Hydrobiologia **292/293**: 31–39, 1994.
F. D. Ferrari & B. P. Bradley (eds), Ecology and Morphology of Copepods.

Copepod phylogeny: a reconsideration of Huys & Boxshall's 'parsimony versus homology'

Ju-shey Ho
Department of Biology, California State University, Long Beach, CA 90840–3702, USA

Key words: Cladistics, copepoda, homology, parsimony, phylogeny, polarity, tree length

Abstract

A hypothesis of copepod phylogeny proposed by Huys & Boxshall (1991) was reanalyzed with the PAUP computer program by using TOPOLOGY and DELTRAN options to remove many character reversals. The remaining reversal characters were removed employing the Camin-Sokal Optimization Procedure. The resulting tree is two steps longer than the best fit tree (Tree C) obtained strictly by using parsimony criterion and optimizd with the same procedure. Further analysis revealed that the differences in these two cladograms are essentially due to the differential treatment of nine characters and varied assumptions of 'missing state' in another six characters as being apomorphies or plesiomorphies. Huys & Boxshall's hypotheses of character polarity in the copepods are shown not to be in conflict with the principle of parsimony. According to the principle of parsimony, their Tree C is the best supported phylogenetic hypothesis for Copepoda.

Introduction

In organismal biology, cladistics is a systematic methodology that attempts to produce phylogenetic hypotheses of genealogical (ancestor-descendant) relationships among monophyletic groups of organisms. The popularity of cladistics has grown steadily among systematists since the mid-1960s because it provides 'scientific method' for the conversion of taxonomic observations into 'testable' hypotheses.

To produce a new phylogeny for copepoda, Huys & Boxshall (1991) examined over 200 species of copepods among ten orders, and selected 54 characters for cladistic analysis. Although there were some difficulties in polarizing and coding some character states, the analysis of Huys & Boxshall is the most rigorous one ever attempted for the orders of Copepoda. Although cladistic analysis was initially employed by Huys & Boxshall, the cladograms (phylogenetic trees) that resulted were rejected on grounds that the primary criterion of cladistics – parsimony – conflicted with the authors' concept of character 'homology'. Thereupon, Huys & Boxshall (1991: 400) proposed a new scheme of phylogenetic relationships based strictly on the *a priori* assumptions about character evolution in the Copepoda.

The purpose of this paper is to show that the concepts of parsimony and homology need not conflict in conducting a phylogenetic analysis using cladistics, and, furthermore, there are provisions in PAUP program (Swofford, 1985) for *a priori* assumptions about character evolution if one does not believe in character reversability. The necessity of invoking the principle of parsimony in tree selection will also be discussed to point out the inadequacy of the copepod phylogeny proposed by Huys & Boxshall (1991). The character matrix of Huys & Boxswhall is reproduced here in Table 1 with a correction of character 51 (absence of praecoxa on male fifth legs) for the Calanoida. It was erroneously assigned a plesiomorphic state in their Table 13 (see Huys & Boxshall, 1991: 394).

Parsimony – the principle in phylogeny reconstruction

By employing cladistics, Huys & Boxshall (1991) obtained four shortest (most parsimonious) trees

Table 1. Character matrix (modified after Huys & Boxhall, 1991)

TAXON	CHARACTER										
	1 5	10	15	20	25	30	35	40	45	50	54
HYPAN	00000	00000	00000	00000	00000	00000	00000	00000	00000	00000	0000
PLATY	00000	00000	01000	00110	09990	99011	00000	00090	00000	10000	0010
CALAN	00110	01011	00100	00000	10000	00000	11111	11100	00000	00100	0000
MISOP	01110	01012	01001	11110	11000	10000	11111	11101	11010	10100	1011
HARPA	01111	11112	01010	11110	11000	10001	11111	11111	11011	10110	1010
MONST	11111	11112	11190	99999	99999	99999	11111	11111	11111	11111	1119
MORMO	01111	91112	11110	10110	21000	99910	11191	11922	22999	99999	9999
GELYE	01111	91112	11101	11110	21110	11111	11991	19999	99999	99999	9999
CYCLO	01111	01112	11101	99010	21110	11111	11111	11101	11010	09100	1001
SIPHO	11111	11112	11110	99111	29991	99912	11111	11101	11111	11111	1111
POECI	01111	11122	11110	99111	29991	99911	11111	11111	11011	19110	1111

Abbreviations: HYPAN = hypothetical ancestor (= outgroup); PLATY = Platycopioida
CALAN = Calanoida; MISOP = Misophrioida; HARPA = Harpacticoida; MONST = Monstrilloida
MORMO = Mormonilloida; GELYE = Gelyelloida; CYCLO = Cyclopoida; SIPHO = Siphonostomatoida
POECI = Poecilostomatoida

Table 2. Number of steps (= tree length) and number of reversals for the trees analyzed with and without the DELTRAN command in PAUP analysis.

Cladogram (Tree)	Without DELTRAN		With DELTRAN		
	Tree length	No. of reversals	Tree length	No. of reversals	Adjusted length*
Tree A	82	17	82	8	90
Tree B	82	18	82	9	91
Tree C	82	18	82	5	87
Tree D	82	18	82	7	89
HB-Tree	86	22	86	4	89
HO-Tree	97	39	97	8	112
STOCK-Tree	136	53	136	48	136

* Tree lengths after adjusting with Camin-Sokal Optimization.

(cladograms) for Copepoda, each with a length of 82 steps. These trees were respectively called Trees A, B, C and D and are illustrated as Fig. 4.3.1 in their book *Copepod Evolution* (1991). Because all four of these trees contained 'a significant number of character reversals' (p. 397) that conflicted with their interpretations of character state polarity (direction of evolutionary change of the homologous character states), all were rejected by Huys & Boxshall. Seventeen or 18 characters on each of these four trees had reversals (see Table 2), which denote a change in a character state from a plesiomorphy (= 0) to an apomorphy (= 1) and back to a plesiomorphy, or from a more apomorphic state (= 2) back to a less apomorphic state (= 1). Huys & Boxshall (1991: 397–398) discussed the apparent conflicts between their postulated evolutionary tree and that implied by Tree A of the two characters (17 and 18) exhibiting character reversals. However, similar discussions were not attempted for the other reversed characters. Ultimately, they concluded that 'evolution within the Copepoda has proceeded primarily by oligomerization... and character

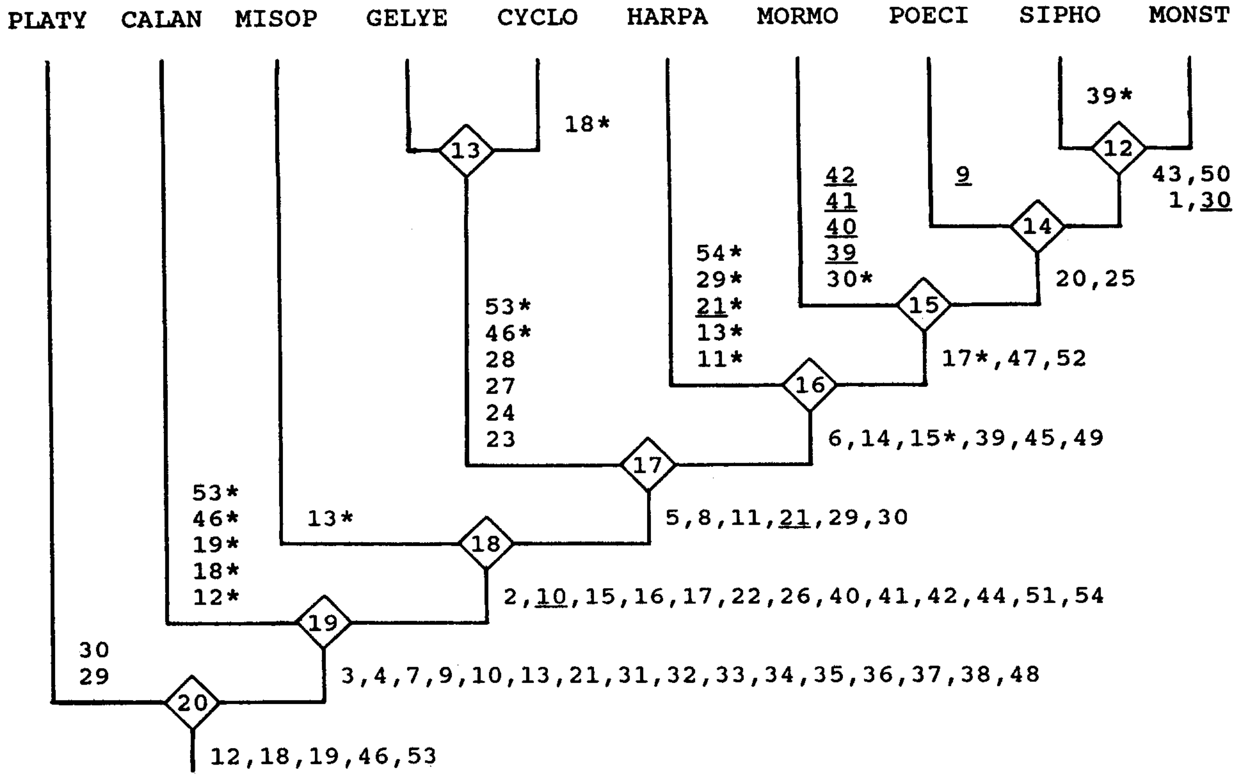

Fig. 1. Tree C generated by PAUP program with 18 character reversals. Asterisk indicates reversal; underline indicates a state coded with '2' in the data matrix. (Tree length = 82 steps)

reversal appears to be an extremely rare event.' (Huys & Boxshall, 1991: 399).

It is possible to reduce the number of reversals on a cladogram by a process called 'optimization' using DELTRAN in a cladistic analysis employing the PAUP program (Swofford, 1985). Via this process, the number of reversals in the aforementioned four trees of Huys & Boxshall (1991) were reduced from either 17 or 18 to 8, 9, 5 and 7 reversals, respectively, for Trees A, B, C and D (see Table 2). The remaining reversals can subsequently be removed employing the Camin-Sokal Optimization Procedure. The resulting trees are longer in length; nevertheless, they are not only free of reversals but also retain the original topology. This procedure of removing reversals without altering the tree topology is explained below using Tree C.

With the default ACCTRAN optimization, the PAUP analysis produces a tree with 18 reversals (see Fig. 1). By using DELTRAN, which reduces the number of reversals by replacing earlier transformations with parallel transformations in later branches and forces the residual reversals to the terminal branches, the character states on this tree are modified as shown in Fig. 2. The reversals are now seen only on four branches: 18* on 13-Cyclopoida, 54* on 16-Harpacticoida, 17* and 30* on 15-Mormonilloida and 39* on 12-Siphonostomatoida. To remove these five residual reversals, one needs to refer to the states listed in the character matrix (see Table 1). For instance, character 17 has an apomorphic state (= 1) found in three orders: Misophrioida, Harpacticoida and Gelyelloida. To remove 17* from the clade 15-Mormonilloida one needs to keep in mind that only each of these three orders with the apomorphic state for this character can receive an assignment of 17. When 17 is eliminated from the clade 18–19, the 17* on the clade 15-Mormonilloida is also removed. The obvious next step is to add 17 to each clade terminating in those three taxa bearing the known apomorphic state for this character. Notice that the topology of the tree is not altered, but one step is added to the tree (i.e., 2 steps were taken away and 3 steps were added). Figure 3 shows the new pattern of character distribution when similar

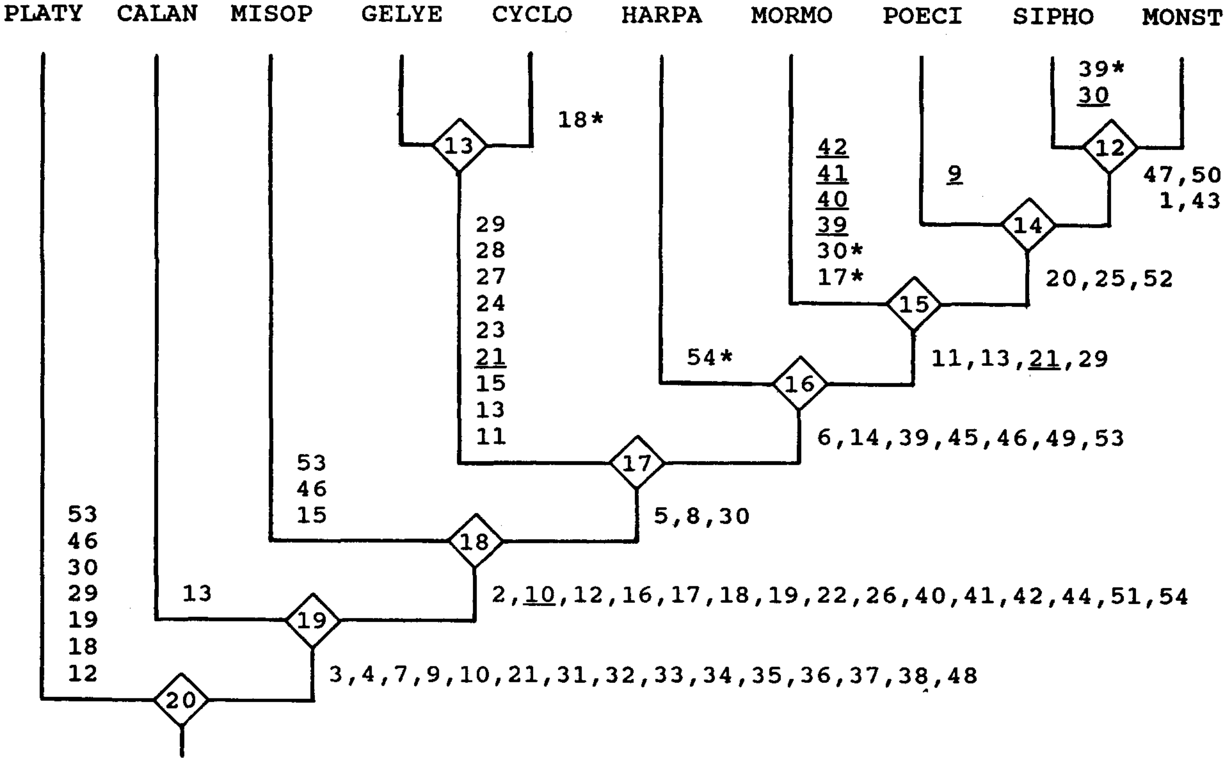

Fig. 2. Tree C generated by PAUP program with DELTRAN in effect. Notice the number of character reversals is reduced to 5. Asterisk indicates reversal; underline indicates a state coded with '2' in the data matrix. (Tree length = 82 steps)

adjustments of all reversals are completed. Again, note that this tree is free of reversals.

Homology – principle in Huys & Boxshall's copepod phylogeny

The copepod phylogeny endorsed by Huys & Boxshall (1991) is reproduced in Fig. 4 (for apomorphic character states supporting this tree, see Huys & Boxshall, 1991: 402–405). Certain branches on this tree are characterized by some uniquely derived character states. For instance, fusion of antennary exopodal segments IV and V on Clade 20-PLATY, asymmetrical male fifth legs on 19-CALAN, loss of first exopodal segment of mandible on Clade 13-CYCLO, etc. However, none of these characters was used by Huys & Boxshall in their cladistic analysis and, therefore, cannot be placed on this tree. The phylogeny depicted by this tree was accepted by Huys & Boxshall (1991: 400) because it made 'a positive statement on the homology of every character and character states in the data matrix'. But, what happens if there is more than one hypothesis showing no conflicts with the perceived character state polarity? Huys & Boxshall (1991) did not pursue this possible problem.

Essentially, the cladogram reproduced in Fig. 4 was obtained by Huys & Boxshall (1991: 400) with 'differential weighting of two characters' in PAUP analysis. However, they referred to it as being generated by hand, 'since to present it as generated by PAUP (but using arbitrary weights) would give it an apparent but undesired objectivity'. I reconstructed Huys & Boxshall's choice of cladogram by using the PAUP command TOPOLOGY. The tree length and the number of reversals before and after the use of DELTRAN on this reconstructed cladogram are given in Table 2 under HB-Tree. Notice that the phylogenetic scheme considered by Huys & Boxshall (1991) as acceptable contains 22 reversals. However, by applying the procedure described above to remove reversals the tree could be converted into one with exactly the same tree topology and pattern of character distribution as given by Huys & Boxshall (see Fig. 4). The length of this tree

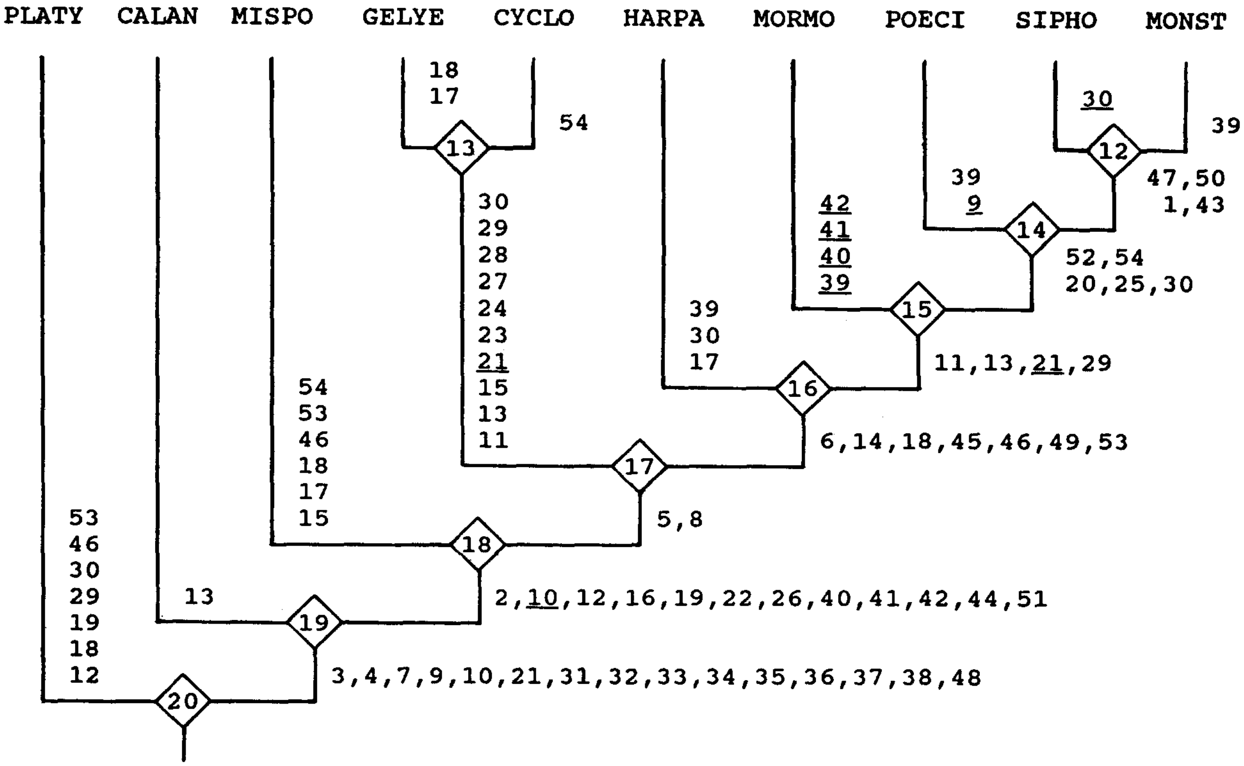

Fig. 3. Tree C generated by PAUP program with DELTRAN in effect and 5 residual reversals removed by employing Camin-Sokal Optimization Procedure. Notice the tree topology remains unchanged. Underline indicates a state coded with '2' in the data matrix. (Tree length = 87 steps)

(Fig. 4) is 89 steps, which is two steps longer than the best fit tree (i.e., Tree C) produced by PAUP-analysis (Fig. 3).

As mentioned above, Tree C was rejected by Huys & Boxshall (1991) because it contained reversals and conflicted with their 'rigorous definition of homology'. (p. 397) (Huys & Boxshall erred in using the word 'homology' in this context. They should have said: rigorous analysis of the polarity of character states). However, the newly fitted Tree C is free of reversals. It is important to determine if the character transformations depicted in the refitted Tree C are congruent with Huys & Boxshall's concepts of character polarity provided in their data matrix (Table 1). To perform this examination, I compared the differences in the pattern of character distribution between the accepted (HB-Tree) and the rejected (Tree C) hypotheses.

Parsimony versus homology?

HB-Tree shows 39 convergences for 14 characters and Tree C, 37 convergences for 13 characters (see Table 3). Due to topological difference between these two trees, only those convergences shown by characters 12, 18 and 39 are found in both trees on the branches with the same component taxa (cf. Figs 3 and 4). In other words, these three characters were treated in the same manner by these two phylogenetic hypotheses. Thus, characters 12, 18 and 39 are excluded from further scrutiny. Three other characters, 6, 45 and 49, occur only once in both trees. However, the state of 'missing' (coded with a '9' in the PAUP-analysis) in Mormonilloida for these three characters is interpreted differently in the two trees. These three characters along with the other 12 characters need further examination.

Differences in the 15 aforementioned characters can be classified into two types (see Table 4): (1) difference in the interpretation of a 'missing state' and (2) difference in the treatment of an apomorphy. The dis-

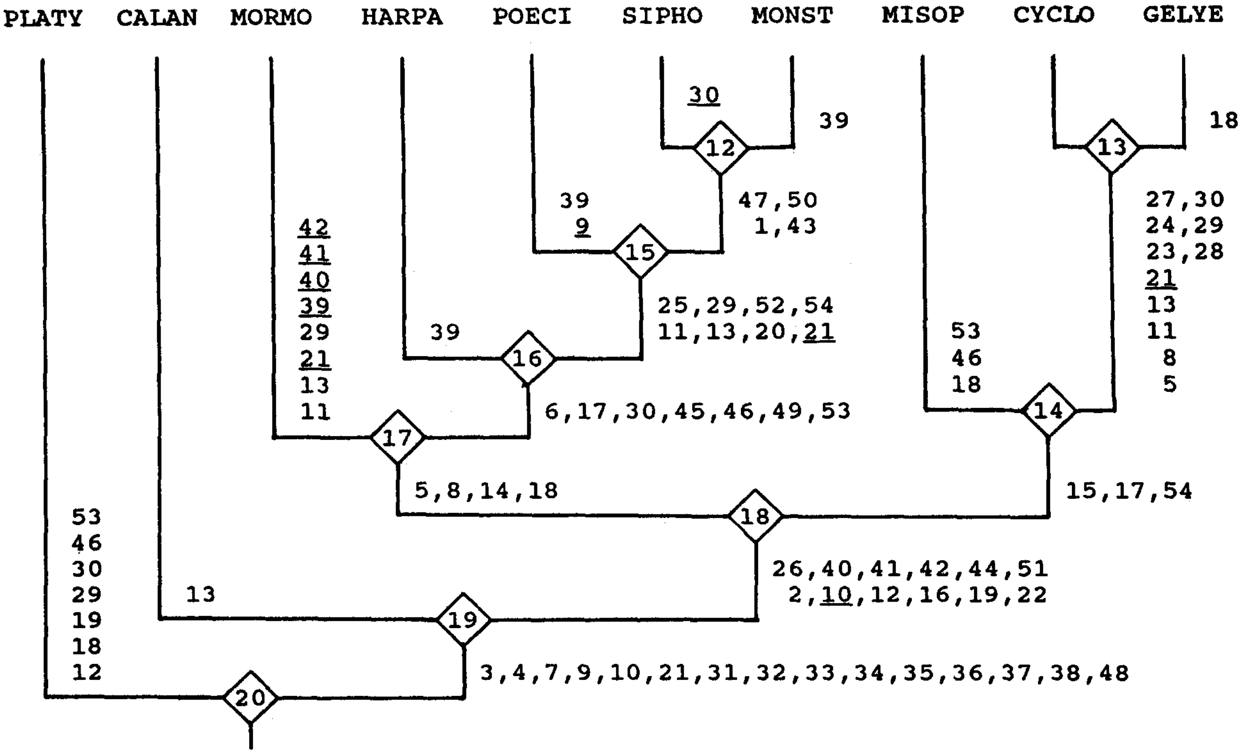

Fig. 4. The phylogenetic hypothesis proposed by Huys & Boxshall (1991) with the distribution of characters on the tree. Underline indicates a state coded with '2' in the data matrix. (Tree length = 89 steps)

crepancies between the two considered trees for characters 6, 17, 45, 46, 49 and 53 are of type (1), those for characters 5, 8, 11, 13, 15, *21*, 29 and 30 are of type (2). Both type (1) and type (2) differences are found in character 54.

Contrary to Huys & Boxshall's (1991: 387) designation, in conducting character analysis, the missing state does not necessarily denote the 'absence' of that character. Instead, it indicates that the state of a character is 'unknown' for the taxon (or taxa). Therefore, the interpretation of '9' (missing data) either as '0' (plesiomorphy) or '1' (apomorphy) in a phylogenetic hypothesis is beyond dispute. Essentially, either way is acceptable until it is disproved. For instance, character 17 refers to the segmentation of IX and X on the exopod of the antenna. The missing state is interpreted as 'fused' for the Poecilostomatoida, Monstrilloida, Cyclopoida and Siphonostomatoida in HB-Tree but as 'free' for the same four orders in Tree C. However, as far as our current knowledge goes, the antennal exopod is lacking in the first two orders and represented by a lobe (or single sgment) in the other two orders. We know nothing about the exopodal segments IX and X on the antenna of the copepods of these four orders. Therefore, under this circumstance, it is impossible to state with certainty that the segments IX and X are fused or separated in these four orders of Copepoda. In other words, different assumptions of the missing states for characters 6, 17, 45, 46, 49 and 53 are not in conflict with Huys & Boxshall's (1991) understanding of character evolution in the copepods.

Huys & Boxshall (1991) divided the ten orders of Copepoda into two infraclasses: Progymnoplea and Neocopepoda. The former accommodates only one order (Platycopioida). The Neocopepoda was further divided into two superorders: Gymnoplea and Podoplea, with the former consisting of only one order, the Calanoida. This scheme of classification is supported by the phylogeny implied not only in HB-Tree but also in Tree C. Thus, the major discrepancy between these two phylogenetic hypotheses lies within the Podoplea.

The chief topological difference between Figs. 3 and 4 lies in the subdivision of the superorder

Table 3. Clades (branches) in HB-Tree (Fig. 4) and Tree C (Fig. 3) showing convergence.

Character	in HB-Tree		in Tree C	
5	17–18	13–14	–	–
8	17–18	13–14	–	–
11	17-MORMO 15–16	13–14	13–17	15–16
12	20-PLATY	19–18	20-PLATY	18–19
13	19-CALAN	17-MORMO	19-CALAN	13–17
15–16	15–16			
15	–	–	18-MISOP	13–17
17	14–18	16–17	18-MISOP 13-GELYE	16-HARPA
18	20-PLATY 14-MISOP	18–17 13-GELYE	20-PLATY 16–17	18-MISOP 13-GELYE
‘*21*’	17-MORMO 13–14	15–16	13–17	15–16
29	20-PLATY 15–16	17-MORMO 13–14	20-PLATY 15–16	13–17
30	20-PLATY	16–17	20-PLATY	13–17
13–14	12-SIPHO	16-HARPA		
39	16-HARPA 12-MONST	15-POECI	16-HARPA 12-MONST	14-POECI
46	20-PLATY 14-MISOP	16–17	20-PLATY 16–17	18-MISOP
53	20-PLATY 14-MISOP	16–17	20-PLATY 16–17	18-MISOP
54	14–18	15–16	18-MISOP 14–15	13-CYCLO
Total No. of Convergences	39		37	

Podoplea: with Misophrioida occurring as the sister-group of Gelyelloida-Cyclopoida in HB-Tree but as the first diverged order of the Podoplea in Tree C. Examination reveals that the cause of such difference is due to the varied treatment of nine characters (5, 8, 11, 13, 15, ‘*21*’, 29, 30 and 54), convergence in one tree and synapomorphy in the other. For instance, characters 5 and 8 (referring to the fusion of female and male antennule of segment XXIV to its distal adjacent segment) are treated as a convergent event in HB-Tree. In Tree C these same characters are considered shared novelties among all orders of Podoplea except Misophrioida. Character 15 (presence of a male antennular sheath) shows just the opposite: synapomorphy in HB-Tree and convergence in Tree C. Which is correct for these three characters, synapomorphy or convergence? With our current knowledge of copepods, there is no answer to this question. We do not know how a genome controls the formation of the male antennular sheaths in Copepoda. However, there is one thing we can say with certainty: whether synapomorphy or convergence, it does not confilict with the hypotheses of character polarity proposed by Huys & Boxshall (1991: 394) in their data matrix.

Different treatment of characters 11, 13, *21*, 29 and 30 between the two hypotheses results in another topological change. A treatment of these as the result of convergent events places the Mormonilloida as an immediate outgroup of the Harpacticoida (as in HB-Tree), whereas treating them as the result of a shared novel evolution reverses the relationship of these two orders (as in Tree C). Again, with our current knowledge of copepods, either treatment is acceptable and

Table 4. Comparison of convergent character states with distributions differing between the best fit tree of PAUP analysis (Tree C) and the tree proposed by Huy & Boxshall in 1991 (HB-Tree).

Character	HB-Tree	Tree C
5 & 8	Convergence between (CYCLO+GELYE) and (MORMO + HARPA POECI + SIPHO = MONST)	Synapomorphy of CYCLO + GELYE + HARPA + MORMO + POECI + SIPHO + MONST
6, 45, 46, 49 & 53	Assuming 9 = 0 in MORMO	Assuming 9 = 1 in MORMO
11, 13, *21* & 29	Convergence between MORMO and (MONST + SIPHO = POECI)	Synapomorphy of MORMO + MONST + SIPHO = POECI
15	Synapomorphy of MISOP + CYCLO + GELYE	Convergence between (GELYE + CYCLO) and MISOP
17	Assuming 9 in POECI, SIPHO, MONST and CYCLO = 1	Assuming 9 in POECI, SIPHO, MONST and CYCLO = 0
30	Convergence between (CYCLO + GELYE) and (HARPA + POECE + SIPHO + MONST)	Convergence between CYCLO + GELYE), HARPA and (POECE + SIPHO + MONST)
54	Synapomorphy of MISOP + CYCLO GELYE, assuming 9 in GELYE = 1	Convergence between MISOP and CYCLO, assuming 9 in GELYE = 0

Definition of characters in the table:
5 – female antennulary segment of XXVI fused to terminal double segments
8 – male antennulary segment of XXVI fused to terminal double segments
11 – male antennulary segment of XXIII fused to XXI–XXII
13 – male antennulary segments XXIV and XXV fused
15 – sheath on antennulary segment of XX enclosing segment XXI in male
21 – basis of maxillae without seta
29 – praecoxal seta on maxilliped lost
30 – basis of maxilliped with 2 setae
54 – male fifth leg with seta b on exopod

does not conflict with the concept of character polarity stressed by Huys & Boxshall (1991).

Tree length as the criterion for hypothesis selection

Cladistic methodology consists of two steps. The first step involves an evaluation of characters and an analysis of the direction of character state transformation (i.e., polarization). It is in conducting this part of a cladistic analysis that the concept of homology is employed. Theoretically, in the second step of cladistic analysis, the construction of cladograms, all shared derived character states (synapomorphies) should fall into the same groups of taxa. In practice, however, this does not always occur (Ridley, 1986). There are three reasons for this:

(1) mistakes in judging homology,
(2) mistakes in polarizing charactr states (such as ancestral character states being mistaken for the derived ones), and
(3) the presence of unrecognized convergence.

These errors are often inevitable due to the inherent difficulty in identifying the true homnologous characters and character states. The existence of such errors in phylogenetic analysis has prompted some phylogeneticists to devise means for evaluating obtained cladograms. The most popular method is an application of the principle of parsimony to seek the cladogram that minimizes homoplasy (Felsenstein, 1983). In other words, based on the principle of parsimony, the best fit cladogram is the one requiring the fewest evolutionary events (steps) on the depicted phylogeny (cladogram).

If the phylogenetic hypothesis is free of reversals, the evolution of character states depicted in the tree would not be in direct conflict with *a priori* assumptions about character state transformation set forth in the matrix used for obtaining the cladogram. Here, therefore, the next question becomes which hypothesis we should selected for Copepoda based on the concept of polarity expressed in Table 1: HB-Tree (proposed by

Huys & Boxshall, 1991) or Tree C (rejected by Huys & Boxshall, 1991)? This is where the 'Principle of Parsimony' comes into play. Tree C (87 steps) should be selected, because after removal of all reversals it is two steps shorter than the HB-Tree (89 steps).

Parsimony is a good methodological criterion for evaluating the fitness of a proposed phylogeny. For example, two other competing phylogenetic hypotheses are currently available for orders of Copepoda (see Ho, 1990 and Stock, 1991). We can use the procedure described above to reconstruct these two trees and then fit them to be free of conflicts with Huys & Boxshall's (1991) assumptions about character state transformation. Some statistics are listed in Table 2 for these two additional hypotheses based on the 54 sets of characters provided by Huys & Boxshall (1991). Neither of these two trees, however, is competitive with Tree C, because the refitted Ho-Tree has 112 steps and the Stock-Tree has 136 steps.

Conclusion

To recognize a derived character state (apomorphy) as a result of common ancestry (synapomorphy) or independent evolution (convergence) is a classical problem in the study of phylogeny. The cladistic analysis was formulated to minimize this problem by dividing the study process into two procedures:

(1) character analysis (including character evaluation and polarization) based on 'homology' and

(2) cladogram construction based on 'parsimony'.

The two procedures are independent and are not necessarily contradictory. If Huys & Boxshall's (1991) hypotheses concerning character homology and character state polarity are correct, the best fit phylogenetic hypothesis for the ten orders of the Copepoda is the one depicted in Fig. 3 (Tree C).

Acknowledgements

I am indebted to Frank D. Ferrari, Daniel R. Brooks and three anonymous reviewers for providing constructive comments and valuable suggestions for improving this paper.

References

Felsenstein, J., 1983. Parsimony in systematics. Biological and statistical issues. Annu. Rev. Ecol. Syst. 14: 313–333.

Ho. J. S., 1990. A phylogenetic analysis of copepod orders. J. crust. Biol. 10: 528–536.

Huys, R. & G. A. Boxshall, 1991. Copepod Evolution. The Ray Society (London), 468 pp.

Ridley, M., 1986. Evolution and Classification, the reformation of cladism. Longman, London & New York, 201 pp.

Stock, J. H., 1991. Some reflections on the antiquity of the copepod lineages. Bull. Plankton Soc. Japan, Special Volume: 1–7.

Swofford, D. L., 1985. PAUP (Phylogenetic Analysis Using Parsimony). Documentation for version 2.4.1. Illinois Natural History Survey, Champaign, 81 pp.