Systematic Biology (2021), **0**, 0, *pp.* 1–**??**
doi:10.1093/sysbio/output

An Evaluation of FastME and IQ-TREE on an Avian Data Set

Olasubomi Efuniyi
Department of Computer Science, University of Maryland - College Park, College Park, 20740, USA

oefuniyi@umd.edu

Abstract

FastME and IQ-TREE represent two different approaches to estimating phylogenetic trees. As a distance method, FastME is generally faster than IQ-TREE, a Maximum Likelihood (ML) method. This work evaluates the reliability of and the similarity between the tree topologies generated by both approaches in the context of the time it took to generate them. It uses bootstrap branch support values and bipartitions to perform these evaluations, building trees from avian intron and exon sequences respectively. Additionally, this work evaluates the improvement in tree topology reliability achieved when a FastME estimated tree is further explored by IQ-TREE to find a higher likelihood tree. The results suggest that the longer time taken by IQ-TREE for longer length sequences pays off in the form of a sizable increase in the reliability of its result tree compared to FastME. Similar gains and time costs are observed when a FastME tree is processed by IQ-TREE to produce a higher likelihood tree.

Key words: FastME, IQ-TREE, Distance Methods, Maximum Likelihood, Bootstrap Branch Support, Bipartitions

Maximum Likelihood (ML) is an important tool for estimating the evolutionary history of sequences from a common ancestor sequence. Given a set of sequences (e.g.,

avian exon sequences) D, the goal of Maximum Likelihood is to find a model tree T, and its edge weights θ, such that (T, θ) maximizes the log likelihood function:

$$\log L(T, \theta|D) = \Sigma_{d \in D} \log p(d|T, \theta)$$

However, the ML problem is NP-hard for simple models of evolution (Roch 2006). Therefore, most ML implementations estimate a starting tree using a distance method or maximum parsimony, and from this heuristic, search for a tree with higher likelihood. The tree search phase can be computationally intensive depending on multiple factors like the size of the neighborhood explored for different search moves (e.g., Nearest Neighbor Interchange (NNI) and Subtree Pruning and Regraft (SPR)) and the stopping criteria.

Distance methods are also useful in estimating phylogenetic trees from a set of sequences. They work by computing a dissimilarity matrix D where the rows and columns are labelled by the name of the sequences under consideration, and entry $D[i, j]$ is the distance between sequences i and j. The distance between two sequences i and j is the expected number of nucleotide substitutions per site (Nahkleh). The distance entry $D[i, j]$ can simply be a p-distance calculated by counting how many positions the sequences i and j differ, then dividing by their sequence length (the sequences are assumed to have gone through a multiple sequence alignment, and therefore have the same sequence length). However, p-distances can be an underestimate due to back and parallel substitutions (Nahkleh). To account for this, a probabilistic model of site evolution is adopted to correct the p-distances using a correction formula for that model. Therefore, the model choice affects the accuracy of the result. Furthermore, starting with an inaccurate multiple sequence alignment will lead to an unreliable dissimilarity matrix. Once the dissimilarity matrix is computed, algorithms like Neighbor Joining and Unweighted Pair Group Method with Arithmetic Mean can be used to transform the dissimilarity matrix into a phylogenetic tree. Not withstanding, distance methods are computationally tractable, which makes them an appealing method for phylogenetic tree estimation.

Given that ML and distance methods have their trade-offs, this work attempts to

compare their accuracy using the bootstrap branch support value of the inner nodes in the phylogenetic trees they estimate. This comparison is carried out using aligned avian exon and intron data sets. The three approaches under consideration are FastME for distance methods and IQ-TREE with both a Maximum Parsimony start tree (IQ-TREE-MP) and a FastME start tree (IQ-TREE-FME).

Methods

Setup

This work uses data produced by the avian phylogenomics project (Jarvis 2015). The project published multiple sets of avian genome data, including a protein-coding exon gene set, a protein amino acid alignment set, and an intron gene set to name a few. A follow up to this work that provides the data from the project, separated into intron and exon data sets, and aligned in a multiple sequence alignment, was used. It can be found here.

The exon and intron data sets included many fasta files with the number of sequences per file in the range of 38 to 48 sequences for introns, and 42 to 48 sequences for exons. Furthermore, the sequence lengths varied from 58 to 38,848 base pairs for introns, and 99 to 15,777 base pairs for exons (https://github.com/Naturhistoriska/birdscanner/blob/master/doc/Jarvis_et_al_2015/README.md). The python script *max_length_sequences_finder.py* was used to identify the exon and intron alignments with the longest sequences. These were selected to give FastME and IQ-TREE a better chance of accurately estimating the model tree topology, as both approaches theoretically perform better on longer sequences.

Next, the selected fasta files were converted into a phylip file format because the FastME implementation used only accepted phylip file formats for aligned DNA sequences. A conversion tool provided by Laboratoire d'Informatique, de Robotique et de

Microélectronique de Montpellier (found here) was used to perform the conversion. The selected fasta files can be found in the folders $exon\_fasta\_files$ and $intron\_fasta\_files$, and the corresponding phylip files can be found in the folders $exon\_phylip\_files$ and $intron\_phylip\_files$ respectively. A taxon translation table can be found in Table 1. All tables and figures are located in the Appendix section.

Estimating Phylogenetic Trees with FastME

The FastME implementation used can be found here. It was run with 100 bootstrap re-sampling, and the consensus trees produced are shown in Figure 1 and Figure 2 for the exon and intron alignments respectively. The F84 substitution model was used because it was the recommended substitution model, and the default gamma shape parameter of 1 was maintained. For the tree building phase, the recommended balanced taxon addition algorithm (the TaxAdd_BME option) was used without a starting tree topology. Finally, for the tree refinement, Subtree Prunning and Regraft (SPR) was used because it was stated in the user guide to generally find better tree topologies compared to Nearest Neighbor Interchange (NNI)(http://www.atgc-montpellier.fr/fastme/usersguide.php).

Estimating Phylogenetic Trees with IQ-TREE using Maximum Parsimony Start Trees

IQ-TREE was used to estimate ML phylogenetic trees from the converted intron and exon sequence files using the command $./iqtree\ -s < path\_to\_phylip\_file > -b\ 100\ -nt\ 4$ from the iqtree binary directory. It performed 100 bootstrap re-sampling to produce a consensus tree. The resulting consensus trees are shown in Figure 3 and Figure 4.

The starting trees were computed with Maximum Parsimony for both data sets. The best fit models of DNA evolution found and used by IQ-TREE based on a Baysian Information Criterion were GTR+F+R3 for the exon data set, and GTR+F+R4 for the intron data set.

Estimating Phylogenetic Trees with IQ-TREE using FastME Start Trees

The command $./iqtree\ -s < path\_to\_phylip\_file > -t < path\_to\_starting\_tree\_file > -b\ 100\ -nt\ 4$ was used to search for higher likelihood phylogenetic trees for both data sets using their FastME results as the start trees. GTR+F+R4 was found to be the best fit model of DNA evolution for both data sets. The resulting consensus trees are shown in Figure 5 and Figure 6.

Results and Discussion

The python script *analysis.py* was used to find the percentage of internal nodes in the generated consensus trees that have a bootstrap support value of 75 or more. This work uses a cut-off of 75 because it was found to be a generally acceptable cut-off value, although this is debatable. The results are summarized in Table 2.

According to the table, for the shorter length exon sequences, FastME and IQ-TREE-MP generate trees with a comparable amount of reliable inner nodes (which can be considered as clades in this discussion). Furthermore, IQ-TREE-FME generated about 6% more reliable clades than the other two approaches. One reason for this could be IQ-TREE-FMEs use of the same start tree for all the bootstrap replicates, whereas IQ-TREE-MP computes a new MP start tree for each replicate. To investigate this, a Maximum Parsimony tree was generated for the exon data set using MPBoot, a Maximum Parsimony tree inference tool found on the IQ-TREE website. This was then used as the start tree to IQ-TREE to ensure that it used the same MP start tree every time. The initial MP tree had 43% reliability, although this was computed using ultrafast bootstrap with 1000 replicates because regular bootstrap was not available. The final consensus tree produced by IQ-TREE also had 43% reliability, although the both trees had 11 different non trivial bipartitions. Therefore, using the same MP start tree for each bootstrap replicate results in a more reliable final tree than IQ-TREE-MP. However this increase is

artificial since it is the result of a tree selection step that needs further justification. More work needs to be done to understand why IQ-TREE handles start trees the way it does, and the consequences of such decisions on the bootstrap support values of the resulting consensus tree. It should also be noted that an attempt was made to access the initial MP start tree generated by IQ-TREE-MP, and feed that back into a new run as the start tree, to investigate its behavior. However, this was not found to be possible.

For the longer length intron sequences, IQ-TREE-MP provides 39% more reliable clades than FastME. Furthermore, its clades are equally as reliable as IQ-TREE-FME. Interestingly, FastME produces 7% less reliable clades with the longer length intron sequences than it did with the shorter length exon sequences, while both IQ-TREE-MP and IQ-TREE-FME see large increases in their clade reliability.

Further analysis of the trees (see Table 3 and 4) using *analysis.py* shows that for the shorter length exon sequences, all three approaches produced similar trees in terms of the number of bipartitions they have in common. However, for the longer length intron sequences, the similarity between FastME and both approaches to IQ-TREE drops, while the latter produces the same trees.

Put together, the results suggest that IQ-TREE and FastME produce considerably different trees in terms of reliability and shared bipartitions as the sequence length of the input data increases, with IQ-TREE producing more reliable trees than FastME. Furthermore, there does not appear to be any difference in tree reliability between IQ-TREE-MP and IQ-TREE-FME when the sequence length is long enough. In this case, there is no observed tree reliability benefit to using a FastME produced start tree as opposed to a MP produced start tree. However, there is a remarkable improvement in the reliability of the FastME start tree when it was used.

The time taken by the three approaches can be found in Table 5. For the shorter length exon sequences, FastME was the fastest, with an approximately 48 minute gap between it and IQ-TREE-MP. IQ-TREE-FME took the longest time. Contrary to initial

expectations, using the FastME start tree did not save time in the ML tree search phase, even though it meant that IQ-TREE did not have to search for a MP start tree for each bootstrap replicate. Upon inspecting the logs, it was found that it only took about 0.013 seconds to find the MP start tree, which would explain why in the approximate time reported in Table 2, there was no time gained. However, it is not clear why it took 10 minutes longer with the FastME start tree. We may be observing the result of this longer process in the increased reliability of the resulting tree recorded in Table 1.

Furthermore, there was a large time difference of about 14 hours between FastME and IQ-TREE for the longer length intron sequences. Considering this with the data from Table 1, we payed 14 more hours to IQ-TREE to get 39% more reliability in our consensus tree. Also, since it took a few minutes to generate a FastME start tree to feed into IQ-TREE, IQ-TREE-FME took a bit longer and was more tedious than IQ-TREE-MP, but the result was only slightly better than IQ-TREE-MP for the shorter length exon sequences, and the same for the longer length intron sequences.

At the beginning of the ML process, IQ-TREE performs a chi-squared test to check for homogeneity of character composition, and a sequence is denoted "failed" if its character composition significantly deviates from the average character composition of the alignment (http://www.iqtree.org/doc/Frequently-Asked-Questions). The formula used for this calculation is:

$\chi^2 = \sum_{i=1}^{k}(O_i - E_i)^2/E_i$ where $k = 4$, the number of nucleotides, O_i is the frequency of nucleotide i in the sequence under consideration, and E_i is the frequency of the nucleotide i in the entire alignment. For the exon sequences, 24 of the 45 sequences failed the composition chi-square test, while 18 of the 46 intron sequences failed. The failure rate was the same for both start tree options.

Gaps are not directly considered in this calculation. However, they form part of the sequence length, parts that may be filled with nucleotides in the same position for other sequences. As such, we may expect that the more gaps a sequence has, the more likely it is

to fail the chi-squared test since the above sum will be greater, making it a more extreme value in the chi-square distribution. However, an investigation into the the frequency of gaps in each sequence in the intron data set did not support this. Geospiza fortis had the most gaps of all the sequences in this data set and indeed failed the test. However, Tinamus guttatus which had the next highest number of gaps did not fail the test, while some of the other sequences with far less gaps did.

Another idea was that the sequences that were closer to the average gap count, assuming nucleotide composition homogeneity, would pass the chi-squared test since the frequency of the non gap entries will also be around the average (expected) value. However, the average gap count in the intron data set was 6931, yet sequences close to this count, like Buceros rhinoceros with a count of 6982, failed the test. The failures may therefore be due to reported heterogeneity of nucleotide composition among Aves(Symonová and Suh, 2019).

Conclusion

Although FastME has its place in phylogeny estimation, the results of this work suggest that on long sequences of avian biological data, the reliability of its estimated tree pales in comparison with the tree estimated by IQ-TREE, both when using a Maximum Parsimony start tree and the FastME result as the start tree. These gains come at a very significant time cost which appears to be worthwhile based on the highly reliable result.

The results of this work can be contested on multiple grounds. First, bootstrap branch support as proposed by Felsenstein has been criticized in the genetics literature as being biased (Hillis 1993). However, since modifications have been proposed to make Falsenstein's approach "better agree with standard ideas of confidence levels and hypothesis testing" (Efron 1996), the same criticism does not apply to every bootstrap approach. More work needs to be done to understand how FastME and IQ-TREE approach bootstrapping.

Additionally, the methodology should be further scrutinized. One observation was that IQ-TREE used the same start tree for each bootstrap replicate when given a start tree as input, whereas it computed a new Maximum Parsimony start tree for each bootstrap replicate when a start tree was not given. This might have had a positive effect on bipartition repetition in the trees produced by each bootstrap replicate, thereby artificially increasing the bootstrap support for some internal nodes in the final consensus tree. More work should be done to understand what happens under the hood of IQ-TREE in this scenario.

Supplemental Materials

The sripts and data files mentioned in this report can be found here.

Acknowledgements

I would like to thank Professor Erin Molloy of the University of Maryland - College Park for inspiring this project and providing useful technical knowledge to improve the quality of the findings.

References

Efron B., Halloran E., Holmes S. 1996. Bootstrap confidence levels for phylogenetic trees. National Academy of Science. Volume 93. Number 23. Pages 13429 - 13429. https://www.pnas.org/content/93/23/13429

Hillis D. M., Bull J. J. 1993. An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. Systematic Biology. Volume 42. Number 2. Pages 182 – 192. https://doi.org/10.2307/2992540

Hoang D.T., Vinh L.S., Flouri T., Stamatakis A., von Haeseler A., and Minh B.Q. 2018.

MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. BMC Evolutionary Biol. Volume 18. Issue 1. https://doi.org/10.1186/s12862-018-1131-3

Jarvis E.D., Mirarab S., Aberer A.J. et al. 2015. Phylogenomic analyses data of the avian phylogenomics project. GigaScience. Volume 4. Issue 1. https://doi.org/10.1186/s13742-014-0038-1

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods. Volume 14. Pages 587 - 589. https://doi.org/10.1038/nmeth.4285

Lefort V., Desper R., Gascual O. 2015. FastME 2.0 - A comprehensive, accurate and fast distance-based phylogeny inference program. Molecular Biology and Evolution. Volume 32. Number 10. Pages 2798 - 2800.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. Volume 37. Issue 5. Pages 1530 - 1534. https://doi.org/10.1093/molbev/msaa015

Nahkleh. https://www.cs.rice.edu/ nakhleh/COMP571/Slides/Phylogenetics-DistanceMethods-Full.pdf

Roch S. 2006. A short proof that phylogenetic tree reconstruction by maximum likelihood is hard. IEEE/ACM Transactions on Computational Biology and Bioinformatics. Volume 3. Number 1. Pages 92-94.

Symonová R., Suh A. 2019. Nucleotide composition of transposable elements likely contributes to AT/GC compositional homogeneity of teleost fish genomes. Mobile DNA. Volume 10. Issue 1. https://doi.org/10.1186/s13100-019-0195-y

Appendix

Fig. 1. Phylogenetic tree estimated by FastME for 45 avian EXON sequences, each of length 15,777. Figure was generated using ETE Toolkit

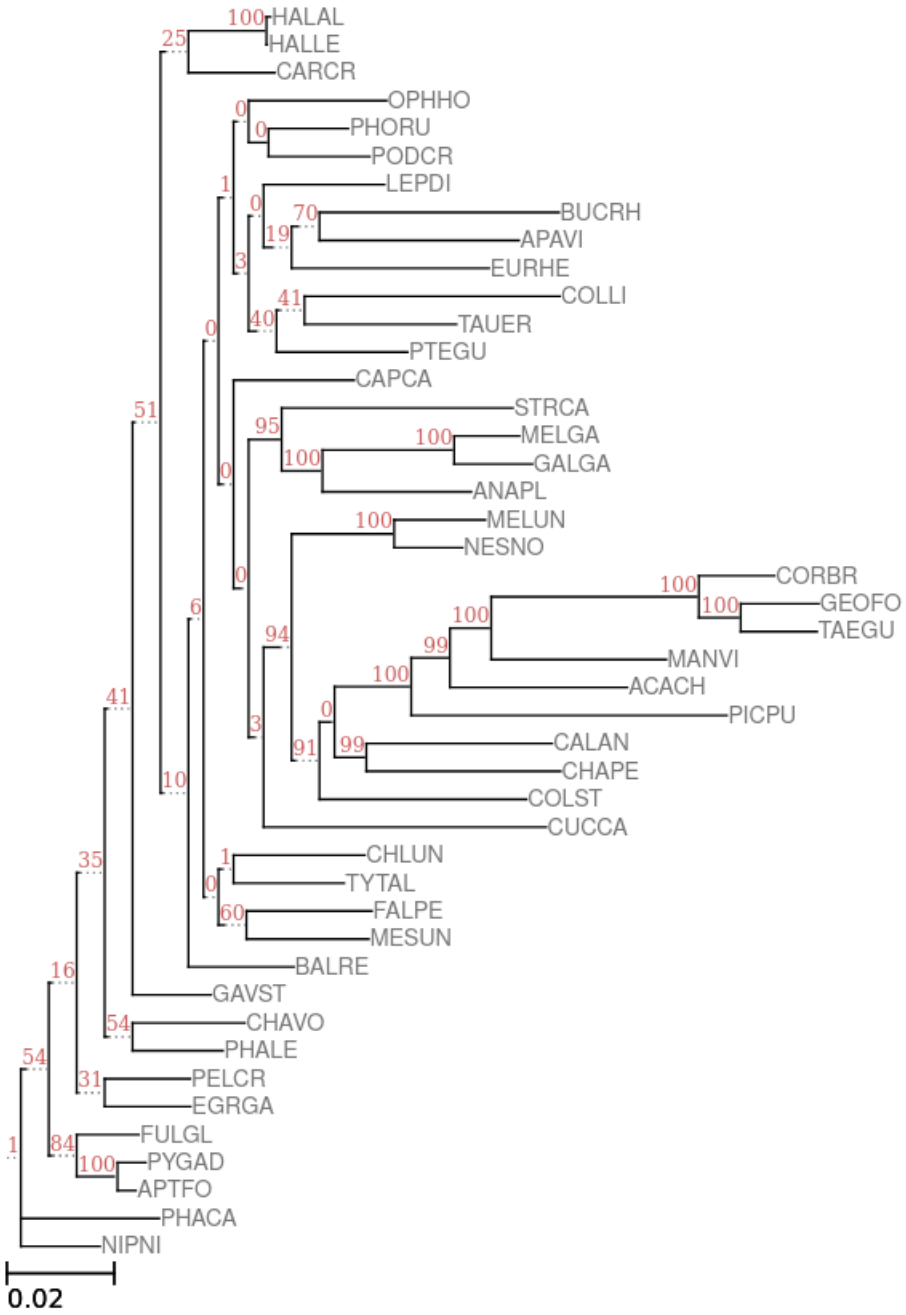

Fig. 2. Phylogenetic tree estimated by FastME for 44 avian INTRON sequences, each of length 38,848. Figure was generated using ETE Toolkit

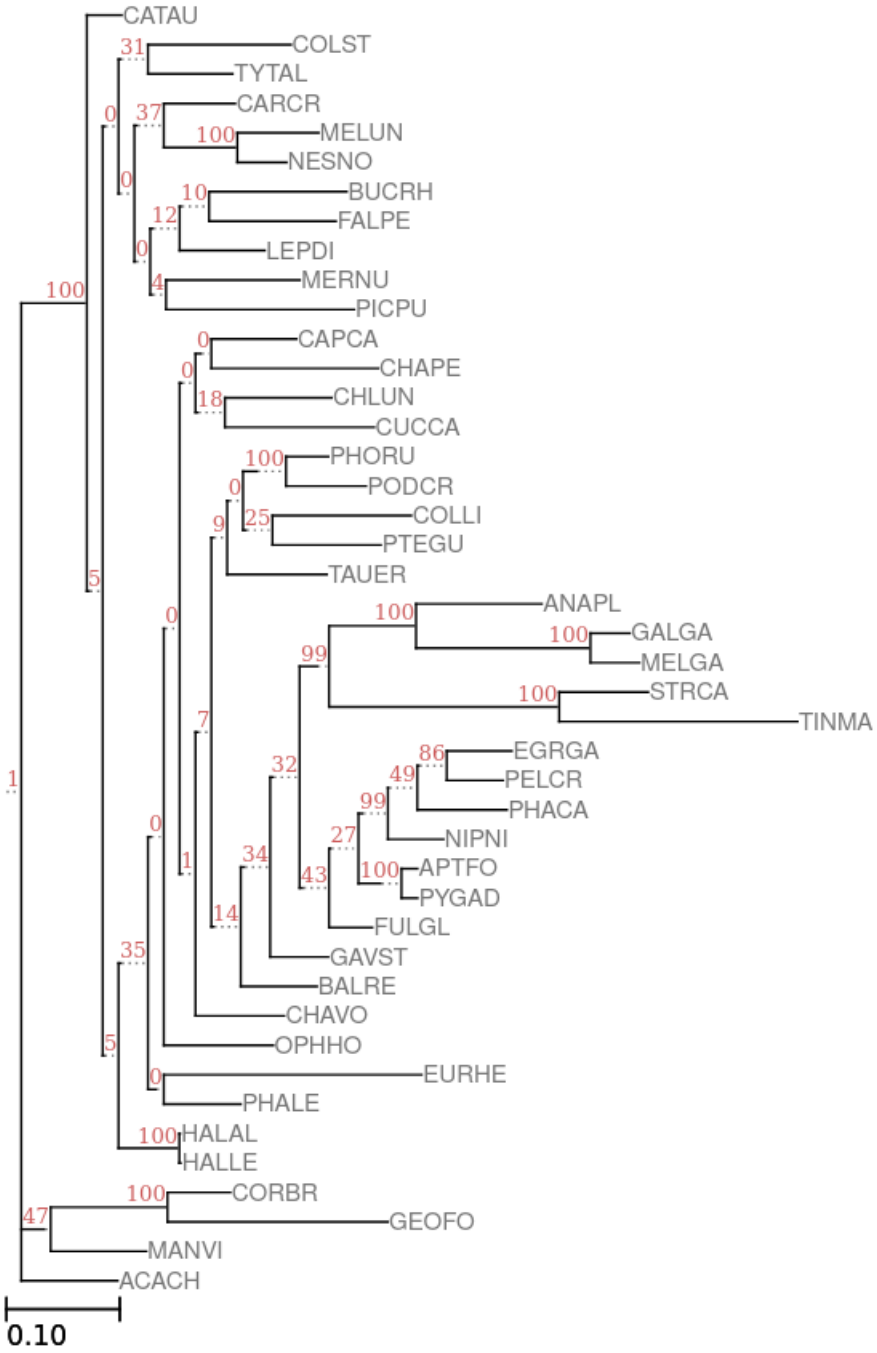

Table 1: Taxon Translation Table

Beginning of Table		
Taxa	**Family**	**Species**

Continuation of Table 1		
Taxa	**Family**	**Species**
ACACH	Acanthisittidae	Acanthisitta chloris
HALAL	Accipitridae	Haliaeetus albicilla
HALLE	Aciipitridae	Haliaeetus leucocephalus
ANAPL	Anatidaei	Anas platyrhynchos
CHAPE	Apodidaei	Chaetura pelagica
EGRGA	Ardeidaei	Egretta garzetta
BUCRH	Bucerotidae	Buceros rhinoceros
CAPCA	Caprimulgidae	Antrostomus carolinensis
CARCR	Cariamidae	Cariama cristata
CATAU	Cathartidae	Cathartes aura
CHAVO	Charadriidae	Charadrius vociferus
COLST	Coliidae	Colius striatus
COLLI	Columbidae	Columba livia
CORBR	Corvidae	Corvus brachyrhynchos
CUCCA	Cuculidae	Cuculus canorus
EURHE	Eurypygidae	Eurypyga helias
FALPE	Falconidae	Falco peregrinus
GAVST	Gaviidae	Gavia stellata
BALRE	Gruidae	Balearica regulorum
LEPDI	Leptosomidae	Leptosomus discolor
MERNU	Meropidae	Merops nubicus
MESUN	Mesitornithidae	Mesitornis unicolor
TAUER	Musophagidae	Tauraco erythrolophus
OPHHO	Opisthocomidae	Opisthocomus hoazin
CHLUN	Otididae	Chlamydotis macqueenii
TAEGU	Passeridae	Taeniopygia guttata
PELCR	Pelecanidae	Pelecanus crispus
PHALE	Phaethontidae	Phaethon lepturus
PHACA	Phalacrocoracidae	Phalacrocorax carbo
GALGA	Phasianidae	Gallus gallus
MELGA	Phasianidae	Meleagris gallopavo
PHORU	Phoenicopteridae	Phoenicopterus ruber
PICPU	Picidae	Picoides pubescens
MANVI	Pipridae	Manacus vitellinus
PODCR	Podicipedidae	Podiceps cristatus
FULGL	Procellariidae	Fulmarus glacialis
MELUN	Psittacidae	Melopsittacus undulatus
NESNO	Psittacidae	Nestor notabilis
PTEGU	Pteroclidae	Pterocles gutturalis
APTFO	Spheniscidae	Aptenodytes forsteri
PYGAD	Spheniscidae	Pygoscelis adeliae

Continuation of Table 1		
Taxa	**Family**	**Species**
STRCA	Struthionidae	Struthio camelus
GEOFO	Thraupidae	Geospiza fortis
NIPNI	Threskiornithidae	Nipponia nippon
TINMA	Tinamidae	Tinamus guttatus
CALAN	Trochilidae	Calypte anna
APAVI	Trogonidae	Apaloderma vittatum
TYTAL	Tytonidae	Tyto alba

Table 2. Percentage of Internal Nodes with Bootstrap Support Value >= 75

Approximate Percentage (%)	FastME	IQ-TREE (with Maximum Parsimony)	IQ-TREE (with FastME)
Exon Sequences of Length 15,777	36%	37%	43%
Intron Sequences of Length 38,848	29%	68%	68%

Table 3. Fraction of bipartitions in common for the EXON data set, trivial bipartitions. An "x" means the number was skipped because it is irrelevant. The denominator is the number of bipartitions in the tree produced by the method of the row in question.

Fraction of bipartitions in Common	FastME	IQ-TREE (with MP)	IQ-TREE (with FastME)
FastME	x	25/42	26/42
IQ-TREE (with MP)	25/41	x	39/41
IQ-TREE (with FastME)	26/42	39/42	x

Fig. 3. Phylogenetic tree estimated by IQ-TREE-MP for 45 avian EXON sequences, each of length 15,777. Figure was generated using ETE Toolkit

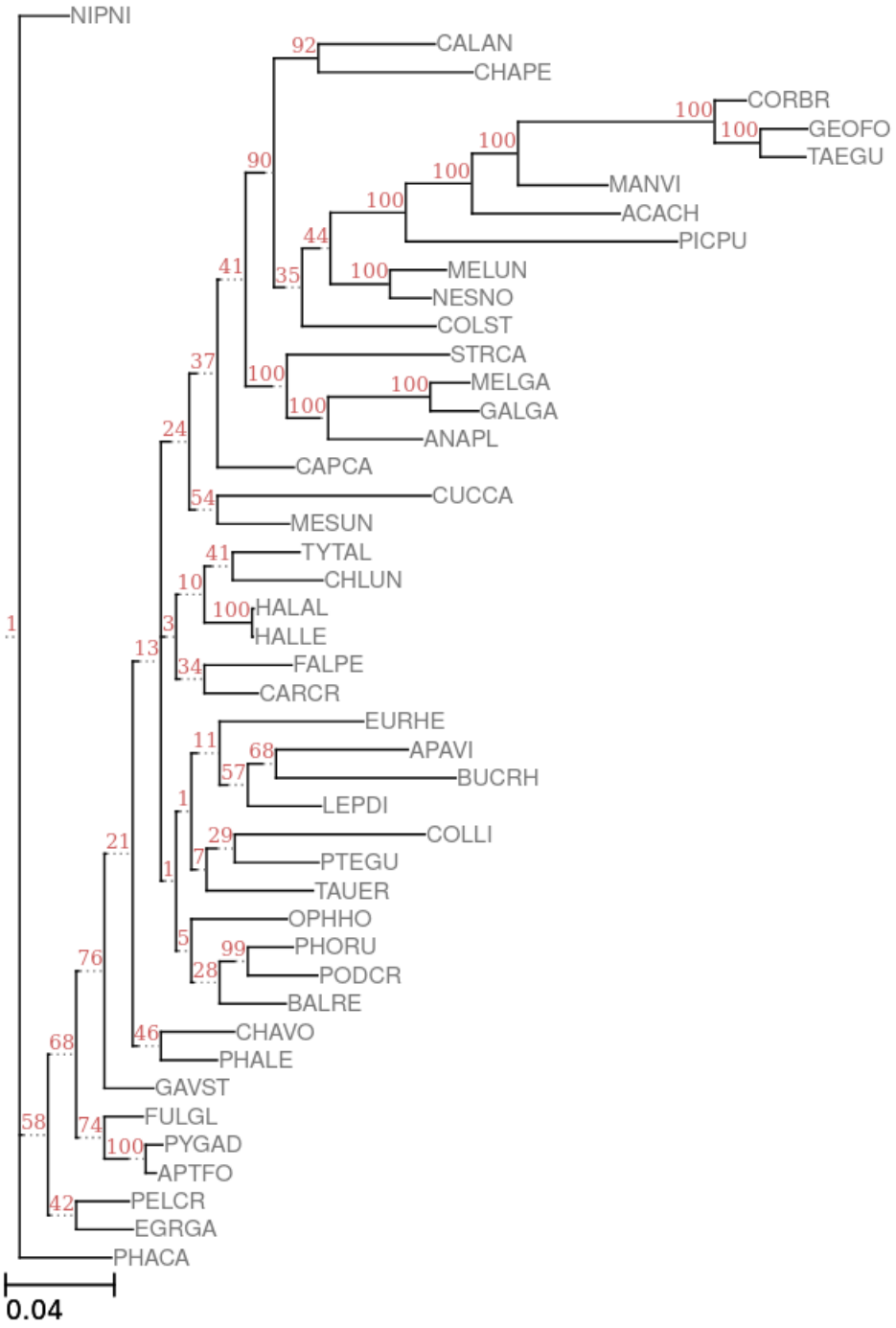

Fig. 4. Phylogenetic tree estimated by IQ-TREE-MP for 44 avian INTRON sequences, each of length 38,848. Figure was generated using ETE Toolkit

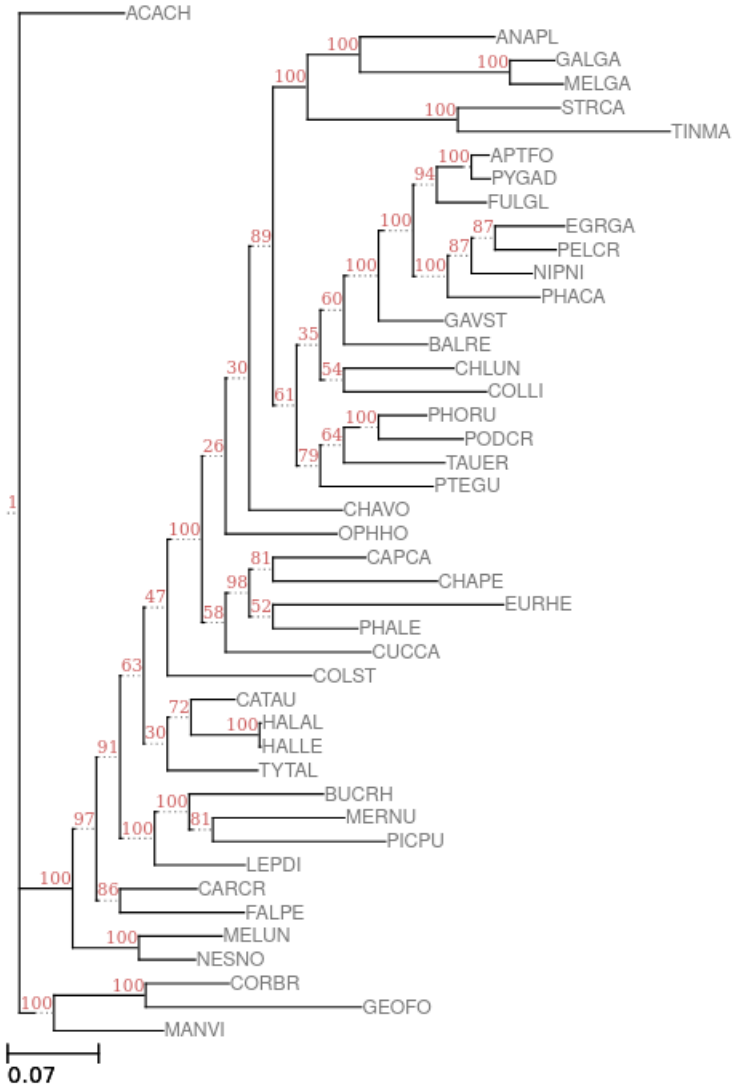

Fig. 5. Phylogenetic tree estimated by IQ-TREE-FME for 45 avian EXON sequences, each of length 15,777. Figure was generated using ETE Toolkit

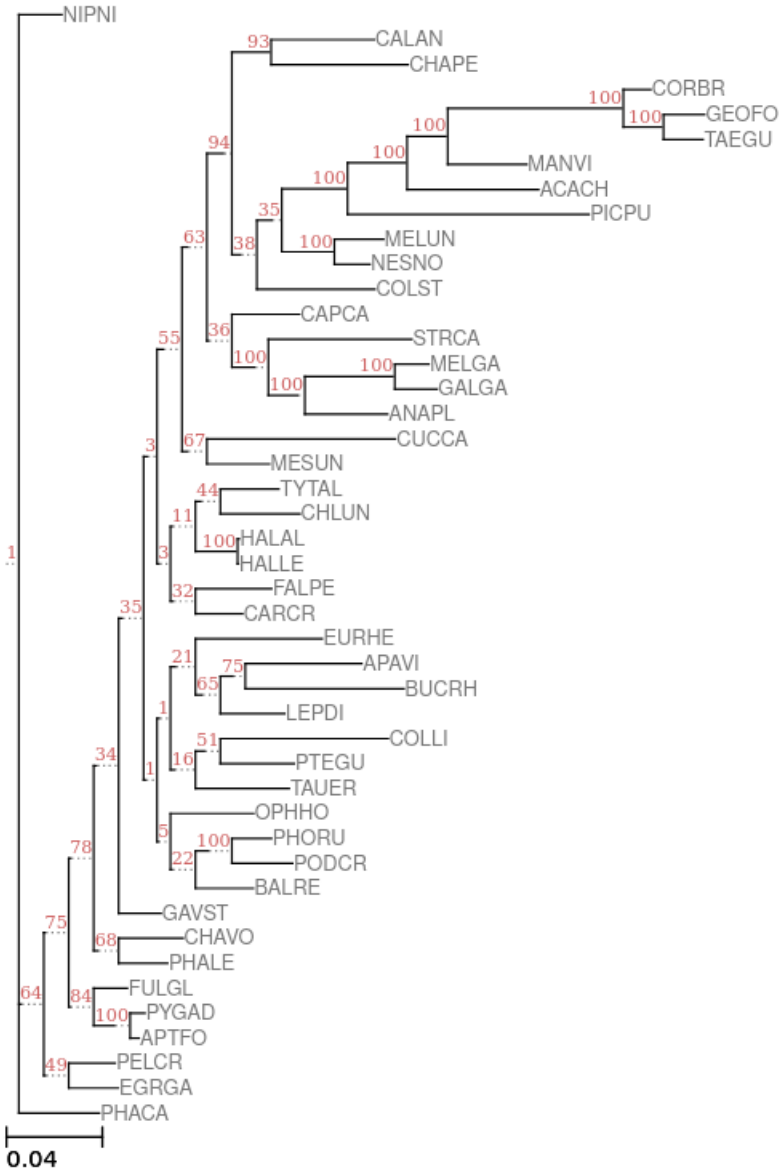

Fig. 6. Phylogenetic tree estimated by IQ-TREE-FME for 44 avian INTRON sequences, each of length 38,848. Figure was generated using ETE Toolkit

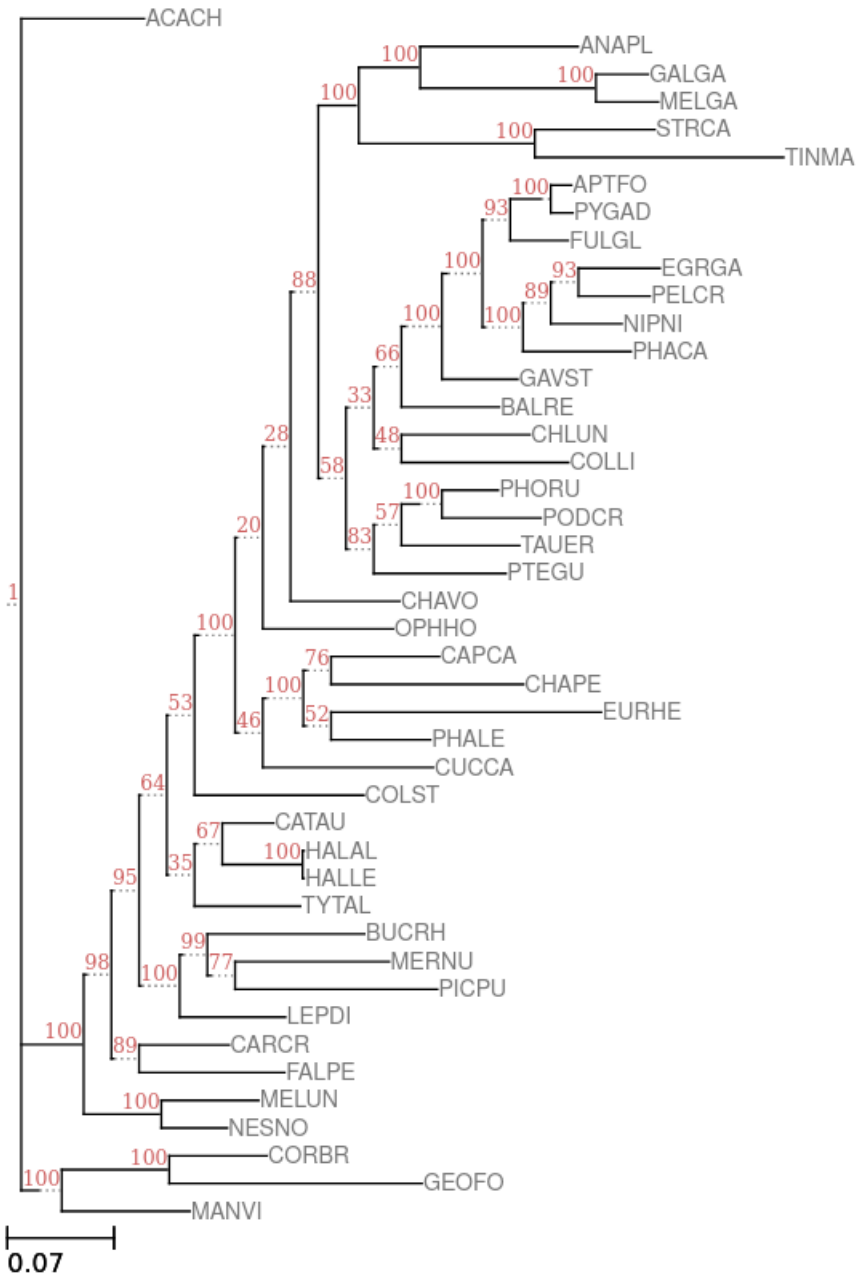

Table 4. Fraction of bipartitions in common for the INTRON data set, excluding trivial bipartitions. An "x" means the number was skipped because it is irrelevant or shown elsewhere in the table. The denominator is the number of bipartitions in the tree produced by the method of the row in question.

Fraction of Clades in Common	FastME	IQ-TREE (with MP)	IQ-TREE (with FastME)
FastME	x	18/41	18/41
IQ-TREE (with MP)	x	x	41/41
IQ-TREE (with FastME)	x	x	x

Table 5. Time taken to build consensus trees. FastME was run on a remote server, while IQ-TREE was run on a intel i5 12 core computer using 4 threads.

Approximate Time	FastME	IQ-TREE (with Maximum Parsimony)	IQ-TREE (with FastME)
Exon Sequences of Length 15,777	2 mins	50 mins	1 hr
Intron Sequences of Length 38,848	3 mins	14 hrs	14 hrs