Impact of outgroup inclusion on estimates by parsimony of undirected branching of ingroup phylogenetic lines

Bernard R. Baum[1] & George F. Estabrook[2]

Summary

Baum, B. R. & Estabrook, G. F: Impact of outgroup inclusion on estimates by parsimony of undirected branching of ingroup phylogenetic lines. – Taxon 45: 243-257. 1996. – ISSN 0040-0262.

Some of the effects of including outgroup taxa on the branching pattern of the ingroup taxa are revealed by an artificial example, and illustrated with a more complex natural example involving the grass genus *Kengyilia*. To reduce the total number of changes required by a branching pattern, parsimony may prefer to reduce the number of changes on temporally long phyletic lines to more distant outgroups while increasing changes on temporally short phyletic lines within the ingroup. This may have the effect of bringing the ancestors that define monophyletic groups within the ingroup down to the phyletic line to the outgroup, which converts them to paraphyletic groups. Thus the inclusion of outgroup taxa during parsimony may alter the branching pattern of the ingroup to destroy distinct monophyletic groups and instead create nested series of monophyletic groups, reminiscent of the chaining properties of some phenetic methods. If similarities between ingroup and outgroup are true homoplasies, then removal of these homoplasies will produce error. However, parsimony estimates of ingroup may produce bogus monophyletic groups because no account has been made of the ingroup's most recent common ancestor.

Introduction

Much has been written and debated about the merits of using outgroups to estimate direction of evolutionary trends (Lundberg, 1972; Watrous & Wheeler, 1981; Stevens, 1981; Meacham, 1984, 1986; Donoghue & Maddison, 1986) and to estimate ancestral states (Crisci & Stuessy, 1980; Stevens, 1980; Nixon & Carpenter, 1993 provide a recent review). Here we discuss the effects of including outgroups on the estimates by parsimony of the branching of phylogenetic lines within the ingroup. Including outgroups during parsimony not only estimates the direction of evolutionary change but may also change the estimated branching patterns of the phylogenetic lines. We reveal the nature of some of the effects of the inclusion of outgroups, illustrate how the mechanism might work, and discuss what this might mean for students of natural history concerned with the reconstruction of phylogeny.

There are basically two approaches to the use of outgroups in making parsimony estimates of phylogenetic branching pattern. One approach (the "outgroup after" method, Lundberg, 1972) first makes an undirected estimate of the branching pattern of the phylogenetic lines of just the in-group and then somehow uses information about the outgroup to estimate the most primitive place on this undirected tree to give it direction. The other approach (the "outgroup during" method) first includes one (or more or a composite or a primitive) member(s) of the outgroup with the ingroup and then makes an undirected estimate of the branching pattern of the phylogenetic lines of this larger group. The most primitive place on this undirected tree is taken to

---

[1] Centre for Land and biological Resources Research, Agriculture Canada, K. W. Neatby Building, Central Experimental Farm, Ottawa, Ontario, K1A 0C6, Canada.

[2] University of Michigan Herbarium, Ann Arbor, MI 481091057, U.S.A.

be on the line that connects the outgroup member(s) with the rest of the tree (Dono-ghue & Maddison, 1986).

Here we first compare the inclusion of outgroups before making an undirected parsimony estimate of the branching pattern of phylogenetic lines, with the undi-rected parsimony estimate that is made when only members of the ingroup are included, and examine whether the resulting estimated branching pattern for the ingroup is the same in both cases. To do this, we first present a simple artificial example that illustrates some of the mechanisms, and then a more complex but natural example that enables argument and speculation to be made, based on specific real data and context. When inclusion of an outgroup results in a different estimate of the undirected branching pattern of the phylogenetic lines of the ingroup, under what conditions is one, or the other, or neither of these estimates more credible? How could outgroup information be used to estimate direction of evolutionary change or primitive states?

A simple contrived example

This example is artificial and hypothetical. It serves to illustrate the concepts and the consequences of applying methods. We cannot use it to find out whether methods produce historically correct estimates of relationships. Bearing this in mind, we will still speak of hypothetical historical realities and discuss their consequences. With a clearer understanding of how methods work, we can better choose those most appro-priate to our own work. The example is of an ingroup with 5 species, *a, b, c, d, e* in one genus, and two species *g, h* in another (sister?) genus (the outgroup). The data are shown in Table 1, where 0 and 1 simply indicate that species in the same state for a character, by that character cannot be distinguished from each other but from species in the other state; no presumptions of presence, absence, primitive, advanced, etc. are made. The less common state has been arbitrarily designated with 1, and the more common with 0. Parsimony trees were constructed using the branch-and-bound method (Hendy & Penny, 1982) using the program PENNY in PHYLIP version 3.5c (Felsenstein, 1993), so all parsimonious trees are always found.

There are 20 characters. Six distinguish outgroup species from ingroup species. Two distinguish *a, b, c* from the rest. Species *d* and *e* are identical except for two "uninformative" autapomorphic characters that distinguish them from each other. Ten characters depict conflicting relationships among the sub-ingroup *(a, b, c)* and some also suggest various conflicting (spurious?) relationships with one or both of the distant outgroup species *(g, h)*.

There are three groups of species: the virtually identical *(d, e)*, sharing the same state for 18 among the 20 characters and differing only for their two "uninformative" autapomorphies; the unresolved *(a, b, c)* pair-wise sharing states for 13 or 14 of the 20 characters; and the outgroup *(g, h)* sharing 12 states. Pairs of species with one in each of the two sub-ingroups *(d, e)* and *(a, b, c)* share states for 10 or 11 characters. Species of the outgroup share states with *(a, b, c)* for variously 4, 5, or 7 characters and with either of *d* or *e* for 8 characters. If the ingroup and the outgroup are sister genera, then an intuitively plausible directed tree is shown in Fig. 1. It has a long phyletic line extending from the outgroup *(g, h)* to some point, within the ingroup, that lies on a phyletic line segment between *(a, b, c)* and *(d, e)*. Consideration of

Table 1. State matrix for 7 hypothetical species *(a-e, g-h)* and 20 characters with states 0 or 1.

a	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0
b	1	1	1	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0
c	1	1	1	1	1	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0
d	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
e	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
g	1	1	1	0	0	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0
h	0	0	0	1	1	1	0	0	0	1	0	0	1	1	1	1	1	1	0	0

states shared might suggest that point lies somewhat closer to *(d, e)* than to *(a, b, c)*. Except for *(a, b, c)*, the tree of Fig. 1 is resolved. It has not been explicitly constructed by a method, only presented as a (the?) intuitively reasonable tree suggested by "eye balling" the data. Below we examine trees constructed from the data by parsimony, using outgroup species to estimate the direction of evolutionary change within the ingroup. We will be especially interested to see how *(a, b, c)* are resolved, and whether the structure of the intuitive tree of Fig. 1 is preserved.

With the "outgroup after" method, ingroup branching pattern is estimated by parsimony applied only to the ingroup species. The ingroup branching pattern of Fig. 2 is uniquely chosen because four characters support the potential monophyly of *(a, b)* and only three characters each support the potential monophyly of *(a, c)* or *(b, c)*. The ingroup branching pattern is left intact while a place to attach the phyletic line leading to the outgroup is somehow determined.

This determination is often somewhat intuitive or informal. In our case twelve characters are invariant in the outgroup, and thus clearly suggest an outgroup based primitive state. Of these twelve only six vary within the ingroup. Two of these six are the autapomorphisms of *d* and *e*. Of the four that remain, two support the attachment of the outgroup between *(a, b, c)* and *(d, e)* or between d and e, one supports the attachment between *(a, b)* and *(c, d, e)* or among *(c, d, e)*, and one supports the attachment between *(a, b)* and *(c, d, e)* or between *a* and *b*. Thus these four characters are all consistent only with the "outgroup after" tree of Fig. 2.

The discomfort that some workers may have felt with the somewhat arbitrary process of attaching an outgroup to an extant branching pattern may have contributed to the practice of including a representative outgroup species with the ingroup species during the process of making parsimony estimates of the branching pattern for the ingroup. In this way an outgroup species would already be attached to the ingroup branching by a phyletic line on which could be assumed to lie the most primitive point, which would direct the tree. This is the advocacy of Nixon & Carpenter (1993).

If outgroup species *g* were gratuitously chosen to represent the outgroup for an "outgroup during" parsimony analysis, then the unique directed tree of Fig. 3 would result. By contrast, if outgroup species *h* happened to be chosen to represent the outgroup for an "outgroup during" parsimony analysis, then the unique directed tree of Fig. 4 would result.

Notice that the inclusion of an outgroup species during the parsimony estimation of the branching pattern of the ingroup can actually change the branching pattern estimated for the ingroup while giving the same estimate of the direction of evolu-
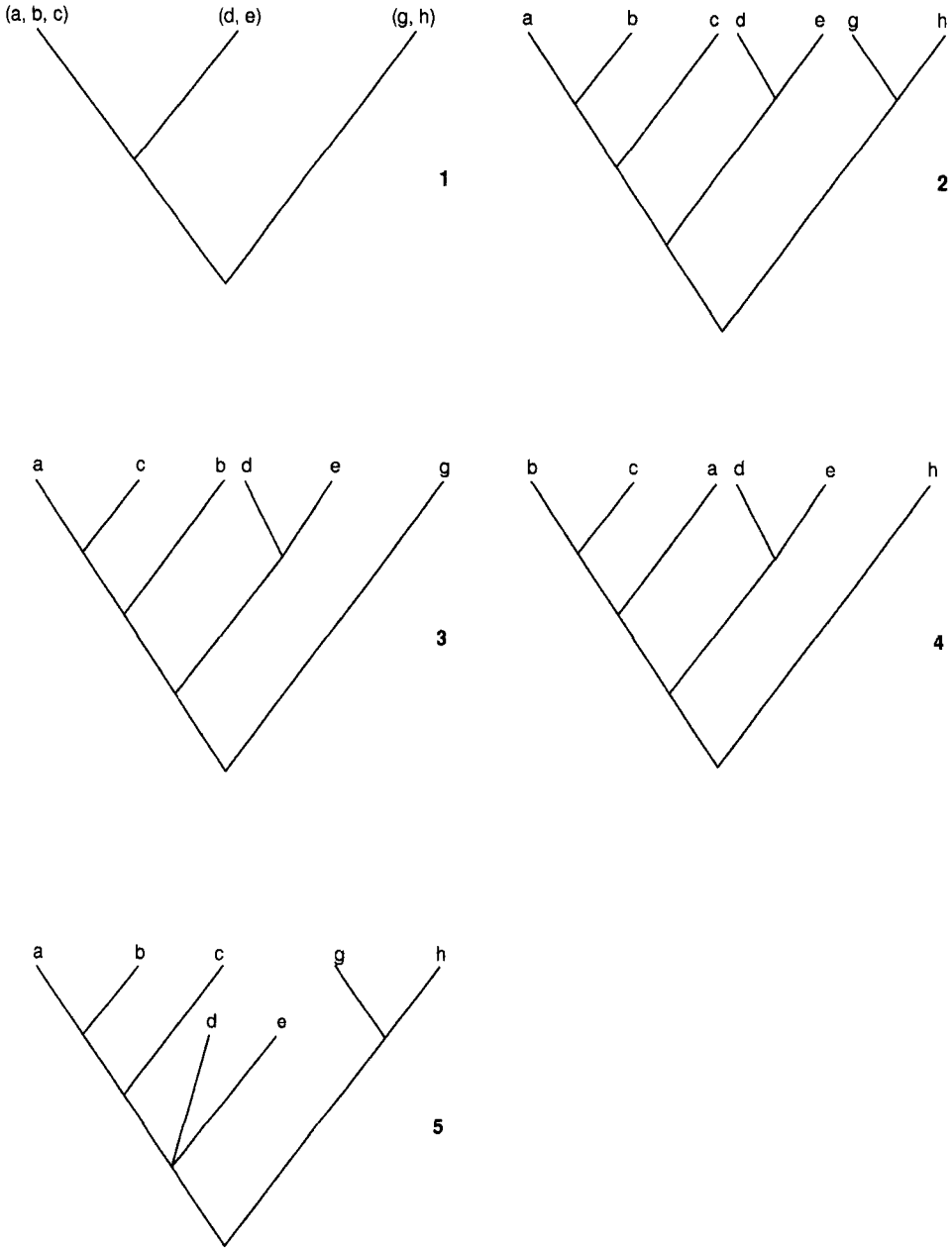
Fig. 1-5. Cladograms involving 7 hypothetical species. – 1: intuitive outgroup-directed unresolved tree; 2: single most parsimonious ingroup tree, plausibly directed by "outgroup after"; 3: single most parsimonious "outgroup *g* during" tree; 4: single most parsimonious "outgroup *h* during" tree; 5: unresolved tree, one of the four equally most parsimonious "outgroups *g* and *h* sharing" trees along with the trees of Fig. 2-4.

tionary change. Furthermore, inclusion of a different representative outgroup species can change the estimated ingroup branching pattern in yet a different way. In "outgroup during", a representative outgroup species can go beyond its role to estimate direction of evolutionary change, and actively participate in the determination of ingroup branching.

If both *g* and *h* are included with the ingroup for an "outgroup during" parsimony analysis, then the result is four equally most parsimonious trees: the three we have seen in Fig. 2, 3, and 4 (with both *g* and *h* included, or with *h* or *g* omitted in Fig. 3 and 4), plus the tree of Fig. 5. This equally parsimonious tree is partially unresolved. It does not show the monophyly of *(d, e)* shown in Fig. 1-4. Although *d* and *e* are identical except for one autapomorphy each, consistent with this is the possibility that the most recent common ancestor of *(d, e)* lie on the phyletic line from *(a, b, c)* to the sister genus outgroup, as shown in Fig. 5. This possibility would become more likely if the sharing of character states by some of the distantly related outgroup species with some but not all ingroup species were considered to be most likely explained homoplastically. In Fig. 3 and 4, it is always more parsimonious to carry these shared states without change over the long phyletic line between genera, creating somewhat artificial homoplastic synapomorphies for *(d, e)* that never occur in the data as exclusively shared character states.

The differences, including resolutions or dissolutions, between the various "outgroup during" parsimony estimates of ingroup branching and the "outgroup after" parsimony estimate of ingroup branching result from parsimony seeking to minimize character state changes between the ingroup and the outgroup with weight equal to minimizing character state changes within these two groups. If we assume that both the ingroup and the outgroup are monophyletic in the narrow sense, then the most recent ancestor common to both groups is distinct from and "primitive" to the (distinct) most recent common ancestors of each of the two groups. Therefore, evolution of change with which to compare species from separate groups has been going on for longer than evolution of change with which to compare species within the same group. We would expect more change between groups than within groups; we expect distant relatives to be more distant than close relatives. When distant relatives are included in a parsimony analysis, parsimony will seek to minimize global parsimony, which can result in different estimates of branching pattern that allow distant relatives to be closer and close relatives to be more distant if by so doing global parsimony is minimized. When this happens in "outgroup during" analyses, the outgroup serves not only to estimate direction of change within the ingroup, but also participates in the estimation of branching pattern within the ingroup.

Change of ingroup branching pattern as a result of including outgroup representatives during parsimony estimations is more likely to occur when there are large numbers of incompatible character pairs within the ingroup, because more incompatibility often indicates a larger number of conflicts to resolve and a larger number of possible ways to resolve them, whether or not distant relatives are included. When included, outgroup representatives can never decrease, only increase or leave unchanged, the amount of character incompatibility to be resolved. When they increase character incompatibility, in some cases new more parsimonious resolutions with different ingroup branching patterns are made possible.

The inclusion of distant relatives can also produce other effects that may or may not be desirable. In this example, parsimony "distinguished" the nearly identical *d*

Table 2. Data matrix of 23 characters of the 15 species of *Kengyilia* C. Yen & J. L. Yang and the outgroup genera (see the Appendix for character coding).

	1	2	3	4	5	6	7	8	9	1 0	1 1	1 2	1 3	1 4	1 5	1 6	1 7	1 8	1 9	2 0	2 1	2 2	2 3
K. alaica (Drobow) J. L. Yang & al.	2	2	0	0	0	0	0	0	0	0	2	2	0	0	1	2	0	0	0	1	0	5	0
K. gobicola C. Yen & J. L. Yang	0	1	4	4	3	3	0	3	0	8	2	3	2	3	0	0	0	1	0	0	0	5	2
K. guidenensis J. L. Yang & al.	0	1	1	0	5	3	0	2	0	0	2	2	6	1	1	1	2	1	0	0	2	0	0
K. habahenensis B. R. Baum & al.	1	0	2	0	8	1	0	0	0	3	2	2	2	4	1	0	3	3	0	0	0	6	9
K. hirsuta (Keng & S. L. Chen) J. L. Yang & al.	1	1	1	0	1	1	3	1	0	8	2	7	3	2	1	2	2	4	0	3	0	0	4
K. kokonorica (Keng & S. L. Chen) J. L. Yang & al.	0	2	1	0	6	8	0	2	0	2	2	2	2	5	0	1	5	4	0	0	2	0	5
K. laxiflora (Keng & S. L. Chen) J. L. Yang & al.	0	2	2	0	8	3	1	2	1	4	2	2	5	3	1	0	1	5	0	0	3	5	0
K. melanthera (Keng & S. L. Chen) J. L. Yang & al.	1	3	2	0	5	1	0	0	0	4	0	6	4	0	1	0	2	0	0	0	1	2	6
K. mutica (Keng & S. L. Chen) J. L. Yang & al.	0	1	1	0	7	1	0	2	0	5	2	4	3	2	0	0	2	0	1	0	4	1	7
K. nana J. L. Yang & al.	0	2	4	4	9	3	0	0	0	0	2	5	6	2	1	1	1	0	0	0	1	2	8
K. pamirica J. L. Yang & C. Yen	0	2	4	0	8	3	0	3	0	8	1	5	6	3	0	1	3	1	0	0	2	2	2
K. rigidula (Keng & S. L. Chen) J. L. Yang & al.	1	3	2	0	8	7	2	1	2	6	2	2	5	5	1	0	1	4	0	0	4	0	7
K. stenachyra (Keng & S. L. Chen) J. L. Yang & al.	2	3	2	0	8	6	0	3	0	6	2	6	3	7	1	1	2	0	0	0	1	5	6
K. tahelacana J. L. Yang & al.	2	1	1	0	9	8	0	3	1	5	2	3	2	3	0	1	1	7	0	0	4	4	0
K. zhaosuenensis J. L. Yang & al.	1	1	2	0	5	8	0	3	2	1	2	3	7	3	1	2	2	1	0	0	1	5	9
Roegneria K. Koch	1	1	2	0	4	4	2	1	0	4	1	4	4	3	0	1	2	4	0	0	2	3	5
Agropyron Gaertn.	1	2	2	2	5	4	2	2	0	4	2	3	3	6	1	0	3	4	0	0	2	7	0

and *e* from the ancestral line between ingroup and outgroup in all but one case, because some states, which occur in some species from the ingroup and some from the outgroup, remained unchanged through the ancestral line from ingroup to outgroup, which then created homoplastic "synapomorphies" for *(d, e)*.

Inclusion of outgroup members during parsimony may be more reasonable when the outgroup is not clearly distinctly and separately evolved from the ingroup. Here, unrelated parallel evolution in the two groups may be a less plausible explanation for characters incompatible with the ingroup-outgroup distinction. In this case, inclusion of the outgroup within the ingroup can be construed as simply enlarging the ingroup for a more complete ingroup analysis. In cases where the outgroup is clearly distant, it might seem more reasonable to assume that incompatible state sharing between distant relatives is unrelated to resolving ingroup branching pattern, and so such relatives should not participate explicitly in estimation of ingroup branching pattern. Which of these cases is most reasonable requires a scientific judgement that is specific to the biology of the groups studied, and the environmental context in which the groups evolved. It is best made by a scientist who understands both his organisms and his methods for drawing inferences about them.

A natural example

The genus *Kengyilia (Gramineae)* consists of 20 known species. The genera most closely related to it are *Roegeneria* and *Agropyron*. Relatedness was established mainly by chromosome similarities in karyotypes, conventionally called genomic similarities. For instance, *Kengyilia* and *Agropyron* share the P genome, and *Kengy-*

ilia and *Roegneria* share the S genome (Baum & al., 1991). Thus both genera are obvious candidates for sister groups in a cladistic analysis. In addition, some morphological traits of each of the sister groups are shared with *Kengyilia*.

The species of *Kengyilia* occur in Central and W. China and adjacent countries (Kazakhstan, Tadzhikistan, Nepal, and Sikkim). The species are rare but some may be locally abundant. From the 20 species described, the 15 species selected for this example were those with no missing values. The genus *Kengyilia* is currently under investigation, but a key was provided for the identification of the 16 species known at that time (Yang & al., 1992). Each outgroup is represented by a single entity (terminal taxon) with character states common or central in the genus. The characters are presented in the Appendix. The data matrix is given in Table 2. The characters for both *Roegneria* and *Agropyron* were compiled from averaging the states present in most of the species in *Roegneria* and all of the species in *Agropyron*. The delimitation of *Roegneria, Agropyron,* and *Kengyilia* has been presented in Baum & al. (1995), where a key is provided for their identification, and their distiction from *Elymus* s. str.

The following analyses were carried out using HENNIG86© version 1.5 (Farris, 1988). Four runs were executed: the first with the fifteen ingroup species and the two outgroups, i.e. seventeen terminal taxa; the second with *Roegneria* only as outgroup, and the third with *Agropyron* only as outgroup, i.e. sixteen terminal taxa; the fourth with the ingroup species alone, i.e. fifteen terminal taxa. For each run the characters were declared unordered, i.e. non-additive, and character states were allowed to change in any direction. Calculation of trees was done as follows for each run: mh*, i.e. finding the trees with branch swapping to each initial tree while retaining no more than one tree for each initial one; then bb*, i.e. generating all possible maximally parsimonious trees starting with the results obtained from mh* as a basis. In addition, strict (Nelsen option in HENNIG86) and majority rule (using NTSYS-PC version 1.7; Rohlf, 1992) consensus trees were used as an aid for inferring the results.

Data analysis resulted in 3 equally most parsimonious trees from the run with the two outgroups (Fig. 6-8). The two outgroups always clearly grouped together as a sister group to the ingroup species. The strict consensus tree is shown in Fig. 9. With *Roegneria* as the single outgroup, the result was a single most parsimonious tree (Fig. 10). With *Agropyron* as the single outgroup, the analysis yielded three most parsimonious trees (Fig. 11-13) with relationships different from those obtained with *Roegneria* as outgroup. The strict consensus tree is shown in Fig. 14. The ingroup analysis yielded a single most parsimonious tree (Fig. 15). A simplified summary of the common elements in the trees is presented in Fig. 16.

Based on these results it is clear from Fig. 16 that including *Agropyron* as outgroup, either alone or with *Roegneria,* changes the undirected branching of the ingroup phylogenetic lines in two places. *Kengyilia guidenensis* is being described by Yen & al. (1996) as a new species apparently close to *K. kokonorica*. These two would be a monophyletic group if the ingroup phylogenetic lines were directed away from any region between them. *K. habahenensis* and *K. alaica* are also close and would be monophyletic by any directing of the ingroup branching pattern that did not originate between them. Each species pair makes a plausible hypothesis of monophyly supported by the ingroup branching pattern. Inclusion of *Agropyron,* with or without *Roegneria,* in the parsimony estimate of branching brings the ancestors of these two hypothesized monophyletic groups down onto the long phyletic lines lead-
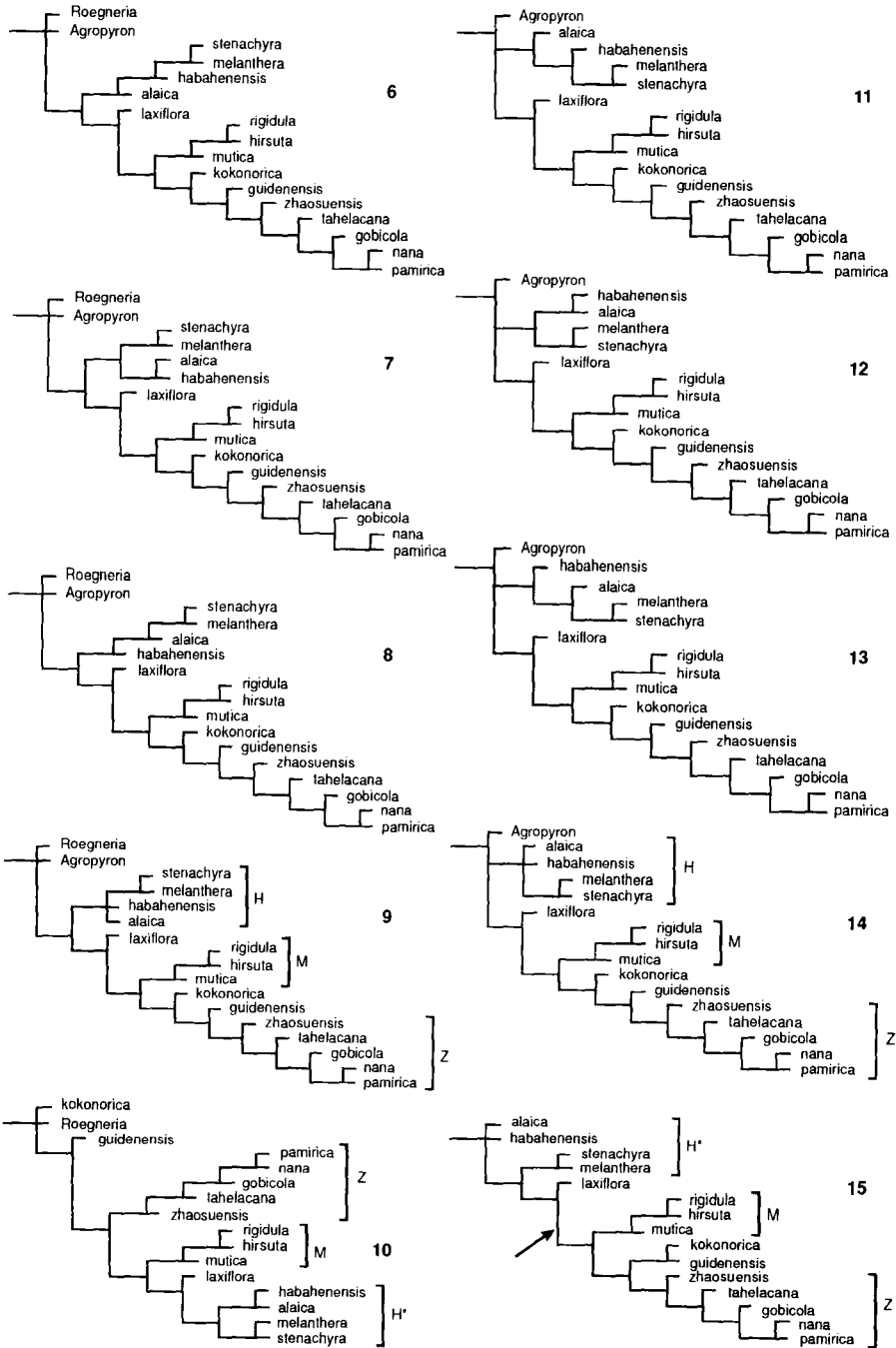
Fig. 6-15. Most parsimonious trees obtained for *Kengyilia*, using the data of Table 2. – 6-9: with both *Roegneria* and *Agropyron* as "outgroups during"; 10: with only *Roegneria* as "outgroup during"; 11-14: with only *Agropyron* as "outgroup during"; 15: with neither outgroup. In the single most parsimonious trees (10, 15) and consensus trees (9, 14), consistently recurrent cladogram branches (species groups) are lettered H (or H*), M, and Z (see Fig. 16).

ing from the outgroup past them to more distantly related species. With outgroups included, it became more parsimonious to keep unchanged the character states shared by *Agropyron* and subtree Z, move homoplasy into the evolution of the closely related *K. kokonorica* and *K. guidenensis,* and destroy their potential monophyly. Similarly, it became more parsimonious to keep unchanged the character states shared by *Agropyron* and the ancestor of *K. stenachyra* and *K. melanthera* in subtree H*, move the homoplasy into the evolution of the closely related *K. alaica* and *K. habahenensis,* and destroy their potential monophyly.

Consider possible natural histories of *Kengyilia.* The genus, which is homogeneous from the genomic point of view, has evolved from the genetic contributions of an *Agropyron*-like and a *Roegneria*-like ancestor. During the course of evolution a number of species retained characteristics closest to *Agropyron,* or their characteristics evolved to become more similar to those of species in *Agropyron.* The other species of *Kengyilia* similarly evolved with traits more *Roegneria*-like. Another possibility is that the genus is diphyletic in the sense that it originated once from an *Agropyron*-like and once from a *Roegneria*-like ancestor, at a time when the two did not differ so much as they do today. Other scenarios are possible, such as multiple hybrid origin involving different parental taxa. Work based on DNA sequencing of these taxa is progressing and will hopefully shed light on the origin and relationships among the various species of *Kengyilia* and on their bonds with *Roegneria* and *Agropyron.*
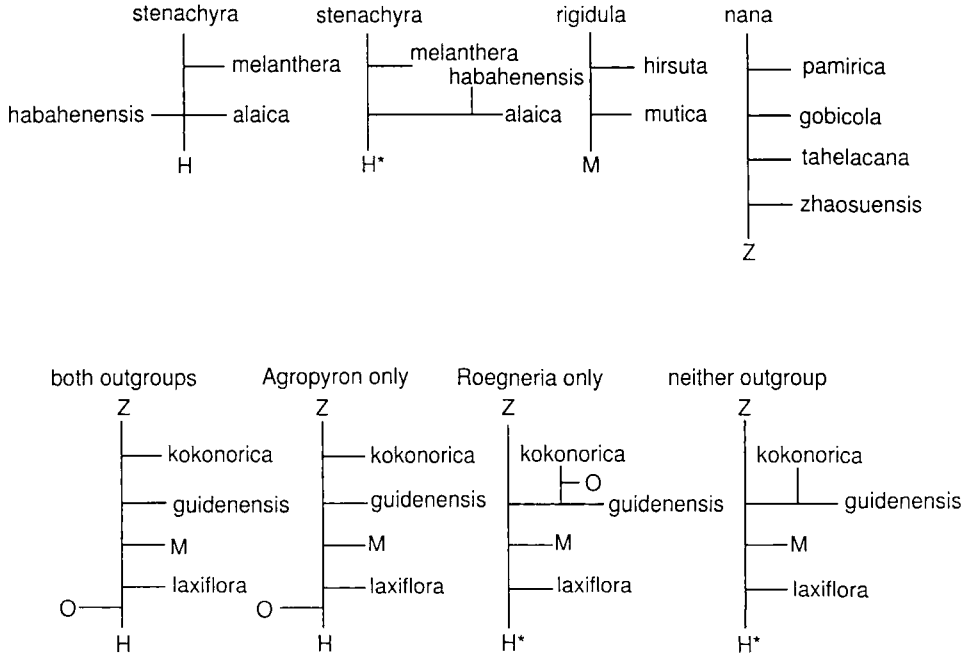

Fig. 16. Summary of Fig. 6-15. Constant pieces of the preceding trees are represented by single letters (as in Fig. 9-10, 14-15). Group H represents all 4 resolutions of the trichotomy shown; Group H* represents just the one shown, as do groups M and Z. Resolutions of H proliferate the trees of Fig. 6-9 and 11-14; when they are aggregated, the variation in the four outgroup cases can be easily seen. O = place of attachment of the outgroup(s).

There seem to be two groups in the genus: species below the arrow in Fig. 15 are more like *Roegneria,* and species above this arrow are more like *Agropyron.* Significantly, *Agropyron* and *Roegneria* are more like each other than either is like the ingroup, so that they are placed near each other when both are included in a parsimony estimate. This makes unlikely the hypothesis of two separate (diphyletic) origins for the ingroup, *Kengyilia,* which is only weakly supported by the present geographic distribution of the species. Roughly, the species more *Agropyron*-like are distributed in north-west China further to the north west, whereas the *Roegneria*-like species are distributed more to the south-east, south and south-west in a crescent-like area, the Tibetan plateau.

Another problem is the resolution of the trichotomy near *Kengyilia habahenensis* and *K. alaica. K. habahenensis* shares some characteristics of spike morphology and habit with *Agropyron* (Baum & al., 1991). Thus the monophyly of the species pair suggested in the absence of *Agropyron* (H* in Fig. 10, 15, 16), when *Agropyron* is included in the estimate becomes equally parsimonious with the trichotomy that results when the defining ancestor is brought down to the longer phyletic line to *Agropyron* (H in Fig. 9, 14, 16).

Discussion

One of the features of the artificial example is that distantly related taxa (the outgroup taxa) are connected to the ingroup taxa by a temporally long phyletic line, over which there has passed more time during which changes can take place. This explains in part why outgroup taxa are generally more dissimilar to ingroup taxa than ingroup taxa are to each other. It also explains in part why we might expect character states held in common between ingroup and outgroup taxa to be truly homoplastic. Parsimony seeks to minimize all homoplasy. As a result, when distant entities are included with the ingroup for parsimony analysis, sometimes ingroup homoplasies are increased (altering the undirected branching pattern of the ingroup) if homoplasies to the outgroup can be reduced. This can have the effect of bringing the ancestor taxa that gave rise to monophyletic subgroups of the ingroup taxa "down" onto the main phyletic line joining the outgroup with other distant parts of the ingroup, which destroys the monophyly of that subgroup. Subgroup *(d, e)* of the artificial example illustrates this, while the altered branching pattern for *(a, b, c)* illustrates the effect of higher ingroup homoplasy. Two instances of these effects were illustrated in our natural example.

One of the main arguments for the "outgroup during" approach is that it takes into consideration conflicts of characters that exist outside the ingroup, and thus it ensures that a cladogram will be maximally parsimonious when the ingroup and outgroup are combined. Provided that true homoplasy is not eliminated in favour of false homoplasy, this is certainly a valid argument. If true homoplasy is suspected between the ingroup and the outgroup, undirected phylogenetic lines could be estimated for the ingroup and then the outgroup attached to it without altering the ingroup branching pattern. For parsimony users, it would be straightforward to implement this in contemporary parsimony computer programs, if it is not already a feature.

Another common argument favouring the inclusion of the outgroup during parsimony analysis is that it eliminates the somewhat subjective step of somehow attaching an outgroup afterward. In so far as it does allay the discomfort that some may

feel in carrying responsibility for directing an undirected tree, this may be a valid argument. However, an additional problem with the practice of including outgroups during parsimony is that the choice of outgroup may be itself very subjective. In some cases the choice is based on existing classifications, many of which are not based on phylogenetic reconstruction and some of which were erected with no explicit consideration of evolution at all. Outgroups ought to be phylogenetically near or next to ingroups. Thus, recognizing an outgroup is impossible unless phylogenetic relationships at the next higher taxonomic level are already known. But to study them with an outgroup requires that the next higher level be known, etc., until finally all organisms would have to be subject to phylogenetic reconstruction, which itself would have to be done without an outgroup because there would be no organisms left.

One solution to this problem, suggested by Eldredge (1979) with whom Watrous & Wheeler (1981) agreed, is to take as a practical starting point those existing classifications that are the results of centuries of progress. But many formerly well studied classifications are at variance with recent phylogenetic reconstructions based on molecular data. Donoghue & Cantino (1984) suggested that a solution might be to use all possible sister groups as outgroups, alone and in different combinations, to assess polarities and then construct directed cladograms for the ingroups, in the spirit of "outgroup before" parsimony. Wheeler (1990) suggests using a random outgroup if none is available.

Some difficulties in the choice of an outgroup are functional. If it is too close it may attach to many places, and if it is too distant it may not attach anywhere very convincingly. This seems to be a common problem in large data sets and with molecular data. This problem is discussed by Donoghue & Cantino (1984), who recognized the need to develop methods to cope with uncertainty of outgroup selection.

It is important to realize that the destruction of potentially monophyletic groups by the inclusion of outgroups in parsimony estimates may not always introduce errors in the branching of phyletic lines. Without the (often extinct and unavailable) most recent common ancestor of the ingroup included, parsimony may create bogus potentially monophyletic groups whose destruction would improve the estimate. The problem is in recognizing true homoplasy, not using or failing to use outgroups. Even if an outgroup could be identified with complete historical accuracy, the problem of parsimony eliminating possibly true homoplasy, as illustrated in our examples, persists. In contrast to the advocacy of Maddison & al. (1984) who would include many different species of the outgroup, what we really want to include with the ingroup is the most recent common ancestor of the ingroup. Many true homoplasy problems are caused because a very important member of the ingroup is missing, namely its most recent common ancestor.

Using an outgroup is clearly not preferable to using the ingroup's most recent common ancestor, if it exists, but may be justified when this ingroup ancestor is not available. In this case, what we would clearly want from the outgroup would be its most recent common ancestor. When it is also unavailable then the relevant treatment of the outgroup should have as its goal the plausible estimation of its most recent common ancestor. This is why in our natural example we use one single composite entity to represent *Agropyron* as an outgroup, and another to represent *Roegneria* as an outgroup. At this point it is natural to wonder when we would expect to make an estimate of the most recent common ancestor (which we don't really want) of a less

well studied outgroup, that would be better than an estimate of the most recent common ancestor of our well studied ingroup, which is what we really want. Perhaps what we should do with outgroups is use them along with other data to make a plausible, character-by-character estimate of the ingroup, most recent common ancestor, for inclusion in the estimate of the branching pattern of the ingroup phyletic lines.

Another reason for including outgroups during parsimony is philosophical, and strongly argued by followers of the dictum that only monophyletic groups be recognized in biological classification. The philosophical argument considers a cladistic analysis as one indivisible step. Since the goal of cladistic analysis is classification, and since classification is based on monophyletic groups that can only be discovered in a "complete" analysis through rooted cladograms, outgroups must be included as an integral part of the analysis to avoid passing through a step that could not be considered cladistic analysis. However, Colless (1985) suggests that the use of outgroups may introduce more error than value, and it is the opinion of Lorenzen (1993) that "outgroup during" parsimony is completely useless and may even be misleading; useless when it can be applied unambiguously because then it is not needed, and possibly misleading when it cannot be applied unambiguously.

Certainly, a goal of virtually all modern systematists, whatever their philosophies may otherwise be, is to incorporate the history of evolution somehow in their classifications. Thus, accurate estimates of that history are important to all systematists. Methodological understandings that can improve accuracy are in everyone's interest. Our goal here is to contribute to those understandings.

Literature cited

Baum, B. R., Yen, C. & Yang, J. L. 1991. *Kengyilia habahenensis (Poaceae: Triticeae)* – a new species from the Altai mountains, China. *Pl. Syst. Evol.* 174: 103-108.
– , Yang, J. L. & Yen, C. 1995. Taxonomic separation of *Kengyilia (Poaceae: Triticeae)* in relation to nearest related *Roegneria, Elymus,* and *Agropyron,* based on some morphological characters. *Pl. Syst. Evol.* 194: 123-132.
Colless, D. H. 1985. On the status of outgroups in phylogenetics. *Syst. Zool.* 16: 284-295.
Crisci, J. F. & Stuessy, T. F. 1980. Determining primitive character states for phylogenetic reconstruction. *Syst. Bot.* 5: 112-135.
Donoghue, M. J. & Cantino, P. D. 1984. The logic and limitations of the outgroup substitution approach to cladistic analysis. *Syst. Bot.* 9: 192-202.
– & Maddison, W. P. 1986. Polarity assessment in phylogenetic systematics: a response to Meacham. *Taxon* 35: 534-545.
Eldredge, N. 1979. Cladism and common sense. Pp. 165-198 *in:* Cracraft, J. & Eldredge, N. (ed.), *Phylogenetic analysis and paleontology.* New York.
Farris, J. S. 1988. *HENNIG86*, version 1.5. New York.
Felsenstein, J. 1993. *PHYLIP (Phylogeny Inference Package),* version 3.5c. Seattle.
Hendy, M. D. & Penny, D. 1982. Branch and bound algorithms to determine minimal evolutionary trees. *Math. Biosci.* 59: 277-290.
Lorenzen, S. 1993. The role of parsimony, outgroup analysis, and theory of evolution in phylogenetic systematics. *Z. Zool. Syst. Evolutionsforsch.* 31: 1-20.
Lundberg, J. G. 1972. Wagner networks and ancestors. *Syst. Zool.* 21: 398-413.
Maddison, W. P., Donoghue, M. J. & Maddison, D. R. 1984. Outgroup analysis and parsimony. *Syst. Zool.* 33: 83-103.
Meacham, C. A. 1984. The role of hypothesized direction of characters in the estimation of evolutionary history. *Taxon* 33: 26-38.

– 1986. More about directed characters: a reply to Donoghue & Maddison. *Taxon* 35: 538-540.

Nixon, C. & Carpenter, J. M. 1993. On outgroups. *Cladistics* 9: 413-426.

Rohlf, F. J. 1992. *NTSYS-pc: numerical taxonomy and multivariate analysis system,* version 1.70. Setauket, NY.

Stevens, P. F. 1980. Evolutionary polarity of character states. *Annual Rev. Ecol. Syst.* 11: 333-358.

– 1981. On ends and means, or how polarity criteria can be assessed. *Syst. Bot.* 6: 186-188.

Watrous, L. E. & Wheeler, Q. D. 1981. The outgroup comparison method of character analysis. *Syst. Zool.* 30: 1-11.

Wheeler, Q. D. 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6: 363-371.

Yang, J. L., Yen, C. & Baum, B. R. 1992. *Kengyilia:* synopsis and key to species. *Hereditas (Beijing)* 116: 25-28.

Yen, C., Yang, J. L. & Baum, B. R. 1996. *Kengyilia guidenensis (Poaceae: Triticeae),* a new species from Guide, Qinghai, China. *Novon (in press).*

Appendix: Characters and character states in *Kengyilia*

1. Plants (roots or rhizomes):
   - 0. with fibrous roots;
   - 1. with short rhizomes;
   - 2. without creeping underground shoots and without fibrous roots.

2. Culms (tuftedness):
   - 0. forming small tufts;
   - 1. tufted;
   - 2. densely tufted;
   - 3. loosely tufed.

3. Culms (pubescence):
   - 0. minutely pubescent below spikes;
   - 1. pubescent just below spikes;
   - 2. at least glabrous below spikes;
   - 3. densely retrorsely pubescent;
   - 4. white pubescent below spikes.

4. Leaf sheaths (vesture):
   - 0. glabrous;
   - 1. rarely pubescent at base;
   - 2. at least the lower ones long pilose;
   - 3. puberulent (slightly hairy);
   - 4. densely pubescent.

5. Leaf blades (cross shape):
   - 0. involute or subinvolute;
   - 1. flat or involute;
   - 2. flat or convolute;
   - 3. complanate or subulate-involute;
   - 4. complicate;
   - 5. flat or slightly involute;
   - 6. subulate-involute;
   - 7. subulate (stiff);
   - 8. involute;
   - 9. flat.

6. Leaf blade, upper surface (vesture):
   - 0. pubescent or with very short scattered spinules;
   - 1. glabrous;
   - 2. glabrous or scabrous;
   - 3. villous or white pubescent;
   - 4. slightly scabrous;
   - 5. glabrous or sparsely pilose;
   - 6. puberulent;
   - 7. sparsely hairy;
   - 8. scabrous.

7. Leaf blade, lower surface (vesture):
   - 0. glabrous;
   - 1. scabrous;
   - 2. puberulent at least towards tip;
   - 3. densely puberulent;
   - 4. pubescent.

8. Leaf blade, margins (vesture):
   - 0. sparsely pilose;
   - 1. ciliated;
   - 2. scabrous;
   - 3. smooth.

9. Spikes (shape):

    0. straight;
    1. curved;
    2. straight or slightly curved.

10. Rachis internodes (vestiture):

    0. densely villose;
    1. scabrous;
    2. pubescent;
    3. glabrous or with sparse short hairs;
    4. glabrous with scabrous edges;
    5. puberulent;
    6. glabrous with ciliated edges;
    7. glabrous;
    8. densely pubescent.

11. Spikelets (disposition in the spikes):

    0. secund;
    1. slightly secund;
    2. not secund.

12. Spikelet (colour):

    0. green to brown;
    1. green or irregularly colored;
    2. greenish-purple;
    3. green;
    4. pale yellow-green;
    5. purple;
    6. straw colored;
    7. yellow-brown or green.

13. Spikelet rachilla internodes (vestiture):

    0. densely minutely hairy;
    1. scabrous;
    2. minutely pubescent;
    3. slightly hairy (puberulent) on back only;
    4. slightly hairy;
    5. with appressed minute hairs;
    6. densely hairy;
    7. pubescent.

14. Glume (shape):

    0. oblong-lanceolate;
    1. wide-lanceolate;
    2. ovate-oblong;
    3. oblong;
    4. triangular;
    5. ovate-lanceolate;
    6. lanceolate;
    7. narrow-oblong.

15. Glumes (relative size):

    0. equal;
    1. unequal.

16. Glume back (vestiture):

    0. glabrous;
    1. hairy;
    2. glabrous, but scabrous or ciliate on the nerves.

17. Glume (tip):

    0. tapering to a point 0.5-1 mm long;
    1. mucronate;
    2. acuminate or mucronate;
    3. pointed to mucronate;
    4. ending with a 1-2 mm long awn;
    5. ending with a 2-3 mm long awn.

18. Lemma back (vestiture):

    0. densely hairy;
    1. covered with 1-1.5 mm long white hairs;
    2. densely hairy toward tip and at margins but with glabrous back;
    3. covered with short soft hairs;
    4. hairy;
    5. covered with appressed hairs;
    6. densely appressed pubescent or with short hairs;
    7. with 1-1.5 mm long hairs toward tip and at margins but the back with appressed hairs;
    8. glabrous.

19. Lemma (tips):

    0. awned;
    1. not awned but acute or mucronate.

20. Lemma awn (shape):

    0. straight;
    1. slightly bent (subdivaricate);
    2. slightly recurved or contorted;
    3. straight or slightly bent.

21. Palea (size relative to lemma):

    0. shorter or equal to lemma;
    1. shorter than lemma (sometimes only slightly);
    2. equal to lemma;
    3. slightly longer than lemma;
    4. equal to or longer than lemma, may be only slightly longer.

22. Palea (apex shape):
- 0. emarginate;
- 1. obtuse or slightly emarginate;
- 2. truncate or emarginate;
- 3. emarginate or bilobed;
- 4. slightly emarginate;
- 5. truncate;
- 6. retuse;
- 7. bidentate.

23. Anther (colour):
- 0. yellow;
- 1. yellow or red;
- 2. yellow or purple;
- 3. black or dark greenish brown;
- 4. grey;
- 5. pale yellow or brown;
- 6. black;
- 7. yellow or black;
- 8. purple;
- 9. pale yellow.