Theor Appl Genet (2005)
DOI 10.1007/s00122-005-0061-9

ORIGINAL PAPER

O. Dobrovolskaya · U. Saleh · L. Malysheva-Otto
M.S. Röder · A. Börner

Rationalising germplasm collections: a case study for wheat

Received: 26 March 2005 / Accepted: 14 July 2005

Abstract In total 70 genebank accessions comprising 50 hexaploid, 12 tetraploid and 8 diploid wheats of the Gatersleben collection were selected based on the screening of the passport data for identical cultivar names or accession numbers of the donor genebanks. Twelve potential duplicate groups consisting of three to nine accessions with identical names/numbers were selected and analysed with DNA markers (microsatellites). A bootstrap approach based on re-sampling of both microsatellite markers and alleles within marker loci was used to test for homogeneity. Although several homogeneous groups were identified it became clear that cultivar name identity alone did not allow the determination of duplicates. A combination of SSR-analysis followed by the bootstrap method and database survey considering the botanical classification and other data (origin, growth habit and donor) available is recommended in order to determine duplicates. A procedure for the identification of duplicates and their further handling in *ex situ* genebanks is discussed.

Keywords Duplicates · *ex situ* genebank collections · Microsatellites · Wheat · *Triticum* spp.

Communicated by H. Becker

O. Dobrovolskaya · U. Saleh · L. Malysheva-Otto
M.S. Röder · A. Börner (✉)
Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstraße 3, 06466 Gatersleben, Germany

O. Dobrovolskaya
Siberian Branch of the Russian Academy of Sciences, Institute of Cytology and Genetics, Lavrentyeva Ave. 10, 630090 Novosibirsk, Russia

U. Saleh
Regional Agricultural Research Station, 8211 Rahmatpur, Barisal, Bangladesh

Introduction

Based on the FAO statement (1998) it is estimated that, worldwide, existing *ex situ* collections contain approximately 6 million accessions of plant genetic resources. This number may have been increased even further by now. There exists, however, duplication within and between collections. The number of unique accessions within global collections is, therefore, estimated to be 1–2 million (FAO 1998). In order to improve the efficiency of conservation there is a need to rationalise the collections by identifying and minimising unnecessary duplications. Whereas an elimination of duplicates between collections needs international rules and is often hampered by national interests, the identification and reduction of duplications within collections can be done individually.

The first step in the identification of probable duplicates is based on the available passport data (Hintum and Knüpffer 1995). For this purpose reliable documentation is necessary. However, accessions with identical passport data are not necessarily duplicates as was demonstrated already for rice (Sahu 1989), barley (Hintum and Knüpffer 1995; Hintum and Visser 1995) and durum wheat (Ruiz and Aguiriano 2004). Further investigations are necessary by obtaining and employing agro-morphological, protein, isozyme and/or DNA data.

Nearly 50% of all genebank holdings are cereals, of which wheat (*Triticum*), with nearly 800,000 accessions, is the largest genus (FAO 1998). Correspondingly, in the German *ex situ* genebank in Gatersleben, wheat is the leading crop. In total about 28,000 wheat accessions are preserved after fusion with the collection formerly maintained at the Federal Centre for Breeding Research on Cultivated Plants in Braunschweig (Anonymous 2005). In order to identify probable duplicates the fields 'Cultivar/Accession name' and 'Accession number donor genebank' of the passport database were screened for identical names or numbers. Several groups

containing three to nine accessions with identical names/numbers were selected and analysed with DNA markers (microsatellites). For further data analysis we applied the bootstrap approach based on re-sampling of microsatellite data, which was already successfully used for the identification of potential duplicates in spring barley (Lund et al. 2003). The aim of the study was to prove whether this approach is generally applicable for rationalising genebank collections.

Materials and methods

Plant materials

Fifty, twelve and eight accessions of hexaploid, tetraploid and diploid wheats, respectively, maintained in the genebank of the IPK Gatersleben were selected for analysis (Tables 1 and 2). Thirty-eight hexaploid wheats (*Triticum aestivum* L.) formed six groups of potential duplicates having identical cultivar names and containing 3–9 accessions (Table 1). One duplicate group having the cultivar name 'Vernal' represents 7 accessions of the tetraploid wheat *T. dicoccon* Schrank. Three groups ('Autonomia', 'Risciola' and 'Regina') consist of both hexaploid (*T. aestivum*) and tetraploid (*T. durum*) accessions. Two duplicate groups containing three and five diploid accessions were identified by Hammer et al. (1998). This material was described to have identical accession numbers of the donor genebank but was introduced several times (Table 2). Regarding the numbers of the donor genebank, which was St. Petersburg, Russia, the duplicate groups represent *T. baeoticum* Boiss. em. Schiem. (three accessions of which one was classified as *T. urartu* Thum. ex Gandil at IPK) and *T. sinskajae* Filat et. Kurk. (five accessions). The duplicate groups will be designated regarding their cultivar (hexaploid and tetraploid wheats) or species (diploid wheat) names.

DNA extraction and microsatellite analysis

Total genomic DNA was isolated from pooled leaf material of five to eight plants of each accession according to Plaschke et al. (1995). Twenty-eight Gatersleben wheat microsatellites (GWM) detecting 31 loci (Table 3) were selected for the examination of the hexaploid accessions (Röder et al. 1998). At least one marker was present for each chromosome. The tetraploid accessions were investigated together with the hexaploid ones considering the A and B genome markers only. For the diploid wheat species, 18 GWM representing all chromosomes of the A genome (Röder et al. 1998) were chosen (Table 4). Microsatellites GWM 1099, GWM 1217 and GWM 1293 were kindly provided by Dr. M. Ganal, Trait Genetics GmbH, Gatersleben, Germany.

The PCR analysis and fragment detection were performed as described by Röder et al. (1998). Fragment analysis was carried out in an ALF (automated laser fluorescence)-express sequencer (Amersham-Biosciences, Freiburg, Germany), and fragment sizes were calculated using the computer program Fragment Analyser Version 1.02 (Amersham-Biosciences), by comparison with internal size standards. The varieties 'Chinese Spring' and 'Aztec' were used as controls in each run to ensure size accuracy and to minimise run-to-run or gel-to-gel variation.

Data analysis

Fragments amplified by microsatellite primers were scored for their size in base pairs. Genetic dissimilarity (distance) coefficients were calculated for each pair of accessions in groups of potential duplicates using the program 'Diversity' kindly provided by Dr. S. B. Andersen (The Royal Veterinary and Agricultural University, Copenhagen, Denmark). In case of intra-accession heterogeneity, more than one allele for a microsatellite marker was detected. Similarity between two accessions within one locus was calculated as the number of common alleles relative to total alleles in the locus observed for the two accessions. Similarity between two accessions is the average similarity over all loci and the coefficient of dissimilarity between them is one minus similarity. Average genetic distances for each pair of accessions in all groups of potential duplicates were calculated. The dissimilarity matrix was subjected to cluster analysis using the UPGMA (unweighted pair-group method with arithmetic mean) algorithm on MEGA Version 3.0 (Kumar et al. 2004). The average gene diversity value (polymorphism information content) of markers was calculated according to the formula of Nei (1973):

$$\mathrm{PIC} = 1 - \sum P_{ij}^2$$

where P_{ij} is the frequency of the jth allele for the ith locus summed up across all alleles for the locus.

Testing of genotypes within potential duplicate groups for homogeneity was performed with the bootstrap approach (Felsenstein 1985) using the program 'Diversity' as described by Lund et al. (2003). Marker loci and alleles for each accession at each locus were sampled randomly with replacement. The number of pseudo-replicate samples was equal to 1000. Maximum average genetic distance from each pseudo data set generated during the re-sampling process was used as a statistic test. The accessions with the maximum average genetic distance in the original data set were determined and the probability for the largest value of maximum average genetic distance within the group (P-value) was calculated from this re-sampled distribution. Subsequently, accessions with the statistically largest average genetic distances were removed and the procedure iter-

Table 1 Duplicate groups of hexaploid and tetraploid (*) wheat cultivars, the genebank catalogue numbers, species, morphological groups, countries of origin, growth habits (*W* winter type, *S* spring type) and donors, extracted from the Gatersleben wheat database

Cultivar/ line	Catalogue no.	Species	Morphological group	Origin	Growth habit	Donor
Ceres	TRI 395	*T. aestivum*	var. *aestivum*	USA	S	Plant Breeding Station Halle, Germany
Ceres	TRI 411	*T. aestivum*	var. *aestivum*	USA	S	Genebank Beltsville, MD, USA
Ceres	TRI 412	*T. aestivum*	var. *aestivum*	USA	S	Genebank Beltsville, MD, USA
Ceres	TRI 414	*T. aestivum*	var. *aestivum*	USA	S	Genebank Beltsville, MD, USA
Ceres	TRI 417	*T. aestivum*	var. *aestivum*	USA	S	Genebank Beltsville, MD, USA
Ceres	TRI 418	*T. aestivum*	var. *aestivum*	USA	S	Genebank Beltsville, MD, USA
Ceres	TRI 12914	*T. aestivum*	var. *aestivum*	Unknown	W	Research Institute Fundulea, Romania
Ceres	TRI 15976	*T. aestivum*	var. *aestivum*	Unknown	S	Genebank Beltsville, MD, USA
Ceres	TRI 28629	*T. aestivum*	var. *erythrospermum*	USA	S	Genebank Braunschweig, Germany
Autonomia	TRI 7113	*T. aestivum*	var. *lutescens*	Italy	S/W	Botanical Garden, University of Ferrara, Italy
Autonomia	TRI 13051	*T. aestivum*	var. *lutescens*	Italy	W	Collection mission Genebank Gatersleben, 1980
Autonomia	TRI 14051	*T. aestivum*	var. *milturum*	Italy	W	Collection mission Genebank Gatersleben, 1982
Autonomia	TRI 14062	*T. aestivum*	var. *lutescens*	Italy	W	Collection mission Genebank Gatersleben, 1982
Autonomia	TRI 14139	*T. durum**	var. *leucurum*	Italy	S	Collection mission Genebank Gatersleben, 1982
Autonomia	TRI 14225	*T. aestivum*	var. *aestivum*	Italy	W	Collection mission Genebank Gatersleben, 1982
Autonomia	TRI 16504	*T. aestivum*	var. *aestivum*	Italy	W	Collection mission Genebank Gatersleben, 1987
Autonomia	TRI 16505	*T. aestivum*	var. *lutescens*	Italy	W	Collection mission Genebank Gatersleben, 1987
Risciola	TRI 13053	*T. aestivum*	var. *aestivum*	Italy	W	Collection mission Genebank Gatersleben, 1980
Risciola	TRI 13071	*T. aestivum*	not determined	Italy	W	Collection mission Genebank Gatersleben, 1980
Risciola	TRI 14063	*T. aestivum*	var. *aestivum*	Italy	W	Collection mission Genebank Gatersleben, 1982
Risciola	TRI 14067	*T. aestivum*	var. *aestivum*	Italy	W	Collection mission Genebank Gatersleben, 1982
Risciola	TRI 14155	*T. durum**	Not determined	Italy	S	Collection mission Genebank Gatersleben, 1982
Risciola	TRI 16459	*T. durum**	Not determined	Italy	W	Collection mission Genebank Gatersleben, 1987
Risciola	TRI 16460	*T. aestivum*	var. *ferrugineum*	Italy	W	Collection mission Genebank Gatersleben, 1987
Risciola	TRI 16461	*T. aestivum*	var. *hostianum*	Italy	W	Collection mission Genebank Gatersleben, 1987
Gabo	TRI 7244	*T. aestivum*	var. *aureum*	Australia	S	C.S.I.R.O., Canberra, Australia
Gabo	TRI 9684	*T. aestivum*	var. *aureum*	Australia	S	INIA Londres, Mexico
Gabo	TRI 28627	*T. aestivum*	var. *aureum*	Australia	S	Genebank Braunschweig, Germany
Gabo	TRI 28622	*T. aestivum*	var. *ferrugineum*	Unknown	S	Genebank Braunschweig, Germany
Gabo	TRI 28630	*T. aestivum*	var. *albidum*	Unknown	W	Genebank Braunschweig, Germany
Gabo	TRI 28628	*T. aestivum*	var. *albidum*	Australia	S	Genebank Braunschweig, Germany
Gabo	TRI 28631	*T. aestivum*	var. *albidum*	Australia	S	Genebank Braunschweig, Germany
Vernal	TRI 445	*T. dicoccon**	var. *dicoccon*	USA	S	Plant Breeding Station Halle, Germany
Vernal	TRI 9525	*T. dicoccon**	var. *dicoccon*	USA	S	Institute of Phytopathology, Halle, Germany
Vernal	TRI 18200	*T. dicoccon**	var. *volgense*	USA	S	VIR, St. Petersburg, Russia
Vernal	TRI 18211	*T. dicoccon**	var. *volgense*	USA	S	VIR, St. Petersburg, Russia
Vernal	TRI 28616	*T. dicoccon**	Not determined	Unknown	S	Genebank Braunschweig, Germany
Vernal	TRI 28615	*T. dicoccon**	Not determined	Unknown	S	Genebank Braunschweig, Germany
Vernal	TRI 28611	*T. dicoccon**	var. *farrum*	USA	S	Genebank Braunschweig, Germany
Penjamo 62	TRI 7543	*T. aestivum*	var. *aestivum*	Mexico	S	INIA Londres, Mexico
Penjamo 62	TRI 9713	*T. aestivum*	var. *aestivum*	Mexico	S	INIA Londres, Mexico
Penjamo 62	TRI 11454	*T. aestivum*	var. *aestivum*	Unknown	S	Dr. E. Bosca, Kompolt, Hungary
Penjamo 62	TRI 28623	*T. aestivum*	var. *erythrospermum*	USA	S	Genebank Braunschweig, Germany
Penjamo 62	TRI 28621	*T. aestivum*	Not determined	Unknown	S	Genebank Braunschweig, Germany
Penjamo 62	TRI 28626	*T. aestivum*	Not determined	Mexico	S	Genebank Braunschweig, Germany
Thatcher	TRI 1308	*T. aestivum*	var. *lutescens*	USA	W	DSG-Saatzucht Schlanstedt, Germany
Thatcher	TRI 3117	*T. aestivum*	var. *lutescens*	USA	S	Genebank Beltsville, MD, USA
Thatcher	TRI 4532	*T. aestivum*	var. *lutescens*	Canada	S	Botanical Garden Ottawa, Canada
Thatcher	TRI 5350	*T. aestivum*	var. *lutescens*	USA	S	Genebank Beltsville, MD, USA
Thatcher	TRI 9710	*T. aestivum*	var. *lutescens*	USA	S	INIA Londres, Mexico
Thatcher	TRI 12688	*T. aestivum*	var. *lutescens*	Unknown	S	Wheat Pool Saskatoon, Canada
Orlando	TRI 9322	*T. aestivum*	var. *lutescens*	GDR (Germany)	W	VEG-Saatz. Leutewiz
Orlando	TRI 11162	*T. aestivum*	var. *lutescens*	GDR (Germany)	W	VEG-Saatz. Leutewiz
Orlando	TRI 28624	*T. aestivum*	Not determined	Unknown	W	Genebank Braunschweig, Germany
Orlando	TRI 29481	*T. aestivum*	Not determined	GDR (Germany)	W	ZfS Nossen, Germany
Regina	TRI 3855	*T. durum**	var. *leucurum*	Italy	S	Collection mission R. Maly, 1950
Regina	TRI 14962	*T. aestivum*	var. *lutescens*	Czechoslovakia	W	Research Institute of Crop Production, Prag-Ruzyne, Czechoslovakia
Regina	TRI 28618	*T. durum**	var. *leucurum*	Italy	S	Genebank Braunschweig, Germany
Regina	TRI 29419	*T. aestivum*	Not determined	Czechoslovakia	W	ZfS Nossen, Germany
Apollo	TRI 5244	*T. aestivum*	var. *lutescens*	The Netherlands	W	University of Agriculture Wageningen, The Netherlands
Apollo	TRI 28625	*T. aestivum*	var. *lutescens*	The Netherlands	W	Genebank Braunschweig, Germany
Apollo	TRI 28632	*T. aestivum*	Not determined	Germany	W	Genebank Braunschweig, Germany

Table 2 Duplicate groups of diploid wheats, the Gatersleben and donor genebank catalogue numbers, morphological groups, countries of origin, growth habits and donors, extracted from the Gatersleben wheat database

Donor genebank no.	Catalogue no.	Species	Morphological group	Country of origin	Growth habit	Donor
K-48993	TRI 11525	*T. sinskajae*	var. *sinskajae*	Russia	S	VIR, St. Petersburg, Russia
K-48993	TRI 12910	*T. sinskajae*	Not determined	Russia	S	VIR, St. Petersburg, Russia
K-48993	TRI 14732	*T. sinskajae*	Not determined	Russia	S	VIR, St. Petersburg, Russia via. Bot. Garden, Kisniev, Moldavia
K-48993	TRI 17415	*T. sinskajae*	Not determined	Russia	S	VIR, St. Petersburg, Russia
K-48993	TRI 17701	*T. sinskajae*	Not determined	Russia	S	VIR, St. Petersburg, Russia
K-28239	TRI 6734	*T. urartu*	Not determined	Soviet Union	S	VIR, St. Petersburg, Russia
K-28239	TRI 7357	*T. baeoticum*	var. *buluchevskajae*	Soviet Union	S	VIR, St. Petersburg, Russia
K-28239	TRI 11355	*T. baeoticum*	var. *subbuluchevskajae*	Azerbaijan	S	VIR, St. Petersburg, Russia

Table 3 Name, chromosomal location, alleles, and PIC values of microsatellite markers used in analysis of hexaploid/tetraploid accessions (L = long arm, S = short arm, C = centromere region)

Marker	Chromosome arm location	Size range of alleles (bp)	Number of detected alleles	PIC
Xgwm357	1A(C)	120–130	6	0.74
Xgwm18	1BS	179–193	7	0.76
Xgwm458	1D(C)	110–192	10	0.66
Xgwm71	2AS, 2AL	111–133	12[a]	0.87
Xgwm95	2AS	110–130	9	0.79
Xgwm122	2A(C)	124–162	8	0.8
Xgwm619	2BL	134–172	13	0.88
Xgwm1099	2DS	117–145	14	0.86
Xgwm1217	3AL	115–150	5	0.5
Xgwm480	3AL	158–192	7	0.58
Xgwm155	3AL	120–148	13	0.83
Xgwm389	3BS	113–147	16	0.86
Xgwm3	3DL	76–86	5	0.59
Xgwm160	4AL	164–186	9	0.74
Xgwm192	4AS, 4BL, 4DL	129–199	9[b]	0.65
Xgwm513	4BL	138–148	6	0.70
Xgwm126	5AL	188–202	7	0.69
Xgwm186	5AL	99–141	12	0.85
Xgwm810	5BL	142–156	7	0.65
Xgwm190	5DS	199–217	8	0.80
Xgwm1293	6AL	107–117	5	0.47
Xgwm680	6BS	107–139	7	0.69
Xgwm325	6DS	137–151	8	0.8
Xgwm60	7AS	189–217	9	0.76
Xgwm631	7AS	176–212	8	0.68
Xgwm1061	7AL	146–176	9	0.7
Xgwm46	7B(C)	141–179	13	0.84
Xgwm437	7DL	95–119	9	0.86
Mean			8.96	0.74

[a,b]Scored as two or three different loci with the total number of alleles, respectively

ated until a single accession remained. Average genetic distances of the first accepted and the last separated accessions were recorded as maximum non-significant and minimum significant differences, respectively.

Results

Genetic diversity

Among the hexaploid and tetraploid accessions, a total of 251 different alleles of 31 microsatellite loci was detected using the 28 microsatellite primer pairs. Primer pairs GWM71 and GWM192 amplified alleles at two and three separate loci, respectively (Table 3). The average number of alleles per locus equaled 8.96. The most polymorphic marker *Xgwm389* had 16 different alleles; the least polymorphic markers *Xgwm1217* and *Xgwm1293* had five alleles each. The gene diversity varied from 0.47 to 0.88 with an average of 0.74. The 18 primer pairs used for the analysis of the diploid accessions detected 19 loci (Table 4). In total, 53 alleles were amplified with an average of 3.3 alleles per locus. Markers *Xgwm1217* and *Xgwm1293* were monomorphic. The gene diversity of the remaining markers varied between 0.40 and 0.77, with an average of 0.57.

Table 4 Name, chromosomal location, alleles and PIC value of microsatellite markers used in analysis of diploid accessions (L = long arm, S = short arm, C = centromere region)

Marker	Chromosome arm location	Size range of alleles (bp)	Number of detected alleles	PIC
Xgwm33	1AS	113–147	4	0.67
Xgwm71	2AS, 2AL[a]	115–133	5	0.75
Xgwm122	2A(C)	126, 162	2	0.47
Xgwm480	3AL	180, 182	2	0.47
Xgwm757	3AS	66, 100	2	0.40
Xgwm1217	3AL	105	1	0.00
Xgwm192	4AS	203–219	4	0.72
Xgwm601	4AS	130, 140, 142	3	0.53
Xgwm126	5AL	Null, 178, 194	2	0.59
Xgwm186	5AL	99–141	3	0.53
Xgwm1009	6AS	89, 103, 129	3	0.53
Xgwm1150	6AL	156–170	4	0.62
Xgwm1293	6AL	111	1	0.00
Xgwm60	7AS	142–152	4	0.66
Xgwm260	7A(C)	145–157	2	0.47
Xgwm276	7AL	85–121	5	0.77
Xgwm631	7AS	186–196	2	0.47
Xgwm942	7AL	105,149,153	3	0.53
Mean			3.3[b]	0.57[b]

[a]Scored as two different loci with the total number of alleles
[b]Among polymorphic markers

Table 5 Results of sequential testing for genetic homogeneity in potential duplicate groups

Duplicate group name	No. of accessions	Minimum significant difference	Maximum nonsignificant difference
Ceres	9	0.172*	0.109
Autonomia	8	0.283***	0.069
Risciola	8	0.333*	–[b]
Gabo	7	0.208*	0.104
Vernal	7	0.330***	0.074
Penjamo 62	6	0.208***	0.139
Thatcher	6	0.242***	0.068
Orlando	4	–[a]	0.115
Regina	4	0.434**	0.125
Apollo	3	0.408***	0.109
T. sinskajae	5	–[a]	0.166
T. baeoticum	3	0.500***	0.0

*, or ***Significance at the 5 and 0.1% levels, respectively
[a]No accessions were separated from their potential duplicate groups
[b]All accessions were separated from their potential duplicate groups

Testing for homogeneity

Hexaploid and tetraploid accessions (groups with identical cultivar names)

Sequential testing followed by the elimination of the most deviating accessions identified statistically homogeneous groups in nine of ten investigated potential duplicates groups of hexaploid/tetraploid accessions (Table 5). The exception was the group 'Risciola', where all eight accessions were significantly different. The maximum values of average genetic distances within statistically homogeneous groups of hexaploid accessions varied from 0.068 ('Thatcher') to 0.139 ('Penjamo 62'), and minimum significant differences of distinct accessions within the groups of potential duplicates were in the interval of 0.172–0.434 (Table 5).

The 'Ceres' group consisting of nine accessions formed separate groups with close pair-wise clustering of accessions TRI 411 and TRI 412, and also TRI 15976 and TRI 28629 (Fig. 1a). Applying sequential bootstrapping for the subclusters separately, it was found that both pairs were statistically homogeneous whereas all other accessions were significantly different (Table 6).

Bootstrap testing of group 'Autonomia' containing eight accessions revealed one genetically homogeneous pair (TRI 7113 and TRI 13051). The distances of the six other accessions were significantly large and varied from 0.283 to 0.631. The largest genetic distance was found for TRI 14139 (*T. durum*).

All accessions of the 'Risciola' group, consisting of six hexaploids and two tetraploids, were significantly different to each other. The average genetic distance varied from 0.333 to 0.673. The UPGMA cluster analysis indicated heterogeneity within the group, partitioning the accessions into subclusters and clearly separating the durum wheats (data not shown). The hexaploid accessions belonged to at least three different botanical varieties (var. *aestivum*, var. *ferrugineum* and var. *hostianum*) indicating that all members of 'Risciola' are distinct accessions (Table 1).

The analysis of the seven accessions of 'Gabo' exhibited one genetically homogeneous pair (TRI 7244 and TRI 9684), whereas all other accessions were distinct from each other, with significantly large average genetic distances (Table 6). The dendrogram constructed confirmed these data (Fig. 1b). In total, three different botanical varieties were described (Table 1).

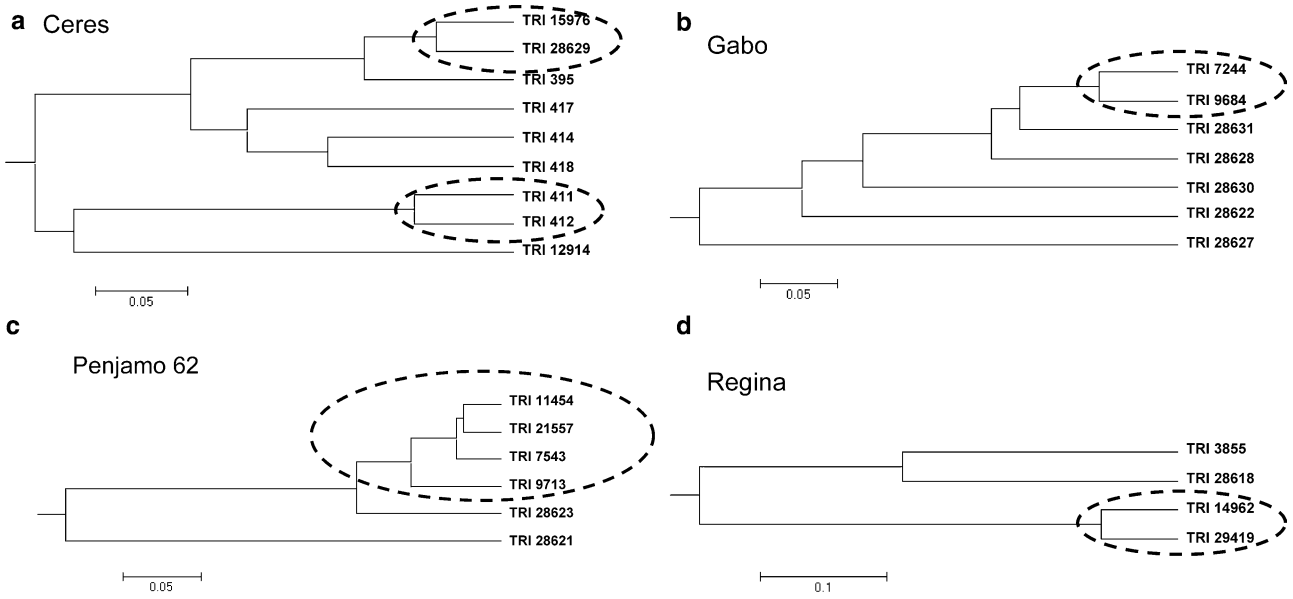

Fig. 1 Dendrograms constructed for the duplicate group 'Ceres' (**a**), 'Gabo' (**b**), 'Penjamo 62' (**c**), and 'Regina' (**d**). Genetically homogeneous groups are *circled*

Another set of seven accessions having an identical cultivar name was formed by the tetraploid duplicate group 'Vernal'. Here, four accessions built a homogeneous group (TRI 445, TRI 28616, TRI 9525 and TRI 18200). Cluster analysis showed that all accessions formed one unique cluster in which members of the statistically homogeneous group clustered closely to each other whereas the other accessions were step-wise separated (data not shown).

Analysis of the 'Penjamo 62' group revealed that four out of six accessions (TRI 9713, TRI 7543, TRI 11454 and TRI 28626) formed a statistically homogeneous group. Two other members of this group (TRI 28621 and TRI 28623) had significantly large average genetic distances (Table 6). Cluster analysis grouped all six accessions of this group into a single cluster, in which accessions were separated gradually (Fig. 1c).

Sequential testing of the group 'Thatcher' consisting of six accessions identified three accessions with significantly large average genetic distances (TRI 1308, TRI 3117 and TRI 4532) and one genetically homogeneous group (TRI 5350, TRI 9710 and TRI 12688). Values of average genetic distances are given in Table 5. Also, for this group, the accessions formed a single cluster from which members were separated one by one in the dendrogram (data not shown).

For 'Orlando' the bootstrap test detected that all four accessions of this group formed a statistically homogeneous group (Table 5). The average genetic distance of the most deviating accession TRI 9322 (0.115) was not significant. The dendrogram resulting from the cluster analysis demonstrated that all accessions clustered closely together which was consistent with the results of the sequential testing (data not shown). The homogeneity of

Table 6 Examples of sequential testing, followed by elimination of the most deviating accessions in the potential duplicate groups 'Ceres', 'Gabo', 'Penjamo 62', 'Regina' and the diploids *T. sinskajae* and *T. baeoticum*

Duplicate group/cluster	Accession no.	Maximum average genetic distance	P^b
Ceres (cluster 1)	TRI 12914	0.479*	0.000
	TRI 411	0.109	0.580
	TRI 412		
Ceres (cluster 2)	TRI 395	0.172*	0.104
	TRI 15976	0.086	0.080
	TRI 28629		
Ceres (cluster 3)	TRI 414	0.290*	0.029
	TRI 417	0.204*	0.026
	TRI 418	0.204*	0.026
Gabo	TRI 28627	0.619***	0.000
	TRI 28622	0.480***	0.000
	TRI 28630	0.408**	0.004
	TRI 28628	0.243*	0.006
	TRI 28631	0.208*	0.033
	TRI 7244	0.104	0.067
	TRI 9684		
Penjamo 62	TRI 28621	0.577***	0.000
	TRI 28623	0.208*	0.075
	TRI 9713	0.139	0.164
	TRI 7543	0.0806	0.239
	TRI 11454	0.072	0.137
	TRI 28626		
Regina (*T. durum*)	TRI 3855	0.435***	0.000
	TRI 28618		
Regina (*T. aestivum*)	TRI 14962	0.125	0.069
	TRI 29419		
T. sinskajae	TRI 14732	0.160	0.227
	TRI 17415	0.103	0.346
	TRI 12910	0.076	0.246
	TRI 17701	0.053	0.141
	TRI 11525		
T. baeoticum	TRI 6734	0.500***	0.000
	TRI 7357	0.000	
	TRI 11355	*0.000*	

*, **, *** indicate significance at the 5, 1, and 0.1% levels, respectively; P^b is the probability of obtaining a larger maximum distance by chance

this duplicate group was confirmed by the morphological classification (Table 1).

The dendrogram for the group 'Regina' showed that the four accessions were split into two subclusters (Fig. 1d). These two subclusters represent the two different species (*T. aestivum* and *T. durum*) having different growth habits as well. Results of sequential bootstrap of the two subclusters detected one genetically homogeneous pair and two statistically different accessions (Table 6). The homogeneous pair was formed by the two hexaploid accessions.

The final potential duplicates group 'Apollo' consisted of one genetically homogeneous pair (TRI 5244 and TRI 28625) and one accession (TRI 28632) with a significantly large average genetic distance (Table 5). Morphological analysis did not detect any difference between accessions of the group. They belonged to the var. *lutescens*.

Diploid accessions

For the diploid wheats analysed, all five accessions of group '*T. sinskaja*' were statistically homogeneous having a maximum value of average genetic distance of 0.160 (Table 7). In the dendrogram they clustered closely together. The other diploid group consisting of three accessions was found to contain one homogeneous pair (TRI 7357 and TRI 11355), whereas one accession (TRI 6734) was significantly distant (0.500). Although having identical numbers of the donor institute, TRI 6734 was classified in the Gatersleben collection as *T. urartu* instead of *T. baeoticum*.

Discussion

Molecular marker technology is a powerful tool for investigating genebank collections because of the high possible throughput. Besides studying the genetic diversity *per se* as demonstrated by Donini et al. (1998, 2000), Hammer et al. (2000), Huang et al. (2002), Khlestkina et al. (2004a, b) and Alamerew et al. (2004), molecular markers have been successfully used for the examination of the genetic integrity of self pollinating (wheat) and open pollinating (rye) crops (Börner et al. 2000; Chebotar et al. 2003).

Furthermore, molecular markers (microsatellites) have been used for the analysis of potential duplicates in barley genebank collections (Lund et al. 2003). A bootstrap approach based on re-sampling of both microsatellite markers and alleles within markers was used to test for homogeneity. The largest average genetic distance accepted in any homogeneous group was smaller than the smallest distance declared significant in any group, with a threshold average genetic distance of 0.14.

The aim of the present study was to use microsatellites for the validation of probable duplicates selected from the Gatersleben wheat (hexaploid, tetraploid and diploid) collection based on common cultivar names or identical accession numbers of the donor genebank. From the results of sequential bootstrap, it may be assumed that the threshold value for analysed hexaploid accessions is approximately equal to 0.139, which is highly comparable to that described for barley (Lund et al. 2003).

It was clearly demonstrated that a cultivar name alone is not sufficient for the detection of identical accessions. In the case of the Italian local cultivar 'Risciola' all eight accessions were shown to differ significantly from each other by applying microsatellite marker analysis. Some of the differences were confirmed by the morphological classification.

Within the 'Ceres' group two pairs of duplicated accessions differing from each other and all other accessions were detected, whereas in other duplicate sets of hexaploid/tetraploid wheats groups of two, three or even four statistically homogeneous accessions were identified by applying the marker analysis. The data were verified by the botanical classification where available.

By investigating the diploid wheats, which usually have no cultivar names, it was shown that even the accession number of the donor genebank may be misleading in detecting duplicates. In the database of the VIR St. Petersburg accession K-28339 was listed as *T. baeoticum*. On the other hand, the botanical classification of TRI 6734, which was performed at IPK Gatersleben resulted in *T. urartu*. The reason for this disagreement is not clear at the moment. Probably it is simply a mistake in one of the databases.

Another interesting discovery was made in the *T. sinskajae* group. The five accessions did not differ significantly although the genetic distances varied slightly with TRI 14732 having the highest value. Regarding the passport data all accessions originated from VIR St. Petersburg; however, TRI 14732 was introduced to the Gatersleben collection from the Botanical Garden in Kishinev, Moldavia. This finding confirms the rule to always consider the most original samples for the maintenance in genebanks.

In summary, the method of re-sampling microsatellite data was shown to be a powerful tool for identifying potential duplicates in our wheat collection. In combination with botanical classification and other data available from the database (origin, growth habit and donor), duplicates can be identified. For this purpose we propose the following procedure:

1) to form groups of accessions having the same or similar cultivar names and/or accession numbers of donor genebanks upon passport data;
2) to apply SSR analysis and sequential bootstrap to detect statistically homogeneous groups and separated accessions;
3) to compare any available passport data for accessions of statistically homogeneous groups for further separation;

4) to perform morphological evaluations of accessions from homogeneous groups;
5) to determine the duplicates. After identification of the duplicates we propose to use the most original sample for maintenance and to eliminate the others. If this is not possible the accessions of the homogeneous groups should be pooled in order to reduce and therefore rationalise the genebank holdings.

Acknowledgements We thank Marina Schäfer and Michael Grau for supplying the seed samples as well as Dr. S. B. Anderson, The Royal Veterinary and Agricultural University, Copenhagen, Denmark for providing the computer program 'Diversity' and helpful discussions. U. Saleh was supported by the 'Internationale Weiterbildung und Entwicklung gGmbH' (InWent).

References

Alamerew S, Chebotar S, Huang XQ, Röder MS, Börner A (2004) Genetic diversity in Ethiopian hexaploid and tetraploid wheat germplasm assessed by microsatellite markers. Gen Res Crop Evol 51:559–567

Anonymous (2005) Annual Report 2004, Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, 192pp

Börner A, Chebotar S, Korzun V (2000) Molecular characterization of the genetic integrity of wheat (*Triticum aestivum* L.) germplasm after long term maintenance. Theor Appl Genet 100:494–497

Chebotar S, Röder MS, Korzun V, Saal B, Weber WE, Börner A (2003) Molecular studies on genetic integrity of open pollinating species rye (*Secale cereale* L.) after long term genebank maintenance. Theor Appl Genet 107:1469–1476

Donini P, Stephenson P, Bryan GJ, Koebner RMD (1998) The potential of microsatellites for high throughput genetic diversity assessment in wheat and barley. Gen Res Crop Evol 45:415–421

Donini P, Law JR, Koebner RMD, Reeves JC, Cooke RJ (2000) Temporal trends in the diversity of UK wheat. Theor Appl Genet 100:912–917

FAO (1998) The state of the world's plant genetic resources for food and agriculture. Food and Agriculture Organization of the United Nations, Rome, 510 pp

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Hammer K, Knüpffer H, Filatenko A, Zuev E (1998) Diploid wheats in the Gatersleben genebank. In: Jaradat AA (ed) Triticeae III. Science publishers Inc., NY, pp 37–44

Hammer K, Filatenko AA, Korzun V (2000) Microsatellite markers—a new tool for distiguishing diploid wheat. Gen Res Crop Evol 47:497–505

Hintum TJL, Knüpffer H (1995) Duplication within and between germplasm collection. I. Identifying duplication on the basis of passport data. Gen Res Crop Evol 42:1127–1133

Hintum TJL, Visser DL (1995) Duplication within and between germplasm collection. II. Duplication in four European barley collections. Gen Res Crop Evol 42:135–145

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

Khlestkina EK, Röder MS, Efremova TT, Börner A, Shumny VK (2004a) The genetic diversity of old and modern Siberian varieties of common spring wheat determined by microsatellite markers. Plant Breed 123:122–127

Khlestkina EK, Huang XQ, Quenum FJB, Chebotar S, Röder MS, Börner A (2004b) Genetic diversity in cultivated plants—loss or stability? Theor Appl Genet 108:1466–1472

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Lund B, Ortiz R, Skovgaard IM, Waugh R, Andersen SB (2003) Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. Theor Appl Genet 106:1129–1138

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91:1001–1007

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Ruiz M, Aguiriano E (2004) Analysis of duplication in the Spanish durum wheat collection maintained in the CRF-INIA on the basis of agro-morphological traits and gliadin proteins. Gen Res Crop Evol 51:231–235

Sahu RK (1989) Screening for duplicates in the germplasm collection. Int Rice Res Newsl 14:4