



# HOKKAIDO UNIVERSITY

|                  |                                                                                                                                                                                                                                                     |
|------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title            | Biological Classification and Identification System (BioCIS)                                                                                                                                                                                        |
| Author(s)        | Toda, Masanori J.; Matsushita, Kotaro; Mawatari, Shunsuke F.                                                                                                                                                                                        |
| Description      | International Symposium on "Dawn of a New Natural History - Integration of Geoscience and Biodiversity Studies". 5-6 March 2004. Sapporo, Japan.                                                                                                    |
| Relation         | Edited by Shunsuke F. Mawatari, Hisatake Okada.                                                                                                                                                                                                     |
| Citation         | Neo-Science of Natural History: Integration of Geoscience and Biodiversity Studies : Proceedings of International Symposium on "Dawn of a New Natural History - Integration of Geoscience and Biodiversity Studies" March 5-6, 2004, Sapporo, 41-44 |
| Issue Date       | 2004                                                                                                                                                                                                                                                |
| Doc URL          | <a href="https://hdl.handle.net/2115/38489">https://hdl.handle.net/2115/38489</a>                                                                                                                                                                   |
| Type             | conference paper                                                                                                                                                                                                                                    |
| File Information | 41-44-neo-scinece.pdf                                                                                                                                                                                                                               |



# Biological Classification and Identification System (BioCIS)

Masanori J. Toda<sup>1</sup>, Kotaro Matsushita<sup>1</sup> and Shunsuke F. Mawatari<sup>2</sup>

<sup>1</sup>*Institute of Low Temperature Science, Hokkaido University, Sapporo 060-0819, Japan*

<sup>2</sup>*Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan*

## ABSTRACT

The Biological Classification and Identification System (BioCIS) as a basic, general-purpose database system, though still under construction, will provide taxonomists with a very useful tool to construct multiple-entry keys based on character matrices. In turn, such multiple-entry keys give various end-users, non-specialists in taxonomy, a great deal of freedom and ease in species identification. The system is designed to be applicable to any group of organisms and to grow through interactions between taxonomists and end-users, in the prospect of developing a hierarchical system of keys for all taxonomic levels of organisms on earth, as a “hub” database system for biological classification and identification.

**Keywords:** Biodiversity, Character matrix, Database, Multiple-entry key, Species identification

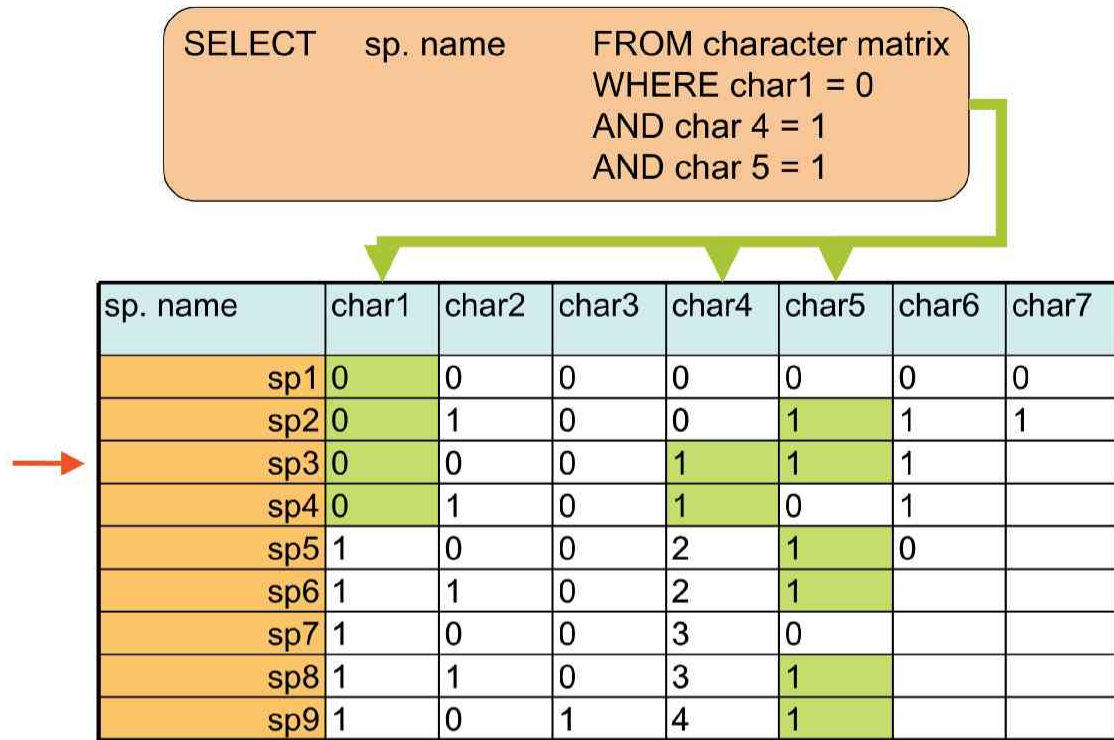
## INTRODUCTION

Accurate identification of biological species is currently in high demand, not only by biodiversity researchers but also by natural-resource managers implementing applied measures for assessing, conserving, and using biodiversity. However, there are too few trained taxonomists to meet this high demand. The BioCIS database is a basic, general-purpose system designed to help taxonomists, experts in biological classification, construct useful tools with which a variety of end-users can identify biological species. The Biological Classification and Identification System (BioCIS) is intended to be applicable to any group of organisms. It is expected that species-identification systems for many groups of organisms will, in the future, be constructed as components of this general system.

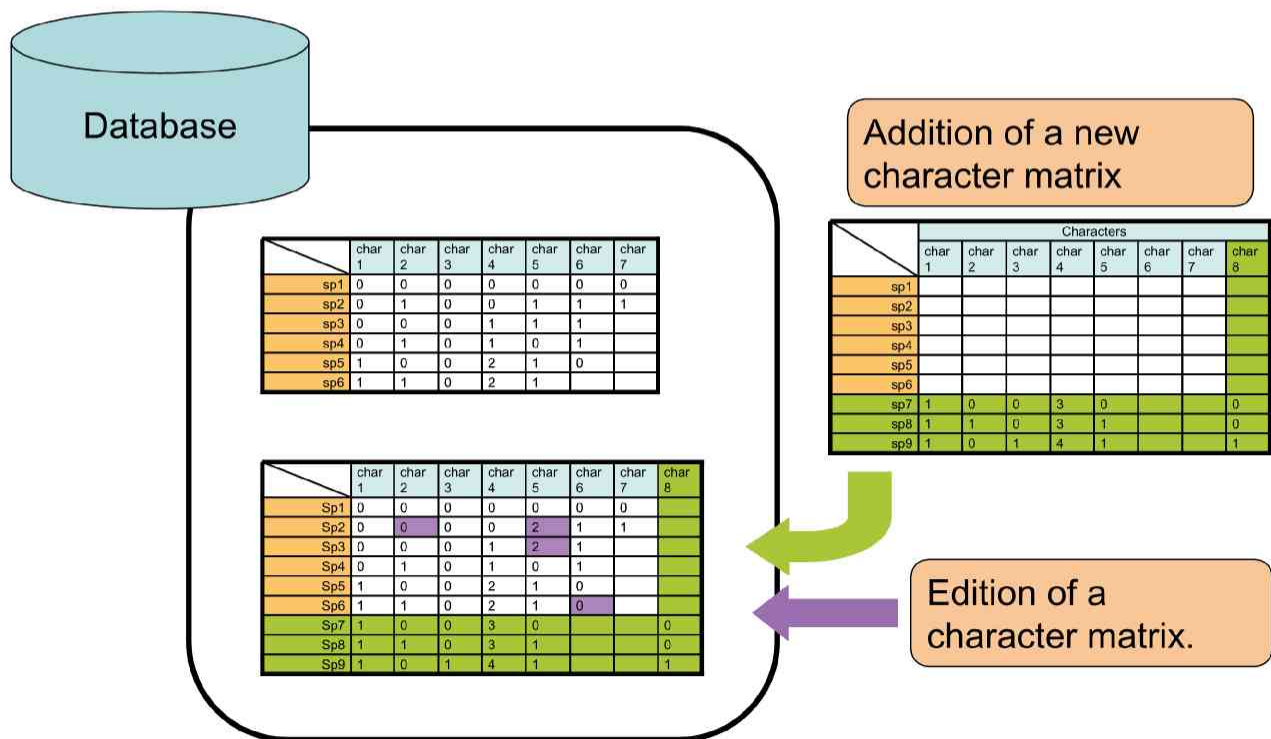
## WHAT IS BioCIS?

### For Taxonomists

BioCIS is designed so that any taxonomist, even without any knowledge of information technology, can easily construct a multi-taxa reference system based on a character matrix (Fig. 1). Provided a character matrix and pictures, if any, this system automatically constructs a multiple-entry key for identifying species or higher taxa. The system is also designed to grow, in response to interactions between taxonomists and end-users, to cover increasing numbers of taxa and broader geographic areas (Fig. 2). Since adding a new character matrix to the system is easy, a taxonomist could, for example, start with a small database covering a single genus at a single locality, and subsequently include additional genera from other localities. Also, since it is possible to edit a character matrix even after a basic character matrix has been constructed, the system can update the matrix according to new findings and/or revision of taxa. Furthermore, the system is



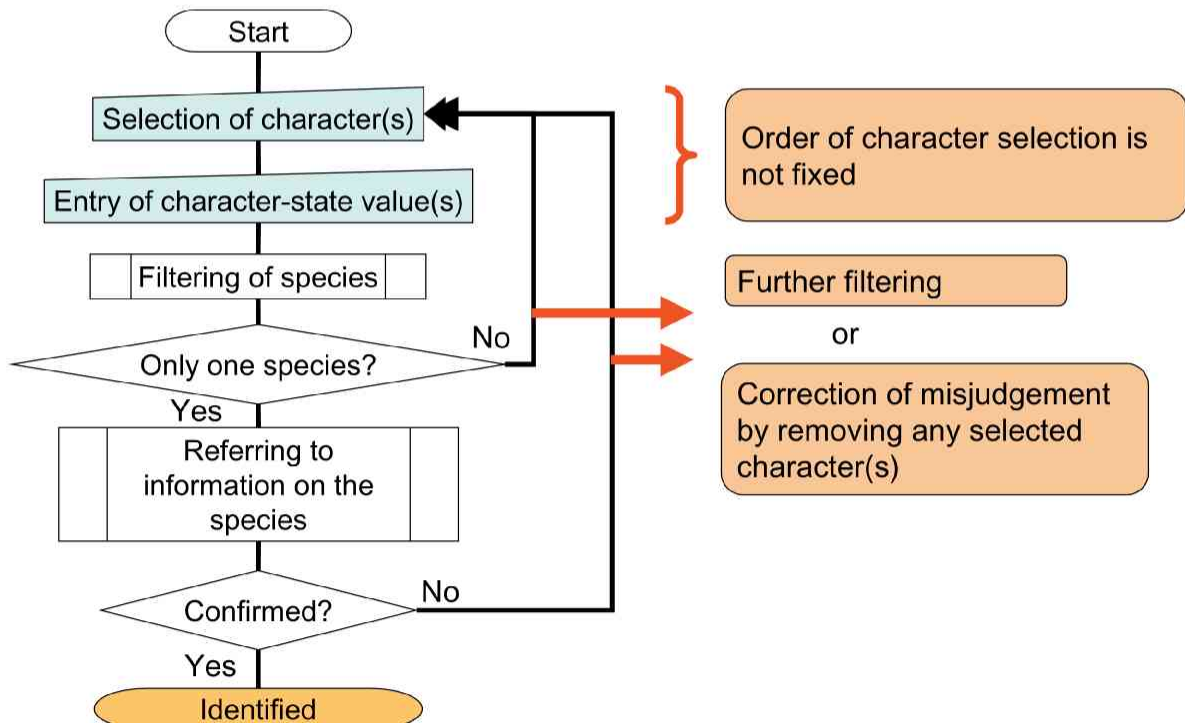
**Fig. 1** Multiple-entry key system based on a character matrix. The system extracts species for which character states match user's entry for selected character set. For example, species 3 is extracted when user chooses character state 0 for character 1, 1 for 4 and 1 for 5.



**Fig. 2** Expansion of a character matrix. This system is designed to grow, allowing taxonomists to easily add a new matrix for additional taxa or locality to the existing matrix and to change individual cell values or the character state definitions.

| sp. name | char1 | char2 | char3 | char4 |
|----------|-------|-------|-------|-------|
| sp1      | 0     | 0     | 0     | 0     |
| sp2      | 0     | 1     | 1     | 1     |
| sp3      | 0     | 0     | 1     | 1,2   |
| sp4      | 0     | 1     | 2     | 1,2   |
| sp5      | 1     | 1     |       | 2     |
| sp6      | 1     | 0     |       | 2     |

**Fig. 3** Multiple and no entry of character-state values. If the state of character 4, for example, is subtle or difficult to judge precisely for species 3 and species 4, multiple values are to be input (red cells). If the state of character 3 is unknown or unexamined for species 5 and species 6, the corresponding cells can be left blank (blue cells). Species 3 and species 4 are not excluded when users choose either character state 1 or 2 for char 4. Species 5 and species 6 are not excluded when users choose either character state 0, 1 or 2 for char 3.



**Fig. 4** Flow chart of identification using BioCIS. As the order of character selection is not fixed, users can start from any character set. The system filters species matching user's entry of character-state values for the character set chosen. If there remain more than one species, users can continue to select another character set for further filtering. When only one species is eventually selected, the user can refer to information on the species to confirm the identification. If the user loses him-/herself in the searching process or cannot get the correct answer, he/she can return to any step of filtering by removing any selected character(s).

designed to allow input of multiple character-state values in a single cell of the matrix (Fig. 3). If the state of a particular character is subtle or difficult to judge precisely in some species, or if multiple states of a character are observed within a supra-specific taxon, multiple values like (0, 1) or (0, 1, 2) etc. are to be input to the corresponding cells in the matrix. This copes, to some extent, with a difficulty in applying a multiple-entry key to supra-specific taxonomic levels [1]. In addition, the system allows no entry (Fig. 3) of character state for a part of the taxa and characters. This enables taxonomists to develop the character matrix gradually by inputting character states step by step and adding new taxa easily. For example, closely related species can be usually distinguished from each other by only a few, very particular characters. When adding such species to the matrix, examination of the relevant, diagnostic characters for all the other unrelated species is almost meaningless and time-consuming for taxonomists. BioCIS allows taxonomists to leave those cells blank.

### For Users

The system design gives users a great deal of freedom in species identification. Traditional dichotomous keys allow no freedom, and it is usually not easy for beginners to identify species correctly. Matrix-based multiple-entry keys allow users to start from any character set and to follow any route. In the process of identification, users are allowed to misjudge character states to some extent, especially those subtle or difficult for beginners to judge correctly, because of the multiple state-values input for those characters in the matrix. Furthermore, when users lose themselves, they can go back to look for the correct route by removing any one of the characters so far selected, the list of which is always presented on the window. Thus, the system leads them eventually to the correct answer (Fig. 4). Finally, it provides information (images, references and so on) with which users can themselves confirm the correctness of their identifications.

### Prospects

Although the system is still under construction, the features of multiple-entry keys to be created in this system will allow convenient, easy access to identifications by various non-specialists in taxonomy, e.g., biodiversity researchers, technicians, students, environmental managers, and the general public wishing to learn more about nature and the environment. Furthermore, this system applicable to any taxonomic levels and allowing any-sized keys to coexist can develop to a hierarchical system of keys covering all organisms on earth, as a "hub" database system for biological classification and identification, being linked with all taxonomic databases publicized in the Internet.

### ACKNOWLEDGMENTS

The Biological Classification and Identification System was constructed with financial support from the Hokkaido University 21<sup>st</sup> Century Center of Excellence (COE) Program for the "Neo-Science of Natural History," and by the Special Coordination Fund for Promoting Science and Technology of the Japan Ministry of Education, Culture, Sports, Science and Technology. We thank Dr. Matthew H. Dick (Hokkaido University) for revision of the manuscript.

### REFERENCES

1. White, I., 1997. CABIKEY & TAXAKEY, In: S. Gallagher (ed.), *Proceedings of the Inaugural Meeting of the BioNET-INTERNATIONAL Group for Computer-Aided Taxonomy (BIGCAT)*, The Technical Secretariat (TECSEC) of BioNET-INTERNATIONAL, Surrey, UK, 59.

### Relevant URLs

BioCIS:

<http://biokey.museum.hokudai.ac.jp:8080/Classification/index.html>

BIGCAT:

<http://www.bionet-intl.org/html/structure/bigcat/bigcat.htm>