAA treatment. In both cases the RIL population $Q_{Col-0 \times C24}$ shows a slight shift to lower RGI ranges compared to the $R_{C24 \times Col-0}$ RIL population. Root growth assays are normally distributed according to the Shapiro-Wilk test with exception of the RIL population $R_{C24 \times Col-0}$ after IAA treatment. To get an estimate of how much phenotypic variance can be explained by the genotype, the broad-sense heritability (BSH, H^2) was estimated. H^2 showed considerably high values for RGI ranging from 0.660 ± 0.023 to 0.858 ± 0.011 (see Table T3.3). This H^2 values indicate that the measured phenotypic variation relies on strong genotypic factors and only on weak environmental factors. In this case, the high H^2 values reflect the controlled experimental conditions that have been chosen and exhibit better chances to unravel chromosomal regions and underlying genes causative for the analyzed traits by QTL mapping.

Auxin response traits leading to HGI do likewise show a high BSH ranging from 0.669 ± 0.023 to 0.790 ± 0.016 (see Table T3.3). For the HGI assays three experiments were not normally distributed according to the Shapiro-Wilk test ($\alpha = 0.05$). For the 2,4-D treatment the parental phenotypic mean values fall into the same RGI classes (25% HGI) (see Figure F3.16 A). For IAA and NAA treatments the phenotypic mean value of Col-0 ($17.67 \pm 2.18\%$ HGI and $33.00 \pm 8.93\%$ HGI) is in both cases lower than for C24 ($34.34 \pm 9.01\%$ HGI and $45.25 \pm 10.44\%$ HGI) (see Figure F3.16 B + C). The phenotypic RIL mean value for the 2,4-D treatment in RIL $Q_{Col-0 \times C24}$ is shifted to higher HGI values ($34.47 \pm 8.16\%$ HGI) compared to the mean value of RIL population $R_{C24 \times Col-0}$ ($26.88 \pm 8.29\%$ HGI). Likewise, the range is slightly higher for the $Q_{Col-0 \times C24}$ RIL population. For these observed shifts between the two RIL populations maternal and paternal effects could play a role. For NAA treatment the mean values are comparable to the 2,4-D treatment with $34.68 \pm 11.09\%$ HGI for $Q_{Col-0 \times C24}$ and $27.94 \pm 11.80\%$ HGI for $R_{C24 \times Col-0}$. Considering the % standard deviations, there is an overlap of all phenotypic mean values when both RILs are compared. The differences in phenotypic mean distributions can be caused by strong allelic effects (one parental allele has a strong decreasing effect on the measured phenotypic value which would cause a one sided shift of the distribution). Alternatively, maternal and paternal effects might have an influence. In addition to these genotypic explanations, another possibility might simply be variability between experiments.

The descriptive statistics for auxin response traits in both RIL populations revealed that parental lines

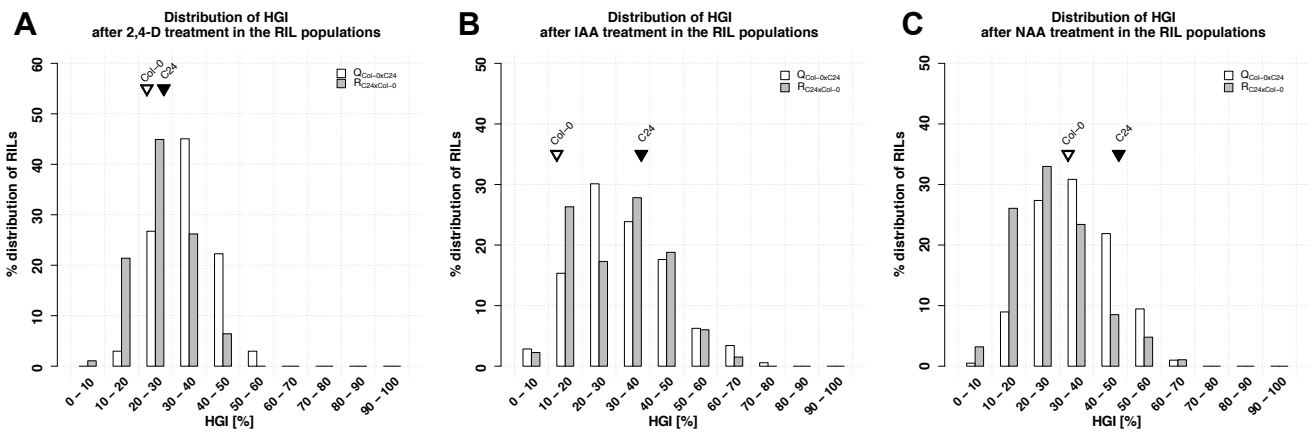


Figure F3.16: Percent frequency distribution of HGI for the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations after different auxin treatments. Histograms show the percent frequency distribution of HGI in two RIL populations after treatment with (A) 375 nM 2,4-D, (B) 500 nM IAA and (C) 500 nM NAA. Filled black triangle (C24) and open triangle (Col-0) indicate HGI category into which parental phenotypic mean values fall.

fall into different phenotypic classes only in three out of six cases, which is not ideal for the validation of QTLs for these traits. In four out of twelve cases, the phenotypic data was not normal distributed and has to be analyzed by an additional non-parametric QTL mapping method. However, the H^2 values were high and exhibit good chances to detect phenotypic to genotypic correlations.

In the next section, one additional aspect of the phenotypic data is considered prior the statistical part of QTL mapping. To assess whether the same genomic regions affect several of the investigated auxin response traits simultaneously, I will next perform a correlation analysis on the phenotypic data.

Phenotypic correlations among root and hypocotyl growth phenotypes in two RIL populations

To examine possible relationships among the auxin response traits within each RIL population, the pairwise Pearson correlation coefficients were calculated as shown in Table T3.4. Surprisingly, no significant correlation could be detected, when the RGI assays for both RIL populations are compared. For the $Q_{Col-0 \times C24}$ RIL population the 2,4-D root and hypocotyl growth inhibition are significantly correlated, but with a rather low Pearson correlation coefficient of 0.236.

For both RIL populations the pairwise correlation coefficients are higher for HGI assays. Here, the comparison of the 2,4-D and NAA treatment showed a significant correlation coefficient of 0.419 for the $Q_{Col-0 \times C24}$ RIL population and 0.418 for the $R_{C24 \times Col-0}$ RIL population. For both populations also a significant correlation for NAA and IAA could be observed, but again with a rather low correlation coefficient of 0.246 for the $Q_{Col-0 \times C24}$ RIL population and 0.247 for the $R_{C24 \times Col-0}$ RIL population (see Table T3.4).

The fact that phenotypic correlations between the auxin response traits were rather low for both RIL populations suggests distinct genetic causes for these traits in the genetic background of Col-0 and C24. Nevertheless, the observed transgression and high values of H^2 were promising to identify QTLs contributing to the phenotypic variation in these traits without suspecting high overlap between possible QTL regions for the different auxin treatments (2,4-D, IAA and NAA).

Table T3.4: Phenotypic correlations among auxin response traits in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations.

$Q_{Col-0 \times C24}$		root			hypocotyl		
		2,4-D	IAA	NAA	2,4-D	IAA	NAA
root	2,4-D						
	IAA	0.027					
	NAA	-0.114	0.033				
hypocotyl	2,4-D	0.236*	0.060	0.112			
	IAA	0.178	0.124	0.149	0.158		
	NAA	0.057	-0.038	0.099	0.419***	0.246*	

$R_{C24 \times Col-0}$		root			hypocotyl		
		2,4-D	IAA	NAA	2,4-D	IAA	NAA
root	2,4-D						
	IAA	-0.034					
	NAA	-0.053	0.166				
hypocotyl	2,4-D	0.100	-0.093	0.059			
	IAA	0.153	-0.062	0.035	0.103		
	NAA	-0.110	0.005	0.104	0.418***	0.247*	

Pairwise pearson correlation coefficients of root and hypocotyl traits within one RIL population.

***, **, * significant after Benjamini-Hochberg correction at $P < 0.001$, $P < 0.01$, $P < 0.05$, respectively.

Single- and two-QTL models applied on auxin response traits in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations

Using standard interval mapping (IM) and composite interval mapping (CIM) to identify main effect QTLs in the two evaluated RIL populations phenotyped for auxin response traits (in total six traits per RIL population) resulted in the detection of 22 significant QTLs with interval mapping and 13 significant QTLs with composite interval mapping.

As one example, Figure F3.17 illustrates LOD profiles obtained by IM and CIM of both populations for RGI after NAA treatment on chromosome five. In addition to the LOD profiles (see Figure F3.17 A + C), the phenotype by genotype plots (see Figure F3.17 B + D) can visualize the QTL effects at a certain marker position. A complete list of the significant detected *main effect QTLs* using a single QTL model for each trait and RIL population is given in Supplementary Table ST10. In case for one RIL population and a given auxin response trait the phenotypic value distribution was not normally distributed (according to Shapiro-Wilk test described in Table T3.3), an additional non-parametric single QTL mapping model was applied on this data. The calculated LOD plots for the three examined auxins and their effects on RGI and HGI and the significant thresholds calculated by 1000 permutations for both RIL populations are visualized in Supplementary Figure S17 and Supplementary Figure S18. In most cases the same *main effect QTLs* were detected with both the IM and CIM method showing the LOD peak at almost the same chromosomal positions. In cases where the LOD peak of the IM

method hardly crosses the determined significant threshold, the CIM method rejected these QTLs (see Supplementary Figure S17 B + C + E + F for RGI and Supplementary Figure S18 A + B + D + E for HGI). Furthermore, the CIM method detected an additional *main effect QTL* for HGI on chromosome 5 at 65.5 cM in the $R_{C24 \times Col-0}$ RIL population after IAA treatment (see Supplementary Figure S18 E). This can be explained by the consideration of covariates by the CIM method (for a detailed description see Broman and Sen (2009) pages 205-206).

The non-parametric QTL mapping model confirmed in all cases the QTL positions identified with the IM model, suggesting that IM can be applied even on the not normally distributed traits. Only in one case (HGI and the $R_{C24 \times Col-0}$ RIL population after NAA treatment) an additional weak QTL could be detected by the non-parametric mapping model. The highest LOD peak value of 13.63 for *main effect QTL* analysis was observed for the $Q_{Col-0 \times C24}$ RIL population and the NAA treatment on chromosome five on position 77 cM as shown in Figure F3.17 A. Figure F3.17 B shows the LOD profile obtained in the $R_{C24 \times Col-0}$ RIL population with overlapping 95% CIs. As described in detail in the materials and methods section, the significant QTLs detected by IM were then used as an initial QTL model to find multiple QTL models with the `stepwiseqtl` function of the R/qtl software (Broman et al., 2003).

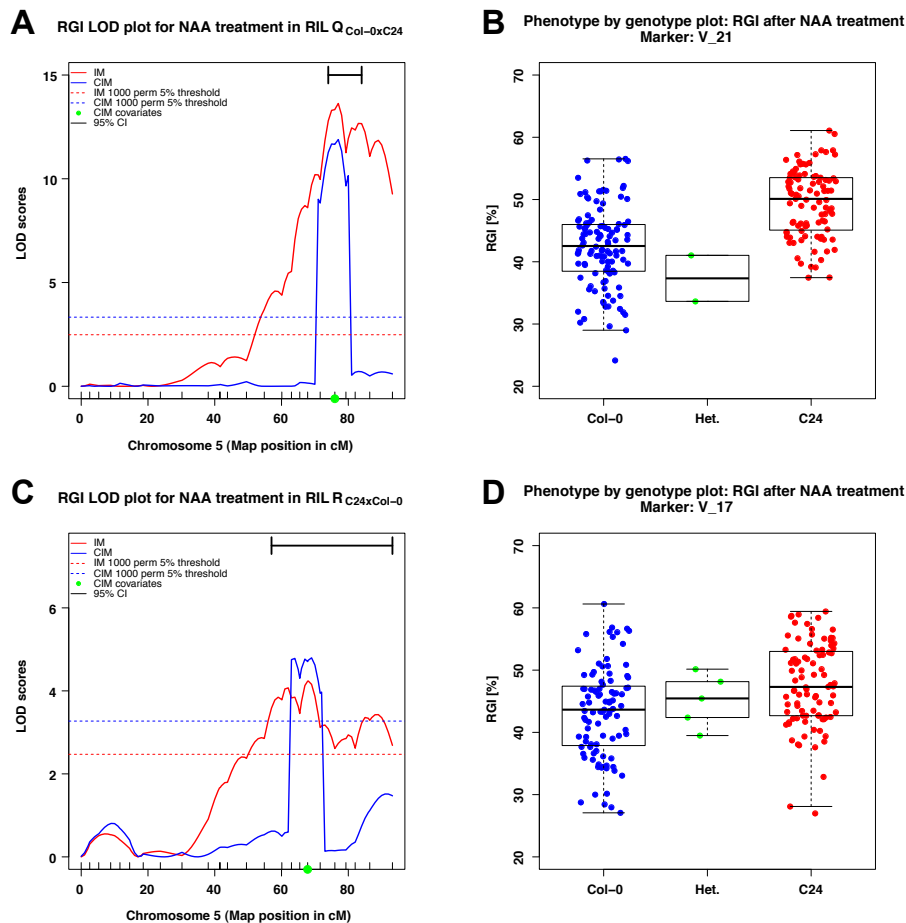


Figure F3.17: QTL associated with RGI after NAA treatment in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL population. LOD profiles from standard interval and composite interval mapping are shown for chromosome five for both RIL populations including the 5% significance thresholds (dashed lines) as calculated by 1000 permutations. Green dots indicate covariates selected by a forward approach. The black line on the top of each plot indicates 95% confidence interval based on results from a multiple QTL model (A + C). Phenotype by genotype plots for the nearest marker of the QTL LOD peak position based on the multiple QTL model are shown as boxplots and grouped by their corresponding genotype: either homozygous for the Col-0, homozygous for the C24 allele or heterozygous (Het.) for the $Q_{Col-0 \times C24}$ (B) or the $R_{C24 \times Col-0}$ RIL population (D).

To unravel possible additive or epistatic interactions between chromosomal regions which could contribute to the phenotypic variation of the investigated complex auxin response traits, a *two-QTL genome scan* was performed for each trait and each RIL population. To detect significant interactions of two evaluated loci, 10000 permutations were used for each trait and RIL population.

Figure F3.18 A shows a *two-QTL genome scan* for RGI for the $Q_{Col-0 \times C24}$ RIL population after NAA treatment. Here, chromosome one and five are divided by one cM steps and the corresponding parental alleles at these positions were calculated. As outlined in the materials and methods section, one can test each of these chromosomal positions for additive or epistatic interactions. Figure F3.18 A shows an additive interaction for which some conditions have to be true. For an additive interaction the additive M_{add} QTL model LOD score (below diagonal) and the conditional additive M_{av1} QTL model LOD score (above diagonal) need to cross the LOD score threshold. In Figure F3.18 A chromosomal positions were subsequently tested for pairwise additive (M_{add} QTL model) or conditional additive interactions (M_{av1} QTL model) and a significant additive interaction of position 92 of chromosome one and position 77 of chromosome five could be detected ($M_{add} = 16.068$ and $M_{av1} = 2.439$).

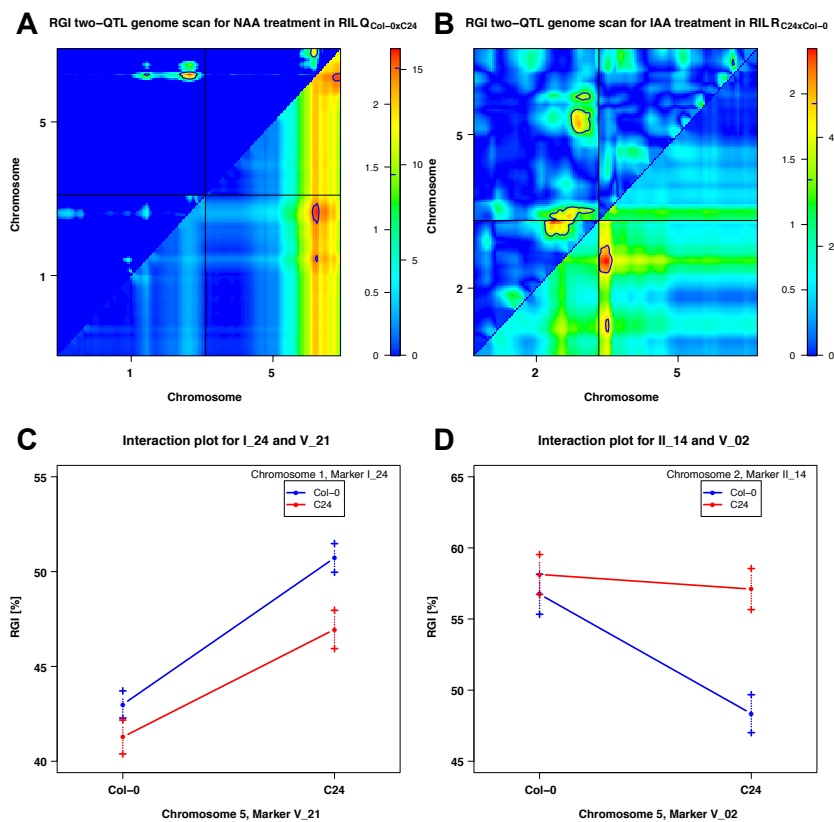


Figure F3.18: Example of significant *two-QTL genome scan* interactions in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations for RGI. A *scantwo* plot with x- and y-axis indicating cM positions along the given chromosomes and the calculated LOD values for the additive QTL model (M_{add} below diagonal) and conditional additive QTL model (M_{av1} above diagonal) is shown for RGI after NAA treatment in the $Q_{Col-0 \times C24}$ RIL population (A). LOD values for the additive QTL model (M_{add} below diagonal) and the epistatic interaction QTL model (M_{int} above diagonal) are shown for RGI after IAA treatment in the $R_{C24 \times Col-0}$ RIL population (B). 10000 permutations were used to calculate significance thresholds (see also Supplementary Table ST11 and materials and method section for a detailed description). Colored scales on the right indicate observed LOD value spectrum, blue tick marks (A + B) highlight 1-LOD below LOD maximum as used to draw contours in the *scantwo* plot. Additive interaction of two markers (C) and possible epistatic interaction (D) of two markers nearest to positions indicated in the *scantwo* summary table (see Supplementary Table ST11) are shown as *effectplots*. Y-axes denote RGI in % of groups either having Col-0 or C24 allele at a given marker locus. Error bars denote \pm single standard error of the group phenotypic mean.

To illustrate how these interactions influence the observed phenotypic values, Figure F3.18 B shows an effect plot for the nearest markers for both positions (I_24 and V_21). The phenotypic mean value for the analyzed RILs, which carry the Col-0 allele on both markers is $42.95 \pm 0.71\%$ RGI (lower left blue dot). Likewise, when marker I_24 is Col-0 and marker V_21 is C24, the phenotypic mean value for this set of RILs is $50.72 \pm 0.75\%$ RGI (upper right blue dot). If marker I_24 is C24 and marker V_21 is Col-0, the phenotypic mean value for this RIL group is $41.32 \pm 0.90\%$ RGI (lower left red dot). In the group of RILs which carry on both markers the C24 allele, the phenotypic mean value is $46.95 \pm 1.01\%$ RGI (upper right red dot). The effect of the chromosome five locus (V_21) is the same for each of the two genotypes at the chromosome one locus (I_24), and vice versa, which can be seen by the parallel blue and red line in Figure F3.18 B. Hence, this observed effect for the markers I_24 and V_21 indicates that both loci act approximately additively (see page 227 in Broman and Sen (2009)).

In Figure F3.18 C a *two-QTL genome scan* for RGI for the $R_{C24 \times Col-0}$ RIL population after IAA treatment is shown. Here, in addition to a significant additive interaction ($M_{add} = 5.620$ and $M_{av1} = 2.759$) as explained above, a slight epistatic interaction of position 51 on chromosome two and position 2 of chromosome five can be seen ($M_{fv1} = 4.379$ and $M_{int} = 1.620$). To be a true epistatic interaction, the full M_{full} QTL model LOD score and the interaction M_{int} QTL model LOD score need to cross the LOD score threshold. Here, this is not true but a slight epistatic interaction can not be rejected, since the conditional full M_{fv1} QTL model LOD score is above the LOD score threshold and the interaction M_{int} QTL model LOD score is rather high.

More clearly, this slight epistatic interaction can be illustrated by the effect plot of the nearest markers II_14 and V_02 for both positions (see Figure F3.18 D). The phenotypic mean value for the analyzed RILs which carry the Col-0 allele on both markers is $56.75 \pm 1.41\%$ RGI (upper left blue dot). When marker II_14 is Col-0 and marker V_02 is C24, the phenotypic mean value decreases to $48.41 \pm 1.33\%$ RGI (lower right blue dot). If marker II_14 is C24 and marker V_02 is Col-0, the phenotypic mean value for this RIL group is $58.16 \pm 1.40\%$ RGI (upper left red dot). In the group of RILs, which carry the C24 allele on both markers, the phenotypic mean value is $57.07 \pm 1.45\%$ RGI (upper right red dot). Hence, the chromosome five locus (V_02) has only an effect in the presence of the Col-0 genotype at the chromosome two locus (II_14). Similarly, the chromosome two locus (II_14) has only an effect in the presence of the C24 genotype at chromosome five. Again, only individuals that are Col-0 at the chromosome two locus (II_14) and C24 at the chromosome five locus (V_02) show decreased RGI. The other three genotype groups have similar high RGI values, which can be seen by the blue and red line drifting apart in Figure F3.18 D. Hence, this observed effect for the markers II_14 and V_02 indicates an epistatic interaction between the markers (see page 227 in Broman and Sen (2009)).

An even stronger additive interaction was observed for HGI after NAA treatment in the $Q_{Col-0 \times C24}$ RIL population between position 42 of chromosome two and position 56 on chromosome five (see Supplementary Figure S19 A + B). Another weak epistatic interaction for HGI after NAA treatment in the $R_{C24 \times Col-0}$ RIL population between position 51 on chromosome three and position 51 on chromosome four is shown in Supplementary Figure S19 C + D. A summary of the complete *two-QTL genome scan* analysis is given in Supplementary Table ST10. In total, 21 significant additive interactions could be found by the *two-QTL genome scan*, two out of 21 show a slight epistatic interaction effect. Among these, the interaction for HGI after NAA treatment in the $R_{C24 \times Col-0}$ RIL population is most pronounced. All identified significant interactions are also highlighted in Figure F3.19.

Taken together, at least one *main effect QTL* for each of the auxin response traits in both RIL populations could be detected by IM method. The CIM method rejected the weak *main effect QTLs* for RGI IAA in both RIL populations. The *main effect QTLs* together with the detected additive interactions showed once again the complexity of auxin response traits for the investigated RIL populations. To further evaluate the complex nature of the auxin response traits, the detected *main effect QTLs* were taken as starting points for multiple QTL mapping to get an estimate how much phenotypic variation could be explained by complex QTL models.

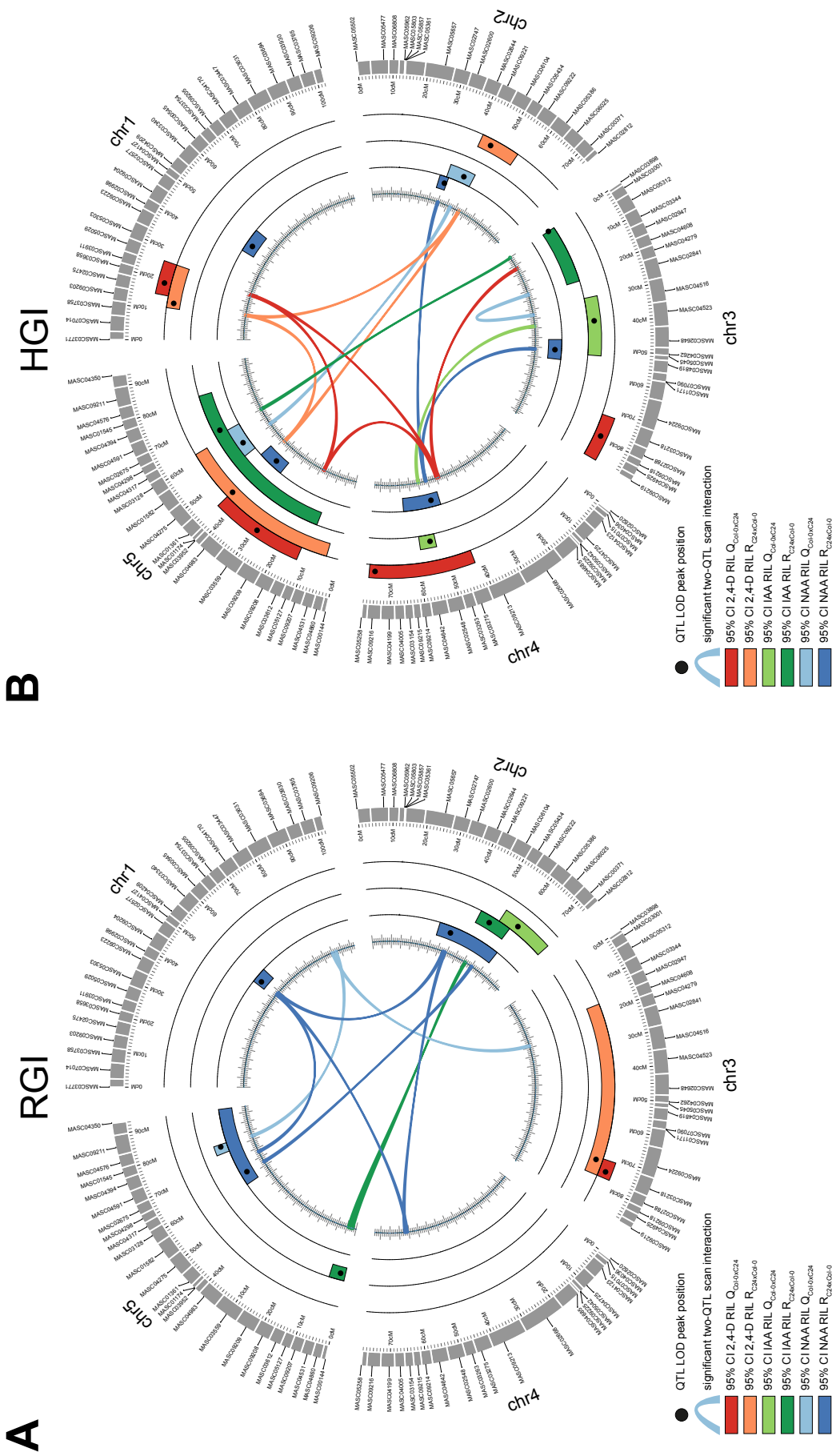


Figure F3.19: QTL loci determined by multiple QTL mapping for auxin response traits in the $Q_{Col-0} \times C_{24}$ and $R_{C24} \times Col-0$ RIL populations. The circles represent the *A. thaliana* genetic map based on marker data, which is equivalent for both RIL populations (Törjék et al., 2006; Meyer et al., 2010). For each chromosome the position is given in centimorgan (cM) and all 113 markers are plotted next to the specified chromosome IDs. In total, 9 QTLs for auxin response traits influencing RGI (A) and 18 QTLs influencing HGI (B) could be detected by multiple QTL mapping. Each QTL LOD peak position is highlighted by a black dot and the corresponding 95% confidence interval is plotted as colored boxes according to the analyzed traits and RIL population ($Q_{Col-0} \times C_{24}$: RGI / HGI: 20 nM / 375 nM 2,4-D red, 40 nM / 500 nM IAA light-green, 75 nM / 500 nM NAA light-blue; $R_{C24} \times Col-0$: RGI / HGI: 20 nM / 375 nM 2,4-D light-red, 40 nM / 500 nM IAA green, 75 nM / 500 nM NAA blue). Significant interactions between chromosomal positions according to a *two-QTL genome scan* are plotted as colored lines with same colors for treatment and RIL population.

Multiple QTL mapping results for auxin response traits

The multiple QTL mapping analysis detected a total of nine QTLs and one QTL interaction for RGI experiments. For the HGI experiments 18 QTLs and one QTL interaction were observed and are listed in Table T3.5. For each trait and RIL population the detected multiple QTL model and the corresponding estimated phenotypic variance explained by the full QTL model was further evaluated for each single QTL and QTL interaction to get estimates about their contribution to phenotypic variance and to get estimates about their additive effect. To find possible overlaps of the detected QTLs between the three auxin treatments as well as for the two RIL populations, the QTL LOD peak positions, their 95% confidence interval and the observed *two-QTL genome scan* interactions were plotted as a circle plot in Figure F3.19.

First, the RGI results obtained by the multiple QTL mapping approach will be discussed. As expected, QTLs detected for each trait in both RIL populations show a high concordance not only on QTL positioning (see Figure F3.19 A) but also in LOD score height (see Table T3.5). For the 2,4-D treatment (see reddish boxes Figure F3.19 A) the detected QTLs r24D-III.1 and r24D-III.2 map to nearly the same position on chromosome three (72 cM versus 74.872 cM). Both QTLs have the same negative additive effect direction so that the C24 allele decreases RGI in both RIL populations. The 95% confidence interval (CI) region overlaps, but clearly the CI region for r24D-III.2 has a smaller range compared to r24D-III.1, which can be explained by the LOD profile on chromosome three for the $R_{C24 \times Col-0}$ RIL population (see Supplementary Table ST10 and Supplementary Figure S16). Unfortunately, the LOD scores of 4.131 and 4.902 are rather low and the phenotypic variance explained is $\sim 10\%$ for r24D-III.1 and $\sim 12\%$ for r24D-III.2, which make it difficult to validate this QTL. However, the *a priori* CG screen between the markers III.17 and III.19 as given in Supplementary Table ST12 detected 12 CGs of which five have non-synonymous changes between Col-0 and C24. With *PLEITROPIC DRUG RESISTENCE 9 (PDR9)* and *P-GLYCOPROTEIN 20 (PGP20)* two auxin transporters are affected between the two accessions and it needs to be investigated if this altered amino acid sequences could explain the phenotypic differences seen for RGI after 2,4-D treatment in Col-0 and C24.

For the IAA treatment (see greenish boxes Figure F3.19 A), likewise, only small effect QTLs could be detected explaining about 6% for the $Q_{Col-0 \times C24}$ RIL population and about 16% for the $R_{C24 \times Col-0}$ RIL population. Interestingly, there is evidence for a weak epistatic interaction of the QTLs rIAA-II.1 and rIAA-V.1. If one considers the single QTL effects for both QTLs, rIAA-II.1 has a positive additive effect, indicating that the Col-0 allele decreases RGI, whereas rIAA-V.1 shows a negative additive effect indicating that the C24 allele decreases RGI. Considering both nearest markers II.14 and V.02 as indicated in Figure F3.18 D, RGI is decreased in presence of the Col-0 allele at the II.14 marker and C24 at the V.02 marker with a stronger influence of the Col-0 allele. However, the rather small LOD scores in both RIL populations makes it once more difficult to validate this QTL.

For the NAA treatment regarding RGI (see blueish boxes Figure F3.19 A) the highest LOD score of all conducted QTL assays in the analyzed RIL populations could be observed. In the $Q_{Col-0 \times C24}$ RIL population, evidence for a single QTL (rNAA-V.2) on chromosome five at position 77 was emphasized by the composite interval and multiple QTL mapping methods (see Figure F3.18 A). rNAA-V.2 explains about 25% of the measured phenotypic variation with a positive additive effect. Here, the Col-0 allele present at marker V.21 decreases RGI, which is illustrated in Figure F3.18 B. For the $R_{C24 \times Col-0}$ RIL population a more complex QTL model for NAA treatment explaining about 23% of the phenotypic variation was found with smaller individual effects of the single QTLs. Also here the QTL rNAA-V.1 on chromosome five shows the highest contribution to the explained phenotypic variation of about 10% with a broader 95% CI than its counterpart rNAA-V.2 for the $Q_{Col-0 \times C24}$ RIL population. The good overlap of rNAA-V.1 and rNAA-V.2 with same additive effect signature in both RIL populations proposed a good candidate for QTL validation which is shown in the next subsection.

Table T3.5: Summary of QTL detected by multiple QTL mapping of auxin response traits in the RIL $Q_{Col-0} \times C_{24}$ and $R_{C_{24} \times Col-0}$ populations.

Trait	Treatment	RIL population	QTL	stepwiseqtl model ^a	Total R ² (%) ^b	Chr	Position ^c	Closest marker	LOD peak	95% Cjd	R ² (%) ^e	Additive effect ^f	F-test p-value
RGI	2,4-D 20nM	$Q_{Col-0} \times C_{24}$ $R_{C_{24} \times Col-0}$	r24D-III.2	y~ Q1	12.686	3	74.872	III_18	4.920	71 - 77	12.686	-4.136 (0.845)	0.00000
			r24D-III.1	y~ Q1	10.645	3	72	III_18	4.131	17 - 77	10.645	-5.740 (1.287)	0.00001
RGI	IAA 40nM	$Q_{Col-0} \times C_{24}$ $R_{C_{24} \times Col-0}$	rIAA-II.2	y~ Q1	6.749	2	61	II_16	2.959	54 - 73.941	6.749	3.138 (0.840)	0.00024
			rIAA-II.1	y~ Q1 + Q2 + Q1:Q2	16.330	2	51	II_14	4.659	46 - 59	10.172	2.665 (0.710)	0.00003
			rIAA-V.1	y~ Q1 + Q2 + Q1:Q2		5	2	V_02	4.442	0 - 6	9.673	-2.517 (0.707)	0.00004
				y~ Q1 + Q2 + Q1:Q2				II_14 x V_02	1.761		3.709	2.082 (0.731)	0.00490
RGI	NAA 75nM	$Q_{Col-0} \times C_{24}$ $R_{C_{24} \times Col-0}$	rNAA-V.2	y~ Q1	25.941	5	77	V_21	13.629	74 - 84	25.941	3.721 (0.437)	0.00000
			rNAA-II.1	y~ Q1 + Q2 + Q3	23.481	1	53	I_15	4.094	51 - 58	7.939	-2.116 (0.480)	0.00002
			rNAA-II.1	y~ Q1 + Q2 + Q3		2	38	II_11	3.481	32 - 61	6.699	2.001 (0.495)	0.00008
			rNAA-V.1	y~ Q1 + Q2 + Q3		5	62	V_17	5.102	57 - 93.22	10.016	2.406 (0.486)	0.00000
			r24D-I.2	y~ Q1 + Q2 + Q3 + Q4	27.048	1	19	I_06	5.604	15 - 26	9.942	2.773 (0.535)	0.00000
			r24D-III.1	y~ Q1 + Q2 + Q3 + Q4		3	79.056	III_19	3.577	72 - 86.829	6.199	-2.151 (0.526)	0.00006
			r24D-IV.1	y~ Q1 + Q2 + Q3 + Q4		4	75.098	IV_20	4.173	42 - 77	7.281	-2.232 (0.503)	0.00002
			r24D-V.1	y~ Q1 + Q2 + Q3 + Q4		5	29	V_09	3.810	13 - 44	6.620	2.236 (0.529)	0.00004
HGI	2,4-D 375nM	$Q_{Col-0} \times C_{24}$ $R_{C_{24} \times Col-0}$	r24D-I.1	y~ Q1 + Q2 + Q3	20.584	1	12.83	I_04	3.339	11 - 26	6.807	2.185 (0.552)	0.00011
			r24D-II.1	y~ Q1 + Q2 + Q3		2	46	II_13	3.974	42 - 55.666	8.164	-2.563 (0.591)	0.00002
			r24D-V.2	y~ Q1 + Q2 + Q3		5	43.808	V_13	2.750	4 - 62	5.565	1.970 (0.550)	0.00044
			hIAA-III.2	y~ Q1 + Q2	20.056	3	39	III_10	2.538	30 - 52	5.490	3.359 (0.975)	0.00071
HGI	IAA 500nM	$Q_{Col-0} \times C_{24}$ $R_{C_{24} \times Col-0}$	hIAA-IV.1	y~ Q1 + Q2		4	55	IV_14	7.198	53 - 59	16.568	5.932 (0.991)	0.00000
			hIAA-III.1	y~ Q1 + Q2	17.924	3	0	III_01	3.235	0 - 23	9.728	4.198 (1.069)	0.00014
			hIAA-V.1	y~ Q1 + Q2		5	65.461	V_18	2.658	11 - 81	7.912	3.679 (1.039)	0.00056
			hNAA-II.2	y~ Q1 + Q2	25.224	2	42	II_12	8.798	35 - 46	16.699	-4.660 (0.701)	0.00000
			hNAA-V.2	y~ Q1 + Q2		5	56	V_15	5.157	51 - 64	9.378	3.622 (0.727)	0.00000
			hNAA-I.1	y~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q2	34.391	1	46	I_13	4.852	42 - 53	8.280	-0.621 (0.752)	0.00002
			hNAA-II.1	y~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q2		2	35	II_10	11.228	32 - 36	20.771	-4.322 (0.7526)	0.00000
			hNAA-III.1	y~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q2		3	50.736	III_13	3.846	46 - 55	6.481	3.263 (0.772)	0.00004
			hNAA-IV.1	y~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q2		4	50.82	IV_13	3.320	47 - 64	5.559	2.871 (0.733)	0.00013
			hNAA-V.1	y~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q2		5	40	V_11	3.818	35 - 47	6.431	3.195 (0.759)	0.00004
	y~ Q1 + Q2 + Q3 + Q4 + Q5 + Q1:Q2				I_13 x II_10	4.763		8.119	3.886 (0.821)	0.00000			

^a Formula of the multiple QTL model defined by the stepwiseqtl function. Bold characters indicates specific QTL or QTL x QTL interaction which effect were estimated dropping it from the full model.

^b Estimated percent variance explained by full QTL model.

^c Positions in centimorgan (cM).

^d 95% Bayes credible interval range from low to high positions in centimorgan (cM).

^e Estimated percent variance explained by specified QTL.

^f Estimated additive effect, positive sign indicates that the Col-0 allele decreases growth inhibition, negative sign indicates that C24 allele decreases growth inhibition.

Nineteen *a priori* auxin related CGs fall within the borders of the markers V_20 and V_22 (see Supplementary Table ST13). Eight of 19 showed nonsynonymous changes between the two accessions Col-0 and C24. Interestingly, all three *IAA-LEUCINE RESISTANT (ILR)-LIKEs (ILLs)* in that region have at least two nonsynonymous changes. *ILLs* play an important role in IAA metabolism and can hydrolase amino acid IAA conjugates as shown by Bartel and Fink (1995).

For HGI experiments 18 QTLs were detected which is twice the number of QTLs detected for RGI experiments. The higher number of detected QTLs is reflected by more complex full QTL models as given in Table T3.5.

The 95% CI of h24D-I.1, h24D-I.2, h24D-V.1 and h24D-V.2 overlap in both populations (see reddish boxes Figure F3.19 B). The full QTL models including four QTLs for the $Q_{Col-0 \times C24}$ RIL population and three QTLs for the $R_{C24 \times Col-0}$ RIL population explain $\sim 27\%$ and $\sim 20\%$ of the phenotypic variation, respectively. Here, each QTL in the model has single $R^2 < 10\%$. Overlapping QTLs have a positive sign for additive effects. To further elucidate the four QTLs detected in the $Q_{Col-0 \times C24}$ RIL population, the RILs were grouped according to their alleles at the corresponding nearest markers (see Figure F3.20 A + B + C + D) and the HGI for each group was plotted as boxplots. As detected by the multiple QTL mapping, the additive effect for h24D-I.2 and h24D-V.1 is positive, indicating that the Col-0 allele decreases HGI for the marker I_06 and V_09 (black dashed lines). Likewise, the negative additive effects of h24D-III.1 and h24D-IV.1 can be seen for the markers III_19 and IV_20 (red dashed lines). Next, the RILs were grouped according to all 16 possible allelic classes and differences of the groups were determined by a pairwise Wilcoxon test as plotted in Figure F3.20 E + F. Here, clearly the allelic combination of Col-0 at h24D-I.2 and h24D-V.1 and C24 at h24D-III.1 and h24D-IV.1 (black box Figure F3.20 E) shows a decreased HGI, whereas the opposite allelic combination shows the highest HGI (red box Figure F3.20 E). To validate this kind of QTL combination with altering allelic effects for both parental alleles, NILs with suitable introgressions need to be crossed, propagated and further analyzed for the observed phenotypic variation. Due to this time consuming steps a detailed analysis for this trait was not conducted in this thesis.

For HGI analysis after IAA treatment (see greenish boxes Figure F3.19 B) no overlap for the detected QTLs after multiple QTL mapping was observed between the RIL populations. The high 95% CI range from 11 to 88 cM on chromosome five for the $R_{C24 \times Col-0}$ RIL population indicates only weak support for the hIAA-V.1 QTL. One explanation might be, that for the $R_{C24 \times Col-0}$ RIL population only 133 RIL lines were evaluated due to bad germination rates. Hence, the rather high R^2 value for the full QTL model has to be interpreted carefully. This and the fact that for both RIL populations complex QTL model schemes were proposed, further selection of NILs to validate this QTLs was neglected.

For the NAA experiments multiple QTL mapping revealed QTLs on chromosome two with overlapping 95% CI in both RIL populations (see blueish boxes Figure F3.19 B). For the $Q_{Col-0 \times C24}$ RIL population a QTL model consisting of two QTLs (hNAA-II.2 and hNAA-V.2) explains $\sim 25\%$ of the measured phenotypic variance. The QTL hNAA-II.2 alone explains $\sim 16\%$ and has a negative additive effect. In the $R_{C24 \times Col-0}$ RIL population, a good overlap for this QTL is found by the QTL hNAA-II.1, which by itself explains $\sim 20\%$ of the phenotypic variance in the $R_{C24 \times Col-0}$ RIL population after NAA treatment. However, the full multiple QTL model explains about 34%, but is rather complex in nature with five QTLs and one QTL interaction as given in Table T3.5. The high R^2 values for QTL hNAA-II.1 and hNAA-II.2 looked promising with regard to validation of these QTLs. A closer look at *a priori* CGs related to auxin biology which fall between the markers II_10 and II_12 revealed for eleven out of 24 CGs nonsynonymous nucleotide changes regarding to the Col-0 and C24 genotypes (see Supplementary Table ST14). These eleven CGs also contained genes for which a function in hypocotyl growth was shown previously. Chapman et al. (2012) identified that *GH3.3*, which is an early auxin responsive gene, shows increased gene expression in hypocotyls after picloram (another auxin derivate) treatment. For *EARLY FLOWERING 3 (ELF3)* at least an involvement in shade avoidance responses between the *A. thaliana* accessions Bay-0 and Shadara was shown by Jiménez-Gómez et al. (2010).

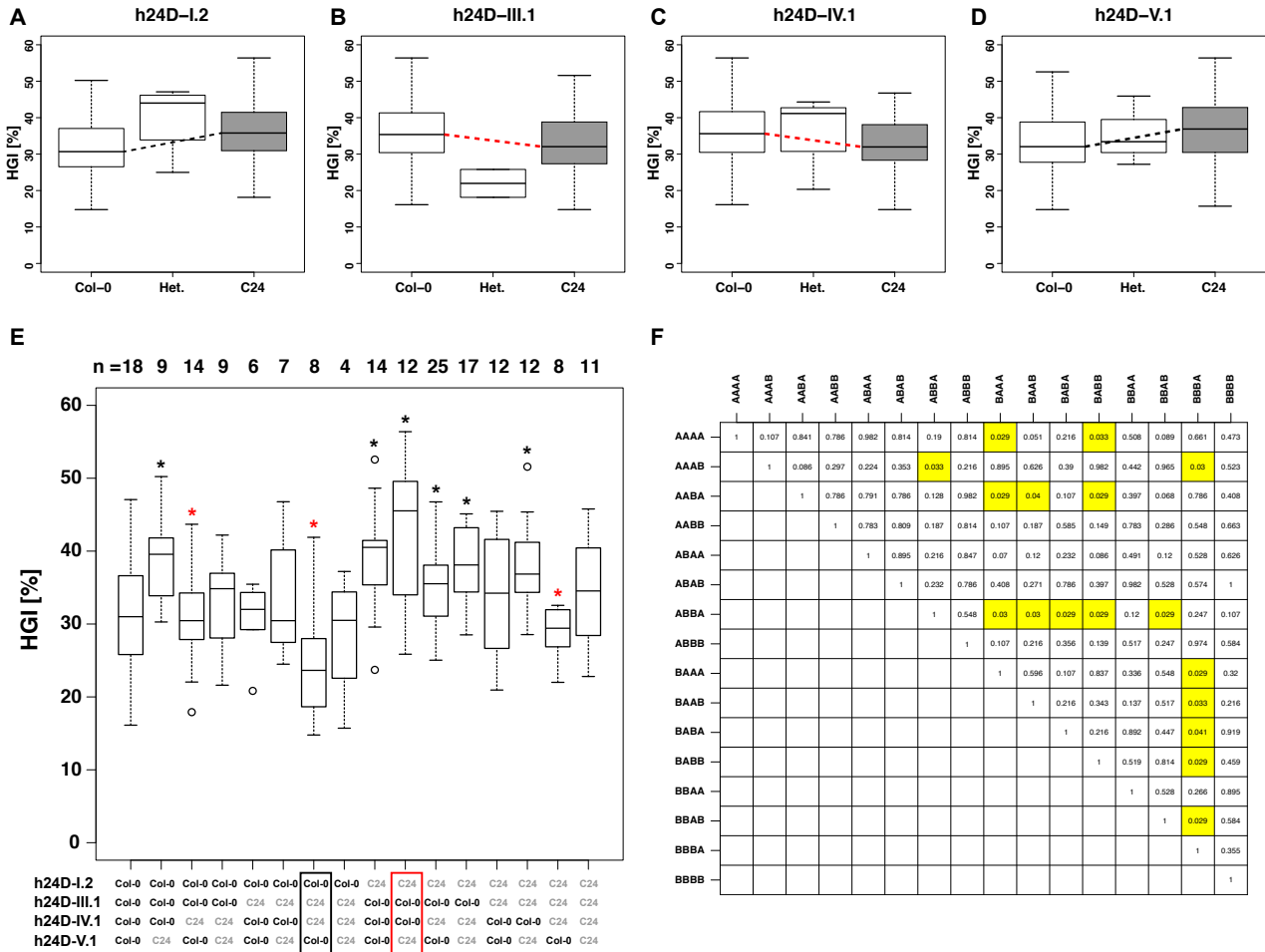


Figure F3.20: Combinatorial QTL effects result either in decreased or increased HGI after 2,4-D treatment. Phenotype by genotype plots for the nearest markers (I.06, III.19, IV.20 and V.09) of the four detected QTLs (h24D-I.2, h24D-III.1, h24D-IV.1 and h24D-V.1) in the $Q_{Col-0 \times C24}$ RIL population are shown as boxplots and grouped by their corresponding genotype, either homozygous for the Col-0, homozygous for the C24 allele or heterozygous (Het.) (A + B + C + D). A black dashed line indicates a positive additive effect, a red dashed line a negative additive effect. (E) Genotype by phenotype plots grouped into 16 possible allelic classes at the 4 detected QTLs. Allelic combinations are indicated below the x-axis. Number of RILs for each group is shown above the boxplot. Black asterisks highlight significant differences between allelic group Col-0 - C24 - C24 - Col-0 and other marked groups. Red asterisks highlight significant differences between allelic group C24 - Col-0 - Col-0 - C24 and other marked groups based on a pairwise Wilcoxon test corrected for multiple testing as shown in F (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$). (F) Names correspond to specific allelic groups at the four indicated QTLs "A" stands for the Col-0 allele and "B" stands for the C24 allele.

However, no association of *ELF3* and shade avoidance response was found when Filiault and Maloof (2012) conducted GWA mapping.

After evaluating all calculated multiple QTL models and their individual allelic effects for each of the auxin response treatments, QTLs identified after NAA treatment promised good chances to be validated by investigating a set of NILs. In both cases for RGI and HGI not only the highest LOD peak values, but also the highest R^2 values for QTLs were observed. Phenotypic analyses of these treatments were further conducted and are explained in the next subsection to validate and to possibly narrow down the genotypic region which is correlated to RGI and HGI.

Allelic effects and their influence on transgression

Based on the allelic effects determined by multiple QTL models it should be possible to predict not only if transgression has to be observed, but also in which direction transgression develops.

For HGI IAA treatment the detected QTLs showed only additive effects with the Col-0 allele decreasing HGI. In this case no transgression would be expected for a trait. The same is true for traits with only one strong main effect QTL like for RGI after 2,4-D treatment. However, looking at the phenotypic distribution data analyzed, only RGI after IAA treatment showed a shift in one direction with a lower RIL mean value, but here compared to both parental lines, which marginally showed the same phenotypic mean values.

The still observed transgression for the RGI 2,4-D trait and the HGI IAA trait might be explained by undetected minor QTLs with opposite effects. One fact which underlies this hypothesis is that the overall R^2 values obtained were only of about 10% for the RGI 2,4-D trait and about 20% for two QTLs for the HGI IAA trait. Another explanation for the detected transgression might be random environmental variation in the RILs for the experiments. They should be low due to the experimental design, but were not addressed in the multiple QTL model (as discussed in another QTL study on light response in *A. thaliana* (Wolyn et al., 2004)).

Colocalization of QTLs

Comparing the detected QTLs and their 95% CI for both RIL populations and both trait categories RGI and HGI facilitates the assessment of colocalization. Such a colocalization can be expected for highly correlated traits and traits which might be influenced by the same genetic pathways. Since only spurious correlations within one RIL population was identified, only an overlap for HGI with higher Pearson correlation coefficients were predicted based on the findings given in Table T3.4.

On chromosome four such an overlap of 95% CIs could be observed for the $Q_{Col-0 \times C24}$ RIL population for all conducted HGI treatments, indicating a genotypic region which might contribute globally to HGI. However, the same genotypic region at the end of chromosome four was not found in the $R_{C24 \times Col-0}$ RIL population, so it can not be rejected that maternal or paternal effects have an influence here.

On chromosome five a similar observation can be made for the $R_{C24 \times Col-0}$ RIL population. Here, the 95% CIs are more scattered, which is also illustrated by the different LOD peak positions (see Figure F3.19). In total, for almost all different auxin treatments influencing HGI one QTL could be detected on chromosome five, resulting in a high overlap of the 95% CIs. However, the LOD peak values were low and also the LOD peak positions are as mentioned scattered over the whole chromosome. Therefore, a reasonable comparison of these loci is quite difficult.

Taking RGI and HGI simultaneously into account, an overlap of the 95% CIs on chromosome two can be seen. Here, a genomic region encompassing the markers II_10, II_11, II_12, II_13 and II_14

is affected. Interestingly, the QTLs on chromosome two for the RGI traits show a positive additive effect, whereas the QTLs for HGI have a negative additive effect (see Table T3.5). The contrasting effects might explain the weak phenotypic correlations between RGI and HGI measurements within each RIL population (see Table T3.4).

3.2.3 Validation of QTLs for auxin related traits by using near isogenic inbred line (NIL) populations

Several QTLs were identified by the QTL mapping approach described above. Each QTL represents a correlation for a genotypic marker or its surrounding genomic region with phenotypic observations for the investigated traits; here, auxin response traits in young *A. thaliana* seedlings. As already outlined in the previous subsection, the genomic regions detected by the QTL mapping approach can be very large including up to thousands or even more genes. To further elucidate QTL regions and to identify the underlying genes which cause the observed phenotypic variation, first the QTL region needs to be reduced to a size suitable for QTL cloning (Alonso-Blanco et al., 2005).

To narrow down the initial QTL region, near isogenic lines (NILs) or heterogeneous inbred families (HIFs) are used. HIFs in contrast to NILs are found within the investigated RIL population itself, carrying residual heterozygosity at genomic regions of interest. NILs contain a single fragment or a small number of genomic introgression fragments from a donor parent into an otherwise homogeneous genetic background (Keurentjes et al., 2007). A first important step in the scheme of QTL cloning is the validation of the identified QTL effects in selected NILs or HIFs. Only if the validation in NILs or HIFs was successful, the next step to further fine-map the QTL region and in the end clone the underlying gene can be taken.

To use HIFs for QTL cloning, in a first step HIFs which segregate in the QTL region need to be selfed. Selfing progenies which are homozygous at the complete QTL region of interest, for either the maternal or paternal allele, can be used to validate the QTL region. HIFs which show recombination events at the segregating QTL region of interest can be further used to narrow down the QTL region with the aim to reduce the number of CGs within this genomic region. Due to probability this includes screening of a very large number of progeny plants.

A population of different NILs with altering positions of parental introgression in a homogenous donor parent genetic background can be used to validate and refine the QTL region of interest. After the genomic region is validated, NILs need to be back-crossed to the donor parent to get heterozygous plants at the QTL region. To further fine-map and narrow down the QTL region, these plants then need to be selfed and screened for recombination events within this region. Again, due to probability this includes screening of a very large number of progeny plants.

As observed in the previous subsection, multiple QTL mapping detected QTL effects of different sign within one RIL population: signatures of epistatic interaction between QTLs and complex QTL models for which QTLs fall on different chromosomes. In case two QTL influence phenotypic variation with the same strength but opposite sign of evaluated QTL effect, a single introgression at one QTL would not be sufficient to distinguish between the phenotypic classes. Here, preferably NILs carrying an introgression either at one or the other QTL position would have to be crossed to get all possible allelic combinations to explain the QTL effects. In contrast to NILs, within the population of HIFs, there might be plants with a shuffled genetic background at these hypothetical QTL positions, which might then be suitable to unravel these complex QTL interactions.

Even if a CG is detected by further fine-mapping and the identified polymorphisms would argue for this CG, a functional characterization is mandatory to explain the observed QTL effects.

However, in this thesis only the validation step was made to elucidate whether further fine-mapping would be possible. For validation, auxin response traits with one strong *main effect* QTL were chosen.

In this case, a NIL population should be sufficient to take the first step on the long way towards cloning the target gene.

Validation of QTLs on chromosome five influencing root growth inhibition (RGI) after NAA treatment

To validate the presence of the overlapping QTL region on chromosome five that was detected for RGI experiments, a subset of NILs introduced by Törjék et al. (2008) carrying either C24 introgressions in Col-0 background or Col-0 introgressions in C24 background were selected and further studied.

The strongest QTL rNAA-V.2 was identified in the $Q_{Col-0 \times C24}$ RIL population after 75 nM NAA treatment (see Figure F3.17 A, Table T3.5, Supplementary Table ST10). The QTL rNAA-V.1 in the $R_{C24 \times Col-0}$ RIL population showed a broader 95% CI (see Figure F3.17 B and Table T3.5) and a slight increase for LOD score values at the end of chromosome five for the CIM method.

To be able to unravel the possibility of a so called "ghost" QTL, two linked QTLs of the same effect sign may lead to a maximum in the LOD profile between the two true QTL locations (Martinez and Curnow, 1992), NILs were selected based on the 95% CI determined for both detected QTLs (rNAA-V.1 QTL and rNAA-V.2) (see Figure F3.21 A).

The selected NILs were quantified and analyzed for significant auxin response differences after 75 nM NAA treatment as given in Figure F3.21 B. Both QTL, rNAA-V.1 and rNAA-V.2, act as *main effect* QTLs in the complete multiple QTL models. In Figure F3.17 B + C the positive additive effects for the nearest markers V_17 and V_21 to QTL rNAA-V.1 and rNAA-V.2 are shown. With that it was supposed that NIL lines with C24 background carrying an introgression at the specified region would result in a decreased RGI.

The opposite effect was expected for NILs in the Col-0 background that carry an introgression of the C24 allele at the specified region. As given in Figure F3.21 B, this expectation could be validated. The NIL N32 with Col-0 background covers with its introgression of the C24 allele the major part of the 95% CI of QTL rNAA-V.2 from 65.461 to 86.424 cM. N32 showed a RGI of $45.71 \pm 4.40\%$ RGI and differs in its growth response significantly from Col-0 ($37.93 \pm 4.12\%$ RGI).

However, the introgression size of NIL N32 is large so that the possibility of a "ghost" QTL or in other words two QTLs, one on the lower and one on the upper part of this introgression can not be rejected.

The QTL region could be further validated by the NIL M02 with C24 background and Col-0 introgression spanning the complete 95% CI showing a percent RGI of 39.27 ± 5.07 , which is significantly different to C24 after correcting for multiple comparisons ($47.39 \pm 5.18\%$ RGI) (see Figure F3.21).

Interestingly, the NIL N10 (Col-0 background), which carries only an introgression at the bottom part of the highlighted 95% CI, differed also significantly from Col-0 (see Figure F3.21 A+B). A closer look at *a priori* CGs related to auxin biology which are located downstream the marker V_24 revealed for four out of 12 CGs nonsynonymous nucleotide changes between the Col-0 and C24 genotypes (see Supplementary Table ST8).

To examine possible NAA concentration effects, the NILs N32 and M02 were investigated in more detail (see Figure F3.22). For the lowest applied NAA concentration of 25 nM, there was no significant difference between N32, M02 and C24 whereas Col-0 showed almost no RGI at this concentration (see Figure F3.22 A). For 75 nM NAA and 125 nM NAA treatment, NIL M02 did not differ significantly from C24 in its growth response, but showed a slight reduction in RGI. Consistent with the previous results, N32 showed a significant difference from Col-0 also at a higher concentration of 125 nM NAA (see Figure F3.22 B + C).

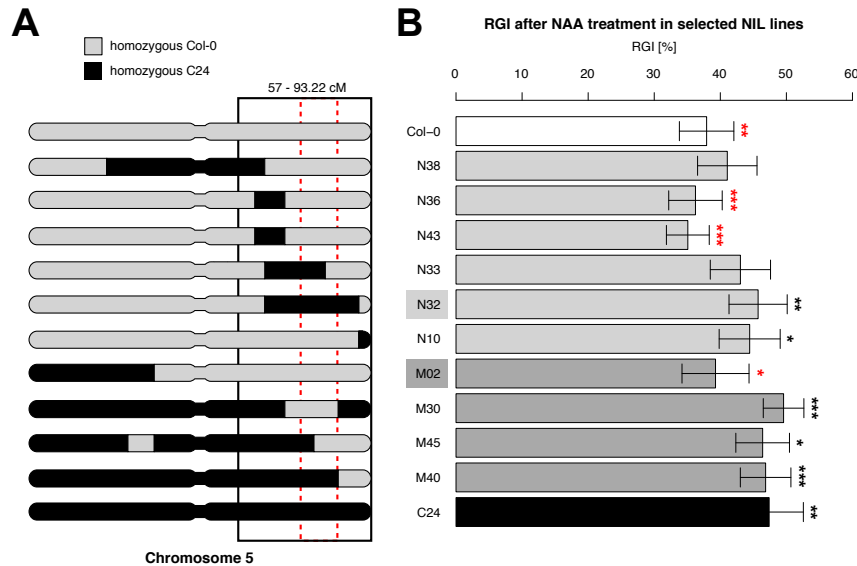


Figure F3.21: Validation of the QTLs corresponding to the NAA response trait on RGI. (A) The allelic composition of the analyzed NILs and the 95% confidence interval based on the multiple QTL mapping results for chromosome five. The black box highlights the 95% CI for the rNAA-V.1 QTL in the $R_{C24 \times Col-0}$ RIL population, the red dashed lines highlight the 95% CI for rNAA-V.2 QTL in the $Q_{Col-0 \times C24}$ RIL population. Allelic regions homozygous for Col-0 are highlighted in grey, whereas C24 is highlighted in black. NILs with the prefix N (lightgrey) have a Col-0 genetic background, NILs with the prefix M (darkgrey) have a C24 genetic background. (B) RGI of the analyzed NILs after 75 nM NAA treatment. NILs N32 and M02 were subsequently analyzed in more detail (see Figure F3.22). Error bars show % standard deviations. Black asterisks highlight significant response differences between Col-0 and the marked NILs based on a two-way ANOVA analysis. Red asterisks denote significant response differences between C24 and the marked NILs (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$).

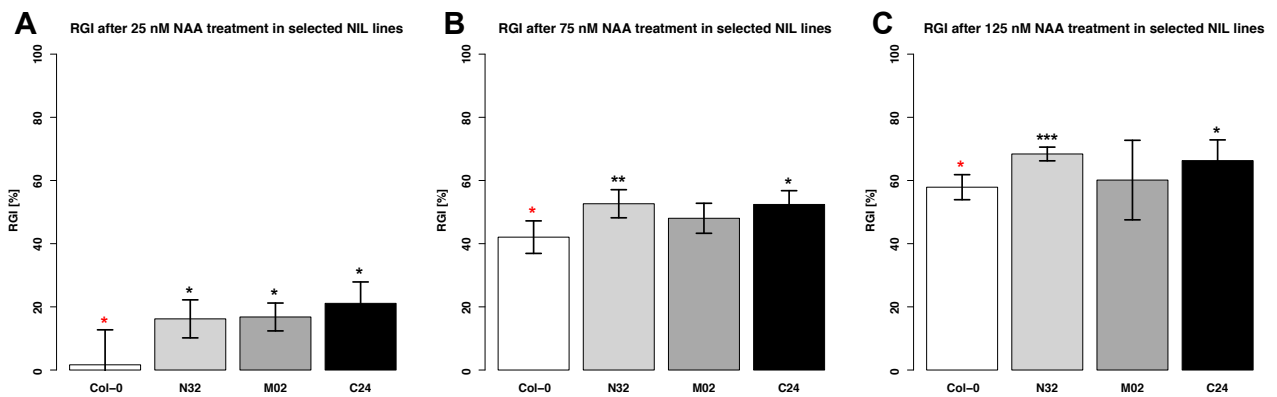


Figure F3.22: Dose response analysis in selected NILs to validate QTL rNAA-V.2 corresponding to the NAA response trait on RGI. RGI in response to the indicated NAA concentrations (25 nM, 75 nM, 125 nM) for the selected NILs (N32 and M02) and the parental lines (Col-0 and C24) are given as barplots showing phenotypic mean values of the 10 longest plants evaluated. Error bars denote % standard deviation. Black asterisks highlight response differences between Col-0 and the marked NILs based on a two-way ANOVA analysis. Red asterisks denote significant response differences between C24 and the marked NILs (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$).

An attempt to validate QTLs on chromosome two influencing hypocotyl growth inhibition (HGI) after NAA treatment

Another promising candidate QTL region for validation encompasses the QTLs hNAA-II.1 and hNAA-II.2. The NILs N59 and M56 were selected to validate hNAA-II.1 (see Supplementary Figure S21 for allelic composition).

Similar to the validation of the QTL for the RGI assay, the NILs N59 and M56 were quantified for HGI after treatment with 3 different NAA concentrations (see Supplementary Figure S21). The expected effect predicted that NILs with Col-0 background carrying a C24 introgression at the QTL hNAA-II.1 region should lead to a decreased HGI and NILs with C24 background carrying a Col-0 introgression, which would lead to an increased HGI.

However, this could not be validated reliably at the given concentrations. For a 250 nM NAA treatment no differences in growth response between NIL N59 and M56 to their respective parental lines with same genetic background could be observed. For a 500 nM NAA treatment, a not significant decrease of HGI could be seen for N59 and a not significant increase of HGI for NIL M56. After treatment with 750 nM NAA, a significant increase of HGI could be measured for NIL M56. However, no reduction of HGI was observed for NIL N59.

One of two QTL regions, which have been identified by multiple QTL mapping in the previous subsection, could be validated by the use of NILs. To further work on cloning the QTL region of rNAA-V.1 and rNAA-V.2, the most promising NIL N32 should be back-crossed to Col-0 plants. After a first round of back-crossing, which would introduce heterozygosity at the given genomic region, the progenies of an additional selfing step could be genotyped and used for further fine-mapping.

3.3 Natural genetic variation of genes involved in auxin biology evaluated by association mapping

Here, genome-wide association (GWA) mapping as another forward genetic approach with regards to the previously described QTL mapping (see section 3.2) was applied on 80 diverse *A. thaliana* accessions.

QTL mapping has some disadvantages, the analysis is specific to the parental lines used for the experimental segregating population (in this thesis *Col* – 0 × *C24*). The chosen QTL mapping population may not be representative for the genetic variation in a broader gene pool of the investigated species (*A. thaliana*) on which natural selection acts (Bergelson and Roux, 2010).

To overcome this limitation of traditional QTL mapping including the generally low resolution with large 95% CI obtained for a detected QTL, GWA mapping has the following advantages over QTL mapping. First, one is not restricted to a small set of parental lines, but can use numerous natural occurring accessions to conduct phenotype genotype correlation analyses. The accessions have accumulated recombination events over the historic lifetime of the species, in this case *A. thaliana*. Between the accessions the mentioned recombination events can be identified by numerous polymorphisms, which have accumulated during this historic lifetime, and can facilitate fine-mapping to a trait of interest (Assmann, 2012). Second, if the population, which is used for GWA mapping, has evolved under natural environments and the origin of all individuals is known, it is more likely to uncover polymorphisms that are causative under field conditions and may reflect natural occurring selection (Assmann, 2012).

3.3.1 80 *Arabidopsis thaliana* accessions screened for phenotypic and genotypic relatedness

Descriptive statistics and geographic correlations for root and hypocotyl growth phenotypes in 80 *A. thaliana* accessions

The impact of population structure on the investigated traits is of major concern, when conducting GWA mapping. At first, important information for GWA analyses can be examined by evaluating possible correlations for the traits of interest and the geographical origin of the used *A. thaliana* accessions. Additionally hints, how strong the impact of population structure might be on the analyzed traits, can be further examined by evaluating the genome-wide relationships together with the above mentioned correlations.

The same growth and treatment conditions as conducted for the QTL assays were applied for the 80 *A. thaliana* accessions. The descriptive statistics are summarized in Table T3.6. All six auxin response traits (2,4-D, IAA, NAA for RGI and HGI) showed a normal distribution for the measured phenotypic values. Hence, no data transformation for the GWA analysis was necessary (see Table T3.6 and Supplementary Figures S23, S24, S25, S26, S27 and S28).

Figure F3.23 shows the auxin dependent root and hypocotyl growth inhibition pattern represented as the relative phenotypic mean values for each evaluated *A. thaliana* accession. In addition, a UPGMA (Unweighted Pair Group Method with Arithmetic Mean) clustering tree highlights the genome-wide relationships among these accessions based on a **K-matrix**. In what follows, the results of the genome-wide relationships will be highlighted.

The **K-matrix** is based on pairwise haplotypes, either calculated on the full set of known SNPs or on a representative subset of SNPs. Here, the subset was defined as SNPs which are not in close LD to each other. However, genome-wide relationship can be also calculated based on random genomic

Table T3.6: Descriptive statistics of auxin response traits among 80 *A. thaliana* accessions.

Name	ABRC stock number	Ecotype ID	Array ID	Latitude	Longitude	Anastasio Table S1 ^a	RGI ^b			HGI ^c		
							2,4-D	IAA	NAA	2,4-D	IAA	NAA
An-1	CS76091	6898	55	51.22	4.40	GREEN	44.38	55.39	51.74	18.22	40.25	52.03
Arby-1	CS28051	6998	472	59.43	16.80	GREEN	37.10	81.47	48.79	26.87	28.33	45.65
Bay-0	CS76094	6899	134	49.00	11.00	GREEN	35.66	78.98	39.18	22.85	51.95	25.41
Bg-2	CS76096	6709	214	47.65	-122.31	RED 4	33.32	83.10	45.52	20.39	61.19	49.89
Bih-2	CS28090	7035	336	48.00	19.00	YELLOW	59.61	70.31	55.77	44.18	28.79	65.75
Bor-1	CS76099	5837	56	49.40	16.23	GREEN	37.95	69.71	47.55	5.42	20.96	47.52
Bor-4	CS76100	6903	135	49.40	16.23	GREEN	20.99	61.22	44.29	32.86	26.64	57.44
Bur-0	CS76105	6905	137	54.10	-6.20	YELLOW	52.08	79.56	NA	37.30	NA	57.89
C24	CS76106	6906	138	40.21	-8.43	GREEN	39.36	63.68	43.00	29.99	25.81	43.74
CAM-16	CS76107	23	1075	48.27	-4.58	GREEN	22.59	74.24	40.04	25.12	43.43	54.85
CAM-61	CS76108	66	208	48.27	-4.58	GREEN	13.69	58.28	31.04	29.07	-7.90	31.27
CIBC-17	CS76111	6907	60	51.41	-0.64	GREEN	38.94	47.97	40.23	24.88	42.08	30.83
CIBC-5	CS28142	6730	377	51.41	-0.64	GREEN	37.25	68.35	50.07	39.71	51.53	85.44
Col-0	CS76113	6909	139	38.30	-92.30	RED 2	40.33	71.18	46.44	21.44	16.95	39.29
Ct-1	CS76114	6910	62	37.30	15.00	RED 4	8.91	68.69	45.11	32.73	25.11	26.86
Cvi-0	CS76116	6911	143	15.11	-23.62	GREEN	25.84	59.51	44.82	30.03	45.60	50.87
Ede-1	CS28217	7110	448	52.03	5.67	GREEN	NA	59.28	49.46	NA	42.52	71.74
Es-0	CS28241	7126	379	60.20	24.57	GREEN	45.62	76.69	47.94	35.61	29.22	57.13
Est-1	CS76127	6916	146	58.30	25.30	YELLOW	19.15	70.22	43.42	3.93	29.17	41.61
Fei-0	CS76129	8215	147	40.50	-8.32	GREEN	43.05	76.05	45.64	10.48	44.11	53.65
Ga-0	CS76133	6919	66	50.30	8.00	GREEN	30.55	53.51	35.00	7.17	44.21	37.10
Gr-5	CS28326	7158	169	47.00	15.50	RED 1	78.11	66.96	50.39	43.01	NA	64.54
Gy-0	CS76139	8214	67	49.00	2.00	GREEN	20.55	68.02	52.12	39.32	43.98	31.81
Hau-0	CS28343	7164	353	55.67	12.57	GREEN	32.93	77.69	31.48	46.17	31.28	42.15
Hh-0	CS28345	7169	354	54.42	9.89	GREEN	34.06	76.72	40.89	34.16	52.64	50.56
Hi-0	CS76140	8304	1138	52.00	5.00	GREEN	39.83	65.72	40.88	30.71	29.53	41.90
HR-5	CS76144	6924	68	51.41	-0.64	YELLOW	36.66	71.34	49.10	22.17	39.73	54.19
Hs-0	CS76145	8310	262	52.24	9.44	GREEN	41.43	68.02	47.89	24.38	25.28	25.93
Ji-3	CS28369	7424	405	49.20	16.62	GREEN	49.78	78.84	47.80	27.27	38.33	56.85
Kas-2	CS76150	8424	69	35.00	77.00	YELLOW	43.13	49.95	46.33	26.94	37.94	32.23
Kn-0	CS28395	7186	450	54.90	23.89	YELLOW	50.22	68.94	31.51	28.01	44.18	40.61
Ler-1	CS76164	6932	150	47.98	10.87	RED 2	-13.46	61.01	-33.62	40.79	36.99	NA
Lip-0	CS76168	8325	270	50.00	19.30	YELLOW	45.53	68.36	36.66	34.31	34.75	34.84
LL-0	CS76172	6933	72	41.59	2.49	GREEN	34.06	50.98	28.79	31.47	51.36	37.57
Lp2-2	CS76176	7520	73	49.38	16.81	GREEN	6.93	58.99	51.09	26.98	41.85	33.84
Lp2-6	CS76177	7521	74	49.38	16.81	GREEN	28.15	60.29	48.54	30.56	19.76	30.92
Lz-0	CS76179	6936	75	46.00	3.30	GREEN	18.37	71.36	52.03	22.76	30.83	18.73
Mrk-0	CS76191	6937	77	49.00	9.30	GREEN	30.95	69.56	37.04	26.54	27.44	45.94
Mt-0	CS76192	6939	78	32.34	22.46	RED 4	31.94	68.84	35.29	15.47	24.18	31.10
Mz-0	CS76193	6940	92	50.30	8.30	GREEN	5.05	59.05	45.09	35.46	35.21	33.13
N4	CS28510	7446	357	61.36	34.15	GREEN	30.68	66.80	53.49	45.30	NA	67.24
Nd-1	CS76197	6942	93	50.00	10.00	YELLOW	NA	64.33	55.13	12.93	29.86	34.34
NFA-10	CS76198	6943	94	51.41	-0.64	GREEN	40.82	NA	52.68	37.70	NA	52.63
NFA-8	CS76199	6944	154	51.41	-0.64	GREEN	53.29	68.57	50.91	32.15	29.48	36.30
Nz1	CS28578	7263	414	-37.79	175.28	YELLOW	35.83	63.68	42.03	33.59	46.75	69.48
Ob-1	CS28580	7277	304	50.20	8.58	GREEN	44.16	80.48	46.08	25.37	47.69	72.49
Ors-2	CS28849	7284	341	44.72	22.40	GREEN	25.80	73.44	43.86	44.97	NA	NA
Oy-0	CS76203	6946	96	60.23	6.13	YELLOW	22.49	60.14	32.09	15.86	32.24	84.02
Per-1	CS76210	8354	282	58.00	56.32	YELLOW	1.18	67.78	47.29	26.65	29.55	47.30
PHW-28	CS28628	7498	338	50.35	-3.58	GREEN	58.10	70.02	43.44	20.89	38.58	68.93
PHW-33	CS28633	7504	427	52.25	4.57	GREEN	57.14	68.06	38.35	39.86	57.20	NA
Pna-17	CS76213	7523	119	42.09	-86.33	YELLOW	65.10	81.33	48.80	27.72	45.58	62.93
Pog-0	CS28650	7306	308	49.27	-123.21	RED 1	35.29	67.74	41.44	43.00	44.28	41.05
Pro-0	CS76214	8213	97	43.25	-6.00	GREEN	35.90	64.86	35.99	31.39	40.71	47.10
Pu2-23	CS76215	6951	98	49.42	16.36	GREEN	56.42	61.04	43.62	35.74	36.82	52.73
Ra-0	CS76216	6958	99	46.00	3.30	GREEN	46.99	69.39	46.01	40.01	43.53	41.20
Ren-1	CS76218	6959	100	48.50	-1.41	GREEN	47.75	66.92	46.10	28.35	35.49	41.04
Rmx-A180	CS76220	7525	101	42.04	-86.51	GREEN	43.78	59.38	39.15	27.17	60.25	51.06
RRS-10	CS22689	7515	155	41.56	-86.43	YELLOW	NA	NA	32.88	27.20	41.31	40.92
RRS-7	CS28713	7514	156	41.56	-86.43	GREEN	16.58	70.52	48.13	10.42	54.85	62.54
Rsch-4	CS76222	8374	285	56.30	34.00	RED 4	8.82	51.90	30.63	25.76	31.63	72.38
Se-0	CS76226	6961	102	38.33	-3.53	GREEN	41.20	76.87	28.21	28.57	45.34	74.54
Sei-0	CS28729	7333	233	46.54	11.56	RED 1	17.56	78.37	53.25	47.19	77.22	69.40
Shahdara	CS76227	6962	157	38.35	68.48	GREEN	39.66	66.35	45.21	7.51	30.21	43.27
Sq-8	CS76230	6967	103	51.41	-0.64	GREEN	-3.29	60.18	42.09	30.52	34.74	68.69
St-0	CS76231	8387	289	59.00	18.00	RED 1	21.92	70.59	47.49	22.98	53.42	66.20
Ts-1	CS76268	6970	161	41.72	2.93	GREEN	28.48	66.12	40.80	27.12	37.47	45.91
Tscha-1	CS28779	7372	173	47.07	9.90	GREEN	47.33	53.46	38.48	20.31	25.79	44.05
Tsu-0	CS28780	7373	480	34.43	136.31	NA	66.34	73.33	51.54	22.89	12.07	24.47
Uk-1	CS28787	7378	206	48.03	7.77	GREEN	24.63	63.18	31.71	12.98	64.55	NA
UII2-3	CS76293	6973	104	56.06	13.97	GREEN	-0.55	67.21	42.07	14.94	36.72	25.28
Uod-7	CS76296	6976	106	48.30	14.45	RED 4	48.23	69.95	56.52	26.64	29.12	53.89
Van-0	CS76297	6977	162	49.30	-123.00	GREEN	42.39	64.11	34.95	28.27	39.18	27.62
Wa-1	CS28804	7394	174	52.30	21.00	RED 4	NA	71.20	66.26	41.81	NA	NA
Wag-3	CS28808	7390	340	51.97	5.67	GREEN	26.26	67.38	47.70	37.14	57.85	76.02
WAR	CS28812	7477	420	41.73	-71.28	GREEN	42.40	77.50	41.73	33.27	51.33	76.42
Wei-0	CS76301	6979	108	47.25	8.26	GREEN	15.06	51.42	45.52	28.48	36.81	62.77
Ws	CS28823	7397	175	52.30	30.00	RED 2	0.35	65.71	41.56	34.19	NA	47.02
Wt-5	CS76304	6982	110	52.30	9.30	GREEN	44.40	64.16	37.50	8.19	16.99	31.81
Zdr-6	CS76306	6985	112	49.39	16.25	GREEN	5.77	77.74	39.03	25.07	31.40	58.67

^a Supplemental Table S1 results from Anastasio et al., 2011.

^b Root growth inhibition. Phenotypic values are indicated in % inhibition.

^c Hypocotyl growth inhibition. Phenotypic values are indicated in % inhibition.

^d Shapiro-Wilk test.

	Mean ± SD	33.35 ± 17.56	67.30 ± 8.07	43.69 ± 7.33	28.04 ± 10.14	37.82 ± 13.19	48.99 ± 15.80
Range (Min)		-13.46	47.97	28.21	3.93	-7.90	18.73
Range (Max)		78.11	83.10	66.26	47.19	77.22	85.44
SW test ^d		0.44551	0.14826	0.33008	0.10018	0.18481	0.14322
Skewness		-0.312	-0.276	-0.005	-0.355	-0.092	0.304
Kurtosis		0.160	-0.230	0.151	-0.136	1.724	-0.696

fragments as conducted by Nordborg et al. (2005). Since the **K-matrix** should represent the genome-wide relationships between accessions, no big changes are expected when different global subsets of SNPs are inspected. Consistent with the previously published data of Atwell et al. (2010), Col-0 was defined by the UPGMA clustering as an outlier (see Figure F3.23). This, contradicts the findings of Nordborg et al. (2005), who placed Col-0 as a Central European accession based on another population structure analysis. However, it was shown by Atwell et al. (2010), that the **K-matrix** with Col-0 as an outlier is capable to correct for population structure when conducting a GWA analysis. Here, I decided to also use the **K-matrix** for further GWA mapping. Consistent with previous studies, other accessions clustered according to their origin or were placed as outliers like Cvi-0 and Bur-0 (see Aranzana et al. (2005), Atwell et al. (2010) and DeRose-Wilson and Gaut (2011)).

One fact that strengthens the usability of the **K-matrix** to correct for relatedness in the investigated population of *A. thaliana* accessions is illustrated in Figure F3.24. Here, the calculated pairwise relatedness based on genomic data declines looking at correlation to pairwise geographical distances of the *A. thaliana* accessions. This suggests that individuals from distant locations are less related to each other as already shown by Sharbel et al. (2000). For some of the used *A. thaliana* accessions the annotated geographical origin is not clear, which was previously examined by Anastasio et al. (2011). Hence, to account for falsely annotated geographical origins, only accessions that fall into the "GREEN" class of Supplementary Table S1^a of Anastasio et al. (2011) were retained in the geographical distance analysis (see Table T3.6).

Aranzana et al. (2005) identified not randomly distribution for flowering time and DeRose-Wilson and Gaut (2011) identified such not random phenotype patterns for salinity tolerance. In addition to the genome-wide relatedness, the scaled phenotypic values are also shown in Figure F3.23. In contrast to previous studies, which showed similar representations of phenotype data and UPGMA clustering in one plot, here no correlation of phenotype patterns and UPGMA clustering is observed. Here, the phenotype distribution for the auxin response traits seems to be of random nature. In cases where phenotypic traits, for example flowering time, are strongly correlated with the geographic origin, the standard null hypothesis in association mapping, independence between marker genotypes and investigated trait, is false (Aranzana et al., 2005). A consequence is that one might find elevated false positive associations due to population structure.

However, in the case of auxin response traits such a correlation of phenotype pattern and genome-wide relatedness was not seen in Figure F3.23. This, in first instance, indicates no strong influence of population structure on the investigated traits. To take a closer look at this, further it was examined, if the correlation between the auxin response traits and the origin of the evaluated accessions exists. The pairwise phenotype distances were plotted against the pairwise geographical distances and correlation coefficients were calculated (see Figure F3.25).

For RGI after 2,4-D and IAA treatments no significant correlations of pairwise phenotype distances and pairwise geographical distances were identified (see Figure F3.25 A + B). This suggests that for these traits the population structure should have no strong influence on the phenotype genotype association analyses. For RGI after NAA treatment, surprisingly, a weak, but significant negative correlation was identified (see Figure F3.25 C). As outlined by Zuther et al. (2012), this indicates an influence of genetic population structure on this trait.

A similar weak negative correlation with the same consequences on GWA mapping was also identified for HGI after 2,4-D treatment (see Figure F3.25 D). For HGI after IAA and NAA treatment a significant positive correlation was detected (see Figure F3.25 E+F). Here, again, population structure might have an influence on phenotype-genotype correlation analyses, which need to be accounted for by the use of suitable GWA methods that can handle such situations. However, in total the Pearson correlation coefficients measured were very low indicating only a spurious correlation of the auxin response traits with the origin of the analyzed *A. thaliana* accessions. With that, population structure might only have minor - if any - influence on GWA mapping. On the contrary, population structure correction schemes might overcorrect the applied GWA mapping models.

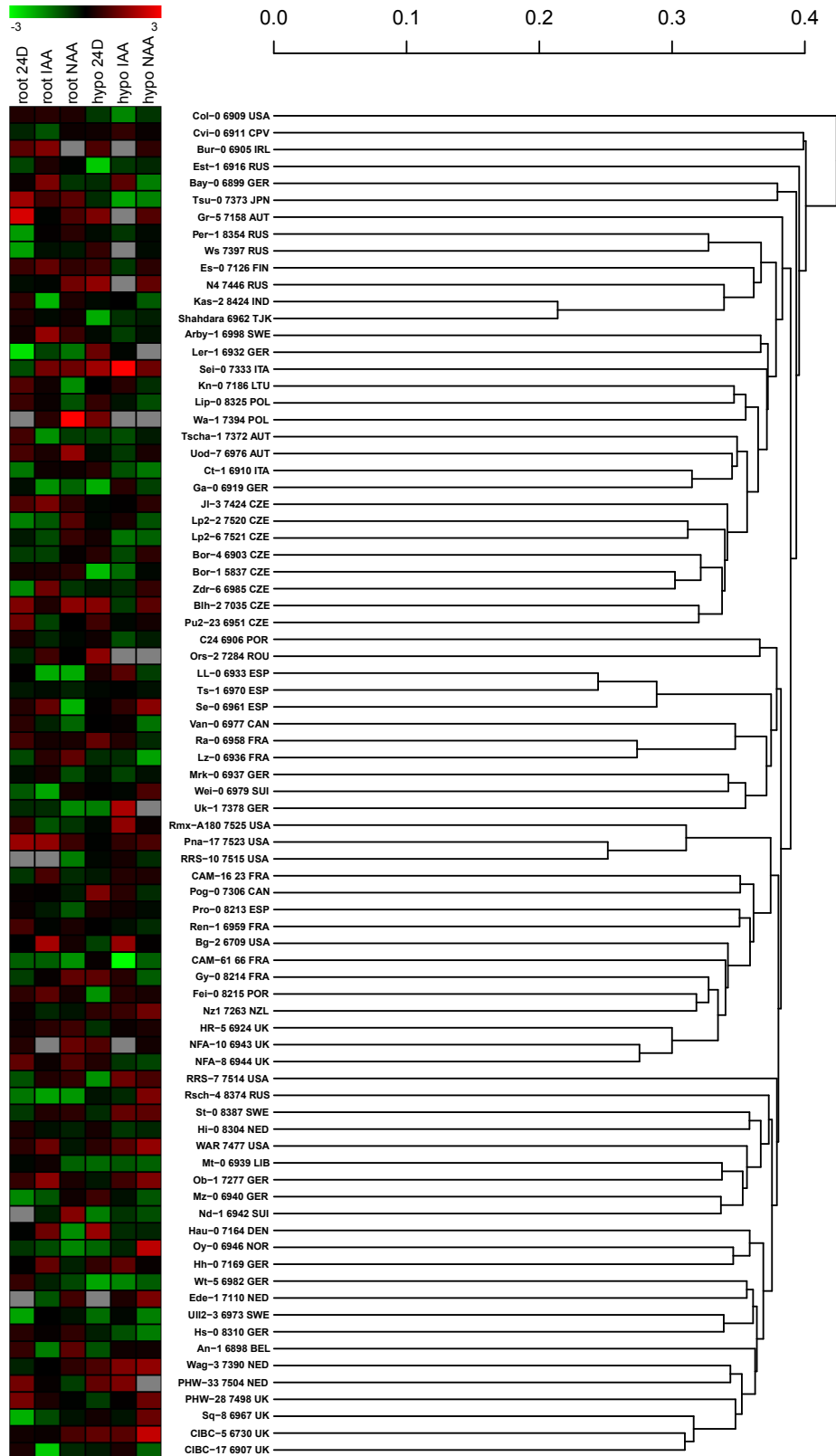


Figure F3.23: Auxin dependent root and hypocotyl growth inhibition pattern with pairwise genotype distance. The tree on the right shows a UPGMA clustering on 80 selected *A. thaliana* accessions based on identity by state kinship distance calculated on the pruned **1307 MAF01 SNP data** set to correct for SNPs in high linkage disequilibrium. On the left side, the investigated auxin treatments and the relative phenotypic mean values for each accession are indicated with colors (green indicates low, red high growth inhibition; grey indicates missing data).

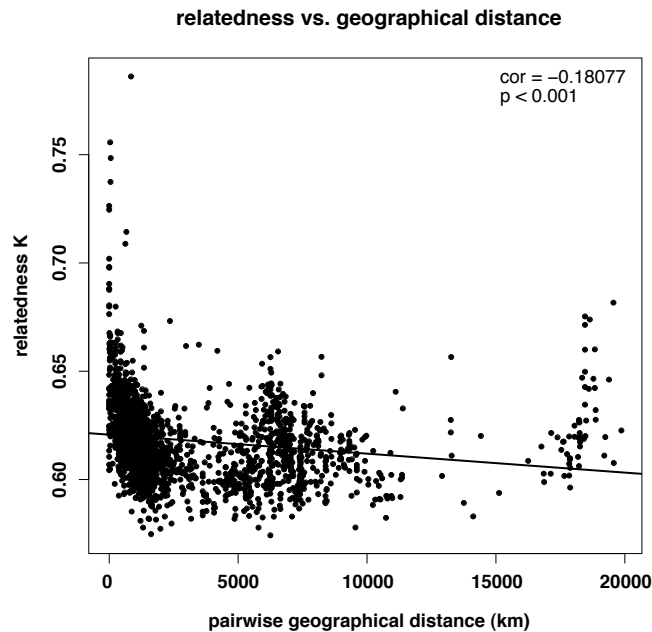


Figure F3.24: Correlation analysis between pairwise genome-wide relationships and pairwise geographic distance for a subset of 67 *A. thaliana* accessions. In addition to the pearson correlation coefficient (cor) and the calculated p-value (p), the black line represents a regression line fitted with the `lm` function in R to the data.

In addition to the correlation of geographical distance, the overall correlations between individual treatments were inspected (see Table T3.7). As already observed for the RIL populations (see section 3.2), the pairwise pearson correlation coefficients were higher for the HGI treatments than for the RGI treatments (see Table T3.4 for comparison). Consistent with the above mentioned results of pairwise phenotypic distances and pairwise geographical distances, the HGI experiments after IAA and NAA treatment showed a significant correlation. Since only one significant correlation between the individual treatments and only low pearson correlation coefficients were identified, no overlaps for the GWA mapping results were expected.

3.3.2 Unravelling the correlation of phenotype and genotype for auxin response traits by the use of genome-wide association (GWA) mapping

Choosing appropriate models for genome-wide association testing

To identify appropriate models, which can be used for the analyses of the investigated auxin response traits in 80 *A. thaliana* accessions and which in case sufficiently correct for population structure, different GWA methods were applied and quantile-quantile plots were inspected for model choosing. As already outlined, the MLM method outperforms other methods by taking population structure into account (Yu et al., 2006).

The analysis of pairwise phenotypic distance and pairwise geographic distance revealed only small effects of population structure on the auxin response traits. For GWA mapping different models can be applied to identify phenotype-genotype correlation. The GLM GWA mapping methods is based on a parametric model, it corrects for possible population structure with a PCA based **P-matrix** (detailed description in the materials and methods section). The MLM GWA mapping is based on a parametric model, in addition to a **P-matrix**, the use of an above mentioned **K-matrix** to correct for

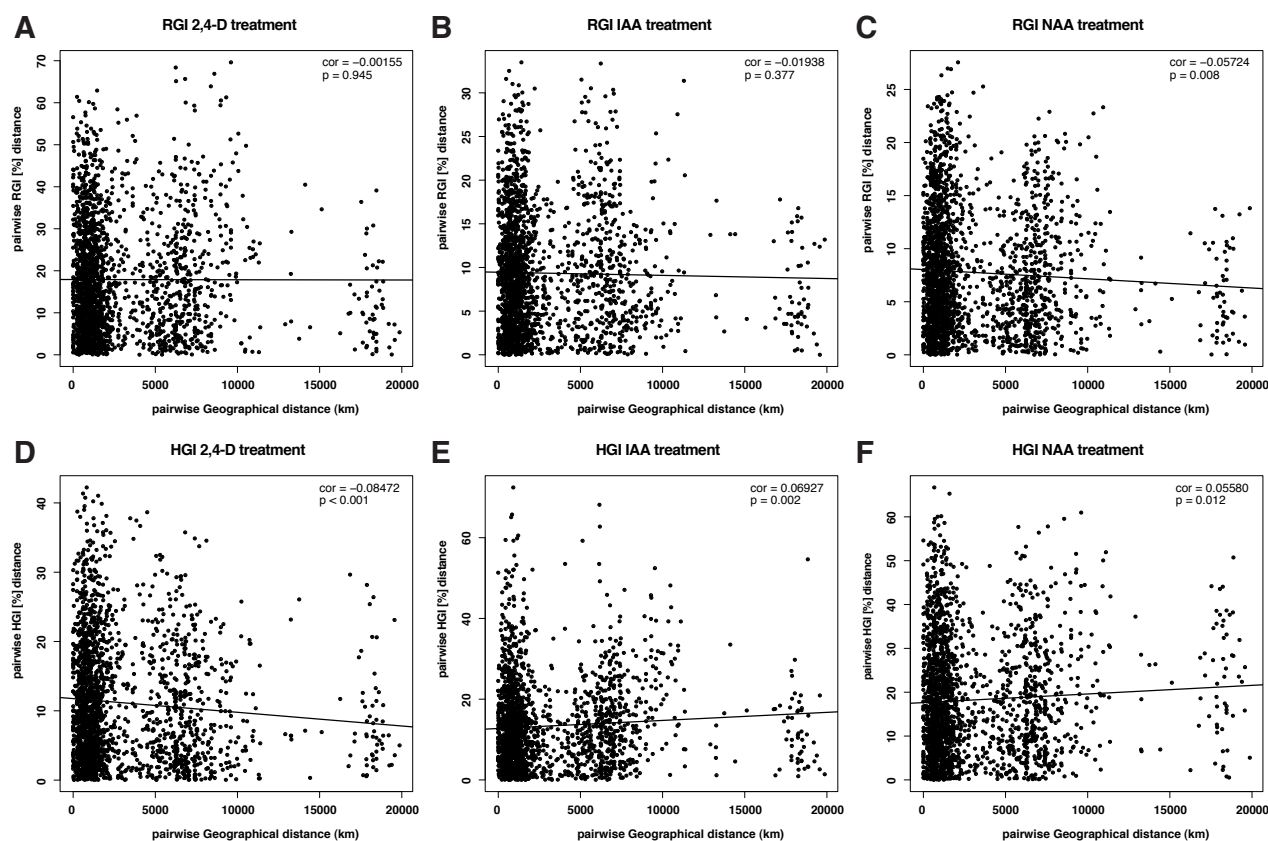


Figure F3.25: Correlation analysis between pairwise phenotypic distances and pairwise geographic distance for a subset of 67 *A. thaliana* accessions. Correlations were calculated for RGI experiments after 20 nM 2,4-D (A), 40 nM IAA (B) and 75 nM NAA (C) treatments and for HGI experiments after 375 nM 2,4-D (D), 500 nM IAA (E) and 500 nM NAA (F) treatments. In addition to the Pearson correlation coefficient (cor) and the calculated p-value (p), the black line represents a regression line fitted with the `lm` function in R to the data.

population structure is mandatory. In contrast to these mapping methods, the WILCOXON GWA mapping is based on a non-parametric test, here, the population structure is not taken into account. These methods were compared to choose appropriate model settings.

As an example, Figure F3.26 illustrates the GLM, MLM and WILCOXON GWA mapping methods for the RGI 2,4-D trait. For the MLM method different options like the P3D option were included, which determines the population parameters only once and with that speeds up the whole fitting process. The `compression` level reduces and clusters the **K-matrix**, which should also improve model fitting.

The MLM method with P3D option showed an underfitting of the model independent from the `compression` level. Hence, no association was detected (blue and red dashed lines in Figure F3.26), which is why the P3D option was rejected and not used for further GWA mapping.

The MLM method with `optimal` `compression` and `reestimate` option showed an overfitting, which would result in the detection of elevated false positive associations (solid red line in Figure F3.26). As a consequence, the `optimal` `compression` option was likewise rejected and not used for further GWA mapping.

The MLM method with `no` `compression` and `reestimate` option showed good fitting results comparable to the GLM method based on the QQ-plot (solid blue line in Figure F3.26). Since there was no indication of strong influence of population structure according to the results of the previous section,

Table T3.7: Phenotypic correlations among auxin response traits among 80 *A. thaliana* accessions.

GWA ₈₀ accessions		root			hypocotyl		
		2,4-D	IAA	NAA	2,4-D	IAA	NAA
root	2,4-D						
	IAA	0.228					
	NAA	0.191	0.231				
hypocotyl	2,4-D	0.059	0.112	0.161			
	IAA	0.005	0.163	-0.009	0.199		
	NAA	0.060	0.166	0.082	0.229	0.421**	

Pairwise Pearson correlation coefficient of root and hypocotyl traits between 80 *A. thaliana* accessions.

***, **, * significant after Benjamini-Hochberg correction at $P < 0.001$, $P < 0.01$, $P < 0.05$, respectively.

the comparable results of MLM and GLM were expected.

The `no compression` and `reestimate` option used with MLM revealed to be an appropriate model for all analyzed auxin response traits, and it was chosen to be the primary GWA mapping method (data not shown). To further correct for population structure, different numbers of first principle components (PCAs) of the **P-matrix** were included in the MLM GWA method. Here, based on QQ-plots (data not shown), the best model fitting was observed for the first 3 PCAs. Consistent with the inspected QQ-plots, the 80 accessions that have been selected out of 1307 *A. thaliana* accessions (orange dots) can be easily divided into at most three different subgroups (see Supplementary Figure S22). As already outlined in the previous section, no strong effects based on population structure would be expected in this set of *A. thaliana* accessions. Applying more PCAs would lead to spurious associations. To keep the advantage of correcting for population structure, but also avoid overcorrecting GWA mapping models, it was decided to use the first 3 PCAs in all further conducted GWA analyses for GLM and MLM GWA methods.

In addition to the GLM and MLM mapping methods, a non-parametric test, which does not correct for population structure, was applied on the data. As outlined in Filiault and Maloof (2012), a non-parametric method, like the Wilcoxon rank-sum test (WILCOXON), presents no risk of p-value over correction, when applied to traits that are correlated with population structure.

Keeping the advantages and disadvantages of both strategies in mind, MLM and WILCOXON GWA mapping methods were successfully combined in other GWAS in *A. thaliana* (Atwell et al. (2010), Brachi et al. (2010) and Filiault and Maloof (2012)).

Recently, Shen et al. (2012) published a GWA method which is capable to identify phenotype-genotype associations based on phenotypic variance differences (*v*GWAS). In contrast to that, GLM, MLM and WILCOXON can detect phenotypic mean differences among the evaluated *A. thaliana* accessions. With this new method, Shen et al. (2012) showed that new associations based on the so far largest existing GWA dataset for the species *A. thaliana* published by Atwell et al. (2010) can be identified. Since completely different associations are detected by these two approaches, the auxin response traits were additionally also screened for phenotypic variance differences.

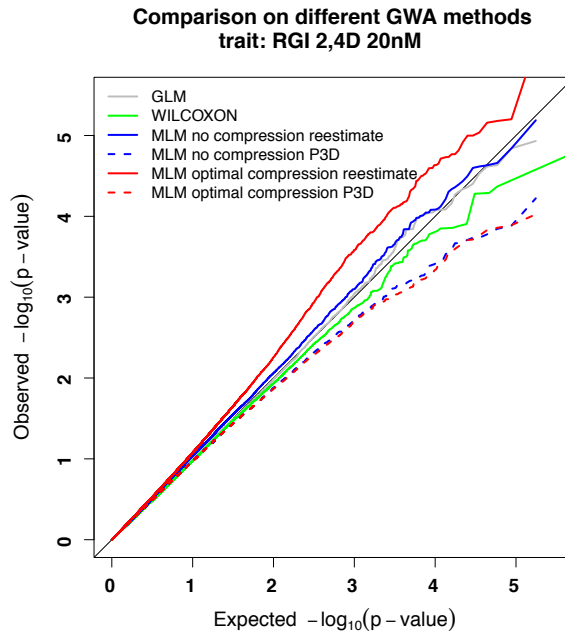


Figure F3.26: Exemplary quantile-quantile plot (QQ-plot) to investigate appropriate GWA methods for RGI after 20 nM 2,4-D treatment. Comparison of different GWA methods by a QQ-plot. Observed $-\log_{10}$ p-values are plotted against expected $-\log_{10}$ p-values to illustrate the fit of the model applied. Mixed linear models were calculated with additional population structure information based only on the **K-matrix** (see materials and methods section for a detailed description).

Conservative analysis of GWA mapping results

For each of the traits and all conducted GWA mapping methods, a so called genome-wide manhattan plot, representing phenotypic genotypic correlations, was calculated. For GLM, MLM and WILCOXON, the manhattan plots represent at a certain SNP position in the *A. thaliana* genome the likelihood, that two groups of accessions ordered by their allele composition differ from each other by their phenotypic mean values. In contrast to that, the manhattan plots for *v*GWAS represent at a certain SNP position the likelihood, that two groups differ by their phenotypic variance. In addition, QQ-plots were generated to highlight model fitting. All global GWA results are given in Supplemental Figures S9 to S14.

Using a conservative significance threshold (nominal 5% with Bonferroni correction), only one SNP appeared to be significant for the RGI experiment after 2,4-D treatment and *v*GWAS mapping method (see Figure F3.27). However, many peaks of moderate significance were observed rather than single distinct peaks like observed for traits effected by a single gene (Filiault and Maloof, 2012). These findings again show the complexity of the auxin response traits and emphasize its polygenic nature.

In addition to a conservative analysis of GWA mapping results, which only concentrates on significant identified SNPs, another approach also takes non-significant SNPs into account. Here, *a priori* CGs related to the trait of interest are investigated. The hypothesis is, that for a trait of interest, for which *a priori* CGs related to this trait are known, such *a priori* CGs show an enrichment in top-ranked associated SNPs.

Since only one SNP appeared to be significant under the conservative point of view, *a priori* CG enrichment in top-ranked SNP categories was evaluated (as also conducted by Atwell et al. (2010), Brachi et al. (2010), Filiault and Maloof (2012) and Shen et al. (2012)). The enrichment of *a priori* CGs, calculated as odd ratios, revealed for the top 100 ranked SNPs in almost all auxin response traits a value above one (see Supplementary Table ST16). For HGI 2,4-D with MLM and WILCOXON, the

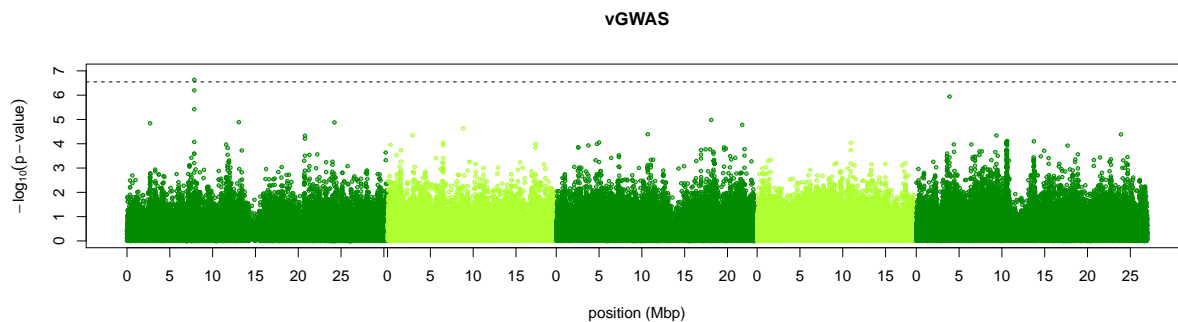


Figure F3.27: Manhattan plot of the *v*GWAS result for RGI 2,4-D. Genome-wide distribution of $-\log_{10}$ p-values of SNP/phenotype association for the variance-heterogeneity GWA method is presented as a manhattan plot. SNP positions on the five *A. thaliana* chromosomes are indicated in Mbp. Their colors are altered to better visualize the chromosomal borders. Dashed lines represent the nominal 5% significance threshold with Bonferroni correction for 175655 tests.

top 50 ranked SNPs showed an enrichment above one, whereas for HGI NAA only the WILCOXON GWA mapping results showed an enrichment for the 50 top ranked SNPs. Here, significantly under-representation of CGs was identified for MLM and WILCOXON GWA mapping methods. For RGI 2,4-D, significantly overrepresentation was identified for MLM and WILCOXON GWA mapping methods. However, to take both mapping methods into account, the top 50 ranked SNPs in both methods were further evaluated. A closer look at these *a priori* CGs was conducted to unravel possible effects on the auxin response traits and are described in the following subsection.

3.3.3 A closer look at *a priori* candidate genes (CGs)

Phenotypic mean differences evaluated by mixed linear model (MLM) and WILCOXON GWA mapping methods

Phenotypic mean differences can be described as follows. One allelic group, which displays a certain SNP position for example an adenine, shows a significantly higher RGI or HGI than the other allelic group, which carries at the same SNP position a cytosine. The MLM and WILCOXON GWA mapping methods firstly group the phenotypic values according to their genotypes and then test for phenotypic mean differences. Here, the MLM and WILCOXON GWA mapping methods were analyzed for *a priori* CGs.

As described and reported elsewhere (Atwell et al. (2010), Brachi et al. (2010), Filiault and Maloof (2012) and Shen et al. (2012)), a CG is defined by a certain rule as follows. A CG must have at least one top-ranked SNP among the highest 50 ranked SNPs within a 10 kb window around their annotated gene position. As outlined by Clark et al. (2007) a 10 kb window is conservative given that linkage disequilibrium decays per 10 kb on average (Brachi et al., 2010).

By applying this rule, 36 CGs associated with auxin were detected and are listed in Table T3.8. In addition to the number of SNPs, the CGs are ordered in Table T3.8, Here, CGs for which the associated SNPs fall directly within the cDNA region of a CG, are bold written and appear on top of the table.

For seven CGs (*NDL1*, *JAZ1*, *SAR1*, *LDL3*, *ABCB22*, *PIN8* and *MYB12*) this holds true, and the underlying SNPs were investigated by checking if they fall into intron or exon regions of the CGs. If the latter was the case, SNPs were checked for synonymous or nonsynonymous status to evaluate the possible influence on the protein sequence level (see Supplementary Table ST17). Five out of 13

SNPs, which fall into exon regions cause an amino acid change according to TAIR10. The indicated differences between the *A. thaliana* accessions in these genes might have direct influence on protein structure and function, and might lead to the observed phenotypic variation.

Interestingly, *PIN8* is one of the genes, which carry a nonsynonymous SNP. *PIN8* belongs to the intensive studied auxin transporter family and within the PINs to the subgroup of short type *PINs*, which localize to the endoplasmic reticulum (Ding et al., 2012). In addition to *PIN8*, another hypothetical auxin transporter (Carraro et al., 2012), was detected with *ABCB22*, which also carries a nonsynonymous SNP. Potentially, both might have an influence on different auxin distribution patterns between the analyzed *A. thaliana* accessions. Here, protein structure prediction analyses might give a hint, whether the specified amino acid changes could have an influence on protein function and integrity. However, this analysis has not been done in this thesis.

Table T3.8: List of *a priori* candidate genes associated with auxin found among the highest correlated SNPs with MLM and WILCOXON GWA mapping methods.

Locus	Gene name ^a	Overlapping QTLs ^b	RGI 2,4-D		RGI IAA		RGI NAA		HGI 2,4-D		HGI IAA		HGI NAA		SNP RANK ^e
			MLM ^c	W ^d	MLM ^c	W ^d	MLM ^c	W ^d	MLM ^c	W ^d	MLM ^c	W ^d	MLM ^c	W ^d	
AT5G56750	NDL1	rNAA-V.2,rNAA-V.1,hIAA-V.1	-	-	-	-	-	-	-	-	-	-	1	1	7,24
AT1G19180	JAZ1	h24D-I.2,h24D-I.1	-	-	-	-	-	-	4	-	-	-	-	-	9,10,14,15
AT1G33410	SAR1		5	-	-	-	-	-	-	-	-	-	-	-	9,33,34,35,36
AT4G16310	LDL3	h24D-IV.1,hNAA-IV.1	-	-	-	-	-	2	-	-	-	-	-	-	14,15
AT3G28415	ABCB22	r24D-III.1,hIAA-III.2	-	2	-	-	-	-	-	-	-	-	-	-	22,38
AT5G15100	PIN8	h24D-V.1,h24D-V.2,hIAA-V.1	-	-	-	-	-	-	-	-	1	-	-	-	22
AT2G47460	MYB12	rIAA-II.2	1	-	-	-	-	-	-	-	-	-	-	-	46
AT1G04680	AT1G04680		-	-	-	-	-	-	-	-	1	-	-	-	3
AT1G22220	AUF2	h24D-I.2,h24D-I.1	-	-	-	-	1	1	-	-	-	-	-	-	4,6
AT5G47370	HAT2	rNAA-V.1,hIAA-V.1	1	1	-	-	-	-	-	-	-	-	-	-	45,10
AT1G19220	ARF19	h24D-I.2,h24D-I.1	-	-	-	-	-	-	1	-	-	-	-	-	10
AT5G20730	ARF7	h24D-V.1,h24D-V.2,hIAA-V.1	-	-	-	-	-	-	-	-	2	-	-	-	12,13
AT3G26810	AFB2	r24D-III.1,hIAA-III.2	-	-	-	-	-	1	-	-	-	-	-	-	12
AT4G18010	5PTASE2	h24D-IV.1,hIAA-IV.1,hNAA-IV.1	-	-	-	-	1	1	-	-	-	-	-	-	14, 29
AT1G34170	ARF13		1	2	-	-	-	-	-	-	-	-	-	-	25,16,48
AT1G63720	AT1G63720		-	2	-	-	-	-	-	-	-	-	-	-	21,25
AT1G71090	AT1G71090		-	-	-	-	-	-	-	-	-	-	1	-	22
AT5G20990	CNX	h24D-V.1,h24D-V.2,hIAA-V.1	1	-	-	-	-	-	-	-	-	-	-	-	23
AT3G61830	ARF18	h24D-III.1	-	-	-	-	-	-	1	-	-	-	-	-	24
AT5G27520	ATPNC2	h24D-V.1,h24D-V.2,hIAA-V.1,hNAA-V.1	-	1	-	-	-	-	-	-	-	-	-	-	27
AT5G56650	ILL1	rNAA-V.2,rNAA-V.1,hIAA-V.1	-	-	-	-	-	-	-	-	-	-	1	-	27
AT5G56660	ILL2	rNAA-V.2,rNAA-V.1,hIAA-V.1	-	-	-	-	-	-	-	-	-	-	1	-	27
AT2G02560	CAND1		-	-	-	-	1	-	-	-	-	-	-	-	29
AT1G25490	EER1		-	-	1	-	-	-	-	-	-	-	-	-	29
AT5G13220	JAS1	h24D-V.2	-	1	-	-	-	-	-	-	-	-	-	-	31
AT5G54500	FQR1	rNAA-V.2,rNAA-V.1,hIAA-V.1	-	-	-	-	1	-	-	-	-	-	-	-	35
AT5G54490	PBP1	rNAA-V.2,rNAA-V.1,hIAA-V.1	-	-	-	-	1	-	-	-	-	-	-	-	35
AT4G36760	APP1	h24D-IV.1	-	-	2	-	-	-	-	-	-	-	-	-	37,44
AT2G01420	PIN4		-	-	-	1	-	-	-	-	-	-	-	-	38
AT4G38630	MBP1		-	-	-	-	1	-	-	-	-	-	-	-	43
AT4G11280	ACS6	h24D-IV.1	-	-	-	-	1	-	-	-	-	-	-	-	44
AT1G35540	ARF14	hNAA-I.1	-	-	-	1	-	-	-	-	-	-	-	-	44
AT4G32880	ATHB8	h24D-IV.1	-	1	-	-	-	-	-	-	-	-	-	-	46
AT4G00220	JLO		-	-	-	-	-	-	-	-	-	-	1	-	48
AT4G01370	MPK4		-	-	-	-	-	-	-	-	2	-	-	-	49,50
AT1G49010	AT1G49010	rNAA-I.1	-	-	-	-	-	-	1	-	-	-	-	-	50

^a Bold written gene names indicate that the found SNP lies within defined gene start and end position.

^b Overlapping with 95% CI for detected QTLs, abbreviation corresponds to Table 4.3.

^c Indicates the number of times that the gene was detected among the top 50 ranked associations with MLM mapping method in a 10 kb window.

^d Indicates the number of times that the gene was detected among the top 50 ranked associations with WILCOXON mapping method in a 10 kb window.

^e Indicates the rank of the SNPs among the top 50 ranked SNPs.

In addition to CGs related to auxin transport, *JAZ1* was identified. *JAZ1* plays an important role in jasmonate (JA) signaling and acts within the JA signaling pathway as a transcriptional regulator (Pauwels and Goossens, 2011). *JAZ1* is an early auxin responsive gene, whose transcriptional activation is independent of JA signaling (Grunewald et al., 2009). As outlined in the review of Vanstraelen and Benkova (2012), *JAZ1* might be a link that enables auxin to attenuate JA activity via a feedback mechanism. A closer look at the four associated SNPs of *JAZ1* and its surrounding genomic region revealed that these SNPs have lower p-values compared to the SNPs in vicinity. This is not only true for the WILCOXON mapping method, but also for the MLM mapping (see Figure F3.28 A). The detailed SNP structure analysis and following LD plots verified the high linkage of these four SNPs (see F3.28 C).

By further inspecting *JAZ1* haplogroups in a more dense SNP data set using 14 *A. thaliana* accessions, which were used in this thesis and overlap with the data of Clark et al. (2007), linkage with a nonsynonymous SNP at position 6622990 (according to TAIR10) was detected (see Supplementary Figure S29 A). This SNP causes an amino acid change from Leucine to Valine, which most likely have no strong effect on protein structure changes. However, the *JAZ1* gene sequence was investigated in the data produced for the population genetic analysis.

The CDS haplotypes were analyzed with the SplitsTree4 software (Huson and Bryant, 2006) and a median joining network was computed according to Bandelt et al. (1999). The median joining network represents the individual haplogroups of a multiple sequence alignment and connects each node (haplogroup) by an edge according to the sequence changes steps from one to another haplogroup. To have a closer look at the *JAZ1* SNPs, the data of Cao et al. (2011) were used, since the SNP density is higher than the data of Clark et al. (2007). Here, one major haplogroup (blue circle) and an outlier group (yellow circle) could be detected for *JAZ1* (see Supplementary Figure S29 B). The outlier group shares the same allelic composition as detected in the SNP data of Clark et al. (2007).

Taken together, the SNPs identified in *JAZ1*, which associate with 2,4-D HGI treatment, seem to separate *A. thaliana* into two main haplogroups, which might also show different responses to 2,4-D. A further phenotypical characterization of HGI after 2,4-D treatment with additional evaluation of the expression differences of *JAZ1* between this outgroup against the major haplogroup might clarify, if *JAZ1* has an influence on the observed phenotypic differences.

Surprisingly, five *ARFs* but no *AUX/IAAs* were detected as CGs among the top 50 ranked SNPs. As shown by Fukaki et al. (2005) and Fukaki and Tasaka (2009), *ARF7* and *ARF19* contribute to lateral root formation and are expressed mainly in the roots (Teale et al., 2006). Here, both *ARFs* appeared on the list of CGs in the context of HGI. For the CG *ARF7* two SNPs were associated with HGI IAA, this can be explained by the fact, that the detected SNPs are in close LD as given in Supplementary Figure S30.

However, if the top 500 ranked SNPs were evaluated, *AUX/IAAs* (*IAA4*, *IAA10*, *IAA28* and *IAA32*) were also detected, so that conclusions out of this bias toward more detected *ARFs* at the top 50 ranked SNPs would be doubtful.

Among the 36 detected *a priori* CGs, five CGs showed up with direct support of both GWA mapping methods. Sawa et al. (2002) could show that over-expression of *HAT2* leads to decreased RGI after NAA treatment. Interestingly, it was recently shown for *HAT2*, that its expression is altered upon picloram treatment (Chapman et al., 2012). However, here *HAT2* was detected as a CG for RGI after 2,4-D treatment. If the association of *HAT2* with the 2,4-D response trait is true, it might have an influence as a global player, not specific to a single auxin derivative.

In addition to *HAT2*, *NDL1*, *5PTASE2*, *ARF13* and *AUF2* were detected with MLM and WILCOXON GWA mapping. Since *AUF2* was also detected with *v*GWAS mapping, this CG will be discussed in more detail in the next subsection.

Eleven CGs were found with MLM only and 20 CGs with WILCOXON only. As expected, the 36

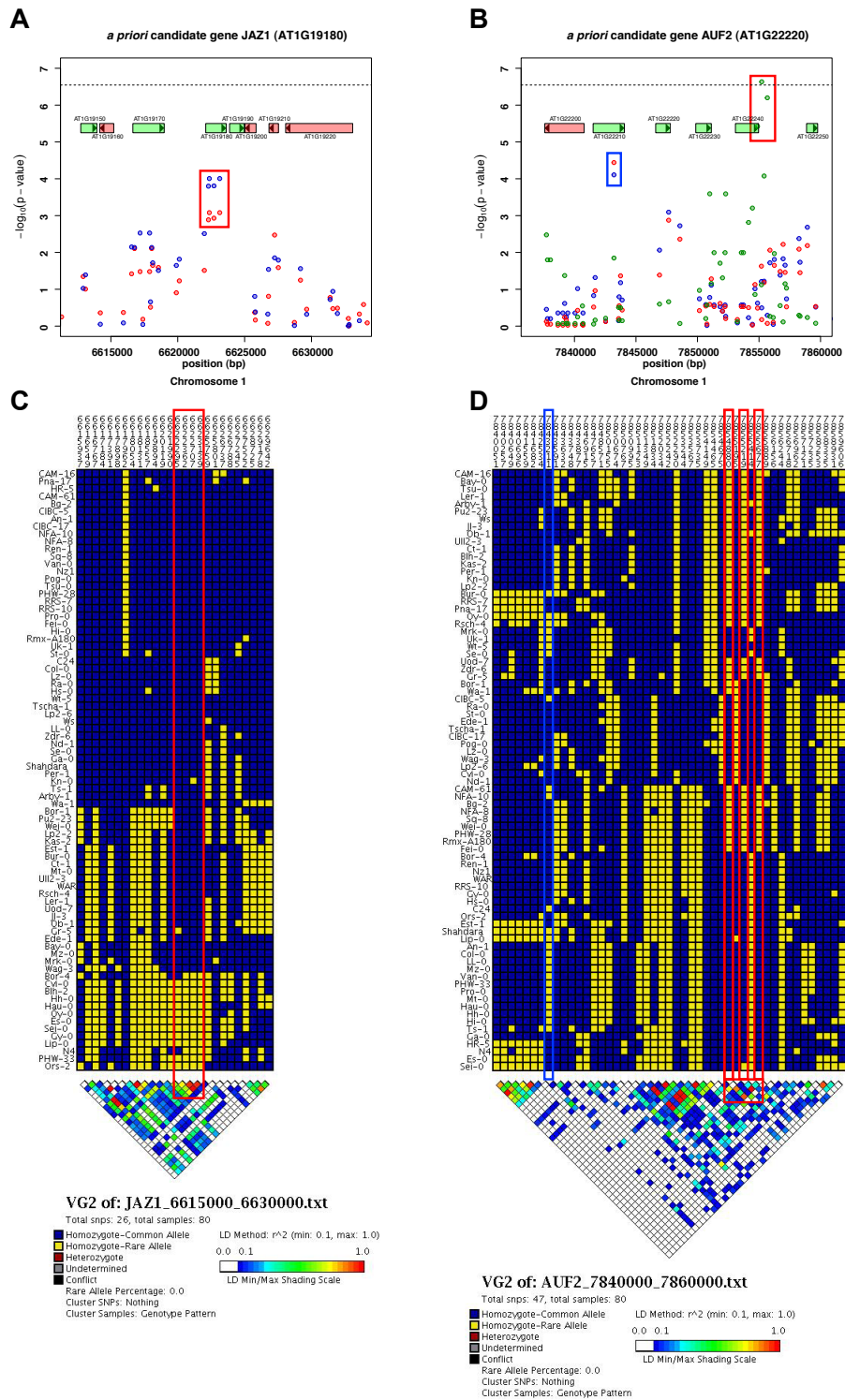


Figure F3.28: Detailed SNP structure analysis of the *a priori* candidate genes *JAZ1* and *AUF2*. Manhattan plots of GWA mapping results for the *a priori* CGs *JAZ1* (A) and *AUF2* (B) and their genomic vicinity. Distribution of $-\log_{10}$ p-values calculated with MLM (red dots) or WILCOXON (blue dots) mapping methods. Annotated genes according to TAIR10 are plotted on top of each manhattan plot (arrowhead indicates gene orientation). (C + D) On top, the SNP positions for the CGs are shown. Each row represents the SNP composition in the 80 *A. thaliana* accessions. Minor alleles are shown in yellow, major alleles in blue, accessions were clustered according to their haplotypes. At the bottom, LD calculated for pairwise SNP comparison is shown. Blue colors indicate low, red colors high r^2 values. Colored boxes highlight same SNPs in manhattan and LD plots.

CGs are not specific to one kind of auxin biology. They have an influence on auxin transport, auxin metabolism, auxin signaling and even on crosstalk between auxin and other plant hormone pathways.

Phenotypic variance evaluated by vGWAS mapping

The recently published GWA mapping method vGWAS (Shen et al., 2012), which considers differences in phenotypic variance between two allelic groups, was applied to the same phenotypic data, and accordingly, the associated SNPs were analyzed for *a priori* CGs.

Applying the same rules as for the MLM and WILCOXON GWA mapping methods, 20 CGs could be detected, for which at least one associated SNP falls under the top 50 ranked SNPs in a window of 10 kb around the annotated gene position. The CGs are listed in Table T3.9.

For one CG (*ANT1*) the detected SNP falls within the gene position. It was investigated, whether it causes an amino acid change. As given in Supplementary table ST17, the SNP is nonsynonymous and might therefore alter the protein function of *ANT1*. However, as outlined by Long and Barton (1998), the ant mutant phenotype is limited to flowers and the detected SNP might just be a false positive. Alternatively, the ant mutant could be phenotyped under the same growth conditions as applied in this thesis to evaluate a possible effect of *ANT1* on RGI.

Table T3.9: List of *a priori* candidate genes associated with auxin found among the highest correlated SNPs with the vGWAS GWA mapping method.

Locus	Gene name ^a	Overlapping QTLs ^b	RGI 2,4-D vGWAS ^c	RGI IAA vGWAS ^c	RGI NAA vGWAS ^c	HGI 2,4-D vGWAS ^c	HGI IAA vGWAS ^c	HGI NAA vGWAS ^c	SNP RANK ^d
AT3G11900	ANT1	hIAA-III.1	1	-	-	-	-	-	43
AT1G22220	AUF2	h24D-I.2,h24D-I.2	4	-	-	-	-	-	1,2,4,21
AT1G34410	ARF21	hNAA-I.1	-	-	-	1	-	-	4
AT3G17185	TAS3	r24D-III.1,hIAA-III.1	-	-	-	-	8	-	6,18,19,20, 21,22,27,29
AT1G48350	EMB3105	rNAA-I.1	-	-	-	1	-	-	8
AT3G01220	ATHB20		-	1	-	-	-	-	9
AT4G25960	ABC2	h24D-IV.1,hNAA-IV.1	-	1	-	-	-	-	10
AT5G59220	HAI1	rNAA-V.1,rNAA-V.2,hIAA-V.1	1	-	-	-	-	-	12
AT1G23320	TAR1		-	-	1	-	-	-	12
AT1G69960	PP2A		-	-	-	1	-	-	15
AT1G22920	CSN5A	h24D-I.2,h24D-I.1	-	-	1	-	-	-	16
AT1G34390	ARF22	hNAA-I.1	-	-	-	1	-	-	22
AT2G47460	MYB22	rIAA-II.2	-	-	-	-	1	-	24
AT1G49010	AT1G49010	rNAA-I.1	-	-	-	-	1	-	25
AT2G25930	ELF3	rNAA-II.1,h24D-II.1,hNAA-II.2	-	-	-	-	-	1	28
AT3G22942	AGG2	r24D-III.1,hIAA-III.2	-	1	-	-	-	-	35
AT3G19580	ZF2	r24D-III.1,hIAA-III.1	-	2	-	-	-	-	40,48
AT1G72430	AT1G72430		-	-	-	-	2	-	41,47
AT4G26200	ACS7	h24D-IV.1,hNAA-IV.1	-	-	-	1	-	-	41
AT3G63010	GID1B	h24D-III.1	-	-	-	-	-	-	43

^a Bold written gene names indicate that the found SNP lies within defined gene start and end position.

^b Overlapping with 95% CI for detected QTLs, abbreviation corresponds to Table 4.3.

^c Indicates the number of times that the gene was detected among the top 50 ranked associations with vGWAS mapping method in a 20 kb window.

^d Indicates the rank of the SNPs among the top 50 ranked SNPs.

As already mentioned, only one significant SNP could be detected by applying conservative threshold values for GWA mapping results. This significant SNP falls in the vicinity of *AUF2*. Three other SNPs appeared under the top 50 ranked SNPs with close position to the significant SNP for RGI 2,4-D. This could be explained by LD analysis as given in Figure F3.28. The *A. thaliana* accessions

can be divided into two allelic groups by the SNP at position 7855219 of chromosome one. The group carrying a thymidine (white boxplot) shows a wider phenotypic range compared to the allelic group carrying a guanine (grey boxplot) at this position as highlighted in figure F3.29 A. According to the model of the *v*GWAS mapping method this SNP can explain $\sim 15\%$ of the variance differences between these two groups.

Since the SNPs fall not directly into the annotated gene position of the CG *AUF2*, one can hypothesize that the SNP influences its gene expression levels. As I illustrated in figure F3.28 B, there might be other genes influenced by this SNP. Furthermore, the possibility of being a false positive can not be ruled out. Zheng et al. (2011) proposed that *AUF2*, if active, plays only a minor, more constitutive role in the context of root elongation. These conclusions were based on finding that no exaggerated phenotypes of the *auf1-2 auf2-1* double mutants were identified compared to the *auf1-2* single mutant. Interestingly, *AUF2* was also detected as a CG by MLM and WILCOXON mapping methods. Here, the SNP causes a difference between allelic groups based on phenotypic mean values influencing RGI after NAA treatment (see Figure F3.29 B).

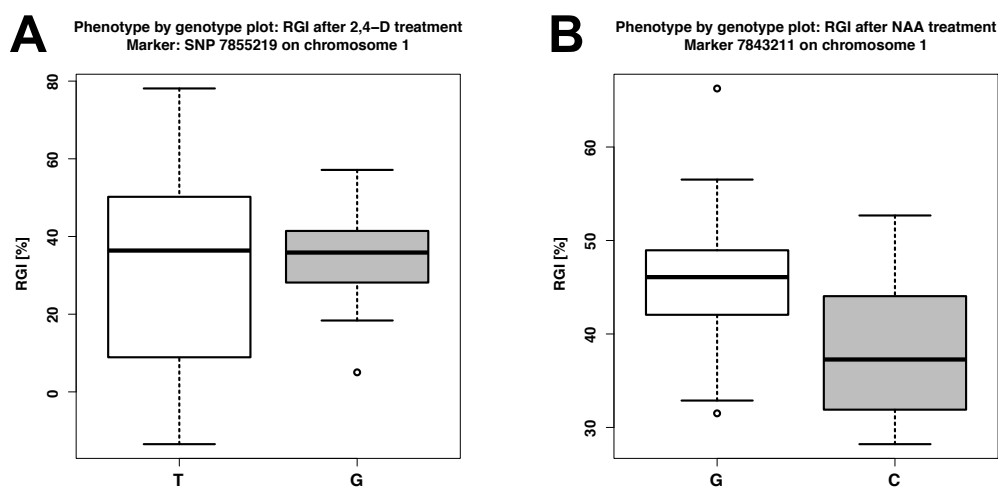


Figure F3.29: Phenotypic variance and phenotypic mean differences in 80 *A. thaliana* accessions.

(A) Phenotype by genotype plots for SNP 7855219 on chromosome one in the 80 *A. thaliana* accessions are shown as boxplots and grouped by their corresponding genotype. Either carrying a thymidine "T" (white) or a guanine "G" (grey) at the specified SNP position. (B) Phenotype by genotype plots for SNP 7843211 on chromosome one in the 80 *A. thaliana* accessions are shown as boxplots and grouped by their corresponding genotype. Either carrying a guanine "G" (white) or a cytosine "C" (grey) at the specified SNP position.

The enrichment analysis of *a priori* CGs in the 50 and 100 top-ranked SNP categories showed, that there is significant enrichment upon 2,4-D treatment influencing RGI. For the other traits a slight but not significant enrichment of CGs, which are related to auxin biology, were identified. However, also under-representations of CGs were detected for NAA HGI treatment.

Only one CG among the top 50 ranked SNPs, namely *AUF2*, was detected to have an influence on different auxin response traits. Among the detected CGs based on the top 50 ranked SNPs and within a window of 10 kb around the annotated CG position, *ARFs* but no *AUX/IAAs* showed up.

Also some of the identified CGs overlap with QTL regions identified in this thesis within RIL populations with Col-0 and C24 genotypes (see Table T3.8 and T3.9). Consistent, according to the auxin treatment, the CG *JAZ1* overlaps with QTLs detected after 2,4-D treatment measuring HGI in dark-grown seedlings. In close vicinity, *ARF19* is located and falls also within the same QTL region.

Discussion

The aim of the thesis was to consider natural occurring variation upon auxin response traits in *A. thaliana* from various angles. In young seedlings the plant morphology, established during embryogenesis, is formed and alterations in this process might lead to selection constraints in interconnection with the surrounding environment.

First, by conducting a population genetic analysis, the question whether genetic natural variation exists for auxin network genes was answered. A closer look at auxin signaling genes was undertaken, to answer the question, which gene family among the auxin signaling sub-group of auxin network genes might contribute at most to phenotypic variation upon auxin response traits.

To gain knowledge about the complex auxin response traits in young seedlings, quantitative genetic approaches were used to identify loci and their impact on natural variation. The goal was to identify genomic regions and to validate their impact on natural occurring phenotypic variation in the evaluated *A. thaliana* accessions.

Since classical QTL approaches, using a two parental cross, are restricted to the genetic material of these parental lines, a GWA in 80 *A. thaliana* accessions was conducted, to answer the question, whether the genomic regions identified by the QTL approach also play a role in the global population of *A. thaliana*. In addition, the question was, whether *a priori* CGs linked to auxin biology associate with the observed phenotypic variation.

4.1 Major SNPs, conserved pathways and amino acid changes

A population genetic analyses for auxin network genes was performed in this thesis in which on the one hand genomic fragments of 97 auxin related genes for 18 *A. thaliana* accessions were re-sequenced by Sanger-sequencing and compared to control fragments published by Nordborg et al. (2005). However, the detailed description of population genetic parameters was restricted to a meta-analysis of full-length CDS for auxin network genes based on sequence data of 80 *A. thaliana* accessions (Cao et al., 2011). It is worth to notice, that the main findings of the meta-analysis with full-length CDS were also made for the re-sequenced data, showing that previous studies on partial CDS produced valid findings (Bakker et al., 2006, 2008; Ramos-Onsins et al., 2008; Sterken et al., 2009; Puerma and Agudé, 2013). However, evaluating full-length CDS produce more reliable data.

In this meta-analysis, population genetic parameters for a large fraction of all annotated representative gene models were conducted. By considering 21325 representative protein coding genes in at least 78 *A. thaliana* accessions, a bias towards lower MAF values was observed (see Figure F3.6). This is consistent with previous findings that population genetic parameters within the species *A. thaliana* do not follow the expected distribution according to the neutral model of molecular evolution (Nordborg et al., 2005; Schmid et al., 2005; Wright and Gaut, 2005; Ramos-Onsins et al., 2008). This suggests, that in addition to natural selection also demographic history (recent bottlenecks, extinction/recolonization events and population structure) has shaped the observed allele frequencies.

To be able to address questions related to selection patterns for 124 auxin network genes, the remaining 21201 genes and their population genetic parameters were taken as empirical null distributions.

Major SNPs

As already noted by others (Clark et al. (2007), Cao et al. (2011), Gan et al. (2011)), a large fraction of representative gene models harbour major SNPs between *A. thaliana* accessions. Although the number of loci with major SNPs among the auxin network genes was not significant from the empirical null distribution, several of them were affected in numerous accessions and, thus, justify further discussion. It is important to mention, that most of the genes discussed in this section were not part of the later population genetic analysis, because the number of accessions affected by major SNPs exceeded the filtering thresholds (see materials and methods section 2.1 and results section 3.1.1).

The auxin biosynthesis gene *ASA1* for example, displayed major SNPs in 22 out of 80 investigated accessions. Considering the root growth phenotypes that are the subject of this thesis, it is interesting to note, that *ASA1* also functions in jasmonate-induced auxin biosynthesis (Sun et al., 2009). The authors revealed a role for jasmonate in the attenuation of auxin transport in the root and the fine-tuning of local auxin distribution in the root basal meristem. However, the second splice variant of *ASA1* was not affected at all in these accessions and it needs to be investigated, how much the loss of this splice variant may contribute to natural occurring phenotypic variation. For the remaining auxin synthesis genes, which were excluded from the data, including *YUCCA5* and *YUCCA10*, only a small number of accessions were affected. Together with the high redundancy of *YUCCA* genes in *A. thaliana* (Cheng et al., 2007), a strong impact on possible phenotypic variation in the whole population is unlikely.

Among the auxin signaling gene families, the *AUX/IAAs* were in most cases only affected in a minor fraction of accessions or - other than the annotated representative splice variant - were not affected at all. This indicates, that the affected accessions might have lost a single splice variant, but other splice variants still can be produced and translated to the functional protein.

In contrast to the *AUX/IAAs*, all Class I' *ARFs* (Remington et al., 2004) were affected by major SNPs. Among this group, *ARF23* lacks domain III and IV (Guilfoyle and Hagen, 2007) and is supposed to be a pseudogene (Okushima et al., 2005a). Class I' *ARFs* are exclusively expressed in the endosperm of the pro-embryo in the direct surrounding of the embryo (Rademacher et al., 2011). Interestingly, *ARF17*, which is expressed in contrast to the Class I' *ARFs* throughout the whole endosperm, is not affected by major SNPs, indicating distinct functions for *ARF17* in embryo development.

Seven out of eight Class I' *ARFs* are located near to each other on chromosome one and appear to be products of a recent series of tandem duplications (Remington et al., 2004). According to Ohno (1970), after a gene duplication event, the duplicated gene may retain the same set of functions, retain only a subset of the original set of functions (subfunctionalization), obtain a new function (neofunctionalization), or degrade into a nonfunctional gene (nonfunctionalization) (outlined by Rodgers-Melnick et al. (2012)).

Until the publication of Force et al. (1999), in which the authors introduced the duplication-degeneration-complementation (DDC) hypothesis, the nonfunctionalization hypothesis was favoured for duplicated genes. Due to findings in empirical data, in the DDC model it is predicted, that degenerative mutations in regulatory elements can increase, rather than reduce, the probability of duplicate gene preservation (Force et al., 1999). If a regulatory element is affected by a mutation, one of the two duplicated genes needs to remain unaffected in this element to make sure that the full function of the ancestral gene is retained. Birchler and Veitia (2007, 2010) introduced the gene balance hypothesis. In this hypothesis it is predicted, that an increasing number of protein-protein interactions should favour retention of whole-genome-duplication (WGD) pairs, while disfavoring the fixation of tandem duplication (TD) pairs (Rodgers-Melnick et al., 2012).

As mentioned above, the difference for the Class I' *ARFs* to other *ARFs* is, that they are products of a recent tandem duplication. Interestingly, Paponov et al. (2009) investigated the *AUX/IAAs* and the *ARFs* for site-specific selection and identified for the Class I' *ARFs* hints for positive selection.

Okushima et al. (2005a) showed that single knockouts did not show obvious aberrant phenotypes, indicating functional redundancy among these *ARFs*. The findings in this thesis are consistent with the above mentioned previous results, suggesting subfunctionalization with a possible redundancy and even signatures of nonfunctionalization in the Class I' *ARFs*.

In addition to the Class I' *ARFs*, *ARF7* and *ARF16* were affected in more than two accessions by major SNPs and further experimental approaches - which I will suggest in the following - could validate, whether these SNPs do indeed affect the functionality of the gene(s). As shown by Okushima et al. (2005b), *ARF7* plays a role in lateral root formation and gravitropism in the root and hypocotyl. Since the auxin-induced gene expression is severely impaired in the single mutant *arf7*, the affected accessions should functionally be analyzed for this kind of phenotype.

ARF16 plays a role in distal stem cell differentiation and columella organogenesis, its expression is regulated by miRNA160 (Wang et al., 2005). However, with six accessions, only a minor fraction of 80 accessions was affected. *ARF10* and *ARF17* show high sequence similarity and have similar functions in the same developmental processes. They are both regulated by miRNA160. Here, to proof the loss of *ARF16* function in these six accessions, a knockdown of *ARF10* should lead to a phenotype with a retarded root and insensitivity to gravity, like the *arf10 arf16* double mutant (Lee et al., 2012).

In both cases, for *ARF7* and *ARF16*, only a minor fraction of the 80 accessions is affected. Since the identified major SNPs are deletions which alter the open reading frame (ORF), small insertions, which have not been included in the analyzed data set of Cao et al. (2011), could restore the ORF of these *ARFs* in these accessions. Sequencing of cDNA could show, if the ORF of *ARF7* and *ARF16* is indeed affected.

Regarding the auxin transport, only two genes were excluded. Here, in addition no strong contribution to possible phenotypic variation between accessions is expected from these affected ones (see results section 3.1.1).

Regarding auxin metabolism *GH3.6* showed an altered stop codon in 44 accessions resulting in a longer protein. *GH3.6* is involved in the formation of IAA-aspartate (IAA-ASP) and plays important roles in auxin homeostasis, since IAA-ASP is degraded and with that removes active IAA from the plant cell's auxin pool (Staswick et al., 2005). The important role of *GH3.6* suggests, that the longer protein of *GH3.6* in these affected accessions is functional. However, a contribution to phenotypic variation caused by these altered *GH3.6* protein can not be ruled out.

Furthermore, *ILL5* and *ILR2* were affected in eleven accessions. It was shown previously by Magidin et al. (2003), that *ILR2* is polymorphic for different *A. thaliana* accessions with one long protein version for twelve accessions including Col-0 and a short protein version for the accessions Ws and Ler. Considering these facts, *ILR2*, which was also detected within the re-sequenced data as a gene containing major SNPs, represents a strong candidate that can contribute to phenotypic variation between the investigated accessions.

Nucleotide variation and genetic variation patterns in auxin network genes

The population genetic analysis can be divided into three different sub-categories. First, the level of nucleotide variation. Here, the nucleotide diversity (π), which is largely independent of fragment length, was used as the summary of the level of nucleotide variation between the *A. thaliana* accessions (Puerma and Aguadé, 2013). Second, the pattern of genetic variation within one species. Here, neutrality tests, like Tajima's D conducted within the species *A. thaliana*, were used as the summary of the pattern of genetic variation within one species. Third, the pattern of genetic variation between two species. Here, the McDonald-Kreitman test, which is based on the polymorphism to divergence ratio, was used as a summary of the pattern of genetic variation between two species. All these categories have to be considered to propose, whether adaptive processes have been involved in shaping

the observed genetic variation for one gene.

One hundred twenty-four auxin network genes could be analyzed for genetic variation patterns. If hints for adaptive selection were detected, in most cases they could be assigned to purifying selection. In addition to signatures of purifying selection, in some cases also the evidence for balancing selection became apparent.

Consistent with the assumption that the majority of synonymous sites is neutral, π_{SYN} is not significantly different from the empirical null distribution for any evaluated auxin network group (see Figure F3.3). For π_{NONSYN} the auxin network genes treated as one group tend to maintain lower levels (but not significant) as compared to the empirical null distribution (see Figure F3.2 and F3.3). The same kind of pattern is also observed when the minor allele frequencies are taken into account. Here, the auxin network genes treated as one group show no significant difference to the empirical null distribution for the synonymous MAF, whereas the nonsynonymous MAF is significantly shifted towards rare alleles (see Figure F3.6 and F3.7).

These observations are possible indications for negative selection. In concordance with that, the comparison of *A. thaliana* auxin network genes to their orthologs in closely related Brassicaceae revealed, that all auxin network genes showed $\frac{K_{NONSYN}}{K_{SYN}}$ -ratio values < 1 , which suggests purifying selection for the majority of auxin network genes, at least in comparison to these Brassicaceae (see Table T3.1).

However, a high degree of variation is apparent among the auxin network genes and outliers, which fall in the extreme tails of the empirical null distribution, were detected for within and between species comparison.

When the auxin network group was divided into sub-groups, auxin transport and auxin signaling genes are significantly different from the empirical null distribution for certain parameters. For both groups the combined π_{NONSYN} values are lower than the empirical null distribution, which most likely suggests purifying selection acting on these genes. The auxin transport genes additionally show significantly low levels for S_{NONSYN} and protein variants per codon, which underpins the detected evidence for purifying selection. Furthermore, over-representation of rare alleles for the nonsynonymous MAF supported these conclusions (see Figure F3.7).

As reviewed by Vanneste and Friml (2009), local auxin gradients are strongly dependent on auxin transporter function and lead to differential developmental processes. Potentially, different accessions may fine-tune auxin gradients in slightly different ways, which might then also lead to different downstream auxin responses. Since the auxin transporters are the key players in maintaining local auxin gradients, they could be directly causing such differential auxin gradients.

However, strong signatures of purifying selection for the auxin transporters make such a scenario highly unlikely. Of course, single nonsynonymous mutations might lead to different flux rates, which then might lead to auxin gradient differences between accessions, and can not be ruled out as a possible source for phenotypic variation. However, the results for the within species $\frac{\pi_{NONSYN}}{\pi_{SYN}}$ -ratio (see Figure F3.4) and the between species $\frac{K_{NONSYN}}{K_{SYN}}$ -ratio (see Table T3.1) suggest a strongly conserved nature of the auxin transporters. In addition, also the analysis of transcript diversity showed, that the auxin transport genes are not significantly different from the empirical null distribution.

Considering these facts, another hypothesis might be, that the distribution pattern of auxin transporters might be differentially regulated, leading to differences in auxin gradients and downstream processes. In this scenario, the auxin transporters would retain the same function, same expression level and flux rates for auxin, but the arrangement of the auxin transportes facilitated by other proteins would be different. But such genes are either unknown or did not yield signatures that would support such a scenario (*PID*).

Based on nucleotide diversity and genetic variation patterns, the auxin synthesis group showed a broad spectrum of genetic variation. There was no bias for one type of selection pattern among these genes throughout the assessed population genetic parameters. The MAF distribution with a over-representation of rare alleles for nonsynonymous sites suggested negative selection acting on the auxin synthesis group. However, some auxin synthesis genes fall into extreme tails of the empirical null distribution, one of them is *NIT1*. *NIT1* showed in all comparisons, conducted to its Brassicaceae orthologs, evidence for positive selection.

As reviewed by Mano and Nemoto (2012), the IAA biosynthesis can be split into two parts. The first part is the TRP synthetic pathway, starting with chorismate and ending with TRP. The second part consists of four TRP-dependent distinct pathways, two of them lead to IAA via the intermediates IPA and IAM. One route branches into the synthetic pathway of indole alkaloids and Serotonin via the intermediate TAM. Another is supposed to be a Brassicaceae specific IAA biosynthesis route.

It was shown previously for the anthocyanin pathway that genes, which encode downstream processes within the pathway, show an increased rate of change at nonsynonymous sites when compared with upstream acting loci. This might be due to relaxed constraint (Rausher et al., 2008). Livingstone and Anderson (2009) could show a similar pattern also for the carotenoid biosynthetic pathway.

By introducing the pathway pleiotropy index, Ramsay et al. (2009) identified a correlation of the $\frac{K_{NONSYN}}{K_{SYN}}$ -ratio and the position along pathways converting glucose to the terpenoid phytohormones (e.g. abscisic acid, brassinosteroid and gibberelic acid). Here, the connectivity and pathway position seem to play major roles of defining selection constraint.

It would be interesting to see, whether these above mentioned findings also apply for the different routes of auxin biosynthesis. However, testing this hypothesis would be complicated by possible functional redundancy and interconnectivity, which makes an explicit position assignment along the pathways quite complicated.

Since for most of the enzymatic reactions in these parallel pathways, there exist more than one gene, which can catalyze the needed reaction. One explanation, that the auxin synthesis genes appear neutral in comparison to the empirical null distribution might be, that there is redundancy among genes in the first and second part of the biosynthesis, like the *PAI* and *YUCCA* genes.

Hence, all pathway positions are integrated in the auxin synthesis group, this might explain the broad range of genetic variation pattern in this auxin network group. Further elucidation of the data considering sub-pathways and genes with high connectivity may show the same kind of relaxed constraint in downstream enzymes as identified for the above mentioned synthesis pathways.

Although the auxin metabolism genes did not reveal significant differences for nucleotide diversity and genetic variation patterns, they showed the highest values for π_{SYN} (significance was lost due to multiple testing correction). These observations were supported by a significant enrichment of intermediate frequency alleles for synonymous sites as identified by MAF distribution analysis, which is a hallmark of positive selection (see Figure F3.7). In addition, the auxin metabolism group showed the highest values for π_{NONSYN} , S_{NONSYN} , protein variants per codon (see Figure F3.3), H_d and D_{CDS} . As this seems intuitively mutually exclusive, how can the concurrent appearance of high synonymous and nonsynonymous nucleotide diversity rates be explained?

As outlined by Ludwig-Müller (2011), auxin conjugating enzymes can regulate auxin levels and are the source of auxin degradation. Among these auxin conjugating genes, one important group is the *GH3* gene family. Several of the *GH3* genes are auxin-inducible and are supposed to act in a negative feedback loop to remove free auxin by their auxin conjugating activity. In addition to the *GH3* gene family, other proteins facilitate individual auxin conjugation functions. Here, redundancy might explain the concurrent appearance of high synonymous and nonsynonymous nucleotide diversity

rates. Some auxin conjugating enzymes need to keep their function and are conserved, whereas others might need to evolve subfunctions to constantly adapt to the local environment. Such more diverse metabolism genes might facilitate then the fine-tuning of auxin conjugation.

For *GH3.4*, the evaluated population genetic parameters fall into the upper extreme tail of the empirical null distribution. According to Supplementary Figure S13, *GH3.4* is located in a genomic region, which shows high values for the three evaluated nucleotide diversities (π_{SYN} , π_{NONSYN} and $\pi_{introns_flanking}$) and rather low gene density. Here, these extreme cases might contribute to phenotypic variation and further investigation of their contribution needs to be done. A first step to unravel the concurrent appearance of high π_{SYN} and π_{NONSYN} values would be to also analyze codon usage of the auxin metabolism genes and compare it to an empirical null distribution to detect possible selection signatures even for synonymous sites. Controversially, for *GH3.4* the Mc-DonaldKreitman-Test suggests purifying selection between *A. thaliana* and the evaluated Brassicaceae. This might be due to the excess of mutations and this finding needs to be seen with caution.

IAR1, which is a metal-ion transporter and might alter hydrolase activity in certain compartments of the cell (Rampey et al., 2013), shows signatures of positive selection (see Table T3.2) due to an excess of intermediate frequency alleles, which was also identified by elucidating the individual MAF distribution (see Supplementary Figure S3). In addition to *IAR1*, also for *GH3.4* an excess of intermediate frequency alleles was identified.

In addition to the high nucleotide diversity values, the auxin metabolism genes showed significant different TRCDIV values. Here, again the highest values among the auxin network groups were identified, which suggests a possible role for the auxin metabolism genes in shaping phenotypic variation between *A. thaliana* accessions.

The more global survey of auxin network genes treated as functional groups (biosynthesis, metabolism, transport and signaling) revealed already signatures of negative selection for the auxin signaling genes. This was reflected by significant lower values for π_{NONSYN} (see Figure F3.4) and lower, but not significant values for genetic variation patterns (see Figure F3.5).

Here, considering the MAF distribution, for the auxin signaling genes the opposite as compared to the auxin metabolism genes was true. In all categories, the MAF distribution was shifted significantly to lower values as compared to the empirical null distribution, indicating an excess of rare alleles.

If this kind of pattern is observed for a single gene, the shift of all MAF values towards low values, might be then due to genetic hitchhiking ("selective sweep"), the loss of genetic variation at neutral loci when a new beneficial allele arises nearby and is fixed in the population (Kim, 2006). However, a scenario, in which multiple sweeps occur in the close vicinity of the majority of auxin signaling genes and cause the low MAF values, is very unlikely.

To further elucidate, which auxin signaling group contributes to this pattern and to answer the question, which gene family might contribute at most for phenotypic variation among the auxin signaling group, a deeper analysis restricted to auxin signaling was conducted.

The auxin signaling group: Inside and insights

By assessing nucleotide diversity levels, the auxin co-receptor family, the *TIR1/AFBs*, showed significantly lower values for π_{NONSYN} , S_{NONSYN} and protein variants per codon (see Figure F3.8). This was followed by the *ARFs* and then *AUX/IAAs*, among which the latter showed the highest values for these parameters. Consistently, the *AUX/IAAs* showed the highest values for D_{CDS} , indicating that the *AUX/IAAs* are the most variable gene family among the auxin signaling group.

A closer look at MAF values for the auxin signaling gene families revealed, that despite the *TIR1/AFBs*, which are not significantly different from the empirical null distribution, all other MAF values were significantly lower than the null distribution, which is a hallmark of negative selection (see Figure F3.7). However, this strong bias towards rare alleles needs to be evaluated further, and a codon usage analysis might unravel selection patterns for synonymous sites.

Consistent with the nucleotide diversity values, the *AUX/IAAs* showed also the highest TRCDIV values among the auxin signaling genes. Maere et al. (2005) could not only demonstrate the existence of three whole-genome-duplications (WGDs) (α , β and γ) for *A. thaliana*, but could also demonstrate, that genes involved in signal transduction are more likely to be retained after such events, like the *AUX/IAAs* and *ARFs*. As shown by Remington et al. (2004), within the *AUX/IAAs* and the *ARFs*, so called "sister pairs" exist, which are paralogs with high sequence similarity. As outlined by Van de Peer et al. (2009), according to Freeling and Thomas (2006), after a genome duplication event, entire functional modules are inherently retained duplicate through non-adaptive dosage balance effects.

To consider the complex scheme of this "sister pairs" among auxin signaling genes, which were retained after whole-genome duplications events, a detailed analysis should be considered for these paralogs with high sequence similarity based on population genetic parameters and expression profiles between *A. thaliana* accessions.

General conclusion for the population genetic analysis

Taken together, I identified several layers of genetic variation on the sequence level, which potentially might contribute to phenotypic variation. One source might be the loss of function of some genes in a subset of accessions via major SNPs.

The auxin transporters show clear evidence for negative selection, suggesting if only an indirect contribution to phenotypic variation by their cellular distribution pattern and not their conserved function to build local auxin gradients.

The auxin synthesis genes showed a broad range of genetic variation, with a possible constraint to conserved enzymatic function, at least for positions in the pathway which show no gene redundancy. A further elucidation of the whole biosynthesis pathway, by taken pathway position and connectivity into account, will shed light on this issue.

Despite the fact that the Class I' *ARFs* were heavily affected by major SNPs. Among the major SNP filtered data, auxin signaling genes treated as one group show hints for negative selection, which represents their conserved function in the regulation of auxin responsive gene expression. In all groups outliers were detected, which fall into the lower tails for population genetic parameters. Consistent with these results, within the auxin signaling group, the auxin co-receptors showed the highest degree of conservation.

Delker et al. (2010) proposed a model about the transcriptional regulation of auxin dependent responses. In this model, the basal auxin response is caused by the equilibrium between *AUX/IAAs* and *ARFs* to regulate auxin dependent downstream responses. Upon auxin stimulus, a shift in the composition and relative amount of *AUX/IAA* and *ARF* proteins enables downstream responses (Delker et al., 2010).

The *AUX/IAAs* showed the highest values for nucleotide diversity levels and genetic variation patterns. In concordance with that, they showed the highest TRCDIV values upon auxin treatment among six *A. thaliana* accessions. This findings support the model in such a way, that the *AUX/IAAs* contribute at most to variation in downstream responses and following the model also to variation in physiological responses between different accessions. Here, it is suggested that differences in gene expression between accessions can fine-tune the negative feedback role of *AUX/IAAs* in the auxin signaling pathway.

The auxin metabolism genes showed the highest values for the investigated population genetic parameters, which hints for positive selection acting at least on some genes for these group. In concordance with the high TRCDIV and the hints for positive selection, the auxin metabolism genes most likely contribute to phenotypic variation in *A. thaliana*.

Here, natural variation was assessed based on sequence information differences of many genotypes. In the following part, the results for a classical QTL analysis are discussed. By conducting QTL analysis, the natural variation of auxin response traits is investigated on a different level, one might say on a functional level. If one genomic region, which is correlated with the observed phenotypic variation, is identified by the QTL approach, the above mentioned population genetic approach can assist to pinpoint possible CGs within this genomic region. However, overlaps between the approaches are not necessarily conclusive, since the focus of the population genetic analysis is on a global scale, whereas the classical QTL approach is based merely on two distinct genotypes and not predictive but functional information is gained.

4.2 QTL analysis of auxin response traits

Many genes that influence plant fate have been identified, which play a direct or indirect role in auxin biology (Vanstraelen and Benkova, 2012). Most of these genes were identified by mutation analyses. The *A. thaliana* genotypes used for these mutation analyses were restricted to only a small number of accessions. To take advantage of the naturally occurring variation within *A. thaliana*, a quantitative genetic approach was conducted. Two RIL populations were used to unravel genomic regions via QTL analysis, which might cause the observed natural variation upon auxin response traits in *A. thaliana*.

I analyzed three different auxins in two RIL populations. The successful development of a software for high-throughput phenotyping (RootDetection, see results chapter) enabled the measurement of plant features like primary root length and hypocotyl length for more than 50.000 single plants.

Nevertheless, the analysis of both endogenous and synthetic auxins clearly went on the costs to conduct replicates of the QTL analysis for the same auxin species. However, as the two RIL populations were derived by a reciprocal cross (Törjék et al., 2006), the same genetic background was screened twice per treatment. The benefit of using two RIL populations derived by a reciprocal cross lies in the ability to identify possible maternal or paternal effects.

Tan et al. (2007) showed for the native auxin IAA and the synthetic auxins 2,4-D and NAA, that they are perceived in a similar manner by the co-receptor *TIR1*. However, they show different binding affinities to the co-receptor (IAA > NAA > 2,4-D) (Tan et al. (2007) and Kepinski and Leyser (2005)). As outlined by Rahman et al. (2007), albeit overlapping pathways with the native auxin IAA, the sensitivity to the synthetic auxin 2,4-D, which can influence primary root growth in *A. thaliana*, can be influenced by the protein *SMAP1*, which does not alter the sensitivity to IAA. While IAA and NAA, activate in first instance cell elongation, 2,4-D activates the division pathway (Campanoni and Nick, 2005). NAA, not IAA or 2,4-D can diffuse through membranes (Delbarre et al., 1996; Yamamoto and Yamamoto, 1998). Despite these contrasting modes of action, the exogenous treatment of young seedlings with these auxin species all lead to RGI. The same mode of actions might be true for the treatment of dark-grown seedlings leading to HGI. In concordance with these different modes of actions, only low correlation coefficient values were obtained between the conducted experiments (see Table T3.4).

In two RIL populations ($Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$), I observed high values for broad-sense heritability for the quantified auxin response traits RGI and HGI, ranging from 0.65 to 0.85 (see Table T3.3). For almost all quantified traits also signatures of transgression were observed, which is a common phenomenon due to the allele combinations in a RIL population. The high values of broad-sense heritability demonstrate the strong impact of genetic factors, which contribute to the observed phenotypic

variation upon auxin response traits between the Col-0 and C24 accession.

By multiple QTL mapping, 27 QTLs in both RIL populations and all conducted treatments could be detected. The magnitude of the identified QTLs explained between 5 and 10% of the phenotypic variation. In addition to these small effect QTLs, also three large effect QTLs were detected. Salvi and Tuberosa (2005) outlined that small effect QTLs are identified more often than large effect QTLs in complex traits. As recently reviewed by Weigel (2012), for the successful cloning of the detected QTLs, it is the other way around. The vast majority of QTLs and their underlying genes have not been cloned so far, with a bias for successful cloning towards large effect QTLs. To increase the chance of QTL cloning of the identified QTLs, these promising QTLs were evaluated further. Among these large effect QTLs, the amount of explained phenotypic variance was higher than 15% (see Table T3.5).

However, depending on marker density and RIL population size, QTL regions can harbour thousands of genes. To be able to clone a QTL, first the QTL region needs to be validated and narrowed down to a size to do fine-mapping. As already mentioned in the results section, NILs and HIFs can be used for this crucial step.

One large effect QTL was found in the $Q_{Col-0 \times C24}$ RIL population (hIAA-IV.1). For the same treatment there was no QTL detected in this region for the $R_{C24 \times Col-0}$ RIL population. This can be due to maternal or paternal effects, or, which is the most likely scenario, just due to experimental variation. A lot of samples were removed from this growth experiment in the $R_{C24 \times Col-0}$ RIL population, so that the results need to be reflected carefully. Here, to rule out the very interesting possibility of maternal or paternal effects, repeating this experiment is mandatory but could not be conducted during the course of this thesis. Preferably, the experiment would be done for both RIL populations in one run by using a random design. Hence, a further selection of NILs and validation was not conducted.

Nevertheless, meta-analysis for the QTL hIAA-IV.1 revealed one interesting candidate among the 95% CI. The candidate gene is *IRON-REGULATED TRANSPORTER 1 (ITR1)*, which showed by far the highest TRCDIV in the expression data of Delker et al. (2010) among six *A. thaliana* accessions (among these six accessions also Col-0 and C24 are present; data not shown). This indicates at least the evidence of differential transcriptional regulation upon IAA treatment for *ITR1*. *ITR1* is a transporter essential for iron uptake from the soil. As outlined by Wu et al. (2012) there are contradicting results, if exogenous auxin treatment might mimic iron deficiency. However, because the validation of this QTL region was not done so far, the contribution of *ITR1* and other genes in this QTL region are just speculative.

Two QTL regions were identified, which harbour large effect QTLs in both RIL populations. Hence, the focus for the validation step was set on these two QTL regions for the corresponding auxin treatments.

The QTL region on chromosome two, which was identified after NAA treatment for HGI, harbours QTL hNAA-II.1 and QTL hNAA-II.2. This region could not be validated reliably. Here, for both RIL populations an additional weaker QTL was identified on chromosome five with an opposite effect for the allelic composition (see Table T3.5). These observed opposite allelic effects might complicate the validation with a NIL population, since only one genomic region is introgressed in another genomic background. Here, two NILs with corresponding introgressions for the QTLs on chromosome two and five would firstly need to be crossed to generate a NIL with two introgression. This step was not done during the thesis. In addition, multiple small effect QTLs were detected in the $R_{C24 \times Col-0}$ RIL population. The individual contribution of these scattered QTLs to the whole phenotypic variation might explain, why this promising QTL region could not be validated.

One QTL region on chromosome five, which was identified after NAA treatment for RGI, could be validated by the use of appropriate NILs (see Figure F3.22). For the $Q_{Col-0 \times C24}$ RIL population the QTL rNAA-V.2 showed the highest LOD peak among all identified QTLs, explaining ~26% of the phenotypic variation. The 95% CI harbours more than 1000 genes. This makes a prediction of possible candidate genes, which might cause the phenotypic variation, very difficult. For the $R_{C24 \times Col-0}$ RIL

population, the LOD peak position of the QTL rNAA-V.1 was shifted and also the 95% CI was larger. Here, additional experiments need to be done to narrow down the validated QTL region.

Since the 95% CI were too large to identify possible candidates, a closer look at *a priori* candidates, genes related to auxin, was conducted. The analysis of *a priori* genes, which are located between the markers V_20 and V_22, revealed some candidates, which show nonsynonymous changes and might therefore yield proteins with functional variation between Col-0 and C24 (see Supplementary Table ST13). Interestingly, three genes related to auxin metabolism are located between these markers. *ILL1*, *ILL2* and *ILL3* all show nonsynonymous changes between Col-0 and C24. Even unique nonsynonymous changes (a mutation is only present in one accession, all other accessions show the same allele like the reference sequence, here Col-0) between Col-0 and C24 for *ILL3* exist in the data from Clark et al. (2007) (see Supplementary Table ST13). As identified in this thesis, the auxin metabolism genes show a high nucleotide diversity between *A. thaliana* accessions, suggesting at least the possibility to influence phenotypic variation. As outlined by Goren and Bukovac (1973), NAA, like IAA, can be conjugated to amino acids. NAA conjugates like NAA-Asp and NAA-Glu can be formed. Rampey et al. (2004) could show for auxin-conjugate hydrolases, including *ILL3*, that auxin metabolism can influence hypocotyl growth, lateral root formation and responsiveness to exogenous applied IAA. Here, the *ILLs* are good candidates for a possible genotype to phenotype correlation.

However, until the QTL region on chromosome five is not narrowed down by back-crossing of suitable NILs and further phenotypic inspection, the possible influence of *ILL3* on the observed phenotypic variation between the accessions Col-0 and C24 remains speculative.

Epistatic interactions are another layer of genotypic relatedness. For RGI after IAA treatment and HGI after NAA treatment, signatures of epistatic interactions were identified by multiple QTL mapping. In addition many additive interactions were present among the evaluated traits in both RIL populations. In most cases complex QTL models with individual small effects were identified, which demonstrates once again the complexity of auxin response traits most likely with many genomic regions involved in the phenotypic outcome of a trait.

To validate these individual small effects, additional information like expression data for all RILs of the evaluated RIL population could shed light on possible candidates within each QTL region. However, such eQTL approaches are cost intensive.

Taken together, the QTL approach in this thesis was restricted to the allelic composition of the accessions Col-0 and C24. The results of QTL validation are not as promising as expected by QTL mapping and heritability calculation. Since many small effect QTLs and QTL interactions have been detected, it is rather unlikely that the phenotypic approach on its own is capable to unravel the evaluated complex auxin response traits between the accessions Col-0 and C24.

A comparison of the QTL analysis results with other QTL studies carried out with other RIL populations under the same experimental conditions might show overlaps for such small effect QTL regions, which then might unravel the important interactions for the auxin response traits.

4.3 GWA analyses of auxin response traits

With GWA mapping, a quantitative genetic approach was chosen to gain insights in natural occurring phenotypic variation within *A. thaliana*. Like QTL analyses, GWA studies try to correlate phenotypic and genotypic data. In contrast to QTL analyses, GWA studies profit from natural occurring recombination events, that shape the genotypes of one species during its life history. Therefore, phenotype-genotype correlations can be identified on a global population scale.

As already outlined in the previous section, the mechanisms, how the evaluated auxin derivatives (2,4-D, IAA and NAA) influence root growth, are partially distinct from each other (Tan et al. (2007), Kepinski

and Leyser (2005), Campanoni and Nick (2005) and Rahman et al. (2007)). However, whether the same distinct mechanisms also apply for hypocotyl growth and the growth inhibiting effects of the different auxin derivatives, is not investigated for natural variation in detail until now.

Here, for GWA analyses, which try to correlate mean phenotypic differences and genotype data, no significant association was found for any of the auxin response traits. The abundance of many moderate significant peaks in all conducted auxin response traits once again emphasizes the polygenic nature of these traits.

A UPGMA clustering of the genetic relatedness of the *A. thaliana* accessions, plotted together with the investigated traits, revealed no visual pattern between the evaluated auxin response traits and the genetic relatedness (see Figure F3.23). Only one significant phenotypic correlation was identified between all pairwise comparison, which would support the above mentioned findings, that partially distinct mechanisms lead to the observed growth inhibition effects (see Table T3.7).

The pairwise geographical distance and pairwise phenotypical distance was used for correlation analyses to unravel possible influence of population structure on the investigated traits. Here, two traits (RGI 2,4-D and RGI IAA) did not show a significant correlation, which suggests hardly any influence of population structure on these phenotypic values. For four traits, a significant correlation was identified (RGI NAA, HGI 2,4-D, HGI IAA and HGI NAA). However, the correlation coefficients ranged from -0.1 to 0.1, suggesting only little contribution of population structure to influence these traits. The information about a possible influence of population structure and genetic relatedness can assist to identify appropriate models for GWA mapping.

Here, in addition to linear models, which take population structure into account, also a non-parametric GWA mapping method was used. As outlined by Filiault and Maloof (2012), the non-parametric method does have two advantages as compared to the linear models. First, it is more robust than the linear models to detect phenotypic mean differences and second, it includes no correction for population structure and with that presents no risk to over-correct p-values for traits that are correlated with population structure (Filiault and Maloof, 2012).

To be able to gain knowledge about the complex auxin response traits, a *a priori* CG approach was used to overcome the limitation of such a situation, where many peaks might contribute to the observed phenotypic variation. To test, whether true associations were identified by the GWA studies, an enrichment analysis of *a priori* CG was conducted in this thesis (see Supplementary Table ST16). A slight enrichment of top ranked SNPs was identified for all traits, but only the 2,4-D RGI trait showed a significant over-representation, which would suggest true association.

Considering the top 50 ranked SNPs and the *a priori* CGs related to auxin, 36 genes were identified, which might contribute to the phenotypic variation upon auxin response traits. However, since the enrichment analysis was only significant for the 2,4-D RGI trait, the possibility of false positives associations among the traits can not be ruled out.

If both, the parametric and the non-parametric method, are combined and only those *a priori* CGs are considered, which fall for both methods in the upper ranked SNP classes, these might be good candidates to further concentrate on in upcoming validation experiments.

Among this kind of CGs, *HAT2* was identified, for which a direct connection to RGI after NAA treatment was already shown (Sawa et al., 2002). However, here, *HAT2* was identified after 2,4-D treatment as a possible CG, which influences RGI. Further expression analyses between different *A. thaliana* accessions might validate this association.

Interestingly, *ARF7* and *ARF19* were identified by the non-parametric model as possible CGs. For both genes an involvement in lateral root formation was already shown (Fukaki and Tasaka (2009) and Teale et al. (2006)). Here, *ARF7* was identified for HGI after IAA treatment and overlaps with the QTL region of hIAA-V.1. *ARF19* was identified for HGI after 2,4-D treatment and falls within the QTL region of h24-D-I.1 and h24-D-I.2. Since the treatment for which both genes were identified

with GWA mapping and QTL mapping is identical, this might suggest a possible role of these genes not only on a global population scale, but also between the accessions Col-0 and C24 (investigated by QTL mapping in this thesis). However, a comparison to other RIL populations, treated exactly in the same manner as for the GWA and QTL analysis in this thesis might clarify, whether true associations exist for *ARF7* and *ARF19*.

Among the top associated SNPs, four SNPs in close vicinity to the transcriptional regulator *JAZ1* were identified. *JAZ1* might play important roles in connecting the auxin and jasmonate phytohormone pathways (Vanstraelen and Benkova, 2012). A closer look at the CDS of *JAZ1* in the dense SNP data of Cao et al. (2011) revealed one major haplogroup and one outlier haplogroup. Here, the SNPs, which were detected by the GWA analysis after 2,4-D HGI, are the same SNPs, which separate the *A. thaliana* population into this two haplogroups. A functional validation, including more accessions of this haplogroup based on the 1001 genome project, with additional expression data, should clarify, whether a true association exist for *JAZ1* and the observed phenotypic variation upon 2,4-D treatment in dark-grown seedlings.

By applying a GWA mapping method, which in contrast to the above mentioned GWA mapping methods can identify variance controlling genes (Shen et al., 2012), *AUF2* was detected for the 2,4-D RGI treatment. Here, the phenotype-genotype correlation was significant according to a conservative significance threshold value. As calculated by the *v*GWAS mapping method, the identified SNP can explain $\sim 15\%$ of the variance differences between the investigated *A. thaliana* accessions. The identified SNP and SNPs in high linkage to the significant SNP (see Figure F3.28) do not fall directly into the annotated gene position of the *a priori* CG *AUF2*. A possible influence on phenotypic variation might be caused by expression level differences, which has to be investigated to verify, whether the identified association exists for *AUF2*. As outlined by Zheng et al. (2011), *AUF2* most likely plays a constitutive role in the context of root elongation. However, since in most studies the mean phenotypic differences and not the variance differences are investigated, the factors which contribute to variance differences might just be overlooked until now.

Like for the QTL analysis, the lack of experimental replicates is a crucial point in the GWA analysis. Here, the experiment for the synthetic auxin 2,4-D, with the promising CG *AUF2*, should be repeated to reduce the possibility of detecting false positive candidates. By repeating the experiment, additional *A. thaliana* accessions could be included, which might be chosen due to their habitat or their pre-calculated subpopulation membership to be able to investigate questions related to local adaption and the CG *AUF2*.

4.4 Concluding remarks

The aim of the thesis was to consider natural occurring variation upon auxin response traits in *A. thaliana* from various angles. In young seedlings the plant morphology, established during embryogenesis, is formed and alterations in this process might lead to selection constraints in interconnection with the surrounding environment. At this developmental stage of *A. thaliana* seedlings, the response to exogenous applied auxin derivatives was investigated by quantitative genetic approaches.

Genomic sequences of diverse *A. thaliana* accessions were used to infer selection signatures for auxin network genes by a population genetic analysis. This analysis revealed among 151 auxin network genes high variability for auxin metabolism genes. Consistent with these findings, the auxin metabolism genes also showed the highest TRCDIV for a set of six diverse *A. thaliana* accessions, based on a previously published gene expression data set (Delker et al., 2010). Auxin transport genes seem to be very conserved within *A. thaliana* and between the evaluated Brassicaceae.

In addition, the population genetic analysis and the transcript diversity analysis revealed, that the *AUX/IAA* gene family is the most variable gene family among the auxin signaling gene families.

Based on the findings of the auxin signaling gene families, the co-receptor *TIR1/AFB* gene family showed signatures of negative selection. Overall, for the auxin signaling, the nucleotide diversity patterns and TRCDIV values revealed the following hierarchy starting from the conserved gene family to the more relaxed gene family: *TIR1/AFBs* → *ARFs* → *IAs*. Here, a consideration of distinct subpopulations of *A. thaliana*, including climatic and geographic data, might be even more suitable to unravel signatures of local adaptation and to pinpoint the genes, which might cause phenotypic differences. However, the conducted population genetic analysis could give first hints for further studies on adaptational processes acting on auxin network genes.

The software RootDetection assisted the phenotypic data acquisition for QTL analyses and GWA analyses. Both quantitative genetic approaches revealed for the investigated populations, that the observed phenotypic variation seems to be regulated by a very complex genetic architecture. Small effect loci seem to dominate this complex architecture and additional quantitative approaches, like large scale expression analyses, might help to unravel these networks. However, the QTL with the strongest genotype-phenotype correlation could be validated in this thesis. Interestingly, this QTL also harbours three auxin metabolism genes, which show non-synonymous SNPs between the evaluated populations. Here, these findings support the population genetic analysis, that the auxin metabolism genes are good CGs to contribute to the observed phenotypic variation.

The molecular basis of phenotypic variation upon auxin response traits could not be yet revealed. The contributions to the observed phenotypic variation of some CGs and QTL regions could be unraveled by quantitative genetic approaches. Interesting CGs, like *AUF2*, *JAZ1* and *HAT2*, were identified by the GWA analyses and further experiments are needed to test for their individual contribution to the observed phenotypic variation upon auxin response traits.

References

- Abel, S. and Theologis, A. (2010). Odyssey of auxin. *Cold Spring Harbor perspectives in biology* 2, a004572.
- Alonso-Blanco, C., Aarts, M. G. M., Bentsink, L., Keurentjes, J. J. B., Reymond, M., Vreugdenhil, D. and Koornneef, M. (2009). What has natural variation taught us about plant development, physiology, and adaptation? *The Plant Cell* 21, 1877–1896.
- Alonso-Blanco, C. and Koornneef, M. (2000). Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. *Trends in Plant Science* 5, 22–29.
- Alonso-Blanco, C., Mendez-Vigo, B. and Koornneef, M. (2005). From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *The International Journal of Developmental Biology* 49, 717–732.
- Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410.
- Anastasio, A. E., Platt, A., Horton, M., Grotewold, E., Scholl, R., Borevitz, J. O., Nordborg, M. and Bergelson, J. (2011). Source verification of mis-identified *Arabidopsis thaliana* accessions. *The Plant Journal* 67, 554–566.
- Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., Lister, C., Molitor, J., Shindo, C., Tang, C., Toomajian, C., Traw, B., Zheng, H., Bergelson, J., Dean, C., Marjoram, P. and Nordborg, M. (2005). Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics* 1, e60.
- Armengaud, P., Zambaux, K., Hills, A., Sulpice, R., Pattison, R. J., Blatt, M. R. and Amtmann, A. (2009). EZ-Rhizo: integrated software for the fast and accurate measurement of root system architecture. *The Plant Journal* 57, 945–956.
- Assmann, S. M. (2012). Natural variation in abiotic stress and climate change responses in *Arabidopsis*. *International Journal of Plant Sciences* 174, 3–26.
- Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., Hu, T. T., Jiang, R., Muliyati, N. W., Zhang, X., Amer, M. A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., de Meaux, J., Ecker, J. R., Faure, N., Kniskern, J. M., Jones, J. D. G., Michael, T., Nemri, A., Roux, F., Salt, D. E., Tang, C., Todesco, M., Traw, M. B., Weigel, D., Marjoram, P., Borevitz, J. O., Bergelson, J. and Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631.
- Bakker, E. G., Toomajian, C., Kreitman, M. and Bergelson, J. (2006). A genome-wide survey of R gene polymorphisms in *Arabidopsis*. *The Plant Cell* 18, 1803–1818.
- Bakker, E. G., Traw, M. B., Toomajian, C., Kreitman, M. and Bergelson, J. (2008). Low levels of polymorphism in genes that control the activation of defense response in *Arabidopsis thaliana*. *Genetics* 178, 2031–2043.
- Balsubramanian, S., Schwartz, C., Singh, A., Warthmann, N., Kim, M. C., Maloof, J. N., Loudet, O., Trainer, G. T., Dabi, T. and Borevitz, J. O. (2009). QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. *PLoS One* 4, e4318.

- Bandelt, H.-J., Forster, P. and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* *16*, 37–48.
- Barbez, E., Kubeš, M., Rolčík, J., Béziat, C., Pěňčík, A., Wang, B., Rosquete, M. R., Zhu, J., Dobrev, P. I., Lee, Y., Zažímalová, E., Petrášek, J., Geisler, M., Friml, J. and Kleine-Vehn, J. (2012). A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants. *Nature* *485*, 119–122.
- Bartel, B. and Fink, G. R. (1995). ILR1, an amidohydrolase that releases active indole-3-acetic acid from conjugates. *Science* *268*, 1745–1748.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* *57*, 289–300.
- Benjamins, R. and Scheres, B. (2008). Auxin: the looping star in plant development. *Annual Review of Plant Biology* *59*, 443–465.
- Bentsink, L., Hanson, J., Hanhart, C. J., Blankestijn-de Vries, H., Coltrane, C., Keizer, P., El-Lithy, M., Alonso-Blanco, C., de Andrés, M. T., Reymond, M., van Eeuwijk, F., Smeekens, S. and Koornneef, M. (2010). Natural variation for seed dormancy in *Arabidopsis* is regulated by additive genetic and molecular pathways. *PNAS* *107*, 4264–4269.
- Bentsink, L., Jowett, J., Hanhart, C. J. and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *PNAS* *103*, 17042–17047.
- Bergelson, J., Kreitman, M., Stahl, E. a. and Tian, D. (2001). Evolutionary dynamics of plant R-genes. *Science* *292*, 2281–2285.
- Bergelson, J. and Roux, F. (2010). Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nature Reviews. Genetics* *11*, 867–879.
- Birchler, J. a. and Veitia, R. a. (2007). The gene balance hypothesis: from classical genetics to modern genomics. *The Plant Cell* *19*, 395–402.
- Birchler, J. A. and Veitia, R. A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytologist* *186*, 54–62.
- Boerjan, W., Cervera, M. T., Delarue, M., Beeckman, T., Dewitte, W., Bellini, C., Caboche, M., Van Onckelen, H., Van Montagu, M. and Inzé, D. (1995). Superroot, a recessive mutation in *Arabidopsis*, confers auxin overproduction. *The Plant Cell* *7*, 1405–1419.
- Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J. and Roux, F. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics* *6*, e1000940.
- Bradbury, P., Zhang, Z., Kron, D., Casstevens, D., Ramdoss, Y. and Buckler, E. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics (Oxford, England)* *23*, 2633–2635.
- Broman, K. W. and Sen, S. (2009). *A Guide to QTL Mapping with R/qtl*. Springer, New York.
- Broman, K. W., Wu, H., Sen, S. and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics (Oxford, England)* *19*, 889–890.
- Buckler, E., Casstevens, T., Bradbury, P., Zhang, Z., Ramdoss, Y., Oak, M. E., Holmberg, K. J., Stevens, N. and Zhang, Y. (2011). User manual for TASSEL - Trait Analysis by aSSociation, Evolution and Linkage.

- Burton, A. L., Williams, M., Lynch, J. P. and Brown, K. M. (2012). RootScan: software for high-throughput analysis of root anatomical traits. *Plant and Soil* 357, 189–203.
- Calderón, L. I. A. V., Lee, S., Oliveira, C. D., 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, 477–485.
- Campanoni, P. and Nick, P. (2005). Auxin-dependent cell division and cell elongation. 1-naphthaleneacetic acid and 2,4-dichlorophenoxyacetic acid activate different pathways. *Plant Physiology* 137, 939–948.
- Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., Wang, X., Ott, F., Müller, J., Alonso-Blanco, C., Borgwardt, K., Schmid, K. J. and Weigel, D. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics* 43, 956–963.
- Carraro, N., Tisdale-orr, T. E., Clouse, R. M., Knöller, A. S. and Spicer, R. (2012). Diversification and expression of the PIN, AUX/LAX, and ABCB families of putative auxin transporters in *Populus*. *Frontiers in Plant Science* 3, 10.3389/fpls.2012.00017.
- Chapman, E. J. and Estelle, M. (2009). Mechanism of auxin-regulated gene expression in plants. *Annual Review of Genetics* 43, 265–285.
- Chapman, E. J., Greenham, K., Castillejo, C., Sartor, R., Bialy, A., Sun, T.-P. and Estelle, M. (2012). Hypocotyl transcriptome reveals auxin regulation of growth-promoting genes through GA-dependent and -independent pathways. *PLoS One* 7, e36210.
- Cheng, Y., Dai, X. and Zhao, Y. (2007). Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis*. *The Plant Cell* 19, 2430–9.
- Churchill, G. A. and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.
- Clark, R., Schweikert, G., Toomajian, C. and Ossowski, S. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 338, 338–342.
- Clark, R. T., MacCurdy, R. B., Jung, J. K., Shaff, J. E., McCouch, S. R., Aneshansley, D. J. and Kochian, L. V. (2011). Three-dimensional root phenotyping with a novel imaging and software platform. *Plant Physiology* 156, 455–465.
- Darwin, C. and Darwin, F. (1880). *The power of movement in plants*. John Murray, London.
- Davarsi, A. and Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141, 1199–1207.
- De Smet, I., Lau, S., Mayer, U. and Jürgens, G. (2010). Embryogenesis - the humble beginnings of plant life. *The Plant Journal* 61, 959–970.
- De Smet, I., Voss, U., Lau, S., Wilson, M., Shao, N., Timme, R. E., Swarup, R., Kerr, I., Hodgman, C., Bock, R., Bennett, M., Jürgens, G. and Beeckman, T. (2011). Unraveling the evolution of auxin signaling. *Plant Physiology* 155, 209–221.
- Delbarre, A., Muller, P., Imhoff, V. and Guern, J. (1996). Comparison of mechanisms controlling uptake and accumulation of 2,4-dichlorophenoxy acetic acid, naphthalene-1-acetic acid, and indole-3-acetic acid in suspension-cultured tobacco cells. *Planta* 198, 532–541.

- 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*, 2184–2200.
- Delker, C., Raschke, A. and Quint, M. (2008). Auxin dynamics: the dazzling complexity of a small molecule’s message. *Planta* *227*, 929–41.
- Depaulis, F. and Veuille, M. (1998). Neutrality tests based on the distribution of haplotypes under an infinite-site model. *Molecular Biology and Evolution* *15*, 1788–1790.
- Depuydt, S. and Hardtke, C. S. (2011). Hormone signalling crosstalk in plant growth regulation. *Current Biology* *21*, R365–73.
- DeRose-Wilson, L. and Gaut, B. S. (2011). Mapping salinity tolerance during *Arabidopsis thaliana* germination and seedling growth. *PLoS One* *6*, e22832.
- Ding, Z., Wang, B., Moreno, I., Dupláková, N., Simon, S., Carraro, N., Reemmer, J., Pe, A., Chen, X., Tejos, R., Petrášek, J., Zaz, E., Honys, D., Rolc, J., X, C., R, T., P, S., S, P., J, M., J, P., E, Z., D, H., J, R., A, M., A, O., M, G. and J, F. (2012). ER-localized auxin transporter PIN8 regulates auxin homeostasis and male gametophyte development in *Arabidopsis*. *Nature Communications* *3*, 941.
- Ettingshausen, S. (2010). Populationsgenetische Analysen von Auxin-assoziierten Genen in *Arabidopsis thaliana*.
- Filiault, D. L. and Maloof, J. N. (2012). A genome-wide association study identifies variants underlying the *Arabidopsis thaliana* shade avoidance response. *PLoS Genetics* *8*, e1002589.
- Force, A., Lynch, M., Pickett, F., Amores, A., Yan, Y. and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* *151*, 1531–1545.
- Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J. and Wilczek, A. M. (2011). A map of local adaptation in *Arabidopsis thaliana*. *Science* *334*, 86–89.
- Freeling, M. and Thomas, B. C. (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Research* *16*, 805–814.
- French, A., Ubeda-Tomás, S., Holman, T. J., Bennett, M. J. and Pridmore, T. (2009). High-throughput quantification of root growth using a novel image-analysis tool. *Plant Physiology* *150*, 1784–1795.
- Friml, J., Benková, E., Blilou, I., Wisniewska, J., Hamann, T., Ljung, K., Woody, S., Sandberg, G., Scheres, B., Jürgens, G. and Palme, K. (2002). AtPIN4 mediates sink-driven auxin gradients and root patterning in *Arabidopsis*. *Cell* *108*, 661–673.
- Fu, Y.-X. and Li, W.-H. (1993). Statistical tests of neutrality of mutations. *Genetics* *133*, 693–709.
- Fukaki, H., Nakao, Y., Okushima, Y., Theologis, A. and Tasaka, M. (2005). Tissue-specific expression of stabilized SOLITARY-ROOT/IAA14 alters lateral root development in *Arabidopsis*. *The Plant Journal* *44*, 382–395.
- Fukaki, H. and Tasaka, M. (2009). Hormone interactions during lateral root formation. *Plant Molecular Biology* *69*, 437–449.
- Gan, X., Stegle, O., Behr, J., Steffen, J. G., Drewe, P., Hildebrand, K. L., Lyngsoe, R., Schultheiss, S. J., Osborne, E. J., Sreedharan, V. T., Kahles, A., Bohnert, R., Jean, G., Derwent, P., Kersey, P., Belfield, E. J., Harberd, N. P., Kemen, E., Toomajian, C., Kover, P. X., Clark, R. M., Rättsch, G. and Mott, R. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* *477*, 419–423.

- Gillespie, J. H. (1991). *The Causes of Molecular Evolution*. Oxford University Press, Oxford.
- Goren, R. and Bukovac, M. J. (1973). Mechanism of naphthaleneacetic acid conjugation. *Plant Physiology* *51*, 907–913.
- Granier, C., Aguirrezabal, L., Chenu, K., Cookson, S. J., Dauzat, M., Hamard, P., Thioux, J.-j., Rolland, G., Bouchier-combaud, S., Lebaudy, A., Muller, B., Simonneau, T. and Tardieu, F. (2006). PHENOPSIS , an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist* *169*, 623–635.
- Grunewald, W., Vanholme, B., Pauwels, L., Plovie, E., Inze, D., Gheysen, G. and Goossens, A. (2009). Expression of the *Arabidopsis* jasmonate signalling repressor JAZ1/TIFY10A is stimulated by auxin. *EMBO Reports* *10*, 923–928.
- Guilfoyle, T. J. and Hagen, G. (2007). Auxin response factors. *Current Opinion in Plant Biology* *10*, 453–60.
- Günther, T. and Schmid, K. J. (2010). Deleterious amino acid polymorphisms in *Arabidopsis thaliana* and rice. *Theoretical and Applied Genetics* *121*, 157–68.
- Haley, C. S. and Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* *69*, 315–324.
- 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*, 95–98.
- Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., Toomajian, C., Roux, F. and Bergelson, J. (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* *334*, 83–86.
- Hartl, D. L. and Clark, A. G. (2007). *Principles of population genetics*. 4th edition, Sinauer Associates, Inc. Publishers, Sunderland, MA 01375 U.S.A.
- Hartmann, A., Czauderna, T., Hoffmann, R., Stein, N. and Schreiber, F. (2011). HTPPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics* *12*, 148.
- Hiller, M., Huse, K., Szafranski, K., Rosenstiel, P., Schreiber, S., Backofen, R. and Platzner, M. (2006). Phylogenetically widespread alternative splicing at unusual GYNGYN donors. *Genome Biology* *7*, 1–16.
- Hoffmann, M. H. (2002). Biogeography of *Arabidopsis thaliana* (L .) Heynh. (Brassicaceae). *Journal of Biogeography* *29*, 125–134.
- Horton, M. W., Hancock, A. M., Huang, Y. S., Toomajian, C., Atwell, S., Auton, A., Mulyati, N. W., Platt, A., Sperone, F. G., Vilhjálmsson, B. J., Nordborg, M., Borevitz, J. O. and Bergelson, J. (2012). Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics* *44*, 212–216.
- Huang, X., Paulo, M.-j. a., Boer, M., Effgen, S., Keizer, P. and Koornneef, M. (2011). Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. *PNAS* *108*, 4488–4493.
- Hudson, R. and Kaplan, N. L. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* *111*, 147–164.
- Huson, D. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* *23*, 254–267.

- Isard, M. and Blake, A. (1998). CONDENSATION - conditional density propagation for visual tracking. *International Journal of Computer Vision* 29, 5–28.
- Janitza, P., Ullrich, K. K. and Quint, M. (2012). Toward a comprehensive phylogenetic reconstruction of the evolutionary history of mitogen-activated protein kinases in the plant kingdom. *Frontiers in Plant Science* 3, 271.
- Jiménez-Gómez, J. M., Wallace, A. D. and Maloof, J. N. (2010). Network analysis identifies ELF3 as a QTL for the shade avoidance response in *Arabidopsis*. *PLoS Genetics* 6.
- Kang, H. M., Zaitlen, N. a., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J. and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723.
- Karaletsos, T., Stegle, O., Dreyer, C., Winn, J. and Borgwardt, K. M. (2012). ShapePheno: unsupervised extraction of shape phenotypes from biological image collections. *Bioinformatics (Oxford, England)* 28, 1001–1008.
- Kent, W. J. (2002). BLAT - the BLAST-like alignment tool. *Genome Research* 12, 656–664.
- Kepinski, S. and Leyser, O. (2005). The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature* 435, 446–451.
- Keurentjes, J. J. B., Bentsink, L., Alonso-Blanco, C., Hanhart, C. J., Blankestijn-De Vries, H., Effgen, S., Vreugdenhil, D. and Koornneef, M. (2007). Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* 175, 891–905.
- Kim, Y. (2006). Allele frequency distribution under recurrent selective sweeps. *Genetics* 1978, 1967–1978.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* 217, 624–626.
- Kimura, M. (1983). *The neutral theory of Molecular Evolution*. Cambridge University Press, Cambridge, Massachusetts, USA.
- King, J. J., Stimart, D. P., Fisher, R. H. and Bleecker, a. B. (1995). A mutation altering auxin homeostasis and plant morphology in *Arabidopsis*. *The Plant Cell* 7, 2023–2037.
- Koornneef, M., Alonso-Blanco, C. and Vreugdenhil, D. (2004). Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Biology* 55, 141–172.
- Korte, A., Vilhjálmsson, B. J., Segura, V., Platt, A., Long, Q. and Nordborg, M. (2012). A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nature Genetics* 44, 1066–1071.
- Kosambi, D. D. (1943). The estimation of map distance from recombination values. *Annals of Eugenics* 12, 172–175.
- Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., Durrant, C. and Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics* 5, e1000551.
- Kowalski, S. P., Lan, T. H., Feldmann, K. A. and Paterson, A. H. (1994). QTL mapping of naturally-occurring variation in flowering time of *Arabidopsis thaliana*. *Molecular Genetics* 245, 548–555.
- Kruglyak, L. and Lander, E. S. (1995). A nonparametric approach for mapping quantitative trait loci. *Genetics* 139, 1421–1428.

- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J. and Marra, M. a. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research* *19*, 1639–1645.
- Lapin, D., Meyer, R. C., Takahashi, H., Bechtold, U. and Van den Ackerveken, G. (2012). Broad-spectrum resistance of Arabidopsis C24 to downy mildew is mediated by different combinations of isolate-specific loci. *New Phytologist* *196*, 1171–1181.
- Lee, Y. P., Babakov, A., de Boer, B., Zuther, E. and Hinch, D. K. (2012). Comparison of freezing tolerance, compatible solutes and polyamines in geographically diverse collections of *Thellungiella* sp. and *Arabidopsis thaliana* accessions. *BMC Plant Biology* *12*, 131.
- Li, J., Li, H., Jakobsson, M., Li, S., Sjödin, P. and Lascoux, M. (2012). Joint analysis of demography and selection in population genetics: Where do we stand and where could we go? *Molecular Ecology* *21*, 28–44.
- Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)* *25*, 1451–2.
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S. and Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics (Oxford, England)* *28*, 2397–2399.
- Livingstone, K. and Anderson, S. (2009). Patterns of variation in the evolution of carotenoid biosynthetic pathway enzymes of higher plants. *The Journal of Heredity* *100*, 754–61.
- Lobet, G., Pagès, L. and Draye, X. (2011). A novel image-analysis toolbox enabling quantitative analysis of root system architecture. *Plant Physiology* *157*, 29–39.
- Long, J. a. and Barton, M. K. (1998). The development of apical embryonic pattern in *Arabidopsis*. *Development (Cambridge, England)* *125*, 3027–35.
- Ludwig-Müller, J. (2011). Auxin conjugates: their role for plant development and in the evolution of land plants. *Journal of Experimental Botany* *62*, 1757–1773.
- Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. and Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. *PNAS* *102*, 5454–9.
- Magidin, M., Pittman, J. K., Hirschi, K. D. and Bartel, B. (2003). ILR2, a novel gene regulating IAA conjugate sensitivity and metal transport in *Arabidopsis thaliana*. *The Plant Journal* *35*, 523–534.
- Manichaikul, A., Moon, J. Y., Sen, S., Yandell, B. S. and Broman, K. W. (2009). A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics* *181*, 1077–1086.
- Mano, Y. and Nemoto, K. (2012). The pathway of auxin biosynthesis in plants. *Journal of Experimental Botany* *63*, 2853–2872.
- Martinez, O. and Curnow, R. (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics* *85*, 480–488.
- Mashiguchi, K., Tanaka, K. and Sakai, T. (2011). The main auxin biosynthesis pathway in *Arabidopsis*. *PNAS* *108*, 18512–18517.
- McDonald, J. H. and Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* *351*, 652–654.

- Meyer, R. C., Kusterer, B., Lisec, J., Steinfath, M., Becher, M., Scharr, H., Melchinger, A. E., Selbig, J., Schurr, U., Willmitzer, L. and Altmann, T. (2010). QTL analysis of early stage heterosis for biomass in *Arabidopsis*. *Theoretical and Applied Genetics* *120*, 227–237.
- Muday, G. K., Rahman, A. and Binder, B. M. (2012). Auxin and ethylene: collaborators or competitors? *Trends in Plant Science* *17*, 181–195.
- Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, New York, USA.
- Nei, M. and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York, New York, USA.
- 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*, e196.
- Ohno, S. (1970). *Evolution by gene duplication*. Springer-Verlag, New York, New York, USA.
- Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology & Systematics* *23*, 263–286.
- Okushima, Y., Mitina, I., Quach, H. L., Theologis, A., Gene, P. and Street, B. (2005a). AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. *The Plant Journal* *43*, 29–46.
- 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. and Smith, A. (2005b). 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*, 444–463.
- Page, R. D. M. and Holmes, E. C. (1998). *Molecular Evolution: A Phylogenetic Approach*. Blackwell Science Ltd, Malden, MA, U.S.A.
- Paponov, I. a., Teale, W., Lang, D., Paponov, M., Reski, R., Rensing, S. a. and Palme, K. (2009). The evolution of nuclear auxin signalling. *BMC Evolutionary Biology* *9*, 126.
- Patterson, N., Price, A. L. and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics* *2*, e190.
- Pauwels, L. and Goossens, A. (2011). The JAZ proteins: a crucial interface in the jasmonate signaling cascade. *The Plant Cell* *23*, 3089–3100.
- Pavlidis, P., Huter, S. and Stephan, W. (2008). A population genomic approach to map recent positive selection in model species. *Molecular Ecology* *17*, 3585–3598.
- Pavlidis, P., Jensen, J. D. and Stephan, W. (2010). Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics* *185*, 907–922.
- Platt, A., Horton, M., Huang, Y. S., Li, Y., Anastasio, A. E., Mulyati, N. W., Bossdorf, O., Byers, D., Donohue, K., Dunning, M., Holub, E. B. and Hudson, A. (2010). The scale of population structure in *Arabidopsis thaliana*. *PLoS genetics* *6*, e10000843.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. a. and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* *38*, 904–909.

- Puerma, E. and Aguadé, M. (2013). Polymorphism at genes involved in salt tolerance in *Arabidopsis thaliana* (Brassicaceae). *American Journal of Botany* *100*, 384–390.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. a. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. and Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* *81*, 559–575.
- Quint, M., Drost, H.-G., Gabel, A., Ullrich, K. K., Bönn, M. and Grosse, I. (2012). A transcriptomic hourglass in plant embryogenesis. *Nature* *490*, 98–101.
- Quint, M. and Gray, W. M. (2006). Auxin signaling. *Current Opinion in Plant Biology* *9*, 448–453.
- Rademacher, E. H., Möller, B., Lokerse, A. S., Llavata-Peris, I. C., van den Berg, W. and Weijers, D. (2011). A cellular expression map of the *Arabidopsis* AUXIN RESPONSE FACTOR gene family. *Plant Journal* *68*, 597–606.
- Rahman, A., Bannigan, A., Sulaman, W., Pechter, P., Blancaflor, E. B. and Baskin, T. I. (2007). Auxin, actin and growth of the *Arabidopsis thaliana* primary root. *The Plant Journal* *50*, 514–528.
- Ramos-Onsins, S. E., Puerma, E., Balañá-Alcaide, D. ., Salguero, D. . and Aguadé, M. (2008). Multilocus analysis of variation using a large empirical data set: phenylpropanoid pathway genes in *Arabidopsis thaliana*. *Molecular Ecology* *17*, 1211–1223.
- Rampey, R. a., Baldrige, M. T., Farrow, D. C., Bay, S. N. and Bartel, B. (2013). Compensatory mutations in predicted metal transporters modulate auxin conjugate responsiveness in *Arabidopsis*. *G3 (Bethesda, Md.)* *3*, 131–41.
- Rampey, R. A., LeClere, S., Kowalczyk, M., Ljung, K., Sandberg, G. and Bonnie Bartel (2004). A family of auxin-conjugate hydrolases that contributes to free indole-3-acetic acid levels during *Arabidopsis* germination. *Plant Physiology* *135*, 978–988.
- Ramsay, H., Rieseberg, L. H. and Ritland, K. (2009). The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. *Molecular Biology and Evolution* *26*, 1045–53.
- Rausher, M. D., Lu, Y. and Meyer, K. (2008). Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *Journal of Molecular Evolution* *67*, 137–44.
- Remington, D. L., Vision, T. J., Guilfoyle, T. J. and Reed, J. W. (2004). Contrasting modes of diversification in the Aux/IAA and ARF gene families. *Plant Physiology* *135*, 1738–1752.
- Rodgers-Melnick, E., Mane, S. P., Dharmawardhana, P., Slavov, G. T., Crasta, O. R., Strauss, S. H., Brunner, A. M., Difazio, S. P., Virginia, W. and Bioinformatics, V. (2012). Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. *Genome Research* *22*, 95–105.
- Salvi, S. and Tuberosa, R. (2005). To clone or not to clone plant QTLs: present and future challenges. *Trends in Plant Science* *10*, 297–304.
- Sawa, S., Ohgishi, M., Goda, H., Higuchi, K., Shimada, Y. and Yoshida, S. (2002). The HAT2 gene, a member of the HD-Zip gene family, isolated as an auxin inducible gene by DNA microarray screening, affects auxin response in *Arabidopsis*. *Plant Journal* *32*, 1011–1022.
- Schmid, K. J., Ramos-Onsins, S., Ringys-Beckstein, H., Weisshaar, B. and Mitchell-Olds, T. (2005). A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism. *Genetics* *169*, 1601–1615.

- Schmid, K. J., Törjék, O., Meyer, R., Schmuths, H., Hoffmann, M. H. and Altmann, T. (2006). Evidence for a large-scale population structure of *Arabidopsis thaliana* from genome-wide single nucleotide polymorphism markers. *Theoretical and Applied Genetics* *112*, 1104–14.
- Schumann, N., Navarro-Quezada, A., Ullrich, K., Kuhl, C. and Quint, M. (2011). Molecular evolution and selection patterns of plant F-box proteins with C-terminal kelch repeats. *Plant Physiology* *155*, 835–50.
- Sharbel, T., Haubold, B. and T. Mitchell-Olds (2000). Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology* *9*, 2109–2118.
- Shen, X., Pettersson, M., Rönnegård, L. and Carlborg, O. (2012). Inheritance beyond plain heritability: variance-controlling genes in *Arabidopsis thaliana*. *PLoS Genetics* *8*, e1002839.
- Staswick, P. E., Serban, B., Rowe, M. and Tiryaki, I. (2005). Characterization of an *Arabidopsis* enzyme family that conjugates amino acids to indole-3-acetic acid. *The Plant Cell* *17*, 616–627.
- Stepanova, A. N., Yun, J., Robles, L. M., Novak, O., He, W., Guo, H., Ljung, K. and Alonso, J. M. (2011). The *Arabidopsis* YUCCA1 flavin monooxygenase functions in the indole-3-pyruvic acid branch of auxin biosynthesis. *The Plant Cell* *23*, 3961–3973.
- Sterken, R., Kiekens, R., Coppens, E., Vercauteren, I., Zabeau, M., Inzé, D., Flowers, J. and Vuylsteke, M. (2009). A population genomics study of the *Arabidopsis* core cell cycle genes shows the signature of natural selection. *The Plant Cell* *21*, 2987–98.
- Sun, J., Xu, Y., Ye, S., Jiang, H., Chen, Q., Liu, F., Zhou, W., Chen, R., Li, X., Tietz, O., Wu, X., Cohen, J. D., Palme, K. and Li, C. (2009). *Arabidopsis* ASA1 is important for jasmonate-mediated regulation of auxin biosynthesis and transport during lateral root formation. *The Plant Cell* *21*, 1495–511.
- Szemenyei, H., Hannon, M. and Long, J. A. (2008). TOPLESS mediates auxin dependent transcriptional repression during *Arabidopsis* embryogenesis. *Science* *319*, 1384–1386.
- Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* *105*, 437–460.
- Tajima, F. (1989a). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* *123*, 585–95.
- Tajima, F. (1989b). The effect of change in population size on DNA polymorphism. *Genetics* *123*, 579–601.
- Tan, X., Calderon-villalobos, L. I. A., Sharon, M., Zheng, C., Robinson, C. V. and Estelle, M. (2007). Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* *446*, 640–645.
- Teale, W. D., Ditengou, F. a., Dovzhenko, a. D., Li, X., Molendijk, a. M., Ruperti, B., Paponov, I. and Palme, K. (2008). Auxin as a model for the integration of hormonal signal processing and transduction. *Molecular Plant* *1*, 229–37.
- 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*, 847–859.
- Thimann, K. V. (1938). Hormones and the analysis of growth. *Plant Physiology* *13*, 437–449.
- Thomine, S., Lelièvre, F., Boufflet, M., Guern, J. and Barbier-Brygoo, H. (1997). Anion-channel blockers interfere with auxin responses in dark-grown *Arabidopsis* hypocotyls. *Plant Physiology* *115*, 533–542.

- Thornton, K. (2003). libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* (Oxford, England) *19*, 2325–2327.
- Thornton, K. R., Jensen, J. D., Bequet, C. and Andolfatto, P. (2007). Progress and prospects in mapping recent selection in the genome. *Heredity* *98*, 340–348.
- Törjék, O., Meyer, R. C., Zehnsdorf, M., Teltow, M., Strompen, G., Witucka-Wall, H., Blacha, A. and Altmann, T. (2008). Construction and analysis of 2 reciprocal *Arabidopsis* introgression line populations. *The Journal of Heredity* *99*, 396–406.
- Törjék, O., Witucka-Wall, H., Meyer, R. C., von Korff, M., Kusterer, B., Rautengarten, C. and Altmann, T. (2006). Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theoretical and Applied Genetics* *113*, 1551–1561.
- van Berloo, R. (2008). GGT 2.0: versatile software for visualization and analysis of genetic data. *The Journal of Heredity* *99*, 232–236.
- Van de Peer, Y., Maere, S. and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nature Reviews. Genetics* *10*, 725–732.
- Vanneste, S. and Friml, J. (2009). Auxin: a trigger for change in plant development. *Cell* *136*, 1005–1016.
- Vanstraelen, M. and Benkova, E. (2012). Hormonal Interactions in the Regulation of Plant Development. *Annual Review of Cell and Developmental Biology* *28*, 463–487.
- Vaughn, M. W., Tanurdzić, M., Lippman, Z., Jiang, H., Carrasquillo, R., Rabinowicz, P. D., Dedhia, N., McCombie, W. R., Agier, N., Bulski, A., Colot, V., Doerge, R. W. and Martienssen, R. a. (2007). Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biology* *5*, e174.
- Viane, T., Delwiche, C., Rensing, S. and Friml, J. (2012). Origin and evolution of PIN auxin transporters in the green lineage. *Trends in Plant Science* *18*, 5–10.
- Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review* *22*, 88–93.
- Wang, B.-b. and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *PNAS* *103*, 7175–7180.
- Wang, J.-w., Wang, L.-j., Mao, Y.-b., Cai, W.-j., Xue, H.-w. and Chen, X.-y. (2005). Control of root cap formation by MicroRNA-targeted auxin response factors in *Arabidopsis*. *The Plant Cell* *17*, 2204–2216.
- Watterson, G. (1975). Number of segregating sites in genetic models without recombination. *Theoretical Population Biology* *7*, 256–276.
- Weigel, D. (2012). Natural variation in *Arabidopsis* : from molecular genetics to ecological genomics. *Plant Physiology* *158*, 2–22.
- Weigel, D. and Mott, R. (2009). The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biology* *10*, 107.
- Weigel, D. and Nordborg, M. (2005). Natural variation in *Arabidopsis*. How do we find the causal Genes? *Plant Physiology* *138*, 567–568.
- Weijers, D. and Jürgens, G. (2005). Auxin and embryo axis formation: the ends in sight? *Current Opinion in Plant Biology* *8*, 32–37.

- Went, F. and Thimann, K. (1937). *Phytohormones*. The Macmillan Company., New York.
- Wolyn, D. J., Borevitz, J. O., Loudet, O., Schwartz, C., Maloof, J., Ecker, J. R., Berry, C. C. and Chory, J. (2004). Light-response quantitative trait loci identified with composite interval and eXtreme array mapping in *Arabidopsis thaliana*. *Genetics* *167*, 907–917.
- Won, C., Shen, X., Mashiguchi, K., Zheng, Z., Dai, X., Cheng, Y., Kasahara, H., Kamiya, Y., Chory, J. and Zhao, Y. (2011). Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in *Arabidopsis*. *PNAS* *108*, 18518–18523.
- Wright, S. I. and Gaut, B. S. (2005). Molecular population genetics and the search for adaptive evolution in plants. *Molecular Biology and Evolution* *22*, 506–519.
- Wu, T., HT, Z., Y, W., WS, J., XF, X., XZ, Z. and ZH, H. (2012). Induction of root Fe(III) reductase activity and proton extrusion by iron deficiency is mediated by auxin-based systemic signalling in *Malus xiaojinensis*. *Journal of Experimental Botany* *63*, 859–870.
- Yamamoto, M. and Yamamoto, K. (1998). Differential effects of 1-naphthaleneacetic acid, indole-3-acetic acid and 2,4-dichlorophenoxyacetic acid on the gravitropic response of roots in an auxin-resistant mutant of *Arabidopsis*, *aux1*. *Plant Cell Physiology* *39*, 660–664.
- Yu, J., Pressoir, G., Briggs, W., Vroh Bi, I., Yamasaki, M., Doebley, J., McMullen, M., Gaut, B., Nielsen, D., Holland, J., Kresovich, S. and Buckler, E. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* *38*, 203–208.
- Zeng, Z. (1994). Precision mapping of quantitative trait loci. *Genetics* *136*, 1457–1468.
- Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. a., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M. and Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* *42*, 355–360.
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C. and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics (Oxford, England)* *28*, 3326–3328.
- Zheng, X., Miller, N. D., Lewis, D. R., Christians, M. J., Lee, K.-H., Muday, G. K., Spalding, E. P. and Vierstra, R. D. (2011). AUXIN UP-REGULATED F-BOX PROTEIN 1 regulates the crosstalk between auxin transport and cytokinin signaling during plant root growth. *Plant Physiology* *156*, 1878–1893.
- Zhu, J., Ingram, P. a., Benfey, P. N. and Elich, T. (2011). From lab to field, new approaches to phenotyping root system architecture. *Current Opinion in Plant Biology* *14*, 310–317.
- Zuther, E., Schulz, E., Childs, L. H. and Hinch, D. K. (2012). Clinal variation in the non-acclimated and cold-acclimated freezing tolerance of *Arabidopsis thaliana* accessions. *Plant, Cell & Environment* *35*, 1860–1878.

Appendix

List of Figures

F1.1	Overview of auxin biology.	2
F1.2	Molecular sources of phenotypic variation.	4
F1.3	Distribution of 1307 genotyped <i>A. thaliana</i> accessions.	7
F3.1	Overview of the population genetic analysis workflow.	30
F3.2	Nucleotide diversity comparing the empirical null distribution with combined auxin network groups as one group.	34
F3.3	Nucleotide diversity of auxin network gene groups considering coding sequences.	35
F3.4	Ratio plot of π_{NONSYN} and π_{SYN} for auxin network gene groups.	36
F3.5	Nucleotide variation patterns of auxin network gene groups considering coding sequences.	38
F3.6	Minor allele frequency distribution of auxin network genes grouped as one group.	39
F3.7	Minor allele frequency distribution of auxin network groups presented as boxplots.	40
F3.8	Nucleotide diversity of auxin signaling gene families considering coding sequences.	41
F3.9	Nucleotide variation patterns of auxin signaling gene families considering coding sequences.	42
F3.10	Transcript diversity among six <i>A. thaliana</i> accessions comparing auxin network gene groups to the empirical null distribution.	47
F3.11	Screenshots from RootDetection.	50
F3.12	Comparing mean root length values of automatically measured raw data with manually curated data.	52
F3.13	Root growth inhibition (RGI) on different auxin concentrations in the core subsets of the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations and the parental lines Col-0 and C24.	53
F3.14	Hypocotyl growth inhibition (HGI) on different auxin concentrations in the core subsets of the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations and the parental lines Col-0 and C24.	54
F3.15	Percent frequency distribution of RGI for the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations after different auxin treatments.	56
F3.16	Percent frequency distribution of HGI for the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations after different auxin treatments.	57
F3.17	QTL associated with RGI after NAA treatment in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL population.	59
F3.18	Example of significant <i>two-QTL genome scan</i> interactions in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations for RGI.	60
F3.19	QTL loci determined by multiple QTL mapping for auxin response traits in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations.	62
F3.20	Combinatorial QTL effects result either in decreased or increased HGI after 2,4-D treatment.	66
F3.21	Validation of the QTLs corresponding to the NAA response trait on RGI.	70
F3.22	Dose response analysis in selected NILs to validate QTL rNAA-V.2 corresponding to the NAA response trait on RGI.	70
F3.23	Auxin dependent root and hypocotyl growth inhibition pattern with pairwise genotype distance.	75
F3.24	Correlation analysis between pairwise genome-wide relationships and pairwise geographic distance for a subset of 67 <i>A. thaliana</i> accessions.	76
F3.25	Correlation analysis between pairwise phenotypic distances and pairwise geographic distance for a subset of 67 <i>A. thaliana</i> accessions.	77

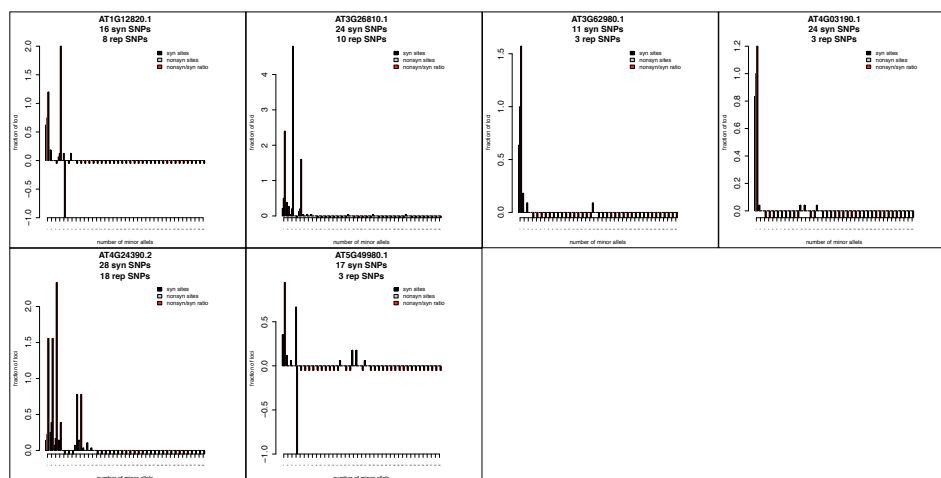
F3.26	Exemplary quantile-quantile plot (QQ-plot) to investigate appropriate GWA methods for RGI after 20 nM 2,4-D treatment.	79
F3.27	Manhattan plot of the <i>v</i> GWAS result for RGI 2,4-D.	80
F3.28	Detailed SNP structure analysis of the <i>a priori</i> candidate genes <i>JAZ1</i> and <i>AUF2</i>	83
F3.29	Phenotypic variance and phenotypic mean differences in 80 <i>A. thaliana</i> accessions.	85
S1	Individual MAF distribution plot for the auxin network genes.	115
S2	Individual MAF distribution plot for the auxin network genes.	116
S3	Individual MAF distribution plot for the auxin network genes.	117
S4	Individual MAF distribution plot for the auxin network genes.	118
S5	Individual MAF distribution plot for the auxin network genes.	119
S6	Individual MAF distribution plot for the auxin network genes.	120
S7	Nucleotide diversity of auxin network gene groups considering coding sequences for the re-sequenced data.	121
S8	Ratio plot of π_{NONSYN} and π_{SYN} for auxin network gene groups for the re-sequenced data.	122
S9	Nucleotide variation patterns of auxin network gene groups considering coding sequences for the re-sequenced data.	122
S10	Nucleotide diversity of auxin signaling gene families considering coding sequences for the re-sequenced data.	123
S11	Nucleotide variation patterns of auxin signaling gene families considering coding sequences for the re-sequenced data.	124
S12	Minor allele frequency distribution of auxin signaling families presented as boxplots.	124
S13	Chromosome-wide transcript diversity among six <i>A. thaliana</i> accessions compared to genome-wide nucleotide diversity among 80 <i>A. thaliana</i> accessions.	125
S14	Exemplary workflow measuring seedling root length with RootDetection.	126
S15	Screenshots from RootDetection's post processing mode.	127
S16	Biplots of 6 auxin response traits analyzed in two RIL populations.	128
S17	<i>Main effect QTLs</i> associated with RGI after different auxin treatments.	129
S18	<i>Main effect QTLs</i> associated with HGI after different auxin treatments.	130
S19	Example of significant <i>two-QTL genome scan</i> interactions in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations for HGI.	131
S20	Supporting information for the validation of a QTL corresponding to the NAA response trait on RGI.	132
S21	An attempt to validate the QTL corresponding to the NAA response trait on HGI.	133
S22	Comparing the first 3 principle components of the P-matrix calculated based on 165984 SNPs in 1307 <i>A. thaliana</i> accessions.	134
S23	GWA results overview for RGI after 20nM 2,4-D treatment.	136
S24	GWA results overview for RGI after 40nM IAA treatment.	137
S25	GWA results overview for RGI after 75nM NAA treatment.	138
S26	GWA results overview for HGI after 375nM 2,4-D treatment.	139
S27	GWA results overview for HGI after 500nM IAA treatment.	140
S28	GWA results overview for HGI after 500nM NAA treatment.	141
S29	Haplogroup representation of the <i>a priori</i> candidate gene <i>JAZ1</i>	142
S30	Detailed SNP structure analysis of the <i>a priori</i> candidate gene <i>ARF7</i>	143

List of Tables

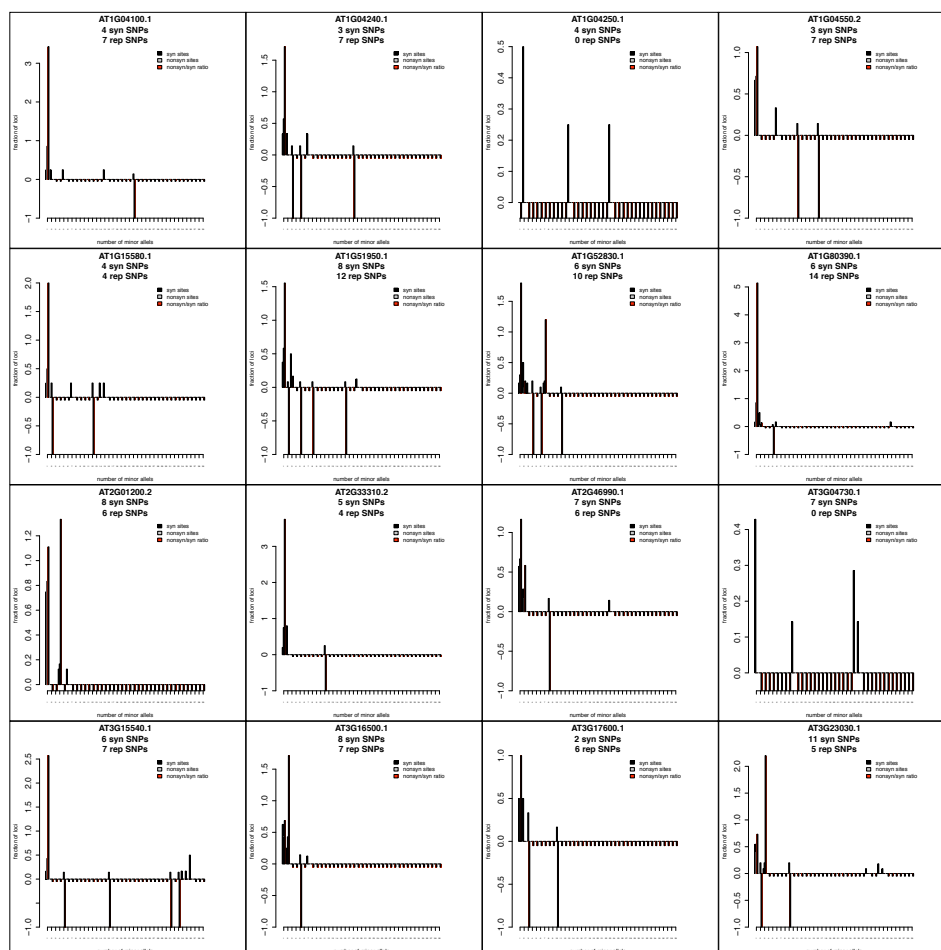
T3.1	Nucleotide divergence between <i>A. thaliana</i> and three other Brassicaceae.	44
T3.2	Extreme auxin network genes in the tails of the empirical null distribution.	46
T3.3	Descriptive statistics of auxin response traits in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations.	55
T3.4	Phenotypic correlations among auxin response traits in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations.	58
T3.5	Summary of QTL detected by multiple QTL mapping of auxin response traits in the RIL $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ populations.	64
T3.6	Descriptive statistics of auxin response traits among 80 <i>A. thaliana</i> accessions.	73
T3.7	Phenotypic correlations among auxin response traits among 80 <i>A. thaliana</i> accessions.	78
T3.8	List of <i>a priori</i> candidate genes associated with auxin found among the highest correlated SNPs with MLM and WILCOXON GWA mapping methods.	81
T3.9	List of <i>a priori</i> candidate genes associated with auxin found among the highest correlated SNPs with vGWAS GWA mapping method.	84
ST1	Primer sequences used for genomic fragment sequencing.	144
ST2	Distribution of TAIR10 annotated splice donor and acceptor sites.	145
ST3	Major single nucleotide polymorphisms in the re-sequenced data for auxin network genes.	146
ST4	Major single nucleotide polymorphisms in the MPICao2010 data for auxin network genes.	147
ST5	Nucleotide diversity among different nucleotide categories within the re-sequenced data.	148
ST6	Summary statistics of within species population genetic parameters in auxin network gene groups within the MPICao2010 data.	149
ST7	Summary statistics of within species population genetic parameters in auxin network gene groups within the re-sequenced data.	150
ST8	McDonald-Kreitman test between <i>A. thaliana</i> and three Brassicaceae species.	151
ST9	Exemplary root length data for NAA treatment in 80 <i>A. thaliana</i> accessions measured with RootDetection.	152
ST10	Main effect QTL analysis in the RIL $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ populations for auxin response traits.	153
ST11	Two-QTL genome scan analysis in the RIL $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ populations for auxin response traits.	154
ST12	Possible SNP effects in <i>a priori</i> auxin related candidate genes within markers III_17 and III_19 on chromosome three.	155
ST13	Possible SNP effects in <i>a priori</i> auxin related candidate genes within markers V_20 and V_22 on chromosome five.	156
ST14	Possible SNP effects in <i>a priori</i> auxin related candidate genes within markers II_10 and II_12 on chromosome two.	157
ST15	Possible SNP effects in <i>a priori</i> auxin related candidate genes from marker V_24 to the end of chromosome five.	158
ST16	SNP enrichment analysis of <i>a priori</i> candidate genes related to auxin detected under different top ranked associated SNP categories with MLM and WILCOXON GWA mapping method.	159
ST17	Possible SNP effects in candidate genes detected by GWA mapping.	160

Supplementary figures

TIR1/AFBs

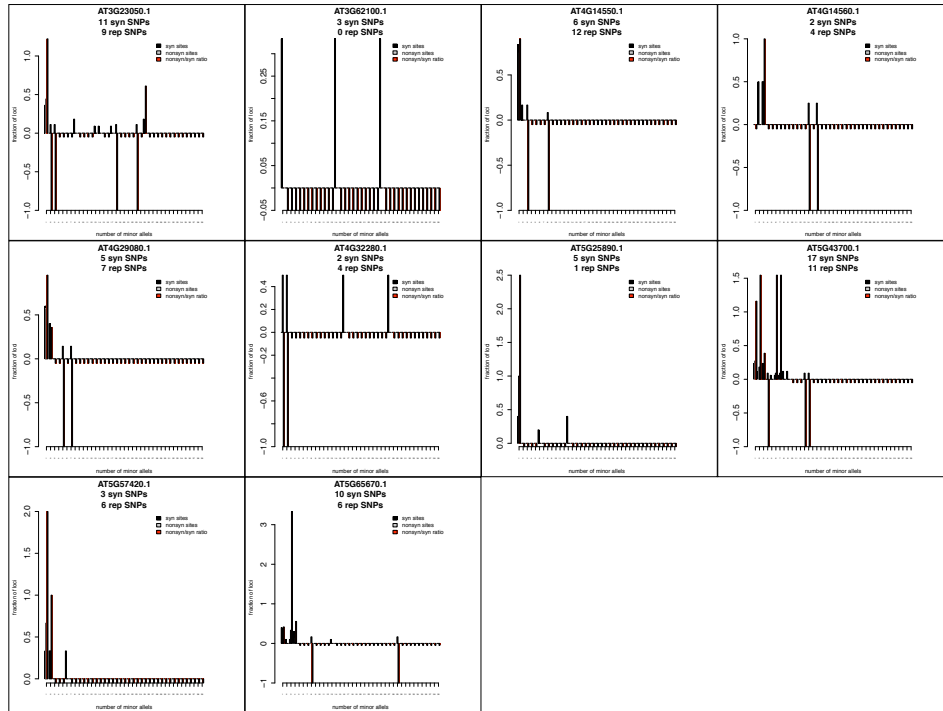


AUX/IAAs

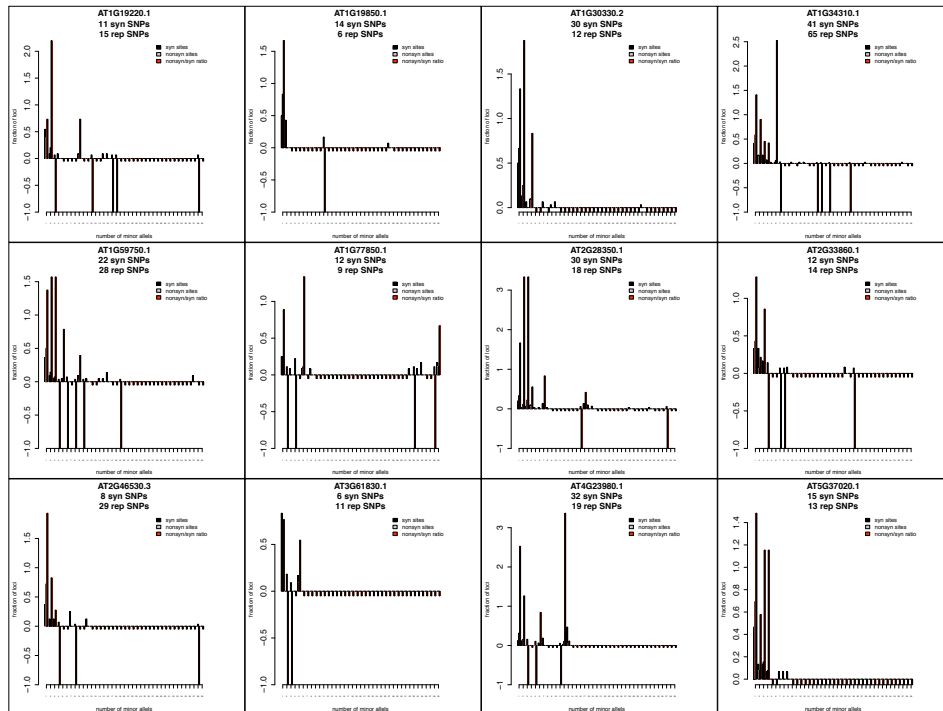


Supplementary Figure S1: Individual MAF distribution plot for the auxin signaling genes. Black: synonymous MAF; grey: nonsynonymous MAF; red: nonsynonymous / synonymous - ratio. If the ratio could not be calculated, values were set for each number of minor allele class to -0.1, if the ratio resulted in 'Inf', the values for the number of minor allele class was set to -1.

AUX/IAAs - continued

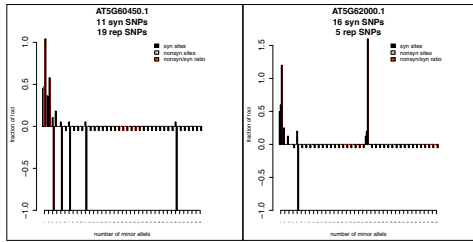


ARFs

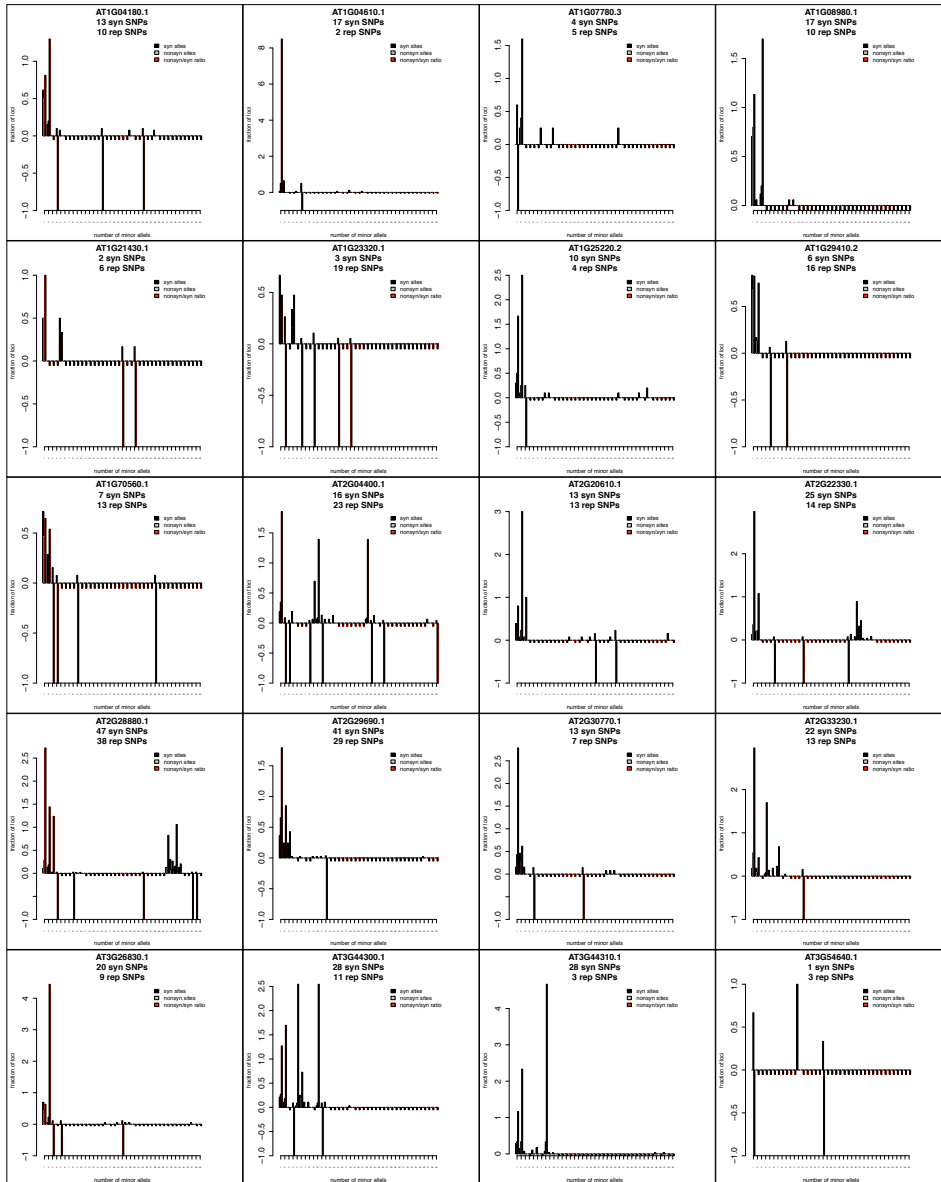


Supplementary Figure S2: Individual MAF distribution plot for the auxin synthesis genes. Black: synonymous MAF; grey: nonsynonymous MAF; red: nonsynonymous / synonymous - ratio. If the ratio could not be calculated, values were set for each number of minor allele class to -0.1, if the ratio resulted in 'Inf', the values for the number of minor allele class was set to -1.

ARFs - continued

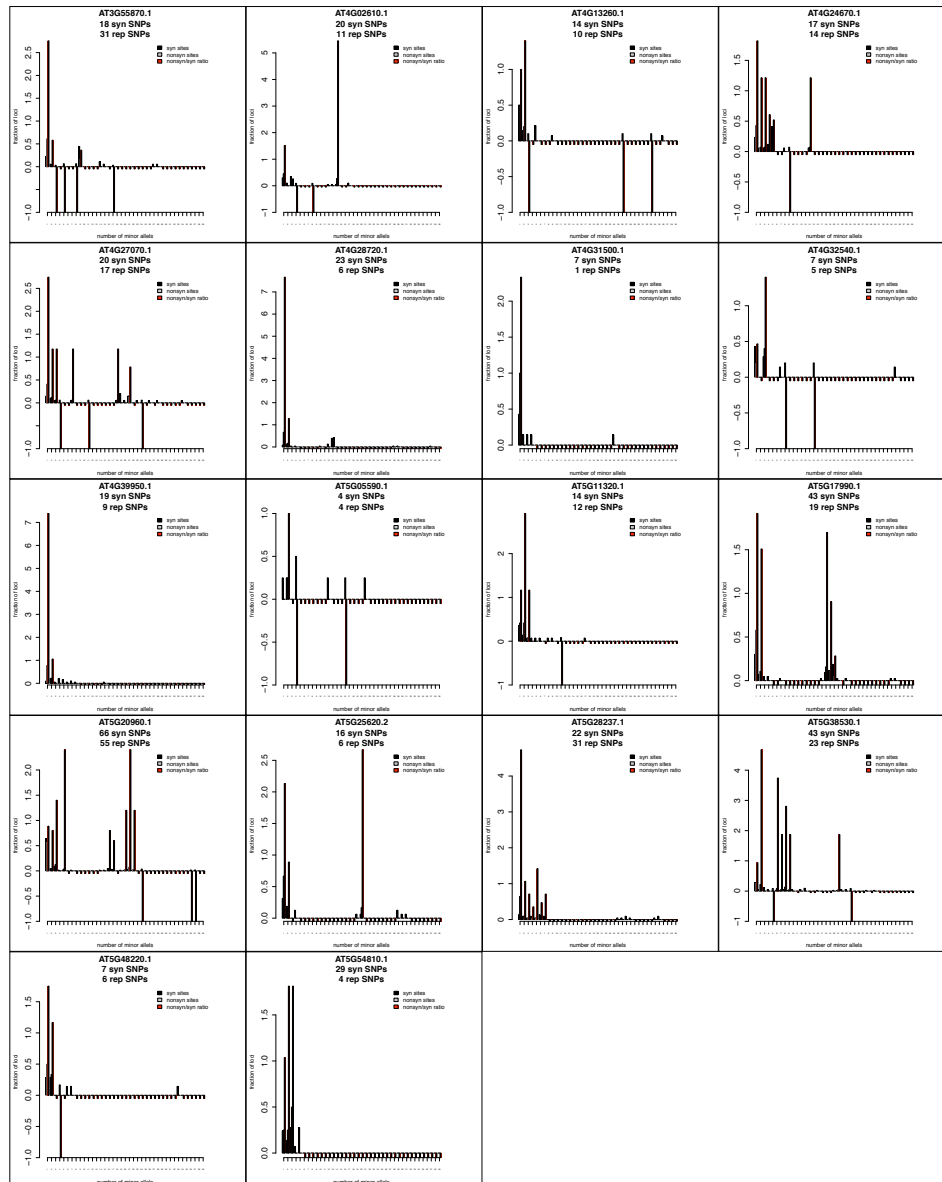


auxin synthesis

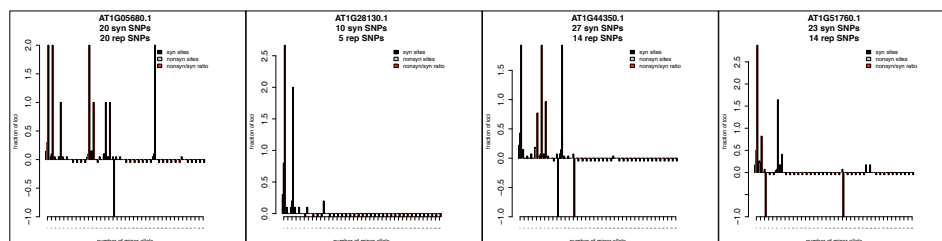


Supplementary Figure S3: Individual MAF distribution plot for the auxin metabolism and transport genes. Black: synonymous MAF; grey: nonsynonymous MAF; red: nonsynonymous / synonymous - ratio. If the ratio could not be calculated, values were set for each number of minor allele class to -0.1, if the ratio resulted in 'Inf', the values for the number of minor allele class was set to -1.

auxin synthesis - continued

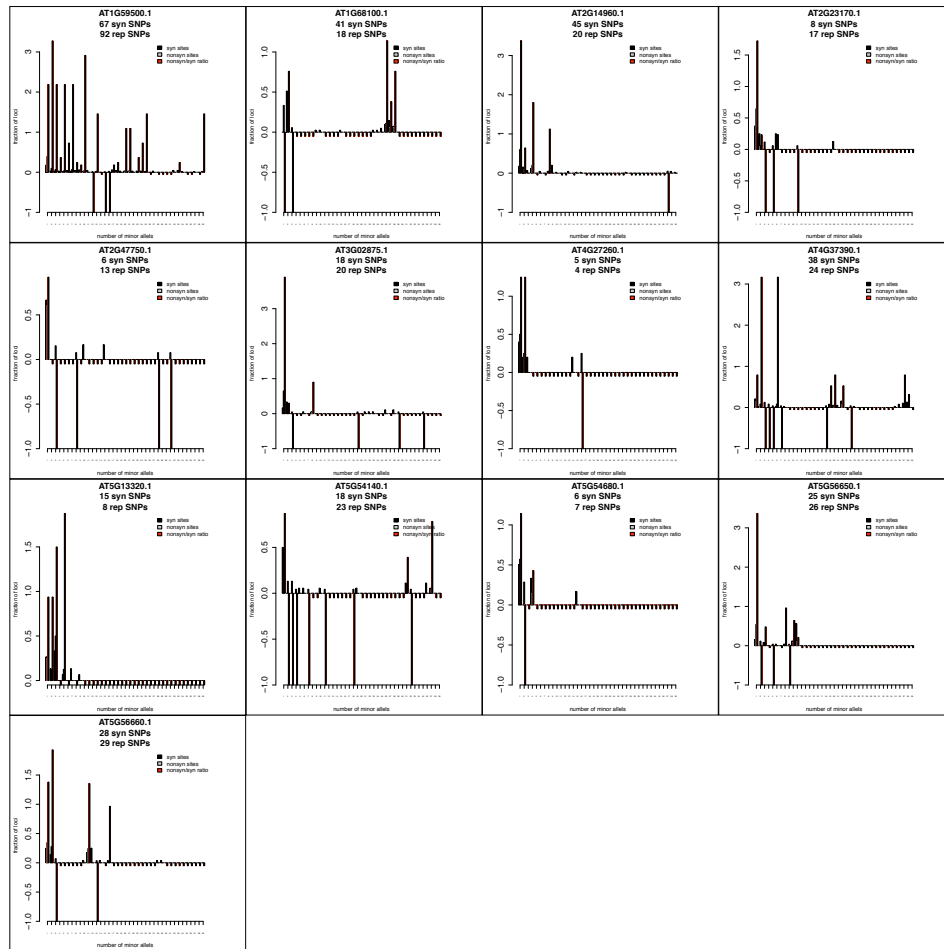


auxin metabolism

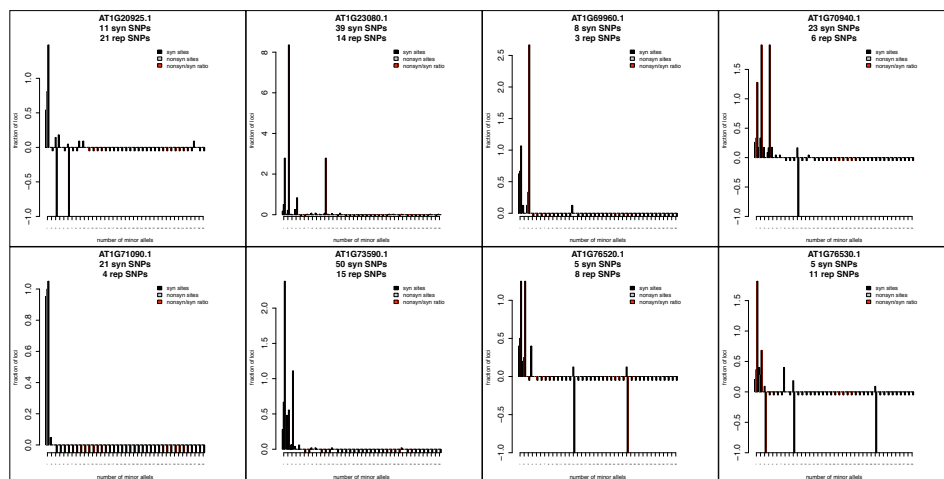


Supplementary Figure S4: Individual MAF distribution plot for the auxin metabolism and transport genes. Black: synonymous MAF; grey: nonsynonymous MAF; red: nonsynonymous / synonymous - ratio. If the ratio could not be calculated, values were set for each number of minor allele class to -0.1, if the ratio resulted in 'Inf', the values for the number of minor allele class was set to -1.

auxin metabolism - continued

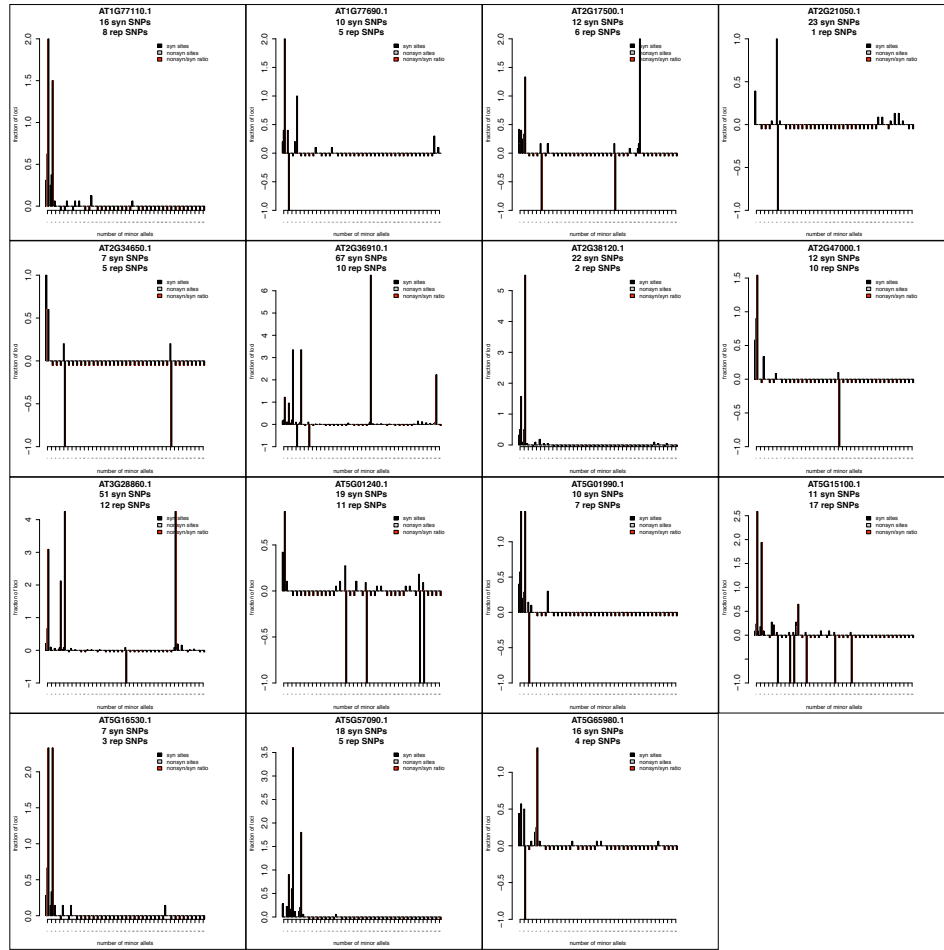


auxin transport

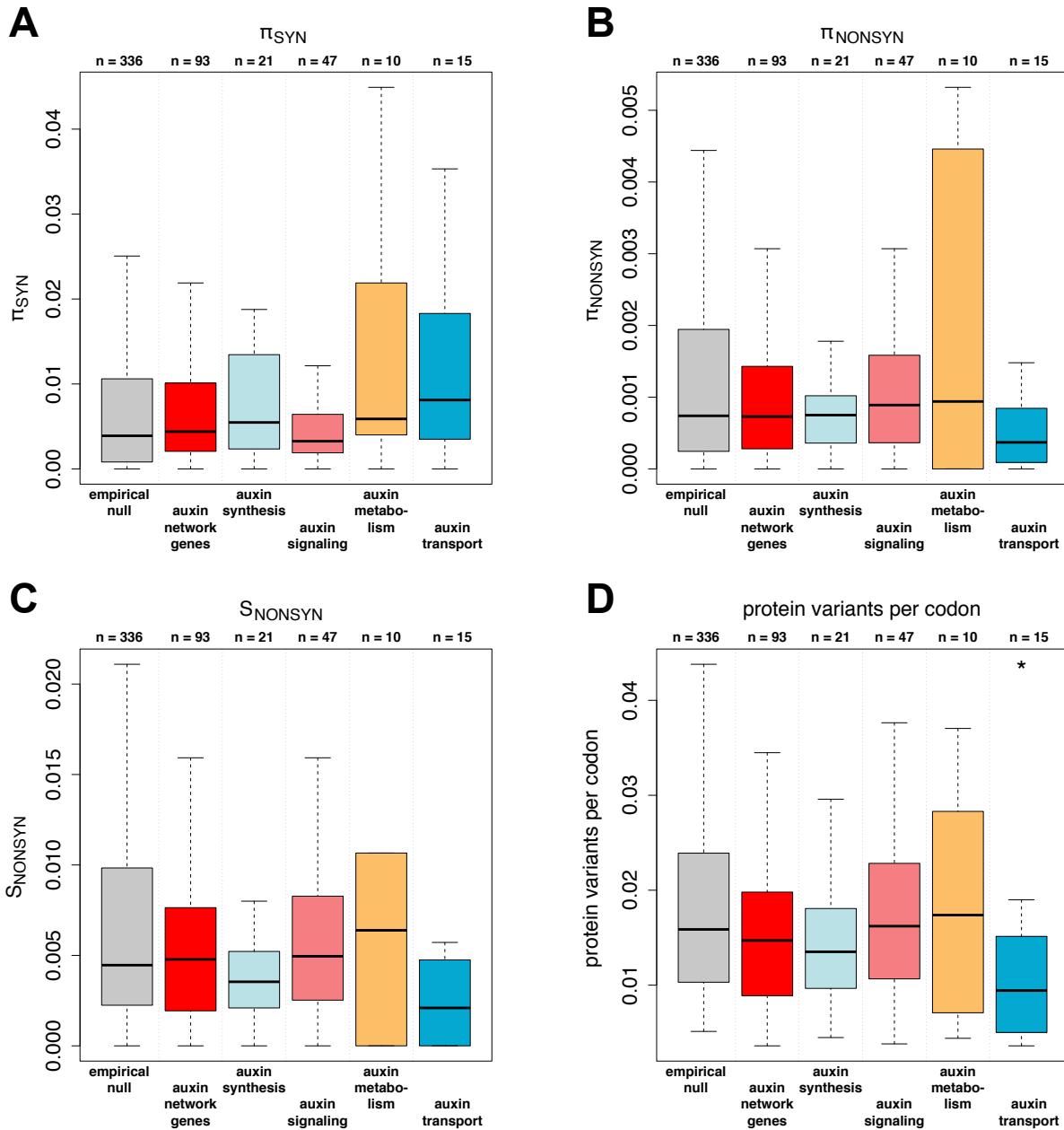


Supplementary Figure S5: Individual MAF distribution plot for the auxin metabolism and transport genes. Black: synonymous MAF; grey: nonsynonymous MAF; red: nonsynonymous / synonymous - ratio. If the ratio could not be calculated, values were set for each number of minor allele class to -0.1, if the ratio resulted in 'Inf', the values for the number of minor allele class was set to -1.

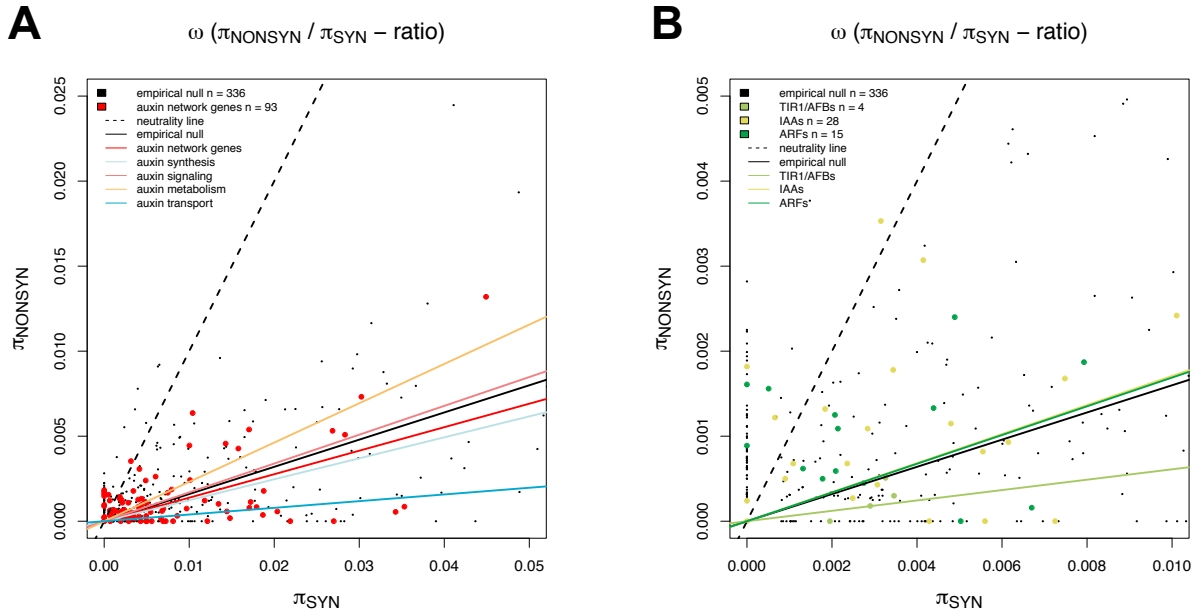
auxin transport - continued



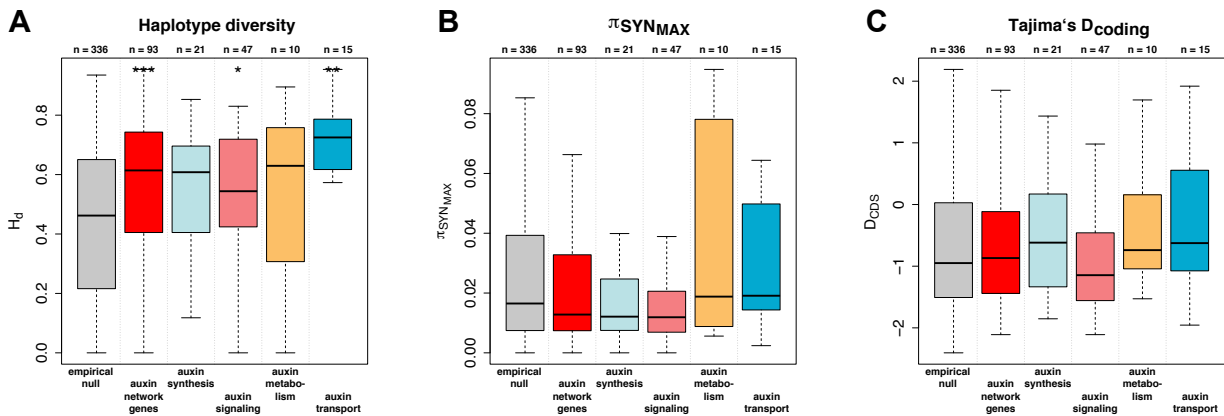
Supplementary Figure S6: Individual MAF distribution plot for the auxin metabolism and transport genes. Black: synonymous MAF; grey: nonsynonymous MAF; red: nonsynonymous / synonymous - ratio. If the ratio could not be calculated, values were set for each number of minor allele class to -0.1, if the ratio resulted in 'Inf', the values for the number of minor allele class was set to -1.



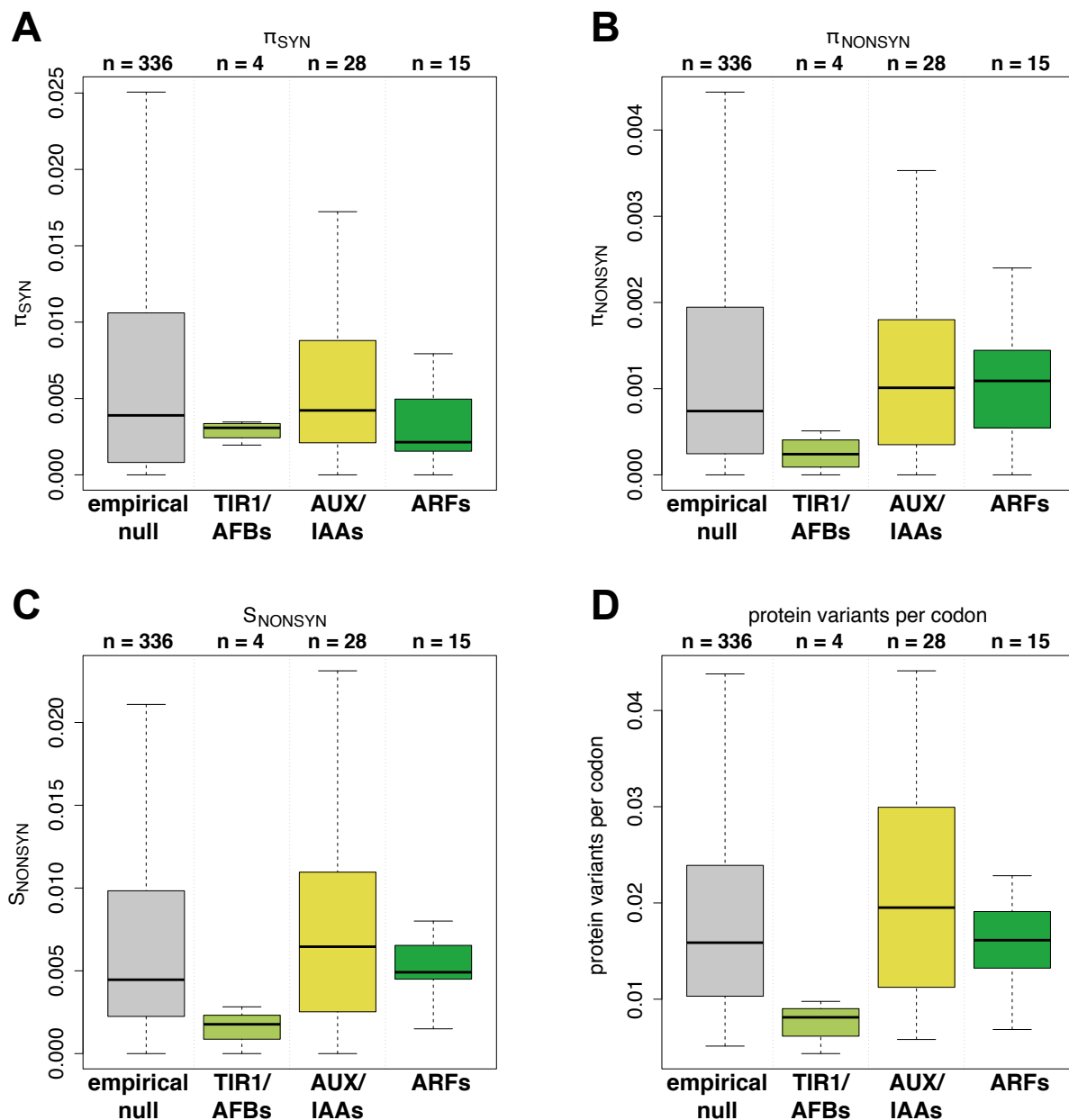
Supplementary Figure S7: Nucleotide diversity of auxin network gene groups considering coding sequences for the re-sequenced data. (A) π_{SYN} , (B) π_{NONSYN} , (C) S_{NONSYN} , (D) protein variants per codon of auxin network gene groups in the re-sequenced data compared to the corresponding empirical null distribution. Grey: empirical null distribution; red: auxin network genes as one group; light blue: auxin synthesis group; light red: auxin signaling group; orange: auxin metabolism group; blue: auxin transport group. Asterisk indicates a significant difference between empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$).



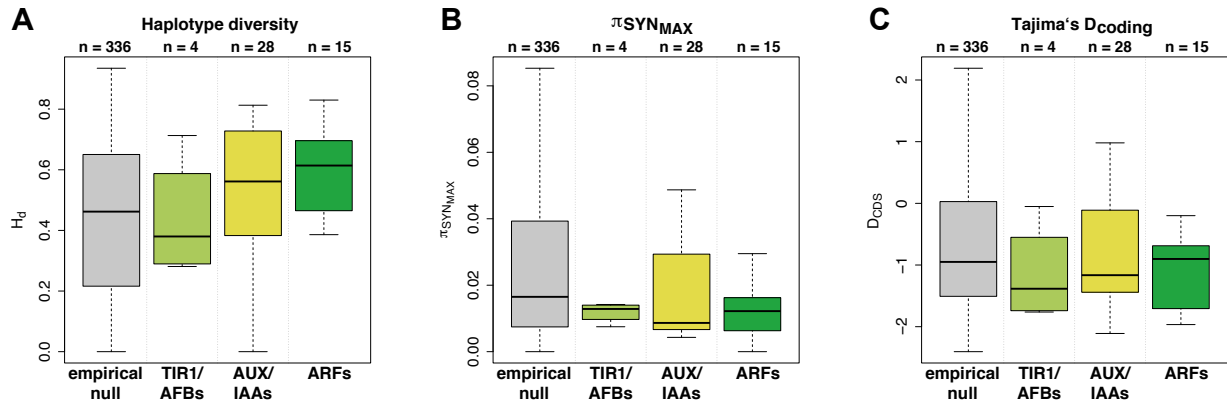
Supplementary Figure S8: Ratio plot of π_{NONSYN} and π_{SYN} for auxin network gene groups for the re-sequenced data. (A) π_{NONSYN} / π_{SYN} - ratio for auxin network gene groups in the re-sequenced data and the corresponding empirical null distribution, (B) π_{NONSYN} / π_{SYN} - ratio for auxin signaling group in the re-sequenced data and the corresponding empirical null distribution. Each black dot represents one representative gene model of the empirical null distribution, each red dot represents one of the auxin network genes representative gene models. Dashed line represents the neutrality line, if synonymous and nonsynonymous mutations would occur at same frequencies during evolution. Black line represents regression line through zero of the empirical null distribution. Colors given in the legend explain assignment of regression lines for the individual auxin network groups. Red: auxin network genes as one group; light blue: auxin synthesis group; light red: auxin signaling group; orange: auxin metabolism group; blue: auxin transport group.



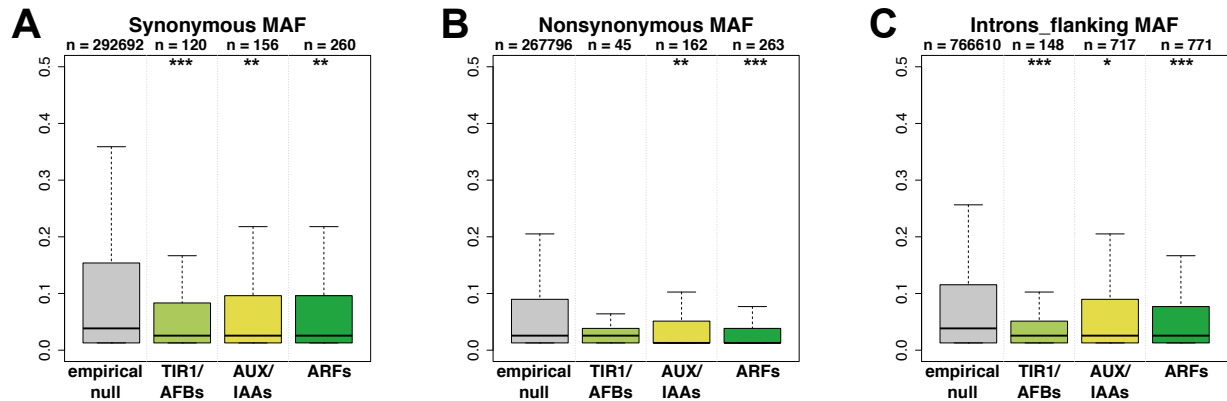
Supplementary Figure S9: Nucleotide variation patterns of auxin network gene groups considering coding sequences for the re-sequenced data. (A) Haplotype diversity (H_d), (B) π_{SYN_MAX} , (C) Tajima's D for coding sites (D_{CDS}) of auxin network gene groups in the re-sequenced data compared to the corresponding empirical null distribution. Grey: empirical null distribution; red: auxin network genes as one group; light blue: auxin synthesis group; light red: auxin signaling group; orange: auxin metabolism group; blue: auxin transport group. Asterisk indicates a significant difference between empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).



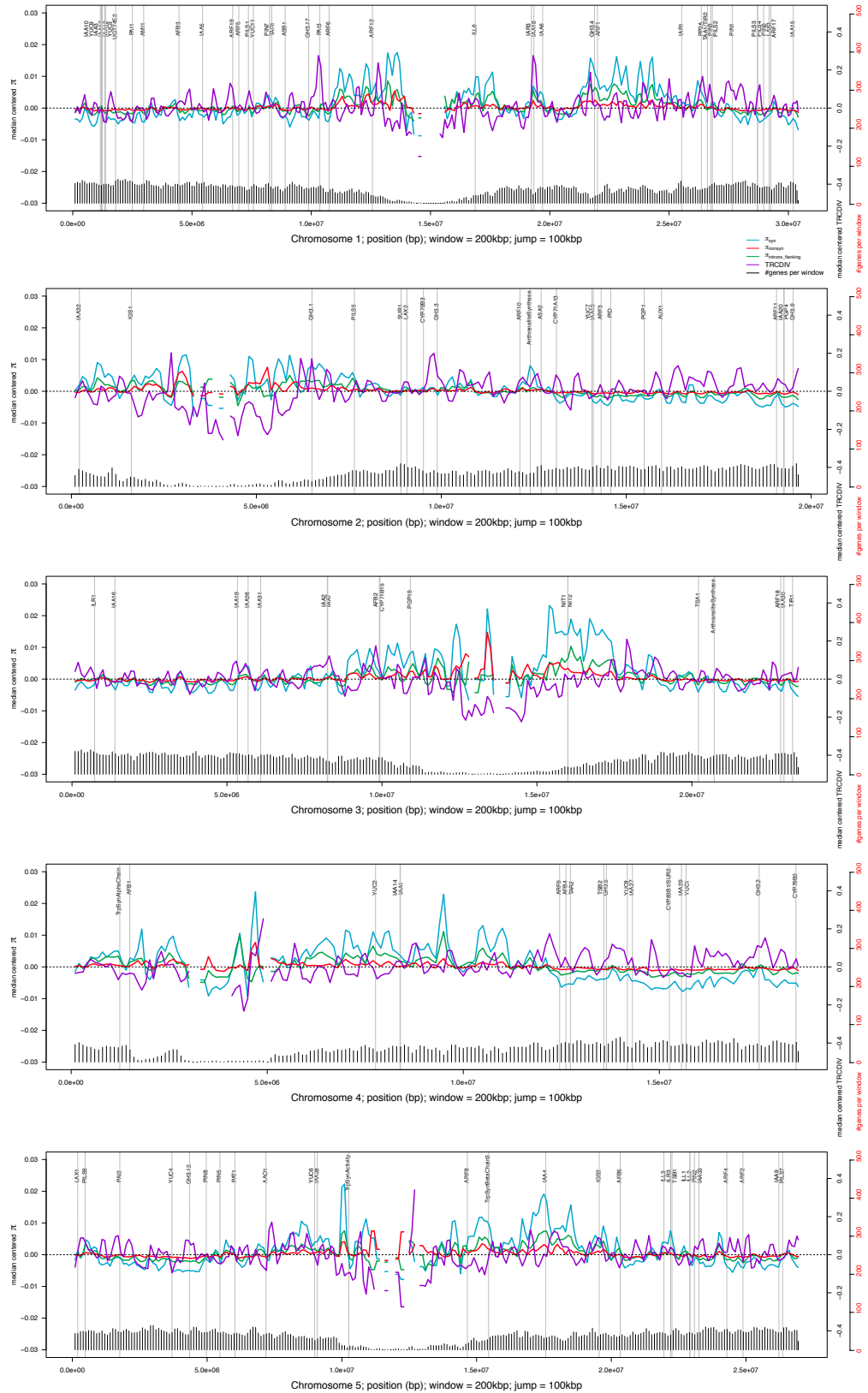
Supplementary Figure S10: Nucleotide diversity of auxin signaling gene families considering coding sequences for the re-sequenced data. (A) π_{SYN} , (B) π_{NONSYN} , (C) S_{NONSYN} , (D) protein variants per codon of auxin signaling gene families in the re-sequenced data compared to the corresponding empirical null distribution. Grey: empirical null distribution; green: *TIR1/AFBs*; light green: *AUX/IAAs*; dark green: *ARFs*. Asterisk indicates a significant difference between empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).



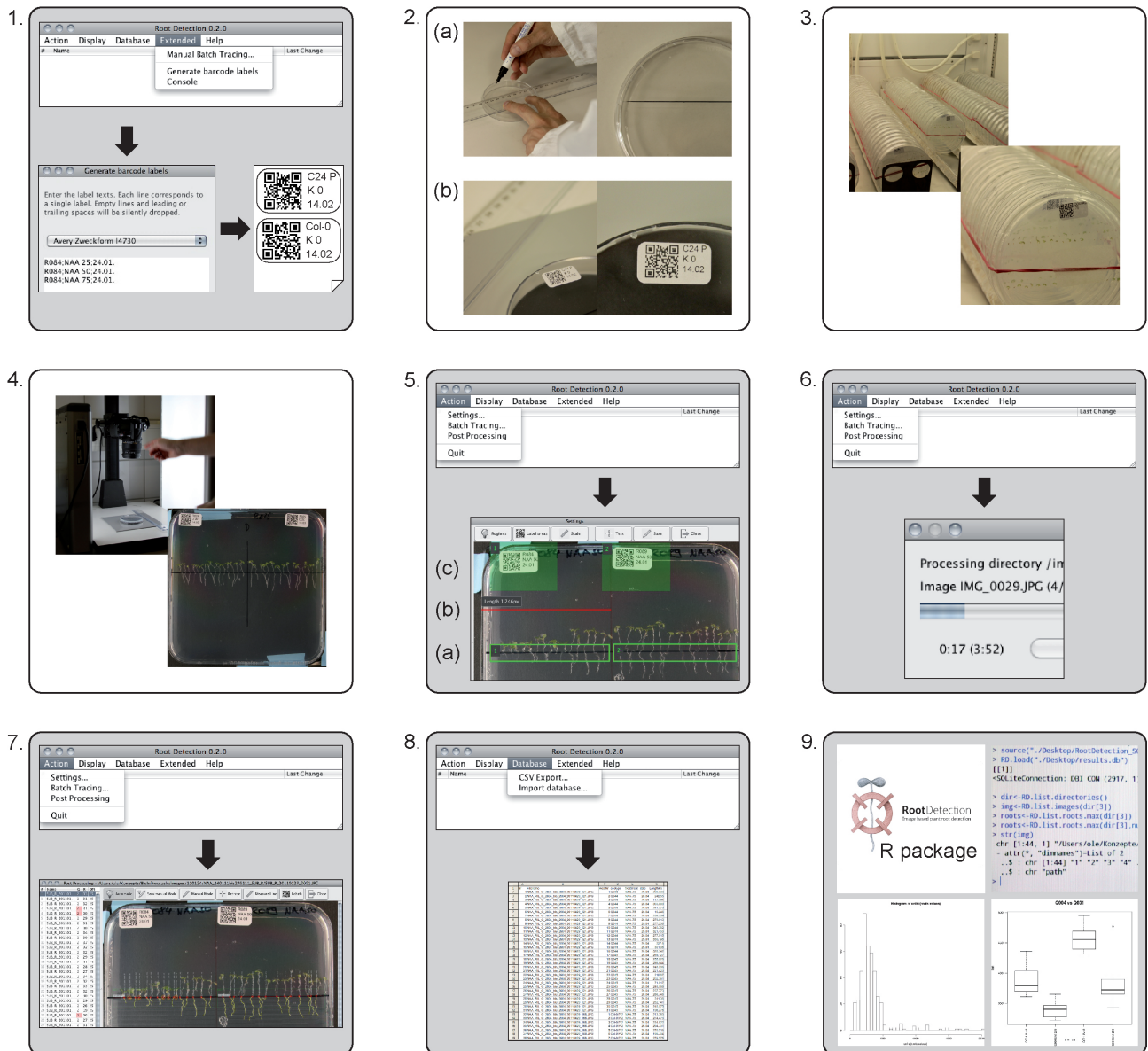
Supplementary Figure S11: Nucleotide variation patterns of auxin signaling gene families considering coding sequences for the re-sequenced data. (A) Haplotype diversity (H_d), (B) π_{SYNMAX} , (C) Tajima's D for coding sites (D_{CDs}) of auxin signaling gene families in the re-sequenced data compared to the corresponding empirical null distribution. Grey: empirical null distribution; green: *TIR1/AFBs*; light green: *AUX/IAAs*; dark green: *ARFs*. Asterisk indicates a significant difference between empirical null distribution and auxin network gene groups (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).



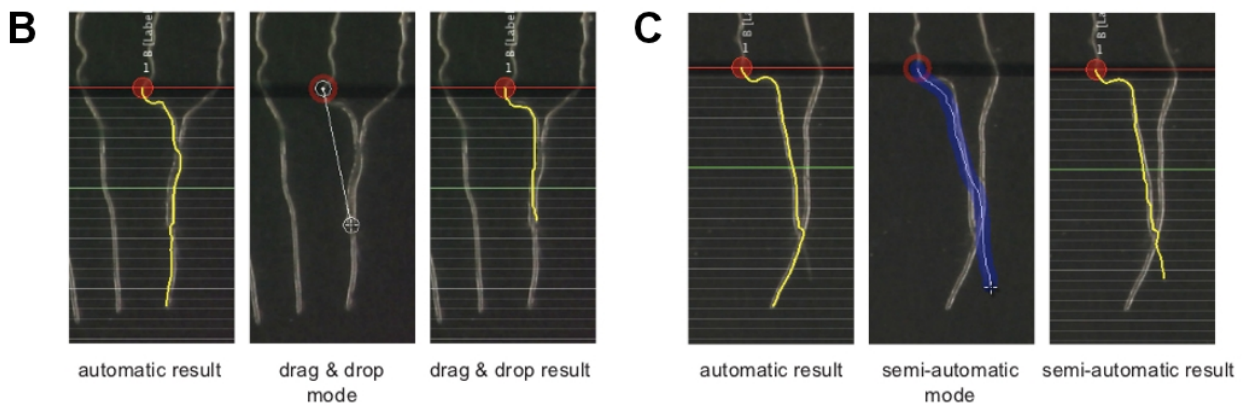
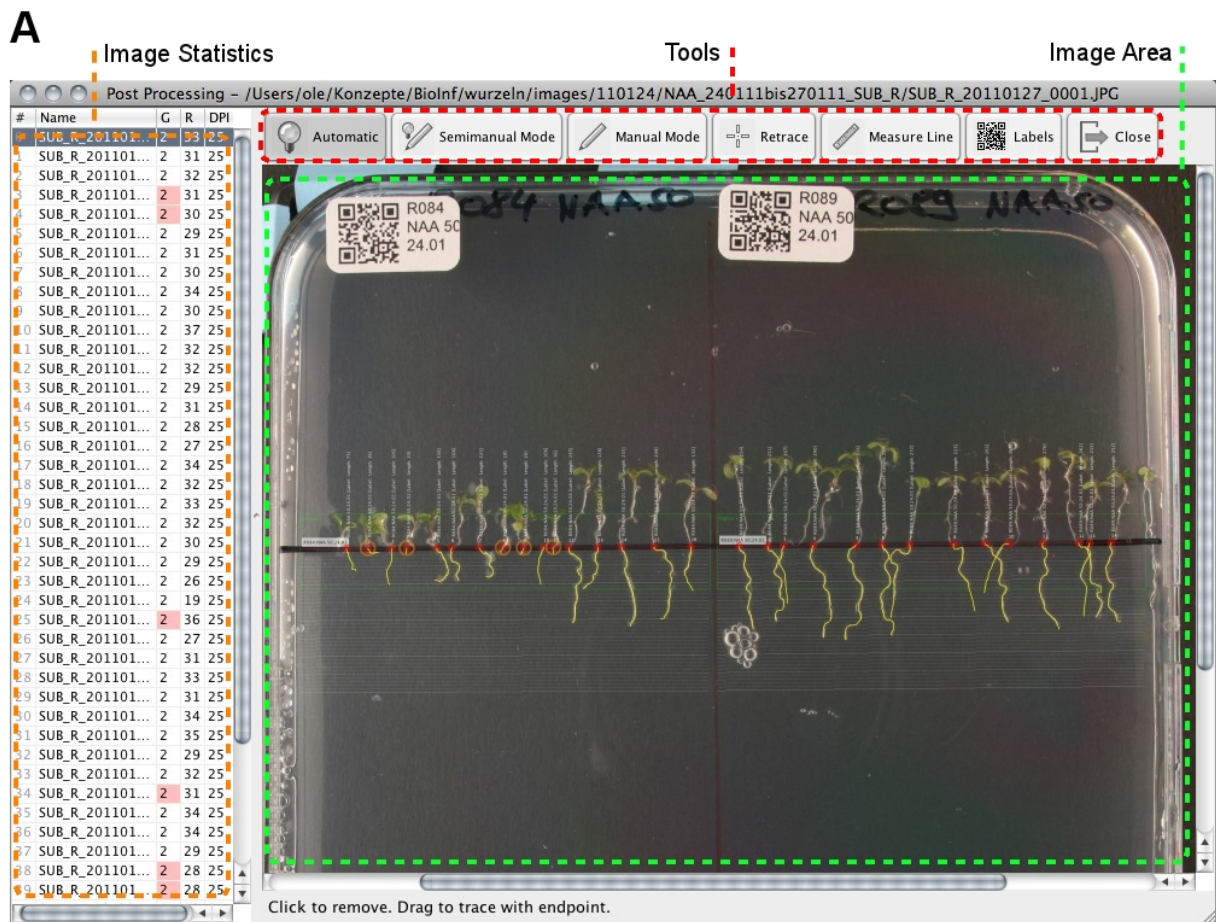
Supplementary Figure S12: Minor allele frequency distribution of auxin signaling families presented as boxplots. (A) Synonymous MAF (B) Nonsynonymous MAF (C) Introns_flanking MAF of individual auxin network groups. Grey: empirical null distribution; green: *TIR1/AFBs*; light green: *AUX/IAAs*; dark green: *ARFs*. The number of analyzed sites is given under each boxplot. Asterisk indicate significant differences between the empirical null distribution and auxin signaling families (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$)).



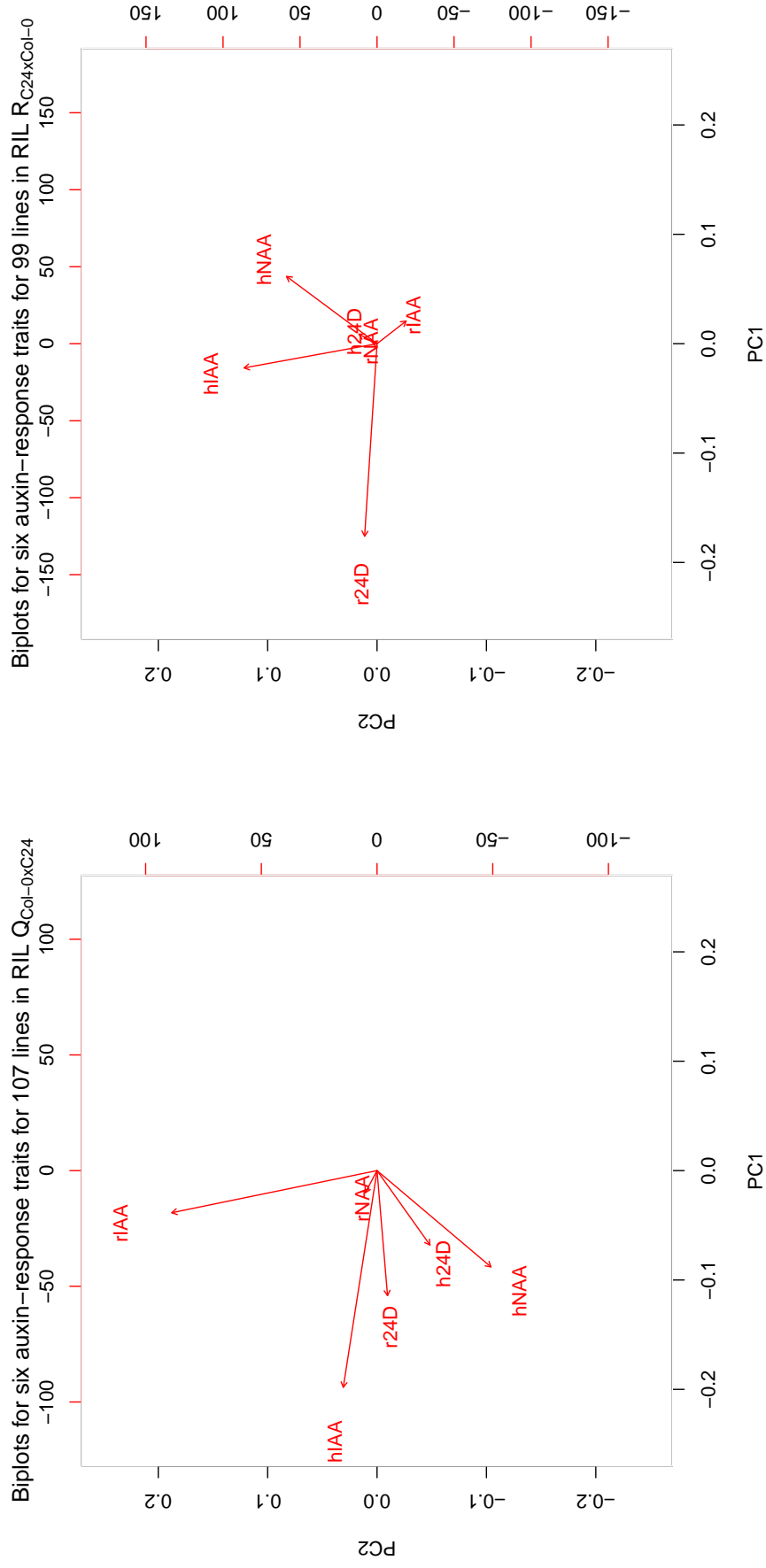
Supplementary Figure S13: Chromosome-wide transcript diversity among six *A. thaliana* accessions compared to genome-wide nucleotide diversity among 80 *A. thaliana* accessions. Median centered values calculated for windows of 200 kbp length and a jump of 100kbp. Blue: π_{syn} , red: π_{nonsyn} , green: $\pi_{introns_flanking}$ and purple: TRCDIV. Black bars on the bottom represent number of gene models in the corresponding sliding window.



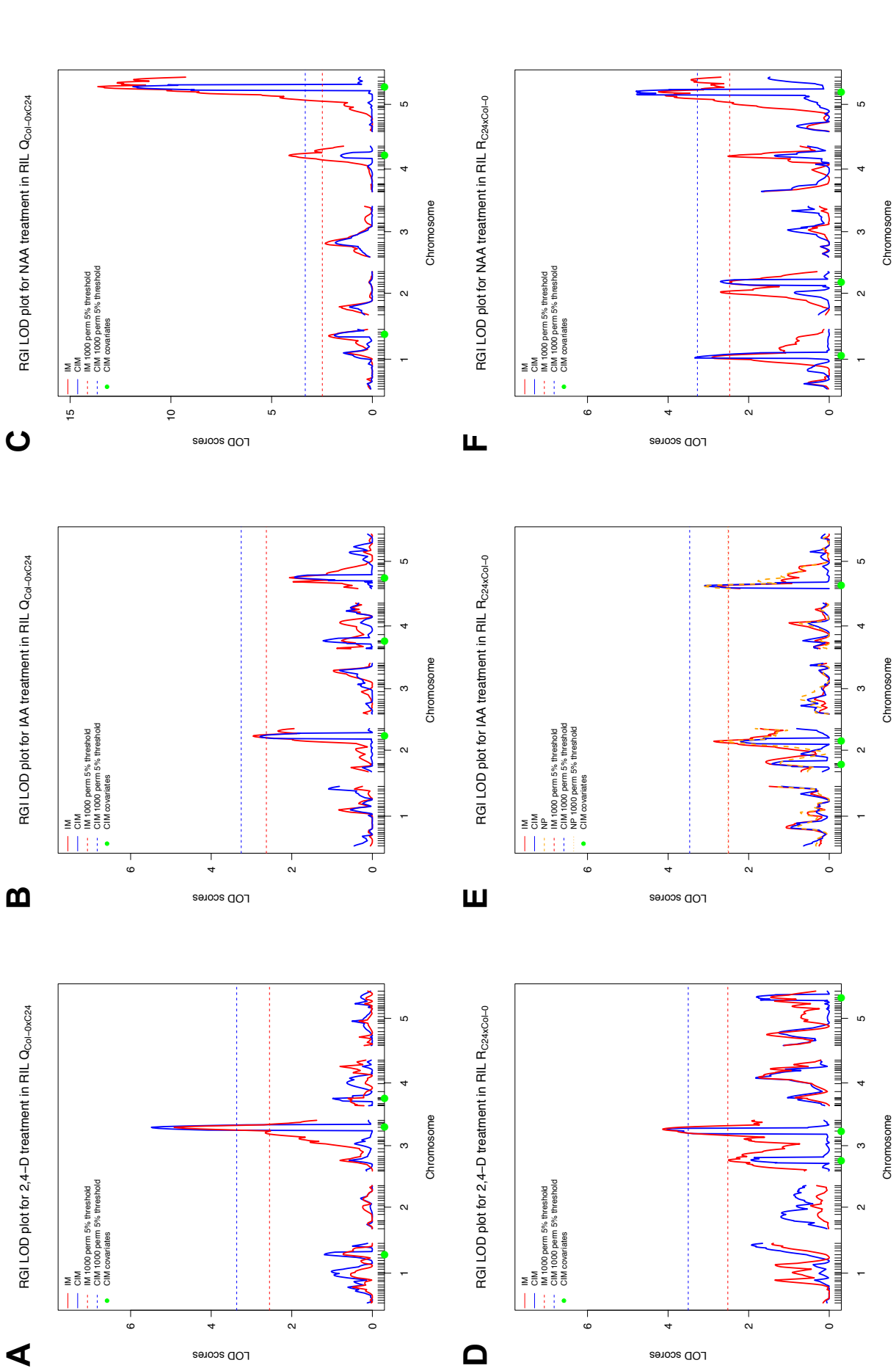
Supplementary Figure S14: Exemplary workflow measuring seedling root length with RootDetection. Boxes with grey background indicate steps which are performed with the help of RootDetection. (1) Generate QR-code labels with RootDetection. (2a) Mark plates with a black start line from which RootDetection will start measuring seedling root length. (2b) Stick QR code labels (generated with RootDetection) printed on label pads to plates. (3) Perform seedling root growth transfer assay. (4) Photograph plates keeping the same camera settings for all pictures. (5a) Define image regions to be evaluated. (5b) Define the scale of the images. (5c) Define label regions for QR code recognition. (6) Perform fully automatic batch trace mode. (7) Perform optional RootDetection post-processing mode. (8) Export results from the SQLite database. (9) Visualize results with the RootDetection R package by connecting to the SQLite database.



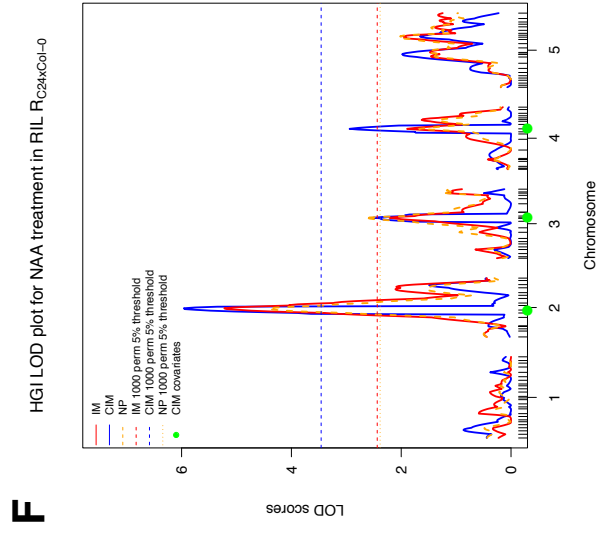
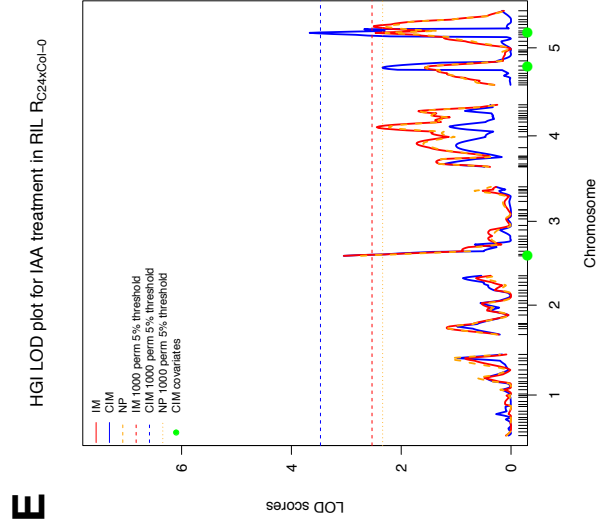
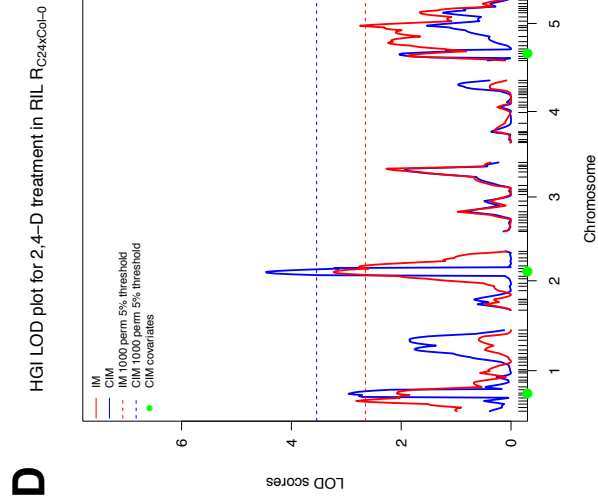
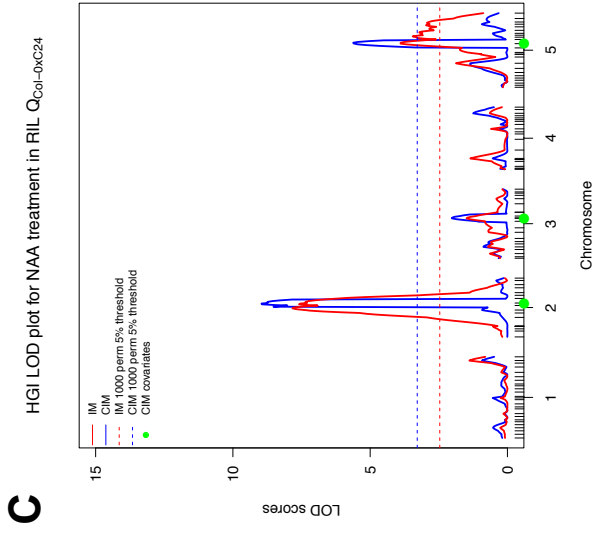
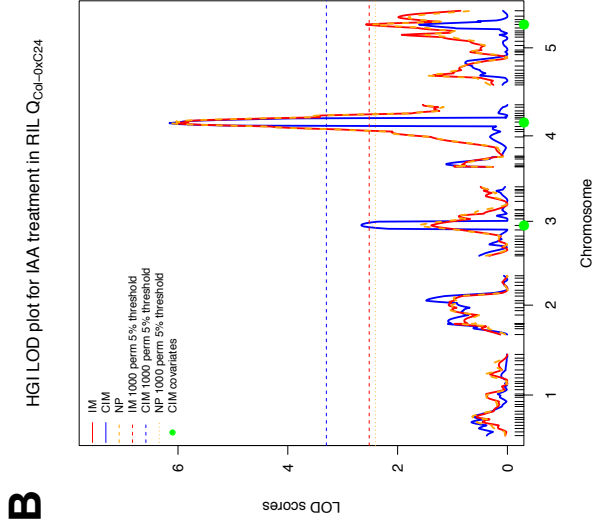
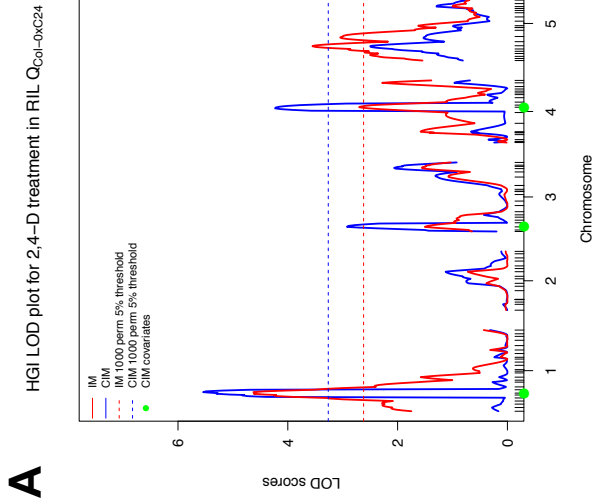
Supplementary Figure S15: Screenshots from RootDetection's post processing mode. (A) Overview of the post processing window. (B) Root length correction with the drag & drop mode ('Automatic'). (C) Root length correction with the semi-automatic mode ('Semimanual Mode').



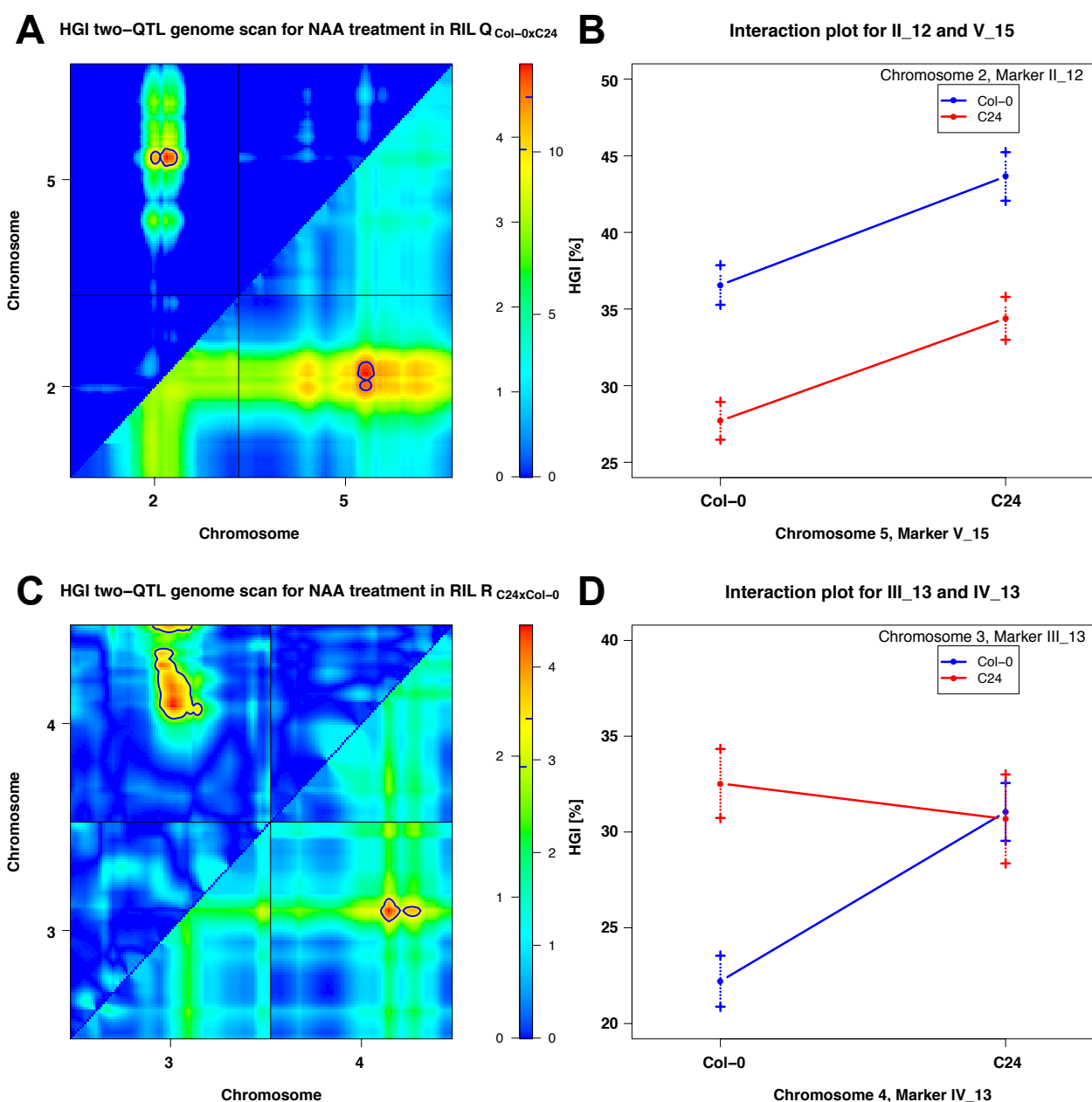
Supplementary Figure S16: Biplots of 6 auxin response traits analyzed in two RIL populations. (A) Biplots for 103 RIL lines for which phenotypic data is available throughout all 6 conducted auxin response experiments in the $Q_{C0L-0 \times C24}$ RIL population. (B) Biplots for 99 RIL lines in the $R_{C24 \times C0L-0}$ RIL population. Each auxin response trait influencing root growth inhibition (RGI) is indicated as r24D for 2,4-D treatment, rIAA for IAA and rNAA for NAA treatment. Hypocotyl growth inhibition assays are indicated for the same auxins with the prefix "h".



Supplementary Figure S17: Main effect QTLs associated with RGI after different auxin treatments. LOD profiles from standard interval (red), composite interval mapping (blue) and nonparametric mapping (orange) are shown for all five *A. thaliana* chromosomes for auxin response traits (20 nM 2,4-D, 40 nM IAA and 75 nM NAA) in the $Q_{Col-0} \times C24$ (A + B + C) and $R_{C24 \times Col-0}$ (D + E + F) RIL population. For both RIL populations and all mapping methods also the 5% significant thresholds (dashed lines) are given as calculated by 1000 permutations. Green dots indicate covariate covariates used by the composite interval mapping method selected by a forward approach. Y-axis denote calculated LOD scores.

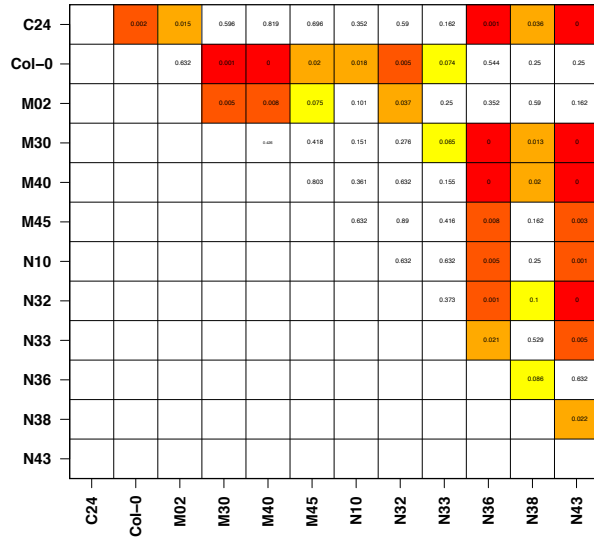


Supplementary Figure S18: Main effect QTLs associated with HGI after different auxin treatments. LOD profiles from standard interval (red), composite interval mapping (blue) and nonparametric mapping (orange) are shown for all five *A. thaliana* chromosomes for auxin response traits (375 nM 2,4-D, 500 nM IAA and 500 nM NAA) in the $Q_{Col-0} \times C24$ (A + B + C) and $R_{C24} \times Col-0$ (D + E + F) RIL population. For both RIL populations and all mapping methods also the 5% significant thresholds (dashed lines) are given as calculated by 1000 permutations. Green dots indicate covariates used by the composite interval mapping method selected by a forward approach. Y-axis denote calculated LOD scores.

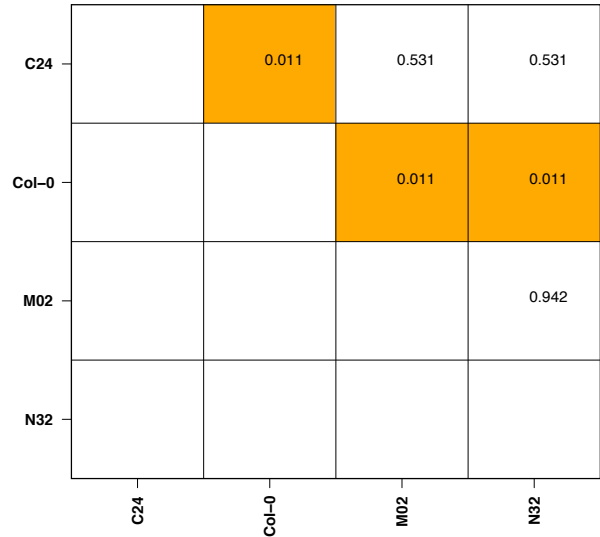


Supplementary Figure S19: Example of significant *two-QTL genome scan* interactions in the $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ RIL populations for HGI. A *scantwo* plot with x- and y-axis indicating cM positions along the given chromosomes and the calculated LOD values for additive QTL model (M_{add} below diagonal) and conditional additive QTL model (M_{av1} above diagonal) is shown for HGI after NAA treatment in the $Q_{Col-0 \times C24}$ RIL population (A). LOD values for the additive QTL model (M_{add} below diagonal) and the epistatic interaction QTL model (M_{int} above diagonal) are shown for HGI after NAA treatment in the $R_{C24 \times Col-0}$ RIL population (C). 10000 permutations were used to calculate significance thresholds (see also Supplementary Table ST17 and materials and method section for a detailed description). Colored scale on the right indicates observed LOD value spectrum, blue tick marks highlight 1-LOD below LOD maximum as used to draw contours in the *scantwo* plot. Additive interaction of two markers (B) and possible epistatic interaction (D) of two markers nearest to positions indicated in the *scantwo* summary table (see Supplementary Table ST17) are shown as *effectplots*. Y-axis denote HGI in % of groups either having Col-0 or C24 allele at a given marker locus. Error bars denote \pm single standard error of the group phenotypic mean.

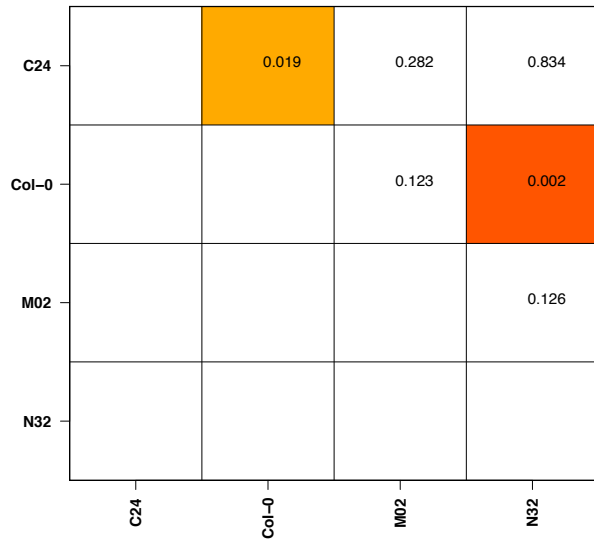
A Differences in growth responses determined by two-way ANOVA
75 nM NAA



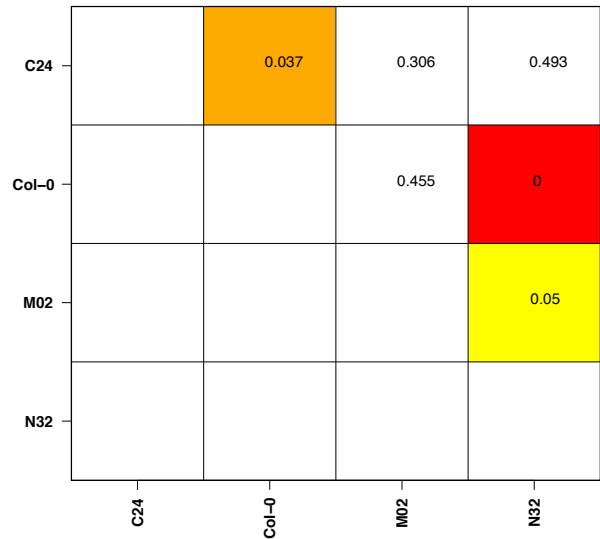
B Differences in growth responses determined by two-way ANOVA
25 nM NAA



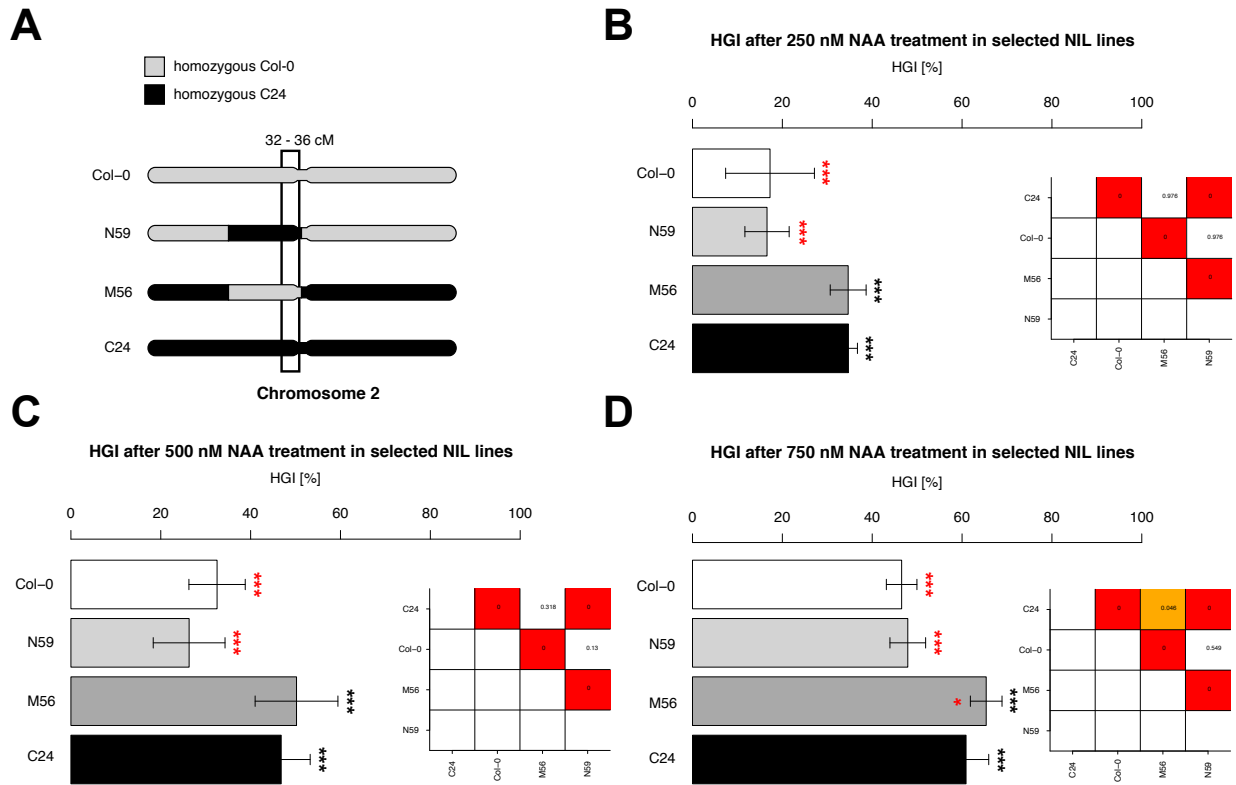
C Differences in growth responses determined by two-way ANOVA
75 nM NAA



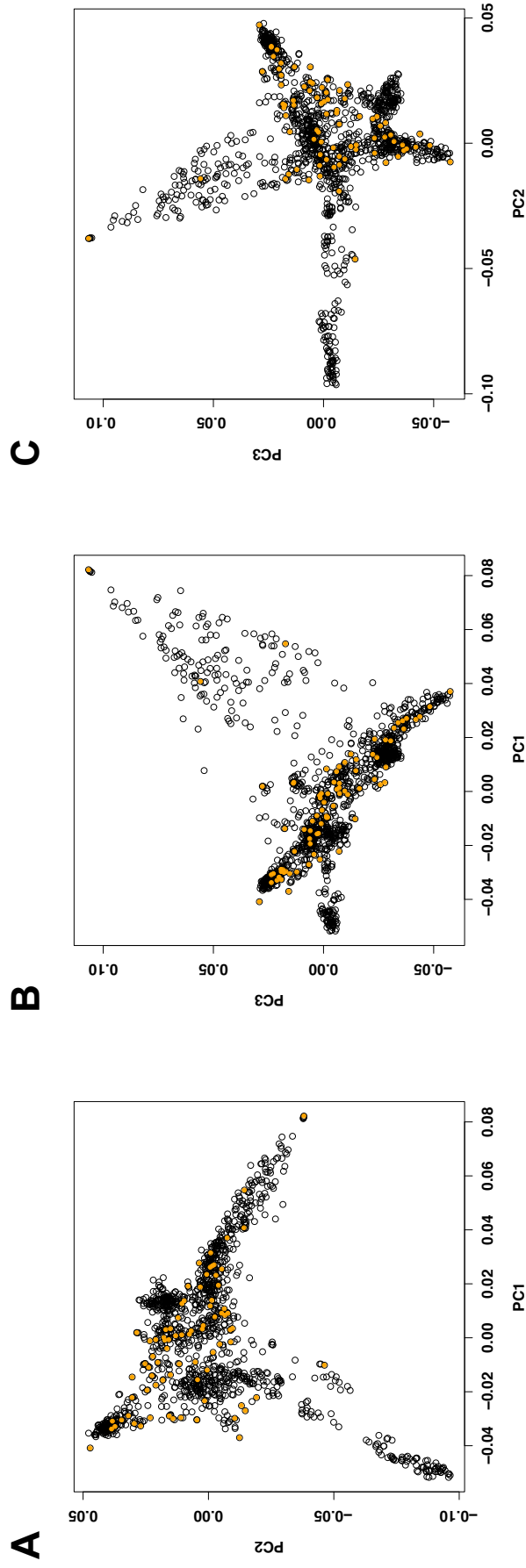
D Differences in growth responses determined by two-way ANOVA
125 nM NAA



Supplementary Figure S20: Supporting information for the validation of a QTL corresponding to the NAA response trait on RGI. Based on a two-way ANOVA analysis the differences of growth responses to NAA treatment were assessed by comparing NILs and parental lines in a pairwise fashion. All analyses were performed on log-transformed data, the corresponding p-values are indicated as numbers. A p-value of 0 corresponds to a p-value $< 2e^{-16}$. Significant differences after Benjamini-Hochberg correction ($P < 0.05$, $P < 0.01$, $P < 0.001$) are highlighted in light-orange, orange and red, respectively. Yellow indicates $0.05 < P < 0.1$.



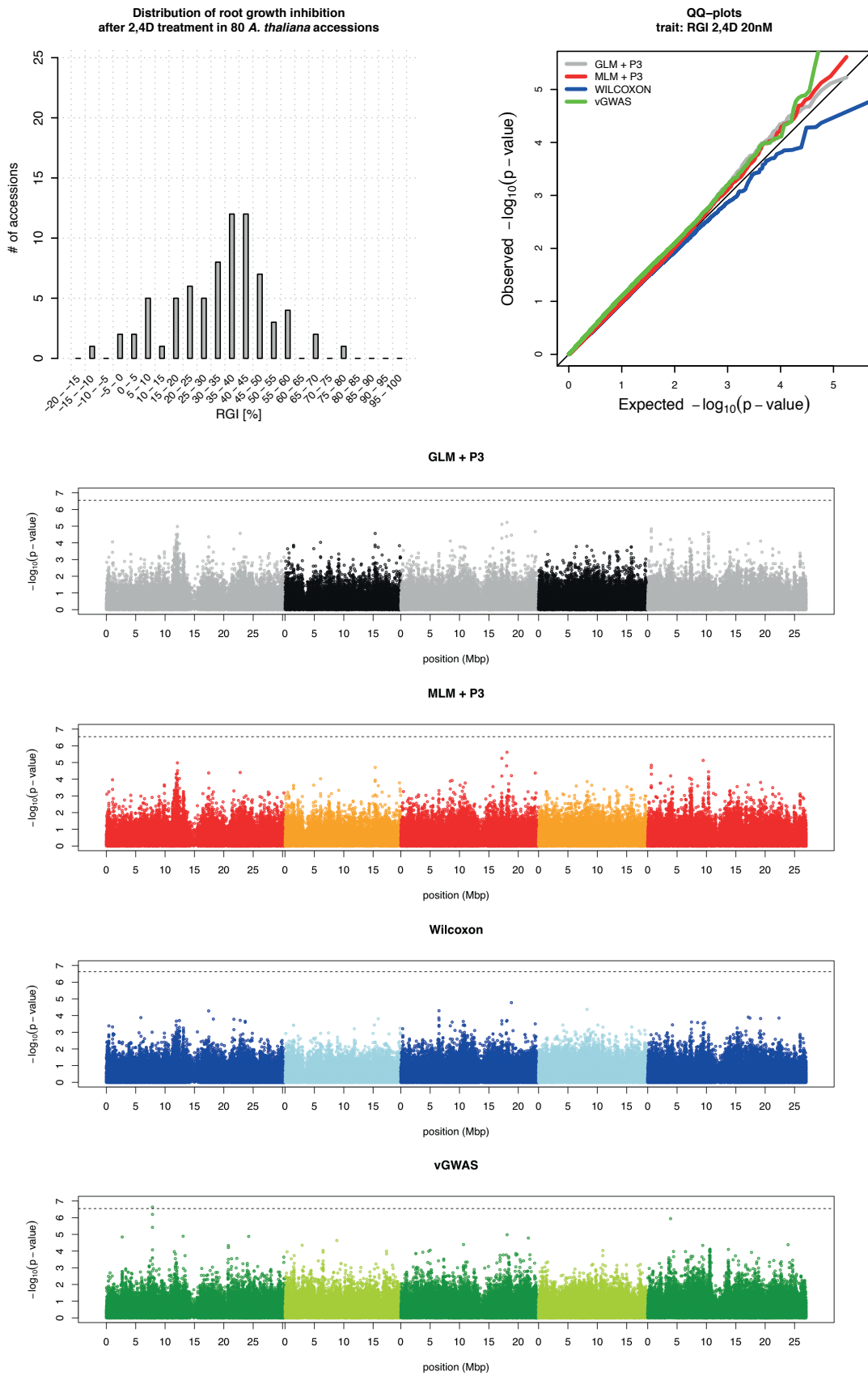
Supplementary Figure S21: An attempt to validate the QTL corresponding to the NAA response trait on HGI. In A the allelic composition of the analyzed NILs is visualized next to the 95% confidence interval based on the multiple QTL mapping results for chromosome 2 in the $R_{C24 \times Col-0}$ RIL population indicated by the black box (A). Allelic regions homozygous for Col-0 are highlighted in grey, whereas C24 is highlighted in black. NILs with the prefix N (lightgrey) have Col-0 genetic background, NILs with the prefix M (darkgrey) have C24 genetic background. Barplots represent the HGI after indicated NAA concentrations on the selected NILs N59 and M56. Error bars show % standard deviations. Black asterisks highlight significant response differences between Col-0 and the marked NILs based on a two-way ANOVA analysis shown as pairwise plot next to each barplot. Red asterisks denote significant response differences between C24 and the marked NILs (***, **, * significant after Benjamini-Hochberg correction at $P < 0.05$, $P < 0.01$, $P < 0.001$). A p-value of 0 corresponds to a p-value $< 2e^{-16}$.



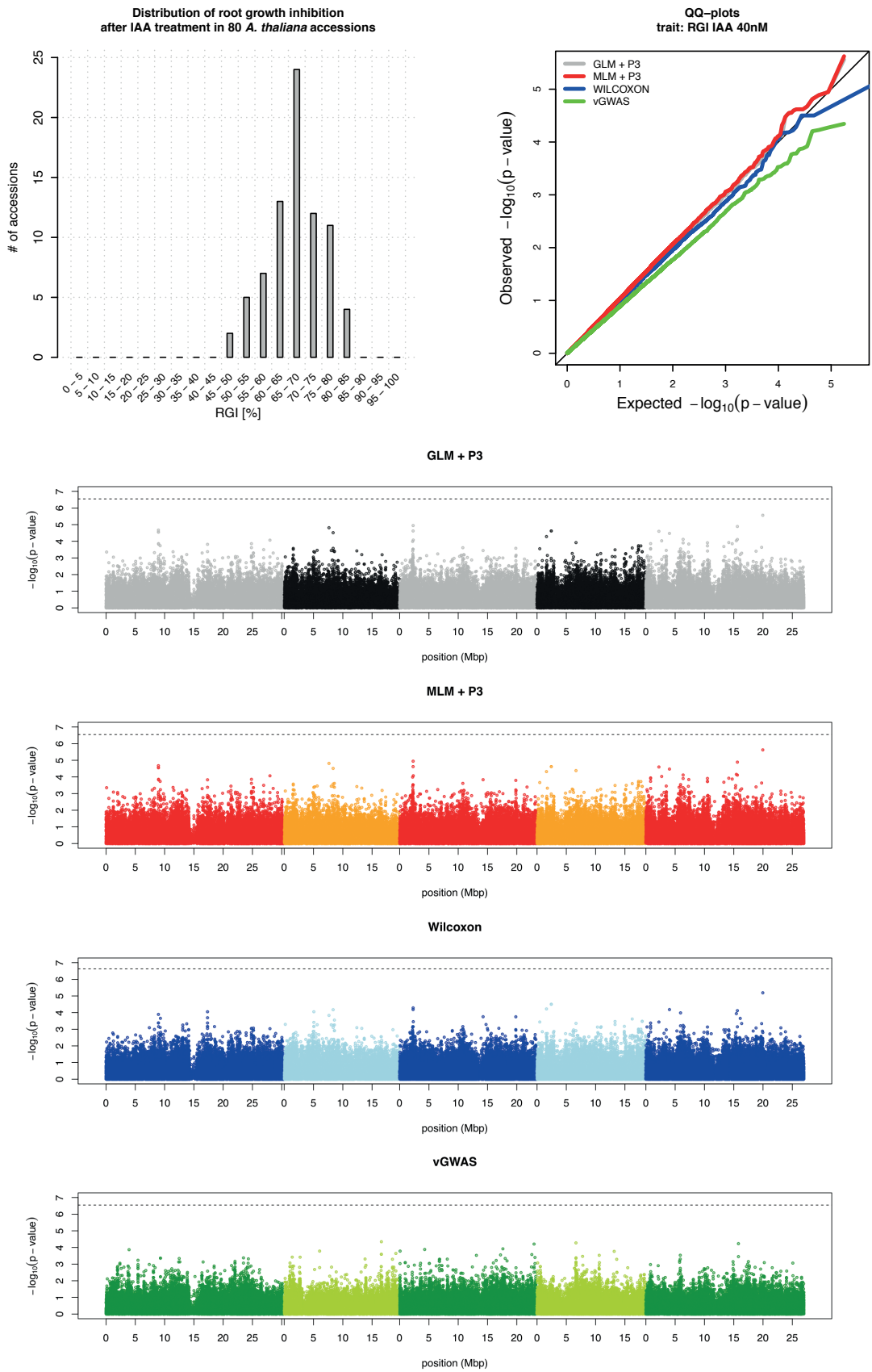
Supplementary Figure S22: Comparing the first 3 principle components of the P-matrix calculated based on 165984 SNPs in 1307 *A. thaliana* accessions. (A) Plot of principle component 2 (PC2) against principle component 1 (PC1). (B) Plot of principle component 3 (PC3) against PC1. (C) Plot of PC3 against PC2. Each dot represents one accession, dots with orange color highlight 80 *A. thaliana* accessions used in this study.

The following figure caption applies to all GWA results overviews (Supplementary Figure S20 to S25):

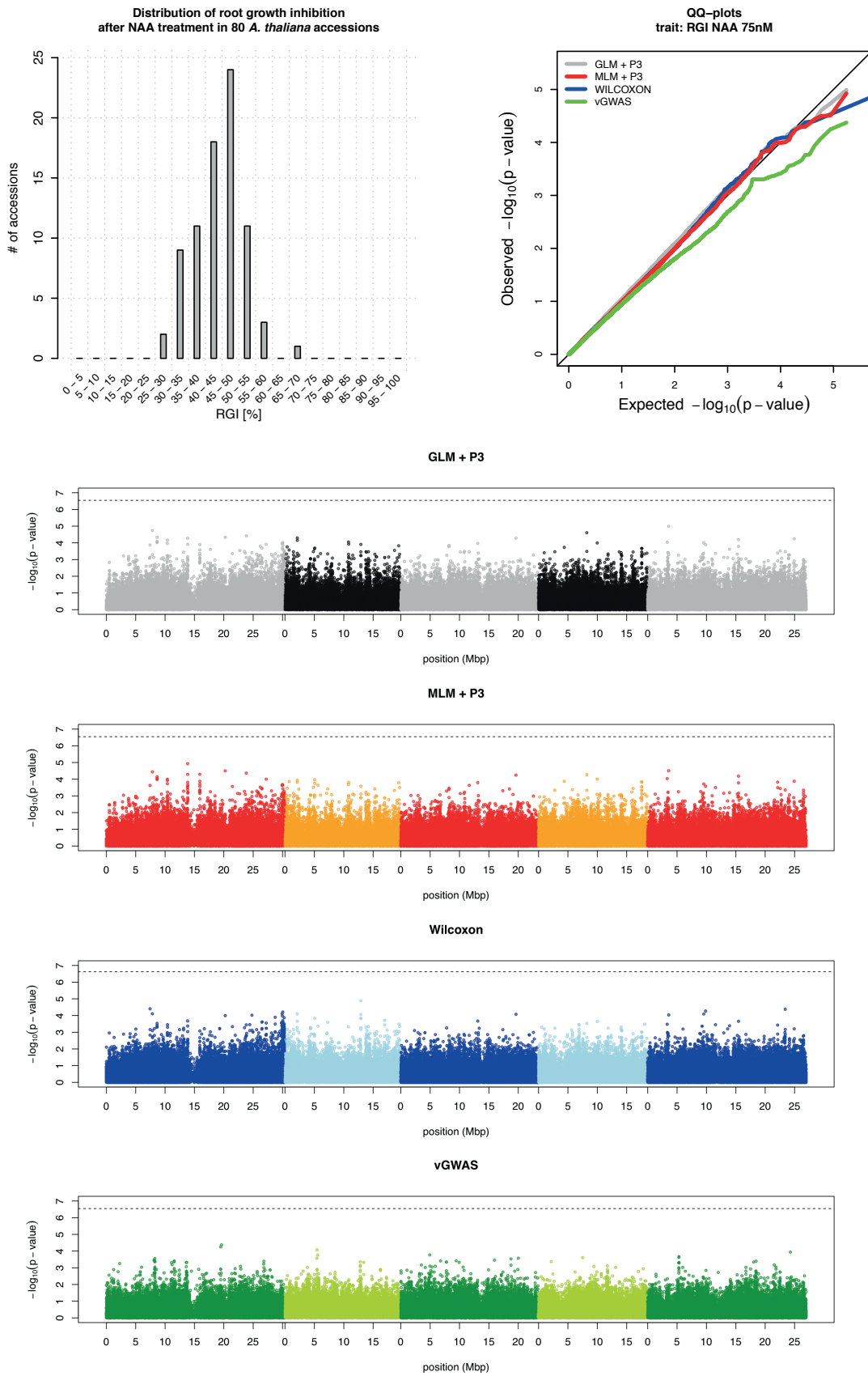
Top left: Histograms showing the frequency distribution of auxin response treatments (RGI or HGI) of the analyzed *A. thaliana* accessions. Top right: Comparison of different GWA methods by a QQ-plot. Observed $-\log_{10}$ p-values are plotted against expected $-\log_{10}$ p-values to illustrate the fit of the model applied. General linear models (GLMs) and Mixed linear models (MLMs) were calculated with additional population structure information based on the first 3 PCAs (+ P3). Genome-wide distribution of $-\log_{10}$ p-values of SNP/phenotype association for GLM, MLM, WILCOXON and variance-heterogeneity GWA methods are presented as manhattan plots. SNP positions on the five *A. thaliana* chromosomes are indicated in Mbp. Their colors are altered to better visualize the chromosomal borders. Dashed lines represent the nominal 5% significance threshold with Bonferroni correction for 175655 tests for the GLM, MLM and vGWAS, and for 214051 tests for the WILCOXON GWA method.



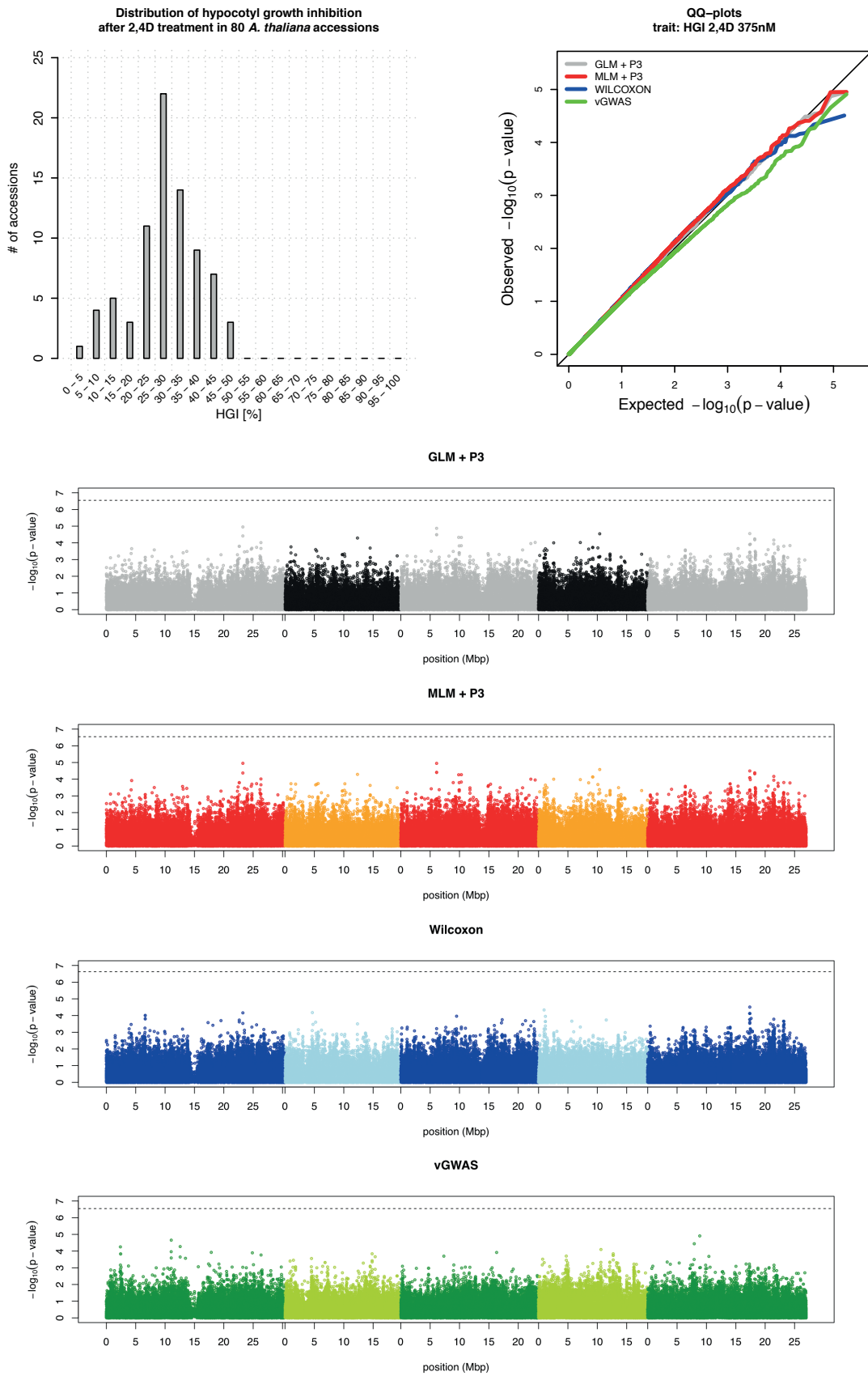
Supplementary Figure S23: GWA results overview for RGI after 20nM 2,4-D treatment.



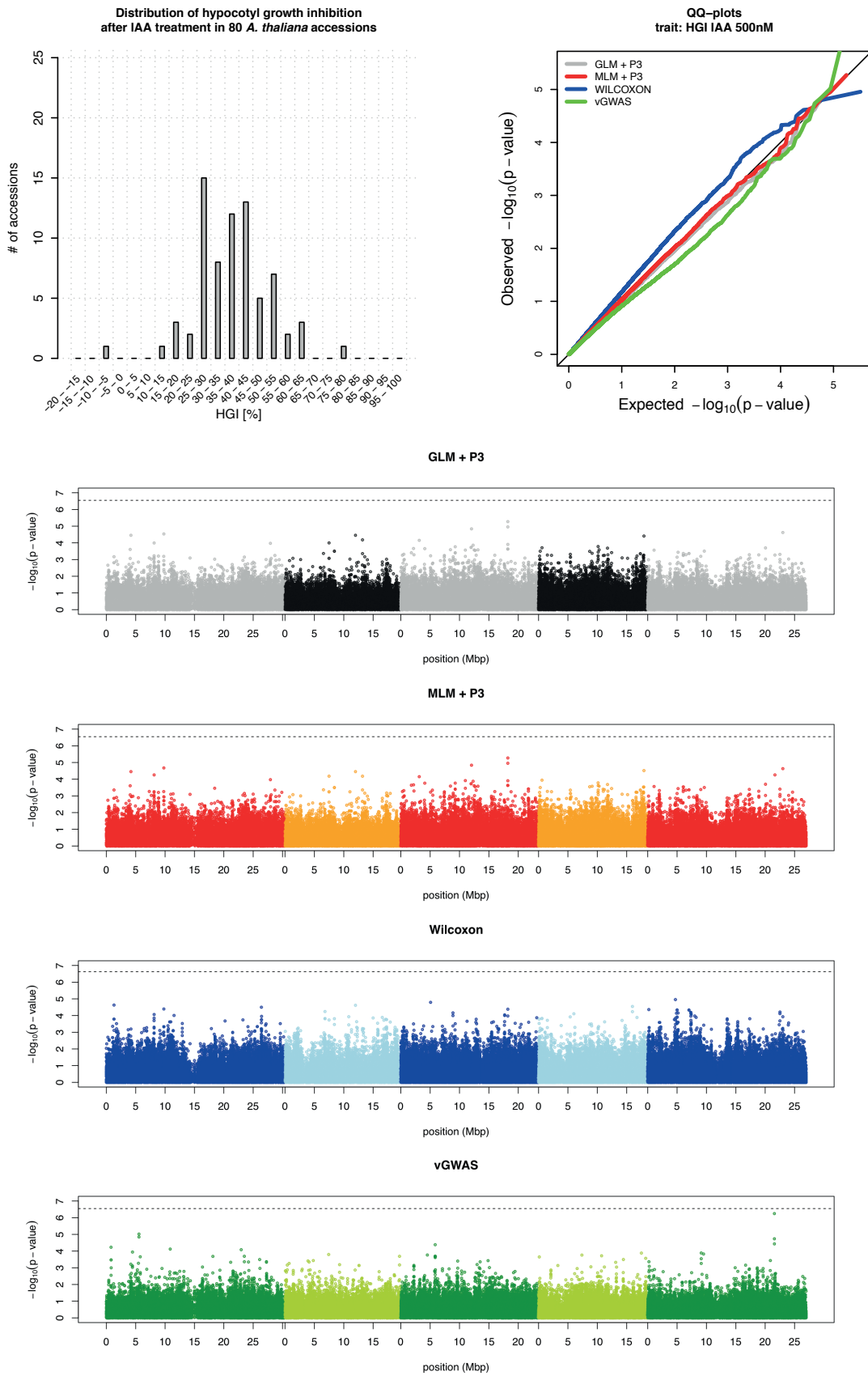
Supplementary Figure S24: GWA results overview for RGI after 40nM IAA treatment.



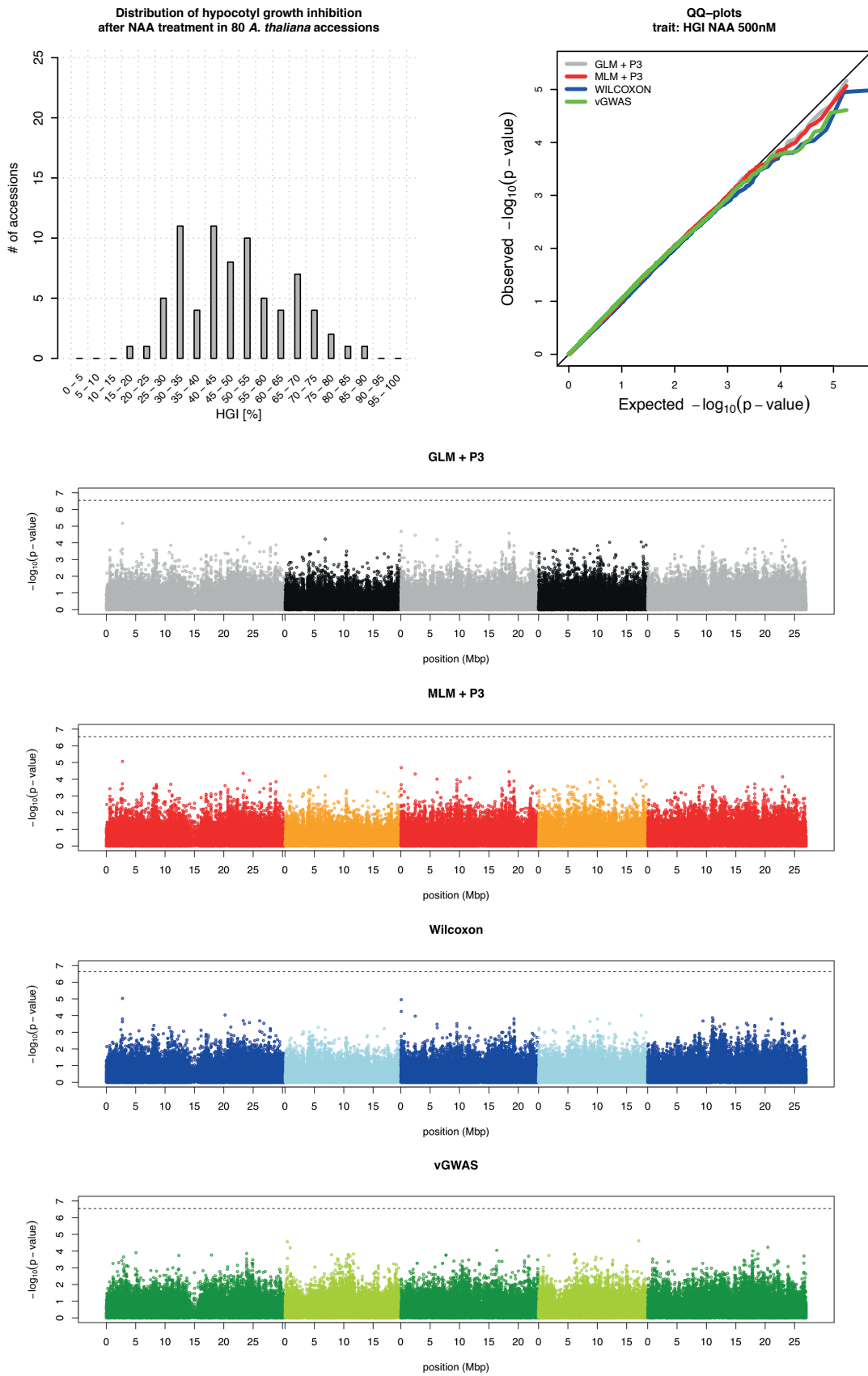
Supplementary Figure S25: GWA results overview for RGI after 75nM NAA treatment.



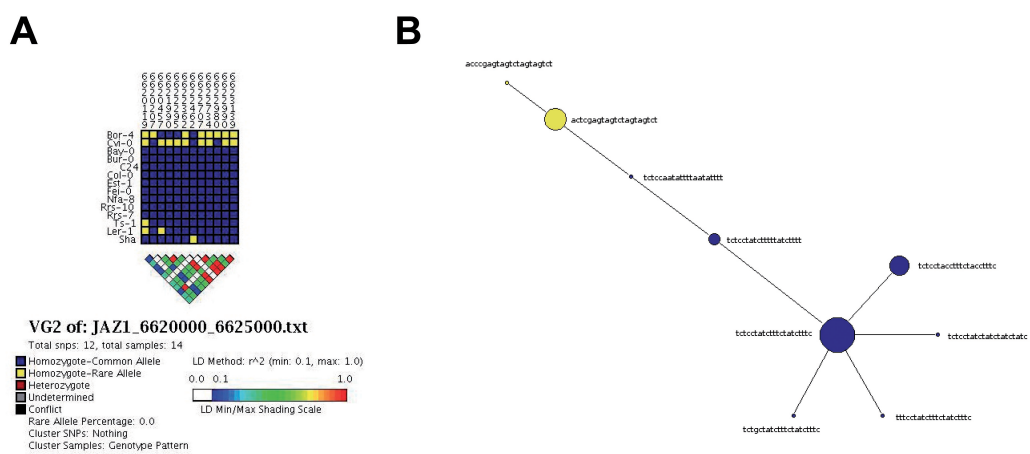
Supplementary Figure S26: GWA results overview for HGI after 375nM 2,4-D treatment.



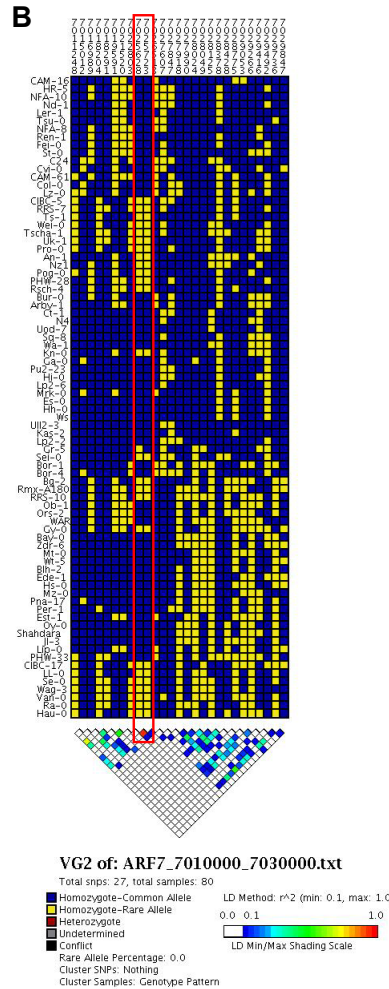
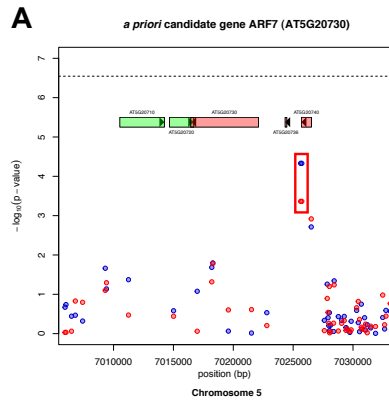
Supplementary Figure S27: GWA results overview for HGI after 500nM IAA treatment.



Supplementary Figure S28: GWA results overview for HGI after 500nM NAA treatment.



Supplementary Figure S29: Haplogroup representation of the *a priori* candidate gene *JAZ1*. (A) On top the SNP positions are shown for the CG. Each row represents the SNP composition in 14 *A. thaliana* accessions according to Clark et al. (2007). Minor alleles are shown in yellow, major alleles in blue. Accessions were clustered according to their haplotypes. At the bottom LDs calculated for pairwise SNP comparison are shown, blue colors indicate low, red colors high r^2 values. (B) Median joining network calculated according to Bandelt et al. (1999) for haplogroups based on CDS of *JAZ1* obtained for 80 *A. thaliana* accessions. Blue circle represents haplogroup, which consists of 69 accessions with thymidine at bp position 6622990 on chromosome one (Leucin), yellow circle represents eleven accession with guanine at this bp position (Valin).



Supplementary Figure S30: Detailed SNP structure analysis of the *a priori* candidate gene *ARF7*. Manhattan plots of GWA mapping results for the *a priori* CG *ARF7* and its genomic vicinity (A). Distribution of $-\log_{10}$ p-values calculated with MLM (red dots) or WILCOXON (blue dots) mapping methods. Annotated genes according to TAIR10 are plotted on top of the manhattan plot (arrowhead indicates gene orientation). (B) On top the SNP positions are shown for the CG, each row represents the SNP composition in the 80 *A. thaliana* accessions. Minor alleles are shown in yellow, major alleles in blue. Accessions were clustered according to their haplotypes. At the bottom LD calculated for pairwise SNP comparison are shown, blue colors indicate low, red colors high r^2 values. Colored boxes highlight same SNPs in manhattan and LD plots.

Supplementary tables

Supplementary Table ST1: Primer sequences used for genomic fragment sequencing.

AGI	ID	Forward strand	Reverse strand
AT3G62980	TIR1	GGTCTTGAGGTGCTTGCTTC	TACCCACCAGGATCTCTCA
AT4G03190	AFB1	CCTCGTTGAAGTTGAGCTACG	GGGAACAAATGGAAGGTGAAT
AT3G26810	AFB2	CGCCGCAGATAGTTGACTTAG	GAAGACGACCAATCAGAACCA
AT1G12820	AFB3	CCCAATTTGCCAAAACCTTA	CAATGGCTGAATTACGGTGA
AT4G14560	IAA1	ACACACACAAGCATTTTCAAGG	ATTTGTTTTTGCCTCGACCA
AT3G23030	IAA2	CAACCAGCTCACCAAGAACA	AAAACCCCGAAGTTTCGTCT
AT1G04240	IAA3	CCTCGAAGCCTCTCATCTTC	CAACCCAAGCACAGACAGAG
AT5G43700	IAA4	CAATCTTCTTCACAAAGCCTCTT	GAACCCGAGCAGTTTCAGAG
AT1G15580	IAA5	GGGCCGACAAGATAAATCAA	CCGAAGAGAAATATTGAGTAACTGC
AT1G52830	IAA6	AAGAAATAAATGGCAAAGGAAGG	TCAACGGTTTTCAAACTTTATGG
AT3G23050	IAA7	ACGGTGCTCCATATCTGAGG	CCCCTCATATTGTTGATCATTG
AT2G22670	IAA8	TGTTGTTGTGATGCTAAGCAGA	AAGCATCCAATCACCGTCTT
AT5G65670	IAA9	TTTGTGCTGTGTAATCATCAGG	AGCGCTTATACCAATCACCA
AT1G04100	IAA10	TGATCGGACTACCAAGCTGAA	GTCACAGACCCGAGGAACAT
AT4G28640	IAA11	GATTTTACAAGCCCTGAAGTG	TTTGTGCATTGCTCTGAACC
AT1G04550	IAA12	TTTCTCTGCTGGTATCATTGGA	GAAGCCTTTTCACCGAGTTG
AT2G33310	IAA13	TGCTTCCTTTCCAGGTTG	CGTAGCCTTTTCACCGAGTT
AT4G14550	IAA14	TCCCCTTTATCAAACCTCACC	TTGGTCATAATAATGAAAGCAAGAA
AT1G80390	IAA15	CACTTTCCGCACTTTTGTG	ATCCGATGCGTTTACGTTGT
AT3G04730	IAA16	ACGAAATTTCAAGGCACAAG	GAACACGAAATGTCGTCGAG
AT1G04250	IAA17	CCATCTTCTCATCACCTTCC	CGGTAATTCATTTGGGACTGA
AT1G51950	IAA18	TGTCGAAATCTTCTTCTTTTGG	ACTGCCATGGAACATCTCCT
AT3G15540	IAA19	TGTCTCCCCACAAAACCTGA	GGGGTGTGGTTTGGTTGTA
AT2G46990	IAA20	ATGGCGGCTCATTACTCATT	TTTTCTTGGCATTACTTCGATTT
AT3G16500	IAA26	GGGTTCTTAGGGTTTAGTGAAGC	TCAAGAATCTAATTCTGCAAAACA
AT4G29080	IAA27	AAATGCTTCAAAGTTTGTCCA	AAACAGTAAGAGAGCTTCACTGATCTA
AT5G25890	IAA28	CTTGCCAACCCCATATAA	TCAAAGCCAAACCCCATTAG
AT4G32280	IAA29	CAACACCAACGAACACAACC	AAATAATGGTCCGATTTGAACG
AT3G62100	IAA30	CACACGTTCAAATCATTAAATCACC	AACTAACTTTTTCCCCACCA
AT3G17600	IAA31	GAAGAGAAGGAAGCATGAAGC	TGCATCAATCAATTCTCCAA
AT2G01200	IAA32	TGAGCTCTCTTTTCTCATCTCAC	GCTTAAAGCTTCCATTACATCAGA
AT5G57420	IAA33	CATGTAAGCGTCCGTTGTCTT	TCGTACGTGAACAAGAAGCA
AT1G15050	IAA34	CAGGTTTTGACTTGAGACAATGA	TTCAAGGCTGACTTGGAAT
AT1G59750	ARF1	CCCAGAAGCTGTCTGTAGGC	AGAGCGAGCGATACAAGAGG
AT5G62000	ARF2	CCCAGAAGCTGTCTGTAGGC	AGAGCGAGCGATACAAGAGG
AT2G33860	ARF3	TTCAGGTAAGGTGGGACGAC	TCCAAAAGTTAGCAGCAAGTCA
AT5G60450	ARF4	ATGCGCTGGTGTAGTACTG	TGCAGGTGTCTCCACTGGTA
AT1G19850	ARF5	GCAGCTGGTGAACTGAATG	TGCCGCAGACTACAATCATC
AT1G30330	ARF6	GCAATCTCCGTCCTTAGTGC	AAACATTTCCGGAATCATCG
AT5G20730	ARF7	GGGCAGAATCAGCAAACACT	ACTGAGCCTCGTTTTTGCAC
AT5G37020	ARF8	GAAGGGGTGATTTGGGAAGT	GTAGCTGCTGAATCGGGAGA
AT4G23980	ARF9	AAGATGCGTTTTGAGGGAGA	CTGCACCTGTTTCAGCGATAA
AT2G28350	ARF10	AATGGCGTTTGAACAGAGG	CCTGAAAGGTTCTGCTCCAA
AT2G46530	ARF11	CAAACGGTCCAGACCAATCT	CAACAACGACTTCGACCTCA
AT1G34170	ARF13	GGTATTAGGCGAGCAAAGCA	TGCATGATTCAATTGGACAGA
AT4G30080	ARF16	ATCCGTTGGTCTCAGGTAT	CTGCGTATTTGATGGCTCCT
AT1G77850	ARF17	TGAGCGTCCGTGATGTAGAG	CCTCAAGGCGACTGATTACTG
AT3G61830	ARF18	CGATTGTGGGAAGTGGAGAT	CCCACACAACAGAACATTGC
AT1G19220	ARF19	TTCAGCAGCAACTGGAAATG	CCTCCACCATTATGATTCC

Supplementary Table ST2: Distribution of TAIR10 annotated splice donor and acceptor sites.

splice donor sites							
dimer	Chr1	Chr2	Chr3	Chr4	Chr5	SUM	%
AA	0	0	0	0	0	0	0.0000
AT	30	16	24	9	22	101	0.0626
AG	1	0	1	1	0	3	0.0019
AC	0	0	0	0	0	0	0.0000
TA	1	0	0	0	1	2	0.0012
TT	2	0	2	0	1	5	0.0031
TG	0	0	1	0	0	1	0.0006
TC	0	0	0	0	0	0	0.0000
GA	4	2	5	3	0	14	0.0087
GT	43624	23961	30316	24522	37143	159566	98.8429
GG	0	0	0	0	0	0	0.0000
GC	506	338	311	238	343	1736	1.0754
CA	1	0	0	1	0	2	0.0012
CT	2	0	1	0	0	3	0.0019
CG	0	1	0	0	0	1	0.0006
CC	0	0	0	0	0	0	0.0000
SUM	44171	24318	30661	24774	37510	161434	

splice acceptor sites							
dimer	Chr1	Chr2	Chr3	Chr4	Chr5	SUM	%
AA	0	2	2	0	0	4	0.0025
AT	3	0	3	0	2	8	0.0050
AG	44131	24301	30634	24763	37483	161312	99.9244
AC	27	14	20	8	22	91	0.0564
TA	1	0	0	0	0	1	0.0006
TT	0	1	0	0	1	2	0.0012
TG	2	0	0	2	1	5	0.0031
TC	0	0	0	0	1	1	0.0006
GA	0	0	0	0	0	0	0.0000
GT	0	0	1	0	0	1	0.0006
GG	4	0	0	0	0	4	0.0025
GC	1	0	0	0	0	1	0.0006
CA	0	0	0	0	0	0	0.0000
CT	2	0	0	0	0	2	0.0012
CG	0	0	1	1	0	2	0.0012
CC	0	0	0	0	0	0	0.0000
SUM	44171	24318	30661	24774	37510	161434	

Supplementary Table ST3: Major single nucleotide polymorphisms in the re-sequenced data for auxin network genes.

Gene name ^a	AGI ^b	Accessions	SNP position	SNP type	effect
ARF1	AT1G59750.4	Bay-0, Bur-0, Cvi-0, Est-1, Fei-0, Got-7, Ler-1, Lov-5, Nfa-8, Rrs-7, Sha, Tamm-2, Ts-1, Tsu-1	318	GT → CG	altered splice site
<u>ARF13</u>	AT1G34170.1	Bay-0 Tamm-2, Tsu-1	377-380 391	deletion CGA → TGA	frameshift altered STOP codon
<u>ARF13</u>	AT1G34170.2	Bay-0 Tamm-2, Tsu-1	377-380 391	deletion CGA → TGA	frameshift altered STOP codon
<u>ARF13</u>	<u>AT1G34170.3</u>	Bay-0 Tamm-2, Tsu-1	377-380 391	deletion CGA → TGA	frameshift altered STOP codon
<u>GH3.6</u>	<u>AT5G54510.1</u>	Bay-0, Sha, Ts-1	827	TAA → CAA	altered STOP codon
<u>IAA11</u>	<u>AT4G28640.2</u>	Bay-0, Bor-4, Br-0, Bur-0, C24, Cvi-0, Est-1, Fei-0, Got-7, Lov-5, Nfa-8, Rrs-7, Rrs-10, Sha, Tamm-2, Ts-1, Tsu-1	715	AG → AA	altered splice site
<u>ILR2</u>	<u>AT3G18485.1</u>	Bur-0, Got-7, Nfa-8 Bur-0, Got-7, Nfa-8 Cvi-0	622 623 450	deletion deletion deletion	frameshift altered splice site frameshift

^a Underlined gene names mark representative gene models, which were excluded from further analysis.

^b Underlined AGIs mark representative gene models according to TAIR10.

Supplementary Table ST4: Major single nucleotide polymorphisms in the MPICao2010 data for auxin network genes.

Gene name ^a	AGI ^b	#accessions affected	other splice variants
Anthranalite Synthase	<u>AT3G55870.1</u>	1	
ARF1	AT1G59750.4	70	(.1, .2, .3 not affected)
<u>ARF7</u>	<u>AT5G20730.1</u>	9	
ARF12	<u>AT1G34310.1</u>	1	
<u>ARF13</u>	<u>AT1G34170.1</u>	27	
<u>ARF13</u>	<u>AT1G34170.2</u>	24	
<u>ARF13</u>	<u>AT1G34170.3</u>	32	
<u>ARF14</u>	<u>AT1G35540.1</u>	13	
<u>ARF15</u>	<u>AT1G35520.1</u>	26	
<u>ARF16</u>	<u>AT4G30080.1</u>	6	
<u>ARF20</u>	<u>AT1G35240.1</u>	17	
<u>ARF21</u>	<u>AT1G34410.1</u>	6	
<u>ARF22</u>	<u>AT1G34390.1</u>	5	
<u>ARF23</u>	<u>AT1G43950.1</u>	77	
<u>ASA1</u>	<u>AT5G05730.2</u>	22	(.1 not affected)
<u>ASB2</u>	<u>AT5G57890.1</u>	4	
Gh3.2	<u>AT4G37390.1</u>	1	
Gh3.4	<u>AT1G59500.1</u>	2	
<u>Gh3.6</u>	<u>AT5G54510.1</u>	44	
IAA2	<u>AT3G23030.1</u>	1	
IAA6	<u>AT1G52830.1</u>	1	
<u>IAA8</u>	<u>AT2G22670.4</u>	38	(.1, .2, .3 not affected)
<u>IAA11</u>	<u>AT4G28640.2</u>	72	(.1, .3 not affected)
IAA18	<u>AT1G51950.1</u>	1	
IAA26	<u>AT3G16500.1</u>	2	
<u>IAA34</u>	<u>AT1G15050.1</u>	6	
ILL2	<u>AT5G56660.1</u>	2	
<u>ILL5</u>	<u>AT1G51780.1</u>	36	
<u>ILR2</u>	<u>AT3G18485.1</u>	11	
IGS2	<u>AT5G48220.1</u>	2	
IGS2	AT5G48220.2	2	
IGS2	AT5G48220.3	2	
LAX2	<u>AT2G21050.1</u>	1	
PAI3	AT1G29410.1	1	
PAI3	<u>AT1G29410.2</u>	2	
<u>PGP2</u>	<u>AT4G25960.1</u>	58	
PILS1	<u>AT1G20925.1</u>	1	
PILS4	<u>AT1G76530.1</u>	2	
<u>PIN4</u>	<u>AT2G01420.2</u>	3	(.1 not affected)
PIN7	AT1G23080.2	18	(.1, .3 not affected)
PIN8	<u>AT5G15100.1</u>	1	
<u>UGT74B1</u>	<u>AT1G24100.1</u>	7	
<u>UGT84B1</u>	<u>AT2G23260.1</u>	3	
YUC4	AT5G11320.2	1	(.1 not affected)
<u>YUC5</u>	<u>AT5G43890.1</u>	4	
YUC7	<u>AT2G33230.1</u>	2	
<u>YUC10</u>	<u>AT1G48910.1</u>	5	

^a Underlined gene names mark representative gene models, which were excluded from further analysis.

^b Underlined AGIs mark representative gene models according to TAIR10.

Supplementary Table ST5: Nucleotide diversity among different nucleotide categories within the re-sequenced data.

group	nucleotide category	#fragments (#genes)	average analyzed sites (in bp)	median π
auxin network genes	introns + flanking ^{1, m}	58 (58)	271.5	0.00344
	coding ^{3, m}	93 (93)	611.55	0.00160
	syn. ^{3, s}	93 (93)	143.01	0.00439
	nonsyn. ^{3, s}	93 (93)	468.54	0.00073
	total ^{5, m}	93 (93)	821.42	0.00197
auxin synthesis	introns + flanking ^{1, m}	12 (12)	290.75	0.00474
	coding ^{3, m}	21 (21)	577.14	0.00179
	syn. ^{3, s}	21 (21)	138.09	0.00547
	nonsyn. ^{3, s}	21 (21)	439.05	0.00075
	total ^{5, m}	21 (21)	802.33	0.00159
auxin signaling	introns + flanking ^{1, m}	32 (32)	275.5	0.00235
	coding ^{3, m}	47 (47)	605.55	0.00142
	syn. ^{3, s}	47 (47)	138.86	0.00326
	nonsyn. ^{3, s}	47 (47)	466.69	0.00089
	total ^{5, m}	47 (47)	826.64	0.00168
auxin metabolism	introns + flanking ^{1, m}	6 (6)	233	0.01617
	coding ^{3, m}	10 (10)	636.9	0.00204
	syn. ^{3, s}	10 (10)	149.39	0.00589
	nonsyn. ^{3, s}	10 (10)	487.51	0.00094
	total ^{5, m}	10 (10)	834.5	0.00224
auxin transport	introns + flanking ^{1, m}	8 (8)	216.5	0.00770
	coding ^{3, m}	15 (15)	661.6	0.00209
	syn. ^{3, s}	15 (15)	158.66	0.00811
	nonsyn. ^{3, s}	15 (15)	502.94	0.00037
	total ^{5, m}	15 (15)	823.07	0.00261
Nordborg2005 data	introns + flanking ^{1, m}	396 (393)	253.49	0.00317
	coding ^{3, m}	336 (335)	463	0.00148
	syn. ^{3, s}	336 (335)	107.26	0.00390
	nonsyn. ^{3, s}	336 (335)	355.74	0.00074
	total ^{5, m}	930 (918)	483.97	0.00242

¹ only fragments with at least 100 bp intron sites analyzed.

² only fragments with at least 360 bp exon sites analyzed.

³ only fragments with at least 120 codons analyzed.

⁴ only fragments with at least 100 bp silent sites analyzed.

⁵ only fragments with at least 200 bp sites analyzed.

^m DnaSP Ver. 5.10.01: MultiDomain Analysis.

^s DnaSP Ver. 5.10.01: Syn. and NonSyn. Substitutions.

^{*} Nucleotide diversity π (Nei 1987 equation 10.5).

Supplementary Table ST6: Summary statistics of within species population genetic parameters in auxin network gene groups within the MPICao2010 data.

group	auxin network genes	synthesis	signaling	metabolism	transport	TIR1/AFBs	AUX/IAAs	ARFs	PINs	PILSs	MPICao2010			
#rep. gene models	124	38	46	17	23	6	26	14	7	7	21201	2.5 per. ¹	med.	97.5 per. ¹
length(bp)^a														
introns + flanking	1064.50	1054.00	1084.50	901.00	1166.00	955.00	843.00	1840.50	1419.00	1116.00	0.00	878.00	3370.00	
coding	1320.00	1278.00	804.00	1407.00	1416.00	1768.50	582.00	2036.52	1857.00	1245.00	186.00	1053.00	3393.00	
syn.	293.10	277.24	175.84	316.12	311.10	393.96	123.20	445.04	412.90	274.33	38.00	225.25	722.79	
nonsyn.	1033.86	1001.24	637.94	1090.88	1109.32	1375.85	460.35	1591.48	1444.10	970.67	145.00	827.39	2666.19	
total	2371.00	2348.00	2303.00	2296.00	2544.00	2743.50	1399.00	4210.50	3335.00	2336.00	303.00	1983.00	6189.00	
length(aa)^b														
codons	441.00	427.00	269.00	470.00	473.00	590.50	195.00	680.00	620.00	416.00	63.00	352.00	1132.00	
S^c														
introns + flanking	35.0	41.0	29.5	32.0	36.0	24.0	22.5	46.0	48.0	26.0	0.0	27.0	133.0	
coding	23.5	27.5	18.0	40.0	24.0	25.5	11.5	29.0	28.0	18.0	2.0	20.0	88.0	
syn.	13.5	17.0	9.0	20.0	16.0	20.5	5.5	15.0	18.0	11.0	0.0	10.0	50.0	
nonsyn.	9.5	10.5	7.0	18.0	7.0	5.5	6.0	14.5	8.0	7.0	0.0	9.0	45.0	
total	61.0	66.0	50.0	80.0	58.0	48.5	34.5	75.5	82.0	46.0	6.0	50.0	192.0	
S per number of sites^d														
coding	0.01792	0.02113	0.01558	0.02945	0.01515	0.01462	0.01721	0.01347	0.01510	0.01515	0.00513	0.01979	0.06204	
syn.	0.04827	0.05746	0.04202	0.07056	0.04512	0.05256	0.04354	0.02890	0.04733	0.03756	0.00000	0.04489	0.15474	
nonsyn.	0.01016 (*)	0.01056	0.01025	0.01452	0.00451 (***)	0.00403 (*)	0.01226	0.00917	0.00605	0.00700	0.00000	0.01167	0.04730	
θ_n^e														
introns + flanking	0.00317	0.00421	0.00245	0.00479	0.00298	0.00177	0.00258	0.00199	0.00315	0.00283	0.00000	0.00323	0.01687	
coding	0.00183	0.00222	0.00140	0.00460	0.00146	0.00137	0.00153	0.00119	0.00126	0.00141	0.00026	0.00209	0.01262	
syn.	0.00593	0.00740	0.00363	0.01300	0.00476	0.00558	0.00397	0.00256	0.00476	0.00296	0.00000	0.00499	0.03732	
nonsyn.	0.00087 (*)	0.00109	0.00088 (*)	0.00157	0.00034 (**)	0.00021 (*)	0.00092	0.00089	0.00034	0.00073	0.00000	0.00104	0.00793	
total	0.00253	0.00296	0.00208	0.00581	0.00260	0.00159	0.00220	0.00146	0.00262	0.00197	0.00067	0.00285	0.01391	
promoter_500	0.00351	0.00416	0.00315	0.00480	0.00365	0.00233	0.00313	0.00402	0.00365	0.00335	0.00063	0.00388	0.01658	
promoter_1000	0.00407	0.00478	0.00368	0.00479	0.00326	0.00270	0.00389	0.00399	0.00287	0.00306	0.00087	0.00404	0.01545	
promoter_3000	0.00410	0.00468	0.00398	0.00510	0.00328	0.00372	0.00416	0.00347	0.00274	0.00328	0.00104	0.00405	0.01384	
θ_{ES}max^f	0.02974	0.04055	0.02360	0.03881	0.02328	0.03267	0.02838	0.01048 (*)	0.02253	0.02480	0.00000	0.02797	0.10560	
Tajima's D^g	-1.32662	-1.23327	-1.60919	-1.04638	-1.32396	-1.82794	-1.46773	-1.84605	-1.71914	-1.44917	-2.34012	-1.29027	1.44968	
Fu and Li's D^h	-1.56503	-1.29969	-1.98777	-1.06378	-1.50119	-2.54723	-1.74571	-2.55931	-0.62346	-1.90301	-4.67668	-1.49283	1.00749	
Fu and Li's F^h	-1.56176	-1.50974	-2.13821	-1.21734	-1.20618	-2.49735	-1.83919	-2.69519	-1.08346	-2.07055	-4.46387	-1.60338	1.02873	
Rminⁱ	1.0	2 (*)	0.0	2 (*)	2.0	0.5	0 (*)	1.5	2.0	1.0	0.0	1.0	12.0	
hapdiv^j	0.78560	0.79070	0.67848	0.87722	0.83038	0.75570	0.62294 (*)	0.85095 (*)	0.84258	0.81685	0.09778	0.74937	0.86994	
pv per codon^k	0.02339 (**)	0.02291	0.02609	0.03046	0.01299 (***)	0.01115 (**)	0.03381	0.02165	0.01401	0.01852	0.00615	0.02685	0.11364	

^a Median number of nucleotides analyzed within one group with the analysis software package, values in bp.

^b Median number of codons analyzed within one group, values in amino acids.

^c Median number of segregating sites found within one group with the polydNdS software.

^d Median number of segregating sites found divided by the number of analyzed sites within one group.

^e Median pairwise nucleotide diversity within one group calculated with polydNdS according to Tajima, F. (1983).

^f Median of the maximum value of synonymous nucleotide diversity values calculated for all possible pairwise comparisons for each gene within one group using a custom R-script.

^g Median Tajima's D neutrality test statistic calculated with compute according to Tajima, F. (1989) within one group.

^h Median Fu and Li's neutrality test statistic calculated with compute according to Fu and Li (1993) within one group.

ⁱ Median 4-Gamete test calculated with compute according to Hudson and Kaplan (1985) within one group.

^j Median haplotype diversity calculated with compute according to Depaulis and Veuille (1998) within one group.

^k Median protein variants divided by the analyzed codon length within one group using a custom R-script.

Asterisks denote significant differences between one group and the empirical null distribution (* P < 0.05; ** P < 0.01; *** P < 0.001) assessed by a two-sided Wilcoxon rank sum test Benjamini Hochberg corrected.

Supplementary Table ST7: Summary statistics of within species population genetic parameters in auxin network gene groups within the re-sequenced data.

group	auxin network genes	synthesis	signaling	metabolism	transport	TIR1/AFBs	AUX/IAAs	ARFs		Nordborg205	
#fragments (#genes)	93 (93)	21 (21)	47 (47)	10 (10)	15 (15)	4 (4)	28 (28)	15 (15)		336 (335)	
									2.5 per. ¹	med.	97.5 per. ¹
length(bp)^a											
coding	588.00	567.00	558.00	600.00	660.00	739.50	513.00	744.00	363.00	456.00	600.75
syn.	137.93	137.93	134.68	142.51	152.83	174.60	113.38	173.30	80.54	106.04	143.81
nonsyn.	450.07	428.57	432.51	457.49	507.17	564.90	395.71	570.70	278.32	352.67	463.35
total	828.00	836.00	817.00	855.50	870.00	804.50	795.50	837.00	397.38	510.00	623.25
length(aa)^b											
codons	196.00	189.00	186.00	200.00	220.00	246.50	171.00	248.00	121.00	152.00	200.25
S per number of sites^c											
coding	0.00846	0.00798	0.00761	0.00994	0.01099	0.00489	0.00971	0.00747	0.00000	0.00775	0.05274
syn.	0.02118	0.02530	0.01982	0.02460	0.03669	0.01628	0.02028	0.01982	0.00000	0.01937	0.13352
nonsyn.	0.00478	0.00354	0.00495	0.00639	0.00210	0.00177	0.00646	0.00491	0.00000	0.00446	0.03312
θ_{pi}^d											
coding	0.00160	0.00179	0.00142	0.00204	0.00209	0.00093	0.00175	0.00133	0.00000	0.00148	0.01523
syn.	0.00439	0.00547	0.00326	0.00589	0.00811	0.00308	0.00422	0.00214	0.00000	0.00390	0.04268
nonsyn.	0.00073	0.00075	0.00089	0.00094	0.00037	0.00024	0.00101	0.00109	0.00000	0.00074	0.00916
θ_{nsmax}^e	0.01280	0.01210	0.01190	0.01880	0.01910	0.01285	0.00865	0.01220	0.00000	0.01650	0.12261
Tajima's D^f											
coding	-0.86819	-0.61820	-1.14576	-0.73968	-0.62583	-1.38463	-1.16480	-0.90185	-2.16280	-0.94972	2.05284
syn.	-0.70350	-0.49607	-0.99174	-0.24042	-0.44690	-1.00270	-0.72914	-1.01834	-2.10998	-0.56216	2.17149
nonsyn.	-1.11995	-0.96788	-1.14238	-1.29402	-1.14191	-1.16480	-1.16480	-0.99407	-1.87119	-1.09212	1.64017
hapdiv^g	0.614 (***)	0.60800	0.544 (*)	0.62950	0.725 (**)	0.38000	0.56150	0.614 (*)	0.00000	0.46200	0.86313
pv per codon^h	0.01471	0.01351	0.01622	0.01738	0.00943 (*)	0.00812	0.01952	0.01613	0.00562	0.01587	0.04508

^a Median number of nucleotides analyzed within one group with DnaSP Ver. 5.10.01, values in bp.

^b Median number of codons analyzed within one group, values in amino acids.

^c Median number of segregating sites found divided by the number of analyzed sites within one group.

^d Median pairwise nucleotide diversity within one group calculated with DnaSP Ver. 5.10.01.

^e Median of the maximum value of synonymous nucleotide diversity values calculated for all possible pairwise comparisons for each gene within one group using DnaSP Ver. 5.10.01.

^f Median Tajima's D neutrality test statistic calculated with DnaSP Ver. 5.10.01 within one group for either all coding sites, synonymous or nonsynonymous sites.

^g Median haplotype diversity calculated with DnaSP Ver. 5.10.01.

^h Median protein variants divided by the analyzed codon length within one group using a custom R-script.

Asterisks denote significant differences between one group and the empirical null distribution (* P < 0.05; ** P < 0.01; *** P < 0.001) assessed by a two-sided Wilcoxon rank sum test Benjamini Hochberg corrected.

Supplementary Table ST8: McDonald-Kreitman test between *A. thaliana* and three Brassicaceae species.

<i>A. thaliana</i> vs. <i>A. lyrata</i>													
species_A_ID	species_B_ID	gene name	blastp_evalue_A_B	reciprocal_best_hit_A_B ^a	K _{NSYN}	K _{SYN}	K _{NSYN} /K _{SYN} - ratio	N _F	S _F	N _P	S _P	fisher.test corrected p-value ^b	
AT1G20925.1 PACId:19655523	313001 PACId:16056219	PILS1	0e+00	1	0.00977	0.09539	0.10118	8	26	17	7	0.0271	
AT1G23320.1 PACId:19655437	335577 PACId:16052925	TAR1	0e+00	1	0.03007	0.13938	0.21593	23	30	20	3	0.0248	
AT1G76520.1 PACId:19649934	895536 PACId:16062586	PILS3	0e+00	0	0.00765	0.22607	0.03388	6	54	8	5	0.0163	
AT2G04400.1 PACId:19643382	480086 PACId:16043332	IGS1	0e+00	1	0.01483	0.11395	0.13070	6	27	23	16	0.0319	
AT2G20610.1 PACId:19643554	480895 PACId:16036602	SUR1	0e+00	1	0.00675	0.08405	0.08038	4	30	13	13	0.0499	
AT2G23170.1 PACId:19638788	481188 PACId:16035757	GH3.3	0e+00	1	0.01757	0.12162	0.14449	21	52	17	8	0.0348	
AT2G46530.3 PACId:19640017	904120 PACId:16045535	ARF11	0e+00	0	0.02908	0.16109	0.18054	35	55	28	10	0.0255	
AT2G47000.1 PACId:19643294	483871 PACId:16048231	PGP4	0e+00	1	0.00730	0.15350	0.04757	20	140	10	12	0.0306	
AT3G28830.1 PACId:19660318	484459 PACId:16064767	CYP71B15	0e+00	1	0.07548	0.09121	0.82818	69	25	9	20	0.0094	
AT3G44310.1 PACId:19664832	323132 PACId:16053666	NIT1	0e+00	0	0.06429	0.28303	0.22771	44	49	2	28	0.0062	
<i>A. thaliana</i> vs. <i>B. rapa</i>													
species_A_ID	species_C_ID	gene name	blastp_evalue_A_C	reciprocal_best_hit_A_C ^a	K _{NSYN}	K _{SYN}	K _{NSYN} /K _{SYN} - ratio	N _F	S _F	N _P	S _P	fisher.test corrected p-value ^b	
AT1G59500.1 PACId:19655856	Bra041046 PACId:22724500	GH3.4	0e+00	0	0.06855	0.64906	0.10304	64	142	89	66	0.0006	
AT1G59750.1 PACId:19658021	Bra035427 PACId:22704834	ARF1	0e+00	0	0.02816	0.28663	0.09824	35	106	28	23	0.0170	
AT2G23170.1 PACId:19638788	Bra039186 PACId:22712889	GH3.3	0e+00	1	0.04686	0.46680	0.10038	54	148	17	8	0.0123	
AT2G46530.3 PACId:19640017	Bra004530 PACId:22685326	ARF11	0e+00	0	0.06900	0.41095	0.16792	81	120	26	9	0.0173	
AT3G44310.1 PACId:19664832	Bra028932 PACId:22709042	NIT1	4e-172	0	0.09916	0.44427	0.22345	66	72	2	28	0.0040	
AT3G55870.1 PACId:19659144	Bra009130 PACId:22692233	AnthranilateSynthase	0e+00	0	0.15219	4.35778	0.03449	137	247	31	18	0.0196	
AT4G27070.1 PACId:19646872	Bra019075 PACId:22707183	TSB2	0e+00	1	0.02486	0.54275	0.04585	22	126	16	19	0.0193	
<i>A. thaliana</i> vs. <i>T. parvula</i>													
species_A_ID	species_D_ID	gene name	blastp_evalue_A_D	reciprocal_best_hit_A_D ^a	K _{NSYN}	K _{SYN}	K _{NSYN} /K _{SYN} - ratio	N _F	S _F	N _P	S _P	fisher.test corrected p-value ^b	
AT1G23320.1 PACId:19655437	Trp1g20760	TAR1	0e+00	1	0.05788	0.41475	0.13961	44	73	20	3	0.00501	
AT1G59500.1 PACId:19655856	Trp7g35050	GH3.4	0e+00	0	0.07076	0.38100	0.18109	63	100	92	67	0.04140	
AT2G23170.1 PACId:19638788	Trp4g02220	GH3.3	0e+00	1	0.03310	0.31774	0.10417	38	114	17	8	0.00832	
AT2G46530.3 PACId:19640017	Trp4g28570	ARF11	0e+00	0	0.07309	0.39530	0.18490	87	117	28	10	0.03551	
AT3G44310.1 PACId:19664832	Trp5g16960	NIT1	4e-173	0	0.10379	0.45564	0.22805	72	73	3	28	0.00740	
AT3G55870.1 PACId:19659144	Trp6g37170	AnthranilateSynthase	0e+00	0	0.15771	4.32504	0.03601	145	246	31	18	0.03440	
AT4G27070.1 PACId:19646872	Trp6g16750	TSB2	0e+00	1	0.03328	0.66130	0.05034	29	148	15	20	0.04755	
AT5G49980.1 PACId:19669305	Trp2g15950	AFB5	8e-169	0	0.41513	4.50231	0.09220	375	356	2	16	0.03460	
AT5G4680.1 PACId:19665574	Trp6g16640	ILR3	1e-133	1	0.01234	0.28108	0.04395	6	44	8	6	0.04814	

^a Indicates, if the best blastp hit is also the reciprocal best hit.

^b P-values calculated on a 2 x 2 table, if the (N_P/S_P) ratio is significant different from the (N_F/S_F) ratio. P-values were corrected for multiple testing with Benjamini Hochberg on all conducted tests.

Supplementary Table ST10: Main effect QTL analysis in the RIL $Q_{Col-0 \times C24}$ and $R_{C24 \times Col-0}$ populations for auxin response traits.

Trait	Treatment	RIL population	Method ^a	Model ^b	Covariates	Chr	Position ^c	95% CI ^d	LOD	pvalue
RGI	2,4-D 20nM	$Q_{Col-0 \times C24}$	hk	normal	none	3	74.9	71 - 77	4.92	0
			cim	normal	3	3	74.9	71 - 77	5.48	0.001
		$R_{C24 \times Col-0}$	hk	normal	none	3	72	17 - 77	4.13	0.001
			cim	normal	3	3	72	65 - 73	4.03	0.023
RGI	IAA 40nM	$Q_{Col-0 \times C24}$	hk	normal	none	2	61	54 - 73.9	2.96	0.023
			cim	normal	3	There were no LOD peaks above the threshold.				
		$R_{C24 \times Col-0}$	hk	normal	none	2	52.3	13 - 73.9	2.86	0.026
						5	4	0 - 26	2.77	0.03
			cim	normal	3	There were no LOD peaks above the threshold.				
			-	np	none	2	52.3	12 - 73.9	2.71	0.031
			5	4	0 - 16	3.15	0.014			
RGI	NAA 75nM	$Q_{Col-0 \times C24}$	hk	normal	none	4	62	56.4 - 71	4.14	0.001
						5	77	74 - 84	13.63	0
			cim	normal	3	5	77	74 - 78	11.9	0
		$R_{C24 \times Col-0}$	hk	normal	none	1	54	45 - 73	2.91	0.012
						2	39	33 - 64	2.7	0.025
						4	61	0 - 78.1	2.51	0.041
						5	68	55 - 92	4.24	0
			cim	normal	3	1	53	53 - 60	3.34	0.041
						5	69	63 - 71	4.8	0.006
HGI	2,4-D 375nM	$Q_{Col-0 \times C24}$	hk	normal	none	1	24	17 - 27	4.63	0.001
						4	44.4	12 - 78.1	2.7	0.042
						5	18	5 - 35	3.55	0.01
			cim	normal	3	1	24	19 - 26	5.54	0.001
						4	43	40 - 47.1	4.23	0.009
		$R_{C24 \times Col-0}$	hk	normal	none	1	12.8	0 - 29	2.82	0.033
						2	48	43 - 59	3.23	0.012
						5	43.8	4 - 63	2.75	0.04
			cim	normal	3	2	48	44 - 51	4.46	0.006
HGI	IAA 500nM	$Q_{Col-0 \times C24}$	hk	normal	none	4	55	52 - 59	6.02	0
						5	76	10 - 93.2	2.58	0.048
			cim	normal	3	4	55	53 - 59	6.16	0.001
			-	np	none	4	55	52 - 59	6.14	0
		$R_{C24 \times Col-0}$				5	76	10 - 93	2.46	0.045
			hk	normal	none	3	0	0 - 83	3.05	0.015
			cim	normal	3	5	65.5	20 - 70	3.67	0.04
			-	np	none	3	0	0 - 86.8	2.73	0.011
HGI	NAA 500nM	$Q_{Col-0 \times C24}$	hk	normal	none	2	36	33 - 46	7.83	0
						5	55	52 - 83	3.89	0.003
			cim	normal	3	2	42	38 - 46	8.96	0
		$R_{C24 \times Col-0}$				5	56	52 - 58	5.62	0
			hk	normal	none	2	36	32 - 40	5.22	0
			cim	normal	3	2	36	33 - 38	5.96	0.001
			-	np	none	2	35	31 - 40	4.42	0.002
						3	50.7	32 - 86.8	2.67	0.025

^a Interval mapping with Haley-Knott regression (hk); Composite interval mapping with Haley-Knott regression (cim).

^b Nonparametric QTL mapping model (np).

^c Positions in centimorgan (cM).

^d 95% Bayes credible interval range from low to high positions in centimorgan (cM).

Supplementary Table ST11: Two-QTL genome scan analysis in the RIL $Q_{Col-0} \times C_{24}$ and $R_{C_{24} \times Col-0}$ populations for auxin response traits.

Trait	Treatment	RIL population	10000 permutations 5% thresholds ^a				Summary scantwo				pos1 ^c	pos2 ^c	pos1a ^c	pos2a ^c	lod.add	pval ^d	lod.av1	pval ^d					
			full	fv1	int	add	av1	one	interaction ^b	pos1 ^c									pos2 ^c	lod.full	lod.fv1	lod.int	lod.add
RGI	2,4-D 20nM	$Q_{Col-0} \times C_{24}$	5.357	3.986	3.381	4.201	2.388	2.548	-	-	There were no pairs of loci meeting the criteria.				4	5.620	0.004	2.759	0.023				
			5.352	4.011	3.403	4.168	2.342	2.531	-	-	There were no pairs of loci meeting the criteria.				24	5.108	0.011	2.769	0.018				
			5.329	3.998	3.378	4.175	2.388	2.530	-	-	There were no pairs of loci meeting the criteria.				77	16.068	0.000	2.439	0.041				
			5.410	4.099	3.398	4.194	2.436	2.498	c2:c5	51	2	7.240	0.002	4.379	0.029	1.620	0.846	51	4	5.620	0.004	2.759	0.023
			5.356	4.001	3.381	4.216	2.371	2.556	c1:c3	90	24	5.130	0.076	2.791	0.468	0.022	1.000	90	24	5.108	0.011	2.769	0.018
RGI	IAA 40nM	$R_{C_{24} \times Col-0}$	5.334	3.989	3.359	4.154	2.373	2.528	c1:c5	93	77	16.410	0.000	2.781	0.475	0.342	1.000	92	77	16.068	0.000	2.439	0.041
			5.329	3.998	3.378	4.175	2.388	2.530	c1:c2	55	38	6.073	0.013	3.162	0.258	0.022	1.000	54	38	6.051	0.001	3.140	0.007
			5.410	4.099	3.398	4.194	2.436	2.498	c1:c4	53	61	5.612	0.031	2.701	0.539	0.139	1.000	53	61	5.472	0.004	2.562	0.030
			5.356	4.001	3.381	4.216	2.371	2.556	c1:c5	53	68	7.951	0.001	3.713	0.088	0.084	1.000	53	68	7.867	0.000	3.628	0.003
			5.329	3.998	3.378	4.175	2.388	2.530	c2:c4	39	62	5.482	0.038	2.783	0.480	0.035	1.000	39	62	5.447	0.004	2.747	0.019
HGI	2,4-D 375nM	$Q_{Col-0} \times C_{24}$	5.265	3.967	3.342	4.167	2.349	2.534	c1:c4	24	44	7.848	0.000	3.222	0.223	0.002	1.000	24	44	7.847	0.000	3.221	0.005
			5.329	3.998	3.378	4.175	2.388	2.530	c1:c5	23	18	7.897	0.000	3.271	0.204	0.315	1.000	24	18	7.582	0.000	2.956	0.009
			5.410	4.099	3.398	4.194	2.436	2.498	c3:c4	7	45	5.299	0.047	2.614	0.603	0.085	1.000	7	44	5.214	0.007	2.528	0.030
			5.356	4.001	3.381	4.216	2.371	2.556	c4:c5	46	32	6.564	0.005	3.013	0.324	0.409	1.000	45	18	6.154	0.002	2.603	0.024
			5.338	3.974	3.354	4.166	2.388	2.530	c1:c2	13	47	6.741	0.004	3.508	0.134	0.058	1.000	13	47	6.683	0.000	3.451	0.003
HGI	IAA 500nM	$R_{C_{24} \times Col-0}$	5.403	4.075	3.465	4.226	2.393	2.535	c1:c5	13	57	5.668	0.028	2.861	0.426	0.328	1.000	13	44	5.340	0.006	2.534	0.033
			5.418	4.022	3.450	4.238	2.302	2.564	c2:c5	48	6	6.493	0.006	3.260	0.212	0.491	1.000	47	44	6.002	0.002	2.769	0.018
			5.325	4.017	3.412	4.156	2.344	2.502	c3:c4	39	55	8.571	0.000	2.555	0.667	0.017	1.000	39	55	8.555	0.000	2.538	0.035
			5.418	4.022	3.450	4.238	2.302	2.564	c3:c5	0	65	5.664	0.031	2.618	0.610	0.067	1.000	0	65	5.598	0.004	2.551	0.024
			5.325	4.017	3.412	4.156	2.344	2.502	c2:c5	42	56	12.688	0.000	4.856	0.008	0.001	1.000	42	56	12.687	0.000	4.855	0.000
HGI	NAA 500nM	$Q_{Col-0} \times C_{24}$	5.426	4.055	3.367	4.201	2.426	2.533	c3:c3	22	33	4.524	0.200	3.048	0.309	0.001	1.000	22	33	4.522	0.024	3.046	0.007
			5.329	3.998	3.378	4.175	2.388	2.530	c2:c4	35	51	7.845	0.001	2.623	0.621	0.141	1.000	35	51	7.705	0.000	2.483	0.042
			5.410	4.099	3.398	4.194	2.436	2.498	c3:c4	51	51	6.532	0.006	4.325	0.030	2.089	0.520	51	51	4.443	0.031	2.236	0.078
			5.356	4.001	3.381	4.216	2.371	2.556															
			5.329	3.998	3.378	4.175	2.388	2.530															

^a LOD score thresholds calculated for the different QTL scantwo models: full (q1 + q2 + q1 x q2); fv1 (q2 + q1 x q2); int (q1 x q2); add (q1 + q2); one (q1).

^b Interaction between chromosomes is indicated between first and second, also includes interactions on a single chromosome.

^c Position of tested loci in centimorgan (cM) corresponding to the chromosome mentioned in the interaction column. The evaluated positions can differ between full (pos f) and additive (pos a) QTL model.

^d P-values calculated based on the permutations.

Supplementary Table ST12: Possible SNP effects in *a priori* auxin related candidate genes within markers III_17 and III_19 on chromosome three.

Locus ^a	Gene name	Chr	Position ^b	SNP alleles ^c	SNP effect ^d
MASC09224	III_17	3	18817123		
AT3G51060	SRS1	3	18964896	C/T	synonymous
AT3G51200	SAUR-like	3	19019044	A/G	non-synonymous
AT3G51840	ACX4	-			
AT3G52400	SYP122	3	19427001	G/C	non-synonymous
		3	19426826	C/T	synonymous
		3	19426723	A/C	non-synonymous
		3	19426676	T/C	synonymous
		3	19426576	T/C	non-synonymous
		3	19426210	C/T	synonymous
		3	19426102	G/A	synonymous
		3	19426050	G/T	synonymous
AT3G53020	RPL24	3	19660848	C/A	synonymous
AT3G53250	SAUR-like	-			
AT3G53480	PDR9	3	19825470	C/T	synonymous
		3	19825573	G/A	non-synonymous
		3	19831003	T/G	non-synonymous
		3	19831492	C/G	non-synonymous
MASC03218	III_18	3	20180613		
AT3G55120	TT5	-			
AT3G55320	PGP20	3	20510674	A/T	synonymous
		3	20509673	C/T	non-synonymous
AT3G55730	MYB109	-			
AT3G56280	pseudogene				
AT3G56370	LRR-kinase family	3	20902349	T/C	synonymous
		3	20902208	T/A	synonymous
		3	20902175	A/G	synonymous
		3	20902166	A/G	synonymous
		3	20902127	G/C	synonymous
		3	20901517	A/C	non-synonymous
		3	20901008	C/G	non-synonymous
MASC02788	III_19	3	21060362		

^a Locus name or marker name.

^b Position in bp according to TAIR10.

^c SNP alleles between Col-0 or C24 *A. thaliana* accessions. Bold: unique allele for C24.

^d Indicates possible SNP effect on protein sequence based on TAIR10 annotation.

Supplementary Table ST13: Possible SNP effects in *a priori* auxin related candidate genes within markers V_20 and V_22 on chromosome five.

Locus ^a	Gene name	Chr	Position ^b	SNP alleles ^c	SNP effect ^d
MASC04591	V_20	5	21429747		
AT5G53590	SAUR-like	5	21772259	A/T	synonymous
AT5G54140	ILL3	5	21966352	G/A	non-synonymous
		5	21967105	C/G	non-synonymous
		5	21967474	A/T	non-synonymous
AT5G54490	PBP1	5	22121471	A/C	non-synonymous
		5	22121577	C/T	synonymous
AT5G54500	FQR1	-			
AT5G54510	DFL1	-			
AT5G55120	VTC5	-			
MASC04394	V_21	5	22403190		
AT5G55250	IAMT1	5	22409968	A/G	synonymous
AT5G55540	TRN1	5	22499545	A/G	synonymous
		5	22499086	A/T	non-synonymous
		5	22498662	T/C	non-synonymous
		5	22498621	T/C	synonymous
		5	22498360	C/T	synonymous
		5	22498171	A/G	synonymous
		5	22497134	G/T	non-synonymous
AT5G55910	D6PK	5	22641486	T/C	non-synonymous
AT5G56010	AtHSP90.3	5	22683446	G/A	synonymous
		5	22683458	T/A	synonymous
AT5G56030	ERD8	5	22687476	T/C	synonymous
		5	22688449	G/A	synonymous
		5	22689058	C/T	synonymous
		5	22689182	T/C	synonymous
		5	22689310	C/T	synonymous
AT5G56290	EMB2790	5	22787102	C/G	non-synonymous
AT5G56650	ILL1	5	22930897	A/G	non-synonymous
		5	22931404	C/A	non-synonymous
		5	22932098	A/C	synonymous
		5	22932149	T/C	synonymous
AT5G56660	ILL2	5	22933482	C/T	non-synonymous
		5	22933544	T/G	synonymous
		5	22933836	G/C	non-synonymous
		5	22933850	T/C	synonymous
		5	22934121	C/T	synonymous
		5	22934169	T/C	synonymous
		5	22934265	C/T	synonymous
AT5G56750	NDL1	5	22958980	A/G	synonymous
		5	22960042	C/T	synonymous
		5	22960495	C/A	synonymous
AT5G57090	AGR1	5	23100791	T/C	synonymous
		5	23103022	T/C	synonymous
AT5G57420	IAA33	-			
AT5G57560	TCH4	-			
AT5G57740	XBAT32	5	23396528	A/C	non-synonymous
		5	23395234	A/G	non-synonymous
		5	23395048	T/C	non-synonymous
MASC01545	V_22	5	23417207		

^a Locus name or marker name.

^b Position in bp according to TAIR10.

^c SNP alleles between Col-0 or C24 *A. thaliana* accessions. Bold: unique allele for C24.

^d Indicates possible SNP effect on protein sequence based on TAIR10 annotation.

Supplementary Table ST14: Possible SNP effects in *a priori* auxin related candidate genes within markers II_10 and II_12 on chromosome two.

Locus ^a	Gene name	Chr	Position ^b	SNP alleles ^c	SNP effect ^d
MASC02600	II_10	2	9532715		
AT2G22670	IAA8	2	9637168	T/G	non-synonymous
		2	9637180	G/A	synonymous
		2	9637221	A/C	non-synonymous
		2	9637513	T/C	synonymous
		2	9637591	T/C	synonymous
AT2G22810	ACC4	2	9719426	A/T	non-synonymous
		2	9718706	C/A	synonymous
AT2G23050	NPY4	2	9811998	C/G	synonymous
AT2G23170	GH3.3	2	9866276	A/G	non-synonymous
		2	9864556	C/T	non-synonymous
AT2G23560	MES7	-			
AT2G23580	MES4	-			
AT2G23590	MES8	-			
AT2G23600	MES2	2	10043312	T/A	synonymous
		2	10042586	A/G	synonymous
		2	10042460	G/T	non-synonymous
		2	10042429	C/G	non-synonymous
AT2G23610	MES3	-			
AT2G23620	MES1	2	10049104	T/A	non-synonymous
		2	10047637	A/G	non-synonymous
AT2G24400	SAUR-like	2	10378263	A/G	synonymous
		2	10378098	T/C	synonymous
MASC02644	II_11	2	10577823		
AT2G24850	TAT3	-			
AT2G25170	SSL2	2	10715874	A/G	synonymous
		2	10716128	G/A	synonymous
		2	10720591	T/A	synonymous
		2	10722795	C/T	non-synonymous
		2	10723433	G/T	synonymous
AT2G25790	LRR-kinase family	2	11000915	C/G	synonymous
		2	11001129	G/T	non-synonymous
		2	11002687	A/G	non-synonymous
AT2G25930	ELF3	2	11060163	G/A	non-synonymous
		2	11060537	G/C	synonymous
		2	11062985	C/T	non-synonymous
		2	11063124	A/G	synonymous
AT2G26170	MAX1	2	11142609	G/A	non-synonymous
		2	11143000	A/G	synonymous
		2	11143240	G/T	synonymous
AT2G26710	CYP72B1	-			
AT2G26730	LRR-kinase family	2	11390644	G/A	synonymous
AT2G26740	SEH	-			
AT2G27690	CYP94C1	2	11809588	C/A	synonymous
AT2G28085	SAUR-like	-			
AT2G28350	ARF10	2	12114609	C/G	synonymous
		2	12115510	G/A	synonymous
		2	12115643	C/G	non-synonymous
		2	12116206	C/A	synonymous
AT2G28780	unknown protein	-			
AT2G29450	TAU1	2	12625060	A/T	non-synonymous
		2	12624999	C/T	synonymous
MASC09221	II_12	2	11573651		

^a Locus name or marker name.

^b Position in bp according to TAIR10.

^c SNP alleles between Col-0 or C24 *A. thaliana* accessions. Bold: unique allele for C24.

^d Indicates possible SNP effect on protein sequence based on TAIR10 annotation.

Supplementary Table ST15: Possible SNP effects in *a priori* auxin related candidate genes from marker V_24 to the end of chromosome five.

Locus ^a	Gene name	Chr	Position ^b	SNP alleles ^c	SNP effect ^d
MASC09211	V_24	5	25596187		
AT5G64770	RGF3	-			
AT5G64890	PROPEP2	5	25935234	G/A	synonymous
		5	25935307	G/A	non-synonymous
AT5G65670	IAA9	-			
AT5G65940	CHY1	5	26378779	A/G	synonymous
AT5G65980		5	26393054	G/T	synonymous
AT5G66260	SAUR-like	-			
AT5G66700	ATHB53	-			
AT5G67250	VFB4	-			
AT5G67300	ATMYB44	-			
MASC04350	V_25	5	26886066		
AT5G67440	NPY3	5	26914396	T/C	non-synonymous
AT5G67480	ATBT4	5	26932292	C/T	non-synonymous
AT5G67580	TBP3	5	26956070	T/C	non-synonymous

^a Locus name or marker name.

^b Position in bp according to TAIR10.

^c SNP alleles between Col-0 or C24 *A. thaliana* accessions. Bold: unique allele for C24.

^d Indicates possible SNP effect on protein sequence based on TAIR10 annotation.

Supplementary Table ST16: SNP enrichment analysis of *a priori* candidate genes related to auxin detected under different top ranked associated SNP categories with MLM and WILCOXON GWA mapping method.

Number of Top SNPs	RGI 2,4-D		RGI IAA		RGI NAA			
	MLM ^a	WILCOXON ^b	MLM ^a	WILCOXON ^b	MLM ^a	WILCOXON ^b		
50	2.935	**	3.261	**	0.978	0.652	1.304	1.304
100	2.283	**	1.956		1.304	1.304	1.141	1.304
200	1.875	**	1.793	*	1.304	1.630	1.223	1.141
300	1.685	**	1.848	**	1.304	1.413	1.196	1.087
400	1.712	**	1.834	**	1.182	1.223	1.182	1.060
500	1.467	*	1.696	**	1.043	1.141	1.239	1.109
1000	1.370	**	1.581	***	1.109	1.076	1.109	0.978
2000	1.296	**	1.353	**	1.158	1.084	1.052	1.003
3000	1.288	**	1.234	**	1.076	1.054	1.087	1.033
175677 214051	1.000		1.000		1.000	1.000	1.000	1.000

	HGI 2,4-D		HGI IAA		HGI NAA	
	MLM ^a	WILCOXON ^b	MLM ^a	WILCOXON ^b	MLM ^a	WILCOXON ^b
50	1.304	1.630	0.000	2.283	0.326	1.304
100	0.815	0.978	1.141	1.630	0.815	0.815
200	0.571	1.060	0.897	1.549	0.571	0.734
300	0.815	0.924	1.359	1.467	0.598	0.543
400	0.815	0.856	1.304	1.386	0.489	* 0.489 *
500	1.011	0.848	1.109	1.206	0.489	* 0.554 *
1000	0.962	0.799	1.076	1.043	0.505	** 0.668 *
2000	0.856	1.019	1.052	0.995	0.652	** 0.636 ***
3000	0.962	0.946	1.033	0.956	0.766	** 0.685 *
175677 214051	1.000	1.000	1.000	1.000	1.000	1.000

^a 10774 SNPs within 10 kb of 642 a priori genes detected within total analyzed 175655 SNPs.

^b 13377 SNPs within 10 kb of 642 a priori genes detected within total analyzed 214051 SNPs.

Bold characters indicate odd ratios above one, normal font indicate odd ratios below one.

Fisher-Test: ***, **, * significant after Benjamini-Hochberg correction at P < 0.001, P < 0.01, P < 0.05, respectively.

Supplementary Table ST17: Possible SNP effects in candidate genes detected by GWA mapping.

Locus	Gene name	Chr	SNP position ^a	SNP alleles ^b	MAF	SNP location ^c	SNP effect ^d
AT5G56750	NDL1	5	22958883	G/C	0.325	intron	-
AT1G19180	JAZ1	1	6622295	A/G	0.15	intron	-
		1	6622362	T/A	0.1625	exon	synonymous
		1	6622707	G/C	0.175	exon	synonymous
		1	6623139	C/T	0.1625	exon	synonymous
AT1G33410	SAR1	1	12115793	G/T	0.1	exon	non-synonymous
		1	12117101	C/T	0.1	exon	non-synonymous
		1	12117452	G/C	0.1	intron	-
AT4G16310	LDL3	4	9218929	A/T	0.1125	exon	non-synonymous
		4	9224334	A/G	0.1125	intron	-
AT2G47460	ATMYB12	2	19478133	C/T	0.3	intron	-
AT5G15100	PIN8	5	4892517	C/T	0.375	exon	non-synonymous
AT3G28415	ABCB22	3	10648433	C/T	0.4	exon	non-synonymous
AT3G11900	ANT1	3	3759103	A/G	0.4375	exon	non-synonymous

^a SNP position based on TAIR10.

^b SNP alleles between 80 *A. thaliana* accessions.

^c Indicates if SNP lies within exon or intron region.

^d Indicates possible SNP effect on protein sequence based on TAIR10 annotation.

Glossary

General glossary

μM	micromolar
θ	Watterson estimator
π	Nucleotide diversity
D	Tajima's D statistic
FLD	Fu and Li's D* statistic
F_{LS}	Fu and Li's F* statistic
F_S	Fu's F _s statistic
H^2	Broad sense heritability
H_d	Haplotype diversity
noS	Number of segregating sites
R_{min}	Minimum number of recombination events
T_{noS}	Total number of sites
2,4-D	2,4-Dichlorophenoxyacetic acid
4-Cl-IAA	4-chloride indole-3-acetic acid
ANOVA	Analysis of variance
AMPRIL	<i>Arabidopsis</i> multiparent RIL
ASP	Aspartate
AuxRE	Auxin response element
bp	Base pair
BSH	Broad sense heritability
cDNA	complementary DNA
CDS	Coding sequence
CG	Candidate gene
CI	Confidence interval
CIM	Composite interval mapping
cM	Centimorgan
DBD	DNA-binding domain
DDC	Duplication-degeneration-complementation
DNA	Deoxyribonucleic acid
e.g.	"exempli gratia" - for example
EMMA	Efficient mixed-model association
et al.	"et alii" - and others
FC	Failure candidate
GFF3	Genetic feature format version 3
GLM	General linear model
GO	Gene ontology
GWA	Genome-wide association
GWAS	Genome-wide association study
HGI	Hypocotyl growth inhibition
HIF	Heterogenous inbred family
IAA	Indole-3-acetic acid
IAM	Indole-3-acetamine
IAOx	Indole-3-acetaloxime
IBA	Indole-3-butyric acid
IM	Interval mapping
IPA	Indole-3-puryvic acid
JA	Jasmonate
K-matrix	Kinship matrix
kb	Kilo base

General glossary
- continued

LD	Linkage disequilibrium
LOD	Logarithm of odds
MAF	Minor allele frequency
MAGIC	Multiple advanced generation intercross
miRNA	micro ribonucleic acid
MLM	Mixed linear model
mRNA	Messenger ribonucleic acid
MSB	Mean square between
MSW	Mean square within
NAA	Naphtaleneacetic acid
NB-LRR	Nucleotide-binding leucine-rich repeat
NIL	Near isogenic line
nM	nanomolar
ORF	Open reading frame
P-matrix	PCA matrix
PCA	Principle component analysis
PCR	Polymerase chain reaction
QR-code	Quick response code
QTL	Quantitative trait loci
QQ-plot	Quantile-quantile plot
R-gene	Resistant gene
RGI	Root growth inhibition
RGM	Representative gene model
RIL	Recombinant inbred line
RNA	Ribonucleic acid
RSA	Root system architecture
SNP	Single nucleotide polymorphism
TAM	Tryptamine
tasiRNA	Trans-acting small interference RNA
TD	Tandem duplication
TRCDIV	Transcript diversity
TRP	Tryptophan
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
URL	Uniform resource locator
<i>v</i> GWAS	variance-heterogeneity GWAS
WGD	Whole-genome duplication

Gene glossary

<i>5PTASE2</i>	<i>MYO-INOSITOL POLYPHOSPHATE 5-PHOSPHATASE 2</i>
<i>ABCB</i>	<i>ATP-BINDING CASSETTE FAMILY B</i>
<i>AFB</i>	<i>AUXIN F-BOX PROTEIN</i>
<i>ANT1</i>	<i>AROMATIC AND NEUTRAL TRANSPORTER 1</i>
<i>AUF</i>	<i>AUXIN UP-REGULATED F-BOX PROTEIN</i>
<i>AUX1</i>	<i>AUXIN RESISTANT 1</i>
<i>AUX/IAA</i>	<i>AUXIN/INDOLE-3-ACETIC ACID</i>
<i>ARFs</i>	<i>AUXIN RESPONSE FACTORs</i>
<i>ASA1</i>	<i>ANTHRANILATE SYNTHASE ALPHA SUBUNIT 1</i>
<i>ASB1</i>	<i>ANTHRANILATE SYNTHASE BETA SUBUNIT 1</i>
<i>DOG1</i>	<i>DELAY OF GERMINATION 1</i>
<i>ELF3</i>	<i>EARLY FLOWERING 3</i>
<i>GH3</i>	<i>GRETCHEN HAGEN 3</i>
<i>HAT2</i>	<i>HOMEODOMAIN FROM ARABIDOPSIS THALIANA 2</i>
<i>IAAs</i>	<i>INDOLE-3-ACETIC INDUCIBLEs</i>
<i>IAR1</i>	<i>IAA-ALANINE RESISTANT 1</i>
<i>IGS2</i>	<i>INDOLE-3-GLYCEROL PHOSPHATE SYNTHASE 2</i>
<i>ILLs</i>	<i>IAA-LEUCINE RESISTANT (ILR)-LIKEs</i>
<i>ILRs</i>	<i>IAA-LEUCINE RESISTANTs</i>
<i>ITR1</i>	<i>IRON-REGULATED TRANSPORTER 1</i>
<i>JAZ1</i>	<i>JASMONATE-ZIM-DOMAIN PROTEIN 1</i>
<i>LAX</i>	<i>LIKE AUX1</i>
<i>LDL3</i>	<i>LSD1-LIKE 3</i>
<i>MYB12</i>	<i>MYB DOMAIN PROTEIN 12</i>
<i>NDL1</i>	<i>N-MYC DOWNREGULATED-LIKE 1</i>
<i>NIT1</i>	<i>NITRILASE 1</i>
<i>PAIs</i>	<i>PHOSPHORIBOSYLANTHRANILATE ISOMERASEs</i>
<i>PAT1</i>	<i>PHOSPHORIBOSYLANTHRANILATE TRANSFERASE 1</i>
<i>PDR9</i>	<i>PLEIOTROPIC DRUG RESISTANCE 9</i>
<i>PGPs</i>	<i>P-GLYCOPROTEINs</i>
<i>PILS</i>	<i>PIN LIKES</i>
<i>PIN</i>	<i>PIN FORMED</i>
<i>SAR1</i>	<i>SUPPRESSOR OF AUXIN RESISTANCE1</i>
<i>SCF</i>	<i>SKP1/CDC53/F-BOX</i>
<i>SUR1</i>	<i>SUPERROOT 1</i>
<i>TAR1</i>	<i>TRYPTOPHAN AMINOTRANSFERASE RELATED 1</i>
<i>TIR1</i>	<i>TRANSPORT INHIBITOR RESISTANT 1</i>
<i>TSB2</i>	<i>TRYPTOPHAN SYNTHASE BETA-SUBUNIT 2</i>
<i>TSA1</i>	<i>TSK-ASSOCIATING PROTEIN 1</i>
<i>YUCCA</i> s	

German abstract

Pflanzenwachstum und -entwicklung wird primär durch Pflanzenhormone beeinflusst. Das Pflanzenhormon Auxin spielt in fast allen diesen Wachstums- und Entwicklungsprozessen eine Rolle. In der vorliegenden Arbeit wurde die natürliche Variation von Wachstumsprozessen nach Auxinbehandlung in jungen Keimlingen von *Arabidopsis thaliana* betrachtet. Im Kontext der adaptiven Selektion wurden quantitativ genetische Methoden genutzt und mit einer populationsgenetischen Analyse kombiniert, um Gene zu identifizieren, die zur beobachteten phänotypischen Variation beitragen. In der populationsgenetischen Analyse wurden 151 Gene betrachtet, die die Auxin-Biosynthese, den Auxin-Metabolismus, den Auxin-Transport und die Auxin-Signaltransduktion regulieren. Die Auxin-Transportgene erscheinen sehr konserviert, wohingegen die Auxin-Metabolismogene als die diverseste Auxin-Netzwerkengruppe identifiziert wurden. In Übereinstimmung mit diesen Resultaten, zeigten die Auxin-Metabolismogene die höchsten Transkriptionsdiversitätswerte, die mit einem veröffentlichten Expressionsdatensatz (Delker et al., 2010) für 6 *A. thaliana* Akzessionen berechnet wurden. Die populationsgenetische Analyse deutete neben den eben beschriebenen Resultaten für die Auxin-Metabolismogene darauf hin, dass die *AUXIN/INDOLE-3-ACETIC ACID*-Genfamilie die variabelste Genfamilie der Auxin-Signaltransduktionsgene ist. Im Gegensatz dazu zeigte die *TRANSPORT INHIBITOR RESISTANCE 1/AUXIN F-BOX PROTEIN* Genfamilie, wie die Auxin-Transportgene, Hinweise auf negative Selektion. Um Merkmale der Auxin-Antwort auf funktioneller Ebene zu analysieren, wurden quantitativ-genetische Untersuchungen wie "Quantitative Trait Loci" (QTL) Analysen und genomweite Assoziationsstudien (GWAS) durchgeführt. Nach erfolgreicher Entwicklung der Software RootDetection, die im Hochdurchsatz-Verfahren die Primärwurzellänge bestimmen kann, konnte die Software genutzt werden, um bei der phänotypischen Datenerfassung zu assistieren. Im Allgemeinen scheint die genetische Architektur, die die phänotypische Variation innerhalb der untersuchten Populationen beeinflusst, sehr komplex und überwiegend reguliert von genetischen Regionen mit kleinem Effekt zu sein. Indes unterstützen die Resultate für den QTL mit der höchsten Korrelation von Phänotyp- und Genotypdaten die populationsgenetische Analyse in soweit, dass drei Auxin-Metabolismogene innerhalb dieser QTL Region liegen. Daher erscheint es wahrscheinlich, dass Auxin-Metabolismogene zur gemessenen phänotypischen Variation beitragen. Mittels der GWAS zur Erforschung von Merkmalen der Auxin-Antwort konnte *AUXIN UP-REGULATED F-BOX PROTEIN 2* als Kandidatengen identifiziert werden, das die phänotypische Variation innerhalb der untersuchten *A. thaliana* Akzessionen beeinflusst. Zusätzlich konnten viele Korrelationen von Phänotyp- und Genotypdaten mit moderatem Signifikanzniveau ermittelt werden. Diese Resultate deuten wiederum auf die komplexe Natur der untersuchten Merkmale hin. Während womöglich die komplexe Architektur des Auxinnetzwerkes die Identifizierung von genomischen Regionen mit starkem Effekt verhinderte, wurden einige vielversprechende Kandidatengene und genomische Regionen identifiziert, die zukünftig funktionell validiert werden müssen.

Acknowledgements

The time in Halle at the IPB have passed away so fast. During this great period of my life many people played a role, on professional and social level. Here, I would like to take the opportunity to thank some people, without whom this all would not have been possible.

First, I would like to thank my supervisor Prof. Dr. Dierk Scheel and Prof. Dr. Steffen Abel for giving me the opportunity to work in this great institute. I had the chance not only to experience the fun part of science but also to get to know the political side of science during my period as a PhD Speaker. It was really great to have the opportunity to look behind the scene of this necessary part of science, which keeps things going.

Second, I would like to thank my co-supervisor Dr. Marcel Quint for his support during the years at the IPB. I exactly remember the first day in Halle, it was the day of my PhD job interview. A little nervous, due to the forthcoming talk I had to give, I walked up the Weinberg to the IPB and stopped at the sliding front door of the institute. It was just by coincidence that Marcel came by on his bicycle, just in time to open the gate to this life changing time.

I am very thankful to Dr. Carolin Delker for showing me so many biological aspects of the plant *A. thaliana*. Carolin, with your critical assistance in any kind of scientific related question and your sensitive way to communicate and bring someone back on the right track, I have learned a lot from you. I always enjoyed our inspiring conversations.

I am very thankful to Prof. Dr. Ivo Grosse and Yvonne Pöschl. They showed me the importance of statistics, dealing with thousands of data points. Without the informative and mind breaking meetings, I think, I would have never managed to find the shortest path between SNP positions or get to know the "Bell numbers". The collaboration was really a stimulating factor during my studies.

Without scientific collaborators also the quantitative part of this thesis would not have been successful. I am grateful to Prof. Dr. Thomas Altman not only for providing the RIL seeds but also to get to know the great phenotyping platform at the IPK Gatersleben, which opened up my mind for high-throughput image processing. I am thankful to Dr. Rhonda Meyer and Dr. Renate Schmidt for scientific advices concerning the QTL analysis part of this thesis.

The whole RootDetection software development would not have been possible without Ole Trenner, who is solid as a rock in questions related to Java programming. He implemented most of the core system of the RootDetection software and I am glad that our software development experience will continue.

During my studies I interacted with many people, who helped to broaden my horizon. Here, I would like to thank Luz Irina Calderón Villalobos and Clayton Douglas Grubb for the great conversations about science, after these discussions I was always inspired by their enthusiasm. A special thanks go to the bioinformatic group of Dr. Steffen Neuman. Whenever there was need, I could drop by and have a seat in your office to discuss my "R" problems with you. It was great to get to know Carsten Kuhl, Sebastian Wolf and Michael Gerlich, not only to be able to talk about IT stuff but also to spend some "out of institute" time in the evenings.

Without the help of the gardeners, the propagating of the different *A. thaliana* accessions would not have been possible. You are doing a great job for the whole institute. People are going to miss this facility, when they have to take care of their plants alone again.

I also would like to thank various assistant students, who helped planting thousands of plants, seed per seed. Here, special thanks go to Philipp Janitza and Max Hoffmann, I am grateful to get to know you.

From Marcel's group I would like to thank several people. Nadine, it was really fun with you, "beta propeller". You are the most organized person in the world I know, sometimes I miss this skill on my side. Anja, I am very thankful for your scientific advices, you are the living pubmed, unbelievable. Kathrin and Jana, what can I say, I am looking forward for every meeting with our small families.

Finally, I would like to thank my family.

I am very thankful to my sister Janne, my brother Niels and my father for their never ending support.

Christine, thank you for your support in any circumstance.

We not only gained experience from Halle but also Ida appeared in our lives.

Thanks the evantuality of evolution that both of you exist.

Curriculum vitae

PERSONAL DETAILS

Full Name: Kristian Karsten Ullrich

Nationality: German

Day of birth: 15.10.1980

Place of birth: Kiel, Germany

EDUCATION

2013 – now Scientific researcher, Philipps-Universität Marburg, Germany

2012 – 2012 Predoctoral fellow, VIB Department of Plant Systems Biology Gent, Belgium

2009 – 2012 PhD student, Leibniz Institute of Plant Biochemistry Halle, Germany

2009 Diploma degree: Diplom Biologe

2008 – 2009 Diploma at the institute of Molecular Phytopathology,
Christian-Albrechts University (CAU) Kiel, Germany

Diploma thesis:

”Klonierung von organspezifischen Genen aus der Kartoffel und Untersuchung
ihres Potenzials zur Erzeugung von gentechnischer Pathogenresistenz”

2001 – 2008 Subject of study: Biology, CAU Kiel, Germany

List of publications

Delker, C., Pöschl, Y., Raschke, A., Ullrich, K., Eттingshausen, S., Hauptmann, V., Grosse, I. and Quint, M. (2010). Natural variation of transcriptional auxin response networks in *Arabidopsis thaliana*. *The Plant Cell* 22, 2184–2200

Janitza, P., Ullrich, K. K. and Quint, M. (2012). Toward a comprehensive phylogenetic reconstruction of the evolutionary history of mitogen-activated protein kinases in the plant kingdom. *Frontiers in Plant Science* 3, 271

Quint, M., Drost, H.-G., Gabel, A., Ullrich, K. K., Bönn, M. and Grosse, I. (2012). A transcriptomic hourglass in plant embryogenesis. *Nature* 490, 98–101

Schumann, N., Navarro-Quezada, A., Ullrich, K., Kuhl, C. and Quint, M. (2011). Molecular evolution and selection patterns of plant F-box proteins with C-terminal kelch repeats. *Plant Physiology* 155, 835–50

Statutory declaration

Halle/Saale, 2013

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe. Es wurden keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt. Die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen sind als solche kenntlich gemacht.

Hiermit erkläre ich, dass ich noch keine vergeblichen Promotionsversuche unternommen habe und die vorliegende Dissertation nicht in der gegenwärtigen bzw. in einer anderen Fassung bereits einer anderen Fakultät / anderen wissenschaftlichen Einrichtung vorgelegt habe.

Teile dieser Arbeit sind veröffentlicht in:

Delker, C., Pöschl, Y., Raschke, A., Ullrich, K., Eттingshausen, S., Hauptmann, V., Grosse, I. and Quint, M. (2010). Natural variation of transcriptional auxin response networks in *Arabidopsis thaliana*. *The Plant Cell* 22, 2184–2200

Janitza, P., Ullrich, K. K. and Quint, M. (2012). Toward a comprehensive phylogenetic reconstruction of the evolutionary history of mitogen-activated protein kinases in the plant kingdom. *Frontiers in Plant Science* 3, 271