

Systems for making NIAS Core Collections, single-seed-derived germplasm, and plant photo images available to the research community

Masaru Takeya · Fukuhiro Yamasaki ·
Sachiko Hattori · Akito Kaga · Norihiko Tomooka

Published online: 3 August 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract The National Institute of Agrobiological Sciences (NIAS) Genebank coordinates the conservation of plant, microorganism, and animal genetic resources related to food and agriculture in Japan. It also coordinates the distribution of genetic resources in the public domain for research, breeding, and educational purposes. To operate the NIAS Genebank efficiently, we have developed a genetic resources database, data management software, and web-based data retrieval systems to make the data available worldwide. This article describes the NIAS Genebank's Core Collections of global and Japanese soybean (*Glycine max* (L.) Merr.), Japanese azuki bean (*Vigna angularis* (Willd.) Ohwi et Ohashi), and Japanese wheat (*Triticum aestivum* L. s. l.), all of which are available through the Genebank website. This article also describes new features of the NIAS Genebank database, such as the ability to select single-seed-derived germplasm of soybean in the plant search system and to download photographic data on accessions. By using the downloaded plant image PDF files,

users can obtain detailed passport and agronomic information by clicking on the image of an accession of interest.

Keywords Core collection · Database · Genebank · Plant image · Single-seed-derived germplasm · Web-based retrieval system

Introduction

The National Institute of Agrobiological Sciences (NIAS) Genebank Project conserves and promotes the use of plant, microorganism, and animal genetic resources related to food and agriculture (Okuno et al. 2005; Takeya et al. 2012). The NIAS Genebank also distributes accessions in the public domain for research, breeding, and educational purposes. To operate the NIAS Genebank efficiently, it is important to manage data such as passport data, characterization data, evaluation data, and storage data. For that purpose, we have developed several databases, data management software, and web-based data retrieval systems.

It is important for gene banks to prepare and provide genetic materials to meet the needs of the research community. One such research tool is a core collection, a limited set of accessions representing, with a minimum of repetition, the genetic diversity of a crop species and its wild relatives (Frankel 1984). We have modified this concept and developed the

Electronic supplementary material The online version of this article (doi:10.1007/s10722-013-0005-1) contains supplementary material, which is available to authorized users.

M. Takeya (✉) · F. Yamasaki · S. Hattori ·
A. Kaga · N. Tomooka
Genetic Resources Center, National Institute
of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602,
Japan
e-mail: katu@affrc.go.jp

NIAS Core Collections which we designed to be of a size that can be analysed in a 96-well microplate. The NIAS Core Collections of global and Japanese cultivated rice (*Oryza sativa* L.) and Japanese maize (*Zea mays* L.) landraces have already been introduced in a previous publication (Takeya et al. 2011; http://www.gene.affrc.go.jp/databases-core_collections_en.php). In this paper, we introduce the recently developed NIAS Core Collections of Japanese and world soybean, Japanese azuki bean, and Japanese wheat (Table S1 added as Electronic Supplementary Material). The NIAS Soybean Core Collections were developed in accordance with the original core collection concept but are much smaller than a typical core collection. The NIAS Japanese Azuki Bean Core Collection was developed from original collections from our own explorations; these accessions therefore have detailed information on collection site, agricultural information from farmers, and ecological information on the natural habitat of wild species. The NIAS Japanese Wheat Core Collection consists of a set of Japanese landraces from a wide geographic region and another set of popular Japanese released cultivars.

Another type of material useful for genetic research is single-seed-derived germplasm. Here, we describe the development and availability single-seed-derived germplasm of soybean genotyped by using single-nucleotide polymorphism (SNP) markers. We also describe a system for bulk download of seed photographs from which users can access passport and evaluation data.

NIAS World and Japanese Soybean Core Collections

Approximately 11,300 soybean accessions are conserved at the NIAS Genebank. By examining the passport information and agronomic evaluation data compiled in the NIAS Genebank database, we selected 1,603 soybean accessions, consisting of 832 Japanese landraces, 109 old and 57 recent Japanese cultivars, 341 landraces from 16 Asian foreign countries, and 264 wild soybean accessions, and characterized them by using 191 SNP markers (Kaga et al. 2012). By assessing SNP marker genotypes using the programs PowerMarker (Liu and Muse 2005), Structure (Pritchard et al. 2000), and PowerCore (Kim et al. 2007) and assessing several agro-morphologic traits, we

have developed two NIAS Soybean Core Collections, each consisting of 96 accessions from Japanese and world germplasm. These accessions were selected to retain 100 % of the gene diversity of the complete set of NIAS Genebank accessions; this selection was based primarily on SNP variation, but it also considered morpho-agronomic trait variation, population structure, and geographic origin. Seeds in these core collections have been multiplied from SNP-genotyped single plants.

Single-seed-derived germplasm of soybean

Single plants from seeds each of the selected 1,603 accessions described above were grown in 2009. Among them, 1,250 cultivated accessions are registered as separate accessions of single-seed-derived germplasm in the NIAS Genebank. We are planning to advance the material by the single-seed-descent method and to accumulate accurate phenotype data for these accessions. We have developed the search option ‘Single-seed-derived germplasm only’ to enable users to limit their search results to these materials.

NIAS Japanese Azuki Bean Core Collection

From more than 2,000 azuki bean accessions conserved at the NIAS Genebank, 616 accessions originating from eight Asian countries were selected on the basis of passport information, and their genetic diversity was analysed by using 13 simple sequence repeat (SSR) primers (Xu et al. 2008). The results showed that cultivated azuki beans from East Asia (China, Korea, and Japan) harboured the highest genetic diversity among cultivated accessions and that accessions of these three East Asian countries were genetically distinct from one another. This suggested a long and relatively isolated history of cultivation in each East Asian country. In addition, the wild azuki bean germplasm from Japan showed higher genetic diversity than cultivated accessions and represented much of the allelic variation found in the cultivated germplasm. The SSR results, together with recent archaeobotanical evidence, support the view that Japan is the centre of domestication of azuki bean (Crawford 2005). Therefore, we used PowerMarker

(Liu and Muse 2005) to develop the NIAS Japanese Azuki Bean Core Collection, which consists of 80 cultivated and 38 wild azuki bean accessions.

Because selected accessions of the NIAS Japanese Azuki Bean Core Collection were directly collected as part of the NIAS Genebank domestic exploration project, detailed collection site information, including latitude and longitude, is available. By clicking the ‘Collection site’ button on the accessions list table, a KML-formatted file can be downloaded and the collection site of each accession can be plotted by using the Google Earth program. Cultivated and wild accessions are shown in different colours on the Google Earth map. Each accession is linked to detailed passport and evaluation information. Users can obtain seed photos by clicking on the ‘Image’ button. A larger seed photo image can be obtained by clicking on the ‘Seed’ link within the ‘Image’ column. Agronomic data for the accessions, such as plant height, seed weight, and days to flowering, are also available as an Excel file, which can be downloaded from a link at the bottom of the page.

NIAS Japanese Wheat Core Collection

The NIAS Japanese Wheat Core Collection has been developed on the basis of a concept slightly different from the original core collection definition. It consists of 45 traditional Japanese landraces, 51 cultivars released in Japan, and ‘Chinese Spring’, which is included as a standard reference cultivar. The Japanese landraces represent materials collected from throughout Japan. The released cultivars represent popular cultivars covering the breeding history of wheat in Japan. We are now accumulating genotype information for several genes useful in agriculture.

Improvement of genetic resources database to manage core collections

A unique identifier for accessions—the ‘JP number’—is assigned to each of the plant genetic resources in the genetic resources database. The database table ‘Plant Genetic Resources’ contains the JP number as the primary key and includes passport data such as registration date and plant code of scientific name. The ‘Conservation ID’ is a unique accession ID

identifying each accession and associated conservation site. When an accession is conserved at two different sites, two Conservation IDs are assigned to that accession. The table ‘Conservation Accessions’, which has Conservation ID as the primary key, is linked to the table ‘Plant Genetic Resources’ via the JP number. Figure 1 shows the relationship between the main tables containing passport and characteristics/evaluation data.

We have developed several new tables to manage the core collections or special research sets of germplasm. The table ‘Collection’ is designed to register the type of collection and includes not only the plant section but also the microorganism and animal sections in the NIAS Genebank. The type of collection (e.g. ‘core collection’) and section (e.g., ‘plant’) are entered into the columns ‘Collection category’ and ‘Section code’, respectively. The table ‘Core Collection’ has been developed to register specialized information pertaining to each core collection, such as core collection name, development organisation, developer, explanatory notes, and references. The associated accessions of core collection are registered into the table ‘Core Collection Accessions’, for which the candidate key is a combination of Conservation ID and Core collection ID.

Database for managing plant photo images and evaluation dataset

Image data provides the most direct means for users to obtain information on the appearance of plant genetic resources. NIAS Genebank manages image data and characteristics/evaluation data similarly. The evaluation data contain categories such as numeric, data states, date, and class. For example, the evaluation items ‘grain length of rice’ and ‘seed coat colour of cowpea’ (*Vigna unguiculata* (L.) Walpers cv-gr. Unguiculata E. Westphal) belong to the numeric and class data categories, respectively. The colour item has nine classes, such as 1 = white, 2 = yellowish white and 3 = yellow. An evaluation research manual, including item name, method of measurement, measurement unit, and remarks is maintained for each evaluation group (e.g., ‘rice’) including close species. The number of evaluation groups is 125, the largest of which has 121 items. To manage a large number of items efficiently, a characteristics/evaluation dataset

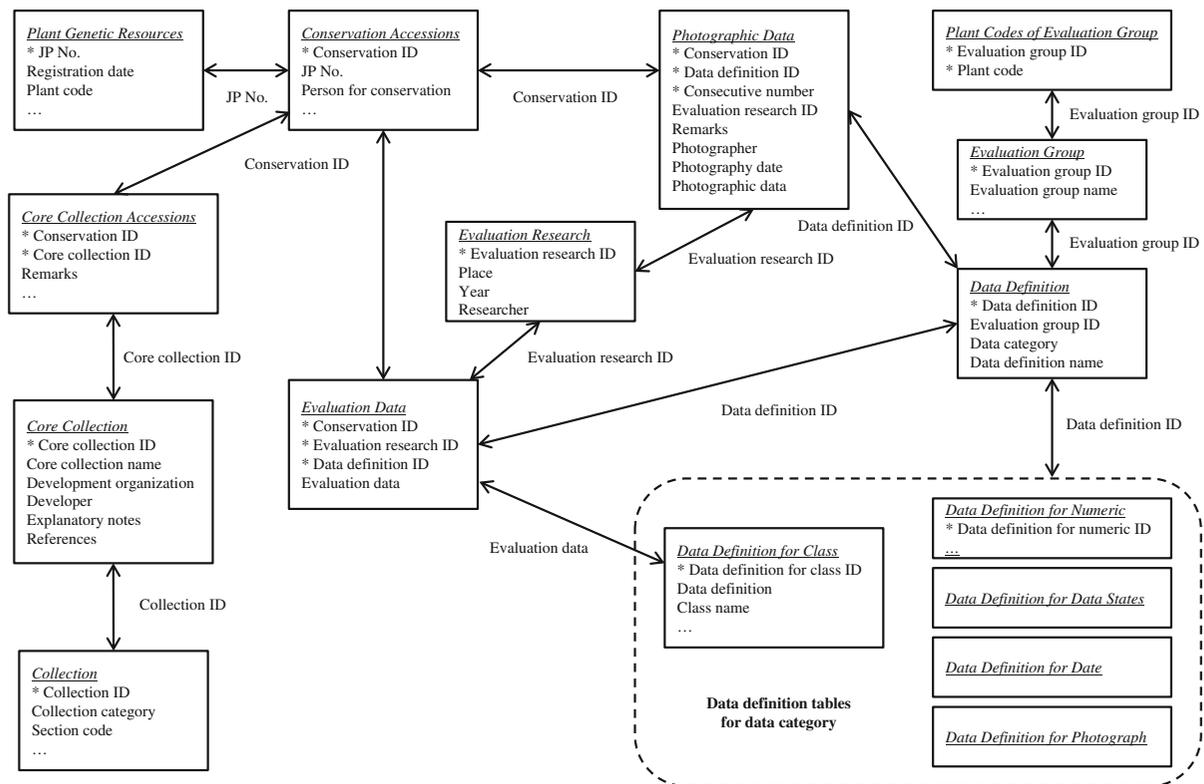


Fig. 1 Schematic diagram of the relationships between the database tables of the core collection and image data. Table names are in *italics*; *asterisks* indicate primary or candidate keys

has been developed to register the manual as metadata into the database. Each evaluation item in an evaluation group is registered into the table ‘Data Definition’ and given the unique identifier ‘Data definition ID’. The specialized data definition for each data category is registered into specialized ‘Data Definition’ tables. For example, the table ‘Data Definition for Numeric’ consists of the number of digits, number of decimal places, maximum, minimum, and units. The table ‘Data Definition for Photograph’ contains information on the plant part in each photo and the assigned photo number for laboratories. The table ‘Evaluation Data’ has been developed to register evaluation data for any category except image data. The table contains a candidate key consisting of three columns (Conservation ID, Evaluation research ID and Data definition ID) and the column ‘Evaluation data’, where evaluated values are registered. Plant photo images are registered into the table ‘Photographic Data’. The table contains a candidate key consisting of three columns (Conservation ID, Data definition ID and Consecutive number) along with

other data such as remarks, photographer and photography date. The photo image data are registered into the column ‘Photographic data’ as a Binary Large Object (BLOB).

Improvement of system for providing image data

Before this work, NIAS Genebank had already developed an illustrated plant genetic resources database to provide images and characteristics of rice, legumes, vegetables, flowers and ornamental plants, millet and forage crops (http://www.gene.affrc.go.jp/databases-plant_images_en.php). We have now developed a function to efficiently display a diverse set of plant images each time users access the web-based illustrated database. There is indirect linkage between image data and evaluation data in the genetic resources database (Fig. 1). The new system uses morphological evaluation data to select photo images for display. The registered image data are divided into classes of data definition (e.g., particular colours). Within each class

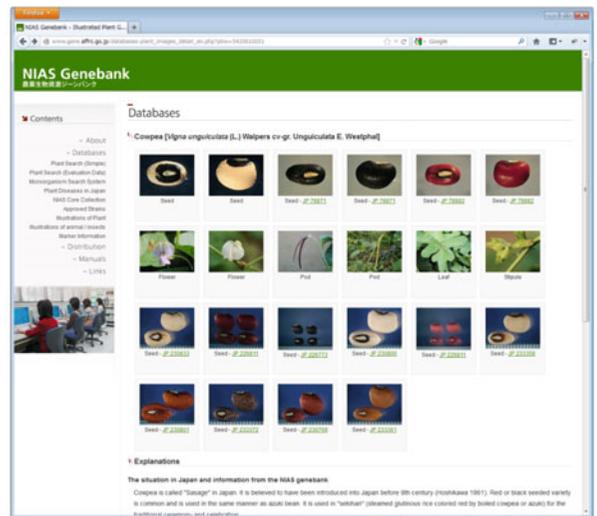
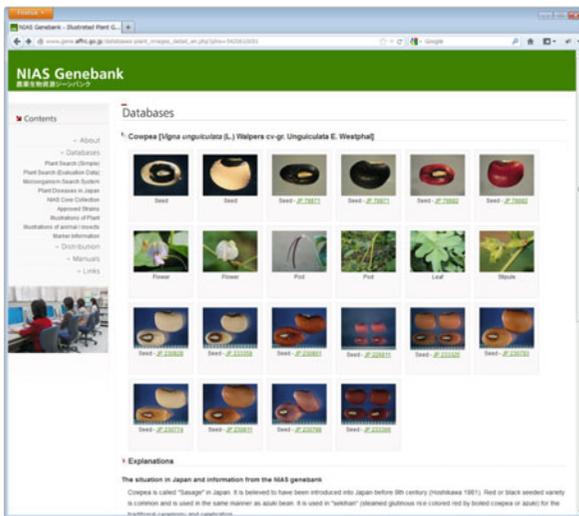
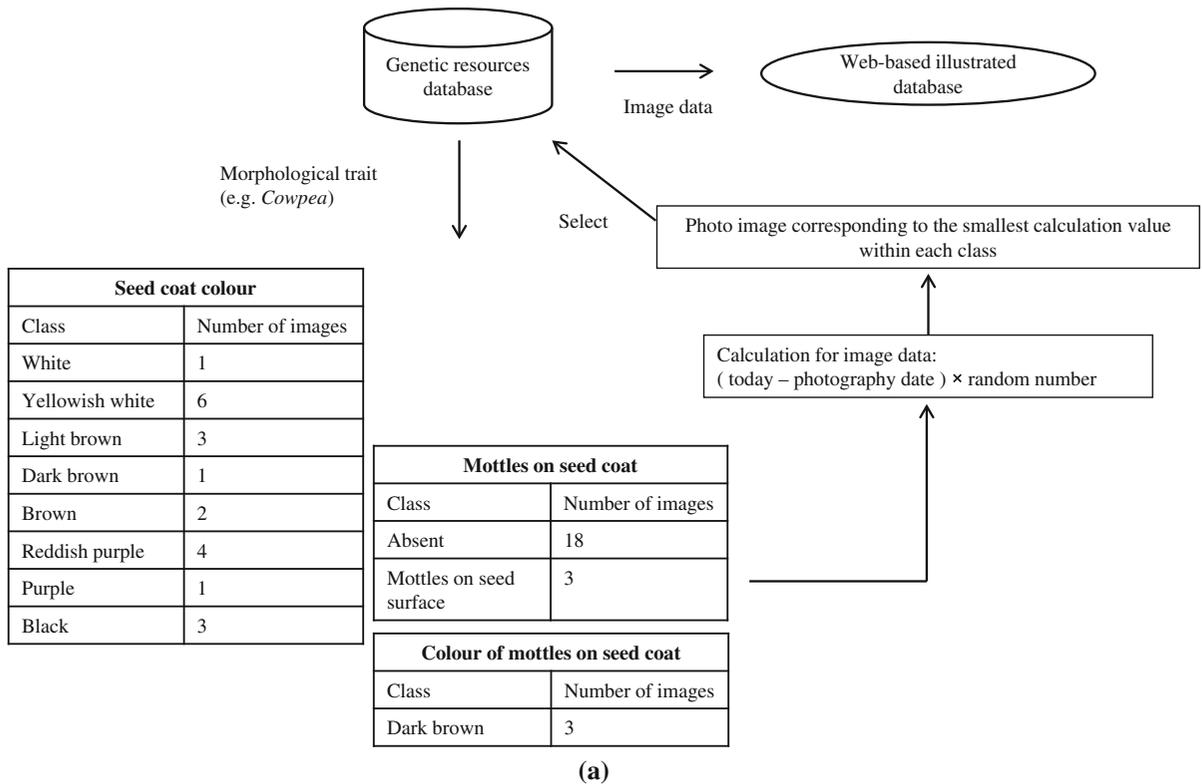


Fig. 2 Use of evaluation data to provide greater diversity in images displayed from plant genetic resources database. **a** Schematic diagram of the process for selecting image data,

using cowpea as an example. **b** Images displayed without taking evaluation data into consideration. **c** Evaluation data taken into consideration for last 10 images

(e.g., ‘yellow seed colour’), the photo image with the smallest value (calculated as the difference between today’s date and the photography date, multiplied by a

random number) is selected for display on the web page (Fig. 2a). The value for this calculation tends to be smallest for the most recently photographed images,

but the use of random numbers ensures that the identical photos are not displayed each time. Using cowpea as an example, when evaluation data are not taken into consideration, many light-coloured images (e.g., white or light brown) appear because of the large number of registered images in these colours (Fig. 2b). On the other hand, a greater variety of photo images including dark seed coat colour and mottles on the seed surface are displayed by the new system (Fig. 2c).

System to download PDF files of plant images

To increase user convenience, we have developed a function to construct PDF files of plant photos of selected accessions obtained as search results. The PDF file can be downloaded by selecting an icon at the top of the search page. Detailed information on each accession can be displayed by clicking the corresponding thumbnail in the PDF file.

Discussion

The NIAS Core Collections have been developed on the basis of a modification of the original core collection concept (Frankel 1984) and consist of about 100 accessions each, regardless of the size of the whole collections they represent. These collections are suitable for obtaining rough information on the diversity of morphological, physiological, agronomic and DNA levels in a species but are not suitable for conserving the genetic diversity of whole collections. In addition, the NIAS Core Collections may be too small to screen for resistance to biotic and abiotic stresses. It will be necessary to develop core collections based on the original concept for species represented by more than 10,000 accessions.

Although selected only on the basis of geographical and agronomic data, the single-seed-derived cultivated soybean germplasm represents about 10 % (1,250 accessions) of the whole soybean collection and thus could be considered as a core collection on the basis of the original concept of Frankel (1984). The accessions are genotyped with 191 SNP markers and are used as a base collection for selecting the NIAS Soybean Core Collections. Each line derived by single-seed descent is registered as a new accession. This approach could result in considerable increase in

the genebank's accession size and presents the risk of losing a rare allele in a heterogeneous accession (Nelson 2011). However, an advantage of this approach is that the genebank can supply seeds with a known genotype that can be phenotyped to allow more accurate genotype–phenotype association analysis (such as Genome Wide Association Analysis (GWAS)). In addition, single-seed-derived accessions accompanied by genotype data can be checked for purity, enabling the genebank manager to minimize the risk of genetic contamination and outcrossing. We plan to obtain more genomic and phenomic information on these accessions. NIAS Core Collections and single-seed-derived germplasm for other important species will be added to the genetic resources database as they are developed. We are now developing the NIAS Eggplant Core Collection and a set of single-seed-derived germplasm of 5,000 Asian rice landraces with corresponding SNP data.

Acknowledgments We thank the National Agriculture Research Organization (NARO) Institute of Crop Science for developing the NIAS Japanese Wheat Core Collection.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Crawford GW (2005) East Asian plant domestication. In: Stark MT (ed) *Archaeology of Asia*. Blackwell Publishing, Oxford, pp 78–95
- Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber W, Llimensee K, Peacock WJ, Starlinger P (eds) *Genetic manipulation: Impact on man and society*. Cambridge University Press, Cambridge, pp 161–170
- Kaga A, Shimizu T, Watanabe S, Tsubokura Y, Katayose Y, Harada K, Vaughan DA, Tomooka N (2012) Evaluation of soybean germplasm conserved in NIAS Genebank and development of mini core collections. *Breed Sci* 61:566–592
- Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwang JG, Kim TS, Cho EG, Park YJ (2007) PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* 23:2155–2162
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129
- Nelson RL (2011) Managing self-pollinated germplasm collections to maximize utilization. *Plant Genet Resour* 9:123–133
- Okuno K, Shirata K, Niino T, Kawase M (2005) Plant genetic resources in Japan: platforms and destinations to conserve and utilize plant genetic diversity. *JARQ—Jpn Agric Res Q* 39:231–237

- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Takeya M, Yamasaki F, Uzuhashi S, Aoki T, Sawada H, Nagai T, Tomioka K, Tomooka N, Sato T, Kawase M (2011) NIASGBdb: NIAS Genebank databases for genetic resources and plant disease information. *Nucleic Acids Res* 39:D1108–D1113
- Takeya M, Yamasaki F, Uzuhashi S, Kumagai M, Sawada H, Nagai T, Tomioka K, Sato T, Aoki T, Kawase M (2012) Development of a database of plant diseases in Japan and a system for making microorganism genetic resources and their DNA sequence data available to the research community. *JARQ—Jpn Agric Res Q* 46:193–198
- Xu HX, Jing T, Tomooka N, Kaga A, Isemura T, Vaughan DA (2008) Genetic diversity of the azuki bean (*Vigna angularis* (Willd.) Ohwi & Ohashi) gene pool as assessed by SSR markers. *Genome* 51:728–738