Theor Appl Genet (2003) 106:1129–1138
DOI 10.1007/s00122-002-1130-y

B. Lund · R. Ortiz · I. M. Skovgaard · R. Waugh
S. B. Andersen

Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data

Received: 10 March 2002 / Accepted: 20 August 2002 / Published online: 30 October 2002

Abstract Redundant duplication among putative Nordic spring barley material held at 12 gene banks worldwide was studied using 35 microsatellite primer pairs covering the entire barley genome. These microsatellite markers revealed an average of 7.1 alleles per locus, and a range of 1 to 17 different alleles per locus. Similarity of accession name was initially used to partition the 174 repatriated accessions into 36 potential duplicate groups, and one group containing 36 apparently unique or unrelated accessions. This partitioning was efficient to produce a distribution of mainly small average genetic distances within potential duplicate groups compared to distances from the group of unique accessions. However, comparisons within potential duplicate groups still contained large genetic distances of the same size as distances between unique accessions indicating classification errors. A bootstrap approach based on re-sampling of both microsatellite markers and alleles within marker loci was used to test for homogeneity within potential duplicate groups. The test was used in each group for sequential elimination of accessions with a significantly large average genetic distance to identify a homogeneous group. Such genetically homogeneous groups of two or more accessions were identified in 22 among the 36 potential duplicate groups studied. Results from the genetic analysis of some potential duplicate groups supported previous conclusions based on passport data through inclusion of the historically most-original accession in the genetically homogeneous group. In other potential duplicate groups the apparently most-original accession according to passport data was not included in the homogeneous set of accessions, indicating that this most-original accession does not have duplicate accessions in the group. During the present study 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 approximately 0.14. The results are discussed with respect to the identification of duplicate accessions within potential duplicate groups, as well as the elimination of genetic off types in such groups. Furthermore, large barley gene bank collections may be screened for potential duplicates with genetic distances below the suggested threshold of 0.14.

Keywords Barley *Hordeum vulgare* L. · Bootstrap · Duplicates · Genetic resources · Microsatellite markers

Communicated by H.C. Becker

B. Lund · S.B. Andersen (✉)
Department of Agricultural Sciences,
The Royal Veterinary and Agricultural University,
Thorvaldsensvej 40, DK-1871 Frederiksberg C, Copenhagen,
Denmark
e-mail: sba@kvl.dk
Tel.: +45-35-28-34-44, Fax: +45-35-28-34-60

R. Ortiz
International Institute of Tropical Agriculture,
c/o L.W. Lambourn and Co., Carolyn House, 26 Dingwall Road,
Croydon, CR9 3EE, UK

I.M. Skovgaard
Department of Mathematics and Physics,
The Royal Veterinary and Agricultural University,
Thorvaldsensvej 40, DK-1871 Frederiksberg C, Copenhagen,
Denmark

R. Waugh
Genome Dynamics Program, Scottish Crop Research Institute,
Invergowrie, Dundee DD2 5DA, Scotland

Introduction

More than 6-million accessions of plant genetic resources are maintained in ex situ germplasm collections worldwide (FAO 1998). An important aim in the Global Plan of Action for Conservation and Sustainable Utilization of Plant Genetic Resources for Food and Agriculture (FAO 1996) is to increase efficiency for conservation of ex situ collections due to their expansion during recent decades. This includes the development of core collections (Frankel and Brown 1984; Brown 1989; Hintum et al. 2000) and reduction in redundant duplication within and between collections.

Some of the latest estimates of the total number of unique accessions maintained among world gene banks have been reported to be between 1 and 2 million only (FAO 1998; Hammer et al. 1999). Barley (*Hordeum vulgare* L.) is presently represented with approximately 486,000 accessions (FAO 1998). Among barley world collections, large numbers of apparently duplicate accessions have been reported in two investigations (Lyman 1984; Plucknett et al. 1987). Exact figures for the degree of such duplicate accessions, however, are still uncertain (FAO 1998). Furthermore, an unknown number of accessions may be misclassified because of errors in passport data and seed mixing, or exchange during activities of maintenance (Hintum and Knüpffer 1995; Willner et al. 1998; Gilbert et al. 1999). Redundant duplication introduces a waste of resources through increased costs of general seed and documentation management activities without contributing to the maintenance of genetic diversity. For users of ex situ collections (e.g. plant breeders, researchers and gene bank curators), documentation errors in passport data reduce the practical value of germplasm collections because access to, and exploitation of, desired genetic resources become inefficient.

Different methods have been described to identify potential duplicate accessions within and between collections. Potential duplicate accessions may be identified based on passport data comprising identity or similarity of the name or designation of the accession and geographical origin (Hintum and Knüpffer 1995; Greene and Pederson 1996). Furthermore, passport data (Hintum and Knüpffer 1995) combined with a review of historic documentation (Greene and Pederson 1996; Willner et al. 1998) may be used. Passport information as the only means for identification of duplicate accessions may be uncertain (Sahu 1989; Hintum and Knüpffer 1995; Hintum and Visser 1995; Willner et al. 1998). Therefore, potential duplicate accessions should be verified through analysis of genetic similarity. Genetic comparisons based on phenotypic measurements have been used for such verification of potential duplicate accessions (Sahu 1989; Greene and Pederson 1996; Ortiz et al. 1998; Ortiz et al. 2002). These approaches can now be complemented by the use of molecular markers at the protein (Hintum and Visser 1995; Hintum et al. 1996) or DNA level (Waycott and Fort 1994; Virk et al. 1995; Zeven et al. 1998). In particular, DNA based markers have the advantage of being unaffected by the environment. Approaches for the identification of duplicate accessions based on DNA markers have used studies of dendrograms (Cervera et al. 1998; Gilbert et al. 1999) and molecular analysis of within and between accession variation (Phippen et al. 1997; Dean et al. 1999; Treuren et al. 2001).

Barley microsatellite markers or simple sequence repeats (SSRs) show high levels of polymorphism within loci (Russell et al. 1997a) and therefore enable discriminating between closely related genotypes (Russell et al. 1997b; Struss and Plieske 1998; Pillen et al. 2000; Russell et al. 2000). These markers have a high degree of reproducibility between laboratories (Jones et al. 1997) and are suitable for automation (Ziegle et al. 1992; Mitchell et al. 1997). This paper reports results with a re-sampling approach based on microsatellite marker data, to study redundant duplication among barley accessions obtained from germplasm collections worldwide for repatriation to the Nordic Gene Bank.

Materials and methods

Plant material

The 174 spring barley accessions used were part of a larger barley material with a supposed origin in one of the five Nordic countries obtained for repatriation from non-Nordic gene banks (Table 1). Selection for repatriation was based on the search in available passport information for the country of origin, followed by the accession name, the year of inclusion in the collection and the type of material (variety or landrace). Only potentially unique accessions of Nordic origin, not held by the Nordic Gene Bank, were selected for repatriation. For microsatellite analysis, nine Nordic reference accessions (Table 2) previously cultivated on large acreages in the Nordic region were selected and requested too, if available. Based on identical or similar accession names, the 174 selected accessions were grouped into potential duplicate groups (Table 2) and a remaining group consisting of apparently "unique accessions" without obvious duplicates (data not shown).

DNA extraction and microsatellite analysis

Young leaf material from six single plants representing one spike of each individual accession was used for extraction of genomic DNA with the CTAB procedure modified by Saghai-Maroof et al. (1984). The analysis used 48 barley microsatellite primer pairs with the approximate map locations described in Macaulay et al. (2001). Microsatellite markers were selected to cover all 14 barley chromosome arms. Polymerase chain reactions (PCR) were performed with minor modifications according to Russell et al. (1997b). Subsequent analysis of PCR-amplified microsatellites using silver-stained polyacrylamide gels was performed as described by Christiansen et al. (2002).

Data analysis

For accessions in each potential duplicate group, and for the group of apparently unique accessions, a genetic dissimilarity matrix was calculated from marker scores. If an accession was genetically heterogeneous, i.e. a mixture of different genotypes, more than one allele could be observed for a microsatellite marker locus. Similarity between two accessions within one locus was calculated as the number of common alleles relative to the total number of alleles in the locus. Similarity between two accessions was calculated as the average similarity over all marker loci and dissimilarity as 1-similarity (Diwan and Cregan 1997). Dendrograms based on dissimilarity matrices were constructed using PROC CLUSTER in SAS (Anonymous 1990) by the unweighted pair-group method with an arithmetic mean (UPGMA).

For each group of potential duplicate accessions and the group of unique accessions, an average genetic distance for each accession with other accessions in the group was calculated. To investigate whether a group of potential duplicate accessions contained deviating members the hypothesis that differences occur at random, or more precisely that accessions are identically distributed with independence between loci, was tested. For simplicity this was referred to as the hypothesis of homogeneity, since it corresponds to sporadic differences in contrast with one or a few accessions accounting for a higher proportion of differences within the

Table 1 Gene banks contributing material for repatriation of spring barley. In the last three columns the total number of SSR-analysed accessions are divided into unique, Nordic reference and potential duplicate accessions, respectively

Donor gene bank[a]	Donor country	Putative Nordic accessions	Repatriated accessions	Analysed with SSR	Unique accessions	Reference accessions	Potential duplicates
VIR	Russia	Yes	56	29	19	6	4
NSGC	USA	Yes	77	44	28	9	7
PGRC	Canada	Yes	–	–	–	–	–
BGC	Japan	Yes	7	5	2	1	2
IHAR	Poland	Yes	29	14	10	4	0
RICP-GP	Czech Republic	Yes	18	8	5	2	1
JIC	United Kingdom	Yes	58	23	12	8	3
BAZ	Germany	Yes	23	14	9	2	3
IPK	Germany	Yes	25	16	9	5	2
CGN	The Netherlands	Yes	–	–	–	–	–
INRA-Clermont	France	Pending	–	–	–	–	–
TAMAWC	Australia	Yes	7	1	0	0	1
ICGR-CAAS	China	No	0	0	–	–	–
NBPGR	India	Pending	–	–	–	–	–
CIMMYT	Mexico	No	0	0	–	–	–
GRU-ICARDA	Syria	No	0	0	–	–	–
NGB	DNK, FIN, NOR, SWE		(21)	20	0	9	11
Total			321	174	94	46	34

[a] VIR = N.I. Vavilov Research Institute of Plant Industry, St. Petersburg, Russian Federation; NSGC = National Small Grains Collection (USDA-ARS), Aberdeen, Idaho, United States; PGRC = Plant Gene Resources of Canada, Saskatoon, Saskatchewan, Canada; BGC = Barley Germplasm Center, Okayama University, Kurashiki, Japan; IHAR = Plant Breeding and Acclimatization Institute, Radzikow, Poland; RICP-GP = Genebank Depart. Div. Genet. and Plant Breed. Res. Inst. Crop Production, Prague, Ryzyne, Czech Republic; JIC = John Innes Centre Norwich Research Park, Norwich, United Kingdom; BAZ = Federal Centre for Breeding Research on Cultivated Plants (BAZ) Gene Bank, Braunschweig, Germany; IPK = Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany; CGN = Centre for Genetic Resources, The Netherlands (CGN), Wageningen, the Netherlands; INRA-Clermont = Station d'Amelioration des Plantes, Clermont-Ferrand Cedex, France; TAMAWC = Australian Winter Cereals Collection, Agricultural Research Centre, Tamworth, New South Wales, Australia; ICGR-CAAS = the National Crop Gene Bank of China, Institute of Crop Germplasm Resources CAAS, Beijing, China; NBPGR = National Bureau of Plant Genetic Resources in New Delhi and Regional Station Akola, India; CIMMYT = Spanish acronym for "International Maize and Wheat Improvement Center", Mexico City, Mexico; GRU-ICARDA = Genetic Resources Unit of ICARDA, Syria; NGB = Nordic Gene Bank, Alnarp, Sweden

group. The hypothesis of homogeneity was tested with a bootstrap approach (Efron 1979; Felsenstein 1985). To account for the uncertainty caused by analysis with a limited number of genetic markers, marker loci were sampled randomly with replacement, leaving some loci out and duplicating others for creation of a new or pseudoreplicate data set of the same size as the original. To simulate the situation with all accessions in the duplicate group being equally different, alleles from different accessions were sampled from the observed allele distribution at each locus, again with replacement. Thus, the bootstrap method was used first to sample the marker loci and then again to sample alleles for each accession at each locus. Re-sampling of alleles in each marker also accounted for the uncertainty due to a limited number of alleles in the locus. Maximum average genetic distance for any accession in the group was used as a test statistic recorded for 1,000 pseudoreplicate (bootstrap) samples. The probability (P) for the observed maximum average genetic distance within the group was then calculated from this re-sampled distribution. If the observed maximal average genetic distance within the group turned out to be significant at 5% (i.e. $P = 0.05$) the most-deviating accession was removed and the procedure iterated until a single accession remained. A windows-based computer programme to perform the calculation has been written and is available upon request.

Results

Passport data (accession name) partitioned the original 174 repatriated barley accessions into 36 groups of potential duplicate accessions and another group of 36 apparently unrelated or unique accessions. Among the 36 potential duplicate groups, 15 groups contained only two accessions each, while the remaining duplicate groups each held between three to ten accessions (Table 2).

Microsatellite markers

Analysis of all 174 barley accessions with 48 different microsatellite primer pairs revealed eight primer pairs (Bmag0211, Bmag0384, HvMLO3, HvLEU, Bmag0218, Bmac0018, EBmac0806 and Bmag0007) that failed to amplify any product. Furthermore, five primer pairs (Bmag0013, HVM3, EBmac0701, EBmac0684 and Bmac0223) produced only weak amplification products uneasy to score. The remaining 35 markers used in the present study showed a high average degree of polymorphism with an average of 7.1 alleles per locus and a total average diversity of 0.56 in the entire material. One of these 35 microsatellite markers (HvLOX) was monomorphic in all accessions, while the most-polymorphic marker (Bmag0156) had 17 different alleles. Primer pairs generally amplified only one allele in each accession, as expected for a highly self-pollinating species like barley. Only six accessions showed genetic heterogeneity with

Table 2 Results of sequential testing for genetic homogeneity in potential duplicate groups

Duplicate group name	No. of accessions	Country of origin	Minimum significant difference[d]	Maximum non-significant difference[e]
Danisch Insel	3	DNK	0.642***	0.088
Erhard Frederiksen	4	DNK	0.181*	0.032
Lofa Abed[c]	6	DNK	0.333***	0.0
Kenia Abed[c]	6	DNK	0.271*	0.114
Pomo[c]	6	FIN	0.490***	0.046
Balder J	3	FIN	–[a]	0.029
Halikko[c]	3	FIN	0.177*	–[b]
Uurainen	3	FIN	0.400**	–[b]
Perttu	6	FIN	–[a]	0.021
Vankurri	8	FIN	–[a]	0.059
Olli	10	FIN	0.220*	0.135
Asplund	3	NOR	0.283**	0.0
Mjøs[c]	3	NOR	0.333**	–[b]
Bjørneby[c]	3	NOR	0.344**	–[b]
Maskin[c]	4	NOR	0.512***	0.037
Anita	5	NOR	0.392**	–[b]
Opal B	3	SWE	0.144*	0.0
Nemex	4	SWE[f]	–[a]	0.020
Svanhals[c]	7	SWE	0.337***	0.051
Bonus[c]	9	SWE	0.195*	0.135
Clara	9	SWE	0.280***	0.113
Loti	2	DNK	–[a]	0.0
Rosie Abed	2	DNK	–[a]	0.029
Griar	2	NOR	–[a]	0.0
Amble Sogne Fjord	2	NOR	–[a]	0.0
Selecta	2	SWE	0.575**	–[b]
Goud	2	SWE	0.313**	–[b]
Skånes 6-row	2	SWE	0.356	–[b]
Skåneslet	2	SWE	0.370**	–[b]
Gotlands	2	SWE	0.531***	–[b]
Ölands	2	SWE	0.196*	–[b]
Luleå	2	SWE	0.575***	–[b]
Seru	2	SWE	0.225*	–[b]
Black hull-less	2	SWE	0.692***	–[b]
Marlies	2	SWE[f]	–[a]	0.047
Cross	2	SWE[f]	–[a]	0.031

*, ** or *** indicate significance at the 5%, 1% and 0.1% levels, respectively
[a] No accessions were separated from its potential duplicate group
[b] All accessions were separated from its potential duplicate group
[c] Nine Nordic standard reference accessions
[d] Average genetic distance of the last accession separated
[e] Average genetic distance of the first accession accepted
[f] According to passport data from donor

more than two alleles in one locus. In addition, primer pairs Bmac0067 and Bmac0273 amplified alleles from two separate microsatellite loci in all accessions (Table 3).

Genetic-distance distributions

Identification of potential duplicate groups based on passport data was efficient because the average genetic distance from other group members within the group of unique accessions, and within groups of potential duplicate accessions, produced widely different distributions (Fig. 1). Average genetic distances of unique accessions formed a near normal distribution (Fig. 1A). In contrast, within-group average genetic distances for accessions in potential duplicate groups produced a very different distribution with a major density peak close to zero. However, this distribution of within duplicate-group comparisons also contained larger average genetic distances of up to 0.70.

Bootstrap testing

Sequential testing for genetic homogeneity, followed by elimination of the most-deviating accession, identified statistically homogeneous groups in 16 of the 21 potential duplicate groups with more than two accessions (Table 2). Accessions in the remaining five of these potential duplicate groups all separated from each other at the 5% level. Similar testing for genetic homogeneity within each of the 15 potential duplicate groups each containing only two accessions declared six of these pairs to be genetically homogeneous (Table 2). Inspection of average genetic distances in each potential duplicate group revealed that the smallest average distance declared significant in any group was 0.144 (Table 2). Accordingly, the largest average genetic distance accepted in any potential duplicate group declared genetically homogeneous was 0.135 (Table 2). Thus, although the hypothesis of homogeneity did not directly refer to differences within a group being small or large, it turned out in this study that homogeneous groups were, in fact, separated from heterogeneous groups by smaller average genetic distances (Table 2). For comparison the smallest

Table 3 Name, chromosome location, alleles and PIC-value for microsatellite markers

Locus	Chrosomal location	Number of alleles	Allele size range (bp)	PIC[c]
Bmac0399	5(1H)[a]	13	119–170	0.80
Bmac0032	5(1H)	11	210–270	0.76
Bmag0382	5(1H)	3	95, 107, 113	0.44
HvHVA1	5(1H)	2	140, 142	0.06
WMC1E8	5(1H)	2	190, 233	0.44
Bmac0134	2(2H)	12	138–182	0.73
HVM36	2(2H)	7	109–141	0.72
Bmag0378	2(2H)	5	136–152	0.40
Bmac0093	2(2H)	6	153–163	0.78
Bmag0125	2(2H)	11	121–149	0.69
HVM54	2(2H)	5	149–166	0.70
EBmac0415	2(2H)	5	228–249	0.65
HvLTTB	3(3H)	4	209–224	0.22
Bmac0067	3(3H)	7[b]	213–228	0.61
Bmac0209	3(3H)	7	181–202	0.72
Bmag0136	3(3H)	3	204,206,208	0.35
Bmag0225	3(3H)	14	143–173	0.82
HVM62	3(3H)	7	233–266	0.29
HVM40	4(4H)	6	146–165	0.45
Bmag0353	4(4H)	7	99–137	0.49
HVM67	4(4H)	2	230, 233	0.44
EBmac0970	7(5H)	2	198, 200	0.17
Bmac0113	7(5H)	9	181–215	0.80
Bmag0222	7(5H)	5	151–166	0.53
HvLOX	7(5H)	1	153	0.00
Bmac0316	6(6H)	8	156–194	0.32
Bmag0173	6(6H)	12	126–177	0.80
Bmag0009	6(6H)	6	171–181	0.28
Bmac0040	6(6H)	11	177–234	0.76
Bmag0021	1(7H)	6	137–148	0.66
HVCMA	1(7H)	3	137,142,145	0.50
Bmac0273	1(7H)	14[b]	141–221	0.81
Bmag0120	1(7H)	7	234–270	0.77
Bmac0156	1(7H)	17	136–212	0.91
Bmag0135	1(7H)	10	126–164	0.85
Average		7.1		0.56

[a] Chromosome numbering follows barley nomenclature, with the homologous chromosome group given in brackets
[b] Scored as two different loci with the total number of alleles
[c] Polymorphic information content (PIC), i.e. genetic diversity of marker

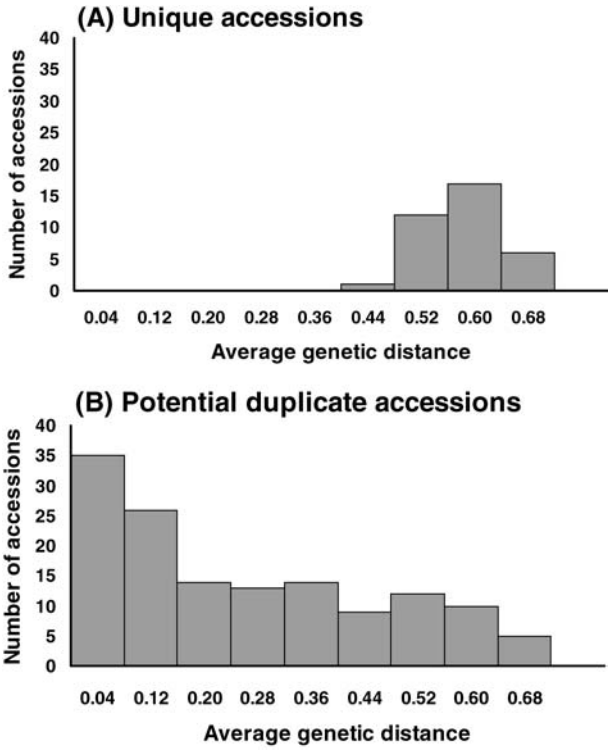

Fig. 1A–B Average genetic distances for each accession within a group of unique accessions (**A**) and within groups of potential duplicate accessions (**B**)

genetic distance observed between any pair of unique accessions in this study was 0.137, while the largest genetic distance between any pair of unique accessions was 0.857 (data not shown).

Conclusions from the sequential testing approach were in good accordance with the appearance of the dendrograms of each potential duplicate group. Analysis of the duplicate group, consisting of ten accessions named 'Olli', identified four accessions designated 3011, 2015, 3013 and 3012 with significantly large average genetic distances (Table 4). The remaining six accessions in the potentially duplicate group could be regarded as a genetically homogeneous group. The dendrogram for the duplicate group showed that these ten accessions formed a single cluster (Fig. 2A), from which accessions were separated one at a time. The original cultivar 'Olli' was developed through single-plant selection from a Finnish landrace by Hankkija Plant Breeding Institute, Finland, and released in 1927 (Aikasalo 1988). From passport data the most original, i.e. the accession with the supposed original integrity, of these ten accessions named 'Olli' were accessions 2019, 2016 and 2017, which have all been donated from the Hankkija breeding company to the National Small Grains Collection of the USA, through different donors (Table 4). This idea was supported by the fact that these three accessions of 'Olli' were part of the statistically homogeneous group identified.

Analysis of the group named 'Bonus' with nine accessions, declared average genetic distances of two accessions 7012 and 1010 as significantly different (Table 4). The remaining seven accessions formed a statistically homogeneous group. Also for this group the accessions formed a single cluster from which members were separated one by one in the dendrogram (Fig. 2B). The original Swedish cultivar 'Bonus' {pedigree: [[Sv.34/22 [Opal × Seger (Gull × Hannchen)] × Maja]]} (Torp et al. 1978) was released in 1950 by the Swedish Seed Association (Osvald 1959). Accordingly, the seven accessions of 'Bonus' constituting the genetically homogeneous group all had Sweden as the country of origin. Therefore, the most original accessions of 'Bonus' from the Swedish Seed Association might be identified in this group, because passport data indicated that accessions 1010, 6021 and 7017 had all been donated directly from

Table 4 Accessions of four potentially duplicate groups ranged according to elimination with the bootstrap test

Name: accession no.	Country of origin[c]	Max. ave. genet. dis.[a]	P[b]	Donor to NGB	Donor (s) history	Accession no.	Year received
Olli							
3011	FIN	0.677***	0.0	AUSAWCC	P3058, WUB25	AUS 401631	1986
2015	FIN	0.637***	0.0	INTNGB	KVL, DNK→Smith, M. CAN in 1952 selec. CA739=CI6251	NBG9520	1991
3013	FIN	0.576***	0.0	POLIHAR	Instutute of Genetics and Plant Breeding, Warsaw, POL	IHAR 41859	1961
3012	FIN	0.220*	0.016	DEUIPK	Jokioinen, Finland	HOR 1433	1949
2019	FIN	0.135	0.124	USANSGC	Laitinen, A., Hankkija Plant Breeding Institute, FIN	PI 467411	1982
2021	FIN	0.084	0.166	GBRJIC	Canada Dept of Agriculture, Research St., Winnipeg, CAN	B 8056	1973
2018	FIN	0.039	0.341	USANSGC	Salekhard Exp. St., Sibiria, RUS through USA Off. Agr. Relat.	PI 156627	1946
2020	FIN	0.032	0.293	JPNRIB	Source: KYO	OUU 359	
2016	FIN	0.0	0.0	USANSGC	Hankkija FIN, through USA consul J.L. Bouchal	PI 102157	1933
2017	FIN		0.0	USANSGC	Dr. R. Shands, USDA→ CAN (CAN739)→Dr. Pessola, FIN	CIho 6251	1937
Bonus							
7012	NOR	0.491***	0.0	RUSVIR	Collected by Russian scientist in Norway	K-17017	1947
1010	SWE	0.195*	0.048	INTNGB	Svalöf AB, Sweden	NGB1489.1	1986
6021	SWE	0.135	0.104	USANSGC	Dr. K. Froier, K., SWE at request of Dr. E. Åberg, SWE	PI 189763	1950
7015	SWE	0.103	0.122	POLIHAR	Maison Andre Blondeau, Bersee, FRA	IHAR 41964	1959
7011	SWE	0.065	0.125	JPNRIB	Sweden	OUU 406	
7014	SWE	0.029	0.251	GBRJIC	E. Craigs, Agric Scientific Services, Edinburgh, UK	B 7758	1966
7013	SWE	0.0	0.0	GBRJIC	National Inst of Agricul Botany, Cambridge (NIAB), UK	B 3421	1952
7016	SWE	0.0	0.0	DEUBAZ	DEUTUMPZFS	BAZ 5612	
7017	SWE		0.0	DEUIPK	Svalöf, Sweden	HOR 2496	1958
Kenia							
1012	DNK	0.584**	0.008	USANSGC	Dr. J. Pryborowski, Dep. Plant Breed., Univ. Krakow, POL	PI 129425	1938
1014	DNK	0.547**	0.006	GBRJIC	Abed Plant Breeding Station, DNK	B 4020	1967
1013	DNK	0.626***	0.0	JPNRIB	Source: UTK	OUU 105	
1002	DNK	0.271*	0.005	INTNGB	KVL, DNK	NGB6935.1	1988
1011	DNK	0.114	0.055	DEUIPK	F. Heine, Schnega/Hann., DEU	HOR 2534	1960
1015	DNK		0.0	POLIHAR	Instutute of Genetics and Plant Breeding, Warsaw, POL	IHAR 41542	1961
Anita							
4022	NOR	0.620*	0.017	INTNGB	From Gunnarn by, Sweden	NGB8185.1	1981
5011	SWE	0.634***	0.0	DEUBAZ	DEUFUNIOB	BAZ 17674	
4021	NOR	0.507**	0.001	INTNGB	NGB (90-62), from Högsby, Sweden	NGB263.2	1981
5012	NOR	0.392**	0.005	RUSVIR	Statens Forsøgsgård, Møystad, NOR	K-19447	1965
5013	NOR		0.0	CZERICP	Norway	03C0600936	1964

*, ** or *** indicate significance at the 5%, 1% and 0.1% levels, respectively

[a] Maximum average genetic distance

[b] P is the probability of obtaining a larger maximum distance by chance

[c] According to passport data from the donor

the breeding company. Concerning the deviating accession 7012, inspection of field books of the N.I. Vavilov Research Institute of Plant Industry revealed Norway to be registered as the country of origin for this seed sample of 'Bonus' collected by a Russian scientist in Norway in 1947. 'Bonus' accession 7012 might therefore derive for the cultivar 'Bonus' from Forus (pedigree: Asplund × Maskin) bred by the State Experimental Station Forus,

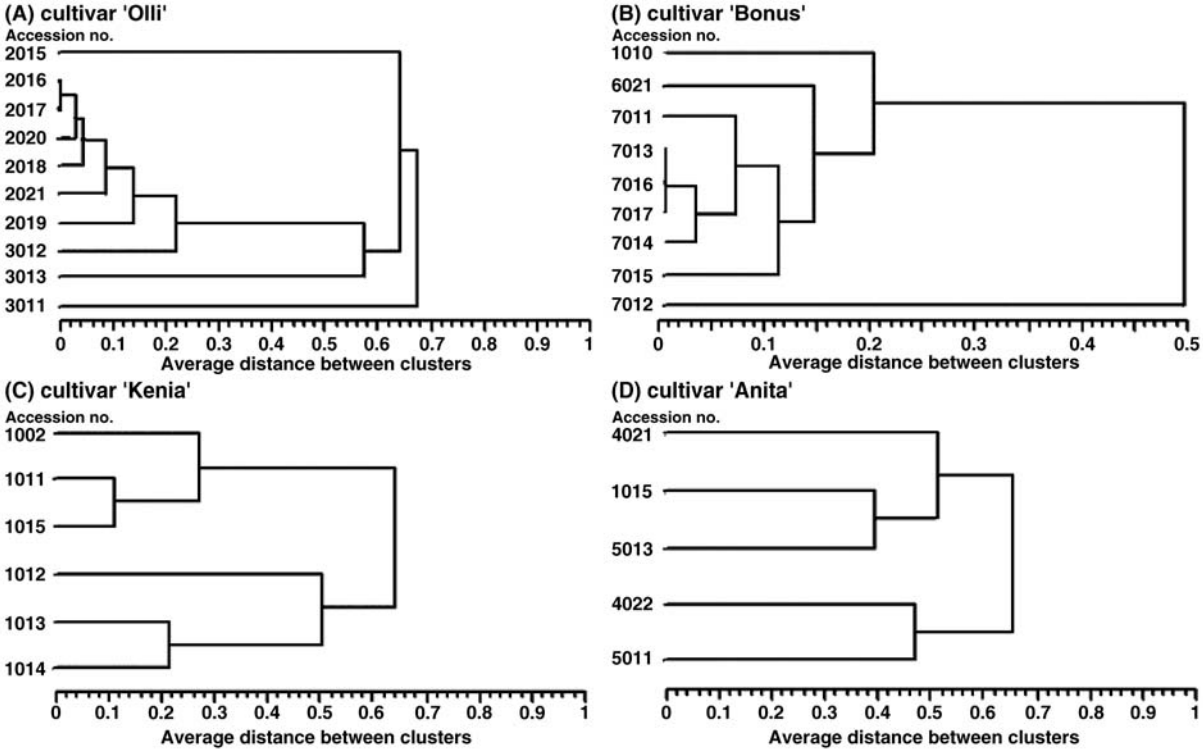

Fig. 2A–D Dendrograms of potential duplicate accessions for four barley cultivars

Norway, and released in 1939 (Strand 1979). This name-duplication of cultivars with completely different pedigrees may be the explanation of why accession 7012 was so different from the rest of the potential duplicate group of the cultivar 'Bonus'.

Analysis of the six accessions named 'Kenia' declared average genetic distances of accessions 1012, 1014, 1013 and 1002 as significant (Table 4). Accession numbers 1011 and 1015 formed a statistically homogeneous pair with a maximum average distance of 0.11. Cluster analysis of 'Kenia', separated the six accessions into two sub-clusters (Fig. 2C), indicating heterogeneity in the potential duplicate group. According to history, the cultivar 'Kenia' has been bred in Denmark by the Abed Plant Breeding Station and released in 1932 (Anonymous 1978). Accession 1012 has been maintained at the National Small Grains Collection of the USA since 1938, when it was donated from Poland. Accessions 1011 and 1015 were donated to Gatersleben, Germany, in 1960 and to Poland in 1961, respectively, from other local institutions with no information about the original Danish donors. In this potentially duplicate group of the cultivar 'Kenia' the most original accession according to passport data could be accession 1002 donated directly from the breeding company to the Nordic Gene Bank through the Royal Veterinary and Agricultural University, Denmark; or accession 1014 was also donated directly from the breeding company to the barley genebank collection of the John Innes Centre in the United Kingdom in 1967. The present analysis based on testing for homogeneity did not support the idea that both these accessions represent the original cultivar 'Kenia'.

In the potential duplicate group named 'Anita' the bootstrap test declared all five accessions 4022, 5011, 4021, 5012 and 5013 to be significantly different (Table 4). The dendrogram of this group of accessions formed two sub-clusters (Fig. 2D) indicating heterogeneity of the potential duplicate accessions. According to historical information on 'Anita' it was originally bred by K. Vik and E. Strand, and released from Vollebekk in Norway in 1962 [pedigree of 'Anita': (Asplund × Ds 295) × Varde] (Strand 1979). Passport data for accession 5011 repatriated from Braunschweig in Germany has Sweden as the country of origin, and accessions 4021 and 4022 have been donated to the Nordic Gene Bank by local Swedish farmers. Accessions 5012 and 5013 have both been delivered directly from Norway according to passport data, indicating that they might both represent the most-original accession. Accession 5012 was donated from the State Experimental Station, Møystad, in Norway in 1965 to the N.I. Vavilov Research Institute of Plant Industry (VIR), Russia, according to original field books in VIR. Accession number 5013 was delivered directly from Norway (donor unknown) to the Czech Republic in 1964 (receiver unknown).

Discussion

Initial identification of potential duplicate groups of crop accessions, as recommended by Hintum and Knüpffer (1995), Hintum and Visser (1995), Greene and Pederson

(1996), Hintum et al. (1996), Phippen et al. (1997), Dean et al. (1999) and Treuren et al. (2001), also proved successful in our research with repatriated barley accessions. Grouping potentially duplicate accessions based on identical or similar names formed a distribution of average genetic distances within groups containing many small distances compared to the distribution formed by unique accessions (Fig. 1). This identification of potential duplicate accessions, however, is not without problems because the distribution of comparisons within the duplicate groups still contains large genetic distances, indicating misclassifications.

An exact definition of the term 'duplicate accession' is still under development. Most authors distinguish between historical duplicates and genetic (Hintum and Knüpffer 1995; Hintum and Visser 1995) or biological (Willner et al. 1998) duplicates. Identification of historical duplicate accessions relies primarily on passport data, indicating their common origin from the same originally collected or bred material without undergoing intentional selection (Willner et al. 1998). Historical duplicate accessions may also be identified as genetic duplicate accessions if, in addition to passport data, genetic comparisons indicate nearly identical genetic composition (Sahu 1989; Hintum and Knüpffer 1995; Hintum and Visser 1995; Virk et al. 1995; Hintum et al. 1996; Willner et al. 1998; Treuren et al. 2001). Historical duplicate accessions, however, may diversify from their original accession in genetic composition during maintenance in ex situ collections (Hintum and Knüpffer 1995; Willner et al. 1998). During seed handling and regeneration, accessions may be subjected to genetic drift, natural or unintentional selection, contamination through cross pollination and seed mixing, or exchange of passport data through incorrect labelling of seed bags or trial plots (Hintum and Knüpffer 1995; Willner et al. 1998), or mutation at the microsatellite loci. For these reasons historical duplicate accessions are seldom completely genetically identical (Hintum and Knüpffer 1995; Hintum and Visser 1995; Treuren et al. 2001).

The amount of genetic diversity acceptable between genetically duplicate accessions is still not well-defined. Such genetic variation between genetically duplicate accessions will depend on both the type of plant material and the methods of maintenance. For highly allogamous species or for accessions maintained under less controlled conditions, such genetic variation between genetically duplicate accessions will be higher than for strictly autogamous species or accessions maintained under well-controlled conditions. For a species like barley it may be assumed that original genetic heterogeneity and diversification, due to conditions of maintenance in bred cultivars, is approximately constant among accessions in collections. For these reasons genetically duplicate accessions originating from a common original accession would be expected to be genetically equally different from each other and from their common original accession, and thus form a genetically homogeneous group. This type of genetic homogeneity is the basis for identification of duplicate accessions forming statistically homogeneous groups in a previous study (Treuren et al. 2001), using the test of between-accession variation against within-accession variation based on the analysis of molecular variance. The present approach using a re-sampling method to distinguish homogeneous groups from groups with deviating accessions avoids the resource-demanding estimation of within-accession variation.

Results from the four selected groups of potential duplicate accessions in the present study illustrate the type of additional information to be gained from genetic analysis of potentially duplicate accessions based on passport data. In some cases like the group of 'Olli', a homogeneous group will include the most-original accessions identified and thus support the conclusion from passport data. A similar situation is found with the group of 'Bonus' where the most-original accession from Sweden according to passport data is included in a homogeneous subgroup of accessions. In this group, however, one of the accessions excluded from the homogeneous group could also be an original accession derived from another cultivar 'Bonus' from Norway with a similar name, but of different origin. The separation of this accession 'Bonus' from the Sweden accessions in the test, support such a hypothesis. The group of 'Kenia' is an example where genetic analysis of homogeneity does not support a candidate of the most-original accession identified by passport data, because this accession is not included in the homogeneous group. The candidate based on passport data may still be the best choice of the most-original accession, but the two accessions comprising the homogeneous group may be considered an alternative for further studies. In the group of 'Anita' two accessions, both identified as the potentially most-original, were tested as significantly different, as did all other accessions in the group. In that case the genetic analysis seriously questions the hypothesis that these accessions identified by passport data are genetic duplicates. Future efforts to answer questions raised by such genetic analyses may include additional passport data and further genetic information obtained from supplementary accessions belonging to the same duplicate group or from related cultivars.

Another advantage of the re-sampling procedure for the detection of statistically homogeneous groups is the fact that the largest average genetic distance accepted in a statistically homogeneous group was below the smallest average genetic distance declared significant in any potentially duplicate group. The cut-off level with an average genetic distance of approximately 0.14 is well in accordance with the visual impression based on the two histograms in Fig. 1. Furthermore, the observed apparent threshold value of approximately 0.14, as the maximum genetic distance between accessions in a homogeneous group, may be a general threshold for declaring two accessions different in barley collections, with a Nordic background analysed with this set of microsatellite markers. This idea, in accordance with the assumption of

a constant level of diversification between duplicate accessions, is supported by data from the present study, where genetic distances between pairs of unique accessions were almost always larger than this threshold. The threshold of the 0.14 genetic distance between pairs of accessions may therefore be used to screen large genebank collections of this barley gene pool for potentially duplicate accessions, using this set of microsatellite markers.

Acknowledgements This research was funded by the Nordic Gene Bank and the Royal Veterinary and Agricultural University of, Denmark (KVL). The authors thank all donors supplying seed samples and information (Table 1). Professor Roland von Bothmer, SLU, Sweden, is thanked for his comments to the manuscript, and Ms Mette Sylvan, Mr B.K. Hansen and Mr J. Bertelsen are thanked for their technical assistance.

References

Aikasalo R (1988) The results of six-row barley breeding and the genetic origin of varieties released. J Agric Sci Finland 60:293–305

Anonymous (1978) Abed Planteavlsstation 1903–1978. Udgivet i anledning af Abed Planteavlsstations 75 års jubilæum af Planteavlsstationen, Dansk Landbrugs Grovvareselskab og Fyens Andels-Foderstofforretning, Denmark

Anonymous (1990) SAS/STAT user's guide, version 6, 3rd edn. SAS Institute Inc, Cary, North Carolina

Brown, AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Cervera M-T, Cabezas JA, Sancha JC, Martinez de Toda F, Martinez-Zapater JM (1998) Application of AFLPs to the characterization of grapevine *Vitis vinifera* L. genetic resources. A case study with accessions from Rioja (Spain). Theor Appl Genet 97:51–59

Christiansen MJ, Andersen SB, Ortiz R (2002) Diversity changes in an intensively bred wheat germplasm during the 20th century. Mol Breed 9:1–11

Dean RE, Dahlberg JA, Hopkins MS, Mitchell SE, Kresovich S (1999) Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with simple sequence repeat (SSR) markers. Crop Sci 39:1215–1221

Diwan N, Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor Appl Genet 95:723–733

Efron B (1979) Bootstrap methods: another look at the jacknife. Ann Statistics 7:1–26

FAO (1996) Report on the state of the world's plant genetic resources for food and agriculture. Food and Agriculture Organization of the United Nations, Rome

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

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

Frankel OH, Brown AHD (1984) Plant genetic resources today: a critical appraisal. In: Holden JHW, Williams JT (eds) Crop genetic resources: conservation and evaluation. Allen and Unwin, Winchester, Massachusetts, USA, pp 249–268

Gilbert JE, Lewis RV, Wilkinson MJ, Caligari PDS (1999) Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theor Appl Genet 98:1125–1131

Greene SL, Pederson GA (1996) Eliminating duplicates in germplasm collections: a white clover example. Crop Sci 36:1398–1400

Hammer K, Diederichsen A, Spahillari M (1999) Basic studies toward strategies for conservation of plant genetic resources. In: Serwinski J, Faberová I (eds) Proc Technical Meeting on the Methodology of the FAO World Information and Early Warning System on Plant Genetic Resources, Food and Agriculture Organization of the United Nations, Rome, pp 1–11

Hintum ThJL van, Knüpffer H (1995) Duplication within and between germplasm collections. I. Identifying duplication on the basis of passport data. Genet Res Crop Evol 42:127–133

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

Hintum ThJL van, Boukema IW, Visser DL (1996) Reduction of duplication in a *Brassica oleracea* germplasm collection. Genet Res Crop Evol 43:343–349

Hintum ThJL van, Brown AHD, Spillane C, Hodgkin T (2000) Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3. International Plant Genetic Resources Institute, Rome, Italy

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Wiel C van de, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazque A, Karp A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3:381–390

Lyman JM (1984) Progress and planning for germplasm conservation of major food crops. FAO/IBPGR Plant Genet Resourc Newslett 60:3–21

Macaulay M, Ramsay L, Powell W, Waugh R (2001) A representative, highly informative 'genotyping set' of barley SSRs. Theor Appl Genet 102:801–809

Mitchell SE, Kresovich S, Jester CA, Hernandez CJ, Szewc-McFadden AK (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. Crop Sci 37:617–624

Ortiz R, Madsen S, Vuylsteke D (1998) Classification of African plantain landraces and banana cultivars using a phenotypic distance index of quantitative descriptors. Theor Appl Genet 96:904–911

Ortiz R, Nurminiemi M, Madsen S, Rognli OA, Bjørnstad A (2002) Cultivar diversity in Nordic spring barley breeding (1930–1991). Euphytica 123:111–119

Osvald H (1959) Åkerns nyttoväxter. AB Svensk Litteratur, Stockholm, pp 101–119

Phippen WB, Kresovich S, Candelas FG, McFerson JR (1997) Molecular characterization can quantify and partition variation among genebank holdings: a case study with phenotypically similar accessions of *Brassica oleracea* var. *capitata* L. (cabbage) "Golden Acre". Theor Appl Genet 94:227–234

Pillen K, Binder A, Kreuzkam B, Ramsay L, Waugh R, Förster J, Léon J (2000) Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. Theor Appl Genet 101:652–660

Plucknett DL, Smith NJH, Williams JT, Anishetty NM (1987) Gene banks and the world's food. Princeton University Press, Princeton, New Jersey, USA

Russell J, Fuller J, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997a) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W (1997b) Discriminating between barley genotypes using microsatellite markers. Genome 40: 442–450

Russell JR, Ellis RP, Thomas WTB, Waugh R, Provan J, Booth A, Fuller J, Lawrence P, Young G, Powell W (2000) A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Mol Breed 6:553–568

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

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

Strand E (1979) Norsk kornforedling og korndyrkning. In: Strand E (ed) Korn er liv Statens Kornforretning, pp 57–137

Struss D, Plieske J (1998) The use of microsatellite markers for detection of genetic diversity in barley populations. Theor Appl Genet 97:308–315

Torp J, Jensen HP, Jørgensen JH (1978) Powdery mildew resistance genes in 106 Northwest European spring barley varieties. Kgl Vet- og Landbohøjsk Årsskr:75–102

Treuren R van, Soest LJM van, Hintum ThJL van (2001) Marker-assisted rationalisation of genetic resource collections: a case study in flax using AFLPs. Theor Appl Genet 103:144–152

Virk PS, Newbury HJ, Jackson MT, Ford-Lloyd BV (1995) The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. Theor Appl Genet 90:1049–1055

Waycott W, Fort SB (1994) Differentiation of nearly identical germplasm accessions by a combination of molecular and morphologic analyses. Genome 37:577–583

Willner E, Sackville Hamilton NR, Knüpffer H (1998) Duplications in forages collections. In: Maggioni L, Marum P, Sackville Hamilton R, Thomas I, Gass T, Lipman E (eds) Report of a Working Group on Forages, 6th Meeting, 6–8 March 1997, Beitostølen, Norway. International Plant Genetic Resources Institute, Rome, Italy, pp 92–95

Zeven AC, Dehmer KJ, Gladis T, Hammer K, Lux H (1998) Are the duplicates of perennial kale (*Brassica oleracea* L. var. *ramosa* DC.) true duplicates as determined by RAPD analysis? Genet Res Crop Evol 45:105–111

Ziegle JS, Su Y, Corcoran KP, Nie L, Mayrand E, Hoff LB, McBride LJ, Kronick MN, Diehl SR (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14:1026–1031