Pl. Syst. Evol. 175: 39–54 (1991)

—Plant—
Systematics
and
Evolution

Restriction data from chloroplast DNA for phylogenetic reconstruction: is there only one accurate way of scoring?

BIRGITTA BREMER

Received April 27, 1990; in revised version October 29, 1990

Key words: Angiosperms, *Rubiaceae*. – Chloroplast DNA, fragment analysis, restriction analysis, site analysis, Wagner parsimony.

Abstract: Information from the same restriction analysis of chloroplast DNA of 33 taxa of *Rubiaceae* was scored in four different ways, two of which were based on fragments, and two on restriction sites, and they were subsequently analysed with Wagner parsimony. The methods resulted in different phylogenetic trees. The inherent differences between the methods relate to the amount of non-homologous characters and dependent characters, but none of the methods will systematically bias the resulting cladograms. The fragment analyses are much less time-consuming, but probably less accurate, than the site analyses. The choice of method is dependent on a trade-off between accuracy and resources (time). One important recommendation is made: all phylogenetic analyses of chloroplast DNA data should be accompanied by a data matrix and contain information on how the matrix was compiled.

Cladistic methodology, combined with information from nucleic acids, has focused an increased interest on systematics, phylogeny, and evolution. In plants, perhaps the most fruitful studies have been based on restriction endonuclease data analysis of chloroplast DNA (cpDNA). Although DNA sequencing is becoming more common, restriction data from cpDNA will probably continue to be very useful for phylogenetic analysis. The usefulness of this approach is evident, above all, from the work of PALMER and coworkers (PALMER & ZAMIR 1982; PALMER & al. 1985, 1988; PALMER 1985, 1987; but also from many other studies (KUNG & al. 1982; HOSAKA & al. 1984; PERL-TREVES & GALUN 1985; SYTSMA & SCHAAL 1985; HOSAKA 1986; SYTSMA & GOTTLIEB 1986 a, b; JANSEN & PALMER 1987 a, b; RIESEBERG & al. 1988; ROSE & al. 1988; SYTSMA & SMITH 1988; SOLTIS & al. 1989; SMITH & SYTSMA 1990; SORENG & al. 1990; SYTSMA & al. 1990).

In general, restriction data are very suitable for phylogenetic analysis, as fragments or sites are discrete units; they either do or do not occur, and thus are easily coded as binary characters. However, the difference between probability of a restriction site gain and a restriction site loss (TEMPLETON 1983) is a problem with use of restriction data, and it could justify assignment of different weights to the two state changes (ALBERT & al. 1991). There has been some dispute over which parsimony principle that should be used in an analysis to handle the "gain and

loss" problem (cf. DEBRY & SLADE 1985, ALBERT & al. 1991). In Wagner analysis (PAUP package, SWOFFORD 1989, HENNIG86, FARRIS 1988) all gains and losses are treated equally, while in Dollo analysis (PAUP package, SWOFFORD 1989, FARRIS 1977) parallel gains are not allowed. However, it is possible to use step matrices for different weighting of gains and losses (PAUP package, SWOFFORD 1989). The resulting phylogenetic pattern will depend on the choice of method (cf. JANSEN & al. 1990, BREMER & JANSEN 1991). In the present study I focus on another methodological issue, namely how the data set for phylogenetic analysis of a restriction pattern should be constructed. Is there only one accurate way of scoring data? Are there any inherent differences between the different methods of scoring? Will different scoring of data from the same molecular study, analysed with the same parsimony program, yield different cladograms? To answer the questions I have used results from a cpDNA investigation of the coffee family, *Rubiaceae* (BREMER & JANSEN 1991).

Data matrices are rarely included in publications, although there are exceptions (e.g., SYTSMA & SMITH 1988, SORENG & al. 1990). Because of that it is usually difficult, if not impossible, to reconstruct them from the results of analyses, and impossible to find out how the scoring was done, or to reanalyze the data set. Students of restriction data usually state that they have used restriction site mutations as characters, whereas length mutations are usually not considered in most analyses (cf. PALMER & al. 1988). However, my interpretation is that site occurrences rather than mutations were scored in most of the studies published. The question of scoring patterns (occurrences) or processes (mutations), is intimately linked with the more general question of homology. I argue for the principle that characters which appear similar should be tentatively treated as homologous (Hennig's auxiliary rule, HENNIG 1966: 121), and that final hypotheses of homology and homoplasy should be identified from character distributions after the tree is constructed. The main object of the present paper is to provide an explicit comparison of four different methods for scoring restriction data. I will evaluate the methods with regard to the effort needed to produce data, the risks of incorrectly interpreting characters as homologous and including characters that are causally "dependent", and with regard to their basic assumption of inferring processes (mutations) from pattern (phylogeny). Since the methods may yield different results I will argue that the procedure of character scoring should be explicitly described, and that data matrices should be published.

Restriction data of cpDNA for phylogenetic analysis

Chloroplast DNA is digested with restriction enzymes and the fragments separated by gel electrophoresis. The molecule is about 150 kilobases (kb) long, and after digestion with an enzyme that recognizes six nucleotides, it is common to get about 10 to 60 different fragments. An analysis in which possessions of these fragments are used directly as characters, will be referred to as fragment "direct" analysis (FDA).

In order to reduce the risk of non-homology between similar-sized fragments of different taxa, one usually tries to localize the area of origin on the cpDNA molecule of a particular fragment. This is done by heterologous or homologous probe hybridizations. These cpDNA probes have a known size and position on

the cpDNA molecule and together they usually cover most of the cpDNA molecules. Restriction fragments of the same size and position on the cpDNA molecule can then be used as characters in a phylogenetic analyis, in the following called fragment occurrence analysis (FOA) (cf. WAUGH & al. 1990).

In some of the early studies an "unspecified fragment analysis" was used (BERTHOU & al. 1983, CLEGG & al. 1984, ERICKSON & al. 1983). Fragment based distance analysis (NEI & LI 1979) has also been applied by some students (COATES & CULLIS 1987, LEHVÄSLAIHO & al. 1987, HANTULA & al. 1989). However, most systematists using cpDNA for phylogenetic analyses have adopted parsimony instead of distance methods, and restriction site comparison instead of fragment comparison (cf. PERL-TREVES & GALUN 1985).

Restriction site comparison is very different from FOA and FDA when dealing with restriction data. The original fragment order is reconstructed as more or less detailed restriction maps and the restriction sites are then aligned and used as characters. If one taxon has a fragment of 8 kb at a particular position and another taxon has two fragments of 2 and 6 kb, respectively, at the same position, this is interpreted either as a site loss in the first taxon, or as a site gain in the latter. Length mutations (insertions or deletions) can also be identified.

There are two different methods of dealing with sites as characters. In site occurrence analysis (SOA) the pattern of restriction sites is used. The occurrence or absence of a site at a particular position is used as a character in the analysis. However, the underlying process behind the pattern causing occurrence or absence may be different: various substitutions among the six positions of the site, a deletion, or an insertion, may cause the change in site occurrence.

The other method, here called site mutation analysis (SMA), makes some assumptions about the processes. Gain or loss of a site is used as a character only when it is supposed to be caused by the same mutation. If a particular site loss is supposed to be the result of a substitution then species without the site but with a deletion within the same region should not be scored for a site loss (0) but with a question mark (?), as the suspected underlying process may be different. In several studies restriction site mutations are said to be used as characters (cf. PALMER 1988, JANSEN & al. 1990). As matrices are usually not published, the method of scoring is, however, not explicitly shown. According to the terminology in the present paper, these studies are probably SOA.

In SMA, changes that can not be homologized directly should be scored as "unknown". Consider a hypothetical case where the outgroup and some of the ingroup taxa have two fragments 2 and 6 kb, respectively, and the remaining taxa have an 8 kb fragment. The latter taxa should be scored as 0 in the matrix, indicating a site loss mutation. But if a third taxon possesses a fragment that is only 7 kb long, the mechanism causing the loss of the site could also be a non-homologous mutation, i.e. a deletion including the site. Hence, that taxon should be scored as unknown for that site character.

Material and methods

I have compared four data sets derived from an investigation of 33 genera of the *Rubiaceae* (BREMER & JANSEN 1991). The restriction pattern was obtained by digestion with eight restriction enzymes and subsequent hybridization with 19 different heterologous probes from *Lactuca* and *Petunia*.

Table 1. FDA matrix of the 412 phylogenetically informative restriction fragments. 1 indicates presence of fragment, 0 indicates absence of fragments, and − indicates incertainty in autoradiogram reading. The numbers correspond to the same numbers in the FOA matrix. All fragments of equal length have been pooled into the first fragment number of each length, a number including an × (eg. × 1) indicates characters which were autapomorphic in the FOA analysis. Characters are restriction fragments obtained from digestion with the following enzymes Nco I (1 − 55, incl. x1), Bcl I (58 − 148, incl. 1x1 − 1x5), Hind III (149 − 212, incl. 2x1 − 2x8), Bam HI (218 − 296, incl. 2x1, 2x2), Bst XI (298 − 357, incl. 3x1), Eco RV (358 − 418, incl. 4x1, 4x2), Hae II (420 − 486, incl. 4x1 − 4x5), and Sac I (487 − 548, incl. 5x1, 5x2)

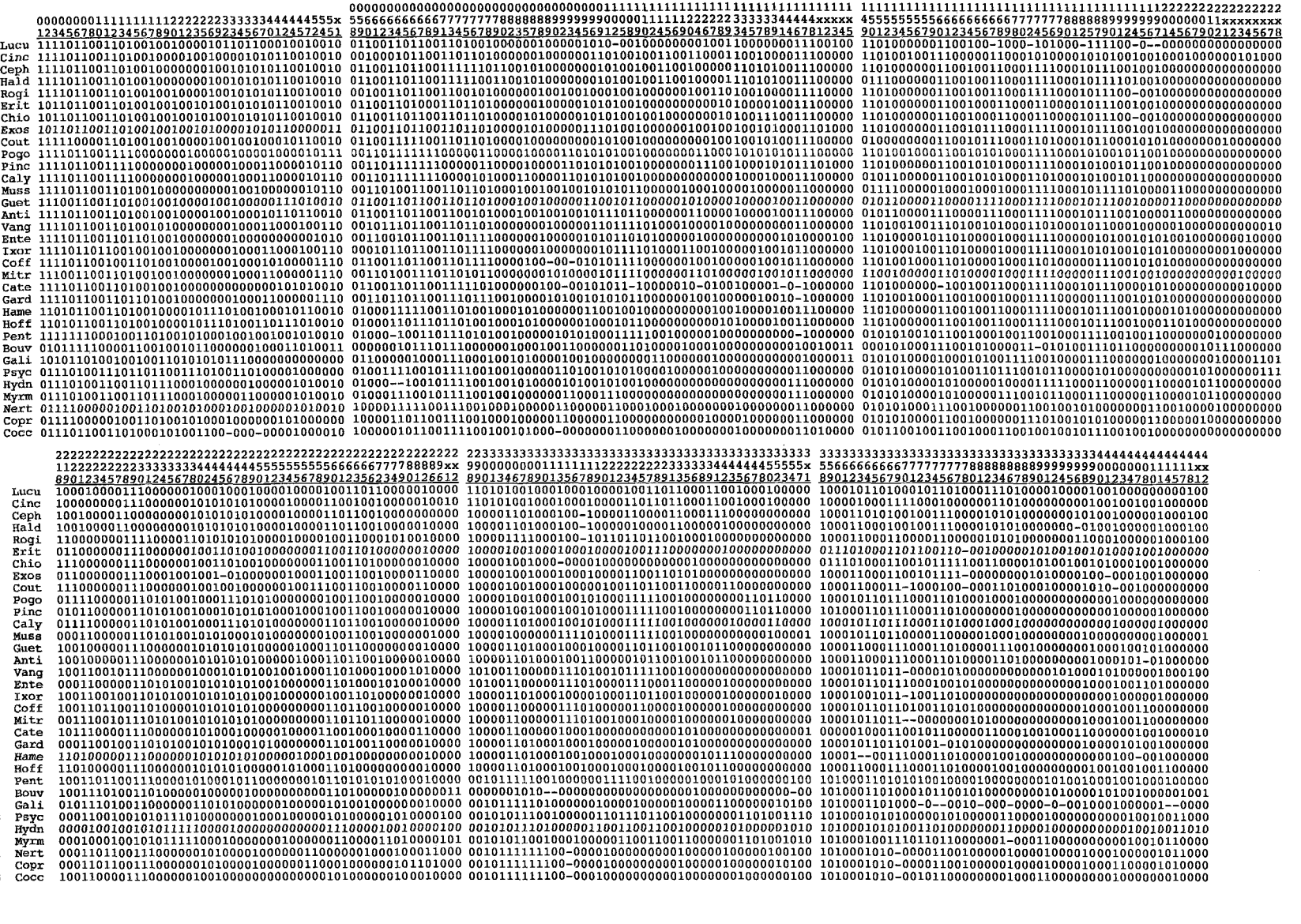

Table 1. (continued)

		44444444444444444444444444444444444444444444444	44444444444445555555555555555555555555555555555
		2222222223333333334444555555666666667777788888xxxxx	8889999999999900000001111222222333333344444xx
		01234567801234567893678024567801245679147892456123455	78901234567890123679127912457823456789027812
1	Lucu	0001100100100101010100000000000000000000110010000001000	0001000-1101-0000100010000--00010110-0011000
2	Cinc	000001110001010111000000000100000000011001110000000001	00011001001110000000010001000001100010010100
3	Ceph	0110110100010101100000000101000000010001001000000000000	10011100101110000000010000110001000000100100
4	Hald	0000111100010101100000000010000000100010010010000000000	0001100010110000000010000110001000100-110000
5	Rogi	0000111100010101100000000000000000100010010010000000000	00011000101110000000010000110001011001000100
6	Erit	0000--01000101011000000000000000000100001001010100000000	01011000010100000000011100110001011001010000
7	Chio	0000--01000101011000000000000000000100001001010100000000	01011000010100000000011100110001011001010000
8	Exos	00001-11000101011000000000000000000100001001000110000000	01011000101110000000010000110001011001010100
9	Cout	000011110001010110000000100100000001000100100011000000	01011000100100000000010100110001100010010110
10	Pogo	0000111100100100000000000000000000100010011000000000000	01010001100110000100010000000001011010010000
11	Pinc	0000111100100101000000000000000000100010011000000000010	01011001110110000100010000000001100010010000
12	Caly	000011110010010100000000100000000010001001101000000000	0101100011011000010001000000000010-1010010000
13	Muss	0000111100100101000000000000000100010010001001000000000	-00110011101110001000100000010011000100100000
14	Guet	000011110001010111000001000100010000000100100100000000	01011000110110000000010000110011011010010100
15	Anti	000011110001010111000001000100010000000100100100000000	01011000110101000000010000110011011010010100
16	Vang	10000-110010000000110001010000000000101011101000000000	01011100100100000100010011000001011010000000
17	Ente	0100001100100011O-100001100000000000100000110100010000	00011000100100010100010010000001011010010000
18	Ixor	0010110100100100001100000000000000000101011100000000000	11011000100100000100010010000000011010010001
19	Coff	1000011100100101001000001100000000000000001101000000000	00011000100100010100010010000001011010010000
20	Mitr	0001111100100010000000000000000000001000001101000000000	10011010100100010100010010000000010010010000
21	Cate	0000111100010100000000010100000000100001001001000000000	01011000010110000000010001110001011100000100
22	Gard	1000011100100010001000001100000000001000001101000000000	00011100100100010100010111000001001000010000
23	Hame	0000011100010101100000000001000000000010010010000000000	01011001100110000000010001110111011100000100
24	Hoff	0000111100010100100000000000100000010001001001000000000	01111001000000000000010001110111011100000110
25	Pent	0100010110001001101101010001000000000010010000010000	0-11110000010100011100000100000000001001000
26	Bouv	00001101101000001111100100000011000001010000100101100	00011000000101000111O-01000000010000000000001
27	Gali	000101010010000110100001010001011000001101001000000000	000010000000101-0011000101100100100000011010
28	Psyc	0100110011010001101011100000101000000011001000000000000	01011100001100100010000001000001100000001000
29	Hydn	0-0---0---10000-0--0-0-00000-0-0000000-00-00-0000000	01011110000100101110000101000001100001001000
30	Myrm	0000110011100001011010100000111110000011010001000000000	01011110000100101110000101000001100001001000
31	Nert	0000010100100001111000010101010010000011110010000000000	00011-00000-010001-010000-000101-00001001000
32	Copr	0100010100100000111001010100010010000001000001000000001	00011-00000-010001-010000-000101-00001001000
33	Cocc	01001001000010010100010101000000000000011010010010100	000-1100000100001-000000-0000001-0000100100

FDA matrix (Table 1). Fragments from different taxa of equal length, without considering the positions on the cpDNA molecule, were treated as homologous. In this study the FDA matrix was derived from the FOA matrix (Table 2) by pooling all fragments of equal length. Of the 548 fragments in the FOA, 386 were of different length. Furthermore, several autapomorphies in the FOA became 26 shared characters. Consequently there were 412 phylogenetically informative characters (occurring in at least two but not all of the ingroup taxa) in this analysis.

FOA matrix (Table 2). The fragment measurements were made directly on the autoradiograms from the different hydridizations (Bremer & Jansen 1991). There were 944 different fragments identified. Of these, 548 were phylogenetically informative.

SOA matrix (Table 5 in Bremer & Jansen 1991). There were 268 restriction sites identified, 161 of which were phylogenetically informative. In all taxa where a particular site did not occur, it was scored as absent (0).

SMA matrix (Table 3). Only those taxa where the supposed homologous site mutation could be identified were scored for presence or absence of the mutation. In taxa with length mutations, or where direct site mutations could not be identified, the character state was scored as unknown, as the process behind the change is unknown. The SMA matrix differs from the SOA matrix by the way the lack of sites is scored when homology is dubious. A question-mark (unknown) in the SMA matrix in Table 3 should be read as a 0 (absent) to get the SOA matrix. The number of informative sites is the same as in the SOA, i.e. 161.

The polarization of gains and losses of restriction sites or fragments was done using the genus *Luculia* as outgroup (Bremer & Jansen 1991).

The four matrices were analyzed with the Wagner parsimony program Hennig86 (Farris 1988) on a microcomputer. The initial trees were calculated by the mhennig method, and the options mhennig* and bb*, were used to construct the cladograms.

Results

The Wagner parsimony analysis of the 33 *Rubiaceae* spp. gave the following results. The 412 fragments in the FDA gave two equally parsimonious trees, 1 386 steps

Table 2. FOA matrix of the 548 phylogenetically informative restriction fragments. 1 indicates presence of fragment, 0 indicates absence of fragments, and − indicates incertainty in autoradiogram reading. Characters are restriction fragments obtained from digestion with the following enzymes Nco I (1 − 56), Bcl I (57 − 148), Hind III (149 − 217), Bam HI (218 − 297), Bst XI (298 − 357), Eco RV (358 − 419), Hae II (420 − 486), and Sac I (487 − 548)

Table 2. (continued)

		33333333333333333333333333333333333333333333333333	33333333333333333333333333333333333333333333333333	44444444444444444444444444444444444444444444444444
		00000000001111111111222222222233333333334444444444	55555555556666666666777777777788888888889999999999	00000000001111111111222222222233333333334444444444
		01234567890123456789012345678901234567890123456789	01234567890123456789012345678901234567890123456789	01234567890123456789012345678901234567890123456789
1	Lucu	01001100000001100010000100011001011000010100110000	10000000100010110100001011010000010-01000010000001	00100100000001000010000110010101001000000000100001
2	Cinc	01001100000001100010000110011001011000010100110000	10000000100001000100110001000000010101000001000001	00100100000100000001000000110100101001000000100000
3	Ceph	00001100000001100-10000110100001111000010110000001	00000000100011000100010011100100001000000000001001	00100100000100000001001001001010010100100000100000
4	Hald	00001100000001100-10000010100001111000000001000001	0000000010001100010001001110010000100000000000-001	00100100000100000001000001011010010100100000100000
5	Rogi	00001100000001100-10000110011000111000010000100000	00000000100011000100000011000000100101000001001010	00100100000100000001000001011010010100100000100001
6	Erit	00001100000001100010000100011100000000000010000001	00000000111010001000110010001001001000100001010001	010000-101010000000010000-00101001010010000100000
7	Chio	00001-0000000-1000000000000000000001000000000000000	00000000111010001000010110101001101000100001010001	0100001101010000000010000-00101001010010000100000
8	Exos	00001100000001100010000110001100011010000000000000	00000000100011000100001011010100000100010001000010	0-0000-1010100000000100001011010010100100000100001
9	Cout	00001100000001100000100110011000010100000001000001	00000000100011000100-1000100000001010000000001000010	00-000-1010010000000100001011010010100100000100000
10	Pogo	00001100000001100101000110011010010000000000101100	10100010100010110100100011010000100101000001000100	00100000000010000000000001011010101001000000000000
11	Pinc	00001100000001100101000110011010010000000000101100	101000101010001101001000110100000000-01000001000100	00100000000010000000000001011010101001000000100000
12	Caly	00001100000001100101000110111010110000000001100000	10100010100010110100100011010000100101000001000100	00100000000010000000000001011010101001000000100000
13	Muss	00001100000001011001000110011010010000000000100000	10000000100010110100000011000000100-01000001000010	00000000000010000000000001011010101001000000100001
14	Guet	00001100000001100010000110011000111000100110000001	00000000100011000100100011010000010100010001000001	00010100100100000000000001011010010100100000100000
15	Anti	00001100000001100110000010111000111000100110000001	00000000100011000100100011010000010101000001000001	00010101-00100000000000001011010010100100000100000
16	Vang	00101000000011011010000110011010010000000000000000	00000000100010110100-00001010000000000000011000000	10100000000010000000101000001101010000000000000000
17	Ente	00101100000011011010000110100011110000000001000001	00000000100010110100100010010010000000000001000010	00100001100100000000001000011010100010000100100000
18	Ixor	00001100000001100001000110011000110000000001000001	00000000100010010100--00110100000000000001000100	00100000000100000000000010100101010010000000000000
19	Coff	00001100000001011010000010100010110000000001000001	00000000100010110100010011010010000000000001000010	00100001100000000000001000001101010010000000000000
20	Mitr	00001100000001011010000000100010110000000001000000	00000000100010110100--00000001000000000000001000010	00100001100001000000000000011010100010000100000001
21	Cate	00001100000001000010000000100000110010000000000000	00000000000010001000010110000000110000100001000010	00000001010100000000100001011010010100000000000000
22	Gard	00001100000001100010000010100000110010000000000000	0000000010001011010001001-01001000000000001000010	00010001000100000000100000110101000100001000000000
23	Hame	00001100000001100100000001000000101000100110000001	0000000010001--00100100011010000010101000001000100	00100000000010000000-000000110100101001000000100000
24	Hoff	00001100000001100100010001000000111000100110000001	00000000100011000100100011010000010101000001000001	00100000000010000000100001011010010100100000000000
25	Pent	10011011100100100000111100010000000100000010010000	01001000101000110100101001000000100000000100100010	00100001000000100001010000010000010001010111100100
26	Bouv	00000010-0--00100000000000000000000100000000000000	0000-000101000110100001011001001000000000100100000	10100001000000000-0000011010001000001010111011001
27	Gali	10011011101000100001000010100000110001000110000001	00001000101000110100000-0--00100-0000-0000-00-00010	00100000000100--0000000101010001000001010010000000
28	Psyc	10010011100100100000101110011000010001000000011010	01010111101000101011000000000000000101000101000100	00001000001000011100010001001010100011010000101011
29	Hydn	10010011101000100001000100011000010001001010010000	01010111101000101011001100000000001010000101000100	00001000001000011100 0-0---0--0-10000-0--0000-0-0--
30	Myrm	10010011001000000000000100011000110101001000011010	0101011110100010101100110010010000 10-0001101000100	00001000001000011000000001001011000010100000001011
31	Nert	10011011011000 1-0000100000100000110001000010000000	10001000101000110100000011001000000001000010-000010	00100000000100010100000001010001000000011 0010-00000
32	Copr	100110110110001-0000100000100000110000000001000001	00101000101000110100000011001000000001000010-000010	00100000000100010000010001010001000000110010000100
33	Cocc	100110110110001-0000000000100000110000000000000000	01001-00101000110100001011000000000001000110-000100	00100000000000010000010000010000010010100000110000

		44444444444444444444444444444444444444444444444444	5555555555555555555555555555555555555555555555555
		55555555556666666666777777777788888888889999999999	0000000001111111111222222222233333333334444444444
		01234567890123456789012345678901234567890123456789	012345678-012345678901234567890123456789012345678
1	Lucu	00010000000000000000000100010010000000000000-00000100-	000001100000110010000000---100010010110-0001000000
2	Cinc	01000010000000000000101000100100000000000001000000101	000000000001100000000010000001001100010001000101
3	Ceph	01100100000000001000001000000100100000100000000000101	000000000001100101000001110001001000000010001 0101
4	Hald	01000100000000001000001000000100000010000000000000100	00000000000110010100000111000100100100-101000100
5	Rogi	01000000000000001000001000000100000010000001000000101	000000000001100101000001110001001011001000000101
6	Erit	0-000000000000001000000100000010001001001010000001000	000000000001100011010001110001001011001001000100
7	Chio	0-000000000000001000000100000010001001001010000001000	000000000001100011010001110001001011001001000100
8	Exos	0-000000000000001000000100000010000001101010000000101	000000000001100101000001110001001011001001000101
9	Cout	01100010000000001000001000000100000000110-0100001000 0	000000000001100001010001110001001100010001000101
10	Pogo	01000000000000001000001000000110000000010100000000001	0000011000001100101000010010011001011010001000000
11	Pinc	01000000000000001000001000000110000000010100000001001	0000011000001100101000010010011001100010001000100
12	Caly	01100000000000001000001000000110000100010100000001001	000001100000110010100000001000100 10-101000100 0100
13	Muss	00000000001000010001000000000100000100-00100000001001	0000011000001100101000010011001101100010001000100
14	Guet	01000010010100000001001000100100000100010100000001001	000000000000110010100000111000101101101000100 0101
15	Anti	01000010010100000001001000100100000100010100000001000	100000000000110010100000111000101101101000100 0101
16	Vang	0-011000000000000000100010000111100001000101010000000	0000011000001100101001000010011001011010010000100
17	Ente	10100000000000000001000000000-1011000010000001100000000	0010011000001100100001000010011001011010001000100
18	Ixor	01010000000000000001000100001111010000001100000000000	00000110000011001010010000000-10000110100010001 00
19	Coff	01101000000000000000000000101100001000000110000000	0010011000001100101001000010011001011010001000100
20	Mitr	01000000000000000001000000000000110000100100110000000	0010001000001100101001000010011000010010001000100
21	Cate	11001000000000001000000010000000100000010010100000 1001	00001000000011000010000001110001001011100010000101
22	Gard	01101000000000000001000000000010110000100000101000000 0	0010011000001100101101000010011001001000001000100
23	Hame	01000010000000000000001110000001000000100000100010000 1	0000100000001100101000001110100011011100010000101
24	Hoff	01000010000000001000000100000010000000100001100010000 0	0000100000001101000000001110100011011100010000101
25	Pent	11000010000000000000000011000000101000000000-1100000000 0	1000011101000001001000100010000000000001000010110
26	Bouv	10000000001000000000100011100000000000100000-000000000	100000110100000100110000000100-00010000000000101000
27	Gali	10001000100011000000001111000100000010000000000000000	0-00001100000000011000100001000101000000001000110
28	Psyc	01000001000000000000000001000000010000100000001000000010	01001001001000000010001000100000011000000000100110
29	Hydn	0-00000-0000000000000-0000-0-0000--00010001000010	0101011110000010001000100010000001100001000011110
30	Myrm	01000001000000000000000101001010000100001000010000 10	0101011110000010001000100010000001100001000011110
31	Nert	10001000100011000000011101000100000010000010000000 00	1000011-00-1000000-000-000-0100001-00001000100110
32	Copr	10001000100011000000010000000000000100000100000000 00	1000011-00-1000000-000-000-0100001-00001000000110
33	Cocc	10000000000000000000000100100010100000000000000000 0	0000001-00100000001000-000-0000001-00001000100110

Table 3. SMA data matrix of the 161 phylogenetically informative restriction site mutations. 1 indicates restriction site gains, 0 indicates restriction site loss, – indicates incertainty in mapping. Question-marks (?) in this matrix should be read as a 0 to get the SOA matrix, see material and methods. Characters are obtained from digestion with the following enzymes Sac I (1 – 18), Bst XI (19 – 40), Hae II (41 – 62), Nco I (63 – 85), Bam HI (86 – 108), Hind III (109 – 136), Eco RV (137 – 143), and Bcl I (144 – 161)

				11111111111111111111111111111111111111111111111111	111111111111
		000000000111111111122222222223333333333444444444	55555555556666666666777777777788888888889999999999	00000000001111111111222222222233333333334444444444	555555555566
		123456789012345678901234567890123456789012345678 9	01234567890123456789012345678901234567890123456789	01234567890123456789012345678901234567890123456789	012345678901
1	Lucu	0011110110010001111111101010100001001000100100000	00001001000000011100010010111000010001111111000000	0011101000---11110000000010010001011011000000000010	100010011110
2	Cinc	0011010110010101011111101010100001001000100100000	00101001010001111100010010111010010001110111000010	00011010001011111000000000100?00010110110000000000011	101010111110
3	Ceph	00110?111001010 0?1101110101000000100100010010000 0	00001001010001111100010010111000010001110111000010	0001001000101111100000000100100010110110000000000011	101110011110
4	Adin	00111?11100100000?110111010?000000100100010010000 0	00101001010001111100010010111000010001110111000010	00010010001011111000000001001100101101100000000000011	101110011110
5	Rogi	0011110110010101011111101010100001001000100100000	00101001010001111100010010111000010011110111000010	00011010001011111000000001001000101101100000000000010	10101011111?0
6	Erit	0011110100010001011110100000000001001100100100000	0100?00101000111110001001011100001001?10?011010010	01?11010001010111000000001001000101101100001000000010	101110011110
7	Chio	001111010001000101????0010000000001001100100100000	0100?00101000111110001001011100001001?10?011010010	01?11010001010111000000001001000101101100001000000010	101110011110
8	Exos	0011110110010101011111100100000001001100100100000	01101001010001111100010010111000010011??111010010	01?1101000101111100000000100100010110110?00000000000010	101110011110
9	Cout	001101011001?010101111110000000000100110010010000 0	11101001010000011?000100101110000100011??111010010	01?11010001011111?0110000?0010001011011000000000000010	101110011110
10	Pogo	101111 0?000101011111111001001100011001010100100000	101010010000001001001?10101100100101001111111000100	0011111000101111110110000000010001011011000100010010	101011010110
11	Pinc	1011010100010101111111100100110001100101010010000 0	101010010000011001001?10101100100101001111111000100	0011111000101111110110000000010001011011000100010010	101011010110
12	Caly	0011110100010101111111100100010001100101010010000 0	001010010000011001001?10101100100101001101111000100	0011111000101111110110000000101010110110001000100 10	101011010110
13	Muss	0011010100010101111111100100010000100100110010000 0	0010100100000000?01001?10101110100101001111111000000	0011111000101111100001001001100101101100010000000010	100001011110
14	Guet	0011110110010101011111111001000000100101011010010 0	0010100101000-111?000100101110000100011101?1000010	000110100010111110000100100101000110-1000000000000010	101100011110
15	Anth	0011110110010101011111111001000000100101011010010 0	00101001010001111?00010010111000010001110111000010	000110100010111110000100100101000110-1000000000000010	1011100111-0
16	Vang	10111110100 0?0101111111100110000000100100 1?00100011	101010010000011001000?1010111000010001110111000000	101110100010111110010000100100010110 0?0001001000 10	00100-011110
17	Ente	101111010000110111110110010000000001001001100000011	1010100100000000?0111001010111000000001110111000000	0011111000011111110010001?0010001011000000100000010	1010?1011110
18	Ixor	10111101000?0001111111?0100000000001001000100100011	100010010000011001000?101101100001000111??11000000	10111111000101111100100 01?0010001011 00?0001000000010	00100101111 0
19	Coff	1011110100001101111011?010000000001001001?00?00001	10101001000001110111001010?11000010001110111000000	101111000 0?11111100100001001000101100000 00100100010	101000011110
20	Mitr	1010110100001 10?1110110010000000001001001?00000010	10101001000101100111001010111000010001110111000000	101111100 0?1111110010000100100010110000000100100010	101001011110
21	Cate	0011110110010101011011??01000000001001001100100000	0110100101000111?0000100101110000100111?1111000000	10?110100010111110010000100100010110 1?0000000000010	1011?0011110
22	Gard	10111?01000?010111011??0100000000100100 0?00000011	10101001000101100111001010111000010000?110111000000	101111100 0?11111100100001001000101101000010010001 0	101001011101
23	Hame	01011101100101010 10??111001000100100101010010100 0	0010100101000000 11?000100101110001?010111??1100000-	-0?110100010111111000101010010001011011000000000000010	101110011010
24	Hoff	010111011001000101 0?1111001000000100101010010000 0	00101001010001101000010010111000 1?01011101?100000-	-0?110100010111111000101010010001011011000000000000010	101100011010
25	Pent	00?001010101000110 11??0000001010111001000010100 0	0000?10010101010?00001001010110101100?0??11000000	001111001010111 1?1000000101000110101 1?011000001 1?0	-000?0011001
26	Bouv	0--10?0?0101000?10?????00100000000 11100---00111000	0010010?0010100?10000101101?110101100?0???1000000	001110?010?01111?1000000101?0011010 11?100000001 1?0	-0000-011?01
27	Gali	00?10?0100110 00?101011?00010000101110010?01111000	0010010?0010100??0000101001?11010?100?1???11000000	101110?010?01111?1000000101000 1?01011?10000000 1000	?000?0011001
28	Psyc	00?1010100110101001111?0001101111011001000011000 0	0001011001001 00??0000101010111000?010?1???11001001	001011110 0?00111?0000001?00?00111010??1000010 01000	?000?0001010
29	Hydn	00?1010100110111101111?00-0001111011001 0?00110000	0001?11?0000100??00001010101110000000?1??111001001	00101111 00?0?1110000000 1?11000111010??10000 1001000	?100?0001010
30	Myrm	00?101010011011111011 11?0000101111011?010?0011000 0	0001011?0000100??00001010101110000000?1??111001001	00101111 00?0011100000001?11000111010??10000 1001000	?100?0001010
31	Nert	010??1010101000110101 1?00000100101 11?010?0110100 0	0010010?0010100??0000101?0101101011 00?1???-?100000	0011110011001100?1000001?0100011010 11?011000001000	?000-0011001
32	Copr	010??101010100011010 11?000100000101 11?0100011000 00	001001000010100???000?01?01011010110 0?1???-?100000	0011110011001100?10000001001001101011?011000001000	?000-0011001
33	Cocc	00???101010?010010111 1??000001010111?0100001010 00	0010?1001000100??000010110 11?10001000?1???11000000	001110100100111100000000 100?00-101101?1000000010?0	?000-0011010

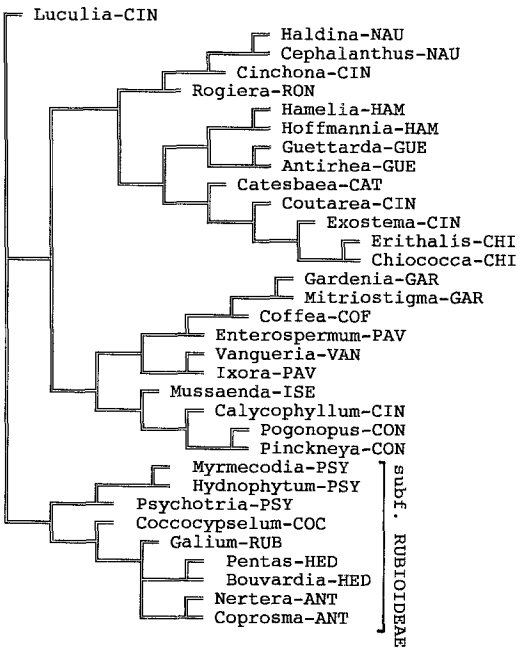

Fig. 1. A strict consensus tree from two equally parsimonious trees of *Rubiaceae*, based on restriction fragment analysis (FDA), each 1 386 steps long, with a consistency index of 0.29, and a retention index of 0.56

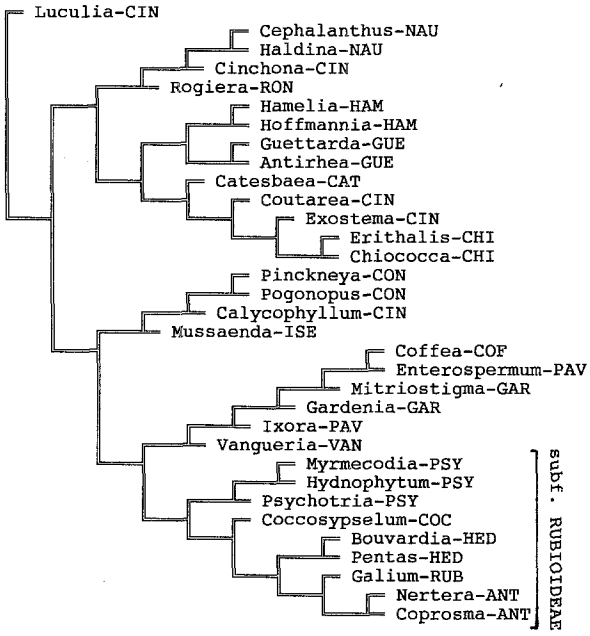

Fig. 2. A phylogenetic tree of *Rubiaceae*, based on restriction fragment analysis (FOA), 1 486 steps long, with a consistency index of 0.36 and a retention index of 0.63

long, with a consistency index of 0.29 and a retention index of 0.56 (Fig. 1). The 548 phylogenetically informative fragments (FOA) resulted in one tree, 1 486 steps long, with a consistency index of 0.36 and a retention index of 0.63 (Fig. 2). The site analysis (SOA) of the 161 informative sites (BREMER & JANSEN 1991) gave six equally parsimonious trees, 348 steps long with a consistency index of 0.46 and a retention index of 0.78 (Fig. 3), while the same 161 informative sites with the SMA analysis gave 60 equally parsimonious trees, 291 steps long with a consistency index of 0.53 and a retention index of 0.79 (Fig. 4).

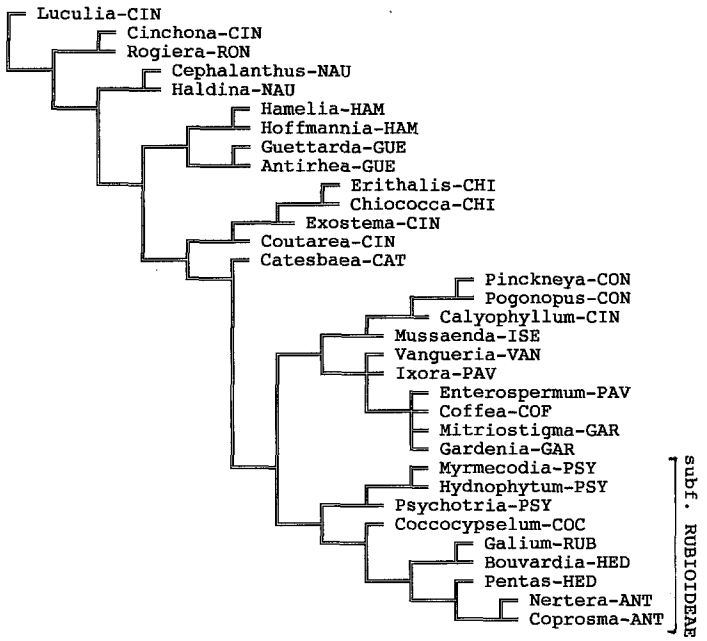

Fig. 3. A strict consensus tree from six equally parsimonious trees of *Rubiaceae*, based on restriction site occurrences (SOA), each 348 steps long, with a consistency index of 0.46 and a retention index of 0.78

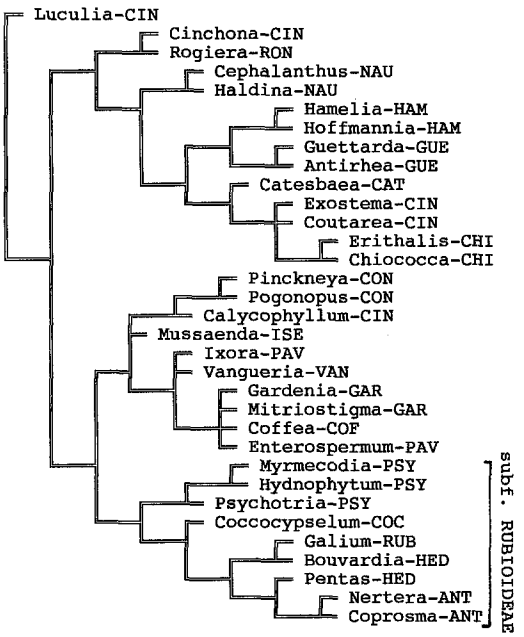

Fig. 4. A strict consensus tree from 60 equally parsimonious trees of *Rubiaceae*, based on restriction site mutations (SMA), each 291 steps long, with a consistency index of 0.53 and a retention index of 0.79

The resulting trees differ in topology. The FDA, with the two different solutions, resulted in a strict consensus tree (Fig. 1) with all branches except one, a trichotomy, resolved in dichotomies. The single FOA tree (Fig. 2) is totally resolved in dichotomies. The SOA, with the six different solutions, resulted in a strict consensus tree (Fig. 3) with one trichotomy and one tetratomy. The SMA with the 60 different

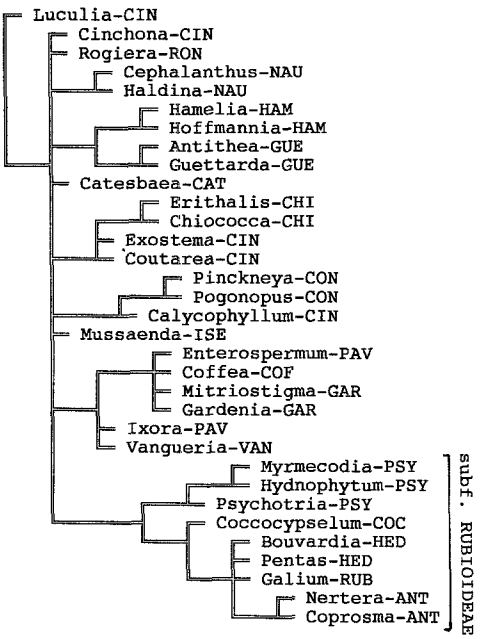

Fig. 5. A strict consensus tree for all four analyses (FOA, FDA, SOA, SMA), showing all replicated parts

possible solutions gave a consensus tree (Fig. 4) with five polytomies, of which three were trichotomies and two were tetratomies.

Despite these differences many clades were identical between the analyses. The subfamily *Rubioideae* (9 taxa) is similar in all four analyses, and many terminal clades are congruent between the four analyses. A strict consensus tree for all four analyses (Fig. 5) shows the replicated parts in all four analyses. The main differences between the trees were found in the basal branchings.

Discussion

The results from the four different scoring methods yielded solutions with different topologies, resolutions, number of trees, numbers of steps, consistency and retention index values. If we had the "true" phylogeny it would be easy to decide the best method, at least in this particular case. Instead, there are two different, indirect ways of evaluating the methods. The first is to compare the results to see whether these give indications of the quality of the methods. The other way would be to analyse inherent differences between the methods. If such inherent differences exist, they are essential to consider before a certain method is chosen.

As the number of characters differs between the analyses, the number of steps for each tree is not an adequate criterion of the quality of the methods. For similar reasons, the consistency index (CI) is also not a useful indicator of quality (ARCHIE 1989, SANDERSON & DONOGHUE 1989). The CI is strongly dependent on the degree of question-marks in the matrix (in a SMA there will always be many question-marks and thus the CI will be higher), and it may also depend on the number of characters (ARCHIE 1989, although this was not found by SANDERSON & DONOGHUE 1989). The retention index (RI) of FARRIS (1989), or the homoplasy excess ratio minimum index (HERM) of ARCHIE (1989), should not be sensitive to different numbers of characters and may thus be a better criterion when evaluating different

trees. In this study the site analyses yielded distinctly higher RI values. However, the main discrepancy between the trees concerns the basal branchings. If we consider the character distribution on the trees (BREMER & JANSEN 1991, and unpubl.), we find that few characters support these branchings, and therefore they are very sensitive to small matrix changes. Even though insufficiency of data may influence the differences between the trees, there are some important inherent differences between the methods.

A difference between the SMA and the other analyses is that SMA makes some assumptions about the presumed processes (mutations), while the others are founded on scoring of pattern only.

Concerning fragment analyses (FOA, FDA) the information from large length mutations will also be used as informative characters as fragments of equal length are considered homologous. However, one must be cautious that these length mutations will not be separated from the site mutations and thus it would be impossible to give, e.g., length mutations different weight.

Without map construction there is a risk that two fragments of equal length come from different parts of the genome. In a FDA this risk is not negligible, and I would therefore avoid the FDA method. In the *Rubiaceae* study, 30% of the initially 548 informative characters were shown to be non-homologous, e.g., the fragments were from different parts of the cpDNA molecules. However, with the rather small probes normally used in a FOA (1 to 10 kb) there is a low probability that two non-homologous fragments should have exactly the same size. In the *Rubiaceae* study, including 944 different fragments, not a single one of the fragments identified directly from the autoradiogram were shown to be non-homologous between taxa after mapping. The danger of including non-homologous fragments in FOA thus seems to be over-estimated.

Another objection for using fragments as characters is that they are not "evolutionarily" independent of each other. Dependent characters should be avoided if they mirror the same "evolutionary event", i.e. if they originated from a single mutation (deletion, insertion etc.). If some characters are dependent but treated as if they were independent, the corresponding evolutionary event is assigned a higher weight than other such events. Nevertheless, both restriction sites and fragments may be dependent characters. Consider the effect of a single 200 bp deletion affecting, say, ten different sites. Scoring each distinct fragment as an independent character multiplies the effect of this single deletion ten times in the data matrix. Similarly, if restriction sites are scored, all taxa with the deletion will share ten site losses, although all dependent on a single event, the deletion. Restriction fragments are also "dependent" in a very special way. Consider a hypothetical case where an outgroup has a 10 kb fragment and two ingroup taxa share a site that has split the former 10 kb fragment into two smaller ones, 4 and 6 kb, respectively. In SOA and SMA this site is scored as a gain of one character shared by the two taxa and it will give one step in the resulting phylogenetic tree. In FDA and FOA the same site mutation will be represented as three characters (the occurrence or not of the 10 kb, 4 kb, and 6 kb fragments, respectively) in the data matrix and on the resulting tree. Since all three characters will normally be placed at the same node, they are congruent. In the cladogram, the 10 kb fragment character will be represented by a loss in the ingroup taxa, while the 4 and 6 kb fragments will be represented as gains. If the only difference between a tree built from restriction sites and one built

from fragments is that there are always three characters instead of one in the fragment tree, the resulting phylogeny would not differ between methods. However, not all site characters are represented by three fragments. In the *Rubiaceae* study there are 548 fragments and 161 sites, differing from the expected ratio 3 : 1. There are different reasons for the quotient not being exactly 3 : 1. One is that a length mutation, ignored in site analysis, will result in two characters in a fragment analysis, (one fragment representing the unmutated state and the other the mutated fragment). Another reason is that a site mutation may not be phylogenetically informative (an autapomorphy or synapomorphy for all ingroup taxa) while the surrounding fragments are informative. If we do not have restriction maps we can not determine whether a fragment is a result of a length or a point mutation, and consequently whether characters really are dependent. Since the occurring deviations (if there are any) from the 3 : 1 ratio in the character distribution probably are randomly distributed on the resulting tree, there is no reason to suspect that these will systematically bias the results. However, a weakly supported clade may of course be affected, and this is probably the case with the basal branchings of the *Rubiaceae*.

Turning to restriction site analyses (SOA and SMA) one may note that if restriction site maps were constructed from totally sequenced DNA, we could be rather certain whether sites were homologous between taxa. However, cpDNA from two plants only have been totally sequenced (SHINOZAKI & al. 1986, OHYAMA & al. 1986), and if such sequences were available we should use these and not the restriction sites only. To construct restriction maps we have to rely on restriction enzyme digestion, fragment separation, hybridization to known cpDNA fragments, and the following estimation of where the sites belong. It is necessary to stress that the position of a site on the map is reconstructed from measurements of the surrounding fragments, so there will be no more accuracy in its position than the confidence by which the fragment lengths have been estimated.

If we construct restriction maps and use sites as characters, most problems about non-homologous or dependent characters will be eliminated since most sites are independent of each other. On the other hand, all phylogenetic information from length mutations will be lost unless we include length mutations as separate characters, which is of course possible.

What about the differences between SOA and SMA? In SOA only the pattern of restriction sites is considered (sites are treated as homologous if they occur at the same location in the genome of different taxa). In SMA also information about the inferred process, mutation, is used when scoring a character, and if we suspect different processes we would not treat them as homologous. In SOA all characters are scored, independently of each other, as 0 or 1. In SMA analysis all question-marks (?) in the matrix will be "scored by the program" as 0 or 1 in the most parsimonious way for that particular analysis. This means that the other characters determine the "scoring". Different length mutations or base substitutions (processes) can result in the same pattern of restriction sites or fragments. We should avoid stating a priori the processes, since they belong to the possible explanations which we can use if we want to interpret homoplasy in a resulting cladogram. I would avoid SMA for these reasons.

In conclusion, the methods differ in the degree of problems with determining homology, dependent characters, and time-consumption, but none of the methods

will systematically bias the resulting cladograms. However, the methods probably differ in accuracy. When choosing between FOA and SOA I think one should consider the time and expenses. The SOA is associated with the smallest risk of error, and is thus probably most accurate. However, length mutations are usually omitted and therefore potential information is lost. Furthermore, it is a more time-consuming method than FOA and thus more expensive. In a field like systematic botany, where a vast number of taxonomical groups should be investigated, perhaps a cheaper or faster method could be accepted as long as we are aware of the weakness of the method. So far, I think most students of cpDNA for phylogenetic reconstruction have used SOA, although the method used is not clearly stated. However, as the scoring method may influence the resulting hypothesis of phylogeny, it is absolutely necessary that each published restriction analysis is accompanied by a data matrix and contains information on how the matrix was compiled. Without such additional information, it is not possible to determine the reliability, and hence value, of a study.

I am grateful to K. BREMER, O. ERIKSSON, B. MISHLER, H.-E. WANNTORP, and one anonymous reviewer for constructive criticism on the manuscript. This study was financially supported by the Swedish Natural Science Research Council.

References

ARCHIE, J. W., 1989: Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. – Syst. Zool. **38**: 253–269.

ALBERT, V. A., MISHLER, B. D., CHASE, M. W., 1991: Character-state weighting for restriction site data in phylogenetic reconstruction, with an example from chloroplast DNA. – In SOLTIS, D., SOLTIS, P., DOYLE, J., (Eds.): Plant molecular systematics. – New York: Chapman & Hall (in press).

BERTHOU, T., MATTHIEU, C., VEDEL, F., 1983: Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. – Theoret. Appl. Genet. **65**: 77–84.

BREMER, B., JANSEN, R. K., 1991: Comparative restriction site mapping of chloroplast DNA implies new phylogenetic relationships within *Rubiaceae*. – Amer. J. Bot. (in press).

CLEGG, M. T., BROWN, A. H. D., WHITFELD, P. R., 1984: Chloroplast DNA diversity in wild and cultivated barley: implication for genetic conservation. – Genet. Res. **43**: 339–343.

COATES, D., CULLIS, C. A., 1987: Chloroplast DNA variability among *Linum* species. – Amer. J. Bot. **74**: 260–268.

DEBRY, R. W., SLADE, N. A., 1985: Cladistic analysis of restriction endonuclease cleavage maps within a maximum-likelihood framework. – Syst. Zool. **34**: 21–34.

ERICKSON, L. R., STRAUS, N. A., BEVERSDORF, W. D., 1983: Restriction patterns reveal origins of chloroplast genomes in *Brassica* amphiploides. – Theoret. Appl. Genet. **65**: 201–206.

FARRIS, J. S., 1977: Phylogenetic analysis under Dollo's law. – Syst. Zool. **26**: 77–88.

– 1988: Hennig 86. Version I.5 [computer software and manual]. – Port Jefferson Station, New York: (publ. by author).

– 1989: The retention index and the rescaled consistency index. – Cladistics **5**: 417–419.

HANTULA, J., UOTILA, P., SAURA, A., LOKKI, J., 1989: Chloroplast DNA variation in *Anemone* s. lato (*Ranunculaceae*). – Pl. Syst. Evol. **163**: 81–85.

HENNIG, W., 1966: Phylogenetic systematics. – Urban: University of Illinois Press.
HOSAKA, K., 1986: Who is the mother of the potato? – Restriction endonuclease analysis of chloroplast DNA of cultivated potatoes. – Theoret. Appl. Genet. **72**: 606 – 618.
– OGIHARA, Y., MATSUBAYASHI, M., TSUNEWAKI, K., 1984: Phylogenetic relationship between the tuberous *Solanum* species as revealed by restriction endonuclease analysis of chloroplast DNA. – J. Genet. (Japan) **59**: 349 – 369.
JANSEN, R. K., PALMER, J. D., 1987a: Chloroplast DNA from lettuce and *Barnadesia* (*Asteraceae*): structure, gene localization, and characterization of a large inversion. – Curr. Gent. **11**: 553 – 564.
– – 1987b: A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (*Asteraceae*). – Proc. Natl. Acad. Sci. U.S.A. **84**: 5818 – 5822.
– HOLSINGER, K. E., MICHAELS, H. J., PALMER, J. D., 1990: Phylogenetic analysis of chloroplast DNA restriction site data at higher taxonomic levels: an example from the *Asteraceae*. – Evolution (in press).
KUNG, S. D., ZHU, Y. S., CHEN, K., 1982: *Nicotiana* chloroplast genome. III. Chloroplast DNA evolution. – Theoret. Appl. Genet. **61**: 73 – 79.
LEHVÄSLAIHO, H., SAURA, A., LOKKI, J., 1987: Chloroplast DNA variation in the grass tribe *Festuceae*. – Theoret. Appl. Genet. **74**: 298 – 302.
NEI, M., LI, W. H., 1979: Mathematical model for studying genetic variation in terms of restriction endonucleases. – Proc. Natl. Acad. Sci. U.S.A. **76**: 5269 – 5273.
OHYAMA, K., FUKUZAWA, H., KOHCHI, T., SHIRAI, H., SANO, T., SANO, S., UMESONO, K., SHIKI, Y., TAKEUCHI, M., CHANG, Z., AOTA, S., INOKUCHI, H., OZEKI, H., 1986: Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. – Nature **322**: 572 – 574.
PALMER, J. D., 1985: Comparative organization of chloroplastomes. – Ann. Rev. Genet. **19**: 325 – 354.
– 1987: Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. – Amer. Naturalist **130**: S 6 – S 29.
– JANSEN, R. K., MICHAELS, H. J., CHASE, M. D., MANHART, J. A., 1988: Chloroplast DNA variation and plant phylogeny. – Ann. Missouri Bot. Gard. **75**: 1180 – 1206.
– JORGENSEN, R. A., THOMSON, W. F., 1985: Chloroplast DNA variation and evolution in *Pisum:* patterns of change and phylogenetic analysis. – Genetics **109**: 195 – 213.
– ZAMIR, D., 1982: Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. – Proc. Natl. Acad. Sci. U.S.A. **79**: 5006 – 5010.
PERL-TREVES, R., GALUN, E., 1985: The *Cucumis* plastome: physical map, intrageneric variation and phylogenetic relationships. – Theoret. Appl. Genet. **71**: 417 – 429.
RIESEBERG, L. D., SOLTIS, D., PALMER, J. D., 1988: A molecular reexamination of introgression between *Helianthus annus* and *H. bolanderi* (*Compositae*). – Evolution **42**: 227 – 238.
ROSE, R. J., SCHLARBAUM, S. E., SMALL, E., JOHNSON, L. B., 1988: Chloroplast genomic variation and phylogeny in *Medicago* section *Intertextae*. – Canad. J. Bot. **66**: 1352 – 1358.
SANDERSON, M. J., DONOGHUE, M. J., 1989: Patterns of variation in levels of homoplasy. – Evolution **43**: 1781 – 1795
SHINOZAKI, K., OHME, M., TANAKA, M., WAKASUGI, T., HAYASHIDA, N., MATSUBAYASHI, T., ZAITA, N., CHUNWONGSE, J., OBOKATA, J., YAMAGUCHI-SHINOZAKI, K., OHTO, C., TORAZAWA, K., MENG, B. Y., SUGITA, M., DENO, H., KAMOGASHIRA, T., YAMADA, K., KUSUDA, J., TAKAIWA, F., KATO, A., TOHDOH, N., SHIMADA, H., SUGIURA, M., 1986: The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. – EMBO J. **5**: 2043 – 2049.
SMITH, R. L., SYTSMA, K. J., 1990: Evolution of *Populus nigra* (sect. *Aigeiros*): introgressive hybridization and the chloroplast contribution of *Populus alba* (sect. *Populus*). – Amer. J. Bot. **77**: 1176 – 1187.

SOLTIS, D. E., SOLTIS, P. S., NESS, B. D., 1989: Chloroplast-DNA variation and multiple origins of autopolyploidy in *Heuchera micrantha* (*Saxifragaceae*). – Evolution **43**: 650–656.

SORENG, R. J., DAVIS, J. I., DOYLE, J. J., 1990: A phylogenetic analysis of chloroplast DNA restriction site variation in *Poaceae* subfamily *Pooideae*. – Pl. Syst. Evol. **172**: 83–97.

SWOFFORD, D. L., 1989: PAUP, Phylogenetic analysis using parsimony. Version 3.0 [computer software and manual]. – Champaign, Illinois: Illinois Natural History Survey.

SYTSMA, K. J., GOTTLIEB, L. D., 1986 a: Chloroplast DNA evidence for the derivation of the genus *Heterogaura* from *Clarkia* (*Onagraceae*). – Proc. Natl. Acad. Sci. U.S.A. **83**: 5554–5557.

– – 1986 b: Chloroplast DNA evolution and phylogenetic relationships in *Clarkia* sect. *Peripetasma* (*Onagraceae*). – Evolution **40**: 1248–1262.

– SCHAAL, B. A., 1985: Phylogenetics of the *Lisianthus skinneri* (*Gentianaceae*) species complex in Panama using DNA restriction fragment analysis. – Evolution **39**: 594–608.

– SMITH, J. F., 1988: DNA and morphology: comparisons in the *Onagraceae*. – Ann. Missouri Bot. Gard. **75**: 1217–1237.

– – GOTTLIEB, L. D., 1990: Phylogenetics in *Clarkia* (*Onagraceae*): restriction site mapping of chloroplast DNA. – Syst. Bot. **15**: 280–295.

TEMPLETON, A. R., 1983: Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. – Evolution **37**: 221–244.

WAUGH, R., VAN DE VEN, G. T. W., PHILLIPS, M. S., POWELL, W., 1990: Chloroplast DNA diversity in the genus *Rubus* (*Rosaceae*) revealed by Southern hybridization. – Pl. Syst. Evol. **172**: 65–75.

Address of the author: Dr BIRGITTA BREMER, Department of Systematic Botany, Uppsala Universtiy, Box 541, S-75121 Sweden.

Accepted October 29, 1990 by F. EHRENDORFER