

**Reconstructing Biogeographic History Using Phylogenetic-Tree Analysis of
Community Structure**



Pierre Legendre

Systematic Zoology, Vol. 35, No. 1 (Mar., 1986), 68-80.

Stable URL:

<http://links.jstor.org/sici?sici=0039-7989%28198603%2935%3A1%3C68%3ARBHUPA%3E2.0.CO%3B2-X>

Systematic Zoology is currently published by Society of Systematic Biologists.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ssbiol.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

For more information on JSTOR contact jstor-info@umich.edu.

©2003 JSTOR

RECONSTRUCTING BIOGEOGRAPHIC HISTORY USING PHYLOGENETIC-TREE ANALYSIS OF COMMUNITY STRUCTURE

PIERRE LEGENDRE

*Département de Sciences biologiques, Université de Montréal,
C.P. 6128, Succursale A, Montréal, Québec H3C 3J7, Canada*

Abstract.—The dispersal of a community in a territory can be reconstructed using various more or less specific methods of data analysis. When the data are in the form of presence/absence of species for the various units of territory, or in the form of frequencies of occurrence, they are analogous to the data used for reconstructing evolutionary trees. Furthermore, when one is looking for a treelike structure of dispersal, issued from a common trunk, the tree-reconstruction methods may represent an appropriate analytical tool. Four tree-reconstruction methods are found to be appropriate in their hypotheses. They lead to satisfactory trees when applied to a data set describing the geographic distribution of the stenohaline freshwater fishes of the Québec peninsula, previously divided into 21 regions. For binary data, the method of Camin and Sokal and the polymorphism parsimony method were used with success, while the continuous character data maximum-likelihood method and the least-squares method of Cavalli-Sforza and Edwards were applied to frequency-of-occurrence data. When compared to a method of pairwise comparison of adjacent regions, it is shown that the tree-reconstruction methods provide the user with information that is not found otherwise. I also show that external information (i.e., geographic adjacencies, and the degree of asymmetry of the faunal vectors of regions connected by the tree), available in biogeographic data but not taken into account by the tree-reconstruction algorithms, must be considered to render the results meaningful. [Biogeography; cladistic methods; freshwater fishes; methods of phylogenetic inference; postglacial dispersal; tree-reconstruction methods.]

Biogeographic reconstruction of the dispersal of a community through a territory can be studied by several methods, whose interest and complexity grow as more assumptions have to be made, and more constraints are built into the analysis. The easiest but least-informative way consists of using the well-known general techniques of data analysis (clustering and ordination). The method of spatial autocorrelation (Cliff and Ord, 1981) is useful when a process can be hypothesized that may have generated gradients, or a patchy structure; this method cannot handle multidimensional data, however, so that its usefulness in community analysis is limited, except through the use of synthetic parameters of community structure (but see also Wartenberg, 1985). Finally, community dispersal can be approached through a physical model of diffusion flow, for instance by the pairwise comparison of the fauna or flora found in adjacent and relatively homogeneous units of territory

(as in the "coefficient of dispersal direction" of Legendre and Legendre, 1984).

Behind any such analysis lies the assumption that the present distribution of organisms contains information relating to the process by which they got to where they are now found. The presence or absence of a species in a region can be regarded as a descriptor that changes according to some model, rather than at random. One can postulate, for example, that the diffusion of species from one region to its neighbors is a process that can explain the appearance of species in a given region, while the disappearance of a species from an area causes the reversal of the species descriptor.

There are circumstances where the methods designed for inferring phylogenetic trees can be seen as appropriate for biogeographic analysis. This is the case when one is looking for a treelike structure of geographic dispersal, issued from a single trunk, and describing the move-

ment of species through adjacent areas. The purpose of the present paper is to investigate this application of phylogenetic-reconstruction methods. The assumptions and conditions of application of these methods will be presented. Data on the present-day distribution of freshwater fishes in Québec will be used as illustration, after which conclusions will be drawn as to the application of the tree-reconstruction methods in biogeographic studies.

MATERIALS AND METHODS

Based on the actual distribution maps of 109 species of freshwater fishes in the Québec peninsula, the presence and absence of these species was noted for each of 289 1°-square quadrats, or pixels ("picture cells") of the peninsula. In a previous paper, the pixels were clustered into 21 relatively homogeneous regions (Legendre and Legendre, 1984). Among the 109 species, 85 are stenohaline (restricted to fresh water) so that they can be thought of as having reinvaded this recently deglaciated territory in a dendritic pattern, through inland routes, from the glacial refugia of the Great Lakes and New England. Only those 85 species will be used below to decipher which roads were mainly followed during this postglacial dispersal process.

The operational geographic units (OGUs; Crovello, 1981) of this study are the 21 regions. When frequencies of occurrence will be required, these will be taken as the proportion of the given region's pixels where the species occurs; the data set (regions \times species) is presented in table 2 of Legendre and Legendre (1984). When using phylogenetic methods for binary data, these frequencies will be transformed into binary form by cutting at the 50% point, because this division has been found to be the most informative (Legendre and Legendre, 1984:1790). Furthermore, since the root of the dispersal tree is assumed to be the glacial refugia where all species occurred (reference is given in the Results), the "ancestral" state (coded 0) will be given to an area where the species is found

with a frequency higher than 50%, while the "derived" state (coded 1) corresponds to a frequency smaller than or equal to 50%.

When the data are in the form of presence/absence of species in the various units of territory, they resemble the binary data used by several methods for inferring phylogenies. These methods are often called *cladistic methods*, but I prefer to reserve the term *cladism* to refer to the Hennigian philosophy of classification. Most of these methods pertain to two families, *parsimony* and *compatibility*, designed to deal with the incompatibility of some characters with the phylogeny reconstructed by a subset of compatible characters. When *homoplasy*, that part of the similarity among objects (called OTUs in evolutionary studies) which is due to parallelism or convergence, is concentrated in only a few characters, compatibility methods should be used to eliminate incompatible characters before reconstructing the phylogeny. If one can reasonably assume that homoplasy is distributed at random among characters, parsimony methods, which minimize the total number of evolutionary steps in the tree, should be used instead. This is indeed the case with biogeographic data related to a single taxocene of organisms, such as the fish distribution data described above.

Most methods of phylogenetic-tree reconstruction assume that different characters evolved independently; this may be only partly true in ecological communities, because of the trophic links and other factors determining the connectedness of species (Allen and Starr, 1982; Leps et al., 1982; Pimm, 1984). The methods further assume that different branches of the tree evolved independently from one another, thus preventing reticulate interactions between branches after their point of splitting off. Taking this as a simplifying working assumption, its consequences will be discussed below. Several tree-reconstruction methods will be considered, some which call for binary data, while others can handle quantitative information.

The Camin and Sokal (1965) parsimony

algorithm searches for the tree that minimizes the number of mutations from ancestral to derived states, which corresponds in this study to the number of species lost along the lineages. Other kinds of evolutionary changes, such as retention of polymorphism, are not taken into account because they are seen as far less probable than a 0-to-1 change (loss of a species). This method requires that the ancestral state be known for each descriptor, in order to produce a rooted tree. In the present study, the ancestral state is "species present," as explained above. The method does not allow for reversals, which means here the reacquisition of a species (above the 50% mark) after it has been lost (below or equal to 50%). Since these conditions represent acceptable working assumptions, the Camin and Sokal parsimony method will be used in this study.

The Wagner parsimony method (Kluge and Farris, 1969) does not require that ancestral states be known, so that it can produce an undirected or unrooted tree, also called a network. The Wagner network was rooted by including a vector of ancestral states (0, presence of all species) in the analysis, and by drawing the tree from that trunk. Since character polarity is not taken into account, reversals are permitted, as well as repeated 0-to-1 mutations of the same species in different branches of the tree; actually, the algorithm tries to find the tree minimizing the number of these evolutionary events. In the present study, a reversal can be the result of three different biogeographic events: (a) a loss of a species (below or equal to 50%) followed by a reacquisition (occupancy of more than 50% of the pixels in the next region); (b) a loss in an intermediate region, after the species has reached regions located further up the tree; or (c) a reticulate pattern of dispersal if a species has reached a region from another area that is geographically adjacent but is not adjacent along the tree. In the example being analyzed, reticulate dispersal seems to have occurred in a few cases and would have to be represented by a single-rooted acyclic directed graph instead of a tree. These graphs were

named *reticulograms* by Lefkovitch (see Legendre, 1984).

The polymorphism parsimony method (Farris, 1978; Felsenstein, 1979) allows incompatibilities to be explained by the existence of polymorphism within evolving populations, noted as an intermediate state between 0 and 1. The algorithm tries to find the tree that minimizes the duration of polymorphism. This model is well adapted to biogeographic data, especially when coded by reference to a 50% threshold as in the present study, despite the fact that a single 0-to-1 change is allowed for each descriptor. Although this is too restrictive for biogeographic studies (a species should be allowed to be lost more than once), the method will be used nevertheless in order to assess the influence of this limitation.

Two methods of phylogenetic reconstruction that require quantitative species frequency data will also be investigated. The first is the continuous character data maximum-likelihood method (CONTML) of Felsenstein (1973, 1981). The method is based on the genetic model of Edwards and Cavalli-Sforza (1964), which explains the independent evolution of characters in different lineages by genetic drift. Drift is modeled by a random Brownian motion. This method was designed to handle gene frequencies in populations, which become species frequencies in communities, in the present study. The estimation of the tree topology for which the observed data are most likely is obtained by a maximum-likelihood statistical procedure. The branch-length estimates are measured in amounts of expected accumulated variance along each branch, after an arcsine-square-root transformation of the original frequencies, as recommended by Sokal and Rohlf (1981). Each branch length is accompanied by a confidence-interval estimate, so that a zero length may be attributed to branches with negative lower bounds. This tree was rooted by postulating an ancestral region in which all species were present in 100% of the pixels.

Finally, the least-squares distance method of Cavalli-Sforza and Edwards (1967)

was used. This method was originally designed as a workable alternative to the problem of finding the shortest network connecting OTUs in a gene-frequency space. This network actually represents the projection, in the phenetic hyperplane, of a cladistic tree whose temporal axis is orthogonal to the plane. Instead of a matrix of genetic distances, I used a matrix of ecological distances among regions, computed with Odum's (1950) percentage difference distance, which is actually the one-complement of Steinhaus' similarity coefficient (Motyka, 1947). This distance function is well adapted to species-frequency data, as discussed by Legendre and Legendre (1983) and by Gower and Legendre (1986). This tree was rooted as above.

The various methods briefly described above are discussed at length by Felsenstein (1982). Computations were carried out using version 2.2 of Felsenstein's PHYLIP computer package for inferring phylogenies.

In contrast to phylogenies, one expects biogeographic dispersal to proceed first to adjacent areas. Therefore, geographically adjacent regions should also be adjacent on the tree. This information may help to assess the adequacy of the classification results. Regions that were not geographically adjacent were sometimes found to be linked on the tree, especially in the case of peripheral regions with similarly poor faunas. This process, analogous to homoplasy, was possibly the result of the same selective forces producing convergence of faunal composition, by allowing only the most ubiquitous species to reach the climatically harshest regions. The geographic adjacencies could be used when building the trees, if the phylogenetic analysis programs were constrained to allow links to be formed only between geographically adjacent areas. In the absence of such programs, the study was broken down into parts and the parts were run separately; actually, all the trees presented in the Results have been run as three overlapping segments: western regions 31, 41, 42, 43, 52, 53, 56, 57, 58, 59, 60; eastern regions

21, 22, 23, 32, 33, 53, 54, 55; and northern regions 11, 12, 51, 52, 53, 54, 55, 59.

RESULTS

The results from five tree-reconstruction methods are presented as Figures 1 to 3. In each figure, the 21 regions are linked by the treelike structure issued from the common root, and lengths are indicated on the branches. They do not have to be proportional to the geographic distance between region gravity centers, because the dispersal of a fauna depends on factors other than geographic distance; in the case of fishes, these factors include the abundance of waterways and the harshness of climate.

In the Camin and Sokal method (Fig. 1A), segment lengths measure the number of 0-1 changes (loss of species) that occurred along segments, according to this parsimony method. On the tree, the most-parsimonious solution (shortest tree) poses a problem of geographical adjacency, in the area 53-54-55, since the 54 fauna would come from region 55 which is not adjacent. The next most-parsimonious solution, involving a link between 53 and 54, is only one step longer and seems more plausible. The two competing solutions are drawn as dashed lines.

In the rooted Wagner tree (Fig. 1B), the segment lengths are measured by the number of 0-1 changes (loss of species) plus the number of reversals (reacquisitions). This tree differs from the previous one in the topology linking the western regions. These differences will be discussed below.

For the polymorphism parsimony method (Fig. 2A), I have chosen to express the lengths of segments as the number of changes from presence to the polymorphic state hypothesized by the method, plus the number of fixations to presence or to absence, produced by selection of species during the travel of the fish community. The CONTML method (Fig. 2B) produces segment lengths that represent the expected amounts of variance, given the transformation into arcsine (in degrees) of the square root of the relative

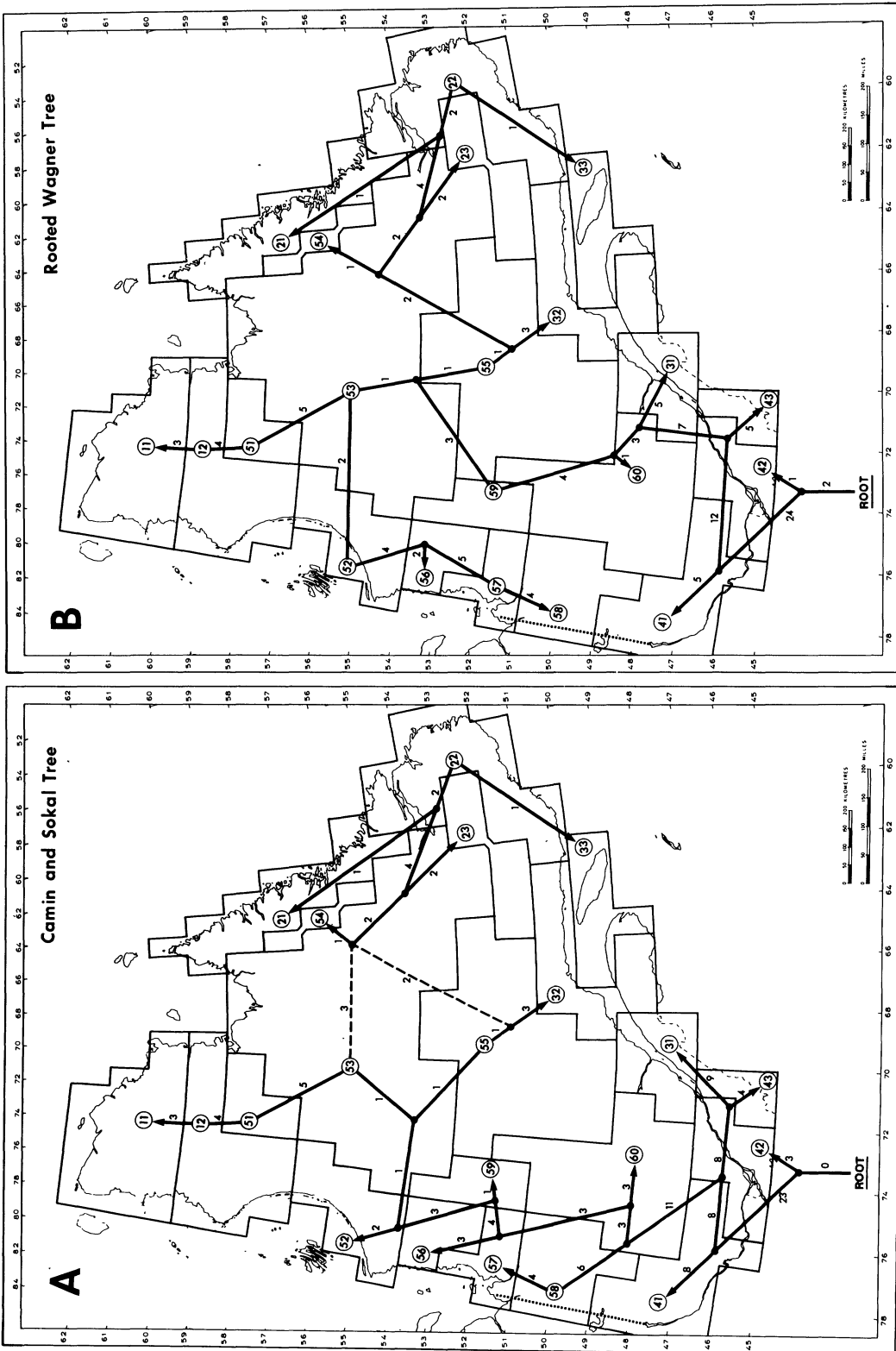


FIG. 1. (A) Camin and Sokal tree. Lengths are number of 0-1 changes (loss of species). (B) Rooted Wagner tree. Lengths are number of 0-1 changes and reversals.

frequencies, originally expressed on a 0-to-1 scale. For instance, a species with a frequency of 1 (100%) becomes a 90° angle following this transformation. If angles had been expressed in radians instead of degrees, all segment lengths would have been reduced by a constant proportionality factor. The CONTML approach also can be used on the original species percentage data; this produced only minor changes compared to the solution presented as Figure 2B.

Finally, for the method of Cavalli-Sforza and Edwards (Fig. 3A), the segment lengths are fitted distances. Since the original distances fed into the program were Odum's percentage differences, that are bound by a scale from 0 to 1, the fitted distances are also fractional numbers. They are multiplied by 100 in the figure.

DISCUSSION

For comparison purposes, the reconstruction of postglacial dispersal routes in the diffusion-flow analysis of Legendre and Legendre (1984), obtained by their "coefficient of dispersal direction," is presented in Figure 3B. Since it results from pairwise comparisons of all adjacent regions (but only the statistically significant arrows are drawn), it leads to an acyclic directed graph, or linkage diagram, which is not necessarily a tree because two arrows can point to the same region. In Figure 3B, the "length" L of an arrow from region i to region j , which represents an amount of faunal divergence ("distance"), is

$$L(i, j) = \left[\frac{|DD_{\max}| - |DD(i, j)|}{|DD_{\max}|} \right] \times 100,$$

where $DD(i, j)$ is the "coefficient of dispersal direction" between regions i and j in table 4 of Legendre and Legendre (1984), and DD_{\max} is the maximum value of DD in the same table. The longest of these lengths (e.g., 50 or more) are of little interest.

This is not to say that I evaluated the results from the tree-reconstruction method only by their concordance with Figure

3B. The methods for phylogenetic-tree reconstruction use a different part of the information than the pairwise comparison of adjacent regions. Indeed, the "coefficient of dispersal direction" of Legendre and Legendre (1984) looks for small-scale gradients between adjacent regions only, with no account taken of the large-scale flow pattern. It is to be applied as a first approximation, when the structure connecting regions is still unknown. The methods of phylogenetic-tree reconstruction can be used legitimately if these first results indicate that: (1) there is a single major root for the dispersal phenomenon under study; and (2) the pattern is mostly dendritic instead of reticulate. Furthermore, Figure 3B is a fairly crude first approximation of dispersal history. Indeed, since my purpose is to try to reconstruct history from present-day faunas, a dispersal arrow is drawn only if there is strong indication for it. The indication is considered strong enough (1) if the common part of the fauna is large enough, and (2) if the gradient is strong enough to pass the statistical test of asymmetry described below.

In contrast to pairwise comparisons, the methods of phylogenetic-tree reconstruction (Figs. 1A–3A) try to find the dendritic structure that *minimizes faunal changes* throughout the tree. Of course, these trees represent a simplification of the relationships that exist in the data set, because the constraint is imposed that only one incoming arrow can point to any given region, but this arrow is likely to correspond to the most important dispersal pathway (under the assumptions of each particular method). This is the first type of information that tree-reconstruction methods can provide.

In some instances, the "leaves" of the tree (the regions) are located very close to nodes (junction points), or even right on top of them, while in other cases the distance between a region and the closest node is important. This information is the second type that only tree-reconstruction methods can supply. It can be interpreted in the following two ways: (1) A node

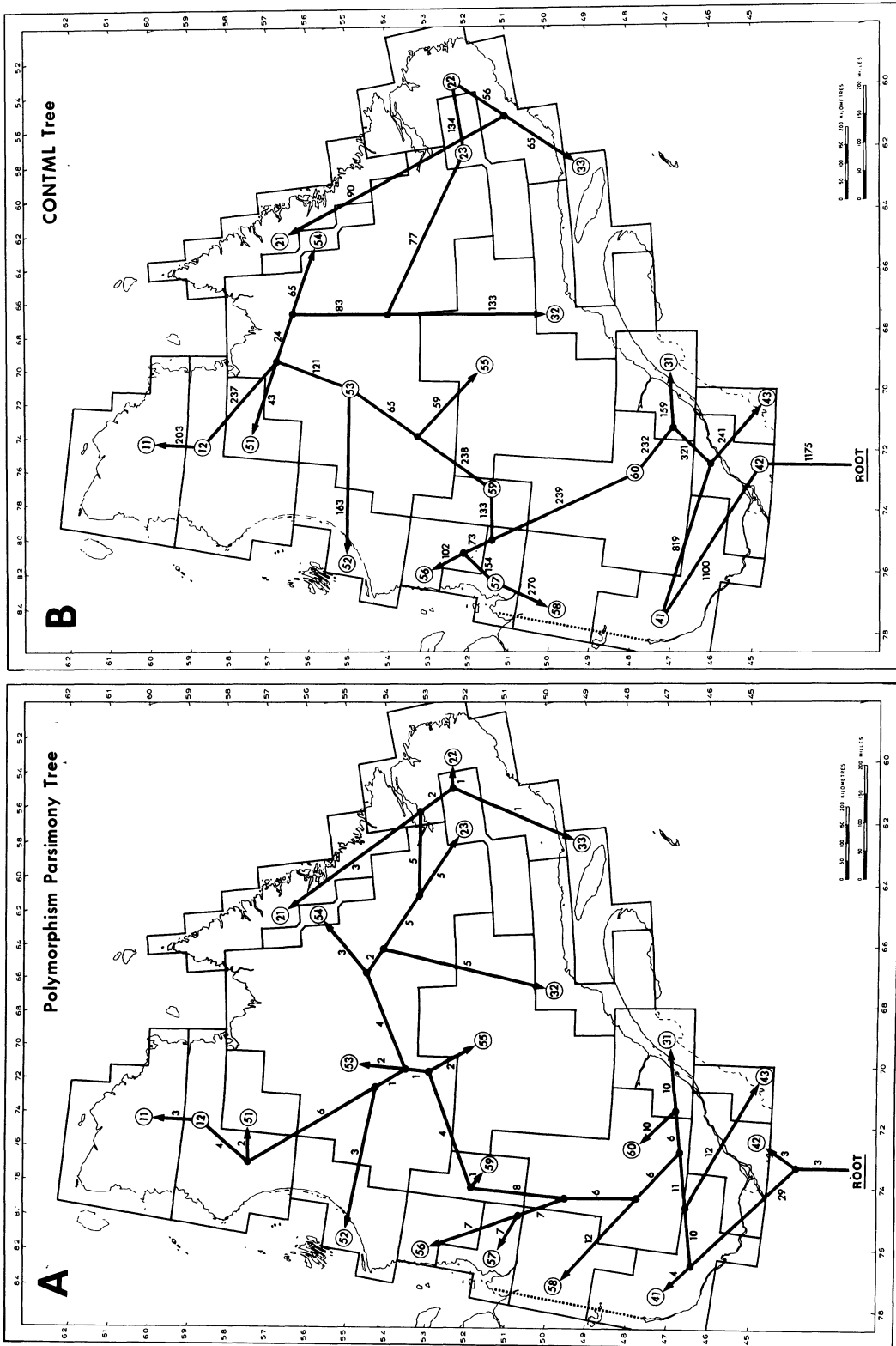


FIG. 2. (A) Polymorphism parsimony tree. Lengths are number of changes from presence to the polymorphic state, plus number of fixations through selection. (B) Continuous-character data maximum-likelihood tree (CONTML), after an arcsine transformation of the square root of the data. Lengths are expected amounts of variance.

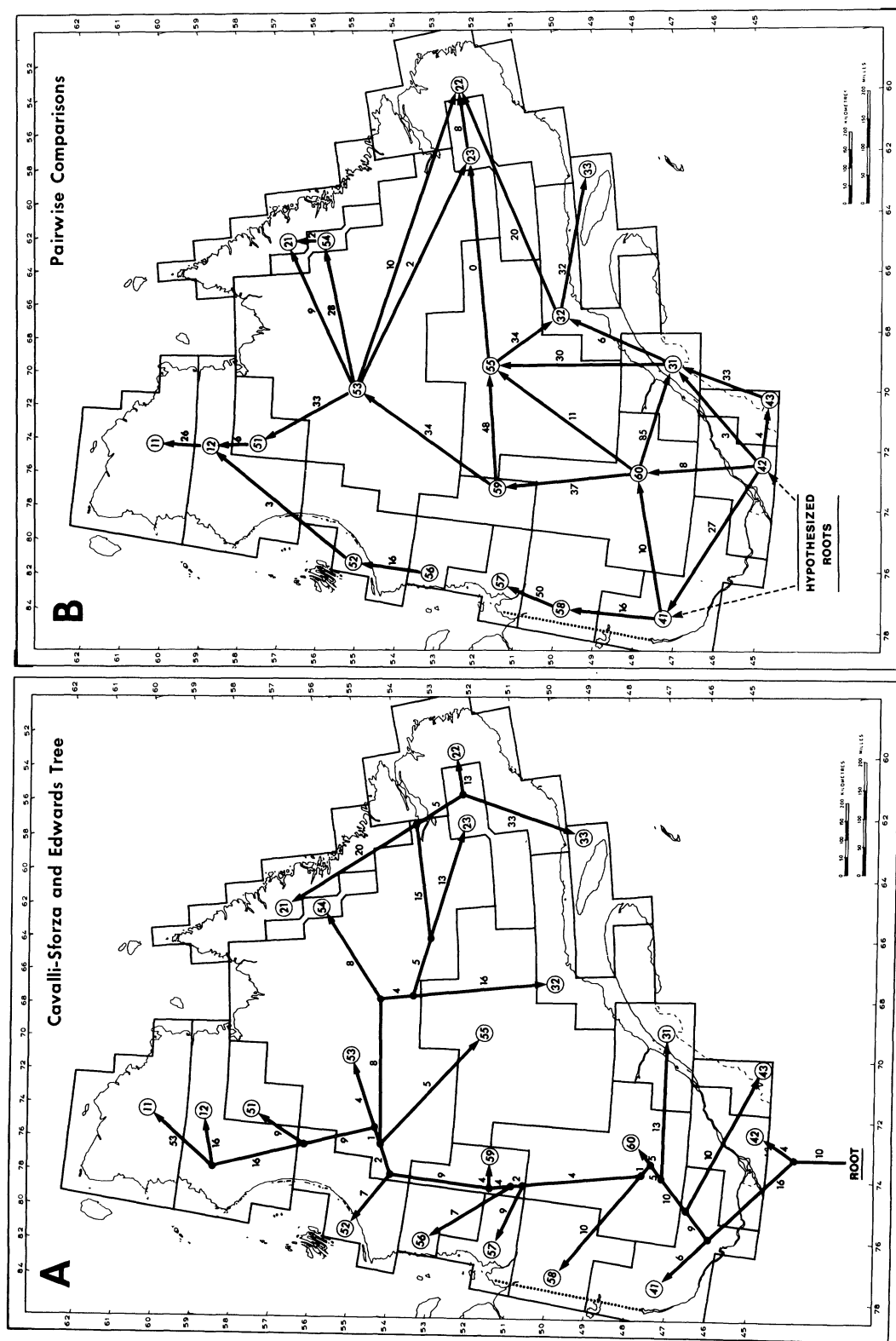


FIG. 3. (A) Least-squares Cavalli-Sforza and Edwards tree from a distance matrix of Odum's percentage differences. Lengths are fitted distances, multiplied by 100. (B) Pairwise comparison of adjacent regions. Modified from figure 5 of Legendre and Legendre (1984), which is represented here as an acyclic directed linkage diagram. Arrows indicate pairwise comparisons of adjacent regions that were found to have a statistically significant asymmetry. Lengths are transformations of the "coefficients of dispersal direction" (see text).

hanging far from the "leaves" may indicate a divergence of the regions after a long interruption of contact between geographically adjacent faunas. In other words, the node would represent a fauna of the past, from which the present-day "leaves" are derived. This may well be the case with the faunas of regions 56 to 60 in Figures 1A, 2A, 2B and 3A. (2) Alternatively, it may indicate that the treelike model breaks down, because of multiple and equally important spatial influences on the fauna of a given region. This is likely to be the case with region 23, for instance, since the multiple inputs shown in Figure 3B cause nodes to hang midway between 23, 53, 54 and 55 in Figures 1A to 3A. This is also the case with the multiple inputs to region 60. The pairwise comparison of adjacent regions can help decide between these two alternatives since it shows what regions have received the influence of multiple inputs.

The features common to the output of almost all tree-reconstruction methods (Figs. 1A-3A) and to the pairwise-comparison method as well (Fig. 3B) are:

(1) The root of the dispersal pattern is located in region 42, or between regions 42 and 41. This fits previous knowledge of glacial refugia for stenohaline freshwater fishes of the Québec peninsula. When the Laurentide Ice Sheet retreated from the peninsula approximately 14,000 to 5,000 years ago, these species reinvaded the territory from United States refugia through the Great Lakes, the province of Ontario, and New England states, using watershed interconnections, as described by Legendre and Legendre (1984:1792).

(2) South of the St. Lawrence River, the faunas of regions 41 and 42 gave rise to those of regions 43 and 31. In all the trees, the node leading to regions 43 and 31, and then to all other regions, is located at a large distance away from region 42, because the St. Lawrence River fauna in region 42 (Montréal area) is the richest of the peninsula and contains several species found nowhere else in the territory under study.

(3) With all methods but the rooted

Wagner tree, there is a clear dispersal movement northward, from regions 41 and 42 in the south to region 11 at the northern edge of the peninsula. For the three northernmost regions (51, 12 and 11), and perhaps for region 52 as well, this northward faunal movement seems to have gone through the large central region 53.

(4) From that northward stem, the fish community branched off eastward to reach the center of the peninsula (regions 53 to 55).

(5) Finally, the fish reached the east coast regions (21 to 23) from the center of the peninsula, and came down into the rivers of the St. Lawrence north shore (region 32). Mechanisms allowing the species to cross the center of the peninsula are described in Legendre and Legendre (1984).

The main point of disagreement between all tree-reconstruction methods on one hand, and the pairwise-comparison method (Fig. 3B) on the other, is the origin of the stenohaline fish fauna of region 33. All tree-reconstruction methods suggest that it derives from region 22, while the pairwise-comparison method traces it from region 32. This last solution is more plausible, since contact between opposite shores of the Gulf of St. Lawrence is largely prevented for stenohaline freshwater species by the sea water. The freshwater ichthyofauna of region 33 (Gulf of St. Lawrence removed) is composed mostly of euryhaline species (15 species), compared to four stenohaline species. Only one (*Salvelinus fontinalis*) is found everywhere in the region, the other three being found only in a few spots of one or two of the seven 1°-square pixels of this region. To the tree-reconstruction methods, this places region 33 closer to 22 than to 32, because 32 has a much richer fauna than either 22 or 33, and because region 33 is more similar to region 22 for these four species. This is a case of convergence of two impoverished regions, 22 and 33. The pairwise-comparison method did not suffer the same pitfall because the asymmetry in species composition was not large enough between regions 22 and 33 to pass

McNemar's test (Sokal and Rohlf, 1981), which was a condition for an arrow to be drawn. McNemar's test is a statistical test of asymmetry for a 2×2 contingency table, which is used here to compare pairs of region faunal vectors for the presence or absence of species. Between regions 32 and 33, the asymmetry was strong enough to pass the test, and an arrow was drawn despite the relatively large distance between faunal composition vectors. This case suggests that tree-reconstruction programs could be constrained to connecting only region pairs whose faunal vectors could pass McNemar's test of asymmetry. In the absence of constrained programs, one can remove from the final tree connections that do not pass the test, even though it means leaving regions unlinked to the tree because of lack of evidence.

Among the tree-reconstruction methods that were acceptable (see Methods), the rooted Wagner tree is the only one that produced a topology markedly different from all others. The Wagner tree is much shorter (total length: 122 steps) than the Camin and Sokal tree (total length: 140 steps) in particular, because the Wagner method imposes fewer constraints. Using reversals as well as repeated 0-to-1 mutations to account for incompatibilities in the data set, the Wagner algorithm created a western branch with a fish community that progressed southward from region 53 to 58, reacquiring species that had previously been lost during its march northward. Although the presence of some reversals is not in contradiction with the dispersal process, it is very unlikely that the community could have developed a whole branch of the tree, drawing mainly on reversals; between regions 53 and 52, there is one loss of a species and one reversal in the rooted Wagner tree; four reversals between 52 and the node; two reversals between 56 and the node; two losses and three reversals between 57 and the node; four reversals between regions 57 and 58. Besides the unlikelihood of such a series of species reacquisitions, the geomorphological history of the area also makes this pattern very unlikely, since it

has been especially easy for a large part of the fish fauna of region 41 and of north-eastern Ontario to reach regions 58 and 60, then 57, 59 and 56, using glacial Lake Ojibway-Barlow as its main waterway. This lake formed along the margin of the ice sheet as it receded towards Hudson Bay (Anonymous, 1970).

All four other tree-reconstruction methods lead to comparable trees. In the Camin and Sokal tree, distances are particularly easy to interpret in biogeographic terms because each step making up internode distances corresponds to the loss of a species (below or equal to the 50% mark in these data). Although this is felt to be a simplifying assumption because it excludes all the other processes described in the Methods, this tree can be taken as a template on which to list the species lost during each internode. This results in Figure 4, which represents a crude approximation to what may have happened during the 14,000 years the freshwater fish species took to reinvade this territory since the last glaciation. Lost species are represented by numbers, which correspond to the list in Legendre and Legendre (1984: table 2, which also includes the list of species remaining in each region, with their frequencies of occurrence). The arrow linking region 33 to region 22 has been removed since it does not pass McNemar's test of asymmetry comparing the stenohaline fish faunas of these two regions. Fifteen stenohaline species are not present with sufficient frequency (50%) in any one of the regions; they are listed as disappearing along the root branch.

CONCLUSIONS

Tree-reconstruction methods can be used for biogeographic reconstruction of the dispersal of a community if there is some evidence that: (1) there is a single major root for the dispersal phenomenon; and (2) the pattern is mostly dendritic. Such evidence may be obtained by the method of pairwise comparison of adjacent regions (Fig. 3B), for instance.

Four tree-reconstruction methods were found to be based upon acceptable as-

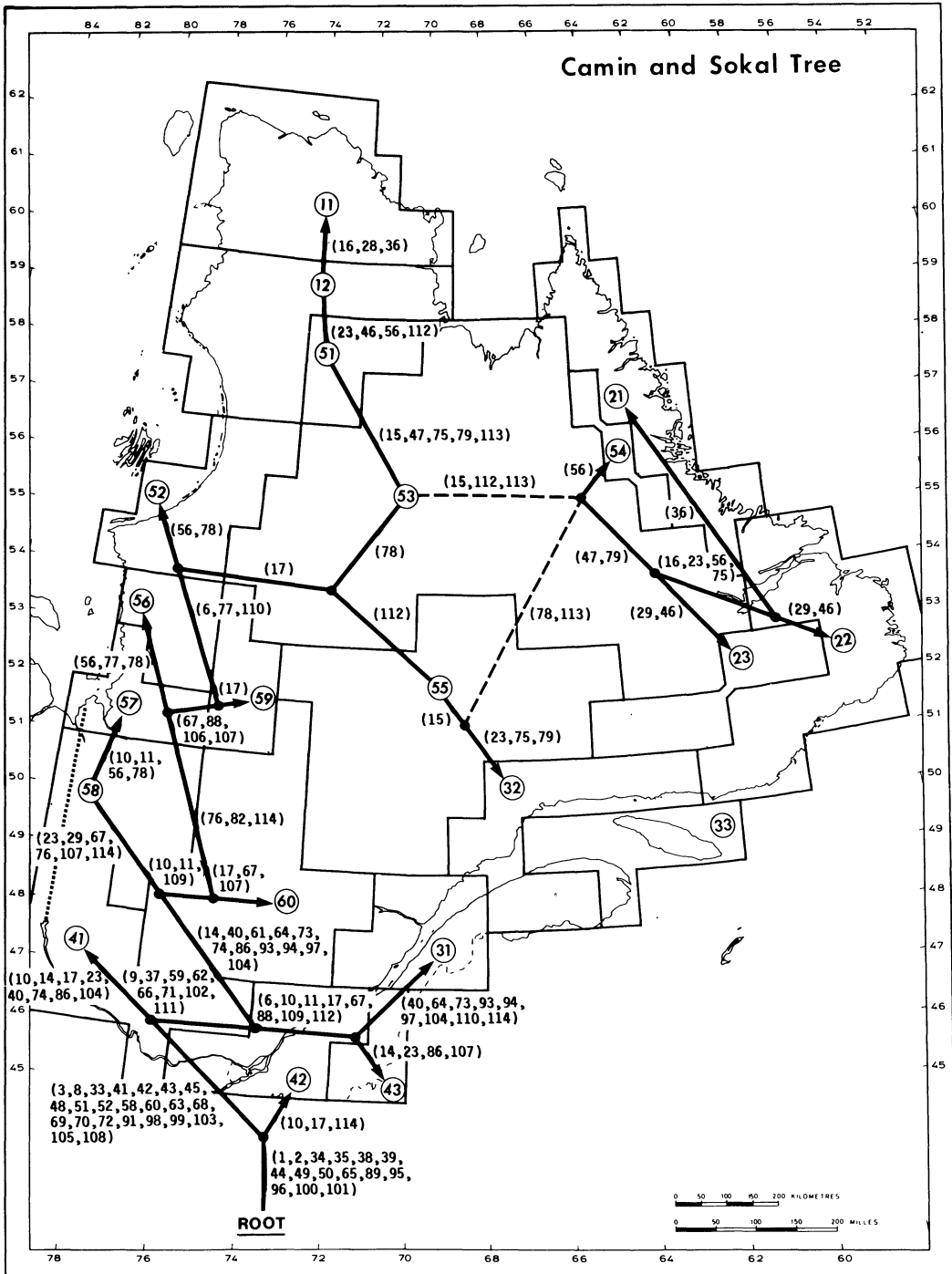


FIG. 4. Camin and Sokal tree, showing in parentheses the species lost along each branch. Species numbers as in table 2 of Legendre and Legendre (1984). Region 33 is not connected for lack of evidence.

sumptions for this type of study, and to lead to satisfactory trees: for binary data, the method of Camin and Sokal and the polymorphism parsimony method; for frequency data, the CONTML method, with or without arcsine-square-root transformation, and the least-squares distance method of Cavalli-Sforza and Edwards after computing an Odum's distance matrix among regions. The results produced by these four methods are very much alike; I have not discovered criteria to tell them apart and find one to be superior to all others. As with many other methods of multivariate data analysis, the user may find it useful to use several of these methods on the same data set in order to discover the common part of the outcoming structure. The Camin and Sokal results are easily interpreted in biogeographic terms because internode distances measure the number of lost species.

Geographic adjacency of regions found to be adjacent along the tree is an assumption of biogeographic dispersal that can be used as a criterion to point out unrealistic solutions. All methods of tree reconstruction tended, with data used here, to cause the convergence of nonadjacent peripheral regions with similarly impoverished faunas. Programs constrained to link only adjacent regions would be very welcome. In the absence of such programs, it was felt necessary to run the data set as three separate but overlapping segments, from which the final tree was assembled.

The tree-reconstruction methods provide the researcher with two types of information that did not come out of a simple pairwise analysis of adjacent regions. (1) The tree is a skeleton representing the most important pathways of dispersal in a territory; as such, it is a simplification of the phenomenon because it leaves aside reticulate interconnections among regions and focuses only on the most important dispersal flows. (2) Furthermore, when nodes of the tree are located far away from terminal regions (or "leaves"), such nodes may represent a fauna of the past, from which present-day faunas have diverged. Alternatively, these nodes may indicate

areas where the reticulate mode of dispersal became more important than the dendritic mode. A pairwise analysis of adjacent regions, for that area, can help decide between these interpretations.

The tree-reconstruction algorithms are built to include in the tree all regions fed into the computer run, even when data become very sparse in highly impoverished regions as found in the present study. McNemar's test of asymmetry was found to be a good criterion; region pairs should not be linked if their faunal vectors do not show enough asymmetry to pass this test. Indeed, and contrary to phylogenetics, one is not interested in biogeography to derive a fauna from another region's fauna if the two faunal vectors do not show enough asymmetry. In the absence of computer programs capable of taking this constraint into account, I recommend removing such weak links from the final tree, even if it means leaving regions unlinked to the tree.

ACKNOWLEDGMENTS

I am grateful to J. Felsenstein, University of Washington in Seattle, who supplied the PHYLIP package of tree-reconstruction programs used in this study. He, as well as D. Currie, University of Ottawa, kindly reviewed the manuscript. Computing time was provided by the Centre de Calcul, Université de Montréal. This research was supported by grants from the National Sciences and Engineering Research Council of Canada and from the Ministère de l'Éducation du Québec.

REFERENCES

- ALLEN, T. F. H., AND T. B. STARR. 1982. *Hierarchy—Perspectives for ecological complexity*. Univ. Chicago Press, Chicago.
- ANONYMOUS. 1970. Carte glaciaire du Canada. Commission géologique du Canada, No. 1253A, Ministère de l'Énergie, des Mines et des Ressources, Ottawa, Ontario.
- CAMIN, J. H., AND R. R. SOKAL. 1965. A method for deducing branching sequences in phylogeny. *Evolution*, 19:311–326.
- CAVALLI-SFORZA, L. L., AND A. W. F. EDWARDS. 1967. Phylogenetic analysis: Models and estimation procedures. *Evolution*, 32:550–570.
- CLIFF, A. D., AND J. K. ORD. 1981. *Spatial processes: Models and applications*. Pion Limited, London.
- CROVELLO, T. J. 1981. Quantitative biogeography: An overview. *Taxon*, 30:563–575.
- EDWARDS, A. W. F., AND L. L. CAVALLI-SFORZA. 1964.

- Reconstruction of evolutionary trees. Pages 67-76 in *Phenetic and phylogenetic classification* (V. H. Heywood and J. McNeill, eds.). Systematic Association Publication No. 6, London.
- FARRIS, J. S. 1978. Inferring phylogenetic trees from chromosome inversion data. *Syst. Zool.*, 27:275-284.
- FELSENSTEIN, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.*, 25:471-492.
- FELSENSTEIN, J. 1979. Alternative methods of phylogenetic inference and their interrelationship. *Syst. Zool.*, 28:49-62.
- FELSENSTEIN, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: Finding maximum likelihood estimates. *Evolution*, 35:1229-1242.
- FELSENSTEIN, J. 1982. Numerical methods for inferring evolutionary trees. *Q. Rev. Biol.*, 57:379-404.
- GOWER, J. C., AND P. LEGENDRE. 1986. Metric and Euclidean properties of dissimilarity coefficients. *J. Classification*, 3:5-48.
- KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.*, 18:1-32.
- LEGENDRE, L., AND P. LEGENDRE. 1983. Numerical ecology. *Developments in Environmental Modelling*, 3. Elsevier Scientific Publishing Co., Amsterdam.
- LEGENDRE, P. 1984. Report on Seventeenth International Numerical Taxonomy Conference. *Syst. Zool.*, 33:117-121.
- LEGENDRE, P., AND V. LEGENDRE. 1984. Postglacial dispersal of freshwater fishes in the Québec peninsula. *Can. J. Fish. Aquat. Sci.*, 41:1781-1802.
- LEPS, J., J. OSBORNOVA-KOSINOVA, AND M. REJMANEK. 1982. Community stability, complexity and species life history strategies. *Vegetatio*, 50:53-63.
- MOTYKA, J. 1947. O zadaniach i metodach badan geobotanicznych. Sur les buts et les méthodes des recherches géobotaniques. *Ann. Univ. Mariae Curie-Sklodowska Sect. C, Suppl.* 1.
- ODUM, E. P. 1950. Bird populations of the Highlands (North Carolina) plateau in relation to plant succession and avian invasion. *Ecology*, 31:587-605.
- PIMM, S. L. 1984. The complexity and stability of ecosystems. *Nature*, 307:321-326.
- SOKAL, R. R., AND F. J. ROHLF. 1981. *Biometry. The principles and practice of statistics in biological research*. Second edition. W. H. Freeman and Co., San Francisco.
- WARTENBERG, D. 1985. Multivariate spatial correlation: An exploratory tool for geographical analysis. *Geograph. Anal.*, 17:263-283.

Received 6 June 1985; accepted 19 November 1985.