Aliso: A Journal of Systematic and Evolutionary Botany

Volume 23 | Issue 1 Article 21

2007

An Electronic World Grass Flora

Steve Renvoize
Royal Botanic Gardens, Kew, UK

Derek Clayton
Royal Botanic Gardens, Kew, UK

Tom Cope
Royal Botanic Gardens, Kew, UK

Helen Williamson
Royal Botanic Gardens, Kew, UK

Aliso 23, pp. 267–270

AN ELECTRONIC WORLD GRASS FLORA

STEVE RENVOIZE,[1] DEREK CLAYTON, TOM COPE, AND HELEN WILLIAMSON

The Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK
[1]*Corresponding author (s.renvoize@rbgkew.org.uk)*

ABSTRACT

The development of an electronic world grass flora database is described in which data for 1090 morphological characters gathered for 11,000 species in 700 genera organized according to accepted names. This descriptive information is linked to a synonym database of 60,000 names. Authors, literature references, and the status of each name are also recorded in the database along with geographical distribution and type information. The list of accepted species is linked to a global herbarium of 350,000 specimens at Kew arranged in a phylogenetic sequence at the generic and species levels and according to broad phytogeographic divisions. From the database, descriptions can be generated for species, genera, and tribes. Character similarities or differences can be identified and character sets generated as an aid to key writing. Taxonomic and geographic subsets can be generated and specimens can be identified using an interactive key.

Key words: DELTA, descriptive data, geography, identification, Poaceae, synonyms, world grass flora.

INTRODUCTION

The herbarium collection of grasses at Kew has developed over a long period of time, starting with the core collections of Bentham and Hooker in the mid-nineteenth century and continuing through to the present day. The collection now numbers in the region of 350,000 specimens. In the past it has served as a resource for the preparation of numerous floristic accounts for many parts of the world, in particular the British Isles, the Middle East, tropical Africa, the Indian subcontinent, Australasia, and tropical South America. This history of floristic research has influenced the growth of the collection, and since its inception has attracted gifts from all over the world. As a result of this influx of material the collection is now very large, of global coverage, and of immense richness. While a resource such as this can continue to provide a tool for further piecemeal research of either a floristic or monographic nature, its size and near comprehensive geographic coverage offer an exceptional opportunity to explore on a global scale other aspects of diversity within the grass family. When a collection of this magnitude is combined with the volume of available literature and a wealth of experience there is the opportunity to undertake a synthesis of the family as a whole. However, a baseline collection, reference resources, and ambition to carry out such a work are not alone sufficient to facilitate its execution. The challenge is how to handle the vast amount of information as a single unit, which is what a global synthesis would require, and to avoid, as has been done in the past, the natural desire to break it down into more easily handled components on a regional or taxonomic basis. Furthermore, responding to user requirements of the collection is difficult as the information is widely scattered and time consuming to retrieve. These problems arise because the information that we handle is mostly descriptive with a bulky and complex structure, and using traditional methods of analysis the volume of text would be overwhelming and difficult to index. The answer is to convert the species descriptions to a numerical matrix that can be handled electronically.

An opportunity to realize this ambition came with the development by Mike Dallwitz (Dallwitz 1980; Dallwitz et al. 1993 onwards) in Canberra, Australia, of the DELTA System (DEscription Language for Taxonomy), a flexible method of encoding and computer processing of taxonomic descriptions. This program facilitates the handling of large amounts of descriptive data through a numerical format. Over the past 15 years the Grasses Section of the Herbarium at the Royal Botanic Gardens, Kew, has been developing two databases of grass information, one for descriptive data, based on the DELTA system, and another for synonyms, which utilizes Microsoft Access. In addition to the morphological characters, information on geographical distribution is included, which uses the regional divisions of the Taxonomic Database Working Group (TDWG; Brummitt 2001). Altitude, habitat, and uses are omitted at the present time as adequate information is difficult to obtain from herbarium specimens.

These databases were devised to serve as standard global lists of grass species and as resources for future floristic and systematic research.

DESCRIPTIVE DATA—THE DELTA SYSTEM

Descriptive data are handled by DELTA. This is a flexible suite of programs, which can store and search descriptive data to perform numerous operations. Some of the programs and their functions are as follows:

Confor
: Writes readable descriptions of species.
: Converts data into the formats used by other programs, including classification programs (e.g., PAUP*, Swofford 2001).

Key (Dallwitz 1974)
: Generates conventional keys.

Intkey (Dallwitz et al. 1995 onwards)
: Identification of specimens using interactive keys (Dallwitz et al. 2000 onwards).

Provides lists of similarities, differences, or diagnostic character sets, which can be used as support for hand written keys.
Amalgamates species descriptions into generic or tribal descriptions.
Produces taxonomic or geographic subsets.

The data set on which DELTA currently operates contains 11,000 species and records for 1090 characters, although not all characters are scored for each species. Improvement of the data continues both locally and through contributions from the wider community of grass taxonomists.

The database attempts to emulate the traditional descriptive portrayal of morphological diversity and diagnostic detail. Secondary objectives are to provide a discipline for testing comparative and diagnostic properties, and a framework for accumulating observations of potential interest. The database is confined to morphological characters. Anatomical and other cryptic characters are less numerous, sparsely recorded, and arguably better handled at the generic level.

The general principles governing the choice of characters and states are:

(1) Terminology should conform to traditional usage. A few unfamiliar expressions (e.g., *companion sterile spikelets*) have been used to secure generality, but are not intended as new terms.

(2) Characters should be broken down to simple elements, e.g., *principal lemma awn* has four states—*apical* (an implicit state), *subapical, from a sinus,* and *dorsal.* The main exceptions are a few syndrome characters useful for tribal separation, e.g., *bamboo life form,* which has two states—*herbaceous, or if woody then unlike a bamboo* and *woody bamboo*—and *panicoid,* which has two states—*spikelets two-flowered, the lower floret male or barren, the upper fertile* and *spikelets one–many-flowered, if two-flowered then both florets fertile or the upper sterile.* In addition, exceptions are made for mixed collections of unusual states, e.g., *roots,* which has four states—*simple* (an implicit state), *bearing tubers, aromatic,* and *woolly.* An implicit state is assumed to apply to any species for which the character is not recorded

(3) All diagnostic characters should be retrievable, but there is a problem with fine distinctions between critical species, which are difficult to accommodate in an open-ended system. These are handled by three text characters, which can be used for a precisely worded mini-key, e.g., *Agrostis delicatula* Pourr. ex Lepeyr.—*ligule longer than wide, acute; lemma apex lobed, the lobes 0.15 the length of the lemma; leaf-blades involute*; *A. durieui* Boiss. & Reut. ex Merino—*ligule as long as wide, truncate; lemma apex lobed, the lobes 0.25 the length of the lemma; leaf-blades flat.* When keying out species in the interactive key (Intkey) and the last few species are reached, these small diagnostic differences can be displayed.

(4) There should be a place for all characters, or at least suitable pegs on which to hang qualifying comments. Some loosely defined facies characters fall into this category. A few of the commonest qualifiers have been given character status, e.g., *clumped,* which has three states—*loosely, moderately,* and *densely.*

(5) The wording should be such that the characters can stand alone, but otherwise no attempt is made to cater for the special requirements of key writing.

(6) Strict standardization of subjective character states is regarded as impractical for data obtained from such a wide range of sources, but consistent distinctions within a genus should be attainable.

The character set is based upon phrases encountered during our wide experience of reading and writing descriptions, supplemented by a study of keys in most of the principal Floras, but inevitably it is a compromise between conflicting demands, not the least of which is the desirability of keeping it to a manageable size. Each character may be two-state (present or absent), multistate, a number, or a measurement. For example:

Two-state: *presence of glands on pedicels—eglandular/glandular*
Multistate: *leaf-blade color—yellowish green/light green/mid-green/dark green/glaucous/gray-green/red/purple*
Number: *bud complement on culm node—enter number of buds*
Measurement: *rachilla internode length—enter measurement*

One drawback of the database is that the descriptions were initially taken from the literature, including monographs and Floras, with minimal herbarium checking, many from a time when the character set was considerably shorter. Sourcing data from literature also led to weaknesses where regions or genera lacked an up-to-date account. However, as new work proceeds on various floristic and monographic projects where descriptive information is gathered directly from herbarium specimens, there is the opportunity for the descriptions to be updated accordingly. Thus, there is ongoing improvement to the descriptions based upon the latest research. The database is never static.

Identifications can be undertaken, but their effectiveness in this open-ended system depends upon having enough knowledge of grasses to cope with characters liable to subjective interpretation (e.g., shape) or an incompletely recorded range of states (e.g., leaf hairiness).

SYNONYM DATABASE

The synonym data set attempts to place all 11,000 known grass species in a systematic context by uniquely identifying each accepted species, genus, and tribe. This is a consensus taxonomy that is under continuous review according to emerging literature and in-house research. As a result there will inevitably be some changes to accepted names. In order to take full account of past and future literature it is important that a link be made to alternative names for the taxa. This is achieved through a synonym database.

This database, which utilizes Access, contains 60,000 names. The data are organized in three separate tables, Names, Geography, and Types, with a link to the descriptive database provided through the accepted name. It is supported by extensive validity checks for spelling mistakes or logical inconsistencies in the linkage.

Names Table

The names data are gathered according to 19 fields:

(1) Genus
(2) Species
(3) Rank [of infraspecific name]
(4) Infraspecific name
(5) Hybrid status (g = genus, s = species, v = infraspecific category)
(6) (Author) [Parenthetical author]
(7) Author [Publishing author]
(8) Publication reference
(9) Date of publication
(10) Remarks
(11) Status
A = accepted (requires matching the name to a description in the DELTA database)
B = available
C = illegitimate (other than homonym)
H = later homonym
O = orthographic variant
R = accepted hybrid
V = invalid (nom. inval., pro syn., nom. nud.)
X = uncertain application
Z = uncertain and invalid
(12) Homonym flag (distinguishing number for homonyms if there is more than one entry for the same name: 0 for available or only name, numbering of later homonyms is arbitrary)
(13) Basionym (earliest legitimate name, if available, otherwise choice is arbitrary; its function is to link homotypic synonyms together)
(14) Basionym flag (homonym number to match homonymous basionyms to their key name)
(15) Adopted name (accepted name)
(16) Adopted name flag (homonym number of accepted name: normally 0, but used when a name of uncertain application is a homonym and the ''Adopted name'' field can only be itself)
(17) Key name (full name as first published, not the accepted name)
(18) Source of data (comment on provenance of the record)
(19) Notes (memo field for discussion and comment)

Geography Table

The geographical data are organized in three fields:

(1) Full (accepted) name
(2) Region
(3) Country

For region and country the TDWG codes to level 3 are used. The codes comprise a two-digit number and a three-letter code, e.g., 90 SGE (90 = Sub-Antarctic Islands, SGE = South Georgia).

Types Table

For this table, data are recorded in the following seven fields:

(1) Key name
(2) Country
(3) Locality
(4) Collector name and number
(5) Herbarium and type status
(6) Verification (space for ''seen by'' comment)
(7) Notes (author of lectotype or other useful information)

The database is menu-driven, which provides several options. Data can be provided in two categories, synonyms and lists, and may be viewed in two modes, on-screen (form mode) or printed (report mode). Thus, information can be provided according to a variety of requirements, e.g., list of accepted names for a genus/region/genus within a region, list of accepted names with their synonyms for any input name, sorted by date or basionym, with options to include literature reference and type.

In addition, a list of accepted genera (currently 702) can be examined, as well as a list of all generic names. Genera are not presented as a synonymy; this is available in Clayton and Renvoize (1986).

The synonymy is currently a subjective assessment based upon the most reliable literature; ideally it should be moved forward towards a consensus nomenclature.

FUTURE DEVELOPMENTS

Taxonomy is not a precise science and there is always the facies factor to be taken into account. The experienced taxonomist may rely on a specimen or an illustration more than they are prepared to admit. No matter how detailed the description, the addition of an image is worth as much if not more than all the words. Usually a glance at an authentic specimen in a herbarium is all that is needed, but this may not be an option. The next best thing is a recorded image.

A program of image capture is currently under way at Kew. Initially the emphasis is on type specimens as they are unique and vulnerable, but types are primarily for nomenclature and do not necessarily help in identification. For this purpose an image of a representative specimen taken from the herbarium collection is the preferred solution and could be included in the future.

Both the descriptive and synonym databases can be downloaded from the Kew website (*www.rbgkew.org.uk*). Each is under continuous review with additions and corrections incorporated on a regular basis.

The project is open to contributions from all grass taxonomists and any changes or additions will be attributed to the persons providing information. In this way it will become an international database, giving all agrostologists access to morphological and nomenclatural information on a worldwide scale.

ACKNOWLEDGMENTS

The authors would like to thank Mike Dallwitz and Bryan Simon for helpful comments on the manuscript and also Kehan Harman for assistance in the final stages of preparation.

LITERATURE CITED

BRUMMITT, R. K. 2001. Geographical scheme for recording plant distributions, Ed. 2. International Working Group on Taxonomic

Databases for Plant Sciences, Hunt Institute, Pittsburgh, Pennsylvania, USA. 137 p.

CLAYTON, W. D., AND S. A. RENVOIZE. 1986. Genera graminum: grasses of the world. *Kew Bull., Addit. Ser.* **13:** 1–389.

DALLWITZ, M. J. 1974. A flexible computer program for generating identification keys. *Syst. Zool.* **23:** 50–57.

———. 1980. A general system for coding taxonomic descriptions. *Taxon* **29:** 41–46.

———, T. A. PAINE, AND E. J. ZURCHER. 1993 onwards. User's guide to the DELTA System: a general system for processing taxonomic descriptions, Ed. 4. *http://biodiversity.uno.edu/delta/* [Now at: *http://delta-intkey.com* Jun 2005].

———, ———, AND ———. 1995 onwards. User's guide to Intkey: a program for interactive identification and information retrieval. *http://delta-intkey.com* [Jun 2005].

———, ———, AND ———. 2000 onwards. Principles of interactive keys. *http://delta-intkey.com* [Jun 2005].

SWOFFORD, D. L. 2001. PAUP*: phylogenetic analysis using parsimony (*and other methods), vers. 4.0. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.