

## Korea Barcode of Life Database System (KBOL)

Sungmin Kim<sup>a</sup>, Chang-Bae Kim<sup>b</sup>, Gi-Sik Min<sup>c</sup>, Youngbae Suh<sup>d</sup>, Jong Bhak<sup>e</sup>, Taeha Woo<sup>f</sup>, Hyeyoung Koo<sup>g</sup>, Jun-Kil Choi<sup>h</sup>, Mann Kyoon Shin<sup>h</sup>, Jongwoo Jung<sup>i</sup>, Kyo-Hong Song<sup>j</sup>, Han-Il Ree<sup>k</sup>, Ui Wook Hwang<sup>l</sup>, Yungchul Park<sup>m</sup>, Hae-Seok Eo<sup>n</sup>, Joopil, Kim<sup>o</sup>, Seong Myeong Yoon<sup>p</sup>, Hyun Soo Rho<sup>q</sup>, Sa Heung Kim<sup>r</sup>, Hang Lee<sup>s</sup>, Mi-Sook Min<sup>s</sup>, Jin Koo Kim<sup>t</sup>, Chungja Sim<sup>u</sup>, Yong-Jin Won<sup>v</sup>, Jongwook Lee<sup>w</sup>, Soowon Cho<sup>x</sup>, Ji Eun Seo<sup>y</sup>, Seunghwan Lee<sup>z</sup>, Jun-Im Song<sup>ai</sup>, Ho-Yeon Han<sup>bi</sup>, Sook Shin<sup>ci</sup>, Sung Joon Song<sup>di</sup>, Joong-Ki Park<sup>ei</sup>, Keeseon S. Eom<sup>ei</sup>, Jung Hee Park<sup>fi</sup> and Won Kim<sup>gf\*</sup>

<sup>a</sup>Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-747, Korea; <sup>b</sup>Department of Green Life Science, Sangmyung University, Seoul 110-743, Korea; <sup>c</sup>Department of Biological Sciences, Inha University, Incheon 402-751, Korea; <sup>d</sup>Natural Products Research Institute, Seoul National University, Seoul 151-742, Korea; <sup>e</sup>Theragen Bio Institute, Suwon 443-270, Korea; <sup>f</sup>Omicis Inc., 211, Bio Venture Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea; <sup>g</sup>Department of Biological Science, Sangji University, Wonju 220-702, Korea; <sup>h</sup>Department of Biological Sciences, University of Ulsan, Ulsan 680-749, Korea; <sup>i</sup>Department of Science Education, Ewha Womans University, Seoul 120-750, Korea; <sup>j</sup>Wildlife Genetic Resources Center, National Institute of Biological Resources, Incheon 404-708, Korea; <sup>k</sup>Department of Environmental Medical Biology, Institute of Tropical Medicine, and Korean National Arthropods of Medical Importance Resource Bank, Yonsei University College of Medicine, Seoul 120-752, Korea; <sup>l</sup>Department of Biology, Teachers College & Institute for Phylogenomics and Evolution, Kyungpook National University, Daegu 702-701, Korea; <sup>m</sup>Department of Forest Environment Protection, College of Forest & Environmental Science, Kangwon National University, Chuncheon 200-701, Korea; <sup>n</sup>Future IT R&D Laboratory, LGE Advanced Research Institute, Seoul 137-130, Korea; <sup>o</sup>Joopil Spider Museum, Joan-myeon, Namyangju-si, Gyeonggi-do 472-870, Korea; <sup>p</sup>Department of Biology Education, Chosun University, Gwangju 501-759, Korea; <sup>q</sup>East Sea Research Institute, Korea Ocean Research & Development Institute, Uljin 695-1, Korea; <sup>r</sup>Marine Biodiversity Research institute, IN THE SEA Korea Co., Ltd., Jeju 697-110, Korea; <sup>s</sup>Conservation Genome Resource Bank for Korean Wildlife and Research Institute for Veterinary Science, College of Veterinary Medicine, Seoul National University, Seoul 151-742, Korea; <sup>t</sup>Department of Marine Biology, Pukyong National University, Busan 608-737, Korea; <sup>u</sup>Department of Biological Sciences, College of Life Sciences and Nano Technology, Hannam University, Daejeon 305-811, Korea; <sup>v</sup>Division of EcoScience, Ewha Womans University, Seoul 120-750, Korea; <sup>w</sup>Department of Life Sciences, Yeungnam University, Gyeongsan 214-1, Korea; <sup>x</sup>Department of Plant Medicine, Chungbuk National University, Cheongju 361-763, Korea; <sup>y</sup>Department of Rehabilitation, Woosuk University, Jeonbuk 565-701, Korea; <sup>z</sup>Laboratory of Insect Biosystematics, Division of Entomology, Research Institute for Agricultural and Life Sciences, School of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Korea; <sup>ai</sup>Department of Life Sciences, Ewha Womans University, Seoul 120-750, Korea; <sup>bi</sup>Division of Biological Science and Technology, College of Science and Technology, Yonsei University, Gangwon-do 220-710, Korea; <sup>ci</sup>Department of Life Science, Sahmyook University, Seoul 139-742, Korea; <sup>di</sup>National Park Research Institute, Namwon-si, Jeollabuk-do, 590-81, Korea; <sup>ei</sup>Graduate Program in Cellular Biology and Genetics, and Department of Parasitology, College of Medicine, Chungbuk National University, Cheongju 361-763, Korea; <sup>fi</sup>Department of Life Science, The University of Suwon, Hwaseong 445-743, Korea; <sup>gf</sup>School of Biological Sciences, Seoul National University, Seoul 151-747, Korea

(Received 10 March 2011; accepted 29 May 2011)

A major concern regarding the collection and storage of biodiversity information is the inefficiency of conventional taxonomic approaches in dealing with a large number of species. This inefficiency has increased the demand for automated, rapid, and reliable molecular identification systems and large-scale biological databases. DNA-based taxonomic approaches are now arguably a necessity in biodiversity studies. In particular, DNA barcoding using short DNA sequences provides an effective molecular tool for species identification. We constructed a large-scale database system that holds a collection of 5531 barcode sequences from 2429 Korean species. The Korea Barcode of Life database (KBOL, <http://koreabarcode.org>) is a web-based database system that is used for compiling a high volume of DNA barcode data and identifying unknown biological specimens. With the KBOL system, users can not only link DNA barcodes and biological information but can also undertake conservation activities, including environmental management, monitoring, and detecting significant organisms.

**Keywords:** DNA barcoding; species identification; Korea Barcode of Life database system (KBOL); biodiversity

### Introduction

DNA barcoding is a molecular diagnostic method for identifying biological species using a short DNA

sequence derived from a standardized position on the genome (Hebert and Gregory 2005). Recent studies have promoted the DNA barcoding technique as an

\*Corresponding author. Email: wonkim@plaza.snu.ac.kr

approach to surpass the limits of conventional taxonomic research (Hebert et al. 2003; Hajibabaei et al. 2007). DNA barcoding allows for excellent taxonomic resolution and the rapid identification of organisms in several research fields, which could be critically important to our environment, food supply, and economy (Hajibabaei et al. 2005; Dasmahapatra and Mallet 2006; Park et al. 2008; Kim et al. 2010). As the DNA barcoding initiative was being universalized, a large-scale biodiversity database based on DNA barcodes was urgently required to create a library of new information about species diversity and to make it easier to access the vast number of plant and animal records (Rubinoff 2006; Ratnasingham and Hebert 2007; Radulovici et al. 2010).

The Korea Barcode of Life database system (KBOL) was developed to integrate and manage all biological data, thereby combining molecular and morphological research from various taxa. The KBOL collects biological data such as specimen information, taxonomic status, geographic data, DNA sequence trace files, primer information, and images, along with DNA barcodes. The Korean database also makes it easier to access other databases, including the Barcode of Life Data Systems (BOLD), which supports large storage, management, analysis, and use of DNA barcodes (Ratnasingham and Hebert 2007), and GenBank, which publicly provides genetic information resources (Benson et al. 2008). In addition to providing these basic features, the KBOL database system introduces a molecular identification procedure that is carried out using a profile hidden Markov model (profile HMM) (Krogh et al. 1994; Eddy 1998; Kim et al. 2010) and BLAST (Altschul et al. 1990). The KBOL allows users to manage and search various biological data and acquire new information via directed searches. Therefore, this database can be utilized more frequently in several areas of research, including systematic biology, conservation biology, biodiversity informatics, and population genetics. It will provide a strong foundation for understanding many important biological resources and launching national databases for biodiversity conservation.

Here, we introduce the KBOL as the first large-scale DNA barcode database system in Korea and as a good role model for networking national biodiversity with the international community that is interested in species diversity.

## Materials and methods

### *System architecture and project establishment*

The KBOL database system was completed via four procedures: assembling a high volume of DNA

barcodes, developing a molecular identification system, identifying unknown samples, and analyzing DNA barcodes (Figure 1). The first procedure involved projects aimed at uploading DNA barcodes and related biological information. However, all registered users who conduct various types of research that involve the DNA barcoding technique can create projects. Project managers or a group of collaborators can acquire the rights to secure their private data until they decide to release the data to the public. The second procedure created statistical species models and optimized parameters to complete a molecular identification system based on HMM. DNA barcodes can then be linked to related biological information. The last procedure involved generating new information such as discovering cryptic species, clarifying species relationships, and monitoring species diversity.

### *Data management*

To facilitate data management and access, the KBOL data have been grouped into four categories: barcode overview, general specimen information, taxonomic information, and barcode sequence (Figure 2). All barcode information is processed by the verified BioBarcode scheme (Lim et al. 2009). During the automated upload procedure, an additional barcode-filtering function checks whether the input barcode sequence is a nuclear copy of mitochondrial DNA (NUMT) (Bensasson et al. 2001; Song et al. 2008). The KBOL database currently holds a collection of 5531 barcode sequences from 2429 Korean species (Table 1) that were made using several genetic markers such as the mitochondrial cytochrome *c* oxidase subunit I (*COI*), nuclear internal transcribed spacer (*ITS*), or 18S ribosomal RNA for animals and the plastid *trnH-psbA* intergenic spacer or chloroplast genes (*matK* and *rbcL*) for plants. The KBOL system uses taxonomic structures and codon usage tables that are maintained by the taxonomy group at the National Center for Biotechnology Information (NCBI). Any user can search all of the barcode lists and sequences during an analysis; however, only registered users can freely download the barcode sequences (FASTA format).

### *Development of molecular identification system*

Statistical species models were constructed using the *hmmbuild* and *hmmcalibrate* programs from the HMMER v2.3.2 software (Eddy 1998). The species models and the *hmmfam* program were used to identify the species-specific patterns of unknown DNA barcode sequences. A three-fold cross-validation

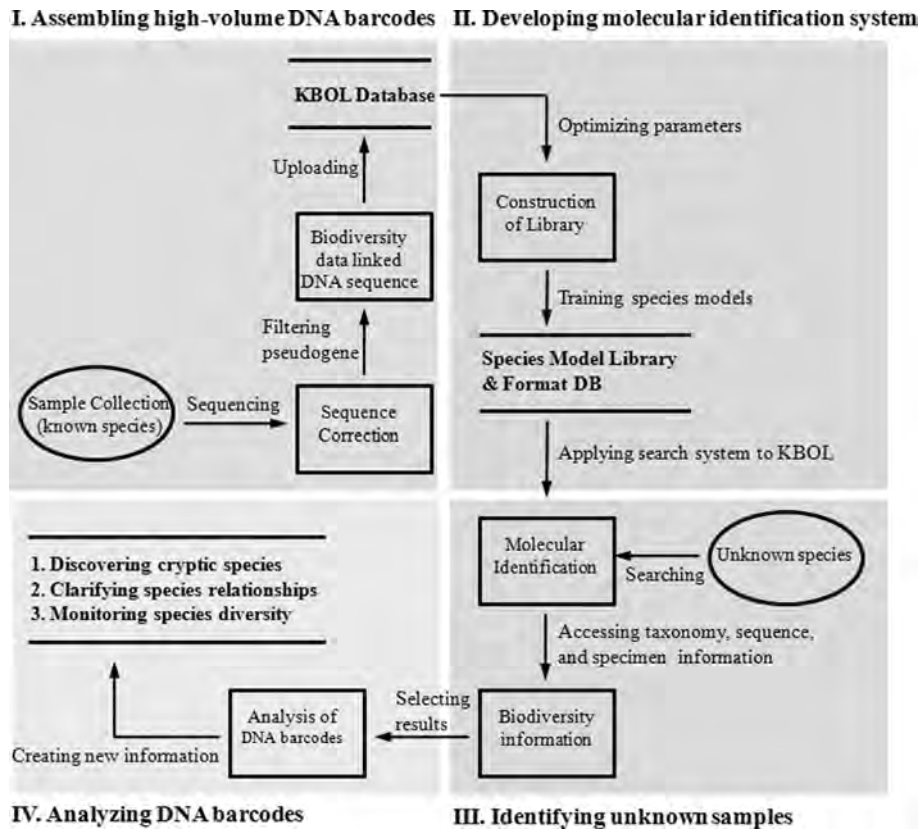


Figure 1. Workflow of the Korean Barcode of Life database system (KBOL) for assembling DNA barcodes, developing species identification system, identifying unknown samples, and analyzing DNA barcodes.

approach was adopted to evaluate the performance of the profile HMMs. The *blastall* program was used to execute BLAST (Altschul et al. 1990), which completed the rapid sequence alignment of unknown query sequences and sequences from reference barcode databases. The results are listed in ascending order according to the match score and E-value (expectation value). All of the steps for optimum condition and cutoff value were automatically processed by Perl scripts and controlled by cutoff values. The cutoff values indicate the standard deviations of the HMMER bit scores when constructing a profile HMM (Kim et al. 2010).

#### Development of web-based KBOL

The KBOL system was developed with open source software. PHP script, Java, Perl, and MySQL RDBMS 5.0 were used for most of the web pages and the main system. The web-based KBOL system operates on an Apache web-server using Linux. The easy-to-use and quick-loading interface ensures rapid

data loading even over low-bandwidth Internet connections.

## Results and discussion

### User interface

The user interface for the KBOL database system contains five items: the main menu bar, login, category, statistics, and links to related databases. An upload management system is provided on the Project page, with search systems provided on the Search page and Lineage page. The Project page shows integrated information consisting of five sections, including the project overview, geographical distribution, project statistics, member list, and entries. The Search page allows fast, reliable, and automated molecular identification using the HMM and BLAST algorithms (Figure 3 and Supplementary 10). The Lineage page presents specimen objects belonging to different taxa, allowing users to search through categories of interest (Figure 4). The simple KBOL interface permits a researcher to identify and access all of the biological

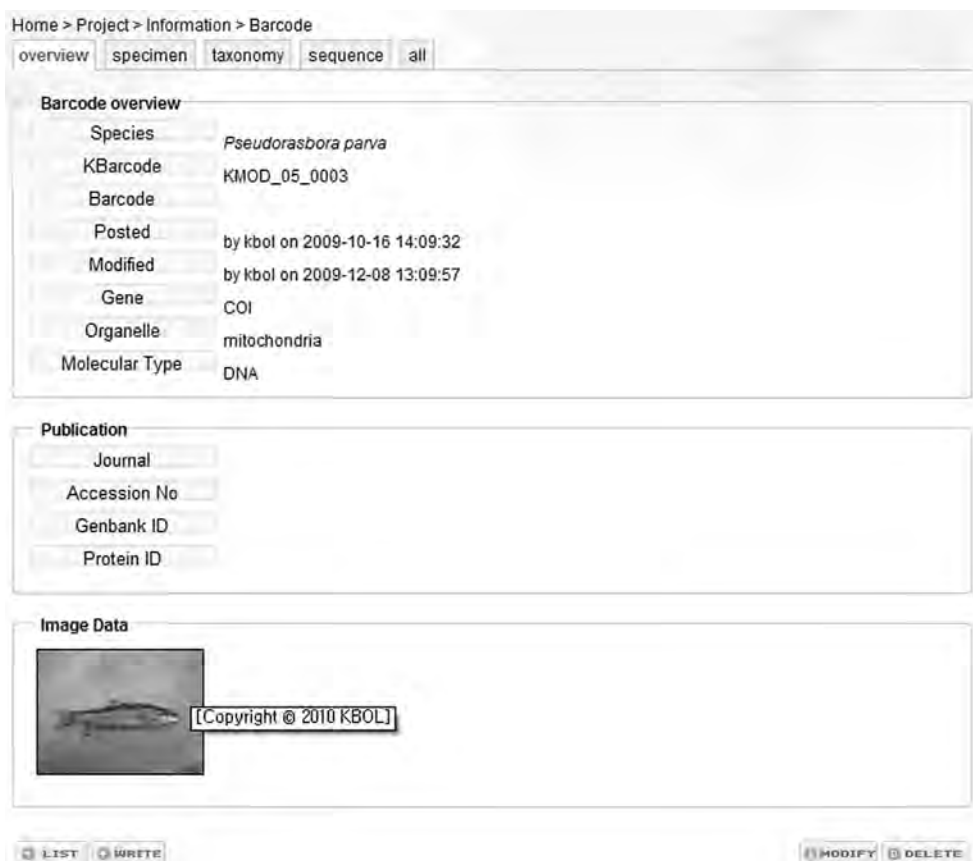


Figure 2. A view of biological information in each species entry, including DNA barcode sequence, specimen, taxonomy, image, and geographic data.

Table 1. Summary of the Korean DNA barcode resources.

Name of DNA barcoding projects	No. of sequences	No. of species
<b>Korean vertebrates (KVER)</b>		
Korean birds 2007 (KVER_01)	148	49
Korean mammals, amphibians, and reptiles 2007 (KVER_02)	131	31
Korean mammals, amphibians, and reptiles 2008 (KVER_03)	90	28
Korean birds 2008 (KVER_04)	141	45
Korean birds 2009 (KVER_05)	136	55
Korean mammals, amphibians, and reptiles 2009 (KVER_06)	201	51
Korean marine fish 2009 (KVER_07)	56	20
Total	903	279
<b>Korean insects (KINS)</b>		
Korean Hymenopteran 2007 (KINS_01)	84	34
Korean Hemiptera 2007 (KINS_02)	62	36
Korean Hymenopteran 2008 (KINS_03)	77	30
Korean Hemiptera 2008 (KINS_04)	76	31
Korean Lepidoptera 2007 (KINS_05)	112	112
Korean Lepidoptera 2008 (KINS_06)	75	70
Korean Diptera 2007 (KINS_07)	65	55
Korean Diptera 2008 (KINS_08)	75	59
Korean Lepidoptera 2009 (KINS_09)	91	88
Korean Hymenopteran 2009 (KINS_10)	84	37
Korean Hemiptera 2009 (KINS_11)	88	70
Korean Diptera 2009 (KINS_12)	88	69
Total	977	691

Table 1 (Continued)

Name of DNA barcoding projects	No. of sequences	No. of species
Korean model ecosystems (KMOD)		
Korean freshwater ecosystems 2007 (KMOD_01)	216	83
Korean freshwater ecosystems 2008 (KMOD_02)	391	107
Korean tidal flat ecosystems 2008 (KMOD_03)	155	50
Korean tidal flat ecosystems 2009 (KMOD_04)	91	37
Korean freshwater fish 2007-2009 (KMOD_05)	399	54
Korean freshwater ecosystems 2009 (KMOD_06)	194	82
Total	1446	413
Korean plant (KPLA)		
Korean vascular plants 2007 (KPLA_01)	478	249
Korean vascular plants 2008 (KPLA_02)	420	236
Korean vascular plants 2009 (KPLA_03)	311	158
Total	1209	643
Korean invertebrates (KINV)		
Korean sponge 2007 (KINV_01)	30	20
Korean Copepoda 2007 (KINV_02)	32	15
Korean ciliate 2007 (KINV_03)	31	16
Korean Bryozoa 2007 (KINV_04)	12	10
Korean sea squirt 2007 (KINV_05)	11	6
Korean jellyfish 2007 (KINV_06)	2	1
Korean Ctenophora 2007 (KINV_07)	2	1
Korean Hydra 2007 (KINV_08)	10	5
Korean Urochordata 2007 (KINV_09)	2	1
Korean sponge 2008 (KINV_10)	30	16
Korean Copepoda 2008 (KINV_11)	38	18
Korean ciliate 2008 (KINV_12)	30	15
Korean Bryozoa 2008 (KINV_13)	31	16
Korean jellyfish, Ctenophora, Hydra, Urochordata, sea squirt 2008 (KINV_17)	10	9
Korean coral 2008 (KINV_19)	22	7
Korean coral 2008 (KINV_20)	106	28
Korean Platyhelminthes 2007 (KINV_21)	24	8
Korean Platyhelminthes 2008 (KINV_22)	12	4
Korean Mollusca 2007 (KINV_23)	61	24
Korean Mollusca 2008 (KINV_24)	48	26
Korean Echinodermata 2007 (KINV_25)	40	20
Korean Echinodermata 2008 (KINV_26)	29	15
Korean sponge 2009 (KINV_27)	30	24
Korean coral 2009 (KINV_28)	178	13
Korean Bryozoa 2009 (KINV_30)	33	14
Korean Mollusca 2009 (KINV_31)	63	30
Korean Copepoda 2009 (KINV_32)	29	16
Korean Echinodermata 2009 (KINV_33)	19	8
Korean ciliate 2009 (KINV_34)	31	17
Total	996	403

resources using a particularly powerful but easy-to-use search interface.

### ***Molecular identification system***

The KBOL system provides various searches to explore a large quantity of biodiversity information, such as biological information searches, molecular identification searches, and keyword searches (Supplementary Figures 8–12). During the molecular identi-

fication process, an unknown DNA barcode sequence can be progressively assigned to reliable ranked hits by BLAST or to a known species by the profile HMMs, which statistically estimate species-specific patterns and yield clear statistical significance values and good prediction accuracy (Table 2); however, the profile HMMs have a limited barcode data pool. The statistical species models are automatically generated (see Methods) and opened in the Search page. The final results are directly linked to the relevant

**Korea Barcode Of Life**  
DNA barcode system for species identification and taxonomy

Home Project News Lineage Search Organization Feedback

**1. Species Identification System**

Database: KBOL

BLAST Search (Sequence or FASTA format) [Example]  
 HMM Search (FASTA format) [Example]

>Unknown\_Sequence  
 ATCATAAAGATATTGGCACCCTTATCTTATATTTGGTGCCTGGGCGCGATAGTGGGGACT  
 GCTTTAAGCCTCCTATCCGGGCTGAGCTTGCCAGCCCGGATCATTTCTAGGTGATGATCA  
 AATTTATAATGTCATCGTTACTGCCACGCGCTCGTAATAATTTCTTTATAGTAATGCATA  
 TTCTTATCGGCGGTTTCGAAATTBACTCGTGCCACTTATAATCGGGGACCCGACATGGCA  
 TTCCCTCGAATGAATAACATAAGCTTTTGGCTTCTCCACCATCATTTCTTCTGCTACTAGC  
 CTCTTCTGGTGTGAAGCTGGAGCCGGGACGGGATGAACAGTTTACCCACCACTTGCAGGCA  
 ATCTCGCCACCGAGGAGCATCCGTAGATCTTACAATTTTCTCTGCACTTAGCAGGTGTC  
 TCATCAATCTAGGAGCAATTAATTTAATTACTACTATCATTAAATAAAACCCCAAGCCAC  
 ATCCCAATATCAAACACCCCTATTGTATGGGCGCTTCTAGTAACAGCCGACTTCTGCTTQ

Filtering : cutoff | 10E-10 | results | 50 | hits

**Category**

- Vertebrate (KVER)
- Invertebrate (KINY)
- Insect (KINS)
- Plant (KPLA)
- Model organisms (KMOD)

**Rank**    **Genus**    **Score**    **E-value**

1	<i>Synthliboramphus</i> [Analysis]	756.1	2.7e-197
2	<i>Falco</i> [Analysis]	588.1	1.1e-180
3	<i>Scolopax</i> [Analysis]	585.3	7.8e-180
4	<i>Larus</i> [Analysis]	583.4	2.8e-179
5	<i>Egretta</i> [Analysis]	582.4	9.9e-178

**Home > Project > Information > Barcode**

overview specimen taxonomy sequence all

**Barcode overview**

Species: *Synthliboramphus antiquus*  
 KBarcode: KVER\_01\_00041  
 Barcode: [Sequence]  
 Posted: by bird on 2008-08-26 14:08:06  
 Modified: by bird on 2009-02-27 16:09:31  
 Gene: COI  
 Organelle: mitochondria  
 Molecular Type: DNA

>333 KVER\_01\_00042 *Synthliboramphus antiquus*  
 ATTGGCACCCCTATACCTAATCTTCGGGCGATGAGCTGGCATAGTGGCACCGCCCTAAGCCTACTCATCC  
 GCGCAGAACTAGGCAACCAAGGACCTCTAGGAGACGACCAATCTACAAGTAACTGTCACCGGCCA  
 CGCTTCGTAAATAATCTCTTTATAGTTATACCGATCATAATCGGTGGCTTGGAACTGACTGGTCCA  
 CTCATAATCGGTGCGCCTGACATAGCTTCCCGGTATGAACAACAATAAGCTTTGACTACTTCTCCCT  
 CCTTCCTTCTCTAGCCTCTCTACAGTAAAGCTGGTCTGGTACAGGGTGAACCTATATACCCCC  
 CTAGCTGGTAACTAGGCCACCGCGGCGCTCAGTAGACCTAGCAATCTTCTCTCCACTAGCAAGT  
 GTATCTCTATCTAGGCTATCACTCATCAACAGCCGATCAACATAAAACCCCGCCCTCTGCA  
 AATACCAAAACCCCTTATTTGTATGATCACTTATTACTGCGCTACTACTTACTTCACTTCACT  
 ACTGCTGCTGGCATTACTATGCTACTAAGACCGAACTAAACACAACTTTTGAACCACTGGG  
 GCGGTGATCCAGTACTATACCAACACCTTTTCTGATTTTTTGGTCATCCAGAAGTATATATCTAACT  
 TACCTGGATTCCGA

Figure 3. Molecular identification system using profile hidden Markov model (HMM) and BLAST. Users can identify a species from an unknown sequence (1–3) and then search the relevant biological information (4) or download sequence datasets for further analysis (5).

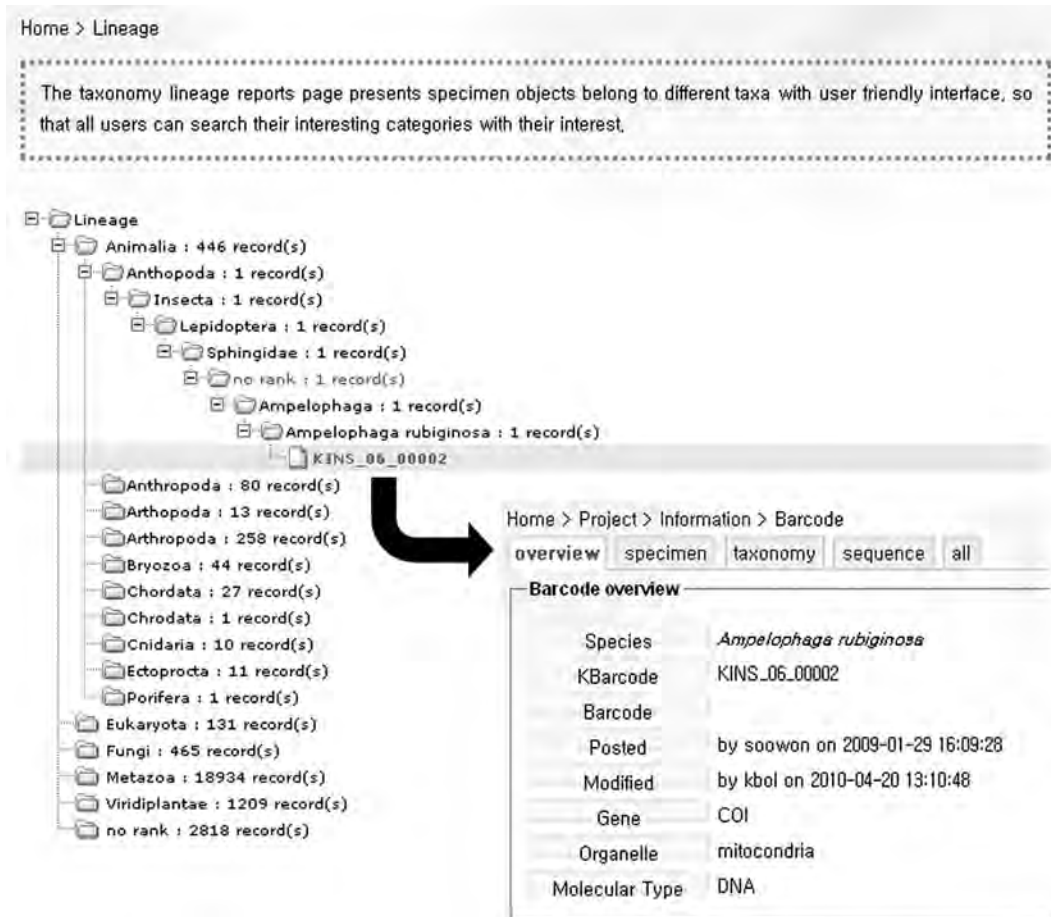


Figure 4. Accessing biological information via the taxonomy lineage search page of the menu bar.

biological information, including the collection date, specimen source, specimen images, DNA barcode sequences, trace files, specimen taxonomy, and geographical location (Supplementary Figures 1–7). The Java-based ATV (the successor to A Tree Viewer) package is used to visualize the species relationships from the predicted results or barcodes chosen by the user (Zmasek and Eddy 2001) (Figure 3). Furthermore, the KBOL system can directly link DNA barcodes and biological information in some species groups that were only initially identified via their

signature sequences, rather than their morphologies, especially where species diversity cannot be handled with conventional approaches (Markmann and Tautz 2005; Pfenninger et al. 2007) (Figure 5).

**Future development**

The future development of the KBOL system will include the utilization of global standard protocol formats and sophisticated knowledge-based processes. Utilizing a protocol based on the Darwin Core standard (Core 2004) allows the KBOL system to share information on biological diversity and inter-operate with related biodiversity databases, such as the marine arthropod depository bank of Korea (MADBK, <http://www.madbk.org>), the Korean Biodiversity Information Facility (KBIF, <http://www.kbif.re.kr>), and BOLD. As the volume of data continues to increase, well-organized sophisticated databases will help to increase our understanding of the enormous

Table 2. Number of DNA barcodes in each dataset and performance of the HMM library at species and genus levels.

	KBOL	BOLD	GenBank
No. of trained genera	984	300	739
No. of trained species	1135	431	1527
Accuracy at the species level (%)	92.72	91.45	90.32
Accuracy at the genus level (%)	96.53	97.20	97.31

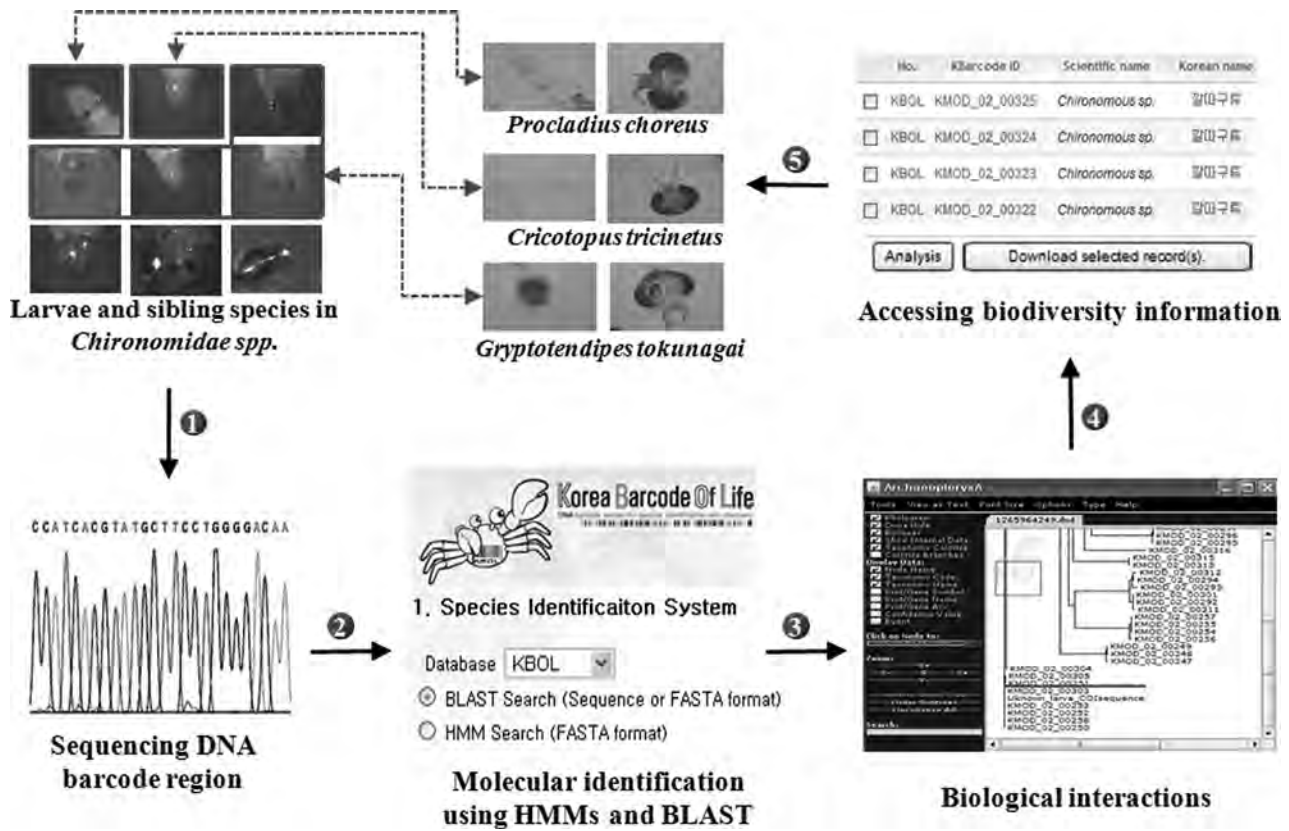


Figure 5. Example of reverse taxonomic research of *Chronomidae* spp. using the KBOL.

amount of information that has already been collected about the world’s biodiversity.

**Conclusions**

As a principal reference system for biodiversity research, the KBOL database system was developed to manage various types of biological data using sophisticated analysis tools and powerful search engines. All of the processes in the KBOL database system can be visualized using a convenient graphical user interface, and the DNA barcode data released to the public domain are freely available to all registered users. The web-based DNA barcoding system is essential for a descriptive understanding of species diversity, along with the conservation and management of major domestic organisms. In the future, the KBOL system will be upgraded to secure more biological barcode data, include bioinformatics analysis tools, and share biodiversity data with relevant international organizations.

**Acknowledgements**

This work was supported by a grant from the Eco-Technopia 21 Project, which is funded by the Ministry of Environment of Korea (2008-05002-0065-0).

**References**

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution’s misplaced witnesses. *Trends Ecol Evol.* 16(6):314–321.

Benson DA, Karsch-Mizrachi IK, Lipman DJ, Ostell J, Wheeler DL. 2008. GenBank. *Nucleic Acids Res.* 36 Suppl1:D25–D30.

Core D. 2004. Darwin Core Schema (version 1.3), a draft standard of the Taxonomic Database Working Group (TDWG).

Dasmahapatra KK, Mallet J. 2006. Taxonomy: DNA barcodes: recent successes and future prospects. *Heredity.* 97(4):254–255.

Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics.* 14(9):755–763.

Hajibabaei M, Dewaard JR, Ivanova NV, Ratnasingham S, Dooh RT, Kirk SL, Mackie PM, Hebert PDN. 2005. Critical factors for assembling a high volume of DNA barcodes. *Phil Trans R Soc B.* 360(1462):1959–1967.

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *TIG.* 23(4):167–172.

Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. *Syst Biol.* 54(5):852–859.

Hebert PDN, Ratnasingham S, de Waard JR. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond B.* 270 Suppl1:S96–S99.



- Krogh A, Brown M, Mian IS, Sjolander K, Haussler D. 1994. Hidden Markov models in computational biology. Applications to protein modeling. *J Mol Biol.* 235(5):1501–1531.
- Kim S, Eo HS, Koo H, Choi JK, Kim W. 2010. DNA barcode-based molecular identification system for fish species. *Mol Cells.* 30(6):507–512.
- Lim J, Kim SY, Kim S, Eo HS, Kim CB, Paek WK, Kim W, Bhak J. 2009. BioBarcode: a general DNA barcoding database and server platform for Asian biodiversity resources. *BMC Genomics.* 10 Suppl 3:S8.
- Markmann M, Tautz D. 2005. Reverse taxonomy: an approach towards determining the diversity of meio-benthic organisms based on ribosomal RNA signature sequences. *Phil Trans R Soc B.* 360(1462):1917–1924.
- Park HC, Lee YM, Choi WY. 2008. A study on technical development of DNA barcoding analyses. *Anim Cells Syst.* 12 Suppl:G601–G700.
- Pfenninger M, Nowak C, Kley C, Steinke D, Streit B. 2007. Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic Chironomus (Diptera) species. *Mol Ecol.* 16(9):1957–1968.
- Radulovici AE, Archambault P, Dufresne F. 2010. DNA Barcodes for marine biodiversity: moving fast forward? *Diversity.* 2(4):450–472.
- Ratnasingham S, Hebert PDN. 2007. BOLD: the Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol Ecol Notes.* 7(3):355–364.
- Rubinoff D. 2006. Utility of mitochondrial DNA barcodes in species conservation. *Conserv Biol.* 20(4):1026–1033.
- Song H, Buhay JE, Whiting MF, Crandall KA. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *PNAS.* 105(36):13486–13491.
- Zmasek CM, Eddy SR. 2001. ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics.* 17(4):383–384.