Organic farming - Deep genotyping reveals specific selection footprints in barley populations

Michael Schneider
University of Bonn   https://orcid.org/0000-0002-6491-8852

Marissa Barbosa
University of Bonn

Agim Ballvora
University of Bonn

Jens Leon (✉ j.leon@uni-bonn.de)
University of Bonn   https://orcid.org/0000-0002-6514-3112

Abstract

Organic farming has the potential to tackle the imminent task of sustainable food production, if the yields could be raised. Here, the benefits of additional exotic alleles, the necessity of increased genetic heterogeneity in organically farmed systems and the buffering capacities by the pronounced plasticity of root traits are demonstrated. Two barley populations, naturally adapted for more than two decades to organic and conventional farming systems, were compared by a novel strategy of whole genome resequencing of pooled samples. Substantial allele frequency deviations between the farming systems were uncovered (for various agronomically relevant chromosomal regions) by testing multiple generations. In contrast to the organic adapted population, an early equilibrium in the conventional population was observed, accompanied reduced genetic diversity. Differences between the populations were revealed in root morphology, developmental processes and abiotic stress responses. These findings indicate that wild genetic resources play a critical role in the development of organically adapted varieties and distinct variations in ecosystems demand different genetic compositions.

Main

The cultivation of barley (*Hordeum vulgare*) and human development have been closely associated over the last 10,000 years[1]. Due to this tight bond, barley has accompanied human migration across almost all continents[2,3]. This movement was associated with early adaptation of genetic material to a wide range of local environments and formed a highly diverse gene pool across the globe[3,4]. As compared to this passive process, modern breeding strategies and the development of new farming technologies have shaped the allele frequency of modern high-yielding varieties at an extraordinary pace and extent[5]. These varieties have adapted to high-quantity application of agro-chemicals and the implementation of modern management strategies and harvesting processes[6]. As the demand of substantial yields continues to increase in the coming decades[7], the status of environmental factors will become more influential in modern societies[8]. The loss of biodiversity, pollution, and climate change are relevant to the future development of strategies for farmers and policy makers. Based on shifts in priorities, organic farming has become ever more relevant over the past few years[9]. To the benefit of ecosystem services, organic farming is reported to result in a yield reduction of 20%−30% as compared to conventional farming[10]. Therefore, it is imperative to counteract with drastic increases in yield until the year 2050[7,11]. As the genetic material used in organic farming is either derived from old varieties or varieties developed for conventional farming[12], the yield gap might be reduced by the development of organically adapted varieties[13]. However, conventionally adapted germplasm seems unsuitable for use as a single source of variation to achieve this goal[13,14]. Hence, the integration of wild-type (WT) genetic resources, which continuously undergo natural selection, could reverse the depletion of genetic diversity[15] and might enrich alleles suitable for organic farming. To test these hypotheses and identify relevant alleles for organic farming, a long-term selection experiment has been ongoing over the last two decades. As no artificial selection occurred, natural selection processes in conventionally and organically characterised

farming systems have altered the composition of WT introgression populations. Therefore, the aim of the present study was to investigate variant molecular genetic requirements and adaptation strategies on different generations of barley.

Results

Characterisation of genetic diversity A second backcross of the former elite cultivar Golf (recurrent) and the wild type ISR42-8 (donor) was performed to establish the starting population BC_2F_3. Their seeds were sawn in two agro-ecological environments. For each following year, the organic and conventional populations were harvested separately and seeds were used to establish the next generation. The two parents were genotyped by whole genome resequencing (WGrS). Five progeny generations (BC_2F_3, BC_2F_{12}, BC_2F_{16}, BC_2F_{22} and BC_2F_{23}) were separately pooled for WGrS, which resulted in 400 million high-quality paired-end reads for each library with an average length of 150 bp. The generated reads were mapped against the barley reference genome (version V2). Overall, 3,991,259 single nucleotide polymorphisms (SNPs) and 621,434 indels were identified. The genome-wide coverage was uniform with an average of nine reads per base (**Supplementary Figure 1**). As many as 3,632,910 SNPs were annotated to 34,343 high-confidence genes. The boundaries of the high-confidence genes were extended into the intergenic region for enrichment with closely located SNPs (**Extended Data Figure 1**), which led to a median enrichment of 55 SNPs per gene and a median haplotype size of 71,585 base pairs. The calculated allele frequency of all SNPs related to the same gene was aggregated to one haplotype frequency with a median read coverage of 434 reads per gene-haplotype. To predict the functional effects of SNP variants, the coding sequences with annotated start and stop codons were aligned to the parental haplotypes. A total of 28,601 complete coding DNA alignments were generated and 640,611 amino acid variants in 16,570 predicted proteins were identified between the parents. As compared to the reference coding DNA, 3,831 and 3,305 genes in the WT and the cultivar genomes, respectively, showed evidence of a deletion or insertion. A high degree of variation was observed at the DNA as well as the predicted amino acid level.

Variation among farming systems Genomic variation was estimated by genetic diversity analysis of the populations to assess the agro-environmental selection process. The principal component analysis (PCA) results revealed large genetic differences between the offspring generations due to differences in microevolution between the two agro-ecological environments. The first principal component (PC) described the highly significant ($p < 0.001$) differences between the organic and conventional environments, while the second PC separated the effects of different years (**Figure 1**). The PCA based variation between the two samples, pooled for each generation and environment separately, indicate a three to 13-fold higher genetic variation over the first five PCs for the 12th, 16th and 23rd generation in the organic environment. Only the 22nd generation revealed identical values between the environments (**Table 1**, **Supplementary Table 1**). Based on the first PC, the conventionally farmed population evolved until the 12th generation, but changed little afterward. In contrast, the generations of organic populations were clearly distinguished from one another. With each subsequent generation, the genetic distance between

the organic and conventional populations increased, as also shown by the increase in the population differentiation fixation index (F_{ST}) and genetic distance (**Extended Data Figure 2**). While the F_{ST} value of the 12th generation was 0.012, values successively increased to 0.043, 0.052 and 0.113 in the 16th, 22nd and 23rd generations, respectively. When the initial generation BC_2F_3 was included in the PCA, the first PC described the difference between the starting generation and offspring. The second PC split the organic from the conventional environment (**Extended Data Figure 3**).

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 BC_2F_{23} generation. 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 (**Supplementary Figure 2**, green line). To ensure these observations were not only or majorly the result of genetic drift, its effect was simulated in 1,000 iterations on the average gwAF as well as on all gene-haplotypes individually. Both were compared to the observed haplotype allele frequency in the 12th, 16th, 22nd and 23rd generations. Compared to the drift simulation, ten times larger 95% confidence intervals and a deviating average gwAF in the conventional and organic system were observed (**Supplementary Figure 3**). This indicates significant variations between the drift effect estimate and the conventional and organic environment for each generation (**Supplementary Table 2**).

Of the 34,343 analysed gene-based haplotypes, 7,115 WT haplotypes in the BC_2F_{23} generation of the conventional system were positively selected, as were 15,124 cultivar-derived haplotypes. Meanwhile, 3,066 WT alleles were not found in the BC_2F_{23} generation of the conventional system, of which 524 were already missing by the 12th generation onwards. In the organic system, only 547 of the gene-based haplotypes did not carry WT alleles present in any of the analysed generations (**Supplementary Table 3, Supplementary Figure 2**).

In order to assess the fitness of chromosomal regions, a genetic map was created based on the physical distance of two neighbouring haplotypes and associated allele frequency variation. A contracted genetic map indicated a higher selection pressure towards one haplotype and was associated with reduced genetic diversity in the population. The calculation resulted in a consistently more contracted map in the conventional system across the generations and chromosomes, as compared to the organic population. The results showed that genetic diversity, calculated as the chromosome map length, has 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. In contrast to the general findings, the organic population showed evidence of an increased map length for chromosomes 3H and 7H in the later generations (**Figure 3, Supplementary Figure 4**). The reliability of the constructed genetic map was estimated by simulating a cross population according to given genetic distances. A high correlation between the true and estimated genetic map ($r^2 > 0.93$) was obtained, indicating the genetic map build from pooled samples is accurate (**Supplementary Figure 5**).

To validate the observation of increased genetic variation in the organic environment, additional experiments were performed to assess the phenotypic difference of root traits in the BC_2F_{24} generation, which highlighted the greater phenotypic variation in the organic population (**Extended Data Figure 4**). In that regard, a significantly higher root length (p = 0.02) was observed in the organic population, supported by a 4-degree lower root angle (p = 0.005) (**Supplementary Table 4**)

Identification of hotspots undergoing selection Genome-wide variation in allele frequency was observed in both farming systems. While the wild type allele frequency (wtAF) was reduced in most pericentromeric regions in both environments, many loci in the telomeric regions indicated increasing wtAF (**Extended Data Fig. 5 &6**). A strong increase in wtAF was observed in both environments on the short arm of chromosome 1H, the long arms of 2H, 3H and 6H, as well as the centromere of 5H (**Figure 2**). The only centromere regions positive selected for WT alleles were on chromosomes 3H and 5H. These two regions were characterised by large linkage blocks (approx. 250 and 108 million bp for 3H and 5H, respectively).

The increased wtAF on chromosome 5H was associated with QGL5H / QGW5H (**Supplementary Table 5**), a quantitative trait loci (QTL) for grain length and width[16]. A search for candidate genes for this region resulted in several hits with a wtAF of ~0.5 for both environments in the late generations (**Supplementary Table 6**). Remarkably, this entire region of chromosome 5H (393–501 Mbp) had a very low wtAF in the BC_2F_3 generation (0.005 wtAF), which had increased to 0.33 in the conventional and 0.27 in the organic environments over successive generations. A similar pattern of lower than the average wtAF values in the initial population was observed at the centromere of chromosome 2H.

The centromere of chromosome 3H revealed a major differentiation between the organic and conventional populations. The wtAF reached 0.2 in the organic system, but remained below 0.05 in the conventional system. In multiple QTL studies, this region was characterised by enriched loci for various traits, including plant height[17], root morphology[18,19] and flowering time[20,21] (**Supplementary Table 5**). The genetic variation on chromosome 3H allowed for the study of selection and recombination events in the pericentromeric regions. The brittleness genes b*rt1* and b*rt2* are located in this region[22], and the WT had the physiological trait of ear brittleness, which is undesirable in modern cultivars and negatively selected by machine harvesting. Both brittleness genes were reduced in the wtAF to 0 and 0.005 in the conventional and organic systems, respectively. In addition, this region was characterised by four blocks undergoing positive selection in the organic system, but negative selection in the conventional system. During the increase in wtAF in these regions of up to 0.2 in the F_{23} generation, the smooth transition from down- to up-selected regions became more distinct with each subsequent generation tested (**Extended Data Figure 7**). In total, 82 QTL loci and 95 related candidate genes were retraced (**Supplementary Table 5, 6**).

The impacts of climatic variations on yield trend and genetic composition As indicated by the PCA results, the climate and weather influenced the selection process. The combination of annual weather (**Supplementary Figure 6**) and estimation of the actual evapotranspiration (ET_a, **Extended Data Figure 8**) of barley revealed the presence of different periods with varying selection pressure. The years 1998, 2001,

2002, 2005, 2007, 2008, 2009, 2012 and 2013 were characterised by no limitations of water availability. The years 2007, 2012, 2016 and 2018 have shown high levels of humidity in the anthesis stage, why these have been assumed to be beneficial years for above average fungal infection levels (**Supplementary Figure 7A & B**). Additionally, the ratio of days with precipitation to days without in the anthesis stage was found to be above average (0.3) for 2007 (0.41), 2012 (0.39) and 2016 (0.38). Based on evapotranspiration (ET) calculations, drought stress periods occurred in the years 1999, 2000, 2006, 2011, 2015, 2017, 2018 and 2019, highlighting the other end of the spectrum (**Extended Data Figure 8**). Multiple consecutive days in 2011, 2017, 2018 and 2019 were characterised by an actual ET_a to potential ET_C ratio of < 0.7 (**Extended Data Figure 9**).

The yield in both environments showed evidence of annual variation, while the yield of the conventional system was on average 30% higher over the period than the yield in the organic system. Furth details can be found in the **Supplementary Note and Supplementary Figure 8**.

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 described before, 2007 and 2016 were notably moist, giving chances for higher fungal pressure. Therefore, 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 signalling functions. The most manifested gene ontology classes were observed for proteolysis[23], ADP and ATP binding[24], serine/threonine protein kinase activity[25] and kinase activities[26] (**Extended Figure 10**).

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[27], transcription factor activity[28–30], nucleus[31], DNA binding[32] and sequence-specific DNA binding[33]), five additional ontologies were identified in the conventional environment (protein binding, oxidation reduction, oxidoreductase activity[34], zinc ion binding[35] and nucleic acid binding) (**Extended Figure 11**). Hence, the climatic conditions in exceptional years influenced the genetic composition of the observed populations.

Linking agro-environmental selection pressure with QTL for key trait categories To transfer the genome-wide variation to single traits, meta-data analysis of reported QTL data was performed of various agronomically important traits from the following six categories: nutrient uptake, biotic stress, drought tolerance, yield physiology, yield components and root morphology. Many of these trait associations had direct links to the WT ISR42-8 genotype. The wtAF of a QTL region was assessed (**Supplementary Table 5**) and compared between the two environments and among generations. All of the reported QTL were clustered into the above-mentioned six trait categories (**Figure 4**). The wtAF of the yield components, the biotic stress and yield physiology categories were increased in both environments above the level of the

BC_2F_3 generation, while significant differences in drought tolerance, yield physiology and root morphology between the two environments were observed. Even though the wtAF of drought tolerance decreased in the 12th generation, the process was reversed in the organic system in the following generations, analogous to the more frequent occurrence of water stress. In contrast, this cluster was further reduced in the conventional system. Root-related traits demonstrated the most dominant variation between the systems. While the wtAF followed a linear increase in the organic system across generations, there was no variation as compared to the first generation in the conventional system.

The wtAF of all clusters increased in the organic system across generations, while the conventionally farmed population benefited from the WT donor alleles, particularly in biotic resistance, yield physiology and yield components.

Discussion

We used whole-genome sequencing of pooled samples to trace allele frequency trajectories in a biparental cross derived from an elite cultivar and a wild barley. Based on the biparental origin, parental haplotypes were retraced within the populations. Unlike other species, such as *Arabidopsis thaliana*, the genome size and structure of barley results in pronounced challenges in terms of sequencing depth, technique selection and costs. Regular coverage pool sequencing of crops with exceptionally large genomes is possible and can produce reliable output down to the gene allele frequency level, as based on WGrS of pool sizes of 300 genotypes. Using a regular sequencing coverage can successfully be compensated by haplotype-based analysis, as reported previously[36]. Pool sequencing approaches have been successfully applied to small genomes[37,38], but relatively few studies have attempted to estimate genetic variation at the gene level[39,40]. Pool sequencing has proven to be an essential tool for the analysis of structural and genetic variations in population studies[40] and can be vital in both breeding strategies and population genetics.

Both organically and conventionally farmed populations can benefit from the genetic diversity of distinct WT alleles (**Supplementary Table 5, 6**). ISR42-8, which was used as the WT donor, was reported to have an expanded root system[41], increased drought tolerance[42–44], increased seed and ear length and kernel weight[45], reduced thresh-ability[46], reduced plant height[41], late flowering[47], enhanced biotic stressor resistance genes[48–50] and ear brittleness[46].

Few major alleles seemed to have dominated the selection process in the first generations. This resulted in the most remarkable changes of the allele frequency patterns in the first generations. The lowest variation between both environments was observed in the 12th generation. In the following generations, a cumulative distinction of the populations was observed. While the conventionally farmed population seemed to have reached a state of equilibrium in the later generations, the organic population was portrayed by diversity and genetic variation throughout all tested generations. This process was observed

across the entire genome (**Figure 2**). Furthermore, alleles derived from the WT were observed at higher rates in the organic population.

Across the genome, 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 (**Extended Data Figure 7**). These often small segments of sharp transitions may have resulted from breeding behaviour (backcrossing and self-pollinating) combined with selection. Backcrossing as well as selfing will reduce the length of heterozygous chromosome segments, which are the only source of recombination between nearby genes[51]. A following selection will address the remaining segment rather than the responding allele which result in the sharp transitions.

The genetic length of all chromosomes was estimated based on the reference sequence and the haplotype allele frequency. In contrast to a physical map, a genetic map depends on linkage as well on genotypic frequencies and can alter from generation to generation when the mating behaviour or allele frequencies change[52]. The mating behaviour did not alter between the organic and conventional population. Allele frequencies changes depend on genetic drift, mutation, migration or selection[53]. While genetic drift was identified as minor effect in the present study (**Supplementary Figure 3, Supplementary Table 2**), mutations were negligible and migration was avoided by the experimental design. Therefore, mostly selection can have influenced the allele frequencies. Therefore, linkage and selection can alter the frequency of gamete combinations. Based on these assumptions, the estimate of linkage disequilibrium (LD) is not only dependent on linkage, but also on selection in these populations, as the estimation of LD is based on the distribution of respective allele frequencies. The expected value of LD at equilibrium decreases in the presence of selection[54,55]. Consequently, an even moderate but continuous selection over consecutive generations will reduce the allele frequency variation and subsequently LD (**Figure 3**). Condensing LD chromosome-wise summarises the level of selection on allele frequency reduction and, consequently, reduced variation (**Supplementary Figure 4**). A more condensed map was generated for the conventional population. This is in line with the reduced system parameter variability, such as nutrients, growth regulators and pesticides, as compared to organic farming. Theoretically, the genetic map should extend further beyond the BC_2F_3 before it reaches a maximum[52]. Since both populations show condensed maps, we assume a strong selection in both environments, which, however, was more continuous and consequently stronger in the conventional agro-ecosystem.

Besides the farming system, weather variations had the second greatest impact on the changes in allele frequencies (**Figure 1**). The populations endured varying weather conditions, defined by water and heat stress as well as biotic attack. In both populations, differences in gene ontology resulting from variations in allele frequencies were found, with more extensive adaptations for the abiotic stress response being made in the conventional and extensive biotic stress adaptations in the organic one. Contrasting, resistance genes show no statistical difference between organic and conventional environment (p = 0.07) (Figure 4). Furthermore, the repeated fungal pressure resulted in a comparable yield reduction in the organic environment in the years 2007 and 2016 (**Supplementary Figure 8**), which highlights the

assumption of non-persistent adaptations to biotic stress. Although unexpected, this finding is in line with a report by Allard[56].

In terms of agronomically relevant traits, the most prevalent variation between the farming systems was related to the root system. Due to expected nutrient-deficient growing conditions, a distinct root system may have been a key trait in the organic environment. While other trait-associated allele frequencies have risen above the level of 0.5 for the WT allele, most root-related alleles were characterised by intermediate fitness (**Supplementary Table 5, 6**), suggesting high variability in the root morphology within the organic population. Phenotyping of root morphological traits of a representative sample size revealed the same results (**Extended Data Figure 4**, **Supplementary Table 4**), indicating advanced genetic and phenotypic diversity and pronounced habit of root characteristics in the organic population.

For the given population size and backcrossing level, the expected wtAF in the BC_2F_3 was 10.5%[57]. A minor deviation of 0.43% was observed. This underlines the power of the conduced crossing scheme as well as pooling and sequencing strategy. This deviation was most likely associated with the reduced allele frequency in the centromeres of chromosomes 2H and 5H (**Extended Data Figure 5**), but local deviations from the expected range could also be the result of copy number variations allocated to a single reference location. An incomplete representation of the ISR42-8 alleles on 5H was observed earlier[46]. The mean wtAF of these regions was <0.02 in the initial BC_2F_3 generation. Interestingly, the wtAF of this particular region on chromosome 5H exceeded >0.3 in both farming systems after less than 10 generations. A reported yield-related QTL[16] and the seed dormancy gene *SD1*[58] are located in this region, which led to the assumption that the linkage between these loci had a strong impact on the genetic composition. Once a linkage is broken, evolutionary selection processes are very fast and even rare haplotypes can be up-selected to the major haplotypes within a few generations. Similar patterns were observed for the brittleness genes brt*1* and brt*2*[22], which are responsible for an important domestication trait[3] (**Extended Data Figure 7**). The WT donor carried an unfavourable allele, which could explain the negative selection in both environments. This observation may have been due to an artificial influence, as predominant genotypes with stable ears were harvested by the combiner.

The observation of yield and allele frequency in the conventional and organic environments indicate the potential ability of such populations to buffer variant environmental stressors. An adaptation of the allele frequency towards increased (a)biotic resistance was observed. Analogous to the extended root system observed in the organic population, the drought stress response on candidate QTL loci level was over represented by WT alleles. In contrast, the molecular drought stress response of the organic system was less pronounced by WT alleles compared to the conventional population. These observations might indicate that the more diverse root architecture in the organic population enables enhanced adaptation to water limitation events. Nevertheless, climatic conditions vary widely from year to year, especially in the last couple of years, resulting in potentially unfavourable adaptation, as observed in the organic system for the year 2016 and afterward (**Supplementary Figure 8**).

In conclusion, pool sequencing is a powerful method for the characterisation of populations when the breeding behaviour, 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 localise candidate genes, without analysing phenotyping data. With regard to the selection forces under organic or conventional farming, the organic population showed a much higher genetic and phenotypic variability. 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 the availability of resources, thus increased genetic heterogeneity is recommended. Genetic resources that have not undergone selection in farming systems with high resources should be used to establish adapted cultivars.

Declarations

Acknowledgements

Author´s contribution

JL and AB conceptualized the research. MS analysed the data. MS, AB and JL have written the manuscript. MB performed the root phenotyping.

Competing interests

The authors declare no competing interests.

Data availability

Raw variant calling data and processed data can be found under doi:10.5281/zenodo.4304046

Code availability

Code used to perform the analysis can be found at https://github.com/mischn-dev/HAFcall

References

1. Zohary, D., Hopf, M. & Weiss, E. *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. (Oxford University Press, 2012).

2. Malysheva-Otto, L. V, Ganal, M. W. & Röder, M. S. Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (Hordeum vulgare L.). (2006) doi:10.1186/1471-2156-7-6.

3. von Bothmer, R., van Hintum, T., H., K. & Sato, K. *Diversity in Barley (Hordeum vulgare), Volume 7 – 1st Edition*. (2003).

4. Morrell, P. L. & Clegg, M. T. Genetic evidence for a second domestication of barley (Hordeum vulgare) east of the Fertile Crescent. *Proc. Natl. Acad. Sci. U. S. A.***104**, 3289–3294 (2007).

5. Khush, G. S. Green revolution: The way forward. *Nature Reviews Genetics* vol. 2 815–822 (2001).

6. Hedden, P. The genes of the Green Revolution. *Trends Genet.***19**, 5–9 (2003).

7. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nat. Commun.***3**, 1–7 (2012).

8. Tilman, D. The greening of the green revolution. *Nature* vol. 396 211–212 (1998).

9. Willer, H., Yussefi-Menzler, M. & Sorensen, N. *The world of organic agriculture: Statistics and emerging trends 2008. The World of Organic Agriculture: Statistics and Emerging Trends 2008* (2008). doi:10.4324/9781849775991.

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

11. Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS One***8**, 66428 (2013).

12. Shelton, A. C. & Tracy, W. F. Participatory plant breeding and organic agriculture: A synergistic model for organic variety development in the United States. *Elem. Sci. Anthr.***4**, 000143 (2016).

13. Murphy, K. M., Campbell, K. G., Lyon, S. R. & Jones, S. S. Evidence of varietal adaptation to organic farming systems. *F. Crop. Res.***102**, 172–177 (2007).

14. Le Campion, A., Oury, F. X., Heumez, E. & Rolland, B. Conventional versus organic farming systems: dissecting comparisons to improve cereal organic breeding strategies. *Organic Agriculture* vol. 10 63–74 (2020).

15. Nevo, E. Genetic diversity in wild cereals: Regional and local studies and their bearing on conservation ex situ and in situ. *Genet. Resour. Crop Evol.* **45**, 355–370 (1998).

16. Watt, C., Zhou, G., McFawn, L. A., Chalmers, K. J. & Li, C. Fine mapping of qGL5H, a major grain length locus in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* **132**, 883–893 (2019).

17. Marzec, M. & Alqudah, A. M. Key hormonal components regulate agronomically important traits in barley. *Int. J. Mol. Sci.* **19**, 1–12 (2018).

18. Reinert, S., Kortz, A., Léon, J. & Naz, A. A. Genome-wide association mapping in the global diversity set reveals new QTL controlling root system and related shoot variation in barley. *Front. Plant Sci.* **7**, (2016).

19. Oyiga, B. C. *et al.* Genetic components of root architecture and anatomy adjustments to water-deficit stress in spring barley. *Plant Cell Environ.* **43**, 692–711 (2020).

20. Guan, J. C. *et al.* Diverse roles of strigolactone signaling in maize architecture and the uncoupling of a branching-specific subnetwork. *Plant Physiol.* **160**, 1303–1317 (2012).

21. Xu, Y. *et al.* Characterization of the sdw1 semi-dwarf gene in barley. *BMC Plant Biol.* **17**, 1–10 (2017).

22. Komatsuda, T., Maxim, P., Senthil, N. & Mano, Y. High-density AFLP map of nonbrittle rachis 1 (btr1) and 2 (btr2) genes in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* **109**, 986–995 (2004).

23. Valueva, T. A. & Mosolov, V. V. Role of inhibitors of proteolytic enzymes in plant defense against phytopathogenic microorganisms. *Biochemistry (Moscow)* vol. 69 1305–1309 (2004).

24. DeYoung, B. J. & Innes, R. W. Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature Immunology* vol. 7 1243–1249 (2006).

25. Lin, Z. J. D., Liebrand, T. W. H., Yadeta, K. A. & Coaker, G. PBL13 is a serine/threonine protein kinase that negatively regulates arabidopsis immune responses. *Plant Physiol.* **169**, 2950–2962 (2015).

26. Goff, K. E. & Ramonell, K. M. The role and regulation of receptor-like kinases in plant defense. *Gene Regul. Syst. Bio.* **1**, 167–75 (2007).

27. Talamè, V., Ozturk, N. Z., Bohnert, H. J. & Tuberosa, R. Barley transcript profiles under dehydration shock and drought stress treatments: A comparative analysis. in *Journal of Experimental Botany* vol. 58 229–240 (Oxford Academic, 2007).

28. Marè, C. *et al.* Hv-WRKY38: A new transcription factor involved in cold- and drought-response in barley. *Plant Mol. Biol.* **55**, 399–416 (2004).

29. Al Abdallat, A. M., Ayad, J. Y., Abu Elenein, J. M., Al Ajlouni, Z. & Harwood, W. A. Overexpression of the transcription factor HvSNAC1 improves drought tolerance in barley (Hordeum vulgare L.). *Mol. Breed.* **33**, 401–414 (2014).

30. Alexander, R. D., Wendelboe-Nelson, C. & Morris, P. C. The barley transcription factor HvMYB1 is a positive regulator of drought tolerance. *Plant Physiol. Biochem.***142**, 246–253 (2019).

31. Qin, T., Zhao, H., Cui, P., Albesher, N. & Xionga, L. A nucleus-localized long non-coding rna enhances drought and salt stress tolerance. *Plant Physiol.***175**, 1321–1336 (2017).

32. Xue, G. P. 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 J.***33**, 373–383 (2003).

33. Abe, H. *et al.* Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell***9**, 1859–1868 (1997).

34. Moran, J. F. *et al.* Drought induces oxidative stress in pea plants. *Planta***194**, 346–352 (1994).

35. Yang, Z., Wu, Y., Li, Y., Ling, H. Q. & Chu, C. OsMT1a, a type 1 metallothionein, plays the pivotal role in zinc homeostasis and drought tolerance in rice. *Plant Mol. Biol.***70**, 219–229 (2009).

36. Tilk, S. *et al.* Accurate allele frequencies from ultra-low coverage Pool-seq samples in evolve-and-resequence experiments. *G3 Genes, Genomes, Genet.***9**, 4159–4168 (2019).

37. Schlötterer, C., Tobler, R., Kofler, R. & Nolte, V. Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.***15**, 749–763 (2014).

38. Gautier, M. *et al.* Estimation of population allele frequencies from next-generation sequencing data: Pool-versus individual-based genotyping. *Mol. Ecol.***22**, 3766–3779 (2013).

39. Burke, M. K. *et al.* Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature***467**, 587–590 (2010).

40. Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T. & Nuzhdin, S. V. Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nat. Genet.***42**, 260–263 (2010).

41. Naz, A. A., Arifuzzaman, M., Muzammil, S., Pillen, K. & Léon, J. Wild barley introgression lines revealed novel QTL alleles for root and related shoot traits in the cultivated barley (Hordeum vulgare L.). *BMC Genet.***15**, 1–12 (2014).

42. Honsdorf, N., March, T. J., Berger, B., Tester, M. & Pillen, K. High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. *PLoS One***9**, (2014).

43. Honsdorf, N., March, T. J. & Pillen, K. QTL controlling grain filling under terminal drought stress in a set of wild barley introgression lines. *PLoS One***12**, (2017).

44. Sayed, M. A., Schumann, H., Pillen, K., Naz, A. A. & Léon, J. AB-QTL analysis reveals new alleles associated to proline accumulation and leaf wilting under drought stress conditions in barley (Hordeum vulgare L.). *BMC Genet.***13**, 1–12 (2012).

45. Zahn, S., Koblenz, B., Christen, O., Pillen, K. & Maurer, A. Evaluation of wild barley introgression lines for agronomic traits related to nitrogen fertilization. *Euphytica* **216**, 1–14 (2020).

46. Schmalenbach, I., March, T. J., Pillen, K., Bringezu, T. & Waugh, R. 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, Genet.* **1**, 187–196 (2011).

47. Wang, G. *et al.* 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. *Theor. Appl. Genet.* **120**, 1559–1574 (2010).

48. Dragan, P. *et al.* 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* 269–284 (Springer Netherlands, 2013). doi:10.1007/978-94-007-4682-4_23.

49. Von Korff, M., Wang, H., Léon, J. & Pillen, K. AB-QTL analysis in spring barley. I. Detection of resistance genes against powdery mildew, leaf rust and scald introgressed from wild barley. *Theor. Appl. Genet.* **111**, 583–590 (2005).

50. Schmalenbach, I., Körber, N. & Pillen, K. Selecting a set of wild barley introgression lines and verification of QTL effects for resistance to powdery mildew and leaf rust. *Theor. Appl. Genet.* **117**, 1093–1106 (2008).

51. Hanson, W. D. Early Generation Analysis of Lengths of Heterozygous Chromosome Segments around a Locus Held Heterozygous with Backcrossing or Selfing. *Genetics* **44**, 833–837 (1959).

52. Teuscher, F., Guiard, V., Rudolph, P. E. & Brockmann, G. A. The map expansion obtained with recombinant inbred strains and intermated recombinant inbred populations for finite generation designs. *Genetics* **170**, 875–879 (2005).

53. Wright, S. Evolution in mendelian populations. *Bull. Math. Biol.* **52**, 241–295 (1990).

54. Qu, J., Kachman, S. D., Garrick, D., Fernando, R. L. & Cheng, H. Exact distribution of linkage disequilibrium in the presence of mutation, selection, or minor allele frequency filtering. *Front. Genet.* **11**, 1–10 (2020).

55. Bejarano, D. *et al.* Linkage disequilibrium levels and allele frequency distribution in blanco orejinegro and romosinuano creole cattle using medium density snp chip data. *Genet. Mol. Biol.* **41**, 426–433 (2018).

56. Allard, R. W. Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. *J. Hered.* **79**, 225–238 (1988).

57. Cox, T. S. Expectations of means and genetic variances in backcross populations. *Theor. Appl. Genet.* **68**, 35–41 (1984).

58. Pourkheirandish, M. & Komatsuda, T. The importance of barley genetics and domestication in a global perspective. *Annals of Botany* vol. 100 999–1008 (2007).

59. Comadran, J. *et al.* Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.***44**, 1388–1391 (2012).

Table

Table 1: **Sum of principal component distance between the two samples of the same generation and environment.** The distance between these samples was calculated for each PC separately, transformed to an absolute value and multiplied with the explanatory effect that has been observed for the corresponding PC. All values were summed up for the first five PCs and illustrated in the table.

	Generation			
Environment	*F12*	*F16*	*F22*	*F23*
organic	140.23	123.67	31.6	265.42
conventional	45.1	17.06	31.41	21.73

Material And Methods

Plant material and population construction

More than 25 years ago, a long-term selection experiment was launched using barley as a model. The crossing scheme used in this study is described elsewhere[57]. The population was derived from an initial cross of the cultivar Golf (*H. vulgare*) and WT ISR42-8 (*Hordeum vulgare L. ssp. spontaneum)* as the recurrent and donor parents, respectively. Golf was selected as an elite cultivar and used as the maternal plant in the initial barley cross, while the WT was used as the donor to increase the genetic diversity of the population. The F_1 generation was backcrossed to the cultivar and six BC_1F_1 plants were randomly selected and further backcrossed. Two plants of each BC_2F_1, in a total of 12 sublines, were used to produce BC_2F_2 plants. All crossings were performed in a greenhouse under controlled conditions to avoid admixtures. An equal number of obtained seeds (BC_2F_3) from each (sub)cross in BC_2F_2 were then sown to ensure that all the BC_2F_2 introgression lines had equal contributions for the following selection process.

Agro-environmental selection experiments

Cultivation occurred under organic and conventional cropping conditions for 22 generations (from 1997 to 2019) at two field stations managed by the University of Bonn: Dikopshof (1997 to 2010; 50°48'22.6"N 6°57'05.5"E) and Klein−Altendorf (2009 to 2019; 50°36'47.2"N 6°59'39.3"E). The organically farmed system employed a seven-field crop rotation without application of pesticides or growth regulators. In contrast, the conventionally farmed system was based on a three-field crop rotation, complemented by on-demand plant and fungicide control in accordance with professional practice standards. 80−100 kg/ha of nitrogen (calcium ammonium nitrate) fertiliser were applied at the booting and flowering stages.

Cattle manure (200 dt/ha) was applied to two crops in the seven-field rotation of the organic system. Further details of the crop management strategy are presented in **Supplementary Table 7**. The pH value was adjusted to 7 and the organic matter content of both systems ranged from 1.5% to 2%. The sowing density was 320 and 360 grains/m^2 for the conventional and organic systems, respectively. Based on a plot size of 9 × 15 m, the annual population size was approximately 40,000 individual plants. To ensure that selection pressure is caused only by diseases, availability/lack of nutrients and minerals, crop rotation and weather conditions, the populations rotated in defined plots. The plants were harvested using a plot combine harvester. Seeds were harvested from the centre of the plots for the following generation.

Weather data and yield

Climate data of the experimental fields were collected from local weather stations. Radiation, wind speed, humidity, precipitation and temperature were measured every 10 min at 1 m above ground level. As there were deviations in the measured data of the two sites (Dikopshof and Klein−Altendorf), annual average values were calculated for the two data sets separately. The weather data was aggregated to daily maximum, minimum, and mean values. The potential evapotranspiration during the barley vegetation period was estimated by calculating the reference evapotranspiration value using the Penman−Monteith equation[60]. In addition, the barley crop evapotranspiration (ET_C) value[61,62] was calculated and the soil−water balance was determined to estimate periods with water deficits (ET_A)[63,64]. The limit of thermal time for the initial, crop development, mid and end stages were set at 220, 650, 1300 and 1650 growing degree days, respectively, using 0°C as the base temperature. The water balance was based on a maximal root depth of 75 cm. When the usable field capacity decreased to less than 30%, the evapotranspiration value was corrected as described previously[63]. The correct estimation of ET_C by the FAO *Penman-Monteith* equation was illustrated in [65]. Precipitation events, minimal humidity and the deviation of ET_A to ET_C were used to predict and categorise periods with high fungal pressure and excessive physiological drought conditions by grouping years based on the Tukey test[66].

The texture of the soil, which was composed predominantly of loam, was characterised by a high-field capacity of approximately 25%. The soil type is a Luvisol.

Annual yield data was collected and reported as metric decitons per hectare (dt/ha) with a corrected water content of 14%. A yield ratio value was calculated based on the annual yield, compared to the mean value of the first three generations, national and state-wide yields of spring barley.

Preparation of samples for genotyping

WGS was performed for five offspring generations (BC_2F_3 [initial offspring generation grown in a greenhouse], BC_2F_{12}, BC_2F_{16}, BC_2F_{22} and BC_2F_{23}) from both the conventional and organic environments.

Additionally, the genomes of the two parental cultivars were sequenced. The required sample size was estimated by applying Cochran's Formula[67] with an aimed precision level of $p < 0.05$ and a narrow confidence interval of $p > 0.996$ in order to determine the correct allele frequency with a maximum deviation of 5% from the true value. Based on a theoretically expected allele frequency of an infinite population size - 12.5%[57] of the donor WT alleles, a sample size of 296 genotypes per generation and environment was required for an accurate estimate of the total genetic diversity of the population. Based on these estimates, 300 genotypes were pooled. Pool samples at the leaf tissue level were constructed by harvesting equal amounts of leave discs per genotype utilising a hole puncher, which were then pooled. To avoid bias, tissue materials were vigorously homogenised using a TissueLyser II ball mill (Qiagen GmbH, Hilden, Germany). Total genomic DNA was extracted using the Plant DNA Mini kit (VWR International GmbH, Darmstadt, Germany) in accordance with manufacturer's instructions. Two samples of each generation and environment with 300 genotypes were pooled, creating a test set of 600 genotypes.

Genotyping of parents and populations using pooled samples

Three sequencing approaches, namely WGrS (Novogene Co., Ltd., Beijing, China)[68], genotyping by sequencing (GBS; LGC Genomics GmbH, Berlin, Germany)[69] and massive analysis of coding DNA ends (MACE; GenXPro GmbH, Frankfurt am Main, Germany)[70], were tested and compared. Sequencing with a moderate depth of 10 (WGrS) to 30 (GBS) reads per locus was applied. Haplotypes were constructed based on data of the cultivar and WT genotypes. The SNPs located within or close to an annotated gene were aggregated into one single haplotype. The annotated gene position was expanded up- and downstream of the 3´and 5' ends, respectively, to include adjacent SNPs into the "gene-based haplotype". The developed algorithm extended the haplotype size by 45% of the flanking intergenic region on each side of the genes, leaving a minimum gap of 1000-bp non-coding DNA sequence between two neighbouring genes. Furthermore, the same approach was utilised to construct haplotypes based on markers of the Barley 9K iSelect SNP Chip[59] (Illumina, Inc., San Diego, CA, USA). 3,539,767 SNPs were linked to 5,948 markers included on the Barley 9K iSelect SNP Chip. Details on the applied bioinformatics pipelines for reads alignment, variant calling and allele frequency estimations, and haplotype construction are described in the **Supplementary Notes** and the following sections.

Evaluation of the pool genotyping approaches

The genotypes of the pooled samples were compared to the genotype data obtained from the corresponding individual pooled samples. Competitive allele specific PCR (KASP) genotyping assays were designed for three SNPs of each of the seven barley chromosomes. Meanwhile, seeds from 288 individuals of one pool were germinated and leaf tissues were collected from each. Subsequently, total

genomic DNA was extracted using the Plant DNA Mini kit (Qiagen GmbH). The individual samples were genotyped using KASP technology by SGS S.A. (Geneva, Switzerland).

The SNPs and haplotype allele frequencies of the pooled sequences were associated with the allele frequencies determined with the KASP assay. All three sequencing methods were applied to sequence this exact same set of genotypes via a pooling approach (GBS, MACE and WGrS). The KASP markers and haplotypes were correlated if the marker was located within the haplotype boundaries. The data obtained by the WGrS approach for gene-based haplotyping had the highest Pearson correlation coefficient of the pooled allele frequency determined with the KASP assay of 0.97. MACE RNA sequencing revealed a comparably high correlation of 0.9, while GBS-based pooled genotyping failed to produce a reliable correlation (-0.035) (**Supplementary Figure 9**). Based on a pool size of 288 genotypes per sample, the most accurate allele frequency estimate was obtained by WGrS. On account of these findings, the WGrS procedure was chosen for further pooled genotyping. Above a level of 100 reads coverage per haplotype, no impact of the read coverage and the accuracy were identified.

Genetic diversity determination

More than 20 million variants were identified between the cultivar and WT donor. Variant detection of next-generation sequencing data is exceptionally error prone[71,72]; therefore, variant loci not reported in the Ensembl online database were removed[73]. Following this approach, about 4 million SNPs were retained for calling of pooled SNPs. Although a potential source of variation, insertions and deletions (indels) were omitted from haplotyping.

The allele frequency for each variant was calculated as $AF_{Parent1} = RD_{Allele\ Parent1} / (RD_{Allele\ Parent1} + RD_{Allele\ Parent2})$ and $AF_{Parent2} = 1 - AF_{Parent1}$, where RD is the read depth. The allele frequency and the read depth were further used to calculate the haplotype allele frequency (HAF), as illustrated below. As the allele frequency was corrected by the read coverage of the SNP, a minimum cut off of coverage per SNP was not calculated and all SNPs with a coverage of >1 were included for analysis.

Based on the fact that the population has two founders (Golf and ISR42-8) and their haplotypes are known, this information can be incorporated in the pool sequencing of the population. On short to medium distances (< 200,000 bp), we assume that recombination should have rarely occurred. Gene annotation derived from IPK database[74] with positional annotation for low and high confidence genes was used as anchor to generate haplotypes with a direct link to a function. 80,554 low and high confident genes are listed and can be utilized as potential origin of a haplotype. From this set of genes, only the high confident genes have been used for haplotype construction. The gene bounds were extended on both 3' and 5' ends to create haplotypes and include close-by mapped SNPs for haplotype allele frequency (HAF) calculation. 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 any neighbouring gene. Exemplarily, gene Q with position 200 to 500 and gene W with

position 800 to 850 both are expanded. The intergenic region has a size of 300 base pairs. 10 percent of this, 30 bp, are removed, leaving 270 bp. This is divided by 2, giving 135 bp. The new end position of haplotype Q is 500 + 135 = 635, and the new start position of haplotype W is 800 − 135 = 665. The gap between these two haplotypes has been closed from 300 to 30bp. If a SNP is at position 590, it is now considered as part of haplotype Q. Without the extension, the information of this particular SNP would have been lost for the improvement of the haplotype frequency. If genes overlap, no extension is performed in the direction of overlap for both genes. Besides the genes, markers present in the 9KiSelect SNP chip[59] were mapped against the reference genome and a range around them was defined as a haplotype region. This was done according to the same procedure as highlighted for the genes and shown in the **extended figure 1**. 6,080 markers could be mapped to the reference genome. All reads within a haplotype were used to calculate the haplotype allele frequency,

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

Equation 1: Haplotype allele frequency calculation

where p represents the allele for a specific parent. The k is the SNP, for which the coverage rd and $freq$ allele frequency is given. Marker information comes along with the genetic position, which were utilized to plot the wtAF on a genetic map (**Figure 2**).

Genetic map construction

As no standard haplotype allele frequency (HAF) can be obtained from the pooled genotyping data, the recombination (D') and linkage (r^2) values were calculated based on the HAF variation (*delta)* and the physical distance (*Dist*) between two neighbouring haplotypes. Haplotypes with a read coverage below 100 were excluded, as these were more likely to present a biased or false allele frequency.

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

Equation 2: linkage estimation from pool sequencing data

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

Equation 3: recombination estimation from pool sequencing data

The recombination D' was used to calculate a genetic map. The haplotype's physical position was used to arrange the haplotypes and D' was calculated for each pair of neighbouring haplotypes. To generate a genetic map, pairwise calculated D' was continually summarized for all haplotypes. For each

chromosome, its length was calculated from the sum of D' over the chromosome. This calculation has been repeated separately for each generation and each agro-ecosystem, resulting in a number of map length estimates. The rationale behind this idea is that the gamete phase of two neighbouring haplotypes should be identical, unless a recombination has occurred. The majority of the retraceable crossovers has accumulated in the initial crossing and two backcrossing steps[52] and should not be significantly different between the two populations, if no further forces affect the allele frequencies. Therefore, it is assumed that the observed differences in crossovers' probabilities between environments are related to a fitness effect of one of regarded haplotypes.

To validate these assumptions, a simulation was performed using AlphaSimR[75] package to create a simple F4 population. The genetic positions of the 9KiSelect chip were used, after all markers with a unique physical position and a distance of at least 10kb between the markers were selected. 3,999 markers in total were used. Homozygote, polymorph, diploid and inbred parental genotypes for all these loci were created. Those were crossed by using the *makeCross2* function with 4,000 progenies output. Those 4,000 progenies were selfed 3 times, creating a single progeny per input genotype for each round. The allele frequency for each marker was calculated over all 4,000 F4 genotypes. The genetic map was estimated based on D' (formula presented above), using the physical position of the marker and the allele frequency variation, multiplied with 12,500,000. This value was subjectively chosen to align the genetic maps and has no impact on the shape of the curve. All steps were iterated 20 times and the mean genetic position was calculated for each marker over all iterations. The correlation of the approximated genetic map to the 9KiSelect chip derived map was calculated by a Pearson correlation (**Supplementary figure 4**).

Estimation of the donor alleles rate

Introgression level assessment at the SNP level was prone to errors, as the read depth was rather low. Therefore, gene haplotypes of each generation were tested to determine the allele frequency of the WT. Of 34,344 HC genes, 531 (1.55%) had a trans-generational WT allele frequency of <0.005 and, thus, were not expected to be introgressed into the population. Most of these genes were located on chromosome 2H and the short arms of chromosomes 3H and 4H.

Overall genetic structure analysis

The potential influence of genetic drift was simulated by a Julia script, related to the R function *sim.drift*[76]. The following settings were applied: population size, 10,000 (equals the center part of the field plot); ploidy level, 2; number of generations, 21; as initial allele frequency, 0.1; number of loci, 34,237; and iterations, 1,000. Additionally, the wtAF of the BC_2F_3 was used as start point to calculate the variation based on drift effects for the following generations. 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[77] to identify genetic structures and variations between generations and environments. The explanatory effect of each principal component on the generations, replicates and environments was calculated by the *stats*[78] *cor.test* function, based on a Pearson correlation test. Samples were categorised by 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*[79] 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 *vcftools*[80] with default settings and *vcf* and BAM files as input.

Determination of root phenotype variation

Hydroponic and field experiments on root morphological characteristics were performed to complete the observations made at the genetic level. In a hydroponic experiment, 150 BC_2F_{24} genotypes of the conventional and organic environment were sown, as described previously[81]. 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 6 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.

The field evaluation was conducted at Campus Klein–Altendorf Research Facility under rain-fed conditions without additional application of agro-chemicals. A total of 100 barley BC_2F_{24} lines from the organic and conventional populations were grown until anthesis. The roots of three plants per line were carefully harvested and the soil was washed off, in accordance with the guidelines described by Neumann et al.[82].

Root trait measurements were performed using WinRHIZO image analysis software (ver. 2008a; Regent Instruments Inc., Québec, QC, Canada). The root mass density was calculated as described by Nakhforoosh et al.[83]. The root anatomical parameters were measured on cross-section slides. A digital microscope (Keyence's VHX-1000D with 100X magnification) was used to capture images of the section area. At least three to five roots per genotype were considered for measurements of root anatomical traits. Captured images were then analysed with ImageJ software[84].

In the hydroponic experiment, significant variations between the organic and conventional environments were observed in 14 (51.9%) of 27 traits. Similarly, significant variations were observed in 10 (50%) of the

20 traits measured in the field experiment between the populations ($p < 0.05$) (**Supplementary Table 4**).

Predictions of coding and putative amino acid sequences

The splice variants of the coding DNA reference from the EnsemblPlants genome-centric portal for plant species[73] were filtered for the presence of start and stop codons and used to map the obtained reads. If multiple splice variants had a start and stop codon, the longest was used as a reference. Information of the physical location of the gene was utilised to extract a consensus sequence for the cultivar and WT donor from BAM files using *samtools* and *bcftools*[85]. If necessary, the reverse complement strand was assessed with the FASTX toolkit[86]. The obtained consensus sequences of the two parents were aligned to the selected coding DNA splice variant by MEGACC muscle[87] alignment implementation and *mafft*[88] alignment using default settings. All steps were implemented in a custom Julia script for processing of 127,149 gene splice variants. Introns as well as up- and downstream start and stop codons were removed. The presence of start and stop codons in the parental coding DNA alignment was tested and deletions counted. Additionally, the DNA sequence was translated into a putative sequence of amino acids with *Bio.Seq*[89] to estimate mismatches and stop codons. The most similar alignment of *muscle* and *mufft* to the coding DNA was selected. Some genes characterised by poor alignment quality were reprocessed on splice variant level. All splice variants of these genes were aligned to the cDNA and the most similar splice variant (compared to the reference coding DNA) was selected and added to the coding sequence list. In total, 28,601 complete CDS alignments were utilised for the identification and selection of candidate genes.

Selection signature–candidate genes and gene ontology

Meta-analysis of published QTL loci was performed. Several QTL studies have observed phenotype to genotype associations for numerous traits, including drought, yield physiology, resistance to biotic stressors and yield components. As many of these studies employed the 9K iSelect genotyping chip, a direct link of the genetic map of the QTL regions with marker-HAF was established. Other genetic maps and specifically designed markers reported in previous publications were matched to the physical or genetic map by sequence alignment using NCBI *Blastn* algorithm[90]. Potential candidate genes related to QTL were selected based on significant ($p < 0.001$) allele frequency variations between the environments or generations, the functional prediction[74] and the coding DNA variation. The identified candidate loci were clustered in the six agronomically relevant categories root morphology, biotic resistance, nutrient related, drought tolerance, yield components and yield physiology. The wtAF were compared generations wise for significant variations between the environments by an unpaired Wilcoxon test, using the *ggpubr* R package[91].

Using GO term annotations from the agriGo v2.0 database[92], gene ontologies (GO) were classified for two predominate weather impacts with potential to promote selection pressure. Fungal pressure was

investigated by comparing the allele frequency patterns of the F_{12} (2007) and F_{22} (2017) generations from the organic and conventional environments. These two generations were selected as they followed years with observed high fungal pressure. The conventional environment was used as a control, as minor or no selection in regard to fungi was expected. Variation analysis of the allele frequency was performed in accordance with the statistical approach highlighted in the **Supplementary Notes** section. Genes with significantly different allele frequency patterns were selected ($p < 0.05$). Only genes with significant differences in both years' allele frequency patterns were further processed. Genes were annotated to GO terms and a Mann–Whitney-U test was performed to assess the allele frequency variation between the F_3 and F_{12} / F_{22} generations as well as the conventional and organic populations to validate a targeted selection process.

To determine the impact of drought on the selection processes, the F_{12} generation was selected as a control or baseline (2007). This generation follows several years of sufficient water supply during the vegetation period, while water deficiencies occurred in 2011 (F_{16}), 2017 (F_{22}) and 2018 (F_{23}). The gene allele frequency pattern of these three generations was compared to the F_{12} generation of the organic and conventional systems, which created a test set of six comparisons. Variation analysis of the allele frequency was performed in accordance with the statistical approach highlighted in the **Supplementary Notes** section. Genes with significant difference in haplotype allele frequency profiles ($p < 0.05$, HAF level) were selected and the intersections of the six comparisons was identified. Those genes were annotated to GO terms and the Mann–Whitney U test was applied for each GO term to identify variations in the HAF of the F_3 and F_{12} vs. F_{16}, F_{22} and F_{23} generations, respectively. Only GO terms with a coverage of at least six genes were considered as potential candidates.

References Material and Methods

60. Chapter 2 - FAO Penman-Monteith equation. (2015).

61. Chapter 7 - ETc - Dual crop coefficient (Kc = Kcb + Ke). (2015).

62. Zotarelli, L. & Dukes, M. Step by step calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). *Inst. Food …* 1–10 (2010).

63. Meyer, W. S. & Green, G. C. Water Use by Wheat and Plant Indicators of Available Soil Water 1. *Agron. J.***72**, 253–257 (1980).

64. Minasny, B. & McBratney, A. B. Limited effect of organic matter on soil available water capacity. *Eur. J. Soil Sci.***69**, 39–47 (2018).

65. Pozníková, G., Fischer, M., Pohanková, E. & Trnka, M. Analyses of spring barley evapotranspiration rates based on gradient measurements and dual crop coefficient model. *Acta Univ. Agric. Silvic. Mendelianae Brun.***62**, 1079–1086 (2014).

66. de Mendiburu, F. agricolae: Statistical Procedures for Agricultural Research. (2020).

67. Cochran, W. G. *et al.Sampling Techniques third edition*. (1977).

68. Belkadi, A. *et al.* Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. U. S. A.***112**, 5473–5478 (2015).

69. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J. L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One***7**, (2012).

70. Zawada, A. M. *et al.* Massive analysis of cDNA Ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. *Epigenetics***9**, 161–172 (2014).

71. Nielsen, R., Paul, J. S., Albrechtsen, A. & Song, Y. S. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* vol. 12 443–451 (2011).

72. Olson, N. D. *et al.* Best practices for evaluating single nucleotide variant calling methods for microbial genomics. *Frontiers in Genetics* vol. 6 (2015).

73. Cunningham, F. *et al.* Ensembl 2019. *Nucleic Acids Res.***47**, D745–D751 (2019).

74. Barley Gene Function Annotation. https://webblast.ipk-gatersleben.de/barley_ibsc/downloads/.

75. Gaynor, C., Gorjanc, G. & Hickey, J. AlphaSimR: Breeding Program Simulations. (2020).

76. Sim Drift calculator.

77. Blighe, K. PCAtools: PCAtools: Everything Principal Components Analysis. (2019).

78. R Core Team. R: A Language and Environment for Statistical Computing. (2020).

79. Suzuki, R., Terada, Y. & Shimodaira, H. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. (2019).

80. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics***27**, 2156–2158 (2011).

81. Hoagland, D. R. & Arnon, D. I. Growing plants without soil by the water-culture method. *Grow. plants without soil by water-culture method.* (1938).

82. Neumann, K., Verburg, P. H., Stehfest, E. & Müller, C. The yield gap of global grain production: A spatial analysis. *Agric. Syst.***103**, 316–326 (2010).

83. Nakhforoosh, A., Grausgruber, H., Kaul, H. P. & Bodner, G. Wheat root diversity and root functional characterization. *Plant Soil***380**, 211–229 (2014).

84. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* vol. 9 671–675 (2012).

85. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics***25**, 2078–2079 (2009).

86. Gordon, A. & Hannon, G. Fastx-Toolkit. https://dl.acm.org/doi/10.1093/bioinformatics/bts507.

87. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.***32**, 1792–1797 (2004).

88. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.***30**, 772–780 (2013).

89. Ward, B. Bio.Seq Julia Package. https://biojulia.net/Bio.jl/stable/ (2020).

90. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

91. Kassambara, A. 'ggpubr': 'ggplot2' Based Publication Ready Plots. *R Packag. version 0.2.5.* (2020).

92. Tian, T. *et al.* AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **45**, W122–W129 (2017).
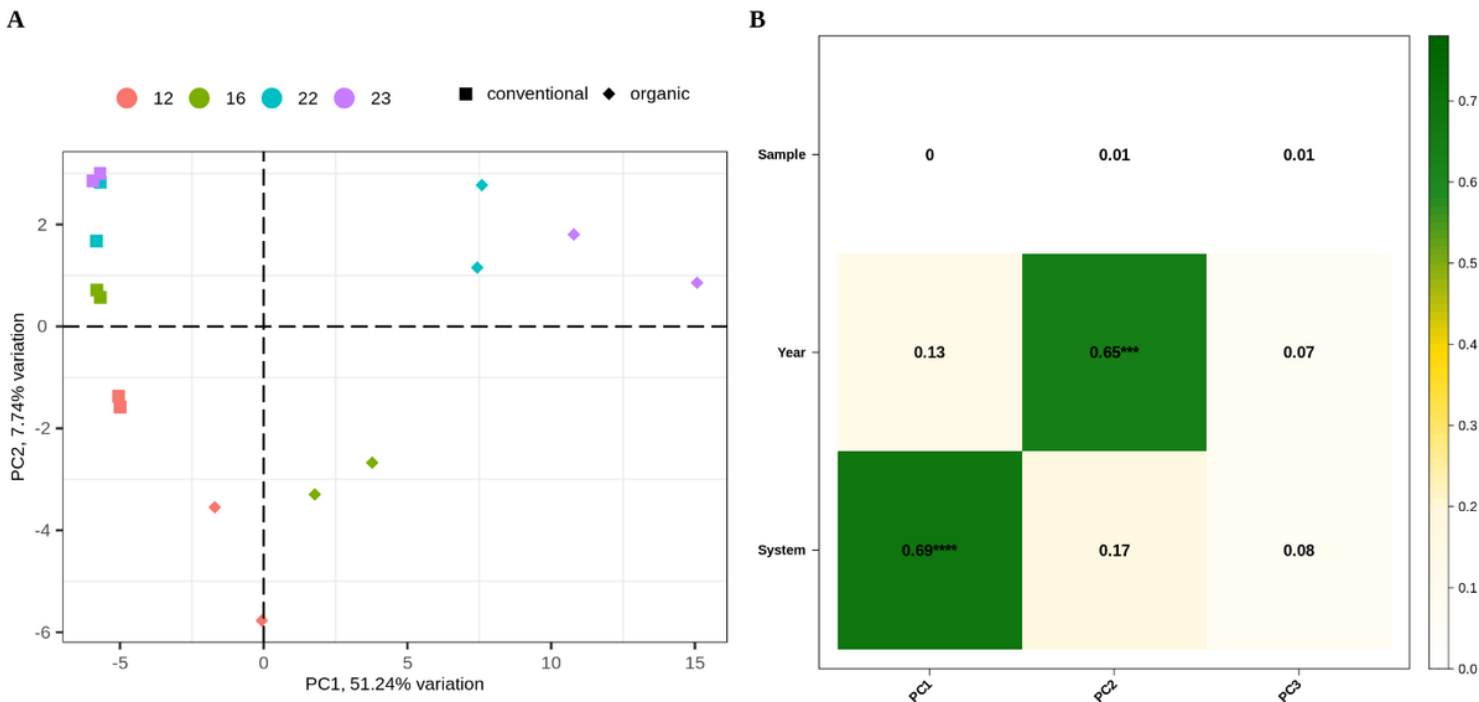
Figures

Figure 1

PCA of the tested populations from the 12th, 16th, 22nd and 23rd generations with both samples per population. The generations are presented as different colours, while different shapes indicate the different systems (A). The explanatory effect of the principle components is illustrated in (B), where the first PC explains the system impact, while the second PC highlights the generation / year effect. The samples had no 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%.
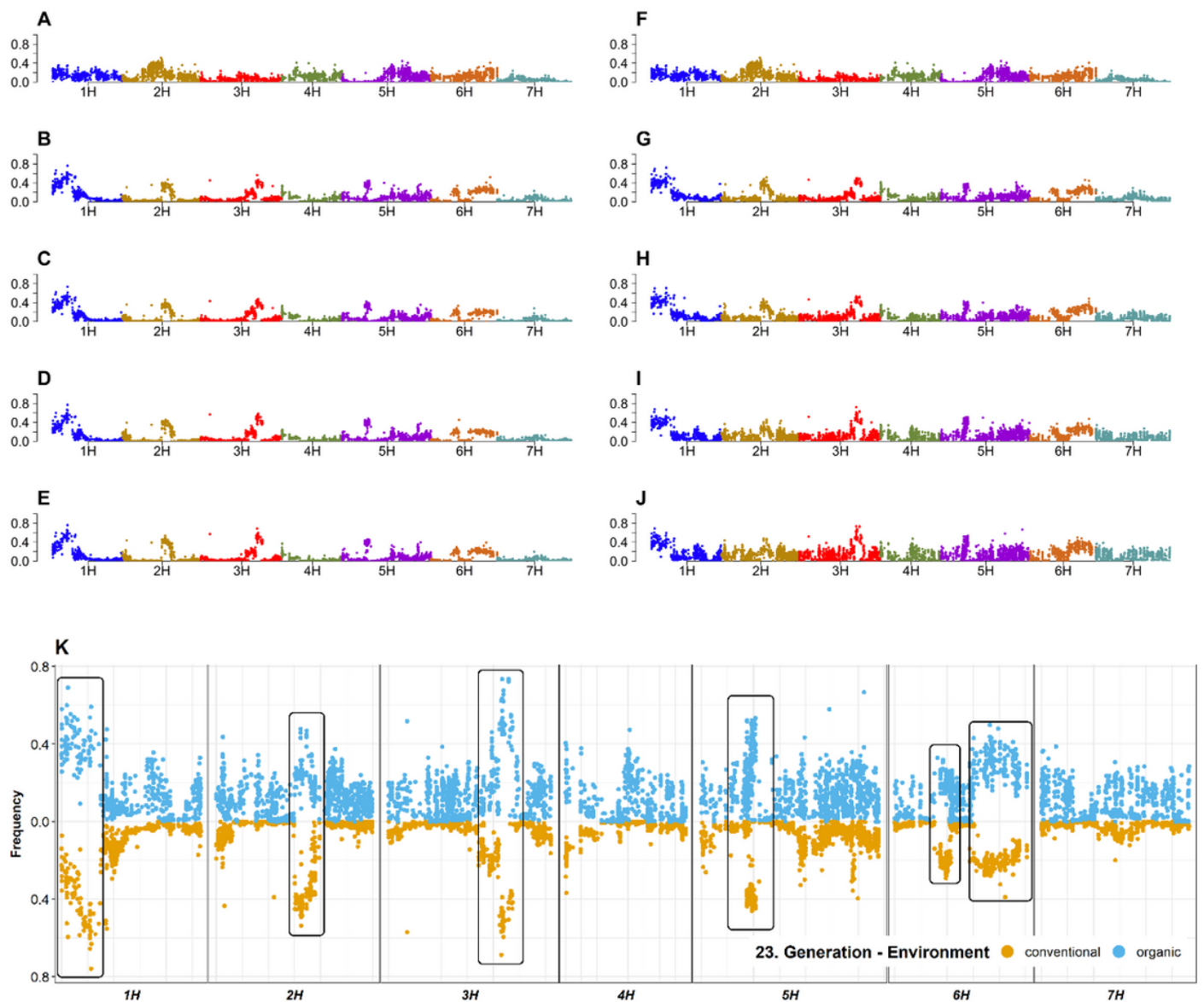
Figure 2

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 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 consensus genetic map (x-axis, 9KiSelect map origin59), where each of seven barley chromosomes is indicated by different colours. (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 colours. Boxes indicate regions with similar positive wild type allele frequency selection in both systems.
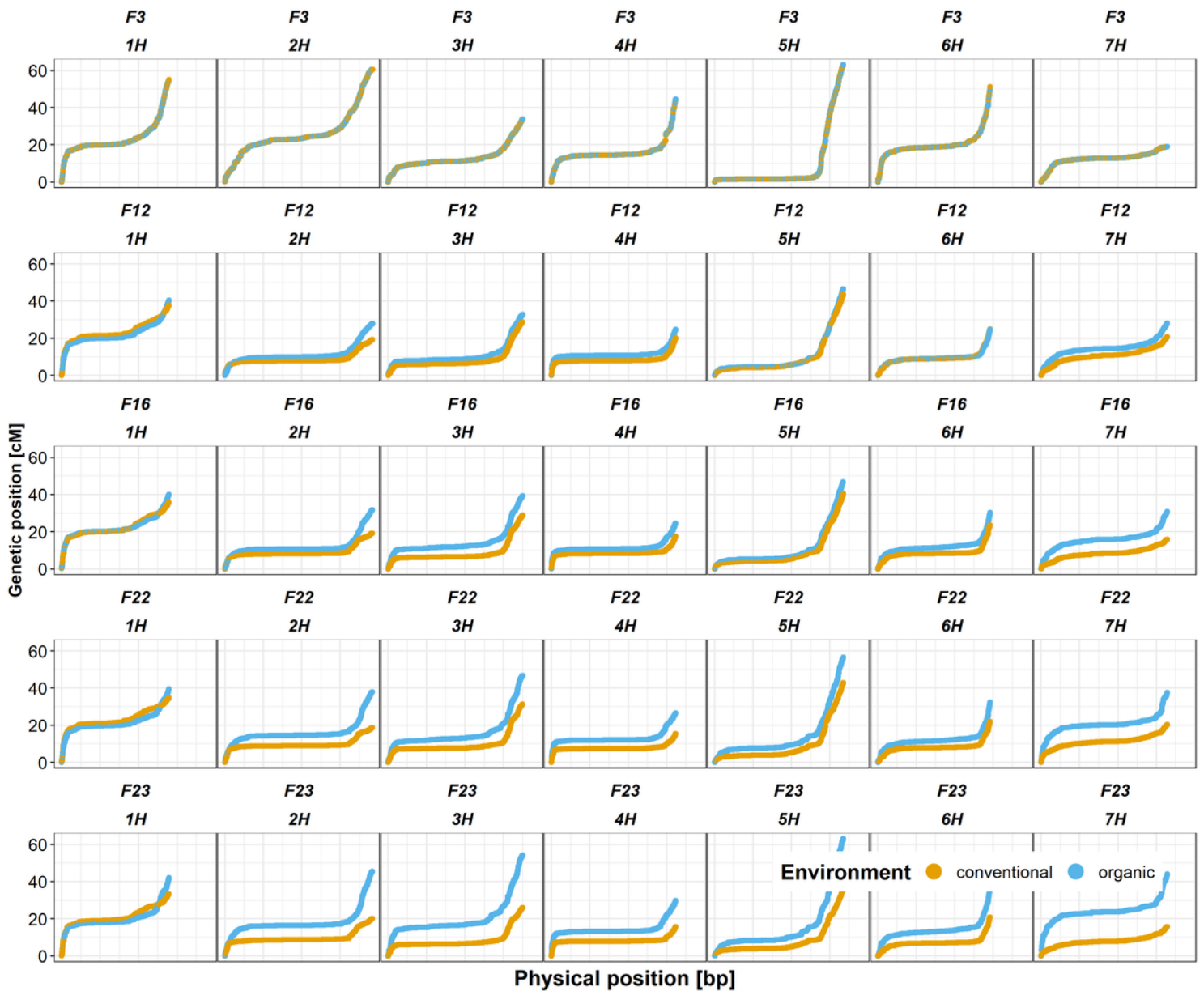
Figure 3

The genetic distance compared to the physical distance to investigate the recombination event likelihood. The environments are separated by colour. 9,000 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 higher genetic map length compared to the yellow curve of the conventional population. Extended map length indicates more variant selection processes, and a higher genetic variability.
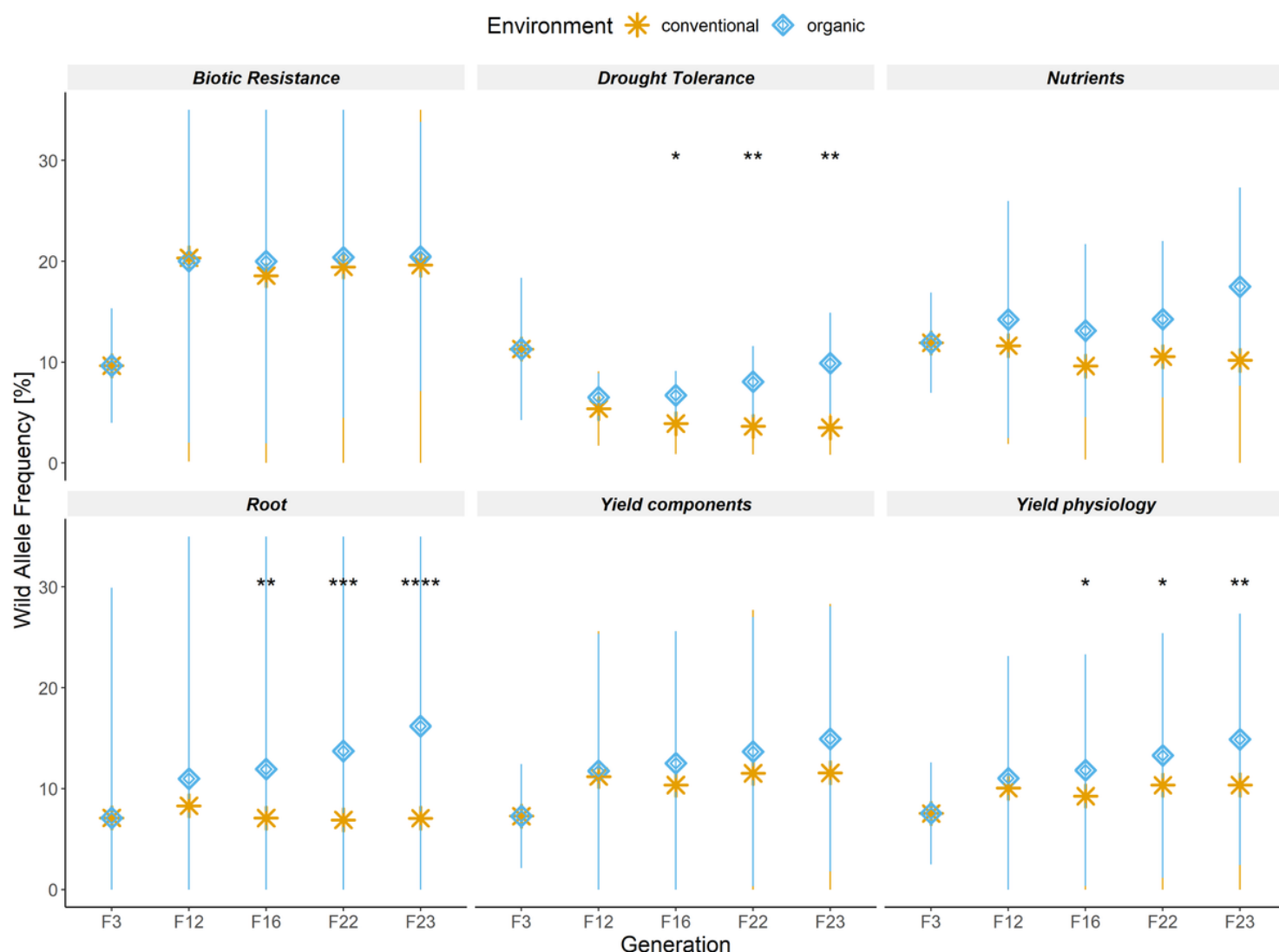
Figure 4

The wild donor allele frequencies in the QTL regions (based on Sup. Table 5), 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 and colour (organic, blue square; conventional, yellow star). The blue and yellow line give the standard deviation for the respective population, trait and generation. Significant levels between the populations are indicated by stars. 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 analysed studies used ISR42-8 as donor parent to create introgression lines.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ExtendedData.docx
- SupplementaryData.docx
- SupplementaryTable5.xlsx
- SupplementaryTable6.xlsx