Institute of Plant Breeding and Agronomy I

Department of Plant Breeding

Justus Liebig University Giessen

Germany

Systems biological analysis of seedling vigour and osmotic stress tolerance in oilseed rape (*Brassica napus* L., Brassicaceae)

A thesis submitted for the requirement of the doctoral degree in Agriculture from the Faculty of Agricultural Sciences, Nutritional Sciences and Environmental Management, Justus Liebig University Giessen, Germany

Submitted by

KIDIST BOGALE KIBRET

Giessen, 2017

Supervisor: Prof. Dr. Rod Snowdon

Co-supervisor: Prof. Matthias Frisch

Examiners: Prof. Dr. Sylvia Schnell and Prof. Karl-Heinz Kogel

Chairperson: Prof. Dr. Gesine Lühken

Contents

List of Figures

List of Tables

1. Introduction

1.1 Taxonomy and breeding of *B. napus*

Oilseed rape (*Brassica napus* L.), Brassicaceae is a globally important oilseed crop. It has become increasingly important for edible oil, livestock feed and biofuel production after development of double zero varieties with low erucic acid for human consumption as well as low glucosinolates for animal feed (Obermeier and Friedt 2015). *B. napus* is a polyploid (AC, n=19) formed by hybridization of the paleopolyploids *B. rapa* (turnip) (A, n=10) and *B. oleracea* (cabbage) (C, n=9) (U 1935). Other polyploid crops of *Brassica* species described by (U 1935) are *B. carinata* (Ethiopian rape) formed by hybridization of *B. nigra* (B, n=8) and *B. oleracea* (C, n=9), and *B. juncea* (Indian mustard) formed by hybridization of *B. nigra* (B, n=8) and *B. rapa* (A, n=10). *B. carinata* is cultivated in Ethiopia as a vegetable and oilseed crop (Getinet et al. 1997). *B. rapa* and *B. oleracea* are used as vegetables while *B. juncea* and *B. nigra* are used as spices.

B. napus has two subspecies: *B. napus* ssp. *napobrassica* (the Swedes or the root vegetable) and *B. napus* ssp. *napus* (kale vegetable and fodder). Spring types are sown in spring and do not require vernalization whereas winter types are sown in autumn and flower in spring (Friedt and Snowdon 2010). *B. napus* is mainly self pollinating but facultative out crossing when pollinators are abundant. To increase genetic variability of *B. napus* use of hybrids and re-synthesized rapeseed is used. Haploid and doubled haploid plants have been developed to produce homozygous lines for hybrid breeding (Friedt and Snowdon 2010, Obermeier and Friedt 2015).

The recent availability of reference genome sequences for *B. napus* (Chalhoub et al. 2014) and its two diploid progenitors *B. rapa* (Wang et al. 2011) and *B. oleracea* (Liu et al. 2014), coupled with their close relationship to the Brassicaceae model plant *Arabidopsis thaliana* has facilitated the use of next generation sequencing techniques for genotyping and SNP discovery (Trick et al. 2009, Schmutzer et al. 2015). This has made way for the development of high density SNP array for genetic mapping and genome-wide association analysis (GWAS) studies of several *B. napus* traits such as seed germination and vigour (Hatzig et al. 2015), seedling development (Körber et al. 2015) and drought stress tolerance (Zhang et al. 2015). The use of next generation sequencing for RNA-sequencing has led to numerous possibilities of studies that contribute to the breeding of *B. napus*. Some examples are to

investigate seed development (Obermeier et al. 2009), seed quality traits (Harper et al. 2012), seedling development (Körber et al. 2015) and drought stress (Liu et al. 2015).

1.2 Overview of seedling vigour and emergence

Seed germination and seedling vigour are among the important agronomical traits for *B. napus* (Obermeier and Friedt 2015). Seed germination and seedling vigor are related processes in plants that affect seedling development and yield (Ellis 1992). In addition to seed quality, several environmental factors affect seedling growth (Egli and Rucker 2012). Seedling vigor can be strongly influenced by multiple environmental influences and the mechanisms for the variation are unclear. Variation in germination and emergence is a major challenge for oilseed rape breeders, because strong influences from both the seed production environment and the sowing environment infer low heritability (Hatzig et al. 2015). Identification of genetic factors underlying this variation is further complicated in *B. napus* by its complex amphipolyploid genome. In such cases, utilisation of recent advancements in genome and transcriptome sequencing can provide important insight beyond conventional genetic analyses (Snowdon et al. 2012, Edwards et al. 2013). The recent availability of reference genome sequences for *B. napus* (Chalhoub et al. 2014) and its two diploid progenitors *B. rapa* (Wang et al. 2011) and *B. oleracea* (Liu et al. 2014), coupled with their close relationship to the Brassicaceae model plant *Arabidopsis thaliana*, open the way for systems-analytical studies of complex traits like emergence in oilseed rape.

Systems analysis approaches combining phenome, genome, transcriptome and metabolome data are useful to gain an understanding of the regulatory processes underlying seed germination and seedling vigour. Biochemical (de los Reyes et al. 2003) and genetic (Zhou et al. 2007) markers for seedling vigour have been identified in some crops. However, to date systems-based approaches have generally been limited to model plant species with small genomes that are easy to handle in controlled-environment phenotyping studies. Unfortunately such approaches ignore the strong genotype by environment interactions that are of most interest for harnessing emergence-related traits in field crops.

1.3 Overview of drought response in plants

Drought has occurred frequently in many areas of the world. Due to climate change, it is predicted to persist in even more parts of the world in the future. (Dai 2011) made climate simulation models based on the past couple of decades and predicted areas affected by severe drought to increase within this century. Drought severely reduces the yield of crops (Boyer

2010) including *B. napus* (Good and Maclagan 1993). Thus there is a growing need to breed for drought resistant plants, i.e, plants able to produce stable yields in water-deficient environments (Fang and Xiong 2015).

Drought leads to osmotic stress. Similarly, osmotic stress can also be caused by salt stress. It is caused by cell dehydration and disrupts many cellular functions (XIONG and ZHU 2002). Early responses to drought and salt stress are similar. Improving osmotic adjustment capabilities can improve crops in terms of withstanding water shortage (Seeraj and Sinclair 2002).

Environmental stresses that cause osmotic changes and damage to cells are first perceived by osmosensors and lead to the activation of various signalling pathways. Signal perception pathways include inositol phosphates, Ca2+ and reactive oxygen species (Xiong et al. 2002, Vinocur and Altman 2005). Signalling transduction pathways include kinases (Boudsocq and Laurière 2005), ubiquitination pathway (Stone 2014) and hormones. The most important signalling hormone is abscisic acid (ABA), which results not only in activation of stress responsive transcription factors (TFs) and genes, but also other physiological responses to drought such as stomatal closure to minimize water loss via transpiration (Lee and Luan 2012).

The drought stress perception and transduction signals affect transcription factors which in turn lead to activation of genes that have several functions such as protection of the cells against osmotic stress. Osmotic stress responsive transcription factors (TFs) can be classified as ABA dependent and independent. ABA dependent TFs include the key ABA-responsive element (*ABRE*), *MYC* and *MYB*. ABA independent TFs includes drought responsive element binding (*DREB*) (Yoshida et al. 2014, Shinozaki et al. 2003).

Responses to osmotic stress include a variety of techniques which render the plant osmotic stress tolerant by maintaining cellular homeostasis, and protecting proteins and membranes from damage. These processes include removal of reactive oxygen species, detoxification, water and ion movement and osmoprotection via osmolyte accumulation (Vinocur and Altman 2005). One of such responses is metabolite accumulation in order to stabilize the cells and maintain cell turgor pressure. Proline and γ-aminobutyric acid (GABA) were reported to be accumulated due to osmotic stress in several plant species. Other metabolites which accumulate include sugars, myo-inositol and its derivatives, glycine betaine, asparagine, glycine and alanine (Krasensky and Jonak 2012, D'Mello 2015, Merewitz et al. 2012). In *B.*

napus leaves, the concentration of alanine, arginine, asparagines, aspartate, betaalanine, GABA, glutamate, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, threonine, valine (Good and Zaplachinski 1994), *myo*-inositol, sucrose and fructose (Hatzig et al. 2014, Koh et al. 2015) were found to be increased in response to osmotic stress.

1.4 Approaches of drought response study in *B. napus*

Due to the complexity of drought stress resistance mechanisms in plants, the use of systems analysis to integrate data at multiple levels of omics has been proposed to better understand the holistic responses to drought rather than the components (Chawla et al. 2011). Drought related traits from multiple levels of omics have been useful for selection of drought resistant plants (Fang and Xiong 2015). Such studies using omics approaches of drought stress responses have also been performed in *B. napus*. For example, one *B. napus* variety was investigated for expression, proteomic and osmolyte changes due to drought stress (Koh et al. 2015), and another cultivar for its amino acid accumulation under osmotic stress (Good and Zaplachinski 1994). Differential expression analysis of stressed and control samples of a single *B. napus* genotype also provided insight into drought stress responses (Liu et al. 2015, Shamloo-Dashtpagerdi et al. 2015). However, for breeding purposes, a comparison between different genotypes in terms of osmotic stress responses would be more useful. In a study by (Hatzig et al. 2014), drought tolerant and susceptible *B. napus* genotypes were investigated for physiological, metabolite and hormone adjustments which showed different responses. Moreover, in a genome-wide association (GWAS) study that studied water stress tolerance in seedlings of *B. napus*, 16 loci were significantly associated with water stress response induced by using polyethylene glycol. Those loci contained differentially expressed genes in response to the osmotic stress (Zhang et al. 2015).

Gene co-expression network provides additional information to differential expression studies involving a few genotypes because the network analysis clusters genes with according to their responses to drought in a diverse set of genotypes. Moreover, such systems analysis approach allows the integration of multiple levels of transcriptome, metabolome and phenome data across a range of genotypes with varying drought stress responses.

1.5 Transcriptional profiling (RNA-Seq) using next generation sequencing (NGS) and third-generation sequencing technology

One of the applications of NGS technologies is transcriptional profiling (RNA-Seq). Next-generation sequencing is rapidly becoming the method of choice for transcriptional profiling

experiments. In contrast to microarray technology, high throughput sequencing allows identification of novel transcripts, does not require a sequenced genome and avoids background noise associated with fluorescence quantification. Furthermore, unlike hybridization-based detection, RNA-seq allows genome-wide analysis of transcription at single nucleotide resolution, including identification of alternative splicing events and post-transcriptional RNA editing events. Transcriptomic research has benefited from greater accessibility to NGS (Trapnell et al. 2010).

All RNA-seq experiments using NGS follow a similar protocol: Isolation of RNA or polyA mRNA, RNA fragmentation, synthesis of high-quality double-stranded cDNA and library preparation. Total RNA is isolated from a sample of interest which, depending on the type of RNA to be profiled, may be purified to enrich for mRNAs, microRNAs or long non-coding RNA (lincRNAs) prior to preparing an RNA library. Library preparation may involve such steps as reverse transcription to cDNA, PCR amplification and may or may not preserve strandedness information. Sequencing can produce one read in a single-end sequencing reaction, or two ends separated by an unsequenced fragment in paired-end reactions (Mortazavi et al. 2008).

Up to date, NGS research is increasingly being conducted within the Illumina suite of instruments. The following are general details of the standard illumina procedures followed for the transcriptome sequencing. For preparation of 3' fragment cDNA libraries, RNA quality control is done, polyA+ mRNA from each total RNA sample is isolated and 3'-fragment cDNA library is prepared from mRNA. Cluster generation is done to bind single molecules on the flow cell with oligos, and bridge amplification and sequencing by synthesis is done to incorporate fluorescent nucleotides that are imaged. Two different oligos called adapters that are partly complementary are ligated to the fragmented DNA. One copy turns into thousands-fold clonal copies distributed in discrete clusters across the surface. The adapter ligated DNA hybridizes randomly to a complementary oligo on the surface of the flow cell. A surface coated with millions of oligos that are complementary to the adapter-sequences are introduced into the workflow. A universal sequencing primer is annealed to the millions of templates scattered across the surface. The annealing site is located within the adapter sequence. The templates are sequenced with fluorescent nucleotides and base calling occurs automatically while sequencing. Random fragmentation and hybridization of the template to the oligo-coated surface ensure unbiased sequencing across all regions of the template. Each discrete cluster of clonal copies of the template represents a tiny fraction of the

whole genome. And when millions of templates are being sequenced in parallel across the oligo-coated surface most of the genome is covered. The number of reads for a specific region is denoted "depth" or "coverage." In the infancy of RNA-seq, technical replicates (libraries prepared from the same RNA sample) were commonly used. However, it has been shown that biological variation far outweighs technical variation, at least when coverage of at least 5 reads per nucleotide is obtained. Technical replicates, therefore, are most useful when the goal is to compare performance of two or more competing sequencing technologies (McIntyre et al. 2011).

RNA expression analysis is most often performed by NGS sequencing of cDNA copies. To date, RNA-seq has been mostly done using short read platforms. A major limitation of short-read data is that it is difficult to accurately predict full-length splice isoforms. Alternative splicing results in a single gene coding for multiple proteins when exons of a gene may be included within or excluded from the final, processed messenger RNA (mRNA) produced from that gene. Thus, the proteins translated from alternatively spliced mRNAs will contain differences in their amino acid sequence and in their biological functions. The NGS reads are too short to represent a full transcript with all the isoforms, thus requiring assembly of cDNA reads into full-length transcripts. This is an issue for the accurate characterization of RNA splice isoforms because there is often insufficient information to quantification of the different transcripts properly. Long reads provide superior performance for detecting transcriptomic structure. A solution to the current problem of isoform identification is a limiting factor for RNA-Seq and requires technology capable of providing long reads in the range of 1.5–10 kbp. The median length of human gene transcripts is about 2.5 kbp; long reads should be able to provide full-length mRNA isoforms, detect new isoforms, and bypass the transcript reconstruction process by identifying isoforms directly. Recently, Roche announced the shutdown of the 454 pyrosequencing technology, and Life Technologies replaced ABI SOLiD technology in favor of the Ion Torrent sequencing technology. With those changes, there are only two major companies of next-generation sequencing: Illumina and Ion Torrent. Both technologies provide relatively short reads, Illumina with 100–150 bp reads, and Ion technology with 200–400 bp reads. The third-generation sequencing technology promises affordable, real-time sequencing with long reads (>5 kbp). Oxford Nanopore and Pacific Biosciences are the two companies that fill the void for such technology. Full-length cDNA reads would avoid this problem and can be executed with either the PacBio or MinION platforms (Gonzalez-Garay 2015).

PacBio Iso-Seq library preparation and sequencing, also called the isoform sequencing (Iso-Seq) application is a relatively well established method and has been applied recently to two crops: Sorghum and Maize. Iso-Seq generates full-length cDNA sequences from the 5' end of transcripts to the poly-A tail eliminating the need for transcriptome reconstruction using isoform-inference algorithms to improve transcriptome sequencing of complex genomes. The method depends on the isolation of high-quality total RNA. The majority of the applications require the isolation of polyA mRNA, but the Iso-Seq method is flexible enough to allow the sequencing of different types of RNA. cDNA synthesis is done with adapters. By doing size partitioning of cDNA, three different fractions are isolated: 1–2 kilobase pair (Kb) or up to 2000 bp, 2–3 Kb and 3–6 Kb. Optionally, a fourth fraction of 5–10 Kb could be included. After size selection, the double-stranded cDNA is PCR amplified for SMRTbell library construction (Gonzalez-Garay 2015).

Iso-Seq has the potential to improve transcriptome sequencing of complex genomes: demonstrated by isolating thousands of molecules that differ by a small number of variations introduced during the recombination events associated with B-cell maturation of the bovine immunoglobulin G (IgG) repertoire. Iso-Seq has also been used to analyse full-length splice isoforms in human organs and embryonic stem cells: a long-read profile of the human transcriptome showed that >10% of the reads represented novel splice isoforms. Recently, the sorghum transcriptome was sequenced using Pacific Biosciences single-molecule real-time long-read isoform sequencing to identify Alternative splicing and alternative polyadenylation (APA) of pre-mRNAs. And a pipeline called TAPIS (Transcriptome Analysis Pipeline for Isoform Sequencing) to identify full-length splice isoforms and APA sites developed. This study found transcriptome-wide full-length isoforms with over 11,000 novel splice isoforms, APA of 11,000 expressed genes and more than 2,100 novel genes. Pacific Biosciences single-molecule long reads obtained using the Iso-Seq protocol offer a considerable advantage in transcriptome-wide identification of full-length splice isoforms and other forms of post-transcriptional regulatory events such as APA (Abdel-Ghany et al. 2016). This technology was also applied to the study of Maize transcriptomics. Using PacBio long-read single-molecule sequencing technology to transcriptome sequencing in maize, 111,151 transcripts from 6 tissues captured around 70% of the genes annotated in maize reference genome. A large proportion of transcripts (57%) represented novel, sometimes tissue-specific, isoforms of known genes and 3% correspond to novel gene loci. In other cases, the identified transcripts improved existing gene models. Averaging across all six tissues, 90% of the splice junctions were supported by short reads from matched tissues. In addition, the study identified

a large number of novel long non-coding RNAs and fusion transcripts and found that DNA methylation plays an important role in generating various isoforms (Wang et al. 2016).

In 2015, Oxford Nanopore Technologies introduced a portable sequencer called the MinION, a 90-g portable device. The nanopore sequencing system consists of motor and nanopore reader proteins set in on electrical current flowing membranes. While a double strand DNA escorted by a motor protein is unraveling, one of the strands goes through a nanopore reader that sends electrical current signal each base. Prior to sequencing, adapters are ligated to both ends of genomic DNA or cDNA fragments. Upon capture of a DNA molecule in the nanopore, the DNA passes through the pore, the sensor detects changes in ionic current caused by differences in the shifting nucleotide sequences occupying the pore. These ionic current changes are segmented as discrete events that have an associated duration, mean amplitude, and variance. This sequence of events is then interpreted computationally. The MinION's 1D sequencing reads only one strand, leaving the other complementary strand not sequenced. A 2D sequencing reads both strands to improve sequencing quality. Oxford Nanopore Technologies has introduced a concept of 'Run until'. The MinION could be instructed to run until a pre-specified read coverage has reached. If each RNA-seq experiment is configured to stop when reaching a certain threshold of sequence depth, the sequencing depth bias could be alleviated. The company would also to develop technology that sequences RNA and protein molecules, and the reverse-transcription of RNAs for complementary DNAs would not be necessary. Therefore, sequencing and detection of protein molecules will help the direct measure of gene expression. The nanopore technology primarily measures electrical current differences along the membrane due to molecular-level changes in the membrane-bound channel, through which target molecules pass. The technology can detect not just nucleotides or amino acids but also any other modification thereof. Thus, proteins after post-translational modification could be distinguished. To illustrate the application in RNAseq, it was used to determine RNA splice variants and to detect isoforms for four genes in Drosophila. Among these is, the most complex alternatively spliced gene known in nature, with 18,612 possible isoforms ranging in length from 1806 bp to 1860 bp over 7000 isoforms for with >90% alignment identity were detected. Identifying these isoforms would have been impossible with 450-base-long NGS reads (Jain et al. 2016).

1.6. Next generation sequencing transcriptional profiling (RNA-Seq) data analysis

Wide varieties of tools are available for mapping, normalization and calculating the abundance of each gene expressed in a sample. These tools are split in to two groups.

Commercial tools such as those offered by CLC bio (http://www.clcbio.com/) and Partek (http://www.partek.com/) are user friendly. The second type of tools are tools which are open source, developed, supported and published by the scientific community. The second types of tools were reviewed in this section. One type of such tools is open, web-based platform such as Galaxy (Goecks et al. 2010) and GeneProf (Halbritter et al. 2014). There are several others based on command line interface, shell commands and scripting. These are different specialized tools for each step of data analysis: preprocessing, normalization, mapping, differential expression analysis and clustering or network analysis. The most commonly used tools for these steps are FastQC and FASTX tools for quality control, Burrows-Wheeler transform (BWT) for mapping and the TUXEDO protocol for normalization. These were described in detail in the following sections. However, several new tools are emerging recently to further facilitate data analysis. For instance, MultiQC is a tool to create a single report visualizing output from multiple tools across many samples (Ewels et al. 2016). HTSeq is a Python library to facilitate the rapid development of custom scripts for analyzing high-throughput sequencing data that are needed once a project deviates from standard workflows. HTSeq offers parsers for many common data formats in high-throughput sequencing (HTS) projects, as well as classes to represent data, such as genomic coordinates, sequences, sequencing reads, alignments, gene model information and variant calls, and provides data structures that allow for querying via genomic coordinates. Htseq-count is a tool developed with HTSeq that preprocesses RNA-Seq data for differential expression analysis by counting the overlap of reads with genes (Anders et al. 2015). The major drawback of RNA-seq analysis was the high time consumption and requirement of high computational capacity. Recently methods for rapid analysis of RNAseq data have been developed such as kallisto (Bray et al. 2016). The latest program developed is Salmon with an ultra-fast read mapping procedure. It is the first transcriptome-wide quantifier to correct for fragment GC-content bias, which substantially improves the accuracy of abundance estimates and the sensitivity of subsequent differential expression analysis (Patro et al. 2017).

Transcriptome profiling data analyzed using the established and commonly used methods Tuxedo protocol (which includes cufflinks for normalization and cuffdiff for differential gene expression analysis) and HTSeq were proved to be highly reproducible and highly correlated to wet-lab validation using Real-Time Quantitative PCR (qPCR). (Everaert et al. 2017) performed an independent benchmarking study using RNA-sequencing reads that were processed using five workflows (Tophat-HTSeq, Tophat-Cufflinks, STAR-HTSeq, Kallisto and Salmon) and resulting gene expression measurements were compared to expression data

generated by wet-lab validated qPCR assays for all protein coding genes. All methods showed high gene expression correlations with qPCR data.

1.6.1 Preprocessing

Currently, all commercially available RNA-seq platforms rely on reverse transcription and PCR amplification prior to sequencing and sequencing is therefore subject to the biases inherent to these procedures. First, annealing of random hexamer primers to fragmented RNA is not random, which results in depletion of reads at both 5' and 3' ends (Roberts et al. 2011). This makes the identification of the true start and end of novel transcripts a challenge, as well as underestimating expression level of short genes. Second, PCR can introduce bias based upon GC content and length due to non-linear amplification (Risso et al. 2011). A number of data analysis tools to correct these biases are available.

Sequencing errors can be identified from the quality score. Most tools expect Sanger format by default so the first step would be to convert the Phred score into Sanger format. For the quality control of the FastQ data, FastQC, a quality control tool for high throughput sequence data will be used. It is used to identify biases before analysis and spots problems originating either in the sequence or in starting library material. Graphical representation of various control parameters facilitates quality control checks (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Per base sequence quality shows quality values (y-axis) across all bases at each position in the FastQ file. High quality well above 20 is expected for all bases. The per sequence quality score reports a subset of sequences with low quality values and high quality values. It helps to visualize which majority of sequences have high or low quality. It helps to decide whether to trim all sequences or filter out low quality ones if most of them are good quality. Per base sequence content should run parallel with each other because base content in a random library should not have difference in AT and GC content. Per base GC content is expected not to change across the read. A normal distribution of GC content is expected. This graph is used to check the presence of contamination. If there are secondary peaks at top of a smooth peak, it's an indication of contamination. Number of uncalled base (N) and level of duplicated sequences should be low. Overrepresented sequences module lists the over represented sequences if they represent greater than 0.01% of the library. It's another indication of contamination. Overrepresented Kmers can be plotted and viewed with the FastQC module. K-mer is a short oligonucleotide of length k. K-mers with low frequency are either sequencing errors or low abundance transcripts, and in both cases need to be removed for decreasing the amount of

random access memory (RAM) (Martin and Wang 2011). After identification of errors, the next step is to remove adapters, trim reads and remove low quality bases or sequences.

1.6.2 Mapping

Challenges involved in mapping RNA seq reads against a reference genome are short reads from Illumina, repetitive regions of a genome, alternative splicing, sequencing errors and very large data produced up to tera bases. Reference based assembly needs good quality reference genome and a computation tool to identify splice junctions (Martin and Wang 2011). Burrows-Wheeler transform (BWT) is one of the algorithms used for fast and accurate short read alignment to a large reference sequence and is suitable for Illumina reads (Li and Durbin 2009).

Bowtie is a fast, memory efficient aligner based on Burrows-Wheeler transform with downstream tools and is available as an open source. It can align 35-base-pair reads to the human genome at a rate of 25 million reads per hour, a significant improvement over the earlier aligners and can be used on a standard desktop. It supports standard FASTQ and FASTA input formats. The speed is achieved at the cost of failure to align small number of reads with multiple mismatches and minimizing mismatches leads to increased computational costs. However, if one or more best matches exist, Bowtie will report one. It can be made to report multiple reads and increased sensitivity using command line. The first valid alignment reported is not always the best so command option ‘best’ can be used to instruct for selection based on the number of mismatches. The user can make Bowtie report all alignments up to a specified number (option -k) or all alignments with no limit on the number (option -a) for a given read (Langmead et al. 2009). The output is as a SAM file which is a TAB delimited text format, an optional header starting with @ and alignment without @. For each alignment line there is a mandatory field for alignment information such as mapping position and other optional fields (The SAM format specification workgroup 2011). Paired end alignment is possible with versions starting from 0.9.9.2 release. Bowtie 1 and Bowtie 2 have been developed with different applications (http://bowtie-bio.sourceforge.net). The other popular mapping tool for short reads of up to 50 bp in length from Illumina is SOAP. It is used to align small RNA discovery and mRNA tag sequence mapping (Li et al. 2008).

1.6.3 Normalization

Given that the total number of reads per transcript is proportional to the level of a transcript multiplied by transcript length, a long transcript will be sequenced more often than a short

transcript when expressed at equivalent levels. Since statistical power is closely linked to sample size, a long transcript is more likely to be found differentially expressed than a short transcript. To mitigate this problem, expression levels are frequently expressed by calculating the number of reads or fragments per kilobase per million reads (RPKM and FPKM, respectively). The FPKM transformation also allows direct comparison of transcript expression between two libraries with different sequencing depth as well as an indication of relative expression levels between two or more transcripts in a single library (Mortazavi et al. 2008). To report abundance of transcript and compare between two or more samples, normalization is needed. The normalization method that normalizes for RNA length and for total read number called Reads per Kilobase of exon or transcript per Million mapped reads (RPKM) or Fragments per Kilobase of exon or transcript per Million mapped reads (FPKM). The difference between RPKM and FPKM is the nomenclature change from RPKM to FPKM to better reflect what RNA-Seq actually measures in case of paired-end sequencing. A fragment corresponds to a single cDNA molecule, which can be represented by a pair of reads from each end for paired-end sequencing or reads from one end in case of single-end sequencing. Fragment means fragment of DNA, so the two reads that comprise a paired-end read count as one. Thus, the preferred term by which transcript abundances are reported for both single and paired end sequencing is FPKM. Per kilobase of exon means the counts of fragments are then normalized by dividing by the total length of all exons in the gene (or transcript). This makes it possible to compare expressed genes even if they are of different lengths. Per million reads means this value is then normalized against the library size. This makes it possible to compare expressed genes in different with different number of reads (Mortazavi et al. 2008).

TopHat and Cufflinks are free tools as part of the Tuxedo protocol developed for RNA seq analysis of illumina data, identify new genes and splice variants, and compare expressions in different conditions. They can be used both in UNIX shell or a computing cloud such as galaxy (Trapnell et al. 2012). TopHat is a tool that is used to identify not only already known splice junctions but also identifies novel splice junctions from short reads. These normalization tools use Bowtie as an alignment tool. The outputs are .BAM for data, .BAI for index and .bed file. Cufflinks contains FPKM tracing files called Gene expression.fpkm_tracking which shows gene expression values and estimated isoform levels are given in Transcript expression.fpkm_tracking files (Trapnell et al. 2010). Cufflinks uses the output file from Tophat (bam) or Bowtie (sam) and maps it against the genome to assemble the reads into transcripts for each condition. Cufflinks package includes Cuffmerge,

which merges the assemblies together. As part of Tuxedo tools, Cuffdiff, which reports genes and transcripts that are differentially expressed from reads and merged assembly of many conditions. Cuffdiff can also identify genes differentially regulated at the transcriptional or post-transcriptional level. CummeRbund gives output of Cuffdiff in figures and plots and can connect Cufflinks to the R statistical computing environment (Trapnell et al. 2012).

Other established normalization methods based on genome alignments are Augustus, Exonerate, GSTRUCT, mGene, mTim, NextGeneid, SLIDE, Transomics, Trembly, Tromer. In addition, another method to estimate the abundances of known and novel isoforms iReckon was published later than Cufflinks and TopHat (Mezlini et al. 2013).

Including Tuxedo tools, most of the well-developed methods for detecting differentially expressed genes take count data as the input to their statistical model. Count data of reads are normalized to account for the biases before the analysis of gene expression. Recommendation for normalizing gene counts includes quantile-based normalization and DESeq or edgeR Bioconductor package's scaling factor. Count-based data can be fed to either DESeq or edgeR package to quantify transcript or gene expression. DESeq and edgeR are available as packages of the Bioconductor software development project (Anders et al. 2013).

1.7 Third-generation sequencing transcriptional profiling (RNA-Seq) data analysis

The MinION performance was optimized by the improved MinION sequencing chemistry and base-calling software. Combined with new MinION-specific bioinformatics tools, the identity of sequenced reads, that is, the proportion of bases in sequencing read that align to a matching base in a reference sequence were improved from 66% to 92% De novo base-calling for MinION data is performed using hidden Markov model (HMM) based methods by Metrichor, a cloud-based computing service provided by Oxford Nanopore Technologies. Metrichor requires an active internet connection and its base-calling source code is available to registered MinION users under a developer license. To create a fully open-source alternative, Nanocall, an HMM-based base-caller that performs efficient 1D base-calling locally without requiring an internet connection at accuracies comparable to Metrichor-based 1D base-calling and Deep-Nano, a recurrent neural network framework which performs base-calling and yields better accuracies than HMM-based methods have been recently developed. The first attempts at aligning MinION reads to reference sequences used conventional alignment programs. Most of these were designed for short-read technologies, such as the 250-nucleotide highly accurate reads produced by the Illumina platform. MarginAlign was later developed to improve alignments of MinION reads to a reference genome by better

estimating the sources of error in MinION reads. Jain et al., Improved data analysis for the MinION nanopore sequencer (Jain et al. 2016).

(Abdel-Ghany et al. 2016) sequenced the sorghum transcriptome using Pacific Biosciences single-molecule real-time long-read isoform sequencing to identify full-length splice isoforms and alternative polyadenylation (APA) sites. To analyze this data set, (Abdel-Ghany et al. 2016) developed TAPIS (Transcriptome Analysis Pipeline for Isoform Sequencing), a computational pipeline to correct errors, align reads to the reference genome, identify all splice isoforms and alternative splicing events generated from a gene and 3′ heterogeneity because of alternative polyadenylation sites. As the accuracy of read mapping is affected by read quality, especially for the correct identification of splice junctions, the first step in the TAPIS pipeline consists of an iterative process that alternates read mapping and error correction on the basis of the reference genome. Using this approach, nearly 95% of the reads were aligned to the reference genome after filtering alignments for splice junction accuracy on the basis of a model of sorghum splice junctions (Abdel-Ghany et al. 2016). The most recently developed method for analysis of Pacific Biosciences long read RNA sequencing is the Iso-Seq Browser, a Web-based visual analytics tool for long-read RNA sequencing data produced by Pacific Biosciences isoform sequencing (Iso-Seq) techniques. Key features of the Iso-Seq Browser are: an exon-only web-based interface with zooming and exon highlighting for exploring reference gene transcripts and novel gene isoforms, automated grouping of transcripts and isoforms by similarity, many customization features for data exploration and creating publication ready figures, and exporting selected isoforms into fasta files for further analysis. Iso-Seq Browser was written in Python using several scientific libraries. The application and analyses described in this paper are freely available to both academic and commercial users at https://github.com/goeckslab/isoseq-browser. Iso-Seq Browser enables interactive genome-wide visual analysis of long RNA sequence reads. Through visualization, highlighting, clustering, and filtering of gene isoforms, ISB makes it simple to identify novel isoforms and novel isoform features such as exons, introns and untranslated regions (Hu et al. 2017).

1.8. Methods for profiling and analysis of metabolome

Plants produce large numbers of metabolites of diversified structures and abundance that play important roles in plant growth, development, and response to environments. The diversity of plant metabolites and the likely complicated regulatory mechanism highlight the necessity to explore the underlying biochemical nature. The output of plant metabolomics depends largely

on its methodologies and instrumentations to comprehensively identify, quantify, and localize every metabolite. The methodologies and instrumentations of plant metabolomics have been developing rapidly. At present large scale analysis of highly complex mixtures are enabled by a series of integrated technologies and methodologies, such as non-destructive NMR (nuclear magnetic resonance spectroscopy), mass spectrometry (MS) based methods including GC–MS (gas chromatography–MS), LC–MS (liquid chromatography–MS) and CE–MS (capillary electrophoresis–MS), and FI-ICR–MS (Fourier transform ion cyclotron resonance–MS). Because of the limitation of each analytical platform, combined approaches are increasingly used in metabolomics analysis (Hong et al. 2016). In case of *B. napus*, amino acids can be extracted from ground plant material using a mixture of methanol–chloroform–water containing a known concentration of DL-3-aminobutyric acid (BABA) as an internal standard. Amino acids can then be derivatized with a Waters kit (Waters Corporation, Milford, MA, USA) for analysis by UPLC (Ultra Performance Liquid Chromatography). Individual amino acids can then be identified by co-chromatography with pure synthetic compounds (Sigma-Aldrich, St. Louis, MO, USA) and quantified with respect to the BABA signal and individual external calibration curves (Albert et al. 2012, Deleu et al. 2013). Analysis of non-structural carbohydrates (NSCs) and organic acids (OAs) in *B. napus* is possible using a gas chromatography-flame ionization detector (GC-FID) System. Metabolites can be identified by comparison of sample chromatograms to standard mixtures of known concentration and quantified in absolute amount after normalization against internal standards and plant material dry weight (Lugan et al. 2009).

Metabolomics is used to obtain a large amount of valuable information for the discovery of genes and pathways through accurate and high throughput corollary peak annotation by snapshot of the plant metabolome. There is a complicated regulatory network among these small molecules in plants, and by detecting the interactions among these metabolites, metabolomic analysis contributes significantly to the understanding of the relation between genotype and metabolic outputs by tackling key network components. Such kinds of metabolomic analysis, integrated with transcriptomic analysis, have been successfully applied to investigate the coordinated rules of metabolic fluxes and metabolite concentrations in plants. The temporal changes in the metabolic state can be characterized by the metabolite concentrations and flux rates. Metabolic flux analysis (MFA) has accelerated the development of computational methods for analysis of metabolic networks. Flux balance analysis (FBA) together with its extension, dynamic FBA, have proven instrumental for analyzing the robustness and dynamics of the metabolome. Flux balance analysis (FBA), as one of the most

prominent of the MFA methods, is based on linear programming (LP) whereby a given objective function such as biomass yield is optimized under the assumption that the system operates at steady state under the constraints given by the stoichiometric matrix. Consequently, the biological implications of the optimal flux distribution depend on the choice of the objective function. Recently, new methods for analyzing the dynamics of metabolic networks and for quantifying their robustness have been developed. Following the biochemically meaningful premise that metabolite concentrations exhibit smooth temporal changes, the new methods rely on minimizing the significant fluctuations of metabolic profiles to predict the time-resolved metabolic state, characterized by both fluxes and concentrations. By conducting a comparative analysis with a kinetic model of the Calvin-Benson cycle and a model of plant carbohydrate metabolism, the principle of regulatory on/off minimization coupled with dynamic FBA can accurately predict the changes in metabolic states (Kleessen and Nikoloski 2012).

Recent advances in metabolomics mentioned above such as flux analysis have provided a large amount of highly reproducible data, allowing reconstruction and analysis of genome-scale metabolic networks. These developments in metabolomics technologies have challenged computational systems biology. A network reconstructed from metabolite profiling data is a collection of nodes, which represent the metabolites, and edges, which capture the relationships between the metabolites. This network definition is similar to network analysis of other biochemical components such as proteins and gene transcripts. A relationship between the data profiles of two metabolites can be quantified by applying different similarity measures and/or principles from probability theory. A correspondence between graph theory-based networks and biochemical networks obtained from KEGG indicated that graph theoretic properties contain biochemically meaningful information. The choice of a similarity measure depends on the biological question to be answered. Different relationships are generally captured by applying different measures to pairs of metabolic profiles. For instance, correlation-like similarity measures extract linear relationships and the Euclidean distance accounts for differences in relative abundance between metabolites. After applying similarity measures and/or principles from probability theory to all pairs of metabolic profiles, a similarity matrix is obtained. Constructing the weighted network from the similarity matrix is done by the application of a statistically sound threshold (Toubiana et al. 2013).

As a result of high throughput metabolite profiling, mQTL (metabolite quantitative trait loci) and mGWAS (metabolome-based genome-wide association study) studies in plants has been

possible using the metabolic phenotype to study genetic variation. In Arabidopsis, the analysis of 369 recombinant inbred lines and 41 introgression lines indicated that the metabolite heterosis is primarily contributed by epistasis. In tomato, metabolite profiling in seeds of 76 introgression lines in two consecutive harvest seasons revealed the presence of 30 metabolite quantitative trait loci (mQTLs) and dissected partial mechanisms, underlying the variational contents of main primary metabolites. In this study, 74 metabolites of known chemical structure were quantified. Similar mQTL analyses have been performed in other plant species, such as wheat, rice, and oilseed rape. In QTL analysis, the heritability of given phenotype (broad-sense heritability) and the r^2 of the individual locus linked to a given phenotype (the effect size of locus) are usually evaluated (Hong et al. 2016).

1.9 Review of methods used in systems biology

Computational methods in systems biology aim at normalization and reduction of multidimensional data into networks to provide insight into the various aspects of plant biology. Several types of biological networks are used in systems biology to study mechanisms of complex developmental and physiological processes. These include protein-protein interaction networks, metabolic networks, gene regulatory networks and gene co-expression networks. Gene regulatory networks are used to study interactions among genes underlying biological processes. Genes that interact with many other genes are called hub genes (Allen et al. 2012).

Gene expression network based on similarity are called gene co-expression networks. In addition to functional annotation of unknown genes, this method can be used to infer gene regulatory relationships using prior knowledge and data integration. This method is useful to cluster genes with similar expression patterns across different samples and/or conditions. The method of choice to analyze large transcription datasets is to construct co-expression networks. Co-expression network analysis is useful for clustering thousands of genes with similar expression patterns across multiple conditions or across many samples. Prioritization or ranking technique of genes within the same cluster allows finding hub genes or a few nodes with high connectivity. Gene annotation tools are also important aspects of gene prioritization. Modules are clusters of potentially biologically interrelated genes and functional annotations further confirm biological meaning. Correlation networks have undirected edges, since no causality can be inferred from two connected genes. Regression methods find directed edges and Bayesian networks also allow the inclusion of prior

knowledge, but their application is computationally challenging and not feasible for large datasets (Yuan et al. 2008, Serin et al. 2016).

There are two main steps in the analysis of gene co-expression networks. The first step is to calculate a similarity score between two pairs of genes. This step of co-expression network construction measures the relationship between each pair of expressed gene to infer the strength or weakness of relationship between the expressed genes, called a similarity score. A similarity score is a correlation coefficient calculated from the pairwise comparison of the gene expression patterns for each possible pair of genes. There are different methods of correlation. However, simple Pearson correlation is often used and performs well in finding gene relationships on large data sets (Serin et al. 2016). The next step is to find a suitable threshold to select expressed genes to be connected to each other. Soft thresholds produce weighted gene networks that preserve the underlying continuous nature of the correlation (Zhang and Horvath 2005).

Several systems biological network databases are available for the model plant *A. thaliana* such as metnet online (Sucaet et al. 2012), AraCyc metabolic network, protein-protein interaction tools (www.arabidopsis.org) which integrate pathway information from prediction and also those proven knock out etc experiments. A recently developed A. *thaliana* gene co-expression database integrates data from all expression studies available (Aoki et al. 2016). Data visualization tools are also important. Some tools such as cytoscape (Cline et al. 2007) integrate network construction, annotation and visualization tools. *B. napus* is a recently sequenced crop so annotation depends mostly on other organisms. Blast2GO tool for annotation (Conesa et al. 2005) and GOstats for gene ontology enrichment (Gentleman and Falcon 2013) are suitable tools for non model organisms such as *B. napus*.

1.10 Weighted Gene Co-expression Network Analysis (WGCNA)

Several gene co-expression network databases exist for model plant *A. thaliana* and crops such as rice (Serin et al. 2016). Some tools such as *CoExpNetViz* work on any species but input only a subset of genes. For non-model polyploid crops such as *B. napus*, custom co-expression network construction tool is needed for analysis of a large dataset of expressed genes.

Weighted Gene Co-expression Network Analysis (WGCNA) is a method for construction of gene co-expression networks in large dataset of expressed genes (Allen et al. 2012). It is also useful to reduce large amounts of data into module eigengenes as a summary of the

expression of interconnected gene networks. Module eigengene (ME) can be defined as the first principal component of the expression profile of a module (cluster), a value obtained computationally to reduce the data dimension without loss of information and representing much of the variation in the data. Moreover, the method is useful to correlate module eigengenes to trait data. Furthermore, connectivity based on module eigengene values (kME) can be obtained by correlating the gene expression profile for each gene individually with the module eigengene.

Eigengene-based connectivity (kME) can be used as a measure of intramodular connectivity (kIM) in detecting highly interconnected hub genes within a cluster. Since kME and kIM are very strongly correlated, kME is preferred because it solves computational challenges with large datasets, has an associated p-value, and kMEs can be compared across modules since they all lie between -1 and 1. The membership of a gene with respect to the module is also measured by kME. If kME is close to 0, the gene is not part of the module. If kME is close to 1 or -1, it is highly connected to the module. The sign of kME encodes whether the gene has a positive or a negative relationship with the module eigengene. To relate individual genes to each trait, gene significance (GS), obtained as absolute value of the correlation between the trait and the expression profile of each gene, can be used. Although all genes in a given module are interconnected with each other, the kME values can be used to rank interconnectivity and identify network drivers (hub genes) (Langfelder and Horvath 2008).

To relate individual genes to each trait, gene significance (GS), obtained as correlation between the trait and the expression profile of each gene, can be used. Thus, GS and kME values can be used to identify genes that have a high significant correlation of expression profile with the traits of interest (Langfelder and Horvath 2008).

In addition to the approach of analyzing a single network as mentioned above, WGCNA also offers the option where one can compare the networks of expression profiles of genes in different networks called consensus analysis. The resulting modules are present in all datasets and called consensus modules (Langfelder and Horvath 2008). This method was later on expanded to include meta-analysis of kME and GS (Langfelder et al. 2013). Such analysis is useful to study gene co-expression network among different tissues and different species. The resulting groups of genes are called consensus modules, consisting of genes with similar expression profiles in all the datasets. Consensus analysis keeps the same gene in a single module, but different sets of consensus module eigengenes are generated for each dataset. This is because a given module (set of genes) has a unique expression profile in the different

datasets. The final step of the analysis is to correlate the eigengenes to the traits of interest. Correlation coefficients with the same sign, that is, when both eigengenes from corresponding modules which are positively or negatively correlated to the trait in both the tissues are considered. Then, the lowest correlation coefficient could be considered for screening modules of interest that relate to traits in both the gene expression networks.

WGCNA was first developed as a tool for systems biology approach in medicine. It has been applied to medical research to find gene expression modules associated to diseases and link these to quantitative trait loci (QTL) in Humans (Voineagu et al. 2011, Horvath et al. 2006) and mice (Ghazalpour et al. 2006, Fuller et al. 2007). Later on, this method has also been useful to relate gene expression networks of plants to agronomic traits of interest in the model plant *A. thaliana* (Weston et al. 2008), and in crops such as rice (Ficklin et al. 2010, Smita et al. 2013) and oilseed rape (Harper et al. 2012, Körber et al. 2015). It has also been used to link expression networks of oilseed rape to seedling development trait QTL (Körber et al. 2015).

1.11 Aims of the thesis

In the first part of the study a transcriptome-based systems analysis of germination and seedling emergence in a panel of 42 winter-type *B. napus* genotypes with high, low and intermediate germination rate was performed. Using a WGCNA approach seedling root and shoot gene expression modules were correlated to multiple germination traits, assayed in automated, high-throughput phenotyping experiments, and to multi-environment field emergence data. Seed lots from different seed production environments were also considered, adding a further layer of environmental complexity to the phenotypic correlation analyses. Genome positions of hub genes with high intramodular connectivity, identified by the WGCNA analysis, were compared to quantitative trait loci (QTL) for emergence-related traits. Candidate genes with putative network regulatory roles during seedling emergence, that also lie within QTL intervals for seed germination and vigour, contribute to understanding the control of this important trait complex (Figure 1).

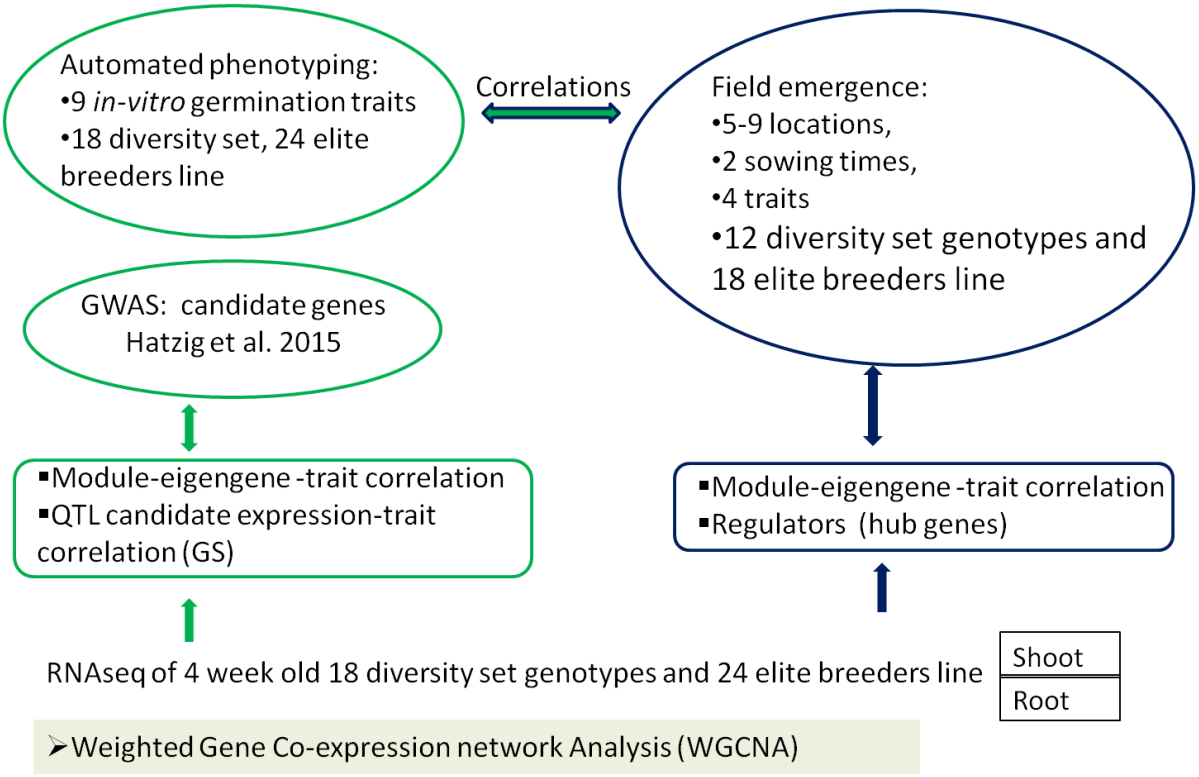


Figure 1. Schematic representation of the material and methods used in the study 'Complex multifunctional gene expression networks influence seedling development and vigour in *Brassica napus* '.

The aim of the second study was to investigate osmotic stress related gene expression and metabolite responses in the same diverse set of 42 *B. napus* genotypes used in the first part of the study by using weighted gene co-expression network analysis. In this part of the thesis, gene expression and metabolite profiling in four week old seedlings subjected to one week of osmotic stress induced by polyethylene glycol (PEG). Expressed genes with similar fold change patterns in osmotic stress were clustered using the WGCNA approach. The resulting module eigengenes were correlated to fold changes of the metabolites and biomass in osmotic stress. This will contribute to identify the genes and metabolites affecting osmotic stress response as potential markers for crop improvement (Figure 2).

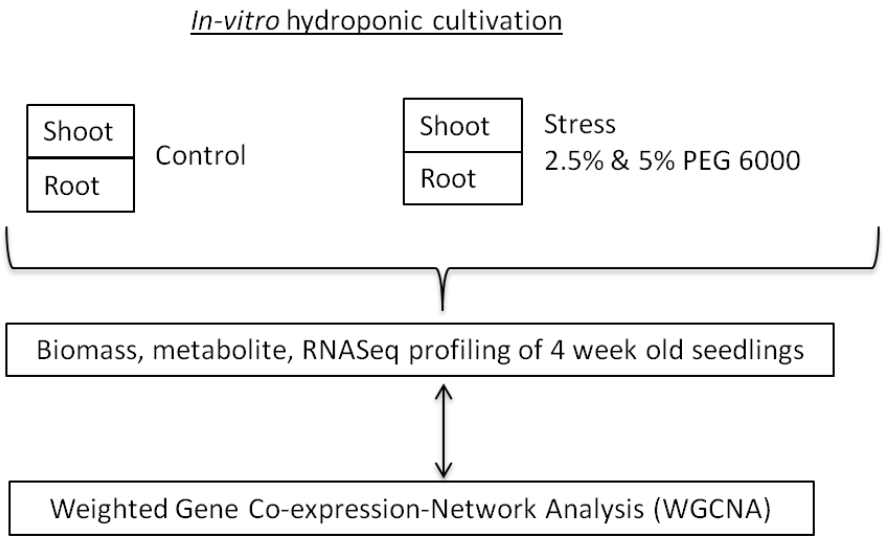


Figure 2. Schematic representation of the material and methods used in the study 'Osmotic stress responsive gene expression networks in *Brassica napus* seedlings *'*.

2. Materials and methods

2.1 Selection of plant material based on high-throughput germination phenotypes

Automated high-throughput phenotyping of germination and seedling vigour was performed for a set of 248 genetically diverse winter oilseed rape genotypes with seeds from two independent production environments (Hatzig et al. 2015) and for 24 breeder's elite lines using the imaging-based system of the French national seed testing laboratory (Groupe d'Etude et de Contrôle des Variétés et des Semences) in Angers, France (Wagner et al. 2011). Hereafter the measured traits using this automated high-throughput phenotyping system are called *in-vitro* germination traits. All plant materials used were advanced inbred lines generated by repeated self-pollination to ensure maximal homozygosity. For all parameters the mean value from 100 biological replications (100 single seeds) per seedlot was used as the trait input value for downstream analyses. Phenotyping data from 42 genotypes of this experiment (18 genotypes from the diversity set and 24 breeder's elite lines) were used in this study. These included time necessary to reach 50% of germination (T50), germination rate after 72 hours (GR72) and uniformity (T80-T20), germination rate at 36 h (GR36), elongation speed (ES), mean germination time (MGT), volume increase after 8 hours (VI), thousand seed weight (TSW) and imbibition speed (IS).

Based on the high-throughput phenotyping of the diversity panel mentioned above, a set of 18 genotypes from a diversity set and 24 breeder's elite lines spanning the observed broad range

of germination speed and vigour were selected for detailed field phenotyping and for the transcriptome study (Appendix 1). Selection of high, intermediate and low vigour accessions was based on the overall performance of genotypes with regard to time necessary to reach 50% of germination (T50), germination rate after 72 hours (GR72) and uniformity (T80-T20).

2.2 Multi-environment phenotyping of field emergence

The 36 of the 42 RNA sequenced genotypes were assessed for field emergence in field trials at 5 to 9 locations across Germany (Einbeck, Hohenlieth, Leutewitz, Malchow, Seligenstadt and Asendorf) and France (Le Rheu, Prémesques and Vierville) in the 2011/2012 growing season. To capture broad variation caused by differences in soil moisture and temperature, two replications per genotype and location were sown in observation plots in August (early sowing time, E_) and September (late-sown variants, L_), respectively. At 2 and 4 weeks after emergence the plant number (NPL) and leaf biomass (LBIO) were recorded for each replicate. Development before winter (DBW) was scored when plants stopped growing. Development after winter survival was measured as DAW. For each location and sowing date, rainfall and soil moisture data were collected during the field trials.

2.3 *In-vitro* cultivation

Plant samples from each of the genotypes were generated by *in vitro* cultivation in a fully climate-controlled growth chamber (day period: 16 h light, 16°C, 75% relative humidity; night period: 8 h/12°C/65% RH), using a protocol described in detail by (Hatzig et al. 2014). The osmotic stress treatment was also done according to the protocol of (Hatzig et al. 2014) using polyethylene glycol (PEG 6000). After three days of germination, seedlings were transferred to the hydroponic cultivation system and grown in MS medium for 24 days without stress. While control samples were continued to be cultivated in MS medium only, stressed samples were treated with 2.5% PEG from day 24 to 27 and 5% PEG from day 27 to day 33 (Figure 3). Samples for transcriptome and metabolome profiling were harvested at day 31 (giving 7 days in stress in total for the stressed samples) and immediately frozen using liquid nitrogen. Because our aim was to assess whole-plant responses on a global network scale in a large number of different genotypes, whole shoot samples from a single time-point at 4-5 leaf stage were used rather than time series or more differentiated tissue samples.

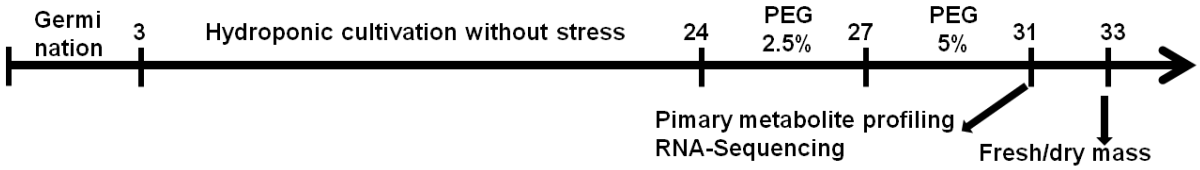


Figure 3. Plant cultivation in hydroponic medium and controlled environment. Control plants were cultivated without the addition of PEG throughout the growth period. Numbers indicate the number of days after sowing.

Leaf surface temperature and chlorophyll content were measured from day 24-32 according to (Hatzig et al. 2014). Samples for biomass measurement were taken at day 33. Relative values of biomass were obtained by dividing the value of stress treated samples by the value of control treated samples. Significance of biomass reduction in stress was tested using Welsch t-test in R (R Core Team 2015).

2.4 Metabolite profiling

Samples for metabolite profiling were taken at day 31. The metabolites measured included sugars, amino acids and organic acids (Table 1). The method of metabolic profiling is described in detail in (Hatzig et al. 2014). Metabolites were measured as µmol/g of dry weight and the mean of two independent extractions and quantifications was used for subsequent analysis.

Table 1. List of metabolites measured for this study

Amino acids	Histidine, Asparagine, Serine, Glutamine, Arginine, Glycine, Aspartate, Glutamate, Beta-alanine, Threonine, Alpha-alanine, Proline, Methylcysteine, Ornithine, Lysine, Tyrosine, Methionine, Valine, Isoleucine, Leucine, Phenylalanine, Tryptophane, GABA
Sugars	Fructose, Glucose, Sucrose, *Myo*-inositol
Organic acids	Succinate, Glycerate, Malate, Quinate

Relative values of metabolites were obtained by dividing the value of stress treated samples by the value of control treated samples. The relative values of the metabolites in osmotic stress were correlated with each other using in R statistical software pearson correlation and using the option "pairwise.complete.obs" to account for missing values. Hierarchical

clustering of metabolites fold change in osmotic stress was done using pearson correlation as distance. Significance of biomass reduction in stress was tested using Welsch t-test in R (R Core Team 2015).

2.5 Transcriptome sequencing and data processing

Samples for RNAseq were taken at day 31, from the same aliquots used for metabolite profiling. Quantitative transcriptome sequencing by 100 bp 3'-end expressed sequence tag (3'-EST) profiling on an Illumina HiSeq 2000 sequencer (Illumina Inc., San Diego, CA, USA) was performed by Eurofins Genomics (Ebersberg, Germany). Briefly, polyA+ mRNA isolated from each total RNA sample was used to prepare cDNA libraries of 3'-EST fragments. Bar-coded 3'-cDNA libraries were sequenced in pools of 12 libraries per channel with a target read depth of approximately 150 million reads per channel (12.5 million reads per genotype). Quality control, mapping, normalization and network construction were done as part of this project (Figure 4).

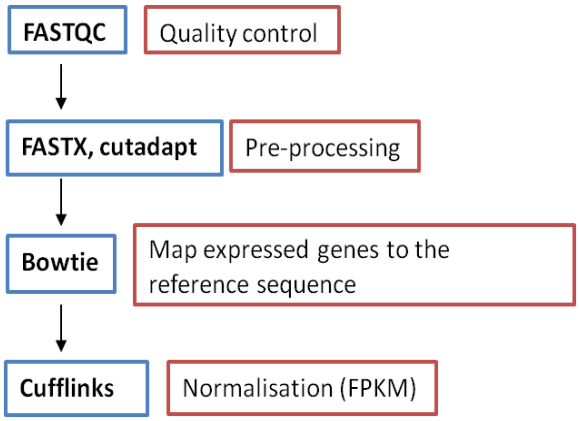


Figure 4. Overview of the RNAseq data analysis workflow

Quality control of the raw sequencing data was performed using the FastQC tool (Babraham Bioinformatics. 2011). The preprocessing tools cutadapt (Martin 2011) and fastx (http://hannonlab.cshl.edu/fastx_toolkit/) were used to remove barcode adapters and improve the quality of the reads, respectively.

Pre-processed reads were mapped to the *B. napus* cv. Darmor-*bzh* reference genome sequence v.4.2 (Chalhoub et al. 2014) using the software Bowtie 2 (Langmead and Salzberg 2012), with slight modifications of the default settings to increase sensitivity. Specifically, the length of the seed substrings was lowered to 15 and the number of mismatches allowed in a seed alignment was set to 1. The setting that gave the highest proportion of unique mapping of

about 50% was chosen. After the release of the draft *B. napus* genome, the standard method to tackle the issue of its polyploid nature is to utilize only uniquely mapped reads that aligned exactly one time to the reference genome. (Chalhoub et al. 2014) have on the basis of this method quantified differential expression between A and C homeologue genes. Using the same approach in the thesis, it can be concluded that the expression of A and C homeologues was accurately quantified. Normalisation and quantification of the expression was performed using the default settings of Cufflinks and assemblies were merged using Cuffmerge (Trapnell et al. 2012). Numbers of fragments per kilobase of transcript per million mapped reads (FPKM values) were used in the subsequent network analysis.

Basic local alignment search tool (BLAST) matching to gene and protein databases, gene ontology (GO) searches, functional protein analysis, pathways and enzyme code annotations from the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.kegg.jp) were used for annotation via the Blast2GO tool (Conesa et al. 2005) (https://www.blast2go.com). BLAST results were linked to the GO database (http://geneontology.org/page/go-database), which contains functionally annotated gene products from different species. Functional protein analysis using the tool InterProScan 5 (Jones et al. 2014) was performed with all relevant databases available via Blast2GO, and then merged with the GO results. Pathway maps were loaded from the KEGG database via Blast2GO.

2.6 Construction of gene co-expression networks for vigour study

For the seedling vigour study, FPKM values of expressed genes from the plant samples cultivated without PEG treatment (controls) were used for the network construction. The Weighted Gene Co-expression Network Analysis (WGCNA) systems analysis tool (Ghazalpour et al. 2006, Langfelder and Horvath 2008), was used for the construction of gene co-expression networks and to correlate co-expression networks to germination and emergence traits. The analysis was implemented under the R language (R Core Team 2015) using R scripts structured for WGCNA analysis (Langfelder and Horvath 2008).

The FPKM values of the root and shoot tissues were analysed separately. Interconnected genes were clustered based on expression values, leading to weighted gene co-expression networks for 4-week old shoots and roots, respectively. After preprocessing of the expression data into a format suitable for network analysis, the presence of outlier samples was checked. Genes that have either too many missing entries or zero variance were removed. Subsequently, the soft thresholding power β was determined through scale free topology,

independently for shoot and root expression networks, as described by (Zhang and Horvath 2005). A soft-threshold power of 4 was applied in both cases. Block-wise module detection was performed with the lowest computationally possible number of blocks of 4. Pairwise Pearson correlation was used for co-expression similarity measure between all expressed genes. Other parameters were assigned default values according to (Langfelder and Horvath 2008). Expressed genes were clustered into distinct groups, hereinafter referred to as (co-expression) modules, each represented by an arbitrary color.

Genes within a co-expression module have high interconnectivity that might reflect a biological interrelation, such as a common biochemical pathway. Hence, functional enrichment analysis was performed for further inference of the biological meaning of each co-expression module. For this purpose a gene ontology enrichment analysis using the R package GOstats was applied (Gentleman and Falcon 2013). The GO enrichment was determined for all genes in each module which showed a GO mapping result from Blast2GO. Hypergeometric tests were performed for overrepresentation, with a *p*-value cutoff of 0.05. Highly significant GO terms for modules of interest were summarized using the REVIGO tool (Supek et al. 2011) with the options *medium similarity* and *Arabidopsis GO database*. GO network images were drawn using Cytoscape (Cline et al. 2007), while pathway overrepresentation of *A. thaliana* orthologues was performed using MetNet (Wurtele et al. 2007).

All of the 104 individual traits measured in the automated germination phenotyping and the field emergence trials with different locations and sowing dates were included as input traits in the network analysis. For both root and shoot gene expression networks, module eigengenes (MEs) representing the gene expression profiles of a given module were correlated with the phenotypic data from each trait. Modules of particular relevance to trait expression were selected based on ME correlations to traits and subsequent co-localization of genes in the modules to previously determined *B. napus* QTL for germination-related traits was investigated (Hatzig et al. 2015; see below). In addition, in order to further rank the importance of modules with respect to the traits, module significance which is the average absolute gene significance (GS) for all genes in the module, was calculated.

Module membership, also known as eigengene-based connectivity or kME, which measures the degree of connection of each gene within each module, was generated using WGCNA. Genes with high module eigengene-based connectivity (kME), in modules with strong

correlations to traits, are of particularly high importance and were determined as hub genes (Horvath and Dong 2008).

2.7 Co-localization of trait-associated expressed genes to QTL for germination

Hatzig et al. (2015) mapped QTL for germination and seedling vigour, by genome wide association studies, in the same *B. napus* diversity panel (248 genotypes) were used to select the materials for the present study. The GWAS analysis used the same phenotyping platform, so that the detected QTL represent traits corresponding to those analysed in the present co-expression network analysis. The genome positions of all genes from modules with high trait-ME correlation to identify genes that also showed linkage disequilibrium to germination-related QTL were analyzed. Due to computational limitations, the Cytoscape plugin *ExpressionCorrelation* (Saito et al., 2012) was used to compute the intra-modular connectivity between genes localised within germination QTL and hub genes (high kME genes, kME>0.8) in modules of interest, which ranges between 6-22% of total genes in modules. *Hierarchical Clustering Explorer* was used to visualize the FPKM values of hub and QTL genes (Seo and Shneiderman, 2002).

2.8 Construction of gene co-expression networks for osmotic stress study

Relative gene expression values (log2 fold changes) for each of the 41 genotypes listed in Appendix 1, excluding for KWS 10, for each of the shoot and root tissues were obtained as follows:

$$Relative\ gene\ expression = log2[\frac{[(FPKM+1)\text{osmotic stress treated sample}]}{[(FPKM+1)\text{control sample}]}]$$

The log2 fold changes in gene expression for shoot and root were analysed using WGCNA package in R (Langfelder and Horvath 2008). The fold changes were clustered after checking for outlier samples as well as genes and samples with excessive numbers of missing entries. Subsequently, the soft thresholding power β was determined through scale free topology. A soft-threshold power of 4 was applied. Block-wise module detection was performed with the lowest computationally possible number of blocks of 4. Pearson correlation coefficients were determined for all pair-wise comparisons of fold change values across all genotypes. Based on the correlations between pair-wise expression levels, unsigned networks (based on the absolute value of correlations) were constructed in which the connection strength between two genes was weighted with power β. Other parameters were assigned default values

according to (Langfelder and Horvath 2008). Fold changes of different treatments of expressed genes were clustered into modules, each represented by an arbitrary colour.

Genes within a co-expression module have high interconnectivity that reflects a biological interrelation, such as a common biochemical pathway. Hence, functional enrichment analysis was performed for further inference of the biological meaning of each module. For this purpose a gene ontology enrichment analysis using the R package GOstats was applied (Falcon and Gentleman 2007). The GO enrichment was determined for all genes in each module which showed a GO mapping result from Blast2go. Hypergeometric tests were performed for overrepresentation, with a *p*-value cut-off of < 0.05. Highly significant GO terms for modules of interest were summarised using the REVIGO tool (Supek et al. 2011) with the options *medium similarity* and *Arabidopsis GO database*.

All of the metabolites profiled (Table 1) were included as input traits in the network analysis. Module eigengenes (MEs) representing the gene expression profiles of a given module were correlated with the relative values of metabolites. Modules of particular relevance to osmotic stress responses were selected based on functional enrichment analysis and/or correlation of MEs to metabolites.

2.9 Construction of consensus gene co-expression networks for shoot and root

The WGCNA function 'blockwiseConsensusModules' was used to construct consensus network to find modules with similar network between shoot and root gene expression. A soft-threshold power of 7 was applied. Other parameters were assigned default values. Expressed genes with similar network in both tissues were clustered into distinct groups, hereinafter referred to as consensus modules, each represented by an arbitrary color. Genes whose expression profile does not match between shoot and root were kept in the grey module. The overall density of the preservation network was estimated as the mean preservation of eigengenes with each other in the two datasets of shoot and root expression. For each set of gene expression, connectivity based on module eigengene (kME) and correlation of gene expression of each gene with the trait (GS) were analyzed. The Z scores of kME and GS from each set of root and shoot expression profile was used to form a meta-Z score of kME and GS and the corresponding p-value (Langfelder et al. 2013). Genes with highest connectivity within each module were extracted using the WGCNA package function "chooseTopHubInEachModule".

3. Results

3.1 Complex multifunctional gene expression networks influence seedling development and vigour in *Brassica napus*

3.1.1 *In-vitro* germination traits were correlated with field emergence traits

In-vitro germination traits measured in an automated high-throughput phenotyping system show significant medium to high correlations to a number of field emergence traits from several locations. For instance, time taken to reach 10-60% germination (T10-T60) and germination rate at 36 and 72 hours is significantly correlated to leaf biomass at early and late sowing under field conditions. This is obvious when using average values for field emergence traits across all locations for correlation analysis with *in-vitro* germination traits. GR36 measured in the automated high-throughput germination *in-vitro* phenotyping experiments significantly correlated with the average field emergence parameters number of plants (NPL) and leaf biomass (LBIO) assessed in replications at multiple field locations throughout Europe (r = 0.50-0.78). GR36 was also correlated with development before winter (DBW) (r = 0.50-0.65). Germination rate after 72 hours (GR72) correlated with LBIO and NPL (r = 0.50-0.67). Correlation of time necessary to reach 50% of germination (T50) measured in the lab in the automated high-throughput germination phenotyping platform to LBIO and NPL was r= - 0.42 to - 0.76. The highest correlation was GR36 with late sowing LBIO2 (r=0.78).

Pairwise *in-vitro* trait to field trait correlations for MGT, ES, TSW, IS, FG, VI, T50, GR36 and GR72 are significantly correlated to number of plants and leaf biomass field traits in up to 9 of the 9 tested locations (range in Table 1). TSW was significantly correlated to emergence traits in some locations: location Vierville, V_E_DAW (r=0.60), V_L_NPL2 (r=0.42), V_L_NPL4 (r=0.34); location Prémesques, P_L_NPL4 (r=0.41); location Einbeck, E_L_NPL2 (r=0.36), E_L_NPL4 (r=0.35); location Hohenlieth H_E_LBIO4 (r=0.36). In addition, TSW was significantly correlated (r=0.35) to mean development before and after winter (L_DBW and L_DAW).

Table 2. Range across 5 to 9 locations of Pearson correlations (r, upper row = maximum, lower row= minimum) of field emergence traits in different environments (top-to-bottom) and *in-vitro* germination traits (left to right). [#]For some traits, only one location is significant so only one value is shown. [**] Highly significant (P<0.01); [*] significant (p<0.05); - not significant. E_= early sowing time, L_= late sowing time. [§]For the other abbreviations of traits and samples see below.

Trait	MGT	ES	TSW	IS	FG	VI	T50	GR36	GR72
E_LBIO2	-0.70**	0.62**	-	-	-0.49**	-	-0.67**	0.70**	0.54**
	-0.54**	0.35*			-0.40**		-0.57**	0.53**	0.46**
L_LBIO2	-0.71**	0.62**	-	-	-0.64**	-	-0.70**	0.76**	0.55**
	-0.38*	0.34*			0.37*		-0.40*	0.35*	0.46**
E_NPL2	-0.59**	0.43$^{*\#}$	-	-0.44**	-0.36$^{*\#}$	-0.49**	-0.60**	0.58**	0.65**
	-0.34*			-0.42**		-0.36*	-0.35*	0.35*	0.40**
L_NPL2	-0.67**	0.35$^{*\#}$	0.42*	-0.38$^{*\#}$	-0.37$^{*\#}$	-0.46**	-0.68**	0.67**	0.65**
	0.37*		0.36*			-0.40*	-0.38*	0.38*	0.36*
E_LBIO4$^{\#}$	-0.64**	0.71**	0.36$^{*\#}$	-	-0.45**	0.39$^{*\#}$	-0.60**	0.63**	0.58**
	-0.40*	0.36*			-0.32*		-0.38*	0.40*	0.35*
L_LBIO4	-0.68**	0.71**	-	-	-0.57**	-	-0.64**	0.64**	0.48**
	-0.50**	0.43*			-0.36*		-0.49**	0.36*	0.41*
E_NPL4	-0.55**	0.48$^{*\#}$	-	-0.42**	-	-0.50**	-0.56**	0.53**	0.70*
	-0.43*			-0.34*		-0.35*	-0.39*	0.43*	0.34*
L_NPL4	-0.67**	0.43**	0.41*	-	-0.41$^{*\#}$	-0.47**	-0.67**	0.69**	0.65**
	-0.37*	0.38*	0.34*	0.45$^{**\#}$		-0.38*	-0.38*	0.35*	0.52**
E_DBW	-0.57**	0.56**	-	-	-0.44**	-	-0.55**	0.58**	0.40$^{*\#}$
	-0.36*	-0.39*			-0.37*		-0.34*	0.28*	
L_DBW	-0.68**	0.58**	0.35$^{*\#}$	-	-0.50**	-	-0.65**	0.65**	0.49**
	-0.34*	0.35*			-0.41*		-0.40*	0.34*	0.42*
E_DAW	-0.44$^{**\#}$	0.46$^{**\#}$	0.60$^{**\#}$	0.37$^{*\#}$	-	-0.44$^{*\#}$	-0.43$^{*\#}$	0.47$^{**\#}$	-
L_DAW	-0.44**	0.52**	0.49**	0.37$^{*\#}$	-	-	-0.43*	0.44**	0.47$^{**\#}$
	-0.39*	0.41*	0.34*				-0.38*	0.32*	

[§]MGT=Mean germination time (h), TSW= Thousand seed weight (g), ES=Elongation speed (mm/h), T50=Time necessary to reach 50% of germination (h), IS=Imbibition speed (mm3/h), FG= Time when germination begins after a lag period corresponding to imbibition and metabolic repairs (h), VI=Volume increase 8h (%), GR36=Germination rate after 36 hours (%), GR72=Germination rate after 72 hours (%), LBIO2= leaf biomass 2 weeks after emergence, LBIO4= leaf biomass 4 weeks after emergence, NPL2= number of plants 2 weeks after emergence, NPL4= number of plants 4 weeks after emergence, DBW= development before winter, DAW= development after winter.

Since they are correlated to the field emergence traits, the automated germination phenotyping measurements represent a viable proxy to supplement extremely time-consuming, expensive and work-intensive field evaluations of emergence from different winter oilseed rape seedlots. Furthermore, the strong correspondence of *in-vitro* data to field phenotype data from plants 4 weeks after sowing make particularly these data points from the germination *in-vitro* phenotyping relevant for network correlation analyses to expressed genes from 4-week old seedlings grown under field conditions. However, there are outliers in the German location Seligenstadt. This might be due to extreme climate in Seligenstadt as the lowest rainfall from all locations was recorded at two weeks after sowing in Seligenstadt. Within the French locations, field trait data for P_E_NPL4 (location Prémesques, early sowing, number of plants 4 weeks after emergence) shows no significant correlation with any *in-vitro* trait while field trait data for only one out of nine locations showed a significant correlation with an *in-vitro* trait (V_L_NPL4, location Vierville, late sowing, number of

plants 4 weeks after emergence). Vierville had the highest rainfall of all 9 locations. Several locations of development after winter (DAW) did not significantly correlate to *in-vitro* traits.

3.1.2 Co-expression modules from *in-vitro* grown seedlings were correlated with *in-vitro* seed vigour and field emergence traits

For both shoot and root tissue, the number of reads that passed quality control ranged from 5,659,366 to 20,688,573 reads with an average of about 11 million reads and a coefficient of variation of 3 %. The reads aligned to the genome reference ranged from 71 % to 86 % of the total reads that passed the quality control. They mapped for the shoot and root tissue to a total number of 85,498 genes expressed at least in one genotype. This accounts for 85.6% of genes of the total number of 101,040 genes predicted to be encoded by the *B. napus* genome overlapping in expression between the two tissues. After pre-processing of the expression data into a format suitable for network analysis, the presence of outlier samples was checked and none of the genotypes were detected as outlier. After further filtering out of outlier genes, 84,363 expressed genes for root and 82,897 expressed genes for shoot were identified and used for further analysis. Of the total number of 101,040 genes predicted to be encoded by the *B. napus* genome, this represents 83.5% of genes expressed in roots and 82 % of genes expressed in shoots in this study.

Weighted gene co-expression networks were constructed for the shoot and root expressed genes separately. For each network, all possible pairwise Pearson correlations were calculated for 84,363 expressed genes for the root and 82,897 expressed genes for the shoot dataset. Finally, 69 modules were obtained for the shoot expression network and 85 modules for the root expression network. The average number of expressed genes in the shoot expression modules was 1,075 and the median 475 (minimum 50 and maximum 7,590). The average number of expressed genes in the root expression modules was 809 and median was 362 (minimum 33 and maximum 7,276). The number of genes that did not belong to any module due to very unique expression profiles was 9,738 in the shoot and 16,373 in the root expression network. All of the modules have genes in common to both the root and shoot expression network.

The resulting large tables of module eigengene-trait correlations of modules were summarized as a colour-coded matrix according to the correlation coefficients of module eigengenes (ME) with traits. Figure 5 shows the module eigengene -trait correlations of shoot modules that were summarized as a colour-coded matrix according to the correlation coefficients given in

the table as an example. The figure shows that for some gene expression modules (colour coded on the left) a strong conspicuous correlation to many *in-vitro* germination and field emergence traits is present (boxed, represented by dark blue and red colours ordered line-by-line in the matrix). Using this approach a number of shoot and root gene expression networks were identified in four week old seedlings from hydroponic culture associated strongly with both, with *in-vitro* germination and vigour traits of up to 3 day old plants as well as with field emergence traits at two and four weeks after sowing.

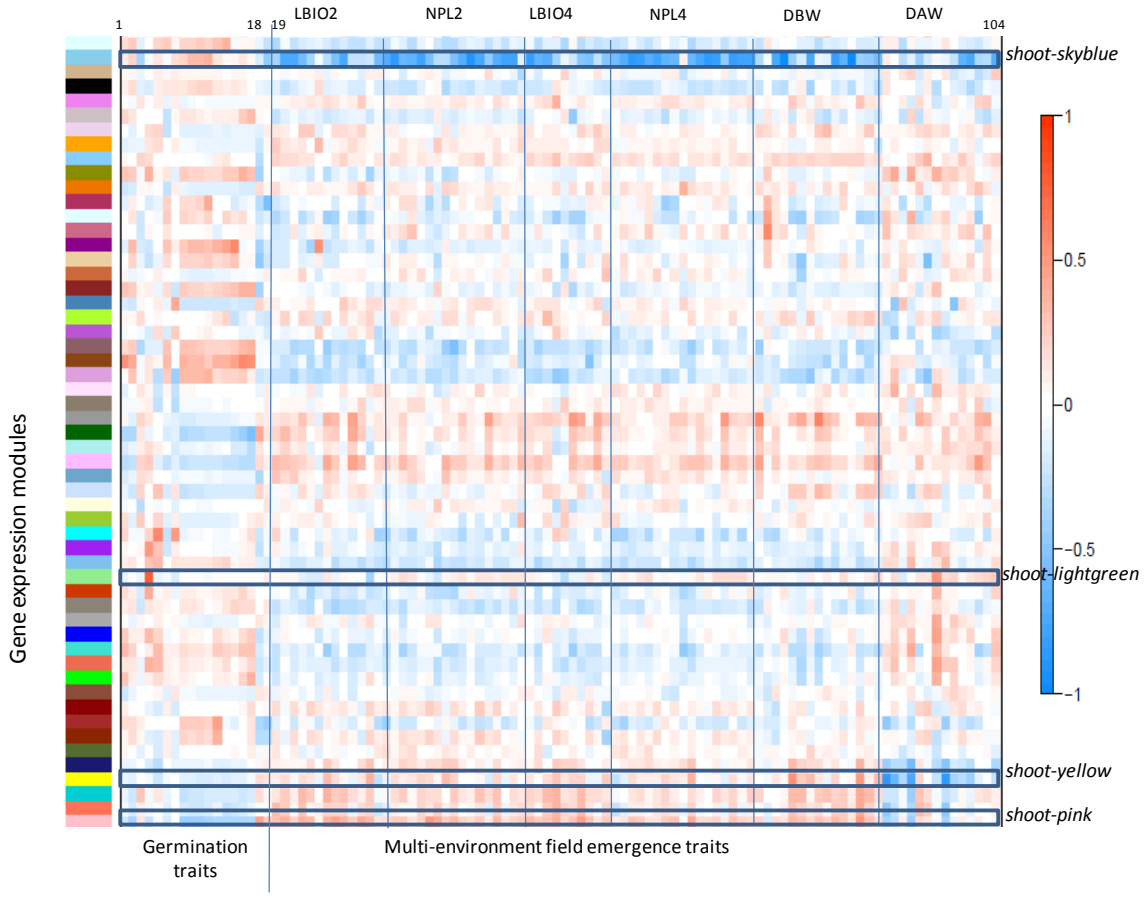


Figure 5. Correlation table showing shoot module eigengene (ME, summarizing seedling expression clusters) correlated with traits.

Each correlation was colour coded by the correlation value to display the data as a table. Each square represents a correlation value between ME and trait. Red represents positive correlation and blue represents negative correlation. Scale of correlation corresponds to intensity of color.

In total, 18 *in-vitro* germination traits (Number 1-18) and 86 field emergence traits (Number 19-104) were included in the network. (LBIO2= leaf biomass 2 weeks after emergence, NPL2= number of plants 2 weeks after emergence, LBIO4= leaf biomass 4 weeks after

emergence, NPL4= number of plants 4 weeks after emergence, DBW= development before winter, DAW= development after winter).
For all the field measurements, every field location and sowing date was included as one trait.
Modules *shoot-skyblue*, *shoot-yellow*, *shoot-lightgreen* and *shoot-pink* were chosen for detailed analysis.

3.1.3 Co-expression network analysis of candidate genes selected from QTL for *in-vitro* germination traits

In an earlier genome-wide association study in a germination *in-vitro* automated phenotyping experiment that included 18 of the 42 accessions which are included in the present study grown in *in-vitro* and in multi-location field studies, (Hatzig et al. 2015) 562 genes were found inside intervals of linkage disequilibrium (LD) containing significant marker-trait associations for a number of germination-related and seed vigour traits, namely: TSW, VI, FG, T50, GR36, GR72 and ES (Hatzig et al. 2015). Out of these genes, a high proportion of 367 genes (65 %) were also found in this study in modules of the shoot expression network and 374 genes (66 %) were found in modules of the root expression network. In total, 54 modules of the shoot and 51 modules of the root expression networks were identified which include at least one gene in LD with the QTL for germination-related and seed vigour traits described by (Hatzig et al. 2015). Among these 54 identified shoot network modules containing at least one gene from the QTL intervals, 8 modules were enriched for regulation of germination and/or seedling development. Similarly, among the 51 identified root network modules containing at least one gene from the QTL intervals, 12 modules were enriched for germination and/or seedling development. The trait thousand seed weight (TSW) mapped by QTL analysis from *in-vitro* germination experiments before was identified in this study to show the highest module eigengene correlation compared to any of the 9 measured *in-vitro* germination traits (*shoot-lightgreen*, r=0.79). The identified expression module *shoot-lightgreen* also had the highest module eigengene correlation r=0.60 to field emergence V_E_DAW (Vierville, early sowing, development after winter). The traits TSW and DAW were significantly correlated (Table 2).

To investigate if the genes selected from within QTL intervals based on their annotation as candidate genes for strong involvement in seed emergence and seedling vigour traits by (Hatzig et al. 2015) can be further supported by WGCNA, the individual effects of all expressed genes on traits (gene significance, GS), their interconnectivity with other genes

(module eigengene based connectivity, kME) in modules correlated with QTL traits (Eigengene-trait correlation) and their co-localization within QTL intervals were compared (Table 3). Some genes such as, *Bna.ARR4*, *Bna.ATE1* and *Bna.SCO1*, were selected as candidate genes for germination-related and early vigour traits by (Hatzig et al. 2015) because of their annotation to germination. For candidate gene *Bna.ARR4*, localized in a QTL region for the trait T50, the ME-trait correlation for module *shoot-pink* with T50 was the highest negative correlation of -0.33 from all identified modules with a gene significance of 0.21 (Table 3). Gene significance (GS) is describing the correlation of each gene expression profile across 42 genotypes with a particular trait. Also gene *Bna.ARR4* is localized in a QTL region for the trait GR72. The WGCNA analysis revealed a gene significance of 0.27 for correlation of the expression profile with this trait within the shoot tissue. Thus, WGCNA analysis is further supporting the role of this candidate gene for involvement in traits T50 and GR72 which was initially identified by QTL analysis and selected as a candidate gene based on germination-related annotation information. However, for other candidate genes selected from QTL intervals by (Hatzig et al. 2015), e.g. for *Bna.SCO1* (GS 0.15 for shoot and root tissue) and *Bna.ATE1* (GS -0.08 to 0.05 for shoot and root tissue) selected as candidate genes from two QTL intervals for elongation speed (ES) the ME-trait correlation GS were found to be rather low (Table 3) suggesting that other genes from the QTL interval might be more relevant as candidate genes involved in the respective traits.

Table 3. Comparison of measured WGCNA parameters for candidate genes suggested from QTL mapping of *in-vitro* germination experiments (Table 4 in (Hatzig et al. 2015) and network analysis in this study. * Description of trait-location is given below. NA=not found in the network or without *A. thaliana* orthologs. #Genes found in the grey module were genes that did not belong to any module.

WGCNA clusters	Trait information			Gene and QTL information			
Module	Trait	ME-trait correlation	Highest ME-trait correlation for this trait among all 69 identified modules	GeneID of gene co-localizing in QTL with the same trait in (Hatzig et al. 2015)	Gene description	Gene significance (correlation with trait)	Connectivity within module (kME)
shoot-pink	T50	-0.33	0.47, -0.33	*BnaA09g48160D* (*AT1G10470*) (*Bna.ARR4*)	response regulator 3	0.20	-0.47
Root-grey#	T50	-0.10	0.47, -0.29	*BnaA09g48160D (AT1G10470) (Bna.ARR4)*	‘’	0.00	0.45
Root-black	T50	-0.03	0.47, -0.29	*BnaC06g31150D (AT1G70070)*	dead-box atp-dependent rna helicase chloroplastic-like	0.16	0.55
Shoot-grey#	T50	-0.33	0.47, -0.33	*BnaC06g31150D (AT1G70070)*	‘’	0.21	-0.28
shoot-pink	GR72	0.26	-0.53, 0.26	*BnaA09g48160D (AT1G10470) (Bna.ARR4)*	response regulator 3	-0.27	-0.47
Root-grey#	GR72	0.16	-0.48, 0.25	*BnaA09g48160D (AT1G10470) (Bna.ARR4)*	‘’	0.04	0.45
Root-black	GR72	0.02	-0.48, 0.25	*BnaC06g31150D (AT1G70070)*	dead-box atp-dependent rna helicase chloroplastic-like	0.09	0.55
Shoot-grey#	GR72	0.20	-0.53, 0.26	*BnaC06g31150D (AT1G70070)*	‘’	-0.25	-0.28
NA	GR72	NA	NA	*BnaC09g40630D* (NA)	transcriptional factor b3 family protein	NA	NA
NA	GR72	NA	NA	*BnaC09g40640D* (NA)	transcriptional factor b3 family protein	NA	NA

Table 3 continued.

WGCNA clusters	Trait information			Gene and QTL information			
Module	Trait	ME-trait correlation	Highest ME-trait correlation for this trait among all 69 identified modules	GeneID of gene co-localizing in QTL with the same trait in (Hatzig et al. 2015)	Gene description	Gene significance (correlation with trait)	Connectivity within module (kME)
shoot-darkgreen	ES	0.15	0.40, -0.40	*BnaA09g12770D (AT1G62750) (Bna.SCO1)*	elongation factor chloroplastic-like	0.15	0.67
Root-black	ES	0.14	-0.40, 0.39	*BnaA09g12770D (Bna.SCO1) (AT1G62750)*	”	0.15	0.77
shoot-red	ES	0.04	0.40, -0.40	*BnaA10g24850D (AT5G05700) (Bna.ATE1)*	arginyl-trna-protein transferase 1-like	-0.08	0.55
root-yellow	ES	-0.01	-0.40, 0.39	*BnaA10g24850D (AT5G05700) (Bna.ATE1)*	”	0.05	0.75
Root-green	ES	0.00	-0.40, 0.39	*BnaA10g24780D (AT5G05780)*	probable 26s proteasome non-atpase regulatory subunit 8a	0.01	-0.71
Shoot-turquoise	ES	-0.01	0.40, -0.40	*BnaA10g24780D (AT5G05780)*	‘’	-0.16	0.81
NA	ES	NA	NA	*BnaA09g12820D (AT1G62680)*	pentatricopeptide repeat-containing protein	NA	NA
Root-turquoise	ES	0.04	-0.40, 0.39	*BnaA09g12800D (AT1G62700)*	nac domain containing protein 26	0.04	0.18
Shoot-pink	ES	0.12	0.40, -0.40	*BnaA09g12800D (AT1G62700)*	‘’	0.19	0.40
Root-turquoise	ES	0.04	-0.40, 0.39	*BnaA09g13040D* (NA)	mitochondrial import inner membrane translocase subunit tim9-like	0.37	0.65
Shoot-blue	ES	0.12	0.40, -0.40	*BnaA09g13040D* (NA)	‘’	-0.02	0.84
Root-salmon	ES	0.32	-0.40, 0.39	*BnaA09g13040D* (NA)	‘’	0.37	0.92
Shoot-tan	ES	-0.09	0.40, -0.40	*BnaA09g13040D* (NA)	‘’	-0.02	0.58

Table 3 continued.

WGCNA clusters	Trait information			Gene and QTL information			
Module	Trait	ME-trait correlation	Highest ME-trait correlation for this trait among all 69 identified modules	GeneID of gene co-localizing in QTL with the same trait in (Hatzig et al. 2015)	Gene description	Gene significance (correlation with trait)	Connectivity within module (kME)
Root-turquoise	TSW	-0.22	0.73	*BnaA06g34100D (AT2G02560)*	cullin-associated nedd8-dissociated protein 1-like	0.20	0.46
Shoot-black	TSW	-0.18	0.79, -0.25	*BnaA06g34100D (AT2G02560)*	‘’	-0.004	0.62
NA	TSW	NA	NA	*BnaA06g33970D (AT2G02850)*	basic blue	NA	NA
Root-turquoise	TSW	-0.22	0.73	*BnaA06g33830D (AT2G03120)*	signal peptide peptidase isoform 1	0.21	0.81
Shoot-lightgreen	TSW	0.79	0.79, -0.25	*BnaA06g33830D (AT2G03120)*	‘’	-0.28	-0.35
Root-blue	TSW	-0.12	0.73	*BnaA06g33830D (AT2G03120)*	‘’	-0.11	0.49
Shoot- black	TSW	-0.18	0.79, -0.25	*BnaA06g33830D (AT2G03120)*	‘’	-0.28	0.32
Root-brown2	TSW	-0.05	0.73	*BnaA06g33880D (AT2G03050)*	transcription termination factor family protein	0.10	-0.77
Shoot-green	TSW	0.10	0.79, -0.25	*BnaA06g33880D (AT2G03050)*	‘’	0.10	0.28
Root-green	TSW	-0.08	0.73, -0.36	*BnaC06g19160D (AT1G80410)*	n-alpha-acetyltransferase auxiliary subunit-like	0.02	0.64
Shoot-cyan	TSW	0.22	0.79, -0.25	*BnaC06g19160D (AT1G80410)*	‘’	0.12	0.73
Root-tan	FG	-0.17	0.37, -0.28	*BnaC07g31020D (AT5G61130)*	glucan endo- -beta-glucosidase-like protein 3	0.24	0.61
Shoot-pink	FG	-0.18	-0.37, 0.34	*BnaC07g31020D (AT5G61130)*	‘’	-0.16	0.42

* TSW= Thousand seed weight (g), ES=Elongation speed (mm/h), T50=Time necessary to reach 50% of germination (h), GR36=Germination rate after 36 hours (%), GR72=Germination rate after 72 hours (%).

3.1.4 Identification of new candidate genes from QTL for *in-vitro* germination and seedling vigour traits by co-expression network analysis

To investigate if WGCNA can support filtering of candidate genes from QTL intervals by adding a non-biased quantitative filtering criteria the gene significance of all genes harboured within the QTL intervals was compared with the WGCNA candidate genes selected by (Hatzig et al. 2015) (Table 4).

The expression profiles for 12 of the genes harboured in QTL intervals correlated significantly (r=0.65 to 0.40) to the trait of the respective QTL, with higher Pearson correlation compared to the candidate genes selected by (Hatzig et al. 2015) (compare Table 3 and Table 4). QTL genes with GS was higher than 0.40 could be found from the network analysis (Table 4) compared to GS between 0 and 0.37 for the genes selected from QTL regions purely based on annotation (Table 3). These 12 genes represent a small fraction of 2.1 % of all the genes harboured within the QTL intervals. Module eigengene-based connectivity (kME) was analysed to determine the hub genes defined as genes with highest connectivity to other genes in the same module. The kME of these genes within QTL trait-associated modules ranged from 0.50 to 0.85 suggesting that many of them are hub genes within the respective modules. The highest GS for any *in-vitro* germination traits was found for TSW with r= 0.65 in shoot expression networks. For TSW, GS greater than 0.40 was found for all chromosome regions, except for TSW chromosome C06 QTL, which has the highest GS.TSW=0.35 for *BnaC06g19020D* (Figure 6). These results suggest that the efficiency of filtering and identification of relevant candidate genes from QTL regions can be strongly improved for all *in-vitro* germination traits by applying WGCNA.

Table 4. Constructed shoot modules filtered first for the 562 genes identified within QTL regions for *in-vitro* germination traits by (Hatzig et al. 2015), then filtered by gene significance (GS) for these associated traits. Genes with GS>0.35 or GS<-0.35 are shown. The table was sorted based on the value of ME-trait correlation for each trait. NA=no *A.thaliana* ortholog. Description of traits: TSW= Thousand seed weight (g), ES=Elongation speed (mm/h), T50=Time necessary to reach 50% of germination (h), GR36=Germination rate after 36 hours (%), GR72=Germination rate after 72 hours (%).

WGCNA clusters	Trait information			Gene and QTL information			
Module	Trait *	ME-trait correlation	Highest ME-trait correlation for this trait among all 69 identified modules	GeneID of gene co-localizing in QTL to same trait in (Hatzig et al. 2015)	Gene name	Connectivity within module (kME)	Gene significance (correlation with trait)
shoot-lightgreen	TSW	0.79	0.79, -0.25	*BnaA05g25150D (Bna.PME3) (AT3G14310)*	pectin methylesterase	0.60	0.60
shoot-purple	TSW	0.48	0.79, -0.25	*BnaA05g25310D (AT3G14170)*	uncharacterized loc101211274	0.50	0.58
shoot-purple	TSW	0.48	0.79, -0.25	*BnaA05g25320D (AT3G14160)*	2-oxoglutarate-dependent dioxygenase family protein	0.50	0.58
shoot-purple	TSW	0.48	0.79, -0.25	*BnaA01g04540D (AT4G32470)*	cytochrome bd ubiquinol oxidase	0.80	0.44
root-black	TSW	0.44	0.73, -0.36	*BnaA06g33960D (AT2G02860)*	sucrose transporter	0.52	0.49
root-black	TSW	0.44	0.73, -0.36	*BnaA01g04540D (AT4G32470)*	cytochrome bd ubiquinol oxidase	0.84	0.47
root-black	TSW	0.44	0.73, -0.36	*BnaC06g19020D (NA)*	coatomer subunit beta -1	0.73	0.35
shoot-coral2	TSW	0.37	0.79, -0.25	*BnaA06g34030D (Bna.UBC2) (AT2G02760)*	ubiquitin-conjugating enzyme e2 2-like	0.83	0.43
shoot-coral2	TSW	0.37	0.79, -0.25	*BnaA06g34020D (AT2G02780)*	probable lrr receptor-like serine threonine-protein kinase	0.83	0.43
shoot-cyan	TSW	0.22	0.79, -0.25	*BnaA06g33780D (AT2G03260)*	exs (erd1 xpr1 syg1) family protein	0.78	0.65
root-blue	TSW	-0.12	0.73, -0.36	*BnaA01g04540D (AT4G32470)*	cytochrome bd ubiquinol oxidase	0.60	0.47
shoot-navajowhite2	T50	0.18	0.47, -0.33	*BnaA09g48170D (AT1G10460)*	germin-like protein subfamily 1 member 1-like	0.60	-0.44
shoot-brown	GR72	-0.41	-0.53, 0.26	*BnaA03g09940D (AT5G58040)*	atp binding	0.84	-0.45
root-turquoise	GR72	0.16	-0.48, 0.25	*BnaA03g10030D (AT5G57850)*	branched-chain-amino-acid aminotransferase-like protein chloroplastic-like	0.68	0.41
root-brown	ES	0.19	0.39, -0.40	*BnaA09g12890D (AT1G62640)*	3-ketoacyl-acyl carrier protein synthase iii	0.70	0.41

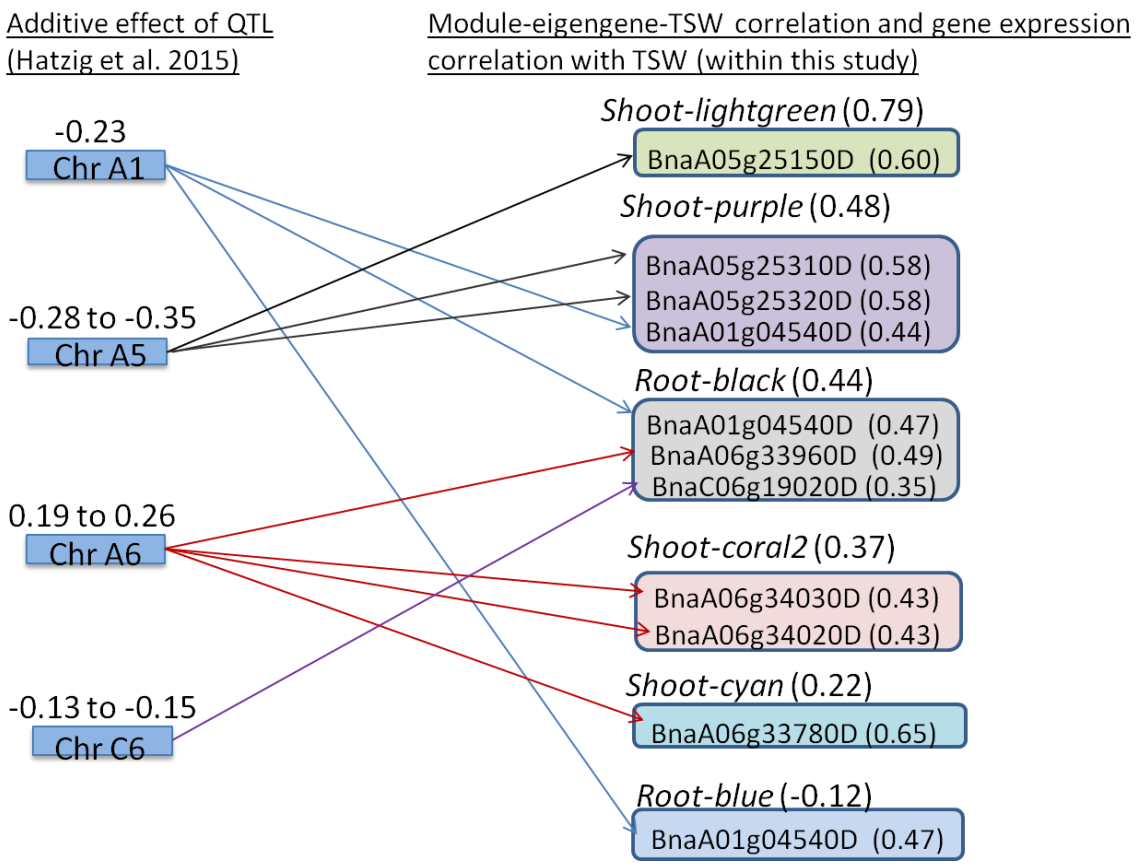


Figure 6. QTL for thousand seed weight (TSW) in *in-vitro* germination experiments, co-expression networks with strongest correlations of ME to TSW and genes with highest gene significance to TSW. The numbers in brackets next to the module color represent correlation r values of the module eigengne to TSW. The numbers in brackets next to the genes represent correlation r values of the genes expression to the trait TSW.

3.1.5 Comparison of QTL associated co-expression networks with *in-vitro* germination traits and field emergence traits

As it has been shown above, a number of co-expression networks and hub-genes from plants grown in hydroponic culture are strongly correlated to *in-vitro* germination traits and a number of the traits from *in-vitro* germination experiments are correlated to field emergence and vigour traits. It was further evaluated if WGCNA can help to identify new key regulator genes associated with these field traits.

To select modules for high multi-trait field correlations, for each of the 104 trait-location combinations investigated, modules with the top 10 highest correlations of the traits with their module eigengenes were extracted. Based on high multi-trait correlations to their module eigengenes, subsets of three modules from shoot tissue were selected for deeper investigations (*shoot-pink*, *shoot-skyblue*, *shoot-yellow*, see boxes in Figure 5).

Shoot-skyblue module correlated to 60 seed vigour and field emergence traits as the top 10 highest. MEs from module *shoot-skyblue*, with a total of 561 expressed genes, were found to be highly correlated to NPL (number of plants), LBIO (leaf biomass), DBW (development before winter) and DAW (development after winter) field emergence traits in multiple environments (r= -0.99 to -0.41). This pattern of ME-trait correlation was in turn correlated to mean rainfall (r=0.45) and mean soil moisture content (r=0.55) corresponding to different locations and sowing dates indicating that expression of genes from module *shoot-skyblue* are highly influenced by the environment. The module *shoot-yellow*, which consisted of 4,212 expressed genes, exhibited the highest ME correlations to DAW, development after winter (r= -0.88 to -0.64). Module *shoot-pink* with 3,742 expressed genes, correlated to 48 seed vigour and field traits as the highest top 10 (r=-0.42 to 0.52).

Similar to the shoot module, based on top 10 trait correlations to module eigengene a subset of four modules from root tissue were also selected for deeper investigation (*root-plum2, root-brown4*, *root-green*, *root-yellow*). The ME of module *root-plum2* with 127 expressed genes showed highest correlations to were found to be highly correlated to NPL (number of plants), LBIO (leaf biomass), DBW (development before winter) and DAW (development after winter) field emergence traits in multiple environments (r= -0.99 to -0.41). The ME of module *root-green* with 3,523 expressed genes showed highest correlations to DAW, development after winter (r= -0.87 to -0.65). ME of module *root-brown4* with 313 expressed genes had top highest correlations to T80.T20 (Time necessary to complete full germination), T50 (Time necessary to reach 50% of germination) and MGT (Mean germination time) with range of r=0.60 to 0.40. ME of module *root-yellow* with 3875 expressed genes showed top 10 highest correlations to 47 seed vigour and field emergence traits (r=-0.63 to 0.30).

The highest correlation of ME with trait for *shoot-skyblue* and *root-plum2* was development before winter with late sowing M_L_DBW (Malchow location, late sowing, development before winter), r= -0.99, with *shoot-yellow* was E_E_DAW (Einbeck location, early sowing, development after winter), r= -0.88 and *shoot-pink* P_L_DBW (Prémesques, late sowing, development before winter), r= 0.52. *Root-green* had highest ME-trait correlation with

E_E_DAW (Einbeck location, early sowing, development after winter), r= -0.87, *root-yellow* with M_E_DAW (Malchow location, early sowing, development after winter), r= -0.63 and *root-brown4* time necessary to reach 80% of germination (T80), r=0.59. Then the mean GS was calculated which describes the gene significance across all investigated traits (104 germination-related and trait-location-sowing date combinations in total). For both the shoot and root networks, the module mean GS of two of the modules with highest ME-field emergence correlation is the highest of all modules and the module mean gene significance of the other 4 is among the top highest 10 and 18 for the shoot and root networks, respectively. GS and kME correlation was high for many of the traits within the selected modules that had highest ME-trait correlation for the module which shows that expression profile of hub genes was highly correlated to the traits.

The modules with high module eigengene correlation to the field emergence traits contained genes from the QTL interval with moderate to low correlation of QTL gene expression (gene significance or GS) with QTL trait and moderate to low module eigengene-QTL trait correlation. Several of the field emergence traits had high GS>0.50 with the QTL gene. On the other hand, lower GS was found with the QTL genes and the QTL traits within the modules highly correlated to field emergence. The gene with the highest GS with QTL trait associated to QTL genes found within module eigengenes with high correlation to field traits in different environments was the QTL gene associated to TSW, *BnaA01g04540D* (*AT4G32470*, cytochrome bd ubiquinol oxidase) with GS.TSW=0.44 found in *shoot-skyblue* module. Even though the module eigengene of *shoot-skyblue* showed very low correlation to TSW (r= -0.06), the QTL gene *BnaA01g04540D* had high intramodular connectivity to top hub genes in *shoot-skyblue*. The rest of the GS to the QTL genes within the field emergence correlated modules was less than absolute value of 0.35. Among the modules highly correlated to the field traits, *shoot-skyblue* had the highest module eigengene correlation to the QTL trait ES, r= -0.40. However, the ES associated QTL gene *BnaA09g12780D* (*Bna.HOP*) within this module had GS of only -0.18. This result suggests that although both the *in-vitro* as well as field germination traits exhibited strong pairwise correlations and correlations with co-expression networks, the co-expression networks were slightly different and putatively different QTL are involved in *in-vitro* and field expression of the emergence and vigour traits. Thus, co-expression networks correlated with field emergence trials were independently investigated in more detail.

3.1.6 Putative regulatory hub genes in field emergence trait-associated modules were enriched for auxin signalling pathway

Functional annotation based on GO and pathway enrichment subsequently provides a means to confirm the biological relevance and significance of the modules of interest. The modules were enriched for different pathways and GO terms. This shows they are comprised of groups of genes that are co-expressed but possibly involved in different, sometimes unrelated pathways. In modules with eigengenes correlated to field emergence traits, enrichment of several highly significant GO terms associated with seed germination, seedling development, auxin and other plant development and growth hormones, growth and plant development were found reinforcing the expectation that these trait-associated modules should be biologically meaningful for seedling growth. Interestingly, stress tolerance GO terms such as 'response to stress' were also enriched, indicating a putative role of stress responses in early plant performance.

The module eigengenes highly correlated to the field seedling emergence were enriched for GO terms associated to auxin and at the same time had hub genes among highest connectivity based on module eigengene (kME) annotated for auxin response and signalling pathway. The modules were also enriched for the pathways 'Auxin signaling' and 'IAA signaling'. Hub genes, or genes with high interconnectivity to other genes in the module, were measured using connectivity based on module eigengene (kME). The proportion of genes with kME>0.80 ranges between 6-22% of total genes in the selected modules. Genes with high intramodular connectivity based on their kME values that were within the top 5% highest kME for each module were specified as hub genes.

Among the root modules selected based on ME-trait correlation, *shoot-yellow* was highly significantly enriched for GO terms 'response to auxin stimulus'. Several hub genes with top 5% highest kME were annotated for 'response to auxin stimulus' and 'auxin mediated signaling pathway'. Module *shoot-pink* showed enrichment for GO terms: auxin mediated signaling pathway. Several hub genes with top 5% highest kME were annotated for 'response to auxin stimulus' and 'auxin mediated signaling pathway'. For module *shoot-skyblue*, the term 'response to stimulus' was highly significantly enriched. Several hub genes with top 5% highest kME were annotated for 'response to auxin stimulus' in this module. *Shoot-skyblue* was also enriched for 'aerobic respiration' and had top hub genes annotated for aerobic respiration. *Root-green* was enriched for GO terms 'auxin transport', 'cellular response to auxin stimulus' and 'auxin mediated signaling pathway'. Several hub genes with top top 5%

highest kME annotated for ‘response to auxin stimulus’ and ‘auxin mediated signaling pathway’. *Root-yellow* was enriched for GO terms ‘response to auxin stimulus’. Several hub genes with top top 5% highest kME annotated for ‘response to auxin stimulus’. Interconnectivity of genes within a module was measured by expression correlation. The highest correlations between expression of two hub genes (r=0.99) were found between all of the auxin pathway ascribed top 5% hub genes in *shoot-skyblue,* two in *shoot-yellow* and *root-green* modules. Very high correlations between expression of two hub genes (r>0.90) were found between most of the auxin pathway ascribed top 5% hub genes: in all of the edges in *shoot-skyblue* (Figure 7), all but one in *shoot-yellow* and *root-green* (Figure 8), two in *root-yellow* and in one in case of *shoot-pink* module.

Several overlapping genes were found between shoot and root modules. Among these, *BnaA06g06460D* (an orthologue of the Arabidopsis gene *AT1G10588*) was identified as a top 5% highest hub gene in modules with high ME-trait correlation *shoot-skyblue* and *root-plum2* (Figure 7). *BnaA09g46830D* an orthologue of *AUXIN SIGNALING F-BOX 3* (*AFB3*), BnaA01g13250D (*SGT1A*) and *BnaA01g00370D* an orthologue of *HOOKLESS 1 (HLS1)* were identified as a common hub gene in modules *shoot-yellow* and *root-green*. The expression of *BnaA06g06460D* (r=0.91), *BnaA01g13250D* (r=0.99) and *BnaA09g46830D* (*AFB3*) (r=0.99), hub genes in the root and shoot were found to be highly correlated and perhaps may have a coordinated function in different tissues (Figure 8).

Auxin signalling genes hub genes were found in modules with high module eigengene correlation to development after (DAW), in addition to other field traits in different environments. Module eigengene trait correlations are summarized in Table 5.

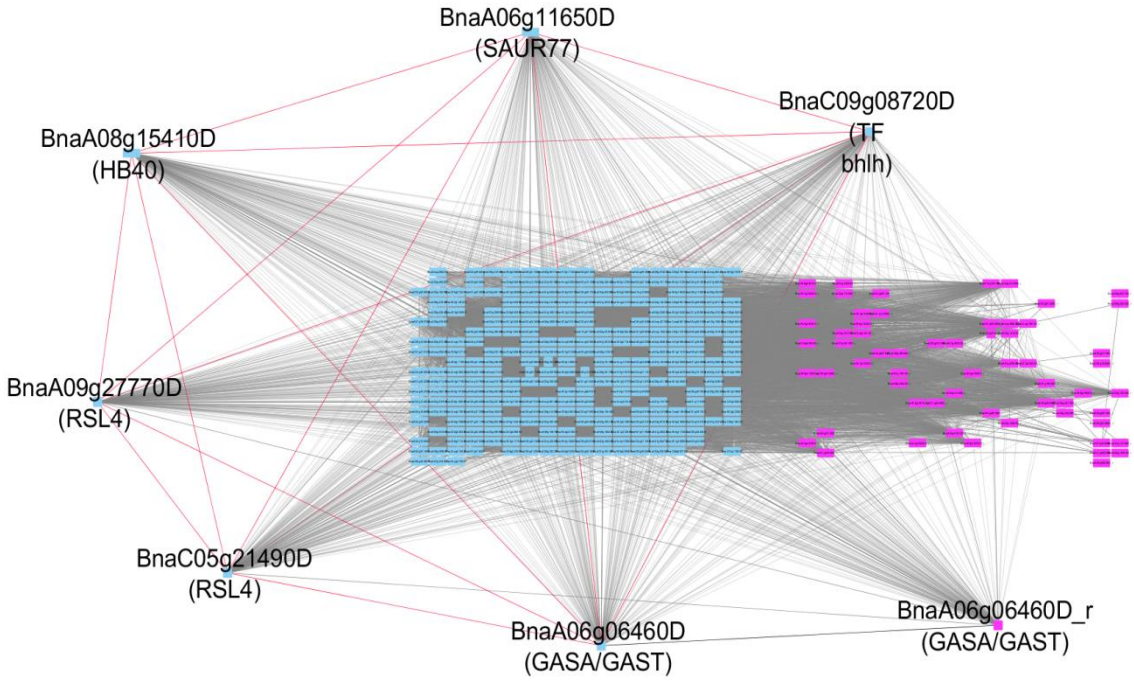


Figure 7. *ExpressionCorrelation* software analyses of gene expression correlation of module shoot *skyblue* and *root-plum2* hub genes ascribed to auxin pathway. *Shoot-skyblue* and *root-plum2* module eigengenes highly correlated to several field emergence traits in different locations. Blue nodes represent *shoot-skyblue* hub genes and pink nodes are hub genes from *root-plum2* (kME>0.8). Label of edges: Red edges represent the high interconnections among hub genes of auxin pathway with r>0.99. The black edge is for the common gene between shoot and root r=0.91. Purple edges represent first neighbours of the hub genes of auxin pathway to other hub genes (r> 0.7). _r=genes from root

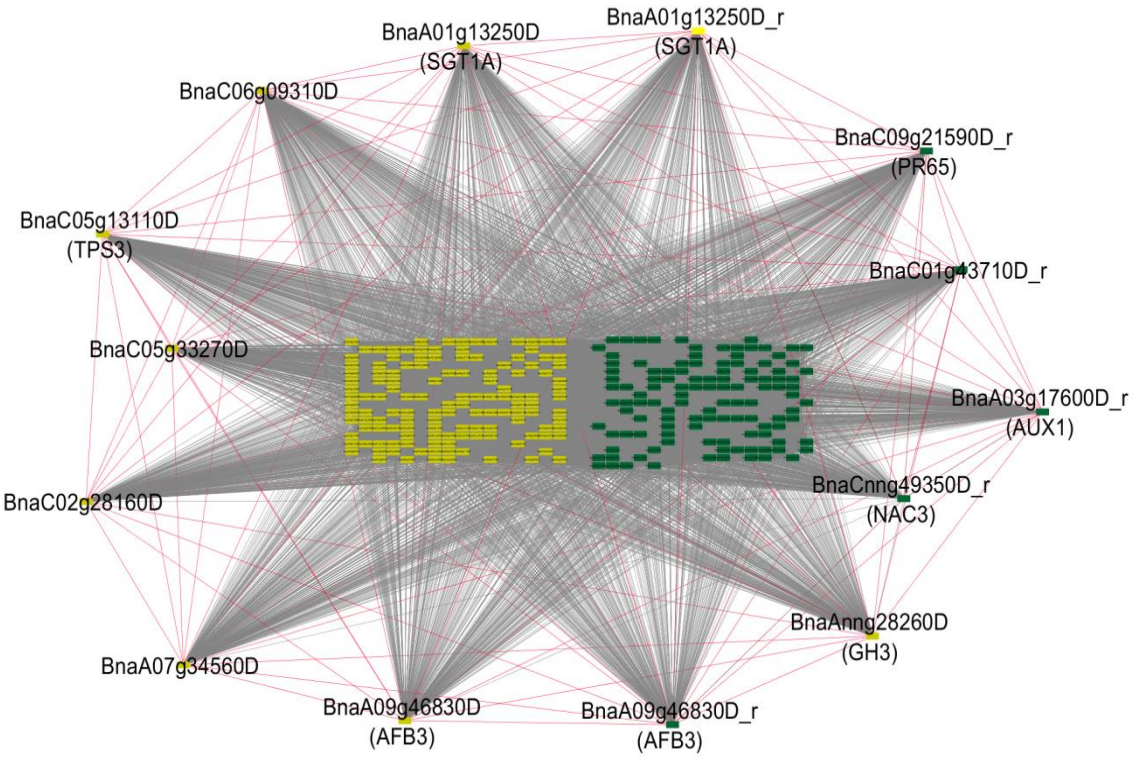


Figure 8. *ExpressionCorrelation* software analysis of gene expression correlation of modules *shoot-yellow* & *root-green* modules hub genes ascribed to auxin pathway. Module eigengenes of *shoot-yellow* and *root-green* highly correlated to development after winter (DAW). Yellow nodes represent *shoot-yellow* hub genes and green nodes are hub genes from *root-green* (top 5%). Label of edges: Red edges represent interconnections among top 5% hub genes of auxin pathway (r>0.81). Grey edges represent first neighbours of hub genes of auxin pathway to other hub genes (r>0.70). AUX1 and GH3 (r=0.95) connected in a KEGG pathway. _r=genes from root.

Table 5. List of shoot and root expression hub genes (top 5%) associated to auxin along with range of correlation to different locations and sowing dates of the module eignengene (ME) of the top 10 highest correlated traits. TF=transcription factor. NA= there is no *A. thaliana* ortholog available, and Blast2go gene description is given. NPL= number of plants measured at 2 and 4 weeks after sowing, early and late sowing; LBIO= leaf biomass measured at 2 and 4 weeks after sowing, early and late sowing; M_E =Malchow location early sowing. Scale for leaf biomass [1-9]: 9 = 0cm, 8 = 5cm, 7 = 10cm, 6 = 15cm, 5 = 20cm, 4 = 25cm, 3 = 30cm, 2 = 35cm, 1 = 40cm and bigger. DAW=development after winter measured in March; DBW=development before winter measured when growing of plants in the field stops because of cold around November. Scale for DBW and DAW [1-9]: 9=best development.

WGCNA clusters	Trait information		Gene information of top 5% high kME hub genes from auxin signalling pathway	
Module	Field Trait*	Range of ME-trait correlation	*B. napus* Gene ID	*A. thaliana* ortholog Gene ID, Gene name and Alias
Shoot-skyblue	DBW DAW NPL LBIO	-0.59 to -0.99 -0.65 to -0.65 -0.44 to -0.89 -0.49 to -0.82	*BnaA08g15410D*	*AT4G36740*, homeobox protein 40 TF (*HB40*)
	‘’	‘’	*BnaA06g11650D*	*AT1G17345*, SAUR-like auxin-responsive protein family (*SAUR77*)
	‘’	‘’	*BnaA09g27770D*	*AT1G27740*, basic helix-loop-helix family protein TF (root hair defective, *RSL4*)
	‘’	‘’	*BnaC05g21490D*	*AT1G27740*, basic helix-loop-helix family protein TF (root hair defective, *RSL4*)
	‘’	‘’	*BnaA06g06460D*	*AT1G10588*, gibberellin-regulated protein
Shoot-yellow	DAW NPL	-0.24 to -0.88 0.23 to 0.37	*BnaAnng28260D*	*AT5G51470*, Auxin-responsive GH3 family protein
	‘’	‘’	*BnaC02g28160D*	*AT4G12500*, Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein (*pEARLI1*)
	‘’	‘’	*BnaA07g34560D*	*AT1G78990*, HXXXD-type acyl-transferase family protein
	‘’	‘’	*BnaA01g13250D*	*AT4G23570* , phosphatase (*SGT1A*)
	‘’	‘’	*BnaC06g09310D*	NA, myrosinase-binding protein
	‘’	‘’	*BnaA09g46830D*	*AT1G12820*, auxin signaling F-box3 (*AFB3*)
	‘’	‘’	*BnaC05g33270D*	NA, pyk10-binding protein 1
	‘’	‘’	*BnaC05g13110D*	*AT1G17000*, trehalose-phosphatase/synthase 3 (*TPS3*)
Shoot-pink	DBW NPL LBIO M_E_DAW	-0.15 to -0.52 0.17 to 0.49 0.20 to 0.48 -0.42	*BnaA03g54130D*	*AT4G37610*, BTB and TAZ domain protein 5 (*BT5*)
	‘’	‘’	*BnaC07g46630D*	*AT4G37610*, BTB and TAZ domain protein 5 (*BT5*)
	‘’	‘’	*BnaC06g40530D*	NA, pyrophosphate-energized vacuolar membrane proton pump 1
	‘’	‘’	*BnaA08g00900D*	*AT1G54100* aldehyde dehydrogenase 7B4 (*ALDH7B4*)
	‘’	‘’	*BnaA06g12480D*	*AT1G18330*, Homeodomain-like superfamily protein (*RVE7/EPR1*), MYB TF

Table 5 continued

WGCNA clusters	Trait information		Gene information of top 5% high kME hub genes from auxin signalling pathway	
Module	Field Trait*	Range of ME-trait correlation	*B. napus* Gene ID	*A. thaliana* ortholog Gene ID, Gene name and Alias
Root-green	DAW DBW NPL LBIO	-0.28 to -0.87 0.34 to 0.53 0.19 to 0.35 0.24 to 0.33	*BnaCnng49350D*	*AT3G29035*, *NAC* domain containing protein 3, TF NAC
	‘’	‘’	*BnaA09g46830D*	*AT1G12820*, auxin signaling F-box 3 (*AFB3*)
	‘’	‘’	*BnaA01g13250D*	*AT4G23570*, phosphatase-related (*SGT1A*)
	‘’	‘’	*BnaC01g43710D*	*AT3G21380*, Mannose-binding lectin superfamily protein
	‘’	‘’	*BnaA03g17600D*	*AT2G38120*, auxin transporter-like protein 2-like (*AUX1*)
	‘’	‘’	*BnaC09g21590D*	*AT3G25800*, protein phosphatase 2A subunit A2 (*PR 65*)
Root-yellow	DAW DBW NPL LBIO	-0.23 to -0.63 0.25 to 0.49 0.17 to 0.41 0.22 to 0.41	*BnaC05g40070D*	*AT3G13300*, Transducin/WD40 repeat-like superfamily protein (*VCS*)
	‘’	‘’	*BnaC05g36990D*	*AT3G16350*, Homeodomain-like superfamily protein, myb transcription factor myb52
	‘’	‘’	*BnaA09g40710D*	*AT2G25930*, protein early flowering 3-like, hydroxyproline-rich glycoprotein family protein (*PYK20/ELF3*)
	‘’	‘’	*BnaC06g12550D*	*AT5G39610*, NAC domain containing protein 6 (*ORE1, NAC2/NAC6 TF*)
	‘’	‘’	*BnaC06g40530D*	NA, pyrophosphate-energized vacuolar membrane proton pump 1
	‘’	‘’	*BnaC06g24340D*	*AT1G72430*, SAUR-like auxin-responsive protein family
	‘’	‘’	*BnaC04g52510D*	*AT2G42620* (*MAX2, MORE AXILLARY BRANCHES 2, ORE9, ORESARA 9, PLEIOTROPIC PHOTOSIGNALING, PPS*)
	‘’	‘’	*BnaA01g06750D*	*AT4G30270*, xyloglucan endotransglucosylase/hydrolase 24 (*XTH24*)
	‘’	‘’	*BnaA07g22640D*	*AT1G73500*, mitogen-activated protein kinase 9 (*MKK9*)
	‘’	‘’	*BnaA02g35860D*	*AT1G73500*, mitogen-activated protein kinase 9 (*MKK9*)
	‘’	‘’	*BnaC05g52020D*	*AT3G14050*, RELA/SPOT homolog 2 (*RSH2*)

3.2 Osmotic stress responsive gene expression networks in *Brassica napus* seedlings

3.2.1 Genotypes showed different responses to osmotic stress

The genotypes reacted differently to osmotic stress in respect to their relative values of one or more metabolites. The metabolites which showed the highest correlations with each other for relative values in shoots under stress were fructose, glucose, *myo*-inositol, and glycine (r>0.9). Alpha alanine also highly correlated to these metabolites (r>0.6). The relative value of shoot biomass in stress was correlated to these metabolites (r=0.5). The other group of metabolites that highly correlated with each other are lysine, tyrosine, valine, isoleucine and leucine (r>0.6). Phenylalanine highly correlated to these metabolites (r>0.6). Beta-alanine with ornithine (r=0.71), phenylalanine (r=0.55), valine (r=0.69) were also correlated with each other. Shoot fructose, glucose, *myo*-inositol, glycine, ornithine, lysine, tyrosine, valine and isoleucine in stress were moderately correlated with shoot biomass.

For roots, the relative values of fructose, glucose and *myo*-inositol under stress were highly correlated with each other (r>0.66). The other group of root metabolites that were highly correlated with each other were isoleucine, leucine, phenylalanine and tryptophan (r>0.7). Methionine and histidine were also highly correlated to these metabolites (r>0.63). Root fructose, glucose and *myo*-inositol in stress were moderately correlated with biomass (r=0.4).

3.2.2 Osmotic stress responsive shoot gene expression networks were correlated to biomass and metabolites

For both shoot and root tissues that were control and PEG treated, the number of reads that passed quality control ranged from 5.6 to 17.3 million reads with a mean value of about 11 million reads and a coefficient of variation of 0.19%. The alignment rate to the genome reference ranged from 71.05% to 86.7% of the total reads that passed the quality control. The reads mapped to a total number of 85,498 genes. After removing the expressed genes with too many missing values, around 83,000 expressed genes were kept for further analysis. This represents 82% of the total number of 101,040 genes predicted to be encoded by the *B. napus* genome (Chalhoub et al. 2014).

The fold change of osmotic stress treated sample over control for 41 genotypes was clustered using WGCNA package in R, generating 70 modules (Figure 9). The number of genes in most of the modules ranged from about 72 to 2000, with a mean value of 1010 and a median of 541. The number of genes that do not belong to any module was 12,987 (grey module).

Several metabolites were found to be highly correlated to module eigengenes (MEs) of the modules and also correlated to biomass. The groups of metabolites which were highly correlated with each other also correlated to the same module eigengene.

GOstats analysis and REVIGO summary revealed that 12 modules were highly significantly enriched for osmotic stress response related terms. Most of the modules enriched for osmotic stress response terms had a significantly high correlation of MEs with one or more of the metabolites. Modules of interest were selected based on the criteria that they were correlated to biomass, that the correlation of their module eigengene with at least one of the metabolites is greater than 0.5 and that the module gene ontology enrichment is related to osmotic stress response. These can be broadly classified as modules whose module eigengene highly correlated to sugars, *myo*-inositol and alpha-alanine (*Shoot-skyblue3*), proline modules (*Shoot-darkorange2* and *Shoot-thistle2*) and the module *Shoot-palevioletred3* whose module eigengene highly correlated to beta-alanine and ornithine (Table 6).

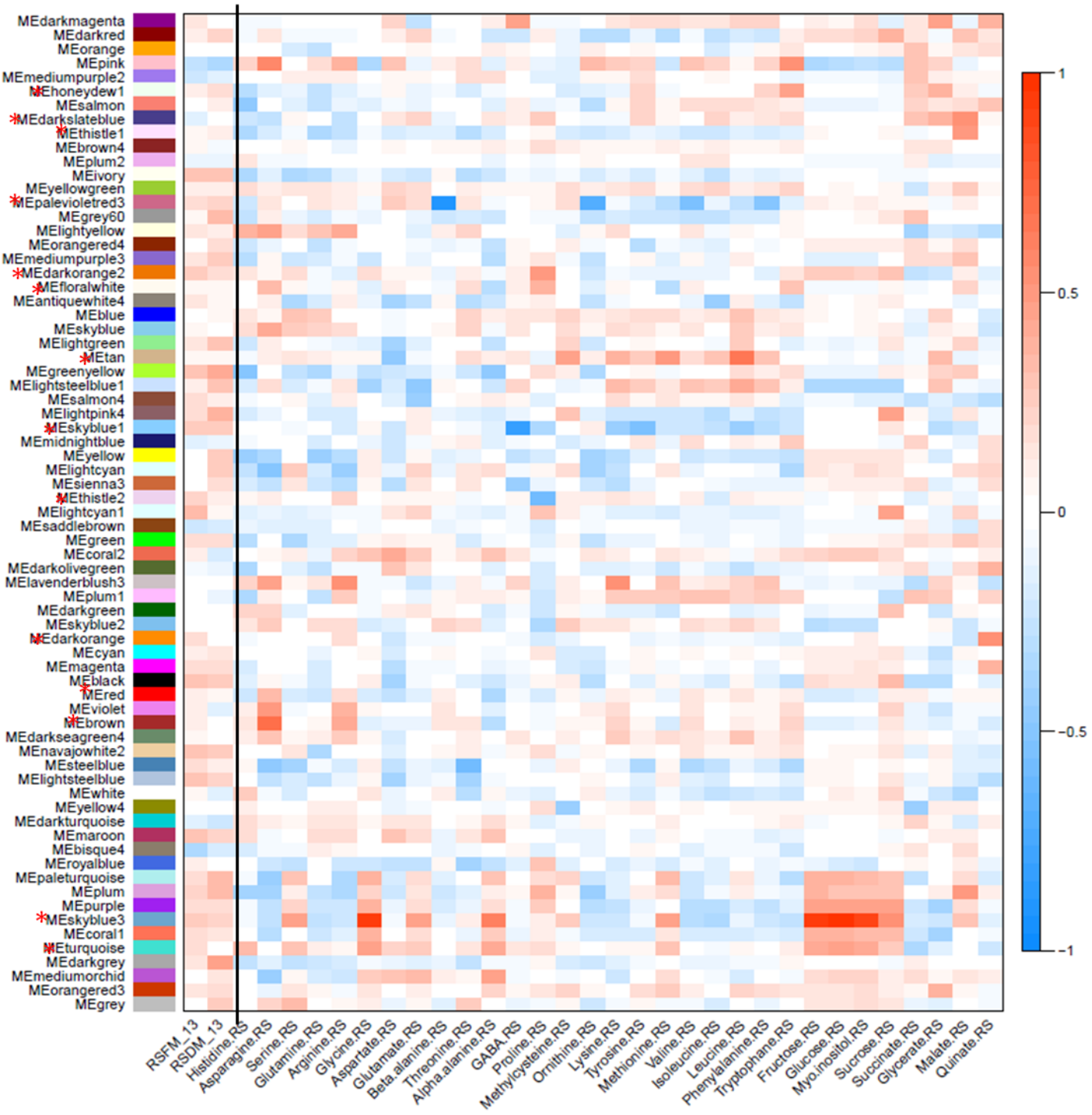


Figure 9. Expression network of fold changes in osmotic stress of shoot tissue correlated with fold change of metabolites, fresh mass (RSFM) and dry mass (RSDM) in osmotic stress. The correlation coefficient is color coded, with blue being a negative correlation, red a positive correlation and white no correlation. *modules with GO enrichment related to osmotic stress responses.

Table 6. Modules with highly significant osmotic stress response related terms enriched and module eigengenes correlated to metabolites and biomass. The correlation of the module eigengene with the metabolite and biomass is given in the third column.

Module	Relevant GO term enriched	Correlation and significance
Shoot-Skyblue3	Drought recovery, homeostasis, ion transport, calcium-dependent protein serine/threonine phosphatase activity, indole acetic acid metabolism	Fructose (r=0.94, p=1e-19), glucose (r=0.96, p=3e-23), glycine (r=0.95, p=1e-20), *myo*-inositol (r=0.92, p=3e-17), alpha-alanine (r=0.62, p=1e-02), sucrose (r=0.52, p=5e-04), glutamate (0.46 p=0.002), serine (r=0.45, p=0.003), methionine (r=0.44, p=0.004), biomass (r=0.5, p=0.0008)
Shoot-Thistle2	Glutamine biosynthesis and metabolism, response to H_2O_2, proline transport	Proline (r= -0.59, p=5e-05), biomass (r=0.2).
Shoot-Darkorange2	Calcium mediated signalling, Response to osmotic stress, response to salt stress, regulation of DNA-templated transcription in response to stress	Proline (r=0.51, p=7e-04), biomass (r=0.26)
Shoot-Palevioletred3	Response to abscisic acid	Beta-alanine (r= -0.91, p=2e-16), Ornithine (r=-0.68, p=9e-07), Valine (r= -0.52, p=5e-04), Phenylalanine (r= -0.50, p=8e-04), biomass (r=0.21)
Shoot-turquoise	Cellular water homeostasis, anion transport, hyperosmotic response, *myo*-inositol biosynthesis, sucrose biosynthesis, glycine biosynthesis	Glycine (r=0.48, p=0.001), alpha-alanine (r=0.42, p=0.007), fructose (r=0.44, p=0.004), glucose (r=0.47, p=0.002), *myo*-inositol (r=0.43, p=0.005), biomass (r=0.25)

3.2.3 Genes in osmotic stress responsive shoot expression networks related to metabolites and biomass

The genes in the selected modules in general consist of genes in response to osmotic stress and genes for the metabolism of the metabolites to which the module expression profile is correlated to.

The module eigengene (ME) of *shoot-skyblue3* is correlated to fructose, glucose, glycine, *myo*-inositol, alpha-alanine and sucrose. The genotype KWS12 had the highest fold change of these metabolites. The genotype whose module eigengene contributed to the variation in this module was KWS12 (Figure 10). When subjected to osmotic stress, this genotype increased in these metabolites and reduced in biomass in one of the cultivations in the osmotic stress experiment. This module contains hub genes that are responsive to osmotic stress. These include *BnaC03g13890D*, a transcription factor involved in MAPK (mitogen-activated protein kinase) and calcium mediated signaling, a transcription factor *BnaAnng29870D* or *DEHYDRATION RESPONSE ELEMENT-BINDING PROTEIN 26 (Bna.DREB26*), which are down-regulated in KWS12. The up-regulated hub genes include *BnaA10g20420D* annotated for hyperosmotic response and others annotated for response to abscisic acid stimulus *(BnaC05g29210D* and *BnaA05g08680D*). In addition, lower kME genes involved in glycine and *myo*-inositol biosynthesis were found (Table 7).

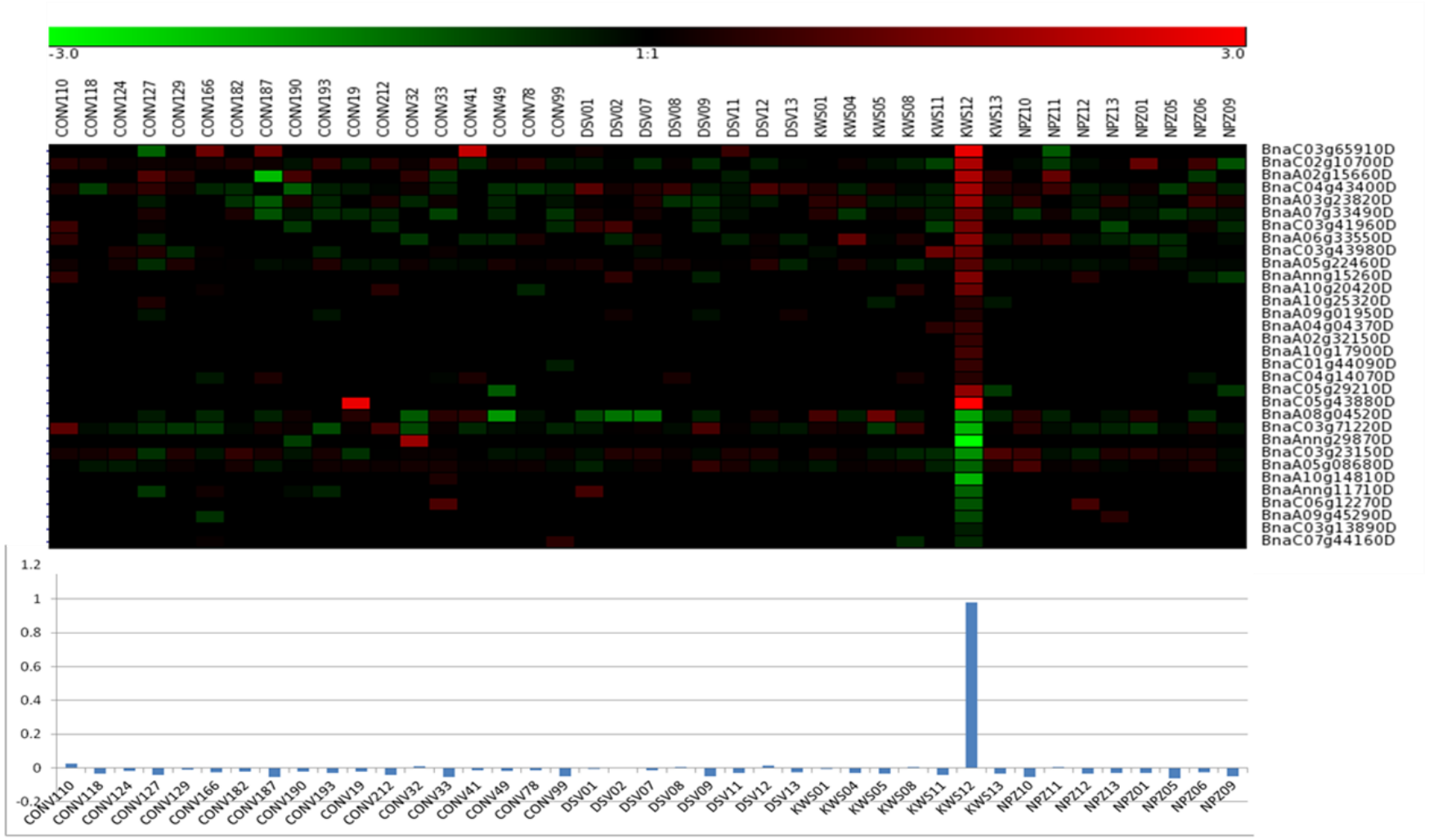


Figure 10. Expression heatmap of genes that are responsive to osmotic stress and genes involved in glycine and myo-inositol biosynthesis. The barplot shows the module eigengenes for the module *shoot-skyblue3* whose expression profile correlated to biomass, fructose, glucose, glycine, myo-inositol, alpha-alanine and sucrose

Table 7. Top hub genes and metabolism genes from the module *shoot-skyblue3* whose expression profile correlated to fructose, glucose, glycine, myo-inositol, alpha-alanine and sucrose

Shoot-Skyblue3 Fructose, glucose, glycine, *myo*-inositol, alpha-alanine, sucrose module				
Gene ID	A.th.orth	Gene description	GO enrichment	kME
BnaA02g32150D (top hub)	*CYP71B12*	cytochrome p450	oxidation-reduction process	0.99
BnaC03g13890D	*CBL2*	calcineurin b-like protein 3	TF, MAPK, calcium mediated signaling	-0.99
BnaA10g17900D	*LRR1*	atp binding	response to abscisic acid stimulus, kinase activity	0.99
BnaA10g14810D	*DUF581*	-	response to salt stress	-0.98
BnaC01g44090D	-	pentatricopeptide repeat-containing protein	response to abscisic acid stimulus	0.88
BnaA10g20420D	*AT5G12250*	beta-6 tubulin	hyperosmotic response	0.84
BnaAnng29870D	*DREB26*	ap2 domain-containing TF family protein	transcription factor	-0.84
BnaC05g29210D	*NLM9*	protein brevis radix-like	response to abscisic acid stimulus	0.8
BnaAnng15260D	*AtMYB107*	myb domain protein 9	response to abscisic acid stimulus	0.8
BnaA05g08680D	*LEA4-2*	seed maturation protein	response to water deprivation, response to osmotic stress, response to abscisic acid stimulus	0.63
BnaC03g71220D	-	nac domain ipr003441	drought recovery	0.58
BnaC04g14070D	*AT2G29360*	tropinone reductase	myo-inositol hexakisphosphate biosynthetic process	0.54
BnaA08g04520D	-	cytochrome p450 monooxygenase	glycine catabolic process	-0.46

The module eigengene (ME) of *shoot-thistle2* had the highest correlation to proline. This module had hub genes involved in proline biosynthesis pathway (*BnaA04g08160D* or *Bna.ATGLN1* and *BnaC04g11450D*), proline transport (*BnaA10g26830D* and *BnaC04g16720D*) as well as genes involved in osmotic stress response (Table 8). Conv33 (Lisek) genotype had the highest fold change of proline in osmotic stress and reduced biomass in stress. This genotype contributed to variation in module due to the high absolute value of the module eigengene. In this genotype, up-regulated genes were involved in proline biosynthesis (*BnaC04g11450D*) and response to osmotic stress (*BnaC07g38190D*). Down-regulated genes high in kME were genes involved in response to salt stress (*BnaA09g12960D*

and *BnaC07g46400D*), response to water deprivation (*Bna.PIP1-1*, *Bna.ABF4* and *Bna.AP2/ERF1*).

Table 8. Top hub genes and metabolism genes from the module *Shoot-Thistle2* whose expression profile correlated to proline.

Shoot-Thistle2 (Proline module)				
Gene ID	A.th.orth	Gene description	GO enrichment	kME
BnaA01g06730D (top hub)	*ATXTH19*	probable xyloglucan endotransglucosylase hydrolase protein 17	response to auxin stimulus	1
BnaA10g26830D	*AT5G03030*	chaperone -domain containing protein	proline transport, heat shock protein binding, response to stress	-1
BnaA09g39170D	*PIP1-1*	aquaporin pip1-2	response to water deprivation, water channel activity	1
BnaC04g16720D	*AT2G27660*	cysteine histidine-rich c1 domain-containing protein	proline transport	1
BnaC03g09270D	*HSP*	dnaj heat shock n-terminal domain-containing protein	-	1
BnaA09g12960D	*AT1G63010*	major facilitator superfamily with spx domain-containing protein	response to salt stress	1
BnaC07g46400D	*MES9*	acetone-cyanohydrin lyase	response to salt stress, TF	0.83
BnaA04g08160D	*AT5G37600A TGLN1*	glutamine synthetase	glutamine biosynthetic process	0.72
BnaC05g33570D	*ABF4*	abscisic acid responsive elements-binding factor	abscisic acid mediated signaling pathway, calcium mediated signaling, response to water deprivation , TF	0.71
BnaC06g40040D	*AP2/ERF 1*	ap2-like ethylene-responsive transcription factor	response to water deprivation, response to abscisic acid stimulus	0.71
BnaC04g11450D	*At5g14800*	pyrroline-5-carboxylate reductase	proline biosynthetic process, pyrroline-5-carboxylate reductase activity	-0.50
BnaC01g10570D	*AT4G18170*	wrky dna-binding protein 28	cellular response to hydrogen peroxide, response to abscisic acid stimulus	-0.53
BnaC07g38190D	*PIP1;5*	water channel-like protein	response to water deprivation, water transport, hyperosmotic response	-0.48

The ME of *shoot-darkorange2* is correlated to proline. The top hub gene in this module is *BnaC02g06660D* (*Bna. RGP5*) annotated for response to salt stress. This module had hub genes annotated for abscisic acid mediated signaling pathway and drought stress tolerance. Some lower kME genes such as *BnaC03g03050D* (*Bna.GDH2*) are from the proline biosynthesis pathway (Table 9). Module eigengenes differentiated the genotypes that reduced and did not reduce biomass in stress. The stressed genotypes NPZ09 and DSV09 decreased in biomass as well as the module eigengene. On the other hand, Gross-Luesewitzer (Conv124) and SWGospel (Conv49) did not show any change in biomass under stress, nor did the module eigengenes, but they showed a high fold change in proline. The top hub gene *BnaC02g06660D* and the hub gene *BnaA01g02740D* were up-regulated for Gross-Luesewitzer and SWGospel but down-regulated for NPZ09 and DSV09. These genes were both annotated for osmotic stress response.

Table 9. Top hub genes and other osmotic stress related genes from the module *Shoot-Darkorange2* whose expression profile correlated to proline

Shoot-Darkorange2 (Proline module)				
Gene ID	A.th.orth	Gene description	GO enrichment	kME
BnaC02g06660D (top hub)	*AT5G16510*	alpha- -glucan-protein synthase	response to salt stress	0.89
BnaA01g02740D	*AT4G34390*	extra-large gtp-binding protein 2	hyperosmotic salinity response, abscisic acid mediated signaling pathway, response to water deprivation	0.74
BnaC07g44770D	-	v-type proton atpase subunit h-like	response to salt stress	0.71
BnaA08g09910D	-	carbonic anhydrase family protein	response to salt stress	0.69
BnaC06g21690D	*AT1G76650*	calcium-binding protein cml38	abscisic acid mediated signaling pathway	0.68
BnaA08g10990D	*AT4G34710*	arginine decarboxylase	response to water deprivation, abscisic acid mediated signaling pathway	0.65
BnaA07g14610D	*AT5G38960*	oxalate oxidase	response to salt stress	-0.59
BnaC01g02670D	*MCA1*	mid1-complementing activity 1	osmosensory signaling pathway	0.59
BnaC05g31510D	*ATEIN3*	ethylene insensitive 3-like	response to salt stress, response to abscisic acid stimulus, TF	-0.57
BnaC05g28450D	*ATMYC2*	transcription factor myc2-like	response to desiccation, abscisic acid mediated signaling pathway, sequence-specific DNA binding transcription factor activity	0.54
BnaA08g10700D	*IDH1*	isocitrate dehydrogenase	response to salt stress	-0.54
BnaA05g23110D	*AT3G16785*	phospholipase d p1-like	abscisic acid mediated signaling pathway, cellular response to water deprivation	-0.48
BnaA04g13540D	*ATMYB70*	transcription factor myb44-like	calcium-mediated signaling, response to abscisic acid stimulus	0.44
BnaC03g03050D	*GDH2*	glutamate dehydrogenase 2	response to salt stress	-0.43

The ME of the module shoot-palevioletred3 is correlated to ornithine, betaalanine, valine and phenylalanine. Among the top hub genes is *BnaA10g17910D* or *Bna.GLN1* from the proline biosynthesis pathway. The other hub genes are involved in response to osmotic stress such as *Bna.CBP60G*. Several osmotic stress related hub genes were abscisic acid regulated:

Bna.ATMKK2, BnaC03g63530D and *Bna.ATRDUF2*. Other osmotic stress related genes in the module regulated by abscisic acid are *BnaA02g14740D* (*Bna.SIN3-LIKE*), *BnaA03g18930D* (*Bna.ATEM6*) and *BnaC08g10490D* (*Bna.DI21*) (Table 10). The genotype whose ME contributed to the variation in this module was Laser (Conv41), which reduced in biomass in the osmotic stress experiment. It also had high fold change of metabolites. For this genotype, the hub genes involved in hyperosmotic response (*Bna.ATMKK2*) and response to water deprivation (*BnaC03g63530D*) were up-regulated while glutamine biosynthesis (*BnaA10g17910D*) and another gene involved in response to water deprivation (*Bna.ATRDUF2*) were down-regulated.

Table 10. Top hub genes and metabolism genes from the module *Shoot-palevioletred3* whose expression profile correlated to ornithine, betaalanine and phenylalanine

Shoot-palevioletred3 (ornithine, betaalanine, valine and phenylalanine module)				
Gene ID	A.th.orth	Gene description	GO enrichment	kME
BnaC06g35620D (top hub gene)	*AT1G74790*	hipl1 protein		0.99
BnaA10g17910D	*GLN1;4*	glutamine synthetase	glutamine biosynthetic process	0.99
BnaC01g08650D	*ATMKK2*	mitogen-activated protein kinase 2	hyperosmotic response, MAPK, abscisic acid mediated signaling pathway, TF	-0.99
BnaA02g31430D	*CBP60G*	cam-binding protein 60-like g	MAPK cascade	0.94
BnaC03g63530D	-	gtp binding protein	response to water deprivation, abscisic acid mediated signaling pathway	-0.99
BnaA10g12520D	*ATRDUF2*	e3 ubiquitin-protein ligase ring1-like	response to water deprivation, abscisic acid mediated signaling pathway	0.99
BnaA02g14740D	*SIN3-LIKE 4.*	sin3-like 4-like	response to abscisic acid stimulus, TF	0.64
BnaA03g18930D	*ATEM6*	em protein	response to abscisic acid stimulus, acquisition of desiccation tolerance in seed	0.53
BnaA10g26360D	*HD2C*	histone deacetylase hdt1-like	response to water deprivation, response to abscisic acid stimulus	0.42
BnaA05g10200D	*RD20*	caleosin 3	response to desiccation, response to abscisic acid stimulus	0.42
BnaA04g08450D	*AT5G38200*	-	glutamine metabolic process	0.42

3.2.4 Consensus modules conserved between root and shoot expression networks correlated with traits

The expression profile network preservation analysis between shoot and root fold changes in osmotic stress showed that they were highly preserved. The overall density of the preservation network was found to be high: D= 0.88. The analysis resulted in 83,012 genes clustered into 75 consensus modules. The number of genes varied from 33 to 400 in most of the consensus modules with an average of 290 genes, with a total of 21,459 genes (25.8% of the total expressed genes). The remaining genes that did not correspond to networks common between the shoot and root expression were 61,553 and were kept in the grey module. Several shoot and root metabolites were found to be highly correlated to MEs of the consensus modules. GOstats analysis and REVIGO summary showed that 12 modules were annotated for GOs related to osmotic stress response. Out of these, the modules correlated to metabolites and to biomass were selected for a deeper investigation (Table 11).

Relative values of shoot proline, fructose, glucose, glycine, *myo*-inositol, beta-alanine and ornithine in osmotic stress were highly correlated to module *Consensus-Lightcoral* enriched in osmotic stress response and metabolite biosynthesis. Tyrosine, methionine, arginine and leucine were the metabolites of root that showed highest correlations to modules of interest.

Hub genes with high meta-kME from the selected consensus modules are involved in osmotic stress resistance. These include *Bna.CBL2*, *Bna.AZF2*, *Bna.ATERF-1*, *Bna.ATLEA4-2*, *Bna.P5CS1*, *Bna.ERD7* and *Bna.ATPI4K-GAMMA4* (Table 12).

Table 11. Modules with ME-trait correlation to both root and shoot expression network. The low correlation is taken except for some traits which correlated to ME of only one tissue. *Only the root ME correlation coefficient.

Cons. ME-trait	Relevant GO enriched	Correlation of ME to root trait
Consensus-Darkolivegreen	threonine biosynthetic process, isoleucine metabolic process, isoleucine biosynthetic process	Threonine (0.44*), Glycerate (0.44*), Glutamine (-0.42), biomass (0.26)
Consensus-Lightcoral	response to osmotic stress, response to salt stress, proline biosynthesis	Tyrosine (0.57), Methionine (0.56), Arginine (0.53), Leucine (0.53), Tryptophane (0.51), Histidine (0.49), Glycine (0.47), Threonine (0.48), Valine (0.5), Phenylalanine (0.45), biomass (0.49)
Consensus-Yellow	response to desiccation	Methylcysteine (-0.53), biomass (0.23)
Cons. ME-trait	Relevant GO enriched	Correlation of ME to shoot trait
Consensus-Lightcoral	response to osmotic stress, response to salt stress, proline biosynthesis	Fructose (0.94), glucose (0.96), glycine (0.95), *myo*-inositol (0.92), alpha-alanine (0.62), sucrose (0.50), glutamate (0.48), serine (0.43), methionine (0.41), biomass (0.2)
Consensus-Yellow	response to desiccation	Ornithine (-0.22)

Table 12. Top hub genes and other genes from the selected consensus modules involved in osmotic stress.

Cons. module	Gene ID	A.th.orth	Gene description	GO enrichment	Meta-kME
Consensus-Lightcoral	*BnaA06g33090D* (tophub)	-	glycosyl hydrolase family 5 protein	carbohydrate metabolic process	24.8
	BnaC03g13890D	*CBL2*	calcineurin b-like protein 3	calcium-mediated signaling	-24.8
	BnaC07g21220D	-	leucine-rich repeat receptor-like protein kinase	transmembrane receptor protein serine/threonine kinase activity	10
Consensus-Darkolivegreen	*BnaC06g14080D (tophub)*	*ZPR3*	protein little zipper 3	regulation of meristem growth, leaf shaping	14.8
	BnaC05g33170D	*AZF2*	zinc-finger protein 2	response to water deprivation, abscisic acid mediated signaling pathway	9.8
	BnaA09g24240D	*ATLEA4-2*	seed maturation protein	response to water deprivation, response to osmotic stress	7.3
	BnaC02g44370D	-	receptor-like protein kinase	protein serine/threonine kinase activity	6.5
Consensus-Yellow	*BnaC03g25840D*	*ATPI4K GAMMA 4*	phosphoinositide 4-kinase gamma 4	abscisic acid mediated signaling pathway, response to water deprivation	-23.7
	BnaA07g03060D	*ERD7*	dehydration-associated protein	response to abscisic acid stimulus, response to water deprivation	-11
	BnaAnng14540D	-	armadillo beta-catenin repeat family protein	response to water deprivation, hyperosmotic salinity response, MAPK cascade, response to cold, abscisic acid mediated signaling pathway	10.6
	BnaC04g05620D	*P5CS1*	delta1-pyrroline-5-carboxylate synthase 1	response to desiccation, response to abscisic acid stimulus	8

4. Discussion

In the first part of the study a transcriptome-based systems analysis of germination and seedling emergence in a panel of 42 winter-type *B. napus* genotypes with high, low and intermediate germination rate was performed. Gene co-expression analysis was done on gene expression of shoot and root of seedlings grown without any treatment (control samples) and co-localized with QTL for *in-vitro* germination and seedling vigour from another study (Hatzig et al. 2015). The first outcome of this analysis was a type of eQTL confirmation of candidate genes from automated phenotype seed germination QTL (genes showing high expression correlation to QTL trait in QTL trait-correlated gene co-expression modules). This was a new approach of eQTL to confirm genes controlling QTL traits by RNAseq of only a subset of the GWAS diversity set and additional genotypes. Module selection for this part was driven by the correlation of module eigengenes to the QTL trait from the *in-vitro* germination phenotyping and contained QTL candidate genes at the same time. The principle behind this is because of allelic variation, genes are differentially expressed in genotypically diverse individuals of a species. These variations could arise from allelic variants of transcription factors and other regulators, cis-elemental variation in promoter sequences, differences in mRNA stability, copy number variation and genomic rearrangements such as translocations, insertions and deletions. The latter include gene loss and duplication, resulting in neo- and sub-functionalization (Rockman and Kruglyak 2006). Gene co-expression modules from a subset of genotypes from a GWAS diversity set represent the variation of high and low performing genotypes.

Most of the genotypic variations in DNA structure will result in eQTLs but depending on the position of the causal polymorphism, they could be local and distant eQTLs. Local eQTLs can be the result of closely linked trans-acting factors but in the majority of cases result from cis-regulatory variation in the genes under study. eQTLs acting in cis affect transcription initiation, rate and/or transcript stability in an allele-specific manner. In addition, cis-regulated genes might encode regulators affecting the expression of downstream target genes in trans (Gibson and Weir 2005). Because of the difficulties in cloning QTLs and the large biological relevance of hotspots, additional sources of information such as information on gene ontology, co-expression with transcription factor are often used to reduce the number of candidate genes and predict the causal regulator to reduce the number of candidate genes and prioritize remaining candidates for further experimentation (Zhu et al. 2008). Thus modules

which contain a group of genes whose expression profile is correlated to the QTL traits, including the QTL candidate genes can be useful to elucidate regulation of the QTL traits.

The second outcome of the first project was to describe hub genes regulating field seedling emergence. Since the *in-vitro* germination and field emergence traits correlated, the expression profile of the *in-vitro* grown seedlings was correlated to field emergence traits. Thus it describes transcriptome-based network analysis of field germination and seedling emergence in roots and shoots from a panel of 42 genetically diverse winter oilseed rape genotypes with good, poor and intermediate germination. Co-expressed transcript clusters were correlated to field emergence data from multi-environment field trials. Gene significance (GS) describing the correlation of each gene expression profile across 42 genotypes to each trait was analyzed and was compared for all identified shoot and root network modules. Then the mean GS was calculated which describes the gene significance across all field traits. The other criteria for modules selection in this section is highest module-field emergence correlations, in order to find hub genes regulating field emergence. The modules with high module eigengene correlation to the field emergence traits contained genes from the QTL interval with moderate to low correlation of QTL gene expression (gene significance or GS) with QTL trait and moderate to low module eigengene-QTL trait correlation. *In-vitro* germination QTL trait-correlated gene co-expression modules had moderate to low correlation to field emergence traits. Nonetheless, there were modules associated to both *in-vitro* germination and field emergence. This implies that there were genes whose expression was correlated with multiple traits. Evidence is accumulating that pleiotropy is even more pervasive than previously imagined and also occurs between traits that are not thought to be functionally related. Furthermore, the pleiotropic effects of different genes that affect pairs of traits are often not in the same direction and therefore do not result in significant genetic correlations between the traits (Mackay et al. 2009).

The second project adds another variable to the first dataset: an additional gene expression profiling of simulated osmotic stress treated samples in *in-vitro* PEG treatment, and additional metabolite profiling in stress and control and biomass in stress and control. Although not an exhaustive list of osmotic stress response enriched modules generated during the analysis, the modules were chosen primarily based on module eigengene correlation to metabolites and biomass to decipher the osmotic stress gene co-expression network with respect to changes in metabolites and biomass. Due to the diverse response of the genotypes to osmotic stress in

terms of *in-vitro* biomass and field emergence, there were few extremes that could be selected when the *in-vitro* biomass and field emergence were taken into consideration.

4.1 Complex multifunctional gene expression networks influence seedling development and vigour in *Brassica napus*

4.1.1 Co-expression modules from in-vitro grown seedlings correlated to in-vitro seed vigour as well as field emergence traits

Gene expression studies in plants often compare the differential expression of one or a few genotypes under different treatments or conditions. Such studies can be valuable to gain knowledge about the genetic control of specific phenotypes in a given genetic background, whereas they are less suitable for making general conclusions about the regulation of complex traits on a species-wide scale. On the other hand, global gene expression network analysis in genetically diverse populations gives less precise information about specific aberrant phenotypes, but is much more suitable to give general insight into the underlying processes for complex biological processes like germination and emergence. Systems biology approaches like WGCNA have been demonstrated to be useful in crops for relating gene expression networks in diverse genotypes to agronomic traits of interest (Ficklin et al. 2010, Weston et al. 2008) and QTL (Körber et al. 2015).

In the present study high correlations of module eigengene with *in-vitro* germination (QTL traits) as well as field emergence traits for both root and shoot gene expression networks were observed, demonstrating the validity of this approach to identify groups of functionally related genes whose cumulative expression in a diverse genotype panel associates with broad phenotypic variation for germination and emergence traits. Functional annotation based on GO and pathway enrichment subsequently provides a means to confirm the biological relevance and significance of the modules of interest. Candidate genes inside QTL intervals of linkage disequilibrium (LD) containing significant marker-trait associations was confirmed with the network analysis, which further validated the potential regulatory role of the genes in seed germination and seedling vigour.

In this study gene expression networks related to both germination and field emergence traits, some of which showed correlations to each other were identified. Close correlations were found between germination traits and field phenotype data relating to the emergence at two and four weeks after sowing. This relationship presumably explains why gene expression networks in 4 weeks old seedlings that showed significant correlations to germination and

field emergence traits were identified. In addition, several genes in modules that showed such correlations are ubiquitous genes that are expressed and/or involved in growth and development at various developmental stages. According to gene ontology information (Conesa et al. 2005) and information about *A. thaliana* orthologues (Lamesch et al. 2012), many of the genes that were identified are expected to be expressed during seed germination as well as during subsequent seedling growth stages. Thus WGCNA from 4 week old plants grown hydroponically can be used for identification of genes involved in *in-vitro* and field germination performance.

4.1.2 WGCNA allows unbiased filtering for candidate genes from QTL regions and identified new candidate genes overlooked by annotation filtering

QTL analysis in combination with WGCNA identified candidate genes that could only be filtered out by supporting the QTL study with the network analysis. The network analysis showed stronger support to additional QTL gene candidates than mentioned in the study of (Hatzig et al. 2015). Based on the correlation of gene expression and QTL trait, it can be concluded that the QTL genes with higher values are important candidate genes. The high connectivity (kME) within the modules that had high correlation of ME (expression profile) with the QTL trait serves as an additional confirmation of candidate genes by the network analysis in this study.

Among the QTL traits, module eigengene and TSW trait correlations were the highest for both root and shoot. Several TSW candidate genes were supported by the network analysis. Among the genes from the QTL region (Hatzig et al. 2015) whose expression correlated to the associated trait TSW, *BnaA05g25150D* (*Bna.PME3*) is a pectin methylesterase gene in a chromosome region with the strongest QTL effect on TSW on chromosome A05. Pectin methylesterases (*PME*s) disassemble pectin in the cell wall at different developmental stages of several plants including Brassica species (Duan et al. 2016). According to gene ontology information (Conesa et al. 2005), it is involved in growth and development at various developmental stages. Based on its *A. thaliana* orthologue (Lamesch et al. 2012), it is expressed in seed and as well as during subsequent seedling growth stages. These evidences support its importance in development from the seed to seedling stage in *B. napus*. This gene was found within the module highly correlated to TSW. Regarding GO terms, the module was highly enriched with the term ‘plant-type cell wall’ which was ascribed to *Bna.PME3*. This indicates that in addition to *Bna.PME3* the module contains other genes acting within the cell wall. Within this module, *BnaC04g51830D* (*AT2G47550*), a hub gene with one of the highest

kME (0.90) is from the plant invertase/pectin methylesterase inhibitor superfamily. Within this module, *BnaA10g03540D* (*AT1G05805*, transcription factor bhlh128-like) is a hub gene (kME=0.84) that is co-expressed with *Bna.PME3*. BnaC01g34790D (invertase inhibitor-like protein) (kME=0.80) a hub gene with potential interaction with *Bna.PME3*. BnaC01g34790D (invertase inhibitor-like protein) (kME=0.80), which does not have the *A. thaliana* ortholog was annotated with pectinesterase activity via Blast2go and is similar to pectin methylesterase inhibitor (*PMEI*) (Conesa et al. 2005). Similarly, the hub gene BnaC04g51830D (AT2G47550) with kME (0.90) was from the plant invertase/pectin methylesterase inhibitor superfamily (Lamesch et al. 2012). *PMEI* was reported to be co-expressed with and regulate *PME* genes for tuning of plant development (Hocq et al. 2017). The co-expression of the *A. thaliana* orthologs of the transcription factor hub gene BnaA10g03540D (AT1G05805, transcription factor bhlh128-like) and *PME3* was confirmed by other studies (r=0.5) and deposited in a co-expression database (Aoki et al. 2016).

The other candidate genes strongly supported in the chromosome region with the strongest QTL effect on TSW *BnaA05g25310D* (*AT3G14170*) and *BnaA05g25320D* (*AT3G14160*) were not well characterized even in *A. thaliana* (Lamesch et al. 2012). In addition to the TSW QTL trait associated modules, the TSW QTL gene *BnaA01g04540D* (*AT4G32470*, cytochrome bd ubiquinol oxidase). Within this module, this QTL gene annotated to 'aerobic respiration' had high connectivity with two hub genes annotated to 'aerobic respiration' with r=0.65 and 0.61. For the *A. thaliana* orthologues, the co-expression of QTL gene *AT4G32470*, cytochrome bd ubiquinol oxidase with the hub gene *BnaC05g45770D* (*AT3G06050* peroxiredoxin- mitochondrial-like, *PRXIIF*) had already been confirmed in other studies and the information deposited in the *A. thaliana* gene co-expression database as r= 0.48 (Aoki et al. 2016).

In a recent study involving GWAS and transcriptome study of TSW in *B. napus*, the QTL was found on chromosome A03 and C09 (Lu et al. 2017), and thus had no overlapping genes with the QTL genes confirmed by the network analysis in this study. In other crops such as wheat, dwarfing genes, photoperiodism genes, cell wall invertase genes and RING-type E3 ubiquitin ligase affecting grain filling and grain size were known from TSW QTL (Zanke et al. 2015). All the genes identified as TSW candidate genes in this study were not reported to be associated to seed weight QTL in earlier studies. In other crops, E3 ubiquitin ligase was known to be associated to TSW QTL (Zanke et al. 2015). *BnaA06g34030D* or *Bna.UBC2* (AT2G02760, ubiquitin-conjugating enzyme e2), which forms a complex with E3 ubiquitin

ligase during ubiquitination (Mazzucotelli et al. 2006) was found to be a candidate gene for TSW QTL supported by WGCNA in this study.

Seed weight is known as an essential yield component in a number of crops, especially in winter type crops such as wheat and known to be generally under strong genetic control (Zanke et al. 2015). Also in oilseed rape the thousand seed weight (TSW) shows a high heritability of h^2=0.76 and h^2=0.84 for seed lots of a diversity panel of 248 produced in two different environments (Hatzig et al. 2015). But in general, in oilseed rape seed quality is more determined by production factors for the seeds than by genetics. A large number of seed quality traits including seed size (TSW) and environmental effects affect establishment (germination and vigour) in the field with each of these individual factors having a relatively small effect. However, larger oilseed rape seeds show under controlled conditions higher germination rates and are more resistant to drought stress (BLAKE et al. 2004). Seed weight affects seedling emergence and tolerance to flea beetles stress in *B. napus* (Elliott et al. 2008). In this study, which contains 18 of the 248 diversity panel from the QTL study of (Hatzig et al. 2015), TSW was correlated to number of plants, leaf biomass, and development before and after winter in some of the locations of the field trials. Thus, breeding for TSW and combining favourable alleles for TSW and other vigour-related QTL and candidate genes might be one approach to increase germination and vigour traits in oilseed rape.

4.1.3 Putative regulation for field seedling emergence and stress adaptation

In the present study high module eigengene and trait correlations up to r=0.99 for root and shoot gene expression networks were observed, demonstrating the validity of this approach to identify groups of functionally related genes whose cumulative expression in a diverse genotype panel associates with broad phenotypic variation for germination and field emergence traits.

Regarding the modules associated to field emergence traits, several hub genes, including transcription factors controlling seedling growth were found. Although not all found in one known pathway, QTL and hub genes with related processes involved in seed germination, seedling growth and development were found to be interconnected in the network analysis. Several hub genes and all of the common hub genes between shoot and root expression networks were involved in germination and growth regulating hormones, namely, absicic acid, cytokinin, giberellin and auxin. These hormones influence growth in a changing environment by being involved in development and stress tolerance, by interacting with each

other by cross-talk (Peleg and Blumwald 2011) and modulate growth by responding to environmental cues (Wolters and Jurgens 2009). Pathways involving hormones such as auxin and cytokinin are important targets to improve traits in many crop species, including germination and seedling growth (Khurana et al. 2007).

Auxin signaling and response related genes were found in the module eigengenes highly correlated to field emergence traits and the top hub genes were highly interconnected. Within the *root-yellow* (module eigengene highly correlated to DAW) confirmation of interconnection of such top hub genes *BnaA01g06750D* (*XTH24*), *BnaA02g35860D* (*MKK9*) and *BnaC05g52020D* (*RSH2*) with their *A. thaliana* orthologues was possible from the *A. thaliana* co-expression database (Aoki et al. 2016).

Three of the top 5% hub genes annotated for auxin pathway in the *shoot-skyblue* module with high eigengene correlation to the field emergence traits are from the helix-loop-helix TF family. There were also several high connectivity genes from the helix-loop-helix TF family in this module. Regulatory helix-loop-helix TF are known to be involved in plant growth and development processes like cell elongation (Zhiponova et al. 2013), while others are specialized in root hair cell growth (Yi et al. 2010). Among the top 5% hub genes in *shoot-skyblue*, *BnaA09g27770D* and *BnaC05g21490D* are homeologous genes and ortholog of *A. thaliana AT1G27740* (*RSL4*). *RSL4* was previously known to be expressed in root and function in root formation and growth in *A. thaliana*, but was expressed in *B. napus* shoot (Hwang et al. 2017). One of the hub genes identified in the module *BnaC06g35430D* (*Bna.ATBS1)*, is a member of the latter type of TFs. The *ATBS1* gene family is known to function in giberellin-dependent development, including germination and elongation of hypocotyls in *A. thaliana* (Lee et al. 2006). In *A. thaliana, ATBS1* has also been implicated in regulation of auxin-dependent embryonic root initiation (Schlereth et al. 2010), early development modulated through brassinosteroid signaling (Wang et al. 2010) and regulation of light signaling and root development (Castelain et al. 2012). This supports the putative identification of *Bna.ATBS1* as a candidate hub gene for regulation of gene expression networks determining germination and seedling establishment.

BnaA06g06460D (*AT1G10588*) is a hub gene in common with both *root-plum2* and *shoot-skyblue* modules with high ME-trait correlation with most field traits. It encodes a giberellin regulated GASA/GAST/Snakin family protein (Lamesch et al. 2012). Snakin/GASA proteins are known to be regulated by giberellin, auxin, brassinosteroids and absicic acid. Although

precise roles are not yet known, they were shown to play roles in growth and development including root formation and stem growth in several plant species (Nahirñak et al. 2012).

BnaA09g46830D an orthologue of *AUXIN SIGNALING F-BOX 3* (*AFB3*) and *BnaA01g00370D* an orthologue of *HOOKLESS 1 (HLS1)* were identified as a common hub gene in modules *shoot-yellow* and *root-green*. Furthermore, the expression of *BnaA06g06460D* and *BnaA09g46830D* (*AFB3*) hub genes in the root and shoot were found to be highly correlated (r>0.9) and perhaps may have a coordinated function in different tissues. *Bna.AFB3* was found to be among top hub genes for both *shoot-yellow* and *root-green* modules correlated to DAW. Its *A. thaliana* orthologue is known to play a role in seedling as well as root development by regulating auxin response (Dharmasiri et al. 2005, Parry et al. 2009, Vidal et al. 2013). In the same *root-green* module, another hub *BnaA03g17600D* gene has *A. thaliana* orthologue *AUX1* (*auxin influx1*) which also regulates auxin dependent developmental processes via auxin transport. This gene works in coordination with *AFB3* as part of auxin signaling pathway (Hayashi 2012). The other common hub between *root-green* and *shoot-yellow* is *BnaA01g00370D* orthologue of *A. thaliana HLS1*, known to involve in seedling emergence via auxin signaling (An et al. 2012). *BnaAnng28260D* (*GH3*) from shoot-yellow and *BnaA03g17600D* (*AUX1*) from root green were found in the auxin hormone signal transduction pathway in KEGG pathway database. *AUX1* activates *GH3* via intermediary steps, and they indirectly influence cell enlargement and plant growth (Zhang et al. 2013). In *A. thaliana* they are both expressed in whole plants (Lamesch et al. 2012). The roles of the top 5% hub genes within the development after winter (DAW) and other field emergence trait associated modules in the auxin pathway is summarized as follows: *AUX* is induced by auxin and ethylene, may integrate auxin signal into ethylene signal. *SGT1A*, required for *SCF-TIR1/AFB* mediated auxin responses. *SGT1A* is a phosphatase required for ubiquitination,. Auxin promotes degradation of *AUX/IAA* transcriptional repressors via proteasome pathway enhancing ubiquitination of the *AUX/IAA* protien. Auxin induced proteolysis of the *AUX/IAA* repressors leads to activation of *ARF* to induce early auxin responsive gene expression. *GH3* and *SAUR* are auxin responsive genes. *GH3* controls negative feedback regulation of *IAA* homeostasis. *HLS1* regulates *IAA* (Hayashi 2012). In addition to Arabidopsis, auxin-responsive genes *Aux/IAA*, *SAUR*, and *GH3* have been identified and characterized from different plant species including pea, soybean, tobacco, mung bean, and cucumber (Hagen and Guilfoyle).

Auxin signaling pathway was one of the pathways enriched for the shoot module. The emerging trend from the recent experiments suggest that cold stress induced change in the plant growth and development is tightly linked to the intracellular auxin gradient, which is regulated by the polar deployment and intracellular trafficking of auxin carriers. Numerous studies link the plant hormone auxin both at cellular and molecular levels to regulate the developmental plasticity, which powers the plant to adapt to continuous environmental changes (Rahman 2013). Auxin is unique among the plant hormones with its capacity to move both long and short distances. The long-distance transport is rapid and source-to-sink type, where auxin moves from biosynthetically highly active young tissues to sink tissue such as root through phloem. The short-distance transport is slower, occurs in a cell-to-cell manner and is regulated by specific influx and efflux carrier proteins. In Arabidopsis, two polar transport streams function in facilitating the intracellular auxin transport. The unidirectional transport from shoot apex to root, root-ward auxin transport, is regulated by *PINFORMED* (*PIN1)*. In the root, auxin transport is more complex, with two distinct polarities. IAA moves toward the root tip (rootward direction), through the central cylinder cells with the aid of *AUX1* (auxin influx), *ABCB19* (ATP-binding cassette group B) and *PIN1,3,7* (*PINFORMED*) (Muday et al. 2012). Once it reaches the root tip, IAA moves in a reverse direction toward the shoot through the outer layers of root cells. This shootward movement of IAA is regulated by *AUX1*, *PIN2* and *ABCB4*. The growth and development of plant is regulated by a complex web of hormonal interactions. Interestingly, auxin has been found to be a common factor in majority of these interactions. The process that has been shown to be intrinsically involved in all the existing hormonal crosstalk is auxin transport.

In module *shoot-pink* with ME-trait correlation to germination traits and field emergence, the hub genes *Bna.EPR1 BnaA06g12480D* (*AT1G18330, EPR1*) was a TF annotated for germination, ABA response and auxin signalling pathway. *EPR1* is also involved in light induced cotyledon opening (Kuno 2003).

4.1.4 Genes in LD with QTL for germination and vigour are interconnected to network hubs in multi-environment field emergence trait associated modules

Among the top hub genes in the *shoot-skyblue* module with high module eigengene correlation to multi-environment field emergence traits and ES, several transcription factors (TF) hub genes involved in regulation of seedling growth and development were found to be highly correlated to the two genes in the module from the QTL interval (r=0.65-0.55) (S16 Table). Three hub genes in the *shoot-skyblue* module with high expression correlation to the

QTL genes colocalized to the same module are from the helix-loop-helix TF family. Regulatory helix-loop-helix TF are known to be involved in plant growth and development processes like cell elongation (Zhiponova et al. 2013), while others are specialized in root hair cell growth (Yi et al. 2010). One of the top hub genes identified in the present study, *Bna.ATBS1*, is a member of the latter type of TFs. The *ATBS1* gene family is known to function in giberellin-dependent development, including germination and elongation of hypocotyls in *A. thaliana* (Lee et al. 2006). In *A. thaliana*, *ATBS1* has also been implicated in regulation of auxin-dependent embryonic root initiation (Schlereth et al. 2010), early development modulated through brassinosteroid signaling (Wang et al. 2009) and regulation of light signalling and root development (Castelain et al. 2012). This supports the putative identification of *Bna.ATBS1* as a candidate hub gene for regulation of gene expression networks determining germination and seedling establishment. Another top hub in the same module correlated to QTL gene expression is *BnaA06g06460D* (AT1G10588), which encodes a giberellin regulated GASA/GAST/Snakin family protein (Lamesch et al. 2012). Snakin/GASA proteins are known to be regulated by giberellin, auxin, brassinosteroids and absicic acid. Although precise roles are not yet known, they were shown to play roles in growth and development including root formation and stem growth in several plant species (Nahirñak et al. 2012) and also thought to be involved in stress tolerance in addition to growth and development (Sun et al. 2013). Another hub gene whose expression is highly correlated to the QTL genes *BnaC05g45770D* (AT3G06050) is a peroxiredoxin which is involved in oxidative stress tolerance (Xia et al. 2015).

In the *shoot-yellow* module, several TF hub genes annotated for regulation of seed germination, response to ABA and auxin stimulus and regulation of hormone mediated signaling pathways were highly correlated (r=0.84-0.70), to the genes from QTL region found in this module. For instance, within this module, *Bna.AL6*, found in thousand seed weight QTL region is also among hub genes in the same module and is involved in seed germination and seedling growth. It acts in repression of seed developmental genes which in turn facilitates seed germination and early seedling growth (Molitor et al. 2014). It has high connectivity with three of the top hub genes *BnaC06g09310D* (myrosinase-binding protein), *BnaA02g34700D* (*Bna.ENY*) and *BnaC06g43220D* (transcription factor ice) enriched for regulation of seed germination.

In the *shoot-pink* module, top hub genes included ones with roles in seed germination and seedling growth with medium to high correlation to genes from QTL interval within the same

modules (r=0.7-0.5). These *shoot-pink* module hub genes include *Bna.EPR1*, which is involved in light induced cotyledon opening (Kuno 2003). *Bna.UGT76C1* that modulates cytokinin responses in germinating seeds and young seedlings, (Wang et al. 2013) which is found in the QTL interval and correlated to hub genes annotated to *Bna.EPR1* a TF hub gene annotated for germination, ABA response, auxin signaling pathway and *Bna.AIM1* annotated for germination. The homeolog of *Bna.UGT76C1 BnaC09g49610D* is a hub gene within the same module and has a homeolog gene in the QTL region (Hatzig *et al.*, 2015). The homeologue of the hub gene *Bna.UGT76C1* from A genome was found in QTL region whereas the C homeologue was found to be among top hub genes in the module correlated to germination and field traits and is a hub in both *root-yellow* and *shoot-pink* which could mean regulatory role across tissues.

Similarly, in *yellow-root* module the hub genes *Bna.RACK1A* and *Bna.RCAR2* annotated for germination showed high correlation to QTL genes in the same module (r=0.75 - -0.66) (Table S16). The expression of the hub gene *Bna.RACK1A* is negatively correlated to the QTL expression. In *A. thaliana RACK1A* is known to suppress GA induced germination (Fennell et al. 2012). *Bna.AE3*, another gene from the QTL is known to be involved in seedling growth and development (Lamesch et al. 2012). *Bna.AE3* was co-localized to the *root-green* module and showed connectivity to *Bna.AFB3* and *Bna.AUX1* (r=0.66 and 0.62), top hub genes involved in auxin mediated seedling growth and development. *Bna.AFB3* was found to be among top hub genes for both *shoot-yellow* and *root-green* modules correlated to DAW. Its *A. thaliana* orthologue is known to play a role in seedling as well as root development by regulating auxin response (Dharmasiri et al. 2005, Parry et al. 2009, Vidal et al. 2013). In the same *root-green* module, another hub gene has *A. thaliana* orthologue *AUX1* which also regulates auxin dependent developmental processes via auxin transport. This gene works in coordination with *AFB3* as part of auxin signaling pathway (Hayashi 2012). This suggests a role in germination and seedling development for the QTL genes.

4.2 Osmotic stress responsive gene expression networks in *Brassica napus* seedlings

4.2.1 Osmotic stress responsive shoot gene expression networks were associated to biomass and metabolites responses to osmotic stress in diverse *B. napus* genotypes

In this study, several genotypes were decreased in biomass and increased in leaf temperature under osmotic stress. Plants increase leaf temperature in osmotic stress, which indicates that

transpiration is low due to stomatal closure (Hanson and Hitz 1982, Munns and Tester 2008). In another study, significant decrease in biomass and increase in leaf temperature was found in drought sensitive *B. napus* genotypes (Hatzig et al. 2014). Fold changes of metabolites in stress such as fructose, glucose and proline were also high. This indicates that plant response by producing osmolytes to avoid negative effect of water deficit on cells (Good and Zaplachinski 1994, Hatzig et al. 2014, Koh et al. 2015). On the other hand, genotypes which did not show significant fold changes in biomass were found. These genotypes could be classified as not sensitive to osmotic stress. The genotypes in this study showed varying responses in biomass and metabolite change in osmotic stress. Therefore, the correlation of the metabolites and the biomass to the module eigengenes, which is the summary of expression profile of the genes inside a module, resulted in a group of genes that are involved in osmotic adjustment. The modules selected based on correlation of module eigengene to biomass and metabolites were enriched for GO terms related to osmotic stress resistance. This is also further evidence for the relevance of the genes in the modules for osmotic stress response.

The module highly correlated to glucose, fructose, *myo*-inositol, alpha-alanine and other glucogenic amino acids (*shoot-skyblue3*) was enriched for terms such as drought recovery and calcium dependent protein phosphatase activity. Calcium dependent protein phosphatase is known to be involved in stress signalling responses (Vinocur and Altman 2005, Schulz et al. 2013). This links the drought responsive group of expressed genes in the module to the metabolites acting as osmolytes for resistance to osmotic stress. In other studies, glucose, fructose and *myo*-inositol concentrations increased as a result of osmotic stress in drought sensitive *B. napus* genotypes (Hatzig et al. 2014, Koh et al. 2015).

The modules correlated to proline (*shoot-thistle2* and *shoot-darkorange2*) were enriched for the GO terms such as calcium mediated signalling, response to osmotic stress, regulation of transcription in response to stress, glutamine biosynthesis and proline transport. Calcium mediated signalling is known to be involved in stress signalling in response to osmotic stress. This in turn activates transcription factors in response to stress (Vinocur and Altman 2005). Proline is important for osmotic adjustment of plant cells. Glutamine biosynthesis is involved in proline metabolism (Sharma and Verslues 2010). Thus, the genes in the proline correlated modules are involved in stress signalling and proline biosynthesis as a result of the stress stimuli perception. The module highly correlated to beta-alanine and ornithine (*Shoot-*

Palevioletred3) was enriched for response to abscisic acid (ABA). Ornithine is involved in proline biosynthesis. ABA regulates proline accumulation (Verslues and Bray 2006).

Among the root metabolites that correlated with consensus modules were leucine, valine, threonine and methionine. Leucine and valine (branched chain amino acids) share enzymes in their biosynthesis and are co-ordinately regulated. Leucine and valine are useful as drought stress osmolytes and produced in plants in abiotic stress (Joshi et al. 2010). High correlation of shoot metabolites to consensus modules was to fructose, glucose, glycine and *myo*-inositol. These metabolites are also known as osmolytes produced in plants under stress (Good and Zaplachinski 1994, Hatzig et al. 2014, Koh et al. 2015). Both root and shoot methionine was moderately correlated to the consensus module. Threonine and methionine regulate leucine and valine homeostasis. Leucine and valine are known to be osmolytes in plant osmotic stress resistance (Joshi et al. 2010). High fold changes of these metabolites were found in the root of the genotype KWS12.

4.2.2 Co-localization of osmotic stress responsive gene co-expression networks to QTL associated with water stress response

In this study, WGCNA was used to compare fold changes of expressed genes in 41 genotypes and relate it to metabolites and biomass. Several 'omics' studies on osmotic response in *B. napus* focused on one or a few genotypes (Koh et al. 2015, Liu et al. 2015, Shamloo-Dashtpagerdi et al. 2015). Recently, in a genome-wide association (GWAS) study that studied water stress tolerance in seedlings of *B. napus*, 16 loci were significantly associated with water stress response induced by using polyethylene glycol. 79 differentially expressed genes (DEGs) in PEG versus control in 12 old day seedlings were found in the GWAS loci out of which 11 genes were selected as candidate genes (Zhang et al. 2015). When the study of (Zhang et al. 2015) is compared with this study, 61 genes were found in the clustered shoot modules. Out of these, 16 genes were found in metabolite associated modules. This includes 12 genes, including *BnaA05g10920D* from the module moderately correlated to glycine, glucose, fructose and *myo*-inositol (*shoot-turquoise*) that overlapped with (Zhang et al. 2015). *BnaA05g10200D* was also found in a module highly correlated to ornithine and beta-alanine. Yet another candidate gene, *BnaC03g12400D*, from (Zhang et al. 2015) was found in a module highly correlated to Asparagine. The module highly correlated to ornithine and beta-alanine (*Shoot-Palevioletred3*) also contained candidate gene *BnaA05g10200D* (*Bna.RD20*) from (Zhang et al. 2015). From this module, *BnaA05g10200D* (*Bna.RD20*) was up-regulated in the genotypes such as NPZ13 and KWS08 which did not decrease in biomass in osmotic

stress and did not show significant increase in leaf temperature. *Bna.RD20* was down-regulated in the genotypes which decreased in biomass in stress such as Laser. In *A. thaliana*, *RD20* was found to be responsive not only to osmotic stress but also to disease stress (Sham et al. 2015). In general, this gene was up-regulated by the high performance genotypes both in terms of not changing in *in-vitro* biomass in stress as well as high performance genotypes in multiple year, multiple environment field emergence trials and germination tests. In contrast it was down-regulated in the low performance genotypes in osmotic stress, germination and field emergence.

In this study, it was found that around 26% of the expressed genes had similar expression patterns in the root and shoot. Consensus network analysis function of the WGCNA package has been useful to compare expression network profiles of different datasets such as tissues, species and multiple conditions. When the results of this study are compared with the GWAS study of water stress associated DEGs in *B. napus* (Zhang et al. 2015), 48 out of 79 of the DEGs in the GWAS loci from Zhang et al., 2015 were found in the consensus modules in this study, of which one module was correlated to glutamine.

4.2.3 Genes in osmotic stress responsive shoot expression networks related to metabolites

The gene expression network analysis mines out the key responses of the genotypes. For this purpose, the modules correlated to biomass, metabolites and osmotic stress responses were selected for deeper investigation. The hub genes and other genes in these modules were from osmotic stress response and metabolite biosynthesis pathways.

The diverse genotypes included in the study showed different gene expression responses to osmotic stress. Quality control steps to check for gene expression bias from outlier genotypes have already been done before construction of modules and calculation of module eigengenes. Before the decision to use WGCNA for the osmotic stress dataset, hierarchical clustering was done with treatment as binary trait and did not show any globally different genotypes. The genotype differences overwhelm the treatment differences. The winter fodder genotypes clustered together based on the gene expression (Appendix 2). In a GWAS association study that contained the genotypes in this study showed that the winter fodder genotypes are closely and formed distinct cluster than winter oilseed rape (Hatzig et al. 2015). In addition, scale free topology criterion was utilized to evaluate the appropriateness of the WGCNA method for the osmotic stress datasets. In the osmotic stress dataset, all the gene co-expression networks

satisfy scale free topology criterion. The scale-free topology fit index reached above 0.8 for reasonable powers (less than 7) and the mean connectivity remained low (below 500) (Appendix 3). It means that the data does not exhibit a strong driver that makes a subset of the samples globally different from the rest. In other words, there is no difference that causes high correlation among large groups of genes which invalidates the assumption of the scale-free topology approximation. (Zhang and Horvath 2005) provide empirical evidence that the scale free topology criterion results in adjacency function parameter estimates that in turn result in networks with a high biological signal: in yeast network, the connectivity distribution was better modeled using an exponentially truncated power law. Instead of focusing on the significance of the correlation or the network size, (Zhang and Horvath 2005) propose to pick the threshold by making use of the fact that despite significant variation in their individual constituents and pathways. Metabolic networks have been found to display approximate scale free topology (Jeong et al. 2000, Bergmann et al. 2004). This may indicate that metabolic organization is not only identical for all living organisms, but also complies with the design principles of robust and error-tolerant scale-free networks, and may represent a common blueprint for the large-scale organization of interactions among all cellular constituents. This provides strong evidence that these datasets represent true biological networks. This was the reason for the choice of WGCNA to analyze these datasets.

The absolute value of the eigengene for the genotype KWS12 was high in the module that was highly correlated to fructose, glucose, glycine and *myo*-inositol (*shoot-skyblue3*). The genotype KWS12 increased in fructose, glucose, glycine and *myo*-inositol concentration in osmotic stress, except one cultivation, where it decreased in biomass. High fold changes of these metabolites associated to decrease in biomass were reported earlier for drought sensitive *B. napus* genotypes because of resource mobilization from plant growth to production of the sugars (Hatzig et al. 2014). Among the hub genes in this module is *Bna.DREB26*. The expression of this gene was down-regulated in the genotype KWS12. In *A. thaliana*, down-regulation of the expression of *DREB26* is known to promote osmotic stress tolerance (Kazama et al. 2014). *BnaA05g08680D* (*Bna.LEA4-2*) was down–regulated in KWS12 although annotated to be responsive to osmotic stress in seedlings. The *A. thaliana* group 4 LEA (late embrogenesis abundant) proteins were found to be accumulated in seedlings in response to osmotic stress (Olvera-Carrillo et al. 2010). The up-regulated hub genes were annotated for hyperosmotic response and response to ABA. Among these, the transcription factor *Bna.MYB107* annotated for response to ABA was up-regulated in this genotype and could possibly be involved in ABA mediated response to osmotic stress. In *A. thaliana*, *MYB*

genes were found to be responsive to stress treatments (Yanhui et al. 2006). It has been reported that transcription factors such as *MYB* and *DREB* increase glucose, fructose and glycine betaine levels in *A. thaliana*. Glycine is a precursor of glycine betaine (Krasensky and Jonak 2012). In the genotype KWS12, the gene *BnaA08g04520D* annotated for glycine catabolism was down-regulated supporting the increase in glycine.

In the module with highest correlation to proline (*shoot-thistle2*), the genotype Lisek had the highest absolute value of module eigengene. This genotype also had the highest fold change of proline content in osmotic stress. The genotype decreased in biomass under stress although it increased in proline content. Stronger fold changes of proline associated to decrease in biomass were reported earlier for drought sensitive *B. napus* genotypes (Hatzig et al. 2014). Among the genes in this module involved in the proline biosynthesis pathway, *BnaA10g26830D* was annotated for proline transport and up-regulated in this study. Proline transport is an important process during osmotic stress induced proline accumulation. The flux of proline into and out of compartments is needed for movement of proline among different parts of the plant (Verslues and Sharma 2010). *BnaA04g08160D* or glutamine synthase is also from the proline biosynthesis pathway (Verslues and Sharma 2010) and was down-regulated in the genotype Lisek. This is because glutamate availability, a precursor to proline, is controlled by the action of glutamine synthetase, which uses glutamate and ammonium to produce glutamine. Glutamate is then regenerated by the action of another enzyme called glutamate synthase (Forde and Lea 2007). *Bna.P5CS* was up-regulated. This gene produces an enzyme that is intermediate in proline biosynthesis from glutamate (Verslues and Sharma 2010). Osmotic response hub gene *BnaA09g39170D* or *Bna.PIP1* was down-regulated. Plasma membrane intrinsic proteins (PIPs) mediate water transport in plants. They are down-regulated to reduce osmotic hydraulic conductance in roots (Mahdieh et al. 2008). The transcription factors *BnaC05g33570D (Bna.ABF4)* and *BnaC06g40040D* (*Bna.AP2)* involved in drought stress signaling in plants (Shinozaki and Yamaguchi-Shinozaki 2007) were down-regulated.

Similarly, the other module correlated to proline (*Shoot-Darkorange2*) had the top hub gene *BnaC02g06660D* (*AT5G16510*) that was annotated as response to salt stress (Dixon et al. 2005). Among the hub genes, *BnaA01g02740D, BnaC06g21690D* and *BnaA08g10990D* are hub genes annotated for ABA mediated signaling response to osmotic stress. ABA and calcium mediated signaling is known to be signaling for the accumulation of proline in response to drought stress (Verbruggen and Hermans 2008, Krasensky and Jonak 2012). The

gene *Bna.GDH2* (EC 1.4.1.3) from the proline biosynthesis pathway (Verslues and Sharma 2010) was also found in this module correlated to proline. The stressed genotypes NPZ09 and DSV09 decreased in biomass while Gross-Luesewitzer (Conv124) and SWGospel (Conv49) did not show change in biomass in stress but showed fold change in proline. This indicates that the genotypes adjusted for osmotic stress are using the osmolyte proline and prevented loss in biomass due to the stress. It was found in this study that in addition to proline biosynthesis, these genotypes resisted osmotic stress using ABA mediated signaling response to osmotic stress. The top hub gene *BnaC02g06660D* annotated as response to salt stress (Dixon et al. 2005) and the hub gene *BnaA01g02740D* annotated for ABA mediated signaling response to osmotic stress were up-regulated for Gross-Luesewitzer and SWGospel. In contrast, these genes were down-regulated for NPZ09 and DSV09 suggesting these genotypes lack transcriptomic adjustment to osmotic stress thus decrease in biomass. *BnaC05g31510D* (*Bna.ATEIN3*) ethylene insensitive 3-like was up-regulated by the high performance genotypes both in terms of not changing in *in-vitro* biomass in stress as well as high performance genotypes in multiple year, multiple environment field emergence trials and germination tests. In contrast it was down-regulated in the low performance genotypes in osmotic stress, germination and field emergence. The ethylene signaling pathway is more complex, due to the large number of genes that are involved and the multiple processes and responses that are regulated. Different elements of this proposed ethylene signaling pathway were reported to be associated with stress responses after drought and salt stress. Ethylene interferes with ABA mediated closure of stomata. Thereby mutants showed a delayed stomatal closure, resulting in a higher degree of water loss by transpiration during drought. The signaling network includes the receptor complexes I (*ETR1/ERS1*) and II (*ETR2/ERS2/EIN4*), that are membrane associated and connected to a negative constitutive triple response 1 (*CTR1*) protein kinase. Complex I consists of ethylene response 1 (*ETR1*) and ethylene response sensor 1 (*ERS1*), while complex II has three components consisting of *ETR2, ERS2* and ethylene insensitive 4 (*EIN4*). In absence of ethylene *CTR1* negatively regulates the ethylene insensitive 2 (*EIN2*) protein. After binding of ethylene to one of the receptor subfamilies, a signal peptide (*EIN2C*) is cleaved from EIN2 and transported into the nucleus, where the ethylene insensitive 3 (*EIN3*)/ethylene insensitive like 1 (*EIL1*) dependent gene transcription is activated (Krannich et al. 2015).

The module highly correlated to ornithine and beta-alanine (*Shoot-Palevioletred3*) also contains genes involved in osmotic stress resistance. Ornithine is a precursor to the alternate pathway of proline metabolism (Verslues and Sharma 2010). *Bna.GLN1;4* or glutamine

synthase 1.4 is also from the proline biosynthesis pathway (Verslues and Sharma 2010) and was down-regulated in the genotype Laser, causing increase in proline in the process mentioned above. This genotype reduced in biomass in stress and increased in fold change of the metabolites ornithine, betaalanine, valine and phenylalanine. This suggests response to osmotic stress via osmolyte adjustment in the genotype Laser. However, the concentration of ornithine in osmotic stress was decreased in drought sensitive genotypes in the study of (Hatzig et al. 2014). The two top hub genes annotated for osmotic stress response and ABA mediated signaling pathway*: Bna.MKK2* and *BnaC03g63530D* were up-regulated in the genotype. In *A. thaliana*, *MKK2* is a mitogen-activated protein kinase (MAPK) that mediates activation of MAPK cascade in response to osmotic stress (Teige et al. 2004). *BnaC03g63530D* is a GTP-binding protein. GTP-binding proteins are involved in many functions in the cell including signal transduction of osmotic stress stimuli (Agarwal et al. 2008).

A calcineurin b-like protein *Bna.CBL2* was the top hub gene in the consensus module correlated to both root and shoot traits (*Consensus-lightcoral*). In *A. thaliana*, the *CBL2* gene is known to be expressed in both shoot and root during seedling development. This gene interacts with genes up-regulated in response to osmotic stress as calcium sensor in calcium mediated signaling (Pandey et al. 2015). However in this study, there was no change in expression profile of this gene in the genotypes except for down-regulation in KWS12.

A zinc finger protein, *Bna.ZF2* was a top hub gene in the consensus module correlated to root metabolites (*Consensus-Darkolivegreen*). In *A. thaliana*, *ZF2* functions as transcriptional repressor to reduce plant growth under abiotic stress conditions and expressed in both root and shoot (Kodaira et al. 2011). This gene was up-regulated in both root and shoot of KWS12 and DSV9 but down-regulated in Laser (Conv41), NPZ9 and Lisek (Conv33). And all of these genotypes decreased in biomass in osmotic stress. In case of the genotypes Gross-Luesewitzer (Conv124) and SWGospel (Conv49), the gene was up-regulated in one tissue and down-regulated in another tissue. These genotypes did not change in biomass when stressed.

5. Conclusions

In the first part of the study, WGCNA was used to reduce a large dataset initially containing about 82,000 expressed genes from 4 week old seedlings, into modules correlated to germination traits up to 3 days old seedlings and multi environmental field emergence traits at 2-4 week old seedlings. Combining QTL analysis with WGCNA has helped to prioritize additional candidate genes other than the candidate genes in (Hatzig et al. 2015) from the QTL region in QTL trait associated modules. This presents a new approach to confirm genes controlling TSW in a subset of GWAS diversity set and additional genotypes. Such genes strongly supported as a candidate by the network analysis was *Bna.PME3* and *BnaA01g04540D*, a cytochrome bd ubiquinol oxidase gene that were associated to thousand seed weight and found in modules associated to the QTL *in-vitro* thousand seed weight trait as well as field emergence traits. Seed quality traits including thousand seed weight and environmental effects affect establishment in the field. Thus, breeding for thousand seed weight and combining favourable alleles for TSW and other vigour-related QTL and candidate genes might be one approach to increase germination and vigour traits in oilseed rape. In addition, inside multi-environment field emergence associated modules, connectivity of top hub genes especially from the auxin signalling pathway suggest a role in germination and seedling development for the genes. This study will aid in selection of markers for improving seedling vigour in *B. napus* by complementing genomic studies through identification of regulatory and QTL candidate genes involved in seedling development.

In the second study, WGCNA was used to correlate gene expression profiles in different genotypes to metabolites and biomass. This method was useful to pinpoint the key variations in the response to osmotic stress of a diverse set of genotypes. This study was further supported by the study of (Zhang et al. 2015), whose candidate genes from the QTL associated to water stress was supported by the network analysis. The genotypes were found to have varied response to osmotic stress at both transcriptional and metabolic level. Activation of genes leading to drought tolerance was observed in genotypes that showed increase in fold change of osmolytes and reduced in biomass in osmotic stress. The genotypes that decreased in biomass in the stress treatment also went through activation of osmotic stress responsive and osmolyte biosynthesis genes. The other group of genes found to be important hub genes are involved in ABA mediated signaling response to osmotic stress. Hub genes that are in the metabolite biosynthesis pathway of the corresponding metabolite to which the module correlated to could be used as confirmation of the network. Similar activation of genes was also found in genotypes that did not reduce in biomass in stress. The

genotype with the highest fold change of proline decreased in biomass although osmotic responsive genes were activated. No decrease in biomass occurred when proline increased moderately coupled with up-regulation of genes involved in ABA mediated response to osmotic stress. *In-vitro* biomass in osmotic stress and gene expression can be useful as early markers for stress tolerance/susceptibility and field emergence at the same time.

In addition to previously reported osmolytes such as proline and sugars, glycine and betaalanine were highly correlated to shoot modules with clusters of genes involved in osmotic stress response. Overlapping expression networks between shoot and root are related to different metabolites in the two tissues: leucine, valine, and threonine important in root and fructose, glucose, glycine and *myo*-inositol in root. Methionine in both root and shoot was correlated to the module containing genes with a common expression pattern in both the tissues.

6. Summary

The aim of the first part of the thesis was to confirm the candidate genes with putative network regulatory roles during seedling emergence. A transcriptome-based systems analysis of seedling emergence was performed in a panel of 42 winter-type *Brassica napus* L. accessions with broad genetic and phenotypic diversity. Using weighted gene co-expression network analysis (WGCNA) shoot and root gene co-expression modules from four weeks old seedlings were correlated to *in-vitro* germination parameters and multi-environment field emergence traits. Combining QTL analysis with gene co-expression network analysis helped to prioritize candidate genes that lie within QTL intervals for seed germination and vigour. Several newly reported genes that were associated to the QTL were strongly supported as candidates by the network analysis. Such genes associated to QTL thousand seed weight were *Bna.PME3* and *BnaA01g04540D*, a cytochrome bd ubiquinol oxidase gene found in gene co-expression module eigengenes correlated to thousand seed weight trait as well as field emergence traits. Seed quality traits including thousand seed weight as well as environmental effects affect establishment in the field. In addition, inside multi-environment field emergence associated modules, top hub genes especially from the auxin signalling pathway suggest a role in germination and seedling development for the genes. This study will aid in selection of markers for improving seedling vigour in *B. napus* by complementing genomic studies through identification of regulatory and QTL candidate genes involved in seedling development.

In the second part of the thesis, fold change of expressed genes in osmotic stress (PEG) in the same panel of diverse 42 winter-type *B. napus* accessions were clustered using a WGCNA approach and the module eigengenes were correlated to metabolites and biomass. Clusters of expressed genes associated to metabolites and biomass were enriched for osmotic stress resistance. They contained hub genes involved in metabolite biosynthesis, osmotic stress responses and resistance mechanisms. These results were further supported by another previously published study whose candidate genes from the QTL associated to water stress was supported by the network analysis. The present study has pinpointed the key variations in the response to osmotic stress of a diverse set of genotypes. The genotypes were found to have varied response to osmotic stress at both transcriptional and metabolic level. In addition to previously reported osmolytes such as proline and sugars, glycine and betaalanine were found to be associated to osmotic stress related gene co-expression networks. The study provides insight into the responses of different *B. napus* genotypes to osmotic stress to contribute to the improvement of the crop.

7. Zusammenfassung

Das Ziel des ersten Teils dieser Arbeit war die Bestätigung von Kandidatengenen mit putativ regulatorischer Funktion in Gennetzwerken während der Keimlingsentwicklung. In einer Auswahl von 42 Wintertypen von *Brassica napus* L. mit hoher genetischer und phänotypischer Diversität wurde eine transkriptombasierte Systemanalyse der Keimlingsentwicklung durchgeführt. Durch gewichtete Gen-Koexpressions-Netzwerkanalyse (WGCNA) wurden Sproß-und Wurzel-Koexpressionsmodule von vier Wochen alten Keimlingen mit *in-vitro*-Keimlingsparametern und Feldaufgangsdaten aus verschiedenen Umwelten korreliert. Die Kombination von QTL-Analyse und Koexpressions-Netzwerk-Analyse half bei der Priorisierung von Kandidatengenen, die in QTL-Intervallen für Samenkeimung und Triebkraft lagen. Einige der neu beschriebenen Gene, die mit QTLs assoziierten, wurden durch die Netzwerkanalyse als Kandidatengene deutlich unterstützt. Solche Gene, die mit Tausendkorngewicht-QTL assoziiert waren, waren *Bna.PME3* und *BnaA01g04540D*, eine Cytochrom-bd-Ubiquinol-Oxidase, die in Eigengenen von Koexpressionsmodulen gefunden wurde, die sowohl mit Tausendkorngewicht als auch mit Feldaufgang assoziieren. Parameter für Samenqualität wie Tausendkorngewicht sowie Umwelteffekte beeinflussen die Etablierung auf dem Feld. Außerdem scheinen in Modulen, die mit Feldaufgang in vielen Umwelten assoziieren, zentrale Steuerungsgene des Auxinsignalwegs eine Rolle für Keimung und Feldaufgang zu spielen. Diese Studie wird helfen, Marker für die Verbesserung der Keimlingstriebkraft in *B. napus* zu selektieren, indem sie genomische Studien durch die Identifizierung von regulatorischen und QTL-Kandidatengenen aus der Samenentwicklung ergänzt.

Im zweiten Teil dieser Arbeit wurde die Änderung der Genexpressionsmuster unter osmotischem Stress in derselben Auswahl von 42 diversen *B. napus* Genotypen mit einem WGCNA-Ansatz gruppiert und die Eigengene der Module wurden mit Metaboliten und Biomasse korreliert. Cluster von exprimierten Genen, die mit Metaboliten und Biomasse assoziierten, wurden angereichert. Sie enthielten Steuerungsgene, die an der Biosynthese von Metaboliten, der osmotischen Stressantwort und an Resistenzmechanismen beteiligt sind. Diese Ergebnisse wurden weiterhin durch eine andere bisher unveröffentlichte Studie unterstützt, deren Kandiatengene aus einem Wasserstress-QTL von der Netzwerkanalyse bestätigt werden. Die vorliegende Arbeit hat die wesentlichen Variablen der osmotischen Stressantwort in diversen Genotypen herausgestellt. Demzufolge haben die Genotypen sowohl auf der transkriptionellen wie auch auf der metabolischen Ebene unterschiedliche Arten, zu reagieren. Zusätzlich zu den bislang identifizierten Osmolyten wie Prolin und verschiedenen Zuckern, wurden Glycin und beta-Alanin mit Genen assoziiert, die sich in Koexpressionsnetzwerken für osmotischen Stress befinden. Diese Studie gibt einen Einblick in die osmotische Stressantwort verschiedener *B. napus*-Genotypen und kann damit zu einer Verbesserung dieser Kulturpflanze beitragen.

8. References

Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy ASN (2016) A survey of the sorghum transcriptome using single-molecule long reads. Nature communications 7:11706. doi: 10.1038/ncomms11706

Agarwal PK, Agarwal P, Jain P, Jha B, Reddy MK, Sopory SK (2008) Constitutive overexpression of a stress-inducible small GTP-binding protein PgRab7 from Pennisetum glaucum enhances abiotic stress tolerance in transgenic tobacco. Plant cell reports 27(1):105–115. doi: 10.1007/s00299-007-0446-0

Albert B, Le Cahérec F, Niogret M-F, Faes P, Avice J-C, Leport L, Bouchereau A (2012) Nitrogen availability impacts oilseed rape (Brassica napus L.) plant water status and proline production efficiency under water-limited conditions. Planta 236(2):659–676. doi: 10.1007/s00425-012-1636-8

Allen JD, Xie Y, Chen M, Girard L, Xiao G (2012) Comparing statistical methods for constructing large scale gene networks. PLoS ONE 7. doi: 10.1371/journal.pone.0029348

An F, Zhang X, Zhu Z, Ji Y, He W, Jiang Z, Li M, Guo H (2012) Coordinated regulation of apical hook development by gibberellins and ethylene in etiolated Arabidopsis seedlings. Cell research 22(5):915–927. doi: 10.1038/cr.2012.29

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nature protocols 8(9):1765–1786. doi: 10.1038/nprot.2013.099

Anders S, Pyl PT, Huber W (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics (Oxford, England) 31(2):166–169. doi: 10.1093/bioinformatics/btu638

Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T (2016) ATTED-II in 2016: A Plant Coexpression Database Towards Lineage-Specific Coexpression. Plant & cell physiology 57(1):e5. doi: 10.1093/pcp/pcv165

Babraham Bioinformatics. (2011) FastQC [Software]: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bergmann S, Ihmels J, Barkai N (2004) Similarities and differences in genome-wide expression data of six organisms. PLoS biology 2(1):E9. doi: 10.1371/journal.pbio.0020009

BLAKE J, SPINK J, BULLARD M (2004) Successful establishment of oilseed rape,. HGCA conference 2004: Managing soil and roots for profitable production:11.1-11.12

Boudsocq M, Laurière C (2005) Osmotic signaling in plants: multiple pathways mediated by emerging kinase families. PLANT PHYSIOLOGY 138(3):1185–1194. doi: 10.1104/pp.105.061275.

Boyer JS (2010) Drought decision-making. Journal of Experimental Botany 61(13):3493–3497. doi: 10.1093/jxb/erq231

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nature biotechnology 34(5):525–527. doi: 10.1038/nbt.3519

Castelain M, Le Hir R, Bellini C (2012) The non-DNA-binding bHLH transcription factor PRE3/bHLH135/ATBS1/TMO7 is involved in the regulation of light signaling pathway in Arabidopsis. PHYSIOLOGIA PLANTARUM 145(3):450–460. doi: 10.1111/j.1399-3054.2012.01600.x

Chalhoub B, Denoeud F, Liu S, Parkin, Isobel A. P., Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Corréa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Wang X, Canaguier A, Chauveau A, Bérard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Wang X, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science (New York, N.Y.) 345(6199):950–953. doi: 10.1126/science.1253435

Chawla K, Barah P, Kuiper M, Bones A (2011) Systems Biology: A Promising Tool to Study Abiotic Stress Responses. Omics and Plant Abiotic Stress Tolerance. Bentham Science Publishers

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang P-L, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2(10):2366–2382. doi: 10.1038/nprot.2007.324

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676. doi: 10.1093/bioinformatics/bti610

Dai A (2011) Drought under global warming: a review. WIREs Clim Change 2(1):45–65. doi: 10.1002/wcc.81

de los Reyes BG, Myers SJ, McGrath JM (2003) Differential induction of glyoxylate cycle enzymes by stress as a marker for seedling vigor in sugar beet (*Beta vulgaris*). Molecular Genetics and Genomics 269(5):692–698. doi: 10.1007/s00438-003-0875-6

Deleu C, Faes P, Niogret M-F, Bouchereau A (2013) Effects of the inhibitor of the γ-aminobutyrate-transaminase, vinyl-γ-aminobutyrate, on development and nitrogen metabolism in Brassica napus seedlings. Plant physiology and biochemistry : PPB 64:60–69. doi: 10.1016/j.plaphy.2012.12.007

Dharmasiri N, Dharmasiri S, Weijers D, Lechner E, Yamada M, Hobbie L, Ehrismann JS, Jürgens G, Estelle M (2005) Plant development is regulated by a family of auxin receptor F box proteins. Developmental cell 9(1):109–119. doi: 10.1016/j.devcel.2005.05.014

Dixon DP, Skipsey M, Grundy NM, Edwards R (2005) Stress-induced protein S-glutathionylation in Arabidopsis. PLANT PHYSIOLOGY 138(4):2233–2244. doi: 10.1104/pp.104.058917

D'Mello JPF (2015) Amino acids in higher plants. Amino acid synthesis under abiotic stress. CABI

Duan W, Huang Z, Song X, Liu T, Liu H, Hou X, Li Y (2016) Comprehensive analysis of the polygalacturonase and pectin methylesterase genes in *Brassica rapa* shed light on their different evolutionary patterns. Scientific reports 6:25107. doi: 10.1038/srep25107

Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet 126(1):1–11. doi: 10.1007/s00122-012-1964-x

Egli DB, Rucker M (2012) Seed Vigor and the Uniformity of Emergence of Corn Seedlings. Crop Science 52(6):2774–2782. doi: 10.2135/cropsci2012.01.0064

Elliott RH, Franke C, Rakow GFW (2008) Effects of seed size and seed weight on seedling establishment, vigour and tolerance of Argentine canola (Brassica napus) to flea beetles, Phyllotreta spp. Can. J. Plant Sci. 88(1):207–217. doi: 10.4141/CJPS07059

Ellis R (1992) Seed and seedling vigour in relation to crop growth and yield. Plant growth and regulation 11: 249-255

Everaert C, Luypaert M, Maag JLV, Cheng QX, Dinger ME, Hellemans J, Mestdagh P (2017) Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Scientific reports 7(1):1559. doi: 10.1038/s41598-017-01617-3

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics (Oxford, England) 32(19):3047–3048. doi: 10.1093/bioinformatics/btw354

Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. Bioinformatics (Oxford, England) 23(2):257–258. doi: 10.1093/bioinformatics/btl567

Fang Y, Xiong L (2015) General mechanisms of drought response and their application in drought resistance improvement in plants. Cellular and molecular life sciences : CMLS 72(4):673–689. doi: 10.1007/s00018-014-1767-0

Fennell H, Olawin A, Mizanur RM, Izumori K, Chen J-G, Ullah H (2012) Arabidopsis scaffold protein RACK1A modulates rare sugar D-allose regulated gibberellin signaling. Plant Signaling & Behavior 7(11):1407–1410. doi: 10.4161/psb.21995

Ficklin SP, Luo F, Feltus FA (2010) The Association of Multiple Interacting Genes with Specific Phenotypes in Rice Using Gene Coexpression Networks. PLANT PHYSIOLOGY 154(1):13–24. doi: 10.1104/pp.110.159459

Forde BG, Lea PJ (2007) Glutamate in plants: metabolism, regulation, and signalling. Journal of Experimental Botany 58(9):2339–2358. doi: 10.1093/jxb/erm121

Friedt W, Snowdon R (2010) Oilseed rape. In: Vollmann J, Rajcan I. (eds) Handbook of Plant Breeding, Vol. 4: Oil crops breeding. Springer Verlag, Berlin/Heidelberg, pp 91-126

Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S (2007) Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm. Genome 18(6-7):463–472. doi: 10.1007/s00335-007-9043-3

Gentleman R, Falcon S (2013) Package 'GOstats'. Tools for manipulating GO and microarrays. https://www.bioconductor.org/packages/release/bioc/html/GOstats.html

Getinet A, Rakow G, Raney J, Downey R (1997) Glucosinolate content in interspecific crosses of *Brassica carinata* with *B. juncea* and *B. napus*. Plant Breeding 116:39–46

Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S (2006) Integrating Genetic and Network Analysis to Characterize Genes Related to Mouse Weight. PLoS Genet 2(8):e130. doi: 10.1371/journal.pgen.0020130

Gibson G, Weir B (2005) The quantitative genetics of transcription. Trends in genetics : TIG 21(11):616–623. doi: 10.1016/j.tig.2005.08.010

Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome biology 11(8):R86. doi: 10.1186/gb-2010-11-8-r86

Gonzalez-Garay ML (2015) Introduction to Isoform Sequencing Using Pacific Biosciences Technology (Iso-Seq). In: Jiaqian Wu (ed) Transcriptomics and Gene Regulation Volume 9 of Translational Bioinformatics. springer, pp 141–160

Good A, Zaplachinski S (1994) The effects of drought stress on free amino acid accumulation and protein synthesis in Brassica napus. PHYSIOLOGIA PLANTARUM 90:9–14

Good AG, Maclagan JL (1993) Effects of drought stress on the water relations in Brassica species. Can. J. Plant Sci 73:525–529

Hagen G, Guilfoyle T Auxin-responsive gene expression: genes, promoters and regulatory factors:373–385. doi: 10.1007/978-94-010-0377-3_9

Halbritter F, Kousa AI, Tomlinson SR (2014) GeneProf data: a resource of curated, integrated and reusable high-throughput genomics experiments. Nucleic acids research 42(Database issue):D851-8. doi: 10.1093/nar/gkt966

Hanson AD, Hitz WD (1982) Metabolic Responses of Mesophytes to Plant Water Deficits. Annual review of plant physiology 33:163–203

Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I (2012) Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nature biotechnology 30(8):798–802. doi: 10.1038/nbt.2302

Hatzig S, Zaharia LI, Abrams S, Hohmann M, Legoahec L, Bouchereau A, Nesi N, Snowdon RJ (2014) Early osmotic adjustment responses in drought-resistant and drought-sensitive oilseed rape. Journal of integrative plant biology 56(8):797–809. doi: 10.1111/jipb.12199

Hatzig SV, Frisch M, Breuer F, Nesi N, Ducournau S, Wagner M-H, Leckband G, Abbadi A, Snowdon RJ (2015) Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. Front. Plant Sci. 6. doi: 10.3389/fpls.2015.00221

Hayashi K (2012) The Interaction and Integration of Auxin Signaling Components. Plant and Cell Physiology 53(6):965–975. doi: 10.1093/pcp/pcs035

Hocq L, Sénéchal F, Lefebvre V, Lehner A, Domon J-M, Mollet J-C, Dehors J, Pageau K, Marcelo P, Guérineau F, Kolšek K, Mercadante D, Pelloux J (2017) Combined Experimental and Computational Approaches Reveal Distinct pH Dependence of Pectin Methylesterase Inhibitors. PLANT PHYSIOLOGY 173(2):1075–1093. doi: 10.1104/pp.16.01790

Hong J, Yang L, Zhang D, Shi J (2016) Plant Metabolomics: An Indispensable System Biology Tool for Plant Science. International journal of molecular sciences 17(6). doi: 10.3390/ijms17060767

Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, Laurance MF, Zhao W, Qi S, Chen Z, Lee Y, Scheck AC, Liau LM, Wu H, Geschwind DH, Febbo PG, Kornblum HI, Cloughesy TF, Nelson SF, Mischel PS (2006) Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. PNAS 103(46):17402–17407. doi: 10.1073/pnas.0608396103

Hu J, Uapinyoying P, Goecks J (2017) Interactive analysis of Long-read RNA isoforms with Iso-Seq Browser. doi: 10.1101/102905

Hwang Y, Choi H-S, Cho H-M, Cho H-T (2017) Tracheophytes Contain Conserved Orthologs of a Basic Helix-Loop-Helix Transcription Factor That Modulate ROOT HAIR SPECIFIC Genes. The Plant Cell 29(1):39–53. doi: 10.1105/tpc.16.00732

Jain M, Olsen HE, Paten B, Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome biology 17(1):239. doi: 10.1186/s13059-016-1103-0

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L, (None) (2000) The large-scale organization of metabolic networks. Nature 407:651–654

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics (Oxford, England) 30(9):1236–1240. doi: 10.1093/bioinformatics/btu031

Joshi V, Joung J-G, Fei Z, Jander G (2010) Interdependence of threonine, methionine and isoleucine metabolism in plants: accumulation and transcriptional regulation under abiotic stress. Amino acids 39(4):933–947. doi: 10.1007/s00726-010-0505-7

Kazama D, Kurusu T, Mitsuda N, Ohme-Takagi M, Tada Y (2014) Involvement of elevated proline accumulation in enhanced osmotic stress tolerance in Arabidopsis conferred by chimeric repressor gene silencing technology. Plant Signaling & Behavior 9(3):e28211. doi: 10.4161/psb.28211

Khurana JP, Mukesh Jain, Akhilesh K. Tyagi (2007) Auxin and Cytokinin Signaling Component Genes and Their Potential for Crop Improvement. In: Rajeev K. Varshney,, Roberto Tuberosa (eds) Genomics-assisted crop improvement. springer, pp 289–314

Kleessen S, Nikoloski Z (2012) Dynamic regulatory on/off minimization for biological systems under internal temporal perturbations. BMC systems biology 6:16. doi: 10.1186/1752-0509-6-16

Kodaira K-S, Qin F, Tran L-SP, Maruyama K, Kidokoro S, Fujita Y, Shinozaki K, Yamaguchi-Shinozaki K (2011) Arabidopsis Cys2/His2 zinc-finger proteins AZF1 and AZF2 negatively regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions. PLANT PHYSIOLOGY 157(2):742–756. doi: 10.1104/pp.111.182683

Koh J, Chen G, Yoo M-J, Zhu N, Dufresne D, Erickson JE, Shao H, Chen S (2015) Comparative Proteomic Analysis of Brassica napus in Response to Drought Stress. Journal of proteome research 14(8):3068–3081. doi: 10.1021/pr501323d

Körber N, Bus A, Li J, Higgins J, Bancroft I, Higgins EE, Parkin IAP, Salazar-Colqui B, Snowdon RJ, Stich B (2015) Seedling development traits in *Brassica napus* examined by gene expression analysis and association mapping. BMC Plant Biology 15:136. doi: 10.1186/s12870-015-0496-3

Krannich CT, Maletzki L, Kurowsky C, Horn R (2015) Network Candidate Genes in Breeding for Drought Tolerant Crops. International journal of molecular sciences 16(7):16378–16400. doi: 10.3390/ijms160716378

Krasensky J, Jonak C (2012) Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. Journal of Experimental Botany 63(4):1593–1608. doi: 10.1093/jxb/err460

Kuno N (2003) The Novel MYB Protein EARLY-PHYTOCHROME-RESPONSIVE1 Is a Component of a Slave Circadian Oscillator in Arabidopsis. THE PLANT CELL ONLINE 15(10):2476–2488. doi: 10.1105/tpc.014217

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic acids research 40(Database issue):10. doi: 10.1093/nar/gkr1090

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

Langfelder P, Mischel PS, Horvath S (2013) When Is Hub Gene Selection Better than Standard Meta-Analysis? PLoS ONE 8(4). doi: 10.1371/journal.pone.0061505.t001

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

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10. doi: 10.1186/gb-2009-10-3-r25

Lee S, Lee S, Yang K-Y, Kim Y-M, Park S-Y, Kim SY, Soh M-S (2006) Overexpression of PRE1 and its homologous genes activates Gibberellin-dependent responses in *Arabidopsis thaliana*. Plant & cell physiology 47(5):591–600. doi: 10.1093/pcp/pcj026

Lee SC, Luan S (2012) ABA signal transduction at the crossroad of biotic and abiotic stress responses. Plant, cell & environment 35(1):53–60. doi: 10.1111/j.1365-3040.2011.02426.x

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 25(14):1754–1760. doi: 10.1093/bioinformatics/btp324

Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. Bioinformatics (Oxford, England) 24(5):713–714. doi: 10.1093/bioinformatics/btn025

Liu C, Zhang X, Zhang K, An H, Hu K, Wen J, Shen J, Ma C, Yi B, Tu J, Fu T (2015) Comparative Analysis of the Brassica napus Root and Leaf Transcript Profiling in Response to Drought Stress. International journal of molecular sciences 16(8):18752–18777. doi: 10.3390/ijms160818752

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin, Isobel A. P., Zhao M, Ma J, Yu J, Huang S, Wang X, Wang J, Lu K, Fang Z, Bancroft I, Yang T-J, Hu Q, Wang X, Yue Z, Li H, Yang L, Wu J, Zhou Q, Wang W, King GJ, Pires JC, Lu C, Wu Z, Sampath P, Wang Z, Guo H, Pan S, Yang L, Min J, Zhang D, Jin D, Li W, Belcram H, Tu J, Guan M, Qi C, Du D, Li J, Jiang L, Batley J, Sharpe AG, Park B-S, Ruperao P, Cheng F, Waminal NE, Huang Y, Dong C, Wang L, Li J, Hu Z, Zhuang M, Huang Y, Huang J, Shi J, Mei D, Liu J, Lee T-H, Wang J, Jin H, Li Z, Li X, Zhang J, Xiao L, Zhou Y, Liu Z, Liu X, Qin R, Tang X, Liu W, Wang Y, Zhang Y, Lee J, Kim HH, Denoeud F, Xu X, Liang X, Hua W, Wang X, Wang J, Chalhoub B, Paterson AH (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Comms 5. doi: 10.1038/ncomms4930

Lu K, Peng L, Zhang C, Lu J, Yang B, Xiao Z, Liang Y, Xu X, Qu C, Zhang K, Liu L, Zhu Q, Fu M, Yuan X, Li J (2017) Genome-Wide Association and Transcriptome Analyses

Reveal Candidate Genes Underlying Yield-determining Traits in Brassica napus. Front. Plant Sci. 8:3389. doi: 10.3389/fpls.2017.00206

Lugan R, Niogret M-F, Kervazo L, Larher FR, Kopka J, Bouchereau A (2009) Metabolome and water status phenotyping of Arabidopsis under abiotic stress cues reveals new insight into ESK1 function. Plant, cell & environment 32(2):95–108. doi: 10.1111/j.1365-3040.2008.01898.x

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nature reviews. Genetics 10(8):565–577. doi: 10.1038/nrg2612

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nature reviews. Genetics 12(10):671–682. doi: 10.1038/nrg3068

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j. 17(1):10. doi: 10.14806/ej.17.1.200

Mazzucotelli E, Belloni S, Marone D, DE Leonardis A, Guerra D, Fonzo N, Cattivelli L, Mastrangelo A (2006) The E3 Ubiquitin Ligase Gene Family in Plants: Regulation by Degradation. Current Genomics:509–522

McIntyre LM, Lopiano KK, Morse AM, Amin V, Oberg AL, Young LJ, Nuzhdin SV (2011) RNA-seq: technical variability and sampling. BMC Genomics 12:293. doi: 10.1186/1471-2164-12-293

Merewitz EB, Du H, Yu W, Liu Y, Gianfagna T, Huang B (2012) Elevated cytokinin content in ipt transgenic creeping bentgrass promotes drought tolerance through regulating metabolite accumulation. Journal of Experimental Botany 63(3):1315–1328. doi: 10.1093/jxb/err372

Mezlini AM, Smith EJM, Fiume M, Buske O, Savich GL, Shah S, Aparicio S, Chiang DY, Goldenberg A, Brudno M (2013) iReckon: simultaneous isoform discovery and abundance estimation from RNA-seq data. Genome Research 23(3):519–529. doi: 10.1101/gr.142232.112

Molitor AM, Bu Z, Yu Y, Shen W-H (2014) Arabidopsis AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes. PLoS genetics 10(1):e1004091. doi: 10.1371/journal.pgen.1004091

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods 5(7):621–628. doi: 10.1038/nmeth.1226

Muday GK, Rahman A, Binder BM (2012) Auxin and ethylene: collaborators or competitors? Trends in Plant Science 17(4):181–195. doi: 10.1016/j.tplants.2012.02.001

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annual review of plant biology 59:651–681. doi: 10.1146/annurev.arplant.59.032607.092911

Nahirñak V, Almasia N, Hopp, E, Vazquez-Rovere C (2012) Snakin/GASA proteins Involvement in hormone crosstalk and redox homeostasis. Plant Signaling & Behavior 7(8):1004–1008

Obermeier C, Friedt W (2015) Applied Oilseed Rape Marker Technology and Genomics. In: Poltronieri P, Hong Y (eds) Applied Plant Genomics and Biotechnology. Woodhead Publishing Limited, Elsevier, Cambridge, UK, pp 253–295

Obermeier C, Hosseini B, Friedt W, Snowdon R (2009) Gene expression profiling via LongSAGE in a non-model plant species: a case study in seeds of *Brassica napus*. BMC Genomics 10:295. doi: 10.1186/1471-2164-10-295

Olvera-Carrillo Y, Campos F, Reyes JL, Garciarrubio A, Covarrubias AA (2010) Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in Arabidopsis. PLANT PHYSIOLOGY 154(1):373–390. doi: 10.1104/pp.110.158964

Pandey GK, Kanwar P, Singh A, Steinhorst L, Pandey A, Yadav AK, Tokas I, Sanyal SK, Kim B-G, Lee S-C, Cheong Y-H, Kudla J, Luan S (2015) Calcineurin B-Like Protein-Interacting Protein Kinase CIPK21 Regulates Osmotic and Salt Stress Responses in Arabidopsis. PLANT PHYSIOLOGY 169(1):780–792. doi: 10.1104/pp.15.00623

Parry G, Calderon-Villalobos LI, Prigge M, Peret B, Dharmasiri S, Itoh H, Lechner E, Gray WM, Bennett M, Estelle M (2009) Complex regulation of the TIR1/AFB family of auxin receptors. PNAS 106(52):22540–22545. doi: 10.1073/pnas.0911967106

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Meth 14(4):417–419

Peleg Z, Blumwald E (2011) Hormone balance and abiotic stress tolerance in crop plants. Current Opinion in Plant Biology 14(3):290–295. doi: 10.1016/j.pbi.2011.02.001

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical computing

Rahman A (2013) Auxin: a regulator of cold stress response. PHYSIOLOGIA PLANTARUM 147(1):28–35. doi: 10.1111/j.1399-3054.2012.01617.x

Risso D, Schwartz K, Sherlock G, Dudoit S (2011) GC-content normalization for RNA-Seq data. BMC Bioinformatics 12:480. doi: 10.1186/1471-2105-12-480

Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. Genome biology 12(3):R22. doi: 10.1186/gb-2011-12-3-r22

Rockman MV, Kruglyak L (2006) Genetics of global gene expression. Nature reviews. Genetics 7(11):862–872. doi: 10.1038/nrg1964

Schlereth A, Moller B, Liu W, Kientz M, Flipse J, Rademacher EH, Schmid M, Jurgens G, Weijers D (2010) MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature 464(7290):913–916. doi: 10.1038/nature08836

Schmutzer T, Samans B, Dyrszka E, Ulpinnis C, Weise S, Stengel D, Colmsee C, Lespinasse D, Micic Z, Abel S, Duchscherer P, Breuer F, Abbadi A, Leckband G, Snowdon R, Scholz U (2015) Species-wide genome sequence and nucleotide polymorphisms from the model allopolyploid plant Brassica napus. Sci. Data 2:150072. doi: 10.1038/sdata.2015.72

Schulz P, Herde M, Romeis T (2013) Calcium-dependent protein kinases: hubs in plant stress signaling and development. PLANT PHYSIOLOGY 163(2):523–530. doi: 10.1104/pp.113.222539

Seeraj R, Sinclair TR (2002) Osmolyte accumulation: can it really help increase crop yield under drought conditions? Plant, Cell and Environment 25:333–341

Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W (2016) Learning from Co-expression Networks. Possibilities and Challenges. Front. Plant Sci. 7(444). doi: 10.3389/fpls.2016.00444

Sham A, Moustafa K, Al-Ameri S, Al-Azzawi A, Iratni R, AbuQamar S (2015) Identification of Arabidopsis candidate genes in response to biotic and abiotic stresses using comparative microarrays. PLoS ONE 10(5):e0125666. doi: 10.1371/journal.pone.0125666

Shamloo-Dashtpagerdi R, Razi H, Ebrahimie E (2015) Mining expressed sequence tags of rapeseed (Brassica napus L.) to predict the drought responsive regulatory network. Physiology and molecular biology of plants : an international journal of functional plant biology 21(3):329–340. doi: 10.1007/s12298-015-0311-5

Sharma S, Verslues PE (2010) Mechanisms independent of abscisic acid (ABA) or proline feedback have a predominant role in transcriptional regulation of proline metabolism during low water potential and stress recovery. Plant, cell & environment 33(11):1838–1851. doi: 10.1111/j.1365-3040.2010.02188.x

Shinozaki K, Yamaguchi-Shinozaki K (2007) Gene networks involved in drought stress response and tolerance. Journal of Experimental Botany 58(2):221–227. doi: 10.1093/jxb/erl164

Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. Current Opinion in Plant Biology 6(5):410–417. doi: 10.1016/S1369-5266(03)00092-X

Smita S, Amit K, Dev MP, Viswanathan C, Sunil A, Kailash CB (2013) Identification of conserved drought stress responsive gene-network across tissues and developmental stages in rice. Bioinformation 9:72–78

Snowdon RJ, Iniguez L, Federico L (2012) Potential to improve oilseed rape and canola breeding in the genomics era. Plant Breeding 131(3):351–360. doi: 10.1111/j.1439-0523.2012.01976.x

Stone SL (2014) The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling. Frontiers in plant science 5:135. doi: 10.3389/fpls.2014.00135

Sucaet Y, Wang Y, Li J, Wurtele ES (2012) MetNet Online: a novel integrated resource for plant systems biology. BMC Bioinformatics 13:267. doi: 10.1186/1471-2105-13-267

Sun S, Wang H, Yu H, Zhong C, Zhang X, Peng J, Wang X (2013) GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation. Journal of Experimental Botany 64(6):1637–1647. doi: 10.1093/jxb/ert021

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE 6(7):e21800. doi: 10.1371/journal.pone.0021800

Teige M, Scheikl E, Eulgem T, Doczi R, Ichimura K, Shinozaki K, Dangl JL, Hirt H (2004) The MKK2 pathway mediates cold and salt stress signaling in Arabidopsis. Molecular cell 15(1):141–152. doi: 10.1016/j.molcel.2004.06.023

The SAM format specification workgroup (2011) The SAM Format Specification

Toubiana D, Fernie AR, Nikoloski Z, Fait A (2013) Network analysis: tackling complex data to study plant metabolism. Trends in biotechnology 31(1):29–36. doi: 10.1016/j.tibtech.2012.10.011

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7(3):562–578. doi: 10.1038/nprot.2012.016

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals

unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology 28(5):511–515. doi: 10.1038/nbt.1621

Trick M, Long Y, Meng J, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid Brassica napus using Solexa transcriptome sequencing. Plant Biotechnology Journal 7(4):334–346. doi: 10.1111/j.1467-7652.2008.00396.x

U N (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Japan. J. Bot 7:389–452

Verbruggen N, Hermans C (2008) Proline accumulation in plants: a review. Amino acids 35(4):753–759. doi: 10.1007/s00726-008-0061-6

Verslues PE, Bray EA (2006) Role of abscisic acid (ABA) and Arabidopsis thaliana ABA-insensitive loci in low water potential-induced ABA and proline accumulation. Journal of Experimental Botany 57(1):201–212. doi: 10.1093/jxb/erj026

Verslues PE, Sharma S (2010) Proline metabolism and its implications for plant-environment interaction. The Arabidopsis book / American Society of Plant Biologists 8:e0140. doi: 10.1199/tab.0140

Vidal EA, Moyano TC, Riveras E, Contreras-Lopez O, Gutierrez RA (2013) Systems approaches map regulatory networks downstream of the auxin receptor AFB3 in the nitrate response of *Arabidopsis thaliana* roots. PNAS 110(31):12840–12845. doi: 10.1073/pnas.1310937110

Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. Current opinion in biotechnology 16(2):123–132. doi: 10.1016/j.copbio.2005.02.001

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH (2011) Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474(7351):380–384. doi: 10.1038/nature10110

Wagner M, Ducournau S, Dürr C, Léchappé J (2011) Computer vision for monitoring seed germination from dry state to young seedlings. Seed Test. Int 142:49–51

Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D (2016) Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. Nature communications 7:11708. doi: 10.1038/ncomms11708

Wang H, Zhu Y, Fujioka S, Asami T, Li J (2010) Regulation of Arabidopsis Brassinosteroid Signaling by Atypical Basic Helix-Loop-Helix Proteins. The Plant Cell 21(12):3781–3791. doi: 10.1105/tpc.109.072504

Wang H, Zhu Y, Fujioka S, Asami T, Li J, Li J (2009) Regulation of Arabidopsis brassinosteroid signaling by atypical basic helix-loop-helix proteins. The Plant Cell 21(12):3781–3791. doi: 10.1105/tpc.109.072504

Wang J, Ma X-M, Kojima M, Sakakibara H, Hou B-K (2013) Glucosyltransferase UGT76C1 finely modulates cytokinin responses via cytokinin N-glucosylation in *Arabidopsis thaliana*. Plant Physiol. Biochem. 65:9–16. doi: 10.1016/j.plaphy.2013.01.012

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park B-S, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Di Shen, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin, Isobel A P, Batley J, Kim J-S, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon S-J, Choi S-R, Lee T-H, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y, Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43(10):1035–1039. doi: 10.1038/ng.919

Weston DJ, Gunter LE, Rogers A, Wullschleger SD (2008) Connecting Genes, Coexpression Modules, and Molecular Signatures to Environmental Stress Phenotypes in Plants. BMC Syst Biol 2(1):16. doi: 10.1186/1752-0509-2-16

Wolters H, Jurgens G (2009) Survival of the flexible: hormonal growth control and adaptation in plant development. Nature reviews. Genetics 10(5):305–317. doi: 10.1038/nrg2558

Xia X-J, Zhou Y-H, Shi K, Zhou J, Foyer CH, Yu J-Q (2015) Interplay between reactive oxygen species and hormones in the control of plant development and stress tolerance. Journal of Experimental Botany 66(10):2839–2856. doi: 10.1093/jxb/erv089

Xiong L, Schumaker KS, Zhu J-K (2002) Cell Signaling during Cold, Drought, and Salt Stress. The Plant Cell 14(S1):S165–S183

XIONG L, ZHU J-K (2002) Molecular and genetic aspects of plant responses to osmotic stress. Plant, Cell and Environment(25):131–139

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q (2006) The MYB transcription factor superfamily of

Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant molecular biology 60(1):107–124. doi: 10.1007/s11103-005-2910-y

Yi K, Menand B, Bell E, Dolan L (2010) A basic helix-loop-helix transcription factor controls cell growth and size in root hairs. Nat Genet 42(3):264–267. doi: 10.1038/ng.529

Yoshida T, Mogami J, Yamaguchi-Shinozaki K (2014) ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. Current Opinion in Plant Biology 21:133–139. doi: 10.1016/j.pbi.2014.07.009

Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart CN (2008) Plant systems biology comes of age. Trends in Plant Science 13(4):165–171. doi: 10.1016/j.tplants.2008.02.003

Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann F, Eichhorn A, Polley A, Jaenecke C, Ganal MW, Roder MS (2015) Analysis of main effect QTL for thousand grain weight in European winter wheat (Triticum aestivum L.) by genome-wide association mapping. Frontiers in plant science 6:644. doi: 10.3389/fpls.2015.00644

Zhang B, Horvath S (2005) A General Framework for Weighted Gene Co-Expression Network Analysis. Statistical Applications in Genetics and Molecular Biology 4(1)

Zhang J, Mason AS, Wu J, Liu S, Zhang X, Luo T, Redden R, Batley J, Hu L, Yan G (2015) Identification of Putative Candidate Genes for Water Stress Tolerance in Canola (Brassica napus). Front. Plant Sci. 6(221):403. doi: 10.3389/fpls.2015.01058

Zhang T, Jiang M, Chen L, Niu B, Cai Y (2013) Prediction of Gene Phenotypes Based on GO and KEGG Pathway Enrichment Scores. BioMed Research International 2013(3):1–7. doi: 10.1155/2013/870795

Zhiponova MK, Vanhoutte I, Boudolf V, Betti C, Dhondt S, Coppens F, Mylle E, Maes S, González-García M-P, Caño-Delgado AI, Inzé D, Beemster, Gerrit T S, Veylder L de, Russinova E (2013) Brassinosteroid production and signaling differentially control cell division and expansion in the leaf. The New phytologist 197(2):490–502. doi: 10.1111/nph.12036

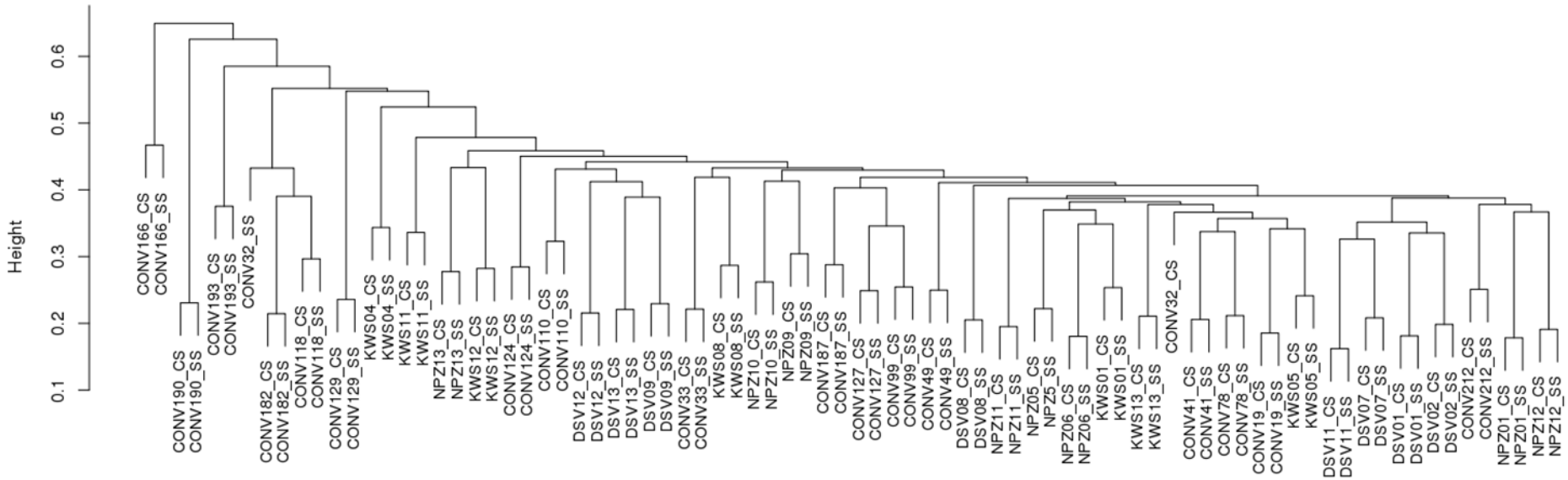
Zhou L, Wang J-K, Yi Q, Wang Y-Z, Zhu Y-G, Zhang Z-H (2007) Quantitative trait loci for seedling vigor in rice under field conditions. Field Crops Research 100(2-3):294–301. doi: 10.1016/j.fcr.2006.08.003

Zhu J, Zhang B, Smith EN, Drees B, Brem RB, Kruglyak L, Bumgarner RE, Schadt EE (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nature genetics 40(7):854–861. doi: 10.1038/ng.167

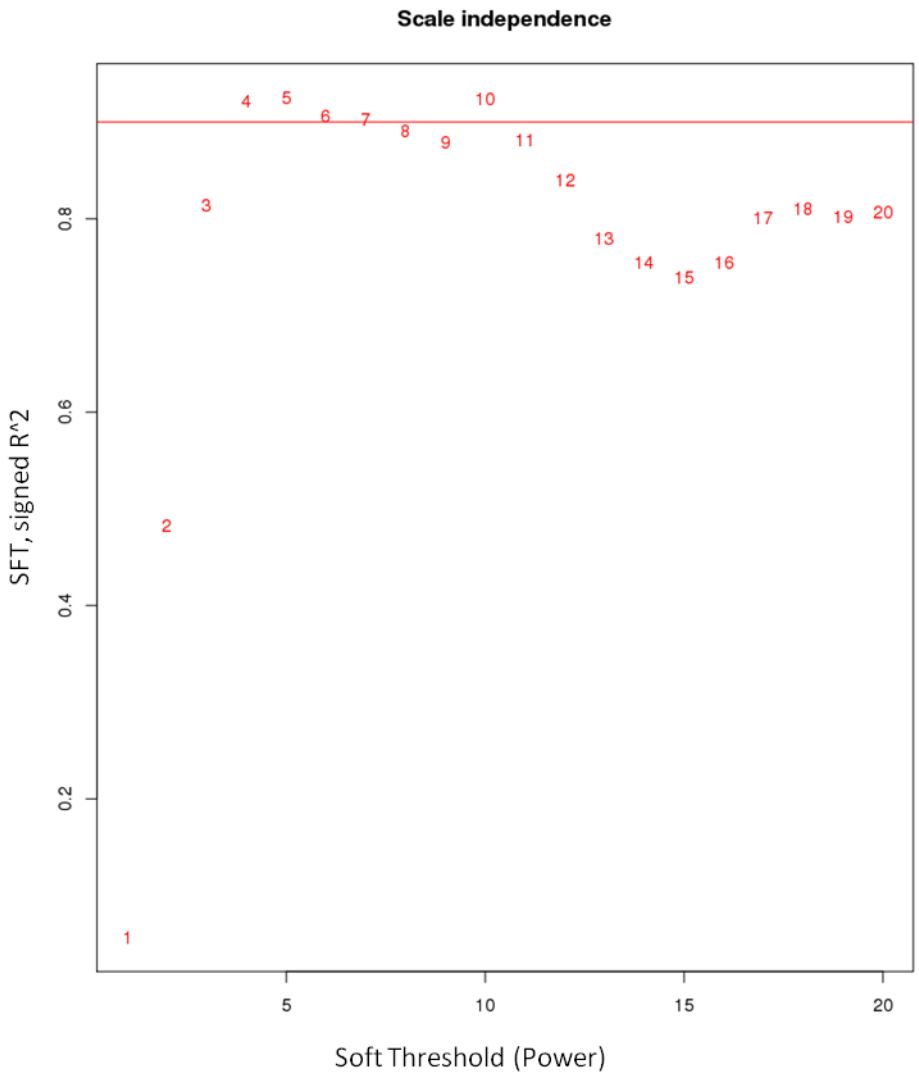
Appendix

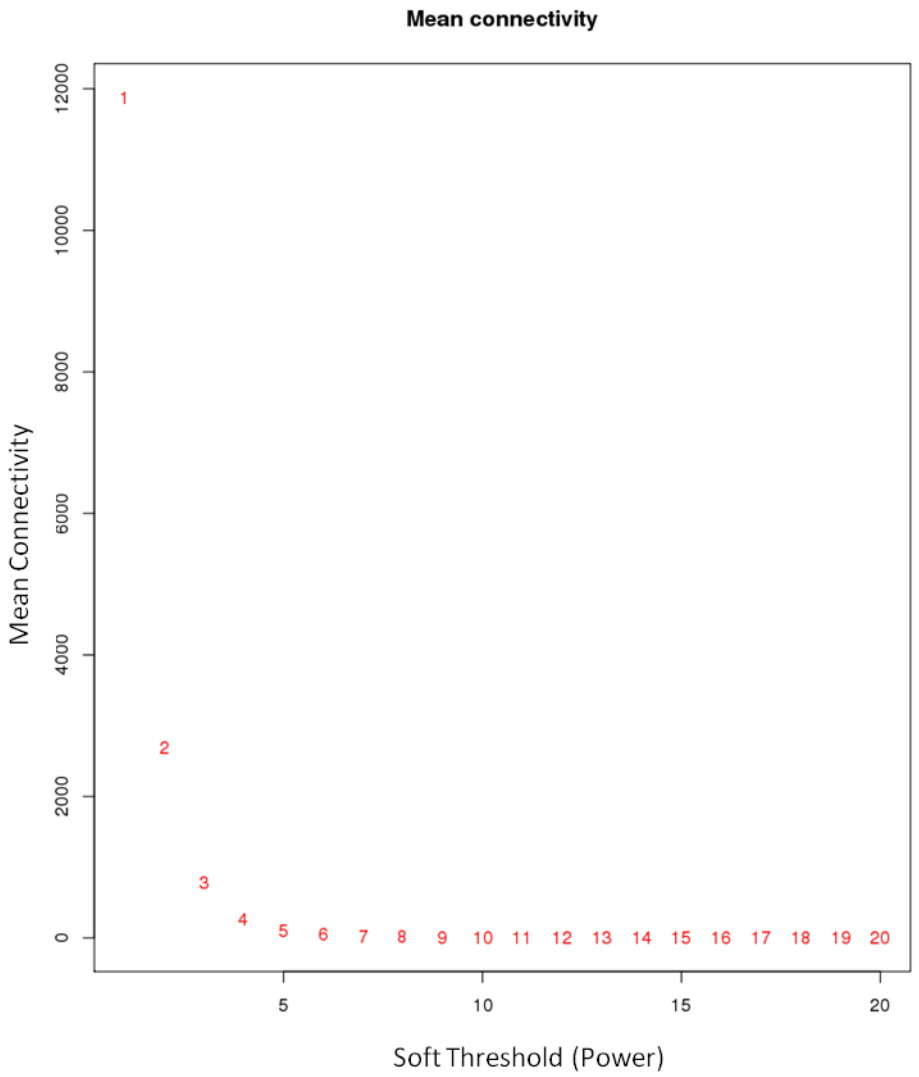
Appendix 1. List of the 42 genotypes used in the study for the RNAseq. 18 diversity set genotypes (also included in the GWAS, Hatzig et al. 2015) and 24 breeder‘s elite lines. *included in the field trial, metabolite profiling and biomass measurement. NA=not applicable

Diversity set		Breeder‘s elite lines
Genotype code	Genotype name	Genotype code
CONV110*	Zenith	DSV01*
CONV118*	Coriander	DSV02*
CONV124*	Gross-Luesewitzer	DSV07*
CONV127	Janpol	DSV08*
CONV129*	Jupiter	DSV09*
CONV166*	Aphid Resistant Rape	DSV11*
CONV182*	Abukuma natane	DSV12*
CONV187	Wild accession	DSV13*
CONV19*	Savannah	KWS01*
CONV190	Alesi	KWS04*
CONV193	Arvor	KWS05*
CONV212	Hektor	KWS08*
CONV32*	Lipton	KWS10*
CONV33*	Lisek	KWS11*
CONV41*	Laser	KWS12*
CONV49*	SWGospel	KWS13*
CONV78*	Pollen	NPZ01*
CONV99	Sollux	NPZ05*
		NPZ06*
		NPZ09*
		NPZ10*
		NPZ11*
		NPZ12*
		NPZ13*



Appendix 2. Hierarchical clustering of shoot control samples without treatment (CS) and shoot osmotic stressed samples (SS) with treatment as binary trait (control sample=0, stress sample=1). The treatments clustered closely together, and genotypic differences were greater than treatment differences. The winter fodder genotypes (CONV166 and CONV190) clustered with height greater than six.


Scale independence
SFT, signed R^2
0.2
0.4
0.6
0.8
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
5
10
15
20
Soft Threshold (Power)


Mean connectivity
Mean Connectivity
0
2000
4000
6000
8000
10000
12000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
5
10
15
20
Soft Threshold (Power)

Appendix 3. In case of the osmotic stress responsive gene co-expression network, The scale-free topology fit index reached above 0.8 for reasonable powers (less than 7) and the mean connectivity remained low (below 500). It means that the data does not exhibit a strong driver that makes a subset of the samples globally different from the rest.

Declaration

I declare: this dissertation submitted is a work of my own, written without any illegitimate help by any third party and only with materials indicated in the dissertation. I have indicated in the text where I have used texts from already published sources, either word for word or in substance, and where I have made statements based on oral information given to me. At any time during the investigations carried out by me and described in the dissertation, I followed the principles of good scientific practice as defined in the "Statutes of the Justus Liebig University Giessen for the Safeguarding of Good Scientific Practice".

Giessen, 2017 KIDIST BOGALE KIBRET

Signature: .....

Acknowledgements

First of all I would like to express my deepest gratitude to my first supervisor Prof. Dr. Rod Snowdon for giving me the chance to work on the project and his continuous guidance throughout the study. I am thankful for Dr. Birgit Samans, Dr. Sarah Hatzig and Dr. Christian Obermeier for their guidance at various stages of my study. I am also grateful for the support during the multiple *in-vitro* cultivations by Dr. Sarah Hatzig, Christine Nolte, Lukas Fehse, Nelly Weis, Swetlana Renner, Petra Degen, Annette Plank and Stavros Tzigos. The CONVIGOUR project partners from NPZ Innovation GmbH, Holtsee, Germany, RAGT Semences, Rodez, France, KWS SAAT SE, Einbeck, Germany, DSV Deutsche Saatveredelung AG (Lippstadt, Germany), GEVES (Groupe d'Etude et de Contrôle des Variétés et des Semences, Beaucouzé, France) and Institute for Genetics, Environment and Plant Protection, INRA, Le Rheu, France are acknowledged for providing the seeds and for conducting the field trials. The automated *in-vitro* germination profiling was done at GEVES (Groupe d'Etude et de Contrôle des Variétés et des Semences, Beaucouzé, France). Metabolite profiling was done by Dr. Laurie Legoahec and Dr. Nathalie Nesi from Institute for Genetics, Environment and Plant Protection, INRA, Le Rheu, France.