DELTA Newsletter 1. Originally published in hard copy.
This reformatted electronic version is available at http://delta-intkey.com

DELTA Newsletter

Number 1, February 1988

Compiled by Mike Dallwitz and Richard Pankhurst

In response to demand, we are producing a newsletter, of interest to all botanical and zoological taxonomists using computers. We expect to produce the next issue in October, so please send copy to reach us by the end of September. This first issue is being sent to anyone that we think might like to receive it, but to receive further issues, please fill in and return the registration form.

DELTA (DEscription Language for TAxonomy) is a computer format for taxonomic descriptions of plants or animals, and was invented by Mike Dallwitz. These descriptions can be used for expert interactive identification programs, to make identification keys, to write descriptions in various natural languages and as an exchange format generally.

Mike Dallwitz and Toni Paine have written a set of programs to process DELTA, and PANKEY is the name of the similar package by Richard Pankhurst. These packages are compatible and complementary, but not identical. They are highly useful for the preparation of monographs, Floras and Faunas, and a number of important publications have been already been produced in this way. DELTA is simply a data format and not (yet) a database system, but see below.

The DELTA System

Mike Dallwitz

Currently available programs

1. CONFOR. Converts DELTA into natural-language descriptions, or into other formats. The other formats currently available are:
   (a) KEY, DIST, and INTKEY formats, for other programs in the package (see below).
   (b) PAUP format, for cladistic analysis.
   (c) DELTA format, to tidy the data files, to re-order characters or character states, or to produce subsets of the data.
2. KEY. Constructs diagnostic keys.
3. DIST. Constructs a distance matrix for phenetic analysis.
4. INTKEY. Interactive identification and information retrieval.

Developments

It is planned to replace the present system, which is based on sequential data files as the primary means of data storage, by a system using random-access data files. This will enable much quicker access to any part of the data, for data entry, retrieval, and correction. Auxiliary information, such as character dependencies, character types, and masks will be automatically kept consistent with the main data when operations such as character re-ordering are carried out. A hierarchy of items (taxa) will be maintained, going down to specimen level if required, and there will be provision for automatic passing of information up and down the hierarchy. The new system will be able to read the current DELTA format, and will use an enhanced version of it for data exchange.

The PANKEY programs

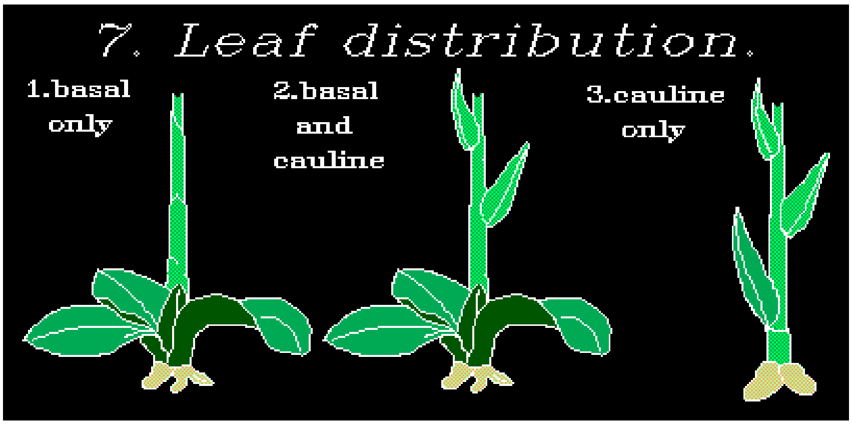
Richard Pankhurst

Currently available programs

1. KEY3M2. Construction of diagnostic keys. Reads DELTA data and constructs keys without user guidance.
2. KCONI. Interactive key construction. Writes keys with the user making his own choice at every stage to produce a very polished key.
3. ONLIN6. Interactive expert identification program. Question and answer at the keyboard. Now includes colour graphics character images.
4. DESCRIP3. Converts DELTA to written descriptions in natural language.
5. SPD1. Finds diagnostic descriptions, the smallest character sets which will distinguish a taxon from all others.
6. MATCH. Uses similarity coefficients to make an accurate comparison, suitable for numerous similar taxa.
7. Conversion programs SC3 to give a matrix of similarity coefficients for cluster analysis, and DELPAUP1 to convert to PAUP format for cladistics.
8. DEDIT, DELTA editor. Current version is simply a utility for re-ordering and/or deleting characters and tidying the data.

Developments

The package is usually provided for IBM compatible micros, but is suitable for any kind of computer. RP would like to hear from anybody willing to help implement the package on Macintosh, Vax etc. The programs are gradually being rewritten in C, starting with KCONI and ONLIN6. Work is underway on an interactive editor for DELTA, in the style of Wordstar. This is a step towards moving DELTA into a database system. This is highly desirable, and many users have asked for it, but commercial database systems are inadequate (Revelation?), except perhaps for some of the most sophisticated ones, e.g. Empress, which we are not able to afford. The BAOBAB system (Bob Allkin) is also planned to provide a proper taxonomic database. A database system for DELTA must be able to cope with variable length and repeated fields, recognise the dependencies between characters, and incorporate the taxonomic hierarchy.



Example of image from ONLIN6 identification program.

Representation of extreme and normal values in DELTA

Mike Dallwitz

When entering numeric values in DELTA format, it is often important to decide whether to enter the range of values normally found, or the most extreme values ever found. Both types of information are of interest for descriptions, but for identification, one or the other may be preferable. For absolute accuracy in identification, the extreme values are clearly necessary. But if they are used, the character often loses much of its diagnostic power, because extreme ranges overlap more than normal ranges.

It is possible to code either the normal or extreme range, and record the other in a comment, e.g. 10,2.5–3.4<rarely 2.1–4.2> or 10,2.1–4.2<usually 2.5–3.4>. However, only the coded information is accessible for identification, and once the information has been coded one way, there is no easy way to convert to the other if that should later seem desirable.

Another method, which has always been possible in DELTA but has seldom been used, is to code a ‘range’ containing all of the values, e.g. 10,2.1–2.5–3.4–4.2. (It is also possible to place a fifth value in the middle, to indicate the mean, median, or mode.) However, the programs made no special interpretation of such extended ranges. For purposes requiring a range (such as identification), the outermost values were used, and for purposes requiring a single value (such as the calculation of distances), the mean of all the coded values was used. In natural-language descriptions, all the values appeared exactly as coded, so that the reader would have had to be aware of the intended interpretation. (This did have the advantage that the compiler of the data was free to choose any interpretation, e.g. the values might represent certain percentiles.)

In recent discussions with DELTA users, I proposed to formalise the scheme described in the last paragraph. If four or five values appeared in a range, the outermost were to be interpreted as the extreme range, the next innermost as the normal range, and the central one (if present) as the mean, median, or mode. In natural-language descriptions, parentheses would be used in the conventional way to denote the extreme values, e.g. (2.1–)2.5–3.4(–4.2). If an extreme value was required at only one end of the range (as happens quite frequently with integer values), the normal and extreme values would be coded the same at the other end of the range, e.g. 1–1–2–4 would represent l–2(–4). Joseph Kirkbride, of the U.S.D.A., thought that separate symbol for extreme values would be preferable, e.g. 1–2*4 instead of 1–1–2–4. When Leslie Watson drew my attention to the need to represent situations such as l(–2), I realised that both methods were unsatisfactory. My own proposal lead to the very clumsy 1–1–1–2, and Kirkbride’s was unable to distinguish l(–2) from (l–)2.

I therefore decided that it would be best to code the extreme values in the conventional way, using parentheses, e.g. 10,(2.1)–2.5–2.8–3.4(–4.2). CONFOR has now been modified to accept this coding. Applications requiring a single value use the middle normal value, if present (2.8 in the above example), or otherwise the mean of the normal values. Applications requiring a range of values use the extreme values by default, but the use of normal values for any characters can be specified by means of a USE NORMAL VALUES directive.

Appeal to users

HELP! MD and RP are greatly overstretched by the demands of maintaining the existing software and writing more, not to mention other commitments and slender resources. Users have made many sensible suggestions for improving the software and the documentation, and if we respond slowly or not at all, it is not because of lack of commitment. We are producing this newsletter with slight diffidence, since we would prefer there to be an independent DELTA USER GROUP. Would somebody out there like to volunteer to organise it, please? An editor would need to gather and edit copy, organise reproduction and distribution, and coordinate a response from us to the users.

In the meantime, articles and letters for inclusion in the Newsletter will be gratefully received.

R. J. Pankhurst, Taxonomic Systems, 203, Sheen Lane, London SW14 8LE, England.
M. J. Dallwitz, CSIRO Division of Entomology, PO Box 1700, Canberra ACT 2601, Australia.