



## DELTA Newsletter

Number 3, April 1989

**Note from the Editor** — The DELTA Newsletter is designed to promote communication among scientists developing and applying computer technology in the collection, storage, analysis, and presentation of taxonomic data for the production of descriptions, keys, interactive identification, and information retrieval. To achieve this goal the DN will be issued in April and October of each year. Contributions in the form of short comments or long discussions and explanations are encouraged from all developers and users of DELTA format and similar systems. Comments on methods of application, suggestions for improvements, or criticisms of current technology are encouraged. This issue contains contributions from Bob Allkin on ALICE, Richard Zander on TAXACOM, Fred Rhoades on ASKATAXA, Doyle Anderegg on TAXADAT, Mike Dallwitz on INTKEY and CONFOR, and Richard Pankhurst on PANKEY. In addition, a set of literature citations which directly or indirectly apply DELTA technology is included. This is not a complete list, but it is a good start and I will be glad to present updates in future editions of the newsletter. Finally, an update on the 1988 directory is presented. — Robert D. Webster, USDA/ARS/SBML, Bldg. 265, BARC-East, Beltsville, MD 20705, USA.

### Introduction to ALICE

**Bob Allkin**

ALICE is a database system designed for botanists and zoologists wishing to establish or use, species diversity and checklist databases. We aim to provide biologists with a system having the flexibility and control that properly structured databases can offer, while at the same time shielding them from the complexities inherent in managing and using such databases. To this end, ALICE incorporates an easily used biologists' interface. Users select options from menus with single key strokes and need no previous database experience. ALICE is taxon-based. For each species, subspecies or variety entered, information can be stored about synonymy, distribution, uses, common names and habitats, as well as multi-state taxon descriptors defined by the user. Any fact can be referred to an entry in a citation list (e.g., the protologue of a name, or who said that plant x occurs in India?)

The program has a degree of biological 'sense' which it uses to validate data at entry and prevent logical blunders during editing. Output is in the form of standard lists or user defined reports. Users can ask questions about particular taxa or frame more general enquiries such as 'Which poisonous species are native to India?'

### History and philosophy

ALICE has been developed over the last five years primarily by Peter Winfield and myself. It is the intellectual property of, and is marketed by, the ALICE Software Partnership consisting of Peter, Frank Bisby and myself. We believe that biologists need not spend time learning how to write programs or use commercial database systems effectively. Nor should they have to undertake large data coding exercises. Today's technology should be made accessible to all biologists, and as painlessly as possible! For certain tasks, the data management requirements of many biologists are sufficiently similar for it to be possible to write generic software solutions, just as Pankhurst and Dallwitz have provided identification programs for botanist and zoologist alike. Producing good quality database software with documentation and support is expensive, but we believe the development costs will be repaid many times over 1) by the saving of man-hours as biologists are freed from developing their own system, on their own machine, for their own data – thereby duplicating the efforts of colleagues in many other institutes and 2) by the

increased opportunities for the exchange of biological data (Allkin, 1988 'Taxonomically intelligent database systems'. In 'Prospects in Systematics', ed. Hawksworth, D. L., Oxford Univ. Press).

### **Example uses of ALICE**

The International Legume Database and Information Service (ILDIS) is using ALICE to build its monographic database for the Leguminosae involving research groups in more than 20 countries. ILDIS is also to use ALICE to distribute their data more widely. Other actual or potential users of ALICE include: species inventories and floras for individual countries or reserves (e.g. 'Orchids of Vanuatu'); ethnobotanical or ecological studies (e.g. 'Ethnobotany of the Colombian Amazon') listing species, together with their common names and uses or with habitat data and ecological descriptors. WCMC (IUCN) hope to distribute ALICE to field workers at remote sites for preparing conservation inventories. A CITES sponsored world checklist of the Cactaceae is being built at Kew.

### **The data**

The heart of any ALICE database is the Species Checklist. ALICE 'knows' about binomials, authorities, homonyms, misapplied names etc. but shields users from unnecessary complexity. Enquiries made using synonyms, for example, are automatically redirected via the 'accepted' name to the correct information. If homonyms occur, users are prompted to select from among them.

In entering distribution data users build their own gazetteer around a 3-tier hierarchy (e.g. Continent/Country/Area OR State/County/Town). The gazetteer is used to facilitate data capture and ensure logical consistency: plants recorded from Alabama must also occur in the USA! Users define their own categories of introduction status, conservation etc. at each geographic level.

Other datatypes include vernacular names, uses, habitat and notes. Any number of taxon descriptors maybe defined to record species attributes that are appropriate to organisms in the database e.g. 'Wing span' or 'Flower colour'. Apart from being used to back up individual facts in the database, the Citation List is also used to indicate sources of additional information (e.g. where to find Illustrations, Maps or Descriptions).

### **ALICE program modules and functions**

A family of programs operate on ALICE databases. Each has a different function and set of users in mind.

- ALICE – The workhorse program for those building databases: database design, data capture, database maintenance, standard list output and taxon based enquiries.
- QUERY – flexible data retrieval with simple interface for intermittent use by 'end users' with superficial knowledge of what is in the database.
- ALICE REPORT GENERATOR – for flexible control of reports.
- ALICE FOR TYPESETTING – special routines to post process ALICE reports and embed typesetter control statements.
- IMPORT and EXPORT – for exchange of data with other databases.
- MERGE and SPLIT – for manipulation of entire ALICE databases.

### **Database design and performance**

The ALICE system was designed using the relational database model – for whatever that's worth! Benefits to the user are i) efficient use of data store, ii) the degree of data validation and data consistency checks made possible and iii) that database performance does not degrade significantly with larger data sets.

The African legume checklist (c. 6,000 taxa – see separate panel) with ALL the above datatypes (e.g. distribution records from 139 countries in 13 sub-continent linked to 580 citations) with many text notes, occupies c. 10.5 Mb. The entire ILDIS checklist (23,000 taxa) is estimated to require 40 Mb.

## Cost and support

ALICE version 1 has been in use during the last three years and demonstration packs (including a tutorial manual, a test data set for 7 plant species and a demonstration copy of ALICE) are available for 20 pounds sterling. ALICE version 2 is in Beta test by a number of projects and should be ready in May 1989. Documentation is being prepared. The price of a single user registration is 600 sterling. The price is negotiable for those working in developing countries. Institutional licenses are available. Demonstration copies of AQUERY for flexible database searches are also available.

## Machine requirements

ALICE requires an IBM compatible microcomputer (XT, AT, 386 clone etc.) with a hard disk, 512 K RAM and PC DOS V.2.1 or later. Performance is very adequate on a 286 machine and can best be improved on older machines with a larger or better quality hard disc.

## Final note

I thank the editor for his invitation to submit this contribution. We hope that it will become a regular feature of the DELTA Newsletter and I am grateful for the opportunity to reach a wider audience. This time I have tried to give an overview of ALICE and in subsequent issues will describe components of the system in more detail and report recent developments. Those requiring further information please contact me at the following address.

*Bob Allkin, Royal Botanic Gardens KEW, Richmond, TW9 3AB, UK. Tel. + 44-1-940-1171 ext. 4715. Telex 296694 KEWGAR G. Fax +44-1-948-1197. Email BTGOLD 81:bio023.*

## Comments on TAXACOM

### Richard L. Zander

TAXACOM, established January 12, 1987, is a free service for collections-oriented biologists. It is completely user-supported, meaning that no services are sold, and users simply share data files and programs with each other. Contact is by microcomputer and asynchronous modem through standard telephone lines at (716) 896-7581, 300 or 1200 or 2400 bps, 8 data bits, no parity, one stop bit. (Users outside USA should note use of CCITT protocol at 2400 bps, but Bell protocols at 300 and 1200 bps.) Any interested person may call and examine most portions of TAXACOM anonymously by logging on as a 'GUEST', after signing up and validation, full access is granted.

The purpose of TAXACOM is to explore opportunities in the field of electronic communications among systematic biologists. We intend to provide innovative means of access to computerized data and biologically oriented programs, and to demonstrate the utility of online searchable databases and formal electronic publication of scientific papers.

Conferences on TAXACOM are mini-symposia in which particular topics may be discussed publicly. Messages may have binary or text files 'attached' to them (uploading is by Xmodem protocol). The present public Conferences include: Botanical Latin, Bryology & Lichenology, Curator's (specimen exchange and conservation techniques), DELTA, Membership, Mycology, Niagara Frontier, Offers of Positions, Online Communications, Open, Ornithology, Phylogenetics, Questionnaires, Research, and Share Programs. An additional three conferences, DBMLS Design, Geographic Information Systems and Technical, are hosted by J. Beach (Field Museum, Chicago). There are several libraries of downloadable text files and programs, e.g. Flora Online, Cyclopedia (Miscellaneous publications, including two electronic magazines), Botanical Programs, Communications Programs, and Database Help.

The DELTA Conference should be of considerable interest to DELTA users in that it is moderated by M. Dallwitz, who calls periodically from Canberra. He is willing to answer questions and encourages discussion of the DELTA system and its potentiality. It is expected that software updates of DELTA and associated packages will be posted in this conference gratis for downloading by TAXACOM users.

The Botanical Latin Service, run by P. M. Eckel, provides systematists with Latin translations of descriptions of new taxa. Fine points and questions of language may be discussed in the Latin Conference.

Flora Online is the first electronic publication to be assigned an ISSN number from the US Library of Congress (ISSN 0892-9106). Twenty monograph-style issues have been published to date, mainly of data-intensive 'papers' and botanically related programs. It is available both online on TAXACOM and on diskette (5.25- or 3.5-inch, DS-DD, MS-DOS format) by subscription. Write Shaun Hardy, Research Library, Buffalo Museum of Science, 1020 Humboldt Pkwy, Buffalo, NY 14211 USA, for ordering information. Readers of the DELTA Newsletter are invited to contribute to Flora Online; instructions to authors are in the second issue available, of course, online.

Searchable database on TAXACOM include the Field Museum of Natural History Botanical Type Photograph Database, consisting of more than 62,000 records. This is, as far as we are aware, the first substantial searchable database of collection information to be made freely available by modem to researchers and the public. A simple query system is used to retrieve records of specimen label information and the appropriate Field Photograph number. This database can also be used to ascertain the whereabouts of many type specimens of vascular plants at the Field Museum and certain other herbaria. The ease of access to this large type database should make tracking down the location of types less arduous. For types deposited at Berlin-Dahlem (B) but destroyed during WW II, the Field photograph collection is the only remaining record.

TAXACOM users may also communicate by private electronic mail that allows appending of downloadable files.

Long-distance phone charges should not exceed \$.15 to .20 per minute (about \$10–12 per hour) for evening access from any part of the USA. This is similar to costs of accessing commercial networks such as CompuServe; TAXACOM, unlike commercial networks, supports both 1200 and 2400 bps connection from anywhere in the country. Although Internet may appear free to most users, the (considerable) cost is picked up by the institution. Internet access to TAXACOM would cost the Buffalo Museum of Science at least \$7000 per year. This is too expensive for a small institution, but we are investigating alternative possibilities.

We will be glad to help other institutions set up similar electronic services with functions appropriate to their collection character and the particular specialties of their staff. It may be expected that biologists will eventually be able to share data and programs of importance in an orderly and responsible fashion through a series of electronic 'bulletin boards,' using both dedicated lines (Internet, Bitnet) and standard phone lines (e.g. TAXACOM)

*Richard H. Zander, Buffalo Museum of Science, 1020 Humboldt Pkwy, Buffalo, NY 14211, USA.  
Telephone +1 716 896-5200.*

## **Comments on ASKATAXA**

### **Fred Rhoades**

Four years ago, someone handed me an Apple-II disk with an early attempt at a synoptic key generator for microcomputers. I saw the potential for introducing beginners to mushrooms and other groups of cryptogams, but was so frustrated with the program's non-user-friendliness that I decided to attempt programming my own. Unaware of the approaches taken in main-frame and other microcomputer programs by Pankhurst and others at the time, I plowed ahead with a different scheme. The result is the synoptic key package, PC-TAXON, distributed for the IBM-PC and compatibles by COMPress (Queue, Inc., 562 Boston Avenue, Bridgeport, CT 06610, U.S.A.

PC-TAXON is written in Turbo Pascal and consists of two main programs. TAXON and ASKATAXA. TAXON is used to create and edit interactive keys. The key may be queried from TAXON during creation, or the public domain program, ASKATAXA, can be distributed with completed keys, allowing key-access without editing. Keys may have up to 255 taxa and up to 99 characters, each with up to 99

character-states. Character-states may be coded as descriptive text to clarify the characters or states. Brief, supplemental text additionally describing each taxon can be stored and recalled. Queries of a key can include searches, comparison of two taxa, descriptions of individual taxa and listing of 'field keys' to printer or disk. Searches allow AND, OR and NOT qualifiers and 'Back up', 'allowance of error (in character selection)' and 'find best character' options. Context-sensitive messages are available at each step of key-building or querying. Because of extensive error-trapping, programs rarely, if ever, dump the user into never-never land.

The in-memory structure of C-TAXON keys is simpler than that used by DELTA. To keep memory use to a minimum and to allow rapid filtering through characters, the memory image is composed of sets of taxa exhibiting each character-state. This does not allow for scoring percentage presence of character-states among members of taxa (a taxon either exhibits a character state or doesn't), but results in small program size, rapid access to keys, easy coding of AND, OR and not links in key searches (using Pascal's set-handling operators), and a very fast algorithm for finding the best character. By using data compression, the data matrix for each key needs very small disk space. Names of taxa, character and character-state descriptions and the compressed data matrix are all stored in a single file. Supplemental text descriptions of taxa are stored in a separate file.

The aim in developing PC-TAXON was to produce a foolproof utility to generate written synoptic keys, and a program for use by undergraduate students being introduced to the use of all kinds of keys. PC-TAXON is easy to use but is somewhat limited in its abilities. Specifically, two features are not present in PC-TAXON: dichotomous (or polychotomous) key building and the ability to score percentage presence of character-states among taxa. In order to take advantage of these capabilities in the DELTA-format KEY and CONFOR programs, and to allow 'up grading' keys beyond the limits imposed by the PC-TAXON system, I recently have completed two conversion programs allowing exchange between PC-TAXON- and DELTA- formats.

If you are unfamiliar with PC-TAXON and would like to experiment with approach, sets of keys (and the ASKATAXA program to access them) are available to fungi and lichens/bryophytes. If you have a modem, the keys are available for down-loading from TAXACOM, a free bulletin board service for biosystematics and biogeography at the Buffalo (New York) Museum of Science (716 896-7581, use standard protocol). Access the FLORA-ONLINE section of TAXACOM. Otherwise, the keys and conversion programs are available from me (\$5.00 for each disk: fungi keys, lichen/bryophyte keys, conversion programs (These are not yet available on TAXACOM; send requests to Fred Rhoades, Mycena Consulting, 4320 Dumas Avenue, Bellingham, WA 98228, U.S.A. Information on PC-TAXON from COMPRESS will be included with each request.

## **Comments on TAXADAT**

**Doyle E. Anderegg**

TAXADAT is a software program to assist taxonomists in writing descriptions and constructing diagnostic keys. The program is written in interpreted BASICA and runs on IBM PCs with BASICA and compatible microcomputers with GWBASIC. Minimum machine requirements are at least one floppy disk drive and 128K memory.

Program options are: 1) CONSTRUCT/EDIT DATABASE, 2) INTERACTIVE IDENTIFICATION, 3) CONSTRUCT KEYS (Linear or Synoptic), and 4) UTILITIES (for writing descriptions of taxa and altering data set dimensions). Use of large data sets is feasible since most of the data is maintained as random access files on disk. Use of RAM disk drives and/or a hard disk offers a significant speed advantage.

The program is completely menu-driven in all of its operations. Output is to the screen, disk, and to the printer (if a printer is in the system and turned on) during key construction and/or description writing. Files may be dumped to the printer for checking, as well as to word processor text files (for modification and reentry). Any word processor which reads and writes an ASCII file may be used.

The program data files are set up under program control and may be entered and edited in the same manner. Characters, character states, and taxa (object names) are random access text files. A binary table contains the 'relationships' between the taxa and character/character states. As the user selects the character states exhibited by each taxon (from none to all), the program calculates the decimal equivalent of the binary number and enters it into a table.

Once set up, any of the files may be modified by deletion and/or addition of characters or character states, using the program utilities alone or in conjunction with a word processor. Files may be tested with partial data sets.

The Interactive Keying program allows the user to select characters in any of several ways: by character number, from a list of characters, or by keying one (or more) letters of the character name. Characters are entered singly, in any order. When a single taxon is arrived at, by elimination of nonmatching taxa, the name appears on the screen. At any intermediate step, the user may examine the list of taxa still under consideration. A complete description of any taxon may be obtained and/or any two taxa may be compared (only characters in which they differ are presented).

Also, at any step before identification, the program will present a list of 'BEST' characters with suggestions for 'EFFECTIVE' character states.

The 'Linear Key' option allows construction of a key to the whole data set or a selected subset. Individual characters may be 'weighted' to influence the character sequence within the key. The 'Synoptic Key' option uses the whole data set. In either case, the keys are written to the screen, to disk, and to a printer. The disk file is an ASCII file readily processed by any of several word processors which accept ASCII input files.

The programmer's objective is to optimize the user's total time and effort by combining data entry, interactive keying, description writing, and key generation in a single program package which lends itself to manipulation with a word processor during database construction and in the final processing of program output.

*TAXADAT, 825 West C St, Moscow, ID 83843, USA.*

## **Diagnostic descriptions from INTKEY and CONFOR**

**Mike Dallwitz**

[An updated version of this article is available at <http://delta-intkey.com/diagnostic-descriptions.htm>.]

INTKEY is a program for interactive identification and information retrieval. It was briefly described in DELTA Newsletter No. 2. This article illustrates INTKEY's facilities for generating diagnostic descriptions, and for saving information for later use by other programs. The data used in the examples are the sample set which is provided with our programs. Note that the data and examples were chosen purely to illustrate the operation of the programs, and therefore no biological significance should be attached to the results.

A diagnostic description for a taxon is a description which distinguishes any representative of it from representatives of all other members of a given group of taxa. That is, the description is true for representatives of the given taxon, but not for any other representatives of the group. Richard Pankhurst's PANKEY package has a program, SPD1, specifically for finding diagnostic sets of characters, and INTKEY has a command, DIAGNOSE, for this purpose. Unlike SPD1, DIAGNOSE does not guarantee to find the smallest set of diagnostic characters, and, for given settings of various parameters, it finds only a single set. However, DIAGNOSE does have the advantage that character reliabilities (weights) are taken into account in selecting characters to be included in the diagnostic set.

In the following INTKEY dialogues, the user's responses are shown in a bold face. We start by examining the list of character keywords, which have been provided to allow easy reference to commonly required subsets of the characters.

INTKEY version: 13-MAR-89.

M.J. Dallwitz and T.A. Paine, CSIRO Division of Entomology, Australia.  
Sample Grass Genera 09:23:39 10-MAR-89

Enter command: **keywords characters**

CHARACTER KEYWORDS

Used

Available

ALL

Nomenclature

Vegetative form Habit

CULms (form)

LEAVes (form)

Reproductive organization

INFlorescence form

FSs - female-sterile spikelets

FFs - female fertile spikelets

GLumes

INComplete florets

FLorets (female-fertile)

LEMmas (female-fertile)

AWNs of female-fertile lemmas

PAleas (female-fertile)

ANDroecium of female-fertile florets

Gynoecium

FRuit

LBa - leaf blade anatomy

PHotosynthetic pathway, etc

Biochemistry

TS - transverse-section anatomy of the leaf blade

Diagnostic features of individual taxa

CLassification and number of species

SUBfamilies

SUPERtribes

TRibes and subtribes

GEography

PReset for diagnostic descriptions

The characters which specify the classification and geographical distribution would not normally be used in diagnostic descriptions, so we exclude them. (Note the abbreviations used in the response. Any command words and keywords can be abbreviated by truncation, as long as enough letters are left to distinguish them from other words which could validly appear in the same context. The 'c' in the response is an abbreviation of 'characters'. The keywords could have been further abbreviated to 'cl' and 'ge'.)

Enter command: **exclude c class geog**

10 characters excluded.

We now ask for a diagnostic description of taxon 12. (The 'Enter preset characters' prompt will be explained later.)

Enter command: **diagnose**

Enter taxon range, keyword or combination: **12**

Enter preset characters: *(Press 'Enter')*

12. Phragmites <Adans.>

52: palea

2. conspicuous but relatively short

8: leaf blades

1. broad

The diagnostic description can be strengthened by increasing the value of the TOLERANCE parameter, which is 0 by default. A diagnostic description generated with TOLERANCE set to  $n$  will differ from the corresponding description of any other taxon in at least  $n+1$  respects.

Enter command: **set tolerance 1**

TOLERANCE set to 1.

14 taxa remain.

Enter command: **diag 12**

Enter preset characters:

12. Phragmites <Adans.>

52: palea

2. conspicuous but relatively short

8: leaf blades

1. broad

11: adaxial ligule

3. a fringe of hairs

26: spikelets

1. compressed laterally

The 'Enter preset characters' prompt allows the user to specify a set of characters which will be added to the diagnostic set before the program starts its search for characters. This could be done to ensure that all diagnostic descriptions contain a common core of characters, or simply to speed up the generation of the descriptions for large sets of data. A suitable set of characters for the sample data is defined by the PRESET keyword. (Note that the name of the keyword does not have to be 'PRESET' – any name could have been used.)

Enter command: **diag 12**

Enter preset characters: **pr**

12. Phragmites Adans.

4: culms

1. woody and persistent

2. herbaceous

9: leaf blades

2. not pseudopetiolate

11: adaxial ligule

3. a fringe of hairs

13: inflorescence

5. paniculate

19: spikelets

2. not in distinct long-and-short combinations

27: spikelets

1. disarticulating above the glumes

41: female-fertile florets

2–10

57: ovary

1. glabrous

62: hilum

1. short

64:

2. C3

73:

2. fruiting inflorescence not as in Zea

32: glumes

1. very unequal

Character 73 is the last of the preset characters. Although there were already many more characters in use than in the previously obtained diagnostic set, the program still had to add character 32 to make the set diagnostic (at the given TOLERANCE).

The diagnostic descriptions produced by INTKEY are in a crude format, not suitable for publication. Furthermore, some of the information in the original DELTA-format descriptions is not available to INTKEY. The program CONFOR has better facilities for formatting descriptions, including the automatic insertion of typesetting marks, but cannot generate diagnostic sets of characters. However, INTKEY can be made to output its diagnostic sets in a form suitable for input to CONFOR.

The first step in this process is to make INTKEY direct its output to a file.

Enter command: **output on diag.tmp**

All output to file.

The output produced by *any* command will now be placed in the file DIAG.TMP, in addition to being displayed on the screen. To avoid cluttering this file with unwanted material, we can give the command

Enter command: **output off**

Output to file halted.

This suppresses output to the file, *except* for output produced by the commands REMARK and SAVE. The sole purpose of these commands is to place information in output files. The REMARK command is used to insert comments in the file. The SAVE command outputs various information in formats suitable for input to other programs.

To save diagnostic character sets for all of the taxa in the sample data set, we could then proceed as follows.

Enter command: **remark Tolerance 1, no preset characters**

Tolerance 1, no preset characters

Enter command: **save diag**

Enter taxon range, keyword or combination: **all**

Enter preset characters:

Diagnosis for taxon 1 is incomplete.

#1 7 11 27 33 36

#2 16 27 32

#3 32-33 57 63

#4 15 27 68

#5 11 13 32 37 41 65

#6 13 27 32-33 54 59

Diagnosis for taxon 7 is incomplete.

#7 13 26-27 32 42

#8 32 61 63

#9 11 32 37 50 62

#10 29 37 56 62

Diagnosis for taxon 11 is incomplete.

#11 12-13 26-27 32 42

#12 8 11 26 52

Diagnosis for taxon 13 is incomplete.

#13 3 13 25 32-33 50 56 62

#14 12 73

We can also make use of the KEYWORDS facility (which CONFOR does not have) to output a set of character numbers to be used by CONFOR to produce descriptions of the required form. We then finish the INTKEY session.

Enter command: **include c all**

83 characters included.

Enter command: **rem Characters to be included in descriptions**

Characters to be included in descriptions

Enter command: **save c preset geog class**

4 9 11 13 19 24 27 41 57 62 64 73-83

Enter command: **finish**

Output files–

DIAG.TMP

The output file DIAG.TMP looks like this:

Tolerance 1, no preset characters

#1 7 11 27 33 36

#2 16 27 32

#3 32-33 57 63

#4 15 27 68

#5 11 13 32 37 41 65

#6 13 27 32-33 54 59

#7 13 26-27 32 42

#8 32 61 63

#9 11 32 37 50 62

#10 29 37 56 62

#11 12-13 26-27 32 42

#12 8 11 26 52

#13 3 13 25 32-33 50 56 62

#14 12 73

Characters to be included in descriptions

4 9 11 13 19 24 27 41 57 62 64 73-83

By means of a text editor, we could use this file to make up the following CONFOR directives:

\*INCLUDE CHARACTERS

\*ADD CHARACTERS

#12 8 11 26 52

#13 3 13 25 32-33 50 56 62

#14 12 73

(All of the diagnostic character sets would normally be included in the ADD CHARACTERS directive. Here, we are using only three taxa, to save space.)

By incorporating these with other appropriate CONFOR directives, generating the descriptions, and typesetting the result, we obtain

**Phragmites** Adans.

Leaf blades broad. Adaxial ligule a fringe of hairs. Spikelets compressed laterally. Palea conspicuous but relatively short.

**Poa** L.

Culms 4–150 cm high. Inflorescence paniculate. Spikelets 2–11 mm long. Glumes more or less equal; decidedly shorter than the adjacent lemmas. Lemmas 5 nerved (usually), or 7–11 nerved (rarely: e.g. *Neuropoa* Clayton). Stamens 3, or 0 (when dioecious). Hilum short.

**Zea** L.

Plants monoecious with all the fertile spikelets unisexual. Fruiting inflorescence a massive, spatheate cob, the fruits in many rows.

Alternatively, we could use the following directives

\*INCLUDE CHARACTERS 4 9 11 13 19 24 27 41 57 62 64 73-83

\*ITALICIZE CHARACTERS #12 8 11 26 52

to obtain

***Phragmites*** Adans.

(4) Culms woody and persistent to herbaceous (often somewhat persistent). (8) *Leaf blades broad*. (11) *Adaxial ligule a fringe of hairs*. (13) Inflorescence paniculate. (19) Spikelets not in distinct long-and-short combinations. (26) *Spikelets compressed laterally*; (27) disarticulating above the glumes (at least above the L1). (41) Female-fertile florets (2–)3–10. (52) *Palea conspicuous but relatively short*. (57) Ovary glabrous. (62) Hilum short. (64) C3. (74) Arundinoideae; (77) Arundineae. (81) ‘Nearest neighbours’ *Crinipes* 0.0810, *Arundo* 0.0957, *Festuca* 0.1056, *Panicum* 0.1057. (82) 3 species. (83) Holarctic Kingdom, Paleotropical Kingdom, Neotropical Kingdom, Australian Kingdom, and Antarctic Kingdom.

Character numbers have been included in this description, to make it easier to see its relationship to the directives used to generate it. Notice that the ‘PRESET’ characters have been included (along with the geography and classification characters) via the INCLUDE CHARACTERS directive. The ‘PRESET’ characters were not (in this instance) used in the generation of the diagnostic characters, which have been included via the ITALICIZE CHARACTERS directive. It is also possible to use the ADD CHARACTERS and ITALICIZE CHARACTERS together, each with different sets of characters. And, of course, the characters used can come from any source — not necessarily from INTKEY. The possibilities are almost unlimited. *(To be continued)*

## **Differences between PANKEY and CONFOR**

**Richard Pankhurst**

The PANKEY programs still use Version 2 of DELTA. The differences are not great in practice, and since DELTA 2 is upward compatible with DELTA 3, you can simply use DELTA 2 for both PANKEY and CONFOR. There are also a number of differences of practice between PANKEY and CONFOR which were not intentional, but there has nevertheless been some divergence. I am committed to converting to Version 3, but this will take a while, since there are more than 10 different programs to convert. Also, higher priority is being given to developing the DELTA editor (DEDIT) as this is seen to be more important.

### **FORMAT differences**

The following differences have been observed:

- 1) Running together character numbers in directives. In DELTA 2 you need to write out CHARACTER TYPES and NUMBERS OF STATES separately for each character; e.g. 23,RN 24,RN 25,RN and not 23–25,RN, and similarly 6,3 7,3 8,3 and not 6–8,3
- 2) IMPLICIT VALUES are not implemented in PANKEY, except in DEDIT and in KCONP (for ONLIN5 and 6 and KCONI).
- 3) Qualitative character states may not be put into ranges in DELTA 2, so that if character 5 has states 1, 2 or 3, you must put 5,1/2/3 and not 5,1–3
- 4) PANKEY expects the directive CHARACTER DESCRIPTIONS instead of CHARACTER LIST.
- 5) DEPENDENT CHARACTERS are programmed as a type 5 directive in PANKEY, so it appears after the CHARACTER LIST and before the ITEM DESCRIPTIONS.
- 6) Quantitative KEY STATES in PANKEY are not programmed to accept truncated ranges at the beginning and end of a sequence; i.e. the forms ~t and t~ (DELTA manual p 54) are not accepted. Also, PANKEY does not yet accept KEY STATE data for qualitative (IN) characters, only for quantitative (RN).
- 7) The feature of V.3 where quantitative characters in descriptions can have bracketted parts of their ranges is not in version 2. e.g. 5,2–9.5 instead of 5,(2–)4–7.5(–9.5)
- 8) PANKEY only computes (where applicable) with CHARACTER WEIGHTS which are integers.

- 9) The '+' feature for ITEM DESCRIPTIONS is not in Version 2.
- 10) PANKEY expects every line of the DELTA file to have a blank in column 1 in order to show that there are no sequence numbers. Sequence numbers, if present, are ignored.
- 11) PANKEY expects the \*HEADING directive to end with a '/
- 12) '&' in item descriptions is treated as an error (see Newsletter 2).

### **Differences of practice**

These differences are in the way the programs work, which affect the data.

- 1) DELTA files. In PANKEY the normal practice is to keep all the DELTA data in one file, and not in several files as for CONFOR.
- 2) Dependent characters. PANKEY is stricter about the use of dependent characters than CONFOR is. If the DEPENDENT CHARACTER rules imply that a character state should be inapplicable, then both CONFOR and PANKEY check that the state is inapplicable (if scored), or record it automatically as inapplicable if it is not scored. If the controlling character is variable, then the dependent characters might be either inapplicable or scored. This is automatically allowed for and it is never necessary to write characters such as 5,-/2 if character 5 depends on 4 (say) which is variable. This means that you can simply leave out inapplicable characters and there is no need to score them. PANKEY checks for the consistent use of inapplicables and issues error messages accordingly. CONFOR allows the use of the 'inapplicable' coding in contexts where it is not implied by a character dependency.
- 3) MAXIMUM NUMBER OF ITEMS. PANKEY treats this number as the actual number of items to be expected, and gives an error message if the actual number of items encountered is not correct.

If anybody knows of any further problems, please let me know.

## **MACINTOSH versions of PANKEY programs**

**Richard Pankhurst**

Macintosh versions of PANKEY programs are now available, except for the graphics versions of ONLIN5 and ONLIN6 which will take a little longer. The conversion has proved far more difficult than for any other machine I have so far encountered! The famous Macintosh desktop environment is splendid for beginners to computing, but it is a real headache for program developers, and especially so when, as in PANKEY, the programs to be converted already have a graphics interface which is not the same as that in the Macintosh. When writing for copies, please specify whether you have the 'old' Macintosh or the MAC 2. The so-called toolbox interface is supposed to be compatible between the two versions, but I have evidence to the contrary! The MACPANKEY programs have been developed in my own time and at my own expense, and enquiries should be addressed to RJP at 203, Sheen Lane, London SW14 8LE, England.

## **Myths about DELTA**

**Richard Pankhurst**

Over a great many years, I have been explaining and demonstrating software for identification to a great variety of people. The reasons they have given from time to time for NOT adopting the DELTA approach to taxonomic computing are intriguing, and I have tried to set out some of them below.

- 1) 'The differences between my species are so subtle that only a human expert can tell them apart, and it would be impossible to put them down in a form that a computer could understand'. If that is so, then I doubt that the species are genuinely distinct!

2) 'The data on my species is incomplete and I am still changing my taxonomic views, so I can't use DELTA until I have finished collecting and analysing the data'. This is a complete misunderstanding; DELTA programs are generally designed to deal with incomplete and missing data and can easily be used to recast keys and descriptions when new data becomes available. Also, it is much better to get used to the beneficial rigour of the DELTA approach at any early stage, so that you get used to the extra precision of the methods.

3) 'Those methods are fine in Botany but could not be applied to Zoology'. This one is particularly pernicious and persistent. There is no difference that I have been able to detect between botanical and zoological taxonomy when it comes to writing keys and descriptions, but there is a widespread belief to this effect. It is true that zoologists are more often interested in the rather academic problems of classification (cladistics) than botanists are, and that this does seem to distract them from the practical business of identification, but there is no reason why the two disciplines should not run side by side. It is distressing to see taxonomists creating data matrices for what they see as a different purpose and then not putting them into DELTA.

4) 'DELTA will not be of any use to my taxonomic studies because it does not allow for probabilities'. As far as I can tell, identification applications which require probabilistic data are rare, except in microbiology and medicine. Generally speaking, with plants and animals, data on the probability of occurrence of characters and taxa are unobtainable or not very meaningful. The 'need' for using probability in medical applications is rather dubious; key methods have never really been tried in medicine. In the DELTA programs, uncertainty is catered for in a different way, by allowing specimens and taxa to disagree in a controlled manner, as in the 'online' and 'matching' programs. \*

5) 'DELTA can't be used for my group because the phylogenetic relationships have not been clarified yet'. This is simply nonsense; identification has no need of any theories about evolution. \*\*

6) 'The DELTA format is too complicated for me to be able to apply it to my group'. This is naive; DELTA allows for all the actual kinds and ranges of variation which real organisms show, and would be at fault if it did not.

7) 'This special purpose identification software is all very well; why not use one of the expert systems which are now so popular and readily available?' In fact, the interactive DELTA programs do already possess the important features of expert systems, and have the important advantage of being based on a character matrix, instead of a set of rules. Interestingly, programs such as key constructing programs produce rules as output, instead of requiring them as input. Another interesting point is that the earliest online identification programs preceded expert systems by many years.

8) 'I am not going to use DELTA because its format is so horrible. You must provide a proper database, data capture screens, a help system and a fully-featured editor'. This is perfectly valid criticism, but otherwise just an excuse! Taxonomists have been successfully using DELTA format or its predecessors for about twenty years already. However, RJP and MJD are working on improved user-friendliness. Users should realise that the kind of software that micro users are accustomed to use comes from commercial software organisations with huge resources, and that it isn't very reasonable to expect us to be able to do the same. What this argument really comes down to is that, given the increasing realisation of the importance of taxonomic software, the managers of taxonomic institutes ought to be devoting much greater resources to encouraging it.

I would be interested to hear from anyone who has any other myths which they would like to add to my collection.

\*Provision for coding probabilities is planned for our DELTA database system (see DNL #2). — MJD

\*\*In addition to serving as a tool in identification, DELTA provides an efficient means of accumulating data, generating mechanically superior descriptions, and information retrieval. Furthermore, DELTA coded data is available for analysis by various phenetic and cladistic analysis software. — Editor

## Literature citations

- Boswell, K. and A. J. Gibbs. 1983. *Virus of Legumes 1983. Descriptions and Keys from VIDE*. Australian National University, Canberra.
- Boswell, K., M. J. Dallwitz, A. J. Gibbs, and L. Watson. 1986. The VIDE (Virus Identification Data Exchange) project: a data bank for plant viruses. *Rev. Pl. Path.* 65:221–31.
- Boswell, K. and A. J. Gibbs. 1986. The VIDE data bank for plant viruses. In *Development and Applications in Virus Testing* (eds. R. A. C. Jones and L. Torrance), pp. 283–7. Association of Applied Biologists, UK.
- Britton, E. B. 1986. A revision of the Australian Chafers (Coleoptera: Scarabaeidae: Melalonthinae). Vol. 4. Tribe Liparetrini: genus *Colpochila*. *Aust. J. Zool., Suppl. ser.* 118:1–135.
- Büchen -Osmond, C, K. Crabtree, A. Gibbs, and G. McLean (eds). 1988. *Viruses of plants in Australia – descriptions and lists from the VIDE database*. 590 pp. Research School of Biological Sciences, Australian National University, Canberra.
- Clifford, H. T. and L. Watson. 1977. *Identifying Grasses. Data, Methods and Illustrations*. University of Queensland Press, Brisbane.
- Dallwitz, M. J. 1974. A flexible program for generating identification keys. *Syst. Zool.* 23:50–7.
- Dallwitz, M. J. 1980. A general system for coding taxonomic descriptions. *Taxon* 29:41–6.
- Dallwitz, M. J. 1984. Automatic typesetting of computer-generated keys and descriptions. In *Systematics Association Special Volume No. 26, Databases in Systematics* (eds. R. Allkin and F. A. Bisby), pp. 279–90. Academic Press, London.
- Dallwitz, M. J. and T. A. Paine. 1986. *User's Guide to DELTA System. A general system for coding taxonomic descriptions* (3rd. edition). CSIRO Aust. Div. Entomol. Rep. No. 13.
- Dallwitz, M. J. and E. J. Zurcher. 1988. *User's guide to TYPSET. A computer typesetting program* (2nd. edition). CSIRO Aust. Div. Entomol. Rep. No. 18.
- Ellis, D. V. 1988. Quality control of biological surveys. *Marine Pollution Bulletin*, in press.
- Lazarides, M. and L. Watson. 1986. *Cyperochloa*: A new genus in the Arundinoideae Dumortier (Poaceae). *Brunonia* 9:215–21.
- Macfarlane, T. D. and L. Watson. 1982. The classification of Poaceae subfamily Pooideae. *Taxon* 13:178–203.
- Pankhurst, R. J. and J. M. Allinson. 1985. British Grasses – a punched card key to grasses in the vegetative state. Field Studies Council Occasional Publication No. 10 and BM (NH). 76 pp and 124 cards.
- Pankhurst, R. J. 1986. A package of computer programs for handling taxonomic databases. *CABIOS* 2:33–9.
- Pankhurst, R. J. and A. O. Chater. 1988. Sedges of the British Isles. BSBI Computer key No. 1, 40 pp. and floppy disc. BSBI Publications. Version 2 is imminent with Hercules graphics for 56 illustrated characters.
- Partridge, T. R., M. J. Dallwitz, and L. Watson. 1988. *A primer for the DELTA system on MS-DOS and VMS* (2nd. edition). CSIRO Aust. Div. Entomol. Rep. No. 38.
- Rao, C. K. and R. J. Pankhurst. 1986. A polyclave to the monocotyledonous families of the world. A computer generated identification key. BM(NH) 59 pp. and 235 cards. Also sold in computer program version (ONLIN5).
- Shaw, R. B. and R. D. Webster. 1987. The Genus *Eriochloa* (Poaceae: Paniceae) in North and Central America. *SIDA* 12(1):165–207.
- Taylor, R. W. 1978. A taxonomic guide to the ant genus *Orectognathus*. CSIRO Aust. Div. Entomol. Rep. No. 3.
- Taylor, T. W. 1979. New Australian ants of the genus *Orectognathus*, with summary descriptions of the twenty-nine known species (Hymenoptera: Formicidae). *Aust. J. Zool.* 27:733–88.
- Wallace, C. C. and M. J. Dallwitz. 1981a. *Key to Species of the Coral Genus Acropora from the Great Barrier Reef (Prototype)*. 17pp. James Cook University: Townsville, Australia.
- Wallace, C. C. and M. J. Dallwitz. 1981b. Writing coral identification keys that work. In *Proceedings of the Fourth International Coral Reef Symposium*, Manila, Vol. 2, pp. 187–90.

- Watson, L. 1981. An automated system of generic descriptions for Caesalpinioideae, and its application to classification and key-making. In *Advances in Legume Systematics* (eds. R. M. Polhill and P. H. Raven), pp. 65–81. Royal Botanic Gardens, Kew, with two microfiches.
- Watson, L. and M. J. Dallwitz. 1980. *Australian Grass Genera. Anatomy, Morphology and Keys*. Research School of Biological Sciences, Australian National University, Canberra.
- Watson, L. and M. J. Dallwitz. 1981. An automated data bank for grass genera. *Taxon* 30:424–9.
- Watson, L. and M. J. Dallwitz. 1983. *The genera of Leguminosae-Caesalpinioideae. Anatomy, Morphology Keys and Classification*. Research School of Biological Sciences, Australian National University, Canberra.
- Watson, L. and M. J. Dallwitz. 1985a. *Australian Grass Genera. Anatomy, Morphology, Keys and Classification* (2nd. edition). Research School of Biological Sciences, Australian National University, Canberra.
- Watson, L., H. T Clifford, and M. J. Dallwitz. 1985b. The classification of Poaceae: subfamilies and super-tribes. *Aust. J. Bot.* 33:433–84.
- Watson, L., S. G. Aiken, M. J. Dallwitz, L. P. Lefkovitch, and M. Dub. 1986a. Canadian grass genera: keys and descriptions in English and French from an automated data bank. *Can. J. Bot.* 64:53–70, with two microfiches.
- Watson, L. and M. J. Dallwitz, and C. R. Johnston. 1986b. Grass genera of the world: 728 detailed descriptions from an automated database. *Aust. J. Bot.* 34:223–30, with four microfiches.
- Watson, L. 1988a. Automated descriptions of grass genera. In *Grass Systematics and Evolution* (eds. T. E. Soderstrom, K. W. Hilu, C. S. Campbell and M. E. Barkworth). Smithsonian Institution, Washington.
- Watson, L. and M. J. Dallwitz. 1988b. *Grass Genera of the World – Illustrations of Characters, Classification, Interactive Identification, Information Retrieval. With microfiches, and floppy disks for MS-DOS microcomputers*. Research School of Biological Sciences, Australian National University, Canberra.
- Watson, L., M. J. Dallwitz, A. J. Gibbs, and R. J. Pankhurst. 1988c. Automated taxonomic descriptions. In *Prospects in Systematics*, ed. (D. L. Hawksworth), pp. 292–304. Clarendon Press, Oxford.
- Watson, L., M. Damanakis, and M. J. Dallwitz. 1988d. *The grass genera of Greece – descriptions, classification, keys*. 231 pp. University of Crete, Iraklion.
- Webster, R. D. 1983. A revision of the genus *Digitaria* Haller (Paniceae: Poaceae) in Australia. *Brunonia* 6:131–216.
- Webster, R. D. 1987a. *The Australian Paniceae (Poaceae)*. J. Cramer, Berlin & Stuttgart.
- Webster, R. D. 1987b. Taxonomy of *Digitaria* Section *Digitaria* in North America (Poaceae: Paniceae). *SIDA* 12(1):209–22.
- Webster, R. D. 1988a. Genera of the North American Paniceae (Poaceae: Panicoideae). *Syst. Bot.* 13(4):576–609.
- Webster, R. D. 1988b. Genera of Mesoamerican Paniceae (Poaceae: Panicoideae). *SIDA* 13(2):187–221.