Phenotypic characters of static homology increase phylogenetic stability under direct optimization of otherwise dynamic homology characters

Samuli Lehtonen*

Biodiversity Unit, University of Turku, FI-20014 Turku, Finland

*Corresponding author:
E-mail address: samile@utu.fi

Abstract

Direct optimization of unaligned sequence characters provides a natural framework to explore the sensitivity of phylogenetic hypotheses to variation in analytical parameters. Phenotypic data, when combined into such analyses, are typically analyzed with static homology correspondences unlike the dynamic homology sequence data. Static homology characters may be expected to constrain the direct optimization and thus, potentially increase the similarity of phylogenetic hypotheses under different cost sets. However, whether a total-evidence approach increases the phylogenetic stability or not remains empirically largely unexplored. Here, I studied the impact of static homology data on sensitivity using six empirical data sets composed of several molecular markers and phenotypic data. The inclusion of static homology phenotypic data increased the average stability of phylogenetic hypothesis in five out of the six data sets. To investigate if any static homology characters would have similar effect, the analyses were repeated with randomized phenotypic data, and with one of the molecular markers fixed as static homology characters. These analyses had, on average, almost no effect on the phylogenetic stability, although the randomized phenotypic data sometimes resulted in even higher stability than empirical phenotypic data. The impact was related to the strength of the phylogenetic signal in the phenotypic data: higher average jackknife support of the phenotypic tree correlated with stronger stabilizing effect in the total-evidence analysis. Phenotypic data with a strong signal made the total-evidence trees topologically more similar to the phenotypic trees, thus, they constrained the dynamic homology correspondences of the sequence data. Characters that increase phylogenetic stability are particularly valuable for phylogenetic inference. These results indicate an important role and additive value of phenotypic data in increasing the stability of phylogenetic hypotheses in total-evidence analyses.

Introduction

Maximization of homology provides a rationale for parsimony-based phylogenetic inference (De Laet, 2005, 2015). Hypotheses of homologous character correspondences are logically tree dependent, and the optimization of these correspondences on a phylogenetic tree is known as the Tree Alignment Problem (Sankoff, 1975; Wheeler, 1996; Varón and Wheeler, 2012). The most widely used heuristic solution to this problem is direct optimization (DO; Wheeler, 1996).

Tree topologies typically vary when sensitivity analysis (sensu Wheeler, 1995) is conducted under DO, with a general understanding that less variable topological configurations are desired outcome of such exercise (Sharma et al., 2010; but for a contrasting view, see Grant and Kluge, 2005). Hence, characters that increase stability, i.e. make the data to converge towards a single optimal resolution under variable analytical conditions, can be seen particularly valuable for the phylogenetic inference. Sensitivity may result, for example, from a lack of clear signal to resolve a particular node in a tree, thus potentially resulting in alternative topological resolutions under different analytical cost sets. A character providing information for the resolution of such a node

may guide the optimization towards a common resolution under a variety of cost sets and this way improve the stability of the phylogenetic hypothesis. Such a behavior might be expected from characters evolving at a rate different from the other characters involved. Morphological (or more broadly, phenotypic) data undoubtedly evolve in a manner different from molecular sequence data, and therefore could help DO to converge into a more stable hypothesis (Titus and Frost, 1996). Indeed, Lehtonen et al. (2016) reported increased stability of phylogenetic hypotheses in two different data sets after static homology phenotypic data were incorporated into a sensitivity analysis of otherwise dynamic homology data. They noted strongest stabilizing effect when phenotypic data were given higher weight, clearly indicating that these data were influential, but were uncertain whether this was because the phenotypic data were constraining DO towards a phylogenetically more reasonable result, or if any static homology characters, supporting a reasonable tree or not, would impose similar constraints for DO.

Total-evidence, or simultaneous analysis of phenotypic and genotypic data (Kluge, 1989), is a generally advocated approach by the proponents of DO (e.g. Giribet et al., 2001; Wheeler et al., 2006a). Although some authors have applied DO to optimize phenotypic character homologies (Robillard et al., 2006; Ramírez, 2007; Agolin and D'Haese, 2009; Japyassú and Machado, 2010), the general approach is to co-optimize static homology phenotypic data with dynamic homology sequence data (Wheeler et al., 2006a). Under this kind of analysis, the homology correspondences of molecular data are affected by the phenotypic data, whereas the sequence data have no effect on the homology correspondences in the phenotypic data. How the static homology characters affect the optimization of dynamic homologies, and thus, the outcome of sensitivity analysis, remains largely unexplored. However, Aagesen (2005) noted that adding the length invariant *rbcL* partition into an analysis of Triticeae did not generally improve stability – this analysis did not fix the homology correspondences in the length invariant partition, but nevertheless indicated that invariable homology scheme does not automatically translate into increased stability. The same study found that stability increases, in general, when data partitions are added (Aagesen, 2005). These findings would suggest that static homology characters may not increase stability any more than dynamic homology characters. This is not necessarily surprising, given that the topological variation in a sensitivity analysis is only partially derived from altered homology correspondences and to an unknown degree from altered transformation costs (e.g. relative costs of transitions versus transversions) within unaltered homology scheme. Because of the varying transformation costs, topological variation is expected in the sensitivity analysis even if the underlying homology scheme remains the same (i.e. static). In this respect, however, the static phenotypic data behaves quite differently, because the transformation costs within the phenotypic data set typically are not altered in a sensitivity analysis, even if the weight of the phenotypic data set as a whole may be varied.

In this study my main goal is to evaluate the impact of phenotypic data analyzed as static homology characters on phylogenetic sensitivity when combined with dynamic homology sequence data. I explore whether fixed sequence alignments or random static homology characters impact the analyses in the same manner as empirical phenotypic data and discuss the possible benefits and pitfalls of adding static phenotypic data into DO sensitivity analysis.

Materials and methods

Data sets

Six published empirical data sets combining several molecular loci with phenotypic data were extracted from the literature. These data sets represented both plants and animals and varied by their phylogenetic depth. Each original data set was standardized to 19 terminal taxa, including one

outgroup taxon, and to four or five molecular loci by subsampling the original matrices. Terminal taxa were selected semi-randomly, so that they covered broadly the original phylogenetic space and were not missing any molecular locus selected for this study. However, not all sequences were complete (the sequence lengths below are reported for the complete sequences only). The phenotypic data were coded and treated as originally, i.e. the original transformation costs and ordering were retained. Taxa and loci used with their GenBank codes are listed in Appendix 1, and the data sets analysed with the resulting trees are available as supplementary Data S1.

Alismataceae. The phenotypic Alismataceae data were published by Lehtonen (2009) and consist of ten continuous characters coded as ranges and analyzed as such (Goloboff et al., 2006), and 67 characters coded as having discrete states. Four molecular loci, the nuclear *ITS* (656–767 bp), and plastid *matK* (1264–1274 bp), *psbA* (759–760 bp), and *rbcL* (1091 bp) were taken from the data set compiled in Lehtonen (2017).

Clematis. The *Clematis* (Ranunculaceae) data set was published by Lehtonen et al. (2016). The phenotypic data consisted of nine continuous characters analyzed as such and 31 discrete characters. The molecular data consisted of five loci: the nuclear *ITS* (567–613 bp), and plastid *atpB-rbcL* (700–757 bp), *matK* (2502–2520 bp), *psbA-trnH-trnQ* (480–573 bp), and *rbcL-accD* (1609–1643 bp). The *ITS*, *matK*, and *rbcL-accD* were further split into three, two, and three partitions, respectively, following the original analyses.

Euphasiacea. The Euphasiacea (Crustacea) data set was published by Vereshchaka et al. (2018). The phenotypic data consisted of 168 discrete characters and the molecular data of four loci: *COI* (658 bp), *16S* (511–517 bp), *18S* (1809–1861 bp), and *H3* (305 bp).

Lindsaeaceae. The Lindsaeaceae data set was published by Lehtonen et al. (2010). The phenotypic data consisted of 55 discrete characters and the molecular data of five loci: *trnL-trnF* (429–496 bp), *trnH-psbA* (447–480 bp), *rpoC1* (732 bp), *rps4* (560–568 bp), and *rps4-trnS* (215–266 bp).

Protodrilidae. The Protodrilidae (Annelida) data were published by Martínez et al. (2014). The phenotypic data consisted of 55 discrete characters and the molecular data of four loci: *28S* (992–1072 bp), *16S* (327–362 bp), *18S* (1720–1753 bp), and *H3* (329–330 bp).

Riama. The *Riama* (Squamata: Gymnophthalmidae) data set was published by Sánchez-Pacheco et al. (2017). The phenotypic data consisted of 35 discrete characters and the molecular data of four loci: *C-mos* (362–374 bp), *12S* (340–368 bp), *16S* (453–472 bp), and *ND4* (621 bp).

The possible impact of any kind of static homology characters on the phylogenetic sensitivity was explored by re-running the analyses ten times with randomized instead of empirical phenotypic data. The random phenotypic character matrices were created for each of the six data sets by randomly re-assigning the observed character states to terminal taxa, independently within each character. For the discrete characters this was done with the *reshuffle states within characters* command in Mesquite (Maddison and Maddison, 2018). In addition, the DNA-only analyses were repeated by fixing the homology scheme of the most conservative locus of each data set aligned with MAFFT v.7.215 (Katoh and Standley, 2013), and re-running the sensitivity analyses with the fixed locus as static homology characters. In these analyses the same substitution costs were applied for the static and dynamic homology characters. Gap extension costs cannot be applied to static homology characters, but only in the case of *Clematis* data (*matK*) did the fixed alignment have gaps spanning several bases.

Tree searches

All the phylogenies were inferred in POY 5.1.1 (Wheeler et al., 2015) with parsimony as the optimality criterion. The search strategy was based on TBR swapping of 50 Wagner trees using commands "build(50)" and "swap()"; this simple search strategy is here considered sufficient given that the trees have only 19 terminals and therefore do not exhibit composite optima (Goloboff, 1999). The strict consensus topologies were used for downstream analyses. For each data set the phenotypic data, as well as each molecular marker, were first analyzed separately, then all the molecular markers combined, and finally a total-evidence analysis was completed. For the phenotypic data sets a jackknife analysis was run with 100 pseudoreplicates.

Sensitivity analyses

Sensitivity analyses under DO have generally varied the transversion-transition and gap-substitution costs, but sometimes also the gap opening-extension cost (Aagesen, 2005; Aagesen et al., 2005). Theoretically, this parameter space is infinite, but it can be narrowed down for more practical solutions. The triangle inequality constrains transformation costs to be symmetrical, transversion-transition cost to a minimum of 0.5, and the cost of a gap to at least one half the cost of substitutions (Wheeler, 1993). The upper limits for these costs are not so obvious, but it has been suggested that gap costs should not exceed substitution costs by more than five times (Spagna and Álvarez-Padilla, 2008). Even more complicated is the case of affine gap costs. Strings of gaps can be most parsimoniously explained as single events, and this can be emulated in POY by defining a low gap extension cost (Aagesen, 2005; Aagesen et al., 2005; De Laet, 2005). A gap extension cost set too low may violate the triangle inequality and result in uninformative alignments, however, whether an affine gap cost situation is non-metric depends on the gap length (Aagesen et al., 2005). In any case, affine gap costs sound biologically reasonable and seem to improve congruence (Petersen et al., 2004; Aagesen, 2005; Aagesen et al., 2005; Pons and Vogler, 2006).

Given these limits for the parameter space the following transformation cost sets were applied. The transversion-transition ratio was set to 1 or 2. The unit gap cost (sum of opening and extension cost) was set to 1 or 2 times the highest substitution cost. The gap extension cost was set to 1 or 0.25 the cost of a unit gap. In total, this resulted in eight cost sets that were explored in sensitivity analyses (Table 1). In the total-evidence analyses, the phenotypic data were given the weight of the highest transformation cost. The results of the sensitivity analyses were examined with Cladescan (Sanders, 2010).

Sensitivity analyses were run for combined analyses of the molecular loci as dynamic homology characters (DNA-only), for total-evidence analyses of static homology phenotypic data combined with dynamic homology molecular loci, for combined analyses of the molecular loci with the most conservative locus treated *prealigned*, and for total-evidence analyses with randomized phenotypic data. Furthermore, in order to estimate how much of the topological variation is due to different homology correspondences and how much of it rather represents alternative transformation costs within the same homology scheme, the POY implied alignments resulting from the sensitivity analyses of the combined molecular data were analyzed as static homology characters (*prealigned*) under equal weights. The topological variation observed among the trees from the latter analyses thus purely reflects the variation in the underlying homology scheme.

Congruence measures

Sensitivity analysis can be used to measure the sensitivity of any hypothesis and is therefore independent of hypothesis selection (Giribet, 2003). However, typically one cost set in the sensitivity analysis is selected as the preferred one (or simply as a reference) on some basis. The choice can be done *a priori*, in which case it must be based on some philosophical judgement, or the preferred cost set can be selected *a posteriori*. In the latter case, the choice can be based on congruence (Giribet and Wheeler, 1999; Giribet, 2003; Wheeler et al., 2006a; Giribet and Wheeler, 2007; Spagna and Álvarez-Padilla, 2008; Sharma et al., 2010). Congruence, on the other hand, can be measured in phylogenetics either as character congruence, or topological congruence (Mickevich and Farris, 1981). Most authors using DO have relied on *a priori* choice of the preferred cost set (e.g. Frost et al., 2001), or have based their choice on character congruence (e.g. Giribet et al., 2001; Sharma et al., 2010).

The Meta-Retention Index (MRI) was proposed as a character congruence measure suitable for selecting the optimal cost set in a sensitivity analysis (Wheeler et al., 2006b).

$$\text{MRI} = (\textstyle\sum \text{Maximum cost of Fragments} - \text{cost combined Fragments}) / (\textstyle\sum \text{Maximum cost of Fragments} - \textstyle\sum \text{Minimum cost of Fragments})$$

The maximum and minimum costs of each fragment were calculated following Wheeler et al. (2006b); the fragments were optimized on unresolved bush and the sum of minimum costs was taken as the sum of maximum cost of fragments, whereas the sum of minimum cost of fragments was simply summed from the best costs found in separate analyses of the genes.

Alternatively, phylogenetic congruence can be evaluated topologically by comparing the trees obtained under different cost regimes. Nodal Stability (NS), as defined by Giribet (2003), is the extent to which alternative cost sets support the topology in question. Here, NS is summed over all the nodes for the trees obtained under each cost set and, hence, the topological congruence of each cost set can be directly compared. This approach has sometimes been used to select the reference cost set *a posteriori* (Lehtonen et al., 2013; 2016). Here, the topology most supported by alternative cost sets is referred to as Maximum Stability Tree (MST) and the NS value of a tree is simply the number of investigated cost sets supporting a node summed over all the nodes. Hence, in the present case of 19 taxa, the fully resolved topology has 18 nodes (including the root node) and eight cost sets were evaluated. The maximum stability is achieved when every cost set support the same fully resolved topology, in which case each of the 18 nodes is supported by eight cost sets, thus, the maximum NS value is $18 \times 8 = 144$. This measure takes into account both the resolution (tree with only a few resolved nodes receives a low value, even if the resolved nodes are congruent among all the trees) and the congruence (fully resolved but incongruent trees receive a low value). It is important to note that parameter sets resembling each other will likely produce more congruent topologies than more distant parameter sets. Therefore, MST will likely be located in the parameter space most densely sampled irrespective of whether this represents a biologically reasonable part of the parameter space. For this reason, it is important to explore the parameter space evenly.

Furthermore, topological distances between the obtained trees were heuristically estimated as SPR distances (Goloboff, 2008) in TNT version 1.5 (Goloboff and Catalano, 2016). This aimed to measure the topological variation among gene trees and their distance to the phenotypic tree, and to find out if phenotypic data guides total-evidence analysis towards the phenotypic tree topology.

Statistical tests

The possible correlation between different congruence measures was investigated with Pearson correlation test. The impact of empirical phenotypic data on stability was compared to that of randomized phenotypic data using Mann-Whitney U test. U test was also used to test if SPR distances from gene trees to phenotypic trees were different from SPR distances among the gene trees. The relative roles of the signal strength in phenotypic data (jackknife support) and sensitivity of the DNA data (NS) in explaining the impact of phenotypic data on topology under total-evidence (change in SPR distance from the phenotypic tree to DNA tree in comparison to SPR distance from the phenotypic tree to total-evidence tree) were examined with multivariate regression analysis. Statistical tests were performed in Microsoft Excel.

Results

Character-based (MRI) and topology-based (NS) congruence measures were positively correlated (r = 0.375; n = 96; p = 0.000; correlation computed for all the analyses, including both empirical and randomized replicates). As well, MRI was negatively correlated with the average SPR-distances among the gene trees (r = –0.307; n = 48; p = 0.034), indicating that the higher the character-based congruence, the more topologically similar are the gene trees. However, MRI and NS did not always prefer the same cost sets (Fig. 1) and sometimes the preferred cost set of MRI was found to be worse with NS, or vice versa (Table S1). Both of the congruence measures behaved similarly in preferring cost sets with affine gap costs approximately three times as often as cost sets without affine gap costs.

The Mann-Whitney U test indicated that phenotypic trees were topologically more distant to gene trees (Mdn = 6) than gene trees were from each other (Mdn = 5, U = 27415, p = 0.000). This was not always the case; in Protodrilidae and Euphasiacea data sets gene trees were on average topologically more distant from each other than they were from the phenotypic tree (Fig 2). Strength of the signal in phenotypic data varied from a data set to another, with average jackknife support varying from 23 to 100 and the number of equally parsimonious trees from 1 to 26. The randomized phenotypic data sets did not yield a much higher number of equally parsimonious trees than empirical data sets, but had a much lower average jackknife support (Table 2).

The impact of static homology data on sensitivity was measured by computing the NS value for each cost set, thus, the results are not computed for the preferred trees only, but are general over all the cost sets explored. Stability varied greatly among the studied data sets (Fig. 3; Fig. S1). The least stable was the *Clematis* data, with an average NS value of 83 computed over all the studied parameter sets in the DNA-only analysis. The highest average NS value in DNA-only analyses was 131 in the *Riama* data set. On average, the phenotypic data increased the stability in all the data sets (average impact +8) except in Protodrilidae, in which the NS was reduced from an average of 123 to 122 (Fig. 3). A multivariate regression was computed to explore whether the signal strength in the phenotypic data (average jackknife support) or the sensitivity of the molecular data (NS) is a more important factor in explaining the observed increased topological similarity of the total-evidence trees with the phenotypic trees. A significant regression (F(2, 45) = 6.898, p = 0.002) had an R^2 of 0.235. Average jackknife support of the phenotypic tree significantly predicted (p = 0.000) the higher topological similarity of total-evidence trees to phenotypic trees, whereas sensitivity of the DNA trees (NS) had no significant effect (p = 0.317) (Fig. 4).

On average, the randomized phenotypic data increased NS by 0.5 (Fig. 3). However, of the random phenotypic data analyses, 15% increased the stability even more than the empirical data, whereas 22% of the random data analyses increased sensitivity compared to DNA-only analyses. The Mann-

Whitney *U* test indicated that empirical phenotypic data increased stability significantly more (*Mdn* = 6.5) than the randomized data (*Mdn* = 0.7, *U* = 1402.5, *p* = 0.002).

Results of the sensitivity analyses in DNA-only analyses with the most conservative locus aligned and analyzed as static homology data did not differ at all from the sensitivity analyses where the homologies of all the sequence data were dynamically optimized.

The instability related to different homology correspondences explained 100% of the instability observed in the DNA-only analysis of Alismataceae, 89% of the observed instability in *Clematis*, 80% of the observed instability in Euphasiacea, 88% of the observed instability in Lindsaeaceae, 81% of the observed instability in Protodrilidae, and 46% of the observed instability in *Riama*. Only in two data sets and under a single cost set in each of these data sets did DO trees have a higher stability than corresponding trees inferred from the implied alignments, indicating that the different change costs made alternative homology schemes to converge onto topologies more similar to each other than in the case of equal cost static homologies.

Discussion

It has been argued that phenotypic data are not distinct from genotypic data and should be analyzed together (Wheeler et al., 2006a). In practice, however, phenotypic data typically are treated as static homology characters even when analyzed in an otherwise dynamic homology context. The different underlying homology concepts make these data behave differently in DO total-evidence analysis; the homology correspondences and the phylogenetic signal of sequence characters will depend on the transformation costs applied, whereas the static homology data will provide the same signal irrespective of the cost sets applied for sequence characters. Therefore, it is not surprising that such phenotypic data make total-evidence topologies inferred under different cost sets more similar to each other, and more similar to the phenotypic tree, as observed in this study.

A similar stabilizing effect was not observed when one of the 4–5 molecular loci was aligned prior to analysis and treated as static homology data in an otherwise dynamic homology analysis. This is likely because the most conservative locus, which in most cases was length invariant, was selected to be prealigned. Apparently, these sequences were optimized to have the same homology correspondences also when homologies were dynamic in sensitivity analyses. The effect would probably have been different if some of the length variable characters had been prealigned instead. This behavior can also be partly explained by the different transformation costs applied in the sensitivity analyses: even if the homology scheme of these selected loci remained static, the substitution costs were nevertheless varied the same way as in dynamic homology characters.

The complex interplay of variable homology correspondences on one hand, and the different transformation costs even when the homology correspondences remain the same, have remained unexplored. It may be argued that these cannot logically be separated, as the search for optimal homology correspondences and topology can be seen as a single tree alignment problem (Sankoff, 1975; Wheeler, 1996). Despite this, it may still be of great interest to know how much the observed sensitivity is related to differences in the inferred homology correspondences, indicating 'difficult alignment', or if the results remain sensitive regardless of largely similar homology correspondences. Here, by using implied alignments produced under different cost regimes and analyzing them with equal costs as static homology characters, it was shown that most of the topological variation in sensitivity analyses of these data sets could be linked with alternating homology correspondences, meaning that the 'alignment' indeed is of crucial importance and the

increased stability in total-evidence analyses is probably due to alignment constraints imposed by the phenotypic data.

Randomized phenotypic data had little impact in the analyses, although in general they slightly improved the stability. In some cases, however, random data had even a stronger stabilizing effect than empirical data. This reveals that any static homology data, phylogenetically relevant or not, may constrain DO and result in false stability. Of course, the same problem applies to any data, as the old phrase – garbage in, garbage out – points out. DO is particularly prone to pick up any signal and amplify it, especially in the lack of strong and clear signal in the sequence characters (Simmons et al., 2010; Yoshizawa, 2010) and it has been suggested that congruence-based criteria will pick up any alignment scenario that happens to support the same topology as morphology or other static homology data (Simmons, 2004). In the present data sets randomized data behaved "better" than empirical data in some random replicates of Lindsaeaceae, Protodrilidae, and *Riama* data. In these cases, the empirical phenotypic tree had a low average jackknife support and generally many equally parsimonious solutions indicating absense of clear phenotypic signal. Furthermore, the Protodrilidae gene trees were topologically highly distinct, suggesting a lack of clear signal also in the molecular data, and the phenotypic *Riama* data supported a topology highly different from the gene trees. Thus, in all of these cases the empirical data (phenotypic or sometimes also genotypic) appears somehow problematic and seems to behave to some degree like random data.

Obviously, this does not imply that most phenotypic data are useless, or that genotypic data are automatically more trustworthy, as some authors have suggested (e.g. Scotland et al., 2003). Although the gene trees and phenotypic trees differed by a median distance of 6 SPR moves in contrast to median SPR distance of 5 moves among the gene trees, this difference, even if statistically significant, indicates rather that both types of data generally reflected the same underlying phylogenetic history. After all, two random topologies of 19 taxa with a fixed rooting differ on average by 11.7 SPR moves (1000 random tree pairs compared in TNT). A previous study focusing on the congruence between morphological and molecular trees similarly concluded that molecular and morphological trees are much more similar than expected by a chance, but congruence appears to be somewhat higher within the molecular trees (Pisani et al., 2007). These results support the idea that phenotypic data improve the phylogenetic stability by providing extrinsic and hopefully reasonable constrains for the dynamic homology correspondences (Titus and Frost, 1996). Although some authors might argue that this behavior illuminates the problems of DO rather than benefits of total-evidence DO analysis (Simmons et al., 2010; Yoshizawa, 2010), it should be noted that phylogenetic sensitivity to input parameters is by no means restricted to DO approach and phenotypic data can be expected to have similar impact regardless of the analytical method applied (see Mongiardino Koch and Gauthier, 2018). Particularly interesting avenues would be simultaneous dynamic optimization of all character data (for a stimulating discussion on phenotypic data optimization see Vogt, 2018), and the development of more biologically realistic optimizations. De Laet (2005, 2015) has demonstrated that maximization of sequence character homologies under the equal weighting scheme for length variable gap characters can be approximated in DO by using his weighting scheme (commonly referred to as 3221, but note that this corresponds with a cost set of 2221 in POY versions 4 and 5). However, it has been demonstrated that different cost sets should likely be applied to each data fragment in DO (Sharma et al., 2010). How the sequence data should be appropriately fragmented and what actually should be optimized within each fragment (consider for example the secondary structures, Kjer et al., 2007) certainly deserve further consideration. Nevertheless, it should be clear that optimality of structural alignments (or whatever type of alignment) vary just like in any kind of alignments, the preference for suboptimal alignment can hardly be justified (whatever alignment criteria is applied),

and that homologies are by definition tree-dependent (i.e. optimality of an alignment depends on the topology and therefore alignment and tree search are logically connected).

In this study I did not consider nodal support (sensu Giribet, 2003) beyond investigating the strength of phylogenetic signal within the phenotypic data sets. Resampling techniques aim to test the support of the secondary homology scheme (sensu de Pinna, 1991) through pseudoreplicating the character congruence test of primary homologies. Since primary homologies do not exist at the nucleotide level under dynamic homology, resampling at this level is meaningless also (Wheeler et al., 2006a). The rather commonly used approach to resample POY implied alignments by jackknifing should therefore never be conducted, as it logically results in heavily inflated support values (Simmons et al., 2010). This problem can be avoided by properly resampling characters at the primary homology level, that is, at the fragment level with dynamic resampling methods in DO, but this would have required more fragments than used in the present study (see Simmons et al., 2010). As well, whether character-based or topology-based congruence measures should be used to select the reference cost set, and which cost sets should be investigated in the first place, are questions largely beyond the scope of this paper. Unlike in some previous studies (e.g. Pons and Vogler, 2006), a significant correlation between character-based and topology-based congruence measures was found here, but the correlation was not particularly strong. Both approaches, however, more often preferred affine gap costs providing further evidence that these scenarios should be explored under DO sensitivity framework (Petersen et al., 2004; Aagesen, 2005; Aagesen et al., 2005; Pons and Vogler, 2006).

Acknowledgements

References

Aagesen, L., 2005. Direct optimization, affine gap costs, and node stability. Mol. Phylogenet. Evol. 36, 641–653.

Aagesen, L., Petersen, G. and Seberg, O., 2005. Sequence length variation, indel costs, and congruence in sensitivity analysis. Cladistics 21, 15–30.

Agolin, M. and D'Haese, C.A., 2009. An application of dynamic homology to morphological characters: direct optimization of setae sequences and phylogeny of the family Odontellidae (Poduromorpha, Collembola). Cladistics 25, 353–385.

De Laet, J., 2005. Parsimony and the problem of inapplicables in sequence data, in: Albert, V.A. (Ed.), Parsimony, Phylogeny and Genomics. pp. 81–116.

De Laet, J., 2015. Parsimony analysis of unaligned sequence data: maximization of homology and minimization of homoplasy, not minimization of operationally defined total cost or minimization of equally weighted transformations. Cladistics 31, 550–567.

de Pinna, M.C.C., 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7, 367–394.

Frost, D.R., Rodrigues, M.T., Grant, T. and Titus, T.A., 2001. Phylogenetics of the lizard genus *Tropidurus* (Squamata: Tropiduridae: Tropidurinae): direct optimization, descriptive efficiency, and sensitivity analysis of congruence between molecular data and morphology. Mol. Phylogenet. Evol. 21, 352–371.

Giribet, G., 2003. Stability in phylogenetic formulations and its relationship to nodal support. Syst. Biol. 52, 554–564.

Giribet, G., Edgecombe, G.D. and Wheeler, W.C., 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413, 157–161.

Giribet, G. and Wheeler, W.C., 2007. The case for sensitivity: a response to Grant and Kluge. Cladistics 23, 294–296.

Giribet, G. and Wheeler, W.C., 1999. On gaps. Mol. Phylogenet. Evol. 13, 132–143.

Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics 15, 415–428.

Goloboff, P.A., 2008. Calculating SPR distances between trees. Cladistics 24, 591–597.

Goloboff, P.A. and Catalano, S.A., 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. Cladistics 32, 221–238.

Goloboff, P.A., Mattoni, C.I. and Quinteros, A.S., 2006. Continuous characters analyzed as such. Cladistics 22, 589–601.

Grant, T. and Kluge, A.G., 2005. Stability, sensitivity, science and heurism. Cladistics. 21, 597–604.

Japyassú, H.F. and Machado, F.A., 2010. Coding behavioural data for cladistic analysis: using dynamic homology without parsimony. Cladistics 26, 625–642.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kjer, K.M., Gillespie, J.J. and Ober, K.A., 2007. Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. Syst. Biol. 56, 133–146.

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

Lehtonen, S., 2017. Splitting *Caldesia* in favour of *Albidella* (Alismataceae). Aust. Syst. Bot. 30, 64–69.

Lehtonen, S., 2009. Systematics of the Alismataceae—A morphological evaluation. Aquat. Bot. 91, 279–290.

Lehtonen, S., Christenhusz, M.J.M. and Falck, D., 2016. Sensitive phylogenetics of *Clematis* and its position in Ranunculaceae. Bot. J. Linn. Soc. 182, 825–867.

Lehtonen, S., Tuomisto, H., Rouhan, G. and Christenhusz, M.J.M., 2013. Taxonomic revision of the fern genus *Osmolindsaea* (Lindsaeaceae). Syst. Bot. 38, 887–900.

Lehtonen, S., Tuomisto, H., Rouhan, G. and Christenhusz, M.J.M., 2010. Phylogenetics and classification of the pantropical fern family Lindsaeaceae. Bot. J. Linn. Soc. 163, 305–359.

Maddison, W. P. and Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. Version 3.51 http://www.mesquiteproject.org

Martínez, A., Di Domenico, M., Rouse, G.W. and Worsaae, K., 2014. Phylogeny and systematics of Protodrilidae (Annelida) inferred with total evidence analyses. Cladistics 31, 250–276.

Mickevich, M.F., Farris, J.S., 1981. The implications of congruence in *Menidia*. Syst. Zool. 30, 351–370.

Mongiardino Koch, N. and Gauthier, J.A., 2018. Noise and biases in genomic data may underlie radically different hypotheses for the position of Iguania within Squamata. PLoS ONE 13, e0202729.

Petersen, G., Seberg, O., Aagesen, L. and Frederiksen, S., 2004. An empirical test of the treatment of indels during optimization alignment based on the phylogeny of the genus *Secale* (Poaceae). Mol. Phylogenet. Evol. 30, 733–742.

Pisani, D., Benton, M.J. and Wilkinson, M., 2007. Congruence of morphological and molecular hylogenies. Acta Biotheor. 55, 269–281.

Pons, J. and Vogler, A.P., 2006. Size, frequency, and phylogenetic signal of multiple-residue indels in sequence alignment of introns. Cladistics 22, 144–156.

Ramírez, M.J., 2007. Homology as a parsimony problem: a dynamic homology approach for morphological data. Cladistics 23, 588–612.

Robillard, T., Legendre, F., Desutter Grandcolas, L. and Grandcolas, P., 2006. Phylogenetic analysis and alignment of behavioral sequences by direct optimization. Cladistics 22, 602–633.

Ross, T.G., Barrett, C.F., Soto Gomez, M., Lam, V.K.Y., Henriquez, C.L., Les, D.H., Davis, J.I., Cuenca, A., Petersen, G., Seberg, O., Thadeo, M., Givnish, T.J., Conran, J., Stevenson, D.W. and Graham, S.W., 2016. Plastid phylogenomics and molecular evolution of Alismatales. Cladistics 32, 160–178.

Sanders, J.G., 2010. Program note: Cladescan, a program for automated phylogenetic sensitivity analysis. Cladistics 26, 114–116.

Sánchez-Pacheco, S.J., Torres-Carvajal, O., Aguirre-Peñafiel, V., Nunes, P.M.S., Verrastro, L., Rivas, G.A., Rodrigues, M.T., Grant, T. and Murphy, R.W., 2017. Phylogeny of *Riama* (Squamata: Gymnophthalmidae), impact of phenotypic evidence on molecular datasets, and the origin of the Sierra Nevada de Santa Marta endemic fauna. Cladistics 34, 260–291.

Sankoff, D., 1975. Minimal mutation trees of sequences. SIAM J. Appl. Math. 28, 35–42.

Scotland, R.W., Olmstead, R.G. and Bennett, J.R., 2003. Phylogeny reconstruction: the role of morphology. Syst. Biol. 52, 539–548.

Sharma, P.P., Vahtera, V., Kawauchi, G.Y. and Giribet, G., 2010. Running WILD: the case for exploring mixed parameter sets in sensitivity analysis. Cladistics 27, 538–549.

Simmons, M.P., 2004. Independence of alignment and tree search. Mol. Phylogenet. Evol. 31, 874–879.

Simmons, M.P., Müller, K.F. and Norton, A.P., 2010. Alignment of, and phylogenetic inference from, random sequences: the susceptibility of alternative alignment methods to creating artifactual resolution and support. Mol. Phylogenet. Evol. 57, 1004–1016.

Spagna, J.C. and Álvarez-Padilla, F., 2008. Finding an upper limit for gap costs in direct optimization parsimony. Cladistics 24, 787–801.

Titus, T.A., Frost, D.R., 1996. Molecular homology assessment and phylogeny in the lizard family Opluridae (Squamata: Iguania). Mol. Phylogenet. Evol. 6, 49–62.

Varón, A., Wheeler, W.C., 2012. The tree alignment problem. BMC Bioinformatics. 13, 293–36.

Vereshchaka, A.L., Kulagin, D.N. and Lunina, A.A., 2018. A phylogenetic study of krill (Crustacea: Euphausiacea) reveals new taxa and co-evolution of morphological characters. Cladistics 35, 150–172.

Vogt, L., 2018. Towards a semantic approach to numerical tree inference in phylogenetics. Cladistics 34, 200–224.

Wheeler, W., 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–9.

Wheeler, W., Aagesen, L., Arango, C.P., Faivovich, J., Grant, T., D'Haese, C., Janies, D., Smith, W.L., Varón, A. and Giribet, G., 2006a. Dynamic Homology and Phylogenetic Systematics: A Unified Approach Using POY. American Museum of Natural History, New York.

Wheeler, W.C., 1993. The triangle inequality and character analysis. Mol. Biol. Evol. 10, 707–714.

Wheeler, W.C., 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst. Biol. 44, 321–331.

Wheeler, W.C., Lucaroni, N., Hong, L., Crowley, L.M. and Varón, A., 2015. POY version 5: phylogenetic analysis using dynamic homologies under multiple optimality criteria. Cladistics 31, 189–196.

Wheeler, W.C., Ramírez, M.J., Aagesen, L. and Schulmelster, S., 2006b. Partition-free congruence analysis: implications for sensitivity analysis. Cladistics 22, 256–263.

Yoshizawa, K., 2010. Direct optimization overly optimizes data. Syst. Entomol. 35, 199–206.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1. Data sets analysed with the resulting trees.
Figure S1. Trees from the analyses of combined molecular data (DNA-only; in left panel), phenotypic data (middle panel), and total-evidence analyses (right panel). Jackknife support values are shown for the phenotypic trees and sensitivity plots for the other trees. Reference topologies for the sensitivity analyses are the ones maximizing nodal stability (Maximum Stability Tree, MST). (a) Alismataceae. (b) *Clematis*. (c) Euphasiaceae. (d) Lindsaeaceae. (e) Protodrilidae. (f) *Riama*.
Table S1. Minimum and maximum tree lengths, calculated Meta-Retention Index (MRS) and Nodal Stability (NS) under the studies cost sets. Separate tables for empirical data analyses and replicates or randomized phenotypic data (rnd1–rnd10).

Appendix 1. Molecular data used in the analyses.

Alismataceae. *Albidella nymphaeifolia*: EF088077, KX980066, KU499837, EF088125; *Alisma canaliculatum*: DQ339081, JF975413, JF781042, AB040179; *Alisma plantago-aquatica*: JF780977, JF975410, L08759, JF781065; *Butomus umbellatus*: JF780965, JF975415, AY149345, AY952416; *Caldesia grandis*: DQ207881, JF975417, JF781043, JF781068; *Caldesia parnassifolia*: JF780984, JF975416, KU499838, EF088140; *Damasonium alisma*: JF780988, JF975419, U80678, JF781070; *Damasonium minus*: JF780987, JF975418, JF781063, JF781069; *Echinodorus berteroi*: EF088087, KX980068, KU499839, EF088134; *Echinodorus cordifolius*: EF088078, KX980067, DQ859164, EF088126; *Helanthium bolivianum*: EF088060, KX980064, KU499841, EF088109; *Hydrocleys nymphoides*: JF780985, JF975425, JF781047, JF781074; *Limnocharis flava*:JF780986, JF975426, JF781048, JF781075; *Limnophyton angolensis*: JF780991, JF975427, JF781049, JF781076; *Luronium natans*: DQ339093, (*psbA* taken from Ross et al., 2016), U80680, HQ456465; *Ranalisma rostrata*: JF780983, JF975431, JF781051, JF781078; *Sagittaria guayanensis*: JF780968, JF975434, JF781054, JF781081; *Sagittaria latifolia*: JF780975, JF975441, L08767, JF781087; *Sagittaria trifolia*: JF780970, JF975436, JF781056, JF781083. **Clematis.** *Anemone flaccida*: AB115462, AB120212, AB117604, AB110530, AB117001; *Clematis alternata*: AB115440, AB120190, AB117582, AB110509, AB116979; *Clematis crassifolia*: AB115444, AB120194, AB117585, AB110513, AB116983; *Clematis eichleri*: AB115459, AB120209, AB117601, AB110527, AB116998; *Clematis florida*: AB115436, AB120186, AB117578, AB110505, AB116975; *Clematis fusca:* AB115429, AB120179, AB117571, AB110535, AB116968; *Clematis gentianoides*: AB115460, AB120210, AB117602, AB110528, AB116999; *Clematis hancockiana*: KR909735, KR909772, KR909882, KR909833, KR909668; *Clematis ispahanica*: KR909718, KR909777, KR909884, KR909837, KR909711; *Clematis lasiandra*: AB115435, AB120185, AB117577, AB110504, AB116974; *Clematis ligusticifolia*: AB115451, AB120201, AB117593, AB110519, AB116990; *Clematis nobilis*: AB115456, AB120206, AB117598, AB110524, AB116995; *Clematis patens*: AB115434, AB120184, AB117576, AB110503, AB116973; *Clematis potaninii*: AB115448, AB120198, AB117590, AB110517, AB116987; *Clematis stans*: AB115438, AB120188, AB117580, AB110507, AB116977; *Clematis tangutica*: AB115445, AB120195, AB117587, AB110514, AB116984; *Clematis villosa*: AB115461, AB120211, AB117603, AB110529, AB117000; *Clematis vitalba*: AB115457, AB120207, AB117599, AB110525, AB116996; *Pulsatilla cernua*: GU732566, GU732647, GU732728, AB110531, AB117002. **Euphasiaceae.** *Acanthephyra purpurea*: KP076173, KP075887, KP075819, KP076127; *Euphausia krohnii*: MG669391, MG677869, MG677832, MG699136; *Euphausia pacifica*: MG669392, MG677870, MG677833, MG699137; *Euphausia pseudogibba*: MG669393, MG677871, MG677834, MG699138; *Euphausia recurva*: MG669394, MG677872, MG677835, MG699139; *Euphausia tenera*: MG669395, MG677873, MG677836,

MG699140; *Meganyctiphanes norvegica*: MG669397, MG677875, MG677838, MG699142; *Nematobrachion boopis*: MG669398, MG677876, MG677839, MG699143; *Nematobrachion flexipes*: MG669399, MG677877, MG677840, MG699144; *Nematoscelis atlantica*: MG669401, MG677879, MG677842, MG699146; *Nematoscelis megalops*: MG669402, MG677880, MG677843, MG699147; *Nematoscelis tenella*: MG669403, MG677881, MG677844, MG699148; *Stylocheiron maximum*: MG669408, MG677886, MG677849, MG699153; *Thysanoessa gregaria*: MG669410, MG677888, MG677851, MG699155; *Thysanoessa inermis*: MG669411, MG677889, MG677852, MG699156; *Thysanoessa longicaudata*: MG669412, MG677890, MG677853, MG699157; *Thysanopoda aequalis*: MG669415, MG677893, MG677856, MG699160; *Thysanopoda cornuta*: MG669416, MG677894, MG677857, MG699161; *Thysanopoda cristata*: MG669417, MG677895, MG677858, MG699162. **Lindsaeaceae.** *Lindsaea blotiana*: GU478813, GU478510, GU478633, GU478679, GU478381; *Lindsaea botrychioides*: FJ360994, FJ360904, FJ360949, GU478714, GU478416; *Lindsaea lancea*: FJ361006, FJ360915, FJ360961, GU478651, GU478353; *Lindsaea lapeyrousei*: FJ361007, FJ360916, FJ360962, GU478680, GU478382; *Lindsaea linearis*: FJ361008, FJ360917, FJ360963, GU478721, GU478423; *Lindsaea lobata*: FJ361009, FJ360918, FJ360964, GU478674, GU478376; *Lindsaea multisora*: GU478849, GU478534, GU478631, GU478676, GU478378; *Lindsaea pacifica*: FJ361015, FJ360924, FJ360970, GU478663, GU478365; *Lindsaea pallida*: GU478778, GU478479, GU478624, GU478705, GU478407; *Lindsaea pectinata*: FJ361021, FJ360930, FJ360976, GU478686, GU478388; *Lindsaea plicata*: GU478768, GU478471, GU478593, GU478661, GU478363; *Lindsaea propinqua*: GU478825, GU478537, GU478601, GU478662, GU478364; *Lindsaea rigidiuscula*: GU478794, GU478476, GU478621, GU478691, GU478393; *Lindsaea tenuifolia*: FJ361030, FJ360939, FJ360985, GU478677, GU478379; *Nesolindsaea kirkii*: KC155881, KC155835, HQ157323, HQ157327, KC155826; *Odontosoria chinensis*: MG561415, MG561405, HQ157313, HQ157328, MG561412; *Odontosoria flexuosa*: GU478742, GU478464, GU478583, GU478648, GU478350; *Osmolindsaea odorata*: GU478760, GU478431, GU478589, GU478646, GU478348; *Tapeinidium luzonicum*: GU478751, GU478444, GU478576, GU478643, GU478345. **Protodrilidae.** *Claudrilus corderoi*: KJ451189, KJ451231, KJ451272, KJ451387; *Claudrilus draco*: JX402098, KJ451237, KJ451278, KJ451361; *Claudrilus hypoleucus*: KJ451196, KJ451240, KJ451282, KJ451365; *Claudrilus n. sp. 1*: KJ451184, KJ451226, KJ451267, KJ451354; *Claudrilus n. sp. 2*: KJ451186, KJ451228, KJ451269, KJ451356; *Claudrilus ovarium*: KJ451188, KJ451230, KJ451271, KJ451386; *Claudrilus tenuis*: KJ451205, KJ451248, KJ451291, KJ451372; *Megadrilus hochbergi*: JX402096, KJ451238, KJ451279, KJ451362; *Megadrilus n. sp. 1*: KJ451208, KJ451251, KJ451294, KJ451375; *Megadrilus purpureus*: AY527057, EU418874, AY340474, DQ779760; *Megadrilus schneideri*: KJ451190, KJ451232, KJ451273, KJ451358; *Meiodrilus adhaerens*: KJ451197, KJ451241, KJ451283, KJ451366; *Meiodrilus gracilis*: KJ451194, KJ451239, KJ451280, KJ451363; *Protodrilus affinis*: KJ451207, KJ451250, KJ451293, KJ451374; *Protodrilus ciliatus*: KF954464, KF954420, KF954442, KF954505; *Protodrilus oculifer*: KJ451203, KJ451246, KJ451289, KJ451370; *Protodrilus pythonius*: KJ451212, KJ451255, KJ451298, KJ451379; *Protodrilus smithsoni*: JX402097, KJ451236, KJ451277, KJ451347; *Saccocirrus pussicus*: KF954481, KF954439, KF954460, KF954500. **_Riama._** *Ameivula ocellifera*: AF420862, AF420706, AF420759, AF420914; *Anadia rhombifera*: KU902052, KU902133, KU902214, KU902289; *Andinosauria aurea*: KY670647, KY670682, KY681100, KY710831; *Andinosauria crypta*: KY670649, KY670684, KY681102, KY710833; *Andinosauria vespertina*: KY670657, KY670693, KY681111, KY710839; *Bachia flavescens*: AF420859, AF420705, AF420753, AF420869; *Cercosaura argula*: AF420838, AF420698, AF420751, AF420896; *Kentropyx calcarata*: AF420864, AF420707, AF420760, AF420913; *Petracola ventrimaculata*: AY507910, AY507863, AY507883, AY507894; *Pholidobolus macbrydei*: AY507896, AY507848, AY507867, AY507886; *Placosoma cordylinum*: AF420823, AF420673, AF420734, AF420879; *Potamites ecpleopus*: AF420829, AF420656, AF420748, AF420890; *Proctoporus lacertus*:

AY507897, AY507850, AY507868, AY225180; *Ptychoglossus brevifrontalis*: AY507911, AY507865, AY507884, AY507895; *Rhachisaurus brachylepis*: AF420853, AF420665, AF420737, AF420877; *Riama anatoloros*: KY670664, KY670702, KY681120, KY710843; *Riama orcesi*: KY670675, KY670713, KY681131, KY710851; *Riama striata*: KY670678, KY670718, KY681136, KY710855; *Riama unicolor*: AY507907, AY507862, AY507880, AY507893.

Table 1. Parameter cost sets explored in this study (gap opening : transversion : transition : gap extension).

0:1:1:1	3:4:4:1	3:2:2:1	0:1:1:2
0:2:1:2	3:4:2:1	3:2:1:1	0:2:1:4

Table 2. Key statistics of the phenotypic trees. The impact of phenotypic data on stability is reported as an average over the explored eight cost sets. For randomized data sets (rnd data) minimum and maximum values are given in the parenthesis after the averages of ten randomizations.

| | Minimum tree length | | Average jackknife support | | Number of MPTs | | Impact on stability | |
	empirical	rnd data	empirical	rnd data	empirical	rnd data	empirical	rnd data
Alismataceae	169.564	237.922 (233.071–240.930)	97	46 (21–66)	1	1 (1–1)	+12	+4 (-2…+12)
Clematis	91.67	99.67 (97.35–102.66)	70	46 (5–66)	1	1 (1–3)	+9	+4 (-1…+9)
Euphasiaceae	180	288 (284–294)	100	0 (0–0)	1	3 (1–5)	+21	-6 (-19…+4)
Lindsaeaceae	105	120 (117–122)	52	38 (24–45)	26	5 (1–11)	+6	+2 (-2…+6)
Protodrilidae	114	164 (161–166)	56	25 (15–38)	1	3 (1–8)	−1	+1 (-6…+7)
Riama	62	84 (82–86)	23	0 (0–0)	9	4 (1–10)	+1	-2 (-5…+2)

Figures
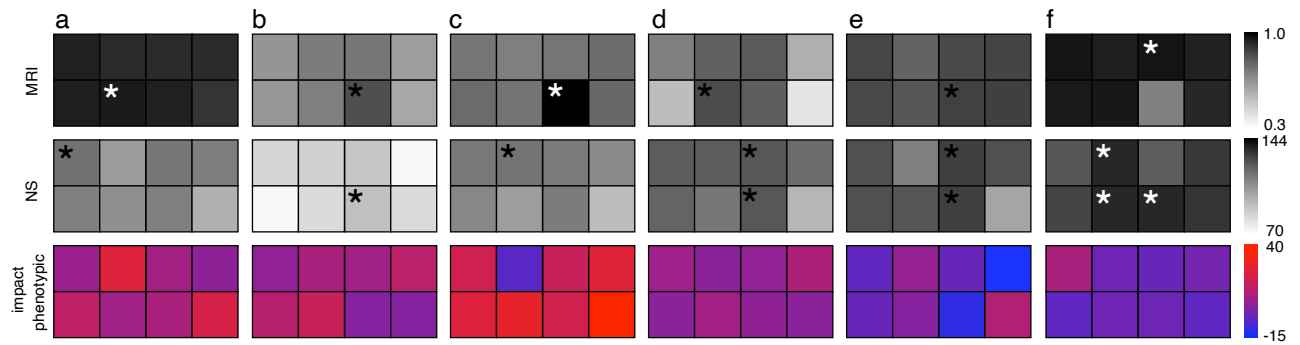
Fig. 1. (a) Sensitivity plots illustrating the Meta-Retention Index (MRI) and Nodal Stability (NS) in sensitivity analyses of the combined molecular data (DNA-only). Cost sets are in the same order as in the Table 1 and best cost sets are indicated with asterisks. The lower most heat maps show the change in NS after the phenotypic data were added. (a) Alismataceae. (b) *Clematis*. (c) Euphasiaceae. (d) Lindsaeaceae. (e) Protodrilidae. (f) *Riama*.
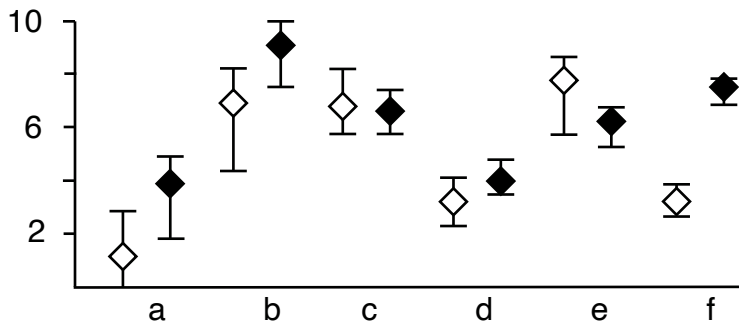

Fig. 2. Median SPR distances between gene trees (white symbols) and from gene trees to phenotypic tree (black symbols), whiskers indicate one standard deviation above and below the mean of the data. (a) Alismataceae. (b) *Clematis*. (c) Euphasiaceae. (d) Lindsaeaceae. (e) Protodrilidae. (f) *Riama*.
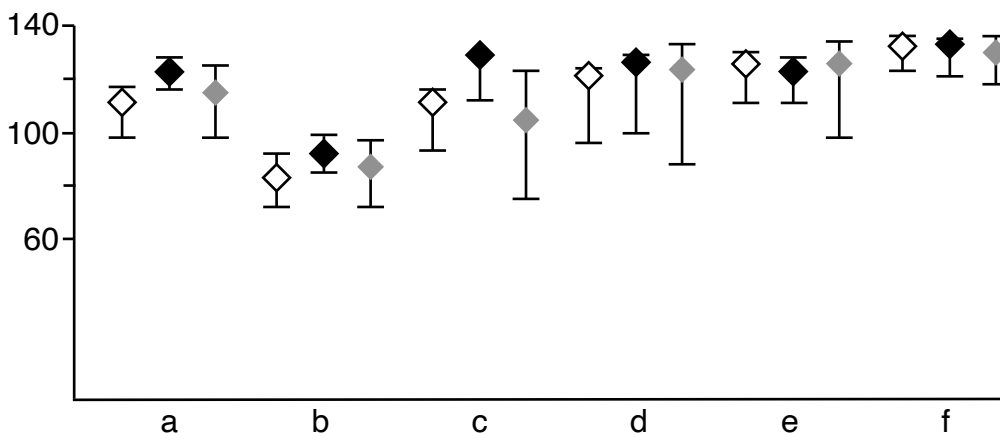

Fig. 3. Median nodal stability across sensitivity analyses of DNA-only data (white symbols), total-evidence (black symbols), and total-evidence with randomized phenotypic data. Whiskers indicate observed minimum and maximum values. (a) Alismataceae. (b) *Clematis*. (c) Euphasiaceae. (d) Lindsaeaceae. (e) Protodrilidae. (f) *Riama*.
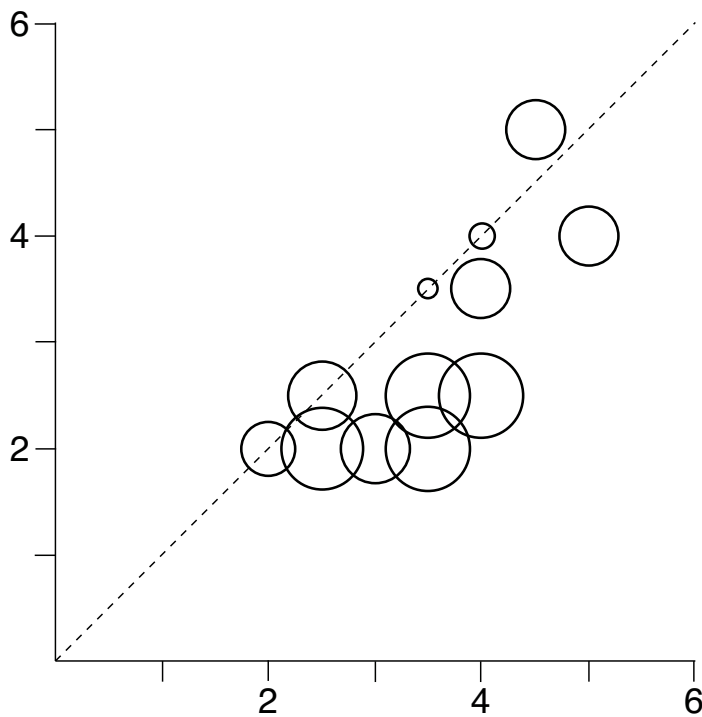
Fig. 4. Total-evidence topologies are more similar with the phenotypic trees than DNA-only topologies. SPR distance from the phenotypic tree to DNA-only tree on x-axis, and SPR distance from the phenotypic tree to total-evidence tree on y-axis. Size of the symbol corresponds with the average jackknife support of the phenotypic tree.