Multivariate methods in taxonomy

Karol Marhold

contact:

karol.marhold@savba.sk

Presentations available at:

https://botany.natur.cuni.cz/brassiploidy

Phenetic approach (multivariate methods; "pattern"; cluster analysis, ordination methods, discriminant analysis)

Cladistic approach (parsimony analysis)

Alternative approaches to the phylogenetic reconstruction (neighbour joining method, maximum likelihood, Bayesian statistical methods)

Geometric morfometrics (Booksteinove shape coordinates, Procrustes analysis, thin plate spline method)

Software: MorphoTools2, SYN-TAX 2000

Phenetic approach

Michel Adanson (1727-1806)

Familles des Plantes (1763)

65 different classifications, each based on one character (e.g. placentation, type of inflorescence ...)

PREFACE. cclxxix

55. Blitons.

17e CLASSE.

Etamines sur le Calice & sur l'Overe ensemble.

12. Elaeagnus.
13. Onagres.
14. Mirtes.
15. Ombellifères.
16. 2 Chevrefeuille.

32. Q. Pourpiers.
42. Q. Rosiers.

18e CLASSE.

Etamines sur l'Overe.

9. 4 Jonjoubres.
11. 12 Aristoloches.

19e CLASSE.

Etamines sur le stil de l'Overe.

10. Orchis.
11. 3 Aristoloches.

20e CLASSE.

Etamines sur la Corole.

9. 10 Jonjoubres.
16. Composées.
17. Campanules.
18. Briones.
19. Aparines.
20. Scabieuses.

21. P. Chevrefeuilles.
22. M. Aireles.
23. Apocins.
24. Bourraches.
25. Labiées.
26. Verveines.
27. Personées.
28. Solanons.
29. Jasmens.
30. Anagallis.
32. Q. Pourpiers.
33. Q. Joubarbes.
34. Q. Alsines.

42. Systême. *Etamines; leur figure respective.*

1re CLASSE.

Plantes sans Etamines.

1. Bissus.
2. Champignons.
3. 6 Fucus.

2e CLASSE.

Etamines distinctes les unes des autres.

4. Eparites.
5. Foujères.
7. Gramens.
8. 72 Liliacées.
9. Jonjoubres.
10. Orchis.
11. 11 Aristoloches.
12. Elaeagnus.
13. Onagres.
14. Mirtes.
15. Ombellifères.
16. 6 Composées.
17. 6 Campanules.
18. 2 Briones.
19. Aparines.
20. Scabieuses.

21. Chevrefeuilles.
22. Aireles.
23. 26 Apocins.
24. Bourraches.
25. Labiées.
26. Verveines.
27. Personées.
28. 9 Solanons.
29. Jasmens.
30. Anagallis.
31. Salicaires.
32. Pourpier.
33. Joubarbes.
34. Alsines.
35. Blitons.
39. Persicaires.
40. Garou.
41. Rosiers.
42. Jujubiers.
43. 22 Légumineuses.
44. 16 Pistachiers.
45. 12 Titimales.
46. Anones.
47. Chataigners.
48. Tilleuls.
52. 48 Crucifères.
53. 16 Pavots.
54. 64 Cistes.
55. Renoncules.
56. Arons.
58. Mousses.

3e CLASSE.

Etamines réunies toutes ensemble par les filets en un faisceau.

5. 3 Fucus.
6. Palmiers.
8. 2 Liliacées.
11. 4 Aristoloches.
36. Jalaps.
37. Amarantes.
38. Espargoutes.
43. 24 Légumineuses.
44. 15 Pistachiers.
45. 21 Titimales.
49. Geranions.
50. Mauves.
51. Capriers.
57. Pins.

4e CLASSE.

Etamines réunies par les filets en 2 corps.

43. 61 Légumineuses.

5e CLASSE.

Etamines réunies par les filets en plus de 2 corps.

43. 2 Légumineuses.
52. 1 Crucifère.
54. 7 Cistes.

6e CLASSE.

Etamines réunies par les anteres seulement.

16. 112 Composées.
17. 3 Campanules.
18. 13 Briones.
23. 2 Apocins.
28. 1 Solanon.
53. 1 Pavot.

7e CLASSE.

Etamines réunies par les filets & les anteres ensemble.

23. 1 Apocin.

Tte

Phenetic approach

Department of Entomology, University of Kansas, Lawrence, U.S.A.

Michener, Ch.D. & Sokal, R.R. 1957. A quantitative approach to a problem in classification. *Evolution* 11: 130-162.

Department of Microbiology, University of Leicester, U.K.

Sneath, P.H.A. 1957. Some thoughts on bacterial classification. *J. Gen. Microbiol*. 17: 184-200.

Sokal, R.R. & Sneath, P.H.A. 1963. *Principles of numerical taxonomy*. W. H. Freeman and comp., San Francisco & London.

Sneath, P.H.A. & Sokal, R.R. 1973. *Numerical taxonomy, the principles and practice of numerical classification.* W. H. Freeman and comp., San Francisco.

Neo-Adansonian principles

(1) The ideal taxonomy is that in which the taxa have the greatest content of information and which is based on as many characters as possible.

(2) A priori, every character is of equal weight in creating natural taxa.

(3) Overall similarity (or affinity) between any two entities is the function of the similarity of the many characters in which they are being compared.

(4) Correlation of characters differ in the groups of organisms under study. Thus distinct taxa can be recognized.

(5) Phylogenetic conclusions can be drawn from the taxonomic structure of a group and from character correlations, assuming some evolutionary mechanisms and pathways.

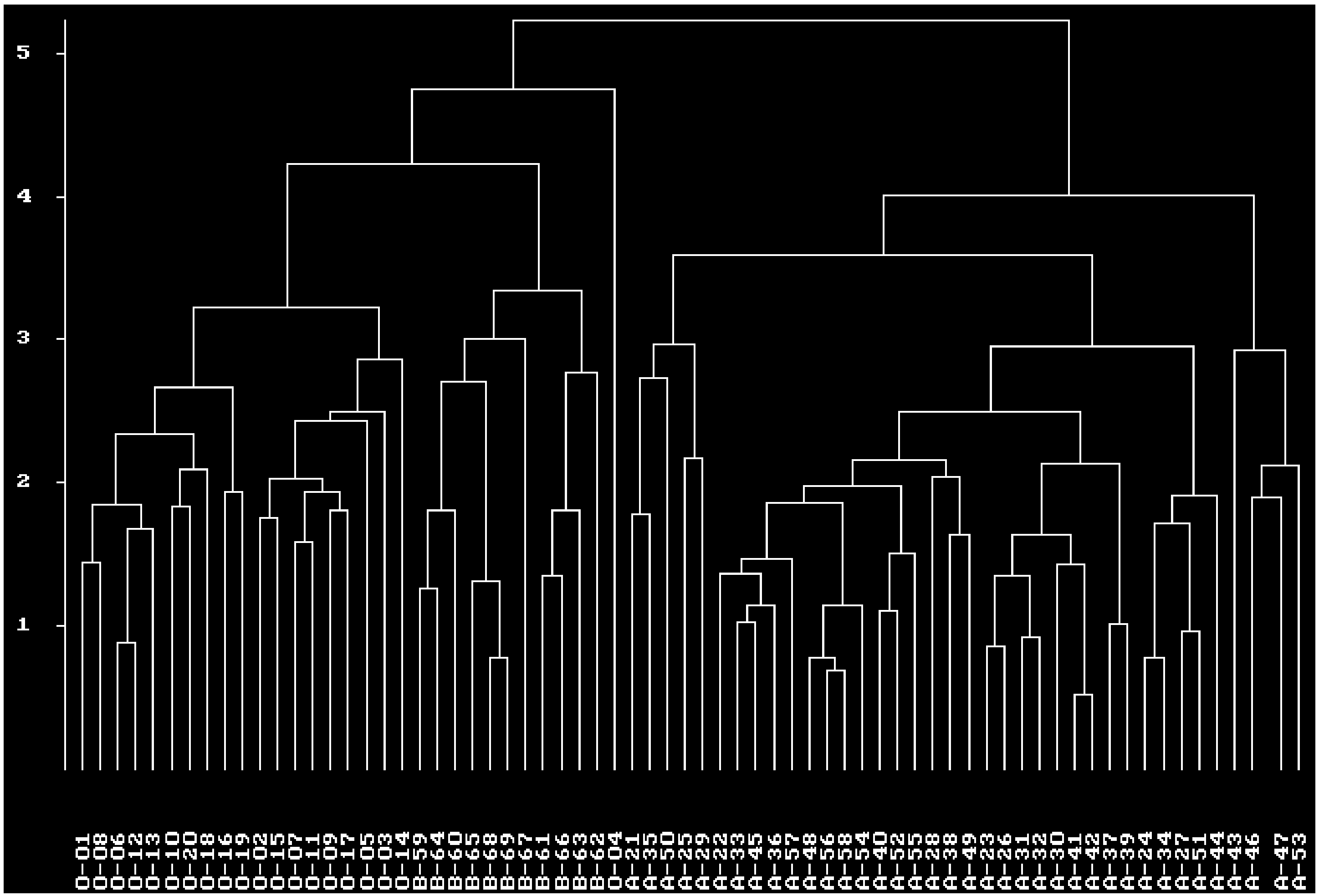
(6) The science of taxonomy is viewed and practiced as an empirical science.

(7) Phenetic similarity is the base of classifications.

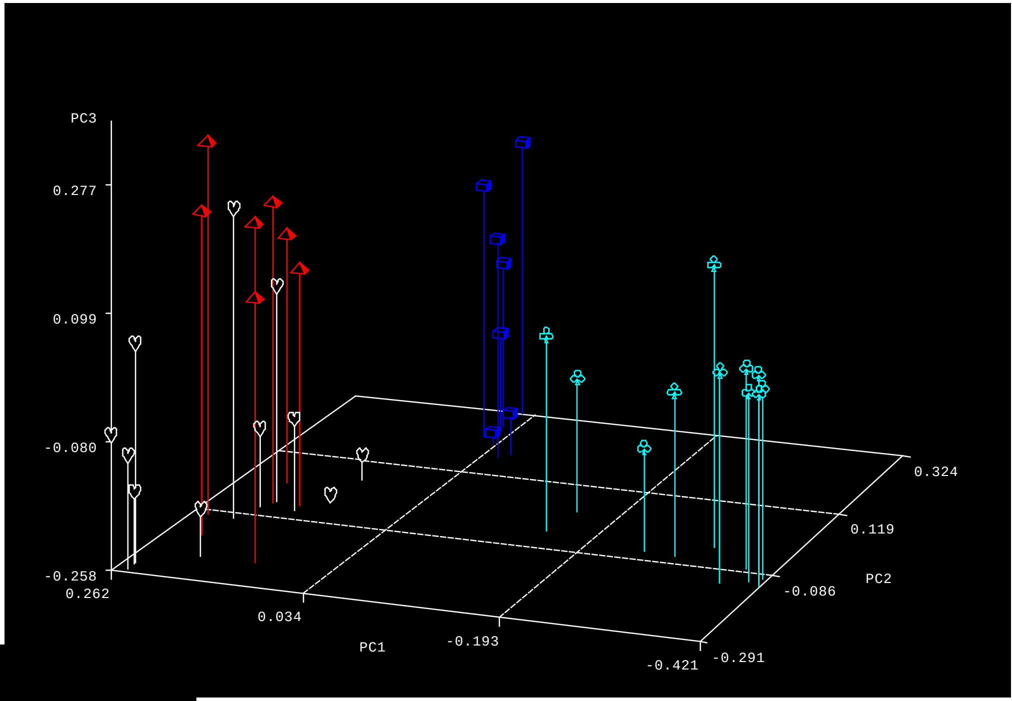
Phenetic approach

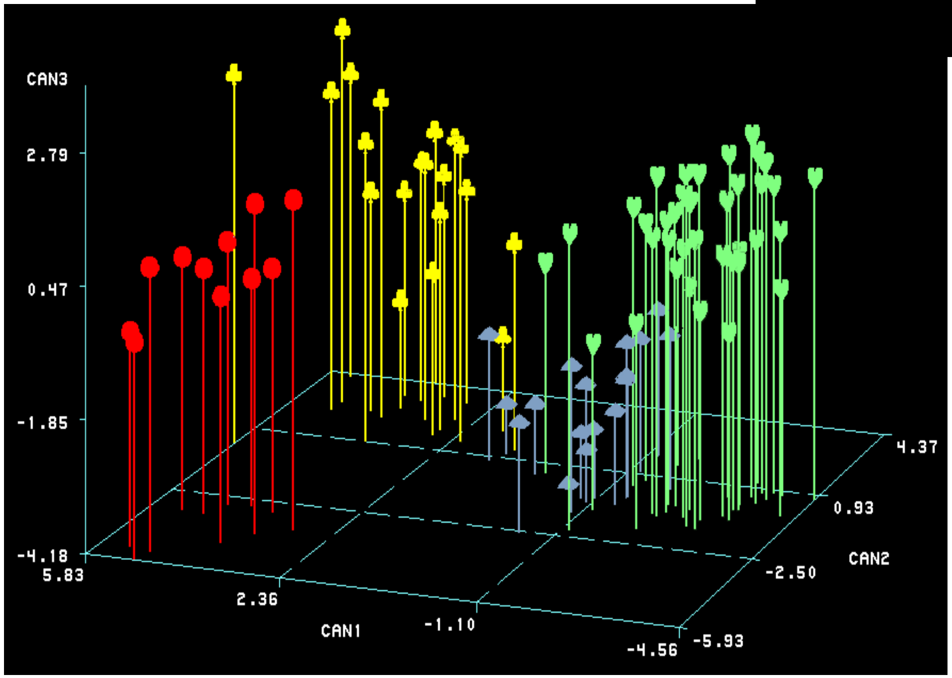
- Terms: numerical taxonomy (Sokal & Sneath), statistical systematics (Solbrig), numerical phenetics (Duncan & Baum), multivariate morphometrics (Blackith & Reyment)
- Operational Taxonomic Units (OTU)
- Characters, primary matrix, number of characters, correlations
- Coefficients expressing relationships between characters or objects, secondary matrix
- Multivariate methods (clustering methods, ordination methods, discriminant analysis)
- Different methods may yield different results
- Use of methods in the past and in current taxonomic practice (infraspecific variability, polyploid complexes, study of morphological variability in large areas, molecular data)

Cluster analysis



Principal component analysis



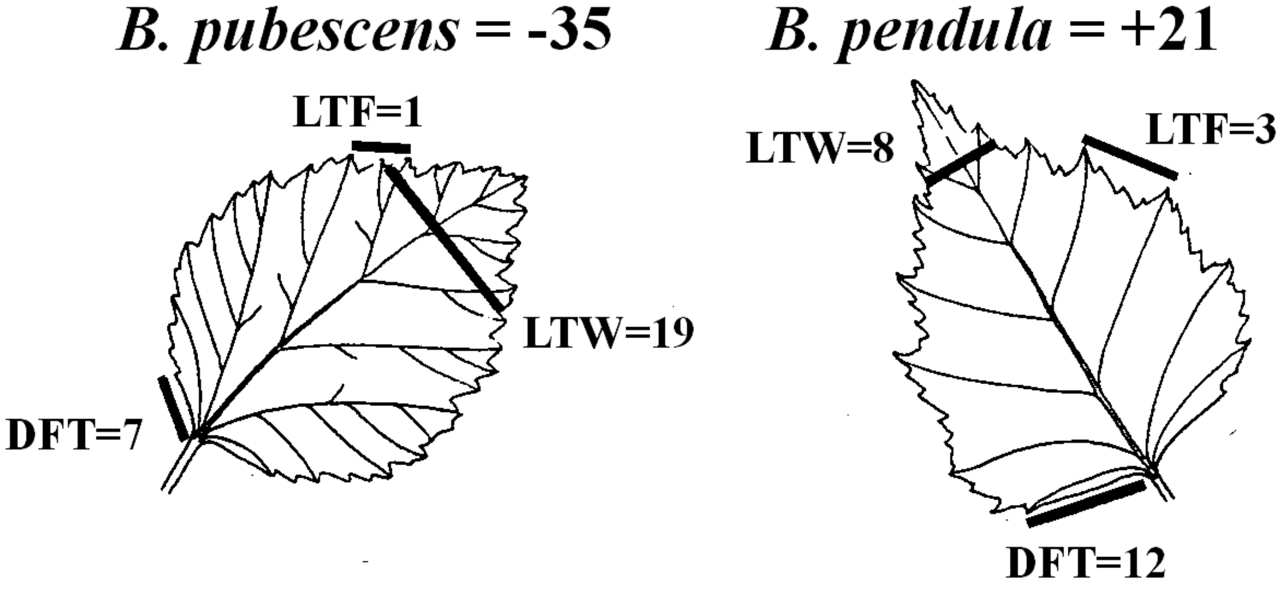


Canonical discriminant analysis

Classificatory discriminant analysis

group — The affiliation of plants to defined groups predicted based on the established classification criterion (the absolute number and percentage of plants classified into individual groups)

	amara	**austr.**	**olot.**	**opicii**	**pyren.**	Total
amara	**349**	20	3	1	7	380
	91.84	5.26	0.79	0.26	1.84	100.00%
austriaca	51	**302**	1	6	8	368
	13.86	**82.07**	0.27	1.63	2.17	100.00%
olotensis	2	0	**99**	0	0	101
	1.98	0.00	**98.02**	0.00	0.00	100.00%
opicii	1	9	0	**326**	42	378
	0.26	2.38	0.00	**86.24**	11.11	100.00%
pyrenaea	1	11	0	19	**207**	238
	0.42	4.62	0.00	7.98	**86.97**	

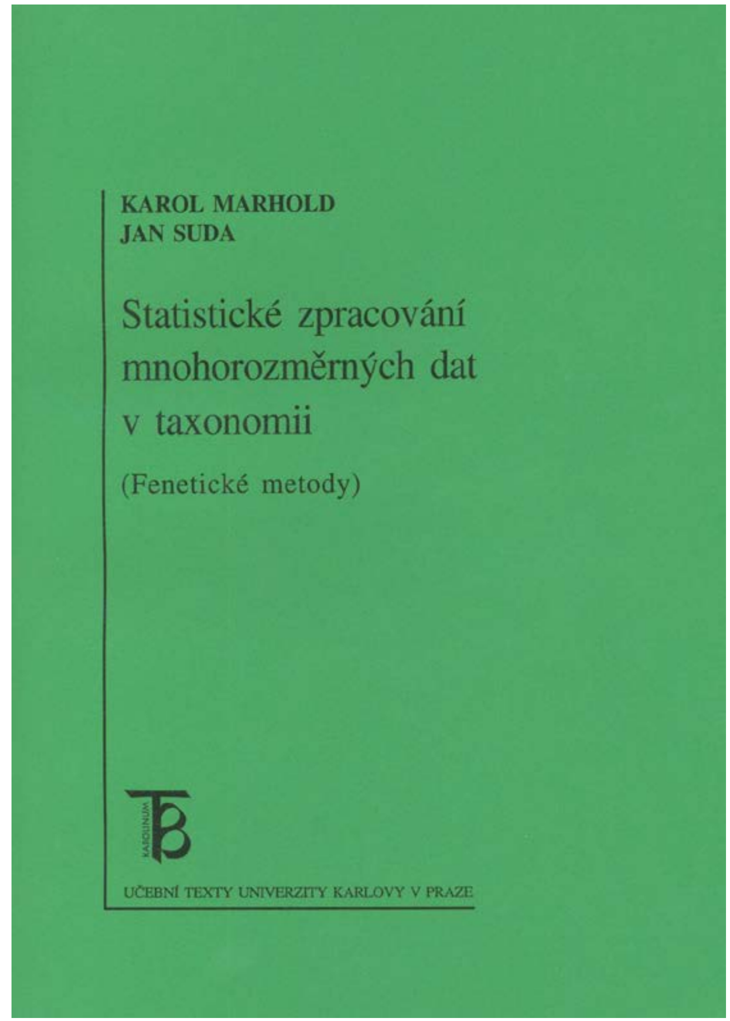


Discriminant function for determining the species Betula pubescens and B. pendula:

12LTF + 2DFT – 2LTW – 23
positive values: B. pendula
negative values: B. pubescens
probability of correct identification: 93%

(Stace, C. A., 1991, New Flora of the British Isles

Marhold, K. & Suda, J. 2002: *Statistické zpracování mnohorozměrných dat v taxonomii*. Karolinum, Praha.

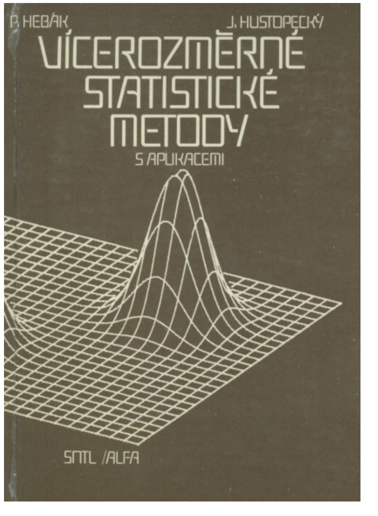


Hebák, P. & Hustopecký, J. 1987: *Vícerozměrné statistické metody s aplikacemi*. SNTL – nakladatelství technické literatury, Alfa, vydavateľstvo technickej a ekonomickej literatúry, Praha.

Hebák, P., Hustopecký, J., Jarošová, E. & Pecáková, I. 2007. *Vícerozměrné statistické metody (1)*. Ed. 2. Informatorium, Praha.

Hebák, P., Hustopecký, J. & Malá, I. 2005. *Vícerozměrné statistické metody (2)*. Informatorium, Praha.

Hebák, P., Hustopecký, J., Pecáková, I., Průša, M., Řezanková, H., Svobodová, A. & Vlach, P. 2007. *Vícerozměrné statistické metody (3)*. Ed. 2. Informatorium, Praha.

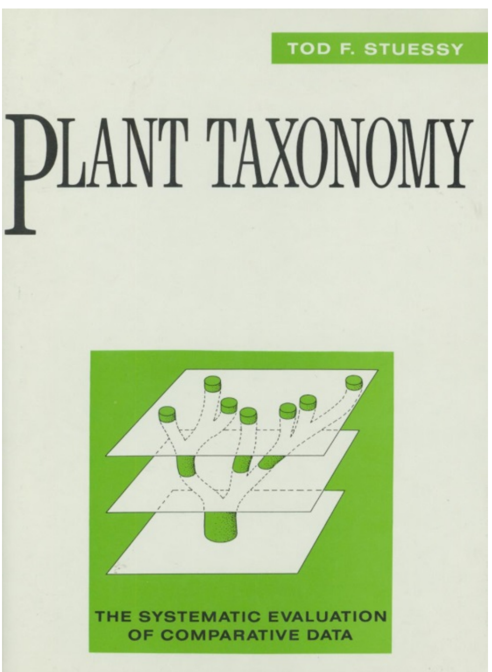


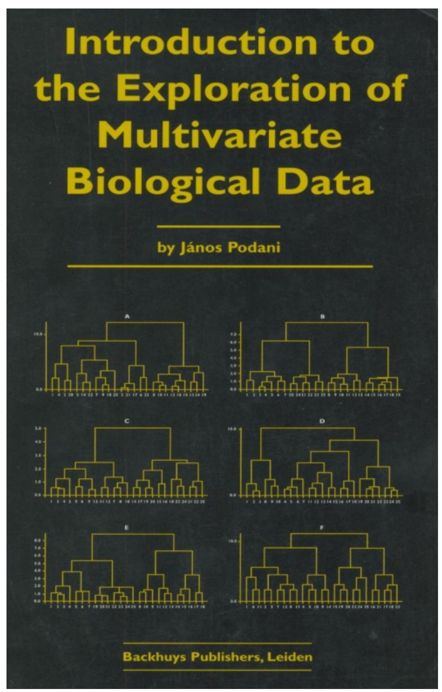
Legendre, P. & Legendre, L. 1998. *Numerical ecology*. Second English edition. Elsevier, Amsterdam.

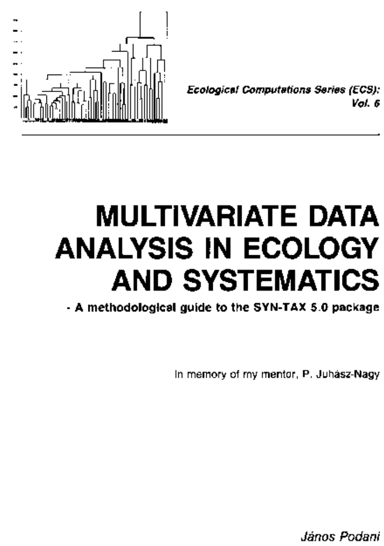
Podani, J. 1994. *Multivariate data analysis in ecology and systematics*. SPB Academic Publishing bv, The Hague.

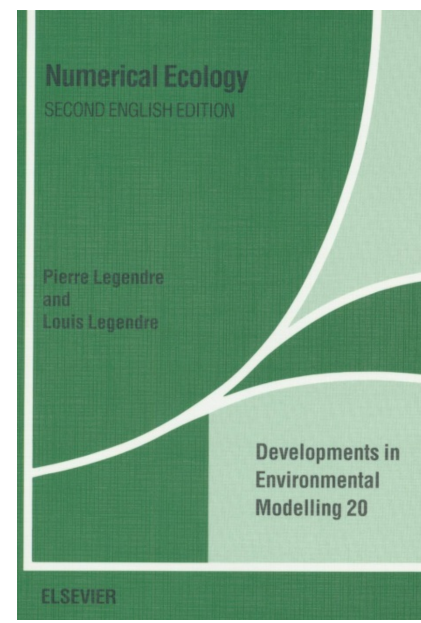
Podani, J. 2000. *Introduction to the exploration of multivariate biological data*. Backhuys Publishers, Leiden.

Stuessy, T. F. 1990. *Plant taxonomy: the systematic evaluation of comparative data*. Columbia University Press, New York.









Kladistický prístup

Hennig, W.

1950: *Grundzüge einer Theorie der phylogenetischen Systematik*. Deutsche Zentralverlag, Berlin.
1965: Phylogenetic systematics. *Annual Review of Entomology* 10: 97-116.
1966: *Phylogenetic systematics*. University of Illinois Press, Urbana.

Botany:

Koponen, T., 1968: Generic revision of Mniaceae Mitt. (Bryophyta). *Ann. Bot. Fenn.* 5: 117-151.
Funk, V. & Stuessy, T. F. 1978: Cladistics for practicing plant taxonomist. *Syst. Bot.* 3: 159-178.
Bremer, K. & Wantorp, H.- E. 1978: Phylogenetic systematics in botany. *Taxon* 27: 317-329.

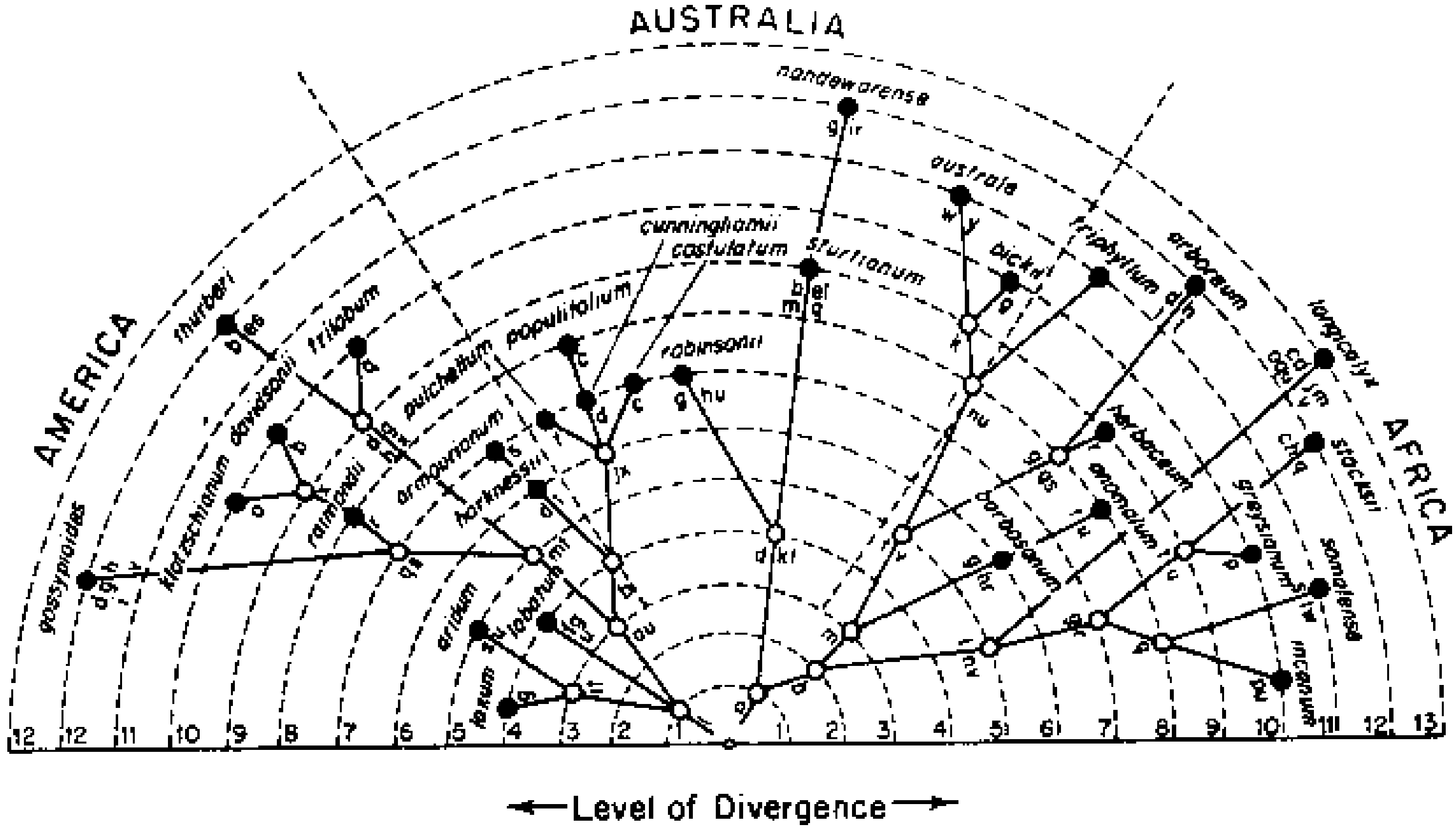


Fig. 2.17 Cladogram (Wagner tree) of 30 species of *Gossypium* (Malvaceae), modified from Frywell[142].

W.H. Wagner, University of Michigan - Groundplan/divergence method

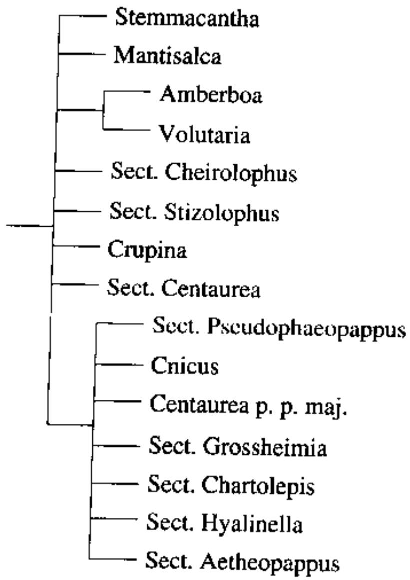


FIGURE 8-4. Strict consensus tree of six equally parsimonious cladograms of *Centaurea* sections and related genera based on cypsela characters from Dittrich (1966, pp. 138–139). The data matrix is given in Table 8-4.

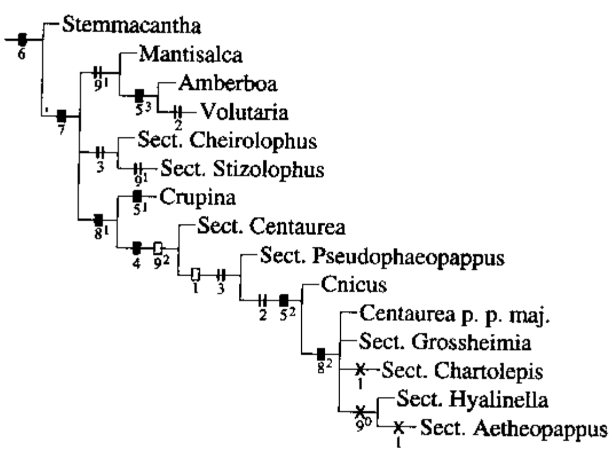


FIGURE 8-5. One of six equally parsimonious cladograms of *Centaurea* sections and related genera based on cypsela characters from Dittrich (1966, pp. 138–139). The characters are given in Table 8-3 and the data matrix in Table 8-4. Solid bars indicate nonhomoplastic synapomorphies; open bars indicate homoplastic synapomorphies with reversals; double bars indicate parallelisms; crosses indicate reversals.

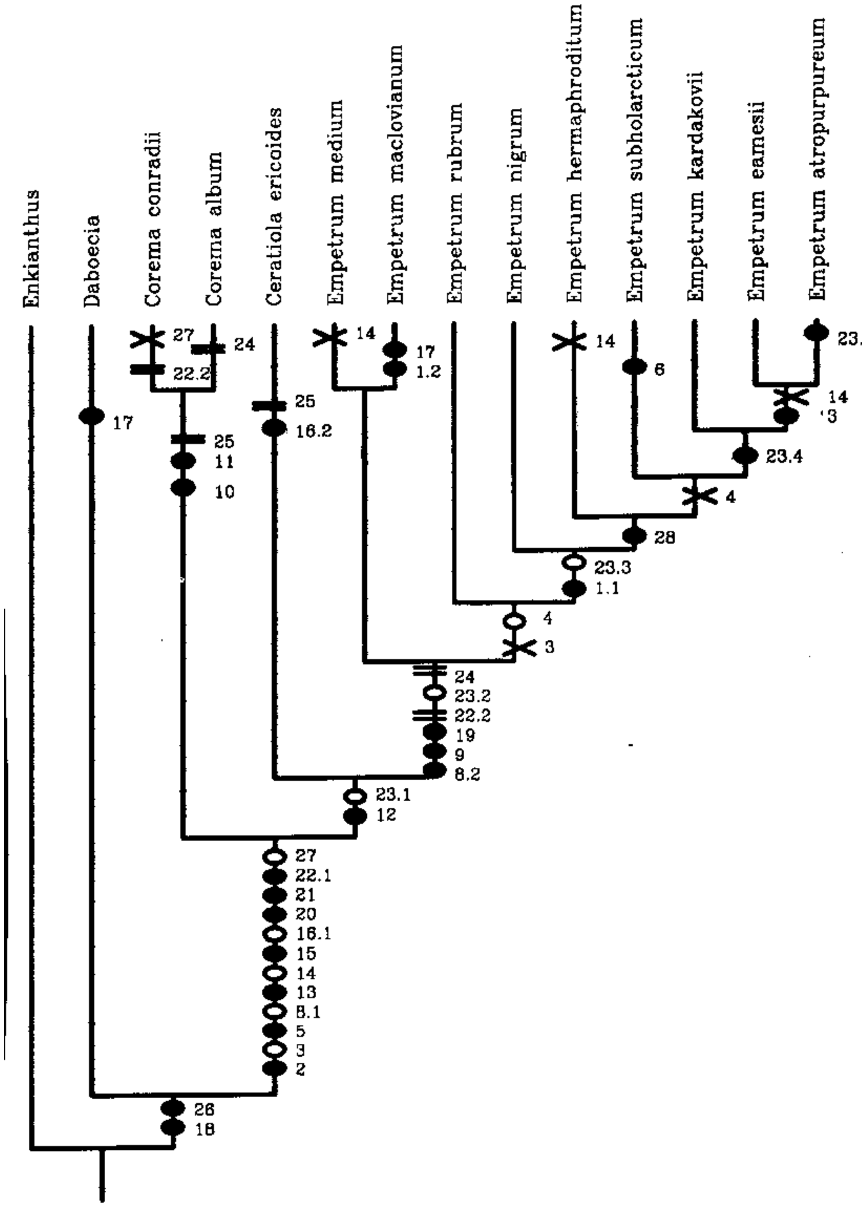


FIG. 2. One of five equally parsimonious cladograms of the Empetraceae. *Enkianthus* and *Daboecia* are outgroup taxa. Characters are numbered in accordance with the text, Appendix 1, and with Table 1. Black dots = synapomorphies (ci = 1), white dots = synapomorphies (ci < 1), parallel lines = parallelisms, crosses = reversals.

A Y-Z, X-Y-Z monophyletic groups

B X-Y paraphyletic groups

C X-Y polyphyletic groups, paralelisms

D X-Y polyphyletic groups, convergences

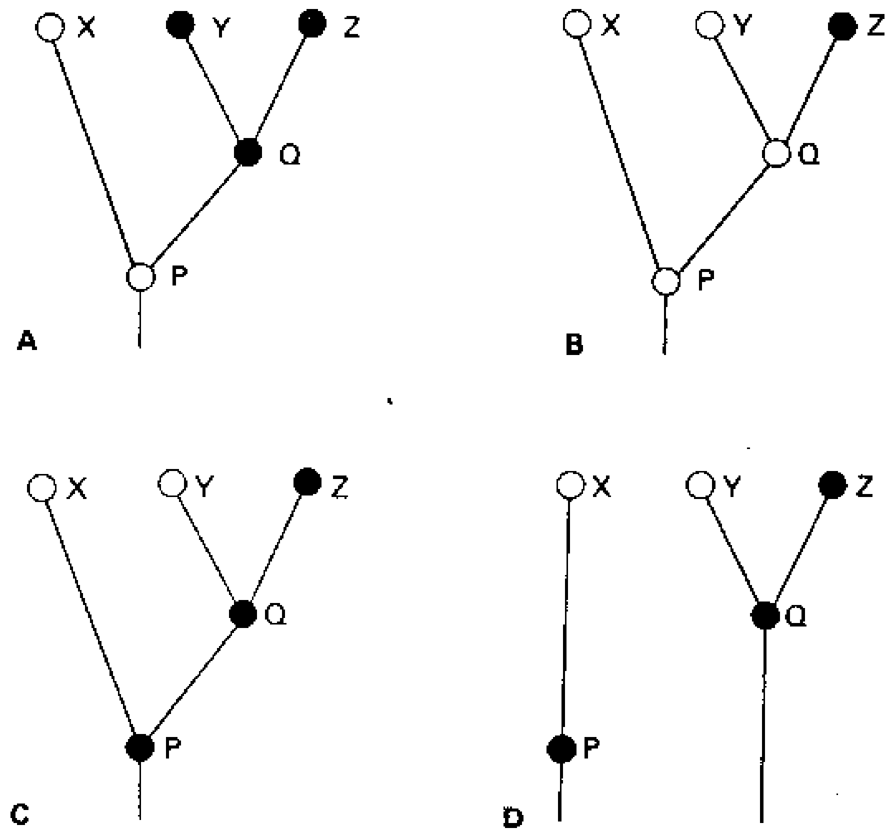


Fig. 2.6 Four diagrams showing different origins of three species (X, Y, Z) from the ancestral taxa P and Q in order to illustrate the concepts of monophyly, paraphyly, polyphyly, parallelism and convergence. The possession of one or other of two contrasting character-states by each of the five taxa is indicated by an open or closed circle respectively. **A.** Groups YZ and XYZ are both monophyletic; the similarity between Y and Z is a synapomorphy; the difference between X and YZ is due to divergence. **B.** Group XY is paraphyletic; group XYZ is monophyletic; the similarity between X and Y is a symplesiomorphy; the difference between Y and Z is due to divergence. **C.** Group XY is polyphyletic; group XYZ is monophyletic; the similarity between X and Y is a false synapomorphy caused by parallelism. **D.** Groups XY and XYZ are both polyphyletic; group YZ is monophyletic; the similarity between X and Y is a false synapomorphy caused by convergence.

- **Primitive character state**

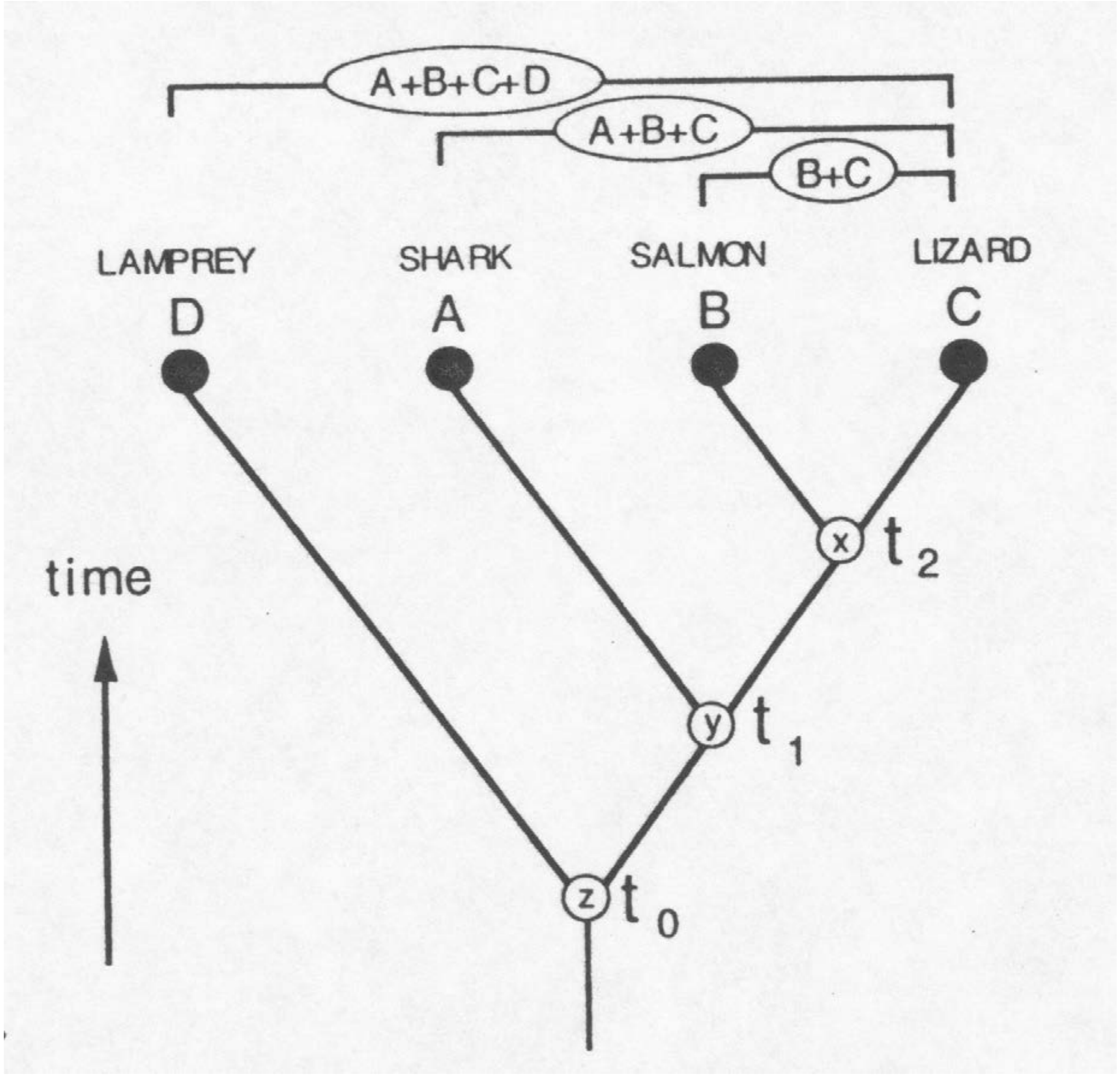
Plesiomorphy
Symplesiomorphy

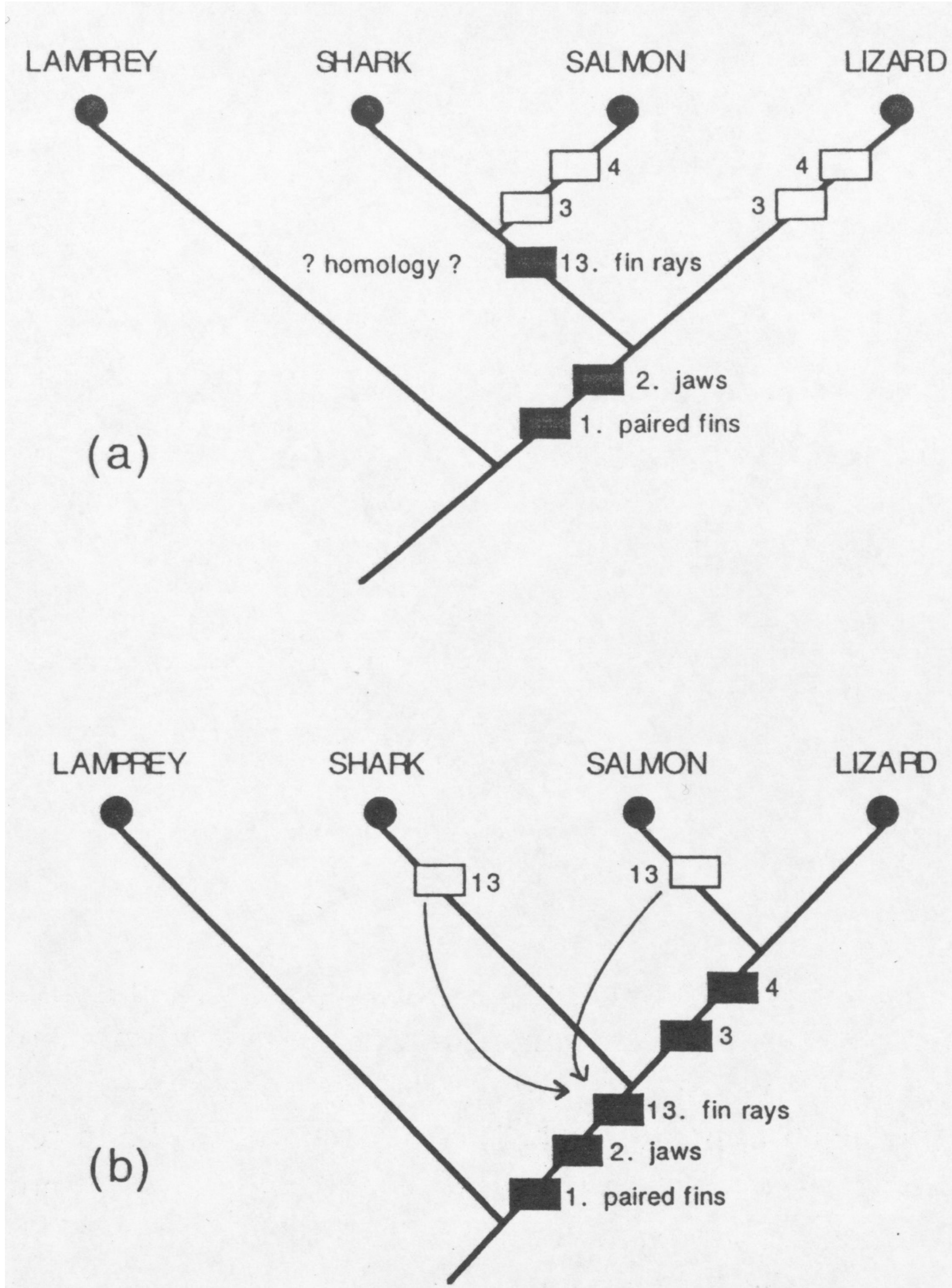
- **Derived character state**

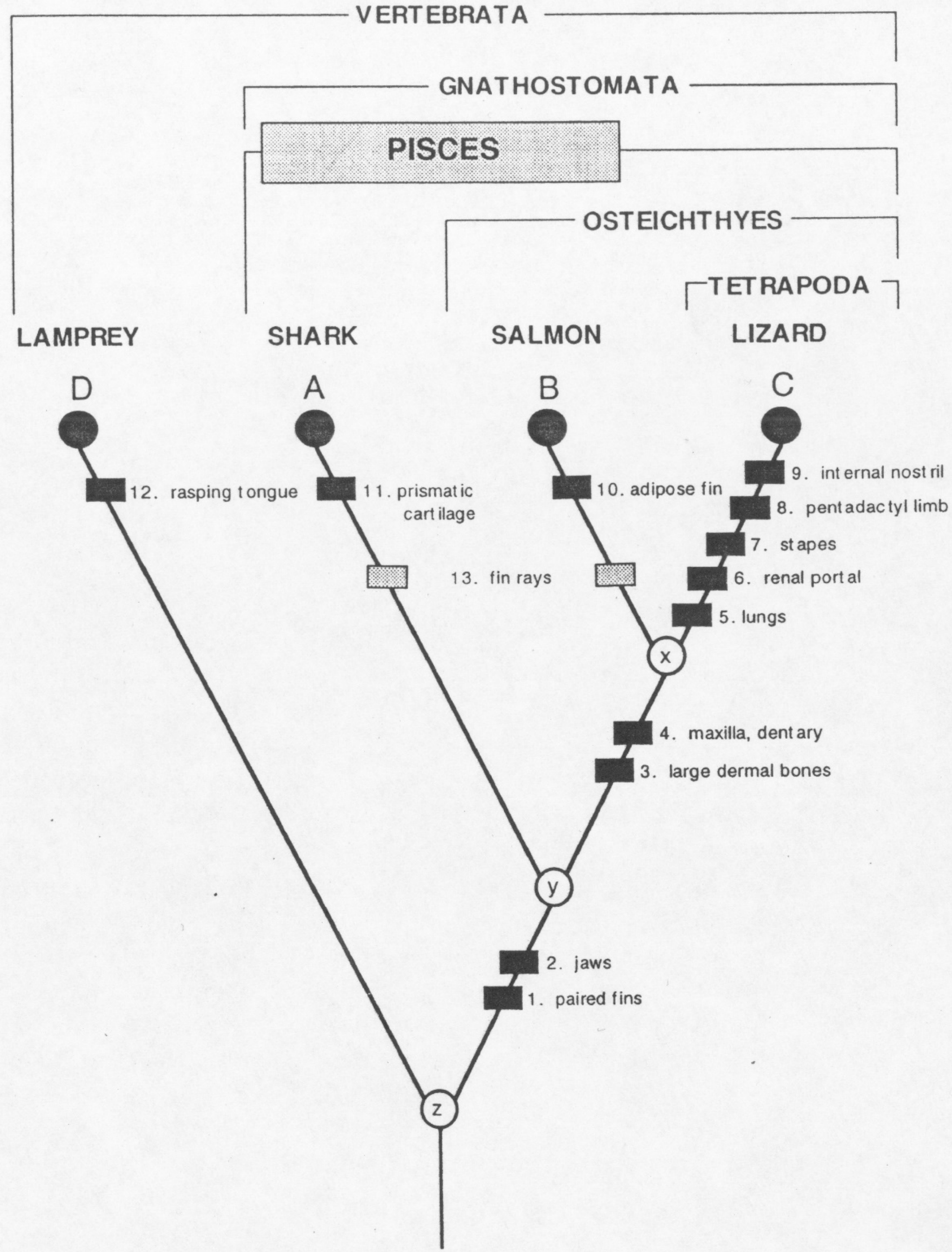
Apomorphy
Autapomorphy
Synapomorphy

- **Homoplasy** = convergences + parallelisms

- **Outgroup comparison**


A+B+C+D
A+B+C
B+C
LAMPREY
SHARK
SALMON
LIZARD
D
A
B
C
x
t_2
y
t_1
z
t_0
time


LAMPREY
SHARK
SALMON
LIZARD
4
3
4
3
? homology ?
13. fin rays
2. jaws
1. paired fins
(a)
LAMPREY
SHARK
SALMON
LIZARD
13
13
4
3
13. fin rays
2. jaws
1. paired fins
(b)


VERTEBRATA
GNATHOSTOMATA
PISCES
OSTEICHTHYES
TETRAPODA
LAMPREY
SHARK
SALMON
LIZARD
D
A
B
C
12. rasping tongue
11. prismatic cartilage
10. adipose fin
9. internal nostril
8. pentadactyl limb
7. stapes
6. renal portal
5. lungs
13. fin rays
x
4. maxilla, dentary
3. large dermal bones
y
2. jaws
1. paired fins
z

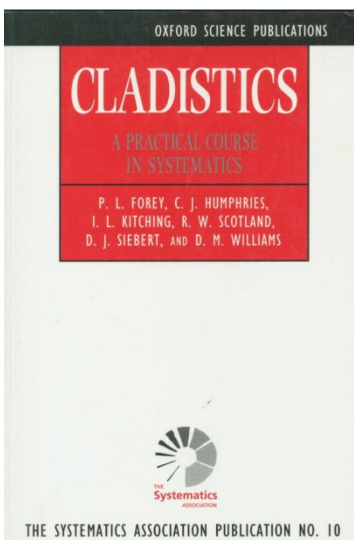
Forey, P.L., Humphries, C.J., Kitching, I.J., Scotland, R.W., Siebert, D.J. & Wiliams, D.M., 1992. *Cladistics. A practical course in systematics*. Clarendon Press, Oxford.

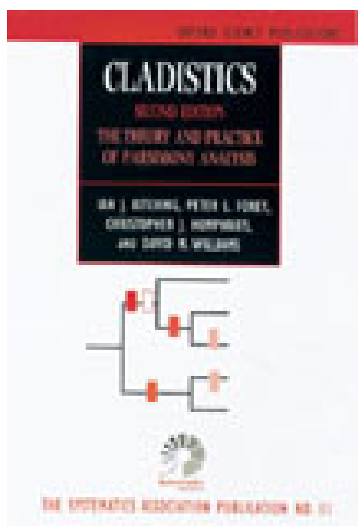
Kitching, I.J., Forey, P.L., Humphries, C.J. & Williams, D.M., 1998. *Cladistics. The theory and practice of parsimony analysis*. Ed. 2. Oxford University Press, Oxford.

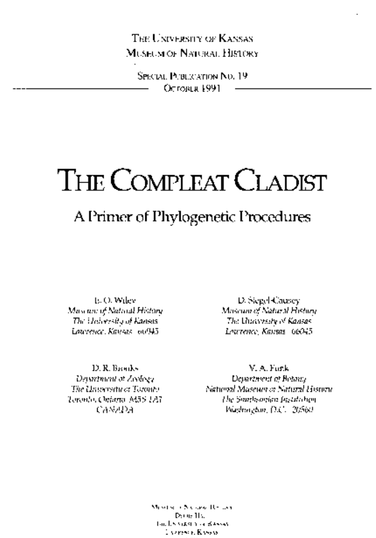
Stuessy, T. F. 1990. *Plant taxonomy: the systematic evaluation of comparative data*. Columbia University Press, New York.

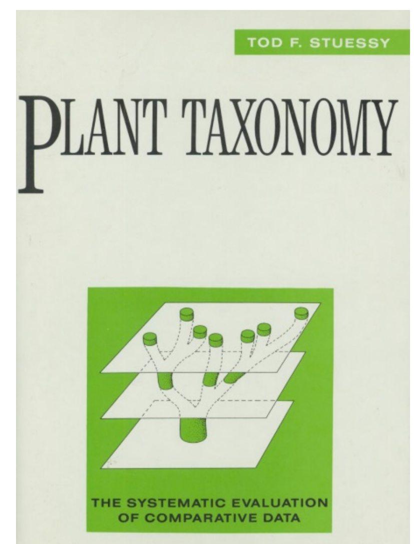
Wiley, E.O., Siegel-Causey, D., Brooks, D.R. & Funk, V.A. 1991. *The compleat cladist, a primer of phylogenetic procedures*. The University of Kansas, Museum of Natural History, Lawrence.

K dispozícii na www stránke: http://nhm.ku.edu/cc.html



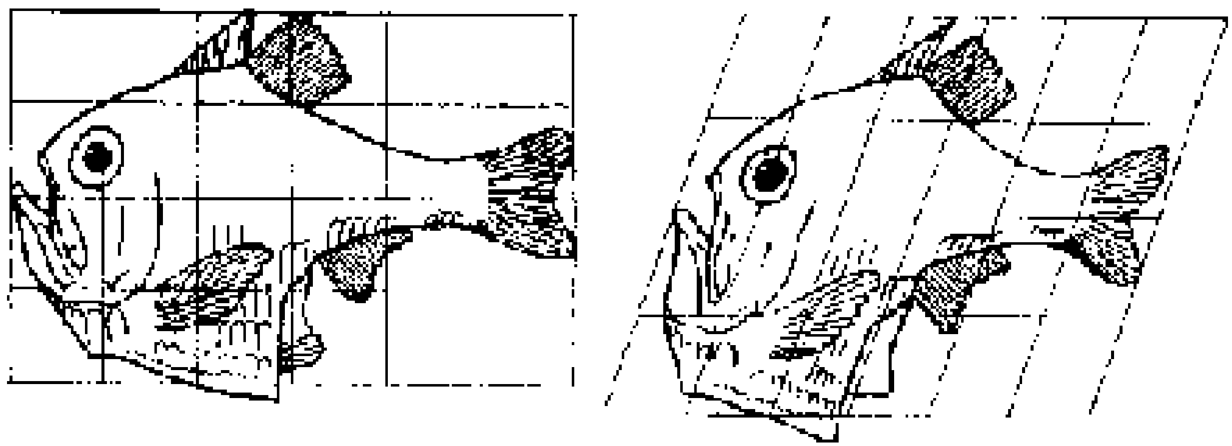






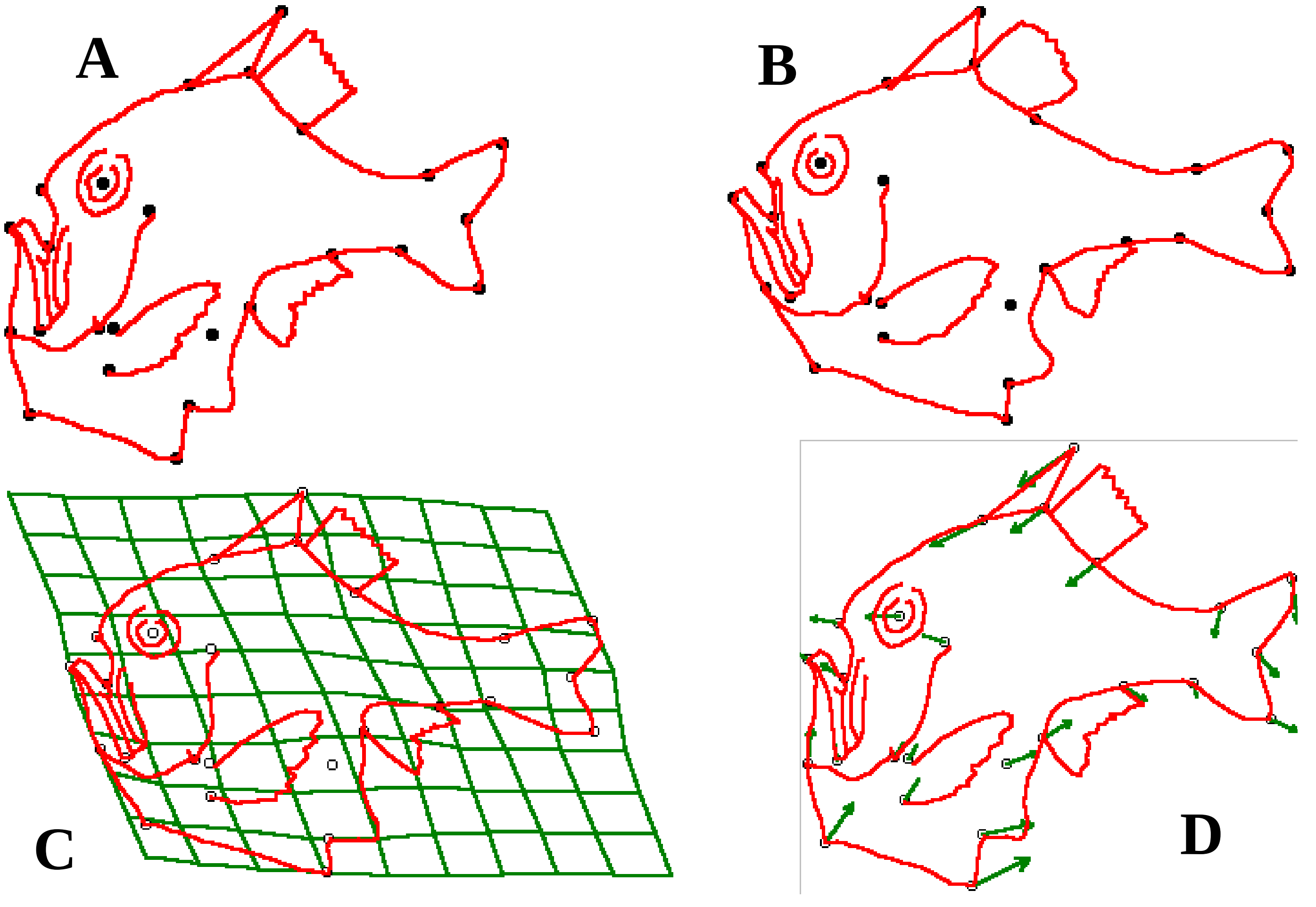
Geometric morphometrics

Thompson, A. W. 1917. ***On growth and form.*** **Cambridge University Press, Cambridge.**



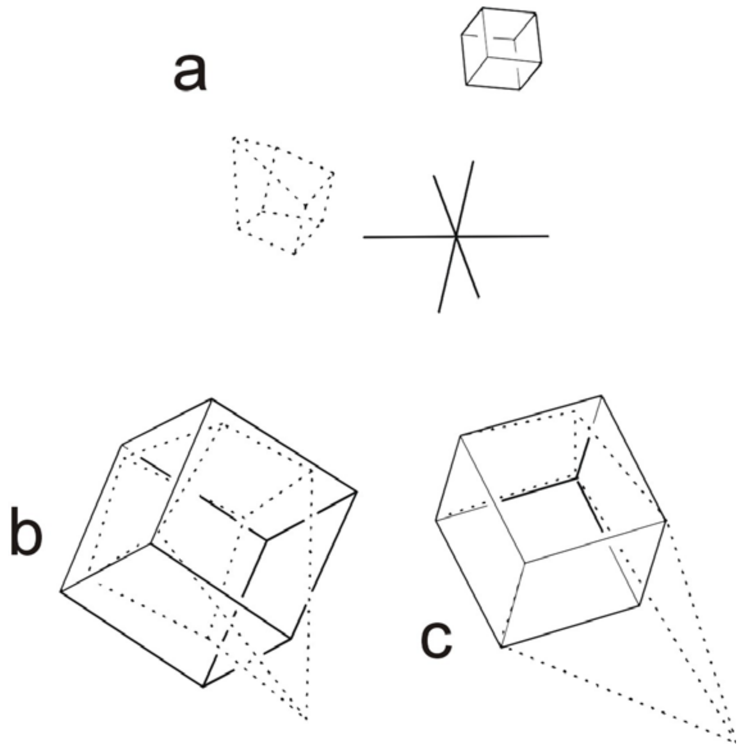
Argyropelecus olfersi. *Sternoptyx diaphana.*

Geometrická morfometrika



Mutual relationships of the shapes of the species *Stenoptyx diaphana* (A) and *Argyropelecus olfersi* (B) – sample data from the program tpsSpline (http://life.bio.sunysb.edu/morph/), C – visualization of the overall transformation using thin-plate spline, D – the same expressed using vectors.

Geometric morphometrics



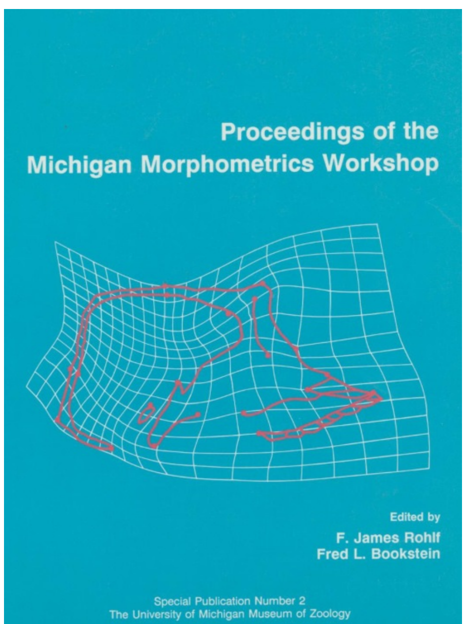
a – consensus configuration with a solid line, individual object dotted;
b – superposition using the GLS method (differences in the position of corresponding landmark points are comparable);
c – superposition using the resistant fitting method (objects differ significantly in the position of a single point).

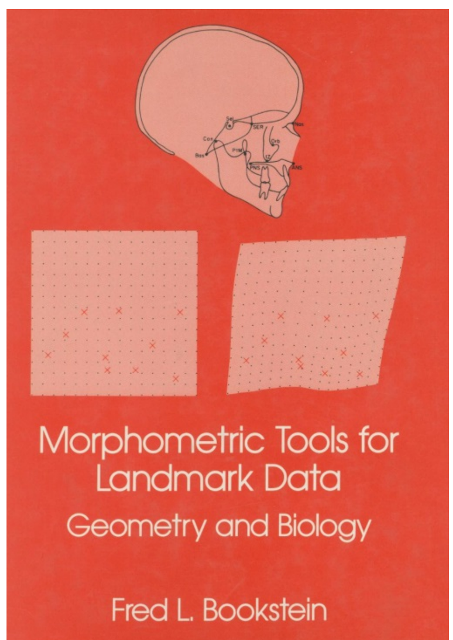
Rohlf, F.J. & Bookstein, F.L., eds., 1990. Proceedings of the Michigan morphometric workshop. *Special Publ. No. 2, The University of Michigan Museum of Zoology*. [Blue book]

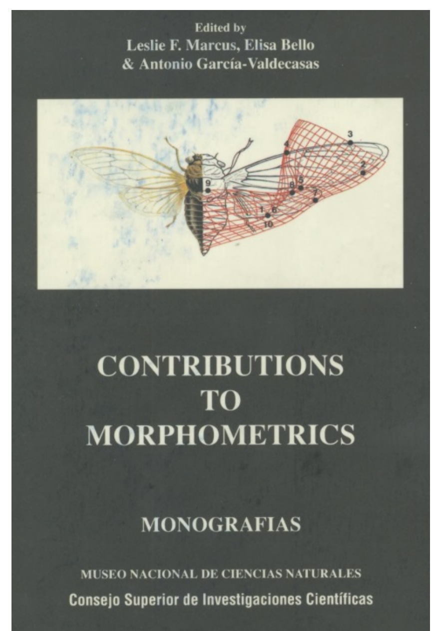
Bookstein, F.L. 1991. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, New York. [Red book]

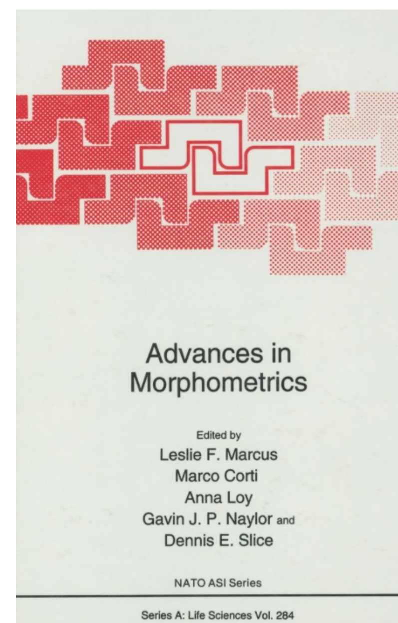
Marcus, L.F., Bello, E. & García-Valdecasas, A., eds., 1993. *Contributions to morphometrics*. Museo Nacional de Ciencias Naturales, Madrid. [Black book]

Marcus, L.F., Corti, M., Loy, A., Naylor, G.J.P. & Slice, D.E., eds., 1996. Advances in morphometrics. *NATO ASI Series A: Life Sciences* 284. [White book]









Macleod, N & Forey, P. 2002. *Morphology, shape and phylogeny*. Taylor and Francis, London, New York.

Macholán, M. 1999. Prokrustes, deformace a nová morfometrie. *Vesmír* 78: 35-39.