METHODS & TECHNIQUES

ActKey: a Web-based interactive identification key program

Anthony R. Brach[1] & Hong Song[2]

[1] *Missouri Botanical Garden, c/o Harvard University Herbaria, 22 Divinity Avenue, Cambridge, Massachusetts 02138, U.S.A. brach@oeb.harvard.edu (author for correspondence)*
[2] *Saint Louis University, DuBourg Hall, Room 345, 221 North Grand Boulevard, St. Louis, Missouri 63103 U.S.A. hongsong2k@yahoo.com*

Interactive keys are useful multiple-entry identification tools for science. A web-based program called *ActKey* (URL: http://flora.huh.harvard.edu:8080/actkey/) was developed to enable online, ready-access to interactive identification keys. A number of keys have been imported into a centralized, relational database, and are now available for the floras of China, North America, Madagascar, Borneo, and the world. Plants from more than 40 of the largest genera of China can be identified using ActKey. Datasets containing taxa and their characters can be imported from a variety of database formats and updated by taxonomic specialists and editors.

KEYWORDS: ActKey, computer-assisted identification, Flora of China, interactive key.

INTRODUCTION

Interactive keys are excellent identification tools for systematic botany, inventory, and conservation (Dallwitz, 1980; Jarvie & Stevens, 1998; Dallwitz & al., 2000; Heidorn, 2001). In printed publications, indented dichotomous keys are the most common forms of identification keys, while multiple-entry and bracketed keys are used less often. In both dichotomous and bracketed keys, the identification (ID) process must follow a predefined path, at each step asking: does the plant have traits "A" or "B"? Each lead of a couplet (e.g., 1a, 1b) provides contrasting, diagnostic characters. If the diagnostic character used in the key is unavailable or obscure, e.g., in flowering, fruiting, or sterile specimens, it becomes difficult, if not impossible, to use this type of key.

Taxonomic keys could benefit from "decision trees" from rule-based "expert systems" of computer science (see Pankhurst, 1991; but also: Dallwitz, 1992). Recently, artificial neural networks have been applied as tools for identification (Clark, 2003). These offer potential advantages over other types of keys when the data are limited and the species are difficult to differentiate.

An interactive key provides multiple entry points that allow the user to choose any characters available on the plant at hand to arrive at a name. The identification process involves using available or easily observed characters, and then selecting the appropriate character state or numerical value (i.e., from a measurement or count). Interactive keys are thus more suitable for the identification of unknown plants.

Interactive identification keys in DELTA-format (Description Language for Taxonomy: Dallwitz & al., 1993–) have been developed for several taxa (e.g., Watson & Dallwitz, 1981; Loizeau, 1994; Aiken & al., 1998; Coles & al., 1998; Chen & Kuoh, 2000a, b; see URL: http://www.delta-intkey.com/ for an extensive list of available INTKEY datasets); however, their use requires the installation of the DELTA INTKEY application on a personal computer (PC). Additional programs for building interactive keys (e.g., LucID, URL: http://www.lucidcentral.org/; Meka, URL: http://ucjeps.berkeley.edu/meacham/meka/index03.html; Navikey, URL: http://www.huh.harvard.edu/databases/legacy/navikey/; PollyClave, URL: http://prod.library.utoronto.ca:8090/polyclave/; and XID, URL: http://www.xidservices.com/) have been catalogued and reviewed elsewhere (see Dallwitz, 1996, 2000). Early web-based interactive keys (e.g., Navikey, see Dallwitz & al., 2002) often required helper applications, plug-ins, Java applets, and web forms, resulting in lengthy download times from the server to the client computer.

To make the electronic version of the *Flora of China* (Wu & Raven, 1994–, see Brach & Song, in press), a powerful tool for identifying the plants of China, we developed interactive keys for a number of large genera. The program was developed by Hong Song. It enables a user to access identification keys readily and to arrive at identifications more quickly than by using a traditional key. The scope of the project has since expanded beyond the *Flora of China* to provide a centralized database where specialists can share their taxonomic keys with the international community.

ActKey is a Java-based program that uses MySQL as the database server. The program and database reside on servers in the Harvard University Herbaria and Missouri

Botanical Garden. The client (web browser) sends a request; then, the server processes it, and sends the results back. ActKey provides a centralized database of interactive keys, and a browser interface (split frames) to character attributes (state-value statistics, notes, images) and remaining taxa (descriptions, images, dynamic links), and translations of character lists in available languages.

Selected keys have been prepared from the floras of China (also see URL: http://www.efloras.org/), North America, Madagascar, Borneo, and the world (ActKey URL: http://flora.huh.harvard.edu:8080/actkey/).

USING ACTKEY

Each ActKey is divided into two frames (see Fig. 1): character-attributes and taxa remaining (once character selection begins). For best results, we recommend using the latest version of a web browser (e.g., Mozilla Firefox (c), Microsoft Internet Explorer (c)). An *online help* feature assists the user with basic operational instructions. Clicking on a character in the left frame opens a window (see Fig. 2) to provide a list of character attributes (i.e., states) followed by the number of taxa having that particular state, character statistics (min., max.) for counts and measurements, and character notes (definitions and images). The client (web browser) expands and collapses the character tree. Clicking on a plus sign (+) expands the character "tree" to display the available states, and a minus sign (-) collapses the character "tree". Clicking on a checkbox selects a state. To address uncertainty, a checkbox is provided to include (i.e., retain) taxa with unknown character states (as the set default).

Taxa are displayed based on the characters selected, with at least one matched state for each selected character. After entering one or more state(s) or numeric value(s) for any character(s) in the left frame and then clicking on the search button at the top or bottom, a list of taxa with matching criteria will be displayed in the right frame. Clicking on a taxon name in the right frame opens a new window (see Fig. 3) with a description of the taxon and dynamic links to related data and images.

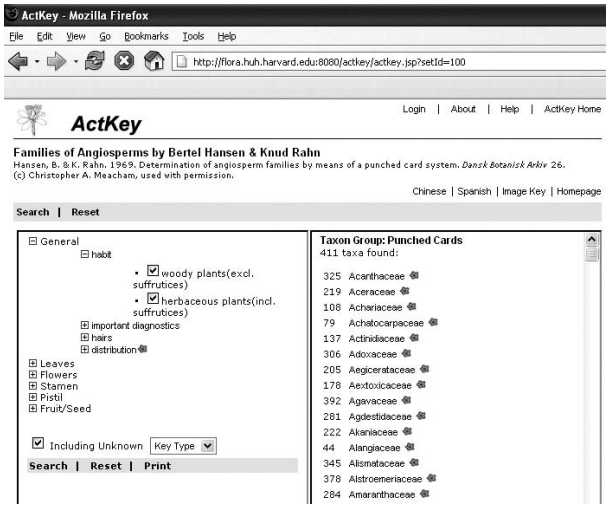

Fig. 1. ActKey screen for "Families of angiosperms" (based on Hansen & Rahn, 1969).

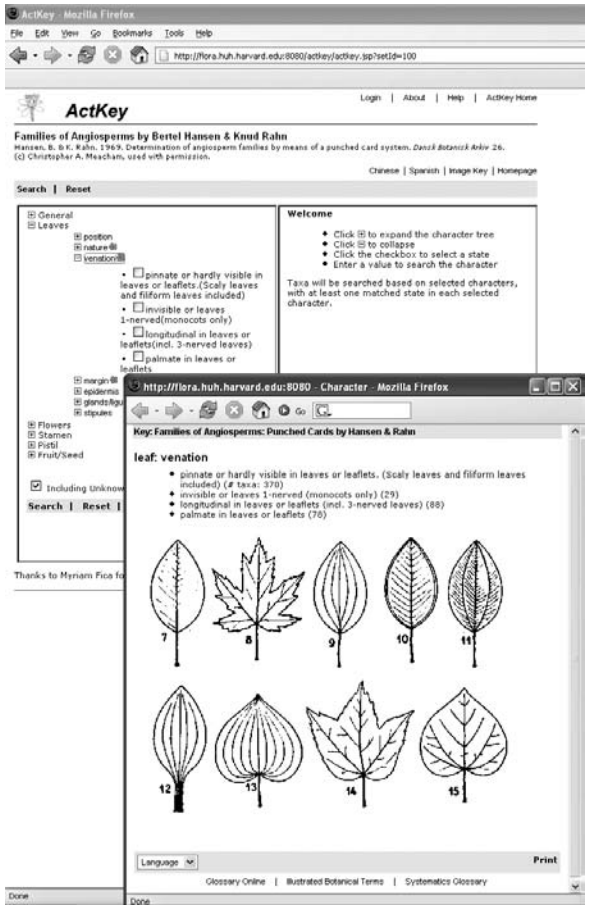

Fig. 2. ActKey screen for "Families of angiosperms" with expanded list of leaf venation character states, plus second popup window enumerating number of taxa in each state and providing illustrations.

In addition to English, the character list can be displayed in other languages, such as Chinese and Spanish (see Table 1). The character list can also be sorted according to the order of *best* characters (e.g., as in Figs. 4 and 5) for efficiency of identification as in DELTA INTKEY or according to a specialist's recommendations (currently static and optimal only at the start of the key since the optimal order differs at later steps).

ACTKEY HOME PAGE

The ActKey home page (URL: http://flora.huh.harvard.edu:8080/actkey/) provides access to several taxo-

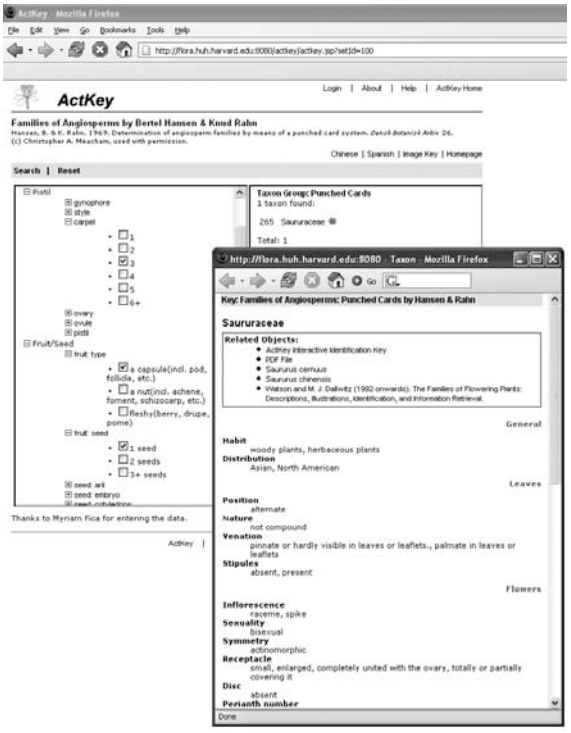

Fig. 3. ActKey screen for "Families of angiosperms" denoting possible identification as within the family Saururaceae, plus second popup window summarizing character attributes and providing dynamic links to related illustrations and treatments.

nomic keys within the database, particularly to large- to medium-sized genera of China (with at least 50 species) where treatments have been prepared, plus keys to families and floras including:

- *Angiosperm Families* by Bertel Hansen & Knud Rahn (1969; see Figures 1–3, also in Chinese and Spanish) (also see URL: http://ucjeps.berkeley.edu/meacham/meka/index03.html).

- *Families of Dicotyledons of the Western Hemisphere South of the United States* by Donald R. Simpson & David Janos (1974).

- *Genera of Brassicaceae of the World* by Ihsan A. Al-Shehbaz (also in Chinese, URL: http://flora.huh.harvard.edu:8080/actkey/actkey.jsp?setId=2051).

- *Generic Tree Flora of Madagascar* by George E. Schatz (2001) (also see URL: http://www.mobot.org/MOBOT/Madagasc/welcome.html).

- *Salix* L. (Salicaceae) of North America by George W. Argus (and in Chinese; also homepage at URL: http://aknhp.uaa.alaska.edu/willow/).

- *Trees and Shrubs of Borneo* by James K. Jarvie & Ermayanti. (1995–1996, also in Indonesian) (URL: http://www.phylodiversity.net/borneo/).

- *Trilliaceae* (*Trillium* L. and *Paris* L.) *of the World* by Susan B. Farmer (URL: http://flora.huh.harvard.edu:8080/actkey/actkey.jsp?setId=3001).

For an ActKey example from the *Flora of China*, the genus *Cotoneaster* Medik. (Rosaceae subfam. Maloideae) is represented by 59 species (Lu & Brach, 2003). By selecting inflorescences with 3 flowers, leaf blades ca. 3 × 1.5 cm (i.e., 2:1 length:width ratio) and petioles ca. 5 mm long, this narrows the number of species and varieties with that set of characteristics to six (see Fig. 4). The available taxon images then can be browsed, or further character values and/or states can be entered. From here, if "red" is selected for petal color, three species remain (*C. acutifolius* Turcz., *C. melanocarpus* Lodd., and *C. tenuipes* Rehder & E. H. Wilson). Finally, if the value of 10 is entered for number of stamens, or the state of the abaxial surface of the leaf blade is selected as villous, the plant is tentatively identified as *Cotoneaster acutifolius* (Fig. 5).

Table 1. Example of language display options for leaf margin (based on Hansen & Rahn, 1969).

English: margin
entire (without any lobes or teeth) (# taxa: 358)
lobed or divided (118)
dentate (197)
Spanish: Margen
Hojas o folíolos entero (sin ningún lóbulo o diente) (# taxa: 358)
Hojas o folíolos lobulados o divididos (118)
Hojas o folíolos dentados (197)
Chinese: 叶缘
叶或小叶全缘 (不具裂片或齿) (# taxa: 358)
叶或小叶浅裂或分裂 (118)
叶或小叶具齿 (197)

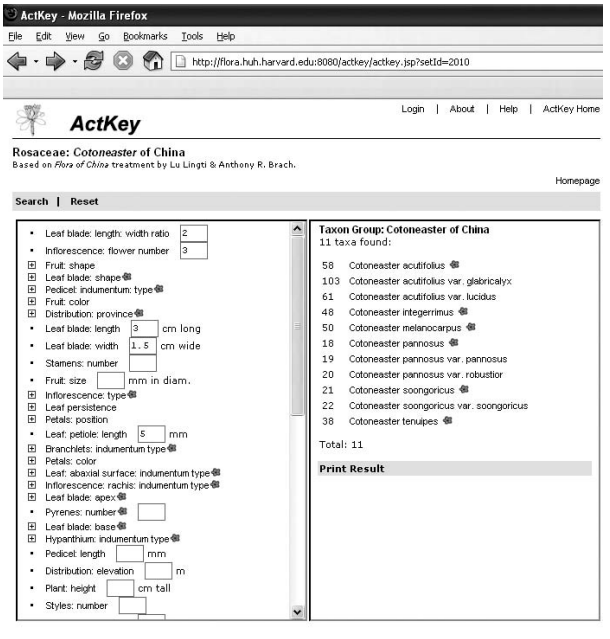

Fig. 4. Example of ActKey identification of *Cotoneaster*.

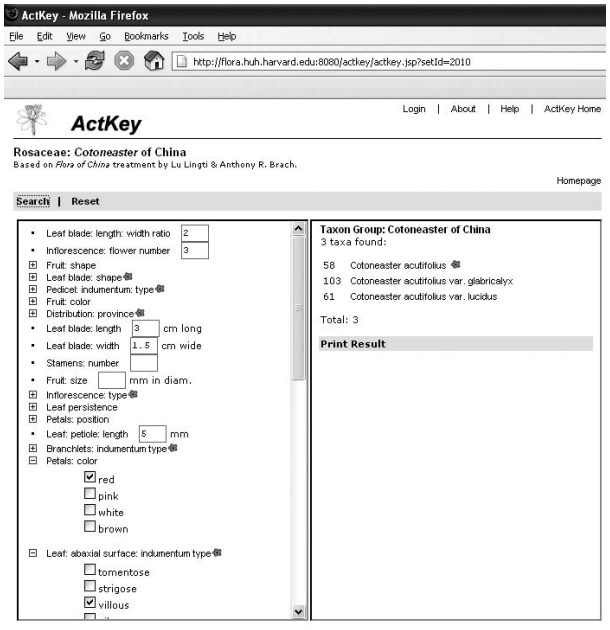

Fig. 5. Example of ActKey identification of *Cotoneaster acutifolius*.

CREATING A NEW DATASET FOR ACTKEY

Data have been successfully imported into ActKey from a variety of database and spreadsheet formats including DELTA and Microsoft Excel. The DELTA Editor and associated DELTA programs including INTKEY can be found online (URL: http://delta-intkey.com/). The following primary files (or their equivalents) are necessary to create a new ActKey Dataset: CHARS (character list), ITEMS (taxa with character states scored), SPECS (specifications about the above data), and INTKEY.INK (or an equivalent list of character headings or groupings of common characters, e.g., all leaf characters). It is better to input data while working with specimens; but, if it is necessary to work from descriptions, the character list, including characters and their attributes (i.e., states, values), can be developed (e.g., with the DELTA Editor) as one reads through the descriptions for the taxon group, and the states and values can be filled-in (scored) for each taxon. Often, the length:width ratio can provide a useful "best" character indicating relative shape particularly of the leaf blade (see Radford & al., 1976; Jensen, 2003). We have found that a split-screen with both a spreadsheet and taxonomic treatment open simultaneously provides a good, working, desktop environment. Character notes (definitions and images), and taxon images (illustrations, photos, type images) can be added to the ActKey database. After keys have been imported into the centralized database, authors and editors (with permissions) can update the data.

Taxonomists are welcome to share their datasets (e.g., DELTA, Microsoft Excel, Microsoft Access formats) via this Web-based system. Persons interested in translating character sets into another language may wish to contact us.

Additionally, eFloras (www.efloras.org, see Brach & Song, in press) provides an online tool to build interactive keys. Authorized users can create and edit character sets, build descriptions of taxa using characters, and create interactive searches of taxa using the same basic principles as in ActKey (e.g., URL: http://www.efloras.org/flora_page.aspx?flora_id=1001). The same technique is being used for a new project based on a fixed dataset (*A Catalogue of the Vascular Plants of Madagascar*, URL: http://www.efloras.org/key_page.aspx?set_id=10001&flora_id=12).

DISCUSSION

The centralized, relational structure of the ActKey database offers new possibilities and challenges to facilitate identification of plants and other organisms. Characters and their corresponding states can be dynamically linked to online glossaries and character images (e.g., Radford & al., 1976, URL: http://www.ibib lio.org/botnet/glossary/; Rossi-Wilcox, 2001, URL: http://flora.huh.harvard.edu/Glossary/). Similarly, names of taxa can be dynamically linked to online checklists, floras, and images.

With the centralized database structure, character state attributes, and numerical measurements and counts can be summarized statistically. Similarly, web access can be summarized statistically, e.g., to learn which keys, and character notes and images are frequently accessed, to better improve identification tools for those taxonomic groups.

Future ActKey use and testing with other interactive key programs will indicate its effectiveness and necessary improvements. Tools are being developed to export DELTA-format files to enable users to further explore the data with the DELTA INTKEY suite of programs. ActKey performs best using complete data matrices; however, floristic treatments sometimes lack parallel data for all taxa. Thus, the default "include unknowns" should be retained throughout an identification.

Because of some instances of instability of the Java server, and the greater potential functionality of ActKey within another programming interface (i.e., additional features possible with ASP.Net and C#) as evidenced by a similar move of the earlier "Flora Accounts Online" to eFloras (see Brach & Hong, in press), we are considering moving the application and the central database to eFloras.org where users can import and update their data in ActKey.

For the future, it is worthwhile to consider that a key, in a sense, is a collection of descriptions, wherein each lead is a mini-description, especially when combined with the higher levels of the key(s) and a classification system. Combining the pertinent leads of a key to infraspecies with the key to species, and higher ranks (genus, family) forms a hierarchy of character data (i.e., the description) about the taxon. Information about the diagnostic characters within each corresponding lead should be most helpful for identification because precision in identification should increase as the final lead for identifying a taxon is approached. Thus, besides an improvement in manageability, databased keys offer new options for usability and searches. Since keys can be considered hierarchical when the higher (or lower) keys are considered, it should be possible to link the keys to higher and lower taxa in an ontological manner.

Interactive keys provide much-needed identification tools for science. The ActKey program provides freely available web-based interactive keys to researchers and students of botany worldwide. Keys have been imported into a centralized, relational database and are available for the floras of China, North America, Madagascar, Borneo, and the world. Datasets containing taxa and their characters can be imported from a variety of database formats and updated by specialists and database editors. As more data become available, further ActKey applications will be developed.

ACKNOWLEDGEMENTS

We are thankful to all who contributed to the development of ActKey: Ihsan Al-Shehbaz (MO), George W. Argus (CAN), David E. Boufford (A), Susan B. Farmer (TENN), Myriam Fica (MO), Chris Freeland (MO), Mike Gilbert (MO at K), Porter P. Lowry II (MO at P), Robert Magill (MO), Chuck Miller (MO), Kristin Pierce (MO), Armand Randrianasolo (MO), George Schatz (MO), Peter F. Stevens (MO), Carmen Ulloa Ulloa (MO), Mark F. Watson (E), Xiangying Wen (MO), Yong Yang (MO), James L. Zarucchi (MO), and Guanghua Zhu (MO). Thanks to Chris Meacham (UC) for permission to use Hansen & Rahn's data. We thank Mike Dallwitz (Giralang, Australia) and two anonymous reviewers for helpful review comments. We are grateful to Anne Marie Countie (GH), systems administrator, for her support of the servers at the Harvard University Herbaria. Support of the U.S. National Science Foundation (BSR-8906215, BSR-9201378, DEB 9626806, DEB 0072682, and DBI 0343439) is gratefully acknowledged.

LITERATURE CITED

Aiken, S. G., Dallwitz, M. J., McJannet, C. L., Gillespie, L. J. & Consaul, L. L. 1998. Saxifragaceae of the Canadian Arctic Archipelago: a contribution to a DELTA database for interactive identification and illustrated information retrieval. *Canad. J. Bot.* 76: 2020–2036.

Brach, A. R. & Song, H. In press. eFloras: New directions for online floras exemplified by the Flora of China Project. *Taxon* 55.

Chen, C. H. & Kuoh, C. S. 2000a. The genus *Poa* L. (Poaceae) in Taiwan: a DELTA database for generating key and descriptions. *Taiwania* 45: 147–157.

Chen, C. H. & Kuoh, C. S. 2000b. The genus *Bromus* L. (Poaceae) in Taiwan: a DELTA database for generating key and descriptions. *Taiwania* 45: 311–322.

Clark, J. Y. 2003. Artificial neural networks for species identification by taxonomists. *Biosystems* 72: 131–147.

Coles, S., Maxted, N. & Van Der Maesen, L. J. G. 1998. Identification aids for *Cicer* (Leguminosae, Cicereae) taxa. *Edinb. J. Bot.* 55: 243–265.

Dallwitz, M. J. 1980. A general system for coding taxonomic information. *Taxon* 29: 41–46.

Dallwitz, M. J. 1992. A comparison of matrix-based taxonomic identification systems with rule-based systems. Pp. 21–58 in: Xiong, F. L. (ed.), *Proceedings of IFAC Workshop on Expert Systems in Agriculture.* International Academic Publishers, Beijing. URL: http://delta-intkey.com/

Dallwitz, M. J. 1996 onwards. *Programs for Interactive Identification and Information Retrieval.* URL: http://delta-intkey.com/

Dallwitz, M. J. 2000 onwards. *A Comparison of Interactive Identification programs.* URL: http://www.delta-intkey.com/

Dallwitz, M. J., Paine, T. A. & Zurcher, E. J. 1993 onwards. *User's Guide to the DELTA System: a General System for Proces-sing Taxonomic Descriptions*, ed. 4. URL: http://delta-int key.com/ CSIRO Division of Entomology, Canberra.

Dallwitz, M. J., Paine, T. A. & Zurcher, E. J. 2000 onwards. *Principles of Interactive Keys.* URL: http://delta-intkey.com/

Dallwitz, M. J., Paine, T. A. & Zurcher, E. J. 2002 onwards. *Interactive Identification Using the Internet.* URL: http://delta-intkey.com/

Hansen, B. & Rahn, K. 1969. Determination of angiosperm families by means of a punched card system. *Dansk Botanisk Arkiv* 26: 1–46 + 172 punched cards.

Heidorn, P. B. 2001. *A Tool for Multipurpose Use of Online Flora and Fauna: The Biological Information Browsing Environment (BIBE)*. URL: http://firstmonday.org/issues/issue6_2/heidorn/index.html

Jarvie, J. K. & Ermayanti, T. M. 1995–1996. *Trees of Borneo*, English edition [computer file]; Pohon di Borneo, edisi Bahasa Indonesia. Conservation International-Indonesia, Natural Resources Management Program, Jakarta.

Jarvie, J. K. & Stevens, P. F. 1998. Interactive keys, inventory, and conservation. *Conserv. Biol.* 12: 222–224.

Jensen, R. J. 2003. The conundrum of morphometrics. *Taxon* 52: 663–671.

Loizeau, P. A. 1994. Peruvian Aquifoliaceae (elements for a revision of the Neotropical Aquifoliaceae). *Boissiera* 48: 1–306.

Lu, L. T. & Brach, A. R. 2003. *Cotoneaster*. Pp. 85–108 in:

Wu, Z. Y. & Raven, P. H. (eds.), *Flora of China*, vol. 9. Science Press, Beijing, Missouri Botanical Garden Press, St. Louis.

Pankhurst, R. J. 1991. *Practical Taxonomic Computing.* Cambridge Univ. Press, Cambridge.

Radford, A. E., Dickison, W. C., Massey, J. R. & Bell, C. R. 1976. *Vascular Plant Systematics*. Harper and Row, New York. URL: http://www.ibiblio.org/ botnet/glossary/

Rossi-Wilcox, S. M. 2001. *Chinese-English Botanical Glossary*. URL: http://flora.huh.harvard.edu/Glossary/

Schatz, G. E. 2001. *Generic Tree Flora of Madagascar.* Royal Botanic Gardens, Kew; Missouri Botanical Garden, St. Louis.

Simpson, D. R. & Janos, D. 1974. *Punch Card Key to the Families of Dicotyledons of the Western Hemisphere south of the United States.* Field Museum of Natural History, Chicago.

Watson, L. & Dallwitz, M. J. 1981. An automated data bank for grass genera. *Taxon* 30: 424–429.

Wu, Z. Y. & Raven, P. H. (eds.). 1994–. *Flora of China*. Science Press, Beijing, Missouri Botanical Garden Press, St. Louis.