

—Review—

Review Series: Animal Bioresource in Japan

National BioResource Project Information Center

Yukiko YAMAZAKI and Hideaki SUGAWARA

National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

Abstract: The information center is the hub and glue of the National BioResource Project (NBRP). The center provides the NBRP portal site and has also contributed to the development of databases for diverse types of bioresources. The program covers information on experimental living organisms as the core of NBRP, and on specimens of biodiversity related to the activities of the Japan node of the Global Biodiversity Information Facility (GBIF). The framework of the former and the information facility of the latter are introduced.

Key words: barcode of life, biodiversity, bioresource, cross-species, database

Introduction

The National BioResource Project (NBRP) was launched in 2002 to facilitate the collection, preservation, and distribution of bioresources, especially for the purpose of research. Currently, in the second five-year term of the NBRP, twenty-seven core resource facilities and one information center (NBRP information center) are promoting the project. The main task of the information center is the construction of databases containing genetic information, biological characteristics, and the locations of bioresources gathered at core facilities. Another task is public relations through the NBRP home page. The information group has two additional sub-programs: one is to strengthen the Japan node activities of the Global Biodiversity Information Facility (GBIF) and the other is the establishment of the Great Ape Information Network (GAIN). Since the activities of GAIN will be introduced in a separate issue of this series, here we will focus on the introduction of the main program and GBIF.

The definition of “bioresources” is not always clear. For example, one could say that all living things are bioresources of the earth. Since the GBIF is primarily concerned with specimen records and observations, this definition of bioresources is most applicable. NBRP primarily manages bioresources that can be propagated and distributed to users as experimental materials for scientific research. All animal resources of the NBRP will be introduced in detail for each core facility in this series. Therefore, an overview of NBRP information, as well as services specific to the information center, is introduced in this manuscript.

The State of Bioresources in Japan

Several government ministries of Japan have been participating in different areas of bioresource preservation and distribution. The Ministry of Health, Labour and Welfare (MHLW) focuses on bioresources for medical treatment and drug development. The Ministry of Economy, Trade and Industry (METI) is in charge of

(Received 7 January 2009 / Accepted 16 February 2009)

Address corresponding: Y. Yamazaki, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

industrial microbes. The Ministry of Agriculture, Forestry and Fisheries (MAFF) secures agricultural resources, and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) is responsible for the bioresources utilized for academic research.

All of the ministries listed below have resource centers:

- MHLW: Health Science Research Resource Bank (http://www.jhsf.or.jp/index_b.html),
- METI: National Institute of Technology and Evaluation, Biological Resource Center (<http://www.nbrc.nite.go.jp/e/index.html>),
- MAFF: National Institute of Agrobiological Sciences (<http://www.nias.affrc.go.jp/gb/gb.html>).

The funding of the first basic science of bioresources was secured in 1952 by the predecessor of MEXT. However, these bioresources were maintained separately by researchers at different universities and research institutes for a long time. Career changes of key personnel over time and the influx of new researchers resulted in resources and their records becoming scattered and lost. In 2001, MEXT established the RIKEN BioResource Center, and then in 2002, MEXT started the NBRP to create a permanent system in which one central group was responsible for the collection, maintenance, and distribution of specific resources. This plan entered its second phase in 2007 after the first phase of 5 years produced reliable results.

In addition to ministry specific resource facilities, we have a BioResource Committee (<http://www.shigen.nig.ac.jp/shigen/grc/grc.jsp>, in Japanese), which consists of advisers and representatives from each of the resource communities in Japan. This structure facilitates the dissemination of information and communication among interested parties beyond the boundaries of government ministries. In fact, these committee discussions influenced the establishment of the NBRP.

We now have successful examples of integrated databases, such as the mouse (JMSR: <http://www.shigen.nig.ac.jp/mouse/jmsr/>) and microbe (JSCC: <http://www.jscc-home.jp>) databases as a result of collaborations between different databases funded by different ministries. Thus, the bioresource projects in Japan have become more unified.

NBRP Databases

Mission of the NBRP information center

Bioresources are undoubtedly important as experimental materials. Materials of high quality and availability are essential for guaranteeing the reproducibility of experimental results.

A high priority of the NBRP information center is to support dissemination of resource information held by the core facilities. As a result of the collaboration between the information center and core facilities, 75% of the database has been constructed. Some core facilities, such as the RIKEN BioResource Center, constructed their databases independently since they were established before NBRP. By combining existing databases, NBRP was able to make all of its resource information available to the public by the end of March 2006. Further integration of individual databases was undertaken to establish a single site for all NBRP resources. The resultant database (BRW: BioResource World) housed more than 3.6 million biological collections as of December 2008. A synchronization system collating the integrated database and individual databases has also been developed to keep the integrated database up-to-date.

Another mission is to create a resource portal site where users can obtain all necessary information about bioresources throughout the world. We have collected the URLs of bioresource-related web sites outside the NBRP, both within Japan (JGR: Japan Genetic Resources, <http://www.shigen.nig.ac.jp/wgr/jgr/jgrUrlList.sjp>) and outside Japan (WGR: Worldwide Genetic Resources, <http://www.shigen.nig.ac.jp/wgr/top/>). The resulting WGR currently contains 682 URLs.

There are 3,507 journal references in which authors have used bioresource(s) distributed by the NBRP. We are continuously collecting these papers to add depth to the resource information. The information center plans to develop tools to provide more user-friendly and advanced applications to maximize the usefulness of all resources.

How to use the NBRP database(s)

Figure 1 shows the NBRP home page. There are 28 organism-specific databases that are accessible directly

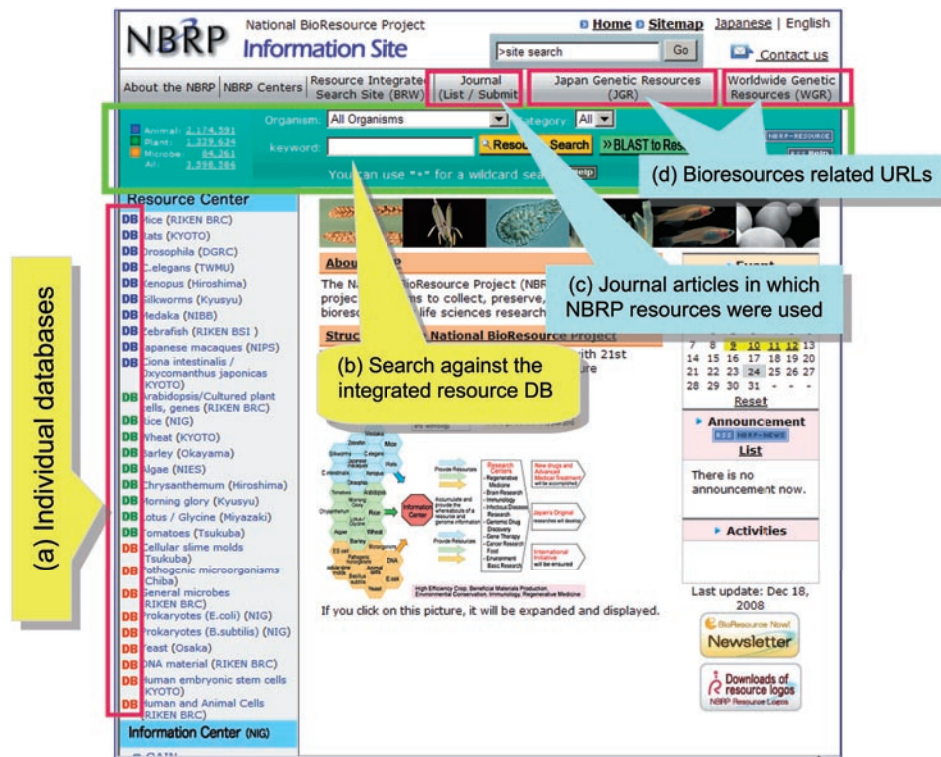


Fig. 1. The NBRP home page is a bioresources hub. Individual databases are accessible from the menu on the left (a), a search can be made of all resources via (b), journal articles in which NBRP resources were used are listed under (c), and bioresource related URLs are available under (d).

from the menu on the left (Fig. 1a). Although database usage and page layout vary depending on the collection types and requirements of each user community, services for browsing, searching, and ordering bioresources are available for all databases.

There are several characteristic gateways to bioresources. Sometimes, DNA clones and mutants are displayed on a physical map (Fig. 2) and wild type strains on a phylogenetic tree (Fig. 3). Some mutants can be accessed from a phenotype list and/or from images (Fig. 4). Mutual links between the resource database and the relevant database of biology allow users to retrieve more information.

Most resource centers have formatted a material transfer agreement (MTA), and an electronic MTA file is usually attached to e-mails during the ordering process, so that users can complete the ordering procedure on the web site as easily as they would conduct general online shopping.

Furthermore, from the single site (BRW Figs. 1b and 5a), users can search for resources from more than 3.6 million biological collections using either a query keyword, DNA sequence homology or gene ontology (Figs. 5b–d). More than 60% of all collections are animal resources; 5% are wild derived, inbred, mutant, and genetically modified strains of organisms, and 95% are DNA clones.

We encourage researchers who use NBRP resources and publish papers to feed journal information back to the RRC (Research Resource Circulation) site (Fig. 1c).

Potential of the NBRP database(s)

Since the NBRP databases were established, the number of users has increased year-on-year, reaching an average of 150,000 per month in 2008. At present, most users access an individual database directly because the idea of cross-species searches for bioresources is new

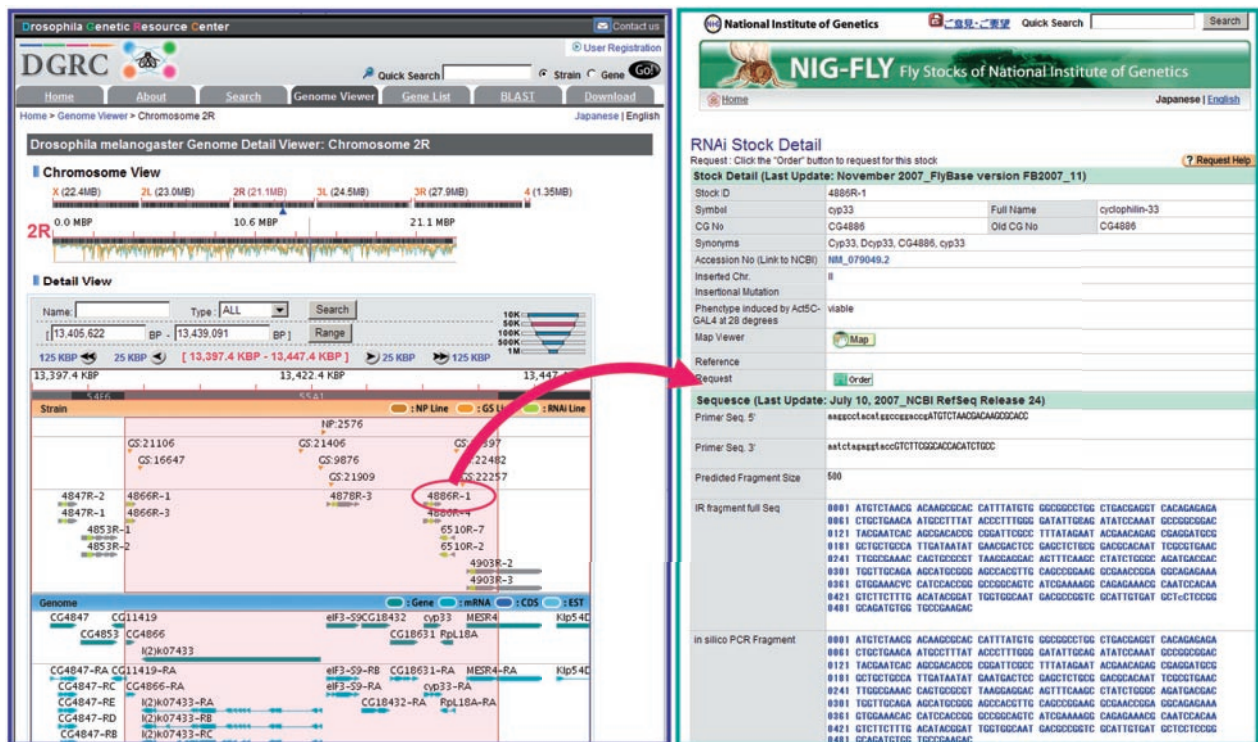


Fig. 2. Drosophila bioresources mapped on the physical map (left), and a page of a resource (RNAi stock) linked to the map (right).

and unconventional. However, if all future resources contain comprehensive and relevant information, including journal articles, we can easily imagine that cross-species searches will provide more appropriate bioresource candidates for specific research purposes. In the current BRW, users can conduct searches using keywords, sequences, and gene ontology (GO). We would like to extend the search function to characters such as reference, phenotype and trait ontology, anatomical ontology, developmental ontology, pathway, and research field.

Because bioresources are used for most biological experiments, the resultant experimental knowledge becomes resource-relevant information, meaning that there is no limitation in the research fields to which bioresources can be applied. As the number of users of the bioresource database increases, more research results will be accumulated. By extension, if the range of resource selections becomes wider, the chance to find the most suitable resource will increase.

Although this serial “Animal Bioresources in Japan” is focused primarily on animals, researchers may be able

to use a variety of bioresources, including plants or microbes, to thoroughly investigate specific animal attributes. For this exact purpose, the NBRP Information Center was invited to introduce its activities in this series. For example, assume a young researcher uses various bioresources such as drosophila mutants and/or yeast strains efficiently in his/her study of human disease.

It is also important to provide courses for explaining how new resources should be handled or for providing new users with the necessary protocols. Some NBRP centers have already initiated these services.

GBIF

What is GBIF?

GBIF is an international project that has been established to provide comprehensive and accurate biodiversity data for fundamental biology research, conservation of biological resources, environmental protection, and human welfare, primarily using specimen data from biological collections and observation data. GBIF also

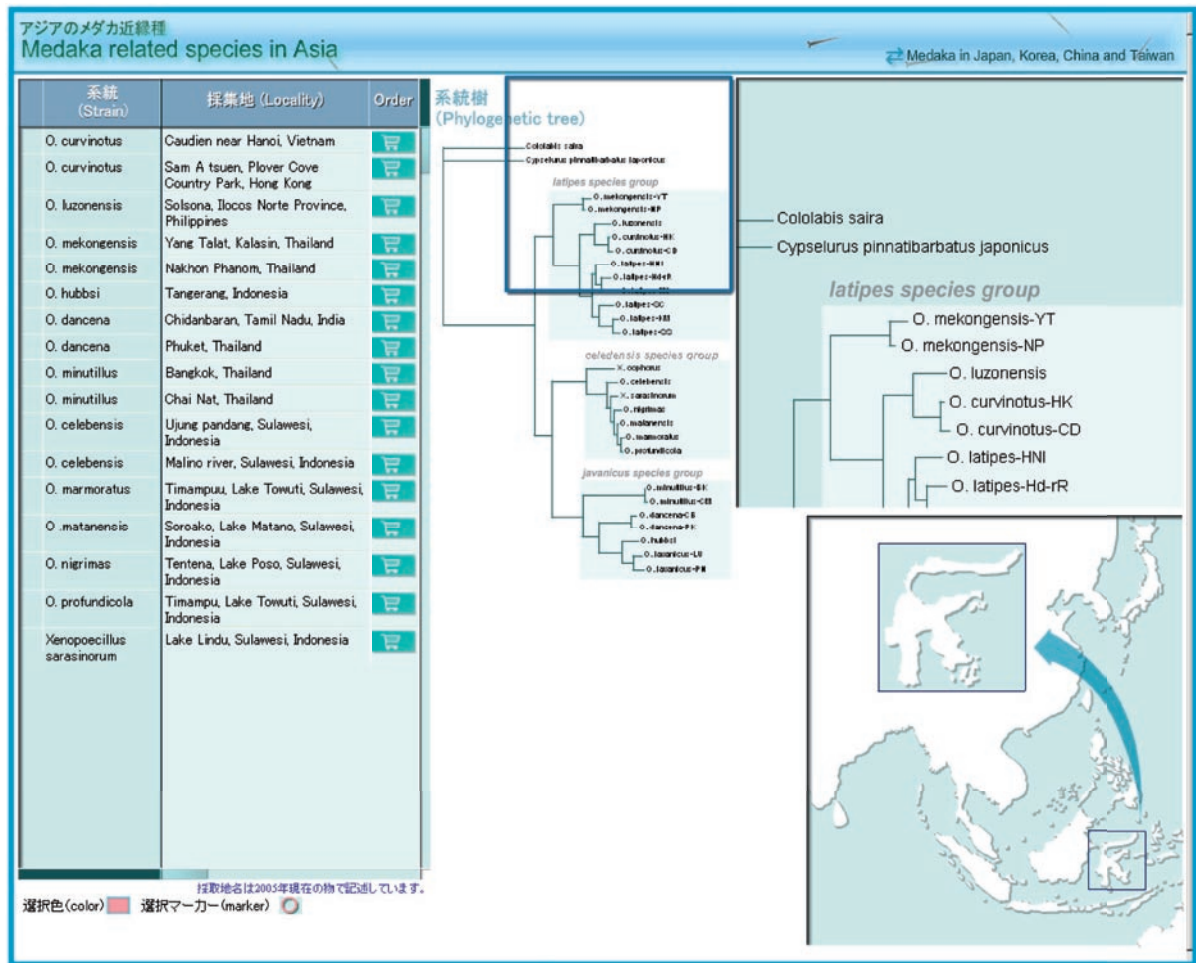
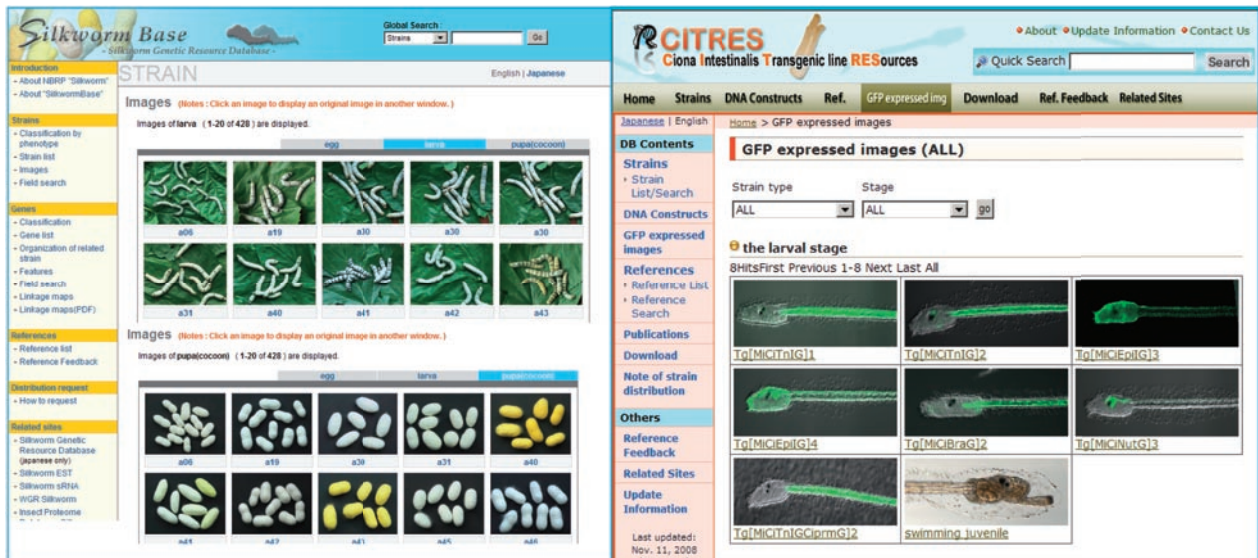


Fig. 3. Medaka wild strains mapped on the phylogenetic tree.



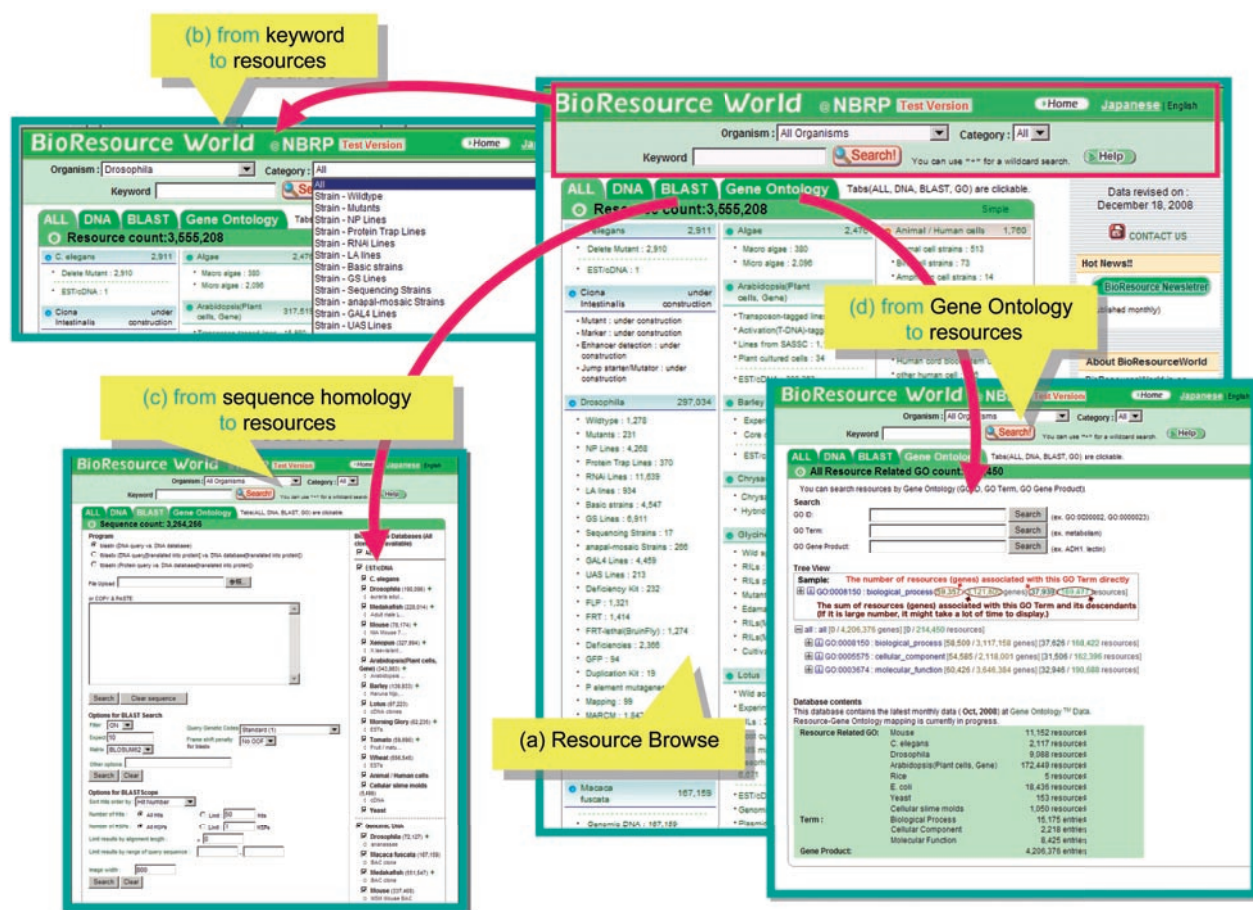


Fig. 5. NBRP resource single site (BRW)(a). This site provides a search service by keywords (b), sequence homology (c), and gene ontology (d) of all NBRP resources.

intends to construct an information platform that will mesh with data provided by public databases independent of GBIF, e.g., genes and genome sequences by the International Nucleotides Sequence Database Collaboration (<http://www.insdc.org/>), geological data, climate and environmental data, and even socio-economic data.

The concept and plan of the GBIF was proposed by the Organization for Economic Cooperation and Development (OECD) Megascience Forum Working Group on Biological Informatics and endorsed by the OECD Committee for Scientific and Technological Policy at the ministerial level in 1999. GBIF was set up independent of the OECD in 2001. The first Governing Board meeting of GBIF was held in Montréal, Canada in March 2001 and Denmark was selected to host the GBIF Sec-

retariat at the second Governing Board meeting in Bonn in June 2001. The GBIF is now financially supported by 29 countries that signed the Memorandum of Understanding with the GBIF Secretariat. Japan has been one of the major participants from the beginning, both conceptually and financially.

The GBIF conducted a feasibility study by networking nodes distributed geographically and thematically. The feasibility study earned the following accolades by third-party evaluation teams [6].

"It is the right initiative with the right goals at the right time."

"In our view, if it did not exist, it would need to be created."

The GBIF entered its second 5-year phase, from 2007 to 2011, declaring that it would cover 1.8 million species

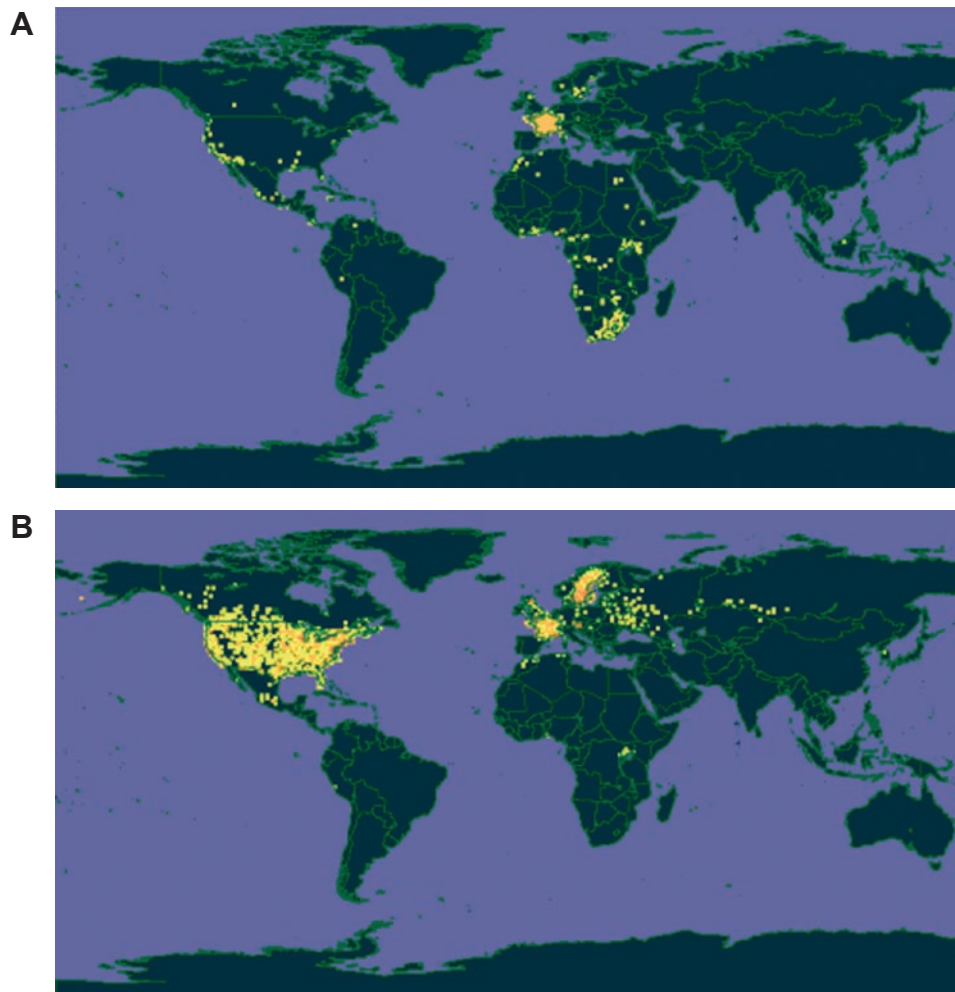


Fig. 6. Map of all available georeferenced records of *Hirundo rustica* (barn swallow) in January (A) and July (B). The one degree cells of the map are color-coded depending on the density of occurrences.

and one billion specimen and observation records.

Utilization of GBIF

As of December 2008, a GBIF user could search more than 160 million records in the GBIF query system. These data were registered at GBIF by 274 data providers. The GBIF Japan node constructed a portal site in Japanese to utilize the GBIF data. Customization of the GBIF data portal by the GBIF member countries is recommended by the GBIF community to improve the accessibility and visibility of GBIF. The GBIF Japan data portal site is located at <http://gbif.ddbj.nig.ac.jp/>. The user is prompted to enter either a common name, scientific name, country name, or the name of a dataset in the

box in the third block from the top to start the search. The query is actually processed by a server maintained at the GBIF Secretariat.

Figure 6 shows the result of a sample query. The user entered “swallow” as a keyword, selected “Barn swallow” in the list suggested by the system, and overlaid records retrieved for January and July, respectively, in Fig. 6A and 6B by using the mashup implemented in the GBIF system. Although this example is a simple mashup of GBIF data and a Google map maintained outside the GBIF, it clearly demonstrates that bird migration is dependent on season.

Figure 7 is a map of *Mus musculus* distribution. It should be noted that there are many occurrences of *Mus*

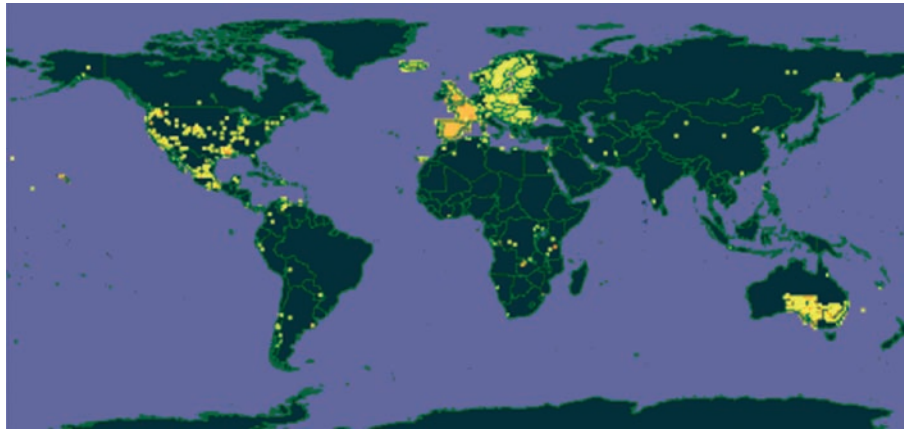


Fig. 7. Map of all available georeferenced records for *Mus musculus*. Central Europe is a reddish color, indicating there are numerous occurrences of *Mus musculus*. The map indicates that there has been no occurrence of *Mus musculus* in Japan.

musculus in Europe but none in Japan. GBIF is unable to determine whether these results reflect the real world situation or not. There are three possibilities for the state of data in the system: the data is precise, no data have been collected, or no data have been uploaded to GBIF. Therefore, we must be careful when we use GBIF data.

The GBIF combines data from multiple data providers in many countries. The databases are constructed in a variety of ways and often cannot communicate with each other. GBIF has prepared standard data formats and data transfer protocols to enable compatibility of databases. DarwinCore is a typical example of a GBIF standard and many GBIF data providers have implemented DarwinCore on top of their own databases to make their databases compatible with the GBIF data portal.

GBIF activities in Japan

Table 1 summarizes the number of data providers and the number of records contributed by individual countries. The United States is the largest in terms of the number of data providers and records, and has contributed approximately 30% of GBIF records. The US is followed by Great Britain and four European Union countries. These five countries in Europe have long histories of promoting natural history museums, biological collections, and also taxonomy. Australia and Costa Rica, both of which are very active in matters related to biodiversity, occupy the seventh and eighth

Table 1. Number of data providers and data records by country

Country	Data providers	Records
United States	72	55,386,642
Great Britain	8	17,243,292
Sweden	1	15,756,511
France	12	9,661,578
Germany	19	6,917,871
Netherlands	3	5,274,715
Australia	7	3,060,228
Costa Rica	2	2,828,062
Austria	10	2,616,381
Spain	5	2,550,328
Norway	3	2,420,012
Canada	9	1,534,267
New Zealand	1	1,460,967
Japan	4	1,343,453
Poland	27	1,128,357
South Korea	16	1,063,900
Mexico	4	980,597
Denmark	2	702,586
Switzerland	2	484,675
Finland	4	461,518
Iceland	1	458,396
Israel	1	430,857
Belgium	4	404,961
Slovenia	3	257,831
Taiwan	2	175,788
Colombia	2	136,155
Argentina	8	73,108
Peru	3	40,078
Italy	1	34,060
India	1	29,449
Portugal	3	25,215
Hong Kong	1	1,901
Pakistan	1	853
Tanzania	1	72

positions. Japan climbed to 14th position in 2008 due to the efforts of the Japan Science and Technology Agency and NBRP of MEXT. NBRP has promoted the networking of natural history museums in Japan by developing an infrastructure of data digitization and publication of specimen data. The National Museum of Nature and Science (NMNS) in Tokyo serves as the hub for 31 networked museums [2].

In the framework of the NBRP, the National Institute of Genetics, NMNS and the University of Tokyo have collaboratively organized annual workshops on biodiversity studies in the 21st century since 2006. The goals of the workshops are to publicize GBIF, promote digitization and capture of biodiversity data, and to promote biodiversity informatics, e.g., simulation of the spread of organisms by combining observation data for organisms, geological data, and environmental data [1, 3]. Figure 8 shows an example of niche modeling of the spread of smallmouth bass in the United States and Japan. With this kind of analysis, we will be able make rational management decisions that will impact future generations.

Barcode of life

The 2011 target of the GBIF is to collect specimen and observation data for 1.8 million species. The number of species has been estimated to be approximately eight million worldwide and confronted with the enormous challenge of efficiently recording and storing identification information for 8 million species. Professor Hebert of the University of Guelph, Canada, proposed barcoding organisms using a gene sequence [4, 5]. For animals, the gene sequence used for such a barcode is the cytochrome c oxidase subunit 1 mitochondrial region. More than 520,000 barcode sequences of 50,000 species of bird, fish, fungi, and other taxa have been captured in the barcode database. The Japanese portal site for the GBIF provides an identification system based on barcode genes and barcode-like genes at <http://bol.ddbj.nig.ac.jp/>.

In the near future, all of the analyses introduced in Figs. 2–5 will be conducted using barcodes derived from genes. The barcode database will be a more powerful resource for biodiversity studies, particularly given the rapid progress that has been made in the area of genetic

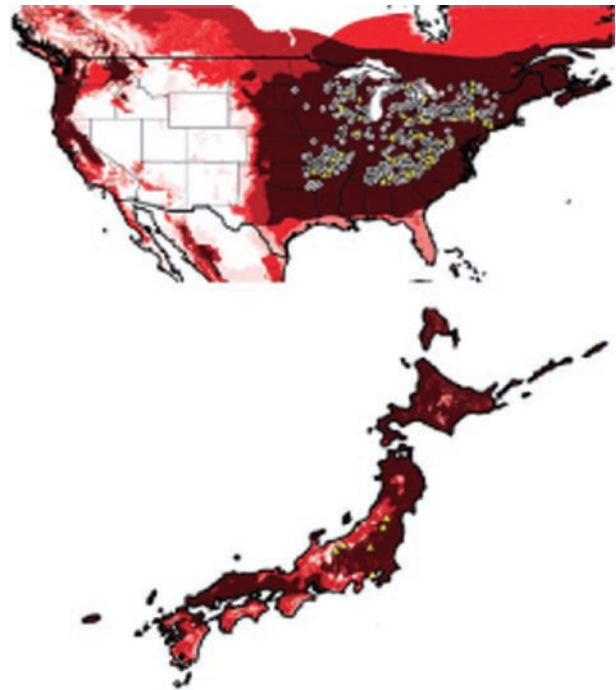


Fig. 8. Integrated results of ten niche models showing the spread of smallmouth bass in the US and Japan. Yellow triangles indicate occurrences. Dark red areas show locations at which the probability of increased distribution is higher.

sequencing.

Conclusion

The activities of the NBRP information center consist of two different programs related to bioresources. These two programs have certain common characteristics, and both the GBIF and some NBRP databases are made available through international collaboration. All resources described in the NBRP databases and some GBIF specimens can be distributed and used for experimental research. While the approaches of the two programs differ, both are very important.

Projects such as these require a long time to become established and succeed in achieving their original goals. However, failure to implement such programs would result in some information being lost very quickly. We believe that the most important task is to continue with these programs properly, improving on successive generations of these databases with proper periodic evaluations.

References

1. Iguchi, K., Matsuura, K., McNyset, K.M., Peterson, A.T., Scachetti-Pereira, R., Powers, K.A., Vieglais, D.A., Wiley, E.O., and Yodo, T. 2004. Predicting invasions of North American basses in Japan using native range data and a genetic algorithm. *Trans. Am. Fish. Soc.* 133: 845–854.
2. Matsuura, K. 2007. The networking of natural history museums and biodiversity information. *Iden* 61: 36–41 (in Japanese).
3. Mitsuhashi, H. 2007. Prediction of distribution of wild organisms based on biodiversity information. *Iden* 61: 30–35 (in Japanese).
4. Ratnasingham, S. and Hebert, P.D. 2007. BOLD: The barcode of life data system. *Mol. Ecol. Notes* 7: 355–364.
5. Stoeckle, M.Y. and Hebert, P.D. 2008. Barcode of life. *Sci. Am.* 299: 82–88.
6. The Review Committee. 2005. “Global Biodiversity” The GBIF 3rd-Year Review Report from the Review Committee Executive Summary. [Online] http://www2.gbif.org/3YR_summary.pdf.