Institut für Nutzpflanzenwissenschaften und Ressourcenschutz (INRES)

Adaptation assessment of a spring barley population to organic and conventional agro-ecosystems using genome sequencing approaches

Dissertation

zur Erlangung des Grades

Doktor der Agrarwissenschaften (Dr. agr.)

der Landwirtschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Michael Schneider

aus

Hagen, Deutschland

Bonn 2021

Referent:	Prof. Dr. Jens Léon
Korreferent:	Prof. Dr. Thomas Döring

Tag der mündlichen Prüfung: 04.06.2021

Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität Bonn

Acknowledgments

I am most grateful to Prof. Jens Léon for allowing me to graduate in such a challenging and interesting field of research. Furthermore, his guidance and the casual work environment he created gave me the chance to uncover my full potential and tackle various demanding tasks in the duration of my Ph.D. program. Furthermore, I would like to thank Agim Ballvora for the fruitful discussions, his guidance, and his help. Time in a Ph.D. period is rushing by like an inter-city train through a local train station, so I want to express my gratefulness to all supporting hands who helped me finish my workload. This includes Asis Shrestha, who helped me a lot in the laboratory and was my number one correspondent concerning wet lab work. Lea Hördemann, spending countless hours sampling leaves. Elisabeth Kukanov, as reliable and always available to help with all kinds of work. Karola Müller and Winfried Bungert for taking care of seed preparation, running the permeant selection trial for all these years, and their hands-on attitude. I saw the Ph.D. time as a chance to collect new knowledge in various backgrounds related to breeding and good scientific practice. Therefore, I want to thank my colleagues Said Dadshani, Diana Duarte, Benedict Oyiga, Ali Naz, Boby Mathew, Nazanin Afsharyan, Patrice Koua, Nurealam Siddiqui, Andreas Honecker, and Henrik Schumann for their collaborations and meaningful discussions over all the years. Many things would not have been possible without the help and support of my dear colleagues Sandra Markwitz, Anne Reiners, and Karin Woitol. As I not only worked on the field in Klein-Altendorf, whose stuff I would like to thank here as well, I also worked in the green house several times. For the support from the green house team around Josef Bauer, I would like to express my gratefulness as well. My acknowledgment also goes to the soil science, pathology, and plant nutrient departments. I want to express my gratefulness, especially to Gerd Welp, Erich-Christian Oerke, for the collaborations and for giving me the chance to test soil mineral variations and pathologic resistance variations. This would not have been possible without the guidance and support from Kerstin Lange and Lilli Wittmaier. Lastly, I would like to thank my bachelor and master students, who supported me in various experiments.

Summary

Over the past few decades, plant breeding has contributed significantly to increased crop yields in industrial agriculture. Besides, alternative, environmentally friendly farming approaches became more popular in the past years. However, questions arise which breeding goals are relevant to breed new varieties which are specially adapted for organic farming. The common practice of transferring conventionally adapted material into organic farming might lead to yield losses due to lacking adaptation of such varieties to organically farmed environments. So far, little is known about physiologically relevant characteristics for organically adapted varieties.

To answer this question, a long-term selection experiment in a spring barley population was established in 1998. A twice backcrossed population with a cultivar as recurrent and a wild-type as donor parents was established and cropped in conventionally and organically managed farming environments for more than two decades. Mainly the farming environments with their adjustments in fertilization, crop rotation, and plant protection as well as weather impacts should have driven the selection of individual genotypes in these populations. Therefore, the artificial selection was reduced to a minimum so that only natural selection should lead to changes in the allele frequencies of the populations.

In this thesis, complete populations for different generations and environments were genotyped entirely by applying a novel pool-based deep genotyping using a whole-genome resequencing approach. Implementing a haplotyping strategy makes it possible to dissect allele frequency variations on gene-level at low sequencing depth. Additionally, allele frequency variations between neighboring haplotypes have been used to calculate a consistent genetic map.

Comparing the organically and conventionally adapted populations, a distinct variation between the systems was observed. The organic population was characterized by a high variation in the population, with a conspicuous tendency to evolve apart from the conventional population. The latter showed evidence of an early equilibrium state from the twelfth generation onwards. Nevertheless, these latest allele frequency changes were small compared to significant changes in the early generations. The calculation of a genetic map additionally indicated a more pronounced selection in the conventional population, leading to the assumption that the heterogeneity is lower in this population. In general, the wild alleles showed a higher fitness effect in the organic farming system. Some wild-type alleles were selected in the conventional system, whereas those alleles with a depressing effect on the yield were negatively selected. Exemplarily, this has been observed for two brittleness alleles and a dormancy allele. A metastudy of identified QTL regions for agronomically relevant traits revealed several candidate loci with variant allele frequencies between environments or generations. Those were clustered in yield components, yield physiology, biotic stress resistance, drought tolerance, root morphology, and nutrient uptake. Significant variations between the organically and conventionally adapted populations were observed for the root morphology, yield physiology, and drought tolerance. The variations in the root system were confirmed by two phenotypic experiments, which revealed higher root length, lower angle, and increased heterogeneity in the organic population. Pronounced root growth and coverage of the rhizosphere by the entire population, with each individual in its unique niche, might enable a better accumulation of nutrients and, at the same time, will increase resilience against drought events.

Concluding, evolutionary adaptation processes in large populations undergoing long-term natural selection processes of a cereal crop were genetically examined by an innovative pool sequencing and haplotyping approach. Based on these findings, breeding goals for organic varieties could be adjusted and propagated.

Kurzfassung

Die Pflanzenzüchtung der vergangenen Dekaden hat in außerordentlichem Maß zur Ertragssteigerung in der industriellen Landwirtschaft geführt. Durch die zunehmende Nachfrage und Popularität alternativer und umweltschonender Anbaumethoden ergibt sich die Frage, wie Sorten speziell für diese ökologisch nachhaltigen Anbauweisen selektiert und entwickelt werden können. Die aktuell weit verbreitete Praxis, in welcher Saatgut von Sorten aus konventionellen Züchtungsumwelten im ökologischem Landbau Verwendung findet, lässt die Vermutung zu, dass diese Sorten nicht optimal an ihre Anbauumgebung angepasst sind und somit das volle Ertragspotential des organischen Landbaus nicht ausgeschöpft werden kann. Wenig ist über potentiell relevante physiologische Eigenschaften ökologisch angepasster Sorten bekannt. Um diese Fragen beantworten zu können, wurde zum Ende des letzten Jahrtausends ein Langzeitversuch etabliert, in welchem der selektive Einfluss des Anbausystems auf die genetische Zusammensetzung von einstmals identischen Populationen untersucht werden sollte. Diese Population, entstanden aus einer zweifachen Rückkreuzung einer Kulturgerste und einer Wildform, wurde über mehr als 20 Jahre in einem konventionell und einem organisch geprägten Anbausystem angebaut. Keine gerichtete menschliche Selektion wurde durchgeführt, so dass nur das Anbausystem und die Wettereinflüsse eine Veränderung der Allelfrequenz innerhalb der Populationen bewirken konnten. In dieser Arbeit wird erstmals eine komplette Population in mehreren Generationen mit vergleichsweise geringem Einsatz von Ressourcen vollständig genotypisiert. Durch das Sequenzieren des gesamten Genoms, das Wissen der elterlichen Allele und einen fortschrittlichen Haplotypisierungansatz ist es möglich, Allelfrequenzänderungen auf Einzelgenebene mit hoher Sicherheit schätzen zu können. Darüber hinaus wurde ein neuartiger Ansatz entwickelt eine genetische Karte, basierend auf Allelfrequenzunterschieden, zu schätzen.

Der Vergleich der beiden unterschiedlich angepassten Populationen hat mehrere Erkenntnisse geliefert. Eine deutlich erhöhte Heterogenität im organischen System konnte, verglichen zum konventionellen System, speziell in den späteren Generationen identifiziert werden. Die größten Unterschiede in den Allelfrequenzen wurden in den frühen Generationen erfasst. Nach etwa 10 Generationen hatte die konventionelle Population ein Gleichgewichtsstadium erreicht, in dem die Allelfrequenz keine signifikanten Unterschiede über die folgenden Generationen mehr aufzuweisen hatte. Dagegen deutet die organische Population eine zunehmende Diversifizierung an, die sich neben zunehmender Heterogenität zwischen Wiederholungen derselben Population und Generation in der Zunahme der Wildform Allelfrequenz an vielen Loci kennzeichnet. Die Kalkulation einer genetischen Karte aus den Allelfrequenzunterschieden und der physikalischen Distanz benachbarter Haplotypen deutet zudem auf eine stärker gerichtete Selektion hin. Allgemein kann beobachtet werden, dass Wildformallele in der organischen Umwelt eine deutlich höhere Fitness besitzen, als es in der konventionellen der Fall ist. Nichts desto trotz gibt es auch einige Gene, deren Wildformallele auch in der konventionellen Variante positiv selektiert wurden. Wildformallele mit negativen Effekten auf den Ertrag wurden in beiden Umwelten negativ selektiert. Dies wurde sowohl für zwei Spindelbrüchigkeitsallele als auch für ein ausgeprägtes Dormanzallel erfasst. Durch eine Metastudie und wurden Kandidaten QTL Regionen ermittelt, die in den Bereichen Ertragsphysiologie, Wurzelwachstum, Ertragskomponenten, biotische Resistenz, Nährstoffaufnahme und Trockentoleranz relevante Unterschiede zwischen den Anbausystemen oder Generationen gezeigt haben. Bei der Betrachtung der Wildformallelfrequenz dieser Regionen offenbart sich, dass die konventionelle und die organische Population statistisch signifikante Abweichungen im Wurzelwachstum, der Ertragsphysiologie und der Trockentoleranz zeigen. Die Abweichung im Wurzelsystem wurde mittels zweier phänotypischer Experimente bestätigt. Demnach wachsen die Wurzeln in der organischen Population deutlich tiefer, haben dazu aber auch eine erhöhte Heterogenität. Eine breiter gefächerte Durchwurzelung der Rhizosphäre durch eine diverse Zusammensetzung der Population, bestehend aus Tief- und Flachwurzlern, ermöglicht potentiell eine

bessere Ausnutzung von Nährstoffen im Boden und sorgt gleichzeitig für eine höhere Trockenstressresilienz.

Zusammenfassend bietet diese Arbeit erstmals einen sehr detaillierten Einblick in die Evolution einer Kulturart über mehrere Generationen durch die Verwendung eines innovativen Poolsequenzierungs- und Haplotypisierungsansatzes. Mit den Ergebnissen dieser Arbeit lassen sich Zuchtziele für organische Sorten ableiten.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

BC	–	back cross
GB	–	giga base
ddNTP	–	dideoxide nucleotide triphosphate
H	–	hydrogen
OH	–	oxyhydrogen group
PCR	–	polymerase chain reaction
SNV	–	single nucleotide variant
SNP	–	single nucleotide polymorphism
WT	–	wild-type
AF	–	allele frequency
ET	–	evapotranspiration
KASP	–	competitive allele specific PCR
WGrS	–	whole genome sequencing
MACE	–	massive analysis of C-terminal ends
GBS	–	genotyping by sequencing
HAF	–	haplotype allele frequency
GO	–	gene ontology
CDS	–	coding sequence of a gene
PCA	–	principal component analysis
HAF	–	haplotype allele frequency
Vcf	–	variant calling file
BAM	–	alignment files
HC	–	high confidence (gene)
QTL	–	quantitative trait loci
Bp	–	base pairs
Phred score	–	quality score of sequencing results
GH	–	gene haplotype

MH	–	marker haplotype
CH	–	contig haplotype
RMSE	–	root mean square error
P value	–	probability value
REML	–	Restricted maximum likelihood
SPAD	–	none destructive chlorophyll measurement
F_{ST}	–	fixation index
gwAF	–	genome-wide allele frequency
wtAF	–	wild-type allele frequency
IAF	–	ISR42-8 allele frequency
D’	–	coefficient of linkage disequilibrium
cM	–	centi Morgan
Anova	–	analysis of variance
Glm	–	generalized linear model
Lm	–	linear model
Mlm	–	mixed linear model
Vcf	–	variant calling format
Bam	–	compressed alignment format
Fasta	–	sequence file format
Fastq	–	sequence file format with read quality

Chapter 1 - Introduction

1.1 What defines organic farming, its origin and today's standing

In the history of crop management, organic farming is a relatively new approach in food production. The three-field system has marked the first stage of yield increase, followed by the purposeful application of fertilizer. Another step in yield increase was enabled with the discovery of mineral fertilizer production, especially ammonia, by the Haber-Bosch method. As the plant material of those days was poorly adapted to higher fertilizer levels, the crop yield in the early 20th century did not increase or even declined. It was suspected that the use of mineral fertilizer led to a disturbance of the plant metabolism, acidification of soils, soil fatigue, and soil compaction by the extended use of machines. These observations and hypotheses canalized in the origin of organic farming theories, which resulted in a successful application of these methods in the 1920s and 1930s. However, the success of organic farming was disturbed by World War II and the deployed Marshall plan for Europe, with the main focus on yield expansion for a starving European population. As a result, organic farming was not a factor in European and American agriculture until the late 1970s. In 1972, the International Federation of Organic Agriculture Movements (IFOAM) was founded (Lockeretz, 2007). IFOAM has the goal to promote organic farming all over the world. A transformation of the acceptance in science and society was supported by previous studies (Lockeretz et al., 1981). Those studies negated the widely spread opinion of organic farming as an old-fashioned, low-yielding farming approach. IFOAM has grown from that time on, having more than 500 members. Among other achievements, IFOAM established a definition of organic farming that found its way into the EU-Eco-regulations in 2007 (Council et al., 2008). The definition of organic agriculture is based on four core requirements: the principle of health, the principle of ecology, the principle of fairness, and the principle of care.

"Organic Agriculture is a production system that sustains the health of soils, ecosystems, and people. It relies on ecological processes, biodiversity and cycles adapted to local conditions rather than the use of inputs with adverse effects. Organic Agriculture combines tradition, innovation, and science to benefit the shared environment and promote fair relationships and good quality of life for all involved." IFOAM General Assembly, 2008 (Anonymous VI, 2008)

These four requirements can also be called well-being, environment, production, and economics. Where organic farming considers all four on equal levels, the conventional farming approach tends to overemphasize the production aspect. Water pollution, soil degradation, loss of biodiversity and the impact on climate change have become emerging factors in people's food choices. All these points create costs that are not compensated by the product's price and so have to be paid by the society as a whole (Pretty et al., 2005). By adding manure instead of mineral fertilizer, the organic matter can be increased significantly, leading to a higher CO_2 fixation in the soil and better resistance to dry periods. Furthermore, the biodiversity in soils can be enriched (Anonymous V, 2013; Fließbach et al., 2007; Pulleman et al., 2003). The well-being aspect of the organic farming requirements highlights the reduction of poverty and pressure within the production chain. Opposing to this, the conventional farmers are heavily pressured by the market, where a standardized product can be sold before it is produced. Thus, the price of the product mainly defines the circumstance of the production. Figure 1 illustrates the weights and trade-offs that are mediated for both farming systems.

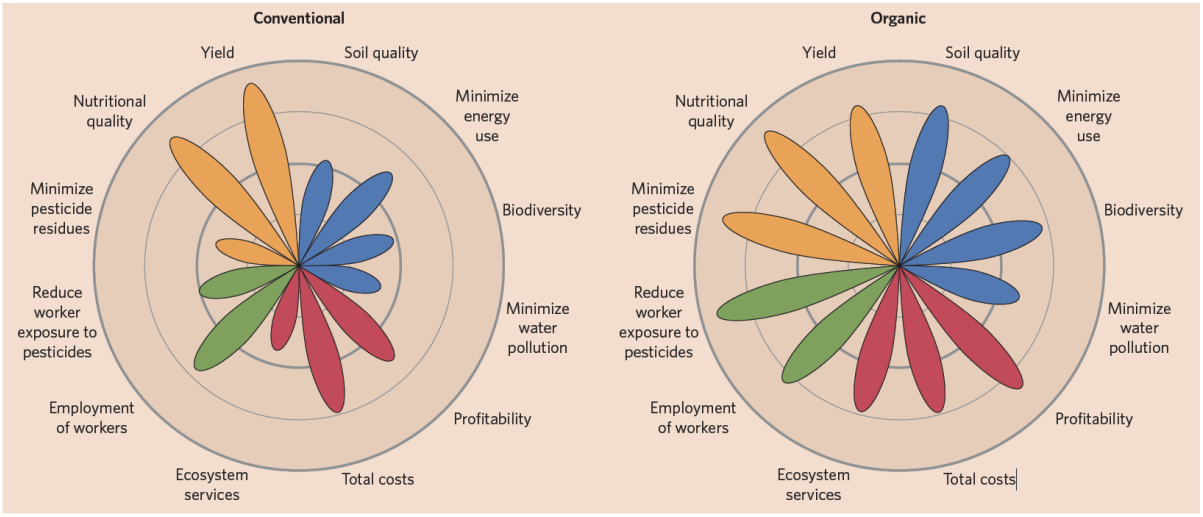


Figure 1: The four primary areas of sustainability and the trade-off in conventional and organic farming between these areas, from Reganold et Wachter 2016 (Reganold & Wachter, 2016)

All these principles have in common that they include the environment and social components as the defining factors. High-quality food with effects on preventive health care and well-being must be produced in living eco systems. In a manner of reuse and recycling, the environment should retain its health and productivity unlimited beyond the boundaries of the field plot. The instrumentalization of new technologies is a crucial feature to match the production with the well-being aspect (Anonymous V, 2013). This transfers into a significant variation between the organic and conventional farming systems. While conventional farming has benefited from developing new, high-yielding varieties, the organic community could not benefit from these. On one side, the rules and guidelines for the selection of varieties are strict. On the other side, modern breeding has not produced a decent number of adapted varieties for organic farming. This happened because the organic seed market was too small, and modern materials were not common in the community. Furthermore, the use of genetically modified material is strictly forbidden. The genome as a unit has to be respected as well as the cell. Genetic diversity creation must be performed by crossing within species boundaries, and the farmer has to have the ability to reuse his seed material without major drawbacks (farmers privilege). New varieties have to be created the traditional way, based on phenotypic selection and field-based testing. Modern genetic information is allowed to be implemented, but the focus should remain on the whole organism selection (Anonymous V, 2017).

This raises the question, how such a variety should be characterized. The demand for organically adapted seeds exceeds production; therefore, a relevant fraction of the seed material comes from conventionally adapted and produced varieties (Döring et al., 2012). Nevertheless, the application of adapted genetic material for organic farming makes sense, as this should be better adapted to the needs of such systems. This should also help to reduce the observed yield reduction in organic systems by 25-33% for cereals (Crespo-Herrera & Ortiz, 2015). The attributes of such adapted varieties have been postulated by several reviews, indicating that resistance to weeds and diseases, an enhanced nutrient use efficiency, stable seed quality, and high yields are beneficial to achieve a reduction of the yield gap (Crespo-Herrera & Ortiz, 2015; Lammerts Van Bueren et al., 2011; Osman et al., 2016; Wolfe et al., 2008). Weed suppression has been observed to be enhanced when the early vigor and allelopathic interactions are promoted (Bertholdsson, 2005). Besides, an indication for improved nutrient uptake and use efficiency was given by studies that highlighted the positive effect of deep rooting and overall higher root mass and weight in organically adapted varieties (Bertholdsson et al., 2016; Bhaskar et al., 2019; Finckh & Bhaskar, 2019).

Besides all these variations, it has been expressed that many breeding goals overlap between organic and conventional environments. Therefore, advanced conventionally selected generations are postulated to be the best starting material for selection under organic environments or in low input systems (Löschenberger et al., 2008; Osman et al., 2016). Contrasting to these opinions, overall higher breeding success was observed in varieties selected in organic environments alone, without prior selection in conventional systems (Reid et al., 2009).

1.2 Genetic diversity in wild-types and modern cultivars

The first domestication of barley was reported to be approximately 10.000 years ago, making it one founder crop of agriculture. In the fertile crescent, *Hordeum vulgare* was domesticated from its wild-type *Hordeum spontaneum* (Badr et al., 2000). Barley has a wide range of diversity distributed all over the world. Due to its adaptation to high altitude, drought and salinity, it can cover many variant niches. The USDA-ARS is the second most extensive grain collection and stores more than 33,000 barley accessions of modern cultivars and landraces. By genotyping a fraction of this set, five major sub-populations could be identified, located in the Ethiopian high land, middle to the far east, near east to Mediterranean area, central Europe and eastern Europe/North America (Muñoz-Amatriaín et al., 2014). The most isolated group of this is the group located from Iran over India to China and Japan. This finding is in line with the observation of a second domestication process of barley, located 1,500 to 3,000 km farther east of the Fertile Crescent (Morrell & Clegg, 2007). This genetic diversity in cultivated barley has faced bottleneck events, which resulted in a decrease in genetic diversity compared to wild barley. To reverse this process of genetic depression, the introgression of wild barley material in breeding programs can lead to a major increase in genetic diversity. It can introduce beneficial alleles that can deal with future challenges of soil salinification, drought tolerance and rare resistance to biotic stressors (Kilian et al., 2006; Muñoz-Amatriaín et al., 2014). Furthermore, the genetic diversity within the subgroups is high as well. Within the Fertile Crescent, the genetic composure of selected sites can be highly variant and even within a single location, the primarily homozygote stored genetic diversity can be widely dissimilar (Ceccarelli et al., 1987; Nevo E. et al., 1979; Nevo et al., 1984). As this observation was also made for economically significant traits, the genetic resources from the origin of barley can be utilized in barley crop improvement (Nevo et al., 1984).

In the past couple of years, driven by genotyping approaches, many gene trait interactions have been described. Crosses were established between wild genotypes and cultivars to increase the genetic variation of the crossing population on which the gene trait associations were studied. Following this widely used approach, many candidate regions and genes for multiple agronomic relevant traits have been characterized. Diseases have been investigated as a significant reduction factor for yield and resistance genes for barley yellow dwarf virus infections (Dragan et al., 2013; Lüpken et al., 2014; Niks et al., 2004; Riedel et al., 2011), Fusarium (Bedawy et al., 2018), Powdery Mildew (Lyngkjær et al., 2000; Schmalenbach et al., 2008; M. Von Korff et al., 2005) or leaf rust (Castro et al., 2012; Schmalenbach et al., 2008; M. Von Korff et al., 2005) were found in the genetic background of barley. Furthermore, the genetic background of root morphology has been investigated in the recent past, discovering variations in root length, xylem density and nodal root development, as well as variation in response to drought (Naz et al., 2014; Oyiga et al., 2020; Reinert et al., 2016). Besides the extensive yield-related field of traits, the genetic variation in drought response has become an increasingly important topic. Studies have revealed the potential of drought stress adaptation strategies in European cultivars by quantitative trait analysis (Honsdorf et al., 2014; Mohammed & Léon, 2004). The last major field of variation studied is related to plant physiology and yield defining traits. Significant yield increases have promoted the green revolution. It is based on the introgression of dwarfing genes in major

cereal cultivars (Hedden, 2003). Additionally, many other growth-regulating genes have been identified. Other dwarfing and semi-dwarfing genes have been found (Guan et al., 2012; Ren et al., 2016; Vu et al., 2010; Y. Xu et al., 2017), but also the influence of hormones on growth regulation has been studied (Itoh et al., 2004; Marzec & Alqudah, 2018). Besides the growth, the flowering and time until maturity have an enormous impact on adapting to a location. Wild-types originated in the Fertile Crescent tend to have a much more extended vegetation period than domesticated European varieties. Several genes controlling the length of the growth period have been identified (Jones et al., 2008; Laurie et al., 1995). Characterizing genes defining the yield levels is a challenging task. Numerous genes interacting with the environment have an impact on the yield level. This makes a depiction of all genes demanding. Nevertheless, several studies tried to find a genetic fundament for the deviation in yield levels, and some genes have been described (Liller et al., 2017; Nadolska-Orczyk et al., 2017; Saade et al., 2016; Schmalenbach et al., 2009; Sharma et al., 2018; Watt et al., 2019; X. Xu et al., 2018). The diversity in the barley gene pool is high, not only for quantitative but also for qualitative traits. Alterations have been reported also for protein content of the grain (Cai et al., 2013; Schmalenbach & Pillen, 2009), anther extrusion (Honda et al., 2006), threshability (Schmalenbach et al., 2011) and ear brittleness (Komatsuda et al., 2004; Pourkheirandish et al., 2015).

The barley gene pool is highly diverse. Especially in wild forms, many alleles lost in the domestication process are still present. These alleles can be used to improve modern cultivars by introgression breeding. This way, barley varieties could be adapted to challenging climate and weather variations. The values of these wild relatives are eminent in organic breeding programs. Nevertheless, the burden of introgression breeding is the unintended transfer of unbeneficial alleles. Therefore, new breeding techniques, based on genome editing, could be helpful to overcome the time and cost-intensive introgression of new alleles in common varieties (Andersen et al., 2015).

1.3 Barley as a model crop for studying the variation between farming systems

Barley is one of the most relevant crops. It is essential for animal feeding, brewing, and in some regions of the world also as a staple food (Muñoz-Amatriaín et al., 2014). As illustrated in figure 2, almost 20% of the whole cereal yield in the European Union is accommodated by barley.

Barley seems to be the perfect fit to characterize the impact of the farming system. Among the cereals, barley is reported to have the highest yield gap between organic and conventional farming. Furthermore, the discrepancy of yield tends to widen in regions with high water availability (De Ponti et al., 2012). Generally, a yield gap of 8-25% is reported, with barley on an average level of 31% (De Ponti et al., 2012; Reganold & Wachter, 2016).

The rich genetic background of wild types and the excellent adaptation of these to arid climate conditions, characterized by dry and hot summers, are the ideal base to study the varying yield levels and adaptation strategies to climate change (Kilian et al., 2006; Muñoz-Amatriaín et al., 2014). As Europe and Germany, in particular, have faced successive dry springs followed by hot summers with little rain, the proclaimed advantage of conventional farming in terms of productivity might be reduced. The yield gap has shown to be highest in areas where water is not a limiting factor and even can be overcompensated in dry years (Lotter et al., 2003; Neumann et al., 2010). Since 2012, the physical, genetic and functional sequence assembly of the barley genome is available. Although its genome is massive (5.1 Gb), its diploid characteristic makes it simpler to study compared to the hexaploidy wheat. Information or techniques established in barley can be transferred to other crop species with a more complicated genome assembly. Almost 80.000 transcript clusters as potential genes and about 40.000 genes with coding DNA information as well as a growing library of gene function and ontologies are

available (Mayer et al., 2012). The mainly self-pollinating character leads to a homozygote state of individual genotypes, conserving the genetic variability of a land race in the combination of different genotypes. With this comes several advantages. With a homozygote state, dominant and recessive or intermediate state do not need to be addressed, as only one allele in the genotype is present. Selfing species have been reported to promote and maintain adaptation processes towards their environment (Allard, 1988). Unbeneficial alleles cannot be compensated by each other. This has direct consequences for the fitness of a genotype. Furthermore, producing identical seeds for the following generation is easy to achieve, plus germination of seeds is usually not affected by stratification needs.

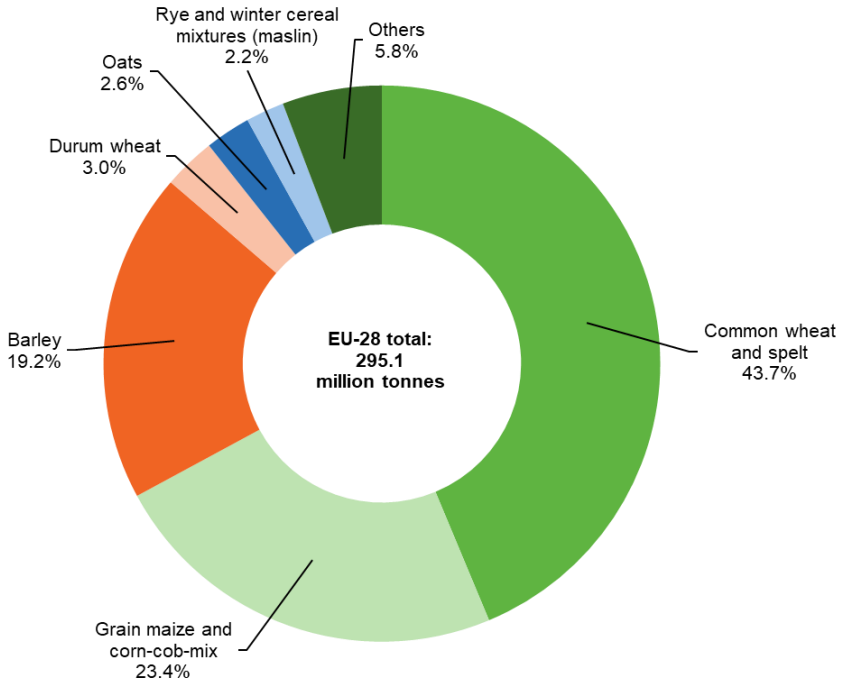
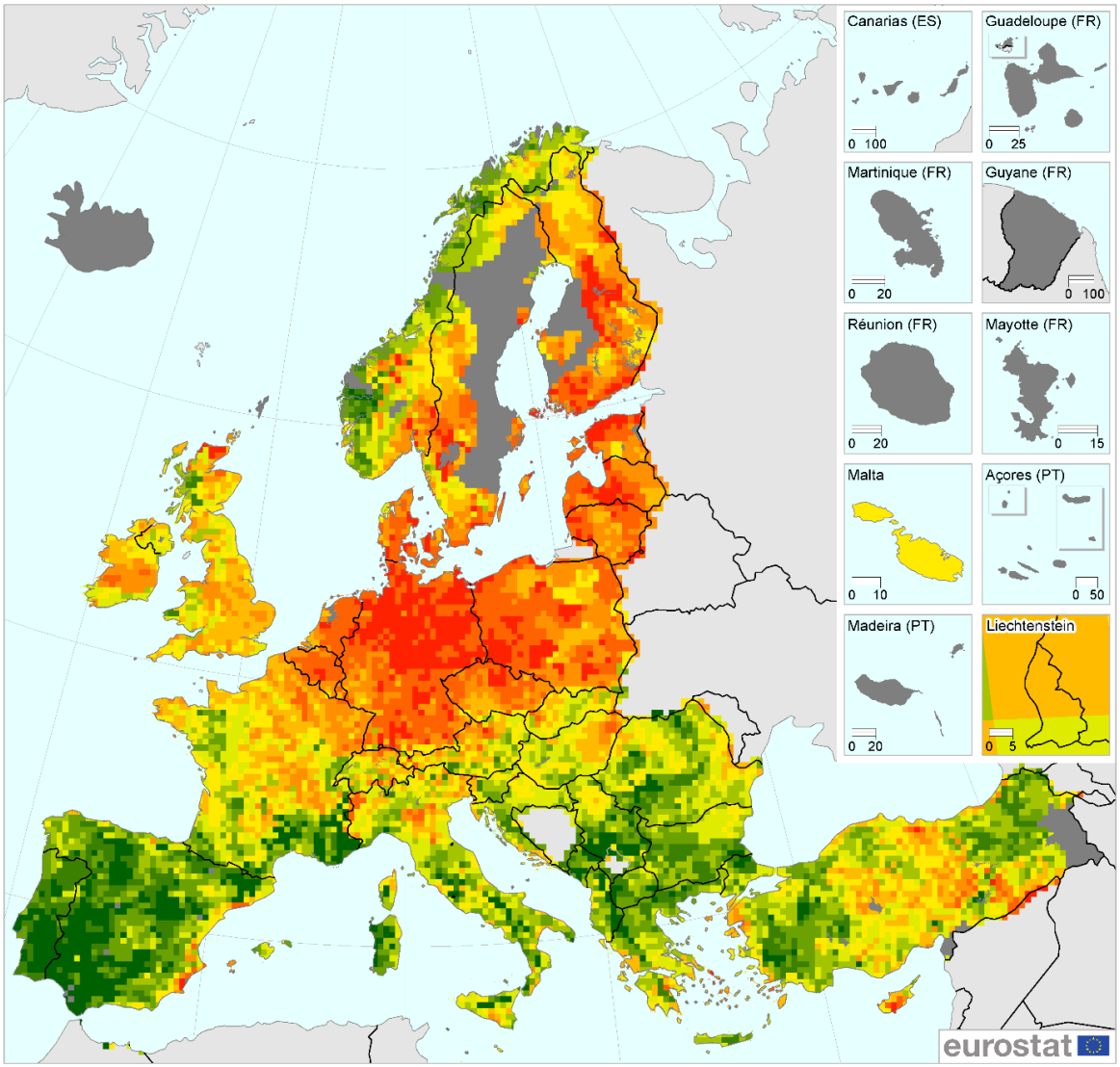


Figure 2: Grain Yield of Cereals grown in the European Union in 2018, Source: Eurostat (online data code: apro_cpnh1)

1.4 Climate change as an impact on future farming

Agriculture itself has a major impact on global warming. About 20 to 34% of the global emission of greenhouse gases can be directly linked to agricultural activity. Methane emitted by cows and rice production, carbon dioxide emitted by deforestation, animal production and intensive application of conventional fuels and nitrous oxide emissions by mineral fertilizer production, paddy rice production and not correctly applied fertilizer adding up to these numbers (Aydinalp & Cresser, 2008). Estimating the impact of climate change on agriculture is difficult, as the effects might differ completely depending on the region. Among the most discussed negative impacts, the reduction of water availability could be the most challenging part (Aydinalp & Cresser, 2008). As also illustrated in figure 3, central Europe has faced a severe drought in 2018, followed by an equally dry season in 2019. The reduction of water availability can result in severe declines in grain yields (Stone & Schlegel, 2006). The German federal ministry of agriculture reported a yield decline of 15% for cereals compared to the previous year for 2018 and a 16.3% decline compared to the mean of the previous six years (Daten-analysen, 2018). The missing of rain events in the growing period becomes more severe with declining soil quality. Sandy soils tend to have a low water storage capability (Ritchie, 1981). These locations demand regular rain events. Otherwise, a loss of yield is the consequence.

The adaption of farming strategies as lower sowing densities, soil conservation methods like the increase of organic matter, decreasing evaporation losses, and adapted genetic material or crops have to be considered (Lal et al., 2012). While organic matter itself can store much water, the impact on the field is reported to be relatively small. Sandy soils can benefit the most with up to 3 mm water per 10 cm soil at an increase of organic matter of 1% (Minasny & McBratney, 2018).



Note: This map shows the Standardised Precipitation Evapotranspiration Index (SPEI) used to estimate drought conditions affecting Europe in spring-summer (March-August 2018. Colours are associated with values of the SPEI.

* This designation is without prejudice to positions on status, and is in line with UNSCR 1244/1999 and the ICJ Opinion on the Kosovo declaration of independence.

Source: JRC

Figure 3: Drought conditions in central and northern EU Member States; March to August 2018. Red indicates severe drought conditions, while green indicates a positive water balance. Source: Eurostat/JRC https://ec.europa.eu/eurostat/statistics-explained/index.php/Agricultural_production_-_crops#cite_ref-1

Besides the local shortage of water, a more general problem with the loss of arable land due to salinity and rising sea level is associated with climate change. Especially regions defined by a river delta and tides might be strongly affected by both these problems. These regions usually are highly populated, leading to an increase in migration from flooded and arid areas to cities or more humid regions.

However, there are also positive aspects of climate change. The higher CO_2 concentration in the air might lead to a CO_2 fertilization effect. C_3 plants have always been less efficient in the fixation of CO_2, which might lead to an increased productivity gain for these compared to C_4 plants (Lobell & Field, 2008). Furthermore, the need for pest control could be reduced due to lower moisture levels, and in some regions, the cropping of some crops with higher temperature demands might get possible (Aydinalp & Cresser, 2008).

1.5 Adaptation to biotic and abiotic stressors of plant communities

Evolution is based on fixed adaptation effects that lead to an advantage over other individuals of the same or different species. In the process of evolution, land plants have evolved a vast variation based on a conserved core of genes. The complex response to the environment and its interaction makes it hard to define key regulators and beneficial alleles. Significant adaptation and evolution drivers of land plants can be found in the five groups phenolics, plant hormones, Superoxide dismutase, isoprene and heat shock proteins. Phenolics protect the plant from many different aggressors like UV radiation, herbivores, microbes and retain their structural rigidity. Plant hormones play a crucial role in the developmental processes of the plant, because they regulate cell division, senescence, dormancy, flowering and seed or fruit development. Plants need to be well adapted to their environment in terms of flowering time and seed production. An unknown genotype with a too-long growing period might face floral damage due to frost or an incomplete seed or fruit production. Superoxide dismutase activity can prevent the plant from the adverse effects of radicals and related oxidative stress. The integrity of the chloroplast is based on the emission of isoprene. Additionally, they influence the tolerance of thermal stress, as heat shock proteins have as well. They are responsible for the prevention of irreversible thermal aggregation of proteins. Plants are immobile, and therefore they only have the chance to avoid or escape the stress factor. Escape strategies are related to the production of seeds before the stress occurs and avoidance strategies to physiological adaptations (Sierra et al., 2015; Waters, 2003). The adaptation to a niche is based on the allelic variation. A local adaptation to serpentine soils of *Arabidopsis lyrate* could be identified as a variation of alleles of few genes (Turner et al., 2010). In another study, 175 genes were identified that were associated with a topo-climatic factor (Fischer et al., 2013a). The ability to adapt to new environments of a maize landrace was illustrated in literature (Wisser et al., 2019). They could achieve a 26-day decrease of flowering time within ten generations to adapt the population towards the new temperate environment. Allard described a more extended period of evolution and adaptation in cultivated plants (1988). Sixty generations of a barley composite cross II have been pheno- and genotyped. The set was synthesized in 1928 by crossing 28 barley varieties. Equal numbers of F1 hybrid seeds of the 378 possible pairwise crosses were used to establish the population. It was managed according to standard agricultural practices without artificial selection, and the seeds of each harvest were sown out in the following year. They could watch a rapid change of the allele frequency in the first couple generations, which increased the fitness of the population. This was supported by the fact that the seed yield increased after 15 generations from 60% to 95%, compared to commercial varieties. After these generations, the yield level remained stable. As the kernel size and the number of kernels increased by about 20%, they summarized that natural selection favors genotypes with high reproduction fitness. Besides the yield traits, they found other traits to adapt to their environment. The population remained highly diverse but developed its mean clearly towards higher reproduction fitness in any kind of trait.

This included, among others heading date, leaf and culm size, as well as awn length. Furthermore, they tested the genetic composition. They identified an increased frequency of six-row barley compared to the number of two-rowed genotypes. This had shown to be the most relevant effect, as it also influenced many other traits. Disease resistance did not play a significant role in the selection process. Susceptible alleles were more frequent in the population than resistant. The group of resistant plants tent to produce smaller numbers of seeds. Allard called this process of short to mid-term adaptation to the environment *microevolution* (Allard, 1988). This study indicates the capability of adaptation processes towards high reproduction fitness in a new environment of a diverse population. This population does not even need to show lower yields than equivalent breed varieties.

1.6 Genotyping approaches

The story of sequencing started in the mid of the 1970s when Sanger sequencing was developed. Sanger sequencing was a synonym for DNA sequencing for the coming three decades. The two basic principles are the ability to separate DNA fragments according to size variations by polyacrylamide gels and the fact that modified nucleotides could be integrated by the polymerase. Therefore required ddNTPs varied in their 3' End, containing an H instead of an OH. This led to a break of the polymerase activity. If a fluorescence marker is added to the ddNTP, the base can be identified on the gel. The sequencing reaction mix for this required a DNA polymerase, Primer of a specific region, and dNTP and ddNTPs, which are usually mixed in a ratio of 100 to 1. Up-to-date, Sanger sequencing is the most accurate sequencing method allowing for large fragments to be sequenced. The first next-generation sequencing technique was 454 Pyrosequencing, developed in the late 1990s. Compared to Sanger sequencing, it is possible to sequence the read while it is synthesized. Nucleotides are released one after the other to a chip with wells. The polymerase incorporates these if possible. This creates a detectable light signal. The nucleotides are removed, and a new set is added. Compared to the following released sequencing methods, the 454 Pyrosequencing is relatively slow, but yield reads of an average length of 700 bp and has high accuracy. The Reversible terminator sequencing (Illumina) was developed a decade later. It is at the moment the by far most prominent sequencing method and follows a sequencing-by-synthesis approach, just like Pyrosequencing. The method itself is more related to Sanger sequencing. All four bases plus a polymerase are added to flow cells. Additionally to standard dNTPs, modified ones are added. These stop the polymerase, which can be detected by a laser. The libraries are created with two different adapters, which are ligated to the end of the sheared, single-stranded DNA. With this method, paired-end sequencing became possible. The accuracy is slightly lower than Sanger sequencing, overcome by the ability to produce millions of reads in short amounts of time. These days, it is the standard sequencing method for transcriptome and genome resequencing and is used in this study as well.

Table 1: Overview of the most prominent sequencing methods and their characteristics, from Phylogenomics (Bleidorn, 2017)

Platform	Roche 454 FLX plus	Illumina MiSeq	Illukina Next Seq 500	Illumina HiSeq 2500 RR	Illumina HiSeq 4000	Illumina HiSeq X
Reads: (M)	1,25	25	400	600	5000	6000
Read length: (paired-end*)	700	300*	150*	100*	150*	150*
Run time: (d)	0,9	2	1,2	1,125	3,5	3
Yield: (Gb)	0,7	15	120	120	1500	1800
Rate: (Gb/d)	0,75	7,5	100	106,6	400	600
Reagents: ($K)	6,2	1	4,41	6,145	29,9	12,75
per-Gb: ($)	8K	93	36,75	51,2	20	7
hg-30x: ($)	--	11160	4410	6144	2400	840
Machine: ($)	500K	99K	250K	740K	900K	1M
Platform	Ion Torrent 318 HiQ 520	Ion Proton P1	PacBio RS P6-C4	PacBio Sequel	MinIoN MK1	PromethION
Reads: (M)	3-5	165	5,5	38,5	0,05	--
Read length: (paired-end*)	200 400	200	15K	12K	10K	10K
Run time: (d)	0,37	--	4,3	4,3	2	--
Yield: (Gb)	1.2-2	10	12	84	2,75	3100
Rate: (Gb/d)	5,5	--	2,8	19,5	1,375	--
Reagents:($K)	0,6	1	2,4	11,2	0,5	2,5
per-Gb: ($)	--	100	200	80	180	20
hg-30x: ($)	--	12000	24000	9600	21800	2400
Machine: ($)	50K	243K	695K	350K	1K	75K

PacBio or Single-Molecule Real-Time Sequencing is a sequencing method developed in the past decade. It underlies the fact that the polymerase activity can be monitored without an interruption of the sequencing process. It bases on a method where each dNTP is tagged by a fluorescent. When the phosphate group is cleaved away by the polymerase, light is emitted. This can be detected by real-time imaging. This happens in special zero-mode waveguide microwells, and four to five bases per sequenced can be processed. With this technique, it is possible to create reads of long length with an average of 15 kb. The drawbacks of this method are the low accuracy and high demand in quality and fragment size of the DNA, as it does not require a PCR step (Bleidorn, 2017). Another long-read sequencing approach is the nanopore. This method is known since the 1990s and is based on an application of electric current, which forces the DNA through a nanopore. The movement through the pore creates specific variations of the electric current for each base. Although it relies on a completely different approach, it shares the high error rate of the PacBio method (Bleidorn, 2017). A comparison of all introduced methods can be found in table 1.

Concluding on all the presented methods, Sanger Sequencing is the most accurate method, while the NGS methods provide a much cheaper and fast alternative. It can provide information on the entire transcriptome and does not demand a highly complex and time-intensive DNA preparation. The limitations of NGS are a sufficient knowledge of bioinformatics to process the sequence data because of the short length of the reads. This can create problems in highly repetitive regions (Raza & Ahmad, 2016). The short-read sequencing methods have a different demand for DNA amount and quality, as they depend on a polymerase chain reaction step (PCR) (Tsuchihashi & Dracopoli, 2002).

The presented sequencing techniques are applied in different strategies. One of these strategies is shotgun sequencing. It is widely used for Illumina whole-genome resequencing and was already

developed in the early 1980s. Large fragments of DNA are cut into small pieces. After library preparation, random pieces of the fragments are sequenced until a redundant amount of sequencing data has been generated. RAD seq or restriction site-associated DNA sequencing is another method that produces a reduced representative fraction of the genome. It is based on restriction enzymes that digest the DNA at random or specific positions. The application of asymmetric cutters produces sticky ends. On these ends, adapters with known primer sequences can be ligated. Genotyping by sequencing is based on this method and can be applied with one or two restriction enzymes. It has proven to be of use for the identification of SNP in plant genomes (Bleidorn, 2017). Instead of sequencing the whole genome or a random fraction, exome capture is a more targeted approach. With it, only the coding regions of a genome can be sequenced. This leads to a drastic decrease in data while still retaining most of the coding regions (Raza & Ahmad, 2016). RNA sequencing can give the same kind of information, as only transcriptomes are sequenced. The RNA is a reverse transcript of CDS. Adapters are attached, and further library preparation steps are performed. Different RNA seq methods exist, allowing for the whole transcriptome sequencing as well as the sequencing of the 5' end. The method allows for the identification of gene expression variations but requires a robust experimental setup (Bleidorn, 2017). Another approach is the Chromatin immunoprecipitation sequencing (CHIP-seq). It allows identifying the identification of regulation events (Raza & Ahmad, 2016).

1.7 Quality of the reference genome, available information

Barley is a diploid species with a large genome of 5.1 Giga base size. The first assembly of the genome was published in 2012, where a genome-wide physical map of the cultivar Morex was presented. It consisted of a high proportion of repetitive sequences, where more than 79,000 transcript clusters were identified. 84% of these were reported to be comprised of mobile elements or repeated structures, mainly consisting of retrotransposons. A first estimate of the gene count was made, estimating a gene count of 30,400 genes on the genome. Within a chromosome confirmation study, the number of genes was corrected to 39,734 high confidence genes and 41,949 low confidence genes. Most of the high confidence genes show evidence of alternative splicing and consist of multiple exons. The later introduced linear sequence order of the barley genome on chromosomes made the evolutionary analysis of morphological and physiological innovations in the genome visible. By testing various cultivars, slight variation between these could be observed. This can be addressed as a signature of intense selection within breeding and little recombination due to the selfing habit of barley. Wild-types have proven to have a much higher diversity, leaving them as an essential source of genetic variation. Additionally, a decreased frequency of single nucleotide variants (SNV) can be observed towards the centromeres. Cluster aggregations revealed a big group of colocalized response defense domain proteins all over the chromosome with a hot spot on the short arm of chromosome 1H (Mascher et al., 2017; Mayer et al., 2012).

The chromosomally divided reference genome, as well as the functional prediction and position of barley, are under free access and can be used for any kind of genetic analysis. Furthermore, the web-based application *BARLEX* can be used as a database for any so far related information in respect to the reference genome. The reference genome is enriched with sequence assemblies and genetic maps. Additional information can be obtained from exome capture data, the information of annotated genes, as well as their expression profiles in 8 developmental stages (Colmsee et al., 2015). The reference genome is a high-quality resource for aligning information short read sequencing results directly to genes. Among others, single nucleotide variations, insertions or deletions, and expression patterns can be linked directly to genes. The location information added to this can give vital information on the expected recombination in the region.

1.8 Pool genotyping – Genotyping approaches for pools

Next-generation sequencing (NGS) technologies have become very popular in the past couple of years (Metzker, 2010). They have been applied to various studies to identify variations between genotypes on genomic and transcriptomic levels and can identify high numbers of polymorphisms, even without a reference genome. NGS can be applied for various tasks and has been promoted to be of tremendous use when it comes to pooled experimental designs. Especially when the number of samples to sequence becomes very high, pooling approaches can lead to a cost reduction of several magnitudes. Population studies, breeding process observation, evolution and adaptation detection, and allele mining approaches can capitalize from the application of pooled sequencing (Bélanger et al., 2016; Byrne et al., 2013). For small to medium size genomes, pool sequencing approaches were already applied to identify variations on gene-level between populations or treatments (Burke et al., 2010; Ehrenreich et al., 2010; Fischer et al., 2013a; Turner et al., 2010).

Studies published in the past couple of years validated different sequencing methods, including genotyping by sequencing (Anand et al., 2016; Bélanger et al., 2016; Gautier et al., 2013), pyrosequencing (Rellstab et al., 2013), RNA sequencing (Konczal et al., 2014) and whole-genome resequencing (Burke et al., 2010; Fischer et al., 2013b; Y. Zhu et al., 2012) approaches. Most of these had in common that the count of samples in the pool was relatively small, and the sequencing coverage of the pool was relatively high. High coverage for species with a small genome might be unproblematic in terms of costs. Species like barley or wheat, having a genome size of approximately 6 GB and 16 GB, double the sequencing depth results in severe cost inflation. This makes high coverage sequencing an issue. Besides the benefits of a pooling approach has, a couple of disadvantages exist as the information of the haplotypes and the heterozygosity gets lost. Some approaches to recover the haplotype information exist (Kessner et al., 2013; Long et al., 2011; Rode et al., 2018). The haplotype information of single genotypes has to be derived, and reads need to be rather long to estimate the haplotypes correctly. On top, the identification of rare alleles might be challenging (Rellstab et al., 2013).

Nevertheless, pooling approaches are useful in many fields, but the workflow is more sensitive to errors compared to single genotype sequencing. Four primary error sources have been highlighted. The first is the potential unequal contribution of individuals to the pool. None uniformity of individual contribution of DNA or RNA leads to a biased allele frequency estimation (Guo et al., 2013). The process level of pooling does not have a relevant influence on the uniformity of the sample. This leaves the choice for tissue or DNA/RNA pooling (Rode et al., 2018). The second source of error can derive from the sequencing depth. When the depth is low, it becomes difficult to estimate the correct allele frequency, especially when the frequency of an allele is low. Numerous studies gave recommendations on the minimum coverage to set, with the average suggestion ranging from 50 to 100 reads coverage per polymorphism (Schlötterer et al., 2014). In contrast to a static coverage level, Rellstab et al. proposed a minimum coverage adapted to the number of individuals in a pool and the ploidy level (2013). Another interesting approach tries to assess the correct allele frequency on low coverage levels with only 5x coverage per locus (Tilk et al., 2019). The third source of error derives from the sequencing, especially by the amplification step. PCR duplicates can lead to a substantial bias in the frequency estimation. This makes it necessary to remove duplicates from the data, which is not always possible (Gautier et al., 2013). The last main error identified is the pool size itself. Smaller pools with fewer individuals tend to have a more considerable variation of the individual contribution than bigger pools (Anand et al., 2016; Rode et al., 2018). The sequencing technique, however, has not been tested as a potential origin of error so far by any study.

Further requirements to give a decent overview of a population are a high number of tracked polymorphisms. These need to be uniformly distributed over the whole genome with a delicate coverage.

Furthermore, the number of individuals pooled should be high to cover the populational variation as precisely as possible. The accuracy of pooling approaches has been reported to increase marginally in accuracy when a level of 200 genotypes per pool is exceeded (Gautier et al., 2013).

1.9 Objectives and hypothesis of this study

Food security has been the major driver for industrial agriculture in the world and still is.
Alternative aspirations in agriculture, like the reduction of the carbon footprint or the conservation of ecosystems, have existed for a long time. But the majority of the global food production is driven by the objective of yield expansion. To reach that, the application of artificial fertilizer, pest, and weed control has dominated crop breeding in the past decades.
The biodiversity reductions, water pollution, loss of arable land, and the impact of agriculture on climate change are significant negative consequences of yield maximization strategies in industrial agriculture. Food production and the environment are in close interaction, and agriculture needs to obey ecosystem services.
Sustainability has become an essential topic for consumers. Therefore, the demand for it has continuously amplified over the last decades. Organic farming is one of the most popular ways to accomplish the goal of ecosystem service integration. A 20% yield reduction in organic farming compared to the conventional one is reported, conflicting with the need for global food production to rise substantially in the upcoming years. Crop breeding can have a significant influence on mediating the needs of yield increase and the ecosystem services. Up to date, organic farming is often conduced and limited by using outdated or not well adapted genetic material developed by conventional breeding strategies. So far, little is known about the requirements organic farming does have on the genetic material.
Focusing breeding programs to create and select genetic material tailored for organic farming can be the key to close or narrow the yield gap between highly intensive and sustainable farming. For that, the deep analysis of genetic constitution affecting plant resilience to (a)biotic stress factors is compulsory.
To gain closer insights, a long-term selection experiment was launched using a barley population created by crossing one cultivar with a wild species, backcrossed and grown for 25 consecutive generations under organic and conventional agricultural systems without performing further artificial selection. By gaining insights into the genetic composure and evolution processes, we wanted to increase the knowledge about the use of exotic lines in modern agriculture, the effect of selection pressure by pests and competition for light and nutrients, loss of unfavorable trait characteristics, adaptation to increasing climatic variation and drought stress. All these effectors have a direct influence on the fitness of the genotypes and the population as a whole, which may or may not lead to equilibrium after generations of adaptation.
This study tries to answer some questions about agricultural practice and the genetic requirements of different farming approaches. As indicated above, pool sequencing has been applied in several studies to identify variations between different populations. So far, no study has applied pool sequencing to crop species with big genomes. Pool sequencing is the designated tool to gain insights into population genetics on a genome-wide scale. As individual genotyping of a population or plant community comes short in either tagged loci or is far costlier, a reliable pool sequencing approach can reduce costs and cover many loci at the same time. With this base, questions regarding the genetic requirements of the cropping system shall be addressed. Conventional and organic farming follow substantially different approaches concerning fertilization, pest management and weed control. This could influence an interactively performing population, where beneficial alleles should lead to an advanced fitness and consequently to an increase of the allele frequency. The barley genome has almost 40,000 reported genes, from which some are expected to vary between the founders. The genetic variation can lead to a better performance in a given environment and thus an adaptation by an allele frequency shift. As these processes are expected to happen after some generations of selection, we expect to identify selection

sweeps by applying pool sequencing. The causal base of those sweeps, genes driving the selection, shall be identified and characterized. We want to see the influence of wild genetic material on the fitness of conventional and organic systems and characterize performance gains and variations between the systems.

The overall goal of this study is to identify special genetic requirements of organic farming. The assumptions are based on the different selection patterns of the organic and conventional systems.

Based on the formulated goals, hypotheses have been postulated.

- The allele frequency changes by none random effects are fastest in the first generations
- The majority of alleles contributed to the population originating from the cultivar parent, but wild alleles can add important alleles for conventional and organic farming systems – especially for the organic environment
- Double backcrossing leads to a decreased linkage drag, and the genetic diversity declines over the generations
- The organic system benefits more from gene related to nutrient uptake and tolerance to biotic stressors than the conventional system does, and the overall genetic diversity is higher in the organic population after years of selection

Chapter 2 – Material and Methods

2.1 Genetic compound - Founder and population

The experiment is based on an intercross of the two spring Barley genotypes Golf *(Hordeum vulgare vulgare)* and ISR42-8 *(Hordeum vulgare* ssp. *spontaneum)*. Golf is a former German elite breed selected from the cross (Armelle x Lud) x Luke (Landwirtschaft, 2013), while ISR42-8 is an undomesticated wild-type from Israel (Schmalenbach et al., 2009). Golf has been used as the mother plant, while ISR42-8 has acted as pollen donor. The initial cross was backcrossed twice. Therefore, Golf was used as a recurrent parent. The recommendations given by Cox (1984) were considered to reduce the risk of losing donor wild-type alleles during the backcrossing steps. Due to Mendelian law, the allele frequency of the wild-type should be 50% after the first cross. The two backcrossing steps should have reduced the wild allele frequency to 25% and finally 12.5%. The actually expected donor allele frequency should be approximately 11%, according to the finite size of the performed crossing. The double backcrossing was performed as the wild-type has not undergone domestication and therefore carries multiple unwanted alleles in it. Backcrossing can reduce the wild alleles in the population and eliminate linkage drags between beneficial and detrimental alleles. This way, more crossing over between the elite and the wild genetic material can be achieved. The crossing scheme is illustrated in figure 4.

Two rounds of selfing were performed in addition to the backcrossing to create a sufficient amount of seeds for a field experiment. The seed multiplication steps were performed in a greenhouse, where each individual was restricted to have the same amount of seeds in the multiplication to ensure the same number of progenies for each plant. By these steps, a BC_2F_3 was developed. One sample of this was used for long-term storage, while the second portion was sown out on a field plot of 9 x 15 m at a density of 330 and 360 for the conventional and organic systems, respectively. The annual population size of approximately 40,000 genotypes was the result of plot size and sowing density.

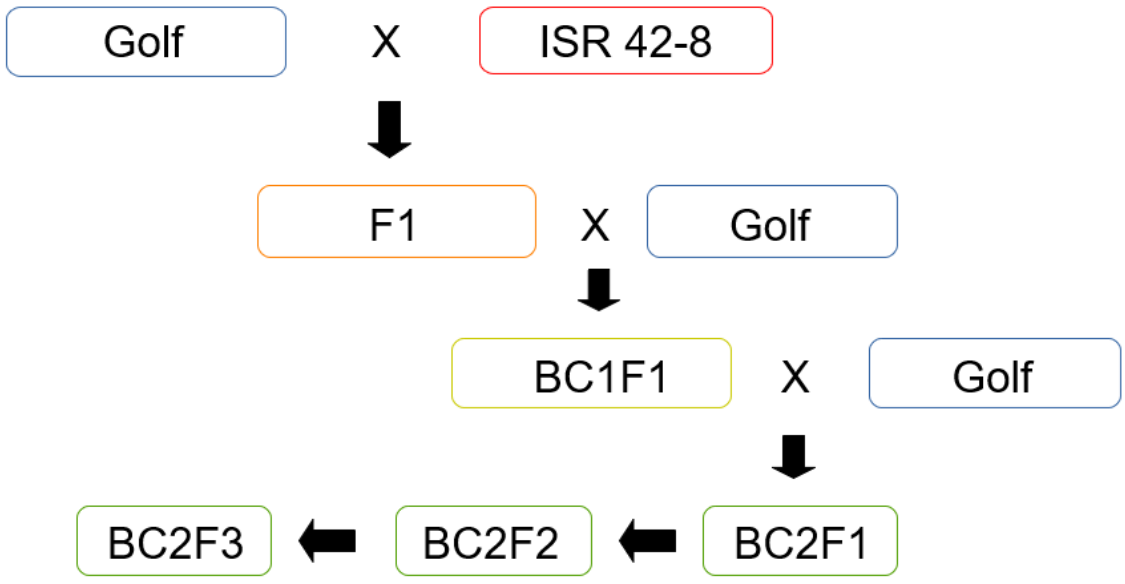


Figure 47: The crossing scheme of the tested spring barley population – a double backcross population with the donor ISR42-8, a barley wild-type from Israel, and the former elite cultivar Golf. Golf was used as a recurrent parent and mother plant. The wild allele frequency in the population should be about 12.5% for infinite population sizes due to the mendelian law.

2.2 Environmental compound - Crop rotation, treatment and plot design

The created double backcross population of Golf and ISR42-8 was sown to a field trial 1998 for the first time. In the period from 1998 to 2010, the experiment was located at the field experiment station Dikopshof (1997 to 2010; 50°48'22.6"N 6°57'05.5"E). Due to some policy decisions of the University, the station Dikosphof was replaced by the Campus Klein-Altendorf (2009 to 2019; 50°36'47.2"N 6°59'39.3"E). Based on this development, the experiment was moved to a new location at the Campus Klein-Altendorf, where it was established after a transition interval of 3 years. In this period, crop rotation was established. From 2011 until now, the experiment is located at the Campus Klein-Altendorf. The experiment remains in the same position for all the years despite the described change of location. This should ensure the adaptation of the soil to the farming system as well as avoid unintended contamination of the seed material.

Table 2: Treatments and application in the two contrasting cropping environments.

		Organic	Conventional
	Times of Application	once per year for Potato and Rape seed	Booting and Flowering Stage
	Amount added	200 dt/ha – 156kg N, 78 kg P2O5, 374 kg K2O, 36 KG MgO,*	80 – 100 kg N by calcium ammonium nitrate*
Growth Control	Seed Treatment	Seeds coated for sowing	-
	Growth Regulation	none	on demand[2]
	Fungicide Treatment	none	on demand[2]
	Competition Control	groom between rows	Herbicide on demand[2]
System Managed	Crop Rotation	Rape Seed – Winter Rye – Winter Wheat – Broad Beans – Oat – Spring Barley – Potato	Rape Seed – Winter Wheat – Spring Barley
	Soil Balance	pH, Nmin, Pmin, organic matter	pH, Nmin, Pmin, organic matter
	pH Adjustment	lime to pH of 7	lime to pH of 7

* – Phosphate fertilization was not required due to the high saturation level of the soil

2 - only products with a valid registration certificate of the year of use were utilized

In 1998, the same set of seeds were sown in an organic and a conventional environment. The environments are cultivated concerning good practice for the respective system, as described in table 2. For the organic system, this goes together with a crop rotation of seven crops. These crops are canola, winter wheat, oat, potato, acre beans, winter rye, and spring barley (Table 3). Manure is added annually to rapeseed in the autumn. Chemical weed or pest control is not applied. Instead, weeds are controlled by manual weeding until the crop has created a complete canopy. Contrasting to this, mineral fertilizer and chemical pest and weed control are applied in the conventional system. The crop rotation consists out of three crops rapeseed, winter wheat, and spring barley. The soil mineral content as well as the pH and organic matter are surveyed annually.

Table 3: Field trial design of the permanent selection trial. On the left, three parcels, each row belong to the same crop rotation. The studied population is population 1. The conventionally treated population can be found in row two (W2), while the organically treated population can be found in row 1 (W1). The position of the crop rotation remains the same over the years so that soil-borne contamination with seed material from other populations (P2-P4) is prevented. The parcels are 15 meters in length and 9 meters in width. The crop rotations are separated row-wise by a 9 meters long border. The border is free of any vegetation. The crops move in rotation every year, one slot to the right. [WW – winter wheat, SG – spring barley, WRa – winter rapeseed, Ka – Potato, AB – common bean, WR – winter rye, Ha – spring oat]

	conventional			**organic**						
	3 m border			3 m border						
W 4	**WRa P2**	**WW P2**	**SG P2**	**AB**	**WRa P4**	**SG P4**	**Ha**	**WR**	**WW P4**	**Ka**
	652	653	654	658	659	660	661	662	663	664
	9 m border			9 m border						
W 3	**SG P3**	**WRa P3**	**WW P3**	**WR**	**Ka**	**Ha**	**AB**	**WW P2**	**SG P2**	**WRa P2**
	636	637	638	642	643	644	645	646	647	648
	9 m border			9 m border						
W 2	**WW P1**	**SG P1**	**WRa P1**	**SG P3**	**Ha**	**WRa P3**	**WR**	**Ka**	**AB**	**WW P3**
	620	621	622	626	627	628	629	630	631	632
	9 m border			9 m border						
W 1	**WRa P4**	**WW P4**	**SG P4**	**Ka**	**AB**	**WR**	**SG P1**	**WRa P1**	**WW P1**	**Ha**
	604	605	606	610	611	612	613	614	615	616
	3 m border			3 m border						

93 m

15 m

9 m

90m

The plot design should prevent unintended outcrossing with other varieties or contamination of seed material (Figure 5). Therefore, the parcel of each crop in the rotation consists of three sub parcels of three meters wide rows of 15 meters in length. The seeds used for the coming generation are harvested from the center row. This row is harvested after the outer two rows to clean other remaining seeds from the combine harvester.

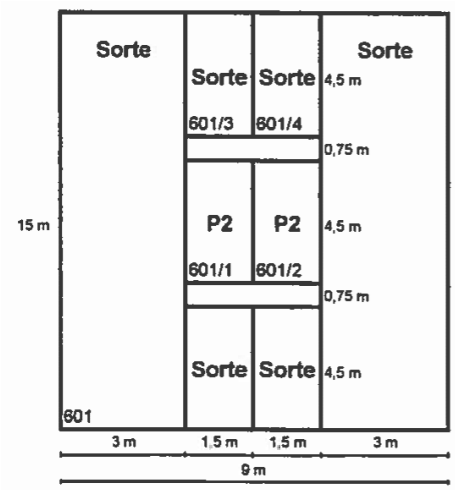


Figure 5: Design of a single plot with the three rows shown. The parcel has a size of 9 meters in width and 15 meters in length. The center part is used for seed production to avoid admixture with other barley varieties or populations by cross-pollination. The plot is harvested from the outside to the inside. This should reduce the seed contamination in the combiner with seeds from other trials.

The populations are grown from 1998 until now sequentially. This means that each year's seeds for the organic and conventional system were harvested separately, and these seeds were sown out in the following generation in the respective system.

2.3 Phenotypic characterization of populations and founders

2.3.1 Phenotyping and seed multiplication for parents

The phenotypic characterization of the founders and populations was performed in the greenhouse or a field trial. The parents' Golf and ISR42-8 have been phenotypes during a seed multiplication step in the greenhouse from July 2019 until February 2020. Morphological and phenological traits and the response to drought have been observed. Four plants were grown in a 25x25cm sized pot filled with a sandy substrate. Fertilizer and water have been applied to an optimum state to ensure maximum seed yield. Additionally, the resistance to *Puccinia hordei, Blumeria graminis,* and *Fusarium graminearum* was tested. The plants were grown until maturity, and variation in the phenotypes was noted.

2.3.2 Determination of root phenotype variation

We performed a hydroponic and field experiment on root morphological characteristics to complete the observations made on the molecular genetic level. 150 BC_2F_{24} genotypes of the conventional and organic environment were sown out in a hydroponic experiment, following instructions mentioned here (Hoagland & Arnon, 1938). PH was mediated to 5.9-6.0 every second day, and the nutrient solution was replaced once a week. One seed per genotype was grown for six weeks, and complete root and shoot material were harvested. Plants were grown under greenhouse conditions with 16h light and 18/12°C temperature regime. Besides root morphological traits, shoot height, the number of leaves and tillers were measured. Leaf number per tiller was used to correct for variant growth speed.

The field evaluation was conducted at Campus Klein-Altendorf Research Facility under rain-fed conditions without additional application of agrochemicals. A total of 100 barley BC_2F_{24} lines from organic and conventional populations were grown until anthesis. The roots of three plants per line were carefully harvested and soil was washed off, following the guidelines mentioned here (Neumann et al., 2010).

Root trait measurement was performed by WinRHIZO image analysis software (ver. 2008a, Regent Instruments Inc., Quebec, Canada). The root mass density was calculated according to Nakhforoosh et al. (Nakhforoosh et al., 2014). Root anatomical parameters were measured on cross-section slides of root tissue. A digital microscope (Keyence's VHX-1000D with 100X magnification) was used to create images of the section area. At least three to five roots per genotype were considered for measuring root anatomical traits. Captured images were then analyzed with ImageJ software (Schneider et al., 2012).

2.3.3 Field-based phenotyping of genotype offspring lines – representation of the phenotypic variation of the two populations

Field-based individual phenotyping was performed in the season 2020 for 500 organic and 500 conventionally evolved offsprings of genotypes harvested as a whole plant in the previous year. The conceptional idea behind this experiment was to elaborate anatomical, physiological, and morphological variations between the organic and conventional adapted populations. The hypothesis was that the mean phenotype would vary between the organic and conventional populations concerning the described upper and below ground phenotype.

Therefore, 11 to 45 seeds per genotype were sown out, depending on the amount of harvested seeds for the genotype in the previous generation. Scarlett and Golf were sown out additionally as control. ISR42-8 was excluded from the experiment as its growth period is longer than the vegetation period at the field location at Campus Klein-Altendorf Research Facility (50°370 N, 6°590 E), University of Bonn, Germany. This experiment was performed to measure and score multiple traits. Besides plant height in different growth stages, the emergency, score of different diseases impact, flowering time, chlorophyll leaf content, ears per plant, anther extrusion, ear size, growth type, lodging, tiller number were measured (Table 4).

The chlorophyll content was measured by a SPAD meter, other physiological parameters by a Polypen (RP 410, PSI (Photon Systems Instruments) spol. s r.o.), the plant height and ear length in centimeters, and the health status of the plants by the percentage of infected leaf area.

Table 4: Measured traits for upper body phenotypic characteristics

Score	**Date**	**Schemata**
Emergence	21.04.20	*Day of the year, ear fully pushed out of the shoot*
Emergence	24.04.20	*Day of the year*
Growth type	01.05.20	*1 to 9 – 1 plant close to the ground, 9 shoots straight up, tall plant*
Yellow dwarf virus	20.05.20	*1 to nine, nine high infection level*
Yellow dwarf virus	26.05.20	*1 to nine, nine high infection level*
Plant height	26.05.20	*cm from soil top to maximal height*
Flowering	26.05.20	*Day of the year; ear has to fully emerge from the stem*
Plant height	07.06.20	*cm length from stem to top row of seed*
Flowering	07.06.20	*Day of the year; ear has to fully emerge from the stem*
Flowering	13.06.20	*Day of the year; ear has to fully emerge from the stem*
Flowering	15.06.20	*Day of the year; ear has to fully emerge from the stem*
Powdery Mildew	21.06.20	*0 to 5, five high infection pressure*
Health status	24.06.20	*% infected leaf tissue*
Ear position	24.06.20	*1 straight up, 2 lodged*
Spad measurement	24.06.20	*Spadmeter*
Polypen	24.06.2020/26.06.2020	*all Parameters given by the sensor*
Ear position	30.06.20	*1 straight up, 2 lodged*
Ear length	30.06.20	*cm length from stem to top row of seed*

The yellow dwarf virus disease level was monitored on a scale from two to seven. The score of two was related to no infection at all, while a linear increase on the scoring line was followed. 3 and 4 were related to increasing yellowish coloring of leaves, while five was equal to an explicit infection, described by circular brown dots on the leaf tissue. A score of seven as the highest score was given for parcels where all plants were observed to be highly infected. Analogous to this, the infection with powdery mildew was scored from zero to five. The score of zero was related to no infection at all, while one was associated with plants that do show clear evidence of resistance. Genotypes were rated as resistant if the

genotypes showed apparent infection but no productive colonialization or sporulation by the fungus. Score 2 was related to low infection levels, characterized by black spots on yellow leaves. Score 3 was given if leaves were covered with black mycelia on mostly dead leaves. 4 and 5 are the same as 3, just on more and top leaves.

On a closely located field trial, the root morphology variation at the flowering stage of 100 organic and 100 conventionally evolved genotypes was tested. The plants were harvested as a whole plant, including the roots. Therefore, the plant was carefully dug out of the soil. Similar measurements as described for the hydroponics were performed. Hydroponic and field experiments on root morphological characteristics were performed to complete the observations made at the genetic level. In a hydroponic experiment, 150 BC2F24 genotypes of the conventional and organic environment were sown, as described previously (Hoagland & Arnon, 1938). The soil pH value was adjusted to 5.9–6.0 every second day and the nutrient solution was replaced once per week. One seed per genotype was grown for six weeks, and the complete root shoot materials were harvested. Plants were grown in a greenhouse under a 16-h light/8-h dark cycle at a constant temperature of 12°C. Besides the root morphological traits, the shoot height and the numbers of leaves and tillers were measured. The leaf number per tiller was used to correct for variations in growth speed.

Phenotyping was performed as described in the previous section. For both experiments, no additional fertilizer and no weed or pest control were performed.

2.3.4 Statistical evaluation of phenotypic characteristics

The phenotypic traits assessed were analyzed by a classic test of variation (ANOVA). Tukey Test was performed to obtain the variation between the organic and conventional set of genotypes and the parents ISR42-8, Golf, and Scarlett. Scarlett was used as a check variety. A *Pearson* correlation was calculated between the traits measured, and a principal component analysis was conducted on either all traits or only selected traits. Furthermore, the variation of the groups was estimated. A binomial distribution in a generalized linear model was applied for binomially assessed traits like the ear position.

2.4 Weather information and evapotranspiration calculation

Phenotypic data for all generations is not always available, and if it is, then in an unpleasing depth. The weather data can give additional information. It can answer questions regarding physiological stress phases in the vegetation period. The plant was stressed by a fungal attack more severely in years with constant or high precipitation, while it was confronted with drought stress in years with little to low precipitation. Drought stress or fungal pressure can occur in various developmental stages and intensities. For example, high moisture and precipitation events during and after flowering can result in severe ear blindness caused by *fusarium*, while a wet period in the early growth stages might result in a significant infection with powdery mildew. The same holds true for drought events. Such events can lead to yield reduction on various levels, like tiller number reduction or reduced corn filling. It might be possible to estimate the infection pressure just by basic weather data like moisture content, air temperature and precipitation. When it comes to identifying drought stress events, unprocessed weather data will not result in appropriate estimates of drought events. A more robust method to estimate such things is an advanced evapotranspiration model that combines weather information, together with plant physiology parameters and soil water content. Using these three information sources, the estimate can still be not as precise as measuring the actual evapotranspiration.

Evapotranspiration is the combined water loss from soil by evaporation and from the plant by transpiration. Evaporation describes the process whereby liquid water is removed from the evaporating surface (vapor removal). Water evaporates from various surfaces like lakes, rivers, pavements, soils, and wet vegetation. Plants cannot utilize evaporated water. Evaporation requires energy to transform

water from its liquid to its vapor form. The origin of this energy required for the transformation can be found in solar radiation or air temperature. The extend of evaporation is also limited by the vapor pressure of the atmosphere. The higher the atmospheric pressure, the more energy is required to vaporize liquid water. Transpiration consists of the vaporization of liquid water contained in plant tissues and the vapor removal to the atmosphere. Plants predominantly lose water by stomata. The water is taken up together with nutrients by the roots and is transported to the upper tissues, where it follows the vapor pressure gradient towards the atmosphere (Zotarelli & Dukes, 2010).

Nevertheless, an adequate estimation of the correct actual evapotranspiration is possible by applying the dual crop coefficient model proposed by the *FAO*. Pozníková et al. (2014) have validated this model based on real field measurements of spring barley evapotranspiration rates. They found a high correlation between the measured and calculated evapotranspiration of > 0.9, indicating that the dual crop method can produce reliable estimates of the actual evapotranspiration (Pozníková et al., 2014). The evapotranspiration model of the *FAO* is based on the *Penman-Monteith equation* (Equation 1) to calculate the evapotranspiration ET_0 of grass under optimal water supply.

$$ET_0 = \frac{0.408\Delta(R_n - G) + y\frac{900}{T+273}u_2(e_s - e_a)}{\Delta + y(1 + 0.34u_2)}$$

Equation 1: Penman-Monteith equation for calculating the reference evapotranspiration of grass in mm day^{-1}. R_n is the net radiation [MJ m^{-2} day^{-1}], G is the soil heat flux [MJ m^{-2} day^{-1}], T the air temperature at 2 meters height in °C, u_2 the wind speed at 2 meters height [m s^{-1}], e_s the saturation vapor pressure [kPa], e_a the actual vapor pressure [kPa], e_s-e_a the saturation vapor pressure deficit, Δ the slope of the vapor pressure curve [kPa °C^{-1}] and y the psychrometric constant [kPa °C^{-1}] (Zotarelli & Dukes, 2010).

Based on this, the crop evapotranspiration ET_c can be calculated (Equation 2). Two ways can give the ET_c value, while the dual crop coefficient used here is respected to be more accurate. The dual crop coefficient divides the evaporation from the transpiration and calculates both separately. This model is based on an irrigated cropping system or a system where water is not in shortage (Zotarelli & Dukes, 2010). Therefore, a soil water balance has to be included in the estimation. Reduced water availability will result in physiological responses of the plant-like increased stomatal conductance to reduce water loss (F. Liu et al., 2003).

$$ET_c = (K_{cb} + K_e)ET_0$$

Equation 2: crop evapotranspiration calculation by the dual crop coefficient $K_c = K_{cb} + K_e$. The procedure is based daily and split in a basal crop coefficient (Kcb) and a soil evaporation efficient (K_e). The basal crop coefficient (K_{cb}) is defined as the ratio of the crop evapotranspiration over the reference evapotranspiration and represents the transpiration component primarily. Water is not a limiting factor. K_e describes the evaporation component of ET_c. K_e depends on the wetness of the topsoil (Zotarelli & Dukes, 2010).

Including a soil-water balance allows for the calculation of the actual evapotranspiration ET_a. ET_c requires the calculation of the basal crop coefficient K_{cb} and the evaporation component K_e. K_{cb} is calculated as illustrated in equation 3. K_{cb} depends on the implementation of the crop growth stage prediction. The FAO model characterizes five Stages. These are the initial stage, crop development, anthesis stage (mid-stage), ripening, and maturity (end). The initial, mid, and end-stage are characterized by constant values for the crop stage coefficient $K_{cb(Tab)}$ (0.15,1.1,0.15), while the development and ripening stages follow linear regressions, ranging from 0.15 to 1.1 and 1.1 to 0.15 for development and ripening stage, respectively. The thermal time for the initial phase has been set to 200° days, development stage to 650° days, anthesis stage to 1,300° days and ripening to 1,650° days. These values mark the end of the stages. The limits of the stages have been selected based on the mean vegetation

period length reported by sowing and harvesting dates. The actual growing degree days have been calculated as the sum of the daily mean for all days of the vegetation period before the i^{th} day.

$$K_{cb} = K_{cb(stage)} + [0.04(u_2 - 2) - 0.004(RH_{min} - 45)]\left(\frac{h}{3}\right)^{0.3}$$

Equation 3: calculation of the basal crop coefficient where RH_{min} is the value for daily minimum relative humidity, u_2 is the mean value for daily wind speed at 2 m height over grass, and h is the mean plant height in meters, which has been set to 1 meter (Zotarelli & Dukes, 2010).

The soil evaporation K_e describes the evaporation component. It is defined as described in equation 4. The total evaporable water from the topsoil has been set to a value of 22 mm, while the value at the start of the vegetation period was set to 9 mm. The sowing was always performed in moist but not wet soil, so the assumption that the top soil is dried out has been made. The soil evaporation calculation was performed each time after a major rain event with more than 3 mm rain in 24 hours occurred.

$$K_e = minimum[K|r(K_{cmax} - K_{cb}) \vee f_{ew}K_{cmax}]$$

$$where\ K_r = 1\ \ if\ D_e < REW$$

$$else\ K_r = \frac{TEW - D_e}{TEW - REW}$$

Equation 4: calculation of the soil evaporation coefficient K_e. K_e is the soil evaporation coefficient, K_{cb} is the basal crop coefficient after equation 3, $K_{c\ max}$ the maximum limit of evaporation and transpiration following rain or irrigation, f_{ew} the fraction of wetted soil, and K_r the dimensionless evaporation reduction coefficient that depends on the depleted water from the topsoil. K_r depends on the total evaporable water in the 10cm top soil TEW [mm], the readily evaporable water on day$_i$ [mm], and the cumulative depth of evaporation D_e [mm]. D_e depends on rain events (Zotarelli & Dukes, 2010).

The exposed and wetted soil fraction f_{ew} is calculated as shown in equation 5. Based on the calculations, the lowest value is selected for soil evaporation.

$$f_{ew} = 1 - f_c$$

$$where f_c = \frac{K_{cb} - K_{cmin}}{K_{cmax} - K_{cmin}}^{1+0.5h}$$

Equation 5: calculation of the exposed and wet soil fraction $f_{ew.}$ The average exposed soil fraction f_c is calculated by the plant height h and the crop coefficient.

Having calculated K_e and K_{cb}, the crop evapotranspiration ET_C can be calculated for each day. As already mentioned, the evapotranspiration calculation is based on a regular water supply, resulting in enough water for the plant to remain in full photosynthetic activity mode. As the experiment was only based on irrigation from rainfall, water shortage in some periods might have occurred. When the utilizable soil water decreased below a field capacity of 30%, the plant initializes water-saving processes. For a file capacity bigger than 30%, the actual evaporation ET_A equals the potential crop evaporation ET_C. In case the field capacity is lower than 30%, the transpiration is decreased based on linear regression, and the ratio of ET_A /ET_C decreases towards 0. At approximately 10% field capacity, the wilting is reached. At this point, ET_A equals 0. In mid-European areas, the field capacity level of 10% is not likely to be reached, even in periods dominated by dry and hot weather (David J. Connor, Robert S. Loomis, 2011; Meyer & Green, 1980). Based on measurements performed by Pätzold et al. (2005), the field capacity of the field trial was measured to be 25vol%. The rooted area has been defined as 75cm of soil depth. This resulted in a total field capacity of 180mm. The main soil fraction is silt and the soil type is a *Parabraunberde (Inceptisol)*(Pätzold et al., 2005). The evapotranspiration level is adjusted to the

available soil water and gradually reduced if the soil water content decreases below a level of 30% usable field capacity.

The required weather data was collected for the whole period by a weather station located in the direct proximity of the filed trial. Daily mean values were calculated from measurements made every 10 minutes. The weather station measured air pressure, radiation, humidity, air temperature in 2 meters height, windspeed, the direction of wind, temperature of the top soil, photosynthetic active radiation and precipitation. As the trial was transferred permanently from the Campus Dikopshof to Campus Klein-Altendorf, the weather data of these two stations was used. For the first period until 2010, weather data was used from Dikopshof, while the weather data for the second period was used by the station of Klein-Altendorf.

Performing all these steps enables an approximate estimation of the evapotranspiration of the barley populations and elaborates the climatic variations over the years.

2.5 Assessment of genetic diversity by pool sequencing

As described above, the annual population size of each year and system was calculated to be 40,000 genotypes. Testing the whole set of genotypes for each environment and year is far from possible. A representative sample had to be drawn out of this population. By this sample, the wild-type allele frequency should be estimated on the whole genome. This required an extensive set of identified polymorphisms. Sequencing many loci to get a representative genotyping output demands a proper tool. Two possible ways can be followed. Genotyping by a chip with distinct loci is one way of characterizing a population. This method tests each sampled genotype for thousands of designated polymorph loci. The second method is related to sequencing the genome of these samples in a more untargeted manner. Methods like RNA sequencing or whole genome sequencing are genotyping by sequencing approaches, following a more untargeted way of genotyping a sample. While genotyping by a chip is limited by genotyping single genotypes on designated locations, genotyping by sequencing can deliver information on unspecified loci for multiple genotypes in one sample. Through the sequencing coverage of a locus, this method has a quantitative component. This allows for the calculation of frequencies directly from the sequencing data. The sequencing of pool samples can utilize the significant advantage of this quantitative information. A pooled sample consists of genetic material derived from multiple genotypes. This makes it possible to gain insights into a set of genotypes by sequencing only one single sample. Establishing a reliable and cost-efficient procedure to access the allele frequency of an enormous population has been defined to be one major goal of this study. The established procedure is based on the steps described in the following sections.

2.5.1 Sample size estimation for pools

The major problem for uncovering the total genetic diversity of a population lies in testing the accurate sample size. We aimed to hit a sampling error (precision level) of 0.05 with a low-risk level (confidence interval) of 0.99. To calculate the necessary minimal sample size, we applied Cochran‘s Formula to estimate the correct sample size, shown in equation 6 (Israel, 1992).

$$N_0 = \frac{Z^2 * p * (1 - p)}{e^2}$$

Equation 6: Cochran's Formula to calculate an ideal sample size given a desired level of precision to test a large population by a representative sample

N_0 is the sample size, Z is the z-score of the confidence interval, p is the preliminary information about the allele frequency of *ISR42-8* in the population, and e is the desired level of precision. Based on an

expected allele frequency of 12.5% of the donor alleles, we end up with a sample size of 291 genotypes per generation and environment. To round up, we used 300 genotypes.

2.5.2 Sampling genotypes

The pool samples were collected for selected generations of the permanent selection trial. Seeds of the initial, unselected generation F_3 and late selected generations were chosen for both conventional and organic systems. The late stages include the F_{12}, F_{16}, $F_{21,}$ and F_{22} from the years 2007, 2011, 2017 and 2018. Two samples with different sets of genotypes were collected for each generation. For the F_{22} conventional population, one additional sample was collected. In addition to the pool samples, the two parents Golf and ISR 42-8 were collected, and DNA was extracted for the two genotypes separately. These two should build reference haplotypes for the population.

For the F_{22} conventional population, one set was collected that contained 288 genotypes instead of 300 genotypes. Leaf tissue for these 288 genotypes were collected separately for each genotype as well. This set should be used for the pool validation. DNA was extracted from these genotypes individually from KASP array sequencing on designated polymorph loci chosen based on the pool sequencing results.

The DNA extraction was performed using fresh leaf tissue. Seeds of the generations have been sown out on pit soil in a greenhouse at 22°C with a 16/8-hour day-night setting. The leaf tissue was harvested seven days after sowing out.

Preparing a pooled sample has some pitfalls. One of these happens right at the sampling. As Sham et al. (2002) have stressed out, pooling is an effective way to reduce the costs in large-scale genotyping studies, but it requires adjusted amounts of DNA/RNA for each sample (Sham et al., 2002). The identical amounts of DNA should result in an equal contribution of genetic material for each genotype in the pool. Two ways of pool sampling were described. A sampling of DNA/RNA required an individual DNA/RNA extraction for each genotype separately. After this is accomplished, the same amount of DNA for each genotype is added to a pool. This method is the most commonly used but can result in problems as well. These methods do not assess the DNA/RNA quality. Variant degradation can result in unbiased sequencing probabilities. Because of this and the fact that individual DNA/RNA extraction goes along with a tremendous amount of resource, cost and labor, we decided to pool the samples on the tissue level.

Equal amounts of leaf tissue for each genotype should be added to a pooled sample. We accomplished this by using a hole plier. This cut a leaf fraction for each sample of approximately equal size. Three hundred genotypes per pool have been sampled. Tissue samples were collected in rounds of 60 to 80 genotype tissue samples to prevent DNA/RNA degradation. The samples were collected in a bowl and placed in a container with ice to ensure a decreased speed of degradation. At the point where 60 to 80 samples were collected, the leaf discs were carefully transferred to a 5 ml Eppendorf container. This container was placed in a rack on a vessel that contained liquid nitrogen. The filled-in leaf discs were instantly frozen and stored afterward in a -80°C freezer.

The DNA and RNA extraction was performed by using the *peqlab plant DNA/RNA Mini Kit.* Prior, the samples have been vigorously homogenized by a vibration mill. Six metal grinding balls have been added to the 5 ml Container. The metal rack in which the 5 ml containers were placed allowed a simultaneous homogenizing of up to 20 samples. The rack itself was chilled with liquid nitrogen prior before been loaded with the containers. The samples were homogenized for 6 minutes at 25,000 U/s. The grinding balls had a diameter of five and three millimeters. Three of each were used. The tissue needed to be crushed to a very fine powder to ensure non-biased sampling. Following, 15 µg of this tissue were extracted from the container for DNA and RNA extraction. For the extraction, we followed

the instructions in the *peqlab plant DNA/RNA Mini Kit.* The quality and amount were checked on a 1.5% agarose gel as well as by Nanodrop measurement.

2.5.3 Sequencing approaches

Literature highlights the successful application of genotyping by sequencing, whole-genome sequencing, and RNA sequencing for pool genotyping. We compared these three common sequencing methods for their use in pool sequencing of barley. Each has advantages and disadvantages regarding coverage, costs, or polymorphism yield, as illustrated in table 5. Three replicates for the F_{22} conventional population were sequenced using MACE RNA sequencing (Zawada et al., 2014) and genotyping by sequencing (GBS) (Poland et al., 2012). Two replicates of this population were sequenced using a whole-genome resequencing approach (Belkadi et al., 2015). One of these sets was sequenced as a pool and individually for 21 SNP-specific competitive allele-specific PCR **(**KASP) assays (followed referred to as KASP marker) - three for each chromosome covered the short and the long arm as well as the centromeric region. The markers for single genotype testing were selected based on the MACE RNAseq sequencing results. Representative markers for the whole genome were selected. This pool was sequenced by all three different techniques. The applied genotyping by sequencing (GBS) approach is based on a rare and frequent cutter enzyme combined with paired end-setting of 145bp long reads. Besides this, another untargeted DNA-based sequencing approach was applied by using whole-genome resequencing (WGS), resulting in 150 bp long paired-end read output. Lastly, we validated a transcriptional approach by MACE RNA sequencing with a single-end read yield of varying lengths. MACE sequencing was chosen instead of classic RNA sequencing because it claims to deliver higher coverage rates on fewer loci costs and is based on the TRUEQuant method that can identify PCR duplicates. The methods vary in costs, but none of them exceeded a value of 600€ per sample.

Table 5: Overview of strengths and weaknesses of the applied methods on different levels. MACE = transcriptome sequencing; GBS = enzyme-based genotyping by sequencing; WGS = whole-genome resequencing, (Schlötterer et al., 2014)

	MACE	**GBS**	**WGS**
Costs	med	low	high
Genes covered	yes	partly	yes
Level	RNA	DNA	DNA
SNP count	low	med	high
Coverage	varying	med	low
Genome representation	biased	random	all
potential for Sampling error	high	low	low

2.5.4 Mapping and Polymorphism detection

All reads were mapped to the *Hordeum vulgare* reference genome *top-level v2* (downloaded 30.05.2019 (Mayer et al., 2012)) using *bwa mem* (version 0.7.1 (Li, 2013)). Sequences were trimmed if the read quality required it (Bolger et al., 2014). Aligned reads were strictly filtered by *sambamba view* (version 0.6.6 (Tarasov et al., 2015)), only retaining single mapped reads with a mapping quality higher than 30.

sambamba_v0.6.6 view -h -f bam -p -F "mapping_quality >= 30 and not (unmapped or secondary_alignment) and not ([XA]!= null or [SA]!= null)"

Following this, the duplicates were removed for the MACE RNA seq and WGS data sets by *sambamba makrdup* and then sorted by *sort.* Variant calling was performed using *samtools mplieup and bcftools call (*version 1.8 (Li et al., 2009)) on minimum variant base quality of 25 phred score. Only single

nucleotide polymorphisms (SNP) were called, and Indels were omitted from calling. Polymorphisms between both parents were identified and tested for presence in the reference SNP database for *Hordeum vulgare*. By retaining only SNP that have already been found in this database, the chances of falsely discovered SNP should be reduced. If a SNP was not listed in this, it was deselected to omit false-positive SNP calls. The sequence quality was obtained by *fastqc* (Andrews, 2010). The SNP calling reveals various information from which the reference and alternative base count are used to calculate the allele frequency for each polymorphism. In combination with the knowledge of whether the reference or alternative allele describes Golf or ISR42-8, the allele frequency of these can be calculated in the pooled sample, as illustrated in equation 7. The allele frequency analysis was performed only on single nucleotide polymorphisms, and deletions or insertions are omitted.

$$AF_{ISR42-8} = \frac{RDAllele_{ISR42-8}}{RDAllele_{ISR42-8} + RDAllele_{Golf}}$$

Equation 7: Calculation of the allele frequency of the parental alleles in the pool progeny sample. The read count (RD) of the parental allele is extracted (either the reference base or an alternative base) and divided by the sum of reads covering the locus at a given quality threshold. This is performed for each polymorphism separately.

2.5.5 Haplotype estimation

Based on the fact that the population has two founders (Golf and ISR42-8) and their haplotypes were known, this information can be incorporated into the pooled sequencing of the population. On short to medium distances, we assumed that recombination should not have occurred. Gene annotation derived from the IPK database with positional annotation for low and high confidence genes was used as an anchor to generate haplotypes with a direct link to a function (https://webblast.ipk-gatersleben.de/barley_ibsc/). 80,554 low and high confident genes were listed and utilized as the potential origin of a haplotype. From this set of genes, only the high confident genes were used for haplotype construction. The gene bounds were extended on both 3' and 5' ends to include close-by mapped SNP. The extension algorithm extended the gene by 45 percent of the intergenic region size if the latter was bigger than 1,000 bp. A gap of 10 percent the size of the intergenic region was not annotated to a gene. Exemplarily, gene Q with position 200 to 500 and gene W with position 800 to 850 are both expanded. The intergenic region has a size of 300 base pairs. Ten percent of this, 30 bp, are removed, leaving 270 bp. This is divided by 2, giving 135 bp. The new end position of gene Q is 500 + 135 = 635, and the new start position of gene W is 800 – 135 = 665. The gap between these two genes has been closed from 300 to 30bp. If a SNP is at position 590, it is now considered as part of Gene Q. Without the extension, the information of this particular SNP would have been lost for the improvement of the haplotype frequency. Further details can be obtained from figure 6.

If genes overlap, no extension was performed in the direction of overlap for both genes. Besides the genes, markers present in the 9K SNP chip were mapped against the reference genome, and a range around them was defined as a haplotype region. According to the same procedure as highlighted for the genes and shown in figure 6. 6,080 markers could be mapped to the reference genome. The extension for these has also been performed, analogous to the genes. All reads within a haplotype were used to calculate the haplotype allele frequency (HAF),

$$HAF_p = \frac{\sum rd_k * freq_{pk}}{\sum_k rd}$$

Equation 8: Calculation of the haplotype frequency HAF of parent p. rd is the coverage of a SNP k, while freq is the frequency of SNP k for parent p.

where *p* represents the allele for a specific parent. *k* is the SNP, for which the coverage *rd* and *freq* allele frequency for SNP *k* and parent *p* is given (Equation 8). Marker information comes along with the genetic position, which can be used here because it includes genetic distance estimates.

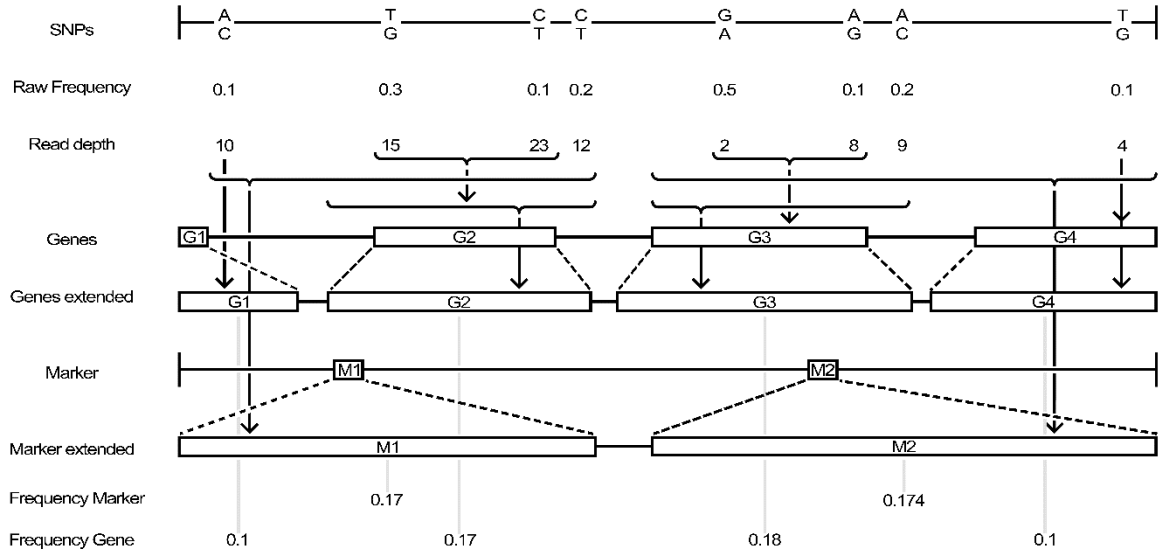


Figure 6: Scheme of haplotype allele frequency estimation by aggregating allele frequency information of multiple SNP to one single haplotype allele frequency. The haplotypes are regions on the genome that are surrounding a marker or a gene. The haplotypes based on genes have been expanded regarding the neighboring genes into the intergenic regions. SNP located outside coding sequences were utilized by this approach to increase the coverage and frequency estimation accuracy. It is assumed that linkage disequilibrium is a factor and SNP of a given region are linked. The same procedure was applied for haplotypes based on markers. Markers were derived from the 9Kiselect chip. Contrasting to a gene, a marker does not cover a region but only a single SNP. For the marker-based haplotypes, single SNP positions were replaced by a piece of region information. The regions aimed not to overlap each other and have a gap between them.

The haplotyping approach applied is comparable to the method used by Tilk et al. (2019). They used a constant window size, while the approach followed in this work is based on a variable haplotype size. The window size can vary over the genes, depending on the size of the intergenic region. Aggregating frequency information of multiple SNP was reported to increase the accuracy of the frequency (Tilk et al., 2019).

2.5.6 Pool accuracy estimation

KASP markers were designed to estimate the precision of the applied pool sequencing. The individual genotyping was performed for 21 designed *KASP* markers. These have been selected based on *MACE RNAseq* results. The requirements for a polymorph locus to be used as a *KASP* marker are based on reading depth in the *MACE RNAseq*, the missing of other variants closely neighboring, full coverage of the genome, where three markers for each genome at the short arm, pericentromeric region, and long arm are used as well as a decent coverage of all kinds of allele frequencies represented. The allele frequency location of a single *KASP* SNP and a haplotype is expected to be the same if the *KASP* SNP is located inside the boundaries of the haplotype.

Two hundred eighty-eight genotypes, representing one pool, were individually genotyped with these KASP markers, and the allele frequency was compared against the pool with the same genotypes on GBS, MACE and WGS sequencing levels for single SNP, gene-based and marker-based haplotype approaches. Based on a zero-inflated negative binomial distribution, a linear model was applied as a statistical model to compare replicates and sequencing methods. For this, we used the *pscl* package (Jackman et al., 2013) in R 3.5.2 (R Core Team, 2020). Furthermore, the *Pearson* correlation and the

root means square error (equation 9) between individually genotyped and pool genotyped samples and between the replicates and the sequencing methods was calculated.

$$RMSEestimated = \sqrt{\frac{\sum((pestimated|-ptrue))^2}{n}}$$

Equation 9: The squared sum of the observed allele frequency substrate by the derived true allele frequency divided by the sample size and square rooted (Tilk et al., 2019).

To investigate the potential unintended bias impact of the chromosomal position and the extension size, we applied a mixed linear model of R package lme4 (Bates et al., 2015), as follows

$$mlm(|pestimated - ptrue|(extension + position))$$

Equation 10: Estimating the impact of the haplotype extension by a mixed linear model, where the position is fixed and the extension a variable factor

A haplotype or contig was associated with a KASP marker if the ladder was located inside the boundaries of the constructed haplotype.

2.6 Algorithm for the most accurate frequency extraction

The entire analysis, including haplotype construction, allele and haplotype frequency estimation, was performed in *Julia* (version 1.3.0, (Bezanson et al., 2017)), where a minimum coverage level was set to one read, and the minimum SNP quality score of *mpileup* needs to be above 30 for each used polymorphism. The allele frequency estimation for the populations was based on polymorphisms detected between the two parents of the populations.

The code for the allele frequency analysis was designed to be published as a *Julia* package. With this intention in mind, the code should work on various kinds of data sets. Different generations, environments, replications, and sequencing methods can be applied to the code, resulting in an advanced variation analysis based on polymorphisms. To perform all analyses, the *Julia* packages *DelimitedFiles, DataFrames, Statistics, CSV, ProgressMeter, HypothesisTests, GLM, Crayons, Gadfly, Compose, CuArrays, RCall, StatsBase,* and *Cairo* were required.

2.6.1 Required input files

Three files were required as input to accomplish allele frequency calling, information aggregation, and cross-comparison. The first file is a *vcf* file. All polymorphisms called were stored in this file for all the samples included. This file, generated by the *samtools* code shown below, produces information on each identified polymorphism. Insertions and deletions have been excluded from the frequency calling. This file type gives information on the SNP position on the reference genome, the reference and alternative base, and a quality score for the obtained SNP. A reference base was defined as a base found in the reference genome at the particular position, while the alternative base is found in the tested sample for the same position, instead of or in addition to the reference base. This quality score ranges from 0 to 999. An additional *INFO* column gave an overview of reading depth, average mapping quality, the number of reference and alternative base calls and much more. By applying the *–output-tags AD*, each aligned sequence file gets a unique column with information on the genotype, phred-scaled genotype likelihoods, and the number of called reference and alternative bases (Anonymous III, 2020). The latter mentioned information is essential and used to generate the allele frequency for each sample tested. Reads of a sample were included if the quality of the read was higher than a phred score of 25 (maximum

is 35), and the base of interest had to have a minimum quality of 30. By these high numbers, secondary or mismatch alignments should be omitted from further downstream analysis.

$$samtools mpileup - uvf\ reference_{genome}.fa - I - q25 - Q30 - t\ AD \\ - -positions\ Positions_{tocheck}\ file1.bam\ file2.bam \dots fileX.bam \mid bcftools call - vm \\ - Ov > output.vcf$$

The second file is an overview file of the samples tested (Table 6). It is an additional file to the variant calling file and stores information of the filename, the environment (conventional or organic) and the generation. The parents are also included in the table and presented as a zero value in the environment and generation. The information given in this table is used in further progress to classify parental alleles, separate environments, and generations and identify replicates. It is applied in various functions of the code to order the files and create reasonable output.

Table 6: Overview of the included files in the variant calling file (vcf) in the same order as it has been used in the vcf file. The first column gives the names to the dictionary entries in Julia*; the second column gives information on the environment: 0 is used for the parents, one is used for the conventional environment, and two is used for the organic environment. The third column gives an overview of the rested generations, where the parents again are displayed by 0.*

Name	*Env*	*Generation*
Golf	0	0
Isr	0	0
F3P1K1	1	3
F3P1K2	1	3
F3P1O–1	2	3
F3P1O–2	2	3
F12P1K1	1	12
F12P1K3	1	12
F12P1O–1	2	12
F12P1O–2	2	12
F16P1K1	1	16
F16P1K2	1	16
F16P1O–1	2	16
F16P1O–2	2	16
F22P1K1	1	22
F22P1KEP	1	22
F22P1O–1	2	22
F22P1O–2	2	22
F23P1K1	1	23
F23P1K2	1	23
F23P1O–1	2	23
F23P1O–2	2	23

The third file is a reference file with information on genes or markers. Two different types of files exist and can be used in the code to generate information on or gene allele frequencies. The marker reference file is based on the sequence alignment of the marker sequence against the reference genome. The markers originate from the barley 9K iSelect chip (Comadran et al., 2012a). The obtained position was

linked to a region on the genome. Based on this, the marker reference file delivers information on the marker name, the chromosome and base position in base pairs and centi Morgan. The other reference file contributes information on genes. 39,734 high confidence genes were included in this list. Besides the name of the gene, a position on the genome was defined and a potential function, GO term identifier, and PFAM and InterPro IDs. This reference was derived from the IPK blast website *(https://webblast.ipk-gatersleben.de/barley_ibsc/downloads/)* and has been described elsewhere (Mascher et al., 2017).

2.6.2 Model for calculation variations of backcross population variations

As the SNP allele frequency distribution varied widely over different genes, we applied a dynamic algorithm to select the appropriate test statistic. A Mann-Whitney-U test was applied if the sample size was smaller than 30 and a proper estimate of the distribution could not be done or if one of both test samples consisted of nothing but zero values. If less than 15% of all values were zero or one of both test samples did not have any zero AF entries, a generalized linear model based on a negative binomial distribution was used to calculate statistics. The most significant fraction of genes was characterized by a zero-inflated, negative binomial distribution. All those genes were analyzed by *pscl* R package implementation in *Julia* if less than 75% of all values were zero. Lastly, if the zero counts exceeded a ratio of 75%, a zero-inflated *Poisson*-based linear model was applied. *Pscl* package calculated two test statistics for both Poisson/negative binomial and zero-inflated statistics, separately. Therefore, both probability values were extracted and saved in the statistics summary table along with the information of the used test statistic, the average haplotype allele frequency and variation, read depth, SNP and zero value count.

2.6.3 Allele frequency examination code

In the following, the main functional components of the code are illustrated. Some adjustments to the code have been performed to increase the speed and reduce memory resources. Since *Julia* version 1.3.0, straightforward application of multi-threaded scripting was enabled.

In the function *Allelfreqcalling*, the variant calling file created by *samtools/bcftools* is processed. Additionally to the variant file, the information described in table 6 has to be added (followingly referred to as *genolist*). The minimum accepted coverage and quality and the maximum quality and the row of the parents in the *genolist* can be adjusted in the function. After the files are read in the *Julia* environment, the variant call file is split according to the samples listed in the *genolist*, and each sample goes in its unique dictionary entry. The dictionary equals a database in its structure and accessibility. The count of the allele observations, which is stored in a string (*1/1:235,18,0:0,6*), is extracted and stored as a numeric entry in two columns – one for each allele. Following this, a normalization index is calculated. It was reported that the alternative allele is less likely to be called compared to the reference allele (Brandt et al., 2015). The sum of alternative allele counts is divided by the sum of all reads to get an overview of this bias. An unbiased ratio should be around 0.5. In the next step, SNPs with low coverage or monomorphic for the parents are removed from the analysis. Furthermore, SNPs that appear as heterozygote for a parent are excluded. When sequencing reads are mapped to the reference genome, sometimes it is not possible to separate between heterogeneous and heterozygote. In another filtering step, SNPs with low coverage and quality are muted for the population samples. When the filtering is finished, the allele frequency of the populations for each SNP individually is calculated. Therefore, the allele of the parents is identified and based on this, the frequency calculation of the alternative or reference base as *Golf* or *ISR42-8* allele is classified. This results in a dictionary, where the information for each SNP is stored. The pooled samples are stored in separate dictionary entries each.

The next function *genloc* connects the gene or marker information to the allele frequency called in the previous function. It can annotate SNPs to a gene or a marker based on the selected input reference file. Besides the reference genome file, the *genolist* file, and the *freqfile* calculated by the previously applied function *Allelfreqcalling*. Furthermore, it has to be specified if a marker or gene file as reference is used and if genes shall be extended or not. The extension of the genes or marker, described in the chapter *Haplotype estimation*, is performed before connecting the SNP to a related gene or marker. The connection is performed based on the position. If a SNP is located in the boundaries of a gene, it is annotated to this. The function adds a dictionary entry to the *freqfile* that contains the information of the gene or marker of each SNP. As all dictionary entries for the SNP allele frequency have the same dimensions, the created *location* annotation information can be stored in a separate dictionary entry. This does save a reasonable amount of memory and retains all information at the same time.

Following this, the core compound of the analysis can be performed. The *haplotyping* function aggregates the allele frequency of a single SNP which are denoted to a gene or marker together to one value for the whole region. As the information on GO terms and PFAM IDs are also given, the aggregation of the allele frequency can also be performed for these two levels. SNP not annotated to a gene or marker are removed from the set. The genomic area that is not covered by a gene or marker is associated with the gap left between two genes or markers. In the haplotyping, the allele frequency of the *ISR42-8* and *Golf* allele is calculated by multiplying the allele frequency times the coverage, summing up these values for all SNP in the window and dividing this by the sum of the coverage. A SNP with higher coverage and expected higher accuracy would be weighted more heavily than a SNP with lower coverage by this method. This should prevent a bias towards low coverage SNP. Additionally, the haplotype coverage as the sum of the coverage of all SNP and the minimum and maximum allele frequencies of a single SNP is calculated. This step is time-critical, as it requires much time to search for all entries related to a gene or marker in a list of millions of SNP. Therefore, a small helper function was included to decrease the computation time by magnitudes. The dictionary entries of the location and a given pool sample are merged and sorted according to the gene or marker name. The first couple of rows are checked until an entry with a variant haplotype is found. This indicates the end of the list for the haplotype and makes further check-up in the data frame unnecessary. The head with the SNP for the gene or marker is transferred into another file and processed further to the haplotype identifier values mentioned above. This part was not only copied to another variable but also chopped off the table. This way, the following iteration again only needed to check the first couple of rows. With this simple implementation, the processing time of the *haplotype* function could be reduced from days to minutes.

Besides the *haplotype* function, a simpler contig function is included. The contig function calculates the allele frequency and coverage for a given window, as has been proposed by Tile et al. (2019). The so far presented functions build the core set of functions each of the following functions depends.

With the function *equaltest*, the replicates are compared to each other. The comparison of the replicates is performed on the haplotype or the contig level and is based on the output generated by the function *haplotyping*. This level of processed data is chosen instead of the single SNP because the standard deviation of the processed layer is less explicit and less prone to errors. Furthermore, two additional layers in the function compare the generations against each other, and a third level compares the generations between the systems against each other. The function can take up to five replicates per sample and does not demand an equal number of replications for the samples. The statistics are based on the *R* package *pscl* (Jackman et al., 2013), which allows applying a generalized linear model combined with a zero-inflated model. The data for two sets to test is transformed and pushed to the *R* environment by the package *RCall* to use the *pscl* functions. After the calculation is performed, the output is transferred back to the *Julia* environment.

This function can answer the question on genome-wide variation processes, while the more sophisticated task to test individual genes or markers generation-wise has not been addressed so far. The application of the function *evol* can close this gap. It is written to calculate the gene or marker-wise variations between the initial generation F_3 and each following generation. The function utilizes the information produced in the *Allelfreqcalling* and *genloc* functions. It collects the SNP for a given gene or marker haplotype and works on the level of GO terms and PFAM-IDs, too. Average allele frequency, standard deviation, coverage, marker count, and count of zero calls are reported. Additionally to this, a highly adaptive algorithm performs the matching of correct statistical tool and data. The replicates were merged in one dataset to increase the coverage and value of the data. Each haplotype has its unique characteristics and is settled in a dedicated spot. Four possible statistics are created based on the variation of observed distributions to address all possible kinds of allele frequency distributions for a haplotype. The first statistical test is a simple non-parametric *Mann-Whitney-U Test*. It is applied to all haplotypes with less than 30 SNP for the first and test generation together. This test is also applied if one of both sets only has zero values in its allele frequency. For this kind of haplotypes, a precise estimation of the actual distribution cannot be made. In case all SNPs indicated a wild allele frequency of 0, an application of the following methods is prone to produce unexpected errors. If none of these two statements hold true for the two sets, three options remain. If less than 15% of all SNP of the haplotypes are 0% wild allele frequency or all zero values origin from a single test generation, a generalized linear model based on a binomial distribution is applied. If more than 15%, but less than 75% SNP of the haplotypes characterized by a zero-allele frequency call for the wild allele, a zero-inflated generalized linear model, based on the binomial distribution, is applied. If none of the previously described distributions fit the haplotypes, a zero-inflated, Poisson-based linear model is applied to distinguish variation between the initial generation and the following generation. If this failed due to missing or incomplete data, a *catch* loop was applied to keep the function alive. In this catch, the haplotype calculation will be noted as not possible. This last branch is implemented as the last defense to keep a long-running function alive and is a rarely expected outcome. The four described statistical tests should accurately cover 99.99% of all possible distributions in the data sets. As this function is based on processing natural SNP allele frequency data, like the *haplotyping* function, the processing time can become extraordinary long. Especially with millions of SNP, the selection of all SNP located in a gene becomes a bottleneck. This problem was solved in the same manner as described for the *haplotyping* function. A sorted list of SNP is checked from top to bottom, and processed SNPs are deleted from the table. By applying this workaround, the processing time of the function could be reduced from several weeks to few hours when applied on data with millions of SNPs. This function creates the backbone information of any further selection or evolution identification on the gene level.

Besides the comparison of the first to all following generations, a comparison between other generations and the environments was desired. The functions *scomp* and *gcomp* were written to fit this demand. For the environment comparison by *scomp*, the dictionary name, the *genolist*, and the organic and conventional generation to test have to be specified. The level of haplotyping and whether a transformation step should be performed or not also needs to be addressed. If transformation is selected, the allele frequency is multiplied with the read depth of the SNP and log-transformed. The default setting runs the code without a transformation. The *scomp* function created comparable output to the described *evol* function and uses the same decision tree for the statistical test. Analogously, the function *gcomp* was created to compare generations within an environment to each other.

A further implemented function (*investRegio2*) was intensively used for the extraction of candidate genes in a QTL region, defined by prior findings published. It can be used to identify candidate genes in a QTL region that are positive selected in one of the environments, positively selected in both environments compared to the initial generation, or negatively selected for both. For the correct

application of this function, the regionwide allele frequency has to be estimated priory. Several settings have to be specified. Besides the chromosome, the region on the chromosome have to be specified. It can be specified either as a genetic or physical position. The physical position is multiplied by 10^6. Therefore, the position is only entered as a digit of 0.1 to 800. The genetic map position is not adjusted to be entered as a number from 0 to 200. The genetic position has to be in line with the *9KiSelect* chip. Some investigated publications used genotyping chips based on the *BOPA1* markers or created markers on their own. The sequences of the *BOPA1* markers were aligned to the reference genome to transfer the information of the *BOPA1* makers to the *9Kiselect* map to get a physical position. The physical positions of the *9KiSelect* map and the *BOPA1* markers were used to match the two maps and align them to each other. This way, a genetic position on the *BOPA1* map could be translated to a genetic position in the population. Individually created markers obtained from other publications were aligned to the reference genome and extracted regions based on the physical map position. If only a single marker characterizes a region, an educated guess is performed, and a region surrounding the marker is defined. Additionally to the position, a range of allele frequencies is qualified. A numeric value defines the direction of the allele frequency for both environments. Exemplary, a value of 1.1 indicates a gene should have a wild-type allele frequency above 10%. The other direction was described by 0.1, where genes are selected with a wild-type allele frequency of less than 10%. By combining one value for each environment, the candidate genes can be pre-selected (E.g., the potential candidate has an allele frequency of less than 10% in the conventional and higher than 20% in the organic system *[rangek=0.1, rangeo=1.2]*). Some genes might be selected in the same direction. Because of this, an option *equal=true* can be specified. If the allele frequency of a region evolved in the same direction, this setting has to be set as *true*. It is used in filtering for the *p-value*. If *equal=true* is specified, the algorithm searches for entries with a *p-value* greater than the selected one. On the reverse side, if both environments have evolved in opposing directions, the *p-value* is set as *cut-off level*, where entries with a higher *p-value* are eliminated. For genetic regions, the physical position was extracted, and all genes in the region are selected as candidate genes. The identification of candidate genes is performed to systematically compare the 23rd generation of the organic and conventional environment, calculated by *scomp*. After the above-mentioned filtering steps are completed, an additional file with the protein code information is included in the final table. The protein information was prepared according to chapter 2.9. Based on the allele frequency, the QTL candidate region, the predicted gene function, and the protein variation between the parents, a most likely candidate gene was selected for the trait of interest.

The package contains further small functions that allow to save/reload the database, merge replicates, and create plots on genome broad or gene levels.

2.7 Validation of replicates

To test if biological pool replicates of the same population result in the same allele frequency pattern, we applied a negative binomial and zero-inflated generalized linear regression model from the R package *pscl* (Jackman et al., 2013). Together with the expected decline of wild donor AF for some genes, the double back-crossing made it necessary to include a zero-inflated model. The smaller the wild donor AF is, the more likely there is an underestimation of the actual AF, leading to false-positive estimates. The negative binomial distribution was chosen concerning the fit to the AF distribution. As haplotyping with a none constant window size has not been reported yet, we tried to estimate the effect of the extension size and the position on the allele frequency estimate. Therefore, we calculated the deviation of real to pool estimated allele frequency and estimated the effect of extension in a linear model. Chromosomal position and coverage were used as factors on the AF correctness.

2.8 Genetic map construction from pooled genotyping data

We aimed to illustrate the genome-wide genetic variation within generations and systems and compared these against each other. Genetic maps can give higher resolutions in telomeric regions than physical maps (Russell et al., 2016). As no classic haplotype allele frequency can be obtained from pool data, we estimated *D'* and r^2 by multiplying the haplotype allele frequency (HAF) deviation of two neighboring HC gene haplotypes with 1/physical distance multiplied with 10,000. r^2 was calculated by dividing the negative logarithm of the HAF deviation by the haplotype physical distance squared to two. Both formulas are illustrated below. The procedure was performed for each generation and system separately, merging the information of both replicates.

$$r^2 = \left[\frac{-\log(\Delta\, HAF_{H1H2})}{\log(\Delta\, Dist_{H1H2})}\right]^2$$

$$D' = \Delta\, HAF_{H1H2} * \frac{1}{\Delta\, Dist_{H1H2}} * 10{,}000$$

Equation 11: genetic map construction equations to calculate the recombination r^2 and the linkage D'

2.9 Prediction of Protein variation

The prediction of the protein variation between the parents is a critical element in detecting candidate genes. Such a variation caused by a deletion, insertion, or base change can result in a loss of function (Comadran et al., 2012a). The experimental design can lead to a wide range of possible adaptations. Contrasting to studies like Comadran et al. (2012a), the experiment was not started to investigate a specific protein for its functional variation (Comadran et al., 2012a). Therefore, a database for all known barley proteins with a start and stop codon was created. It should highlight the protein variation between the two parents *Golf* and *ISR42-8* and could be implemented in a candidate gene analysis. The protein variation database was created based on a reference *CDS* file (coding DNA sequence), downloaded from the ensemble plants database (Version 2 CDS fasta sequence file). 236,194 splice variants of 37.705 genes were reported in the list. From this set of genes, 26.173 genes have a splice variant with a known start and stop codon. This file was processed in a way that the multi-line sequence was transformed to a single line sequence. The name of the splice variant, the locus, direction of the gene as well as the length, gene name and the sequence were stored in a file. The locus defines the region of the smallest start to largest end position for all splice variants available. This file was used as annotation to estimate the protein for the two parental genotypes *Golf* and I*SR42-8.* Chloroplast and mitochondrial information are omitted.

A multi-threaded *julia* function runs a cascade of external *bash* functions to align the *Golf* and *ISR42-8* alignments to the reference protein. The first step in this cascade was the call of all bases on the locus of the gene. Regarding the gene orientation, forward or reverse sequences were called for *Morex* (reference genotype), *Golf* and *ISR42-8.* A consensus sequence can be created from the aligned whole genome resequencing data. The information of the chromosomal location of the gene was utilized to extract a consensus sequence for the two parents Golf and ISR42-8 from reference aligned *bam* files using *samtools* and *bcftools* (Li et al., 2009). If necessary, the reverse complement strand was assessed by the *FASTX* toolkit (Gordon & Hannon, n.d.). The obtained consensus sequences of the two population founders were aligned to the designated CDS reference splice variant. The alignment to the coding region was performed by *Muscle* (Edgar, 2004) and *Mafft* (Katoh & Standley, 2013) alignment. Two aligners perform the alignment to increase the alignment quality. *Mafft* alignment was performed with a maximal iteration rate of 10, using the default option for the remaining options. *Muscle* alignment was performed on default options with a memory limitation of 2,048 MB with a maximum of five iterations. *Mafft* has shown to be much more memory efficient than *Muscle*. After the alignment was performed,

the *fasta* alignment file with the information of *Morex, Golf,* and *ISR42-8* was transferred to the *Julia* environment and converted to a DNA sequence by the package *Bio.Seq*. Deletions or excessive bases before the start and after the end of genes were displayed by a '-'. This character was removed from the sequence and counted when not related to an intron (before the start or past the stop codon). The match of the start as well as the stop codon is inspected and the count of stop codons calculated. Besides these characteristics, the count of base mismatches, deletions in *Golf* and *ISR42-8* sequence and the protein size were reported. Lastly, the amino acid sequence of *Golf, ISR42-8* and the reference *CDS* were reported. All those sequences build the information of allelic variations between the parents and gave a prediction for base variations with an effect on the protein function or structure.
For some genes, a well-matched alignment was not achieved with the proposed procedure. For this reason, genes with high levels of mismatching bases are reprocessed for all available splice variants. The best alignment with the lowest number of mismatches is filtered out and is further used for the protein function variation.

2.10 Genome-wide structure and variation

The potential influence of genetic drift was simulated by a Julia script, related to the R function *sim.drift* (Anonymous II, n.d.). The following settings were applied: population size, 10,000 (center part of the plot); ploidy level, 2; the number of generations, 21; as initial allele frequency, 0.1; the number of loci, 34,237; and iterations, 1,000. Additionally, the wtAF of the third generation was used as a start point to calculate the variation based on drift effects. Over the 1,000 iterations, a median value and the 95% confidence interval were calculated and compared against the gene-based haplotype allele frequencies of the organic and conventional population. Corresponding generations were tested against one another. A negative binomial linear model was applied to calculate the probability (*p*) value.

PCA was performed with the R Bioconductor package PCAtools (Blighe, 2019) to identify genetic structures and variations between generations and environments. The explanatory effect of each principal component on the generations, replicates, and environments were calculated by the *stats* (R Core Team, 2020) *cor.test* function, based on a *Pearson* correlation test. Samples were categorized by the number of replicates, environments, and generations. The diversity estimation within each environment and generation was tested based on the variation between the two samples. The distance of the samples was calculated for each principal component separately and multiplied with the explanatory weight of the PC. Based on these values, a single value over the first five components was summarized and compared between the environments. A dendrogram of genetic distances based on the HC gene HAF was created with the *pvclust* (Suzuki et al., 2019) package with the *ward.D2* agglomeration and *canberra* distance method. Gene-based haplotypes with a read depth above 100 were used for this calculation. The F_{ST} value was calculated using *vcftool*s (Danecek et al., 2011) with default settings and *vcf* and BAM files as input.

To estimate the potential influence of random genetic drift, we estimated the drift effect in a population by R function *sim.drift* (Anonymous II, n.d.). The population size was set to 6,000 according to our estimate of genotypes per generation, with a ploidy of two and 25 generations to test on an initial allele frequency of 0.125 over 35,000 loci, 20 times iterated. For each iteration and generation, the variation generation was tested by HSD.test of *agricolae* (de Mendiburu, 2020) package. PCA was performed with the R Bioconductor package *PCAtools* (Blighe, 2019). Samples were categorized into replicates, environments, and generations. The dendrogram, based on the HC gene HAF was created by stats (R Core Team, 2020) *hclust* function with *ward.D2* agglomeration after scaling of HAF values by *Euclidean* method. All gene haplotypes, regardless of the read coverage, were used. The F_{ST} value was calculated using *vcftools* (Danecek et al., 2011), which used raw *vcf* and *BAM* files as input.

The overall introgression level on all loci was assessed at the gene haplotype level. If a haplotype was characterized by an ISR42-8 allele frequency of bigger than 0.005 in any of the observed generations, it was assumed to be introgressed.

2.11 Selection signature – gene ontology and candidate genes

We aimed to distinguish candidate genes as a causal reason for selection signals. QTL studies observed phenotype to genotype correlations for various traits like drought, plant physiology, resistance to biotic stressors, or yield. As many of these studies used the 9KiSelect genotyping chip, a direct link to the genetic map created for the population marker-haplotype allele frequency could be established. For other publications using designed markers or other genotyping chips, the exact position was identified by mapping the marker to the reference genome or aligning genetic maps to each other by using the physical position. Putative candidate genes were selected based on the HC gene HAF variation between the organic and conventional systems in the 23rd generation. The number of potential candidate genes was reduced by giving the genetic or physical position of the QTL region, a potential range of HAF below or above a given percentage, as well as the information if both systems were equally selected for the trait. The output was selected for the most promising candidate(s) based on the raw haplotype allele frequency, the variation between the systems, the amino acid variation of CDS, and the functional prediction (Anonymous, n.d.).

Gene ontologies were classified for the two predominate weather impacts with the potential to promote selection pressure. Fungal pressure was investigated by comparing the organic and the conventional environment for the F12 and F22 generations. These two generations were selected as they follow two years with expected high fungal pressure. Significant genes for comparing the environments were extracted ($p < 0.01$) and the intersect between the two generations was selected. Genes were annotated to GO terms, and a Mann-Whitney-U Test was performed on the allele frequency. The list of GO terms was ranked according to the p-value of organic to initial generation for GO terms with coverage of at least six genes.

The twelfth generation was selected as a baseline (2007) to dissect the impact of drought on the selection processes in the populations. This generation follows several years of sufficed water supply during the vegetation period. Contrasting to this generation, 2011 (F16), 2017 (F22), and 2018 (F23) are described by water deficiency periods. These three generations were compared to the F12. The significant genes intersection of these six comparisons were identified on a significance level of $p < 0.01$ and GO terms annotated. GO terms with a significant variation in at least two of three generations were selected as candidates for the organic and conventional system's drought stress response separately.

2.12 Yield characteristics

For each year in the period of 1999 to 2020, several yields defining parameters were collected. Besides the yield, the plants per meter and the tiller per plant were measured. Furthermore, the thousand kernel weight at a dry weight of 86%, emergence, flowering, and ripening date of the population are available. If lodging occurred, this was reported, too. For the last generations, the protein content of the corn was measured.

As each population has only one replication per generation, the analysis of this data was somewhat challenging and prone to errors. Therefore, relative yield data was calculated to the average spring barley yield of Germany as a whole and North Rhine-Westphalia separately. Additionally, spring barley yield trials were performed in the years 2011 to 2016 at Campus Klein-Altendorf. The relative yield was

calculated by dividing the deci tons per hectare yield of the plot by the yield obtained by the mean yield of Germany, North Rhine-Westphalia, and the field trial plots mentioned.

The other yield defining characteristics were omitted, as data was not available for each year.

Chapter 3 Results

3.1 Pool sequencing – comparison of different methods

The sequencing of the populations is one major challenge to identify variant developments over the generations of selection. A key might be the application of pool sequencing. Population genetics without pool sequencing can be considered a rather costly approach. Therefore, three next-generation sequencing methods have been tested to figure out the best fit. The aim is to discover many SNP at moderate to high coverage with a good representation of the genome. Ideally, the variants found are related to a gene so that a direct link between the function and the allele frequency variation can be drawn.

3.1.1 Genotyping by sequencing

The sequencing of the two parents Golf and ISR42-8 and three pool samples of the 22nd generation of conventional farming was performed by *LGC genetics.* Twenty million raw paired-end reads with a length of 145bp could be found on average for each sample. This data were trimmed before the alignment by applying a sliding window approach with the setting 30:15. Additionally, the first 15 bases were removed, and the following bases needed to exceed the quality of 30 phred score. The minimum length of the reads must not decline below 70, and the end of the reads needed to have at least a sequencing quality phred score of 15. The sequencing quality declines with the length of the read, as prior potential sequencing errors add up to the quality score of each following base. The sequencing quality before and after trimming is shown in figure 7.

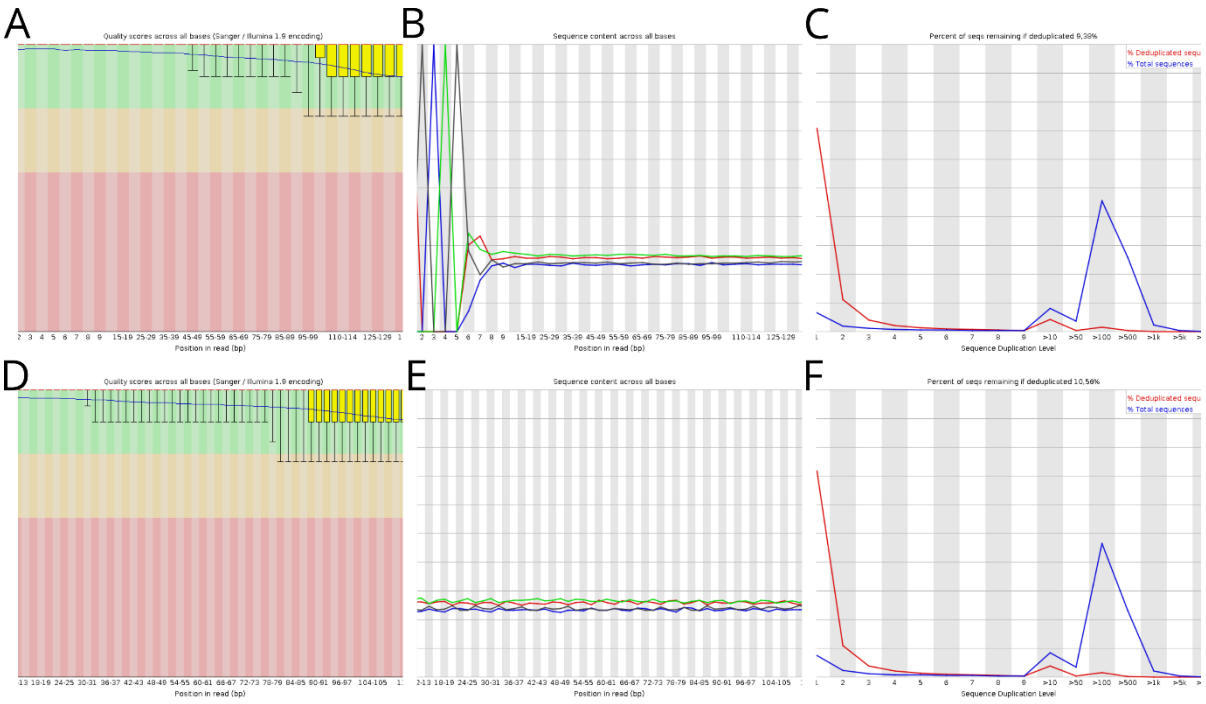


Figure 7: Quality score overview of GBS sequencing results before and after trimming the sequences; ***A, B,*** *and* ***C*** *before trimming;* ***D, E,*** *and* ***F*** *after trimming. The* ***A*** *and* ***D*** *illustrate the median quality of each base position over a sample of tested sequences with a boxplot. The background color gives an estimate of which phred level should be accepted or where low quality can be expected. Furthermore, the duplicate level before and after trimming is given in the figure* ***C*** *and* ***F****, giving the assumption that a high share of duplicates is present in the data set.* ***B*** *and* ***E*** *illustrate the percentage base count for each position of the reads. While the first nine bases do have an inhomogeneous pattern before trimming, the read appears much more consistent after the removal of the first bases. Generally, the reads are of high quality, while they contain adapters at the start of the reads and the duplication level is very high.*

The alignment filtering was done according to the material and method chapter. The enzyme-based digestion created a problem that was not expected. Due to the exact digestion, all reads of a locus started

at the same base and mostly ended at the same base position. This made PCR duplicate removal challenging. When the same settings were applied as for *MACE RNAseq* or *WGrS*, more than 90% of all reads were removed. Due to this, the PCR duplicate removal was not performed. The allele frequency of the progenies was called with default settings. A SNP was considered if it had a minimum quality of 30 over all tested samples and at least one read coverage. The SNP calling revealed 82,435 polymorph SNPs between Golf and ISR42-8. The sequencing coverage is 28 reads on average, with a median of six reads and a variance of 3,329 reads per polymorphism. Most polymorph loci were characterized by low coverage, while a distinct group exists with more than 220 reads per locus. The three biological replicates did not differ for the distribution of sequencing depth, as shown in figure 8.

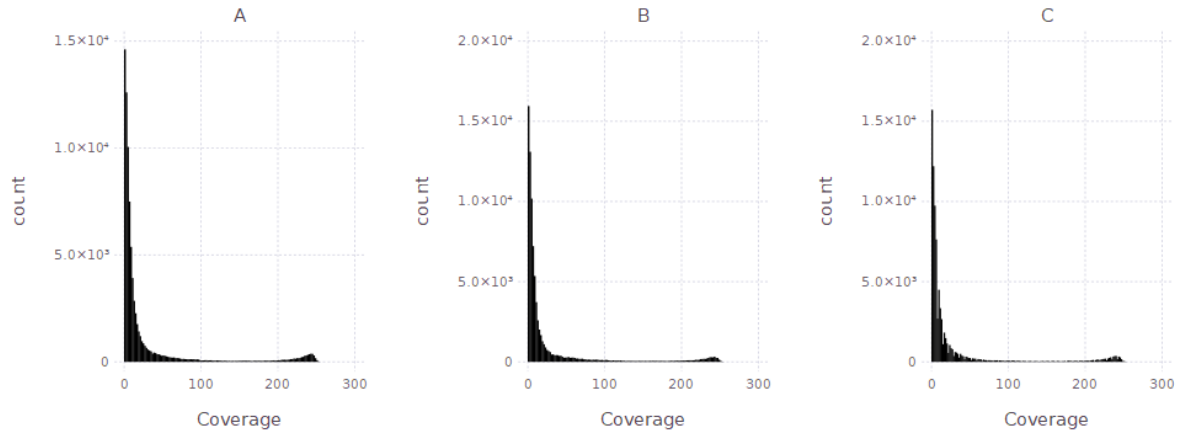


Figure 8: Read depth of each polymorphism locus for the three replicates of the 22nd conventional generation. The histogram illustrates the presence of two distributions – one close to zero, the other bigger than 200. Over the three biological replicates, the distribution is similar.

The measured allele frequency of ISR42-8 in the population over all replicates and loci was 7.48%. Assuming an allele frequency of 12.5% in the initial generation F_3, this was equivalent to a frequency reduction of 4.5 to 5.5%. The three replicates were correlated by 0.55, and the alternative to reference base call ratio is equivalently 0.55. The result of this ratio indicates that more alternative base calls than reference base calls were observed. The statistical test, performing a zero-inflated negative binomial linear regression model on the allele frequency, points out that the replicates differ significantly on a p-value < 0.001 for both distributions.

The coverage of the loci over the genome was uniform. No region with a deviating pattern was identified. Generally, the pericentromeric region of each chromosome showed fewer occurrences of above-average sequencing depth. More details can be obtained from figure 9.

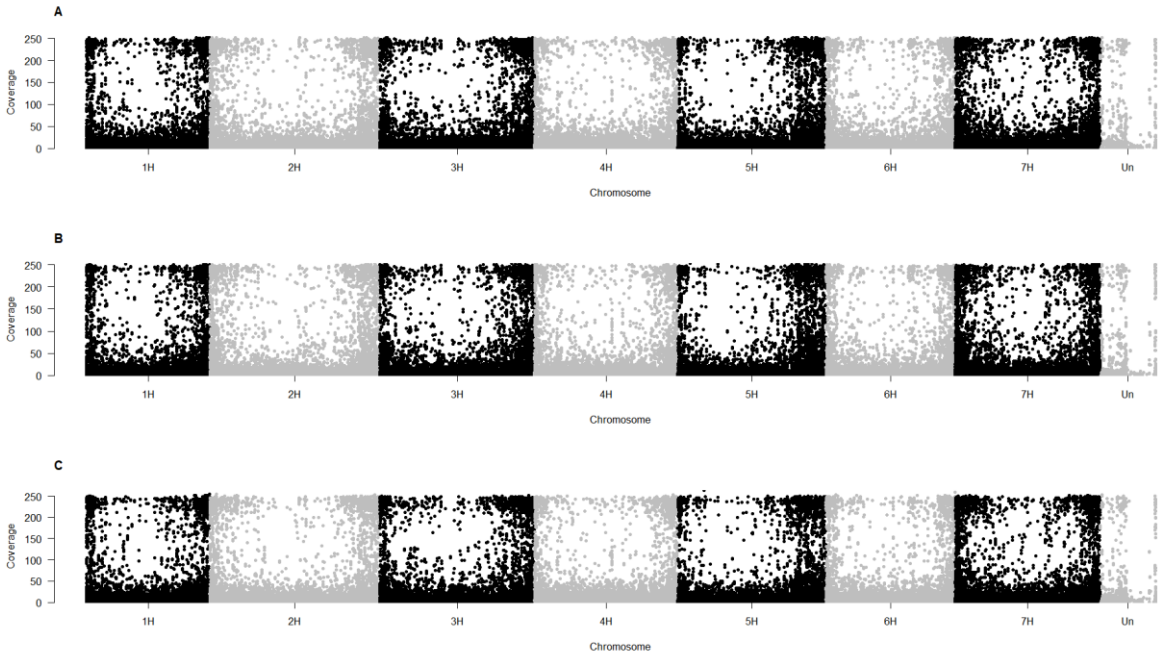


Figure 9: The genome-wide coverage distribution for the three biological replicates (A, B, C). The majority of loci have a coverage of smaller 50, while a small group of loci has a coverage of approximately 250 reads per locus. This pattern is uniform over the whole genome, and no positional bias can be identified.

When calculating the gene annotation-based haplotypes (GH), 21,369 haplotypes were compiled. About 4,800 of these were constructed from more than four SNP. Approximately 6,500 GH consisted of only one SNP. About 4,200, 2,750, and 1,820 GH were constructed that consist of two, three, and four SNPs, respectively. By constructing GH, the average read coverage per locus was increased to 110, while the median was 28 reads. The allele frequency of ISR42-8 remained close to the estimate from the SNP data, with a low deviation (7.54%). 13,595 SNP were disregarded from the haplotype aggregation, as these were too far away from any gene they could be annotated to. The GHs were typified by an average polymorphism count of 3.4 and a median of 2 SNP. The variation of the SNP count was 10.45. figure 10 showed the distribution of the read depth per haplotype (A-C) and the corresponding SNP count (D-F) for the three replicates. The observations indicated a *Pareto distribution* for the read and SNP count.

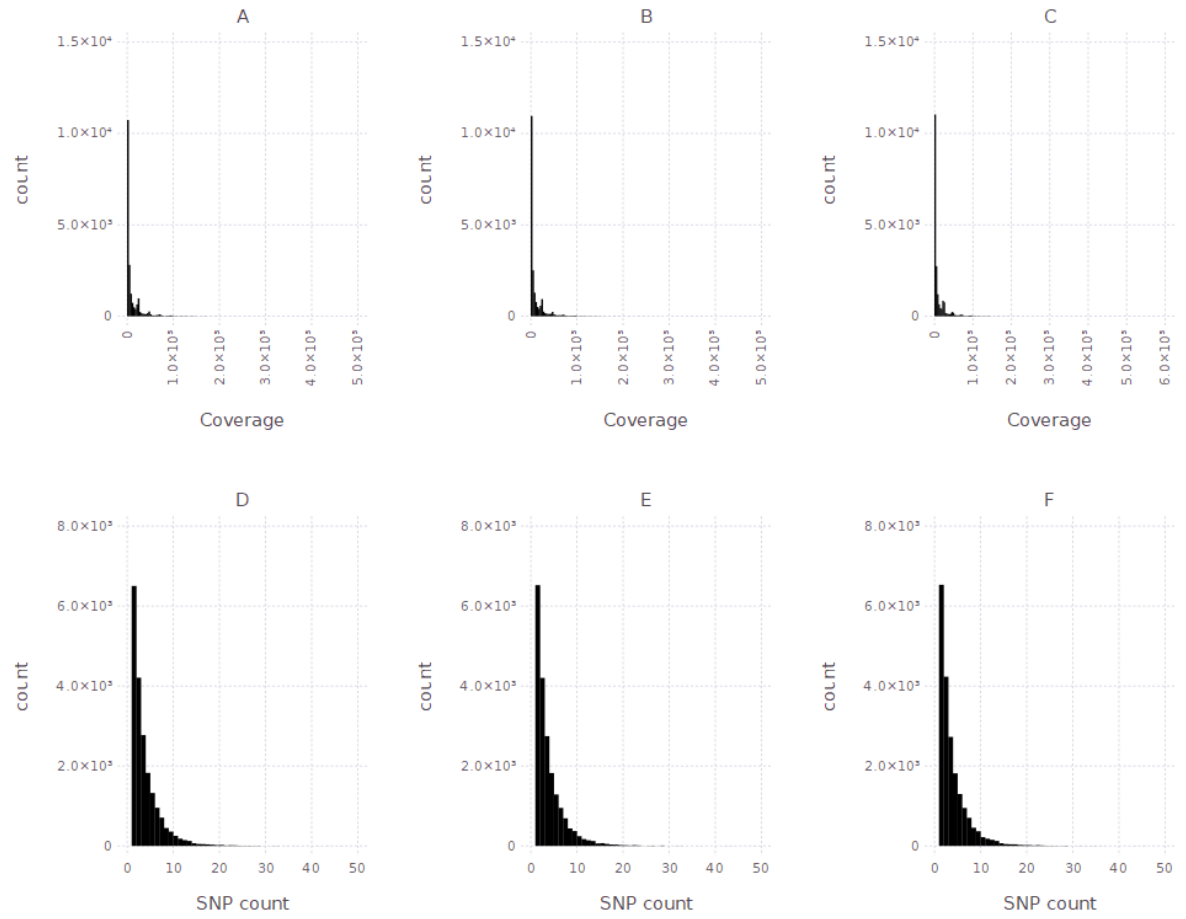


Figure 10: Distribution of the read depth per sample (A-C) and the corresponding SNP count for the gene haplotypes (D-F). The replicates are printed in rows. The distributions are related to Pareto distributions. For most of the haplotypes, the information of one to four SNP is aggregated together.

The correlation of the replicates increased to a value of 0.71, compared to a previous level of 0.55 on the SNP level. In the negative binomial zero-inflated linear regression, the p-values increased in direct comparison to the SNP level. Nevertheless, the samples were still significantly different ($p < 0.0001$). The difference between the replicates was reduced even further if the contig approach was applied to the data. 10,000,000-bases was the defined window size in which the SNP information was aggregated together (Contig haplotype – CH). Four hundred eighty-three genome-wide contigs with an average SNP count of 154 per contig were constructed. The median was close to the mean (142), and the variance is 5,081 SNP. The coverage of reads per CH was increased to 4,837 reads on average, with a variance of 2.9×10^9. The median coverage was 2645 reads. This was an apparent increase to the other haplotyping approaches. Along with the increase of coverage came a highly insignificant p-value of 0.91 for the negative binomial model and 0.48 for the zero-inflated model for all three replicates. Analogous to this, the correlation of the replicates increased to levels above 0.75. Thus, the three replicates were equal on the CH level. Figure 11 illustrates the deviating pattern of the contig haplotype construction when compared to figure 10.

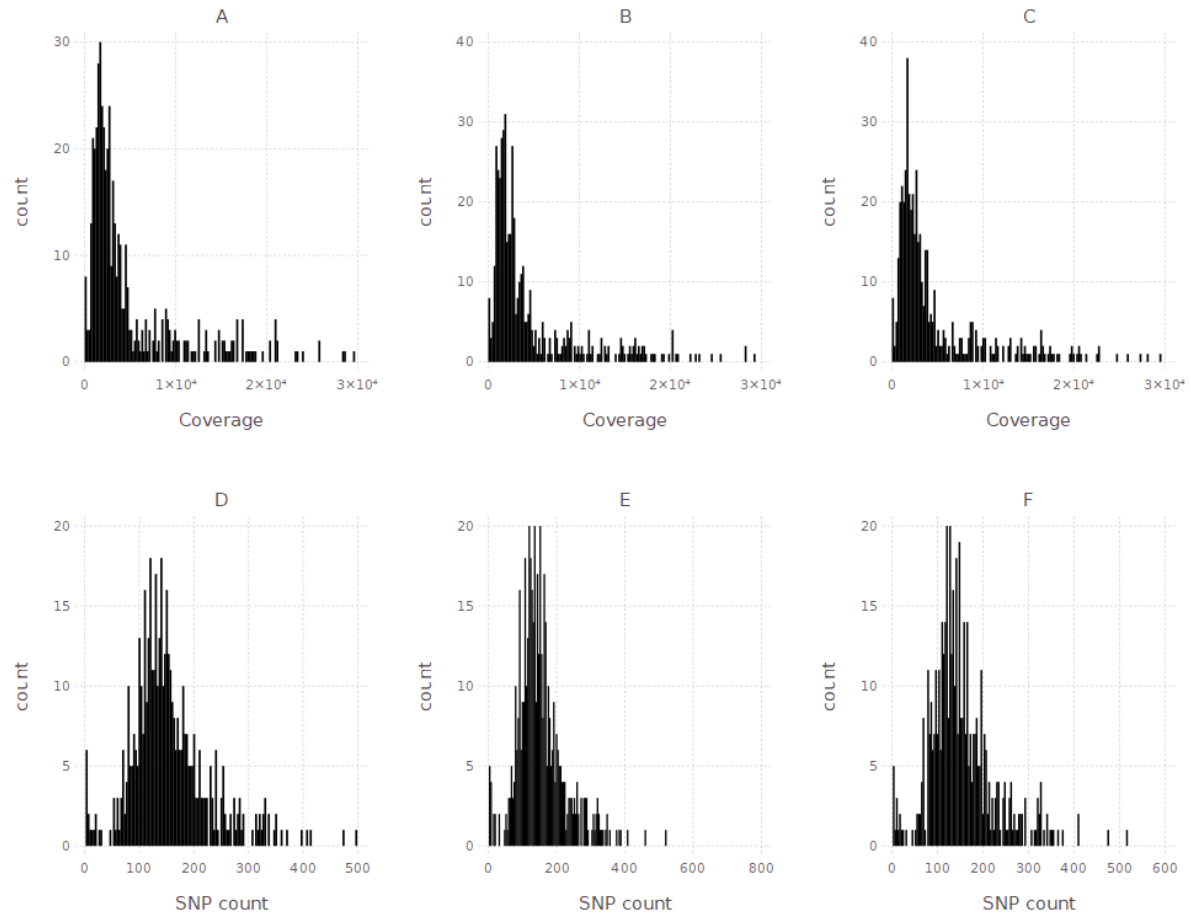


Figure 11: Distribution of the read depth per replicate (A-C) and the corresponding SNP count for the contig haplotypes (D-F). The replicates are printed in rows. The distribution pattern is totally different from the pattern observed for the marker and gene haplotypes. There is almost no zero inflation present in the data anymore.

Similar haplotyping was applied based on markers of the *9KiSelect* genotyping chip (marker interfered haplotypes – MH). Four thousand seven hundred two marker-related haplotypes were constructed, which on average had a SNP count of 15 with a coverage of 450 reads per MH. The median value was 8 SNP and 234 reads coverage, while the variance was 890 SNPs and 430,575 reads coverage. The allele frequency of 7.43% was calculated based on 65,600 SNP, which constructed the haplotypes. The correlation of the replicates was 0.75 and comparably high, as mentioned for the gene haplotypes. For the replicate comparison, MH showed a non-significant p-value of $p=0.17$ (negative binomial distribution), while it was highly significant for the zero-inflated distribution ($p<0.001$). Figure 12 illustrates the construction distribution of the marker haplotypes. It is comparable to the previously described gene haplotype pattern.

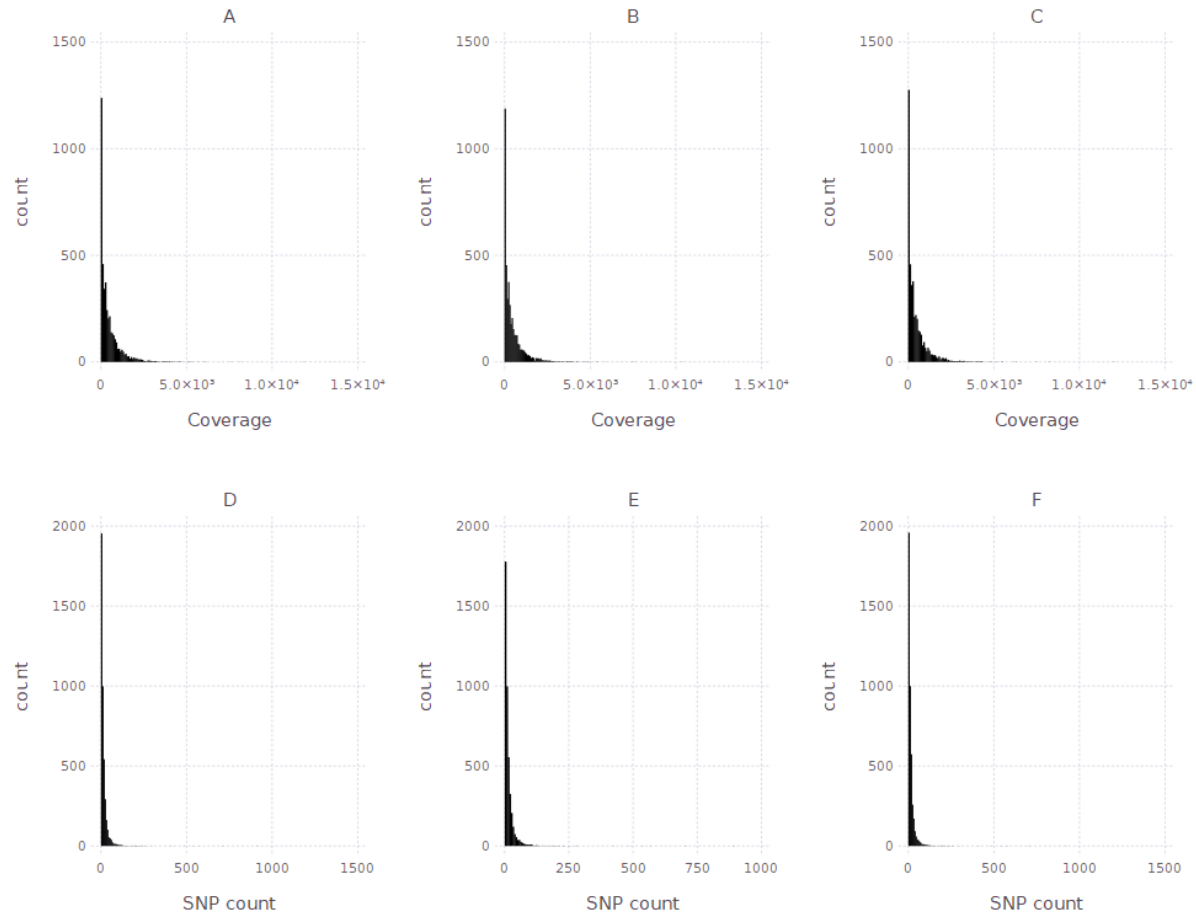


Figure 12: Distribution of the read depth per replicate (A-C) and the corresponding SNP count for the marker haplotypes (D-F). The replicates are printed in rows. The histogram highlights the fact that most of the haplotypes consist of few loci with shallow coverage.

The comparison of the pooled sample to individual genotyping resulted of the pool members was necessary to validate the accuracy of the pooling method using GBS sequencing. *21 KASP* markers were designed, from which one failed completely and another was classified by an error rate above 50%. These two markers were discarded from any further analysis. The remaining 19 markers had a failure rate of < 0.015 and were compared to the belonging pool of genotypes by *Person correlation*, and root means square error and a negative binomial zero-inflated linear model.

The intersection of *GBS* and *KASP* marker was only a single variant position for the single SNP level. The sequencing resulted only in one direct match. For the GH approach, nine matches were observed. No correlation was found in the direct comparison of pooled to individual genotyping (-0.06). Nevertheless, the other two replicates showed a correlation above 0.75. It had to be mentioned that these samples were not identical to the individually tested genotypes. With a median coverage of 56 reads, the coverage was marginally higher compared to the single SNP level (41). The *Pearson* residual width was 3.4%. Although the correlation was low, an evident variation between the three replicates could not be distinguished (neg. bino. dist. p = 0.174; zero infl. dist. p = 0.179). The root means square error is 0.109.

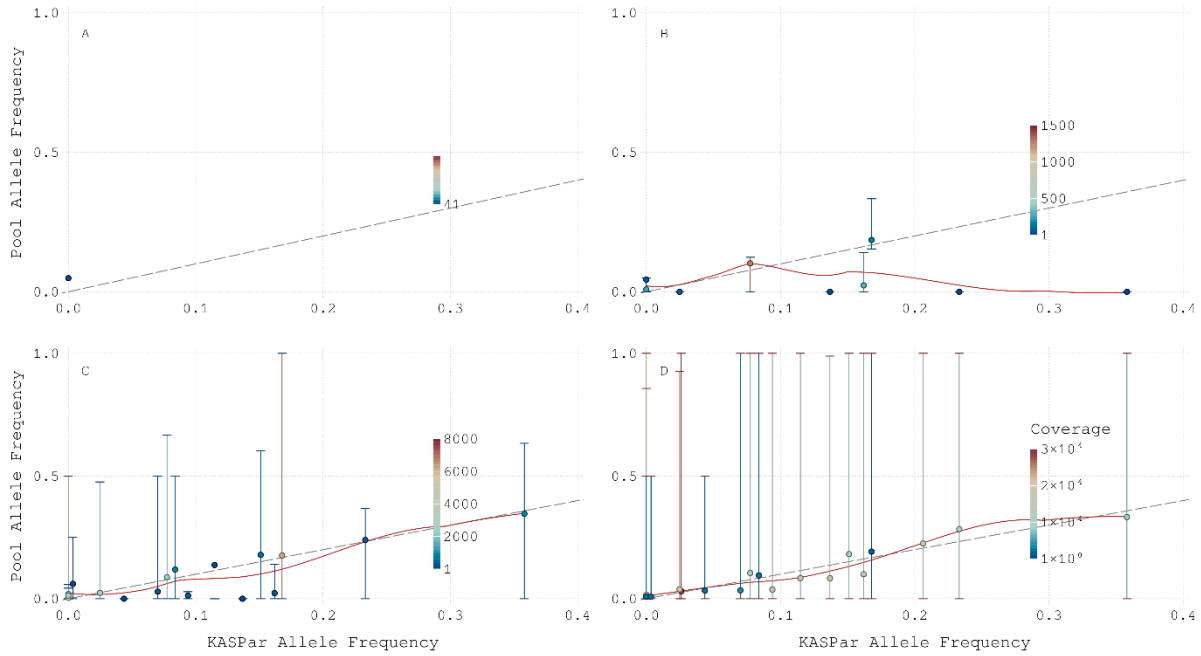


*Figure 13: KASPar individual genotyping allele frequency results against the measured allele frequency in the corresponding pool sample for the **GBS** data. The dashed line indicates the optimal match, where pool obtained values match the individual genotyping ideally. The red curve is a smooth regression curve using all points. The color of the points indicates the coverage per locus, which ranges from 1 to several thousand. Additionally to the points, error bars are drawn. This highlights the margin of single SNP allele frequency that contributed to the overall allele frequency of the haplotype. If no error bar is visible, this means there is only one SNP contribution information to the haplotype. **A** – the single SNP comparison. Only one SNP was detected having the exact same position as the KASP markers. **B** – the gene-based haplotype allele frequency compared to the individual genotyping. **C** – marker-based allele frequency. **D** – contig haplotype-based comparison to the individual genotyping. The pool sequenced sample contains the exact same 288 genotypes that have been tested individually for the 21 KASP loci. As two KASP markers did not meet the quality threshold, they were omitted.*

For the MH, all 19 markers could be matched (figure 13C). Compared to the GH, the median coverage per haplotype was ten times higher. Analogous to this, the correlation of the pool to individual data was 0.834. The statistical test highlights that there was no variation between the two samples (p_{negbin} = 0.466/$p_{zeroinfl}$ = 0.62). The *Pearson* residual width was 3.77%, and the root means square error is 0.057. The allele frequency of the marker haplotypes was close to the real allele frequency over the whole range of frequencies. Especially when the coverage was high, the pool frequency matches the individual frequency. This is illustrated by the red smoothing regression curve in figure 13C. The error bar illustrates the whole variety of single SNP allele frequencies. For one marker, it ranges from 1.0 to 0.0 and covers the entire palette of the allele frequencies possible. However, the calculated mean ends up to be exactly on the dashed line, indicating a correct estimation of the allele frequency for the pooled sample. This particular point has high coverage of >6,000 reads. On the other end of the spectrum, some haplotypes only had a shallow coverage, which seems to be associated with an increased likelihood for false allele frequency estimation. Analogous to 28B, the haplotypes consisting of only one SNP showed high variability and poor reliability.

The contig haplotyping of the pooled sample resulted in the best correlation to the individual genotyping of 0.944. Homologous to this, there could not be any significant variation reported between the two sets (p_{negbin} = 0.59/$p_{zeroinfl}$ = 0.446). The root means square error was 0.032, and the *Person* residual width 3.6%. These values were based on the match of all 19 KASP loci on a median coverage of 10,122 reads per haplotype. As it is illustrated in figure 13D, the haplotypes were constructed from SNP that range from 0 to 1 for the same haplotype. The combination of the SNP information to the haplotype lead to a

nearly correct estimation of the real allele frequency by the pool. No coverage-related correlation can be conducted from this data.

The contig haplotypes showed the best performance for the pool genotyping by using genotyping by sequencing. The replicates did not differ, and the correlation to the individual genotyping was the highest. The other three levels of information aggregation suffer from low and biased sequencing.

3.1.2 MACE RNA sequencing

Besides the two parents Golf and ISR42-8, the 22nd conventionally farmed generation was sequenced for three biological pool replicates, which consist of 300, 300, and 288 genotypes each. The sequencing was performed by *GenXPro*, generating on average 8 million reads over the three replicates with a sequencing length ranging from 17 to 68 bp. The relatively short reads needed to be trimmed before alignment, removing the ten bases on the head of each read and omitting all reads shorter than 40 bp.

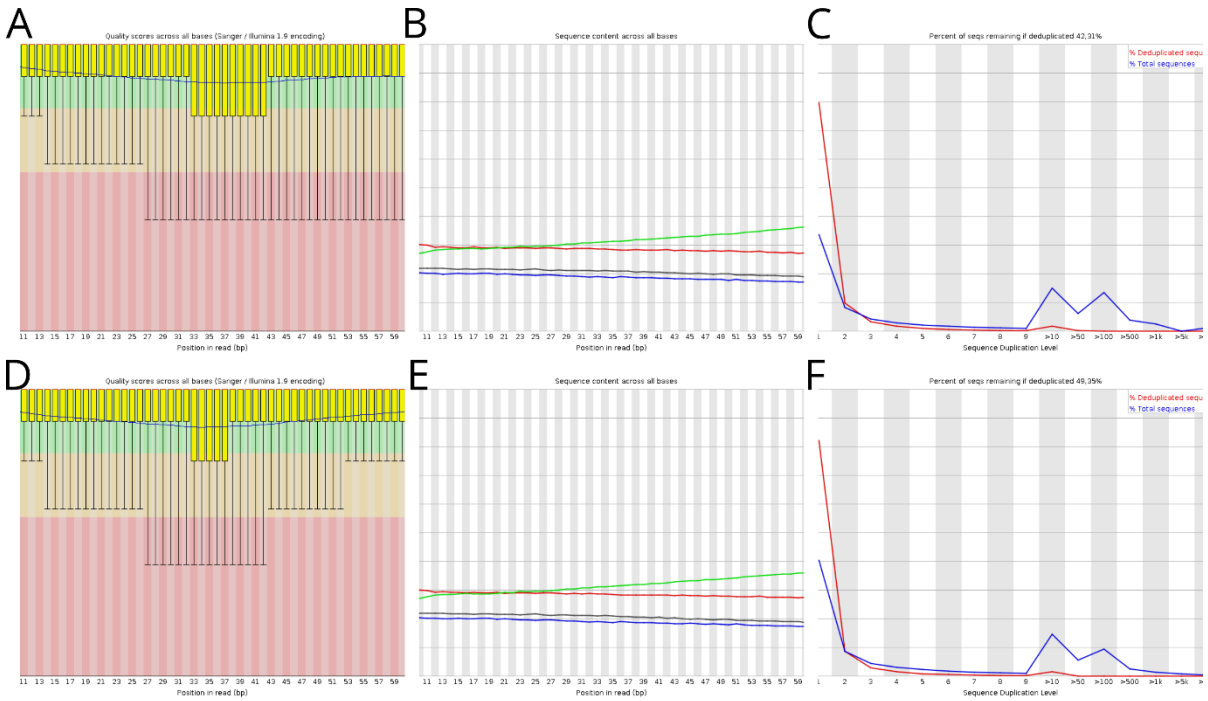


Figure 14: Quality score overview of MACE RNA sequencing results before and after trimming the sequences; ***A, B,*** *and* ***C*** *before trimming;* ***D, E,*** *and* ***F*** *after trimming. The* ***A*** *and* ***D*** *illustrate the median quality of each base position over a sample of tested sequences with a boxplot. The overall sequencing quality could not be improved by trimming, as illustrated by comparing* ***A*** *to* ***D****. The mean base composition over the read is biased towards thymine and adenosine (****B, E****) and a particular fraction of reads shows evidence for duplications (****C, F****).*

The sequencing quality scoring revealed the need for duplication removal, as before trimming, 58% of all reads are marked as duplicate. After the trimming process, this value goes down to 51%. By running the duplicate identification and removal pipeline of *samtools*, these were removed from the data sets before calling the SNP (figure 14). The allele frequency of the progenies was called with default settings. A SNP was considered if it had a minimum quality of 30 over all tested samples and at least one read coverage. The SNP calling detected 13,079 polymorph SNP between Golf and ISR42-8. The allele frequency for these was calculated. The coverage per locus was 25 reads on average, with a median value of 11 reads and a variance of 348. As illustrated by figure 15, the replicates differ according to the coverage. While one sample was characterized by a shallow coverage of many loci, the other two replicates show evidence of higher sequencing depth. The three samples had the read depth range of 1 to 100 reads in common, as well as the decreasing shape of the curve towards higher coverage levels. The allele frequency of ISR42-8 in the pooled sample was, on average, 7.06% over all replicates. This

was about 0.42% less than estimated by the genotyping by sequencing approach. The correlation of the replicates was 0.5. The alternative to reference base ratio of the samples ranged from 0.43 to 0.47, indicating that the reference base was slightly more frequently called than the alternative base. Testing the replicates for significant variation in between the replicates, a variation of $p_{negbin} < 0.001$ was observed, while $p_{zeroinfl} = 0.011$.

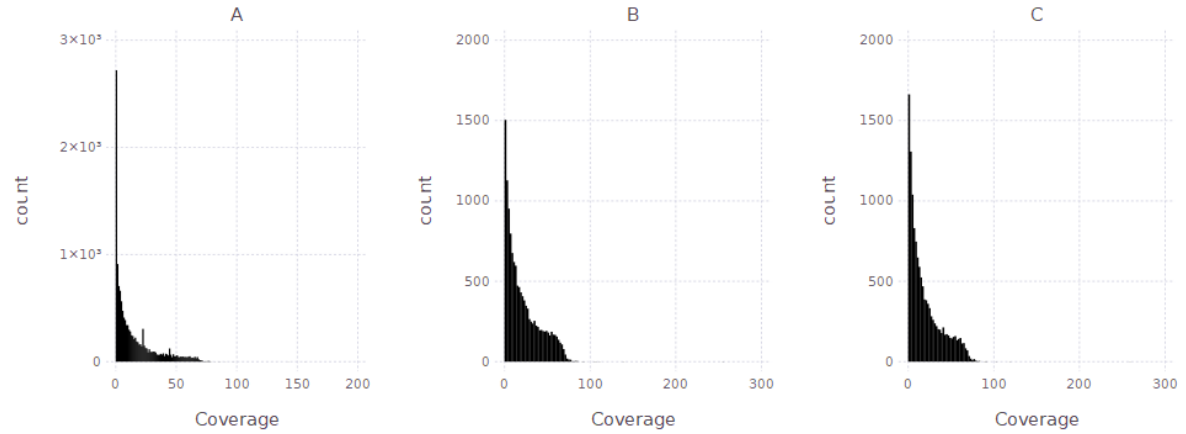


Figure 15: Read depth for each polymorphism locus of the three replicates of the 22nd conventional generation, sequenced by MACE RNAseq. All three replicates have the range of 1 to 100 in common, while they differ in the overall coverage. The replicate in A is sequenced more shallowly, leading to a high amount of low sequenced loci, while the replicates of B and C share the same coverage pattern.

By investigating the genome coverage success of MACE RNAseq, the pericentromeric regions of chromosomes 2H, 5H, and 6H were poorly covered. For the remaining chromosomes, the coverage was improved but still less compared to the telomeres. Both the count of detected polymorphism as well as the coverage per polymorphism were reduced. These results are illustrated in figure 16.

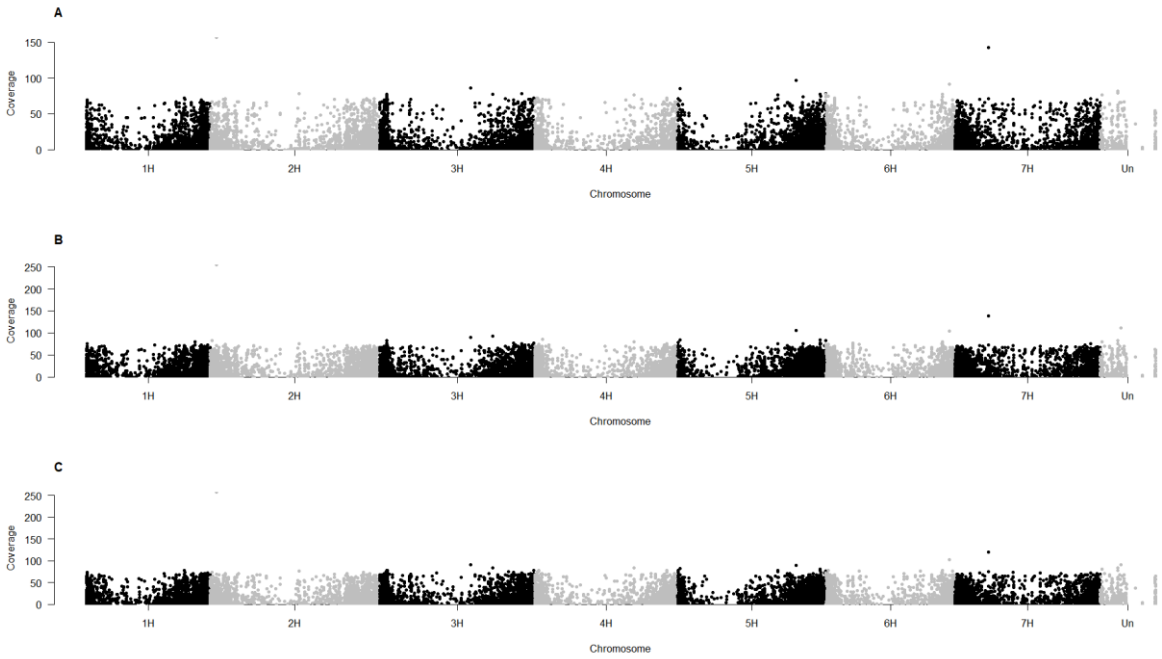


Figure 16: The genome-wide coverage distribution for the three biological replicates (A, B, C) of MACE RNA sequencing. The centromere regions of 2H, 4H, and 5H are poorly covered, while the telomeres are saturated with SNP on an adequate level.

Based on the SNPs allele frequency, the gene annotation-based haplotypes (GH) were compiled. From the 13,079 SNP loci, 6,129 GH were derived. These were classified by an average read coverage of 41, while the median value was 20 and the variance was 3,760. Compared to the SNP level, the coverage was increased by factor two. The observed allele frequency increases to 7.11%. The slight deviation to the SNP level results from the minor disregard of 283 SNPs from the haplotype construction. The median SNP count per haplotype was one, while the mean was 3.5. Most of the haplotypes could not benefit from the haplotyping, as the gene windows are too small for the amount of found polymorphisms. Therefore, the replication correlation did not increase dramatically and had to be reported as 0.59. On both negative binomial as well as zero-inflated regression, the replicates differed on a level of $p < 0.001$. The SNPs per haplotype (D-F) and the read coverage per haplotype are illustrated in detail in figure 17.

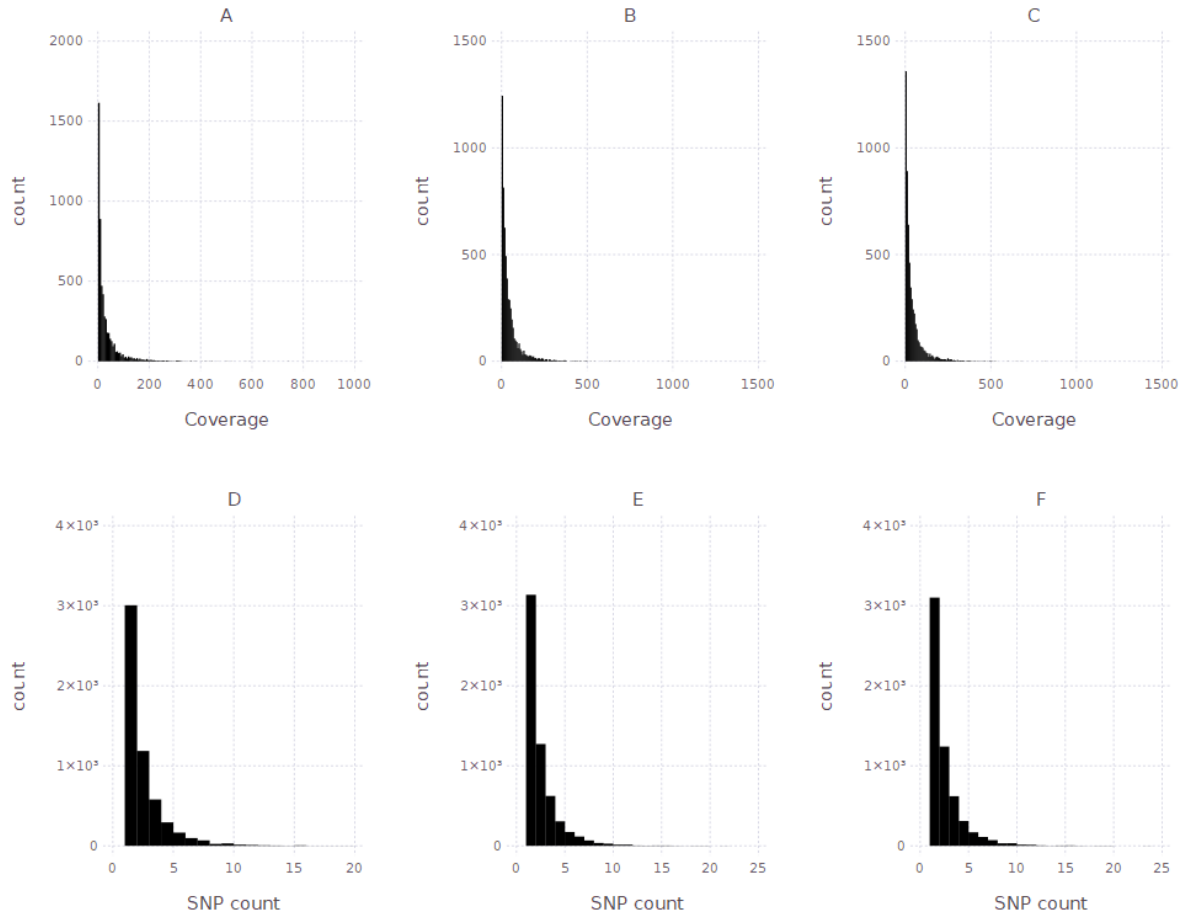


Figure 17: Distribution of the read depth per sample (A-C) and the corresponding SNP count for the gene-based haplotypes (D-F). The replicates are printed in rows. The distribution over the replicates is comparable, highlighting that most of the haplotypes are constructed of only one polymorphism, and analogous to this, only have a shallow read coverage.

The marker-based haplotyping created 3,275 marker haplotypes (MH). These were constructed of 12 SNP on average and two as the median. The allele frequency was 7.02%, and 1161 SNP had not been considered for the compilation. The correlation of the replicates increased to 0.67, compared to 0.5 on the SNP level. The statistics revealed that there was no significant variation between the replicates ($p_{negbin} = 0.346$, $p_{zeroinf} = 0.758$). The average read coverage per haplotype was three times higher than on the SNP level (70), while the median was 36, and the variance was 9,482. The advanced coverage and SNPs per haplotype count were shown in figure 18. Compared to the gene haplotypes, the MH had a much more robust coverage that ultimately resulted in a higher correlation of the replicates.

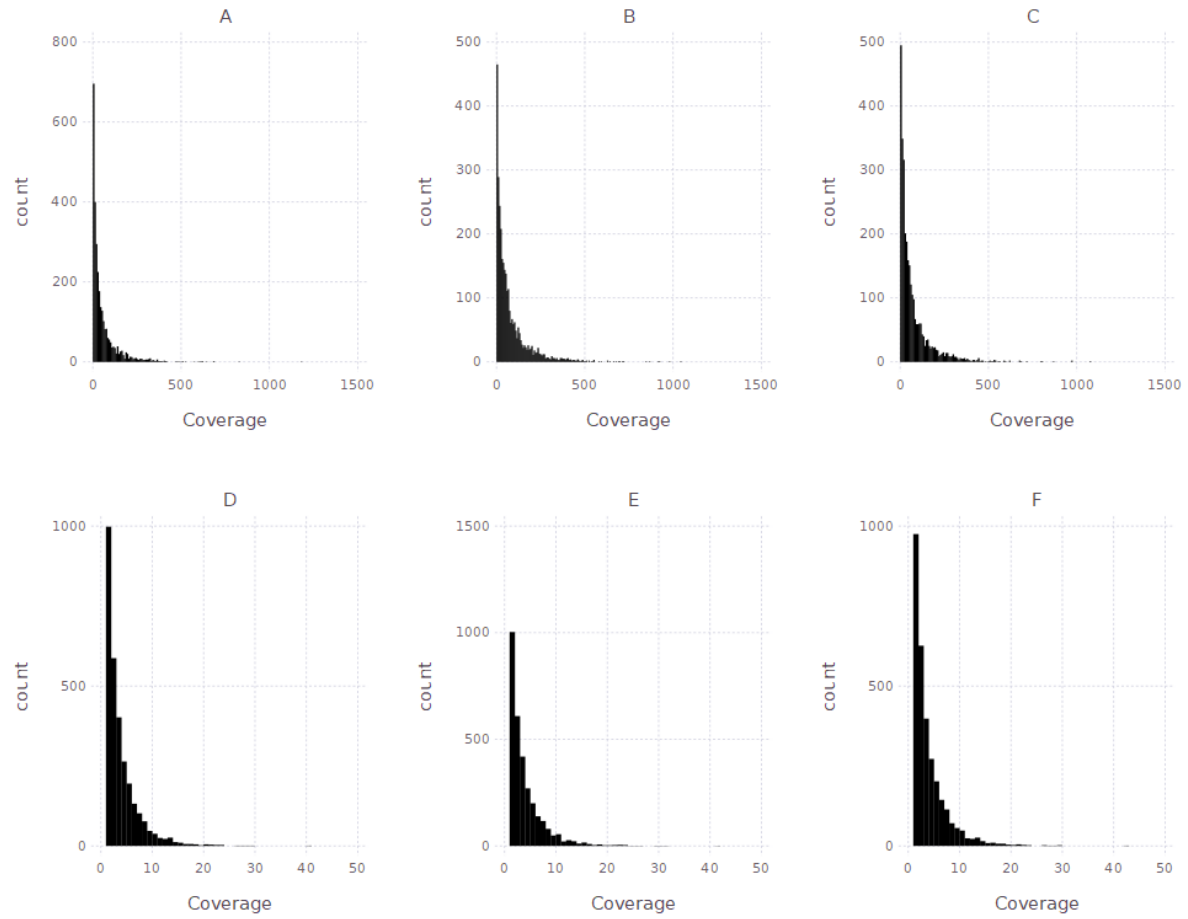


Figure 18: Distribution of the read depth per sample (A-C) and the corresponding SNP count for the marker-based haplotypes (D-F). The replicates are printed in rows. The distributions over the replicates are comparable, highlighting that most of the haplotypes are constructed of only one polymorphism, and analogous to this, only have a shallow read coverage. Nevertheless, the coverage is increased compared to the gene-based haplotypes.

The last remaining haplotype level is the contig level (CH). The CH was classified by 460 haplotypes with an average read depth of 520 and a median value of 238. The variance was high (469,851). On average, 26 SNPs constructed a CH. The allele frequency was 7.06%, and with that, precisely the same value as for the SNP level. The replicates showed a high correlation of 0.9 on the CH level. This was also reflected in the statistical test, where $p_{negbin} = 0.485$ and $p_{zeroinfl} = 0.06$. Compared to the other two haplotyping levels, the CH histogram was typified by a wider distribution of coverage (figure 19). Still, many CH consisted of only one SNP, but this fraction was reduced to less than 15%. This meant more than 85% of all contigs were compiled of two or more SNPs.

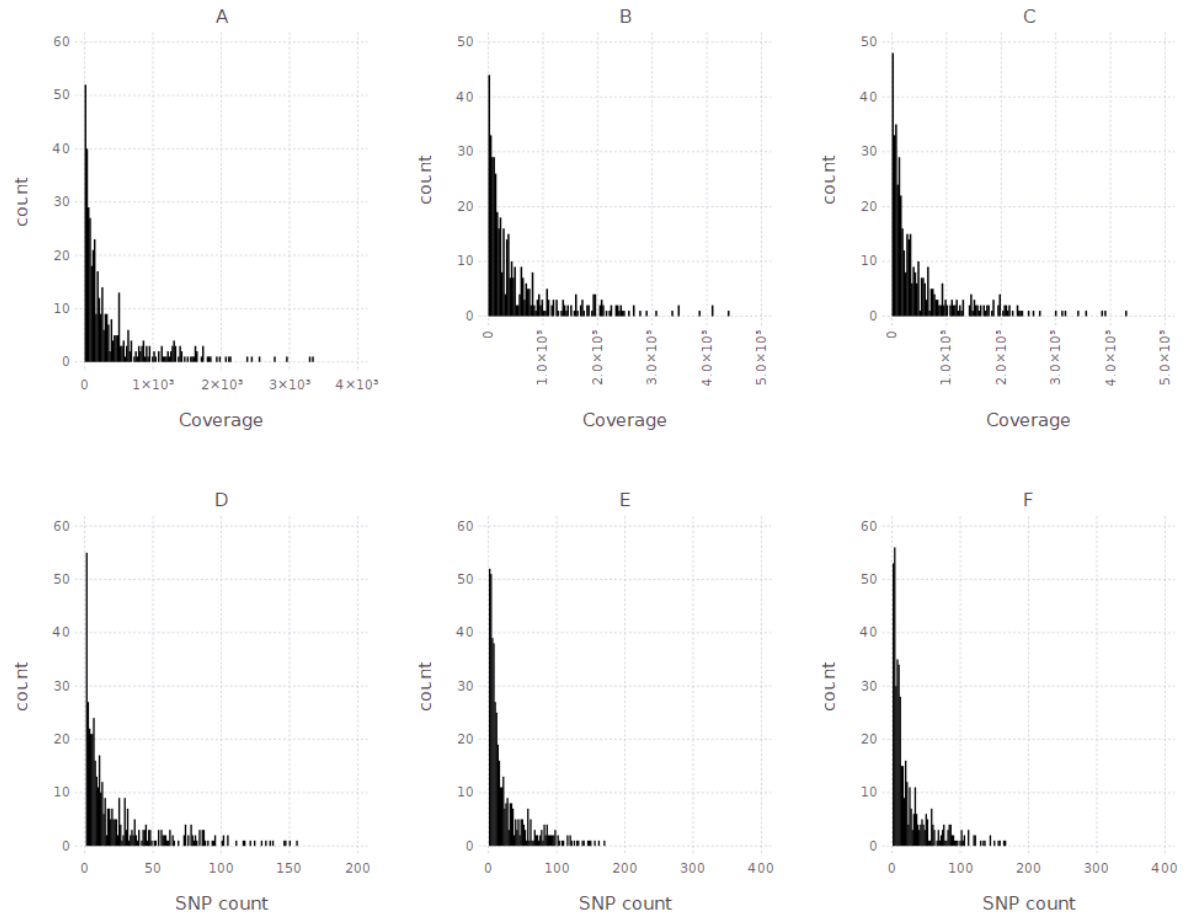


Figure 19: Distribution of the read depth per sample (A-C) and the corresponding SNP count for the contig haplotypes (D-F). The replicates are printed in rows. The distribution over the replicates is comparable, highlighting that most of the haplotypes are constructed of more than one polymorphism, and analogous to this have high read coverage.

By the comparison of the individual to the pooled genotyping approaches, the accuracy of the sequencing method was estimated. The sample with the identical genotype composition as the individual genotyping was tested for correlation and differences in the distributions. 19 KASP markers were linked to a contig, gene haplotypes, and the single SNP. Two markers were unmatched on the marker haplotype level, so 17 were compared to the KASP marker. A correlation of 0.79 on the SNP level, based on a median coverage of 51 reads per locus, was observed. The *Pearson* residual width was 3.85%, while the root means square error was 0.069. The statistical test revealed no significant variation between the observed pool-derived allele frequency and the true KASP assessed allele frequency of the individual genotyping. As illustrated in figure 20A, the correlation was visible but characterized by some outliers. The correlation of the real to observed allele frequency was further increased on the gene-based haplotype level (r^2=0.88). The median coverage value was slightly increased to 61 reads, as along with the residual width, which was increased to 3.92%. The RMSE is 0.0478, and the statistical test proved that no deviation was found between the observed and true allele frequencies (p_{negbin} = 0.9, $p_{zeroinfl}$ = 0.67). In figure 20B, a slight improvement of the correlation could be observed as well. Nevertheless, affected by the marginal increase of the coverage per haplotype, some allele frequencies were still not aligned properly. The marker-based haplotypes, shown in figure 20C, had a further improved correlation of 0.93. The 17 markers were described by a median coverage of 158 reads and a residual width of 3.89%. The RMSE and the statistical test were close to the results of the GH (RMSE = 0.04, p_{negbin} = 0.94, $p_{zeroinfl}$ = 0.67). The red smoothed regression curve in plot 35C follows the dashed optimal line tightly, while still, some allele frequencies deviate. Nevertheless, this deviation is marginal. The contig

haplotypes, shown in figure 20D, were described by a correlation of 0.88, and 917 reads were aggregated to a haplotype on median level for the 19 loci. The statistical test revealed a non-significant deviation of the true and observed data ($p_{nergbin}$ = 0.28, $p_{zeroinfl}$ = 0.62).

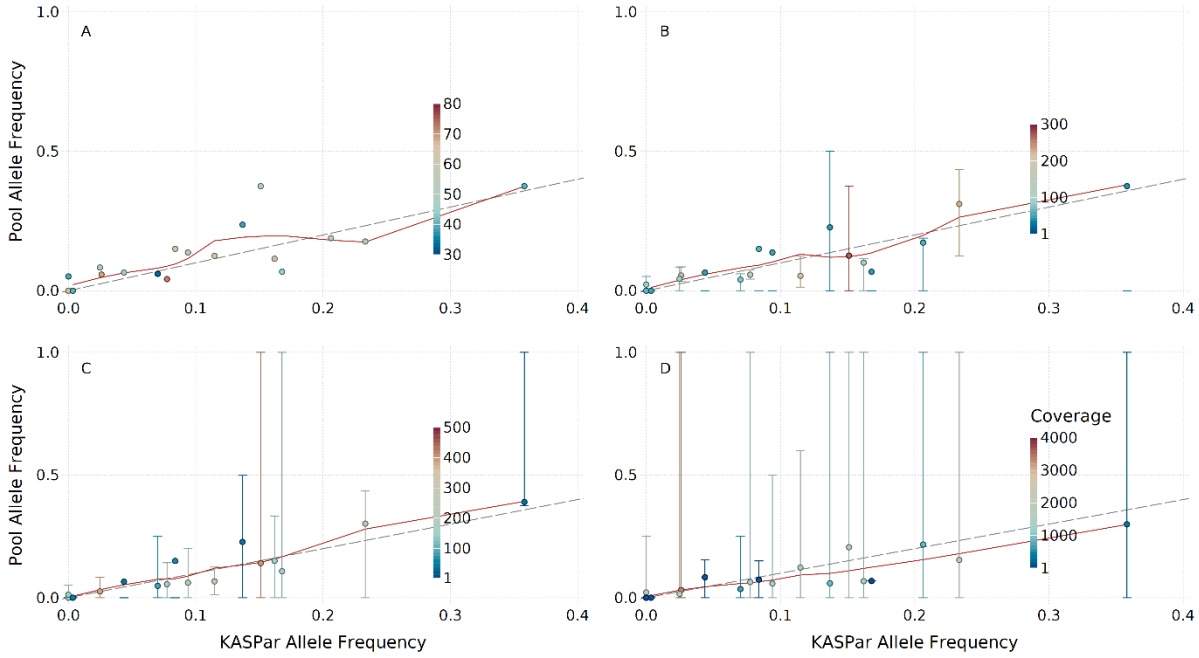


Figure 20: KASPar individual genotyping allele frequency results against the measured allele frequency in the corresponding pool sample for the MACE RNA data. The dashed line indicates the optimal match, where pool obtained values match the individual genotyping ideally. The read curve is a smooth regression curve using all points. The color of the points indicates the coverage per locus, which ranges from 1 to several thousand. Additionally to the points, error bars are drawn. This highlights the margin of single SNP allele frequency that contributed to the overall allele frequency of the haplotype. If no error bar is visible, there is only one SNP contribution information to the haplotype. A – the single SNP comparison. Only one SNP is detected having the exact same position as the KASP markers. B – the gene-based haplotype allele frequency compared to the individual genotyping. C – marker-based haplotype pool allele frequency comparison to real allele frequency. D – contig haplotype-based comparison to the individual genotyping. The pool sequenced sample contains the exact same 288 genotypes that have been tested individually for the 21 KASP loci. As two KASP markers did not meet the quality threshold, they are omitted.

For the MACE RNA sequencing data, the marker-based haplotypes had shown the best performance. The additional coverage created by the contig approach did not lead to an increased inaccuracy.

3.1.3 Whole-genome resequencing

Besides the genotyping by sequencing approach, another untargeted DNA-based sequencing approach was performed. The whole-genome resequencing was accomplished for the two parents Golf and ISR42-8, as well as for two biological replicates of the 22nd conventionally evolved population. One of these two replicates is identical to the sample set, which was genotyped individually for each genotype by KASP markers. Sequencing steps were performed by *Novogene* and generated 400 million high-quality paired-end reads with a length of 150 bp. A quality adjustment was not necessary because the quality of the reads is already high, as shown in figure 21. The high read quality that can be obtained by figure 21A was summarized by a low duplication rate, as shown in figure 21C. After the duplication removal, more than 2/3 of all reads remained unique.

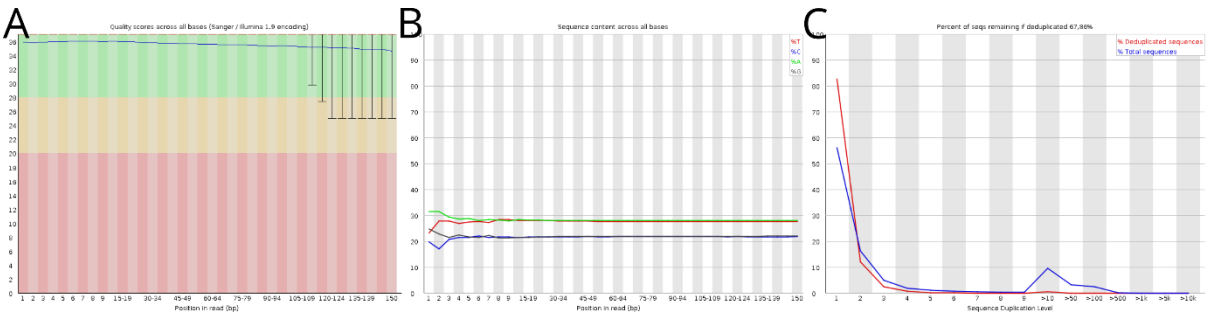


*Figure 21: Sequencing quality overview. **A** illustrates the median quality of each base position over a sample of tested sequences with a boxplot. The overall sequencing quality could not be improved by trimming, was therefore not performed. The mean base composition over the read is unbiased (**B**), and a particular fraction of reads shows evidence for duplications (**C**).*

The SNP calling with default settings was performed using an additional step prior to the pool sample SNP calling. The variants within the parents were called prior, and these polymorphisms were selected, which are homozygote for the genotypes and polymorphic between the two parents. This step revealed more than twenty million polymorphisms. As the chance for false-positive calls was given, the polymorph set was tested against a reference base of SNP positions ((Cunningham et al., 2019)). By including this step, only verified SNP positions were considered for the pooled analysis. This reduced the polymorphisms down to 3,991,259 SNP with a normal distribution of the coverage with a mean of 10 reads per locus (Figure 22A/B), which were identified and clustered to 34,344 GH 5,946 MH and 485 CH. The genome-wide allele frequency measured for the two replicates was 6.54%, with a variance of 0.02 on the SNP level.

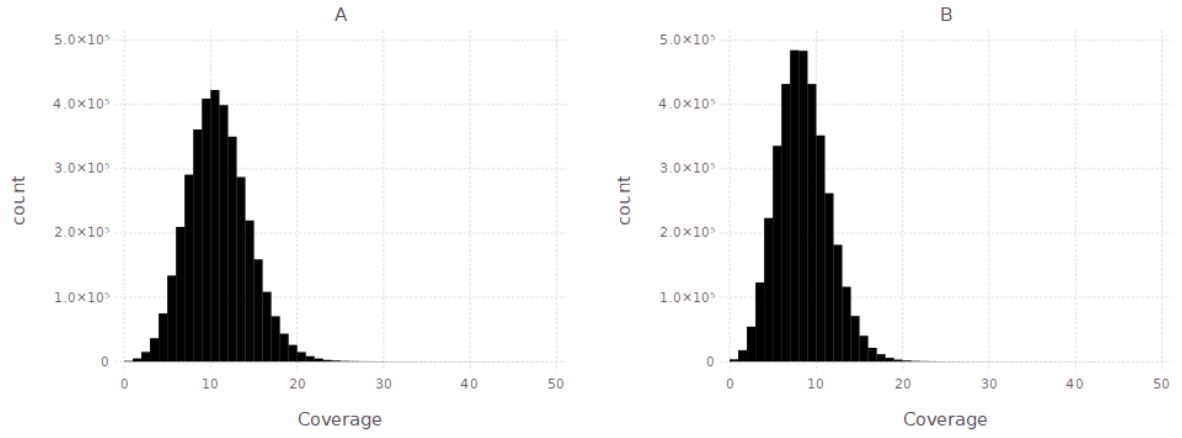


*Figure 22: Read depth distribution for each polymorph locus of the two replicates for the 22nd conventional generation; in summary, more than three million SNP are called. The replicates are separated in **A** and **B** and are comparable in their shape. The replicate in **B** shows evidence of a slightly higher overall sequencing depth, which might cause variation in the allele frequency between the replicates.*

Compared to the other two sequencing approaches, the distribution differs clearly. While a normal distribution could be assumed here, the other two had a more Pareto-like distribution with an excessive amount of single read genotyped loci. Additionally to the desirable locus coverage distribution, the genome-wide coverage showed a very homogeneous pattern. All genomic regions were covered uniformly. Most of the regions were described by coverage levels lower than 50, while some few loci tend to produce more reads than others. Generally, the distribution of the sequencing provides excellent coverage over the entire genome and, therefore, could deliver information about each chromosomal

region. These findings are illustrated in figure 23, where the two sequenced replicates were plotted for the SNPs distribution over the entire genome. The median coverage was nine reads per SNP, while the mean was 9.1 reads, and the variance was 14.91.

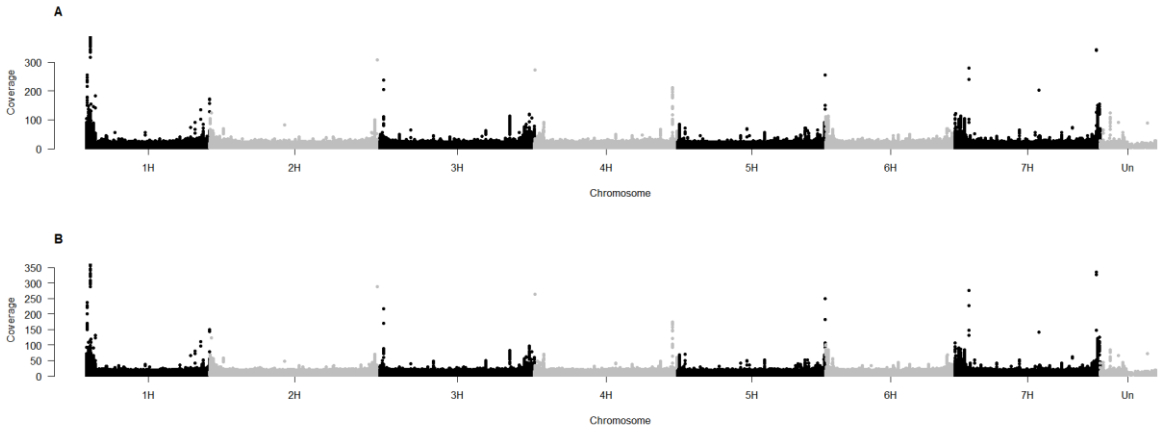


Figure 23: The genome-wide coverage distribution for the two biological replicates (A, B) of whole-genome resequencing. The genome is very uniformly covered on a coverage level of 9 reads per locus on the median, with some regions in the telomeres having a higher coverage of up to 350 reads (Chromosome 1H, short arm).

Although the red coverage of the loci is desirable, no correlation between the replicates was found on the SNP level. The correlation on the other two levels was at least 0.5 while having a comparable read coverage per locus of 6 (GBS) and 11 (MACE RNAseq) median value.

A significantly higher correlation for the replicates was found on the gene haplotype level. The correlation increased to 0.908, while the p-values increase to 0.67 for the negative binomial distribution. The 34,344 gene haplotypes were characterized by a median read count of 494 reads, while the average read count was 963, and the variance was $2x10^9$. This goes along with a median SNP per haplotype count of 55, where the average value was 106 SNP per haplotype. The allele frequency increased to 6.41%, based on more than 3.5 million SNPs. Figure 24 illustrates the read coverage and SNP distribution over the haplotypes. The distribution was dominated by small values and followed the shape of a Pareto distribution.

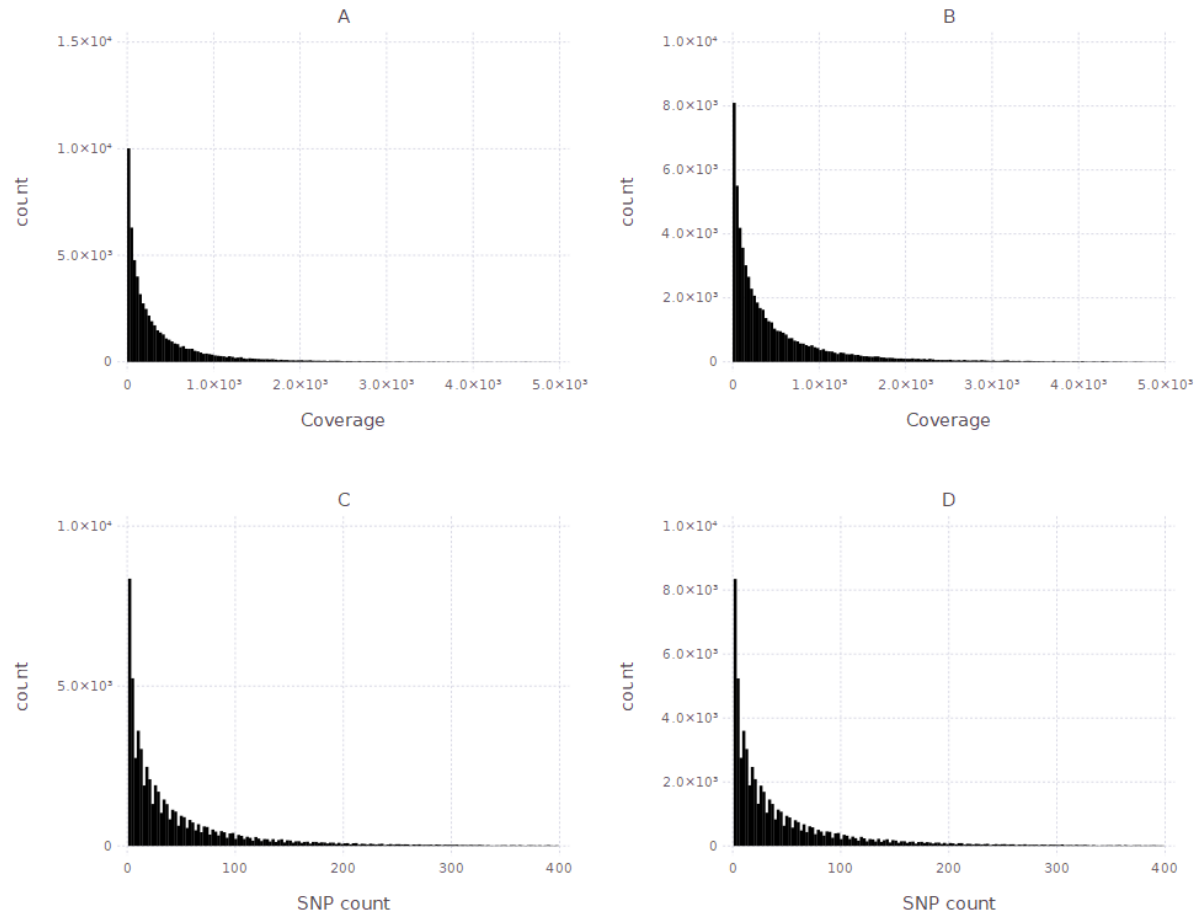


Figure 24: Distribution of the read depth per sample (A – B) and the corresponding SNP count for the gene-based haplotypes (C-D). The replicates are printed in rows. The distribution over the replicates is comparable, highlighting that most of the haplotypes are constructed of only one polymorphism, and analogous to this have only a shallow read coverage.

The marker-based haplotypes were defined by a high correlation of the replicates (0.959). The statistical test supports this statement, as the p_{negbin}= 0.7 and the $p_{zeroinfl}$ = 0.39. The mean read coverage per haplotype was 5,443, while the median was 2,440, and the variance was $1.46x10^8$. The calculated allele frequency for the genotypes was 6.56%, which was based on 3,248,956 SNP, leaving 414,922 unconsidered. The values deviate relevantly from the expected wild-type allele frequency in the third generation (10.5%). Each haplotype was constructed of 598 SNPs (average, median = 272 SNPs). Only 6.12% of all haplotypes contain less than 100 reads. Overall, 5,946 haplotypes were constructed. Figure 25 highlights the distribution of the read coverage and SNP count per marker haplotype. It indicated a much more linearly decreasing slope compared to the gene haplotype distribution, supporting the statement that most marker haplotypes had a higher read coverage than 100.

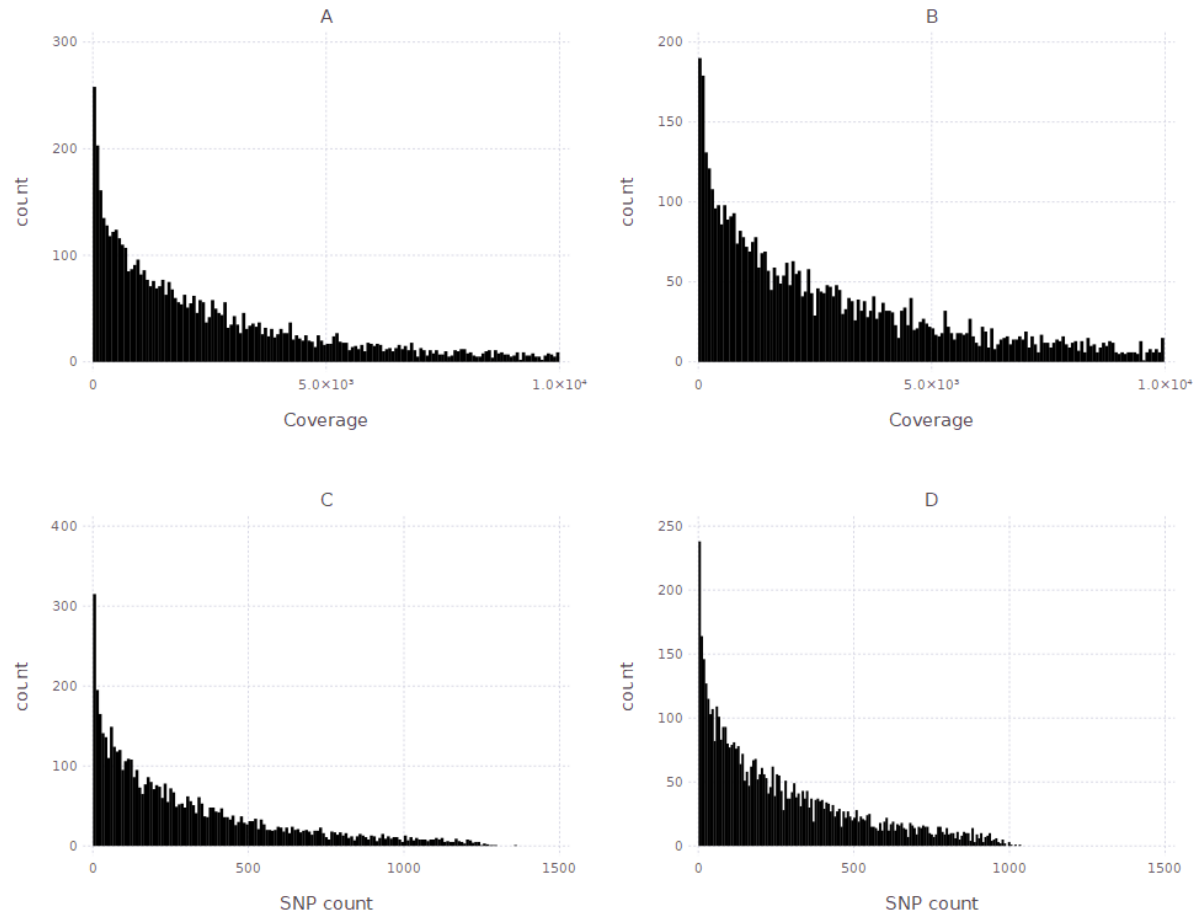


Figure 25: Distribution of the read depth per sample (A – B) and the corresponding SNP count for the marker-based haplotypes (C-D). The replicates are printed in rows. The distribution over the replicates is comparable. A linear decrease of the SNP and read coverage count can be obtained. The coverage is increased compared to the gene-based haplotypes.

The contig haplotypes were represented by a primarily normal shape-like distribution (Figure 26). Four hundred eighty-five contigs were constructed, which were classified by a median read count of 74,134 reads and 74,854 reads on average. The variance was 1.03×10^9. A contig was constructed of 8,250 SNP (median), and the average value for the composition of SNP per contig was 8216. The calculated allele frequency was identical to the allele frequency of the SNP level, as all SNPs were considered for the calculation (6.54%). When the replicates were compared, a high correlation of 0.986 was identified. This statement was supported by the statistical test, where $p_{negbin} = 0.352$ and $p_{zeroinfl} = 0.035$. As can be obtained from figure 26, almost all haplotypes had a coverage higher than 100 reads per haplotype.

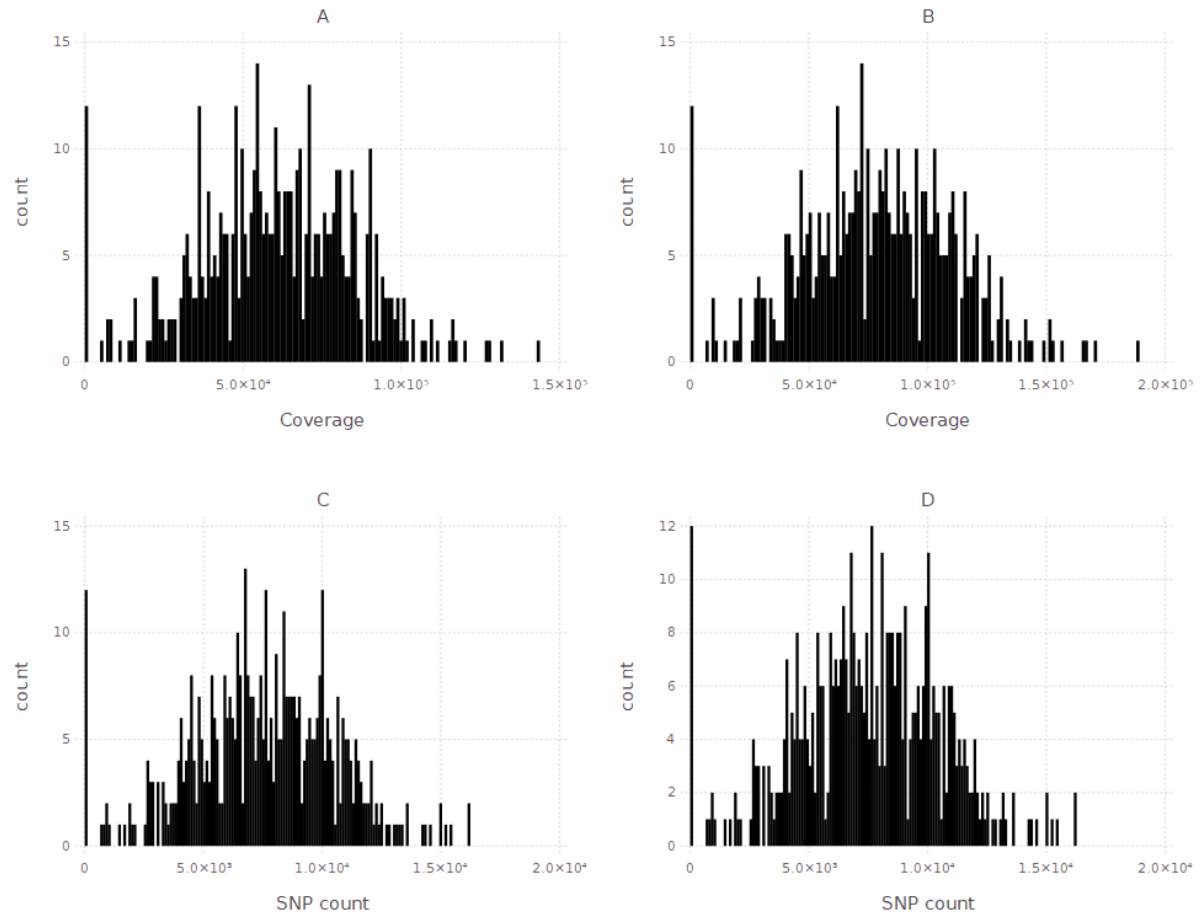


Figure 26: Distribution of the read depth per sample (A – B) and the corresponding SNP count for the contig haplotypes (C-D). The replicates are printed in rows. The distribution over the replicates is comparable, highlighting that most of the haplotypes are constructed of a high number of SNP and reads.

The pooled sample was compared to the allele frequency call generated by individual genotyping to determine the accuracy of the pool estimated allele frequency. Ten SNPs could be linked from the pooled to KASP sequencing directly (Figure 27A). These SNPs were characterized by a median read coverage level of seven reads per SNP. Further information about the coverage distribution was given in figure 26. The correlation was reported to be 0.93, while the negative binomial p-value was 0.878, and the zero-inflated p-value was 0.991. The high correlation might be affected by the low number of regarded data points. When having a look at figure 27A, a high variation from the ideal fit can be observed. Due to the high variance, the accuracy of the pool allele frequency on the SNP level had to be concluded as low. For the haplotypes based on gene annotations, the correlation of the pool and individual genotyped allele frequency was > 0.95, based on 15 haplotypes. When the haplotypes with lower coverage than 100 reads were omitted, the correlation increased to 0.96. All 19 KASP markers could be matched to a haplotype. On a median, a gene-based haplotype was constructed by 197 reads, and the residual width was 3.49%. The root means square error was 0.0255. The variance of the pool estimated to real allele frequency was low, indicated by the red regression line in figure 27B following the direction of the optimal match (dashed line) over the entire range of frequencies. For the marker-based haplotypes, the correlation between the pooled estimate and the real frequency was 0.96. The high correlation was supported by the statistical test, where $p_{negbin} = 0.39$ and $p_{zeroinfl} = 0.63$. The residual width was almost 1% higher than the one of the gene-based haplotypes (4.39%), while the RMSE is 0.029. The median read count per haplotype was 7,749 reads. This leads to a high correlation with a low variance from the optimal estimate value. The contig-based haplotypes had a comparable character. The correlation of the

pooled estimate to the real allele frequency was 0.95. The p-value of the negative binomial distribution, as well as the zero-inflated distribution, supported this high representation of real to pool estimated values (0.46/0.53). All 19 KASP markers were matched, and the median read coverage per haplotype was 81,186. The residual width was 3.91%, and the RMSE is 0.0313. The variation of the estimated value to real value was very low over the whole spectrum of the allele frequency, as indicated by the red regression line in figure 27D.

The aggregation of the allele frequency information from many linked SNPs to haplotype values increased the accuracy significantly when comparing the haplotyping approaches to the single SNP approach. Especially when only regarding the haplotypes with a coverage of more than 100 reads, the accuracy of all methods is on the same level.

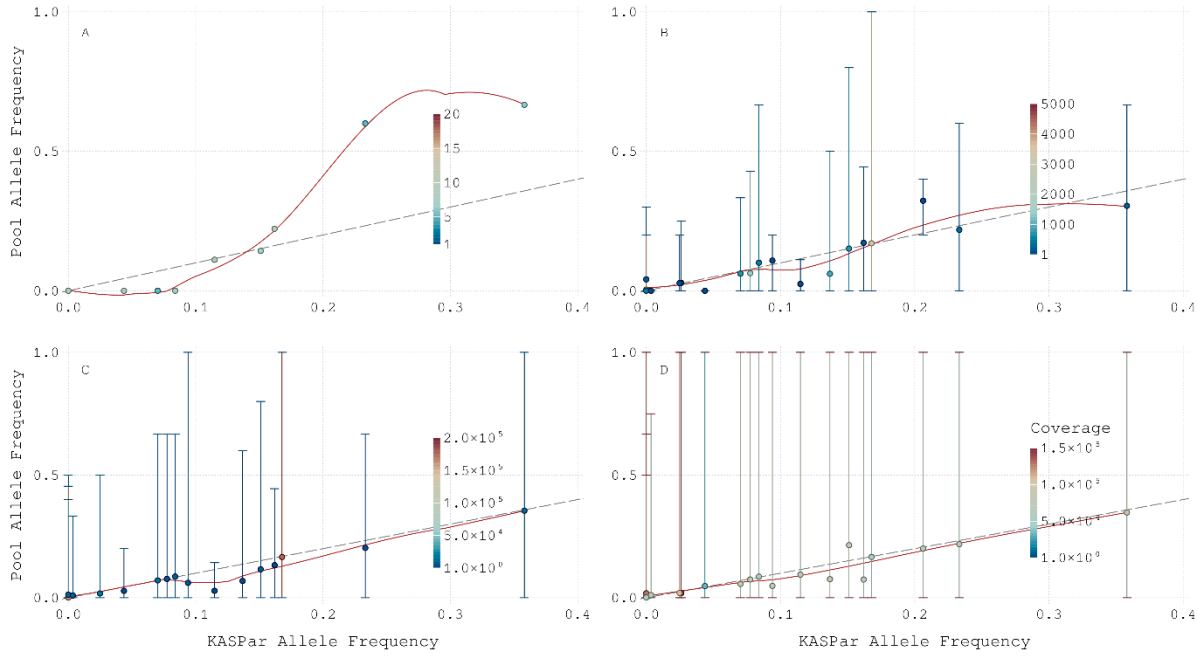


*Figure 27: KASPar individual genotyping allele frequency results against the measured allele frequency in the corresponding pool sample for the **WGrS** data. The dashed line indicates the optimal match, where pool obtained values match the individual genotyping ideally. The read curve is a smooth regression curve using all points. The color of the points indicates the coverage per locus, which ranges from 1 to several thousand. Additionally to the points, error bars are drawn. This highlights the margin of single SNP allele frequencies that contributed to the overall allele frequency of the haplotype. If no error bar is visible, there is only one SNP contribution information to the haplotype. **A** – the single SNP comparison. 10 SNP are detected to have the exact same position as the KASP markers. **B** – the gene-based haplotype allele frequency compared to the individual genotyping. **C** – marker-based haplotype pool allele frequency comparison to real allele frequency. **D** – contig haplotype-based comparison to the individual genotyping. The pool sequenced sample contains the exact same 288 genotypes that have been tested individually for the 21 KASP loci. As two KASP markers did not meet the quality threshold, they are omitted.*

3.4.4 Comparison of the three methods

To elaborate on the benefits of the haplotyping approach based on any of the used sequencing approaches, the size of constructed haplotypes might play a crucial role in terms of read coverage. As the extension of haplotypes did not follow a static sliding window approach or related ideas, the question arises if the accuracy of the haplotype frequency depended on the extension size and the chromosomal position. The deviation of the pool estimated frequency to the real frequency was tested in a linear regression against the extension size and the chromosomal position to uncover this question. The position had not shown any effect on the deviation for both WGS and MACE haplotypes. The same was true for the extension size for the WGS haplotypes and for the marker MACE haplotypes. For the gene haplotypes of MACE, a regression of $p=0.0065$ was detected. As MACE only covers expressed 3' ends

of genes, the bias might be related to this biological background. For GBS, no correlation could be calculated (Figure 28).

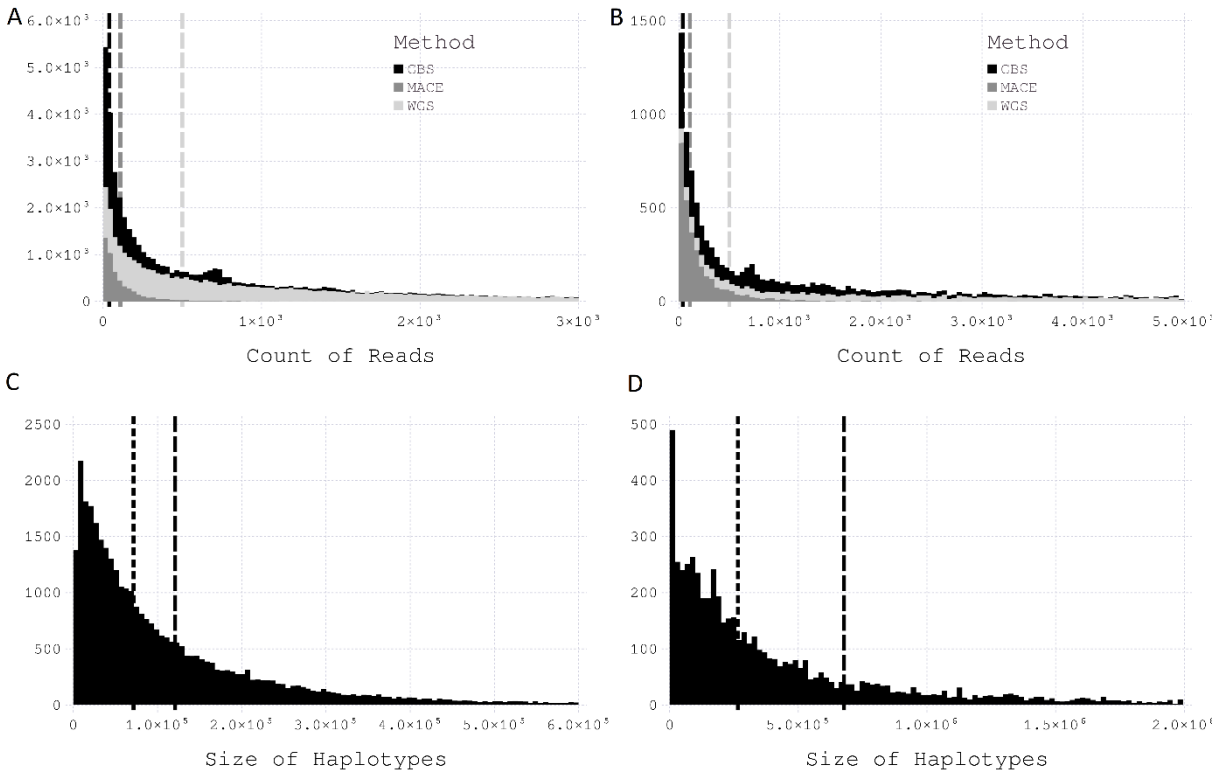


Figure 28: Read count per haplotype for the three applied sequencing methods with their mean value as a vertical line and the distribution of the haplotype window size for gene-based and marker-based haplotypes. ***A*** *– Count of reads per gene-based haplotype for the three applied sequencing methods.* ***B*** *– same as for A, but based on marker-based haplotypes.* ***C*** *– Size of the gene-based haplotypes in base pairs. The dotted line indicates the median value, and the dashed line represents the mean value.* ***D*** *– same as C for marker-based haplotypes.*

Furthermore, it is assumed that 300 genotypes in a pool are already a good representation of the whole population. Based on this, the hypothesis is constructed that the replicates should be highly correlated if the sequencing technique did not introduce a significant error. As shown in table 7, the hypothesis could generally hold true for WGrS, where the correlation of the replicates was best with a value of 0.91 and 0.958 for GH and MH levels, respectively. For the other two approaches, the replicates showed lower correlations on any given level. Both MACE and GBS showed evidence of significant variation between the replicates on the negative binomial level of the model, where MACE > GBS. The zero-inflated model showed high significant variation for the GH level of WGrS while being far from significant on the MH level. GBS showed high significance on all given levels, indicating a high variation between the three tested pool samples. Generally, the correlation of samples was lowest on a single SNP level while increasing with the size of haplotype blocks (SNP < GH < MH < contig haplotype). Further details can be obtained from table 7.

Table 7: Comparison of sequencing methods for different haplotyping levels on the comparability over the replicates. ***nC*** *– not calculated due to excess of time penalty. Replicates were correlated by Pearson correlation, and significant variations were calculated between the replicates in a generalized linear model with a zero-inflated and negative binomial model. Correlation and p values were calculated for all 4 applied haplotyping levels for all three sequencing methods.*

Replicate comparison		**GBS**	**MACE**	**WGrS**
Pearson correlation	SNP level	*0.55*	*0.5*	*0.132*
	GENE-BASED HAPLOTYPE level	*0.64*	*0.59*	*0.908*
	MARKER-BASED HAPLOTYPE level	*0.75*	*0.67*	*0.958*
	Contig Haplotype level	*0.75*	*0.9*	*0.986*
Negative Binomial	SNP level	*< 0,001*	*< 0,001*	*nC*
	GENE-BASED HAPLOTYPE level	*< 0,001*	*0.007*	*0.67*
	MARKER-BASED HAPLOTYPE level	*0.17*	*0.346*	*0.522*
	Contig Haplotype level	*0.91*	*0.485*	*0.352*
Zero-inflated	SNP level	*< 0,001*	*0.011*	*nC*
	GENE-BASED HAPLOTYPE level	*< 0,001*	*0.006*	*<0,001*
	MARKER-BASED HAPLOTYPE level	*< 0,001*	*0.758*	*0.2036*
	Contig Haplotype level	*0.48*	*0.06*	*0.035*

On the genome-wide scale, the methods differ clearly in terms of genome-wide coverage and distribution. Using GBS revealed 17,026 gene haplotypes, three times more than MACE (5,919), but only half the count detected by WGS (34,344). Besides these qualitative differences, the quantitative aspect of each haplotype showed to have a crucial role in the accuracy of the allele frequency estimate. While GBS showed a median read count of 39 reads per gene haplotype, MACE decreased to the half value (21), while WGrS has produced 494 reads per gene haplotype. The variation becomes more distinct when the average values are reported (Table 8). As the read coverage per SNP is more or less static, the coverage per haplotype is strongly influenced by the SNP count per haplotype. While on the median, only a single SNP per haplotype could be found for MACE, three SNP were detected for GBS and 55 SNP for WGrS. The sequencing coverage can be adjusted based on the expected polymorphism amount – higher numbers of polymorphism can result in a lower sequencing depth, while low rates of polymorphisms have to be overcome by an increased sequencing depth.

Table 8: General information on sequencing coverage, locus tags, and haplotype saturation for GBS, MACE and WGrS. Besides basic statistics for read and SNP count per SNP/haplotypes, the count of tagged SNP and haplotypes is reported.

***Count over the* whole sample**	SNP	*82,435*	*13,079*	*3,991,259*
	GH	*17,026*	*5,919*	*34,344*
	MH	*4,702*	*3,275*	*5,946*
	Contig	*483*	*460*	*485*
Median read count	SNP	*6*	*11*	*9*
	GH	*39*	*21*	*494*
	MH	*234*	*36*	*2,440*
	Contig	*2645*	*238*	*74,134*
Average read count	SNP	*28*	*25*	*9.1*
	GH	*137*	*42*	*963.2*
	MH	*450*	*70*	*5,443.26*
	Contig	*4,837*	*520*	*74,854.8*
Variance read count	SNP	*3,329*	*348*	*14.91*
	GH	*62,413*	*4,109*	*2x10⁶*
	MH	*430,575*	*9,482*	*1,46x10⁸*
	Contig	*2,9*10^7*	*469,851*	*1,03x10⁹*
Allele frequency	SNP	*7.48 %*	*7.06 %*	*6.54 %*
	GH	*7.48 %*	*7.09 %*	*6.56 %*
	MH	*7.43 %*	*7.02 %*	*6.56 %*
	Contig	*7.48 %*	*7.06 %*	*6.54 %*
Median SNP count	GH	*3*	*1*	*55*
	MH	*8*	*2*	*272*
	Contig	*142*	*2*	*8,250*
Mean SNP count	GH	*4.2*	*2*	*106*
	MH	*14.33*	*12*	*598*
	Contig	*153.94*	*26*	*8,216*

Besides the quantitative statistics, the qualitative aspect was crucial for the selection of the appropriate sequencing strategy. The allele frequency comparison of pooled and individual genotyping obtained for 19 selected KASP markers gave a valid illustration of this attribute. As illustrated in figure 29, the single SNP-based allele frequency was associated with a high deviation of the ideal 1:1 correlation. The deviation was decreased when the gene-based haplotyping approach is applied. The best correlation can be found for WGrS gene haplotyping, followed by MACE gene haplotyping. Generally, it can be highlighted that increased coverage of a haplotype will result in a higher pool obtained allele frequency correlation to the real allele frequency.

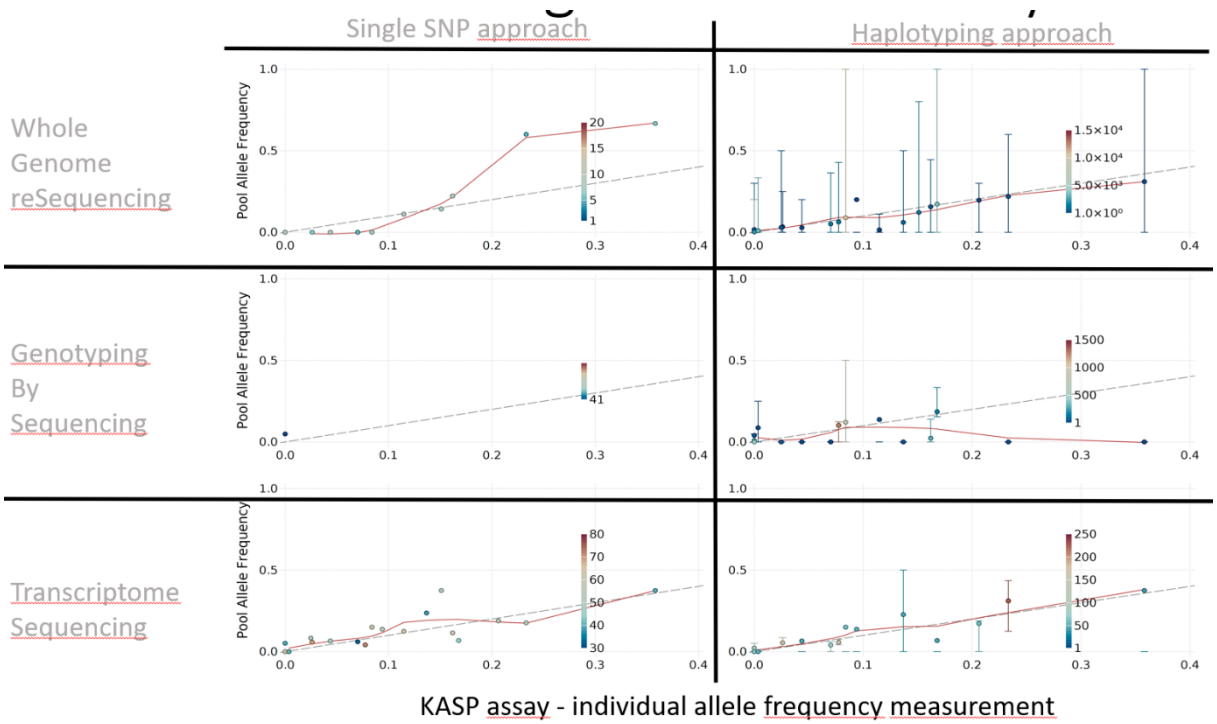


Figure 29: Comparison of all three sequencing methods to individual KASP assay sequencing for single SNP and gene haplotyping approaches. The dashed line indicates the optimal match, where pool obtained values match the individual genotyping ideally. The read curve is a smooth regression curve using all points. The color of the points indicates the coverage per locus, which ranges from 1 to several thousand. Additionally to the points, error bars are drawn on the right side for the haplotypes. This highlights the margin of single SNP allele frequencies that contributed to the overall allele frequency of the haplotype. If no error bar is visible, there is only one SNP contribution information to the haplotype. The pool sequenced sample contains the exact same 288 genotypes that have been tested individually for the 21 KASP loci. As two KASP markers did not meet the quality threshold, they are omitted.

As all of the presented statistics did not illustrate the allele frequency on the genome, the haplotype allele frequency for marker and gene-based haplotypes were illustrated in figures 30 and 31. Both plots reveal that the highest accuracy was obtained from WGrS. The accuracy was defined by the deviation of the allele frequency of neighboring haplotypes. Due to linkage disequilibrium, neighboring haplotypes should have a comparable allele frequency with low deviation from each other. If one single haplotype is highly different from surrounding haplotypes in its allele frequency, it has to be assumed to be a false positive call. MACE and especially GBS show evidence of an increased amount of such false positive hits all over the genome, while few of those deviating haplotypes can be observed for WGrS. Besides these errors, the general allele frequency pattern was comparable over all three methods. In comparison to the physical map, the genetic map (based on the *9KiSelect* chip (Comadran et al., 2012b)) was characterized by an advanced resolution of the telomeres. As illustrated by figure 31, the pericentromeric regions were described by a highly reduced recombination, resulting in big linkage blocks of several hundred million base pairs.

Concluding, all three tested sequencing methods revealed pool sequencing results with a rather high correlation to individual genotyping results. While MACE and GBS were distinguished by a medium correlation of pool samples and to individual sequencing results, WGrS was portrayed by a high correlation of pool to individual sequencing, high correlation of replicates, and a superior genome-wide coverage and polymorphism detection level. Furthermore, WGrS illustrated little error in respect to expected linkage disequilibrium of haplotype allele frequency correlations. The higher errors in GBS and MACE could be related to technical or biological backgrounds.

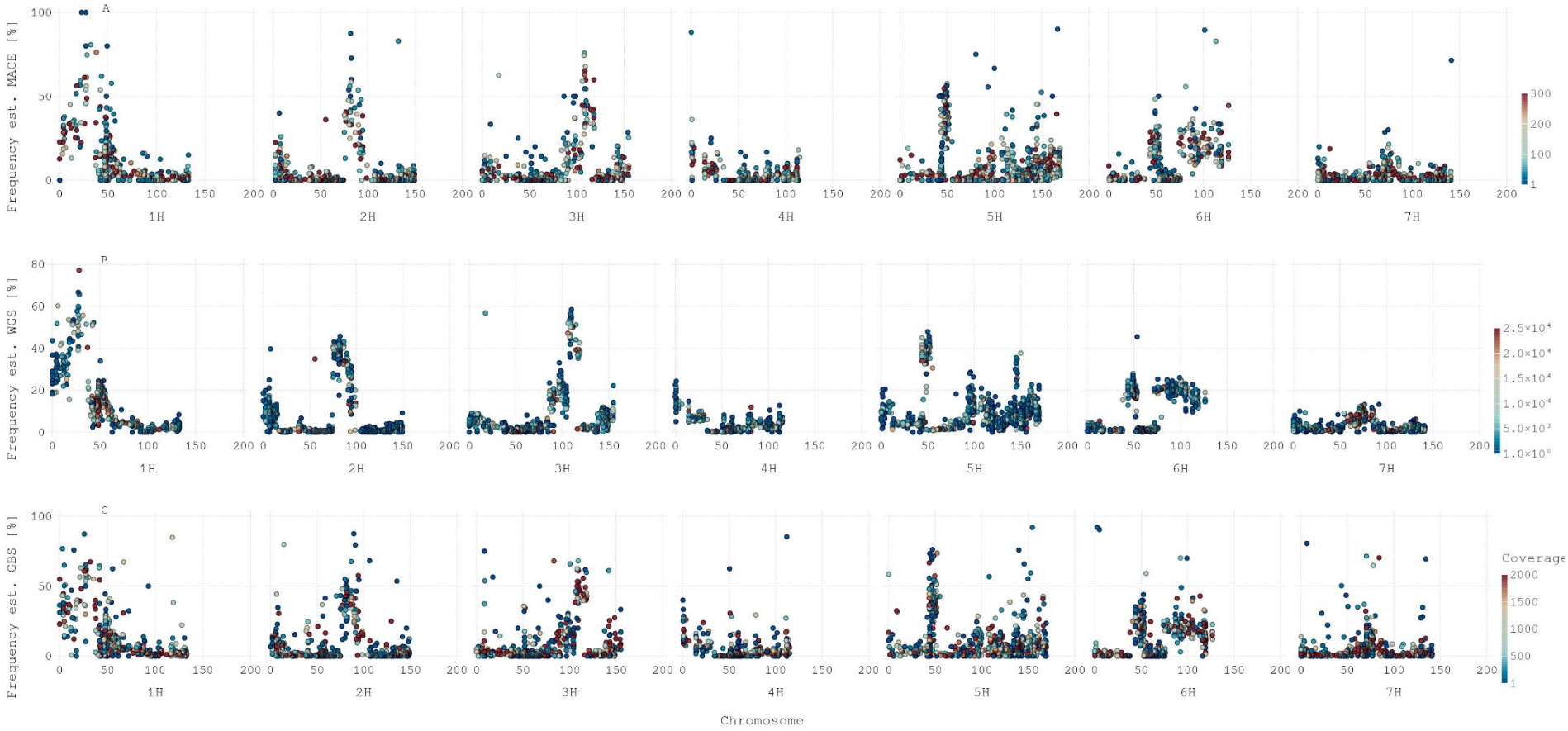


Figure 30: Genome-wide allele frequency on a genetic map for marker-based haplotypes. The ISR42-8 allele frequency is plotted in % (y-axis) against the genomic position (x-axis), split by chromosome, and illustrated in centi morgan. Each dot represents a marker haplotype, and the color is related to the read coverage. ***A*** *– MACE RNA sequencing output,* ***B*** *– WGrS,* ***C*** *– GBS.*

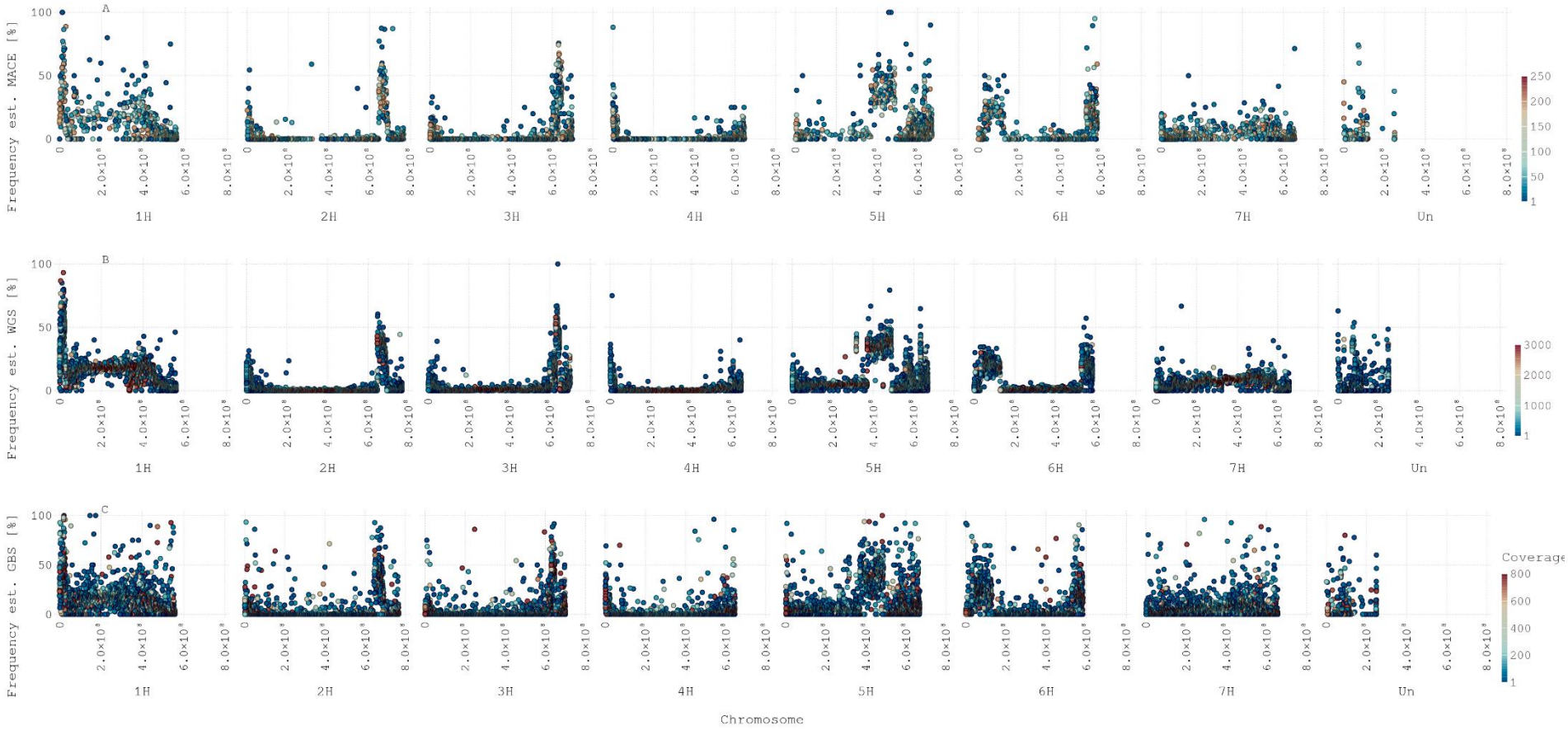


Figure 31: Genome-wide allele frequency on a physical map for gene-based haplotypes. The ISR42-8 allele frequency is plotted in % (y-axis) against the genomic position (x-axis), split by chromosome, and illustrated in base pairs. Each dot represents a gene haplotype, and the color is related to the read coverage. ***A*** *– MACE RNA sequencing output,* ***B*** *– WGrS,* ***C*** *– GBS.*

3.2 Phenotypic characterization of founders

The two genotypes Golf and ISR42-8, diverge in many phenotypic characteristics. The most obvious is the morphological variation. Golf is a tall and fast-growing variety, while ISR42-8 is a more grasslike genotype. In the initial leaf building phase, the leaves stick close to the ground. In the following tillering stage, tillers emerge slower, and the internodes are much longer, compared to Golf, while characterized by lower stem stability. ISR42-8 growth morphology is more related to a classical grass type that can be found on pasture land. It has many leaves close to the ground, from which single stems with ears emerge. Contrasting to this, Golf shoots up high relatively fast and has an advanced growth height and rate compared to ISR42-8 and the more modern cultivar Scarlett. Furthermore, the root system of ISR42-8 tends to be more distinct than Scarlett (Naz et al., 2014).

The variation on the phenological level is adequately high. While Golf is a typical barley spring type with low requirements for stratification, ISR42-8 has a higher demand for stratification. Additionally, the growth period of ISR42-8 is adapted to a high thermal environment. This results in a long growth period until seeds are ready to be harvested. Compared to Golf, the flowering is delayed by several weeks and is initialized at a time when Golf genotypes have reached the ripening stage. The late flowering is accompanied by an untypical extrusion of the anthers while actively flowering. Interestingly, ISR42-8 tends to produce many leaves under optimal water supply, which leads to high transpiration and water usage. If the water supply is reduced, these leaves can be reduced quickly by apoptotic processes.

The yield characteristics are an additional variation between the two genotypes. While ISR42-8 has a habit of producing few tillers with long ears, the ears of Golf are rather short, but the number of tillers is notably higher. Moreover, the corn contrasts in its morphology. Golf is classified by thick and normal-sized kernels in comparison to other cultivars. Contrasting to this, ISR42-8 produces longer and thinner kernels. When it comes to harvesting, Golf shows clear evidence of domestication. This genotype has a stable ear with easy thresh grains. In contrast to this, ISR42-8 has a highly unstable and brittle ear with awns and rachis that are hard to separate from the corn (Schmalenbach et al., 2011). This makes automated threshing without additional processing of the kernel impossible.

Besides the physiology, morphology, and yield characteristics, an important aspect of variation is based on the resistance and susceptibility of barley against biotic stressors. Like powdery mildew or leaf rust, fungi can result in a severe yield decrease (Lim & Gaunt, 1986; Walters et al., 1984). Advanced performance regarding disease response has been captured for ISR42-8 regarding leaf rust, powdery mildew, and *Fusarium graminearum* tolerance or resistance (figure 32).

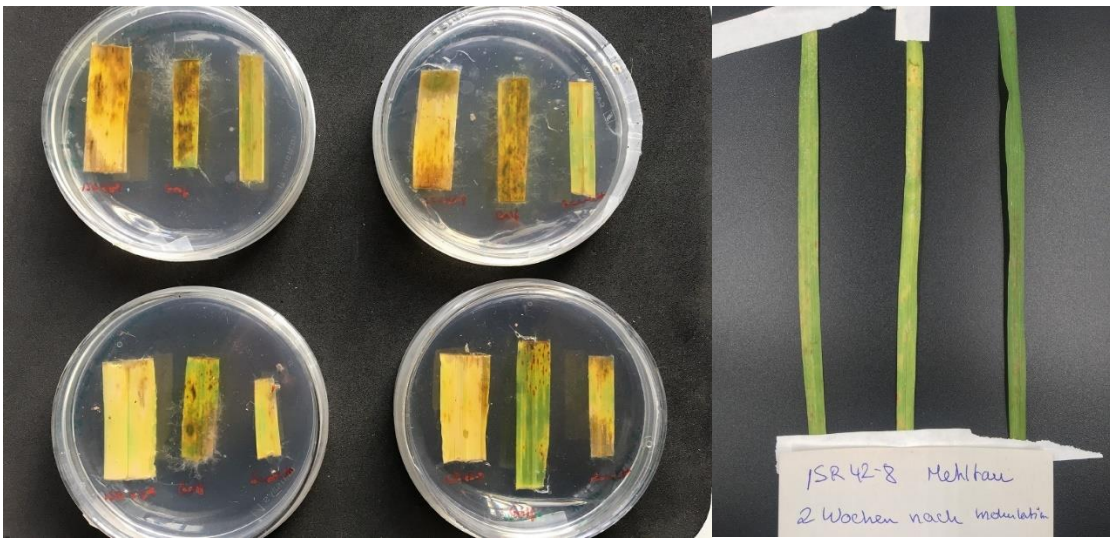


Figure 32: Infection level on fresh leaf tissue. Left – infection of leaf tissue with Fusarium graminearum spores (inoculated with 100,000 spores per µl suspension), ISR42-8 left, Golf in central, and Scarlett in the right position of the tube. Leaf tissue applied on agarose gel, enriched with Kinetin, according to (Bedawy et al., 2018). *Right – inoculation with powdery mildew on seedling stage ISR42-8 plants. Low infection level observed.*

Concluding, the founders of the population are highly diverse in their genetic background. The combination of these genotypes can lead to a diverse combination of genotypes in the population.

3.3 Characterization of populations

3.3.1 Root phenotype variation in a hydroponic experiment

In this part of the root analysis, twenty-four traits were measured. To ensure the variation in a trait was not associated with increased overall growth in the first couple of weeks, normalization was performed based on the leaf per tiller count. After six weeks, the average plant observed in the experiment was characterized by a mean leaf per tiller ratio of three. Therefore, the leaf number was divided by the tiller number and used as a corrector to address the bias introduced by the growth speed, as illustrated in equation 12.

$$\frac{1}{\frac{\frac{Leaf\ number\ per\ plant}{Tiller\ number\ per\ plant}}{median(\frac{Leaf\ number\ per\ plant}{Tiller\ number\ per\ plant})}}$$

Equation 12: Correction by growth – equalizing variation, which is based on growth speed variations.

The corrected traits were tested for significant variations between the organic and the conventional group. Among the traits, 14 traits revealed a significant variation between the two groups based on adjusted data. These include aerenchyma area, average late metaxylem area, late metaxylem area, the ratio of length per volume, the root average diameter, the root cross-section, root length, root mass density, root surface area, stele area, tiller number, plant height, total cortical area, and vascular bundle area. Table 9 illustrates the variation between the organic and conventional groups, the identified 95% confidence interval as well as the standard deviation for organic and conventional groups separately. Ultimately, the p-value is presented in the last column.

Table 9: Analysis of variation for all traits observed based on adjusted data. For each trait, the abbreviation, the difference between organic and conventional (positive values are related to higher in organic systems), the upper and lower 95% confidence interval, and the adjusted p-value for the root trait comparison between the organic and conventional populations are illustrated.

			Conf. Interval		
Abbreviation	**Trait**	**Difference**	**low**	**Up**	**P adjusted**
days.of.Emergence	*Days to emergence*	0.951	0.511	1.392	0.000
TLR	*Tiller number [1/plant]*	-0.298	-1.195	0.600	0.827
LN	*Leaf number [1/plant]*	2.473	-0.846	5.792	0.219
TPH	*Tiller number [1/plant]*	3.118	0.552	5.684	0.010
RSD	*Root system depth [cm]*	5.671	1.203	10.138	0.006
SDW	*Shoot dry weight [g]*	0.312	0.011	0.612	0.039
RL	*Root length [mm]*	584.480	326.312	842.647	0.000
PA	Projected Area [cm²]	11.102	-0.069	22.273	0.052
RSA	*Root surface area [mm²]*	24.433	-10.253	59.120	0.265
RAD	*Root average diameter [mm]*	-0.008	-0.045	0.029	0.938
RV	*Root volume [mm³]*	0.216	-0.235	0.666	0.603
L.V	*Ratio length per volume*	341.173	220.528	461.818	0.000
RDW	*Root dry weight [g]*	0.025	-0.094	0.146	0.947
SRL	*Specific root length [mm]*	418.488	-332.368	1169.344	0.474
RMD	*Root mass density [g/cm³]*	-0.003	-0.034	0.029	0.996
RSR	*Root surface area [cm]*	-0.080	-0.150	-0.010	0.019
RXA	*Root cross section [mm²]*	0.005	-0.034	0.045	0.988
LMN	*Late metaxylem number*	-0.258	-0.575	0.059	0.154
ALMA	*Average Late metaxylem area [mm]*	-0.001	-0.001	0.000	0.001
TCA	*Total cortical area [mm²]*	0.004	-0.025	0.032	0.988
CCA	*Cortical cell area [mm²]*	0.008	-0.020	0.036	0.898
AA	*Aerenchyma area [µm²]*	-0.005	-0.006	-0.004	0.000
SA	Stele area [mm]	0.002	-0.013	0.017	0.991
VBA	*Vascular bundle area [mm²]*	0.006	-0.033	0.046	0.978

The highest significant variation between the organic and conventional groups could be found for the aerenchyma area. These air channels in the root were much more prominent in the conventional (ϕ = 0.041 µm²) compared to the organic genotypes (ϕ = 0.036µm²). Remarkably, the standard deviation in both systems was small. Therefore the 95% confidence interval did not overlap. The second most

significant trait was the average late metaxylem area. Similar to the aerenchyma area, the standard deviations were particularly low, and the conventional system was characterized by a higher area than the organic group (0.0033 to 0.0027). The root length and the root length per volume were described by higher values in the organic group. Besides the higher values, the organic group was characterized by a far higher standard deviation when compared to the conventional group. Further details could be obtained from figure 33 and table 9**.** For most adjusted traits, the lowest value had to be reported for ISR42-8. This effect was not present when values without adjustment were considered (table 10, figure 34). A comparable observation was made for the plant height and the leaf number. While the organic system was characterized by a 2.5 higher leaf count per plant and a 3.1 cm taller growth habit based on uncorrected values, the effect was reversed by the adjustment (1.45 less leaf count in organic, 7.35 cm lower growth than conventional). This was accompanied by a reduced tiller number in the organic system (2.97 fewer tillers on unadjusted, 1.39 fewer on adjusted base).

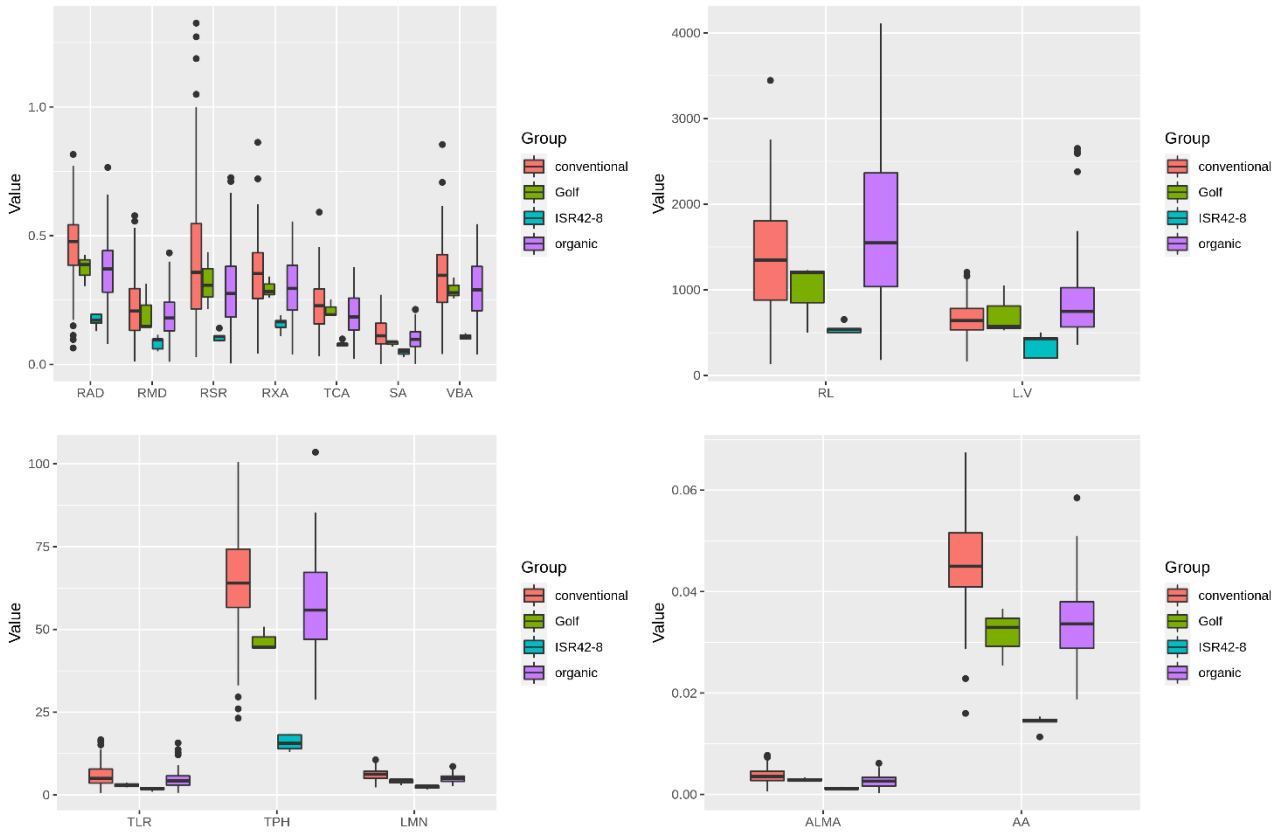


Figure 33: Overview of the significantly variant traits between the organic and conventional groups, based on adjusted data. The color indicates the lines of organic or conventional populations, as well as for the two parental lines Golf and ISR42-8. Boxplots illustrate the observed range of variation and the median value for each group. RAD – average root diameter; RMD – root mass density; RSR - root surface area; RXA – root cross-section area; TCA – total cortical area; SA – stele area; VBA – vascular bundle area; RL – root length; L.V – root to volume ratio; TLR – tiller number; TPH – total plant height; LMN – late metaxylem number; ALMA - Average Late metaxylem area; AA – aerenchyma area.

Compared to the adjusted values, the raw values revealed only ten significantly variant traits between the organic and conventional systems. These included days to emergence, plant height, root length, length per volume ratio, average late metaxylem area, aerenchyma area, root surface area, shoot dry weight, and root system depth.

Table 10: Analysis of variation for all traits observed based on unadjusted data. Positive difference values indicate an increased value in the organic population. Also, the confidence interval is observed from the organic population's perspective. For each trait, the abbreviation, the difference between organic and conventional (positive values are related to higher in organic systems), the upper and lower 95% confidence interval, and the adjusted p-value for the root trait comparison between the organic and conventional populations are illustrated.

		Conf. Interval		
	Difference	**low**	**Up**	**P adjusted**
Days to emergence	-0.15	-0.75	0.44	0.90
Tiller number [1/plant]	-1.38	-2.50	-0.28	0.01
Leaf number [1/plant]	-1.45	-5.00	2.09	0.71
Tiller number [1/plant]	-7.35	-12.25	-2.45	0.00
Root system depth [cm]	-1.51	-6.80	3.77	0.88
Shoot dry weight [g]	0.06	-0.27	0.39	0.97
Root length [mm]	316.53	35.41	597.66	0.02
Projected Area [cm²]	-0.54	-12.94	11.85	0.99
Root surface area [mm²]	-12.95	-50.33	24.42	0.80
Root average diameter [mm]	-0.08	-0.12	-0.03	8.19E-05
Root volume [mm³]	-0.23	-0.72	0.25	0.5816218
Ratio length per volume	228.03	99.65	356.42	4.22E-05
Root dry weight [g]	-0.06	-0.19	0.06	0.50
Specific root length [mm]	-151.95	-951.15	647.25	0.96
Root mass density [g/cm³]	-0.04	-0.07	-0.00	0.05
Root surface area [cm]	-0.14	-0.22	-0.06	3.18E-05
Root cross section [mm²]	-0.05	-0.10	-0.01	0.02
Late metaxylem number	-1.25	-1.75	-0.75	3.69E-09
Average Late metaxylem area [mm]	-0.00	-0.00	-0.00	8.86E-09
Total cortical area [mm²]	-0.03	-0.07	-0.00	0.03
Cortical cell area [mm²]	-0.02	-0.06	0.01	0.18
Aerenchyma area [µm²]	-0.01	-0.01	-0.01	3.12E-14
Stele area [mm]	-0.01	-0.04	-0.00	0.04
Vascular bundle area [mm²]	-0.05	-0.09	-0.00	0.02

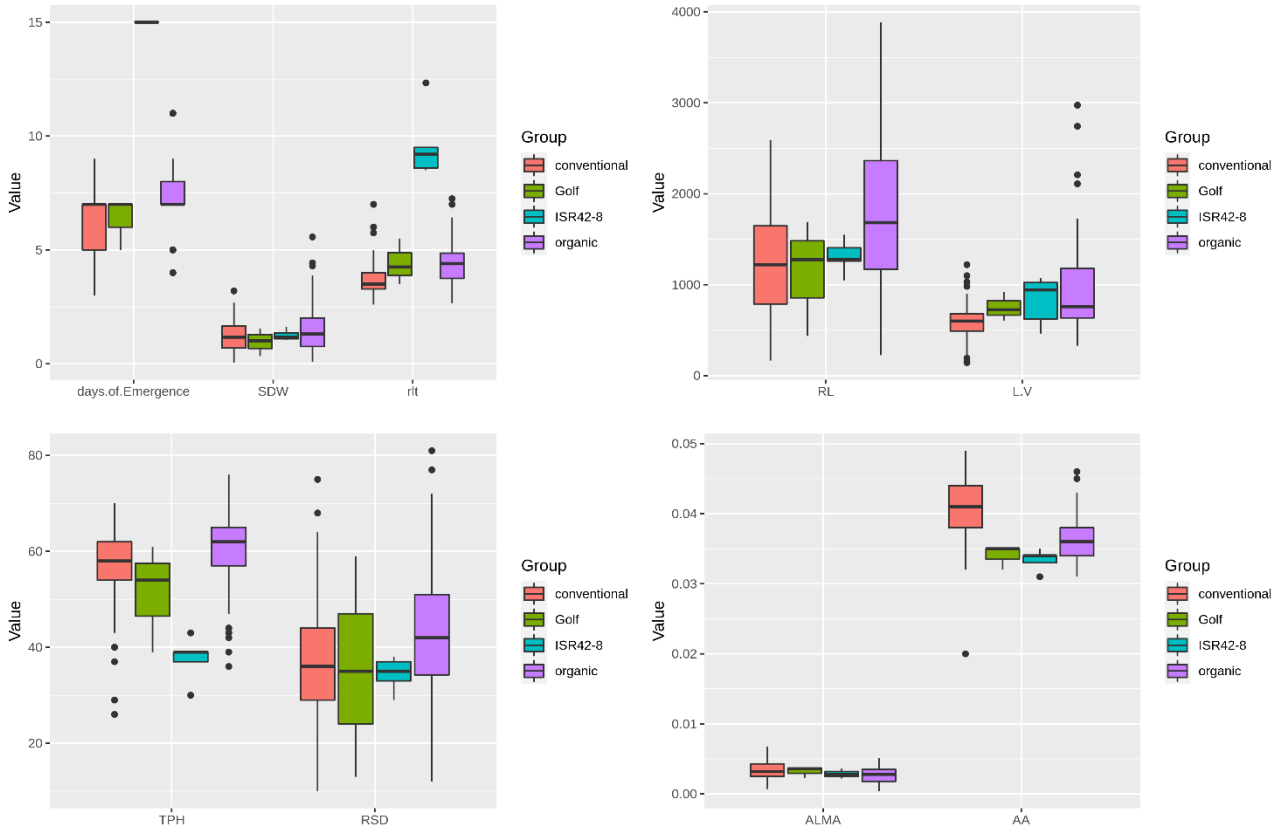


Figure 34: Overview of the significantly variant traits between the organic and conventional groups, based on the unadjusted data. The color indicates the lines of organic or conventional populations, as well as for the two parental lines Golf and ISR42-8. Boxplots illustrate the observed range of variation and the median value for each group. Days.of.Emergence – days from planting to seed germination; SDW – shoot dry weight; rlt – leaf per tiller ratio; RL – root length; L.V – root to volume ratio; TLR – tiller number; TPH – total plant height; RSD – root system depth; ALMA - Average Late metaxylem area; AA – aerenchyma area.

Genotypes of the organic system tend to emerge about one day later than the group of conventional genotypes. Further details can be obtained from table 7.
Concluding these observations, the organic genotypes were on average observed to have more leaves but fewer tillers.

3.3.2 Root phenotype variation field experiment – root analysis

From the measured 20 traits, ten were found to be significantly different between the organic and the conventional group. Among these significant traits, one can find average root diameter, aerenchyma area, cortical call area, late metaxylem area, the length per volume ratio, root angle, root cross-section area, root mass density, tip number, and total cortical area.

Table 11: Analysis of variation for all traits observed on the field-grown genotypes. Positive difference values indicate an increased value in the organic population. Also, the confidence interval is observed from the organic population's perspective. For each trait, the abbreviation, the difference between organic and conventional (positive values are related to higher in organic systems), the upper and lower 95% confidence interval, and the adjusted p-value for the root trait comparison between the organic and conventional populations are illustrated.

			Conf. Interval		
Abbreviation	**Trait**	**Difference**	**low**	**Up**	**P adjusted**
days.of.Emergence	*Days to emergence*	0.951	0.511	1.392	0.000
TLR	*Tiller number [1/plant]*	-0.298	-1.195	0.600	0.827
LN	*Leaf number [1/plant]*	2.473	-0.846	5.792	0.219
TPH	*Tiller number [1/plant]*	3.118	0.552	5.684	0.010
RSD	*Root system depth [cm]*	5.671	1.203	10.138	0.006
SDW	*Shoot dry weight [g]*	0.312	0.011	0.612	0.039
RL	*Root length [mm]*	584.48	326.312	842.647	0.000
PA	Projected Area [cm²]	11.102	-0.069	22.273	0.052
RSA	*Root surface area [mm²]*	24.433	-10.253	59.120	0.265
RAD	*Root average diameter [mm]*	-0.008	-0.045	0.029	0.938
RV	*Root volume [mm³]*	0.216	-0.235	0.666	0.603
L.V	*Ratio length per volume*	341.173	220.528	461.818	0.000
RDW	*Root dry weight [g]*	0.025	-0.094	0.146	0.947
SRL	*Specific root length [mm]*	418.488	-332.368	1169.344	0.474
RMD	*Root mass density [g/cm³]*	-0.003	-0.034	0.029	0.996
RSR	*Root surface area [cm]*	-0.080	-0.150	-0.010	0.019
RXA	*Root cross section [mm²]*	0.005	-0.034	0.045	0.988
LMN	*Late metaxylem number*	-0.258	-0.575	0.059	0.154
ALMA	*Average Late metaxylem area [mm]*	-0.001	-0.001	0.000	0.001
TCA	*Total cortical area [mm²]*	0.004	-0.025	0.032	0.988
CCA	*Cortical cell area [mm²]*	0.008	-0.020	0.036	0.898
AA	*Aerenchyma area [µm²]*	-0.005	-0.006	-0.004	0.000
SA	Stele area [mm]	0.002	-0.013	0.017	0.991
VBA	*Vascular bundle area [mm²]*	0.006	-0.033	0.046	0.978

The highest significant variation was observed for the length per volume ratio (p< 0.0001). While the organic genotype group was classified by an average value of 200.96 and a standard deviation of 53, the conventional genotype group was observed to have a mean of 158.34 and a standard deviation of 77. The root angle indicated an average 4-degree lower angle in the organic group. Analogous to this, the tip number of the roots was 166 tips higher in the organic group, supported by a 95% confidence interval ranging from 19 to 312. For six out of ten traits, the variation within the conventional group was higher than in the organic system. Further details were reported in table 11 and figure 35.
In the group of significant traits, the ISR42-8 value is the highest value for eight of ten traits. Only the root angle and the length per volume ratio were described by equal or reduced values. When the organic group was related to the ISR42-8 genotype, only four of the ten traits were illustrated by a value of the organic group related to the observed values of ISR42-8 (root tip count, late metaxylem number, root mass density, root angle). For the root cross-section area, total cortical area, cortical cell area, aerenchyma area, and root average root diameter, the mean value of the organic group was lower than the average of Golf and ISR42-8 genotypes.

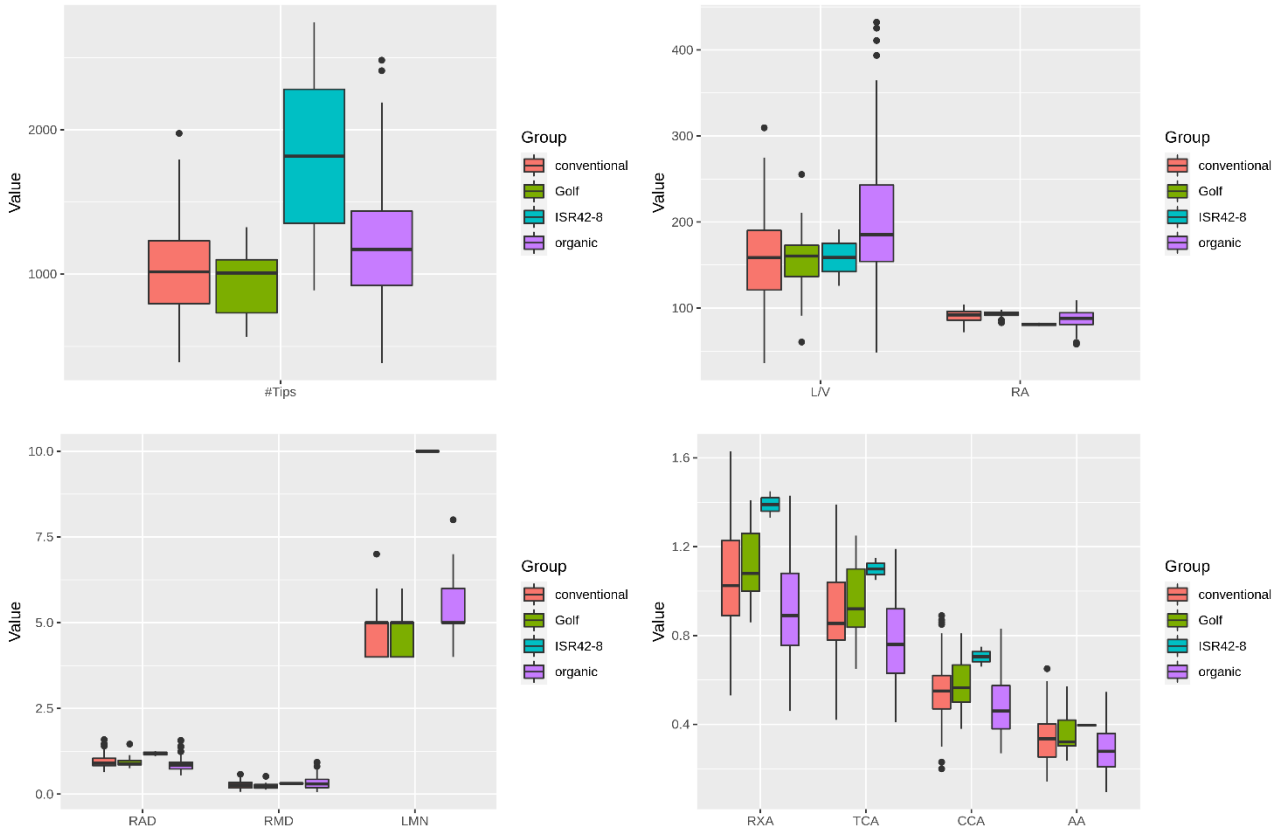


Figure 35: Overview of the significantly variant traits between the organic and conventional group, based on the field-grown genotypes. #Tips – root tip count; L.V – root to volume ratio; RA – root angle; RAD – root average diameter: RMD – root mass density; LMN – late metaxylem number; RXA – root cross-section area; TCA – total cortical area; CCA – cortical cell area; AA – aerenchyma area.

The variation in the populations was compared by a PCA plot. As illustrated in figure 36, the in-group variation was highest in the organic group. Both experiments, hydroponic and field-based, revealed the same outcome. While the variation of the conventional population is marginally broader compared to

the Golf genotype under field conditions, the organic population is represented by a higher variation in comparison to both parental genotypes.

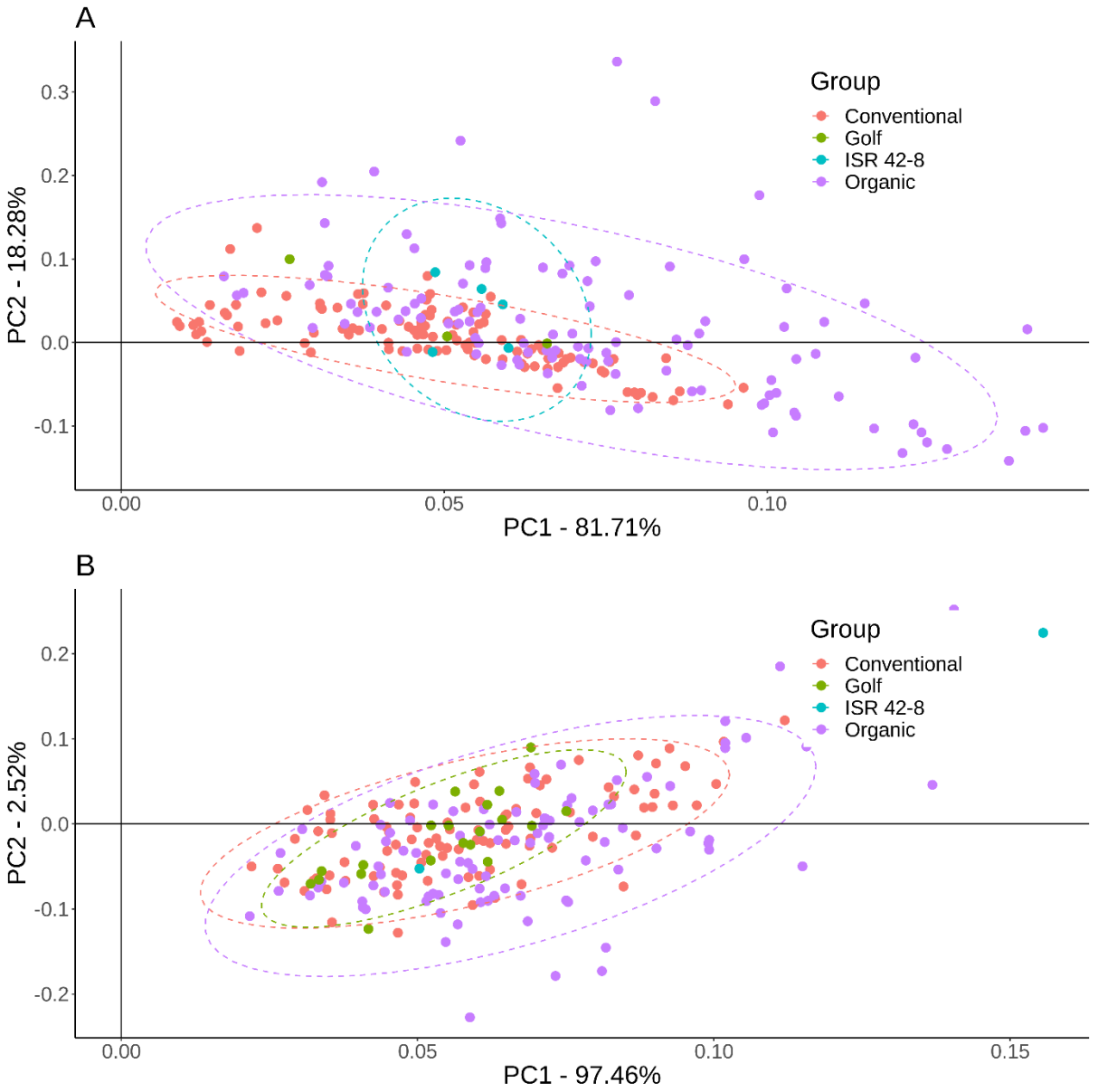


Figure 36: The principal component output of the root morphology assessment in a hydroponic (A) and a field-based environment (B). The first two components are plotted, with the first PC on the x-axis and the second PC on the y-axis. The colors differentiate the conventional (magenta) and organic (purple) systems and the wild donor (ISR 42-8, turquoise) from the cultivar (Golf, green). Each point represents an individual genotype. In (A), 150 genotypes were tested from the organic and conventional systems, while 100 genotypes each were tested in (B). Dashed ellipses are plotted based on the points if the sample size was sufficient. Both experiments illustrate higher genetic variance in the organic system compared to the conventional system.

3.3.3 Field experiment – upper body observations

Phenotyping of upper body characteristics resulted in a homogeneous picture of entire populations. Due to unexpected complications in the sowing process, emergence was more or less mainly related to the varying depth of seed placement. Therefore, both measurements for this trait were excluded.

The correlation matrix for the remaining traits revealed a high correlation of >0.5 among the four flowering score dates. The flowering was negatively correlated with the plant height, while the plant height and the growth type were positively correlated. The health status was slightly correlated with the

measurements of yellow dwarf virus and powdery mildew. The ear-related traits could not be correlated to any other trait (figure 37A).

When all these traits were included in a principal component analysis (PCA), the first two principal components (PC) explained about 45% of the total variation. No cluster variation was identified between the tested groups' Golf, Scarlett, organic and conventional (figure 37B).

On a single trait level, the variation between the organic and conventional genotypes did not show evident variations. While for three of four flowering scores, the population mean deviated by 0.2 days earlier flowering in the organic population, the first flowering score described a 0.3 days later heading of the organic compared to the conventional genotypes.

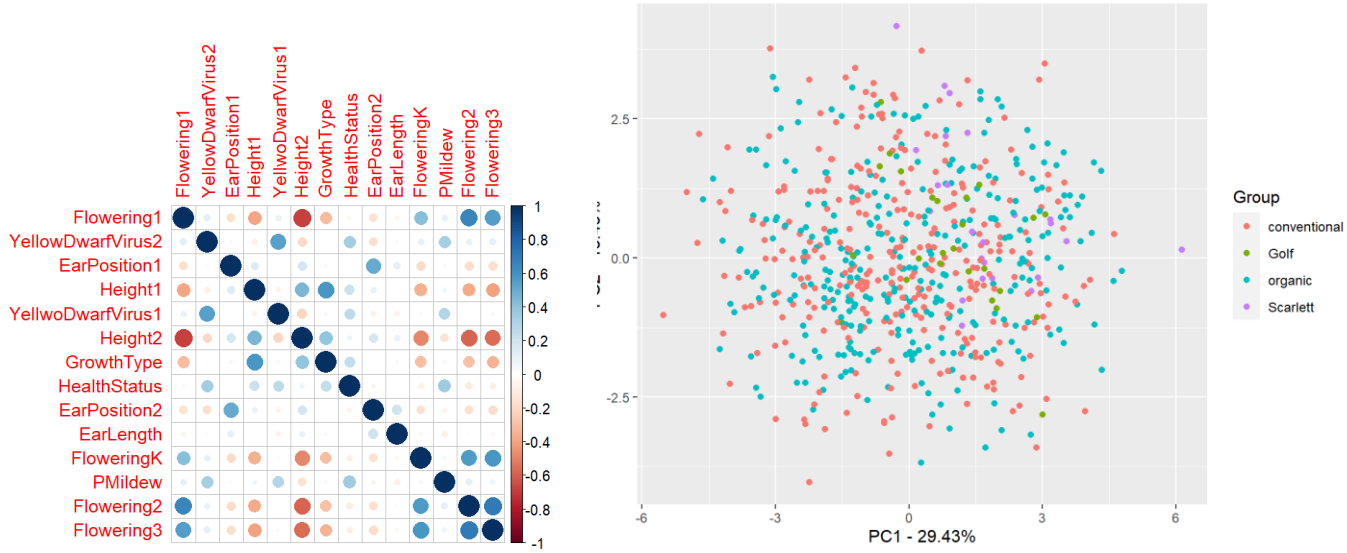


Figure 37: Correlation plot of all measured field traits (left) and the PCA of these traits for the four clusters, illustrated by color. The correlation plot on the left illustrates potential correlations between traits. Exemplarily, high plant height is correlated with early flowering. Each trait is compared to each other trait to highlight relations. The principle component plot on the right was created based on all these above-ground phenotypes and separated by color for the lines and parental genotypes. As illustrated, no variation can be observed between the groups.

The variation within the group was equivalently high between the conventional and organic population, with a tendency of reduced variability in the set of organic genotypes. This observation was consistent over all four flowering scores. The Tukey variation test finally concluded on a p-value of 0.26 that there was no variation between the organic and conventional groups.

The two scorings of the yellow dwarf virus identified a marginal variation between the organic and the conventional group (0.05 higher infection for organic genotypes). The variation within the groups was comparable of approximately one score point each. Overall, the average score was 3.78 and 3.72 at the second date of measurement for the organic and conventional groups, respectively. The first measurement was lower compared to the first measurement by 1.6 scoring points, which might be an indication of a progressing infection. For both scores, the Tukey test revealed no significant variations between the organic and the conventional group.

Powdery mildew was scored once. No variation between the groups could be identified. Both the average value and the variations within the groups were characterized by equal values between the organic and the conventional group.

The health status of the plants gave a summary of all fungal attacks resulting in any kind of advanced tissue degradation. Necrosis was mainly a result of the yellow dwarf virus, powdery mildew, or leaf rust attack. It was measured in percent of leaf tissue infected. The average health status was about 0.5% lower in the organic group. The variation within the groups was almost equal on a level of 23%. Finally, the Tukey test could not reveal any significant variations ($p = 0.54$).

The plant height measurements reported an average decreased plant height of 0.7 cm in the organic group. On the second day of measurement, the reduced plant height is reversed to an average superior plant height of 0.2 cm. The variation within the groups on the second date of measurement is observed to be 4 to 5 cm for both organic and conventional groups. The Tukey test finalized the picture of no significant variations.

The growth type observations were related to the plant height. Groups were established based on the growth habit. Some plants have shown to produce very short internodes, resulting in a highly dwarfed growth type. The other end of the scale was presented by plants with very long internodes. This observation revealed a highly significant variation between the organic and conventional groups ($p < 0.001$). The organic genotypes had 0.4 points reduced internode length, resulting in an average score of 5.5 (conventional 5.9). Besides the reduced score, the organic system was characterized by a twice as high variation within the group compared to the conventional group.

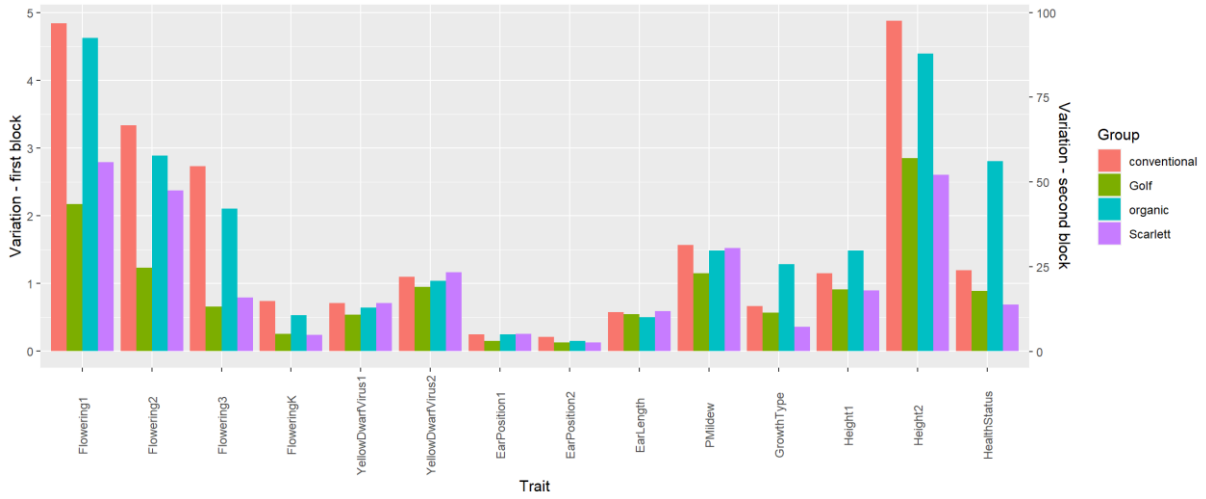


Figure 38: Variation of all yield physiological traits for all four groups/Genotypes, scored on the field. Values give the standard deviation for each group and trait separately. Height and health status are represented by the y-axis on the right, while all other traits are related to the left y-axis.

Finally, the ear position and ear length were measured. While the ear length was measured and reported in metric centimeters, the ear posture was reported in a binomial pattern. The ear length was characterized by an average 0.12 cm shorter ear in the organic system. This information was accompanied by a lower variation compared to the conventional group. The Tukey test did not reveal a significant variation between the groups. Nevertheless, compared to Golf, the ear was characterized by a higher variation (8 to 13 cm in the conventional group, 9-13 cm in the organic group, Golf 9.5 to 12 cm). The ear posture was assessed on two dates. Both highlight that the ears tend to be lowered more in the organic group (0.07, 0.13 higher score for the organic group). While the first measurement could not identify significant variations, the second measurement highlighted a significant variation ($p < 0.05$, based on a binomial generalized linear model). A more detailed classification of the traits was reported in table 12. The variation within the groups was almost identical for both the organic and conventional groups (figure 38).

Based on these observations, the most interesting traits, ear length, growth type, and ear posture, were plotted for the organic, conventional group, Golf, and Scarlett. In figure 39, the increased variation of the groups compared to the single genotypes is visualized. Furthermore, the ear posture shows clearly an advanced lodging of the ears in the organic group before any other group or genotype.

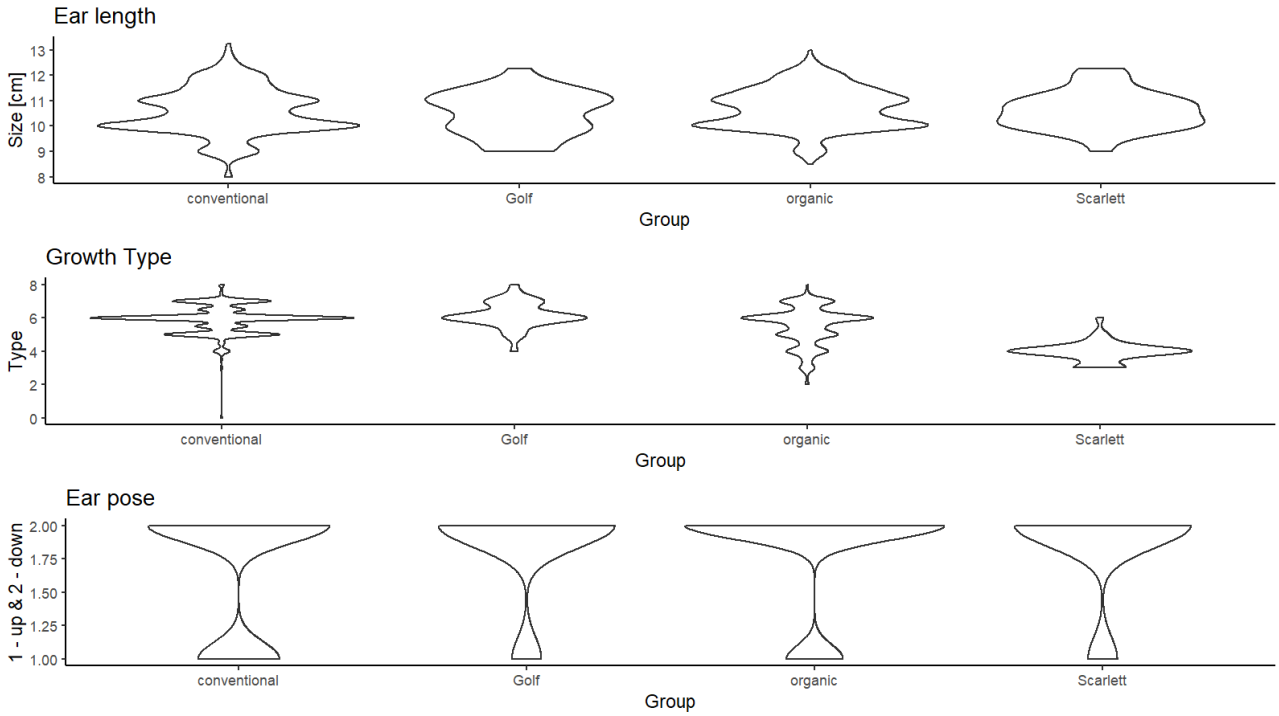


Figure 39: Illustration of the three most interesting traits observed, which are characterized by a significant difference between the organic and conventional group or a high variation within the group. The ear length, growth type, and ear posture are presented, separated for each genotype.

Table 12: Statistics for the scored traits. If the difference is negative, the organic population mean was smaller than the mean of the conventional population. Furthermore, the upper and lower 95% confidence interval is illustrated. The last column gives the p-value between the organic and conventional groups. Two traits have been identified as significant, the ear position and the growth type. For the ear position, a binomial distribution was assumed.

	Difference	lwr	upr	p adjusted
Ear Length	0.106	-0.098	0.311	0.5342
Ear Position	-0.069	-0.152	0.012	0.029
Flowering 1	0.274	-0.113	0.662	0.2633
Flowering 2	-0.163	-0.453	0.126	0.4642
Flowering 3	-0.229	-0.524	0.064	0.1843
Growth Type	-0.390	-0.552	-0.229	4.01E-09
Health Status	-0.150	-1.348	1.047	0.9882
Height 1	-0.711	-1.551	0.129	0.1300
Height 2	0.261	-1.306	1.828	0.9735
Powdery Mildew	0.119	-0.082	0.319	0.4226
Yellow Dwarf Virus 2	0.048	-0.123	0.220	0.8842
Yellwo Dwarf Virus 1	0.059	-0.076	0.195	0.6721

The measurement with the polypen revealed 25 indices. Nor the PCA, neither Tukey test nor the variation comparison identified differences between the organic and the conventional group for any of the 25 indices (figure 40).

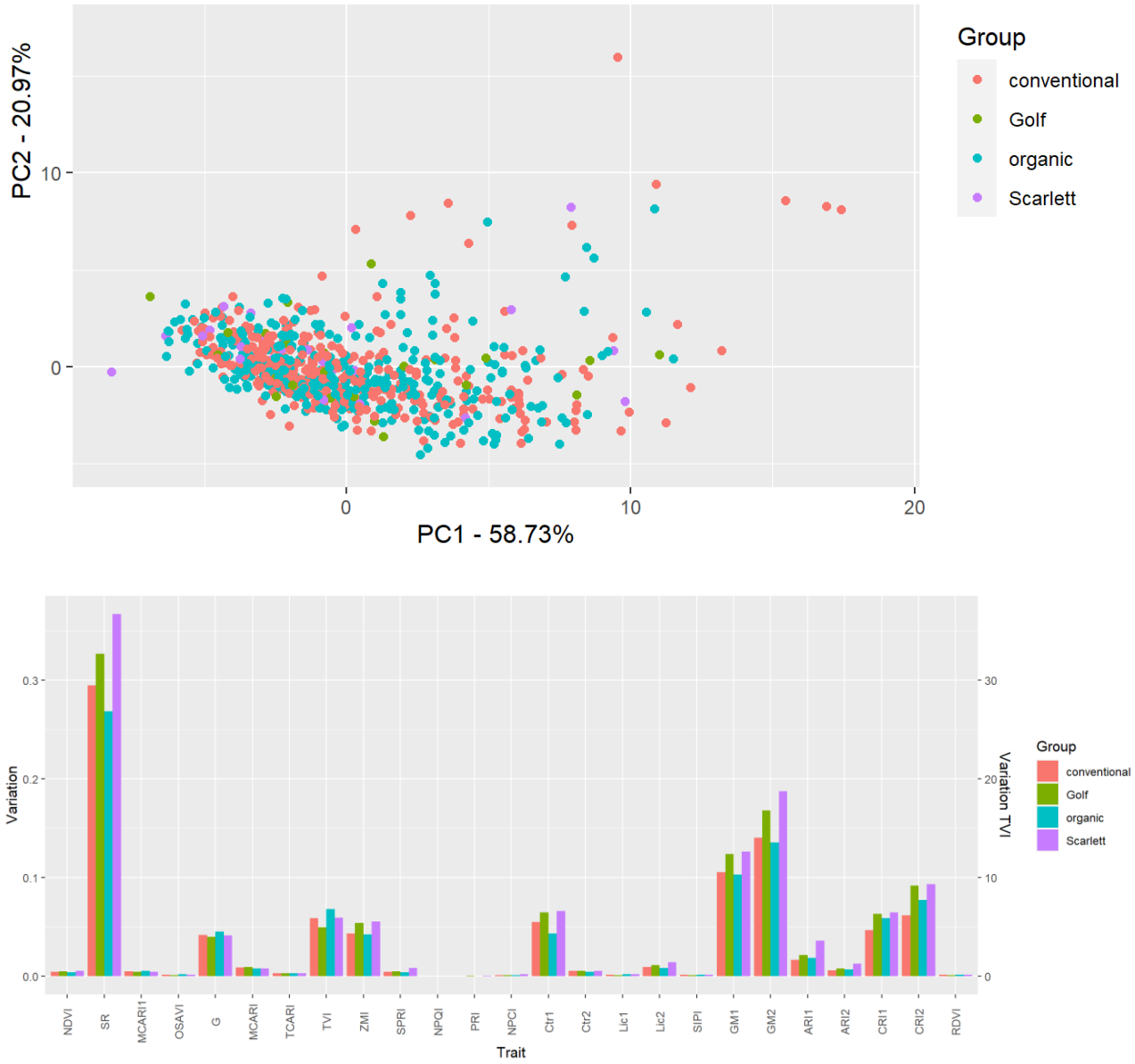


Figure 40: Polypen leaf measurements. A - PCA of all polypen generated traits, separated by color for the populations and parental genotypes; B – the variation in the genotype classes for the observed spectrum measured.

Finally, the scores obtained by SPAP chlorophyll measurement were classified by no significant variations between the groups. Nevertheless, the organic group had a 15% higher average score compared to the conventional group. Both conventional and organic systems were characterized by substantially higher average values in comparison to Golf and Scarlett (figure 16).

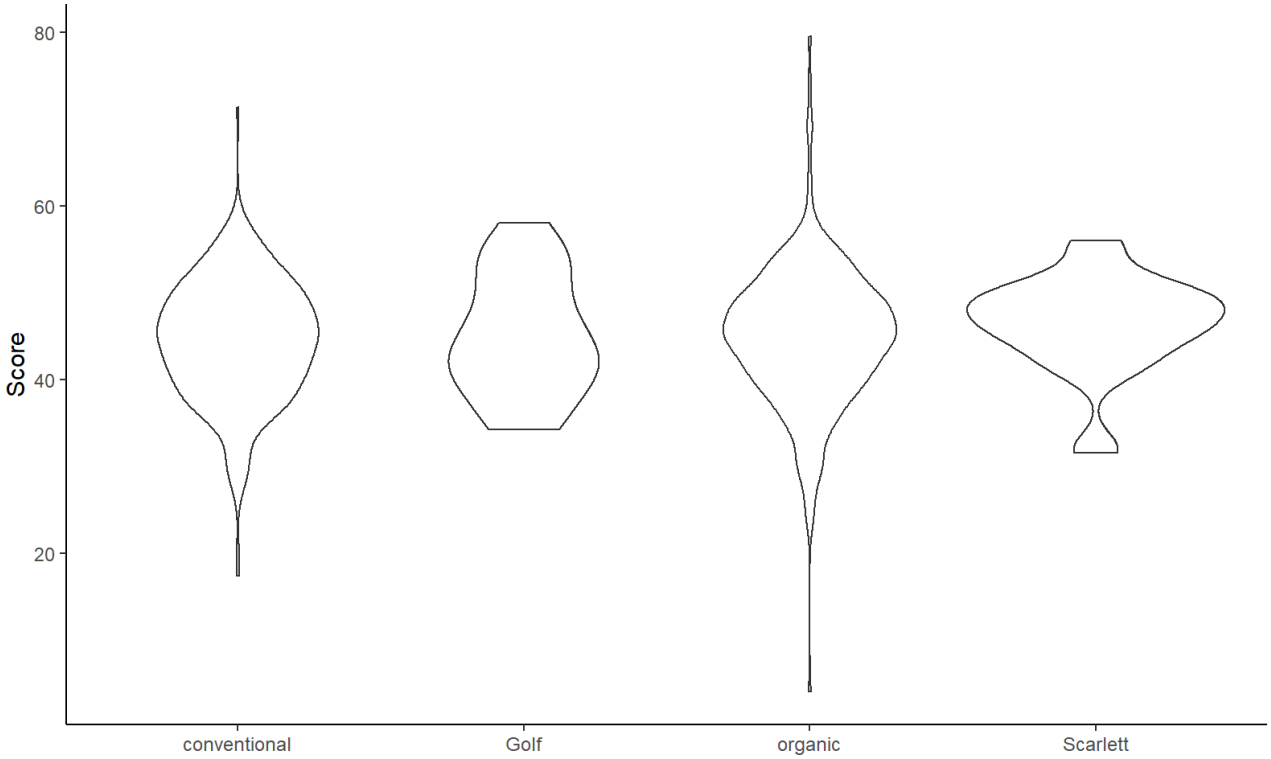


Figure 41: Box plot of SPAD chlorophyll content measurements for all four groups.

3.4 Climate, weather, and evapotranspiration

The climate and annual weather events are expected to have a major influence on the allele frequency variation over the duration of the experiment. It is helpful to know as much as possible about the climate, to understand particular adaptations or changes from one year to another. From available measurements of the weather station, the actual evapotranspiration for the growth period of spring barley has been calculated. The days under water stress can be estimated when the evapotranspiration is extended by a soil water balance. The correlation of the measured traits, shown in figure 42, highlights the fact that the correlation of the evapotranspiration ET_0 was highest to the radiation, followed by the minimum air humidity and the maximum air temperature.

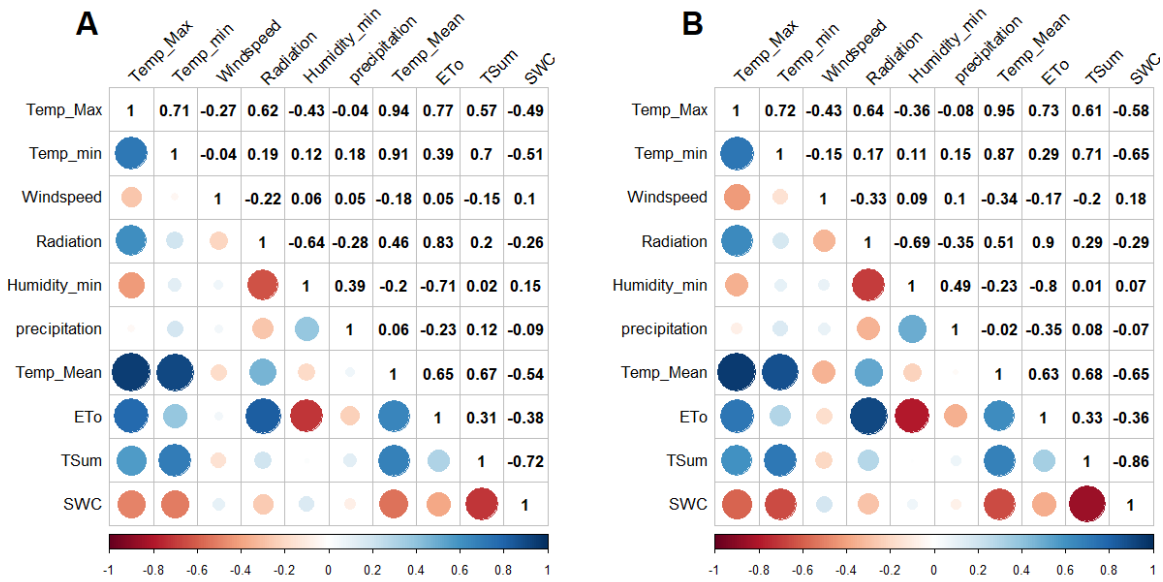


Figure 42: Correlation plot of the most important measurements made by the weather station over all the years. ***A*** *– the measurements from the weather station at Campus Klein-Altendorf from 2010 to 2019;* ***B*** *- the identical measurements from Campus Dikopshof from 1996 to 2009. The correlation of the calculated average daily values of maximum, minimum and mean temperature, as well as windspeed, minimal humidity, the sum of precipitation, radiation. Additionally, the correlation to the calculated base evapotranspiration* $\boldsymbol{ET_0}$*, the accumulated temperature sum* ***Tsum,*** *and the soil water capacity* ***SWC*** *are shown.*

A linear mixed model was performed to get a better estimate for the dependencies of the evapotranspiration. The regression of the actual evapotranspiration ET_A against the year and growth stage as a fixed effect, minimum humidity, maximum temperature, radiation, soil water capacity, and drought event as random effects was estimated in this lmm model. The overall *REML* score is 8,020.7, where the random effects had an explanatory power of ~72% together. The most variation was explained by the humidity (0.367), followed by the soil water content (0.260) and the radiation (0.205). The temperature had the smallest variance with 0.08, leaving 0.36 residuals unexplained. On the fixed effect side, all growth stages had a highly significant variation from each other with a p-value < 0.001. The same holds true for the occurrence of drought stress. The year factors are much more differentiated. In relation to the year 1996, the years 2004, 2009, 2017, and 2019 were highly significant on a p-level below 0.001. The years 2005, 2006, and 2018 are significant on a level of 0.01. The remaining years are above this threshold.

Figure 45 illustrates the deviation of the annual mean or sum in the experimental period for multiple traits. The period of the experiment had to be split into two subperiods, as the weather stations did measure slightly differently. Therefore, the means and sums of the period from 1996 to 2009 were calculated separately from the period of 2011 to 2019. Besides the temperature, the humidity, radiation, precipitation, as well as the calculated actual evapotranspiration and the days of drought were plotted. Over all the years, the climate was observed to be highly variable. The average evapotranspiration in the vegetation period of barley was 312 – 334 millimeters. The maximum daily temperature mean was about one degree Celsius lower in the first period compared to the second, where the value was 20.36 °C as mean over the growth period. The minimum temperature indicated the other way. 9.44°C in the first period could be compared to 8.97°C as the daily mean over all days in the growing period. The average day temperature does not deviate too much and is about 14.5°C high. The sum of radiation in the growth period differs. In the primary period, on average, 2,209 W/m² were measured, while 1,726 W/m² were measured in the period from 2011 onwards. Besides the radiation, also the minimum humidity was variant. While 42.5% was calculated for the first period, the second period had a seven

percent higher value. The mean precipitation for the years varied by 30 mm. The higher rain fall could be observed at Campus Dikopshof. Lastly, the average count of drought days over both periods was 14 days. Two weeks of drought seem like a lot, but the variance for this is very high. While there was no drought event in ten out of all years, ten years have shown to lie above the mean of 14 days. This value does not give an idea about the extent of drought days with little evapotranspiration reduction were considered equal to days with severe drought. The total sum of the variation of potential to actual evaporation is plotted at the bottom right. This can serve as a more informed idea of the drought extend of each year. A more detailed resolution for the extend of drought is illustrated in figure 44. For each year, the daily sum of evapotranspiration is plotted. The time point of drought stress can be obtained by figure 43. Three curves are presented, from which the blue curve is related to the evapotranspiration over a reference grassland ET_0. The yellow curve presents the actual evapotranspiration of barley, adapted to the growth stages and soil coverage. Lastly, the red curve is the actual evapotranspiration, extended by a soil water model. If the field capacity was lower than 30%, the plant was drought-stressed and increased the stomatal conductance successively, the lower the field capacity declines. Several years were present where the yellow and red lines were separated from each other. The years that stand out most were 1996, 1999, 2000, 2006, 2011, 2015, 2017, 2018, and 2019. While for the first of these years, the drought was characterized by a moderate variation between ET_A and ET_C, the year 2011 was the first year where a variation peak of 4 mm per day was observed. This was equivalent to the fact that on two days in this year, more than 6 millimeters of water could be transpired, but less than 2 millimeters actually were. This had not happened before and was a severe drought event. Less than 30% of the potential transpiration can be reported for more than two days in 2011. 2017, 2018, but especially 2019 have this pattern in common as well. 2019 has to be highlighted in this regard, as the deviation of the actual to potential transpiration increased over the half period of July to a maximum of 5 millimeters. Over all 23 years of observations, the potential evapotranspiration exceeded a value of 8 millimeters per day. One day was in 2017, the other in 2019. Out of the last five years, four could be observed with severe drought conditions that hold the potential to reduce the growth potential of the crop. In comparison to the whole period, the climate has shifted towards dry and hot climates. On the other hand, years without a shortage of water could be observed. In the recent past, 2016 was a moist year, but also 1998, 2001, 2002, 2005, 2007, 2008, 2009, 2012, and 2013 are characterized by no limitations of water availability. These years are potentially interesting, as a low water shortage usually goes along with higher precipitation in the productive period. The precipitation events, as well as the average daily temperature and the evapotranspiration ET_0 are shown in figure 46. These values are illustrated for each day of the cropping period. Many precipitation events in the tillering to anthesis stage make infection by fungi more likely. 1998, 2007, and 2016 were years considered to be optimal years for above-average infection levels, as the time before and during anthesis was classified by many precipitation events. To summarize, the period can be divided into two to three different groups. The grouping is illustrated in figure 44, where the deviation of the actual and potential evapotranspiration was shown. Years associated with the letter *a* were regarded as humid and moist years, with little to no drought stress in the vegetation period. Contrasting to this, years with more or less distinct drought stress were characterized by *d, e,* and *f* letters. Together, twelve years were associated with category *a*, while eight years were associated with the drought stress letters *d, e,* and *f.* A third, intermediate group, could be described for the years 2003, 2006, and 2015.

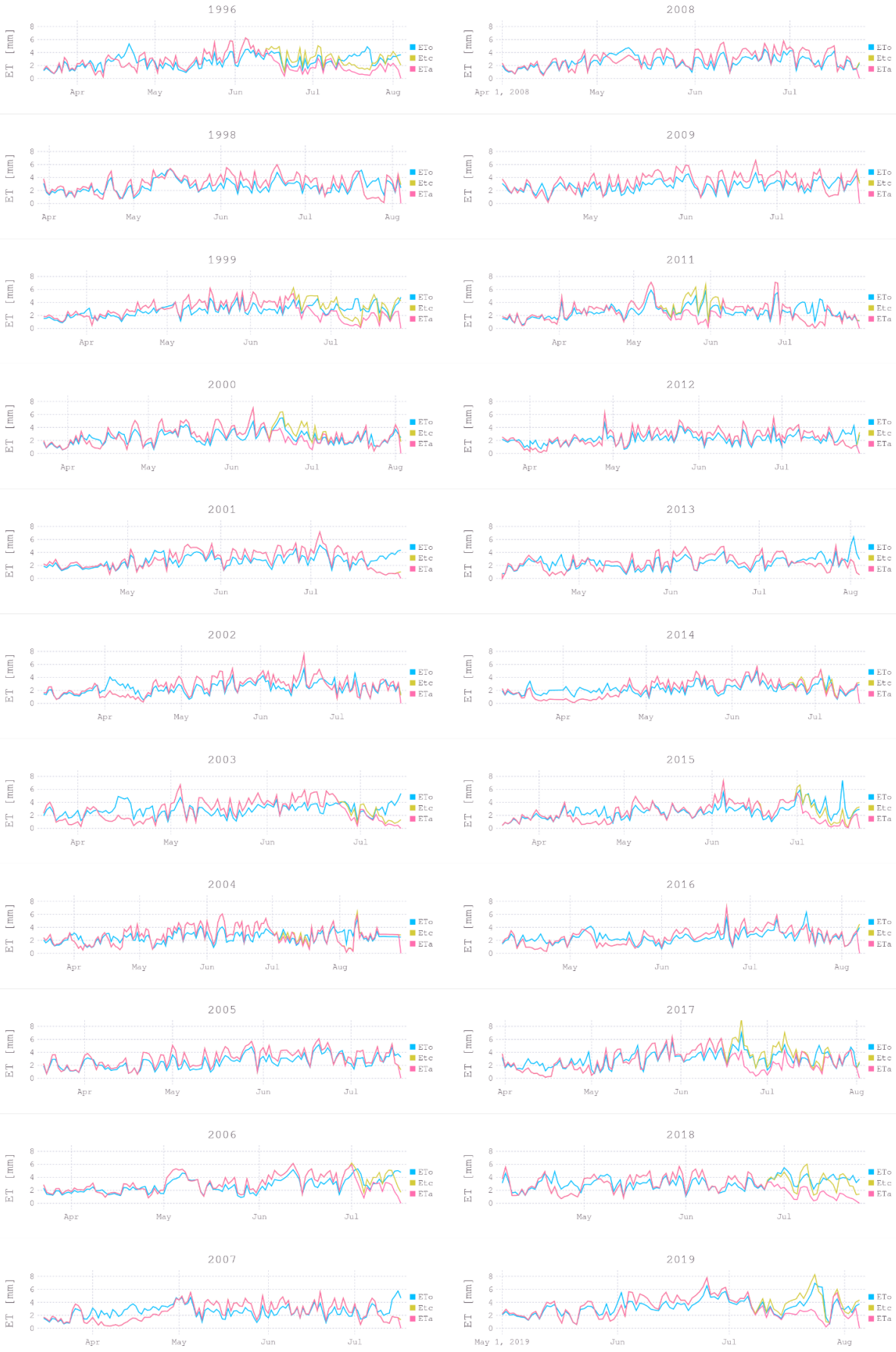

1996
2008
1998
2009
1999
2011
2000
2012
2001
2013
2002
2014
2003
2015
2004
2016
2005
2017
2006
2018
2007
2019
ET [mm]
ETo
Etc
ETa
Apr
May
Jun
Jul
Aug
Apr 1, 2008
May 1, 2019

Figure 43: The evapotranspiration in the vegetation period for each day. Evapotranspiration is the sum of water lost from soil to the atmosphere in millimeters. The curves are plotted, showing three levels of evapotranspiration. The blue line ***ET_0*** *is equivalent to the evapotranspiration over the grassland. The* ***yellow*** *line is the actual evapotranspiration, which is corrected to spring barley growth stages and development. The* ***red*** *line is the actual evapotranspiration plus a soil water balance model. If the soil water is depleted to less than 30% field capacity, the plant starts to save water. The yellow and red lines run identically, as long as there is no shortage of water. As soon as the yellow line becomes visible, the plant suffers drought stress. The variation of the blue and red lines correlates with the developmental stages' emergence, crop development, anthesis, ripening, and maturity. The figure illustrates the crop development and at which stage the plant might have suffered water shortage.*

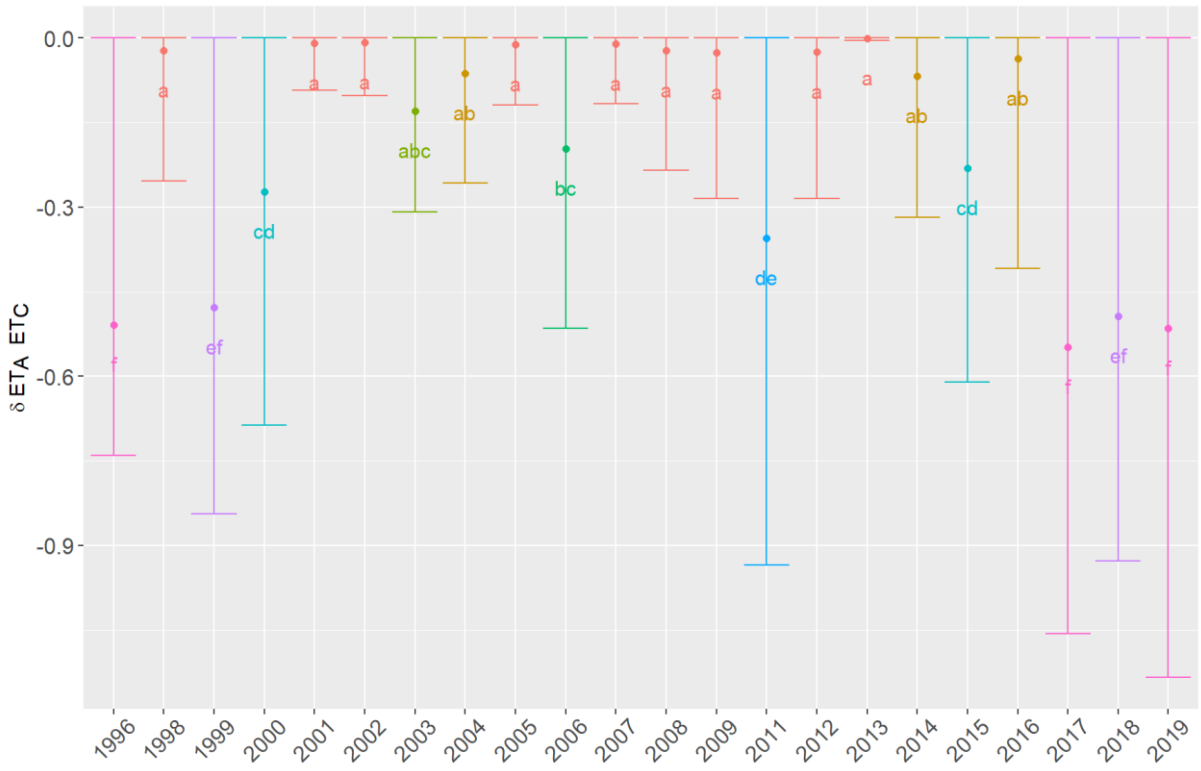


Figure 44: Grouping of delta ET_A to ET_C. Delta is equivalent to the transpiration reduction of the active biomass due to water stress. The grouping has been performed by HSD.test of agricolae. Groups with low to little water limitation effects in the vegetation period are characterized by a little deviation from 0, while the years with a strong impact of a drought are illustrated by a high deviation of > 0.5 reductions of the actual evapotranspiration.

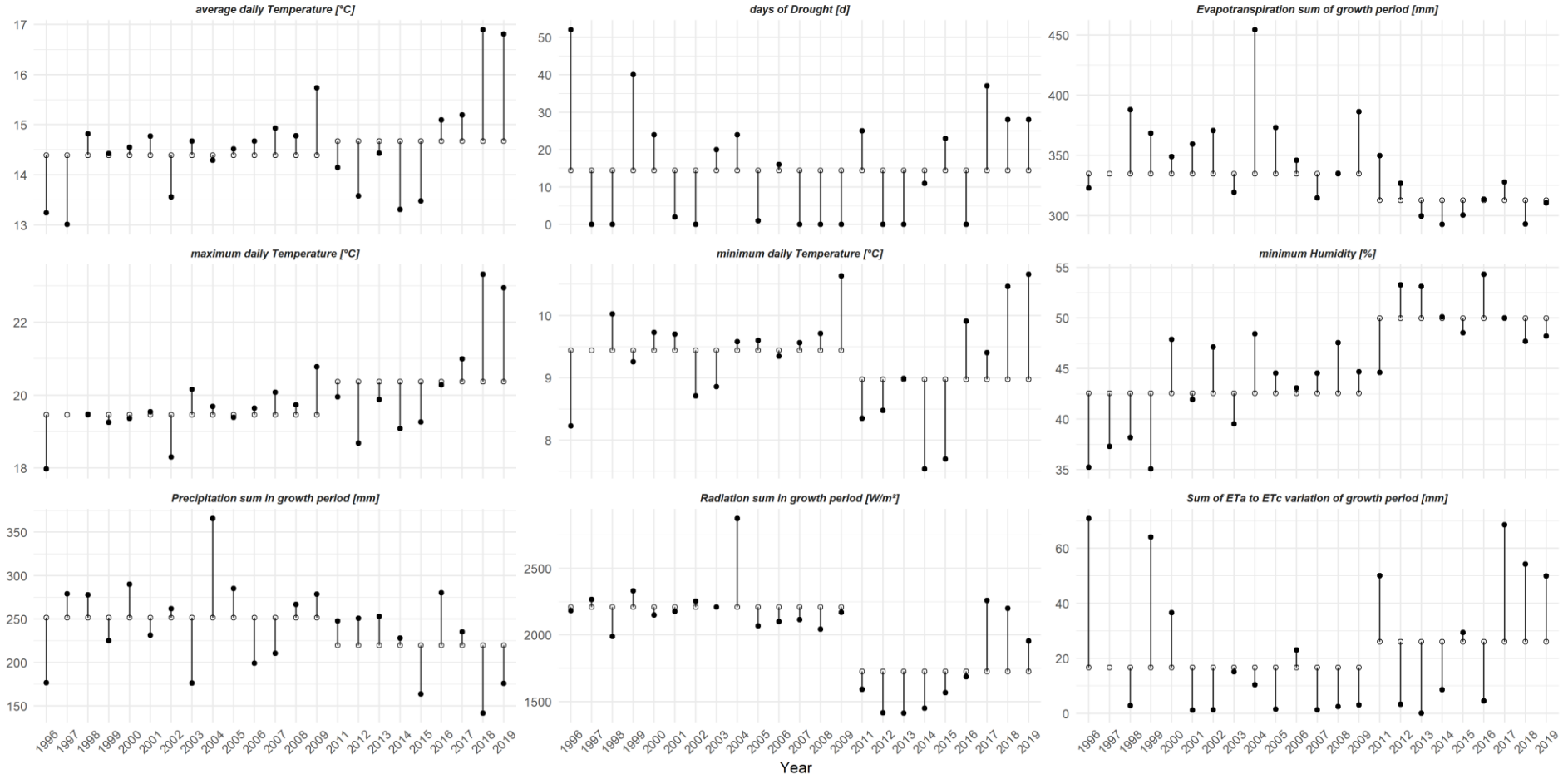


Figure 45: Average or sum values for different traits over the years. Each year has one value (black) which is compared to the mean overall years (white point). The mean over the years has been calculated separately for the two weather stations, as these are measured a bit differently. The line between the black and the white point indicates the difference between the mean over the years and the actual average value for the particular year. Besides the average, maximum, and minimum daily temperature, the evaporation sum, the estimated days under drought stress, humidity, radiation, the variance of potential to actual evapotranspiration, and precipitation are plotted.

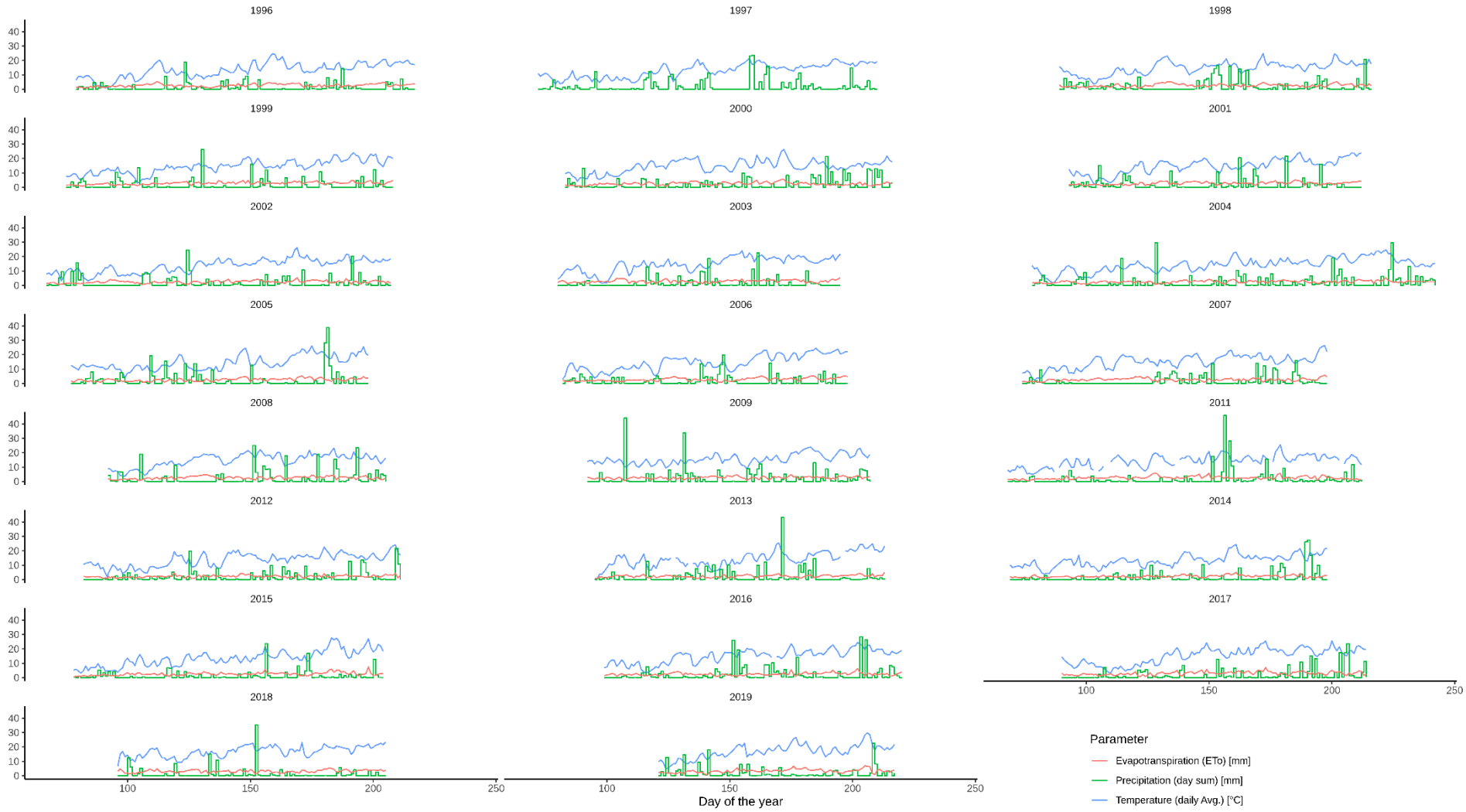


Figure 46: The daily average temperature, the daily precipitation sum as well as the evapotranspiration ***ET_0*** *are presented for each year and each day of the vegetation period from sowing to harvesting day. When combining this information, an estimation of the growth conditions in the year, as well as the comparison between the years, can be made.*

3.5 Observed evolution of populations enforced by farming environment

Based on the observations made by comparing the accuracy of different sequencing methods, whole-genome resequencing has been used to sequence five generations and the parents of the population. The selected generations were BC_2F_3, BC_2F_{12}, BC_2F_{16}, BC_2F_{22}, and BC_2F_{23} for organic and conventional farming systems, respectively. Initially, more than 20 million variant base pairs were found between the parents Golf and ISR42-8. Due to uncertainties in regard to the quality and reliability of those variant detections, only those variants were kept that could be found in the ensemble SNP variant deposit (Cunningham et al., 2019). This action resulted in a decrease of variant positions down to 3,991,259 – or, on average, every 1,500 base pair a variant base. The ratio of alternative to reference base call ranges from 0.4 to 0.45. This ratio was in the expected range; therefore, no correction of the allele frequency had been applied for the data sets. SNPs were annotated to 34,344 gene-based haplotypes and 5,948 marker-based haplotypes.

3.5.1 Distribution of allele frequency calls in haplotypes

The highlighted selection of statistical models and distributions to represent the allele frequencies has not been described before and had to be adjusted to the observed distributions. Therefore, the SNP-based allele frequency of haplotypes was collected, and the variation over all defined haplotypes was observed. This revealed four major groups the allele frequency patterns of the haplotypes could be clustered to. The first group is the “blackbox”, a group of haplotypes described by a low SNP count (lower than 15, figure 47A). Based on this value, no conclusions of the distribution could be made. The second group is described by a negative binomial distribution, as illustrated in figure 47B. The third group, the biggest group, is presented by a divided distribution. While one set is described by a negative binomial distribution, the other set is characterized by zero values. These zero values are expected to occur more often at the applied sequencing level, especially with the double backcrossing. The lower the real allele frequency of this locus is, the more often this case will be true (figure 47C). Lastly, for loci with allele frequencies lower than 5%, the distribution of the SNP allele frequency was majorly based on the zero-inflated distribution (figure 47D). The variation of distributions made it necessary to implement an algorithm that adjusts to the distribution of the haplotype and sets the appropriate regression model to calculate the statistics. This was done in accordance with the described model in the material and methods section.

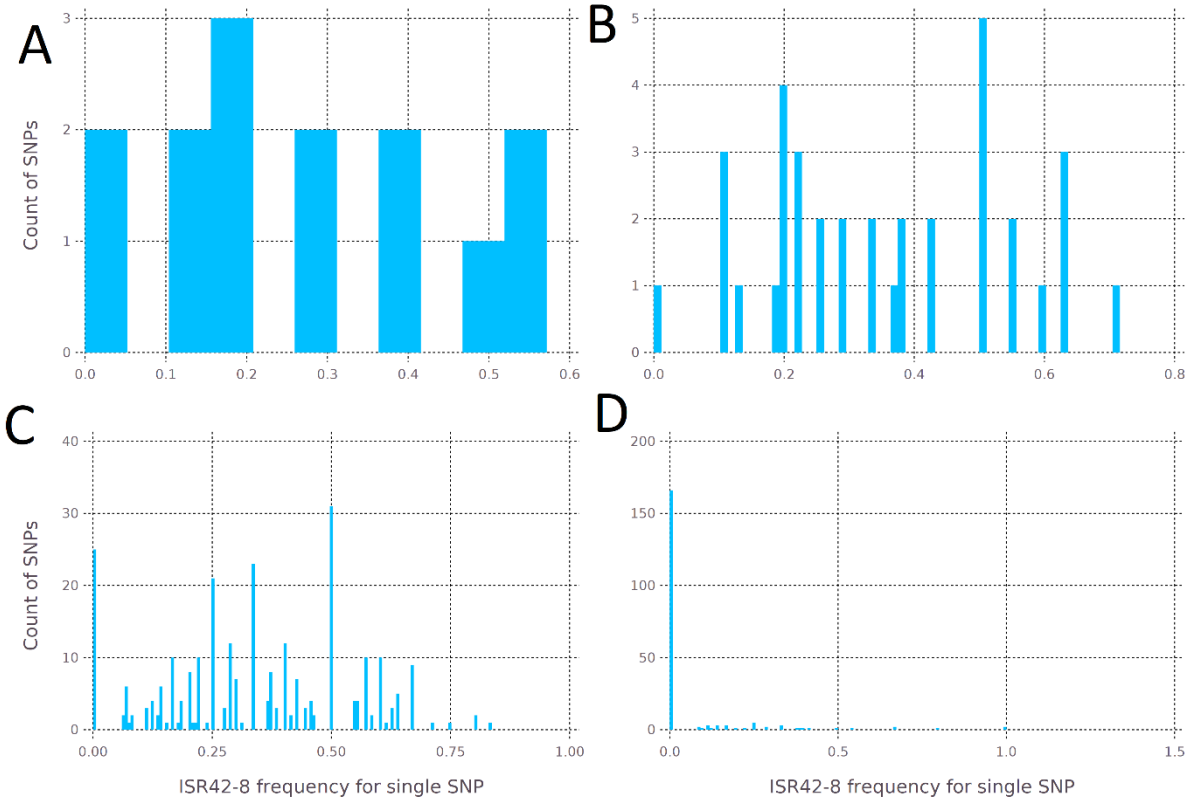


Figure 47: Comparison of the allele frequency distribution for three representative haplotypes, based on all SNP forming a haplotype.

3.5.2 Molecular genetic variation on genome-wide scale

The genomic variation was estimated by a genetic diversity analysis of the populations to dissect the undergone agro-environmental selection. Principal-component analysis (PCA) showed a sizeable genetic differentiation between the offspring generations due to microevolution in the different agro-ecological environments. The first principal component (PC) describes the highly significant ($p < 0.001$) distinction of the environments (organic to conventional), while the second PC separates the effect of the years (Figure 48). Furthermore, the first PC explains more than 51% of the variation, while the second PC only explains 7.74%. Based on this observation, the agro-environmental farming system has had the most dominant impact on the development and adaptation of the differentially farmed populations. As already indicated in figure 48A, the molecular genetic variation between the two samples of the same generation in the organic environment tended to be higher compared to the conventional population. This can be observed for the evolved populations of F_{12}, $F_{16,}$ and F_{23}.

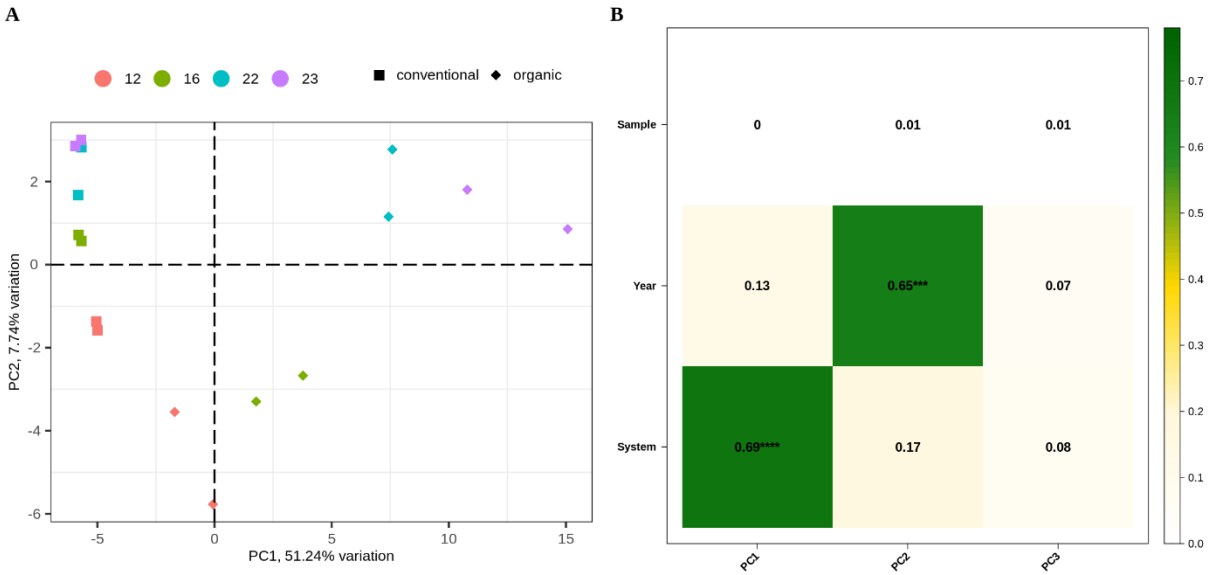


Figure 88: ***Principal component analysis of the tested populations*** *from the 12th, 16th, 22nd, and 23rd generations with both samples per population. The generations are presented in different colors, while the shape is split for the systems (**A**). The explanatory effect of the principle components is illustrated by **B** – the first PC explains the system impact, while the second PC highlights the generation/year effect. The samples do not have any effect on the clustering of the samples. **A** – the majority of variation is explained by the first PC (51.24%), while the second PC explains 7.74%.*

The variation between the two samples of the same generation and agro-system was tested to dissect this variation in more detail and test the statistical significance of this observation. The calculated variations on haplotype level (illustrated in table 13) underline the observations made in the PCA. Besides the 12th generation of the conventional environment, the 12th, 16th, and 23rd generation in the organic environment showed highly significant variations between the two samples over the 34,344 gene-based haplotypes. When the marker-based haplotypes were considered, the variation between the samples becomes less significant. For the 5,948 marker-based haplotypes, only the 12th generation in both environments and the 23rd generation in the organic environment indicate a significant variation between the samples ($p < 0.001$). On the contig-based haplotype level, no variation between the samples of the same generation and environment could be identified. Concluding this statistic, the haplotype level had an effect on the comparison of replicates.

Table 13: Variation between samples of the same population and generation. Two samples per combination were sequenced, and the variation between these was calculated by a zero-inflated negative binomial distribution in a generalized linear model, based on the contig-based, marker-based, and gene-based haplotype frequencies. Values smaller than 0.01 are assumed to highlight a distinct variation between the two tested samples and are highlighted by orange color.

		Organic System		**Conventional System**	
		negative binomial p-value	*zero-inflated p-value*	*negative binomial p-value*	*zero-inflated p-value*
Contig based haplotypes	*F3*	0.796	0.818	0.796	0.818
	F12	0.101	0.811	0.191	0.743
	F16	0.309	0.068	0.995	0.700
	F22	0.165	0.050	0.352	0.034
	F23	0.021	0.363	0.786	0.929
Marker based haplotypes	*F3*	0.442	0.553	0.442	0.553
	F12	0.000	0.991	0.000	0.950
	F16	0.126	0.110	0.102	0.485
	F22	0.561	0.257	0.436	0.695
	F23	5.56E-11	0.000	0.219	0.825
Gene Based haplotypes	*F3*	0.706	0.821	0.706	0.822
	F12	1.79E-29	2.23E-15	5.56E-11	0.001
	F16	0.003	5.82E-10	0.159	0.056
	F22	0.078	0.251	0.675	7.12E-09
	F23	4.09E-35	2.22E-08	0.845	2.00E-06

Including the initial 3rd generation in the PCA figure, a clear and distinct evolution away from this population was observed for organic and conventional populations. Based on the first PC, which explains more than 42% of the variation, the conventionally farmed population evolved until the 12th generation but changed little after that. In contrast, the generations of organic populations could clearly be distinguished from each other. Both environments had the same development until the 12th generation in common. The following generations differed much more distinctly between the environments. With each studied subsequent generation, the genetic distance between the organic and the conventional populations increased. Generally, the most significant variation was observed between the 3rd and 12th generations of both environments. The two samples of the 3rd generation were far different from all other populations tested, indicating a distinct development away from the original composition of the populations. Besides this system effect, the first principle component was highly significant for the yearly effect (Figure 48B). The sample did not show any effect in the PCA, indicating that samples of the same generation and environment were not relevant for any observed evolutionally effect. The second PC mainly differentiates the organic and the conventional system from each other. Concluding, both systems become more and more distinct with each subsequent generation.

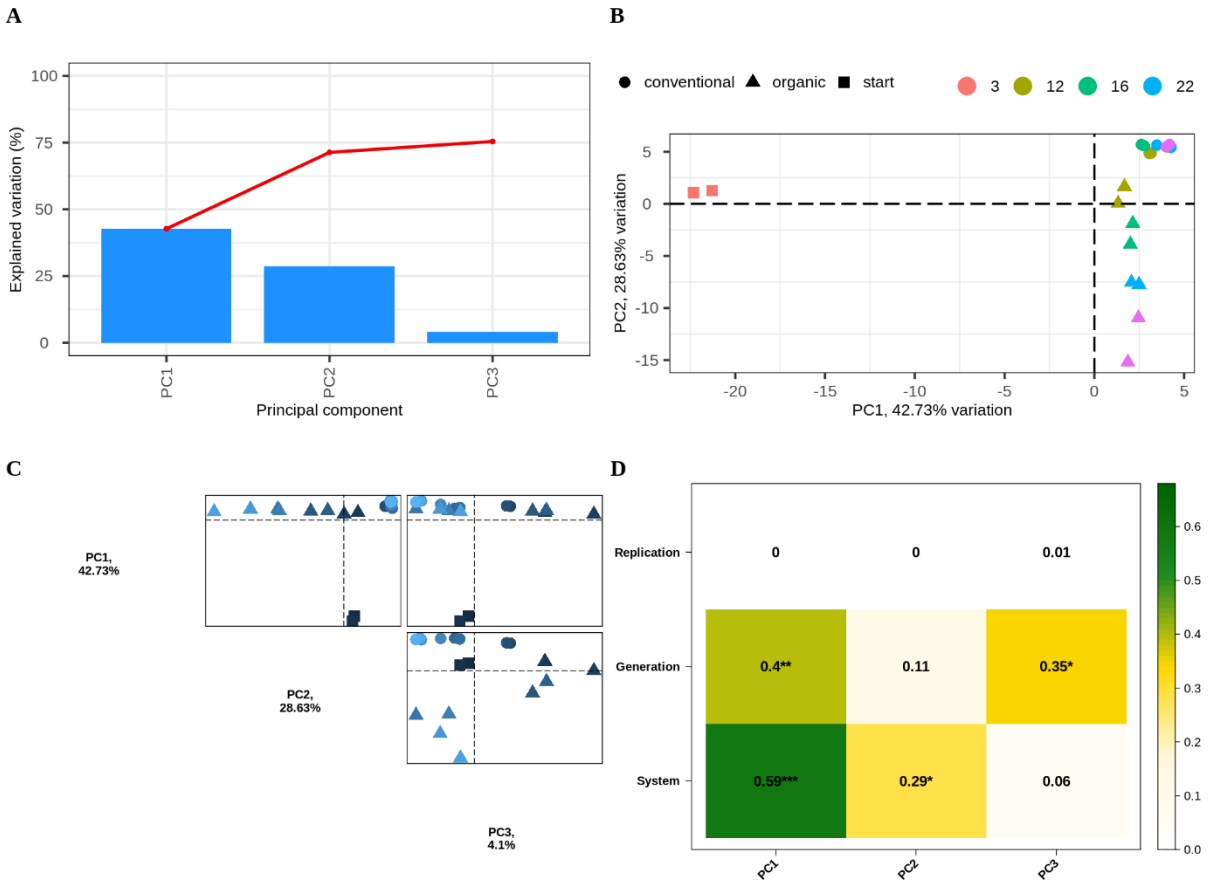


Figure 49: Principal component clustering of all tested generations (including F3). The principal component analysis of the 3rd, 12th, 16th, 22nd, and 23rd generations with both samples per population (replication), split by shapes for the systems (B). The explanatory effect of the principal components is illustrated by D – the first PC explains both the year and the system impact, while the second PC highlights the system's (environmental) effect, and the third PC depicts the year effect. The replications do not have any effect on the clustering of the samples. B – the majority of variation is explained by the first PC (42.73%), while the second PC explains 28.63% and the third PC 4.1%. The offspring generations are clustered in three major groups. C – The first three principal components plotted against each other. A – The explanation power of the first three components.

This was underlined by the increasing population differentiation fixation index (F_{ST}) and the genetic distance assessment, illustrated by the dendrogram in figure 49. While the F_{ST} revealed a value of 0.012 for the 12th generation, it successively increased to F_{ST}=0.043 in the 16th, F_{ST}=0.052 in the 22nd, and F_{ST}=0.113 in the 23rd generation. Five clusters can be reported based on the genetic distance calculation. The first and most distinct cluster included both BC_2F_3 samples. The second and third clusters were the two subgroups of the organic populations. In comparison, the 12th and 16th generations form one cluster. The other consisted of the 22nd and 23rd generations. As these generations were also more related to each other in regard to time, the observation made is in line with the expectation. For all the mentioned clusters, the two replicated samples of the same population and generation were closer to each other than to the different generations. The last two clusters observed were related to the conventionally evolved populations. While the two replicate samples of the 12th generation form one of these two clusters, the second cluster includes the 16th, 22nd, and 23rd generations. Contrasting to the organically evolved populations of the 22nd and 23rd generations, these two generations could not be clearly separated from each other by this cluster analysis (figure 50). The four samples of these two generations were highly similar to each other, which was in line with the observations made in the PCA (figure 48).

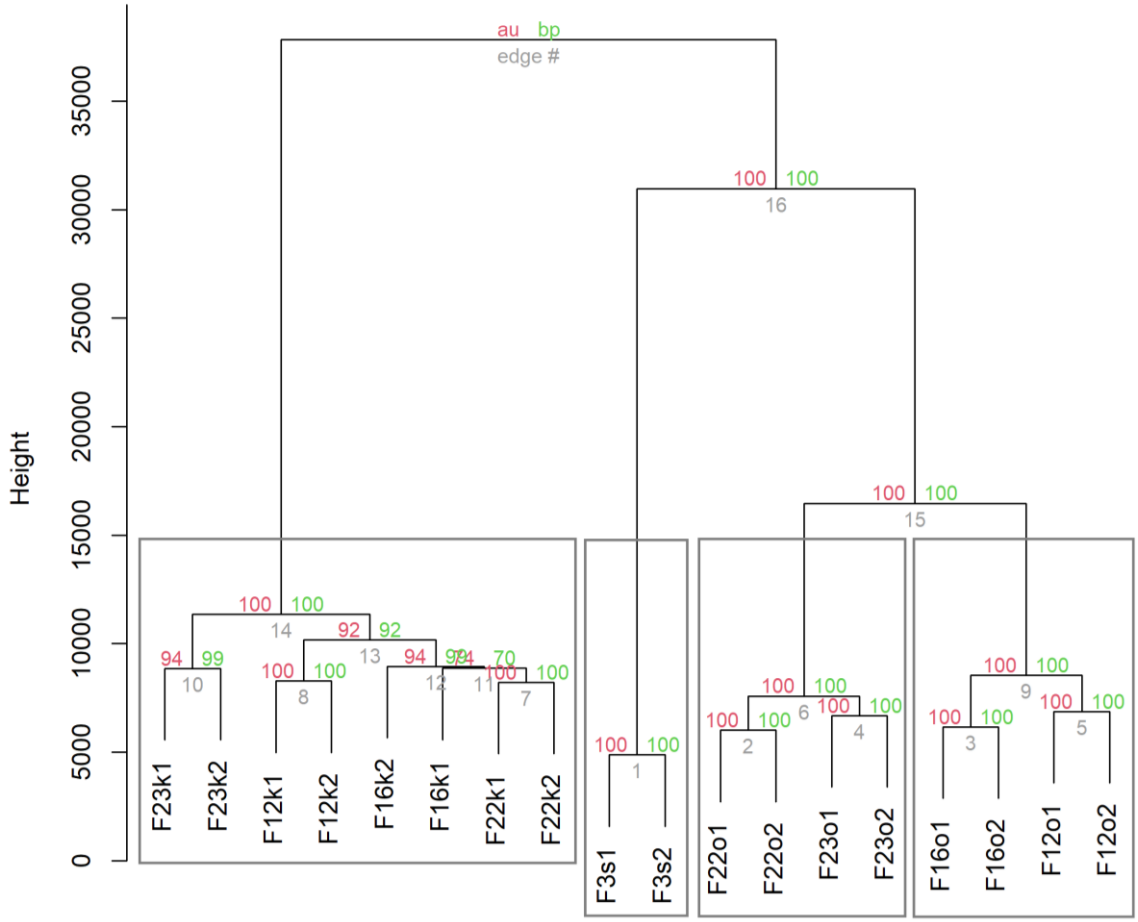


Figure 50: Genetic distance of all tested pool samples. The dendrogram illustrates the relation of test samples to one another. The name of the samples is constructed by the generation ($F_{3/12/16/22/23}$), environment (k – conventional; o – organic), sample (1 / 2). Each node is surrounded by three numbers – the grey number specifies the cluster rank; the red indicates the approximately unbiased (au) calculated p-value, while the green number indicates the bootstrap probability (bp) value. As closer the value is towards 100, as more valid the cluster. On a cut-off level of 15,000, four clusters can be identified: the initial F_3, as well as the early (F_{12}, F_{16}) and the late (F_{22}, F_{23}) organic populations, and finally, the conventional populations with all samples in one cluster.

The observed donor genome-wide allele frequency (gwAF) of the BC_2F_3 generation was 10.07%. In the conventionally selected populations, the gwAF was continuously reduced to 6.95% in the BC2F23 generation. In contrast, in the organic system, the donor gwAF was reduced to 8.96% in the 12th generation but subsequently increased to 10.57% in the 23rd generation, which was 0.5% greater than in the initial BC_2F_3 generation (figure 51). While for the given population size during the backcrossing, the expected donor gwAF in the BC2F3 was 10.5% (Cox, 1984), a minor deviation of 0.43% was observed compared to the 3rd generation, 2093 ISR42-8 related gene haplotypes were positively selected in the 12th generation of the organic environment ($p < 0.01$). The average allele frequency of these ISR42-8 alleles is 8.67%. In the 16th generation, 3404 genes on an average allele frequency of 13.44% were positively selected for the ISR42-8 allele. The value was increased to 5432 genes in the 22nd generation with a further increased ISR42-8 related allele frequency of 14.96%. Finally, the 23rd generation was characterized by 6,495 genes positively selected on an average ISR42-8 allele frequency level of 18.34%. It can be reported that both the average ISR42-8 related allele frequency as well as the count of genes were constantly increased over the last generations in the organic system.

Compared to the organic environment, the conventionally farmed populations show much lower numbers of positively selected genes for the ISR42-8 allele. While 1,005 genes on an average ISR42-8 allele frequency of 21.7% were observed in the 12th generation, the gene count in the 16th generation

drops down to 702 genes (average ISR42-8 allele frequency of 24.70%). Contrasting to this, the count of genes was increased in the 22nd and 23rd to 2,378 and 2,545 genes, respectively (21.28%, 22,52%). Concluding these observations, the count of genes with an increased ISR42-8 allele frequency was twice as high in the organic compared to the conventional environment across all tested generations. Contrasting to the gene count, the ISR42-8 frequency level of these positively selected genes was higher in the conventional system. This might indicate a more distinct selection towards these fewer genes in the conventional environment compared to a less distinct and heterogeneous selection in the organic system.

Besides the genes indicating any signs of selection forces, the majority of genes were classified by a neutral selection potential. 19,467 genes in the 12th generation were found to be not significantly different from the 3rd generation. According to this, the observed ISR42-8 allele frequency of this group (9.69% organic, 9.38% conventional) was close to the genome-wide observed ISR42-8 allele frequency of the BC_2F_3. (10.07%). The value of non-selected genes decreases over duration of the experiment and generations (BC_2F_{16} – 17,827; BC_2F_{22} – 14,912; BC_2F_{23} – 13,476). For these genes, the ISR42-8 allele frequency was almost equal between the environments over all tested generations.

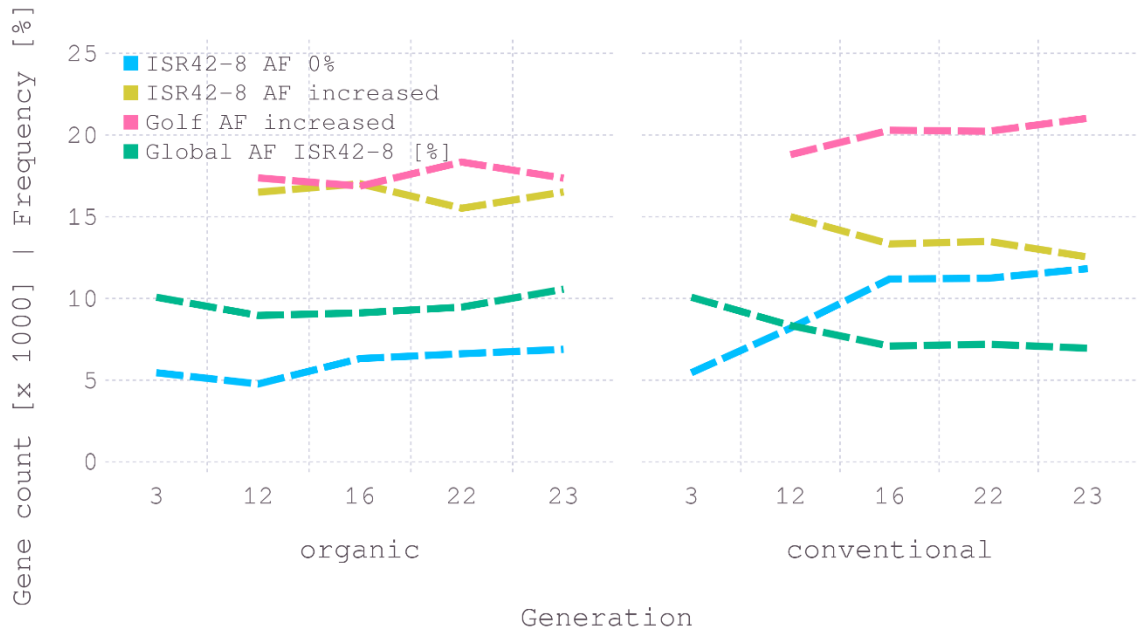


Figure 51: Genome-wide overview of the allele frequency evolution over the generations as well as the count of positive selected wild donor and cultivar alleles. The x-axis shows the generations split by the organic and conventional systems. The y-axis presents the gene count (multiplied by 1000) and the allele frequency in %. The blue line highlights the count of genes for which no ISR42-8 alleles can be observed in the population. AF is an acronym for allele frequency. ISR42-8 is the wild-type, while Golf is the cultivar parent. The green line represents the genome-wide wild donor allele frequency for each generation. The yellow line indicates the count of genes for which the wild-type allele frequency has increased in the population, compared to the F_3. Analogous to this, the red line represents the same information for the cultivar alleles in the population.

From 34,343 analyzed gene-based haplotypes, 7,115 wild-type specific ones were actively and passively positively selected in the BC_2F_{23} generation of the conventional system, where the same holds true for 15,124 cultivar-derived haplotypes. Three thousand sixty-six wild alleles were not found in the BC_2F_{23} of the conventional system. Five hundred twenty-four of these were already missing by the 12th generation onwards. In the organic system, the count of genes, which increased in wild-type allele frequency (wtAF), is 51% higher than in the conventional system, and only 547 of the gene-based haplotypes did not show wild-type alleles present in any of the analyzed generations. (Table 14, Figure 51).

Observed evolution of populations enforced by farming environment

		Count of Genes	
	Generations	*conventional env.*	*organic env.*
	Total number of Genes	*34343*	*34343*
count of genes variant to F3	*F12*	*22308*	*20834*
	F16	*22046*	*21598*
	F22	*22940*	*21839*
	F23	*22240*	*22773*
Wild-type allele positively selected to F3	*F12*	*9179*	*9566*
	F16	*7548*	*10361*
	F22	*8115*	*9658*
	F23	*7115*	*10721*
Cultivar allele positively selected to F3	*F12*	*13128*	*11268*
	F16	*14498*	*11235*
	F22	*14824*	*12180*
	F23	*15124*	*12052*
0% wild-type AF	*F12*	*1857*	*442*
	F16	*2105*	*728*
	F22	*2318*	*622*
	F23	*3066*	*912*
0% wild-type AF	*All generations from F12-F23*	*524*	*127*
	All generations from F16-F23	*806*	*188*
	All generations from F22-F23	*1379*	*376*

Table 14: Overview of significantly changed genes in their allele frequency compared to the F_3 generation ($p< 0.001$). The first column differentiates major blocks; the second column differentiates the generations, and the third column differentiates the count of gene data, separated for the conventional and organic environment. Count of genes variant to F_3 – all genes that show a significant variation ($p < 0.001$) compared to the initial generation. Wild-type allele positively selected to F_3 – the fraction of prior identified genes which has a sig. higher wild-type allele frequency compared to the F_3. Cultivar allele positively selected to F_3 – same as above, but for the cultivar allele frequency positively selected. 0% wild-type AF (first block) – the count of observed genes with no evidence of wild-type allele frequency. 0% wild-type AF (second block) – the count of genes for which wild-type alleles are non-existent in more than one generation; All generations from F_{12} toF_{23}, including F_{12}, F_{16}, F_{22}, F_{23}.

The genome-wide introgression level of the donor parent ISR42-8 in the population was calculated on the gene-based haplotype level, as low coverage sequencing would result in too many false positive hits. 531 of 34,344 HC genes (1,55%) are observed to have a transgenerational wild-type allele frequency below 0.005 and therefore are not expected to be introgressed in the population. Most of these genes are located on chromosome 2H and the short arm of chromosome 3H and 4H. This finding indicates an almost complete introgression of donor alleles, though it does not say anything about the frequency of the introgressed loci.

3.5.3 Impact of genetic drift on allele frequency variations

To ensure that the observed assemblage was not only affected by genetic drift, we simulated the allele frequency pattern of a double backcross population after twenty-three generations. The observed pattern of allele frequency heterogeneity and evolution significantly differentiates between the simulated and the discovered data.

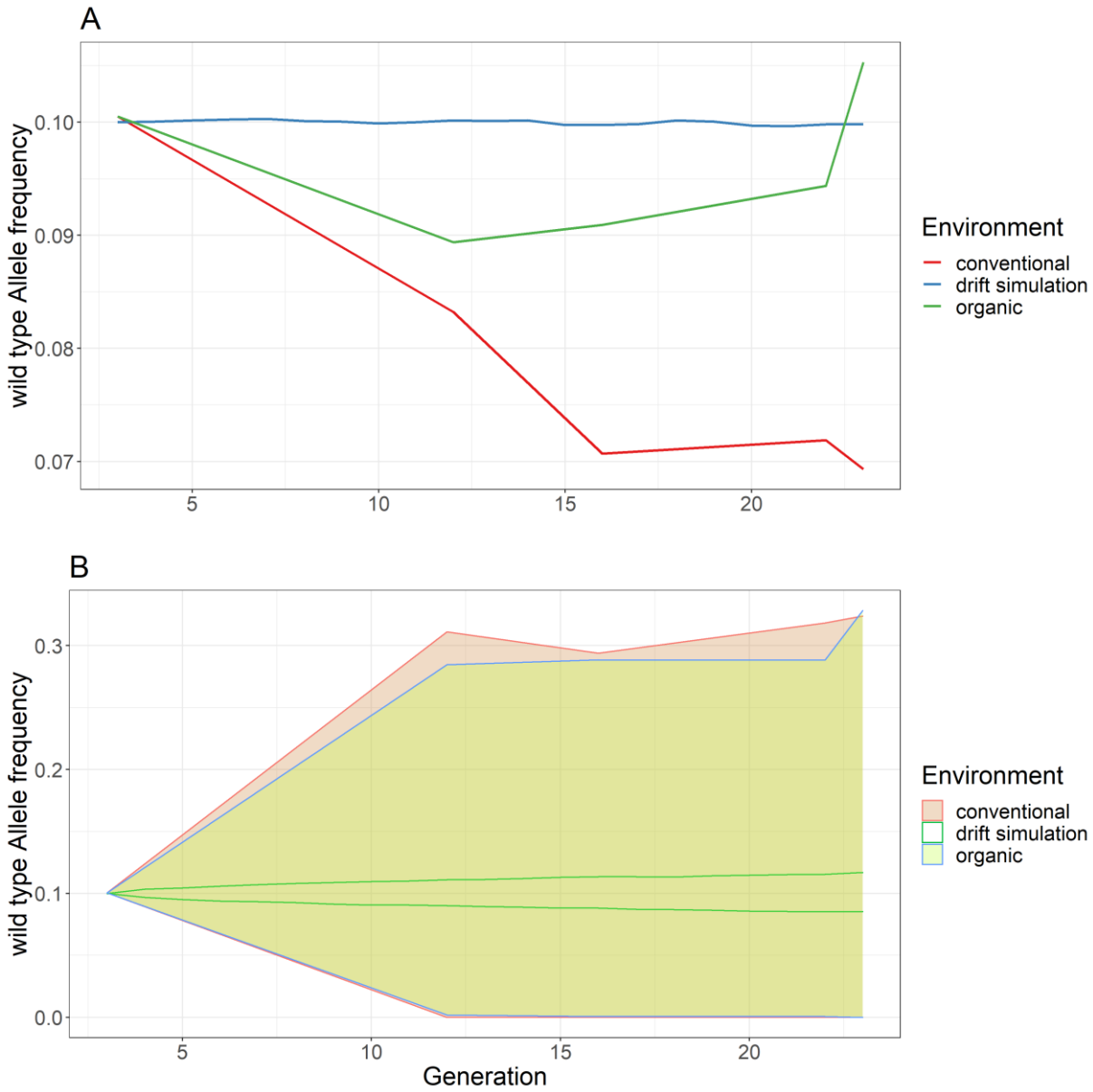


Figure 52: Effect size estimation of genetic drift, compared to the allele frequency evolution in the organic and conventional population over all generations. A – the average genome-wide wild-type allele frequency value (y-axis) against the generations (x-axis). The conventional, organic, and drift are separated by colors. B – the 95% confidence interval of the gene-based haplotypes for conventional and organic populations. Additionally, the confidence interval of the drift is plotted and compared against both field experiment-based environments. When compared, the drift as a single factor showed a ten times smaller effect on the allele frequency variation compared to the conventional and organic population. The conventional and organic lines and areas are the results of the real data, while the drift line is the result of a simulation.

As illustrated in figures 52 and 53, the random effect of simulated genetic drift resulted in a much less pounced and untargeted allele frequency variation over the generations compared to the observed allele frequency shifts reported by the measured populations. Both organic and conventional populations

showed a much higher range of variation, ranging from 0.0 to 0.3. Based on variance analysis, two results were obtained. First, when all 1,000 AF drift simulated iterations were used as a dataset and compared to another generation, no variation below $p < 0.05$ was observed for any generation to another. The second result is illustrated in table 14. The same set of allele frequency observations made in the BC_2F_3 was used for the simulation calculation as baseline (as illustrated by the exact same gwAF in table 14A). The gwAF changed little in the drift simulation in the following generations, contrasting to the conventional and organic population gwAF. Almost 30% wild-type gwAF decrease was observed in the conventional system in the last generation, while the drift simulation was described by a variation of 0.5% (F_3 to F_{23}). The comparison of the median allele frequency for all genes over the 1,000 iterations (drift) to the organic and conventional environment is illustrated in table 14A on the left-hand side. The low probability values indicated a severe variation in the allele frequency pattern of the organic and conventional populations that exceeded the drift effect by far. The upper and lower end of the confidence interval comparison of the drift simulation compared to the organic and conventional environment indicates that the increase of the wtAF was overrepresented in the real data. The same holds true for the low-end confidence interval but on a less high significant level. Taken together, when the allele frequency pattern of the simulation was compared generation-wise to the real allele frequency pattern, highly significant variations between these two sources of data were described (table 15). For all generations and both environments, a high deviation of simulated and real data was observed. Drift might have caused the loss or change of some alleles but cannot explain the majority of the observed changes in the allele frequency.

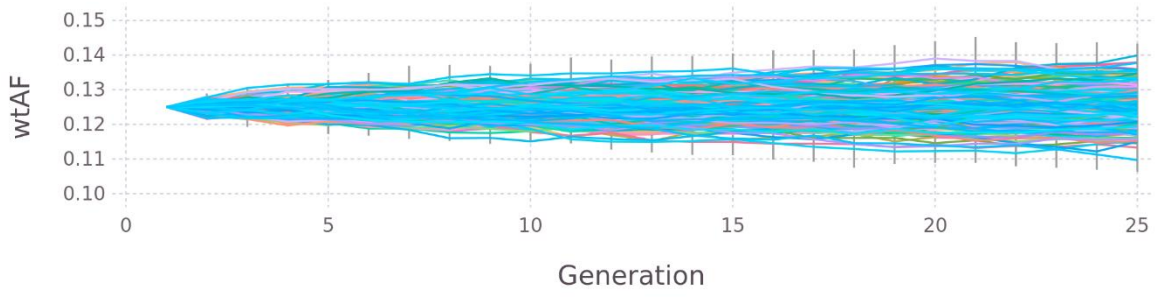


Figure 53: Donor allele frequency evolution for 5,500 simulated genes under the influence of genetic drift. The same data as in figure 52 is visualized on a higher resolution on the y-axis.

Table 15: Validation of random genetic drift of the observed allele frequencies of 34,237 haplotypes. The average allele frequency of 1,000 replicates of drift simulation was tested with a negative binomial linear model against the actually observed allele frequencies of genes for each generation. This test was used to check if the simulated and actual data were equal and if so, the observations were based on random drift only. Values of < 0.001 are assumed to highlight a distinct variation between the simulated and actual allele frequencies, indicating that the genetic drift did not have the power to explain the majority of observed allele frequency changes over the generations. A – The p-value reports the difference between the organic and conventional against the drift simulated population for each generation separately. The drift allele frequency value of each corresponding haplotype is calculated as the median value over the 1,000 iterations. Low p-values indicate high divergence between the drift and the real population sample. Next to the p-values, the average genome-wide wild-type allele frequency values are reported in percent for organic, conventional, and drift populations. Drift simulation used the BC2F3 gene-haplotypes allele frequency as starting value. B – The low and upper limits of the drift simulation 95% confidence interval compared to the organic and conventional population for each generation. It follows the same logic as stated for A.

A

			AVERAGE GWAF [%]		
	Organic environment	Conventional environment	organic	conventional	simulated drift
F3	1	1	10.05	10.05	10.05
F12	9.61E-92	3.13E-34	8.93	8.32	10.05
F16	3.09E-27	4.62E-71	9.09	7.06	10.05
F22	0.000114009	2.13E-51	9.43	7.18	10.06
F23	1.99E-38	1.29E-54	10.53	6.93	10.06

B

	Organic environment		*Conventional environment*	
Comparison to	Low-end confidence interval	Upper-end confidence interval	Low-end confidence interval	Upper-end confidence interval
F12	3.30E-23	5.48E-207	0.005139584	9.16E-104
F16	0.068741754	4.47E-121	2.87E-11	1.33E-184
F22	9.44E-27	1.75E-74	0.045137162	9.50E-173
F23	5.98E-159	0.307200274	0.007000865	4.75E-178

3.5.4 Selection pressure in regard to chromosomal positions

We observed a genome-wide variation in allele frequency in both farming systems. The first generation tested was the 3rd generation. As illustrated in figure 54, most chromosomal regions were defined by a slight deviation from the expected allele frequency of 12.5% for infinite population sizes. Chromosomes 1H, 3H, 4H, and 6H showed a very close link to the 12.5% line. Besides single gene allele frequency deviations, the expectation was matched. Contrasting to these findings, the 2H, 5H, and in a smaller extend also chromosome 7H showed substantial deviations from the expected value. While the ends of the telomeres on chromosome 2H were found to match the expectations in allele frequency, the pericentromeric regions were classified by an increased ISR42-8 allele frequency above the expected value. Contrasting, the centromere was underrepresented in ISR42-8 allele frequency. In the centromere, the allele frequency of ISR42-8 was found to be close to 0%. The chromosome-wide average allele frequency of ISR42-8 is 18.4%. Besides the centromere of chromosome 2H, a distinct and unexpected reduction in allele frequency was observed on chromosome 5H. Besides the first few million base pairs, the fifth chromosome was classified by an underrepresentation of the ISR42-8 allele frequency from 10 million up to 550 million base position. The average allele frequency in this region was far below 1.55%, which presents an eight-fold deviation from the expected value. This observation concludes that more than 80% of the fifth chromosome was characterized by a high underrepresentation of ISR42-8 allele on this chromosome. The 7th chromosome is also classified by a lower allele frequency than expected. Over the entire 7th chromosome, the ISR42-8 allele frequency in the population was 3.8%. Compared, the chromosome-wide ISR42-8 allele frequency on the 1st, 3rd, 4th, and 6th chromosomes was 14%, 5.7%, 10.4%, and 11.7%, respectively.

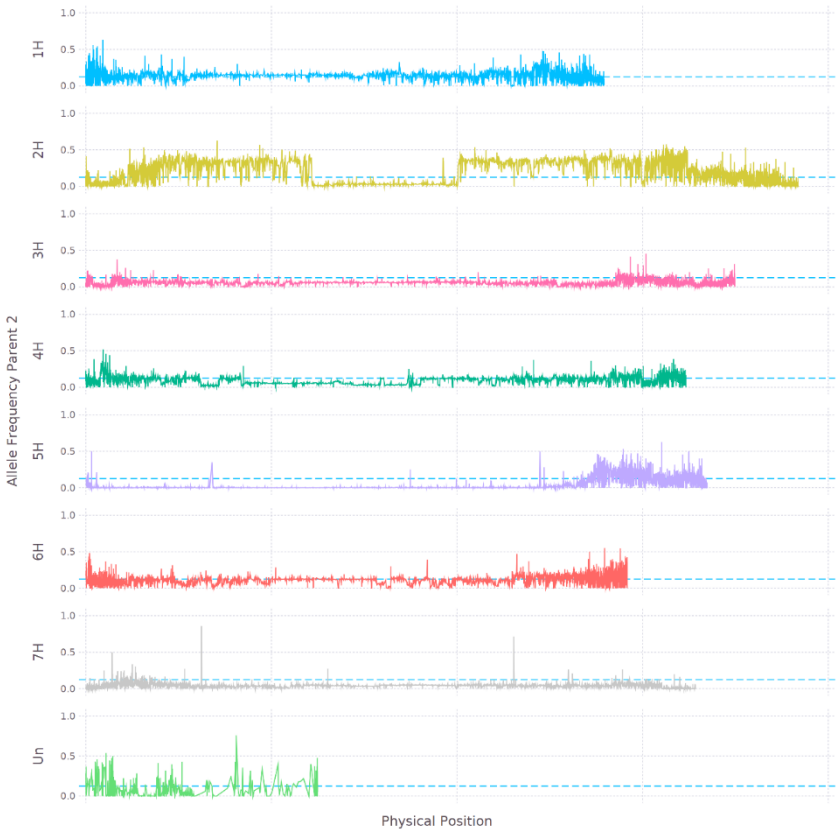


*Figure 54: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The dashed line indicates the expected allele frequency of 12.5% for infinitely sized populations in the presented **3rd generation**.*

Compared to the BC_2F_3, the ISR42-8 allele frequency pattern of the 12th generation had changed severely. As illustrated in figure 55 for the conventional population, most regions were characterized by a clear negative selection of the wild ISR42-8 alleles. As the initial BC_2F_3 had evolved for nine consecutive generations in either an organic and a conventional system, the 12th generation is the first to compare these two farming systems and their impact on the allele frequency patterns.

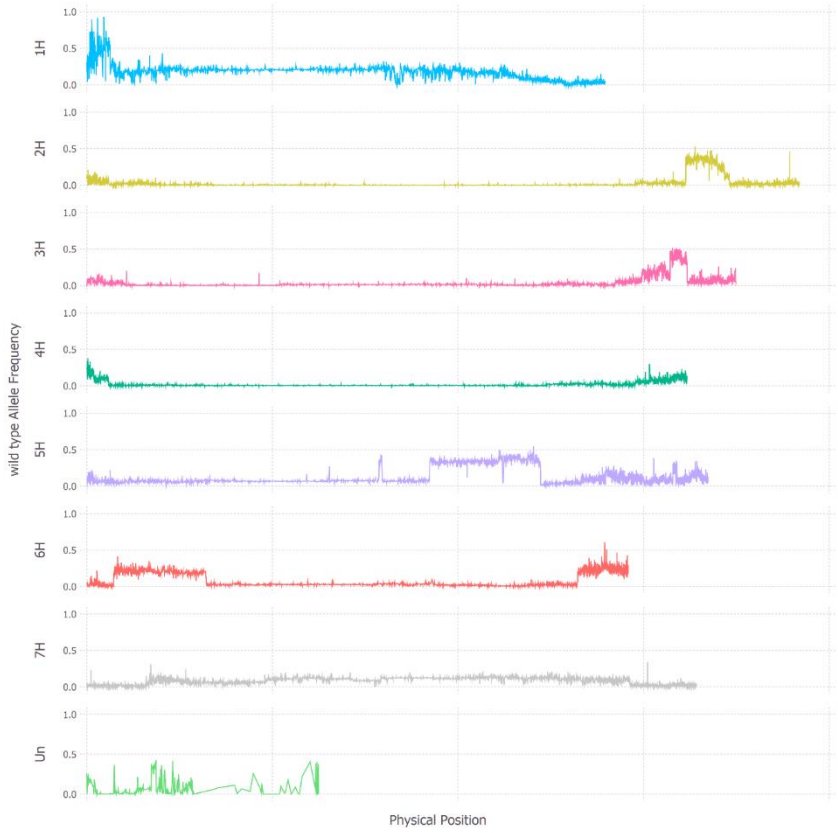


*Figure 55: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The observed genome-wide wild-type allele frequency (ISR42-8) of the **F12 conventional** population is illustrated. The x-axis illustrated the physical genome position, while the y-axis presents the allele frequency (0 to 1). The illustration is based on the haplotype allele frequency observations.*

The evolution in the conventional system led to the negative selection of ISR42-8 alleles in many regions of the genome. While the ISR42-8 allele frequency (IAF) was almost unchanged on chromosome 1H (16.24% to 14% in F_3), the IAF was reduced by 14% on chromosome 2H (4.1%). Besides one region on the telomere on the long arm, the entire chromosome was characterized by a reduction of the allele frequency. The rather high IAF in the pericentromeric regions observed in the 3rd generation was reduced to nearby 0% in only nine generations. The positively selected region on chromosome 2H reaches its peak with an IAF above 40%. A slight variation can be overserved on chromosome 3H. Analogous to the second chromosome, most regions are described by a very low IAF, and a region on the telomere of the long arm indicated a positive selection of IAF of up to 50%. Compared to these distinct variations on micro levels, the chromosome-wide IAF had almost not changed (6.05%). The 4th chromosome indicated a reduction of the IAF over the entire chromosome. Excluding the outer ends of the telomeres, the entire chromosome showed an IAF close to 0%. The chromosome-wide IAF was reduced to 4.1%. Chromosome 5H had changed in the IAF in a remarkable way. In the priory low IAF observed interval from 10 million to 550 million base pairs, a block from 370 million to 490 million base pairs had increased in IAF from below 3% to above 40%. This was associated with a redouble of IAF to 12.5% over the entire chromosome. For chromosome 6H, the chromosome-wide IAF had changed little

between the 12th and 3rd generations. Contrasting to this, the changes of IAF in distinct regions were significant. On the short arm, a block of 110 million base pairs indicated an IAF positive selection of up to 30%. The same pattern could be observed for the distal end of the long arm. For chromosome 7H, a chromosome-wide increase of the allele frequency from 3.7 to 6.1% could be reported. Contrasting to the other chromosomes, no region was specifically positively selected. In fact, the IAF on the whole chromosome was slightly positively selected.

Little variation between the conventional and organic populations was observed based on the IAF up and negative selection pattern in the 12th generation. Based on the observations made in the 12th generation, clear variation from the 3rd generation was identified, but the different farming systems did not result in a different IAF pattern.

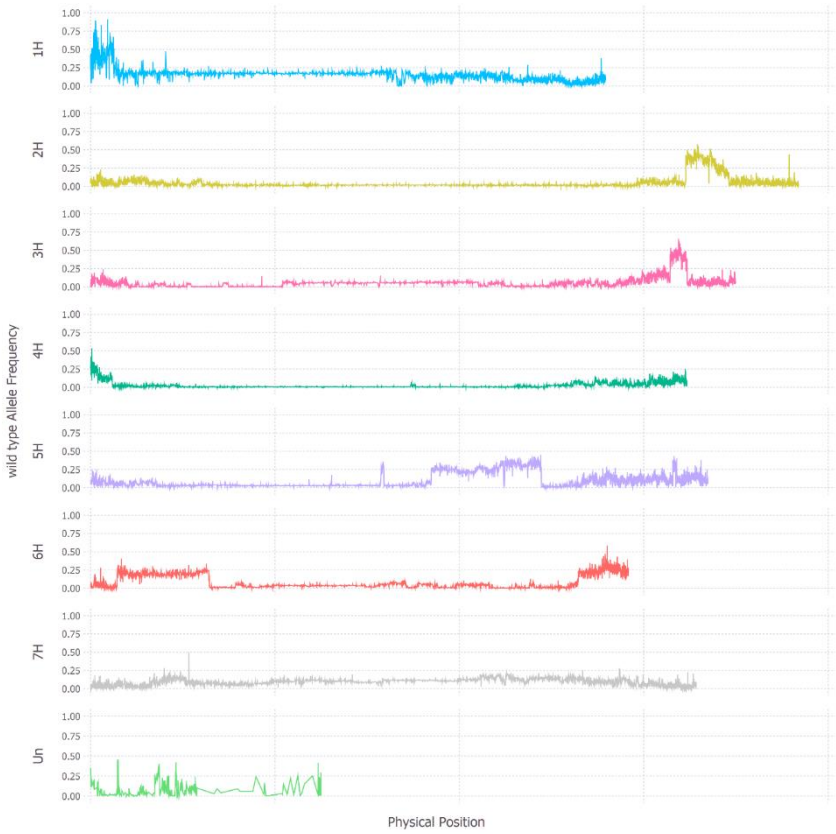


*Figure 56: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The observed genome-wide wild-type allele frequency (ISR42-8) of the **F12 organic** population is illustrated. The x-axis illustrated the physical genome position, while the y-axis presents the allele frequency (0 to 1). The illustration is based on the haplotype allele frequency observations.*

The conventional 16th generation showed a follow-up of the observed allele frequency changes made in the 12th conventional generation. No significant variation in the allele frequency patterns was identified (figure 57). Contrasting to the stagnation of allele frequency variation from one tested generation to another, the organically farmed populations showed apparent variations between the 12th and 16th generations. While the patterns on chromosomes 1H, 2H, and 4H had not changed much, chromosome 3H was described by an increasing IAF in the pericentromeric regions. The IAF in this region increased to a level of 12%, covering more than 150 million base pairs in size. Despite the pericentromeric region, distinct IAF increased peaks could be found on the short arm in the pericentromeric regions of chromosome 3H. Chromosome 5H had not changed as much as 3H, but the previously positively selected IAF block was reduced in its IAF for fractions of this region. This might be an indication that unbeneficial alleles linked to one or more beneficial alleles were deselected. Two peaks remain on the

same level as before, while the region closer to the centromere was described by a slightly decreased IAF. Analogous to chromosome 3H, the centromere of chromosome 6H was positively selected. The IAF in this region was a little lower than reported for 3H (8%) but indicating a clear tendency of positive selection. In the pericentromeric region on the long arm of chromosome 6H, three further peaks in IAF are observed to emerge (Figure 58).

Figure 57: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The observed *genome-wide wild-type allele frequency* (ISR42-8) of the ***F16 conventional*** *population* is illustrated. The x-axis illustrated the physical genome position, while the y-axis presents the allele frequency (0 to 1). The illustration is based on the haplotype allele frequency observations.

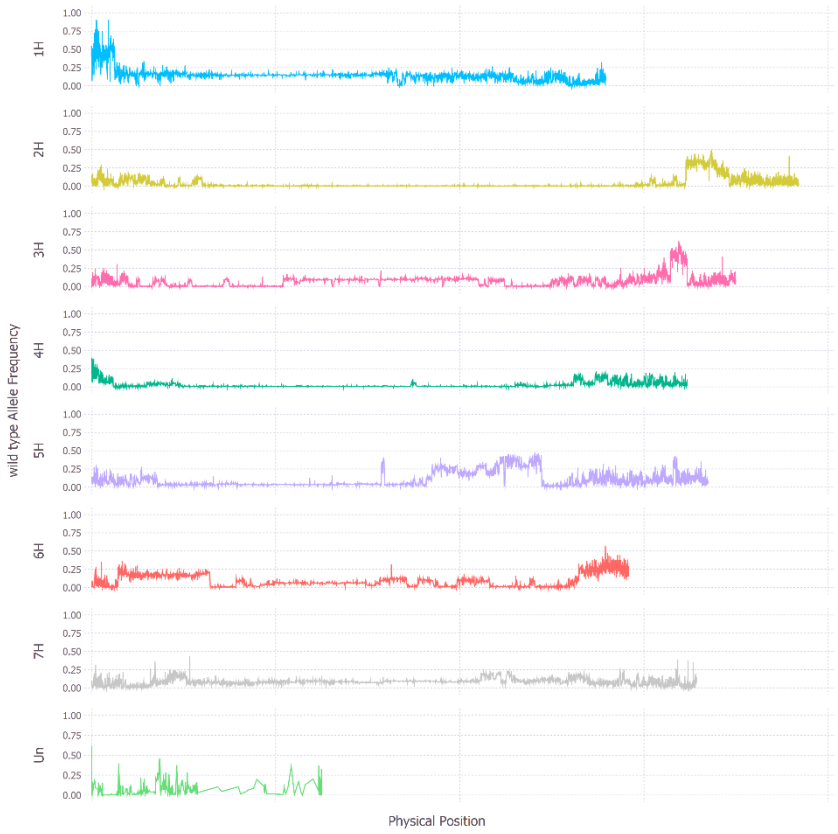


Figure 58: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The observed *genome-wide wild-type allele frequency* (ISR42-8) of the ***F16 organic population*** is illustrated. The x-axis illustrated the physical genome position, while the y-axis presents the allele frequency (0 to 1). The illustration is based on the haplotype allele frequency observations.

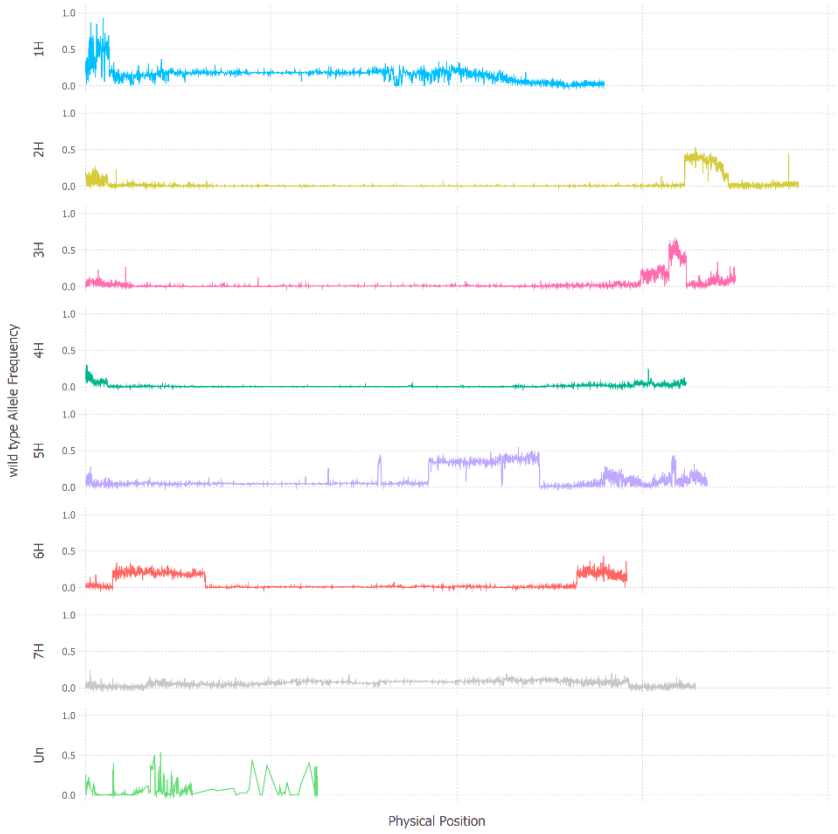


Figure 59: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The observed genome-wide wild-type allele frequency (ISR42-8) of ***the F22 conventional*** *population is illustrated. The x-axis illustrated the physical genome position, while the y-axis presents the allele frequency (0 to 1). The illustration is based on the haplotype allele frequency observations.*

Figure 60: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The observed genome-wide wild-type allele frequency (ISR42-8) of the F22 organic population is illustrated. The x-axis illustrated the physical genome position, while the y-axis presents the allele frequency (0 to 1). The illustration is based on the haplotype allele frequency observations.

Similar to the conventional 16th generation, little variation was observed in the 22nd generation. The most relevant variation was found on chromosome 3H, where the IAF of the block on the telomere on the long arm was further increased to more than 50%. All other regions remained more or less identical to the 12th and 16th generation (Figure 59). Compared to the conventional system, the 22nd generation in the organic system changed more clearly compared to the 16th generation. The centromere of chromosome 3H increased further in IAF up to 25%, while also the telomeric region on the long arm was described by an increase of the IAF above 75%. Furthermore, the IAF on the telomere on chromosome 4H was increased to 50% at its maximum. This region was characterized by many small peaks, indicating a positive selection of many genes, separated by linkage breaks. The previously observed differentiation in the centromere of chromosome 5H was further progressed. In the 22nd generation, this region was characterized by at least four clearly distinguishable peaks, separated by regions with a distinct negative selection of the IAF. The most remarkable positive selection of IAF can be observed on chromosome 6H. More than two-thirds of the chromosome can be found to be increased in the IAF. The most relevant variations compared to the 16th generation were observed in the pericentromeric region, where the IAF increased to a maximum level of 25%. This was divided into at least seven subregions. On chromosome 7H, some regions in the telomeres on both arms were positively selected in IAF. Eight regions that exceed an IAF of 25% were reported (Figure 60).

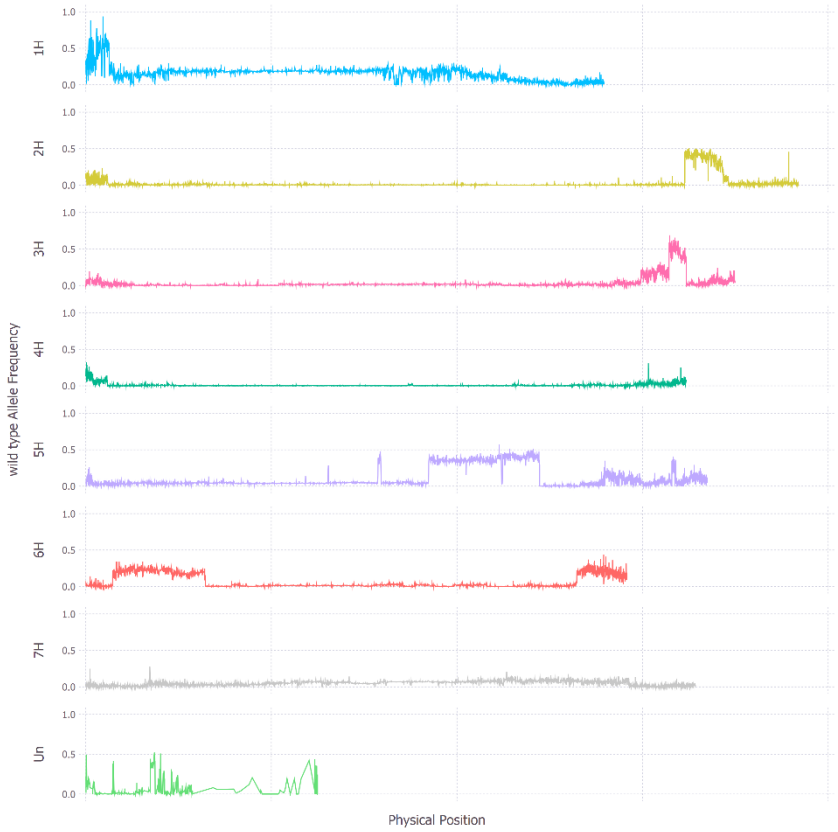


Figure 61: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The observed genome-wide wild-type allele frequency (ISR42-8) of the F23 conventional population is illustrated. The x-axis illustrated the physical genome position, while the y-axis presents the allele frequency (0 to 1). The illustration is based on the haplotype allele frequency observations.

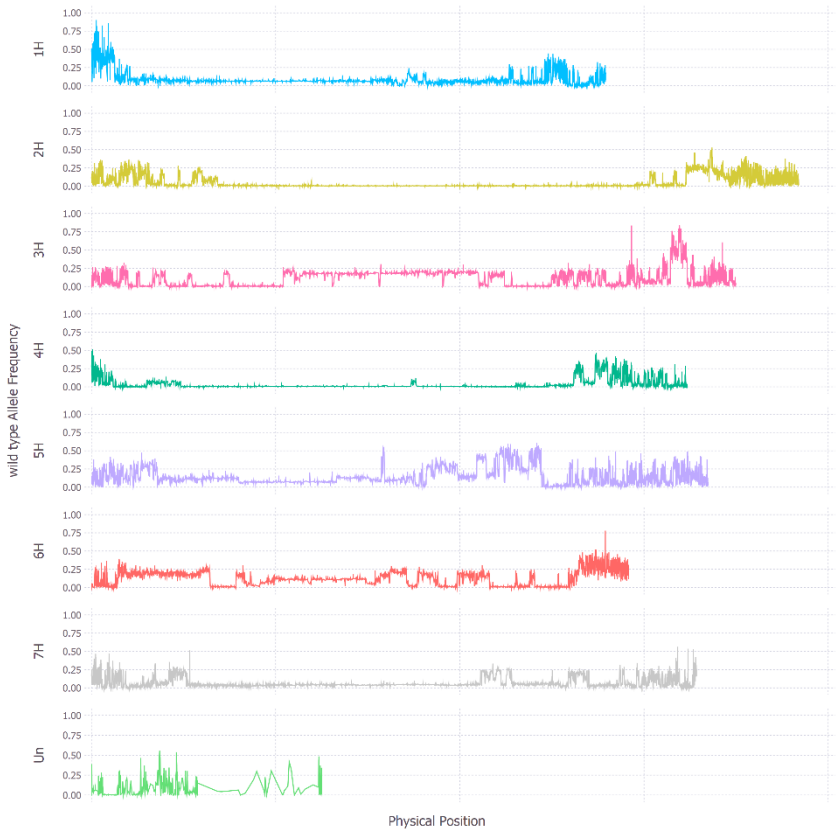


Figure 62: Genome-wide ISR42-8 allele frequency against the physical position, horizontally stacked by chromosome. The observed genome-wide wild-type allele frequency (ISR42-8) of ***the F23 organic*** *population is illustrated. The x-axis illustrated the physical genome position, while the y-axis presents the allele frequency (0 to 1). The illustration is based on the haplotype allele frequency observations.*

The pattern of little variation to the previous generation was also observed for the 23rd generation in the conventional environment (Figure 61). As a matter of fact, the last two analyzed populations were coming from consecutive years. Therefore this observation might be expected. Contrasting to the conventional environment, the organic population of the 23rd generation showed deviations from the previous year's generation. The huge centromeric block on chromosome 1H was negatively selected, while a region in the telomeric region on the long arm was positively selected in its IAF. The same positive selection pattern was reported for the telomere on the short arm of chromosome 2H. All blocks indicating a positive selection on chromosome 3H in previous generations were further increased in their IAF. The same holds true for the 4th chromosome. On chromosome 5H, the variations in IAF of the single regions in the pericentromeric block became more distinct. On top, both telomeres were described by an increase of the IAF. An increase of IAF was observed in the centromere and telomere on the long arm of chromosome 6H, too. The same holds true for distinct regions on the telomeres of chromosome 7H (Figure 62).

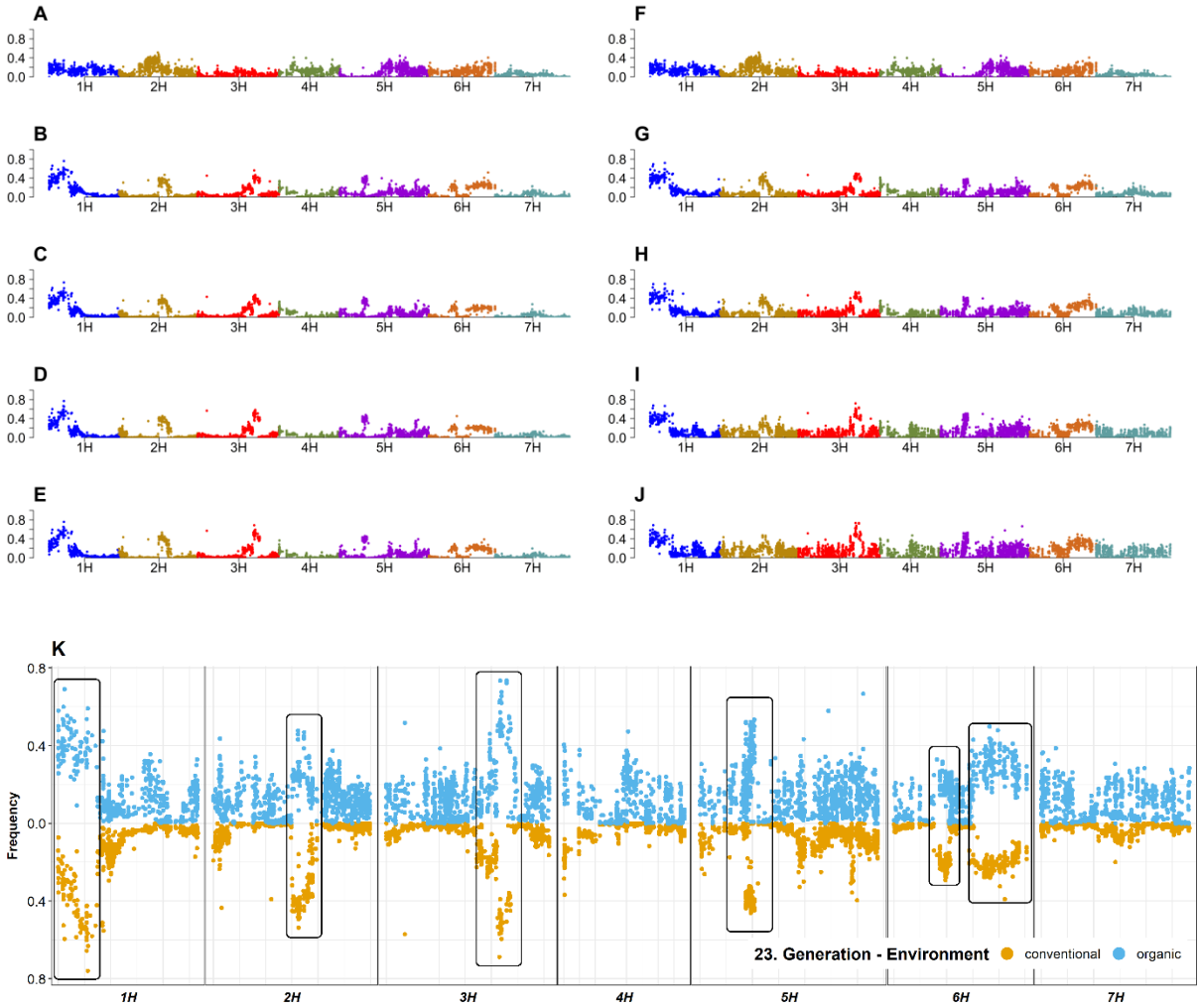


Figure 62-1: Genome-wide wild-type allele frequency on the barley genetic map. The left part of the plot represents the conventional system, where each subplot corresponds to one generation. (A) F3; (B) F12; (C) F16; (D) F22; (E) F23. The right part illustrates the generations that evolved in the organic system. (F) F3; (G) F12; (H) F16; (I) F22; (J) F23. Both samples per generation and environment were merged into one dataset. The wild-type allele frequency of the (y-axis) haplotypes is shown on the genetic map (x-axis), where each of seven barley chromosomes is indicated by different colors. (K) direct wild-type allele frequency comparison between the F23 organic and conventional population. The allele frequency (x-axis) is plotted against the genetic position chromosomes-wise (y-axis), the environments are separated by colors. Boxes indicate regions with similar positive wilt type allele frequency selection in both systems.

Comparing the organic and conventional populations directly to each other (Figure 62-1), variations between the systems became obvious. As illustrated in figure 62-1, six major regions with similar wtAF development over the generations are shown by boxes in subfigure K. Beyond these, the organic population showed evidence for a higher and more diverse wtAF in most loci. While the wtAF was low in most regions (wtAF < 0.2), almost the whole genome, illustrated on the genetic map, indicated increased and heterogeneous wtAF values. Over all generations, a reduction of wild-type alleles in the conventional system was reported, while the reduction of the wild alleles was overturned in the organic population after the 12th generation onwards.

It can be concluded that the organic and conventional populations evolved similarly until the 12th generation. While the conventional populations showed little variation in the later generations tested, compared to this 12th generation, the organic system started to differentiate from the 12th generation. The IAF of numerous regions was increased in the following generations in the organic population, which resulted in an evident variation between the evolution of the conventional and organic systems.

3.5.5 From selection sweeps to candidate genes of agronomic traits

The identification of changed allele frequency patterns had to be accommodated with biological meaning. Numerous QTL studies were screened to identify relevant regions and traits to give meaning to the observed allele frequency shifts. The telomere of chromosome 1H had shown high IAF above both environments and several generations. In this region, resistance genes against powdery mildew and *fusarium graminearum* were identified (Bedawy et al., 2018; M. Von Korff et al., 2005). The positive selection of IAF might be associated with these two loci. While the locus of the powdery mildew resistance gene *Mla* was rather unspecific (0-28cM), the resistance to fusarium was reported to be on 30cM. For *Mla*, the genes *HORVU1Hr1G001490* and *HORVU1Hr1G001500* were identified as potential candidates. The first was reported to be functional as a disease resistance protein and is characterized by 44 amino acid variants between Golf and ISR42.8. Furthermore, the IAF in both farming systems was high, with a value above 60%. The second candidate for the powdery mildew locus was a *TRICHOME BIREFRINGENCE-LIKE 8* protein, characterized by a single base pair deletion in Golf, resulting in 48 variant amino acids. The IAF of this gene was 64% in the conventional and 70% in the organic 23rd population. The candidate gene related to the fusarium resistance was *HORVU1Hr1G009800*. Its function is a bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein. The parents vary in the amino acid sequence by 44 AS. Additionally, the Golf related allele was described by a six bp deletion. The IAF is 55% in the conventional and 48% in the organic environment.

Another resistance locus against powdery mildew was reported to be on chromosome 2H at position 715 to 720 million base pairs. The potential candidate gene for this locus is *HORVU2Hr1G109840.* The function of this gene is so far undescribed but characterized by a three-base pair deletion in Golf which results in 210 variant amino acids between Golf and ISR42-8. The candidate gene of this locus was described by an IAF of 0% in the conventional and 26% in the organic environment (M. Von Korff et al., 2005).

In the region of 100 to 130 cM on the 3rd chromosome, Perovic et al. identified the QTL Ryd2, which is associated with a barley yellow dwarf virus resistance(Lüpken et al., 2014). The identified candidate gene was *HORVU3Hr1G089990*. Its function is characterized as a *RING/FYVE/PHD zinc finger superfamily protein*, and Golf was found to have a four base pair insertion, which results in 29 amino acid variations. The IAF of this gene was 51% in the conventional and 74% in the organic population in the 23rd generation, compared to 24% in the third generation. Another barley yellow dwarf virus resistance is reported on chromosome 6H on 50 cM. This Ryd3 locus could be related to *HORVU6Hr1G024780*. The function of this gene is reported as *Pathogenesis-related thaumatin superfamily protein.* The ISR42-8 allele frequency of this gene in the conventional environment was 25%, while it was 16% in the organic environment. The gene is characterized by a two-base pair deletion, resulting in a frame shift and 295 amino acid variation.

Two relevant leaf rust resistance loci have been described by von Korf – one on chromosome 4H (613-620 million bp), the other on chromosome 7H (637-655 million bp)(M. Von Korff et al., 2005). The most promising candidate gene on the 4H locus is *HORVU4Hr1G081400*, whose function is related to *AP2-like ethylene-responsive transcription factor AIL5.* The allele frequency of this gene was 1% in the conventional system, while 11% in the organic system. The parental alleles varied in a single base pair deletion in Golf, which results in 72 variant amino acids and a preliminary stop codon in Golf. The leaf rust locus on chromosome 6H was likely to be related to *HORVU7Hr1G121290, a flavin-dependent monooxygenase one* enzyme. The structural variation between the parents was two amino acids. The IAF is 6% in the conventional and 53% in the organic environment in the 23rd generation.

Twelve relevant QTL loci were identified to be active in drought response (Anonymous, n.d.; Honsdorf et al., 2014, 2017; Mohammed & Léon, 2004; Muzammil et al., 2018). P5CS1 on chromosome 1H at 500 million bp position acts in proline accumulation and can buffer oxidative stress caused by reduced water availability(Muzammil et al., 2018). The organic system was characterized by a 3-fold higher IAF compared to the conventional system (3.5%). A locus on chromosome 2H at 38-58cM was associated with plant growth and growth habit. The candidate gene *HORVU2Hr1G017270* is a *WUSCHEL related homeobox 12* and was described by an 18 bp deletion in the ISR42-8 allele, resulting in 116 variant amino acids. The IAF is highly variant – 27% in the organic and 1% in the conventional system. The drought-sensitive locus *Qtil.S42IL-3H* was identified to have an impact on the tiller number and biomass production (Honsdorf et al., 2014). The locus of this QTL is on chromosome 3H at 122-154 cM. The identification of a single candidate gene was difficult, but two genes showed evidence of being related to the functionality reported in the QTL studies. These genes are *HORVU3Hr1G096510 and HORVU3Hr1G108150*, functional as *auxin response factor 2* and *Cytochrome P450 superfamily protein*, respectively. They were selected as candidate genes based on allele frequency variations between the systems and on structural variations. While the IAF of the auxin response factor 2 was 3% in the conventional system, it was 44% in the organic – both regarding the 23rd generation. This protein of this gene was variant in six amino acids. The other candidate, the cytochrome protein, was defined by a less pronounced variation between the parental alleles. While the 23rd conventional generation is observed to have an IAF of 14%, the IAF of the corresponding organic population is 36%. Additionally, the protein structure varies by 15 amino acids as a result of a five base pair deletion in the Golf allele of this gene. A bit further downstream on the 3rd chromosome, at position 152-167 cM, a QTL locus was identified playing a crucial role in yield reduction under drought effects (*Hv13GEIII/ Bmac 0357*) (Mohammed & Léon, 2004). The proclaimed candidate gene for this locus was found to be *HORVU3Hr1G072920*, a *Y-family DNA polymerase H*. A deletion of 3 base pairs in the ISR42-8 allele of the gene resulted in 71 variant amino acids. The observed IAF in the organic is 23%, compared to 2% in the conventional system. The last observed drought related QTL on chromosome 3H was found on position 43-54 cM by Honsdorf et al. (Honsdorf et al., 2017). The QTL *Qhei.S42IL-3H* was found to be active in plant height reduction under drought stress. The position was also reported by Wang, related to a dwarfing gene (*HvFT2*) (Wang et al., 2010b). The region is close to the centromere of chromosome 3H. Therefore the specified genetic position was associated with a large covered area on the physical map. That made the identification of a candidate gene challenging, but the most promising candidate gene was *HORVU3Hr1G019210*. The amino acid structure of this *Diacylglycerol kinase* was variant between the parental alleles, based on a single base pair deletion in the ISR42-8 allele. The observed IAF of this gene in the 23rd generation of the organic system was 20%, compared to 2% in the conventional system.

Three QTL related to drought response were found on chromosome 4H. QTL *Qhei.S42IL-4H* on position 52-81 cM is associated with plant height(Honsdorf et al., 2014). The candidate gene *HORVU4Hr1G070940* is a *Zinc finger MYM-type protein 1* and was characterized by an IAF of 0% and 38% in the conventional and organic 23rd generation, respectively. The parental alleles of this gene vary by 3 amino acids. At position 111 cM, the QTL *Qtil.S42IL-4H* was detected, related to adaption in tiller number and the integral of plant height (Honsdorf et al., 2017). The most likely candidate gene of this region is *HORVU4Hr1G086500*, a *respiratory burst oxidase homologous B*. The parental alleles of this gene vary by three amino acids. The observed IAF in the conventional and organic 23rd generation is 3% and 13%, respectively. The last QTL in regard to drought tolerance found on chromosome 4H is *Ebmac 0701* (Mohammed & Léon, 2004). This locus is associated with relative leaf water content. The candidate gene of this region is regarded as interesting as both parents have a deletion in comparison to the reference sequence. While the ISR42-8 allele was found to have a five base pair deletion, Golf was observed to have a three base pair deletion. This leads to 292 variant amino acids of the *Ras-related*

protein Rab-6A. On the IAF level, the systems are highly different. While the conventional 23rd generation is characterized by 5% IAF, the organic has shown to have 23% IAF. Compared to the IAF observed in the 3rd generation, the conventional system has not changed.

One single locus relevant under drought stress was identified on chromosome 5H - Bmag 0357(Mohammed & Léon, 2004). The locus is regarded to be active in multiple phenotypic traits related to leaf per tiller, yield components, and biomass. Two potential candidate genes were observed for this locus. The *DCD (Development and Cell Death) domain protein HORVU5Hr1G091160* was characterized by a two-fold higher IAF in the organic compared to the conventional system (31% to 15%). The protein code also varies for 638 amino acids, based on a six-base pair deletion in the Golf allele. The other candidate gene, *HORVU5Hr1G088460*, is functional as a *Mitochondrial glycoprotein family protein*. The protein code varies in 194 amino acids, based on a 12 base pair deletion in Golf. The IAF pattern of this gene was similar to the previous gene, showing a higher IAF in the organic (36%) compared to the conventional system (12%).

Another drought candidate locus was reported for chromosome 6H – *Qagri.S42IL-6H*(Honsdorf et al., 2014). This locus is located in the interval of 49 to 76 cM and has an effect on *Biomass, Growth Plant, Height, Shoot Area, and water use efficiency.* Two candidate genes with a comparable IAF pattern were selected for this region, based on the fact that the IAF was 12 times higher in the organic compared to the conventional system (24%, 2 & 3%). The first of both, *HORVU6Hr1G046850*, a *Photosystem II D2 protein,* was found to have a single base pair deletion in the ISR42-8 allele and a two-base pair deletion in the Golf allele, resulting in 13 variant amino acids in the protein code. The second candidate is a *Peroxidase superfamily protein* (*HORVU6Hr1G076260*), typified by a single base pair deletion in the Golf allele, which results in 53 variant amino acids.

On chromosome 7H, a candidate locus in regard to plant height effect of drought stress was observed. The position of this locus is reported to be between 90 and 127 cM. The large locus of this QTL *Qhei.S42IL-7H* resulted in the identification of two candidate genes (Honsdorf et al., 2017). The *F-box protein* (*HORVU7Hr1G113340*) was observed to have an IAF of 56% in the organic and 12% in the conventional 23rd generation. The parental alleles varied based on six amino acid variations. The second candidate gene of this region is *HORVU7Hr1G097510,* an *ENTH/ANTH/VHS superfamily protein*. The IAF of this gene is 4% in the conventional and 26% in the organic system. The parental alleles were found to differ for 23 amino acids, based on a three-base pair deletion in the Golf allele.

Nutrient uptake or allocation-related genes were rarely described by association studies; therefore, the reference gene database was used additionally(Anonymous, n.d.). Nitrate transporter or reduction elements were identified on chromosomes 5H and 6H. The NRT3.1 gene was described to have a vital role in nitrate transportation and is located on position 647 million base pairs on chromosome 5H. The IAF of the 3rd generation was 12.8%. Compared to this generation, the IAF in the 23rd generation changed in both organic and conventional systems. The IAF increased to 18.3% in the organic 23rd generation, while it reduced to 7.8%in the conventional system. For the NRT2.6 nitrate transporter, a reduction of the IAF in the conventional system could be reported, as well. The 3rd generation was characterized by an IAF of 10.2%, while the conventional was observed to have an IAF of 1.1%. Contrasting, the organic system was typified by an IAF of 11.1%. A nitrate reductase is also located on the 6th chromosome at position 538 million base pairs. The IAF for this gene was positively selected in both environments to 22.1% and 27.7% in the conventional and organic systems, respectively (3rd generation IAF 14.9%). Another group of nutrient-related genes that shows relevant variations was observed for potassium. One gene on chromosome 2H and two genes on chromosome 3H were

identified. The potassium transporter on chromosome 2H was characterized by a negative selection of the ISR42-8 allele down to 1.1%, compared to a start value of 17.7% in the 3rd generation. Contrasting to this, the IAF in the 23rd generation of the organic system almost did not change (16.7%). The two potassium transporter genes on chromosome 3H are located at 655 and 667 million base pairs positions. These had the same selection pattern in common, starting with an IAF of around 4% in the 3rd generation and being positively selected in the organic system to 9.9 and 7.9%. Compared to the organic system, the conventional system was described by a negative selection down to ~ 2%.

Finally, two copper transporter genes were identified by a relevant change of the allele frequency over the generations. While the copper transporter on the short arm of chromosome 4H was observed to have an IAF of 7.6%, the IAF was increased in both environments. The conventional population was characterized by a 2-fold increase, while the organic population increased by 4-fold. The copper transporter on the short arm of chromosome 6H had a different pattern. While minor changes were observed in the organic system, the conventional population was described by a decrease of the allele frequency to 1% (3rd generation 9%).

Two other genes were also related to nutrients – a secondary metabolism-related ortholog from rice and a transporter-related gene on chromosome 6H(Takahashi et al., 2012). The transporter-related gene on chromosome 6H was observed to have an IAF increase of 2-fold in the organic population, while the conventional population remains unchanged. The rice ortholog from chromosome 2H had the opposite pattern. Despite the 2-fold increase in the conventional population, the organic system remained unchanged.

18 QTLs of relevance in this population were identified by dissecting known root-related QTL loci (Naz et al., 2014; Oyiga et al., 2020; Reinert et al., 2016). Most of these traits were related to tiller number, root length, dry weight, volume, root to shoot ratio or shoot dry weight. Ten QTLs have been identified with ISR42-8 used as an introgression line. Two of these genes were identified on chromosome 1H. The first gene is located at 15-28cM, related to root dry weight. For this large locus, four potential candidate genes were selected. All of these had a positive selection of the IAF in both systems compared to the 3rd generation in common. Furthermore, all these genes were described by deletions in the Golf allele of the respective gene. Gene *HORVU1Hr1G005230* is an undescribed protein with nine bp deletions in the Golf allele. The IAF is increased by two-fold in both farming systems compared to the 3rd generation. The next candidate gene on this locus is *HORVU1Hr1G009430*, a *Single-stranded nucleic acid-binding protein R3H*. A single base pair deletion in Golf might have caused the IAF change from 24% in the 3rd generation to 60% and 70% in the organic and conventional 23rd generation, respectively. The *Plant Tudor-like RNA-binding protein* gene *HORVU1Hr1G008510* was described by 18 base pair deletions in the Golf allele, resulting in a preliminary stop codon in this allele. The allele frequency changed from 17% in the 3rd generation to 45% and 64% in the organic and conventional systems, respectively. The last candidate found in this QTL locus is the *Potassium transporter family protein*-coding gene *HORVU1Hr1G009100*. A three-base pair deletion in Golf leads to 84 variant amino acids between the two parental alleles. The allele frequency increased from 11% in the 3rd generation to 52% and 62% in the organic and conventional environment, respectively. The next QTL found on chromosome 1H, *QRL.S42IL.1H.b*, located in the interval of 128-132 cM, is related to the root length. The potential candidate gene was *HORVU1Hr1G094070*, a *Cytoplasmic membrane protein*. Two base pair deletions in Golf resulted in 187 amino acid variations between the parental alleles. While the IAF of the conventional 23rd generation was reduced to 2%, the IAF is increased to 19% in the organic system.

One QTL for the root volume was identified on the 2nd chromosome. The locus *QRV.S42IL.2H* was placed in the region of 118 to 121 cM. The potential candidate gene is *HORVU2Hr1G111680,* an *Ethylene-responsive transcription factor 10.* Four base pair deletions in the ISR42-8 allele resulted in eleven amino acid variations. The observed IAF decreased from 14% in the 3rd generation to 1% in the conventional system while increasing to 23% in the organic system.

A QTL in regard to root dry weight was identified on the 3H chromosome at position 38-40 cM. The candidate gene of this locus, *HORVU3Hr1G015840, is a Serine/threonine-protein kinase* and was described by an IAF positive selection of 50% in the organic system, while the conventional system was described by a 66% reduction. The parental alleles of this gene differ by 309 amino acids, caused by a five base pair deletion in the ISR42-8 allele. Two overlapping loci for root length (*QRL.S42IL.4H.a*) and root dry weight (*Qrdw.S42IL.4H*) were identified for chromosome 4H at position 1-15 cM. The candidate gene *HORVU4Hr1G001420* is described as a *GDSL esterase/lipase*. A 21 base pair deletion in the ISR42-8 allele result in 365 variant amino acids between the parental alleles. The IAF was positively selected in both systems, with a more pronounced positive selection observed in the organic system (15% conventional, 19% organic). Two loci were observed on chromosome 5H, on position 52-58 cM (*QRL.S42IL.5H.a*, root lengt*h*), and on position 122-138 cM (*Qgh.S42IL.5H,* Growth Habit & Root dry weight & Root length & Root Volume & Tiller number). While the identification of a candidate gene failed for *QRL.S42IL.5H.a*, a *Pentatricopeptide repeat-containing protein* (*HORVU5Hr1G095110*) was identified for *Qgh.S42IL.5H*. While the IAF in the conventional system was decreased to 6%, the IAF was increased to 28% in the organic population (21% in the 3rd generation). Finally, two QTL were identified on chromosome 7H. QTL locus *Qrdw.S42IL.7H.a* at position 12-30 cM is related to root dry weight and root volume. Three candidate genes were identified for this locus. All of these have a positive selection in the organic population in common. The first candidate is the *Disease resistance protein RPM1* (*HORVU7Hr1G013300*), varying between the parental alleles in 14 amino acids. The IAF was observed to be 1% in the 3rd generation, increased to 5% in the conventional and 47% in the organic system. The second candidate is *HORVU7Hr1G022980*, an *exocyst subunit exo70 family protein G1*. A single base pair deletion in the Golf allele resulted in 18 variant amino acids. The IAF is 1% in the conventional, but 25% in the organic system. The last candidate gene of this locus was the auxin response factor 16 (*HORVU7Hr1G022980*). The parental alleles were variant by 57 amino acids, caused by three-base pair deletions. The organic IAF was 19%, while 1% in the conventional system. The other candidate locus on 7H, *Qrdw.S42IL.7H.b*, is located in the interval of 76-78 cM and associated with the root dry weight. Two candidate genes for this locus were found. The first, a *Leucine-rich receptor-like protein kinase family protein (HORVU7Hr1G085790*), was characterized by three variant amino acids. The starting IAF was 3%, increased to 12% and 28% in the conventional and organic systems of the 23rd generation, respectively. The second candidate is the *ABC transporter G family member 42* gene *HORVU7Hr1G088600*. The parental alleles vary by 1302 amino acids, based on a three-base pair deletion in the Golf allele. While the IAF remains unchanged in the conventional system, it increased by 4.5-fold in the organic system.

Six QTLs relevant in regard to yield components were observed to have variations in the allele frequency. The first QTL was identified on chromosome 1H, the Thresh-1 locus for grain threshability (Schmalenbach et al., 2011). It is located in the interval of 28 to 37 cM. Two candidate genes for this locus were identified. The first, the *WUSCHEL related homeobox 1* gene *HORVU1Hr1G010580*, was observed to have a significant deletion in the ISR42-8 allele, indicating that this allele was non-functional. Although this allele did not seem to be functional, an increase of the IAF was observed in both environments by a 2 to 3.5-fold change. The second candidate gene, *HORVU1Hr1G008360*, is a *laccase 7*. The variation between the alleles was six amino acids. The IAF was high in the 3rd generation

(49%) and increased to 59% in the conventional while being reduced in the organic system (39%). The second and third loci are related to the two brittleness genes Brt1 and Brt2 (Komatsuda et al., 2004). Both are closely located near the 48 million base pair position on chromosome 3H. Two candidate genes were identified for these two loci. For Brt1, the gene *HORVU3Hr1G018300*, a *Leucine-rich repeat receptor-like protein kinase family protein*, is characterized by 316 variant amino acids caused by a single base pair deletion in the Golf allele. The IAF of this gene is down to 0% in both systems. The candidate gene for Brt2 is *HORVU3Hr1G016990*, a *cellulose synthase like G3*, which was characterized by a 24 base pair deletion in Golf, resulted in 366 variant amino acids. The IAF for this gene was 1% in both environments. On chromosome 5H, the loci *QGL5H/QGW5H* were identified in the pericentromeric region (393-501 million base pairs) (Watt et al., 2019). This QTL locus is related to grain length and width. This large region was hard to reduce to a single candidate. Besides the gene related to the phenotype of increased grain length and width, multiple other candidate genes can be identified as relevant in this particular region. Four genes have been associated with this group. The first is *HORVU5Hr1G061160*, a *heat shock protein 21*. It was described by a six-base pair deletion in the ISR42-8 allele. The IAF of this gene is 0% in the 3rd generation, while increased to 45% and 54% in the conventional and organic 23rd generation, respectively. The second candidate gene is *HORVU5Hr1G057100*, a *BnaC03g66760D protein*. The parental alleles vary by a single base pair deletion in the Golf allele, resulting in 14 variant amino acids. The IAF pattern was identical to the prior gene. The third candidate gene is a *Protein kinase superfamily protein* (*HORVU5Hr1G061760)*, defined by ten variant amino acids. The IAF in the 3rd generation is 0%, but increased up to 44% and 40% in the conventional and organic system, respectively, towards the 23rd generation. Finally, the *Cytoplasmic tRNA 2-thiolation protein 1* (*HORVU5Hr1G061960*) was observed to have an IAF of 1% in the 3rd generation and 39% and 42% in the conventional and organic system. The gene varies between the parental alleles in 20 amino acids. The last reported QTL in regard to yield components was observed on chromosome 7H at position 134 cM(Liller et al., 2017). This QTL was observed to be relevant in the awn length. In both organic and conventional system, a low IAF throughout all generations was observed (IAF < 0.03). The corresponding candidate gene is a zinc finger protein 3 gene (*HORVU7Hr1G118490*). Contrasting to the low and similar IAF for the QTL region, the IAF for this gene was highly variant between the organic and conventional population. The IAF was observed to be 1% in the conventional and 54% in the organic 23rd generation. The candidate gene was observed to be variant in four amino acids.

Thirty-two QTLs were associated as relevant due to IAF changes of the generations or variations between the farming systems for yield physiological traits. The QTLs were found in 17 different association or QTL mapping studies. These QTLs are related to phenotypic traits like the count of tillers, plant height, flowering time, vernalization, or dormancy. The loci are spread all over the genome, detecting at least one locus per chromosome. On chromosome 1H, loci associated with tiller number, plant height, and leaf production were identified. The tiller-related QTL *Qtil.1H* candidate gene *HORVU1Hr1G091460* is a *sugar transporter 9*. Ten variant amino acids can be observed for the parental alleles(Reinert et al., 2016). While the IAF of the conventional system was reduced by 50%, the IAF in the organic system was increased by 80% compared to the 3rd generation. The *HvCMF10* locus, associated with plant height, might be related to *HORVU1Hr1G053610*. Its function was undescribed so far, but the protein sequence varies by 109 amino acids, caused by a single base pair deletion in the Golf allele. While the IAF was positively selected to 25% in the conventional system, it was negatively selected to 9% in the organic system. The *HvGA20Ox3/Qgh.S42IL.1H* locus is described to be related to *Tillering + LAI/Gibbereline Oxidase*(Marzec & Alqudah, 2018). The candidate gene *HORVU3Hr1G089980* is a *gibberellin 20 oxidase 2* with a two amino acid variation between the parental alleles. The IAF was positively selected in both systems by 4 to 6-fold. The *PpdH1* locus on chromosome 2H at 20cM is a *pseudo-response regulator 7* gene (*HORVU2Hr1G013400*). The gene

varied in 60 amino acids, based on a nine base pairs deletion in the Golf allele. The ISR42-8 allele is negatively selected to 0% in both systems. On chromosome 3H, three loci were found close to each other, all in the region of 100 to 110 cM. *HvHeading-3H-HA* locus was found to be associated with the candidate gene *HORVU3Hr1G089250*, an *Ethylene insensitive 3 family protein*. The protein was variant by 119 amino acids, based on a three-base pair deletion in the ISR42-8 allele. The IAF was not changed in the conventional but reduced by 3-fold in the organic system. The *HvGA20Ox3/Qgh.S42IL.1H* locus found in this region might be related to the candidate gene *HORVU3Hr1G089980*, a *gibberellin 20 oxidase 2*. This locus, related to tillering, showed substantial IAF variation between the generations. The conventional and the organic system were positively selected by 4 to 6-fold change compared to the 3rd generation. For the *sdw1/denso* locus, two candidate genes were found(X. Xu et al., 2018). While the *gibberellin 20-oxidase 3* gene *HORVU3Hr1G090980* indicated a homologous IAF increase in both systems to 45%, the second candidate is a *Bidirectional sugar transporter N3* with an increase of the IAF to 60% in both systems, compared to 3% in the 3rd generation. As the Oxidase only varies by two amino acids, but the sugar transporter is characterized by a single base pair deletion in the Golf allele, the real cause of the IAF is questionable. Chromosome 3H is the most relevant chromosome in terms of yield physiological traits, as 13 QTLs were identified on this chromosome. Besides the region around 100 cM, another hot spot of these QTLs was identified in the region of 40 to 60 cM. Four dwarfing or plant height loci were reported in this region. A candidate gene related to the QTL locus *HvHXK9* was reported as an undescribed protein(Alqudah et al., 2016). This gene (*HORVU3Hr1G065350*) was described by an IAF of 0% in both systems and a 15 base pair deletion in the Golf allele. Because of the size of this block and its strong linkage, the identification of further candidate genes was biased.

On chromosome 4H, three QTLs were identified. *HvPRR59* is a dormancy locus at position 54 cM(Alqudah et al., 2016). The IAF of the candidate gene *HORVU4Hr1G027970*, a *SelT/selW/selH selenoprotein domain-containing protein* is negatively selected to 0% in both environments. The gene allele of both parents differs by a six base pairs deletion in the Golf allele. The second candidate gene of this locus is an *Ethylene-overproduction protein 1*, also characterized by a negative selection in both systems and major structural variations in the protein sequence. Another important dormancy locus was reported for chromosome 5H. The locus SD1 is located at 489 million base pairs and interestingly directly neighboring the grain length locus *QGL5H/QGW5H.* No *coding DNA* alignment is available for the candidate gene *HORVU5Hr1G062990*, coding for an *alanine aminotransferase 2*(Sato et al., 2016). For all generations of both environments, starting from the 3rd to the 23rd, the IAF was 0% or close to 0%. The last candidate locus of this region is *HvD4* on 60 cM position (Alqudah et al., 2016). This locus is related to reduced plant height. The candidate gene of this locus is *HORVU4Hr1G067270*, a NAC domain protein. The protein sequence is variant by two amino acid changes between the two parents. A decrease of the IAF for the conventional system can be reported (15 IAF), while the organic system has shown a positive selection of this gene (28%). On chromosome 5H at position 43 to 50 cM, a QTL region related to the plant stature is reported. The candidate gene for the *HvBC12/GGD1* locus could be *HORVU5Hr1G048180*, a *2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein*. The ISR42-8 allele was classified by 13 base pair deletions, and the IAF was increased from 0% in the 3rd generation to 36% and 41% in the 23rd generation of the conventional and organic system, respectively.

Three dwarfing genes were identified on the 7H chromosome (Alqudah et al., 2016; Itoh et al., 2004). The candidate locus D35, associated with a dwarfing phenotype, was reported on position 76-79 cM. The candidate gene of this locus, a ribosomal protein-like gene (*HORVU7Hr1G088120*), was observed to have a six base pairs deletion in the ISR42-8 allele. Furthermore, the IAF pattern varies between the systems. While the IAF is positively selected from 3% in the 3rd generation to 20 % in the organic population, the increase in the conventional system was more marginal, up to 7%.

To abstract the information collected for all these QTL loci and candidate genes, the QTL loci were clustered in six groups – biotic resistance, including all resistance loci against fungal attacks; drought tolerance – any observed phenotypic reaction of genotypes when they were grown under reduced water availability conditions; nutrients – all genes or loci reported to be functional regarding nutrient uptake or transportation; root – all loci which have been found to be related to root length, volume, root to shoot ratio or comparable root related traits; yield components – all yield defining phenotypic characteristics like grain size or threshability; and finally yield physiology – all traits related to forming yield including plant height, flowering time. The results of this clustering were grouped for generations and environments. This gives a timeline where developmental steps or the evolution of the ISR42-8 related allele frequency can be visualized in a meaningful way.

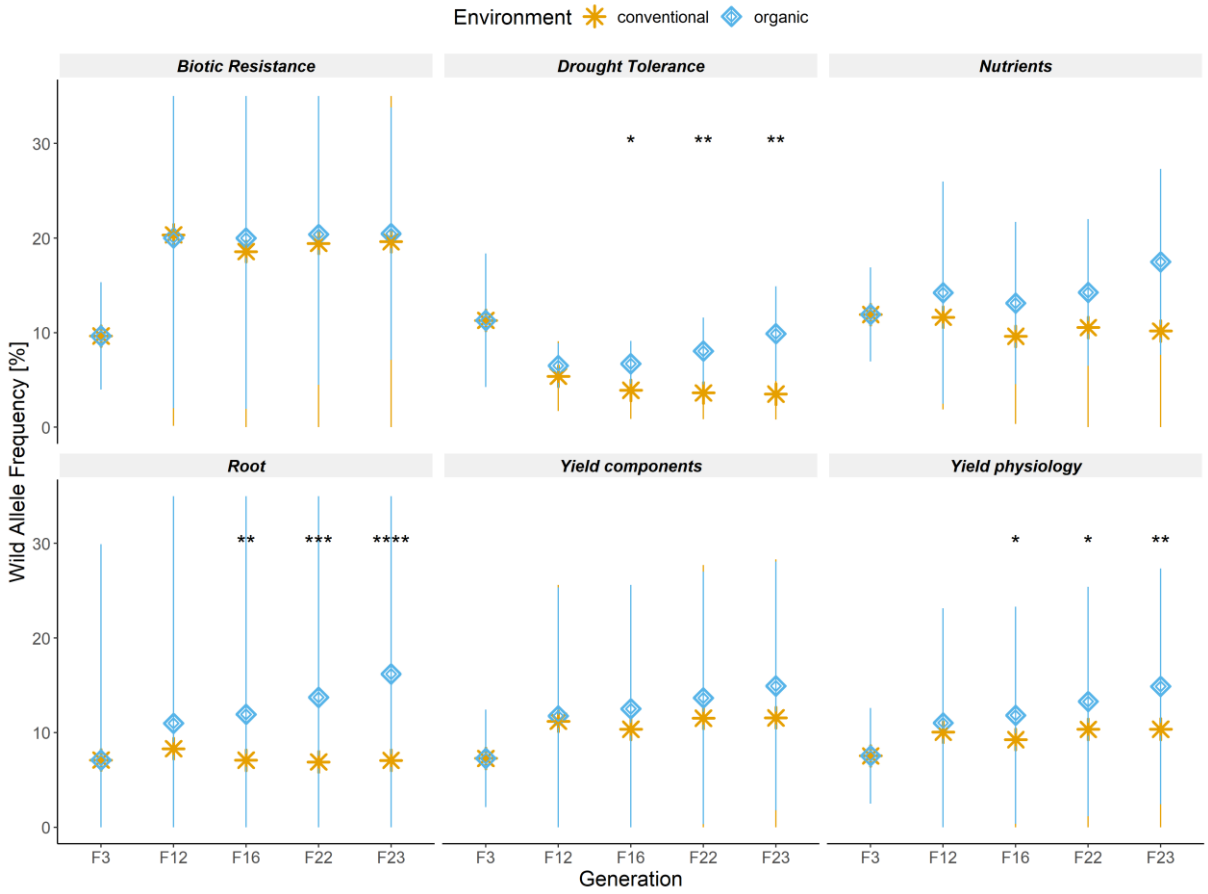


Figure 63: The wild donor allele frequencies in the QTL regions, clustered in six groups and plotted for each generation and environment. The y-axis illustrates the wild-type allele frequency in %, while the generations are plotted on the x-axis. The environments are separated by shape (organic, square; conventional, star). The QTLs were identified for various traits (e.g., root length and fusarium resistance). The QTLs were clustered into six groups: biotic stress (resistance genes), drought tolerance (reaction to drought), nutrients (QTLs or genes identified and related to the transport or uptake of nutrients), yield physiology (e.g., plant height and heading dates), root morphological traits (e.g., root length) and yield components (e.g., kernel size). Many of the analyzed studies used ISR42-8 as donor parent to create introgression lines. QTL clusters were compared generation-wise between the environments by a Mann-Whitney-U test to identify significant variations in the wtAF patterns.

As illustrated in figure 63, the ISR42-8 allele frequency in the populations changed in different manners, depending on the trait class and environment. While the average IAF in the 3rd generation for the observed biotic resistance genes was found to be 9.6%, the IAF increased in both organically and conventionally farmed systems comparably to a level of 20 - 23%. The average IAF of the organic groups was a bit higher, but the variance of both groups overlapped, indicating no statistical variation between the organic and conventional population for all generations by the 12th and following. This

observation could indicate that the resistance alleles were selected fast in both environments and remained on a plateau after this. As these are not down or positively selected anymore in the later generations, one could suspect that the resistance against the fungi was still unbroken, but the physiological costs of this resistance might be high, and therefore, a further positive selection is prohibited by the reduced fitness of genotypes with a functional resistance allele in years without high disease pressure. A variant pattern between the environments was observed for the drought tolerance-related QTL loci. While the IAF was 11% in the 3rd generation, it decreased to 5% in the conventional and 6.7% in the organic 12th generation. In the successive generations, the IAF decreased further in the conventional system to 3.5% in the 23rd generation. Contrasting to this development, the IAF increased in the 16th, 22nd, and 23rd generation to levels to reach 9.9% in the last generation tested. A similar pattern was observed for the nutrients. The average IAF of the QTL loci related to nutrients is 11.5%. In the 12th generation, little variation from this value was observed, resulting in 10.6% and 11.6% IAF in the conventional and organic environments, respectively. The variation between the generation became more distinct in the 16th generation (8.7 to 12.2%), reaching the highest level of variation in the 23rd generation (9.5, 15.7%). Generally, a slow but constant increase in the IAF of the nutrient-related genes was observed in the organic population. The root-related QTLs were found to have an IAF of 8.2% in the 3rd generation. The IAF was increased to 9.3% and 11.7% in the 12th generation in the conventional and organic systems, respectively. With every subsequent generation, the IAF of the root-related QTLs is further increased in the organic population. In the 23rd generation, the IAF can be observed at 16.5%. Contrasting to this evolution, the IAF remained on the same level overall generations in the conventional environment. The IAF evolution pattern of the yield components was similar to the one observed for the biotic resistance. The IAF in the 3rd generation is 5.3%. In all following generations, the IAF ranges between 15% and 19%. For the 12th and 16th generations, no significant variation was observed between the organic and conventional systems. In the 22nd and 23rd generations, the two systems became more distinct. While the IAF in the organic system was slightly reduced to 15.2%, the conventional system was defined by a stable IAF of 19%. The last of the six groups was related to yield physiology. Compared to the IAF of the 3rd generation, both systems showed and positive selection. While the IAF in the 3rd generation was 7%, the IAF in the 12th generation was 10.6% and 11.6% for the conventional and organic systems, respectively. Even though both systems were close together in this generation, the variation between them became more and more distinct with each subsequent generation. In the 23rd generation, the IAF for the organic system was 4.5% higher compared to the conventional system (15.6%). Concluding this, apparent variations and similarities were identified between the two farming approaches in terms of agronomically relevant traits.

For most of the reported QTLs, single candidate genes could be identified. Details on the candidate QTL loci can be found in table 16. The corresponding candidate genes were presented in table 17.

Table 16 (following pages): The allele frequency of the 3rd and 23rd generation for selected QTL regions identified by meta-analysis. QTL studies related to fungal resistance genes (Bedawy et al., 2018; Dragan et al., 2013; Lüpken et al., 2014; M. Von Korff et al., 2005)*, grain yield* (Komatsuda et al., 2004; Liller et al., 2017; Schmalenbach et al., 2011; Watt et al., 2019)*, nutrient uptake and translocation* (Anonymous, n.d.; Takahashi et al., 2012)*, upper body yield physiology, such as flowering, dormancy or dwarfing genes* (Afsharyan et al., 2020; Alqudah et al., 2016; Anonymous, n.d.; Guan et al., 2012; Itoh et al., 2004; Marzec & Alqudah, 2018; Mulki et al., 2018; Sato et al., 2016; Wang et al., 2010c; Y. Xu et al., 2017; Yan et al., 2004)*, associated root traits* (Naz et al., 2014; Oyiga et al., 2020; Reinert et al., 2016) *and drought resistance* (Anonymous, n.d.; Honsdorf et al., 2014; Mohammed & Léon, 2004; Muzammil et al., 2018) *were identified and classified by color. Besides the QTL name, the position, map source, category, reference publication of the QTL, actual trait, and crossing background are reported. This information is linked to the wild-type allele frequency of the F_3 and the F_{23} generations of both the organic and conventional systems for the region of the QTL. In general, a specified QTL region was regarded as relevant and listed in the table if a significant variation of the generations or environments was observed in the pooled sample data.*

Gene/QTL	***Position***	***Map Source***	***Category***	***Reference***	***Trait***	***Cross***	***wtAF conv F23***	***wtAF org F23***	***wtAF F3***
Qpm.S42-1H.a	*1H 0-28cM/1-120mio bp*	*98 SSR Markers*	*Resistance*	*Von Korf 2005*	*Powerdy Mildew/Mla*	*Scarlett x ISR 42-8*	*0.256*	*0.214*	*0.148*
Qtil.1H	*1H 118-127 cM*	*9KiSelect*	*Yield physiology*	*Reinert 2016*	*Tiller number*	*Assosiation Panel*	*0.038*	*0.070*	*0.130*
QRL.S42IL.1H.b	*1H 128-132 cM (199-205 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root length*	*Scarlett x ISR 42-8*	*0.038*	*0.127*	*0.087*
Qrdw.S42IL.1H	*1H 15-28 cM (24-40 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root dry weight*	*Scarlett x ISR 42-8*	*0.458*	*0.351*	*0.172*
SCRI-RS239784	*1H 30 cM*	*9KiSelect*	*Resistance*	*Bedawy 2018*	*Fusarium graminearum resistance*	*Assosiation Panel*	*0.633*	*0.480*	*0.217*
Thresh-1	*1H 28 - 37 cM (54 cM BOPA1)*	*ref Genome*	*Yield components*	*Schmalenbach 2011*	*Thresability*	*Scarlett x ISR 42-8*	*0.581*	*0.437*	*0.172*
HvCMF10	*1H 47.5-48 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Plant Height*	*Assosiation Panel*	*0.164*	*0.070*	*0.122*
P5CS1	*1H 500mio bp*	*ref Genome*	*drought tolerance*	*Muzammil 2018*	*Proline accumulation*	*Scarlett x ISR 42-8*	*0.035*	*0.090*	*0.141*
Qgh.S42IL.1H	*1H 86-101 cM (134-162 cM BOPA1)*	*BOPA1*	*Yield physiology*	*Naz 2014*	*Growth Habit & Tiller number*	*Scarlett x ISR 42-8*	*0.016*	*0.140*	*0.13513*
HvGA20Ox4	*1H 94 cM*	*9KiSelect*	*Yield physiology*	*Marzrc 2018*	*Tillering + LAI/Gibbereline Oxidase*	*Assosiation Panel*	*0.007*	*0.170*	*0.156*
-	*1H 95.9-97.9 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2018*	*LAI at pre anthesis stage*	*Assosiation Panel*	*0.002*	*0.085*	*0.120*
Qsdw.2H.b	*2H 104-111 cM*	*9KiSelect*	*Root System*	*Reinert 2016*	*Shoot dry weight*	*Assosiation Panel*	*0.010*	*0.153*	*0.131*
QRS.2H	*2H 105-108 cM*	*9KiSelect*	*Root System*	*Reinert 2016*	*Root-Shoot Ratio & Root dry weight*	*Assosiation Panel*	*0.010*	*0.136*	*0.126*

QRV.S42IL.2H	*2H 118-121 cM (197-206 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root Volume*	*Scarlett x ISR 42-8*	*0.013*	*0.193*	*0.063*
Qpm.S42-2H.c	*2H 139-146 cM/715-720mio bp*	*98 SSR Markers*	*Resistance*	*Von Korf 2005*	*Powerdy Mildew & CHS/M1La*	*Scarlett x ISR 42-8*	*0.006*	*0.126*	*0.086*
PpdH1	*2H 20 cM/29 mio bp*	*ref Genome*	*Yield physiology*	*Deng 2007*	*Photoperiod sensitivity – both parents sensitive type*	*-*	*0.002*	*0.039*	*0.049*
Qhei.S42IL-2H	*2H 38-58 cM (63-110 cM BOPA1)*	*BOPA1*	*drought tolerance*	*Honsdorf 2014/Naz 2014*	*Plant Height/Growth Habit*	*Scarlett x ISR 42-8*	*0.007*	*0.047*	*0.243*
	2H 55mio bp	*ref Genome*	*Nutrient uptake*	*IPK database*	*Potassium transporter*	*-*	*0.011*	*0.167*	*0.177*
-	*2H 6.5 -9 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*productive Tiller number*	*Assosiation Panel*	*0.082*	*0.141*	*0.036*
-	*2H 677mio bp (675-680)*	*ref Genome*	*Nutrient uptake*	*Takahashi 2012*	*secondary metabolism*	*Rice*	*0.351*	*0.199*	*0.171*
-	*2H 73.7-83.8 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*productive Tiller number/Vrs 1*	*Assosiation Panel*	*0.358*	*0.220*	*0.222*
Ryd2	*3H 100-130 cM*	*9KiSelect*	*Resistance*	*Petcovic 2013*	*Barley Yellow Dwarf Virus Resistance*	*MBR1012 x Scarlett*	*0.238*	*0.280*	*0.064*
HvHeading-3H-HA	*3H 100.71-109.84 cM*	*9KiSelect*	*Yield physiology*	*Afsharyan 2020*	*Flowering*	*MAGIC Population*	*0.325*	*0.332*	*0.081*
HvGA20Ox3	*3H 106 cM*	*9KiSelect*	*Yield physiology*	*Marzrc 2018*	*Tillering/Gibbereline Oxidase 20*	*Assosiation Panel*	*0.258*	*0.221*	*0.089*
sdw1	*3H 108-110*	*9KiSelect*	*Yield physiology*	*Kuczyńska 2012*	*Semi dwarf 1/denso*	*-*	*0.506*	*0.534*	*0.076*
Flow Loc T3	*3H 119mio bp*	*ref Genome*	*Yield physiology*	*Mulki 2018*	*Flowering*	*Ubi::HvFT3(Golden Promix) x Igri*	*0.001*	*0.004*	*0.052*
Qrdw.3H	*3H 120-124 cM*	*9KiSelect*	*Root System*	*Reinert 2016*	*Root dry weight*	*Assosiation Panel*	*0.026*	*0.188*	*0.053*

Qtil.S42IL-3H	*3H 122-154 cM (204-255 cM BOPA1)*	*BOPA1*	*drought tolerance*	*Honsdorf 2014*	*Tiller number & Tiller & Biomass*	*Scarlett x ISR 42-8*	*0.040*	*0.119*	*0.041*
-	*3H 137-139 cM*	*9KiSelect*	*Yield physiology*	*IPK database*	*Dwarf 5*		*0.032*	*0.105*	*0.044*
QTL3H_4	*3H 138-148 cM*	*9KiSelect*	*Root & Shoot System*	*Oyiga 2019*	*main shoot axis nodal root cortical area*	*Assosiation Panel*	*0.061*	*0.157*	*0.038*
Hv13GEIII/ Bmac 0357	*3H 152-167cM/550-601mio bp*	*97 SSR Markers*	*drought tolerance*	*Hamam 2004*	*Yield components*	*Scarlett x ISR 42-8*	*0.102*	*0.165*	*0.073*
QTL3H_1	*3H 35-38 cM*	*9KiSelect*	*Yield physiology*	*Oyiga 2019*	*Nodal root system size & plant height & harvest index & grain Yield components*	*Assosiation Panel*	*0.017*	*0.160*	*0.082*
Qrdw.S42IL.3H	*3H 38-40 cM (64-66 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root dry weight*	*Scarlett x ISR 42-8*	*0.014*	*0.083*	*0.060*
HvFT2	*3H 43-47 cM*	*9KiSelect*	*Yield physiology*	*Wang 2010*	*Dwarf 2*		*0.006*	*0.057*	*0.062*
Qhei.S42IL-3H	*3H 43-54 cM (63-98 cM BOPA1)*	*BOPA1*	*drought tolerance*	*Honsdorf 2014*	*Plant Height*	*Scarlett x ISR 42-8*	*0.012*	*0.115*	*0.057*
-	*3H 44.3-46.2 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Plant Stature*	*Assosiation Panel*	*0.008*	*0.057*	*0.058*
Brt2	*3H 40-50 mio bp*	*ref Genome/99 RILs*	*Yield components*	*Komatsuda 2004*	*Ear brittleness*	*Assosiation Panel*	*0.003*	*0.021*	*0.052*
Brt1	*3H 40-50 mio bp/48mio bp*	*ref Genome/99 RILs*	*Yield components*	*Komatsuda 2004*	*Ear brittleness*	*Assosiation Panel*	*0.003*	*0.021*	*0.052*
Dwarf61	*3H 50-53 cM*	*9KiSelect*	*Root & Shoot System*	*Oyiga 2019*	*Plant height & semi-dwarf single recessive gene uzu gene*	*Assosiation Panel*	*0.013917068*	*0.14143*	*0.05643*

QRS.3H	*3H 51-53 cM*	*9KiSelect*	*Root System*	*Reinert 2016*	*Root-Shoot Ratio*	*Assosiation Panel*	*0.014*	*0.147*	*0.057*
HvHXK9	*3H 54-59 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Plant Height*	*Assosiation Panel*	*0.011*	*0.100*	*0.040*
QTL3H_3	*3H 56-61 cM*	*9KiSelect*	*Yield physiology*	*Oyiga 2019*	*Flowering/Dwarf 61*	*Assosiation Panel*	*0.015*	*0.120*	*0.029*
-	*3H 62-64 cM*	*9KiSelect*	*Yield physiology*	*Guan 2012*	*Dwarf 10*	*Maize*	*0.015*	*0.134*	*0.035*
Denso	*3H 634mio bp (630-640)*	*ref Genome*	*Yield physiology*	*Xu 2017*	*Plant Height/Gibbereline Oxidase*	*-*	*0.513*	*0.535*	*0.073*
-	*3H 655mio bp (650-660)*	*ref Genome*	*Nutrient uptake*	*IPK database*	*Potassium transporter*	*-*	*0.019*	*0.099*	*0.045*
-	*3H 655mio bp (650-660)*	*ref Genome*	*Yield physiology*	*IPK database*	*Chloroplast activity*	*-*	*0.019*	*0.099*	*0.045*
-	*3H 667mio bp*	*ref Genome*	*Nutrient uptake*	*IPK database*	*Potassium transporter*	*-*	*0.027*	*0.079*	*0.038*
Qrdw.S42IL.4H	*4H 1-14 cM (5-27 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root dry weight*	*Scarlett x ISR 42-8*	*0.153*	*0.195*	*0.075*
QRL.S42IL.4H.a	*4H 1-15 cM (5-26 cM)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root length*	*Scarlett x ISR 42-8*	*0.110*	*0.164*	*0.080*
VrnH2	*4H 100 cM (95-105cM Region)*		*Yield physiology*	*Yan 2004*	*Vernalisation & Flowering*	*-*	*0.026*	*0.075*	*0.135*
Qtil.S42IL-4H	*4H 111 cM (171-183 cM BOPA1)*	*BOPA1*	*drought tolerance*	*Honsdorf 2014*	*Tiller & Height integral*	*Scarlett x ISR 42-8*	*0.047*	*0.039*	*0.142*
Qlr.S42-4H.a	*4H 125-132 cM/613-620mio bp*	*98 SSR Markers*	*Resistance*	*Von Korf 2005*	*Leaf rust*	*Scarlett x ISR 42-8*	*0.023*	*0.107*	*0.098*
Ebmac 0701	*4H 130-150 cM/613mio bp*	*97 SSR Markers*	*drought tolerance*	*Hamam 2004*	*relative leaf water content*	*Scarlett x ISR 42-8*	*0.023*	*0.012*	*0.062*

HvPRR59	*4H 51 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Dormancy*	*Assosiation Panel*	*0.005*	*0.029*	*0.098*
Qhei.S42IL-4H	*4H 52-81 cM (74-119 cM BOPA1)*	*BOPA1*	*drought tolerance*	*Honsdorf 2014*	*Plant Height*	*Scarlett x ISR 42-8*	*0.010*	*0.118*	*0.116*
	4H 52,000 bp	*ref Genome*	*Nutrient uptake*	*IPK database*	*copper ion binding*	*-*	*0.152*	*0.263*	*0.076*
HvD4	*4H 59.6-59.8 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Plant Height/reduced plant height*	*Assosiation Panel*	*0.012*	*0.150*	*0.112*
Qgh.S42IL.5H	*5H 122-138 cM (204-231 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Growth Habit & Root dry weight & Root length & Root Volume & Tiller number*	*Scarlett x ISR 42-8*	*0.043*	*0.131*	*0.150*
-	*5H 143.7-146.1 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Photoperiodic tillering*	*Assosiation Panel*	*0.116*	*0.167*	*0.102*
QTL5H_1	*5H 23-26 cM*	*9KiSelect*	*Root & Shoot System*	*Oyiga 2019*	*Tiller nodal root xylem vessel area*	*Assosiation Panel*	*0.024*	*0.123*	*0.010*
QGL5H/QGW5H	*5H 393–501mio bp*	*712 DArT marker*	*Yield components*	*Watt 2019*	*Grain length/Grain width*	*Vlamingh x Buloke*	*0.331*	*0.269*	*0.005*
-	*5H 43.7-50 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Plant Stature*	*Assosiation Panel*	*0.163*	*0.190*	*0.003*
HvBC12/GGD1	*5H 46-47 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Aspargine Synthase/ Gibbereline Deficit DWARF*	*Assosiation Panel*	*0.352*	*0.232*	*0.002*
Bmag 0357	*5H 47 cM/580mio bp*	*97 SSR Markers*	*drought tolerance*	*Hamam 2004*	*leaf per tiller & Yield components & Biomass*	*Scarlett x ISR 42-8*	*0.098*	*0.086*	*0.183*
SD1	*5H 488-510 Mio bp*	*ref Genome*	*Yield physiology*	*Sato 2016*	*seed dormancy*	*-*	*0.023*	*0.034*	*0.016*
QRL.S42IL.5H.a	*5H 52-58 cM (106-109 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root length*	*Scarlett x ISR 42-8*	*0.140*	*0.148*	*0.006*
NRT3.1	*5H 647mio bp*	*ref Genome*	*Nutrient uptake*	*IPK database*	*Nitrate transport*	*-*	*0.078*	*0.183*	*0.128*

NRT2.6	*6H 13mio bp*	*ref Genome*	*Nutrient uptake*	*IPK database*	*Nitrate transport*	*-*	*0.011*	*0.111*	*0.102*
-	*6H 16 Mio bp*	*ref Genome*	*Nutrient uptake*	*IPK database*	*copper transporters*	*-*	*0.008*	*0.104*	*0.090*
Qagri.S42IL-6H	*6H 49-76 cM (74-134 cM BOPA1)*	*BOPA1*	*drought tolerance*	*Honsdorf 2014*	*Biomass & Growth Plant Height & Shoot Area & WUE*	*Scarlett x ISR 42-8*	*0.061*	*0.124*	*0.106*
Ryd3	*6H 50 cM*	*Marker Set*	*Resistance*	*Lübken 2014*	*Barley Yellow Dwarf Virus*	*L94 x L94-QTL3*	*0.228*	*0.198*	*0.100*
-	*6H 531mio bp*	*ref Genome*	*Nutrient uptake*	*IPK database*	*Transport*	*-*	*0.131*	*0.276*	*0.142*
Nitrate Reductase 1	*6H 538mio bp*	*ref Genome*	*Nutrient uptake*	*IPK database*	*Nitrate reduction*	*-*	*0.221*	*0.277*	*0.149*
-	*6H 9cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*Tillering*	*Assosiation Panel*	*0.017*	*0.050*	*0.115*
Qrdw.S42IL.7H.a	*7H 12-30 cM (17-45 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root dry weight & Root Volume*	*Scarlett x ISR 42-8*	*0.013*	*0.118*	*0.048*
AL7.1	*7H 133.9 cM*	*384 BOPA Markers*	*Yield components*	*Liller 2017*	*Awn length*	*Morex x Wild-types*	*0.015*	*0.029*	*0.003*
HvBRD2	*7H 140 cM*	*9KiSelect*	*Yield physiology*	*Alqudah 2016*	*DWARF2/Tillering*	*Assosiation Panel*	*0.014*	*0.116*	*0.001*
QLR.S42-7H.a	*7H 166-181 cM/637-655mio bp*	*98 SSR Markers*	*Resistance*	*Von Korf 2005*	*Leaf rust*	*Scarlett x ISR 42-8*	*0.017*	*0.055*	*0.002*
QRI.7H	*7H 3-4 cM*	*9KiSelect*	*Root System*	*Reinert 2016*	*Root length*	*Assosiation Panel*	*0.019*	*0.141*	*0.012*
Qrdw.S42IL.7H.b	*7H 76-78 cM (119-121 cM BOPA1)*	*BOPA1*	*Root & Shoot System*	*Naz 2014*	*Root dry weight*	*Scarlett x ISR 42-8*	*0.073*	*0.149*	*0.038*
D35	*7H 76-79 cM*	*9KiSelect*	*Yield physiology*	*Itoh 2004*	*Dwarf 35*	*Rice*	*0.073*	*0.143*	*0.038*

Qhei.S42IL-7H	*7H 90-127 cM (134-193 cM BOPA1)*	*BOPA1*	*drought tolerance*	*Honsdorf 2014*	*Plant Height*	*Scarlett x ISR 42-8*	*0.014*	*0.126*	*0.025*

Table 17: Identified candidate genes related to the QTL region of Table 16. Besides the associated QTL, the structural variation observed on amino acid and/or DNA level are reported, as well as the function of this gene, the gene name in the reference, the wild-type allele frequency for the 23rd generation in conventional (wtAFcon F23) and organic system are shown. Furthermore, the wild-type allele frequency of the third generation, the SNP count and read count of the haplotype, and the information of preliminary stop codons are reported.

QTL	Variation	Function	Candidate Gene	wtAF con F23	wtAF org F23	Wt AF F3	SNP count	Read Coverage	PrelStop Codon
1H95-98	*122 AS variant; 14 bp deletion Golf*	*undescribed protein*	*HORVU1Hr1G078120*	*0.00*	*0.19*	*0.12*	*21*	*363*	*Golf*
2H677	*299 AS variant; 12 bp deletion Golf*	*Peroxisome biogenesis protein 19-2*	*HORVU2Hr1G097360*	*0.38*	*0.22*	*0.26*	*32*	*512*	*no*
2H6-9	*3 AS variant*	*Starch synthase 2, chloroplastic/amyloplastic*	*HORVU2Hr1G090980*	*0.41*	*0.24*	*0.39*	*173*	*2864*	*no*
2H6-9	*346 AS variant; 3 bp deletion Golf*	*auxin response factor 1*	*HORVU2Hr1G076920*	*0.00*	*0.00*	*0.47*	*5*	*60*	*no*
2H73-84	*25 AS variant; 13 bp deletion ISR42-8*	*GRAS transcription factor*	*HORVU2Hr1G094020*	*0.43*	*0.27*	*0.06*	*149*	*2261*	*no*
3H44-46	*6 AS variant; 1 bp deletion ISR42-8*	*Diacylglycerol kinase*	*HORVU3Hr1G019210*	*0.02*	*0.20*	*0.09*	*514*	*8090*	*no*
3H62-64	*48 AS variant; 30 bp deletion Golf*	*early nodulin-like protein 1*	*HORVU3Hr1G071370*	*0.02*	*0.22*	*0.02*	*83*	*1315*	*no*

3H62-64	*7 AS variant*	*Protein NRT1/ PTR FAMILY 4.3*	*HORVU3Hr1G071510*	*0.03*	*0.31*	*0.08*	*36*	*525*	*no*
5H143-147	*163 AS variant; 3 bp deletion ISR42-8*	*Ethylene-responsive transcription factor 10*	*HORVU5Hr1G111550*	*0.06*	*0.30*	*0.09*	*43*	*802*	*no*
5H43-50	*21 AS variant; 1 bp deletion ISR42-8*	*Plant/F1M20-13 protein*	*HORVU5Hr1G048410*	*0.36*	*0.33*	*0.00*	*68*	*1116*	*no*
5H43-50	*150 AS variant; 6 bp deletion ISR42-8; 12 bp deletion Golf*	*Auxin-responsive protein IAA30*	*HORVU5Hr1G014300*	*0.04*	*0.21*	*0.00*	*120*	*2081*	*no*
5H43-50	*17 AS variant; 1 bp deletion Golf*	*Phosphatidylinositol-glycan biosynthesis class X protein*	*HORVU5Hr1G016340*	*0.05*	*0.30*	*0.00*	*15*	*235*	*no*
6H520_540	*87 AS variant; 3 bp deletion ISR42-8*	*EPIDERMAL PATTERNING FACTOR-like protein 1*	*HORVU6Hr1G077840*	*0.18*	*0.28*	*0.09*	*58*	*1175*	*no*
7H1-3	*8 AS variant*	*MLO-like protein 1*	*HORVU7Hr1G002390*	*0.03*	*0.26*	*0.01*	*21*	*376*	*no*
7H84bp	*2 AS variant*	*unknown function*	*HORVU7Hr1G036880*	*0.05*	*0.22*	*0.05*	*36*	*604*	*no*
AL7.1	*4 AS variant*	*zinc finger protein 3*	*HORVU7Hr1G118490*	*0.01*	*0.54*	*0.00*	*56*	*1594*	*no*
Bmag 0357	*638 AS variant; 6 bp deletion Golf*	*DCD (Development and Cell Death) domain protein*	*HORVU5Hr1G091160*	*0.15*	*0.31*	*0.24*	*167*	*2580*	*no*

Bmag 0357	*194 AS variant; 12 bp deletion ISR42-8*	*Mitochondrial glycoprotein family protein*	*HORVU5Hr1G088460*	*0.12*	*0.36*	*0.07*	*42*	*758*	*no*
Brt1	*316 AS variant; 1 bp deletion Golf*	*Leucine-rich repeat receptor-like protein kinase family protein*	*HORVU3Hr1G018300*	*0.00*	*0.00*	*0.06*	*151*	*2067*	*Golf*
Brt2	*366 AS variant; 24 bp deletion Golf*	*cellulose synthase like G3*	*HORVU3Hr1G016990*	*0.01*	*0.01*	*0.06*	*118*	*1993*	*no*
D35	*19 AS variant; 6 bp deletion ISR42-8*	*Ribosomal protein-like*	*HORVU7Hr1G088120*	*0.07*	*0.20*	*0.03*	*678*	*11387*	*no*
Denso /sdw1	*2 AS variant*	*gibberellin 20-oxidase 3*	*HORVU3Hr1G090980*	*0.46*	*0.42*	*0.07*	*187*	*3300*	*no*
Denso /sdw1	*8 AS variant; 1 bp deletion Golf*	*Bidirectional sugar transporter N3*	*HORVU3Hr1G091230*	*0.59*	*0.63*	*0.03*	*43*	*732*	*no*
Dwarf61	*418 AS variant; 36 bp deletion Golf*	*calmodulin-binding family protein*	*HORVU3Hr1G049640*	*0.02*	*0.17*	*0.06*	*99*	*1632*	*no*
Ebmac 0701	*292 AS variant; 5 bp deletion ISR42-8; 3 bp deletion Golf*	*Ras-related protein Rab-6A*	*HORVU4Hr1G078690*	*0.05*	*0.23*	*0.05*	*62*	*1134*	*no*
Flow Loc T3	*873 AS variant; 6 bp deletion ISR42-8*	*ABC transporter family protein*	*HORVU3Hr1G027580*	*0.00*	*0.00*	*0.03*	*30*	*543*	*no*

Hv13GEIII/ Bmac 0357	*71 AS variant; 3 bp deletion ISR42-8*	*Y-family DNA polymerase H*	*HORVU3Hr1G072920*	*0.02*	*0.23*	*0.00*	*233*	*3513*	*no*
HvBC12/GGD1	*21 AS variant; 13 bp deletion ISR42-8*	*2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein*	*HORVU5Hr1G048180*	*0.36*	*0.41*	*0.00*	*21*	*357*	*no*
HvBRD2	*3 AS variant*	*Flavin-containing monooxygenase family protein*	*HORVU7Hr1G121260*	*0.01*	*0.22*	*0.00*	*36*	*591*	*no*
HvCMF10	*109 AS variant; 1 bp deletion Golf*	*unknown function*	*HORVU1Hr1G053610*	*0.25*	*0.09*	*0.17*	*123*	*1853*	*Golf*
HvD4	*2 AS variant*	*NAC domain protein,*	*HORVU4Hr1G067270*	*0.01*	*0.28*	*0.13*	*84*	*1309*	*no*
HvGA20Ox3/Qgh.S42IL.1 H	*2 AS variant*	*gibberellin 20 oxidase 2*	*HORVU3Hr1G089980*	*0.38*	*0.57*	*0.10*	*80*	*1196*	*no*
HvHeading-3H-HA	*119 AS variant; 3 bp deletion ISR42-8*	*Ethylene insensitive 3 family protein*	*HORVU3Hr1G089250*	*0.12*	*0.05*	*0.14*	*69*	*1133*	*no*
HvHXK9	*96 AS variant; 15 bp deletion Golf*	*unknown function*	*HORVU3Hr1G065350*	*0.00*	*0.00*	*0.01*	*43*	*715*	*no*
HvPRR59	*934 AS variant; 6 bp deletion Golf*	*SelT/selW/selH selenoprotein domain containing protein*	*HORVU4Hr1G027970*	*0.00*	*0.01*	*0.05*	*236*	*3984*	*no*

HvPRR59	*904 AS variant; 54 bp deletion ISR42-8; 55 bp deletion Golf*	*Ethylene-overproduction protein 1*	*HORVU4Hr1G055900*	*0.01*	*0.02*	*0.10*	*252*	*3745*	*Golf*
NRT2.6	*2 AS variant*	*high affinity nitrate transporter 2.6*	*HORVU6Hr1G005930*	*0.02*	*0.20*	*0.08*	*104*	*1649*	*no*
NRT3.1	*6 AS variant*	*unknown function*	*HORVU5Hr1G116240*	*0.08*	*0.43*	*0.00*	*28*	*496*	*ISR*
PpdH1	*69 AS variant; 9 bp deletion ISR42-8*	*pseudo-response regulator 7*	*HORVU2Hr1G013400*	*0.00*	*0.01*	*0.04*	*87*	*1656*	*no*
Qagri.S42IL-6H	*13 AS variant; 1 bp deletion ISR42-8; 2 bp deletion Golf*	*Photosystem II D2 protein*	*HORVU6Hr1G046850*	*0.02*	*0.24*	*0.07*	*3*	*52*	*no*
Qagri.S42IL-6H	*53 AS variant; 1 bp deletion Golf*	*Peroxidase superfamily protein*	*HORVU6Hr1G076260*	*0.03*	*0.24*	*0.19*	*201*	*3536*	*no*
Qgh.S42IL.5H	*2 AS variant*	*Pentatricopeptide repeat-containing protein*	*HORVU5Hr1G095110*	*0.06*	*0.28*	*0.21*	*19*	*313*	*no*
Qgh.S42IL.5H	*64 AS variant; 11 bp deletion Golf*	*Pentatricopeptide repeat-containing protein*	*HORVU5Hr1G100670*	*0.04*	*0.28*	*0.07*	*30*	*536*	*Golf*
Qgh.S42IL.5H	*47 AS variant; 1 bp deletion Golf*	*Pentatricopeptide repeat-containing protein*	*HORVU5Hr1G094560*	*0.02*	*0.27*	*0.19*	*63*	*1074*	*no*
QGL5H/QGW5H	*123 AS variant; 6 bp deletion ISR42-8*	*heat shock protein 21*	*HORVU5Hr1G061160*	*0.45*	*0.54*	*0.00*	*50*	*789*	*no*

QGL5H/QGW5H	*14 AS variant; 1 bp deletion Golf*	*BnaC03g66760D protein*	*HORVU5Hr1G057100*	*0.45*	*0.53*	*0.00*	*65*	*1188*	*no*
QGL5H/QGW5H	*10 AS variant*	*Protein kinase superfamily protein*	*HORVU5Hr1G061760*	0.44	0.40	0.00	240	3894	*no*
QGL5H/QGW5H	*20 AS variant*	*Cytoplasmic tRNA 2-thiolation protein 1*	*HORVU5Hr1G061960*	0.39	0.42	0.01	18	256	*no*
Qhei.S42IL-2H	*116 AS variant; 18 bp deletion ISR42-8*	*WUSCHEL related homeobox 12*	*HORVU2Hr1G017270*	*0.01*	*0.27*	*0.16*	*74*	*1085*	*no*
Qhei.S42IL-2H	*420 AS variant; 6 bp deletion ISR42-8*	*Rho GTPase-activating protein 3*	*HORVU2Hr1G094320*	*0.31*	*0.23*	*0.10*	*25*	*389*	*no*
Qhei.S42IL-2H	*75 AS variant; 1 bp deletion Golf*	*ATP synthase subunit b*	*HORVU2Hr1G096060*	*0.45*	*0.46*	*0.04*	*200*	*3206*	*Golf*
Qhei.S42IL-4H	*3 AS variant*	*Zinc finger MYM-type protein 1*	*HORVU4Hr1G070940*	*0.00*	*0.38*	*0.18*	*80*	*1254*	*no*
Qhei.S42IL-7H	*6 AS variant*	*F-box protein*	*HORVU7Hr1G113340*	*0.12*	*0.56*	*0.00*	*42*	*1113*	*ISR*
Qhei.S42IL-7H	*23 AS variant; 3 bp deletion Golf*	*ENTH/ANTH/VHS superfamily protein*	*HORVU7Hr1G097510*	*0.04*	*0.26*	*0.01*	*141*	*2115*	*no*
Qlr.S42-4H.a	*72 AS variant; 1 bp deletion Golf*	*AP2-like ethylene-responsive transcription factor AIL5*	*HORVU4Hr1G081400*	*0.01*	*0.11*	*0.09*	*200*	*2991*	*Golf*
QLR.S42-7H.a	*2 AS variant*	*flavin-dependent monooxygenase 1*	*HORVU7Hr1G121290*	*0.06*	*0.53*	*0.00*	*7*	*199*	*no*

Qpm.S42-1H.a	*44 AS variant*	*Disease resistance protein*	*HORVU1Hr1G001490*	*0.62*	*0.62*	*0.22*	*26*	*427*	*no*
Qpm.S42-1H.a	*48 AS variant; 1 bp deletion Golf*	*TRICHOME BIREFRINGENCE-LIKE 8*	*HORVU1Hr1G001500*	*0.64*	*0.70*	*0.41*	*4*	*120*	*ISR*
Qpm.S42-2H.c	*210 AS variant; 3 bp deletion Golf*	*Chromosome 3B, genomic scaffold, cultivar Chinese Spring*	*HORVU2Hr1G109840*	*0.00*	*0.26*	*0.15*	*99*	*1581*	*no*
Qrdw.3H	*107 AS variant; 9 bp deletion ISR42-8*	*EamA-like transporter family protein*	*HORVU3Hr1G095930*	*0.03*	*0.27*	*0.07*	*53*	*870*	*no*
Qrdw.3H	*6 AS variant*	*auxin response factor 2*	*HORVU3Hr1G096510*	*0.03*	*0.44*	*0.10*	*11*	*154*	*no*
Qrdw.3H	*6 AS variant*	*auxin response factor 2*	*HORVU3Hr1G096510*	*0.03*	*0.44*	*0.10*	*11*	*154*	*no*
Qrdw.S42IL.1H	*115 AS variant; 9 bp deletion Golf*	*undescribed protein*	*HORVU1Hr1G005230*	*0.40*	*0.35*	*0.22*	*94*	*1384*	*no*
Qrdw.S42IL.1H	*122 AS variant; 1 bp deletion Golf*	*Single-stranded nucleic acid-binding protein R3H*	*HORVU1Hr1G009430*	*0.70*	*0.60*	*0.24*	*84*	*1646*	*no*
Qrdw.S42IL.1H	*230 AS variant; 18 bp deletion Golf*	*Plant Tudor-like RNA-binding protein*	*HORVU1Hr1G008510*	*0.64*	*0.45*	*0.17*	*51*	*871*	*Golf*
Qrdw.S42IL.1H	*84 AS variant; 3 bp deletion Golf*	*Potassium transporter family protein*	*HORVU1Hr1G009100*	*0.62*	*0.52*	*0.11*	*90*	*1625*	*no*

Qrdw.S42IL.3H	*309 AS variant; 5 bp deletion ISR42-8*	*Serine/threonine-protein kinase*	*HORVU3Hr1G015840*	*0.03*	*0.13*	*0.09*	*67*	*1117*	*ISR*
Qrdw.S42IL.4H	*11 AS variant*	*Triacylglycerol lipase SDP1*	*HORVU4Hr1G001470*	*0.20*	*0.30*	*0.07*	*101*	*1827*	*no*
Qrdw.S42IL.7H.a	*14 AS variant*	*Disease resistance protein RPM1*	*HORVU7Hr1G013300*	*0.05*	*0.47*	*0.01*	*164*	*2832*	*no*
Qrdw.S42IL.7H.a	*18 AS variant; 1 bp deletion Golf*	*exocyst subunit exo70 family protein G1*	*HORVU7Hr1G022980*	*0.01*	*0.25*	*0.20*	*17*	*256*	*no*
Qrdw.S42IL.7H.a	*57 AS variant; 3 bp deletion Golf*	*auxin response factor 16*	*HORVU7Hr1G021880*	*0.01*	*0.19*	*0.06*	*244*	*4485*	*no*
Qrdw.S42IL.7H.b	*3 AS variant*	*Leucine-rich receptor-like protein kinase family protein*	*HORVU7Hr1G085790*	*0.12*	*0.28*	*0.02*	*53*	*820*	*no*
Qrdw.S42IL.7H.b	*1302 AS variant; 3 bp deletion Golf*	*ABC transporter G family member 42*	*HORVU7Hr1G088600*	*0.06*	*0.22*	*0.05*	*480*	*7521*	*no*
QRI.7H	*4 AS variant*	*B3 domain-containing protein*	*HORVU7Hr1G004350*	*0.03*	*0.25*	*0.02*	*363*	*6022*	*no*
QRL.S42IL.1H.b	*187 AS variant; 2 bp deletion Golf*	*Cytoplasmic membrane protein*	*HORVU1Hr1G094070*	*0.02*	*0.19*	*0.05*	*63*	*1036*	*no*
QRL.S42IL.4H.a	*365 AS variant; 21 bp deletion ISR42-8*	*GDSL esterase/lipase*	*HORVU4Hr1G001420*	*0.15*	*0.19*	*0.08*	*101*	*1636*	*no*

QRS.3H	*145 AS variant; 3 bp deletion Golf*	*Protein kinase family protein*	*HORVU3Hr1G055620*	*0.02*	*0.18*	*0.06*	*245*	*3987*	*no*
QRS.3H	*64 AS variant; 2 bp deletion ISR42-8*	*Serine/threonine-protein kinase*	*HORVU3Hr1G038690*	*0.01*	*0.12*	*0.03*	*560*	*9561*	*ISR*
QRV.S42IL.2H	*11 AS variant; 4 bp deletion ISR42-8*	*Ethylene-responsive transcription factor 10*	*HORVU2Hr1G111680*	*0.01*	*0.23*	*0.14*	*244*	*4006*	*no*
Qsdw.2H.b	*64 AS variant; 6 bp deletion Golf*	*unknown protein*	*HORVU2Hr1G095840*	*0.46*	*0.30*	*0.15*	*65*	*1061*	*no*
Qtil.1H	*10 AS variant*	*sugar transporter 9*	*HORVU1Hr1G091460*	*0.04*	*0.16*	*0.09*	*50*	*867*	*no*
Qtil.S42IL-3H	*6 AS variant*	*auxin response factor 2*	*HORVU3Hr1G096510*	*0.03*	*0.44*	*0.10*	*11*	*154*	*no*
Qtil.S42IL-3H	*6 AS variant*	*auxin response factor 2*	*HORVU3Hr1G096510*	*0.03*	*0.44*	*0.10*	*11*	*154*	*no*
Qtil.S42IL-3H	*15 AS variant; 5 bp deletion Golf*	*Cytochrome P450 superfamily protein*	*HORVU3Hr1G108150*	*0.14*	*0.36*	*0.00*	*4*	*79*	*no*
Qtil.S42IL-4H	*3 AS variant*	*respiratory burst oxidase homologous B*	*HORVU4Hr1G086500*	*0.03*	*0.13*	*0.14*	*93*	*1607*	*no*
QTL3H_1	*109 AS variant; 12 bp deletion ISR42-8*	*NAC domain protein,*	*HORVU3Hr1G014140*	*0.03*	*0.21*	*0.12*	*87*	*1446*	*no*
QTL3H_1	*294 AS variant; 441 bp deletion ISR42-8*	*Protein of unknown function (DUF1664)*	*HORVU3Hr1G015180*	*0.05*	*0.33*	*0.12*	*55*	*703*	*no*

QTL3H_4	*380 AS variant; 3 bp deletion ISR42-8; 11 bp deletion Golf*	*Chromosome 3B, genomic scaffold, cultivar Chinese Spring*	*HORVU3Hr1G111140*	*0.07*	*0.27*	*0.04*	*48*	*828*	*Golf*
Ryd2	*29 AS variant; 4 bp deletion ISR42-8*	*RING/FYVE/PHD zinc finger superfamily protein*	*HORVU3Hr1G089990*	*0.51*	*0.74*	*0.14*	*26*	*399*	*no*
Ryd3	*295 AS variant; 2 bp deletion Golf*	*Pathogenesis-related thaumatin superfamily protein*	*HORVU6Hr1G024780*	*0.25*	*0.16*	*0.10*	*13*	*211*	*ISR*
SCRI-RS239784	*44 AS variant; 6 bp deletion Golf*	*Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein*	*HORVU1Hr1G009800*	*0.55*	*0.48*	*0.09*	*184*	*2928*	*no*
SD1	*no cds alignment*	*alanine aminotransferase 2*	*HORVU5Hr1G062990*	*0.00*	*0.01*	*0.00*	*57*	*956*	*-*
Thresh-1	*37 AS variant; 724 bp deletion ISR42-8; 3 bp deletion Golf*	*WUSCHEL related homeobox 1*	*HORVU1Hr1G010580*	*0.50*	*0.36*	*0.14*	*95*	*1535*	*no*
Thresh-1	*6 AS variant*	*laccase 7*	*HORVU1Hr1G008360*	*0.59*	*0.39*	*0.49*	*99*	*1602*	*no*
VrnH2	*190 AS variant; 12 bp deletion Golf*	*Auxin efflux carrier family protein*	*HORVU4Hr1G082680*	*0.04*	*0.00*	*0.13*	*129*	*2168*	*no*

3.5.6 Genetic distance evolution (genetic map construction)

The physical map overweight's the pericentromeric region in the graphical illustration. Genetic maps have been used as a tool to illustrate a linear map according to recombination likelihoods. This method overrepresents the telomeres of chromosomes, as these are more likely being recombined between the gametes (Muñoz-Amatriaín et al., 2011). A genetic map is based on the allele and haplotype frequency of two neighboring markers. In the constructed pooling approach, the allele frequency is known, but the haplotype frequency is hidden. The haplotype frequency can only be estimated when genotypes are sequenced individually. A workaround has been constructed to overcome the problem of the missing haplotype frequency information. This is based on the allele frequency deviation of two neighboring markers and the physical distance on the chromosomal map. The physical distance between two haplotypes was used as a replacement for the missing haplotype frequency information. With this, the recombination likelihood *D'* and the linkage r^2 were estimated. Based on *D'*, a genetic map was created in order to assess the fitness of chromosomal regions. The genetic map was calculated for each generation and environment separately so that each population has a unique map. A contracted genetic map indicates a higher selection pressure towards one allele or haplotype and is associated with reduced genetic diversity in the population. The precision of the applied genetic map calculation can be acquired from figure 64. The pattern of the linkage value r^2 was highly related to patterns observed in classic genetic map construction scenarios - the higher the inter haplotype distance, the lower the chances of a linkage between these two. A value of one is associated with a complete linkage and observed in the pool data by an exactly identical allele frequency value for two neighboring haplotypes. The overall linkage is highest in the 3rd generation, and it decreases in the following generations.

The sum of the genetic map length for each population is illustrated in figures 65 and 66. The genetic distance in the 3rd generation was highest on chromosomes 2H and 5H. Both chromosomes had a length of approximately 60 cM. The two shortest regular chromosomes were 3H (35 cM) and 7H (20 cM). The comparison of the genetic and physical map highlights distinct regions with high and low recombination events. Across all chromosomes, the centromere region was illustrated by a low recombination rate. Remarkably, the telomere on the short arm of chromosome 5H was also represented by a very low recombination rate. In the 12th generation, a decrease in the genetic map extension was observed for both environments. While little variation to the 3rd generation can be observed for the centromere, the telomeres were remarkably shorter. Contrasting to all other chromosomes, the 5th chromosome decreases in the centromere's length but increases its length in the telomeres. The overall length of the chromosome was still decreased to 45 cM. The most remarkable decrease in genetic map length was observed for chromosome 2H in the conventional system. Little variation in both environments was reported for the 3rd chromosome, although the allele frequency pattern clearly changed (Figure 65). In the following generations, the conventional and organic populations evolved in the same manner for chromosome 1H and the unknown chromosome. For the remaining chromosomes, the variation between the systems became more distinct with each subsequent generation. While the genetic map length did not change as drastically from the 12th generation onwards, the map length increased in the organic system in each subsequent generation. While the initial length of chromosome 5H was reached once again in the 23rd generation, the map length of chromosome 3H exceeded the length observed in the 3rd generation already in the 16th generation. Onward, the length increased further to 54 cM, compared to the initially observed 34 cM. All chromosomes have in common that the genetic map length is higher in the organic compared to the conventional system. Compared to the initial generation (BC_2F_3), the organic population showed evidence of an increased map length for chromosomes 3H and 7H in the later generations (Figure 65). The results showed that genetic diversity, calculated as the sum of the genetic distance across all chromosomes, were decreased in both farming systems over the generations from 300 centi Morgan (cM) for the BC_2F_3 generation to 130 cM and 250 cM for the 23rd generations of the conventional and organic systems, respectively.

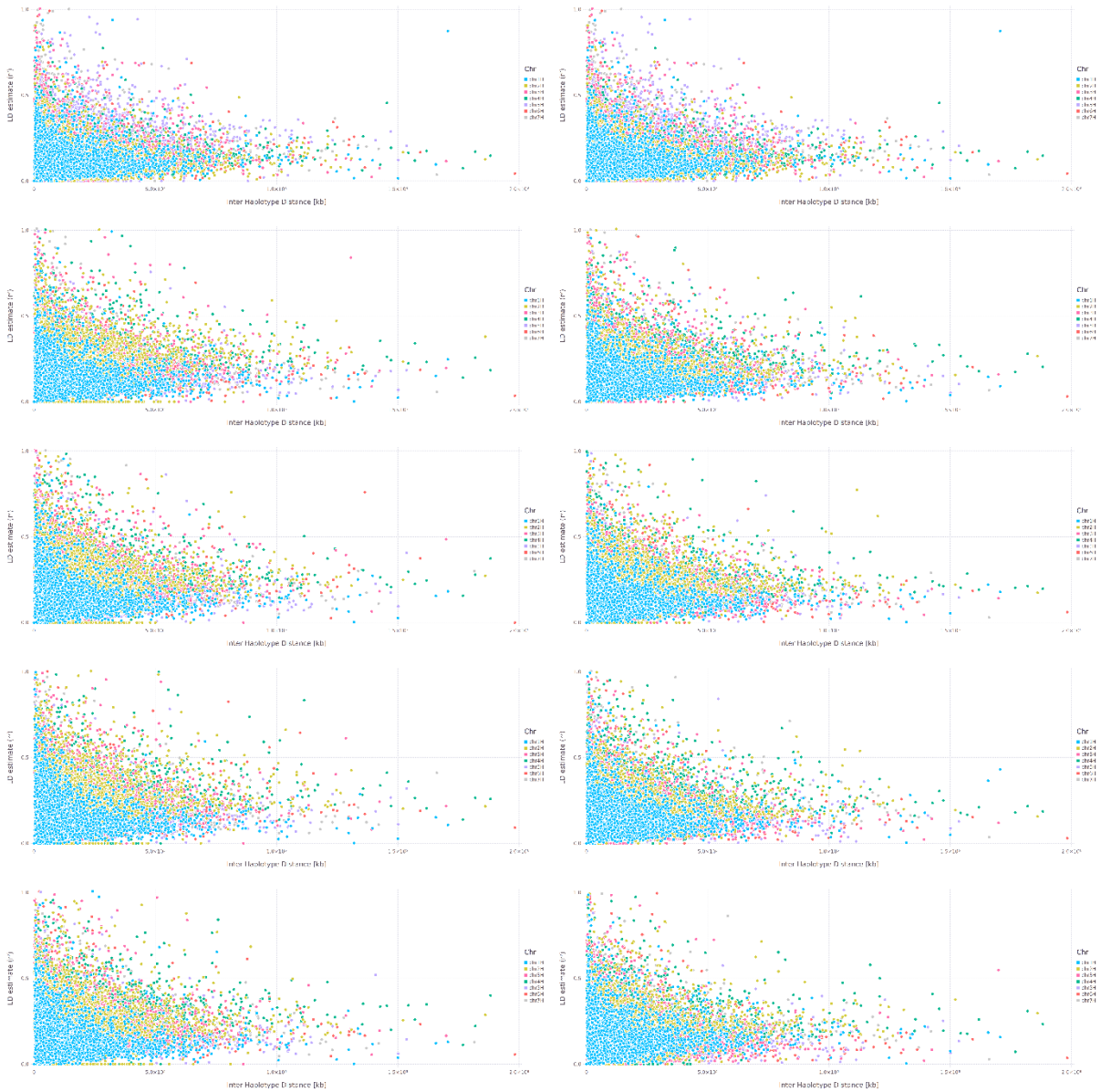


Figure 64: r² linkage value as LD estimation rate (y-axis) compared to the physical distance of the two neighboring haplotypes in kb (x-axis). The chromosomes are separated by color. First row – BC2F3 generations, equal in both columns; 2nd row – the 12th generation of the barley populations, the left side is conventional, the right side the organic population; 3rd row – 16th generation; 4th row – 22nd generation; 5th row - 23rd generation.

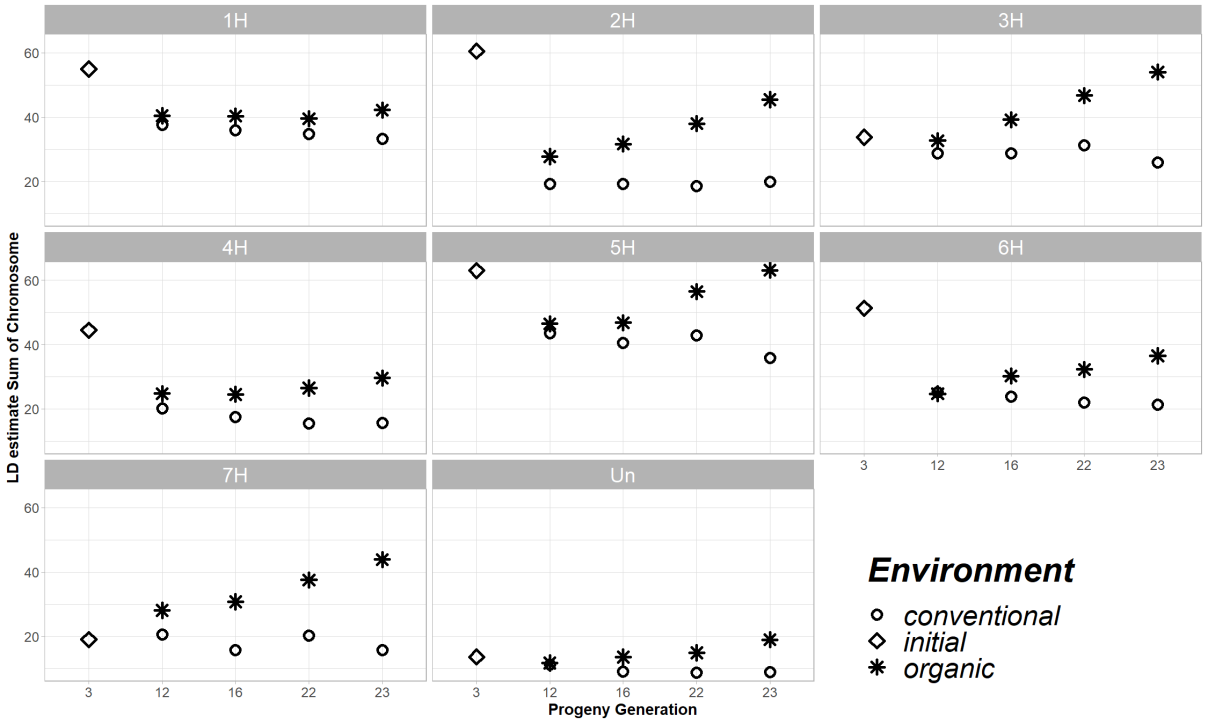


Figure 65: Genetic map length per each chromosome for the conventional, organic, and initial populations. The chromosome-wise sum of LD is facetted for chromosomes and plotted against the generation (x-axis). A high value is associated with high variability. Decreasing values indicate reduced variability, as observed for the conventional system.

The comparison of the genetic and physical map, illustrated in figure 66, describes interesting variations between chromosomes, generations, and environments. The first chromosome followed a transposed sigmoidal function curve. The centromere can clearly be separated from the telomeres. The short arm was characterized by a rapid increase of genetic distance, compared to the low increase in physical distance. Contrasting to this, the transformation of the centromere to the long arm telomere is much smoother with a less drastic transition. In the following generations of the conventional population, this transition becomes more and more linear with each subsequent generation. In comparison to this, the organic population retains the shape observed for the initial 3rd generation. The genetic to physical map ratio on chromosome 2H was much more homogeneous when the two telomeres were compared to each other (F3 generation). Balanced to this generation, the transition of centromere to telomere became more distinct in the later generations. This holds true for both environments, while the telomere on the long arm in the conventional system was described by another break. When the two telomeres were compared to each other, the short arm was much shorter compared to the long arm telomere. This observation could also be made for the 3rd, 5th, and 6th chromosomes. The telomere on chromosome 3H extended in comparison to the F3 generation in both environments and all following generations. Besides a harsher transition towards the long arm telomere in the conventional populations, little variation could be observed based on the shape. The same holds true for the 4H chromosome. Chromosome 5H was characterized by an exceptional shape. Especially in the 3rd generation, no transition from the telomere on the short arm towards the centromere was visible. More than 500 million base pairs seem not to be recombinant – more than 70% of the entire chromosome. The telomere on the long arm was characterized by a high degree of recombination, equivalent to 50 centi Morgan on 120 million base pairs. The telomere on the short arm was just minorly getting more extended in size in the following generations. Besides the sharp transition of centromere to long-arm telomere, the conventional and organic systems were described by an identical shape of the curve. No major variations could be described for the 6th chromosome either. Contrasting, the 7th chromosome deviations in between both

farming systems. While the shape of the curve in the conventional system is almost linear, the organic system is described by a distinct transposed sigmoidal function.

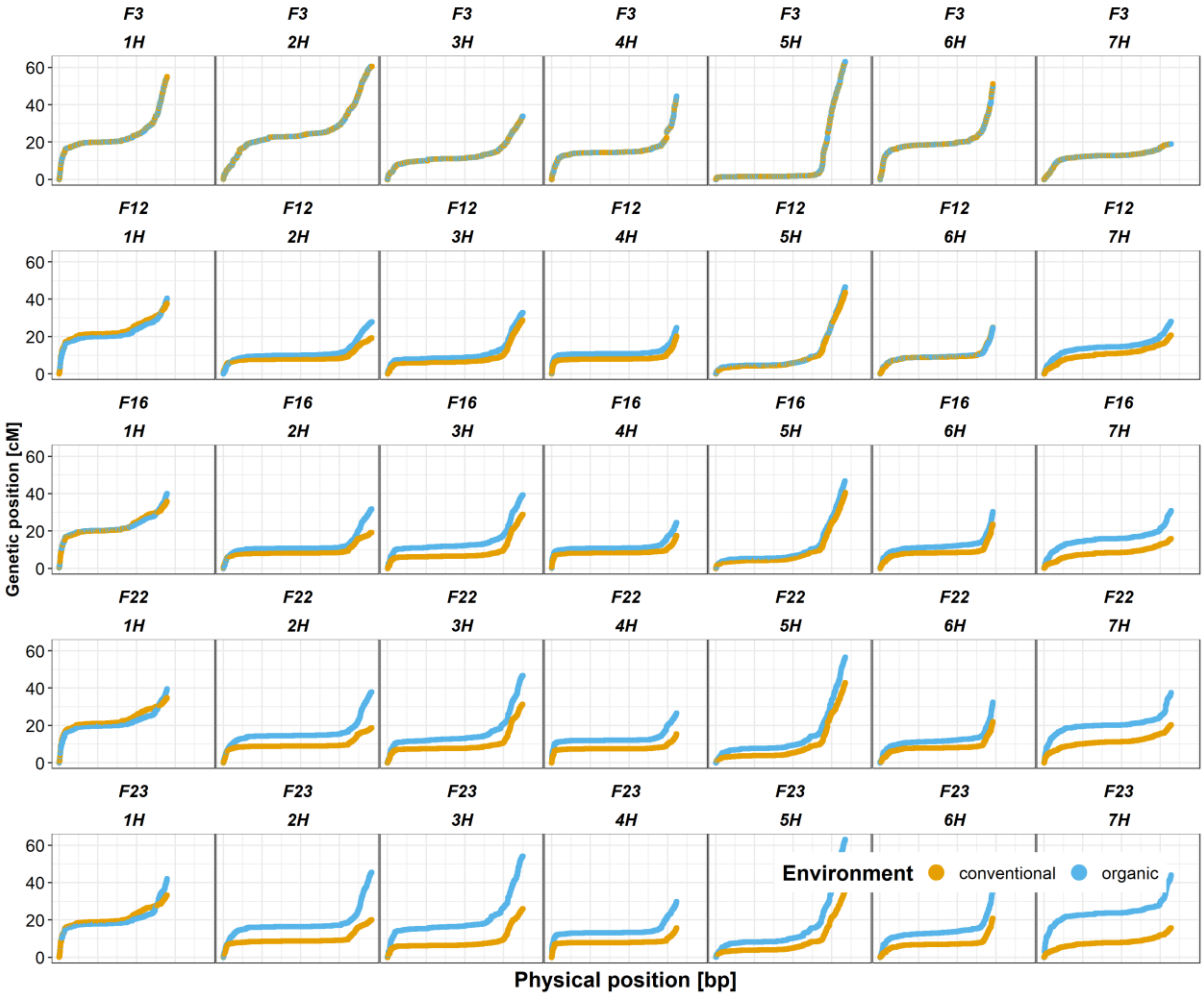


Figure 66: The genetic distance compared to the physical distance to investigate the recombination event likelihood. The environments are separated by color. Nine thousand random haplotypes are drawn and used to generate the plot. The genetic position (y-axis) is plotted against the physical position (x-axis). The generation is given above each plot, as well as the identifier for the chromosome. The blue curve of the organic population reveals a higher genetic map length compared to the yellow curve of the conventional population. Extended map length indicates more variant selection processes and higher genetic variability.

When the allele frequency is set into perspective to the genetic distance assessment, the results illustrated in figure 67 can be analyzed. A more or less unorganized pattern in the 3rd generation was replaced by clearly distinguishable patterns in the following generations. Remarkably, the most prominent allele frequency variant loci were present from the 12th generation onward. Additionally, the conventional system was described by less variation in the allele frequency – also represented by smaller genetic map size. The results showed that genetic diversity, calculated as the chromosome map length, was decreased in both farming systems over the generations from 300 centi Morgan (cM) for the BC2F3 generation to 130 cM and 250 cM for the 23rd generations of the conventional and organic systems, respectively.

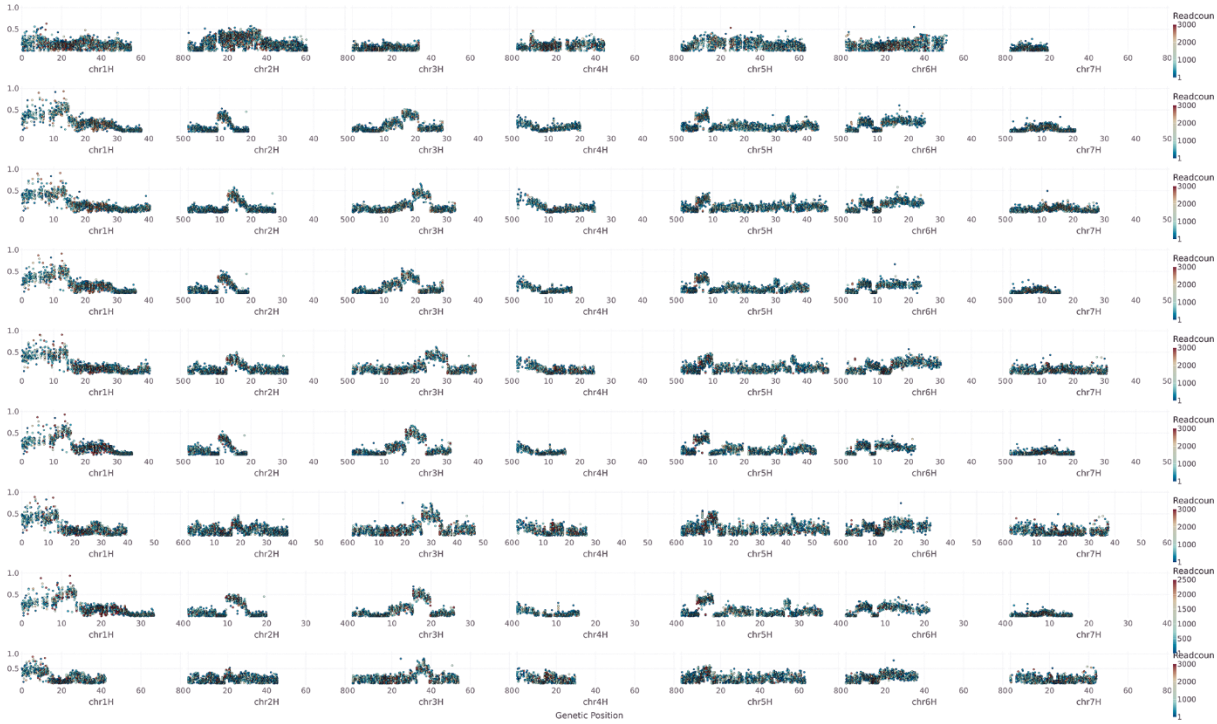


Figure 67: The HAF (y-axis) plotted against the calculated genetic map (x-axis). Each point represents one haplotype, while the color of the point is related to the read coverage of the haplotype. A genetic map was generated for each set of samples separately (generation × environment), resulting in variant length per chromosome for each presented sample set. F3, F12E1, F12E2, F16E1, F16E2, F22E1, F22E2, F23E1 and F23E2 is the order of the rows. E1 and E2 represent the conventional and organic systems, respectively.

The calculation resulted in a consistently more contracted map in the conventional system across the generations and chromosomes, while the organic population indicates increased variation for various regions and chromosomes.

3.5.7 Drought response of populations

A prediction of drought adaptation was performed by observing the common WT allele frequency shift in the years 2011 (F16), 2017 (F22), and 2018 (F23) as compared to those in 2007 (F12). While five significant ($p < 0.05$) gene ontologies were identified in all tested conventional and organic environments (DNA-dependent regulation of transcription(Talamè et al., 2007), transcription factor activity (Al Abdallat et al., 2014; Alexander et al., 2019; Marè et al., 2004), nucleus (Qin et al. 2017), DNA binding (Xue, 2003) and sequence-specific DNA binding (Abe et al., 1997)), five additional ontologies were identified in the conventional environment (protein binding, oxidation-reduction, oxidoreductase activity(Moran et al., 1994), zinc ion binding(Yang et al., 2009) and nucleic acid binding).

The DNA-dependent regulation of transcription gene ontology was observed to have an identical allele frequency of the wild-type of 10.6% in both organic and conventional environments in the 12th generation. Compared to the 3rd generation, a minor increase of 2% was observed. In contrast to this, the wild-type allele frequency had decreased to 6.6% and 7% in the organic and conventional 16th generations, respectively. The ISR42-8 allele frequency had further decreased in the 22nd and 23rd generations to 5.2% and 3.5% for the conventional and organic systems. While the first years with a sufficient amount of evaporable water were characterized by an increase of the ISR42-8 allele frequency, the years observed to be drought effected showed evidence of wild-type allele frequency reduction for

the genes related to transcription regulation. A widespread pattern was observed for the transcription factor activity gene ontology class. An increase of the allele frequency in the 12th generation was followed by a decrease in the 16th, 22nd, and 23rd generations. The other three gene ontologies, nucleus, DNA binding, and sequence-specific DNA binding, all fell into the same category. Non-surprisingly, the pattern of the GO allele frequency was equivalent to the first two described classes.

Five GO classes were only relevant in the conventional population. The oxidoreductase activity was classified by slight variation between the 12th and 3rd generation but changes afterwards (9.5%). The ISR42-8 allele frequency was decreased by almost 50% in the 22nd and 23rd generations (4.1%). Analogous, the pattern of zinc ion binding genes was observed. A significant change in nucleic acid binding GO term class was only found in the conventional environment. While the 12th generation was classified by an increase of 5% in the ISR42-8 allele frequency, the allele frequency of the wild-type was reduced on the level observed in the 3rd generation. The pattern of this wild-type allele frequency variation of the generations is presented in figure 68.

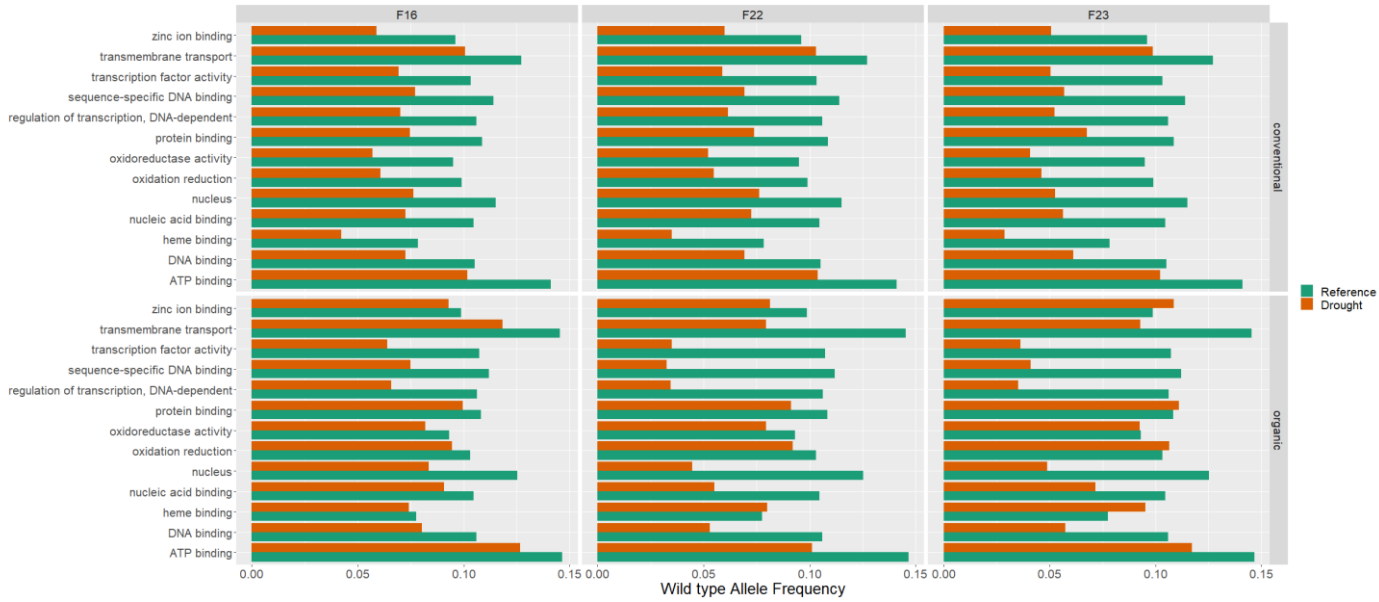


Figure 68: Gene ontology of significant genes identified between the 12th (Reference) and 16th, 22nd, and 23rd generation (drought). The generations and environments are faceted – the color represents the 12th generation (moist reference, green) and the drought candidate (orange). The bars indicate the average wild-type allele frequency over all genes related to the same GO term.

Generally, a group of wild alleles has been negatively selected in drought environment conditions. Hence, the climatic conditions in exceptional years influenced the genetic composition of the observed populations.

3.5.8 Disease resistance response

To estimate the effect of weather conditions on the genetic constitution of barley, the allele frequency patterns were compared between the years 2007 (F_{12}) and 2017 (F_{22}). As 2007 and 2016 were notably moist and high fungal pressure years, the conventional (control) and organic system (stress) were compared and referenced against the BC_2F_3 generation to ensure selection processes had occurred. In total, 11 commonly selected gene ontology classes were observed for both generations, which were related to oxidative stress responses and signaling functions. In addition, 34 gene ontologies were identified for the 22nd generation. In the 12th generation, all gene ontologies showed a minor increase in the wild-type allele frequency in the organic system. The most relevant variation could be reported for the protein serine/threonine kinase activity. For this class, the wild-type allele frequency was 11% in the organic, while 4% in the conventional system of the 12th generation. In the 22nd generation, this

variation was even further extended. In this generation, the wild-type allele frequency was 13.6% and 3.3% in the organic and conventional systems, respectively. Another serine-type carboxypeptidase activity observed to be significant in the 22nd generation was found to have an ISR42-8 allele frequency of 15.6% in the organic and 9.3% in the conventional environment. Contrasting to the kinase, the ISR42-8 allele frequency was also increased in the conventional system when compared to the 3rd generation. The overall protein kinase activity had also been found to be significantly different between the environments for both generations tested. While the 12th generation was characterized by a minor deviation between the systems, in the organic system, the 22nd generation presented an ISR42-8 allele frequency increased by 3.7%, compared to the conventional system. In this group, 205 genes were found to be significantly variant between the organic and conventional systems. An even higher gene count per GO term was observed for ATP binding (341) and protein binding (452). The pattern was equivalent to the one observed for the kinase activity – little variation between the mean of the systems in the 12th generation but more discrete variations in the 22nd generation. The variation between the systems in the 22nd generation was 4% for both ATP and protein binding class. A constant variation between the environments over the generations could be reported for the hydrolase activity, acting on ester bonds gene ontology.

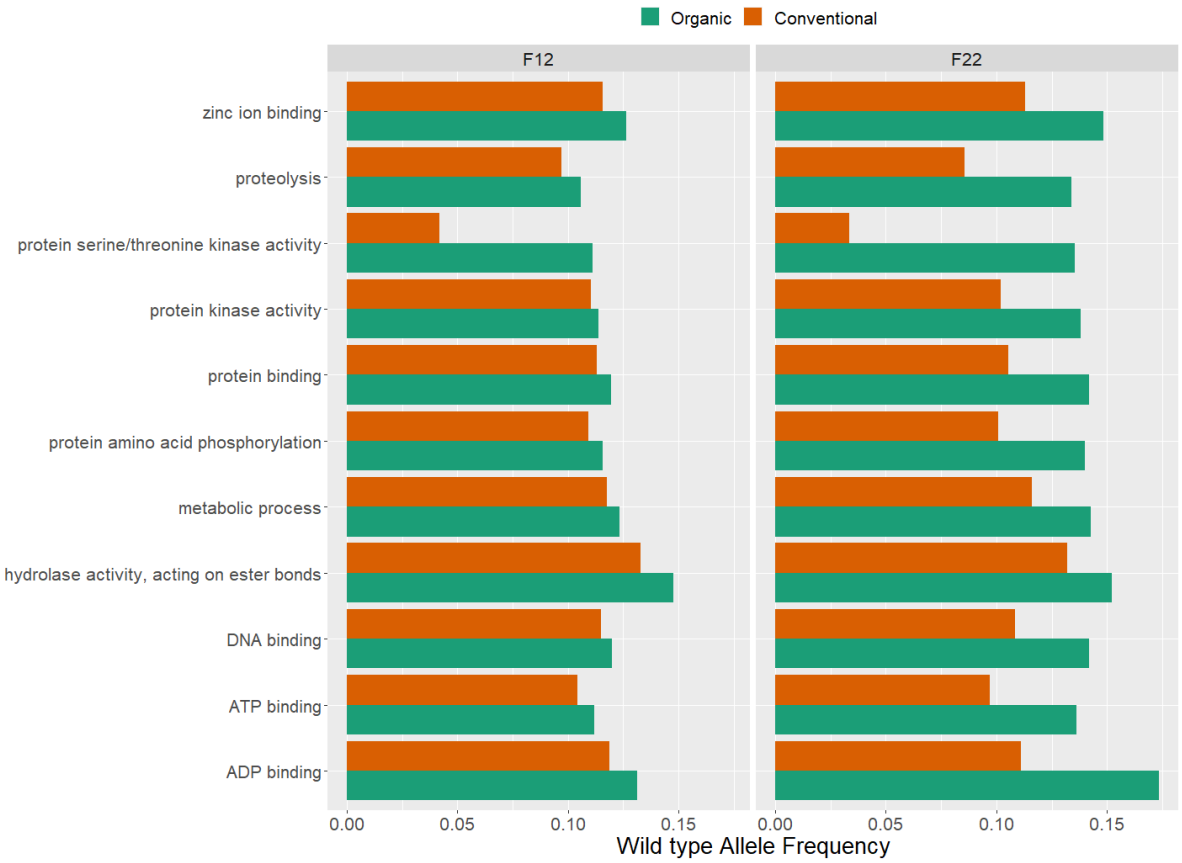


Figure 69: Gene ontology (GO) intersection between the 12th and 22nd generation for fungal attack related classes. The organic (green) and conventional (orange) environments are separated by color. The bars indicate the average wild-type allele frequency over all genes related to the same GO term.

In both years, the variation was ~1.5%, while both environments showed an increase in the wild-type allele frequency compared to the 3rd generation (7.6%). A genetic modification of the metabolic processes was also observed, as the ISR42-8 allele frequency for these genes was ~0.8% and ~3% higher in the organic compared to the conventional system. Eighty-eight genes fell into this category. Further details can be obtained from figure 69.

3.6 Predicted or putative protein variation observation

To predict the functional effects of SNP variants, the coding sequences (CDS) with annotated start and stop codons were aligned to the parental haplotypes. A total of 28,601 complete CDS alignments were generated, and 640,611 amino acid variants in 16,570 predicted proteins were identified between the parents. As compared to the reference CDS, 3,831 and 3305 genes in the WT and the cultivar genomes, respectively, showed evidence of deletion or insertion.

On a median, each CDS had one amino acid mismatch, contrasting to an average value of 22.4. In total, 75,874 stop codons were observed for the 28,601 proteins for the Golf allele, while there were 74,955 for the ISR42-8 allele. On a median level, each gene had one single stop codon for both genotypes, while the average value was 2.65 and 2.62 for Golf and ISR42-8, respectively. The most dominant impact on the amino acid sequence of a gene was expected from deletions and insertions. Genes that were characterized by a deletion in one genotype, while the other genotype was typified by the reference sequence, were likely candidates for causal allele frequency variation of a chromosomal region. One thousand two hundred fifty-two genes with observed deletions in the ISR42-8 genotype, leading to amino acid variations between the two parental genotypes, were observed. On the contrasting side, 730 genes were identified where ISR42-8 was characterized by the reference sequence, but Golf is typified by a deletion that resulted in a frameshift of the amino acid reading frame. In regard to the start and stop codons, variations were also observed between the genotypes. While 2,521 genes were classified by a missing start codon in ISR42-8, the same holds true for 2,435 genes in the Golf genotype. In regard to the stop codons, 886 were missing in ISR42-8, while 723 were absent in Golf. As (multiple) preliminary stop codons could also be the result of a frameshift mutation, the genes were screened for these as well. While 74,955 stop codons were found for ISR42-8, marginally more stop codons were counted for Golf (75,874). This resulted in an average stop codon count per gene of 2.62 for both genotypes, while the median value was one. Generally, the average protein size was observed to be 364 amino acids, and the chance to observe a mismatch of an amino acid was 0.062% per amino acid. A high degree of variation was observed on the DNA level as well as the predicted amino acid level.

3.7 Grain yield evolution in the populations

The yield in both systems was affected by abiotic and biotic effectors, leading to a highly variable yield in the period of the experiment.

The yield of the conventional population ranged from 0.67 to 1.2 times the average yield of the first three years (43 to 77 dt/ha). Compared to a country and statewide comparison, the yield was observed to be above the level of one, indicating a higher yield in this experiment compared to the average yield on other fields. 2004, 2014, 2016, and 2017 have shown to be exceptions from this law, as these years were reparented by a lower yield level compared to the country or state average yield. The lowest ratios were observed in 2004 and 2017 with 0.75 and 0.76, respectively. Analogous to these observations, the yield ratio to the first three years showed a very similar pattern, indicating a high correlation between these curves. The yield pattern can be divided into five time periods based on their yield pattern. While the seasons from 2001 to 2004 were represented by a constant decrease of yield, 2005 stands out as a year with an above-average yield, compared to the surrounding years. Contrasting to 2005, the period from 2006 to 2007 is described by a constant low yield level. From 2007 to 2013, the yield is observed to increase constantly from 46 dt/ha to 76 dt/ha – an increase of 50% without a variation in cropping strategy or artificial adaptation of the seed material. Its peak in 2013 is followed by a yield decreasing period until 2017, where the yield level decreases to 41 dt/ha – an absolute minimum of the entire experimental duration. In the following two years, a recovery of the yield level can be observed, ending in a yield level of 59 dt/ha and 50 dt/ha in the years 2018 and 2019, respectively (Figure 70A).

Grain yield evolution in the populations

The organic population yield ranged from 0.6 to 1.4 times the yield of the first three years until 2015 (27 to 59 dt/ha). The years 2007 and 2016 were characterized by massive yield reductions in the organic system of 41% and 31%, respectively, as compared to the previous years. In both years, mild and wet periods were observed during the crop development and anthesis stages. The yield reduction to 0.89 in 2016 was followed by three extraordinarily dry years, which reduced the yield ratio to an absolute minimum of the entire experimental period in 2019 (0.48 times; 20.53 dt/ha, figure 70B).

Once again, the duration of the experiment can be separated into three periods. While the first period from 1999 to 2006 was characterized by a constant yield, 2007 marks a first local minimum in the curve of the yield with a value of 27.5 dt/ha. The following years were characterized by a swinging increase in the yield. From 2008 to 2015, the yield increases solidly with an average of 44 dt/ha. This period was followed by a yield depletion period from 2016 to the last year of observation. In 2016, the yield decreased to 38 dt/ha, most likely based on fungal pressure prior to and during the anthesis stage. In the years following, the yield decreased further to an absolute minimum of the entire experiment in 2019 (20.53 dt/ha). This amount equates to only 40% of the yield obtained in the conventional system. As compared with the organic system, the yield levels in the conventional system recovered relatively quickly after 2016 (yield reduction of 27% in 2016). In 2017, the ratio decreased to 0.64 times (41 dt/ha), followed by an increase to 0.92 times (59 dt/ha) in the dry year of 2018.

Compared to the country and state-wide yield, the ratio remained below one for all years except for 2010. The average ratio over the period is 0.8 for the organic population, while 1.15 for the conventional population. The raw average yield is 58.65 dt/ha and 41.08 dt/ha for the conventional and organic systems, respectively. Based on these values, the average yield in the organic system was 30.1% lower compared to the conventional system. Only in the years 2010 and 2014, the yield of the organic population was observed to be higher in comparison to the conventional. Compared to the average yield of the first three years, the yield ratio overall generations was 0.91 and 0.96 in the conventional and organic environments.

In general, the yield of the organic system was 30% lower than the yield of the conventional system, and a more distinct effect of the (a)biotic stressors on the yield level might be a relevant result.

Figure 70: Grain yield evolution over the years of the experiment. Actual seed yield in metric deci tons per hectare are illustrated by bars, while the lines illustrate the ratio of seed yield measured in the experiment and the comparison towards average yields obtained in the first three years of the experiment (purple), Germany wide (red) and in

North Rhine-Westphalia only (green). The lines use the left y-axis, while the bars are put into the perspective of the right y-axis. ***A*** *– the seed yield obtained from conventionally farmed spring barley population;* ***B*** *- seed yield obtained from the organically farmed spring barley population.*

Chapter 4 – Discussion

4.1 Variation in parental physiology and morphology

The two genotypes used to create the cross are characterized by significant variations in their physiological and morphological habit. Golf and most other domesticated cultivars release their anthers before the ear emerges from the stem, which leads to the predominant selfing of barley (Allard, 1988; Kahler et al., 1975; Morrell et al., 2005). Contrasting to this, ISR42-8 is observed to have an open pollination status, where anthers are exposed widely under greenhouse conditions. This might enhance cross-pollination in this wild-type, though it has been reported that little variation in the outcrossing status was detected between *Hordeum vulgare* and *Hordeum spontaneum* (Abdel-Ghani et al., 2004). The enhanced extrusion might predominantly be affected by the missing precipitation on flowers, as irrigation was done by applying water directly to the soil. This observation has been reported by (Abdel-Ghani et al., 2004) as well.

A pronounced late flowering was observed for the wild-type ISR42-8, which is unsurprising as this genotype evolved in arid conditions with high thermal time during an extended vegetation period. The occurrence of early flowering genes in wild-types has been reported by (Cockram et al., 2011), but many of these were also reported as mutant genes not found in wild-types but only in adapted cultivars.

Besides the late flowering, studies have reported the pronounced tolerance of the wild-type ISR42-8 to drought events (Honsdorf et al., 2014; Muzammil et al., 2018). Under greenhouse conditions, the phenotype was characterized to produce many leaves and having a high water demand, meaning under optimal water conditions, the genotype will use a lot of water to produce fresh biomass. The overproduction in the vegetative stage is not accompanied by increased generative production of seeds. Therefore the water use efficiency under a high water supply can be assumed as low. Under the water-reduced status of soils, the wild-type ISR42-8 quickly responds with a reduction of fresh leaf biomass. Various adaptation strategies in response to drought stress developed by wild-types are summarized by (Nevo & Chen, 2010), indicating, beyond others, a hypersensitive reaction to drought stress and natural adaptations to local environments and climates of wild-type populations.

Contrasting to the domesticated cultivar allelic diversity, the genetic variation of wild-types has been reported to be relatively high in terms of disease resistance (Nevo E. et al., 1979; Nevo et al., 1984; Williams, 2003). Similar results are reported in QTL studies revealing the advanced resistance of ISR42-8 alleles to powdery mildew and leaf rust (Schmalenbach et al., 2008; M. Von Korff et al., 2005). The observed phenotypic variation of biotic attack response between Golf and ISR42-8 indicates an advanced resistance or tolerance response of the wild-type.

The advanced seed dormancy observed in the wild-type is definitely unbeneficial in terms of commercial use of this seed material. As reported in (Sato et al., 2016), major seed dormancy habits are beneficial in wild habits and environments, where seeds gradually start to germinate, and some seeds might even germinate a year later. This habit was observed for the ISR42-8 as well, leading to a highly increased demand for chill treatment before sowing.

Yield components play a crucial role in the genetic formation of the populations. Therefore, yield components were investigated and characterized. The wild-type has shown evidence of ear brittleness and low threshability. These are reported to be common alleles in wild-types (Pourkheirandish et al., 2015; Schmalenbach et al., 2011). Both phenotypes were observed for ISR42-8. Furthermore, long ears with elongated seeds were observed, resulting in ear lodging.

Pronounced drought tolerance is reported to be genetically accommodated by an extended root system, with a low root growth angle and deep rooting physiology (Nevo & Chen, 2010). These characteristics were also observed for the wild-type ISR42-8.

In conclusion, ISR42-8, which was used as the WT donor, is reported to have an expanded root system (Naz et al., 2014), increased drought tolerance (Honsdorf et al., 2014, 2017; Sayed et al., 2012), increased seed and ear length and kernel weight (Zahn et al., 2020), reduced threshability (Schmalenbach et al., 2011), reduced plant height (Naz et al., 2014), late-flowering (Wang et al., 2010b), enhanced biotic stressor resistance genes (Dragan et al., 2013; Schmalenbach et al., 2008; M. Von Korff et al., 2005) and ear brittleness (Schmalenbach et al., 2011).

4.2 Phenotypic composition of populations

The phenotypic composition of the population was examined on both the above and below-ground traits. Below ground traits related to root morphology and physiology were tested in a hydroponic and field-based experiment. In both experiments, many traits were measured identically so that two replicates in one year could be reported. Both indicate identical findings in root diameter and length. While the aerenchyma and the metaxylem area were found to be more prominent in the conventional test set of genotypes, the root length and angle indicated a prolonged root growth towards deep soil layers. Additionally to this observation, the variation in the organic population was characterized by a higher heterogeneity. The variation for the aerenchyma area was low but highly significant. The importance of aerenchyma area in regard to drought or nutrient stressed plants is described before (Fan et al., 2003; Saengwilai et al., 2014; J. Zhu et al., 2010). These studies revealed the importance of increased aerenchyma area to cope with nutrient deficiency and drought events. This is in contrast to the observations and hypotheses formed in this experiment. It would have been expected that the organic population faces nutrient-deficient environmental conditions, wherefore the aerenchyma area should have been increased in the organic instead of the conventional system. Nevertheless, variation for this trait between the systems is not extraordinarily high. Furthermore, a potential trade-off in regard to fitness has not been understood and estimated (Saengwilai et al., 2014). The conclusion on this trait in regard to the observations opens up three potential explanations – either the fitness trade-off is high for the formation of pronounced aerenchyma, the genetic variation contributed by the parents was not sufficient to contribute beneficial alleles for the organic environment, or the conventional and organic population react in a different manner to nutrient and drought stress. Another approach to deal with limited nutrient availability and drought is also presented by the measurement of the root length and angle. As illustrated by (Hodge et al., 1999), extended root profiling results in increased nitrogen capture – an essential nutrient to promote plant growth and yield formation. Furthermore, a pronounced root system is observed to increase drought resistance in crops (D. Chen et al., 2018; Manschadi et al., 2008). As the organic population is observed to not only have an increased root length, but also a reduced root growth angle, it can be assumed that the root grows towards deep soil layers to capture nutrients and water. The fertilization with manure instead of mineral fertilizer might delay the availability of the nutrients, as the peak mineralization period is described to be between mid-August to early September (Lehrsch et al., 2016). Based on this assumption, the majority of fertilizer is plant accessible after flowering, or even in the next generation. As described by (Jannoura et al., 2013), the application of manure has a positive effect on the following crop in the crop rotation. As nitrogen is easily transportable and leachable in the soil, the mineralized manure might leach to deep soil layers in winter periods. Additionally, the mineralization is reduced but still ongoing in winter periods, which might add up the leaching potential to deep soil layers (Cookson et al., 2002). Therefore, roots need to grow deep to promote access to the fertilizer applied in previous years. As the fertilizer applied in previous generations is the single source of added minerals for the barley population, pronounced root development and growth is assumed as beneficial. Such advantages of pronounced root development have been observed

before (Bertholdsson et al., 2016; Finckh & Bhaskar, 2019). A positive side effect of this nutrient deficiency adaptation is increased drought tolerance (D. Chen et al., 2018; Manschadi et al., 2008). These studies describe drought-tolerant varieties by a low root angle and pronounced root growth straight below the upper body biomass center. This is in line with the observed reduced molecular genetic adaptation of the organic population towards drought events compared to the conventional population.

Contrasting to the root-related traits, no significant variations in terms of upper body physiology had been detected between the organic and conventional populations. Contrasting to hypotheses of variations in pronounced weed suppression by early vigor in the organic systems (Bertholdsson, 2005; Crespo-Herrera & Ortiz, 2015), none of both could have been specifically detected. A promoted plant biomass production, measured by the observation of the growth type, was also not scientifically different between the groups. Nevertheless, the growth type in both farming systems revealed an increased biomass production compared to the control genotype Scarlett. Furthermore, the conventional and the organic group were significantly variant in terms of growth habits, indicating a more erected growth in the conventional population. The other trait found to be significant was the ear posture. The ears in the organic group tended to head slightly downwards early, which might be associated with an increased ear size or reduced shoot stability. As no lodging was observed for the entire set of genotypes, no assumption can be made for this trait. The ear length was just marginally longer, which might show that this was not the single source of variation.

In regard to yield formation, also the SPAD measurement of the chlorophyll content indicated an increased score compared to the Golf parent – in both farming systems. The correlation between high chlorophyll leaf content and high yields has been examined in rice (Lu et al., 2017; Yamamoto et al., 2017). Furthermore, the ear length was observed to be marginally longer in the organic compared to the conventional population. None of these results were significantly different between the groups. Generally, some reviews have postulated relevant breeding goals for organic farming (Crespo-Herrera & Ortiz, 2015; Lammerts Van Bueren et al., 2011), but few studies have proven these organic breeding goals and adaptation of varieties. Some publications postulate the importance of weed suppression, resistance genes, and nutrient uptake (Wolfe et al., 2008). While weed suppression usually should be accomplished by a fast and high growth (Bertholdsson, 2005), the observed populations did not reveal any variations in this regard. Also, early or late flowering could not be detected. The second major breeding goal, the introgression of resistance genes, seems to be essential for both organic and conventional populations, as both groups revealed a positive selection of wild-type derived resistance genes. As prior illustrated elsewhere, the active expression of resistance genes and biomass growth to suppress weeds come with a cost of yield reduction (Bergelson & Purrington, 1996). The growth habit showed significant variations, but no extensive tillering or leaf production was observed throughout the entire vegetation period. Therefore, the costs of additional biomass formation have to be assumed as too high to gain a substantial fitness effect for genotypes in the organic population. The mechanical treatment of weeds in the early vegetation period might be sufficient to suppress weeds effectively. The same might be true for the resistance genes. In contrast to the weed suppression by early vigor, the root mass has been observed to be higher in the organic compared to the conventional population. This underlines the findings of Finckh et al. (2019) and Bertholsson et al. (2016), who showed an increased root mass, weight, and rooting depth in organically adapted populations and varieties.

In conclusion, the organic and conventional phenotypic evaluation revealed significant variations in regard to root morphologies, indicating an adaptation to opportune fertilizing strategies. Contrasting, upper body characteristics were found to be insignificantly different.

4.3 Climate and weather influential characteristics

The climate and weather had the second biggest impact on the evolution of the populations. To estimate the actual effect and the overrepresentation of some weather events, the evapotranspiration was calculated using the weather data collected by local weather stations. The base evapotranspiration ET_0 after the Penman-Monteith equation used to calculate the reference evapotranspiration above a grass canopy revealed the most significant correlation of ET_0 to the radiation, followed up by the humidity and the temperature. This was observed for the radiation and the humidity (Hidalgo et al., 2005), but not so much for the temperature. The possibility to utilize evapotranspiration as a drought indicator has been described before (Vicente-Serrano et al., 2015). Therefore, the crop-specific evapotranspiration ET_C and the actual crop evapotranspiration ET_A were calculated. The most critical factors of explanation for ET_A were the humidity, the soil water content, and the radiation - in descending order. The variation of ET_0 to ET_A explanatory effects highlights the fact that drought events are mainly driven by low humidity and an insufficient soil water content, which results in reduced transpiration. Both are counteracted by plants with an increased stomatal conductance to avoid excessive water loss, which was described before (Saliendra et al., 1995). As described by (Sánchez-Díaz et al., 2002), low humidity alone, resulting in high vapor pressure, does not lead to reduced transpiration. In fact, plants can be even more productive in biomass production and yield formation when the vapor pressure of the atmosphere is low. These publications indicate that the presence of low humidity and a low soil water content at the same time are the most relevant factors to indicate a drought event. As these two are highly represented by the ET_A, the calculated values give a reasonable estimate of the presence of water stress scenarios.
Climate change is a major factor in media, policy decisions, and people's life. Uncountable studies such as (Nerem et al., 2018) have shown that the increased emission of CO_2 to the atmosphere will result in increasing global temperatures, more severe drought, or heavy rain events. These changing climatic conditions have a direct impact on natural systems (Parmesan & Yohe, 2003) and, therefore, also on food production and security (Wheeler & Von Braun, 2013). In the observed time period from 1996 to 2019, on average, 1°C higher temperature in the last ten years was observed compared to the first ten-year period. Additionally, for the entire period, ten years without any drought and ten years with drought were observed. Remarkably, the most severe drought events were observed in the last decade in the years 2011, 2017, 2018, and 2019. As the data does not originate from the exact same weather station overall years, the comparison between time periods is challenging and cannot be used to extract a solid conclusion. Nevertheless, the last three years (2017 to 2019) were characterized by severe drought periods with a substantial impact on the productivity of crops. The obtained data also revealed that the quality of the drought events has changed. While moderate drought could be observed in the early years, with few days of drought and a moderate variation of potential to actual evapotranspiration, late generations showed evidence of prolonged drought periods with increasing deviations of actual to potential evapotranspiration. The expectancy of these kinds of climate change effects in central Europe was illustrated before the drought period of 2017 pp. was observed (Gudmundsson & Seneviratne, 2016; Hlásny et al., 2014).

Climate and weather also massively impact plant development by creating favorable conditions for fungi to infect crops. By infections with fusarium, leaf rust, or powdery mildew, the yield can be decreased significantly (Kiesling, 2015; Lim & Gaunt, 1986; Whelan et al., 1997). These fungi favor a moist and humid environment to infect their host (Bolton et al., 2008; Newton, 1993). Years with especially many days around the anthesis stage were observed for the years 2007 and especially 2016. The infection pressure could have an equal, maybe even bigger effect, on the constitution of the barley populations. The yield decline observed in the organic system in these two years indicates the effect of fungal pressure on the population. While the conventional population was protected by the application of fungicides, the organic population needed to balance the attack untreated. Resistance genes towards these diseases

can prevent a massive yield reduction in case of favorable attack conditions, as illustrated for powdery mildew (Helms Jørgensen, 1994).

Concluding, the weather and climate were highly variable and had a big impact on the development of the barley populations.

4.4 Pooling approaches for low coverage pool sequencing

When it comes to pool sequencing, so far, rather small genomes have been used as model species to validate the use of this method. The chosen coverage often exceeded reasonable levels in terms of costs and computational time for big genomes. The proposed method can be the key to close the gap between big and small genomes. As highlighted by Tilk et al. (2019), even 5x coverage could be sufficient. As a serious estimation of the allele frequency is impossible with low coverage, the implementation of parental information for haplotype construction is a valuable tool. By this, the coverage can be increased by magnitudes. For haplotype construction, using annotated genes or markers derive two major benefits to this idea. When haplotypes are constructed of genes, one can directly calculate and estimate the haplotype frequency of this gene itself. Based on the haplotype frequency variation and utilizing the information derived from the gene sequences of the founder, gene function loss, caused by a stop codon or amino acid change, can directly be detected. This way, false-positive candidates can be excluded, which only indicate the same frequency pattern due to linkage. For two reasons, marker position information can be useful as well. First, it creates the direct link to already identified QTLs by other studies. Chip SNP markers are still popular tools for quantitative trait detection (Naz et al., 2014; Oyiga et al., 2020; Schmalenbach & Pillen, 2009; Wang et al., 2010c). The information can directly transfer into the pooling results to highlight frequency changes caused by relevant agronomic traits. The second advantage is the potential presence of a genetic map. Physical maps have become more and more popular, but for this particular case, genetic maps still seem to have a high value to relate linkage information.

SNP calling is a highly complex task (Mielczarek & Szyda, 2016; Sandmann et al., 2017). Not all SNP identified as such by a variant caller are true polymorphisms. As it has been described before (Yao et al., 2020), SNP detection is a pretty challenging task that will always result in a fraction of false-positive SNPs. We handled this issue by only using SNPs in the analysis that are already reported in the ensemble SNP database for *Hordeum vulgare*. After this adjustment, we obtained an overall alternative base to reference base ratio in the pools of about 0.45 for WGS and MACE, which is in line with the expectation of a reasonably unbiased SNP calling (Brandt et al., 2015).

Extending the haplotypes from a point or location to a region showed benefits for this inbreeding crop in terms of accuracy. For outcrossing species or events, the window size might need to be reduced compared to the applied method. As most variants can be found in the intergenic regions, this approach saturates the haplotype with SNP and reads coverage. Therefore, it increases the accuracy of the allele frequency estimation.

We tested three different pool sequencing methods, including popular enzyme-based genotyping by sequencing, transcriptome sequencing, and whole-genome sequencing. When possible, we compared these to 21 KASP assay-designed SNP that were used for individual genotyping. This way of assessing the method's correctness was already used before (Rellstab et al., 2013). The transcriptome and whole-genome sequencing approach showed a high correlation without significant deviation between the pool and individual genotyping on haplotype level. The root means square error, but also the linear model, clearly indicated an improvement by the applied haplotyping approach compared to the single SNP correlation, especially for WGS. When the replicates of the pools were compared, transcriptome and genotyping by sequencing showed low correlation and significant deviation on the haplotyping gene

level ($p < 0.01$). The deviation of transcriptomic data might be related to variation in transcription levels, while the proper estimation of the frequency by GBS was unbalanced by the impossibility of PCR duplicate removal. The sequencing output of the GBS method revealed identical start and end positions of reads, which made it impossible to remove duplicates without removing more than 90% of the sequences.

The statistic test is based on two distributions. The population showed many loci typified by a low allele frequency, so the inclusion of a zero-inflated distribution makes sense. When the overall allele frequency of the haplotype is low (< 0.05), and it is constructed of multiple SNPs with low coverage, it is very likely to receive a substantial number of SNPs where the obtained frequency is 0. The zero-inflated model can handle those false negative detections. These kinds of distributions are commonly used in human resource studies. The data generated in these studies show similarities to the observed problem in this sequencing data (Famoye & Singh, 2006). It shows high sensitivity for haplotypes with low coverage per SNP and can contribute to the analysis. The negative binomial model handles the higher resolution data compared to the binary zero-inflated model and has to be assumed as the primary distribution with a more enormous impact on the interpretation of variation.

A general issue when starting such an analysis is the settlement with an appropriate sample size. According to Cochran's sampling guideline, we evaluated 300 genotypes per pool sample and created two replicates with different genotypes in each of the replicates. The sampling of the genotypes was performed using a hole punch with the intention to generate the same amount of leaf tissue for each genotype. This was regarded as a crucial step to ensure an unbiased representation of the population (Bélanger et al., 2016; Schlötterer et al., 2014). Pool sequencing approaches have been applied in various population studies for model plants (Fischer et al., 2013a; Gautier et al., 2013; Schlötterer et al., 2014; Turner et al., 2010), using high coverage sequencing approaches. RNAseq-based pool sequencing has also been described (Konczal et al., 2014), indicating in-line results with the generated results that have been made in this work. Unlike other species, such as *Arabidopsis thaliana*, the genome size and structure of barley results in pronounced challenges in terms of sequencing technique selection and costs.

Trans-generational allele variation was assessed by applying a haplotype-based pool sequencing approach. Based on the biparental origin, parental haplotypes were retraced within the populations. Pool sequencing has proven to be an essential tool for the analysis of structural and genetic variations in population studies (Turner et al., 2010) and can be vital in both breeding strategies and population genetics.

4.5 Evolution of the populations

The overall climatic conditions in central Germany are quite different from the climate the wild-type genotype has evolved in. Based on this fact, the flowering time response of both populations has adapted towards the cultivar genotype, which is adapted to local climate by early flowering. Such adaptations are expected to happen (Verhoeven et al., 2008), and little variation between the organic and conventional populations was observed, indicating a sweet spot for optimal flowering loci combinations. Nevertheless, wild alleles for flowering time are still present in the populations, indicating some kind of pleiotropic and epistatic effects of the alleles on flowering time and yields, as reported by (Maria von Korff et al., 2010; Wang et al., 2010a). Such effects might compensate for the occurrence of wild flowering alleles in the genotypes.

Few other studies have tried to investigate the effect of farming systems on the genetic constitution of a population (Knapp et al., 2020), but none have generated high-resolution output of a long-term experiment like demonstrated here. No variation in genetic constitutions has been observed in the study

by Knapp et al. (2020), most likely because too few genes and too short time intervals were observed. Alike this study, no variation in phenotype observations of flowering time and plant height were observed between the conventional and organic population. Beyond this, the nutrient transport, biotic resistance genes, and yield components were statistically not variant between the systems. Still, there can be a trend towards higher IAF reported in the organic population. Exemplarily, the here reported selection of candidate resistance alleles was statistically not more pronounced in the organic population (p = 0.07), although the average of resistance genes wtAF was 13.72% higher in the 23rd organic population (42.57% organic, 28.85% conventional). Although unexpected, this finding is in line with a prior report (Allard, 1988). Contrasting to these observations, highly variant IAF patterns were observed for the groups' root morphology, drought tolerance, and yield physiology. All these variations could be detected by the 16th generation onward, indicating a long observation period is required to generate genetic variations between an organic and conventional system on statistically significant levels. This was also highlighted by the PCA reported. The conventional and organic population deviated little until the 12th generation but tend to differentiate from each other after this generation. While the conventional population varied little by the 16th generation onward, indicating a state of equilibrium, the organic system still changes much after the 12th generation. This observation indicates that majorly important genes suppress environmentally relevant adaptation in the first couple of generations. This might be related to flowering time or yield formation adaptations, as these might have the biggest impact on the genetic fitness of the individuals of the population. These physiological groups might be the causal factor for the major IAF changes observed in both systems until the 12th generation. Once the optimal equilibrium of these genes is set, other, less relevant genes are altered in their allele frequency pattern of the population. Notably, the variation within the replicates is much higher in all organic population generations compared to the conventional population. The potential benefits of increased heterogeneity were expressed before as beneficial for organic systems (Wolfe et al., 2008). Although observed in multiple generations, it has to be mentioned that the variation in a population becomes smaller and smaller the more extensive the haplotypes windows are set. Large haplotypes seem to overcome any observed variation. Therefore a moderate level of haplotyping has to be chosen to not overcorrect any real variations.

From a genome-wide perspective, almost all loci have been found to be introgressed in the population. This highlights the power of the applied crossing scheme. Additionally, the observed IAF in the BC_2F_3 was only 0.43% apart from the theoretically expected value. Nevertheless, on special genomic regions, high deviations of the IAF were observed from the expected mean of 10.5%. This was especially remarkable for the chromosomes 2H and 5H, but also chromosome 7H was characterized by a deviation from the expected value. The telomeres on chromosome 5H and 2H were underrepresented, while both telomeres of chromosome 2H and the long arm telomere on 5H were overrepresented in the IAF. An underrepresentation was observed for the centromere of 7H as well. In the following generations, 25,000 genes had changed in the IAF significantly in both farming systems. This is more than 50% of all high confidence genes known in the reference. From this set of 25,000 genes, only a few had shown a positive selection of the ISR42-8 allele in the conventional system. Contrasting, many genes in the organic population were characterized by an increased IAF. This underlines the assumption that alleles derived from wild types were more likely to be beneficial in organically farmed environments. Nevertheless, conventionally farmed populations or varieties could benefit from wild alleles as well. All these observed changes might be the result of random effects, like drift or mutations (Wright, 1990). Mutations are expected to be rare, and the drift effect on the allele frequency variations over the generations was calculated as minor – both the time span as well as the number of individuals in the population were to be too big to result in substantial effects of random genetic drift on the population.

The selected testing interval of different years was applied to check if allele frequency variations can only be observed on long time spans or if the variation can be identified between subsequent generations,

too. Based on the obtained data, both patterns could be reported. The comparison of the IAF pattern between the organic 22nd and 23rd generation leads to the assumption that short-term evolutionary processes occurred and were traceable in the population IAF variation.

To finally give the IAF variation over the generations and between the environments a meaning, a candidate gene search was performed, based on QTLs identified by previous studies. By this, IAF variations of regions should be linked to genes and functions. Usually, a candidate gene selection is either error-prone or requires a lot of work effort to prove a gene as a real candidate. In this work, all information available was incorporated to get a good estimate. Besides the information of the QTL locus and its function, the wild-type allele frequency, as well as the functional prediction of the gene and the protein variation between the two parental haplotypes, were used for this purpose. 82 QTLs were identified. From these 82 QTLs, the most promising 95 candidate genes were identified and reported. Some regions had several highly likely candidates, while for others, no candidate could be identified. On chromosome 1H, a powdery mildew and fusarium resistance was reported. Two candidate genes were identified for the mildew locus *Mla*. One of these is already indicated as a disease candidate gene. The other is a *TRICHOME BIREFRINGENCE-LIKE 8* protein. This protein has been reported to be active in a plant's response to a fungal attack (Gao et al., 2017). The candidate gene for the fusarium resistance, a bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein, was reported likewise to be active in plant response to fungal attack (R. Liu et al., 2019). A yellow dwarf virus-related QTL on chromosome 6H with the candidate gene *HORVU6Hr1G024780* was reported as *Pathogenesis-related thaumatin superfamily protein*. The function of thaumatin proteins in pathogen resistance has been described before (J. Zhang et al., 2018). Leaf rust resistance loci were identified on chromosomes 4H and 7H. The candidate on chromosome 4H is *HORVU4Hr1G081400*, an *AP2-like ethylene-responsive transcription factor AIL5*. Ethylene responsive transcription factors were described in disease resistance reactions before (Gutterson & Reuber, 2004). The other candidate gene on chromosome 6H was a *flavin-dependent monooxygenase 1*, which was reported as an essential component in systemic resistance (Mishina & Zeier, 2006). Remarkably, none of the resistance genes were positively selected to high ratios of above 70% in the populations. This might indicate the potential costs of resistance (Bergelson & Purrington, 1996).

In regard to drought resistance, several candidate loci were identified. One candidate locus related to a *WUSCHEL related homeobox 12* protein was identified on chromosome 2H. The functional activity in drought resistance by formation and development of root systems was described by various studies before (Cheng et al., 2016). Another locus on chromosome 3H revealed an auxin response factor 2 as a likely candidate for drought stress response. The role of auxin in drought and salt stress response has been described before (Kang et al., 2018). Another candidate gene was found for the *Hv13GEIII/ Bmac 0357* locus on the same genome. The *Y-family DNA polymerase H* could contribute to increased seed yields under water limitation conditions (Nelson et al., 2007). A *Diacylglycerol kinase* is assumed to be the most likely candidate on the centromere of chromosome 3H. This kind of kinases has been reported to be active in drought tolerance response in Arabidopsis (Tan & Wang, 2020). All these genes were found on chromosome 3H, a chromosome that showed significant variation between the organic and conventional populations. Many of the identified candidate genes showed an effect in regard to root formation, though a positive relation towards nutrient uptake is assumed as well.

Chromosome 4H harbors further more drought-related genes. The drought-induced plant height reduction locus *Qhei.S42IL-4H* was associated with a candidate gene coding for a *Zinc finger MYM-type protein 1*. This gene was found to be 40-fold higher expressed in erect growth genotypes compared to prostrate growth type genotypes, indicating a role in the growth habit of the plant (Bing Zhang et al., 2017). As the two parental genotypes, Golf and ISR42-8, share the same deviation in growth type, a regulation based on the Zinc finger might be possible. Furthermore, the tiller and plant height-related

locus *Qtil.S42IL-4H* was described. The identified candidate gene is coding for a *respiratory burst oxidase homologous B* and has been described to be part of a superoxide catalyzing complex, which is related to reactive oxygen species (ROS). These have been shown to play a significant role in abiotic stress response (Rohollahi et al., 2018). Further candidate loci and genes have been detected on chromosome 4H and 5H, but a much more important one was found on chromosome 6H. The locus *Qagri.S42IL-6H* (Honsdorf et al., 2014) has an effect on many phenotypic traits, among which are biomass, growth plant, height, shoot area, and water use efficiency. For this locus, two candidate genes were identified, from which the first one is a *Photosystem II D2 protein*. The effect of drought stress on photosystem II has been described before (Y. E. Chen et al., 2016). The reduction under drought conditions can be an indication that the populations might respond in a different manner on the photosynthesis level. The plant in the organic population might be better supplied with water by the deeper root system, which results in less pronounced drought stress and so also less reduction of the photosystem II proteins. The other candidate gene of this locus is a *Peroxidase superfamily protein*. These proteins have been annotated to oxidative stress reduction processes under drought conditions (Veljovic-Jovanovic et al., 2006).

Several QTLs and candidate genes were identified for root morphological traits. The QTL *QRL.S42IL.1H.b*, related to the root length, was found on chromosome 1H. The potential candidate gene of this locus was identified to be a *Cytoplasmic membrane protein*. These proteins have been reported to be an essential part of root morphology in various functionalities (Boonsirichai et al., 2003; Nakagawa et al., 2007). A QTL locus identified on chromosome 2H was associated with the candidate gene *Ethylene-responsive transcription factor 10*. This locus was described as affecting the root volume. The impact of such ethylene-responsive transcription factors in the root system has been characterized before (Kawaharada et al., 2017), indicating that these play a major role in the infection and formation of nitrogen-fixation root nodules. The decreased IAF in the conventional but increased IAF in the organic system indicated why a potential role of this nitrogen fixation adaptation in the organic population might make sense.

A root dry weight locus on chromosome 3H was associated with the candidate gene *HORVU3Hr1G015840*, a *Serine/threonine-protein kinase*. Kinase activities have been reported to promote root growth and elongation, especially under drought conditions (Conley et al., 1997; Sheremet et al., 2010). Another candidate kinase gene was identified on the same chromosome (*Leucine-rich receptor-like protein kinase family protein (HORVU7Hr1G085790*)), indicating the vital role of these proteins in the formation of the root system. Two overlapping loci on chromosome 4H are related to root length and root dry weight. The underlying candidate gene is a *GDSL esterase/lipase*, which was described to have an essential influence on root growth and formation (Zhao et al., 2020). Furthermore, candidate genes for more loci were identified, like an *ABC transporter G family member 42* for the QTL locus *Qrdw.S42IL.7H.b*. ABC transporters were already described to be functional in the control of root development (Gaedeke et al., 2001).

Another group of QTLs was identified for yield components. For the Thresh-1 locus that was identified for reduced threshability of ears, a candidate gene coding for a *WUSCHEL related homeobox 1* was identified. This group of proteins was identified to play a major role in developmental processes, including spikelet formation (Cho et al., 2013). The brittleness locus brt2 on chromosome 3H can be related to a *cellulose synthase like G3*. This protein has been reported to be active in the secondary wall regulation, and therefore potentially might have an effect on the stability of the ear (Olins et al., 2018). For the major yield QTL on chromosome 5H, related to the ear size, the region is very big, and the identification of a single candidate gene is complex. Several wild alleles of various genes seem to have a beneficial effect on the fitness, as more than four loci within the cluster were observed to be positively

selected in the organic population. Contrasting, the conventional population did not show evidence of further selection and recombination in the region.

Most candidate genes and loci were related to yield physiological traits. The *QTil.1H* QTL locus was identified to be associated with the tillering of the plant. The candidate gene for this locus, a *sugar transporter 9*, was identified based on amino acid variations. Sugar transporters have been reported to be vital in cell wall biosynthesis and therefore promote plant growth (Baocai Zhang et al., 2011). The *HvGA20Ox3/Qgh.S42IL.1H* locus on chromosome 1H was also reported to be functional in tillering and plant height. The identified candidate gene for this is a *gibberellin 20 oxidase 2*, which has been reported to be a major gene in the green revolution (Spielmeyer et al., 2002). Various other candidate genes with relation to physiological response were identified. As these aspects underlying yield components are rather unspecific, more general candidate genes were identified, which might be active in several other traits and biological processes as well. Therefore, further candidate genes are not listed, as a reliable link of a candidate gene to a trait cannot be drawn based on state-of-the-art microbiological science in plants.

4.6 Genetic map from pooled sequencing

Before physical reference genomes based on a linear base-pair contig for each chromosome were available for species and crops, genetic maps were ubiquitous to determine variations in associations or QTL mapping approaches. Genetic maps give a highly informative impression of recombination events and crossing-over likelihoods. Several mapping functions are available and try to illustrate these recombination events in a linear order.

As most crossing-overs happen in the telomeres, a genetic map outweighs these telomeric regions in terms of resolution. Getting to know the recombination likelihood of two neighboring regions might be desired in crossing and breeding approaches. The usually utilized genetic maps are the consensual result of multiple crosses generated from various combinations to illustrate recombination likelihoods. To generate a genetic map, the haplotype and the allele frequency are required. This information was utilized to calculate the deviation of observed and expected haplotype combinations (Wei et al., 2020).

In this works' applied approach, only the allele frequency was available. Therefore, a classic calculation of a genetic map was not possible. To overcome the burden of missing haplotype information, the physical map was used as a replacement. The rationale behind this idea is mainly based on the allele frequency deviation. As the investigated system is a mainly self-pollinating species, recombination is unlikely to occur after the initial backcrossing between Golf and ISR42-8. The establishment of the cross indicates the primary source of recombination, and therefore, crossing-over counts should be highest in the first generations. As the recombination count cannot be increased anymore, crossing-overs can only be reduced over the generations.

Besides the recombination, the allele frequency is another factor highly affected by the crossing. Contrasting to the recombination, the allele frequency can be increased or decreased. Four factors could alter the allele frequency: mutation, migration, drift, and selection (Wright, 1990). These factors usually do not actively change the allele frequency of all genes. Selection by drought might just have a fitness effect on ten to fifty genes, but due to linkage, many more genes will be passively selected as well. The more crossing-overs have occurred in the establishment of the population, the smaller these selected genomic windows might become. Ultimately, the bigger the distance of two genes, the more likely these two will be recombined. Linked genes under selection will have the same allele frequency, while recombined genes will vary in their allele frequency – one gene will remain stable in its frequency, while the other will change. On a population base, we can assume that the more recombination is present between a particular selected and unselected gene, the higher the allele frequency difference can become. As illustrated before, the overall recombination cannot become higher over the generations. However,

the share of genotypes with such recombination at the particular locus can become higher, and therefore, the allele frequency deviates between the loci. A more targeted selection pressure will apply to a few genes and alleles in a less diverse environment. This will result in the selection of few genotypes with the best adaption towards the system. When few genotypes are selected, equivalently few recombinations will persist in the population. This will make it much harder for alternative selection forces to shift the allele and genotype frequency. One can relate that such a population is settled in an equilibrium state, with minor allele frequency changes due to few crossing overs present in the population.

This concept is known as soft sweep or hitchhiking (Stephan, 2019). Hitchhiking describes the allele frequency change, caused by selection, of neutral variants at linked sides to a selected locus. As more far such a locus is positioned to a selected one, as less likely this effect will become. Therefore, directional selection results in hitchhiking (Berry et al., 1991; Braverman et al., 1995). The soft sweep concept applies in a comparable way. The alleles were present in the population on a moderate frequency and got selected after the environmental conditions had changed.

There will be a more diverse selection pressure applied to the population in a more diverse system, with a less stable equilibrium in selection forces. These changing external selection forces prevent the reduction of crossing-overs in the population. Therefore, the diversity in the population should remain high and adjustable.

Based on the illustration above, the genotype frequency, recombination, physical distance on the genome, and the allele frequency are tightly associated. Therefore, the physical distance between loci is used to overcome missing haplotype information in the pooled sample.

In the populations, sharp transitions in allele frequencies between regions were observed. These changes were not coincidental, as such changes became more and more distinct over generations in this self-pollinating crop. Genetic drift has been calculated as a minor effect in the present study, mutations were negligible, and migration was avoided by the experimental design. Therefore, only selection can have altered the allele frequencies in the observed manner. Furthermore, the combination of selection and linkage can alter the frequency of gamete combinations. Based on these assumptions, the estimate of linkage disequilibrium (LD) depends on linkage and selection in these populations, as LD is a function of the allele frequency distribution. The expected value of LD at equilibrium decreases in the presence of selection (Bejarano et al., 2018; Qu et al., 2020).
Consequently, a rectified selection over consecutive generations will change the allele frequency diversity and subsequently LD. A more condensed map could have been generated for the conventional population, in line with the reduced variability of system parameters (such as nutrients, growth regulators, and pesticides), compared to organic farming.
The pericentromeric regions appeared to have the lowest recombination activity. The calculated genetic map additionally highlights variant recombination likelihoods on different chromosomes and even on the same chromosome between the two telomeres. The pericentromeric region and short telomere arm of chromosome 5H are fascinating, as little to no recombination was observed in the starting population. An incomplete representation of the ISR42-8 alleles on 5H was observed earlier (Schmalenbach et al., 2011). Contrasting to this observation, the telomere on the long arm of chromosome 5H is represented by a high degree of recombination. Such variant recombination events on the genome are relevant aspects that have to be considered when generating new lines by crossing. The mating intensity should be increased, having the information of low recombination events for some parental lines. With a higher crossing intensity, recombination events could be forced to happen more often, which ultimately results in a reduced linkage in the desired region. Again, chromosome 5H gives an excellent example of this pattern. A reported yield-related QTL (Watt et al., 2019) and the seed dormancy gene *SD1*

(Pourkheirandish & Komatsuda, 2007) are located in the telomeric region, tightly linked to each other. Few genotypes have overcome the linkage that used to prevent genotypes with these alleles from germinating. For those genotypes that recombine the Golf seed dormancy allele and the ISR42-8 yield alleles, the fitness advantage leads to a rapid increase of these genotypes' representation in the population.

Interestingly, not only macro but also micro-adjustments by recombination events could be found. An example of this can be found at the brittleness locus on chromosome 3H. The ISR42-8 allele was unfavorable in the population, so it was negatively selected over the generations. While the block in the conventional system was extensive and all genes around the brittleness locus have a Golf allele frequency close to 100%, the population developed in the organic system promotes genotypes with ISR42-8 alleles in this region. Genes close by the brittleness locus indicated recombination. While the brittleness alleles of ISR42-8 were down-selected, neighboring genes were up-selected in their ISR42-8 allele frequency. This indicates crossing-overs in this region, which allows closely located alleles to a locus with highly negative selection pressure to be positively selected.

In conclusion, a genetic map calculated from pool sequencing data can be a valuable tool to estimate the recombination of a crossing population in breeding approaches. It indicated sufficient or insufficient amounts of recombination in distinct regions of interest. Based on this information, unwanted allele combinations could be observed, and adjustments could be initiated.

4.7 Molecular answer to (a)biotic stress

Additionally to the observation of variant IAF for defined QTL loci, the adaptation of populations was assessed by applying a GO term enrichment approach commonly used in transcriptome analysis. Molecular genetic arrangements towards drought were observed in both populations on the allele frequency level. Five gene ontology classes were observed in both organic and conventional populations, indicating a strong impact of these classes on a drought stress response. The classes were DNA-dependent regulation of transcription (Talamè et al., 2007), transcription factor activity (Al Abdallat et al., 2014; Alexander et al., 2019; Marè et al., 2004), nucleus (Qin et al., 2017), DNA binding (Xue, 2003) and sequence-specific DNA binding (Abe et al., 1997). All these ontologies have been reported to be active in drought stress response. Mainly transcription factors have been observed to have a crucial role in the drought response reactions in both populations. While the reaction in the organic population was stronger, both environments seem to promote the reduction of wild transcription factor alleles under drought stress conditions. The role of transcription factors in drought stress response is known. It has many regulatory effects, e.g., on the root development (D. Chen et al., 2018) or a general drought resistance (Al Abdallat et al., 2014).

As this wild-type origins from an arid area, it should be naturally adapted to such conditions. In this background, this observation is somewhat unexpected. On a closer look, this reduction of wild-type allele frequency makes more sense. The wild-type, as described earlier, was observed to have an extraordinary growth of vegetative tissues without a generative growth increase. This unproductive habit might be associated with the previously mentioned gene ontologies. While useful in rainy and moist conditions to suppress weeds by increased vegetative growth, this habit is unfavorable in dry conditions. Furthermore, a regulation towards low vegetative growth under high water availability is unnecessary under constant drought conditions and, therefore, might not be selected for the wild-type.

Furthermore, five additional ontologies were identified in the conventional environment. These are related to protein binding, oxidation-reduction, oxidoreductase activity (Moran et al., 1994), zinc ion binding (Yang et al., 2009), and nucleic acid-binding. Especially oxidation-reduction is an interesting class, as the variation between the organic and conventional population is high. While there were minor variations overall three generations (16^{th}, $22^{nd,}$ and 23^{rd} generation) in the organic population compared

to the 12th generation in the IAF of the oxidative stress response, the conventional population was characterized by an increasing variety of IAF over the generations. While the IAF of the oxidative stress response is almost 10% in the 12th generation, it decreased to levels below 5% in the conventional population. This adaption might indicate that the conventionally farmed population suffers more form drought stress, leading to a higher level of reactive oxygen substances (ROS)(Miller et al., 2010), which have to be oxidized to prevent cell damages. Furthermore, zinc fingers have already been described by a potential candidate gene of a drought resistance locus in a previous chapter and observed to be transcribed in drought stress response (W. X. Liu et al., 2013).

The other gene ontology biotic stress response was only observed in the organic population. As it is assumed that the conventional population was treated with fungicides on demand, this group of genotypes should not have undergone a significant selection towards biotic stress. Nevertheless, application of fungicide in the conventional population was applied when a certain level of infection was exceeded, indicating that a small amount of infection will have occurred. 11 gene ontology classes were identified in both generations tested (12th and 22nd), but 34 additional classes were identified in the 22nd generation. More ontology classes might have been observed to be significant as the stress was much more severe in 2016 compared to 2007. The most manifested gene ontology classes in both generations were observed for proteolysis (Valueva & Mosolov, 2004), ADP and ATP binding (DeYoung & Innes, 2006), serine/threonine-protein kinase activity (Lin et al., 2015), and kinase activities (Goff & Ramonell, 2007), which have been reported to be active in biotic stress response mechanisms. This indicates a response beyond resistance genes which might modulate the reaction on a less energy-intensive base. As described previously, the costs of resistance can vary, depending on the background activities it causes. The enhancement of the observed gene ontology classes might lead to a less yield costly immune response of the plant (Bergelson & Purrington, 1996).

4.8 Yield evolution in the perspective of allele frequency variation and climate change

In general, the yield of the organic population was observed to be 30% lower over the entire period compared to the conventional population. The variation in the yield level was very high during the period, also in relation to the yield in Germany or state-wide. Two very interesting years are 2007 and 2016. In both years, the yield in the organic system was drastically decreased compared to the previous year. In contrast, only a moderate reduction in the yield level was observed in the conventional system. These two years were characterized by high fungal pressure and low temperatures at the beginning of the vegetation period, affecting the organic and the conventional population. This yield decrease has probably resulted in a rearrangement of the allele frequencies, especially concerning the biotic attack. In combination with the gene ontology response to biotic attacks, one must assume a non-persistent allele frequency adaption of resistance alleles in the organic population.

The highest yield in the organic system was observed in 2014, where almost 60 dt/ha were harvested. An average radiation sum described the year, one °C lower than average temperature, few drought days, and a slightly above average humidity level. As a comparable weather pattern characterized the previous years, an adaption of the population on the allele frequency level towards these conditions is assumed. Analogous to the organic population, 2010 to 2015 was characterized by a stable and high yield level. In this period, the populations might have adapted to moderate or no drought stress, with moderate levels of biotic attack. In the following year, 2016, the above-average precipitation and humidity resulted in a yield decrease and adaptation to high biotic attacks in the organic population. It has to be assumed that minor genetic adjustments have happened in the conventional population, as the PCA of the IAF did not reveal significant variations compared to the prior or following generations. Therefore, only the yield was decreased without a reshuffling of the allele frequency. The years 2017, 2018, and 2019 are the driest and warmest years in the entire period of the experiment. While the temperature in the vegetation period of 2017 was only moderately higher, the deviation of actual to potential evapotranspiration was

more than 60 mm. Comparable values only have been observed for the years 1996 and 1999. While the temperature was only moderately increased, the radiation was high. The following years 2018 and 2019, were characterized by more than two °C higher average temperatures, low precipitation, and a high deviation of the actual to potential evapotranspiration. This drastic variation of weather characteristics compared to 2016 might have led to an adaptation to biotic stress and low water stress but resulted in little adaptation towards drought conditions. In the years from 2017 to 2019, the yield declined in an almost linear way, indicating that the organic population of these generations might not be well suited to produce high yields under water-limited conditions. In contrast, the conventional population produced marginally lower yield in 2017, while the yield increased in 2018 and 2019 in direct comparison to 2016.

Concluding, the equilibrium state reached by the conventional population on the genetic level might make it more robust towards variant climatic and weather conditions. The organic population, still changing a lot, seems to suffer a lot in its ability to produce decent yield amounts under variant climatic conditions. Nevertheless, the ongoing adaption over more than the observed three years of yield decline could potentially produce a population with optimal adaption to drought stress.

4.9 Outlook: The use of gained information in future breeding and potential application of pool sequencing in related fields

This work is the first of its kind, highlighting the adaptation of populations towards their environment on a whole-genome scale. This analysis was conducted on two populations – one adapting to organic and conventional farming practices over twenty generations of consecutive selection towards these systems. Several studies have promoted the development of new organic varieties in the environment of their intended use (Anonymous V, 2017; Karn et al., 2020; Le Campion et al., 2020; Murphy et al., 2007; Shelton & Tracy, 2016). A transfer of this knowledge into breeding projects is limited so far, potentially because of big agrochemical companies' low financial interest in the field of organic cropping. The reuse and access of breeding material, free of charge or limited access, are pitfalls generated by the organic farming community that prevented extensive development of adapted genetic material for organic farming. This can be assumed to be one reason why the yield level is about ~30% lower in organic compared to conventional systems (Le Campion et al., 2020). A comparable lower yield level was also observed in the performed long-term experiment, indicating that even adapted varieties cannot make up for the lower yield caused by the reduced application of fertilizer or pesticides when the allelic diversity is not suitable or big enough. Nevertheless, the cross-generational yield reduction was less than 25% when extreme years were omitted from the yield data. Good soil with high cation exchange capacity might extend the yield gap between organic and conventional farming, as more of the applied fertilizer can be bound and exchanged between the root and soil. Generally, a primary concern of organic farmers has been the dependence of agro-chemical companies and seed material produced by these. A common approach is the reuse of seed material harvest in the previous year. This technique, combined with a postulated utilization of heterogeneous varieties might result in severe yield declines in extreme years. During the cropping process, a fight for resources leads to selection based on the highest fitness under the given environmental effects. As the climatic conditions in central Europe have shown to be highly variable in the past couple of years, an adaptation towards drought conditions might result in poor yield performance in a wet year with high fungal pressure. Regarding this approach, the reuse of seed material could be assumed to be prone to unintended and undesirable yield reduction events. Therefore, best practice could be the use of new seed material each year or at least for the following year, when the actual period was characterized by major deviations in climate conditions to previous year(s), to prevent unintended over-adaptations of heterogeneous varieties.

For sure, we defined some, not all, relevant aspects future approaches in organic breeding might consider focusing on. As it has been illustrated before, breeding and selection should be applied in an organically

farmed environment (Wolfe et al., 2008). Nevertheless, the gained knowledge in this study could enable breeders to target specific aspects of physiological habits. The often-postulated importance of biotic stress resistance seems to play a role in the development of a variety. However, the cost of resistance has to be quantified as well as the potential use in the variety to make intelligent decisions if the introgression of this resistance will finally pay off or not. The observed patterns in the two populations indicate that both the organic and the conventional system can benefit from biotic resistance. Though not statistically different, it has to be concluded that the organic system can benefit a bit more from the genetic implementation of resistance genes.

Nevertheless, multi-generational fixation of resistance responses could not have been observed, underlining the hypothesis of yield costs of resistance genes. Besides the classic response by resistance genes, a second layer of resistance response has been identified, indicating a wide range of micro genetic adjustments. Several gene ontologies positively selected for a limited time span of generations show evidence of adaptions and reactions towards biotic stressors apart from the classic resistance gene concept.

A much more important role seems to be the adaptation to nutrient uptake and mineralization processes in the soil. As described in a previous chapter, nutrients were applied by manure, which leads to late mineralization of nutrients. The main period of mineralization of this type of fertilizer was reported to be after the cropping period, or at least after the developmental demand of the spring barley is highest. The remaining mineralized nutrients in the soil potentially were washed out over the winter, down to deeper soil layers. This may result in unavailability for crops or genotypes with a shallow and flat root system (Lehrsch et al., 2016). Therefore, the genetic adjustment of the root system seems to play a significant role in the development of higher-yielding organic varieties. Deep rooting genotypes can access the leached nutrients in deeper soil layers, enabling them to produce more or bigger seeds.

Regarding climate change, this development might be a beneficial development, as these plants can access more rhizosphere areas and therefore more water, which leads to a better resilience of drought events. As described earlier, a drought effect only occurs if low air humidity and reduced water availability happen simultaneously. Therefore, deep root systems might be the key to increasing nutrient uptake and drought stress resistance and ultimately increasing the yield in arid growth conditions. Besides having a shallower root architecture, the conventional population was also observed to have a more distinct genetic adjustment reaction towards micro genetic adjustments. Several additional gene ontology classes were identified to have changed in their allele frequency after a drought year in the conventional population, compared to the organic population. This indicates the ability of genotypes to buffer certain extents of drought exposure and highlights the fact that a deep root system might be a desirable trait for conventional varieties in the future.

The suppression of weeds or other concurrences for light was also postulated as an essential aspect for the development of new organic varieties (Anonymous V, 2017). Like biotic resistances, the trade-off between increased vegetative biomass to shade weeds and the required input of energy and nutrients in a nutrient-limited environment has to be considered a vital aspect of yield maximization. According to our findings, there was no majorly increased biomass production of tillers observed in the organic population, neither on phenotypic nor on a genotypic scale. Many of these aspects remained undefined, but a significant variation between the conventional and organic systems could not be detected. Therefore, it has to be assumed that minor adjustments in the organic population have led to an adaptation. This might include an early emergence of leaves, as observed in the hydroponic experiment.

Besides the consideration of these classic traits, a high heterogeneity in the organic population has been observed. It seems to have formed a unique ecosystem, where each genotype covers its niche. The high

heterogeneity was observed both on genetic and phenotypic levels, indicating that a diverse set of genotypes is more robust in terms of stress adaption. This status might enable the population to adapt to variant environmental impacts without losing its genetic diversity. Compared to the organic population, the conventional one is not characterized by this extend of heterogeneity. This might indicate that the extent of environmental variation should be accompanied by an equal amount of heterogeneity in the genetic material. In terms of exporting this approach to breeding schedules or farmers, several things have to be mentioned. First, though it is allowed under European law to register a population variety these days, a robust phenotypic identification of such varieties is not possible. Therefore, genetic approaches like the one presented in this work should be applied. Second, the population can change in its allele frequency from one generation to another when it has undergone environmental stressors events. This might lead to a positive adaptation to a location but can also result in a negative over-adaption towards a particular environment. We observed this progress in the organic population twice, especially in the years following 2016. A population adapted to biotic attacks in 2016 was grown in an environment with little biotic stress but increased drought stress (2017). This drastic change resulted in further decreased yield, also in the following years, 2018 and 2019. Compared to this, the conventional system was able to retain its yield potential in the drought years, as no adaption to biotic attach in 2016 was necessary. This observation might highlight the need for new seed application every year to prevent suboptimal yield levels. It could be handled in a similar way hybrid varieties are created, just with one or two additional generations of segregation.

In conclusion, pool sequencing is a powerful method for the characterization of populations when the breeding behavior, population size, gene flow, and a reference sequence are known. A comparison of allele frequencies between populations provides information about different selection forces. Together with an expansion or condensation of the genetic map distances relative to the reference map, the selected regions indicate relevant traits and localize candidate genes without analyzing phenotyping data. The organic population showed a much higher genetic and phenotypic variability concerning the selection forces under organic or conventional farming. Since most of the differences between the farming systems were related to soil attributes, functional and morphological roots traits should be the focus of breeding for organic farming. Especially in organic farming, unpredictable weather strongly influences resource availability. Thus increased genetic heterogeneity is recommended. Genetic resources that have not undergone selection in high resource farming systems should be used to establish adapted cultivars.

4.10 General conclusion and evaluation of hypothesizes

Finally, the hypothesizes constructed before the analysis of the data are compared to the identified results. The hypothesis made on the allele frequency changes by non-random effects can be concluded as correct. The given data cannot answer the second hypothesis on the deviating speed of adaptions over the period of the experiment. Instead, it has to be assumed that we see a non-linear evolution of the allele frequency in the organic system. While significant fitness genes have influenced the allelic shift in the first couple of generations, minor adjustments towards more diversity have been observed in later generations. To answer this question, at least one, maybe two generations in the early period have to be sequenced.

The third hypothesis is that most of the alleles contributed to the populations derived from the cultivar parent is valid. Overall analyzed generations, the cultivar allele frequency was highest for most of the genes in both farming systems.

The fourth hypothesis, double backcrossing leads to a decreased linkage drag, can only partially be answered. A direct comparison with the same pooling strategy would be needed for a normal cross-population to evaluate this statement fully. Nevertheless, a high degree of recombination over most of

the chromosomes and regions has been observed. Therefore, the assumption of a sufficient recombination rate in the populations can be made.

The fifth hypothesis, wild allele can add important alleles to the conventionally and organically farmed genetic pool, can be answered positively. Both systems, but more dominantly the organic population, can benefit from the derived allelic variation added by the wild-type.

The sixth hypothesis that the genetic diversity declines over the generations to an equilibrium state has to be regarded as true for the conventional population. Contrasting to the observations made in the conventional system, the genetic diversity was increased in the organic population over the late generations, while being decreased in early generations.

The seventh hypothesis, the organic system can benefit more from wild alleles, was found to be true. Compared to the conventional system, the organic system is more related to the wild habitus and environment with low fertilization and increased overall heterogeneity of endogen stressors.

The eighth hypothesis, an advanced benefit concerning nutrient uptake and tolerances to biotic stressors in the organic system, cannot be fully supported. While the organic system has a more expressed and diverse root system, which might lead to improved nutrient uptake, the intra-plant variation of nutrient transporting genes was not observed to vary between the populations. The genetic variation for the traits might have been too slight or irrelevant in the empirical comparison. Furthermore, the resistance alleles known to be present in the wild-type were positively selected in the organic and conventional systems. Based on this observation, it has to be assumed that the importance of resistance genes in organic populations might not be too high.

The ninth hypothesis, the weather and climate impact the allelic formation of the populations, has to be considered significant. It was observed to be the second most crucial factor forming the allele frequency patterns in both populations in the late generations. It might have been the primary driver in the first generations. As these were not tested, no assumption can be made for this hypothesis. Nevertheless, the impact of the climate was relevant.

Lastly, the hypothesis made about the increased genetic diversity in the organic compared to the conventional system had been observed. The organic population had shown evidence of increased genetic diversity overall tested generations, underlining that the observations were not made by chance.

Chapter 5 – References

Abdel-Ghani, A. H., Parzies, H. K., Omary, A., & Geiger, H. H. (2004). Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. *Theoretical and Applied Genetics*, *109*(3), 588–595. https://doi.org/10.1007/s00122-004-1657-1

Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., & Shinozaki, K. (1997). Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell*, *9*(10), 1859–1868. https://doi.org/10.1105/tpc.9.10.1859

Afsharyan, N. P., Sannemann, W., Léon, J., & Ballvora, A. (2020). Effect of epistasis and environment on flowering time in barley reveals a novel flowering-delaying QTL allele. *Journal of Experimental Botany*, *71*(3), 893–906. https://doi.org/10.1093/jxb/erz477

Al Abdallat, A. M., Ayad, J. Y., Abu Elenein, J. M., Al Ajlouni, Z., & Harwood, W. A. (2014). Overexpression of the transcription factor HvSNAC1 improves drought tolerance in barley (Hordeum vulgare L.). *Molecular Breeding*, *33*(2), 401–414. https://doi.org/10.1007/s11032-013-9958-1

Alexander, R. D., Wendelboe-Nelson, C., & Morris, P. C. (2019). The barley transcription factor HvMYB1 is a positive regulator of drought tolerance. *Plant Physiology and Biochemistry*, *142*, 246–253. https://doi.org/10.1016/j.plaphy.2019.07.014

Allard, R. W. (1988). Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. *Journal of Heredity*, *79*(4), 225–238. https://doi.org/10.1093/oxfordjournals.jhered.a110503

Alqudah, A. M., Koppolu, R., Wolde, G. M., Graner, A., & Schnurbusch, T. (2016). The genetic architecture of barley plant stature. *Frontiers in Genetics*, *7*(JUN), 117. https://doi.org/10.3389/fgene.2016.00117

Anand, S., Mangano, E., Barizzone, N., Bordoni, R., Sorosina, M., Clarelli, F., Corrado, L., Boneschi, F. M., D'Alfonso, S., & De Bellis, G. (2016). Next generation sequencing of pooled samples: Guideline for variants' filtering. *Scientific Reports*, *6*(November 2015), 1–9. https://doi.org/10.1038/srep33735

Andersen, M. M., Landes, X., Xiang, W., Anyshchenko, A., Falhof, J., Østerberg, J. T., Olsen, L. I., Edenbrandt, A. K., Vedel, S. E., Thorsen, B. J., Sandøe, P., Gamborg, C., Kappel, K., & Palmgren, M. G. (2015). Feasibility of new breeding techniques for organic farming. In *Trends in Plant Science* (Vol. 20, Issue 7, pp. 426–434). Elsevier Ltd. https://doi.org/10.1016/j.tplants.2015.04.011

Andrews, S. (2010). FastQC - A quality control tool for high throughput sequence data. *Babraham Bioinformatics*, http://www.bioinformatics.babraham.ac.uk/projects/. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anonymous. (n.d.). *Barley Gene Function Annotation*. https://webblast.ipk-gatersleben.de/barley_ibsc/downloads/

Anonymous II. (n.d.). *Sim Drift calculator*. https://rdrr.io/github/rknx/drift/src/R/sim.drift.R

Anonymous III. (2020). *The Variant Call Format (VCF) Version 4.1 Specification*.

Anonymous V. (2013). Best Practice Guideline for Agriculture & Value Chains. *IFOAM*, 43.

Anonymous V. (2017). Position Paper 4. *IFOAM*. https://www.ifoam.bio/sites/default/files/position_paper_v01_web_0.pdf

Anonymous VI, U. (2008). Principles of Organic Agriculture. *Ifoam*, 5. https://doi.org/10.1300/J064v21n04_06

Aydinalp, C., & Cresser, M. S. (2008). Agriculture Land use change (including biomass burning) The Effects of Global Climate Change on Agriculture. *Agric. & Environ. Sci*, *3*(5), 672–676.

Badr, A., Müller, K., Schäfer-Pregl, R., El Rabey, H., Effgen, S., Ibrahim, H. H., Pozzi, C., Rohde, W., & Salamini, F. (2000). On the origin and domestication history of barley (Hordeum vulgare). *Molecular Biology and Evolution*, *17*(4), 499–510. https://doi.org/10.1093/oxfordjournals.molbev.a026330

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bedawy, I. M. A., Dehne, H. W., Léon, J., & Naz, A. A. (2018). Mining the global diversity of barley for Fusarium resistance using leaf and spike inoculations. *Euphytica*, *214*(1). https://doi.org/10.1007/s10681-017-2103-1

Bejarano, D., Martínez, R., Manrique, C., Parra, L. M., Martínez Rocha, J. F., Gómez, Y., Abuabara, Y., & Gallego, J. (2018). Linkage disequilibrium levels and allele frequency distribution in blanco orejinegro and romosinuano creole cattle using medium density snp chip data. *Genetics and Molecular Biology*, *41*(2), 426–433. https://doi.org/10.1590/1678-4685-gmb-2016-0310

Bélanger, S., Esteves, P., Clermont, I., Jean, M., & Belzile, F. (2016). Genotyping-by-Sequencing on Pooled Samples and its Use in Measuring Segregation Bias during the Course of Androgenesis in Barley. *The Plant Genome*, *9*(1), plantgenome2014.10.0073. https://doi.org/10.3835/plantgenome2014.10.0073

Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., Antipenko, A., Shang, L., Boisson, B., Casanova, J. L., & Abel, L. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5473–5478. https://doi.org/10.1073/pnas.1418631112

Bergelson, J., & Purrington, C. B. (1996). Surveying patterns in the cost of resistance in

plants. *American Naturalist*, *148*(3), 536–558. https://doi.org/10.1086/285938

Berry, A. J., Ajioka, J. W., & Kreitman, M. (1991). Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. *Genetics*, *129*(4).

Bertholdsson, N. O. (2005). Early vigour and allelopathy - Two useful traits for enhanced barley and wheat competitiveness against weeds. *Weed Research*, *45*(2), 94–102. https://doi.org/10.1111/j.1365-3180.2004.00442.x

Bertholdsson, N. O., Weedon, O., Brumlop, S., & Finckh, M. R. (2016). Evolutionary changes of weed competitive traits in winter wheat composite cross populations in organic and conventional farming systems. *European Journal of Agronomy*, *79*, 23–30. https://doi.org/10.1016/j.eja.2016.05.004

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Bhaskar, V., Weedon, O. D., & Finckh, M. R. (2019). Exploring the differences between organic and conventional breeding in early vigour traits of winter wheat. *European Journal of Agronomy*, *105*, 86–95. https://doi.org/10.1016/j.eja.2019.01.008

Bleidorn, C. (2017). *Phylo- genomics*.

Blighe, K. (2019). *PCAtools: PCAtools: Everything Principal Components Analysis*. https://github.com/kevinblighe/PCAtools

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolton, M. D., Kolmer, J. A., & Garvin, D. F. (2008). Wheat leaf rust caused by Puccinia triticina. In *Molecular Plant Pathology* (Vol. 9, Issue 5, pp. 563–575). John Wiley & Sons, Ltd. https://doi.org/10.1111/j.1364-3703.2008.00487.x

Boonsirichai, K., Sedbrook, J. C., Chen, R., Gilroy, S., & Masson, P. H. (2003). *The Plant Cell ARG1 Is a Peripheral Membrane Protein That Modulates Gravity-Induced Cytoplasmic Alkalinization and Lateral Auxin Transport in Plant Statocytes*. https://doi.org/10.1105/tpc.015560

Brandt, D. Y. C., Aguiar, V. R. C., Bitarello, B. D., Nunes, K., Goudet, J., & Meyer, D. (2015). Mapping bias overestimates reference allele frequencies at the HLA genes in the 1000 genomes project phase I data. *G3: Genes, Genomes, Genetics*, *5*(5), 931–941. https://doi.org/10.1534/g3.114.015784

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, *140*(2).

Burke, M. K., Dunham, J. P., Shahrestani, P., Thornton, K. R., Rose, M. R., & Long, A. D. (2010). Genome-wide analysis of a long-term evolution experiment with Drosophila.

Nature, *467*(7315), 587–590. https://doi.org/10.1038/nature09352

Byrne, S., Czaban, A., Studer, B., Panitz, F., Bendixen, C., & Asp, T. (2013). Genome Wide Allele Frequency Fingerprints (GWAFFs) of Populations via Genotyping by Sequencing. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0057438

Cai, S., Yu, G., Chen, X., Huang, Y., Jiang, X., Zhang, G., & Jin, X. (2013). Grain protein content variation and its association analysis in barley. *BMC Plant Biology*, *13*(1), 35. https://doi.org/10.1186/1471-2229-13-35

Castro, A. J., Gamba, F., German, S., Gonzalez, S., Hayes, P. M., Pereyra, S., & Perez, C. (2012). Quantitative trait locus analysis of spot blotch and leaf rust resistance in the BCD47×Baronesse barley mapping population. *Plant Breeding*, *131*(2), 258–266. https://doi.org/10.1111/j.1439-0523.2011.01930.x

Ceccarelli, S., Grando, S., & Van Leur, J. A. G. (1987). Genetic diversity in barley landraces from Syria and Jordan. *Euphytica*, *36*(2), 389–405. https://doi.org/10.1007/BF00041482

Chen, D., Chai, S., McIntyre, C. L., & Xue, G. P. (2018). Overexpression of a predominantly root-expressed NAC transcription factor in wheat roots enhances root length, biomass and drought tolerance. *Plant Cell Reports*, *37*(2), 225–237. https://doi.org/10.1007/s00299-017-2224-y

Chen, Y. E., Liu, W. J., Su, Y. Q., Cui, J. M., Zhang, Z. W., Yuan, M., Zhang, H. Y., & Yuan, S. (2016). Different response of photosystem II to short and long-term drought stress in Arabidopsis thaliana. *Physiologia Plantarum*, *158*(2), 225–235. https://doi.org/10.1111/ppl.12438

Cheng, S., Zhou, D.-X., & Zhao, Y. (2016). *WUSCHEL* -related homeobox gene *WOX11* increases rice drought resistance by controlling root hair formation and root system development. *Plant Signaling & Behavior*, *11*(2), e1130198. https://doi.org/10.1080/15592324.2015.1130198

Cho, S.-H., Yoo, S.-C., Zhang, H., Pandeya, D., Koh, H.-J., Hwang, J.-Y., Kim, G.-T., & Paek, N.-C. (2013). The rice *narrow leaf2* and *narrow leaf3* loci encode WUSCHEL-related homeobox 3A (OsWOX3A) and function in leaf, spikelet, tiller and lateral root development. *New Phytologist*, *198*(4), 1071–1084. https://doi.org/10.1111/nph.12231

Cockram, J., Hones, H., & O'Sullivan, D. M. (2011). Genetic variation at flowering time loci in wild and cultivated barley. *Plant Genetic Resources*, *9*(2), 264–267. https://doi.org/10.1017/S1479262111000505

Colmsee, C., Beier, S., Himmelbach, A., Schmutzer, T., Stein, N., Scholz, U., & Mascher, M. (2015). BARLEX - The barley draft genome explorer. *Molecular Plant*, *8*(6), 964–966. https://doi.org/10.1016/j.molp.2015.03.009

Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., Shaw, P., Bayer, M., Thomas, W., Marshall, D., Hedley, P., Tondelli, A., Pecchioni, N., Francia, E., Korzun, V., Walther, A., & Waugh, R. (2012a). Natural variation in a homolog of Antirrhinum

CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nature Genetics*, *44*(12), 1388–1391. https://doi.org/10.1038/ng.2447

Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., Shaw, P., Bayer, M., Thomas, W., Marshall, D., Hedley, P., Tondelli, A., Pecchioni, N., Francia, E., Korzun, V., Walther, A., & Waugh, R. (2012b). Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nature Genetics*, *44*(12), 1388–1391. https://doi.org/10.1038/ng.2447

Conley, T. R., Sharp, R. E., & Walker, J. C. (1997). Water deficit rapidly stimulates the activity of a protein kinase in the elongation zone of the maize primary root. *Plant Physiology*, *113*(1), 219–226. https://doi.org/10.1104/pp.113.1.219

Cookson, W. R., Cornforth, I. S., & Rowarth, J. S. (2002). Winter soil temperature (2-15°C) effects on nitrogen transformations in clover green manure amended or unamended soils; A laboratory and field study. *Soil Biology and Biochemistry*, *34*(10), 1401–1415. https://doi.org/10.1016/S0038-0717(02)00083-4

Council, T. H. E., The, O. F., Union, E., & Stabilisation, a. (2008). (Acts adopted under the EC Treaty/Euratom Treaty whose publication is obligatory). *Regulation*, *2007*(June), 1–9.

Cox, T. S. (1984). Expectations of means and genetic variances in backcross populations. *Theoretical and Applied Genetics*, *68*(1–2), 35–41. https://doi.org/10.1007/BF00252308

Crespo-Herrera, L. A., & Ortiz, R. (2015). Plant breeding for organic agriculture: Something new? In *Agriculture and Food Security* (Vol. 4, Issue 1, pp. 1–7). BioMed Central Ltd. https://doi.org/10.1186/s40066-015-0045-1

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., Bennett, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K. J., Gall, A., Girón, C. G., Gil, L., Grego, T., Haggerty, L., Haskell, E., … Flicek, P. (2019). Ensembl 2019. *Nucleic Acids Research*, *47*(D1), D745–D751. https://doi.org/10.1093/nar/gky1113

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Daten-analysen, R. (2018). *Besondere Ernte- und Qualitätsermittlung 2018*. https://www.bmel-statistik.de/fileadmin/daten/EQB-1002000-2018.pdf

David J. Connor, Robert S. Loomis, K. G. C. (2011). *Crop Ecology: Productivity and Management in Agricultural Systems*. Cambridge University Press.

de Mendiburu, F. (2020). *agricolae: Statistical Procedures for Agricultural Research*. https://cran.r-project.org/package=agricolae

De Ponti, T., Rijk, B., & Van Ittersum, M. K. (2012). The crop yield gap between organic and

conventional agriculture. *Agricultural Systems*, *108*, 1–9. https://doi.org/10.1016/j.agsy.2011.12.004

DeYoung, B. J., & Innes, R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. In *Nature Immunology* (Vol. 7, Issue 12, pp. 1243–1249). NIH Public Access. https://doi.org/10.1038/ni1410

Döring, T. F., Bocci, R., Hitchings, R., Howlett, S., Lammerts van Bueren, E. T., Pautasso, M., Raaijmakers, M., Rey, F., Stubsgaard, A., Weinhappel, M., Wilbois, K. P., Winkler, L. R., & Wolfe, M. S. (2012). The organic seed regulations framework in Europe-current status and recommendations for future development. In *Organic Agriculture* (Vol. 2, Issues 3–4, pp. 173–183). Springer. https://doi.org/10.1007/s13165-012-0034-7

Dragan, P., Kopahnke, D., Steffenson, B. J., Förster, J., König, J., Kilian, B., Plieske, J., Durstewitz, G., Korzun, V., Kraemer, I., Habekuss, A., Johnston, P., Pickering, R., & Ordon, F. (2013). Genetic Fine Mapping of a Novel Leaf Rust Resistance Gene and a Barley Yellow Dwarf Virus Tolerance (BYDV) Introgressed from Hordeum bulbosum by the Use of the 9K iSelect Chip. In *Advance in Barley Sciences* (pp. 269–284). Springer Netherlands. https://doi.org/10.1007/978-94-007-4682-4_23

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Ehrenreich, I. M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J. A., Gresham, D., Caudy, A. A., & Kruglyak, L. (2010). Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature*, *464*(7291), 1039–1042. https://doi.org/10.1038/nature08923

Famoye, F., & Singh, K. P. (2006). Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *Journal of Data Science*, *4*(May 2014), 117–130. https://doi.org/Yes

Fan, M., Zhu, J., Richards, C., Brown, K. M., & Lynch, J. P. (2003). Physiological roles for aerenchyma in phosphorus-stressed roots. *Functional Plant Biology*, *30*(5), 493–506. https://doi.org/10.1071/FP03046

Finckh, M. R., & Bhaskar, V. (2019). *Anpassungsprozesse in der Frühentwicklung von Weizenpopulationen über 11 Generationen an das Anbausystem*. http://orgprints.org

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., Holderegger, R., & Widmer, A. (2013a). Population genomic footprints of selection and associations with climate in natural populations of Arabidopsis halleri from the Alps. *Molecular Ecology*, *22*(22), 5594–5607. https://doi.org/10.1111/mec.12521

Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., Holderegger, R., & Widmer, A. (2013b). Population genomic footprints of selection and associations with climate in natural populations of Arabidopsis halleri from the Alps. *Molecular Ecology*, *22*(22), 5594–5607. https://doi.org/10.1111/mec.12521

Fließbach, A., Oberholzer, H. R., Gunst, L., & Mäder, P. (2007). Soil organic matter and biological soil quality indicators after 21 years of organic and conventional farming. *Agriculture, Ecosystems and Environment*, *118*(1–4), 273–284. https://doi.org/10.1016/j.agee.2006.05.022

Gaedeke, N., Klein, M., Kolukisaoglu, U., Forestier, C., Müller, A., Ansorge, M., Becker, D., Mamnun, Y., Kuchler, K., Schulz, B., Mueller-Roeber, B., & Martinoia, E. (2001). The Arabidopsis thaliana ABC transporter AtMRP5 controls root development and stomata movement. *EMBO Journal*, *20*(8), 1875–1887. https://doi.org/10.1093/emboj/20.8.1875

Gao, Y., He, C., Zhang, D., Liu, X., Xu, Z., Tian, Y., Liu, X. H., Zang, S., Pauly, M., Zhou, Y., & Zhang, B. (2017). Two trichome birefringence-like proteins mediate xylan acetylation, which is essential for leaf blight resistance in rice. *Plant Physiology*, *173*(1), 470–481. https://doi.org/10.1104/pp.16.01618

Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhué, C., & Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: Pool-versus individual-based genotyping. *Molecular Ecology*, *22*(14), 3766–3779. https://doi.org/10.1111/mec.12360

Goff, K. E., & Ramonell, K. M. (2007). The role and regulation of receptor-like kinases in plant defense. *Gene Regulation and Systems Biology*, *1*, 167–175. http://www.ncbi.nlm.nih.gov/pubmed/19936086

Gordon, A., & Hannon, G. (n.d.). *Fastx-Toolkit*. https://dl.acm.org/doi/10.1093/bioinformatics/bts507

Guan, J. C., Koch, K. E., Suzuki, M., Wu, S., Latshaw, S., Petruff, T., Goulet, C., Klee, H. J., & McCarty, D. R. (2012). Diverse roles of strigolactone signaling in maize architecture and the uncoupling of a branching-specific subnetwork. *Plant Physiology*, *160*(3), 1303–1317. https://doi.org/10.1104/pp.112.204503

Gudmundsson, L., & Seneviratne, S. I. (2016). Anthropogenic climate change affects meteorological drought risk in Europe. *Environmental Research Letters*, *11*(4), 44005. https://doi.org/10.1088/1748-9326/11/4/044005

Guo, Y., Cai, Q., Li, C., Li, J., Li, C. I., Courtney, R., Zheng, W., & Long, J. (2013). An evaluation of allele frequency estimation accuracy using pooled sequencing data. *International Journal of Computational Biology and Drug Design*, *6*(4), 279–293. https://doi.org/10.1504/IJCBDD.2013.056709

Gutterson, N., & Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. In *Current Opinion in Plant Biology* (Vol. 7, Issue 4, pp. 465–471). Elsevier Current Trends. https://doi.org/10.1016/j.pbi.2004.04.007

Hedden, P. (2003). The genes of the Green Revolution. *Trends in Genetics*, *19*(1), 5–9. https://doi.org/10.1016/S0168-9525(02)00009-4

Helms Jørgensen, J. (1994). Genetics of powdery mildew resistance in barley. *Critical*

Reviews in Plant Sciences, *13*(1), 97–119. https://doi.org/10.1080/07352689409701910

Hidalgo, H. G., Cayan, D. R., & Dettinger, M. D. (2005). Sources of variability of evapotranspiration in California. *Journal of Hydrometeorology*, *6*(1), 3–19. https://doi.org/10.1175/JHM-398.1

Hlásny, T., Mátyás, C., Seidl, R., Kulla, L., Merganičová, K., Trombik, J., Dobor, L., Barcza, Z., & Konôpka, B. (2014). Climate change increases the drought risk in Central European forests: What are the options for adaptation? *Forestry Journal*, *60*(1), 5–18. https://doi.org/10.2478/forj-2014-0001

Hoagland, D. R., & Arnon, D. I. (1938). Growing plants without soil by the water-culture method. *Growing Plants without Soil by the Water-Culture Method.*

Hodge, A., Robinson, D., Griffiths, B. S., & Fitter, A. H. (1999). Why plants bother: Root proliferation results in increased nitrogen capture from an organic patch when two grasses compete. *Plant, Cell and Environment*, *22*(7), 811–820. https://doi.org/10.1046/j.1365-3040.1999.00454.x

Honda, I., Seto, H., Turuspekov, Y., Watanabe, Y., & Yoshida, S. (2006). Inhibitory effects of jasmonic acid and its analogues on barley (Hordeum vulgare L.) anther extrusion. *Plant Growth Regulation*, *48*(3), 201–206. https://doi.org/10.1007/s10725-006-0003-9

Honsdorf, N., March, T. J., Berger, B., Tester, M., & Pillen, K. (2014). High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. *PLoS ONE*, *9*(5). https://doi.org/10.1371/journal.pone.0097047

Honsdorf, N., March, T. J., & Pillen, K. (2017). QTL controlling grain filling under terminal drought stress in a set of wild barley introgression lines. *PLoS ONE*, *12*(10). https://doi.org/10.1371/journal.pone.0185983

Israel, G. D. (1992). Determining Sample Size. *Agricultural Education and Communication Department Florida*.

Itoh, H., Tatsumi, T., Sakamoto, T., Otomo, K., Toyomasu, T., Kitano, H., Ashikari, M., Ichihara, S., & Matsuoka, M. (2004). A rice semi-dwarf gene, Tan-Ginbozu (D35), encodes the gibberellin biosynthesis enzyme, ent-kaurene oxidase. *Plant Molecular Biology*, *54*(4), 533–547. https://doi.org/10.1023/B:PLAN.0000038261.21060.47

Jackman, S., Tahk, A., Zeileis, A., Maimone, C., & Fearon, J. (2013). *Political Science Computational Laboratory, Stanford University: Package ' pscl .'* http://pscl.stanford.edu/

Jannoura, R., Bruns, C., & Joergensen, R. G. (2013). Organic fertilizer effects on pea yield, nutrient uptake, microbial root colonization and soil microbial biomass indices in organic farming systems. *European Journal of Agronomy*, *49*(August), 32–41. https://doi.org/10.1016/j.eja.2013.03.002

Jones, H., Leigh, F. J., Mackay, I., Bower, M. A., Smith, L. M. J., Charles, M. P., Jones, G., Jones, M. K., Brown, T. A., & Powell, W. (2008). Population-based resequencing

reveals that the flowering time adaptation of cultivated barley originated east of the fertile crescent. *Molecular Biology and Evolution*, *25*(10), 2211–2219. https://doi.org/10.1093/molbev/msn167

Kahler, A. L., Clegg, M. T., & Allard, R. W. (1975). Evolutionary Changes in the Mating System of an Experimental Population of Barley (Hordeum vulgare L.). *Proceedings of the National Academy of Sciences*, *72*(3), 943–946. https://doi.org/10.1073/pnas.72.3.943

Kang, C., He, S., Zhai, H., Li, R., Zhao, N., & Liu, Q. (2018). A Sweetpotato Auxin Response Factor Gene (IbARF5) Is Involved in Carotenoid Biosynthesis and Salt and Drought Tolerance in Transgenic Arabidopsis. *Frontiers in Plant Science*, *9*, 1307. https://doi.org/10.3389/fpls.2018.01307

Karn, R., KC, M., Lamichhane, A., & Bhandari, D. (2020). A Review On Corn Breeding For Organic And Sustainable Agriculture. *Maize Genomics and Genetics*, *10*(0). http://www.cropscipublisher.com/index.php/mgg/article/view/3748

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kawaharada, Y., James, E. K., Kelly, S., Sandal, N., & Stougaard, J. (2017). The ethylene responsive factor required for nodulation 1 (ERN1) transcription factor is required for infection-thread formation in lotus japonicus. *Molecular Plant-Microbe Interactions*, *30*(3), 194–204. https://doi.org/10.1094/MPMI-11-16-0237-R

Kessner, D., Turner, T. L., & Novembre, J. (2013). Maximum likelihood estimation of frequencies of known haplotypes from pooled sequence data. *Molecular Biology and Evolution*, *30*(5), 1145–1158. https://doi.org/10.1093/molbev/mst016

Kiesling, R. L. (2015). *The Diseases of Barley* (pp. 269–312). John Wiley & Sons, Ltd. https://doi.org/10.2134/agronmonogr26.c10

Kilian, B., Özkan, H., Kohl, J., Von Haeseler, A., Barale, F., Deusch, O., Brandolini, A., Yucel, C., Martin, W., & Salamini, F. (2006). Haplotype structure at seven barley genes: Relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. *Molecular Genetics and Genomics*, *276*(3), 230–241. https://doi.org/10.1007/s00438-006-0136-6

Knapp, S., Döring, T. F., Jones, H. E., Snape, J., Wingen, L. U., Wolfe, M. S., Leverington-Waite, M., & Griffiths, S. (2020). Natural Selection Towards Wild-Type in Composite Cross Populations of Winter Wheat. *Frontiers in Plant Science*, *10*(February), 1–11. https://doi.org/10.3389/fpls.2019.01757

Komatsuda, T., Maxim, P., Senthil, N., & Mano, Y. (2004). High-density AFLP map of nonbrittle rachis 1 (btr1) and 2 (btr2) genes in barley (Hordeum vulgare L.). *Theoretical and Applied Genetics*, *109*(5), 986–995. https://doi.org/10.1007/s00122-004-1710-0

Konczal, M., Koteja, P., Stuglik, M. T., Radwan, J., & Babik, W. (2014). Accuracy of allele frequency estimation using pooled RNA-Seq. *Molecular Ecology Resources*, *14*(2), 381–392. https://doi.org/10.1111/1755-0998.12186

Lal, R., Delgado, J. A., Gulliford, J., Nielsen, D., Rice, C. W., & Scott Van Pelt, R. (2012). Adapting agriculture to drought and extreme events. *Journal of Soil and Water Conservation*, *67*(6), 162–166. https://doi.org/10.2489/jswc.67.6.162A

Lammerts Van Bueren, E. T., Jones, S. S., Tamm, L., Murphy, K. M., Myers, J. R., Leifert, C., & Messmer, M. M. (2011). The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: A review. *NJAS - Wageningen Journal of Life Sciences*, *58*(3–4), 193–205. https://doi.org/10.1016/j.njas.2010.04.001

Landwirtschaft, B. für. (2013). *Abstammungs-Katalog der Gerstensorten Abstammungs-Katalog der Gerstensorten*.

Laurie, D. A., Pratchett, N., Bezant, J. H., & Snape, J. W. (1995). RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter X spring barley (Hordeum vulgare L.) cross. *Genome*, *38*(3), 575–585. https://doi.org/10.1139/g95-074

Le Campion, A., Oury, F. X., Heumez, E., & Rolland, B. (2020). Conventional versus organic farming systems: dissecting comparisons to improve cereal organic breeding strategies. In *Organic Agriculture* (Vol. 10, Issue 1, pp. 63–74). Springer. https://doi.org/10.1007/s13165-019-00249-3

Lehrsch, G. A., Brown, B., Lentz, R. D., Johnson-Maynard, J. L., & Leytem, A. B. (2016). Winter and growing season nitrogen mineralization from fall-applied composted or stockpiled solid dairy manure. *Nutrient Cycling in Agroecosystems*, *104*(2), 125–142. https://doi.org/10.1007/s10705-015-9755-9

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *00*(00), 1–3. http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liller, C. B., Walla, A., Boer, M. P., Hedley, P., Macaulay, M., Effgen, S., von Korff, M., van Esse, G. W., & Koornneef, M. (2017). Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. *Theoretical and Applied Genetics*, *130*(2), 269–281. https://doi.org/10.1007/s00122-016-2807-y

Lim, L. G., & Gaunt, R. E. (1986). The effect of powdery mildew (Erysiphe graminis f. sp. hordei) and leaf rust (Puccinia hordei) on spring barley in New Zealand. II. Apical development and yield potential. *Plant Pathology*, *35*(1), 54–60. https://doi.org/10.1111/j.1365-3059.1986.tb01980.x

Lin, Z. J. D., Liebrand, T. W. H., Yadeta, K. A., & Coaker, G. (2015). PBL13 is a

serine/threonine protein kinase that negatively regulates arabidopsis immune responses. *Plant Physiology*, *169*(4), 2950–2962. https://doi.org/10.1104/pp.15.01391

Liu, F., Jensen, C. R., & Andersen, M. N. (2003). Hydraulic and chemical signals in the control of leaf expansion and stomatal conductance in soybean exposed to drought stress. *Functional Plant Biology*, *30*(1), 65–73. https://doi.org/10.1071/FP02170

Liu, R., Weng, K., Dou, M., Chen, T., Yin, X., Li, Z., Li, T., Zhang, C., Xiang, G., Liu, G., & Xu, Y. (2019). Transcriptomic analysis of Chinese wild Vitis pseudoreticulata in response to Plasmopara viticola. *Protoplasma*, *256*(5), 1409–1424. https://doi.org/10.1007/s00709-019-01387-x

Liu, W. X., Zhang, F. C., Zhang, W. Z., Song, L. F., Wu, W. H., & Chen, Y. F. (2013). Arabidopsis Di19 functions as a transcription factor and modulates PR1, PR2, and PR5 expression in response to drought stress. *Molecular Plant*, *6*(5), 1487–1502. https://doi.org/10.1093/mp/sst031

Lobell, D. B., & Field, C. B. (2008). Estimation of the carbon dioxide (CO2) fertilization effect using growth rate anomalies of CO2 and crop yields since 1961. *Global Change Biology*, *14*(1), 39–45. https://doi.org/10.1111/j.1365-2486.2007.01476.x

Lockeretz, W. (2007). Organic farming: An international history. In *Organic Farming: An International History*. https://doi.org/10.1093/erae/jbp005

Lockeretz, W., Shearer, G., & Kohl, D. H. (1981). Organic farming in the corn belt. *Science*, *211*(4482), 540–547. https://doi.org/10.1126/science.211.4482.540

Long, Q., Jeffares, D. C., Zhang, Q., Ye, K., Nizhynska, V., Ning, Z., Tyler-Smith, C., & Nordborg, M. (2011). PoolHap: Inferring haplotype frequencies from pooled samples by next generation sequencing. *PLoS ONE*, *6*(1), 1–7. https://doi.org/10.1371/journal.pone.0015292

Löschenberger, F., Fleck, A., Grausgruber, H., Hetzendorfer, H., Hof, G., Lafferty, J., Marn, M., Neumayer, A., Pfaffinger, G., & Birschitzky, J. (2008). Breeding for organic agriculture: The example of winter wheat in Austria. *Euphytica*, *163*(3), 469–480. https://doi.org/10.1007/s10681-008-9709-2

Lotter, D. W., Seidel, R., & Liebhardt, W. (2003). The performance of organic and conventional cropping systems in an extreme climate year. *American Journal of Alternative Agriculture*, *18*(3), 146–154. https://doi.org/10.1079/AJAA200345

Lu, G., Casaretto, J. A., Ying, S., Mahmood, K., Liu, F., Bi, Y. M., & Rothstein, S. J. (2017). Overexpression of OsGATA12 regulates chlorophyll content, delays plant senescence and improves rice yield under high density planting. *Plant Molecular Biology*, *94*(1–2), 215–227. https://doi.org/10.1007/s11103-017-0604-x

Lüpken, T., Stein, N., Perovic, D., Habekuß, A., Serfling, A., Krämer, I., Hähnel, U., Steuernagel, B., Scholz, U., Ariyadasa, R., Martis, M., Mayer, K., Niks, R. E., Collins, N. C., Friedt, W., & Ordon, F. (2014). High-resolution mapping of the barley Ryd3 locus

controlling tolerance to BYDV. *Molecular Breeding*, *33*(2), 477–488. https://doi.org/10.1007/s11032-013-9966-1

Lyngkjær, M. F., Newton, A. C., Atzema, J. L., & Baker, S. J. (2000). The Barley mlo-gene : an important powdery mildew resistance source Plant Genetics and Breeding. *Crop Research*, *20*, 745–756.

Manschadi, A. M., Hammer, G. L., Christopher, J. T., & DeVoil, P. (2008). Genotypic variation in seedling root architectural traits and implications for drought adaptation in wheat (Triticum aestivum L.). *Plant and Soil*, *303*(1–2), 115–129. https://doi.org/10.1007/s11104-007-9492-1

Marè, C., Mazzucotelli, E., Crosatti, C., Francia, E., Stanca, A. M., & Cattivelli, L. (2004). Hv-WRKY38: A new transcription factor involved in cold- and drought-response in barley. *Plant Molecular Biology*, *55*(3), 399–416. https://doi.org/10.1007/s11103-004-0906-7

Marzec, M., & Alqudah, A. M. (2018). Key hormonal components regulate agronomically important traits in barley. *International Journal of Molecular Sciences*, *19*(3), 1–12. https://doi.org/10.3390/ijms19030795

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P. E., Russell, J., Bayer, M., Ramsay, L., Liu, H., Haberer, G., Zhang, X. Q., Zhang, Q., Barrero, R. A., Li, L., Taudien, S., … Stein, N. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature*, *544*(7651), 427–433. https://doi.org/10.1038/nature22043

Mayer, K. F. X., Waugh, R., Langridge, P., Close, T. J., Wise, R. P., Graner, A., Matsumoto, T., Sato, K., Schulman, A., Ariyadasa, R., Schulte, D., Poursarebani, N., Zhou, R., Steuernagel, B., Mascher, M., Scholz, U., Shi, B., Madishetty, K., Svensson, J. T., … Stein, N. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, *491*(7426), 711–716. https://doi.org/10.1038/nature11543

Metzker, M. L. (2010). Sequencing technologies the next generation. In *Nature Reviews Genetics* (Vol. 11, Issue 1, pp. 31–46). https://doi.org/10.1038/nrg2626

Meyer, W. S., & Green, G. C. (1980). Water Use by Wheat and Plant Indicators of Available Soil Water 1. *Agronomy Journal*, *72*(2), 253–257. https://doi.org/10.2134/agronj1980.00021962007200020002x

Mielczarek, M., & Szyda, J. (2016). Review of alignment and SNP calling algorithms for next-generation sequencing data. *Journal of Applied Genetics*, *57*(1), 71–79. https://doi.org/10.1007/s13353-015-0292-7

Miller, G., Suzuki, N., Ciftci-Yilmaz, S., & Mittler, R. (2010). Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant, Cell and Environment*, *33*(4), 453–467. https://doi.org/10.1111/j.1365-3040.2009.02041.x

Minasny, B., & McBratney, A. B. (2018). Limited effect of organic matter on soil available

water capacity. *European Journal of Soil Science*, *69*(1), 39–47. https://doi.org/10.1111/ejss.12475

Mishina, T. E., & Zeier, J. (2006). The Arabidopsis flavin-dependent monooxygenase FMO1 is an essential component of biologically induced systemic acquired resistance. *Plant Physiology*, *141*(4), 1666–1675. https://doi.org/10.1104/pp.106.081257

Mohammed, K. A. H., & Léon, J. (2004). Improving crop varieties of spring barley for drought and heat tolerance with AB-QTL-analysis. *Dissertation University of Bonn*, 150. http://www.secheresse.info/article.php3?id_article=2706

Moran, J. F., Becana, M., Iturbe-Ormaetxe, I., Frechilla, S., Klucas, R. V., & Aparicio-Tejo, P. (1994). Drought induces oxidative stress in pea plants. *Planta*, *194*(3), 346–352. https://doi.org/10.1007/BF00197534

Morrell, P. L., & Clegg, M. T. (2007). Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3289–3294. https://doi.org/10.1073/pnas.0611377104

Morrell, P. L., Toleno, D. M., Lundy, K. E., & Clegg, M. T. (2005). Low levels of linkage disequilibrium in wild barley (Hordeum vulgare ssp. spontaneum) despite high rates of self-fertilization. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(7), 2442–2447. https://doi.org/10.1073/pnas.0409804102

Mulki, M. A., Bi, X., & von Korff, M. (2018). Flowering locus T3 controls spikelet initiation but not floral development. *Plant Physiology*, *178*(3), 1170–1186. https://doi.org/10.1104/pp.18.00236

Muñoz-Amatriaín, M., Cuesta-Marcos, A., Endelman, J. B., Comadran, J., Bonman, J. M., Bockelman, H. E., Chao, S., Russell, J., Waugh, R., Hayes, P. M., & Muehlbauer, G. J. (2014). The USDA barley core collection: Genetic diversity, population structure, and potential for genome-wide association studies. *PLoS ONE*, *9*(4), 1–13. https://doi.org/10.1371/journal.pone.0094688

Muñoz-Amatriaín, M., Moscou, M. J., Bhat, P. R., Svensson, J. T., Bartoš, J., Suchánková, P., Šimková, H., Endo, T. R., Fenton, R. D., Lonardi, S., Castillo, A. M., Chao, S., Cistué, L., Cuesta-Marcos, A., Forrest, K. L., Hayden, M. J., Hayes, P. M., Horsley, R. D., Makoto, K., … Close, T. J. (2011). An Improved Consensus Linkage Map of Barley Based on Flow-Sorted Chromosomes and Single Nucleotide Polymorphism Markers. *The Plant Genome*, *4*(3), 238–249. https://doi.org/10.3835/plantgenome2011.08.0023

Murphy, K. M., Campbell, K. G., Lyon, S. R., & Jones, S. S. (2007). Evidence of varietal adaptation to organic farming systems. *Field Crops Research*, *102*(3), 172–177. https://doi.org/10.1016/j.fcr.2007.03.011

Muzammil, S., Shrestha, A., Dadshani, S., Pillen, K., Siddique, S., Léon, J., & Naz, A. A. (2018). An Ancestral Allele of Pyrroline-5-carboxylate synthase1 Promotes Proline Accumulation and Drought Adaptation in Cultivated Barley. *Plant Physiology*, *178*(2),

771–782. https://doi.org/10.1104/pp.18.00169

Nadolska-Orczyk, A., Rajchel, I. K., Orczyk, W., & Gasparis, S. (2017). Major genes determining yield-related traits in wheat and barley. *Theoretical and Applied Genetics*, *130*(6), 1081–1098. https://doi.org/10.1007/s00122-017-2880-x

Nakagawa, Y., Katagiri, T., Shinozaki, K., Qi, Z., Tatsumi, H., Furuichi, T., Kishigami, A., Sokabe, M., Kojima, I., Sato, S., Kato, T., Tabata, S., Iida, K., Terashima, A., Nakano, M., Ikeda, M., Yamanaka, T., & Iida, H. (2007). Arabidopsis plasma membrane protein crucial for Ca2+ influx and touch sensing in roots. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3639–3644. https://doi.org/10.1073/pnas.0607703104

Nakhforoosh, A., Grausgruber, H., Kaul, H. P., & Bodner, G. (2014). Wheat root diversity and root functional characterization. *Plant and Soil*, *380*(1), 211–229. https://doi.org/10.1007/s11104-014-2082-0

Naz, A. A., Arifuzzaman, M., Muzammil, S., Pillen, K., & Léon, J. (2014). Wild barley introgression lines revealed novel QTL alleles for root and related shoot traits in the cultivated barley (Hordeum vulgare L.). *BMC Genetics*, *15*(1), 1–12. https://doi.org/10.1186/s12863-014-0107-6

Nelson, D. E., Repetti, P. P., Adams, T. R., Creelman, R. A., Wu, J., Warner, D. C., Anstrom, D. C., Bensen, R. J., Castiglioni, P. P., Donnarummo, M. G., Hinchey, B. S., Kumimoto, R. W., Maszle, D. R., Canales, R. D., Krolikowski, K. A., Dotson, S. B., Gutterson, N., Ratcliffe, O. J., & Heard, J. E. (2007). Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(42), 16450–16455. https://doi.org/10.1073/pnas.0707193104

Nerem, R. S., Beckley, B. D., Fasullo, J. T., Hamlington, B. D., Masters, D., & Mitchum, G. T. (2018). Climate-change–driven accelerated sea-level rise detected in the altimeter era. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9), 2022–2025. https://doi.org/10.1073/pnas.1717312115

Neumann, K., Verburg, P. H., Stehfest, E., & Müller, C. (2010). The yield gap of global grain production: A spatial analysis. *Agricultural Systems*, *103*(5), 316–326. https://doi.org/10.1016/j.agsy.2010.02.004

Nevo, E., Beiles, A., Gutterman, Y., Storch, N., & Kaplan, D. (1984). Genetic resources of wild cereals in Israel and vicinity. II. Phenotypic variation within and between populations of wild barley, Hordeum spontaneum. *Euphytica*, *33*(3), 737–756. https://doi.org/10.1007/BF00021901

Nevo, E., & Chen, G. (2010). Drought and salt tolerances in wild relatives for wheat and barley improvement. *Plant, Cell and Environment*, *33*(4), 670–685. https://doi.org/10.1111/j.1365-3040.2009.02107.x

Nevo E., Zohary, D., Haber, M., & Brown, A. H. D. (1979). Genetic Diversity and

Environmental Associations of Wild Barley, Hordeum spontaneum, in Israel. *Society*, *33*(3), 815–833. https://www.jstor.org/stable/2407648

Newton, A. C. (1993). The effect of humidity on the expression of partial resistance to powdery mildew in barley. *Plant Pathology*, *42*(3), 364–367. https://doi.org/10.1111/j.1365-3059.1993.tb01513.x

Niks, R. E., Habekuß, A., Bekele, B., & Ordon, F. (2004). A novel major gene on chromosome 6H for resistance of barley against the barley yellow dwarf virus. *Theoretical and Applied Genetics*, *109*(7), 1536–1543. https://doi.org/10.1007/s00122-004-1777-7

Olins, J. R., Lin, L., Lee, S. J., Trabucco, G. M., MacKinnon, K. J.-M., & Hazen, S. P. (2018). Secondary Wall Regulating NACs Differentially Bind at the Promoter at a CELLULOSE SYNTHASE A4 Cis-eQTL. *Frontiers in Plant Science*, *9*, 1895. https://doi.org/10.3389/fpls.2018.01895

Osman, A. M., Almekinders, C. J. M., Struik, P. C., & Lammerts van Bueren, E. T. (2016). Adapting spring wheat breeding to the needs of the organic sector. *NJAS - Wageningen Journal of Life Sciences*, *76*, 55–63. https://doi.org/10.1016/j.njas.2015.11.004

Oyiga, B. C., Palczak, J., Wojciechowski, T., Lynch, J. P., Naz, A. A., Léon, J., & Ballvora, A. (2020). Genetic components of root architecture and anatomy adjustments to water-deficit stress in spring barley. *Plant Cell and Environment*, *43*(3), 692–711. https://doi.org/10.1111/pce.13683

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, *421*(6918), 37–42. https://doi.org/10.1038/nature01286

Pätzold, S., Vetterlein, D., & Jahn, R. (2005). *"Crop Sequence and the Nutrient Acquisition from the Subsoil" Description of the Reference Soil Profile*.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0032253

Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., Chen, G., Sameri, M., Azhaguvel, P., Sakuma, S., Dhanagond, S., Sharma, R., Mascher, M., Himmelbach, A., Gottwald, S., Nair, S. K., Tagiri, A., Yukuhiro, F., Nagamura, Y., Kanamori, H., Matsumoto, T., … Komatsuda, T. (2015). Evolution of the Grain Dispersal System in Barley. *Cell*, *162*(3), 527–539. https://doi.org/10.1016/j.cell.2015.07.002

Pourkheirandish, M., & Komatsuda, T. (2007). The importance of barley genetics and domestication in a global perspective. In *Annals of Botany* (Vol. 100, Issue 5, pp. 999–1008). Oxford Academic. https://doi.org/10.1093/aob/mcm139

Pozníková, G., Fischer, M., Pohanková, E., & Trnka, M. (2014). Analyses of spring barley evapotranspiration rates based on gradient measurements and dual crop coefficient model. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *62*(5),

1079–1086. https://doi.org/10.11118/actaun201462051079

Pretty, J. N., Ball, A. S., Lang, T., & Morison, J. I. L. (2005). Farm costs and food miles: An assessment of the full cost of the UK weekly food basket. *Food Policy*, *30*(1), 1–19. https://doi.org/10.1016/j.foodpol.2005.02.001

Pulleman, M., Jongmans, A., Marinissen, J., & Bouma, J. (2003). Effects of organic versus conventional arable farming on soil structure and organic matter dynamics in a marine loam in the Netherlands. *Soil Use and Management*, *19*(2), 157–165. https://doi.org/10.1079/sum2003186

Qin, T., Zhao, H., Cui, P., Albesher, N., & Xionga, L. (2017). A nucleus-localized long non-coding rna enhances drought and salt stress tolerance. *Plant Physiology*, *175*(3), 1321–1336. https://doi.org/10.1104/pp.17.00574

Qu, J., Kachman, S. D., Garrick, D., Fernando, R. L., & Cheng, H. (2020). Exact distribution of linkage disequilibrium in the presence of mutation, selection, or minor allele frequency filtering. *Frontiers in Genetics*, *11*(April), 1–10. https://doi.org/10.3389/fgene.2020.00362

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Raza, K., & Ahmad, S. (2016). Principle, Analysis, Application and Challenges of Next-Generation Sequencing: A Review. *ArXiv Preprint*, *19*(October 2017), 1–29.

Reganold, J. P., & Wachter, J. M. (2016). Organic agriculture in the twenty-first century. In *Nature plants*. https://doi.org/10.1038/nplants.2015.221

Reid, T. A., Yang, R. C., Salmon, D. F., & Spaner, D. (2009). Should spring wheat breeding for organically managed systems be conducted on organically managed land? *Euphytica*, *169*(2), 239–252. https://doi.org/10.1007/s10681-009-9949-9

Reinert, S., Kortz, A., Léon, J., & Naz, A. A. (2016). Genome-wide association mapping in the global diversity set reveals new QTL controlling root system and related shoot variation in barley. *Frontiers in Plant Science*, *7*(2016JULY). https://doi.org/10.3389/fpls.2016.01061

Rellstab, C., Zoller, S., Tedder, A., Gugerli, F., & Fischer, M. C. (2013). Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PLoS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0080422

Ren, X., Wang, J., Liu, L., Sun, G., Li, C., Luo, H., & Sun, D. (2016). SNP-based high density genetic map and mapping of btwd1 dwarfing gene in barley. *Scientific Reports*, *6*(July), 1–7. https://doi.org/10.1038/srep31741

Riedel, C., Habekuß, A., Schliephake, E., Niks, R., Broer, I., & Ordon, F. (2011). Pyramiding of Ryd2 and Ryd3 conferring tolerance to a German isolate of Barley yellow dwarf virus-PAV (BYDV-PAV-ASL-1) leads to quantitative resistance against this isolate.

Theoretical and Applied Genetics, *123*(1), 69–76. https://doi.org/10.1007/s00122-011-1567-y

Ritchie, J. . T. . (1981). SOIL WATER AND NITROGEN in Mediterranean. *Plant and Soil*, *58*(1), 327–338. https://www.jstor.org/stable/42933793

Rode, N. O., Holtz, Y., Loridon, K., Santoni, S., Ronfort, J., & Gay, L. (2018). How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Molecular Ecology Resources*, *18*(2), 194–203. https://doi.org/10.1111/1755-0998.12723

Rohollahi, I., Khoshkholghsima, N. A., Nagano, H., Hoshino, Y., & Yamada, T. (2018). Respiratory burst oxidase-D expression and biochemical responses in Festuca arundinacea under drought stress. *Crop Science*, *58*(1), 435–442. https://doi.org/10.2135/cropsci2017.07.0416

Russell, J., Mascher, M., Dawson, I. K., Kyriakidis, S., Calixto, C., Freund, F., Bayer, M., Milne, I., Marshall-Griffiths, T., Heinen, S., Hofstad, A., Sharma, R., Himmelbach, A., Knauft, M., Van Zonneveld, M., Brown, J. W. S., Schmid, K., Kilian, B., Muehlbauer, G. J., … Waugh, R. (2016). Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nature Genetics*, *48*(9), 1024–1030. https://doi.org/10.1038/ng.3612

Saade, S., Maurer, A., Shahid, M., Oakey, H., Schmöckel, S. M., Negraõ, S., Pillen, K., & Tester, M. (2016). Yield-related salinity tolerance traits identified in a nested association mapping (NAM) population of wild barley. *Scientific Reports*, *6*(September), 1–9. https://doi.org/10.1038/srep32586

Saengwilai, P., Nord, E. A., Chimungu, J. G., Brown, K. M., & Lynch, J. P. (2014). Root cortical aerenchyma enhances nitrogen acquisition from low-nitrogen soils in maize. *Plant Physiology*, *166*(2), 726–735. https://doi.org/10.1104/pp.114.241711

Saliendra, N. Z., Sperry, J. S., & Cotock, J. P. (1995). Influence of leaf water status on stomatal response to humidity, hydraulic conductance, and soil drought in Betula occidentalis. *Planta: An International Journal of Plant Biology*, *196*(2), 357–366. https://doi.org/10.1007/BF00201396

Sánchez-Díaz, M., García, J. L., Antolín, M. C., & Araus, J. L. (2002). Effects of soil drought and atmospheric humidity on yield, gas exchange, and stable carbon isotope composition of barley. *Photosynthetica*, *40*(3), 415–421. https://doi.org/10.1023/A:1022683210334

Sandmann, S., De Graaf, A. O., Karimi, M., Van Der Reijden, B. A., Hellström-Lindberg, E., Jansen, J. H., & Dugas, M. (2017). Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Scientific Reports*, *7*, 1–12. https://doi.org/10.1038/srep43169

Sato, K., Yamane, M., Yamaji, N., Kanamori, H., Tagiri, A., Schwerdt, J. G., Fincher, G. B., Matsumoto, T., Takeda, K., & Komatsuda, T. (2016). Alanine aminotransferase controls seed dormancy in barley. *Nature Communications*, *7*(1), 1–9.

https://doi.org/10.1038/ncomms11625

Sayed, M. A., Schumann, H., Pillen, K., Naz, A. A., & Léon, J. (2012). AB-QTL analysis reveals new alleles associated to proline accumulation and leaf wilting under drought stress conditions in barley (Hordeum vulgare L.). *BMC Genetics*, *13*(1), 1–12. https://doi.org/10.1186/1471-2156-13-61

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*(11), 749–763. https://doi.org/10.1038/nrg3803

Schmalenbach, I., Körber, N., & Pillen, K. (2008). Selecting a set of wild barley introgression lines and verification of QTL effects for resistance to powdery mildew and leaf rust. *Theoretical and Applied Genetics*, *117*(7), 1093–1106. https://doi.org/10.1007/s00122-008-0847-7

Schmalenbach, I., Léon, J., & Pillen, K. (2009). Identification and verification of QTLs for agronomic traits using wild barley introgression lines. *Theoretical and Applied Genetics*, *118*(3), 483–497. https://doi.org/10.1007/s00122-008-0915-z

Schmalenbach, I., March, T. J., Pillen, K., Bringezu, T., & Waugh, R. (2011). High-resolution genotyping of wild barley introgression lines and fine-mapping of the threshability locus thresh-1 using the illumina goldengate assay. *G3: Genes, Genomes, Genetics*, *1*(3), 187–196. https://doi.org/10.1534/g3.111.000182

Schmalenbach, I., & Pillen, K. (2009). Detection and verification of malting quality QTLs using wild barley introgression lines. *Theoretical and Applied Genetics*, *118*(8), 1411–1427. https://doi.org/10.1007/s00122-009-0991-8

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. In *Nature Methods* (Vol. 9, Issue 7, pp. 671–675). Nature Publishing Group. https://doi.org/10.1038/nmeth.2089

Sham, P., Bader, J. S., Craig, I., O'Donovan, M., & Owen, M. (2002). DNA pooling: A tool for large-scale association studies. In *Nature Reviews Genetics* (Vol. 3, Issue 11, pp. 862–871). Nature Publishing Group. https://doi.org/10.1038/nrg930

Sharma, R., Draicchio, F., Bull, H., Herzig, P., Maurer, A., Pillen, K., Thomas, W. T. B., & Flavell, A. J. (2018). Genome-wide association of yield traits in a nested association mapping population of barley reveals new gene diversity for future breeding. *Journal of Experimental Botany*, *69*(16), 3811–3822. https://doi.org/10.1093/jxb/ery178

Shelton, A. C., & Tracy, W. F. (2016). Participatory plant breeding and organic agriculture: A synergistic model for organic variety development in the United States. *Elementa: Science of the Anthropocene*, *4*(0), 000143. https://doi.org/10.12952/journal.elementa.000143

Sheremet, Y. A., Yemets, A. I., Vissenberg, K., Verbelen, J. P., & Blume, Y. B. (2010). Effects of Inhibitors of Serine/Threonine Protein Kinases on Arabidopsis thaliana Root

Morphology and Microtubule Organization in Its Cells. *Cell and Tissue Biology*, *4*(4), 399–409. https://doi.org/10.1134/S1990519X10040139

Sierra, H., Cordova, M., Chen, C. S. J., & Rajadhyaksha, M. (2015). Confocal imaging-guided laser ablation of basal cell carcinomas: An ex vivo study. In *Journal of Investigative Dermatology* (Vol. 135, Issue 2, pp. 612–615). https://doi.org/10.1038/jid.2014.371

Spielmeyer, W., Ellis, M. H., & Chandler, P. M. (2002). Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(13), 9043–9048. https://doi.org/10.1073/pnas.132266399

Stephan, W. (2019). Selective Sweeps. *Genetics*, *211*(1), 5–13. https://doi.org/10.1534/genetics.118.301319

Stone, L. R., & Schlegel, A. J. (2006). Yield-water supply relationships of grain sorghum and winter wheat. *Agronomy Journal*, *98*(5), 1359–1366. https://doi.org/10.2134/agronj2006.0042

Suzuki, R., Terada, Y., & Shimodaira, H. (2019). *pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling*. https://cran.r-project.org/package=pvclust

Takahashi, R., Bashir, K., Ishimaru, Y., Nishizawa, N. K., & Nakanishi, H. (2012). *Takahashi ('12) - The role of heavy-metal ATPases, HMAs, in zinc*. *December*, 1605–1607.

Talamè, V., Ozturk, N. Z., Bohnert, H. J., & Tuberosa, R. (2007). Barley transcript profiles under dehydration shock and drought stress treatments: A comparative analysis. *Journal of Experimental Botany*, *58*(2), 229–240. https://doi.org/10.1093/jxb/erl163

Tan, Y., & Wang, L. (2020). MpDGK2, a Novel Diacylglycerol Kinase from Malus prunifolia, Confers Drought Stress Tolerance in Transgenic Arabidopsis. *Plant Molecular Biology Reporter*, *38*(3), 452–460. https://doi.org/10.1007/s11105-020-01209-y

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics*, *31*(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Tilk, S., Bergland, A., Goodman, A., Schmidt, P., Petrov, D., & Greenblum, S. (2019). Accurate allele frequencies from ultra-low coverage Pool-seq samples in evolve-and-resequence experiments. *G3: Genes, Genomes, Genetics*, *9*(12), 4159–4168. https://doi.org/10.1534/g3.119.400755

Tsuchihashi, Z., & Dracopoli, N. C. (2002). Progress in high throughput SNP genotyping methods. *Pharmacogenomics Journal*, *2*(2), 103–110. https://doi.org/10.1038/sj.tpj.6500094

Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T., & Nuzhdin, S. V. (2010). Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine

soils. *Nature Genetics*, *42*(3), 260–263. https://doi.org/10.1038/ng.515

Valueva, T. A., & Mosolov, V. V. (2004). Role of inhibitors of proteolytic enzymes in plant defense against phytopathogenic microorganisms. In *Biochemistry (Moscow)* (Vol. 69, Issue 11, pp. 1305–1309). https://doi.org/10.1007/s10541-005-0015-5

Veljovic-Jovanovic, S., Kukavica, B., Stevanovic, B., & Navari-Izzo, F. (2006). Senescence- and drought-related changes in peroxidase and superoxide dismutase isoforms in leaves of Ramonda serbica. *Journal of Experimental Botany*, *57*(8), 1759–1768. https://doi.org/10.1093/jxb/erl007

Verhoeven, K. J. F., Poorter, H., Nevo, E., & Biere, A. (2008). Habitat-specific natural selection at a flowering-time QTL is a main driver of local adaptation in two wild barley populations. *Molecular Ecology*, *17*(14), 3416–3424. https://doi.org/10.1111/j.1365-294X.2008.03847.x

Vicente-Serrano, S. M., Van der Schrier, G., Beguería, S., Azorin-Molina, C., & Lopez-Moreno, J. I. (2015). Contribution of precipitation and reference evapotranspiration to drought indices under different climates. *Journal of Hydrology*, *526*, 42–54. https://doi.org/10.1016/j.jhydrol.2014.11.025

Von Korff, M., Wang, H., Léon, J., & Pillen, K. (2005). AB-QTL analysis in spring barley. I. Detection of resistance genes against powdery mildew, leaf rust and scald introgressed from wild barley. *Theoretical and Applied Genetics*, *111*(3), 583–590. https://doi.org/10.1007/s00122-005-2049-x

von Korff, Maria, Léon, J., & Pillen, K. (2010). Detection of epistatic interactions between exotic alleles introgressed from wild barley (H. vulgare ssp. spontaneum). *Theoretical and Applied Genetics*, *121*(8), 1455–1464. https://doi.org/10.1007/s00122-010-1401-y

Vu, G. T. H., Wicker, T., Buchmann, J. P., Chandler, P. M., Matsumoto, T., Graner, A., & Stein, N. (2010). Fine mapping and syntenic integration of the semi-dwarfing gene sdw3 of barley. *Functional and Integrative Genomics*, *10*(4), 509–521. https://doi.org/10.1007/s10142-010-0173-4

Walters, D. R., Paul, N. D., & Ayres, P. G. (1984). Effects of mildew and nitrogen on grain yield of barley artifically infected in the field. *Annals of Botany*, *54*(1), 145–148. https://doi.org/10.1093/oxfordjournals.aob.a086768

Wang, G., Schmalenbach, I., von Korff, M., Léon, J., Kilian, B., Rode, J., & Pillen, K. (2010a). Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC2DH population and a set of wild barley introgression lines. *Theoretical and Applied Genetics*, *120*(8), 1559–1574. https://doi.org/10.1007/s00122-010-1276-y

Wang, G., Schmalenbach, I., von Korff, M., Léon, J., Kilian, B., Rode, J., & Pillen, K. (2010b). Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC2DH population and a set of wild barley introgression lines. *Theoretical and Applied Genetics*, *120*(8), 1559–1574.

https://doi.org/10.1007/s00122-010-1276-y

Wang, G., Schmalenbach, I., von Korff, M., Léon, J., Kilian, B., Rode, J., & Pillen, K. (2010c). Association of barley photoperiod and vernalization genes with QTLs for flowering time and agronomic traits in a BC2DH population and a set of wild barley introgression lines. *Theoretical and Applied Genetics*, *120*(8), 1559–1574. https://doi.org/10.1007/s00122-010-1276-y

Waters, E. R. (2003). Molecular adaptation and the origin of land plants. *Molecular Phylogenetics and Evolution*, *29*(3), 456–463. https://doi.org/10.1016/j.ympev.2003.07.018

Watt, C., Zhou, G., McFawn, L. A., Chalmers, K. J., & Li, C. (2019). Fine mapping of qGL5H, a major grain length locus in barley (Hordeum vulgare L.). *Theoretical and Applied Genetics*, *132*(4), 883–893. https://doi.org/10.1007/s00122-018-3243-y

Wei, K. H. C., Mantha, A., & Bachtrog, D. (2020). The theory and applications of measuring broad-range and chromosome-wide recombination rate from allele frequency decay around a selected locus. *Molecular Biology and Evolution*, *37*(12), 3654–3671. https://doi.org/10.1093/molbev/msaa171

Wheeler, T., & Von Braun, J. (2013). Climate change impacts on global food security. In *Science* (Vol. 341, Issue 6145, pp. 508–513). American Association for the Advancement of Science. https://doi.org/10.1126/science.1239402

Whelan, H. G., Gaunt, R. E., & Scott, W. R. (1997). The effect of leaf rust (Puccinia hordei) on yield response in barley (Hordeum vulgare L.) crops with different yield potentials. *Plant Pathology*, *46*(3), 397–406. https://doi.org/10.1046/j.1365-3059.1997.d01-23.x

Williams, K. J. (2003). The molecular genetics of disease resistance in barley. In *Australian Journal of Agricultural Research* (Vol. 54, Issues 11–12, pp. 1065–1079). CSIRO PUBLISHING. https://doi.org/10.1071/ar02219

Wisser, R. J., Fang, Z., Holland, J. B., Teixeira, J. E. C., Dougherty, J., Weldekidan, T., De Leon, N., Flint-Garcia, S., Lauter, N., Murray, S. C., Xu, W., & Hallauer, A. (2019). The genomic basis for short-term evolution of environmental adaptation in maize. *Genetics*, *213*(4), 1479–1494. https://doi.org/10.1534/genetics.119.302780

Wolfe, M. S., Baresel, J. P., Desclaux, D., Goldringer, I., Hoad, S., Kovacs, G., Löschenberger, F., Miedaner, T., Østergård, H., & Lammerts Van Bueren, E. T. (2008). Developments in breeding cereals for organic agriculture. In *Euphytica* (Vol. 163, Issue 3, pp. 323–346). https://doi.org/10.1007/s10681-008-9690-9

Wright, S. (1990). Evolution in mendelian populations. *Bulletin of Mathematical Biology*, *52*(1–2), 241–295. https://doi.org/10.1007/BF02459575

Xu, X., Sharma, R., Tondelli, A., Russell, J., Comadran, J., Schnaithmann, F., Pillen, K., Kilian, B., Cattivelli, L., Thomas, W. T. B., & Flavell, A. J. (2018). Genome-Wide Association Analysis of Grain Yield-Associated Traits in a Pan-European Barley

Cultivar Collection. *The Plant Genome*, *11*(1), 1–11. https://doi.org/10.3835/plantgenome2017.08.0073

Xu, Y., Jia, Q., Zhou, G., Zhang, X. Q., Angessa, T., Broughton, S., Yan, G., Zhang, W., & Li, C. (2017). Characterization of the sdw1 semi-dwarf gene in barley. *BMC Plant Biology*, *17*(1), 1–10. https://doi.org/10.1186/s12870-016-0964-4

Xue, G. P. (2003). The DNA-binding activity of an AP2 transcriptional activator HvCBF2 involved in regulation of low-temperature responsive genes in barley is modulated by temperature. *Plant Journal*, *33*(2), 373–383. https://doi.org/10.1046/j.1365-313X.2003.01630.x

Yamamoto, T., Suzuki, T., Suzuki, K., Adachi, S., Sun, J., Yano, M., Ookawa, T., & Hirasawa, T. (2017). Characterization of a genomic region that maintains chlorophyll and nitrogen contents during ripening in a high-yielding stay-green rice cultivar. *Field Crops Research*, *206*, 54–64. https://doi.org/10.1016/j.fcr.2017.03.001

Yan, L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., Bennetzen, J. L., Echenique, V., & Dubcovsky, J. (2004). The Wheat VRN2 Gene Is a Flowering Repressor Down-Regulated by Vernalization. *Science*, *303*(5664), 1640–1644. https://doi.org/10.1126/science.1094305

Yang, Z., Wu, Y., Li, Y., Ling, H. Q., & Chu, C. (2009). OsMT1a, a type 1 metallothionein, plays the pivotal role in zinc homeostasis and drought tolerance in rice. *Plant Molecular Biology*, *70*(1–2), 219–229. https://doi.org/10.1007/s11103-009-9466-1

Yao, Z., You, F. M., N'Diaye, A., Knox, R. E., McCartney, C., Hiebert, C. W., Pozniak, C., & Xu, W. (2020). Evaluation of variant calling tools for large plant genome re-sequencing. *BMC Bioinformatics*, *21*(1), 360. https://doi.org/10.1186/s12859-020-03704-1

Zahn, S., Koblenz, B., Christen, O., Pillen, K., & Maurer, A. (2020). Evaluation of wild barley introgression lines for agronomic traits related to nitrogen fertilization. *Euphytica*, *216*(3), 1–14. https://doi.org/10.1007/s10681-020-2571-6

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*(1), 161–172. https://doi.org/10.4161/epi.26931

Zhang, Baocai, Liu, X., Qian, Q., Liu, L., Dong, G., Xiong, G., Zeng, D., & Zhou, Y. (2011). Golgi nucleotide sugar transporter modulates cell wall biosynthesis and plant growth in rice. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(12), 5110–5115. https://doi.org/10.1073/pnas.1016144108

Zhang, Bing, Xiao, X., Zong, J., Chen, J., Li, J., Guo, H., & Liu, J. (2017). Comparative transcriptome analysis provides new insights into erect and prostrate growth in bermudagrass (Cynodon dactylon L.). *Plant Physiology and Biochemistry*, *121*(October), 31–37. https://doi.org/10.1016/j.plaphy.2017.10.016

Zhang, J., Wang, F., Liang, F., Zhang, Y., Ma, L., Wang, H., & Liu, D. (2018). Functional analysis of a pathogenesis-related thaumatin-like protein gene TaLr35PR5 from wheat induced by leaf rust fungus. *BMC Plant Biology*, *18*(1), 1–11. https://doi.org/10.1186/s12870-018-1297-2

Zhao, H., Ma, B., Duan, K. X., Li, X. K., Lu, X., Yin, C. C., Tao, J. J., Wei, W., Zhang, W. K., Xin, P. Y., Lam, S. M., Chu, J. F., Shui, G. H., Chen, S. Y., & Zhang, J. S. (2020). The GDSL lipase MHZ11 modulates ethylene signaling in rice roots. *Plant Cell*, *32*(5), 1626–1643. https://doi.org/10.1105/tpc.19.00840

Zhu, J., Brown, K. M., & Lynch, J. P. (2010). Root cortical aerenchyma improves the drought tolerance of maize (Zea mays L.). *Plant, Cell and Environment*, *33*(5), 740–749. https://doi.org/10.1111/j.1365-3040.2009.02099.x

Zhu, Y., Bergland, A. O., González, J., & Petrov, D. A. (2012). Empirical validation of pooled whole genome population re-sequencing in Drosophila melanogaster. *PLoS ONE*, *7*(7), 1–7. https://doi.org/10.1371/journal.pone.0041901

Zotarelli, L., & Dukes, M. (2010). Step by step calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). *Institute of Food and …*, 1–10. http://fawn.