



## DELTA Newsletter

Number 4, October 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, project descriptions, or criticisms of current technology are encouraged. In this issue, Bob Allkin compares ALICE and DELTA, J. Lebbe and R. Vignes introduce XPER, R. Pankhurst discusses topics on PANKEY, F. Rhoades describes the ASKATAXA interactive key program, and M. Dallwitz continues his discussion on the application of INTKEY. In addition, Richard Spjut submitted a review of TAXADAT, and D. Anderegg responded to the review. Finally, introductory notes are included by J.-P. Frahm on a computer software library and Laurence Hatch on the Microcomputer Plant Identification (MPI) System. — Robert D. Webster, USDA/ARS/SBML, Bldg. 265, BARC-East, Beltsville, MD 20705, USA.

### ALICE and DELTA: where's the link?

Bob Allkin

Following the description of the ALICE database system that appeared in the last DELTA Newsletter, we received a number of enquiries that made it clear that we should clarify the relationship between ALICE and DELTA. Here, therefore, I concentrate on describing the descriptive component of ALICE.

First of all we should clarify that ALICE is not an alternative to DELTA. DELTA is a data format concerned primarily with descriptive data. ALICE is a program for biologists creating species diversity databases that has two core modules handling i) taxonomy and nomenclature (subspecies, varieties, synonymy, homonyms, authorities, misapplied names etc.) and ii) distribution records and introduction status in a hierarchical gazetteer.

ALICE differs from the DELTA manipulation programs of Pankhurst and Dallwitz in that it doesn't provide specialist identification functions but instead offers database management facilities not found in those programs. Confusion may arise, however, since some ALICE modules do handle descriptive data about organisms. I list here the types of descriptive data contained in ALICE databases and then describe some of the things that ALICE allows you to do with that data.

#### Types of descriptive data in ALICE

Descriptive data occurs in two program modules: common knowledge and user-defined descriptors. Here I give only the briefest of descriptions of the common knowledge module since it is the user-defined descriptors that relate most directly to DELTA.

##### *Common-knowledge descriptors*

These currently include 'uses', 'habitats', 'common names' and 'literature pointers'. These are preset by ALICE but the list of alternative 'states' (uses, habitats or whatever) is defined by the user for every individual database. There is no limit on the number of 'states' (some ALICE databases have very long lists of habitats for example) and these are always maintained in alphabetical sequence by

ALICE. Users can create their own lists of uses etc. either by i) importing existing lists of standard terms, or ii) by entering their own reference lists by adding new terms as necessary when species descriptions are recorded.

A taxon may have any number of recorded uses or habitat etc. and each individual observation may have any number of bibliographic references attached. Thus it is possible to record, for example, which three books contain information about the use of species X as a forage crop.

The limitation of common knowledge descriptors is that users are unable to define new descriptors. A more straightforward parallel exists, however, between DELTA characters and ALICE user-defined descriptors.

#### *User-defined descriptors*

ALICE allows for the following DELTA character types.

- UM Unordered multistate characters
- EUM Exclusive unordered multistate characters
- OM Ordered multistate characters
- EOM Exclusive ordered multistate characters

Each character may have any number of states. ALICE V2 also supports free text notes and will allow word processor facilities with free text taxon descriptions of any length. In the future ALICE V3 will enable users to define 'text characters' to store complex information such as 'chromosome no.' separately.

ALICE does NOT currently support integer or real numeric characters. Thus for 'tree height' users must define an ordered multistate character using arbitrary size ranges (e.g. 0–1m; 1–2m; 2–4m; greater than 4m).

Missing observations may be made explicit using the pseudostate 'missing' or simply left unrecorded. Inapplicable characters can be treated similarly. The database structure underlying ALICE supports character dependence but as yet the user interface software has not been implemented other than in the geographic module.

### **ALICE functions for manipulating descriptive data**

#### *Data input*

Data can be entered into ALICE databases either by data capture at the terminal or in batch mode via the XDF data exchange format using the program SAM. Either route involves much data validation which given the space cannot be described here and two simple examples must suffice. 1) It is impossible to enter more than one observation per taxon for characters designated as 'Exclusive'; and 2) constant use is made of look up tables e.g. the states of character X OR the current bibliography. This saves considerably on computer storage and protects users from error prone and time consuming typing. During data entry or retrieval users need only to select from existing lists.

In the medium term ALICE Software is committed to providing for DELTA data sets to be imported into ALICE databases. To prioritise our development it would be useful to know what interest there is in such a mechanism.

As elsewhere in ALICE databases individual observations (taxon 't' has state 'x') can be linked to any number of source references in the ALICE bibliographic database.

#### *Data editing*

The following editing functions are available: to correct or delete individual observations; to add, reorder, delete or rename character states; to add, delete, rename or modify characters.

The validation rules used during data input apply equally during editing.

*Data retrieval*

Using the program AQUERY you can search ALICE databases using any combination of characters e.g. list those plants with 'Leaf = compound' AND 'Habit = Tree'.

While building up enquiries users are prompted with a list of available characters and then of the alternative states for that character. The logical operators available are AND, OR and NOT. It is also possible to build searches combining descriptive information with other types of data e.g. 1 'Country = Mexico OR Guatemala' AND 'Wing spot colour = red' or e.g. 2 'Genus = Vicia' AND 'Use = forage'.

Users can define their own report formats using the ALICE Report Writer and can select which datatypes appear, in what order, and in what typeface as well as controlling the inclusion of abbreviated names, bibliographic references etc.

It will be possible to output data subsets in a number of different formats for use with other programs e.g. in XDF, dBASE or DELTA formats.

**Conclusion**

ALICE provides first and foremost a database tool for biologists building and disseminating their own checklists, florulas, or species inventories. ALICE has knowledge of the complex sets of rules that govern taxonomy, nomenclature, geographical hierarchies etc. but also allows users to define their own multistate characters for recording descriptive data.

ALICE does NOT provide for key writing or other specialist identification functions. Other authors have spent much effort in developing very useful programs that already have that functionality and we see no sense in duplicating their effort. There is, inevitably, a slight functional overlap with online identification since database retrieval as supported by ALICE (e.g. 'What plants have red berries, are poisonous and are found in Spain') may resemble a primitive interactive identification sequence. ALICE does not, however, provide any of the many specialist functions (e.g. best diagnostic character) provided in existing identification programs. We thus for-see, and encourage, ALICE users periodically exporting appropriate subsets of their data into DELTA format for use with other people's identification programs.

In the next DELTA Newsletter we will examine the taxonomic and nomenclatural core of ALICE.

*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.*

**Introduction to XPER****Jacques Lebbe and Regine Vignes**

The XPER program is software to create interactive knowledge bases (taxonomic data matrix) and to run Computer Aided Identification (interactive identification)<sup>1</sup>. It was written five years ago specifically for micro- computers. XPER is commercially distributed and continually improved.

It includes four modules:

- the **Editor** to edit, modify, duplicate, and merge taxa, characters and character states. Many useful commands allow for quick creation of a knowledge base.
- the **CAI** module for interactive identification, with free access, possibility to go back in the identification sequence, to correct possible errors, to give more than one state to describe a character (doubt), with advice to select the most efficient character (to discriminate all the remaining taxa or to confirm an hypothesis), with summary of the identification process, deduced characters, reasons of elimination of taxa, closest taxa, etc. (functions similar to ONLINE or INTKEY).

- the **Printer** to present the data in detailed or condensed form on paper.
- the **Reorganizer** to modify the order of taxa or characters, and to merge several knowledge bases.

The software is very easy to use even by a user without any computer experience. XPER (Release 2.2) is commercially available in French and English for any IBM compatible with MS/DOS at DEVNET 6 rue Jules SIMON 92100 Boulogne Billancourt FRANCE. It costs 3000 FF + shipping (about 500 US\$). It is delivered with example and full documentation.

### **The data**

The data managed by XPER are taxonomic or medical descriptions, or any data which form an exhaustive list of descriptions about one particular domain. The characters are only qualitative, with possible dependence relations between characters. Numerical data need to be split in ranges.

### **The applications**

The easy accessibility of the system, the large number of micro-computers for which it was available, and its translation into several languages have all contributed to its wide diffusion. In France it forms part of the software recommended for the Computing for All program run the French National Education System.

Current applications created with XPER include the program Superior Fungi of France<sup>2</sup>, accessible on the French MINITEL<sup>3</sup> system (big public network), and numerous knowledge bases in taxonomy, biology and medicine<sup>4-10</sup>. An application (XPERphlebo), about phlebotomine sandflies (vectors of leishmaniasis diseases) of French Guiana, was produced by the Pasteur Institute with the financial backing of the World and Panamerican Health Organizations<sup>11-12</sup> and is freely distributed to entomologists and epidemiologists interesting by this problem. The extension of this application to all south and central America is in course with collaboration of experts in the different countries. It is sure that the generalization of this type of technique can be of considerable importance in the domains of medicine and veterinary diagnosis.

### **The Toolbox**

Many new programs using XPER knowledge bases were created. They form the XPER TOOLBOX. Two of them allow automatic translation from XPER format to DELTA format (with Pankhurst and Dallwitz particularities) and the contrary. It is very easy to manage a knowledge base with XPER, so if you have already created a DELTA data file you can easily complete, modify, correct, and check it with the editing module of XPER. In the contrary, if you create a knowledge base with XPER, you can use, for example, Pankhurst key maker. The XPER TOOLBOX contains also programs to evaluate the quality of a knowledge base (discriminant power, character correlations, etc.): CHKBASE, to create keys (even key of groups); MAKEY, to extract minimizing set of characters and diagnose; MINSET, to compute distance matrix in the format of Rohlf numerical taxonomic system (NTSYS); DISRXP (distances between taxa or between characters), to automatically convert numerical data from spreadsheet to XPER; NUMTOXP, and specific extended DOS command for XPER; DIRXP, COPYXP, DELXP, etc.

The XPER TOOLBOX is available free of charge to owner of any version of XPER from Jacques LEBBE 4 bis passage LEGRAND 92100 Boulogne Billancourt FRANCE. The documentation is very short in English, but the programs are easy to use. Commercial version with full documentation will be available during 1990.

### **References**

1. J. Lebbe, 1984. Manuel d'utilisation du logiciel XPER. Micro Application, Rueil Malmaison.
2. J. Lebbe, 1986. Les champignons identifiées par ordinateur. Science et vie 829: 126-129

3. J. Lebbe and R. Vignes, 1988. Le système XPER. Implantation télématique et applications. Actes des Journées Systèmes experts et télématique. EC2 et Télésystèmes. 88–97.
4. P. M. Forget, J. Lebbe, H. Puig, R. Vignes and M. Hideux, 1986. Microcomputer-aided identification: an application to trees from French Guiana. *Botanical Journal of the Linnean Society* 93: 205–223.
5. P. Esterre, R. Vignes and J. Lebbe, 1987. Identification assistée par ordinateur: application aux levures pathogènes. *Bulletin de la Société Française de Mycologie Médicale* 16: 117–120.
6. R. Tronche, J. Lebbe and R. Vignes, 1987. Présentation d'un Système expert d'aide au diagnostic des causes de surdit. Démonstration logiciel. Congrès O.P.A., Paris.
7. P. Esterre, R. Vignes and J. Lebbe, 1987. Identification assistée par ordinateur des principaux oeufs d'helminthes en coprologie humaine. *Bulletin de la Société Française de Parasitologie* 5: 241–244.
8. S. Darmoni, R. Vignes, T. Poynard, J. Lebbe and J.C. Chaput, 1988. Diagnostic sérologique de l'hépatite virale aigüe sur MINITEL. Actes des Journées Systèmes experts et télématique. EC2 et Télésystèmes.
9. J. Lebbe, S. Nilsson, J. Praglowski, R. Vignes and M. Hideux, 1988. The morphology of airborne pollen grains and spores from northern Europe in relation to allergenic function: a microcomputer-aided identification. In S. Blackmore and I.K. Ferguson (organisers): "Pollen and spores form and function" Symposium (Poster and exhibition, March 1985). *Grana* 26: 223–229.
10. J. Lebbe, R. Vignes and M. Hideux, 1988. Base de connaissances sur les pollens et spores atmosphériques de l'Europe du Nord et tempérée: identification et applications taxonomiques. *Annales des Sciences Naturelles, Botanique* 107–123.
11. J. Lebbe R. Vignes and J. P. Dedet, 1987. Identification assistée par ordinateur des phlébotomes de la Guyane Française. Editions Institut Pasteur de la Guyane Française. 165 p.
12. J. Lebbe, R. Vignes and J. P. Dedet, 1989. Computer Aided Identification of insect vectors. *Parasitology Today* 5: 301–304.

*Jacques Lebbe, 4 bis passage Legrand, 92100 Boulogne Billancourt, France.*

## **Comments on PANKEY**

**Richard Pankhurst**

### **MacPankey – Macintosh PANKEY programs**

Since announcing the Macintosh versions of the PANKEY programs in the last newsletter, I have not received a single request. Where are all those people who were pressing me to do versions for Mac? MacPankey programs are SHAREWARE. This is to say, that you can have the programs for just for the asking, but at the same time you can optionally contribute £30 or \$50 towards what I personally spent on software in order to develop them.

### **DELTA editor – DEDIT**

A preliminary MSDOS version of the DELTA editor is now available. Please send formatted 5 1/4 inch (one high density 1.2 Mb or two 360K) or 3 1/2 inch (one at 720 K or 1.2MB) floppy discs.

It is \*\*\*\*\*FREE\*\*\*\*\*

It does not do everything that it is planned to do, nor will I claim that it is the last word in user-friendly software (yet) but it does do the following:

- It operates in full screen mode with windows, menus and online help, and uses colour (EGA or VGA) if you've got it.
- It reads as input DELTA data files in full version 3 format.
- It outputs DELTA data files as text, but with all the abbreviations expanded i.e implicit characters are made explicit, variable items are output in full, and all character number contractions are expanded.

- It outputs DELTA files in binary form for use with the new versions of the PANKEY interactive programs, see below.
- It will reorder and/or delete characters and renumber them, like TIDY.
- It acts as a data capture program, like PREP/COND4 already does i.e. it will interactively prompt you for unscored characters, for new or existing items and code the data, and will print reports of unscored characters.
- In addition, it will let you rescore characters which are already scored, and deals correctly with changes in character dependency as it does so.
- SPECIAL STORAGE is implemented either on regular disk, RAM disk (on memory expansion card) or in EXPANDED memory (whichever you have). This saves space for a large matrix.

Important features which are being tested or are planned, are:

- online definition of new characters and states
- online editing of existing characters and slates, especially reordering and inserting/deleting states over existing item descriptions
- copying and combining item descriptions e.g. to make a species description from a set of specimen descriptions

\*\*\*\*IMPORTANT\*\*\*\*

PANKEY users have for some time been asking me to update PANKEY programs to accept DELTA version 3 data format instead of version 2 format as at present. DEDIT gets you over this problem. Read your version 3 file into DEDIT and output it again, and it will then be in version 2 format and readable by PANKEY programs.

### **New versions of PANKEY programs**

New versions of ONLIN6 (the one with graphics) and KCONI are now available. Apart from various corrections and improvements the important difference is that they have been translated into C and reprogrammed with dynamic memory allocation. This means that you no longer have to have different versions for machines with different RAM memory sizes, and that you would not have to have a recompiled version if the internal arrays are not set large enough for a large data set. The programs adapt themselves to the amount of memory available in your machine at the time. Since the compiler is quite different, this also means that the intermediate (binary) data files are different, and INCOMPATIBLE with the old ones.

Users who have licensed versions of ONLIN6 and/or KCONI please return their original distribution disks to me, and I will send the new versions FREE. Please note that you will also need to have DEDIT (see above), in order to prepare the binary data files.

If you have versions of other PANKEY programs which are more than one year old, it is time that you had an update. Return to me the original distribution disks and I will send new versions. It would be a considerable help to me if you could also send me a suitable pre-addressed packet for the return mail.

### **An unadvertised feature of DELTA?**

In an article in DELTA Newsletter 3, I listed the differences between DELTA version 2 and DELTA version 3. There is one that I missed, and thanks are due to Robert Webster for pointing it out. It does not seem to be in the manual. When giving characters in ITEM DESCRIPTIONS it is possible to run two or more consecutive characters together, provided they have the same states e.g.

5,2 6,2 7,2

can be written as

5-7,2

Presumably it doesn't matter if the characters are of different types.

## The illustrated key to sedges

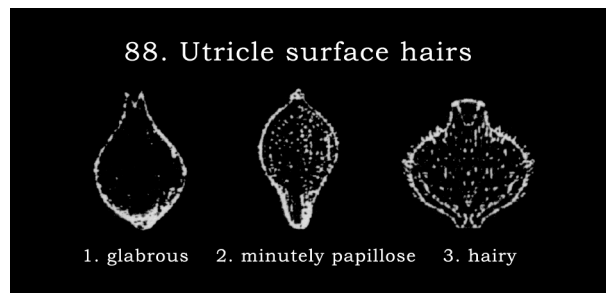
This key makes an important precedent. It is the first DELTA application that I have participated in where the taxonomic expert involved is not a DELTA partisan, or even a computer user. Arthur Chater is an acknowledged expert in the genus *Carex* (sedges), and his unqualified support for this project has made the key greatly more acceptable in the eyes of skeptical users. There were three particular advantages:

- The choice of characters and their states and the scoring of the items were all made with expert advice, and the images which go with the characters were also criticized by the expert.
- The expert provided subjective character weighting which are incorporated in the program and data.
- There exists an independent and well-established illustrated publication on the group which complements the computer key. The DELTA descriptions are in exact correspondence with the text of the book. In particular, we were able to use a scanner on the illustrations to prepare the computer images more easily.

The *Carex* key uses a modified version of ONLIN6. There are 75 species and 97 characters of which 56 have computer images. The images are monochrome, and there are versions for both EGA colour and Hercules monochrome monitors. The standard ONLIN6 program has been altered slightly for this key. The new features are:

- character weighting. The results of BEST and DIAG commands are ordered according to 3 levels of subjective weighting (good, fair and poor). No characters are hidden if they happen to be assigned a low weight. I am generally skeptical about expert assessment of character weights, as experience shows that the useful practical characters are often different from the taxonomically 'good' characters.
- in addition to the usual help screens, each character can have its own help screen. In most cases, the character is presented as an image and by pressing a function key, you can toggle between help in words and help as a picture.
- searching for strings in CHAR and TAXON commands. You can search for particular characters or taxa by typing part of the name strings e.g  
CHAR = hair  
gets you a list of characters involving hair.

The illustration shows the image for hairy fruit.



The keys to Angiosperm families (Hansen and Rahn) and to monocot families (Rao and Pankhurst) are being enhanced with image data in a similar fashion.

JERMY, A. C, CHATER, A. O. & DAVID, R. W. (1982). *Sedges of the British Isles*. BSBI Handbook No. 1, BSBI, London.

PANKHURST R. J. & CHATER A. O. (1989). *Sedges of the British Isles*. BSBI Computer Key No. 1, version 2 (illustrated). 40pp and 2 floppy disks.

Both available from BSBI Publications, 24, Glapthorne Road, Oundle, Peterborough, PE8 4JQ, UK.

## A general purpose floristic database system

Some of you may have seen my paper in TAXON 37: 733. about a database for floristics, which I am applying to Rosaceae for the Flora Mesoamericana. I have now translated this from dBASE III (of

which I hold a rather low opinion) into Advanced Revelation (which is heaps better). The original database included data on nomenclature, bibliography, geographical distribution and specimen labels. To this has been added DELTA data capture (with the help of David French at the BM). Quite a bit of work needs to be done on it yet, but when we have the bare bones of a complete database for monographic work. Notice that this AREV system overlaps with some of what DEDIT does, and also overlaps with the ALICE software (although that does not cater for specimens).

Advanced Revelation (AREV) is a powerful and sophisticated database system based on PICK. It has the great virtue of working with variable length data strings and also allows multi-valued fields. It has an excellent user interface; windows and menus and online help. It is programmed in R/BASIC; a version of BASIC for databases. I'm not so keen on that part of it. AREV has been adopted by both the US and the British Nature Conservancy organizations for databases of plant and animal records, and by a chain of botanic gardens in the US for garden records (Kerry Walter).

*R. Pankhurst, Dept. of Botany, British Museum (NH), Cromell Road, London SW7 5BD, England.*

## A description of the ASKATAXA interactive key program

**Fred Rhoades**

ASKATAXA is the synoptic-key driver in the PC-TAXON package for IBM computers distributed by COMPress. The program may be distributed with keys created in the PC-TAXON format.

Background on the PC-TAXON package, including the key editor, was given in the first article in this series in DELTA NEWSLETTER #3. PC-TAXON is available for \$75.00 from COMPress, Queue. Inc., 338 Commercial Avenue, Fairfield, CT 06430.

The user interface with ASKATAXA is designed to be useful in field or laboratory identification situations with a minimum of computer knowhow. I use ASKATAXA to introduce students to computerized keys in a course covering 'Lower Organisms'. Students use a self-booting disk including ASKATAXA and 18 keys to genera and species of mushrooms, lichens, molds and other lower organisms. The AUTOEXEC.BAT file on the disk starts ASKATAXA and loads the mushroom key into programs memory from the same drive as the program. Student users can easily load other keys while running the program (choice '5' from the main menu — below).

A PC-TAXON key file includes the details of each key: the terms used for the taxonomic level, the names of individual taxa, the phrases describing characters and character states and the data matrix in binary form. In addition, the key file stores locations in an optional information file with additional descriptive text about the individual taxa.

When the program begins, the main ASKATAXA menu is displayed:

```

<< ASKATAXA - ACCESS SYNOPTIC KEYS >>

Presently loaded:
MUSHROOMS OF WHATCOM COUNTY: GENERA OF GILLED MUSHROOMS

1 Identify a mushroom genus
2 Compare two mushroom genera
3 Describe a mushroom genus
4 List Key to printer or file
5 Load a new Key
← End the program

Press NUMBER or ←
Response:

```

[The ASKATAXA screens were represented as text in the original version of the Newsletter, and lacked the symbol for the 'Enter' key. The screen images used here were captured from the version of the key available at <http://www.mobot.org/plantscience/ResBot/FO/FloraOnline.htm>.]

Activities are chosen by pressing the numbered key:

'1' The identification module is described in detail below.

'2' Descriptions of individual taxa can include both the character-states described for each taxon and a descriptive paragraph from the supplemental file. In the mushroom key, these paragraphs discuss the diagnostic features, habitats and common species for each genus.

'3' In a comparison of taxa, a list of the similarities and differences between the character-states describing the two taxa is presented character by character. When two taxa are particularly well separated by a character, this fact is high-lighted.

'4' The user can list a 'field-ready' synoptic key to a printer or disk file, or list parts of a key (Characters, taxa or descriptions of taxa). The complete key can include the contents of a preface file (produced with a word processor) with introductory material, glossary, references, etc. Default printer routines use Epson printers. Other printers can be initialize into the program. By printing to a disk file, a user can modify the key with a word processor and add longer phrases descriptions, illustrations, etc.

### Identification searches using ASKATAXA

To begin an identification search, the user presses '1'. The following screen appears, listing the characters which describe taxa.

```

MUSHROOMS OF WHATCOM COUNTY: GENERA OF GILLED MUSHROOMS

mushroom genera are described by States of these Characters:

 1 Sporeprint color                16 Color of unbruised, fresh flesh
 2 Lamella (gill) attachment       17 Stains or color-changes
 3 Annulus (ring) or stipe surface  18 Smell or taste
 4 Stipe (stem) base or volva      19 Presence of milk or juice
 5 Pileus (cap) shape              20 Spore ornamentation
 6 Habitat (general type of substrate) 21 Spore amyloid reaction
 7 Habit (how related to substrate) 22 Cystidia
 8 Overall texture (cap/stipe flesh) 23 Stature type (see Largent, 1978)
 9 Pileus (cap) surface texture
10 Pileus (cap) margin
11 Pileus (cap) diameter
12 Gill closeness
13 Gill surface or texture
14 Gill cross-section
15 Gill edge or margin

Error Allowance = 0  Auto Listing = OFF
62 mushroom genera remain. Press NUMBER, "A", "C", "D", "E", "T", "U" or "Q"

Response for the First Step: _

```

The INPUT window below the list indicates the condition of the search options and the activities available. These are selected by number or letter: a more complete description of these activities is obtained by pressing the context-sensitive help key (Alt-H) which would display the following:

```

Press a NUMBER to evaluate a Character in the unknown mushroom genus.
Characters used at least once are in red. Press: ← - review, A -
toggle automatic listing of mushroom genera and used States, B - back up one
step, C - find a good Character (best separates remaining mushroom genera),
D - describe a remaining mushroom genus, E - change the Error Allowance,
Q - quit, T - list mushroom genera and U - list used States.

```

During a search, ASKATAXA maintains a list of the taxa described by the previous choices in the search which can be listed automatically each step or by choice. Some of the activities allowed during a search are explained more fully below.

'**CHARACTER NUMBER**' Once a character is chosen, its states are listed and an indication is given in the INPUT window of how well the remaining taxa are separated by the character (whether remaining taxa are each described by one or several unused states and whether there is any overlap in the description). [The following screen was not in the original version of the Newsletter.]

```

MUSHROOMS OF WHATCOM COUNTY: GENERA OF GILLED MUSHROOMS

mushroom genera can have these States of Stipe (stem) base or volva:

 1 Stipe lacking (6)
 2 Stipe lateral/off center (8)
 3 No volva, base even (59)
 4 No volva, base tapered (31)
 5 No volva, base club-shaped (18)
 6 No volva, base abruptly bulbous (4)
 7 No volva, base deeply rooting (3)
 8 Base strigose/tomentose (hairy) (13)
 9 Volva saccate (deep cup) (2)
10 Volva scaly (uneven scales) (1)
11 Volva annular (concentric rings) (2)

Press ←, NUMBER, "N", "A" or "O"
Which State does the mushroom genus have?

Some remaining mushroom genera are each described by several unused States
and the mushroom genera overlap in their descriptions.

```

While viewing the list of character-states the user can:

- Press the **'State NUMBER'** to choose a single character-state. Press **'N'** (NOT) and then the **'State NUMBER'** of a state to cause the search to include taxa which DON'T exhibit that state.
- Press **'O'** (OR) and several **'State NUMBERS'** to have ASKATAXA include taxa which exhibit at least one of the states in the range of states entered.
- Press **'A'** (AND) and several **'State NUMBERS'** to have ASKATAXA include taxa which exhibit every one of the states in the range of states entered.

**'C'** Finding a 'good character' suggests the best, unused character which separated the remaining taxa. The algorithm was arrived at empirically to use the format of the memory image (sets of taxa showing each character-state). It is quite fast and almost always picks the character an expert identifier would use when considering the state of an identification (remaining taxa and unused characters).

**'D'** Describing a taxon works like the 'Describe' feature from ASKATAXA's main menu, but the user selects from the list of taxa currently remaining in the search. This is handy as the user narrows in on a group of similar taxa.

**'E'** Setting an error allowance (at Step 1) enables a search to allow (up to 9) errors in character-state choice and still include a taxon in the search. If errors are allowed, the listing of taxa indicates how many errors have been made for each taxon remaining. Using this feature with 'OR' state range selections provides considerable flexibility in searches and allows the identification procedure to be especially forgiving.

Once a taxon is identified, ASKATAXA will review the list of character-states used to make the identification. If the error allowance feature or the 'OR' state range entry have been used, the list of used characters will high-light the states actually displayed by the identified taxon. The user can also choose to review a complete description of the taxon, including the supplemental descriptive information.

Some of the more 'friendly' aspects of ASKATAXA are:

1. Choosing activities with minimum key pressing from menu where possible.
2. Context-sensitive help at all program locations.
3. Ease of recovery from user mistakes with informative error messages.
4. Small program size allows both program and several keys to comfortably fit on one disk.
5. Inclusion of identification, comparison, description and key-printing utilities in one program.
6. Ability to print keys to a disk file.

If you are unfamiliar with PC-TAXON and would like to experiment with the approach, sets of keys along with ASKATAXA are available to fungi and lichens/bryophytes. Programs are also available to translate DELTA format to/from PC-TAXON format. If you have a modem, the keys, key driver and conversion programs are available for downloading 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 (Issue 11: lichen/bryophyte keys; Issue 14: keys to fungi; Issue 21: 'PC-TAXON - DELTA data format conversion utilities') Keys and conversion programs are also available from me for \$5.00 each set. Information on PC-TAXON will be included with each request. [Flora Online is now available at <http://www.mobot.org/plantscience/ResBot/FO/FloraOnline.htm>.]

One of the disadvantages of distributing the programs through COMPRESS has been my isolation from users. I find myself not knowing who is using PC-TAXON to do what. I am trying to build a library of the location of keys developed with PC-TAXON. COMPRESS will include the library listing with the programs when sold and publish it in their newsletter. Please let me know of any keys you have to share with other users. For each key, indicate, 1) author, 2) taxonomic group, 3) key title, 4) number of taxa, 5) .KEY size, 6) .DES size, and 7) a description of for whom the key is written or for what applications it would be useful. All contributors will receive the library listing and be put on a mailing list for periodic updates.

*Fred Rhoades, Biology Department, Western Washington University, Bellingham, WA 98225, USA.*

## **Diagnostic descriptions for groups of taxa**

**Mike Dallwitz**

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

An article in DELTA Newsletter No. 3 showed how to use INTKEY and CONFOR to generate diagnostic descriptions. These descriptions were for taxa or 'items' at the taxonomic level at which the original descriptions were recorded (in DELTA format). The current version of INTKEY cannot automatically generate diagnostic descriptions for *groups* of taxa, but it can greatly assist in this process.

The previous article contained basic information that will not be repeated here, so you should read it first. (Reprints are available from the author, and back issues of the Newsletter from the Editor.) Note that if you try to reproduce the results given in the first article, you may find some minor discrepancies, depending on what version of INTKEY you have.

In the following INTKEY dialogues, the user's responses are shown in a bold face, and explanatory comments are shown in italics. As in the first article, the data used in the examples are the sample set which is provided with our programs.

The taxa in the sample data are genera. Suppose that we want to generate diagnostic descriptions for subfamilies. The first step is to define appropriate taxon keywords so that we can conveniently refer to the groups of taxa involved. A taxon keyword represents a set of taxon numbers, and can be directly defined as such (e.g. DEFINE TAXA "Pooideae" 1 9 13). However, this is not usually a good method, because the numbers may change if new taxa are added to the data. It is better to define the keyword in terms of taxon names (e.g. DEFINE NAMES "Pooideae" Agrostis Festuca Poa), or, where possible, by making use of information coded in the data. In the sample data, there is a character giving the subfamily, so we can use this as shown below.

INTKEY version: 11-AUG-89.

M.J. Dallwitz and T.A. Paine, CSIRO Division of Entomology.

GPO Box 1700, Canberra ACT 2601, Australia  
Fax +61 6 246 4000. Internet [delta@ento.csiro.au](mailto:delta@ento.csiro.au)

Sample Grass Genera 12:14 19-SEP-89

Enter command or HELP INTRODUCTION: **74**    *Use character 74*

74: <subfamily: Watson et al. 1985>

1. Pooideae
2. Bambusoideae
3. Arundinoideae
4. Chloridoideae
5. Panicoideae

Enter value: **1**

3 taxa remain

1. Agrostis
9. Festuca
13. Poa

Enter command: **define taxa "pooideae" remaining**    *Define the keywords*

Enter command: **define taxa "not pooideae" eliminated**

Enter command: **taxa po**    *Check the keywords*

1. Agrostis
9. Festuca
13. Poa

Enter command: **taxa notpo**

- (1) 2. Andropogon
- (1) 3. Anisopogon
- (1) 4. Bambusa
- (1) 5. Chloris
- (1) 6. Cynodon
- (1) 7. Echinochloa
- (1) 8. Eleusine
- (1) 10. Oryza
- (1) 11. Panicum
- (1) 12. Phragmites
- (1) 14. Zea

If we want to use the keywords in several sessions, it may save time if we save the lists of numbers (by using SAVE TAXA) and edit them into appropriate DEFINE commands for later use. Better still, we can use a text editor make up a command file, SUBFGET.INP, which will generate and save all of the necessary information. This command file can be used again if new taxa are added.

```

output on subfdef.inp
output off
restart
remark define t Pooideae
74,1
save t rem
remark define t not Pooideae
save t elim
restart
remark define t Bambusoideae
74,2
save t rem
remark define t not Bambusoideae
save t elim
. . .
restart

```

```
remark define t Panicoideae
74,5
save t rem
remark define t not Panicoideae
save t elim
```

We execute this command file in INTKEY by entering INPUT SUBFGET.INP. This produces the file SUBFDEF.INP:

```
define t Pooideae
1 9 13
define t not Pooideae
2-8 10-12 14
define t Bambusoideae
4 10
define t not Bambusoideae
1-3 5-9 11-14
. . .
define t Panicoideae
2 7 11 14
define t not Panicoideae
1 3-6 8-10 12-13
```

This requires a little editing to get it into the correct format for input to INTKEY:

```
define t "Pooideae" 1 9 13
define t "not Pooideae" 2-8 10-12 14
define t "Bambusoideae" 4 10
define t "not Bambusoideae" 1-3 5-9 11-14
define t "Arundinoideae" 3 12
define t "not Arundinoideae" 1-2 4-11 13-14
define t "Chloridoideae" 5-6 8
define t "not Chloridoideae" 1-4 7 9-14
define t "Panicoideae" 2 7 11 14
define t "not Panicoideae" 1 3-6 8-10 12-13
```

Whenever we need these definitions in INTKEY, we can invoke them by entering INPUT SUBFDEF.INP.

The most obvious way of generating diagnostic descriptions of the groups is to first make appropriate new DELTA-format descriptions of the groups, which can then be treated as single 'taxa', as described in the previous article. The group descriptions can be generated by means of the SAVE SUMMARY command in INTKEY, as follows.

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

All output to file.

Enter command: **out off**

Output to file halted.

Enter command: **display nooutput**     *Suppress screen display of file output*

Output not displayed.

Enter command: **input subfdef.inp**

Enter command: **keywords taxa**     *Display taxon keywords*

TAXON KEYWORDS

Remaining

Eliminated

ALL

POoideae  
 NOT POoideae  
 Bambusoideae  
 NOT Bambusoideae  
 ARundinoideae  
 NOT Arundinoideae  
 Chloridoideae  
 NOT Chloridoideae  
 PANicoideae  
 NOT PANicoideae

Enter command: **rem # Pooideae/**

Enter command: **save summary**

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

Enter character ranges or keywords: **all**

Enter command: **rem # Bambusoideae/**

Enter command: **save sum ba all**

...

Enter command: **rem # Panicoideae/**

Enter command: **save sum pa all**

The output file, SUMMARY.TMP, contains valid DELTA-format items (it does not require editing):

```
#Pooideae/
2,1<2>/2 3,2-84.83-200 4,2 5,2 6,2<2/2> 7,1/2<2> 8,2 9,2 10,1 11,1/2<1> 12,2/3<1> 13,5 14,1/2
15,2 16,2 19,2 25,0.8-6.8-20 26,1 27,1 28,2<2/2> 29,1/2<1> 30,1 31,2 32,1<1>/2<2>
33,1<2>/2<1> 34,2 35,1/2<2> 36,1<2>/2<2> 37,2<2/2> 41,1<3>/2-14<2>/15<1> 42,1/2<2>
43,1<2/2>/2<1/2> 44,1/2<2>/3<2> 45,1<2/2>/3<1/2>/5<1/2> 46,1<1/2>/2<1/2>/3<1/2>
47,1<1/2>/2<1/2> 48,1<2/2>/2<2/2>/3<1/2> 49,2<1/1> 50,3-4<2>/5<3>/6<1>/7<2>/8-11<1>
51,1 52,1/2<2>/3<2> 53,1 54,2 55,1<1>/2 56,0<1>/3<3> 57,1/2<1> 58,2<3> 59,1
60,1/2<1>/3<1> 61,2 62,1<2>/2<1> 63,2 64,2 66,1 67,2 68,2 69,1/2<2> 70,1 71,2 72,1/2 73,2
74,1 75,2 76,4<1>/7<2> 82,170-223-300 83,1/2/3/4<2>/5/6

# Bambusoideae/
2,1<1>/2 3,30-1008-3500 4,1<1>/2<1> 5,1<1/1> 6,2 8,1/2 9,1/2<1> 10,1 11,1/2<1> 12,2 13,5
14,1 15,1<1>/2<1> 16,2 19,2 25,4-26.5-80 26,1/2<1> 27,1 28,2<1/1> 29,1<1>/2<1> 30,1/2<1>
31,2 32,2 33,1
...
```

We can add these items to the bottom of the original items file: they will be items 15-19. We also need to change the MAXIMUM NUMBER OF ITEMS in the SPECS file before running CONFOR TOINT to convert the data to INTKEY format. The lines

```
define t "genera" 1-14
define t "subfamilies" 15-19
```

should be added to INTKEY.INI, so that we can conveniently restrict our attention to the data for the genera or for the subfamilies.

We can now generate the diagnostic descriptions for the subfamilies.

```
Enter command: set tolerance 0
TOLERANCE set to 0.
```

```
Enter command: set stopbest 40
STOPBEST set to 40.
```

```
Enter command: exclude characters nomenclature classification geography
```

11 characters excluded.

Enter command: **include taxa subfamilies**

5 taxa included.

Enter command: **diagnose all**

Enter preset characters: *None, so just press 'Enter'*

15. Pooideae

27: spikelets

1. disarticulating above the glumes

5: culms

2. unbranched above

13: inflorescence

5. paniculate

Diagnosis for taxon 15 is incomplete.

16. Bambusoideae

56: stamens

6

17. Arundinoideae

7: leaf blades

2. linear-lanceolate

3. lanceolate

56: stamens

3

Diagnosis for taxon 17 is incomplete.

18. Chloridoideae

13: inflorescence

2. of spike-like main branches

Diagnosis for taxon 18 is incomplete.

19. Panicoideae

27: spikelets

2. falling with the glumes

3. not disarticulating

Diagnosis for taxon 19 is incomplete.

We have obtained a complete diagnosis for only one of the subfamilies (16. Bambusoideae) at tolerance 0, although the data for *genera* allow complete diagnoses of all the genera at tolerance 0, and of all but 4 of them at tolerance 1. (Furthermore, these 4 failures are between pairs in the same subfamilies: *Agrostis–Poa* and *Echinochloa–Panicum*.) The failures to diagnose the subfamilies are because information was lost when the descriptions of the genera were combined into the descriptions of the subfamilies.

A simple example will clarify the way in which information is lost. Suppose that we have taxa A and B, with DELTA descriptions #A/ 1,1 2,1 3,1 and #B/ 1,1 2,2 3,2. The group description for A&B, produced as above, is #A&B/ 1,1 2,1/2 3,1/2. However, taxa C and D, with descriptions #C/ 1,1 2,1 3,2 and #D/ 1,1 2,2 3,1, would also conform to this group description; so, in using this description to represent the group, we have lost sight of the fact that C and D do not exist, or, at least, do not belong to the group. What we really need is a description like #A&B/ 1,1 {2,1 3,1}/{2,2 3,2}, which specifies only those combinations of attributes which actually occur in the group. (Note that this is not valid DELTA syntax. The 'variant item' facility of CONFOR can be used to produce a similar effect, but KEY is the only other program that can make full use of the information.)

However, there is a fairly simple way to improve on the diagnoses obtained above. Instead of trying to separate each subfamily from the other subfamilies, we try to separate each subfamily from the *genera* of the other subfamilies. This approach gives complete diagnoses for the subfamilies at tolerance 0. At tolerance 1, as shown below, one diagnosis is incomplete.

Enter command: **set tolerance 1**

TOLERANCE set to 1.

Enter command: **include t 15 notpo**     *Include the subfamily description of the Pooideae, 12 taxa included plus the genera not belonging to the Pooideae*

Enter command: **diag 15**

Enter preset characters:

- 15. Pooideae
  - 54: lodicules
    - 2. membranous
  - 70: midrib
    - 1. with one bundle only
  - 63: pericarp
    - 2. fused
  - 13: inflorescence
    - 5. paniculate
  - 26: spikelets
    - 1. compressed laterally
  - 56: stamens
    - 0/3
  - 25: spikelets
    - 0.8–20 mm long

Enter command: **inc t 16 notba**

13 taxa included.

Enter command: **diag 16**

Enter preset characters:

- 16. Bambusoideae
  - 56: stamens
    - 6
  - 70: midrib
    - 3. with complex vascularization

Enter command: **inc t 17 notar**

13 taxa included.

Enter command: **diag 17**

Enter preset characters:

- 17. Arundinoideae
  - 25: spikelets
    - 9–60 mm long
  - 29: spikelets
    - 1. with the rachilla prolonged apically
  - 35: glumes
    - 2. not carinate
  - 11: adaxial ligule
    - 2. a fringed membrane
    - 3. a fringe of hairs
  - 7: leaf blades
    - 2. linear-lanceolate
    - 3. lanceolate
  - 50: lemmas
    - 1–3 nerved
  - 12: plants
    - 2. bisexual, with bisexual spikelets

Diagnosis for taxon 17 is incomplete.

Enter command: **inc t 18 notch**

12 taxa included.

Enter command: **diag 18**

Enter preset characters:

18. Chloridoideae

13: inflorescence

2. of spike-like main branches

65: biochemical type

1. PCK

2. NAD-ME

64:

1. C4

26: spikelets

1. compressed laterally

27: spikelets

1. disarticulating above the glumes

3. not disarticulating

Enter command: **inc t 19 notpa**

11 taxa included.

Enter command: **diag 19**

Enter preset characters:

19. Panicoideae

27: spikelets

2. falling with the glumes

3. not disarticulating

29: spikelets

2. without an apically prolonged rachilla

36: spikelets

2. with incomplete florets

38: proximal incomplete florets

1

41: female-fertile florets

1

18: spikelets

2. in pairs

To see where the diagnosis for the Arundinoideae is incomplete, we can enter the diagnostic description in an 'identification'.

Enter command: **inc t 17 notar**

13 taxa included.

Enter command: **match iuo**

MATCH set to Inapplicables, Unknowns, Overlap.

Enter command: **25,9-60 29,1 35,2 11,2/3 7,2/3 50,1-3 12,2**

Also setting –

30: glumes

1. present

Also setting –

10: adaxial ligule

1. present

2 taxa remain

(1) 9. Festuca

(0) 17. Arundinoideae

Thus, we see that *Festuca* is the genus that could not be separated from the Arundinoideae. We can find the difference between these taxa as follows.

Enter command: **diff (9 17)**  
 Enter character ranges or keywords: *The default is 'all'*  
 11: adaxial ligule <form – avoid seedlings>  
   9. Festuca  
     1. an unfringed membrane  
   17. Arundinoideae  
     2. a fringed membrane  
     3. a fringe of hairs  
 1 difference between taxa.

Note that the 'match' setting is IUO (Inapplicables, Unknowns, Overlap), which is required for identification and diagnosis. The 'overlap' setting means that taxa are not regarded as differing in a given character if they have any state values in common. With 'match' set to IU, there are 30 differences. Alternatively, we can leave 'match' at IUO, and look at the differences between *Festuca* and the *genera* of the Arundinoideae.

Enter command: **include t genera**  
 14 taxa included.

New identification started.

Enter command: **diff (3 9) all**  
 11: adaxial ligule <form – avoid seedlings>  
   3. Anisopogon  
     2. a fringed membrane  
   9. Festuca  
     1. an unfringed membrane  
 25: <female-fertile> spikelets <approximate length, excluding any awns: data unreliable for large genera>  
   3. Anisopogon  
     40–60 mm long  
   9. Festuca  
     3–20 mm long  
 32: glumes <of female-fertile spikelets, whether markedly unequal in the intact spikelet; regardless of any differences in form>  
   3. Anisopogon  
     2. more or less equal  
   9. Festuca  
     1. very unequal  
 33: glumes <of female-fertile spikelets, lengths relative to proximal (adjacent) lemmas; refers to the longer glume when glumes unequal>  
   3. Anisopogon  
     2. long relative to the adjacent lemmas  
   9. Festuca  
     1. decidedly shorter than the adjacent lemmas  
 45: awns <of female-fertile lemmas, number>  
   3. Anisopogon  
     3  
   9. Festuca  
     1  
     (when applicable)  
 47: awns <of female-fertile lemmas, whether straight or geniculate when dry>  
   3. Anisopogon  
     2. geniculate

- 9. Festuca
  - 1. non-geniculate  
(when applicable)
- 48: awns <main, median of the female-fertile lemmas, relative length>
  - 3. Anisopogon
    - 3. much longer than the body of the lemma
  - 9. Festuca
    - 1. much shorter than the body of the lemma
    - 2. about as long as the body of the lemma  
(when applicable)
- 63: pericarp <whether fused or loose (or free)>
  - 3. Anisopogon
    - 1. loose
  - 9. Festuca
    - 2. fused
- 8 differences between taxa.
- Enter command: **diff (12 9) all**
- 8: leaf blades <whether broad or narrow (specify the true range)>
  - 9. Festuca
    - 2. narrow
  - 12. Phragmites
    - 1. broad
- 11: adaxial ligule <form – avoid seedlings>
  - 9. Festuca
    - 1. an unfringed membrane
  - 12. Phragmites
    - 3. a fringe of hairs
- 37: the incomplete <male or sterile> florets <position in spikelet>
  - 9. Festuca
    - 2. distal to the female-fertile florets
  - 12. Phragmites
    - 3. both distal and proximal to the female-fertile florets
- 52: palea <female-fertile, relative size>
  - 9. Festuca
    - 1. relatively long
  - 12. Phragmites
    - 2. conspicuous but relatively short
- 54: lodicules <of female-fertile florets, texture>
  - 9. Festuca
    - 2. membranous
  - 12. Phragmites
    - 1. fleshy
- 59: stigmas <colour, in chasmogamous spikelets>
  - 9. Festuca
    - 1. white
  - 12. Phragmites
    - 3. brown
- 62: hilum <form>
  - 9. Festuca
    - 2. long-linear
  - 12. Phragmites
    - 1. short
- 67: <leaf blade> mesophyll <presence of arm cells (= ‘ratchet’ cells)>
  - 9. Festuca

- 2. without arm cells
  - 12. Phragmites
    - 1. with arm cells
  - 70: midrib <of the mid leaf blade, vascularization>
    - 9. Festuca
      - 1. with one bundle only
    - 12. Phragmites
      - 2. with a conventional arc of bundles
- 9 differences between taxa.

Clearly, there are plenty of ways of distinguishing the *members* of the Arundinoideae from *Festuca*. The difficulty in diagnosing the group arises from the loss of information in the group description. To get around this, we can attempt to find a diagnostic character set by diagnosing the individual genera, and accumulating the characters used. We can start from an empty character set, or from any set that seems suitable. In the example below, we start from the first 4 characters found by the group diagnosis.

Enter command: **define c "d17" 25 29 35 11**     *Define the starting set*

Enter command: **include t 3 notar**  
13 taxa included.

Enter command: **diag 3**

Enter preset characters: **d17**

- 3. Anisopogon <R.Br.>
  - 11: adaxial ligule
    - 2. a fringed membrane
  - 25: spikelets
    - 40–60 mm long
  - 29: spikelets
    - 1. with the rachilla prolonged apically
  - 35: glumes
    - 2. not carinate
  - 63: pericarp
    - 1. loose

Enter command: **define c "d3" d17 63**     *Add the extra character found*

Enter command: **include t 12 notar**  
13 taxa included.

Enter command: **diag 12 d3**

- 12. Phragmites <Adans.>
  - 11: adaxial ligule
    - 3. a fringe of hairs
  - 25: spikelets
    - 9–16 mm long
  - 29: spikelets
    - 1. with the rachilla prolonged apically
  - 35: glumes
    - 2. not carinate
  - 63: pericarp
    - 2. fused
  - 37: the incomplete florets
    - 3. both distal and proximal to the female-fertile florets

Enter command: **define c "d12" d3 37**     *Add the extra character found*

We now have a diagnostic character set containing 6 characters. This set is capable of distinguishing all the genera belonging to the Arundinoideae from all the genera *not* belonging to the Arundinoideae. The description of the Arundinoideae, as represented by taxon 17, in terms of this set is:

11,2/3 25,9–60 29,1 35,2 37,2/3 63,1/2.

This, of course, is *not* diagnostic (at tolerance 1). We can make a diagnostic description for the Arundinoideae by combining the diagnostic descriptions of the genera, as follows:

{11,2 25,40–60 29,1 35,2 37,2 63,1}/{11,3 25,9–60 29,1 35,2 37,3 63,2}.

We can immediately see that the attributes 29,1 and 35,2 are common to the two parts of the description, and, by feeding in the descriptions as ‘identifications’, we can discover that the attributes 37,2 and 63,2 are redundant. Thus, the description can be simplified a little to:

29,1 35,2 {11,2 25,40–60 63,1}/{11,3 25,9–16 37,3}.

There is no escaping the fact that neat diagnostic descriptions or keys for higher taxa are often not possible. This is because ease of identification is not considered a very important factor in the construction of classifications (nor should it be). Corollaries are: (1) By the time we have found to which high-level taxon a specimen belongs, we may have got part or all of the way towards finding to which lower-level taxon it belongs. (2) It is often more efficient to identify directly to a low-level taxon, rather than by passing through identifications to higher levels.

*M. J. Dallwitz, CSIRO Division of Entomology, G.P.O. Box 1700, Canberra, A.C.T. 2601, Australia. Telephone 61–6–246–4911. Facsimile 61–6–246–4000. Telex AA 62309.*

## Application for DELTA software to studies on Australian grasses

**Bryan Simon**

My involvement in using the DELTA software as an aid to taxonomic research on Australian grasses dates from April 1985 when I attended a DELTA workshop at the Taxonomy Unit of the Research School of Biological Sciences (Australian National University). This was run by Mike Dallwitz and Les Watson and I received funding to attend from the Australian Biological Resources Study. Since then I have been using it constantly in my work on the Flora of Australia account of the tribe Andropogoneae, revisionary work on the genus *Aristida*, and in the compilation of a character set for an projected account of the grasses of Queensland. I have also used the Watson and Dallwitz database of world genera, in addition to my own, to run phenetic and cladistic programs from the DIST and PAUP formats of DELTA. The INTKEY program has also been used both as an on-line tool for identification and as a means of compiling generic descriptions from a number of included species.

Initially the programs had been run on a Pyramid microcomputer after some difficulty had been experienced in getting the software installed. However, all work is now run on an IBM-PC, mainly as computer operation is not dependent on external management and discs are conveniently exchanged by mail. Recently, I received the PREP and COND4 programs of Richard Pankhurst and find I have saved much time in the assembling of data using the menu-driven options of these programs.

Overall I have found the DELTA programs to be of great value in assembling data, particularly in the production of botanical descriptions. Prior to using DELTA, I found the writing of descriptions to be a tedious task at the best of times, whereas the interaction of researcher with computer using Pankhurst's COND4 programs (Pankhurst 1988) results in the operation being more manageable and enjoyable.

Although I have also used the KEY programs many times in my work I have reservations at this stage on relying on them as the only method of key compilation. This perhaps comes from a bias of having compiled numerous grass keys in a more classical way (Simon 1980; in press) and find the DELTA keys producing couplets I was unfamiliar with or difficult to put into practice. Consequently at this stage I have opted to use the DELTA keys as a tool for improving keys I have already generated

intuitively. However, as I gain more experience in the use of the key options I shall be able to discern at the outset those characters better for use in key generation from those for description preparation.

When using DELTA for my *Andropogoneae* work I created separate CHARS files for the species of each genus and another one for the genera. A more recent trend has been to use one CHARS file for both genera and species (Webster 1987; Lazarides, Clayton & Palmer ms.) and then when compiling the ITEMS file to use only the characters applicable to the taxa in question. However, at the time I found the production of descriptions and keys to be more manageable by the smaller CHARS files for the different groups of taxa.

In the gathering of characters for the account of Queensland grasses I intend to use one CHARS file, after conducting a close scrutiny of the character model at world level for species released recently on a limited scale for trial (Lazarides, Clayton & Palmer ms.).

## References

- Lazarides, M., W.D. Clayton & Joanne Palmer (MS). A morphological character model for the Poaceae.
- Pankhurst, R. (1988). Data capture for DELTA. DELTA Newsletter 2: 5.
- Simon, B.K. (1980). A Key to Queensland grasses. Botany Branch, Department of Primary Industries, Technical Bulletin 4.
- Simon, B.K. (in press). A Key to Australian grasses. Queensland Department of Primary Industries.
- Webster, R.D. (1987). The Australian Paniceae (Poaceae). Berlin and Stuttgart: J. Cramer.
- B. K. Simon (Senior Botanist), Queensland Herbarium, Department of Primary Industries, Meiers Road, Indooroopilly 4068, Australia.*

## Software review – TAXADAT

**Richard Spjut**

Although utility software for producing taxonomic descriptions and keys has been made available to the MS DOS personal computer (PC) user within the last few years, this software is still in its infancy with regard to ease-of-learning, perhaps at the same level as the early versions of the popular commercial wordprocessing or database programs. The early versions of WORDSTAR, PERFECT WRITER, DBASE, and PERFECT FILER often required special training for the user who often had to consult technical manuals and memorize numerous combinations of key strokes to execute program functions. MULTIMATE or WANG Wordprocessing became popular because these were easy to learn and use, while other current products such as PCWRITE and VENTURA have become quite sophisticated in their visual displays of help screens and pull-down menus. Since taxonomists do not have large sums of money to invest in developing help menus and other forms of visual documentation, what should the user expect before acquiring one or more of products being promoted as an aid for organizing taxonomic data? TAXADAT was selected for this review, but the selection is not meant to be indicative of the taxonomic software available.

In DELTA Newsletter No. 3, Doyle E. Anderegg briefly described the features of his product-TAXADAT. The author acknowledges Richard Pankhurst for some of his TAXADAT concepts (TAXADAT manual). TAXADAT was tested on an 80386 with 2000K RAM, a 40 MB HardDrive and GWBASIC; however, it reportedly requires a system with only a single floppy drive and 128K RAM. The author indicates that TAXADAT was originally written for the Apple II, and he has received reports of the program running successfully on an Atari 1040 (with PC-DITTO emulator), Tulipocomputer, and Phillips computer. The programs came with 2 tutorial databases, a small one on 'critters' (6 species by 10 characters), and another with 26 coded characters for 156 moss genera. I also produced a database of 10 species by 32 characters, belonging to a genus of lichens.

## Features

TAXADAT is a menu operation program with strict limitations on data entry, editing, keying, and format; it lacks the flexibility and many features that are offered in the DELTA programs. One cannot directly view data files, establish dependent character relationships, insert comments in the data files, utilize numeric characters, or distinguish between character types such as integers, real numerics and text characters. Linear keys generated by TAXADAT will not have strictly comparative alternatives at the dichotomies, nor does it have the options of including or excluding characters in the keys and descriptions. The strongest features of TAXADAT are its fast interactive key and conversion of data to text description.

In using the interactive key of TAXADAT, the character of choice can be selected while viewing the list of characters on screen, and by typing the character number at the prompt, the character states appear from which the user selects the appropriate state by entering its number. After executing this feature, the number of taxa selected is indicated, and the user can view their names by selecting the TAXA LIST option. Characters and states selected in an identification sequence may be displayed, and in the process of selecting new characters, the previous ones used are indicated by an 'x'. A contrast feature allows comparison between taxa, and after eliminating all but one taxon the name appears highlighted and flashing at the top of the screen.

There are relatively few steps involved in generating a natural language taxon description. Characters can be linked to preceding characters in output descriptions by inserting a slash before the character in the character list. Character order can be changed by character weighting as prompted by the program, a confusing procedure in contrast to Dallwitz's CONFOR where the character order is explicitly shown in a directive prior to program execution.

## Performance

The interactive key and conversion to text descriptions performed satisfactorily. I had no problem in keying out the moss genus *Polytrichum* or generating descriptions of species from of my lichen database.

The 'Best Character' option was tested on the moss tutorial database, and 24 of the 26 were considered as 'best', although their order under this option hardly differed from viewing the characters under the 'List' option. Effects of character weightings were not evaluated.

Production of the linear key resulted in many noncomparative statements, and in some dichotomies it appeared that a complete description was used. This may explain the slow execution time of the program. In a test run on 10 taxa from the lichen database, 265 items requiring 85K were generated, where a maximum of 15 items using about 20K of disk space would have been expected. In another run, the output was reduced to 21 items by weighting some of the characters. An acceptable key to the lichen taxa could not be generated; however, a species could be satisfactorily keyed out using the same data in the interactive key program.

When BASIC errors were encountered, program functions were often resumed by pressing the F2 key rather than rebooting the system. This technique is not described in the manual.

## Documentation

The manual for TAXADAT is written clearly, but it is incomplete. Many of the menu or submenu screens are not described, and experimentation was required to understand the application of the various options. For example, under the title screen 'Input/Edit Menu', '1. Manual flag (On/Off)' was the first in a list of nine options. In my first attempt under this option, no change was noticed, but after repeated selection, the words 'Manual Flag Set' appeared and disappeared under the screen title (not the selection title). After more experimentation, I learned that making this selection allows one enter the edit mode for re-entering character text. While some of the options may be self explanatory, e.g., '3. Input Characters', their relative sequence and function in the entire program was not clear, especially in view of the fact that changes can be made by ASCII word-processors or editors.

## Ease of learning and use

Characters and character states have to be entered as separate files (random access text files). When inputting characters through TAXADAT, one cannot return to a previous character or character state without exiting the input mode, which in this case requires going through the entire sequence of entering all characters, or aborting the input operation by rebooting the system.

When editing text via TAXADAT, corrections have to be entered carefully; for example, if you accidentally choose the wrong character number in the edit mode, you have to re-enter the text for the entire character, or all of the character states because a change will be made regardless. On the other hand, TAXADAT offers the alternative of setting up a character list in a ASCII wordprocessing file that can be converted to the TAXADAT format by a TAXADAT utility. There is a BASIC limitation in that the descriptive text for a character or character state cannot exceed 80 characters.

Character states for each item are scored through the 'Input Data Matrix' option. After entering a character number at the prompt, the character states appear; one may then highlight the character states that are not applicable to the character by pressing the space bar. A 'Repeat Evaluation' option can be executed by pressing the left arrow key after moving the cursor down the list of character states.

Checking data for accuracy is best done by conversion of data to a text description through one of the utility features. There are a number of other optional features that require an understanding of the mathematical relationships between decimal equivalents and binary numbers, and a learned ability to differentiate between numbers assigned to taxa, characters, character states and matrix values. Under these alternative features, data are visualized through the data matrix selection by either species or characters. The array of numbers is difficult to translate into meaningful coded taxonomic information. Producing a synoptic key is another option, but the key may be difficult to read in large databases. TAXADAT does not offer conversion to DELTA format, a feature that could be useful.

## Summary

Requirements: 128k, Atari DOS or MS-DOS (no longer available for Apple DOS), BASIC or GWBASIC, \$75 U.S.

Pros: Fast interactive-key program. Data entry menus have potential for development into a Program Editor.

Cons: Incomplete documentation: characters and character states are stored in separate files; limited format for outputting taxonomic descriptions; cannot escape data entry menus when inputting characters and character states; could not produce an acceptable linear key; many options in menus that need re-organization.

*Richard Spjut, Botanist, USDA/ARS/SBML, Bldg. 265, BARC-East, Beltsville, MD 20705, USA.*

## Programmer's reply to review of TAXADAT

### Doyle E. Anderegg

I acknowledge Dr. Spjut's thorough and fair analysis of TAXADAT. In interpreting my responses to his review please be aware that I have not seen DELTA software and therefore may miss the significance of some of his statements.

Dr. Spjut's criticism of the lack of an 'escape' during the re-entry (Editing) of characters and character states is a good one. I will add that feature immediately.

The comment concerning 'linking' is unclear to me, perhaps because I am not familiar with DELTA or it may be that Dr. Spjut misunderstood the program's use of the slash, due to poor documentation. There is no enforced linking of characters in TAXADAT. All characters are independent. Characters

which are not applicable are scored as '-1'. A slash in a character field begins a comment and comments are not printed in the key. Rather, comments help the user select characters for interactive keying. Example of such usage would be 'SPORE/NUMBER', 'SPORE/COLOR', or 'SPORE/ORNAMENTATION'. The comment would be superfluous in the key because of the associated character state.

Dr. Spjut's comment regarding 'BEST CHARACTER' suggests possible poor documentation rather than poor programming. BEST CHARACTER is to be used late in interactive programming when the more obvious characters have been used and there are still too many 'active taxa' to make the COMPARE facility useful. At that time BEST CHARACTER will list those characters/character states that are shared (and scored) for the taxa still under consideration.

Dr. Spjut comments that 'One cannot directly view data files...'. It is true that BASIC random access files cannot be viewed outside the program. However, all of the files may be viewed and/or printed from within the editing program that is used to create them. Also, even though the characters and character states are in separate files the editor shows them and prints them out in a combined report. Perhaps my failure to understand the comment is related to my not having seen the DELTA program.

It is disappointing that he 'could not produce an acceptable linear key.' 'Acceptable' will vary from user to user, but every user has a right to expect acceptable results. I don't claim that the program produces publishable results (it produces a printed copy, a screen 'copy', and an ASCII file for further processing with a word processor). The form of the key is dependent upon the quality of the data and weighting of the characters. The smallest, most compact key should result from unweighted characters. Unless the user requests that 'remaining characters be dumped,' only characters used in keying will appear in the key. A database with 32 characters would indeed make a very bulky key, if the unused characters were to be listed in each 'final' lead (but, even so, it might be simpler to strip these out with a word processor than to locate and add the few 'auxiliary' characters that one might want in the key).

Reviewing software is time consuming, frustrating, and nearly as rewarding as writing it. Review and criticism is essential for improvement and very hard to find. My thanks to Dr. Spjut for his review.

## **Microcomputer Plant Identification (MPI) system**

**Laurence C. Hatch**

### **Microcomputer Plant Identification (MPI) system**

\$69.00 per volume. The MPI system is a comprehensive reference and development package for identifying native and cultivated plant taxa on an IBM PC or PS/2 compatible microcomputer. The Master Menu processes both single access (linear, dichotomous) and multiple access (character grouping) keys. Individual plant groups and genera may be loaded from the menu by entering a code (e.g. t for trees, v for vines) or the first three letters of the genus (e.g. jun for *Juniperus*, vio for *Viola*, ace for *Acer*). Main functions operate from function keys F1-F12 and include such commands as list all taxa in key, scroll through key, preview decision path, find keyword/epithet, list non-excluded taxa, look up term definition, undo last choice, or backtrack to one of three marked points in the key. The display supports traditional key formats plus charts, full page descriptions, bibliographic references, herbarium citations, and any ASCII text up to 20 lines x 60 characters per page. Screen results can be output to wordprocessor ASCII textfiles to document student use or merging with manuscripts and letters. The system operates under BASICA.

MPI Volume No. 1 contains the complete menu system, development package, 25 keys to app. 300 taxa, and an expert system program to debug keys developed by users. Volume No.2 includes 35 additional keys to app. 400 taxa and offers additional software for key development. Volume No. 2 requires Volume No. 1.

TCR will distribute 1–2 volumes of keys per year to support the system. We also distribute public domain and shareware keys developed by users at cost (\$3.00 postage, disk cost) to registered MPI users. Key developers can also have free advertisements listed in the MPI Newsletter which covers new programming ideas, product release notices, and essays on basic ID skills.

The MPI system operates on any IBM PC or PS/2 compatible with one 5.25" disk drive and 256K RAM or more. A harddisk is required to operate more than one MPI volume. The system operates on portable PC's for field identification.

### **Graduate student grant program for MPI software**

Taxonomic Computer Research is offering a free grant of MPI system software to current graduate students in botanical, horticultural, entomological, zoological, and general taxonomy. The grant requires approval of the graduate faculty adviser or taxonomic adviser and runs for the duration of the course of study. The grant is made on the condition that the student create and original MPI format key and make the key available in the public domain. Neither TCR or any other party will distribute the key for profit or commercial interest. The key or keys created should reflect new research related to the student's plan of study and offer effective communication of new taxonomic concepts and delimitations. A copy of the grant application may be obtained by writing to Taxonomic Computer Research, Student Software Program, P.O. Box 12011, Raleigh, NC 27605, USA.

## **Computer software for botany**

### **Jan-Peter Frahm**

The International Association of Bryologists has built up a software library for IBM PCs and compatibles. This software library contains at present more than 170 diskettes. The programs cover all computer applications in Institutes for Systematic Botany or Herbaria:

- general computer application such as wordprocessing, database programs, chart-programs, integrated software packages, utilities, etc. selected from the public domain.
- general scientific applications such as statistic packages, multivariate analysis.
- special applications for Systematic Botany such as herbarium database programs, herbarium labels, herbarium loans, cladistics, phenetics, identification programs, mapping.
- applications for related fields such as ecology, plant sociology, etc.

All programs may be copied free. They can be obtained free for members of the International Association of Bryologists by sending the appropriate number of blank disks. For non-IAB-members and institutes, a fee of \$8 per disk is charged to cover the costs of copying, mailing and maintenance of the software library.

For a free catalogue send a blank 5.25" disk to: Dr. Jan-Peter Frahm, Universitaet Duisburg, Fachbereich 6, Botanik, Postfach 101629, D-4100 Duisburg, F.R.G.