

Current taxonomic composition of European genebank material documented in EURISCO

Theo van Hintum^{1*} and Helmut Knüpffer²

¹Centre for Genetic Resources, Wageningen University and Research Centre, PO Box 16, NL-6700 AA Wageningen, The Netherlands and ²Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), D-06466 Gatersleben, Germany

Received 14 April 2010; Accepted 26 May 2010 – First published online 21 June 2010

Abstract

Taxonomy plays an essential role in genebank documentation. It is often the first level at which users search material, and it determines the protocols used in the management of collections. Especially, when plant genetic resources information is pooled in systems such as EURISCO, the European catalogue of *ex situ* plant genetic resources, problems regarding technical handling of taxonomic nomenclature, such as lack of standardization and low quality of data, become apparent. These problems were studied by analysing the content of EURISCO and mapping the taxon names in EURISCO on those used in the United States Department of Agriculture genebank system GRIN-Tax. Thus, the number of spelling errors and the level of standardization could be quantified and improved. An analysis of the content of EURISCO was made, showing a highly unbalanced distribution over crops: 50% of the accessions belong to ten genera only. Mapping EURISCO on the crops listed in Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture showed that 67% of the accessions in EURISCO belong to crops in that list.

Keywords: EURISCO; genebanks; GRIN-Tax; taxonomy

Introduction

Taxonomy, the science of classification, plays an essential role in genebank documentation. The Linnaean taxonomy of plants provides the first key to users seeking material of a particular species or group of species for use in their scientific or breeding programme, but it also determines the protocols to be applied in the maintenance of the material by the genebank curators. A proper classification and naming of the genebank material is thus a prerequisite for its utilization and handling.

In any attempt to bring together information about germplasm holdings from numerous sources, the problem of harmonising or standardising scientific nomenclature between the data sources arises. This is true for

both taxon-level data sources (for example, when compiling a continental flora from numerous country floras that use different taxonomy and nomenclature) and accession-level data sources (e.g., in the development of crop-specific databases from accession data provided by individual genebanks that use different classification systems) and is due to the use of diverse nomenclatures in different collections (Knüpffer, 2009).

The main problem regarding the use of taxonomic classification in genebanks is the fact that there is disagreement among scientists in selecting the taxonomic system, and the nomenclature, to be used. New scientific insights urge taxonomists to create new and better classifications, but not all colleagues will agree, and not all users might adopt the new and possibly better systems. A recent example is the renaming of tomato from *Lycopersicon esculentum* to *Solanum lycopersicum* (Spooner *et al.*, 1993), resulting in one-fourth of the tomato accessions in European genebanks being called

*Corresponding author. E-mail: theo.vanhintum@wur.nl

Solanum, and the rest *Lycopersicon*. The resulting problems for genebank documentation were also noted by van Veller *et al.* (2008). Other well-known examples of complexity due to synonymy involve the genus *Aegilops* that, according to some taxonomists, had to be included in *Triticum* (Bowden, 1959), which a few others did not agree (Gupta and Baum, 1986). The continuing regrouping of wild potato species in the genus *Solanum* (e.g., Spooner and van den Berg, 2004) is another example. The result is 'confusion' amongst users and curators. This confusion is further increased by the frequent occurrence of spelling mistakes and plain errors in the use of the nomenclature, such as the use of a family name instead of the genus name.

These problems with taxonomy in the context of plant genetic resources (PGR) became very visible in 2003 when the passport data of the germplasm collections in Europe were combined into a single system: EURISCO. EURISCO is a web-based catalogue that provides information about *ex situ* plant collections maintained in Europe (EURISCO, 2010). As of 1 March 2010, a total of 1,119,348 accessions from 39 European National Inventories covering 304 individual genebank collections were documented in this system. EURISCO includes five fields for taxonomic information: a field for the genus, one for the species and one for the infra-specific name, called here sub-taxon, plus two fields for the author citations for the latter two. The genus name was described on the list of the Food and Agricultural Organization (FAO)/International Plant Genetic Resources Institute (IPGRI) multi-crop passport descriptors on which EURISCO was based as '*Genus name for taxon, in Latin. Initial uppercase letter required*', the species as '*Specific epithet portion of the scientific name, in Latin, in lowercase letters. Following abbreviation is allowed: "sp."*' and, finally the sub-taxon as '*Sub-taxa can be used to store any additional taxonomic identifier, in Latin. Following abbreviations are allowed: "subsp." (for subspecies); "convar." (for convariety); "var." (for variety); "f." (for form)*' (FAO/IPGRI, 2001). The genus name is one of four mandatory descriptors in EURISCO; if this field is empty, the record is rejected; if it is not empty, the record can be accepted. The content of this field, as that of the other four taxonomic fields, is not checked against controlled vocabularies that would provide correct spelling and grouping of synonyms under a preferred scientific name.

The present study is part of an effort to improve the searchability in EURISCO by standardising the scientific names through enhancement of their quality, which will eventually lead to the development of a tool for EURISCO to map most occurring scientific names to preferred ones. In addition, it aims at providing a consistent classification of EURISCO accessions into 'crops' or 'crop groups' as a

prerequisite for proper handling of characterization and evaluation data, and for compatibility with GENESYS (a worldwide PGR information system under development) in this respect. Finally, it also aims at classifying the material documented in EURISCO into Annex-1 crops and non-Annex-1 crops. This classification is based on the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) (FAO, 2002), a legally binding instrument aiming at the conservation and sustainable use of Plant Genetic Resources for Food and Agriculture and the fair and equitable sharing of benefits derived from their use. The crops covered by this Treaty are listed in its Annex 1.

EURISCO is far from being complete. The draft Report on the State of the World's PGRFA (FAO, 2010) estimates the number of PGR accessions in Europe at 1,735,407. A list of over one million European genebank accessions in EURISCO can, however, give a good overview of the situation regarding the taxonomic classification in European genebanks: the *lingua franca* or Babylonian confusion?

Material and methods

Data

The complete content of EURISCO was made available to the authors on request on 27 January 2010 by Milko Skofic of Bioversity International (Rome, Italy) as a zipped comma separated file. It included 1,049,460 accessions, from 36 national inventories. These data were loaded in Excel; all manipulations and calculations were done in Excel 2007, when necessary using Visual Basic for Applications (VBA).

As an external reference for the taxonomical nomenclature, the taxonomy of the Genetic Resources Network (GRIN) of the United States Department of Agriculture National Plant Germplasm System was used (GRIN-Tax, 2010). This well-curated system of taxa and synonyms is the most authoritative and most complete system for cultivated and other economically important plant taxa available and used as taxonomic reference in GRIN. There are a number of other taxonomic databases that could have been used as checklist, e.g. 'Mansfeld's World Database of Agricultural and Horticultural Crops' (IPK, 2010) based on the book edition (Hanelt and IPK, 2001); however, GRIN-Tax was considered very appropriate for the purpose. Mansfeld's database deals only with cultivated species (except for ornamental and forestry plants), but contains some species that are not documented in GRIN-Tax, and also includes some synonyms not found in GRIN-Tax. Other extensive online lists of scientific plant names, such as the International Plant Names Index (IPNI, 2010) or the Integrated Taxonomic

Information System (ITIS, 2010) do not have a particular focus on economic plants and have therefore been consulted in singular cases only to check the spelling of scientific names of wild plants.

The GRIN-Tax genus and taxon tables were downloaded on 5 March 2010, and loaded in Excel. The genus table concerned 27,855 records with 25,934 distinct genera, some occurring more than once with different author citations and some with multiple entries to accommodate associated infrageneric names. The taxon table contained 95,089 taxa, each with its name, author citation and a link to the 'preferred taxon name', the GRIN-accepted name for that taxon. Some genera in the taxon list were listed without species but with the addition 'sp.', and this designation sometimes existed together with other named species; however, not all genera were listed this way.

Given the importance of the ITPGRFA, the content of EURISCO was matched with the crops listed in its Annex 1. Since Annex 1 is not a clean list as it does not include the scientific names of all genera and species, and since it uses terms such as '*Artocarpus*, Breadfruit only' and 'genus *Solanum* Section *melongena* included', a list of taxa related to each of the Annex 1 Crops was created. This list includes those taxa of genera completely covered by Annex 1, and an additional list with genus–species combinations in the cases where particular species of certain genera were either included or excluded.

Data processing

From the 1,049,460 accessions in EURISCO, all 44,584 distinct genus–species–subtaxon combinations were extracted with their frequencies using custom-made VBA procedures.

These genus–species–subtaxon combinations were 'cleaned' in a four-step procedure:

- (1) The format was corrected in terms of case; all fields were transformed to lower case, except for the genus with an initial capital. For example, the genus name 'AEGILOPS' was replaced with '*Aegilops*'.
- (2) The structure of the genus and species fields was corrected by deleting everything after a space, unless this field contained an 'x' (or 'X') preceding a hybrid genus or species name. This usually implied the removal of author citations or other undesirable additions. For example, '*Vicia* L.' became '*Vicia*'.
- (3) In cases where the species field was empty, 'sp.' was added.
- (4) The genera and most frequent genus–species combinations were checked against names occurring in

GRIN Taxonomy with the Taxonomic Nomenclature Checker (Bioversity, 2010a), and the most obvious mistakes were corrected. For example, '*Phaselous*' was replaced with '*Phaseolus*'.

The genus–species–subtaxa combinations on the resulting list were matched with the GRIN-Tax data. In the cases without match, a manual inspection followed. The most frequent non-matching cases were corrected if the deviation from GRIN-Tax was obvious. Especially, the hybrid genera and species required much attention in this process. For example, the hybrid genus 'X *Triticosecale*' occurred in the list, after the first three steps of cleaning as 'X *Triticosecale*', '*Triticosecale*', '*Triticale*', '*Triticocecale*', '*Xtriticosecale*' and '*Xtriticale*'. In the cases of non-matching hybrid genera, it was checked whether the genus name without the preceding 'X' could be found in GRIN-Tax. If this was the case the genus name was replaced accordingly, e.g. 'x *Sorghum*' was replaced with '*Sorghum*'. Finally, it appeared that not all generic names in the GRIN-Tax genus name list appeared in the taxon list; therefore, a match of the latter list with genus names was also made.

The match with the taxa of the Annex 1 of the ITPGRFA could be easily made based on the list that was compiled with genera and species included in this Annex 1. However, the creation of this list was not obvious. For example, the genus *Aegilops* is not explicitly mentioned, whereas 'Wheat – *Triticum et al.* including *Agropyron*, *Elymus* and *Secale*' is. In this case, *Aegilops* was considered part of '*Triticum et al.*'

Results

The downloaded EURISCO dataset contained 1,049,460 accessions, with 5,385 distinct genus names, 34,668 distinct genus–species combinations and 44,584 genus–species–subtaxon combinations. After cleaning, as described above, these numbers had decreased to 5,264 genus names, 33,463 genus–species combinations and 42,661 genus–species–subtaxon combinations. A match with the taxa in GRIN-Tax, where the genus, species and subtaxon names were simply concatenated with a space in between, showed that 37% of the uncleaned taxa and 41% of the cleaned taxa matched, respectively corresponding to 57 and 76% of the accessions.

The cleaning was an exercise that could be performed automatically (with a VBA script), whereas the correcting of the spelling errors requires time and some knowledge of taxonomy. Frequently occurring spelling errors were based on:

- (1) doubling of consonants (e.g. *Ocimum* should be *Ocimum*)
- (2) the use of 'i' instead of 'y' (e.g. *Pinus silvestris* should be *P. sylvestris*, *Poligonum* should be *Polygonum*) and other single letter alterations, apparently most frequently occurring due to misinterpretation of Latin letters in genebanks with a working language using the Cyrillic alphabet
- (3) the use of the ending 'i' instead of 'ii' or *vice versa* (e.g. *Helianthus maximiliani* should be *H. maximilianii*, *Aegilops vavilovi* should be *Ae. vavilovii*, but *Abutilon theophrastii* should become *A. theophrasti*)
- (4) the use of the ending 'ae' instead of 'aea' (e.g. *Althae* should be *Althaea*)
- (5) the use of the wrong gender in species epithets, e.g. 'um' or 'a' instead of 'us' (e.g. *Cucumis sativum* should be *C. sativus*)
- (6) the use of the Latin ending 'um' instead of the Greek 'on' or *vice versa* (e.g. *Agropyrum* should be *Agropyron*).

Sometimes, more structural changes were required such as the correction of the probably erroneous genus name for cotton *Gossypeae*, that was either the misspelled tribe name *Gossypieae* that should not have been used here, or simply the misspelled proper genus name *Gossypium*. It might also be based on a locally used taxonomic system, but in any case *Gossypea* was not

used apart from one genebank where the 6,181 accessions with that name are maintained.

At the end of the cleaning and correcting process, the genus name of 3.7%, the species name of 15.8% and the subtaxon of 8.9% of the accessions had changed. In total, the names of 24.8% of the accessions were corrected.

The distribution of accessions by genus was highly uneven, as could be expected. Fifty percent of the EURISCO accessions belong to only ten genera (*Triticum* 16%, *Hordeum* 9%, *Zea* 4%, *Phaseolus* 4%, *Avena* 3%, *Pisum* 3%, *Solanum* 3%, *Vicia* 2%, *Vitis* 2% and *Malus* 2%); the 60 largest genera with respect to number of accessions in EURISCO are listed in Table 1. With only 191 genera, 95% of the EURISCO accessions can be covered; all these genera are accepted in GRIN-tax except for the genus 'Melo' (usually included in *Cucumis*) with 316 accessions (from three east European countries) on position 139. This implies that the remaining 5,073 genera, or 96% of all genera in EURISCO, cover only 5% of the accessions; 1,655 of these only with one accession.

Obviously, all accessions had a genus name, since it is a mandatory field (three accessions were of the genus 'Mixture'). The number of accessions without a species name was 86,989, or 8.3% of the EURISCO accessions.

The frequency distribution of species names was uneven, similar to that of genera. The top ten genus–species combinations comprised 40% of the accessions, with *Triticum aestivum* in the lead with 12% of the accessions followed by *Hordeum vulgare* (8%) and

Table 1. List of the 60 largest genera in EURISCO^a

Genus	Number of accessions	Genus	Number of accessions	Genus	Number of accessions
<i>Aegilops</i>	10,059	<i>Festuca</i>	12,478	<i>Phaseolus</i>	44,382
<i>Agrostis</i>	1,683	<i>Fragaria</i>	2,423	<i>Phleum</i>	5,287
<i>Allium</i>	9,507	<i>Glycine</i>	15,834	<i>Pisum</i>	31,462
<i>Amaranthus</i>	1,684	<i>Gossypium</i>	8,574	<i>Poa</i>	4,656
<i>Anethum</i>	1,418	<i>Helianthus</i>	6,735	<i>Prunus</i>	28,426
<i>Arabidopsis</i>	1,636	<i>Hordeum</i>	99,503	<i>Pyrus</i>	9,927
<i>Arachis</i>	2,827	<i>Juglans</i>	1,293	<i>Raphanus</i>	3,196
<i>Avena</i>	33,857	<i>Lactuca</i>	10,354	<i>Ribes</i>	3,029
<i>Beta</i>	7,677	<i>Lathyrus</i>	4,480	<i>Secale</i>	12,931
<i>Brassica</i>	21,746	<i>Lens</i>	7,827	<i>Sesamum</i>	1,941
<i>Bromus</i>	1,505	<i>Linum</i>	19,450	<i>Solanum</i>	45,378
<i>Cannabis</i>	1,326	<i>Lolium</i>	11,519	<i>Sorghum</i>	6,572
<i>Capsicum</i>	9,385	<i>Lotus</i>	1,704	<i>Spinacia</i>	1,513
<i>Cicer</i>	10,039	<i>Lupinus</i>	11,827	<i>Trifolium</i>	21,812
<i>Citrullus</i>	3,961	<i>Malus</i>	23,753	<i>Triticum</i>	167,603
<i>Cucumis</i>	11,064	<i>Medicago</i>	10,338	<i>Vicia</i>	25,441
<i>Cucurbita</i>	10,879	<i>Nicotiana</i>	6,258	<i>Vigna</i>	4,642
<i>Dactylis</i>	12,144	<i>Oryza</i>	6,954	<i>Vitis</i>	23,885
<i>Daucus</i>	4,646	<i>Panicum</i>	10,529	<i>X Triticosecale</i>	13,150
<i>Fagopyrum</i>	3,527	<i>Papaver</i>	5,065	<i>Zea</i>	46,075

^a Since the total number of accessions is 1,049,460, these 60 genera cover 89.4% of the accessions in EURISCO, with a range of 16.0% belonging to the genus *Triticum* to 0.1% of the genus *Juglans*.

Zea mays (4%). To cover 95% of the accessions, 2,412 genus–species combinations were required. Of these, 258 are not known in GRIN-Tax, corresponding to only 9,696 accessions. The most frequent unknown combinations are ‘*Sorghum birse*’ (1,742 accessions from one country), *Fragaria ananassa* (832 accessions from two countries referring to *Fragaria* × *ananassa*) and *Brassica capitata* (668 accessions from a number of national inventories belonging to *Brassica oleracea*).

Matching the taxa to GRIN-Tax showed that out of the 5,264 genus names, 186 are not found in GRIN-Tax, representing only 352 accessions (an average of 1.9 accessions/genus). At the species level, out of the 33,463 distinct genus–species combinations, less than half, 16,457 combinations, were in GRIN-Tax (of which 6% consisted of a genus name only). However, these 49% of the names represented 96.8% of the accessions. Similar to the genus names, the 14,867 species names not known in GRIN-Tax represented on average only 2.1 accessions per species. The genera with the largest number of species that could not be matched with GRIN-Tax are: *Eucalyptus* (368 accessions in 344 species), *Silene* (303/156), *Acacia* (270/232), *Carex* (211/150) and *Senecio* (201/159). The majority of these species names (972 represented by 1,258 accessions) is found in a single genebank from the UK, the Millennium Seed Bank at Kew, focussing on wild species from various regions of the world, aiming at covering half of the known plant species worldwide (Ian Thomas, pers. commun.). At the subtaxon level, the number of names that could be matched was low; of the 33.3% of accessions that had a sub-specific epithet only 12.5% could be matched.

To determine the applicability of the ‘preferred taxon name’ concept for the improvement of the access to

EURISCO, the largest part of the taxon name that could be matched to GRIN-Tax was determined for each accession, and the corresponding ‘preferred taxon name’ was determined. In 41% of the names, corresponding to 76% of the accessions, the complete taxon name, i.e. the combination of genus, species and subtaxon as far as available, was found in GRIN-Tax. When only the largest part that could be matched was considered, 99% of the accessions could be matched with at least the genus name. This concerned 17,821 distinct taxa, some consisting of only a genus name (6%), most of a genus–species combination (84%), or of a complete triplet including a sub-specific epithet (9% of the names). Since for each of the taxa appearing in GRIN-Tax also a ‘preferred taxon name’ was listed in GRIN-Tax, it was possible to replace the taxon name with the preferred taxon name; this decreased the total number of distinct taxa in EURISCO by only 8% to 16,380. However, it could be observed that for some of the larger agronomically important taxa, the ‘preferred taxon name’ brought together some important synonyms. This is illustrated in Table 2 for the taxon ‘*T. aestivum* subsp. *aestivum*’.

If the cleaned EURISCO taxon list was matched with the species of Annex 1 of the ITPGRFA, it could be shown that 66.7% of the accessions in EURISCO belong to species occurring on Annex 1. The list of the 25 largest crops of Annex 1 with respect to number of accessions in EURISCO is provided in Table 3.

Discussion

Taxonomy in genebanks is considered a problematic area by many genebank staff. For example, in the 1998

Table 2. Illustration of the use of the GRIN-Tax ‘preferred taxon name’ for bringing synonyms together in one taxon (example: *Triticum aestivum* L. subsp. *aestivum*^a)

EURISCO taxon names	Number of accessions
<i>Triticum aestivum</i> subsp. <i>aestivocompactum</i> E. Schiem.	2,406
<i>T. aestivum</i> L. subsp. <i>aestivum</i>	16,133
<i>T. aestivum</i> var. <i>cinereum</i> (Dekapr.) Mansf.	14
<i>T. aestivum</i> var. <i>graecum</i> (Körn.) Hayek	2,208
<i>T. aestivum</i> var. <i>leucospermum</i> (Körn.) Farw.	222
<i>T. aestivum</i> var. <i>nigraristatum</i> (Flaksb.) Filat.	8
<i>T. aestivum</i> var. <i>pseudoturcicum</i> (Vavilov) Mansf.	79
<i>T. aestivum</i> var. <i>pyrothrix</i> (Alef.) Mansf.	172
<i>T. aestivum</i> var. <i>suberythrospermum</i> (Vavilov) Mansf.	154
<i>Triticum hybernum</i> L.	1
<i>Triticum muticum</i> Schübl.	15
<i>Triticum vulgare</i> Vill.	1

^a Author citations according to GRIN-Tax.

Table 3. The 25 crops of Annex 1 with the highest number of accessions in EURISCO, and an 'other crops' category^a

Crop	Number of accessions
Wheat	178,407
Barley	99,503
Maize	45,957
Grass forages	44,739
Beans	42,394
Legume forages	38,778
Oat	33,857
Pea	31,462
Brassica complex	29,359
Faba bean/vetch	25,441
Apple	23,753
Potato	17,412
Triticale	13,150
Rye	12,931
Chickpea	10,041
Lentil	7,827
Beet	7,677
Rice	6,954
Sunflower	6,735
Sorghum	6,572
Carrot	4,646
Cowpea <i>et al.</i>	4,642
Strawberry	2,423
Eggplant	1,892
Grass pea	1,042
Other crops	2,396

^a The crop names are the names used in the International Treaty on Plant Genetic Resources for Food and Agriculture. The material on this list covers 66.7% of the accessions in EURISCO.

publication about the creation of the European *Brassica* Central Crop Database (Boukema *et al.*, 1998), a database that aimed at combining passport data of all accessions in European *Brassica* collections, the authors list the names under which they received the broccoli accessions: *B. oleracea botrytis italica*, *B. botrytis italica*, *B. oleracea botrytis cymosa*, *B. oleracea* convar. *botrytis* var. *italica*, *B. oleracea italica* and *B. oleracea* var. *italica*. When the EURISCO database was created in 2003, this became very visible; a search of EURISCO was and still is quite difficult since the desired accession might appear under a number of different names, a highly undesirable situation that needs to be resolved.

In the analysis described in this paper, it appears, however, that the problem is not as big as it might seem. The distribution of accessions over taxa is highly uneven, which implies that in order to improve the situation for most of the accessions, attention needs to be given only to a limited number of taxa. Furthermore, GRIN-Tax provides a freely accessible system of synonymy pointing to 'preferred taxa', which is also implemented

in the Taxonomic Nomenclature Checker (Bioversity, 2010a) where large lists of names can be checked for synonyms. This system could be used as a reference for searches in EURISCO that would allow 'translating' misspelled names and synonyms into a preferred name, thus avoiding the problems caused by the use of different classification systems and the occurrence of spelling errors. Names not found in GRIN-Tax can also be checked against the Mansfeld Database (IPK, 2010) using the Taxonomic Nomenclature Checker (Bioversity, 2010b). Thus, 352 genus–species combinations (corresponding to 1,683 accessions in EURISCO) that could not be found in GRIN-Tax could be matched directly, or after correcting obvious spelling errors, with names occurring in the Mansfeld Database. Other available online nomenclature checkers, such as that of TAXAMATCH (2010) that allows fuzzy matching (both phonetic and non-phonetic) of lists of scientific names with various lists of organism names (Rees, 2008) were not used so far.

An important observation in the study was the low quality of a considerable part of the taxonomic data; errors of all imaginable types could be observed. EURISCO might play a role in the reduction of such errors; identifying the errors and giving feedback to the data donors is expected to act as an incentive to correct mistakes and to, perhaps, adopt standard nomenclature. Clear recommendations regarding the formatting of the names of problematic taxonomic groups, such as hybrid taxa (with hybrid names such as *xTriticosecale* or *Allium x proliferum* or hybrid formulas like *Citrus aurantium x Fortunella japonica*) could also help in improving the standardization and quality of taxonomic names in genebanks.

The newest 'Report on the State of the World's PGRFA' estimates the number of PGR accessions in Europe at 1,735,407 (FAO, 2010). The true number is likely to be lower, since this report is largely based on the FAO database (WIEWS, 2010) that, due to inherent curation problems, includes material that is either not publicly available or does not exist anymore. Over 60% of these listed accessions are included in EURISCO. An important known omission in EURISCO is France, which only included 3,589 of its 249,389 accessions (estimation of the previous Report on the State of the World's PGR (FAO, 1996)). Harmonization of taxon names allowed the creation of a 'cleaned' overview of the content of EURISCO (Table 1), and thus an overview on Europe's PGR. This overview shows a remarkable distribution over crop groups. The small grains with 31% of the accessions show a large domination; however, this domination is not as large as could be expected, given the ease of conservation and their importance in scientific research. The position of *Zea* in third place is quite noteworthy, being a cross-pollinated large-seeded genus, and thus

difficult to maintain in PGR collections. Also, the dominant position of the legumes is remarkable, with *Phaseolus* at the fifth place, *Pisum* at the seventh place and *Vicia* at the ninth, together covering 9.7% of the European accessions. Also notable is the fact that the fruit tree genera *Prunus* and *Malus* have many accessions (52,179).

Acknowledgements

The authors acknowledge Milko Skofic and Sonia Dias, both working for EURISCO at Bioversity International, Rome, for their support for and comments to the paper, Dag Terje Filip Endresen (Nordic Genetic Resources Centre, Alnarp, Sweden) and Renato di Giovanni (CRIA – Centro de Referência em Informação Ambiental, Campinas, Brasil) for providing useful information, and John Wiersema (USDA) and an anonymous reviewer for their helpful comments on an earlier version of this manuscript.

References

- Bioversity (2010a) Taxonomic nomenclature checker – GRIN. Available at <http://pgrdoc.bioversity.cgiar.org/taxcheck/grin/> (accessed 2 February 2010).
- Bioversity (2010b) Taxonomic nomenclature checker – Mansfeld. Available at <http://pgrdoc.bioversity.cgiar.org/taxcheck/mansfeld/> (accessed March 2010).
- Boukema IW, van Hintum TJL and Astley D (1998) The European Brassica database. *Acta Horticulturae* 459: 249–254.
- Bowden WM (1959) The taxonomy and nomenclature of the wheats, barleys and ryes and their wild relatives. *Canadian Journal of Botany* 37: 130–136.
- EURISCO (2010) Web catalogue providing access to all *ex situ* PGR information in Europe. Available at <http://eurisco.ecpgr.org/> (accessed 15 March 2010).
- FAO (1996) *Report on the State of the World's Plant Genetic Resources*. Rome: Food and Agricultural Organization of the UN, p. 511.
- FAO (2002) *International Treaty on Plant Genetic Resources for Food and Agriculture*. Rome: Food and Agricultural Organization of the UN, p. 45. Available at <ftp://ftp.fao.org/docrep/fao/011/i0510e/i0510e.pdf>
- FAO (2010) DRAFT Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture. Available at http://typo3.fao.org/fileadmin/templates/agphome/images/PGR/cgrfa_wg_pgr_409_Inf1dr.pdf
- FAO/IPGRI (2001) *FAO/IPGRI Multi-Crop Passport Descriptors December 2001*. Rome: FAO, p. 4.
- GRIN-Tax (2010) *USDA, ARS, National Genetic Resources Program. Germplasm Resources Information Network – (GRIN) (Online Database)*. Beltsville, MD: National Germplasm Resources Laboratory. Available at <http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl?language=en> (accessed 5 March 2010).
- Gupta PK and Baum BR (1986) Nomenclature and related taxonomic issues in wheats, triticales and some of their wild relatives. *Taxon* 35: 144–149.
- Hanelt P, Institut für Pflanzengenetik und Kulturpflanzenforschung Gatersleben (eds) (2001) *Mansfeld's Encyclopedia of Agricultural and Horticultural Crops*, 6 vols, 1st Engl. ed. Springer, Berlin, Heidelberg, New York, etc. (LXX + 3645 pp.)
- IPK (2010) Mansfeld's world database of agricultural and horticultural crops. Available at <http://mansfeld.ipk-gatersleben.de/> (accessed March/April 2010).
- IPNI (2010) The international plant names index. Available at <http://www.ipni.org/ipni/plantnamesearchpage.do> (accessed March/April 2010).
- ITIS (2010) Integrated taxonomic information system. Available at <http://www.itis.gov/> (accessed March/April 2010).
- Knüpfper H (2009) *Triticeae* genetic resources in *ex situ* genebank collections. In: Feuillet C and Muehlbauer G (eds) *Genetics and Genomics of the Triticeae. Plant Genetics and Genomics: Crops and Models 7*. Springer/Dordrecht/Heidelberg/London/New York: Springer Science + Business Media, LLC 2009, pp. 31–79.
- Rees T (2008) Applications of fuzzy (approximate string) matching in taxonomic database searches, with an example multi-tiered approach (Extended abstract). In: Worcester T, Bajona L and Branton B (eds) *Proceedings of a Conference on Ocean Biodiversity Informatics*, 2–4 October 2007. Bedford Institute of Oceanography (CSAS/SCCS Proceedings Series 2008/024), pp. 12–14.
- Spooner DM and van den Berg RG (2004) An analysis of recent taxonomic concepts in wild potatoes (*Solanum* sect *Petota*). *Genetic Resources and Crop Evolution* 39: 23–37.
- Spooner DM, Anderson GJ and Jansen RK (1993) Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (*Solanaceae*). *American Journal of Botany* 80: 676–688.
- TAXAMATCH (2010) TAXAMATCH – fuzzy matching algorithm for genus and species scientific names. Available at <http://www.cmar.csiro.au/datacentre/taxamatch.htm> (accessed March/April 2010).
- van Veller MGP, Hoekstra R and van Dooijeweert W (2008) Implications of taxonomic inconsistencies for plant genetic resources documentation. *Acta Horticulturae* 799: 143–146.
- WIEWS (2010) *World information and early warning system on PGRFA*. Rome: FAO. Available at <http://apps3.fao.org/wiews/>