Genetic Resources and Crop Evolution **47:** 147–152, 2000.

An information retrieval system for VIR's herbarium, St. Petersburg, Russia, as a tool for cultivated plant research

Natalia N. Luneva*, Irena G. Chukhina & Elizaveta G. Lebedeva
*N.I. Vavilov All-Russian Research Institute of Plant Industry, 42, B. Morskaya St., St. Petersburg 190000, Russia (*Author for correspondence)*

Received 28 September 1998; accepted in revised form 27 April 1999

Key words: cultivated plants, databases, herbarium, taxonomy

Abstract

The herbarium collection of the N.I. Vavilov Institute of Plant Industry (WIR) contains specimens of cultivated plants, their wild relatives and weedy plants and provides ample opportunities for research. It is a unique special herbarium collection and one of those rare collections having a card index (catalog). For the catalog to be functional, it should reflect the current changes in the taxa names. Only a computerized card file may meet this requirement. The creation of a database for VIR's herbarium was launched in 1995. VIR is carrying out a range of work necessary for providing a wider access to its herbarium and related information for the international scientific community.

Abbreviations: VIR – Russian abbreviation of N.I. Vavilov Institute of Plant Industry; WIR – index of Herbarium of VIR; WRCP – wild relatives of cultivated plants; RFFR – Russian Foundation for Fundamental Research.

Introduction

The herbarium of the N.I. Vavilov All-Russian Research Institute of Plant Industry (WIR) contains over 250,000 specimens of cultivated plants, their wild relatives and weedy plants. It is a unique herbarium collection for all CIS countries and Europe, which enjoys the status of specialized herbarium of global importance. In the VIR herbarium of the cultivated flora, collections of the majority of cultivated plants and their wild relatives have been assembled (Agaev et al., 1994).

With the exception of small-sized herbarium collections of cultivated plants preserved at some botanical gardens and universities (USA: National Arboretum, Arnold Arboretum, Washington University, etc.), only three large herbaria which specialize in cultivated plants exist in the world, namely the herbarium at VIR in St. Petersburg (Russia), Bailey Hortorium in Ithaca (USA) and the herbarium of cultivated plants at IPK Gatersleben (Germany). The outstanding researcher Academician N.I. Vavilov was the initiator of collecting the VIR herbarium. According to his idea, the herbarium was meant to depict all the existing variability within each crop, taking into account taxonomy and geography. Collections of cultivated plants and their wild relatives as well as material brought from exploration expeditions were used as sources for establishing the herbarium. Great contribution to this activity was made by N.I. Vavilov himself and his co-workers: E.N. Sinskaya, P.M. Zhukovsky, V.V. Markovich, S.M. Bukasov, I.D. Bukinich, E.G. Chernyakovskaya, N.V. Kovalev, M.G. Popov and others who contributed hundreds of herbarium accessions from their expeditions.

Structure of herbarium

At present the collection is divided into the following sections:

(1) The main herbarium of cultivated plants and their wild relatives. The collection numbers about 250,000 sheets. Over 100 families, about 500 genera and over 4600 species are present in the collection. The greatest part of the main herbarium material

comprises such families as Fabaceae, Poaceae and Rosaceae, which amount to about two-thirds of the collection.

(2) The type herbarium represented by a collection of nomenclatural types, the majority of which belong to taxa of intraspecific rank. Of 182 type specimens, only 17 represent species, the remainder subspecies and varieties. They embrace 10 families and 20 genera.

(3) The general herbarium, or herbarium of plants exotic for Russia, comprising samples sent from foreign herbaria and which do not match the main herbarium theme (about 20,000 sheets). This is an auxiliary herbarium used for wide-scale research on supraspecific taxa, for preparing taxonomic monographs of genera in which relations between taxa should be taken into account, for providing a wider range of services to the users, lending included, for preparing volumes of the 'Flora', as well as for specific research programmes on agricultural, nectariferous, or ornamental plants, etc.

(4) The herbarium of weeds, collected during the last 86 years and which at present is the only herbarium in Russia, CIS countries and Europe that reflects the whole range of plant species occurring as weeds in crops in the CIS countries. It comprises 60,000 specimens belonging to 93 families, 845 genera and 3507 species.

(5) Exchange herbarium (delectus) of exchange materials set up to replenish VIR's Herbarium through exchange with other botanical institutions. This herbarium was compiled by means of collecting from 40 to 60 duplicate specimens of the same taxon from the same collecting site. The herbarium numbers about 14,000 sheets. The contents of the exchange stock is published in the 'Delectus plantarum', from which orders can be made.

VIR's herbarium as a basis for research

The herbarium of VIR accumulates data on the composition of cultivated floras and distribution of species and forms, and may provide exhaustive and reliable information on changes within cultivated floras of different regions at various periods.

Most noteworthy are the collections of numerous old varietal populations which were cultivated in the past and are lost at present. Of great significance are the herbarium specimens which represent diversity of forms from the centres of origin of cultivated plants, established by N.I. Vavilov. The special importance of these materials is also due to the ongoing extinction of local varieties due to their replacement with new bred varieties.

In various types of plant research, especially those dealing with plant introduction and breeding, a herbarium specimen is a document showing exactly the plants which had been investigated. The herbarium of cultivated plant specimens makes it possible to judge the identity of a reproduction with the original. After years of reproduction, a cultivated plant may undergo changes which may be traced in herbarium specimens collected at different periods of its crop development. The history of a crop may be traced back thanks to the herbarium collection which stays unchanged, since seed collection may get contaminated mechanically or biologically and thus be lost. Moreover, an original herbarium specimen of a variety/cultivar is the document required for the approbation procedure.

If the VIR herbarium contained only cultivar/varietal specimens, the range of tasks which may be solved with its help would have been significantly limited. However, VIR's herbarium pursued a wider objective of including wild relatives of cultivated plants as well. A thorough study of a cultivated plant is impossible without tracing its origin, evolution, potential variation range; that is, without a simultaneous study of all wild relatives of this plant. The herbarium of numerous wild relatives of cultivated plants from many collecting sites within their distribution aids in developing intraspecific classifications and in judging the usefulness of these relatives in breeding programmes. From this point of view, VIR's herbarium provides ample opportunities for research.

For example, *Prunus* L. is represented by over 8000 herbarium sheets. Of these, about 7000 are specimens of wild *Prunus cerasifera* Ehrh. sampled throughout its distribution area. Since this species, together with *Prunus spinosa* L. is hypothesized to be an ancestor of *Prunus domestica* L., it is very important to study its intraspecific variability. Research on leaf, flower and stone characters of *P. cerasifera* L. carried out on herbarium materials at VIR, applying statistical and statusmetrical methods[1], formed the basis for an infraspecific taxonomy of the species and an un-

[1] The system of the complex scientific analysis of collections of genetic resources has been created in the N.I. Vavilov Institute. The main tool of such analysis are information mathematical models, which connect complex quantitative characters of efficiency, stability and quality of cultivated plants and their wild relatives with biological parameters of plants and agroclimatic conditions of an environment. We used these models for the analysis of biodiversity of plants in general and intraspecific variability, particularly.

derstanding of the emergence of the varietal types of the cultivated myrobalan plum (Luneva, 1985, 1995; Luneva & Razorjenova, 1997). Many other investigators use this collection for research in taxonomy of cultivated plants and their wild relatives and floristic (Chukhina, 1994; Chukhina et al., 1998; Potokina, 1997; Seferova, 1995, 1997; Smekalova, 1995, 1997).

Creation of the information retrieval system

The Herbarium at VIR is one of those rare collections with a card file. Setting up a card file, if not initiated together with the collection, may pose certain difficulties later and thus was not undertaken by most herbaria. The development of VIR's herbarium card file is determined by the specificity of the collection of taxa represented by large numbers of specimens.

For example, the herbarium of the myrobalan plum (*Prunus cerasifera* Ehrh.) numbers over 7000 specimens sampled from various locations within the large area of distribution of this species, which embraces North Caucasus, Transcaucasus and Central Asia. Several dozens of boxes occupying several cabinets are required to store this herbarium. The herbarium of blackthorn (*Prunus spinosa* L.) numbers over 500 sheets, that of apple (*Malus domestica* Borkh.) over 1000 sheets, of pear (*Pyrus communis* L.) over 3600 sheets, etc. It is obvious that the search for a certain specimen from a particular region requires a tool to facilitate this work. The card file was designed to meet this requirement, but fails to accelerate the search because of its bulkiness.

More than 25 years ago, computer applications to herbarium studies were initiated in international practice, and advantages and drawbacks identified (Crovello, 1967). The boom of computerizing herbarium collections soon slowed sharply when significant problems, mostly associated with the great volume of data to be manually computerized, were encountered. Dr. Shetler from the National Herbarium of the USA, one of the enthusiasts of computer application for herbarium needs, pointed out that in fact the computer was increasing the volume of manual work. The data from labels have to be edited and made uniform. Computerization of information requires precision and attention (Shetler, 1974). Besides, the major information contained by a herbarium is not in the labels but in the herbarium specimens themselves. Proceeding from this, many herbarium curators believe creating of a herbarium catalogue to be scientifically non-profitable (Skvortsov, 1977). Nevertheless, computerization of herbarium collections has been initiated by many botanical institutions. Databases are being created not only for collections of nomenclature types, but for different sections of herbarium collections as well (Dzhus & Tichomirov, 1997; Jezniakowsky, 1997; Konstantinova & Savchenko, 1997; Krasinsky, 1997; Nikitin at al., 1997).

Another important yet problematic aspect in creating a herbarium catalogue must be resolved. It concerns the taxonomic placement of a herbarium specimen which, on the one hand, depends on the qualification of the botanist who carried out the initial determination, and on the other hand, on the opinion of other botanists who share different views on the taxon size and its position in the taxonomic system. As a result, a series of different names may be attributed to the same specimen, and sometimes even the generic name of a specimen is changed. It is impossible to keep renaming herbarium specimens, following the most recent taxonomic developments, even more so to regroup them to fit new taxonomic systems.

Nevertheless, to be functional, the herbarium's catalog should reflect the current changes in the taxa names. So we started a database implementation, the database embodied the primary names of permanently stored specimens (Luneva & Chukhina, 1995). Using the database, we implemented the information retrieval system (IRS), which employs additional taxonomic and nomenclature information. This IRS is a versatile guide to the herbarium collection (Luneva et al., 1997).

Materials and methods

To eliminate labeling imperfections, we used a principle of selection and discrimination of the information. No data input to IRS are a mere copy of a label, but include only most important information of a herbarium sheet and additional information on the taxa. A number of approaches and procedures were developed to facilitate IRS productivity and to exclude occasional errors while entering data. Among them are: a program to speed up recurrent data input; usage of ready reference data; impose referential database integrity on genera and species lists.

Such an IRS database design provides for speeding up operations of information search and retrieval, and simplifies user operations required for entering and editing herbarium data. The 'WIR Herbarium'

information retrieval system was implemented with FoxPro 2.6 RDBMS (MS-DOS version). We choose this software instrument to maintain data compatibility with passport and assessment databases of specimens of WIR cultivated plant collections, for the taxonomic diversity of these specimens is reflected in the herbarium collection. The FoxPro application developed has an elaborated user interface and a help system, but is very compact and requires only 395 Kb of hard disk space.

The set of queries built in 'WIR Herbarium' IRS is identical to sets of other herbarium management software systems: one can retrieve taxonomic, historical, personal and other relevant data.

Description of the database and information retrieval system

The IRS enables one either to browse the specimen list or to query the list and to filter the retrieved data according to specified criteria. When in the browse mode, one can obtain information on all the families represented in the WIR collection. The family records include information on the specimens of which genera are in the collection (the number of sheets included), where the specimens were collected, on which shelves and in which boxes they are stored. Within a genus, one can retrieve similar information on smaller taxonomic units down to a cultivar. As far as the information on the specific taxa has been input into the IRS, data on all relevant specimens can be retrieved in the browse mode. The first information screen displays the herbarium sheet number, the nomenclature part of the record (with all re-identifications, if present), collection date and region, collector's name, specimen status (cultivated, wild, weed), expedition number (in case the specimen was collected during an expedition). The second screen combines characteristics of the place of collection: the geographical point (current and old names), altitude above sea level, latitude and longitude, ecological and cenotic characterization.

In search mode, one can retrieve information on the geographic habitat of the specimen either starting from the name of the taxon ('taxon' search) or starting from the information on the specimen collection place ('collection place' search). 'Taxon' search yields two kinds of data:

- which species of the specific genus from which countries are in the collection;
- which species from the specific country of the specific region are in the collection.

In both cases, the place of the specimen storage is retrieved.

Within a country, one can search by a territory (state, province) and retrieve specimen lists with the specimens' shelf indices along with all specimen data. A 'geography' search allows one to get answers to the following questions:

- which specimens of which families from the queried region (country, territory, etc.) are in the collection;
- which specimens from the queried region represent a specific family (genus etc. down to a cultivar) in the collection and to retrieve the corresponding quantity of the herbarium sheets and the place of their storage.

One can also obtain estimates of the sizes of herbaria, collected by a particular collector or processed by a particular monographer (either counting taxa of all ranks or counting the quantity of the herbarium sheets).

General information (year, time span, region, members, aim, etc.) on the expeditions with the participation of WIR researchers is also available.

The IRS has a special feature called 'WIR Herbarium of Types'. This feature allows browsing the list of families and genera, represented in the WIR Herbarium of Types. One can query for:

- the type specimens identified by a particular author;
- the type specimens belonging to a particular family (of a particular continent);
- herbarium sheets of a particular type class.

In case the search has been performed down to a particular herbarium sheet, the search results are output as two display screens. The first screen is the herbarium label citation, the second one refers to the primary description and literature reference.

The 'WIR Herbarium' IRS is used both for herbarium management and for research and it appeared to be helpful for the everyday chores of the herbarium specialists. It has been built as an open software system with a modular structure, the design facilitating ease of system maintenance.

Discussion

Upon completing the IRS development and filling up database fields, the following research tasks may be solved:

- identification of gaps in the herbarium in terms of representation of large taxa and smaller systematic units of the cultivated plants and wild relatives of the cultivated plants (WRCP) in the collection, and in sampling from their distribution areas;
- elaboration of theoretical and methodological aspects of territoriy identification for the *in situ* preservation of WRCP;
- elaboration of strategies for plant introduction and core-collections proceeding from the results of eco-geographical analysis of the distribution of WRCP;
- systematics- and breeding-oriented morphological studies of species widely represented in the VIR herbarium;
- approbation of variety (cultivar) using the VIR herbarium database as a reference source.

Conclusions

Thanks to Internet technologies, integration of botanical research on nomenclature, taxonomy and floristics using herbarium collections is underway on an international scale. In this relation, VIR is carrying out a scope of work necessary for providing wider access to its herbarium and related information for the international scientific community. With the support from the RFFR, the institute is realizing a project entitled 'Virtual Herbarium'.

Though an image of a herbarium sheet cannot replace the real specimen, it may be very useful at many stages of taxonomic research. Computer images of herbarium sheets may be multiplied in necessary quantities, and easily sent to a user thus allowing him to gather at one location images of herbarium sheets actually dispersed among different herbaria through the world. Computer images may be very helpful for assembling sample series from different parts of the area of distribution, compiling them on the basis of such morphological characters as size, shape, number of denticles (veins) in the leaf, pattern of the leaf blade, flower, stone, etc. This information may be quite sufficient for deriving preliminary results. Viewing of virtual herbarium specimens (with magnification of necessary fragments), studying of morphological characters along with analyzing information from databases on geography and ecology allows revealing the structure of a species. All this comprehensive information accessible via the Internet will allow any researcher, a potential visitor to VIR herbarium, to solve various research problems in the areas of taxonomy and plant science.

Opportunities provided by computer and telecommunication technologies in retrieval and use of herbarium specimens information by remote users will make it possible to intensify, and bring to a new level, classical botanical investigations.

Acknowledgements

Creation of dataware for VIR herbarium was launched in 1995 with support from the RFFR.

With the support from the RFFR the institute is realizing the project entitled 'Virtual Herbarium'.

References

Agaev, M.G., T.N. Uljanova & N.N. Luneva, 1994. A unique herbarium. Bull. Ross. Akad. Selskochos. nauk 3: 22–25 (In Russian).

Chukhina, I.G., 1994. Ecogeographical analysis of useful plants in the North-Eastern Altai. Nauchnyi Bull. VIR. 233: 101–104 (In Russian).

Chukhina, I.G., N.N. Luneva & E.G. Lebedeva, 1998. Collection of the type specimens in the Herbarium of N.I.Vavilov Institute of Plant Industry (WIR). Bot. Zhurn. 3: 74–79 (In Russian).

Crovello, T.J., 1967. Problems in the use of electronic data processing in biological collections. Taxon 16: 481–494.

Dzhus, M.A. & V.N. Tichomirov, 1997. Computerised database of Belarusian State University herbarium. In: Komputernye basy dannych v botanicheskich issledovaniyach. Botanical Institute Press, St. Petersburg, p. 20 (In Russian).

Jezniakowsky, S.A., 1997. Orchids database of Novgorod Province. In: Komputernye basy dannych v botanicheskich issledovaniyach. Botanical Institute Press, St. Petersburg. p. 24 (In Russian).

Konstantinova, N.A. & A.N. Savchenko, 1997. The database 'Bryophyta herbarium of Murmansk province'. In: Komputernye basy dannych v botanicheskich issledovaniyach., Botanical Institute Press, St. Petersburg. p. 38 (In Russian).

Krasinsky, V.I., 1997. Reliable computer input and checking of herbarium labels. In: Komputernye basy dannych v botanicheskich issledovaniyach. Botanical Institute Press, St. Petersburg. p. 43 (In Russian).

Luneva, N.N., 1985. On the taxonomic status of *Prunus caspica* myrobalan plum (Rosaceae) from Eastern Transcaucasus. Bot. Zhurn. USSR. 70, 11: 1487–1491 (In Russian).

Luneva, N.N., 1995. Variations of petal morphological characters and diversity of forms in the genus *Prunus* L. (Rosaceae). Bot. Zhurn. 80, 8: 79–84 (In Russian).

Luneva, N.N. & I.G. Chukhina, 1995. Implementation of a database and an information retrieval system of the herbarium of cultivated plants and their wild relatives (WIR). In: Komputernye basy dannych v botanicheskich issledovaniyach. St. Petersburg. pp. 26–27 (In Russian).

Luneva, N.N., I.G. Chukhina & E.G. Lebedeva, 1996. The database and information retrieval system 'Herbarium of WIR'. In: Komputernye basy dannych v botanicheskich issledovaniyach. Botanical Institute Press, St. Petersburg. p. 31 (In Russian).

Luneva, N.N. & T.S. Razorjenova, 1997. Evaluation databases as the foundation for the determination of intraspecific differences in cherryplum using the methods of functional modelling. In: ADBIS 97, Proceedings of the first East-European symposium on advances in databases and Information Systems. St. Petersburg, September 2–5, 1997. 2: 123–124.

Nikitin, V.V., A.E. Borodina-Grabovskaja & M.S. Novoselova, 1997. Creation of the computer catalogue of type specimens of the Komarov Botanical Institute herbarium (LE). In: Komputernye basy dannych v botanicheskich issledovaniyach. Botanical Institute Press, St. Petersburg. p. 75 (In Russian).

Potokina, E.K., 1997. *Vicia sativa* L. aggregate (Fabaceae) in the flora of former USSR. Genet. Resour. Crop Evol. 44: 199–209.

Seferova, I.V., 1995. The synopsis of the system of genus *Cicer* (*Fabaceae*). Bot. Zhurn. 80, 8: 96–104 (In Russian).

Seferova, I.V., 1997. Varieties of chick-pea – *Cicer arietinum* L. Trudy po prikladnoj botanike, genetike i selekcii. 152: 9–18 (In Russian).

Shetler, S.G., 1974. Demythologizing biological data banking. Taxon 23: 71–100.

Skvortzov, A.K., 1977. The herbarium. 200 p. (In Russian).

Smekalova, T.N., 1995. On the question of factors of differentiation of intraspecific taxa in *Lathyrus sativus* L. In: Faktory taxonomicheskogo i biochorologicheskogo rasnoobrasija. St. Petersburg, p.75 (In Russian).

Smekalova, T.N., 1997. Major annual species from gen. *Lathyrus* subgen. *Cicercula* (systematics, geography, ecology, variability). Autoref. of Ph.D.Thesis. St. Petersburg, 21 p. (In Russian).