Identification of duplicates by comparing passport data of *Avena* germplasm collections

[NOT REVIEWED PAPER]

Igor G. Loskutov[1], Andrew Ya. Omelchenko[1] & Siegfried Harrer[2]

[1]N.I. Vavilov Institute of Plant Industry, 42-44, Bolshaya Morskaya str., St. Petersburg, 190000, Russia; [2]German Centre for Documentation and Information in Agriculture, Villichgasse 17, 53177 Bonn, Germany.

i.loskutov@vir.nw.ru

Key words: *Avena* species, botanical varieties, collections, duplicates, inter- and intra-specific classification, taxonomy.

Abstract

Rationalization of genebank collections is one of the most significant problems to improve the efficiency of PGR conservation. Duplicates identification in collections held by genebanks is one of the aspect of optimising and systematizing the existing collections. In result of analysing of *Avena* collections of VIR (Russia), IPK and BAZ (Germany) and NGB (Sweden) number of probable duplicates of advanced cultivars were identified. It is concluded that practical work with databases for unification and identification duplicates of the collections have to be used taxonomic approaches inter-specific and intra-specific classification.

Introduction

Creation and appropriate use of passport, characterization and other databases for the *ex situ* collections of plant genetic resources (PGR) should be among the priorities of any genebank. Value of any collection strictly depends on the completeness and correctness of this information. An item (accession) of a collection in any genebank is a plant form, which must be registered and precisely identified. The problem of conservation and identification of global PGR is currently becoming the major one for

many genebanks of the world (Steiner et al., 1997; MacKenzie, 1998; Diederichsen et al., 1999). Most of European genebanks that were set up in the mid-60s and 70s of the 20th century face the problem of accessions regeneration after long-term storage for 20 to 30 years. At the same time, many of them lack strict criteria for accessions identification at the level of a species, to say nothing of their capability to maintain intra-specific diversity.

The intra-specific criteria or botanical varieties are based on clearly visible and easily distinguishable morphological characters. The majority of these criteria systems were created under the guidance of N.I.Vavilov in the 20s and 30s and later developed and refined on the basis of new data from studies of plant diversity (Loskutov, 1999; 2000; 2002). Using botanical varieties it helps safety to maintain morphological purity of each accession.

In view of the above, duplicates identification in the process of databases comparison has recently been receiving an increasing attention (van Hintum, 1994; van Hintum & Knüpffer, 1995a; van Hintum & Visser, 1995b; van Hintum, 2000). At present, for solving this problem mostly the approaches that involve the analysis of computerized database information alone, with no account of characteristics of the conserved seed material, are being employed. This problem is directly linked with the necessity to optimise and systematize the existing collections, and duplicates identification in collections held by genebanks is an aspect of special importance.

Botanical classification is one of the most important component for identifying duplicates. It is supported by the fact that data on specific, and especially intra-specific classification are gaining significance not only for botanical research and breeding purposes, but also for genebanks seeking genetic purity of their maintained *ex situ* seed collections.

According to Wesenberg (1992), 30 main (more than 200 accessions) world collections maintain about 94,500 accessions belonging to the genus *Avena* L. The European *Avena* Database contains, with complete data for the VIR collection, more than 30,500 entries (Bücken & Frese, 1998). This PGR diversity has to be structuring for safety conservation in different genebanks.

Material and methods

In the end of 1999, Vavilov Institute of Plant Industry (VIR, Russia), German Centre for Documentation and Information in Agriculture (ZADI, Germany) and Nordic Gene Bank (NGB, Sweden) agreed on the implementation of a pilot project on the comparison of *Avena* collections conserved in Russia, Germany and Sweden. For identifying duplicates in the collections, the passport databases of four European *Avena* germplasm collections have been compared using the Corel Paradox8 for Windows98. There are 11,918 accessions in the VIR collection, 2,921 in IPK, 1,825 in BAZ and 478 in NGB. The total of over 17,100 entries have been analysed. It equals 62% from the European *Avena* Database and 18% from the total *Avena* collections in the whole world.

For comparing the passport data, the following principles have be applied in this work. The search for duplicated material between germplasm collections of different countries is necessary, in the first place, for the identification of advanced cultivars which have the same name and are stored in different collections. The retrieval of duplicate accessions with the same identification numbers by the cultivar name supposes their morphological identity. It is intra-specific systematics that is used for establishing identity of such accessions. Quite often, accessions of national collections were received from other genebanks. When comparing collections for possible duplication, it is desirable to have the most complete information about the origin of an accession, that is, original name of the accession, catalogue number from the donor genebank, catalogue numbers for this accession in other genebanks of the world, the place of origin and reproduction of the accession.

All the above-mentioned principles were used in our work for comparing all passport data between collections. For the analysis of passport data, a special structure for a joint database has been developed to facilitate accessions comparison by their names and additional information (Table 1 and 2).

This structure allowed the development of a compact database that ensured the comparative analysis of accessions in the mentioned collections.

Results and discussion

For analysing and identifying duplicates, the most important fields are ACCNAME, DONORNUM, OTHERNUM, as well as the SCINAM field for determining identity in terms of classification units (i.e., genus, species, subspecies, etc.).

It was found that in the IPK database the fields DONORNUM (15%) and OTHERNUM (1%) are not sufficiently filled in, while the SCINAM field is populated well, but subtaxa is identified in only 88% of all cases, and 180 accessions have no species identification (i.e. genus *Avena* only). The database at NGB featured all of the above mentioned drawbacks. All accessions are identified at the species level only, and without botanical varieties. In BAZ, the database has sufficiently complete fields DONORNUM (85%) and OTHERNUM (21%), but the field SCINAM (40% for botanical varieties only) is very poor filled, besides that the content is often confusing and/or incorrect.

We found several duplicates within collections (in field ADD identified as D). Within the analysed collections, 130 duplicates have been found: 70 accessions in the IPK collection, 54 in BAZ and only 6 in NGB (Table 3).

The analysis of passport data showed 464 accessions in the joint database to have some confusing information. These accessions require additional checking and consultations of genebank experts.

- 394 accessions have been revealed to coincide by accession name and all other numbers, but differ considerably by the scientific name (in the field ADD marked by "+"): 160 accessions in IPK, 90 in BAZ, 109 in VIR and 35 in NGB (Table 3);
- 70 accessions were with incomprehensible geographical origin and botanical varieties (in field ADD marked by "?"): 42 very strange accessions of the Soviet or Russian origin (40 accessions in BAZ and 2 in IPK), 28 accessions - with not understandable botanical identification (26 accessions in BAZ and 2 in IPK).

The conclusion from the analysis of 4 databases shows that 90% of the VIR database is unique accessions between four compared collections; the amount of these in the IPK database is 70%, and is equal to 50-60% in databases of BAZ and NGB. It

is worth mentioning that quantity of unique accessions were following: VIR – 10770 accessions, IPK – 2080, BAZ – 1120 and NGB – 270.

As we mentioned previously this compact database is about 18% from International *Avena* Database. The quantity of unique entries was analysed by using DONORNUM and OTHERNUM fields, when there was available information about identification numbers from different genebanks and especially from American genebank. This analysis showed that percent of unique accessions compare between other genebanks was lower: VIR – 80%, IPK – 60%, NGB – 45% and BAZ – 25%.

From the analysis of databases, 1,276 accessions have been found to have the same scientific name and be duplicated in 2 (1034 entries), 3 (207 entries), or sometimes in 4 (35 entries) genebanks at a time (Tables 4, 5 and 6).

It was found out that some collections are contaminated, some accessions are misidentified, but most often they are not comprehensively taxonomically determined. In some cases, the computerized botanical identification, which using of most genebanks officers in their database analyses, does not relate to the morphological characterisation of the stored accessions. It shows that some genebanks are not maintained original entries of seeds or herbarium seeds, which could compare with regenerated seed accession and check contaminated material.

Our opinion is that some formal approach targeted at revealing similarly sounding names should be used by the crop expert with utmost care, as absolutely different cultivars may have similar or the same names.

Determination of genetic identity of advanced cultivars with the same parents is also problematic, as the contents of Pedigree fields in databases leaves much to be desired.

Talking of genetic identity of populations or landraces would be premature, as relevant information for these materials is very scanty in passport databases, or quite often is absent in most genebanks of the world.

All discussions about duplicates of gene alleles in any accessions, or about parental duplication are unfounded due to the extreme scarcity of information.

The use of various molecular-biological techniques for the identification of duplicates in collections is hardly reasonable, if these techniques are inefficient and

costly, while present-day collections in the world number dozens and hundreds of thousands of accessions. For example, from more than 31, 000 accessions of European Barley Database only 148 accessions of barley were analysed using molecular marker technique (van Hintum & Visser, 1995b).

Conclusions

Our opinion that duplicates identification within and between germplasm collections of particular crops must be performed by an expert or collection curator specializing in this crop, as these are the people who can, thanks to their knowledge and experience, understand value and significance of a duplicate and come to a well-weighed decision in each particular case.

And so practical work with databases for unification of the collections using of taxonomic approaches inter-specific and intra-specific classification helps us better understand, properly evaluate and carefully maintain the global diversity of plant genetic recourses for future generations.

Acknowledgements

We thank very much Dr. S.Alexanian, VIR, Russia, Dr. F.Begemann and Dr. T.Gladis, ZADI, Germany for helpful comments on the manuscript.

References

Bueken S. & L. Frese. 1998. The European Avena database (EADB) – state of the art. Report of a Working Group on Avena. Fifth meeting, Vilnius, Lithuania. IPGRI, Rome, Italy. 1998, 20-25.

Diederichsen A., Kessler D., Williams D. 1999. Germplasm of the genus Avena at Plant Gene Resources of Canada (PGRC). Oat Newsletter, v.45.

Hintum van T.J.L. 1994. Drowning in the genepool. Managing genetic diversity in genebank collections. 360 pp.

Hintum van T.J.L. 2000. Duplication within and between germplasm collections. III. A quantitative model. Genetic Resources and Crop Evolution, 47, 507-513.

Hintum van T.J.L. & H. Knupffer 1995a. Duplication within and between germplasm collections. I. Identifying duplication on the basis of passport data. Genetic Resources and Crop Evolution, 42, 127-133.

Hintum van T.J.L. & D.L. Visser 1995b. Duplication within and between germplasm collections. II. Duplication in four European barley collections. Genetic Resources and Crop Evolution, 42, 135-145.

Loskutov I.G. 1999. Vavilov and his institute. A history of the world collection of plant genetic resources in Russia. IPGRI, Rome, Italy, 189 p.

Loskutov I.G. 2000. The problems of maintenance of plant genetic recourses and intra-specific taxonomy of plants. Proc. III International Crop Science Congress, Germany, p.128.

Loskutov I.G. 2002. Systematical approaches to the genus Avena L. http://www.vir.nw.ru/avena/sys_app.htm

Steiner A.M., Ruckenbauer P., Goecke E. 1997. Maintenance in genebanks, a case study: contaminations observed in the Nurnberg oats of 1831. Genetic Resources and Crop Evolution., 44: 6, 533-538.

MacKenzie D. 1998. Growing problems in seed banks. New Scientist, 157: 2121.

Wesenberg D.M., Briggle L.W., Smith D.H. 1992. Germplasm collection, preservation and utilization. In: Oat science and technology. Ed. By A.R.Marshall and M.E.Sorrels. Agronomy 33, Am. Soc. Agron., Madison, WI, USA, 793-820.

Table 1. Structure of cultivated joint database.

VIRCATNUM	VIR Catalogue Number
BAZCATNUM	BAZ Catalogue Number
IPKCANTUM	IPK Catalogue Number
NGBCATNUM	NGB Catalogue Number
ACCENAME	Accession name
SCINAM	Scientific name – full botanical name with authority. Following abbreviations are used: sp., subsp., var., convar.
OTHERNUM	Catalogue number(s) in others genebanks of the world
DONORNUM	Catalogue number(s) of donor organization
ORIGCTY	Country of origin
DONCTY	Country of donor
YEAR	Year of accessions entry
ADD	Comments: + - problem accessions, D - duplicates within collection, ? – not understandable origin and botanical identification.

Table 2. Structure of wild oat joint database.

VIRCATNUM	VIR Catalogue Number
BAZCATNUM	BAZ Catalogue Number
IPKCANTUM	IPK Catalogue Number
NGBCATNUM	NGB Catalogue Number
WPBSNUM	Welsh Plant Breeding Station Catalogue Number (UK)
PINUM	Plant Introduction Catalogue Number (USA)
CAVNUM	Canadian Avena Catalogue Number (Canada)
ACCENAME	Accession name
SCINAM	Scientific name – full botanical name with authority. Following abbreviations are used: sp., subsp., var., convar.
OTHERNUM	Catalogue number(s) in others genebanks of the world
DONORNUM	Catalogue number(s) of donor organization
ORIGCTY	Country of origin
DONCTY	Country of donor

Table 3. Duplicates within and between cultivated oat collections

VIR	BAZ	IPK	NGB	ACCENAME	SCINAM	OTHENUM	DONONUM	ORIGCTY	DONORCTY	ADD
	52939			ACACIA	Avena sativa L.		PI220867	AUS	USA	+
11300				ACACIA	A.BYZANTINA C.KOCH.	PI 220867		AUS		+
		AVE 1319		ALAMO	A. saliva L. var. montana Alef.			USA	USA	D
		AVE 975		ALAMO	A. saliva L. var. aurea Koern.		CI 5371	USA	USA	D
10300			NGB 9798	ALGERIBEE	A.BYZANTINA C. KOCH.			AUS		
		AVE 1156		ALGERIBLE	A. sativaj, var. mutica Alef				HUN	
		AVE 901		ANDERS	A. saliva L. var. aristata Krause		6-149		HUN	+
13498				ANDERS	A.BYZANTINA C. KOCH			SWE		+
		AVE 606		ANTHONY	A. saliva L var mulica Aiof	CI 7001	BBA 2627	USA	DEU	D
		AVE 781		ANTHONY	A saliva L var. mulica Alef.		BBA 2556	USA	DEU	D
8665	16815	AVE 231		ANTHONY	A.SATIVA L. VAR.UUTICA	C.I. 2143		USA		D
		AVE 679		APPLER	A. bvzantina Koch subsp. bvzantina	CI 7003	BBA 2629	USA	DEU	D
1874	52328	AVE 786		APPLER	A.BYZANTINA C.KOCH.	C.I. 0775	BBA 2607	USA	USA	D
	41660			ARTEMOWSK	Avena saliva var. mutica ALEF.		94	SUN	DEU	?
	41700			ARTEMOWSK. KRIM 90	Avena sativa var. aurea KOERN.		106	SUN	DEU	?
12239				AVE 0448/63	A:STRIGOSA SCHREB.	AVE 0448		URY	GERMANY	
			NGB 4726	AVOINE NUE GROSSE	Avena nuda L.			DNK		?
2122				AVOINE NUE GROSSE	A.SATIVA L. VAR.INERMIS			FRA		?
	52566			BAGE	Avena sativa L.		PI 185657	ARG	USA	D
	52581			BAGE	Avena sativa L.		PI 189625	BRA	USA	D
10881				BAGE SEL. KLEIN	A BYZANTINA C.KOCH.	D.I.V 040		ARG		
	52931		NGB 9795	BALLIDU	Avena sativa L.		PI 193031	AUS	USA	+
11305		AVE 876		BALLIDU	A.BYZANTINA C. KOCH.	PI 193031		AUS		+
		AVE 651		BAMBO II	A. sativa L. var. obtusata Alef.				HUN	
		AVE 600		BAMBU	A sativa L. var. mutica Alef.			SWE	DDR	?
	16631			BAMBU 1	Avena sativa var. mutica ALEF.		11 12 42	SWE	DEU	?
	52423			BANCROFT	Avena sativa L.		CI 04468	USA	USA	D
	51440			BANCROFT (OW03)	Avena sativa L.		CI 04468	USA	USA	D
		AVE 1232		BEEDEE	A. sativa L. var brunnea Koern.		CI 6752	USA	USA	+
11210				BEEDEE	A.SATIVA L VAR.MUTICA	C.I. 6752		USA		+
			NGB 6356	BELAR	Avena sativa L-			AUS		D
		AVE 1158		BELAR	A. bvzantina Koch subsp. bvzantina			IND	DDR	D
		AVE 227		BELAR	A. sativa L. var. aurea Koern.			GER	DEU	D
		AVE 877		BELAR	A. bvzantina Koch subsp. bvzantina			AUS	AUS	D
8669		AVE 1159		BELAR	A.BYZANTINA C. KOCH.	C.I. 2760	BBA 2559	USA		D

Table 4. Structure of duplicates in joint database.

Genebanks				Number of duplicates
VIR	BAZ	IPK	NGB	
VIR	BAZ			409
VIR		IPK		414
VIR			NGB	87
	BAZ	IPK		93
	BAZ		NGB	18
		IPK	NGB	13
Double duplicates				1034
VIR	BAZ	IPK		153
VIR	BAZ		NGB	23
VIR		IPK	NGB	23
	BAZ	IPK	NGB	8
Triple duplicates				207
VIR	BAZ	IPK	NGB	35
Total duplicates				1276

Table 5. Duplicates between collections of oats wild species.

VIR	BAZ	IPK	NGB	WPBS	USDA	CAV	SCINAM	ORIGCTY	ORIG LOCAL
292	053741			CC 7041		CAV 3862	CANARIENSIS	SPAIN	FUERTEVENT, PTA. ROSA, 4KM
1917	053750					CAV 3874	CANARIENSIS	SPAIN	FUERTEVENTURA, TETIR
1860	053751			CC 7042		CAV 0001	CLAUDA	IRAN	BISTOON, 16KMNOERDL
1907	053754					CAV 0046	CLAUDA	TURKEY	CEYLANPINAR
1862	053756			CC 7045		CAV 0258	DAMASCENA	SYRIA	DAMASKUS, 60 KM NOERDL.
1984	053757					CAV 0259	DAMASCENA	SYRIA	DAMASKUS, 60 KM NOERDL.
1890	053795			CC 7058		CAV 0138	ERIANTHA	SYRIA	DAMASKUS, 40 KM NOERDL.
1872	051490			CC 7046		CAV 0330	HIRTULA	TURKEY	ANTALYA 25 E
1849	051491			CC 7048	PI 337734	CAV 2284	HIRTULA	GREECE	IRAKLION 10S
2034	051492			CC 7050		CAV 4490	HIRTULA	TUNISIA	SEDJENANE
1878	051489			CC 3678			HIRTULA	SPAIN	
1912	052962				PI 282730		LONGIGLUMIS	ISRAEL	
1909	053797					CAV 2835	VENTRICOSA	CYPRUS	
1859	051503			CC 7066		CAV 0150	BARBATA.	IRAN	GHAZVIN, 125 KM N
1919	052946				PI 282709		BARBATA	ISRAEL	
1920	052947				PI 282710		BARBATA	ISRAEL	
1921	052948				PI 282712		BARBATA	ISRAEL	
1926	052953				PI 282720		BARBATA	ISRAEL	
1928	052955				PI 282722		BARBATA	ISRAEL	
1929	052956				PI 282723		BARBATA	ISRAEL	
1934	052961				PI 282729		BARBATA	ISRAEL	
1851	053190			CC 7069		CAV 3908	MAROCCANA	MOROCCO	MAAZIA, 8KM WESTL.
1863	051493			CC 7070		CAV 3926	MAROCCANA	MOROCCO	TIFLET, 20KM SUEDL.
1852	051494			CC7071		CAV 3927	MAROCCANA	MOROCCO	ROMMANI, 10KMOESTL.
1871	051495			CC 7072		CAV 4390	MAROCCANA	MOROCCO	KENITRA, 20 KM NORDWESTL.
1853	051496			CC 7073		CAV 4397	MAROCCANA	MOROCCO	SOUK-EL-AR-BA-DURHARB.
161	052989				Cl 8330		MAROCCANA	MOROCCO	

Table 6. Duplicates of cultivated oat maintained in four genebanks.

VIR	BAZ	IPK	NGB	ACCENAME	SCNAM	OTHERNUM	DONORNUM	ORIGCTY	DONCTY
1994	16755	AVE618	NGB 8712	GOLDEN GIANT	A.SATIVA L. VAR.ELIGULATA	C.I. 1366	11 22 30	FRANCE (FRA)	USA
2036	53108	AVE 928	NGB 9731	JOANETTE	A SATIVAL VAR.BRUNNEA	C.I. 1880	6 -493	FRANCE (FRA)	
7684	41531	AVE 202	NGB 5123	GOPHER OATS	A.SATIVA L. VAR.MUTICA	C.I. 0047.	250	UNITED STATES	
1206	16599	AVE 969	NGB 8756	EARLY MILLER OAT	A.SATIVA L. VAR.ARISTATA		11 12 10	SCOTLAND (GBR)	
9243	16421	AVE 29	NGB 9737	FLAEMINGSTREUE	A.SATIVA L.VAR.AUREA	16421	11 11 5	GERMANY	DEU
9941	41 670	AVE 950	NGB 9747	AJAX	A.SATIVA L. VAR.MUTICA	CN 04562	385	CANADA (CAN)	CAN
10028	41587	AVE 814	NGB 9754	CLINTON	A SATIVA L. VAR AUREA	C.I. 3971		ALASKA (USA)	USA
10265	41578	AVE 954	NGB 9802	GARRY	A.SATIVA L. VAR ARISTATA	CAN 0809	402	CANADA (CAN)	
10296	41524	AVE 320	NGB 5122	EXPRESS	ASATIVA L. VAR.MUITCA		167	NETHERLANDS	DEU
10386	16610	AVE 709	NGB 4900	BLENDAHAVRE	A SATIVA L. VAR MUTICA		11 12 21	SWEDEN (SWE)	
10892	16673	AVE 698	NGB 7011	CIVENA	A.SATIVA L. VAR AUREA			NETHERLANDS	
10918	16618	AVE 1006	NGB 2707	JUHA	A.SATIVA L. VAR.MUTICA		11 12 29	FINLAND (FIN)	
10984	47088	AVE 1408	NGB 9806	FORWARD	A SATIVA L. VAR.OBTUSATA	C.I. 2242		GREAT BRITAIN	
11379	16667	AVE 1181	NGB 6376	ASTOR	A.SATIVA L. VAR.ARISTATA	AVE 1181		NETHERLANDS	
11406	16642	AVE 704	NGB 8764	CONDOR	A.SATIVA L. VAR.MUTICA		11 12 53	NETHERLANDS	
11523	16630	AVE 1280	NGB 4902	-INDA	A SATIVA L. VAR MUTICA		11 12 41	SWEDEN (SWE)	
11529	16595	AVE 1323	NGB 366	HANNES	ASATIVA L. VAR.MUTICA			FINLAND (FIN)	
13404	30354	AVE 2979	NGB 12270	ALFRED	A.SATIVA L. VAR.MUTICA		3602-07	NETHERLANDS	