Total evidence requires exclusion of phylogenetically misleading data

Guillaume Lecointre & Pierre Deleporte

Accepted: 16 February 2004

Lecointre G. & Deleporte P. (2004). Total evidence requires exclusion of phylogenetically misleading data. — *Zoologica Scripta*, *34*, 101–117.

Treating all available characters simultaneously in a single data matrix (i.e. combined or simultaneous analysis) is frequently called the 'total evidence' (TE) approach, following Kluge's introduction of the term in 1989, quoting Carnap (1950). However, the general principle and one of the possible procedures involved in its application are often confused. The principle, first enunciated within the context of inductive logic by Carnap in 1950, did not refer to a particular procedure, and TE meant using all relevant knowledge, rather than a combined analysis of all available data. Using TE, all relevant knowledge should be taken into account, including the fact that some data are probably misleading as indicators of species phylogeny and should be discarded. Based on the assumption that molecular partitions have some biological significance (process partitions obtained from nonrandom homoplasy or from 'processes of discord'), we suggest that separate analyses constitute an important exploratory investigation, while the phylogenetic tree itself should be produced by a final combined analysis of all relevant data. Given that the concept of process partitions is justified and that reliability cannot be evaluated using any robustness measure from a single combined analysis, the analysis of multiple data sets involves five steps: (1) perform separate analyses without consensus trees in order to assess reliability of clades through their recurrence and improve the detection of artifacts; (2) test significance of character incongruence, using, for example, pairwise ILD tests in order to identify the sets responsible for incongruence; (3) replace likely misleading data with question marks in the combined data matrix; (4) perform simultaneous analysis of this matrix without the misleading data; (5) assess the reliability of clades found by the combined analysis by computing their recurrence within the previous separate analyses, giving priority to repeatability.

Guillaume Lecointre, UMR 7138 CNRS, département 'Systématique et Evolution', Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris cedex 05, France. E-mail: lecointr@mnhn.fr
Pierre Deleporte, UMR 6552, CNRS et Université de Rennes I, Station Biologique de Paimpont, F-35380 Paimpont, France. E-mail: pierre.deleporte@University-rennes1.fr

Total evidence: What's in a word?

Kluge's (1989) introduction of the term 'total evidence' stimulated several years of debate within the field of systematics, as a result of which new tools were formulated to evaluate character congruence (e.g. Farris *et al.* 1995). The debate was not without its ambiguities. Kluge had borrowed the term from Carnap's (1950) *Logical Foundations of Probability*, where it was used within the context of inductive logic. Later on Kluge argued for a Popperian hypothetico-deductive approach to phylogenetic inference, as is clear from his later papers on the topic (e.g. Kluge 1997; Kluge & Wolf 1993; Kluge 2003). While reference to Popper rather than Carnap would probably have been more appropriate in the seminal 1989 paper (Kluge, pers. comm.), Carnap's concept has been very useful for systematists.

While debates about character congruence vs. taxonomic congruence have continued to be plagued by differing interpretations of TE, the Popperian approach to systematics remains highly topical (Sober 1988; Rieppel 2003a,b), and strong epistemological arguments have been proposed in favour of an abductive and non-Popperian notion of 'testability' in phylogenetic inference (Fitzhugh 1997, 1998; Geiger *et al*. 2001; Rieppel 2003a,b). By analysing all the characters together in a single data matrix and calling it TE, many systematists have tended to confuse the procedure and the underlying principle. This ambiguity was implicit from the beginning. In the 1989 paper Kluge (1989: 8) used TE to mean all the available evidence in the text, while in the captions of his figs 1 and 2 his formulations jumped from the general principle to the particular procedure, with TE

applied to the procedure of considering all available data in a single matrix. He did so again, four years later (Kluge & Wolf 1993: 184, fig. 2):

> 'Taxonomic congruence involves partitioning evidence into separate data sets (…), seeking the best-fitting hypothesis for each data set (…), and deriving a consensus of those topologies. Alternatively, total evidence uses character congruence to find the best-fitting phylogenetic hypothesis for an unpartitioned set of synapomorphies, which, ideally, is all of the relevant available data (…)'.

Eernisse and Kluge (1993) also compared taxonomic congruence with TE rather than with character congruence, this time as two competing paradigms: 'Taxonomic congruence and total evidence are competing paradigms in phylogenetic inference'.

We suggest that taxonomic congruence could be contrasted with character congruence as two different approaches: optimally combining the information from different cladograms versus different characters. Various useful terms have been proposed which avoid confusion when naming the different procedures. For instance, separate analysis (Nixon & Carpenter 1996) is used when assessing taxonomic congruence (Mickevich 1978; Kluge 1989); this involves partitioning the data into several matrices and analysing them separately. Combined analysis (Bull *et al.* 1993) or simultaneous analysis (Nixon & Carpenter 1996) are used when assessing character congruence (Kluge 1989), with all the available data treated in a single matrix.

In this paper we do not use TE to describe procedures; rather, we follow Carnap in limiting its use to mean the principle, in the context of inductive logic. As we are dealing with phylogenetic inference, this principle is more precisely applied within the context of abduction (i.e. retro-description of the past, given our knowledge of the result of the evolutionary process and implementing some notion of this process as the explanatory law; Fitzhugh 1997; 1998; Geiger *et al.* 2001). We see no reason why TE should not apply both to abductive logic and to any scientific explanation as well.

Carnap (1950) recommended following the principle (using all the relevant evidence available at a given time) but said nothing about how to apply it (which way to use all the evidence, much less all the data to hand). When applying inductive logic within a given situation, the numerical value of a sample's statistical probability can become the value of the degree of confirmation (Carnap's *logical probability*) only if all of the relevant evidence available at the time has been taken into account. In our view, for the sake of clarity, this notion of TE should be opposed to the notion of 'partial evidence', and not to particular procedures or approaches like taxonomic congruence. TE should not be identified with character congruence, as frequently seen in the literature.

The present position requires a clarification of what is meant by evidence. Carnap (1950: 211) suggested that the degree of confirmation (also called the logical probability) c of a hypothesis h given the evidence e is a given real number r: $c(h, e) = r$, and that the degree of confirmation c has the numerical value r only if the total evidence available has been taken into account:

> 'If e expresses the total knowledge of X at the time t (…) X is justified at this time to believe h to the degree r, and hence to bet on h a betting quotient not higher than r. (…) The total evidence available must be taken as basis for determining the degree of confirmation. (…) The mere fact that X knows e does not entitle him to believe h to the degree r; obviously it is required either that X knows nothing beyond e or that the totality of his additional knowledge i be irrelevant for h with respect to e, i.e. that it can be shown in inductive logic that $c(h, e.i) = c(h, e)$'.

It is clear that for Carnap knowledge constituted both the relevant evidence e as well as the irrelevant evidence i. If h represents a phylogenetic hypothesis, we have to decide what kind of knowledge should be put in i in order to concentrate exclusively on e. The question of neutral knowledge requires little comment, while relevant evidence is empirical knowledge which has the ability to alter the probability of a statement. The problem is less obvious when using such data as inheritable characters, all of which are *a priori* potentially informative and can thus be used in the absence of contrary evidence. However, characters which have undergone 'processes of discord' (Maddison 1997) or aberrant rates of change will likely be misleading for inferring phylogeny.

The knowledge that some characters are probably misleading constitutes perfectly relevant knowledge for phylogenetic analysis. While the data to be excluded have the power to change the probability of a statement, so does the additional knowledge pleading for their exclusion. In the same way, the knowledge that some characters are strictly dependent on others is relevant and should lead to their exclusion as redundant. Rieppel (2003a: 270) emphasized that even Popper included background knowledge within TE in his definition of degree of confirmation: '(…) the total evidence e is to be partitioned into y (new observations) and z (background knowledge); and y and z should be chosen as to give $C(x, y, z)$ the highest value possible for x (the hypothesis or the theory), on the available total evidence'.

Contrary to a number of supporters of the TE approach *sensu* Kluge (1989), who claimed that no additional knowledge beyond the data matrix should be taken into account for phylogeny inference, we believe that the reasons for a data matrix being designed in a particular way and the corresponding background knowledge should always be made explicit. The argument can thus be made in terms of the processes of character evolution and the indicators of such processes.

Table 1 Terminology used in this paper.

Total evidence/partial evidence	Total knowledge (Carnap 1950), as opposed to partial knowledge
Requirement of total evidence	Principle of using all available relevant knowledge in inductive logic (Carnap 1950)
Relevant evidence/irrelevant evidence	Relevant knowledge: relevant evidence that should be used possibly discarding irrelevant evidence *i* (Carnap 1950)
Taxonomic congruence/character congruence	Two different approaches to phylogeny reconstruction (Kluge 1989)
Separate analyses/simultaneous analysis	Procedures for performing the two previous approaches (Nixon & Carpenter 1996)

Following this abductivist rather than purportedly hypothetico-deductivist logic, we propose that two kinds of data be excluded before implementing a simultaneous analysis: (1) those responsible for significant character incongruence, thus excluding the misleading products of a process incompatible with the history of the taxa; (2) those that obviously provoked tree-reconstruction artifacts detected in separate analyses (which probably underwent an aberrant rate of evolution). We thus discard data considered to obscure the history of the taxa. The protocols for identifying likely misleading data will be described in the next section.

Using TE to mean all available *data* instead of all available *knowledge* would merely obscure the possibility of legitimately performing such data exclusion on the basis of the knowledge that some data probably have nothing to do with the history of the taxa. It is thus important to distinguish between principles and their possible procedures, between the requirement of TE and the procedures for analysing the data. Nelson (1979) and Miyamoto & Fitch (1995), when recommending separate analyses, probably considered that they were in fact following the principle of using TE (i.e. they did not *deliberately* discard relevant knowledge or data) while using the procedure they thought to be the best. The utility of different terms for the methodological debate is obvious. One could see the reasoning underlying Nixon & Carpenter's (1996: 223) proposal of a change in the terms:

> 'The term (total evidence) is probably not appropriate to contrast the method of separate analyses of partitioned data followed by consensus of results with the method of simultaneous analysis of multiple combined data sets (…) We therefore prefer to use the terminology of "simultaneous analysis" in place of 'total evidence' following Nixon and Carpenter (1993)'.

According to Carnap (1950), relevant evidence is relevant knowledge, not merely relevant data. In order to provide contrast between simultaneous analysis and separate analyses, we compare the procedures by which total relevant evidence for phylogenetic inference is taken into account at the level of data analysis, notwithstanding its use for data delineation and choice.

To summarize, in this paper TE is used *sensu* Carnap (1950: 211), i.e. all relevant knowledge (Popper in Rieppel, 2003a: 270) (see Table 1). All available potentially useful data are considered, despite the fact that some *a priori* putatively informative data are probably misleading and should be discarded. As for procedures, we use the terms 'simultaneous analysis' and 'separate analyses' following Nixon & Carpenter (1996). We consider two questions. First, what kind of data can be excluded from the analysis without violating the principle of TE? Second, if there is some justification for excluding some kinds of data, which procedure are we inclined to choose?

Excluding misleading data

In this section we discuss three assumptions.

(1) Genes are natural classes relevant to phylogenetic analysis, because they can generally be considered functionally and hence evolutionarily independent, and thus are likely to constitute independent markers of phylogenetic relationships. The proteins they encode are integrated autonomous physical entities, which are subject to peculiar selective pressures. A gene is never chosen at random and we always know something about its role; hiding that knowledge or pretending to ignore it is ill conceived at best.

(2) Misleading signals due to horizontal transfers or changes in selective regimes across taxa in molecular data sets are in the minority. If this assumption is not accepted, trees are useless for phylogenetic inference because noise and misleading signal would be overwhelming. As a consequence, one could favour networks for representing interrelationships, or could prefer that the tree maximizes homoplasy instead of optimizing character congruence. It is not stated often enough in phylogenetics that it is an implicit assumption that horizontal transfers are in the minority of occurrences among genes and across taxa.

(3) Separate analyses are an exploratory step during which no data are excluded. Otherwise we would never summarize our results and obtain the final degree of confirmation of a phylogenetic inference.

Molecular data may contain non-random homoplasies that prevent recovering the phylogeny of taxa. Combining all the available data right from the beginning without discrimination involves the risk of biased reconstruction. Bull *et al.* (1993) explained the circumstances when it could be misleading to combine data:

> 'Molecular studies of viruses and bacteria have revealed cases of horizontal gene transfer, i.e. a small portion of the genome of one "species" (species A) has replaced the homologous portion of the genome in another species (B) (…). The evolutionary history of the species B genome differs for different genes, and reconstruction of the evolutionary histories of species A and B would depend on which characters were analysed. No rational systematist would suggest combining genes with different histories to produce a single reconstruction, because combining the data not only obscures an important feature of history but runs the risk of producing a reconstruction that fails to represent either history'.

We agree that these processes are sources of potentially misleading signal, but there is no reason to rule out data combination in general, provided that the misleading data have been identified and excluded. The problem is how to identify them. Special evolutionary processes that disrupt the phylogeny of markers from the phylogeny of species are called 'processes of discord' by Maddison (1997) and are listed in detail in Doyle (1992, 1997) and Maddison (1997). They are horizontal transfers (e.g. bacterial recombination or introgressive hybridization in plants or teleosts), deep coalescence, and gene duplication/extinction. In such cases, taxonomic incongruence between trees based on different genes is not due to tree reconstruction artifacts but to an erroneous hypothesis of orthology affecting one or more gene copies. If processes of discord have occurred and affect at least one tree, taxonomic incongruence is not due to tree reconstruction errors and thus can be called 'legitimate' and should be supported by significant character incongruence. The tree is 'correct', although simply misinterpreted.

Testing for heterogeneity of 'process partitions' (Bull *et al.* 1993) or 'linkage partitions' (Slowinski & Page 1999) to 'determine whether the rules have been different' (Bull *et al.* 1993) is the 'prior agreement approach' of Chippindale & Wiens (1994), also called 'conditional combination' (Bull *et al.* 1993; Huelsenbeck *et al.* 1996). In other words, it is testing for character incongruence (Farris *et al.* 1995) prior to performing simultaneous analysis (if allowed). If it can be demonstrated that the characters of data set A are significantly incongruent with those of set B (rejection of the null hypothesis of character congruence), simultaneous analysis should be avoided (despite the contrary advice of some authors, Wiens 1998).

'Process partitions' have experienced different evolutionary processes (Bull *et al.* 1993), and should be distinguished from the 'linkage partitions' of Slowinski & Page (1999), which result from 'processes of discord'. Detection of the latter requires prior definition of process partitions. But, process partitions can be evoked without process of discord, for instance, when homoplasy is heterogeneously distributed among genes. Bull *et al.* (1993) argued for conditional combination after checking whether or not these processes led to significantly different histories. It is striking to observe how right they were, when one really finds character incongruence between partitions due to strongly different processes of homoplasy accumulation in each partition and not to the 'processes of discord' listed by Doyle (1992, 1997) and Maddison (1997).

The possibility is now widely admitted in molecular systematics that homoplasy in a particular gene can affect all the characters (all the positions) or a number of them in the same way, rendering any robustness indicator positively misleading for some clades. For instance, a higher GC content in gene X in several unrelated taxa will be responsible for an apparently robust clade grouping them in the molecular phylogeny based on that gene, while no grouping of these taxa would be supported by genes Y and Z. The molecular systematist is often led to recognize the fact that genes are natural partitions in terms of the way homoplasy accumulates within them. For example, the way homoplasy is stocked in cytochrome *b* sequences (Hassanin *et al.* 1998) is different from the way it is stocked in a nuclear ribosomal gene (Philippe *et al.* 1996) or in a coding nuclear one like rbcL (Sennblad & Bremer 2000). Note that the metaphorical use of the term 'homoplasy' here could be thought to be essentialist. In fact, this is simply a way to express the fact that these genes are under very different selective pressures. There is nothing new here, except considering the process partitions of Bull *et al.* (1993) in terms of nonrandom homoplasy. It cannot be excluded that processes at work beyond the detected homoplasy can lead to significant incongruence (Sullivan 1996). For instance, there can be more incongruence between codon positions of the same gene than between analogous codon positions of two different genes (Vidal & Lecointre 1998). This question requires further detailed exploration.

Even without evoking processes of discord, separate analyses could be viewed as a heuristic step to discover tree-reconstruction artifacts provoked by changes in sequence selective regime across taxa. Fruitful use of separate analysis as a tool to discover artifacts is found in Philippe & Adoutte (1998), Philippe & Laurent (1998), Moreira *et al.* (1998), Germot & Philippe (1999) and Philippe *et al.* (2000). They allowed long-branch misplacement in molecular phylogenies, whereas it was not possible to reliably detect them from a single tree. Another example is the reanalysis by Page and Charleston (1999) of the mammalian sequences of Allard and Carpenter (1996). However, the tree resulting from simultaneous analysis can be biased by homoplasy concentrated in only one of the data sets (Chen *et al.* 2000, 2001, 2003). Absence of previous separate analyses would have prevented this diagnosis, as already suggested by Grande (1994):

> 'Since we have no way of demonstrating that all characters are equivalent (any more than we can demonstrate that

Table 2 The four possible situations of multiple data sets analysis with regard to taxonomic incongruence (TI) and character incongruence (CI). TI is dissociated from CI, because it is possible to detect TI without detecting significant CI using recently developed tests such as ILD.

	Taxonomic incongruence	Taxonomic congruence
Significant CI	Processes of discord or positively misleading homoplasy	Coincidence
No significant CI	Tree reconstruction artifacts at work	Corroboration

> they are not), we should not allow a single conflicting pattern from one type of data to obscure a pattern repeated by several other types of data. (…) Any pattern obtained through (taxonomic) congruence is a potential tool in pattern/process studies'.

Recognition of these facts, at least in the field of molecular systematics, leads to recognition of the following: (1) the naturalness of partitions when they are genes, (2) the usefulness of separate analyses both as an aid to detecting the impact of differential nonrandom homoplasy on trees and as a guide on what to delete, according to criteria detailed below.

In recent years, new tools have been actively developed for testing significance of incongruence (Rodrigo *et al.* 1993; Farris *et al.* 1995; Huelsenbeck & Bull 1996; Page & Charleston 1998; Templeton 1983; Larson 1994; Cunningham 1997a,b; Slowinski & Page 1999; see Huelsenbeck *et al.* 1996, for a review). As a consequence, character congruence has been more clearly distinguished from taxonomic congruence than it had been previously. The two concepts were not clearly separated, operationally speaking, before the rise of algorithms allowing conditional combination. The practical separation can now be understood, as illustrated in Table 2, and comes from the fact that taxonomic incongruence can be obtained from noisy data for which the null hypothesis of congruence cannot be rejected with the statistical tools now available. Significant character incongruence is due to processes of discord and possibly to highly constrained homoplasy (for instance a strong decrease of the mutational space in a sequence of some unrelated taxa), while taxonomic incongruence is merely due to the existence of process partitions.

If the source of character incongruence can be identified, there are three possible strategies to infer the phylogeny of taxa. Either no combination of partitions at all, as advocated by many authors (e.g. Bull *et al.* 1993; Miyamoto & Fitch 1995; Slowinski & Page 1999), or data combination despite significant character incongruence (Wiens 1998), or combination with removal of the data responsible for incongruence that are likely to be misleading. The third strategy is explored at length in this paper.

The first strategy should be followed when processes of discord have been so numerous that it is impossible to identify the source of incongruence. It was applied, for instance, when the phylogeny of *Escherichia coli* was inferred from genes that underwent numerous horizontal transfers, like plasmidic genes (*finO*, *traD*) or from genes close to hot spots of recombination and therefore subject to hitch-hiking, like *gnd* (Fig. 1, Lecointre *et al.* 1998).

The second strategy should be avoided, as will be argued below. The third strategy requires identification of a source of incongruence. This may be possible by performing ILD tests (Mickevich & Farris 1981; Farris *et al.* 1995) with species jackknifing, i.e. iterative removal of single taxa, each followed by ILD test (as in Lecointre *et al.* 1998; Johnson *et al.* 2001; Escobar-Paramo *et al.* 2003). The taxon, the removal of which provokes an increase in *P*-value above the 5% threshold, is identified as the one responsible for incongruence. Visual inspection of sequences permits identification of a transferred gene (Fig. 2, Lecointre *et al.* 1998; Escobar-Paramo *et al.* 2003) if punctual homoplasy is low. It is thus possible to propose a simultaneous analysis based on a modified data matrix where the likely foreign gene is replaced by question marks in the taxon where it has been transferred.

It is necessary here to explain in further detail the process of detection of misleading data. Let us suppose that pairwise ILD tests reject the null hypothesis of congruence between the sequence data of genes X, Y and Z, and that for each pairwise comparison, iterative removals of taxa followed by ILD tests (Fig. 2, Lecointre *et al.* 1998) allow us to identify the taxon responsible for incongruence (e.g. taxon alpha). In all cladograms except the one based on X, alpha belongs to the same clade. Visual inspection of sequences and cladograms allows us to identify the transferred stretch of DNA in alpha. The Y, Z, T sequences exhibit synapomorphies of the putative clade to which it probably belongs, while X is a foreign sequence that exhibits none of the synapomorphies (in Fig. 2, the aceK sequence of *E. coli* strain ECOR10A does not belong to group A; Lecointre *et al.* 1998). This stretch of DNA is evidence of a horizontal transfer, for it exhibits synapomorphies of another clade; it tells us about the story of gene X, not the story of taxon alpha (see also Matic *et al.* 2002). Visual inspection of sequences helps only if punctual stochastic homoplasy is low. This protocol permits identification of the misleading data, which can be removed from the simultaneous analysis and replaced with question marks. This third strategy is possible as long as processes of discord are few (for instance, introgressive hybridization in cyprinids or bacterial housekeeping genes' sequence data sets (Milkman 1997, Fig. 2), and punctual random homoplasy is low. This limitation is due to the fact that the ILD test rapidly decreases in power when homoplasy increases (Cunningham 1997a,b; Dolphin *et al.* 2000; Darlu and Lecointre 2002) under certain circumstances analysed in Darlu and Lecointre (2002).

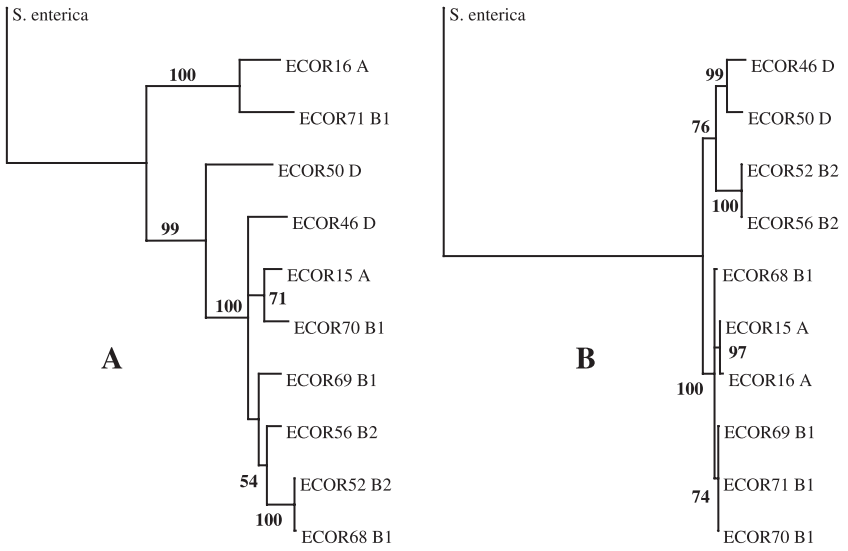


Fig. 1 A, B. Taxonomic incongruence obtained from strongly incongruent characters in the phylogeny of ECOR strains of *Escherichia coli* (Lecointre *et al.* 1998). Only the 11 taxa common to the four genes are sampled here. Strict consensus of all trees obtained using the Branch and Bound search algorithm of PAUP (Swofford 1999). —A. Consensus of two trees obtained from 1335 positions of the locus *gnd*. Number of informative sites = 252, tree length = 0.637, CI = 0.76, RI = 0.66. —B. Consensus of three trees obtained from 3360 positions of three loci of the tryptophane operon, *trpA*, *trpB* and *trpC*. Number of informative sites = 184, tree length = 922, CI = 0.91, RI = 0.77. Branch lengths have been reported under ACCTRAN optimization. Numbers above nodes are bootstrap proportions obtained from 1000 iterations. The *trp* data (B) provide high bootstrap support for three of the four traditional main strain groups (A, D, B2), while the *gnd* tree (A) is completely scrambled: traditional groups A, B1, D, B2 are all split, with high bootstrap support. The gene *gnd* is known to have experienced numerous horizontal transfers through hitch-hiking, as it is localized close to the O antigen gene complex in the bacterial chromosome (Bisercic *et al.* 1991; Nelson & Selander 1994).

It is also possible to remove the genes that exhibit aberrant rates of evolution, provided that the long-branch misplacement artifact (Felsenstein 1978; Huelsenbeck 1997; Siddall 1998) they provoked has previously been detected by performing separate analyses (Fig. 3A). Suppose that separate analyses of sequence data of genes X, Y and Z are performed for the same taxa. Taxon alpha has an aberrant position in tree Z (e.g. *Labeo* in Fig. 3A) with a very long-terminal branch length, different from the position found in tree X (Fig. 3C) and tree Y (Fig. 3D). When separate analyses show unambiguously that a particular mutation rate acceleration obscures the relationships of the taxon, the mainly random signal carried by Z in alpha can be considered as irrelevant at best, or even misleading with regard to the history of alpha.

Separate vs. simultaneous analyses

Many authors have listed advantages and drawbacks of both separate and simultaneous analyses (e.g. de Queiroz *et al.* 1995; Miyamoto & Fitch 1995). To explain and justify the protocol proposed here, a brief review of the arguments seems useful.

Advantages of separate analyses

The main usefulness of separate analyses is as a heuristic tool; they discover repeating patterns and test the properties of each data set. Many authors concur (e.g. Mickevich 1978; Nelson 1979; Grande 1994; Hillis 1995: 11; Page & Holmes 1998: 214). Larson (1994) has reminded us that in fact everybody performs separate analyses before combining. Kluge (1989: 12) considered the sequential performing of separate and simultaneous analyses to be of critical importance:

> 'The methods I use to test for character congruence require the analysis of separate biochemical and morphological data sets; however, the two sets of evidence are combined in assessing the phylogeny of *Epicrates*'.

Performing a simultaneous analysis right from the beginning is of little methodological value. The power of the taxonomic congruence test lies in the fact that the probability of obtaining by chance the same or even similar trees for the same organisms using the different independent data sets is 'vanishingly small for any reasonable number of species' (Page & Holmes, 1998), as a consequence of the very large number of possible phylogenies: 'If different data sets give us similar trees this gives us confidence that both reflect the same underlying cause'.

Nelson (1979) and Grande (1994) stated that the congruence of inferences separately drawn from independent data sets indicates increased support that the clades are likely to be

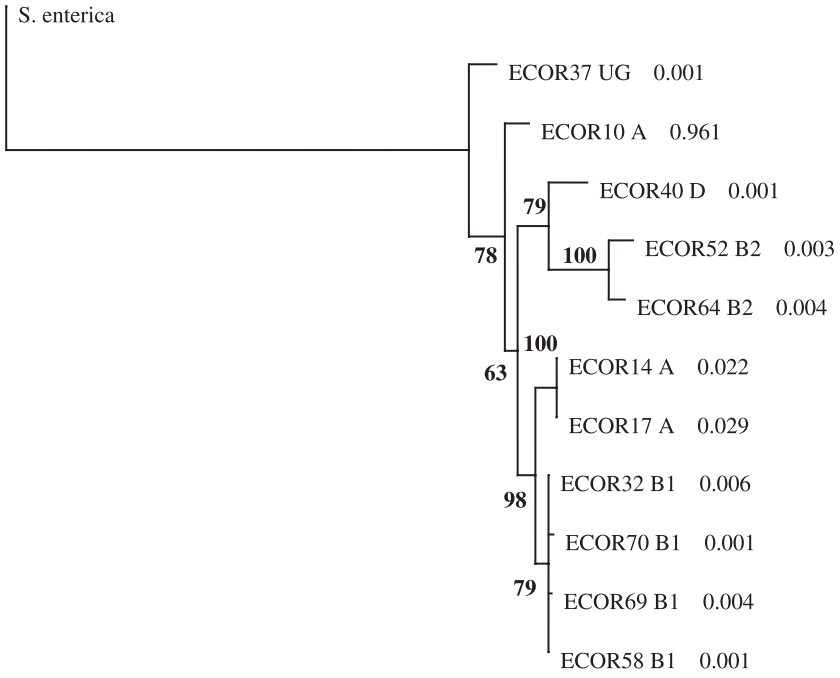


Fig. 2 Strict consensus of four trees obtained using the Branch and Bound algorithm from 1722 positions of the locus *aceK* of the ECOR strains of *Escherichia coli* (Lecointre *et al.* 1998). Nnumber of informative sites = 124, tree length = 482, CI = 0.84, RI = 0.68. Branch lengths have been reported under ACCTRAN optimization. Numbers above nodes are bootstrap proportions obtained from 1000 iterations. The ILD test for incongruence between the *aceK* locus and the 'whole genome data set' produces a *P*-value of 0.002, which is significant; the set contains 320 binary coded characters from allozymes, RAPDs and rrn RFLPs (for a description see Herzer *et al.* 1990; Desjardins *et al.* 1995; Lecointre *et al.* 1998). Numbers on the right are the *P*-values obtained when the ILD test is performed in the absence of the corresponding taxon. It is clear that ECOR10A is the sequence responsible for significant incongruence, since its removal increases the *P*-value above 0.05.

genuine. However, in addition to independence, the distribution and properties of homoplasy must be considered. The fact that a given clade is corroborated by separate analyses, despite various modalities of noise accumulation across genes and the associated various risks of errors in trees, is an additional indication of its reliability.

The complementary proposition of this simple statement is that points of disagreement between trees are either artifactual or due to a process of discord (Maddison 1997). Concerning artifacts, it is often impossible to know from a single data set whether the basal position of a long branch is due to a misplacement artifact (i.e. the long branch being attracted towards the outgroup) or due to common ancestry. Measures of robustness like bootstrap support (Felsenstein 1985), branch length, or branch support (Bremer 1988; 1994) are of little help in this context. It is well known that bootstrap support is sensitive to disparities of unequal rates of evolution among taxa: two long branches will be grouped with a high bootstrap support (for instance see Huelsenbeck 1997: 70), and this artifact will become more likely as the number of taxa decreases (Lecointre *et al.* 1993; Philippe & Douzery 1994). It is so powerful that some authors consider basal paraphylies to be unreliable, even with high bootstrap support (Philippe & Adoutte 1998; Philippe & Laurent 1998; Philippe *et al.* 2000).

This is the reason why there is no means of assessing the reliability of a basal clade even with high bootstrap support, when it has been obtained from a single analysis without previous exploration of its repeatability through separate analyses. Separate analyses are a suitable means of detecting artifactual basal positions or long-branch groupings (Philippe & Laurent 1998; Philippe *et al.* 2000) as they check for the same clades obtained separately from a number of independent genes. The probability is that there is little chance that a taxon which has an accelerated rate of change in a given gene has a similarly accelerated rate of change in all the others, especially if the latter have different functions. With increasing sequencing facilities, there is little doubt that taxonomic congruence will soon be the most powerful tool to evaluate properties of each data set, not as the final step of phylogenetic analysis, but as a tool to detect artifacts and assess reliability of clades.

Concerning process of discord, Wiens (1998) used separate analyses 'to maximize detection of different histories'. They can be used as the first step of detection of these processes through simple examination of taxonomic incongruence (Fig. 1) and then by close reexamination of the data matrix in the light of taxonomic incongruence. This exercise is common in the molecular phylogeny of bacteria where horizontal transfers of stretches of DNA are common (Dykhuizen & Green 1991; Nelson *et al.* 1997; Guttman & Dykhuizen 1994a,b; Boyd *et al.* 1996; Feil *et al.* 1996; Guttman 1997; Boyd & Hartl 1998; Lecointre *et al.* 1998; Denamur *et al.* 2000; Matic *et al.* 2002; Escobar-Paramo *et al.* 2003) and is appealed to be extended to investigations of interrelationships of close potentially hybridizing species like cyprinids (Briolay *et al.* 1998; Zardoya & Doadrio 1999; Gilles *et al.* 2001).

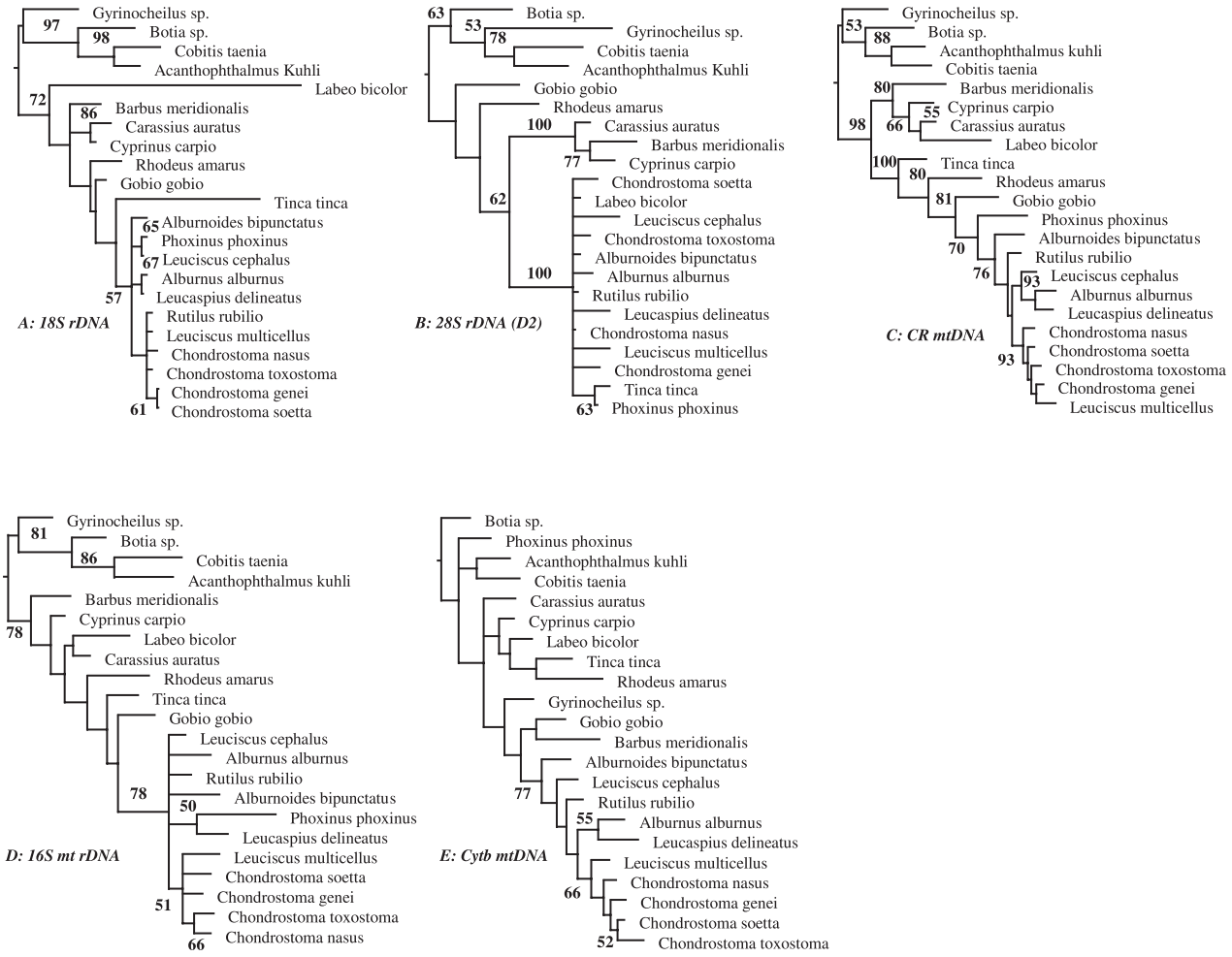


Fig. 3 A—C. MP or strict consensus trees separately obtained from each sequence data set of Table 2, through an unweighted heuristic search of PAUP 4 (100 random stepwise-addition sequences). Branch lengths have been reported under ACCTRAN optimization. Numbers above branches are bootstrap proportions above 50% obtained from 1000 replications. For tree length, CIs and RIs see Table 2. —A. 18S nuclear ribosomal DNA. —B. 28S nuclear ribosomal DNA. —C. Mitochondrial control region. —D. Mitochondrial ribosomal DNA 16S. —E. Mitochondrial cytochrome *b* DNA.

Separate analyses are an efficient exploratory step if process partitions result from homoplasy being heterogeneous among genes while being homogeneous within a single gene. If homoplasy were stored within molecular partitions (i.e. with selective forces affecting the structure of genes) in exactly the same way, the advantages of preliminary separate analyses over simultaneous analysis would decrease. Nevertheless, separate analyses are always recommended in order to check for possible heterogeneity of homoplasy distribution among partitions.

Problems with separate analyses

The choice of characters and data sets (arbitrary or justified) is a sort of weighting. The weightings involved in separate analyses are of three kinds: (1) for the delineation of the global data set with the unused data weighted zero; (2) for the definition of each character and each character state, because differing options could give different phylogenetic signals; (3) for the delineation of the data subsets.

It has often been pointed out that separate analyses involve arbitrariness in the delineation of data subsets, because delineation is often the product of tradition, and not always supported by biologically relevant assumptions. Each possible partitioning of data sets carries some arbitrariness, but even simultaneous analysis has its own limits: the amount of available data at a certain time. The particular weakness of separate analyses is that they involve more data subset limits (delineation of the overall data set and of each subset). These limits have been perceived as arbitrary (Barrett *et al.* 1991;

Siddall 1997), but they are not necessarily so. Separate analyses require more external biological knowledge (in terms of independence of partitions) than simultaneous analysis when delineation of data subsets is justified, and leads to arbitrariness only when it is not justified.

However, even if delineation is justified, separate analyses introduce an uncontrolled implicit weighting, because the same importance is given to each partition cladogram despite the fact that different trees or clades in trees may be supported by different numbers of characters. In fact the separate data sets are generally not identical, either in overall numbers of characters, or in numbers of congruent features (informative character state changes, unambiguous synapomorphies) supporting the optimal cladogram. This uncontrolled differential weighting has a real impact, since separate analyses (or a consensus tree) can provide a result different from a tree based on simultaneous analysis of all available characters, as demonstrated for instance in Barrett *et al.* (1991). Since the data and the tree reconstruction methods are the same for every cladogram, from the point of view of character congruence the sole explanation for these differences is the impact of the uncontrolled differential weighting introduced by data partitions.

The best way to avoid the bias of uncontrolled implicit weighting in separate analyses is either not to perform them at all, using simultaneous analysis instead, or to perform them as a means of data exploration, ending with simultaneous analysis as argued below.

Advantages of simultaneous analysis

The only 'weightings' in simultaneous analysis can be found: (1) in the delineation of global data subsets and (2) in the definition and delineation of characters and character states. This weighting is totally explicit, controlled and can be justified by character description and coding.

While we see separate analyses as a tool for data exploration and sorting, we consider that the cladogram obtained from simultaneous analysis should be the ultimate result. The main reason is that simultaneous analysis allows for maximization of the overall character congruence. It is maximizing explanatory power in terms of homology by descent (character state contiguity on the cladogram) given the *a priori* weights of the characters (Farris 1983) without biasing the analysis by the introduction of uncontrolled implicit differential weighting of characters due to partitioning of the data set. Needless to say, in this context maximization of overall character congruence means maximization of relevant characters, following the removal of data which are likely to be misleading.

Problems with simultaneous analysis

One of the 'arbitrary weighting' problems found in simultaneous analysis lies in the definition of characters and character states (the *bête noire* of systematics, according to Pogue & Mickevich 1990). However, this problem is universal: it is also found in separate analyses. Another problem involves controlling for independence of characters, the impact of which is the same irrespective of the method of analysis.

The problem in performing a simultaneous analysis from scratch is the possibility of a process of discord (Doyle 1992; 1997; Maddison 1997) that would not permit recovery of the history of either the first data set or the second (Bull *et al.* 1993). Obviously, this problem is the same when performing separate analyses of such data sets followed by the use of a consensus tree. These processes of discord are not problematic when separate analyses are not followed by the application of consensus procedures.

An advantageous combination requires a sequential use of separate analyses (to detect taxonomic incongruency as well as recurrent clades) followed by ILD tests (to detect processes of discord), followed by simultaneous analysis, in which misleading data have been replaced by question marks. Ending with a simultaneous analysis of a 'pruned' data set avoids the implicit uncontrolled weighting issuing from the delineation of data sets in separate analyses followed by consensus. However, by performing simultaneous analysis of all characters surviving the process of exclusion, an uncontrolled weighting (with a weaker impact) remains through the choice of the precise boundaries and length of the stretch of misleading DNA that is removed in a particular taxon. In any event, the competing strategy of separate analyses and consensus trees is of little interest because it carries none of the advantages of the sequential performing of separate and simultaneous analyses.

Optimizing the combination of analyses

Combined advantages

By proposing the sequential use of separate analyses (as a means of data exploration) followed by simultaneous analysis (as a means of phylogenetic inference), we intend to combine the advantages of both approaches in the protocol summarized in Fig. 5: (1) detection of artifacts, and (2) optimization of character congruence. As described below, the protocol is a little bit more complex, comprising (1) separate analyses, (2) tests for character incongruence, (3) removal of misleading evidence, (4) simultaneous analysis of relevant evidence, (5) interpreting the resulting clades in the light of separate analyses. As an example, five sequence data sets of 22 cyprinids published elsewhere are used (see Gilles *et al.* 1998: table 2; Gilles *et al.* 2001). Another example is published in Lecointre *et al.* (1998) and Escobar-Paramo *et al.* (2003).

Step 1: perform separate analyses

Do not use consensus trees (for problems with consensus trees, see Barrett *et al.* 1991). Simply compare trees. If a

majority of data sets support the same interrelationships with regard to a particular taxon, and a single gene renders this taxon too basal compared to all other trees, there must be a problem in the rate of changes in this taxon. The gene showing acceleration will be replaced by question marks in the corresponding taxon in the future combined matrix. This can be achieved only in a limited number of cases because attributing a particular basal position to the long-branch attraction artifact is not always obvious. In case of doubt, no data should be removed, for fear of discarding relevant phylogenetic similarities. In our example, comparing trees in Fig. 3, it is clear to any molecular systematist that the taxonomic incongruence and the branch length observed with the 18S sequence of *Labeo bicolor* when compared to the two other trees indicates a possible artifact affecting the taxon *Labeo* in the 18S tree.

Step 2: perform the ILD tests

A long-branch misplacement artifact in one of the two data sets will not necessarily provoke the rejection of the null hypothesis of congruence with another data set free of such artifacts. The reason is that it is most likely due to noise (nucleotides shared by convergence, mostly by superimposed substitutions), rather than to structured signal. Farris *et al.* (1995) mentioned that significant incongruence comes from conflicting structured signals (even punctual ones) between the two data sets (Lecointre *et al.* 1998). As is clear from our example, this is the reason why taxonomic congruence/incongruence and character congruence/incongruence are listed separately in Table 2 and why it is necessary to perform both the separate analysis and the ILD test. In Table 3, the *P*-value of the ILD test performed on the coupled 18S/control region data set (Fig. 3A,C) is 0.16, and 0.64 for 18S/16S. It is obvious that the position of *Labeo* in the 18S tree is unreliable, possibly due to its aberrant branch length. Comparing it with the position of *Labeo* in the 16S tree (Fig. 3D) and in the CR tree (Fig. 3C), as a member of the cyprinine clade close to *Carassius*, suggests an attraction of the branch of *Labeo* toward outgroups in the 18S tree. However, this artifactual taxonomic incongruence is not correlated to significant character incongruence.

If the null hypothesis of congruence is rejected, iterative removals of taxa followed by new ILD tests will permit identification of the taxa responsible for incongruence (as in Lecointre *et al.* 1998: fig. 2). For instance, all the *P*-values in Table 3 involving the 28S sequence data set are below 0.05, leading to the rejection of the null hypothesis of congruence between this and any of the other four sets. There must therefore be something wrong in the 28S sequence data. The same can be said for the cytochrome *b* data set (two pairwise tests with a *P*-value of 0.01 and two with a *P*-value of 0.06 which is very close the the 5% threshold). Let us focus first on the cytochrome *b* data set. Iterative removals of each single taxon followed by new ILD tests for each of the four comparisons identify *L. bicolor* as problematic: removal of *Labeo* leads to the highest increase of the *P*-value above 0.05 in each of the four tests (Cytb/16S: 0.53 instead of 0.06; Cytb/18S: 0.24 instead of 0.06; Cytb/Control region: 1 instead of 0.01; Cytb/28S: 0.22 instead of 0.01). The cytochrome *b* sequence should therefore trace the history of an introgressive hybridization rather than the phylogeny of the cyprinines, or should reflect an artifact from the lab and thus removed from the future combined matrix.

Let us focus now on a slightly more complex situation. Considering the 28S tree, it is remarkable to see *Labeo* and *Tinca* as members of the leuciscine clade with high bootstrap support, which is never the case in the other trees. When all possible iterative removals of single taxa followed by new ILD tests were performed for each pairwise data set, in three out of the four pairwise tests involving 28S sequences, the removal of *Labeo* produced the highest increase of the *P*-value above 0.05 (although this was less clear for the 28S/18S test because of the poor informative content of the 18S sequences). For the 28S/16S comparison, removal of *Tinca* produced the greatest increase of the *P*-value, *ex aequo* with *Phoxinus*. When iterative removals of all possible *couples* were performed,

Table 3 Properties of each of the five cyprinid sequence data used here as an example (22 taxa). SL: sequence length; IS: number of informative positions; N: number of MP trees obtained through an unweighted heuristic search of PAUP 4 (Swofford 1999, 100 random stepwise-addition sequences); TL: tree length; CI: consistency index; RI: retention index, *P*: *P*-values obtained from pairwise ILD tests; T: taxa to be excluded after all possible removals of a single taxon for each pairwise ILD test, completed by removals of pairs of taxa. A: aberrant rate. Taxa mentioned are those which removal increases the *P*-value above 5%, therefore taxa responsible for significant character incongruence.

Gene	SL	IS	N	TL	CI	RI	*P*					T	A
							16S	CYTB	CR	18S	28S		
16S	434	99	25	395	0.53	0.56	1	0.06	0.11	0.64	0.01	none	
CYTB	330	137	1	658	0.35	0.43		1	0.01	0.06	0.01	*Labeo*	
CR	614	267	1	1212	0.56	0.56			1	0.16	0.01	none	
18S	502	44	42	194	0.82	0.73				1	0.02	none	*Labeo*
28S	399	121	150	425	0.68	0.70					1	*Labeo* and *Tinca*	

Labeo + *Tinca* was clearly the one that provoked the highest increase of the *P*-value (to 0.55 for 28S/cytb, to 0.49 for 28S/16S, to 1 for 28S/CR; less clear for 28S/18S: 0.3, while the removal of *Labeo* and *Barbus* yielded 1). This suggests that *Labeo* and *Tinca* 28S sequences could not reflect the phylogeny of the species for some biological or artifactual reasons.

It is easy to understand why it is better to perform pairwise ILD tests rather than a single test performed simultaneously on the five data sets. Pairwise tests detect which sequence is responsible for incongruence, i.e. which portion of a gene for a particular taxon, just by comparing *P*-values of several pairwise tests with exclusion of a particular taxon. Results can be unambiguous (Lecointre *et al.* 1998) but are not always straightforward when homoplasy is high. For example, for the test 28S/18S, while the removal of *Labeo* increases the *P*-value from 0.02 to 0.46, the removal of *Barbus* increases it to 1. When couples are removed, each time the couple contains *Barbus* the *P*-value goes to 1 (and no more than 0.45 for other couples). There must be something problematic about the characters of *Barbus*, but in which sequence: 28S or 18S? If the 28S sequence has experienced a process of discord, the *P*-values of 28S/cytb, 28S/CR and 28S/16S should significantly increase above 0.05 when *Barbus* is removed. But once it is removed, these *P*-values are, respectively, 0.03, 0.01 and 0.01. There are two possible explanations: (1) another taxon must be responsible for the incongruence, obscuring the response when *Barbus* is removed, or (2) it is the 18S sequence of *Barbus* that is problematic. The first explanation is favoured by the fact that *Labeo* was identified as the problematic 28S sequence. The second solution is not obvious, as the removal of *Barbus* in other pairwise tests involving 18S data do not yield clear results (no spectacular increase of the *P*-value). In case of doubt, one should not remove data. Thus, 28S and 18S sequences of *Barbus* will be kept because we cannot clearly demonstrate which one is 'misleading'.

Step 3: remove misleading data

We should discard the data that (1) obscure taxon interrelationships, and (2) tell us a story that has nothing to do with taxon history (Fig. 5). The procedure consists in replacing character states of particular partitions in particular taxa with question marks where it can be shown that they are misleading, and combining all the relevant remaining data into a single matrix. We should replace the following with question marks: (1) the gene(s) in a particular taxon where the rate of change is so extreme that it obviously obscures the relationships of the taxon (e.g. the 18S sequence of *Labeo*, Fig. 3A); (2) the gene(s) responsible for significant incongruence, possibly reflecting a process of discord that leads to a history that is not the phylogeny of the taxa (e.g. the cytochrome *b* and 28S sequence of *Labeo* in the combined matrix, the 28S sequence of *Tinca*, Fig. 5).

Step 4: simultaneous analysis of the 'pruned data'

The cladogram resulting from this 'partial' (Gilles *et al.* 2001) or 'careful' simultaneous analysis (Fig. 6) of the 'pruned data' (i.e. with some sequences removed) can be compared with the cladogram from the simultaneous analysis of the complete data set (Fig. 4). In Fig. 4, *Labeo* is excluded from the cyprinines while it is a member in Fig. 6 (Howes 1991); *Gobio* is the sister-group of *Rhodeus* in Fig. 4 while it is not in Fig. 6. The tree within the leuscicines remains unchanged. When different kinds of consensus trees are calculated from the five cladograms of separate analyses, almost all the nodes collapse: the strict consensus retains one node, the combinable component consensus retains two, the Adams consensus four and the 50%-majority rule consensus six. Comparing Fig. 6 (19 nodes) to these consensus trees (data not shown), it is clear that our approach preserves relevant phylogenetic similarity while a consensus technique loses almost everything. Consensus techniques are far from being able to manage the impact of misleading data.

Step 5: interpretation of the tree

Hillis (1995) recognized the benefit of comparing the tree resulting from the combined data sets with individual trees obtained from separate analyses: 'Although a combined analysis of several data sets (assuming that they are appropriate for combining) may give the single best estimate of phylogeny (…), the conclusion would be greatly strengthened if it were compatible with that of each of the individual data sets as well (…)'. If reliability of a clade is thought in terms of repeatability through separate analyses, this comparison makes sense. Comparing every tree from separate analyses and the tree resulting from data combination, there are theoretically three categories of clades, which are: (1) found in all trees; (2) repeated in several trees but not all of them; (3) never repeated. The first category apparently raises no problem, these clades being in principle automatically found in the tree from the simultaneous analysis. Their strong reliability does not depend on the robustness indicator they have in that tree but on their repeatability across separate analyses. Nevertheless, one should be careful when expecting these clades from the simultaneous analysis. Some clades repeated in separate analyses can be absent from the tree based on the complete data set. This has been shown theoretically (see the clade BCD in Barrett *et al.* 1991: fig. 1) as well as empirically (Dettaï & Lecointre 2004: figs 4 and 5). In such cases the tree from the simultaneous analysis is insufficient to summarize the results: a tree summarizing those clades considered as reliable must be constructed (as in Dettaï & Lecointre 2004). Clades of the third category should have disappeared from the tree based on the 'careful' combination, if misleading data have been correctly identified and removed; for example, the position of *Labeo* in the 18S tree (Fig. 3A) is not found in Fig. 6. How

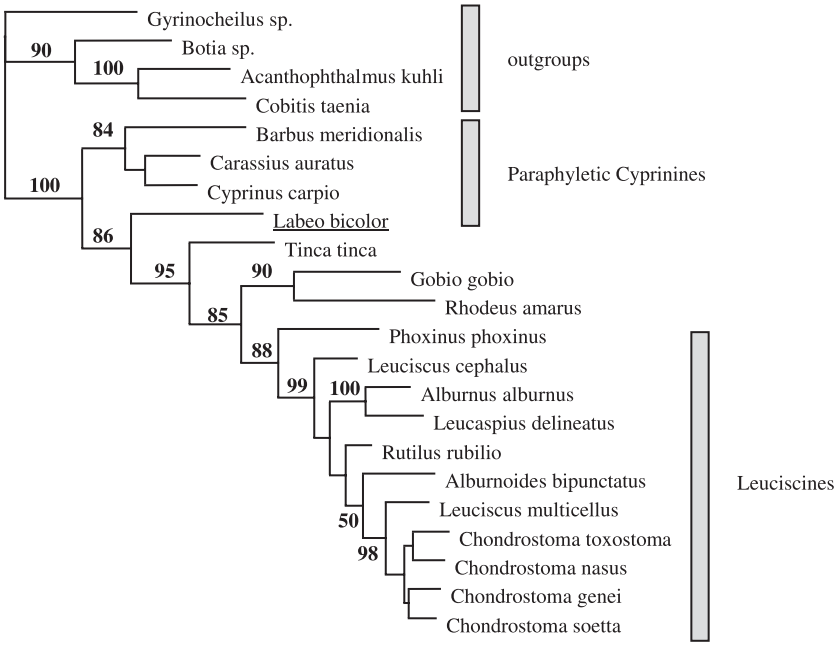


Fig. 4 MP tree obtained from the full combination of the five data sets, using an unweighted heuristic search of PAUP 4 (100 random stepwise-addition sequences). Branch lengths have been reported under ACCTRAN optimization. Numbers above branches are bootstrap proportions above 50% obtained from 1000 replications. Number of characters = 2279, with 668 informative sites. Tree length = 2997 steps, CI = 0.53, RI = 0.52. Here *Labeo* is not a member of the cyprinines.

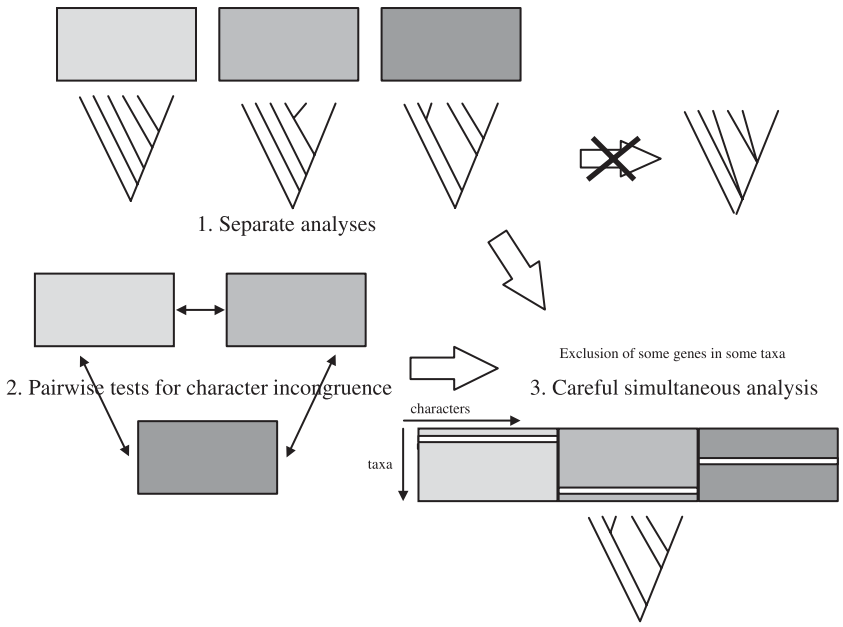


Fig. 5 General protocol proposed to infer phylogenies based on all the available relevant evidence, compatible with Carnap's requirement for TE. This protocol is based on the idea that the requirement for TE (total knowledge) requires the removal of data if it has been reliably established that they are misleading. *Step 1*: separate analyses of biologically justified partitions (without consensus) to explore properties of homoplasy and detect possible reconstruction artifacts. *Step 2*: pairwise ILD tests; in case of significant character incongruence, perform iterative removals of individual taxa to detect which partition in which taxon is responsible (possible removal of all combinations of couples). *Step 3*: replace character partitions for certain taxa responsible for incongruence with question marks in the combined matrix; if necessary, replace stretches of sequences which have an obviously aberrant rate of change, as detected through the previous separate analyses. *Step 4*: simultaneous analysis of this 'pruned' matrix. *Step 5*: interpretation of the reliability of the resulting clades in the light of their repeatability in separate analyses.

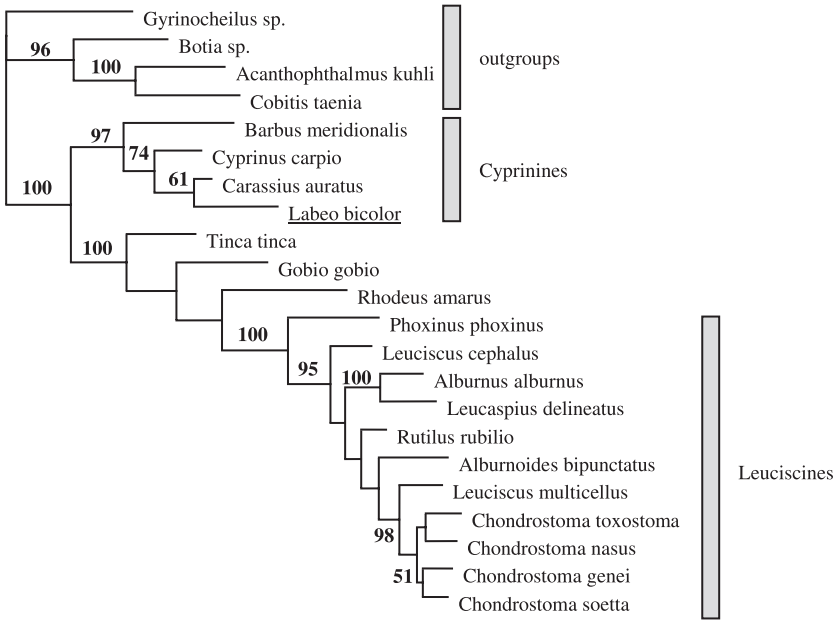


Fig. 6 MP tree obtained from the 'careful' combination of the five data sets. Following the protocol in Fig. 5. The cytochrome *b* and 28S sequences of *Labeo* and the 28S sequence of *Tinca* have been replaced in the full matrix with question marks as they create significant incongruence. The 18S sequence of *Labeo* has also been removed because of its obviously aberrant rate of change (Fig. 3A). This tree was obtained using an unweighted heuristic search of PAUP 4 (100 random stepwise-addition sequences). Branch lengths have been reported under ACCTRAN optimization. Numbers above branches are bootstrap proportions above 50% obtained from 1000 replications. Number of characters = 2279, with 668 informative sites. Tree length = 2864 steps, CI = 0.54, RI = 0.54. *Labeo* is now a member of the cyprinines.

should we manage clades of the second category? In the absence of evidence that data on which these clades rely are misleading, the tree based on the 'careful' combination should provide the clades from which biological conclusions should be drawn and character evolution studied. Obviously, the reliability of these clades should be weaker than that of those of the first category. Another solution is given by Chen *et al*. (2003) and Dettaï & Lecointre (2004); as long as no other alternative clade is itself repeated, the repeated clade is considered as reliable and reasons for non-recovery in some data sets are explored using specific protocols (including, for instance, incomplete combinations in Dettaï & Lecointre, 2004).

To assess reliability, should repeatability be preferred over indicators of robustness? As these indicators do not escape the potential pitfalls of positively misleading signals, they can offer no reliability. Therefore, when the clades found in the cladogram based on the simultaneous analysis are considered, priority is given to their repeatability across previous separated analyses over their statistical robustness (Table 4); repeated clades will be reliable whatever the associated bootstrap support calculated from the simultaneous analysis.

In summary, we use separate analyses to assess the reliability of clades found in the tree from the 'careful' simultaneous analysis. The tree is useful to assess robustness and to obtain the final tree from which character evolution can be studied. Separate analyses are an essential exploratory step to detect artifacts and repeated clades; however, they cannot indicate anything on the clades of the second category and are not suitable for the study of character evolution. Simultaneous analysis of the 'pruned' data is the final step essential for obtaining conclusions with regard to clades of the second category described above, for establishing character evolution and for ensuing biological interpretations. Character evolution must be studied on the tree based on the maximization of congruence of *relevant characters*.

Alternatives and objections

Wiens (1998) suggested identifying processes of discord on the basis of separate analyses, but proposed combining the discordant data into a simultaneous analysis, the interpretation of which was modulated by information from the previous separate analyses. We find it difficult to understand why misleading data should be retained in the analysis, as it leads to the difficult exercise of commenting on a tree likely to be biased. As already argued above, the principle of TE *sensu* Carnap (1950) should not be equated with using all data at hand without discrimination. On the contrary, it consists in taking into account all relevant knowledge, including knowledge based on misleading data. This does not preclude the combination of all remaining putatively informative data, i.e. the careful application of TE without bias.

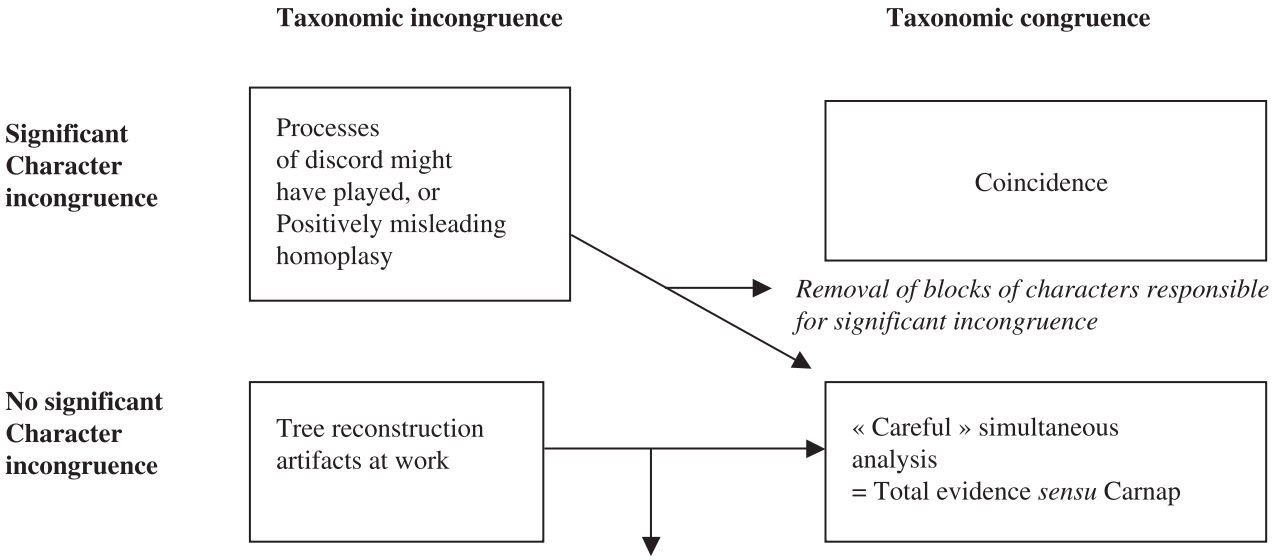
Slowinski & Page (1999) recognized the existence of 'linkage partitions' and kept data sets separated, but proposed the simultaneous analysis of all gene trees based on different linkage partitions using GeneTree (Page 1998). The combination is not produced at the level of data themselves, but is performed at the level of tree searches. This approach can be seen as closer to Carnap in the sense that their kind of simultaneous analysis detects data that are putatively misleading and infers misleading events to explain them.

Siddall (1997) provided arguments against conditional combination, evoking arbitrariness in the choice of partitions,

Table 4 Reliability of a clade according to separate and simultaneous analyses. A clade is considered reliable when it is recurrently found in separate analyses of different data sets.

		SIMULTANEOUS ANALYSIS	
		HIGHLY SUPPORTED:	POORLY SUPPORTED:
SEPARATE ANALYSES	REPEATED CLADE:	Keep as reliable	Keep as reliable
	NON REPEATED CLADE:	Doubtful	Reject

Table 5 Table 2 revisited according to Figure 5. If there are processes of discord and/or tree reconstruction artifacts, phylogeny based on the maximization of congruence of all of the available *relevant* characters is obtained by replacing with question marks the character states obtained from processes of discord or non-random homoplasy.



especially when there is more than one way to identify character incongruence (arbitrariness being understood as 'selected at random or without reason', Siddall 1997: 766). Another criticism is that 'when data are partitioned, the investigator runs the risk of discarding otherwise corroborating information'. Our approach offers solutions to these difficulties.

First, it requires a biological justification for each partition (process partitions of Bull *et al.* 1993; linkage partitions of Doyle 1992; 1997; Slowinski & Page 1999), coupled with the observation that homoplasy accumulates in different manners from one partition (gene) to another.

Second, it considers separate analyses as an exploratory step, employing the final 'careful' combination as the data set on which phylogenetic interpretations should be based. In this approach (summarized in Table 5) corroborating information is not discarded, while only information likely to be misleading is neutralized. This procedure appears to follow the requirement of total *relevant* evidence; it avoids throwing away the baby (putatively informative data) with the bath water (likely misleading data).

Our approach might be judged as circular, or at least as lacking independence from process theories by those who advocate a hypothetico-deductive and refutationist view of phylogenetic systematics. However, if we ignore the mirages of hypothetico-deduction (Rieppel, 2003b) and refutation (Kitts 1977; Ruse 1979; Rieppel 2003a) at the level of phylogeny inference (Sober 1988; Rieppel 2003a,b), and consider it as plainly explanatory, then we are maximizing the internal consistency of our phylogenetic explanation of the data (Fitzhugh, 1997; 1998), taking all relevant background knowledge into account. According to this perspective, knowledge about evolutionary processes must be obtained using all logical means, including separate analyses.

Acknowledgements

We thank André Gilles for allowing us to use his nuclear ribosomal sequence data, and Jean-Pierre Coutanceau for the

help provided. This paper was presented at the 18th meeting of the Willi Hennig Society held at Goettingen, September 1999.

References

Allard, M. W. & Carpenter, J. M. (1996). On weighting and congruence. *Cladistics*, *12*, 183–198.

Barrett, M., Donoghue, M. J. & Sober, E. (1991). Against consensus. *Systematic Zoology*, *40*, 486–493.

Bisercic, M., Feutrier, J. Y. & Reeves, P. R. (1991). Nucleotide sequence of the *gnd* genes from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. *Journal of Bacteriology*, *173*, 3894–3900.

Boyd, E. F. & Hartl, D. L. (1998). Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *Journal of Bacteriology*, *180*, 1159–1165.

Boyd, E. F., Hill, C. W., Rich, S. M. & Hartl, D. L. (1996). Mosaic structure of plasmids from natural populations of *Escherichia coli*. *Genetics*, *143*, 1091–1100.

Bremer, K. (1988). The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution*, *42*, 795–803.

Bremer, K. (1994). Branch support and tree stability. *Cladistics*, *10*, 295–304.

Briolay, J., Galtier, N., Brito, R. M. & Bouvet, Y. (1998). Molecular phylogeny of cyprinidae inferred from cytochrome *b* DNA sequences. *Molecular Phylogenetics and Evolution*, *9*, 100–108.

Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L. & Wadell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Systematic Biology*, *42*, 384–397.

Carnap, R. (1950). *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Chen, W. J. (2001). *La répétitivité des clades comme critère de fiabilité: application à la phylogénie des Acanthomorpha (Teleostei) et des Notothenioidei (acanthomorphes antarctiques)*. Paris: Thèse de l'université Pierre et Marie Curie (Paris VI).

Chen, W. J., Bonillo, C. & Lecointre, G. (2000). Taxonomic congruence as a tool to discover new clades in the acanthomorph (Teleostei) radiation. 80th ASIH meeting, La Paz, Mexico.

Chen, W. J., Bonillo, C. & Lecointre, G. (2003). Repeatability of clades as criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular Phylogenetics and Evolution*, *26*, 262–288.

Chippindale, P. T. & Wiens, J. J. (1994). Weighting, partitioning, and combining characters in phylogenetic analysis. *Systematic Biology*, *43*, 278–287.

Cunningham, C. W. (1997a). Can three incongruence tests predict when data should be combined?. *Molecular Biology and Evolution*, *14*, 733–740.

Cunningham, C. W. (1997b). Is incongruence between data partitions a reliable prediction of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Systematic Biology*, *46*, 464–478.

Darlu, P. & Lecointre, G. (2002). When does the Incongruence Length Difference test fail? *Molecular Biology and Evolution*, *19*, 432–437.

Denamur, E., Lecointre, G., Darlu, P., Tenaillon, O., Acquaviva, C., Sadaya, C., Sunjevaric, I., Rothstein, R., Elion, J., Taddei, F., Radman, M. & Matic, I. (2000). Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. *Cell*, *103*, 711–721.

Desjardins, P., Picard, B., Kaltenböck, B., Elion, J. & Denamur, E. (1995). Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction fragment length polymorphism. *Journal of Molecular Evolution*, *41*, 440–448.

Dettaï, A. & Lecointre, G. (2004). In search of notothenioid relatives. *Antarctic Science*, in press.

Dolphin, K., Belshaw, R., Orme, C. D. L. & Quicke, D. L. J. (2000). Noise and incongruence: interpreting results of the incongruence length difference test. *Molecular Phylogenetics and Evolution*, *17*, 401–406.

Doyle, J. J. (1992). Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany*, *17*, 144–163.

Doyle, J. J. (1997). Trees within trees: genes and species, molecules and morphology. *Systematic Biology*, *46*, 537–553.

Dykhuizen, D. E. & Green, L. (1991). Recombination in *Escherichia coli* and the definition of biological species. *Journal of Bacteriology*, *173*, 7257–7268.

de Queiroz, A., Donoghue, M. J. & Kim, J. (1995). Separate versus combined analysis of phylogenetic evidence. *Annual Review of Ecology and Systematics*, *26*, 657–681.

Eernisse, D. J. & Kluge, A. G. (1993). Taxonomic congruence versus total evidence, and amnioce phylogeny inferred from fossils, molecules, and morphology. *Molecular Biology and Evolution*, *10*, 1070–1195.

Escobar-Paramo, P., Sabbagh, A., Darlu, P., Pradillon, O., Vaury, C., Denamur, E. & Lecointre, G. (2003). Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study. *Molecular Phylogenetics and Evolution*, in press.

Farris, J. S. (1983). The logical basis of phylogenetic analysis. In N. Platnick & V. A. Funk (Eds) *Advances in Cladistics*, Vol. 2. (pp. 1–36). New York: Columbia University Press.

Farris, J. S., Källersjö, M., Kluge, A. & Bult, C. (1995). Testing significance of incongruence. *Cladistics*, *10*, 315–319.

Feil, E., Zhou, J., Maynard-Smith, J. & Spratt, B. G. (1996). A comparison of the nucleotide sequences of the *adk* and *recA* genes of pathogenic and commensal *Neisseria* species: evidence for extensive interspecies recombination within *Adk*. *Journal of Molecular Evolution*, *43*, 631–640.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, *27*, 401–410.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*, 783–791.

Fitzhugh, K. (1997). The abduction of cladistics. *Cladistics*, *13*, 170–171.

Fitzhugh, K. (1998). C (h,be): e =/= synapomorphy. Hennig XVII: 17th meeting of the Willi Hennig Society Abstracts (pp. 34–35).

Geiger, D. L., Fitzhugh, K. & Thacker, C. E. (2001). Matters of the record. Timeless characters: a response to Vermeij (1999). *Paleobiology*, *27*, 179–180.

Germot, A. & Philippe, H. (1999). Critical analysis of eukaryotic phylogeny: a case study based on the HSP70 family. *Eukaryote Microbiology*, *46*, 116–124.

Gilles, A., Lecointre, G., Faure, E., Chappaz, R. & Brun, G. (1998). Mitochondrial phylogeny of the European cyprinids: implications for their systematics, reticulate evolution and colonisation time. *Molecular Phylogenetics and Evolution*, *10*, 132–143.

Gilles, A., Lecointre, G., Miquelis, A., Loerstcher, M., Chappaz, R. & Brun, G. (2001). Partial combination applied to phylogeny of European cyprinids using the mitochondrial control region. *Molecular Phylogenetics and Evolution*, *19*, 22–33.

Grande, L. (1994). Repeating patterns in nature, predictability, and 'impact' in science. In L. Grande & O. Rieppel (Eds) *Interpreting the Hierarchy of Nature*, (pp. 61–84). New York: Academic Press.

Guttman, D. S. (1997). Recombination and clonality in natural populations of *Escherichia coli*. *Trends in Ecology and Evolution*, *12*, 16–22.

Guttman, D. S. & Dykhuizen, D. E. (1994a). Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science*, *226*, 1380–1383.

Guttman, D. S. & Dykhuizen, D. E. (1994b). Detecting selective sweeps in naturally occuring *Escherichia coli*. *Genetics*, *138*, 993–1003.

Hassanin, A., Lecointre, G. & Tillier, S. (1998). The 'Evolutionary Signal' of homoplasy in Protein coding gene sequences and its consequences for a priori weighting in phylogeny. *Comptes Rendus de l'Académie des Sciences*, *321*, 611–620.

Herzer, P. J., Inouye, S., & Whittman, T. S. (1990). Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *Journal of Bacteriology*, *172*, 6175–6181.

Hillis, D. M. (1995). Approaches for assessing phylogenetic accuracy. *Systematic Biology*, *44*, 3–16.

Howes, G. (1991). Systematics and biogeography: an overview. In I. Winfield & J. Nelson (Eds) *Cyprinid fishes: systematic biology and exploitation* (pp. 1–54). New York: Chapman and Hall.

Huelsenbeck, J. P. (1997). Is the Felsenstein zone a fly trap? *Systematic Biology*, *46*, 69–74.

Huelsenbeck, J. P. & Bull, J. J. (1996). A likelihood ratio test to detect conflicting phylogenetic signal. *Systematic Biology*, *45*, 92–98.

Huelsenbeck, J. P., Bull, J. J., Cunningham W. C. (1996). Combining data in phylogenetic analysis. *Trends in Ecology and Evolution*, *11*, 152–158.

Johnson, K. P., Drown, D.M. & Clayton, D. H. (2001). A data based parsimony method of cophylogenetic analysis. *Zoologica Scripta*, *30*, 79–87.

Kitts, R. D. (1977). Karl Popper, verifiability and systematic zoology. *Systematic Zoology*, *26*, 185–194.

Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology*, *38*, 7–25.

Kluge, A. G. (1997). Sophisticated falsification and research circles: Consequences for differential character weighting in phylogenetic systematics. *Zoologica Scripta*., *26*, 349–360.

Kluge, A. G. (2003). On the deduction of species relationships: a précis. *Cladistics*, *19*, 233–239.

Kluge, A. G. & Wolf, A. J. (1993). Cladistics: what's in a word? *Cladistics*, *9*, 183–199.

Larson, A. (1994). The comparison of morphological and molecular data in phylogenetic systematics. In B. Schierwater, B. Streit, G. P. Wagner & R. Desalle (Eds) *Molecular Ecology and Evolution. Approaches and Applications*, (pp. 371–390). Basel, Switzerland: Birkhauser-Verlag.

Lecointre, G., Philippe, H., Lê, H. L. V. & Le Guyader, H. (1993). Species sampling has a major impact on phylogenetic inference. *Molecular Phylogenetics and Evolution*, *2*, 205–224.

Lecointre, G., Rachdi, L., Darlu, P. & Denamur, E. (1998). *Escherichia coli* molecular phylogeny using the Incongruence Length Difference test. *Molecular Biology and Evolution*, *15*, 1685–1695.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*, 523–536.

Matic, I., Tenaillon, O., Lecointre, G., Darlu, P., Radman, M., Taddei, F. & Denamur, E. (2002). Horizontal transfer of mismatch repair genes and the variable speed of bacterial evolution. In M. Syvanen, & C. I. Kado (Ed.) *Horizontal Gene Transfer*, (pp. 147–155). New York: Academic Press.

Mickevich, M. F. (1978). Taxonomic congruence. *Systematic Zoology*, *27*, 143–158.

Mickevich, M. F. & Farris, J. S. (1981). The implications of congruence in *Menidia*. *Systematic Zoology*, *30*, 351–370.

Milkman, R. (1997). Recombination and population structure in *Escherichia coli*. *Genetics*, *146*, 745–750.

Miyamoto, M. M. & Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology*, *44*, 64–76.

Moreira, D., Le Guyader, H. & Philippe, H. (1998). Unusually high evolutionary rate of the elongation factor 1 alpha genes from the Ciliophora and its impact on the phylogeny of Eukaryotes. *Molecular Biology and Evolution*, *16*, 234–245.

Nelson, G. J. (1979). Cladistic analysis and synthesis: principles and definitions, with a historical note on Adanson's *Familles des Plantes* (1763–1764). *Systematic Zoology*, *28*, 1–21.

Nelson, K. & Selander, R. K. (1994). Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proceedings of the National Academy of Science of the USA*, *91*, 10227–10231.

Nelson, K., Wang, F. S., Boyd, E. F. & Selander, R. K. (1997). Size and sequence polymorphism in the isocitrate dehydrogenase kinase phosphatase gene (*aceK*) and flanking regions in *Salmonella enterica* and *Escherichia coli*. *Genetics*, *147*, 1509–1520.

Nixon, K. C. & Carpenter, J. M. (1993). On outgroups. *Cladistics*, *9*, 413–426.

Nixon, K. C. & Carpenter, J. M. (1996). On simultaneous analysis. *Cladistics*, *12*, 221–241.

Page, R. D. M. (1998). GeneTree: Comparing gene and species phylogenies using reconciled trees. *Bioinformatics*, *14*, 819–820.

Page, R. D. M. & Charleston, M. A. (1998). Trees within trees: phylogeny and historical associations. *Trends in Ecology and Evolution*, *13*, 356–359.

Page, R. D. M. & Charleston, M. A. (1999). Comments on Allard and Carpenter (1996), or the 'Aquatic Ape' hypothesis revisited. *Cladistics*, *15*, 73–74.

Page, R. D. M. & Holmes, E. C. (1998).*Molecular Evolution. A Phylogenetic Approach*. Oxford: Blackwell Science.

Philippe, H. & Adoutte, A. (1998). The molecular phylogeny of Eukaryota: solid facts and uncertainties. In G. H. Cooms, K. Vickerman, M. A. Sleigh & A. Warren (Eds) *Evolutionary Relationships among Protozoa*, (pp. 25–56). London: Chapman & Hall.

Philippe, H. & Douzery, E. (1994). The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. *Journal of Mammalian Evolution*, *2*, 133–152.

Philippe, H. & Laurent, J. (1998). How good are deep phylogenetic trees? *Current Opinion in Genetics and Development*, *8*, 616–623.

Philippe, H., Lecointre, G., Lê, H. L. V. & Le Guyader, H. (1996). A critical study of homoplasy in molecular data with the use of a morphologically based cladogram, and its consequences for character weighting. *Molecular Biology and Evolution*, *13*, 1174–1186.

Philippe, H. Lopez, P. Brinkmann, H. Budin, K. Germot, A. Laurent, J. Moreira, D. Müller, M. & Le Guyader, H. (2000).

Early branching or fast evolving Eukaryotes? An answer based on slowly evolving positions. *Proceedings of the Royal Society, London Series B*, *267*, 1213–1221.

Pogue, M. G. & Mickevich, M. F. (1990). Character definitions and character state delineation: the bête noire of phylogenetic inference. *Cladistics*, *6*, 319–361.

Rieppel, O. (2003a). Popper and systematics. *Systematic Biology*, *52*, 259–271.

Rieppel, O. (2003b). Semaphoronts, cladograms and the roots of total evidence. *Biological Journal of the Linnean Society*, *80*, 167–186.

Rodrigo, A. G., Kelly-Borges, M., Bergquist, P. R. & Bergquist, P. L. (1993). A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. *New Zealand Journal of Botany*, *31*, 257–268.

Ruse, M. (1979). Falsifiability, consilience and systematics. *Systematic Zoology*, *29*, 530–536.

Sennblad, B. & Bremer, B. (2000). Is there a justification for differential a priori weighting in coding sequences? A case study from rbcL and Apocynaceae s.l. *Systematic Biology*, *49*, 101–113.

Siddall, M. E. (1997). Prior agreement: arbitration or arbitrary? *Systematic Biology*, *46*, 765–769.

Siddall, M. E. (1998). Success of parsimony in the four-taxon case: long-branch repulsion by likelyhood in the Farris zone. *Cladistics*, *14*, 209–220.

Slowinski, J. B. & Page, R. D. M. (1999). How should species phylogenies be inferred from sequence data? *Systematic Biology*, *48*, 814–825.

Sober, E. (1988). *Reconstructing the Past. Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press.

Sullivan, J. (1996). Combining data with different distributions of among-site rate variation. *Systematic Biology*, *45*, 375–380.

Swofford, D. L. (1997). *PAUP* — Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Test Version 4.0.0d64. Sunderland, MA: Sinauer Associates.

Swofford, D. L. (1999). *PAUP* — Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sunderland, MA: Sinauer Associates.

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.

Vidal, N. & Lecointre, G. (1998). Weighting and congruence: a case study based on three mitochondrial genes in pitvipers. *Molecular Phylogenetics and Evolution*, *9*, 366–374.

Wiens, J. J. (1998). Combining data sets with different phylogenetic histories. *Systematic Biology*, *47*, 568–581.

Zardoya, R. & Doadrio, I. (1999). Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *Journal of Molecular Evolution*, *49*, 227–237.