A PHYLOGENETIC ANALYSIS OF THE CAMINALCULES.
I. THE DATA BASE

ROBERT R. SOKAL
Department of Ecology and Evolution, State University of New York at Stony Brook,
Stony Brook, New York 11794

Abstract.—The Caminalcules are a group of “organisms” generated artificially according to
principles believed to resemble those operating in real organisms. A reanalysis of an earlier
data matrix of the Caminalcules revealed some inconsistencies and errors which necessitated
recoding of some characters. The resulting differences with earlier results are minor. The images
of all 77 Caminalcules are featured, those of the 48 fossil species for the first time. The characters
of the Caminalcules are defined and a data matrix is furnished for all Recent and fossil species.

A new phenetic standard is proposed for the Caminalcules which divides them into five
“genera.” The true cladogram is revealed for the first time. Recent Caminalcules have evolved
over 19 time periods. Five branches correspond to the phenetic genera but originate at greatly
differing time periods. Four lines terminate in fossils.

A series of measures for quantifying evolutionary change is defined, including measures for
homoplasy, parallelism, and reversal. A survey is made of these measures and of other statistics
of relevance to systematics for 19 data sets from the numerical taxonomic literature. The Cami-
nalcules turn out to be compatible to data sets on real organisms with respect to all these
measures, as well as with respect to evolutionary rates and species longevities. Thus, questions
raised by an analysis of the Caminalcules should be of interest to systematists concerned with
the analysis of data sets on real organisms. [Phenetic classifications; cladistic classifications;
estimated cladograms; homoplasy; Wagner trees; Caminalcules; numerical taxonomy.]

This paper, and others to follow, takes ad-
vantage of the opportunity afforded by a
group of artificial organisms with a known
phylogeny, the Caminalcules, to throw light
on some of the questions concerning prin-
ciples and procedures that currently engage
the attention of taxonomists. There is consid-
erable disagreement on the relative merits of
phenetic and cladistic classifications (Sneath
and Sokal, 1973; Eldredge and Cracraft, 1980;
Wiley, 1981). Some workers contend that
classifications based on phylogenetic prin-
ciples are empirically better by various criteria
of optimality (Farris, 1977, 1979a, b; Micke-
vich, 1978a, 1980; Schuh and Polhemus, 1980;
Schuh and Farris, 1981). These claims have
been questioned by others who find the evi-
dence and methodology presented to be flawed (Colless, 1980; Rohlf and Sokal, 1980,
1981; Sokal and Rohlf, 1981a; Rohlf et al.,
1983a, b). The empirical studies, and the argu-
ments pro and con phylogenetic classifica-
tions derived in such investigations, suffer
from a major impediment. All of the phy-
logenetic classifications reported in the lit-

erature are only estimates of the true phy-
logeny, which is unknown for all real
organisms. By contrast the various results of
this study merit serious attention because
they can be examined against the bench-
mark of the true phylogeny of the group.

The Caminalcules are artifacts created by
the late Professor Joseph H. Camin of the
University of Kansas and in effect represent
a single simulation of the evolutionary pro-
cess by rules that have not been made ex-

plicit. However, readers will find that these
organisms, which have the advantage over
other simulations in presenting a visual rec-
dord to the investigator, illustrate a variety of
evolutionary phenomena and are therefore
of considerable pedagogical and heuristic
value. The relevance of this data set to cur-
rently active issues in systematics will read-
ily become evident to the reader of this series.
I shall show that with respect to a substantial
array of measurable properties, the Cami-

alcules are well within the range of empiri-
cally observed values for real taxonomic
groups and that, conversely, for no property
of consequence in numerical taxonomy are
the Caminalcules beyond the range of ob-
served values in real organisms.

At the suggestion of the Editor and some
reviewers, this series of publications is initiated in this paper with the presentation of the data on which previous and succeeding studies have been based. I furnish the images of the previously published 29 Recent species and for the first time the 48 "fossil" species. I also present for the first time the true phylogeny of the Caminalcules as generated by Professor Camin. With these illustrations I provide a list of the descriptions of characters as adopted in my laboratory as well as a data matrix giving the character states for all 106 characters for each of the 77 Recent and fossil Caminalcules. In addition to presenting a new standard phenetic classification, the paper describes a number of measures for taxonomic and evolutionary properties and compares the Caminalcules with data sets on real organisms with respect to these measures. Subsequent studies in this series will treat estimates of the true cladogram, the inclusion of fossils in phenetic and cladistic classifications, congruence and character stability, and OTU stability.

ORIGIN OF THE DATA BASE

The original intention in generating the Caminalcules was to study the nature of taxonomic judgment (eventually published by Sokal and Rohlf, 1980), but work with these animals has led to other developments in numerical taxonomic methodology, such as an early method of numerical cladistics (Camin and Sokal, 1965) and a method for obtaining taxonomic structure by random and systematic scanning of biological images (Sokal and Rohlf, 1966; Rohlf and Sokal, 1967). Other experiments on taxonomic judgment have also been based on the Caminalcules (Moss, 1971; Sokal, 1974; Moss and Hansell, 1980).

Camin drew the Caminalcules using master stencils for ditto machines. The genetic continuity of the Caminalcules was achieved by Camin by tracing successive drawings of the animals, permitting the preservation of all characters except for such modifications as were desired. Xerox copies of the images of the Recent OTUs were made available in the early 1960s, those of the fossil OTUs some years later. Independent xerox copies of all OTUs are in the possession of Dr. Paul A. Ehrlich of Stanford University and Dr. W. Wayne Moss of the Philadelphia Academy of Sciences in addition to myself. The originals drawn on the ditto masters appear to have been lost following the death of Professor Camin in 1979.

All examinations of the Caminalcules for numerical taxonomic studies have been carried out on the xeroxes of the images. Illustrations of all 29 Recent OTUs have been published three times previously (Sokal, 1966; Rohlf and Sokal, 1967; Sokal and Rohlf, 1980). For this purpose, inked copies of the xeroxed images were photographed. Although the artist's copies of the original xeroxes are quite faithful, inevitably some fine detail has been altered or lost. Thus, not every character state described below can be unequivocally recognized in the featured illustrations. The version of the 29 Recent OTUs published in Sokal (1966) was "beautified" by the publisher's artist and cannot be relied upon for detail. The images of the 48 fossils were newly inked for this study and all differentiating characteristics can be observed in them.

The Recent OTUs, numbered 1 to 29, are shown in Figure 1. The fossil OTUs, given different code names by Camin, were randomly assigned numbers 30 to 77 by me. They are shown in Figure 2.

The true cladogram of the group was communicated to me by Camin in 1970. But although this information was employed in the computations leading to the analysis of taxonomic judgment (Sokal and Rohlf, 1980), access to it was restricted even for workers on this project. I did not become intimately familiar with the phylogenetic tree until 1981 during the final analyses leading to this manuscript.

Readers of this paper and of subsequent ones in this series should note that I use the term cladogram in the sense in which I originally coined it (Camin and Sokal, 1965; also independently coined with the same meaning by Mayr, 1965). One definition of this meaning (Sneath and Sokal, 1973:29) is as "A branching . . . network of ancestor-descendant relationships." This definition differs from the several meanings attached to the term cladogram by various cladists (e.g., El-
Fig. 1. Images of the 29 Recent species of the Caminalcules, an imaginary group of animals, generated by J. H. Camin. The phenetic classification adopted in this paper divides these into the following "genera": Genus A comprises species 7, 8, 13, 14, 15, 25, 28; B comprises 6, 9, 10, 11, 21; C comprises 1, 16, 17, 24, 27; DE comprises 2, 3, 4, 5, 12, 18, 22, 23; and F comprises 19, 20, 26, 29. Earlier studies (e.g., Sokal and Rohlf, 1980) placed species 3 and 4 in a genus D separate from E, which comprised the balance of the species in the current genus DE.
Fig. 2. Images of the 48 fossil species of the Caminalcules, an imaginary group of animals, generated by J. H. Camin. The species were assigned numbers 30 through 77 at random. The common ancestor of the entire group is species 73. Their phylogenetic relations can be learned from Figure 4.
dredge and Cracraft, 1980; Nelson and Platnick, 1981; Wiley, 1981; see also Sneath, 1982). Thus, cladogram as used in this and succeeding papers refers to a branching sequence depicting the actual or hypothesized genealogy of the OTUs. It does not include length of branches of the tree and is not a statement about the evolution or pattern of character states akin to “nested synapomorphy schemes.”

Camin did not keep a written record of character changes. In the 1960s, A. J. Boyce and I. Huber prepared a data matrix for a numerical phenetic study of the 29 Recent OTUs, describing 86 characters. This data matrix was the basis of various phenetic analyses of the Caminalcules (Sokal, 1966; Sokal and Rohlf, 1966, 1980; Rohlf and Sokal, 1967). For the present study it became necessary to map the coded characters onto the now known cladogram and during this process each character was re-evaluated for the 29 Recent OTUs. This re-examination of the images led to the discovery of several inconsistencies in the Boyce and Huber data matrix. Altogether 13 of the 2,494 entries in that 29 × 86 matrix seemed to be in error and were corrected. Original character 58 was deleted because it was discovered to be invariant. All measurement characters were re-measured and small differences with earlier measurements were recorded. Because the images available during my current studies are poorer quality xerox copies than those available to Boyce and Huber, the resolution of some of the more minute morphological features was more difficult and in a few cases characters were recoded into fewer, coarser classes.

Remeasurements, changes in logic, and revision of character state coding accounted for changes in 33 of the 85 characters. To account for the features of the fossil OTUs, which had not been coded previously, another 12 characters were redefined and 21 new characters were added. This augmented data matrix comprises 77 OTUs and 106 characters of which 65 are binary, 31 are ordered integer characters, and 10 are measurement characters. From this matrix an 85 character × 29 OTU subset was extracted for the Recent OTUs. This matrix has 48 binary characters, 27 ordered integer characters and 10 measurement characters. Both the 106 character × 77 OTU matrix and the 85 character × 29 OTU matrix contain characters with NCs (no comparison codes). All NCs in this study are logical (i.e., presence of a given state for character h making it impossible to define a state for character i) rather than due to missing information. In the larger matrix, 79 of the 106 characters contain NC codes; in the smaller, 59 of the 85 characters have NCs. The measurement characters are available in two versions. One is as direct measurements in millimeters, the other is coded as ordered integer states but with gaps omitted. The range of measurements for one character was divided into 10 equally wide classes coded 0 to 9. Each character state was then assigned to one of these classes, but classes lacking observed character states were ultimately omitted. Thus, if there were no OTUs with measurements in class 3, the subsequent class 4 was renumbered 3. The number of states in the 10 characters resulting from this procedure ranges from four to nine. A list defining the states of each character is furnished in the Appendix. The actual data matrix (integer coded version) is shown in Table 1.

The consequences of recoding the measurement characters in integer scale were minimal. Matrix correlations for the 29-OTU study between the resemblance matrices based on these two methods of character coding are very high (0.998 and 0.997, respectively, for \( r \) and \( d \), the correlation and taxonomic distance coefficients). The classifications resulting from these two coding methods are identical. In this and subsequent papers, only the data matrix using integer scale coding of measurement character states is featured, since this is the simpler coding preferred for the several cladistic methods employed.

The effects of correcting and recoding the original Boyce and Huber data matrix for the 29 OTUs were minor. Matrix correlations between the two versions of the resemblance matrices are high (0.980 for \( r \), 0.969 for \( d \)). UPGMA phenograms based on these resemblance matrices are close; the strict consensus index \( C_{CI} \) for the classifications based on the
two correlation matrices is 0.889, for those based on the two distance matrices 0.815. This index (Rohlf, 1982) ranges between 0 for no consensus and 1 for perfect consensus. It is identical to the consensus fork index of Colless (1980) and the proportional consensus index of Sokal and Rohlf (1981a).

The differences between the classifications based on correlations and distances computed from the original Boyce and Huber data matrix (see fig. 3 of Rohlf and Sokal, 1967) and those based on the updated, integer-coded version all involve taxonomic affinities at the "species" level. The "generic" classification is the same for both versions. When the new classifications were compared to 49 subjective classifications of the Caminalcules produced by 22 experimental subjects (Sokal and Rohlf, 1980), the new phenograms of the Caminalcules were found to be no more (and no less) similar to these intuitive classifications than the earlier Boyce and Huber classifications.

A NUMERICAL PHENETIC CLASSIFICATION

For phenetic analyses the data were subjected to standard taxometric procedures using the NTSYS program of numerical taxonomic computer programs (Rohlf et al., 1980). Characters were standardized before computation of correlation and taxonomic distance coefficients between OTUs.

Rohlf and Sokal (1967) noted that phenograms based on distances and correlations of the Caminalcules differ substantially in the specific as well as generic affinities indicated. Previous work by Rohlf and Sokal (1967) and by Sokal and Rohlf (1980) showed that the classification based on the correlation matrix corresponded more closely than that based on the distance matrix to the taxonomic structure noted by individual taxonomists grouping the Caminalcules by conventional, intuitive methods. This undoubtedly occurs because correlation coefficients are more sensitive to shape than are distance coefficients (Rohlf and Sokal, 1965), and it is shape rather than size on which taxonomists tend to base their judgments on taxonomic affinity. For this reason, the UPGMA clustering (shown in Fig. 3) of the correlation matrix based on the updated, integer-coded data matrix was the logical choice as a new standard for a phenetic classification of the 29 Recent OTUs to be compared with the true cladogram. This phenogram also has the highest cophenetic correlation coefficient of any phenetic classification of the Caminalcules computed so far (0.965).

Five major clusters, the "genera" of the Caminalcules, can be seen in the phenogram in Figure 3. These fall naturally into two major groups, that containing genera A, B, and F, and a second group containing C and DE. The status of D, consisting of OTUs 3 and 4, is problematical. Only in one previous study, based on a mechanical method of scanning the images (Rohlf and Sokal, 1967: fig. 3A), was D unequivocally separated from E at a level of phenetic similarity equal to that of the other genera—A, B, C, and F. The Boyce and Huber correlation phenogram (Rohlf and Sokal, 1967: fig. 3C) shows D joining E at a level closer than that of at least one member of A joining its genus. Intuitive classifications by various taxonomists (Sokal and Rohlf, 1980) had frequently assigned species 3 and 4 to genus E, although some had singled out the divergence of these two OTUs which are unique in having rudimentary posterior appendages as well as an elongated body obliterating the neck. Since the phenogram in Figure 3 clearly affiliates D with E, it seems appropriate to join the two together into a single genus which I have called DE to preserve continuity with earlier publications.

THE TRUE CLADOGRAM

The true cladogram as furnished by J. H. Camin is shown in Figure 4. The diversity of the taxon was generated over 19 time periods. Lines undergoing evolutionary change are indicated in Figure 4 by slanted (as distinct from vertical) lines. Species that underwent evolutionary change during a given time period are shown as solid circles at the end of that period. Ancestral species that are not continued into the next time period as vertical lines are considered to have become extinct and are indicated by small hollow circles. Five branches leading to Recent forms, corresponding to the five phenetic genera, can be recognized but these originate at
The first 15 characters are from the character code or the Recent OTUs. The last two states have special symbols: X no comparison (NC).
greatly differing time periods. There are also four lineages that underwent evolutionary change before becoming extinct. The extinct terminal species of these fossil lines are shown as large hollow circles. I have indicated the amount of evolutionary change (path length of the internode) based on 85 characters by the length of the thickened bars along the slanted lines. Note that path lengths for lines leading to fossils are based on only 85 characters to make them comparable to path lengths of Recent forms. Path lengths based on all 106 characters will be illustrated in paper III of this series.

MEASURES OF PHENETIC AND EVOLUTIONARY CHANGE

To describe evolutionary changes in this and succeeding papers, various statistics based on the characters and the true cladogram are needed. They are summarized in Table 2. Some terms need to be defined. An entire taxon consists of t OTUs and is described by n characters. OTUs are labelled as 1, ..., j, k, ..., t and characters as 1, ..., h, i, ..., n. The most recent common ancestor of all the OTUs is a. \( \sum \) is summation over all OTUs indexed by j, usually t OTUs.
Let us assume that we know the true cladogram of the taxon, as is the case in the Caminalcules. The nodes of the tree will be the Recent OTUs and their ancestors, and the internodes will be the lines of descent connecting these taxa. Throughout I shall assume that evolutionary (character state) change occurs only along internodes. Let \( I_j \) be the sum of the lengths in character state changes over all internodes along the direct path from OTU \( j \) to \( o \), the most recent common ancestor (root) of the taxon, summed over all characters. Quantity \( I_j \) has been called the patristic length of OTU \( j \) by Farris (1969) and others. I shall refer to it by the neutral term "path length," since it measures homoplastic as well as patristic resemblance (Sneath and Sokal, 1973). Next, let us define \( L_{\text{max}(u)} = \sum_j I_j \) as the sum of such lengths over all \( t \) OTUs. The quantity \( L_{\text{max}(u)} \) is the maximal length which the true cladogram could assume if it were rearranged so that every OTU evolves singly and independently from the ancestor \( o \), with the OTUs diverging from that ancestor in bushlike fashion and repeating separately for each OTU the changes in character states that actually occurred along the common evolutionary stems in the true cladogram. Thus \( L_{\text{max}(u)} \) is a measure which indicates the upper possible bound of parallelism and reversals.

**Fig. 3.** Standard phenogram of the 29 Recent Caminalcules based on the updated data set of this study. It is based on the data matrix with integer codes substituted for the measurements and was obtained by standardization of characters followed by computation of product-moment correlation coefficients and UPGMA clustering. The ordinate is in correlation coefficient scale. The phenetic genera are labeled with capital letters.

**Fig. 4.** True cladogram of the Caminalcules as furnished by J. H. Camin. Morphological change occurred during 19 time periods from time 1 to time 20 along the ordinate. Vertical lines indicate periods without morphological change, slanted lines indicate such change. The amount of evolutionary change (path length of the internode) based on 85 characters is shown by the length of the thickened bars along the slanted lines. To furnish an indication of scale, the length of the internode subtending OTU 54 is one Manhattan distance unit, that subtending OTU 76 is 10 such units. Path lengths based on 106 characters, necessary for differentiating fossil species, are shown in a subsequent publication (Sokal 1983b: fig. 2). Squares identify Recent
species, black circles fossil species. However, note that some fossil species extend into the Recent (e.g., species 8). Hollow circles indicate extinct species. Large hollow circles terminate extinct lineages whereas the tiny ones symbolize extinct species whose lineages continue with evolutionary change. The numbers next to squares and circles identify the species. Recent species have been given the numbers 1 through 29 familiar from the literature. Fossil species were assigned numbers 30 through 77 at random. Note that, although Camin indicated evolution between species 58 and 52, my coding shows no morphological change between these two forms. The internode between species 36 and 55 appears to be a similar case, but it exhibits morphological change for characters numbered 86–106. The phenetic genera of the Caminalcules are identified by brackets across the top of the tree.
Table 2. Summary of formulas of phenetic and evolutionary change.

- $l_{jo} =$ Sum of lengths in character state changes over all internodes along the directed path from OTU $j$ to most recent common ancestor $o$
- $L_{max(a)} = \sum l_{jo}$ where $\sum$ is summation over all OTUs
- $MD_{jo} = \sum |X_{ij} - X_o|$, where $\sum$ is summation over all characters and $X_i$ is the character state for character $i$ and OTU $j$. OTU $o$ is the most recent common ancestor
- $L_{max(i)} = \sum jMD_{jo}$
- $R_i = L_{max(i)}/L_{max(o)}$
- $r_{(i,o)} =$ Range in character states of character $i$ for the assemblage of $t$ OTUs plus the most recent common ancestor $o$
- $L_{min(i)} = \sum i r_{(i,o)}$
- $L_{max(a)} =$ Minimum length for the taxon, given any tree structure originating from the common ancestor $o$, and an evolutionary model for allowable types of character state changes
- $L_{est} =$ Sum of path lengths over all internodes, given knowledge of the true cladogram
- $L_{est} =$ Sum of path lengths over all internodes of an estimated cladogram
- $D_l = (L_{max(a)} - L_{act})/(L_{max(a)} - L_{min(l)})$ (Dendritic index)
- $H = L_{act}/L_{min(l)}$
- $H^* = L_{act}/L_{min(o)} = 1/C$, where $C$ is the consistency index of Kluge and Farris (1969)
- $DR = \sum jk (M_{kk} - MD_{jk})/\sum jk MD_{jk}$, where $\sum jk$ is summation of all OTU pairs except for pairs $jj$ and $kk$
- (deviation ratio of J. S. Farris)
- $L_{br,avg} = L_{act,avg}/(2t_M - 2)$, where the subscript $M$ refers to the subtaxon $M$
- $P_{est} = L_{br,avg}/(L_{pr} + L_{br,avg})$, where $p$ is the most recent common ancestor of subtaxon $M$ and $p'$ is the most recent common ancestor of that subtaxon that is also an ancestor of a Recent nonmember of $M$

*Presented in order of appearance in text. Quantities with asterisks are based on estimated rather than true cladograms.

in the data, given the distribution of character states over Recent OTUs and the character states of the fossil OTUs.

A generally shorter length for each OTU $j$ results if one computes

$$MD_{jo} = \sum |X_{ij} - X_o|,$$

where $X_i$ is the character state of OTU $j$ for character $i$ and $\sum$ is summation over the $n$ characters. Note that $MD_{jo}$ is the Manhattan distance between OTU $j$ and ancestor $o$. It is the minimum possible evolutionary length for each OTU $j$ from the ancestor $o$. We can assemble these lengths to form a new upper bound length of the entire tree $L_{max(l)} = \sum jMD_{jo}$ given a minimal length bush from the ancestor. Note that since $MD_{jo} \leq l_{jo}, L_{max(l)} \leq L_{max(o)}$. For any OTU $j$ other than the ancestor $o$ the ratio $R_{jo} = l_{jo}/MD_{jo} > 1.0$ is a measure of the ratio of unnecessary changes (reversals and repeated forward changes) among the evolutionary steps for that OTU. Consequently,

$$R_1 = L_{max(o)}/L_{max(l)} \geq 1.0$$

is a measure of the amount of reversals and repeats in character state changes for the entire taxon considered as a bush. Note that reversals near the base will be repeated in various OTUs and, hence, weighted more heavily.

Two measures of minimum length of the tree were considered. Define $L_{min(l)} = \sum r_{(i,o)}$, where $r_{(i,o)}$ stands for the range in character states of character $i$ for the assemblage of the $t$ OTUs in the taxon plus its most recent common ancestor $o$. Length $L_{min(l)}$ is the minimum amount of evolution necessary for producing the taxonomic diversity of the $t$ OTUs in the taxon. It is a theoretical quantity, rare-
ancestors, if ever, obtained in real data because of the actual distribution of character states over the OTUs. For $L_{mn(i)}$ to be realized, there would have to exist a cladogram for the taxon for which all characters would have to be compatible. In such a case there could be no reversals or parallelisms. Since this is a hypothetical, largely unattainable minimum length, a second minimum length $L_{mn(u)}$ needs to be defined as the minimum length for the taxon with $t$ OTUs, given any tree structure originating from the common ancestor $o$ and an evolutionary model for the allowable types of character state changes. Necessarily $L_{mn(u)} \geq L_{mn(i)}$ since some homoplasy is usually present.

Given knowledge of the true cladogram, the actual length of the tree, $L_{act}$, is the sum of the lengths over all the internodes. Thus it represents all the evolutionary changes over all characters that have taken place on this tree. Clearly, $L_{mn(u)} \leq L_{act} \leq L_{max(u)}$ but $L_{act}$ could be greater or smaller than $L_{mn(i)}$. The relations among the various quantities are illustrated in Figure 5.

Quantities $L_{act}$ and $L_{max(u)}$ are computable only in those extremely rare cases, such as the present one, in which the true evolutionary sequence is known; quantities $L_{mn(i)}$ and $L_{max(i)}$ require at least knowledge of the most recent common ancestor $o$. Quantity $L_{mn(u)}$ even with knowledge of the common ancestor, can be computed only by enumeration, which is practical for small numbers of OTUs only. For most real data sets, information is available only for recent OTUs. In such cases an estimated cladogram is produced by a numerical cladistic algorithm using an outgroup OTU believed to be close to the most recent common ancestor or a vector of character states believed to be primitive. As a result of the cladistic estimation process, the HTU ultimately specified as the most recent common ancestor of the group may be different from the outgroup first furnished. In these cases, the statistics $L_{max(i)}, L_{max(u)}, L_{mn(i)},$ and $L_{mn(u)}$ must be defined with respect to the estimated cladogram and the HTU representing the hypothesized most recent common ancestor. To distinguish these statistics based on estimated rather than true cladograms, I have added asterisks to their symbol. The equivalent of the quantity $L_{act}$ for the estimated cladogram obtained by a numerical cladistic procedure such as a Wagner tree algorithm is the length $L_{est}$. For any given data set and evolutionary model, given a specified root $ot$, $L_{act} \geq L_{mn(u)}^*$. This estimate by various computational algorithms will frequently not produce the minimum length $L_{mn(u)}^*$. Note that $L_{act}$ can be less than or greater than $L_{act}$.

These quantities permit computation of the following statistics. The saving in evolutionary length by the known tree structure can be defined as $L_{max(u)} - L_{act}$. Reversals and parallelisms are in both coefficients but they are more numerous in $L_{max(u)}$ since this is a bush-like structure repeating the path length of common shared stems separately for each OTU. A dendritic index can be defined

$$DI = (L_{max(u)} - L_{act})/(L_{max(u)} - L_{mn(i)})$$

which expresses the savings in length due to the tree’s dendritic structure departing from that of a bush as a proportion of the tree’s maximum possible range in length. If $DI = 0$, then the taxon is a bush and there is no common evolution. If $DI = 1$, then the character states are fully compatible on the cladogram and there are no parallelisms or reversals. Because basal internodes are more often involved in determining $L_{po}$ than are internodes near the tips of the tree, $DI$ is...
more heavily affected by savings in length in the dendritic structure near the base. For data sets with unknown cladogenies one can define

\[ D_I^* = \frac{(L_{\text{max}(u)} - L_{\text{est}}) - (L_{\text{max}(u)} - L_{\text{min}(l)})}{L_{\text{max}(u)} - L_{\text{min}(l)}} \]  

A second quantity, \( L_{\text{act}} - L_{\text{min}(l)} \) describes the excess in evolutionary length over the absolute hypothetical minimum. It may be useful to partition this excess into two parts as follows:

\[ L_{\text{act}} - L_{\text{min}(l)} = (L_{\text{min}(u)} - L_{\text{min}(l)}) + (L_{\text{act}} - L_{\text{min}(u)}) \]  

The first quantifies the necessary parallelisms to allow for the departure from a fully compatible distribution of character states over OTUs and the second term describes the extra parallelisms and reversals that occurred in the actual evolutionary history of the group over the minimum amount necessary to account for the observed distribution of character states over OTUs. Biologically speaking, however, the two terms may be difficult to distinguish since both describe departures from perfect consistency of character states with the cladogeny. Clearly, \( L_{\text{act}} - L_{\text{min}(l)} \) is a measure of homoplasy as it is conventionally understood and when expressed as a proportion of the maximum possible range of length of the tree \( L_{\text{act}} - L_{\text{min}(l)} \) is simply \( 1 - D_I \) (see Fig. 5).

An alternative statistic, adopted in this paper, is

\[ H = L_{\text{act}} / L_{\text{min}(l)*} \]  

which expresses the homoplasy as a ratio, necessarily greater than 1. The amount by which \( H \) is greater than unity is the extra length of the actual tree, expressed as a proportion of the minimum tree length necessary for evolution of the character states. For data sets with unknown cladogenies

\[ H^* = L_{\text{est}} / L_{\text{min}(l)*} \]  

Note that \( H^* = 1/C \) where \( C \) is the consistency index of Kluge and Farris (1969), also employed by Kluge (1976) and Mickevich (1978a).

A third measure of homoplasy is \( DR \), the deviation ratio featured in Farris’ WAGNER 78 program, which is the sum of the pairwise homoplastic distances divided by the sum of the Manhattan distances among all pairs of OTUs. The pairwise homoplastic distances for OTUs \( j \) and \( k \) are found as \( l_{jk} - MD_{jk} \), whereas the Manhattan distances are \( MD_{jk} \). Therefore,

\[ DR = \sum_{jk} (l_{jk} - MD_{jk}) / \sum_{jk} MD_{jk} \]  

where \( \sum_{jk} \) is summation of all OTU pairs \( jk \) except for pairs \( jj \) and \( kk \). This ratio is affected more by homoplasy at the base of the tree, even though the denominator also counts basal lengths repeatedly, because any excess length due to homoplasy will be counted more often for basal than for terminal internodes.

A second measure of character reversal \( R_2 \) is analogous to the \( DR \) ratio. It is defined as the sum of the pairwise distances due to reversals and nonparallel repeated forward changes divided by the sum of the homoplastic distances among all pairs of OTUs, which is the numerator of \( DR \). It indicates the proportion of homoplasy due to reversals. Its 1-complement is the proportion due to parallelisms.

In studying subtaxa within a larger taxon, as, for example, the genera in the present study, it is useful to define the above statistics also for subtaxa. In such a case when working with subtaxon \( M \), I employ the most recent common ancestor \( p_{M} \), or simply \( p \), of that taxon, replacing \( o \) by \( p \) in the above formulas. Summations over OTUs will then be carried out not over the \( t \) OTUs of the entire study, but only over the \( t_M \) OTUs of subtaxon \( M \).

It is useful to record the path length of the stem of each of the subtaxa. Following the earlier symbolism, this can be defined as \( l_{pp'} \), where \( p' \) is the most recent common ancestor that is also an ancestor of a Recent nonmember of subtaxon \( M \). An average branch length \( L_{br,avg} \) of the internodes subtended by the most recent common ancestor \( p \) is also useful. It can be computed by dividing \( L_{\text{act},M} \), the observed length of the tree representing the subtaxon, by \( 2t_{M} - 2 \), the number of its internodes. Comparing \( l_{pp'} \) with \( L_{br,avg} \) contrasts the evolution on the stem preceding
the most recent common ancestor of the genus with the subsequent evolution within the genus. The average branch length can be expressed as a branch length proportion,

\[ P_{\text{br}} = \frac{L_{\text{br,avg}}}{(l_{pp} + L_{\text{br,avg}})}. \]  

(9)

An additional complication arises when NCs are present as in this study. I have adopted two conventions, which are modifications of the Manhattan distance. In "transparent" measure the NC state is thought of as an unknown. For computation of distances between terminal OTUs, character states for OTU pairs, one or both of which had NCs for a given character, were ignored during the computation of the distance and NCs were passed over during computation of path lengths. Thus a 0 → NC → 3 → 4 path is four units long. In "opaque" measure NCs are considered a distinct character state and changes along the tree from an expressed character state to an unexpressed one (NC) are considered a single step. Successive NC states count as zero steps and a change from an NC to an expressed state is again a single step, regardless of the magnitude of the expressed character state. Thus the path 0 → NC → 3 → 4 is of length 3. Computing distances between terminal OTUs, the difference between two NCs was considered to be zero, that between any numerical state and an NC was considered to be one. Opaque distance was employed mainly because it made it simple to express path length for any internode along the tree and to partition the path length into homoplastic and reversal distances as needed in the next section of this paper. In transparent measure it is not always possible to uniquely estimate path length for each internode.

RELEVANCE OF CAMINALCULES FOR SYSTEMATIC INQUIRY

It is of interest to examine in how many ways the Caminalcules resemble data sets on real organisms, since this will strengthen the relevance of results obtained from them for systematic inquiry. Below I inspect as many separate aspects of the Caminalcules as seem to me relevant to systematic methods and principles. I also examine three of these aspects in 19 zoological data sets, ranging from 8 to 97 OTUs and based on from 20 to 139 characters lacking NCs. None of these data sets was subjected to the exhaustive analysis to which the Caminalcules were treated (recounted in subsequent papers of this series) for the obvious reasons that such a project would have taken several additional years, that finding outgroups for them would have taken much specialized knowledge, and that the true cladogenies of the data are of course unknown. As will be seen below, the results for almost all parameters bracket those obtained for the Caminalcules. It is not expected that the differences in tree topology that might be accomplished by repeated application of a cladistic algorithm to the same data set in the hope of finding a shorter tree would alter these relations with respect to homoplasy and other measures. I have summarized my conclusions in Table 3 for homoplasy, tree symmetry and adequacy of characters.

Homoplasy.—Three indices of homoplasy have been mentioned earlier and in the literature. These are the 1-complement of the dendritic index \((I - DI)\), the homoplasy ratio \(H\), which is the reciprocal of \(C\), the consistency index of Kluge and Farris (1969), and Farris' deviation ratio \(DR\). In the Caminalcules, two values can be obtained for each of these indices. One is based on the true cladogram and should be the correct measure of homoplasy in the group for the given index; the other, comparable to those furnished in the literature on real organisms, is based on the estimated cladogram. As an estimate I employed the approximate Wagner tree obtained with the WAGNER 78 program developed by J. S. Farris, using the distance Wagner procedure and midpoint rooting. Opaque distances were analyzed to permit computation of all homoplasy statistics. However, for the two statistics that can be computed in transparent measure (\(H\) and \(DR\)), these values are reported as well. Statistics based on estimated cladograms are distinguished by affixing an asterisk to their symbols. Approximate Wagner trees were computed for the 19 data sets, again using the WAGNER 78 program and rooted by the midpoint method. Because these data sets lack
NCs, the distinction between opaque and transparent measure does not apply to them. As an experiment, I also tried trees with an outgroup rooting using OTU 1 as the outgroup (clearly not a recommended procedure, although it has been used by some numerical cladists; e.g., Mickevich, 1978b; Farris, 1979a). Only midpoint rooting is reported in Table 3. Results for the OTU 1 rooting procedure were similar.

It is worth emphasizing again that, both for the Caminalcules and the 19 real data sets, the above method provides only approximate Wagner trees. Better estimates could have been obtained with further work and were indeed obtained for the Caminalcules (see Sokal, 1983a). I am employing the "cruder" estimate for the Caminalcules, since it is comparable to the estimates in the 19 data sets. Presumably homoplasy would decrease if the length of the trees could be shortened algorithmically. But, on the average, this would occur proportionately for all data sets and, since the Caminalcules are currently bracketed by the data sets on real organisms, they presumably would still be so bracketed after the length of all trees had been reduced.

In the Caminalcules $1 - DI = 0.1745$ and $1 - DI^* = 0.1326$. In the 19 data sets $1 - DI^*$ ranges from 0.0559 in the Leptopodomorpha (Schuh and Polhemus, 1980) to 0.4646 in 12 orthopteroid insects (Blackith and Blackith, 1968). Because the true trees for these real organisms are not known, the proper comparison with the Caminalcules is with $1 - DI^*$ from the approximate Wagner estimate.

For the true cladogram $H = 2.327$, and for the Wagner estimate $H^* = 1.417$ and 1.261 for opaque and transparent estimates, respectively, yielding corresponding consistency indices of 0.4298, 0.7057, and 0.8048. Mickevich (1978a) reported a range of consistency indices from 0.33 for *Aedes* and papillions to 0.86 for cytochrome C and globin. In the 19 data sets, $H^*$ ranges from 1.160 in the Leptopodomorpha to 6.62 in bees of the *Hoplitis* complex (Michener and Sokal, 1957),

---

**Table 3.** Minima and maxima of tree statistics for 19 zoological data sets and the Caminalcules.*

<table>
<thead>
<tr>
<th>Tree statistics</th>
<th>Low value</th>
<th>Caminalcules</th>
<th>High value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Estimated</td>
<td>True</td>
</tr>
<tr>
<td>Homoplasy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1 - DI^*$</td>
<td>0.0559b</td>
<td>0.1326</td>
<td>0.1745</td>
</tr>
<tr>
<td>$H^*$</td>
<td>1.160b</td>
<td>1.417</td>
<td>2.327</td>
</tr>
<tr>
<td>$C$</td>
<td>0.1495d</td>
<td>0.7057</td>
<td>0.4298</td>
</tr>
<tr>
<td>$DR^*$</td>
<td>0.1124b</td>
<td>0.1795</td>
<td>1.3591</td>
</tr>
<tr>
<td>Symmetry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BSUM2</td>
<td>0.2597d</td>
<td>0.4900</td>
<td>0.3779</td>
</tr>
<tr>
<td>BSUM3</td>
<td>-0.0976f</td>
<td>0.2420</td>
<td>0.0753</td>
</tr>
<tr>
<td>SHAQ2</td>
<td>0.4483h</td>
<td>0.6186</td>
<td>0.7720</td>
</tr>
<tr>
<td>SHAQ3</td>
<td>0.3421h</td>
<td>0.5225</td>
<td>0.7146</td>
</tr>
<tr>
<td>COLLESS2</td>
<td>0.1061f</td>
<td>0.3331</td>
<td>0.1693</td>
</tr>
<tr>
<td>Adequacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n/t</td>
<td>1.16h</td>
<td>2.93</td>
<td>2.93i</td>
</tr>
</tbody>
</table>

*High and low values refer to numerical values, not necessarily to the property described. Thus, high values of C indicate low homoplasy, and high values of BSUM2, BSUM3, and COLLESS2 indicate low symmetry (high asymmetry). The values shown are based on means of three runs of the Wagner 78 program with midpoint rooting for the symmetry measures, and on single estimates for the other measures. Other cladistic estimates were run as well; for details see text. The results conform to those shown in this table. The data sets exhibiting extreme values are identified in footnotes. For explanations of tree statistics see text.

b Leptopodomorpha (Schuh and Polhemus, 1980).

Colonial insects (Blackith and Blackith, 1968).

* Hoplitis complex (Michener and Sokal, 1957).

Western Bufo (Feder, 1979).

1 Drosophila (real set; Throckmorton, 1968).

2 Hemoglobin (Dayhoff, 1969).

3 Dasyuridae (Archer, 1976).

4 Is 5.34 if binary coded data are considered.

5 Pygopodidae (binary coding; Kluge, 1976).
corresponding to consistency indices of 0.8621 and 0.1495, respectively.

The deviation ratio (DR) for the true cladogram is 1.3591, a very high value, but DR* is 0.1795 and 0.3852 for opaque and transparent Wagner estimates, respectively. This great discrepancy can be accounted for by the numerous reversals in the true cladogram, especially along the stems defining the genera. Remember that the estimate optimizes the position of the HTUs and, in consequence, attempts to minimize homoplasy. The formula for computing the deviation ratio involves all pairwise distances between OTUs and, therefore, repeatedly counts these basal relations. Actually the parallelism component of the homoplasy is relatively low in the Caminalcules. For the lower triangular matrix in opaque measure, the sum of the homoplastic distances is 21,225, of which 17,093 is due to reversals and 4,132 is due to parallelisms. The Wagner tree algorithm in trying to obtain a shortest length tree cancels many of the reversals, so that the deviation ratio actually observed is relatively low, 0.1795, whereas the deviation ratio of the true cladogram is much higher. The range of observed values of DR* in the 19 data sets is from 0.1124 in the Leptopodomorpha to 1.7395 in the Hoplitis complex. If reversals are as common in real organisms as in the Caminalcules and if we knew their true cladograms, real organisms might well exhibit generally higher deviation ratios. Real and apparent homoplasy in the Caminalcules is well within the reported range for real organisms.

Symmetry.—The true cladogram of the Caminalcules is fairly symmetrical. No universally accepted criterion of symmetry has yet been established, but K. T. Shao (pers. comm.), who is investigating this problem, has used five separate indices to describe different aspects of symmetry. These are BSUM2 and BSUM3 (modified from Sackin, 1972), SHAO2 and SHAO3 (Shao, pers. comm.), and COLLESS2 (Colless, 1982). These coefficients were computed for the true cladogram of the Caminalcules as well as for several estimated cladograms of the Caminalcules. The estimated cladograms include one obtained by means of Joseph Felsenstein's PHYLIP WAGNER program, rooted at the true ancestor, and an estimated Wagner tree obtained with the WAGNER 78 program with midpoint rooting. The latter was replicated three times with randomly permuted input orders of OTUs.

These same statistics were also calculated for similarly estimated cladograms (except for the impossible true ancestor rooting) for the 19 data sets. Again, rooting using OTU 1 was employed additionally on an experimental basis for trees obtained by both programs. Not all data sets were suitable for each estimation method and statistic, but no reported range of values is based on fewer than 14 data sets. In Table 3 results are reported only for estimated cladograms based on the WAGNER 78 program and midpoint rooting, but below the other results are summarized as well. For WAGNER 78 with midpoint rooting, the true cladogram is contained within the range of observed symmetries for all coefficients except SHAO3. But the true cladogram shows more symmetry than the observed range of symmetry values for estimated cladograms based on the PHYLIP WAGNER program and on WAGNER 78 with OTU 1 as an outgroup. Yet, when estimated cladograms are computed for the Caminalcules, these are almost always less symmetrical than the true cladogram. In Table 3 for midpoint rooted Wagner trees the symmetries of estimated cladograms of the Caminalcules are contained within the range of observed values and, when other estimated cladograms are considered as well, this relation holds in 17 out of 20 comparisons. Since the other 19 classifications are all based on estimated cladograms, it is the estimated cladograms of the Caminalcules that must be compared with these data sets rather than the true cladogram. Thus on this basis, also, it appears that the Caminalcules are not atypical.

 Adequacy of the characters for resolving the cladogram.—The 29 Recent OTUs in a fully bifurcating tree would derive from 27 bifurcations, not counting the basal bifurcation at the root. Since one of the branching points in the true cladogram is a trifurcation, 26 synapomorphies are minimally needed to resolve the tree. How can we measure the
adequacy of the data set for this task? At the simplest level, one can count characters. Fewer than 26 binary characters cannot resolve the tree. Subjected to additive binary coding, the $85 \times 29$ data yield 155 binary characters—superficially a more than adequate number. However, such an approach is too crude since it does not allow for character correlations. In fact, we know that in addition to the single trifurcation, three bifurcations in the true cladogram (19-26, 11-21, 18-23) are not supported by any evident evolutionary change in the stems subtending them. Thus, even though there are more than five times as many binary characters in this data set than are necessary to define the tree, they are distributed across the tree in such a way as to make it impossible to define more than 23 branching points. Most numerical cladistic studies are carried out on far fewer characters and because the true cladograms of real organisms are unknown, it is generally not known whether the data matrix is adequate for resolving the true tree. In the 19 data sets the ratio $n/t$, where $n$ is the number of characters and $t$ the number of OTUs, ranges from 1.16 in *Dasyurus* (Archer, 1976) to 6.62 for binary coded members of the Pygopodidae (Kluge, 1976). These figures compare with 2.93 and 5.34 for multistate and binary coded Caminalcules, respectively (see Table 3). Since the values for the 19 data sets reflect differences in character coding, they must be interpreted with caution. Nevertheless, it is clear that the Caminalcules fall well within the bounds of data from the literature.

One can examine the true cladogram to determine the number of OTUs subtended by any given furcation. It ranges from 2 to 22 OTUs. These figures should minimally be matched by the numbers of OTUs (greater than one) sharing any one character state to provide synapomorphies for the recognition of these furcations. When the required distribution of OTU numbers is compared to the actual distributions of such numbers, one finds at least as many observed frequencies as are required. However, this is insufficient evidence for the adequacy of a data set because we already know that three bifurcations in the Caminalcules are not resolved in terms of the characters available, in addition to the known trifurcation—which goes counter to the assumptions of most cladistic methods. Correlated characters undoubtedly define furcations redundantly, and parallelism allows a single character to define furcations in different parts of the tree. Unfortunately, knowledge of this type is virtually impossible to obtain from real data sets, because the true cladograms are unknown. Using estimated cladograms for such inferences would tend to make the argument circular. In any case, it appears that the Caminalcules are unlikely to be less capable of having their cladogram resolved than most data sets in systematics.

Because of the incompleteness of the fossil record, less is known from real organisms of the following three aspects. However, it may be of some interest to treat these topics at least briefly in an effort to describe where the Caminalcules are located with respect to evolutionary parameters—evolutionary rates, species longevities, and speciation-extinction rates—that might be employed in simulation studies. It will be seen that the Caminalcules are not in contradiction with such findings as are reported in the paleobiological literature.

**Evolutionary rates.**—A frequency distribution of the path lengths (amount of evolutionary change) of each internode for each time period shows extreme clumping. When examined against Poisson expectations (Fig. 6), the overdispersion is highly significant ($P < 0.001$). Thus, there are many more periods of no evolution as well as more periods with substantial amounts of evolution than expected. Two explanations can be advanced for this phenomenon. (1) The Caminalcules are similar to real organisms in that their evolution is of organic form as a whole rather than independent for each character. Changes in form in the Caminalcules involve various correlated characters and thus those segments of the cladogram during which extensive morphological evolution occurred will exhibit greater amounts of change. (2) There also may be local clumping of evolutionary changes on the tree for nonbiological reasons which have an analog in real phylogenetic processes. We must as-
logical analog is to assume that major changes in environment or in genetic architecture leading to adaptive radiation are ongoing processes with a momentum of their own. The data in Figure 6 are based on only those lines leading to Recent OTUs. When all internodes including those leading to extinct terminal species are examined, the results are very similar to those already reported and illustrated in Figure 6. Similar data are hard to obtain in real organisms because of the incompleteness of the fossil record, but the observed pattern of evolutionary change is at least biologically plausible.

In Figure 7, I show a graph of the average evolutionary rates over all evolutionary lines for each of the 19 time periods (again, only for lines leading to Recent OTUs; the results including lines leading to extinct species are quite similar). The rates are computed from the 106 character × 77 OTU data base. There is clear evidence of differential evolutionary rates through time. Between times 7 and 11, rates were generally lower than at other times. By time 7, all genera except for DE and C had already been defined but major within-genus diversity had not yet begun except in genus A. A runs-up-and-down test (Sokal and Rohlf, 1981b), significant at $P <
0.01, demonstrates the alternating nature of the rate changes. Periods of higher change alternated with periods of lower change more frequently than could be expected by chance alone. Relevant comparisons with real organisms are hard to come by. That evolutionary rates differ over the fossil record has been well established since the work of Simpson (1944, 1949, 1953). Yet, because paleontological series are usually based on few characters and because the fossil lineages are not well known, estimates of evolutionary rates based on multiple characters as in Figure 7 are lacking. For single characters, rates do vary with time within species (Stanley [1979:58ff.] illustrates two examples) as well as within and among genera (Gingerich, 1974, 1979). Clearly, the pattern of evolutionary (phenetic) rates observed in the Caminalcules is not uncharacteristic of observations made on real organisms.

Species longevities.—The known cladogeny of the Caminalcules also permits an examination of the distribution of species longevities. These are shown in Figure 8 separately for the 29 Recent species, as well as for all Recent-and-fossil species in the group. The longevities of Recent species are shown in two different ways. In Figure 8A, I show the actual longevity from the time of origin of each species to the Recent or to the time of extinction. Expressed as a cumulative percentage, it can be read as the percentage of living species that extend backwards by various amounts of evolutionary time. Such curves are generally not used by paleontologists (Stanley, 1979:113), because the incompleteness of the fossil record would make estimates of longevities inaccurate. In the Caminalcules, of course, this graph is fully descriptive and accurate. This curve can be contrasted with that in Figure 8B, a so-called Lyellian curve, which depicts the percentage of Recent species that can be found in faunas of a given age. As expected, with an increase in age of the faunal assembly, the percentage of its species that survive to the Recent decreases. None of the five species extant at time period 4 survived to Recent times, the faunal assembly at time period 5 being the first one to contribute a member species to the Recent faunal assemblage. Although the total number of species in faunal assemblages was nondecreasing with time, the number of species surviving to the Recent did not increase proportionately. This explains the lack of monotonicity in Figure 8B. Hence, there are time intervals during which the number of species in the faunal assembly increased without a corresponding increase in the number of species surviving to the Recent, producing a lowering of the percentage of extant species in the fossil fauna.

In Figure 8A, the longevity curve for Recent OTUs can be compared with that including both Recent and fossil OTUs. Both curves resemble the hollow curves of similar data from the paleontological literature, with the Recent-and-fossil curve, comprising more data, presenting the smoother appearance as might be expected. Note that both curves decline to zero at time period 16. No Caminalcule species lived for more than that

![Graph](image-url)

**Fig. 8.** Species longevities. (A) Longevities of species in given time periods plotted as cumulative percent, i.e., the percentage of living species that extended backwards by various amounts of evolutionary time. Solid line—Recent species only; dashed line—Recent and fossil species. (B) Percentage of Recent species that can be found in faunas of a given age.
amount of time. Corresponding graphs in the paleontological literature generally approach an asymptote at a higher percentage because the faunal assemblages are not traced back a sufficient time for all the Recent species to disappear. The Lyellian curve, Figure 8B, is not atypical with respect to those published in the literature (see figs. 5-8 or 9-3 in Stanley, 1979).

The model under which Professor Camin evolved the Caminalcules requires each species to have a longevity of at least one time period. The observed distribution of longevities is highly clumped (P < 0.001 against Poisson expectations); there are many more species than expected becoming extinct after a single time interval and substantial numbers survive for quite long periods of time. The latter phenomenon seems to agree with observations on real organisms. The excess of short-lived species is more difficult to demonstrate in real organisms because there would be an inherent bias against observing these species in fossil assemblages. Moreover, real organisms are not limited to discrete time amounts as were the Caminalcules. That is, species existing for less than one time unit clearly must occur in the fossil record. We may conclude, however, that despite some peculiarities due to their mode of generation, the distribution of longevities in the Caminalcules resembles similar distributions observed in real organisms.

Speciation-extinction ratios.—In the Caminalcules, 56.0% of all evolutionary changes are accompanied by extinction, whereas 54.4% of lines showing no evolutionary change (stasis; vertical lines in Fig. 4) lead to extinctions. Thus there is no preference for one or the other type of process leading to extinction built into the evolutionary tree.

The availability of complete records in the Caminalcules permits computation of the net rate of increase in number of species (R), speciation rate (S) and extinction rate (E) for all time periods. In real data, R and E are usually approximated by various indirect approaches and S is estimated as \( S = R + E \). Here these values can be directly computed and they are \( R = 0.213, \) \( S = 0.538 \) and \( E = 0.324 \). Note that if \( R \) is estimated using an exponential growth model (Stanley, 1979: 104), the resulting value is 0.177, an underestimate of the true rate. The absolute values of \( R, S, \) and \( E \) cannot be contrasted with those of real organisms since the time intervals chosen for the Caminalcules are arbitrary. But the relative magnitudes of these quantities can be compared. The \( S/E \) ratio in the Caminalcules is 1.661. This compares with similar ratios estimated under varying assumptions for Plio-Pleistocene mammals of Europe which range from 1.310 to 2.048, and for temperate and subtropical bivalves of the Pacific which range from 1.667 to 3.333 (Stanley, 1979:117). Both of these groups are considered by Stanley to be undergoing active adaptive radiation into the present. By this criterion, the Caminalcules also do not differ markedly from real organisms.

Conclusions.—By the six criteria examined above, the Caminalcules fall well within the range of observed values for real organisms.

Acknowledgments

This is Contribution No. 463 in Ecology and Evolution from the State University of New York at Stony Brook. The study was greatly aided by two unusually competent assistants. Leora Ben Ami initiated the recoding of the characters. Barbara Thomson completed this task, took over the very considerable amount of computations, and prepared the tables, charts and illustrations, becoming so engrossed in this work that she became transmogrified into a Caminalcule on Halloween’s eve 1982. Throughout this work, I have had the unstinted counsel of my friend and colleague F. James Rohlf. Much advice and computational assistance were given by other members of the Stony Brook Numerical Taxonomy Group: Kent Fiala, Gene Hart, and Kwang-tso (Chet) Shao. I am indebted to Joseph Felsenstein for his valuable PHYLIP program package, to Kent Fiala for use of the CLINCH program, and to Leslie F. Marcus for furnishing our laboratory with a copy of the WAGNER 78 program. The manuscript benefited greatly from a critical reading by Merss. Fiala, Hart, Rohlf and Shao as well as by critical reviews by Raymond B. Phillips, University of Oklahoma, and two anonymous reviewers. The
sections on evolutionary rates, species longevities, and speciation-extinction ratios were read by Michael Bell, Lev R. Ginzburg, and Jeffrey S. Levinton at Stony Brook, who provided useful comments. Barbara McKay uncomplainingly wordprocessed innumerable versions of this manuscript. Joyce Schirmer prepared the final illustrations. Last but far from least I must acknowledge my debt to the late Professor Joseph H. Camin, my former colleague at the University of Kansas. Without his fanciful imagination coupled with a thorough understanding of systematics, there would not have been any Caminacules. The computations were carried out on a UNIVAC computer at the Stony Brook Computer Center and on the MODCOMP minicomputer facility at the Department of Ecology and Evolution. This research was supported by grant DEB 80-03508 from the National Science Foundation, whose continued support of my research is much appreciated.

REFERENCES


SCHUH, R. T., AND J. T. POLHEMUS. 1980. Analysis of taxonomic congruence among morphological, eco-
logical, and biogeographic data sets for the Lep- 
SIMPSON, G. G. 1944. Tempo and mode in evolution.
SIMPSON, G. G. 1949. The meaning of evolution. Yale
Univ. Press, New Haven, Connecticut.
SIMPSON, G. G. 1953. The major features of evolu-
SNEATH, P. H. A. 1982. [Review of] Systematics and
biogeography: Cladistics and vicariance. Syst. Zool.,
31:208–217.
taxonomy. W. H. Freeman, San Francisco.
SOKAL, R. R. 1974. Classification: Purposes, princi-
SOKAL, R. R. 1983a. A phylogenetic analysis of the
Caminalcules. II. Estimating the true cladogram.
SOKAL, R. R. 1983b. A phylogenetic analysis of the
Caminalcules. III. Fossils and classification. Syst.
Zool. (in press).
SOKAL, R. R., and F. J. ROHLF. 1966. Random scan-
congruence in the Leptopodomorpha re-examined.
SOKAL, R. R., and F. J. ROHLF. 1981b. Biometry. 2nd
ed. W. H. Freeman, San Francisco.
STANLEY, S. M. 1979. Macroevolution, pattern and
process. W. H. Freeman, San Francisco.
THROCKMORTON, L. H. 1968. Concordance and dis-
cordance of taxonomic characters in Drosophila
York.

Received 14 January 1983; accepted 11 April 1983.

APPENDIX

List of Characters for 77 Fossil and Recent Caminalcules\(^a,b\)

Head and neck

1. Head junction complex (folded) (1) or not (0). [73, 39]
2. If complex, degree of folding complete (1) or partial (0). [73, 10]
3. If partial, secondary junction narrow (0) or broad (1). [10, 21]
4. P (1) or A (0) of horn. [75, 73]
5. If horn present, pointed (1) or flattened (0). [75, 64]
6. Length of head in mm from rear end of folded section to front, excluding horn and anterior projections.
If head not complex, measure from collar. Recoded as integer in the following manner: (0) 9–10.9; (1) 10.9–12.8; (2) 12.8–14.7; (3) 20.4–22.3; (4) 24.2–26.1; (5) 26.1–28. [73, 65, 60, 28, 25, 39]
7. Anterior end of head concave (0), flat (1) or convex (2). [71, 73, 62]
8. If convex, rounded (0) or sharply pointed (1). [62, 15]
9. P (1) or A (0) of anterior projections. [71, 73]
10. P (1) or A (0) of eyes. [73, 59]
11. If present, states of fusion of eyes: (0) two separate eyes; (1) grown together but two eyes still discernible; (2) grown together into oblong approximately the size of two eyes; (3) grown together into circle approximately the size of one eye. [73, 47, 1, 16]
12. P (1) or A (0) of eye stalks. [37, 73]
13. If stalked, length of stalk (excluding eye) in mm. Recoded as integer in the following manner: (0) 3–4.5; (1) 4.5–5.9; (2) 6–7.5; (3) 10.5–12; (4) 13.5–15; (5) 16.5–18. [37, 38, 70, 48, 72, 19]

---

\(^a\) These characters were originally defined by A. J. Boyce and I. Huber and were revised by B. Thomson and R. R. Sokal.
\(^b\) Abbreviations and conventions: P—presence; A—absence. Numbers in parentheses are character state codes. Numbers in square brackets following definition of each character are the OTU numbers of representative OTUs that exhibit the character states in the order described in the list. Thus for character 1 [73, 39] means that OTU 73 is an example of a complex (folded) head junction and OTU 39 is an example of a not complex head junction. Whenever an "IF" statement is denied, as in character 2 for OTUs whose state for character 1 is 0, an NC is recorded for the OTU for that character. Since all measurement char-
acters were recorded to the nearest mm, there is no ambiguity created by shared, more refined class limits of adjacent classes as listed for some characters (e.g., for character 6, a measurement of 10 mm was coded 0, while one of 11 mm was coded 1).
The character numbers 1–85 employed in this list correspond to those employed in earlier studies and originally defined by Boyce and Huber. Characters numbered 86–106 were needed to describe additional observed differences in fossil species. In order to preserve both the old character numbering system and at the same time the logical order of characters in the list, it became necessary to list some char-
acters out of numerical order. The following paragraph gives the lo-
cation of all characters that are not in strict numerical order in the list.
Characters 26–28 in this list follow upon character 22; 58 follows 13; 59 follows 56; 60 and 61 follow 48; 62 follows 47; 78 follows 80; 84 follows 90 (23); 85 follows 63; 86–90 follow 23; 91 and 92 follow 34; 93–95 follow 35; 96 follows 45; 97 follows 62 (47); 98 follows 61 (48); 99 follows 55; 100 and 101 follow 85 (63); 102 follows 67; 103 follows 69; and 104–106 follow 75.
The following lists of characters comprise the several body regions: head and neck. 1–17, 58 (total of 18); anterior appendages, 18–30, 84, 86–90 (total of 19); posterior appendages, 31–45, 91–95 (total of 13); collar and abdomen, 39–57, 59–65, 85, 96–101 (total of 31); abdominal pores, 64–83, 102–106 (total of 25). The overall total is 106 characters.
The images shown in Figure 1 (OTUs 1–29) were traced from the originals, whereas those in Figure 2 were traced from xerographs of the originals. In the process and in the reduction necessary for illustration here some detail visible to the data coders was necessarily lost. The states of at least one character (17) are quite indistinguishable in the pub-
lished figures. Lengths in millimeters are given in terms of the di-
ensions of the original images. As reduced in Figures 1 and 2, these measurements must be multiplied by 0.837.
58. If stalked, stalks fused (1) or not (0). [20, 37]
   14. If stalks not fused, tips divergent (0), parallel (1) or convergent (2). [37, 19, 31]
15. Top of head depressed (0), flat (1) or crested (2). [73, 41, 53]
   16. If crested, single (0) or lobate (1). [53, 2]
17. P (1) of A (0) of groove in neck. [12, 73]

Anterior appendages

18. P (1) or A (0) of anterior appendages. [73, 15]
   19. If appendages present, length in mm from tangent to posterior end of abdomen (excluding posterior appendages) up to the anterior end of the longer appendage. Recoded as integer in the following manner: (0) 40-46.3; (1) 46.3-52.6; (2) 52.6-58.9; (3) 58.9-65.2; (4) 65.2-71.5; (5) 71.5-77.8; (6) 77.8-84.1; (7) 84.1-90.4; (8) 96.7-103. [73, 74, 45, 43, 66, 24, 17, 19, 20]
20. If appendages present, P (1) or A (0) of flexion (elbow) in appendages. [41, 73]
21. If appendages present, end of appendage divided (1) or not (0). [56, 73]
   22. If not divided, end of appendage is tendril (0), sharply pointed (1), tapered or rounded (2) or knobbed (3). [21, 68, 73, 45]
26. If not divided, P (1) or A (0) of flange. [38, 73]
   27. If flange present, length expressed as sum of right and left sides in mm. Recoded as integer in the following manner: (0) 17-20; (1) 23-26; (2) 32-34.9; (3) 35-37.9; (4) 38-41; (5) 44-47. [38, 70, 48, 50, 20, 29]
28. If flange present, width expressed as sum of right and left sides in mm. Recoded as integer in the following manner: (0) 5-6.9; (1) 10.7-12.6; (2) 14.5-16.4; (3) 16.4-18.3; (4) 18.3-20.2; (5) 20.2-22.1; (6) 22.1-24. [38, 70, 48, 31, 29, 72, 26]
23. If divided, two divisions (0) or three (1). [22, 56]
   24. If two divisions, ending as fingers (0), clamp (pincers without joint) (1), pincers with joint (2) or two wavy spikes (3). [22, 49, 30, 67]
   87. If ending as clamp, closed (0) or open (1). [49, 36]
   88. If open, clamp is small (0) or large (1). [36, 55]
   89. If ending as pincers with joint, serrated edges P (1) or A (0). [57, 30]
   90. If three divisions, one “finger” much longer than the other two (2), one finger slightly longer than the other two (1) or all fingers the same size (0). [65, 76, 56]
34. If divided, bulbs on ends of divisions present on all divisions (2), present on some divisions (1) or absent (0). [18, 66, 56]
   24. If bulbs present on all divisions, P (1) or A (0) of claws. [12, 46]
   25. If bulbs present on all divisions, P (1) or A (0) of pads. [75, 46]
29. If appendages present, P (1) or A (0) of pigment. [9, 73]
   30. If pigment present, distributed in small dots (1), small circles (0), large circles (1) or very broad areas (2). [33, 9, 11, 21]

Posterior appendages

31. Appendage single (0), partially double (1) or completely double (2). (Completely double means that appendage originates from 2 stalks rather than 1; it does not necessarily mean that the 2 halves are separated.) [73, 43, 18]
32. If single, disclike (pedestal like) (0), plate like (1) or propellerlike (2). [73, 40, 36]
   33. If disclike, width of disc in mm. Recoded as integer in the following manner: (0) 5-5.7; (1) 5.7-6.4; (2) 6.4-7.1; (3) 7.8-8.5; (4) 8.5-9.2; (5) 9.9-10.6; (6) 10.6-11.3; (7) 11.3-12. [13, 58, 39, 38, 37, 72, 42, 77]
34. If disclike, stalk long (>3 mm) (2), short (≤3 mm) (1) or absent (0). [74, 73, 28]
   35. If disclike, round (0) or square (1). [73, 74]
   91. If round, P (1) or A (0) of division lines. [49, 74]
   92. If square, P (1) or A (0) of division lines. [49, 74]
35. If plate like, no indentation (1), weakly emarginate (curved indentation) (0), cleft (V-indentation) (1), or deeply cleft (V-indentation continued into division line) (2). [40, 33, 11, 6]
93. If propellerlike, blades rounded (0) or pointed (1). [36, 67]
94. If propellerlike, stalk short (<5 mm) (0) or long (>5 mm) (1). [36, 55]
95. If stalk long, straight (0) or twisted (1). [55, 30]
36. If completely double, ends pointed (0), clubbed (1) or plate like (2). [3, 75, 18]
   37. If plate like, divided (1) or not (0). [66, 18]
   38. If divided, two divisions (0) or three (1). [66, 53]

Collar and abdomen

39. Rim of abdomen plain (0), narrowly raised (1) or broadly raised (2). [64, 45, 73]
40. Anterior margin of abdomen well delineated (0) or imperfectly delineated (1). [73, 31]
41. P (1) or A (0) of abdominal ridge (raised area on central posterior abdomen). [73, 45]
Abdominal

64. P (1) or A (0) of large pore in column 10 (most anterior position). [45, 73]
65. Group II (columns 4 and 5) has zero pores (~1), one pore (0) or two pores (1). [61, 69, 73]
66. If two pores, fused (1) or not (0). [5, 73]
67. If fused, fusion complete (0) or incomplete (1). [48, 5]
68. If not fused, number of pores looking like slits: zero (0), one (1) or two (2). [73, 68, 51]
69. Group III (columns 6–10) has one (0), two (1), three (2) or five (3) pores. [64, 18, 43, 73]
70. If two or three pores, one pore looks like a slit (1) or not (0). [43, 76]
71. If five pores, P (1) or A (0) of fusion. [50, 73]
72. If fusion present, complete (0) or incomplete (1). [29, 50]
73. If incomplete, P (1) or A (0) of a group of three fused pores. [26, 50]
74. Group I (columns 1–3) has zero (0), one (1), two (2) or three (3) pores. [47, 24, 35, 73]
75. If one or two pores, one pore looks like a slit (1) or not (0). [63, 35]
76. If two or three pores, pore nearest posterior enlarged (1) or not (0). [42, 73]
77. If three pores present and posterior one enlarged, enlarged pore apparently see-through (1) or not (0). [30, 42]
78. If see-through, fused with posterior pore from other side's Group I (1) or not (0). [57, 30]
76. If three pores, separated as three open pores of any size (0) or in some way modified (1). [73, 46]
77. If modified, modified by fusion (1) or some other way (0). [64, 46]
79. If fusion, two (0) or three (1) pores fused. [50, 64]
80. If fusion, broad and complete (0), broad and incomplete (1), or slit-like (2). [64, 66, 71]
78. If broad and complete and three pores fused (char. 79), both lateral pores reduced and joining median one (1) or not (0). [22, 64]
81. If modified in some other way, P (1) or A (0) of line (or slit) connecting pores. [46, 65]
82. If present, two (0) or three (1) interconnected pores. [53, 46]
83. If absent, slits in column 1 (0), column 2 (1), columns 1 and 2 (2), or column 3 (3). [60, 65, 8, 51]