

**Constructing general area cladograms
by matrix representation with parsimony :
a western palearctic example**

Jean-Pierre HUGOT¹ & Jean-François COSSON²

¹ Muséum National d'Histoire Naturelle, Paris, Institut de Systématique (FR CNRS 1541), Laboratoire des Mammifères et Oiseaux, 55, rue Buffon, 75231 Paris cedex 05, France (e-mail : hugot@cimrs1.mnhn.fr). To whom correspondence and proofs should be sent.

² Laboratoire de la Faune sauvage, INRA & Laboratoire Génome et Populations, CNRS 9060, Université de Montpellier II, CC63, 34095 Montpellier cedex, France.

Abstract

Since a few years the amount of available phylogeographic data rapidly increased and the question of a general method for comparing the phylogeographic patterns of a great number of taxa over the same areas has been repeatedly addressed. In the present study, published phylogenies of eight terrestrial animal groups living in the Western Palearctic are used to build a general area cladogram, using the "Matrix Representation with Parsimony" method (MRP). The areas are not defined *a priori* but directly deduced from the source phylogenies. The results are discussed regarding previous hypotheses concerning the postglacial recolonization of Europe. The grouping of the areas fit with what is known about the influence of Quaternary glaciations on the phylogeography of European biotas, where : the organisms living in temperate regions are considered to have been restricted to southern refuges, from where these areas in which temperate European flora and fauna get extinct during the cold periods were recolonized when temperature increased. The results also suggest that : (i) the Iberic refuge may be subdivided into a northern part fitting with North Iberia and West France, and a southern part in which exchanges were possible with Northern Africa; (ii) the Italic and Balkanic refuges constituted a North Mediterranean ensemble also connected with the areas surrounding the Black sea; (iii) East France was acting as a part of the North European plain; (iv) North and South Scandinavia had different histories; and finally, (v) close relationships existed between Britain and South Scandinavia. The advantages and limits of the method are also discussed.

Keyword : phylogeography, area cladogram, MRP, Western Palearctic

Introduction

Since a thirty years the question of a general method for analyzing distribution patterns using a vicariance model of analysis has been repeatedly addressed. After HENNIG (1966) and BRUNDIN (1966) an analogy between phylogenetic and biogeographical analysis has been advocated by different authors (see HUMPHRIES, 1992, for an extensive review and discussion). Following KLUGE (1988), this conducted several authors, using different methods and coding strategies, to construct cladograms of areas using taxa as characters (see MORRONE & CRISCI, 1995 and VAN SOESTS & HADJU, 1997, for a recent review and discussion). HOVENKAMP (1997) extensively criticized these approaches and claimed that : "the taxon/area analogy presupposes the existence of "areas of endemism" which should be a subject of investigation, not should be assumed". Consequently, he proposed that " vicariance events, not areas, should be used in biogeographical analysis ", and also pointed out that previous approaches often confused the two different aims of historical biogeography that should be clearly separated methodologically : (i) to elucidate the history of areas on the basis of biological evidences, (ii) to elucidate the history of organisms on the basis of an hypothesized history of areas.

In the present study point (i) only is considered : the distributional data given by the cladistic analysis of eight lineages living in the Western Palearctic are combined in order to produce a general area cladogram. The result is discussed regarding previous hypotheses concerning the postglacial recolonization of Europe. The advantages and limits of the method are also discussed.

Material and methods

Source trees. The source trees are represented on figure 1A-H. Their specifications are given on table 1. As proposed by HOVENKAMP (1997) we define the areas by enumerating from the source cladograms, all these locations in which sister groups of a particular taxon occupy mutually exclusive areas. This also fit with ROSEN (1988) recommendation : " ... decisions about the boundaries [of areas] have usually been arbitrary or generalized. Rather than trying to define such regions at the outset, it is better to work entirely with sample localities". This resulted in the identification of the 19 areas described on figure 1I. When, in a source study, a particular genotype was identified in several different areas, the corresponding areas were represented as an unresolved polytomy in the source cladogram (Figs 1C, F, G & H). This is corresponding with ZANDEE & ROSS (1987) "assumption zero" : "widespread taxa are considered as synapomorphies for the areas inhabited by them" (MORRONE & CRISCI, 1995).

Computation of the area cladogram. Computation of the area cladogram was performed using the method described as "*Construction of supertrees by Matrix Representation with Parsimony (MRP)*" by SANDERSON *et al.* (1998).

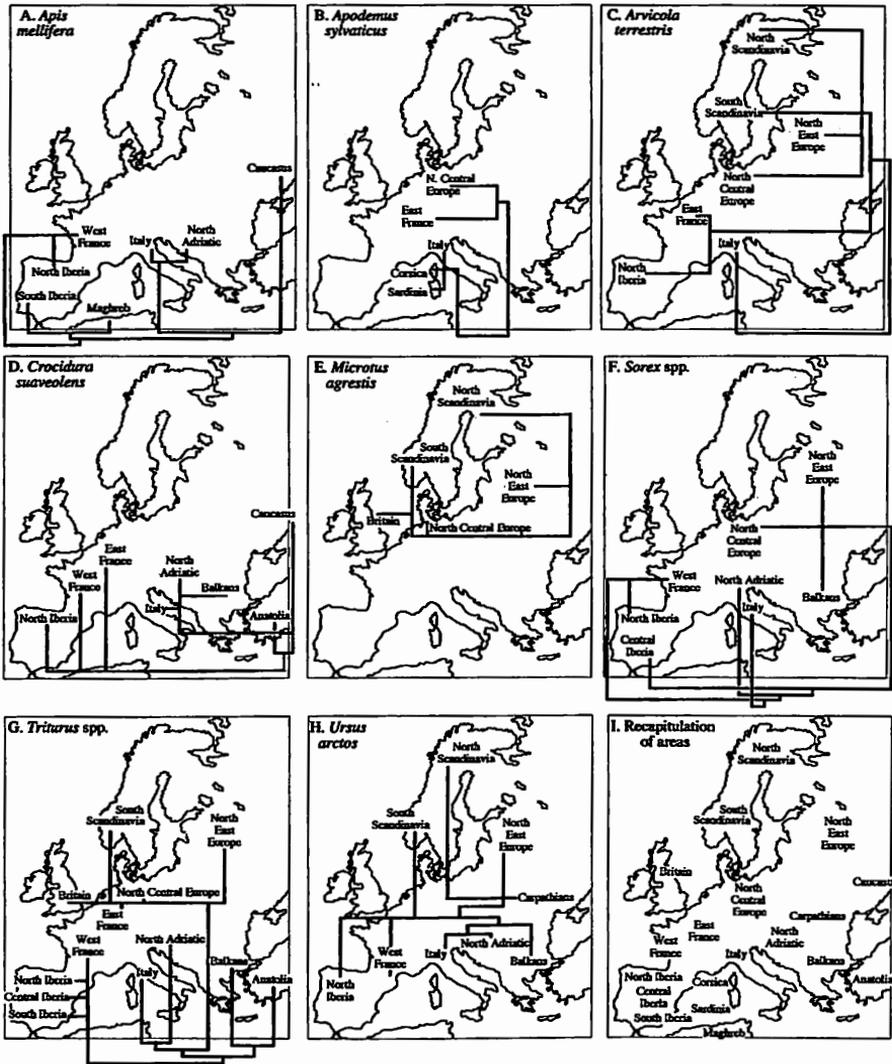


Fig. 1A-H. Area source trees. I, recapitulation of geographic areas. (see Table 1 for specifications).

This method first has been proposed by BROOKS (1981) for analyzing host phylogenies by using the phylogenies of their parasites, as : "*Brooks Parsimony Analysis*" (BPA). Later, WILEY (1987) amended BPA for biogeographical purposes. This transposes to biogeographical areas, DOYLE'S (1992) assumption that the results of different studies of the phylogenetic relationships of a particular group of taxa can be considered as equivalent phylogenetic hypothesis and directly combined using parsimony analysis in

order to produce a composite estimate of the phylogeny of this group. This method also has been independently proposed as the "*Matrix Representation Method*" by BAUM (1992) and RAGAN (1992), for inferring phylogenies of organisms from gene trees.

Table 1. Specification for the source trees. N° = number of areas represented in each particular source study (out of 19).

<i>tree on figure</i>	N°	<i>taxa</i>	<i>source tree after</i>	<i>character type</i>	<i>method for analysis</i>
1.A	7	<i>Apis mellifera</i>	GARNERY <i>et al.</i> (1992)	mtDNA	parsimony
1.B	5	<i>Apodemus sylvaticus</i>	MICHAUX <i>et al.</i> (1996)	mtDNA	UPGMA
1.C	9	<i>Arvicola terrestris</i>	TABERLET <i>et al.</i> (1998)	mtDNA	parsimony
1.D	8	<i>Crocicidura suaveolens</i>	TABERLET <i>et al.</i> (1998)	mtDNA	parsimony
1.E	5	<i>Microtus agrestis</i>	JAAROLA & TEGELSTRÖM (1995)	mtDNA	parsimony
1.F	9	<i>Sorex</i> spp.	TABERLET <i>et al.</i> (1994)	mtDNA	parsimony
1.G	11	<i>Triturus</i> spp.	WALLIS & ARNTZEN (1989)	mtDNA	parsimony
1.H	9	<i>Ursus arctos</i>	TABERLET & BOUVET (1994)	mtDNA	parsimony

Different coding strategies have been experimented and compared by RONQUIST (1996) for matrix representation of trees. Here we follow RONQUIST (1996) recommendations : (i) each original source tree is represented by a character matrix using additive binary coding; (ii) newly coded trees are combined into a single matrix, areas not represented in a particular source tree being coded as having states unknown for that tree, which is corresponding with NELSON & PLATNICK (1981) "assumptions 1 and 2"; (iii) an all-zero outgroup is added to root the trees and the resulting data matrix is subjected to parsimony analysis. Representation of the original source trees in a character matrix using additive binary coding is given in table 2.

Parsimony analysis. Analysis was conducted using test version 4.0d64 of PAUP* (SWOFFORD, 1998). Runs were performed using heuristic search tree-bisection-reconnection (TBR) branch-swapping, addition sequence random = 10 replicates, MULPARS option in effect. MACCLADE 3.07 (MADDISON & MADDISON, 1997) and TreeView 1.5 (PAGE, 1996) were used for data and tree handling and for computation of tree statistics.

Results

The analysis gave 970 trees from which a majority rule consensus tree was computed. Tree length = 50. Consistency index (CI) = 0.72. Retention index (RI) = 0.81. Rescaled consistency index (RC) = 0.59. On figure 2 the majority rule consensus tree is superimposed over a map of the Western Palearctic, on which the mountain barriers, and the southern maximum extension of the permafrost during the last glacial age also have been represented. The majority rule frequency is given for each node.

Table 2. Matrix representation of area source trees as represented on figure 1. Abbreviations : api = *Apis mellifera*; apo = *Apodemus sylvaticus*; arv = *Arvicola terrestris*; cro = *Crocidura suaveolens*; mic = *Microtus agrestis*; srx = *Sorex* spp.; tri = *Triturus* spp.; urs = *Ursus arctos*.

	aaaaa	aa	aaa	cccc	mmm	ssssss	ttttttt	uuuuu
	ppppp	pp	rrr	rrrrr	iii	rrrrrr	rrrrrrr	rrrrr
	iiii	oo	vvv	oooo	ccc	xxxxx	iiiiiii	sssss
OUTGROUP	00000	00	000	00000	000	000000	0000000	00000
SOUTH. IBERIA	01001	??	000	?????	???	???????	0000001	?????
NORTH. IBERIA	10000	??	011	00001	???	000111	0000011	10010
CENTR. IBERIA	?????	??	???	?????	???	011011	0000011	?????
NORTH. ADRIATIC	00111	??	???	11110	???	001011	0110100	01110
BALKANS	?????	??	???	11110	???	111011	0001100	00110
CARPATHIANS	?????	??	???	?????	???	???????	???????	00001
E. FRANCE	?????	01	011	00001	???	111011	1010100	?????
W. FRANCE	10000	??	000	00001	???	000111	0000011	10010
BRITAIN	?????	??	???	?????	110	???????	1010100	?????
N. CENTR. EUROPA	?????	01	101	?????	010	111011	1010100	?????
NE. EUROPA	?????	??	101	?????	001	111011	1010100	00001
SOUTH. SCAND	?????	??	101	?????	110	???????	1010100	10010
NORTH. SCAND	?????	??	101	?????	001	???????	???????	00001
ANATOLIA	?????	??	???	00010	???	???????	0001100	?????
ITALY	00111	10	000	01110	???	000001	0110100	01110
CAUCASUS	00011	??	???	00110	???	???????	???????	?????
CORSICA	?????	10	???	?????	???	???????	???????	?????
SARDINIA	?????	10	???	?????	???	???????	???????	?????
MAGHREB	01001	??	???	?????	???	???????	???????	?????

The basal dichotomies of the area supertree have a frequency of 100% and allow to distinguish four main groups. *Group.1* is composed of North Iberia and West France. *Group.2* is composed of Central Iberia, South Iberia and Maghreb. *Group.3* associates : the peri Tyrrhenean's, peri Adriatic's and peri Aegean's areas including Corsica, Sardinia, Italy, North Adriatic, Balkans and Anatolia, together with Caucasus. *Group.4* associates all the areas of North and Central Europe, separated into three branches also having a frequency of 100% : the first branching isolates East France, the last two branches separate the other areas in two subgroups; *subgroup.1*, includes North Scandinavia, North East Europe and Carpathians, *subgroup.2* includes Britain, South Scandinavia and North Central Europe. Within *Group.2* and *Group.3*. the distribution of the component areas has a lower frequency (<63%).

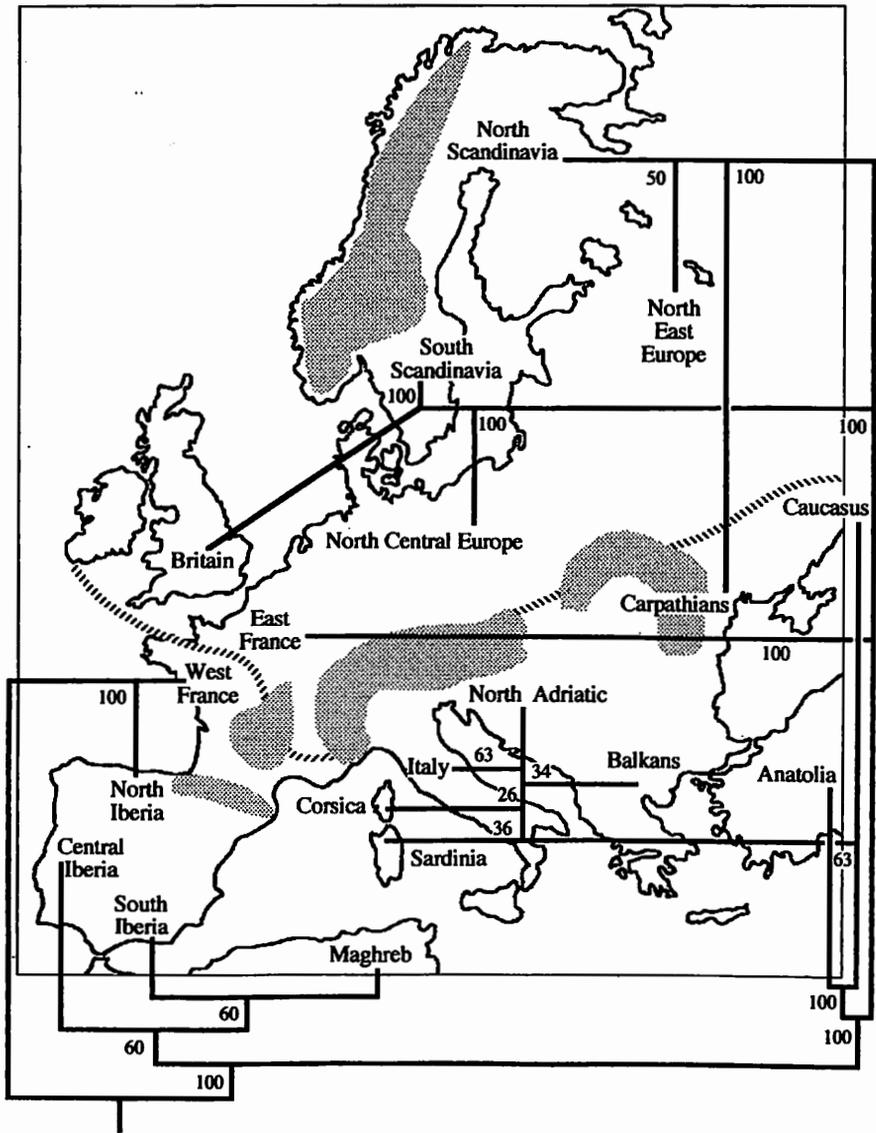


Fig. 2. Area supertree of the Western Palearctic. Numbers indicate the value of the majority rule frequency for each node. Thick dash line : southern maximum extension of permafrost in Europe during the last cold period, 20,000-18,000 years ago after FRENZEL *et al.* (1992) and LUNDQVIST & SAARNISTO (1995). Grey areas : mountain barriers.

Discussion

The grouping of the areas fit with what is known about the influence of Quaternary glaciations on the phylogeography of European biotas (TABERLET *et al.*, 1998). The organisms living in temperate regions are considered to have been restricted to southern refuges during the Quaternary glacial periods (HEWITT, 1996). *Group.1*, *Group.2* and *Group.3* match with the Southern Iberic, Italic and Balkanic refuges which, classically, have been distinguished in Western Europe. *Group.4* is corresponding with these areas covered by the ice sheets and the permafrost during their maximum extension, in which temperate European flora and fauna get extinct during the cold periods (FRENZEL *et al.*, 1992; LUNDQVIST & SAARNISTO, 1995) and which were recolonized when temperature increased, either by populations coming from southern areas or from more eastern areas (HEWITT, 1996). *Group.1* and *Group.2* also suggest that the Iberic refuge may be subdivided into a northern part fitting with North Iberia and West France, and a southern part in which exchanges were possible with Northern Africa. The topology of *Group.3* suggests that the Italic and Balkanic refuges constituted a North Mediterranean ensemble also connected with the areas surrounding the Black sea. The topology of *Group.4* advocates : for East France acting as a part of the North European plain; for the recognition of an eastern and a western parts corresponding respectively with *subgroup.1* and *subgroup.2*; for a different history for North and South Scandinavia and also for close relationships between Britain and South Scandinavia. However, the placement of several areas have to be considered with caution. This is the case for the relationships between South Iberia and Maghreb, between Britain and South Scandinavia and for the placement of Corsica and Sardinia which only rely upon 1 or 2 phylogeographies.

Conclusion

The influence of Quaternary glaciations appears to be the main pattern explaining the current distribution of the eight different taxa used as sources in this study. This relies with HOVENKAMP (1997) assumption that : "At each moment in history, ...[taxa]... are constrained only by the then-operative barriers, and not by the historical origin of the areas they inhabit". This means that generally in this type of study : the closest biogeographical events are from present, the greatest chance they have to be reflected by the results of the analysis. This also probably indicates that tracks of the ancient origin of taxa progressively disappear as more recent biogeographical events accumulate. As underlined by HOVENKAMP (1997) this constitutes a notable difference between the history of characters which, with time, are much more restricted to the clade in which they originated, and the history of the distribution of taxa which first steps can be completely and definitively hidden by more recent episodes.

Using a definition of the areas by enumerating the vicariance events from sets of source cladograms and assembling the source cladograms using the

MRP method, allow to conserve all or most of the information present in the source trees even when each of them only contains a subset of the terminal areas being considered. Thus, this method allows to place a particular area on the cladogram even if this area is represented only once in the source trees and to make statements about the relationships of areas which do not co-occur on any of the source trees. In addition, this method also easily allows to add new source trees and (or) new areas, or to subdivide a previous area into several new components when necessary. Therefore, this method looks to be reliable for producing area cladograms on the basis of biological evidences.

The procedure proposed here allows to easily evidence the main tendencies which have influenced the establishment of complex faunas, to define groups of areas without *a priori*, to investigate which biotas have a divergent pattern of dispersion and to group those organisms which dispersion routes are similar. However, we are appreciative that the validity of the method is dependent of the reliability of the general area cladogram which in turn is linked to the abundance of the available source studies. At the moment, these studies whose results can be represented as reliable area cladograms are relatively few, but their number is currently rapidly increasing. Thus, it would be possible to test the method in better conditions in a close future.

Acknowledgments

We thank Pierre DELEPORTE for help and comments on early drafts of this manuscript. This work was supported by the "Groupe de Travail Morphométrie et Analyse de Forme" du Muséum National d'Histoire Naturelle.

References

- BAUM B.R., 1992. - Combining trees as a way of combining data sets for phylogenetic inference, and the desirability for combining gene trees. *Taxon*, 41 : 1-10.
- BROOKS D.R., 1981. - Hennig's parasitological method : A proposed solution. *Systematic Zoology*, 30 : 229-249.
- BRUNDIN L., 1966 - Transantarctic relationships and their significance as evidenced by chironomid midges with a monograph of the subfamilies Podonominae and Aphroteninae and the austral Heptagyiae. *Kungliga Svenska Vetenskapsakademiens Handlingar, Fjärde Serien*, 11 : 1-472.
- DOYLE J.J., 1992. - Gene trees and species trees : molecular systematics as one-character taxonomy. *Systematic Botany*, 17 : 144-163.
- FRENZEL B., PÉCSI P. & VELICHKO A.A., 1992. - *Atlas of Paleoclimates and paleoenvironments of the Northern hemisphere. Late Pleistocene-Holocene*. Geographical Research Institute, Hungarian Academy of Sciences. Gustav FISCHER Verlag, Budapest, Stuttgart.
- GARNERY L., CORNUET J.-M. & SOLIGNAC M., 1992. - Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, 1 : 145-154.

- HENNIG W., 1966. - *Phylogenetic systematics*. University of Illinois Press, Urbana, Illinois.
- HEWITT G.M., 1996. - Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58 : 247-276.
- HOVENKAMP P., 1997. - Vicariance events, not areas, should be used in biogeographical analysis. *Cladistics*, 13 : 67-79.
- HUMPHRIES C.J., 1992. - Cladistic biogeography. In "*Cladistics a practical course in systematics*". (FOREY, P. L. *et al.* eds.), Chapter 9. Clarendon Press, Oxford.
- JAAROLA M. & TEGELSTRÖM H., 1995. - Colonization history of north European field voles (*Microtus agrestis*) revealed by mitochondrial DNA. *Molecular Ecology*, 4 : 299-310.
- KLUGE A.G., 1988. - Parsimony in vicariance biogeography : a quantitative method and a greater Antillean example. *Systematic Zoology*, 37 : 315-328.
- LUNDQVIST J. & SAARNISTO M., 1995. - Summary of project IGCP-253. *Quaternary International*, 28 : 9-18.
- MADDISON W.P. & MADDISON D.R., 1997. - MacClade : Analysis of phylogeny and character evolution. Version 3.07. Sinauer Associates, Sunderland, Massachusetts.
- MICHAUX J.R., FILIPPUCCI M.-G., LIBOIS R.M., FONS R. & MATAGNE R.F., 1996. - Biogeography and taxonomy of *Apodemus sylvaticus* (the woodmouse) in the Tyrrhenian region : enzymatic variations and mitochondrial DNA restriction pattern analysis. *Heredity*, 76 : 267-277.
- MORRONE J.J. & CRISCI J.V., 1995. - Historical Biogeography : introduction to methods. *Annual Review for Ecology and Systematics*, 26 : 373-401.
- NELSON G. & PLATNICK N.I., 1981. - Systematics and biogeography : cladistics and vicariance. Columbia University Press, New York.
- PAGE R.D.M., 1996. - TreeView : An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, 12 : 357-358.
- RAGAN M.A., 1992. - Phylogenetic inference based on matrix representation of trees. *Molecular Phylogenetic and Evolution*, 1 : 53-58.
- RONQUIST F., 1996. - Matrix representation of trees, redundancy and weighting. *Systematic Biology*, 45 : 247-253.
- ROSEN B.R., 1988. - From fossils to earth history : applied historical biogeography. In "*Analytical Biogeography. An integrated approach to the study of animal and plants distributions*". (MYER & GILLERS eds.), chapter 14. Chapman & Hall, London.
- SANDERSON M.J., PURVIS A. & HENZE C., 1998. - Phylogenetic supertrees : assembling the trees of life. *TREE*, 13 : 105-109.
- SWOFFORD D., 1998. - PAUP Phylogenetic Analysis Using Parsimony, ver. 4.0d64. Smithsonian Institution, Laboratory of Molecular Systematics, MSC, MRC-534, 20560, Washington DC.
- TABERLET P. & BOUVET J., 1994. - Mitochondrial DNA polymorphism, phylogeography and conservation genetics of the brown bear (*Ursus arctos*) in Europe. *Proceedings of the Royal Society, London B.*, 255 : 195-200.
- TABERLET P., FUMIGALLI L. & HAUSSER J., 1994. - Chromosomal versus mitochondrial DNA evolution : tracking the evolutionary history of the southwestern European populations of the *Sorex araneus* group (Mammalia, Insectivora). *Evolution*,

- 48 : 623-636.
- TABERLET P., FUMIGALLI L., WUST-SAUCY A.-G. & COSSON J.-F., 1998. - Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, 7 : 453-464.
- VAN SOESTS R.W.M. & HAJDU E., 1997. - Marine area relationships from twenty sponge phylogenies. A comparison of methods and coding strategies. *Cladistics* 13 : 1-20.
- WALLIS G.P. & ARNTZEN J.W., 1989. - Mitochondrial-DNA variation in the crested newt superspecies : limited cytoplasmic gene flow among species. *Evolution*, 43 : 88-104.
- WILEY E.O., 1987. - Methods in vicariance biogeography. In *Systematics and Evolution : a matter of diversity* (P. Hovenkamp, ed.) Pages 283-306 . Utrecht University, Utrecht.
- ZANDEE M. & ROSS M.C., 1987. - Component-compatibility in historical biogeography. *Cladistics*, 3 : 305-332.