Mol Breeding (2010) 26:229–242
DOI 10.1007/s11032-009-9373-9

Features of SNP and SSR diversity in a set of ICARDA barley germplasm collection

R. K. Varshney · M. Baum · P. Guo · S. Grando · S. Ceccarelli · A. Graner

Received: 31 January 2009 / Accepted: 10 December 2009 / Published online: 7 January 2010

Abstract Detection and utilization of genetic variation available in the germplasm collection for crop improvement have been the prime activities of breeders. Here a set of ICARDA barley germplasm collection comprising of 185 cultivated (*Hordeum vulgare* L.) and 38 wild (*H. spontaneum* L.) genotypes originated from 30 countries of four continents was genotyped with 68 single nucleotide polymorphism (SNP) and 45 microsatellite or simple sequence repeat (SSR) markers derived from genes (expressed sequence tags, ESTs). As two SNP markers provided 2 and 3 datapoints, a total of 71 SNPs were surveyed that yielded a total of 143 alleles. The number of SSR alleles per locus ranged from 3 to 22 with an average of 7.9 per marker. Average PIC (polymorphism information content) value for SSR and SNP markers were recorded as 0.63 and 0.38, respectively. Heterogeneity was recorded at both SNP and SSR loci in an average of 5.72 and 12.42% accessions, respectively. Genetic similarity matrices for SSR and SNP allelic data were highly correlated ($r = 0.75$, $P < 0.005$) and therefore allelic data for both markers were combined and analyzed for understanding the genetic relationships among the germplasm surveyed. Majority of clusters/subclusters were found to contain genotypes from the same geographic origins. While comparing the genetic diversity, the accessions coming from Middle East Asia and North East Asia showed more diversity as compared to that of other geographic regions. Majority of countries representing Africa, Middle East Asia, North East Asia and Arabian Peninsula included the genotypes that contained rare alleles. As expected, *spontaneum* accessions, as compared to *vulgare* accessions, showed a higher number of total alleles, higher number of alleles per locus, higher effective number of alleles and higher allelic richness and a higher number of rare alleles were observed. In summary, the examined ICARDA germplasm set showed ample natural genetic variation that can be harnessed for future breeding of barley as climate change and sustainability have become important throughout all growing areas of the world, drought/heat tolerance being the most important ones.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-009-9373-9) contains supplementary material, which is available to authorized users.

R. K. Varshney · A. Graner
Leibniz Institute of Plant Genetics & Crop Plant Research (IPK), Corrensstrasse 3, 06466 Gatersleben, Germany

M. Baum · S. Grando · S. Ceccarelli
International Centre for Agricultural Research in Dry Areas (ICARDA), P.O. Box 5466, Aleppo, Syria

P. Guo
College of Life Science, Guangzhou University, 510006 Guangzhou, China

R. K. Varshney (✉)
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Greater Hyderabad, AP 502324, India
e-mail: r.k.varshney@cgiar.org

Keywords Genetic diversity · Heterogeneity · Molecular markers · SSR · SNP

Introduction

Barley (*Hordeum vulgare* L.) is an important feed and food cereal crop ranking fourth in the world food production and is grown in a wide range of environments. Large germplasm collections of cultivated (*Hordeum vulgare* subsp. *vulgare*) as well as wild (*Hordeum vulgare* subsp. *spontaneum*) barleys are available in several genebanks e.g. ICARDA (Syria) and IPK-Gatersleben (Germany) that encompass enormous genetic diversity. The systematic evaluation of molecular genetic variation existing in these germplasm collections will be very useful for its efficient harnessing in breeding as well as to define strategies for conservation of genetic diversity (Takeda and Matsuoka 2008).

Different types of molecular markers in large numbers are available in barley (see Varshney et al. 2005b). Microsatellite or simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) marker systems have emerged as the markers of choice in different plant species including barley (Gupta and Varshney 2000; Rafalski 2002). Recently, it has also been possible to develop the markers from expressed sequence tags (ESTs) or genes. Markers developed from genes have been referred as genic molecular markers (GMMs, Varshney et al. 2007b) or functional markers (FMs, Andersen and Lübberstedt 2003) as a putative function based on BLASTX analysis can be deduced for majority of such markers. In case of barley, a large number of EST or gene-derived SSR (Varshney et al. 2006) and SNP (Rostoks et al. 2005, Kota et al. 2008) markers have been developed.

Several genetic diversity studies have been conducted in local, regional or national germplasm collection of barley using SSR markers (e.g. Matus and Hayes 2002; Baek et al. 2003; Koebner et al. 2003; Russell et al. 2003; Karakousis et al. 2003; Malysheva-Otto et al. 2006; Pandey et al. 2006; Kolodinska et al. 2007; Gusami et al. 2008). The majority of these studies employed SSR markers that are derived from anonymous region of the genome. However, due to availability of extensive amounts of EST or candidate gene sequence data, a few recent studies have employed the gene-derived SSR or SNP markers for diversity analysis (e.g. Kota et al. 2001; Kanazin et al. 2002; Russell et al. 2004; Rostoks et al. 2005). Gene derived markers are supposed to assay functional genetic variation in germplasm collection that can be associated with adaptation of germplasm in different climatic regions (Eujayl et al. 2001; Varshney et al. 2005a). It is also noteworthy that use of SNP markers generate a simple binary output which is well suited to automatic data collection systems, their use is gaining momentum (Rostoks et al. 2005; Varshney et al. 2007a). However, SNP genotyping in large germplasm collections is a capital intensive process, unless the markers can be converted into cleaved amplified polymorphic sequence (CAPS) assays (Kota et al. 2008; Varshney et al. 2008) or used in high-throughput genotyping assay e.g. GoldenGate assay of Illumina (Rostoks et al. 2005).

Genetic diversity studies are very useful to understand the genetic structure of a germplasm collection so that linkage disequilibrium (LD) based association genetics studies can be undertaken for mapping traits of interest. In case of barley, a few studies have been undertaken recently to understand the LD around certain regions of genome or genome-wide level and also for association analysis for trait of interest (Kraakman et al. 2004; Morrell et al. 2005; Rostoks et al. 2005; Caldwell et al. 2006; Stracke et al. 2007). The ICARDA genebank contains a collection of nearly 28,000 accessions of barley collected from different countries. Because of availability of high-throughput genotyping assay i.e. GoldenGate assay of Illumina (Close et al. 2009), it would be feasible in coming years to genotype the entre germplasm collection of ICARDA. However, at the time of undertaking this study, this platform was not available.

With an objective of understanding diversity in a set of ICARDA germplasm collection, comprised of 223 accessions sampled from four continents and 30 countries mainly from the Middle East Asia, a total of 68 SNP (71 datapoints) and 45 SSR markers derived from expressed sequence tags (ESTs) were employed in this study. Allelic data for these markers were evaluated mainly to: (a) understand the SSR and SNP-based allelic diversity, and (b) investigate the genetic relationships of the barley accessions of the collection. In addition, some allelic features of SSR markers have also been discussed in the article.

Materials and methods

Plant material

A total of 223 barley (*Hordeum vulgare* L.) accessions containing 185 *H. vulagre* subsp. 'vulgare' L. (*vulgare* or VUL) and 38 *H. vulgare* subsp. 'spontaneum' (*spontaneum* or SPON) lines were used. Details about these lines are given in Table ESM 1 and at website http://pgrc.ipk-gatersleben.de/drought/. In brief, four continents i.e. Africa (52), Asia (123), Australia (2) and Europe (3) representing 30 countries contributed genotypes examined in present study (Table 1).

DNA was isolated from these genotypes as mentioned in Thiel et al. (2003).

Genotyping and marker analysis

A set of 68 SNP (yielding 71 SNP datapoints) and 45 SSR markers including 28 each of SNP and SSR markers of the core set identified earlier (Varshney et al. 2008) and 27 CAPS-based SNP markers developed in another study (Kota et al. 2008) were used for genotyping. Details on the SNP and SSR markers used in the study are provided in Table ESM 2 and 3, respectively.

SNP genotyping of the germplasm collection was done using CAPS and pyrosequencing assays as mentioned earlier (Varshney et al. 2008). Similarly, amplification of microsatellite loci using fluorescent-dye labeled primer pairs, separation of amplification products on an ABI377 fragment analyzer and evaluation of fragments using GenoTyper 3.7 (Applied Biosystems) were carried out as given earlier (Thiel et al. 2003; Varshney et al. 2008).

Allelic data obtained for SSR were scored as allele size for calculating the allelic features (like allele richness, unique alleles, etc.) as well as 0 (absence)- 1 (presence) matrix for preparing and comparing dendrogram with SNP genotyping data. SNP genotyping data were score in binary fashion i.e. 0 and 1 for absence and presence of the alleles for the given SNP locus.

Diversity analysis

The PIC (polymorphic information content) values of SSR and SNP markers were calculated as given in Thiel et al. (2003).

Effective number of alleles in a group were calculated as given in Morgante et al. (1994). The allelic richness per locus and per population was computed using FSTAT version 2.9.3.2 (Goudet 2002).

The genetic similarities for each pair of lines using Jaccard's similarity coefficient were calculated with the NTSYS-pc 2.11 software package (Biostatistics Inc., USA). Cluster analysis (SAHN clustering) was performed based on UPGMA (Unweighted Pair Group Method with Arithmetic Average) for preparing dendrogram. Bootstrapping was carried out using 1000 iterations on PAUP* 4.0 Version 4.0b10 (for McIntosh) to evaluate the reproducibility of nodes of phenograms. Correlations between SSR and SNP matrices were calculated using Mantel test (Mantel 1967) after 10000 random iterations with the help of Mantel Nonparametric Test Calculator (Mantel version 2.0).

Results

SNP-based diversity

In total 68 SNP markers by using CAPS (58), *indel* (5) and pyrosequencing (5) were employed to examine SNP-based diversity (Table 2, Table ESM 2). While CAPS, *indel* and pyrosequencing assays for 55 were developed earlier (Varshney et al. 2008; Kota et al. 2008), CAPS assay were developed for new 13 SNP markers in this study following Thiel et al. (2004). Since two SNP markers namely GBS0461 and GBS0576 in pyrosequencing assay provided 2 and 3 SNP data points, respectively; a total of 71 SNP datapoints were surveyed for the 68 markers examined. As SNP markers are predominantly biallelic markers, 2 alleles each per datapoint were collected. However, an *indel* based marker i.e. GBS0692 yielded 3 alleles as it provided three distinct amplicons of varying length. In total, 143 alleles were observed for 71 SNP datapoints obtained by 68 EST-based markers.

The PIC value for the examined SNPs ranged from 0.06 (GBS0708) to 0.50 (GBS0023, GBS0028, GBS0122, GBS0208, GBS0317, GBS0406, GBS0410, GBS0526, GBS0539, GBS0555 and GBS0576) (Table 2). However, the marker GBS0692 (yielding 3 alleles) provided a PIC value of 0.552. On an average the PIC value for SNP or *in-del* based markers was observed as 0.38 per marker. Markers from the linkage group 4H showed, in general, showed higher PIC values (average

Table 1 Details on origin of barley accessions sampled for genetic diversity

Continent	Geographical region	Country[a]	Number of provinces represented	Number of accessions	Sub-total for region	Total for the continent
Africa	East	Ethiopia (ETH)	2	3	5	52
		Eriteria (ERI)	–	2		
	North	Morocco (MAR)	3	5	47	
		Tunisia (TUN)	4	6		
		Algeria (DZA)	4	9		
		Egypt (EGY)	4	14		
		Libya (LBY)	8	13		
Asia	North East	Afghanistan (AFG)	5	9	68	129
		Tajikistan (TJK)	1	2		
		Turkmenistan (TKM)	8	10		
		Uzbekistan (UZB)	2	2		
		China (CHN)	3	3		
		Pakistan (PAK)	3	9		
		India (IND)	1	2		
		Azerbaijan (AZE)	3	4		
		Cyprus (CYP)	–	1		
		Turkey (TUR)	3	6		
		Iran (IRN)	10	20		
	Middle East	Iraq (IRQ)	3	4	47	
		Jordan (JOR)	7	18		
		Lebanon (LBN)	–	1		
		Palestine (PAL)	1	2		
		Syria (SYR)	8	22		
	Arabian Peninsula	Oman (OMN)	6	7	14	
		Saudi Arabia (SAU)	3	4		
		Yemen (YEM)	3	3		
Australia		Australia (AUS)	–	2	2	2
Europe	Central	Russia (RUS)	–	1	1	3
	Southern	Greece (GRC)	–	1	1	
	Western	France (FRA)	–	1 (1)	1	
Unknown		Unknown (UNK)	–	37	37	37

[a] Standard code for country of origin, e.g. *AFG* = Afghanistan, *AUS* = Australia, *AZE* = Azerbaijan, *CHN* = China, *CYP* = Cyprus, *DZA* = Algeria, *EGY* = Egypt, *ETH* = Ethiopia, *ERI* = Eriteria, *GRC* = Greece, *FRA* = France, *IND* = India, *IRN* = Iran, *IRQ* = Iraq, *JOR* = Jordan, *LBN* = Lebanon, *LBY* = Libya, *MAR* = Morocco, *OMN* = Oman, *PAK* = Pakistan, *PAL* = Palestine, *SAU* = Saudi Arabia, *SYR* = Syria, *TJK* = Tajikistan, *TKM* = Turkmenistan, *TUR* = Turkey, *TUN* = Tunisia, *UZB* = Uzbekistan, *YEM* = Yemen, *UNK* = Unknown

0.44) while the markers from the linkage group 6H had the lower PIC values (0.32).

SSR polymorphism and diversity

A total of 356 alleles were detected by 45 SSR markers with a range from 3 (GBM1020, GBM1404, GBM1456) to 22 (GBM1015) with an average of 7.9 per marker (Table 3). The PIC values for 45 SSR markers in examined germplasm collection ranged from 0.17 (GBM1404) to 0.90 (GBM1015) with an average of 0.63 per marker. Markers from the linkage group 4H, showed higher PIC value (average 0.69) and higher number of alleles (average 9.86) while the

Table 2 Genetic diversity and heterogeneity in germplasm collection based on SNP markers

Linkage group	Marker name	Assay[a]	Number of alleles	PIC per SNP	PIC per marker	Heterogeneity (% accessions)
1H	GBS0131	*Ava* II	2	0.28	0.28	0.45
	GBS0361	*Hha*I	2	0.31	0.31	4.93
	GBS0469	*Rsa* I	2	0.43	0.43	6.28
	GBS0528	*Hpy*CH4 IV	2	0.41	0.41	4.04
	GBS0546	*Sml* I	2	0.35	0.35	0
	GBS0554	*Hha*I	2	0.39	0.39	8.97
2H	GBS0143	*Nsp* I	2	0.30	0.30	0
	GBS0182	IN_DEL	2	0.31	0.31	0.45
	GBS0312	*Alu* I	2	0.35	0.35	3.14
	GBS0379	*Dra* I	2	0.44	0.44	0
	GBS0400	IN_DEL	2	0.19	0.19	1.35
	GBS0535	*Mse* I	2	0.48	0.48	2.69
	GBS0705	PS		0.41	0.41	4.48
3H	GBS0137	*Apo* I	2	0.45	0.45	0
	GBS0419	*Hae* III	2	0.49	0.49	8.52
	GBS0431	*Rsa* I	2	0.49	0.49	8.97
	GBS0526	*Psi* I	2	0.50	0.50	16.14
	GBS0555	*Spe* I	2	0.50	0.50	21.97
	GBS0560	*Hae* III	2	0.38	0.38	0
	GBS0639	*Hpa* II	2	0.08	0.08	1.35
	GBS0667	*Cac8* I	2	0.30	0.30	2.24
4H	GBS0010	*Bts* I	2	0.42	0.42	1.79
	GBS0023	*Dde* I	2	0.50	0.50	0.90
	GBS0192	*Rsa* I	2	0.42	0.42	3.14
	GBS0288	*Hha*I	2	0.48	0.48	8.52
	GBS0349	*Bsa*B I	2	0.34	0.34	7.62
	GBS0461	PS_pos1	2	0.45	0.39	6.73
		PS_pos2	2	0.32		0
	GBS0692	IN_DEL	3	0.55	0.55	5.38
5H	GBS0208	*Nsp* I	2	0.50	0.50	5.38
	GBS0234	IN_DEL	2	0.37	0.37	4.48
	GBS0295	*Cac8* I	2	0.13	0.13	2.24
	GBS0408	*Ava* II	2	0.21	0.21	2.24
	GBS0410	*Ava* I	2	0.50	0.50	0
	GBS0451	*Hpy*CH4 III	2	0.46	0.46	5.83
	GBS0527	*Eco* RV	2	0.44	0.44	7.17
	GBS0539	IN_DEL	2	0.50	0.50	3.59
	GBS0576	PS_pos1	2	0.50	0.48	21.52
		PS_pos2	2	0.48		17.49
		PS_pos3	2	0.47		23.77
	GBS0577	*Dde* I	2	0.39	0.39	7.17
	GBS0669	*Hpy*CH4 III	2	0.22	0.22	0
	GBS0712	*Ava* II	2	0.42	0.42	6.73

Table 2 continued

Linkage group	Marker name	Assay[a]	Number of alleles	PIC per SNP	PIC per marker	Heterogeneity (% accessions)
6H	GBS0136	*Btg* I	2	0.29	0.29	0.45
	GBS0157	*Sal* I	2	0.45	0.45	2.24
	GBS0369	*Hae* III	2	0.39	0.39	6.28
	GBS0388	*Hha* I	2	0.32	0.32	0.49
	GBS0396	*Hinf* I	2	0.27	0.27	1.79
	GBS0489	*Eco*R V	2	0.45	0.45	9.87
	GBS0537	*Nsp* I	2	0.30	0.30	0
	GBS0590	*Hae* III	2	0.34	0.34	0
	GBS0708	PS	2	0.06	0.06	1.35
7H	GBS0028	*Ava* II	2	0.50	0.50	6.73
	GBS0154	IN_DEL	2	0.29	0.29	0.90
	GBS0291	*Hinf* I	2	0.39	0.39	3.14
	GBS0317	*Hha*I	2	0.50	0.50	9.42
	GBS0573	*Dde* I	2	0.21	0.21	0.90
	GBS0591	PS	2	0.36	0.36	6.73
Unassigned	GBS0004	*Dde* I	2	0.47	0.47	4.04
	GBS0054	*Dde* I	2	0.48	0.48	1.35
	GBS0073	*Kpn* I	2	0.36	0.36	11.21
	GBS0110	*Ava* II	2	0.38	0.38	8.97
	GBS0122	*Ava* II	2	0.50	0.50	2.69
	GBS0153	*Dde* I	2	0.12	0.12	2.24
	GBS0172	*Aci* I	2	0.22	0.22	0
	GBS0344	*Hpa* II	2	0.48	0.48	8.07
	GBS0406	*Hha* I	2	0.50	0.50	6.28
	GBS0432	*Apo* I	2	0.17	0.17	2.24
	GBS0578	*Alu* I	2	0.41	0.41	2.24
	GBS0703	*Taq* I	2	0.48	0.48	1.79
	GBS0734	*Eco*R V	2	0.29	0.29	4.04

[a] SNP assays optimized for CAPS are represented by the name of restriction enzyme, *indel* assays have been represented by IN_DEL and Pyrosequencing assays have been designated as PS; genotyping for more than one SNP for a given marker in PS assay are represented as PS_pos1, PS_pos2, etc

lowest PIC value and smaller number of alleles were recorded for markers from the linkage group 5H (average PIC = 0.61, average alleles = 6.67) and 6H (average PIC = 0.60, average alleles = 6.71).

Heterogeneity detected by SSR and SNP markers

A varying level of heterogeneity was detected at 60 SNP and 29 SSR loci, while the remaining 11 SNP and 16 SSR loci did not show any heterogeneity in a single accession (Tables 2, 3; Fig. 1). Heterogeneity was observed in the range of 0. 5–23.7% with an average of 5.7% accessions at SNP loci. In case of SSR loci, heterogeneity was recorded on average in 12.4% accessions (range 0.5–89.2%). While a maximum of two SNP alleles were observed in a heterogeneous accession, up to 4 SSR alleles were observed in heterogeneous cases. For example, the SSR locus GBM1059, revealing highest level of heterogeneity, showed two alleles in 175 accessions, three alleles in 19 accessions and four alleles in 5 accessions. Thus a total of 199 (89.2%) accessions were found heterogeneous for the GBM1059 SSR locus. However, in case of SNP loci, the maximum

Table 3 Genetic diversity and heterogeneity in germplasm collection based on SSR markers

Linkage group	Marker name	SSR motif	Number of alleles	PIC value	Heterogeneity (% accessions)
1HL	GBM1013	(CTG)9	6	0.55	13.00
1HL	GBM1461	(CA)6···(CA)18	17	0.88	2.24
1HL	GBM1002	(CCT)7	11	0.61	8.97
1HS	GBM1029	(AG)10	7	0.56	2.24
1HS	GBM1007	(AC)11	14	0.74	0.45
1HL	GBM1334	(GGC)8	4	0.47	0
1HL	GBM1061	(GGT)6	6	0.61	0
2HL	GBM1047	(AGC)5	7	0.69	5.38
2HS	GBM1459	(AC)7	11	0.69	7.17
2HL	GBM1208	(AG)6	7	0.72	36.32
2HS	GBM1035	(CT)8	5	0.67	0.90
2HL	GBM1036	(CT)8	6	0.50	0
3HS	GBM1031	(AG)15	7	0.65	2.24
3HL	GBM1405	(CGCA)5	5	0.76	7.62
3HL	GBM1110	(AAG)6	12	0.67	11.21
3HL	GBM1059	(GGT)5	10	0.79	89.24
3HL	GBM1043	(AAC)5	4	0.34	0
3HS	GBM1413	(TCATA)6	6	0.64	0
4HL	GBM1003	(CTT)8	11	0.72	19.28
4HL	GBM1015	(ACAT)13	22	0.90	17.04
4HS	GBM1221	(AC)10	15	0.78	7.17
4HS	GBM1323	(GCC)8	7	0.75	7.17
4HL	GBM1020	(AC)7	3	0.53	0
4HL	GBM1018	(CCG)6	6	0.62	0
4HS	GBM1501	(TAGA)6	5	0.54	0
5HL	GBM1483	(GCG)7	14	0.81	17.49
5HL	GBM1054	(CCG)5	7	0.73	14.80
5HL	GBM1064	(AGGG)5	5	0.67	8.07
5HL	GBM1363	(AGG)7	4	0.52	0
5HL	GBM1026	(AC)9	4	0.27	0
5HS	GBM1176	(AT)8	6	0.67	0
6HS	GBM1021	(AC)8	13	0.83	12.56
6HL	GBM1256	(GA)8	7	0.63	6.28
6HS	GBM1212	(AGG)5	4	0.61	7.17
6HL	GBM1008	(AAC)10	7	0.61	9.42
6HS	GBM1075	(GT)6	7	0.60	0
6HL	GBM1404	(TATG)5	3	0.17	0
6HL	GBM1063	(ACAT)7	6	0.73	0
7HS	GBM1033	(AT)9	8	0.66	6.73
7HL	GBM1419	(CTCAT)5	6	0.53	8.52
7HS	GBM1464	(CAG)8···(CAG)5	17	0.85	13.45
7HS	GBM1516	(CT)9	8	0.79	12.56
7HS	GBM1060	(GGT)6	5	0.48	0

Table 3 continued

Linkage group	Marker name	SSR motif	Number of alleles	PIC value	Heterogeneity (% accessions)
7HL	GBM1456	(CGG)6	3	0.26	0
7HS	GBM1326	(CTT)8	8	0.70	5.38

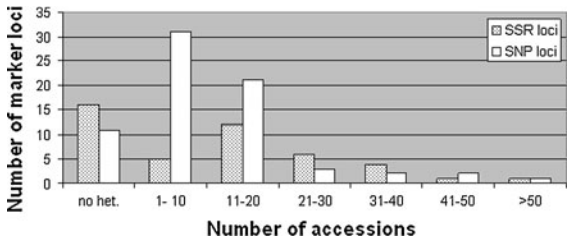


Fig. 1 Structure of heterogeneity in 223 barley accessions. The columns represent heterogeneity in 223 barley accessions detected by SSR (*dotted bars*) and SNP (*white bars*) loci. Number of accessions showing heterogeneity have been grouped in classes like 1–10, 11–20, etc

heterogeneity was recorded for the SNP locus GBS0576 PS_pos3 that revealed both SNP alleles in 53 (23.77%) accessions.

Genetic relationships based on cluster analysis

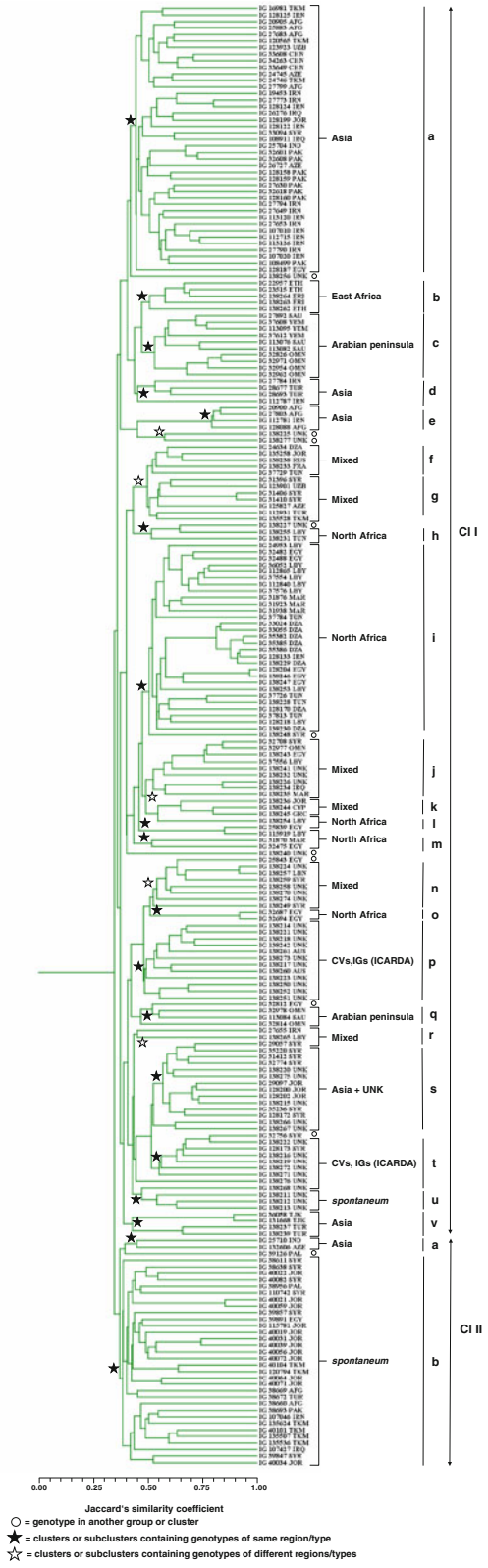
Allelic data for SNP (143 alleles) and SSR (365 alleles) were used individually for preparation of genetic similarity (GS) coefficient matrices for the 223 genotypes. SSR- and SNP- based GS matrices were highly correlated ($r = 0.75$) and highly significant ($P < 0.005$, 2.575, $g = 112.11$) after performing Mantel test using 1,000 iterations. Subsequently, the allelic data for both marker systems were combined and a dendrogram derived from UPGMA cluster analysis (UPGMA) based on the GS coefficient matrix (of 499 alleles) for 223 genotypes (Fig. 2, also available at http://pgrc.ipk-gatersleben.de/drought/). Basically all genotypes could be distinguished. The GS coefficient for all genotypes ranged from 0.263 to 0.832 with an average of 0.385.

Because the germplasm collection included a set of 38 *H. spontaneum* accessions together with 175 *H. vulgare* accessions, two separate clusters i.e. 'Cl I' and 'Cl II' containing majority of *vulgare* (174) and *spontaneum* (32) genotypes were obtained (Fig. 2). However, 1 *vulgare* and 6 *spontaneum* genotypes were grouped in clusters 'Cl II' and 'Cl I', respectively. Under the cluster 'Cl I', several sub-clusters containing genotypes from the same geographic regions were observed. For example, the sub-clusters 'Cl Ia', 'Cl Id', 'Cl Ie', 'Cl Is' and 'Cl Iv' contained genotypes originating from Asia, the sub-clusters 'Cl Ib', 'Cl Ih', 'Cl Ii', 'Cl Il', 'Cl Im' and 'Cl Io' contained genotypes coming from Africa and the sub-cluster 'Cl 1c' and 'Cl 1q' contained genotypes coming from Arabian Peninsula. However, some sub-clusters like 'Cl If', 'Cl Ig', 'Cl Ij', 'Cl Ik', 'Cl In' and 'Cl Ir' contained the genotypes from different geographic regions (Africa, Asia, Europe) together.

Trends in genetic diversity

A comparison of the genetic diversity of barley genotypes was performed after considering six germplasm pools (as per geographic regions), namely Africa (AFR), Middle East Asia (MEA), North East Asia (NEA), Arabian Peninsula (APS), Australia (AUS) and Europe (EUR). The genotypes, for which the country of origin was not known, were grouped together and the pool was designated as Unknown (UNK).

The number of total alleles, alleles per locus, the number of region specific alleles, the number of accessions carrying rare alleles, the countries representing carrying rare and the mean genetic similarity within a region (germplasm pool) were used to assess trend and compare genetic diversity among the regions (Table 4). The accessions from MEA and NEA revealed more diversity compared to other regions. The number of alleles per locus was 6.4 and 5.8 for MEA and NEA, respectively and higher than that of the other regions. In the similar way, effective number of alleles, allelic richness, the number of rare alleles and the number of genotypes carrying rare alleles were higher in MEA and NEA than in the other regions (Table 4). Since the AUS and EUR groups contain only 2 and 3 genotypes, respectively, no rare allele was observed in the genotypes



◀**Fig. 2** UPGMA clustering dendrogram of 223 accessions based on genetic similarity coefficients obtained for 71 SNP and 45 SSR loci (also available at http://pgrc.ipk-gatersleben.de/drought)

representing these groups or continents. Majority of countries representing AFR, MEA, NEA and APA groups included the genotypes that contained rare alleles. However, Eritrea and Morocco of AFR group; Tajikistan, Uzbekistan and India of NEA group; Lebanon of MEA group and Oman of APS groups did not contribute any genotype carrying rare alleles. Further, the NEA and MEA accessions had a relatively low mean genetic similarity (GS) with values of 0.408 and 0.416, respectively. All the above results suggested that the NEA and MEA regions exhibited greater genetic diversity than other regions.

In addition, the comparison of genetic diversity was made after classifying genotypes in *spontaneum* (SPON), *vulgare*- landraces (VUL-LR) and *vulgare*-cultivars or improved genotypes (VUL-IG). A higher number of total alleles, higher number of alleles per locus, higher number of effective alleles and higher allelic richness was observed in SPON genotypes as compared to other genotypes (Table 5). However, relatively higher numbers of rare alleles as well as higher numbers of genotypes containing rare alleles were observed in VUL-LR group. Similarly, a lower genetic similarity (GS = 0.399) was observed in VUL-LR group as compared to other two groups.

Discussion

Marker polymorphism

For understanding the genetic diversity in the ICARDA germplasm collection, a total of 68 SNP (71 datapoints) and 45 SSR markers were analyzed. Generally the SNPs are biallelic and possess the maximum PIC value 0.50. In the present study, 11 markers showed the PIC value 0.50 while the *indel* based marker GBS0692 displayed PIC value 0.55 as the marker yielded three alleles. In general, the SNP markers examined here showed a moderate PIC value. Only 10 (14.1%) SNPs showed the PIC value less than 0.25. These results suggest either the SNPs

selected are highly polymorphic and/or the germplasm assayed in the present study is quite diverse.

As SSR markers are multiallelic, on an average 7.9 alleles per markers and the average PIC value 0.65 was observed for the markers tested. These SSR markers are derived from ESTs and therefore they are supposed to yield relatively lower number of alleles and lower PIC values as compared to genomic SSR studies. However, the reported allele numbers and PIC values for the markers are comparable to several germplasm diversity studies employing the genomic SSRs (Russell et al. 2003; Pandey et al. 2006; Malysheva-Otto et al. 2006). This may be explained as (a) SSR markers used here were already selected for higher polymorphism and part of the core set of highly polymorphic markers (Varshney et al. 2008), (b) the germplasm assayed here are quite diverse as it represents 30 countries of five continents.

It is also interesting to note that in case of SNP as well as SSR markers, the markers coming from the linkage group 4H, as compared from other linkage groups, showed higher PIC value and higher number of alleles. It seems that the markers of linkage group 4H are highly diverse, however a conclusive reason is presently not known.

Higher level of heterogeneity

A varying level of heterogeneity was detected at 84.51% SNP and 64.44% SSR loci in the range of 0.45–89.24% accessions of the germplasm collection. While higher number of SNP loci as compared to SSR loci detected heterogeneity, SSR loci, in general, detected heterogeneity in higher number of accessions as compared to that of SNP loci. These results indicate that probably SSR markers, because of multi-allelic nature, are the preferred markers for detecting heterogeneity. Nevertheless, SNP markers also detect heterogeneity. Indeed observations on heterogenetity are not surprising in the present study, as majority of the genotypes examined represent landraces and not pure lines. Furthermore, SNP and SSR markers analyzed in the present study detected single locus in earlier studies (Varshney et al. 2008; Kota et al. 2008), the observed heterogeneity in the present study can be attributed to bulking DNA from 10 seedlings that most likely are not pure especially in case of 'spontaneum' and 'landrace' accessions. Up to 10.4% accessions showed heterogeneity at genomic SSR loci in the barley germplasm collection in a recent study (Malysheva-Otto et al. 2006). In fact, majority of earlier diversity studies conducted in barley or many self-pollinated species did not care for heterogeneity present in the seed lots of at least wild species and landraces. Several studies used DNA from single plant from a seed lot by assuming all the seeds from a seed lot homogenous. However, present study together with some other studies (e.g. Malysheva-Otto et al. 2006) clearly indicated that a good proportion of germplasm lines have heterogeneity in the seed lots. In fact, a recent study conducted on genetic diversity study in 2,915 accessions of chickpea (Upadhyaya et al. 2008), a self-pollinated species like barley, with 48 SSR markers has shown that the best way to deal with heterogeneity in the seed lots of genebank accessions is to isolate DNA from one seed of each accession and then undertake molecular profiling as well as seed multiplication from the same plant. Subsequently seeds coming from the same plant may be maintained and distributed to the community.

Genetic relationships in germplasm

Based on GS, the cluster analysis classified 99.4% *vulgare* accessions in 'Cl I' while 84.2% *spontaneum* accessions in the cluster 'Cl II'. However, 1 *vulgare* and 6 *spontaneum* genotypes were grouped in clusters 'Cl II' and 'Cl I', respectively. The cluster 'Cl I' could be classified further in several sub-clusters containing genotypes from the same geographic regions. However, some subclusters (like 'Cl If', 'Cl Ig', 'Cl Ij', 'Cl Ik', 'Cl In' and 'Cl Ir) contained the genotypes from different geographic regions. These facts indicate that the genetic diversity of barley is not completely related to geographic distribution. This may be attributed to: (a) occurrence of similar genetic variation independently in the different geographic regions, or (b) artificial transfer of the accessions from one region to another region (Huang et al. 2002).

Comparison of genetic diversity in geographic regions and gene pools

While comparing the genetic diversity as per geographical regions, both, the total alleles as well as unique alleles were higher in the accessions

Table 4 SSR diversity features as per geographic regions of the origin of accessions

	Africa (AFR)	Middle East Asia (MEA)	North East Asia (NEA)	Arabian Peninsula (APS)	Australia (AUS)	Europe (EUR)	Unknown (UNK)
Number of accessions	52	47	68	14	2	3	37
Number of total alleles	222	287	263	146	66	87	212
Number of alleles per locus	4.9	6.4	5.8	3.2	1.5	1.9	4.7
Effective number of alleles	1.84	2.13	1.93	1.55	–	–	1.60
Allelic richness	2.17	2.47	2.40	2.05	–	–	2.22
Number of specific alleles	8	32	20	1	–	–	12
Number of accessions carrying rare alleles	9	24	31	3	–	–	15
Countries of origin of accessions carrying rare alleles	EGY (3), ETH (1), DZA (2), LBY (1), TUN (2)	IRQ (2), JOR (15), PAL (1), SYR (6)	AFG (3), AZE (2), CHN (1), IRN (12), PAK (5), TKM (6), TUR (2)	SAU (1), YEM (2)	–	–	–
Mean GS within the region	0.460	0.416	0.408	0.560	0.850	0.803	0.451

representing Middle East Asia (MEA) and North East Asia (NEA). Lowest genetic similarity (GS) coefficient as compared to other groups reconfirmed higher diversity present in MEA and NEA regions. Such observations were made by Malysheva-Otto et al. (2006) while conducting genetic diversity study in 953 barley accessions with 48 SSR markers. Not only in barley, even in a genetic diversity study on wheat, the accessions from Near East and Middle East exhibited more genetic diversity than those from the other six regions (Huang et al. 2002). These results are expected as these areas have been designated as centre of origin and diversification in several reports earlier (see Pozzi et al. 2004). Although we did not have accessions representing Nepal and Himalayan regions, the observed genetic diversity in MEA (average 0.408) and NEA (average GS 0.416) is higher than the Nepalese and Himalayan regions (average GS 0.500, Pandey et al. 2006).

In terms of the unique alleles, genotypes from Jordan provided highest number of unique alleles. The high genetic diversity in barley genotypes from Eritrea and Ethiopia reported recently by Orabi et al. (2007), were not confirmed in the present study. Nevertheless in the present study, Eritrea and Ethiopia were represented by only 1 and 3 genotypes, respectively. In fact, Bjornstad et al. (1997) reported less diversity in Ethiopian germplasm as compared to that of Europe, North America and Japan.

On comparing different genepools, the total number of alleles as well as the number of alleles per locus was higher in the genotypes of *spontaneum* (SPON). This clearly supports the earlier observations and conclusions on higher genetic diversity in *spontaneum* genotypes (Saghai Maroof et al. 1994; Russell et al. 2004). Nevertheless, the landraces from *vulgare* (VUL-LR) showed relatively more unique alleles (36) as compared to genotypes from SPON (28) and VUL-IG (8) groups. As compared to SPON group, three times higher genotypes of the VUL-LR group yielded rare alleles. Similarly, a lower genetic similarity (GS = 0.399) was observed in VUL-LR group as compared to other two groups. However it is important to note the difference in the numbers of genotypes representing three groups. Higher number of genotypes coming from relatively higher number of countries in case of VUL-LR, as compared to SPON group may be another reason for showing

Table 5 Diversity features in three different genepools of barley

	Number of accessions	Number of total alleles	Number of alleles per locus	Effective number of alleles	Allelic richness	Number of group specific alleles	Number of accessions carrying rare alleles	Countries of origin of genotypes carrying rare alleles	Mean GS within a group
spontaneum (SPON)	38	302	6.7	4.63	1.70	28	29	EGY (1), AFG (2), PAK (1), TKM (5), IRN (2),IRQ (1), JOR (11), SYR (5), PAL (1)	0.433
vulgare-Landraces (VUL-LR))	163	284	6.3	4.51	1.60	36	93	DZA (2), EGY (5), ERI (1), ETH (3), LBY (5), MAR (3), RUS (1), AFG (7), AZE (3), CHN (3), IND (2), PAK (7), UZB (1), TJK (2), TKM (3), TUN (3), TUR (3), IRN (14), SAU (3), YME (2), OMN (4), IRQ (2), SYR (7), UNK (7)	0.399
vulgare-Cultivars or improved genotypes (VUL-IG)	22	186	4.1	3.65	1.51	6	7	Cultivars	0.475

higher genetic diversity in VUL-LR group. Taking all the analyses together, it can be concluded that the genotypes showing specific alleles to the regions and unique alleles to the genotypes may prove very useful for introgressing the resistance traits to biotic stresses and tolerance to abiotic stresses in elite barley breeding programme (Saghai Maroof et al. 1994; Russell et al. 2004; Pandey et al. 2006).

In summary, the gene-based SNP and SSR markers showed a high level of genetic diversity in the germplasm collection that can be harnessed for crop improvement.

For example, as germplasm collections represent a wide range of drought scenarios and diverse selection regimes, association mapping can be used for mapping drought tolerance traits. It is, however, important to note that variation in photoperiod and maturity responses and a wide range of other developmental and morphological traits in germplasm lines may pose the constraints in finding meaningful marker trait associations.

Acknowledgments We are thankful to U. Beier for her excellent technical assistance. This work was financially supported by a GTZ project (no. 2002.7860.6-001.00 and Contract no. 1060503) sponsored by BMZ (Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung), Germany.

References

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 11:554–560

Baek HJ, Beharav A, Nevo E (2003) Ecological-genomic diversity of microsatellites in wild barley *Hordeum spontaneum*, populations in Jordan. Theor Appl Genet 106:397–410

Bjornstad A, Demissie A, Kilian A, Kleinhofs A (1997) The distinctness and diversity of Ethiopian barleys. Theor Appl Genet 94:514–521

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics 172:557–567

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs SatoK, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, Young JD, Marshall DF, Madishetty KM, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582

Eujayl I, Sorrels ME, Baum M, Wolters P, Powel W (2001) Isolation of EST-derived microsatellites markers for

genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407
Goudet J (2002) http://www2.unil.ch/popgen/softwares/fstat.htm (last accessed on 3 June 2009)
Gupta PK, Varshney RK (2000) The development and use of microsatellite arkers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 13:163–185
Gusami F, Touil L, Feres K, Elfelah W, Triki T, Ferchichi A (2008) Genetic diversity of Tunisian barley accessions based on microsatellite markers. Biotechnology 7:781–786
Huang XQ, Börner A, Röder MS, Ganal MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor Appl Genet 105:699–707
Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T (2002) Discovery and assay of single nucleotide polymorphsims in barley (*Hordeum vulgare*). Plant Mol Biol 48:529–537
Karakousis A, Gustafson JP, Chalmers KJ, Barr AR, Langridge P (2003) A consensus map of barley integrating SSR, RFLP, and AFLP markers. Aust J Agric Res 54:1173–1185
Koebner RM, Donini P, Reeves JC, Cooke RJ, Law JR (2003) Temporal flux in the morphological and molecular diversity of UK barley. Theor Appl Genet 106:550–558
Kolodinska AB, Bothmer RV, Dayteg C, Rashal I, Tuvesson S, Weibull J (2007) Genetic diversity changes and relationships in spring barley (*Hordeum vulgare* L.) germplasm of Nordic and Baltic areas as shown by SSR markers. Genet Crop Res Eval 54:749–758
Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A (2001) Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas 135:145–151
Kota R, Varshney RK, Prasad M, Zhang H, Stein N, Graner A (2008) EST-derived single nucleotide polymorphism (SNP) markers for assembling genetic and physical maps of the barley genome. Funct Integr Genom 8:223–233
Kraakman ATW, Niks RE, van den Berg PMMM, Stam P, van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446
Malysheva-Otto LV, Ganal MW, Röder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genet 7:6
Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220
Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. Genome 45:1095–1106
Morgante M, Rafalski A, Biddle P, Tingey S, Oliveri AM (1994) Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome 37:763–769
Morrell PL, Toleno DM, Lundy KE, Clegg MT (2005) Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilizaton. Proc Natl Acad Sci USA 102:2442–2447
Orabi J, Backes G, Wolday A, Yahyaoui A, Jahoor A (2007) The Horn of Africa as a centre of barley diversification and a potential domestication site. Theor Appl Genet 114:1117–1127
Pandey M, Wagner C, Friedt W, Ordon F (2006) Genetic relatedness and population differentiation of Himalayan hulless barley (*Hordeum vulgare* L.) landraces inferred with SSRs. Theor Appl Genet 113:715–729
Pozzi C, Rossini L, Vecchietti A, Salamini F (2004) Genes and genome changes during domestication of cereals. In: Gupta PK, Varshney RK (eds) Cereal genomics. Kluwer, Dordrecht, pp 165–198
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100
Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson JT, Wanamaker SI, Walia H, Rodriguez EM, Hedley PE, Liu H, Morris J, Close TJ, Marshall DF, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Gen Genom 274(5):515–527
Russell JR, Booth A, Fuller JD, Baum M, Ceccarelli S, Grando S, Powell W (2003) Patterns of polymorphism detected in the chloroplast and nuclear genomes of barley landraces sampled from Syria and Jordan. Theor Appl Genet 107:413–421
Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G, Powell W (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47:389–398
Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91:5466–5470
Stracke S, Presterl T, Stein N, Perovic D, Ordon F, Graner A (2007) Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding the *Rym4* locus for *Bymovirus* resistance in barley. Genetics 175:805–817
Takeda S, Matsuoka M (2008) Genetic approaches to crop improvement: responding to environmental and population changes. Nat Rev Genet 9:444–457
Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422
Thiel T, Kota R, Grosse I, Stein N, Graner A (2004) SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. Nucl Acids Res 32:1 e5
Upadhyaya HD, Dwivedi SL, Baum M, Varshney RK, Udupa SM, Gowda CLL, Hoisington DA, Singh S (2008) Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). BMC Plant Biol 8:106
Varshney RK, Graner A, Sorrells ME (2005a) Genic microsatellite markers: features and applications. Trends Biotechnol 23:48–55
Varshney RK, Prasad M, Graner A (2005b) Molecular marker map of barley: a resource for intra- and interspecific genomics. In: Lorz H, Wenzel G (eds) Molecular Markers Systems in Plant Breeding and Crop Improvement,

Biotechnology in Agriculture and Forestry, vol 55. Springer, Dordrecht, pp 229–243

Varshney RK, Grosse I, Hähnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A (2006) Genetic mapping and BAC assignment of EST-derived SSR markers proves non-uniform distribution of genes in the barley genome. Theor Appl Genet 113:239–250

Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A (2007a) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Sci 173:638–649

Varshney RK, Mahender T, Aggrawal RK, Börner A (2007b) Genic molecular markers in plants: development and applications. In: Varshney RK, Tuberosa R (eds) Genomics assisted crop improvement, vol I: genomics approaches, platforms. Springer, Dordrecht, pp 13–30

Varshney RK, Thiel T, Sretenovic-Rajicic T, Baum M, Valkoun J, Guo P, Grando S, Ceccarelli S, Graner A (2008) Identification and validation of a core set of informative genic SSR and SNP markers for assaying functional diversity in barley. Mole Breed 22:1–13