

# The fossil record of sabre-tooth characins (Teleostei: Characiformes: Cynodontinae), their phylogenetic relationships and palaeobiogeographical implications

## Supplementary material

Gustavo A. Ballen\*      Jorge W. Moreno-Bernal†      Carlos Jaramillo‡

## Contents

<b>S1 Parsimony analysis</b>	<b>2</b>
S1.1 Phylogenetic analysis . . . . .	2
S1.1.1 Dependencies . . . . .	2
S1.1.2 Fast run . . . . .	2
S1.1.3 Input files . . . . .	2
S1.1.4 Output files . . . . .	3
S1.1.5 TNT tree format conversion . . . . .	4
S1.1.6 Final steps . . . . .	4
S1.2 <code>complete.run</code> . . . . .	4
S1.3 <code>complete.tnt</code> . . . . .	6
S1.4 <code>automaticPipeline.sh</code> . . . . .	6
S1.5 Results . . . . .	6
S1.5.1 Character definitions . . . . .	6
S1.5.2 List of apomorphies . . . . .	7
S1.6 Characters and dentary anatomy in representatives of the Cynodontidae . . . . .	9
<b>S2 Bayesian analysis</b>	<b>9</b>
S2.1 Taxon sampling . . . . .	9
S2.2 Marker selection . . . . .	9
S2.3 Sequence manual selection . . . . .	9
S2.3.1 <i>CytB</i> . . . . .	9
S2.3.2 <i>COI</i> . . . . .	10
S2.3.3 <i>ATPase6-8</i> . . . . .	10
S2.3.4 <i>rpS7</i> . . . . .	10
S2.3.5 <i>RAG2</i> . . . . .	10
S2.3.6 <i>16S</i> . . . . .	10
S2.3.7 <i>18S</i> . . . . .	10
S2.4 Sequence alignment . . . . .	10
S2.4.1 <i>COI</i> . . . . .	11
S2.4.2 <i>ATPase6-8</i> . . . . .	11
S2.4.3 <i>16S</i> . . . . .	11
S2.4.4 <i>CytB</i> . . . . .	11

\*Museu de Zoologia da Universidade de São Paulo, Smithsonian Tropical Research Institute, Queen Mary University of London

†Smithsonian Tropical Research Institute and Universidad del Norte

‡Smithsonian Tropical Research Institute, U. Montpellier, and University of Salamanca

S2.4.5	<i>RAG2</i>	.....	11
S2.4.6	<i>RAG1</i>	.....	11
S2.4.7	<i>Myh6</i>	.....	11
S2.4.8	<i>rpS7</i>	.....	11
S2.4.9	<i>18S</i>	.....	11
S2.5	Accession numbers	.....	12
S2.6	Substitution models	.....	13
S2.7	MrBayes script	.....	17
	S2.7.1 Analysis output	.....	20

```

## Loading required package: phangorn
##
## Attaching package: 'seqinr'
## The following objects are masked from 'package:ape':
##   as.alignment, consensus

```

## S1 Parsimony analysis

### S1.1 Phylogenetic analysis

#### S1.1.1 Dependencies

Please note that in order to run these scripts and reproduce the results you will need:

- TNT v.1.5 or higher for phylogenetic analysis, support statistics calculation, and tree production. Freely available at <http://www.lillo.org.ar/phylogeny/tnt/>. Please keep in mind that TNT starts numbering from 0, see below.
- R v.3.3.1 or higher for tree format conversion. Freely available at <https://www.r-project.org/>.
- (Optional) FigTree v.1.4.2 or higher for tree edition
- (Optional) Inkscape 0.91 or higher for fine graphical edition of the trees.

#### S1.1.2 Fast run

On a UNIX system just run the `automaticPipeline` script. This can be achieved by running the command `bash automaticPipeline` or `bash ./automaticPipeline`. This script will return no visual output to the screen by itself but will show some internal TNT messages during bootstrapping. These will disappear from the screen after running. The tree file format conversion is completely silent. After running you will have the output files described below. This alternative is designed for fast runs without going through each individual file.

Further details and comments are found in each script for those interested.

#### S1.1.3 Input files

These files must be run following these instructions. These are specific for command-line TNT:

- Navigate to the path where the files are located,`cd PATH` for Unix users.
- Run TNT on the background with the command`tnt bground p complete.run`. Please note that the script was designed as a stand-alone tool so once called from TNT it will exit once the analysis has finished. See the `filecomplete.run` file for further comments and detailed explanation of each step in the phylogenetic analysis and construction of the output tree in `.tre` format.

- If there are already output files these will be rewritten by TNT with the respective warning being recorded in the `complete.out` output file. Otherwise such file should not contain any warning.

#### `complete.tnt`

This file contains the data matrix in tnt format. The penultimate line of the file contains the instruction `collapse [;` in order to collapse branches supported by no apomorphies, contrary to the default in TNT that always presents a dichotomous result.

#### `complete.run`

Script for carrying out the analysis, it contains the following steps as already documented inside the script:

- Tell TNT to use 1 Gb of RAM and store a large number of tree in memory.
- Save all output to `complete.out`
- Read the matrix `complete.tnt`
- Initialize the output tree file `complete.tre`
- Set outgroup to `Acestrorhynchus`
- Use implicit enumeration for exact search
- Plot the most-parsimonious tree (in this case, there was only one MPT) along with node numbers
- Initialize tree annotation
- Calculate CI and RI using the `STATSALL.RUN` script
- Calculate bootstrap statistics
- Calculate jackknife statistics
- Save annotations to tree file in parenthetical notation
- Plot MPT to log file
- Close tree file
- List synapomorphies for each clade. Please note that TNT starts numbering from 0, so you will need to edit the synapomorphies list by adding 1 to all node AND character numbers (e.g., Node 0 is actually node 1, and character 76 is actually character 77).
- Close log file
- Close TNT

#### `STATSALL.RUN`

This script was developed by Peterson Lopes (Universidade de São Paulo) and is available in a number of sources over the internet. For reproducibility purposes I am linking to a thread in the TNT user group on google where such file can be found. Unfortunately there seems to be no official source for this script since the TNT wiki website became unavailable a couple years ago. The script can be found here [https://groups.google.com/d/msg/tnt-tree-analysis-using-new-technology/qPdCzlk\\_at8/YusQvlXCahwJ](https://groups.google.com/d/msg/tnt-tree-analysis-using-new-technology/qPdCzlk_at8/YusQvlXCahwJ). I claim NO AUTHORSHIP for this script. Please contact Peterson Lopes directly for further information.

#### S1.1.4 Output files

The script `complete.run` generates the following files as already noted. Tree edition was done in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) and fine-edited with Inkscape.

#### `complete.out`

This file contains every output generated during analysis including graphic output such as trees and statistics. It will also contain any warning generated during analysis. Further information aside from graphic cladograms can be found here.

#### **complete.tre**

This file contains the output tree file to be converted to newick format in R (see below). Without any reformatting, this file can only be opened by TNT.

#### **S1.1.5 TNT tree format conversion**

An additional step is required before the `complete.tre` file can be edited in FigTree, given that TNT's output tree format differs from a newick standard format. An R script was designed for carrying out format conversion automatically.

In order to use the tree file you need to use the function `tbea::tnt2newick()` in the R console for conversion from TNT to newick format. Afterwards you are ready for converting between to newick format with the `treToNewick` function. Finally, you can open the .newick file in FigTree. Please keep in mind that the tree file has support values as annotations, so when opening FigTree you will be asked to name the labels (the default name labels is enough) so that you can activate them in the 'labels' section of FigTree.

You will need to install the `tbea` R package in order to use the function `tnt2newick`, load the package, and call the function:

```
devtools::install_github("gaballench/tbea")
library(tbea)
tnt2newick("my_TNT_tree_file.tre", "my_TNT_tree_file.newick")
```

Please see the documentation `help(tnt2newick)` in the R console for further details.

#### **S1.1.6 Final steps**

After obtaining a .newick file you will be ready to edit the tree in FigTree in order to add annotations, change fonts, colors, and some other basic edition tasks. For a fully edited and good-looking tree it is suggested to save the tree in the vectorial svg format and then to further edit it with Inkscape. This way the cladogram published in the paper can be obtained.

## **S1.2 complete.run**

```
macro =;

/* Set max RAM to 1 Gb*/
mxram 1000;

/* Collapse all unsupported branches*/
collapse [];

/*
Save all the output to the file 'complete.out'
It can be opened in any text editor
*/
log complete.out;

/*
Read the matrix 'complete.tnt'
*/
proc complete.tnt;

/*
Number of trees to be held in memory
*/
hold 1000000;

/*
Set random seed
*/
rseed 0;
```

```

/*
Open tree file in parenthetical notation
*/
tsave * complete.tre;

/*
Set the composite taxon 'Outgroup' as the root
*/
outgroup Acestrorhynchus;

/*
Given the number of taxa, carry out an exact search
with implicit enumeration
*/
ientum;

/*
Taxon names ON
*/
taxname =;

/*
Plot MPT with node numbers
*/
naked -;
tplot /;
naked =;

/*
Tree tags ON
*/
ttags =;

/*
Overall CI and RI
*/
run STATSALL.RUN 1;

/*
Bootstrap and Jackknife branch support values
*/
resample boot rep 10000;
resample jak rep 10000;

/*
Save tags to tree file in parenthetical notation
*/
save *;

/*
Plot consensus tree to log file
*/
tplot /;

/*
Close tree files
*/
tsave /;

/*
List synapomorphies common to all trees
*/
apo -;

/*
Close log file
*/
log /;

/*
Close TNT
*/
quit;

```

### S1.3 complete.tnt

In order to reconstruct the structure of the dataset, cut the character string (e.g., the line with **???? ... 111** just below `Hydrolycus_sp._Fossil`) and paste it after the previous line that contains the taxon name (e.g., `Hydrolycus_sp._Fossil` in this case). The line should then contain the structure: `taxon_name` and `character_string` separated by a space (e.g., `Hydrolycus_armatus 110210 ... 020`). The matrix includes the characters originally proposed by [Toledo-Piza \(2000\)](#) and the new ones herein proposed.

## S1.4 automaticPipeline.sh

```

#!/bin/bash

# run the TNT script 'complete.run' on the background. This will carry out the phylogenetic
# analysis and draw the annotated cladogram. Please refer to the file 'complete.run' for
# further details and comments.
tnt bground p complete.run

# Load the 'tbea' package in R and convert the 'complete.tre' output file to the newick format
# ('complete.newick'). Afterwards just open it in figtree or your preferred program for tree edition.
Rscript -e 'library(tbea)' -e 'tnt2newick("complete.tre", "complete.newick", subsetting = FALSE)'

```

## S1.5 Results

### S1.5.1 Character definitions

Character definitions and character matrix. New characters are numbered from 70 to 80, characters 1 to 69 are from [Toledo-Piza \(2000\)](#).

**Character 70**—Dentary, symphyseal teeth, number: 0=two; 1=one; 2=six to seven

**Character 71**—Dentary, leading canine, insertion plane: 0=labio-lingual; 1=commisuro-symphysisal; 2=straight, perpendicular to dentary horizontal plane.

**Character 72**—Dentary, leading canine, texture of lingual surface: 0=smoothly curve; 1=with a sharp and apically distinct cutting edge.

**Character 73**—Dentary, area posterior to tooth row, accessory teeth posterior to main tooth row: “-”=inapplicable; 0=absent; 1=extensive patch; 2=1-4 teeth in a restricted patch.

**Character 74**—Dentary-anguloarticular, coronoid process, presence: 0=absent; 1=present.

**Character 75**—Dentary, lateral sulcus on anterior portion below leading canine, presence: 0=absent; 1=present.

**Character 76**—Dentary, lateral depression between leading canine and next hypertrophied canine, presence: 0=absent; 1=present.

**Character 77**—Dentary, leading canine, base of enameloid, labial surface, texture: 0=smooth, 1=present in the form of parallel/oblique ridges.

**Character 78**—Dentary, strong canines between leading canine and next hypertrophied canine number: “-”=inapplicable as dentary teeth are of comparable length; 0=1 canine; 1=2 canines.

**Character 79**—Dentary, anterior margin in lateral view, outline; 0=oblique, 1=straight, 2=round.

**Character 80**—Dentary, canine between leading canine and next hypertrophied canine, relative position: 0=midway between canines, 1=contiguous to posterior canine.

### S1.5.2 List of apomorphies

List of synapomorphies for the most-parsimonious tree. Synapomorphies are listed as transformation series in the form ‘Character: plesiomorphic state → apomorphic state’. Please note that TNT starts numbering from 0 as is the case in computing in general; however, humans usually start counting from 1 as is the case when presenting characters. In order to preserve the numbering of characters in the literature we have added 1 to each of the characters as reported by TNT, so that character 0 for TNT is actually character 1 in our data matrix (and also character 1 in [Toledo-Piza \(2000\)](#)). Nodes are also numbered starting with 0 in TNT. Given that node numbers are never discussed in the manuscript and the fact that we always refer to clade composition (Cynodontidae is Node 16 but we always will say Cynodontidae instead of Node 16) we are not changing node numbers in the list below. These are given in order to aid the reader when checking the log file where the cladogram is plotted in text mode with node numbers.

<i>Hydrolycus</i> MUN 16211:	Char. 51: 1 → 2
No autapomorphies	Char. 52: 0 → 1
<i>Hydrolycus armatus</i> :	Char. 53: 1 → 0
No autapomorphies	Char. 59: 0 → 1
<i>Hydrolycus tatauaia</i> :	Char. 62: 1 → 2
No autapomorphies	Char. 65: 1 → 2
<i>Hydrolycus wallacei</i> :	Char. 67: 0 → 1
Char. 4: 0 → 1	Char. 69: 0 → 1
Char. 6: 0 → 1	Char. 75: 0 → 1
Char. 17: 1 → 0	
Char. 78: 1 → 0	<i>Cynodon gibbus</i> :
<i>Hydrolycus scomberoides</i> :	No autapomorphies
No autapomorphies	<i>Cynodon septenarius</i> :
<i>Rhaphiodon vulpinus</i> :	No autapomorphies
Char. 3: 0 → 1	<i>Acestrorhynchus</i> :
Char. 18: 0 → 2	No autapomorphies
Char. 45: 1 → 3	<i>Myloplus schomburgkii</i> :
Char. 47: 0 → 1	No autapomorphies
Char. 48: 0 → 1	<i>Serrasalmus rhombeus</i> :
Char. 49: 0 → 1	Char. 24: 0 → 1
	Char. 74: 0 → 1

Char. 79: 2 → 3	Char. 44: 0 → 1
<i>H. scomberoides</i> + H. MUN 16211 (Node 12):	Char. 45: 0 → 1
Char. 80: 0 → 1	Char. 50: 0 → 1
<i>Hydrolycus</i> exclusive of <i>H. wallacei</i> (Node 13):	Char. 51: 0 → 1
Char. 4: 0 → 2	Char. 53: 0 → 1
Char. 10: 0 → 1	Char. 54: 0 → 1
Char. 35: 0 → 1	Char. 55: 0 → 1
Char. 70: 0 → 1	Char. 56: 0 → 1
Char. 76: 0 → 1	Char. 57: 0 → 1
<i>Hydrolycus</i> (excl. <i>H. wallacei</i> ) + <i>Rhaphiodon</i> (Node 14):	Char. 58: 0 → 1
Char. 39: 0 → 1	Char. 60: 0 → 1
Char. 46: 0 → 1	Char. 61: 0 → 1
Char. 72: 0 → 1	Char. 62: 0 → 1
Char. 79: 0 → 1	Char. 79: 2 → 0
<i>H. wallacei</i> + <i>Rhaphiodon</i> + <i>Hydrolycus</i> (Node 15):	Cynodontidae + Serrasalmidae (Node 17):
Char. 11: 0 → 1	No unambiguous synapomorphies
Char. 12: 0 → 1	<i>H. armatus</i> + <i>H. tatauaia</i> (Node 18):
Char. 15: 0 → 1	Char. 1: 0 → 1
Char. 26: 0 → 1	Char. 17: 1 → 0
Char. 28: 0 → 1	Char. 20: 2 → 1
Cynodontidae (Node 16):	Char. 77: 0 → 1
Char. 2: 0 → 1	Char. 78: 1 → 0
Char. 8: 1 → 0	Char. 79: 1 → 2
Char. 14: 0 → 1	<i>Cynodon</i> (Node 19):
Char. 19: 0 → 1	Char. 3: 0 → 2
Char. 23: 0 → 1	Char. 9: 0 → 1
Char. 25: 0 → 1	Char. 18: 0 → 2
Char. 27: 0 → 1	Char. 22: 01 → 2
Char. 30: 0 → 1	Char. 24: 0 → 1
Char. 31: 0 → 1	Char. 42: 0 → 1
Char. 33: 0 → 1	Char. 46: 0 → 2
Char. 36: 0 → 1	Char. 49: 0 → 1
Char. 37: 0 → 1	Char. 59: 0 → 1
Char. 40: 0 → 1	Char. 64: 0 → 1
Char. 43: 0 → 1	Char. 74: 0 → 1
	Serrasalmidae (Node 20):

Char. 4: 0 → 1

Char. 34: 1 → 0

Char. 38: 1 → 0

## S1.6 Characters and dentary anatomy in representatives of the Cynodontidae

Figures S1 and S2 illustrate several of the conditions described in characters 70 to 80.

## S2 Bayesian analysis

### S2.1 Taxon sampling

Selection of sequences for *Acestrorhynchus* follows [Pretti et al. \(2009\)](#) that recovered *A. falcirostris* and *A. microlepis* as the successive most basal species in the genus.

### S2.2 Marker selection

A first batch of sequences was downloaded using the string:

```
"Cynodon gibbus"[Organism] OR  
"Hydrolycus"[Organism] OR  
"Rhaphiodon"[Organism] OR  
"Acestrorhynchus microlepis"[Organism] OR  
"Acestrorhynchus falcirostris"[Organism] AND animals[filter]
```

In addition, sequences for *Myloplus schomburgkii* and *Serrasalmus rhombeus* were retrieved in the same way:

```
"Myloplus schomburgkii"[Organism] OR  
"Serrasalmus rhombeus"
```

Next, UGENE was used in order to pick sequences per marker. Given the small original amount of sequences, a survey of representativeness per marker was carried out by picking markers in the sequence title. The following structure was found and allowed to pick a total of nine markers (number of sequences per species between parentheses):

C. gibbus	16S (1),	cytb (1),	Myh6 (1), NA (2), RAG1 (1), RAG2 (1)	
H. armatus	EGR1 (1), EGR2B (1), EGR3 (1), NA (2),		RAG1 (1),	RH (1)
H. scomberoides	mitogen.(2), 16S (3), 18S (1), COI (10), cytb (1),		Myh6 (1), RAG1 (1), RAG2 (2),	sina (1)
R. vulpinus	16S (4), 18S (1), ATP6 (1), COI (15), cytb (1), fkh (1), Myh6 (1), NA (2), RAG1 (1), RAG2 (1), rpS7 (1)			
A. falcirostris	16S (2), ATP8 (2), COI (6),			rpS7 (2)
A. microlepis	16S (4), 18S (2), ATP6-8 (2),			rpS7 (2)

markers selected: 16S, 18S, ATP6, COI, cytb, Myh6, RAG1, RAG2, rpS7, and the mitogenome for scomberoides in all mitochondrial cases

Posterior to this assessment, sequences of Serrasalmids were selected for most of these markers.

A script in R was used for renaming the sequences so that they started with the species name, then the marker name, and finally the accession number.

```
Rscript scripts/ renameAlignments.R
```

### S2.3 Sequence manual selection

A single sequence per species per marker was chosen whenever more than one was present. The following list records all the modifications needed:

#### S2.3.1 *CytB*

- *Hydrolycus\_scomberoides\_CytB\_NC\_015813.1* removed as the other mitogenomic sequence was identical.

- *Hydrolycus\_scomberoides\_CytB\_HQ289558.1* removed as the mitogenome already includes the sequence of CytB

#### S2.3.2 COI

- *Hydrolycus\_scomberoides\_COI\_NC\_015813.1* removed in coordination with the choice in CytB.
- All sequences for *Acestrorhynchus falcirostris* were identical. Picking the first one *Acestrorhynchus\_falcirostris\_COI\_NC\_015813.1*
- From the ten longest sequences of *Rhaphiodon vulpinus*, all were identical. Picking the first one *Rhaphiodon\_vulpinus\_COI\_GU701527.1*

#### S2.3.3 ATPase6-8

- Picking *Acestrorhynchus\_falcirostris\_ATPase6-8\_FJ468304.1* as it was the longest of both for this species.
- Removing *Hydrolycus\_scomberoides\_ATPase6-8\_NC\_015813.1* for the same reason as in other mitochondrial markers.
- Picking *Acestrorhynchus\_microlepis\_ATPase6-8\_FJ468311.1* as it was the longest of both for this species.

#### S2.3.4 rps7

- Picking *Acestrorhynchus\_falcirostris\_rps7\_FJ409851.1* as it was the longest for that species.
- Removing *Hydrolycus\_scomberoides\_rpS7\_NC\_015813.1* for the same reason as in other mitochondrial markers.
- Picking *Acestrorhynchus\_microlepis\_rpS7\_FJ409859.1* as it was the longest for that species.
- Renaming *Rhaphiodon\_sp.\_rpS7\_FJ409867.1* to *Rhaphiodon\_vulpinus\_rpS7\_FJ409867.1* as there is only one species of *Rhaphiodon*.

#### S2.3.5 RAG2

- Picking *Hydrolycus\_scomberoides\_RAG2\_AY804088.1* as it was the longest for that species.

#### S2.3.6 16S

- Picking *Hydrolycus\_scomberoides\_16S\_HQ171269.1* as above.
- Removing all other *H. scomberoides*.
- Picking *Acestrorhynchus\_microlepis\_16S\_FJ362546.1* as it is the longest sequence.
- Picking *Rhaphiodon\_vulpinus\_16S\_HQ171303.1* as it is the longest sequence.
- Picking *Acestrorhynchus\_falcirostris\_16S\_FJ362540.1* as it is the longest sequence.

#### S2.3.7 18S

- Picking *Acestrorhynchus\_falcirostris\_18S\_FJ362540.1* as it is the longest sequence.
- Picking *Hydrolycus\_scomberoides\_18S\_FJ944806.1* as above.
- Picking *Acestrorhynchus\_microlepis\_18S\_FJ362546.1* as it is the longest sequence.
- Picking *Rhaphiodon\_vulpinus\_18S\_HQ171303.1* as it is the longest sequence.

### S2.4 Sequence alignment

Sequences were aligned with mafft in a server due to extensive use of memory (spent ca. 3-4min running):

```
#!/bin/bash
# send email indicating that everything started running
Rscript beginExec.R
# iterate over the files in order to run mafft on them
# and name the output files accordingly
```

```

for i in `ls *.fasta`; do
    mafft --maxiterate 1000 --thread 24 --localpair $i > $i.aligned.fasta
done
# send email indicating that everything finished
Rscript finishExec.R

```

Manual trimming of alignments was carried out in order to preserve as most as possible from the alignment found with mafft. The steps below were carried out in order to reach the per-marker alingment that will be subject to concatenation downward the analysis.

#### S2.4.1 *COI*

- Positions 16548–6124 removed as flanking region
- Positions 5618–1 removed as flanking region

#### S2.4.2 *APTase6-8*

- Positions 16548–8753 removed as flanking region
- Positions 7969–1 removed as flanking region

#### S2.4.3 *16S*

- Positions 608–600 removed as flanking region.
- Positions 13–1 removed as flanking region.

#### S2.4.4 *CytB*

- Positions 16548–15428 removed as flanking region.
- Positions 14590–1 removed as flanking region.

#### S2.4.5 *RAG2*

- Positions 1224–1110 removed as flanking region.
- Positions 96–1 removed as flanking region.

#### S2.4.6 *RAG1*

- Positions 1500–1359 removed as flanking region.
- Positions 93–1 removed as flanking region.

#### S2.4.7 *Myh6*

- Positions 752–712 removed as flanking region.
- Positions 9–1 removed as flanking region.

#### S2.4.8 *rps7*

- This marker disintegrated during alignment, removing altogether

#### S2.4.9 *18S*

- Cynodon gibbus AY523598.1 and Hydrolycus armatus AY523597.1 removed because disintegrated during alignment.
- Positions 16560–2601 removed as flanking regions.
- Positions 2012–1 removed as flanking regions.

## S2.5 Accession numbers

Finally, the following accession numbers were used for ensambling each taxon:

```
cd ../phylo/bayesian/sequences/alignedSeqs/
grep ">" *.trimmed.fasta | sed 's/.fa.renamed.fasta.aligned.fasta.trimmed.fasta:>/ /g'
cd ../../../../../supplementary
```

```
16S Acestrorhynchus_falcirostris_16S_FJ362540.1
16S Cynodon_gibbus_16S_HQ171241.1
16S Hydrolycus_scomberoides_16S_AP011989.1
16S Acestrorhynchus_microlepis_16S_FJ362546.1
16S Rhaphiodon_vulpinus_16S_HQ171303.1
18S Acestrorhynchus_microlepis_18S_FJ944765.1
18S Cynodon_gibbus_18S_AY523598.1
18S Hydrolycus_armatus_18S_AY523597.1
18S Hydrolycus_scomberoides_18S_AP011989.1
18S Rhaphiodon_vulpinus_18S_FJ944807.1
18Snew Acestrorhynchus_microlepis_18Snew_FJ944765.1
18Snew Hydrolycus_scomberoides_18Snew_FJ944806.1
18Snew Rhaphiodon_vulpinus_18Snew_FJ944807.1
18Snew Serrasalmus_rhombeus_18Snew_FJ944818.1
ATPase6-8 Acestrorhynchus_falcirostris_ATPase6-8_FJ468304.1
ATPase6-8 Hydrolycus_scomberoides_ATPase6-8_AP011989.1
ATPase6-8 Acestrorhynchus_microlepis_ATPase6-8_FJ468311.1
ATPase6-8 Rhaphiodon_vulpinus_ATPase6-8_FJ468317.1
COI Acestrorhynchus_falcirostris_COI_MG953597.1
COI Hydrolycus_scomberoides_COI_AP011989.1
COI Rhaphiodon_vulpinus_COI_GU701527.1
CytB Cynodon_gibbus_CytB_HQ289532.1
CytB Hydrolycus_scomberoides_CytB_AP011989.1
CytB Rhaphiodon_vulpinus_CytB_HQ289592.1
Myh6 Cynodon_gibbus_Myh6_HQ288951.1
Myh6 Hydrolycus_scomberoides_Myh6_HQ288979.1
Myh6 Rhaphiodon_vulpinus_Myh6_HQ289013.1
RAG1 Cynodon_gibbus_RAG1_HQ289148.1
RAG1 Hydrolycus_armatus_RAG1_JX470045.1
RAG1 Hydrolycus_scomberoides_RAG1_HQ289176.1
RAG1 Rhaphiodon_vulpinus_RAG1_HQ289205.1
RAG2 Cynodon_gibbus_RAG2_HQ289339.1
RAG2 Hydrolycus_scomberoides_RAG2_AY804088.1
RAG2 Rhaphiodon_vulpinus_RAG2_HQ289399.1
```

Table S1: Accession numbers for the eight molecular markers used in the bayesian phylogenetic analyses.

Genus	Species	Marker	Accession number
<i>Acestrorhynchus</i>	<i>falcirostris</i>	16S	FJ362540.1
<i>Acestrorhynchus</i>	<i>falcirostris</i>	ATPase6-8	FJ468304.1
<i>Acestrorhynchus</i>	<i>falcirostris</i>	COI	MG953597.1
<i>Acestrorhynchus</i>	<i>microlepis</i>	16S	FJ362546.1
<i>Acestrorhynchus</i>	<i>microlepis</i>	18S	FJ944765.1
<i>Acestrorhynchus</i>	<i>microlepis</i>	ATPase6-8	FJ468311.1
<i>Cynodon</i>	<i>gibbus</i>	16S	HQ171241.1
<i>Cynodon</i>	<i>gibbus</i>	CytB	HQ289532.1

<i>Cynodon</i>	<i>gibbus</i>	<i>Myh6</i>	HQ288951.1
<i>Cynodon</i>	<i>gibbus</i>	<i>RAG1</i>	HQ289148.1
<i>Cynodon</i>	<i>gibbus</i>	<i>RAG2</i>	HQ289339.1
<i>Hydrolycus</i>	<i>armatus</i>	<i>RAG1</i>	JX470045.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>16S</i>	HQ171269.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>18S</i>	FJ944806.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>ATPase6-8</i>	AP011989.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>COI</i>	AP011989.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>CytB</i>	AP011989.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>Myh6</i>	HQ288979.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>RAG1</i>	HQ289176.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>RAG2</i>	AY804088.1
<i>Hydrolycus</i>	<i>scomberoides</i>	<i>RAG2</i>	HQ289366.1
<i>Myloplus</i>	<i>schomburgkii</i>	<i>16S</i>	KX087046.1
<i>Myloplus</i>	<i>schomburgkii</i>	<i>COI</i>	MH411315.1
<i>Myloplus</i>	<i>schomburgkii</i>	<i>CytB</i>	KX086862.1
<i>Myloplus</i>	<i>schomburgkii</i>	<i>Myh6</i>	KX086912.1
<i>Myloplus</i>	<i>schomburgkii</i>	<i>RAG1</i>	KX086965.1
<i>Myloplus</i>	<i>schomburgkii</i>	<i>RAG2</i>	KX086979.1
<i>Rhaphiodon</i>	<i>vulpinus</i>	<i>16S</i>	HQ171303.1
<i>Rhaphiodon</i>	<i>vulpinus</i>	<i>18S</i>	FJ944807.1
<i>Rhaphiodon</i>	<i>vulpinus</i>	<i>ATPase6-8</i>	FJ468317.1
<i>Rhaphiodon</i>	<i>vulpinus</i>	<i>COI</i>	GU701527.1
<i>Rhaphiodon</i>	<i>vulpinus</i>	<i>CytB</i>	HQ289592.1
<i>Rhaphiodon</i>	<i>vulpinus</i>	<i>Myh6</i>	HQ289013.1
<i>Rhaphiodon</i>	<i>vulpinus</i>	<i>RAG1</i>	HQ289205.1
<i>Rhaphiodon</i>	<i>vulpinus</i>	<i>RAG2</i>	HQ289399.1
<i>Serrasalmus</i>	<i>rhombeus</i>	<i>16S</i>	AY788081.1
<i>Serrasalmus</i>	<i>rhombeus</i>	<i>18S</i>	FJ944818.1
<i>Serrasalmus</i>	<i>rhombeus</i>	<i>COI</i>	MZ051158.1
<i>Serrasalmus</i>	<i>rhombeus</i>	<i>CytB</i>	MT372736.1
<i>Serrasalmus</i>	<i>rhombeus</i>	<i>Myh6</i>	MT372746.1
<i>Serrasalmus</i>	<i>rhombeus</i>	<i>RAG1</i>	MT372758.1
<i>Serrasalmus</i>	<i>rhombeus</i>	<i>RAG2</i>	MT372769.1

## S2.6 Substitution models

Before finding substitution models, it was necessary to rename the sequences so that jModelTest won't complain by using the script `taxNamesFasta.R`

```
Rscript scripts/taxNamesFasta.R
```

The following script was used for finding the substitution models

```
#!/bin/bash

for i in $(ls *.trimmed.fasta)
do
    echo "Running jModelTest2 on $i\n"
    java -jar ~/programs/jmodeltest2/dist/jModelTest.jar -BIC -d $i
        -f -i -g 4
        -s 3 -t ML
        -tr 4 -o ${i}date +"%H.%M_%d_%Y".out'
done
```

The following models were chosen:

```
for i in `ls .../phylo/bayesian/sequences/alignedSeqs/substModels/* .out` ; do
  echo "$i" | cut -d '/' -f 4
  tail -n 1 $i | awk '{print $2}'
  printf "\n"
done

sequences
SYM+I

sequences
SYM+G

sequences
K80

sequences
K80

sequences
HKY+I

sequences
HKY+I

sequences
HKY+G

sequences
HKY+G

sequences
HKY+I

sequences
HKY+G

sequences
K80

sequences
K80+I

sequences
K80

sequences
K80+G

sequences
K80

sequences
K80+G
```

Now the matrices need to be converted to nexus in order to concatenate. In this step it is crucial to include a morphological fasta file with all missing data in order to guarantee that both molecular and morphological matrices can be concatenated:

```
>Hydrolycus_MUN_16211  
-----  
>Hydrolycus_armatus  
-----  
>Hydrolycus_tatauaia  
-----  
>Hydrolycus_wallacei  
-----  
>Hydrolycus_scomberoides  
-----  
>Rhaphiodon_vulpinus  
-----  
>Cynodon_gibbus  
-----  
>Cynodon_septenarius  
-----  
>Acestrorhynchus_falcirostris  
-----  
>Acestrorhynchus_microlepis  
-----  
>Myloplus_schomburkii  
-----  
>Serrasalmus_rhombeus
```

These matrices can then be renamed sharing taxon names across partitions with the script `createNexusFiles.R`:

```
Rscript scripts/createNexusFiles.R
```

Then we need to concatenate these nexus files for the only-molecular analysis and then add the morphological partition for the total evidence analysis. The script `concatenateMatricesMolonly.R` will call the `concatNexus.R` function and produce a molecular-only concatenated matrix along with endpoints for partitions.

```
Rscript scripts(concatenateMatricesMolonly.R
```

Checking matrix integrity for the Molecular-only matrix:

```
mb ..../phylo/bayesian/sequences/alignedSeqs/substModels/nexus/concatenatedMolonly.nexus > outmbMolonlyConcat  
cat outmbMolonlyConcat
```

```
MrBayes 3.2.7a x86_64
```

```
(Bayesian Analysis of Phylogeny)
```

```
Distributed under the GNU General Public License
```

```
Type "help" or "help <command>" for information  
on the commands that are available.
```

```
Type "about" for authorship and general  
information about the program.
```

```

Executing file ".../phylo/bayesian/sequences/alignedSeqs/substModels/nexus/concatenatedMolonly.nexus"
UNIX line termination
Longest line length = 1304
Parsing file
Expecting NEXUS formatted file
Reading data block
    Allocated taxon set
    Allocated matrix
    Defining new matrix with 10 taxa and 6061 characters
    Data is Dna
    Missing data coded as ?
    Gaps coded as -
    Data matrix is interleaved
    Taxon 1 -> Acestrorhynchus_falcirostris
    Taxon 2 -> Cynodon_gibbus
    Taxon 3 -> Hydrolycus_scomberoides
    Taxon 4 -> Acestrorhynchus_microlepis
    Taxon 5 -> Rhaphiodon_vulpinus
    Taxon 6 -> Hydrolycus_armatus
    Taxon 7 -> Hydrolycus_MUN_16211
    Taxon 8 -> Hydrolycus_tatauaia
    Taxon 9 -> Hydrolycus_wallacei
    Taxon 10 -> Cynodon_septenarius
Successfully read matrix
Setting default partition (does not divide up characters)
Setting model defaults
Seed (for generating default start values) = 1652999209
Setting output file names to ".../phylo/bayesian/sequences/alignedSeqs/substModels/nexus/concatena
Exiting data block
Reached end of file

Tasks completed, exiting program because mode is noninteractive
To return control to the command line after completion of file processing,
set mode to interactive with 'mb -i <filename>' (i is for interactive)
or use 'set mode=interactive'

```

There is another script named `concatenateMatricesMolmorph.R` for the total evidence analysis.

```
Rscript scripts/concatenateMatricesMolmorph.R
```

Checking matrix integrity for the total-evidence matrix:

```
mb .../phylo/bayesian/sequences/alignedSeqs/substModels/nexus/concatenatedMolmorph.nexus > outmbMolmorph
cat outmbMolmorphConcat
```

MrBayes 3.2.7a x86\_64

(Bayesian Analysis of Phylogeny)

Distributed under the GNU General Public License

Type "help" or "help <command>" for information  
on the commands that are available.

```
Type "about" for authorship and general
information about the program.
```

```
Executing file "../phylo/bayesian/sequences/alignedSeqs/substModels/nexus/concatenatedMolmorph.nexus"
UNIX line termination
Longest line length = 1304
Parsing file
Expecting NEXUS formatted file
Reading data block
    Allocated taxon set
    Allocated matrix
    Defining new matrix with 10 taxa and 6141 characters
    Data is Mixed
        Data for partition 1 is Dna
        Data for partition 2 is Standard
    There are a total of 2 default data divisions
    Missing data coded as ?
    Gaps coded as -
    Data matrix is interleaved
    Taxon 1 -> Acestrorhynchus_falcirostris
    Taxon 2 -> Cynodon_gibbus
    Taxon 3 -> Hydrolycus_scomberoides
    Taxon 4 -> Acestrorhynchus_microlepis
    Taxon 5 -> Rhaphiodon_vulpinus
    Taxon 6 -> Hydrolycus_armatus
    Taxon 7 -> Hydrolycus_MUN_16211
    Taxon 8 -> Hydrolycus_tatauaia
    Taxon 9 -> Hydrolycus_wallacei
    Taxon 10 -> Cynodon_septenarius
    Successfully read matrix
    Setting default partition, dividing characters into 2 parts
    Setting model defaults
    Seed (for generating default start values) = 1652999209
    Setting output file names to "../phylo/bayesian/sequences/alignedSeqs/substModels/nexus/concatena
Exiting data block
Reached end of file

Tasks completed, exiting program because mode is noninteractive
To return control to the command line after completion of file processing,
set mode to interactive with 'mb -i <filename>' (i is for interactive)
or use 'set mode=interactive'
```

## S2.7 MrBayes script

For the molecular-only analysis (Figure S3) we have generated the following script:

```
begin mrbayes;
[Script documentation carried out using comments]

[log the analysis]
log start filename=cynoMolonly.log;
[read the matrix concatenatedMolonly.nexus]
execute concatenatedMolonly.nexus;
```

```

[close analysis at end]
set autoclose=yes;
[set Acestrorhynchus_microlepis as outgroup]
outgroup Acestrorhynchus_microlepis;
[This command shows the status of all the taxa, according to the documentation]
taxastat;

[definition of individual partitions per marker]
charset 16S=1-586;
charset 18S=587-1127;
charset ATPase6_8=1128-1910;
charset COI=1911-2415;
charset CytB=2416-3252;
charset Myh6=3253-3954;
charset RAG1=3955-5219;
charset RAG2=5220-6232;

[definition of combined dataset]
partition combined=8: 16S, 18S, ATPase6_8, COI, CytB, Myh6, RAG1, RAG2;

[specification of substitution models]
set partition=combined;
lset applyto=(1) nst=6 rates=gamma; [SYM+G]
lset applyto=(2) nst=2; [K80]
lset applyto=(3) nst=2 rates=propinv; [HKY+I]
lset applyto=(4) nst=2 rates=gamma; [HKY+G]
lset applyto=(5) nst=2 rates=gamma; [HKY+G]
lset applyto=(6) nst=2 rates=propinv; [K80+I]
lset applyto=(7) nst=2 rates=gamma; [K80+G]
lset applyto=(8) nst=2 rates=gamma; [K80+G]

[unlink parameters across partitions]
unlink shape=(all) pinvar=(all) statefreq=(all) revmat=(all) tratio=(all);

[allow separate gamma parameters for each partition and set the stationary freqs in SYM and K80]
prset applyto=(1) statefreqpr=fixed(equal); [SYM+G in 16S]
prset applyto=(2) statefreqpr=fixed(equal); [K80 in 18S]
prset applyto=(6) statefreqpr=fixed(equal); [K80 in Myh6]
prset applyto=(7) statefreqpr=fixed(equal); [K80 in RAG1]
prset applyto=(8) statefreqpr=fixed(equal); [K80 in RAG2]
prset applyto=(all) ratepr=variable;

[turn off taxa for which there are no molecular data]
delete Hydrolycus_MUN_16211 Hydrolycus_wallacei Cynodon_septenarius Hydrolycus_tatauaia;

[show taxa]
taxastat;

[show the model just specified for each partition]
showmodel;

[set up the MCMC, with this setting the analysis will need not less than 16 threads]
mcmcpr nruns=2 ngen=4000000 nchains=8 samplefreq=4000 printfreq=100;

```

```

[run the MCMC]
mcmc;

[summarize the posterior trees]
sumt nruns=2 relburnin=yes burninfrac=0.25;
plot;

[summarize parameter posteriors]
sump;

log stop;
end;

```

Now, for the total evidence analysis (Figure S4), the following script was used:

```

begin mrbayes;
[Script documentation carried out using comments]

[log the analysis]
log start filename=cynoMolmorph.log;
[read the matrix concatenatedMolmorph.nexus]
execute concatenatedMolmorph.nexus;

[close analysis at end]
set autoclose=yes;
[set Acestrorhynchus_microlepis as outgroup]
outgroup Acestrorhynchus_microlepis;
[This command shows the status of all the taxa, according to the documentation]
taxastat;

[definition of individual partitions per marker]
charset 16S=1-586;
charset 18S=587-1127;
charset ATPase6_8=1128-1910;
charset COI=1911-2415;
charset CytB=2416-3252;
charset Myh6=3253-3954;
charset RAG1=3955-5219;
charset RAG2=5220-6232;
charset morph=6233-6312;

[definition of combined dataset]
partition combined=9: 16S, 18S, ATPase6_8, COI, CytB, Myh6, RAG1, RAG2, morph;

[specification of substitution models]
set partition=combined;
lset applyto=(1) nst=6 rates=gamma; [SYM+G]
lset applyto=(2) nst=2; [K80]
lset applyto=(3) nst=2 rates=propinv; [HKY+I]
lset applyto=(4) nst=2 rates=gamma; [HKY+G]
lset applyto=(5) nst=2 rates=gamma; [HKY+G]
lset applyto=(6) nst=2 rates=propinv; [K80+I]
lset applyto=(7) nst=2 rates=gamma; [K80+G]
lset applyto=(8) nst=2 rates=gamma; [K80+G]

```

```

[unlink parameters across partitions]
unlink shape=(all) pinvar=(all) statefreq=(all) revmat=(all) tratio=(all);

[allow separate gamma parameters for each partition and set the stationary freqs in SYM and K80]
prset applyto=(1) statefreqpr=fixed(equal); [SYM+G in 16S]
prset applyto=(2) statefreqpr=fixed(equal); [K80 in 18S]
prset applyto=(6) statefreqpr=fixed(equal); [K80 in Myh6]
prset applyto=(7) statefreqpr=fixed(equal); [K80 in RAG1]
prset applyto=(8) statefreqpr=fixed(equal); [K80 in RAG2]
prset applyto=(all) ratepr=variable;

[show the model just specified for each partition]
showmodel;

[set up the MCMC, with this setting the analysis will need not less than 16 threads]
mcmcp nruns=2 ngen=4000000 nchains=8 samplefreq=4000 printfreq=100;
[run the MCMC]
mcmc;

[summarize the posterior trees]
sumt nruns=2 relburnin=yes burninfrac=0.25;
plot;

[summarize parameter posteriors]
sump;

log stop;
end;

```

### S2.7.1 Analysis output

```

Analysis completed in 42 mins 38 seconds
Analysis used 2485.60 seconds of CPU time on processor 0
Likelihood of best state for "cold" chain of run 1 was -16693.94
Likelihood of best state for "cold" chain of run 2 was -16693.94

```

Chain swap information for run 1:

	1	2	3	4	5	6	7	8
1		0.68	0.42	0.25	0.13	0.07	0.03	0.01
2	142186		0.70	0.46	0.28	0.15	0.08	0.04
3	142972	142680		0.72	0.48	0.30	0.17	0.09
4	142637	142587	142884		0.73	0.49	0.31	0.18
5	142683	143807	142387	143018		0.74	0.50	0.31
6	142833	142619	142866	142306	142724		0.73	0.50
7	142831	142957	142709	143133	143195	143199		0.73
8	142398	143079	143396	143044	142770	143295	142805	

Chain swap information for run 2:

1	2	3	4	5	6	7	8

1		0.68	0.43	0.25	0.13	0.06	0.03	0.01
2	143065		0.70	0.45	0.27	0.15	0.07	0.03
3	142334	143275		0.72	0.47	0.29	0.15	0.07
4	142416	142367	142417		0.73	0.48	0.28	0.14
5	142321	143286	143162	142729		0.72	0.47	0.27
6	143234	143414	143465	142956	142904		0.71	0.45
7	143287	142386	142730	143732	142829	142425		0.70
8	142671	143345	143120	142666	142757	142676	142031	

Upper diagonal: Proportion of successful state exchanges between chains  
Lower diagonal: Number of attempted state exchanges between chains

Chain information:

```
ID -- Heat
-----
1 -- 1.00 (cold chain)
2 -- 0.91
3 -- 0.83
4 -- 0.77
5 -- 0.71
6 -- 0.67
7 -- 0.62
8 -- 0.59
```

```
Heat = 1 / (1 + T * (ID - 1))
(where T = 0.10 is the temperature and ID is the chain number)
```

```
Setting sumt nruns to 2
Using relative burnin (a fraction of samples discarded).
Setting burnin fraction to 0.25
Summarizing trees in files "concatenatedMolmorph.nexus.run1.t" and "concatenatedMolmorph.nexus.run1.tstat"
Using relative burnin ('relburnin=yes'), discarding the first 25 % of sampled trees
Writing statistics to files concatenatedMolmorph.nexus.<parts|tstat|vstat|trprobs|con>
Examining first file ...
Found one tree block in file "concatenatedMolmorph.nexus.run1.t" with 1001 trees in last block
Expecting the same number of trees in the last tree block of all files
```

Tree reading status:

```
0      10     20     30     40     50     60     70     80     90     100
v-----v-----v-----v-----v-----v-----v-----v-----v-----v-----v
*****
```

```
Read a total of 2002 trees in 2 files (sampling 1502 of them)
(Each file contained 1001 trees of which 751 were sampled)
Overwriting file "concatenatedMolmorph.nexus.parts"
Overwriting file "concatenatedMolmorph.nexus.tstat"
Overwriting file "concatenatedMolmorph.nexus.vstat"
Overwriting file "concatenatedMolmorph.nexus.con.tre"
Overwriting file "concatenatedMolmorph.nexus.trprobs"
```

Summary statistics for informative taxon bipartitions

(saved to file "concatenatedMolmolph.nexus.tstat"):

ID	#obs	Probab.	Sd(s)+	Min(s)	Max(s)	Nruns
13	1502	1.000000	0.000000	1.000000	1.000000	2
14	1502	1.000000	0.000000	1.000000	1.000000	2
15	1502	1.000000	0.000000	1.000000	1.000000	2
16	1502	1.000000	0.000000	1.000000	1.000000	2
17	1502	1.000000	0.000000	1.000000	1.000000	2
18	1483	0.987350	0.004708	0.984021	0.990679	2
19	1392	0.926764	0.001883	0.925433	0.928096	2
20	1359	0.904794	0.014123	0.894807	0.914780	2
21	1010	0.672437	0.005649	0.668442	0.676431	2
22	387	0.257656	0.017890	0.245007	0.270306	2

+ Convergence diagnostic (standard deviation of split frequencies) should approach 0.0 as runs converge.

Summary statistics for branch and node parameters

(saved to file "concatenatedMolmolph.nexus.vstat"):

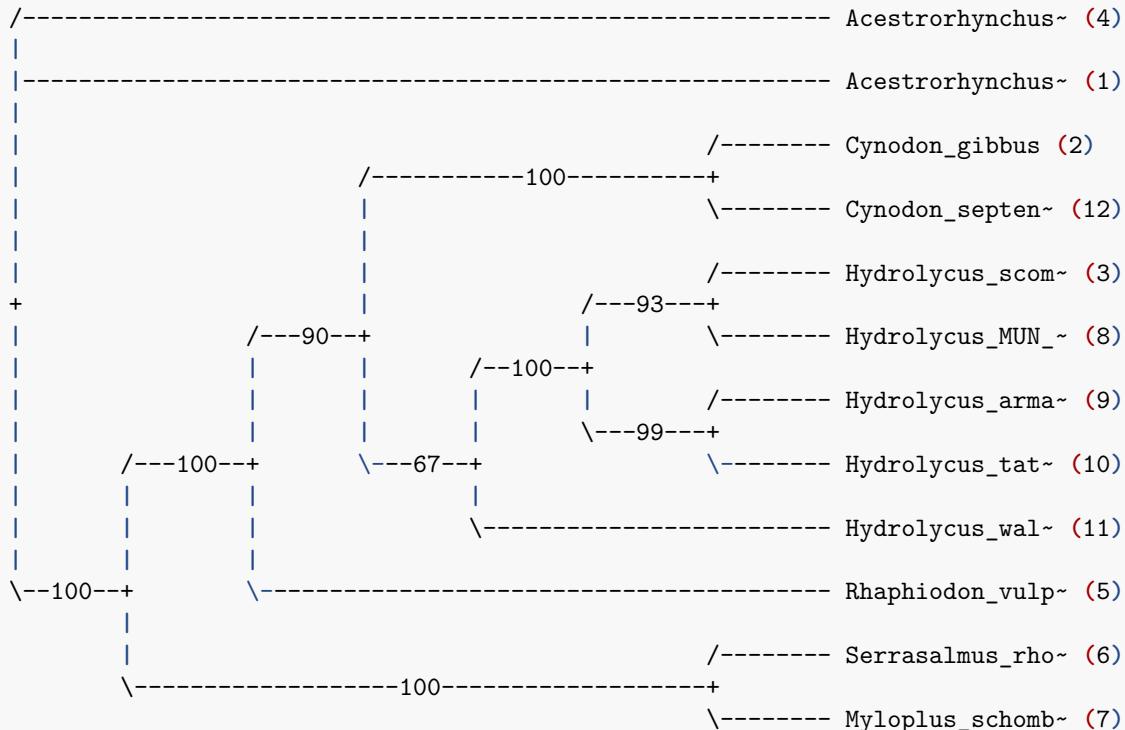
Parameter	95% HPD Interval						PSRF+	Nruns
	Mean	Variance	Lower	Upper	Median			
length{all}[1]	0.121283	0.000647	0.076079	0.168235	0.119123	1.000	2	
length{all}[2]	0.003304	0.000011	0.000005	0.009570	0.002368	1.000	2	
length{all}[3]	0.006582	0.000034	0.000007	0.018537	0.005021	1.000	2	
length{all}[4]	0.106678	0.000490	0.064165	0.145773	0.104678	0.999	2	
length{all}[5]	0.123063	0.000415	0.089677	0.158913	0.120216	0.999	2	
length{all}[6]	0.053902	0.000123	0.034775	0.074554	0.052914	0.999	2	
length{all}[7]	0.066541	0.000145	0.046582	0.091292	0.065228	0.999	2	
length{all}[8]	0.017230	0.000328	0.000011	0.051679	0.011725	0.999	2	
length{all}[9]	0.002073	0.000004	0.000002	0.005984	0.001411	1.000	2	
length{all}[10]	0.005258	0.000033	0.000000	0.015477	0.003577	1.000	2	
length{all}[11]	0.031454	0.000449	0.000003	0.073582	0.027998	1.000	2	
length{all}[12]	0.005225	0.000028	0.000000	0.015899	0.003522	1.001	2	
length{all}[13]	0.093332	0.000513	0.051032	0.138028	0.090221	1.000	2	
length{all}[14]	0.107790	0.000890	0.056422	0.165460	0.104002	0.999	2	
length{all}[15]	0.075369	0.000215	0.050442	0.103065	0.074390	1.000	2	
length{all}[16]	0.061229	0.000323	0.028559	0.092868	0.060664	0.999	2	
length{all}[17]	0.129496	0.000689	0.087184	0.181383	0.126789	0.999	2	
length{all}[18]	0.015813	0.000072	0.001222	0.031710	0.014454	1.000	2	
length{all}[19]	0.017380	0.000067	0.002449	0.033320	0.016695	1.000	2	
length{all}[20]	0.022256	0.000068	0.004804	0.036643	0.022130	1.002	2	
length{all}[21]	0.021295	0.000175	0.000274	0.046515	0.019714	0.999	2	
length{all}[22]	0.015604	0.000101	0.000071	0.033194	0.015121	0.999	2	

+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when deviation of parameter values within all runs is 0 or when a parameter value (a branch length, for instance) is not sampled in all runs.

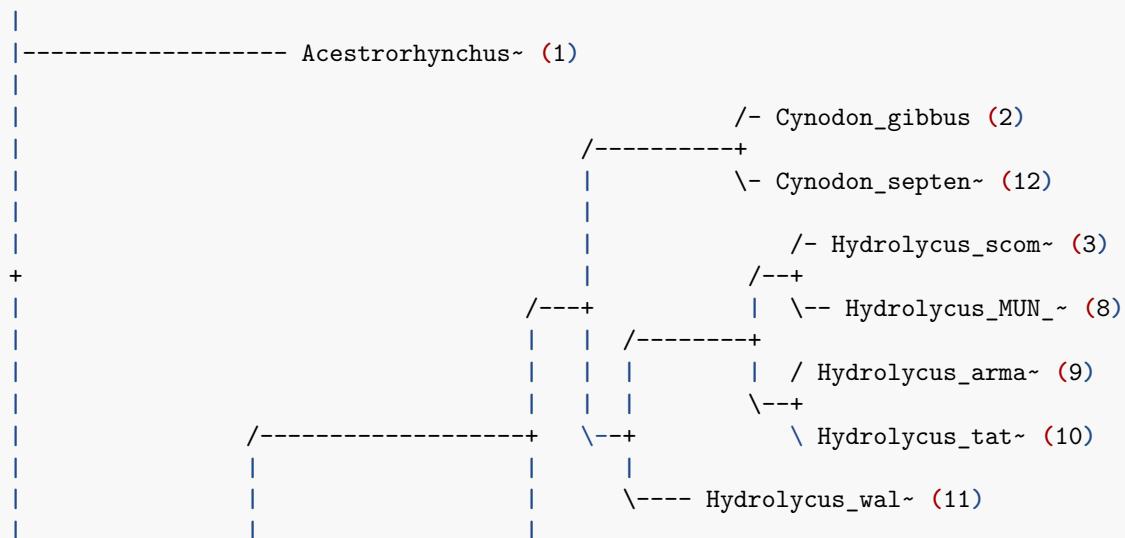
Summary statistics for partitions with frequency  $\geq 0.10$  in at least one run:

- Average standard deviation of split frequencies = 0.004425
- Maximum standard deviation of split frequencies = 0.017890
- Average PSRF for parameter values (excluding NA and  $>10.0$ ) = 1.000
- Maximum PSRF for parameter values = 1.002

### Clade credibility values:



Phylogram (based on average branch lengths):



```

\-----+                               \----- Rhaphiodon_vulp~ (5)
|           |
|           /----- Serrasalmus_rho~ (6)
\-----+
           \----- Myloplus_schomb~ (7)

```

|-----| 0.100 expected changes per site

Calculating tree probabilities...

Credible sets of trees (21 trees sampled):

90 % credible set contains 5 trees

95 % credible set contains 7 trees

99 % credible set contains 15 trees

Plotting parameters in file concatenatedMolmorph.nexus.run1.p ...

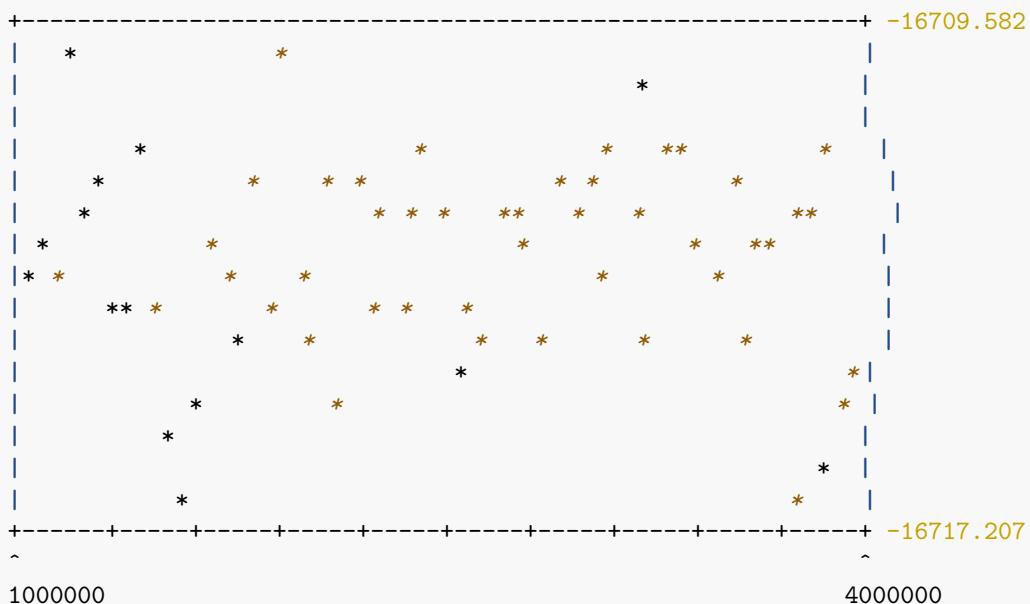
Found 1000 parameter lines in file "concatenatedMolmorph.nexus.run1.p"

Of the 1000 lines, 751 of them will be summarized (starting at line 252)

(Only the last set of lines will be read, in case multiple

parameter blocks are present in the same file.)

Rough trace plot of parameter LnL:

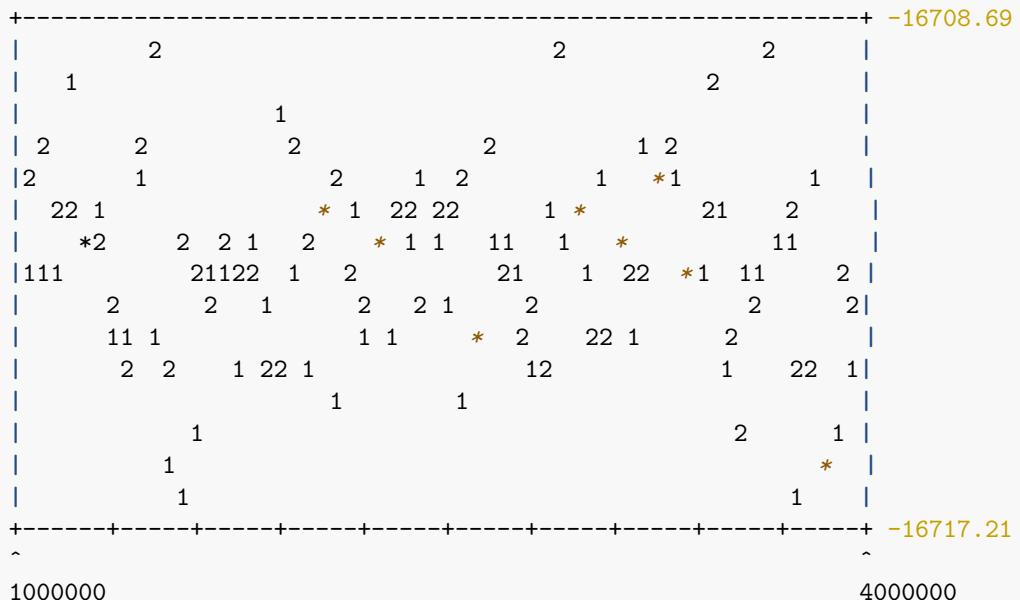


Summarizing parameters in files concatenatedMolmorph.nexus.run1.p and concatenatedMolmorph.nexus.pstat  
Writing summary statistics to file concatenatedMolmorph.nexus.pstat  
Using relative burnin ('relburnin=yes'), discarding the first 25 % of samples

Below are rough plots of the generation (x-axis) versus the log probability of observing the data (y-axis). You can use these graphs to determine what the burn in for your analysis should be. When the log probability starts to plateau you may be at stationarity. Sample trees and parameters after the log probability plateaus. Of course, this is not a guarantee that you are at sta-

tionarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt' commands for all the parameters in your model. Remember that the burn in is the number of samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC analysis.

Overlay plot for both runs:  
(1 = Run number 1; 2 = Run number 2; \* = Both runs)



Overwriting file "concatenatedMolmorph.nexus.lstat"

Estimated marginal likelihoods for runs sampled in files  
"concatenatedMolmorph.nexus.run1.p" and "concatenatedMolmorph.nexus.run2.p":  
(Use the harmonic mean for Bayes factor comparisons of models)

(Values are saved to the file concatenatedMolmorph.nexus.lstat)

Run	Arithmetic mean	Harmonic mean
1	-16704.33	-16727.97
2	-16704.44	-16725.91
TOTAL	-16704.38	-16727.40

Model parameter summaries over the runs sampled in files  
"concatenatedMolmorph.nexus.run1.p" and "concatenatedMolmorph.nexus.run2.p":  
Summaries are based on a total of 1502 samples from 2 runs.  
Each run produced 1001 samples of which 751 samples were included.  
Parameter summaries saved to file "concatenatedMolmorph.nexus.pstat".

Overwriting file "concatenatedMolmorph.nexus.pstat"

95% HPD Interval

Parameter	Mean	Variance	Lower	Upper	Median	min ESS*	avg ESS	PSR
TL{all}	1.081992	0.024825	0.828699	1.359915	1.060861	565.46	575.83	1.00
kappa{2}	2.421763	0.482698	1.326020	3.962307	2.321954	611.26	681.13	1.00
kappa{3}	9.909513	5.021388	6.299223	14.127390	9.601083	601.55	676.28	0.99
kappa{4}	18.269297	70.476883	7.888667	33.629870	16.540860	501.59	585.52	0.99
kappa{5}	11.771211	9.826783	7.091952	17.154490	11.268230	503.47	627.24	1.00
kappa{6}	9.058976	7.792466	4.386020	14.399150	8.611278	680.61	715.81	0.99
kappa{7}	5.451419	0.950469	3.668134	7.309621	5.382256	635.25	693.13	1.00
kappa{8}	4.949995	0.738370	3.444329	6.762524	4.878486	710.56	730.78	1.00
r(A<->C){1}	0.091547	0.000582	0.043856	0.137109	0.090916	566.91	614.45	1.00
r(A<->G){1}	0.201943	0.001146	0.135144	0.267439	0.201132	580.61	665.81	1.00
r(A<->T){1}	0.103613	0.000588	0.059235	0.153394	0.102838	676.17	713.59	0.99
r(C<->G){1}	0.010581	0.000070	0.000016	0.027046	0.008947	556.08	634.04	1.00
r(C<->T){1}	0.583848	0.002741	0.485676	0.682432	0.582281	575.87	618.30	0.99
r(G<->T){1}	0.008467	0.000052	0.000007	0.023403	0.006692	632.09	691.55	1.00
pi(A){3}	0.294309	0.000179	0.266511	0.318523	0.294860	751.00	751.00	1.00
pi(C){3}	0.322309	0.000171	0.296207	0.346717	0.322268	751.00	751.00	1.00
pi(G){3}	0.113658	0.000073	0.097144	0.129874	0.113682	692.97	721.98	1.00
pi(T){3}	0.269723	0.000139	0.246368	0.291832	0.269587	531.18	641.09	1.00
pi(A){4}	0.253823	0.000253	0.221791	0.284381	0.253483	664.93	707.97	1.00
pi(C){4}	0.320107	0.000271	0.287524	0.350807	0.319617	703.06	713.76	0.99
pi(G){4}	0.152532	0.000157	0.129425	0.178021	0.152447	447.88	527.18	0.99
pi(T){4}	0.273538	0.000234	0.245050	0.303697	0.273273	735.21	743.10	1.00
pi(A){5}	0.283648	0.000190	0.257013	0.311128	0.283483	751.00	751.00	0.99
pi(C){5}	0.377336	0.000207	0.349711	0.404988	0.376773	587.91	669.46	1.00
pi(G){5}	0.118121	0.000076	0.102446	0.135929	0.117900	670.19	710.59	1.00
pi(T){5}	0.220894	0.000126	0.199958	0.242083	0.220631	625.74	688.37	1.00
alpha{1}	0.195004	0.001539	0.129338	0.275439	0.190472	751.00	751.00	0.99
alpha{4}	0.137169	0.000516	0.096190	0.184523	0.135719	544.95	606.13	1.00
alpha{5}	0.109087	0.000387	0.072005	0.143524	0.109857	653.91	702.46	1.00
alpha{7}	0.280499	0.029045	0.001010	0.595317	0.251893	751.00	751.00	1.00
alpha{8}	1.161758	0.537155	0.260355	2.624213	0.959879	467.26	609.13	1.00
pinvar{3}	0.496351	0.000511	0.450581	0.539967	0.496751	522.00	561.78	1.00
pinvar{6}	0.729193	0.007206	0.567212	0.866081	0.744308	718.19	734.60	0.99
m{1}	0.656955	0.012828	0.441965	0.867755	0.647204	571.08	598.84	1.00
m{2}	0.142192	0.000736	0.089194	0.192879	0.141289	445.19	554.48	0.99
m{3}	2.281850	0.148166	1.627690	3.080942	2.267581	496.76	563.39	1.00
m{4}	3.177575	0.710836	1.724133	4.791906	3.041214	474.90	481.32	1.00
m{5}	1.919017	0.111336	1.344906	2.624610	1.895170	482.75	525.80	1.00
m{6}	0.188779	0.001290	0.122588	0.258991	0.186690	601.76	627.43	1.00
m{7}	0.230903	0.001119	0.167111	0.297218	0.230352	672.80	674.36	0.99
m{8}	0.277220	0.001705	0.197420	0.361959	0.276121	595.39	599.49	0.99
m{9}	1.838470	0.135857	1.167301	2.584990	1.812737	563.99	573.85	0.99

\* Convergence diagnostic (ESS = Estimated Sample Size); min and avg values correspond to minimal and average ESS among runs.

ESS value below 100 may indicate that the parameter is undersampled.

+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman and Rubin, 1992) should approach 1.0 as runs converge.

## References

- Pretti, V. Q., Calcagnotto, D., Toledo-Piza, M., and de Almeida-Toledo, L. F. (2009). Phylogeny of the Neotropical genus *Acestrorhynchus* (Ostariophysi: Characiformes) based on nuclear and mitochondrial gene sequences and morphology: A total evidence approach. *Molecular Phylogenetics and Evolution*, 52(2):312–320.
- Toledo-Piza, M. (2000). The Neotropical fish subfamily Cynodontinae (Teleostei: Ostariophysi: Characiformes): A phylogenetic study and a revision of *Cynodon* and *Rhaphiodon*. *American Museum Novitates*, 3286(3286):1–88.

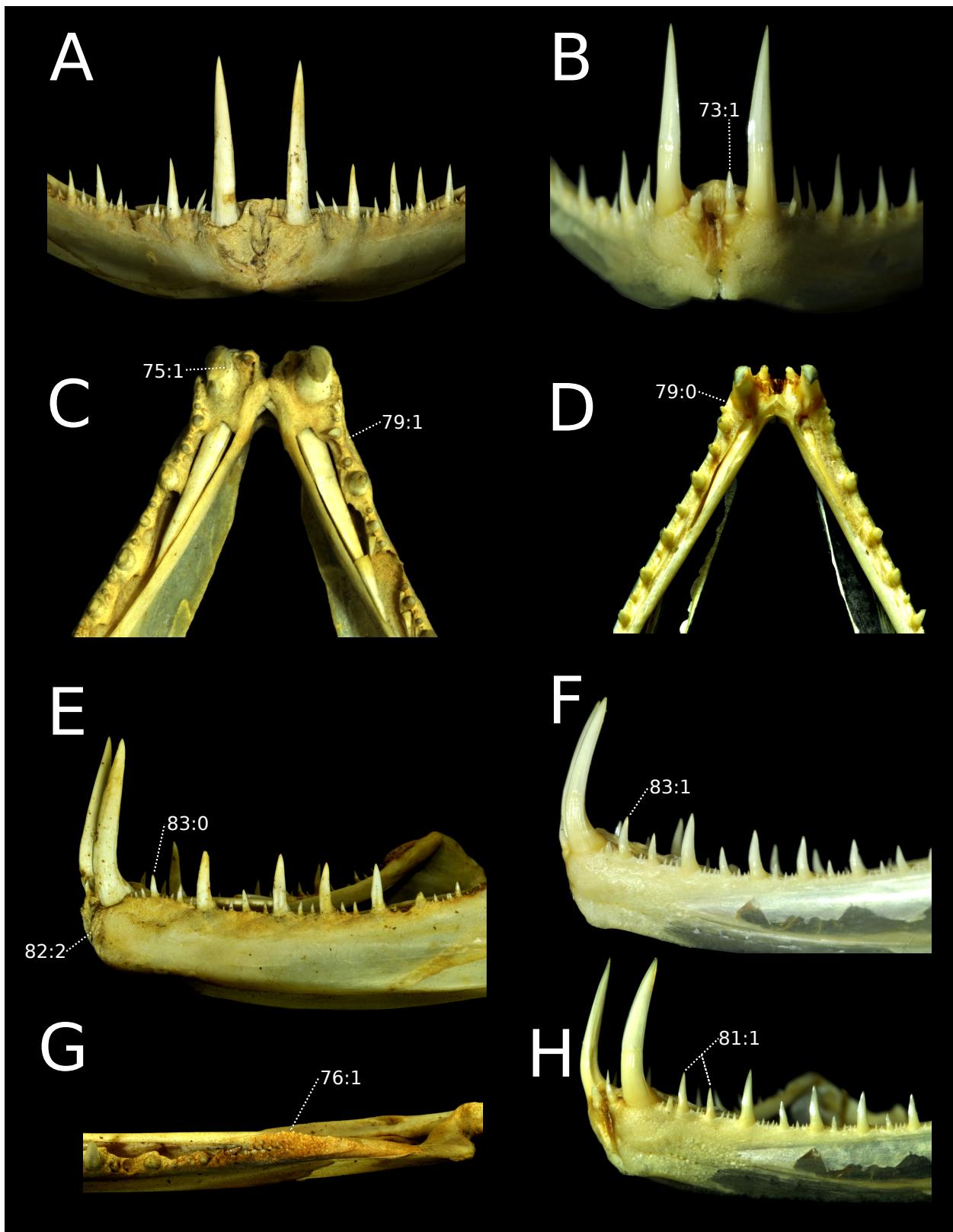


Figure S1: Dentary bones of all species of *Hydrolycus* and *Rhaphiodon*. *Hydrolycus armatus* in A. anterior, C. occlusal, and E. lateral views. G. detail of the posterior portion of the dentary showing the patch of small teeth. *Hydrolycus scomberoides* in B. anterior, D. occlusal, and F. lateral views. H. showing a slightly oblique view where the single symphyseal tooth is more evident. Character states are shown as character:character-state in all cases

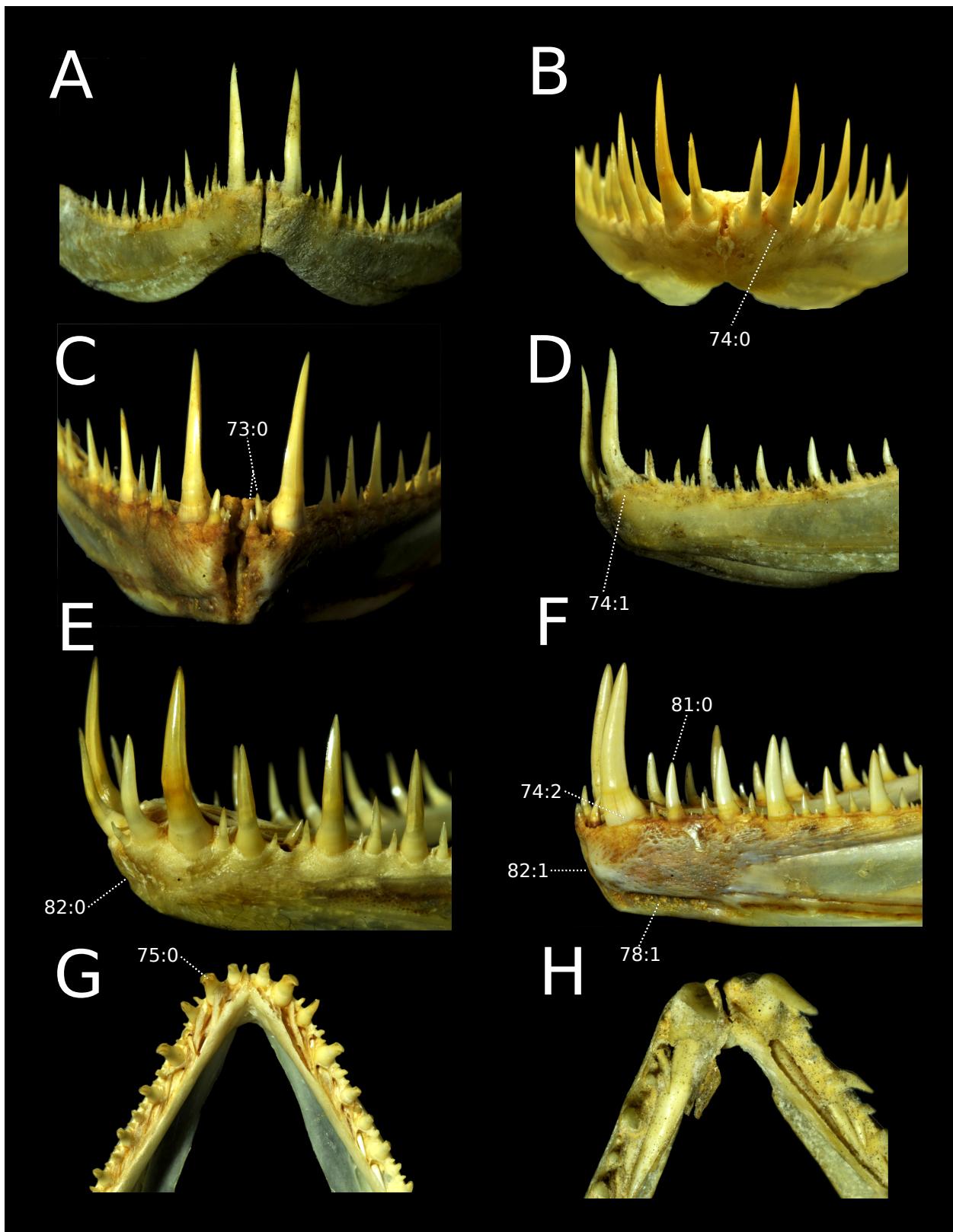


Figure S2: Dentary bones of all species of *Hydrolycus* and *Rhaphiodon*. *Hydrolycus tatauaia* in A. anterior, D. lateral, and H. occlusal views. *Hydrolycus wallacei* in B. anterior, E. lateral, and G. occlusal views. *Rhaphiodon vulpinus* in C. anterior, and F. lateral views. Character states are shown as character:character-state in all instances

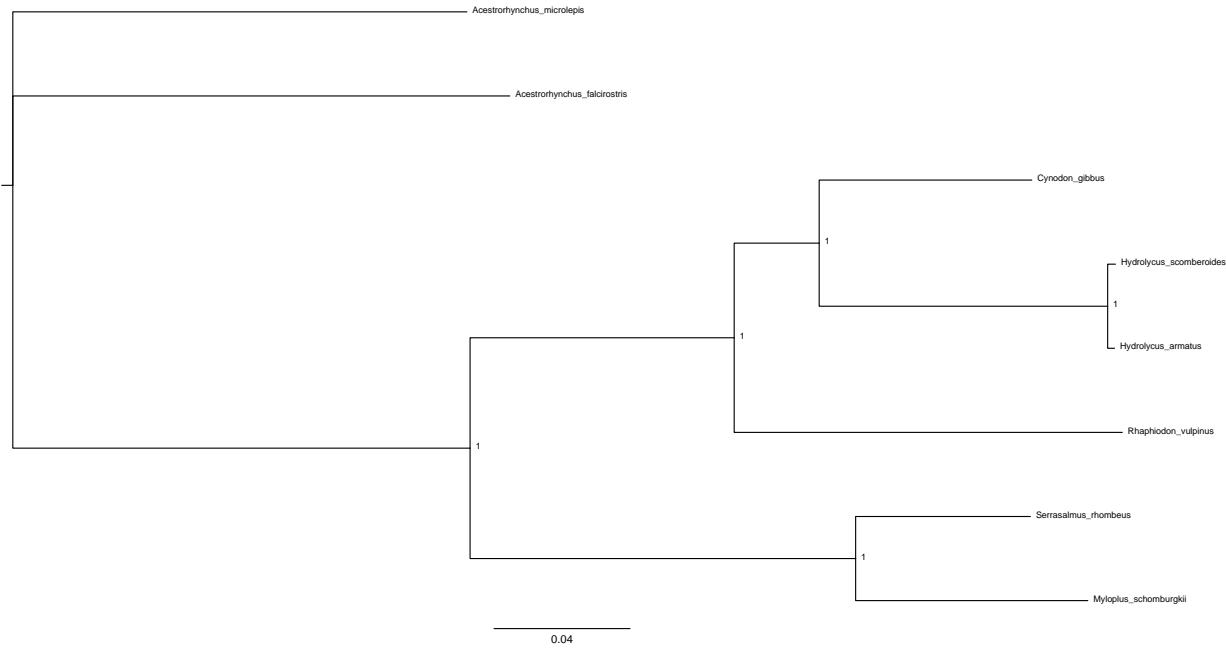


Figure S3: Molecular-only bayesian inference tree. Nodal values are posterior probabilities. Scale is branch length.

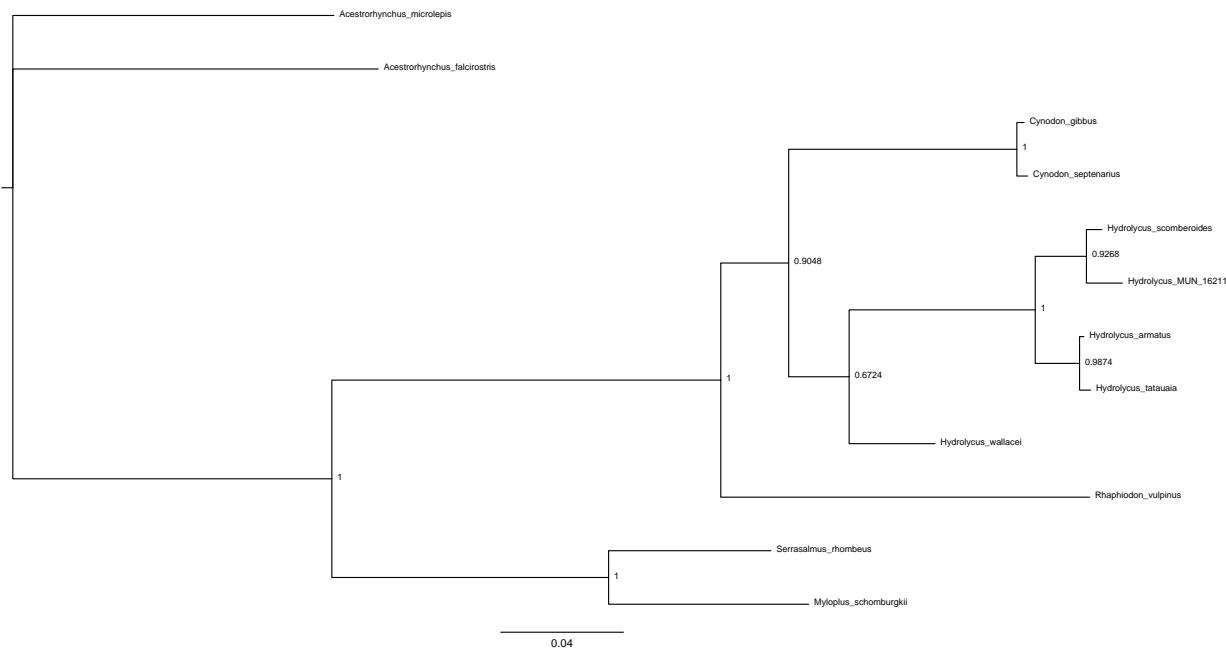


Figure S4: Combined analysis using both the morphological and molecular datasets bayesian inference. Nodal values are posterior probabilities. Scale is branch length.