Theor Appl Genet (2013) 126:1477–1486
DOI 10.1007/s00122-013-2065-1

ORIGINAL PAPER

Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers

Tobias Würschum · Simon M. Langer · C. Friedrich H. Longin · Viktor Korzun · Eduard Akhunov · Erhard Ebmeyer · Ralf Schachschneider · Johannes Schacht · Ebrahim Kazman · Jochen C. Reif

Received: 8 October 2012 / Accepted: 8 February 2013 / Published online: 22 February 2013

Abstract Modern genomics approaches rely on the availability of high-throughput and high-density genotyping platforms. A major breakthrough in wheat genotyping was the development of an SNP array. In this study, we used a diverse panel of 172 elite European winter wheat lines to evaluate the utility of the SNP array for genomic analyses in wheat germplasm derived from breeding programs. We investigated population structure and genetic relatedness and found that the results obtained with SNP and SSR markers differ. This suggests that additional research is required to determine the optimum approach for the investigation of population structure and kinship. Our analysis of linkage disequilibrium (LD) showed that LD decays within approximately 5–10 cM. Moreover, we found that LD is variable along chromosomes. Our results suggest that the number of SNPs needs to be increased further to obtain a higher coverage of the chromosomes. Taken together, SNPs can be a valuable tool for genomics approaches and for a knowledge-based improvement of wheat.

Communicated by X. Xia.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2065-1) contains supplementary material, which is available to authorized users.

T. Würschum (✉) · S. M. Langer · C. F. H. Longin · J. C. Reif
State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany
e-mail: tobias.wuerschum@uni-hohenheim.de

V. Korzun · E. Ebmeyer
KWS LOCHOW GMBH, 29303 Bergen, Germany

E. Akhunov
Department of Plant Pathology, 4024 Throckmorton Plant, Sciences Center, Kansas State University, Manhattan, KS 66506, USA

R. Schachschneider
Nordsaat Saatzuchtgesellschaft mbH, 38895 Langenstein, Germany

J. Schacht
Limagrain GmbH, 31226 Peine-Rosenthal, Germany

E. Kazman
Lantmännen SW Seed Hadmersleben GmbH, 39398 Hadmersleben, Germany

Present Address:
J. C. Reif
Leibniz Institute of Plant Genetics and Crop Plant Research, 06466 Gatersleben, Germany

Introduction

Bread wheat (*Triticum aestivum* L.) is of fundamental importance for human nutrition (Bonjean and Angus 2001). It is grown worldwide and shows a high genetic diversity. Bread wheat is a hexaploid crop possessing the A, B, and D genome. Knowledge-based improvement of this important crop requires the availability and the characterization of molecular markers which in addition should be suitable for high-throughput, high-density genotyping at low costs (Eathington et al. 2007).

Single nucleotide polymorphisms (SNPs) are the most common type of sequence variation in the genome (Rafalski 2002). They are consequently well suited for genomics approaches for which a high number of markers

are required such as association mapping (Myles et al. 2009) or genomic selection (Meuwissen et al. 2001).

High-throughput SNP genotyping platforms have been available for diploid crops like maize (e.g., Yan et al. 2010) or barley (e.g., Sako and Takeda 2009; Close et al. 2009) for some time, but not for polyploid species. For wheat and other polyploid crops, SNP genotyping is complicated by the presence of two or more copies of each locus in the nucleus, i.e., one on each genome. Consequently, genotypic data in wheat has so far been obtained with RFLP (Chao et al. 1989), AFLP (Barrett and Kidwell 1998), SSR (Röder et al. 1998), and DArT (Akbari et al. 2006) markers. A breakthrough in wheat genotyping was the development of a SNP array enabling the cost-effective, large-scale acquisition of genotypic data in hexaploid wheat (Akhunov et al. 2009). The utility of the SNP array for genomic analyses in elite wheat, however, needs to be demonstrated.

Diversity studies and evaluation of population structure have commonly been done with multiallelic SSR markers which compared to the biallelic SNPs have a higher information content (e.g., Dreisigacker et al. 2005; Reif et al. 2005; Roussel et al. 2005). This is due to differences in the mutational properties of both types of polymorphism. The mutational rate is much higher for SSRs and new alleles can also be created by the slippage process (i.e., increase or decrease in repeat length due to misalignments during the replication process) (Ellegren 2004). Consequently, SSRs are often characterized by an increased frequency of rare alleles. By contrast, SNPs on genotyping arrays show mainly intermediate frequency alleles, as they have been selected for heterozygosity in a small panel of founder lines, thus creating an ascertainment bias which further exaggerates the difference in the frequency spectrum between SSRs and SNPs.

With the increased availability and use of SNP markers it is important to characterize the performance and the specific properties of these two commonly used marker types. Jones et al. (2007) found that in a panel of maize inbred lines the genetic distance estimates based on SSRs, SNPs, or SNP haplotypes were not significantly correlated, except for individuals related by pedigree. This finding was confirmed by Hamblin et al. (2007) who observed a moderate correlation only for closely related lines. Another comparison in maize showed that similar conclusions on population structure and genetic diversity could be drawn from both marker types, but that a substantially higher number of SNPs must be used as compared to the SSRs (Van Inghelandt et al. 2010).

Linkage disequilibrium (LD) refers to the non-random association of alleles at two separate loci and is the basis for association mapping approaches. LD is affected by different genetic forces such as recombination, selection, mutation and drift (Flint-Garcia et al. 2003). The extent and the pattern of LD determine the mapping resolution and the QTL detection power that can be achieved in association mapping with a given marker density (Myles et al. 2009; Würschum 2012). Generally, association mapping relies on the availability of a high number of markers distributed throughout the entire genome.

The aim of our study was to use the newly developed SNP array and test its utility for genomic approaches in a European elite winter wheat population. In particular, our objectives were to genotype 172 wheat lines with the 9k SNP array as well as with SSR markers to (1) characterize the genome distribution of the SNPs and their information content; (2) assess population structure and genetic diversity with both marker types and explore how the choice of the marker type affects the results; and (3) evaluate the extent and the pattern of linkage disequilibrium based on the SNP markers.

Materials and methods

Plant materials

This study was based on 172 European winter wheat (*T. aestivum* L.) lines, including registered varieties as well as experimental lines. All materials used in this study were provided by the breeding companies KWS LOCHOW GmbH (Bergen, Germany), Limagrain GmbH (Peine-Rosenthal, Germany), Nordsaat Saatzuchtgesellschaft mbH (Langenstein, Germany), and Lantmännen SW Seed Hadmersleben GmbH (Hadmersleben, Germany).

Molecular markers

DNA was extracted according to standard procedures and all lines were genotyped with the 9k SNP array based on the Illumina Infinium assay (Cavanagh et al., in review). 8,630 SNPs were scored on the array. Markers were removed from the data set if they were either monomorphic, showed more than 20 % missing values, had ambiguous SNP calling, or had a minor allele frequency of <5 %. After these checks, 1,395 high-quality SNP markers remained in the data set of which for 518 a map position was available (Cavanagh et al., in review). In addition, the lines were genotyped with 91 SSR markers distributed throughout the genome.

Molecular analyses

The polymorphic information content (PIC) was calculated for every marker as $\mathrm{PIC} = 1 - \sum_{i=1}^{n} p_i^2$, where p_i^2 is the squared frequency of allele i at each locus. Relationships among the 172 genotypes were analyzed by applying principal coordinate analysis (PCoA) (Gower 1966) based

on the modified Rogers' distances (Wright 1978). The modified Rogers' distances (d_w) were calculated as $d_w = \frac{1}{\sqrt{2m}}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}(p_{ij}-q_{ij})^2}$ and genetic similarities were calculated as $1 - d_w$. Relationships among the 172 genotypes were also analyzed using average linkage clustering based on the genetic similarity matrix.

Linkage disequilibrium (LD) was assessed by the LD measure r^2 (Weir 1996) and significance of LD was tested with Fisher's exact tests (Hill and Robertson 1968). The association between LD decay and genetic map distance was summarized by fitting a curve by locally weighted regression to r^2 values that were plotted against the genetic map distance. To obtain a threshold of r^2 above which LD was likely to be caused by genetic linkage the 95 % percentile derived from the distribution of the r^2 values of unlinked loci was taken as a population-specific critical value (Breseghello and Sorrells 2006). The intersection of the fitted curve of r^2 values with this threshold was considered as the estimate of the range of LD. LD along chromosomes was assessed by a sliding window approach with 5 cM windows at 500 positions along the chromosomes. LD and PCoA computations were performed with the software package Plabsoft (Maurer et al. 2008).

Results

Analysis of SNP markers

A total of 8,630 SNPs were genotyped with the 9k Illumina Infinium SNP array. Out of these, 26.9 % (2,323) produced no signal and could thus not be scored, 1.8 % (158) showed more than 20 % missing values, and 20.4 % (1,761) were monomorphic. These markers were removed from the data set. In addition, 460 (5.3 %) SNPs were excluded because they showed a minor allele frequency (MAF) of <5 %. We found that the heterozygosity within each of the 172 lines ranged from 24.1 to 29.9 % (Figure S1A). Considering markers scored as heterozygous (A/C or A/G), we observed that many markers showing a heterozygous scoring were heterozygous in a high number of lines (Figure S1B, C). 246 markers even showed the heterozygous scoring in all 172 lines. As the lines included in this study have been selfed for many generations and can all be expected to be largely homozygous, very little heterozygosity should be present in these lines. We ambiguously set the number of SNPs, which were tolerated as being heterozygous to three. Consequently, SNPs which showed a heterozygous scoring in more than three lines were also removed from the data set (29.4 %, 2,533). After these checks a total of 1,395 high-quality SNPs were left that were used for further analyses.

Properties of SNP and SSR markers

The average number of alleles per locus was 2 for the SNPs and 6.9 for the SSRs illustrating the higher rate of SSR mutation. The spectra of allele frequencies were very different for the SNPs and the SSRs (Fig. 1a). Most SSR alleles were present at low frequency in the population. By contrast, the SNPs showed a large number of alleles with intermediate frequency. For 518 of the SNPs, a genetic map position was available (Cavanagh et al., in review) and these SNPs mapped to 303 unique positions in the wheat genome. Considering unique map positions, the A genome (143) and the B genome (142) were comparable, whereas the D genome (18) was strongly underrepresented (Fig. 1b). The number of SNPs with unique map position on the chromosomes of the D genome ranged from 1 to 4. The PIC values showed an approximate equal distribution, but with an increased number of markers having equal allele frequencies (PIC of 0.5). The average PIC for both, the 1,395 high-quality SNPs and for the 518 mapped SNPs was 0.33 (Fig. 1c). For the SSRs the PIC averaged 0.50.

Analysis of genetic relatedness and population structure

We investigated the genetic relatedness and population structure in the panel of 172 elite European winter wheat lines by cluster analysis and principal coordinate analysis (PCoA) based on the modified Rogers' distances between the individuals (Fig. 2). The cluster plots for SNPs and SSRs were different implying different kinship among the 172 wheat lines (Fig. 2a). PCoA was performed based on the 1,395 high-quality SNPs as well as with 91 SSR markers. The proportion of genotypic variance explained by the first two principal coordinates was higher for the SNPs with 5.1 and 4.3 %, respectively, as compared to 2.0 and 1.9 % for the SSRs. A higher proportion of explained genotypic variance was observed for each of the first ten principal coordinates for the SNPs (Fig. 2c). For both marker types no population structure was apparent for any of the first ten principal coordinates as revealed by their density distribution (Fig. 2c). The genetic similarity, assessed as simple matching coefficient, among the 172 wheat lines was higher for the SNPs than for the SSRs and averaged 0.43 and 0.25, respectively (Fig. 3a). The estimates for the coefficient of variation of genetic similarity were comparable between SNPs (0.10) and SSRs (0.12). We observed no significant correlation between the genetic similarity estimates based on SNPs and SSRs ($r = 0.008$; $P = 0.32$) (Fig. 3b) which was corroborated by the different cluster plots based on both marker types (Fig. 2a). When assessed on a genome-specific basis, we found that the correlation was moderate for the A and D genome, but low for the B genome (Table S1). We next tested the correlation

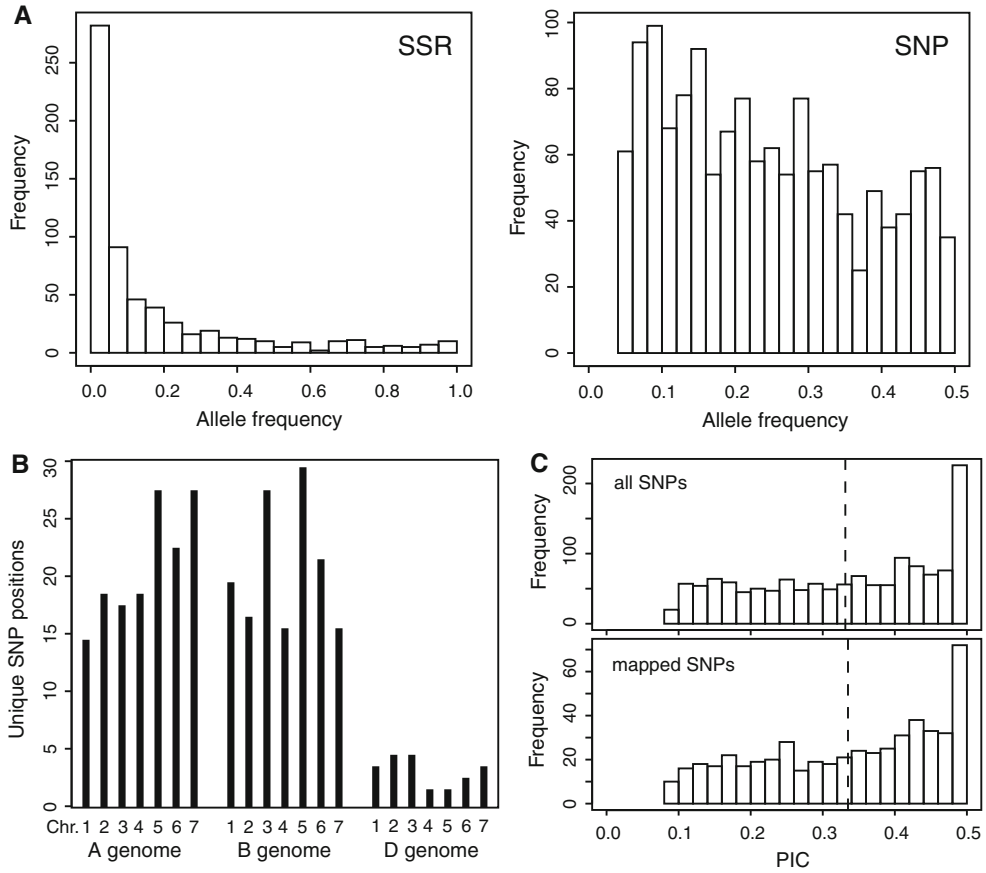

Fig. 1 **a** Allele frequency distributions for the 1,395 SNPs and the 91 SSRs. **b** Number of mapped SNPs with unique position (303) for each of the 21 chromosomes. **c** Polymorphic information content (PIC) of all SNPs (1,395) and of the SNPs with a map position (518)

of genetic similarity estimates based on SNPs and different numbers of randomly sampled SSR alleles as previous studies have shown that a much higher number of SNPs are required to achieve the same accuracy in genetic similarity as the multiallelic SSRs. We found, that the correlation was highest for 100 SSR alleles and then decreased when more SSR alleles were included in the analysis (Fig. 3c).

Evaluation of linkage disequilibrium

The analysis of linkage disequilibrium with the mapped SNP markers revealed that LD decayed with genetic map distance (Fig. 4a). The fitted regression showed that closely linked markers have LD values around 0.2. The 95th percentile of unlinked markers used as a threshold for LD due to linkage was $r^2 = 0.04$. The fitted regression intersected this threshold at approximately 8 cM. We next investigated the LD between linked markers for distinct genetic map distances and confirmed the decay with genetic map distance (Fig. 4b). The average LD between adjacent markers was 0.32. The analysis of the pattern of LD by a sliding window approach revealed that LD is variable between chromosomes but also along the chromosomes (Fig. 5).

Discussion

The rapid advancement in sequencing and genotyping technologies in the last years has enabled the development of SNP arrays even for polyploid crops such as wheat (Edwards et al. 2012). Here, we used the currently available 9k SNP array to genotype and analyze a panel of 172 diverse European elite winter wheat lines.

SNP genotyping

We found that 26.9 % of the SNPs on the array could not be scored which may be due to technical issues, but could also be attributed to certain loci not being present in the material under study. The SNP markers present on the array were discovered in a diverse set of 27 wheat lines originating from the US and Australia. The SNP array must therefore be expected to show a higher degree of ascertainment bias when used on European winter wheat germplasm. This is likely also responsible for the observed degree of monomorphic markers (20.4 %) in our diverse panel of elite lines. A major issue of SNP arrays for polyploid species is the genome-specificity. In wheat, the

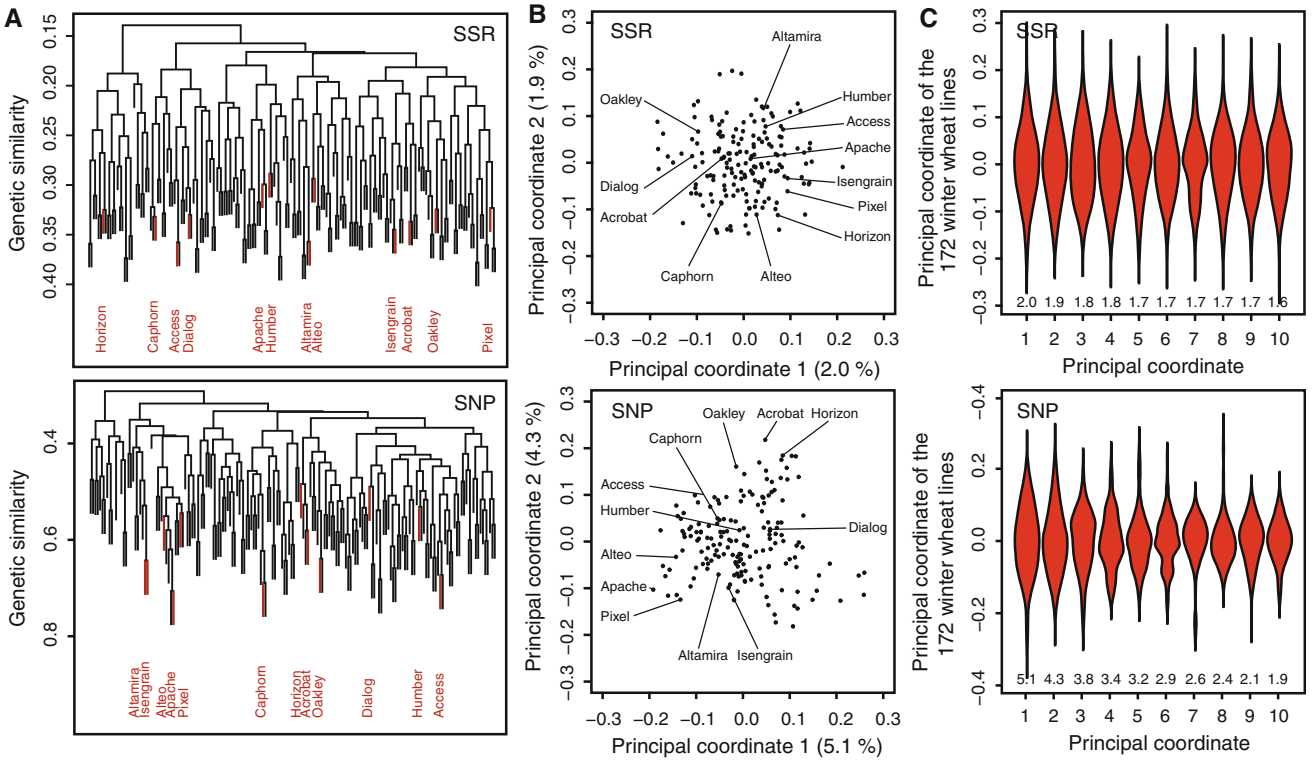

Fig. 2 Genetic relatedness and population structure assessed in the panel of 172 elite winter wheat lines with SSR and SNP markers. **a** Cluster plot for the 172 elite winter wheat genotypes. **b** Principal coordinate analysis based on the modified Rogers' distances estimates. **c** *Violin plot* showing the density distribution of the first ten principal coordinates. *Numbers* refer to the percentage of variance explained by the respective principal coordinate

SNPs on the 9k array were developed based on polymorphisms in transcribed sequences and the genotyping is done based on genome-specific primers (Cavanagh et al., in review). We observed a high degree of heterozygosity in the 172 lines with a mean of 28.3 % (Figure S1). Given that the plants have been selfed for many generations, they should be highly homozygous and one would expect that only a small fraction of loci remain heterozygous. In maize it has been shown that regions exhibiting a reduced recombination frequency, as for example close to the centromere, show a higher degree of heterozygosity (McMullen et al. 2009). This and other genetic forces, however, could only account for a rather small proportion of the observed heterozygosity. The high percentage of heterozygosity observed here must therefore be due to technical issues. These could be difficulties in reading and interpretation of the fluorescence signals or cross-hybridization of the primers to homologous or homeologous genes elsewhere in the genome. In the latter case there could be a polymorphism between the duplicated genes which will be scored as heterozygous even though the plant is homozygous for each of the two loci. Analysis of the genomic regions underlying the SNPs and additional work on the clustering algorithms may in part solve this problem. SNPs which in additional experiments prove not to be locus-specific will need to be removed from future versions of the wheat SNP array.

SNP genome properties

A total of 1,395 high-quality SNPs were left after the quality checks by which we removed SNPs that were unscorable, monomorphic, showed a high degree of heterozygosity, or a minor allele frequency (<5 %). For 518 of these SNPs a map position was available illustrating the need to genotype additional segregating populations to place more SNPs from the array on a genetic linkage map. The 518 mapped SNPs identified 303 unique positions which is in part likely due to different SNPs being derived from the same gene. Such SNPs cannot be resolved in genetic maps based on segregating populations but their order can be deduced from gene sequences. This is in contrast to SSRs which usually identify a unique genomic position. Larger population sizes and additional populations will probably help to separate some of the closely linked SNPs which in the current map locate to the same map position. Markers that have been mapped to the same map position are nevertheless useful for association

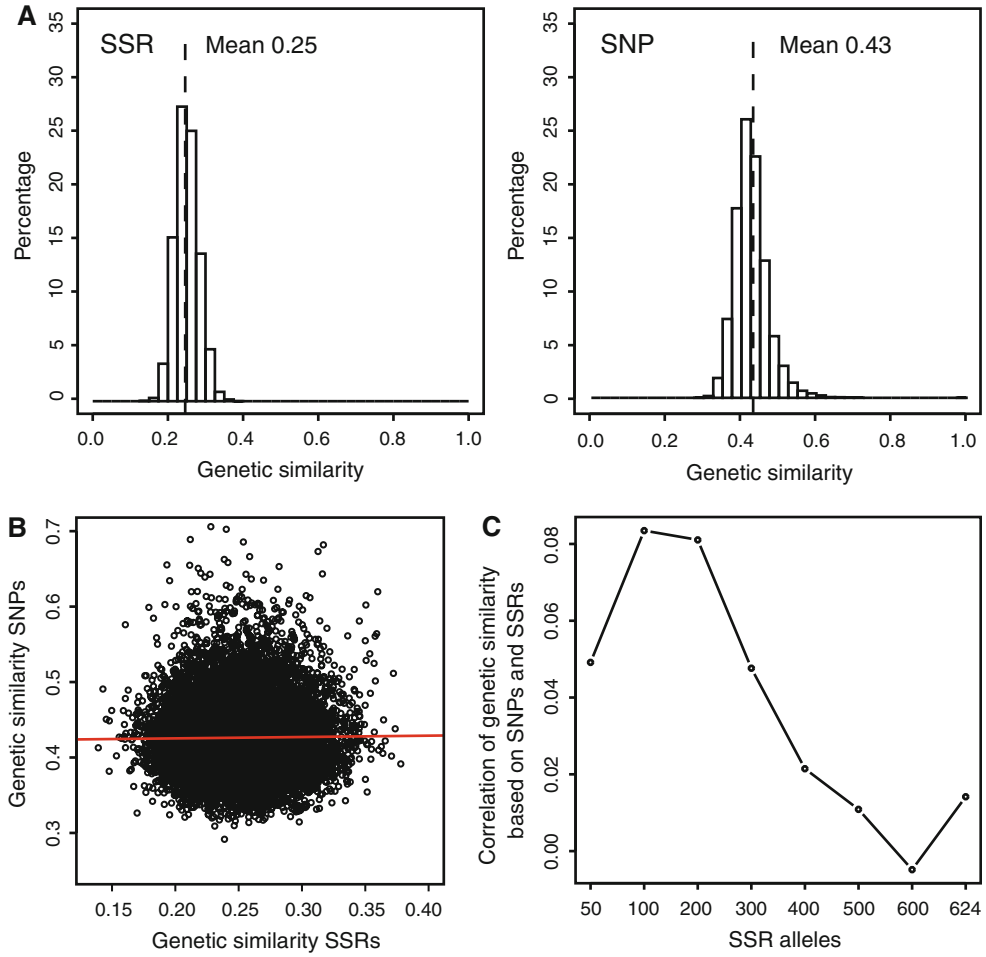

Fig. 3 Genetic similarity assessed in the panel of 172 elite winter wheat lines with SSR and SNP markers. **a** Histogram of the genetic similarity (simple matching coefficient) among the lines. **b** Correlation between genetic similarity based on SNPs and SSRs. **c** Correlation of genetic similarity based on SNPs and SSRs for different numbers of SSR alleles

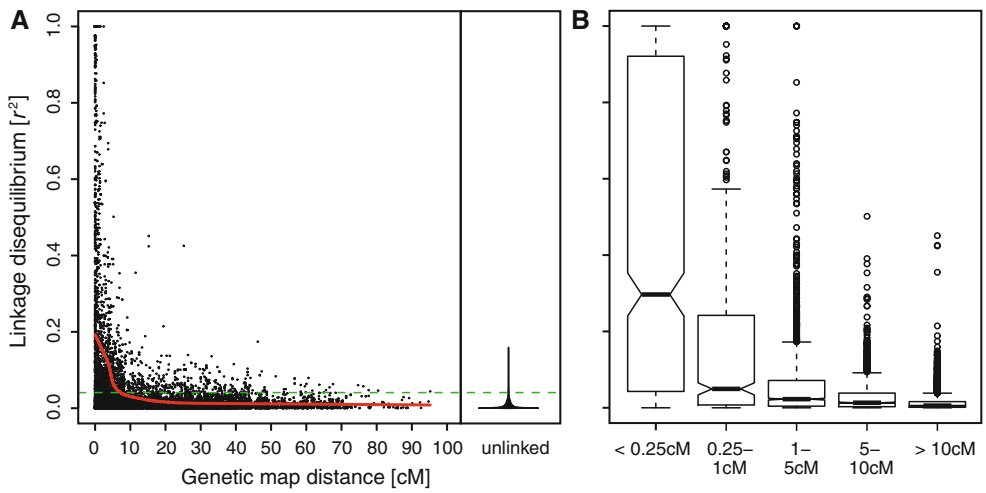

Fig. 4 **a** Linkage disequilibrium as a function of genetic map distance. *Curve* was fitted by robust locally fitted regression and shows that LD decays with genetic map distance. The *violin plot* shows the density distribution of r^2 values for unlinked loci and the *dashed line* indicates the derived threshold for LD due to linkage. **b** Distribution of LD between linked marker pairs for different genetic map distances between the two markers

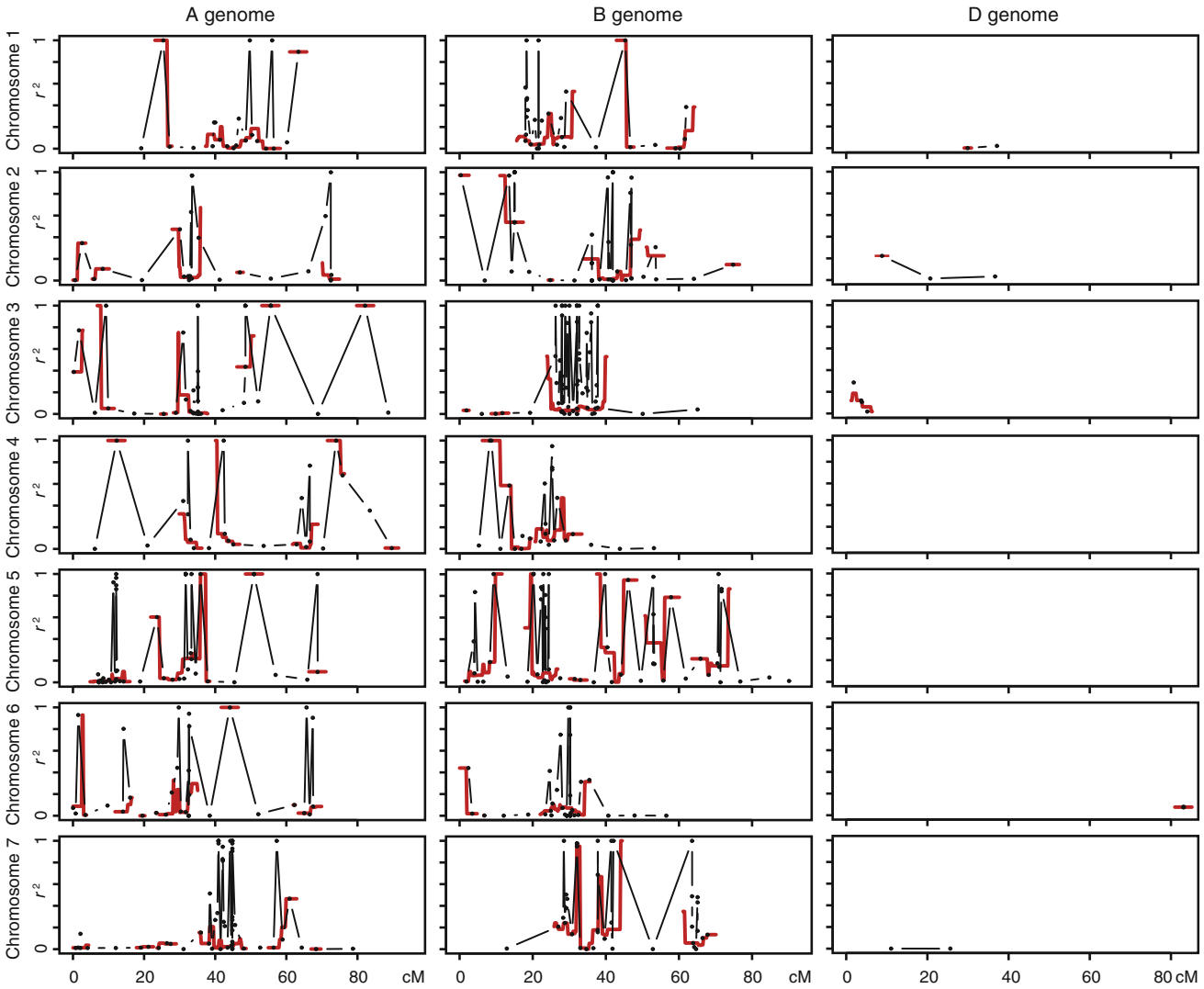

Fig. 5 Distribution of linkage disequilibrium along the chromosomes. The *black points* show the LD of each locus with its neighboring locus. *Red line* represents the median of r^2 values calculated within a 5 cM window at 500 positions along the chromosomes (color figure online)

mapping approaches as they can be derived from distinct but closely linked loci or may identify the SNP most closely to the causal polymorphism within a gene. We observed that only few of the mapped markers were derived from the D genome. This may be due to a bias on the array with regard to polymorphic SNPs from the different genomes and is consistent with previous studies which reported a lower degree of polymorphism for the D genome (Chao et al. 2009; Akhunov et al. 2010; Hao et al. 2011). The number of SNPs per chromosome ranged from 1 to 4 for the D genome which is certainly not enough to warrant genomic studies. The availability of a SNP array with a much higher number of SNPs (90k) and a targeted SNP discovery approach for the D genome will likely alleviate this problem.

Based on the selected high-quality SNPs, we observed an average PIC of 0.33 which is in accordance with the PIC based on 253 SNPs reported previously for a diverse set of wheat lines (0.34) and higher than the average PIC observed for a set of US cultivars (0.23) (Chao et al. 2009). The higher PIC in the European lines compared to the US lines may indicate a higher diversity in our set, but must certainly also be attributed to the larger number of lines included in our study (172) as compared to the 20 lines used by Chao et al. (2009). The observed approximate equal distribution of PIC values illustrates that there are regions in the wheat genome with low gene diversity for which consequently the majority of the lines carry identical alleles. For plant breeding it is thus difficult to create novel haplotypes and diversity in such regions. Selection gain in breeding programs, however, requires genetic variation and our results illustrate that SNP markers can be used to assess the extent of genetic diversity in elite germplasm. A detailed analysis of the genome-wide patterns of gene diversity (Akhunov et al. 2010; Alheit et al.

2012) must await higher-density SNP maps. The average PIC value obtained with the genome-wide distributed SSRs (0.50) was higher than the PIC for the SNPs. It must be noted, however, that for biallelic markers such as SNPs the PIC ranges from 0 (fixation of one allele) to 0.5 (equal allele frequencies), whereas for multiallelic markers like SSRs the PIC values can exceed 0.5 and approach 1. Therefore, PIC values obtained for SNPs and SSRs are not comparable.

Population structure and genetic relatedness assessed with SNP and SSR markers

Our analysis with both, SNPs and SSRs, showed that no population structure appears to be present in the panel of 172 elite wheat lines (Fig. 2). This is in agreement with previous studies of elite European winter wheat (Reif et al. 2011a) and can be explained by the breeding history of wheat. Wheat is improved by line breeding and there has always been a constant exchange of germplasm between breeding programs. We found that with regard to the principal coordinate plot the general pattern appeared similar between the two marker types, but a closer analysis revealed differences among them (summarized in Supplementary Table S2). Lines which with one marker type appear closely related can be widely separated in the plot with the other marker type. This can be attributed to the biallelic (SNPs) and multiallelic (SSRs) nature of the two marker types which consequently capture the existing diversity in a different way. We found that a higher proportion of variance is explained by the first principal coordinates for the SNPs as compared to the SSRs which is in accordance with findings in maize (Van Inghelandt et al. 2010).

The kinship estimates among the lines were higher for the SNPs than for the SSRs which may also be attributed to the multiallelic nature of the latter. SSR markers are the marker type of choice for diversity analyses because of their high degree of polymorphism even among related germplasm. The kinship estimates based on SNPs and SSRs showed no significant correlation, confirming the results of the PCoA. Recent analyses in maize also showed no significant correlation between distance measures based on SNPs and SSRs except for plants that were closely related (Jones et al. 2007; Hamblin et al. 2007). Since the lines included in our analysis represent a major part of the diversity in elite European winter wheat and can thus be expected to be only distantly related, our results are in agreement with these previous studies.

One could reason that the difference between SNPs and SSRs lies in the different spectra of allele frequencies of both marker types, since SSRs are dominated by rare alleles. However, Hamblin et al. (2007) found that the SNPs at intermediate frequency are best correlated with the distances based on SSRs. According to Laval et al. (2002) and Van Inghelandt et al. (2010), a much higher number of biallelic SNPs are required to achieve the same accuracy as a set of multiallelic SSRs. We did indeed observe a dependency of the correlation between genetic similarity measures based on both marker types on the number of SSR alleles included in the analysis (Fig. 3c). The correlation, however, was low even at its maximum. It thus appears that the two marker types with their different mutational properties capture genetic diversity in a different way. It must also be noted here, that the SSR markers have been selected over years to optimally capture genetic diversity, whereas the SNPs are unselected in that regard. In addition, both marker types do not cover the entire genome due to their rather low number. As genetic diversity varies among and along chromosomes (Akhunov et al. 2010; Hao et al. 2011; Alheit et al. 2012), the difference in kinship estimates between SNPs and SSRs may be reduced if more markers are used that represent similar genomic regions.

Consequences for genomics and plant breeding

These findings have consequences for applied plant breeding and genomics approaches. Ancestry among individuals is assessed and accounted for in genomic analyses to prevent spurious results. In addition, genetic relationships are often the basis for the choice of breeding material for crop improvement strategies as well as for the design of experimental crosses. If no pedigree information is available the breeder may select lines to be crossed based on molecular marker data. Taking SSRs as a standard for genetic diversity analyses, the SNPs represented on the array should only be used cautiously. It is, however, well known that most SSR variation is neutral, whereas SNPs, especially those derived from transcribed sequences, are much more likely to be associated with functional variation. Therefore, additional research is required to assess which marker type better captures the variation of interest. With the increased number of SNPs that will be available in the near future, a subset of them may be selected that is particularly suited to assess genetic diversity, or closely linked SNPs can be used to form haploblocks thereby creating a multiallelic system. The use of multiallelic haploblocks instead of the biallelic SNPs would also overcome the inherent problem of SNPs of a high identity-by-state probability and could increase the quality of association mapping approaches.

Extent and pattern of linkage disequilibrium

The analysis of linkage disequilibrium is of utmost importance for association mapping as this mapping

approach mainly detects associations between QTL and the trait indirectly by the LD between markers and the QTL. We used the 95th percentile of LD between unlinked markers as a threshold for LD due to linkage (Breseghello and Sorrells 2006). Based on this threshold LD decayed within approximately 5–10 cM (Fig. 4). This decay of LD with genetic map distance is comparable to recent studies in wheat (Chao et al. 2010; Hao et al. 2011; Reif et al. 2011b). Compared to results from maize (van Inghelandt et al. 2011; Tenaillon et al. 2001) the LD decay in wheat appears slow. Nowadays, for elite germplasm of both crops crosses are made by the breeder rendering the preferred mode of reproduction (selfing or outcrossing) of the species inconsequential. LD analyses, however, also capture historical LD which is much lower in outcrossing species like maize as compared to selfing species like wheat because selfing causes and maintains homozygosity and thus renders recombinations ineffective in reducing LD. The observed comparably slow decay of LD with genetic map distance has consequences for association mapping in wheat. On the one side, fewer markers are required to entirely cover the genome as compared to maize; on the other side, the achievable mapping resolution will be limited.

Genomic selection is emerging as an additional approach that holds great potential for plant breeding and generally also relies on a high number of molecular markers distributed throughout the genome (Heffner et al. 2009). Calus et al. (2008) observed in a simulation study that genomic selection is useful even when the mean r^2 value between adjacent markers is around 0.10. Hayes et al. (2009) recommended a slightly higher threshold of an r^2 value of 0.20 between adjacent markers. Consequently, the average r^2 value of 0.32 observed in our study suggests that the SNP marker density is at the lower level for being useful in genomic selection studies.

Our analysis of the pattern of LD revealed that LD is variable along chromosomes, consistent with results in wheat (Hao et al. 2011), sugar beet (Würschum et al. 2011), and triticale (Alheit et al. 2012). Consequently, the mapping resolution in association mapping will also be variable. Whereas the association of LD with genetic map distance could be interpreted such that a limited number of markers (i.e., <1,000) will already be sufficient for association mapping in wheat, the variability of LD suggests that as many markers as possible should be used to ensure a complete coverage of the genome even in chromosomal regions with rapid LD decay. Our analysis was hampered by the rather low number of mapped SNPs, especially for the D genome. Future experiments based on a much higher number of mapped SNPs will enable a more detailed analysis of the pattern of LD in elite wheat.

Conclusions

In this study, we employed high-throughput SNP genotyping which is nowadays available even for polyploid wheat to investigate the utility of SNP markers for genomic analyses in elite wheat breeding programs. We found that additional research is required if SNPs are to be used for analyses on population structure and genetic diversity. Provided that the number of SNPs will be strongly increased, they will certainly be a valuable tool for plant breeding and also hold great promise for genomic approaches like association mapping and genomic selection in wheat.

Acknowledgments This research was conducted within the BMBF-funded HYWHEAT project (Grant ID: FKZ0315945D). S.M. Langer was supported by the Deutsche Forschungsgemeinschaft under Grant number WU 658/1-1.

References

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113: 1409–1420

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet 119:507–517

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N et al (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11:702

Alheit KV, Maurer HP, Reif JC, Tucker MR, Hahn V et al (2012) Genome-wide evaluation of genetic diversity and linkage disequilibrium in winter and spring *triticale* (× *Triticosecale Wittmack*). BMC Genomics 13:235

Barrett BA, Kidwell KK (1998) AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci 38:1261–1271

Bonjean AP, Angus WJ (2001) The world wheat book: a history of wheat breeding. Lavoisier Publishing, Rue Lavoisier

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Calus MPL, Meuwissen THE, de Roos APW, Veerkamp RF (2008) Accuracy of genomic selection using different methods to define haplotypes. Genetics 178:553–561

Cavanagh C, Chao S, Wang S, Huang BA, Stephen S et al Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars (in review)

Chao S, Sharp PJ, Worland AJ, Warham EJ, Koebner RMD, Gale MD (1989) RFLP-based genetic maps of wheat homoeologous group 7 chromosomes. Theor Appl Genet 78:495–504

Chao S, Zhang W, Akhunov E, Sherman J, Ma Y et al (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. Mol Breed 23:23–33

Chao S, Dubcovsky J, Dvorak J, Luo M-C, Baenziger SP et al (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11:727

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N et al (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582

Dreisigacker S, Zhang P, Warburton ML, Skovmand B, Hoisington D, Melchinger AE (2005) Genetic diversity among and within CIMMYT wheat landrace accessions investigated with SSRs and implications for plant genetic resources management. Crop Sci 45:653–661

Eathington SR, Crosbiea TM, Edwards MD, Reiter RS, Bull JK (2007) Molecular markers in a commercial breeding program. Crop Sci 47(S3):54–163

Edwards D, Batley J, Snowdon RJ (2012) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet. doi:10.1007/s00122-012-1964-x

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Ann Rev Plant Biol 54:357–374

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325–338

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS One 2(12):e1367

Hao C, Wang L, Ge H, Dong Y, Zhang X (2011) Genetic diversity and linkage disequilibrium in Chinese bread wheat (*Triticum aestivum* L.) revealed by SSR markers. PLoS One 6(2):e17279

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Jones ES, Sullivan H, Bhattramakki D, Smith JSC (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet 115:361–371

Laval G, San Cristobal M, Chevalet C (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. Genet Sel Evol 34:481–507

Maurer HP, Melchinger AE, Frisch M (2008) Population genetic simulation and data analysis with Plabsoft. Euphytica 161: 133–139

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H et al (2009) Genetic properties of the maize nested association mapping population. Science 325:737–740

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2202

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Reif JC, Zhang P, Dreisigacker S, Warburton ML, Van Ginkel M, Hoisington D, Bohn M, Melchinger AE (2005) Wheat genetic diversity trends during domestication and breeding. Theor Appl Genet 110:859–864

Reif JC, Gowda M, Maurer HP, Longin CFH, Korzun V et al (2011a) Association mapping for quality traits in soft winter wheat. Theor Appl Genet 122:961–970

Reif JC, Maurer HP, Korzun V, Ebmeyer E, Miedaner T et al (2011b) Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. Theor Appl Genet 123:283–292

Röder MS, Korzun V, Gill BS, Ganal MW (1998) The physical mapping of microsatellite markers in wheat. Genome 41: 278–283

Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet 111:162–170

Sako K, Takeda K (2009) An application of high-throughput SNP genotyping for barley genome mapping and characterization of recombinant chromosome substitution lines. Theor Appl Genet 119:613–619

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF et al (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. mays L.). Proc Natl Acad Sci USA 98:9161–9166

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299

Van Inghelandt D, Reif JC, Dhillon BS, Flament P, Melchinger AE (2011) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor Appl Genet 123:11–20

Weir BS (1996) Genetic data analysis II, 2nd edn. Sinauer Associates, Sunderland

Wright S (1978) Evolution and genetics of populations, variability within and among natural populations, Vol 4. The University of Chicago Press, Chicago

Würschum T (2012) Mapping QTL for agronomic traits in breeding populations. Theor Appl Genet 125:201–210

Würschum T, Maurer HP, Kraft T, Janssen G, Nilsson C et al (2011) Genome-wide association mapping of agronomic traits in sugar beet. Theor Appl Genet 123:1121–1131

Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J et al (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breeding 25:441–451