



# Phylogenetics of Chondrichthyes and the problem of rooting phylogenies with distant outgroups

Chenhong Li <sup>a,1</sup>, Kerri A. Matthes-Rosana <sup>b</sup>, Michael Garcia <sup>b</sup>, Gavin J.P. Naylor <sup>b,\*</sup>

<sup>a</sup> School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA

<sup>b</sup> Department of Scientific Computing, The Florida State University, Tallahassee, FL 32306, USA

## ARTICLE INFO

### Article history:

Received 2 September 2011

Revised 8 January 2012

Accepted 13 January 2012

Available online 24 January 2012

### Keywords:

Chondrichthyes  
Nuclear gene markers  
Phylogenetics  
Compositional bias  
Outgroup  
Rooting

## ABSTRACT

Erroneous estimates of ingroup relationships can be caused by attributes in the outgroup chosen to root the tree. Phylogenetic analyses of DNA sequences frequently yield incorrect estimates of ingroup relationships when the outgroup used to “root” the tree is highly divergent from the ingroup. This is especially the case when the outgroup has a different base composition than the ingroup. Unfortunately, in many instances, alternative less divergent outgroups are not available. In such cases, investigators must either target genes with attributes that minimize the problem (slowly evolving genes with stationary base compositions – which are often not ideal for estimating relationships among the more closely related ingroup taxa) or use inference models that are explicitly tailored to deal with an attenuated historical signal with a superimposed non-stationary base composition. In this paper we explore the problem both empirically and through simulation. For the empirical component we looked at the phylogenetic relationships among elasmobranch fishes (sharks and rays), a group whose closest living outgroup, the holocephalan Ghost fishes, are separated from the elasmobranchs by more than 100 million years of evolution. We compiled a data set for analysis comprising 10 single-copy nuclear protein-coding genes (12096 bp) for representatives of the major lineages within elasmobranchs and holocephalans. For the simulation, we used an evolutionary model on a fixed tree topology to generate DNA sequence data sets which varied both in their distance to the outgroup, and in their base compositional difference between ingroup and outgroup. Results from both the empirical data set and the simulation, support the idea that deviation from base compositional stationarity, in conjunction with distance from the root can act in concert to compromise accuracy of estimated relationships within the ingroup. We tested several approaches to mitigate such problems. We found, that excluding genes with overall faster rates and heterogeneous base compositions, while the least sophisticated of the methods evaluated, seemed to be the most effective.

© 2012 Elsevier Inc. All rights reserved.

## 1. Introduction

Phylogenomic studies based on hundreds of genes are becoming increasingly common in the literature (Baptiste et al., 2002; Blair et al., 2005; Dunn et al., 2008; Philippe et al., 2005b; Qiu et al., 2006; Regier et al., 2010; Rokas et al., 2003; Wolf et al., 2004). The increased character sampling associated with such studies often improves the resolution and support for inferred trees (Rokas et al., 2003). However, such data sets are not immune to systematic error. Indeed, the large size of many phylogenomic data sets can yield misleading inferences that have higher support values for incorrect groupings than those seen for smaller data sets. (Collins

et al., 2005; Gatesy et al., 2007; Jeffroy et al., 2006). For example, upon reanalyzing the dataset of Rokas et al. (2003), which was based on 106 genes from seven species of *Saccharomyces* yeast and an outgroup (*Candida albicans*), nonstationarity (Collins et al., 2005), divergent outgroups (Gatesy et al., 2007) and inappropriate tree reconstruction methods (Jeffroy et al., 2006) were all found to contribute to misleading inferences.

Systematic error can be especially problematic in the context of rooting, particularly when outgroups used to root the tree are highly divergent from the rest of the taxa (Graham et al., 2002; Wheeler, 1990). Although rooting with outgroups usually performs better than alternative methods (Huelsenbeck et al., 2002), the accuracy of inferred trees decreases as the distance from the root to the ingroup increases up to the point at which the root is essentially a randomized sequence relative to the ingroup taxa under investigation (Wheeler, 1990). In cases where some of the ingroup branches are also long, systematic error in the form of

\* Corresponding author. Fax: +1 843 9539199.

E-mail address: [naylorg@cofc.edu](mailto:naylorg@cofc.edu) (G.J.P. Naylor).

<sup>1</sup> Present address: Grice Marine Laboratory, College of Charleston, 205 Fort Johnson, Charleston, SC 29412, USA. Fax: +1 843 9539199.

“long-branch attraction” (Felsenstein, 1978) can artifactually draw the ingroup taxa with long branches to the base of the tree (Philippe et al., 2005b). Finally, nonstationarity of base composition can lead to misleading inference when outgroups have evolved different nucleotide frequencies from ingroups (Tarrio et al., 2000).

A number of methods have been developed to alleviate the adverse effects of systematic error in phylogenetic analysis. Methods generally fall into two categories: (1) those that use parameter-rich models to more realistically capture all of the forces that have shaped the data, (2) those that set out to exclude components of the data such that the remainder can be analyzed with standard time-homogeneous i.i.d. models (Philippe et al., 2005a). We anticipate that data exclusion approaches will become more common in the future as access to next-gen sequencing data eliminates concerns about data exclusion resulting in insufficient number of characters. Data exclusion approaches also can serve as a yardstick against which parameter rich models can be evaluated.

In this study, we explored the effectiveness of different methods to address systematic error of rooting using a dataset of 10 nuclear genes from representatives of some of the major lineages of chondrichthyan fishes (sharks, skates, rays and chimaeras). Chondrichthyans as currently described, constitute 14 orders, 54 families with about 970 species, (Nelson, 2006) although these numbers vary considerably among authors (see Compagno, 2005; Naylor et al., *in press* who suggest at least 1200 species). Two main evolutionary lineages are recognized within chondrichthyans, the subclasses Holocephali and Elasmobranchii. The monophyly of each lineage is well supported (Grogan et al., 1999; Inoue et al., 2010), but the phylogenetic relationships within Elasmobranchii are contentious. Currently, two hypotheses exist. One often seen in molecular studies (Maisey et al., 2004; Naylor et al., 2005) and in traditional classifications (Berg, 1940; Nelson, 1976, 1984), supports reciprocal monophyly of sharks and rays. The other supported by cladistic morphological studies (Compagno, 1973, 1977; de Carvalho, 1996; Shirai, 1992, 1996) suggests that sharks are paraphyletic, with rays nested among them.

Several molecular studies have explored the phylogenetic relationships among chondrichthyans (Douady et al., 2003; Human et al., 2006; Inoue et al., 2010; Lim et al., 2010; Lopez et al., 2006; Maisey et al., 2004; Mallatt and Winchell, 2007; Naylor et al., 2005; Winchell et al., 2004). However, most of these have been based on mitochondrial data. Until the phylogeny of elasmobranchs is explored with a broad suite of independent nuclear markers, it will be hard to determine if the signal emanating from the mitochondrial data represents the true species phylogeny, a mitochondrial gene tree, or neither.

In this study, we used the bioinformatics pipeline of Li et al. (2007, 2010) to identify candidate nuclear gene markers based on a comparison of genome sequences of human (*Homo sapiens*), zebrafish (*Danio rerio*) and elephant shark (*Callorhynchus milii*). We examined 20 candidate single-copy nuclear genes markers in chondrichthyans and were able to amplify 10 of these in all 14 chondrichthyan species tested. We explored the use of the previously published methods to reduce the rooting errors. Finally, we used simulation to explore the effect of base compositional nonstationarity and degree of divergence between ingroup and outgroup on phylogenetic accuracy.

## 2. Materials and methods

### 2.1. Development of nuclear gene markers

The complete genomes of human, zebrafish and the elephant shark were used to select candidate single-copy protein-coding genes following Li et al. (2007, 2010).

### 2.2. Taxon sampling and data collection

Sequence data for 10 single-copy nuclear genes selected using the Li et al. pipeline (Li et al., 2007) were collected for representatives of 14 chondrichthyan families from 11 orders: Chimaeriformes, Carcharhiniformes, Lamniformes, Orectolobiformes, Heterodontiformes, Hexanchiformes, Squaliformes, Squatiniformes, Pristiformes, Rajiformes, and Myliobatiformes. The taxon sampling was not designed to reconstruct a comprehensive phylogeny for chondrichthyans, so much as to test the utility of the newly developed markers. The representative taxa used span most of chondrichthyan diversity and include samples from all of the currently recognized major lineages. Nevertheless, we also sequenced one gene, KBTBD2 for 115 species, representing most of the 54 chondrichthyan families recognized in Nelson (2006) to test if increasing taxon sampling might reduce rooting errors in the resulting phylogeny.

DNA samples were extracted from muscle or liver tissues using DNeasy tissue kit (Qiagen, Valencia, CA, USA). PCR reactions were performed in a total volume of 20  $\mu$ l, including 0.1  $\mu$ l TaKaRa Taq™ (Takara, Mountain View, CA), 2.0  $\mu$ l 10  $\times$  PCR buffer (+MgCl<sub>2</sub>), 1.6  $\mu$ l dNTP Mixture (2.5 mM each dNTP), 0.64  $\mu$ l forward and reverse primers respectively, 0.8  $\mu$ l DNA template and 14.22  $\mu$ l distilled water. The PCR program consisted of a 95 °C initial heating for 30 s, 15 cycles of 98 °C for 10 s, the first annealing temperature (Tm) for 30 s and 72 °C for 45 s, 15 cycles of 98 °C for 10 s, the second Tm for 30 s and 72 °C for 45 s, followed by a final extension at 72 °C for 5 min. The PCR products were visualized on agarose gels to check the quality and size of amplification before sequencing. The primers and annealing temperatures used for the markers are listed in Appendix A.

### 2.3. Sequence alignment and characterization

Sequences were edited and assembled using software Sequencher 4.8 (Gene Codes Corporation, Ann Arbor, MI, USA). Sequences were translated to their corresponding amino acid sequences and aligned using Clustal W, implemented in MEGA 4 (Tamura et al., 2007) and checked with MAFFT (Katoh and Toh, 2008). The aligned sequences were subsequently translated back to nucleotide sequences for further analyses.

The size of the alignment, number of variable sites, number of parsimony informative sites, average *p*-distance among all taxa for each gene were calculated using MEGA 4. The stationarity of base composition was checked using Chi-square test implemented in PAUP\* (Swofford, 2003). Because the Chi-square test is not comparable for genes of different sizes (Phillips and Penny, 2003), we also calculated relative composition variability (RCV) using the equation given by Phillips and Penny (2003).

### 2.4. Phylogenetic methods

Maximum likelihood tree searches were conducted with Tree-Finder (Jobb, 2008), PAUP\* (Swofford, 2003) and RAxML (Stamatakis et al., 2005). Bootstrap analysis with 1000 pseudo replicates was performed to assess the statistical support for each node in the resulting phylogeny. Both partitioning by gene and codon position and concatenated analyses without partitioning were carried out. The evolutionary models used to estimate trees under likelihood were selected using the ModelTest procedure (Posada and Crandall, 1998). Bayesian analyses were carried out using MrBayes 3.1.2 (Nylander et al., 2004) using the same partitioning schemes used for the likelihood analyses. Two independent runs, with 20 Markov Chain Monte Carlo (MCMC) chains for each run were carried out for the Bayesian analyses. The heating parameter was set as “temp = 0.1” and “nswaps = 5” to improve the mixing of the

MCMC chains. The runs were terminated after 10 million generations. The software AWTY (Nylander et al., 2008) was used to evaluate the convergence of MCMC runs. Parameter states were sampled every 1000 generations (10,000 trees save for each run). The majority-rule 50% consensus tree was summarized from the sampled trees after discarding the samples before reaching stationarity as burn-in. The data were also subjected to Maximum parsimony and neighbor-joining analyses (Kimura 2-*p* distance) for comparison. Analyses were carried out both with and without sites exhibiting gaps and missing data.

In addition to analyzing all 10 genes together, ML tree searches were also performed for each individual gene to check the congruence among genes. To test the support of individual gene on alternative phylogenetic hypotheses, for each gene, the topologies resulting from the analysis of individual genes and the topologies based on data concatenating all 10 genes were evaluated by one tailed Shimodaira–Hasegawa (SH) tests (Shimodaira and Hasegawa, 1999) with 50,000 RELL bootstrap replicates as implemented in TreeFinder.

### 2.5. Rooting

We tested several procedures forwarded in the literature to remedy rooting problems: (1) The sites at third codon position were excluded from analyses as they generally have faster evolutionary rates and fluctuating base compositions which can contribute to systematic error and exacerbate long branch attraction problems (Jeffroy et al., 2006). (2) RY-coding (nucleotide A and G as R, C and T as Y) was applied to the data set to reduce problems associated with base compositional bias among taxa (Philippe et al., 2005a; Phillips et al., 2004; Woese et al., 1991), particularly associated with the outgroup (Tarrío et al., 2000). (3) The data were analyzed at the amino acid level to mitigate the influence of fast evolving sites and nonstationarity evident at the nucleotide level. (4) An explicit nonstationary model using a reversible jump algorithm was used in an effort to address the problem of compositional bias. This approach is implemented in the software PHASE 2.1 alpha (Gowri-Shankar and Rattray, 2007). In the PHASE runs, the number of models was set as “variable” and the evolutionary model used was HKY plus gamma. Two million generation of MCMC after one million “burnin” were performed to sample every 500 trees. (5) The effect of taxon sampling on rooting-induced error was evaluated for one of the genes (KBTBD2) by comparing the inferred topologies for this gene based on the sample of 14 species with one based on 115 species. (6) We tested the effects of data exclusion. Firstly, we ranked the 10 genes by their rate (average *p*-distance across taxa) and their stationarity (RCV value). Then we excluded genes, which had either high evolutionary rate or high base-compositional bias among taxa. This strategy was based on the premise that the slow genes would be less “saturated” with multiple hits on long branches than would the fast-evolving genes, so they might retain more phylogenetic signal for rooting with distant outgroups than their fast-evolving counterparts. Similarly, genes with more homogeneous base composition among taxa might suffer less from rooting problems when using distant outgroups.

### 2.6. Simulations

Distant outgroups coupled with long branches in the ingroups can cause rooting problems (Philippe et al., 2005b). Fluctuating compositional bias across taxa can also affect where to place the root (Tarrío et al., 2000). In empirical data sets, these two factors frequently co-occur because fast evolving sites have more opportunities for base compositional drift. Nonetheless, the effects of both factors are rarely studied together, so we investigated the problem

with simulation accounting for both factors. We simulated the sequences under three different lengths of the root branch and four sets of base compositions. The phylogeny resulting from the neighbor-joining analysis exhibiting reciprocal monophyly of sharks and rays was used as the initial starting template for the simulation (Fig. 1, left). Subsequently, two modified trees were derived from this starting tree, one with half the root branch length of the original one and one with a third of the root branch length of the original. Four sets of base compositional schemes were used. (1) All taxa with a base composition set at the average for the ingroup values observed in the data (GC content of 46%). (2)–(4) A set of three different base compositional profiles between ingroup and outgroup (see Table 1). In total, 12 scenarios (three root branch lengths x four sets of base compositions) were used for simulation. A hundred datasets were simulated for each scenario with each sequence being 9000 bp in length. All simulations were done with the MySP software (Rosenberg, 2005). The average base compositions of ingroups were used as the start values for the simulations. The expected base compositions in different clades were tuned to achieve the observed values for each clade as mentioned above. ML trees were estimated for all simulated data sets using the same model used to simulate the data but with stationary base compositions. The frequencies with which the correct tree was inferred under each of the different scenarios were summarized. The command blocks used for data simulation are listed in Online Supplementary 1.

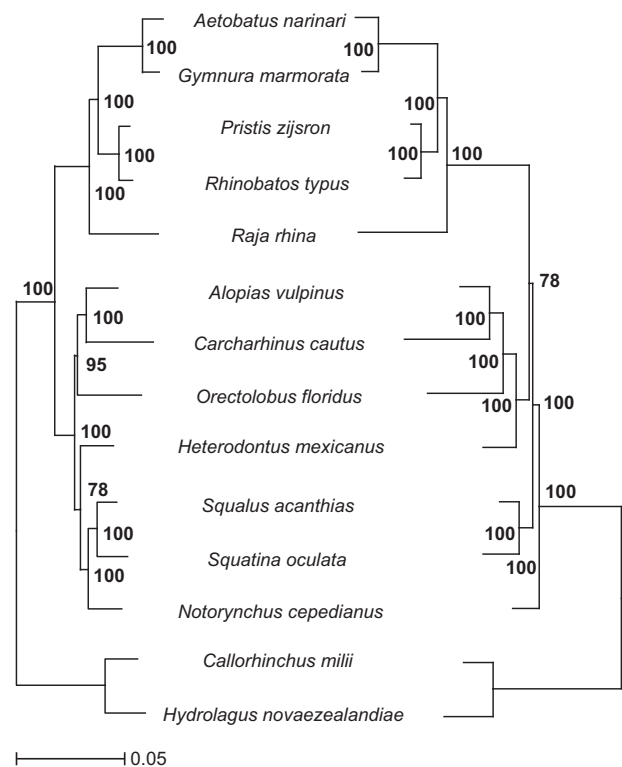


**Fig. 1.** Molecular phylogeny of chondrichthyans based on 10 nuclear genes. On the left is a NJ tree (Kimura 2-parameter distance). The numbers showing on the branches are bootstrap support values. MP tree has the similar topology as the NJ tree, except that *Heterodontus mexicanus* is grouped with *Alopias vulpinus*, *Carcharhinus cautus* and *Orectolobus floridus* instead of being placed with *Notorynchus* and *Squalus*. On the right is a ML tree inferred using TreeFinder, the numbers showing on the branches are bootstrap support values. ML analyses using PAUP\* and RAxML and Bayesian analyses resulted in the same tree. The major difference between (left) and (right) is the rooting. While ML tree (right) puts the root of sharks and rays between *Notorynchus* and the rest elasmobranchs sampled, resulting a paraphyletic shark group, MP and NJ tree (left) separates the sharks and rays as reciprocal monophyletic group.

**Table 1**  
Four different base (G + C) compositional schemes used for simulations.

| Schemes | The outgroup ( <i>C. milli</i> and <i>H. novaezealandiae</i> ) (%) | Ingroups ( <i>N. cepedianus</i> , <i>S. acanthias</i> and <i>S. oculata</i> ) biased toward the outgroup (%) | Remaining ingroup taxa (%) |
|---------|--------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|----------------------------|
| I       | 46                                                                 | 46                                                                                                           | 46                         |
| II      | 49                                                                 | 47                                                                                                           | 46                         |
| III     | 51.5                                                               | 48                                                                                                           | 46                         |
| IV      | 54.5                                                               | 49                                                                                                           | 46                         |

### 3. Results and discussion

#### 3.1. Markers: alignments, variability and compositional bias

The gene name and description of the markers are listed in Table 2. Alignments based on sequences for the 14 chondrichthyan and 2 outgroup taxa ranged from 651 bp in SCFD2 to 1734 bp in ZFPM2. The number of variable sites, number of parsimony informative sites, and average *p*-distance are listed in Table 3. Chi-square tests of base compositional stationarity indicate that genes exhibit significant non-stationarity, except for KBTBD2 and ZFPM2. When these same Chi-square tests are restricted to the third codon position data of each gene, all but KBTBD2 is rejected for stationarity. The RCV values showed similar rankings for stationarity across the 10 genes (Table 3).

#### 3.2. Phylogenies based on all genes and individual genes

Phylogeny reconstruction based on the 10 concatenated genes, including ML carried out using TreeFinder, PAUP\* and RaxML, and Bayesian analysis all resulted in the same topology (Fig. 1, right). NJ analyses (Fig. 1, left), yielded a different topology. Maximum Parsimony yielded a tree (not shown) that differed from the NJ tree in that *Heterodontus mexicanus* grouped with *Alopias vulpinus*, *Carcharhinus cautus* and *Orectolobus floridus* instead of being placed with *Notorynchus* and *Squalus* as in the NJ tree. Analysis of the sequences with and without the sites containing gaps or missing data did not affect the results. Different data partitioning schemes did not lead to different trees. The major difference between the two trees in Fig. 1 appears to be associated with the placement of the root. Fig. 1 (right) places the root between *Notorynchus* and the rest elasmobranchs, making sharks paraphyletic, while, the topology on the left yields a topology in which sharks and rays form reciprocally monophyletic groups.

The individual gene trees show substantial topological incongruence. Two gene trees (C3orf39 and RAG2) yield the same topology seen in the MP analysis of the combined data set, while one gene tree (KBTBD2) yields the same topology as combined data set ML tree (Fig. 1, right). The tree topologies resulting from

**Table 2**  
Nuclear gene markers developed in this study.

| Gene ID             | Short name | Position on <i>C. milli</i> genome | Identity | Coverage | Length | Description                               |
|---------------------|------------|------------------------------------|----------|----------|--------|-------------------------------------------|
| ENSG00000139826     | ABHD13     | AAVX01079547.1:1:833               | 75       | 89       | 907    | Abhydrolase domain-containing protein 13  |
| ENSG00000144647     | C3orf39    | AAVX01027905.1:64:1793             | 70       | 94       | 1642   | Uncharacterized glycosyltransferase AGO61 |
| ENSDARG00000071584  | KBTBD2     | AAVX01028885.1:140:1663            | 71       | 100      | 1551   | Hypothetical protein LOC563254            |
| ENSG00000117707     | PROX1      | AAVX01031237.1:541:2258            | 74       | 100      | 1736   | Prospero homeobox protein 1               |
| ENSDARG000000052122 | RAG1       | AAVX01004067.1:890:2541            | 69       | 95       | 1657   | V(D)J recombination-activating protein 1  |
| ENSG00000175097     | RAG2       | AAVX01103316.1:1420:2              | 60       | 90       | 1430   | V(D)J recombination-activating protein 2  |
| ENSG00000184178     | SCFD2      | AAVX01026251.1:1050:279            | 62       | 86       | 721    | Sec1 family domain-containing protein 2   |
| ENSG00000187714     | SLC18A3    | AAVX01061230.1:876:2011            | 69       | 71       | 1138   | Vesicular acetylcholine transporter       |
| ENSG00000145217     | SLC26A1    | AAVX01051623.1:2665:1564           | 58       | 72       | 1102   | Sulfate anion transporter 1               |
| ENSG00000169946     | ZFPM2      | AAVX01052760.1:1974:17             | 65       | 90       | 2235   | Zinc finger protein ZFPM2                 |

Note: Gene ID is the id number from Ensembl release 53. Identity, Coverage and Length are average identity, coverage rate and average length among human, zebrafish and elephant shark were calculated as Li et al. (2007).

**Table 3**  
Characterization of the 10 nuclear gene markers.

| Markers | # bp | # var | # PI | <i>p</i> -Dis | <i>p</i> (Chi-sq) | RCV    |
|---------|------|-------|------|---------------|-------------------|--------|
| ABHD13  | 816  | 352   | 193  | 0.1256        | 0.0013            | 0.0492 |
| KBTBD2  | 1461 | 660   | 338  | 0.1313        | 0.99              | 0.0252 |
| PROX1   | 1551 | 695   | 428  | 0.1415        | 0.0017            | 0.0522 |
| ZFPM2   | 1734 | 997   | 646  | 0.1867        | 0.5028            | 0.0403 |
| RAG1    | 1344 | 736   | 509  | 0.188         | <0.0001           | 0.0439 |
| SLC18A3 | 930  | 446   | 334  | 0.189         | <0.0001           | 0.0681 |
| C3orf39 | 1398 | 714   | 477  | 0.1934        | <0.0001           | 0.0553 |
| SLC26A1 | 975  | 615   | 343  | 0.195         | <0.0001           | 0.0980 |
| SCFD2   | 651  | 389   | 255  | 0.213         | 0.0017            | 0.0791 |
| RAG2    | 1236 | 795   | 496  | 0.224         | 0.0001            | 0.0590 |

Note: All values were calculated on the sequence alignments of 16 taxa (14 chondrichthyan plus two outgroups). # bp is the number of base pair; # var is the number of variable sites; # PI is the number of parsimony informative sites; *p*-Dis is the average *p*-distance among all taxa; *p* (Chi-sq) is the *p*-value of Chi-square test on base composition; RCV is the relative composition variability (Phillips and Penny, 2003).

analysis of the other seven markers are different from both those seen in Fig. 1 left and right (see Fig. 2). SH-tests fail to reject both trees for all genes but KBTBD2, which rejected the topology of Fig. 1 (left) (*p* < 0.05), but not that of Fig. 1 (right).

#### 3.3. Application of existing methods to reduce rooting errors

##### 3.3.1. PHASE 2.1 alpha

When the data set was subjected to the explicit nonstationary model implemented in PHASE 2.1 alpha, we obtained the same tree topology seen in Fig. 1 (right) with the characteristic paraphyletic grouping for sharks.

##### 3.3.2. Removal of 3rd codon positions, RY coding and amino acid analysis

ML analyses of the data set in which: (1) 3rd codon positions were removed, (2) RY coding was implemented, and (3) the translated amino acid sequence were used as input data, all yielded the topology in which sharks were seen to form a paraphyletic group (Fig. 1, right). This suggests either that the topology depicting paraphyletic sharks is correct, or that existing methods are ineffective in reducing the erroneous signal latent in this particular data set.

##### 3.3.3. Taxon sampling

We explored the effects of increasing the taxon sampling within the ingroup. We sequenced 115 species covering most chondrichthyan families for KBTBD2 gene, the only gene in the set of 10 that appears to have a stationary base composition across the sample of taxa used in the initial tests (Table 3). The ML tree based on expanded KBTBD2 sequences yielded a topology in which sharks were paraphyletic as shown in Fig. 1, right; (see Online Supplementary figure). This may be because the additional taxon-sampling, while improving branch length estimates among

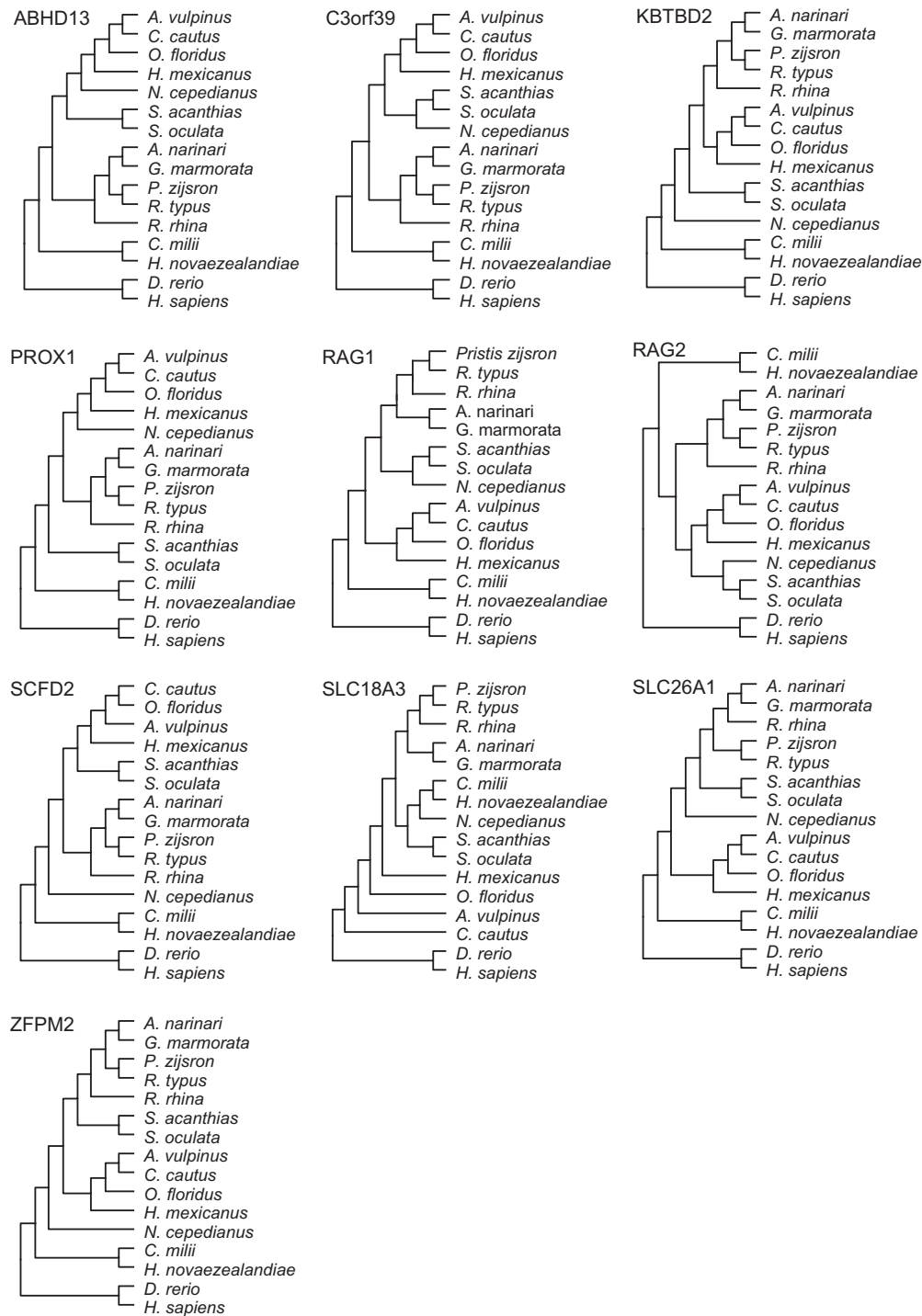


Fig. 2. ML trees of chondrichthyans reconstructed using individual genes (only topologies are shown).

the in-group taxa, did little to improve the estimate of the changes between the outgroup and the base of the ingroup. Because there are no extant taxa for which sequence data can be obtained to break the long branch between elasmobranchs and their chimae-roid outgroups, an alternative approach is needed to address the rooting problem in this particular case.

### 3.3.4. Data exclusion

We tested data exclusion as a method to improve the accuracy of rooting. We ranked the genes by their *p*-distance and RCV values. The genes with smaller *p*-distance were assumed to have over-all slower rates and the genes with smaller RCV values were

assumed to be more homogenous in base composition across taxa. We excluded the five genes with the fastest rates and least homogeneous base compositions and subjected the remaining five to ML analysis. The included genes were ABHD13, KBTBD2, PROX1, RAG1 and ZFPM2 (the genes in bold font in Table 3). The analysis resulted in a topology consistent with reciprocal monophyly of sharks and rays. The topology is similar to that in Fig. 1 (left), except that *Notorynchus cepedianus* moves from its position as the sister taxon to the clade containing *Squalus acanthias* and *Squatina oculata* (Fig. 1, left) to the immediate sister group to a clade containing *Heterodontus mexicanus*, *Orectolobus floridus*, *Alopias vulpinus* and *Carcharhinus cautus* (Fig. 3).

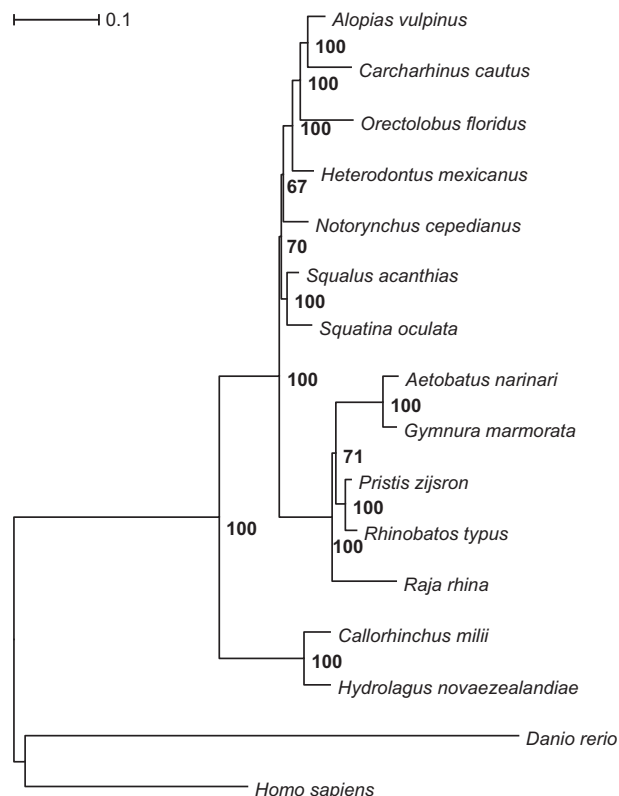


**Fig. 3.** ML trees of chondrichthyans reconstructed using five genes that have both slow rate and homogeneous base composition among taxa. These five genes are ABHD13, KBTBD2, PROX1, RAG1 and ZFPM2.

### 3.4. Results from simulations

Simulations showed that both long root branch and non-stationarity interact to affect the placement of the root. At all three biased base composition settings and even the stationary setting,

data simulated with longer root branches compromised the accuracy of rooting (Fig. 4). The more biased the base composition the stronger the distorting influence on the placement of the root. When the base composition setting changed from GC content of 51.5%–48%–46% to 54.5%–49%–46% (outgroups–ingroups1–ingroups2), the success rate for correct root placement diminished substantially (Fig. 4). It is also clear from Fig. 4 that the more distant outgroup the more pronounced the effect of base compositional bias, the most problematic datasets being those with the longest root branches and the most biased base compositions (Fig. 4).

Distant outgroups are often cited as the cause of rooting problems (Gatesy et al., 2007; Wheeler, 1990). Nonstationarity is also occasionally invoked to account for rooting errors (Tarrío et al., 2000). Our results show that both factors interact to make the problem especially difficult.

### 3.5. Phylogeny of chondrichthyans

One of the more controversial debates associated with the phylogeny of chondrichthyans concerns the relationships between sharks and rays. As mentioned previously, traditional views held that sharks and rays are reciprocal monophyletic groups (Berg, 1940; Nelson, 1976, 1984). This conclusion is also supported by many recent molecular studies (Maisey et al., 2004; Naylor et al., 2005). However, morphological studies using cladistic methods suggest that sharks constitute a paraphyletic group with rays nested within them. (Compagno, 1973, 1977; de Carvalho, 1996; Shirai, 1992, 1996). Parsimony analysis of the morphological data suggest that pristionorhynchiform sawsharks are the immediate sister group to a clade containing the skates and rays, and that this group is nested in turn, within a clade containing squatiniform, squaliform, hexanchiform and echinorhinid to form the “Hypnosqualean” group (Shirai, 1992, 1996). The remaining orders Lamniformes, Carcharhiniformes, Orectolobiformes and Heterodontiformes form the “Galea” group which fall as the sister group to the hypnosqualea. In our results, the Galea clade (*Alopias vulpinus*, *Carcharhinus cautus*, *Orectolobus floridus* and *Heterodontus mexicanus*) is supported by

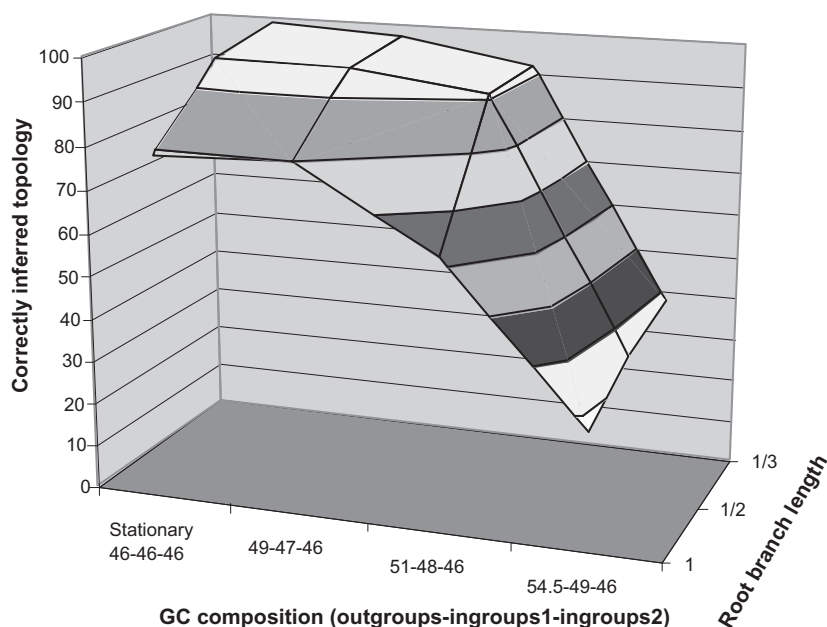


**Fig. 4.** The percentage of correctly inferred topologies resulting from the analysis of simulated data that vary according to the change of base composition and root length. Four categories of base composition and three root branch lengths were used for sequence simulation (see Section 2.6. for detail). A hundred simulations were carried for each combination of different base composition and root branch length.

both ML and MP analyses, but the close relationship between rays and Squatiniformes (*Squatina oculata*), Squaliformes (*Squalus acanthias*) and Hexanchiformes (*Notorynchus cepedianus*) is not seen in any of the trees. Thus though several of the trees presented herein are consistent with a paraphyletic arrangement for sharks, none are consistent with the Hypnosqualean hypothesis. Instead they suggest an arrangement in which rays group with the Galea to the exclusion of the squatiniform, squaliform, hexanchiform taxa. Nonetheless, we do not place much stock in these results because the internal branches for the novel groupings are very short (Figs. 1 and 3). Rather, we favor the idea that rapid patterns of diversification among some of the major lineages have contributed to uncertainty in understanding their interrelationships and that these uncertainties have been exacerbated by the interaction of base compositional non-stationarity and a highly divergent outgroup.

#### 4. Conclusion

In the present study, we present phylogenetic inferences among 14 representatives of the major lineages of chondrichthyans based on 10 protein-coding nuclear genes. We anticipate that these markers will be useful for estimating phylogenetic relationships among chondrichthyan sub-groups. However, more importantly, we bring attention to a more general and perhaps widespread problem in which distant outgroups interact with nonstationarity of base composition between the ingroup and the outgroup to compromise phylogenetic accuracy. We tested a suite of candidate approaches that have been forwarded to address such problems

and found, at least in our case, that none were effective. The approach we found to be most effective was to remove genes with overall faster rates and heterogeneous base compositions a priori and subject the remaining data to standard statistical analyses. Nonetheless this approach, while better than the others we tried, was not especially satisfying. We were unable to come up with criteria or measures to predict the overall degree of divergence allowable for a range of base compositions. We are confident that next-generation sequencing will provide the molecular systematics community with hundreds of potential genes and other markers from which to estimate phylogenetic relationships. However we are less confident that more than a handful of such genes and markers will have the properties necessary to provide unbiased estimates of relationships using standard models, across a wide range of taxa. In light of this we feel it important to continue to improve theoretical approaches to accommodate and model the biases that affect so many molecular phylogenetic data sets.

#### Acknowledgments

This work was supported by National Science Foundation (DEB-1132229). We thank J.-J.M. Riethoven for allowing us to access the computer clusters at Bioinformatics Core Research Facility, University of Nebraska, Lincoln.

#### Appendix A. PCR primers and annealing temperatures used for the 10 nuclear gene markers

| Marker  | Primer name   | Primers sequence                 | PCR Tm | PCR steps |
|---------|---------------|----------------------------------|--------|-----------|
| ABHD13  | ABHD13_F44    | ATTGCTTTGGTCTCRTGGTCYTG          | 5250   | 1st       |
|         | ABHD13_R918   | TAGCCCTGRCACTGCCAWGTGT           |        |           |
|         | ABHD13_F52    | GGTCTCGTGGTCYTGGAAGTTATG         | 5856   | 2nd       |
|         | ABHD13_R894   | TTGTGAGTYCCATCTGGAAAAATG         |        |           |
|         | ABHD13_F127   | CCTTTTGCTTCTYTTGATMCTTGC         | 5250   | 1st       |
|         | ABHD13_R873   | ATGRCTAGTCTYTTAGTCCGTGAYGT       |        |           |
|         | ABHD13_F66    | CCGRATTTCTCTGCTAGCTTTGAT         | 5856   | 2nd       |
| C3orf39 | ABHD13_R830   | GYTTCATCACTACTGGTGGGA            |        |           |
|         | C3orf39_F203  | TCTGCACGGGAAGAATRCACAC           | 5250   | 1st       |
|         | C3orf39_R1709 | CCAAGGAGGTTTATTGAAGATGC          |        |           |
|         | C3orf39_F211  | GGAAGAATACACACWGAACGCATCTG       | 5856   | 2nd       |
|         | C3orf39_R1677 | CCAAATCAGGTATGTGTGAAGGGYTT       |        |           |
|         | C3orf39_F255  | GAAGCTGARGAGTTTGTRTTCTTCCA       | 5856   | 2nd       |
|         | C3orf39_R1619 | GGTAARATGTACGGCATGTARGTGTCTC     |        |           |
| KBTBD2  | KBTBD2_F9     | TGTATTGCATCGTTGTGCRGAATA         | 5856   | 1st       |
|         | KBTBD2_R1522  | GCAATGCAACCATCTCTCCATC           |        |           |
|         | KBTBD2_F49    | AATGCAGAGAACTGYGTGCGAT           | 5856   | 2nd       |
|         | KBTBD2_R1515  | AACCATCTCTCCATCAAAYTCAAATC       |        |           |
| PROX1   | PROX1_F2      | TGCCTGACCATGACAGCAC              | 5250   | 1st       |
|         | PROX1_R1693   | TGTCGTGCAAGTCRCCGCACT            |        |           |
|         | PROX1_F50     | GAAGAGTCGAYATCGGAGTGAA           | 5856   | 2nd       |
|         | PROX1_R1667   | TTAATGAGAGACAGRGAGAGACCTTC       |        |           |
|         | PROX1_F50     | GAAGAGTCGAYATCGGAGTGAA           | 5856   | 1st       |
|         | PROX1_R1693   | TGTCGTGCAAGTCRCCGCACT            |        |           |
|         | PROX1_F133    | ATGAATCCCCACGGYTCAGATCA          | 6058   | 2nd       |
| RAG1    | PROX1_R1667   | TTAATGAGAGACAGRGAGAGACCTTC       |        |           |
|         | RAG1_F1380    | AACACATTCTTGAGYTGCACTCA          | 5250   | 1st       |
|         | RAG1_R2982    | ATTGTGTGATTCAATRAACCTCTG         |        |           |
|         | RAG1_F1385    | ATTCTTRAGTTGCAGTCAATAYCATAAAATGT | 5250   | 2nd       |

(continued on next page)

## Appendix A (continued)

| Marker  | Primer name   | Primers sequence                | PCR Tm | PCR steps |
|---------|---------------|---------------------------------|--------|-----------|
| RAG2    | RAG1_R2954    | TATTTGGAAGTGACARCCAGTGGT        | 5250   | 2nd       |
|         | RAG1_F1407    | CATAAAATGTACAGAACAGTSAAGC       |        |           |
|         | RAG1_R2953    | ATTTTGAAGTGACAGCCARTGRTGCT      |        |           |
|         | RAG2_F51      | CGGGTTCTCCYTATGAATTT            | 5250   | 1st       |
|         | RAG2_R1442    | TGAACATTGCAAAARTACTTG           | 6058   | 2nd       |
|         | RAG2_F94      | GGTCARAAAGGATGGCCAAARAGAT       |        |           |
| SCFD2   | RAG2_R1380    | ACCCATRCCTGGGCATGTACC           | 6058   | 2nd       |
|         | RAG2_F96      | CCARAAAGGATGGCCAAARAGATCCTG     |        |           |
|         | RAG2_R1252    | TGCAYTTAATCCARTAACCTTCCTCATCA   |        |           |
|         | SCFD2_F58     | AAGGTGAAGARGGCGKTGGT            | 6260   | 1st       |
| SLC18A3 | SCFD2_R764    | TGTGGRTAATTRGCAAGYTCTCCTGC      | 6058   | 2nd       |
|         | SCFD2_F61     | GTGAAGARGGCGKTGGTGTT            |        |           |
|         | SCFD2_R757    | AATTAGCAAGYTCTCCTGCAATWATTCT    |        |           |
|         | SLC18A3_F357  | CGAAGACATCCARATAGGAGTGTC        | 5250   | 1st       |
| SLC26A1 | SLC18A3_R1420 | GGCACACGTTTCTGAGCAGYAG          | 5856   | 2nd       |
|         | SLC18A3_F369  | AATAGGAGTGCTCTTYGCATCC          |        |           |
|         | SLC18A3_R1366 | CCATGCCGAGGYTGAGCTG             |        |           |
|         | SLC18A3_F487  | TTCTTTTCCACCCTYACSTTTGTC        | 6058   | 2nd       |
|         | SLC18A3_R1173 | CGGCTATAGCATACACRTRCCRTAYAC     |        |           |
|         | SLC26A1_F12   | AGTTTTCAACCTTGGGATYGTGTC        | 5250   | 1st       |
| ZFP204  | SLC26A1_R1104 | GTAAGGGCATTTCAGRAAGAAGTC        | 5250   | 2nd       |
|         | SLC26A1_F49   | CAACCTTTGCTSGATGGATTTG          |        |           |
|         | SLC26A1_R1089 | CAGGAAGAAGTCTTTATTRGCATA        |        |           |
|         | SLC26A1_F49   | CAACCTTTGCTSGATGGATTTG          | 5250   | 1st       |
|         | SLC26A1_R1089 | CAGGAAGAAGTCTTTATTRGCATA        | 6058   | 2nd       |
|         | SLC26A1_F119  | TTGGAATAAACTWCCACGYAGYCA        |        |           |
| ZFP204  | SLC26A1_R982  | TTGAACCCCTGGATTTGRCCYAACAA      |        |           |
|         | SLC26A1_F84   | AATTACGATCCTAMCAAYGCAAGTCAAGTA  | 6058   | 1st       |
|         | SLC26A1_R1058 | GGGCTTCAAAGCGRAAGAKCTTAAT       | 6058   | 2nd       |
|         | SLC26A1_F197  | AAACYAATTTCTGTGATYTGATAACAAGTGC |        |           |
|         | SLC26A1_R846  | CCARACCAGWGTGTCATTTTGTCT        |        |           |
| ZFP204  | ZFP204_F30    | GGTGCCAACCTTGAAATGYACCATCTG     | 5250   | 1st       |
|         | ZFP204_R1954  | TTTATTGCTCCTGGRTAAAGGCA         | 5856   | 2nd       |
|         | ZFP204_F147   | GGTTTCCAGAGYACACAGGGA           |        |           |
|         | ZFP204_R1945  | CCTGGATAAAGGCAATGWGG            |        |           |
| ZFP204  | ZFP204_F147   | GGTTTCCAGAGYACACAGGGA           | 5250   | 1st       |
|         | ZFP204_R1945  | CCTGGATAAAGGCAATGWGG            | 5856   | 2nd       |
|         | ZFP204_F277   | ATGCAGCAAGYAGRCAGGAAGT          |        |           |
|         | ZFP204_R1813  | GAGACCAGTTTNGAATTTTATTTTCTC     |        |           |

Note: The markers are named with the Ensembl locus tag name or the orthologous gene name of humans; PCR Tm is the annealing temperature for PCR, the first numbers are for the first 15 touchdown PCR cycles while the second numbers are for the last 15 PCR cycles; PCR steps are steps in the nested-PCR.

## Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.ympev.2012.01.013](https://doi.org/10.1016/j.ympev.2012.01.013).

## References

- Baptiste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M., Philippe, H., 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc. Natl. Acad. Sci. USA* 99, 1414–1419.
- Berg, L.S., 1940. Classification of fishes, both recent and fossil. *Trav. Inst. Zool. Acad. Sci. URSS* 5, 87–517.
- Blair, J.E., Shah, P., Hedges, S.B., 2005. Evolutionary sequence analysis of complete eukaryote genomes. *BMC Bioinform.* 6, 53.
- Collins, T.M., Fedrigo, O., Naylor, G.J., 2005. Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. *Syst. Biol.* 54, 493–500.
- Compagno, L.J.V., 1973. Interrelationships of living elasmobranchs. In: Greenwood, P.H., Miles, R.S., Patterson, C. (Eds.), *Interrelationships of Fishes*. Academic Press, London.
- Compagno, L.J.V., 1977. Phyletic relationships of living sharks and rays. *Am. Zool.* 17, 303–322.
- Compagno, L.J.V., 2005. Global checklist of living chondrichthyan fishes. In: Fowler, S.L., Cavanagh, R.D., Camhi, M., Burgess, G.H., Cailliet, G.M., Fordham, S.V., Simpfendorfer, C.A., Musick, J.A. (Eds.), *Sharks, Rays and Chimaeras: The Status of Chondrichthyan Fishes*. International Union for the Conservation of Nature and Natural Resources, Gland and Cambridge, pp. 401–423.
- de Carvalho, M.R., 1996. Higher-level elasmobranch phylogeny, basal squalines, and paraphyly. In: Stiassny, M.L.J., Parenti, L.R., Johnson, G.D. (Eds.), *Interrelationships of Fishes*. Academic Press, San Diego.
- Douady, C.J., Dosay, M., Shivji, M.S., Stanhope, M.J., 2003. Molecular phylogenetic evidence refuting the hypothesis of Batoidae (rays and skates) as derived sharks. *Mol. Phylog. Evol.* 26, 215–221.
- Dunn, C.W., Hejnal, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sorensen, M.V., Haddock, S.H., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M.Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452, 745–749.

- Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Gatesy, J., DeSalle, R., Wahlberg, N., 2007. How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. *Syst. Biol.* 56, 355–363.
- Gowri-Shankar, V., Rattray, M., 2007. A reversible jump method for Bayesian phylogenetic inference with a nonhomogeneous substitution model. *Mol. Biol. Evol.* 24, 1286–1299.
- Graham, S.W., Olmstead, R.G., Barrett, S.C., 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Mol. Biol. Evol.* 19, 1769–1781.
- Grogan, E.D., Lund, R., Didier, D., 1999. Description of the chimaerid jaw and its phylogenetic origins. *J. Morphol.* 239, 45–59.
- Huelsenbeck, J.P., Bollback, J.P., Levine, A.M., 2002. Inferring the root of a phylogenetic tree. *Syst. Biol.* 51, 32–43.
- Human, B.A., Owen, E.P., Compagno, L.J., Harley, E.H., 2006. Testing morphologically based phylogenetic theories within the cartilaginous fishes with molecular data, with special reference to the catshark family (Chondrichthyes; Scyliorhinidae) and the interrelationships within them. *Mol. Phylog. Evol.* 39, 384–391.
- Inoue, J.G., Miya, M., Lam, K., Tay, B.H., Danks, J.A., Bell, J., Walker, T.L., Venkatesh, B., 2010. Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. *Mol. Biol. Evol.*
- Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231.
- Jobb, G., 2008. TreeFinder, Munich, Germany. Distributed by the author at <[www.treefinder.de](http://www.treefinder.de)>.
- Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9, 286–298.
- Li, C., Ortí, G., Zhang, G., Lu, G., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evol. Biol.* 7, 44.
- Li, C., Riethoven, J.J., Ma, L., 2010. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. *BMC Evol. Biol.* 10, 90.
- Lim, D.D., Motta, P., Mara, K., Martin, A.P., 2010. Phylogeny of hammerhead sharks (Family Sphyrnidae) inferred from mitochondrial and nuclear genes. *Mol. Phylog. Evol.* 55, 572–579.
- Lopez, J.A., Ryburn, J.A., Fedrigo, O., Naylor, G.J., 2006. Phylogeny of sharks of the family Triakidae (Carcharhiniformes) and its implications for the evolution of carcharhiniform placental viviparity. *Mol. Phylog. Evol.* 40, 50–60.
- Maisey, J.G., Naylor, G.J.P., Ward, D.J., 2004. Mesozoic elasmobranchs, neoselachian phylogeny, and the rise of modern neoselachian diversity. In: Arratia, G., Tintori, A. (Eds.), *Mesozoic Fishes 3-Systematics, Paleoenvironments and Biodiversity*. Verlag Dr. Friedrich Pfeil, München, pp. 17–56.
- Mallatt, J., Winchell, C.J., 2007. Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star. *Mol. Phylog. Evol.* 43, 1005–1022.
- Naylor, G.J.P., Ryburn, J.A., Fedrigo, O., Lopez, J.A., 2005. Phylogenetic relationships among the major lineages of modern elasmobranchs. In: Hamlett, W.C. (Ed.), *Reproductive Biology and Phylogeny of Chondrichthyes: Sharks, Batoids and Chimaeras*. Science Publishers, Enfield, New Hampshire, pp. 1–25.
- Naylor, G.J.P., Caira, J.N., Jensen, K., Rosana, K.A.M., White, W.T., Last, P.R., in press. A sequence based approach to the identification of shark and ray species and its implications for global elasmobranch diversity and parasitology. *Bull. Am. Mus. Nat. Hist.*
- Nelson, J.S., 1976. *Fishes of the World*. Wiley, New York.
- Nelson, J.S., 1984. *Fishes of the World*. Wiley, New York.
- Nelson, J.S., 2006. *Fishes of the World*. John Wiley and Sons, Inc., New York.
- Nylander, J.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53, 47–67.
- 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., Delsuc, F., Brinkmann, H., Lartillot, N., 2005a. Phylogenomics. *Annu. Rev. Ecol. Syst.* 36, 541–562.
- Philippe, H., Lartillot, N., Brinkmann, H., 2005b. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* 22, 1246–1253.
- Phillips, M.J., Penny, D., 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylog. Evol.* 28, 171–185.
- Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* 21, 1455–1458.
- Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrowska, O., Lee, J., Kent, L., Rest, J., Estabrook, G.F., Hendry, T.A., Taylor, D.W., Testa, C.M., Ambros, M., Crandall-Stotler, B., Duff, R.J., Stech, M., Frey, W., Quandt, D., Davis, C.C., 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. USA* 103, 15511–15516.
- Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzter, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 463, 1079–1083.
- Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.
- Rosenberg, M.S., 2005. MySP: non-stationary evolutionary sequence simulation, including indels. *Evol. Bioinform. Online* 1, 81–83.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- Shirai, S., 1992. *Squalan Phylogeny, a New Framework of "Squaloid" Sharks and Related Taxa*. Hokkaido University Press, Sapporo.
- Shirai, S., 1996. Phylogenetic interrelationships of neoselachians (Chondrichthyes: Euselachii). In: Stiassny, M.L.J., Parenti, L.R., Johnson, G.D. (Eds.), *Interrelationships of Fishes*. Academic Press, San Diego, pp. 9–34.
- Stamatakis, A., Ludwig, T., Meier, H., 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21, 456–463.
- Swofford, D.L., 2003. *PAUP\*. Phylogenetic Analysis Using Parsimony (\* and Other Methods)*. Version 4 Sinauer Associates, Sunderland, Massachusetts.
- 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.
- Tarrio, R., Rodriguez-Trelles, F., Ayala, F.J., 2000. Tree rooting with outgroups when they differ in their nucleotide composition from the ingroup: the *Drosophila* saltans and willistoni groups, a case study. *Mol. Phylog. Evol.* 16, 344–349.
- Wheeler, W.C., 1990. Nucleic acid sequence phylogeny and random outgroups. *Cladistics* 6, 363–367.
- Winchell, C.J., Martin, A.P., Mallatt, J., 2004. Phylogeny of elasmobranchs based on LSU and SSU ribosomal RNA genes. *Mol. Phylog. Evol.* 31, 214–224.
- Woese, C.R., Achenbach, L., Rouviere, P., Mandelco, L., 1991. Archaeal phylogeny: reexamination of the phylogenetic position of *Archaeoglobus fulgidus* in light of certain composition-induced artifacts. *Syst. Appl. Microbiol.* 14, 364–371.
- Wolf, Y.I., Rogozin, I.B., Koonin, E.V., 2004. Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res.* 14, 29–36.