Development and application of a molecular marker resource for improving drought stress tolerance in rye (*Secale cereale* L.)

Dissertation

zur Erlangung des Doktorgrades

der Agrarwissenschaften (Dr. agr.)

der

Naturwissenschaftlichen Fakultät III

Agrar- und Ernährungswissenschaften,

Geowissenschaften und Informatik

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Herrn Matthias Enders

Geb. am 06.10.1983 in Halberstadt

Verteidigt am: 18.10.2021

Gutachter:

Prof. Dr. Frank Ordon

Prof. Dr. Ralf Uptmoor

Table of contents

Abbreviations

6PGDH	– 6-phosphogluconate dehydrogenase (gene family proteins)
ABA	– abscisic acid
AFLP	– amplified fragment-length polymorphism
bp	– base pair
BSA	– bulked segregant analysis
cDNA	– complementary deoxyribonucleic acid
cM	– centimorgan
CMS	– cytoplasmatic male sterility
CNN	– convolutional neural networks
COS	– Conserved Ortholog Set markers
DArT	– diversity array technology
DNA	– deoxyribonucleic acid
EMI	– electromagnetic inductance
ERD	– early-responsive to dehydration (stress)
ERT	– Electrical Resistance Tomography
EST	– expressed sequence tag
g	– gram(s)
gb	– giga bases
GBS	– genotyping by sequencing
GCA	– general combining ability
GO	– gene ontology
GPR	– Ground penetrating radar
GS	– Genomic Selection
GWAS	– genome wide association study
GxE	– genotype by environment (interaction)
ha	– hectare
HSP	– heat shock protein
LEA	– late embryogenesis abundant (proteins)
LD	– linkage disequilibrium
LOD	– logarithm of the odds
m^2	– square meter
MAS	– marker assisted selection
mg	– milligram(s)
µg	– microgram(s)
MAGIC	– multiparent advanced generation intercross (population)
MQM	– multiple-QTL model
N50	– the length of the shortest contig at 50% of the total assembly length
NAM	– nested association mapping (population)
NGS	– next generation sequencing
PCoA	– principal coordinate analysis
PCR	– polymerase chain reaction
PPP	– pentose phosphate pathway
QTL	– quantitative trait loci
RAPD	– randomly amplified polymorphic deoxyribonucleic acid
RFLP	– restriction fragment length polymorphism
RIN	– RNA integrity index

RNA – ribonucleic acid
ROS – reactive oxygen species
RSQ – Rye SNP Quedlinburg
SCA – special combining ability
smHSP – small Heat Shock Proteins
SNP – single nucleotide polymorphism
SSR – simple sequence repeats
t – tones
tpm – tags per million
vit – vitamin
WUE – Water use efficiency
ZEP – zeaxanthin epoxidase

1. Introduction

1.1 Rye

Rye (*Secale cereale* L.) is an allogamous, annual, cereal crop species in the tribe *Triticeae* within the *Poaceae* family. Similar to closely related important crop species like barley (*Hordeum vulgare* L.) and wheat (*Triticum aestivum* L.) the centre of diversity is located in the Fertile Crescent (Zohary et al., 2012).

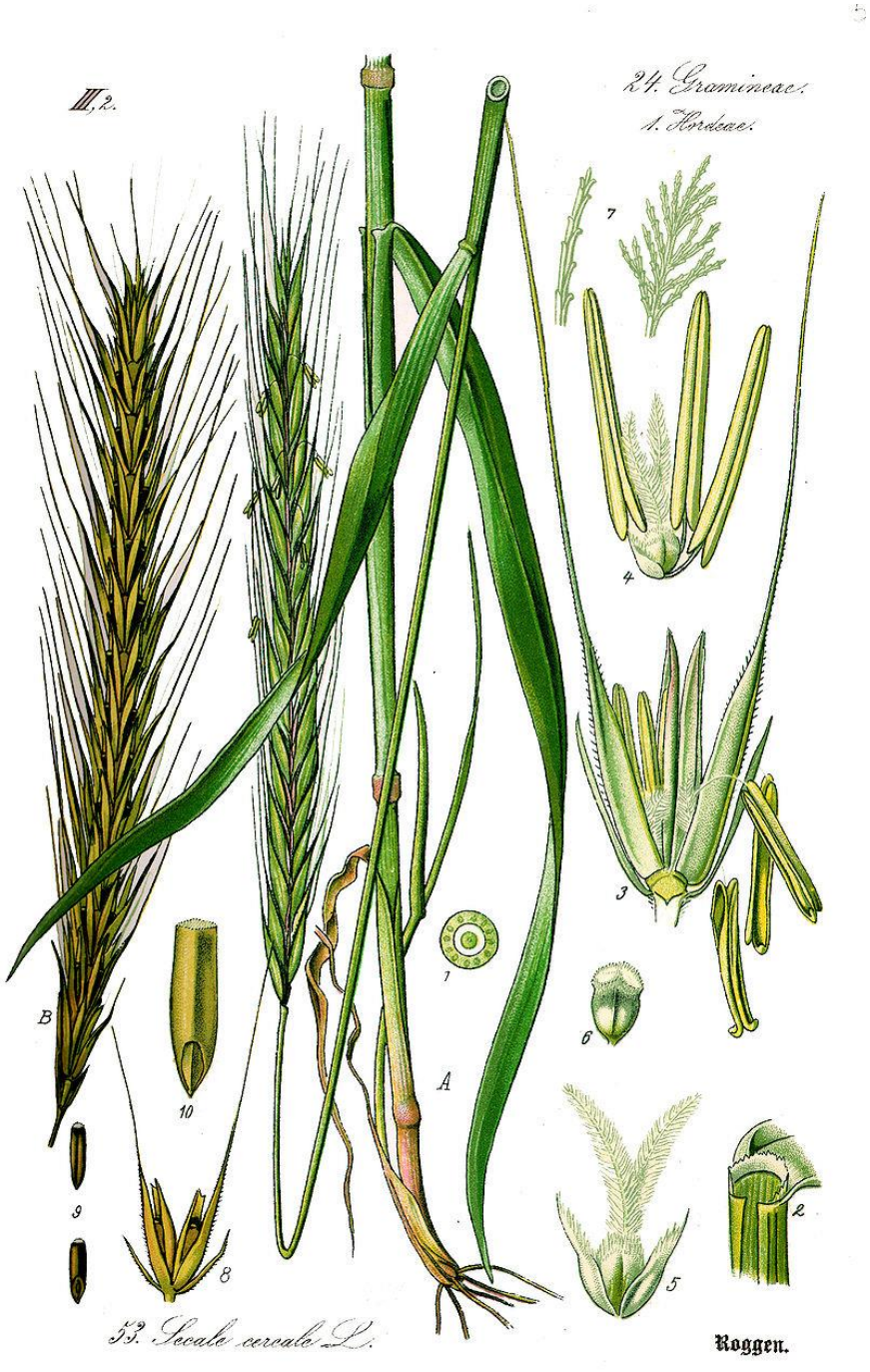

Figure 1. Drawing of Rye - *Secale cereale* L. (Thomé, 1885).

In contrast to direct domestication, rye co-evolved as a weed in barley and wheat fields, which is referred to as secondary domestication (Preece et al., 2017) or Vavilovian mimicry (McElroy, 2014). Domesticated rye (*Secale cereale* L.) together with the wild or weedy species *Secale africanum* Stapf, *Secale anatolicum* Boiss., *Secale ciliatoglume* Boiss., *Secale iranicum* Kobyl., *Secale montaneum* Guss., *Secale segetale* Zhuk., *Secale sylvestre* Host. and *Secale vavilovii* Grossh. forms the genus *Secale* (Govaerts, 2009).

Though the economic importance and current production of 15.2 million tones worldwide in 2017 (FAO, 2019) is not on the same level as for wheat or barley, rye exhibits a range of positive and favourable characteristics. In low input environments (soil quality, fertilization) rye is relatively high yielding (Madej, 1996), and it is known for an outstanding tolerance to a range of abiotic stresses like aluminium (Silva et al., 2012), salt and drought (Geiger and Miedaner, 2009), as well as low temperature (Fowler and Limin, 1987). In particular, rye is known for its outstanding drought tolerance, which highly influences crop productivity (Toker et al., 2007). Some of the genes underlying this favourable characteristics are transferred to other species like the man-made cereal triticale (x *Triticosecale*) or to wheat using translocations, e.g. of the short arm of rye chromosome 1 (1RS) (Hoffmann, 2008; Ko et al., 2002).

Rye is grown mainly as a winter crop in northern and eastern Europe. Germany, Poland, and the Russian Federation are the most important rye producers. The highest yields are achieved in Germany and Denmark (FAO, 2019). For details concerning rye production cf. Table 1, Table 2 and Table 3.

Table 1. Yield (in t/ha) of rye for important rye growing countries from 2010 to 2017 (FAO, 2019).

Year	Germany	Poland	Russia	China	Denmark	Belarus	Ukraine
2010	4.6	2.6	1.1	3.1	4.8	2.1	1.6
2011	4.1	2.3	1.9	3.7	5.1	2.4	2.0
2012	5.4	2.7	1.4	3.4	5.9	2.7	2.2
2013	5.9	2.8	1.8	3.4	6.0	2.0	2.3
2014	6.1	3.1	1.7	3.0	6.3	2.7	2.6
2015	5.6	2.7	1.6	3.2	6.3	3.0	2.6
2016	5.5	2.8	2.0	3.6	5.8	2.7	2.7
2017	5.0	3.0	2.1	3.7	6.5	2.6	2.9

Table 2. Total production (in thousand tones) of important rye growing countries from 2010 to 2017 (FAO, 2019).

Year	Germany	Poland	Russia	China	Denmark	Belarus	Ukraine
2010	2900	2851	1635	570	254	734	464
2011	2520	2600	2970	750	294	800	578
2012	3878	2888	2131	650	384	1082	676
2013	4689	3359	3359	620	526	648	637
2014	3854	2792	3280	520	677	867	478
2015	3487	2013	2086	1311	771	752	391
2016	3173	2199	2541	930	577	650	391
2017	2737	2673	2547	1332	723	669	507

Table 3. Total area harvested (in thousand ha) of important rye producing countries from 2010 to 2017 (FAO, 2019).

Year	Germany	Poland	Russia	China	Denmark	Belarus	Ukraine
2010	627	1059	1367	180	52	342	279
2011	613	1085	1524	200	57	326	279
2012	708	1042	1422	190	64	392	297
2013	784	1172	1774	180	87	323	279
2014	629	886	1858	170	106	319	185
2015	616	725	1250	409	121	250	150
2016	570	775	1249	255	99	241	143
2017	537	873	1174	352	111	255	171

Rye is used mainly for human nutrition (bread), animal feed and as a source of biomass for renewable energy production (Miedaner et al., 2012b). Studies indicate that rye has several positive effects for humans, e.g. decreased risk of diabetes, cardiovascular disease and certain cancers as well as an improved regulation of blood glucose level, especially if whole grain products are used (Jacobs et al., 1998a, 1998b; Liu et al., 2000; Liukkonen et al., 2003; Pereira et al., 2002). On the other hand, some of these components, especially pentosanes, which confer the positive health effects for humans are detrimental in animal feed, e.g. in pigs (Patience et al., 1992) or poultry (Lázaro et al., 2004). A selection of rye ingredients can be obtained in Table 4.

Table 4. Selected ingredients of rye (*Secale cereale* L.) (Souci et al., 2003). All data is the average content in relation to 100g eatable rye grain.

overview		minerals		vitamins		amino acids	
Water	13.7 g	**Sodium**	4.0 mg	**Vit. B1**	360 µg	**Tryptophan**	110 mg
Protein	8.8 g	**Potassium**	510.0 mg	**Vit. B2**	170 µg	**Threonine**	360 mg
Fat	1.7 g	**Magnesium**	90.0 mg	**Vit. B3**	1800 µg	**Isoleucine**	390 mg
Carbohydrates	60.7 g	**Calcium**	35.0 mg	**Vit. B5**	1500 µg	**Leucine**	370 mg
Dietary fiber	13.2 g	**Manganese**	2.9 mg	**Vit. B6**	235 µg	**Lysine**	400 mg
Minerals	1.9 g	**Iron**	2.8 mg	**Folic Acid**	145 µg	**Methionine**	140 mg
		Copper	0.4 mg	**Vit. E**	2000 µg	**Valine**	530 mg
		Zinc	2.9 mg			**Phenylalanine**	470 mg
		Phosphorus	335.0 mg			**Tyrosine**	230 mg
						Arginine	490 mg
						Histidine	190 mg

The genome of rye (R) is diploid with 2n = 2x = 14 chromosomes and of rather large size with 8.1 giga bases (gb) (Dolezel et al., 2007), contains large amounts of repetitive sequences (Flavell et al., 1974) and shows a high degree of synteny with the genomes of barley and wheat (Bauer et al., 2017).

As a cross-pollinating (wind) and self-incompatible species, rye cultivars are bred by means of population breeding (open-pollinated varieties, synthetic varieties) (Schnell, 1982) and since 1985 by hybrid breeding (Geiger and Miedaner, 1999). The later breeding scheme aims to improve crop productivity by exploiting the heterosis in terms of general combining ability (GCA) and special combining ability (SCA) using crosses between genetically distant heterotic

groups. Hybrid rye breeding was facilitated and promoted by three important factors, (i) the development and definition of “Carsten” and “Petkus” as two distinct gene pools followed by the assignment as seed parent and pollinator pool for line development, respectively (Geiger and Schnell, 1970; Hepting, 1978), (ii) the presence of dominant self-fertility genes to overcome the rye self-incompatibility (Wolski, 1970; Wricke, 1969) and (iii) the availability of a cytoplasmatic male sterility (CMS) system (Geiger and Schnell, 1970). For CMS two major groups, namely the V-type (Vavilov) and the P-type (Pampa) are known (Lapiński and Stojalowski, 2003). Most commercial rye hybrid varieties are based on the P-type cytoplasm (Geiger and Miedaner, 1999). The traits targeted by breeding mainly depend on the potential use of new cultivars and are partially contrasting. Grain yield accompanied by a high protein content is by far the most important trait for livestock feed and human consumption (Miedaner et al., 2012a). On the other hand, a maximized dry matter yield is targeted for biomass production, combined with an optimized methane output per unit dry matter for biogas production (Roux et al., 2010). Furthermore, the previously discussed abiotic stress tolerance of rye (Gallego and Benito, 1997; Hubner et al., 2013), enables production on light soils with marginal fertility, not suited for other crops like maize or wheat (Hoffmann, 2008).

1.2 Molecular tools available in rye

Rye has a large and highly repetitive genome and is less economically important than the closely related species barley and wheat. Therefore, also the molecular information generated and available for rye is relatively sparse in comparison to barley and wheat. Nevertheless, a range of different types of molecular techniques and molecular markers are available ranging from allozyme to single nucleotide polymorphism (SNP) as well as sequence information (Bauer et al., 2017; Bolibok et al., 2006; Bolibok-Bragoszewska et al., 2009; Chikmawati et al., 2005; Martis et al., 2013; Matos et al., 2001; Saal and Wricke, 1999). A wide range of those genetic features are also integrated in genetic maps (Korzun et al., 2001; Loarce et al., 1996; Milczarski et al., 2011, 2007).

The first available molecular markers were allozyme assays followed by the assignment of these to specific chromosomes (Adam et al., 1987; Salinas and Benito, 1985, 1983; Wehling et al., 1985). These were followed by restriction fragment length polymorphisms (Devos et al., 1992; Plaschke et al., 1993) and after the availability of robust polymerase chain reaction (PCR) protocols, by randomly amplified polymorphic DNA (RAPD) (Loarce et al., 1996), or amplified

fragment-length polymorphism (ALFP) markers (Saal and Wricke, 2002). These markers can be used without previous knowledge on the genome sequence, which greatly reduces the time and cost of marker development and some of them, e.g. AFLPs, allow analysing a high number of loci at the same time. The set of available marker systems was further extended with simple sequence repeat (SSR) markers, which are still used extensively in many plant species and were also employed in rye (Feuj et al., 2018; Li et al., 2018; Targońska et al., 2016). Respective markers were used for different purposes, e.g. genetic diversity and phylogenetic analysis (Shang et al., 2006), construction of genetic maps (Schlegel et al., 1998), tagging the 1RS rye-wheat translocation (Kofler et al., 2008) or detection of quantitative trait loci (QTL) (Börner et al., 1999) to name just a few. Favourable characteristics, like high reproducibility and robustness in data generation, as well as the multi-allelic nature of SSR markers conferring superior information content per data-point (Saal and Wricke, 1999), made SSR markers a very valuable tool. Also, the high abundance, covering the whole genome leads to an extensive use of SSR markers until very recently. Nevertheless, SSR assays exhibits often a relatively low rate of multiplexing compared to single nucleotide polymorphisms (SNPs) (Rafalski, 2002) resulting in a potentially higher price per data-point and a potentially higher amount of time needed for data generation. Furthermore, valid sequence information is needed in order to create SSR markers (Robinson et al., 2004).

The advent of so called Next Generation Sequencing (NGS) techniques (Schuster, 2007), which is also termed massive parallel sequencing or second generation sequencing, allows the generation of sequence information for large genomes in a short period of time with previously unmatched low costs. Despite challenges, like assembling short NGS sequence reads to large genomic DNA information called contigs and scaffolds, the chromosomal assignment or the difficulties to cover highly repetitive genomic regions, NGS proved to be an excellent tool (Michael and VanBuren, 2015). A detailed and frequently updated list of all sequenced plant species can be obtained at Usadel, B. (2019). NGS was employed to detect SNPs, which are the most simple and most common class of genetic variation (Varshney et al., 2007a). SNPs were detected and used in rye already in the beginning of this century (Varshney et al., 2007a), but the massive number of detected SNPs by sequencing provides information to generate two large scale, highly multiplexed hybridization based SNP arrays (Bauer et al., 2017; Haseneyer et al., 2011), which are valuable resources for both rye research and rye breeding. To avoid high sequencing costs and the challenges of assembling the highly

repetitive rye genome, the study presented in this thesis, as well as the study that generated the first SNP array (Haseneyer et al., 2011) focused on the sequencing of expressed genes via sequencing expressed sequence tags (EST). This transcriptome sequencing approach reduces the number of nucleotides, which needs to be sequenced and assembled down to only the proportion of expressed genes after transcription and splicing. It also allows inducing and quantifying the expression level of transcripts using defined and controlled treatments and sampling expressed sequences from defined tissues. On the base idea of ribonucleic acid (RNA) sequencing a set of different technologies and methodologies arise, e.g. SuperSAGE (Matsumura et al., 2005), DeepSuperSAGE (Matsumura et al., 2012) which was used in this study or MACE (Zawada et al., 2014). Although the EST resources available for rye were increased by a transcriptome study (Haseneyer et al., 2011) and a draft genome sequence was published (Bauer et al., 2017), the knowledge on transcripts or molecular markers directly associated with drought stress responses in rye is still scarce.

Beside the possibility to directly sequence RNA, also so called exome capture methods are available (Chilamakuri et al., 2014), which omit sequencing of intergenic or repetitive DNA sequences using filtering methods. As a downside of this later methods, a quantification of gene expression is not directly possible. Furthermore, polymorphisms like SNPs are detected by sequencing multiple genotypes. A set of methods also uses the basic idea of reducing the genome representation and were also used in rye, like Diversity Array Technology (DArT) markers (Bolibok-Bragoszewska et al., 2009) and Genotyping-by-Sequencing (GBS) (Schreiber et al., 2018). As these methods in general can be applied and used without the availability of a high-quality reference genome, and respective annotations, the use of the resulting data is limited to some extend as e.g. physical mapping cannot be conducted directly. To address some of these challenges so called 'Genome Zippers' (Mayer et al., 2009; Spannagl et al., 2013) may be employed. As toolboxes, which support the reconstruction of large and complex cereal genomes, 'Genome Zippers' integrate the information on genes, genetic maps, sequence data and synteny in the *Triticeae* into a virtual gene order of barley (Mayer et al., 2011), rye (Martis et al., 2013) and wheat (Hernandez et al., 2012). These resources promote a wide range of applications, e.g. fine mapping and marker saturation (Silvar et al., 2013) or phylogenetic and evolutionary analysis (Martis et al., 2013), especially in species and crops with no complete high quality standard reference sequence like rye. This resource makes use of the high degree of synteny between related species like barley, wheat, rice, brachypodium, sorghum, and rye

(Martis et al., 2013) facilitating the alignment of genes into a virtual linear order, which may be seen as an early kind of pan-genomic approach (Tettelin et al., 2005; Vernikos et al., 2015) covering several species within the *Poaceae*. For some of the previously listed species high quality reference genome sequences (Appels et al., 2018; Mascher et al., 2017; Ramírez-González et al., 2018), gene annotations and additional data are available, today. This information may be used in studies on species like rye with still limited genomic information. As sequencing techniques evolve and the overall price for sequencing projects is dropping, the generation of a gold-standard reference sequence for rye comes into reach (Bauer et al., 2017). Currently available and new technologies like long-range third generation sequencing (Rusk, 2009) and optical mapping (Burton et al., 2013) will be used in rye most probably, as it was already done for wheat, to generate a high quality reference genome (Appels et al., 2018). On a long-term perspective, a pan-genome approach (Tettelin et al., 2005) covering the sequence diversity within the rye primary gene pool may be envisioned (Monat et al., 2019; Montenegro et al., 2017).

1.3 Detection and mapping of quantitative trait loci and underlying genes

In plant science, as well as in medicine and biology in general, a very common task is to identify genes, which influence a target trait. This knowledge may be used to gain insight into gene functions, metabolic pathways or to develop molecular markers to predict characteristics of an organism without the need for phenotyping (Agarwal et al., 2008). Furthermore, it saves time and resources for breeding and allows e.g. risk assessment in medicine (Hedenfalk et al., 2001). In the era of 2nd and 3rd generation sequencing, a common way to identify single genes encoding a trait of interest is to conduct a bulked segregant analysis (BSA) based on massive sequence data which is derived from two pools of individuals, which contrast with respect to the trait of interest (Magwene et al., 2011). Nevertheless, this approach is highly depended on a high-quality annotation of the reference genome to reduce the number of potential candidate genes. This is common for model organisms which usually have smaller genomes (Wenger et al., 2010) and for economically important crop plants (Appels et al., 2018; Mascher et al., 2017; Monat et al., 2019) which genomes often are composed of tens of thousands of genes, large intergenic regions and repetitive sequences. This increases the costs for sequencing and potentially hampers the identification of candidate genes, particularly in species with multiple sub-genomes like wheat (Appels et al., 2018).

Many traits are influenced by multiple genes, potentially scattered over the whole genome and further controlled by transcription factors located in genomic positions distant to the encoding gene (Agarwal et al., 2008; Lander and Botstein, 1989). Other types of challenges to find gene-trait-associations comprise gene by gene interactions (Gauderman, 2002), presence/absence variation versus allelic variation (Springer et al., 2009), alternative splicing of transcripts (Barbazuk et al., 2008), epigenetic factors (Mirouze and Paszkowski, 2011) or duplication of genes resulting in copy number variations (Schiessl et al., 2017). All these types of variation can hamper the detection of potentially causal genes by means of a genome sequencing based bulked segregant analysis as well as other analysis approaches.

This gives reasons why even today methods like linkage mapping and association mapping are commonly used, as they detect genomic regions with causal influence on a target trait based on the co-occurrence of genotypic and phenotypic alterations within a set of genotypes, most often in combination with respect to positional information of the used molecular markers (Mackay et al., 2009). For linkage mapping of QTL, a special mapping population with a large progeny of defined parents is needed. While the classical approach uses two contrasting parents (Lander and Botstein, 1989) to generate a bi-parental population, a set of new multi-parent approaches like nested association mapping (NAM) (Yu et al., 2008) or multiparent advanced generation intercross (MAGIC) (Cavanagh et al., 2008) are available (Bandillo et al., 2013) in the meantime. Furthermore, the employed statistical methods were further developed to account for issues like co-founding factors (Zeng, 1994). As a downside of these methods, the mapping resolution in such a population is directly dependent on the number of recombination events, which is a function of population size and the number of meiosis. As both factors are increasing the resource demand for population development and data generation, QTL mapping suffers from low resolution, indirectly (Mackay et al., 2009). Furthermore, some QTL detected in one population, are reported to be not stable and reproducible in other populations, most presumably caused by genetic background effects (Kang et al., 2009). These disadvantages, together with the need to construct dedicated populations are surmounted by genome wide association studies (GWAS). GWAS methods were developed in the context of medical research, as the construction of defined, large populations is not applicable in humans. The available GWAS methods make use of the massive number of historic recombination events in a natural population, enabling a much higher mapping resolution, which, on the other hand needs a higher marker saturation of the

genome (Hirschhorn and Daly, 2005). The detection precision, especially by reducing the risk of false positive marker-associations was further increased by considering the population structure, also called stratification (Price et al., 2006). Although both, GWAS and bi-parental QTL studies are successfully used for traits controlled by a small to medium number of genes (oligogenic traits), the currently available statistical methods perform quite poor for highly polygenic traits with a high number of genes involved (Xu and Crouch, 2008). To further increase the prediction accuracy for highly polygenic traits, most important yield (Goddard and Hayes, 2007; Heffner et al., 2009) and to integrate all available molecular marker data even for loci with no or very small effects on the target trait so called genomic selection (GS) was implemented in both, animal- and plant breeding. Genomic selection methods proof their value in the context of breeding (Jannink et al., 2010; Shikha et al., 2017; Zhao et al., 2015), but are not directly usable to generate knowledge and insight for research on gene functions and metabolic pathways (Heffner et al., 2009).

A methodology often applied in agricultural research is RNA sequencing (Wang et al., 2009), which enables to capture both, the expressed portion of genes in terms of transcripts, as well as the expression level of those transcripts. This may be combined with the sampling of specific tissues or organs and with the exposure to a specific treatment to provoke specific expression profiles (Kakumanu et al., 2012). The combination of these techniques and datasets may be used to achieve several goals within a single experiment. This can comprise: genome complexity reduction of large crop genomes (Haseneyer et al., 2011); recording of gene expression profiles with respect to a specific genotype, tissue, organ, or treatment (Kakumanu et al., 2012); detection of polymorphisms in the gene–space (Kumar et al., 2012) as well as splice variants (Loraine et al., 2013) or antisense RNA (Lu et al., 2012). This combination of factors makes RNA sequencing experiments an extremely economical technique for a range of applications including agricultural research.

In the future, with the prospect of a largely increased knowledge on functions and interactions of genes, as well as genotype by environment interaction, mechanistic approaches like Pathway Breeding, seems to be in reach on the mid- to long term view, even in crops with complex genomes (Reguera et al., 2012). Holistic approaches like system biology, aiming to deterministically modelling the complete organism with all bio-chemical reactions are currently only applicable to extremely simple model microbes like *E. coli* within completely controlled environments like bioreactors (Chubukov et al., 2016). The usage of those models

seems far away from practical breeding applications. Another set of mathematical methods, so called machine learning algorithms (Goldberg, 1989), are available for the general task of modelling (Almeida, 2002). A subset of these methods, so called deep learning approaches; most predominantly deep convolution neural networks (CNN) (LeCun et al., 1998), are used massively in a wide range of applications like speech recognition and generation (Collobert and Weston, 2008), image analysis (Krizhevsky et al., 2012), data mining and predictive analytics in general. They provide a framework to automatically learn a model underlying a specific, real world process, by incrementally adjusting the response of artificial neuronal networks to given observations (Bishop, 2006). This process is often referred to as "training" which has similarities to the animal and human learning process to some extent. These methods mathematically enable the integration of gene-environment interactions, environmental data like sequential weather data and can model higher order gene-gene interactions (Ma et al., 2018; Uppu et al., 2016a, 2016b). This approach further increases the detection power and sensitivity of respective methods for gene identification in comparison with classical approaches which only take limited information into account (Liang et al., 2014). The downside of the currently available tools is the need for a massive amount of data for the training set, which can be hardly generated in one scientific research project. Also some algorithms are lacking interpretability, meaning they are so called "black-box" algorithms, which did not allow a direct extraction of the learned underlying model (Zeiler and Fergus, 2014). To sum up, a range of methods is available to detect QTL and underlying genes, as well as to predict phenotypes out of given genotypes, but the choice for a method greatly depends on the trait of interest, the genome structure of the target species and the resources available for the analysis (Jain et al., 2010; Moose and Mumm, 2008).

1.4 Drought stress response in plants on a molecular level

Drought stress and the negative impact on plant performance and yield is a very active field of research, as prognoses on the global climate change suggest an elevated temperature and alterations in precipitation. Therefore, the number and severity of drought stress periods will increase for many regions (Arneth et al., 2019; Lobell et al., 2011). The impact of drought on yield due to global warming has been simulated for major crops (Lobell et al., 2011; Osborne et al., 2013), predicting significant yield losses also in rye. This holds especially true since, even if the overall precipitation is not reduced, alterations in time, duration or intensity of

precipitation can cause drought stress (Serraj et al., 2005). To meet the goal of global food security for an increasing human population, improving drought stress tolerance is of prime importance in plant breeding. The plant reaction to surmount drought stress can be divided into escape, avoidance and tolerance strategies (Levitt, 1972; Turner, 1986), which are often combined in stress resistant plants (Ludlow, 1989). Furthermore, drought stress often comes along with other abiotic stresses, e.g. higher temperatures, oxidative stress caused by high amounts of reactive oxygen species (ROS), or intense light irradiation (Chaves et al., 2003; Gupta et al., 2013). These factors disguise and superimpose the plant response to drought stress on the molecular level making drought tolerance a complex trait which is controlled by a putatively large number of QTL with small individual effects and often cofounded by differences in plant phenology (Fleury et al., 2010). Several studies focused on drought stress tolerance using different strategies (Mir et al., 2012). These strategies ranged from physiological approaches (Sinclair, 2011) and genetic engineering (Yang et al., 2010) to genomic approaches (Cattivelli et al., 2008; Varshney et al., 2011) in different crop species, e.g. rice (Bernier et al., 2007), wheat (Fleury et al., 2010) or canola (Wan et al., 2009) resulting in a deeper knowledge on the reaction of plants to drought stress (Wehner et al., 2016). In many studies, QTL for drought stress tolerance-related traits in some major crop species were detected (Mir et al., 2012). Further investigations identified phytohormones, transcription factors, and respective genes as well as pathways involved in drought stress tolerance. Well known key factors of the reaction to biotic and abiotic stresses are for instance salicylic acid, jasmonic acid, ethylene, ROS and abscisic acid (ABA) (Fujita et al., 2006; Krasensky and Jonak, 2012). Some drought stress responsive genes are already used for practical breeding purposes, e.g. the “stay green” leaf senescence associated gene (Gous et al., 2013; Jiang et al., 2007). Most of these studies made use of molecular information available for model but also crop species (Hou et al., 2007; Mir et al., 2012; Raney et al., 2014; Vaseva et al., 2010). On the molecular level multiple mechanisms and molecules are known which are involved in the plant reaction to drought stress. As a first important group protein kinases and transcription factors are heavily involved in triggering the biosynthesis of osmolytes or proteins for macromolecular protection (Agarwal et al., 2006). For crops like barley (*Hordeum vulgare* L.) it was shown that the accumulation of sucrose and malate may serve as osmolytes as they are stronger accumulated as a response to drought stress (Templer et al., 2017). Other studies detect the accumulation of the amino acid proline as a reaction to abiotic stresses like drought in barley

(Wehner et al., 2015), wheat (Keyvan, 2010) and also crop species like chickpea (Mafakheri et al., 2010). Furthermore, polyamines are frequently found as osmolytes involved in the plant drought stress reaction complex, like reported for barley (Montilla-Bascón et al., 2017) in dependency of an altered nitric acid oxide level which promotes drought stress tolerance, as well as in rice (Capell et al., 2004) or in the model plant species *Arabidopsis thaliana* (L.) Heynh. (Yamaguchi et al., 2007). Fructans as soluble carbohydrates are also referred to as osmoprotectants in the context of drought and salinity stress (Kerepesi and Galiba, 2000; Murakeözy et al., 2003; Pilon-Smits et al., 1995), but there are also studies which found fructans to be converted to fructose and mobilized at a higher level as a response to drought stress, e.g. for wheat in the kernel filling stage (Wardlaw and Willenbrink, 2000). Furthermore, glycinebetaine is a very extensively studied compatible solute which was found in bacteria, plants and animals with regard to abiotic stresses in general and drought stress (Farooq et al., 2009). Besides the direct effect as osmoprotectant, glycinebetaine was found to be involved in the signalling transduction pathways in pigeon pea (Subbarao et al., 2000). An elevated synthesis of Trehalose as another compatible solute was also found throughout many kingdoms, i.e. bacteria, fungi, animals and lower plants as a response to stress (Wingler, 2002). Although genes for Trehalose biosynthesis were suspected to be absent from higher plants, newer studies found evidence for Trehalose synthesis in cotton (Kosmas et al., 2006), tobacco (Goddijn et al., 1997) and *Arabidopsis thaliana* (L.) Heynh. (Wingler, 2002).

Aquaporins, as membrane proteins which are highly conserved, are another important group of drought stress tolerance related molecules in plants (Tyerman et al., 2002), as they are reported to be involved in soil water uptake efficiency (Javot et al., 2003) as well as membrane water permeability and protein mediated water transport in general (Maurel and Chrispeels, 2001).

As drought stress is often accompanied with other stresses, e.g. heat stress and oxidative stress, other classes of proteins and molecules are directly and indirectly involved in the plant stress reaction to drought (Wang et al., 2003). Well known and studied molecules are for example heat shock proteins (HSP), which were discovered back in 1962 (Ritossa, 1962) and are key components of thermotolerance at the molecular level (Wang et al., 2004). HSP can be distinguished into six major groups with regard to their molecular mass, e.g. small Heat Shock Proteins (smHSP), HSP60, HSP70, HSP90, HSP100 and HSP110 (Kotak et al., 2007). There are multiple effects of the HSP on molecular mechanisms, e.g. acting as chaperones in protein

folding (Miernyk, 1999), protein assembly (Vierling, 1991) or protein aggregation prevention (Kotak et al., 2007), which is of prime importance in the frame of a reduced turgor and cytosol water content which is an effect of water deficiency that can lead to unintended protein-protein interaction, aggregation and denaturation (Hoekstra et al., 2001). Nevertheless, HSPs are reported to be induced by several different stresses and factors, which are not limited to heat or drought stress - like chilling and freezing stress or low oxygen (Coca et al., 1994). An elevated transcription of HSP genes is reported to start at 32°C-33°C tissue temperature for wheat (Vierling, 1991).

Late embryogenesis abundant (LEA) proteins in particular the dehydrin family, known as group 2 LEA proteins (dos Reis et al., 2012; Gupta et al., 2013) are also known to be expressed at a higher level in response to drought stress among others stresses (Mahajan and Tuteja, 2005). These proteins were shown to confer a faster recovery after salt stress and drought stress as well as an elevated osmotic stress tolerance as shown in rice and wheat (Sivamani et al., 2000). Besides these proteins and metabolites, also the underlying genes, networks, transcription factors and other players like micro-RNA or epigenetic factors have been described in various studies (Agarwal et al., 2006; Ahuja et al., 2010; Bartels and Sunkar, 2005; Golldack et al., 2011; Hadiarto and Tran, 2011; Yamaguchi-Shinozaki and Shinozaki, 2006). Also, alterations in the leaf carbon metabolism are reported in the context of drought stress tolerance as for exotic barley accessions from Mediterranean locations (Templer et al., 2017).

1.5 Applications of molecular markers in rye

A very widely used application of molecular markers in agricultural science is the identification and localization of quantitative trait loci, which was also conducted for various traits in rye, e.g. preharvest sprouting (Masojć et al., 2007), in vitro culture response (Bolibok et al., 2007), flowering time (Börner et al., 2000), plant height (Börner et al., 1999; Miedaner et al., 2018), yield, yield components, agronomic traits (Hackauf et al., 2017a) as well as quality parameters (Falke et al., 2009; Miedaner et al., 2012a; Wricke, 2002) and α-amylase activity (Myśków et al., 2012), or morphological traits (Myśków et al., 2014). Beside the extensive use of molecular markers in research projects, also commercial breeding programs are depended on cost efficient genotyping platforms or may benefit substantially from the incorporation of genomic data, especially in the context of hybrid breeding (Eathington et al., 2007; Tanksley, 1983; Zhao et al., 2015). Although modern highly multiplexed genotyping platforms with very low

costs per data point are available also for rye (Bauer et al., 2017; Bolibok-Bragoszewska et al., 2009; Haseneyer et al., 2011) the overall price per sample is challenging for many small to medium sized breeding programs which are common in rye breeding. This is further emphasized, as factors like ascertainment bias (Moragues et al., 2010) or clustering of multiple markers in close genomic neighbourhood (Milczarski et al., 2011) can decrease the overall benefit of these tools for plant breeding (Würschum, 2012) or even have negative impact on results (Stange et al., 2013).

1.6 Objectives

Genomic resources in rye are still scarce compared to wheat and barley and drought stress is gaining increasing importance also in rye. Based on the hypothesis that there is variation in rye with respect to drought stress tolerance which can be related to variation in gene sequence and expression, the aim of this study was the development of a mid-plex genotyping platform for rye. This is based on genomic information generated by sequencing and quantification of drought stress induced transcripts followed by an *in silico* mapping approach using a virtual linear gene order model of the *Triticeae.* This aims to optimize the genome coverage of the platform in terms of an equidistant marker distribution and comprises the validation of this platform by conducting a QTL mapping.

In detail this comprises the following steps:

i) Sequencing and quantifying of drought stress induced transcripts in rye.
ii) *In silico* mapping of drought stress induced transcripts.
iii) Detection of single nucleotide polymorphism (SNPs) within the transcripts.
iv) Creating a mid-plex genotyping platform with optimized genome coverage.
v) Integrate the newly developed markers in the genetic map of rye, including multiple published molecular markers of different types to allow comparisons to other linkage maps of rye.
vi) To proof the quality and usability of the generated marker set by mapping QTL for agronomic traits in a population derived from a commercial hybrid rye breeding program.

2. Material and Methods

2.1 Identification and *in silico* mapping of drought responsive transcripts

The overall strategy of the identification of drought responsive transcripts and the later *in silico* mapping of those as well as the workflow for analysing this data is given in Figure 2.

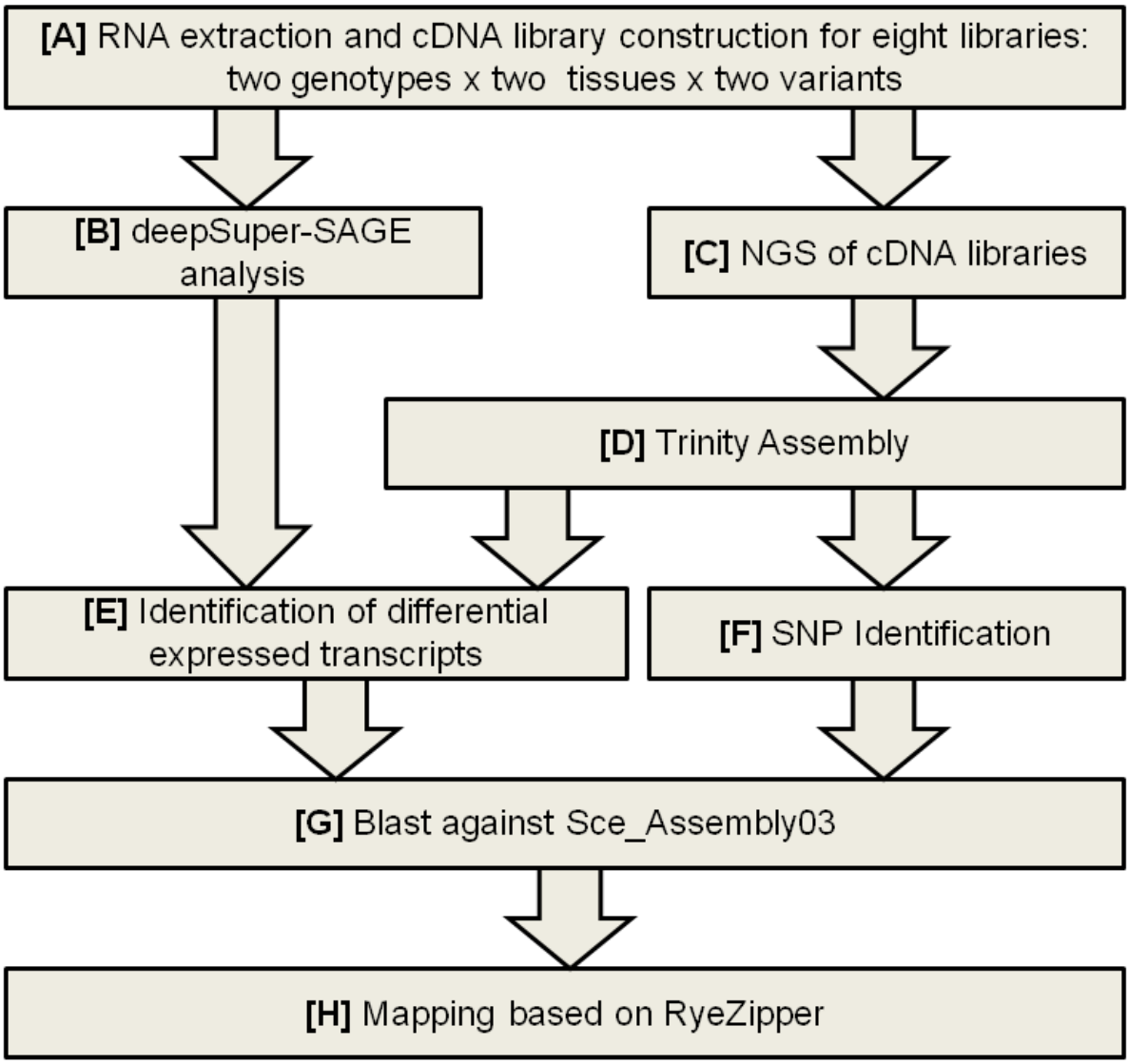

Figure 2. Pipeline for the generation of transcript sequence information and mapping of drought induced transcripts.

2.1.1 Plant material and stress application

Six plants each of two elite inbred lines (Hyb201 and Hyb202) derived from the hybrid breeding program of the HYBRO Saatzucht GmbH & Co. KG (Kleptow, Germany) were vernalized for eight weeks at 4 °C and subsequently planted in a peat substrate. Then, the seedlings were cultivated at short day conditions (12 h light per day, 15 °C day-temperature, 12 °C night-temperature, 40% relative air humidity) in a growth chamber in a complete

randomized design with 6 replications. At the end of tillering (EC 29), growing conditions were changed to long day (16 h light per day, 18 °C day-temperature, 12 °C night-temperature, 30% relative air humidity). After vernalization, plants of the stressed variant were not watered until soil water capacity reached 25%. This soil water capacity was maintained until the end of the experiment. For the well-watered group (control variant) the water capacity was kept at 55% throughout the whole experiment. The water capacity was measured gravimetrically. All stressed and well-watered plants were grown under the same environmental conditions.

Sampling was conducted at two time points: (i) when 25% water capacity was reached, bulked leaf samples from 6 individual plants per genotype were harvested from the stressed and well-watered variant, respectively, (ii) after a further cultivation until the beginning of heading (EC 51) all watering of the stressed variant was stopped for 96 h. At this time point first wilting symptoms became visible. Then a bulked sample of ear, shoot and flag leaf was harvested from six plants each per stressed variant and well-watered control. All samples were frozen in liquid nitrogen and stored at -80 °C until use.

2.1.2 RNA isolation

Total RNA (Figure 2 step A) was isolated following the standard protocol of the NucleoSpin RNA plant kit (MACHEREY-NAGEL GmbH and Co., KG, Düren, Germany). From total RNA, poly(A)-RNA was purified with the Oligotex mRNA Mini Kit (QIAGEN GmbH, Germany) according to the manufacturer's protocol.

DNA traces were removed by digestion with Baseline-ZERO DNase (Biozym Scientific GmbH, Hessisch Oldendorf, Germany). The quality of the total RNA was checked on a Bioanalyzer with a 2100 expert Plant RNA Nano chip (Agilent Technologies, USA).

RNA quality, i.e. RNA Integrity Index (RIN) values (standard developed by Agilent Technologies and Quantiom Bioinformatics) and amount of total RNA used for DeepSuperSAGE and complementary deoxyribonucleic acid (cDNA) library preparation were determined with an Agilent 2100 Bioanalyzer with an expert Plant RNA Nano chip (Agilent Technologies, USA). For the SNP quality analysis, the sequence data sets of control and drought stress were combined. The stress application, as well as the RNA isolation was conducted by the group of Dr. Bernd Hackauf, Julius Kühn-Institute, Institute for Breeding Research on Agricultural Crops, Groß-Lüsewitz, Germany.

2.1.3 cDNA-library preparation and normalization

10 µg of total RNA of each bulk were used for cDNA synthesis. Full-length-enriched cDNA was normalized using the duplex-specific nuclease (Shagin et al., 2002) Trimmer-2 cDNA normalization kit from Evrogen (Moscow 117997, Russia) as described by Zhulidov et al. (2005, 2004) based on nucleic acid hybridization kinetics (Young and Anderson, 1985). After optimization, the cDNA libraries were prepared for Illumina 100 base pairs (bp) paired-end sequencing (Figure 2 step C) on an Illumina HiSeq2000 following standard procedures (Illumina, Inc., San Diego, CA, USA) (Bentley, 2006).

2.1.4 *De novo* assembly of ESTs and SNP discovery

After Illumina paired-end sequencing, raw sequence reads were passed through quality filtering also removing sequencing adapters and cDNA synthesis primers. This quality filtered reads were assembled using the Trinity (Grabherr et al., 2011) RNA-Seq *de novo* assembly tool (Version: trinityrnaseq_r2012-06-08) (Figure 2 step D). The resulting contigs were compared with publicly available sequence resources and annotated via blastx (Altschul et al., 1990) to the Swiss-Prot and TrEMBL database from Uniprot. All plant sequences were retrieved using the web interface. SNPs between the different bulks of the two elite lines were discovered using an in-house tool developed by GenXPro. Only SNPs with a minimum coverage of 5 raw sequencing reads were considered for further analysis.

2.1.5 DeepSuperSAGE library preparation and analysis

DeepSuperSAGE (Matsumura et al., 2012) was conducted with the Standard Operating Procedure protocol of GenXPro GmbH (Frankfurt am Main, Germany) including specific software tools. The detailed workflow is given below. DeepSuperSAGE captures one distinct 26-28 bp long tag each from a defined position of all polyadenylated transcripts, so that the total number of tags is similar to the total number of transcripts examined. Artificial tags eventually arising from PCR steps during sample preparation are eliminated by the TrueQuant technology (also known as unique molecular identifier) developed by GenXPro.

The basic principle of the procedure follows the methodology described in Matsumura et al. (2012, 2010, 2005). For the analysis out of each RNA bulk, 5-10 µg total RNA was used (Figure

2 step B). High-throughput sequencing was performed with the Illumina 50 base fluorescent nucleotide-based system (50 bp reads, Illumina, Inc., San Diego, CA, USA).

Sequences were processed using GenXPro's in-house analysis pipeline. Briefly, libraries were sorted according to their respective index followed by elimination of PCR-derived tags identified by the TrueQuant technology. Distinct DeepSuperSAGE tags were quantified and combined to tag clusters of common origin (UniTags) according to Akmaev and Wang (2004). After TrueQuant removal of artifacts, UniTags were subsequently annotated via blast (Altschul et al., 1990). The DeepSuperSAGE tags were annotated in several rounds at decreasing stringency of homology. The most stringent blast-bitscore for annotation was 52 corresponding to a perfect match between tag and reference sequence in the first step of annotation. Any tag that did not reach this score was re-annotated by lowering the required score in a next annotation step by two points, until in the last step the least stringent acceptable homology score of 44 was reached.

To quantify transcript expression, normalized values of each tag in relation to one million tags were calculated (tpm = tags per million) for all UniTags in the eight libraries. This allowed exact quantification and comparison of mRNA expression in drought stress and control libraries of Hyb201 and Hyb202.

To detect transcripts with significant differences in expression of drought stress and control libraries Fishers-exact test was conducted using an alpha threshold of 0.05.

The preparation, sequencing and analysis of the DeepSuperSAGE libraries was conducted by GenXPro GmbH, Frankfurt am Main, Germany.

2.1.6 *In silico* mapping

The Trinity (Grabherr et al., 2011) assembly identified transcripts, which showed high quality SNPs between both inbred lines. In addition the DeepSuperSAGE (Matsumura et al., 2012) analyses disclosed those transcripts that were significantly differentially expressed under stress. For the later transcripts (Figure 2 step E & F) the putative position in the rye genome was predicted by *in silico* mapping. For mapping, the Sce_Assembly03 (Haseneyer et al., 2011) was converted by the "makeblastdb" tool from the blast+ toolbox (Camacho et al., 2009) version 2.2.29 and used as the database resource for the following blast. Subsequently, a blastn with standard parameters was performed, using the previously selected transcripts as query (Figure 2 step G) and the converted Sce_Assembly03 as database. The information out

of the resulting significant alignments was used to predict the genomic position of the transcripts identified in our study. The genomic position of the Sce_Assembly03 contigs in a virtual linear order is given in the RyeZipper (Martis et al., 2013). To create a link between the information given in the RyeZipper (Martis et al., 2013) and the sequence information from our study, the map position of the non-redundant, stringently selected ESTs, which are presented in the RyeZipper (Martis et al., 2013), was used. This was combined with the information of the significant alignments of the blastn analysis, using an automated in-house tool (available from M. Enders on request), to predict the genomic position of the transcripts in the rye genome (Figure 2 step H). Furthermore, a functional annotation of all transcripts which were used in the *in silico* mapping approach was conducted using the Blast2Go (Conesa et al., 2005) v2.7.0 Gene Ontology pipeline. This comprises a blastx with standard parameters against the non-redundant (nr) protein database followed by GO-Mapping using the b2g_sep13 (Conesa et al., 2005) database.

2.2 Development of a SNP array for QTL detection

2.2.1 Population description

The population used for the following analysis of this study consists of a mapping population with 271 F_2 plants used for genotyping and the derived $S_{2:3}$ lines for field testing and assessment of phenotypic data (Hackauf et al., 2017a). All progenies are derived from a cross of Hyb201 and Hyb202, two rye elite inbreed lines from the pollinator pool (Carsten) of the commercial hybrid rye breeding program of the Hybro Saatzucht GmbH & Co KG (Kleptow, Germany). This population was developed to allow for an integration of newly developed molecular markers together with already known and published markers into a dense genetic map. As this population was also phenotyped, a solid base for QTL detection of the per-se performance was generated. Besides the analyses of this study, another distinct study was conducted (Hackauf et al., 2017b; Haffke, 2015a) which addressed QTL detection in test-crosses using a closely related population. This allows to track QTL detected in both studies, which may have implications for line testing in the scope of hybrid rye breeding.

2.2.2 SNP-array development and genotyping

The dataset generated as described in 2.1 of this work in was used as starting point for the development of a rye SNP array. In brief, the parental genotypes used also in this study (Hyb201 and Hyb202) were exposed to artificial drought stress under controlled conditions. Using an RNA sequencing approach, i.e. DeepSuperSAGE (Matsumura et al., 2010), 4,437 contigs characterized with at least one SNP were identified. Furthermore, an *in silico* mapping was conducted to predict the localization of the contigs in the rye genome.

A stepwise procedure was conducted to select for high quality and well distributed SNPs out of these. In the first step, all contigs with a coverage of less than 20 reads at any position were excluded. The remaining set was further condensed by excluding SNPs closer to each other than 50 bp up- or downstream. The predicted mapping position was used to select a set of 500 SNPs with an optimized coverage of the rye genome. This was achieved by selecting SNPs with a putative equidistant coverage of the complete genome, also considering the relative chromosomal length to adjust the total number of selected SNPs per chromosome. The resulting SNP set was further evaluated using the Illumina Assay Design Tool (Illumina, 2013). The quality scores assigned by the design tool, reflecting e.g. chemical restrictions of the Illumina detection technology, were used to rank the SNPs. The 384 top-scoring SNPs were selected to create a 384-plex GoldenGate/VeraCode® SNP array by Illumina. These 384 SNP will be referred further as RSQ marker (**R**ye **S**NP **Q**uedlinburg) and numbered from RSQ_001 to RSQ_384.

The complete F_2 mapping population, including the parental lines, was genotyped using the previously described RSQ markers, following the standard protocol provided by Illumina using the BeadXpress® plate reader for signal detection and the GenomeStudio Software v2011.1 for data assessment, manually adjusted cluster assignment and generating allele scores per genotype.

The previously described population was further genotyped using 60 SSR markers (Bolibok et al., 2006; Hackauf and Wehling, 2003, 2002; Korzun et al., 2001; Saal and Wricke, 1999), 73 Conserved Ortholog Set markers (COS) (Hackauf et al., 2012, 2009) and 3,117 Diversity Array Technology® markers (DArT) (Bolibok-Bragoszewska et al., 2009). The genotyping was conducted strictly following the producers or published protocols.

The resulting datasets of all marker types were filtered in a very stringent manner, discarding all markers with allele scores in the negative control (purified water without DNA),

heterozygous allele scores or missing data for one or both parental genotypes or more than 10% missing data. Also, markers with significant deviations from the expected ratio of homozygotes and heterozygotes by means of a Chi-Square Goodness-of-fit test in the F_2 mapping population for co-dominant marker types were excluded from further analysis.
To verify the quality of the genotyping, as well as the population itself a principal coordinate analysis was conducted and a respective plot generated (Figure 5) using the Darwin Software, Version 6.0.20 (Perrier and Jacquemoud-Collet, 2006).

2.2.3 Field experiments

Field trials in the years 2011 and 2012 were conducted at the locations Wulfsode (29565 Wriedel, 53°03'46.30''N; 10°14'02.22''O), Groß Lüsewitz (18190 Sanitz, 54°04'15.32''N; 12°20'19.79''O) and Kleptow (17291 Schenkenberg, 53°21'59.73''N; 14°00'07.19''O) all located in Northern Germany. In total 271 $S_{2:3}$ lines derived from the parental genotypes Hyb201 and Hyb202 were evaluated and scored for the trait plant height (cm) at three developmental stages: EC 32, EC 52, and EC 89. Furthermore, the thousand-kernel weight (g), number of spikes and the heading date were assessed as a rating score from one to nine (1=very early, 9=very late). The number of spikes was counted for one row for each entry and normalized to the number of spikes on 50 cm. All field trials were conducted using a randomized incomplete block design (alpha design) with two replicates at each location/year combination (environment). About 15% of check genotypes were included throughout the trials. In all trials the entries were sown in single rows with a sowing density of 25 germinating seeds per row. Plant protection and fertilization was carried out according to local practice.

2.2.4 Statistical analysis

To estimate variance components the following model was used in which the observed phenotype y is composed of the genotypic effect (G), the environmental effects (E), the interaction of both effects (GxE) and residual variance (e).

$$Y = G + E + GxE + e$$

A false positive rate of alpha=0.05 was assumed throughout the statistical analysis of phenotypic data to detect significant deviations from the H_0 hypotheses (no effect on the target trait).

Heritability (h^2) was calculated as the ratio of genotypic variance - Var(G) to phenotypic variance – Var(P) (Fehr, 1987).

$$h^2 = \frac{Var(G)}{Var(P)}$$

All phenotypic statistical analysis were conducted using the software R in Version 2.15 (Ripley, 2001). Furthermore, the R packages 'lme4' and 'arm' were used to calculate adjusted entry means. All field experiments were carried out by Hybro Saatzucht GmbH & Co. KG, Schenkenberg Germany and the Group of Dr. Bernd Hackauf, Julius Kühn-Institute Germany, which also conducted the analysis of variance for the phenotypic data.

2.2.5 Map construction

Out of the 3,250 available markers, 1,395 segregated between both parental genotypes and were used for genetic map construction. This set comprises 1,028 dominant markers. Furthermore, 367 co-dominant markers, comprising 254 newly developed RSQ marker after quality filtering, were integrated in the mapping procedure, following a 2-step approach (Mester et al., 2003) to get a reliable estimate of recombination frequencies and the correct order of the markers along the linkage groups. To achieve this, three datasets (i) a dataset comprising all co-dominant marker information, (ii) a dataset comprising all marker information in coupling phase (iii) a dataset comprising all information in the repulsion phase were created (Hackauf et al., 2017a). Datasets ii and iii were used to create a genetic map for the paternal and the maternal genotype each. These two maps were integrated into one complete map using the dataset derived from the co-dominant markers. The JoinMap® Software Version 4 (Van Ooijen, 2006) was used to construct the genetic map using the Kosambi mapping function (Kosambi, 1943) to calculate genetic distances in centimorgan (cM) from the observed segregation ratios. In total seven linkage groups were created and assigned to rye chromosomes using previously published mapping information of SSR and COS markers.

A plot was generated using the MapChart Software (Voorrips, 2002). The map construction was supported by Dr. Bernd Hackauf, Julius Kühn Institute, Germany.

2.2.6 QTL analysis and validation

The software MapQTL Version 5 (Van Ooijen, 2009) was used for QTL identification. For this purpose, co-segregating markers and those being closer linked than 1 cM were removed from the genetic map (Piepho, 2000a). This skeletal map was the basis for QTL detection together with the adjusted entry means for all six traits to avoid the detection of environmental depended QTL. QTL detection was conducted using the multiple-QTL model (MQM) (Jansen, 1994, 1993; Jansen and Stam, 1994) method with markers as cofactors, integrating additive and dominance effects. Only detected QTLs with peak values exceeding a logarithm of the odds (LOD-score) of 4 were considered as statistically significant.

To get information on the reliability of the QTLs detected in this study a comparison with QTLs known from literature was conducted. To achieve an unequivocal assignment of QTLs between multiple studies, all molecular markers of the QTLs known from the external studies were used to determine their positions in the genetic map created in this study. If at least one marker mapped on the same chromosome as reported in the literature and was located less than 10 cM (Piepho, 2000a) away from a QTL for the same trait on the genetic map of this study, both QTLs were considered as potentially identical in both studies. In total 12 studies were included in the QTL comparison: (Bolibok et al., 2007; Börner et al., 2000, 1999; Falke et al., 2009; Korzun et al., 2001; Masojæ and Milczarski, 2005; Masojć et al., 2017; Masojć and Milczarski, 2009; Miedaner et al., 2018, 2012a; Myśków et al., 2014; Wricke, 2002).

In another study (Haffke, 2015b) a population closely related to the one analysed in this study was used for QTL mapping. To obtain information on the distance of the QTLs identified in both studies, the flanking markers for each QTL reported by Haffke (2015) were integrated in the genetic map generated in this study. In a second step the smallest distance between this mapping position and the QTL interval of this study was calculated based on the genetic map calculated within this study. Only QTLs with distances between intervals less than 10 cM are reported (Piepho, 2000a).

2.2.7 *In silico* mapping quality assessment

The quality of the *in silico* mapping approach is accessed by comparing predicted mapping positions with the genetic mapping calculated from recombination ratios in a bi-parental mapping population (Kosambi, 1943; Van Ooijen, 2006). Three different categories were analysed to access the precision and quality of the *in silico* mapping.

i) At a genome wide scale, the ratio of markers mapped on the same chromosome by both methodologies was determined.

ii) The number of markers, concordantly mapped by both approaches on the same chromosomic region (short-arm, centromeric region, long-arm) was compared with the number of markers mapped on different regions.

iii) On a local scale, the marker order on a specific chromosome given by both approaches was correlated by means of a Spearman rank correlation coefficient, as well as the positioning itself using the Pearson correlation coefficient.

3. Results

3.1 Stress application, RNA extraction and *de-novo* Assembly

The stress application was successfully applied, as the stressed plants showed a reduction in tillers per plant with respect to the first sampling and symptoms of wilting become visible in the cause of no water supply before the second sampling date.

Quality of the RNA used for cDNA library preparation and DeepSuperSAGE is given in Table 5 and quantities of raw sequence reads before and after trimming in Table 6.

Table 5. RNA quality (RIN values) after RNA extraction for genotype (Hyb201 / Hyb202), treatment (well-watered / drought-stressed) and tissue (leaf / ear, shoot, flag leaf).

Genotype	Treatment	Tissue	RNA quality (RIN)
Hyb201	Stress	Leaf	7.2 – 8.1
Hyb201	Control	Leaf	7.0 – 7.8
Hyb201	Stress	Ear, Shoot, Flag Leaf	8.4 – 9.4
Hyb201	Control	Ear, Shoot, Flag Leaf	7.3 – 8.7
Hyb202	Stress	Leaf	7.2 – 7.6
Hyb202	Control	Leaf	7.2 – 7.5
Hyb202	Stress	Ear, Shoot, Flag Leaf	6.4 – 8.4
Hyb202	Control	Ear, Shoot, Flag Leaf	6.4 – 8.6

Table 6. Raw read count before and after quality filtering. Number of raw reads of eight normalized cDNA libraries for genotype (Hyb201 / Hyb202), treatment (well-watered / drought-stressed) and tissue (leaf / ear, shoot, flag leaf).

Genotype	Treatment	Tissue	Number of Raw Reads	Number of Filtered Reads
Hyb201	Stress	Leaf	5,805,456	5,697,014
Hyb201	Control	Leaf	4,192,729	2,676,136
Hyb201	Stress	Ear, Shoot, Flag Leaf	39,582,362	21,976,219
Hyb201	Control	Ear, Shoot, Flag Leaf	40,603,478	22,460,746
Hyb202	Stress	Leaf	9,902,974	5,442,251
Hyb202	Control	Leaf	10,022,771	3,580,975
Hyb202	Stress	Ear, Shoot, Flag Leaf	40,440,781	18,931,776
Hyb202	Control	Ear, Shoot, Flag Leaf	42,322,804	21,783,558

After quality trimming the Trinity assembly was conducted, resulting in a total of 128,010 contigs, with a length between 150 bp and 5,410 bp. The mean contig length was 326 bp

(median 257 bp), the N50 of the assembly was 352 bp. All raw sequences are deposited in the NCBI short read archive under the Bioproject ID PRJNA421405.

Assembled sequences are given in supplementary data S1. Detailed statistics are given in Table 7 and a histogram of the contig length distribution is shown in Figure 3.

Table 7. Statistical measures on a Trinity Assembly of short sequence reads derived from rye (*Secale cereale* L.) RNA.

Statistic	Value
GC Content	47.4%
N50	352
Longest Contig	5,410 bp
Contig Length (mean)	326 bp
Contig Length (median)	257 bp
Shortest Contig	150 bp
Total Number of Contigs	128,010
Total Number of base pairs	41,847,269 bp
Number of Adenine	10,989,000 bp (26.2%)
Number of Thymine	11,041,394 bp (26.4%)
Number of Cytosine	10,003,620 bp (23.9%)
Number of Guanine	9,813,255 bp (23.5%)
Number of N (any base / ambiguous)	0

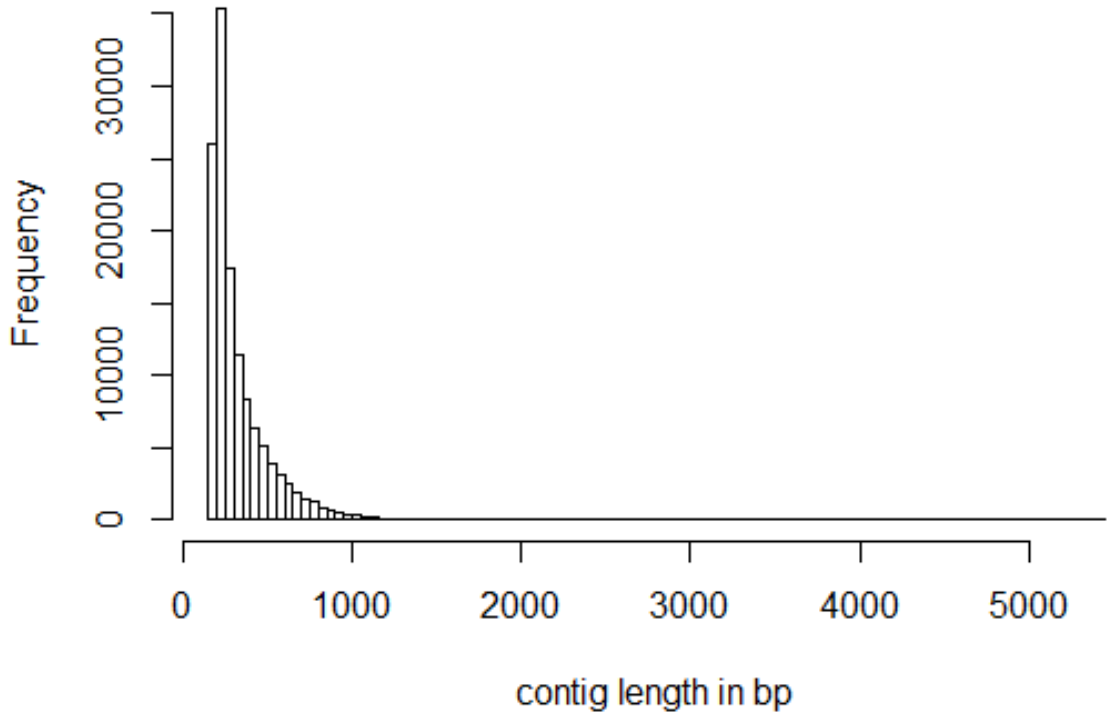

Figure 3. Histogram of the contig length distribution, as a result of a *de novo* assembly of drought induced transcripts in rye (*Secale cereale* L.).

3.2 SNP-Mining and DeepSuperSAGE

Out of in total 128,010 contigs, 7,227 contigs (5.64%) contained at least one SNP resulting in 19,844 SNPs distinguishing both rye genotypes at a rate of 2.74 SNPs per contig. Out of these 7,227 polymorphic contigs, 1,019 derived from leaf tissue cDNA libraries, whereas the remaining 6,208 contigs derived from cDNA libraries from the combined ear, shoot and flag leaf tissues. All high-confidence bi-allelic SNPs together with their flanking sequences are listed in supplementary data S2.

The DeepSuperSAGE analysis revealed expression levels of 408,915 annotated tags in four different treatments (stress/control x leaf/bulk of shoot, ear, flag leaf) for each of the two elite lines. Details on the number and proportion of annotated tags are given in Table 8.
The DeepSuperSAGE information on differential expression and transcript abundance was used jointly together with the hits against the Trinity contigs to predict the expression level of 47,842 contigs in four different treatments for each of the two elite lines. This allows addressing both, the common abundance of the transcripts as well as changes in expression levels after different treatments in both genotypes. The expression levels for all 47,842 contigs per library is given in supplementary data S3. In general, the expression level of all transcripts was higher (1.3 times) in the drought stressed variant than in the well-watered reference genotypes.

Table 8. TAG statistics of DeepSuperSAGE libraries. Statistics include eight different libraries (Hyb202 / Hyb201) x (control /stress) x (leaf / flag leaf, ear, shoot) and are calculated without singletons.

Library	Number of 26 bp Tags	Number of antisense Tags	Antisense Tags annotated	Number of sense Tags	Sense Tags annotated
Hyb201 Leaf Control	18,119,661	12,648	3,595	169,561	45,651
Hyb201 Leaf Stress	33,751,543	15,108	3,738	210,119	50,417
Hyb202 Leaf Control	19,010,951	11,518	3,051	162,816	41,247

Table 8. continued

Library	Number of 26 bp Tags	Number of antisense Tags	Antisense Tags annotated	Number of sense Tags	Sense Tags annotated
Hyb202 Leaf Stress	22,507,986	13,474	3,361	184,681	45,400
Hyb201 Ear, Shoot, Flag Leaf Control	19,143,799	14,177	4,132	184,057	48,538
Hyb201 Ear, Shoot, Flag Leaf Stress	23,146,377	15482	4,138	219,609	52,178
Hyb202 Ear, Shoot, Flag Leaf Control	16,707,100	12395	3,881	178,249	46,283
Hyb202 Ear, Shoot, Flag Leaf Stress	19,672,699	15441	3,904	208,188	49,401

3.3 *In silico* mapping

Mapping to the rye genome zipper was conducted for contigs containing SNPs between the elite lines for which, according to Fishers-exact test ($p<0.05$), DeepSuperSAGE determined significantly different expression levels (Figure 2 step E & F). This filtering resulted in 4,437 polymorphic, stress-responsive contigs. All transcript log-foldchanges, the test statistics for differential expression and the indication of the 4,437 filtered contigs, which are used for further analysis, are given in supplementary data S4. By combining the information deposited in the RyeZipper (Martis et al., 2013) and the significant alignments, the position of 2,754 transcripts in the rye genome was predicted. The resulting *in silico* map is shown in Figure 4 and all mapping positions are deposited in supplementary data S5. With respect to the genetic map of rye given in the RyeZipper (Martis et al., 2013) all seven chromosomes were covered with transcripts.

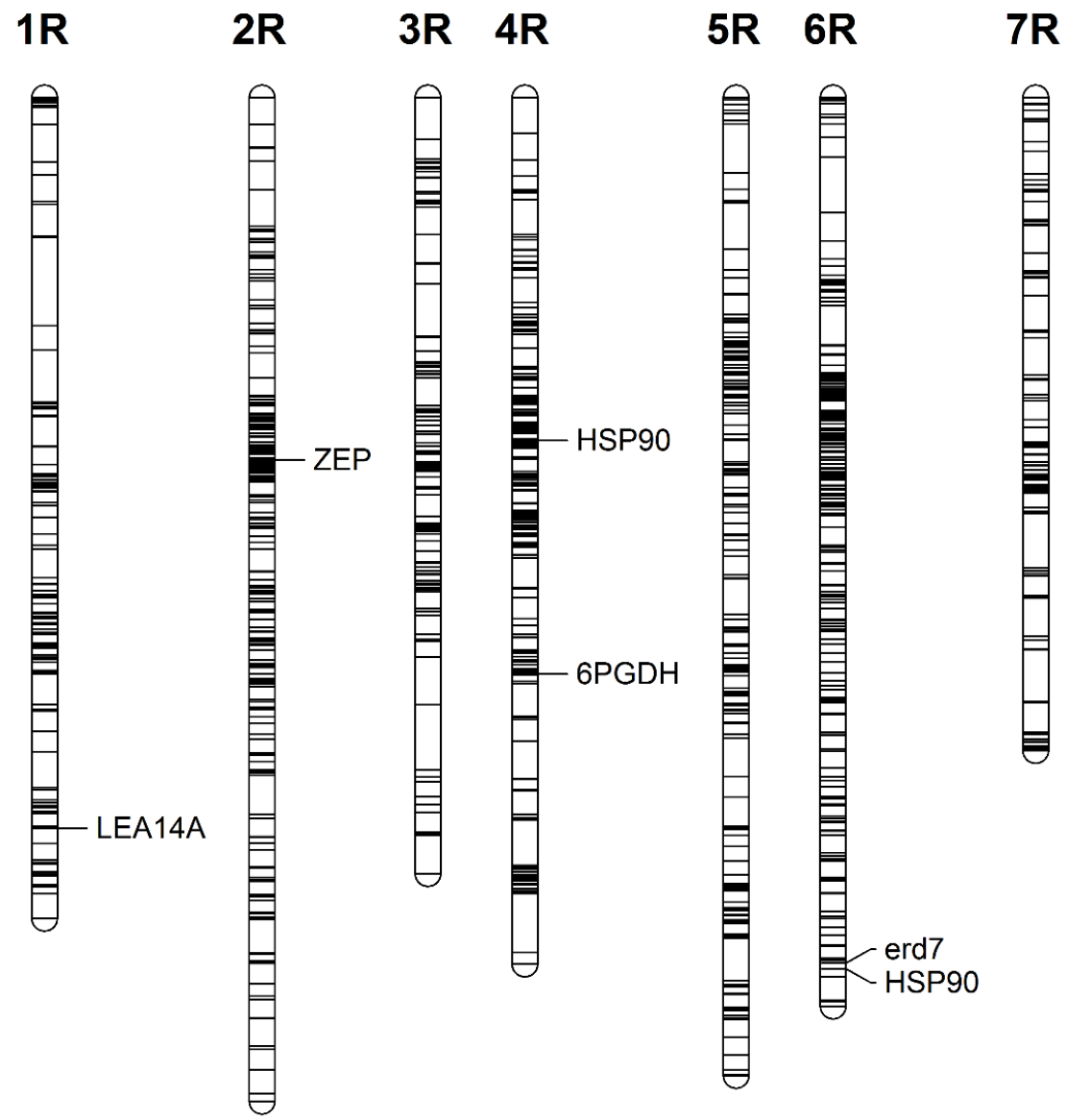

Figure 4. Putative position of 2,754 drought-induced transcripts in the rye genome.
Bold bars indicate the mapping of multiple transcripts in close neighbourhood. Transcripts which revealed high degrees of homology with genes known to be involved in drought stress response are named.

Gene Ontology (GO) mapping of the 4,437 transcripts resulted in 768 transcripts (38.3%) without any hit against the NCBI non-redundant protein (nr) database, which may indicate the presence of interesting new genes, potentially involved in drought stress response. For 737 sequences (16.6%) at least one hit was detected but could not be mapped to any GO term. In total 445 (10.0%) transcripts were mapped to GO-ID 0006950 (response to stress) and are listed in supplementary data S6.

3.4 Genotypic analysis and map construction

As described in 2.2.2 a mid-plex genotyping assay which was build up comprising 384 SNP markers which were used to genotype a set of rye genotypes to assess the quality of the platform. Out of the 384 initially available RSQ markers integrated in the platform, 254 (66%) passed the quality check. 19 out of the 384 markers (5%) showed signals in the negative control, 39 (10%) had missing data in at least one parent, 29 (7.5%) showed significant deviations from the expected segregation ratio and for 3 loci (0.8%) both parental genotypes were scored heterozygous and 40 loci (10.4%) were scored with more than 10% missing data. No monomorphic allele calls were recorded. The remaining dataset contains 1.65% missing data. A principal coordinate analysis (PCoA) using the RSQ marker dataset is shown in Figure 5 revealing the population structure of the mapping population, which is in very good accordance with the expected structure.

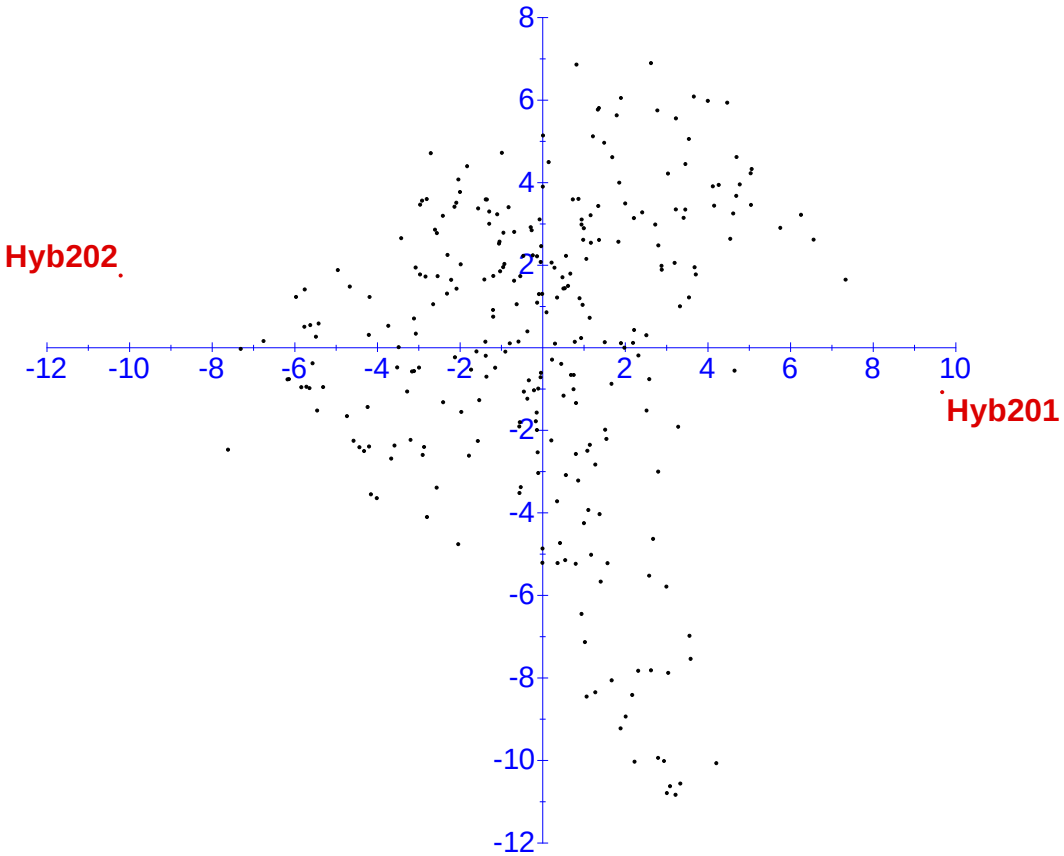

Figure 5. Principal coordinate analysis plot of the bi-parental rye population Hyb201 x Hyb202. Both parents are indicated in red while progenies are displayed as black dots. Results are based on 254 RSQ markers after quality filtering and the PCoA explains 18.36% of the variance.

Out of the 3,117 DArT markers 1,008 markers (32.3%), as well as all 60 SSRs and 73 COS markers passed the quality filtering criteria as described for the RSQ markers.

In total 1,201 of the available 1,395 segregating high-quality markers (86%), including 214 of the 254 RSQ markers, were integrated in a genetic map with a total length of 928.5 cM and assigned to one of the seven linkage groups ranging in length between 80.1 cM and 191.7 cM. Figure 6 shows the resulting complete map, highlighting newly developed RSQ markers. The mapping position of all RSQ markers can be obtained in Appendix 1.

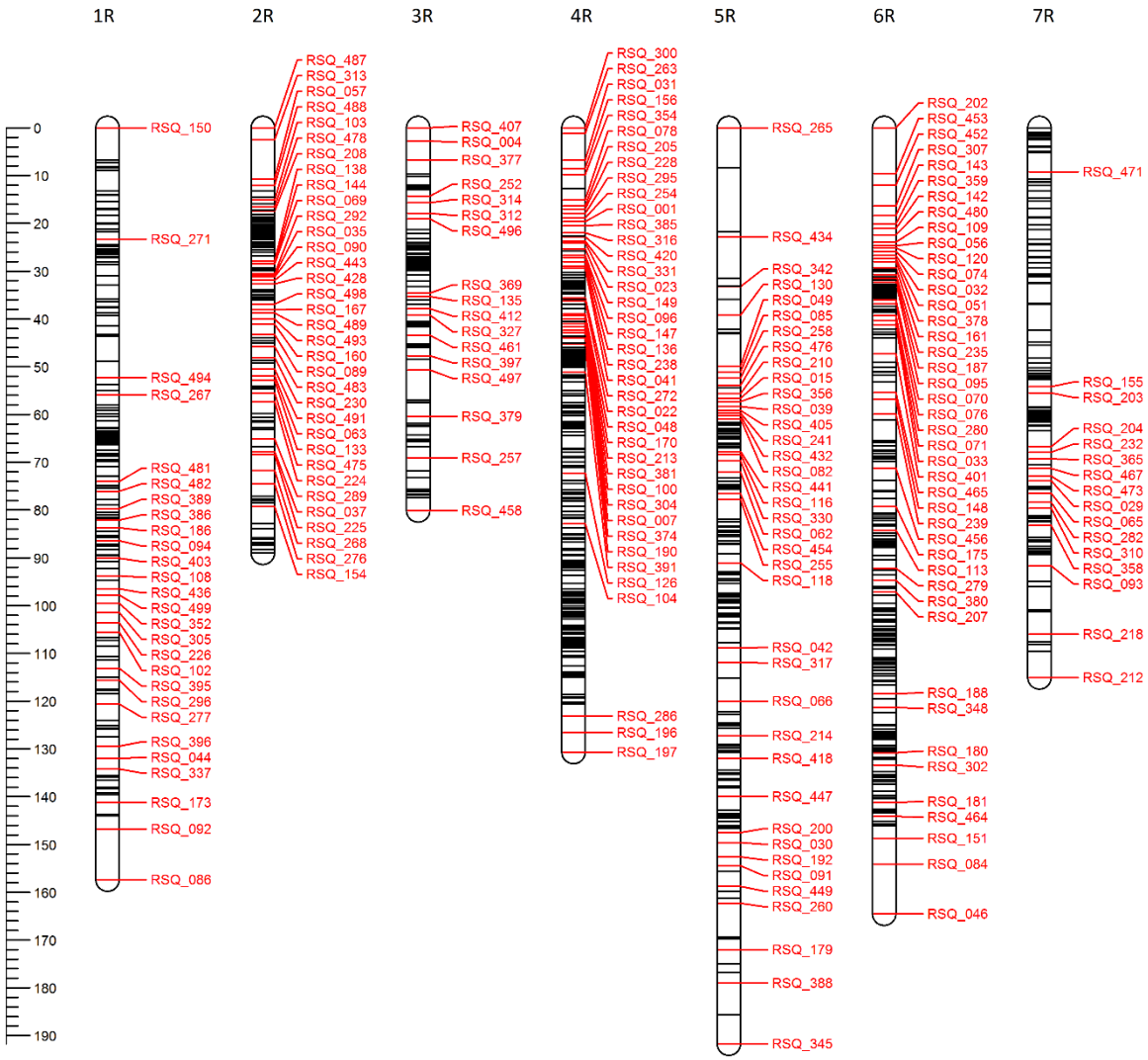

Figure 6. Genetic map of the bi-parental rye population Hyb201 x Hyb202 considering 1,201 molecular markers and 271 individuals. The scale indicates the length of the linkage groups in cM. The localization of newly developed RSQ markers is indicated in red. Black lines indicate the position of other, previously described markers, whereas bold black lines indicate clusters of co-localized previously described markers.

The distribution of DArT markers over the genome is characterized by clusters of multiple markers mapping in close neighbourhood, emphasized by a mean mapping distance (median)

of 0.27 cM between two neighbouring markers. In contrast the RSQ markers were equidistantly distributed, covering a large proportion of the calculated linkage groups with minor gaps on chromosomes 1RS, 4RL and 7RS. A median of 2.31 cM between neighbouring markers was calculated. The genetic mapping positions of all markers is given in supplementary data S7. Descriptive statistics of the genetic map are given in Table 9.

Table 9. Summary of a genetic map of rye integrating 214 newly developed RSQ markers.

Chr.	Length (cM)	DArT	SSR	COS	RSQ	Sum
1R	157.4	117	10	11	27	165
2R	89.0	83	7	4	34	128
3R	80.1	82	5	4	17	108
4R	130.7	194	7	21	39	261
5R	191.7	92	16	17	38	163
6R	164.5	179	11	7	43	240
7R	115.1	107	4	9	16	136
Sum	**928.5**	**854**	**60**	**73**	**214**	**1201**

3.5 Phenotypic Performance

The descriptive statistics for all six measured traits of the adjusted and corrected entry means of the mapping population are listed in Table 10. Box-Whisker-Plots visualizing measures of central tendencies and probable outliers of this data are given in Figure 7. Furthermore, histograms of the same data displaying the frequencies of the respective phenotypes are shown in Figure 8 together with the phenotypes of both parental genotypes.

In general, plant height increases in the course of the three measured developmental stages as expected (Table 10, Figure 8). Furthermore, the ranking of both parental genotypes was persistent throughout all three stages, i.e. Hyb201 was always shorter than Hyb202. Also, an increasing standard deviation was observed from EC 32 via EC 52 to EC 89, which is expected as elevated absolute measures also entail an elevated standard deviation (Pearson, 1894).

For all traits except heading date both parental genotypes showed different phenotypes. As a probable result, major deviations from normal distribution of the adjusted entry means per trait are detected for heading with a Shapiro-Wilk test-statistic resulting in a p-value of $5.669e^{-13}$ rejecting the H_0 hypothesis of a normal distribution. As heading date was measured as a discrete rating score, deviations from a normal distribution by means of the Shapiro-Wilk

test could also arise because of the non-metric character of this trait. All other five traits showed no statistically significant deviation from a Gaussian distribution.

For both yield components, i.e. thousand kernel weight and number of spikes, the performance of the parental genotype Hyb201 was in the same range as the population mean, while Hyb202 revealed a higher thousand kernel weight but a lower number of spikes (Figure 8).

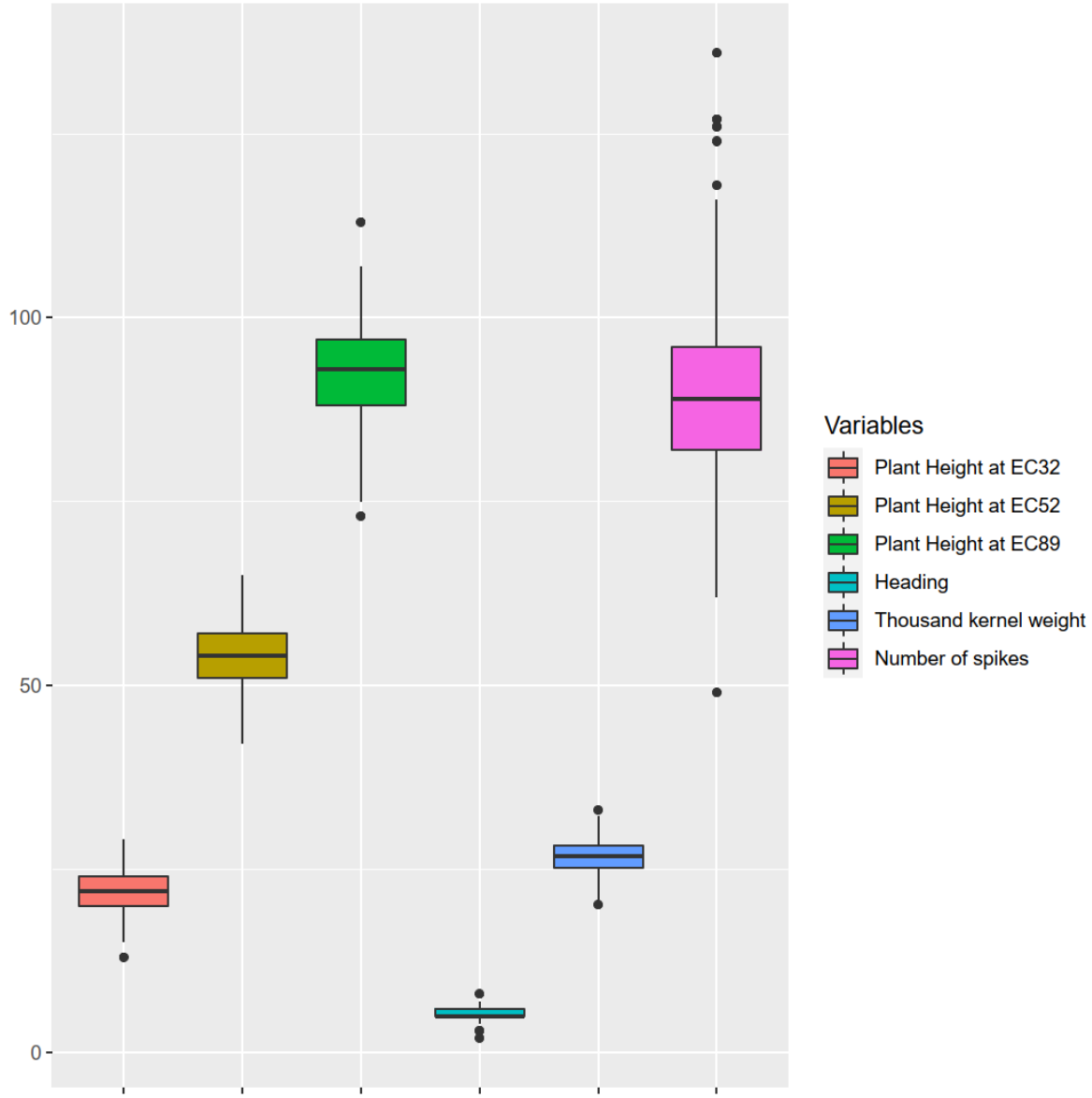

Figure 7. Box - Whisker Plots of adjusted entry means for six traits, measured in the bi-parental rye mapping population (Hyb201 x Hyb202) in two years and three locations per year.

Table 10. Descriptive statistics for six traits (adjusted entry means) for the bi-parental mapping population Hyb201 x Hyb202 including both parents in six environments.

Trait	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. Deviation
Plant Height at EC 32	13	20	22	21.83	24	29	2.67
Plant Height at EC 52	42	51	54	53.78	57	65	4.46
Plant Height at EC 89	73	88	93	92.69	97	113	5.95
Heading	2	5	5	5.44	6	8	0.89
Thousand kernel weight	20.14	25.23	26.74	26.66	28.23	33.02	2.25
Number of spikes	49	82	89	89.80	96	136	12.23

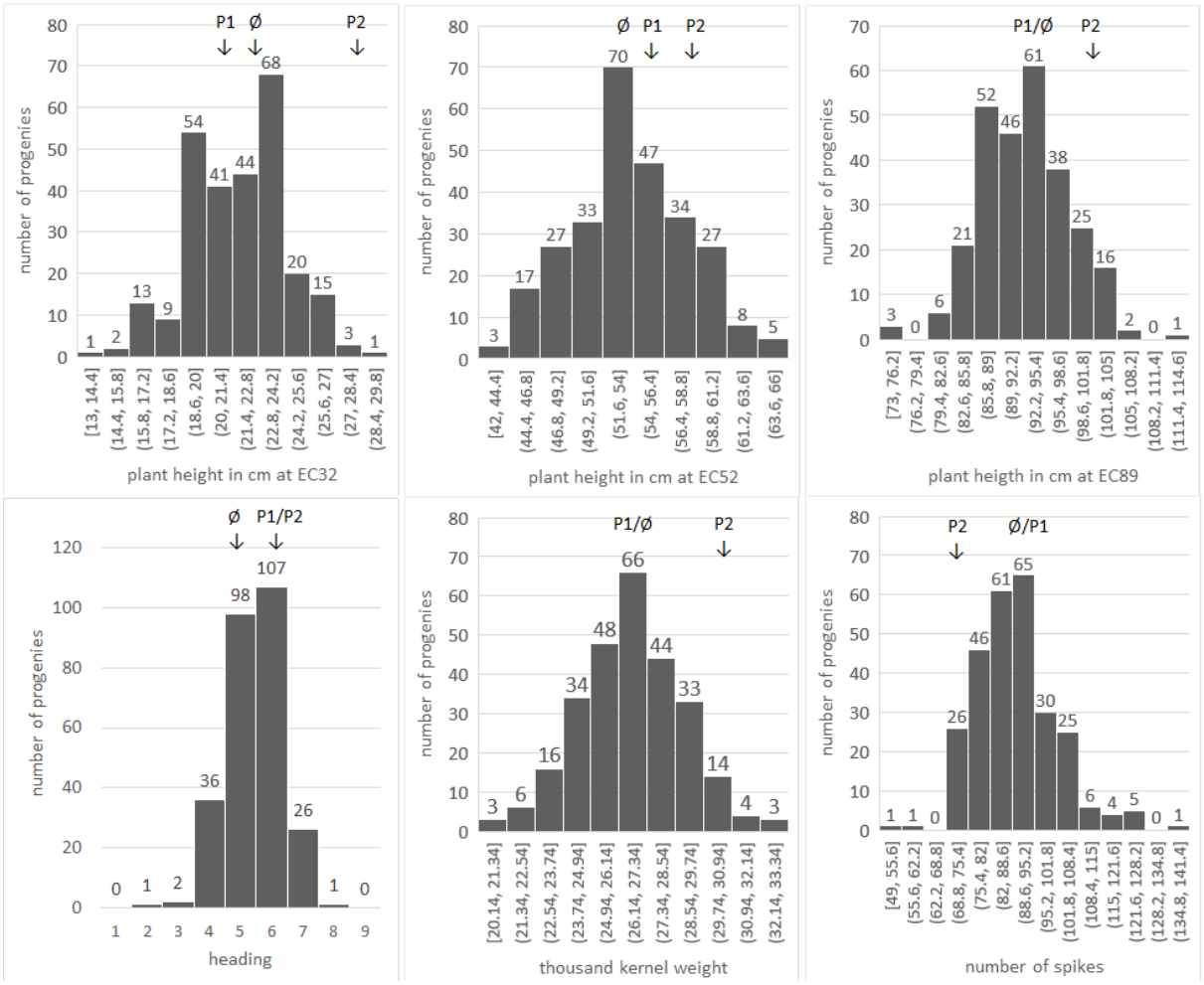

Figure 8. Histograms, showing adjusted entry means for the bi-parental rye population Hyb201 x Hyb202 tested together with the parental genotypes for six different traits in two years and three locations per year. The mean (Ø), as well as the performance of the parental genotypes Hyb201 (P1) and Hyb202 (P2) are indicated at the top of the plots.

A statistically significant correlation at alpha=0.05 was found for all trait combinations expect the combinations plant height at EC 32 versus number of spikes and heading versus number of spikes. The correlation coefficients as well as p-values for each trait combination are given in Table 11 and Table 12, respectively. A scatterplot visualizing the data can be obtained in Figure 9. All correlations are computed as Pearson's *r* and the p-values according to the underlying *t* distribution.

Table 11. Person´s *r* correlation coefficients of six traits measured in the mapping population Hyb201 x Hyb202 of rye.

	Plant Height at EC 32	Plant Height at EC 52	Plant Height at EC 89	Heading	Thousand kernel weight	Number of spikes
Plant Height at EC 32	1.000	0.788	0.401	0.594	0.448	-0.035
Plant Height at EC 52	0.788	1.000	0.718	0.618	0.653	-0.119
Plant Height at EC 89	0.401	0.718	1.000	0.370	0.604	-0.132
Heading	0.594	0.618	0.370	1.000	0.440	-0.058
Thousand kernel weight	0.448	0.653	0.604	0.440	1.000	-0.230
Number of spikes	-0.035	-0.119	-0.133	-0.058	-0.230	1.000

Table 12. Respective p-values of the correlation coefficients given in Table 11.

	Plant Height at EC 32	Plant Height at EC 52	Plant Height at EC 89	Heading	Thousand kernel weight	Number of spikes
Plant Height at EC 32	-	0.000	$6.6e^{-12}$	0.000	$9.3e^{-15}$	0.562
Plant Height at EC 52	0.000	-	0.000	0.000	0.000	0.050
Plant Height at EC 89	$6.6e^{-12}$	0.000	-	$3.3e^{-10}$	0.000	0.029
Heading	0	0.000	$3.3e^{-10}$	-	$3e^{-14}$	0.337
Thousand kernel weight	$9.3e^{-15}$	0.000	0.000	$3e^{-14}$	-	0.000
Number of spikes	0.562	0.050	0.029	0.337	0.000	-

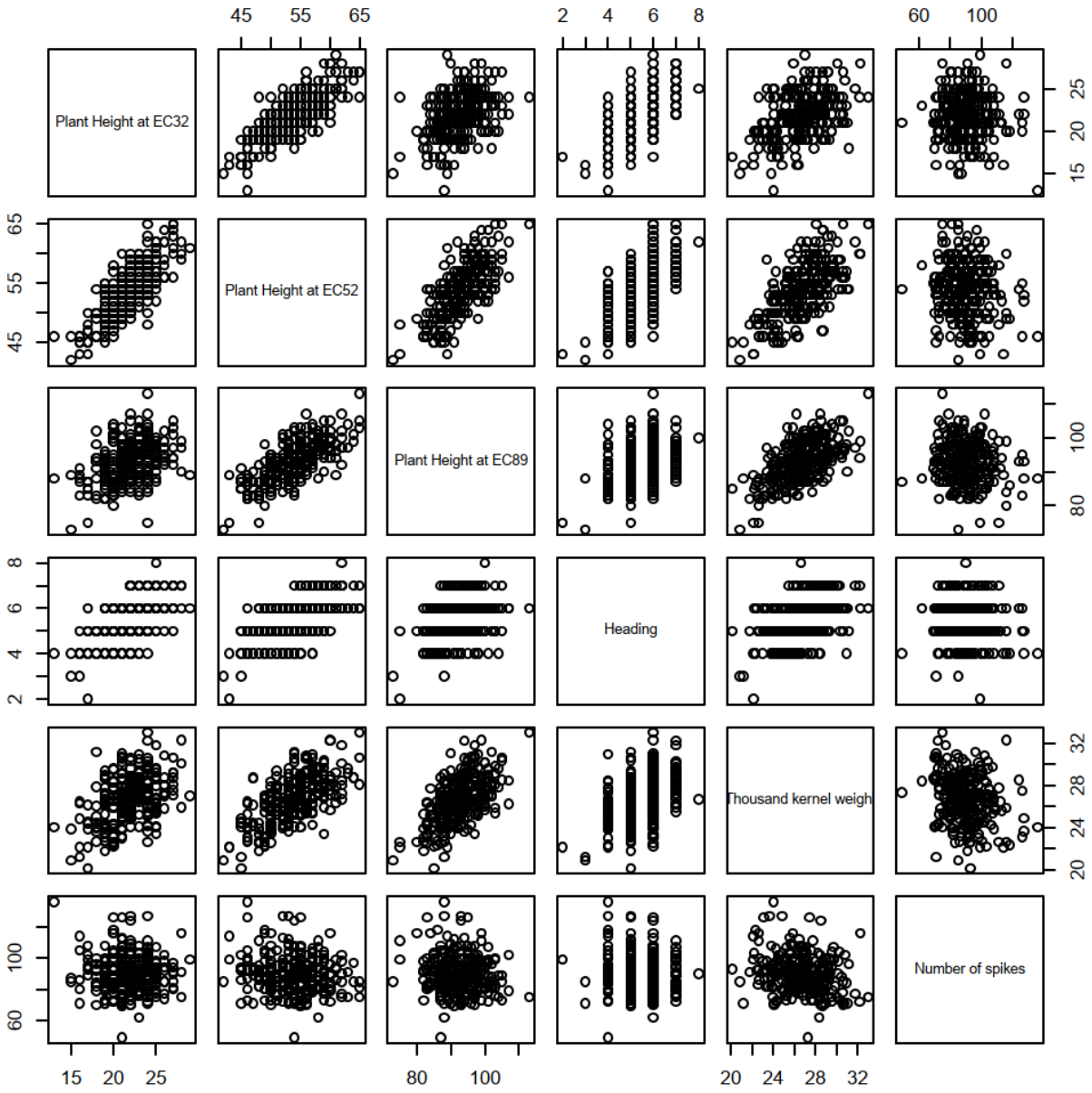

Figure 9. Matrix of scatterplots, visualizing the correlation between the traits measured in a bi-parental rye mapping population Hyb201 x Hyb202, tested together with their parental genotypes in two years and three locations per year. Displayed datapoints are adjusted entry means of the respective measured phenotypes, as described Chapter 3.5.

For all analysed traits, a significant genotypic effect at alpha=0.05 was recognized. Furthermore, all traits except number of spikes showed significant genotype by environment (GxE) interactions. The heritability of the traits ranged from 0.62 to 0.91. Detailed results of the analysis of variance for all six traits can be obtained in Table 13.

Table 13. Analysis of variance of six traits measured on the bi-parental mapping population Hyb201 x Hyb202 including both parents in six environments. Displayed are the parental and population means, the components of variance for genotype (σ^2_G), genotype by environment interaction (σ^2_{GXE}), model error (σ^2_e) as well as the heritability. Asterisks indicate significant effects at an alpha level of 0.05.

Trait	Hyb201	Hyb202	pop. mean	σ^2_G	σ^2_{GXE}	σ^2_e	h^2
Heading	6.43	5.68	5.42	0.58*	0.39*	0.59	0.84
Number of spikes	95.02	75.05	90.02	74.18*	0.00	538.00	0.62
Thousand kernel weight	26.71	30.15	26.90	4.69*	1.17*	2.48	0.91
Plant height EC 32	22.49	28.13	23.29	5.21*	1.59*	10.09*	0.83
Plant height EC 52	56.35	57.96	53.71	15.81*	5.99*	20.65	0.85
Plant height EC 89	99.47	94.11	93.65	31.99*	9.74	25.41	0.90

3.6 Identification of QTL

In total 21 QTLs located on all 7 rye chromosomes were detected as reported in Table 14.

The number of significant QTLs at a LOD > 4 ranged between two for heading date to six for thousand kernel weight. For all traits except plant height at EC 52 QTLs on multiple chromosomes were detected.

Two QTLs were discovered for heading date, one located on chromosome 2R and the other located on chromosome 7R explaining 6% and 13.65% of the phenotypic variance, respectively. The QTL on chromosome 2R is located in very close neighbourhood (1.6 cM) to a QTL for plant height at EC 89. Also, the QTL on chromosome 7R is neighbouring a QTL for plant height at EC 52 with a distance of 1.2 cM. Furthermore, four QTLs for the number of spikes were located on chromosomes 1R, 3R and 5R. The QTL for number of spikes is the only one detected on chromosome 1R. Six QTLs were detected for thousand kernel weight located on chromosome 2R, 4R, 5R and 7R. On chromosome 4R three QTLs for this trait were detected, two QTLs on the short arm of chromosome 4R with a distance of 8.4 cM between them and another QTL on the long arm of chromosome 4R. No QTL of any other trait was detected on chromosome 4R. It is interesting to note, that the QTL on chromosome 7R is located in the same interval as a QTL for plant height at EC 89 and at a distance of 3.2 cM to a QTL for plant height at EC 32. With regard to LOD score and explained phenotypic variance the QTL for

thousand kernel weight on chromosome is the most prominent one out of the six QTLs detected (Table 14). One QTL was detected for plant height at EC 32 on chromosome 5R and in addition two QTLs on chromosome 7R. Although all three measurements of plant height are positively and significantly correlated to each other, no common QTL was detected. However, the only QTL detected for plant height at EC 52 on chromosome 7R is in close neighbourhood (0.6 cM) to a QTL for plant height at EC 32. With regard to plant height at EC 32 three QTLs were detected, one on chromosome 5R and two on chromosome 7R. One of the QTLs on chromosome 7R is located close to the QTL at EC 32 and the other one close to a QTL for plant height at EC 89 at a distance of 3.2 cM. Five QTLs for plant height at EC 89 on chromosomes 2R, 3R, 6R and 7R were discovered with LOD scores between 4.09 and 6.94 and the mean variance explained was estimated between 2.50 % and 5.85 %.

Table 14. Summary of QTLs detected in the population Hyb201 x Hyb202.

Trait	Chr.	Interval in cM	max. LOD of the QTL	Mean explained variance in %*	Peak Marker
Heading	2R	67.7 - 69.3	6.25	6.0	RSQ_225
	7R	61.5 – 66.4	13.01	13.7	XrPT-399361
Number of spikes	1R	62.8 – 63.0	4.06	5.0	Xgwm112-3B
	3R	42.5 – 44.3	6.13	5.8	RSQ_461
	5R	4.0 – 13.3	4.89	7.6	Xscm224
	5R	53.3 – 55.4	5.92	5.8	XrPT-509684
Thousand kernel weight	2R	88.7 – 89.0	4.04	4.5	XrPT-390369
	4R	30.6 – 31.6	4.03	4.7	Xhvpdia
	4R	44.0 – 48.7	6.08	6.6	XrPT-400322
	4R	129.6 – 130.7	4.71	5.3	RSQ_197
	5R	125.7 – 130.9	5.38	5.5	XrPT-411320
	7R	24.3 – 29.1	9.13	9.6	XrPT-509401_r
Plant height EC 32	5R	95.3 – 96.3	4.09	4.4	Xscm74-5R
	7R	32.3 – 36.4	5.15	5.8	XrPT-402607
	7R	70.4 – 74.9	5.99	5.9	RSQ_029
Plant height EC 52	7R	67.6 – 69.8	5.09	4.2	RSQ_232
Plant height EC 89	2R	29.5 – 32.2	5.05	2.5	RSQ_090
	2R	62.8 – 66.1	6.01	5.3	RSQ_289
	3R	29.82 – 33.0	6.94	4.7	Xscm84-3R
	6R	31.4 – 36.0	6.05	3.4	Xscm176-6R
	7R	24.3 – 29.1	6.36	4.9	XrPT-509401_r

*If multiple markers mapped within one QTL interval, the mean of the explained variance of the respective markers was calculated to correct for overestimations.

As an example, the LOD-scores for the plant height QTLs at EC 32, EC 52 and EC 89 located on chromosome 7R are shown in Figure 10.

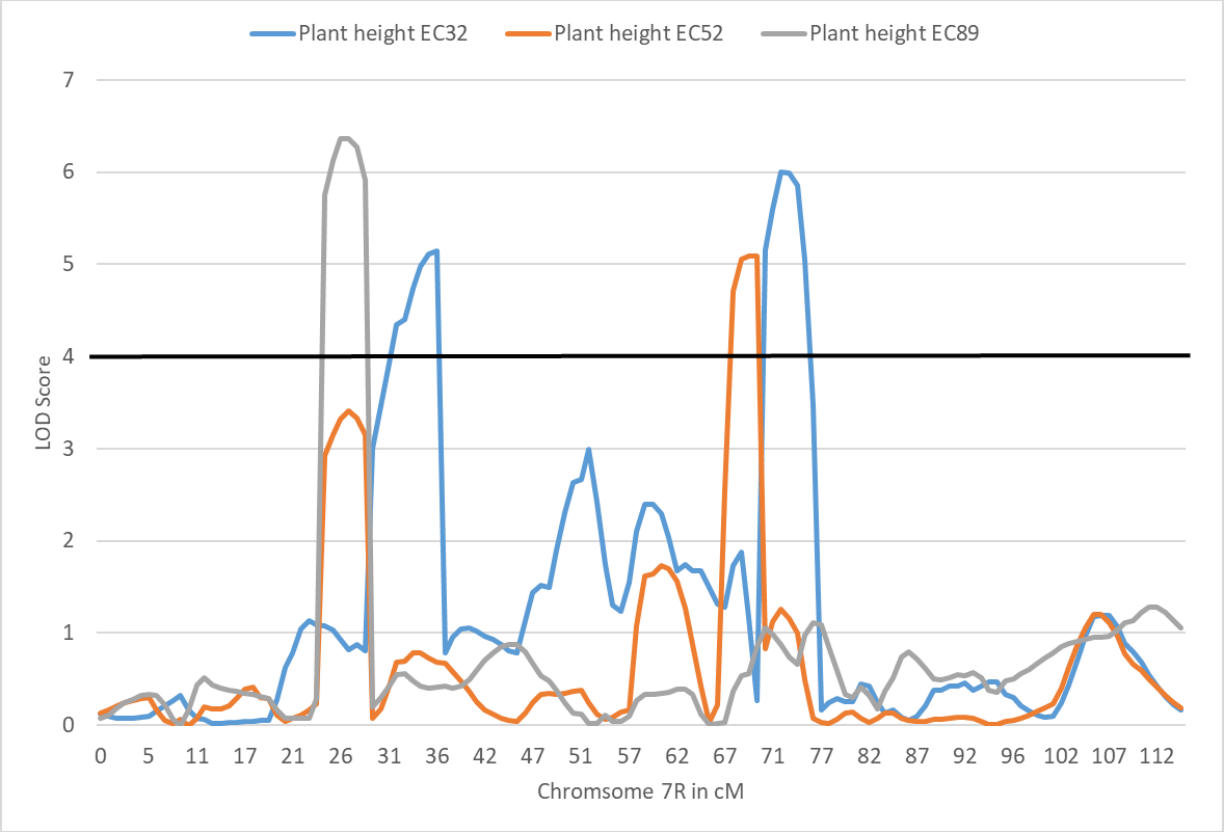

Figure 10. QTLs detected for plant height at different developmental stages located on chromosome 7R. The used threshold (LOD Score 4), to detect significant QTLs is indicated by a bold horizontal black line.

3.7 *In silico* mapping quality assessment

The quality of the *in silico* approach was determined in three different categories: the genome wide level, the intra chromosomal level and on a local scale. In the genome wide assessment 111 of 124 SNP (89.5%) were mapped to the same chromosome by both techniques.

As a next evaluation step, it was observed that 97 of 124 SNP (78%) were mapped to the same region (short-arm, centromeric region, long-arm) of a given chromosome.

Both, the local ordering and positioning of mapping approaches were accessed using suitable correlation coefficients as described hereafter. Therefore, for each chromosome, the Spearman rank correlation coefficient and the Pearson correlation coefficient were calculated. Results can be obtained in Table 15 and visually inspected using Figure 11.

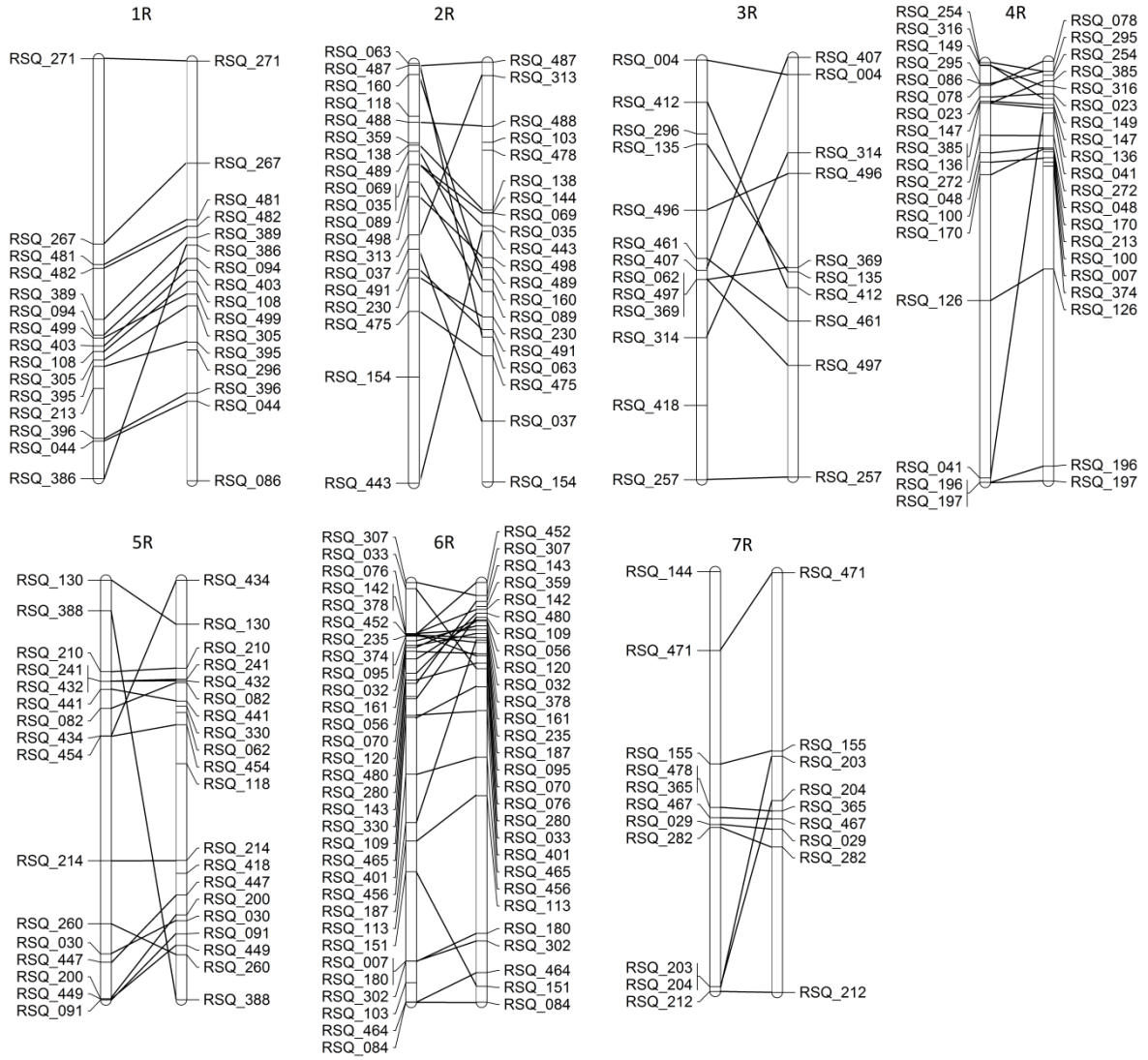

Figure 11. Comparison of an *in silico* map derived from positional information deposited in a synteny based linear virtual gene order of Matris et al. (2013) (left chromosomes) and the genetic map derived from the bi-parental mapping population in rye analysed in this study (right chromosomes).

Table 15. Pearson (r) and Spearman (*p*) correlation coefficients per chromosome, comparing the ordering and position of genetic markers per chromosome between the map derived from *in silico* mapping and the genetic map based on the population Hyb201 x Hyb202.

Chromosome	Number of Markers	Spearman	Pearson
1R	14	0.82	0.85
2R	17	0.45	0.40
3R	10	0.31	0.50
4R	17	0.87	0.81
5R	16	0.56	0.69
6R	27	0.67	0.92
7R	9	0.59	0.66

4. Discussion

4.1 Assessment of differential expression

In accordance with previous studies (Haseneyer et al., 2011; Molina et al., 2008) our dual-purpose transcriptome sequencing approach turned out to be a valuable tool for both, the assessment of gene expression and the detection of sequence variants. In our study the application of the DeepSuperSAGE technology allowed the parallel determination of the relative transcript abundance as well as the level of their differential expression. The raw sequence data in our study is in line with other studies in rye (Haseneyer et al., 2011) as well as other crops (Draffehn et al., 2013; Jain et al., 2014; Trick et al., 2012). The high number of 4,437 SNPs in drought-induced transcripts indicates genetic diversity between both elite inbred lines. Therefore, the DeepSuperSAGE approach proved to be useful to overcome the problem of lacking resources in non-model organism and neglected crops, as has been shown already for chickpea (Molina et al., 2008).

4.2 Drought stress responsive genes in rye

The relatively small rate of transcripts (10%) which were mapped to the GO term “response to stress” in this study exceeds the number of roughly 1% (350 out of 35,356) reported by Haseneyer et al. (2011). The different mode of stress induction used in both studies may be one major factor influencing this finding. Nevertheless, this finding gives an indication for a successful stress application in our study.

A full list of all gene ontology mapping results can be obtained in supplementary data S6. As shown in Appendix 2 several transcripts that were differentially expressed in our study were homologs to proteins known to be involved in the reaction to stress. For example, four homologs to 6-phosphogluconate dehydrogenase gene family proteins (6PGDH, EC 1.1.1.44; NADP+ -dependent (decarboxylating)) were identified which are rate limiting enzymes of the pentose phosphate pathway (PPP) (Kruger and von Schaewen, 2003). As reviewed by Hou et al. (2007), 6PGDH-genes enzymatic activity and/or mRNA expression is significantly increased in response to a wide range of abiotic stresses (salt, zinc, cadmium, aluminium, drought). Specifically, it was shown that Os6PGDH1 (Os06g0111500) is up-regulated in rice seedlings after salt and drought stress. Interestingly, a transcript with homology to the Os6PGDH1 was expressed at a low level in the inbred line Hyb201 but did not respond to stress in any tissue.

However, the same transcript was 44 times more abundant in the genotype Hyb202 already under well-watered conditions. Under drought stress conditions, the transcript was further up-regulated 3.2 times in Hyb202. Another transcript was identified as an Os6PGDH2 (Os11g0484500) homolog. The increase of Os6PGDH2 expression is a reaction to cold and ABA treatment, but especially to salt and drought stress (Hou et al., 2007). This putative regulator of PPP was not detected in Hyb201 but expressed in Hyb202 and in accordance with the known expression pattern of Os6PGDH1 as it was up-regulated in the library of ear, shoot and flag leaf as a reaction to drought. The expression pattern resembles previous findings, which stated that drought-responsive genes are often expressed at a constant high level already prior to the onset of drought. These genes may be controlled by higher endogenous ABA levels (Ingram and Bartels, 1996; Iuchi et al., 2001). This hypothesis is supported by the expression pattern of other transcripts, as well indicating genetic differences between both genotypes in their stress response. The LEA 14A homolog observed in our study and mapped on rye chromosome 1 revealed a minor up-regulation on drought stress in Hyb202 and a strong up-regulation in Hyb201. But, the expression level of the LEA 14A homolog before stress application was 2.25 times higher in Hyb202 which is significantly higher than the detected up-regulated expression in Hyb201.

Three further transcripts homologous to the zeaxanthin epoxidase (ZEP) are of special interest as an elevated ABA level before stress application is assumed to increase drought stress related protein synthesis. ZEP catalyses the conversion of zeaxanthin to violaxanthin which is a main step of the ABA synthesis in the xanthophyll cycle, and therefore may be connected to the plant reaction to drought stress (Agrawal et al., 2001; Audran et al., 1998). All three identified transcripts in our study showed high degrees of sequence homology to each other, and therefore are presumably transcripts of one gene. As a result of the *in silico* mapping, all three transcripts were accordingly mapped at the same position (chromosome 2R, 118.87cM). The transcription patterns of these transcripts in both genotypes are highly similar to the previously described LEA 14A homologous transcript. Hyb201 showed a low abundance of the transcripts before stress and a strong up-regulation after drought stress application. Hyb202 showed no significant up-regulation on drought stress, but exhibits a significantly higher expression rate of the transcripts before stress application in comparison to Hyb201. Thus, ABA synthesis may be an important factor to understand the observed high expression levels of drought stress related genes in Hyb202.

ABA-independent drought stress responses are known as well (Zhu, 2002), e.g. for genes encoding heat shock proteins (HSP) and early-responsive to dehydration stress proteins (ERD), which are up-regulated in response to drought and other abiotic stresses before a significant ABA increase is detected (Kiyosue et al., 1994). We observed six transcripts homologous to HSP90 and one ERD7 homologous transcript, which revealed a differential expression pattern in both rye genotypes. The six HSP90 homologous transcripts mapped to two different loci on rye chromosomes 4R and 6R, respectively, whereas the ERD7 homologous transcript was mapped at chromosome 6R closely linked to one HSP90 homolog. Transcripts of this group show a different expression pattern in comparison to the ABA-dependent group described before. The expression pattern of these transcripts is not significantly different between both genotypes. Both share the same expression level before stress application and the transcripts are up-regulated on drought stress application with the same intensity in both inbred lines. These findings suggest that differences in response to drought stress observed in this experiment may be due to the elevated stress-independent ABA level of Hyb202.

4.3 Fast track SNP-array development

The *in silico* mapping approach presented in this study makes use of information on inter-species syntenic genome regions. This information is based on a broad range of species, particularly of the family of the *Poaceae* comprising major crop species like barley (*Hordeum vulgare* L.), sorghum (*Sorghum bicolor* L.), rice (*Oryza sativa* L.) and the model species *Brachypodium distachyon* L. (Pfeifer et al., 2013; Spannagl et al., 2013). The information gathered from this *in silico* mapping can be useful for different purposes. Primarily, it can support the scientific and breeding community to conduct further studies to understand the drought stress tolerance of rye and the underling mechanisms. Furthermore, the resources generated in this study can be used for the selection of molecular markers derived from sequences based on the mapping information to create high density arrays with a low number of co-segregating markers.

Out of the 4,437 SNPs identified in drought-stress related transcripts, the flanking sequences of 1,006 SNPs (22.5%) showed significant homologies with sequences represented on the publicly available high density 5K rye SNP array (Haseneyer et al., 2011). Furthermore, 2,654 SNPs (59.8%) showed significant homologies with the 600k Rye SNP Array (Bauer et al., 2017). Now a set of 1,783 novel SNPs in drought stress-regulated transcripts as well as about 18,000

SNPs in rye ESTs not differentially responding to drought are available for breeding and research in rye and triticale.

Elite rye genotypes are likely pre-selected for drought stress tolerance, as this small grain cereal crop is traditionally grown on light soils with low water capacity (Hubner et al., 2013). Due to the controlled and uniform environment used in this study, the detected expression level variation is very likely caused by genetic differences. In this study known drought stress related transcripts were observed, supporting the validity of our approach. Furthermore, many transcripts showed high homology to previously described sequences with unknown function. For 768 (17.3%) transcripts, no known homologs were identified. These are eventually specific to rye and thus may contribute to the high stress tolerance of this crop in comparison to other cereals. For the generation of interspecific hybrids with increased drought stress tolerance especially those transcripts/genes showing pre-stress overexpression or post-stress-up-regulation are of utmost interest.

The data presented here on stress-related gene expression led to a better understanding of the response of rye to drought stress whereas the sequence information and resulting SNPs may be directly used for marker-assisted breeding for drought stress tolerance as the underlying genes are already embedded in elite rye breeding material.

4.4 SNP array evaluation

The main aim of the presented work was the development of a high quality mid-plex SNP array optimized for QTL detection in rye including the assessment of quality parameters with special emphasis on the use in commercial rye breeding.

The usability and value of genotyping platforms like the presented 384-plex SNP array are influenced by a wide range of parameters like costs per data point, reproducibility, time needed for generating the results or information content in relation to the analysed plant material. The importance of these parameters may vary to a large extend for different uses ranging from scientific studies to the application in commercial breeding programs. Therefore, the quality evaluation will focus on parameters which are of general importance (Edwards and McCouch, 2007).

The overall proportion of usable markers, i.e. 66% after very stringent filtering, of the newly developed SNP-Array is in very good accordance with other non-optimized SNP arrays based on Illumina technologies, like the 60k Canola SNP array (Edwards et al., 2013; Snowdon and

Iniguez Luy, 2012), the 9k Barley SNP array (Comadran et al., 2012) and the 90k Wheat SNP array (Wang et al., 2014).

The total length of the constructed genetic map is in the same range as previously published maps developed on different rye populations (Devos et al., 1993; Hackauf et al., 2017a, 2009; Korzun et al., 2001; Ma et al., 2001; Miedaner et al., 2012a). But, it is in contrast to a map constructed using DArT markers with a total length of 2,349 cM (Bolibok-Bragoszewska et al., 2009). This large deviation could be a result of an overestimation of the length due to not applying a separate map calculation (Mester et al., 2003), which corrects for the different phases as it was conducted in this study. There are small regions with poor marker saturation like on the short arm of chromosome 5R (Figure 6) which is also known from literature (Milczarski et al., 2011).

The results of the conducted mapping revealed the success of the previously conducted marker selection based on the *in silico* mapping information. On each chromosome, the first marker with a positional information is a newly developed RSQ marker except for chromosome 7R. Also, the last marker on each chromosome is an RSQ marker except for chromosome 2R (Figure 6). Furthermore, the markers show no major patterns of clustering, which in contrast is common for DArT markers as reported before (Milczarski et al., 2011). The equidistant distribution of the RSQ markers is of high value for QTL studies, in which oversaturation or an uneven distribution of marker density does not increase mapping precision (Piepho, 2000a). In contrast, it can reduce detection power (Liu, 1997). Especially for QTL mapping in breeding populations with large extends of linkage disequilibrium at loci under selection, the equidistant genomic distribution is very favourable to not underestimate the variance explained by the QTL (Würschum, 2012). Furthermore, analysis like genetic distance calculations, which are of prime importance for hybrid breeding may benefit from the observed genomic distribution, as individual haplotype blocks or specific sections of the genome are not overrepresented by means of high marker saturation in contrast to other underrepresented genomic regions. On the other hand, the usage of molecular markers in the gene space only, could bear disadvantages for the calculation of the genetic distance in the scope of hybrid breeding.

4.5 QTL regions in rye

As both parental genotypes are elite rye inbred lines derived from the same gene pool of one breeding company, many traits and the overall phenotype is not as contrasting as in many other traditional QTL mapping studies (Bolibok et al., 2007; Masojæ and Milczarski, 2005; Masojć and Milczarski, 2009). This is very likely a result of fixation of major genes influencing traits targeted by multiple selection cycles in the breeding process (Würschum, 2012). This is further supported by the observed phenotypic data for the mapping population and the low differences between both parental genotypes. Therefore, the detection of QTLs with low to medium LOD scores and small amounts of variance explained was expected for this study. As assumed, no major QTL of the investigated traits like the *Ddw1* locus for plant height on chromosome 5RL (Börner et al., 1999; Korzun et al., 1996) or other QTLs for plant height (Börner et al., 2000), thousand kernel weight (Wricke, 2002) or number of spikes (Börner et al., 1999) were detected in this study. Nevertheless, some QTLs, described in literature may be potentially assigned to the QTL regions detected in this study. For details see Table 16.

Table 16. Comparison of QTLs detected in the population Hyb201 x Hyb202 to those known from literature.

Trait	Chromosome / Interval (cM) detected in this study	Reference of QTL study	Marker	Distance
Plant height EC 89	2R 29.5 – 32.2	(Myśków et al., 2014)	XrPT-506926	7.85 cM
Plant height EC 89	2R 29.5 – 32.2	(Myśków et al., 2014)	XrPT-509630	7.55 cM
Plant height EC 89	6R 31.4 – 36.0	(Myśków et al., 2014)	XrPT-401305	0.00 cM*
Plant height EC 32	5R 95.3 – 96.3	(Myśków et al., 2014)	XrPT-506172	8.27 cM
Plant height EC 32	7R 32.3 – 36.4	(Myśków et al., 2014)	XrPT-400783	4.55 cM
Number of spikes	1R 62.8 – 63.0	(Myśków et al., 2014)	XrPT-411519	2.30 cM
Number of spikes	3R 42.5 – 44.3	(Myśków et al., 2014)	XrPT-400319	1.71 cM
Number of spikes	5R 125.7 – 130.9	(Myśków et al., 2014)	XrPT-411109	0.00 cM*

*The marker given in the external QTL study was mapped within the QTL interval in our study.

Table 16. continued

Trait	Chromosome / Interval (cM) detected in this study	Reference of QTL study	Marker	Distance
Plant height EC 89	2R 29.5 – 32.2	(Masojć et al., 2017)	XrPT-509630	7.55 cM
Plant height EC 89	3R 29.8 – 33.0	(Masojć et al., 2017)	XrPT-509013	0.00 cM*
Plant height EC 32	5R 95.3 – 96.3	(Masojć et al., 2017)	XrPT-399462	1.80 cM
Plant height EC 89	7R 24.3 – 29.1	(Masojć et al., 2017)	XrPT-505798	0.00 cM *
Plant height EC 32	7R 70.4 – 74.9	(Masojć et al., 2017)	XrPT-400732	6.34 cM
Plant height EC 52	7R 67.6 – 69.8	(Masojć et al., 2017)	XrPT-509108	5.27 cM
Heading	7R 61.5 – 66.4	(Masojć et al., 2017)	XrPT-399878	0.00 cM*
Thousand kernel weight	4R 44.0 – 48.7	(Masojć et al., 2017)	XrPT-508446	0.00 cM*
Heading	7R 61.5 – 66.4	(Myśków and Stojałowski, 2016)	XrPT-509108	0.00 cM*
Plant height EC 89	3R 29.8 – 33.0	(Miedaner et al., 2012a)	XrPT-400480	4.51 cM
Plant height EC 32	5R 95.3 – 96.3	(Miedaner et al., 2012a)	Xrms1115	6.09 cM

*The marker given in the external QTL study was mapped within the QTL interval in our study.

To further verify the detected QTLs in this study and proof the quality of the newly developed RSQ marker, data from a distinct study (Haffke, 2015b) using a highly related QTL mapping population was used. In brief, a population of 258 three-way hybrids was generated by crossing one CMS line of the Petkus (seed parent) pool with 258 $F_{2:3}$ single-seed descents of a Hyb201 x Hyb202 cross. These inter-pool hybrids were tested at four locations in the same years (2011, 2012) as the mapping population in this study. Therefore, QTLs detected in our intra-pool mapping population should be also detectable in the described material among other QTLs, derived from the Petkus CMS line. The phenotyping of both populations in the same years in adjacent geographic regions (data not shown) further promote the comparability of both studies by minimizing the possibility of environment-specific QTL (Piepho, 2000b). In total, we could clearly assign 4 of the 21 QTLs detected in our study to

QTLs found in the study of Haffke (2015b). Furthermore, 9 additional QTLs are mapped in close vicinity (<10 cM) to QTLs identified in our study, which are very likely to be the same (Piepho, 2000a). In total, 13 of the 21 QTLs detected in our study were most likely also detected by Haffke (2015b). Detailed information on the QTLs found in both studies can be obtained in Table 17. This proves the validity of the QLT mapping approach as well as the veracity of the underlying QLTs.

Table 17. QTLs detected in this study, as well as in the QTLs study of Haffke (2015b), using a closely related QTL mapping population but considering different phenotypic data. Given are the name and the QTL position with respect to the different genetic maps generated in both studies as well as the distance of the nearest flanking marker reported by Haffke (2015b) to the QTLs found in this study.

This Study		Haffke (2015)		
QTL	**Interval (cM)**	**QTL**	**Interval (cM)**	**QTL distance**
Heading – 2R	67.7 - 69.3	QHdt-2R.2	83 – 88	6.31 cM (XrPT-402599)
Heading – 7R	61.5 – 66.4	QHdt-7R	56 – 59	1.20 cM (XrPT-399686)
Number of Spikes – 3R	42.5 – 44.3	QSsm-3R	69 - 72	0.00 cM*
Number of Spikes – 5R	53.3 – 55.4	QSsm-5R	48 – 52	8.16 cM (XtPT—3980)
Thousand kernel weight – 2R	88.7 – 89.0	QTgw-2R	104 - 107	4.79 cM (XrPT-398612_r)
Thousand kernel weight – 4R	30.6 – 31.6	QTgw-4R.2	50 - 52	2.64 cM (XrPT-400488)
Thousand kernel weight – 4R	44.0 – 48.7	QTgw-4R.2	50 - 52	0.00 cM**
Thousand kernel weight – 5R	125.7 – 130.9	QTgw-5R	118 – 125	5.30 cM (Xtnac1394)
Plant height EC 32 – 5R	95.3 – 96.3	QPh1-5R	90 – 96	8.36 cM (XrPT-389427)
Plant height EC 32 – 7R	70.4 – 74.9	QPh1-7R	56 – 60	8.80 cM (XrPT-402149_r)
Plant height EC 89 – 2R	29.5 – 32.2	QPh3-2R.2	54 – 57	0.00 cM**
Plant height EC 89 – 2R	62.8 – 66.1	QPh3-2R.3	87 – 90	3.04 cM (Xtcos5085_2RL)
Plant height EC 89 – 3R	29.8 – 33.0	QPh3-3R	47 - 51	0.00 cM***

* QTL in Haffke (2015b) is completely enclosed in the QTL detected in this study.
** The flanking markers reported by Haffke (2015b) enclose the QTL detected in this study completely.
*** Both QTL are partially overlapping.

In conclusion, we created a rye mid-plex SNP-Array with an optimized marker distribution covering large proportions of the rye genome in equidistant intervals further expanding the genetic toolbox available for scientific studies and rye breeding, especially promoting QTL detection and genetic distance analysis for hybrid breeding.

4.6 Breeding for drought resistance

Drought stress itself and the plant response to this stimulus is complex as well as the factors influencing drought stress. This complexity starts right at the definition of drought stress, which has to take into account the severity of the stress, the time point of onset in relation to plant development, the duration of the water deficiency as well as radiation intensity (Farooq et al., 2009; Zhu, 2002). Therefore, drought stress is very variable and extremely difficult to reproduce in field conditions. In addition, this trait is inherited in a polygenic and quantitative manner, which renders breeding for drought stress tolerant genotypes difficult (Blum, 1983; Ceccarelli and Grando, 1996; Sinclair, 2011). Therefore, drought stress tolerance is most often only assessed as a secondary trait in plant breeding, which influences primary target traits like grain yield (Araus et al., 2002; Hubner et al., 2013). During the breeding process, drought stress appears occasionally in some environments. As candidates for new cultivars are selected for high yield, an indirect selection for drought stress tolerance is applied (Hubner et al., 2013). A complex and difficult trait to asses during the breeding process are roots which are of prime importance for drought stress tolerance (Paez-Garcia et al., 2015). In contrast to the above ground plant parts, the root system architecture, total depth, development over time and many other root traits are not easily accessible for phenotyping under field conditions for a large number of genotypes, which would be a requirement for phenotyping in the context of commercial breeding programs (Atkinson et al., 2019). First scientific attempts to uncover the rhizosphere are ongoing, using e.g. rhizotrons (Nagel et al., 2012) to get access to root traits. In addition soil sensors to measure e.g. soil water content or soil-temperature (Baggio, 2005); or using microbiome analysis to unlock the interaction of plant roots with soil-borne organisms (Zolla et al., 2013) are used. Furthermore, non-destructive imaging techniques for field application are under development like Electrical Resistance Tomography (ERT) (Amato et al., 2008; Srayeddin and Doussan, 2009), electromagnetic inductance (EMI) (Shanahan et al., 2015; Whalley et al., 2017) or ground penetrating radar (GPR) (Liu et al., 2016). Nevertheless, most of this methods are employed in trees, as there is

a lack of resolution to be directly useable for cereals or field crops in general (Atkinson et al., 2019).

Consequently, a large proportion of factors influencing drought stress tolerance are not or at least only indirectly available for efficient selection in plant breeding. This slows down the development of new varieties with an improved performance in water deficient environmental situations, which will be present in many geographic regions due to elevated average temperatures or decreased precipitation forecasted in the context of global climate change (Cox et al., 2000; Misra, 2014; Nelson et al., 2014). Even though irrigation may be employed in a range of agricultural production systems in theory, some negative factors of using artificial irrigation are discussed (Fernández-Cirelli et al., 2009). Therefore, one promising target to improve crop performance in environments with limited water availability by means of targeted breeding approaches is to study the plant stress response on the molecular level in order to identify markers and in a next step favourable genes or alleles (Nguyen et al., 1997). The combination of this knowledge with low cost and high throughput genotyping platforms enables breeders to conduct time and cost-efficient marker assisted selection (MAS), as well as a marker-guided selection of crossing parents to generate populations designated for drought prone environments (Hussain, 2006; Reynolds et al., 2010; Sinclair, 2011). As RNA sequencing allows the identification of genes and their respective expression level depending on a specific treatment at a given time point, it is an efficient tool to identify favourable alleles and develop molecular markers in a one-step approach (Matsumura et al., 2012; Ozsolak and Milos, 2011). This is further emphasized by the possibility to dissect some complex traits into components (Slafer et al., 1996). Similar to yield components like kernels per ear or thousand kernel weight, some components of drought stress tolerance may be controlled by a limited number of genes, e.g. an elevated water use efficiency (WUE) (Blum, 2005) or root architecture (Passioura, 1983). Some of this drought stress tolerance factors may be even independent from environmental factors, meaning these are favourable in some stress environments and not detrimental in other environments (Blum, 2005). One component of drought stress tolerance, bearing these features may be a stress-independent elevated abscisic acid (ABA) level, which was found and described earlier in other studies (Shinozaki and Yamaguchi-Shinozaki, 2007; Zhu, 2002) and may also be involved in the differential, genotype specific drought stress response reported in this study.

Due to the variability of drought stress itself, it is highly plausible to assume that there is not one single idiotype, which performs optimal across each environment with limited water availability (Blum, 2005). Nevertheless, there are multiple possibilities and methods known from literature, which aim to tackle the challenges of water shortage for crops.

One example is the utilization of the natural symbiosis of plants with fungi, termed mycorrhizal symbiosis, which is also present in rye and other cereals (Gollner et al., 2011; Pandey et al., 2005). Positive effects of arbuscular mycorrhizal symbiosis on drought tolerance was found for many crops, e.g. wheat (Al-Karaki et al., 2004), barley (Khalvati et al., 2005) or maize (Boomsma and Vyn, 2008), whereas the underlying mechanisms are reviewed by Augé, (2001). As a recent study on wheat found genetic variation and associated QTLs for the ability of genotypes to create a better mycorrhizal colonization (Lehnert et al., 2017), a MAS breeding scheme may be applied to select for genotypes with an elevated drought tolerance, due to an elevated mycorrhizal colonization (Lehnert et al., 2017). Although, no QTLs or associations with this trait were reported for rye until now, one could assume a high possibility of their existence.

Another method discussed in literature for improving drought stress tolerance, is the use of genomic selection in plant breeding (Heffner et al., 2009). First positive results have been reported for maize (Beyene et al., 2015; Shikha et al., 2017), wheat (Mwadzingeni et al., 2016) or pea (Annicchiarico et al., 2017). Nevertheless, a large proportion of papers apply genomic selection on yield as the target trait and include yield data of water limited environments to estimate the marker effects. Therefore, the (genomic) selection for drought stress tolerant genotypes is indirect or at least mixed with the selection for genomic regions, which confer better yield.

Also, the use of methods to alter the genetic code or incorporate novel genes into the genome of a given species, hereafter named genetic engineering, is a possibility to increase drought stress tolerance. Most prominently the so called “Stay-green” genes (Borrell et al., 2013) are used to increase drought stress tolerance. Despite of some favourable features of genetic engineering, like a short time to incorporate a new trait into a variety in comparison to conventional breeding (Lusser et al., 2012), there are also challenges for the application, like the complex architecture of drought stress tolerance (Mittler and Blumwald, 2010) or legal issues, prohibiting or hampering the utilization of these techniques in many parts of the world (Hartung and Schiemann, 2014).

As a long-term perspective, disruptive technological changes in the fields of genetic engineering and the so-called synthetic biology could give rise to other possibilities to cope with drought stress, e.g. the exogenous control of plant metabolism on-site by farmers using agrochemicals like phytohormones may be employed (Farooq et al., 2009) to adapt the plant development according to weather forecasts. This would allow to change the principle of just reacting on a stress stimulus after it appears, to a paradigm of anticipation to future events like drought stress before it may happen. Basic examples for this idea are already in use, like the application of plant growth regulators, conferring an elevated winter hardiness (Carter and Brenner, 1985).

4.7 *In silico* mapping

In silico mapping provides a very cost, labour and time efficient method to select a set of polymorphisms with optimized genomic coverage. The differences between both maps identified in this study are potentially influenced by many factors; therefore, the precision and quality of the mapping approach will be discussed separately for three categories of possible causes. The first and most important factor is the extent of synteny of the rye genome with the genomes of the other species represented in the genome zipper, which was used as the data source for the *in silico* mapping. The observed 89.5% indicates a high quality of the linear virtual gene order data source (Martis et al., 2013). This percentage is slightly higher than reported for rye DArT markers (85.4%) by Gawroński et al. (2016) and also higher than reported for SSR markers (83.2%) by Li et al. (2018), which both employed a comparable methodology for *in silico* mapping in rye. As a valid chromosomal assignment is not sufficient to optimize the genome coverage, the assignment to chromosomal regions and local ordering of the predicted mapping positions must be evaluated as a second factor. In total, 97 of 124 SNP (78%) are assigned to the correct chromosomal region (short-arm, centromeric region, long-arm), taking all chromosomes into account. The number of mismatched chromosomal assignments may be due to structural variation and chromosomal rearrangements between the two genotypes used in our study (Hyb201, Hyb202) and the varieties used to generate the draft reference sequence used for the genome zipper (Lo7, Lo225, P87, Lo90, Lo115, L2038-N, L285, L290) (Martis et al., 2013). Both sets of genotypes are derived from different breeding companies (KWS Lochow GmbH, Germany for the genome zipper; Hybro Saatzucht GmbH & Co. KG, Germany for Hyb201 and Hyb202). Furthermore, the genotypes used to generate the

data for the genome zipper are derived from both breeding pools (Martis et al., 2013). In contrast, the genotypes of the current study are members of the seed parent pool, only.

On the smallest scale the local ordering and mapping, as the last factor for discussing the precision of the mapping approach, must be evaluated with caution. It is known that the ordering and the distance between polymorphisms within a genetic map can greatly differ with respect to the used mapping population, number of individuals in the mapping population, the population structure itself and the used mapping algorithm (Qi et al., 1996; Ruiz and Asins, 2003). As we used just one mapping population in this study, the quality assessment on this small-scale level is just one sample and does not allow the generalization of the observed precision to a general estimation of quality statistics. To assess the quality, we conducted a correlation analysis per chromosome, using both, the Spearman rank correlation coefficient assessing the similarity in the ordering of polymorphisms and the Pearson correlation coefficient to assess the similarity in the context of the positional information. Of course, both metrics are highly interconnected and correlated, but on the other hand both can be used to get a slightly different view on the quality.

To correctly interpret the results of the correlation analysis, one must keep in mind that only markers which are mapped to the correct chromosome (111 of 124 - 89.5%) were considered, which leads to an overestimation of the correlation. On the other hand, positional information of mapped polymorphisms covering whole chromosomes were used in the analysis. Therefore, variations in mapping positions spanning over complete chromosomal regions (short-arm; centromere; long-arm) are included. This later source of variation accounts for a substantial amount of reduction in mapping quality by means of correlation coefficients. This is demonstrated nicely on chromosome 1R, where the altered mapping position of just two markers (RSQ_386; RSQ_499, Figure 11) disturbs the otherwise perfectly matched ordering of both maps.

These mismatches, especially for RSQ_386 spanning half of the chromosome, reduces the Spearman rank correlation coefficient significantly. This kind of large range mismatches are found with various frequencies on all chromosomes. The difference of the order (Spearman) as well as in the position (Pearson) is used in a quadratic manner to calculate deviations from a perfect correlation, putting high penalties on such large-scale deviations. This results in an overall lower correlation, which is masking the quality on a local scale to some extent, when just evaluating the correlation coefficients. To sum up, a substantial fraction of the

chromosomal assignment was forecasted correctly as well as the mapping distance between marker pairs and the ordering itself. Despite the fact that no exact correlation coefficient was given, Gawroński et al. (2016) suggest a comparable quality of the *in silico* mapping approaches in both studies on all three discussed levels. Nevertheless, the collinearity of the *in silico* predicted position and the position in the genetic map in this study (r = 0.69) is lower, than observed in a barley study (Silvar et al., 2015) (93.2%). This difference potentially results from two sources:

i) To calculate the collinearity, Silvar et al. (2015) used a binning based approach, neglecting errors in the ordering on a very small scale (< 5cM) and calculating a percentage instead of a correlation, minimizing the effect of mismatches, which spans long ranges and therefore have large impact on the correlation, as discussed previously.

ii) Potentially the *in silico* mapping of genomic features in barley using the genome zipper (Mayer et al., 2011) gives higher quality results, as the integrated data basis for barley is of better quantity and quality. Another reason may be the higher synteny of barley with the other species included in the genome zipper, compared with rye (Martis et al., 2013).

Most important and in accordance with the aim of the *in silico* mapping approach, the coverage of each chromosome with molecular markers with an optimized spacing was achieved. This result is further emphasized, when comparing the equidistant genome coverage of the mid-plex SNP genotyping platform with the molecular mapping data of other platforms like DArT markers, as described earlier in this study.

4.8 Evaluation of the molecular marker resource and genotyping platform

As discussed in the previous section, the RNA sequencing in combination with the *in silico* mapping approach was able to detect large quantities of SNPs, e.g. the study of Haseneyer et al. (2011) identified 17,917 SNP candidates after quality filtering using RNA sequencing, which is in the same magnitude as the 19,844 SNPs identified in this study. On the other hand, a more recent study of Bauer et al. (2017) identified 8,405,856 SNPs in total. This substantially

higher number was achieved due to a set of differences used in the approach of their study, e.g.:

i) Genotypes from both pools (seed parent and pollinator) were sequenced, as well as five *Secale vavilovii* G. accessions. Taking this into account, a reduced number of 4,010,067 intra seed parent pool SNPs were detected by Bauer et al. (2017) which have to be compared to the number of SNPs detected by our study.

ii) A whole genome sequencing approach was conducted. Therefore, SNPs in both, the genic and intergenic regions were detected, whereas the RNA sequencing approach in our study is only able to detect SNPs in expressed genes in terms of mature mRNA after splicing and potential post-translational editing. As the gene space spans only about 5% of the complete rye genome (Flavell et al., 1974; O'Neill et al., 2003), a substantial proportion of variation and polymorphisms is not detectable by our approach.

iii) Bauer et al. (2017) used six rye inbreed lines of the seed parent pool for variant detection while in our study only two lines were used. A reduced number of detected variants was expected, therefore.

iv) A set of other parameters is different between both studies, e. g. the length of raw reads from sequencing, the sequencing depth or the sequence analysis pipeline. These differences will most probably also have impact on the number of detected SNPs.

The prediction of the genomic positions of these polymorphisms with sufficient precision, enabled the selection of a subset with optimized genome coverage for the generation of a mid-plex SNP array genotyping platform in rye. To evaluate the usability of this resource for scientific and breeding purposes, a range of indicators can be used (Varshney et al., 2007b). Many features are a direct consequence of the technique used to generate the data, in our case RNA sequencing. As the procedure uses a protocol to capture polyadenylated messenger RNA molecules, only transcripts of genes after splicing and post-processing, so called mature messenger RNA, is represented, skipping the complete non-coding sequence information of the genome including introns and the vast majority of mitochondrial and chloroplast transcripts as poly-adenylated mRNA from both is degraded rapidly in higher plants (Chang

and Tong, 2012). Therefore, a large portion of sequence variability is neglected. In a marker assisted selection scenario, this setting is favourable, in particular when causal genes and variants are represented in the resource itself (Collard and Mackill, 2007). In contrast, assuming a marker assisted background selection for backcrossing or a diversity study, a set of polymorphisms in non-coding regions, which are neutral in the context of selection showing high allelic diversity, would be favourable instead (Servin and Hospital, 2002; Wan et al., 2004). Because only two elite rye genotypes of one genetic pool from one breeding company were analysed, only a very reduced proportion of the intra–species genetic variability was captured by this approach (Geiger and Miedaner, 2009). This is in contrast with the standard procedure of SNP array generation for many crops, in which a preferably large set of diverse genotypes is sequenced to detect and cover the diversity in this set and therefore in the species itself (Ganal et al., 2012). This approach was used for example to develop the 600k Rye SNP Chip (Bauer et al., 2017), the 90k Wheat SNP Chip (Wang et al., 2014) or the 50k Barley SNP Chip (Bayer et al., 2017). This will result in a reduced quality of the SNP resource generated in this study in terms of failed values when applied to distinct and genetically distant genotypes. But, on the other hand there is an elevated average quality when applied to closely related material, most prominently to genotypes of the same genetic pool of the same breeding company (Heslot et al., 2013; Moragues et al., 2010). It was also not possible to identify and select for polymorphisms using allele frequencies due to the use of just two genotypes. This quality parameter is important to avoid the representation of rare alleles within a genotyping array, which would on average result in overall reduced polymorphic information content (PIC) (Botstein et al., 1980) of molecular markers representing these loci. This conclusion can be easily obtained, given the PIC formula, which has a global maximum of 0.5 for bi-allelic markers when exactly half of the analysed individuals exhibit one allele and the other half the other one. Deviations from this 1:1 ratio directly lower the PIC value. Another quality parameter is the pure number of polymorphisms in combination with the coverage of the genome. As reported for some crops, larger genomic stretches so called haplo-blocks, are accumulated in current breeding populations (Gupta et al., 2005; Voss-Fels et al., 2019), due to genetic bottlenecks introduced by domestication and breeding (Ersoz et al., 2007). For many applications in plant breeding, one or a small set of molecular markers addressing each single haplo-block would be sufficient, as a higher number will presumably increase the genotyping costs, reduce the statistical power of most methods due to multiple testing, and

will add little relevant information. This concept is referred to as the haplotype tagging SNPs (Sebastiani et al., 2003). A valid proxy for the genome wide average extend of such haplo-blocks is the decay of linkage disequilibrium (LD) (Hill and Robertson, 1968). As reported for inbreeding species like wheat or barley, large proportions of the genome exhibit genetic linkage, so in comparison to the rather large genome size of the afore mentioned species, only a limited number of molecular markers would be needed in elite material for most breeding purposes (Moragues et al., 2010; Thomson, 2014). Unfortunately, for rye a fast decline in LD was reported, most likely due to the outcrossing mating scheme (Li et al., 2011), resulting in a higher number of polymorphisms needed to access dense genomic information for breeding. This number is most presumably not reached by the genotyping platform generated within this study.

To sum up, the sequence resources of rye, i.e. additional single nucleotide polymorphisms, were developed, evaluated in terms of quality parameters, and combined with predicted positional information, resulting in a resource with high value for downstream applications, like the development of drought stress specific molecular markers.

4.9 Prospects on pan-genomes and the virtual linear order of genes

The collinearity of genes and homology in genomic sequences found within the *Poaceae*, combined in a resource termed genome zipper (Martis et al., 2013; Mayer et al., 2011), is a valuable resource for several applications, like fine mapping of genes (Hofmann et al., 2013; Lüpken et al., 2013) or for predicting genomic positions as shown in this study. This similarity in the genome structure of related species may be employed and combined with novel approaches of sequence data representations, the so-called pan-genome (Tettelin et al., 2005; Vernikos et al., 2015). With the progress in sequencing technologies and the constant cost reduction for data generation (Schwarze et al., 2018), high quality, whole genome sequencing of large cohorts of genotypes will presumably be achievable in the long term (Monat et al., 2019). The current state of few high-quality reference genomes per species with little connection between this reference sequences will be insufficient to unlock the full potential of this data source (The Computational Pan-Genomics Consortium, 2016). Therefore, the use of so-called pan-genomes was proposed (Morgante et al., 2007), to i) integrate multi-genotype sequence information in one resource and ii) to find and separate the so called core-genome from the dispensable genome. These terms refer to the genes and sequence

stretches, which are represented in all individuals of a given species or clade as the core genome, whereas the genes found only within a small proportion of the individuals are termed dispersible genome. From a bioinformatic point of view, this approach extends the currently linear, one-dimensional space of genomic sequence information of one individual genotype to a two-dimensional plane, in which multiple genomes are aligned side by side, covering multiple individuals, highlighting intra-species variance and diversity in the background of stable genomic stretches common to all members of the given species. (The Computational Pan-Genomics Consortium, 2016).

The source of genomic collinearity and synteny, which is exploited by the genome zipper approach, is the common ancestry and relatedness of multiple species (Martis et al., 2013; Mayer et al., 2011). Like the genome zipper approach, this homology may be applied to the pan-genome concept to extend and transfer information from one species to multiple others (Vernikos et al., 2015), resulting in a multi-species pan-genome, which is common for prokaryotic species (Medini et al., 2005). The resulting data source would integrate large scale inter-species genomic rearrangements into the currently discussed intra-species pan-genome concept also for higher plants and crops, resulting in a hierarchical, multi-level search space that extends the graph based representation presented by The Computational Pan-Genomics Consortium (2016). This envisioned resource would put additional value on a large proportion of currently species-specific results and in general is not restricted to the *Poaceae*. QTLs identified in one species could be transferred to other species, functional gene annotation precision may be greatly increased, and molecular markers linked to genes with causal effects in desirable traits could be transferred into other species (Georges, 1997; Parsons et al., 2005). To sum up, the extension and combination of genome zippers and pan-genomes, as two existing concepts, could result in a resource, which greatly promotes the exchange and utilization of biological information and data across species.

5. Summary

Rye (*Secale cereale* L.) is known for its outstanding abiotic stress tolerance within the family of the *Poacea*, especially for its drought tolerance. Nevertheless, there is little knowledge on the molecular mechanisms which confer this elevated tolerance compared to related crop species like barley or wheat. In this study, we generated a significant amount of sequence information to fill this gap by sequencing the transcriptome of two German elite rye genotypes grown in a well-watered and a drought stress inducing variant to get specific access to drought induced transcripts. Furthermore, the sequencing protocol allowed for transcript quantification as a basis for a differential gene expression analysis. Based on this, candidate genes were identified. In a next step, sequence data were mined for SNPs. By applying a stringent quality filtering using only polymorphisms within previously identified drought related transcripts and applying a synteny based *in-silico* mapping to optimize genome coverage, a set of SNPs was defined and used to generate a mid-plex genotyping array. This array was used for QTL detection based on phenotypic data for six agronomic traits estimated in two years and three locations per year on a population of 271 individuals. Out of the 384 SNPs on the array, 254 SNPs were integrated into the genetic map of rye and 21 QTLs for plant height at three different developmental stages, heading date, spikes per plant and thousand kernel weight were detected.

Using the sequencing data, a deeper insight into drought stress tolerance mechanisms in elite rye was achieved on the molecular level. Furthermore, the generated sequence information as well as the mid-plex genotyping array may be employed by scientists or breeding companies for multiple purposes, like QTL and candidate gene discovery, genetic distance analysis or marker assisted selection.

6. Literature

Adam, D., Simonsen, V., Loeschcke, V., 1987. Allozyme variation in rye, Secale cereale L. Theor. Appl. Genet. 74, 560–565.

Agarwal, M., Shrivastava, N., Padh, H., 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep. 27, 617–631.

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

Agrawal, G.K., Yamazaki, M., Kobayashi, M., Hirochika, R., Miyao, A., Hirochika, H., 2001. Screening of the rice viviparous mutants generated by endogenous retrotransposon Tos17 insertion. Tagging of a zeaxanthin epoxidase gene and a novel OsTATC gene. Plant Physiol. 125, 1248–1257. https://doi.org/10.1104/pp.125.3.1248

Ahuja, I., de Vos, R.C.H., Bones, A.M., Hall, R.D., 2010. Plant molecular stress responses face climate change. Trends Plant Sci. 15, 664–674. https://doi.org/10.1016/j.tplants.2010.08.002

Akmaev, V.R., Wang, C.J., 2004. Correction of sequence-based artifacts in serial analysis of gene expression. Bioinformatics 20, 1254–1263. https://doi.org/10.1093/bioinformatics/bth077

Al-Karaki, G., McMichael, B., Zak, J., 2004. Field response of wheat to arbuscular mycorrhizal fungi and drought stress. Mycorrhiza 14, 263–269.

Almeida, J.S., 2002. Predictive non-linear modeling of complex data by artificial neural networks. Curr. Opin. Biotechnol. 13, 72–76.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. BASIC LOCAL ALIGNMENT SEARCH TOOL. J. Mol. Biol. 215, 403–410. https://doi.org/10.1006/jmbi.1990.9999

Amato, M., Basso, B., Celano, G., Bitella, G., Morelli, G., Rossi, R., 2008. In situ detection of tree root distribution and biomass by multi-electrode resistivity imaging. Tree Physiol. 28, 1441–1448.

Annicchiarico, P., Nazzicari, N., Pecetti, L., Romani, M., Ferrari, B., Wei, Y., Brummer, E.C., 2017. GBS-based genomic selection for pea grain yield under severe terminal drought. Plant Genome 10.

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., Pozniak, C.J., Choulet, F., Distelfeld, A., Poland, J., 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361, eaar7191.

Araus, J.L., Slafer, G.A., Reynolds, M.P., Royo, C., 2002. Plant breeding and drought in C-3 cereals: What should we breed for? Ann. Bot. 89, 925–940. https://doi.org/10.1093/aob/mcf049

Arneth, A., Barbosa, H., Benton, T., Calvin, K., Calvo, E., Connors, S., Cowie, A., Davin, E., Denton, F., van Diemen, R., 2019. IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems. Summ. Policy Mak. Geneva Intergov. Panel Clim. Change IPCC.

Atkinson, J.A., Pound, M.P., Bennett, M.J., Wells, D.M., 2019. Uncovering the hidden half of plants using new advances in root phenotyping. Curr. Opin. Biotechnol. 55, 1–8.

Audran, C., Borel, C., Frey, A., Sotta, B., Meyer, C., Simonneau, T., Marion-Poll, A., 1998. Expression studies of the zeaxanthin epoxidase gene in Nicotiana plumbaginifolia. Plant Physiol. 118, 1021–1028. https://doi.org/10.1104/pp.118.3.1021

Augé, R.M., 2001. Water relations, drought and vesicular-arbuscular mycorrhizal symbiosis. Mycorrhiza 11, 3–42.
Baggio, A., 2005. Wireless sensor networks in precision agriculture. Presented at the ACM Workshop on Real-World Wireless Sensor Networks (REALWSN 2005), Stockholm, Sweden, Citeseer.
Bandillo, N., Raghavan, C., Muyco, P.A., Sevilla, M.A.L., Lobina, I.T., Dilla-Ermita, C.J., Tung, C.-W., McCouch, S., Thomson, M., Mauleon, R., 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6, 11.
Barbazuk, W.B., Fu, Y., McGinnis, K.M., 2008. Genome-wide analyses of alternative splicing in plants: opportunities and challenges. Genome Res.
Bartels, D., Sunkar, R., 2005. Drought and salt tolerance in plants. Crit. Rev. Plant Sci. 24, 23–58. https://doi.org/10.1080/07352680590910410
Bauer, E., Schmutzer, T., Barilar, I., Mascher, M., Gundlach, H., Martis, M.M., Twardziok, S.O., Hackauf, B., Gordillo, A., Wilde, P., 2017. Towards a whole-genome sequence for rye (Secale cereale L.). Plant J. 89, 853–869.
Bayer, M.M., Rapazote-Flores, P., Ganal, M., Hedley, P.E., Macaulay, M., Plieske, J., Ramsay, L., Russell, J., Shaw, P.D., Thomas, W., 2017. Development and evaluation of a barley 50k iSelect SNP array. Front. Plant Sci. 8, 1792.
Bentley, D.R., 2006. Whole-genome re-sequencing. Curr. Opin. Genet. Dev. 16, 545–552. https://doi.org/10.1016/j.gde.2006.10.009
Bernier, J., Kumar, A., Ramaiah, V., Spaner, D., Atlin, G., 2007. A large-effect QTL for grain yield under reproductive-stage drought stress in upland rice. Crop Sci. 47, 507–518. https://doi.org/10.2135/cropsci2006.07.0495
Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B., Sehabiague, P., Makumbi, D., Magorokosho, C., Oikeh, S., 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci. 55, 154–163.
Bishop, C.M., 2006. Pattern recognition and machine learning. springer.
Blum, A., 2005. Drought resistance, water-use efficiency, and yield potential - are they compatible, dissonant, or mutually exclusive? Aust. J. Agric. Res. 56, 1159–1168. https://doi.org/10.1071/ar05069
Blum, A., 1983. Genetic and physiological relationships in plant breeding for drought resistance. Agric. Water Manag. 7, 195–205.
Bolibok, H., Gruszczyńska, A., Hromada-Judycka, A., Rakoczy-Trojanowska, M., 2007. The identification of QTLs associated with the in vitro response of rye (Secale cereale L.). Cell. Mol. Biol. Lett. 12, 523.
Bolibok, H., Rakoczy-Trojanowska, M., Wyrzykowska, M., Radecka, M., Orczyk, W., 2006. Identification of microsatellite markers in the rye genome. Cell. Mol. Biol. Lett. 11, 291–298.
Bolibok-Bragoszewska, H., Heller-Uszynska, K., Wenzl, P., Uszynski, G., Kilian, A., Rakoczy-Trojanowska, M., 2009. DArT markers for the rye genome - genetic diversity and mapping. Bmc Genomics 10. https://doi.org/10.1186/1471-2164-10-578
Boomsma, C.R., Vyn, T.J., 2008. Maize drought tolerance: potential improvements through arbuscular mycorrhizal symbiosis? Field Crops Res. 108, 14–31.
Börner, A., Korzun, V., Voylokov, A.V., Weber, W.E., 1999. Detection of quantitative trait loci on chromosome 5R of rye (Secale cereale L.). Theor. Appl. Genet. 98, 1087–1090. https://doi.org/10.1007/s001220051171

Börner, A., Korzun, V., Voylokov, A.V., Worland, A.J., Weber, W.E., 2000. Genetic mapping of quantitative trait loci in rye (Secale cereale L.). Euphytica 116, 203–209. https://doi.org/10.1023/a:1004052505692

Borrell, A.K., Jordan, D.R., Mullet, J., Klein, P., 2013. Drought tolerant plants produced by modification of the stay-green stgx locus. Google Patents.

Botstein, D., White, R.L., Skolnick, M., Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., Shendure, J., 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. 31, 1119.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Capell, T., Bassie, L., Christou, P., 2004. Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. Proc. Natl. Acad. Sci. 101, 9909–9914.

Carter, J.V., Brenner, M.L., 1985. Plant growth regulators and low temperature stress, in: Hormonal Regulation of Development III. Springer, pp. 418–443.

Cattivelli, L., Rizza, F., Badeck, F.-W., Mazzucotelli, E., Mastrangelo, A.M., Francia, E., 2008. Drought tolerance improvement in crop plants: An integrated view from breeding to genomics. Field Crop Res 105. https://doi.org/10.1016/j.fcr.2007.07.004

Cavanagh, C., Morell, M., Mackay, I., Powell, W., 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr. Opin. Plant Biol. 11, 215–221.

Ceccarelli, S., Grando, S., 1996. Drought as a challenge for the plant breeder. Plant Growth Regul. 20, 149–155.

Chang, J.H., Tong, L., 2012. Mitochondrial poly (A) polymerase and polyadenylation. Biochim. Biophys. Acta BBA-Gene Regul. Mech. 1819, 992–997.

Chaves, M.M., Maroco, J.P., Pereira, J.S., 2003. Understanding plant responses to drought - from genes to the whole plant. Funct. Plant Biol. 30, 239–264. https://doi.org/10.1071/fp02076

Chikmawati, T., Skovmand, B., Gustafson, J.P., 2005. Phylogenetic relationships among Secale species revealed by amplified fragment length polymorphisms. Genome 48, 792–801.

Chilamakuri, C.S.R., Lorenz, S., Madoui, M.-A., Vodák, D., Sun, J., Hovig, E., Myklebost, O., Meza-Zepeda, L.A., 2014. Performance comparison of four exome capture systems for deep sequencing. BMC Genomics 15, 449.

Chubukov, V., Mukhopadhyay, A., Petzold, C.J., Keasling, J.D., Martín, H.G., 2016. Synthetic and systems biology for microbial production of commodity chemicals. Npj Syst. Biol. Appl. 2, 1–11.

Coca, M.A., Almoguera, C., Jordano, J., 1994. Expression of sunflower low-molecular-weight heat-shock proteins during embryogenesis and persistence after germination: localization and possible functional implications. Plant Mol. Biol. 25, 479–492.

Collard, B.C., Mackill, D.J., 2007. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. B Biol. Sci. 363, 557–572.

Collobert, R., Weston, J., 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning. Presented at the Proceedings of the 25th international conference on Machine learning, ACM, pp. 160–167.
Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., 2012. Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat. Genet 44. https://doi.org/10.1038/ng.2447
Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.
Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A., Totterdell, I.J., 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408, 184.
Devos, K.M., Atkinson, M.D., Chinoy, C.N., Francis, H.A., Harcourt, R.L., Koebner, R.M.D., Liu, C.J., Masojć, P., Xie, D.X., Gale, M.D., 1993. Chromosomal rearrangements in the rye genome relative to that of wheat. Theor. Appl. Genet. 85, 673–680.
Devos, K.M., Atkinson, M.D., Chinoy, C.N., Liu, C.J., Gale, M.D., 1992. RFLP-based genetic map of the homoeologous group 3 chromosomes of wheat and rye. Theor. Appl. Genet. 83, 931–939.
Dolezel, J., Kubalakova, M., Paux, E., Bartos, J., Feuillet, C., 2007. Chromosome-based genomics in the cereals. Chromosome Res 15. https://doi.org/10.1007/s10577-006-1106-x
dos Reis, S.P., Lima, A.M., de Souza, C.R.B., 2012. Recent Molecular Advances on Downstream Plant Responses to Abiotic Stress. Int. J. Mol. Sci. 13, 8628–8647. https://doi.org/10.3390/ijms13078628
Draffehn, A.M., Li, L., Krezdorn, N., Ding, J., Lübeck, J., Strahwald, J., Muktar, M.S., Walkemeier, B., Rotter, B., Gebhardt, C., 2013. Comparative transcript profiling by SuperSAGE identifies novel candidate genes for controlling potato quantitative resistance to late blight not compromised by late maturity. Front. Plant Sci. 4. https://doi.org/10.3389/fpls.2013.00423
Eathington, S.R., Crosbie, T.M., Edwards, M.D., Reiter, R.S., Bull, J.K., 2007. Molecular markers in a commercial breeding program. Crop Sci. 47, S-154-S-163.
Edwards, D., Batley, J., Snowdon, R.J., 2013. Accessing complex crop genomes with next-generation sequencing. Theor. Appl. Genet. 126, 1–11. https://doi.org/10.1007/s00122-012-1964-x
Edwards, J., McCouch, S., 2007. Molecular markers for use in plant molecular breeding and germplasm evaluation. Marker-Assist. Sel.-Curr. Status Future Perspect. Crops Livest. For. Fish-Rome Food Agric. Organ. U. N. FAO 29–49.
Ersoz, E.S., Yu, J., Buckler, E.S., 2007. Applications of linkage disequilibrium and association mapping in crop plants, in: Genomics-Assisted Crop Improvement. Springer, pp. 97–119.
Falke, K.C., Wilde, P., Wortmann, H., Geiger, H.H., Miedaner, T., 2009. Identification of genomic regions carrying QTL for agronomic and quality traits in rye (Secale cereale) introgression libraries. Plant Breed. 128, 615–623.
FAO, 2019. http://faostat.fao.org [WWW Document].
Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., Basra, S.M.A., 2009. Plant drought stress: effects, mechanisms and management. Agron Sust Dev 29. https://doi.org/10.1051/agro:2008021
Fehr, W.R., 1987. Heritability. Princ. Cultiv. Dev. 1, 95–105.

Fernández-Cirelli, A., Arumí, J.L., Rivera, D., Boochs, P.W., 2009. Environmental effects of irrigation in arid and semi-arid regions. Chil. J. Agric. Res. 69 2009 Nr SUPPL 1 69, 27–40.
Feuj, R., Heidari, B., Dadkhodaie, A., 2018. Association of DREB Genes and Microsatellite Markers Linked to NAX2 with Salt Tolerance in CIMMYT-derived Triticale, Wheat and Rye Genotypes. J. Crop Sci. Biotechnol. 21, 309–319.
Flavell, R.B., Bennett, M.D., Smith, J.B., Smith, D.B., 1974. Genome size and proportion of repeated nucleotide-sequence DNA in plants. Biochem. Genet. 12, 257–269. https://doi.org/10.1007/bf00485947
Fleury, D., Jefferies, S., Kuchel, H., Langridge, P., 2010. Genetic and genomic tools to improve drought tolerance in wheat. J. Exp. Bot. 61, 3211–3222. https://doi.org/10.1093/jxb/erq152
Fowler, D.B., Limin, A.E., 1987. Exploitable genetic variability for cold tolerance in commercially grown cereals. Can J Plant Sci 67, 278–278.
Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K., Shinozaki, K., 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr. Opin. Plant Biol. 9, 436–442. https://doi.org/10.1016/j.pbi.2006.05.014
Gallego, F.J., Benito, C., 1997. Genetic control of aluminium tolerance in rye (Secale cereale L.). Theor. Appl. Genet. 95, 393–399.
Ganal, M.W., Polley, A., Graner, E.-M., Plieske, J., Wieseke, R., Luerssen, H., Durstewitz, G., 2012. Large SNP arrays for genotyping in crop plants. J. Biosci. 37, 821–828.
Gauderman, W.J., 2002. Sample size requirements for association studies of gene-gene interaction. Am. J. Epidemiol. 155, 478–484.
Gawroński, P., Pawełkowicz, M., Tofil, K., Uszyński, G., Sharifova, S., Ahluwalia, S., Tyrka, M., Wędzony, M., Kilian, A., Bolibok-Brągoszewska, H., 2016. DArT markers effectively target gene space in the rye genome. Front. Plant Sci. 7, 1600.
Geiger, H.H., Miedaner, T., 2009. Rye Breeding, in: Carena, M.J. (Ed.), Cereals. Springer, New York, pp. 157–181.
Geiger, H.H., Miedaner, T., 1999. Hybrid rye and heterosis. Genet. Exploit. Heterosis Crops Crop Sci. Soc. Am. Madison Wis. USA 439–450.
Geiger, H.H., Schnell, F.W., 1970. Cytoplasmic Male Sterility in Rye (Secale cereale L.) 1. Crop Sci. 10, 590–593.
Georges, M., 1997. QTL mapping to QTL cloning: mice to the rescue. Genome Res. 7, 663–665.
Goddard, M.E., Hayes, B.J., 2007. Genomic selection. J. Anim. Breed. Genet. 124, 323–330.
Goddijn, O.J., Verwoerd, T.C., Voogd, E., Krutwagen, R.W., De Graff, P., Poels, J., van Dun, K., Ponstein, A.S., Damm, B., Pen, J., 1997. Inhibition of trehalase activity enhances trehalose accumulation in transgenic plants. Plant Physiol. 113, 181–190.
Goldberg, D.E., 1989. Genetic Algorithms in Search. Optimization and Machine Learning. Addism1—Wesley. Read. MA.
Golldack, D., Luking, I., Yang, O., 2011. Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. Plant Cell Rep. 30, 1383–1391. https://doi.org/10.1007/s00299-011-1068-0
Gollner, M.J., Wagentristl, H., Liebhard, P., Friedel, J.K., 2011. Yield and arbuscular mycorrhiza of winter rye in a 40-year fertilisation trial. Agron. Sustain. Dev. 31, 373–378.

Gous, P.W., Hasjim, J., Franckowiak, J., Fox, G.P., Gilbert, R.G., 2013. Barley genotype expressing “stay-green”-like characteristics maintains starch quality of the grain during water stress condition. J. Cereal Sci. 58, 414–419. https://doi.org/10.1016/j.jcs.2013.08.002

Govaerts, R., 2009. World checklist of selected plant families. Royal Botanic Gardens.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.D., Chen, Z.H., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644-U130. https://doi.org/10.1038/nbt.1883

Gupta, K., Dey, A., Gupta, B., 2013. Plant polyamines in abiotic stress responses. Acta Physiol. Plant. 35, 2015–2036. https://doi.org/10.1007/s11738-013-1239-4

Gupta, P.K., Rustgi, S., Kulwal, P.L., 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol. Biol. 57, 461–485.

Hackauf, B., Haffke, S., Fromme, F.J., Roux, S.R., Kusterer, B., Musmann, D., Kilian, A., Miedaner, T., 2017a. QTL mapping and comparative genome analysis of agronomic traits including grain yield in winter rye. Theor. Appl. Genet. 130, 1801–1817.

Hackauf, B., Haffke, S., Fromme, F.J., Roux, S.R., Kusterer, B., Musmann, D., Kilian, A., Miedaner, T., 2017b. QTL mapping and comparative genome analysis of agronomic traits including grain yield in winter rye. Theor. Appl. Genet. 130, 1801–1817.

Hackauf, B., Korzun, V., Wortmann, H., Wilde, P., Wehling, P., 2012. Development of conserved ortholog set markers linked to the restorer gene Rfp1 in rye. Mol. Breed. 30, 1507–1518.

Hackauf, B., Rudd, S., van der Voort, J.R., Miedaner, T., Wehling, P., 2009. Comparative mapping of DNA sequences in rye (Secale cereale L.) in relation to the rice genome. Theor. Appl. Genet. 118, 371–384. https://doi.org/10.1007/s00122-008-0906-0

Hackauf, B., Wehling, P., 2003. Development of microsatellite markers in rye: map construction. Plant Breed. Seed Sci. 48.

Hackauf, B., Wehling, P., 2002. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. Plant Breed. 121, 17–25.

Hadiarto, T., Tran, L.S.P., 2011. Progress studies of drought-responsive genes in rice. Plant Cell Rep. 30, 297–310. https://doi.org/10.1007/s00299-010-0956-z

Haffke, S., 2015a. Quantitative-trait loci (QTL) mapping of important agronomical traits of the grain and biomass production in winter rye (Secale cereale L.).

Haffke, S., 2015b. Quantitative-trait loci (QTL) mapping of important agronomical traits of the grain and biomass production in winter rye (Secale cereale L.).

Hartung, F., Schiemann, J., 2014. Precise plant breeding using new genome editing techniques: opportunities, safety and regulation in the EU. Plant J. 78, 742–752.

Haseneyer, G., Schmutzer, T., Seidel, M., Zhou, R., Mascher, M., Schon, C.C., Taudien, S., Scholz, U., Stein, N., Mayer, K.F.X., Bauer, E., 2011. From RNA-seq to large-scale genotyping-genomics resources for rye (Secale cereale L.). BMC Plant Biol. 11, (28 September 2011).

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Raffeld, M., 2001. Gene-expression profiles in hereditary breast cancer. N. Engl. J. Med. 344, 539–548.

Heffner, E.L., Sorrells, M.E., Jannink, J.-L., 2009. Genomic selection for crop improvement. Crop Sci. 49, 1–12.

Hepting, L., 1978. Analyse eines 7 x 7-Sortendialles zur Ermittlung geeigneten Ausgangsmaterials fur die Hybridzüchtung bei Roggen. Z Pflanzenzücht. 80, 188–197.
Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Galvez, S., Schaaf, S., Jouve, N., Simkova, H., Valarik, M., Dolezel, J., Mayer, K.F.X., 2012. Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. Plant J. 69, 377–386. https://doi.org/10.1111/j.1365-313X.2011.04808.x
Heslot, N., Rutkoski, J., Poland, J., Jannink, J.-L., Sorrells, M.E., 2013. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLoS One 8, e74612.
Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226–231.
Hirschhorn, J.N., Daly, M.J., 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95.
Hoekstra, F.A., Golovina, E.A., Buitink, J., 2001. Mechanisms of plant desiccation tolerance. Trends Plant Sci. 6, 431–438.
Hoffmann, B., 2008. Alteration of drought tolerance of winter wheat caused by translocation of rye chromosome segment 1RS. Cereal Res. Commun. 36, 269–278.
Hofmann, K., Silvar, C., Casas, A.M., Herz, M., Büttner, B., Gracia, M.P., Contreras-Moreira, B., Wallwork, H., Igartua, E., Schweizer, G., 2013. Fine mapping of the Rrs1 resistance locus against scald in two large populations derived from Spanish barley landraces. Theor. Appl. Genet. 126, 3091–3102.
Hou, F.Y., Huang, J., Yu, S.L., Zhang, H.S., 2007. The 6-phosphogluconate dehydrogenase genes are responsive to abiotic stresses in rice. J. Integr. Plant Biol. 49, 655–663. https://doi.org/10.1111/j.1672-9072.2007.00460.x
Hubner, M., Wilde, P., Schmiedchen, B., Dopierala, P., Gowda, M., Reif, J.C., Miedaner, T., 2013. Hybrid rye performance under natural drought stress in Europe. Theor. Appl. Genet. 126, 475–482. https://doi.org/10.1007/s00122-012-1994-4
Hussain, S.S., 2006. Molecular breeding for abiotic stress tolerance. drought perspective. Proc. Pak. Acad. Sci. 43, 189–210.
Illumina, I. 9885 T.C.D., 2013. Assay Design Tool [WWW Document].
Ingram, J., Bartels, D., 1996. The molecular basis of dehydration tolerance in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47, 377–403. https://doi.org/10.1146/annurev.arplant.47.1.377
Iuchi, S., Kobayashi, M., Taji, T., Naramoto, M., Seki, M., Kato, T., Tabata, S., Kakubari, Y., Yamaguchi-Shinozaki, K., Shinozaki, K., 2001. Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in Arabidopsis. Plant J. 27, 325–333. https://doi.org/10.1046/j.1365-313x.2001.01096.x
Jacobs, D.R., Marquart, L., Slavin, J., Kushi, L.H., 1998a. Whole-grain intake and cancer: An expanded review and meta-analysis. Nutr. Cancer 30, 85–96.
Jacobs, D.R., Meyer, K.A., Kushi, L.H., Folsom, A.R., 1998b. Whole-grain intake may reduce the risk of ischemic heart disease death in postmenopausal women: the Iowa Women's Health Study. Am. J. Clin. Nutr. 68, 248–257.
Jain, M., Moharana, K.C., Shankar, R., Kumari, R., Garg, R., 2014. Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. Plant Biotechnol. J. 12, 253–264. https://doi.org/10.1111/pbi.12133

Jain, S.M., Brar, D.S., Ahloowalia, B.S., 2010. Molecular techniques in crop improvement. Springer.
Jannink, J.-L., Lorenz, A.J., Iwata, H., 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9, 166–177.
Jansen, R.C., 1994. Controlling the type I and type II errors in mapping quantitative trait loci. Genetics 138, 871–881.
Jansen, R.C., 1993. Interval mapping of multiple quantitative trait loci. Genetics 135, 205–211.
Jansen, R.C., Stam, P., 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136, 1447–1455.
Javot, H., Lauvergeat, V., Santoni, V., Martin-Laurent, F., Güçlü, J., Vinh, J., Heyes, J., Franck, K.I., Schäffner, A.R., Bouchez, D., 2003. Role of a single aquaporin isoform in root water uptake. Plant Cell 15, 509–522.
Jiang, H.W., Li, M.R., Liang, N.B., Yan, H.B., Wei, Y.L., Xu, X., Liu, J.F., Xu, Z., Chen, F., Wu, G.J., 2007. Molecular cloning and function analysis of the stay green gene in rice. Plant J. 52, 197–209. https://doi.org/10.1111/j.1365-313X.2007.003221.x
Kakumanu, A., Ambavaram, M.M., Klumas, C., Krishnan, A., Batlang, U., Myers, E., Grene, R., Pereira, A., 2012. Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant Physiol. 160, 846–867.
Kang, S.-T., Kwak, M., Kim, H.-K., Choung, M.-G., Han, W.-Y., Baek, I.-Y., Kim, M.Y., Van, K., Lee, S.-H., 2009. Population-specific QTLs and their different epistatic interactions for pod dehiscence in soybean [Glycine max (L.) Merr.]. Euphytica 166, 15.
Kerepesi, I., Galiba, G., 2000. Osmotic and salt stress-induced alteration in soluble carbohydrate content in wheat seedlings. Crop Sci. 40, 482–487.
Keyvan, S., 2010. The effects of drought stress on yield, relative water content, proline, soluble carbohydrates and chlorophyll of bread wheat cultivars. J Anim Plant Sci 8, 1051–1060.
Khalvati, M.A., Hu, Y., Mozafar, A., Schmidhalter, U., 2005. Quantification of water uptake by arbuscular mycorrhizal hyphae and its significance for leaf growth, water relations, and gas exchange of barley subjected to drought stress. Plant Biol. 7, 706–712.
Kiyosue, T., Yamaguchishinozaki, K., Shinozaki, K., 1994. CLONING OF CDNAS FOR GENES THAT ARE EARLY-RESPONSIVE TO DEHYDRATION STRESS (ERDS) IN ARABIDOPSIS-THALIANA L - IDENTIFICATION OF 3 ERDS AS HSP COGNATE GENES. Plant Mol. Biol. 25, 791–798. https://doi.org/10.1007/bf00028874
Ko, J.M., Seo, B.B., Suh, D.Y., Do, G.S., Park, D.S., Kwack, Y.H., 2002. Production of a new wheat line possessing the 1BL.1RS wheat-rye translocation derived from Korean rye cultivar Paldanghomil. Theor. Appl. Genet. 104, 171–176. https://doi.org/10.1007/s00122-001-0783-2
Kofler, R., Bartoš, J., Gong, L., Stift, G., Suchánková, P., Šimková, H., Berenyi, M., Burg, K., Doležel, J., Lelley, T., 2008. Development of microsatellite markers specific for the short arm of rye (Secale cereale L.) chromosome 1. Theor. Appl. Genet. 117, 915–926.
Korzun, V., Börner, A., Melz, G., 1996. RFLP mapping of the dwarfing (Ddw1) and hairy peduncle (Hp) genes on chromosome 5 of rye (Secale cereale L.). Theor. Appl. Genet. 92, 1073–1077.
Korzun, V., Malyshev, S., Voylokov, A.V., Börner, A., 2001. A genetic map of rye (Secale cereale L.) combining RFLP, isozyme, protein, microsatellite and gene loci. Theor. Appl. Genet. 102, 709–717.

Kosambi, D.D., 1943. THE ESTIMATION OF MAP DISTANCES FROM RECOMBINATION VALUES. Ann. Eugen. 12, 172–175. https://doi.org/10.1111/j.1469-1809.1943.tb02321.x
Kosmas, S.A., Argyrokastritis, A., Loukas, M.G., Eliopoulos, E., Tsakas, S., Kaltsikes, P.J., 2006. Isolation and characterization of drought-related trehalose 6-phosphate-synthase gene from cultivated cotton (Gossypium hirsutum L.). Planta 223, 329–339.
Kotak, S., Larkindale, J., Lee, U., von Koskull-Döring, P., Vierling, E., Scharf, K.-D., 2007. Complexity of the heat stress response in plants. Curr. Opin. Plant Biol. 10, 310–316.
Krasensky, J., Jonak, C., 2012. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. J. Exp. Bot. 63, 1593–1608. https://doi.org/10.1093/jxb/err460
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Presented at the Advances in neural information processing systems, pp. 1097–1105.
Kruger, N.J., von Schaewen, A., 2003. The oxidative pentose phosphate pathway: structure and organisation. Curr. Opin. Plant Biol. 6, 236–246. https://doi.org/10.1016/s1369-5266(03)00039-6
Kumar, S., Banks, T.W., Cloutier, S., 2012. SNP discovery through next-generation sequencing and its applications. Int. J. Plant Genomics 2012.
Lander, E.S., Botstein, D., 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121, 185–199.
Lapiński, Stojalowski, 2003. Occurence and genetic identity of male sterility-inducing cytoplasm in rye (secale spp.). Plant Breed. Seed Sci. 48.
Lázaro, R., Latorre, M.A., Medel, P., Gracia, M., Mateos, G.G., 2004. Feeding regimen and enzyme supplementation to rye-based diets for broilers. Poult. Sci. 83, 152–160.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324.
Lehnert, H., Serfling, A., Enders, M., Friedt, W., Ordon, F., 2017. Genetics of mycorrhizal symbiosis in winter wheat (Triticum aestivum). New Phytol. 215, 779–791.
Levitt, J., 1972. Responses of plants to environmental stresses. Responses Plants Environ. Stress. 697pp.
Li, J., Zhou, R., Endo, T.R., Stein, N., 2018. High-throughput development of SSR marker candidates and their chromosomal assignment in rye (Secale cereale L.). Plant Breed. 137, 561–572.
Li, Y., Haseneyer, G., Schon, C.C., Ankerst, D., Korzun, V., Wilde, P., Bauer, E., 2011. High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (Secale cereale L.) genes involved in frost response. BMC Plant Biol. 11, (10 January 2011).
Liang, M., Li, Z., Chen, T., Zeng, J., 2014. Integrative data analysis of multi-platform cancer data with a multimodal deep learning approach. IEEE/ACM Trans. Comput. Biol. Bioinform. 12, 928–937.
Liu, B.H., 1997. Statistical genomics: linkage, mapping, and QTL analysis pp. 163-214. CRC press.
Liu, S., Manson, J.E., Stampfer, M.J., Hu, F.B., Giovannucci, E., Colditz, G.A., Hennekens, C.H., Willett, W.C., 2000. A prospective study of whole-grain intake and risk of type 2 diabetes mellitus in US women. Am. J. Public Health 90, 1409.
Liu, X., Dong, X., Leskovar, D.I., 2016. Ground penetrating radar for underground sensing in agriculture: a review. Int. Agrophysics 30, 533–543.

Liukkonen, K.-H., Katina, K., Wilhelmsson, A., Myllymaki, O., Lampi, A.-M., Kariluoto, S., Piironen, V., Heinonen, S.-M., Nurmi, T., Adlercreutz, H., 2003. Process-induced changes on bioactive compounds in whole grain rye. Proc. Nutr. Soc. 62, 117–122.
Loarce, Y., Gallego, R., Ferrer, E., 1996. A comparative analysis of the genetic relationships between rye cultivars using RFLP and RAPD markers. Euphytica 88, 107–115.
Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate Trends and Global Crop Production Since 1980. Science 333, 616–620. https://doi.org/10.1126/science.1204531
Loraine, A.E., McCormick, S., Estrada, A., Patel, K., Qin, P., 2013. RNA-seq of Arabidopsis pollen uncovers novel transcription and alternative splicing. Plant Physiol. 162, 1092–1109.
Lu, T., Zhu, C., Lu, G., Guo, Y., Zhou, Y., Zhang, Z., Zhao, Y., Li, W., Lu, Y., Tang, W., 2012. Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice. BMC Genomics 13, 721.
Ludlow, M.M., 1989. Strategies of response to water stress, Structural and functional responses to environmental stresses: water shortage. SPB Academic Publishing bv, The Hague, Netherlands.
Lüpken, T., Stein, N., Perovic, D., Habekuß, A., Krämer, I., Hähnel, U., Steuernagel, B., Scholz, U., Zhou, R., Ariyadasa, R., 2013. Genomics-based high-resolution mapping of the BaMMV/BaYMV resistance gene rym11 in barley (Hordeum vulgare L.). Theor. Appl. Genet. 126, 1201–1212.
Lusser, M., Parisi, C., Plan, D., Rodríguez-Cerezo, E., 2012. Deployment of new biotechnologies in plant breeding. Nat. Biotechnol. 30, 231.
Ma, J., Yu, M.K., Fong, S., Ono, K., Sage, E., Demchak, B., Sharan, R., Ideker, T., 2018. Using deep learning to model the hierarchical structure and function of a cell. Nat. Methods 15, 290.
Ma, X.-F., Wanous, M.K., Houchins, K., Milla, M.R., Goicoechea, P.G., Wang, Z., Xie, M., Gustafson, J.P., 2001. Molecular linkage mapping in rye (Secale cereale L.). Theor. Appl. Genet. 102, 517–523.
Mackay, T.F., Stone, E.A., Ayroles, J.F., 2009. The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10, 565.
Madej, L.J., 1996. Worldwide trends in rye growing and breeding. Vortraege Fuer Pflanzenzuechtung Ger.
Mafakheri, A., Siosemardeh, A.F., Bahramnejad, B., Struik, P.C., Sohrabi, Y., 2010. Effect of drought stress on yield, proline and chlorophyll contents in three chickpea cultivars. Aust. J. Crop Sci. 4, 580.
Magwene, P.M., Willis, J.H., Kelly, J.K., 2011. The statistics of bulk segregant analysis using next generation sequencing. PLoS Comput. Biol. 7, e1002255.
Mahajan, S., Tuteja, N., 2005. Cold, salinity and drought stresses: an overview. Arch. Biochem. Biophys. 444, 139–158.
Martis, M.M., Zhou, R., Haseneyer, G., Schmutzer, T., Vrana, J., Kubalakova, M., Konig, S., Kugler, K.G., Scholz, U., Hackauf, B., Korzun, V., Schon, C.C., Dolezel, J., Bauer, E., Mayer, K.F., Stein, N., 2013. Reticulate evolution of the rye genome. Plant Cell 25, 3685–98. https://doi.org/10.1105/tpc.113.114553
Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J., 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature 544, 427.

Masojæ, P., Milczarski, P., 2005. Mapping QTLs for α amylase activity in rye grain. J Appl Genet 46, 115–123.
Masojć, P., Banek-Tabor, A., Milczarski, P., Twardowska, M., 2007. QTLs for resistance to preharvest sprouting in rye (Secale cereale L.). J. Appl. Genet. 48, 211–217.
Masojć, P., Milczarski, P., 2009. Relationship between QTLs for preharvest sprouting and alpha-amylase activity in rye grain. Mol. Breed. 23, 75–84.
Masojć, P., Milczarski, P., Kruszona, P., 2017. Comparative analysis of genetic architectures for nine developmental traits of rye. J. Appl. Genet. 58, 297–305.
Matos, M., Pinto-Carnide, O., BENIT02, C., 2001. Phylogenetic relationships among Portuguese rye based on isozyme, RAPD and ISSR markers. Hereditas 134, 229–236.
Matsumura, H., Ito, A., Saitoh, H., Winter, P., Kahl, G., Reuter, M., Kruger, D.H., Terauchi, R., 2005. SuperSAGE. Cell. Microbiol. 7, 11–18. https://doi.org/10.1111/j.1462-5822.2004.00478.x
Matsumura, H., Yoshida, K., Luo, S.J., Kimura, E., Fujibe, T., Albertyn, Z., Barrero, R.A., Kruger, D.H., Kahl, G., Schroth, G.P., Terauchi, R., 2010. High-Throughput SuperSAGE for Digital Gene Expression Analysis of Multiple Samples Using Next Generation Sequencing. Plos One 5. https://doi.org/10.1371/journal.pone.0012010
Matsumura, Molina, Krüger, Terauchi, Kahl, 2012. DeepSuperSAGE: High-Throughput Transcriptome Sequencing with Now- and Next-Generation Sequencing Technologies, in: Matthias Harbers & Guenter Kahl (Ed.), Tag-Based Next Generation Sequencing.
Maurel, C., Chrispeels, M.J., 2001. Aquaporins. A molecular entry into plant water relations. Plant Physiol. 125, 135–138.
Mayer, K.F.X., Martis, M., Hedley, P.E., Simkova, H., Liu, H., Morris, J.A., Steuernagel, B., Taudien, S., Roessner, S., Gundlach, H., Kubalakova, M., Suchankova, P., Murat, F., Felder, M., Nussbaumer, T., Graner, A., Salse, J., Endo, T., Sakai, H., Tanaka, T., Itoh, T., Sato, K., Platzer, M., Matsumoto, T., Scholz, U., Dolezel, J., Waugh, R., Stein, N., 2011. Unlocking the Barley Genome by Chromosomal and Comparative Genomics. Plant Cell 23, 1249–1263. https://doi.org/10.1105/tpc.110.082537
Mayer, K.F.X., Taudien, S., Martis, M., Simkova, H., Suchankova, P., Gundlach, H., Wicker, T., Petzold, A., Felder, M., Steuernagel, B., Scholz, U., Graner, A., Platzer, M., Dolezel, J., Stein, N., 2009. Gene Content and Virtual Gene Order of Barley Chromosome 1H. Plant Physiol. 151, 496–505. https://doi.org/10.1104/pp.109.142612
McElroy, J.S., 2014. Vavilovian mimicry: Nikolai Vavilov and his little-known impact on weed science. Weed Sci. 62, 207–216.
Medini, D., Donati, C., Tettelin, H., Masignani, V., Rappuoli, R., 2005. The microbial pan-genome. Curr. Opin. Genet. Dev. 15, 589–594.
Mester, D.I., Ronin, Y.I., Hu, Y., Peng, J., Nevo, E., Korol, A.B., 2003. Efficient multipoint mapping: making use of dominant repulsion-phase markers. Theor. Appl. Genet. 107, 1102–1112.
Michael, T.P., VanBuren, R., 2015. Progress, challenges and the future of crop genomes. Curr. Opin. Plant Biol. 24, 71–81.
Miedaner, T., Haffke, S., Siekmann, D., Fromme, F.J., Roux, S.R., Hackauf, B., 2018. Dynamic quantitative trait loci (QTL) for plant height predict biomass yield in hybrid rye (Secale cereale L.). Biomass Bioenergy 115, 10–18.
Miedaner, T., Hubner, M., Korzun, V., Schmiedchen, B., Bauer, E., Haseneyer, G., Wilde, P., Reif, J.C., 2012a. Genetic architecture of complex agronomic traits examined in two testcross populations of rye (Secale cereale L.). Bmc Genomics 13. https://doi.org/10.1186/1471-2164-13-706

Miedaner, T., Koch, S., Seggl, A., Schmiedchen, B., Wilde, P., 2012b. Quantitative genetic parameters for selection of biomass yield in hybrid rye. Plant Breed. 131, 100–103. https://doi.org/10.1111/j.1439-0523.2011.01909.x
Miernyk, J.A., 1999. Protein folding in the plant cell. Plant Physiol. 121, 695–703.
Milczarski, P., Banek-Tabor, A., Lebiecka, K., Stojałowski, S., Myśków, B., Masojć, P., 2007. New genetic map of rye composed of PCR-based molecular markers and its alignment with the reference map of the DS2× RXL10 intercross. J. Appl. Genet. 48, 11–24.
Milczarski, P., Bolibok-Bragoszewska, H., Myskow, B., Stojalowski, S., Heller-Uszynska, K., Goralska, M., Bragoszewski, P., Uszynski, G., Kilian, A., Rakoczy-Trojanowska, M., 2011. A High Density Consensus Map of Rye (Secale cereale L.) Based on DArT Markers. Plos One 6. https://doi.org/10.1371/journal.pone.0028495
Mir, R.R., Zaman-Allah, M., Sreenivasulu, N., Trethowan, R., Varshney, R.K., 2012. Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. Theor. Appl. Genet. 125, 625–645. https://doi.org/10.1007/s00122-012-1904-9
Mirouze, M., Paszkowski, J., 2011. Epigenetic contribution to stress adaptation in plants. Curr. Opin. Plant Biol. 14, 267–274.
Misra, A.K., 2014. Climate change and challenges of water and food security. Int. J. Sustain. Built Environ. 3, 153–165.
Mittler, R., Blumwald, E., 2010. Genetic engineering for modern agriculture: challenges and perspectives. Annu. Rev. Plant Biol. 61, 443–462.
Molina, C., Rotter, B., Horres, R., Udupa, S.M., Besser, B., Bellarmino, L., Baum, M., Matsumura, H., Terauchi, R., Kahl, G., Winter, P., 2008. SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. Bmc Genomics 9, 28. https://doi.org/10.1186/1471-2164-9-553
Monat, C., Schreiber, M., Stein, N., Mascher, M., 2019. Prospects of pan-genomics in barley. Theor. Appl. Genet. 132, 785–796.
Montenegro, J.D., Golicz, A.A., Bayer, P.E., Hurgobin, B., Lee, H., Chan, C.-K.K., Visendi, P., Lai, K., Doležel, J., Batley, J., 2017. The pangenome of hexaploid bread wheat. Plant J. 90, 1007–1013.
Montilla-Bascón, G., Rubiales, D., Hebelstrup, K.H., Mandon, J., Harren, F.J., Cristescu, S.M., Mur, L.A., Prats, E., 2017. Reduced nitric oxide levels during drought stress promote drought tolerance in barley and is associated with elevated polyamine biosynthesis. Sci. Rep. 7, 1–15.
Moose, S.P., Mumm, R.H., 2008. Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol. 147, 969–977.
Moragues, M., Comadran, J., Waugh, R., Milne, I., Flavell, A.J., Russell, J.R., 2010. Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. Theor. Appl. Genet. 120, 1525–1534.
Morgante, M., De Paoli, E., Radovic, S., 2007. Transposable elements and the plant pan-genomes. Curr. Opin. Plant Biol. 10, 149–155.
Murakeözy, É.P., Nagy, Z., Duhazé, C., Bouchereau, A., Tuba, Z., 2003. Seasonal changes in the levels of compatible osmolytes in three halophytic species of inland saline vegetation in Hungary. J. Plant Physiol. 160, 395–401.
Mwadzingeni, L., Shimelis, H., Dube, E., Laing, M.D., Tsilo, T.J., 2016. Breeding wheat for drought tolerance: Progress and technologies. J. Integr. Agric. 15, 935–943.

Myśków, B., Hanek, M., Banek-Tabor, A., Maciorowski, R., Stojałowski, S., 2014. The application of high-density genetic maps of rye for the detection of QTLs controlling morphological traits. J. Appl. Genet. 55, 15–26.
Myśków, B., Stojałowski, S., 2016. Bidirectional selective genotyping approach for the identification of quantitative trait loci controlling earliness per se in winter rye (Secale cereale L.). J. Appl. Genet. 57, 45–50.
Myśków, B., Stojałowski, S., Łań, A., Bolibok-Brągoszewska, H., Rakoczy-Trojanowska, M., Kilian, A., 2012. Detection of the quantitative trait loci for α-amylase activity on a high-density genetic map of rye and comparison of their localization to loci controlling preharvest sprouting and earliness. Mol. Breed. 30, 367–376.
Nagel, K.A., Putz, A., Gilmer, F., Heinz, K., Fischbach, A., Pfeifer, J., Faget, M., Blossfeld, S., Ernst, M., Dimaki, C., 2012. GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons. Funct. Plant Biol. 39, 891–904.
Nelson, G.C., Valin, H., Sands, R.D., Havlík, P., Ahammad, H., Deryng, D., Elliott, J., Fujimori, S., Hasegawa, T., Heyhoe, E., 2014. Climate change effects on agriculture: Economic responses to biophysical shocks. Proc. Natl. Acad. Sci. 111, 3274–3279.
Nguyen, H.T., Babu, R.C., Blum, A., 1997. Breeding for drought resistance in rice: physiology and molecular genetics considerations. Crop Sci. 37, 1426–1434.
O'Neill, R., Snowdon, R., Köhler, W., 2003. Population genetics: aspects of biodiversity, in: Progress in Botany. Springer, pp. 115–137.
Osborne, T., Rose, G., Wheeler, T., 2013. Variation in the global-scale impacts of climate change on crop productivity due to climate model uncertainty and adaptation. Agric. For. Meteorol. 170, 183–194.
Ozsolak, F., Milos, P.M., 2011. RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet. 12, 87.
Paez-Garcia, A., Motes, C.M., Scheible, W.-R., Chen, R., Blancaflor, E.B., Monteros, M.J., 2015. Root traits and phenotyping strategies for plant improvement. Plants 4, 334–355.
Pandey, R., Singh, B., Nair, T.V.R., 2005. Impact of arbuscular-mycorrhizal fungi on phosphorus efficiency of wheat, rye, and triticale. J. Plant Nutr. 28, 1867–1876.
Parsons, C.A., Mroczkowski, H.J., McGuigan, F.E., Albagha, O.M., Manolagas, S., Reid, D.M., Ralston, S.H., Reis, R.J.S., 2005. Interspecies synteny mapping identifies a quantitative trait locus for bone mineral density on human chromosome Xp22. Hum. Mol. Genet. 14, 3141–3148.
Passioura, J.B., 1983. Roots and drought resistance, in: Developments in Agricultural and Managed Forest Ecology. Elsevier, pp. 265–280.
Patience, J.F., Bedford, M.R., Classen, H.L., Inborr, J., 1992. The effect of dietary enzyme supplementation of rye-and barley-based diets on digestion and subsequent performance in weanling pigs. Can. J. Anim. Sci. 72, 97–105.
Pearson, K., 1894. On the dissection of asymmetrical frequency curves. Philos. Trans. R. Soc. A 185, 71–110.
Pereira, M.A., Jacobs Jr, D.R., Pins, J.J., Raatz, S.K., Gross, M.D., Slavin, J.L., Seaquist, E.R., 2002. Effect of whole grains on insulin sensitivity in overweight hyperinsulinemic adults. Am. J. Clin. Nutr. 75, 848–855.
Perrier, X., Jacquemoud-Collet, J.P., 2006. DARwin software.
Pfeifer, M., Martis, M., Asp, T., Mayer, K.F.X., Lubberstedt, T., Byrne, S., Frei, U., Studer, B., 2013. The Perennial Ryegrass GenomeZipper: Targeted Use of Genome Resources for

Comparative Grass Genomics. Plant Physiol. 161, 571–582. https://doi.org/10.1104/pp.112.207282
Piepho, H.-P., 2000a. Optimal marker density for interval mapping in a backcross population. Heredity 84, 437–440.
Piepho, H.-P., 2000b. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156, 2043–2050.
Pilon-Smits, E.A., Ebskamp, M.J., Paul, M.J., Jeuken, M.J., Weisbeek, P.J., Smeekens, S.C., 1995. Improved performance of transgenic fructan-accumulating tobacco under drought stress. Plant Physiol. 107, 125–130.
Plaschke, J., Börner, A., Xie, D.X., Koebner, R.M.D., Schlegel, R., Gale, M.D., 1993. RFLP mapping of genes affecting plant height and growth habit in rye. Theor. Appl. Genet. 85, 1049–1054.
Preece, C., Livarda, A., Christin, P.-A., Wallace, M., Martin, G., Charles, M., Jones, G., Rees, M., Osborne, C.P., 2017. How did the domestication of Fertile Crescent grain crops increase their yields? Funct. Ecol. 31, 387–397.
Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904.
Qi, X., Stam, P., Lindhout, P., 1996. Comparison and integration of four barley genetic maps. Genome 39, 379–394.
Rafalski, J.A., 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci. 162, 329–333.
Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., Van Ex, F., Pasha, A., 2018. The transcriptional landscape of polyploid wheat. Science 361, eaar6089.
Raney, J.A., Reynolds, D.J., Elzinga, D.B., Page, J., Udall, J.A., Jellen, E.N., Bonfacio, A., Fairbanks, D.J., Maughan, P.J., 2014. Transcriptome analysis of drought induced stress in Chenopodium quinoa. Am. J. Plant Sci. 5, 338–357.
Reguera, M., Peleg, Z., Blumwald, E., 2012. Targeting metabolic pathways for genetic engineering abiotic stress-tolerance in crops. Biochim. Biophys. Acta BBA-Gene Regul. Mech. 1819, 186–194.
Reynolds, M.P., Hays, D., Chapman, S., 2010. Breeding for adaptation to heat and drought stress. Clim. Change Crop Prod. 1, 71–91.
Ripley, B.D., 2001. The R project in statistical computing. MSOR Connect. Newsl. LTSN Maths Stats Netw. 1, 23–25.
Ritossa, F., 1962. A new puffing pattern induced by temperature shock and DNP in Drosophila. Experientia 18, 571–573.
Robinson, A.J., Love, C.G., Batley, J., Barker, G., Edwards, D., 2004. Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics 20, 1475–1476.
Roux, S.R., Wortmann, H., Schlathölter, M., 2010. Rye (Secale cereale L.) for biogas production-breeding capability. J. Für Kult. 62, 173–182.
Ruiz, C., Asins, M., 2003. Comparison between Poncirus and Citrus genetic linkage maps. Theor. Appl. Genet. 106, 826–836.
Rusk, N., 2009. Cheap third-generation sequencing. Nat. Methods 6, 244.
Saal, B., Wricke, G., 2002. Clustering of amplified fragment length polymorphism markers in a linkage map of rye. Plant Breed. 121, 117–123.
Saal, B., Wricke, G., 1999. Development of simple sequence repeat markers in rye (Secale cereale L.). Genome 42, 964.

Salinas, J., Benito, C., 1985. Chromosomal location of malate dehydrogenase structural genes in rye (Secale cereale L.). Z. Pflanzenzuchtung J. Plant Breed.
Salinas, J., Benito, C., 1983. Chromosomal location of genes controlling 6-phosphogluconate dehydrogenase, glucose-6-phosphate dehydrogenase and glutamate dehydrogenase isozymes in cultivated rye. Euphytica 32, 783–790.
Schiessl, S., Huettel, B., Kuehn, D., Reinhardt, R., Snowdon, R., 2017. Post-polyploidisation morphotype diversification associates with gene copy number variation. Sci. Rep. 7, 41845.
Schlegel, R., Melz, G., Korzun, V., 1998. Genes, marker and linkage data of rye (Secale cereale L): 5th updated inventory. Euphytica 101, 23–67. https://doi.org/10.1023/a:1018351008889
Schnell, F.W., 1982. A synoptic study of the methods and categories of plant breeding. Z Pflanzenzücht. 89, 1–18.
Schreiber, M., Himmelbach, A., Börner, A., Mascher, M., 2018. Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. Evol. Appl.
Schuster, S.C., 2007. Next-generation sequencing transforms today's biology. Nat. Methods 5, 16.
Schwarze, K., Buchanan, J., Taylor, J.C., Wordsworth, S., 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet. Med. 20, 1122.
Sebastiani, P., Lazarus, R., Weiss, S.T., Kunkel, L.M., Kohane, I.S., Ramoni, M.F., 2003. Minimal haplotype tagging. Proc. Natl. Acad. Sci. 100, 9900–9905.
Serraj, R., Hash, C.T., Rizvi, S.M.H., Sharma, A., Yadav, R.S., Bidinger, F.R., 2005. Recent advances in marker-assisted selection for drought tolerance in pearl millet. Plant Prod. Sci. 8, 334–337. https://doi.org/10.1626/pps.8.334
Servin, B., Hospital, F., 2002. Optimal positioning of markers to control genetic background in marker-assisted backcrossing. J. Hered. 93, 214–217.
Shagin, D.A., Rebrikov, D.V., Kozhemyako, V.B., Altshuler, I.M., Shcheglov, A.S., Zhulidov, P.A., Bogdanova, E.A., Staroverov, D.B., Rasskazov, V.A., Lukyanov, S., 2002. A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Res. 12, 1935–1942. https://doi.org/10.1101/gr.547002
Shanahan, P.W., Binley, A., Whalley, W.R., Watts, C.W., 2015. The use of electromagnetic induction to monitor changes in soil moisture profiles beneath different wheat genotypes. Soil Sci. Soc. Am. J. 79, 459–466.
Shang, H.-Y., Wei, Y.-M., Wang, X.-R., Zheng, Y.-L., 2006. Genetic diversity and phylogenetic relationships in the rye genus Secale L.(rye) based on Secale cereale microsatellite markers. Genet. Mol. Biol. 29, 685–691.
Shikha, M., Kanika, A., Rao, A.R., Mallikarjuna, M.G., Gupta, H.S., Nepolean, T., 2017. Genomic selection for drought tolerance using genome-wide SNPs in maize. Front. Plant Sci. 8, 550.
Shinozaki, K., Yamaguchi-Shinozaki, K., 2007. Gene networks involved in drought stress response and tolerance. J. Exp. Bot. 58, 221–227.
Silva, S., Pinto, G., Dias, M.C., Correia, C.M., Moutinho-Pereira, J., Pinto-Carnide, O., Santos, C., 2012. Aluminium long-term stress differently affects photosynthesis in rye genotypes. Plant Physiol. Biochem. 54, 105–112.

Silvar, C., Martis, M.M., Nussbaumer, T., Haag, N., Rauser, R., Keilwagen, J., Korzun, V., Mayer, K.F., Ordon, F., Perovic, D., 2015. Assessing the barley genome zipper and genomic resources for breeding purposes. Plant Genome 8.

Silvar, C., Perovic, D., Nussbaumer, T., Spannagl, M., Usadel, B., Casas, A., Igartua, E., Ordon, F., 2013. Towards Positional Isolation of Three Quantitative Trait Loci Conferring Resistance to Powdery Mildew in Two Spanish Barley Landraces. Plos One 8, 12. https://doi.org/10.1371/journal.pone.0067336

Sinclair, T.R., 2011. Challenges in breeding for yield increase for drought. Trends Plant Sci. 16, 289–293. https://doi.org/10.1016/j.tplants.2011.02.008

Sivamani, E., Bahieldin, A., Wraith, J.M., Al-Niemi, T., Dyer, W.E., Ho, T.-H.D., Qu, R., 2000. Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley HVA1 gene. Plant Sci. 155, 1–9.

Slafer, G.A., Calderini, D.F., Miralles, D.J., 1996. Yield components and compensation in wheat: opportunities for further increasing yield potential. Increasing Yield Potential Wheat Break. Barriers 101–133.

Snowdon, R.J., Iniguez Luy, F.L., 2012. Potential to improve oilseed rape and canola breeding in the genomics era. Plant Breed. 131, 351–360.

Souci, S.W., Fachmann, W., Kraut, H., 2003. Lebensmitteltabelle für die Praxis. Wissenschaftliche Verlagsgesellschaft, Stuttgart.

Spannagl, M., Martis, M.M., Pfeifer, M., Nussbaumer, T., Mayer, K.F.X., 2013. Analysing complex Triticeae genomes - concepts and strategies. Plant Methods 9, 9. https://doi.org/10.1186/1746-4811-9-35

Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C.-T., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. 5, e1000734.

Srayeddin, I., Doussan, C., 2009. Estimation of the spatial variability of root water uptake of maize and sorghum at the field scale by electrical resistivity tomography. Plant Soil 319, 185–207.

Stange, M., Utz, H.F., Schrag, T.A., Melchinger, A.E., Würschum, T., 2013. High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations. Theor. Appl. Genet. 126, 2563–2574.

Subbarao, G.V., Nam, N.H., Chauhan, Y.S., Johansen, C., 2000. Osmotic adjustment, water relations and carbohydrate remobilization in pigeonpea under water deficits. J. Plant Physiol. 157, 651–659.

Tanksley, S.D., 1983. Molecular markers in plant breeding. Plant Mol. Biol. Report. 1, 3–8.

Targońska, M., Bolibok-Brągoszewska, H., Rakoczy-Trojanowska, M., 2016. Assessment of genetic diversity in Secale cereale based on SSR markers. Plant Mol. Biol. Report. 34, 37–51.

Templer, S.E., Ammon, A., Pscheidt, D., Ciobotea, O., Schuy, C., McCollum, C., Sonnewald, U., Hanemann, A., Förster, J., Ordon, F., 2017. Metabolite profiling of barley flag leaves under drought and combined heat and drought stress reveals metabolic QTLs for metabolites associated with antioxidant defense. J. Exp. Bot. 68, 1697–1713.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial “pan-genome.” Proc. Natl. Acad. Sci. 102, 13950–13955.

The Computational Pan-Genomics Consortium, 2016. Computational pan-genomics: status, promises and challenges. Brief. Bioinform. 19, 118–135.
Thomé, O.W., 1885. Flora von Deutschland. Österr. Schweiz.
Thomson, M.J., 2014. High-throughput SNP genotyping to accelerate crop improvement. Plant Breed. Biotechnol. 2, 195–212.
Toker, C., Canci, H., Yildirim, T., 2007. Evaluation of perennial wild Cicer species for drought resistance. Genet. Resour. Crop Evol. 54, 1781–1786. https://doi.org/10.1007/s10722-006-9197-y
Trick, M., Adamski, N.M., Mugford, S.G., Jiang, C.C., Febrer, M., Uauy, C., 2012. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biol. 12, 17. https://doi.org/10.1186/1471-2229-12-14
Turner, N.C., 1986. CROP WATER DEFICITS - A DECADE OF PROGRESS. Adv. Agron. 39, 1–51. https://doi.org/10.1016/s0065-2113(08)60464-2
Tyerman, S.D., Niemietz, C.M., Bramley, H., 2002. Plant aquaporins: multifunctional water and solute channels with expanding roles. Plant Cell Environ. 25, 173–194.
Uppu, S., Krishna, A., Gopalan, R.P., 2016a. Towards Deep Learning in genome-Wide Association Interaction studies., in: PACIS. p. 20.
Uppu, S., Krishna, A., Gopalan, R.P., 2016b. A deep learning approach to detect SNP interactions. JSW 11, 965–975.
Usadel, B., 2019. PlabiPD Sequenced Plant Genomes.
Van Ooijen, J.W., 2009. MapQTL ® 6. Kyazma B.V., Wageningen, Netherlands.
Van Ooijen, J.W., 2006. JoinMap ® 4. Kyazma B.V., Wageningen, Netherlands.
Varshney, Bansal, K.C., Aggarwal, P.K., Datta, S.K., Craufurd, P.Q., 2011. Agricultural biotechnology for crop improvement in a variable climate: hope or hype? Trends Plant Sci. 16, 363–371. https://doi.org/10.1016/j.tplants.2011.03.004
Varshney, Beier, U., Khlestkina, E.K., Kota, R., Korzun, V., Graner, A., Börner, A., 2007a. Single nucleotide polymorphisms in rye (Secale cereale L.): discovery, frequency, and applications for genome mapping and diversity studies. Theor. Appl. Genet. 114, 1105–1116.
Varshney, Mahendar, T., Aggarwal, R.K., Börner, A., 2007b. Genic molecular markers in plants: development and applications, in: Genomics-Assisted Crop Improvement. Springer, pp. 13–29.
Vaseva, Grigorova, B.S., Simova-Stoilova, L.P., Demirevska, K.N., Feller, U., 2010. Abscisic acid and late embryogenesis abundant protein profile changes in winter wheat under progressive drought stress. Plant Biol. 12, 698–707. https://doi.org/10.1111/j.1438-8677.2009.00269.x
Vernikos, G., Medini, D., Riley, D.R., Tettelin, H., 2015. Ten years of pan-genome analyses. Curr. Opin. Microbiol. 23, 148–154.
Vierling, E., 1991. The roles of heat shock proteins in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 42, 579–620.
Voorrips, R.E., 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93, 77–78.
Voss-Fels, K.P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., Chen, T.-W., Zetzsche, H., Seddig, S., Baig, M.M., 2019. Breeding improves wheat productivity under contrasting agrochemical input levels. Nat. Plants 1.
Wan, J.X., Griffiths, R., Ying, J.F., McCourt, P., Huang, Y.F., 2009. Development of Drought-Tolerant Canola (Brassica napus L.) through Genetic Modulation of ABA-mediated

Stomatal Responses. Crop Sci. 49, 1539–1554. https://doi.org/10.2135/cropsci2008.09.0568
Wan, Q.-H., Wu, H., Fujihara, T., Fang, S.-G., 2004. Which genetic marker for which conservation genetics issue? Electrophoresis 25, 2165–2176.
Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M., Salvi, S., Milner, S.G., Cattivelli, L., 2014. Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol. J. 12, 787–796.
Wang, W., Vinocur, B., Shoseyov, O., Altman, A., 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci. 9, 244–252.
Wang, W.X., Vinocur, B., Altman, A., 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218, 1–14. https://doi.org/10.1007/s00425-003-1105-5
Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57.
Wardlaw, I.F., Willenbrink, J., 2000. Mobilization of fructan reserves and changes in enzyme activities in wheat stems correlate with water stress during kernel filling. New Phytol. 148, 413–422.
Wehling, P., Schmidt-Stohn, G., Wricke, G., 1985. Chromosomal location of esterase, peroxidase and phosphoglucomutase isozyme structural genes in cultivated rye (Secale cereale L.). Theor. Appl. Genet. 70, 377–382.
Wehner, G., Balko, C., Enders, M., Humbeck, K., Ordon, F., 2015. Identification of genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley. BMC Plant Biol 15. https://doi.org/10.1186/s12870-015-0524-3
Wehner, G., Balko, C., Humbeck, K., Zyprian, E., Ordon, F., 2016. Expression profiling of genes involved in drought stress and leaf senescence in juvenile barley. BMC Plant Biol. 16, 1–12. https://doi.org/10.1186/s12870-015-0701-4
Wenger, J.W., Schwartz, K., Sherlock, G., 2010. Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from Saccharomyces cerevisiae. PLoS Genet. 6, e1000942.
Whalley, W.R., Binley, A., Watts, C.W., Shanahan, P., Dodd, I.C., Ober, E.S., Ashton, R.W., Webster, C.P., White, R.P., Hawkesford, M.J., 2017. Methods to estimate changes in soil water for phenotyping root activity in the field. Plant Soil 415, 407–422.
Wingler, A., 2002. The function of trehalose biosynthesis in plants. Phytochemistry 60, 437–440.
Wolski, T., 1970. Studies on the inbreeding of rye. Genet. Pol. 11, 1–26.
Wricke, G., 2002. Two major genes for kernel weight in rye. Plant Breed. 121, 26–28.
Wricke, G., 1969. Untersuchungen zur Vererbung der Selbstfertilität beim Roggen (Secale cereale). Theor. Appl. Genet. 39, 371–378.
Würschum, T., 2012. Mapping QTL for agronomic traits in breeding populations. Theor. Appl. Genet. 125, 201–210. https://doi.org/10.1007/s00122-012-1887-6
Xu, Y., Crouch, J.H., 2008. Marker-assisted selection in plant breeding: from publications to practice. Crop Sci. 48, 391–407.
Yamaguchi, K., Takahashi, Y., Berberich, T., Imai, A., Takahashi, T., Michael, A.J., Kusano, T., 2007. A protective role for the polyamine spermine against drought stress in Arabidopsis. Biochem. Biophys. Res. Commun. 352, 486–490.

Yamaguchi-Shinozaki, K., Shinozaki, K., 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses, in: Annual Review of Plant Biology, Annual Review of Plant Biology. Annual Reviews, Palo Alto, pp. 781–803.
Yang, S.J., Vanderbeld, B., Wan, J.X., Huang, Y.F., 2010. Narrowing Down the Targets: Towards Successful Genetic Engineering of Drought-Tolerant Crops. Mol. Plant 3, 469–490. https://doi.org/10.1093/mp/ssq016
Young, B.D., Anderson, M.L.M., 1985. QUANTITATIVE ANALYSIS OF SOLUTION HYBRIDIZATION, Hames, B. D. And S. J. Higgins.
Yu, J., Holland, J.B., McMullen, M.D., Buckler, E.S., 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178, 539–551.
Zawada, A.M., Rogacev, K.S., Müller, S., Rotter, B., Winter, P., Fliser, D., Heine, G.H., 2014. Massive analysis of cDNA Ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. Epigenetics 9, 161–172.
Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. Presented at the European conference on computer vision, Springer, pp. 818–833.
Zeng, Z.-B., 1994. Precision mapping of quantitative trait loci. Genetics 136, 1457–1468.
Zhao, Y., Mette, M.F., Reif, J.C., 2015. Genomic selection in hybrid breeding. Plant Breed. 134, 1–10.
Zhu, J.K., 2002. Salt and drought stress signal transduction in plants. Annu. Rev. Plant Biol. 53, 247–273. https://doi.org/10.1146/annurev.arplant.53.091401.143329
Zhulidov, P.A., Bogdanova, E.A., Shcheglov, A.S., Shagina, I.A., Wagner, L.L., Khazpekov, G.L., Kozhemyako, V.V., Lukyanov, S.A., Shagin, D.A., 2005. A method for the preparation of normalized cDNA libraries enriched with full-length sequences. Russ. J. Bioorganic Chem. 31, 170–177. https://doi.org/10.1007/s11171-005-0023-7
Zhulidov, P.A., Bogdanova, E.A., Shcheglov, A.S., Vagner, L.L., Khaspekov, G.L., Kozhemyako, V.B., Matz, M.V., Meleshkevitch, E., Moroz, L.L., Lukyanov, S.A., Shagin, D.A., 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res. 32. https://doi.org/10.1093/nar/gnh031
Zohary, D., Hopf, M., Weiss, E., 2012. Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. Oxford University Press on Demand.
Zolla, G., Badri, D.V., Bakker, M.G., Manter, D.K., Vivanco, J.M., 2013. Soil microbiomes vary in their ability to confer drought tolerance to Arabidopsis. Appl. Soil Ecol. 68, 1–9.

7. Supplementary files

All supplementary files are accessible using the digital object identifier (DOI) 10.5073/20200803-174646

S1: 128,010 contigs as a result of a *de-novo* assembly of drought stress induced transcripts in rye (*Secale cereale* L.).

S2: Table of 19,844 single nucleotide polymorphisms (SNP), obtained drought stress induced transcripts of two rye (*Secale cereale* L.) genotypes.

S3: Table of expression level of drought induced transcripts with respect to two rye (*Secale cereale* L.) genotypes, two sampled tissues and two treatments.

S4: Contigs of *de-novo* assembled rye (*Secale cereale* L.) drought induced transcripts, which were filtered for differential expression and high-quality SNPs.

S5: Synteny based virtual map of drought induced rye (*Secale cereale* L.) transcripts.

S6: Gene Ontology analysis of drought induced rye (*Secale cereale* L.) transcripts.

S7: Genetic map of a bi-parental QTL mapping population in rye (*Secale cereale* L.), containing newly developed drought associated SNP markers.

8. Acknowledgments

This thesis would not have been possible without the support, patience and exchange with the people who have accompanied me over the years. I will try to thank as many people as possible at this point, even though it would go beyond the scope of this work to mention them all, unfortunately.

First of all, I would like to thank Prof. Dr. Frank Ordon for the opportunity to write this dissertation in his working group, and beyond this especially for the persistent support and motivation even in difficult phases of the work, which is exceptional in my eyes.

I would like to thank Prof. Dr. Klaus Pillen for supervising this doctorate and for his professional academic and administrative support.

Special thanks go to the entire working group of the RS Institute and many other JKI employees, who have accompanied me on my way with advice, ideas and also practical support and, moreover, have ensured a very cordial, familiar working atmosphere. In particular, I would like to thank Ilona Renneberg, Sandra Färber, Heike Lehnert, Jasmin Philippi, Edgar Schliephake and Ilona Krämer as well as Maike and René Grünwald.

Furthermore, I´m really grateful for the professional expertise and work the project partners contributed. Without this, the doctorate would not have been possible. I would like to make special mention here of Bernd Hackauf, Dörthe Siekmann, Ralf Horres and Barbara Kusterer.

I would also like to thank the employees and managers of the German Seed Alliance, NPZ and NPZ Innovation, who gave me the freedom to combine my career and my doctorate, constantly encouraged me to continue along the path I had chosen and also provided professional support. Thanks to Gunhild Leckband, Markus Wolf and Frank Wolter.

I would also like to thank Sebastian Köhnke, Stefan Unger, Anja Hühnlein and Ulrike Stahl for the intensive exchange with and support from the DV and IB departments at the JKI in Quedlinburg.

This work was carried out within the framework of the project "Entwicklung hochertragreicher Biomassesorten des Roggens unter Nutzung innovativer Transkriptom- und Metabolom-Analysetechniken" (FKZ: 22006210), which was gratefully funded the German Federal Ministry of Food and Agriculture.

Finally, I would like to thank my whole family - Nadine, Paula, Kjell, my grandmother and especially my grandfather and my mother. It is impossible to put into words in how many ways you always supported and encouraged me and how grateful I am for that.

9. Appendix

9.1 Position of 214 newly developed RSQ SNP markers

Linkage Group 1R		Linkage Group 2R		Linkage Group 3R	
Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)
RSQ_150	0.000	RSQ_487	0.000	RSQ_407	0.000
RSQ_271	23.322	RSQ_313	2.488	RSQ_004	2.761
RSQ_494	52.255	RSQ_057	10.686	RSQ_377	6.729
RSQ_267	55.920	RSQ_488	12.079	RSQ_252	14.312
RSQ_481	74.017	RSQ_103	15.125	RSQ_314	15.705
RSQ_482	76.128	RSQ_478	16.595	RSQ_312	17.923
RSQ_389	79.698	RSQ_208	17.495	RSQ_496	19.136
RSQ_386	82.073	RSQ_138	27.913	RSQ_369	34.648
RSQ_186	83.658	RSQ_144	28.382	RSQ_135	35.297
RSQ_094	86.354	RSQ_069	28.525	RSQ_412	37.927
RSQ_403	90.091	RSQ_292	30.453	RSQ_327	39.193
RSQ_108	93.792	RSQ_035	30.910	RSQ_461	43.378
RSQ_436	96.536	RSQ_090	31.200	RSQ_397	47.708
RSQ_499	97.845	RSQ_443	31.909	RSQ_497	50.742
RSQ_352	99.489	RSQ_428	32.661	RSQ_379	60.405
RSQ_305	101.509	RSQ_498	36.922	RSQ_257	69.054
RSQ_226	103.649	RSQ_167	38.063	RSQ_458	80.088
RSQ_102	105.612	RSQ_489	38.761		
RSQ_395	113.142	RSQ_493	39.962		
RSQ_296	115.637	RSQ_160	41.057		
RSQ_277	120.625	RSQ_089	43.276		
RSQ_396	129.390	RSQ_483	45.796		
RSQ_044	131.971	RSQ_230	48.111		
RSQ_337	134.173	RSQ_491	50.457		
RSQ_173	141.205	RSQ_063	51.856		
RSQ_092	146.778	RSQ_133	52.777		
RSQ_086	157.356	RSQ_475	55.486		
		RSQ_224	57.324		
		RSQ_289	65.108		
		RSQ_037	67.771		
		RSQ_225	68.335		
		RSQ_268	71.656		
		RSQ_276	74.544		
		RSQ_154	79.254		

Linkage Group 4R		Linkage Group 5R		Linkage Group 6R	
Marker	Position (cM)	Marker	Position (cM)	Marker	Position (cM)
RSQ_300	0.000	RSQ_487	0.000	RSQ_202	0.000
RSQ_263	1.221	RSQ_313	2.488	RSQ_453	9.599
RSQ_031	6.838	RSQ_057	10.686	RSQ_452	11.960
RSQ_156	8.490	RSQ_488	12.079	RSQ_307	16.383
RSQ_354	9.828	RSQ_103	15.125	RSQ_143	18.428
RSQ_078	15.146	RSQ_478	16.595	RSQ_359	20.142
RSQ_205	16.350	RSQ_208	17.495	RSQ_142	21.103
RSQ_228	16.974	RSQ_138	27.913	RSQ_480	22.395
RSQ_295	17.861	RSQ_144	28.382	RSQ_109	23.864
RSQ_254	18.901	RSQ_069	28.525	RSQ_056	24.726
RSQ_001	19.594	RSQ_292	30.453	RSQ_120	25.107
RSQ_385	20.504	RSQ_035	30.910	RSQ_074	25.894
RSQ_316	21.989	RSQ_090	31.200	RSQ_032	26.732
RSQ_420	22.847	RSQ_443	31.909	RSQ_051	27.440
RSQ_331	23.674	RSQ_428	32.661	RSQ_378	28.022
RSQ_023	24.086	RSQ_498	36.922	RSQ_161	29.258
RSQ_149	25.267	RSQ_167	38.063	RSQ_235	30.659
RSQ_096	26.665	RSQ_489	38.761	RSQ_187	31.494
RSQ_147	27.106	RSQ_493	39.962	RSQ_095	32.271
RSQ_136	28.030	RSQ_160	41.057	RSQ_070	36.135
RSQ_238	28.956	RSQ_089	43.276	RSQ_076	36.794
RSQ_041	29.402	RSQ_483	45.796	RSQ_280	39.271
RSQ_272	35.586	RSQ_230	48.111	RSQ_071	40.308
RSQ_022	36.334	RSQ_491	50.457	RSQ_033	41.166
RSQ_048	38.973	RSQ_063	51.856	RSQ_401	47.345
RSQ_170	39.250	RSQ_133	52.777	RSQ_465	55.383
RSQ_213	39.982	RSQ_475	55.486	RSQ_148	56.759
RSQ_381	40.963	RSQ_224	57.324	RSQ_239	59.885
RSQ_100	41.700	RSQ_289	65.108	RSQ_456	71.177
RSQ_304	42.388	RSQ_037	67.771	RSQ_175	79.329
RSQ_007	42.874	RSQ_225	68.335	RSQ_113	84.154
RSQ_374	44.012	RSQ_268	71.656	RSQ_279	92.170
RSQ_190	45.338	RSQ_276	74.544	RSQ_380	94.748
RSQ_391	51.066	RSQ_154	79.254	RSQ_207	97.157
RSQ_126	72.262			RSQ_188	118.535
RSQ_104	82.863			RSQ_348	121.345
RSQ_286	123.062			RSQ_180	130.792
RSQ_196	126.609			RSQ_302	133.382
RSQ_197	130.715			RSQ_181	141.244
				RSQ_464	144.110
				RSQ_151	148.717
				RSQ_084	154.084
				RSQ_046	164.514

Linkage Group 7R	
Marker	Position (cM)
RSQ_471	9.225
RSQ_155	54.238
RSQ_203	55.535
RSQ_204	66.677
RSQ_232	67.874
RSQ_365	69.298
RSQ_467	71.342
RSQ_473	72.857
RSQ_029	73.917
RSQ_065	76.506
RSQ_282	78.405
RSQ_310	79.634
RSQ_358	83.231
RSQ_093	91.696
RSQ_218	105.96
RSQ_212	115.064

9.2 Filtered Gene Ontology Mapping of contigs generated by sequencing of drought stress induced transcripts in rye.

SeqName	Hit-Desc
comp10001_c0_seq1_A	map kinase kinase kinase
comp1010_c0_seq1_A	protein os-9-like
comp1010_c0_seq7_A	protein os-9 precursor
comp10202_c0_seq2_A	protein-tyrosine sulfotransferase-like
comp10220_c0_seq1_A	ankyrin repeat and kh domain-containing protein 1-like
comp1022_c1_seq1_G	zeaxanthin epoxidase
comp10277_c0_seq1_A	protein serine threonine kinase
comp10287_c0_seq2_A	isoflavone reductase-like protein
comp1028_c0_seq1_A	map kinase
comp10338_c0_seq1_A	pyridoxal biosynthesis protein
comp10484_c0_seq1_A	peroxidase 21
comp104_c0_seq1_A	tpa: aminomethyltransferase
comp10512_c0_seq2_A	purple acid phosphatase-like protein
comp1058_c0_seq1_A	mlo-like protein 4-like
comp10655_c0_seq1_A	hypothetical protein TRIUR3_12561
comp10738_c0_seq1_A	predicted protein
comp10789_c0_seq1_A	crs2-associated factor mitochondrial
comp1085_c0_seq1_G	3-beta hydroxysteroid dehydrogenase isomerase family protein
comp1088_c0_seq2_A	catalase
comp11039_c0_seq1_A	glutamyl-trna amidotransferase subunit a-like
comp11039_c0_seq2_A	outer envelope protein mitochondrial-like
comp1108_c0_seq1_A	protein cobra-like
comp1109_c0_seq1_A	3 (2) -bisphosphate nucleotidase
comp110_c0_seq1_A	thioredoxin h-type
comp11310_c0_seq1_A	endo- -beta-glucanase
comp11325_c0_seq1_A	presenilin family protein
comp113_c0_seq2_G	serine hydroxymethyltransferase
comp11538_c0_seq1_A	chloroplastic mitochondrial-like
comp11553_c0_seq1_A	chaperone protein dnaj
comp11553_c0_seq2_A	chaperone protein dnaj
comp1155_c0_seq1_G	20 kda chloroplastic
comp11638_c0_seq1_A	transducin beta-like protein 3-like
comp11687_c0_seq1_A	chaperone protein chloroplastic-like
comp1184_c0_seq1_A	kinesin-like motor protein heavy chain
comp11988_c0_seq1_A	zeaxanthin epoxidase
comp11988_c0_seq4_A	zeaxanthin epoxidase
comp1201_c0_seq1_A	indole-3-glycerol phosphate chloroplastic-like
comp12030_c0_seq1_A	endo- -beta-glucanase
comp1215_c0_seq6_A	6-phosphogluconate decarboxylating
comp12283_c0_seq1_A	tpa: universal stress protein family protein

comp1231_c0_seq1_A	anther-specific protein
comp12345_c0_seq1_A	lon protease-like mitochondrial
comp12378_c0_seq1_A	cwf19-like protein 2-like
comp12405_c0_seq2_A	endoribonuclease dicer-like 1
comp1249_c0_seq1_A	protein phosphatase 2c
comp1249_c0_seq2_A	protein phosphatase 2c
comp12531_c0_seq1_A	plasma membrane intrinsic protein
comp1259_c0_seq1_G	cold acclimation protein cor413-pm1
comp1276_c0_seq2_G	heat shock factor-binding protein 1
comp12924_c0_seq1_A	isoflavone reductase-like protein irl
comp12981_c0_seq1_A	non-structural maintenance of chromosome element 4-like
comp129_c0_seq1_A	calreticulin precursor
comp13021_c0_seq1_A	receptor-like cytosolic serine threonine-protein kinase rbk1
comp13097_c0_seq2_A	macpf domain-containing protein cad1-like
comp1315_c0_seq1_A	thioredoxin h-type
comp1316_c0_seq1_A	porphobilinogen deaminase
comp13203_c0_seq1_A	heat stress transcription factor a-1-like
comp13213_c0_seq1_A	dna mismatch repair protein mlh1
comp13224_c0_seq3_A	dihydrodipicolinate reductase chloroplastic
comp13286_c0_seq1_A	er molecular chaperone
comp1329_c0_seq2_A	isoflavone reductase homolog irl-like
comp13306_c0_seq1_A	retrotransposon unclassified
comp1332_c0_seq1_A	bri1-kd interacting protein 112
comp13375_c0_seq2_A	peroxiredoxin-2e- chloroplastic
comp13431_c0_seq1_A	plasma membrane atpase 1-like
comp13468_c0_seq1_A	cipk-like protein 1
comp1348_c0_seq1_A	nadh dehydrogenase
comp1356_c0_seq1_A	skp1-like protein 1b
comp1356_c0_seq4_G	omega-3 fatty acid chloroplastic
comp13645_c0_seq1_A	synaptic vesicle 2-related protein
comp13689_c0_seq1_A	cleavage and polyadenylation specificity factor subunit 2-like
comp1368_c0_seq1_A	abscisic stress ripening
comp13811_c0_seq1_A	aba-responsive lea-like partial
comp1381_c0_seq1_A	gdp-l-galactose phosphorylase 2-like
comp1404_c0_seq2_A	glyceraldehyde-3-phosphate dehydrogenase chloroplastic-like
comp14071_c0_seq1_A	calcium-transporting atpase plasma membrane-type
comp14086_c0_seq1_A	transcription initiation factor iia subunit 2-like
comp14141_c0_seq1_A	laccase-12 13-like
comp14232_c0_seq1_A	protein arginine n-methyltransferase 5
comp14251_c0_seq2_A	dnaj homolog subfamily b member 4-like
comp142_c0_seq1_G	alanine glyoxylate aminotransferase

comp1443_c0_seq1_A	chlorophyll a-b binding protein cp24
comp14497_c0_seq1_A	laccase 17
comp1449_c0_seq1_A	cell wall invertase
comp14550_c0_seq1_A	phosphoinositide 3-kinase regulatory subunit 4-like
comp1456_c0_seq1_A	inositol -trisphosphate 5 6-kinase
comp1483_c0_seq1_A	zinc-binding alcohol dehydrogenase domain-containing protein 2-like
comp1495_c0_seq2_A	glutathione peroxidase
comp14970_c0_seq1_A	plasma membrane atpase 1
comp1497_c0_seq1_A	laccase lac12
comp15124_c0_seq1_A	mitogen-activated protein kinase kinase kinase 1-like
comp151_c0_seq1_A	QTLg3-1
comp1523_c0_seq1_A	gtp-binding nuclear protein ran-3-like
comp1537_c0_seq1_A	ferredoxin--nadp leaf isozyme
comp1571_c0_seq1_A	succinyl- ligase
comp1583_c0_seq3_G	1-deoxy-d-xylulose 5-phosphate reductoisomerase
comp1591_c0_seq1_A	ferredoxin-nadp oxidoreductase
comp15_c0_seq1_A	annexin p35
comp15_c0_seq2_A	annexin p35
comp16036_c0_seq1_G	3-oxoacyl-
comp1626_c0_seq1_A	plasma membrane h+-atpase
comp1641_c0_seq3_A	nad h oxidase
comp16495_c0_seq1_A	leucine-rich repeat receptor-like protein kinase family protein
comp16498_c0_seq1_A	nadh dehydrogenase complex assembly factor 6-like
comp16501_c0_seq1_A	vacuolar defense protein
comp1652_c0_seq1_A	aspartate chloroplastic-like
comp1652_c0_seq2_A	aspartate aminotransferase
comp1683_c0_seq1_G	glucan water dikinase
comp1690_c0_seq1_A	hypothetical protein TRIUR3_20539
comp1693_c0_seq1_A	peroxidase 16-like
comp169_c0_seq1_A	hydrophobic protein lti6a
comp169_c0_seq1_G	hydrophobic protein osr8
comp169_c0_seq2_A	hydrophobic protein lti6a
comp17046_c0_seq1_A	staphylococcal nuclease domain-containing protein 1-like
comp1718_c0_seq1_A	peroxisomal acyl-coenzyme a oxidase 1-like
comp17200_c0_seq3_G	chromatin structure-remodeling complex protein syd-like isoform x1
comp1726_c0_seq1_A	3-beta hydroxysteroid dehydrogenase isomerase family protein
comp17319_c0_seq1_A	elongator complex protein 1-like
comp17347_c0_seq1_A	abc transporter c family member 4-like
comp1738_c2_seq3_A	wd-40 repeat-containing protein msi1
comp17708_c0_seq1_A	indole-3-glycerol phosphate chloroplastic
comp17734_c0_seq3_A	chalcone synthase

comp1778_c0_seq1_A	splicing factor 3a subunit 3-like
comp17836_c0_seq2_A	nuclear pore complex protein nup85-like
comp1786_c0_seq1_A	v1 subunit a protein
comp1827_c0_seq1_A	oxygen-evolving enhancer protein chloroplastic
comp1830_c0_seq1_A	outer mitochondrial membrane protein porin
comp18605_c0_seq1_A	tumor suppressor
comp18701_c0_seq1_A	ornithine aminotransferase
comp18805_c0_seq2_A	s-norcoclaurine synthase
comp188_c0_seq1_A	transcription factor btf3
comp19032_c0_seq1_A	acyl- synthetase family member 4
comp1911_c0_seq1_A	nadh dehydrogenase
comp19264_c0_seq1_A	low temperature and salt responsive protein
comp1926_c0_seq2_G	cysteine proteinase inhibitor 3
comp19567_c0_seq1_A	transmembrane protein 53-like
comp1958_c0_seq1_A	synaptic glycoprotein sc2
comp1959_c0_seq1_G	dnaj protein
comp19669_c0_seq1_A	p-type h+-atpase
comp1978_c0_seq1_A	hydrophobic protein osr8
comp19_c0_seq1_G	thioredoxin h-type
comp20038_c0_seq1_A	staphylococcal nuclease domain-containing protein 1-like
comp20099_c0_seq1_A	eh domain-containing protein 1-like
comp2032_c0_seq1_A	universal stress protein a-like
comp20388_c0_seq1_A	nitric oxide synthase-like
comp20413_c0_seq1_A	nucleosome assembly protein 1-like 1-like
comp2042_c0_seq1_A	ribonucleoprotein a
comp20609_c0_seq1_A	dnaj homolog subfamily c member 3 homolog
comp2069_c0_seq4_G	6-phosphogluconate decarboxylating
comp2070_c0_seq2_A	6-phosphogluconate decarboxylating
comp20737_c0_seq1_A	protein fluorescent in blue chloroplastic
comp20835_c0_seq1_A	heat shock protein binding protein
comp20924_c0_seq1_A	kynurenine formamidase-like
comp21340_c0_seq1_A	3-isopropylmalate dehydrogenase
comp2135_c0_seq1_G	fkbp-type peptidyl-prolyl cis-trans isomerase chloroplastic
comp2161_c0_seq2_G	probable sucrose-phosphate synthase 2-like
comp2180_c0_seq1_A	aconitate cytoplasmic-like
comp2193_c0_seq1_G	inorganic pyrophosphatase
comp22110_c0_seq1_A	late embryogenesis abundant protein
comp221_c0_seq1_A	dehydrin- lea group 2-like protein
comp2225_c0_seq1_A	anther-specific protein
comp2232_c0_seq1_A	transmembrane protein 208-like
comp2232_c0_seq2_A	transmembrane protein 208-like
comp2284_c0_seq1_A	v-type proton atpase catalytic subunit a-like
comp2292_c0_seq4_A	smc3 protein

comp229_c0_seq1_A	ru large subunit-binding protein subunit chloroplastic-like
comp2307_c0_seq3_A	hydroxypyruvate reductase
comp231_c0_seq1_A	tonoplast intrinsic protein
comp23288_c0_seq3_A	fact complex subunit spt16-like
comp2334_c0_seq1_A	phosphopantetheine adenylyltransferase
comp23364_c0_seq1_A	f-box protein pp2-a13-like isoform x1
comp2336_c0_seq1_A	tyrosyl-dna phosphodiesterase 1-like
comp23406_c0_seq1_A	poly(adp-ribose) glycohydrolase 1
comp2340_c0_seq1_G	26s protease regulatory subunit 6b
comp2390_c0_seq1_A	heat shock protein 90
comp2390_c0_seq2_A	heat shock protein 90
comp2390_c0_seq5_A	heat shock protein 90
comp2390_c0_seq9_A	heat shock protein 90
comp240_c0_seq3_G	phospholipid hydroperoxide glutathione peroxidase chloroplastic-like
comp24136_c0_seq1_A	callose synthase 10-like
comp2415_c0_seq1_A	mitochondrial import inner membrane translocase subunit tim13-like
comp2427_c0_seq1_A	mlp-like protein 423-like
comp24330_c0_seq1_A	alpha- -glucosyltransferase alg10-a
comp24359_c0_seq1_A	calcineurin b-like protein 4
comp2440_c0_seq1_A	sll1 protein
comp24522_c0_seq1_A	peroxidase 2
comp2462_c0_seq1_A	3-oxo-delta( )-steroid 5-beta-reductase-like
comp2476_c0_seq1_A	cytochrome c oxidase subunit
comp2489_c1_seq1_A	dna repair protein rad23
comp24959_c0_seq1_A	inner membrane protein albino3
comp2500_c0_seq1_A	aquaporin
comp2500_c0_seq4_A	pip1 protein
comp2508_c1_seq1_A	proteasome chain protein
comp252_c0_seq1_A	elongation factor 2
comp25347_c0_seq1_A	disease resistance protein rpm1
comp2552_c0_seq1_A	glucose-6-phosphate isomerase
comp255_c0_seq1_A	pathogenesis-related protein 1
comp255_c0_seq2_A	pathogenesis-related protein 1
comp25608_c0_seq1_A	thioredoxin h-type
comp2560_c0_seq2_A	actin depolymerizing factor 4
comp25844_c0_seq1_A	histone-lysine n-methyltransferase ashh2
comp2584_c0_seq1_A	serine--glyoxylate aminotransferase
comp2595_c0_seq1_A	bidirectional sugar transporter sweet15-like
comp2661_c1_seq2_A	betaine aldehyde dehydrogenase
comp266_c0_seq1_A	translation initiation factor 5a
comp2683_c0_seq1_A	salt tolerance expressed
comp2695_c0_seq1_A	ribosome recycling factor

comp269_c0_seq3_G	glyceraldehyde-3-phosphate dehydrogenase
comp2709_c0_seq1_A	atp synthase subunit mitochondrial-like
comp2728_c0_seq1_A	heat shock factor-binding protein 1
comp2734_c0_seq1_A	20 kda chloroplastic
comp2741_c0_seq1_A	3-oxoacyl-synthase i
comp2752_c0_seq1_A	er lumen protein retaining receptor
comp28348_c0_seq1_A	delta-1-pyrroline-5-carboxylate synthetase
comp2882_c0_seq1_A	tpa: phospholipase d family protein
comp2915_c0_seq1_A	f1f0-atpase inhibitor protein
comp29243_c0_seq1_A	phospholipase d beta 1-like
comp2958_c0_seq1_A	peptidyl-prolyl cis-trans isomerase pin1-like
comp29703_c0_seq1_A	c-type lectin receptor-like tyrosine-protein kinase at1g52310-like
comp2985_c0_seq1_A	fkbp-type peptidyl-prolyl cis-trans isomerase chloroplastic
comp2995_c0_seq1_A	transmembrane bax inhibitor motif-containing protein
comp3003_c0_seq1_G	inositol -trisphosphate 5 6-kinase
comp3016_c0_seq1_A	ring-box protein 1a
comp3016_c0_seq2_A	ring-box protein 1a
comp3044_c0_seq1_G	peptidyl-prolyl cis-trans isomerase pin1-like
comp3048_c0_seq1_A	20 kda chloroplastic-like
comp3055_c0_seq1_A	peroxiredoxin-2e- chloroplastic-like
comp305_c0_seq1_A	cleavage and polyadenylation specificity factor subunit 5-like
comp30_c0_seq1_A	heat shock protein 90
comp30_c0_seq3_A	heat shock protein 90
comp31112_c0_seq1_A	soluble starch synthase iv-2
comp3112_c0_seq1_A	protein h2a
comp3138_c0_seq1_A	plasma membrane atpase 1
comp3150_c0_seq2_A	tyrosine n-monooxygenase
comp3152_c0_seq1_G	1-aminocyclopropane-1-carboxylate deaminase
comp3169_c0_seq1_A	probable v-type proton atpase subunit h-like
comp31908_c0_seq1_A	alpha-glucan water chloroplastic
comp3250_c0_seq1_A	ubiquitin-conjugating enzyme e2 36
comp325_c0_seq1_A	fatty acyl coa reductase
comp3288_c0_seq1_A	guanine nucleotide-binding protein beta subunit-like protein
comp3288_c0_seq5_A	guanine nucleotide-binding protein beta subunit-like protein
comp3418_c0_seq1_A	nad dependent epimerase
comp3421_c0_seq1_G	calcium lipid binding
comp34226_c0_seq1_A	catalase
comp3433_c0_seq1_A	er lumen protein retaining receptor-like
comp3438_c1_seq1_A	atp synthase delta chain
comp3438_c1_seq2_A	atp synthase delta chain
comp3446_c0_seq4_A	proteasome subunit beta type-6

comp3462_c0_seq2_A	26s proteasome regulatory complex atpase rpt3
comp3464_c0_seq1_A	cop9 signalosome complex subunit 1-like
comp3464_c0_seq2_A	cop9 signalosome complex subunit 1
comp3516_c0_seq1_A	tumor suppressor
comp3560_c0_seq1_A	dna repair protein rad23
comp3592_c0_seq1_A	40s ribosomal protein sa
comp3593_c0_seq1_A	delta-1-pyrroline-5-carboxylate dehydrogenase mitochondrial-like
comp3608_c0_seq1_G	6-phosphogluconate decarboxylating
comp3630_c0_seq1_A	protein transparent testa glabra 1
comp3671_c0_seq1_A	stromal cell-derived factor 2-like
comp3683_c0_seq1_A	cold acclimation protein
comp3724_c0_seq1_A	t-complex protein 1 subunit delta-like
comp3726_c0_seq2_A	heat shock factor-binding protein 1
comp378_c1_seq1_G	vacuolar proton-inorganic pyrophosphatase
comp38000_c0_seq1_A	pinus taeda anonymous locus 0_11081_01 genomic sequence
comp3832_c0_seq1_A	snw domain-containing protein 1-like isoform 1
comp3883_c0_seq1_A	mpk14 - mapk
comp3888_c0_seq1_A	rna binding protein rp120
comp3888_c0_seq3_A	rna binding protein rp120
comp388_c0_seq1_A	fiber protein fb15
comp3896_c0_seq2_A	peptide methionine sulfoxide reductase b5-like
comp3951_c0_seq1_A	glucan water dikinase
comp3962_c0_seq1_A	f-box protein skip8
comp3963_c0_seq1_A	hsp20-like chaperone domain family protein
comp3972_c0_seq1_A	adp-ribosylation factor-like protein 8a
comp4007_c0_seq1_A	replication factor a protein 2
comp4018_c0_seq1_A	elicitor-responsive protein 1
comp403_c1_seq1_A	transketolase 1
comp403_c1_seq2_A	transketolase 1
comp4050_c0_seq2_A	macrophage migration inhibitory factor homolog
comp4100_c0_seq1_A	phytochelatin synthetase
comp4100_c0_seq2_G	signal recognition particle 54 kda chloroplastic
comp4185_c0_seq1_A	3-oxoacyl-synthase i
comp4206_c0_seq1_G	serine threonine-protein kinase sapk10
comp4239_c0_seq1_A	cysteine desulfurase chloroplastic
comp425_c0_seq1_A	protein iojap- mitochondrial-like
comp425_c0_seq2_A	protein iojap- mitochondrial-like
comp4311_c0_seq1_A	soluble inorganic pyrophosphatase chloroplastic-like
comp4351_c0_seq1_G	voltage dependent anion channel
comp4370_c0_seq1_A	photosystem ii stability assembly factor chloroplastic-like
comp439_c0_seq1_A	prohibitin- mitochondrial-like
comp439_c0_seq2_A	prohibitin- mitochondrial-like
comp442_c0_seq1_A	cbs domain protein

comp442_c0_seq1_G	rna-binding protein
comp442_c0_seq4_G	nad dependent epimerase
comp4431_c0_seq1_A	methylmalonate-semialdehyde dehydrogenase
comp4431_c0_seq2_A	methylmalonate-semialdehyde dehydrogenase
comp4431_c0_seq4_A	methylmalonate-semialdehyde dehydrogenase
comp4494_c0_seq1_A	peroxisomal membrane protein pex11-1
comp4500_c1_seq1_A	diacylglycerol kinase 7
comp4505_c0_seq1_A	pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta
comp451_c1_seq1_G	cysteine proteinase 1-like
comp4520_c0_seq1_A	glyoxalase i
comp454_c0_seq1_A	serine threonine-protein kinase sapk7
comp4617_c0_seq1_A	hypothetical protein F775_07812
comp4630_c0_seq1_A	usp family protein
comp4659_c0_seq1_A	e3 ubiquitin-protein ligase chfr
comp4682_c0_seq1_A	cop9 signalosome complex subunit 6a-like
comp476_c0_seq2_G	tpa: aminomethyltransferase
comp4783_c0_seq1_A	1-aminocyclopropane-1-carboxylate deaminase
comp480_c0_seq1_G	ferredoxin--nadp leaf isozyme
comp485_c0_seq1_A	calcineurin b-like protein 1
comp4880_c0_seq1_A	coatomer subunit zeta-1-like
comp4920_c0_seq1_A	transmembrane protein 115-like
comp492_c0_seq1_A	cas1 domain-containing protein 1-like
comp4946_c0_seq1_A	chaperone protein dnaj mitochondrial-like
comp499_c0_seq2_A	ap-4 complex subunit sigma-1
comp5041_c0_seq2_A	thioredoxin reductase 2
comp5052_c0_seq1_A	general transcription factor iih subunit 4-like
comp510_c0_seq1_G	copper transport protein atox1
comp5185_c0_seq2_G	hydrophobic protein osr8
comp5191_c0_seq1_A	omega-3 fatty acid desaturase
comp5191_c0_seq2_A	omega-3 fatty acid desaturase
comp5191_c0_seq4_A	omega-3 fatty acid chloroplastic
comp5280_c0_seq1_A	fructan 1-exohydrolase
comp5288_c0_seq1_A	nadh dehydrogenase
comp5399_c0_seq1_A	dnaj protein homolog
comp5562_c0_seq1_A	dna-damage-repair toleration protein drt102
comp5565_c0_seq1_G	pyridoxal biosynthesis protein
comp5618_c0_seq1_A	protein z
comp5632_c1_seq3_A	protein flourescent in blue light
comp5634_c0_seq1_A	tpr repeat-containing thioredoxin tdx
comp5646_c0_seq1_A	charged multivesicular body protein 1-like
comp565_c0_seq1_A	acyl carrier protein
comp567_c0_seq1_A	growth regulator
comp5686_c0_seq1_A	hnh endonuclease
comp5686_c0_seq2_A	hnh endonuclease

comp5723_c0_seq1_G	heat shock protein sti-like
comp572_c0_seq1_A	nucleosome chromatin assembly factor group a
comp5756_c0_seq1_A	peroxidase 24-like
comp5757_c0_seq2_A	chromatin structure-remodeling complex protein syd-like isoform x2
comp5880_c0_seq1_A	er6 protein
comp5923_c0_seq1_A	trehalose-6-phosphate phosphatase
comp5972_c0_seq1_A	ribosome production factor 1-like
comp5990_c0_seq3_G	20 kda chloroplastic
comp5991_c0_seq1_A	universal stress protein a-like protein
comp6183_c0_seq1_A	thylakoid-bound ascorbate peroxidase
comp6183_c0_seq2_A	stromal ascorbate peroxidase
comp621_c0_seq1_A	fructose-bisphosphate aldolase
comp6252_c0_seq1_A	acyl- -binding domain-containing protein 6
comp6280_c0_seq1_A	vesicle-associated membrane protein 713-like
comp6382_c0_seq2_A	calcium-dependent protein kinase 3
comp6397_c0_seq3_A	copper chaperone for superoxide dismutase
comp6397_c0_seq4_A	copper chaperone for superoxide dismutase
comp6566_c0_seq2_A	cold acclimation protein cor413-pm1
comp6592_c0_seq1_A	phospholipid hydroperoxide glutathione peroxidase chloroplastic-like
comp6592_c0_seq2_A	glutathione peroxidase
comp6772_c0_seq1_A	late embryogenesis abundant protein lea14-a
comp6799_c1_seq2_A	fe-only nitrogenase accessory protein
comp6856_c0_seq1_A	mpk14 - mapk
comp6856_c0_seq3_A	mpk14 - mapk
comp6914_c0_seq1_A	phosphoenolpyruvate carboxylase kinase
comp695_c0_seq1_G	actin-depolymerizing factor 7
comp6978_c0_seq2_A	xyloglucan endotransglycosylase
comp6982_c0_seq1_A	dna repair helicase xpb1-like
comp701_c0_seq2_G	superoxide dismutase 2
comp703_c0_seq5_A	acetyl-coenzyme a synthetase
comp7063_c0_seq1_A	copper transport protein atox1
comp7085_c0_seq1_A	1-deoxy-d-xylulose 5-phosphate reductoisomerase
comp7187_c0_seq1_A	nitrate-induced noi protein
comp718_c0_seq1_A	disease resistance rpp13-like protein 1
comp718_c0_seq2_A	disease resistance rpp13-like protein 1
comp71_c0_seq1_G	ferredoxin--nadp leaf chloroplastic
comp7210_c0_seq1_A	transcription factor btf3-like
comp7244_c0_seq1_A	phosphopantothenoylcysteine decarboxylase
comp730_c0_seq1_A	atp synthase subunit mitochondrial-like
comp730_c0_seq2_A	atp synthase subunit mitochondrial-like
comp730_c0_seq3_A	atp synthase beta subunit
comp730_c0_seq4_A	atp synthase subunit mitochondrial
comp7364_c0_seq1_A	glutathione peroxidase

comp7502_c0_seq1_A	erd7 protein
comp7535_c0_seq1_A	protein
comp7540_c0_seq1_A	myb29 protein
comp7549_c0_seq1_A	predicted protein
comp754_c0_seq1_A	pyrroline-5-carboxylate reductase
comp7583_c0_seq1_A	u-box domain-containing protein 35-like
comp7652_c0_seq1_A	nudix hydrolase 8-like
comp7687_c0_seq1_A	cysteine proteinase inhibitor
comp7731_c0_seq1_A	ethylene receptor
comp7743_c0_seq1_A	slt1 protein
comp7749_c0_seq1_A	fatty acyl coa reductase
comp778_c0_seq1_A	succinate dehydrogenase 5
comp7835_c0_seq1_A	chloroplast processing peptidase
comp784_c1_seq1_A	sucrose synthase 1
comp784_c1_seq2_A	sucrose synthase
comp7861_c0_seq1_A	f-box wd-40 repeat-containing protein at5g21040-like
comp7888_c0_seq1_A	ru large subunit-binding protein subunit chloroplastic-like
comp7948_c0_seq1_A	dna-binding protein p24
comp794_c0_seq1_A	pathogen induced protein 2-4
comp799_c0_seq1_A	serine hydroxymethyltransferase
comp8071_c0_seq1_A	wrky dna binding domain containing expressed
comp808_c0_seq1_A	caffeic acid o-methyltransferase
comp8126_c0_seq1_A	u-box domain-containing protein 4-like
comp8145_c0_seq1_A	respiratory burst oxidase protein b
comp8338_c0_seq1_A	soluble starch synthase i
comp8345_c0_seq1_A	chloroplast lumen common protein family-like protein
comp8383_c0_seq1_A	dead-box atp-dependent rna helicase 50
comp838_c0_seq1_A	zinc finger a20 and an1 domains-containing protein
comp841_c0_seq1_A	sucrose synthase 2
comp8453_c0_seq3_A	lipoxygenase 2
comp8469_c0_seq1_A	pinus taeda anonymous locus 2_3934_01 genomic sequence
comp8545_c0_seq1_A	protein kinase
comp854_c0_seq1_A	superoxide dismutase 2
comp8613_c0_seq1_A	s-adenosyl l-homocystein hydrolase
comp871_c0_seq2_A	ras-related protein rab7
comp871_c0_seq4_A	ras-related protein rab7
comp876_c0_seq1_A	glyceraldehyde-3-phosphate dehydrogenase
comp877_c0_seq3_A	vacuolar atp synthase subunit c
comp8862_c0_seq1_A	morc family cw-type zinc finger protein 4
comp8959_c1_seq1_A	probable gmp synthase
comp907_c0_seq1_A	xyloglucan endotransglucosylase hydrolase protein 23
comp907_c0_seq2_A	xyloglucan endotransglucosylase hydrolase protein 23
comp9146_c0_seq1_A	cryptochrome 1b

comp9146_c0_seq2_A	cryptochrome 1b
comp9152_c0_seq1_A	receptor-like cytosolic serine threonine-protein kinase rbk1-like
comp9205_c0_seq2_A	k(+) efflux antiporter chloroplastic-like
comp922_c0_seq1_A	amino acid selective channel protein
comp9346_c1_seq1_A	26s proteasome non-atpase regulatory subunit rpn12a-like
comp935_c0_seq1_A	heat shock protein sti-like
comp936_c1_seq3_G	kynurenine formamidase-like
comp93_c0_seq1_A	ascorbate peroxidase
comp9430_c0_seq1_A	histone
comp943_c0_seq1_G	indole-3-glycerol phosphate chloroplastic
comp9448_c0_seq1_G	bri1-kd interacting protein 112
comp9517_c0_seq1_A	probable dolichyl pyrophosphate glc1man9 c2 alpha- -glucosyltransferase-like
comp958_c0_seq1_A	protein arginine n-methyltransferase 5
comp958_c0_seq3_A	protein arginine n-methyltransferase 5
comp9683_c0_seq2_A	phosphoacetylglucosamine mutase-like
comp9714_c1_seq2_A	glutamyl-trna amidotransferase subunit a
comp971_c0_seq1_A	cr14_horvu ame: full=cold-regulated protein blt14
comp976_c0_seq1_A	thioredoxin chloroplastic-like
comp9791_c0_seq2_A	geranylgeranyl transferase type-1 subunit beta
comp9854_c0_seq1_A	cryptochrome 2
comp9875_c0_seq1_A	chitinase 2

10. Eidesstattliche Erklärung / Declaration under Oath

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

_____ _____

Datum/*Date* Unterschrift/*Signature*

11. Curriculum Vitae

Personal Information

Name:	Matthias Enders
Date and Place of Birth:	06.10.1983, Halberstadt, Germany

Education

2003	A-level, Gymnasium Martineum, Halberstadt, Germany
2010	Diploma in Bioinformatics, University Halle-Wittenberg, Germany

Career

2008 – 2011	Research Assistant, Institute for Breeding Research on Agricultural Crops, Julius Kühn-Institute, Quedlinburg, Germany
2011 – 2014	PhD Student, Institute for Resistance Research and Stress Tolerance, Julius Kühn-Institute, Quedlinburg, Germany
2014 – 2016	Scientist, German Seed Alliance GmbH, Hohenlieth, Germany
2016 – 2020	Scientist, NPZ Innovation GmbH, Hohenlieth, Germany
from 2020	Head of digital Phenotyping, NPZ Innovation GmbH, Hohenlieth, Germany

Publications

Frese, L., Nachtigall, M., Enders, M., & Pinheiro de Carvalho, M. Â. A. (2012). *Beta patula* Aiton: genetic diversity analysis. Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces. CABI Publishing, Wallingford, 45-51.

Wehner, G., Balko, C., Enders, M., Humbeck, K., & Ordon, F. (2015). Identification of genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley. BMC plant biology, 15(1), 125.

Lehnert, H., Serfling, A., Enders, M., Friedt, W., & Ordon, F. (2017). Genetics of mycorrhizal symbiosis in winter wheat (*Triticum aestivum*). New Phytologist, 215(2), 779-791.

Mitterbauer, E., Enders, M., Bender, J., Erbs, M., Habekuss, A., Kilian, B., Weigel, H. J. (2017). Growth response of 98 barley (*Hordeum vulgare* L.) genotypes to elevated CO_2 and identification of related quantitative trait loci using genome-wide association studies. Plant Breeding, 136(4), 483-497.

Enders, Matthias „Extreme Classification Challenge – Seed Sorting“, S. 70 in: Bengio, Samy, et al. "Extreme Classification (Dagstuhl Seminar 18291)." Schloss Dagstuhl-Leibniz-Zentrum für Informatik, (2019).

Lange, T. M., Heinrich, F., Enders, M., Wolf, M., & Schmitt, A. O. (2020). In silico quality assessment of SNPs—a case study on the Axiom® Wheat genotyping arrays. Current Plant Biology, 100140.