Syst. Biol. 59(2):226–241, 2010

DOI:10.1093/sysbio/syp105
Advance Access publication on January 13, 2010

An Evaluation of Phylogenetic Informativeness Profiles and the Molecular Phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae)

SERAINA KLOPFSTEIN[1,*], CHRISTIAN KROPF[1], AND DONALD L. J. QUICKE[2,3]

[1]*Department of Invertebrates, Natural History Museum, Bernastrasse 15, CH-3005 Bern, Switzerland;*
[2]*Division of Biology and Centre for Population Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK; E-mail: d.quicke@imperial.ac.uk; and*
[3]*Department of Entomology, Natural History Museum, London SW7 5BD, UK;*
**Correspondence to be sent to: Department of Invertebrates, Natural History Museum, Bernastrasse 15, CH-3005 Bern, Switzerland; E-mail: klopfstein@nmbe.ch.*

Received 16 January 2009; reviews returned 1 April 2009; accepted 12 November 2009
Associate Editor: Thomas Buckley

Abstract.—How to quantify the phylogenetic information content of a data set is a longstanding question in phylogenetics, influencing both the assessment of data quality in completed studies and the planning of future phylogenetic projects. Recently, a method has been developed that profiles the phylogenetic informativeness (PI) of a data set through time by linking its site-specific rates of change to its power to resolve relationships at different timescales. Here, we evaluate the performance of this method in the case of 2 standard genetic markers for phylogenetic reconstruction, 28S ribosomal RNA and cytochrome oxidase subunit 1 (CO1) mitochondrial DNA, with maximum parsimony, maximum likelihood, and Bayesian analyses of relationships within a group of parasitoid wasps (Hymenoptera: Ichneumonidae, Diplazontinae). Retrieving PI profiles of the 2 genes from our own and from 3 additional data sets, we find that the method repeatedly overestimates the performance of the more quickly evolving CO1 compared with 28S. We explore possible reasons for this bias, including phylogenetic uncertainty, violation of the molecular clock assumption, model misspecification, and nonstationary nucleotide composition. As none of these provides a sufficient explanation of the observed discrepancy, we use simulated data sets, based on an idealized setting, to show that the optimum evolutionary rate decreases with increasing number of taxa. We suggest that this relationship could explain why the formula derived from the 4-taxon case overrates the performance of higher versus lower rates of evolution in our case and that caution should be taken when the method is applied to data sets including more than 4 taxa. [28S rRNA; CO1 mtDNA; information content; nucleotide composition bias; parasitoid; phylogenetic utility; simulation.]

Assessing the information content of a data set is a crucial step in all phylogenetic studies. First, it plays a key role in the estimation of the quality of the data underlying a phylogenetic reconstruction, allowing for a critical evaluation of the credibility of resulting relationships. Second, evaluation of the informativeness of different data sets can act as a basis for the experimental design of future phylogenetic studies. Several factors have been proposed that influence the performance of a gene, such as gene length, stationarity of nucleotide compositions, levels of conflict, symmetry of the transformation rate matrix, and the fit to the applied evolutionary model (e.g., Huelsenbeck and Bull 1996; Conant and Lewis 2001; Posada and Crandall 2001; Jermiin et al. 2004; Lin and Danforth 2004; Collins et al. 2005; Danforth et al. 2005; Sullivan and Joyce 2005; Mueller 2006; Simon et al. 2006; Waegele and Mayer 2007; Fischer and Steel 2009). However, the most often cited factor determining the utility of a gene is its evolutionary rate (e.g., Swofford et al. 1996; Goldman 1998; Yang 1998; Shpak and Churchill 2000; Brinda-Emonds et al. 2001; Sanderson and Shaffer 2002; Lin and Danforth 2004; Danforth et al. 2005; Mueller 2006; Townsend 2007; Jian et al. 2008; Regier et al. 2008). The question "what is the optimum rate of evolution" has been intensely discussed in relation to the influence of saturation (Brown et al. 1982; Graybeal 1994; Meyer 1994; Mindell and Thacker 1996; Källersjö et al. 1999; Wenzel and Siddall 1999; Seo and Kishino 2008). According to simulation studies, the optimum rate should not be too low in order to possess enough information and also not be too high to avoid detrimental effects of homoplasy and saturation (Goldman 1998; Yang 1998; Shpak and Churchill 2000; Brinda-Emonds et al. 2001).

Currently, a number of methods exist that assess whether the characters of a data set evolve at an appropriate rate to resolve a phylogenetic relationship in question. These methods are temporally explicit in that they allow measurement of the expected performance of a data set over different timescales. A posteriori measurements of clade supports, such as bootstrap values, approximate likelihood-ratio test, Bremer support, or Bayesian posterior probabilities (Felsenstein 1985; Bremer 1994; Huelsenbeck et al. 2001; Anisimova and Gascuel 2006), can be viewed as such measures; however, they strongly depend on branch lengths, can be biased in some contexts, and finally incorporate many factors other than evolutionary rates (Efron et al. 1996; Waegele and Mayer 2007; Wiens et al. 2008). Methods to explicitly assess whether the evolutionary rates represented in a data set are likely to resolve any phylogenetic relationship include empirical saturation plots or relative rate comparisons (Brown et al. 1982; Graybeal 1994; Mindell and Thacker 1996). In such analyses, the level of saturation is evaluated by contrasting some measure of observed distance between 2 taxa with their expected distance derived from the time since their divergence. These analyses can provide a measure of the suitability of the rates in the data set, although tests derived in this context have been criticized (Grant and Kluge 2003).

A more quantitative, likelihood-based method for experimental design that can also be used to assess the appropriateness of the rates of evolution in a data set was developed by Goldman (1998).

Recently, Townsend (2007) proposed a new approach that quantifies the informativeness of a character in relation to its evolutionary rate. The informativeness is computed by estimating the probability that a character shows a single change on the internal branch of a symmetric 4-taxon tree and no change afterward on the 4 external branches. By compiling the information from all the characters of a data set (or all the base pairs of a gene), the method assesses a data set's power to resolve the phylogeny at different timescales and accordingly different taxonomic levels. Phylogenetic informativeness (PI) profiles therefore consider the average evolutionary rate of a data set together with information about the extent of among-site rate variation. Data sets in which the different characters show a large variation in their respective rates of evolution show a relatively flat PI profile, whereas data sets with a narrow range of rates concentrate their informativeness on a narrow range in time.

Tools for experimental design promise the highest profit when it comes to very diverse and understudied taxa such as parasitoid wasps. They display very high diversity at all taxonomic levels much of which is still unknown (Townes 1969). Because traditional morphological characters often show high levels of homoplasy (Gauld and Mound 1982; Quicke and Belshaw 1999), the use of molecular tools has already greatly enhanced the understanding of phylogenetic relationships within that group. However, although some informal attempts have been made to compare the utility of different molecular markers for phylogeny reconstruction within the Hymenoptera (Gimeno et al. 1997; Mardulyn and Whitfield 1999; Banks and Whitfield 2006; Murphy et al. 2008), an explicit evaluation of their information content is still missing. Phylogenetic reconstructions of parasitoid wasps have focused on mainly 2 molecular markers: the expansion segments D2 and partial D3 of the large ribosomal subunit 28S ribosomal RNA (28S) and the 5′ region of the mitochondrial cytochrome oxidase subunit 1 (CO1) (Quicke, Lopez-Vaamonde, and Belshaw 1999b; Whitfield et al. 2002; Shi et al. 2005; Quicke et al. 2005; Laurenne et al. 2006; Zaldivar-Riverón et al. 2006, 2008). These 2 gene fragments have also been used widely for phylogenetic reconstructions throughout the animal kingdom, and the CO1 fragment has been proposed as the standard marker for DNA-based species identification ("bar coding"; Hebert et al. 2003).

With approximately 340 described species in 20 genera (Yu and Horstmann 2005), the Diplazontinae are a medium-sized subfamily of Ichneumonidae. Besides some informal notes on possible relationships between some of the genera (Dasch 1964; Fitton and Rotheray 1982), the phylogeny of this group is entirely unknown. Here, we build the first comprehensive phylogeny of Diplazontinae including more than half of the described genera, using 28S and CO1 to evaluate the present genus-level classification. Based on this phylogeny, we derive Townsend's PI profiles of the 2 genes. We find a strong discrepancy between the observed performance of the 2 genes and their informativeness as predicted by the PI profiles, in our own diplazontine data set and also in additional studies that applied the 2 genes. We explore the reasons for the failure of PI profiles to predict the relative performance of the 2 genes. First, violations of the assumptions of the method could lead to a discrepancy between predicted and attained performance, like deviations from a strict molecular clock, phylogenetic uncertainty, and model misspecification. Second, specific properties of the 2 genes might influence their relative performance, like biased and nonstationary nucleotide composition (Foster and Hickey 1999; Conant and Lewis 2001; Collins et al. 2005). Finally, the method was developed in the context of a 4-taxon case; we use simulations to address the question whether the accordingly derived numerical results (Townsend 2007) can be directly applied to larger data sets.

Materials and Methods

Taxon Sampling

We included 63 individuals of 39 Diplazontinae species in our study, covering 11 of the 20 described genera (Appendix 1). For larger genera, we sequenced several species in order to cover as much of the intrageneric morphological variation as possible, and multiple specimens were included in some species to account for part of the intraspecific variation. Only species for which we could obtain sequences from both molecular markers were included in the analyses, although in 2 species (*Promethes sulcator* and *Woldstedtius flavolineatus*), a different specimen was used to obtain both gene sequences. Monophyly of the Diplazontinae is supported by a number of adult (Beirne 1941; Dasch 1964; Townes 1971; Wahl 1990; Wahl and Gauld 1998), larval (Wahl 1990; Wahl and Gauld 1998), and biological characters (e.g., Dasch 1964; Fitton and Rotheray 1982; Yu and Horstmann 2005). They form part of the informal subfamilial clade "Pimpliformes." Within this group, they show an association with the other subfamilies that use Diptera as hosts, the Orthocentrinae and Cylloceriinae (Wahl 1990; Belshaw et al. 1998; Wahl and Gauld 1998; Quicke et al. 2000, 2009). We thus included representatives from these 2 subfamilies and 2 additional pimpliform taxa as outgroups.

Molecular Methods

Genomic DNA was extracted from whole specimens preserved in 80% ethanol using the Promega Wizard kit for blood and tissue extractions. Vouchers and DNA samples are kept at the Natural History Museum in Berne (Appendix 1). Approximately 640 bp of the D2 and partial D3 region of 28S were obtained using the primers designed by Belshaw and Quicke (1997) (fwd: 5′-A AGA GAG AGT TCA AGA GTA CGT G-3′) and

Mardulyn and Whitfield (1999) (rev: 5′-TAG TTC ACC ATC TTT CGG GTC CC-3′). Approximately 660 bp from the 5′ end of the mitochondrial CO1 were amplified using the primers designed by Folmer et al. (1994) (LCO 5′-GGT CAA CAA ATC ATA AAG ATA TTG G-3′, HCO 5′-TAA ACT TCA GGG TGA CCA AAA AAT CA-3′).

Polymerase chain reactions (PCR) were carried out in 20 μl final volumes using Promega GoTaq Flexi DNA Polymerase kits. Final volumes contained 30 pmol $MgCl_2$, 16 pmol of both primers, 4 pmol of each dNTP, 0.3 U Taq polymerase, and 2 μl genomic DNA. PCR conditions were 94 °C for 2 min, 35 cycles of 30 s at 94 °C, 30 s at 53 °C (28S) or 51 °C (CO1), and 1 min at 72 °C, followed by a final extension at 72 °C for 10 min. PCR products were purified with the GFXTM DNA and Gel Purification kit (Amersham Biosciences, Little Chalfont, UK). The PCR products were sequenced on an ABI 377 automated sequencer using Big Dye Terminator technology (Applied Biosystems, Warrington, UK). All sequences have been deposited in the GenBank database under accession nos. FJ556422–FJ556488 (CO1) and FJ556489–FJ556555 (28S) (Appendix 1).

Manual alignment of CO1 after translation into amino acids using Mega 4.0 (Tamura et al. 2007) was straightforward as no indels were detected. The D2–D3 region of the large subunit of 28S was aligned according to published secondary structure maps of ichneumonids (Gillespie et al. 2005), identifying the stem regions for partitioning and the pairing nucleotide positions for the application of the doublet model in MrBayes (see below). Of the identified nonpairing regions, only those that were length conserved across the alignment were included in the analyses, whereas length-variable stretches were excluded. We thus obtained a 657-bp fragment of CO1 and 552 unambiguously alignable base pairs of 28S. Alignments can be downloaded from TreeBASE (study accession number S2511), and a file including secondary structure annotations of 28S can be obtained from S.K. on request.

Phylogenetic Analyses

Phylogenetic reconstructions were conducted using maximum parsimony (MP), maximum likelihood (ML), and Bayesian methods on both the 2 genes separately and jointly. For MP analysis, we used the program TNT (Goloboff et al. 2008), applying both traditional search (tree bisection and reconnection branch swapping and 10,000 random additions) and new technology search (sectorial search, ratchet, drift, and tree fusing with default settings). Clade support was evaluated by a nonparametric bootstrap (Felsenstein 1985) with 1000 replicates and 10 random additions each holding only one tree.

For both the ML and the partitioned Bayesian analyses, we identified the best-fitting nucleotide substitution models using MrModeltest version 2.2 (Nylander 2004), with a neighbor-joining tree as the test tree and applying the Akaike information criterion (following Posada and Buckley 2004). The results of the model choice are shown in Table 1. Likelihood analyses were conducted using phyML 3.0 (Guindon and Gascuel 2003), with default settings and 1000 nonparametric bootstrap iterations. For the Bayesian analysis, we tested different partitioning strategies according to the method proposed by Brandley et al. (2005) and advocated by Brown and Lemmon (2007), which uses $2\times$ln (Bayes factor) > 10 as the criterion for preferring a more partitioned model over a less partitioned one (Kass and Raftery 1995; Brandley et al. 2005). Partitioning schemes are summarized in Table 2 and ranged from an unpartitioned analysis (P1) to a distinction of 4 partitions chosen based on gene identity and prior knowledge of biochemical properties (P4): the pairing stem regions of 28S, its remaining loop regions, combined first and second codon positions of CO1, and finally third codon position of CO1. To obtain an estimate for the Bayes factors associated with each comparison of partitioning strategies, we conducted a Bayesian Markov chain Monte Carlo (MCMC) analysis on MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003) for each partitioning strategy separately. Analyses were run with 2 independent runs of 4 chains each (heating $T = 0.15$), random starting trees and trees sampled every 1000 generations for 5×10^6 generations. Convergence of the 2 runs was checked in 4 ways. The log-likelihood (lnL) scores were plotted over generations and stabilization was determined. The overlay plot of the 2 independent runs was examined for a good mixing of the runs and stabilization of the lnL. Then, we checked whether the standard deviation of split frequencies between the 2 runs fell below the 0.01 threshold (Ronquist and Huelsenbeck 2003). Finally, we studied the behavior of the potential scale reduction factor (PSRF) for the model parameters and clade supports and considered the runs to have converged if the PSRF was less than 5% divergent from 1. Except for partitioning scheme P3a, convergence was reached before generation 2.5×10^6. We then conservatively discarded half of the generations as a burn-in and obtained estimates for the harmonic means of the

TABLE 1. Data partitions, their properties, and estimated models of sequence evolution

Partition	# bp	# var	# pars	AT (%)	Model
CO1	657	288	244	73	GTR + I + Γ
CO1 first and second codon positions	438	105	75	63	GTR + I + Γ
CO1 third codon positions	219	183	169	93	GTR + I + Γ
28S	552	164	112	38	GTR + I + Γ
28S stem	354	112	74	29	GTR + I + Γ
28S loop	198	52	38	54	SYM + I + Γ
Combined data set	1209	452	356	57	GTR + I + Γ

Note: # bp: number of base pairs included in the analysis; # var: number of variable sites; # pars: number of parsimony informative sites; AT: AT content of the respective partition; Model: substitution model chosen by MrModeltest and applied in the partitioned Bayesian analysis; GTR: generalised time-reversible model; SYM: Zharkikh symmetrical model; I: invariant sites; Γ: gamma distribution to model among-site rate variation.

TABLE 2. Partitioning strategies, associated log likelihoods, and Bayes factors

Strategy	# part	Specification	LnL	Bayes factor
P1	1	Unpartitioned data set	−10,451.4	1536.3
P2	2	Partitioned according to gene identity (CO1 and 28S)	−10,232.5	1098.56
P3a	3	28S partitioned into stem and loop, CO1 unpartitioned	−10,276.6	1186.62
P3b	3	28S unpartitioned, CO1 partitioned into first and second versus third codon position	−9903.81	441.12
P4	4	Both genes partitioned as under P3a and P3b, standard nucleotide model	−9878.64	390.78
P4*	4	As P4, but with doublet model for the pairing stem partition of 28S	−9683.25	—

Note: # part: number of partitions; lnL: harmonic mean of the log-likelihoods obtained under the respective partitioning strategy; Bayes factor: Bayes factor calculated from the comparisons with the best-fitting partitioning strategy P4*.

likelihood scores from the remaining generations using the sump command implemented in MrBayes (Ronquist and Huelsenbeck 2003). The same procedure as for the partitioning strategies was applied to test whether the application of the doublet model as implemented in MrBayes (Schoeniger and von Haeseler 1994; Ronquist and Huelsenbeck 2003) significantly improved the phylogenetic estimation (P4*, Table 2).

Convergence diagnostics revealed low convergence even after 5×10^6 generations in the case of partitioning strategy P3a, where 28S was partitioned into stem and loop while CO1 remained unpartitioned. Although the lnL plot seemed to reach a plateau already after 10^6 generations, the overlay plot of the 2 runs revealed that they both stabilized on a different peak, which was supported by the average split frequency that stabilized around 0.009 and did not decrease further until generation 5×10^6. A new analysis with heating set to $T = 0.05$ and the number of generations to 2×10^7 did not produce convergence either. We believe that the reason for this unusual convergence behavior lies in the misspecification of the model which can cause the MCMC search to fail to converge for a long time period (e.g., Ronquist et al. 2006). The likelihood scores of all the runs of P3a were distinctly below the values reached by different models (Table 2), and we thus did not further consider this partitioning strategy.

Final Bayesian analysis of the 2 genes separately and jointly and with the preferred partitioning strategy and model choice were conducted using the MCMC parameters described above but running for 1×10^7 generations. Convergence was reached after 2.27×10^6 generations, but we discarded 5×10^6 generations as a conservative burn-in.

Alternative Hypotheses Testing

Two of the genera that were represented by more than one species were recovered as paraphyletic in all our analyses. To test if this nonmonophyly is statistically supported, we used both a Bayesian and a likelihood-based approach. First, we evaluated whether there was at least one tree in the 95% confidence set of tree topologies that showed either of the 2 genera in question to be monophyletic, which would mean that the Bayesian analysis could not exclude monophyly. To do so, we filtered the Bayesian topologies included in the 95% set of trees with a constraint topology that enforced the monophyly of the genera in question in PAUP* (Swofford 2002). Additionally, we applied the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999) as implemented in PAUP*. Trees for testing were obtained by separately enforcing monophyly of both genera in a Bayesian analysis, using the same settings as employed above. The SH tests were carried out applying the resampling estimated lnL method (Kishino et al. 1990) with 1000 replicates to create the test distribution.

Estimation of PI

PI profiles were constructed using the method described by Townsend (2007). As the test topology, we used the Bayesian topology with the highest posterior probability from the combined analysis. ML branch lengths enforcing a strict molecular clock were obtained from PAUP*, where we also obtained likelihood values for the likelihood ratio test of the molecular clock assumption. To determine site-specific rates of evolution under an ML approach, we used the resulting phylogenetic tree together with the alignment as input for the program HyPhy (Kosakovsky Pond et al. 2005). HyPhy is particularly well suited to estimate site-specific rates because it allows a variety of evolutionary models to be implemented. Furthermore, it correctly assigns a zero rate to invariable positions, which is crucial for obtaining a conservative estimate of the PI profiles at ancient timescales. To estimate site-specific rates based on the set of branch lengths specified in the input tree (and not by reestimating branch lengths from the data as in the built-in function), we compiled a new batch file (available from S.K. on request). We tested different models of evolution and partitioning strategies to evaluate their influence on the estimated PI profiles.

PI profiles were constructed in R (R Development Core Team 2007) according to the formula described in Townsend (2007) (R scripts are available from S.K. on request). As there are no fossils available to date the nodes of the tree, we have not explicitly dated the phylogeny. Instead, we use a relative timescale by assigning the value of 1 to the whole tree length from the basal node of the ingroup to the tips. The PI profiles of the 2 markers as obtained in our study might be influenced by peculiarities of the diplazontine data set; to rule out this possibility, we also retrieved PI profiles using data sets from other studies that used the 2 markers to resolve questions at a similar taxonomic level, 2 from

Hymenoptera (Mardulyn and Whitfield 1999; Zaldivar-Riverón et al. 2008) and 1 from Diptera (Mengual et al. 2008). All 3 data sets used the same region of the 28S gene as we did (D2 and part of D3), but 2 used a larger portion of the CO1 gene (Mardulyn and Whitfield 1999; Mengual et al. 2008). Alignments and tree topologies used were the ones recovered in the respective study, with branch lengths reestimated in PAUP* under a strict molecular clock. For the Diptera data set, however, the alignment was originally conducted under an optimization alignment approach (Mengual et al. 2008). We adopted the 28S alignment as used in this study but realigned CO1 manually after translation into amino acids in order to assure that the different codon positions were correctly aligned. Site-specific rates were in all cases estimated in HyPhy applying a generalized time-reversible model and partitioning CO1 into codon positions.

PI profiles are intended to be used to decide between data sets. They should thus predict which of them performs better in resolving a split at a given time in the past (Townsend 2007; Townsend et al. 2008; Schoch et al. 2009). To test the predictive power of PI profiles on our data set, we compared their predictions to the difference in clade support attained by 28S and CO1 at each node of the total evidence topology with the highest posterior probability, both in the ML and in the Bayesian analyses.

Exploring Reasons for the Low Predictive Power of PI Profiles

Because the PI profiles of 28S and CO1 did not correspond to the observed relative performance of the 2 genes, we aimed to evaluate the robustness of Townsend's criterion for PI against some of its assumptions. First, we tested the impact of a strict molecular clock on the resulting PI profiles. As the assumption of a global molecular clock was significantly rejected for our data set (see Results), we conducted site-rate estimates on 2 additional sets of branch lengths, first under nonparametric rate smoothing as implemented in the ape package of R (Sanderson 1997; Paradis et al. 2004) and second on a nonultrametric tree with unconstrained ML branch lengths.

To test whether the discrepancy between PI profiles and observed support was due to phylogenetic uncertainty, we made use of the Bayesian posterior distribution of tree topologies by sampling every 10,000th tree after the burn-in phase. For each of the 1000 retained trees, we obtained an independent estimation of PI as described above, resulting in a confidence range of PI profiles. The same procedure was repeated with the topologies obtained from the parsimony and ML reconstructions to evaluate the sensitivity of PI profiles against the tree-building method.

Simulating Optimum Rates

Townsend derived his formula for calculating PI based on an idealized 4-taxon tree (figure 1 in Townsend 2007) and assumed that the rate of evolution is optimum when it maximizes the probability of one change occurring on the internal branch and no change occurring on any of the 4 tip branches. We used the same setting to simulate data sets for the 4-taxon case and for data sets of 8, 16, and 32 taxa. Branch lengths were 0.95 for the external branches and 0.05 for the internal branch, and taxa were added with zero branch lengths to the base of each of the 2 respective groups (Fig. 1). Using the evolver module of PAML (version 4.1; Yang 2007), we simulated 10,000 nucleotide positions under a Jukes–Cantor model of evolution, adjusting the total branch length of the tree in order to obtain 0.001–2.5 expected substitution per root-to-tip distance unit. We then counted the number of times that a nucleotide pattern in accordance with one single change on the interior branch (i.e., an unreversed synapomorphy for 1 of the 2 groups) is observed under the different evolutionary rates.

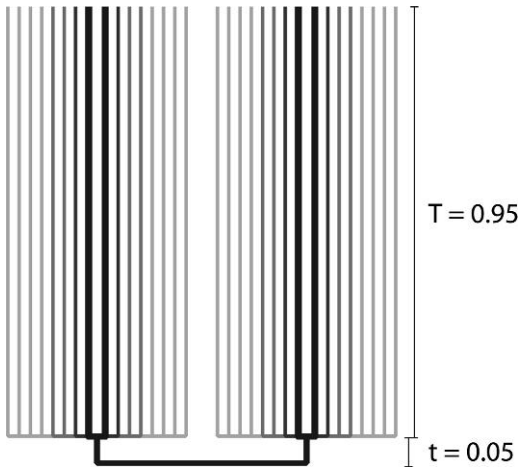

FIGURE 1. Idealized tree topology used in the simulations. *T* denotes the relative length of the external branches and *t* the length of the internal branches. The simulations were conducted using 4 taxa in a setting similar to the one used by Townsend (2007) and adding taxa to a basal polytomy of the 2 groups, thus reflecting an ancient rapid radiation. Total sizes of the data sets were 8, 16, and 32 taxa, represented in the figure by different shades of grey.

RESULTS

Model Selection and Data Partitioning

Table 1 shows the different partitions, their properties, and associated models of evolution as estimated by MrModeltest 2.2 (Nylander 2004). Except for the relatively small 28S loop partition, all partitions were found to fit the GTR + I + G model best. The base compositions of the 2 genes differ markedly, with a strong AT bias in the CO1 gene and a moderate GC bias in 28S, corresponding well to previous observations that have found an AT bias in insect mitochondrial DNA especially at sites under low selective pressure like third codon positions (Foster and Hickey 1999). Base composition differs significantly among taxa at third codon positions of

CO1 (chi-square test of homogeneity as implemented in PAUP*: $P < 0.0001$) while being stationary at the other data partitions and for the combined data set ($P = 1.0$).

Full partitioning of the 2 genes and a doublet model for the stem regions of 28S were preferred by Bayes factor comparisons (Table 2). Globally, we find a significant increase in the likelihood of the respective model when the data were partitioned according to prior knowledge about biochemical properties, except in the case of the nonconverging partitioning strategy P3a (see Materials and Methods). The less partitioned model can be rejected with high confidence in all the cases, a pattern already observed in other partitioned Bayesian analysis (e.g., Nylander et al. 2004; Brandley et al. 2005; Brown and Lemmon 2007). In accordance with previous results, the improvement of the model was greater when CO1 was partitioned according to codon positions than with the ribosomal 28S divided into stem and loop regions (e.g., Pagel and Meade 2004; Aliabadian et al. 2007). Finally, the choice of the doublet model for the pairing region of partial 28S significantly improved the likelihood of the model, as shown by Bayes factor comparisons (Table 2) and as expected for rRNA from other studies (Telford et al. 2005; Leliaert et al. 2007).

Phylogenetic Reconstructions

The topologies obtained from MP, ML, and Bayesian analyses were highly congruent. In MP analysis, both the traditional and the new technology search revealed trees of the same length, although the number of trees found by the different methods differed. The complete data set recovered 33 of the 63 ingroup nodes with a bootstrap support higher than 70%, whereas the 2 genes 28S and CO1 separately resolved 26 and 33 nodes, respectively. Not only the number but also the position of the resolved nodes differed between the 2 genes. 28S resolved more branchings at the genus level, whereas CO1 mainly recovered relationships at the species or intraspecific level (Fig. 2a,b). This pattern is also present in the ML and Bayesian analyses. The ML analyses attained a higher number of significantly supported clades, 41 with the combined data set and 25 and 32 with 28S and CO1, respectively. Of the 22 nonresolved

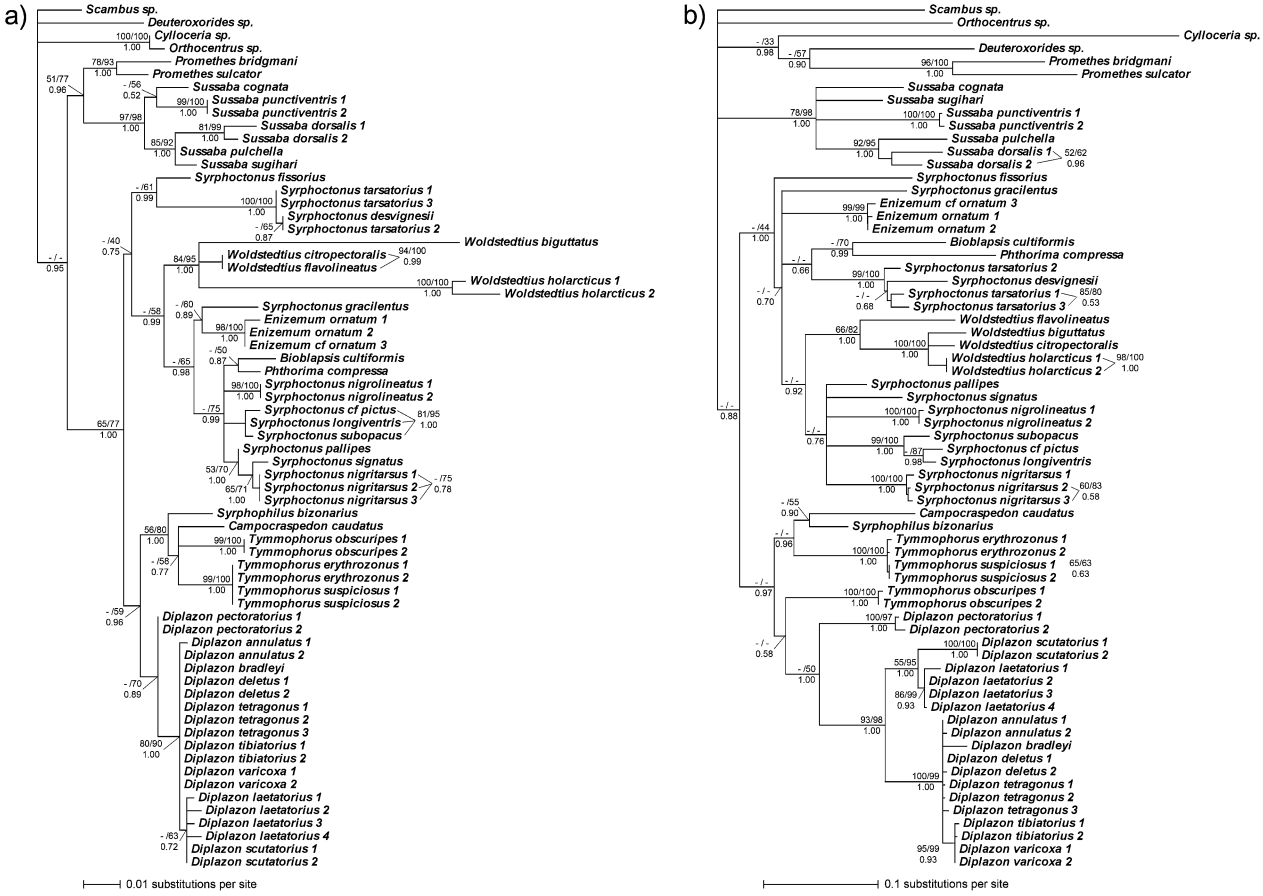

FIGURE 2. Strict consensus of the ML and Bayesian majority-rule consensus trees obtained from individual genes. Branch lengths of the consensus topologies were reestimated under ML. Numbers above branches indicate bootstrap percentages from MP and ML analyses, respectively, and numbers below branches show posterior probabilities from Bayesian reconstructions. Only values >50% are shown. a) Phylogeny obtained from 28S rRNA. b) Phylogeny obtained from CO1 mitochondrial DNA.

nodes in the combined data set, 12 concerned species and intraspecific splits inside the genus *Diplazon*. With the significance level set to 95% posterior probability, the Bayesian analysis resolved 32 (combined set) and 26 (with 28S and CO1 separately) of the ingroup nodes.

The strict consensus of the ML and Bayesian majority-rule consensus topologies of the separate and the combined analyses are shown in Figures 2 and 3, respectively, including MP bootstrap support. All analyses of the full data set consistently supported the split of the subfamily into 3 distinct genus groups: 1) a basal grouping containing the genera *Promethes* and *Sussaba*, 2) a clade comprising the genera *Syrphoctonus*, *Woldstedtius*, *Enizemum*, *Phthorima*, and the highly derived *Bioblapsis*, and 3) the genus *Diplazon* together with *Campocraspedon*, *Tymmophorus*, and *Syrphophilus*. Supports for these 3 groups were highest in the Bayesian analysis (posterior probability of all 3 clades = 1.0), intermediate in ML (63%, 73%, and 84% bootstrap support, respectively), and lowest in MP (17%, 19%, and 41% bootstrap support, respectively). These groupings are also supported by the separate data sets, with the exception of the placement of the genus *Promethes* among the outgroup taxa in the CO1 analysis (Fig. 2b).

All nominal genera that were represented by more than one species were recovered as monophyletic in the combined analysis, with the exception of *Syrphoctonus* and *Tymmophorus*. Filtering the Bayesian topologies revealed that nonmonophyly was only supported in the case of *Syrphoctonus*, for which no tree was in accordance with the hypothesis of monophyly, whereas 7.5% of the Bayesian topologies contained a monophyletic *Tymmophorus*. These results were confirmed by the SH test (*Syrphoctonus*: $P = 0.044$, *Tymmophorus*: $P = 0.591$).

PI Profiles of 28S and CO1

The PI profiles of the 2 molecular markers and their partitions are shown in Figure 4a along with the test topology with clock-enforced ML branch lengths (Fig. 4b). The profiles predict a strong superiority of CO1 compared with 28S over the whole timescale addressed in this study: At the basal node of the subfamily Diplazontinae, the PI of CO1 is still about 4 times higher than that of 28S. The general superiority in PI of CO1 compared with 28S results from the large number of variable sites at the third codon position, whereas first and second codon positions together have an estimated information content comparable to that of the stem partition of 28S. This is in accordance with previous observations of similarly heterogeneous rates of evolution at loop and stem sites of rRNA (Simon et al. 2006). The stem and loop regions of 28S show largely similar, flat profiles, but with the loop region at a lower level. In CO1, however, we find pronounced qualitative differences between the PI profiles of the 3 codon positions that reflect the different evolutionary constraints imposed on them.

The analysis of the difference in the clade support value of 28S versus CO1 at all the nodes of the total evidence consensus tree (Fig. 3), however, shows that 28S performs better on the basal ~0.7 of the diplazontine phylogeny (Fig. 4c). The difference in clade support between the 2 genes was significantly correlated to the timescale (Fig. 4c), with CO1 being superior at recent and 28S at ancient timescales ($P < 0.007$ for both Bayesian and ML support values), although the variance of clade support differences between the genes is very high, reflecting the large stochasticity of phylogenetic processes. The relative point in time at which ML bootstrap support and Bayesian posterior probability of 28S and CO1 are equal is estimated to be 0.23 and 0.28, respectively. This is considerably lower than the point of intersection of the PI values of the 2 genes, which is approximately 4 times the diplazontine tip to root distance (Fig. 7).

When comparing the PI profiles obtained for 28S and CO1 to profiles of the same gene regions obtained from different data sets (Fig. 5), we found each time a very similar picture, with CO1 reaching distinctly higher PI values than 28S. In all these studies, however, 28S was found to perform better, especially for the resolution of splits above the species level (Mardulyn and Whitfield 1999; Mengual et al. 2008; Zaldivar-Riverón et al. 2008).

Assessment of PI Profiles

The PI profiles proved very robust against the assumption of a strict or relaxed molecular clock, against the tree-building method applied, and against phylogenetic uncertainty (Fig. 6a). More precisely, constraining the branch lengths of the tree used to estimate site-specific rates according to a strict molecular clock had virtually no impact on the resulting PI profiles. The same is the case when nonparametric rate smoothing was applied to obtain an ultrametric tree (results not shown). The topologies obtained from parsimony, ML, and Bayesian reconstructions led to almost identical estimates. Similarly, the analysis of a total of 1000 topologies sampled from the Bayesian posterior distribution revealed that the PI profiles estimated from the different topologies varied only very slightly (shaded areas in Fig. 6). Figure 6b shows that the choice of evolutionary model for estimating site-specific rates, and especially the partitioning strategy, had a larger impact on estimate of site-specific rates and the resulting PI profiles. The rates estimated under the different substitution models differed most at the quickly evolving positions of CO1. A larger influence is attributable to the partitioning of the data set in genes and in codon positions for CO1. However, none of these influence the final conclusion from the PI profiles that CO1 is distinctly superior to 28S over the whole timescale studied.

Nucleotide Composition Bias

Moderate amounts of internal conflict caused by a biased nucleotide composition within and among taxa (later referred to as "biased nucleotide composition" and "nonstationary nucleotide composition,"

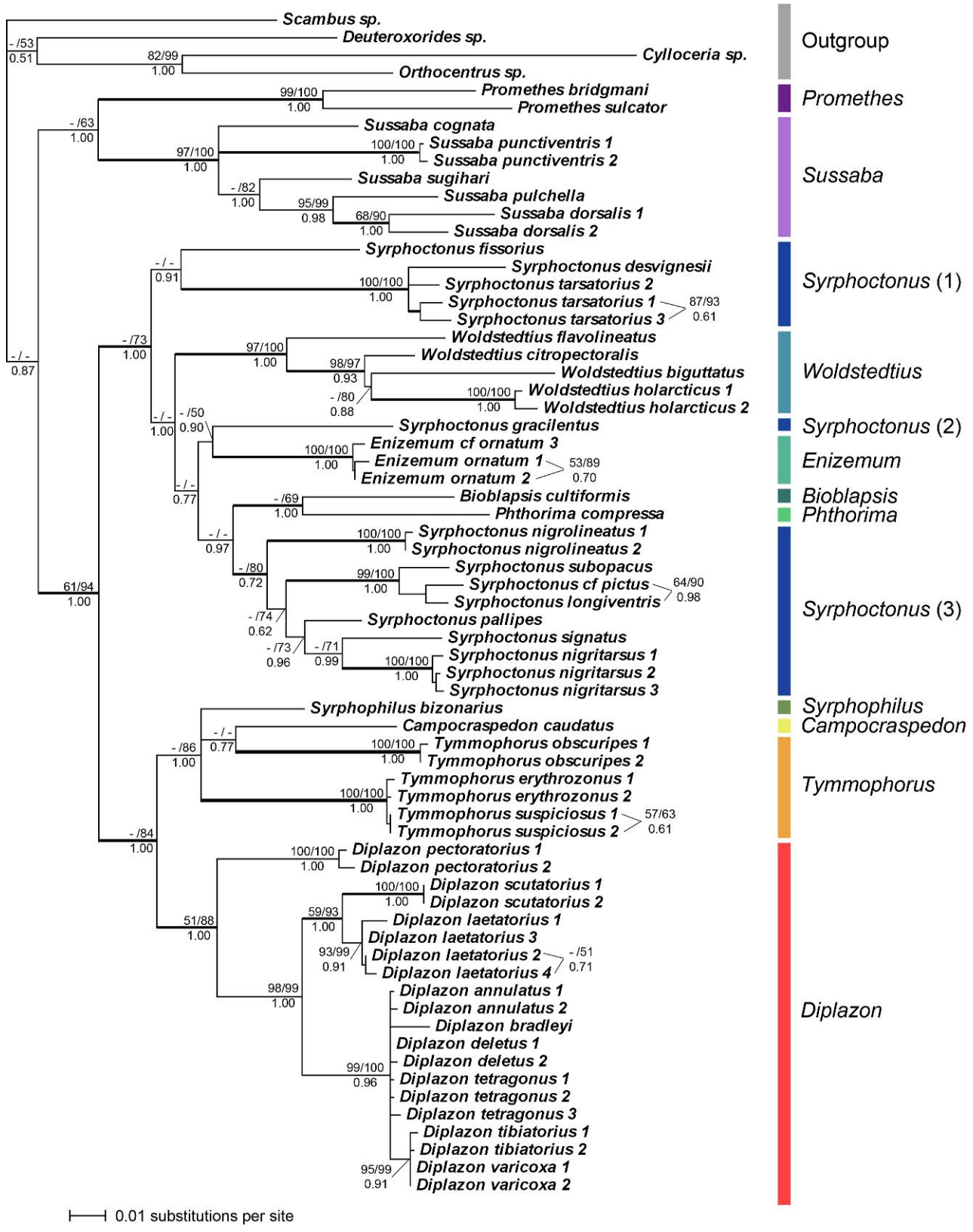

FIGURE 3. Strict consensus of the ML and Bayesian majority-rule consensus tree obtained from combined data set. Branch lengths of the consensus topology were reestimated under ML. Numbers above branches indicate bootstrap percentages from MP and ML analyses, respectively, and numbers below branches show posterior probabilities from Bayesian reconstructions. Only values >50% are shown. Bold branches represent groupings also recovered in the consensus trees of both single-gene analyses in Figure 2.

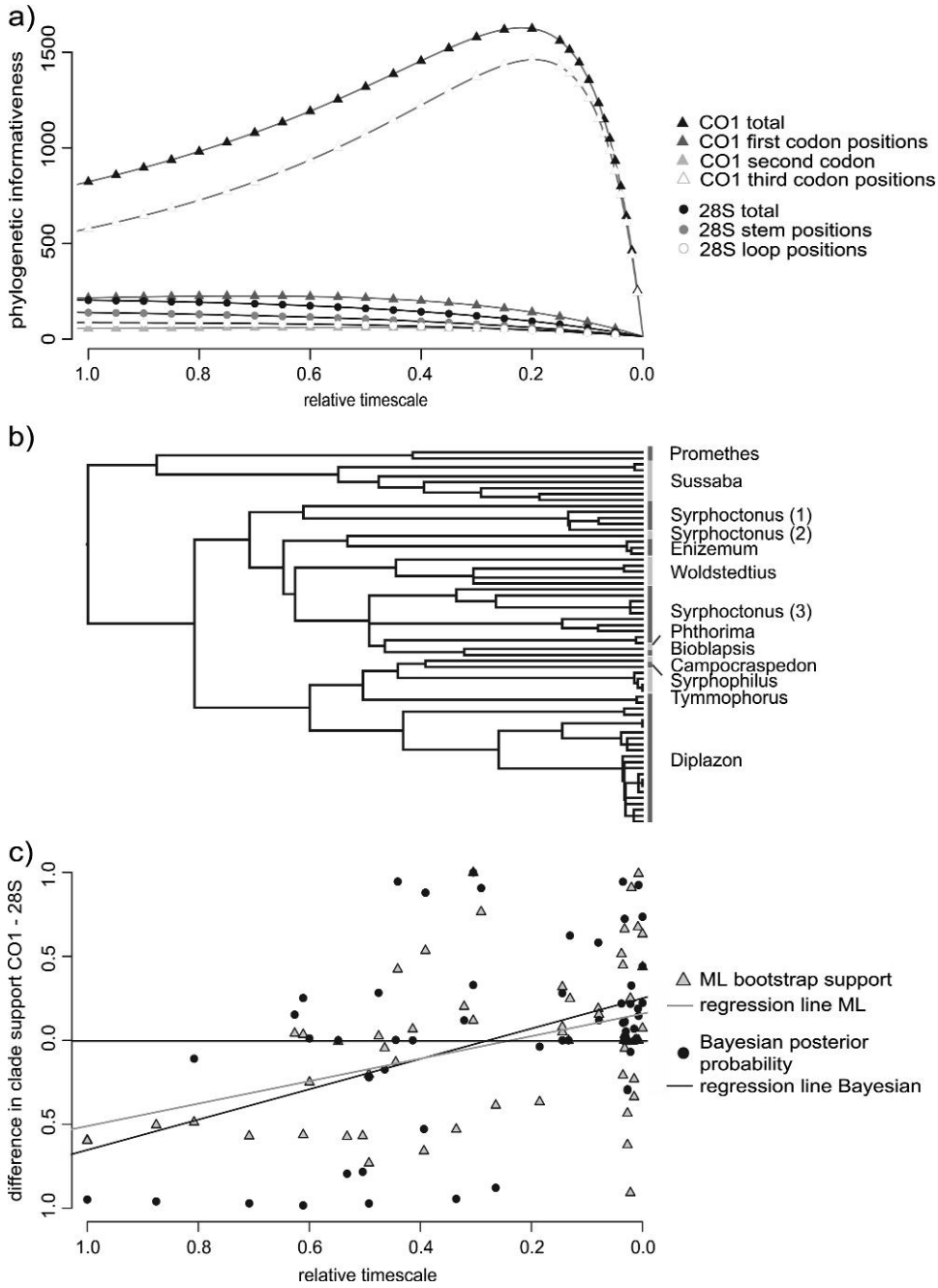

FIGURE 4. Predictive power of Townsend's PI profiles of 28S and CO1 for the phylogeny of Diplazontinae. a) PI of the 2 genes and their respective partitions over the timescale of the ingroup. b) Topology with the highest posterior probability from the Bayesian analysis with clock-enforced branch length. This phylogeny was used to calculate the PI profiles in panel (a) and the difference in clade supports shown in panel (c). Only genus names of the considered taxa are shown, with taxa that belong to one genus shown on top of gray-shaded areas. c) Differences in ML bootstrap support and Bayesian posterior probability between 28S and CO1 over time. The clade support value attained by the CO1 data set minus the support by 28S is shown for each node. Symbols above the zero line represent nodes that are more strongly supported by CO1 than by 28S and vice versa. Superimposed lines represent linear regressions of the differences in ML and Bayesian clade supports, respectively, over time.

respectively) in the CO1 data set might explain why this gene performs worse than expected from its PI profile. To explore this possibility, we needed a data set that is similar in nucleotide composition to 28S. The first and second codon positions of CO1 (abbreviated as $CO1_{12}$ below) represent such a data set, showing a very similar bias in nucleotide frequencies as 28S (63% AT content in $CO1_{12}$ and 62% GC content in 28S, see Table 1). Moreover, neither 28S nor $CO1_{12}$ shows significant nonstationarity of nucleotide compositional. Their PI profiles are similar on the timescale considered in this study, with a slight superiority of $CO1_{12}$ (Fig. 7).

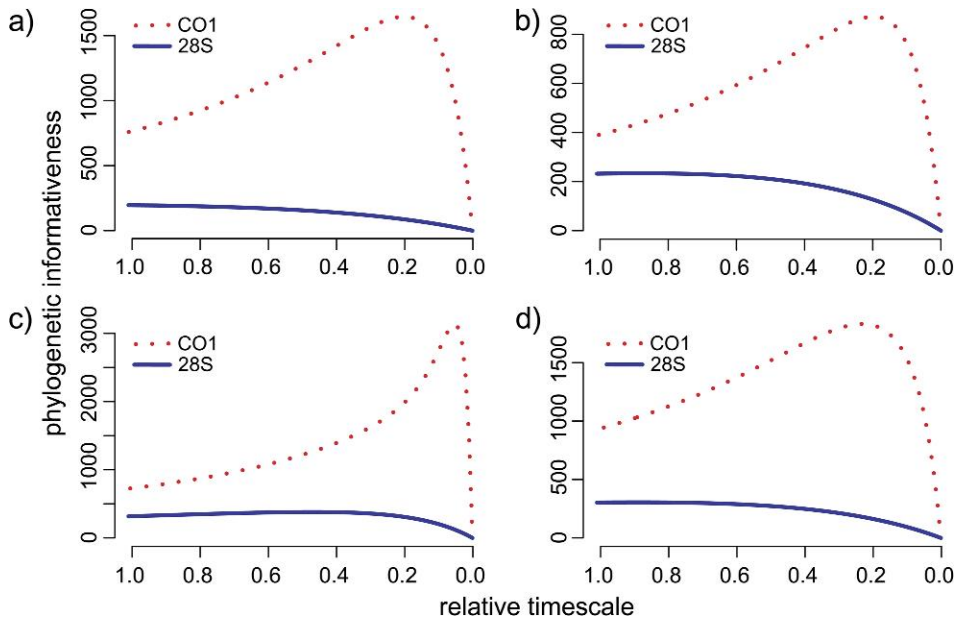

FIGURE 5. PI profiles of 28S and CO1 mitochondrial DNA as estimated from the diplazontine data set and 3 data sets from the literature. a) Phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae) from this study. b) Phylogeny of Rogadinae (Hymenoptera, Braconidae) from Zaldivar-Riverón et al. (2008). c) Phylogeny of Microgastrinae (Hymenoptera, Braconidae) from Mardulyn and Whitfield (1999). d) Phylogeny of Syrphinae (Diptera, Syrphidae) from Mengual et al. (2008).

We thus should observe a similar performance, or if one data set is superior, it should be $CO1_{12}$. However, in the ML analysis, 28S resolved 25 of the ingroup nodes, whereas $CO1_{12}$ resolved only 15 shallow nodes with a bootstrap support higher than 70%. Comparison of 28S with $CO1_{12}$ shows that there must be another explanation than nonstationary nucleotide composition. The difference between the 2 data sets becomes clear when we trace their PI profiles backward in time. The $CO1_{12}$ positions quickly decrease in informativeness, and 28S becomes superior at about 1.4 times our timescale (Fig. 7).

Simulations of Optimum Rates of Evolution

Our simulations on an idealized tree topology show that the frequency of unreversed synapomorphies strongly decreases with increasing numbers of taxa (Fig. 8). More importantly, also the optimum substitution rate decreases and reaches very low values when the number of included taxa is large. The rate that is optimal for a small number of taxa thus is not necessarily optimal for a larger data set. To complement this analysis, we rederived the Townsend's formula (2007) for a very similar setting as he originally used, but based on the 3-taxon instead of the 4-taxon case. The formula that quantifies the informativeness of a data set over time T in relation to its site-specific evolutionary rates λ_i then changes from

$$\rho(T;\lambda_1,\ldots,\lambda_n)=\sum_{i=1}^{n}16\lambda_i^2 T\,\mathrm{e}^{-4\lambda_i T}$$

(Townsend 2007, p. 225) to

$$\rho(T;\lambda_1,\ldots,\lambda_n)=\sum_{i=1}^{n}9\lambda_i^2 T\,\mathrm{e}^{-3\lambda_i T}.$$

Applying this 3-taxon case formula to our data set, we observe an even more pronounced superiority of CO1 compared with 28S. The PI value of CO1 is more than 5 times higher than that of 28S at the basal ingroup node, and the point of intersection is located at more than 5 times the root-to-tip distance (Fig. 7).

DISCUSSION

The Phylogeny of Diplazontinae and the Performance of 28S and CO1

This first phylogenetic analysis of the Diplazontinae shows that this subfamily consists of 3 highly supported monophyletic genus groups. These not only confirm some previous assumptions about generic relationships (Dasch 1964; Fitton and Rotheray 1982) but also reveal some unexpected associations, like that of *Campocraspedon* with the *Diplazon* group and *Bioblapsis* with *Phthorima*. The monophylies of most of the currently accepted genera are confirmed with high support; monophyly could only be rejected significantly in the case of *Syrphoctonus*, whereas more evidence is needed to clarify the status of *Tymmophorus*. *Syrphoctonus* has already been suspected as being polyphyletic based on its morphological variability (Dasch 1964; Wahl 1990). Based on our analyses, the genus would have to be

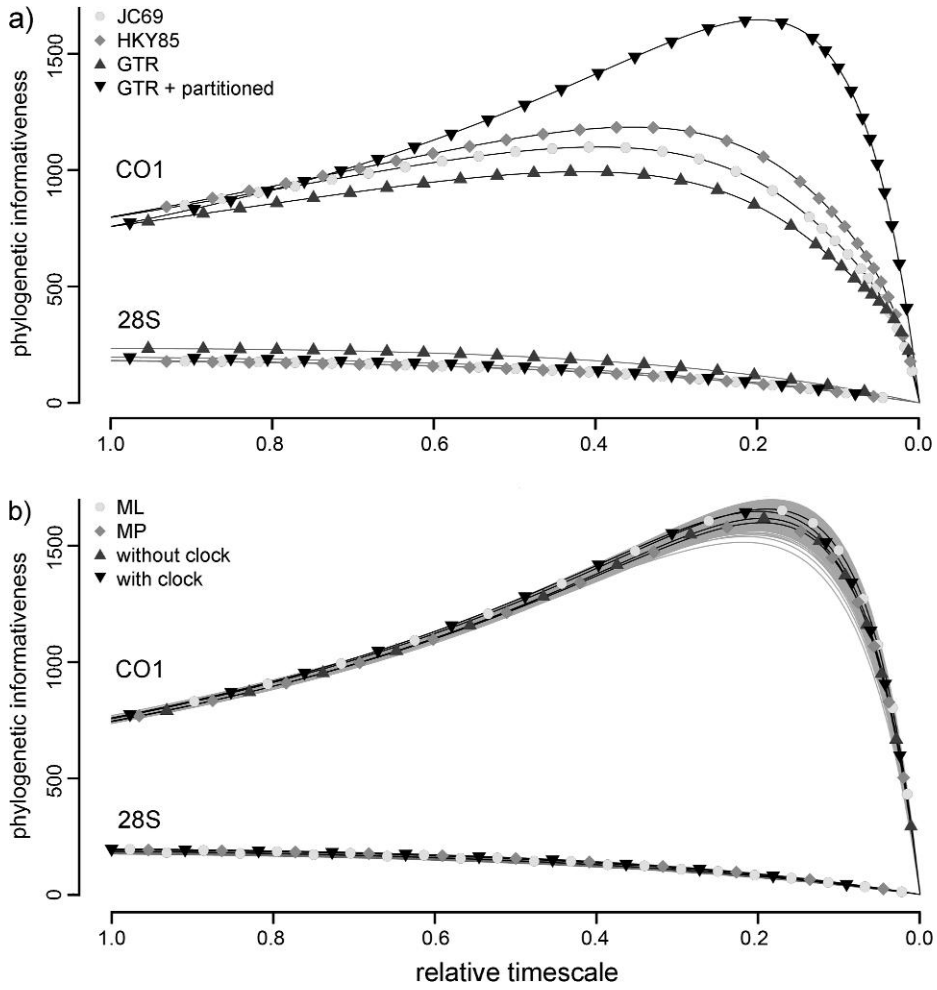

FIGURE 6. Testing the robustness of PI profiles in the case of the phylogeny of Diplazontinae. a) Influence of the evolutionary model and partitioning strategy applied for the estimation of site-specific rates of evolution on the resulting PI profiles (for details, see legend). b) Sensitivity to deviations from the molecular clock assumption, phylogenetic uncertainty, and the tree reconstruction method. The area shaded in light gray represents the CO1 profiles constructed from 1000 trees from the Bayesian distribution, the dark gray shaded area half hidden behind the symbols is the same for 28S.

split in 3 in order to restore monophyly. A denser taxon sampling of the species of *Syrphoctonus* is, however, needed before any formal taxonomic changes can be proposed.

Combined 28S and CO1 data successfully resolved a large part of the phylogeny of the Diplazontinae, with the exception of a few species groups especially in the genus *Diplazon*. CO1 resolves the branching orders around the species level well, whereas 28S also resolves many relationships at genus or higher taxonomic levels. These results are congruent with previous studies using the 2 markers to reconstruct hymenopteran relationships (e.g., Gimeno et al. 1997; Mardulyn and Whitfield 1999; Quicke, Basibuyuk, et al. 1999; Quicke, Lopez-Vaamonde, and Belshaw 1999; Dowton and Austin 2001; Banks and Whitfield 2006; Laurenne et al. 2006; Murphy et al. 2008); they, however, are in discordance with the predictions obtained from Townsend's PI profiles.

Predictive Power of PI Profiles

The PI profiles of the 2 genes predict the superiority of CO1 over the whole time frame studied here, a prediction not met by the actual clade support values. This observation could be repeated with additional data sets from hymenopteran and dipteran subfamilies (Mardulyn and Whitfield 1999; Mengual et al. 2008; Zaldivar-Riverón et al. 2008). Moreover, it is in accordance with observations by Mahon and Neigel (2008). These authors studied PI profiles of arginine kinase and CO1 sequences from 29 species of brachyurans and found that in many cases, the gene with the higher PI at a specific node was actually far less successful at resolving the split.

There are a number of possible explanations for this discrepancy. First, minor assumptions of the method are not met in our case, like the strict molecular clock and the knowledge of the true phylogeny and evolutionary model parameters for the site-rate estimation.

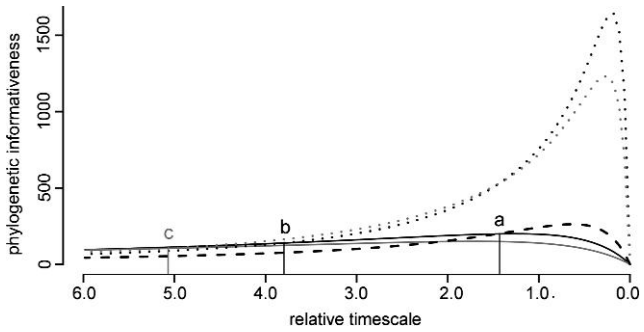

FIGURE 7. The PI profiles of CO1 and 28S are shown in black as calculated with the formula derived by Townsend from the 4-taxon case and in gray using the formula derived from 3 taxa in this study. CO1 profiles are shown as dotted lines, 28S as solid lines. Additionally, the PI profile of the combined first and second codon positions (calculated under Townsend's original formula) is shown as a black dashed line. The 3 intersection points marked by vertical lines and letters a–c denote: a) Intersection point of the PI profiles of 28S and the combined first and second codon positions of CO1. b) Intersection point of 28S with the full CO1 data set. c) Intersection point of 28S and CO1 if Townsend's formula is adapted to the 3-taxon case. The time unit used for the relative timescale corresponds to the root-to-tip distance of the Diplazontine phylogeny.

But our sensitivity analyses have shown that the PI profiles of 28S and CO1 were very robust against deviations from these assumptions. Although the choice of the evolutionary model and the data partitioning strategy had a considerable impact on the estimation of the site-specific rates, neither of these factors provides sufficient explanation for the observed differences. Second, there are attributes of genes other than evolutionary rates that can influence phylogenetic performance.

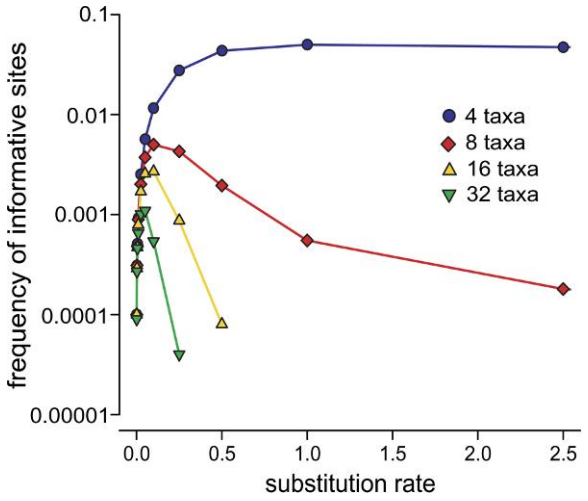

FIGURE 8. Relationship between substitution rate and the frequency of unreversed synapomorphies in the data sets simulated on the tree depicted in Figure 1. Different numbers of taxa sampled are shown with different symbols (see legend). Both the optimum rate and the overall frequency of unreversed synapomorphies decrease with increasing numbers of taxa.

Lin and Danforth (2004) explored causes for the generally lower performance of mitochondrial compared with nuclear genes, such as nucleotide composition bias and nonstationary nucleotide frequencies among taxa (Foster and Hickey 1999; Lockhart et al. 1992; Conant and Lewis 2001; Jermiin et al. 2004). Both are common to most mitochondrial genes and are pronounced at third positions of CO1 in our data set. By comparing the performance and PI profiles of 28S with those of the first and second codon positions of CO1, we have shown that for our data set, biased or nonstationary nucleotide compositions alone are not sufficient to explain the discrepancy between the predictions of the PI profiles and our observations. This result is in congruence with a simulation study by Conant and Lewis (2001) which showed that extreme differences in nucleotide composition among taxa are needed to mislead phylogenetic inference (but see Jermiin et al. 2004). The only obvious differences between 28S and the first and second codon positions of CO1 are their rate profiles, with a higher number of sites evolving at a low rate in 28S than in CO1 (Fig. 7). It thus seems that PI profiles systematically overrate the performance of quickly versus slowly evolving sites.

Optimum Rates of Evolution

Townsend (2007) implicitly based his PI method on 2 fundamental assumptions. First, results obtained from the 4-taxon case should be transformable into a quantitative measure of informativeness for larger data sets. Second, the rate that maximizes the occurrence of unreversed synapomorphies should also maximize the overall PI of a character. Using simulations, we have shown that the first assumption does not hold; increasing the number of taxa led to a decrease in the optimum rate of evolution. The way we have added taxa in our simulations only considers one side of the coin. The addition of taxa to a basal polytomy following the split in question (Fig. 1) means that the phylogenetic problem to resolve becomes much more difficult, with more taxa on long branches to be correctly placed. In real cases, however, the addition of taxa not only increases the total tree length and thus the probability of reversals but also splits up long branches that can facilitate the reconstruction of ancestral states and the estimation of evolutionary parameters (Graybeal 1998; Sullivan et al. 1999; Pollock and Bruno 2000; Hillis et al. 2003; Venditti et al. 2006). The addition of taxa in real cases thus eventually leads to a better instead of a worse performance of phylogenetic inference methods (although not necessarily to an increase of the frequency of unreversed synapomorphies, see below). But although representing an extreme case, the consequences of our result become obvious when we consider that Townsend's formula could just as well have been derived from a setting with 3 instead of 4 taxa in which case the informativeness of CO1 would have been overestimated even more strongly (Fig. 7). Previous studies have already shown

that the optimum rate of evolution changes with different tree shapes and details of the substitution process (Yang 1998; Shpak and Churchill 2000). This implies that it is impossible to predict the performance of a character at resolving any phylogenetic relationship because even if the rate of evolution of this character is constant and known without error, the optimum rate will have to be reestimated for each case again (Sanderson and Shaffer 2002). The rapid radiation of 4 taxa in the distant past that represents the basis for Townsend's (2007) PI profiles even represents a rather extreme case. It probably leads to an overrating of quickly versus slowly evolving sites in most real data sets that include more taxa and a more even branching pattern.

Regarding the assumption that the rate that maximizes unreversed synapomorphies also maximizes PI, this might not be the case in real data sets. Different nucleotide patterns (such as synapomorphies with rare reversals) can also contain valuable information, especially when analyzed with model-based inference methods such as ML or Bayesian inference (e.g., Felsenstein 1981; Simon et al. 2006). More realistic simulation studies will help solve this question. Additionally, the analyses of real data sets could help estimating the relative impact of rate profiles and other attributes like nucleotide compositional bias on the utility of different data sets, thus exploring the possible value of a measure of PI that is based on evolutionary rates alone. Townsend's method, however, despite making valuable contributions to the discussion of how to measure informativeness and identify optimum evolutionary rates, needs some reconsideration before it can directly be applied to real data sets, at least as it is currently practiced (Townsend et al. 2008; Schoch et al. 2009). A good alternative might be the method for phylogenetic experimental design developed by Goldman (1998; see also San Mauro et al. 2009) that relies on the topology and branch lengths of the tree to estimate the optimum evolutionary rate and assesses the phylogenetic performance under ML.

Funding

This work was supported by a grant from the Roche Research Foundation to S.K. and by grant NE/C519583 from the Natural Environment Research Council to D.L.J.Q.

Acknowledgments

We thank Nicolas Salamin, Reto Burri, and Joe Felsenstein for extensive discussion and Nina Laurenne for providing material for this study. We are grateful to Patrick Mardulyn, Ximo Mengual, and Alejandro Zaldivar-Riverón for providing additional data sets for this analysis. For computational assistance, we would like to acknowledge Stefan Bachofner and Daniel Wegmann. The manuscript was greatly improved thanks to comments and suggestions from Jack Sullivan, Thomas Buckley, and 2 anonymous reviewers.

References

Aliabadian M., Kaboli M., Prodon R., Nijman V., Vences M. 2007. Phylogeny of Palaearctic wheatears (genus *Oenanthe*)—congruence between morphometric and molecular data. Mol. Phylogenet. Evol. 42:665–675.

Anisimova M., Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. 55:539–552.

Banks J.C., Whitfield J.B. 2006. Dissecting the ancient rapid radiation of microgastrine wasp genera using additional nuclear genes. Mol. Phylogenet. Evol. 41:690–703.

Beirne B.P. 1941. British species of Diplazonini (Bassini auctt.) with a study of the genital and postgenital abdominal sclerites in the male. Trans. R. Entomol. Soc. Lond. 91:661–712.

Belshaw R., Fitton M.G., Herniou E., Gimeno C., Quicke D.L.J. 1998. Molecular phylogeny of the Ichneumonoidea (Hymenoptera) based on D2 expansion region of 28S rDNA. Syst. Entomol. 23:109–123.

Belshaw R., Quicke D.L.J. 1997. A molecular phylogeny of the Aphidiinae (Hymenoptera: Braconidae). Mol. Phylogenet. Evol. 7:281–293.

Bininda-Emonds O.R.P., Brady S.G., Kim J., Sanderson M.J. 2001. Scaling of accuracy in extremely large phylogenetic trees. Pac. Symp. Biocomput. 6:547–558.

Brandley M.C., Schmitz A., Reeder T.W. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54:373–390.

Bremer K. 1994. Branch support and tree stability. Cladistics. 10:295–304.

Brown J.M., Lemmon A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56:643–655.

Brown W.M., Pragera E.M., Wang A., Wilson A.C. 1982. Mitochondrial DNA sequences in primates: tempo and mode of evolution. J. Mol. Evol. 18:225–239.

Collins T.M., Fedrigo O., Naylor G.J.P. 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. Syst. Biol. 54:493–500.

Conant G.C., Lewis P.O. 2001. Effects of nucleotide composition bias on the success of the parsimony criterion in phylogenetic inference. Mol. Biol. Evol. 18:1024–1033.

Danforth B.N., Lin C.-P., Fang J. 2005. How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? Syst. Entomol. 30:549–562.

Dasch C.E. 1964. Ichneumon-flies of America north of Mexico 5: subfamily Diplazontinae. Mem. Am. Entomol. Inst. 3:1–304.

Dowton M., Austin A.D. 2001. Simultaneous analysis of 16S, 28S, COI and morphology in the Hymenoptera: Apocrita—evolutionary transitions among parasitic wasps. Biol. J. Linn. Soc. 74:87–111.

Efron B., Halloran E., Holmes S. 1996. Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. USA. 93:7085–7090.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.

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

Fischer M., Steel M.A. 2009. Sequence length bounds for resolving a deep phylogenetic divergence. J. Theor. Biol. 256:247–252.

Fitton M.G., Rotheray G.E. 1982. A key to the European genera of diplazontine ichneumon-flies, with notes on the British fauna. Syst. Entomol. 7:311–320.

Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3:294–299.

Foster P.G., Hickey D.A. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. 48:928–941.

Gauld I.D., Mound L.A. 1982. Homoplasy and the delineation of holophyletic genera in some insect groups. Syst. Entomol. 7:73–86.

Gillespie J.J., Yoder M.J., Wharton R.A. 2005. Predicted secondary structure for 28S and 18S rRNA from Ichneumonoidea (Insecta: Hymenoptera: Apocrita): impact on sequence alignment and phylogeny estimation. J. Mol. Evol. 61:114–137.

Gimeno C., Belshaw R., Quicke D.L.J. 1997. Phylogenetic relationships of the Alysiinae/Opiinae (Hymenoptera: Braconidae) and the utility of cytochrome b, 16S and 28S D2 rRNA. Insect Mol. Biol. 6:273–284.

Goldman N. 1998. Phylogenetic information and experimental design in molecular systematics. Proc. R. Soc. Lond. B. 265:1779–1786.

Goloboff P.A., Farris J.S., Nixon K.C. 2008. TNT, a free program for phylogenetic analysis. Cladistics. 24:774–786.

Grant T., Kluge A.G. 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. Cladistics. 19:379–418.

Graybeal A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Syst. Biol. 43:174–193.

Graybeal A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47:9–17.

Guindon S., Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Hebert P.D.N., Cywinska A., Ball S.L., deWaard J.R. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B. 270:313–321.

Hillis D.M., Pollock D.D., McGuire J.A., Zwickl D.J. 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52:124–126.

Huelsenbeck J.P., Bull J.J. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. Syst. Biol. 45:92–98.

Huelsenbeck J.P., Ronquist F., Nielsen R., Bollback J.P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science. 294:2310–2314.

Jermiin L., Ho S.Y.W., Ababneh F., Robinson J., Larkum A.W.D. 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst. Biol. 53:638–643.

Jian S., Soltis P.S., Gitzendanner M.A., Moore M.J., Li R., Hendry T.A., Qui Y.-L., Dhingra A., Bell C.D., Soltis D.E. 2008. Resolving an ancient, rapid radiation in Saxifragales. Syst. Biol. 57:38–57.

Källersjö M., Albert V.A., Farris J.S. 1999. Homoplasy increases phylogenetic structure. Cladistics. 15:91–93.

Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 90:773–795.

Kishino H., Miyata T., Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 30:151–160.

Kosakovsky Pond S.L., Frost S.D.W., Muse S.V. 2005c. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 21:676–679.

Laurenne N.M., Broad G.R., Quicke D.L.J. 2006. Direct optimization and multiple alignment of 28S D2-D3 rDNA sequences: problems with indels on the way to a molecular phylogeny of the cryptine ichneumon wasps (Insecta: Hymenoptera). Cladistics. 22:442–473.

Leliaert F., De Clerck O., Verbruggen H., Boedeker C., Coppejans E. 2007. Molecular phylogeny of the Siphonocladales (Chlorophyta: Cladophorophyceae). Mol. Phylogenet. Evol. 44:1237–1256.

Lin C.-P., Danforth B.N. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30:686–702.

Lockhart P.J., Howe C.J., Bryant D.A., Beanland T.J., Larkum A.W.D. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. J. Mol. Evol. 34:153–162.

Mahon B.C., Neigel J.E. 2008. Utility of arginine kinase for resolution of phylogenetic relationships among brachyuran genera and families. Mol. Phylogenet. Evol. 48:718–727.

Mardulyn P., Whitfield J.B. 1999. Phylogenetic signal in the COI, 16S, and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera; Braconidae): evidence of a high diversification rate in this group of parasitoids. Mol. Phylogenet. Evol. 12:282–294.

Mengual X., Stahls G., Rojo S. 2008. First phylogeny of predatory flower flies (Diptera, Syrphidae, Syrphinae) using mitochondrial CO1 and nuclear 28S rRNA genes: conflict and congruence with the current tribal classification. Cladistics. 24:543–562.

Meyer A. 1994. Shortcomings of the cytochrome b gene as a molecular marker. Trends Ecol. Evol. 9:278–280.

Mindell D.P., Thacker C.E. 1996. Rates of molecular evolution: phylogenetic issues and applications. Annu. Rev. Ecol. Syst. 27:279–303.

Mueller R.L. 2006. Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. Syst. Biol. 55:289–300.

Murphy N.P., Banks J.C., Whitfield J.B., Austin A.D. 2008. Phylogeny of the parasitic microgastroid subfamilies (Hymenoptera: Braconidae) based on sequence data from seven genes, with an improved time estimate of the origin of the lineage. Mol. Phylogenet. Evol. 47:378–395.

Nylander J.A.A. 2004. MrModeltest v2. Program distributed by the author. Uppsala, Sweden: Evolutionary Biology Centre, Uppsala University.

Nylander J.A.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J.L. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

Pagel M., Meade A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Syst. Biol. 53:571–581.

Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20:289–290.

Pollock D.D., Bruno W.J. 2000. Assessing an unknown evolutionary process: effect of increasing site-specific knowledge through taxon addition. Mol. Biol. Evol. 17:1854–1858.

Posada D., Buckley T.R. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53:793–808.

Posada D., Crandall K.A. 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50:580–601.

Quicke D.L.J., Basibuyuk H.H., Fitton M.G., Rasnitsyn A.P. 1999a. Morphological, palaeontological and molecular aspects of ichneumonoid phylogeny (Hymenoptera, Insecta). Zool. Scr. 28: 175–202.

Quicke D.L.J., Belshaw R. 1999. Incongruence between morphological data sets: an example from the evolution of endoparasitism among parasitic wasps (Hymenoptera: Braconidae). Syst. Biol. 48:436–454.

Quicke D.L.J., Fitton M.G., Broad G.R., Crocker B., Laurenne N.M., Miah M.I. 2005. The parasitic wasp genera *Skiapus*, *Hellwigia, Nonnus, Chriodes*, and *Klutiana* (Hymenoptera, Ichneumonidae): recognition of the Nesomesochorinae stat. rev. and Nonninae stat. nov. and transfer of *Skiapus* and *Hellwigia* to the Ophioninae. J. Nat. Hist. 39:2559–2578.

Quicke D.L.J., Fitton M.G., Notton D.G., Belshaw R., Broad G.R., Dolphin K. 2000. Phylogeny of the Ichneumonidae (Hymenoptera). A simultaneous molecular and morphological analysis. In: Austin A.D., Dowton M., editors. Hymenoptera: evolution, biodiversity and biological control. Canberra (Australia): CSIRO. p. 74–83.

Quicke D.L.J., Laurenne N.M., Fitton M.G., Broad G.R. 2009. A thousand and one wasps: a 28S rDNA and morphological phylogeny of the Ichneumonidae (Insecta: Hymenoptera) with an investigation into alignment parameter space and elision. J. Nat. Hist. 43:1305–1421.

Quicke D.L.J., Lopez-Vaamonde C., Belshaw R. 1999b. The basal Ichneumonidae (Insecta, Hymenoptera): 28S D2 rDNA considerations of the Brachycyrtinae, Labeninae, Paxylommatinae and Xoridinae. Zool. Scr. 28:203–210.

R Development Core Team. 2007. R: a language and environment for statistical computing [Internet]. Vienna (Austria): R Foundation for Statistical Computing. Available from http://www.R-project.org.

Regier J.C., Shultz J.W., Ganley A.R.D., Hussey A., Shi D., Ball B., Zwick A., Stajich J.E., Cummings M.P., Martin J.W., Cunningham C.W. 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Syst. Biol. 57:920–938.

Ronquist F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Ronquist F., Larget B., Huelsenbeck J.P., Kadane J.B., Simon D., van der Mark P. 2006. Comment on "Phylogenetic MCMC algorithms are misleading on mixtures of trees". Science. 312:367a.

Sanderson M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.

Sanderson M.J., Shaffer H.B. 2002. Troubleshooting molecular phylogenetic analyses. Annu. Rev. Ecol. Evol. Syst. 33:49–72.
San Mauro D., Gower D.J., Massingham T., Wilkinson M., Zardoya R., Cotton J.A. 2009. Experimental design in Caecilian systematics: phylogenetic information of mitochondrial genomes and nuclear rag1. Syst. Biol. 58:425–438.
Schoch C.L., Sung G.-H., López-Giráldez F., Townsend J.P., Miadlikowska J., Hofstetter V., Robbertse B., Matheny P.B., Kauff F., Wang Z., Gueidan C., Andrie R.M., Trippe K., Ciufetti L.M., Wynns A., Fraker E., Hodkinson B.P., Bonito G., Groenewald J.Z., Arzanlou M., de Hoog G.S., Crous P.W., Hewitt D., Pfister D.H., Peterson K., Gryzenhout M., Wingfield M.J., Aptroot A., Suh S.-O., Blackwell M., Hillis D.M., Griffith G.W., Castlebury L.A., Rossman A.Y., Lumbsch H.T., Lücking R., Büdel B., Rauhut A., Diederich P., Ertz D., Geiser D.M., Hosaka K., Inderbitzin P., Kohlmeyer J., Volkmann-Kohlmeyer B., Mostert L., O'Donnell K., Sipman H., Rogers J.D., Shoemaker R.A., Sugiyama J., Summerbell R.C., Untereiner W., Johnston P.R., Stenroos S., Zuccaro A., Dyer P.S., Crittenden P.D., Cole M.S., Hansen K., Trappe J.M., Yahr R., Lutzoni F., Spatafora J.W. 2009. The Ascomycota tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. Syst. Biol. 58:224–239.
Schoeniger M., von Haeseler A. 1994. A stochastic model and the evolution of autocorrelated DNA sequences. Mol. Phylogenet. Evol. 3:240–247.
Seo T.-K., Kishino H. 2008. Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. Syst. Biol. 57:367–377.
Shi M., Chen X.-X., van Achterberg C. 2005. Phylogenetic relationships among the Braconidae (Hymenoptera: Ichneumonoidea) inferred from partial 16S rDNA, 28S rDNA D2, 18S rDNA gene sequences and morphological characters. Mol. Phylogenet. Evol. 37:104–116.
Shimodaira H., Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.
Shpak M., Churchill G.A. 2000. The information content of a character under a Markov model of evolution. Mol. Phylogenet. Evol. 17:231–243.
Simon C., Buckley T.R., Frati F., Stewart J.B., Beckenbach A.T. 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu. Rev. Ecol. Syst. 37:545–579.
Sullivan J., Joyce P. 2005. Model selection in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 36:445–466.
Sullivan J., Swofford D.L., Naylor G.J.P. 1999. The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. Mol. Biol. Evol. 16:1347–1356.
Swofford D.L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
Swofford D.L., Olsen G.L., Waddell P.J., Hillis D.M. 1996. Phylogenetic inference. In: Hillis D.M., Morowitz C., Mable B.K., editors. Molecular systematics. Sunderland (MA): Sinauer. p. 407–514.
Tamura K., Dudley J., Nei M., Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24:1596–1599.
Telford M.J., Wise M.J., Gowri-Shankar V. 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the Bilateria. Mol. Biol. Evol. 22:1129–1136.
Townes H.K. 1969. The genera of Ichneumonidae, part 1. Mem. Am. Entomol. Inst. 11:1–300.
Townes H.K. 1971. The genera of Ichneumonidae, part 4. Mem. Am. Entomol. Inst. 17:1–372.
Townsend J.P. 2007. Profiling phylogenetic informativeness. Syst. Biol. 56:222–231.
Townsend J.P., López-Giráldez F., Friedman F. 2008. The phylogenetic informativeness of nucleotide and amino acid sequences for reconstructing the vertebrate tree. J. Mol. Evol. 67:437–447.
Venditti C., Meade A., Pagel M. 2006. Detecting the node-density artifact in phylogeny reconstruction. Syst. Biol. 55:637–643.
Waegele J.W., Mayer C. 2007. Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. BMC Evol. Biol. 7:147.
Wahl D.B. 1990. A review of the mature larvae of Diplazontinae, with notes on larvae of Acaenitinae and Orthocentrinae and proposal of two new subfamilies (Insecta: Hymenoptera, Ichneumonidae). J. Nat. Hist. 24:27–52.
Wahl D.B., Gauld I.D. 1998. The cladistics and higher classification of the Pimpliformes (Hymenoptera : Ichneumonidae). Syst. Entomol. 23:265–298.
Wenzel J.W., Siddall M.E. 1999. Noise. Cladistics. 15:51–64.
Whitfield J.B., Mardulyn P., Austin A.D., Dowton M. 2002. Phylogenetic relationships among microgastrine braconid wasp genera based on data from the 16S, COI and 28S genes and morphology. Syst. Entomol. 27:337–359.
Wiens J.J., Kuczynski C.A., Smith S.A., Mulcahy D.G., Sites J.W.J., Townsend T.M., Reeder T.W. 2008. Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Syst. Biol. 57:420–431.
Yang Z. 1998. On the best evolutionary rate for phylogenetic analysis. Syst. Biol. 47:125–133.
Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.
Yu D.S., Horstmann K. 2005. World Ichneumonoidea 2004—taxonomy, biology, morphology and distribution [Internet]. Vancouver, Canada. Available from: www.taxapad.com.
Zaldivar-Riverón A., Belokobylskij S.A., León-Regagnon V., Briceño-G. R., Quicke D.L.J. 2008. Molecular phylogeny and historical biogeography of the cosmopolitan parasitic wasp subfamily Doryctinae (Hymenoptera:Braconidae). Invertebr. Syst. 22:345–363.
Zaldivar-Riverón A., Mori M., Quicke D.L.J. 2006. Systematics of the cyclostome subfamilies of braconid parasitic wasps (Hymenoptera: Ichneumonoidea): a simultaneous molecular and morphological Bayesian approach. Mol. Phylogenet. Evol. 38:130–145.

APPENDIX 1. Diplazontinae and outgroup taxa, their provenance, and GenBank accession numbers

Taxon	Internal code	Country/department/locality/collection date	GenBank accession number 28S	CO1
Diplazontinae				
Bioblapsis cultiformis	SK_2D12	Switzerland/Graubünden/Sur, Alp Flix/05.08.2003	FJ556489	FJ556422
Campocraspedon caudatus	SK_1C3	Switzerland/Graubünden/Sur, Alp Flix/20.06.2003	FJ556490	FJ556423
Diplazon annulatus 1	SK_1A1	Switzerland/Graubünden/Sur, Alp Flix/17.07.2006	FJ556491	FJ556424
D. annulatus 2	SK_1A2	Switzerland/Graubünden/Sur, Alp Flix/17.07.2006	FJ556492	FJ556425
Diplazon bradleyi	SK_2A5	USA/Alaska/Fairbanks, North Star Borough/22.06.2006	FJ556493	FJ556426
Diplazon deletus 1	SK_1A7	Switzerland/Graubünden/Sur, Alp Flix/27.06.2003	FJ556494	FJ556427
D. deletus 2	SK_1A8	Switzerland/Graubünden/Sur, Alp Flix/27.06.2003	FJ556495	FJ556428
Diplazon laetatorius 1	SK_1C4	Switzerland/Graubünden/Sur, Alp Flix/13.09.2006	FJ556496	FJ556429
D. laetatorius 2	SK_1C5	USA/Maryland/Queen Anne County/Centreville/July 2006	FJ556497	FJ556430
D. laetatorius 3	SK_3A2	Spain/Canary Islands/La Gomera, Virgen las Nieves/2006	FJ556498	FJ556431
D. laetatorius 4	SK_3B2	Zambia/Southern Province/Choma/17.05.2006	FJ556499	FJ556432
Diplazon pectoratorius 1	SK_1C6	Switzerland/Graubünden/Sur, Alp Flix/17.07.2003	FJ556500	FJ556433
D. pectoratorius 2	SK_3H2	Finland/Southern Finland/Sipoo, Sipoonkorpi/19.06.2006	FJ556501	FJ556434
Diplazon scutatorius 1	SK_1A11	Finland/Southern Finland/Sipoo, Sipoonkorpi/19.06.2006	FJ556502	FJ556435
D. scutatorius 2	SK_2E3	Switzerland/Bern/Bremgartenwald/25.09.2006	FJ556503	FJ556436
Diplazon tetragonus 1	SK_1B1	Finland/Southern Finland/Sipoo, Sipoonkorpi/07.07.2006	FJ556504	FJ556437
D. tetragonus 2	SK_3E2	Russia/Pskov Province/Sebezhsky/16.07.2006	FJ556505	FJ556438
D. tetragonus 3	SK_3F2	Finland/Southern Finland/Sipoo, Hindsby/21.06.2006	FJ556506	FJ556439
Diplazon tibiatorius 1	SK_1B2	Switzerland/Graubünden/Sur, Alp Flix/23.07.2006	FJ556507	FJ556440
D. tibiatorius 2	SK_1B3	Russia/Primorsk flood plains/2006	FJ556508	FJ556441
Diplazon varicoxa 1	SK_1B4	Switzerland/Graubünden/Sur, Alp Flix/13.09.2006	FJ556509	FJ556442
D. varicoxa 2	SK_3F1	Switzerland/Graubünden/Sur, Alp Flix/11.07.2006	FJ556510	FJ556443
Enizemum ornatum 1	SK_1B5	Switzerland/Graubünden/Sur, Alp Flix/04.07.2003	FJ556511	FJ556444
E. ornatum 2	SK_1B6	Turkey/Bolu/Bolu/06.1999	FJ556512	FJ556445
Enizemum cf. *ornatum* 3	SK_1C8	USA/Alaska/Fairbanks, North Star Borough/19.06.2006	FJ556513	FJ556446
Phthorima compressa	SK_1C9	Switzerland/Graubünden/Sur, Alp Flix/22.07.2006	FJ556514	FJ556447
Promethes bridgmani	SK_1C11	Finland/Southern Finland/Sipoo, Sipoonkorpi/07.07.2006	FJ556515	FJ556448
Promethes sulcator	SK_1C12	Finland/Southern Finland/Sipoo, Hindsby/09.06.2005	FJ556516	
P. sulcator	SK_1D1	Switzerland/Solothurn/Trimbach, Miesernbach/10.06.2002		FJ556449
Sussaba cognata	SK_1D5	Finland/Southern Finland/Sipoo, Sipoonkorpi/07.07.2006	FJ556517	FJ556450
Sussaba dorsalis 1	SK_1D6	Switzerland/Graubünden/Sur, Alp Flix/20.06.2003	FJ556518	FJ556451
S. dorsalis 2	SK_2B1	USA/Alaska/Fairbanks, North Star Borough/19.06.2006	FJ556519	FJ556452
Sussaba pulchella	SK_1D11	Finland/Southern Finland/Sipoo, Hindsby/22.06.2005	FJ556520	FJ556453
Sussaba punctiventris 1	SK_1E1	Switzerland/Graubünden/Sur, Alp Flix/28.07.2003	FJ556521	FJ556454
S. punctiventris 2	SK_1E2	Finland/Southern Finland/Sipoo, Hindsby/09.06.2005	FJ556522	FJ556455
Sussaba sugiharai	SK_1E3	Taiwan/Nanton Ren-ai, Ruligenghih wildlife res./07.2005	FJ556523	FJ556456
Syrphoctonus desvignesii	SK_1E10	Switzerland/Graubünden/Sur, Alp Flix/18.07.2006	FJ556524	FJ556457
Syrphoctonus fissorius	SK_1E11	Switzerland/Bern/Bremgartenwald/25.09.2006	FJ556525	FJ556458
Syrphoctonus gracilentus	SK_1E12	Switzerland/Graubünden/Sur, Alp Flix/25.07.2003	FJ556526	FJ556459
Syrphoctonus longiventris	SK_1G5	Switzerland/Solothurn/Trimbach, Miesernbach/10.06.2002	FJ556527	FJ556460
Syrphoctonus nigritarsus 1	SK_1F2	Switzerland/Graubünden/Sur, Alp Flix/24.06.2003	FJ556528	FJ556461
S. nigritarsus 2	SK_1F3	Switzerland/Graubünden/Sur, Alp Flix/25.07.2003	FJ556529	FJ556462
S. nigritarsus 3	SK_1F4	Switzerland/Graubünden/Sur, Alp Flix/13.06.2003	FJ556530	FJ556463
Syrphoctonus nigrolineatus 1	SK_1F6	Switzerland/Graubünden/Sur, Alp Flix/11.07.2003	FJ556531	FJ556464
S. nigrolineatus 2	SK_1G1	Switzerland/Graubünden/Sur, Alp Flix/18.07.2003	FJ556532	FJ556465
Syrphoctonus pallipes	SK_1G3	Finland/Oulu/Muhos/12.08.2005	FJ556533	FJ556466
Syrphoctonus cf. *pictus*	SK_1F1	Switzerland/Graubünden/Sur, Alp Flix/20.06.2003	FJ556534	FJ556467
Syrphoctonus signatus	SK_1G7	USA/Alaska/Fairbanks, North Star Borough/22.06.2006	FJ556535	FJ556468
Syrphoctonus subopacus	SK_2B9	Finland/Southern Finland/Sipoo, Sipoonkorpi/27.06.2006	FJ556536	FJ556469
Syrphoctonus tarsatorius 1	SK_1G8	Switzerland/Graubünden/Sur, Alp Flix/22.08.2003	FJ556537	FJ556470
S. tarsatorius 2	SK_1G9	Finland/Southern Finland/Sipoo, Sipoonkorpi/27.06.2006	FJ556538	FJ556471
S. tarsatorius 3	SK_1H9	England/Berkshire/Ascot, Silwood Park Campus/2002	FJ556539	FJ556472
Syrphophilus bizonarius	SK_1G10	Switzerland/Graubünden/Sur, Alp Flix/12.06.2003	FJ556540	FJ556473
Tymmophorus erythrozonus 1	SK_1B10	Switzerland/Graubünden/Sur, Alp Flix/03.07.2003	FJ556541	FJ556474
T. erythrozonus 2	SK_1B11	Switzerland/Graubünden/Sur, Alp Flix/23.08.2006	FJ556542	FJ556475
Tymmophorus obscuripes 1	SK_1H2	England/Berkshire/Ascot, Silwood Park Campus/2002	FJ556543	FJ556476
T. obscuripes 2	SK_1H3	Switzerland/Graubünden/Sur, Alp Flix31.07.2003	FJ556544	FJ556477
Tymmophorus suspiciosus 1	SK_2B11	Finland/Oulu/Muhos/12.08.2005	FJ556545	FJ556478
T. suspiciosus 2	SK_3F3	Switzerland/Solothurn/Trimbach, Miesernbach/10.06.2002	FJ556546	FJ556479
Woldstedtius biguttatus	SK_1H5	Switzerland/Graubünden/Sur, Alp Flix/12.09.2006	FJ556547	FJ556480
Woldstedtius citropectoralis	SK_2C4	Finland/Oulu/Muhos/12.08.2005	FJ556548	FJ556481
Woldstedtius flavolineatus	SK_2C1	England/Berkshire/Ascot, Silwood Park Campus/2002	FJ556549	
W. flavolineatus	SK_1H8	Switzerland/Graubünden/Sur, Alp Flix/26.06.2003		FJ556482
Woldstedtius holarcticus 1	SK_1H10	Switzerland/Bern/Bremgartenwald/25.09.2006	FJ556550	FJ556483
W. holarcticus 2	SK_1H11	Finland/Southern Finland/Sipoo, Sipoonkorpi/27.06.2006	FJ556551	FJ556484
Outgroups				
Cylloceria sp.	SK_1F7	Finland/Oulu/Muhos/12.08.2005	FJ556552	FJ556485
Deuteroxorides sp.	SK_3H3	Finland/Southern Finland/Sipoo, Sipoonkorpi/27.06.2006	FJ556553	FJ556486
Orthocentrus sp.	SK_1F10	Finland/Oulu/Muhos/12.08.2005	FJ556554	FJ556487
Scambus sp.	SK_1F9	Finland/Oulu/Muhos/12.08.2005	FJ556555	FJ556488