WILEY-BLACKWELL

Cladistics 27 (2011) 538–549

Cladistics

10.1111/j.1096-0031.2010.00345.x

Running $_{W}$ILD: the case for exploring mixed parameter sets in sensitivity analysis

Prashant P. Sharma*, Varpu Vahtera, Gisele Y. Kawauchi and Gonzalo Giribet

Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

Accepted 4 November 2010

Abstract

The robustness of clades to parameter variation may be a desirable quality or even a goal in phylogenetic analyses. Sensitivity analyses used to assess clade stability have invoked the incongruence length difference (ILD or $_{W}$ILD) metric, a measure of congruence among datasets, to compare a series of most-parsimonious results from re-running analyses under different analytical conditions. It is also common practice to select a single ''optimal'' parameter set that minimizes $_{W}$ILD across all parameter sets. However, the divergent molecular evolution of ribosomal genes and protein-encoding genes—specifically the bias against transversion events in coding genes of conserved function—suggests that deployment of multiple parameter sets could outperform the use of a single parameter set applied to all molecules. We explored congruence in five published datasets by including mixed parameter sets in our sensitivity analysis. In four cases, mixed parameter sets outperformed the previously reported, single optimal parameter set. Conversely, multiple parameter sets did not outperform a single optimal parameter set in a case in which actual strong topological conflict exists between data partitions. Exploration of mixed parameter sets may prove useful when combining ribosomal and protein-encoding genes, due to the relatively higher frequency of single- and double-base pair indel events in the former, and the relatively lower frequency of transversions in the latter.

In phylogenetic systematics, sensitivity analysis is commonly used to assess clade stability, a measure of clade robustness. The justification for sensitivity analysis is that clades insensitive to parameter variation are considered ''well supported'' or good candidates for taxonomic action (Wheeler, 1995, Wheeler and Hayashi, 1998; Giribet, 2003), as evidence for that clade is derived from multiple types of synapomorphies (e.g. indels, transversions). Robustness to parameter variation is therefore a desirable quality of clades, although there are philosophical objections to conducting sensitivity analysis (e.g. Grant and Kluge, 2003, 2005; but see rebuttals by Giribet and Wheeler, 2007; Goloboff et al., 2008).

Operationally, conducting sensitivity analysis requires varying the cost of one or more parameters (or alternatively, analytical method) and scoring monophyletic nodes under each parameter set. The result can be represented visually as a two-axis sensitivity plot (or ''Navajo rug'') corresponding to a given clade (e.g. Wheeler, 1995; Giribet and Wheeler, 1999; Giribet and Wheeler 2002; Janies, 2001; Giribet et al., 2001; Edgecombe et al., 2002; Hormiga et al., 2003; Damgaard et al., 2005; Jarvis et al., 2005; Laamanen et al., 2005; Ogden and Whiting, 2005; Clouse and Giribet, 2007; Richter et al., 2007; Gómez-Zurita et al., 2008; Whiting et al., 2008; Álvarez-Padilla et al., 2009; Arnedo et al., 2009; Cameron et al., 2009; Sharma and Giribet, 2009a,b; Petersen et al., 2010; Sanders, 2010). The outcome of a sensitivity analysis is a series of most-parsimonious trees (MPTs) found under the different parameter sets. Discerning the single MPT from this series as the optimal solution is achieved by selecting the tree that maximizes congruence (Wheeler, 1995). Multiple congruence measures have been proposed and

*Corresponding author:
E-mail address: psharma@fas.harvard.edu

their performance compared (Aagesen et al., 2005), but the metric generally used to evaluate congruence among datasets is the incongruence length difference (ILD) measure (Mickevich and Farris, 1981; Farris et al., 1995) or a derivative, the Wheeler ILD ($_{W}$ILD; Wheeler, 1999):

$$_{W}ILD = (L_C - \sum L_I)(L_C^{-1}),$$

where L_I is the length of the MPT generated by each individual partition, and L_C is the length of the MPT generated by combining all partitions.

The ILD metric has received substantial criticism due to trivial minimal congruence in cases where data partitions are given different weights (Dowton and Austin, 2002), and other underperformance concerns (Dolphin et al., 2000; Barker and Lutzoni, 2002; Hipp et al., 2004; Ramírez, 2006; Wheeler et al., 2006). However, it has been shown that multiple congruence measures tend toward the same optimal solution when combining sufficient data (Aagesen et al., 2005). For this reason, the ILD (or its derivatives, such as the $_{W}$ILD) metric has remained in common use in sensitivity analysis as a meta-optimality criterion for comparing MPTs obtained under various parameter sets (e.g. Cryan et al., 2001; Janies, 2001; Maxmen et al., 2003; Whiting et al., 2003, 2008; Damgaard et al., 2005; Laamanen et al., 2005; Terry and Whiting, 2005; Bocakova et al., 2007; Boyer and Giribet, 2007, 2009; Lindgren and Daly, 2007; Richter et al., 2007; Schulze et al., 2007; Álvarez-Padilla et al., 2009; Arnedo et al., 2009; Sharma and Giribet, 2009a,b; Clouse and Giribet, 2010).

An outstanding concern for sensitivity analysis is the limit of parameter sets explored. If the ILD metric is used as a meta-optimality criterion to select a single best hypothesis, sensitivity analysis may benefit from exploring mixed parameter sets applied to different data partitions as a consequence of the heterogeneous nature of different types of molecular markers (Dowton and Austin, 2002). For example, it may transpire that two different markers yield the same topology under different parameter sets, but uniformly applying a single parameter in fact increases incongruence. This effect may be particularly pronounced when combining sets of genes evolving under very different constraints, such as protein-encoding and ribosomal genes. Protein-encoding genes are often constrained in length variation (indel events have to introduce or excise multiples of three nucleotides in coding sequences for the gene product to remain functional), and substitutions are typically biased toward transitions over transversions (due to patterns of degeneracy in most genetic codes and the wobble effect during translation). By contrast, ribosomal genes are not similarly constrained in length variation (as they are not translated) and substitutions are not as biased toward transitions in stem regions as in protein-coding genes (due to complementarity in strand pairing in stems; substitutions in loop regions of ribosomal genes are more difficult to characterize due to length variability and frequent ambiguity in alignment—although these may be oversimplifications of secondary structure representations that take into account neither tertiary and quaternary structure interactions, nor codon bias).

The majority of datasets used to conduct sensitivity analysis have proceeded by applying a single parameter set uniformly to all molecular loci, calculating the $_{W}$ILD metric and repeating this process for the next parameter set; the parameter set with the lowest $_{W}$ILD metric is selected as the optimal solution (Wheeler and Hayashi, 1998). In the present study, we examined the effect of applying mixed parameter sets to various metazoan datasets to determine whether mixed parameter sets could yield more optimal solutions. We focused on the combination of protein-encoding and ribosomal gene sequences, due to their aforementioned divergent character.

Methods

Simulations

To examine the effect of transition bias on phylogenetic inference, sequences were generated using Seq-Gen ver. 1.3.2 (Rambaut and Grassly, 1997). Sequences of 100 bp were generated for a 16-taxon, fully bifurcating tree with all branch lengths Î¼t = 1.0 (Fig. 1a) under an HKY model (Hasegawa et al., 1985). The branch lengths were kept large to allow for extensive accumulation of mutations. We implemented equal nucleotide base frequencies and equal rates of mutations across sites, varying only the relative costs of transitions and transversions (a *de facto* K80 model; Kimura, 1980). Ten datasets were generated for each of four transition-to-transversion ratios (R): $R = \{0.5, 1, 2, 4\}$. Parsimony analyses of these datasets were based on static homology (as generated sequences were length-invariable), using the ''prealigned'' command in software POY ver. 4.1.2 (Varón et al., 2010). Tree searches were conducted by multiple cycles of (i) random addition sequences with subtree pruning and regrafting (SPR), (ii) tree bisection and reconnection (TBR), (iii) ratcheting (Nixon, 1999), and (iv) tree fusing (Goloboff, 1999, 2002), using the automated timed search feature in POY. Each search was run until the shortest tree was found 500 times (typically, searches were completed in ≤ 10 min, although each search had a set maximum run time of 60 min). For each of the 40 datasets, parsimony searches were conducted for four parameter sets (111, 121, 141, 181) in which the transversion to transition cost ratio ($C_{Tv/Ti}$) was sequentially increased (the indel to base

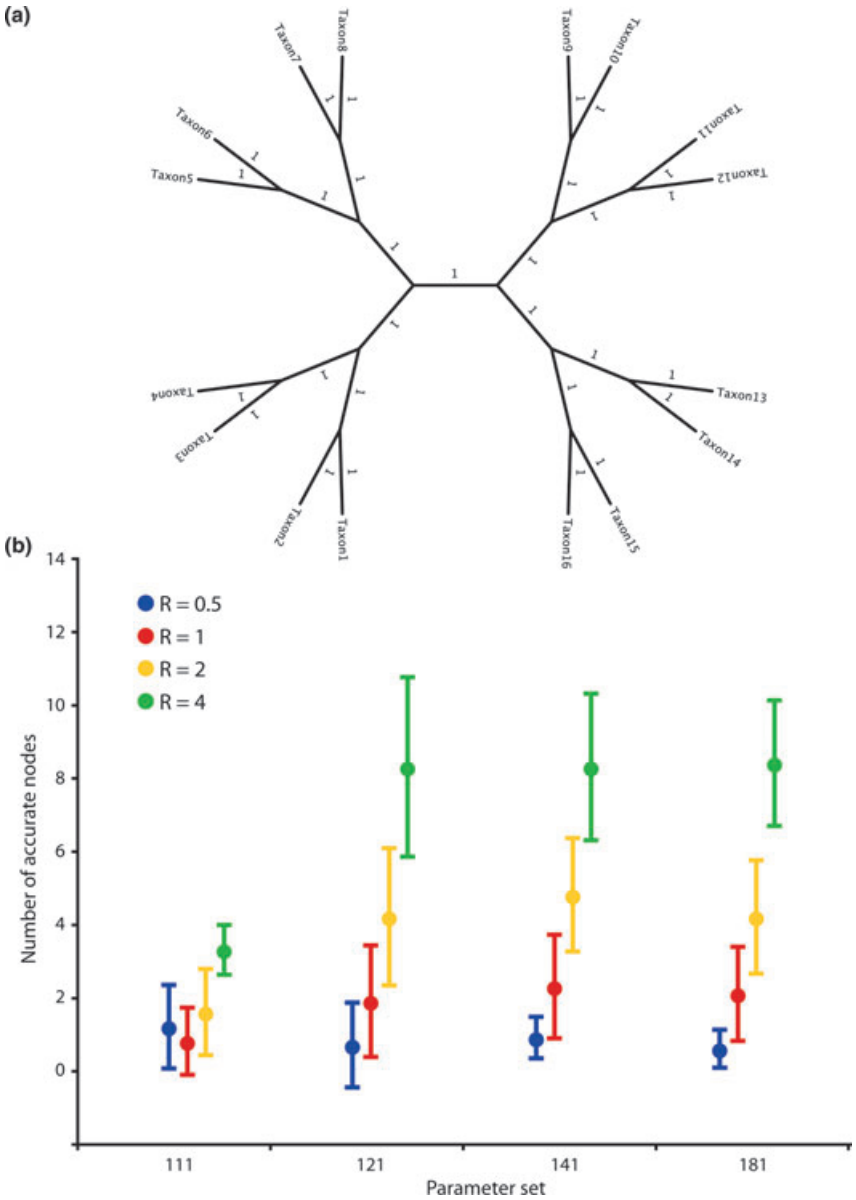

Fig. 1. Testing phylogenetic accuracy with simulated transition bias. (a) True tree used to generate simulated nucleotide sequences. Numbers on nodes indicate branch lengths (all equal to 1). (b) Plot of phylogenetic accuracy under four cost matrices (111, 121, 141, 181; second digit indicates cost ratio of transversions to transitions), for four conditions of transition-to-transversion ratio: $R = \{0.5, 1, 2, 4\}$.

change ratio has no effect in these analyses, as the data are analysed as prealigned and no indels were simulated); parameter set nomenclature follows Boyer and Giribet (2007). We scored the number of nodes successfully reconstructed by each analysis with respect to the true tree.

Histone H3 data

For assessing the effect of parameter variation on congruence between a given protein-encoding gene and its corresponding multilocus phylogeny (the combined molecular tree), we examined the gene histone H3 due to its abundant use in phylogenetic studies of diverse metazoan groups, short sequence length, length invariability, and degree of conservation at the level of peptide sequence. We obtained seven histone H3 datasets (referenced in Table 2) that were constituents of well-resolved multilocus phylogenies and conducted parsimony analyses under four parameter sets (111, 121, 141, 181), as above. We limited the datasets examined to those completed by, or in collaboration with, the Giribet lab to retain access to the original sequence data (chromatograms); this enabled verification of the

Table 1
Histone H3 datasets and number of nodes in histone H3 gene tree congruent with the combined molecular tree under four cost matrices (111, 121, 141, 181; second digit indicates cost ratio of transversions to transitions). Optimal parameter set refers to the single cost matrix that minimizes $_{W}$ILD for the combined molecular tree in the published phylogenies. Asterisks indicate recovery of ingroup monophyly (not found under 111 or 121) only for the Scutigeromorpha dataset

Histone H3 dataset	Reference	Optimal parameter set	Nodes congruent with molecular tree under			
			111	121	141	181
Cephalopoda	Lindgren et al. (2004)	411	10	11	12	11
Opiliones	Giribet et al. (2010)	3221	22	32	27	26
Pettalidae	Boyer and Giribet (2009)	411	16	19	17	14
Sandokanidae	Sharma and Giribet (2009a)	411	12	15	18	18
Scutigeromorpha	Edgecombe and Giribet (2009)	3221	11	11	9*	9*
Sipuncula	Schulze et al. (2007)	111	17	19	19	19
Tetragnathidae	Álvarez-Padilla et al. (2009)	Gap = 4, subst. = 4	5	6	6	5

original sequences' validity. All datasets were treated as prealigned (as before, the lack of indels renders the indel to base change ratio irrelevant). We scored the number of nodes within the ingroup congruent with the corresponding multilocus phylogenies of each dataset.

Mixed parameter sets

To develop and evaluate mixed parameter sets, we examined five multilocus datasets and selected data partitions that were well sampled (sequences for more than 80% of taxa) for both protein-encoding and ribosomal genes. As before, we limited the datasets examined to those completed by, or in collaboration with, the Giribet lab to retain access to the original sequence data (chromatograms). Protein-encoding genes were always treated as prealigned.

The five datasets were as follows (and are summarized in Table 2): we included two gene partitions (18S rRNA and histone H3) from the Opiliones (Arachnida) dataset of Giribet et al. (2010). Within Opiliones, we examined two distantly related clades. First, a dataset for Sandokanidae (Arachnida, Opiliones, Laniatores) included the four most informative gene partitions from a previously published phylogeny (Sharma and Giribet, 2009a): 18S rRNA, 28S rRNA, histone H3, and histone H4. Second, we included a dataset for Pettalidae (Arachnida, Opiliones, Cyphopthalmi) (Boyer and Giribet, 2009) that comprised the ribosomal markers 16S rRNA, 18S rRNA, and 28S rRNA, and the protein-encoding markers cytochrome *c* oxidase subunit I and histone H3 (prealigned).

The fourth dataset was derived from a phylogeny of Cephalopoda (Lindgren et al., 2004) and consisted of the genes 18S rRNA, 28S rRNA, and histone H3. To maintain consistency and comparability with the published phylogeny, we removed the same hypervariable regions of 18S rRNA and 28S rRNA as the authors of the dataset had done.

We specifically included a fifth dataset (Chilopoda; Giribet and Edgecombe, 2006) that documents unambiguous and robust topological conflict between ribosomal and nuclear protein-encoding genes. This dataset consisted of the genes 16S rRNA, 18S rRNA, 28S rRNA, cytochrome *c* oxidase subunit I, elongation factor-1α, elongation factor-2, and RNA polymerase subunit II. As with the Cephalopoda dataset, we removed hypervariable regions of the ribosomal genes, as in the original analyses, to duplicate previous results.

Tree searches were conducted in parallel using the automated timed search feature in POY, and run on 24–30 CPUs of a cluster at Harvard University, Research Computing Group (odyssey.fas.harvard.edu). Individual searches were run for 2–6 h (until the best tree had been found at least ten times). We specified a dynamic homology criterion (Wheeler, 1996) for ribosomal genes. We calculated $_{W}$ILD metrics for 14 single parameter sets, wherein the relative costs of indel, transversion, and transition costs were varied (see Boyer and Giribet, 2007). Two of these parameter sets, designated 3221 and 3211, including explicit costs for gap opening and gap extension events [3221: gap opening = 3, gap extension = 1, substitution = 2 (De Laet, 2005); 3211: gap opening = 3, gap extension = 1, transversion = 2, transition = 1].

For the smallest dataset (Opiliones; 18S rRNA and histone H3), we examined an additional 42 parameter sets in which two different cost matrices were applied to each dataset, calculating the $_{W}$ILD metric for each. We required the transversion cost of histone H3 to match the transition cost of 18S rRNA, effectively down-weighting transition events in histone H3. This convention was adopted for two reasons: first, the parameter space for mixed cost matrices is immense, and for complex datasets, there may be no definable ceiling on the costs to be assessed (as the number of dimensions of the parameter space increase). Second, utilizing a relative cost scheme between data partitions avoids the

problem of trivial solutions, namely minimizing incongruence by down-weighting parameter sets to zero.

For the remaining four datasets (which included more than two data partitions), we grouped the markers into two categories, protein-encoding and ribosomal. We calculated $_{W}$ILD values for the ribosomal genes ($_{W}ILD_{RIB}$) to select their optimal parameter set(s), and thereafter added in protein-encoding genes with different cost matrices (121, 141, 181), such that transversion costs in protein-encoding genes were equal to transition costs in ribosomal genes. This strategy, which explores a smaller parameter space than in the case of the Opiliones dataset, is intended to provide a practical workaround to the problem of an increasing number of parameter space dimensions (each gene partition added increases the number of cost matrix combinations; pursuing the method employed in the Opiliones dataset for datasets with more than five genes becomes untenable). In this manner, we explored three to six mixed parameter sets for each of the remaining datasets.

Table 2
Published multilocus datasets used to explore mixed parameter sets. Gene partitions indicate data that were examined in this study

Dataset	Reference	Gene partitions
Opiliones	Giribet et al. (2010)	18S, H3
Sandokanidae	Sharma and Giribet (2009a)	18S, 28S, H3, H4
Pettalidae	Boyer and Giribet (2009)	16S, 18S, 28S, COI, H3
Cephalopoda	Lindgren et al. (2004)	18S, 28S, H3
Chilopoda	Giribet and Edgecombe (2006)	16S, 18S, 28S, COI, EF1α, EFII, POLII

Results

Simulations

Accuracy of phylogenetic reconstruction of datasets simulated under varying transition-to-transversion ratios indicates that when $R = 0.5$ (no transition bias), the performance of all parameter sets is indistinguishable, but highly prone to inaccuracy (Fig. 1b). As R increases, phylogenetic accuracy under all four parameter sets increases, but the three parameter sets with upweighted transversion costs (121, 141, and 181) outperform a parameter set corresponding to equal weights (111). Curiously, for datasets in which $R = 4$ (eight times as many transitions as transversions), the performance of parameter sets 121, 141, and 181 is very similar; the value of the specified cost of transversions does not affect accuracy, provided that transversions are upweighted.

Histone H3 data

With the exception of one case, analysis of histone H3 datasets under various cost matrices in which transversions were upweighted demonstrated that the parameter set 111 is always outperformed by other cost matrices in which transversions are upweighted (Table 1), as measured by the number of nodes congruent with the combined molecular phylogeny. In the remaining case (Scutigeromorpha dataset; Edgecombe and Giribet, 2009), MPTs under parameter sets 111 and 121 had more nodes congruent with the combined tree than under the remaining parameter sets, but MPTs under the two other parameter sets notably recovered the monophyly of the ingroup, which was not found under 111 or 121. That MPTs under parameter set 111 are less congruent with the combined tree than MPTs under other parameter sets is significant, given that the optimal parameter set (i.e. the parameter set that minimizes $_{W}$ILD) of all of these published phylogenies was one in which transversions and transitions were equally weighted (e.g. 111, 411, 3221).

Mixed parameter sets

Congruence in the Opiliones dataset was interrogated using 14 single parameter sets (applied to both genes), as well as 42 mixed parameter sets. For the single parameter set analyses, $_{W}$ILD was minimized by the parameter set 421 ($_{W}ILD = 0.01998$). This parameter set was outperformed by 17 of the 42 mixed parameter sets (Table 3).

For the Sandokanidae, Pettalidae, and Cephalopoda datasets (Tables 4–6, respectively), we circumvented exploring a massive parameter space by minimizing $_{W}$ILD for particular partitions of genes (e.g. nuclear ribosomal) and using the winning parameter sets for the partitions to construct the mixed parameter set. In all three cases, the mixed parameter set thus constructed outperformed the application of a single parameter set. The improvement was not as drastic in the Pettalidae dataset, but this was expected, given the limited length variation in the ribosomal gene partitions.

The Chilopoda dataset contains well-documented conflict between protein-encoding and ribosomal genes. In this case, a single parameter set (3221) outperformed the mixed parameter sets we explored (Table 7). The change in $_{W}$ILD between the optimal mixed parameter set and the optimal uniform parameter set for all five datasets explored is shown in Fig. 2.

Discussion

The structure and function of a gene play an ineluctable role in its performance as a phylogenetic

Table 3
Tree lengths and calculated incongruence length difference ($_{W}$ILD) values for the Opiliones dataset (Giribet et al., 2010) under singly applied (upper box) and mixed (lower box) parameter sets. Values in bold italics indicate the optimal singly applied parameter set (421, upper box), and the mixed parameter sets that outperform it

		18S	H3	18S + H3	$_{W}$ILD
	111	965	1606	2643	0.02724
	121	1337	2261	3691	0.02520
	141	2060	3527	5752	0.02869
	181	3493	6037	9861	0.03357
	211	1011	1606	2685	0.02533
	221	1421	2261	3771	0.02360
	241	2227	3527	5913	0.02689
	281	3829	6037	10 178	0.03065
	411	1083	1606	2752	0.02289
	421	1563	2261	3902	***0.01999***
	441	2506	3527	6181	0.02394
	481	4374	6037	10 716	0.02846
	3221	1974	3212	5323	0.02574
	3211	1414	2261	3761	0.02287
18S	H3	18S	H3	18S + H3	$_{W}$ILD
111	121	1930	2261	4294	0.02399
	141	3860	3527	7574	0.02469
	181	7720	6037	14 130	0.02640
121	111	1337	1606	3023	0.02646
	141	5348	3527	9094	0.02408
	181	10 696	6037	17 166	0.02522
141	111	2060	1606	3753	0.02318
	121	4120	2261	6507	***0.01936***
	181	16 480	6037	23017	0.02172
181	111	3493	1606	5193	***0.01810***
	121	6986	2261	9388	***0.01502***
	141	13 972	3527	17 769	***0.01520***
211	121	2022	2261	4378	0.02170
	141	4044	3527	7742	0.02209
	181	8088	6037	14 466	0.02357
221	111	1421	1606	3105	0.02512
	141	5684	3527	10 965	***0.15996***
	181	11 368	6037	24 629	0.29331
241	111	2227	1606	3921	0.02244
	121	4454	2261	6846	***0.01914***
	181	17 816	6037	24 350	0.02041
281	111	3829	1606	5536	***0.01824***
	121	7658	2261	10 076	***0.01558***
	141	15 316	3527	19 126	***0.01480***
411	121	2166	2261	4518	0.02014
	141	4332	3527	8022	0.02032
	181	8664	6037	15 026	0.02163
421	111	1563	1606	3238	0.02131
	141	6252	3527	9949	***0.01709***
	181	12 504	6037	18 873	***0.01759***
441	111	2506	1606	4194	***0.01955***
	121	5012	2261	7386	***0.01530***
	181	20 048	6037	26 478	***0.01484***
481	111	4374	1606	6080	***0.01645***
	121	8748	2261	11 157	***0.01327***
	141	17 496	3527	21 292	***0.01263***
3221	121	1974	2261	4322	0.02013
	141	3948	3527	7649	0.02275
	181	7896	6037	14 281	0.02437
3211	111	1414	1606	3098	0.02518
	141	2828	3527	6516	0.02471
	181	5656	6037	12 014	0.02672

Table 4
Tree lengths and calculated incongruence length difference ($_W$ILD) values for the Sandokanidae dataset (Sharma and Giribet, 2009a) under singly applied (upper box) and mixed (lower box) parameter sets. Values in bold italics indicate the optimal parameter set from each group

	RIB	H3	H4	MOL	$_W$ILD MOL
111	1225	484	236	1974	0.01469
121	1780	681	333	2841	0.01654
141	2865	1072	518	4547	0.02023
181	5021	1848	886	7943	0.02367
211	1330	484	236	2078	0.01347
221	1984	681	333	3047	0.01608
241	3268	1072	518	4955	0.01958
281	5825	1848	886	8751	0.02194
411	1510	484	236	2260	***0.01327***
421	2328	681	333	3395	0.01561
441	3945	1072	518	5641	0.01879
481	7163	1848	886	10 121	0.02213
3221	2535	968	472	4029	0.01340
3211	1950	681	333	3011	0.01561

RIB	HIST	RIB	HIST	MIX	$_W$ILD MIX
411	121	3020	1042	4084	***0.00539***
	141	6040	1643	7728	0.00582
	181	12 080	2839	15 012	0.00620
211	121	2660	1042	3726	0.00644
	141	5320	1643	7014	0.00727
	181	10 640	2839	13 590	0.00817

Table 5
Tree lengths and calculated incongruence length difference ($_W$ILD) values for the Pettalidae dataset (Boyer and Giribet, 2009) under singly applied (top box) and mixed (bottom box) parameter sets. Middle box indicates tree lengths and $_W$ILD values for ribosomal genes only, which were used in an intermediate step to construct the mixed parameter set. Values in bold italics indicate the optimal parameter set from each group

	18S	28S	16S	COI	H3	MOL	$_W$ILD MOL
111	226	2172	3016	4713	597	10 953	0.02091
121	326	3590	5174	7090	798	17 340	0.02088
141	520	6322	9335	11 771	1185	29 795	0.02222
211	232	3002	3469	4713	597	12 258	0.01999
221	334	5195	6006	7090	798	19 812	***0.01963***
241	536	9505	10 979	11 771	1185	34 688	0.02053
411	240	4554	4078	4713	597	14 477	0.02038
421	350	8262	7181	7090	798	24 168	0.02015
441	568	15 599	13 304	11 771	1185	43 382	0.02201
3221	460	4129	6351	9426	1194	22 033	0.02147
3211	334	4158	5834	7090	798	18 602	0.02086

			18S	28S	16S	RIB	$_W$ILD RIB
		111	226	2172	3016	5499	0.01546
		121	326	3590	5174	9224	0.01453
		141	520	6322	9335	16 439	0.01594
		211	232	3002	3469	6803	0.01470
		221	334	5195	6006	11 696	0.01377
		241	536	9505	10 979	21 316	0.01389
		411	240	4554	4078	8995	***0.01367***
		421	350	8262	7181	16 031	0.01485
		441	568	15 599	13 304	29 937	0.01557
		3221	460	4129	6351	11 132	0.01725
		3211	334	4158	5834	10 472	0.01394

RIB	PROT	18S	28S	16S	COI	H3	MIX	$_W$ILD MIX
411	121	480	9108	8156	7090	798	26 142	***0.01951***
	141	960	18 216	16 312	11 771	1185	49 444	0.02022
	181	1920	36 432	32 624	21 086	1951	96 081	0.02152

Table 6
Tree lengths and calculated incongruence length difference ($_{W}$ILD) values for the Cephalopoda dataset (Lindgren et al., 2004) under singly applied (upper box) and mixed (lower box) parameter sets. Values in bold italics indicate the optimal parameter set from each group

		RIB	H3	MOL	$_{W}$ILD MOL
	111	1884	1116	3109	0.03506
	121	2771	1507	4439	0.03627
	141	4445	2270	6982	0.03824
	181	7757	3761	12 010	0.04097
	211	2135	1116	3376	0.03703
	221	3233	1507	4916	0.03580
	241	5365	2270	7913	0.03513
	281	9593	3761	13 869	0.03713
	411	2542	1116	3789	***0.03457***
	421	3983	1507	5696	0.03617
	441	6813	2270	9417	0.03547
	481	12 421	3761	16 817	0.03776
	3221	3866	2232	6327	0.03619
	3211	3074	1507	4746	0.03477
RIB	HIST	RIB	HIST	MIX	$_{W}$ILD MIX
411	121	5084	1507	6803	0.03116
	141	10 168	2270	12 814	***0.02934***
	181	20 336	3761	24 830	0.02952
3221	121	3866	1507	5559	0.03346
	141	7732	2270	10 333	0.03203
	181	15 464	2232	19 877	0.10972

marker. The practical necessity of multilocus datasets, sensitivity analysis, differential character weighting, and informativeness measures attests to the variability of molecular loci in phylogenetic analyses (Townsend, 2007). This variability is a dissuasive reason for the application of a single parameter set defined *a priori* in parsimony analysis, such as equal weights (Grant and Kluge, 2003) or the parameter set 3221 (De Laet, 2005), although justifications of these methods have been suggested (Grant and Kluge, 2003).

Sensitivity analysis offers a solution—with its own limitations—by enabling assessment of nodal stability to parameter variation. In addition to indicating which synapomorphy classes support or disfavour particular hypotheses of monophyly, sensitivity analysis also engenders a series of most-parsimonious hypotheses that can be compared by an external criterion, such as congruence measures—although it is not a necessity of a sensitivity analysis to choose a single parameter set (e.g. see Giribet, 2003). However, the practice of applying a single parameter set uniformly to multiple data partitions, even when selected as the best of a series of competing parameter sets, is questionable (Dowton and Austin, 2002). This is because two genes may yield conflicting topologies under any single, uniformly applied parameter set, but a perfectly congruent tree under two different parameter sets (e.g. the Sandokanidae dataset). In such cases, applying a single uniform parameter set will falsely engender incongruence.

We directed our attention to protein-encoding genes, which have previously been reported to engender conflict with ribosomal genes in parsimony analysis (Giribet and Edgecombe, 2006; Sharma and Giribet, 2009a). We observed in simulations that transition bias strongly affects accuracy of phylogenetic reconstruction, such that parameter sets with upweighted transversion costs are better able to reconstruct the true tree as transition bias increases than parameter sets in which transition and transversion costs are equally weighted (e.g. equal weights, 3221). This result is troubling for two reasons. First, datasets of ancient taxa, such as Opiliones, Sipuncula, and Chilopoda, typically have high transition bias in protein-encoding genes (e.g. $R \approx 6.57$ for Opiliones histone H3; Giribet et al., 2010). Secondly, the optimal parameter set selected—either *a priori* or subsequent to sensitivity analysis—in many published phylogenies we have examined is one in which the costs of transitions and transversions are identical (Table 1), probably due to the influence of the relatively larger ribosomal datasets.

We additionally observe that for one archetypal gene, histone H3, the number of nodes congruent with the combined molecular tree is almost always greater for parameter sets in which transversion cost exceeds transition cost (Table 1), consistent with our simulation. These observations suggest that mixed parameter sets may engender better congruence for some datasets than a single parameter set.

In the case of a small, two-gene dataset (Opiliones, Table 3), multiple mixed parameter sets outperformed the single parameter set selected as optimal by the traditional method (*sensu* Wheeler and Hayashi, 1998). When

Table 7
Tree lengths and calculated incongruence length difference ($_{W}$ILD) values for the Chilopoda dataset (Giribet and Edgecombe, 2006) under singly applied (top box) and mixed (bottom box) parameter sets. Middle boxes indicate tree lengths and $_{W}$ILD values for ribosomal genes and for protein-encoding genes, which were used in an intermediate step to construct the mixed parameter set

	16S	18S	28S	COI	EF1α	EF2	POLII	MOL	$_{W}$ILD MOL
111	1108	1968	118	2151	3430	7109	4283	25 118	0.19711
121	1730	2986	157	3191	4776	9931	5994	37 328	0.22940
141	2958	4996	242	5214	7421	15481	9376	61 163	0.25301
181	5276	8988	393	9240	12 678	26 505	16 078	108 507	0.27048
211	1216	2457	129	2151	3430	7109	4283	27 969	0.25721
221	1913	3955	170	3191	4776	9931	5994	42 593	0.29730
241	3255	6919	247	5214	7421	15 481	9376	71 445	0.32937
281	5967	12 799	130	9240	12 678	26 505	16 078	12 8744	0.35223
411	1351	3397	130	2151	3430	7109	4283	32 652	0.33079
421	2199	5796	170	3191	4776	9931	5994	51 718	0.38016
441	3809	10483	248	5214	7421	15 481	9376	89 447	0.41829
481	7041	19 881	404	9240	12 678	26 505	16 078	16 4651	0.44229
3211	1882	3440	169	3191	4776	9931	5994	39 570	0.25744
3221	2330	3779	236	4302	6860	14 218	8566	49 247	***0.18186***

					16S	18S	28S	RIB	$_{W}$ILD RIB
				111	1108	1968	118	3235	0.01267
				121	1730	2986	157	4974	0.02031
				141	2958	4996	242	8329	0.01597
				181	5276	8988	393	14 978	0.02143
				211	1216	2457	129	3900	0.02513
				221	1913	3955	170	6236	0.03175
				241	3255	6919	247	10 720	0.02789
				281	5967	12 799	130	20 042	0.05718
				411	1351	3397	130	5035	0.03118
				421	2199	5796	170	8471	0.03612
				441	3809	10 483	248	15 333	0.05172
				481	7041	19 881	404	28 762	0.04993
				3211	1882	3440	169	5573	0.01471
				3221	2330	3779	236	6403	***0.00906***

		COI	EF1α	EF2	POLII	Nuclear, coding	All coding	ILD nCOD	$_{W}$ILD COD
111		2151	3430	7109	4283	14 941	17 197	0.00796	0.01303
121		3191	4776	9931	5994	20 855	24 156	***0.00738***	***0.01093***
141		5214	7421	15 481	9376	32 543	37 925	0.00814	0.01142
181		9240	12 678	26 505	16 078	55 897	65 405	0.01138	0.01382
211		2151	3430	7109	4283	14 941	17 197	0.00796	0.01303
221		3191	4776	9931	5994	20 855	24 156	***0.00738***	***0.01093***
241		5214	7421	15 481	9376	32 543	37 925	0.00814	0.01142
281		9240	12 678	26 505	16 078	55 897	65 405	0.01138	0.01382
411		2151	3430	7109	4283	14 941	17 197	0.00796	0.01303
421		3191	4776	9931	5994	20 855	24 156	***0.00738***	***0.01093***
441		5214	7421	15 481	9376	32 543	37 925	0.00814	0.01142
481		9240	12 678	26 505	16 078	55 897	65 405	0.01138	0.01382
3211		3191	4776	9931	5994	20 855	24 156	***0.00738***	***0.01093***
3221		4302	6860	14 218	8566	29 882	34 394	0.00796	0.01303

RIB	PROT	16S	18S	28S	COI	EF1α	EF2	POLII	MIX	$_{W}$ILD MIX
3221	121	2330	3779	236	3191	4776	9931	5994	38 911	0.22292
	141	4660	7558	472	5214	7421	15 481	9376	67 345	0.25485
	181	9320	15 116	944	9240	12 678	26 505	16 078	124 189	0.27626

constructing the mixed parameter set, we limited the number of possibilities by setting the cost of transversions in protein-encoding genes equal to the cost of transitions in the ribosomal genes, so as to arrive at absolute costs for all genes in the mixed parameter sets, although other solutions may exist in an infinite parameter space.

Linking the costs of gene partitions has the added advantage of obviating trivial solutions, such as down-weighting incongruent partitions to zero. In previous analyses in which cost matrices explored have included those with zero transition cost (e.g. 110, 220, 440) uniformly applied to all datasets, these parameter sets

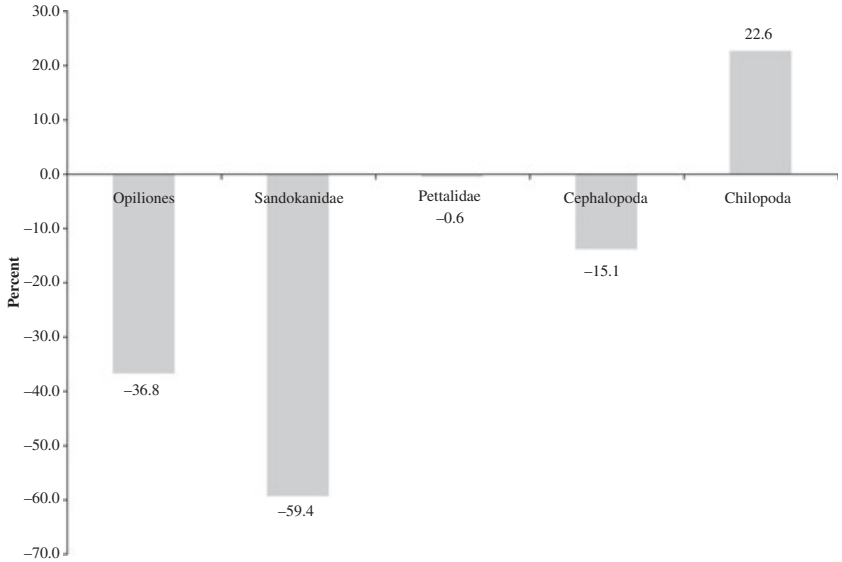

Fig. 2. Change in $_{W}$ILD of the optimal mixed parameter set with respect to the optimal uniform parameter set, expressed as a percentage, for the five datasets explored. The negative change in $_{W}$ILD (first four datasets) indicates that mixed parameter sets explored outperformed the optimal uniform parameter set.

have in fact engendered greater incongruence than those with non-zero transition costs (e.g. Wheeler, 1995; Janies, 2001; Boyer and Giribet, 2007). This suggests that contingent parameter set costs may avoid trivial solutions in cases where removing information from data partitions generates greater incongruence.

However, computing all possible combinations of mixed parameter sets (even with the restrictions we imposed) in this manner becomes untenable for multilocus datasets, as the number of mixed parameter sets increases exponentially. One possible workaround that we executed successfully for three datasets (Tables 4–6) is to divide genes into groups by location and function (nuclear ribosomal, nuclear protein-encoding, mitochondrial ribosomal, etc.), calculate optimal (or nearly optimal) parameter sets for each group, and use these to explore a smaller number of mixed parameter sets. This strategy exploits the observation that datasets dominated by ribosomal genes tend to have incongruence minimized by parameter sets with explicit indel costs, but equal costs for transitions and transversions (e.g. 211, 3221), whereas protein-encoding genes tend to have incongruence minimized by parameter sets with up-weighted transversion costs (e.g. 121, 181; Tables 2 and 7). Parameter sets thus explored outperformed the application of a single parameter set in these cases.

We do not mean to suggest that a single uniformly applied parameter set cannot outperform mixed parameter sets. In some cases of closely related, recently diverging taxa (i.e. short branch lengths), a single parameter set may be sufficient to minimize incongruence among gene partitions because transition bias may not have accumulated in protein-encoding genes, and length variability may likewise not have accumulated in ribosomal genes. Furthermore, in cases where profound topological incongruence occurs among various gene partitions, such that different tree topologies are robustly supported by particular datasets, mixed parameter sets may not be able to outperform a single uniformly applied parameter set using the workaround strategy we employed (i.e. down-weighting transitions in protein-encoding genes). We selected the Chilopoda dataset of Giribet and Edgecombe (2006) as one such counter-case (Table 7).

In addition to finding Chilopoda monophyletic, most of the previously published phylogenies based on molecular, morphological, or combined data have placed the order Scutigeromorpha as sister to the rest of the centipedes (Borucki, 1996; Edgecombe et al., 1999; Giribet et al., 1999; Edgecombe and Giribet, 2002, 2004). In the Chilopoda dataset (Giribet and Edgecombe, 2006), the combined analysis of the ribosomal genes (16S rRNA, 18S rRNA, 28S rRNA) under parameter set 3221 (optimal for ribosomal genes) does not resolve Chilopoda as monophyletic (due to the placement of the diplopod species *Glomeris marginata* in the base of Geophilomorpha), but does place Scutigeromorpha as sister to the rest of the centipedes. By contrast, protein-encoding genes (cytochrome *c* oxidase subunit I, elongation factor-1α, elongation factor-2, RNA polymerase subunit II) under parameter set 121 (optimal for both protein-encoding and nuclear protein-encoding genes) support the monophyly of Chilopoda, but the placement of Scutigeromorpha is derived with respect to the remaining orders (within a paraphyletic Scolopendromorpha). We explored mixed parameter

sets in the same manner as the previous three datasets. As expected, the mixed parameter sets we explored did not outperform 3221, the single optimal parameter set for all genes. This may be because, as suggested by Giribet and Edgecombe (2006), there seems to be an inescapable difference in phylogenetic signal between the protein-encoding and ribosomal genes of centipedes. Altering the weighting scheme by down-weighting transitions in protein-encoding genes cannot conceal this difference if the alternative topology is supported by other classes of synapomorphies (e.g. transversions). However, continued exploration of parameter space, without the cost restrictions and limitations of the workaround strategy we used, should yield a mixed parameter set that outperforms 3221; given our constraints, we simply did not encounter this solution.

Conclusion

In the present study, we conducted an exploration of mixed parameter sets in sensitivity analyses of datasets that included protein-encoding and ribosomal genes. Our exploration of parameter space is not exhaustive, nor do we contend that a single uniformly applied parameter set cannot be optimal with respect to mixed parameter sets. However, at the point where we find better (i.e. more congruent) MPTs with mixed parameter sets than with single uniformly applied parameter sets in multiple cases, as measured by $_{W}$ILD, mixed parameter sets have empirical justification and are worthy of exploration.

Acknowledgements

Sarah L. Boyer and Annie R. Lindgren provided datasets utilized in this study. Feedback from Jessica M. Baker informed the comparisons between gene regions, discussed in the Introduction. Debate with Ronald M. Clouse matured some of the ideas presented in the Discussion. Michael Ethier, Chris Walker and the Harvard Research Computing group facilitated cluster operations. Dennis Stevenson, Lone Aagesen, Ward C. Wheeler, and an anonymous reviewer provided critique and recommendations that improved the quality of an earlier version of this manuscript.

References

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

Álvarez-Padilla, F., Dimitrov, D., Giribet, G., Hormiga, G., 2009. Phylogenetic relationships of spider family Tetragnathidae (Araneae, Araneoidea) based on morphological and DNA sequence data. Cladistics 25, 109–146.

Arnedo, M.A., Hormiga, G., Scharff, N., 2009. Higher-level phylogenetics of linyphiid spiders (Araneae, Linyphiidae) based on morphological and molecular evidence. Cladistics 25, 231–262.

Barker, F.K., Lutzoni, F.M., 2002. The utility of the incongruence length difference test. Syst. Biol. 51, 625–637.

Bocakova, M., Bocak, L., Hunt, T., Teraväinen, M., Vogler, A.P., 2007. Molecular phylogenetics of Elateriformia (Coleoptera): evolution of bioluminescence and neoteny. Cladistics 23, 477–496.

Borucki, H., 1996. Evolution und phylogenetisches System der Chilopoda (Mandibulata, Tracheata). Verh. Naturwiss. Ver. Hamburg 35, 95–226.

Boyer, S.L., Giribet, G., 2007. A new model Gondwanan taxon: systematics and biogeography of the harvestman family Pettalidae (Arachnida, Opiliones, Cyphophthalmi), with a taxonomic revision of genera from Australia and New Zealand. Cladistics 23, 337–361.

Boyer, S.L., Giribet, G., 2009. Welcome back New Zealand: regional biogeography and Gondwanan origin of three endemic genera of mite harvestmen (Arachnida, Opiliones, Cyphophthalmi). J. Biogeogr. 36, 1084–1099.

Cameron, S.L., Sullivan, J., Song, H., Miller, K.B., Whiting, M.F., 2009. A mitochondrial genome phylogeny of the Neuropterida (lace-wings, alderflies and snakeflies) and their relationship to the other holometabolous insect orders. Zool. Scr. 38, 575–590.

Clouse, R.M., Giribet, G., 2007. Across Lydekker's line—first report of mite harvestmen (Opiliones : Cyphophthalmi : Stylocellidae) from New Guinea. Invertebr. Syst. 21, 207–227.

Clouse, R.M., Giribet, G., 2010. When Thailand was an island—the phylogeny and biogeography of mite harvestmen (Opiliones, Cyphophthalmi, Stylocellidae) in Southeast Asia. J. Biogeogr. 37, 1114–1130.

Cryan, J.R., Liebharr, J.K., Fetzner, J.W., Whiting, M.F., 2001. Evaluation of relationships within the endemic Hawaiian Platynini (Coleoptera: Carabidae) based on molecular and morphological evidence. Mol. Phylogenet. Evol. 21, 72–85.

Damgaard, J., Andersen, N.M., Meier, R., 2005. Combining molecular and morphological analyses of water strider phylogeny (Hemiptera-Heteroptera, Gerromorpha): effects of alignment and taxon sampling. Syst. Entomol. 30, 289–309.

De Laet, J.E., 2005. Parsimony and the problem of inapplicables in sequence data. In: Albert, V.A. (ed.), Parsimony, Phylogeny, and Genomics. Oxford University Press, Oxford, pp. 81–116.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting the results of the incongruence length difference test. Mol. Phylogenet. Evol. 17, 401–406.

Dowton, M., Austin, A.D., 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy—the behavior of the incongruence length difference test in mixed-model analyses. Syst. Biol. 51, 19–31.

Edgecombe, G.D., Giribet, G., 2002. Myriapod phylogeny and the relationships of Chilopoda. In: Bousquets, J.L., Morrone, J.J. (Eds.), Biodiversidad, Taxonomía y Biogeografía de Artrópodos de México: Hacia una Síntesis de su Conocimiento. Prensas de Ciencias, Universidad Nacional Autónoma de México, México, pp. 143–168.

Edgecombe, G.D., Giribet, G., 2004. Molecular phylogeny of Australasian anopsobiine centipedes (Chilopoda: Lithobiomorpha). Invertebr. Syst. 18, 235–249.

Edgecombe, G.D., Giribet, G., 2009. Phylogenetics of scutigeromorph centipedes (Myriapoda: Chilopoda) with implications for species delimitation and historical biogeography of the Australian and New Caledonian faunas. Cladistics 25, 406–427.

Edgecombe, G.D., Giribet, G., Wheeler, W.C., 1999. Phylogeny of Chilopoda: combining 18S and 28S rRNA sequences and morphology. In: Melic, A., de Haro, J.J., Mendez, M., Ribera, I.

(Eds.), Evolución y Filogenia de Arthropoda. Bol. Soc. Entomol. Aragonesa 26, 293–331.
Edgecombe, G.D., Giribet, G., Wheeler, W.C., 2002. Phylogeny of Henicopidae (Chilopoda: Lithobiomorpha): a combined analysis of morphology and five molecular loci. Syst. Entomol. 27, 31–64.
Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10, 315–319.
Giribet, G., 2003. Stability in phylogenetic formulations and its relationship to nodal support. Syst. Biol. 52, 554–564.
Giribet, G., Edgecombe, G.D., 2006. Conflict between datasets and phylogeny of centipedes: an analysis based on seven genes and morphology. Proc. R. Soc. B 273, 531–538.
Giribet, G., Wheeler, W.C., 1999. On gaps. Mol. Phylogenet. Evol. 13, 132–143.
Giribet, G., Wheeler, W.C., 2007. The case for sensitivity: a response to Grant and Kluge. Cladistics 23, 294–296.
Giribet, G., Carranza, S., Riutort, M., Baguñà, J., Ribera, C., 1999. Internal phylogeny of the Chilopoda (Myriapoda, Arthropoda) using complete 18S rDNA and partial 28S rDNA sequences. Philos. Trans. R. Soc. B 354, 215–222.
Giribet, G., Edgecombe, G.D., Wheeler, W.C., 2001. Arthropod phylogeny based on eight molecular loci and morphology. Nature 413, 157–161.
Giribet, G., Wheeler, W.C., 2002. On bivalve phylogeny: a high-level analysis of the Bivalvia (Mollusca) based on combined morphology and DNA sequence data. Invertebr. Biol. 121, 271–324.
Giribet, G., Vogt, L., Pérez González, A., Sharma, P., Kury, A.B., 2010. A multilocus approach to harvestman (Arachnida: Opiliones) phylogeny with emphasis on biogeography and the systematics of Laniatores. Cladistics 26, 408–437.
Goloboff, P.A., 1999. Analyzing large data sets in reasonable times: solutions for composite optima. Cladistics 15, 415–428.
Goloboff, P.A., 2002. Techniques for analyzing large data sets. In: Desalle, R., Giribet, G., Wheeler, W.C. (Eds.), Techniques in Molecular Systematics and Evolution. Birkhäuser Verlag, Basel, pp. 70–79.
Goloboff, P.A., Carpenter, J.M., Arias, J.S., Esquivel, D.R.M., 2008. Weighting against homoplasy improves phylogenetic analysis of morphological data sets. Cladistics 24, 758–773.
Gómez-Zurita, J., Hunt, T., Vogler, A.P., 2008. Multilocus ribosomal RNA phylogeny of the leaf beetles (Chrysomelidae). Cladistics 24, 34–50.
Grant, T., Kluge, A.G., 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. Cladistics 19, 379–418.
Grant, T., Kluge, A.G., 2005. Stability, sensitivity, science and heurism. Cladistics 21, 597–604.
Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.
Hipp, A.L., Hall, J.C., Sytsma, K.J., 2004. Congruence versus phylogenetic accuracy: revisiting the incongruence length difference test. Syst. Biol. 53, 81–89.
Hormiga, G., Arnedo, M., Gillespie, R.G., 2003. Speciation on a conveyor belt: sequential colonization of the Hawaiian Islands by *Orsonwelles* spiders (Araneae, Linyphiidae). Syst. Biol. 52, 70–88.
Janies, D., 2001. Phylogenetic relationships of extant echinoderm classes. Can. J. Zool. 79, 1232–1250.
Jarvis, K.J., Haas, F., Whiting, M.F., 2005. Phylogeny of earwigs (Insecta: Dermaptera) based on molecular and morphological evidence: reconsidering the classification of Dermaptera. Syst. Entomol. 30, 442–453.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Laamanen, T.R., Meier, R., Miller, M.A., Hille, A., Wiegmann, B.M., 2005. Phylogenetic analysis of *Themira* (Sepsidae: Diptera): sensitivity analysis, alignment, and indel treatment in a multigene study. Cladistics 21, 258–271.
Lindgren, A.R., Daly, M., 2007. Impact of length-variable data and alignment criterion on the phylogeny of Decapodiformes (Mollusca: Cephalopoda). Cladistics 23, 464–476.
Lindgren, A.R., Giribet, G., Nishiguchi, M.K., 2004. A combined approach to the phylogeny of Cephalopoda (Mollusca). Cladistics 20, 454–486.
Maxmen, A.B., King, B.F., Cutler, E.B., Giribet, G., 2003. Evolutionary relationships within the protostome phylum Sipuncula: a molecular analysis of ribosomal genes and histone H3 sequence data. Mol. Phylogenet. Evol. 27, 489–503.
Mickevich, M.L., Farris, J.S., 1981. The implications of congruence in *Menidia*. Syst. Zool. 30, 351–370.
Nixon, K.C., 1999. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15, 407–414.
Ogden, T.H., Whiting, M.F., 2005. Phylogeny of Ephemeroptera (mayflies) based on molecular evidence. Mol. Phylogenet. Evol. 37, 625–643.
Petersen, G., Knudsen, H., Seberg, O., 2010. Alignment, clade robustness and fungal phylogenetics—Crepidotaceae and sister families revisited. Cladistics 26, 62–71.
Rambaut, A., Grassly, N.C., 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13, 235–238.
Ramírez, M.J., 2006. Further problems with the incongruence length difference test: ''hypercongruence'' effect and multiple comparisons. Cladistics 22, 289–295.
Richter, S., Olesen, J., Wheeler, W.C., 2007. Phylogeny of Branchiopoda (Crustacea) based on a combined analysis of morphological data and six molecular loci. Cladistics 23, 301–336.
Sanders, J.G., 2010. Program note: Cladescan, a program for automated phylogenetic sensitivity analysis. Cladistics, 26, 114–116.
Schulze, A., Cutler, E.B., Giribet, G., 2007. Phylogeny of sipunculan worms: a combined analysis of four gene regions and morphology. Mol. Phylogenet. Evol. 42, 171–192.
Sharma, P., Giribet, G., 2009a. Sandokanid phylogeny based on eight molecular markers—the evolution of a southeast Asian family of Laniatores (Aranchida, Opiliones). Mol. Phylogenet. Evol. 52, 432–447.
Sharma, P., Giribet, G., 2009b. A relict in New Caledonia: phylogenetic relationships of the family Troglosironidae (Opiliones: Cyphophthalmi). Cladistics 25, 279–294.
Terry, M.D., Whiting, M.F., 2005. Mantophasmatodea and phylogeny of the lower neopterous insects. Cladistics 21, 240–257.
Townsend, J.P., 2007. Profiling phylogenetic informativeness. Syst. Biol. 56, 222–231.
Varón, A., Vinh, L.S., Wheeler, W.C., 2010. POY version 4: phylogenetic analysis using dynamic homologies. Cladistics 26, 72–85.
Wheeler, W.C., 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst. Biol. 44, 321–331.
Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–9.
Wheeler, W.C., 1999. Measuring topological congruence by extending character techniques. Cladistics 15, 131–135.
Wheeler, W.C., Hayashi, C.Y., 1998. The phylogeny of the extant chelicerate orders. Cladistics 14, 173–192.
Wheeler, W.C., Ramírez, M.J., Aagesen, L., Schulmeister, S., 2006. Partition-free congruence analysis: implications for sensitivity analysis. Cladistics 22, 256–263.
Whiting, M.F., Bradler, S., Maxwell, T., 2003. Loss and recovery of wings in stick insects. Nature 421, 264–267.
Whiting, M.F., Whiting, A.S., Hastriter, M.W., Dittmar, K., 2008. A molecular phylogeny of fleas (Insecta: Siphonaptera): origins and host associations. Cladistics 24, 677–707.