3

SYSTEMATIC PATTERNS: COMPONENT ANALYSIS

CLADOGRAMS, TREES, AND COMPONENTS

It has been stated above that the cladogram contains information sufficient for specifying groups and subgroups to be recognized as taxa in classifications, and that the information resides in the structure of the cladogram. The unit of information of this kind may be called a component, corresponding to a statement of general synapomorphy. Given two species, A and B, for example, a component in the form "AB" is a statement of general synapomorphy: that species A and B share apomorphy, and are therefore related. What follows is an analysis of the possible components of cladograms and other branching structures—what may be called "component analysis." For convenience, the term "taxon" is used to indicate an entity that may be a single species, or a group of species.

COMPONENT ANALYSIS OF TWO TAXA

The sense of a branching, or dendritic, diagram *seems* immediately obvious. The sense of the alternative diagrams of figure 3.1, for example, could be said to be that taxon A is ancestral to B, or that taxa A and B are separated by some unspecified phenetic or gradistic distance. By way of introduction, the phyletic viewpoint is adopted.

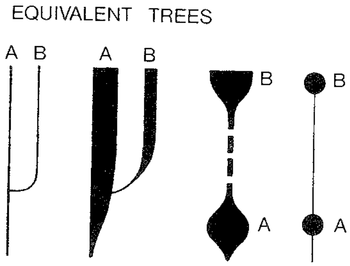


Figure 3.1. Branching diagrams of various styles, showing taxon A as ancestral to taxon B.

Consider an analysis of two taxa (figure 3.2). There are only three ways that two taxa (A and B) might be phyletically interrelated: (1) taxon A is ancestral to B (tree 3.2.1); (2) taxon B is ancestral to A (tree 3.2.2); (3) an unknown taxon X is ancestral to both A and B (tree 3.2.3). But there is another way of stating these three possibilities, and that way consists of what may be called "component analysis." Tree 3.2.1, for example, has two components: (1) taxa A and B are related by common ancestry (the general component); (2) the common ancestor [of A and B] is A (the unique component). Similarly, tree 3.2.2 has two components: (1) taxa A and B are related by common ancestry (general); (2) the common ancestor is B (unique). And finally, tree 3.2.3 has two components: (1) taxa A and B are related by common ancestry (general);

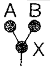
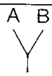
ANALYSIS OF TWO TAXA			
TREES	COMPONENTS		GROUPING
	GENERAL	UNIQUE	
1  B A	AB	A	AB 1
2  A B	AB	B	AB 1
3  A B X	AB	X	AB 1
CLADOGRAM			
1  A B	AB		AB 1

Figure 3.2. Component analysis of two taxa, A and B, showing the three possible trees, one possible cladogram, and one possible grouping (classification).

(2) the common ancestor is neither A nor B but an unknown taxon X (unique).

The general component is called "general" because it occurs as a component of all three trees (3.2.1–3); the unique components, in contrast, are unique to each tree ("A" to tree 3.2.1; "B" to tree 3.2.2; "X" to tree 3.2.3). The general component ("AB") defines, and is the sole component of, a dendritic structure that in this context may be called a "phylogram" (cladogram 3.2.1). The phylogram is defined by the general component: "taxa A and B are related by common ancestry," which is the total information content of the phylogram. This component, however, has itself two parts, which may be called "cladistic" and "phyletic," respectively. The cladistic part is "Taxa A and B are related"; and the phyletic part is "by common ancestry." If the phyletic part is omitted, only the cladistic part remains: "Taxa A and B are related" (by some unspecified relation). In fact, figure 3.2 does not specify the nature of the relation between taxa A and B. For that reason, the trees in themselves are not phyletic trees; they are simply trees. And the cladogram in itself is not a phylogram; it is simply a cladogram.

A cladogram, therefore, may be defined as a branching, or dendritic, structure, or dendrogram, illustrating an unspecified relation (general synapomorphy) between certain specified terms that in the context of systematics represent taxa. If the relation is considered common ancestry, there is reason to call the structure a phylogram; or, alternatively, if the relation is considered phenetic or gradistic similarity (however measured), there is reason to call the structure a phenogram or gradogram, respectively. In either case, the structure itself need not change—only the interpretation of its significance.

Because it is defined by a general component, a cladogram denotes a set of trees. A cladogram, therefore, is not a tree. A tree may be defined as a branching, or dendritic, structure, or dendrogram, having one or more *general* as well as *unique* components (or combination of components). A cladogram, in contrast, is a dendritic structure having only one or more general components. Conceived as a set, a cladogram does not resemble any one of its member-trees more than another: cladogram 3.2.1 does not resemble tree 3.2.3 more than trees 3.2.1 and 3.2.2. It resembles all of its member-trees equally. And the resemblance consists only of the general component shared by all member-trees and the cladogram. A cladogram, therefore, is not *a* tree; it is a *set* of trees.

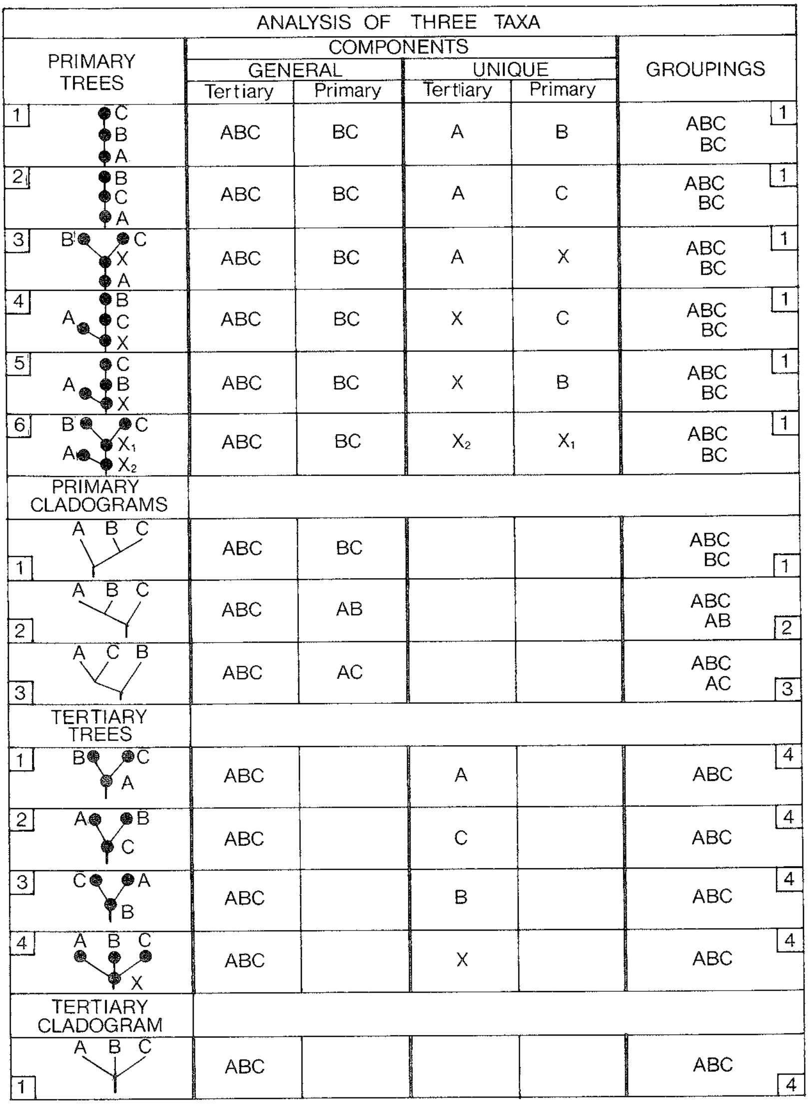
A cladogram is a branching structure joining certain terms (representing taxa) that are related by some unspecified relation. In itself, a cladogram conveys no sense of phylogeny, common ancestry, phenetic resemblance, gradistic resemblance, ecological resemblance, or any other relation that might conceivably join the terms (representing taxa). Of course, it may seem odd to think of a phenogram, for example, as a type of cladogram. But a phenogram does, after all, have a cladistic (branching) aspect that as *branching* is no different than the branching of a phylogram. And the implication of a phenogram, like that of a phylogram, is that certain groups (represented by the components) exist, an implication that can be investigated independently of any implied phenetic (or phyletic) relations.

There is only one way that two taxa might be grouped (classified): as subtaxa of some inclusive taxon. In other words, the two taxa must be classified together: in figure 3.2 the grouping AB is the only one possible. Not surprisingly, its terms ("AB") are those of the general component. Thus, there is a one-to-one correspondence between the grouping and the cladogram (the general component of the trees). Put another way, the groups (inclusive taxa) of a classification correspond on a one-to-one basis to the components of a cladogram (the general components of the member-trees of the set denoted by the cladogram).

COMPONENT ANALYSIS OF THREE TAXA

Other properties of cladograms may be considered in relation to a component analysis of three taxa (fig. 3.3). For three taxa, there are four possible cladograms (primary, or fully resolved, cladograms 1–3, and tertiary, or unresolved, cladogram 1) and, naturally, four corresponding classifications (groupings 1-4). The four cladograms, and the groupings, differ among themselves in their components. The cladograms comprise what may be called three "primary cladograms" and one "tertiary cladogram" (and the corresponding groups comprise what might be called three "primary classifications" [groupings 1-3] and one "tertiary classification" [grouping 4]). Each primary cladogram is a set of six "primary trees," and the tertiary cladogram is a set of four "tertiary trees."

The component analysis of a primary tree (in a phyletic context) is as follows. For tree 3.3.1 there are two general components: (1) taxa A, B, and C are related by common ancestry (a "tertiary component"); (2) taxa



ANALYSIS OF THREE TAXA					
PRIMARY TREES	COMPONENTS				GROUPINGS
	GENERAL		UNIQUE		
	Tertiary	Primary	Tertiary	Primary	
1 C B A	ABC	BC	A	B	ABC BC 1
2 B C A	ABC	BC	A	C	ABC BC 1
3 B C X A	ABC	BC	A	X	ABC BC 1
4 B A C X	ABC	BC	X	C	ABC BC 1
5 C A B X	ABC	BC	X	B	ABC BC 1
6 B C X_1 A X_2	ABC	BC	X_2	X_1	ABC BC 1
PRIMARY CLADOGRAMS					
1 A B C	ABC	BC			ABC BC 1
2 A B C	ABC	AB			ABC AB 2
3 A C B	ABC	AC			ABC AC 3
TERTIARY TREES					
1 B C A	ABC		A		ABC 4
2 A B C	ABC		C		ABC 4
3 C A B	ABC		B		ABC 4
4 A B C X	ABC		X		ABC 4
TERTIARY CLADOGRAM					
1 A B C	ABC				ABC 4

Figure 3.3. Component analysis of three taxa, A, B, C.

B and C are related by common ancestry (a "primary component"). There are two unique components: (1) the ancestor [of ABC] is A ("tertiary component"); (2) the ancestor [of BC] is B ("primary component"). The six primary trees (3.3.1–6) form a set denoted by primary cladogram 3.3.1. There is a similar set of six different trees denoted by primary cladogram 3.3.2, and another set of six different trees denoted by primary cladogram 3.3.3.

Each cladogram corresponds to a classification (grouping). Cladogram 3.3.1 (and trees 3.3.1–6), for example, corresponds to grouping 3.3.1, according to which there is an inclusive taxon (ABC), including two subtaxa: (1) including only A (implied and not listed); and (2) including BC.

There are in addition four "tertiary trees" (3.3.1–4), each having only one general component (ABC) and one unique component (A, B, C, or X). These form a set denoted by the tertiary cladogram, of which the corresponding classification (grouping 3.3.4) is simply an inclusive taxon (ABC) subdivided (by implication) into three noninclusive subtaxa (A, B, and C).

In the three-taxon analysis, both components and trees are sorted into primary and tertiary types. Each primary tree has a maximum number of general components (the maximum number = the number of terms [representing taxa] minus one, or N–1). The more inclusive component is the tertiary; and the less inclusive, the primary. For each general tree-component, there is a corresponding unique tree-component. For primary tree 3.3.1, for example, the general tertiary component (ABC) corresponds to the unique tertiary component (A); and the general primary component (BC), to the unique primary component (B).

The unique components of primary tree 3.3.6 (X_1, X_2) are understood to be different ($X_1 \neq X_2$). If they are the same ($X_1 = X_2$), the tree is a tertiary tree (tertiary tree 3.3.4), defined by only one general tertiary component (ABC) and one unique tertiary component (X).

It is true of the three-taxon analysis, as for the two-taxon analysis, that there is a one-to-one correspondence between classifications (groupings) and cladograms (sets of trees).

COMPONENT ANALYSIS OF FOUR TAXA

An analysis of four taxa (figures 3.4–3.9) is almost excessively complex. There are twelve sets (primary cladograms 3.4.1, 3.5.1–12), each of

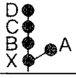
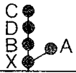
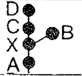
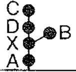
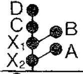
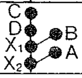
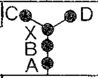
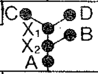
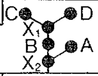
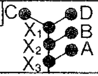
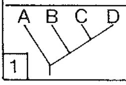
ANALYSIS OF FOUR TAXA							
PRIMARY TREES	COMPONENTS						GROUPINGS
	GENERAL			UNIQUE			
	TERTIARY	SECONDARY	PRIMARY	TERTIARY	SECONDARY	PRIMARY	
1 	ABCD	BCD	CD	A	B	C	ABCD BCD CD 1
2 	ABCD	BCD	CD	A	B	D	ABCD BCD CD 1
3 	ABCD	BCD	CD	X	B	C	ABCD BCD CD 1
4 	ABCD	BCD	CD	X	B	D	ABCD BCD CD 1
5 	ABCD	BCD	CD	A	X	C	ABCD BCD CD 1
6 	ABCD	BCD	CD	A	X	D	ABCD BCD CD 1
7 	ABCD	BCD	CD	X_2	X_1	C	ABCD BCD CD 1
8 	ABCD	BCD	CD	X_2	X_1	D	ABCD BCD CD 1
9 	ABCD	BCD	CD	A	B	X	ABCD BCD CD 1
10 	ABCD	BCD	CD	A	X_2	X_1	ABCD BCD CD 1
11 	ABCD	BCD	CD	X_2	B	X_1	ABCD BCD CD 1
12 	ABCD	BCD	CD	X_3	X_2	X_1	ABCD BCD CD 1
PRIMARY CLADOGRAM							
 1	ABCD	BCD	CD				ABCD BCD CD 1

Figure 3.4. Component analysis of four taxa, A, B, C, D.

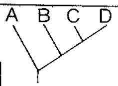
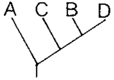
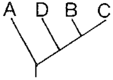
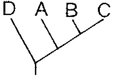
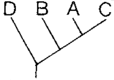
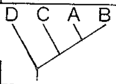
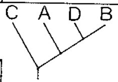
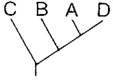
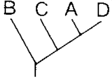
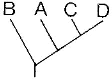
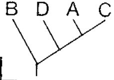
ANALYSIS OF FOUR TAXA (CONT'D)				
PRIMARY CLADOGRAMS	GENERAL COMPONENTS			GROUPINGS
	Tertiary	Secondary	Primary	
1 	ABCD	BCD	CD	ABCD BCD CD 1
2 	ABCD	BCD	BD	ABCD BCD BD 2
3 	ABCD	BCD	BC	ABCD BCD BC 3
4 	ABCD	ABC	BC	ABCD ABC BC 4
5 	ABCD	ABC	AC	ABCD ABC AC 5
6 	ABCD	ABC	AB	ABCD ABC AB 6
7 	ABCD	ABD	AB	ABCD ABD AB 7
8 	ABCD	ABD	BD	ABCD ABD AB 8
9 	ABCD	ABD	AD	ABCD ABD AD 9
10 	ABCD	ACD	AD	ABCD ACD AD 10
11 	ABCD	ACD	CD	ABCD ACD CD 11
12 	ABCD	ACD	AC	ABCD ACD AC 12

Figure 3.5. Component analysis of four taxa (cont'd).

twelve primary trees (3.4.1–12); and in addition, another three sets (primary cladograms 3.6.13–15), each of nine primary trees (3.6.145–153). There are four sets (secondary, or partially resolved, cladograms 3.7.1–4), each of eight secondary trees (3.7.1–8); and in addition, another six sets (secondary cladograms 3.8.5–10), each of nine secondary trees (3.8.33–41). Finally, there is a set (tertiary cladogram 3.9.1) of five tertiary trees (3.9.1–5). The second suite of secondary cladograms (3.8.5–10) are interpreted here (trees 3.8.33–41) in their most restrictive sense; a less restrictive interpretation will be discussed below.

Some aspects of the above analyses are summarized in figure 3.10. Apparent is a one-to-one correspondence between classifications (groupings) and cladograms. Because of the large number of trees, there is no correspondence between them and classifications, except by way of their general components, which form sets (cladograms).

Comparison of the members of a given set of trees shows diversity that in a phyletic context is the cause of controversy, e.g., over the nature of ancestors in general, and the identity of particular ancestors. It may be kept in mind, however, that the cladogram (or phylogram) resembles equally all of the member-trees of its set, as defined by its general components. Of some importance, then, is that cladistic analysis operates at the level of generality of cladograms, not of trees. For many purposes, such as classification, cladistic analysis renders superfluous the analytical morass that often engulfs efforts to resolve trees.

Within the phyletic context, cladistic analysis has great appeal because of its generality. However, its generality is sometimes misunderstood, and cladistic analysis is sometimes confused with tree analysis (and cladograms, such as primary cladogram 3.4.1, are sometimes confused with trees having unknown unique components, such as primary tree 3.4.12). Trees and cladograms are not incompatible; neither is tree analysis incompatible with cladistic analysis. The point is simply that cladograms and cladistic analysis have a generality greater than that of trees and tree analysis. And cladograms and classification stand in a one-to-one relation which trees and classification do not share.

It has been shown that a given cladogram denotes a set of trees. In addition, some "cladograms" denote sets of other cladograms (quotes are used to indicate this quality of some cladograms). Consider a four-

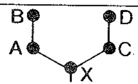
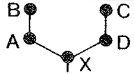
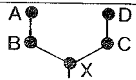
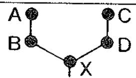
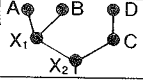
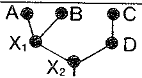
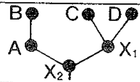
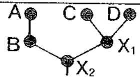
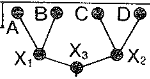
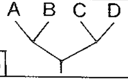
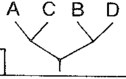
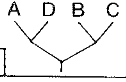
ANALYSIS OF FOUR TAXA (CONT'D)					
PRIMARY TREES	COMPONENTS				GROUPING
	GENERAL		UNIQUE		
	Tertiary	Primary	Tertiary	Primary	
145 	ABCD	AB CD	X	A C	ABCD AB CD 13
146 	ABCD	AB CD	X	A D	ABCD AB CD 13
147 	ABCD	AB CD	X	B C	ABCD AB CD 13
148 	ABCD	AB CD	X	B D	ABCD AB CD 13
149 	ABCD	AB CD	X_2	X_1 C	ABCD AB CD 13
150 	ABCD	AB CD	X_2	X_1 D	ABCD AB CD 13
151 	ABCD	AB CD	X_2	A X_1	ABCD AB CD 13
152 	ABCD	AB CD	X_2	B X_1	ABCD AB CD 13
153 	ABCD	AB CD	X_3	X_1 X_2	ABCD AB CD 13
PRIMARY CLADOGRAMS					
 13	ABCD	AB CD			ABCD AB CD 13
 14	ABCD	AC BD			ABCD AC BD 14
 15	ABCD	AD BC			ABCD AD BC 15

Figure 3.6. Component analysis of four taxa (cont'd).

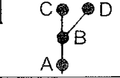
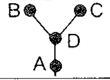
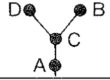
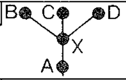
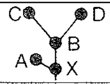
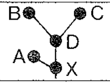
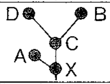
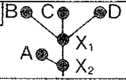
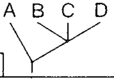
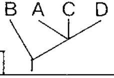
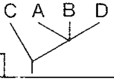
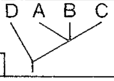
ANALYSIS OF FOUR TAXA (CONT'D)							
SECONDARY TREES	COMPONENTS						GROUPINGS
	GENERAL			UNIQUE			
	TERTIARY	SECONDARY	PRIMARY	TERTIARY	SECONDARY	PRIMARY	
1 	ABCD	BCD		A	B		1 ABCD BCD
2 	ABCD	BCD		A	D		1 ABCD BCD
3 	ABCD	BCD		A	C		1 ABCD BCD
4 	ABCD	BCD		A	X		1 ABCD BCD
5 	ABCD	BCD		X	B		1 ABCD BCD
6 	ABCD	BCD		X	D		1 ABCD BCD
7 	ABCD	BCD		X	C		1 ABCD BCD
8 	ABCD	BCD		X_2	X_1		1 ABCD BCD
SECONDARY CLADOGRAMS							
 1	ABCD	BCD					ABCD BCD 1
 2	ABCD	ACD					ABCD ACD 2
 3	ABCD	ABD					ABCD ABD 3
 4	ABCD	ABC					ABCD ABC 4

Figure 3.7. Component analysis of four taxa (cont'd).

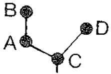
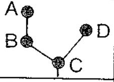
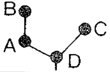
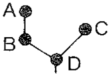
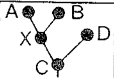
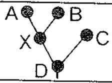
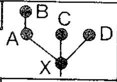
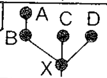
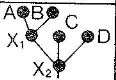
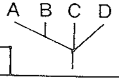
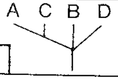
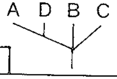
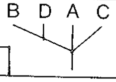
ANALYSIS OF FOUR TAXA (CONT'D)							
SECONDARY TREES	COMPONENTS						GROUPING
	GENERAL			UNIQUE			
	TERTIARY	SECONDARY	PRIMARY	TERTIARY	SECONDARY	PRIMARY	
33 	ABCD		AB	C		A	5 ABCD AB
34 	ABCD		AB	C		B	5 ABCD AB
35 	ABCD		AB	D		A	5 ABCD AB
36 	ABCD		AB	D		B	5 ABCD AB
37 	ABCD		AB	C		X	5 ABCD AB
38 	ABCD		AB	D		X	5 ABCD AB
39 	ABCD		AB	X		A	5 ABCD AB
40 	ABCD		AB	X		B	5 ABCD AB
41 	ABCD		AB	X_2		X_1	5 ABCD AB
SECONDARY CLADOGRAMS							
 5	ABCD		AB				ABCD AB 5
 6	ABCD		AC				ABCD AC 6
 7	ABCD		AD				ABCD AD 7
 8	ABCD		BC				ABCD BC 8
 9	ABCD		BD				ABCD BD 9
 10	ABCD		CD				ABCD CD 10

Figure 3.8. Component analysis of four taxa (cont'd).

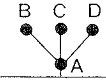
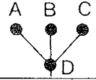
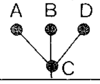
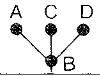
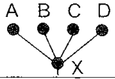
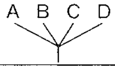
ANALYSIS OF FOUR TAXA							
TERTIARY TREES	COMPONENTS						GROUPING
	GENERAL			UNIQUE			
	Tertiary	Secondary	Primary	Tertiary	Secondary	Primary	
1 	ABCD			A			1 ABCD
2 	ABCD			D			1 ABCD
3 	ABCD			C			1 ABCD
4 	ABCD			B			1 ABCD
5 	ABCD			X			1 ABCD
TERTIARY CLADOGRAM							
 1	ABCD						ABCD 1

Figure 3.9. Component analysis of four taxa (concluded).

taxon tertiary cladogram (3.9.1). It is defined by the tertiary component ABCD, and it denotes a set of five tertiary trees (3.9.1–5). It does so because the tertiary component ABCD is the general component of all five tertiary trees. Yet the tertiary component ABCD is a general component, also, of all secondary cladograms (3.7.1–4, 3.8.5–10), secondary trees (3.7.1–8, 3.8.33–41), primary cladograms (3.4.1, 3.5.1–12, 3.6.13–15), and primary trees (3.4.1–12, 3.6.145–153). The four-taxon tertiary "cladogram" (cladogram 3.9.1), therefore, denotes a set of all other four-taxon cladograms and four-taxon trees that contain the same tertiary component (ABCD).

The subsets of a four-taxon "cladogram" are shown in figure 3.11. The set of all four-taxon cladograms and trees is called a "structure set." The set's notation is a tertiary "structure" (a tertiary "cladogram" with no terms, such as A, B, etc.), of which the node is a circle. The circle denotes a set of five structures: one tertiary (unresolved), two secondary (partially resolved), and two primary (fully resolved). Each of the structures is defined by a "clade type." Of the two secondary structures,

TAXA	TREES				CLADOGRAMS				GROUPINGS
	Primary	Secondary	Tertiary	Total	Primary	Secondary	Tertiary	Total	
2	3			3	1			1	1
3	18		4	22	3		1	4	4
4	171	86	5	262	15	10	1	26	26

Figure 3.10. Summary of component analyses of two, three, and four taxa.

(1) is defined by a tertiary structure to which a primary structure has been added, whereas (2) is defined by a tertiary structure to which a secondary structure has been added. Of the two primary structures, (1) is defined by a tertiary structure to which two different primary structures have been added, whereas (2) is defined by a tertiary structure to which both a secondary and a primary structure have been added. Each of the five structures denotes a set of one or more cladograms, each of which denotes a set of trees. The first of the two secondary structures (defined by a tertiary plus primary clade type) is here again treated in its most restrictive sense.

That a given tertiary or secondary "cladogram" can be a set of other cladograms means that the corresponding classification can be a set of classifications. Interesting as this may be in itself, there are implications

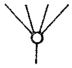
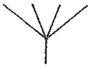
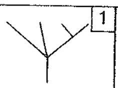
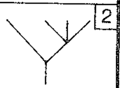
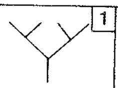
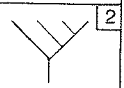
	STRUCTURE SET 				
	TERTIARY	SECONDARY		PRIMARY	
STRUCTURES (sets of cladograms)					
CLADE TYPES	T	T+P	T+S	$T+P_1+P_2$	T+S+P
CLADOGRAMS (all possible)	1	6	4	3	12
TREES (all possible)	5	54	32	27	144

Figure 3.11. Set-logic of 4-taxon analysis.

pertaining to information content. Any given cladogram or classification has an actual information content, specified by its components. If it is a nonprimary cladogram (or classification), it has in addition a potential information content, potentially specifiable by one or more components of its member cladograms (the member cladograms of the set it denotes). For example, tertiary cladogram 3.9.1 (four taxa) is defined only by one component (ABCD). Potentially, it includes two other components (a secondary and a primary, or two primary components), among the various secondary and primary components of cladograms 3.7.1–4, 3.8.5–10, 3.5.1–12, and 3.6.13–15.

That cladograms (and classifications) have an actual and a potential information content means that they can (1) represent the current state of knowledge of relationships at any given time; and (2) evolve in relation to increased knowledge. Put another way, cladograms and classifications evolve as actual information content increases (and potential information content diminishes).

CLADOGRAMS AND INFORMATION

Comparative studies of organisms result in the accumulation of information. Summaries of information in the form of cladograms pose certain problems, stemming from a basic question: how is information integrated into a coherent summary? The simplest case involves two species, let us say a lamprey and a shark (figure 3.12). Certain features may be found in one species and not the other, in both species, or in neither species: a lamprey, for example, has an elongate body (character 1), a single nostril (character 2), a notochord (character 3), etc.; a shark has a more robust body (character 4), two nostrils (character 5), a vertebral column (character 6), etc.; both species have eyes (character 7), cranial nerves (character 8), kidneys (character 9), etc.; neither species has lungs (character 10), a spoken language (character 11), and a bony skeleton (character 12). This information may be summarized in the form of a table (table 3.1). It is easy to see that the information may be integrated into a general form (table 3.2). The corresponding cladogram is figure 3.12, which poses no problems in this case, because it includes only two species (for two species there is only one possible cladogram).

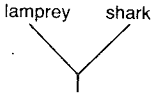


Figure 3.12. A cladogram for two species.

The cladogram does pose a problem, however, in its exact relations with the information of table 3.2: how is the information (character–types) to be understood as represented in the cladogram? Consider character-type A: where is it represented in the cladogram? There are three possibilities: (1) the line extending to the lamprey (figure 3.13.1); (2) the line extending to the shark (figure 3.13.2); (3) the line extending to both species (figure 3.13.3).

Possibility 3.13.3 may be seen immediately to be unsatisfactory, for it implies something true for both species, whereas character-type A asserts that the species differ. Possibility 3.13.2 may also seem unsatisfactory, but the reason is not so obvious. For example, consider the statement, that it is true that the shark does not have the characters found only in the lamprey: namely, an elongate body (character 1), a single nostril (character 2), a notochord (character 3), etc. The statement is true for what it denies to be true for the shark. Possibility 3.13.1 seems intuitively preferable, because it is simple and direct, without the complicated phrasing of possibility 3.13.2, namely that "it is true that the shark does not have. . . ." On intuitive grounds, therefore, one might regard the information (character–types) to be represented in the cladogram as specified in figure 3.14.1, which omits the negative occurrences of character-types. Adding the negative occurrences, however, does not change the nature of the cladogram (figure 3.14.2).

An example with three species poses an additional problem, because there are four possible cladograms. As a third species may be added the lancelet with certain features unique to it, solenocytes (character 13), atriopore (character 14), endostyle (character 15); and certain features

Table 3.1. Characters and Their Occurrence in the Lamprey and Shark

	Characters											
Species	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Lamprey	+	+	+	–	–	–	+	+	+	–	–	–
Shark	–	–	–	+	+	+	+	+	+	–	–	–

Table 3.2. Character-Types and Their Occurrence in the Lamprey and Shark (cf. table 3.1)

	Character-Types			
Species	*A*	*B*	*C*	*D*
Lamprey	+	–	+	–
Shark	–	+	+	–

common to all three species, gill slits (character 16), bilateral symmetry (character 17), and a dorsal nerve tube (character 18). The information may be organized in a general form (table 3.3). The four possible cladograms may be immediately assessed with reference to the positive, rather than the negative occurrences of the character-types (figure 3.15), as represented by single lines. Cladogram 3.15.1 includes all character-types, whereas cladograms 3.15.2–4 do not include character-type D; and cladograms 3.15.2–3 include a line for a character-type (?) unrepresented in the information (table 3.3). On this basis, cladogram 3.15.1 may be accepted as a true summary of the information (table 3.3), and the other cladograms may be rejected.

The cladograms may alternatively be compared with reference only to the relevant positive occurrences, namely character-type D, which distinguish one cladogram (3.15.5) from the others (3.15.6–8). On this basis, too, cladogram 3.15.5 may be accepted as a true summary of the relevant information (table 3.3), and the other cladograms may be rejected.

A full analysis of the four possibilities, with reference both to positive

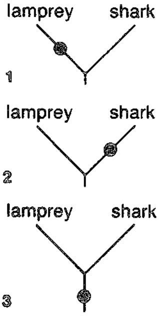


Figure 3.13. Three possibilities for representation of information in a cladogram for two species.

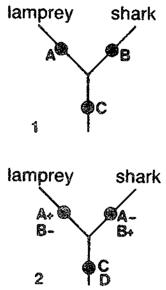


Figure 3.14. Representation of character-types (A, B, C) in a cladogram for two species. 1: Positive occurrences. 2: Positive and negative occurrences (cf. table 3.2).

and negative occurrences of character-types, is given in figure 3.16. As before, only one cladogram (3.16.1) includes all positive occurrences of character-types, as represented by single lines; the other cladograms (3.16.2–4) do not allow for character-type D+ except as multiple occurrences. Also, cladograms 3.16.2–3 include a line for a character-type represented only by a negative occurrence (B- in 3.16.2; C- in 3.16.3); cladogram 3.16.1 also includes a line representing a negative occurrence (A-) but the line also represents a positive occurrence (D+).

A criterion by which to judge the different possible cladograms is not as obvious in the full analysis (figure 3.16) as in the analysis only of positive occurrences (figure 3.15). But regarded as an integration of information, a cladogram may be judged according to how much information it does in fact integrate, or, in other words, how concentrated a summary it is. In the above full analysis (figure 3.16), all of the original information (table 3.3) is related to each of the four possibilities. One way to judge each possibility is simply to count the number of character occurrences (table 3.4). Cladogram 3.16.1 includes five positive, and seven negative, occurrences; cladograms 3.16.2–3, six

Table 3.3. Character-Types and Their Occurrence in the Lancelet, Lamprey, and Shark (cf. tables 3.1, 3.2)

	Character-Types					
Species	*A*	*B*	*C*	*D*	*E*	*F*
Lancelet	+	–	–	–	+	–
Lamprey	–	+	–	+	+	–
Shark	–	–	+	+	+	–

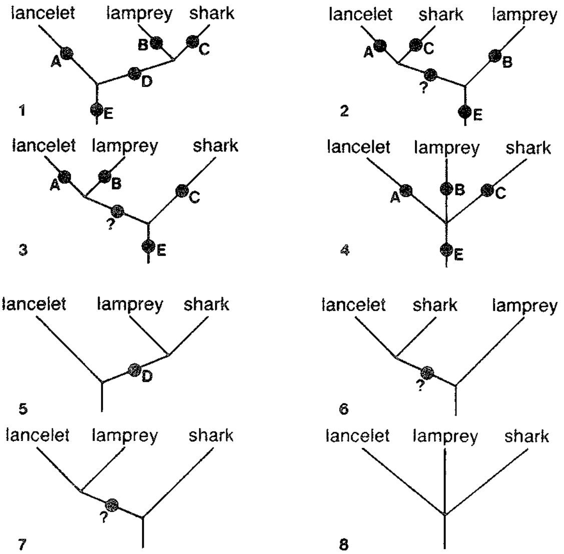


Figure 3.15. Representation of positive occurrences of character-types (A, B, C, D, E) in cladograms for three species. 1–4: All positive occurrences representable by single lines. 5–8: Positive occurrences (D) that differ among the cladograms (cf. table 3.3).

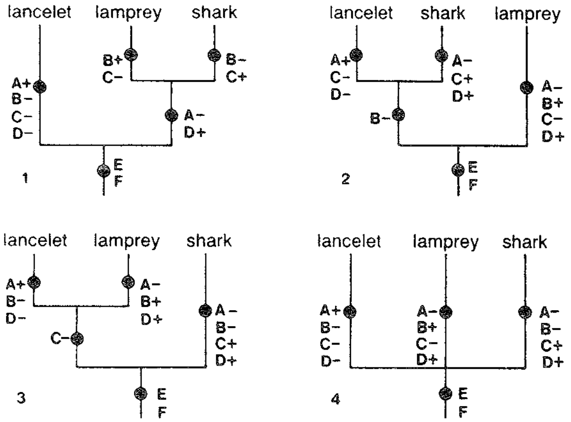


Figure 3.16. Representation of positive (A+, B+, C+, D+, E) and negative (A-, B-, C-, D-, F) occurrences of character-types in cladograms for three species (cf. figure 3.15 and table 3.3).

positive and seven negative; cladogram 3.16.4, six positive and eight negative; the original information (table 3.3) includes eight positive and 10 negative.

The difference in occurrences indicates the degree to which the various cladograms integrate, or generalize, the original information (table 3.5, left). In having the lowest number of occurrences, cladogram 3.16.1 may be judged the most efficient integration. Why such is the case is demonstrated by further analysis. Firstly, the occurrences common to all cladograms (character-types E and F) may be eliminated as irrelevant to the differences among cladograms (table 3.5, center). It is apparent that cladogram 3.16.4 integrates the information no better than table 3.3, and in a manner common also to cladograms 3.16.1–3. Secondly, the occurrences still common to cladograms 3.16.1–3 (character-types A, B, and C, each of which contributes three positive and five negative occurrences among cladograms 3.16.1–3 such that each cladogram receives three positive and five negative occurrences) may be eliminated, as irrelevant to the differences among cladograms 3.16.1–3 (table 3.5, right). It is apparent that the cladograms differ with respect only to character-type D, which appears fewer times in cladogram 3.16.1 than in cladograms 3.16.2–4. The full analysis thus yields the same result as the analysis based only on positive occurrences (figure 3.15), namely that cladogram 3.16.1 (3.15.1) is a true summary of the information (table 3.3), and that the other cladograms may be rejected. Adding the negative occurrences, in short, accomplishes nothing, which is a way of saying that negative occurrences are uninformative with respect to the problem of evaluating the different possible cladograms that might summarize, more or less efficiently, a given sample of information.

But there is more, which derives from the inference to be drawn from a

Table 3.4. Positive (+) and Negative (–) Character-Types and Their Occurrences among Cladograms (cf. figure 3.16 and table 3.3)

Clado-grams	*Positive Occurrences*						*Negative Occurrences*						*Total Occurrences*
	A+	*B+*	*C+*	*D+*	*E+*	*Total*	*A–*	*B–*	*C–*	*D–*	*F–*	*Total*	
3.16.1	1	1	1	1	1	5	1	2	2	1	1	7	12
3.16.2	1	1	1	2	1	6	2	1	2	1	1	7	13
3.16.3	1	1	1	2	1	6	2	2	1	1	1	7	13
3.16.4	1	1	1	2	1	6	2	2	2	1	1	8	14
table 3.3	1	1	1	2	3	8	2	2	2	1	3	10	18

Table 3.5. Positive (+) and Negative (–) Character-Types and Their Occurrences among Cladograms (cf. figure 3.16 and table 3.4)

	All Character-Types			*E and F Eliminated*			*A, B, and C Eliminated*		
Cladograms	+	–	*Total*	+	–	*Total*	+	–	*Total*
3.16.1	5	7	12	4	6	10	1	1	2
3.16.2	6	7	13	5	6	11	2	1	3
3.16.3	6	7	13	5	6	11	2	1	3
3.16.4	6	8	14	5	7	12	2	1	3
table 3.3	8	10	18	5	7	12	2	1	3

cladogram such as 3.16.1—namely that, with cladogram 3.16.1 as a basis for inference, there exists a group including the lamprey and shark and excluding the lancelet. The inference based on cladogram 3.16.2 (that there is a group including the lancelet and shark and excluding the lamprey) and the inference based on cladogram 3.16.3 (that there is a group including the lancelet and lamprey and excluding the shark) conflict with each other, and with the inference based on cladogram 3.16.1. One may anticipate that, of the three inferences, one of them might be true and others false. If cladogram 3.16.1 is accepted as a basis for inference and the others rejected, the acceptance amounts to asserting that cladogram 3.16.1 is true, and the rejection amounts to asserting that cladograms 3.16.2 and 3.16.3 are false. If the decisive factor for acceptance is the occurrence of character-type D+, then the correlated occurrence of character-type A- is coincidental; and the occurrences of character-types B- (cladogram 3.16.2) and C- (cladogram 3.16.3) are also coincidental. In other words, the occurrence of negative characters, if considered informative, is falsely informative in all except one (two of three) of the possible cladograms, if one among them is true.

The above examples are unproblematical in the sense that all positive occurrences are combinable in one cladogram. Real information typically includes positive occurrences that conflict with one another (table 3.6). The four possible cladograms may be compared with

Table 3.6. Character-Types and Their Occurrence in Three Species

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*E*	*F*
1	+	–	–	–	+	+	–
2	–	+	–	+	+	+	–
3	–	–	+	+	–	+	–

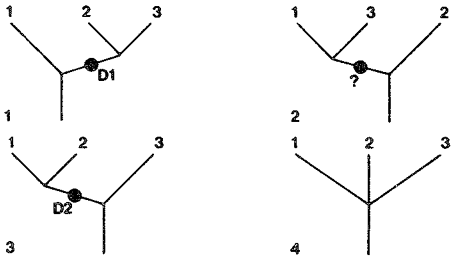


Figure 3.17. Positive occurrences (D1, D2) that differ among cladograms for three species (cf. table 3.6).

reference to the relevant occurrences of the character-types (figure 3.17). Two cladograms (3.17.1 and 3.17.3) integrate the information to the same degree. Although cladogram 3.17.2 might be rejected, no single primary cladogram can be accepted as an efficient summary; the only single cladogram that could be accepted is the tertiary cladogram 3.17.4.

A totally uninformative set of information, which allows for all possible character-types, is presented in table 3.7. Three cladograms integrate the information to the same degree (figure 3.18). No primary cladogram can be rejected; no single primary cladogram can be accepted as an efficient summary; the only single cladogram that could be accepted is tertiary cladogram 3.18.4.

Various procedures have been advocated for the purpose of combining conflicting information into a summary that can be represented in a cladogram. The procedures are sometimes called "clustering procedures" or "clustering algorithms," many of which in recent years have been discussed and compared in the journal *Systematic Zoology*. Clustering procedures vary among themselves and give varying results, any one of which may be used as a basis for inference. That a clustering procedure may be applied to conflicting information, and might yield an unambiguous result in the form of a single cladogram, does not guarantee that the result, or the inference based upon it, is true.

Table 3.7. Character-Types and Their Occurrence in Three Species

	Character-Types							
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*	*F*
1	+	−	−	−	+	+	+	−
2	−	+	−	+	+	−	+	−
3	−	−	+	+	−	+	+	−

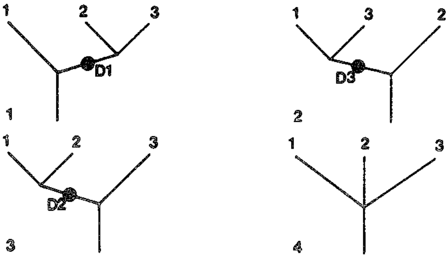


Figure 3.18. Positive occurrences (D1, D2, D3) that differ among cladograms for three species (cf. table 3.7).

Some information with conflicting positive occurrences is contained in table 3.8, and the relevant information is related to the four possible cladograms in figure 3.19 (cladograms 3.19.1–4). A full analysis of positive and negative occurrences (table 3.9) shows that cladogram 3.19.1 is the most efficient summary, and that the decisive factor is the occurrence of character-type D4+ (cf. cladograms 3.19.5–8) which in effect replicates character-type D1+ of the previous example (cf. table 3.7, cladogram 3.18.1). In terms of its efficiency, cladogram 3.19.1 is only 1 positive occurrence more efficient than cladograms 3.19.2–3.

This example illustrates that, for a given sample of information containing conflict in positive occurrences, various cladograms are possible, which differ more or less among themselves in their efficiency in integrating the information. Although one cladogram may be judged most efficient according to a stated criterion, and may be accepted as the best cladogram in that sense, there is no guarantee that, if the cladogram is used as a basis for inference, the inference will be true. Suppose, nevertheless, that cladogram 3.19.1 is used as a basis for inference, namely that there exists a group containing species 2 and 3 and excluding species 1. If so, additional samples of information should tend to give the same result. Suppose also that an additional sample of information is taken, and that the sample is similar enough to the

Table 3.8. Character-Types and Their Occurrence in Three Species

	Character-Types								
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*D4*	*E*	*F*
1	+	–	–	–	+	+	–	+	–
2	–	+	–	+	+	–	+	+	–
3	–	–	+	+	–	+	+	+	–

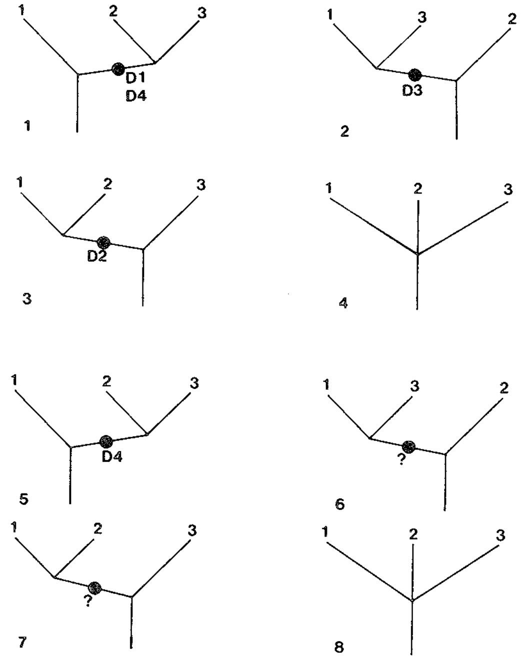


Figure 3.19. Positive occurrences in cladograms for three species. 1–4: All positive occurrences (D1, D2, D3, D4). 5–8: Positive occurrences that differ among cladograms (cf. table 3.8).

information of table 3.8, such that the sample is most efficiently summarized by a cladogram that duplicates cladogram 3.19.1. One may ask, what is the probability of achieving the same cladogram due to chance alone? Because there are three informative cladograms possible (3.19.1–3), the chance is 1 in 3, which is not a particularly low

Table 3.9. Positive (+) and Negative (–) Character-Types and Their Occurrences among Cladograms (cf. figure 3.19 and table 3.8)

	All Character-Types			*A, B, C, E, and F Eliminated*			*D1, D2, and D3 Eliminated*		
Cladograms	+	–	*Total*	+	–	*Total*	+	–	*Total*
3.19.1	10	10	20	6	4	10	1	1	2
3.19.2	11	10	21	7	4	11	2	1	3
3.19.3	11	10	21	7	4	11	2	1	3
3.19.4	12	11	23	8	4	12	2	1	3
table 3.8	14	13	27	8	4	12	2	1	3

probability. However, additional sampling with the same result would eventually produce a significantly low probability due to chance alone. A significantly low probability would offer some basis on which to judge the truth of the inference.

The conclusion to be drawn is not that a particular clustering procedure necessarily leads to the truth, as if truth depends upon mere conformity to procedure. The conclusion is that a particular clustering procedure yields a result that might be true or false, and that a judgment as to its truth or falsity may be considered an independent problem that can be investigated in other ways.

This conclusion should not be taken to imply that all clustering methods are in any sense equally valid. As shown by Farris (1977, 1979), for example, certain phenetic procedures give anomalous results in some circumstances. He supplied a hypothetical set of information for eight species, in which certain characters were multiplied by various factors (table 3.10). With or without the factors, the distribution of positive occurrences, which do not conflict among themselves, yields cladogram 3.20.1 as the most efficient summary. Without the factors, the same result is achieved by a certain phenetic procedure (UPGMA); but with the factors, the phenetic procedure yields no informative cladogram as a summary (figure 3.20.2). The phenetic procedure, in short, seems to behave as if it were clustering according to negative, rather than positive, occurrences. If, as argued above, negative occurrences are mainly false if considered informative, any clustering procedure sensitive to negative occurrences may produce anomalous results for that reason alone.

Conflict among positive occurrences, such as found in tables 3.6–3.8,

Table 3.10. Character-Types and Their Occurrence in Eight Species (modified from Farris 1977: table 1)

Species	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*
1	+	–	–	–	–	–	–	–	+	–	–	–	+	–
2	–	+	–	–	–	–	–	–	+	–	–	–	+	–
3	–	–	+	–	–	–	–	–	–	+	–	–	+	–
4	–	–	–	+	–	–	–	–	–	+	–	–	+	–
5	–	–	–	–	+	–	–	–	–	–	+	–	–	+
6	–	–	–	–	–	+	–	–	–	–	+	–	–	+
7	–	–	–	–	–	–	+	–	–	–	–	+	–	+
8	–	–	–	–	–	–	–	+	–	–	–	+	–	+
Factor	5	1	5	1	5	1	5	1	3	1	3	1	1	1

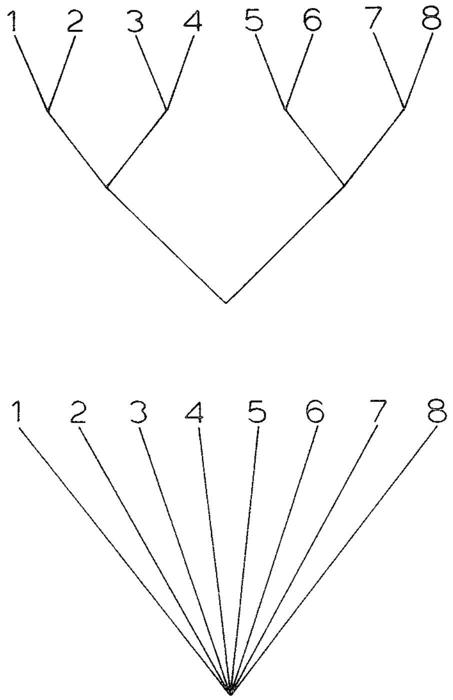


Figure 3.20. Two cladograms for eight species (cf. table 3.10). After Farris (1979), figure 1, p. 201.

may be considered in relation to the inferences for which cladograms serve as the basis. For example, if cladogram 3.19.1 is true (if there exists a group including species 2 and 3 and excluding species 1), then cladograms 3.19.2-3 are false (it is false that there exists a group including species 1 and 3 and excluding species 2; it is false, also, that there exists a group including species 1 and 2 and excluding species 3). It follows that character-types D1 and D4 are true, in the sense that they define a real group; and that character-types D2 and D3 are false, in the sense that they define unreal groups. Given a case of conflict among positive occurrences, therefore, one or more occurrence is false; that is, one or more positive occurrence is not a real occurrence.

Under the assumption that conflict among positive occurrences indicates one or more mistake in their designation as positive occurrences, the conflict may be investigated with this indication in mind. A concrete example may be considered with reference to certain features

Table 3.11. Characters and Their Occurrence in the Lancelet, Lamprey, and Shark

	Characters																				
Species	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*
Lancelet	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	+	+	+	+
Lamprey	+	+	+	–	–	–	+	+	+	–	–	–	–	–	–	+	+	+	+	+	+
Shark	–	–	–	+	+	+	+	+	+	–	–	–	–	–	–	+	+	+	–	–	–

shared by the lancelet and lamprey: an eel-like body (character 19), dorsal and caudal fins confluent (character 20), and a body wall with a smooth exterior surface, uninterrupted by paired fins (character 21). The suite of characters relating to this problem is summarized in table 3.11, and their character-types are summarized in table 3.12, which repeats the structure of the hypothetical example of table 3.6.

If character-type D1 is true, and character-type D2 is false, the implication is that characters 7, 8, and 9 are true, and that characters 19, 20, and 21 are false. "False" does not mean that the lancelet and lamprey truly do not have an eel-like body (character 19), dorsal and caudal fins confluent (character 20), and a body wall with a smooth exterior surface, uninterrupted by paired fins (character 21), for most assuredly they do. "False" means that the characters are not positive occurrences that define a real group including the lancelet and lamprey and excluding the shark.

How might the characters be found to be false positive occurrences? One way would be to discover that the features of the lancelet and lamprey (thought to constitute characters shared by them) are fundamentally different in some way, in which case they would no longer conform to character-type D2, but rather to separate characters of character-types A and B. And it would be correct to assert that characters 19, 20, and 21 are falsely considered as type-D2 characters.

Table 3.12. Character-Types and Their Occurrence in the Lancelet, Lamprey, and Shark

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*E*	*F*
Lancelet	+	–	–	–	+	+	–
Lamprey	–	+	–	+	+	+	–
Shark	–	–	+	+	–	+	–

Another way would be to discover the characters in the shark. If all three were discovered there, the characters would no longer conform to character-type D2, but rather to character-type E. And it would be correct to assert that characters 19, 20, and 21 are falsely considered as type-D2 characters.

Still a third way to find the characters to be false positive occurrences would be to discover that they are not positive occurrences because they are either (a) negative occurrences, in which case they would no longer conform to character-type D2, and they would raise a question about the existence and distribution of the corresponding positive occurrences, which might or might not be discoverable in the shark, or (b) not characters at all. In either case it would be correct to assert that characters 19, 20, and 21 are falsely construed as type-D2 characters.

These various ways are diagrammed in figure 3.21. Each way involves a change in the judgment of the generality of a character. In the first case (figures 3.21.1–2) a character is found to be less general than previously supposed, in the sense that it is found to consist of two separate characters (of types A and B) rather than one (of type D2). In the second case (figures 3.21.3–4) a character is found to be more general than previously supposed, in the sense that it is found to occur in three species (character-type E) rather than in two species (character-type D2). In the third case (figures 3.21.5–6) a character is again found to be less general than previously supposed, in the sense that it is found to occur in one species (character-type C) rather than in two species (character-type D2). In the final case (figures 3.21.7–8), the character is found to have no generality at all, in the sense that it cannot be represented in the cladogram in such a way as to be informative.

What of the characters shared by the lancelet and lamprey? Character 19, "an eel-like body," is a statement that might be questioned on the grounds that "an eel-like body" is a mere gestalt perception, and that the shapes of the lamprey and lancelet are actually two quite different characters that are readily distinguishable from each other (and from the shape of eels) if studied in greater detail. Character 20 "dorsal and caudal fins confluent," is said to occur in the embryonic shark, if in fact the dorsal and caudal fins are represented by the embryonic fin fold (figure 3.22); and the character might well be considered of greater generality (character-type E). Character 21, "a body wall with a smooth

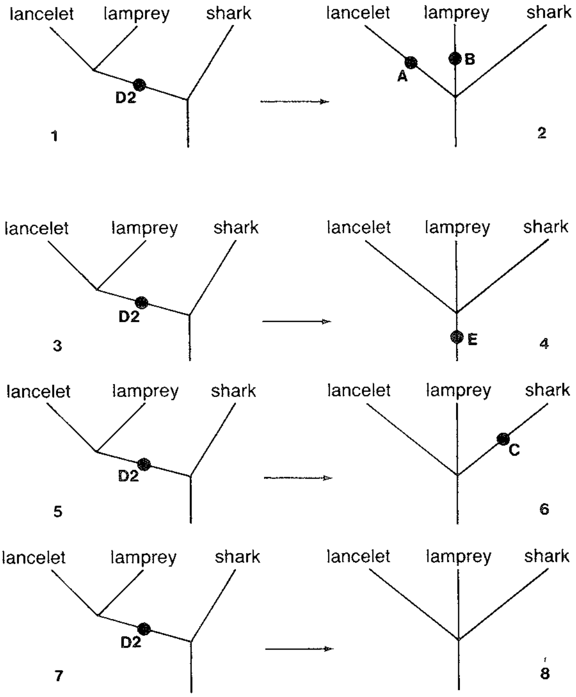


Figure 3.21. Possibilities for reinterpretation of positive occurrences of character-types in cladograms for three species. 1–2: An apparently positive occurrence (D2) reinterpreted as two less general positive occurrences (A, B). 3–4: An apparently positive occurrence (D2) reinterpreted as a more general positive occurrence (E). 5–6: An apparently positive occurrence (D2) reinterpreted as a negative occurrence corresponding to a positive occurrence (C) in the third species. 7–8: An apparently positive occurrence (D2) reinterpreted as no occurrence, either positive or negative.

exterior surface uninterrupted by paired fins," is a statement that is hardly distinguishable from "paired fins absent"; and the character might well be considered a negative occurrence.

Any character, of course, should be open to question and reinterpretation. What of the characters that conflict with those above? Character 7 (eyes), character 8 (cranial nerves), and character 9 (kidneys) each refers to organs that have been extensively studied and found to have (1) a coherent structure that may be identified and studied at gross, histological, and ultrastructural levels; (2) an assignable

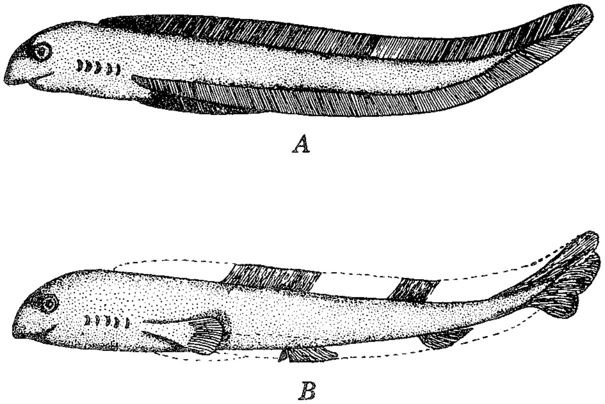


Figure 3.22. The relation of fins to embryonic fin folds according to the fin-fold theory. A: Fin folds as exemplified by embryonic sharks. B: Adult shark fins as remnants of fin folds. After C. K. Weichert (1959), Elements of Chordate Anatomy (New York: McGraw-Hill), figure 10.35, p. 280. Copyright © 1959, McGraw-Hill. Used with permission of McGraw-Hill Book Company.

function as elucidated by physiological studies at behavioral, neural, hormonal, and biochemical levels; and (3) a coherent embryological development that itself may be studied either structurally or physiologically at one or another level. For these reasons characters 7, 8, and 9 seem immune to the kind of reinterpretation applied to characters 19, 20, and 21. There is little reason, therefore, to view the conflict between character-types D1 and D2 as problematical. Character-type D2, as represented by characters 19, 20, and 21, can be rejected in its entirety.

Consider again the sample of information in table 3.8. The use of a clustering procedure sensitive to positive occurrences leads to the acceptance of cladogram 3.19.1. The inference is that character-types D1 and D4 are true and character-types D2 and D3 are false. It follows that the characters represented by types D2 and D3 should be amenable to reinterpretation, whereas the characters represented by types D1 and D4 should not be amenable to reinterpretation, or at least not with an equal facility. In a general sense, therefore, a clustering procedure is merely a means of predicting which positive occurrences, of those that conflict among themselves, are likely to be real and which are not. To test that prediction, additional information is required. A clustering procedure is unnecessary for recognizing conflicting positive occurrences, which may be studied directly without any predictions about which of them are likely to be real and which are not.

To the extent that conflicting positive occurrences can be studied and reinterpreted, conflicting occurrences disappear—if not in fact, at least in one's best judgment. If all conflict is resolved, such that all positive occurrences are combinable in a single cladogram, the choice of the most efficient summary is unproblematical: it is that cladogram that includes all positive occurrences as single lines. Yet as long as there is conflict among positive occurrences, there is a problem that may be investigated: namely, of the conflicting occurrences, which are real and which not? This residual problem cannot be solved, except perfunctorily, through the use of a clustering procedure. Its solution is possible only through study of organisms and new knowledge of, or new insight into, their real characteristics.

TREES AND INFORMATION

As summaries of information, cladograms do not in themselves imply a notion of evolution, or historical descent of species or other taxa. Nevertheless, cladograms, and branching diagrams generally, may be viewed from an evolutionary perspective, and viewed as such they may be termed phyletic trees. Considered as phyletic trees, branching diagrams pose certain problems, stemming from a basic question: how are characters represented in a phyletic tree?

The simplest case involves two species, let us say, again, a lamprey and a shark and certain information about them (figure 3.23 and table 3.2). Two general types of phyletic trees are possible, and each poses a problem in its exact relations with the information of table 3.2. Consider character-type A (characters unique to the lamprey): how are characters of type A represented on the various types of trees? For the bifurcating type of tree (figures 3.23.1–4) there are three possibilities: (1) the line extending to the lamprey (figure 3.13.1); (2) the line extending to the shark (3.13.2); (3) the line extending to both species (figure 3.13.3). Possibilities (1) and (2) imply that characters of type A evolved subsequent to the last occurrence of the species ancestral to the lamprey and shark (subsequent to the bifurcation); possibility (3) implies that the

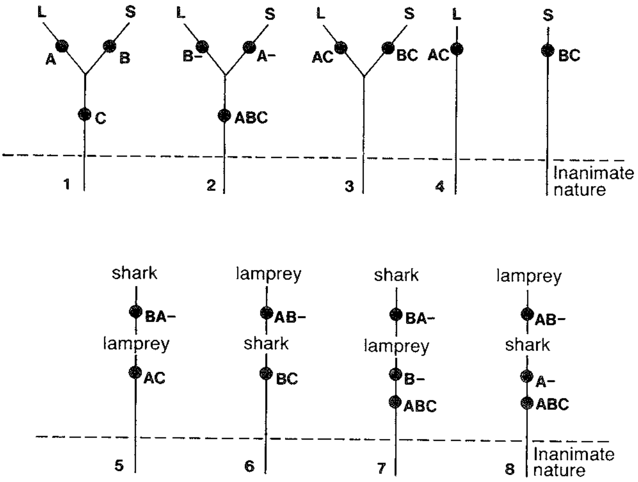


Figure 3.23. Representation of character-types (A, B, C) in phyletic trees for two species, lamprey and shark (cf. table 3.2).

characters evolved prior to the last occurrence of the ancestral species (prior to the bifurcation).

Because characters of type A occur only in the lamprey and not in the shark, possibility (2) may immediately be seen to be unsatisfactory. According to possibility (1), therefore, the characters unique to the lamprey evolved with the lamprey and not before (3.23.1); according to possibility (3), the characters evolved before, and were retained by the lamprey but were lost, or transformed, during the evolution of the shark (figure 3.23.2).

Consider characters of type C (present both in the lamprey and shark). The characters might have evolved either before (figures 3.23.1–2) or after (figures 3.23.3–4) the last occurrence, if any, of the species ancestral to both.

For trees without a bifurcation (figures 3.23.5–8), there are various possibilities for representing character-types. If the lamprey is considered ancestral to the shark (figures 3.23.5 and 3.23.7), characters of type A might have evolved early, and later have been lost or transformed. If the shark is considered ancestral to the lamprey (figure 3.23.6), characters of type A might have evolved only later; or, alternatively (figure 3.23.8), they might have evolved early, then have been lost, or transformed, only to reappear again later. Further possibilities in endless numbers would involve one or more additional cycles of still earlier evolution and loss, or transformation.

With the numerous possibilities for representing characters on trees even in the simplest possible example, how might the numerous (actually limitless) possibilities be evaluated? One way is simply to count the numbers of evolutionary gains and losses specified by each possibility (table 3.13). The possibilities differ among themselves in what they imply about gains and losses. If the true numbers of gains and losses were known, or assumed to be known, the possibilities could be judged according to how well they estimated the true numbers. But no true numbers of gains and losses can be derived from table 3.2. If the different possibilities are to be judged, therefore, they can be judged only in comparison among themselves. But such comparison itself poses a problem, namely, the criterion of judgment.

One possible criterion is parsimony. If this criterion is applied broadly, so as to include both evolutionary gains and evolutionary losses, the members of each class of trees, bifurcating and non-bifurcating, may be compared among themselves. A comparison of trees 3.23.1–4 shows that the character-types are represented more parsimoniously by tree 3.23.1 than by trees 3.23.2–4 (table 3.13); a comparison of trees 3.23.5–8 shows that the character-types are represented more parsimoniously by trees 3.23.5–6 than by trees 3.23.7–8; a comparison of trees 3.23.1 and 3.23.5–6 shows that the character-types are represented more parsimoniously by tree 3.23.1 than by trees 3.23.5–6. Interestingly, tree 3.23.1 corresponds to the cladogram, accepted on intuitive grounds

Table 3.13. Evolutionary Gains and Losses Implied by Various Placements of Character-Types in Phyletic Trees (cf. figure 3.23 and table 3.2)

	Evolutionary Gains				*Evolutionary Losses*				*Total Gains*
Trees	*A*	*B*	*C*	*Total*	*A*	*B*	*C*	*Total*	*and Losses*
3.23.1	1	1	1	3	0	0	0	0	3
3.23.2	1	1	1	3	1	1	0	2	5
3.23.3	1	1	2	4	0	0	0	0	4
3.23.4	1	1	2	4	0	0	0	0	4
3.23.5	1	1	1	3	1	0	0	1	4
3.23.6	1	1	1	3	0	1	0	1	4
3.23.7	1	2	1	4	1	1	0	2	6
3.23.8	2	1	1	4	1	1	0	2	6
table 3.2	?	?	?	?	?	?	?	?	?

as the best summary, in its representation of character-types (figure 3.14.1).

That the representation of one tree (3.23.1) may be found to be more parsimonious than that of other trees (3.23.5–6) is no proof that the one is true, or that the others are false, as statements about historical ancestry and descent. All that may reasonably be claimed is that, with respect to certain data (table 3.2), one representation (3.23.1) is truly more parsimonious than the others (3.23.5–6). Or, in other words, parsimony may be used as a criterion governing mode of representation of characters in phyletic trees. According to this mode, characters are properly represented in a tree at their most parsimonious level of generality. In more concrete terms, a character is represented by one line of a tree better than by two lines.

For a given set of character-types, one may ask if there is one tree in which the set is most parsimoniously represented. Consider the character-types of table 3.14. Two different trees are equally parsimonious in representing the character-types, for each tree implies two evolutionary gains and no losses (figure 3.24). But the trees differ in their complexity: tree 3.24.1 includes three lines; tree 3.24.2, only two. That one tree (3.24.2) represents the character-types by lines fewer than those of the other tree (3.24.1) is also an aspect of parsimony. Tree 3.24.2 is more parsimonious than tree 3.24.1 by one line, even though each tree implies the same number of evolutionary gains and losses of character-types. The most parsimonious phyletic tree, therefore, may be understood in a general sense as that tree implying the fewest evolutionary gains and losses with the fewest lines.

For two taxa, how many different phyletic trees are possible? This question is not easily answered. One may ask instead, how many different sets of character-types are possible for two taxa? Eight different sets exhaust the logical possibilities (table 3.15). For each set a most parsimonious "tree" is specifiable (figures 3.25.1–8); in that sense

Table 3.14. Character-Types and Their Occurrence in the Lamprey and Shark (cf. figure 3.24)

	Character-Types		
Species	*A*	*B*	*C*
Lamprey	–	–	+
Shark	–	+	+

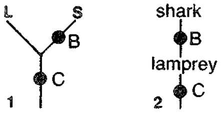


Figure 3.24. Parsimonious representation of character-types in two phyletic trees with different numbers of lines (cf. table 3.14).

there are eight possible "trees." If hybridization is considered, as represented in trees with reticulation, an infinite number of trees is possible, differing in the complexity of their reticulation. All such reticulate trees would be equally parsimonious in their implied evolutionary gains and losses if the character-sets are parsimoniously represented in the trees. But each reticulate tree would be less parsimonious in its number of lines than the corresponding nonreticulate tree. Compare, for example, trees 3.25.1 and 3.25.9. Each tree implies three evolutionary gains and no losses; tree 3.25.1 includes three lines, and tree 3.25.9 includes seven lines. Tree 3.25.1 is more parsimonious than tree 3.25.9 by four lines.

Given a tree most parsimonious in its representation of a certain set of

Table 3.15. Eight Sets of Character-Types for Two Taxa (lamprey and shark) and the Corresponding Most Parsimonious "Trees" (MPT) (cf. figure 3.25)

		Character-Types			
Set	*Species*	*A*	*B*	*C*	*MPT*
1	Lamprey	+	–	+	3.25.1
	Shark	–	+	+	
2	Lamprey	–	–	+	3.25.2
	Shark	–	+	+	
3	Lamprey	+	–	+	3.25.3
	Shark	–	–	+	
4	Lamprey	–	–	+	3.25.4
	Shark	–	–	+	
5	Lamprey	+	–	–	3.25.5
	Shark	–	–	–	
6	Lamprey	–	–	–	3.25.6
	Shark	–	+	–	
7	Lamprey	+	–	–	3.25.7
	Shark	–	+	–	
8	Lamprey	–	–	–	3.25.8
	Shark	–	–	–	

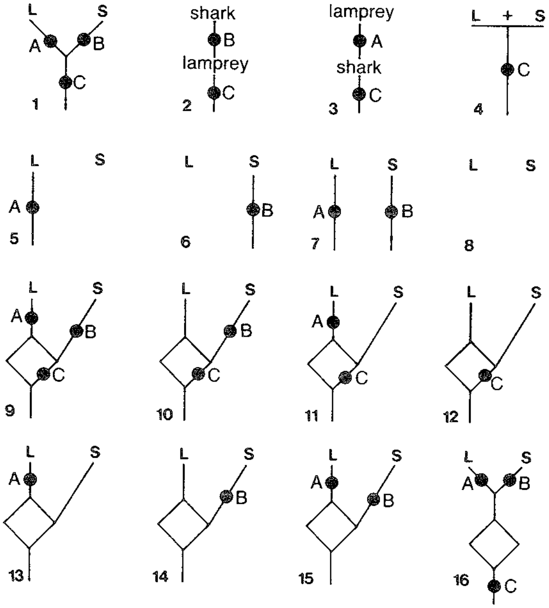


Figure 3.25. Phyletic trees. Trees 1–8: Most parsimonious trees for different sets of character-types (cf. table 3.15: sets 1–8). Trees 9–15: A phyletic tree with reticulation and parsimonious representation of seven sets of character-types (cf. figures 3.25.1–7 and table 3.15: sets 1–7). Tree 16: A phyletic tree with reticulation, additional lines, and parsimonious representation of one set of character-types (cf. figures 3.25.1, 3.25.9, and table 3.15: set 1).

character-types, what inferences may be derived from it? Consider set 3.15.8—no character-types known for the lamprey or shark, and no character-types known to be shared by them. One might infer that the absences are true absences or, alternatively, that the absence of the character-types is a mere artifact of sampling. Consider sets 3.15.2–7. Each set includes some, but not all, of the character-types of set 3.15.1. For each set one might infer that the missing character-types are true absences or, alternatively, that the absence of certain character-types is a mere artifact of sampling.

Two general types of inferences are possible: the absence of character-types in a particular set reflects: (1) the absence of such character-types in the real world or (2) sampling error, with the implication that the missing character-types might exist but are absent from a particular set because of chance alone. Each general type of inference leads to a different prediction: that further search for characters will produce (1) data sets identical to the first or (2) data sets that vary among themselves in the presence or absence of particular character-types. The predictions may be tested by sampling, which in general may be expected to show some variation in the presence and absence of some character-types from sample to sample. It is reasonable to expect that, with sampling sufficiently extensive, all possible character-types will be represented, that the combined samples will conform to set 3.15.1, and that the most parsimonious tree for any two taxa will be a bifurcating tree similar to tree 3.25.1.

That tree 3.25.1 seems to be the general solution to the problem of the most parsimonious tree for any two taxa again does not prove that tree 3.25.1 is true, and the other trees are wrong, as historical statements of ancestry and descent. Tree 3.25.1 might indeed be the most parsimonious tree for a certain set of character-types, or sets of character-types in general, even though another tree, say 3.25.2, might be historically correct. If so, historical truth must be judged by some criterion other than the most parsimonious tree.

Consider the possibility that further sampling shows no variability. Suppose an initial sample of information is summarized by set 3.15.2 and represented by tree 3.25.2, and that some number of additional samples of information all yield identical summaries and trees. At some point one might incline to the judgment that characters of type A do not, in fact, exist in the real world, i.e., that the absence is true. But there is a problem with this judgment, for it would be false if there existed a single character of type A. If that character were discovered, the discovery would automatically change the summary of all samples to set 3.15.1, and would change the most parsimonious tree to 3.25.1. If characters of type A actually exist in very low numbers, so as only rarely to occur in samples of information, their discovery would be improbable because of chance alone, but their occurrence would nevertheless be real. To conclude that characters of type A do not exist simply because none appears in a sample would seem unwarranted. In cases of extensive and exhaustive sampling, however, one might expect that "characters" of type A would normally be "discovered" through misidentification, which, because of chance alone, becomes ever more probable as sampling becomes more extensive.

A most parsimonious tree seems to offer no basis for inference beyond that of the corresponding cladogram for the reasons mentioned above. But one may suppose that a tree such as 3.25.1 is in some sense an aspect of the real world and, as such, deserves an explantion on its own terms. With omission of the "trees" with only one line, or no lines at all 3.25.4–8), there are four possible causal explanations in a historical sense, if trees can be regarded as such: trees 3.25.1–3 and some hybrid scheme of which there is an endless number. If tree 3.25.1 were true historically, the implication would be that evolution worked so as to produce a tree that is not only most parsimonious in a general sense, but also is historically true; if one of the other trees were true, e.g., tree 3.25.2, the implication would be that evolution worked so as to produce a tree that is most parsimonious in a general sense, but also is historically false. With these different possibilities, there seems no reason to equate the notions of most parsimonious tree and historical truth. Indeed, it may be best to divorce these two concepts and to inquire whether historical truth may be estimated by means other than the search for the most parsimonious tree.

The implication of the possible causal explanations, with the exception of reticulate trees, is evolutionary loss or evolutionary transformation of characters. If tree 3.25.1 is truly most parsimonious in a general sense, and if tree 3.25.2 is nevertheless true in a historical sense, characters of type A must have been lost during the evolution from lamprey to shark. Similarly, if tree 3.25.1 is truly most parsimonious, and if tree 3.25.3 is historically true, characters of type B must have been lost during the evolution from shark to lamprey. Searching without success for characters of type A in the shark, or characters of type B in the lamprey, one might judge that the characters are, in fact, truly absent. This judgment would mean also that tree 3.25.1 is truly most parsimonious. The possibilities always exist, however, that characters of type A are present in a transformed, but yet undiscovered, state in the shark; and that characters of type B are present in a transformed, but yet undiscovered, state in the lamprey. Discovery of transformed characters of types A and B would establish their presence in species previously thought to lack them. The discovery would also convert them into characters of type C. If all characters of type A were discovered in a transformed state in the shark, the combined set of all known character-types would change from set 3.15.1 to set 3.15.2, for the transformed characters would be type B; and the most parsimonious tree would change from tree 3.25.1 to tree 3.25.2 (table 3.16:set 1). Similarly, if all characters of type B were discovered in a transformed state in the lamprey, the combined set of all known character-types would change from set 3.15.1 to set 3.15.3, for the transformed characters would be of type A; and the most parsimonious tree would change from tree 3.25.1 to tree 3.25.3 (table 3.16:set 2). Interestingly, if all characters of type A were discovered in a transformed state in the shark, and all characters of type B were discovered in a transformed state in the lamprey, nothing would change, for the transformed characters would constitute a new set of characters of both types A and B (table 3.16:set 2A).

Thus the most parsimonious tree can be false in a historical sense if characters of types A or B have been lost or, which amounts to the same thing, if all the characters of one type have been so greatly changed when transformed into characters of the other type that the relation (transformation) between them is undiscoverable.

Thus the possibility exists that a character of type A is, in reality, a transformed character apparently of type B; and, conversely, that a

Table 3.16. Five Sets of Characters (of types A, B, C) for Two Taxa (lamprey, shark) with Possible Change (⟶) of a Character (+) into a Transformed State (t), and the Resulting Most Parsimonious Tree (MPT) (cf. figure 3.25) and Set of Character Types (cf. table 3.15)

		Character-Types					*MPT*	*Resulting Character-Types*			
Set	*Species*	*C*	*A*		*B*	*C*	*(Figure 3.25)*	*(Table 3.15)*	*A*	*B*	*C*
1	Lamprey	–	+		–	A	3.25.2	3.15.2	–	–	+
			↓								
	Shark	–	t	=	+	A(t)			–	+	+
2	Lamprey	–	+	=	t	B(t)	3.25.3	3.15.3	+	–	+
					↑						
	Shark	–	–		+	B			–	–	+
2A	Lamprey	–	+	=	t	B(t)	3.25.1	3.15.1	+	–	+
			↓		↑						
	Shark	–	t	=	+	A(t)			–	+	+
3	Lamprey	X →	t		–	X(t)	3.25.1	3.15.1	+	–	+
	Shark	X(t)	–		t ←	X			–	+	+
4	Lamprey	–	+	=	?	?	3.25.1	3.15.1	+	–	+
			↓		↑						
	Shark	–	?	=	+	?			–	+	+

character of type B is, in reality, a transformed character apparently of type A. To assume that such might be true in a particular case is to employ the notion of homology in its conventional, and evolutionary, sense. A lamprey, for example, has a single nostril (table 3.2:character 2), and a shark has two nostrils (table 3.2:character 5). To assume that the nostril of the lamprey and the nostrils of the shark are structures homologous in an evolutionary sense is to suppose that both kinds of nostrils have a common origin in some ancestral structure. Possibilities for a common origin include the following: that the ancestral structure was (1) a single nostril (A) like that of the lamprey, which was transformed somehow into the two nostrils of the shark (figure 3.26.1:B); (2) two nostrils (B) like those of the shark, which were transformed somehow into the single nostril of the lamprey (figure 3.26.2:A); (3) a structure (X) other than (1) and (2), which was transformed somehow into the single nostril (A) of the lamprey, on the one hand, and which was transformed into the two nostrils (B) of the

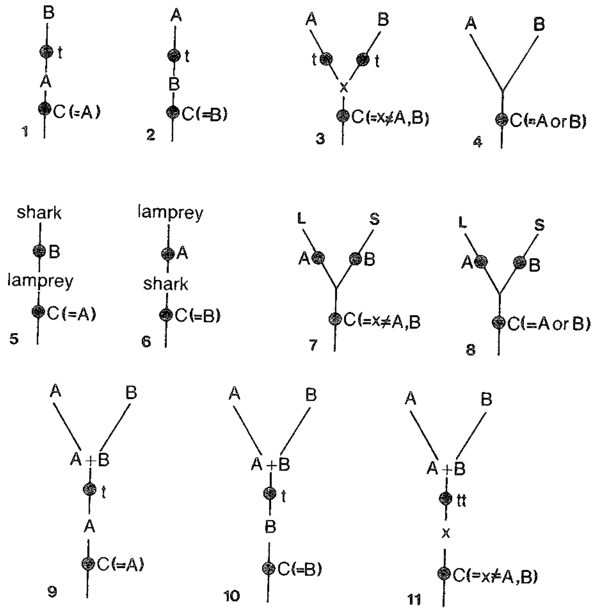


Figure 3.26. Trees implying transformations (t) of characters (A, B, X) for two species, lamprey (L) and shark (S). 1–4: Character trees. 5–8: Phyletic trees. 9–11: Character trees.

shark, on the other. Possibility (1) means that a single nostril is not a character of type A, but rather a character of type C, for it is present (in a transformed state) in the shark (figure 3.26.1); possibility (2) means that two nostrils are not a character of type B, but rather a character of type C, for they are present (in a transformed state) in the lamprey (figure 3.26.2); possibility (3) means that a single nostril is a character of type A, and that two nostrils are a character of type B, for neither character is present in a transformed state in another species, but rather both characters are transformations of a third character (X) of type C, present (in a transformed state) both in the lamprey and in the shark (figure 3.26.3).

According to possibility (1) in other words, the character of a single nostril is primitive or ancestral (plesiomorphic) relative to the character of two nostrils, which is advanced or derived (apomorphic; figure 3.26.1). According to possibility (2), the character of two nostrils is primitive or ancestral (plesiomorphic) relative to the character of one nostril, which is advanced or derived (apomorphic; figure 3.26.2). According to possibility (3), the character of a single nostril and the character of two nostrils are both advanced or derived (apomorphic) relative to yet another character (X) that is primitive or ancestral (plesiomorphic; figure 3.26.3).

A homology statement merely invokes the notion of a character in common or, in this example, a character of type C (nostril or nostrils present). A statement of evolutionary homology adds the possibility, but not the necessity, that a character previously thought to be of type A, or of type B, is in reality, a character of type C. In an evolutionary sense, therefore, a homology statement may be understood generally to imply the possibility of a particular character transformation: that, of two characters, one is primitive relative to the other, which is advanced.

One may compare the three possibilities, each in the form of a character tree (figures 3.26.1–3). In each tree the line leading to both characters represents the homologous relation (a character of type C) supposed to exist between them; and the line leading from one character to another represents an implied transformation (t). In terms of evolutionary gains and losses of characters and their transformations, two trees (3.26.1–2) are more parsimonious than the third (3.26.3), but the two trees are equally parsimonious relative to each other (table 3.17).

Table 3.17. Evolutionary Gains and Losses of Characters (A, B, X) and Character Transformations (A ⟶ B, B ⟶ A, X ⟶ A, X ⟶ B) Implied by Different Character Trees (cf. figure 3.26)

	Gains of Characters and Transformations								*Total Gains*
Trees	*A*	*B*	*X*	*A ⟶ B*	*B ⟶ A*	*X ⟶ A*	*X ⟶ B*	*Losses*	*and Losses*
3.26.1	1	0	0	1	0	0	0	0	2
3.26.2	0	1	0	0	1	0	0	0	2
3.26.3	0	0	1	0	0	1	1	0	3

A parsimonious representation in a single character tree would, therefore, have to represent, and allow for, two alternative and conflicting possibilities (figures 3.26.1–2). The possibilities are allowed by a bifurcating character tree (figure 3.26.4) wherein there are no lines extending from one character to another. In phyletic trees the characters would be represented by lines interconnecting species (figures 3.26.5–8).

There is correspondence between the character trees and the phyletic trees, such that each character tree is included within, and implied by, the corresponding phyletic tree (3.26.1, 3.26.5; 3.26.2, 3.26.6; 3.26.3, 3.26.7; 3.26.4, 3.26.8). The phyletic trees may, therefore, be considered generalized character trees, i.e., trees that represent all character trees combined. The set of characters represented in each pair of corresponding trees is given in table 3.16. The character trees (figures 3.26.1–4) represent only the characters, not the distribution of the characters among the species. The phyletic trees (figures 3.26.5–8) include both types of information, the characters and their distribution among the species. Again, tree 3.26.8 (cf. tree 3.25.1) seems to be the general solution to the problem of the most parsimonious tree for any two taxa, even in cases wherein character transformations occur. In its form and implications it is the same as the cladogram accepted on intuitive grounds (figure 3.14.1).

Tree 3.26.8 does not in itself preclude, or imply the preclusion of, the historical possibility of independent transformation of characters A and B from character X (figure 3.26.7), for this possibility is merely an example of nonparsimonious representation of characters in trees. Nevertheless the most parsimonious tree for this possibility (figure 3.26.7) is the same as the generally most parsimonious tree (figure 3.26.8). Nor does tree 3.26.8 preclude, or imply the preclusion of, the historical possibility that characters of types A and B were merely segregated from an ancestral species in which both types were present (figures 3.26.9–11). Segregation is merely another example of nonparsimonious representation of characters in trees, but in this example there exist the same possibilities for character transformations: A⟶B (figure 3.26.9); B⟶A (figure 3.26.10); X⟶A and X⟶B (figure 3.26.11). Nevertheless, the most parsimonious tree for these possibilities is still the generally most parsimonious tree (figure 3.26.8). Again, that there exists a generally most parsimonious tree does not prove that the tree is correct as a historical statement of ancestry and descent.

For three taxa there are 54 different possible sets of character-types, each of which specifies a different "tree" as its most parsimonious representation. Most of the "trees" are fragmentary in lacking enough lines to form an interconnected structure. In the sense of an interconnected structure in which each of the taxa is separated by at least one line from the other two taxa, there are only 22 trees. Of these, 18 trees are primary trees (figure 3.27) and four trees are tertiary trees (figure 3.28).

The 22 trees may be compared in their representation of certain information about the lancelet, lamprey, and shark (tables 3.3, 3.18). Without consideration of possible character transformations, one tree (3.27.1) is most parsimonious in its implications about evolutionary gains and losses. Interestingly, tree 3.27.1 corresponds to the cladogram that best summarizes the same set of character-types (figure 3.15.1).

Each of the 22 trees is a parsimonious representation of a different set of character-types (table 3.19). The 18 primary trees form three groups, and each group is defined by a different character of type D. In group 1 (table 3.19:left, figure 3.27:left) characters of type D occur in the lamprey and shark; in group 2 (table 3.19:center, figure 3.27:center) characters of type D occur in the lancelet and shark; in group 3 (table 3.19:right, figure 3.27:right) characters of type D occur in the lancelet and lamprey.

Among the six trees of each group, one tree has all possible character-types (e.g., tree 3.27.1), and of the other five trees, each has some but not all possible character-types (e.g., trees 3.27.4–5, 3.27.10, 3.27.13, 3.27.16). For the same reasons given in the discussion of trees for two taxa, the tree in which all character-types are represented is the generally most parsimonious tree of its group, and in that sense corresponds to a cladogram. Consequently the problem of the generally most parsimonious tree for three taxa involves only three primary trees (figures

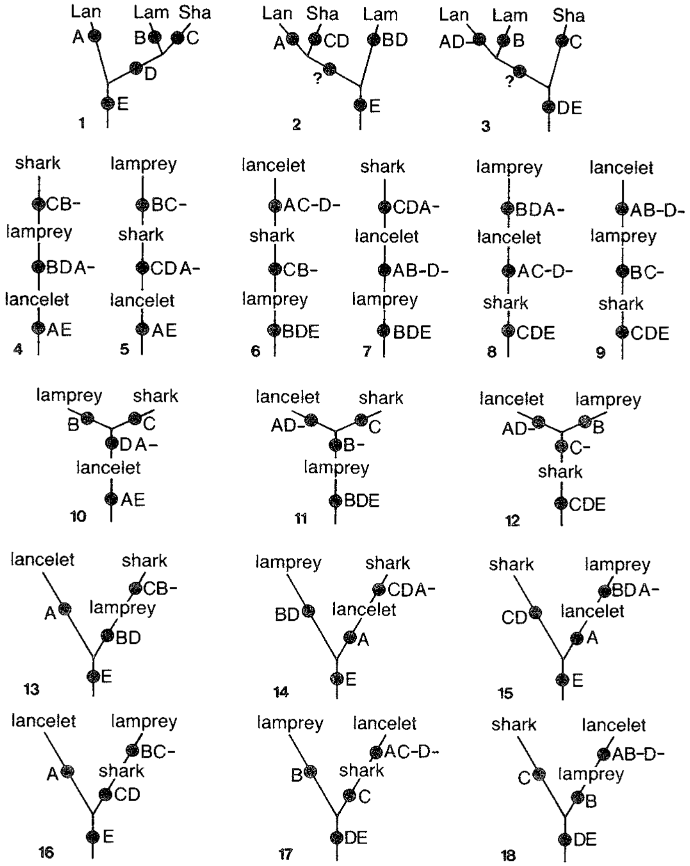


Figure 3.27. Parsimonious representation of one set of character-types in 18 primary trees for three species (cf. tables 3.3, 3.18).

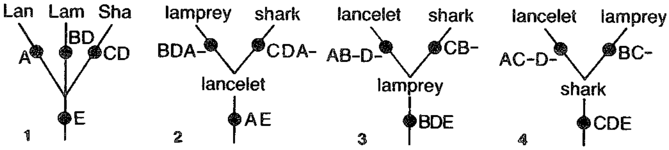


Figure 3.28. Parsimonious representation of one set of character-types in four tertiary trees for three species (cf. tables 3.3, 3.18).

Table 3.18. Evolutionary Gains and Losses Implied by Parsimonious Representation of Characters in Phyletic Trees of Three Species (cf. figures 3.27–3.28 and table 3.3)

Primary	*Evolutionary Gains*						*Evolutionary Losses*						*Total Gains*
Trees	*A*	*B*	*C*	*D*	*E*	*Total*	*A*	*B*	*C*	*D*	*E*	*Total*	*and Losses*
3.27.1	1	1	1	1	1	5	0	0	0	0	0	0	5
3.27.2	1	1	1	2	1	6	0	0	0	0	0	0	6
3.27.3	1	1	1	1	1	5	0	0	0	1	0	1	6
3.27.4	1	1	1	1	1	5	1	1	0	0	0	2	7
3.27.5	1	1	1	1	1	5	1	0	1	0	0	2	7
3.27.6	1	1	1	1	1	5	0	1	1	1	0	3	8
3.27.7	1	1	1	2	1	6	1	1	0	1	0	3	9
3.27.8	1	1	1	2	1	6	1	0	1	1	0	3	9
3.27.9	1	1	1	1	1	5	0	1	1	1	0	3	8
3.27.10	1	1	1	1	1	5	1	0	0	0	0	1	6
3.27.11	1	1	1	1	1	5	0	1	0	1	0	2	7
3.27.12	1	1	1	1	1	5	0	0	1	1	0	2	7
3.27.13	1	1	1	1	1	5	0	1	0	0	0	1	6
3.27.14	1	1	1	2	1	6	1	0	0	0	0	1	7
3.27.15	1	1	1	2	1	6	1	0	0	0	0	1	7
3.27.16	1	1	1	1	1	5	0	0	1	0	0	1	6
3.27.17	1	1	1	1	1	5	0	0	1	1	0	2	7
3.27.18	1	1	1	1	1	5	0	1	0	1	0	2	7
Tertiary													*Total Gains*
Trees	*A*	*B*	*C*	*D*	*E*	*Total*	*A*	*B*	*C*	*D*	*E*	*Total*	*and Losses*
3.28.1	1	1	1	2	1	6	0	0	0	0	0	0	6
3.28.2	1	1	1	2	1	6	2	0	0	0	0	2	8
3.28.3	1	1	1	1	1	5	0	2	0	1	0	3	8
3.28.4	1	1	1	1	1	5	0	0	2	1	0	3	8
table 3.3	?	?	?	?	?	?	?	?	?	?	?	?	?

3.27.1–3), and in that sense is equivalent to the problem of choosing the one cladogram, of three possible cladograms, that best summarizes the information. For choosing the cladogram that best summarizes the information, what is relevant are only positive occurrences of characters of type D. For choosing the most parsimonious tree, what is relevant are the implied total evolutionary gains and losses. But in choosing among the three generally most parsimonious trees, what is relevant are the gains and losses of characters of type D. Hence cladograms and generally most parsimonious trees turn out, at least in cases of three taxa, to be determined by the same factors. Cladograms and generally most parsimonious trees, therefore, seem merely to imply alternative

Table 3.19. Twenty-Two Sets of Character-Types for Three Species (lancelet, lamprey, shark) and the Corresponding Most Parsimonious Trees, Primary (MPPT) and Tertiary (MPTT) (cf. figures 3.27–3.28)

Group 1							Group 2							Group 3						
		Types							Types							Types				
MPPT	*Species*	*A*	*B*	*C*	*D*	*E*	*MPPT*	*Species*	*A*	*B*	*C*	*D*	*E*	*MPPT*	*Species*	*A*	*B*	*C*	*D*	*E*
3.27.1	Lan	+	−	−	−	+	3.27.2	Lan	+	−	−	+	+	3.27.3	Lan	+	−	−	+	+
	Lam	−	+	−	+	+		Lam	−	+	−	−	+		Lam	−	+	−	+	+
	Sha	−	−	+	+	+		Sha	−	−	+	+	+		Sha	−	−	+	−	+
3.27.4	Lan	−	−	−	−	+	3.27.6	Lan	+	−	−	+	+	3.27.8	Lan	−	−	−	+	+
	Lam	−	−	−	+	+		Lam	−	−	−	−	+		Lam	−	+	−	+	+
	Sha	−	−	+	+	+		Sha	−	−	−	+	+		Sha	−	−	−	−	+
3.27.5	Lan	−	−	−	−	+	3.27.7	Lan	−	−	−	+	+	3.27.9	Lan	+	−	−	+	+
	Lam	−	+	−	+	+		Lam	−	−	−	−	+		Lam	−	−	−	+	+
	Sha	−	−	−	+	+		Sha	−	−	+	+	+		Sha	−	−	−	−	+
3.27.10	Lan	−	−	−	−	+	3.27.11	Lan	+	−	−	+	+	3.27.12	Lan	+	−	−	+	+
	Lam	−	+	−	+	+		Lam	−	−	−	−	+		Lam	−	+	−	+	+
	Sha	−	−	+	+	+		Sha	−	−	+	+	+		Sha	−	−	−	−	+
3.27.13	Lan	+	−	−	−	+	3.27.14	Lan	−	−	−	+	+	3.27.15	Lan	−	−	−	+	+
	Lam	−	−	−	+	+		Lam	−	+	−	−	+		Lam	−	+	−	+	+
	Sha	−	−	+	+	+		Sha	−	−	+	+	+		Sha	−	−	+	−	+
3.27.16	Lan	+	−	−	−	+	3.27.17	Lan	+	−	−	+	+	3.27.18	Lan	+	−	−	+	+
	Lam	−	+	−	+	+		Lam	−	+	−	−	+		Lam	−	−	−	+	+
	Sha	−	−	−	+	+		Sha	−	−	−	+	+		Sha	−	−	+	−	+
MPTT	*Species*	*A*	*B*	*C*	*D*	*E*	*MPTT*	*Species*	*A*	*B*	*C*	*D*	*E*	*MPTT*	*Species*	*A*	*B*	*C*	*D*	*E*
3.28.1	Lan	+	−	−	−	+	3.28.2	Lan	−	−	−	−	+	3.28.3	Lan	+	−	−	−	+
	Lam	−	+	−	−	+		Lam	−	+	−	−	+		Lam	−	−	−	−	+
	Sha	−	−	+	−	+		Sha	−	−	+	−	+		Sha	−	−	+	−	+
3.28.4	Lan	+	−	−	−	+														
	Lam	−	+	−	−	+														
	Sha	−	−	−	−	+														

strategies for arriving at the same result—namely a generalization about the distribution of character-types among species in the real world. Among other of its attributes, the generalization restricts the range of possible historical explanations (if phyletic trees may be considered to be such) to a relatively small group of trees that are alike in their agreement with the generalization about the distribution of character-types. In the above example of three species, both the cladogram and the generally most parsimonious tree specify that characters of type D occur in the lamprey and shark, and suggest as a restricted group of possible historical explanations the phyletic trees of group 1, all of which agree, in the above respect (characters of type D), with the cladogram, the generally most parsimonious tree, and each other.

It seems superfluous to analyze the sets of character-types of tables 3.6–3.8 (figures 3.17–3.19), except in a cursory fashion (table 3.20). Of

Table 3.20. Evolutionary Gains and Losses Implied by Phyletic Trees, and the Generally Most Parsimonious Tree (GMPT) for Three Sets of Character-Types (cf. tables 3.6–3.8, figures 3.17–3.19)

Tree	*Gains*	*Losses*	*Total*	*GMPT*
3.17.1 3.17.3	6	1	7	3.17.4
	7	0	7	
3.17.2 3.17.4	6	2	8	
	7	1	8	
	8	0	8	
Table 3.6	?	?	?	
3.18.1 3.18.2 3.18.3	7	2	9	3.18.4
	8	1	9	
	9	0	9	
3.18.4	7	3	10	
	8	2	10	
	9	1	10	
	10	0	10	
Table 3.7	?	?	?	
3.19.1	8	2	10	3.19.1
	9	1	10	
	10	0	10	
3.19.2 3.19.3	8	3	11	
	9	2	11	
	10	1	11	
	11	0	11	
3.19.4	8	4	12	
	9	3	12	
	10	2	12	
	11	1	12	
	12	0	12	
Table 3.8	?	?	?	

Table 3.21. Numbers of Characters of Different Types in Three Species (lancelet, lamprey, and shark) (cf. figure 3.29)

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*
Lancelet	3	-	-	-	4	2	7
Lamprey	-	2	-	8	4	-	7
Shark	-	-	9	8	-	2	7

the diagrams in figure 3.17 (viewed as trees in relation to the character-types of table 3.6), two trees (3.17.1 and 3.17.3) are more parsimonious than the others, but equally parsimonious relative to each other. For each of the two trees (3.17.1 and 3.17.3), alternative representations of the character-types, equally parsimonious, are possible: according to one representation, there are six gains and one loss; according to the other, there are seven gains and no losses. Because there are two primary trees of equal (and maximum) parsimony, the generally most parsimonious tree is the tertiary tree 3.17.4. Of the diagrams in figure 3.18 (viewed as trees in relation to the character-types of table 3.7), all three primary trees (3.18.1–3) are equally parsimonious, and there are three alternative representations of the character-types; the generally most parsimonious tree is the tertiary tree 3.18.4. Of the diagrams in figure 3.19 (viewed as trees in relation to the character-types of table 3.8), one tree (3.19.1) is the most parsimonious, and there are three alternative representations of the character-types. In all three of the above examples, the generally most parsimonious tree duplicates the cladogram previously accepted as the best summary.

If extensive sampling is apt to result in some number of characters of each possible character-type, a realistic problem must concern the relative numbers of characters. Consider the hypothetical set of characters of table 3.21, wherein characters of type D occur in all possible combinations in unequal numbers: eight characters of type D1 (in lamprey and shark); four characters of type D2 (in lancelet and lamprey); two characters of type D3 (in lancelet and shark). Because the characters occur in different numbers, the numbers may be used as a basis for comparing the efficiency of different cladograms, and the parsimony of different trees. For three taxa, counting positive occurrences and counting evolutionary gains and losses give the same result (table 3.22), for no pattern of loss can give a total more parsimonious

Table 3.22. Positive Occurrences (= evolutionary gains) of Characters of Type D in Cladograms (= generally most parsimonious trees) and Some Reticulate Trees for Three Species (lancelet, lamprey, and shark) (cf. table 3.21 and figure 3.29)

Trees	*D1*	*D2*	*D3*	*Total*
3.29.1	8	8	4	20
3.29.2	16	8	2	26
3.29.3	16	4	4	24
3.29.4	8	4	4	16
3.29.5	16	4	2	22
3.29.6	8	8	2	18
3.29.7	8	4	4	16
3.29.8	8	2	4	14
3.29.9	8	2	4	14

than a pattern of multiple gains. Thus branching diagram 3.29.1 may be considered (as a cladogram) as the most efficient summary, and (as a tree) as the generally most parsimonious tree.

If hybridization is considered, as exemplified by reticulate trees (e.g., figures 3.29.4–6), the trees will be found yet more "parsimonious" in number of evolutionary gains (table 3.22). Such trees, however, are

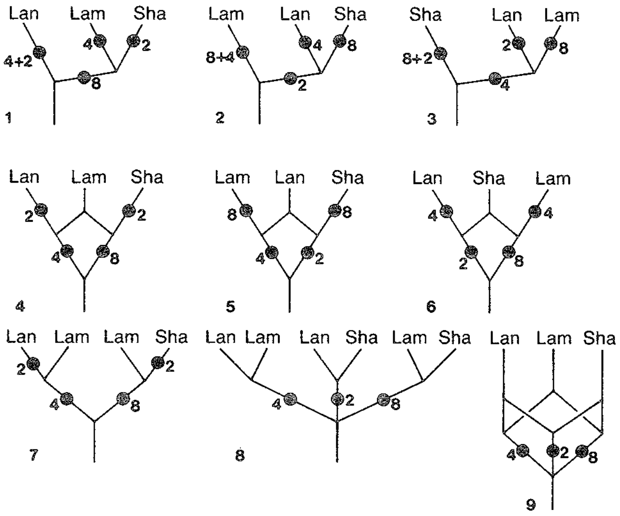


Figure 3.29. Positive occurrences (= evolutionary gains) of characters of type D (D1, D2, D3) in cladograms (= generally most parsimonious trees) and in some reticulate trees for three species, lancelet, lamprey, and shark (cf. tables 3.21–3.22).

equivalent to branching diagrams in which one or more species appears twice. Compare, for example, diagrams 3.29.4 and 3.29.7, which represent the same characters by the same lines. A tree with more complex reticulation is comparable to a branching diagram wherein all taxa appear more than once (3.29.8), and can be represented by a reticulate tree that is best viewed as three-dimensional (figure 3.29.9). Complex reticulation allows all combinations of characters of type D (D1, D2, D3) to be represented by single lines. For this reason, reticulate trees are always most "parsimonious" in their implied numbers of evolutionary gains and losses, even though they include more lines than nonreticulate diagrams. Hence, reticulate trees are not appropriate as possible candidates for the best summary (cladogram) or generally most parsimonious tree. In other words, reticulate trees do not integrate information; they merely reproduce it, and the reticulate tree that reproduces the information most "parsimoniously" (completely) is that tree that allows all possibilities to be represented by single lines (e.g., tree 3.29.9).

For three taxa, there are three possible kinds of characters of type D (D1, D2, D3). The cladogram that is the most efficient summary merely represents the greatest number of characters of type D by the single line available for such representation. Diagrams with more than one line for representing characters of type D for three taxa are equivalent to reticulate trees.

Reticulate trees with two lines available for representing characters of type D may be compared among themselves. Of trees 3.29.4–6, for example, tree 3.29.4 is more parsimonious than trees 3.29.5–6, for tree

Table 3.23. Relations Between Numbers of Characters of Type D (D1, D2, and D3) and (1) Positive Occurrences (total evolutionary gains and losses) and (2) Positive-Negative Occurrences (overall similarity) for Three Species (cf. tables 3.21, 3.24, and figure 3.29)

Tree	*Character-Types Represented by Single Lines*	*Number of Characters Represented*	*Total Positive Occurrences (D1 + D2 + D3)*	*Total Positive-Negative Occurrences (Overall Similarity)*		
				+	−	*Total*
3.29.1	D1	8	20	20	25	45
3.29.3	D2	4	24	24	19	43
3.29.2	D3	2	26	26	26	52

3.29.4 does represent more information with its two lines than do trees 3.29.5–6.

The relation between numbers of characters of type D (D1, D2, D3) and total positive occurrences (total evolutionary gains and losses) is demonstrated in table 3.23: as characters decrease, positive occurrences increase by a like amount. Such being the case, the notion of a most efficient cladogram, or generally most parsimonious tree, may be grasped immediately, at least in cases of three taxa, as indicated by the character-type (D1, D2, or D3) represented by the most characters (e.g., D1).

A POSTSCRIPT ON PARSIMONY

If parsimony and historical truth are best divorced, what is the significance of parsimony? One significance is its role as a criterion of representing characters in a branching diagram. One branching diagram of the many possible diagrams may be judged the most parsimonious for a particular sample of characters. If there is more than one sample, each sample will specify one diagram as its most parsimonious representation. Two or more such diagrams permit comparisons to be made among them, and ultimately permit a judgment as to whether they agree for reasons other than chance alone.

If there is agreement that is non-random, what is the cause? There are only two possibilities: the agreement is either (1) an artifact imposed upon the samples by the methods of the investigator, or (2) a reflection of some real factor that is independent both of method and investigator. Discovery of factors in category (1) constitutes increased knowledge of particular methods and their limitations; discovery of factors in category (2) constitutes increased knowledge of the world.

If our knowledge of the world ultimately stems from comparison of results that need not agree except by chance, but in fact do agree for some reason other than chance, we are well advised to understand, as clearly as possible, the nature of the comparisons that we make. How, then, are different samples of characters to be compared, if each sample, as is the case, may be represented by many different branching diagrams? One procedure is simply to limit comparisons to those few diagrams, one for each sample, that are most parsimonious representations of the samples. The limitation in itself can neither cause nor imply agreement among the diagrams for reasons other than chance alone.

Does parsimonious representation presuppose or imply that evolution (or some other causal factor) is parsimonious? To some persons, such seems to be the case. But let us try to be exact in our argument. If we observe nonrandom agreement, and that agreement is between diagrams most parsimonious for different samples, we conclude that a causal factor is at work. But what do we know of the causal factor? Only that it is the cause of the agreement among the diagrams, not that it is the cause either of parsimony or of the diagrams. If we liken parsimonious representation to a magnifying lens through which we look in order to see what otherwise would be invisible to our naked eye, we may better understand parsimony as a procedure, not a discovery. No one who observes, say, a dividing amoeba would assert that mitosis magnifies rather than multiplies amoebas.

That evolution may seem parsimonious, rather than merely orderly, is easy to understand. If we perceive nonrandom agreement (order) through comparison only of most parsimonious diagrams, our perception might be said to depend upon, or presuppose, parsimony (as our perception of a dividing amoeba might be said to presuppose a magnifying lens). That parsimony is presupposed may be misconstrued to imply that evolution is, or is presupposed to be, parsimonious (as mitosis would be misconstrued to cause magnification rather than multiplication of amoebas).

Does parsimony have a greater significance? If it does the significance lies in the relation between perceived order (as represented, for example, in a cladogram) and some set of historical explanations (as represented, for example, in a set of trees), restricted in their number by the use of a parsimony criterion (as represented, for example, by a set of trees denoted by one cladogram). Thus, the parsimony criterion may be used to specify a restricted set of historical explanations (trees). If one member of the set (of trees) were in fact true, then the notion that evolution is parsimonious may be defined and understood to mean exactly that (namely, that one tree of the set is true; and that an approximation to the truth is achieved first by specifying the set). Here the difficulties that attend the study of historical explanation (trees and their significance) become apparent. The cladogram may be considered to have a truth of its own even if the true historical explanation (the true tree) happens not to be a member of the set denoted by the cladogram (if evolution is truly not parsimonious in a given case). Thus, the notion that evolution is or is not parsimonious, in a given case, always implies comparison between a (true) cladogram and a (true) tree: if the (true) tree is a member of the set denoted by the (true) cladogram, evolution is (truly) parsimonious; if the (true) tree is not a member of the set denoted by the (true) cladogram, evolution is (truly) not parsimonious.

How might one form a judgment that evolution is or is not parsimonious (in the above sense) in a given case or in general? Ideally, some notion of the truth should be in hand, so that the use of a parsimony criterion could be judged according to how well its results match the truth. Yet the truth is not available to us in any particular case or in general. Without a notion of truth, what else is possible? We imagine one possibility, namely that the cladograms (restricted sets of historical explanations) for different groups of organisms might agree in the geographical dimension (biogeographically through area-cladograms, as detailed in Chapters 6–8 below). If there is geographical agreement among cladograms for different groups of organisms, then there is a reason to infer a common causal factor (historical explanation). In such a case there would be reason to infer that evolution was indeed parsimonious (that the true historical explanations lie within the restricted sets of trees).

Suppose that there is no geographical agreement among cladograms for different groups of organisms? Does the lack of agreement mean that evolution is not parsimonious? Failure to find agreement is not decisive unless it is supposed that the reason for failure is that the agreement does not exist to be found. But even under this supposition the true lack of agreement offers no basis for judgment as to which (sets of) trees are true. Failure to find agreement, then, is not evidence that evolution is not parsimonious, only that evolution possibly is not parsimonious.

To show that evolution is nonparsimonious requires that a certain tree and a certain cladogram are true, and that the tree is not a member of the set denoted by the cladogram. Is it possible that evidence could show such in a given case?

If there is geographical agreement among some cladograms, and we infer that there is a common historical explanation, what of a group distributed in a similar way but whose area-relationships conflict with the agreement? One possibility is that the conflict arises from nonparsimonious evolution: namely that there is one (true) historical explanation for all groups, one of which has a (true but) incongruent cladogram.

Thus, the use of a parsimony criterion does not presuppose anything in particular about the nature of evolution. Rather, a parsimony criterion makes possible certain comparisons, according to which evolution may be judged parsimonious, or possibly not parsimonious, or nonparsimonious, as the case may be. This kind of judgment may have little relevance in itself; it arises only as a by-product, as it were, of comparison between cladogram and tree, both of which are assumed true in a given case. Of course, *any* cladogram or tree may be assumed to be true. Judgment of truth is a matter independent of such assumption. We consider that a cladogram may be judged true on the basis of agreement among samples of characters (that agree for reasons other than chance alone or methodological artifact); and that a tree (or set of trees) may be judged historically true on the basis of agreement among cladograms of different groups as considered in the geographical dimension.

Biogeography is often asserted to constitute some of the strongest evidence in favor of evolution. The sense of the assertion is easy to perceive if it stems from agreement of cladograms in the geographical dimension. If agreement is perceived it permits judgment about the historical truth of (a set of) trees—a judgment that is perhaps unattainable through any other considerations independent of parsimony. Such agreement is direct evidence not of evolution, but rather of historical process that is parsimonious, i.e., of historical process that binds cladogram and tree as one.

To the extent that cladograms for different groups agree in the geographical dimension, what is perceived as order (a cladogram) and what is inferred as its historical cause (a tree) are bound together as one and the same notion. To a person with a materialistic philosophy, the notion easily transforms into evolution—as the by-product of mutation and recombination of genes. To a person with an idealistic philosophy, the notion easily transforms into evolution—as the by-product of movement of form or idea. For persons who contemplate final causes, the notion transforms into creation (if final cause is considered supernatural) or to historical process that is inexplicable (if final cause is considered natural but either unknown or unknowable).

Evolution thus seems to depend upon the notion that cladogram and tree are in fact bound as one, as revealed by actual evidence in hand. Our considerations suggest that relevant evidence consists only of agreement of cladograms in the geographical dimension—agreement that is by no means abundantly available at present. In fact, this evidence is notable by its very scarcity. Historically speaking, this evidence seems to have been taken for granted—which is understandable enough if only because cladograms and trees were only recently distinguished as different concepts. We conclude, therefore, that biogeography (or geographical distribution of organisms) has not been shown to be evidence for or against evolution in any sense. The significance of biogeography has been merely that biogeography has raised the possibility of agreement between cladograms as considered in the geographic dimension—a possibility that has been little studied, but one worthy of further investigation.

INFORMATION, PHYLETICS, AND PHENETICS

That there are more characters of type D1 than types D2 or D3 (D1 > D2, D3) in a particular sample of characters need not be considered significant in itself, for what is true of one sample need not be true of other samples. If significant, the relation D1 > D2, D3 should be general, in the sense that the relation is true of samples generally—a matter that is open to empirical investigation through further sampling.

A generalization about the distribution of characters among species (e.g., D1 > D2, D3) is interpretable in an evolutionary sense to indicate relative recency of common ancestry. In other words,

(1) If D1 > D2, D3 is generally true, then
(2) the lamprey and shark had an ancestor in common that was not an ancestor of the lancelet (as exemplified in figure 3.29.1).

According to this mode of interpretation, characters of type D1 are true homologies (synapomorphies, or marks of common ancestry) that evolved prior to the last occurrence of a common ancestor, and were inherited from that ancestor without subsequent loss in any of the descendant species. Also, characters of types D2–3 are interpretable in one of two ways:

(3) as true homologies inherited from a common ancestor with subsequent loss in one or more of the descendant species (if loss is an interpretation more parsimonious than multiple gains); or
(4) as characters not inherited from a common ancestor, but rather as characters developed independently two or more times subsequent to the last occurrence of the common ancestor (if multiple gain is an interpretation more parsimonious than loss).

For three taxa, the two interpretations are equally parsimonious in any given case.

This mode of interpretation, which has been employed commonly enough within systematic biology to be termed "traditional," leads to a seeming difficulty in its implications. The mode may be stated in an abbreviated form:

(5) If D1 > D2, D3 is generally true, then
(6) characters of type D1 were present in a common ancestral species, and characters of types D2–3 either were present in an ancestor and were lost in some descendant species, or they were not present in an ancestor at all.

The difficulty arises if one asks, how might one determine what characters were present in a common ancestor? If the only answer to this question is that

(7) characters present in a common ancestor are those of type D1, if
(8) D1 > D2, D3 is generally true,

then the difficulty is not seeming but real. The implication is circular, leads nowhere, and suggests that the notion of common ancestry, as inferred from character distributions, is meaningless. The difficulty is seemingly multiplied by a reconsideration of an earlier inference:

(1) If D1 > D2, D3 is generally true, then
(2) the lamprey and shark had an ancestor in common that was not an ancestor of the lancelet.

The additional difficulty arises if one asks, how might one determine whether the lamprey and shark had an ancestor in common that was not an ancestor of the lancelet? If the only answer to this question is that

(3) the lamprey and shark had an ancestor in common that was not an ancestor of the lancelet, if

(4) D1 > D2, D3 is generally true,

then the difficulty is not seemingly, but really, multiplied.

The difficulties may be obviated, at least temporarily, if common ancestry is regarded not as real knowledge, in addition to generalizations about character distributions, but merely as a mode of interpretation of generalizations about character distributions. Indeed, common ancestry, or simply "evolution" or "coming into being," may be regarded as a causal explanation, dictated by other factors, of whatever character distributions seem really to exist in a general sense. If, for example, D1 > D2, D3 is generally true, i.e., if the distribution really exists, one may assert simply as an article of faith that the character distribution exists because it (or, more exactly, the species for which it exists) evolved, or came into being.

In this connection, it may be noted that the nature of character distributions is sometimes differently construed. In the above example, only positive occurrences of characters of type D are considered relevant. Sometimes, however, and most commonly in phenetic studies, positive and negative occurrences are combined. Consider, for example, the characters of table 3.21. Three characters are known only for the lancelet, but eight characters (of type D1) are unknown only for the lancelet. In this sense there are 3 + 8 = 11 "characters" known for the lancelet. In table 3.24, the positive and negative occurrences are combined for all character-types of table 3.21.

The relative numbers of "characters" of type D (positive and negative occurrences) may be considered a measure of "overall similarity" among the species (table 3.23). For three species, the effect is to add negative "characters" of type D (D1, D2, D3), in numbers equal to the characters of types A, B, and C. Adding negative "characters" derived in this way can have no effect on the relative numbers of characters of type D (D1,

Table 3.24. Numbers of "Characters" (positive and negative occurrences) of Different Types in Three Species (based on table 3.21; cf. table 3.23)

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*
Lancelet	11	–	–	–	13	4	35
Lamprey	–	4	–	11	13	–	35
Shark	–	–	13	11	–	4	35

D2, D3) unless characters of types A, B, and C occur in different numbers. Adding such negative "characters" implies that, in some sense, they are real information, i.e., that there are real differences in the true relative abundance of characters of types A, B, and C, as reflected in their relative numbers in a particular sample. Thus in table 3.21, there are three characters known for the lancelet, two for the lamprey, and nine for the shark; in table 3.24 each of these values is added to the appropriate character of type D (3+D1 = 11, 2+D3 = 4, 9+D2 = 13).

Considering positive occurrences alone is one approach to systematics, which is sometimes termed phyletic ("cladistic"); considering positive and negative occurrences together is another approach to systematics, which is sometimes termed phenetic. Both approaches assume that, if characters of type D occur in different relative numbers within a sample of information, the differences may be informative. In a phyletic sense, the information is construed to mean that some characters are true homologies and others are parallelisms or convergences, i.e., that some characters are, in a general way, true and others are false. In a phenetic sense, the information is construed to mean that some character-types truly occur more abundantly than others, but that all occurrences of characters are true, and none is false. In a phyletic sense the information relevant for estimating common ancestry is positive occurrences of characters of type D. In a phenetic sense, the information relevant for estimating "overall similarity" is positive occurrences of characters of type D and, in addition, negative occurrences of characters of types A, B, and C.

To the extent that the relative numbers of characters of type D vary randomly in samples, both approaches are liable to mistake random differences for real ones. To the extent that relative numbers of characters of types A, B, and C vary randomly in samples, the phenetic approach is liable to mistake random differences for real ones. Both approaches face one hazard (random variation in numbers of characters of type D); the phenetic approach faces another (random variation in numbers of characters of types A, B, and C).

Interestingly, both approaches would assume that all characters of types A, B, and C, no matter what their relative numbers, are, or might be, real. The approaches differ with respect to characters of type D: in the phyletic sense, some are real and some are not, and the problem is to find out which are real and which are not; in the phenetic sense, all are real, and the problem is to estimate their true abundances.

The "overall similarities" of table 3.23 may be represented by a "most parsimonious" tree (3.29.3; lancelet and lamprey grouped together as most similar), different from that tree most parsimonious for positive occurrences only (3.29.1; lamprey and shark grouped together). Again, that there are more "characters" of type D2 than of types D1 and D3 (D2± > D1, D3) in a particular sample of characters need not be considered significant. If significant, the relation D2± > D1, D3 should be general, in the sense that the relation is true of samples generally—a matter that also is open to empirical investigation through further sampling.

A generalization about "overall similarity" (e.g., D2± > D1, D3) also is interpretable in an evolutionary sense to indicate relative recency of common ancestry. In other words,

(1) if D2± > D1, D3 is generally true, then
(2) the lancelet and lamprey had an ancestor in common that was not an ancestor of the shark (as exemplified in figure 3.29.3).

This mode of interpretation has been employed to some extent within systematic biology, by persons who assume that degree of "overall similarity" reflects relative recency of common ancestry. The mode, not surprisingly, leads to difficulties the same as those discussed above. One may ask, for example, how might one determine whether the lancelet and lamprey had an ancestor in common that was not an ancestor of the shark? If the only answer to this question is that

(3) the lancelet and lamprey had an ancestor in common that was not an ancestor of the shark, if
(4) D2± > D1, D3 is generally true,

then the same circularity is repeated. But there is an additional difficulty about a fundamental principle: how information about "characters" is to be represented by the lines of a phyletic tree.

Consider a simple case of two species (lamprey and shark) and two characters (A and B), each observed to occur in one species (table 3.25:left). Parsimonious representation is achieved in a "tree" having two lines (figure 3.30.1); a third line might be added (figure 3.30.2), but

Table 3.25. Relation Between Number of Characters of Different Types in Two Species (lamprey and shark) and Positive-Negative Occurrences, or "Overall Similarity" (cf. figure 3.30)

	Character-Types			*"Overall Similarity"*		
Species	*A*	*B*	*C*	*A*	*B*	*C*
Lamprey	1	0	0	1+	1–	2
Shark	0	1	0	1–	1+	2

what would it represent if no homology between A and B is implied? Dividing each of the two characters into positive and negative occurrences would produce two "characters" of type C, each present in both species (table 3.25:right). What would constitute parsimonious representation of the "characters"? Two lines, connected or not, would be "unparsimonious" (figure 3.30.3), inasmuch as no line would represent the "characters" of type C. A third line may be added for this purpose (figure 3.30.4). The line would represent both "characters" of type C, but would leave in doubt the nature of such "characters."

Analysis of "character" A may be considered in relation to character trees. If the third line indicates "homology" between A+ and A–, what can be said of the ancestral or primitive (plesiomorphic) condition (figure 3.31.1)? There are three possibilities: A+, A–, and some third condition Ax (figures 3.31.2–6). Parsimonious representation of each possibility is shown in figures 3.31.3, 3.31.5, and 3.31.6, respectively. One ancestral possibility (A–) results in a representation more parsimonious than the other two, in its implied numbers of gains and transformations (table 3.26).

Analysis of "characters" A and B together may be considered in relation to bifurcating "character" trees. What can be said of the ancestral condition? There are nine possibilities (figures 3.32.1–9), two of which (figures 3.32.2–3) are parsimoniously represented by non-

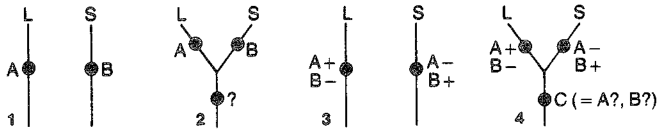


Figure 3.30. Parsimonious representation of character-types (A, B, C) and positive and negative occurrences of character-types (A+, A-, B+, B-) in phyletic trees for two species, lamprey and shark (cf. table 3.25).

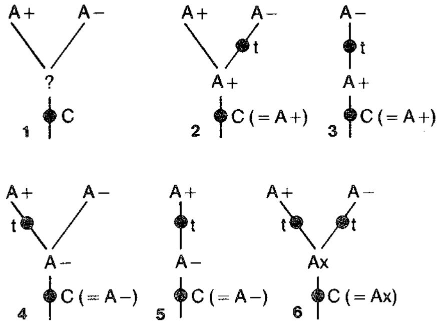


Figure 3.31. Trees implying transformations (t) of characters (A+, A-, Ax; cf. table 3.26).

bifurcating trees, because one line in each tree represents nothing, and could just as well be omitted. One ancestral condition (A–B–; figure 3.32.4) is more parsimonious than the other eight, in its implied gains and transformations (table 3.27). Thus, one step toward parsimonious representation may be achieved by specifying the ancestral condition in a phyletic tree (A–B–; figure 3.33.1). Another step may be achieved by eliminating redundancy of character representation (figure 3.33.2; eliminating B– from the line leading to the lamprey, and A– from the line leading to the shark). The resulting tree (figure 3.33.2) is a contradiction; its third line represents a "character" shared by both the lamprey and the shark, but there is no such "character": neither "character" A– nor B– is found in both taxa. The tree represents no more than a "tree" with two lines (figure 3.33.3), which is the final step toward parsimonious representation.

Examination of the possible "character" trees (figure 3.32) shows that

Table 3.26. Evolutionary Gains of Characters (A, Ax) and Character Transformations (A– → A+, A+ → A–, Ax → A+, Ax → A–) Implied by Different Character Trees (cf. figure 3.31)

	Gains of Characters		*Gains of Character Transformations*				
Trees	*A*	*Ax*	*A– → A+*	*A+ → A–*	*Ax → A+*	*Ax → A–*	*Total*
3.31.2	1	0	0	1	0	0	2
3.31.3	1	0	0	1	0	0	2
3.31.4	0	0	1	0	0	0	1
3.31.5	0	0	1	0	0	0	1
3.31.6	0	1	0	0	1	1	3

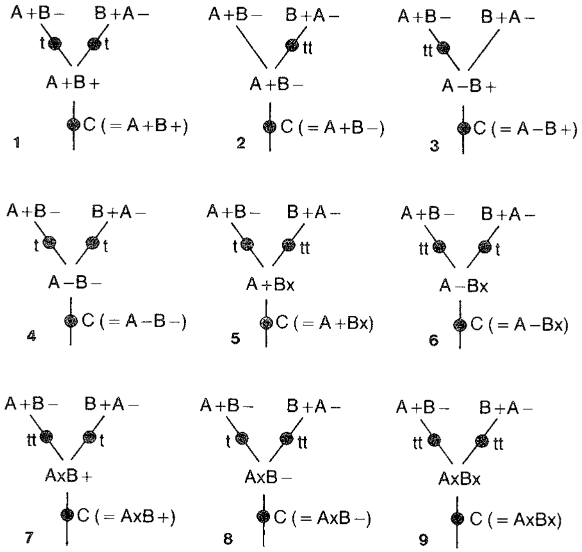


Figure 3.32. Trees implying transformations of characters (A+, A-, Ax, B+, B-, Bx; cf. table 3.27).

a bifurcating form is required only when ancestral conditions include positive occurrences (A+, Ax, B+, Bx). Hence, representation of both positive and negative occurrences in a bifurcating tree (e.g., figure 3.30.4), or trees generally, can be parsimonious only if positive occurrences are ancestral, i.e., if all negative occurrences are transformed positive occurrences (e.g., A+ ⟶ A–, Ax ⟶ A–). The conclusion seems unavoidable that under the assumption that relative "overall similarity" reflects relative recency of common ancestry, positive and negative occurrences may be represented in a phyletic tree only through a criterion other than parsimonious representation.

Is such a criterion possible? If not, then there is no way that positive and negative occurrences considered together can specify one phyletic tree rather than another. If such a criterion is possible, then two kinds of generalizations about character distributions are possible, as exemplified by those discussed above:

(1) D1 > D2, D3
(2) D2± > D1, D3

And it is possible that, in a given case, both might be generally true for the same three species. If so, each would lead to a different and conflicting statement about the relative recency of common ancestry. Such conflict would represent no more than a methodological artifact due to different criteria for representation of characters in trees.

But it is possible that, in a given case, both kinds of generalizations might agree. If, for example,

(1) D1 > D2, D3 is generally true, and if
(2) D1± > D2, D3 is generally true,

then what might explain the agreement? If the agreement in a given case is due to chance alone, then there is no need for further explanation. Hence, at this point an important question is: how may one judge whether the agreement is random (due to chance alone) or nonrandom (due to a cause other than chance)?

Purely random agreement would be expected to occur in 33 percent of pairs of generalizations about character distributions in three taxa. This percentage is too high to allow confidence in a judgment that a particular agreement is nonrandom. One way to decrease the probability is to consider samples of characters rather than pairs of generalizations. If, for example,

(1) D1 > D2, D3 is true for sample 1, and
(2) D1 > D2, D3 is true for sample 2, and
(3) D1± > D2, D3 is true for sample 3, and
(4) D1± > D2, D3 is true for sample 4,

then what is the probability that the agreement among the samples is due to chance alone, if there is no reason to expect agreement due to other factors? Given (1), the probability that (2) would agree is 33 percent; given (1) and (2), the probability that (3) would agree is 33 percent (with a combined probability of 11 percent); given (1), (2), and (3), the probability that (4) would agree is 33 percent (with a combined probability of 4 percent). A probability of 4 percent would indicate that the agreement is probably not due to chance alone.

One might ask how many characters are sufficient to serve as a sample. If the samples are truly independent, the probability that four characters would agree by chance alone is also 4 percent. Thus, four characters are sufficient to indicate a nonrandom distribution of characters in a case with three taxa, under the assumption of independence.

Table 3.27. Evolutionary Gains of Characters (A, Ax, B, Bx) and Character Transformations (A+ → A–, A– → *A*+, Ax → A+, Ax → A–, B+ → B–, B– → B+, Bx → B+, Bx → B–) Implied by Different Character Trees (cf. figure 3.32)

	Characters				*Gains of Character Transformations*								
Trees	*A*	*Ax*	*B*	*Bx*	*A*+ → *A*–	*A*– → *A*+	*Ax* → *A*+	*Ax* → *A*–	*B*+ → *B*–	*B*– → *B*+	*Bx* → *B*+	*Bx* → *B*–	*Total*
3.32.1	1	0	1	0	1	0	0	0	1	0	0	0	4
3.32.2	1	0	0	0	1	0	0	0	0	1	0	0	3
3.32.3	0	0	1	0	0	1	0	0	1	0	0	0	3
3.32.4	0	0	0	0	0	1	0	0	0	1	0	0	2
3.32.5	1	0	0	1	1	0	0	0	0	0	1	1	5
3.32.6	0	0	0	1	0	1	0	0	0	0	1	1	4
3.32.7	0	1	1	0	0	0	1	1	1	0	0	0	5
3.32.8	0	1	0	0	0	0	1	1	0	1	0	0	4
3.32.9	0	1	0	1	0	0	1	1	0	0	1	1	6

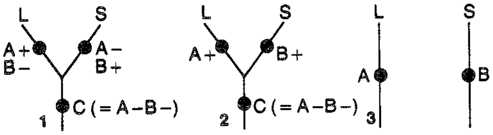


Figure 3.33. Steps toward parsimonious representation of positive and negative occurrences of character-types in phyletic trees for two species, lamprey and shark.

Suppose there is, in a given case, agreement between both factors (positive occurrences alone, and positive and negative occurrences combined). Is the inference of common ancestry more meaningful?

(1) If the agreement is not due to chance alone, then
(2) the lamprey and shark had an ancestor in common that was not an ancestor of the lancelet.

The inference (2) is open to criticism, and may actually be unwarranted in this hypothetical case, for the agreement may itself be an artifact due to nonindependence of the two factors that agree. But the inference illustrates that relative recency of common ancestry may be considered a causal principle not merely of a particular generalization about character distributions, but of an agreement between different kinds of generalizations that need not agree, except by chance, in a given case. If chance agreement can be ruled out, then some other causal principle may be sought. "Evolution" in the sense of "coming into being" by itself is no causal principle of nonrandom agreement, for "evolution" or "coming into being" implies nothing about nonrandom agreements or the lack of them. Relative recency of common ancestry, however, does imply that, if the lamprey and shark had an ancestor in common that was not an ancestor of the lancelet, then the lamprey and shark might be expected to exhibit the marks of that ancestry unique to themselves.

It is interesting to note that generalizations about character distributions relate to current controversies about biological classification. Three points of view may be distinguished, which are sometimes termed phenetic, phyletic ("cladistic"), and gradistic ("evolutionary"). In terms of the above example of three species (lancelet, lamprey, and shark), these points of view may be illustrated as follows:

Phenetic: (1) If $D1\pm > D2, D3$ is generally true, then

(2) the lamprey and shark should be classed together in a group that does not include the lancelet, even if
(3) $D2 > D1, D3$ or $D3 > D1, D2$ is also generally true.

Phyletic: (1) If $D1 > D2, D3$ is generally true, then
(2) the lamprey and shark should be classed together in a group that does not include the lancelet, even if
(3) $D2\pm > D1, D3$ or $D3\pm > D1, D2$ is also generally true.

Gradistic: (1) If D1 defines a biologically meaningful group (grade), then
(2) the lamprey and shark should be classed together in a group that does not include the lancelet,
(3) no matter what character distribution or distributions might generally be true.

At present, gradistic theory is not precisely formulated, and there is some doubt whether it can be, but the notion of a biologically meaningful group (grade) is sometimes said to be essentially the same as a phenetic group, and, if so, gradistic classification would be illustrated by the phenetic formulation above.

Because notions of "overall similarity" seem to be the root of current dispute about systematic philosophies, some additional remarks are appropriate. Consider tables 3.21 and 3.24. The former records information in terms of positive occurrences only; the latter records information in terms of positive and negative occurrences combined. Of interest in the former (table 3.21) are relative numbers of characters of type D (D1, D2, D3), which may serve as a basis of inference about the abundance of characters shared by two of the three species. Of interest in the latter also are relative numbers of characters of type D, which may serve as a basis of inference about the abundance of characters shared by two of the three species, but also of characters unique to each of the three species. Thus in the latter, positive occurrences of characters of type A increase the number of "characters" of type D1; characters of type B increase the number of "characters" of type D3; characters of type C increase the number of "characters" of type D2. In table 3.21 characters of type D1 are more numerous than characters of types D2–3; in table 3.24 "characters" of type D2 are more numerous than "characters" of types D1, D3. The reasons for the discrepancy are that in table 3.21 there are more characters of type C (9) than of types A (3) and B (2), and that when characters of type C (9) are added to those of type D2 (4), the total (13) is larger than the totals of A + D1 = 11 or of B + D3 = 4.

What are the implications of adding characters of types A–C to characters of type D? Consider table 3.21: 3 characters of type A, 2 characters of type B, 9 characters of type C. What might one infer to be true on the basis of these numbers? The only possibility would seem to be that the character-types really differ in their abundance. If the real differences in abundance are reflected in the sample, then

(1) $C > A > B$ is true.

This sort of inference is open to empirical investigation through additional sampling; in short, one may predict that additional samples would show the same relative abundance. But there is a philosophical difficulty with this sort of inference:

(1) if $C > A > B$ is true, then
(2) A, B, and C are unequal quantities and at least A and B (and possibly all three quantities) are therefore finite.

Are characters of a certain type finite in number? And, if so, how might that fact be learned? What seems beyond dispute is that characters are defined, or recognized, by systematists, who regularly count them, or compile lists of them, and then interpret the counts, or lists, in one fashion or another. Are such counts, or lists, theoretically finite in length? Or, alternatively, are such counts and lists always incomplete in the sense that they may always be increased? If so, there is some reason to believe that

(3) $A = B = C = \infty$ is true.

If so, then the different numbers of characters of types A, B, and C in a particular sample are purely the result of chance (sampling error). And adding their numbers to those of characters of type D (D1, D2, D3) adds only random variation to whatever real information characters of type D might contain.

But what of characters of type D? What might one infer to be true on the basis of the numbers of them in table 3.21? One possibility is that

(4) $D1 > D2 > D3$ is generally true.

But this is the same sort of inference, because

(4) if $D1 > D2 > D3$ is true, then
(5) at least D2 and D3 are finite and unequal quantities.

Doubts the same as those mentioned above apply to characters of type D. Thus, there is the possibility that

(6) $D1 = D2 = D3 = \infty$ is true.

But there is an alternative inference, namely that

(7) $D1 = \infty$ and $D2 = D3 = 0$.

Or, in other words,

(8) that the lamprey and shark are members of a group that does not include the lancelet is true, and
(9) that the lancelet and shark are members of a group that does not include the lamprey is false, and
(10) that the lancelet and lamprey are members of a group that does not include the shark is false.

The sense of inference (7) is reflected in the notion of evolutionary homology, in the sense that, for example, characters of type D1 may be interpreted as "true homologies" and characters of types D2 and D3 may be interpreted as "convergences" or "parallelisms." Interpreted as "convergences," they would be equivalent to "false homologies," i.e., "characters" that are not really characters at all, such that

(11) x "convergences" = 0 characters of type D.

Interpreted as "parallelisms," they are equivalent either to "false homologies," such that

(12) x "parallelisms" = 0 characters of type D;

or to characters of type E, present in some sense even in species from which they are absent, such that they are really plesiomorphic:

(13) x "parallelisms" = x characters of type E = 0 characters of type D.

Traditional notions of "convergence" and "parallelism" thus seem to exemplify inference (7), and seem to be at odds with inferences (5), (6), and, by implication, (2). Particular interpretations of "convergence" or "parallelism" may be based on generalizations about character distributions. But,

(14) if $D1 > D2, D3$ is true, and
(15) characters of types D2 and D3 are, therefore, convergences or parallelisms,

then one may always ask, how might one determine what characters are "convergent" or "parallel"? If the only answer is that

(16) characters of types D2, D3 are "convergent" or "parallel," if
(17) $D1 > D2, D3$ is true,

then circularity again emerges. What is needed is another kind of answer, so that there exists the possibility of nonrandom agreement.

With the above analysis certain philosophical differences underlying phenetic and phyletic ("cladistic") approaches seem clarified. Certain general conclusions may be reached. One possibility is that,

(18) if $A = B = C = \infty$ is true, and
(19) if $D1 = D2 = D3 = \infty$ is true, then
(20) all differences in relative numbers of characters of different types in a particular sample are random and due purely to sampling error, and
(21) relative numbers of characters furnish no basis for inference.

If so, then it is futile to use characters as a basis for inference, for no real parameter would be estimated either by a phenetic or by a phyletic ("cladistic") approach. There are, of course, other possibilities:

(22) if $A = B = C = \infty$ is true, and
(23) if $D1 = \infty$ and $D2 = D3 = 0$ is true, then
(24) $D1 > D2, D3$ is true and may be reflected in the relative numbers of characters of type D in a particular sample, so that
(25) relative numbers of characters of type D furnish a basis for inference.

If so, then it is futile to combine characters of types A, B, and C with characters of type D in the hope that a better basis for inference would be obtained. All that could possibly be obtained is greater sampling error (random variation) with no change in the single parameter estimated by both approaches. Still another possibility is that

(26) if A, B, and C are finite and unequal quantities, and
(27) if D1, D2, and D3 are finite and unequal quantities, then
(28) $D1 > D2, D3$ and $D2 \pm > D1, D3$ might both be true in a given case.

If so, then it is futile to compare the results of phenetic and phyletic approaches because each would estimate a different, and equally real, parameter.

How may these different possibilities be evaluated? Items (18)–(21) imply nothing but random variation, which is belied by the history of systematics. Items (22)–(25) and (26)–(28) allow a basis for inference to both phenetic and phyletic approaches. In the former case, (22)–(25), there is a single parameter estimated by both approaches; in the latter case, (26)–(28), there are two parameters, one for each approach. But the two parameters are not independent, for they contain some common elements (D1, D2, D3). If there really are two parameters, however, then

(29) of A, B, and C, some are finite and unequal quantities, such that
(30) characters of each type would not randomly vary in relative frequency among samples,

and the question would seem amenable to empirical investigation through sampling.

INFORMATION AND COMPONENTS

The 3-taxon problem plays a special role in systematics. Firstly, three taxa are the minimum that allows for a choice among cladograms; for two taxa there is only one possible cladogram, but for three taxa there are three possible primary cladograms. Secondly, the number (three) of possible cladograms for three taxa is not so large that a complete analysis (e.g., table 3.21, figures 3.29.1–3) is either prohibitively complex or beyond intuitive appraisal; complete analysis of four taxa (15 possible primary cladograms) is significantly, and prohibitively, more complex—perhaps beyond intuitive appraisal, and complete analysis of five taxa (105 primary cladograms), unaided by a computer, is out of the question. Thirdly, any complex problem, involving four or more taxa, may be reduced to a series of 3-taxon problems. No one would dispute that some systematic problems have been posed and subsequently solved without aid of computers. It is seldom realized, however, that virtually all such problems are, or imply, 3-taxon problems.

Consider a fourth species, the salmon. Suppose that study of the salmon shows that, of the three characters previously known only for the lancelet (table 3.21:A), one character occurs in the salmon (table

Table 3.28. Numbers of Characters of Different Types in Four Species (lancelet, lamprey, shark, and salmon) (derived from table 3.21, except for the 4 characters of type D)

	Character-Types													
Species	*A*	*B*	*C*	*D*	*E1*	*E2*	*E3*	*E4*	*E5*	*E6*	*F1*	*F2*	*F3*	*F4*	*G*
Lancelet	2	-	-	-	3	1	1	-	-	-	1	1	1	-	6
Lamprey	-	1	-	-	3	-	-	2	1	-	1	1	-	6	6
Shark	-	-	6	-	-	1	-	2	-	3	1	-	1	6	6
Salmon	-	-	-	4	-	-	1	-	1	3	-	1	1	6	6

3.28:E3); of the two characters previously known in the lamprey (table 3.21:B), one character occurs in the salmon (table 3.28:E5); of the nine characters previously known in the shark (table 3.21:C), three characters occur in the salmon (table 3.28:E6); of the eight characters previously known in the lamprey and shark (table 3.21:D1), six characters occur in the salmon (table 3.28:F4); of the four characters previously known in the lancelet and lamprey (table 3.21:D2), one character occurs in the salmon (table 3.28:F2); of the two characters previously known in the lancelet and shark (table 3.21:D3), one character occurs in the salmon (table 3.28:F3); of the seven characters previously known in the lancelet, lamprey, and shark (table 3.21:E), six characters occur in the salmon (table 3.28:G). Suppose also that study of the salmon reveals four characters unique to that species (table 3.28:D). The total information known for the four species would be shown in table 3.28, which is a sample "realistic" in the sense that all possible character-types are represented by some characters. Character-types relevant for the most efficient cladogram (generally most parsimonious tree) are increased from three (for three taxa: D1, D2, D3) to ten (table 3.28:E1–6, F1–4). Also the kinds of relevant character-types are increased from one (for three taxa: characters of type D that occur in two of three taxa) to two (for four taxa: characters of type E that occur in two of four taxa, and characters of type F that occur in three of four taxa). Unlike the assembled data for three taxa (table 3.21), for which the most efficient cladogram may be determined at a glance, the data for four taxa seem immune to quick assessment.

An alternative to a complete analysis of the 15 primary cladograms for four taxa is reduction of the four taxa to three. On the basis of previous considerations (table 3.21), one may infer that the lamprey and shark form a group that does not include the lancelet; hence the lamprey

Table 3.29. Numbers of Characters of Different Types in Four Species (lancelet, lamprey, shark, and salmon), Two of Which (lamprey and shark) Are Combined into a Single Taxon (derived from table 3.28)

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*
Lancelet	2	–	–	–	5	1	8
Lam + Sha	–	9	–	10	5	–	8
Salmon	–	–	4	10	–	1	8

and shark may be combined in one taxon, so as to reduce the number of character-types (table 3.29, derived from table 3.28):

Table 3.28 (lamprey + shark)	Table 3.29
A	A
B + C + E4	B
D	C
E5 + E6 + F4	D1
E1 + E2 + F1	D2
E3	D3
F2 + F3 + G	E

The problem may be viewed as beginning with four taxa: lancelet, salmon, lamprey, shark (figure 3.34.1), three of which (lancelet, lamprey, shark) are selected as an initial 3-taxon problem. If the lamprey and shark are grouped together, the result is cladogram 3.34.2. The next step concerns the placement of the salmon, for which there are seven possibilities (figures 3.34.2:1–7), three of which (1, 3, 5) define another 3-taxon problem (figure 3.34.3), for each possibility specifies a different cladogram (figures 3.34.4–6). These three possibilities for placement of the salmon (1, 3, 5) are those subsumed by the secondary cladogram (figure 3.34.2).

A glance at table 3.29 suffices to show that the salmon has its relationships with the lamprey-shark taxon, i.e., that the most efficient cladogram is the secondary cladogram of figure 3.34.6. If true, the secondary cladogram (3.34.6) implies that two other 3-taxon cladograms are also true (figures 3.34.7–8), and the characters relevant to these two cladograms (tables 3.30 and 3.31) may also be derived from table 3.28:

Table 3.28 (- shark)	Table 3.30	Table 3.28 (- lamprey)	Table 3.31
A + E2	A	A + E1	A
B + E4	B	C + E4	B
D + E6	C	D + E5	C
E5 + F4	D1	E6 + F4	D1
E1 + F1	D2	E2 + F1	D2
E3 + F3	D3	E3 + F2	D3
F2 + G	E	F3 + G	E

A glance at tables 3.30 and 3.31 suffices to show that each implied

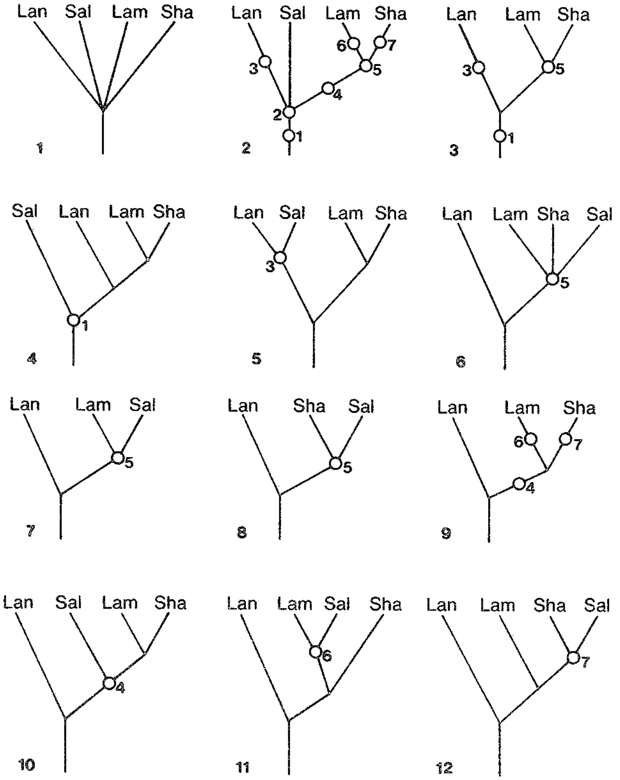


Figure 3.34. Four species (lancelet, lamprey, shark, salmon), considered with reference to three 3-taxon problems. Numbers 1–7 are possibilities for placement of a fourth species, the salmon (cf. tables 3.28–3.32).

Table 3.30. Numbers of Characters of Different Types in Three Species (lancelet, lamprey, and salmon) (derived from table 3.28)

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*
Lancelet	3	–	–	–	4	2	7
Lamprey	–	3	–	7	4	–	7
Salmon	–	–	7	7	–	2	7

Table 3.31. Numbers of Characters of Different Types in Three Species (lancelet, shark, and salmon) (derived from table 3.28)

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*
Lancelet	5	-	-	-	2	2	7
Shark	-	8	-	9	2	-	7
Salmon	-	-	5	9	-	2	7

3-taxon cladogram (figures 3.34.7–8) is the most efficient for the relevant data. The only problem remaining concerns the interrelationships of the lamprey, shark, and salmon. This problem may also be viewed as the placement of a fourth taxon (salmon) in a cladogram previously resolved for three taxa, with the possibilities reduced to three (figure 3.34.9) from the original seven (figure 3.34.2), or in other words to three 3-taxon cladograms (figures 3.34.10–12). These three possibilities for placement (4, 6, 7) are those subsumed by the secondary cladogram (figure 3.34.6). Together with the two possibilities previously considered (figures 3.34.4 and 3.34.5:1,3), these three (4, 6, 7) exhaust the five possibilities for dichotomous resolution allowed under the assumption that the lamprey and shark together form a group that does not include the lancelet.

The relevant data (table 3.32) may also be derived from table 3.28:

Table 3.28 (– lancelet)	Table 3.32
B + E1	A
C + E2	B
D + E3	C
E6 + F3	D1
E4 + F1	D2
E5 + F2	D3
F4 + G	E

A glance suffices to show that the salmon has its relationships with the shark, not the lamprey (figure 3.34.12).

In the above account, all possible placements of the fourth taxon (salmon) have been considered. That placement 2 subsumes possibilities 1, 3, and 5, and that placement 5 subsumes possibilities 4, 6, and 7, mean that placement 2 subsumes all (1–7) possibilities for dichotomous

Table 3.32. Numbers of Characters of Different Types in Three Species (lamprey, shark, and salmon) (derived from table 3.28)

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*
Lamprey	4	–	–	–	3	2	12
Shark	–	7	–	4	3	–	12
Salmon	–	–	5	4	–	2	12

resolution. This conclusion has an important implication. Given the cladogram of figure 3.34.2, for example, one may assert that, if the cladogram is true, the lamprey and shark together form a group; but it is not clear whether both the lancelet and salmon are excluded from the group, or merely the lancelet or the salmon is excluded from the group; and if one and not the other is excluded, it is not clear which is which.

Given information about characters in four taxa (e.g., table 3.28), one may select any three taxa for an initial 3-taxon problem: for example, lancelet, lamprey, shark (figure 3.34.3); lancelet, lamprey, salmon (figure 3.34.7); lancelet, shark, salmon (figure 3.34.8); lamprey, shark, salmon (figure 3.34.12). Whatever the result, there is at least one additional 3-taxon problem, but there is a maximum of two additional 3-taxon problems, in the placement of the fourth species in a dichotomous cladogram. If the initial placement of the fourth species results in a primary (dichotomous) cladogram of the four taxa (e.g., figure 3.34.12), there would have been only one additional 3-taxon problem (for a total of two). If the initial placement of the fourth species results in a secondary (trichotomous) cladogram of the four taxa (e.g., figure 3.34.6), there would have been one, but there would be yet another, additional 3-taxon problem (for a total of three).

Similarly, for five taxa (105 possible dichotomous cladograms), there is a minimum of three 3-taxon problems, and a maximum of six, required for dichotomous resolution (table 3.33). For six taxa (945 possible dichotomous cladograms), there is a minimum of four 3-taxon problems, and a maximum of ten, required for dichotomous resolution. For ten taxa (34,459,425 possible dichotomous cladograms), there is a minimum of eight 3-taxon problems, and a maximum of 36, required for dichotomous resolution. It is apparent that, for a group of many species, no cladogram could possibly be achieved by complete analysis of all possibilities without either the aid of a computer, or the reduction of the

Table 3.33. Number of Taxa in Relation to Number of Possible 3-Taxon Problems (for completely dichotomous resolution), Number of Implied 3-Taxon Cladograms, Number of Character-Types Shared by Two or More Taxa, and Number of Possible Dichotomous Cladograms

	3-Taxon Problems		*Implied 3-Taxon Cladograms*			
Taxa	*Minimum*	*Maximum*	*Maximum*	*Minimum*	*Shared Character-Types*	*Dichotomous Cladograms*
2	0	0	0	0	1	1
3	1	1	0	0	4	3
4	2	3	3	2	11	15
5	3	6	9	7	26	105
6	4	10	19	16	57	945
7	5	15	34	30	120	10,395
8	6	21	55	50	247	135,135
9	7	28	83	77	502	2,027,025
10	8	36	119	112	1013	34,459,425

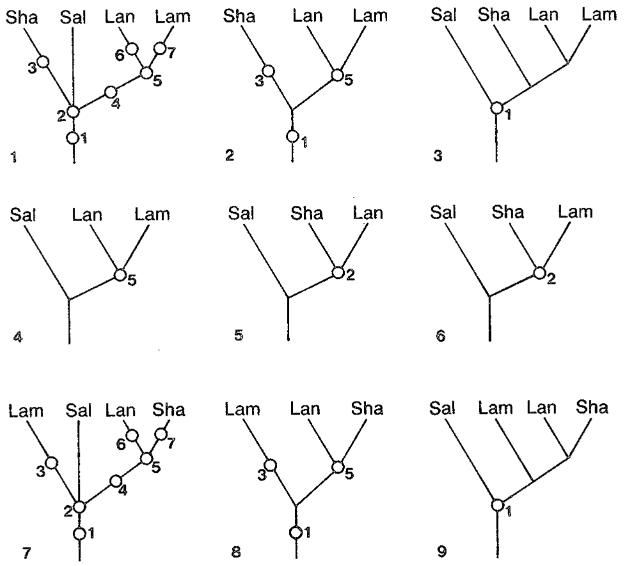


Figure 3.35. Four species (lancelet, lamprey, shark, salmon), considered with reference to two 3-taxon problems. Numbers 1–7 are possibilities for placement of a fourth species, the salmon (cf. tables 3.34–3.35).

many species to a series of 3-taxon problems. That cladograms for large groups have been achieved without the aid of a computer (e.g., Hennig 1969) suggests that the 3-taxon approach was either the method of choice or the method intuitively applied.

Although the above examples include taxa represented by species (lancelet, lamprey, etc.), the 3-taxon approach may be applied generally to taxa of any number of species. Thus, instead of the lancelet, lamprey, shark, and salmon, one may inquire into the relationships of groups more inclusive, such as the Echinodermata, Vertebrata, Mollusca, and Annelida. One may, of course, simply accept such groups as given, in some current classification, without concern for their composition. But if so, problems relating to their composition might possibly be uncovered and go unrecognized. An example of such a problem may be illustrated with reference to the lancelet, lamprey, shark, and salmon, under the assumption that the lancelet and lamprey are members of a group that does not include the shark (figure 3.35.1). Information relevant to this example (table 3.34) may be derived from table 3.28:

Table 3.28 (lancelet + lamprey)	Table 3.34
A + B + E1	A
C	B
D	C
E6	D1
E2 + E4 + F1	D2
E3 + E5 + F2	D3
F3 + F4 + G	E

As before, there are seven possibilities for placement of the salmon, and possibility 2 subsumes all others (figure 3.35.1). The three possibilities of the next 3-taxon problem are those of figure 3.35.2. A

Table 3.34. Numbers of Characters of Different Types in Four Species (lancelet, lamprey, shark, and salmon), Two of Which (lancelet and lamprey) Are Combined in a Single Taxon (derived from table 3.28)

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*
Lan + Lam	6	–	–	–	4	3	13
Shark	–	6	–	3	4	–	13
Salmon	–	–	4	3	–	3	13

glance at table 3.34 shows that possibility 1 specifies the most efficient cladogram (figure 3.35.3), which is a primary (dichotomous) cladogram. Accordingly, there is no additional 3-taxon problem. If cladogram 3.35.3 is true, there are, nevertheless, three implied 3-taxon cladograms that also must be true (figures 3.35.4–6), the data relevant to which have already been assembled for another purpose (tables 3.30–3.32). A glance at tables 3.30–3.32 shows that each implied cladogram is not the most efficient summary of the relevant data; hence the implied cladograms would seem false. That the cladograms seem false suggests in turn that the initial grouping (lancelet + lamprey) is also false. Such is apparent only because the implied cladograms, and the data relevant to them, are at hand.

Study of inclusive taxa, such as Echinodermata, Vertebrata, etc., might similarly lead to an erroneous conclusion for the same reason—namely that one of the groups assumed to exist does not exist. Such would be apparent only if data relevant to the implied cladograms were at hand, so that the implied cladograms, or some sufficient number of them, could be checked.

Another example of the same sort arises if one groups together the lancelet and shark, which would likewise allow for seven possible placements of the salmon (figure 3.35.7). The three possibilities of the next 3-taxon problem are 1, 3, and 5 (figure 3.35.8), and the relevant data (table 3.35) may be derived from table 3.28:

Table 3.28 (lancelet + shark)	Table 3.35
A + C + E2	A
B	B
D	C
E5	D1
E1 + E4 + F1	D2
E3 + E6 + F3	D3
F2 + F4 + G	E

A glance at table 3.35 shows that possibility 1 is the most efficient cladogram (figure 3.35.9), which is a primary (dichotomous) cladogram. Accordingly, there is no additional 3-taxon problem. If cladogram 3.35.9 is true, there are, nevertheless, three implied 3-taxon cladograms that also must be true, and they are the same as those of the previous

Table 3.35. Numbers of Characters of Different Types in Four Species (lancelet, lamprey, shark, salmon), Two of Which (lancelet and shark) Are Combined in a Single Taxon (derived from table 3.28)

	Character-Types						
Species	*A*	*B*	*C*	*D1*	*D2*	*D3*	*E*
Lan + Sha	9	–	–	–	6	5	13
Lamprey	–	1	–	1	6	–	13
Salmon	–	–	4	1	–	5	13

example (figures 3.35.4–6), all of which again seem false and suggest in turn that the initial grouping (lancelet + shark) is also false.

Groupings such as the lancelet + lamprey, which seem false in relation to available information, are sometimes termed "nonmonophyletic." Interpreted within the context of a phyletic tree, such a grouping would include some, but not all, of the descendant species of a common ancestral species. Interpreted in relation to character distributions,

(1) a grouping is false ("nonmonophyletic") if it conforms to D2 or D3, when
(2) D1 > D2, D3 is true.

False groupings, to the extent that they occur in classification, are apt to mislead any investigator who accepts them as true, unless he can examine the cladograms implied by his results in relation to relevant data sufficient to expose the initial grouping(s) as false. Discovery of false groupings is one of the general purposes of systematic research, but it is a goal the value of which is sometimes disputed. False groupings, even when recognized as such, are sometimes valued if they can be defined by "characters" easily perceived. In table 3.28, for example, there are three "characters" shared by the lamprey and lancelet. If these "characters" are conspicuous, or otherwise easily perceived, they might work well in an identification key, by separating the lancelet and lamprey as a "group" distinct from other species. Hence it is helpful to distinguish groups that seem really to exist from "groups" that may be defined. That a "group" may be defined by one or more "characters" does not mean, therefore, that the group has any existence in the real world.

That all implied cladograms are false, as in the above examples

Table 3.36. Character-Types (A, B, C, D, E1, F1, G) and Their Occurrence in Four Taxa (1–4)

	Character-Types						
Taxa	*A*	*B*	*C*	*D*	*E1*	*F1*	*G*
1	+	–	–	–	+	+	+
2	–	+	–	–	+	+	+
3	–	–	+	–	–	+	+
4	–	–	–	+	–	–	+

(figures 3.35.4–6), suggests that the "solutions" of all the 3-taxon problems are also false (figures 3.35.3 and 3.35.9). Such need not always be the case. Consider the data of table 3.36, which is "unrealistic" in the senses that (1) not all possible character-types for the four taxa (1–4) are represented by some characters, and (2) the character-types represented cannot possibly conflict no matter how the taxa might be combined. The four taxa may be reduced to three by combining any two of them, e.g., taxa 3 and 4 (table 3.37; figure 3.36.1). The problem may be visualized as the placement of taxon 2, with three placements possible (1, 3, 5). A glance at table 3.37 shows that taxon 2 is best placed with taxon 1 (figure 3.36.2: placement 3), with the implications that three 3-taxon cladograms must also be true (figures 3.36.3–4 and 3.36.6). Of the three implied cladograms, two cladograms are true (figures 3.36.3–4), and one cladogram is false (figure 3.36.6). The results are not totally false, for they allow taxa 1 and 2 to be grouped (figure 3.36.7), and combined for a new 3-taxon problem (figure 3.36.8). Data relevant to the new problem (table 3.38) show that taxon 3 may be grouped with taxa 1–2 (figure 3.36.9). Interrelationships of taxa 1–3 are another 3-taxon problem, the data relevant to which (table 3.39) show that taxa 1 and 2 may be grouped together relative to taxon 3 (figure 3.36.10).

Table 3.37. Character-Types (A, B, C, D2, E) and Their Occurrence in Four Taxa (1–4), Two of Which (3 + 4) Have Been Combined (derived from table 3.36)

	Character-Types				
Taxa	*A*	*B*	*C*	*D2*	*E*
1	+	–	–	+	++
2	–	+	–	+	++
3 + 4	–	–	++	–	+++

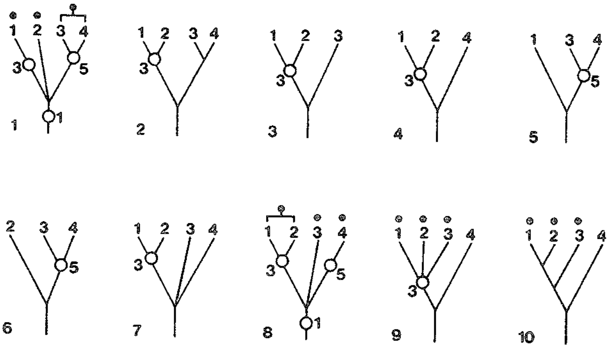


Figure 3.36. Four taxa (1–4), considered with reference to various 3-taxon problems. Circles 1, 3, and 5 are possibilities for placement of a fourth taxon. Black dots specify the three taxa of each 3-taxon problem to be solved (cladograms 1, 8, 9) or the 3-taxon problem solved (cladogram 10).

The minimum and maximum numbers of 3-taxon problems (table 3.33) characterize what may be termed minimum and maximum "modes" of resolving dichotomous cladograms. For six taxa (1–6), an example of resolution in the minimum mode is shown in figure 3.37. The first 3-taxon problem concerns taxa 4, 5, and 6 (shown as solved in figure 3.37.1); the second 3-taxon problem, taxa 3, 4, and 5+6 (solved in figure 3.37.2); the third 3-taxon problem, taxa 2, 3, and 4+5+6 (solved in figure 3.37.3); the fourth 3-taxon problem, taxa 1, 2, and 3+4+5+6 (solved in figure 3.37.4). In this instance, the minimum mode is a unique and stepwise resolution, for it consists of particular 3-taxon problems solved in a particular order. If the final resolution (figure 3.37.4) is correct for the information available, it cannot be reached in four steps other than those of figure 3.37; nor can it be reached in four steps except

Table 3.38. Character-Types (A, B, C, D2, E) and Their Occurrence in Four Taxa (1–4), Two of Which (1 + 2) Have Been Combined (derived from table 3.36)

	Character-Types				
Taxa	*A*	*B*	*C*	*D2*	*E*
1 + 2	++ ++	–	–	++	++
3	–	+	–	+	+
4	–	–	+	–	+

Table 3.39. Character-Types (A, B, C, D2, E) and Their Occurrence in Three Species (1–3; derived from table 3.36)

	Character-Types				
Taxa	*A*	*B*	*C*	*D2*	*E*
1	+	–	–	+	++
2	–	+	–	+	++
3	–	–	+	–	++

as in the order shown. An example of greater-than-minimum resolution is shown in figure 3.38, wherein the first two steps (figures 3.38.1–2) are the same as those of figure 3.37 (3.37.1–2). The third step (figure 3.38.3) differs in concerning taxon 1 rather than taxon 2 (cf. figure 3.37.3). The result is that taxon 2 requires two additional problems (figures 3.38.4–5) for final resolution; and the entire resolution requires five steps instead of four (five 3-taxon problems instead of four).

An example of resolution in the maximum mode is shown in figure 3.39, wherein ten 3-taxon problems are required for resolution of the same dichotomous cladogram (3.39.10; cf. figures 3.37.4 and 3.38.5). The maximum mode is not a unique and stepwise resolution, as is the minimum mode, for the steps toward resolution need not follow in the same sequence. In figure 3.40, for example, the same ten steps occur in a different order.

To determine whether the minimum or maximum mode characterizes a given resolution requires retrospective comparison between the fully resolved cladogram and the various steps toward it. A solved first 3-taxon problem does not in itself indicate either the minimum or maximum mode. A solved second 3-taxon problem will give an indication of the minimum mode (e.g., figures 3.37.2 and 3.38.2) or a greater-than-minimum mode (e.g., figures 3.39.2 and 3.40.2), if the taxa of the first problem are subsumed in the second problem. Taxa 4–6, for

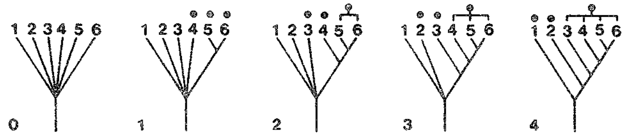


Figure 3.37. A dichotomous cladogram (4) resolved for six taxa (1–6) in the minimum mode, through the solution of four successive 3-taxon problems (shown as solved in cladograms 1–4).

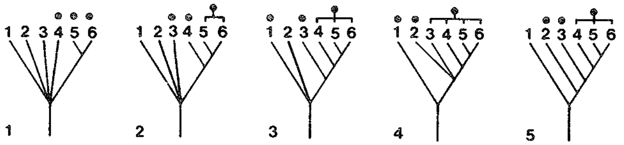


Figure 3.38. The same dichotomous cladogram (5) resolved for six taxa (1–6) in a greater-than-minimum mode, through the solution of five successive 3-taxon problems (shown as solved in cladograms 1–5; cf. figure 3.37).

example, constitute the first problem of figures 3.37 and 3.38, and are subsumed in the second problem of figures 3.37 and 3.38. Similarly, taxa 1–3 constitute the first problem of figures 3.39 and 3.40, and are subsumed in the second problem of figures 3.39 and 3.40.

Resolution of a cladogram in the minimum mode, if such could be done without advance knowledge of the final resolution, would require intuition equivalent in its effect to such knowledge. The role of intuition in systematics, and in science generally, is difficult to assess, mainly because the thrust of scientific investigation is toward nonintuitive analysis, solution, and synthesis of particular research problems. In some general sense, however, intuition might be imagined to play a role in the selection of particular problems to be analyzed, solved, or synthesized. In this sense, resolution in the minimum mode, to the extent that it actually might occur, can be imagined to result from an intuitive sense of the appropriate 3-taxon problem to be solved at the appropriate time within a suite of such problems. With a starting point for a minimum-mode resolution (e.g., figure 3.37.1), what is next

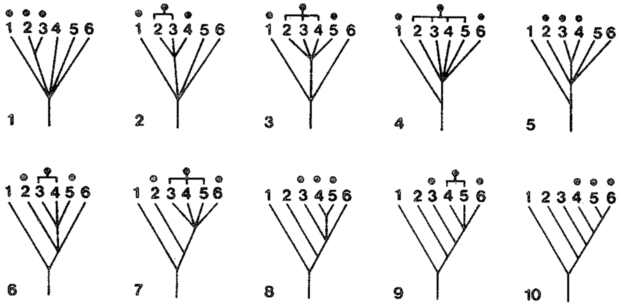


Figure 3.39. The same dichotomous cladogram (10) resolved for six taxa (1–6) in the maximum mode, through the solution of ten successive 3-taxon problems (shown as solved in cladograms 1–10; cf. figures 3.37–3.38).

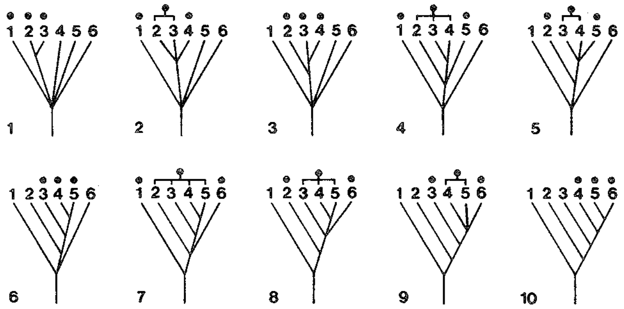


Figure 3.40. The same dichotomous cladogram (10) resolved for six taxa (1–6) in the maximum mode, through the solution of ten successive 3-taxon problems in a different order (shown as solved in cladograms 1–10; cf. figure 3.39).

required is the taxon to be added so as to frame the next 3-taxon problem (e.g., figure 3.37.2: taxon 3), and so on, until the suite of 3-taxon problems gives a final resolution (e.g., figure 3.37.4). In the case of branching of a single taxon at a time (e.g., figure 3.37), the taxon to be added is always single. In a case of more complex branching (e.g., figure 3.41), the taxon to be added will sometimes be two taxa that together form a group of the final resolution (e.g., figure 3.41.3: taxa 1 and 2). Minimum-mode resolution is sufficiently complex so that it probably can never be consistently achieved in practice.

Minimum-mode resolution seems always specifiable retrospectively. For example, consider the final resolution of figure 3.37.4, which is a dichotomous cladogram defined by its five groups, or components: 5,6; 4,5,6; 3,4,5,6; 2,3,4,5,6; 1,2,3,4,5,6. From the components one may derive the suite of four 3-taxon problems for which the informative components are the solutions (figures 3.37.1–4; table 3.40): 4,5,6, with the solution 4(5,6); 3,4(5,6), with the solution 3(4,5,6); 2,3(4,5,6), with the solution 2(3,4,5,6); 1,2(3,4,5,6), with the solution 1(2,3,4,5,6). Consider also the final resolution of figure 3.41.4, which is a dichotomous cladogram defined by five components: 5,6; 4,5,6; 1,2; 1,2,3; 1,2,3,4,5,6. In this case, there are two possible suites of four 3-taxon problems in the minimum mode. One suite (figures 3.41.1–4) is: 4,5,6, with the solution 4(5,6); 3,4(5,6), with the solution 3(4,5,6); 1,2,3, with the solution (1,2)3; (1,2)3(4,5,6), with the solution (1,2,3)(4,5,6). The second suite (figures 3.41.5–8) is: 1,2,3, with the solution (1,2)3; (1,2)3,4,

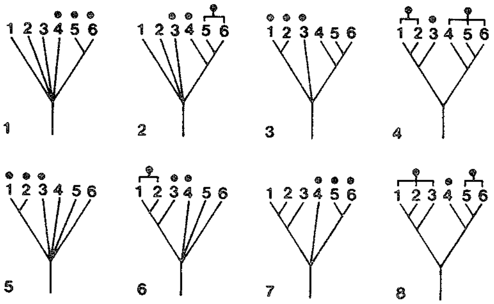


Figure 3.41. Cladograms (4, 8) resolved for six taxa (1–6) in the minimum mode, through the solution of suites of four successive 3-taxon problems (shown as solved in cladograms 1–4 and 5–8).

with the solution (1,2,3)4; 4,5,6, with the solution 4(5,6); (1,2,3)4(5,6), with the solution (1,2,3)(4,5,6).

The minimum mode, then, is a suite of 3-taxon problems that, once solved, result in the informative components of the cladogram that, with respect to a certain sample of information, is the true and final resolution. Thus a cladogram is definable in two different, but related,

Table 3.40. Comparison of Some Cladograms, 3-Taxon Problems and Solutions (minimum mode), and Components of Cladograms

Cladograms	*Problems*	*Solutions*	*Components*
3.37.1	4,5,6	4(5,6)	5,6
3.37.2	3,4(5,6)	3(4,5,6)	4,5,6
3.37.3	2,3(4,5,6)	2(3,4,5,6)	3,4,5,6
3.37.4	1,2(3,4,5,6)	1(2,3,4,5,6)	2,3,4,5,6
3.37.1–4	–	–	1–6
3.41.1	4,5,6	4(5,6)	5,6
3.41.2	3,4(5,6)	3(4,5,6)	4,5,6
3.41.3	1,2,3	(1,2)3	1,2
3.41.4	(1,2)3(4,5,6)	(1,2,3)(4,5,6)	1,2,3
3.41.5	1,2,3	(1,2)3	1,2
3.41.6	(1,2)3,4	(1,2,3)4	1,2,3
3.41.7	4,5,6	4(5,6)	5,6
3.41.8	(1,2,3)4(5,6)	(1,2,3)(4,5,6)	4,5,6
3.41.1–8	–	–	1–6
3.42.1	?	1(2,3,4)	2,3,4
3.42.2	?	1–2(3,4)	3,4
3.42.1–2	–	–	1–4

senses: (1) as a suite of components; (2) as a suite of 3-taxon problems for which the solutions are the suite of components.

For dichotomous cladograms, there is no difficulty of definition in either sense. For nondichotomous cladograms, there are difficulties for the 3-taxon sense of definition. For example, consider the cladogram of figure 3.42.1, defined by two components: 2,3,4 and 1,2,3,4. What is the 3-taxon problem solved? None seems specifiable without reference to a final resolution. In terms of one possible resolution (figure 3.42.3), the problem solved is 1,2(3,4), with the solution 1(2,3,4); in terms of another possible resolution (figure 3.42.4), the problem solved is 1,3(2,4), with the solution 1(2,3,4); in terms of a third possible resolution (figure 3.42.5), the problem solved is 1,4(2,3), with the solution 1(2,3,4). Such being the possibilities, no one of which is specified by figure 3.42.1, the problem solved is unspecifiable except in terms of its solution—1(2,3,4).

Consider the cladogram of figure 3.42.2, defined by two components: 3,4 and 1,2,3,4. What is the 3-taxon problem solved? In this case there are only two possibilities: 1,3,4, with the solution 1(3,4); and 2,3,4, with the solution 2(3,4). Such being the possibilities, neither of which is specified by figure 3.42.2, the problem solved is again unspecifiable except in terms of two possibilities, 1-2,3,4, or their solutions, 1-2(3,4). Thus it would seem that definition of cladograms is not generally possible in terms of 3-taxon problems, but is generally possible in terms of the solutions to such problems. If the solutions and components are compared, it is easy to see that the components are merely abbreviated forms of the solutions, with the addition of one component including all taxa (table 3.40).

Thus, a component may be understood as the solution to a particular 3-taxon problem in the minimum mode; and a cladogram may be understood as the combined solutions to a suite of 3-taxon problems. The information of a cladogram is consequently reflected in the quantity of components that correspond to solved 3-taxon problems. Consider

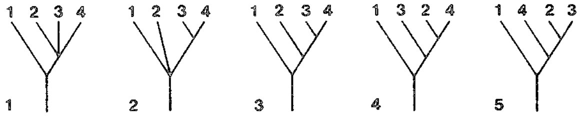


Figure 3.42. Cladograms (1-2) that represent solutions to unspecifiable 3-taxon problems in the minimum mode, with (1) or without (2) reference to final resolutions (3-5).

Table 3.41. Total Components, Informative Components, and Component Information of Some Cladograms (cf. figure 3.41)

Cladograms	*Total Components*	*Informative Components*	*Component Information*
3.41.1	2:1–6; 5,6	1:5,6	1
3.41.2	3:1–6; 4–6; 5,6	2:4–6; 5,6	2
3.41.3	4:1–6; 1,2; 4–6; 5,6	3:1,2; 4–6; 5,6	3
3.41.4	5:1–6; 1–3; 1,2; 4–6; 5,6	4:1–3; 1,2; 4–6; 5,6	4

the cladograms of figure 3.41, with respect to the quantity of their components, and the quantity of components that correspond to solved 3-taxon problems—the component information (table 3.41). The component information is always one less than the total number of components.

The component information is only part of the total information of a cladogram. Another part concerns the taxa, or terms, included in the components. Consider the cladograms of figures 3.42.1–2. Each cladogram has two components (one of which is informative), but the informative components differ in the quantity of taxa, or terms, that they contain, and in the quantity of term information (table 3.42). For reasons mentioned below, the term information is always one less than the total terms of a component.

Whereas the components relate to the number of minimum-mode problems, the terms of a component relate to the number of maximum-mode problems. Cladogram 3.42.1 represents one problem solved in the minimum mode and, at the same time, two problems solved in the maximum mode. Similarly, cladogram 3.42.2 represents one problem solved in the minimum mode, and one problem in the maximum.

Because there is a unit common to both modes (3-taxon problem), the quantities of solved problems of both modes may be summed as a measure of the total information of a cladogram:

Table 3.42. Total Components, Informative Components, Total Terms, and Term Information of Some Cladograms (cf. figure 3.42)

Cladograms	*Total Components*	*Informative Components*	*Total Terms*	*Term Information*
3.42.1	2:1–4; 2–4	1:2–4	3:2,3,4	2
3.42.2	2:1–4; 3,4	1:3,4	2:3,4	1

$$\begin{matrix}\text{Component} \\ \text{information}\end{matrix} + \begin{matrix}\text{Term} \\ \text{information}\end{matrix} = \begin{matrix}\text{Total} \\ \text{information}\end{matrix}$$

The total information is equal also to the sum of the terms of all informative components of a cladogram:

$$\begin{matrix}\text{Terms of} \\ \text{component 1}\end{matrix} + \begin{matrix}\text{Terms of} \\ \text{component 2}\end{matrix} + \cdots \begin{matrix}\text{Terms of} \\ \text{component n}\end{matrix} = \begin{matrix}\text{Total} \\ \text{information}\end{matrix}$$

Considered as some number of 3-taxon problems solved, the total information may be divided by 2, to give the average of minimum- and maximum-mode resolutions:

$$\frac{\text{Total inf}}{2} = \begin{matrix}\text{Average} \\ \text{information}\end{matrix}$$

The component, term, total, and average information of all cladograms of figures 3.37–3.42 are listed in table 3.43.

INCOMPLETE INFORMATION, MULTIPLE BRANCHING, AND RETICULATION

Because secondary cladograms (such as 3.42.1 and 3.42.2) represent solutions to unspecifiable 3-taxon problems, the meaning of the multiple branchings they contain is open to question. An instance of multiple branching in a cladogram may reflect nothing more than ignorance of certain character distributions (those that would be represented by, and allow the resolution of, a primary cladogram). Yet there are situations in which cladograms should exhibit multiple branchings that in some sense reflect real phenomena: for example, character distributions caused by cases of "simultaneous" multiple speciation, instances of hybridization, or groups wherein one species is ancestral to two or more others (as in speciation by the sequential isolation of two or more peripheral populations without change in the central population of a "mother" species, or cases in which studied fossil species are actually the ancestors of other studied species). Multiple branching, therefore, plays an important role in cladistic theory, inasmuch as it is used to represent a variety of character distributions that cannot be represented by a primary (dichotomous) cladogram.

Critics of cladistics have frequently misconstrued it as a theory of

Table 3.43. Component, Term, Total, and Average Information of Some Cladograms (cf. figures 3.37–3.42)

	Information			
Cladogram	*Component*	*Term*	*Total*	*Average*
3.37.1	1	1	2	1.0
3.37.2	2	3	5	2.5
3.37.3	3	6	9	4.5
3.37.4	4	10	14	7.0
3.38.1	1	1	2	1.0
3.38.2	2	3	5	2.5
3.38.3	3	6	9	4.5
3.38.4	3	7	10	5.0
3.38.5	4	10	14	7.0
3.39.1	1	1	2	1.0
3.39.2	1	2	3	1.5
3.39.3	1	3	4	2.0
3.39.4	1	4	5	2.5
3.39.5	2	5	7	3.5
3.39.6	2	6	8	4.0
3.39.7	2	7	9	4.5
3.39.8	3	8	11	5.5
3.39.9	3	9	12	6.0
3.39.10	4	10	14	7.0
3.40.1	1	1	2	1.0
3.40.2	1	2	3	1.5
3.40.3	2	3	5	2.5
3.40.4	2	4	6	3.0
3.40.5	2	5	7	3.5
3.40.6	3	6	9	4.5
3.40.7	3	7	10	5.0
3.40.8	3	8	11	5.5
3.40.9	3	9	12	6.0
3.40.10	4	10	14	7.0
3.41.1	1	1	2	1.0
3.41.2	2	3	5	2.5
3.41.3	3	4	7	3.5
3.41.4	4	6	10	5.0
3.41.5	1	1	2	1.0
3.41.6	2	3	5	2.5
3.41.7	3	4	7	3.5
3.41.8	4	6	10	5.0
3.42.1	1	2	3	1.5
3.42.2	1	1	2	1.0
3.42.3	2	3	5	2.5
3.42.4	2	3	5	2.5
3.42.5	2	3	5	2.5

evolution, rather than a method of systematics. The critics have seized upon the variety of evolutionary events which should produce character distributions that cannot be represented by primary cladograms, and have claimed that these events either falsify cladistics as a theory, or render it useless, for all practical purposes, as a method. In contrast to the critics, we view multiple branching as an essential feature of cladistics—one that is put to a variety of uses.

It is easy to understand how multiple branching of a cladogram can represent character distributions caused by real multiple speciation; all that is necessary is to view the cladogram as a tree (and the lines as evolving lineages). It is almost as easy to understand how multiple branching of a cladogram can represent character distributions caused by an ancestral species that gives rise to two descendant species; all that is necessary is to ask the question (and to realize that it is answerable only in the negative): of the three species, are two of them more closely related to each other than either is to the third? It is less easy to understand how multiple branching can represent character distributions caused by hybridization. As a representation of hybridization, a reticulate pattern (of a tree) seems intuitively meaningful, whereas multiple branching (of a cladogram) seems counterintuitive. Ignorance, finally, is no problem at all if it is complete, but complete ignorance is an ideal seldom achieved in practice. What if ignorance is only partial? If all attempts at generalization about character distribution (by way of a cladogram) reflect partial ignorance, then ignorance is a factor that is perennially present.

Consider cladogram 3.42.1; cladistic interpretation allows for three different primary cladograms (figures 3.42.3–5), each with a component 234. Under the assumption that cladogram 3.42.1 is correct (that component 234 is real), further resolution would seem to be limited to only one of the three primary cladograms.

Consider also cladogram 3.43.1; cladistic interpretation again allows for three different primary cladograms (figures 3.43.2–4), each with a component 34. Under the assumption that cladogram 3.43.1 is correct (that component 34 is real), further resolution would seem to be limited to only one of the three primary cladograms.

The above analysis of possible resolutions, however, is based on the notion (hereafter referred to as interpretation 1) that the information

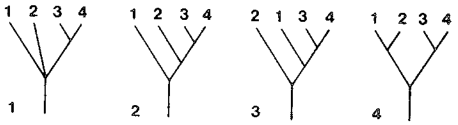


Figure 3.43. Resolution of a secondary cladogram with a basal trichotomy under interpretation 1.

contained in cladogram 3.43.1, for example, is that taxa 3 and 4 are more closely related to each other than either of them is to taxa 1 *and* 2. This would require that two 3-taxon problems had been solved: 1,3,4, with the solution 1(3,4); and 2,3,4, with the solution 2(3,4). Cladogram 3.43.1, however, does not specify that both of these 3-taxon problems have been solved.

There is, therefore, an alternative notion (hereafter referred to as interpretation 2): that the information contained in cladogram 3.43.1, for example, is that taxa 3 and 4 are more closely related to each other than either of them is to taxon 1 *and/or* taxon 2. Under interpretation 2, there are, as before, two possible 3-taxon solutions, only one of which need be true. Each 3-taxon solution (figures 3.44.1 and 3.44.7) allows a fourth taxon to be added at any of five different positions; hence each solution allows a suite of five possible primary cladograms (figures 3.44.2–6 and 3.44.8–12). Three cladograms are common to both suites (figures 3.44.4–6 and 3.44.10–12). If both 3-taxon solutions are true in a given case, they would jointly allow only three primary cladograms (the three cladograms common to both suites, which are the same as those allowed under interpretation 1: figures 3.43.2–4). If only one 3-taxon solution is true, and if it is specified, one (figures 3.44.2–6) or the other

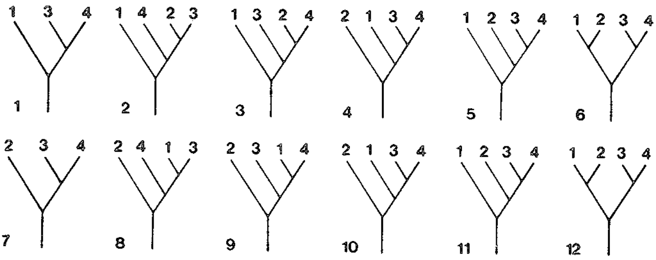


Figure 3.44. Resolution of a secondary cladogram with a basal trichotomy under interpretation 2.

(figures 3.44.8–12) suite of five primary cladograms is allowed. If only one 3-taxon solution is true, and if it is unspecified, a suite of seven primary cladograms is allowed (the three cladograms common to both suites of five, and four other cladograms unique to one suite).

Inasmuch as cladograms may be judged true or false, as the case may be, only on the basis of characters at hand, we may distinguish real and implied information content. Real information in the above sense comprises those 3-taxon solutions that are true on the basis of characters at hand. Implied information comprises those 3-taxon solutions that are derived logically from real information; implied information cannot be determined to be true, except by logical implication, on the basis of characters at hand.

Consider cladogram 3.42.1 and its information in the form of three problems and solutions: 1,2,3, with the solution 1(2,3); 1,2,4, with the solution 1(2,4); 1,3,4, with the solution 1(3,4). Consider the solutions in relation to certain characters (table 3.44): with reference to character A (considered by itself), all three solutions are true (and real information); with reference to characters B–D (considered by themselves), all three solutions are true (and real information); with reference to characters E–F (considered by themselves), only two solutions are true (and real); the third solution is only implied.

Cladogram 3.42.1 might serve as a preferred summary of each of the samples of characters (characters A; B,C,D; E,F), but the real information of the cladogram is lower for one sample (characters E,F). In the case of that sample, cladogram 3.42.1 could serve as a summary under interpretation 1, but only an appeal to logical necessity (rather than characters at hand) could justify the truth of one solution: 1(2,4).

Consider cladogram 3.43.1 and its information in the form of two problems and solutions: 1,3,4, with the solution 1(3,4); 2,3,4, with the solution 2(3,4). Consider the solutions in relation to certain characters

Table 3.44. Occurrence of Some Characters (A–I) in Four Taxa (1–4)

Taxa	*Characters*								
	A;	*B*	*C*	*D;*	*E*	*F;*	*G;*	*H;*	*I;*
1	–;	–	–	–;	–	–;	–;	–;	?;
2	+;	–	+	+;	–	+;	–;	?;	–;
3	+;	+	+	–;	+	+;	+;	+;	+;
4	+;	+	–	+;	+	–;	+;	+;	+;

(table 3.44): with reference to character G (considered by itself), both solutions are true (and real information); with reference to character H (considered by itself), only one solution is true (the other is not even implied); character I gives the same result as character H.

Cladogram 3.43.1 might serve as a preferred summary for each sample (characters G; H; I), but the information of the cladogram is lower for two samples (H; I). For those samples, cladogram 3.43.1 can serve as a summary only under interpretation 2, for there is no basis for an appeal to logical necessity as a justification for the truth of one solution in each case: 2(3,4) for character H; 1(3,4) for character I. Secondary cladograms incorporating basal trichotomies (or multiple branchings) thus differ from those incorporating terminal trichotomies (or multiple branchings) in that they are subject to two different interpretations reflecting varying degrees of completeness in the information they summarize. Basal trichotomies (or multiple branchings), under interpretation 2, can accommodate taxa for which available characters are inadequate to allow their placement on the cladogram under interpretation 1.

The two interpretations can be compared with reference to the primary cladograms that they allow, or, alternatively, that they prohibit. Interpretation 1 (figure 3.43) permits three, and therefore prohibits 12, of the 15 possible dichotomous cladograms for four taxa. Interpretation 2 (figure 3.44) permits seven, and therefore prohibits eight, of the 15 possible dichotomous cladograms for four taxa. Thus, interpretation 2 is less prohibitive, or less restrictive, than interpretation 1. Both interpretations allow the group 34 to be real, but the interpretations differ in their specifications of the limits of the group's reality. Under interpretation 1, neither 1 nor 2 can be a member of the group 34; under interpretation 2, either 1, or 2 (but not both) can be a member of the group 34.

Consider a hypothetical group of three species (1,3,4) whose relationships have been established, for example in the pattern specified by cladogram 3.44.1: solution 1(3,4). Suppose that a fourth species (2) is discovered, that species 2 is determined to be a member of the group already including species 1, 3, and 4, and that the precise relationships of species 2 can be determined with no further degree of accuracy. How may species 2 be added to the cladogram (3.44.1)?

There are two possibilities: under interpretation 1, the only possible placement is the tertiary (unresolved) cladogram (figure 3.9.1); under interpretation 2, the only possible placement is shown in figure 3.43.1. The latter placement (figure 3.43.1) would not be possible under interpretation 1, because that interpretation prohibits dichotomous cladograms (figures 3.44.2–3) that cannot reasonably be prohibited in the absence of evidence pertaining to the precise relationships of species 2 within the group. To arrive at cladogram 3.43.1 under interpretation 1, we would need to know not only that species 2 is a member of the group already including species 1, 3, and 4, but also that species 3 and 4 are more closely related to each other than either is to species 2 (i.e., that cladograms 3.44.2 and 3.44.3 are false). In any particular case, such information may or may not be available.

Species, or taxa generally, which can be placed in a higher taxon, but whose relationships are otherwise obscure, are commonplace. Their occurrence accounts for some of the trichotomies (and multiple branchings) found in the cladogram of any large group. If systematic practice operated exclusively under interpretation 1, any such species (or taxon generally) would effectively collapse the internal cladistic structure of the group to which the species (or taxon) was assigned. An extreme example would be a species that could be recognized as living but whose relationships were otherwise unspecifiable. Under interpretation 1, all cladistic structure would collapse into a basal branching as numerous as there are species.

Interpretation 1 is not universally adopted in systematics (except perhaps in previous theoretical discussion) simply because it is sometimes unworkable. Interpretation 2 seems sometimes to be adopted in systematics; at least it workably merges with, or is implicit within, routine taxonomic practice. If so, then the problem of multiple branching can be considered in a somewhat different light. In a given case one may ask: under which interpretation (1 or 2) is a trichotomy (or multiple branching) to be understood? If under interpretation 2, as occurs sometimes with recent species and perhaps more commonly with fossils, then the trichotomy (or multiple branching) is not a final solution, but rather a problem that, until solved, injects considerable ambiguity into the cladogram. Progress in the taxonomy of a given group may thus involve a gradual shift from interpretation 2 to

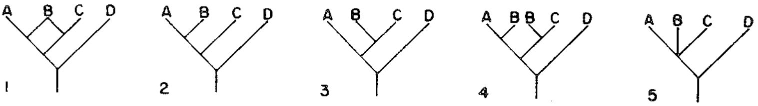


Figure 3.45. Resolution of a terminal hybridization.

interpretation 1 (and finally to a primary cladogram) as additional 3-taxon problems are solved.

Hybridization is sometimes viewed as a problem for cladistics, but hybridization can also be considered in the light of interpretations 1 and 2. A simple case of hybridization involves four species (A, B, C, D), two of which (A and C) hybridize and produce a third species (B; figure 3.45.1). Study of the cladistic relationships of the four species should reveal two conflicting patterns (figures 3.45.2–3), combinable in one branching diagram with reticulation (figure 3.45.1) or in a branching diagram wherein species B (the hybrid) appears twice (figure 3.45.4). The only possible cladogram (without reticulation or repetition of taxa) is one with a terminal trichotomy (figure 3.45.5).

Another simple instance of hybridization involves four species (A, B, C, D), two of which (B and D) hybridize and produce a third species (C; figure 3.46.1). Study of the cladistic relationships of the four species should reveal two conflicting patterns (figures 3.46.2–3), combinable in one branching diagram with reticulation (figure 3.46.1) or in a branching diagram wherein species C (the hybrid) appears twice (figure 3.46.4). Under interpretation 2, the only possible cladogram (without reticulation or repetition of taxa) is one with a basal trichotomy (figure 3.46.5). Under interpretation 1, the cladistic structure collapses (figure 3.9.1) into a tertiary cladogram. In this context, interpretation 2 operates to produce the "consensus tree" (Adams 1973) representing only that information contained in both of the two conflicting patterns (figures 3.46.2–3).

Hybrid species, or taxa generally, presumably occur. Their occurrence may account for some of the trichotomies (and multiple branchings) found in the cladogram of any large group. If systematic practice

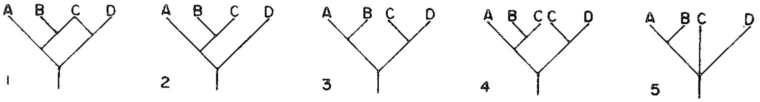


Figure 3.46. Resolution of a basal hybridization.

operated exclusively under interpretation 1, any such hybrid species (or hybrid taxon generally) would effectively collapse the internal cladistic structure of the groups involved in the hybridization. An extreme example would be a species produced by hybridization of the first two primordial species. Under interpretation 1, all cladistic structure would effectively collapse into a basal branching as numerous as there are species.

Multiple speciation, hybridization, and groups including actual ancestors seem cladistically indistinguishable from each other, and from simple ignorance of interrelationships, in that multiple speciation, hybridization, actual ancestors, and ignorance are all represented in cladograms in the same way: by trichotomies (or multiple branchings, be they terminal or basal). In addition, basal trichotomies (or multiple branchings), under interpretation 2, also represent partial ignorance, of whatever degree, of species interrelationships. That the limits of cladistics (Hull 1980) are thereby defined is advantageous in the sense that the discipline of cladistics is rendered intelligible. But what of multiple speciation, hybridization, and actual ancestors? They seem to belong to the suite of problems that arise from a consideration of trees rather than cladograms. If cladistics is that part of systematics concerned with cladograms, then perhaps it is time to speak of "arboristics" as that part concerned with trees and, specifically, modes of speciation in general, as well as particular histories of speciation. One might conceive of an "arboristic analysis" which attempts to determine what tree is the cause of a particular instance of trichotomy or multiple branching, and operates by investigating the particular character distributions found in a given instance and their relative compatibility with various evolutionary scenarios. The results of such an analysis, however, would be of questionable significance for classification if hierarchical classifications convey only the information contained in cladograms, and not the additional information contained in trees.