A laid-back trip through the Hennigian Forests
(Notes from Naturalists)

Mavrodiev E. V.[1]*, Dell C. A.[1]

[1] Florida Museum of Natural History, University of Florida, PO Box 117800,
Gainesville, FL 32611, USA.

*Email: evgeny@ufl.edu

Abstract

Binary matrix with *a priori* defined plesiomorphic character-state may be re-written as a simple set of branching diagrams (as a "Hennigian forest"). The last might be analyzed by the average consensus method. This procedure eventually avoids the taxon-character matrix from the analysis of the data. Within this framework, the criteria of the best topologies based on the character-state changes are unnecessary. The solely 'reversal'-based groups are always appear within the average consensus topologies despite the lack of direct evidence from the primary data.

Key words: Binary and multistate characters, cladograms, matrix-free Cladistics, maximal relationship, character polarity, synapomorphy, average consensus

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2487v3 | CC BY 4.0 Open Access | rec: 6 Oct 2016, publ: 6 Oct 2016

Here it would seem more appropriate to re-write the characters of Matrix 1 from Nelson (1996) in a tree-form representing the relationships exactly

Williams and Ebach, 2006: 414

Introduction

A. Summary of used and implied concepts and selected references:

a. An ordinary (phenetic) matrix and Matrix-free Cladistics (Platnick, 1993; Williams and Ebach, 2006, 2008);

b. The symmetry between binary character and cladogram, as conceptualized by Williams (1994, 1996);

c. Maximal relationship (Nelson and Platnick, 1981; summarized in Williams and Ebach, 2006, 2008);

d. An outgroup comparison (Platnick and Gertsch, 1976; Watrous and Wheeler, 1981, reviewed in Nixon and Carpenter, 1993);

e. *A priori* determination of character's polarity (reviewed in Kitching et al. 1998 and Waegele 2005; see also Williams and Ebach, 2008 and Wiley and Lieberman, 2011);

f. Grouping solely on synapomorphy (Hennig, 1966; see also Platnick, 1985; Nelson, 2004; Williams and Ebach, 2006, 2008);

g. The lack of the equation between Hennigian cladistics and Wagner's algorithm or its optimization-based derivatives (Nelson, 2004; Williams and Ebach 2006, 2008);

h. Average consensus method (Lapointe and Cucumel, 1997; Lapointe and Levasseur, 2004) and its extensions (Creevey, 2004, see also Lapointe and Levasseur, 2004);

i. Matrix representation with Parsimony (Baum, 1992, Ragan, 1992);

j. Three-taxon statement analysis (Nelson and Platnick, 1991);

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2487v3 | CC BY 4.0 Open Access | rec: 6 Oct 2016, publ: 6 Oct 2016

B. Initial propositions

A binary character is a tree with one informative node (Platnick et al., 1996; Williams, 1994; Wiliams and Ebach 2006, 2008). For example, if state 0 is plesiomorphic and state 1 is apomorphic, then the character ABCDE/00011 *is a rooted tree* ABC(DE), where (DE) is a clade (monophyletic group) defined based on the apomorphic character-state. If all of the clades of the tree are defined based on apomorphic character-state, we may call this tree a “Hennigian”. Hennigian trees may also be easily seen as simple hierarchies of two character-states.

Let A be defined as the outgroup. In this case the Hennigian tree ABC(DE) may be re-written as A(BC(DE)) (or as a (A(BC(DE)))), *even if strictly speaking we do not have any formal evidence for the groups* (BCDE) (or (ABCDE)). For example, in the same manner, the three-taxon statement (Nelson and Platnick, 1991) A(BC)/011 may be *technically* re-written as (A(BC)), even if there in no evidence for the group (ABC).

As mentioned, the character state zero *a priori* is defined as plesiomorphic therefore groups (BC) or (ABC) are prohibited. So, below we accept that trees ABC(DE) or A(BC(DE)) are Hennigian, while A((BC)(DE)) or ((ABC)(DE)) are not.

Let us consider tree A(BC(DE)) for one more time. If value of taxon B is missing, two solutions appear to be possible.

First, character ABCDE can be re-written as two trees, assuming that the missing value may be either zero or one: ABC(DE) = A(BC(DE)) or A(C(BDE)). Another, better possibility implies the exclusion of the taxon B from the tree. This reduces the character ABCDE to the tree AC(DE).

Characters like ABCDE/00001, 000?1, ??001, 00000 etc. collapse to polytomies (ABCDE).

The basic idea of Matrix Representation with Parsimony (Baum, 1992, Ragan, 1992) states that the tree (cladogram) can be represented as a binary matrix. Here we propose something opposite – we are showing that the binary character may be re-written as a tree (Williams, 1994; Siebert and Williams, 1998: 342; Williams and Ebach, 2006: 414; see also Platnick et al., 1996). This is should be clear, because “*the* *binary character and the cladogram are one and the same thing*, representation of the relationship" (Williams, 1994: 451-452, italics mine), therefore the binary matrix can be viewed as a set (forest, array etc.) of branching diagrams (cladograms) that may be called as a "Hennigian forest" if one of the character-states (for example, the character-state zero) *a priori* is defined as plesiomorphic.

Hennigian forest might be analyzed by the average consensus method (Lapointe and Cucumel, 1997; Lapointe and Levasseur, 2004), one of the first techniques of the construction of supertrees (reviewed in Bininda-Emonds, 2014 and Lapointe and Levasseur, 2004).

As summarized by Lapointe and Levasseur (2004: 87), the average consensus procedure is a method that takes as input a profile of weighted trees and returns a consensus tree that is in some sense "closest" to the entire profile. This method allows us to operate with the Hennigian trees directly, completely excluding the binary data matrix from the analysis. Within this framework, the criteria of the best topologies, such as a minimal number of character-state changes, as well as the standard optimization procedures, if optimization is defined as a way to reconstruct the sequences of the character-state changes on a tree, all appeared as unnecessary.

Materials and Methods

Despite the practically non-observable ocean of publications within the field of contemporary Bioinformatics, no developed software is able to represent the binary matrix as a Hennigian forest. Given this, we wrote the simple ruby-based script FORESTER version 1.0 (named as a "FORESTER" below) that helps represent the binary data as Hennigian trees for future manipulations.

The FORESTER script (available upon request from the second author) process each input file by storing the text of each line of the file as an element of an array. It finds the beginning and ending location (index) of the characters in the matrix, and loops from the starting index to the ending index, writing a line of the new file(s) at the end of each loop.

For the binary Matrix the usage of the script is: **ruby trees.rb –inputfilename**

1. The out-group taxon should be placed last in the matrix;

2. Output tree files are available as a result of the run: the first file contains polytomies such as (ABCDE), while the second and the third appear without such polytomies, but trees may be rooted relative to the basal polytomy (e.g., (ABC(DE))) or relative to *a priori*-defined outgroup taxon (e.g., (A(BC(DE))))

All tree files are, and the input binaries should be written in the "relaxed" Newick (PHYLIP) format (reviewed in Felsenstein, 1989, see also Maddison and Maddison, 2011).

The minimal trees (or the three-taxon statements (3TS)) are not subjects of our recent considerations, but FORESTER contains options for the speedy rewriting of the 3TS matrices as the forests of the minimal trees. The usage should be: **ruby seedlings.rb –inputfilename** with only one output Newick file saved as a result of the run.

The binary representation of the genomic (DNA) data has been made by script *1001* (Mavrodiev, 2015a). For example, if Method 1 of *1001* (Mavrodiev, 2015a) is selected, as an option the "reduced" polarized matrix (invariant characters removed automatically)(Mavrodiev, 2015a) can be chosen as a source for the Hennigian trees.

The average consensus tables (Lapointe and Cucumel, 1997; Lapointe and Levasseur, 2004) are calculated following the default settings of Clann version 3.00 (Creevey, 2004; Creevey, McInerney, 2005, 2009) for the **avcon** command with the branch lengths of the all input trees assigned to unity (Creevey, 2004) and also used as future inputs for PAUP* 4.0b10 (Swofford, 2002). All trees treated as weighted equally.

In most of the analyses the **optimality criterion of the best tree is defined as distance with non-weighted least squares** (Swofford, 2002). In case of the analysis of the Hennigian forest of Matrix Z (Farris, 1997) we used the weighted least squares (exponential weights with P'=7 (Swofford, 2002)).

Maximum Parsimony (MP) analyses have been conducted in PAUP* 4.0b10 (Swofford, 2002), with a heuristic search of 1000 random addition replicates (saving no more than 100 trees per replicate), and the TBR branch swapping/MulTrees option in effect; the gaps were treated as "missing".

All binary matrices are taken from the literature (Kluge, 1994; Farris, 1997, Farris, Kluge, 1998). The genomic DNA matrix-example has been downloaded from the Online supplement of Barrett et al. (2016: Fig. 4 and S5).

The routine manipulations with the matrices and the tree-files were performed with Mesquite v. 3.01 (Maddison and Maddison, 2011), PAUP* v. 4.0b10 (Swofford, 2002) and FigTree v. 1.4.2 (Rambaut, 2012).

Results and Discussion

In Cladistics, groups must be based solely on "derived" or apomorphic character-states (e.g., Hennig, 1966; Platnick, 1985; Nelson, 2004; Williams and Ebach, 2008). For example, in case of the binary matrix, where the character state 0 is defined as plesiomorphic before the analysis, all groups should be based solely on state 1 (Platnick, 1985). Analyzing all selected matrices (Fig. 1-4) we are followed the general logic of three-taxon statement analysis (3TA) (Nelson and Platnick, 1991): we tried to explicate all possible synapomorphic-based maximal trees *a priori* (1.) in order to find the best-fitting topology as a next step (2.).

However, there are two major differences between our approach and the 3TA as originally proposed and implemented by Nelson and Platnick (1991). The 3TA deals with the three-taxon statements (3TS) or minimal relationships (Williams and Siebert, 2000; Williams and Ebach, 2006, 2008). The set of 3TS must be establishing as a binary matrix and used as an input for the parsimony program (Nelson and Platnick, 1991; see also Mavrodiev and Madorsky, 2012). In this paper, we operate directly with the maximal relationships (Nelson and Platnick, 1981;

summarized in Williams and Ebach, 2006, 2008) *written right as trees*, not as binary matrices, while also using the average consensus method instead of the MP analysis.

In summary:

1. The script FORESTER helps to rewrite every character of the binary matrix as a tree (or as a maximal relationship) in a cladistics way, or in a manner where all of the groups are based only on *a priori* defined apomorphic character state "1" (Fig. 1-4).
2. Next, the forest of this "maximal" or Hennigian trees can be used to calculate their average consensus (Lapointe and Cucumel, 1997; Lapointe and Levasseur, 2004) (Fig. 1-4).

A possible issue of the Hennigian approach to the data is the inability to operate with the putative reversals (reviewed in Farris and Kluge, 1998; Siebert and Williams, 1998 and Mavrodiev, 2015b), unless, however, the 'reversals' are scored as separate apomorphic character-states before the analysis (Mavrodiev, 2015b).

Paradoxically, the average consensus method, if applying to the Hennigian forests, helps to identify the 'reversal'-based clades without the separate scoring of reversals, *despite* the absence of evidence from the primary data (Fig. 1, 2). A similar effect had been described by Nelson and Platnick (1991) for 3TA and discussed in more details by Siebert and Williams (1998).

For example, let consider the modified matrix 17 from Nelson and Platnick (1991) (Fig. 1). According to the MP optimization, group EF is based on the plesiomorphic character-state zero and therefore may be treated as a putative 'reversal'-based clade (Nelson and Platnick, 1991; Mavrodiev, 2015b). However, *neither of the trees from the array c. (Fig. 1) contains any plesiomorphic-based groups, but, the clade (EF) still defined on the topology of the average consensus* (Fig. 1). Therefore, *clade (EF) appears straight through the analysis* of forest c. (Fig. 1) *despite the lack of direct evidence from the raw data* (see also Nelson and Platnick, 1991: 363 for the similar discussion). The same situation is detected in the cases of Table 1 from Kluge (1994) (Fig. 2, IIa. – c.), and the Table 5 from Kluge and Farris (1998) (Fig. 2, IIIa. – c.).

The average consensus technique, if applying to the Hennigian forests, may also help to avoid more potentially negative effects of putative reversals. For example, regarding his Table 3, Kluge (1994: 408-410) mentioned that taxa G and F are highly supported sisters with taxon G exhibiting reversals only in characters one and two (Fig. 3, I). As discovered by Kluge (1994), the 3TA removed taxon G from F, despite the strong evidence of their relationship. A similar situation appears in the case of Matrix Z, designed by Farris (Farris, 1997) simply by duplication of the Table 3 from Kluge (1994) (Farris, 1997; Siebert and Williams, 1998).

However, the topologies of the average consensuses of the Hennigian forests of the Table 3 from Kluge (1994) and Matrix Z from Farris (1997: 136)(Fig. 3, Ia. – c., IIa. – c.) appeared to be identical to the topologies, which were the result of the MP analyses of Table 3 and Matrix Z (Kluge, 1994; Farris, 1997). 3TA, with the fractional weighting (Nelson and Ladiges, 1992; Williams and Ebach, 2008) can also compensate the negative effect of putative reversals in identical situations (Siebert and Willimas, 1998).

As well as canonical 3TA (Nelson and Platnick, 1991), the average consensus analysis of the forests of maximal relationships can successfully recognize groups for which the standard optimization criteria of the MP analysis produce no unequivocal synapomorphies (Nelson, 1996; Williams and Ebach, 2005) (Fig. 2, I).

For example, the group BCD is successfully recognizable after the average consensus analysis of the three trees, each representing the binary character from Matrix 1 from Nelson (1996)(Nelson, 1996; Williams and Ebach, 2005, 2006)(Fig. 2, Ia. – d.).

Williams and Ebach (2006: 414) offered very similar solution for the group (BCD). These authors mentioned that *it would seem more appropriate to re-write the characters of Matrix 1 from Nelson (1996) in a tree-form representing the relationships exactly*, such that the three "characters" are AD(BC), AC(BD) and AB(CD), which, when *combined*, unambiguously provide evidence for the solution A(BCD) (Williams and Ebach 2006: 414, italics mine). But in contrast to the intuitive hand-made solution (Williams and Ebach 2006: 414), as well as to the results of the the average consensus method (Fig. 2.I.), neither of strict, majority rule, Adams's, or combinable component consensuses (see Kitching et al., 1998 and Swofford, 2002 for the reviews and implementation of the methods) of the three trees AD(BC), AC(BD), and AB(CD)(Williams and Ebach 2006: 414) are able to recognize group (BCD).

Williams and Ebach (2006: 414) also noted that the re-writing of the binary characters effectively converts an ordinary (phenetic) matrix into a "Cladistic matrix". In our mind, Cladistics matrix is similar, but not the same entity as a simple forest of the Hennigian trees (see Siebert and Williams, 1998: 242 for the initial simple examples of the Cladistics matrices called by these authors as a "tabular formulations of the 3TA" of the conventional data; see also Williams and Ebach 2006: 412 and Williams and Ebach, 2008).

In numerous cases, the simplest options of search (the least-square criteria of the fit plus un-weighted algorithms) should be sufficient to estimate the reasonable average consensuses. However, the average consensus approach allows various schemes of weighing as well as different optimization criteria (Creevey, 2004; Lapointe and Levasseur, 2004). For example, using of the exponential weights of the least squares may increase the efficiency of the search (Fig. 3, II.).

The proposed approach may also easily handle the real data, for example the alignments of the plastid genomes (Fig. 4). Some ways to do this may be possible.

For example, for someone it may seem to be necessary to:

1. Re-recode the conventional DNA matrix as a binary matrix;
2. Establish the this matrix as a Hennigian forest, and
3. Estimate the average consensus topology that fit this Hennigian forest (Fig. 4).

So, even in the cases of the molecular data, the taxon-character matrices (molecular alignments) may be necessary only as a “soil” for growing the Hennigian forests – the arrays of maximal relationships.

We therefore concluded that, the average consensus method, if applied to the forest of the Hennigian trees, help us to exclude the taxon-character matrix from the analysis of the data. As a result, the standard optimization-based criteria of the best topologies, such as the length of the tree, appear to be unnecessary.

Acknowledgments

Dr. David Williams (London Museum of Natural History) and Dr. Malte Ebach (University of New South Wales and the Sydney's Australian Museum, AU) are greatly acknowledged for their discussion and helpful comments and notes. No agreement is implied from the behalf of both persons listed in this section.

Figure Legends

Figure 1. a. Most parsimonious tree of the length = 5 (CI = 0.80, RI = 0.86) based on modified binary Matrix **b.** (same as modified Matrix 17 from Nelson and Platnik, 1991: 362 with character four excluded)/the average consensus topology of the score 0.22222 calculated based on table (**d**.); **b**. Modified Matrix 17 from Nelson and Platnick (1991: 362)(character 4 excluded); **c**. four characters of the modified Matrix 17 (**b**.) established as rooted trees; **d**. the average consensus table based on **c**. calculated by Clann (Creevey, 2004; Creevey, McInerney, 2005, 2009).

Figure 2. I. a. Matrix 1 from Nelson (1996)(Nelson, 1996; Williams and Ebach, 2005); **b**. three characters of Matrix 1 from Nelson (1996) (Nelson, 1996; Williams and Ebach, 2005) (**a**.) established as a rooted trees; **c**. three average consensuses of the forest (**b**.) of the score zero and their strict consensus (**d**.). **II. a**. Table 2 from Kluge (1994: 408); **b**. nine characters of the Table 2 from Kluge (1994) (**a**.) established as a rooted trees; **c**. strict consensus of three average consensuses of the score 0.27118 resulted the analysis of the forest (**b**.). **III. a**. Figure 5 from Farris and Kluge (1998: 353); **b**. 23 characters of Figure 5 from Farris and Kluge (1998) (**a**.) established as a rooted trees; **c**. the average consensus topology of the score 1.00184 resulted the analysis of the forest (**b**.).

Figure 3. I. a. Table 3 from Kluge (1994: 409); **b**. six characters of Table 3 from Kluge (1994) (**a**.) established as rooted trees; **c**. the average consensus topology of the score 0.29213 resulted in the analysis of the forest (**b**.). **II. a**. Matrix Z from Farris (1997: 136, Fig. 4); **b**. 12 characters of Matrix Z from Farris (1997) (**a**.) established as a rooted trees; **c**. the average consensus topology of the score 1.9603e-08 resulted the analysis of the forest (**b**.).

Figure 4. I. a. Topology resulted MP analysis of 159074 bp alignment of complete plastomes of palms (Arecales), Dasypogonales, and *Typha latifolia* (Poales, Thyphaceae) from Barrett et al. (2016) (Fig. 4 and S5), *a posteriori* rooted relatively

the cattail (*Typha*)(length = 66788, CI = 0.7363, RI = 0.6109). The number of the parsimony-informative characters is equal to 19124. Original parametric tree from Barrett et al. (2016) looks very similar and did not reproduced here by this reason. The same topology resulted MP analysis of the polarized binary matrix with invariant characters removed, recoded from conventional DNA alignment from Barrett et al. (2016) (Fig. 4 and S5) (**a**.) following Method 1 described in Mavrodiev (2015b). *Typha latifolia* assumed as the best all-plesiomorphic group. The length of the topology is equal to 115525, the CI = 0.6743, and the RI = 0.7207. The number of informative characters is equal to 37863, the total number of characters in binary matrix equal to 77895. **II.** The average consensus topology of the score 0.00225 resulted the analysis of the forest of the 77 895 trees derived from the binary representation of the original 159074 bp alignment of complete plastomes of palms (Arecales), Dasypogonales, and *Typha latifolia* (Poales, Thyphaceae) from Barrett et al. (2016) (see **Ia**. For the details).

References

Barrett CF, WJ Baker, JR Comer, JG Conran, SC Lahmeyer, JH Leebens-Mack, J Li, GS Lim, DR Mayfield-Jones, L Perez, J Medina, JC Pires, C Santos, D Wm Stevenson, WB Zomlefer, JI Davis 2016 Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. The New Phytologist 209: 855-870.

Baum BR 1992 Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41: 3-10.

Bininda-Emonds OR 2014 An introduction to supertree construction (and partitioned phylogenetic analyses) with a view toward the distinction between gene trees and species trees. Pages 49-76 *in* L Garamszegi, Z ed. Modern phylogenetic comparative methods and their application in evolutionary biology. Springer, Berlin-Heidelberg. Germany.

Creevey C. 2004. Clan: Construction of supertrees and exploration of phylogenomic information from partially overlapping datasets (version 3.0.0), user manual. The lab of James McInerney, Manchester, Great Britan.

Creevey CJ, JO McInerney 2005 Clann: Investigating phylogenetic information through supertree analyses. Bioinformatics 21: 390-392.

Creevey CJ, JO McInerney 2009 Trees from trees: Construction of phylogenetic supertrees using clann. Bioinformatics for DNA Sequence Analysis 537: 139-161.

Farris JS 1997 Cycles. Cladistics 13: 131-144.

Farris JS, AG Kluge 1998 A/the brief history of three-taxon analysis. Cladistics 14: 349-362.

Felsenstein J 1989 PHYLIP - phylogeny inference package (version 3.2). Cladistics 5: 164-166.

Hennig W 1966 Phylogenetic Systematics. University of Illinois Press, Urbana, IL., USA.

Kitching IJ, PL Forey, CJ Humphries, DM Williams 1998 Cladistics. The theory and practice of parsimony analysis. 2nd edition. Systematics Association

Publication.

Kluge AG 1994 Moving targets and shell games. Cladistics 10: 403-413.

Lapointe FJ, C Levasseur 2004 Everything you always wanted to know about the average consensus, and more. Pages 87-105 *in* Bininda-Emonds, RP. ed. Phylogenetic supertrees. Computational biology. Kluwer Academic Publishers.

Lapointe FJ, G Cucumel 1997 The average consensus procedure: Combination of weighted trees containing identical or overlapping sets of taxa. Systematic Biology 46: 306-312.

Maddison WP, DR Maddison. 2011. Mesquite: A modular system for evolutionary analysis. Version 3.01. http://mesquiteproject.org/

Mavrodiev EV 2015a *1001*-a tool for binary representations of unordered multistate characters (with examples from genomic data). PeerJ PrePrints 3, e1403 https://doi.org/10.7287/peerj.preprints.1153v1

Mavrodiev EV 2015b Three-taxon analysis can always successfully recognize groups based on putative reversals. PeerJ PrePrints 3, e1206. doi.org/10.7287/peerj.preprints.979v1

Mavrodiev EV, A Madorsky 2012 Taxodium version 1.0: A simple way to generate uniform and fractionally weighted three-item matrices from various kinds of biological data. PLoS One 7(11): e48813. doi:10.1371/journal.pone.0048813

Nelson G 1996 ‘Nullius in verba.’ (Self-published: New York)

Nelson G 2004 Cladistics: its arrested development. Pages 127-147 *in* DM Williams and PL Forey, eds. Milestones in Systematics, Systematics Association Special Volume Series 67. CRC Press, Boca Raton, FL, USA.

Nelson G, PY Ladiges 1992 Information-content and fractional weight of 3-item statements. Systematic Biology 41: 490-494.

Nelson G, NI Platnick 1981 Systematics and biogeography: Cladistics and Vicariance. Columbia University Press.

Nelson G, NI Platnick 1991 Three-taxon statements - a more precise use of parsimony? Cladistics 7: 351-366.

Nixon KC, JM Carpenter 1993 On outgroups. Cladistics 9: 413-426.

Platnick NI 1985 Philosophy and the transformation of cladistics revisited. Cladistics 1: 87-94.

Platnick NI 1993 Character optimization and weighting - differences between the standard and three-taxon approaches to phylogenetic inference. Cladistics 9: 267-272.

Platnick NI, WJ Gertsch 1976 The suborders of spiders: A cladistic analysis (Arachnida, Araneae). American Museum Novitates No. 2607: 1-15.

Platnick NI, CJ Humphries, G Nelson, DM Williams 1996 Is Farris optimization perfect?: Three-taxon statements and multiple branching. Cladistics 12: 243-252.

Ragan MA 1992 Phylogenetic inference based on matrix representation of trees. Molecular Phylogenetics and Evolution 1: 53-58.

Rambaut A 2012 FigTree v. 1.4.2 Molecular evolution, phylogenetics and epidemiology. Edinburgh, UK: University of Edinburgh, Institute of Evolutionary Biology. http://tree.bio.ed.ac.uk/software/figtree/

Siebert DJ, DM Williams 1998 Recycled. Cladistics 14: 339-347.

Swofford DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer Associates, Sunderland, Massachusetts.

Waegele JW 2005 Foundations of phylogenetic systematics. Pfeil Verlag, München. Germany.

Watrous LE, QD Wheeler 1981 The out-group comparison method of character analysis. Systematic Zoology 30: 1-11.

Wiley EO, BS Lieberman 2011 Phylogenetics: The theory and practice of phylogenetic systematics. 2nd edition. John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Williams DM 1994 Combining trees and combining data. Taxon 43: 449-453.

Williams DM 1996 Characters and cladograms. Taxon 45: 275-283.

Williams DM, MC Ebach 2005 Drowning by numbers: Rereading Nelson's "Nullius in verba". Botanical Review 71: 415-447.

Williams DM, MC Ebach 2006 The data matrix. Geodiversitas 28: 409-420.

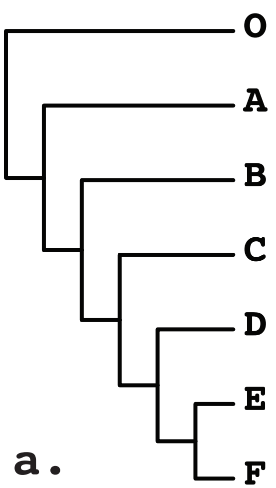
Williams DM, MC Ebach 2008 Foundations of systematics and biogeography.

Springer, New York, USA.

Williams DM, DJ Siebert 2000 Characters, homology and three-item analysis. Pages 183-208 *in* RW Scotland and RT Pennington, eds. Homology and systematics: coding characters for phylogenetic analysis. Systematics Association Special Volume Series 58, Chapman and Hall, London, New York

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2487v3 | CC BY 4.0 Open Access | rec: 6 Oct 2016, publ: 6 Oct 2016

Most parsimonious tree/ Average consensus topology



←

Maximum Parsimony analysis

(e.g., PAUP*)

Polarized Binary Matrix (Nelson and Platnick (1991): modified Matrix 17)

```
A 0000
B 1001
C 1101
D 1111
E 1110
F 1110
O 0000
```

b.

→

FORESTER

Characters of the Binary Matrix (b.), re-written as a trees in Newick (Phylip) notation (”Hennigian forest”)

```
(O(A(BCDEF)))
(O(AB(CDEF)))
(O(ABC(DEF)))
(O(AEF(BCD)))
```

c.

↑

Average consensus Method - step 2 (e.g., PAUP*): calculation of the consensus (a.) based on the Table (d.)

Average consensus Method - step 1 (CLANN): calculation of the the Table (d.) based on the Hennigian forest (c.)

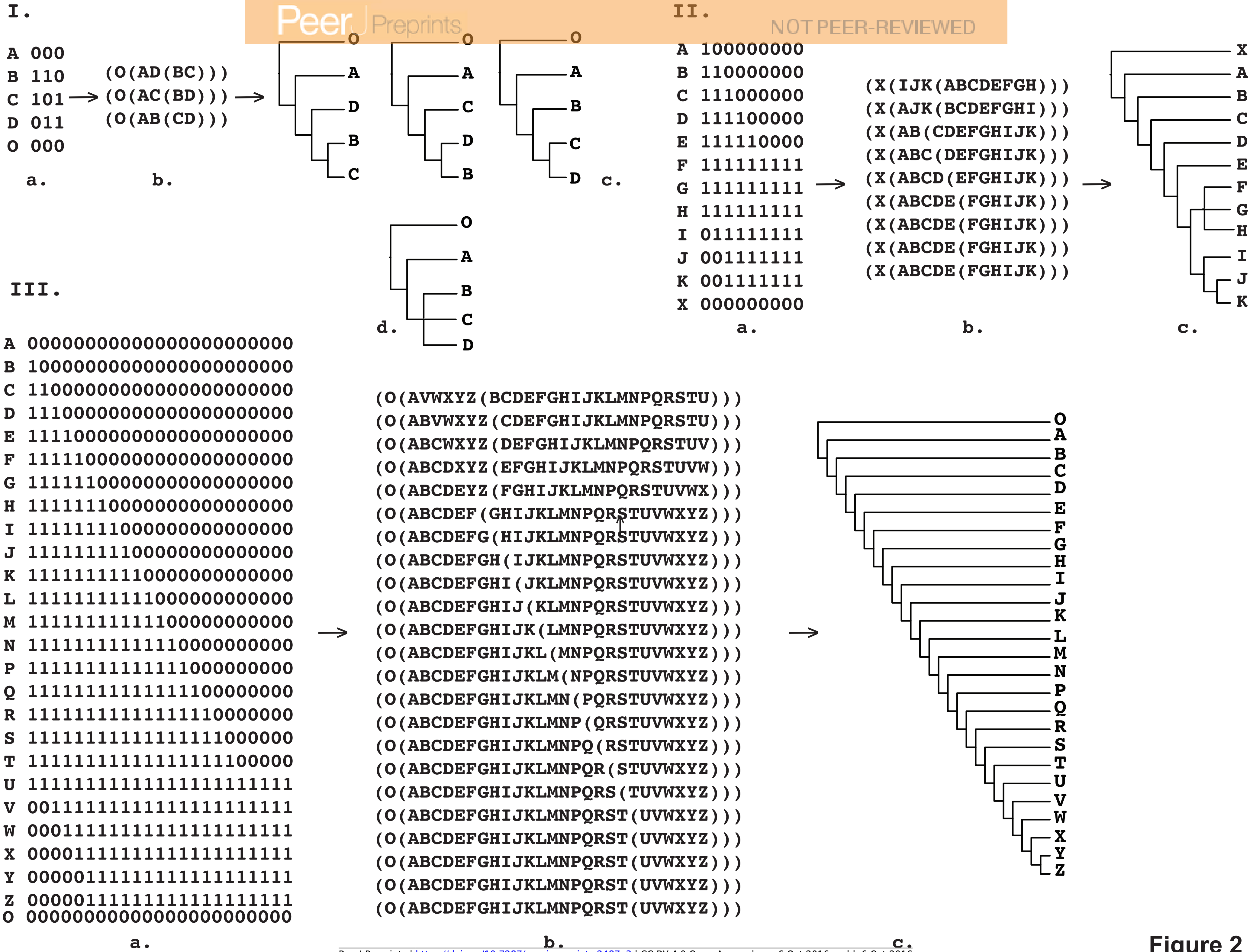
↓

```
O
A 2.000000
B 2.500000 2.500000
C 2.750000 2.750000 2.250000
D 3.000000 3.000000 2.500000 2.250000
E 2.750000 2.750000 2.750000 2.500000 2.250000
F 2.750000 2.750000 2.750000 2.500000 2.250000 2.000000
```

d.

Figure 1

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2487v3 | CC BY 4.0 Open Access | rec: 6 Oct 2016, publ: 6 Oct 2016



I.

A 000
B 110
C 101
D 011
O 000

a.

(O(AD(BC)))
(O(AC(BD)))
(O(AB(CD)))

b.

II.

A 100000000
B 110000000
C 111000000
D 111100000
E 111110000
F 111111111
G 111111111
H 111111111
I 011111111
J 001111111
K 001111111
X 000000000

a.

(X(IJK(ABCDEFGH)))
(X(AJK(BCDEFGHI)))
(X(AB(CDEFGHIJK)))
(X(ABC(DEFGHIJK)))
(X(ABCD(EFGHIJK)))
(X(ABCDE(FGHIJK)))
(X(ABCDE(FGHIJK)))
(X(ABCDE(FGHIJK)))
(X(ABCDE(FGHIJK)))

b.

III.

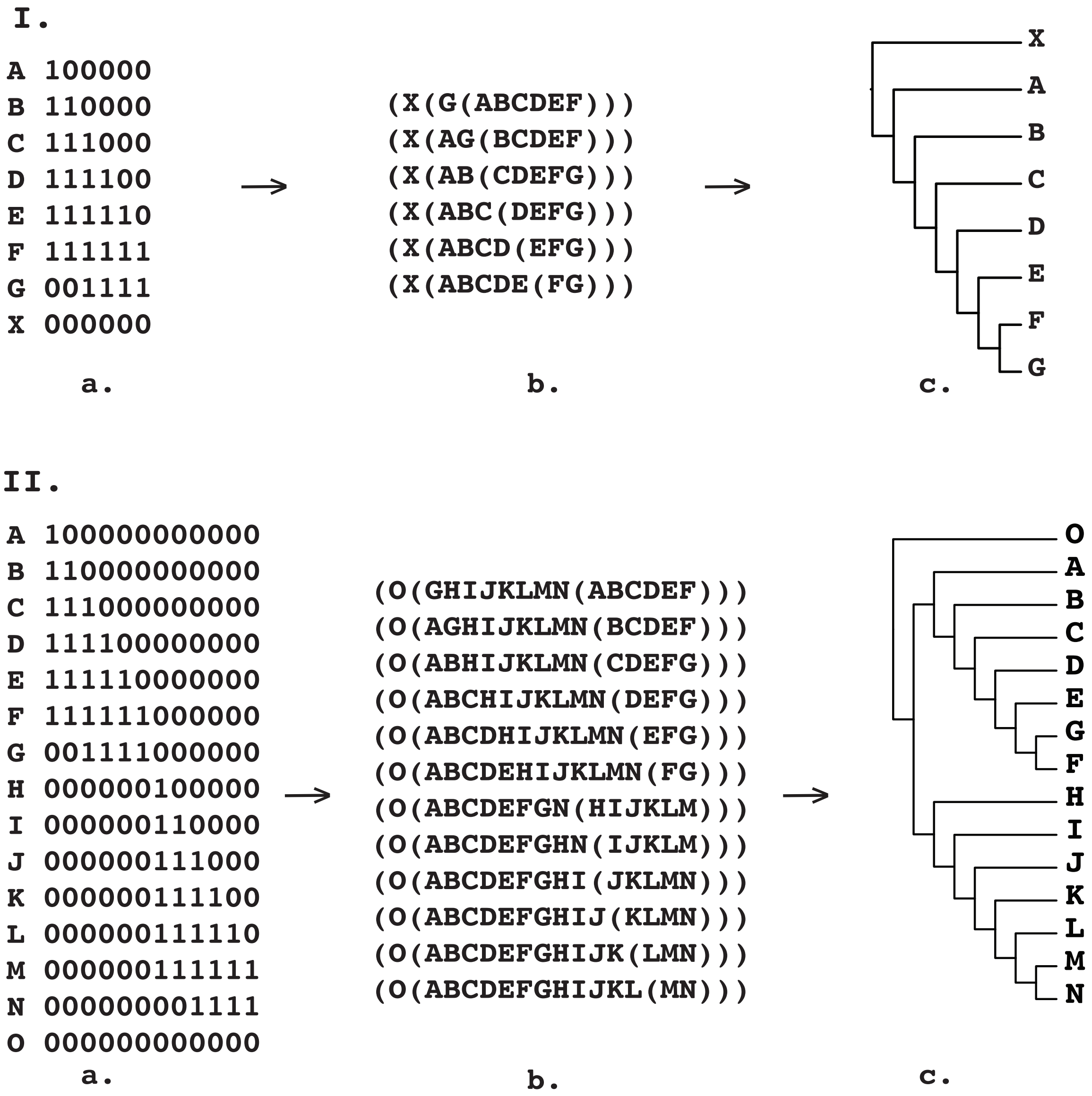
A 00000000000000000000000
B 10000000000000000000000
C 11000000000000000000000
D 11100000000000000000000
E 11110000000000000000000
F 11111000000000000000000
G 11111100000000000000000
H 11111110000000000000000
I 11111111000000000000000
J 11111111100000000000000
K 11111111110000000000000
L 11111111111000000000000
M 11111111111100000000000
N 11111111111110000000000
P 11111111111111000000000
Q 11111111111111100000000
R 11111111111111110000000
S 11111111111111111000000
T 11111111111111111100000
U 11111111111111111111111
V 00111111111111111111111
W 00011111111111111111111
X 00001111111111111111111
Y 00000111111111111111111
Z 00000111111111111111111
O 00000000000000000000000

a.

(O(AVWXYZ(BCDEFGHIJKLMNPQRSTU)))
(O(ABVWXYZ(CDEFGHIJKLMNPQRSTU)))
(O(ABCWXYZ(DEFGHIJKLMNPQRSTUV)))
(O(ABCDXYZ(EFGHIJKLMNPQRSTUVW)))
(O(ABCDEYZ(FGHIJKLMNPQRSTUVWX)))
(O(ABCDEF(GHIJKLMNPQRSTUVWXYZ)))
(O(ABCDEFG(HIJKLMNPQRSTUVWXYZ)))
(O(ABCDEFGH(IJKLMNPQRSTUVWXYZ)))
(O(ABCDEFGHI(JKLMNPQRSTUVWXYZ)))
(O(ABCDEFGHIJ(KLMNPQRSTUVWXYZ)))
(O(ABCDEFGHIJK(LMNPQRSTUVWXYZ)))
(O(ABCDEFGHIJKL(MNPQRSTUVWXYZ)))
(O(ABCDEFGHIJKLM(NPQRSTUVWXYZ)))
(O(ABCDEFGHIJKLMN(PQRSTUVWXYZ)))
(O(ABCDEFGHIJKLMNP(QRSTUVWXYZ)))
(O(ABCDEFGHIJKLMNPQ(RSTUVWXYZ)))
(O(ABCDEFGHIJKLMNPQR(STUVWXYZ)))
(O(ABCDEFGHIJKLMNPQRS(TUVWXYZ)))
(O(ABCDEFGHIJKLMNPQRST(UVWXYZ)))
(O(ABCDEFGHIJKLMNPQRST(UVWXYZ)))
(O(ABCDEFGHIJKLMNPQRST(UVWXYZ)))
(O(ABCDEFGHIJKLMNPQRST(UVWXYZ)))
(O(ABCDEFGHIJKLMNPQRST(UVWXYZ)))

b.

Figure 2



I.

a.

```
A 100000
B 110000
C 111000
D 111100
E 111110
F 111111
G 001111
X 000000
```

b.

```
(X(G(ABCDEF)))
(X(AG(BCDEF)))
(X(AB(CDEFG)))
(X(ABC(DEFG)))
(X(ABCD(EFG)))
(X(ABCDE(FG)))
```

c.

II.

a.

```
A 100000000000
B 110000000000
C 111000000000
D 111100000000
E 111110000000
F 111111000000
G 001111000000
H 000000100000
I 000000110000
J 000000111000
K 000000111100
L 000000111110
M 000000111111
N 000000001111
O 000000000000
```

b.

```
(O(GHIJKLMN(ABCDEF)))
(O(AGHIJKLMN(BCDEF)))
(O(ABHIJKLMN(CDEFG)))
(O(ABCHIJKLMN(DEFG)))
(O(ABCDHIJKLMN(EFG)))
(O(ABCDEHIJKLMN(FG)))
(O(ABCDEFGN(HIJKLM)))
(O(ABCDEFGHN(IJKLM)))
(O(ABCDEFGHI(JKLMN)))
(O(ABCDEFGHIJ(KLMN)))
(O(ABCDEFGHIJK(LMN)))
(O(ABCDEFGHIJKL(MN)))
```

c.

Figure 3

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2487v3 | CC BY 4.0 Open Access | rec: 6 Oct 2016, publ: 6 Oct 2016

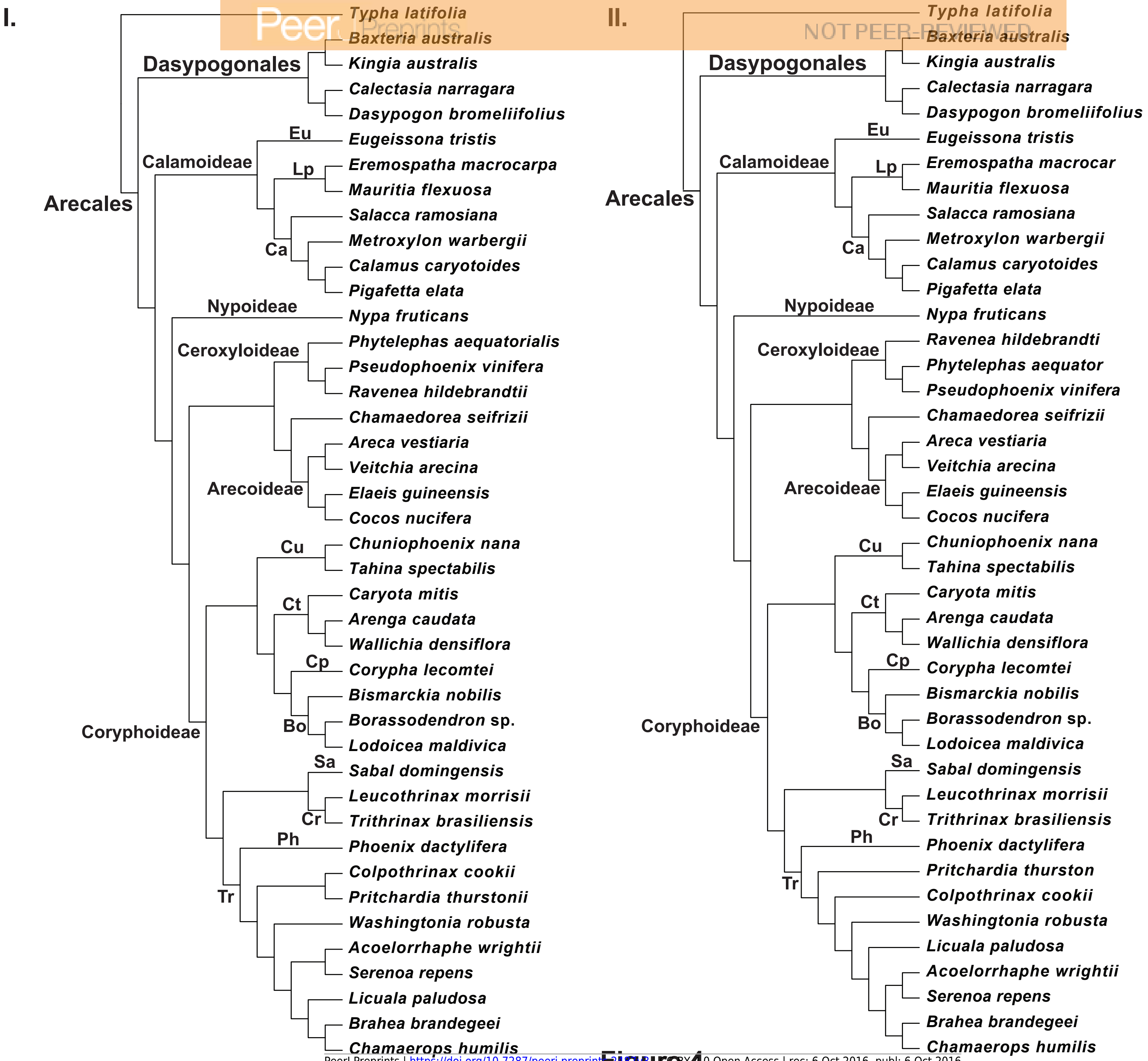

I.
Typha latifolia
Baxteria australis
Kingia australis
Calectasia narragara
Dasypogon bromeliifolius
Dasypogonales
Eu
Eugeissona tristis
Calamoideae
Lp
Eremospatha macrocarpa
Mauritia flexuosa
Arecales
Salacca ramosiana
Metroxylon warbergii
Ca
Calamus caryotoides
Pigafetta elata
Nypoideae
Nypa fruticans
Ceroxyloideae
Phytelephas aequatorialis
Pseudophoenix vinifera
Ravenea hildebrandtii
Chamaedorea seifrizii
Areca vestiaria
Veitchia arecina
Arecoideae
Elaeis guineensis
Cocos nucifera
Cu
Chuniophoenix nana
Tahina spectabilis
Ct
Caryota mitis
Arenga caudata
Wallichia densiflora
Cp
Corypha lecomtei
Bismarckia nobilis
Coryphoideae
Bo
Borassodendron sp.
Lodoicea maldivica
Sa
Sabal domingensis
Leucothrinax morrisii
Cr
Trithrinax brasiliensis
Ph
Phoenix dactylifera
Colpothrinax cookii
Tr
Pritchardia thurstonii
Washingtonia robusta
Acoelorrhaphe wrightii
Serenoa repens
Licuala paludosa
Brahea brandegeei
Chamaerops humilis
II.
Typha latifolia
Baxteria australis
Dasypogonales
Kingia australis
Calectasia narragara
Dasypogon bromeliifolius
Eu
Eugeissona tristis
Calamoideae
Lp
Eremospatha macrocar
Mauritia flexuosa
Arecales
Salacca ramosiana
Metroxylon warbergii
Ca
Calamus caryotoides
Pigafetta elata
Nypoideae
Nypa fruticans
Ravenea hildebrandti
Ceroxyloideae
Phytelephas aequator
Pseudophoenix vinifera
Chamaedorea seifrizii
Areca vestiaria
Veitchia arecina
Arecoideae
Elaeis guineensis
Cocos nucifera
Cu
Chuniophoenix nana
Tahina spectabilis
Ct
Caryota mitis
Arenga caudata
Wallichia densiflora
Cp
Corypha lecomtei
Bismarckia nobilis
Coryphoideae
Bo
Borassodendron sp.
Lodoicea maldivica
Sa
Sabal domingensis
Leucothrinax morrisii
Cr
Trithrinax brasiliensis
Ph
Phoenix dactylifera
Pritchardia thurston
Tr
Colpothrinax cookii
Washingtonia robusta
Licuala paludosa
Acoelorrhaphe wrightii
Serenoa repens
Brahea brandegeei
Chamaerops humilis

Figure 4