

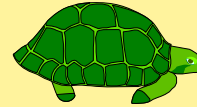
Bioinformatics CSM17 Week 3: Biological Identification

- A fundamental activity
- Traditional methods - keys
- Special problems
- Computer-based methods

JYC: CSM17

Fundamental concepts

- types are often not typical!
- homology



JYC: CSM17



How to identify an organism? Traditional/classical methods...

- Find someone who knows what it is !
- Indented and bracketed Keys
– since the 1600s !
- Floras and monographs
- Mostly phenotypic characters



JYC: CSM17

Traditional Methods ...

- A key to identify Human, Cow, Dog (only!)
 1. Number of legs twoHuman
 1. Number of legs four2.
 2. Stomach chambers four; eats grass
.....Cow
 2. Stomach chambers one; eats meat
.....Dog

JYC: CSM17

Difficulties caused by...

- new taxa (e.g. new species)
- phenotypic variation
- genotypic variation
- maturity
- sexual dimorphism
- incomplete material
- 'incorrect' classification

JYC: CSM17

The *value* of characters

- Ease of observation
- Clarity / unambiguous
- Information content: Entropy (H)

This Entropy value (H) is given by

$$H = -\sum_{i=1}^m p_i \log_2 p_i \quad 3.7$$

for a character with m states, where p_i is the proportion of the total taxa possessing the character state i .

JYC: CSM17

Computer-based methods

- Key generators eg. DELTA
- On-line keys
 - Polyclaves e.g. LucID, CABIKEY
- Expert Systems



JYC: CSM17

DELTA

- Descriptive Language for Taxonomy
- a suite of programs and tools
- a database format
- KEY generator

JYC: CSM17

Main files

- ITEMS
- CHARS
- SPECS

JYC: CSM17

CHARS

- The Characters (attributes)
- Character types
 - Unordered Multistate (UM) e.g. 1. red, 2. blue, 3. green
 - Ordered Multistate (OM) e.g. small, medium, large
 - Integer Numeric (IN) e.g. 1, 2, 5, 3, 8, 9 etc.
 - Real Numeric (RN) e.g. 32.5, 0, 45.2, 3.1 etc.
 - Text (TE) e.g. Collected by J.Smith in 1992

JYC: CSM17

CHARS

```
*SHOW Tilia species - character list.  
*CHARACTER LIST  
#1. Leaf width/  
    cm/  
#2. Axillary tufts/  
    1. absent/  
    2. indistinct or sparse/  
    3. clearly present/  
#3. Flowers per cyme/
```

JYC: CSM17

ITEMS

- the taxa, e.g. species, subspecies, varieties
- Format...

ITEMS

*SHOW Comments are written here

#NAME/

<Char>,<CharState> <Char>,<CharState> ..

JYC: CSM17

ITEMS

ITEMS

*SHOW This is an example for *Tilia*

#HEN/

1,9.9 2,3 3,26

JYC: CSM17

SPECS

- Number of characters
- Maximum number of character states
- Maximum number of items (taxa)
- Character types
- Number of states per character

JYC: CSM17

SPECS

```
*SHOW Tilia species
*NUMBER OF CHARACTERS 22
*MAXIMUM NUMBER OF STATES 7
*MAXIMUM NUMBER OF ITEMS 88
*CHARACTER TYPES 1,RN 2,OM 3,IN
*NUMBERS OF STATES 2,3
```

JYC: CSM17

DELTA KEY Generator

- Creates a text-based identification key
- Chooses 'best' characters first
- Uses a 'comparison' function
- Finds the character which requires fewest questions

JYC: CSM17

TOKEY

```
*SHOW Translate into KEY format
*INPUT FILE specs
*TRANSLATE INTO KEY FORMAT
*COMMENT. EXCLUDE CHARACTERS
*USE NORMAL VALUES 1 3
*COMMENT. CHARACTER RELIABILITIES
*KEY STATES
1,2.1-2.7/4.1-10.8
3,2.8-13.7/15.0-17.8/19.7-22.3/26.0
```

JYC: CSM17

DIANA

- A DELTA shell
- Integrates functionality in Windows

JYC: CSM17

INTKEY

- An interactive *multimedia* on-line key system bundled with DELTA
- Example for Grasses
- Can include pictures
- User chooses order of characters

JYC: CSM17

ETI - Expert center for Taxonomic Identification

- University of Amsterdam, The Netherlands
- Series of Multimedia interactive software
- Includes interactive key, pictures, videos...
- Written by acknowledged experts

JYC: CSM17

AI: Expert Systems, Neural Nets

- EXPERT KEY (Atkinson & Gammerman)
- ISAR (Chesmore *et al.*)
- ANNKEY (Clark & Warwick)

JYC: CSM17

Leading to the Future...

- DNA and RNA
– CATCATCATCATCAT
- eg. Forensic science, Paternity, Maternity
- XDELTA uses XML - eXtensible Markup Language (L.Dodds)
<http://www.ldodds.com/delta/>
- Taxonomic Markup Language (R.Gilmour)



JYC: CSM17

Useful Websites

- DELTA and DIANA: <http://www.delta-intkey.com>
- LUCID: <http://www.lucidcentral.org>
- Digital Taxonomy:
<http://digitaltaxonomy.infobio.net/>

JYC: CSM17

References & Bibliography

- Atkinson & Gammerman (1987). An application of expert systems technology to biological identification. *Taxon* 36 (4), pp. 705-714
- Chesmore, E.D. et al. (1998). Automated analysis of insect sounds. In Bridge, P. et al. (eds.) *Information Technology, Plant Pathology and Biodiversity*, CAB International, pp.273-287.
- Clark, J.Y. (2003). Artificial neural networks for species identification by taxonomists". *BioSystems*, vol. 72, pp. 131-147
- Clark, J.Y. & Warwick, K. (1998). Artificial keys for botanical identification using a multilayer perceptron neural network (MLP). *Artificial Intelligence Review* vol 12, pp.85-115
- Dallwitz, M.J., Paine, T.A. & Zurcher, E.J. (1997). *User's guide to the DELTA system -a general system for processing taxonomic descriptions*, Edition 4.07, CSIRO Division of Entomology; Canberra, Australia. (www.delta-intkey.com)
- Pankhurst, R.J. (1991). *Practical Taxonomic Computing*. University of Cambridge Press; UK.
- Pankhurst, R.J. (1978). *Biological Identification. The Principles and Practice of Identification Methods in Biology*. Edward Arnold, London
- Pankhurst, R.J. (1998). A historical review of identification by computer. In Bridge, P. et al. (eds.) *Information Technology, Plant Pathology and Biodiversity*, CAB International, pp.289-303

JYC: CSM17