

# Cladistics : A practical guide for phylogenetic, biogeographic and cospeciation analyses

Autor(en): **Burckhardt, D.**

Objektyp: **Article**

Zeitschrift: **Entomologica Basiliensia**

Band (Jahr): **22 (2000)**

PDF erstellt am: **22.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-980884>

## **Nutzungsbedingungen**

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

## **Haftungsausschluss**

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Entomologica Basiliensia	22	21-30	2000	ISSN 0253-24834
--------------------------	----	-------	------	-----------------

## INTERNATIONALE ENTOMOLOGEN-TAGUNG BASEL 1999

# Cladistics: A practical guide for phylogenetic, biogeographic and cospeciation analyses

by D. Burckhardt

**Abstract.** An introduction is provided to the use of some computer programmes for cladistic analyses of phylogenetic, biogeographic and coevolutionary data. The paper is intended for beginners providing examples of data input and basic manipulations to get the programmes started. Particular emphasis is given to the biogeographical analyses where procedures are detailed.

**Keywords.** computer programmes - biogeography - cladistics

## 1. Introduction

In biosystematics, cladistic methodology is widely used now. For its implementation, a variety of protocols and computer programmes are available covering aspects of phylogeny, biogeography and cospeciation. Introductions into the theory and practice are given by KITCHING *et al.* (1998) for cladistics, by HUMPHRIES & PARENTI (1999) for biogeography and by SCHUH (2000) for both. The programmes come usually with a manual or an online help function. The present paper is intended as a help for beginners in providing information on how data have to be prepared for input and on the first steps to take for running an analysis. It is not a manual to the various programmes.

Information on software packages can be found on the internet under following addresses:

<http://www.vims.edu/~mes/hennig/software.html>

<http://evolution.genetics.washington.edu/phylip/software.html>

<http://phylogeny.arizona.edu/tree/programs/programs.html>

<http://www.ucmp.berkeley.edu/subway/phylo/phylosoft.html>

These sites provide information on wide range of programmes including how to get them. All the programmes discussed in the following run on PCs. The examples in the phylogenetic analyses are with morphological characters; molecular data are not considered here.

## 2. Phylogeny

In a phylogenetic analysis taxa are grouped based on shared derived characters. To start an analysis, characters have to be selected which constitute potential synapomorphies. Their polarity can be determined by outgroup comparison. The following example taken from a revision of jumping plant-lice (Hemiptera, Psylloidea, Calophyidae) (BURCKHARDT & BASSET, 2000) contains 13 morphological characters with two or three characters states each:

**Larva**

- 1 Anterior margin of head weakly concavely rounded, not cleft in the middle (0), forming two shallow lobes (1), forming two large lobes (2).
- 2 Antenna without (0), with sectasetae (1).
- 3 Antenna straight or weakly curved (0), strongly curved downwards (1), S-shaped (2).
- 4 Abdominal dorsum without (0), with median horns (1).
- 5 Caudal plate ratio > 0.5 (0), < 0.5 (1).
- 6 Circumanal ring oval (0), round (1), absent (2).
- 7 Head margin without or with unspecialised setae only (0), with clavate setae (1), with sectasetae (2).
- 8 Wing bud margin with unspecialised setae only (0), with clavate setae (1), with setasetae (2).
- 9 Abdominal margin with sectasetae (0), without specialised setae (1), with lanceolate setae (2).

**Adult**

- 10 Genal processes short, broadly rounded (0), long, slender (1).
- 11 Forewing with (0), without surface spinules in addition to those in cell  $cu_2$  (1).
- 12 Forewings broadly rounded apically (0), narrowly rounded apically (1).
- 13 Distal segment of aedeagus thick (0), slender with elongate apical dilatation (1), slender with globular apical dilatation (2).

For each taxon the respective character states are organised in a matrix. For missing or inapplicable observations (entries), the symbols '?' or '-' can be used.

<i>rhois</i>	000000000000
<i>andina</i>	0001010021102
<i>clausa</i>	0001010021111
<i>duvauae</i>	0000??0010000
<i>gallifex</i>	1000102201101
<i>mammifex</i>	0001010021102
<i>orbicola</i>	0010121121101
<i>patagonica</i>	0000000021012
<i>rubra</i>	0001010021112
<i>schini</i>	2120102201001
<i>scrobicola</i>	2121002201101

Three programmes are discussed below, i.e. Hennig 86 (FARRIS, 1988), PAUP\* 4.02b (SWOFFORD, 1998) and TAX (NELSON & LADIGES, 1999). The first two are based on parsimony, the last on 'three-item statements' (NELSON & PLATNICK, 1991).

**Hennig 86**

Hennig 86 is a MS-DOS programme. The manuals by LIPSCOMB (1994) and SIDDAL (1999) should be consulted for working with it. TreeGardener (RAMOS, 1997) is a shell to run Hennig 86 interactively under MS Windows with integrated data editor and tree viewer.

The matrix, written with a text editor or with a word-processor (saved as a text-only file), has the following elements:

```
xread
'optional text'
13 11
rhois      00000000000000
andina     0001010021102
clausa     0001010021111
duvauae    0000??0010000
gallifex   1000102201101
mammifex   0001010021102
orbicola   0010121121101
patagonica 0000000021012
rubra      0001010021112
schini     2120102201001
scrobicola 2121002201101
;
proc/;
```

The command 'xread' can be followed by an optional text in ', then the numbers of characters and taxa have to be specified; the matrix has to be finished by a semicolon followed by the command 'proc/;'. Once Hennig 86 is opened, the matrix has to be processed specifying the path. Assuming our example is the file 'cal.dat' on drive D, the command 'process d:cal.dat;' opens the file. Then the proper analysis can be started. If not stated otherwise Hennig 86 treats the first taxon in the matrix as outgroup, and all multistate characters as ordered.

## PAUP

At the moment PAUP\* 4 is available as a beta version including a manual. PAUP uses NEXUS format. The matrix looks as follows:

```
#NEXUS
[optional text]

BEGIN DATA;
DIMENSIONS NTAX=11 NCHAR=13;
FORMAT MISSING=? GAP=- SYMBOLS= " 0 1 2";

MATRIX

rhois      00000000000000
andina     0001010021102
clausa     0001010021111
duvauae    0000??0010000
gallifex   1000102201101
mammifex   0001010021102
orbicola   0010121121101
patagonica 0000000021012
rubra      0001010021112
schini     2120102201001
scrobicola 2121002201101
;
ENDBLOCK;

BEGIN ASSUMPTIONS;
OPTIONS DEFTYPE=unord;

ENDBLOCK;
```

By default multistate characters are treated as unordered. Upon opening the programme prompts for executing or editing a file. Once the file is executed it can be analysed.

## TAX

TAX is a MS DOS programme to translate a data matrix into 'three-item statements'. For input the information of the matrix is entered step by step as demanded by the programme. The resulting matrix of 'three-item statements' can then be analysed by Hennig 86 or PAUP.

### 3. Biogeography

The observed spatial distributions of organisms are a result of historical and ecological factors. Historical biogeography is heavily dependent on systematics. As in the latter, three levels of knowledge can be distinguished: 1. typological: description of phenomena independent of inherent processes; 2. narrative: explanation of empirical data by current geological/geographical models; 3. analytic: empirical data are used to formulate an hypothesis which is compared to geological/geographical data.

#### Analytical historical biogeography

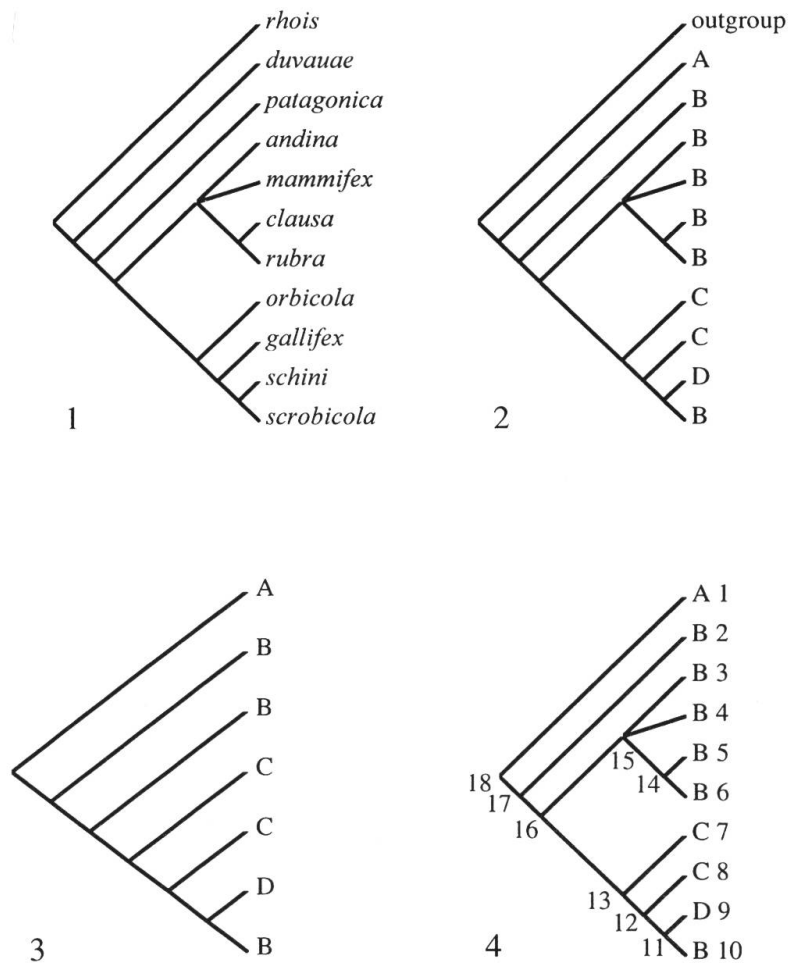
Geographical distributions of organisms can be analysed from a perspective of earth or of taxon history. The former seeks relationships between areas of distribution/endemism reflecting events of geological vicariance. The latter aims at explaining distributions of individual groups in terms of geological vicariance, dispersal, extinction, etc. The two types of cladistically based analyses in historical biogeography are called: 1. cladistic or area biogeography; 2. taxon or phylogenetic biogeography.

All these analyses require taxon cladograms (= TC) and a set of areas of distribution/endemism. These can be defined with geographical/geological criteria or by mapping the known distributions of the taxa within the analysed groups. The first step in each analysis is to create a taxon area cladogram (= TAC) by replacing the taxa in a TC by the areas in which they occur.

*Cladistic biogeography.* The aim of cladistic biogeography is to find a general hypothesis of historical relationships of areas of endemism called fundamental or general area cladogram. Ideally, each taxon is restricted to a single area, and each area contains a single taxon only. Often the situation is more complex with taxa occurring in more than one area, or areas bearing none or more than one taxon, in which case the taxon area cladograms have to be transformed. To deal with some of these problems, NELSON & PLATNICK (1981) formulated the biogeographical 'assumption 1' (areas linked by widespread taxa are treated as monophyletic or paraphyletic) and 'assumption 2' (areas may be monophyletic, paraphyletic or polyphyletic in respect to widespread taxa), and ZANDEE & ROOS (1987) added 'assumption 0', where widespread species are treated as synapomorphies for the areas they inhabit.

The following three methods can be used to translate TACs into resolved area cladograms (= RACs) (also termed "area cladograms" or "fundamental area cladograms"; ENGHOFF, 1996): Brooks Parsimony Analysis (BPA: BROOKS, 1990), 'three-area statements' (TAS: NELSON & LADIGES, 1994, including analysis of paralogy-free subtrees - TASS: NELSON & LADIGES, 1996) and component analysis (COMPONENT version 2.0: PAGE, 1993) (MORRONE & CARPENTER, 1994).

BPA is similar to the assumption 0. With TAS, data can be analysed under assumptions 0 and 1; for assumption 2 the data have to be modified. COMPONENT 2.0 can treat the data under assumption 0 ('Map widespread associates' on) and assumption 1 ('Map widespread associates' off), assumption 2 is not directly implemented.



Figs 1-4: Cladograms of the *rubra*-group; 1, taxon cladogram; *rhois* = outgroup; 2, taxon area cladogram; 3, simplified taxon area cladogram; 4, taxon area cladogram with numbered terminals and nodes.

### Example

1. Get taxon cladogram (Fig. 1).
2. Produce taxon area cladogram (TAC) in substituting the taxa by the areas in which they occur (Fig. 2). In our example the taxa occur in the four areas A, B, C and D.
3. Simplify TAC by reducing branches with multiple taxa occurring in the same area (Fig. 3).

#### 4. Produce RAC with BPA, TAS and COMPONENT.

##### a. BPA

- Get TC (Fig. 1), number interior nodes (Fig. 4), produce matrix taxa/nodes x areas, indicate absence/presence (ancestor of species a in area A and species b in area B occurs in areas AB).

- Prepare matrix for Hennig86 (text file!)

```
xread 'rubra group BPA' 18 5
out      000000000000000000
A        100000000000000001
B        011111000111111111
C        000000110001100111
D        000000001011100111
;
proc/;
```

- Run the file 'calbpa.dat' on drive D:

```
p d:calbpa.dat;ie*;tp;
```

- Combine several cladograms, use all zero outgroup.

##### b. TAS/TASS

- Open TASS for finding paralogy free subtrees.
- Follow instructions and use options for assumptions 0 and 1.
- Open TAS for building three-area statement matrix.
- Follow instructions step by step.
- Then the programme produces a file tas.dat which can be analysed with Hennig 86.

##### c. COMPONENT

- Create NEXUS file (e.g. cal.nex).

```
#NEXUS
[optional text]
BEGIN TAXA;
  DIMENSIONS NTAX=4;
  TAXLABELS
  A1
  A2
  A3
  A4
  ;
ENDBLOCK;

BEGIN DISTRIBUTION;
  TITLE='rubra-group';
  NTAX=10;
  RANGE
  andina      : 2,
  clausa      : 2,
```

```

duvauae      : 1,
gallifex     : 3,
mammifex     : 2,
orbicola     : 3,
patagonica   : 2,
rubra        : 2,
schini       : 4,
scrobicola   : 2;

```

```

TREE T1=(3, (7, ((1, 5, (2, 8)), (6, (4, (9, 10))))));
ENDBLOCK;

```

- Open programme, open file 'cal.nex', click on 'tree'.
- Open parasite tree, if there is more than one tree.
- Go to "Trees" menu, open "Map trees", → "Options" Choose "Map widespread associates" on for assumption 0 and off for assumption 1, → "Heuristic search", click "ok" → choose branch swapping options and criterion to minimise.

*Taxon history.* Many biogeographic papers examine only vicariance. Data which do not fit general patterns are explained *ad hoc* with dispersal or extinction. Only few studies take the last two phenomena into account, such as the search for ancestral areas (BREMER, 1992, 1995; MICHELSEN, 1994). A quantitative approach of dispersal-vicariance analysis was described by RONQUIST (1997) and is implemented in the programme DIVA 1.1 (RONQUIST, 1996). Ancestral distributions are reconstructed based on a three-dimensional step matrix. Optimal ancestral distributions are found in minimising dispersal and extinction events.

DIVA needs fully resolved trees (Fig. 3). The matrix which requires a tree and distributions of taxa, has to be written in text format. The text file (in our example cal.txt) has to be in the same folder as the programme.

```

/* optimal text

tree rubra-group (1, (2, (3, (4, (5, (6, 7)))))) ;

distribution

1 A
2 B
3 B
4 C
5 C
6 D
7 B
;

return;

```

The file is processed with the command 'proc cal.txt;'. The programme prompts 'tree read successfully, distribution read successfully, control returned to console'. Now the optimisations can be calculated.



## 4. Cospeciation

Cospeciation analyses compare independently derived host and parasite trees. They can be performed with COMPONENT version 2.0 (PAGE, 1993) and TREE MAP 1.0 (PAGE, 1995).

In case of incongruent host and parasite trees, COMPONENT constructs a reconciled tree so that the observed host-parasite relationships can be explained by association by descent alone. With COMPONENT, three measures of fit between host and parasite trees can be computed: - 1. Number of duplications of a lineage to reconcile the trees. - 2. Number of terminal taxa (= leaves) added. This measure corresponds to half the items of error as defined by NELSON & PLATNICK (1981). - 3. Minimal number of independent losses. For judging if there is evidence for cospeciation, these measures can be compared to those of  $n$  (100 or 1000) random trees.

The input matrix looks as follows:

```
#NEXUS

[optional text]

BEGIN TAXA;
  DIMENSIONS NTAX=9;
  TAXLABELS
  fasciculatus
  kauselii
  latifolius
  molle
  montanus
  patagonicus
  polygamus
  terebinthifolius
  velutinus
  ;
ENDBLOCK;

BEGIN DISTRIBUTION;
  TITLE = 'rubra group';
  NTAX=10;
  RANGE
  duvauae: fasciculatus,
  patagonica: patagonicus,
  mammifex: latifolius polygamus velutinus,
  andina: montanus,
  rubra: polygamus,
  clausa: kauselii montanus patagonicus,
  orbicola: polygamus,
  gallifex: polygamus,
  schini: molle,
  scrobicola: polygamus
  ;
  TREE                                *                                cal=
  (duvauae, (patagonica, ((clausa, rubra, (mammifex, andina)), (orbicola, (gal-
  lifex, (schini, scrobicola))))));
ENDBLOCK;

BEGIN TREES;
  TREE                                *
```

```
Schinus=(((fasciculatus,polygamus),((latifolius,velutinus),(kauselii,(montanus,patagonicus))))),(molle,terebinthifolius));
```

```
ENDBLOCK;
```

TREE MAP can be used to compare visually host and parasite associations. Reconstructions are made e.g. of the cospeciation history with the greatest number of cospeciation events. For testing the statistical evidence for cospeciation, the number of cospeciation events is compared to those of n (999) random host/parasite trees.

```
#NEXUS
```

```
[optional text]
```

```
BEGIN HOST;
```

```
TREE * Schinus = (((fasciculatus,polygamus),((latifolius,velutinus),(kauselii,(montanus,patagonicus))))),(molle,terebinthifolius));
```

```
ENDBLOCK;
```

```
BEGIN PARASITE;
```

```
TREE * Calrub = (duvauae,(patagonica,((clausa,(rubra,(mammifex,andina))), (orbicola,(gallifex,(schini,scrobicola))))));
```

```
ENDBLOCK;
```

```
BEGIN DISTRIBUTION;
```

```
RANGE
```

```
duvauae: fasciculatus,
patagonica: patagonicus,
mammifex: polygamus,
andina: montanus,
rubra: polygamus,
clausa: kauselii montanus patagonicus,
orbicola: fasciculatus,
gallifex: fasciculatus,
schini: molle,
scrobicola: polygamus
```

```
;
```

```
ENDBLOCK;
```

Contrary to COMPONENT, TREE MAP requires fully resolved trees.

## Acknowledgements

I thank P.Y. LADIGES (Melbourne), G. NELSON (Melbourne), R.D.M. PAGE (Glasgow), T.C. RAMOS (Sao Paulo) and F. RONQUIST (Lund) for letting me use their software packages.

## Literature

- BREMER, K. (1992): *Ancestral areas: a cladistic reinterpretation of the center of origin concept*. Syst. Biol., 41: 436-445.
- BREMER, K. (1995): *Ancestral areas: Optimization and probability*. Syst. Biol., 44: 255-259.
- BROOKS, D.R. (1990): *Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update*. Syst. Zool., 39: 14-30.
- BURCKHARDT, D. & BASSET, Y. (2000): *The jumping plant-lice (Hemiptera: Psylloidea) associated with Schinus (Anacardiaceae): systematics, biogeography and host plant relationships*. J. nat. Hist., 34: 57-155.
- ENGHOFF, H. (1996): *Widespread taxa, sympatry, dispersal, and an algorithm for resolving area cladograms*. Cladistics, 12: 349-364.

- FARRIS, J.S. (1988): *Hennig86, Version 1.5*. Published by the author, Port Jefferson, New York.
- HUMPHRIES, C.J. & PARENTI, L.R. (1999): *Cladistic biogeography: interpreting patterns of plant and animal distributions; Second edition*. Oxford Monographs on Biogeography, 12, 187 pp.
- KITCHING, I.J., FOREY, P.L. HUMPHRIES, C.J. & WILLIAMS, D.M. (1998): *Cladistics. The theory and practice of parsimony analysis*. Second edition. The systematics association publication, 11: 228 pp.
- LIPSCOMB, D. (1994): *Cladistic analysis using Hennig86*. George Washington University, 122 pp.
- MICHELSSEN, V. (1994): *The southern Palearctic genus Enneastigma Stein (Diptera: Anthomyiidae): Taxonomy, of Mediterranean species, phylogeny and ancestral areas analysis*. Ent. scand., 25: 361-375.
- MORRONE, J.J. & CARPENTER, J.M. (1994): *In search of a method for cladistic biogeography: An empirical comparison of Component Analysis, Brooks Parsimony Analysis, and Three-area Statements*. Cladistics, 10: 99-153.
- NELSON, G. & LADIGES, P.Y. (1994): *TAS - Three Area Statements: MS-DOS Program for Cladistic Biogeography, version 3.3*. New York and Melbourne.
- NELSON, G. & LADIGES, P.Y. (1996): *TASS: Three Area Subtrees. 2.1. MS-DOS Program*. New York and Melbourne.
- NELSON, G. & LADIGES, P.Y. (1999): *TAX: Three Taxon Statements. 3.3. MS-DOS Programme available from the authors, Melbourne*.
- NELSON, G. & PLATNICK, N.I. (1981): *Systematics and biogeography: Cladistics and vicariance*. Columbia University Press, New York.
- NELSON, G. & PLATNICK, N.I. (1991): *Three-taxon statements: a more precise use of parsimony?* Cladistics, 7: 351-366.
- PAGE, R.D.M. (1993): *COMPONENT, version 2.0. Tree comparison software for use with Microsoft Windows*. Programme and user's guide, the Natural History Museum, London.
- Page, R.D.M. (1995): *TREEMAP 1.0. Programme and user's guide*. Published by the author, Glasgow, UK.
- RAMOS, T.C. (1997): *Tree Gardener version 2.2.1. Computer programme*: Published by the author, Museu de Zoologia, Universidade de Sao Paulo, Sao Paulo, Brazil.
- RONQUIST, F. (1996): *DIVA version 1.1, Computer program and manual*. Available by anonymous FTP from Uppsala University (ftp.uu.se or ftp.sysbot.uu.se).
- RONQUIST, F. (1997): *Dispersal-Vicariance Analysis: A new approach to the quantification of historical biogeography*. Syst. Biol., 46: 195-203.
- SCHUH, R.T. (2000): *Biological systematics: principals and applications*. Comstock Publishing Associates, Cornell University Press, 239 pp.
- SIDDAL, M. (1999): *Hennig86 Simplified*. <http://www.vims.edu/~mes/mes/henhelp.htm>.
- SWOFFORD, D.L. (1998): *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and other methods). Version 4*. Sinauer Associates, Sunderland, Massachusetts.
- ZANDEE, M. & ROOS, M. C. (1987): *Component compatibility in historical biogeography*. Cladistics, 3: 305-332.

**Author's address:**

Dr. Daniel Burckhardt  
 Naturhistorisches Museum  
 Augustinergasse 2  
 CH-4001 Basel  
 SCHWEIZ