BMC Bioinformatics

BioMed Central

Research

An Integrated Korean Biodiversity and Genetic Information Retrieval System

Jeongheui Lim[1], Jong Bhak[1], Hee-Mock Oh[2], Chang-Bae Kim[3], Yong-Ha Park[4] and Woon Kee Paek*[5]

Address: [1]Korean BioInformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, 305-806, Korea, [2]Biological Resource Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, 305-806, Korea, [3]Major in Life Science, Sangmyung University, Seoul, 110-743, Korea, [4]Department of Applied Microbiology, Yeungnam University, Kyeongsan, 712-749, Korea and [5]Division of Natural History, National Science Museum, Daejeon, 305-705, Korea

Email: Jeongheui Lim - jhlim@kribb.re.kr; Jong Bhak - jongbhak@yahoo.com; Hee-Mock Oh - heemock@kribb.re.kr; Chang-Bae Kim - evodevo@smu.ac.kr; Yong-Ha Park - peter@ynu.ac.kr; Woon Kee Paek* - paekwk@mest.go.kr

* Corresponding author

from Asia Pacific Bioinformatics Network (APBioNet) Seventh International Conference on Bioinformatics (InCoB2008)
Taipei, Taiwan. 20–23 October 2008

Published: 12 December 2008

BMC Bioinformatics 2008, **9**(Suppl 12):S24 doi:10.1186/1471-2105-9-S12-S24

This article is available from: http://www.biomedcentral.com/1471-2105/9/S12/S24

Abstract

Background: On-line biodiversity information databases are growing quickly and being integrated into general bioinformatics systems due to the advances of fast gene sequencing technologies and the Internet. These can reduce the cost and effort of performing biodiversity surveys and genetic searches, which allows scientists to spend more time researching and less time collecting and maintaining data. This will cause an increased rate of knowledge build-up and improve conservations. The biodiversity databases in Korea have been scattered among several institutes and local natural history museums with incompatible data types. Therefore, a comprehensive database and a nation wide web portal for biodiversity information is necessary in order to integrate diverse information resources, including molecular and genomic databases.

Results: The Korean Natural History Research Information System (NARIS) was built and serviced as the central biodiversity information system to collect and integrate the biodiversity data of various institutes and natural history museums in Korea. This database aims to be an integrated resource that contains additional biological information, such as genome sequences and molecular level diversity. Currently, twelve institutes and museums in Korea are integrated by the DiGIR (Distributed Generic Information Retrieval) protocol, with Darwin Core2.0 format as its metadata standard for data exchange. Data quality control and statistical analysis functions have been implemented. In particular, integrating molecular and genetic information from the National Center for Biotechnology Information (NCBI) databases with NARIS was recently accomplished. NARIS can also be extended to accommodate other institutes abroad, and the whole system can be exported to establish local biodiversity management servers.

Conclusion: A Korean data portal, NARIS, has been developed to efficiently manage and utilize biodiversity data, which includes genetic resources. NARIS aims to be integral in maximizing bioresource utilization for conservation, management, research, education, industrial applications, and integration with other bioinformation data resources. It can be found at http://www.naris.go.kr.

Background

Plant and animal specimen data with surveys and observational data stored in museums and herbaria provide a vast information resource; these data include not only the historic information going back several hundred years, but also present day information on the locations of these entities [1]. The availability of this specimen data in online databases is greatly improving science and reducing cost and effort by providing for more efficient and effective biological surveys, which allow scientists to spend more time on research [2].

Traditionally, collections in museums and herbaria were made with only one main purpose in mind, taxonomic study, but their long-term mission has been to document biodiversity and its distribution through time and space for research, education, and service to the public [3]. The introduction of computer databases has opened up this vast data storage to many new uses [4]. These uses include biogeographic studies [5], conservation planning [6], reserve selection [7], climate change studies [8,9], species translocation studies [10], etc.

The Internet's development has allowed new opportunities for interchanging data. Until the Species Analyst project [11], which began in the late 90's, there had been few successful electronic data interchange projects that utilized the Internet. Since then, a number of distributed projects have been initiated: the Mammal Networked Information System (MaNIS) [12], the World Network on Biodiversity (CONABIO) [13], the Australian Virtual Herbarium (AVH) [14], and the European Natural History Specimen Information Network (ENHSIN) [15]. Furthermore, a number of international cooperative projects (e.g., GBIF data portal, Species 2000, Consortium for the Barcode of Life (CBOL), Encyclopedia of Life (EOL), Ocean Biogeographic Information System (OBIS), Scratchpads [16], and WikiSpecies [17]) have recently begun to link and distribute biodiversity data among countries and international organizations. A general trend in bioscience is that advanced bioinformatics analysis and link-up tools are overcoming the challenges of different database types [18,19]. For example, NCBI's databases contain information from molecular to species in different layers [20]. These layers are now becoming linked and integrated through genetic information from fast sequencing projects such as metagenomics; this process also accelerates biotechnological applications. The most important aspect of integration is that once these systems are deployed world wide, it will be possible for machines to process biodiversity information automatically, assuming the data are accurate. This will also accelerate standardization and raise the efficiency of the process of applying biodiversity studies by many magnitudes.

The biodiversity databases in Korea have been scattered among several institutes and local natural history museums, and the data were made up of heterogeneous types of databases with different formats and properties; a centralized standardized portal that enabled access to biodiversity information was necessary. A web portal will be able to speed up the investigation of complicated biological inquiries, allow researchers to develop new knowledge by analyzing large data sets, and provide the appropriate analytical tools [21]. For example, identification studies of predominant areas for conservation [22] and impacts of climate change across natural systems [23] will be increased through the accessibility of large, integrated databases.

We built NARIS, a database and national website for biodiversity, to manage and utilize the biodiversity resources in Korea and beyond. This system can be used in other nations with minimum modifications and language translations. At its core, NARIS is a biodiversity database, but it includes links to molecular and genetic information from the NCBI databases. NARIS aims to promote conservation, management, education, research, and industrial applications in biodiversity. In particular, with the advancements in fast genome sequencing technology, we expect that most of the common species will be completely sequenced before 2015. An integrated biodiversity management system such as NARIS will provide a platform that directly integrates the vast amount of genetic diversity information in the existing species diversity information.

Methods

System architecture

NARIS was built through the cooperation of many organizations from industry, universities, and research institutes in Korea. The participants are presented in Table 1. As shown in Figure 1, the biodiversity data stored in each organization was integrated with the NARIS database using the Darwin Core2.0 metadata standard along with DiGIR protocol [24], which is a request and response message format for communication between data provider, portal engine, and applications. By linking the metadata and the existing databases already available in each organization, any addition, deletion, or modification of the

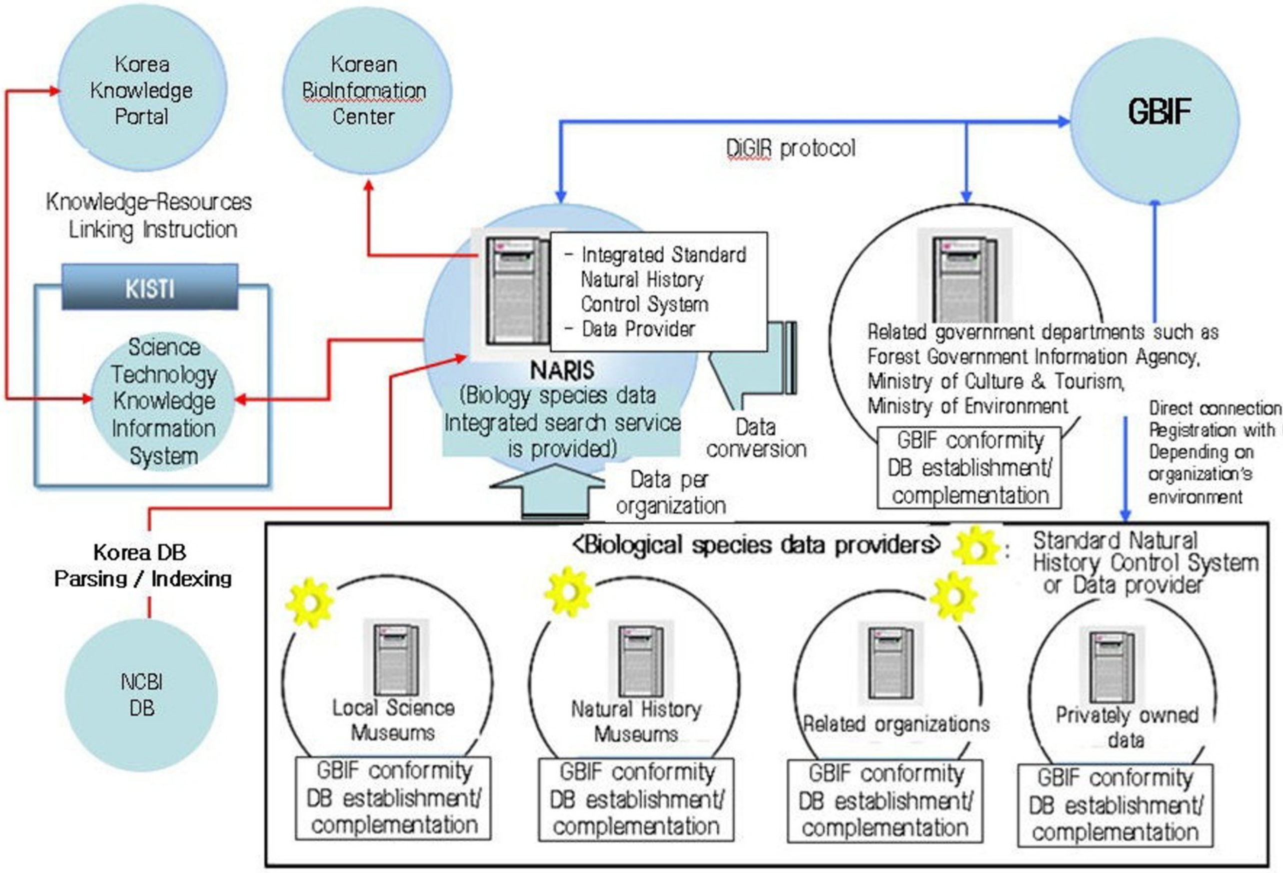

Korea Knowledge Portal
Korean BioInfomation Center
GBIF
DiGIR protocol
Knowledge-Resources Linking Instruction
KISTI
Science Technology Knowledge Information System
- Integrated Standard Natural History Control System
- Data Provider
NARIS (Biology species data Integrated search service is provided)
Data conversion
Data per organization
Related government departments such as Forest Government Information Agency, Ministry of Culture & Tourism, Ministry of Environment
GBIF conformity DB establishment/ complementation
Direct connection Registration with GBIF Depending on organization's environment
Korea DB Parsing / Indexing
NCBI DB
<Biological species data providers>
: Standard Natural History Control System or Data provider
Local Science Museums
Natural History Museums
Related organizations
Privately owned data
GBIF conformity DB establishment/ complementation
GBIF conformity DB establishment/ complementation
GBIF conformity DB establishment/ complementation
GBIF conformity DB establishment/ complementation

Figure 1
NARIS System Architecture. The data from several institutes and museums in Korea was integrated with NARIS by DiGIR protocol for data sharing between data providers, portal engine, GBIF, Korean BioInformation Center, Korea Science Technology Knowledge Information System, and Korea Knowledge Portal.

databases is automatic, so the metadata are integrated into NARIS. The adoption of the DiGIR protocol enables data sharing between GBIF, Korean BioInformation Center (KOBIC), Korea Science Technology Knowledge Information System (KISTI), and Korea Knowledge Portal.

Figure 2 shows the steps for linking NARIS keywords to NCBI's genetic and taxonomic data. First, NARIS retrieves and stores relevant NCBI data in a table using WebCrawler and parsing tool. Second, NCBI's search fields are selected from the NARIS database via SPIDER and Gateway; the search fields are type, category, scientific name, affiliation, sequence, and abstract. Third, the Filter converts the binary data into text. Fourth, the keyword sets are extracted at the Locale stage and stored by Collection Index schema. When a user inputs a search word in the web browser, this is converted to the database Query, and a search engine finally shows the results from the Collection Index.

The NARIS system uses an XML-based database connection module. The search engine supports a logical calculator, wildcards, Processor, API (C, Java-based), web-based management, and database connection with WAS and Oracle. It is equipped with several functions: a multi-database search, a document filter, a synonym/thesaurus dictionary, morpheme analyses, popular and recommended keywords, and statistical analyses. The J2EE (Java 2 Platform, Enterprise Edition) technology has been used for web service, as well as JDBC (Java Database Connectivity), using a database connection pool.

Data format and protocol

The NARIS data format includes 79 fields, with additional field extensions in addition to the Darwin Core2.0 standard format, to incorporate required variables for the comprehensive interchange of biodiversity data. Figure 3 shows the schematic diagram of the DiGIR protocol, which allows NARIS data to be integrated with other

Table 1: Participants of NARIS. As of August 2008, the data generated from 12 institutes and natural history museums in Korea are standardized with the Darwin Core2.0 format and integrated with NARIS to share and exchange the data.

Organizations	Records of Data
National Science Museum	137,680
Korea National Arboretum	980,000
Gyeryongsan Natural History Museum	4,000
Kyunghee University Natural History Museum	6,181
Mokpo Natural History Museum	17,754
Hannam University Natural History Museum	5,000
Seodaemun Museum of Natural History	2,010
Ewha Womans University Natural History Museum	5,050
Jeonam Maritime & Fisheries Science Museum	2,000
Folklore & Natural History Museum Jeju Special Self-Governing Province	54,588
Chungnam University Natural History Museum	1,005
Busan Marine Natural History Museum	2,275

informatics initiatives, such as GBIF and OBIS, to retrieve, link, and integrate the metadata from each relevant organization. The DiGIR protocol allows more precise data searches because it is designed to reduce information overload and extract common characteristics from all data sources that are shown in different formats.

Results and discussion

User interface

The NARIS user-interface mainly consists of a search page and an integrated search result page. The search page has a single input window, and the search result page shows integrated information with various items: species descriptions, the introduction of protective species and endangered species, natural heritages, molecular and genetic information, geographical distribution, specimen and observational data, biodiversity related organizations, videos, images, news, references, etc. As shown in Figure 4, by the integration with NCBI's genetic data, NARIS allows users to search genetic information (e.g., nucleotide, protein, sequence) in a single search result page.

Unique functionalities

The data error report function has been implemented to constantly monitor and control the data quality. If a user finds incorrect data (e.g., nomenclatural, taxonomic, or spelling errors) while searching data and checking the results, then the user can report the errors using the comment box (see Figure 5). In accordance with the data error management procedure shown in Figure 6, the message is reported to a NARIS administrator, and subject area experts (e.g., taxonomic specialists) assess it and revise the incorrect data. However, the records of the modified data are kept and made visible, so qualified researchers can compare and correct them as necessary. This function is necessary for concerted efforts to improve the quality of the data. All data have errors, but that should not be a reason not to use the data, but to ensure that the error is documented and that users are made aware of the errors so that they may determine the fitness for the use of the data [25,26]. As shown in Figure 7, NARIS also has a useful function for statistical analysis that provides the reference materials for scientific research and education on biodiversity, including species diversity, dominance, and similarity.

Future development

The future development of NARIS will contain several new functions. First, a personalized web-service (e.g., an individual database management system) will be included to attract voluntary participation of individual users. Through this system, individual users will be able to upload their own high quality and quantity of data, which

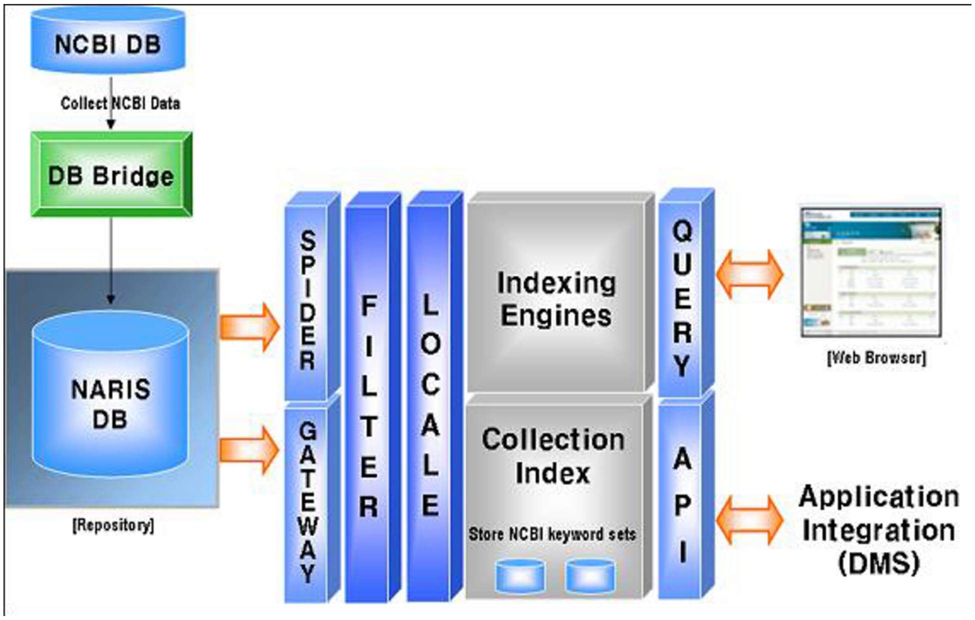


Figure 2
NCBI Data Search and Link Engine.

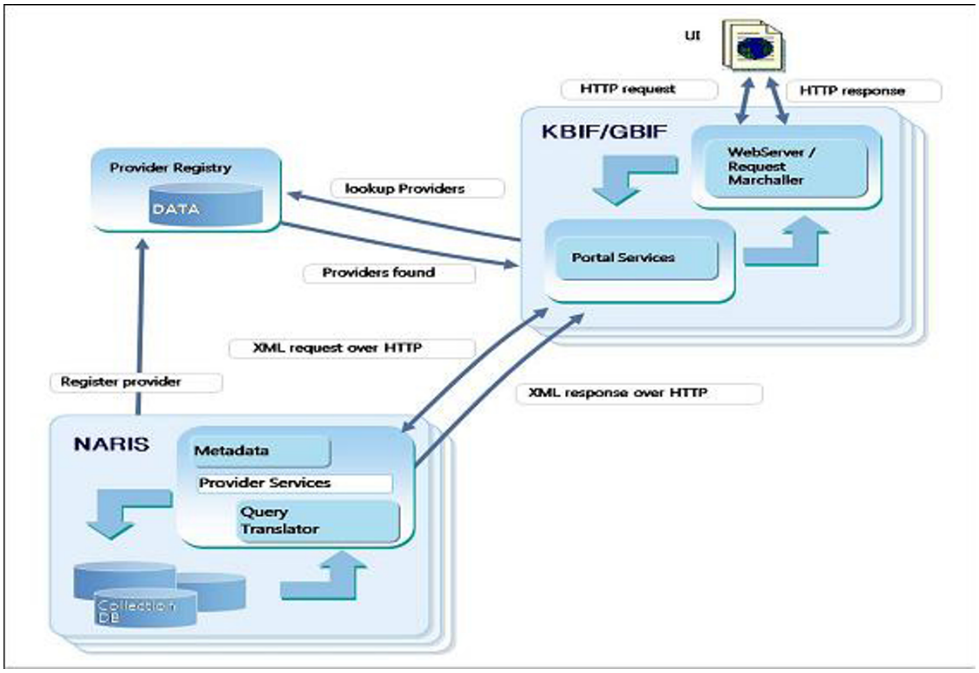
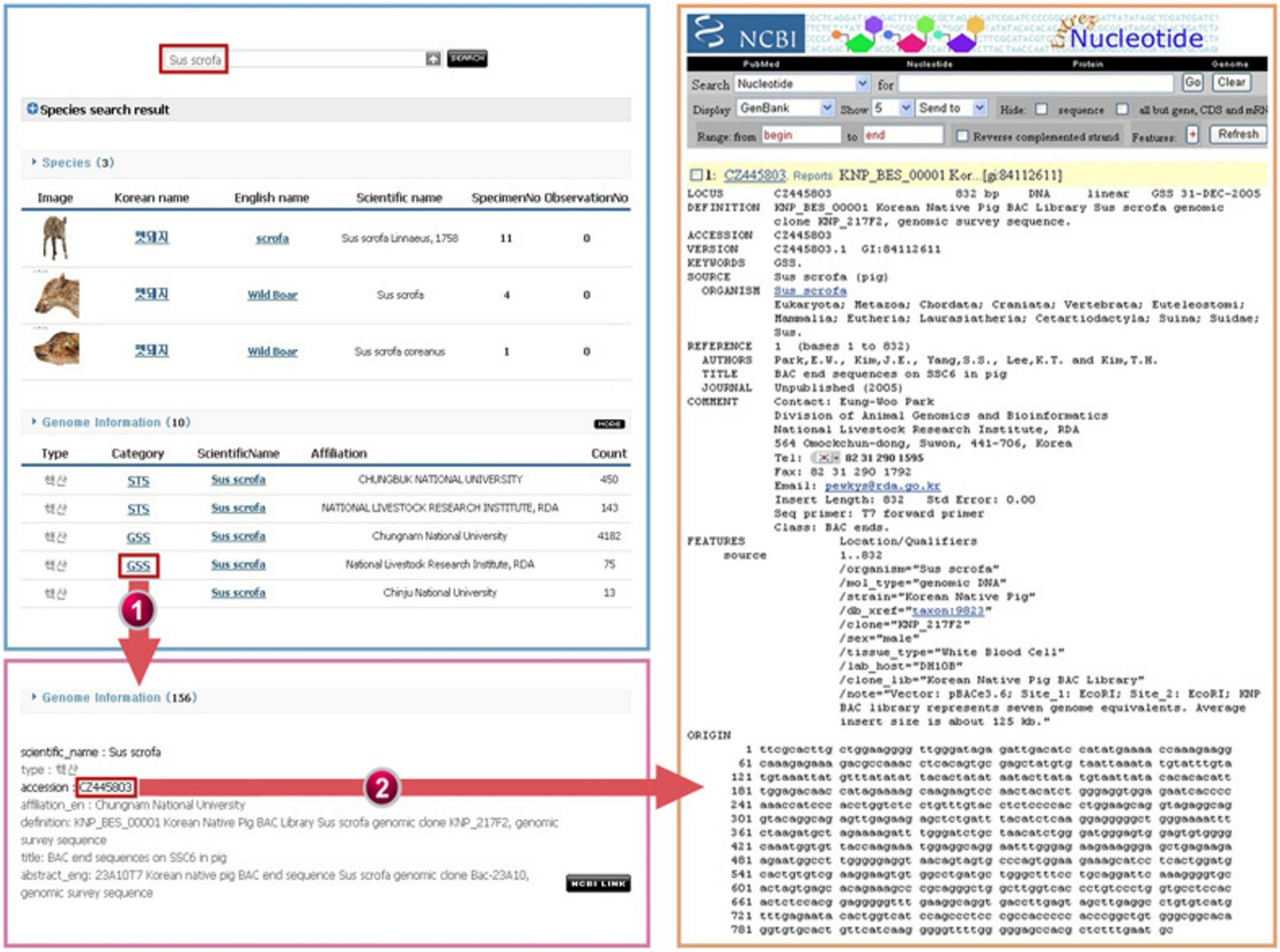


Figure 3
Schematic Diagram of the DiGIR Protocol. The DiGIR protocol is installed in each data provider's server, and the data are extracted after matching with provider databases. The data are saved in a temporary storage of Provider Registry by XML, and then they are transmitted to each portal system, along with corresponding metadata collected through the DiGIR software installed in each data provider's web server.



Sus scrofa

Species search result

Species (3)

Image	Korean name	English name	Scientific name	SpecimenNo	ObservationNo
	멧돼지	scrofa	Sus scrofa Linnaeus, 1758	11	0
	멧돼지	Wild Boar	Sus scrofa	4	0
	멧돼지	Wild Boar	Sus scrofa coreanus	1	0

Genome Information (10)

Type	Category	ScientificName	Affiliation	Count
핵산	STS	Sus scrofa	CHUNGBUK NATIONAL UNIVERSITY	450
핵산	STS	Sus scrofa	NATIONAL LIVESTOCK RESEARCH INSTITUTE, RDA	143
핵산	GSS	Sus scrofa	Chungnam National University	4182
핵산	GSS	Sus scrofa	National Livestock Research Institute, RDA	75
핵산		Sus scrofa	Chinju National University	13

Genome Information (156)

scientific_name : Sus scrofa
type : 핵산
accession : CZ445803
affiliation_en : Chungnam National University
definition: KNP_BES_00001 Korean Native Pig BAC Library Sus scrofa genomic clone KNP_217F2, genomic survey sequence
title: BAC end sequences on SSC6 in pig
abstract_eng: 23A10T7 Korean native pig BAC end sequence Sus scrofa genomic clone Bac-23A10, genomic survey sequence

Search Nucleotide for
Display GenBank Show 5 Send to Hide: □ sequence □ all but gene, CDS and mRN
Range: from begin to end □ Reverse complemented strand Features: +

□ 1: CZ445803. Reports KNP_BES_00001 Kor...[gi:84112611]

```
LOCUS       CZ445803                 832 bp    DNA     linear   GSS 31-DEC-2005
DEFINITION  KNP_BES_00001 Korean Native Pig BAC Library Sus scrofa genomic
            clone KNP_217F2, genomic survey sequence.
ACCESSION   CZ445803
VERSION     CZ445803.1  GI:84112611
KEYWORDS    GSS.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
            Sus.
REFERENCE   1  (bases 1 to 832)
  AUTHORS   Park,E.W., Kim,J.E., Yang,S.S., Lee,K.T. and Kim,T.H.
  TITLE     BAC end sequences on SSC6 in pig
  JOURNAL   Unpublished (2005)
COMMENT     Contact: Eung-Woo Park
            Division of Animal Genomics and Bioinformatics
            National Livestock Research Institute, RDA
            564 Omockchun-dong, Suwon, 441-706, Korea
            Tel: 82 31 290 1595
            Fax: 82 31 290 1792
            Email: pewkys@rda.go.kr
            Insert Length: 832   Std Error: 0.00
            Seq primer: T7 forward primer
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..832
                     /organism="Sus scrofa"
                     /mol_type="genomic DNA"
                     /strain="Korean Native Pig"
                     /db_xref="taxon:9823"
                     /clone="KNP_217F2"
                     /sex="male"
                     /tissue_type="White Blood Cell"
                     /lab_host="DH10B"
                     /clone_lib="Korean Native Pig BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; KNP
                     BAC library represents seven genome equivalents. Average
                     insert size is about 125 kb."
ORIGIN
        1 ttcgcacttg ctggaagggg ttgggataga gattgacatc catatgaaaa ccaaagaagg
       61 caaagagaaa gacgccaaac ctcacagtgc gagctatgtg taattaaata tgtatttgta
      121 tgtaaattat gtttatatat tacactatat aatacttata tgtaattata cacacacatt
      181 tggagacaac catagaaaag caagaagtcc aactacatct gggaggtgga gaatcacccc
      241 aaaccatccc acctggtctc ctgtttgtac ctctccccac ctggaagcag gtagaggcag
      301 gtacaggcag agttgagaag agctctgatt tacatctcaa ggaggggggct gggaaaattt
      361 ctaagatgct agaaaagatt tgggatctgc taacatctgg gatgggagtg gagtgtgggg
      421 caaatggtgt taccaagaaa tggaggcagg aatttgggag aagaaaggga gctgagaaga
      481 agaatggcct tgggggaggt aacagtagtg cccagtggaa gaaagcatcc tcactggatg
      541 cactgtgtcg aaggaagtgt ggcctgatgc tgggctttcc tgcaggattc aaaggggtgc
      601 actagtgagc acagaaagcc cgcagggctg gcttggtcac cctgtccctg gtgcctccac
      661 actctccacg gaggggggttt gaaggcaggt gaccttgagt agcttgaggc ctgtgtcatg
      721 tttgagaata cactggtcat ccagccctcc cgccaccccc acccggctgt gggcggcaca
      781 ggtgtgcact gttcatcaag ggggttttgg gggagccacg ctctttgaat gc
```

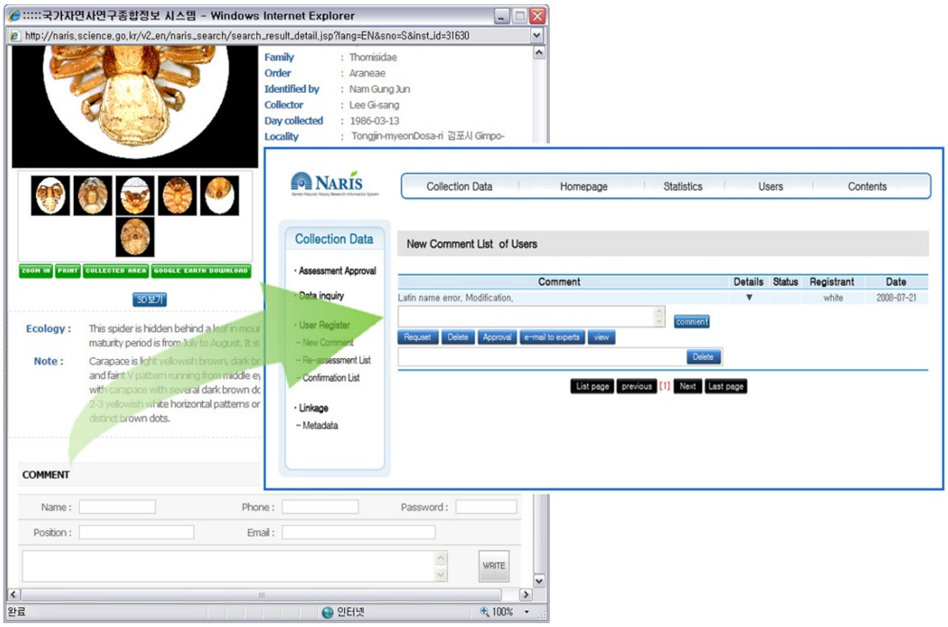
Figure 4
Sample of Linked Genetic Information for *Sus scrofa* (pig). The integrated information for *Sus scrofa*, which included taxonomic data and genetic data from NCBI databases, can be accessed in a single search result page.

will be registered on the NARIS with a verification process. Another multi-media function such as sounds will be provided for access to more exciting services. Two other new functions will be a system to distribute the actual collections and data contents, which can be actively utilized in academic researches or industrial applications, and support for taxonomic studies by providing not only morphological data but also DNA barcode data that have been recently used to identify species. Finally, molecular and genetic information that links to other bioinformatics portals, such as European Bioinformatics Institute (EBI) databases, will be offered.

Conclusion

With the advancement of information technology, the Internet, and rapid gene sequencers, managing biodiversity information in association with genetic information is becoming a global issue. Sharing and disseminating biodiversity information among different countries through relevant international organizations is becoming the standard practice.

In this paper, we presented an effective data management system, NARIS, which is a centralized data portal enabling an integrated information retrieval of the distributed biodiversity data and is linked with genetic information from NCBI. NARIS was specifically developed in Korea but is applicable to other nations. NARIS uses the DiGIR protocol with Darwin Core2.0 format as its metadata standard for data exchange to integrate and share the biodiversity data among organizations in Korea and abroad and to play an important role as a data node of the GBIF as well.



http://naris.science.go.kr/v2_en/naris_search/search_result_detail.jsp?lang=EN&sno=S&inst_id=31630

Family : Thomisidae
Order : Araneae
Identified by : Nam Gung Jun
Collector : Lee Gi-sang
Day collected : 1986-03-13
Locality : Tongjin-myeonDosa-ri 김포시 Gimpo-

Ecology : This spider is hidden behind a leaf in mour
maturity period is from July to August. It is

Note : Carapace is light yellowish brown, dark br
and faint V pattern running from middle ey
with carapace with several dark brown dc
2-3 yellowish white horizontal patterns or
distinct brown dots.

New Comment List of Users

Comment	Details	Status	Registrant	Date
Latin name error. Modification.	▼		white	2008-07-21

Name :
Phone :
Password :
Position :
Email :

Figure 5
Data Error Report Function. A user can report a datum error in the comment box on NARIS's web site. Then the message is reported to a NARIS administrator.

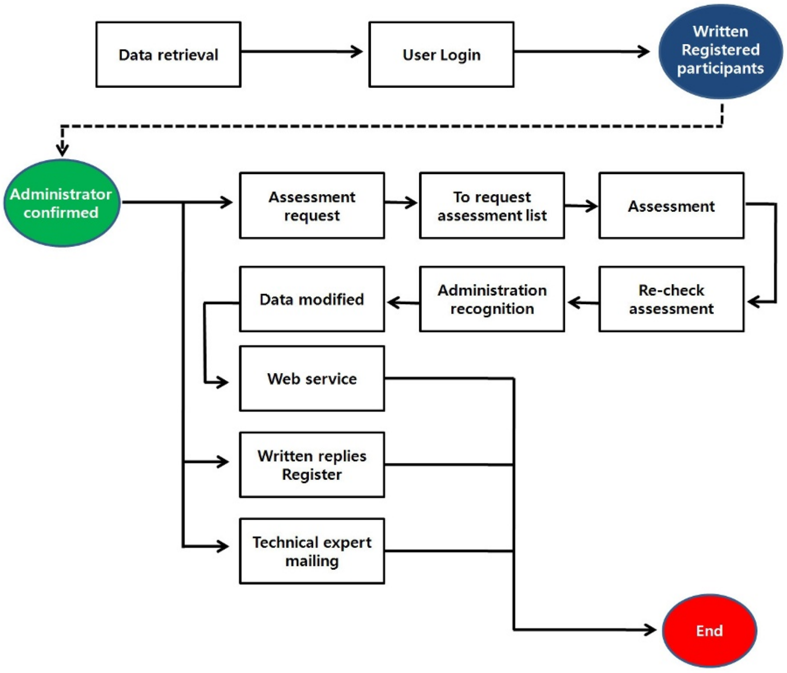


Figure 6
Data Error Management Procedure. An incorrect datum reported from a user on NARIS's web site is reported to the NARIS administrator; the subject area experts assess it; and the incorrect datum is revised.

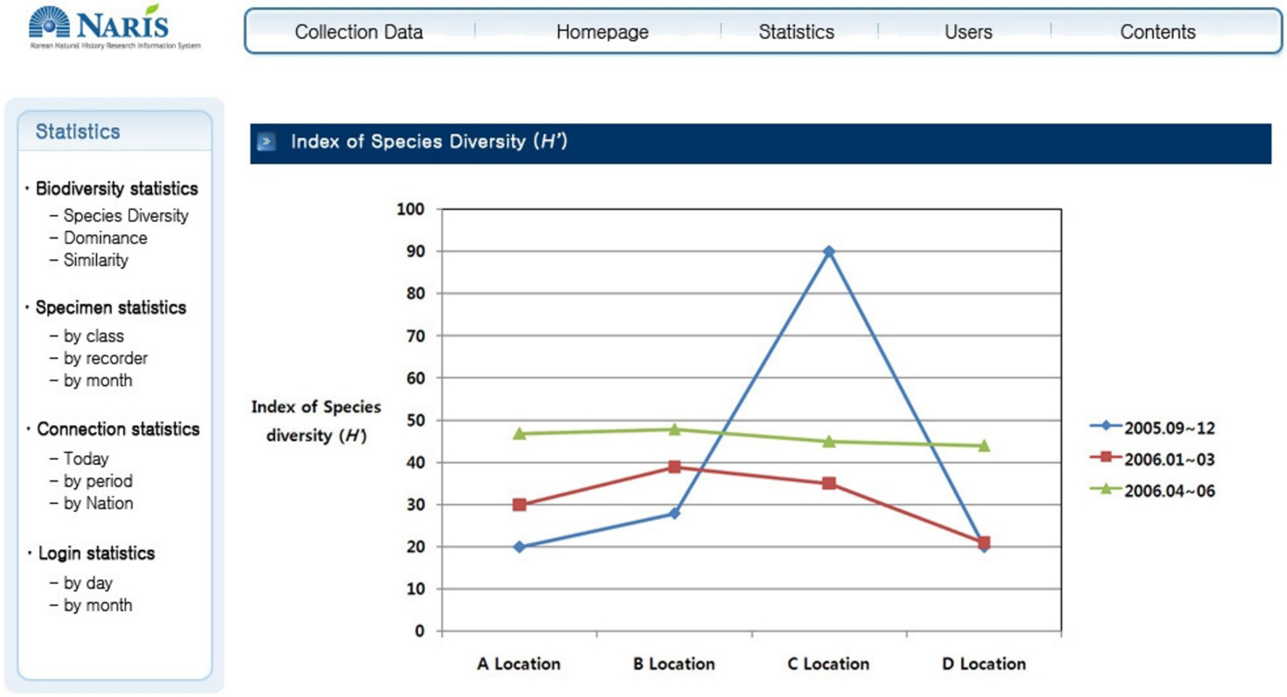


Figure 7
Statistical System. The X axis represents the location of survey, and the Y axis represents the index of species diversity. Each line (triangle, rectangle, and circle) indicates the index of species diversity for each survey period.

Considering the growing importance of biodiversity resources as the original source of bio-industry, establishing this integrated biodiversity information system is expected to be integral in protecting valuable national natural resources. Also, the system will be useful for the study of species and their distributions, conservation, scientific research, education, natural resource management, climate change, social and political uses, and medicinal studies.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

WKP directed the study and helped draft the manuscript. JB, HMO, CBK, and YHP were involved in reviewing and critically revising it for intellectual content. JHL conceived the study and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Seunghun Baek, Hee-Yeun Lee, and Seong-Yong Yang, who provided technical assistance while constructing the database. This work was supported by a grant from the Ministry of Education, Science and Technology of Korea (M10868000001-08N6800-00100), the Korean Research Institute of Bioscience and Biotechnology (KRIBB) Research Initiative Program, Ministry of Environment of Korea (2008-05002-0065-0), and the Yeungnam University Research Program in 2008.

This article has been published as part of *BMC Bioinformatics* Volume 9 Supplement 12, 2008: Asia Pacific Bioinformatics Network (APBioNet) Seventh International Conference on Bioinformatics (InCoB2008). The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2105/9?issue=S12.

References

1. Chapman AD, Busby JR: **Linking plant species information to continental biodiversity inventory, climate and environmental monitoring.** In *Mapping the Diversity of Nature* Edited by: *Miller RI. London: Springer*; 1994:177-195.
2. Chapman AD: **Uses of primary species-occurrence data, version 1.0.** *Report for Global Biodiversity Information Facility. Copenhagen* 2005.
3. Winker K: **Natural history museums in a postbiodiversity era.** *BioScience* 2004, **54(5):**455-459.
4. Chapman AD: **Quality control and validation of point-sourced environmental resource data.** In *Spatial accuracy assessment: Land information uncertainty in natural resources* Edited by: *Lowell K, Jaton A. Chelsea: Ann Arbor Press*; 1999:409-418.
5. Peterson AT, Navarro-Sigüenza AG, Benitez-Diaz H: **The need for continued scientific collecting: A geographic analysis of Mexican bird specimens.** *Ibis* 1998, **140:**288-294.
6. Faith DP, Walker PA, Margules CR, Stein J, Natera G: **Practical application of biodiversity surrogates and percentage targets for conservation in Papua New Guinea.** *Pacific Conservation Biology* 2001, **6:**289-303.
7. Margules CR, Pressey RL: **Systematic conservation planning.** *Nature* 2000, **405:**243-253.

8. Root TL, Price JT, Hall KR, Schneider SH, Rosenzweigk C, Pounds JA: **Fingerprints of global warming on wild animals and plants.** *Nature* 2003, **421:**57-60.
9. Peterson AT, Ortega-Huerta MA, Bartley J, Sánchez-Cordero V, Soberón J, Buddemeier RH, Stockwell DRB: **Future projections for Mexican faunas under global climate change scenarios.** *Nature* 2002, **416:**626-629.
10. Peterson AT, Vieglais DA: **Predicting species invasions using ecological niche modeling.** *BioScience* 2001, **51:**363-371.
11. Kaiser J: **Web tools: Searching museums from your desktop.** *Science* 1999, **284(5416):**888.
12. **MaNIS – Mammal Networked Information System** [http://manisnet.org/]
13. **CONABIO – Comisión national para el conocimiento y uso de la biodiversidad** [http://www.conabio.gob.mx/]
14. **AVH – Australian''s Virtual Herbarium** [http://www.anbg.gov.au/avh/]
15. **ENHSIN – European Natural History Specimen Information Network** [http://www.bgbm.org/BioDivInf/projects/ENHSIN/]
16. **Scratchpads** [http://scratchpads.eu/about]
17. **WikiSpecies** [http://species.wikimedia.org/]
18. Stein LD: **Integrating biological databases.** *Nature Reviews Genetics* 2003, **4:**337-345.
19. Shao KT, Peng CI, Yen E, Lai KC, Wang MC, Lin J, Lee H, Alan Y, Chen SY: **Integration of biodiversity databases in Taiwan and linkage to global databases.** *Data Science Journal* 2007, **6:**2-10.
20. **NCBI Taxonomy** [http://www.ncbi.nlm.nih.gov/Taxonomy/]
21. Neale SH, Pulland MR, Watson MF: **Online biodiversity resources – principles for usability.** *Biodiversity Informatics* 2007, **4:**27-36.
22. Williams P, Gibbons D, Margules CR, Rebelo A, Humphries C, Pressey R: **A comparison of richness hotspots, rarity hotspots, and complementary areas for conserving diversity of British birds.** *Conservation Biology* 1996, **10:**155-174.
23. Parmesan C, Yohe G: **A globally coherent fingerprint of climate change impacts across natural systems.** *Nature* 2003, **421:**37-42.
24. **DiGIR – Distributed Generic Information Retrieval** [http://digir.net/]
25. Chapman AD: **Principles and methods of data cleaning.** *Report for Global Biodiversity Information Facility. Copenhagen* 2005.
26. Chapman AD: **Principles of data quality.** *Report for Global Biodiversity Information Facility. Copenhagen* 2005.