Theor Appl Genet (2011) 123:555–569
DOI 10.1007/s00122-011-1607-7

ORIGINAL PAPER

High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.)

Daniele Trebbi · Marco Maccaferri · Peter de Heer · Anker Sørensen · Silvia Giuliani · Silvio Salvi · Maria Corinna Sanguineti · Andrea Massi · Edwin Andries Gerard van der Vossen · Roberto Tuberosa

Received: 8 January 2011 / Accepted: 26 April 2011 / Published online: 25 May 2011

Abstract We describe the application of complexity reduction of polymorphic sequences (CRoPS®) technology for the discovery of SNP markers in tetraploid durum wheat (*Triticum durum* Desf.). A next-generation sequencing experiment was carried out on reduced representation libraries obtained from four durum cultivars. SNP validation and minor allele frequency (MAF) estimate were carried out on a panel of 12 cultivars, and the feasibility of genotyping these SNPs in segregating populations was tested using the Illumina Golden Gate (GG) technology. A total of 2,659 SNPs were identified on 1,206 consensus sequences. Among the 768 SNPs that were chosen irrespective of their genomic repetitiveness level and assayed on the Illumina BeadExpress genotyping system, 275 (35.8%) SNPs matched the expected genotypes observed in the SNP discovery phase. MAF data indicated that the overall SNP informativeness was high: a total of 196 (71.3%) SNPs had MAF >0.2, of which 76 (27.6%) showed MAF >0.4. Of these SNPs, 157 were mapped in one of two mapping populations (Meridiano × Claudio and Colosseo × Lloyd) and integrated into a common genetic map. Despite the relatively low genotyping efficiency of the GG assay, the validated CRoPS-derived SNPs showed valuable features for genomics and breeding applications such as a uniform distribution across the wheat genome, a prevailing single-locus codominant nature and a high polymorphism. Here, we report a new set of 275 highly robust genome-wide *Triticum* SNPs that are readily available for breeding purposes.

Communicated by P. Langridge.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1607-7) contains supplementary material, which is available to authorized users.

D. Trebbi (✉) · P. de Heer · A. Sørensen · E. A. G. van der Vossen
Keygene N.V., Applied Research, Agro Business Park 90, 6708, PW, Wageningen, The Netherlands
e-mail: trebbida@msu.edu

M. Maccaferri · S. Giuliani · S. Salvi · M. C. Sanguineti · R. Tuberosa
Dipartimento di Scienze e Tecnologie Agroambientali (DiSTA), Università di Bologna, Viale Fanin 44, 40127 Bologna, Italy

A. Massi
Società Produttori Sementi Bologna Spa, Via Macero 1, 40050 Argelato, BO, Italy

Present Address:
D. Trebbi
Dipartimento di Biotecnologie Agrarie, Università degli Studi di Padova, Viale dell' Università 16, 35020 Legnaro, PD, Italy

Introduction

Single-nucleotide polymorphisms (SNPs) are the most common polymorphism among individuals of any species (Brookes 1999; Deschamps and Campbell 2010). In crops, the availability of SNP genotyping platforms would facilitate the genetic dissection of traits of economic importance and the application of marker-assisted and genomic selection (Rafalski 2002; Mackay and Powell 2007; Collard and Mackill 2008; Jannink and Lorenz 2010). Resequencing projects have led to the discovery of thousands to millions of SNPs in *Arabidopsis* (Jander et al. 2002; Zhang and Borevitz 2009), rice (Feltus et al. 2004; McNally et al. 2009; Yamamoto et al. 2010), maize (Ching et al. 2002; Zhao et al. 2006; Liu et al. 2010; Yan et al. 2010) and barley (Close et al. 2009; Waugh et al. 2009; Hayden et al. 2009). Since SNPs can be converted into genetic markers

amenable to high-throughput assays (Gut 2001; Deschamps and Campbell 2010), SNP discovery and the development of SNP-genotyping platforms are receiving increasing attention (Edwards et al. 2008; Ganal et al. 2009; Varshney et al. 2009).

Notwithstanding the remarkable results reported for the above-mentioned species, SNP discovery and detection in wheat have both progressed much more slowly (Somers et al. 2003; Ganal and Röder 2007) due to: (1) lack of a reference genome sequence, (2) availability of gene sequences from only a few genotypes, (3) low polymorphism levels due to a reduced nucleotide diversity of the wheat genome, and (4) the highly repetitive and duplicated nature of the wheat allopolyploid genome (Koebner and Summers 2002; Edwards and Batley 2010). The severe bottlenecks that have characterized the domestication of cultivated wheat species have been suggested as the main cause of the low frequency of marker-exploitable SNP diversity in wheat (Ravel et al. 2006; Ganal et al. 2009; Haudry et al. 2007; Quraishi et al. 2010).

Durum wheat (*Triticum turgidum* spp.) is an allotetraploid crop, derived from the cross between *Triticum urartu* (AA genome, $2n = 14$) and an ancient relative of *Aegilops speltoides* (donor of the BB genome), and is characterized by a high number of paralogous genes (Dubcovsky and Dvorak 2007). In polyploid species as wheat, one of the main problems is that most of the identified SNPs represent sequence variants between homoeologous gene sequences rather than allelic variants between homologous genomes (inter-varietal SNPs). The presence of paralogous and multi-copy sequences adds further complexity, and hence difficulty, toward the correct scoring of SNPs at single locus (Akhunov et al. 2009; Sandve et al. 2010). Collectively, the combination of low nucleotide diversity, polyploidy and repetitiveness has severely limited advances in SNP discovery and genotyping platform development in wheat (Barker and Edwards 2009; Ganal et al. 2009).

Currently, durum wheat marker profiling relies on low- to medium-throughput marker platforms such as simple sequence repeat (SSR; Korzun et al. 1999; Maccaferri et al. 2008, 2011), amplified fragment length polymorphism (AFLP®; Lotti et al. 2000; Song et al. 2005; Akbari et al. 2006), or diversity array technology (DArT; Mantovani et al. 2008), thus hindering the deployment of genomics applications in breeding programs (Akhunov et al. 2009). The advent of next-generation sequencing (NGS) technologies has allowed for the rapid resequencing of genomes when compared with traditional Sanger sequencing technology (Shendure and Li 2008; Varshney et al. 2010). However, efficient SNP discovery by whole-genome resequencing can only be achieved for low-to-moderate complex genomes like *Arabidopsis* or rice (Ossowski et al. 2008; Huang et al. 2009; Yamamoto et al. 2010), while the majority of agronomically important crop genomes present a considerably higher complexity (Adams and Wendel 2005; Bennetzen et al. 2005). In the latter case, a reduction of genomic complexity prior to NGS-based SNP discovery and genotyping is often beneficial and in some cases required. Complexity reduction can be achieved by generating reduced representation libraries (RRLs). The construction of RRLs generates a subpopulation of the same allelic genomic fragments in several samples, thus reducing the genomic complexity of several orders of magnitude. RRLs were originally developed for SNP discovery with Sanger sequencing (Altshuler et al. 2000) and later adapted to NGS technologies (van Tassel et al. 2008; Hyten et al. 2010). Methods used to achieve genomic reduction are based on the use of restriction enzymes (Altshuler et al. 2000), high-C_0t selection (Yuan et al. 2003), methylation filtering (Palmer et al. 2003), microarrays (Albert et al. 2007; Okou et al. 2007) and cDNAs (Barbazuk et al. 2007). However, these methods have several disadvantages, such as they only target repetitive sequences and are ineffective in eliminating low-copy paralogous and homoeologous sequences (C_0t and methylation filtering), limited to exon sequences, laborious and low-throughput (cDNA), or are only applicable to pre-selected sequences (microarrays). Recently, AFLP has efficiently been combined with NGS in the complexity reduction of polymorphic sequences (CRoPS®) method for RRL-based SNP discovery (van Orsouw et al. 2007). CRoPS has the advantage that the extent of genome complexity reduction can be modulated using different enzyme combinations and a varying number of selective nucleotides during the AFLP protocol, allowing for a wide range of polyploid species to be analyzed for high-throughput, genome-wide identification of polymorphic SNPs. The successful application of CRoPS has recently been reported for SNP discovery in maize (Mammadov et al. 2010).

For SNP genotyping, different chemistries have been developed, including dual-labeled hydrolysis probes (TaqMan® probes; Livak et al. 1995; Salvi et al. 2001), fluorescent polarization detection (FP-TDI; Chen et al. 1999), Invader (Olivier 2005), iPLEX (Wright et al. 2008), competitive allele-specific PCR (KASPar, Nijman et al. 2008), SNaPshot (Applied Biosystems; Sanchez et al. 2003), Golden Gate (GG; Fan et al. 2003) and Infinium (Gunderson 2009; Illumina, San Diego, USA). While most of these methods can only reach low- to moderate-throughput (<40-plex), the GG and Infinium assays allow for multiplexing up to 1,536 and >200,000 SNPs, respectively. The GG assay has proved its utility in genotyping a large number of SNPs in several species with repetitive and/or polyploid genomes like soybean (Hyten et al. 2008),

maize (Jones et al. 2009) and bread wheat (Akhunov et al. 2009; Chao et al. 2010).

The purpose of this study was to investigate the suitability and efficiency of applying the CRoPS technology for developing a SNP platform based on the GG assay in durum wheat. Information is also provided on the stringent criteria to be adopted for the selection of highly polymorphic, single locus-based SNP assays. Importantly, we describe a new set of 275 highly robust *Triticum* SNPs that are readily available for breeding purposes.

Materials and methods

Plant material

SNP discovery was carried out on four elite durum cultivars (Claudio, Colosseo, Neodur and Rascon_37/2*Tarro_2) representing the genetic pools of cultivated *T. durum* bred in the Mediterranean and North America regions (Maccaferri et al. 2005, 2006). To estimate the transferability and polymorphism content of the SNPs identified with the CRoPS technology, eight additional cultivars were profiled with a selected subset of 768 SNPs. Cultivar pedigrees, origin and year of release are summarized in Table 1.

SNP mapping was carried out with two recombinant inbred line (RIL) populations derived from the crosses Colosseo × Lloyd (176 RILs) and Meridiano × Claudio (181 RILs); for these two populations, SSR- and DArT-based maps were previously reported (Maccaferri et al. 2008; Mantovani et al. 2008). For each accession, genomic DNA was extracted from leaves of 20 plantlets following the method described in Maccaferri et al. (2005).

SNP discovery with CRoPS

SNP discovery was performed following the CRoPS protocol (van Orsouw et al. 2007). CRoPS is based on the detection of polymorphisms between sequences of different samples by combining the power of reproducible genome complexity reduction of AFLP (Vos et al. 1995) with the sequencing-by-synthesis (pyrosequencing) technology. Briefly, 600 ng DNA per sample was reduced in complexity by digestion with *Pst*I and *Taq*I restriction enzymes (New England BioLabs, Ipswich, MA, USA). The *Pst*I + A/*Taq*I + CT preamplification reaction was used for the preparation of tagged libraries that were sequenced with the GS FLX 454 Roche sequencer (Roche, Basel, Switzerland). Quality control to quantify chloroplast and mitochondrial DNA was performed by cloning and Sanger sequencing of a fraction of the CRoPS libraries. Following a titration run performed to estimate the optimal DNA amount for the emulsion PCR, sequences were clustered to estimate the level of repetitiveness. CRoPS analysis was then performed as described by van Orsouw et al. (2007).

Nuclear DNA reads obtained were clustered and targeted for SNP mining using Keygene's proprietary software. To increase the success rate of the Golden Gate (GG) genotyping assay (Illumina, San Diego, CA, USA), three additional criteria were used for SNP selection: (1) absence of additional polymorphisms (SNPs or short INDELs) in both of the 12-bp-long sequences flanking the putative SNP, (2) homozygosity of the selected SNPs with at least two reads/sample in at least two samples and with allele frequency >0.1, (3) SNPs in homopolymer sequences were discarded. Sequence information of selected SNPs was analyzed with Illumina's proprietary software assay design tool (ADT) for the assessment of their designability rank

Table 1 Pedigree, origin and year of release of the durum wheat cultivars utilized for CRoPS-based SNP discovery (Claudio, Colosseo, Neodur and Rascon_37/2*Tarro_2), genotyping and allele frequency evaluation

Cultivar	Pedigree	Origin	Year of release
Claudio	CIMMYTselection/Durango// IS193b/Grazia	Italy	1999
Colosseo	Mexa mutant/Creso	Italy	1995
Neodur	184-7/Valdur//Edmore	France	1987
Rascon_37/2*Tarro_2	Rascon_37/2*Tarro_2	Mexico	2000
Ardente	Israel durum 303/preliminary77//664	France	1984
Kofa	Dicoccum Alpha population-85 S-1	USA	1998
Svevo	Rok/fg//stil/3/dur1/4/sapi/teal/hui///Zenit	Italy	1996
Levante	G80/Piceno//Ionio	Italy	2002
Lloyd	Cando/Edmore	USA	1983
Meridiano	Simeto/WB881//Duilio/F21	USA	1999
Mohawk	883-22 Alpha population-85 CHA	USA	1998
Simeto	Capeiti 8/Valnova	Italy	1988

scores (http://www.illumina.com). The ADT score indicates the overall probability of genotyping success of a specific SNP based on the length, presence of degenerate nucleotides and repetitiveness on the 50–60 bp flanking sequences. ADT scores range from 0 (worst) to 1 (best). In our analysis, only SNPs with ADT score >0.5 were selected for the design of two 384-plex oligo pool all (OPA) genotyping assays (Online Resource 1). One of the aims of the project was to understand the influence of sequence repetitiveness on the success rate of the GG assay in mapping single-locus SNPs as Mendelian markers in polyploid species. Therefore, only after the design of the two OPA assays, the corresponding 768 SNP-harboring sequences were analyzed for their degree of genomic repetitiveness by comparison in the Triticeae Repeat Sequence databases (TREP; http://wheat.pw.usdagov/ITMI/Repeat).

SNP genotyping with Golden Gate assay

After comparing the genotyping performance of one 384-plex GG assay starting from either genomic DNA or preamplified reaction, the same level of genomic complexity reduction (*Pst*I + A/*Taq*I + CT) achieved during the CRoPS protocol for SNP discovery was also utilized during SNP genotyping, as described previously (van Eijk and van der Poel 2006). The GG assay was performed on the Illumina's bead array (Oliphant et al. 2002), following the methodology described in Fan et al. (2003) and Hyten et al. (2008). Briefly, GG assay utilizes two allele-specific oligos (ASOs) with different nucleotides at the 3′ ends to discriminate between the two SNP alleles. A third locus-specific oligo (LSO) anneals downstream of the ASOs and includes an address sequence (IllumiCode) specific for each SNP. After annealing of the ASOs and LSO to the DNA, polymerase extends the ASOs and ligates the extension to the LSO, simultaneously generating PCR template for the 384 SNPs. PCR is then performed with a set of three primers common to all ASOs and LSOs. The two ASO primers are Cy3- and Cy5-labeled to discriminate between alleles, while a third unlabeled primer is common to all LSOs. The amplicons are then hybridized to beads carrying complementary sequences to the IllumiCodes and the Cy3/Cy5 ratios are calculated and utilized to estimate SNP genotypes.

SNP genotyping was performed on (1) all 12 cultivars to estimate the minor allele frequency (MAF) and (2) the Colosseo × Lloyd and Meridiano × Claudio RIL populations to estimate the frequency of mappable CRoPS-derived markers. SNP genotype calls were also analyzed with the GenomeStudio software (Illumina, San Diego, CA, USA) to estimate the quality of the calls with the Gen_Train_Score (GTS). The GTS indicates how well separated and tight the clusters are of the Cy3/Cy5 ratios of the different genotypes in the GG assay. GTS ranges from 0 (worst cluster separation) to 1 (best cluster separation) in homozygous populations. High-quality SNP markers were designated KBO that stands for "Keygene and University of Bologna".

SNP mapping and estimate of polymorphism content

SNPs segregating in the two RIL populations were integrated into the SSR- and DArT-based linkage maps reported in Maccaferri et al. (2008) and Mantovani et al. (2008). The Colosseo × Lloyd map (2,022 cM) consisted of 162 SSR and 392 DArT markers, while the Meridiano × Claudio map (2,487 cM) included 212 SSR and 716 DArT markers.

SNP data were first imported and inspected in JoinMap 4.0 (Van Ooijen 2006) using the Chi-square goodness-of-fit test to identify the markers showing excessive discrepancies between the observed and the expected (1:1) segregation ratio. SNPs with Chi-square *P* values ≤0.0001 were excluded from the analysis. SNPs showing a Mendelian segregation were integrated into the two SSR- and DArT-based linkage maps and, subsequently, a single combined map was obtained. For each of the two maps, marker grouping was performed using the independence $\log_{10}$ of the likelihood of the odds ratio (LOD) method (incremental LOD thresholds from 2 to 10 with LOD 1.0 steps) and a robust initial linkage group scheme was obtained by selecting the grouping nodes with a stable (non-variant) number of markers in the LOD range between 6 and 10. The linkage groups were used to calculate the corresponding maps based on the maximum likelihood (ML) mapping algorithm. The mapping process was based on the repeated rounds of (1) simulated annealing Monte Carlo map order optimization search, followed by (2) Gibbs sampling (Monte Carlo Expectation Maximization algorithm) that is used to obtain the multipoint recombination frequency estimates. For map building, the number of map optimization rounds was set equal to five. The linkage group maps were gradually constructed by considering spatial samples of loci using the five default recombination frequency spatial sampling thresholds (0.10, 0.05, 0.03, 0.02 and 0.01). In the simulated annealing marker-ordering phase, a chain of 5,000 trials and error steps with constant acceptance probability was used and the system-stop was set after reaching 1,000 chains without further improvement; other parameters were set as default. In the Gibbs sampling recombination frequency estimation phase, the length of the burn-in chain was set equal to 5,000 iterations and the chain length per Monte Carlo Expectation Maximization cycle was set to 1,000 iterations; other parameters were as default.

For each linkage group, the marker order was checked using the plausible position calculation option (simulated annealing method) set to 1,000 random replacement samples of markers. Based on such results, markers causing unstable marker orders were discarded. Potential errors in the final graphical genotypes such as the presence of marker triplets showing double crossovers within 10 cM were substituted with the missing data. However, double crossovers within 10 cM supported by two or more adjacent markers were accepted. After discarding unstable markers, the mapping procedure was repeated. The assignment of linkage groups to wheat chromosomes was carried out by checking for the presence of SSR markers with known map positions. Linkage groups belonging to the same chromosome were merged after the calculation of the genetic distance between the distal markers, while centromere position was inferred from centromeric SSRs (Somers et al. 2003).

The two genetic maps with the integrated SNP markers were then merged in a single consensus map using Carthagene (De Givry et al. 2005). Merging was performed with the *dsmergen* command so that for each marker pair a single recombination rate was estimated based on all available meioses. A framework-mapping method (*buildfw* command) was applied. With the *buildfw* command (*buildfw 2 2 {} 0*), the map procedure started from all possible pairs and triplets of markers, using as criterion the differences in log-likelihood; then, each new marker was inserted in all possible positions in an incremental insertion procedure. Each new marker was kept in the framework only if the difference in log-likelihood was >2 and all the alternative orders with log-likelihood difference >2 were kept for the next analysis steps. Non-framework markers were incorporated in the framework map by building a complete map using the framework map as a fixed order (*buildfw 0 0 {framework map marker order} 0*) and the map orders were improved using the techniques that use flipping of subsections of the map (*greedy* command: *greedy 3 1 1 15 0*) as well as permutations with a sliding window of four markers (*flips* command: *flips 4 1 1*).

The distribution of markers between chromosomes with respect to the distance from the centromeres was estimated. The influence of sequence repetitiveness and GTS on the proportion of mappable SNPs was also estimated.

Results

SNP discovery with CRoPS

A total of 574,665 high-quality reads were obtained from the GS FLX sequencing run, covering more than 88.9 Mbp of genomic sequences (average read length 155 bp). The reads were distributed quite evenly between the four lines of the discovery panel, ranging from 18.4% in Rascon_37/2*Tarro_2 to 29.3% in Neodur. A total of 466,202 reads clustered in 31,665 consensus sequences (average 14.7 reads/consensus), while 100,474 sequences remained as singletons and were excluded from the SNP-mining analysis. Only 1.4% (7,989) of the reads matched chloroplast and mitochondrial DNA. From the SNP-mining pipeline, a total of 2,659 putatively homozygous SNPs on 1,206 consensus sequences were discovered among the four cultivars. The 768 SNPs with the highest ADT scores were selected for the GG assays (Online Resource 1).

In order to estimate the influence of the genomic sequence repetitiveness on the success rate of the genotyping assay and on SNP mappability, the selected 768 SNP-harboring sequences were queried against the TREP database. A total of 247 sequences (32.2%) were classified as repetitive and mainly represented retrotransposons and DNA transposons (Online Resource 1). The designability score was only marginally influenced by the presence of repetitive sequences, whose rate decreased from 40.2% repetitiveness for those SNPs in the 0.5–0.6 ADT score class, to 28.8% repetitiveness for those SNPs in the >0.9 ADT score class (data not shown).

SNP genotyping with Golden Gate assay

Concordance between the SNP genotypes determined with the GG assay and those determined with the CRoPS discovery pipeline was used to assess the accuracy of genotyping. We defined as "correct" and "incorrect" genotypes those SNPs whose alleles were genotyped exactly and differentely from the alleles discovered by the CRoPS pipeline, respectively. The majority of the incorrectly genotyped SNPs showed monomorphic genotypes, though the capacity to discriminate between alleles was retained by the GG assay. We also defined as "undetermined" genotypes those SNPs whose alleles were not identifiable because of insufficient allele clusters separation or because of GG assay failure. We initially tested the performances of the genotyping assay starting from either genomic DNA or preamplified reaction on one 384-plex assay. The proportion of correctly genotyped SNPs increased from 4.2% (16/384) when using genomic DNA to 44.3% (170/384) when using preamplified reaction (*Pst*I + A/*Taq*I + CT) as a starting material for the GG assay (Online Resource 1). The vast majority of undetermined genotypes were observed when starting from genomic DNA (73.7%). Based on these findings, the genotyping analysis was performed starting from the 1:10 diluted preamplification reactions, as in the standard AFLP protocol used in CRoPS.

Samples of the 12 cultivars and of the two Colosseo × Lloyd and Claudio × Meridiano RIL populations

Table 2 Results of the two OPA Golden Gate assays as function of the level of genomic repetitiveness

Genotype	SNPs (*n*)	SNPs in single sequences (*n*)	SNPs in repetitive sequences (*n*)	Repetitive sequences (%)
Correct	275 (35.8%)	245 (47.0%)	30	10.9
Incorrect	172 (22.4%)	85 (16.3%)	87	50.6
Undetermined	321 (41.8%)	191 (36.7%)	130	40.5
Total	768	521	247	32.2

were genotyped with the two 384-plex OPA assays using the GG assay on the BeadExpress platform. Out of the 768 SNPs tested, 275 (35.8%) SNP genotypes matched the expected genotype observed during SNP discovery (correctly genotyped SNPs), while 172 (22.4%) SNPs were incorrectly genotyped. Conversely, for 321 SNPs (41.8%) the GG assay was unable to discriminate between alleles (undetermined genotypes; Table 2).

Despite the reduction of genomic coplexity before GG assay, sequence repetitiveness influenced the accuracy of genotyping, as indicated by the fact that 50.6% of the 172 incorrectly genotyped SNPs were in repetitive sequences, while only 10.9% of the 275 correctly genotyped SNPs were in repetitive sequences (Table 2). In other words, among the 521 SNPs in the non-repetitive sequences, the proportion of correctly genotyped SNPs increases to 47.0%, while the proportion of incorrectly genotyped SNPs drops to 16.3% (Table 2).

Genotypes of the 275 correctly genotyped SNPs of all 12 parental lines were used to estimate the minor allele frequency (MAF; Online Resource 1). A total of 196 (71.3%) SNPs had MAF >0.2, of which 76 (27.6%) had MAF >0.4 (Fig. 1).

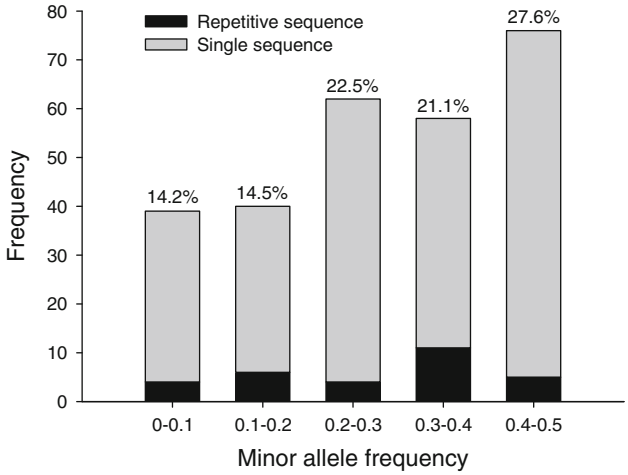

Fig. 1 Minor allele frequency (MAF) of the 275 correctly genotyped SNPs estimated from the genotypes of the 12 cultivars. *Percentages* indicate the proportion of SNPs for each MAF class

Distribution and informativeness of CRoPS-derived SNP markers

As a consequence of the relatively low frequency of rare alleles in the SNPs identified, 204 (74.2%) of the correctly genotyped SNPs showed segregation in at least one RIL population (Table 3). More specifically, 95 and 146 SNPs segregated in the Meridiano × Claudio and Colosseo × Lloyd populations, respectively. Of these segregating markers, 74 (77.9%) and 112 (76.7%) were successfully mapped in the Meridiano × Claudio and Colosseo × Lloyd populations, respectively (Table 3). When only SNPs in the non-repetitive sequences are considered, the proportion of mapped markers increases to 80.7% (71/88) and 79.2% (103/130) in the Meridiano × Claudio and Colosseo × Lloyd populations, respectively (Table 3).

The segregating SNPs were subsequently integrated in the existing maps of Meridiano × Claudio and Colosseo × Lloyd populations, respectively. This resulted in a 2.8 and 10.7% increase of map length for the Meridiano × Claudio (2,556 cM; 25 linkage groups) and Colosseo × Lloyd (2,238 cM; 20 linkage groups) populations, respectively. The merged map spanned 2,967 cM (22 linkage groups) and included 1,479 markers: 312 SSRs, 1,010 DArT markers and 157 SNPs from the CRoPS libraries (Table 4; Fig. 2). A total of 610 (41.2%) markers were mapped as framework markers, while only 156 (10.5%) markers were common between the two populations; among these common markers, 19 were SNPs (Table 4).

Since the SSRs used for mapping were pre-selected for polymorphism content, map position and PCR quality profile (Maccaferri et al. 2005, 2008), the map distribution of the newly mapped SNPs was compared to that of DArT markers only. SNP and DArT markers were differently distributed among chromosomes (Kolmogorov–Smirnov test, $P \leq 0.0001$, Table 4), suggesting a rather limited overlap between the two *Pst*I/*Taq*I genomic DNA representation libraries used for DArT- and CRoPS-based marker discovery. The distribution of SNPs was quite uniform among chromosomes, except for chromosomes 3B and 5B (less abundant), and chromosomes 1B and 2B (more abundant). Differently from SNPs, DArT markers were relatively abundant on chromosomes 3B and 6B and

Table 3 Frequency of segregating and mapped SNPs among the 275 correctly genotyped SNPs

RIL	Segregating SNPs[a] (*n*)	Segregating SNPs (%)	Repetitiveness in segregating SNPs (%)	Mapped SNPs[a] (*n*)	Mapped SNPs[a] (%)
Meridiano × Claudio	95 (88)	34.5	7.4	74 (71)	77.9 (80.7)
Colosseo × Lloyd	146 (130)	53.1	11.0	112 (103)	76.7 (79.2)
Total	204 (182)	74.2	10.8	157 (148)	77.0 (81.3)

[a] Values of SNP derived from single-copy sequence are reported in brackets

Table 4 Distribution of markers in the merged durum wheat map

Chromosomes	Total (*n*)	Framework (*n*)	Common (*n*)	SSR (*n*)	DArT (*n*)	SNP (*n*)	Length (cM)
1A	95	34	4	17	71	7	233
1B	124	42	10	26	79	19	226
2A	94	28	10	21	58	15	139
2B	138	66	23	26	94	18	266
3A	82	44	9	24	51	7	237
3B	157	54	12	26	125	6	254
4A	108	48	17	14	82	12	140
4B	65	27	11	18	36	11	135
5A	48	34	8	25	13	10	198
5B	58	27	4	21	32	5	203
6A	115	43	7	18	83	14	210
6B	145	51	12	27	109	9	192
7A	108	51	11	20	79	9	259
7B	142	61	18	29	98	15	275
Total	1,479	610	156	312	1,010	157	2,967

Markers have been categorized based on their mapping properties and marker classes. Framework refers to those markers mapped with high confidence, while common refers to those markers mapped in both recombinant inbred populations

The marker classes are as follows: *SSR* simple sequence repeat, *DArT* diversity array technology, *SNP* single nucleotide polymorphism

strongly underrepresented on chromosome 5A, a feature that seems common to wheat DArT-based maps (Akbari et al. 2006; Mantovani et al. 2008; Peleg et al. 2008). Both SNP and DArT markers were abundant in chromosome 7B.

To evaluate for the occurrence of overrepresentation of the distal hypomethylated chromosome regions, which has been reported for all the markers obtained from *Pst*I-based representation libraries (Vuylsteke et al. 1999; Wenzl et al. 2006), the distribution of DArT and SNP markers was investigated based on their position relative to the centromeres of the linkage groups (Fig. 3). Based on the Kolmogorov–Smirnov test, DArT and SNP marker distributions differed significantly ($P \leq 0.0001$). As expected, the distribution of the DArT markers was strongly skewed toward the distal hypomethylated region (ca. 30% of the genetic map) of the chromosome arms (cumulative fraction of the total mapped DArT markers is 54.7%), with only 12.5% of the DArT markers located in the proximal centromeric portion (ca. 30%). Conversely, the distribution of the SNPs was more uniform along chromosomes, with only 25.8% of the SNPs being assigned to the hypomethylated distal region, while 27.0% mapped in the proximal centromeric region (Fig. 3). Furthermore, SNPs had a significantly higher proportion of common markers between the two genetic maps when compared to DArT markers (18.5 and 8.1%, respectively; $P \leq 0.05$).

Analysis of the 204 SNPs segregating in the two RIL populations also showed that the GTS score (i.e., the quality index of the genotyping data computed by the Illumina software) can be used as a good predictor of the proportion of mapped markers. With GTS <0.4, only 20% of markers were successfully mapped, while GTS >0.6 ensured that more than 80% of markers were successfully mapped (Fig. 4). These differences in the success of mapping as the function of the GTS can be explained by sequence repetitiveness, which decreased from more than 35% to less than 10% in SNPs with GTS lower and higher than 0.6, respectively (data not shown).

In order to estimate the uniqueness of the SNPs discovered with the CRoPS technology, SNPs were compared with the University of Bristol's Wheat SNP database (http://www.cerealsdb.uk.net). Interestingly, none of the

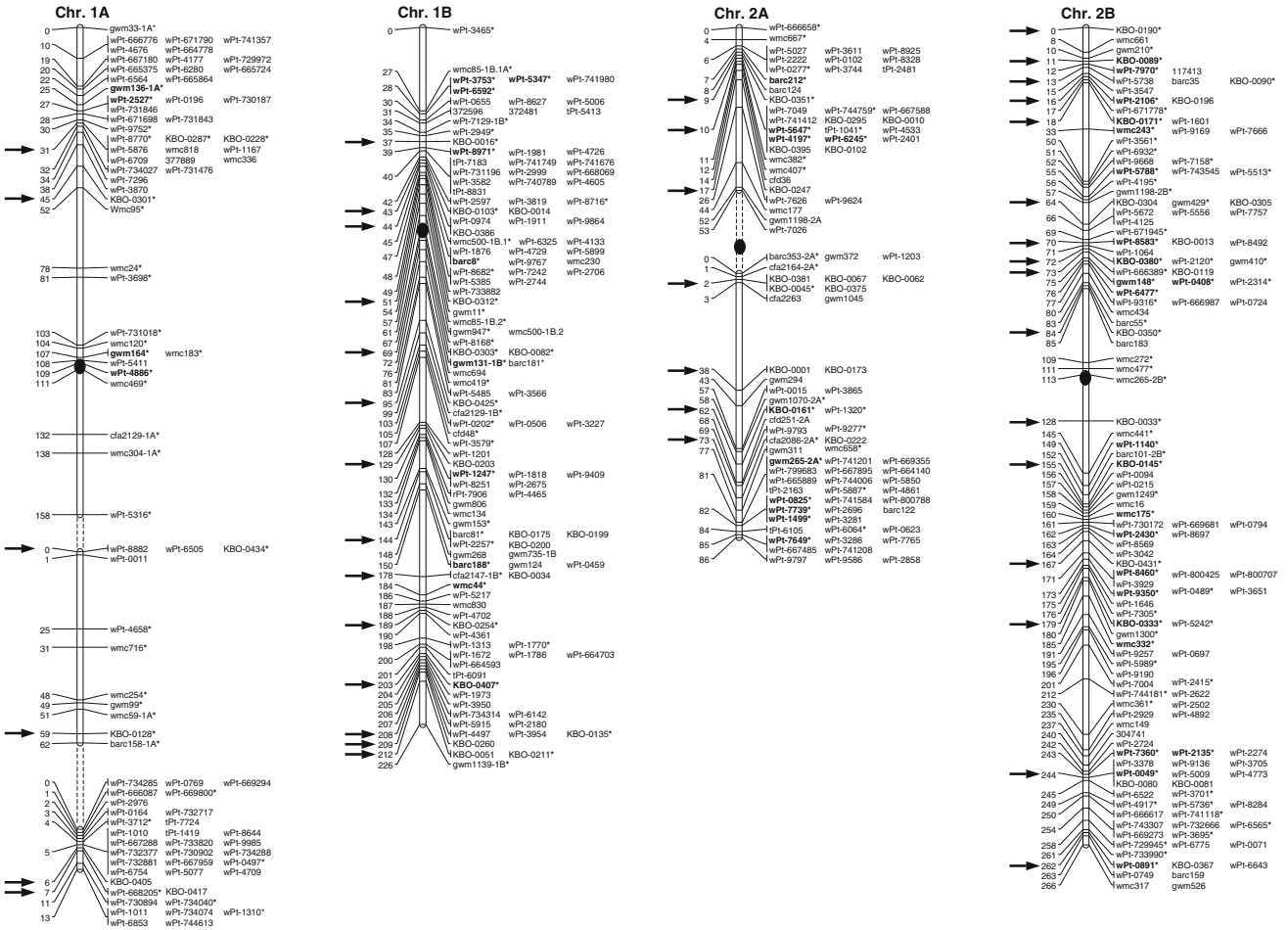

Fig. 2 Integrated genetic map of the 14 *Triticum durum* chromosomes. *Circles* indicate the position of centromeres. *Dashed lines* link linkage groups of the same chromosome. *Stars* indicate framework markers. *Bold* indicates common markers. *Arrows* indicate SNP positions

275 correctly genotyped SNPs was reported in this database, indicating the CRoPS technology reports on genomic regions not yet represented in the Bristol's Wheat SNP database.

Discussion

The repetitive nature of polyploid genomes has been one of the major obstacles to SNP discovery and their utilization in wheat breeding programs (Ganal and Röder 2007). The presence of similar sequences between homoeologous chromosomes reduces both the efficiency of marker discovery and the ability of correctly identifying the allelic state of each individual. Depending on the genomic regions considered, it has been estimated that the SNP frequency between homoeologous chromosomes in wheat (homoeologous sequence variants, also reported as "false SNPs") ranges from 1 in 24 to 61 bp (Somers et al. 2003; Barker and Edwards 2009), while among homologous chromosomes (inter-varietal or "true SNPs") such frequency varies from 1 in 233 to 613 bp depending on the genetic background and the number of accessions considered (Ravel et al. 2006, 2007; Schnurbusch et al. 2007; Barker and Edwards 2009). In a recent study, Barker and Edwards (2009) estimated that only 6% of discovered SNPs in wheat can be used in breeding programs.

Use of CRoPS for genome-wide SNP discovery in durum wheat

To increase the breeding applicability of newly discovered markers, van Orsouw et al. (2007) developed the complexity reduction of polymorphic sequences (CRoPS®), a method that combines genomic complexity reduction based on AFLP and next-generation sequencing. One of the major advantages of CRoPS technology is its low false-discovery rate due to the ability of AFLP to reduce genome complexity mainly based on the discriminating selection between homoeologous regions, thus significantly reducing repetitive and duplicated sequences during marker discovery (Mammadov et al. 2010). In maize, van Orsouw

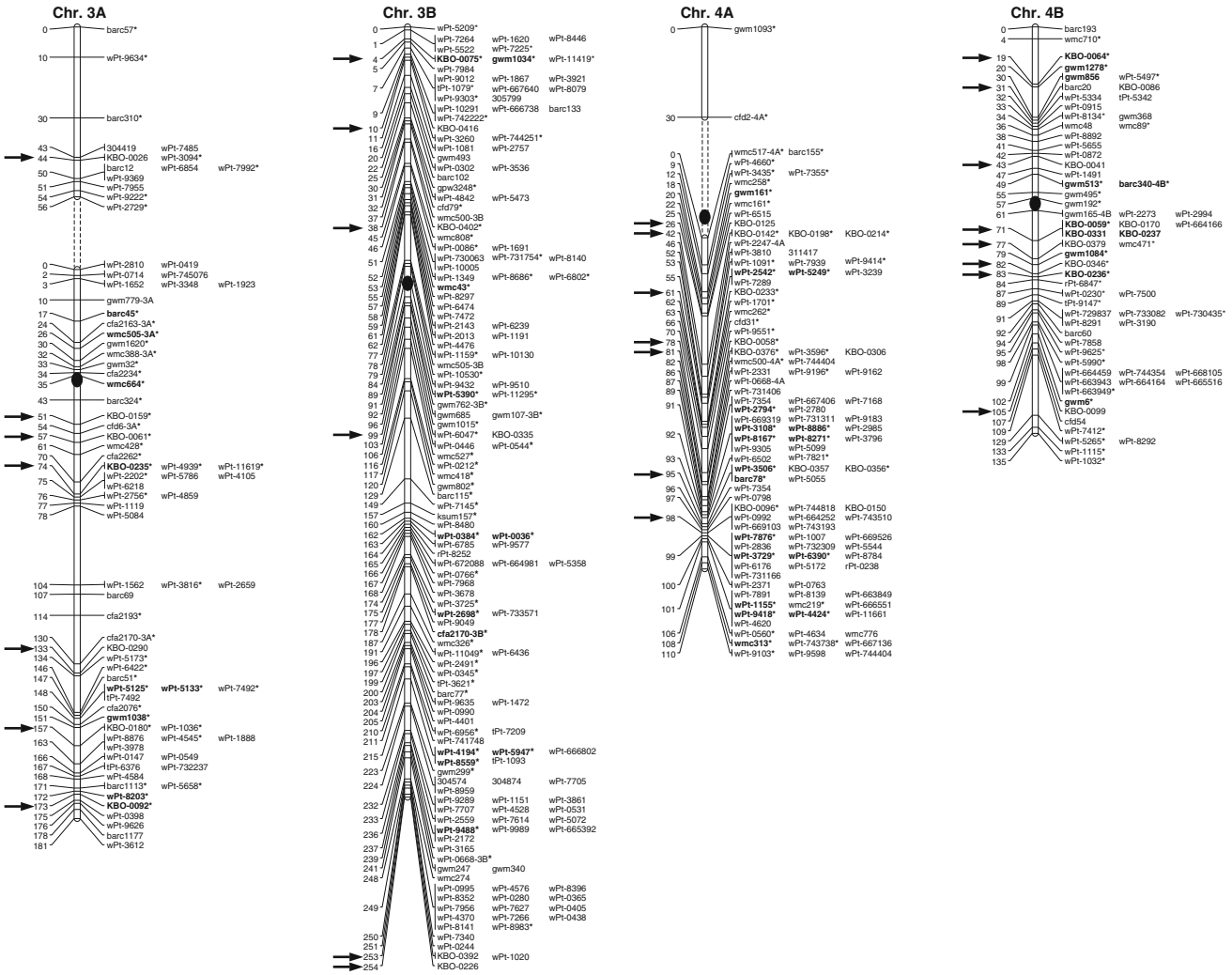

Fig. 2 continued

et al. (2007) validated 77% of the 30 loci tested, while Mammadov et al. (2010) were able to map 63% of the 1,123 CRoPS-derived SNPs. Compared to these studies, the percentage of validated SNP assays was considerably lower in our durum wheat study, most likely due to the high level of repetitiveness of a portion of sequences. This problem can be controlled, at least in part, by in silico filtering for single-copy sequences. Another advantage of CRoPS is the possibility to modulate the level of genome complexity by varying the enzyme combinations and the selective nucleotides during AFLP library preparation. Moreover, the same level of complexity reduction can be applied at the genotyping level, hence partially circumventing the problem of the homoeologous chromosomes in polyploids.

In the AFLP-based reduced representation library preparation, application of the *Pst*I/*Taq*I enzyme combination allowed us to compare the distribution of the CRoPS-derived SNPs with that of the DArT markers previously mapped in the two RIL populations herein considered. The *Pst*I/*Taq*I enzyme combination intrinsically selected for low-repetitive sequences as shown by the fact that only 32.2% of the SNP-harboring sequences were classified as repetitive, while the percentage of repetitiveness in the wheat genome is estimated to be ca. 77% (Flavell et al. 1977), and up to 86% for chromosome 3B (Paux et al. 2006). A similar level of reduction of repetitive sequence frequency was also observed in maize when utilizing the *Pst*I/*Mse*I enzyme combination for the CRoPS library preparation (Mammadov et al. 2010).

The stringency of the SNP-mining parameters was also important to select for SNP assays that complied with the requirement of accurate single-locus Mendelian segregation. The discrimination between SNPs from homoeologous sequences and inter-varietal SNPs was efficiently carried out by selecting only those SNPs that were

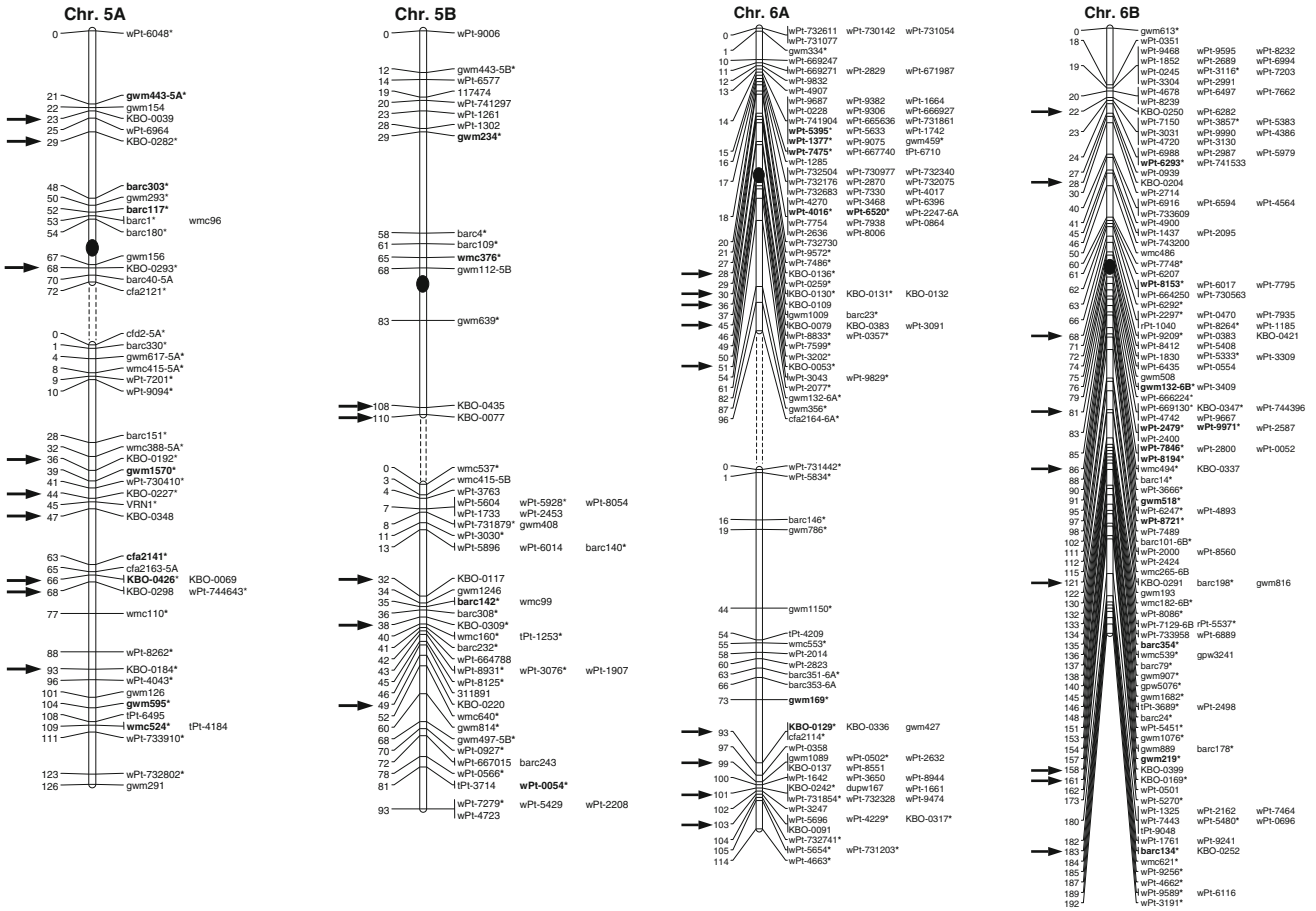

Fig. 2 continued

homozygous in nucleotide stretches without any further polymorphism. In fact, based on the average wheat SNP frequency (Ravel et al. 2006, 2007; Schnurbusch et al. 2007; Barker and Edwards 2009), we would expect from ca. 8,000 to 21,000 "true SNPs" in the 31,665 consensus sequences obtained. However, only 2,659 SNPs in 1,206 consensus sequences met the stringent criteria adopted during the SNP-mining pipeline.

Genotyping of the CRoPS-derived SNPs with the Illumina Golden Gate assay

In our study, the CRoPS technology for SNP discovery in durum wheat was applied by (1) selecting four highly diverse and representative cultivars widely exploited in breeding programs and (2) evaluating allelic information content on a selected sample of 12 diverse cultivars. In order to validate and estimate the breeding quality of the CRoPS-derived SNPs, two RIL populations were genotyped with the Illumina Golden Gate (GG) assay. The OPA genotyping assay is a complex multiplex single-tube reaction (Shen et al. 2005). So far, a high success rate in the multiplex genotyping reaction has been obtained only with SNP assays derived from single-copy sequences. Accordingly, the use of genomic DNA as starting material for the OPA genotyping reaction can be problematic in species with highly complex genomes, where paralogous and homoeologous sequences can interfere with the assay.

The success rate of the OPA genotyping assay when applied to sets of SNP-harboring sequences that passed several selection steps for the reduction of repetitive sequences, was reported to be 89 and 85% in soybean and maize, respectively (Hyten et al. 2008; Jones et al. 2009). Akhunov et al. (2009), using sequences from single-copy genes, demonstrated that the GG assay is very efficient for SNP genotyping in tetraploid wheat, reaching a success rate of 88% on a 96-plex OPA assay. Similar high genotyping conversion rate (85%) was also achieved in tetraploid and hexaploid wheat on 1,536-plex GG assay of EST-derived SNPs amplified with genome-specific primers (Akhunov et al. 2010; Chao et al. 2010). However, restricting the SNP development to the expressed, single-copy-sequence gene portion of the genome may limit

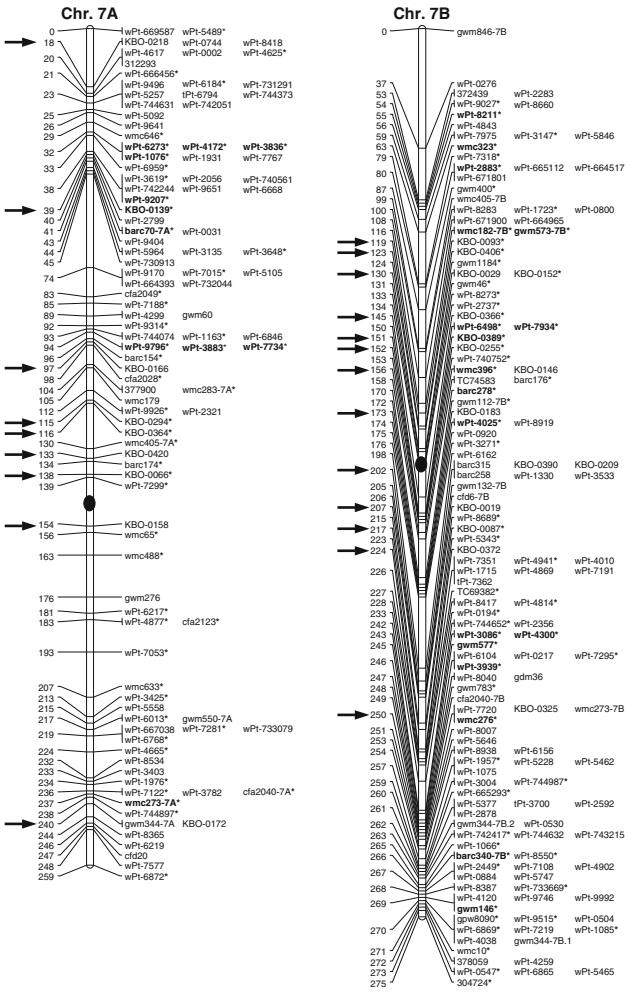

Fig. 2 continued

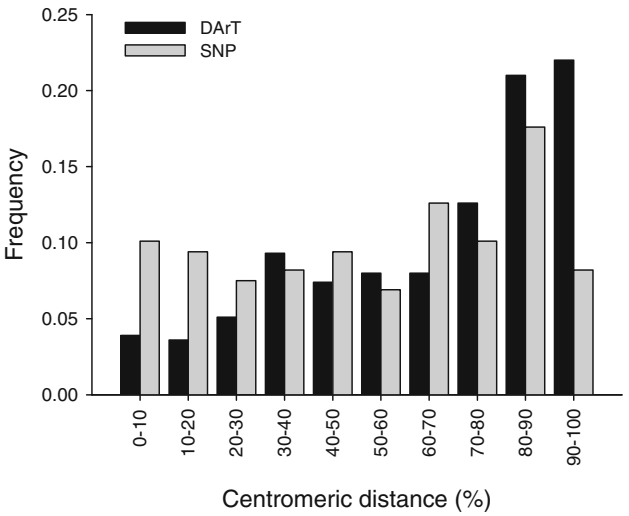

Fig. 3 Chromosome arm distributions of diversity array technology based (DArT) and single-nucleotide polymorphism (SNP) markers expressed as relative distance (%) from the centromere

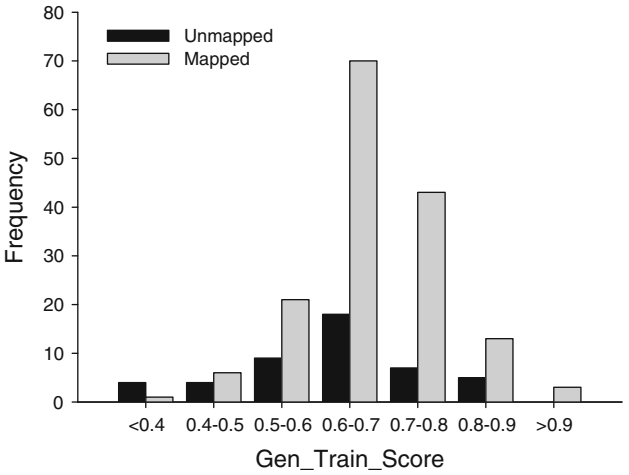

Fig. 4 Frequency of the 204 segregating SNPs in the two RIL populations Claudio × Meridiano and Colosseo × Lloyd as function of the quality score of the Illumina genotyping (Gen_Train_Score, GTS). *Percentages* indicate the proportion of mapped SNPs for each GTS class

marker density and genome coverage offered by SNPs. One peculiarity of the CRoPS technology is its ability to also identify SNPs found in the repetitive sequences due to the reduction of genomic complexity before sequencing the preamplified reactions. Nevertheless, these SNPs may be problematic to be correctly genotyped if the reduction of genomic complexity is not performed. In this study, the experiments were carried out starting with SNP sets in genomic sequences that were not pre-selected for their low repetitiveness or for being expressed. We initially tested SNP genotyping starting directly from genomic DNA, without the AFLP pre-selection step for complexity reduction, but the relatively high number of SNPs simultaneously tested in the OPA assay (384-plex) combined with the high complexity and repetitive nature of the tetraploid wheat genome resulted in a very low success rate of the genotyping assay. Differently, when we tested SNP genotyping starting from the 1:10 diluted preamplifcation reaction as utilized in the CRoPS protocol, we obtained a much higher success genotyping rate. The largest number of undetermined SNP alleles was observed when using genomic DNA, while the proportion of undetermined SNPs considerably dropped when the genomic complexity of the starting material was reduced by the AFLP preamplification step. Our results clearly indicate that the efficiency of high-throughput SNP genotyping on CRoPS-derived SNPs increases when using preamplified reations as template rather than genomic DNA and a complexity-reduction step should be performed to increase the genotyping success rate.

The influence of repetitive and homoeologous sequences on the accuracy of genotyping was still observed despite

the reduction of genomic complexity of the starting material for the GG assay. The RIL populations were derived from several generations of inbreeding and thus only a small portion of residual heterozygosity was observed with SSRs (ca. 4.5 and 6.7% of markers in the Meridiano × Claudio and Colosseo × Lloyd RIL populations, respectively). However, almost 10% of the mapped SNPs showed a small allelic cluster separation (small theta value differences), indicating putative heterozygosity states at these loci, which can be justified by assuming that the allele- and locus-specific oligos of ca. 5% of these SNPs also ligated to and amplified the homoeologous sequences during the OPA assay in the tetraploid genome. Another confounding effect caused by the homoeologous and paralogous sequences was that only slightly more than one-third (35.8%) of the two 384-plex SNPs was correctly genotyped, while the majority of SNPs were undetermined (41.8%) or incorrectly genotyped (22.4%). However, despite the relatively low genotyping efficiency, these results are in keeping with those from Barker and Edwards (2009) who estimated that 74% of discovered SNPs in allohexaploid wheat are derived from homoeologous sequences. Further analysis of the sequences utilized in the GG assays indicated that part of this detrimental effect on genotyping efficiency, particularly on the incorrectly genotyped SNPs, was due to the repetitiveness of the SNP-harboring sequences. By filtering for repetitiveness before designing the GG assay, the percentage of correctly genotyped SNPs would increase to almost half of the total genotypes.

The ploidy effect on the allelic Cy3/Cy5 intensity ratios was also observed on the distribution of the GTSs. Between diploid homozygous lines, GTSs are expected to cluster at 0 and 1 positions based on the absence or presence of polymorphism, respectively. Our results show that the majority of the GTSs for the SNP markers mapped in tetraploid wheat ranged between 0.6 and 0.8. Based on this finding and what is known for species with a complex genome, it has been suggested that a threshold of cluster separation >0.4 should be used for the efficient selection of SNPs for GG assays (Hyten et al. 2008).

Breeding applicability of the discovered SNP markers

The breeding applicability and the effectiveness of the CRoPS-derived genome-wide SNPs is exemplified by the high frequency of segregation and mappability in the two RIL populations, by the high values of MAF and the even distribution of the markers along the wheat genome. Almost three-quarters of the correctly genotyped CRoPS-derived SNPs segregated in at least one RIL population and the vast majority (>75%) were mapped with high confidence. Distribution along chromosomes was more uniform in case of the CRoPS-derived SNPs compared to DArT markers, notwithstanding the use of the same methylation-sensitive enzyme during library construction, which has been suggested to cause a skewed distribution on DArT markers toward the distal hypomethylated chromosomal regions (Wenzl et al. 2006). In bread wheat, Barker and Edwards (2009) detected a small but significant increase (+15%) of SNP markers on chromosomes 4 and 7, similar to what we have shown for chromosome 7B. The MAF estimated on a panel of 12 diverse lines indicated that the majority of the SNPs had MAF >0.2, which is generally considered as a threshold for high quality and useful markers (Akhunov et al. 2009; Mammadov et al. 2010). The selection of SNPs with high MAF is probably due to both the stringent SNP-mining pipeline used during CRoPS discovery and the high genetic diversity of the four genotypes used to identify SNPs. Another positive aspect of the CRoPS-derived SNPs was that none of the newly identified SNPs was previously reported in the University of Bristol's Wheat SNP Database, indicating that our CRoPS-based study explored previously uncharacterized chromosomal regions.

Conclusions

High-throughput methods for streamlining SNP discovery and genotyping are urgently needed in species with high genome complexity like wheat. We have described the application of the high-throughput CRoPS technology for the discovery in tetraploid durum wheat cultivars of SNPs well suited for breeding applications. The CRoPS-based SNP discovery methodology effectively distinguished single-locus inter-varietal SNPs from the vast majority of sequence variants between the two genomes. Additionally, the SNPs identified in this study showed valuable features for marker-assisted and genomic selection applications, such as (1) high proportion of mappability as single-locus codominat assays, (2) high frequency of markers characterized by high informativeness, (3) even distribution along the genome and (4) novelty compared to already known wheat SNPs. Though the final success rate in developing Illumina-based single-locus Mendelian genotyping assays is lower compared to the SNP-discovery methodologies that rely only on the expressed portion of the genome, CRoPS allows for the large-scale discovery and development of SNPs in wheat at the whole-genome sequence level. The genotyping success rate can be increased with the selection of SNPs in non-repetitive sequences and with higher designability scores. Furthermore, our study shows that through genome complexity reduction, CRoPS allows for high-throughput genome-wide identification of polymorphic SNPs in a polyploid species.

Acknowledgments The financial contribution of the European Project "BioExploit Food CT 2005-513959" is gratefully acknowledged. The AFLP® and CRoPS® technologies are subjected to patents and patent applications owned by Keygene N.V. AFLP and CroPS are registered trademarks of Keygene N.V.

References

Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Curr Opin Plant Biol 8:135–141

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

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

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Deal KR, Dubcovsky J, Gill BS, Gu YQ, Hadam J, Heo H, Huo N, Lazo GR, Luo MC, Ma YQ, Matthews DE, McGuire PE, Morrell PL, Qualset CO, Renfro J, Tabanao D, Talbert LE, Tian C, Toleno DM, Warburton ML, You FM, Zhang W, Dvorak J (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11:702

Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle C, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA (2007) Direct selection of human genomic loci by microarray hybridization. Nat Methods 4:903–905

Altshuler D, Pollara VJ, Cowles CR, van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513–516

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51:910–918

Barker GLA, Edwards KJ (2009) A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. Plant Biotechnol J 7:318–325

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Ann Bot 95:127–132

Brookes AJ (1999) The essence of SNPs. Gene 234:177–186

Chao S, Dubcovsky J, Dvorak J, Luo MC, Baenziger SP, Matnyazov R, Clark DR, Talbert LE, Anderson JA, Dreisigacker S, Glover K, Chen J, Campbell K, Bruckner PL, Rudd JC, Haley S, Carver BF, Perry S, Sorrells ME, Akhunov ED (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genomics 11:727

Chen X, Levine L, Kwok PY (1999) Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res 9:492–498

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Close T, Bhat P, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson J, Wanamaker S, Bozdag S, Roose M, Moscou M, Chao S, Varshney R, Szucs P, Sato K, Hayes P, Matthews D, Kleinhofs A, Muehlbauer G, DeYoung J, Marshall D, Madishetty K, Fenton R, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582

Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc B: J Biol Sci 363:557–572

de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T (2005) Carthagene: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 21:1703–1704

Deschamps S, Campbell M (2010) Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. Mol Breed 25:553–570

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid whet under domestication. Science 316:1862–1866

Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnol J 8:2–9

Edwards KJ, Poole RL, Barker GLA (2008) SNP discovery in plants. In: Henry RJ (ed) Plant genotyping II: SNP Technology. Wallingford, Oxfordshire, pp 1–29

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Wang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS (2003) High parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 68:69–78

Feltus FA, Wasn J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. Genome Res 14:1812–1819

Flavell RB, Rimpau J, Smith DB (1977) Repeated sequence DNA relationships in four cereal genomes. Chromosoma 63:205–222

Ganal MW, Röder MS (2007) Microsatellite and SNP markers in wheat breeding. In: Varshney RK, Tuberosa R (eds) Genomics-assisted crop improvement, vol 2: genomics applications in crops. Springer, New York, pp 1–24

Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217

Gunderson KL (2009) Whole-genome genotyping on bead arrays. Methods Mol Biol 529:197–213

Gut IG (2001) Automation in genotyping of single nucleotide polymorphisms. Hum Mutat 17:475–492

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S, David J (2007) Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol Biol Evol 24:1506–1517

Hayden MJ, Tabone TL, Nguyen TM, Coventry S, Keiper FJ, Fox RL, Chalmers KJ, Mather DE, Eglinton JK (2009) An informative set of SNP markers for molecular characterisation of Australian barley germplasm. Crop Pasture Sci 61:70–83

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

Hyten DL, Song Q, Choi IY, Yoon MS, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the Golden Gate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) Arabidopsis map-based cloning in the post-genome era. Plant Physiol 129:440–450

Jannink J, Lorenz AJ (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177

Jones E, Chu WC, Ayele M, Ho J, Bruggeman E, Yourstone K, Rafalski A, Smith OS, McMullen M, Bezawada C, Warren J, Babayev J, Basu S, Smith S (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maze (*Zea mays* L.) germplasm. Mol Breed 24:165–176

Koebner R, Summers R (2002) The impact of molecular markers on the wheat breeding paradigm. Cell Mol Biol Lett 7:695–702

Korzun V, Röder M, Wendehake K, Pasqualone A, Lotti C, Ganal M, Blanco A (1999) Integration of dinucleotide microsatellites from hexaploid bread wheat into a genetic linkage map of durum wheat. Theor Appl Genet 98:1202–1207

Liu S, Chen HD, Makarevitch I, Shirmer R, Emrich SJ, Dietrich CR, Barbazuk WB, Springer NM, Schnable PS (2010) High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. Genetics 184:19–26

Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotide with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. Genome Res 4:357–362

Lotti C, Salvi S, Pasqualone A, Tuberosa R, Blanco A (2000) Integration of AFLP markers into an RFLP-based map of durum wheat. Plant Breed 119:393–401

Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005) Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol Breed 15:271–290

Maccaferri M, Sanguineti MC, Natoli V, Araus Ortega JL, Ben Salem M, Bort J, Chenenaoui S, De Ambrogio E, Garci'a del Moral L, Demontis A, El-Ahmed A, Maalouf F, Machlab H, Moragues M, Motawaj J, Nachit M, Nserallh N, Ouabbou H, Royo C, Tuberosa R (2006) A panel of elite accessions of durum wheat (*Triticum durum* Desf.) suitable for association mapping studies. Plant Genet Resour 4:79–85

Maccaferri M, Mantovani P, Tuberosa R, Deambrogio E, Giuliani S, Demontis A, Massi A, Sanguineti MC (2008) A major QTL for durable leaf rust resistance widely exploited in durum wheat breeding programs maps on the distal region of chromosome arm 7BL. Theor Appl Genet 117:1225–1240

Maccaferri M, Sanguineti MC, Demontis A, El-Ahmed A, Garcia del Moral L, Maalouf F, Nachit M, Nserallah N, Ouabbou H, Rhouma S, Royo C, Villegas D, Tuberosa R (2011) Association mapping in durum wheat grown across a broad range of water regimes. J Exp Bot 62:409–438

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63

Mammadov JA, Chen W, Ren R, Pai R, Marchione W, Yalcin F, Witsenboer H, Greene TW, Thompson SA, Kumpatla SP (2010) Development of highly polymorphic SNP markers from the complexity reduced portion of maize (*Zea mays* L.) genome for use in marker-assisted breeding. Theor Appl Genet 121:577–588

Mantovani P, Maccaferri M, Sanguineti MC, Tuberosa R, Catizone I, Wenzl P, Thomson B, Carling J, Huttner E, DeAmbrogio E, Kilian A (2008) An integrated DArT–SSR linkage map of durum wheat. Mol Breed 22:629–648

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE, Stokowski R, Ballinger DG, Frazer KA, Cox DR, Padhukasahasram B, Bustamante CD, Weigel D, Mackill DJ, Bruskiewich RM, Rãtsch G, Buell CR, Leung H, Leach JE (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA 106:12273–12278

Nijman IJ, Kuipers S, Verheul M, Guryev V, Cuppen E (2008) A genome-wide SNP panel for mapping and associations studies in the rat. BMC Genomics 9:95

Okou DT, Steinberg KM, Middle C, Cutler DJ, Albeert TJ, Zwick ME (2007) Microarray-based genomic selection for high-throughput resequencing. Nat Methods 4:907–909

Oliphant A, Barker DL, Stuelpnagel JR, Chee MS (2002) BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques 32:S56–S61

Olivier M (2005) The Invader assay for SNP genotyping. Mutat Res 573:103–110

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. Genome Res 18:2024–2033

Palmer LE, Rabinowicz PD, O'Shaughnessy A, Balija VLN, Dike S, de la Bastide M, Martienssen RA, McCombie WR (2003) Maize genome sequencing by methylation filtration. Science 302:2115–2117

Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J 48:463–474

Peleg Z, Saranga Y, Suprunova T, Ronin Y, Röder M, Kilian A, Korol A, Fahima T (2008) High-density genetic map of durum wheat × wild emmer wheat based on SSR and DArT markers. Theor Appl Genet 117:103–115

Quraishi U, Abrouk M, Bolot S, Pont C, Throude M, Guilhot N, Confolent C, Bortolini F, Praud S, Murigneux A, Charmet G, Salse J (2010) Genomics in cereals: from genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. Funct Integr Genomics 9:473–484

Rafalski A (2002) Novel genetic mapping tools in plant: SNPs and LD-based approaches. Plant Sci 162:329–333

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhoub B, Brunel D, Beckert M, Charmet G (2006) Single-nucleotide polymorphisms (SNPs) frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). Genome 49:1131–1139

Ravel C, Praud S, Canaguier A, Dufour P, Giancola S, Balfourier F, Chalhoub B, Brunel D, Linossier L, Dardevet M, Beckert M, Rousset M, Murigneux A, Charmet G (2007) DNA sequence polymorphisms and their application to bread wheat quality. Euphytica 158:331–336

Salvi S, Tuberosa R, Phillips RL (2001) Development of PCR-based assays for allelic discrimination in maize by using the 5′-nuclease procedure. Mol Breed 8:169–176

Sanchez J, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N (2003) Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. Forensic Sci Int 137:74–84

Sandve SR, Rudi H, Dørum G, Vigeland MD, Berg PR, Rognli OA (2010) Genotyping unknown genomic terrain in complex plant genomes. In: Huyghe C (ed) Sustainable use of genetic diversity in forage and turf breeding. Springer, New Mexico, pp 455–459

Schnurbusch T, Collins NC, Eastwood RF, Sutton T, Jefferies SP, Langridge P (2007) Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the Bo1 boron toxicity tolerance locus of bread wheat. Theor Appl Genet 115:451–461

Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, Yeakley J, Bibikova M, Garcia EW, McBride C, Steemers F, Garcia F, Kermani BG, Gunderson K, Oliphant A (2005) High-throughput SNP genotyping on universal bead arrays. Mutat Res 573:70–82

Shendure J, Li H (2008) Next-generation DNA sequencing. Nat Biotechnol 26:1135–1145

Somers DJ, Kirkpartick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46:431–437

Song QJ, Shi JR, Singh S, Fickus EW, Costa JM, Lewis J, GB S, Ward R, Cregan PB (2005) Development and mapping of microsatellite (SSR) markers in wheat. Theor Appl Genet 110:550–560

van Eijk MJT, van der Poel (2006) Strategies for high throughput identification and detection of polymorphisms WO2006137733

Van Ooijen JW (2006) JoinMap 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen

van Orsouw NJ, Hogers RCJ, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, van der Poel H, van Oeveren J, Verstegen H, van Eijk MJT (2007) Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. PLoS One 2:e1172

Van Tassel PC, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5:247–252

Varshney RK, Spurthi N, Nayak S, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Varshney RK, Glaszmann JC, Leung H, Ribaut JM (2010) More genomic resources for less-studied crops. Trends Biotechnol 28:452–460

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lubberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935

Waugh R, Jannink J, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. Curr Opin Plant Biol 12:218–222

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesná J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:206

Wright WT, Heggarty SV, Young IS, Nicholls DP, Whittall R, Humphries SE, Graham CA (2008) Multiplex MassARRAY spectrometry (iPLEX) produces a fast and economical test for 56 familial hypercholesterolaemia-causing mutations. Clin Genet 74:463–468

Yamamoto T, Nagasaki H, Yonemaru J, Ebana K, Nakajima M, Shibaya T, Yano M (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267

Yan J, Yang X, Shah T, Sánchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch J, Xu Y (2010) High-throughput SNP genotyping with the Golden Gate assay in maize. Mol Breed 25:441–451

Yuan Y, SanMiguel PJ, Bennetzen JL (2003) High-C_ot sequence analysis of the maize genome. Plant J 34:249–255

Zhang X, Borevitz JO (2009) Global analysis of allele-specific expression in *Arabidopsis thaliana*. Genetics 182:943–954

Zhao W, Canaran P, Jurkuta R, Fulton T, Glaubitz J, Buckler ES, Doebley J, Gaut B, Goodman M, Holland J, Kresovich S, McMullen M, Stein L, Ware D (2006) Panzea: a database and resource for molecular and functional diversity in the maize genome. Nucleic Acids Res 34:D752–D757