Systematic Biology Advance Access published April 17, 2012

DOI:10.1093/sysbio/sys028

Experimental Design in Phylogenetics: Testing Predictions from Expected Information

DIEGO SAN MAURO[1,5,*], DAVID J. GOWER[1], JAMES A. COTTON[2,6], RAFAEL ZARDOYA[3], MARK WILKINSON[1], AND TIM MASSINGHAM[4]

[1]*Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK;* [2]*School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK;* [3]*Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales—CSIC, José Gutiérrez Abascal 2, 28006 Madrid, Spain;* [4]*EMBL—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK;* [5]*Present address: Department of Animal Biology, University of Barcelona, Avenida Diagonal 643, 08028 Barcelona, Spain; and* [6]*Present address: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK*

**Correspondence to be sent to: Department of Animal Biology, University of Barcelona, Avenida Diagonal 643, 08028 Barcelona, Spain; E-mail: dsanmauro@ub.edu.*

Received 13 July 2011; reviews returned 20 December 2011; accepted 6 February 2012
Associate Editor: Tiffani Williams

Abstract.—Taxon and character sampling are central to phylogenetic experimental design; yet, we lack general rules. Goldman introduced a method to construct efficient sampling designs in phylogenetics, based on the calculation of expected Fisher information given a probabilistic model of sequence evolution. The considerable potential of this approach remains largely unexplored. In an earlier study, we applied Goldman's method to a problem in the phylogenetics of caecilian amphibians and made an a priori evaluation and testable predictions of which taxon additions would increase information about a particular weakly supported branch of the caecilian phylogeny by the greatest amount. We have now gathered mitogenomic and *rag1* sequences (some newly determined for this study) from additional caecilian species and studied how information (both expected and observed) and bootstrap support vary as each new taxon is individually added to our previous data set. This provides the first empirical test of specific predictions made using Goldman's method for phylogenetic experimental design. Our results empirically validate the top 3 (more intuitive) taxon addition predictions made in our previous study, but only information results validate unambiguously the 4th (less intuitive) prediction. This highlights a complex relationship between information and support, reflecting that each measures different things: Information is related to the ability to estimate branch length accurately and support to the ability to estimate the tree topology accurately. Thus, an increase in information may be correlated with but does not necessitate an increase in support. Our results also provide the first empirical validation of the widely held intuition that additional taxa that join the tree proximal to poorly supported internal branches are more informative and enhance support more than additional taxa that join the tree more distally. Our work supports the view that adding more data for a single (well chosen) taxon may increase phylogenetic resolution and support in weakly supported parts of the tree without adding more characters/genes. Altogether our results corroborate that, although still underexplored, Goldman's method offers a powerful tool for experimental design in molecular phylogenetic studies. However, there are still several drawbacks to overcome, and further assessment of the method is needed in order to make it better understood, more accessible, and able to assess the addition of multiple taxa. [Bootstrap support; branch lengths; caecilians; experimental design; Gymnophiona; mitochondrial genome; phylogenetic information; *rag1*; taxon sampling.]

Basic experimental design in molecular phylogenetics encompasses primarily the choice of markers (genes, genomic regions, and characters) (Lopez-Giraldez and Townsend 2011). Many studies acknowledge that taxon and character sampling are fundamental (e.g., Graybeal 1998; Pollock et al. 2002; Hillis et al. 2003; Cummings and Meyer 2005; Rokas and Carroll 2005; San Mauro and Agorreta 2010) and that data set completeness is desirable (Cummings and Meyer 2005). However, phylogenetics lacks general rules or agreed strategies for improving the accuracy of, and confidence in, inferred phylogenetic hypotheses through judicious and efficient experimental design. Goldman (Goldman 1998; Massingham and Goldman 2000) proposed a general method for constructing efficient experimental (sampling) designs in phylogenetics based on the calculation of expected Fisher information from a tree assuming a probabilistic model of sequence evolution. Despite its considerable potential for molecular phylogenetics, Goldman's method has rarely been applied to real phylogenetic problems (Goldman 1998; Geuten et al. 2007; San Mauro et al. 2009). Henceforth, we refer to Fisher information simply as "information."

In an earlier study (San Mauro et al. 2009), we applied Goldman's method to the phylogeny of caecilian amphibians (Gymnophiona). We showed that the branch resolving the phylogenetic relationships among Scolecomorphidae, Herpelidae, and all other teresomatans (here dubbed the "controversial branch") received low bootstrap support (54%) from the available mitochondrial (mt) genome and *rag1* data, implying ambiguity about the exact order of branching. Using Goldman's approach, we identified the places to which a single taxon could be added to the caecilian tree in order to maximize the expected information for the controversial branch. A strong correlation was found between the decrease in expected information and the distance between the controversial branch and where each hypothetical new taxon was added. Combining these results with background knowledge of caecilian diversity and phylogeny allowed us to make predictions as to the best candidate extant caecilian taxa to be targeted in future studies. These included intuitive predictions of taxa arising from deep splits within Scolecomorphidae and/or Herpelidae proximal to the controversial branch and a less intuitive prediction of a less proximal taxon

arising from a deep split within Rhinatrematidae (San Mauro et al. 2009). These former (intuitive) predictions were hypothesized to include unsampled species of *Boulengerula* and *Scolecomorphus* as well as species of their respective sister taxa, *Herpele* and *Crotaphatrema*, and the latter included species of the rhinatrematid *Epicrionops*.

Here, we test these predictions through phylogenetic analyses incorporating new data from additional caecilian species, determining how information changes as new taxa are added to the original data, and how these additions affect nonparametric bootstrap support for the controversial branch and for the whole tree. This study thus serves as an empirical test of predictions made in San Mauro et al. (2009), provides the first empirical test of any predictions derived from the application of Goldman's method, and helps improve understanding of the relationship between information and support.

Materials and Methods

Taxon Sampling and DNA Sequencing

Our previous study (San Mauro et al. 2009) included 9 species of caecilian amphibians, representing all 6 families recognized at that time (Wilkinson and Nussbaum 2006) and 8 of the 9 currently recognized families (Wilkinson et al. 2011). For this study, we expanded the original taxon sampling with mt genome and *rag1* sequence data for 14 additional species of caecilian amphibians (Table 1). The new taxon sampling includes representatives of all 9 currently recognized families (Wilkinson et al. 2011). We added one complete caecilian mt genome sequence reported by Zhang et al. (2005) plus the 4 partial and 7 complete caecilian mt genome sequences reported by Zhang and Wake (2009). San Mauro et al. (2009) indicated that caecilians with the greatest chance of increasing information for the controversial branch were those that would join the branch subtended by *Scolecomorphus vittatus* followed by those subtended by *Boulengerula taitanus* and then *Rhinatrema bivittatum*. The study of Zhang and Wake (2009) included 2 such species (*B. boulengeri* and *Epicrionops niger*). Additionally, the complete mt genomes of *Crotaphatrema lamottei* and *Herpele squalostoma* were newly determined for this study, the former from specimens recently collected in dedicated fieldwork (Doherty-Bone et al. 2011). Based on background taxonomic knowledge and previous phylogenetic work (e.g., Taylor 1968; Nussbaum 1985; Nussbaum and Wilkinson 1989; Wilkinson et al. 2003, 2011; Wilkinson and Nussbaum 2006), these latter 2 species were predicted to join the *Scolecomorphus* and *Boulengerula* branches, respectively (San Mauro et al. 2009), and these expectations have received additional support from recent molecular studies (Frost et al. 2006; Roelants et al. 2007; Doherty-Bone et al. 2011; Gower et al. 2011; Loader et al. 2011; Pyron and Wiens 2011). For each of the 14 species with new (not included in our 2009 analyses) mt genome data available, we determined a 1509 bp long fragment of nuclear *rag1* (Table 1).

In all cases, total DNA was purified from ethanol-preserved liver using the QIAamp DNA Mini Kit (QIAGEN) following manufacturer's instructions. Nucleotide sequences of the mt genomes of *C. lamottei* and *H. squalostoma* and of nuclear *rag1* genes were determined using the primers and polymerase chain reaction (PCR) cycling conditions reported by San Mauro et al. (2004, 2005). In all cases, PCR products were purified with Millipore purification plates and cycle sequenced in an automated DNA sequencer (ABI 3730xl DNA Analyzer) using the BigDye Terminator v1.1 Cycle sequencing kit (Applied Biosystems) and the corresponding PCR primers. The obtained sequences averaged 700 bp in length, and each sequence overlapped with the next contig by about 150 bp. No sequence differences within the overlapping regions were observed. GenBank accession numbers of newly determined DNA sequences and voucher specimens are given in Table 1. Distinct structural features of the mt genome of *C. lamottei* are presented in the Supplementary Material (available at http://dx.doi.org, doi:10.5061/dryad.83p1130j). The mt genome of *H. squalostoma* conforms to the vertebrate consensus mt gene arrangement (Lupi et al. 2010) and possesses no distinct structural features.

Sequence Alignments, Phylogeny Reconstruction, and Support

Alignments were prepared separately for each mt gene and nuclear *rag1*. Nucleotide sequences of mt ribosomal genes were aligned using MAFFT version 6.850 (Katoh et al. 2002; Katoh and Toh 2008) considering secondary structure of RNA and revised by eye to correct obvious misalignments. Sequences of each mt tRNA gene were aligned manually based on inferred cloverleaf secondary structures and concatenated to form a single partition. Some mt tRNA genes were either absent or not available for a few taxa and these were coded as missing data. Mitochondrial protein-coding genes were aligned with TranslatorX (Abascal et al. 2010) using MAFFT to compute the protein alignments. In all cases, gaps and alignment ambiguities were excluded from partitions using GBlocks version 0.91b (Castresana 2000) with default parameter settings. This is automated in TranslatorX, so that nucleotide alignments are produced after removal of ambiguously aligned amino acid positions. As in San Mauro et al. (2009), third codon positions of mt protein-coding genes were excluded from the alignments because transitions were judged to be saturated when plots of pairwise uncorrected (transition and transversion) differences versus corrected sequence divergence (measured as maximum-likelihood [ML] distance) were considered (not shown). Nucleotide sequences of nuclear *rag1* were aligned manually against the data of San Mauro et al. (2009), and no positions were excluded. These single-gene alignments were combined into a single master

TABLE 1. Caecilian samples employed in this study

Species	Family	Voucher number	Collection locality	GenBank accession numbers	
				mt genome	*rag1*
Boulengerula boulengeri	Herpelidae	CAS 168822	Lushoto, Tanzania	GQ244464	
		BMNH 2002.95	Amani, Tanzania		**HQ444127**
B. taitanus	Herpelidae	NMK A/3112/1	Wundanyi, Kenya	AY954504	DQ320062
Caecilia volcani	Caeciliidae	MVZ 231242	Fortuna, Panama	GQ244466	**HQ444128**
Crotaphatrema lamottei	Scolecomorphidae	BMNH 2008.274	Mount Oku, Cameroon	**JN089398**	**JN089397**
Dermophis mexicanus	Dermophiidae	MVZ 179061	Finca Santa Julia, Guatemala	GQ244467	
		No voucher	Finca El Faro, Guatemala		**HQ444129**
Epicrionops niger	Rhinatrematidae	MVZ 258040	Near Mount Roraima, Guyana	GQ244468	
		ROM 39682	Mount Ayanganna, Guyana		**HQ444130**
Gegeneophis ramaswamii	Indotyphlidae	MW 331	Thenmalai, India	AY456250	AY456255
Geotrypetes seraphini	Dermophiidae	BMNH 2005.2	Cameroon	AY954505	
		FMNH 256782	Gabon		DQ320063
Grandisonia alternans	Indotyphlidae	MVZ 258026	La Digue, Seychelles	GQ244470	
		UMMZ 240022	Silhouette, Seychelles		**HQ444131**
Gymnopis multiplicata	Dermophiidae	MVZ 171331	Tortuguero, Costa Rica	GQ244471	
		MVZ 203936	Refugio Nacional Tapanti, Costa Rica		**HQ444132**
Herpele squalostoma	Herpelidae	BMNH 2002.97	Cameroon	**HQ456774**	**HQ444133**
Hypogeophis rostratus	Indotyphlidae	MVZ 258025	La Digue, Seychelles	GQ244472	
		UMMZ 240025	Silhouette, Seychelles		**HQ444134**
Ichthyophis bannanicus	Ichthyophiidae	No voucher	Beiliu, China	AY458594	
		VUB 698	Tam Dao, Vietnam		**HQ444135**
I. glutinosus	Ichthyophiidae	MW 1733	Peradeniya, Sri Lanka	AY456251	AY456256
Microcaecilia sp.	Siphonopidae	IWK 0128	Iwokrama, Guyana	GQ244473	
		UMMZ 214081	Iwokrama, Guyana		**HQ444136**
Oscaecilia ochrocephala	Caeciliidae	MVZ 222472	Santa Clara de Arajan, Panama	GQ244474	**HQ444137**
Praslinia cooperi	Indotyphlidae	UMMZ 192934	Silhouette, Seychelles	GQ244475	**HQ444138**
Rhinatrema bivittatum	Rhinatrematidae	BMNH 2002.6	Kaw, French Guiana	AY456252	AY456257
Schistometopum thomense	Dermophiidae	CAS 219292	São Tomé	GQ244476	
		UMMZ 214092	São Tomé		**HQ444139**
Scolecomorphus vittatus	Scolecomorphidae	BMNH 2002.100	Amani, Tanzania	AY456253	AY456258
Siphonops annulatus	Siphonopidae	BMNH 2005.9	Dominguez Martins, Brazil	AY954506	DQ320064
Typhlonectes natans	Typhlonectidae	No voucher	No collection locality details	AF154051	
		BMNH 2000.218	Potreito, Venezuela		AY456260
Uraeotyphlus cf. *oxyurus*	Ichthyophiidae	MW 212	Payyanur, India	AY456254	AY456259

Notes: GenBank accession numbers for new sequences indicated in bold type. BMNH, Natural History Museum, London, UK; CAS, California Academy of Sciences, San Francisco, USA; MW, University of Kerala, India and National Museum, Colombo, Sri Lanka; MVZ, Museum of Vertebrate Zoology, Berkeley, USA; NMK, National Museum of Kenya, Nairobi, Kenya; ROM, Royal Ontario Museum, Toronto, Canada; UMMZ, University of Michigan Museum of Zoology, Ann Arbor, USA; VUB, Vrije Universiteit Brussels, Brussels, Belgium.

alignment that has been deposited in TreeBASE under accession number S11625.

The main caecilian phylogeny (including all 23 caecilian species) was estimated from the master alignment. As in San Mauro et al. (2009), rooted trees assume the Rhinatrematidae to be the sister group of all other caecilians based on extensive molecular (Hedges et al. 1993; San Mauro et al. 2004, 2005; Frost et al. 2006; Roelants et al. 2007; Zhang and Wake 2009; San Mauro 2010; Pyron and Wiens 2011) and morphological (Nussbaum 1977, 1979; Wilkinson 1992, 1996, 1997; Wilkinson and Nussbaum 1996) evidence. Phylogeny was estimated using ML (Felsenstein 1981) and Bayesian inference (BI; Huelsenbeck et al. 2001). ML analysis was performed with RAxML version 7.2.6 (Stamatakis 2006) using the rapid hill climbing algorithm (Stamatakis et al. 2007) and computing 100 distinct ML trees starting from 100 distinct randomized (random stepwise addition of taxa) maximum-parsimony starting trees. BI was performed with MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) running 4 simultaneous Markov chains for 20 million generations, sampling every 2000 generations and discarding the first 1 million samples in order to reduce dependence on the initial starting point ("burn-in period"). Adequate convergence of the BI runs was judged by plots of ln L scores and low standard deviation of split frequencies (as implemented in MrBayes), as well as using the convergence diagnostics implemented in the online tool AWTY (Nylander et al. 2008). Two independent BI runs were performed as an additional check that the chains had mixed well and thus had converged.

Best-fit models of nucleotide substitution were identified using the Akaike information criterion (AIC; Akaike 1973) as implemented in jModeltest version 0.1.1 (Posada 2008). The seven-partition strategy (first codon positions of mt protein-coding genes, second codon positions of mt protein-coding genes, mt ribosomal genes, mt tRNA genes, first codon positions of nuclear *rag1*, second codon positions of nuclear *rag1*, and third-codon positions of nuclear *rag1*) employed in San Mauro et al. (2009) was used in both BI and ML frameworks. For BI, the best-fit models employed for each of the 7 partitions were GTR (Tavaré 1986) + Γ (Yang 1994) + I (Reeves 1992) (first codon positions of mt protein-coding genes), GTR + Γ + I (second codon positions of mt protein-coding genes), GTR + Γ + I (mt ribosomal genes),

GTR + Γ (mt tRNA genes), GTR + Γ + I (first codon positions of nuclear *rag1*), HKY (Hasegawa et al. 1985) + Γ + I (second codon positions of nuclear *rag1*), and GTR + Γ (third codon positions of nuclear *rag1*). In the case of RAxML, which allows only one type of substitution model to be specified, the GTR + Γ + I model was employed for each of the 7 partitions because this is the best-fit model of most partitions (4 of 7).

Support for internal branches was evaluated by nonparametric bootstrapping with 2000 replicates (RAxML, using the exhaustive bootstrap algorithm) and posterior probabilities (MrBayes) in the ML and BI analyses, respectively. Additionally, we used the approximately unbiased test (AU; Shimodaira 2002) to evaluate the 5 alternative topologies assessed by San Mauro et al. (2009) (allowing for our denser taxon sampling). AU tests used the site-wise log-likelihoods calculated by RAxML, independent GTR + Γ + I models assigned to each of the 7 partitions and were performed with CONSEL version 0.1k (Shimodaira and Hasegawa 2001) with 1 million multiscale bootstrap replicates.

Calculation of Expected and Observed Information

Experimental data sets were prepared by pruning taxa from the master alignment. The "original" data set included only the 9 taxa used previously (San Mauro et al. 2009). Fourteen additional data sets each combined the original data set with one additional species of the 14 added in this study, and 4 data sets combined the original data set with multiples of some of the additional 14 species. Table 2 gives the names and taxon composition of each data set. Because all data sets were derived from the same master alignment, they contained the same mitogenomic and *rag1* sequence positions. Each data set was analyzed both with and without partitioning. In all cases, the GTR + Γ + I model was used.

In our previous study, expected information about the controversial branch was calculated for the addition of a hypothetical taxon at various positions in the tree using only unpartitioned data (San Mauro et al. 2009). Our tests of the predictions derived from that analysis are based on our new master alignment, and some use a more realistic partitioning of these data and involve both the calculation of observed information and the recalculation of expected information. The observed information is an *a posteriori* measure of the achieved precision and was calculated after estimating the branch lengths and model parameters for the original data plus the new taxon. In contrast, the expected information is an *a priori* quantity, the precision that we would expect to achieve, and, ideally, we should not have used any of the new sequences to derive the model parameters. Minimally, we only need the length of the new branch and the position where it attaches to the original (San Mauro et al. 2009) tree, but it was unfeasible to estimate these while fixing all other parameters using only currently available software and so we used an alternate approach. Given the complete topological congruence between the original phylogeny of San Mauro et al. (2009) and the main (23 taxa) phylogeny of this study (see Results section), our branch length optimizations used input topologies based on the main phylogeny pruned to contain only the corresponding species of each data set. Information was then calculated in 5 steps to ensure that information scores are comparable (Fig. 1), in that the new taxa are included while keeping the model parameters and tree as nearly identical as possible to those from the original data set:

1. Branch lengths and model parameters of the original data set are optimized.
2. Branch lengths and model parameters of the data set with the new taxon are optimized.

TABLE 2. Names and taxon composition of each data set (cross-referenced with Figs. 3–9)

Data set name	Taxon composition
Original	Original 9 taxa included in San Mauro et al. (2009)
Boubou	Original 9 taxa + *Boulengerula boulengeri*
Caevol	Original 9 taxa + *Caecilia volcani*
Crolam	Original 9 taxa + *Crotaphatrema lamottei*
Dermex	Original 9 taxa + *Dermophis mexicanus*
Epinig	Original 9 taxa + *Epicrionops niger*
Graalt	Original 9 taxa + *Grandisonia alternans*
Gymmul	Original 9 taxa + *Gymnopis multiplicata*
Hersqu	Original 9 taxa + *Herpele squalostoma*
Hypros	Original 9 taxa + *Hypogeophis rostratus*
Ichban	Original 9 taxa + *Ichthyophis bannanicus*
Micsp	Original 9 taxa + *Microcaecilia* sp.
Oscoch	Original 9 taxa + *Oscaecilia ochrocephala*
Pracoo	Original 9 taxa + *Praslinia cooperi*
Schtho	Original 9 taxa + *Schistometopum thomense*
Original plus top 3	Original 9 taxa + *C. lamottei* + *H. squalostoma* + *B. boulengeri*
Original plus top 4	Original 9 taxa + *C. lamottei* + *H. squalostoma* + *B. boulengeri* + *E. niger*
All except top 3	Original 9 taxa + *E. niger* + *Caecilia volcani* + *O. ochrocephala* + *P. cooperi* + *H. rostratus* + *G. alternans* + *D. mexicanus* + *Microcaecilia* sp. + *G. multiplicata* + *S. thomense* + *I. bannanicus*
All except top 4	Original 9 taxa + *Caecilia volcani* + *O. ochrocephala* + *P. cooperi* + *H. rostratus* + *G. alternans* + *D. mexicanus* + *Microcaecilia* sp. + *G. multiplicata* + *S. thomense* + *I. bannanicus*
All	All 23 taxa (original 9 + additional 14 of this study)

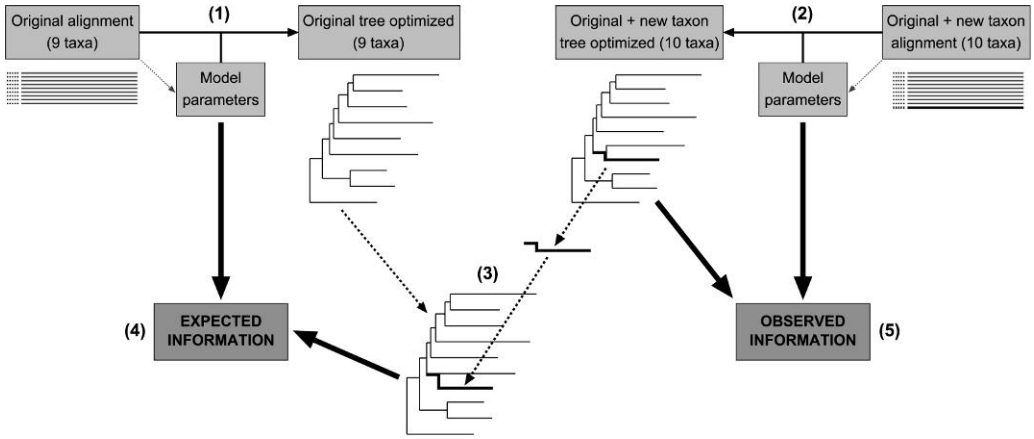

FIGURE 1. Schematic of the taxon-addition procedure employed in the calculation of expected and observed information. See text (Materials and Methods section) for explanation of numbers and steps.

3. The new taxon is added to the tree from Step 1, with the new branch lengths being in proportion to those from Step 2 (the proportion along the branch where the new taxon joins [distance to their most recent common ancestor with their sister taxon] and the relative branch length compared with its sister taxon are both transferred onto the original tree).
4. The expected information is calculated using the model parameters from Step 1 and the tree from Step 3.
5. The observed information is calculated using the tree and model parameters from Step 2.

Expected information (Goldman 1998) for each data set was estimated using EDIBLE (Massingham and Goldman 2000) as employed in San Mauro et al. (2009). As for the observed information, expected information for partitioned data sets involved calculating information for each partition independently, and scaling the information scores by multiplying by the square of the rate, and then by the partition length. The employed version of EDIBLE implementing the GTR model of substitution and site-wise rate variation, as well as the work from Geuten et al. (2007) is available at http://www.ebi.ac.uk/goldman-srv/edible/. PAML (Baseml) version 4.4 (Yang 2007) was modified to allow the GTR + Γ + I model of nucleotide substitution and to calculate the observed information by numerical approximation of the second derivative of the likelihood function using the (central) finite difference method. Observed information is a measure over every site, whereas expected information scores are often reported on a per-site basis (as in EDIBLE), and so need to be scaled by the number of sites to provide a total for an entire alignment that can be compared with the observed information or bootstrap support values. The addition of new taxa made the number of sites in the final alignment different from that of San Mauro et al. (2009), so we scaled the expected information scores from the previous study to make them comparable with the observed information.

Observed information measures the confidence in parameter estimates of the entire alignment and so implicitly makes values comparable for partitioned data sets. Expected information is a per-site measure that varies between partitions but is calculated in terms of the scaled branch lengths for that partition, not the common branch lengths, and so must be transformed to make values comparable with observed information. This is similar to designing experiments for optimal rates as discussed in Goldman (1998), and the appropriate transformation is to multiply by the rate squared, the first partition being arbitrarily chosen to have rate 1. Then, the transformed per-site expected information for each partition has to be multiplied by the number of sites in that partition in order to obtain the total expected information for the partition. The values of all partitions can then be summed to determine total expected information for the alignment and so compared with the observed information.

Bootstrap Support Analyses

In order to assess how support for the controversial branch changes as each new taxon is added, we conducted nonparametric bootstrapping for each of the 15 data sets used to estimate observed information (Table 2). One thousand bootstrap replicates were performed for each data set, using the exhaustive bootstrap algorithm implemented in RAxML. Each data set was analyzed both with and without splitting the data into 7 partitions. In all cases, the GTR + Γ + I model was used. To make bootstrap support comparable among data sets and to the information calculations, the new taxon in each data set was pruned from the bootstrap topologies prior to determining majority-rule consensus.

Because many of the molecular data in this study are of protein-coding genes, trees derived from amino acid data might also be of interest to assess how sensitive the results are to alternative coding of the data. Thus, bootstrap trees derived from amino acid data were also evaluated. For each data set in Table 2, we prepared a parallel alignment of deduced amino acid

sequences of all 13 mt protein-coding genes and nuclear *rag1*. Given that membrane proteins of the mitochondria are known to evolve differently to nuclear globular proteins (Adachi and Hasegawa 1996; Yang et al. 1998; Abascal et al. 2007), 2 partitions were employed (mt proteins combined and nuclear RAG1), and the best-fit model of amino acid substitution for each partition was identified using the AIC as implemented in ProtTest version 2.4 (Abascal et al. 2005). The best-fit models employed were mtREV (Adachi and Hasegawa 1996) + Γ + I for mt proteins and JTT (Jones et al. 1992) + Γ + I for RAG1. As for the nucleotide analyses, 1000 bootstrap replicates were performed in each case using the exhaustive bootstrap algorithm as implemented in RAxML.

In addition to assessing the change in bootstrap support of particular branches (e.g., the controversial branch of San Mauro et al. 2009), we were also interested in assessing the change in overall support of the tree as each new taxon is added. To this end, we calculated geometric means of bootstrap support of all branches in each majority-rule consensus tree (after pruning the new added taxon in each case so that values are comparable) in an attempt to obtain a rough "overall bootstrap support" of each data set's majority-rule consensus tree. We used the geometric mean because it is not as tolerant as the arithmetic mean to badly supported splits. Thus, a dense tree with a single poorly supported split would have a high arithmetic mean if all the other splits are well supported, but it would still have a low geometric mean. The use of the geometric mean was motivated by making the crude assumption that each bootstrap support value is a "probability" that the split is correct and that the splits are independent, so the probability of all splits being correct is their product, and thus related to the geometric rather than the arithmetic mean.

The relationships between information (observed and expected), bootstrap support, and the predictions of San Mauro et al. (2009) were assessed using nonparametric statistical analyses, the Spearman rank correlation test, and Mann–Whitney *U* significance test (2-tailed). Nonparametric tests were used because, for all comparisons, either one or both variables were not normally distributed (based on a Shapiro–Wilk test). All statistical analyses were conducted using R software version 2.11.1 (R Development Core Team 2011).

Other than nonparametric bootstrap, support can also be measured as Bayesian posterior probabilities derived from BI analyses. In our comparisons with information calculations, we chose to use only bootstrap support rather than Bayesian posterior probabilities mainly because bootstrap runs (using RAxML) are considerably faster than BI runs. Our BI analyses with the main (23 taxa) data set yielded a similar result to that obtained with ML (in terms of topology and support; see below). Besides, Bayesian posterior probabilities did not actually pick up the lower support for the controversial node in our previous study (it was 0.99 vs. 54% bootstrap; San Mauro et al. 2009) so there is not much scope for posterior probabilities to be sensitive to relative improvement in support.

Results and Discussion

Caecilian Phylogeny

After exclusion of gaps, alignment ambiguities, and third codon positions of mt protein-coding genes, the final combined alignment includes 11,345 positions, of which 6338 are invariant and 3758 are parsimony-informative. ML ($-\ln L = 91{,}900.09$) and BI ($-\ln L = 92{,}220.43$ for Run 1; $-\ln L = 92{,}222.76$ for Run 2) analyses yield the same inferred phylogenetic relationships among the 23 taxa (Fig. 2), with differences only in branch lengths and levels of support. All BI posterior probabilities are maximal, and ML bootstrap support is maximal or nearly so (>95%) for all internal branches except one, which also receives high support (>85%) (Fig. 2).

The inferred tree is in full agreement with the phylogeny of San Mauro et al. (2009) for shared leaves (Fig. 2) and is broadly congruent with most other recent molecular studies (Wilkinson et al. 2002; San Mauro et al. 2004, 2005; Roelants et al. 2007; Zhang and Wake 2009; San Mauro 2010; Pyron and Wiens 2011) and with morphological evidence (Nussbaum 1979; Wilkinson and Nussbaum 1996; Wilkinson 1997). All caecilian families recognized by Wilkinson et al. (2011) that are represented by more than one taxon are recovered as monophyletic with high or maximal support, as is the monophyly of Teresomata, the clade comprising all caecilian families except Rhinatrematidae and Ichthyophiidae (Wilkinson and Nussbaum 2006). The phylogenetic relationships among caecilian families all receive maximal or nearly maximal support (Fig. 2), with the branching order recovered as indicated by Wilkinson et al. (2011).

The controversial branch of San Mauro et al. (2009) and indeed all branches in that part of the tree receive maximal support from both ML and BI analyses (Fig. 2), and the AU test rejects all constrained topologies (all the alternative resolutions of Scolecomorphidae, Herpelidae, and all other teresomatans) (Table 3). These strongly supported results end the long-standing uncertainty about the deepest splits within Teresomata (Wilkinson and Nussbaum 2006): Scolecomorphidae is the sister group of all other teresomatans. Zhang and Wake (2009) recently used mt genomes to infer the phylogeny of major caecilian lineages but were unable to confidently resolve the relationships among Scolecomorphidae, Herpelidae, and all other teresomatans (bootstrap support values <50%, and topology tests could not statistically reject alternative hypotheses). The study by Zhang and Wake (2009) did not include key lineages, such as representatives of the genera *Crotaphatrema* and/or *Herpele*, identified by San Mauro et al. (2009) as those with the greatest chance of increasing phylogenetic information in this (until now) most weakly supported and controversial part of the caecilian

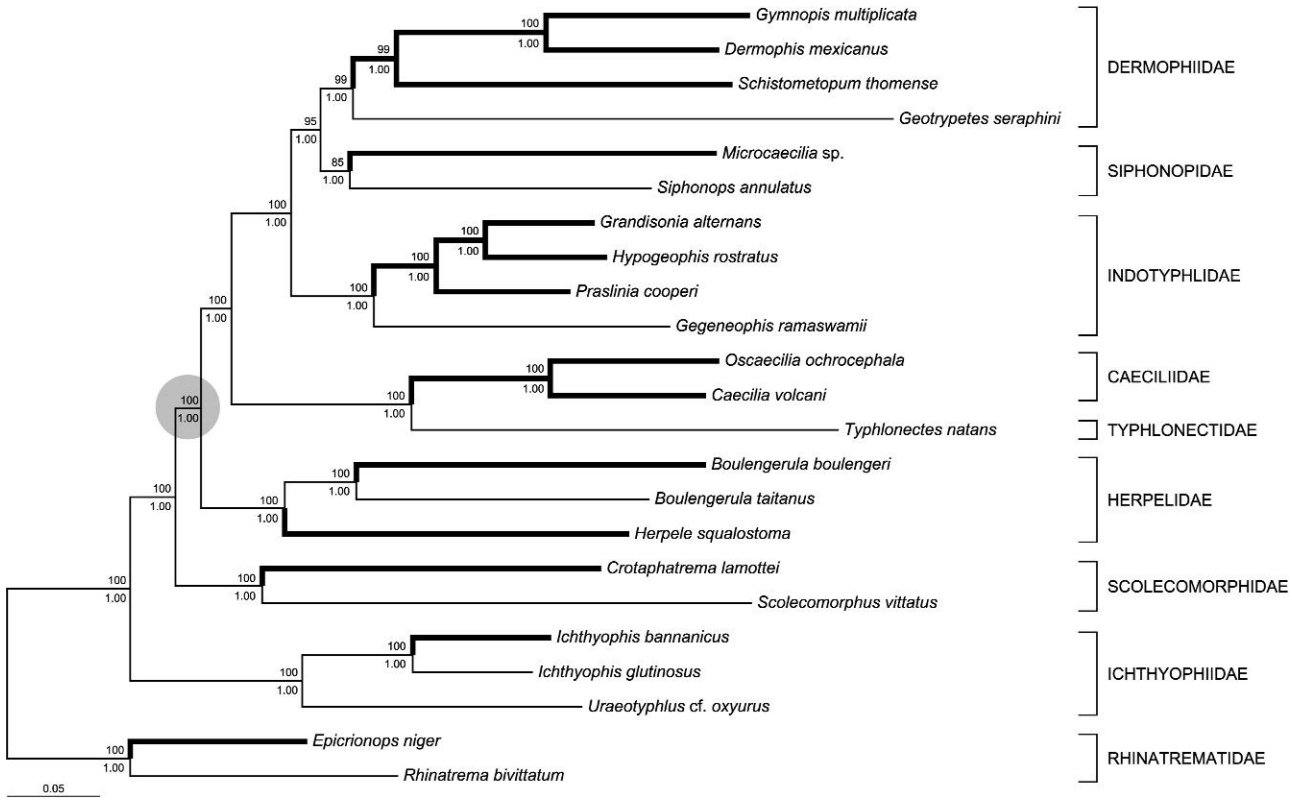

FIGURE 2. ML phylogram of caecilian amphibians inferred from analyses of the full, combined mt genome and nuclear *rag1* data. Numbers above and below branches represent support for ML (bootstrap proportions) and BI (posterior probabilities), respectively. Thick branches correspond to the new taxa added in this study relative to those taxa used in San Mauro et al. (2009) (thin branches). Grey circle indicates the controversial branch of San Mauro et al. (2009). Scale bar indicates substitutions per site.

tree. Compared with previous studies, the high support values obtained (Fig. 2) makes this caecilian phylogeny the most robust thus far for the major lineages of this order.

Observed Versus Expected Information

Figures 3 and 4 show expected and observed information scores for each single-taxon addition (nonpartitioned and partitioned analyses, respectively). In general, expected information scores follow the predictions of San Mauro et al. (2009), with *C. lamottei*, *H. squalostoma*, *B. boulengeri*, and *E. niger* having the highest scores ($U = 0$; $n_1 = 4$ $n_2 = 10$; $P = 0.005$). Both nonpartitioned and partitioned analyses yield highly correlated ranks of taxon additions ($\rho = 0.996$; $P < 0.001$), with the rank obtained corresponding approximately to the order determined by assigning a place (1st–14th) to each new taxon being traced on the plot showed on figure 5b of San Mauro et al. (2009). In contrast to expected information, observed information scores are far more variable and less intuitive (Figs. 3 and 4). Ranks of taxon additions from nonpartitioned and partitioned analyses are correlated ($\rho = 0.837$; $P < 0.001$), with the addition of *C. lamottei* being highly informative in both cases, only marginally outperformed by *E. niger* in the partitioned analyses (Figs. 3 and 4). Rank correlations between expected and observed information are not significant for both nonpartitioned ($\rho = 0.345$; $P = 0.221$) and partitioned ($\rho = 0.257$; $P = 0.374$) analyses.

That observed information is so variable and different to the expected information may be related to the fact that information is highly sensitive to branch length. If branch length estimates change slightly, information can change substantially because the variance of a branch length estimate is proportional to the exponential of the branch length. In turn, the standard deviation of the observed information (considered as an estimator of the expected information) is rather large, and the

TABLE 3. Log-likelihoods and *P* values of AU test for 5 alternative topologies

Alternative topologies	-ln *L*	*P*
Unconstrained (Fig. 2)	91,900.090	0.993
Herpelidae sister to all other teresomatans	91,932.184	0.001
Scolecomorphidae + Herpelidae sister to all other teresomatans	91,928.070	0.012
Tree fully congruent with that of Wilkinson et al. (2003)	91,955.187	< 0.001
Tree fully congruent with that of Frost et al. (2006)	92,144.730	< 0.001

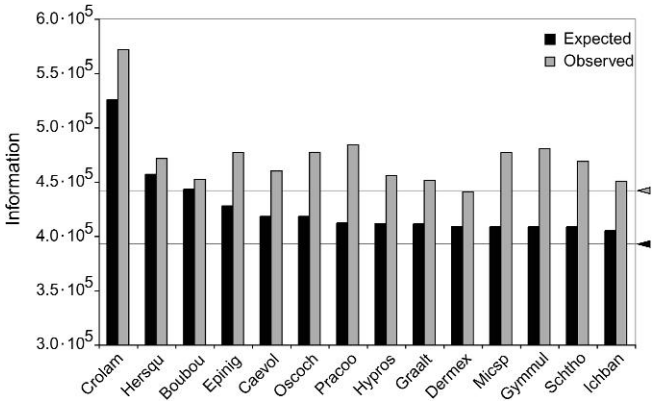

FIGURE 3. Information (expected and observed) of the controversial branch for each single-taxon addition in nonpartitioned analyses. Horizontal lines indicate expected and observed information of the original data set of San Mauro et al. (2009) (correspondence according to color codes of triangles on right). Data sets are arranged in descending order of expected information. See Table 2 for explanation of data set names.

difference between the cases is dwarfed by stochastic error. As an example, for the nucleotide nonpartitioned original data set (without adding any taxa), the expected information is 3.955×10^5, but simulated data sets of the same size (1000 replicates) yield 2.5% and 97.5% quantiles of 2.819×10^5 and 5.132×10^5, respectively, which spans virtually all the variation between the cases (Fig. 3). Other taxon addition data sets yield similar figures (not shown). In contrast, the expected information is "errorless" in the sense that it is a theoretical quantity but assumes particular values for the branch lengths. The observed information divided by the number of sites tends to the expected information, in the same way that the mean of a set of data tends to its expectation, so, if we could collect multiple similar samples of sequence data for a particular set of loci and taxa, there would be variation in the realized values of observed information between any finite samples. Although we would expect that the rank order of taxa will be at least somewhat similar between different samples, there is no guarantee that the order will match for a particular set of data. This phenomenon is what our results appear to show (Figs. 3 and 4).

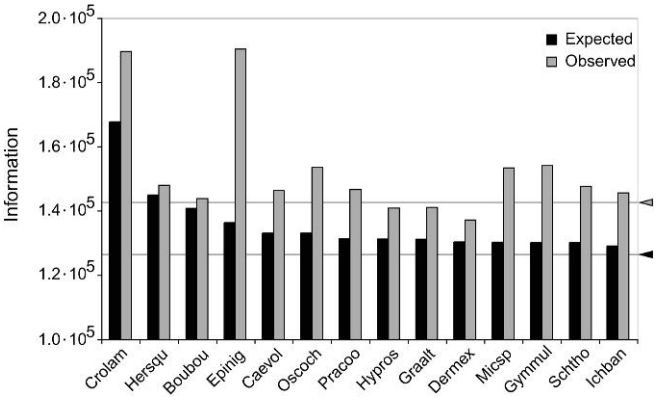

FIGURE 4. Information (expected and observed) of the controversial branch for each single-taxon addition in partitioned analyses. Horizontal lines indicate expected and observed information of the original data set of San Mauro et al. (2009) (correspondence according to color codes of triangles on right). Data sets arranged in the same order as in Figure 3. See Table 2 for explanation of data set names.

Expected and observed information scores of nonpartitioned analyses are about 3 times higher than those of partitioned analyses (Figs. 3 and 4). This difference between the partitioned and nonpartitioned data may well be explained by the trees being scaled differently. The nonpartitioned tree has the usual scale so we expect one mutation per site per unit branch length, whereas the partitioned tree is scaled in terms of the first partition having the usual scaling. The controversial branch in the partitioned analyses has a length of 0.010843 substitutions/site and 0.006120 substitutions/site in the nonpartitioned analyses. This gives a scaling of $(0.010843/0.006120)^2 = 3.14$ to make the partitioned analyses values comparable with the nonpartitioned analyses or $(1.66183/0.94426)^2 = 3.10$ if we calculate the scale from the total tree length.

Change in Bootstrap Support

Figure 5 shows bootstrap support of the controversial branch of San Mauro et al. (2009) with each new taxon addition. The nonpartitioned and partitioned analyses of nucleotide data yield highly correlated ranks of taxon additions ($\rho = 0.974$; $P < 0.001$), with the greatest increases in bootstrap support corresponding to the taxon additions of *B. boulengeri*, *C. lamottei*, and *H. squalostoma* ($U = 0$; $n_1 = 3$ $n_2 = 11$; $P = 0.010$). Interestingly, only these 3 single-taxon additions yield bootstrap support of the controversial branch over 70% (Fig. 5), which is often considered as approximating the 95% significance threshold for branch support (Zharkikh and Li 1992; Hillis and Bull 1993). Ranks of taxon additions of amino acid analyses are correlated with those of nonpartitioned ($\rho = 0.679$; $P = 0.008$) and partitioned ($\rho = 0.675$; $P = 0.008$) nucleotide analyses but show more variation for some taxon additions (Fig. 5).

The overall bootstrap support of the tree (geometric mean of bootstrap proportions) with each new taxon addition is shown in Figure 6. As for the bootstrap support of the controversial branch alone, nonpartitioned and partitioned nucleotide analyses yield highly correlated ranks of taxon additions ($\rho = 0.881$; $P < 0.001$), and the greatest resolutions of the tree correspond again to the single additions of *C. lamottei*, *B. boulengeri*, and *H. squalostoma* ($U = 0$; $n_1 = 3$ $n_2 = 11$; $P = 0.010$). Addition of *C. lamottei* yields the highest increase in overall bootstrap support of the caecilian tree for all nucleotide and amino acid analyses (Fig. 6). Other taxon additions that yield a high increase in overall bootstrap support are those of *Dermophis mexicanus*, *Gymnopis multiplicata*, *Grandisonia alternans*, and, only in the case of the partitioned nucleotide analyses, *Praslinia cooperi* (Fig. 6). With

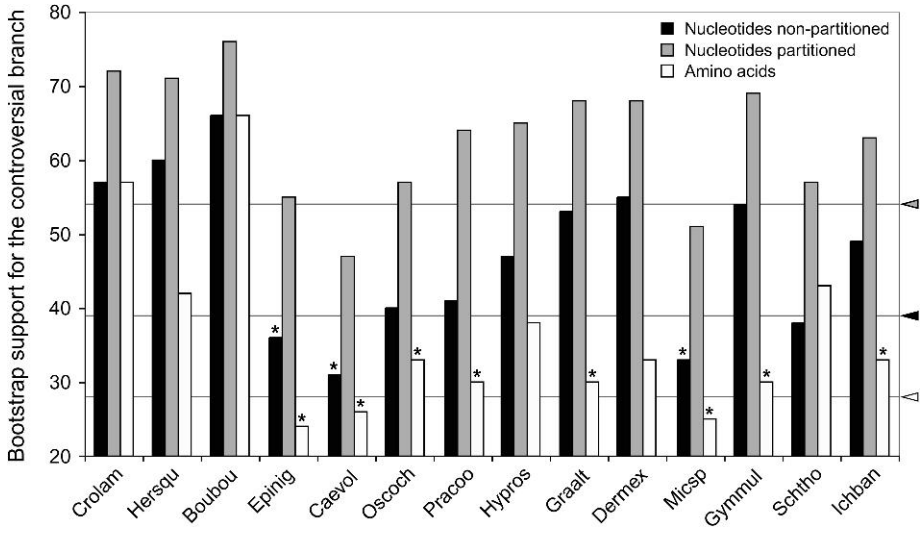

FIGURE 5. Bootstrap support (proportions) of the controversial branch of San Mauro et al. (2009) with each single-taxon addition for the nucleotide (nonpartitioned and partitioned) and amino acid analyses. Horizontal lines indicate the controversial branch's bootstrap support of the original data set of San Mauro et al. (2009) for each of the 3 types of analyses (correspondence according to color codes of triangles on right). Asterisks indicate those cases where the majority-rule bootstrap topology obtained is different from the "optimal" topology (see text). Data sets arranged in the same order as in Figure 3. See Table 2 for explanation of data set names.

overall bootstrap support, ranks of taxon additions of amino acid analyses are well correlated with those of partitioned nucleotide analyses ($\rho = 0.684$; $P = 0.007$) but less clearly with those of non-partitioned nucleotide analyses ($\rho = 0.552$; $P = 0.041$). There is a general strong correlation between the bootstrap support of the controversial branch and the overall bootstrap support of the tree for the nonpartitioned nucleotide ($\rho = 0.938$; $P < 0.001$), partitioned nucleotide ($\rho = 0.895$; $P < 0.001$), and amino acid ($\rho = 0.640$; $P = 0.014$) analyses. These strong correlations are caused mostly by the inclusion of the controversial branch in the calculation of geometric means. Geometric means not including the controversial branch show a different pattern where other single-taxon additions (different from *C. lamottei*, *B. boulengeri*, or *H. squalostoma*) have higher values (see Supplementary Table). The correlations in these cases are all nonsignificant; $P > 0.183$ in all cases. This suggests that the greatest increases in overall bootstrap support following the single additions of *C. lamottei*,

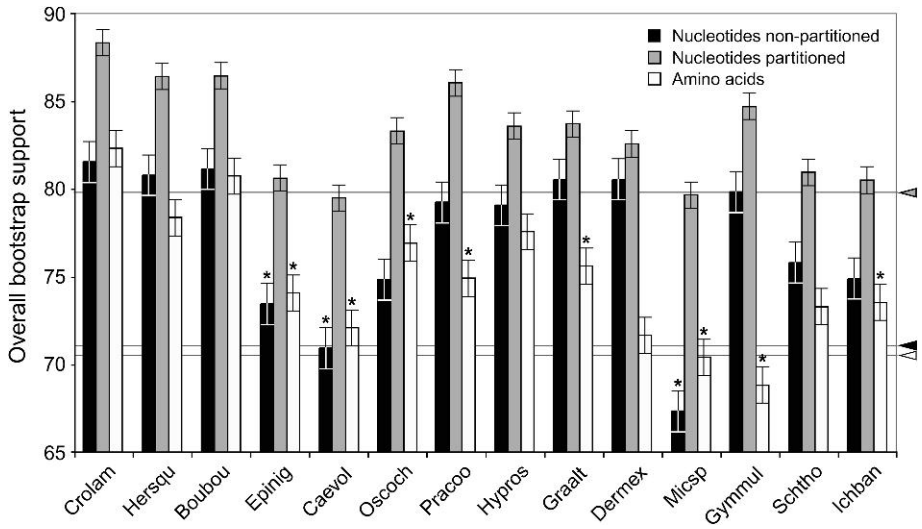

FIGURE 6. Overall bootstrap support (geometric mean of bootstrap proportions ± standard error) of the tree with each single-taxon addition for the nucleotide (nonpartitioned and partitioned) and amino acid analyses. Horizontal lines indicate the overall bootstrap support of the original data set of San Mauro et al. (2009) for each of the 3 types of analyses (correspondence according to color codes of triangles on right). Asterisks indicate those cases where the majority-rule bootstrap topology obtained is different from the "optimal" topology (see text). Data sets arranged in the same order as in Figure 3. See Table 2 for explanation of data set names.

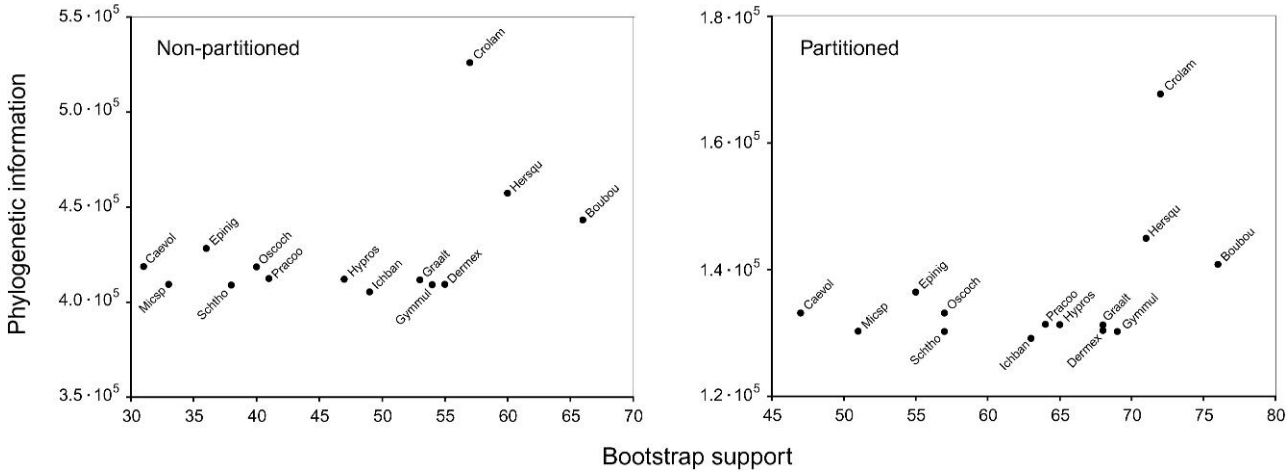

FIGURE 7. Scatter plot of expected information (*y*-AXIS) versus bootstrap support of the controversial branch of San Mauro et al. (2009) (*x*-AXIS) for both nonpartitioned and partitioned nucleotide analyses. See Table 2 for explanation of data set names.

B. boulengeri, and *H. squalostoma* (for all types of analyses) can be explained mainly by stabilization of the controversial branch (most weakly supported part of the tree) but not necessarily to a wider stabilization of other already well-supported parts of the tree.

Although the greatest increases in bootstrap support correspond to the additions of *C. lamottei*, *B. boulengeri*, and *H. squalostoma*, the exact order (rank) of bootstrap support results following single-taxon additions is not always as might be predicted from expected information (Fig. 7). In fact, correlations between bootstrap support values (either overall or of the controversial branch alone) and information scores (either expected or observed) are all nonsignificant ($P > 0.078$ in all cases). This indicates that the relationship between information and support is more complex than we previously (San Mauro et al. 2009) implicitly assumed. Goldman's approach allows us to predict how best to augment a sequence alignment so as to increase the confidence we can have in estimated branch lengths, but this is different from the ability to estimate tree topology with high bootstrap support. For example, accurately inferred short branches may have low support by virtue of their length. In phylogenetics, tree topology is part of the structure of the model (Yang et al. 1995), rather than a parameter of the model, so information matrices do not account for uncertainty in the tree itself (a particular topology needs to be assumed in making information calculations) and the utility of Goldman's method relies on the fact that accurate branch length estimation makes it easier to recover the tree topology (Atteson 1997; Corneli and Ward 2000).

In general, many of our anomalous results (in the sense of strong disparity between information and bootstrap support ranks) correspond to data sets/analyses where the majority-rule bootstrap topology is different from the optimal topology (Fig. 2). Such data sets (indicated with an asterisk in Figs. 5 and 6) yield an alternative arrangement of the Scolecomorphidae and Herpelidae branches (Herpelidae either as sister to all other teresomatans or as sister to Scolecomorphidae), with the bootstrap support of the alternative branching (bipartition) being higher (but still <50% in all cases) than that of the controversial branch bipartition. The differences in tree topology in these cases likely explain why some taxon additions (e.g., *Microcaecilia* sp. or *Caecilia volcani*) yield bootstrap support values even lower than for the original data set (Figs. 5 and 6). Many of these differences in tree topology are for the amino acid results (Figs. 5 and 6) suggesting that the absence of the noncoding partitions (mt ribosomal and tRNA genes) in amino acid analyses may be responsible for the differences in phylogenetic signal. Incidentally, the noncoding partitions were recovered as the most informative by San Mauro et al. (2009).

The case of the taxon addition of *E. niger* is illustrative of the disconnect between information and support. This taxon addition was predicted to be good in our earlier study (San Mauro et al. 2009), and indeed, information scores (both expected and observed) are high (Figs. 3 and 4) and validate the information predictions. In contrast, the addition of *E. niger* makes little to no difference in terms of bootstrap support (Figs. 5–7) and is among those taxon additions leading to a different majority-rule bootstrap topology for the nucleotide nonpartitioned and amino acid analyses (Figs. 5 and 6). Nevertheless, and despite there being no increase in support, the increase in information might be interpreted as increased confidence that the branch (and therefore the tree) is correct. Other discrepancies between information and bootstrap support results occur with the taxon additions of *C. volcani*, *D. mexicanus*, or *G. multiplicata* (Fig. 7). Of all the predictions of most useful taxon additions made by San Mauro et al. (2009), addition of *Epicrionops* was the least intuitive because the distance between where it joins the tree and the controversial branch is longer than for the more intuitive predictions, such as *Crotaphatrema* or *Herpele*. San Mauro

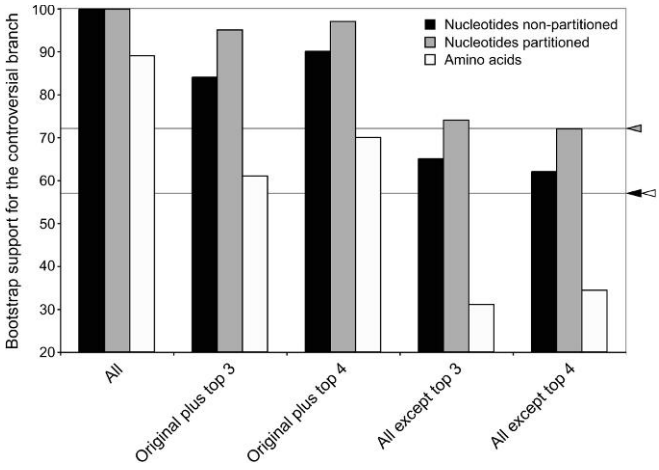

FIGURE 8. Bootstrap support (proportions) of the controversial branch of San Mauro et al. (2009) for multiple-taxon additions for nucleotide (nonpartitioned and partitioned) and amino acid analyses. Horizontal lines indicate the bootstrap support of the top single-taxon addition prediction (Crolam) for each of the 3 types of analyses (correspondence according to color codes of triangles on right). See Table 2 for explanation of data set names.

et al. (2009); fig. 5b) predicted that the places along terminal branches to add new taxa in order to best increase information are those that join as deep splits along the identified branches. Interestingly, *E. niger* is not a particularly deep split within the Rhinatrematidae (Fig. 2). The monophyly of *Epicrionops* is not well established (Pyron and Wiens 2011; Wilkinson et al. 2011), which raises the possibility that another nominal species of this genus (not sampled in this study) perhaps branching off from *R. bivittatum* more deeply might perform better than *E. niger*.

It is noteworthy that although single-taxon additions mostly enhance support as predicted using Goldman's method, none of them yields bootstrap support for the controversial branch approaching that achieved using the full data, indicating the synergism of multiple-taxon additions. We further examined this synergy with analyses of data sets in which multiple taxa were added to the original taxon sampling (Figs. 7 and 8). When the top 3 taxa are added simultaneously, bootstrap support for the controversial branch (Fig. 8) is more substantially enhanced than when they are added separately. In contrast, addition of all but the top 4 taxa has minimal effect on bootstrap support. When *Epicrionops* is also added together with the top 3 taxa, it has a substantial positive impact that contrasts with its impact when added alone or with all the other taxa. A similar pattern is found for overall bootstrap support of the tree (Fig. 9).

Empirical Validation of Predictions

This is the first study to empirically test predictions made using the method of Goldman (1998) for experimental design in phylogenetics. Although the relationship between information (expected and observed) and support is more complex than we had appreciated, the top 3 (more intuitive) taxon addition predictions made in our previous study (San Mauro et al. 2009) (*C. lamottei*, *H. squalostoma*, and *B. boulengeri*) are empirically validated here by both information and bootstrap support results (Fig. 7). These validations are further confirmed when comparing our main (23 taxa; Fig. 2) phylogenetic results with those of Zhang and Wake (2009). These authors used a taxon and gene sampling comparable with that of our study, but did not include representatives of *Crotaphatrema* and/or *Herpele*, and were unable to robustly resolve the phylogenetic relationships among Scolecomorphidae, Herpelidae, and other teresomatans. In contrast, our taxon sampling including newly determined mitogenomic data for *C. lamottei* (clearly confirmed as the top prediction by our information and overall bootstrap results) and *H. squalostoma* yields robust phylogenetic results (Fig. 2; Table 3) with strong support for the phylogenetic relationships in that part of the caecilian tree.

The less intuitive taxon addition prediction (*E. niger*) of San Mauro et al. (2009) is empirically validated by information results (both expected and observed) but not by bootstrap support results (Fig. 7). This and other cases of noncongruence between information and support (see above) reflect the fact that experimental design based on Goldman's method aims to improve the branch length estimation (information about the branch length), and although a correlated increase in bootstrap support might be expected, it is not a necessary consequence. As anticipated by San Mauro et al. (2009), our results here confirm that additional taxa that join the tree closer to controversial internal branches are of greater phylogenetic informativeness and also those

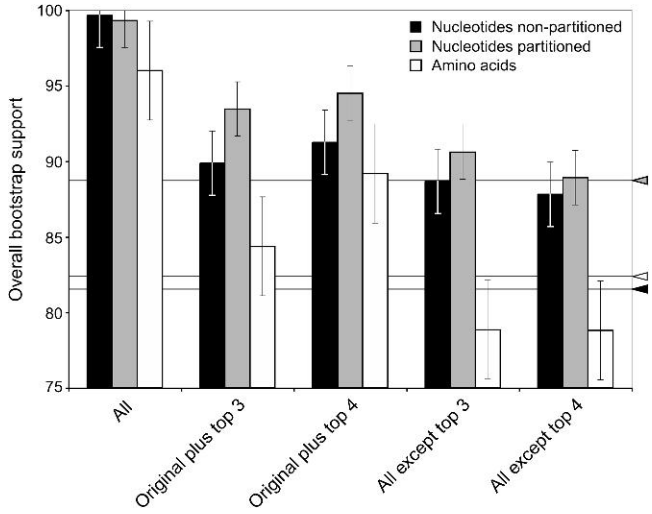

FIGURE 9. Overall bootstrap support (geometric mean of bootstrap proportions ± standard error) of the tree for multiple-taxon additions for nucleotide (nonpartitioned and partitioned) and amino acid analyses. Horizontal lines indicate the overall bootstrap support of the top single-taxon addition prediction (Crolam) for each of the 3 types of analyses (correspondence according to color codes of triangles on right). See Table 2 for explanation of data set names.

yielding higher support and resolution. This is likely a widely held intuition, and (to our knowledge) our results here provide the first explicit empirical test and validation for it.

With respect to the less intuitive (and perhaps therefore more useful) prediction tested here, that of the potential impact of the addition of *Epicrionops*, our results are intriguing. Addition without the more intuitive taxon additions does not improve confidence in the controversial branch as might have been hoped and seems therefore to suggest the prediction was unhelpful. However, when added in combination with the other taxa that comprise the more intuitive best taxon addition predictions, *Epicrionops* has a substantial positive impact on support that seems to confirm the less intuitive prediction.

San Mauro et al. (2009) also predicted that compelling resolution of the controversial branch was attainable by adding taxa to the phylogeny, without the need for more characters/genes. Both the information and the bootstrap support results of this study validate this prediction. Therefore, with respect to the debate about adding more taxa or more characters/genes to increase phylogenetic accuracy (Graybeal 1998; Hillis 1998; Kim 1998; Rannala et al. 1998; Poe and Swofford 1999; Rosenberg and Kumar 2001; Pollock et al. 2002; Zwickl and Hillis 2002; Hillis et al. 2003; Cummings and Meyer 2005; Rokas and Carroll 2005; Hedtke et al. 2006), our work provides a clear example where adding more data for (well chosen) taxa increases phylogenetic resolution and support in previously weakly supported parts of the tree without the need for more characters/genes. Conversely, addition of poorly chosen additional taxa may further destabilize weakly supported parts of the tree. Goldman's method remains underexplored, but our results demonstrate that it offers a powerful tool for experimental design in molecular phylogenetics and a coherent framework for predicting the most efficient (informative) combination of taxa and/or genes to sample in order to improve phylogenetic accuracy. Even in the era of phylogenomics and new technologies of high-throughput DNA sequencing, both taxon and gene choice will remain an important issue in systematics and phylogenetic studies (San Mauro et al. 2009; Philippe et al. 2011). In this context, Goldman's method might be useful to effectively prioritize fieldwork. The method might be also useful in helping to prioritize taxa for genome sequencing in the phylogenomic era.

Concluding Remarks

The combined results of this and our previous (San Mauro et al. 2009) study are highly illustrative and reassure that Goldman's method has great potential for experimental design in molecular phylogenetic studies. However, there are still several drawbacks to overcome in order to make the method more accessible and of broad use for making predictions that go beyond those that are intuitive already. The method remains somewhat complex, computationally demanding, and not yet implemented in a user-friendly interface. Further development and assessment of the method are needed into the benefit of multiple-taxon additions (both for prediction and for empirical testing), assessment of multiple unresolved nodes in the same phylogeny, alternative coding of data (amino acids, RY, etc.), and the more direct prediction of changes in topological support measures using the information calculations. Importantly, the complex relationship between information and support needs to be further clarified. After all, many practising systematists and phylogeneticists are likely to be more interested in support and resolution of branches than in their accurate length estimation. Goldman's method might be key in cases where intuitive predictions fail, and where resolution of weakly supported parts of the tree appears complex. It is in these cases that the potential of Goldman's method is especially worth considering.

Supplementary Material

Supplementary material, including data files and online-only appendices, can be found in the Dryad data repository at http://datadryad.org (DOI: 10. 5061/dryad. 83p1130j).

Funding

This work was supported by the European Union FP7 Marie Curie Mobility and Training Programme (PIEF-GA-2009-237658 to D.S.M.) and grants from the Ministerio de Ciencia e Innovación of Spain (CGL2007-60954 and CGL2010-18216 to R.Z.).

Acknowledgments

We thank R. Cruickshank, R. DeBry, T. Williams, and an anonymous reviewer for insightful comments on an earlier version of the manuscript. We thank T. Doherty-Bone and all the funding sources, people, and Cameroon authorities acknowledged in Doherty-Bone et al. (2011) for obtaining the *Crotaphatrema lamottei* sample. For the provision of tissue loans, we thank MVZ (Museum of Vertebrate Zoology, Berkeley, USA), ROM (Royal Ontario Museum, Toronto, Canada), UMMZ (University of Michigan Museum of Zoology, Ann Arbor, USA; especially R.A. Nussbaum), and VUB (Vrije Universiteit Brussels, Brussels, Belgium).

References

Abascal F., Posada D., Zardoya R. 2007. MtArt: a new model of amino acid replacement for Arthropoda. Mol. Biol. Evol. 24:1–5.

Abascal F., Zardoya R., Posada D. 2005 . ProtTest: selection of best-fit models of protein evolution. Bioinformatics. 21:2104–2105.

Abascal F., Zardoya R., Telford M.J. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acid Res. 38:W7–W13.

Adachi J., Hasegawa M. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. 42:459–468.

Akaike H. 1973. Information theory as an extension of the maximum likelihood principle *in* Second international symposium of

information theory (B. N. Petrov, and F. Csaki, eds.). Akademiai Kiado, Budapest, Hungary.
Atteson, K. 1997. The performance of the neighbor-joining method of phylogeny reconstruction. In Mirkin B., McMorris F.R., Roberts F.S., Rzhetsky A., editors. Mathematical hierarchies and biology. DIMACS Series of Discrete Mathematics and Theoretical Computer Science. Vol. 37. Providence (RI): American Mathematical Society. p. 133–147.
Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17: 540–552.
Corneli P.S., Ward R.H. 2000. Mitochondrial genes and mammalian phylogenies: increasing the reliability of branch length estimation. Mol. Biol. Evol. 17:224–234.
Cummings M.P., Meyer A. 2005. Magic bullets and golden rules: data sampling in molecular phylogenetics. Zoology. 108:329–336.
Doherty-Bone T.M., Kebuh Ndifon R., San Mauro D., Wilkinson M., LeGrand G.N., Gower D.J. 2011. Systematics and ecology of the caecilian *Crotaphatrema lamottei* (Nussbaum) (Amphibia: Gymnophiona: Scolecomorphidae). J. Nat. Hist. 45:827–841.
Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.
Frost D.R., Grant T., Faivovich J., Bain R.H., Haas A., Haddad C.F.B., de Sá R.O., Channing A., Wilkinson M., Donnellan S.C., Raxworthy C.J., Campbell J.A., Blotto B.L., Moler P., Drewes R.C., Nussbaum R.A., Lynch J.D., Green D.M., Wheeler W.C. 2006. The amphibian tree of life. Bull. Am. Mus. Nat. Hist. 297:1–370.
Geuten K., Massingham T., Darius P., Smets E., Goldman N. 2007. Experimental design criteria in phylogenetics: where to add taxa. Syst. Biol. 56:609–622.
Goldman N. 1998. Phylogenetic information and experimental design in molecular systematics. Proc. R. Soc. Lond. B Biol. Sci. 265: 1779–1786.
Gower D.J., Papadopoulou A., Doherty-Bone T.M., Pupin F., San Mauro D., Loader S.P., Wilkinson M. 2011. The systematics of *Boulengerula fischeri* Nussbaum & Hinkel (Amphibia: Gymnophiona: Caeciliidae) based on morphological and molecular data. Zootaxa. 2767:14–24.
Graybeal A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47:9–17.
Hasegawa M., Kishino H., Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.
Hedges S.B., Nussbaum R.A., Maxson L.R. 1993. Caecilian phylogeny and biogeography inferred from mitochondrial DNA sequences of the 12SrRNA and 16S rRNA genes (Amphibia: Gymnophiona). Herpetol. Monogr. 7:64–76.
Hedtke S.M., Townsend T.M., Hillis D.M. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst. Biol. 55:522–529.
Hillis D.M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigatior bias. Syst. Biol. 47:3–8.
Hillis D.M., Bull J.J. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182–192.
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., Ronquist F.R. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.
Huelsenbeck J.P., Ronquist F.R., Nielsen R., Bollback J.P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science. 294:2310–2314.
Jones D.T., Taylor W.R., Thornton J.M. 1992. The rapid generation of mutation data matrices from protein sequences. Comp. Appl. Biosci. 8:275–282.
Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
Katoh K., Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9:286–298.
Kim J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. Syst. Biol. 47:43–60.
Loader S.P., Wilkinson M., Cotton J.A., Measey G.J., Menegon M., Howell K.M., Müller H., Gower D.J. 2011. Molecular phylogenetics of *Boulengerula* (Amphibia: Gymnophiona: Caeciliidae) and implications for taxonomy, biogeography and conservation. Herpetol. J. 21:5–16.
Lopez-Giraldez F., Townsend J.P. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evol. Biol. 11:152.
Lupi R., D'Onorio de Meo P., Picardi E., D'Antonio M., Paoletti D., Castrignanò T., Pesole G., Gissi C. 2010. MitoZoa: a curated mitochondrial genome database of metazoans for comparative genomics studies. Mitochondrion. 10:192–199.
Massingham T., Goldman N. 2000. EDIBLE: experimental design and information calculations in phylogenetics. Bioinformatics. 16:294–295.
Nussbaum R.A. 1977. Rhinatrematidae: a new family of caecilians (Amphibia: Gymnophiona). Occ. Pap. Mus. Zool. Univ. Michigan. 682:1–30.
Nussbaum R.A. 1979. The taxonomic status of the caecilian genus *Uraeotyphlus* Peters. Occ. Pap. Mus. Zool. Univ. Michigan. 687:1–20.
Nussbaum R.A. 1985. Systematics of the caecilians (Amphibia: Gymnophiona) of the family Scolecomorphidae. Occ. Pap. Mus. Zool. Univ. Michigan. 713:1–49.
Nussbaum R.A., Wilkinson M. 1989. On the classification and phylogeny of caecilians (Amphibia: Gymnophiona), a critical review. Herpetol. Monogr. 3:1–42.
Nylander J.A.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 24:581–583.
Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain D. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9:e1000602.
Poe S., Swofford D.L. 1999. Taxon sampling revisited. Nature. 398:299–300.
Pollock D.D., Zwickl D.J., McGuire J.A., Hillis D.M. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51:664–671.
Posada D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.
Pyron R.A., Wiens J.J. 2011. A large-scale phylogeny of Amphibia including over 2,800 species, and a revised classification of extant frogs, salamanders, and caecilians. Mol. Phylogenet. Evol. 61: 543–583.
R Development Core Team 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org.
Rannala B., Huelsenbeck J.P., Yang Z., Nielsen R. 1998. Taxon sampling and the accuracy of large phylogenies. Syst. Biol. 47:702–710.
Reeves J.H. 1992. Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. J. Mol. Evol. 35:17–31.
Roelants K., Gower D.J., Wilkinson M., Loader S.P., Biju S.D., Guillaume K., Moriau L., Bossuyt F. 2007. Global patterns of diversification in the history of modern amphibians. Proc. Natl. Acad. Sci. U.S.A. 104: 887–892.
Rokas A., Carroll S.B. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol. Biol. Evol. 22:1337–1344.
Ronquist F., Huelsenbeck J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572–1574.
Rosenberg M.S., Kumar S. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proc. Natl. Acad. Sci. U.S.A. 98:10751–10756.
San Mauro D. 2010. A multilocus timescale for the origin of extant amphibians. Mol. Phylogenet. Evol. 56:554–561.
San Mauro D., Agorreta A. 2010. Molecular systematics: a synthesis of the common methods and the state of knowledge. Cell. Mol. Biol. Lett. 15:311–341.
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.
San Mauro D., Gower D.J., Oommen O.V., Wilkinson M., Zardoya R. 2004. Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1. Mol. Phylogenet. Evol. 33:413–427.
San Mauro D., Vences M., Alcobendas M., Zardoya R., Meyer A. 2005. Initial diversification of living amphibians predated the breakup of Pangaea. Am. Nat. 165:590–599.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51:492–508.
Shimodaira H., Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–1247.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22:2688–2690.
Stamatakis A., Blagojevic F., Nikolopoulos D., Antonopoulos C. 2007. Exploring new search algorithms and hardware for phylogenetics: RAxML meets the IBM Cell. J. VLSI Sig. Process. 48:271–286.
Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. 17:57–86.
Taylor E.H. 1968. The Caecilians of the World: a taxonomic analysis. Lawrence (KS): University of Kansas Press.
Wilkinson M. 1992. The phylogenetic position of the Rhinatrematidae (Amphibia: Gymnophiona): evidence from the larval lateral line system. Amphib-Reptilia 13:74–79.
Wilkinson M. 1996. The heart and aortic arches of rhinatrematid caecilians (Amphibia: Gymnophiona). Zoomorphology. 105:277–295.
Wilkinson M. 1997. Characters, congruence and quality: a study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. 72:423–470.
Wilkinson M., Loader S.P., Gower D.J., Sheps J.A., Cohen B.L. 2003. Phylogenetic relationships of African caecilians (Amphibia: Gymnophiona): insights from mitochondrial rRNA gene sequences. Afr. J. Herpetol. 52:83–92.
Wilkinson M., Nussbaum R.A. 1996. On the phylogenetic position of the Uraeotyphlidae (Amphibia: Gymnophiona). Copeia. 1996:550–562.
Wilkinson M., Nussbaum R.A. 2006. Caecilian phylogeny and classification. In Exbrayat J.-M., editor. Reproductive biology and phylogeny of Gymnophiona (Caecilians). Enfield (NH): Science Publishers. p. 39–78.
Wilkinson M., San Mauro D., Sherratt E., Gower D.J. 2011. A nine-family classification of caecilians (Amphibia: Gymnophiona). Zootaxa. 2874:41–64.
Wilkinson M., Sheps J.A., Oommen O.V., Cohen B.L. 2002. Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. Mol. Phylogenet. Evol. 23:401–407.
Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306–314.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.
Yang Z., Goldman N., Friday A. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Syst. Biol. 34:384–399.
Yang Z., Nielsen R., Hasegawa M. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. 15:1600–1611.
Zhang P., Wake M.H. 2009. A mitogenomic perspective on the phylogeny and biogeography of living caecilians (Amphibia: Gymnophiona). Mol. Phylogenet. Evol. 53:479–491.
Zhang P., Zhou H., Chen Y.-Q., Liu Y.-F., Qu L.-H. 2005. Mitogenomic perspectives on the origin and phylogeny of living amphibians. Syst. Biol. 54:391–400.
Zharkikh A., Li W.-H. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. II. Four taxa without a molecular clock. J. Mol. Evol. 35:356–366.
Zwickl D.J., Hillis D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51:588–598.