Unlocking the diversity of genebanks: whole-genome marker analysis of Swiss bread wheat and spelt

Thomas Müller¹, Beate Schierscher-Viret², Dario Fossati², Cécile Brabant², Arnold Schori², Beat Keller¹* and Simon G. Krattinger¹,3*

¹Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland
²Department of Plant Production Sciences, Agroscope, Changins, Switzerland
³Current address: King Abdullah University of Science and Technology (KAUST), Biological and Environmental Sciences & Engineering Division (BESE), Thuwal, 23955-6900, Saudi Arabia

*Correspondence:
Beat Keller, E-mail: bkeller@botinst.uzh.ch, telephone: +41 44 6348230
Simon G. Krattinger, E-mail: simon.krattinger@kaust.edu.sa, telephone: +966 12 8082668

E-mail addresses:
Thomas Müller: thomas.mueller@botinst.uzh.ch
Beate Schierscher-Viret: beate.schierscher-viret@agroscope.admin.ch
Dario Fossati: dario.fossati@agroscope.admin.ch
Cécile Brabant: cecile.brabant@agroscope.admin.ch
Arnold Schori: arnold.schori@agroscope.admin.ch

Key message

High-throughput genotyping of Swiss bread wheat and spelt accessions revealed differences in their gene pools and identified bread wheat landraces that were not used in breeding.

Keywords
genebank; high-resolution genotyping; landraces; bread wheat; spelt; SNP

Acknowledgements

We thank Simon Fluckiger and Helen Zbinden for technical assistance with DNA extraction and Prof. Eduard Akhunov for providing the genotypic data of the international wheat accessions. This work was supported by the Swiss Federal Office for Agriculture (BLW) in the framework of NAP-PGREL (national plan of action for the conservation and sustainable utilization of plant genetic resources). S.G.K is supported by an Ambizione fellowship of the Swiss National Science Foundation.
Abstract

Genebanks play a pivotal role in preserving the genetic diversity present among old landraces and wild progenitors of modern crops and they represent sources of agriculturally important genes that were lost during domestication and in modern breeding. However, undesirable genes that negatively affect crop performance are often co-introduced when landraces and wild crop progenitors are crossed with elite cultivars, which often limits the use of genebank material in modern breeding programs. A detailed genetic characterization is an important prerequisite to solve this problem and to make genebank material more accessible to breeding. Here, we genotyped 502 bread wheat and 293 spelt accessions held in the Swiss National Genebank using a 15K wheat SNP array. The material included both spring and winter wheats and consisted of old landraces and modern cultivars. Genome- and sub-genome-wide analyses revealed that spelt and bread wheat form two distinct gene pools. In addition, we identified bread wheat landraces that were genetically distinct from modern cultivars. Such accessions were possibly missed in the early Swiss wheat breeding program and are promising targets for the identification of novel genes. The genetic information obtained in this study is appropriate to perform genome-wide association studies, which will facilitate the identification and transfer of agriculturally important genes from the genebank into modern cultivars through marker-assisted selection.
(2) Introduction

Hexaploid wheat (*Triticum aestivum*) is one of the most important global crops (FAO 2015). The most widely cultivated subspecies is bread wheat (*T. aestivum* ssp. *aestivum*). Spelt (*T. aestivum* ssp. *spelta*) is a close relative of bread wheat. It was the main wheat subspecies grown in Central Europe since the Bronze Age before bread wheat cultivation started to expand from the beginning of the 20th century (Akeret 2005; Jacquemin 2011; Schilperoord 2013). Today, spelt is only grown as a niche product in Central Europe. In contrast to free-threshing bread wheat, spelt is hulled and the kernels are surrounded by tenacious glumes. Both wheat subspecies can be interbred and crosses between winter wheat and spelt were systematically explored to improve the agronomical value of spelt (Winzeler et al. 1991; Siedler et al. 1994). On the other hand, spelt carries genes that were beneficial for bread wheat improvement (Kleijer et al. 2012). For example, the stripe rust resistance gene *Yr5*, first described in spelt accessions, was transferred into bread wheat (Macer 1966; Sun et al. 2002). Similarly, the leaf rust resistance gene *Lr65* was identified in a Swiss spelt from where it was subsequently introduced into the bread wheat gene pool (Mohler et al. 2012).

Domestication of wild progenitors of modern wheat started in the Near East around 10,000 years ago (Salamini et al. 2002). Compared to wild wheat relatives and old landraces, modern cultivars show a lower genetic diversity. This genetic bottleneck is due to the limited number of wild wheat progenitors and landraces that gave rise to modern wheat cultivars ( Tanksley and McCouch 1997; Reif et al. 2005b; Feuillet et al. 2008; Fu and Somers 2009). Several strategies are used to counteract this problem and to increase the genetic diversity in wheat breeding programs. For example, tetraploid...
wheat can be crossed with the wild diploid D-genome progenitor *Aegilops tauschii*, resulting in synthetic hexaploid wheat. This approach was widely explored by the International Maize and Wheat Improvement Center (CIMMYT) and other breeding programs (Mujeeb-Kazi et al. 1996; Dreisigacker et al. 2008). Another approach to increase diversity in modern cultivars is through the use of the genetic diversity present in landraces and wild wheat progenitors (Reynolds et al. 2006). Landraces are of interest to breeders because they often carry genes with beneficial effects that were not introduced into elite cultivars. For instance, the Swiss winter bread wheat landrace Muenstertaler was identified as a source of resistance to snow molds (Gaudet and Kozub 1991) and Swiss barley landraces from mountainous regions were identified as sources of stem rust resistance (Steffenson et al. 2016).

The value of landraces and wild wheat progenitors has long been recognized, which resulted in the systematic collection of plant genetic resources. Today, genebanks worldwide maintain this agricultural diversity by storing and propagating seeds of hundreds of thousands of wheat accessions (Börner et al. 2014). Hence, genebanks provide an enormous resource that can be used in research and breeding to make wheat more resilient to pests, diseases, or adverse climatic conditions. In Switzerland, this task is managed by the Swiss National Genebank, which has been established in the early 1900s. Until 1950, the focus of the bread wheat collection was on Swiss landraces, while spelt was also collected from Germany, Belgium, Austria, Liechtenstein, and Spain. Today, the Swiss genebank contains the largest spelt collection worldwide with more than 2,200 accessions (Kleijer et al. 2012). In addition, the genebank incorporates more than 5,600 bread wheat accessions.
For the most efficient use of genebank material it is important to have detailed genetic information. For example, breeders may be interested in using landraces that are genetically very different from current elite cultivars. Single nucleotide polymorphisms (SNPs) distributed across the entire genome represent a valuable tool to assess genome-wide diversity. It is nowadays possible to detect thousands of SNPs in a large number of accessions with high-throughput technologies like SNP arrays or genotyping-by-sequencing (GBS) in a reasonable time (Elshire et al. 2011; McCouch et al. 2012; Wang et al. 2014). SNP arrays consist of a predefined set of polymorphisms, have low error rates and low computational needs during data analysis. However, data generated by SNP arrays may suffer from an ascertainment bias that is caused by the selection of polymorphisms during array design (Albrechtsen et al. 2010). The 15K SNP array used in this study was mainly designed with polymorphisms identified in bread wheat and durum wheat accessions (Wang et al. 2014). The array consists of 13,006 gene-associated SNPs and was already successfully applied to genotype durum and bread wheat (Merchuk-Ovnat et al. 2016).

Here, we used the 15K wheat SNP array to genotypically characterize a core collection of 502 bread wheat and 293 spelt accessions of the Swiss National Genebank. The collection represents the history of Swiss wheat breeding and farming from the early 20th century to today. We show that bread wheat and spelt accessions represent two separate gene pools based on a large number of genome-wide markers. Based on the genotypic data, we identify two groups of bread wheat landraces that are genetically different from modern bread wheat cultivars. The obtained genotypic data can be used to narrow down the number of accessions to be used in breeding programs or to select accessions for genome-wide association studies based on their genetic diversity.
(3) Material and methods

Plant material

Seven-hundred-ninety-five hexaploid wheat accessions (502 T. aestivum ssp. aestivum and 293 T. aestivum ssp. spelta) from the Swiss National Genebank were used in this study (Online resource 1). Among the bread wheat accessions were 367 landraces, 55 breeder’s lines, and 120 cultivars from Switzerland. The remaining accessions came from Italy, Japan, USA, Russia, Mexico, Turkey (one cultivar per country), and France (one landrace, three cultivars). Accessions are defined as landraces if they were collected prior to 1950. Cultivars are officially registered accessions and breeder’s lines originated from breeding or research projects. Cultivars and breeder’s lines are both representing modern material and were grouped together, i.e., each bread wheat accession belongs to one of the following groups: winter bread wheat cultivar, winter bread wheat landrace, spring bread wheat cultivar or spring bread wheat landrace (Supp. Tab. S1).

The separation of spelt accessions into landraces and cultivars is difficult, because registered spelt cultivars were often collected before 1950 and can also be considered as landraces. In addition, modern spelt lines are in general crosses of spelt with bread wheat accessions. Therefore, we grouped the spelt accessions into winter spelt and spring spelt (242 accessions), representing accessions collected before 1950, and into spelt/wheat crosses (51 accessions) representing accessions originating from breeding or research programs (Supp. Tab. S1).
DNA extraction and genotyping

DNA from one plant per accession was extracted as described previously (Stein et al. 2001). Accessions were genotyped with an Illumina Infinium 15K wheat SNP array (TraitGenetics GmbH, Gatersleben, Germany) consisting of 13,006 SNPs. Haplotype-specific SNP markers were selected by allele frequency, functionality and sub-genome specificity in hexaploid wheat from the wheat 90K iSelect assay (Wang et al. 2014).

Eleven spelt and eleven wheat accessions were genotyped twice. Ninety-eight SNPs returned missing data in all accessions and 26 bread wheat and 11 spelt accessions (including one sample of a replicate) had missing data for all markers. These 37 accessions were excluded from the analysis. For the replicates, we kept the sample with fewer missing data and fewer heterozygous SNP calls for our analyses. Finally, 283 spelt and 476 bread wheat accessions were further analyzed. Genotyping data are deposited on the website of the Swiss National Genebank (bread wheat: https://www.bdn.ch/lists/1701/export/, spelt: https://www.bdn.ch/lists/1699/export/) and in Online resource 2.

A randomly selected set of 29 wheat and 30 spelt accessions were in addition genotyped using GBS (Supp. Tab. S2; Elshire et al. 2011; Poland et al. 2012). GBS was performed by the Genomic Diversity Facility at Cornell University, USA, using PstI as restriction enzyme. SNP calling was performed using the TASSEL 5 GBS v2 Pipeline of the TASSEL package (Glaubitz et al. 2014) using the TGACv1 assembly of wheat as reference (Clavijo et al. 2017).
Data analysis

Most analyses were performed in Python v3.4.3 using the libraries scipy v0.17.0, numpy v1.11.0 (van der Walt et al. 2011), sklearn v0.17.1 (Pedregosa et al. 2011), pandas v0.18.0 (McKinney 2010), matplotlib v1.5.1 (Hunter 2007) and ipython v4.2.0 (Perez and Granger 2007).

Genetic differentiation was determined using a simple Hamming distance (Hamming 1950; Wang et al. 2015), Rogers’ distance (Rogers 1972; Reif et al. 2005a), discriminant analysis of principal components (DAPC), and the fixation index $F_{ST}$. Because the original marker data were notated in the IUPAC notation (Cornish-Bowden 1985), we converted each marker data point into a two character code, e.g., 'A' to 'AA', 'K' to 'GT' and concatenated them in the same order for each accession to calculate the Hamming distance. The distance between two accessions was calculated as the number of mismatches between those converted strings. Missing data were treated like matches. For the calculation of Rogers’ distance missing data were imputed with the mean of all alleles. Rogers’ distance was then calculated with the R-package poppr v2.5 (Kamvar et al. 2014). R package adegenet v2.0.1 was used for DAPC (Jombart et al. 2010). DAPC combines a principal component analysis (PCA) with linear discriminant analysis. While PCA is based on the variation in the whole data set, DAPC results in clustering the data in a way that maximizes the variation between and minimizes variation within clusters. Pairwise $F_{ST}$ values were calculated between the seven classes of accessions using the Weir-Cockerham method implemented in vcftools v0.1.15 (Weir and Cockerham 1984; Danecek et al. 2011). $F_{ST}$ is a measure of population differentiation that, compared to PCA, is found to be less affected by a potential ascertainment bias in
SNP array data (Albrechtsen et al. 2010). Principal coordinate analyses (PCoA) based on pairwise Rogers’ distances and mean pairwise $F_{ST}$ values were performed with R-package *ape* v3.5 (Paradis et al. 2004). Nei’s $G_{ST}$ was calculated using vcfR (Knaus and Grünwald 2017).

Linkage disequilibrium (LD) of SNPs with a minor allele frequency greater than 0.05 was estimated by calculating the squared correlation coefficients ($r^2$) between genotypes with vcftools v0.1.15 (Danecek et al. 2011). To determine LD decay the $r^2$ values were plotted against genetic distances and an exponential curve was fitted in the data.

Phylogenetic trees were constructed using R packages *adegenet* v2.0.1 and *poppr* v2.3.0 with 1,000 bootstrap replicates (Jombart and Ahmed 2011; Kamvar et al. 2014).

**Genome-wide association study**

A genome-wide association study (GWAS) was performed using EMMAX (Kang et al. 2010) and the binary trait ‘type’ (winter - spring) with the bread wheat accessions. Heterozygous SNPs were set to missing, and SNPs with more than 20% missing data were excluded. The applied genetic map consisted of 9,809 SNPs at 2,887 different genetic positions (Wang et al. 2014). Only one SNP was kept, if SNPs at the same genetic position had equal genotypes. Missing data was then imputed by MACH1 (Li et al. 2010) and input files for EMMAX were generated with PLINK 1.9 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al. 2007) filtering out SNPs with minor allele frequencies below 5%. Balding-Nichols kinship matrix was used to account for population structure in the GWAS. Manhattan plot was made with R.
package *qqman* (Turner 2014). SNP reads, i.e., sequences around SNPs, were extracted from Table S5 of Wang et al. (2014).

**(4) Results**

**Genetic distances reveal groups of highly similar accessions**

Genotyping with the 15K wheat SNP array was successful for 476 bread wheat and 283 spelt accessions and resulted in 12,895 polymorphic SNPs across all accessions, 12,892 polymorphic SNPs across the bread wheat accessions alone, and 11,662 SNPs across the spelt accessions alone. Missing data and heterozygous SNP calls were only slightly higher in the spelt than in the bread wheat accessions (Wilcoxon rank sum test: $p < 0.001$ in both cases; Supp. Fig. S1). Mean heterozygosity rates in bread wheat and spelt were 0.5% and 0.4% and the mean missing data rate was 0.6% in both subsets. Missing data rates positively correlated with heterozygosity rates, which likely reflects problems during probe annealing of these accessions rather than actual heterozygosity (Supp. Fig. S1; Mengistu et al. 2016).

Hamming and Rogers’ distances allow to calculate the dissimilarity between individual accessions and consequently provide an estimate for the relatedness of accessions. The mean Hamming and Rogers’ distances between the bread wheat accessions (8,553.7 s.d. 495.7 and 0.335 s.d. 0.053, respectively) were higher than the mean distances between the spelt accessions (4,747.7 s.d. 1027.2 and 0.187 s.d. 0.066, respectively). Winter and spring spelt accessions were more similar to each other than to accessions resulting from crosses of bread wheat and spelt or to bread wheat accessions (Supp. Fig. S2). Replicates of the same accessions showed a high level of reproducibility (Supp.
Tab. S3). The mean Hamming distances of bread wheat and spelt replicates were 9.9 (s.d. 10.7) and 6 (s.d. 4), respectively. Based on these numbers, we selected a Hamming distance threshold of 25 to identify highly similar accessions across the bread wheat and spelt collections. This revealed 18 and 21 groups of highly similar spelt and bread wheats consisting of 63 and 44 accessions, respectively (Online resource 3). Hence, the fraction of highly similar accessions was estimated at 22% among the spelts and 9% among the bread wheats.

Our results might indicate that the spelt wheat gene pool analysed in this study is genetically narrower than the bread wheat gene pool, which is an observation that has been made previously (Siedler et al. 1994; Bertin et al. 2001; Blatter et al. 2004). Alternatively, it is possible that the lower diversity of the spelt wheat gene pool resulted from an ascertainment bias that is associated with the selection of polymorphisms to construct the 15K wheat SNP array. To check for a potential ascertainment bias, we randomly selected and genotyped a subset of 59 bread wheat and spelt accessions using GBS. In contrast to SNP arrays, GBS does not rely on a set of pre-selected SNPs and consequently is less prone to an ascertainment bias comparable to SNP arrays. However, GBS might introduce other biases related to the choice of the restriction enzyme (Arnold et al. 2013). The GBS data confirmed that the spelt accessions showed a lower genetic diversity than the bread wheat accessions (mean Hamming distances wheat 4,012.82 s.d. 137.4; spelt 3,332.95 s.d. 383.42; mean Rogers’ distances wheat 0.179 s.d. 0.016; spelt 0.143 s.d. 0.038).

A comparison of the minor allele frequency distribution between the 15k wheat SNP array and the GBS data revealed differences for bread wheat while the minor allele
frequency distribution was more similar for spelt (Supp. Fig. S3). Nei’s gene diversity index $G_{ST}$ was higher for the SNP array data compared to GBS data (Supp. Fig. S4), indicating that the array may overestimate the diversity, probably due to the choice of SNPs. On the other hand, it has been reported that GBS underestimates diversity (Arnold et al. 2013). In summary, both genotyping methods revealed that the spelt gene pool analyzed in this study was less diverse than the bread wheat gene pool and we conclude that a potential ascertainment bias had no influence on these results.

**Genetic analyses revealed two distinct gene pools for bread wheat and spelt**

The bread wheat and spelt accessions were clearly separated in a PCA along the first axis (Fig. 1a). This separation of bread wheat and spelt remained even after including a worldwide set of six spelt and 393 bread wheat accessions that were previously genotyped with a 90K SNP array (Fig. 1b) (Wang et al. 2014). These data indicate that the separation was not due to the narrow geographic distribution of the bread wheat accessions in the Swiss genebank. Spelt/wheat crosses, i.e., spelt accessions resulting from breeding programs that carry introgressions of bread wheat, located between the bread wheat and spelt clusters. The second axis of the PCA divided the bread wheat accessions into spring and winter types (Fig. 1a). In addition to PCA, we performed DAPC, PCoA of pairwise mean $F_{ST}$ values of the different groups of accessions and PCoA of Rogers’ distances with the SNP array data. The DAPC and both PCoA analyses confirmed the PCA results and revealed a clear separation of bread wheat and spelt (Supp. Fig. S5, S6, and S7). In addition, analysis of the 59 accessions genotyped by GBS confirmed the separation of bread wheat and spelt, indicating that these results are not caused by genotyping biases (Supp. Fig. S8).
Analyses of the bread wheat accessions alone revealed a separation of landraces from cultivars (Fig. 2, Supp. Fig. S9, S10, and S11, Online resources 4 and 5). We identified a cluster of winter bread wheat landraces and a cluster of spring bread wheat landraces that were distinct from the cultivars. The accessions of the two clusters originated mainly from mountainous regions and were most likely not used to generate the cultivars analyzed in our set. The accessions of the two clusters are also grouped together in a phylogenetic tree based on the genotypic data (Online resource 6). A PCA of spelt accessions alone showed no separation between spring and winter spelt accessions (Fig. 3), whereas DAPC and PCoA revealed a minor separation of winter and spring types (Supp. Fig. S12, S13, and S14). Spring spelt accessions were grouped together in the phylogenetic tree (Online resource 6). In summary, these results show that bread wheat and spelt wheat form discrete gene pools using the 15K SNP array and GBS. In addition, the identification of ‘unused’ bread wheat accessions confirms the usefulness of high-throughput genotyping of genebank material for the selection of accessions with potentially novel traits.

We also performed PCA and PCoA for the three wheat sub-genomes individually based on 4,186 A, 5,418 B, and 1,342 D genome-specific SNPs. The results for the sub-genomes were similar to the results using the entire SNP set (Supp. Fig. S15, Supp. Fig. S16). The separation of spelt and bread wheat was observed for each of the three sub-genomes. The separation between spring and winter bread wheat on the other hand was only observed for the A and B but not for the D sub-genome.
The SNP data set is suitable for GWAS

LD decay was calculated because the extent of LD determines the number of markers required for GWAS (Flint-Garcia et al. 2003; Vos et al. 2017). The LD decays in the A and B sub-genome were similar whereas LD decayed slowest in the D sub-genome (Supp. Fig. S17), which is in agreement with previous studies (Chao et al. 2010; Wang et al. 2014). The LD threshold of $r^2=0.1$ was reached after 1.94 cM in the wheat data and after 6.34 cM in the spelt data. A conservative estimation of the genetic map size of wheat is ~4,000 cM based on Wang et al. (2014). The size of the map, the LD decay and the number of polymorphic SNPs lead to the conclusion that it is possible to use our data set for GWAS.

An important difference between spring and winter accessions is the vernalization requirement of winter accessions, i.e., the need of exposure to a cold period to induce flowering. We used the simple binary trait 'type' (spring or winter) to test whether the genotypic information and the SNP-density of the 15K SNP array are indeed sufficient for GWAS. The GWAS was conducted on the bread wheat accessions using 9,737 SNPs (Online resource 1) and returned a peak on chromosome 5A (Fig. 4), in the region where the main vernalization gene VRN1A is located (Yan et al. 2003). The ten SNPs with lowest p-values were in the range of 89.56 cM to 91.3 cM (map positions based on Wang et al. (2014); Supp. Tab. S4). The extended sequence of the top SNP (wsnp_AJ612027A_Ta_2_1) produced a BLAST hit on a T. monococcum BAC clone (GenBank: AF459639) that was used for positional cloning of VRNI, showing the proximity of the GWAS peak to VRNI (Yan et al. 2003). These results demonstrate that
the SNP density in our data set is sufficient to perform GWAS, provided phenotypic
data are available.

**Discussion**

Naturally occurring genetic variation offers an enormous potential for crop
improvement. In addition, landraces preserved in genebanks represent an important
resource to discover and introduce novel genes into modern crop cultivars (Gaudet and
Kozub 1991; Steffenson et al. 2016). However, a systematic phenotypic description of
the many thousand accessions stored in genebanks is not feasible. It is therefore
important that effective choices of genebank accessions can be made by breeders. The
genetic characterization of genebank material, which is relatively cheap and easy, is a
valuable strategy to identify groups of similar accessions or to select landraces for
phenotypic testing (Kilian and Graner 2012; Mason et al. 2015). For example, breeders
might be interested in testing landraces that are very diverse from the cultivars in a
specific breeding program to maximize chances to identify novel genes.

In our study, we found clusters of bread wheat landraces that were very diverse from the
modern wheat cultivars. A possible explanation is that these accessions, which originate
from the mountainous regions Wallis and Graubünden in Switzerland, were collected in
1943, after the Swiss wheat breeding program started around 1900 and these accessions
were then not introduced into the already advanced wheat breeding program (Martinet
1907; Schilperoord 2006). Such accessions might be sources of genes controlling frost
tolerance, snow molds resistance and early maturing. A major drawback that often
limits the use of landraces in modern breeding is the co-introduction of undesired traits
(Feuillet et al. 2008). In comparison to improved cultivars, landraces often show inferior
yield, are tall and thus susceptible to lodging. To make landraces accessible for modern breeding it is essential that desired traits can effectively be separated from genes that negatively affect crop performance. A GWAS allows to identify associations between traits and molecular markers among hundreds of accessions. The identified markers can then be used to introduce desired genes into modern cultivars through methods such as marker-assisted backcrossing, thereby breaking the linkage drag (Collard et al. 2008). A GWAS relies on genome-wide distributed polymorphisms and accurate phenotypes. The SNPs of the 15K SNP array are located in genic regions. In addition, wheat is self-pollinating and has a larger linkage disequilibrium than other cereals (Cavanagh et al. 2013; Wang et al. 2014; Sukumaran et al. 2015). Those factors make the 15K wheat SNP array amenable for GWAS despite the wheat genome size of ~17 Gb and the relatively small SNP density of the array (1 SNP per 1.3 Mb). We showed that the amount and distribution of markers in our panel was sufficient to perform reliable genotype-phenotype analyses.

Our genetic analyses conducted with two different genotyping methods revealed that European spelt represents a gene pool that is distinct from the gene pool of bread wheat landraces and cultivars. This observation is consistent with previous observations (Siedler et al. 1994; Blatter et al. 2004; Dvorak et al. 2012). In contrast to these previous studies that only used few markers or short gene sequences, our analysis assessed the diversity of bread wheat and spelt on a genome-wide level with thousands of polymorphisms. Analyses of minor allele frequencies and Nei’s gene diversity index \( G_{ST} \) showed that the SNP array data may suffer from an ascertainment bias. However, our results were consistent using different analyses, inter alia a \( F_{ST} \)-based method which is reported to be less affected by a potential ascertainment bias (Albrechtsen et al.)
2010), and an additional genotyping method. Therefore, we conclude that our data are not influenced by a strong ascertainment bias that would affect our conclusions.

(6) Conclusion

The main task of genebanks has traditionally been the conservation of agricultural diversity and genebank material was not very frequently used in breeding in the past. Linkage drag and the co-introduction of undesirable genes are probably the two most important reasons for the limited use of old landraces and wild wheat progenitors in modern breeding. We showed that it is now feasible, with the advances in wheat genomics, to genotype large collections of spelt and bread wheat accessions from a genebank and to search for diverse accessions. In combination with high-precision phenotyping, this genotypic information can be used to identify novel genes through GWAS. These genes can then be transferred into modern cultivars through marker-assisted backcrossing, thereby avoiding linkage drag. These genomic advances will help to transform genebanks from 'storage facilities' into active reservoirs for plant breeding.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical standards

The authors declare that this study complies with the current laws of the countries in which the experiments were performed.
Author’s contributions

SGK, BSV, and BK designed the experiment. SGK, DF, CB, and AS provided seeds of accessions and information about these accessions. TM analyzed the data. Manuscript was written by TM, SGK, and BK. All authors read and approved the final manuscript.

(7) References


FAO (2015) FAO statistical pocketbook. Food and Agriculture Organization of the United Nations, Rome


Martinet G (1907) Expériences sur la sélection des céréales.


(8) Figure legends

**Figure 1**: Principal component analysis of bread wheat and spelt accessions. a) Bread wheat (left) and spelt (right) accessions are separated in Swiss material. b) PCA including a set of worldwide bread wheat and spelt accessions. 9,991 SNPs that were present in both the 90K SNP and the 15K SNP array were considered. Each dot represents an accession according to the color coding giving in the legend.
Figure 2: Principal component analysis of bread wheat accessions alone. Two clusters of landraces originating mainly from the Wallis region (winter accessions, circle A, Online resource 4) and from the Wallis and Graubünden regions (spring accessions, circle B, Online resource 5) show no (A) or only little (B) overlap to cultivars.
Figure 3: Principal component analysis of spelt accessions alone.
Figure 4: Manhattan plot of GWAS using EMMAX with trait ‘type’ (winter or spring bread wheat). The blue line corresponds to a p-value of $5.14 \times 10^{-6}$ (Bonferroni correction at a significance level of 0.05).

(9) Supporting information

Online resource 1 — Accession and passport data

Additional information about the accessions used in this study.

Online resource 2 — Genotypes

Genotypes of accessions.

Online resource 3 — Similar accessions

Additional information about highly similar accessions.
Online resource 4 — Winter bread wheat landraces with no overlap with cultivars

Passport data of winter bread wheat landraces of circle A in Figure 2.

Online resource 5 — Spring bread wheat landraces with only little overlap with cultivars

Passport data of spring bread wheat landraces of circle B in Figure 2.

Online resource 6 — Phylogenetic tree

Phylogenetic tree of all accessions. Tree was constructed using R packages adegenet and poppr. Color codes: purple: spring spelt, dark green: winter spelt, light green: wheat/spelt cross, yellow: wheat winter landrace, red: wheat winter cultivar, light blue: wheat spring landrace, dark blue: wheat spring cultivar.

Online resource 7 — Supporting figures and tables

Supporting figures S1-S17 and supporting tables S1-S4.