Theor Appl Genet (2018) 131:407–416
https://doi.org/10.1007/s00122-017-3010-5

ORIGINAL ARTICLE

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

Thomas Müller[1] · Beate Schierscher-Viret[2] · Dario Fossati[2] · Cécile Brabant[2] · Arnold Schori[2] · Beat Keller[1] · Simon G. Krattinger[1,3]

Received: 14 October 2017 / Accepted: 26 October 2017 / Published online: 4 November 2017

Abstract

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.**

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 limit 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.

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*).

Communicated by Andreas Graner.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00122-017-3010-5) contains supplementary material, which is available to authorized users.

✉ Beat Keller
bkeller@botinst.uzh.ch

✉ Simon G. Krattinger
simon.krattinger@kaust.edu.sa

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

[1] Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland

[2] Department of Plant Production Sciences, Agroscope, Changins, Nyon, Switzerland

[3] Present Address: Biological and Environmental Sciences and Engineering Division (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

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 2200 accessions (Kleijer et al. 2012). In addition, the genebank incorporates more than 5600 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 represents 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 identified 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.

Materials 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, 5 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 dataset, 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 1000 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 9809 SNPs at 2887 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).

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 accession, 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 (8553.7 s.d. 495.7 and 0.335 s.d. 0.053, respectively) were higher than the mean distances between the spelt accessions (4747.7 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 analyzed 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 4012.82 s.d. 137.4; spelt 3332.95 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

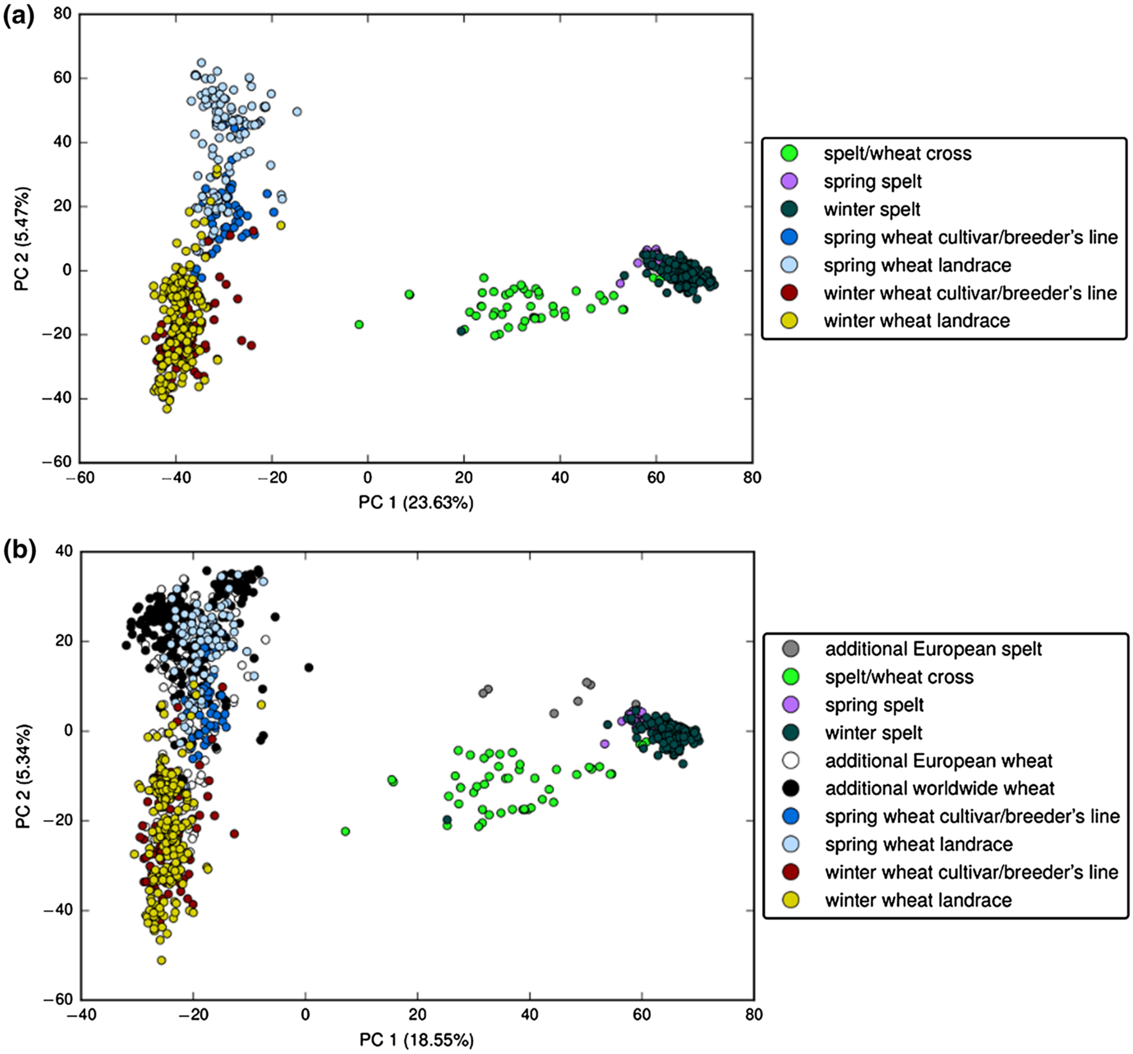

Fig. 1 Principal component analysis of bread wheat and spelt accessions. **a** Bread wheat (left) and spelt (right) accessions are separated in Swiss material. **b** PCA including a set of worldwide bread wheat and spelt accessions. 9991 SNPs that were present in both the 90K SNP and the 15K SNP array were considered. Each dot represents an accession according to the color coding given in the legend

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. Figs. 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. Figs. 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. Figs. S12, S13, and S14). Spring spelt accessions were grouped together in the phylogenetic tree (Online resource 6). In summary, these results show that

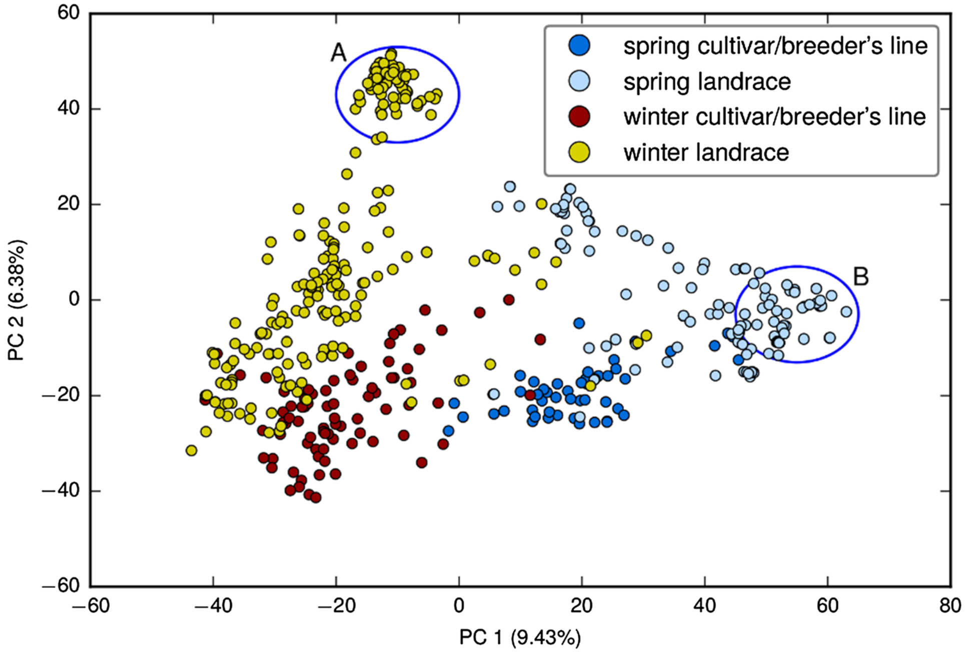

Fig. 2 Principal component analysis of bread wheat accessions alone. Two clusters of landraces originating mainly from the Wallis region (winter accessions, circle A, Online resource 4) and from the Wallis and Graubünden regions (spring accessions, circle B, Online resource 5) show no (A) or only little (B) overlap to cultivars

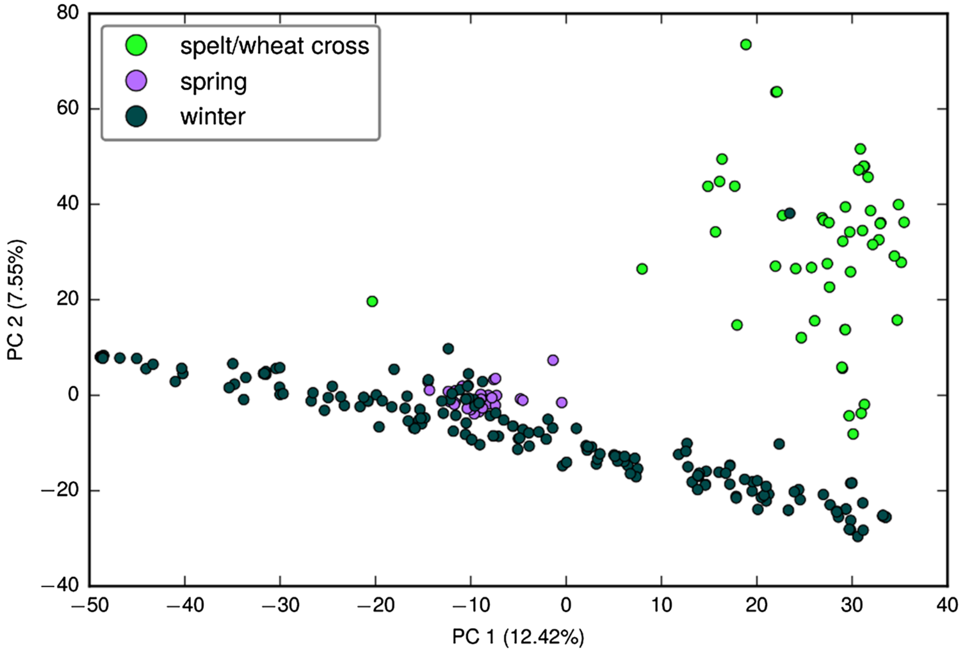

Fig. 3 Principal component analysis of spelt accessions alone

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 4186 A, 5418 B, and 1342 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 ~ 4000 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 dataset 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 9737 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 89.56–91.3 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 *VRN1*, showing the proximity of the GWAS peak to *VRN1* (Yan et al. 2003). These results demonstrate that the SNP density in our dataset 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.

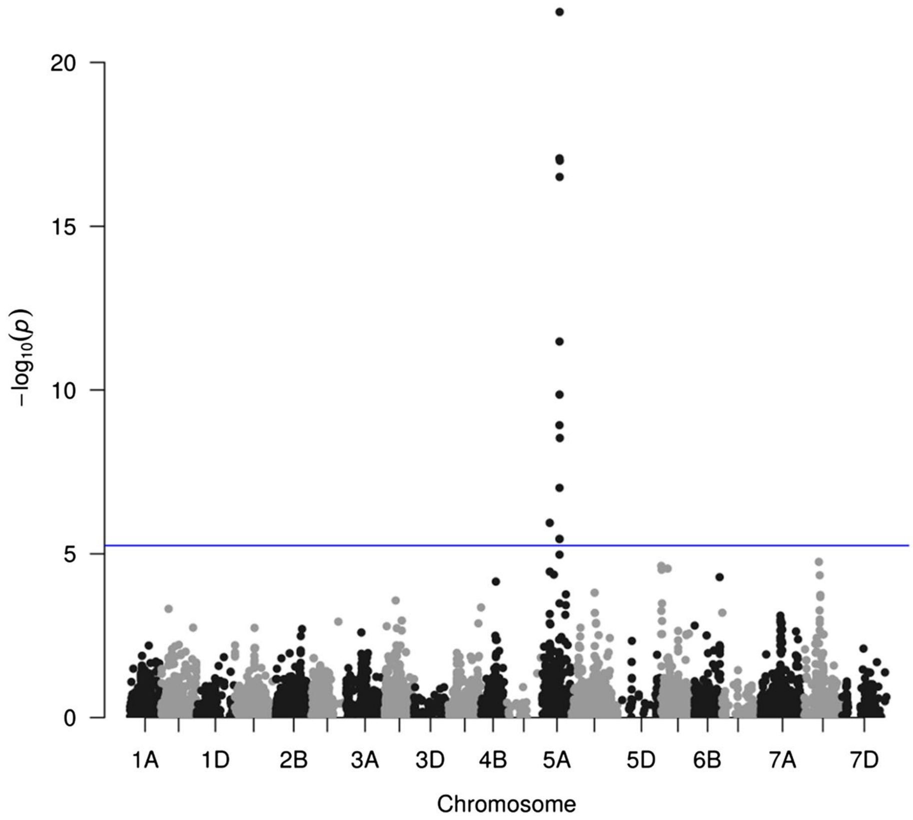

Fig. 4 Manhattan plot of GWAS using EMMAX with trait 'type' (winter or spring bread wheat). The blue line corresponds to a *p* value of 5.14×10^{-6} (Bonferroni correction at a significance level of 0.05)

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 identifying 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.

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.

Author contribution statement 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.

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). SGK is supported by an Ambizione fellowship of the Swiss National Science Foundation.

Compliance with ethical standards

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.

References

Akeret Ö (2005) Plant remains from a Bell Beaker site in Switzerland, and the beginnings of *Triticum spelta* (spelt) cultivation in Europe. Veg Hist Archaeobot 14:279–286. https://doi.org/10.1007/s00334-005-0071-1

Albrechtsen A, Nielsen FC, Nielsen R (2010) Ascertainment biases in SNP chips affect measures of population divergence. Mol Biol Evol 27:2534–2547. https://doi.org/10.1093/molbev/msq148

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol Ecol 22:3179–3190. https://doi.org/10.1111/mec.12276

Bertin P, Grégorie D, Massart S, de Froidmont D (2001) Genetic diversity among European cultivated spelt revealed by microsatellites. Theor Appl Genet 102:148–156. https://doi.org/10.1007/s001220051630

Blatter RHE, Jacomet S, Schlumbaum A (2004) About the origin of European spelt (*Triticum spelta* L.): allelic differentiation of the HMW glutenin B1-1 and A1-2 subunit genes. Theor Appl Genet 108:360–367. https://doi.org/10.1007/s00122-003-1441-7

Börner A, Landjeva S, Nagel M et al (2014) Plant genetic resources for food and agriculture (PGRFA) maintenance and research. Genet Plant Physiol 4:13–21

Cavanagh CR, Chao S, Wang S et al (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci 110:8057–8062. https://doi.org/10.1073/pnas.1217133110

Chao S, Dubcovsky J, Dvorak J et al (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genom 11:727. https://doi.org/10.1186/1471-2164-11-727

Clavijo BJ, Venturini L, Schudoma C et al (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome Res 27:885–896. https://doi.org/10.1101/gr.217117.116

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B 363:557–572. https://doi.org/10.1098/rstb.2007.2170

Cornish-Bowden A (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Res 13:3021–3030. https://doi.org/10.1093/nar/13.9.3021

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dreisigacker S, Kishii M, Lage J et al (2008) Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. Aust J Agric Res 59:413. https://doi.org/10.1071/AR07225

Dvorak J, Deal KR, Luo MC et al (2012) The origin of spelt and free-threshing hexaploid wheat. J Hered 103:426–441. https://doi.org/10.1093/jhered/esr152

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. https://doi.org/10.1371/journal.pone.0019379

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

Feuillet C, Langridge P, Waugh R (2008) Cereal breeding takes a walk on the wild side. Trends Genet 24:24–32. https://doi.org/10.1016/j.tig.2007.11.001

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Fu YB, Somers DJ (2009) Genome-wide reduction of genetic diversity in wheat breeding. Crop Sci 49:161–168. https://doi.org/10.2135/cropsci2008.03.0125

Gaudet DA, Kozub GC (1991) Screening winter wheat for resistance to cottony snow mold under controlled conditions. Can J Plant Sci 71:957–966. https://doi.org/10.4141/cjps91-138

Glaubitz JC, Casstevens TM, Lu F et al (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One. https://doi.org/10.1371/journal.pone.0090346

Hamming RW (1950) Error detecting and error correcting codes. Bell Syst Tech J 29:147–160. https://doi.org/10.1002/j.1538-7305.1950.tb00463.x

Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55

Jacquemin JM (2011) Wheat breeding in Belgium. In: Bonjean AP, Angus WJ, van Ginkel M (eds) The world wheat book: a history of wheat breeding, vol 2. Lavoisier, Paris

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:21

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281. https://doi.org/10.7717/peerj.281

Kang HM, Sul JH, Service SK et al (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354. https://doi.org/10.1038/ng.548

Kilian B, Graner A (2012) NGS technologies for analyzing germplasm diversity in genebanks. Brief Funct Genom 11:38–50

Kleijer G, Schori A, Schierscher B (2012) Die nationale Genbank von Agroscope ACW gestern, heute und morgen. Agrar Schweiz 3:408–413

Knaus BJ, Grünwald NJ (2017) vcfr: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour 17:44–53. https://doi.org/10.1111/1755-0998.12549

Li Y, Willer CJ, Ding J et al (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34:816–834. https://doi.org/10.1002/gepi.20533

Macer RCF (1966) The formal and monosomic genetic analysis of stripe rust (*Puccinia striiformis*) resistance in wheat. In: Mackey I (ed) Proceedings of the second international wheat genetics symposium, Lund, Sweden 1963. Hereditas Suppl 2. pp 127–142

Martinet G (1907) Expériences sur la sélection des céréales

Mason AS, Zhang J, Tollenaere R et al (2015) High-throughput genotyping for species identification and diversity assessment in germplasm collections. Mol Ecol Resour 15:1091–1101. https://doi.org/10.1111/1755-0998.12379

McCouch SR, McNally KL, Wang W, Hamilton RS (2012) Genomics of gene banks: a case study in rice. Am J Bot 99:407–423. https://doi.org/10.3732/ajb.1100385

McKinney W (2010) Data structures for statistical computing in Python. In: Proceedings of the 9th Python in science conference, pp 51–56

Mengistu DK, Kidane YG, Catellani M et al (2016) High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. Plant Biotechnol J 14:1800–1812. https://doi.org/10.1111/pbi.12538

Merchuk-Ovnat L, Barak V, Fahima T et al (2016) Ancestral QTL alleles from wild emmer wheat improve drought resistance and productivity in modern wheat cultivars. Front Plant Sci 7:452. https://doi.org/10.3389/fpls.2016.00452

Mohler V, Singh D, Singrün C, Park RF (2012) Characterization and mapping of *Lr65* in spelt wheat "Altgold Rotkorn". Plant Breed 131:252–257. https://doi.org/10.1111/j.1439-0523.2011.01934.x

Mujeeb-Kazi A, Rosas V, Roldan S (1996) Conservation of the genetic variation of *Triticum tauschii* (Coss.) Schmalh. (*Aegilops squarrosa* auct. non L.) in synthetic hexaploid wheats (*T. turgidum* L. s.lat. × *T. tauschii*; 2n = 6x = 42, AABBDD) and its potential utilization for wheat improvement. Genet Resour Crop Evol 43:129–134. https://doi.org/10.1007/BF00126756

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290. https://doi.org/10.1093/BIOINFORMATICS/BTG412

Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Perez F, Granger BE (2007) IPython: a system for interactive scientific computing. Comput Sci Eng 9:21–29. https://doi.org/10.1109/MCSE.2007.53

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253. https://doi.org/10.1371/journal.pone.0032253

Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1086/519795

Reif JC, Melchinger AE, Frisch M (2005a) Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Sci 45:1. https://doi.org/10.2135/cropsci2005.0001

Reif JC, Zhang P, Dreisigacker S et al (2005b) Wheat genetic diversity trends during domestication and breeding. Theor Appl Genet 110:859–864. https://doi.org/10.1007/s00122-004-1881-8

Reynolds M, Dreccer F, Trethowan R (2006) Drought-adaptive traits derived from wheat wild relatives and landraces. J Exp Bot 58:177–186. https://doi.org/10.1093/jxb/erl250

Rogers JS (1972) Measures of genetic similarity and genetic distance. Stud Genet 7:145–153

Salamini F, Ozkan H, Brandolini A et al (2002) Genetics and geography of wild cereal domestication in the Near East. Nat Rev Genet 3:429–441. https://doi.org/10.1038/nrg817

Schilperoord P (2006) Die Bedeutung des Getreidearchivs der Forschungsanstalt für Agrarökologie und Landwirtschaft Zürich-Reckenholz für die nationale Samenbank. Arbeitsbericht III NAP 02-231

Schilperoord P (2013) Kulturpflanzen in der Schweiz—Dinkel. Verein für alpine Kulturpflanzen, Alvaneu

Siedler H, Messmer MM, Schachermayr GM et al (1994) Genetic diversity in European wheat and spelt breeding material based on RFLP data. Theor Appl Genet 88:994–1003. https://doi.org/10.1007/BF00220807

Steffenson BJ, Solanki S, Brueggeman RS (2016) Landraces from mountainous regions of Switzerland are sources of important genes for stem rust resistance in barley. Alp Bot 126:23–33. https://doi.org/10.1007/s00035-015-0161-3

Stein N, Herren G, Keller B (2001) A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. Plant Breed 120:354–356. https://doi.org/10.1046/j.1439-0523.2001.00615.x

Sukumaran S, Dreisigacker S, Lopes M et al (2015) Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor Appl Genet 128:353–363. https://doi.org/10.1007/s00122-014-2435-3

Sun Q, Wei Y, Ni Z et al (2002) Microsatellite marker for yellow rust resistance gene *Yr5* in wheat introgressed from spelt wheat. Plant Breed 121:539–541. https://doi.org/10.1046/J.1439-0523.2002.00754.X

Tanksley S, McCouch S (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science (80-) 277:1063–1066. https://doi.org/10.1126/science.277.5329.1063

Turner SD (2014) qqman: an R package for visualizing GWAS results using Q–Q and manhattan plots. bioRxiv. https://doi.org/10.1101/005165

van der Walt S, Colbert SC, Varoquaux G (2011) The NumPy array: a structure for efficient numerical computation. Comput Sci Eng 13:22–30. https://doi.org/10.1109/MCSE.2011.37

Vos PG, Paulo MJ, Voorrips RE et al (2017) Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. Theor Appl Genet 130:123–135. https://doi.org/10.1007/s00122-016-2798-8

Wang S, Wong D, Forrest K et al (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796. https://doi.org/10.1111/pbi.12183

Wang C, Kao W-H, Hsiao CK et al (2015) Using Hamming distance as information for SNP-sets clustering and testing in disease association studies. PLoS One 10:e0135918. https://doi.org/10.1371/journal.pone.0135918

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution (New York) 38:1358. https://doi.org/10.2307/2408641

Winzeler H, Schmid J, Winzeler M, Rüegger A (1991) Neue Aspekte der Dinkelzüchtung (*Triticum spelta* L.) in der Schweiz. In: 2. Hohenheimer Dinkelkolloquium, Universität Hohenheim, pp 11–25

Yan L, Loukoianov A, Tranquilli G et al (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proc Natl Acad Sci USA 100:6263–6268. https://doi.org/10.1073/pnas.0937399100