

## **Quantifying the potential utility of phylogenetic characters**

**James A. Cotton\*, Mark Wilkinson**

Department of Zoology, The Natural History Museum, Cromwell Road, London,  
SW7 5BD, UK.

\*Author for Correspondence. Current Address: Bioinformatics Laboratory,  
Department of Biology, National University of Ireland Maynooth, County Kildare,  
Ireland. Tel: +353 (0)1 708 6043, Fax: +353 (0)1 708 3845, E-mail:

[james.cotton@nuim.ie](mailto:james.cotton@nuim.ie)

KEYWORDS: compatibility, cladistic information content, information theory,  
morphology, molecular data.

## **ABSTRACT**

There has been a long history of interest in measuring the information conveyed by phylogenetic data. In one application, recent studies have attempted to compare the informativeness of morphological and molecular data, and of nucleotide and amino acid sequence alignments. While a variety of measures have been proposed to quantify phylogenetic information, most measures are rather unsatisfactory, failing to capture every aspect of the informativeness of a character. One measure, cladistic information content (CIC) is a natural measure of phylogenetic information. We show why CIC is preferable to other, recently introduced, measures, and, as an example, use CIC to compare the information of recent morphological and molecular datasets. This provides new empirical data relevant to the debate about the relative utility of morphology and molecules in phylogenetic inference, a subject of significant interest.

## INTRODUCTION

Recently, Wortley & Scotland (2006) investigated several measures of the potential utility of character data for phylogenetic inference and their use in comparing molecular and morphological matrices. They suggest using a measure based on the number of character states present in a character ( $S_c$ ). One less than this number ( $S_c - 1$ ) is the minimum number of parsimony-informative character state changes that a character must display on any tree, and the sum of this value across characters is a measure of the minimum number of informative changes on an entire matrix (their  $\Delta_{\min}$ , introduced by Baker & al., 1998). Wortley & Scotland suggest that  $\Delta_{\min}$  is the most useful measure of the phylogenetic utility of a data matrix, and so should be more widely reported in phylogenetic research. We agree that  $\Delta_{\min}$  is an improvement over other measures, such as the number of characters or the number of parsimony informative characters, but we suggest that another measure, based on Thorley & al.'s (1998) cladistic information content (CIC), may be more useful still.

To demonstrate the difference between this and previous measures, we investigate 50 recent morphological and 49 recent molecular datasets used for phylogenetic reconstruction that were examined by Wortley & Scotland (2006). The relative importance of molecular and morphological data has been a subject

of controversy within the systematics community for a number of years (Gura, 2000; Hillis & Wiens, 2000). In particular, some authors have claimed that molecular data either has, or soon will, completely supersede morphological evidence for phylogenetic inference (Scotland & al, 2003), a claim wholly rejected by others (Wiens, 2004). With a few notable exceptions (Bledsoe & Raikow, 1990; Wortley & Scotland, 2006; Pisani & al., 2007) this debate has suffered from a lack of much quantitative evidence of the relative merits of different data and an over-reliance on case studies of limited generality. Even for these case-studies, the interpretation has been open to doubt (e.g. Hedges & Maxson, 1996, 1997; Lee, 1997). One general problem with this literature is a lack of agreement on the best ways to quantify the utility of phylogenetic characters. Here, we formulate this problem as one of measuring the amount of information a character or set of characters convey about the relationships of the organisms the character is sampled from. This allows us to use ideas from information theory to propose an elegant measure of phylogenetic information, and clarifies a number of difficult problems with measuring the phylogenetic utility of data matrices.

**Cladistic information of characters.** – In mathematical information theory (Hartley, 1928; Shannon, 1948; Jones & Jones, 2000), the information conveyed by a message is quantified by the reduction in uncertainty upon receiving the message. Phylogenetic characters convey information about the phylogenetic tree linking the taxa that the characters are sampled from, so that the amount of information conveyed by a phylogenetic character is the reduction in uncertainty

about the tree once we know a particular character. For simplicity, we can consider that a character only permits a tree if it fits the tree without homoplasy, and prohibits those trees that would require extra character-state changes. If we consider only tree topology (ignoring branch lengths), and treat all trees as equally likely, then we get a particularly simple measure of the information content, the cladistic information content (CIC), first introduced by Thorley & al. (1998) and Thorley (2000) to study the information conveyed by a consensus tree. This measure has been used a number of times to study character information (Semple & Steel, 2002; Dezulian & Steel, 2004; Thorley & al. 2004; Steel & Penny, 2006) reflecting the conceptual connection between characters and trees (e.g. Rodrigo, 1996).

CIC was originally introduced to quantify the amount of information conveyed by (potentially polytomous) trees on particular leaf sets and defined as the logarithm of the number of binary trees compatible with the tree divided by the total number of possible binary trees on the leaf set of the tree:

$$CIC = -\log\left(\frac{N_{permitted}}{N_{total}}\right)$$

Where the base of the logarithm is arbitrary. The CIC is a dimensionless number, but is conveniently expressed in bits (a standard unit of information) by taking logarithms to base 2.

In this setting it is clear that a polytomous tree is compatible with any binary tree

that is a resolution of that tree, and we could say that the tree only permits its resolutions, as any other tree would contradict it. We can reason similarly for a phylogenetic character: an unordered character induces a partitioning of a taxon set into non-overlapping sets each possessing the same state for the character. Carter & al. (1990) present a simple formula for the proportion of binary trees that fit a particular unordered character without cost (i.e. that the character is convex on, Estabrook, 1978; see Meacham & Estabrook, 1985 for a review of compatibility methods, Semple & Steel, 2003, pp.65-69 for a definition of convexity and some mathematical background to this work), which can easily be transformed into a formula for the CIC of any unordered character (Steel & Penny, 2006).

CIC as defined above is a measure of the informativeness of an individual character, but phylogenetic studies usually employ matrices of many characters on the same set of taxa. For a matrix of characters we employ a measure we call *total CIC*, defined as the sum of the CIC values for the individual characters. As we discuss further below, using total CIC assumes that the information conveyed by characters is independent, which guarantees that the information content of characters is additive (Shannon, 1948). This is certainly not the case in real data but is a useful first-order simplifying assumption. We note that other workers have implicitly made the same assumption, e.g. in adding numbers of character states across characters in a matrix.

## MATERIALS, METHODS AND RESULTS

**Comparing measures of character information.** – In advocating  $\Delta_{\min}$ , Wortley & Scotland (2006: 432) note that "not all parsimony informative characters contain the same information for phylogeny reconstruction" and that their preferred measure takes into account "the amount of information in terms of required branches on a cladogram". However,  $\Delta_{\min}$  does not take into account the uncertainty in the phylogeny relating the sets of taxa that share a particular character state. A character that partitions 120 taxa into 60 sets of 2 taxa each has a higher  $\Delta_{\min}$  but contains much less phylogenetic information than one splitting the 120 taxa into 20 sets of 6 taxa. This is because there are many more possible relationships between the 60 sets than the 20 sets, and this overwhelms the fact that there is uncertainty about the relationships within the sets of 6 but not within the sets of 2 taxa (Fig. 1). A smaller example (on 12 taxa) is shown in Table 1, confirming that  $\Delta_{\min}$  and CIC will rank characters differently in many circumstances, and showing that CIC also discriminates among characters with the same  $\Delta_{\min}$  when they convey different information: for example, characters with more balanced distributions of character states convey more information (Table 2) than those with unbalanced state distributions.

A more obvious limitation of  $\Delta_{\min}$  appears when characters show missing data. Character 4 of Table 2 clearly contains less information than characters 1 or 2: it

is identical except for missing data for two taxa. All three characters have the same  $\Delta_{\min}$  value, while CIC correctly captures the loss of information due to missing data. As Fig. 2 shows, CIC is also well-correlated with other natural measures of character information.

**CIC of molecular and morphological data.** – To illustrate the differences between total CIC and other measures of the character utility, we have re-analysed 99 of the 100 data matrices from TreeBase previously analysed by Wortley & Scotland (2006). Wortley & Scotland (2006) give citations for all of the studies included in this data, and present a table showing, for each study, the values of the various measures they considered. We do not reproduce that information here. The only difference between our data and that of the earlier study is that a single molecular matrix (Treebase matrix M1603, study S966) has been excluded because it is based on RFLP data rather than DNA or amino acid sequence data. While RFLPs are certainly molecular, they are not similar in either number of characters nor character state-space to the molecular sequence data far more commonly used in modern molecular systematics.

Fig. 3 shows the results of using total CIC as a measure of information content of molecular and morphological characters. Total CIC values are non-normally distributed, and a non-parametric Mann-Whitney test confirms that the molecular matrices are significantly more informative than the morphological matrices in our sample ( $p=0.0010$ ), as does a (parametric) two-sample t-test ( $p=0.021$ ). Because



total CIC and  $\Delta_{\min}$  score trees on very different scales, the values are not directly comparable, so, following Wortley and Scotland (2006), we have transformed the scores into ranks under the two criteria. Fig. 3C confirms that most matrices are ranked similarly in information content by both total CIC and  $\Delta_{\min}$ . Extreme differences between  $\Delta_{\min}$  and CIC shown in Table 3 occur for only a minority of datasets – for example, the morphological matrix of Cotton (2001) changes only a single position between the two measures, and more than half of the matrices change by 6 or fewer positions.

## DISCUSSION

Our results do not disagree with Wortley and Scotland's (2006) finding that recent molecular data sets appear to have a greater potential utility than do morphological data sets. However, morphological matrices tend to be ranked more highly under the total CIC measure than under  $\Delta_{\min}$  suggesting that the difference in “informativeness” between the two data types is not as great as other measures might indicate. This further underlines the conclusions of Wortley and Scotland (2006), who found that  $\Delta_{\min}$  shows less difference between morphological and molecular characters than measures like counts of variable characters and counts of parsimony-informative characters.

Wortley & Scotland (2006: 438) noted that "phylogenetic utility as we have measured it here is only one part of the story" because other factors, such as evolutionary conservativeness and levels of homoplasy that are important in the actual (as opposed to potential) phylogenetic utility of character data are not taken into account by their measures. The same limitation affects total CIC. In particular, higher-order methods to quantify the power of a character matrix to reconstruct phylogenetic relationships will need to take account of the interactions between characters. Matrices 1 and 2 from table 4 are two different character matrices for nine taxa of two characters each. In matrix 1, each character permits 351 out of the 135,135 binary trees on nine taxa, so each has

a CIC of 8.59 bits. Considering the matrix as a whole, the two (identical) characters permit the same trees, so 351 trees fit the matrix as a whole with no homoplasy, and the CIC of the whole matrix is also 8.59 bits, while the total CIC of the matrix is 17.18 bits (the sum of the CICs for the two characters). For matrix 2, the two characters have CICs of 8.59 bits and 7.85 bits, but they define different, yet compatible, partitions of the taxa. Only 81 binary trees fit the matrix without homoplasy, giving a CIC of the matrix of 10.70 bits, higher than the CIC of matrix 1, despite the total CIC of matrix 1 being higher than that of matrix 2.

Using the CIC of an entire matrix seems much more natural than using total CIC, but there is a simple problem: almost any real data matrix will contain conflicting characters, and the CIC is not defined for such matrices. If a matrix contains conflict, then, by definition, no tree fits the matrix without homoplasy, so the CIC (the negative logarithm of zero) is undefined. Matrix 3 of Table 4 shows this situation. There are other problems with using this measure. If two characters support the same node in a tree, they will contribute only once to the CIC score of the matrix, but we would usually consider them both to be giving relevant information about the tree topology. This extra information would be reflected in higher values of support measures like Bremer support (Bremer, 1988) and bootstrap proportions (Felsenstein, 1985). Total CIC reflects this kind of extra support, summing the information from each character, and so acts similarly to scores such as  $\Delta_{\min}$  or character-taxon ratio in this respect.

Some measures of character utility avoid this particular problem. The expected information of a particular alignment can, for example, be calculated based on an assumed phylogeny, the rate of evolution of the gene in question, and the parameters of a random process describing its evolution (Goldman, 1998; Massingham & Goldman, 2000). However, all measures of the potential utility of character data that can be easily determined are at best a proxy for actual phylogenetic information. Measures of the latter, such as measuring the power of a matrix to discriminate between alternative phylogenies, or using simulations to assess accuracy, seem to require time-consuming phylogenetic analysis (Ronquist, 1996; Yang, 1998), and so appear inappropriate for large-scale studies comparing many different sources of data. Any tree-independent measure of phylogenetic utility will be approximate, but such measures avoid the uncertainty associated with estimating the phylogenetic tree that produced the data. For example, Wortley & Scotland (2006:437) note that although levels of homoplasy are important they "can only be inferred after a phylogeny has been reconstructed, and never known for certain". However, some proxies for homoplasy, such as pairwise incompatibility between characters, can be calculated without reference to a phylogeny (e.g. Le Quesne, 1969) and merit further attention in this context.

## **ACKNOWLEDGEMENTS**

We thank Davide Pisani and Alex Wortley for helpful discussions, and Claire Pickthall and Davide Pisani for comments on the manuscript. Detailed supplementary results and Perl scripts to calculate CIC and other measures of character information from nexus format data are available from <http://darwin.zoology.gla.ac.uk/~jcotton/charinfo>.

## LITERATURE CITED

(\*indicates references to molecular and morphological datasets discussed in the text or table 3)

**Baker, R.H., Yu, X. & De Salle, R.** 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Molec. Phylog. Evol.* 9: 427--436.

**Bledsoe, A.H. & Raikow, R.J.** 1990. A quantitative assessment of congruence between molecular and nonmolecular estimates of phylogeny. *J. Molec. Evol.* 30: 247--259.

\***Berbee, M.L.** 2001. The phylogeny of plant and animal pathogens in the Ascomycota. *Physiol. and Molec. Pl. Pathol.* 59: 165--187.

\***Bininda-Emonds, O.R. P., Decker-Flum, D. M. & Gittelman, J. L.** 2001. The utility of chemical signals as phylogenetic characters: an example from the Felidae. *Biol. J. Linn. Soc.* 72: 1--15.

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

**Carter, M., Hendy, M.D., Penny, D., Székely, L.A. & Wormald, N.C.** 1990. On the distribution of lengths of evolutionary trees. *SIAM J. Discr. Math.* 3: 38--47.

\***Cotton, T.J.** 2001. The phylogeny and systematics of blind Cambrian ptychoparioid trilobites. *Palaeontology*, 44: 167--207.

\***de Jong, R., Vane-Wright, R.I. & Ackery, P.R.** 1996. The higher classification of butterflies: problems and prospects. *Entomol. Scand.* 27: 65--101.

**Dezulian, T. & Steel, M.** 2004. Phylogenetic closure operations and homoplasy-free evolution, Pp. 395--416 in Banks, D., House, L., McMorris, F. R., Arabie, P. & Gaul, W. (eds.), *Classification, clustering, and data mining applications*. Springer-Verlag, Berlin.

**Estabrook, G.F.** 1978. Some Concepts for the Estimation of Evolutionary Relationships in Systematic Botany. *Syst. Bot.* 3: 146--158.

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

**\*Fliegerová, K., Hodrová, B. & Voigt, K.** 2004. Classical and molecular approaches as a powerful tool for the characterization of rumen polycentric fungi. *Folia Microbiol.* 49: 157--164.

**Goldman, N.** 1998. Phylogenetic information and experimental design in molecular systematics. *Proc. Roy. Soc. London B* 265: 1779--1786.

**Gura, T.** 2000. Bones, molecules ... or both? *Nature* 406: 230--233.

**Hartley, R.V.L.** 1928. Transmission of information. *Bell Sys. Tech. J.* 7: 535--563.

**Hedges, S.B. & Maxson, L.R.** 1996. Molecules and Morphology in amniote phylogeny. *Molec. Phylog. Evol.* 7: 394--395.

**Hedges, S.B. & Maxson, L.R.** 1997. Complementary uses of molecules and morphology: a reply to Lee. *Molec. Phylog. Evol.* 8: 445.

**Hillis, D.M. & Wiens, J.J.** 2000. Molecules versus morphology in systematics: conflicts, artifacts and misconceptions, Pp. 1--19 in Wiens, J. J. (ed.),

*Phylogenetic analysis of morphological data*. Smithsonian Institution Press, Washington DC.

\***Hughes, C.E., Lewis, G.P., Daza Yomona, A. & Reynel, C.** 2004. *Maraniona*, a new dalbergioid legume genus (Leguminosae, Papilionoideae) from Peru. *Syst. Bot.* 29: 366--374.

**Jones, G.A. & Jones, J.M.** 2000. *Information and coding theory*. Springer, London.

\***Koontz, J.A., Soltis, P.S. & Soltis, D.E.** 2004. Using phylogeny reconstruction to test hypotheses of hybrid origin in *Delphinium* section *Diedropetala*. *Syst. Bot.* 29: 345--357.

\***Kropp, B.R. & Matheny, P.B.** 2004. Basidiospore homoplasy and variation in the *Inocybe chelanensis* group in North America. *Mycologia* 96: 295--309.

**Lee, M.S.Y.** 1997. Molecules, Morphology and Phylogeny: A response to Hedges and Maxson. *Molec. Phylog. Evol.* 7: 394--395.

**Le Quesne, M.J.** 1969. A method of selection of characters in numerical taxonomy. *Syst. Zool.* 18: 201—205.

**Massingham, T. & Goldman, N.** 2000. EDIBLE: experimental design and information calculations in phylogenetics. *Bioinformatics* 16: 294--295.

\***Matheny, P.B. & Watling, R.** 2004. A new and unusual species of *Inocybe* (*Inosperma* clade) from tropical Africa. *Mycotaxon* 89: 497--503.

**Meacham, C.A. & Estabrook, G.F.** 1985. Compatibility methods in systematics. *Ann. Rev. Ecol. Syst.* 16: 431--446.



- \*Mostert, L., Halleen, F., Crous, P.W.C. & Creaser, M.L.** 2004. *Cryptovalsa ampelina*, a forgotten shoot and cane pathogen of grapevines. *Australas. J. of Plant Pathol.* 33: 295--299.
- \*Page, R.D.M., Price, R.D. & Hellenthal, R.A.** 1995. Phylogeny of *Geomydoecus* and *Thomomydoecus* pocket gopher lice (Phthiraptera, Trichodectidae) inferred from a cladistic analysis of adult and first instar morphology. *Syst. Entomol.* 20: 129--43.
- Pisani, D., Benton, M.S. & Wilkinson, M.** In Press. The congruence of molecular and morphological trees. *Acta Biotheoretica*.
- \*Prendini, L.** 2000. Phylogeny and classification of the superfamily Scorpionidae Latreille 1802 (Chelicerata, Scorpiones): an exemplar approach. *Cladistics* 16: 1-78.
- Rodrigo, A.G.** 1996. On combining cladograms. *Taxon* 45: 267--274.
- Ronquist, F.R.** 1996. Matrix representation of trees, redundancy, and weighting. *Syst. Biol.* 45: 247--253.
- \*Rouse, G.W.** 2000. Bias? What bias? Gain and loss of down-stream larval feeding in animals. *Zool. Scr.* 29: 213--236.
- Scotland, R.W., Olmstead, R.G. & Bennett, J.R.** 2004. Phylogeny reconstruction: the role of morphology. *Syst. Biol.* 52:539--548.
- Semple, C. & Steel, M.** 2002. Tree reconstruction from multi-state characters. *Adv. Appl. Math.* 28: 169--184.
- Semple, C. & Steel, M.** 2003. *Phylogenetics*. Oxford University Press, Oxford, UK.

- Shannon, C.E.** 1948. A mathematical theory of communication. *Bell Sys. Tech. J.* 27: 379--423,623--656.
- Steel, M. & Penny, D.** 2004. Maximum parsimony and the phylogenetic information in multi-state characters, Pp. 163--178 in Albert, V. (ed.), *Parsimony, phylogeny and genomics*. Oxford University Press, Oxford, UK.
- \*Taylor, J.W. & Purvis, A.** 2002. Have mammals and their chewing lice diversified in parallel?, Pp. 240--261 in Page, R.D.M. (ed.), *Tangled Trees: Phylogeny, Cospeciation, and Coevolution*. University of Chicago Press, Chicago.
- Thorley, J.L.** 2000. *Cladistic information, leaf stability and supertree construction*. PhD thesis, University of Bristol, UK.
- Thorley, J.L., Wilkinson, M. & Charleston, M.A.** 1998. The information content of consensus trees, Pp. 91--98 in Rizzi, A., Vichi, M. and Bock, H.H. (eds.), *Advances in Data Science and Classification: Studies in Classification, Data Analysis and Knowledge Organisation*. Springer, Berlin.
- Thorley, J.L., Cotton, J.A. & Wilkinson, M.** 2004. The information content of trees and their matrix representations. *Syst. Biol.* 53: 898--1001.
- Wiens, J.J.** 2004. The role of morphological data in phylogeny reconstruction. *Syst. Biol.* 53: 653--661.
- Wortley, A.H. & Scotland, R.W.** 2006. Determining the potential utility of datasets for phylogeny reconstruction. *Taxon* 55: 431--442.
- Yang, Z.** 1998. On the best evolutionary rate for phylogenetic analysis. *Syst. Biol.* 47: 125--133.

## FIGURE LEGENDS

Figure 1. CIC and  $\Delta_{\min}$  for different symmetric characters on 120 taxa. Within each character, every character state is present in the same number of taxa, with different characters varying in the number of states. Symmetric characters on 12 taxa are shown in table 1.

Figure 2. Proportion of matrices that uniquely define a tree (i.e. having a single most-parsimonious tree), for different sized matrices of random characters of the same structure as those in Figure 1.

Figure 3. Phylogenetic Informativeness of molecular and morphological data from TreeBase, comparing CIC and  $\Delta_{\min}$ . (A) Scatterplot of CIC vs  $\Delta_{\min}$ . (B) Distribution of CIC scores across datasets. (C) Scatterplot of rank within the 99 datasets under CIC and  $\Delta_{\min}$ . Solid line represents the same rank under both measures. In each panel, black dots represent molecular datasets and grey crosses represent morphological datasets.

## TABLES

Table 1 –  $S_c-1$  and CIC for six symmetric characters on 12 taxa. The characters each have the same number of taxa with each character state, but vary in the number of character states.

	char 1	char 2	char 3	char 4	char 5	char 6
num						
of	1	2	3	4	6	12
states						
A	0	0	0	0	0	0
B	0	0	0	0	0	1
C	0	0	0	0	1	2
D	0	0	0	1	1	3
E	0	0	1	1	2	4
F	0	0	1	1	2	5
G	0	1	1	2	3	6
H	0	1	1	2	3	7
I	0	1	2	2	4	8
J	0	1	2	3	4	9
K	0	1	2	3	5	10
L	0	1	2	3	5	11
$S_c - 1$	0	1	2	3	5	0
CIC						
(bits)	0	9.52	13.32	14.61	13.34	0

Table 2 – Four hypothetical characters with equal number of character states ( $S_c$ ), and so equal contributions to  $\Delta_{\min}$ , but different CIC values.

	char 1	char 2	char 3	char 4
A	0	0	0	?
B	0	0	0	0
C	0	0	1	0
D	1	1	1	1
E	1	1	1	1
F	1	1	1	1
G	2	1	1	?
H	2	2	2	2
I	2	2	2	2
$S_c - 1$	2	2	2	2
CIC (bits)	8.59	7.85	6.63	5.13

Table 3 – Matrices where CIC and  $\Delta_{\min}$  show very different potential utility (20 or more rank places). Rows above the line are ranked more highly by CIC than by  $\Delta_{\min}$ , those below the line are ranked more highly by  $\Delta_{\min}$ , and studies in these sections are ordered by the magnitude of change in rank.

Reference	Ntax	Dataset type	Rank by CIC <sub>char</sub>	Rank by $\Delta_{\min}$	CIC <sub>char</sub> (bits)	$\Delta_{\min}$
Page & al. (1995)	122	morph.	23	61	7164.34	125
Rouse (2000)	91	morph.	29	60	5396.28	133
Berbee (2001)	297	morph.	69	99	1141.06	6
Prendini (2000)	71	morph.	34	59	4534.68	139
De Jong & al. (1996)	75	morph.	28	51	5567.79	159
Taylor & Purvis (2002)	234	morph.	3	25	35474.9	396
Hughes & al. (2004)	95	morph.	22	41	7375.56	205
Bininda-Emonds & al. (2001)	16	morph.	57	31	2107.36	290
Fliegerov & al. (2004)	17	mol.	53	28	2511.07	344
Koontz & al. (2004)	32	mol.	60	26=	1796.51	220
Kropp & Matheny (2004)	16	mol.	45	24	3021.14	403
Mostert & al. (2004)	21	mol.	58	38	1904.03	217
Matheny & Watling (2004)	18	mol.	50	30	2777.76	298

Table 4 - Three different hypothetical two-character matrices showing how different interactions between characters affects CIC. The total CIC of a matrix is the sum of the CICs of the two characters, while the CIC of the matrix is the logarithm of the number of trees permitted by the entire matrix (i.e. that both characters fit without homoplasy) divided by the number of binary trees on nine taxa (see text).

	matrix 1		matrix 2		matrix 3	
	char 1	char 2	char 1	char 2	char 1	char 2
A	0	0	0	0	0	0
B	0	0	0	0	0	1
C	0	0	0	0	1	1
D	1	1	1	1	1	0
E	1	1	1	1	1	0
F	1	1	1	1	1	1
G	2	2	2	1	2	2
H	2	2	2	2	2	2
I	2	2	2	2	2	2
CIC of char	8.59	8.59	8.59	7.85	7.85	8.59
total CIC of matrix	17.18		16.44		16.44	
CIC of matrix	8.59		10.70		undefined	







