



Phylogenetic Placements of HTS Data

Micah Dunthorn, Lucas Czech & Ľubomír Rajter

Today's aim

Today's aim

theory and background of phylogenetic
placements

Today's aim

theory and background of phylogenetic
placements

how to run EPA-ng and make a heattree



...assumed that you know phylogenetics

Why?

Why?

massive metabarcoding datasets
- both short- & long-reads

Why?

massive metabarcoding datasets
- both short- & long-reads

difficult to visualise so much data

Why?

massive metabarcoding datasets
- both short- & long-reads

difficult to visualise so much data

cannot do normal phylogenetic inferences

Is metabarcoding data
phylogenetically meaningful?

Is metabarcoding data
phylogenetically meaningful?

short-read data...yeah, maybe

Molecular Biology and Evolution 2014; 31: 993–1009

Placing Environmental Next-Generation Sequencing Amplicons from Microbial Eukaryotes into a Phylogenetic Context

Micah Dunthorn,^{*,1} Johannes Otto,¹ Simon A. Berger,² Alexandros Stamatakis,^{2,3} Frédéric Mahé,¹ Sarah Romac,^{4,5} Colomban de Vargas,^{4,5} Stéphane Audic,^{4,5} BioMarKs Consortium,[†] Alexandra Stock,¹ Frank Kauff,⁶ and Thorsten Stoeck¹

Is metabarcoding data
phylogenetically meaningful?

short-read data....yeah, maybe

long-read data better

2020; 20: 429–443

DOI: 10.1111/1755-0998.13117





Check for
updates

RESOURCE ARTICLE

MOLECULAR ECOLOGY
RESOURCES

WILEY

Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity

Mahwash Jamy¹  | Rachel Foster² | Pierre Barbera³ | Lucas Czech³ |
Alexey Kozlov³ | Alexandros Stamatakis^{3,4} | Gary Bending⁵ | Sally Hilton⁵ |
David Bass^{2,6}  | Fabien Burki¹

What can it do?

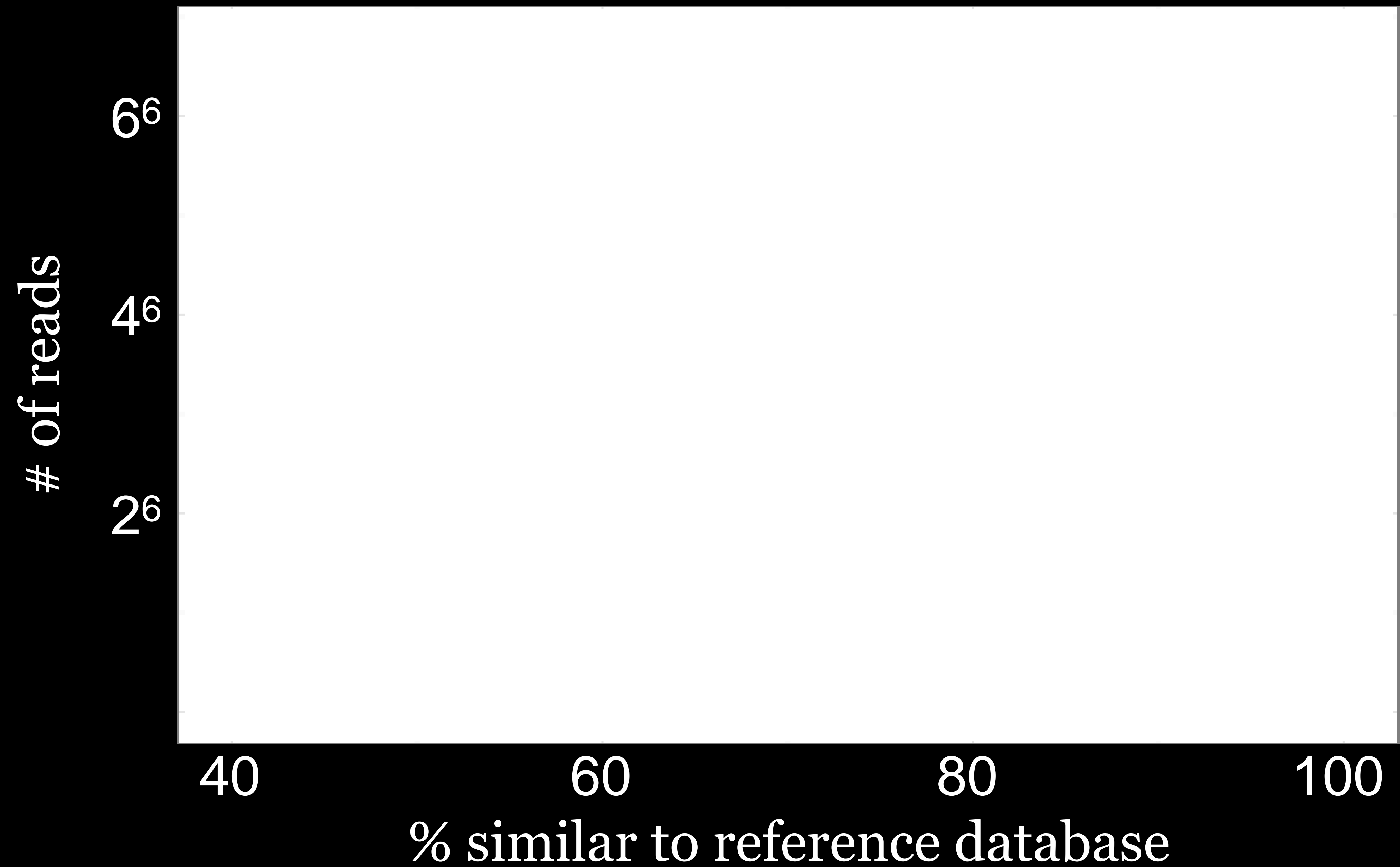
What can it do?

clean your data

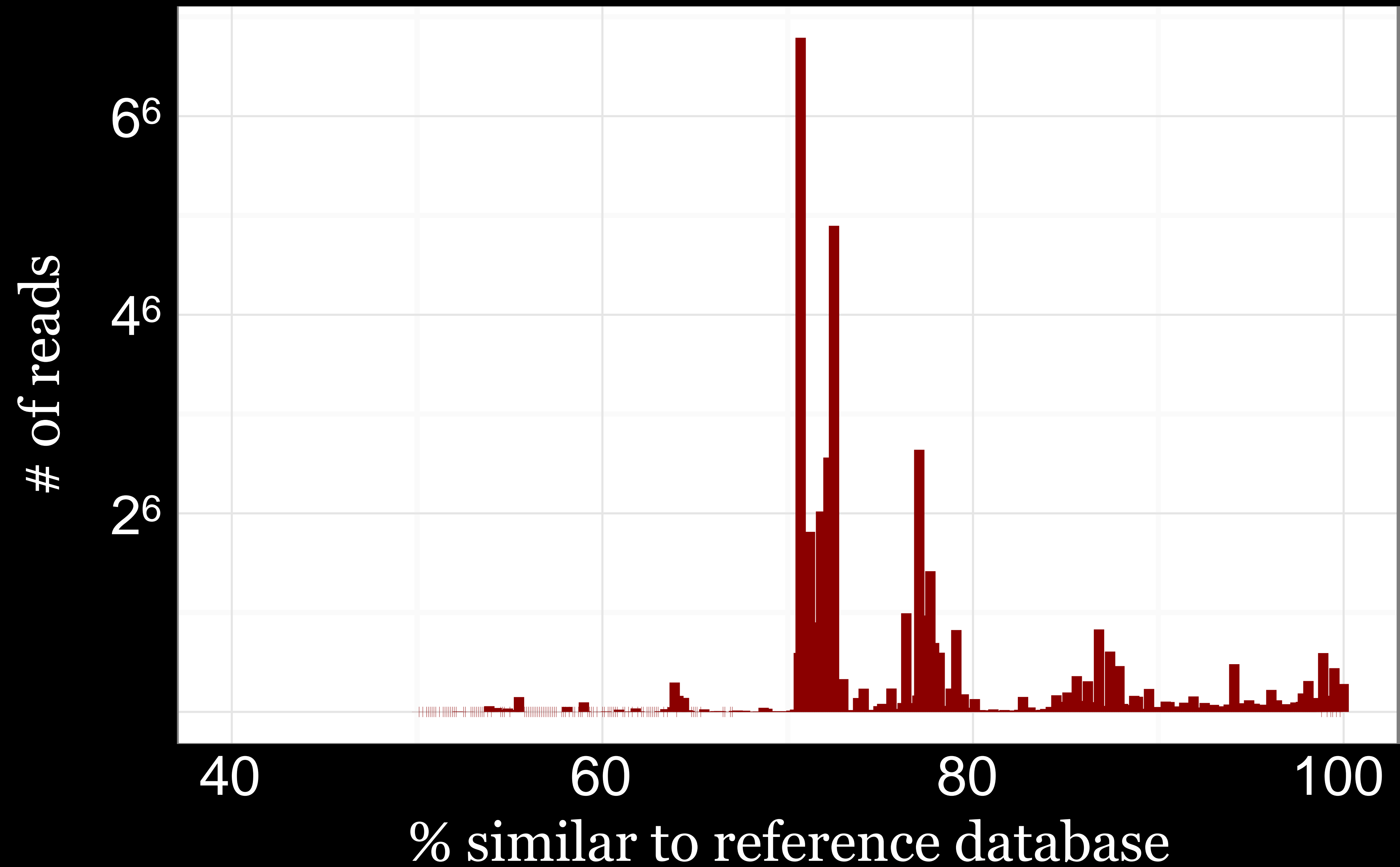
Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests

Frédéric Mahé¹, Colomban de Vargas^{2,3}, David Bass^{4,5}, Lucas Czech⁶, Alexandros Stamatakis^{6,7}, Enrique Lara⁸, David Singer⁸, Jordan Mayor⁹, John Bunge¹⁰, Sarah Sernaker¹¹, Tobias Siemensmeyer¹, Isabelle Trautmann¹, Sarah Romac^{2,3}, Cédric Berney^{2,3}, Alexey Kozlov⁶, Edward A. D. Mitchell^{8,12}, Christophe V. W. Seppey⁸, Elianne Egge¹³, Guillaume Lentendu¹, Rainer Wirth¹⁴, Gabriel Trueba¹⁵ and Micah Dunthorn^{1*}

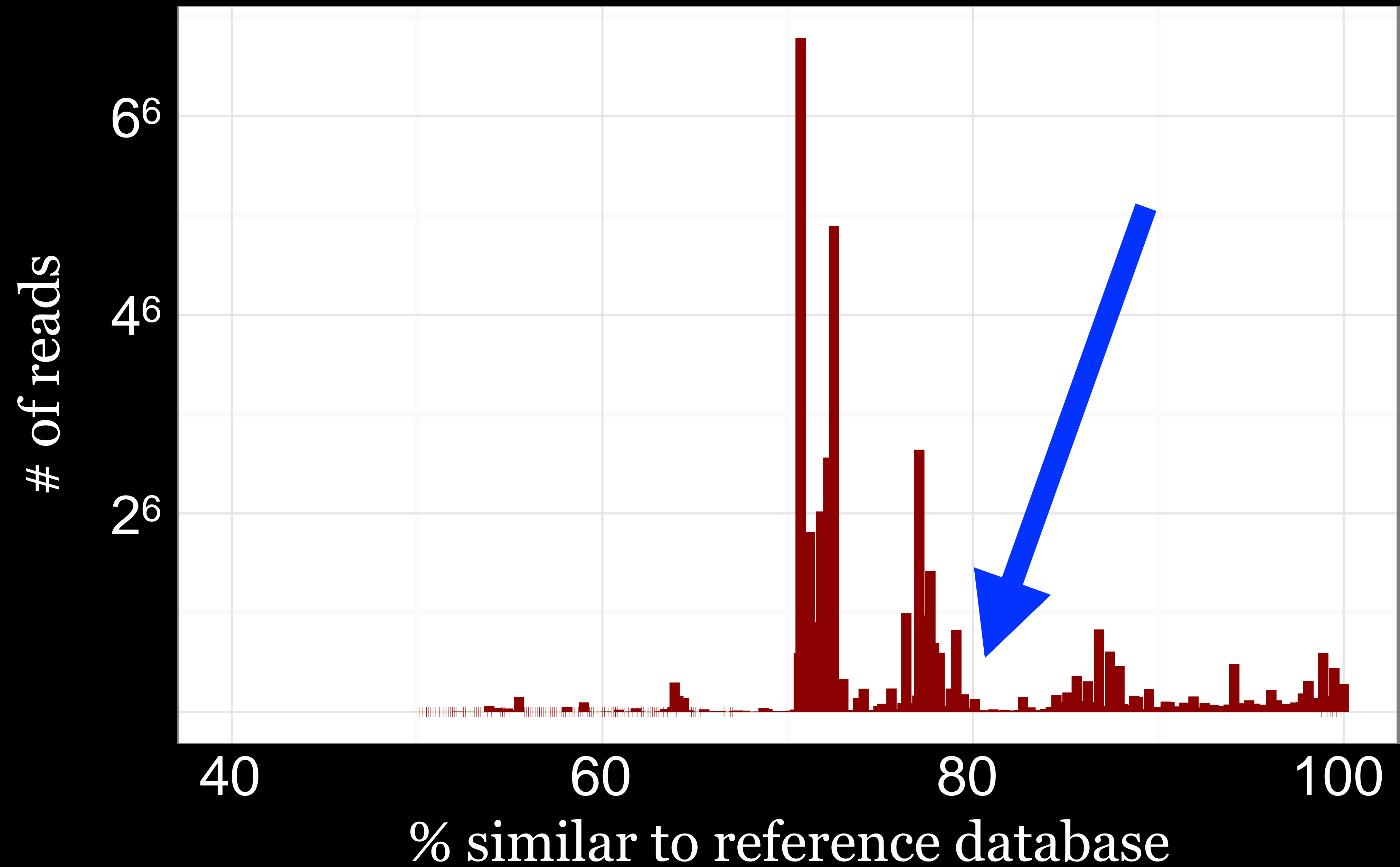
rainforest soil protists



rainforest soil protists



rainforest soil protists

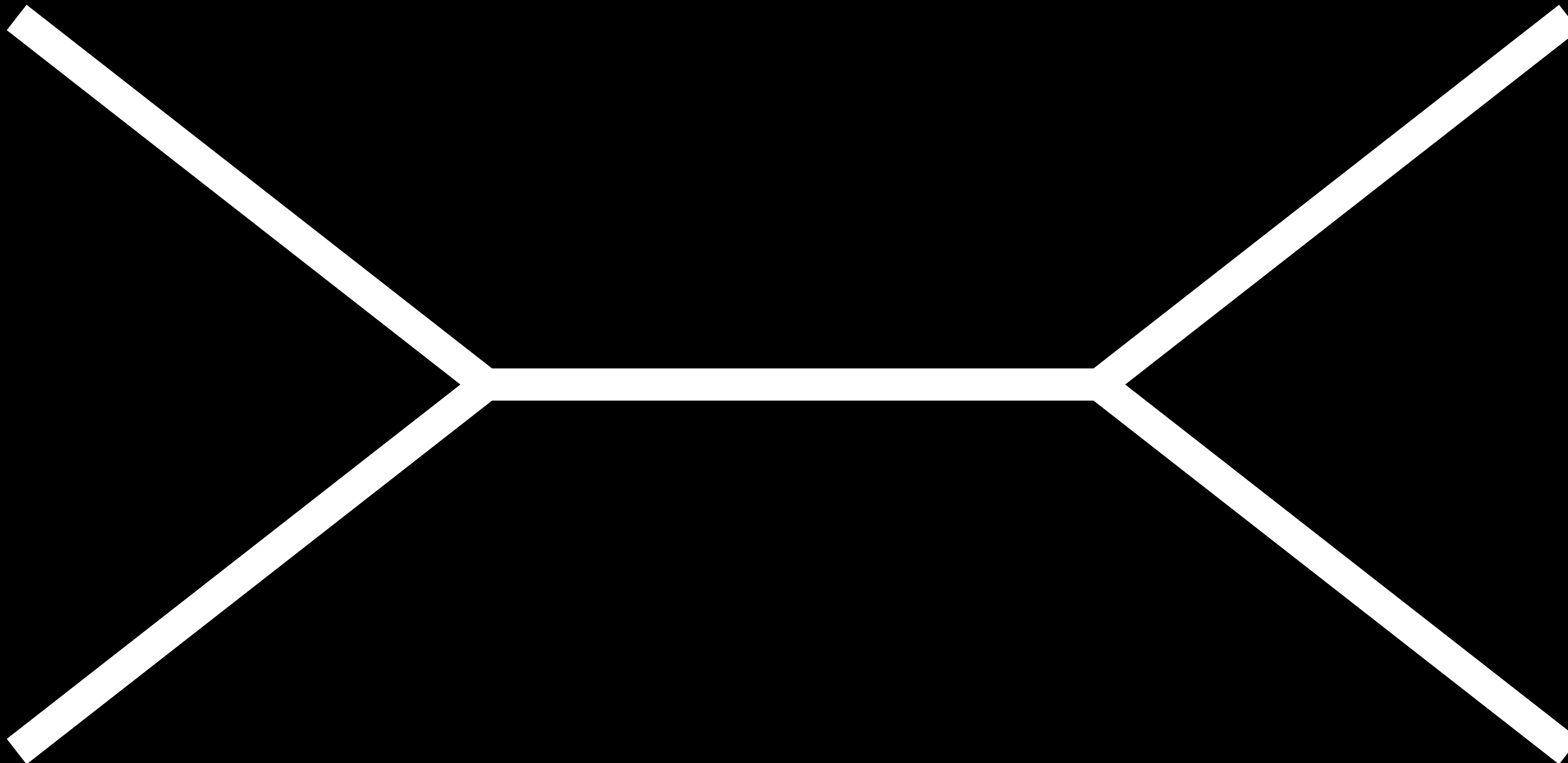


A1

B1

A2

B2

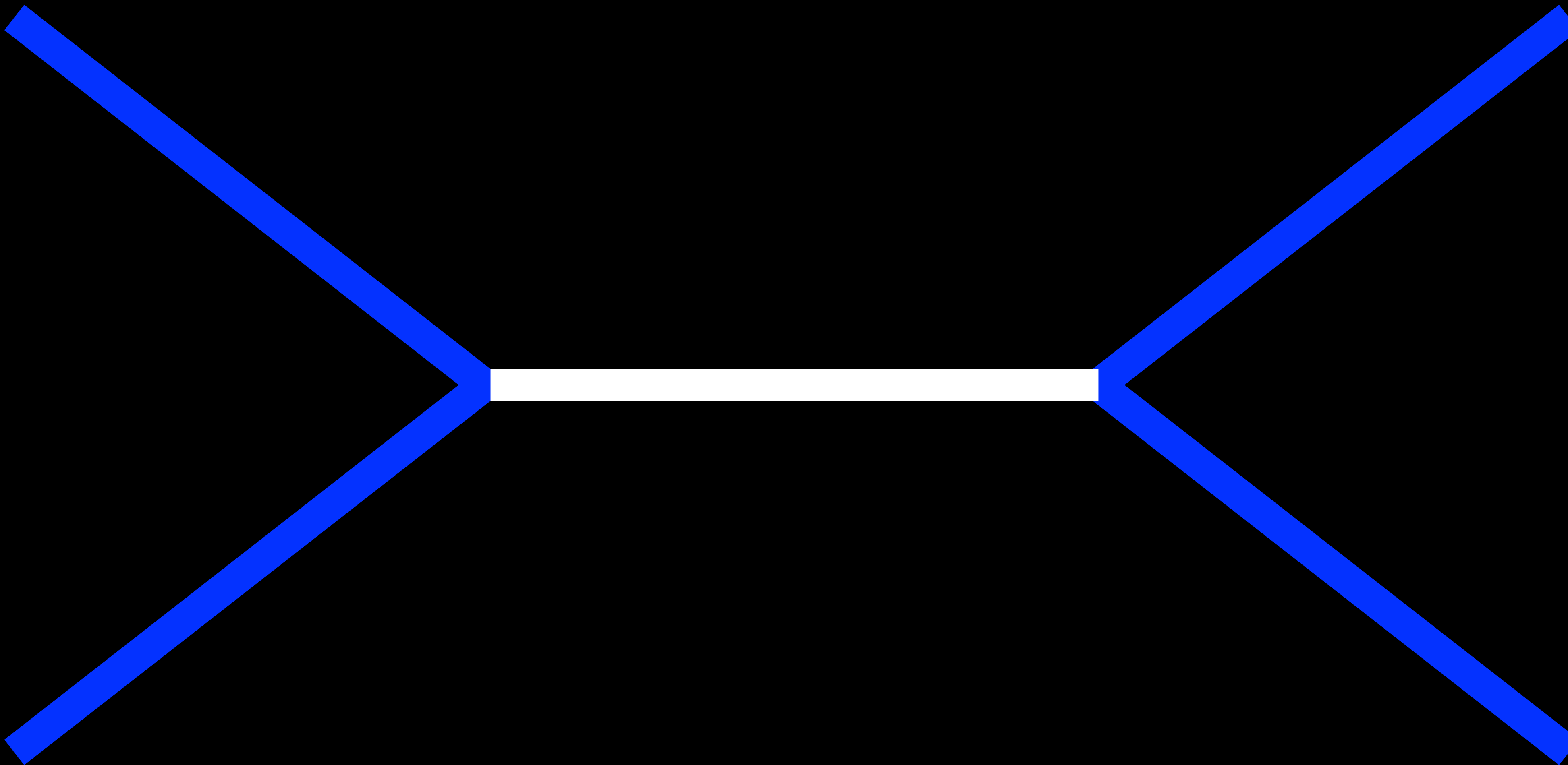


A1

B1

A2

B2

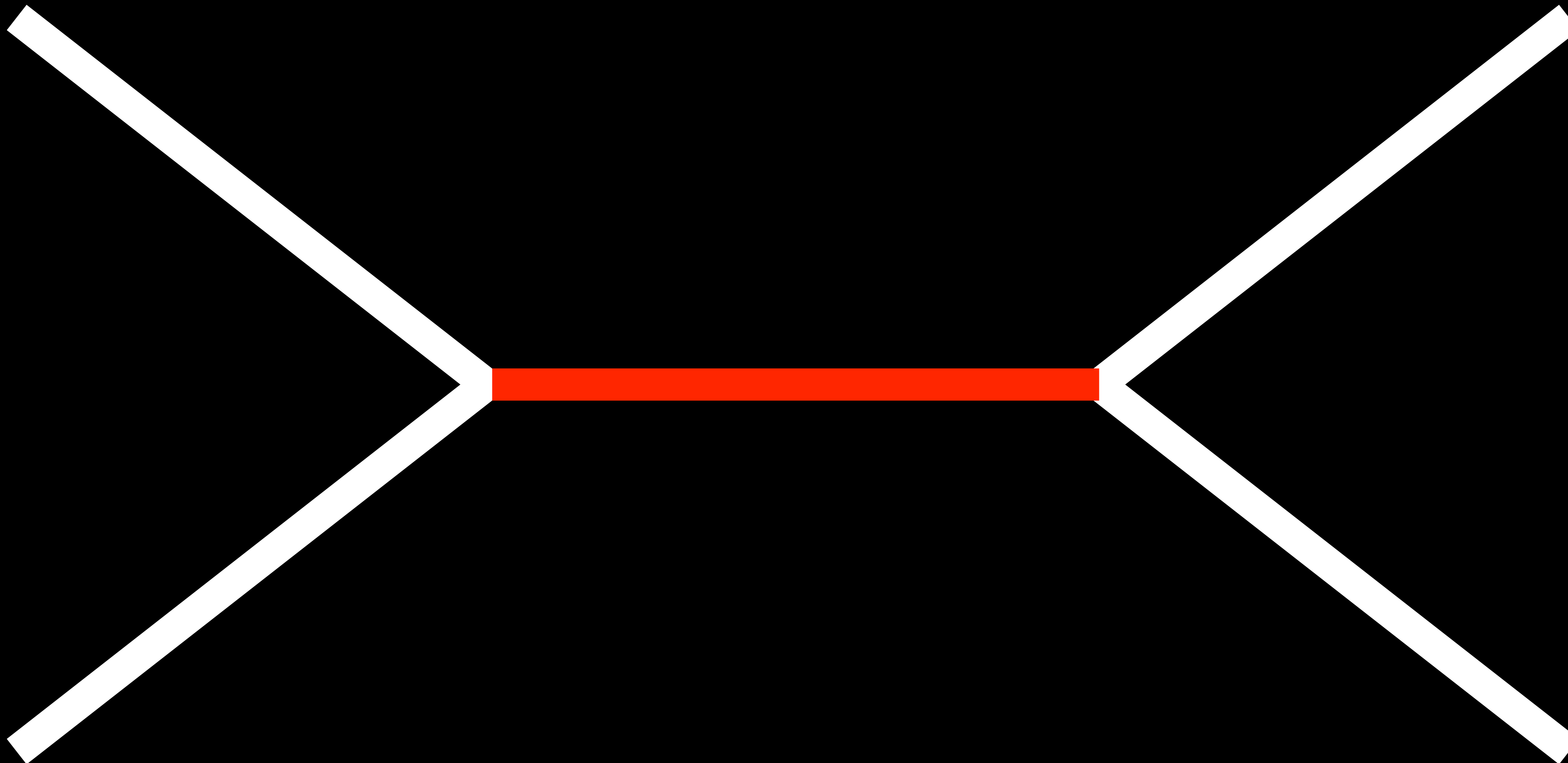


A1

B1

A2

B2

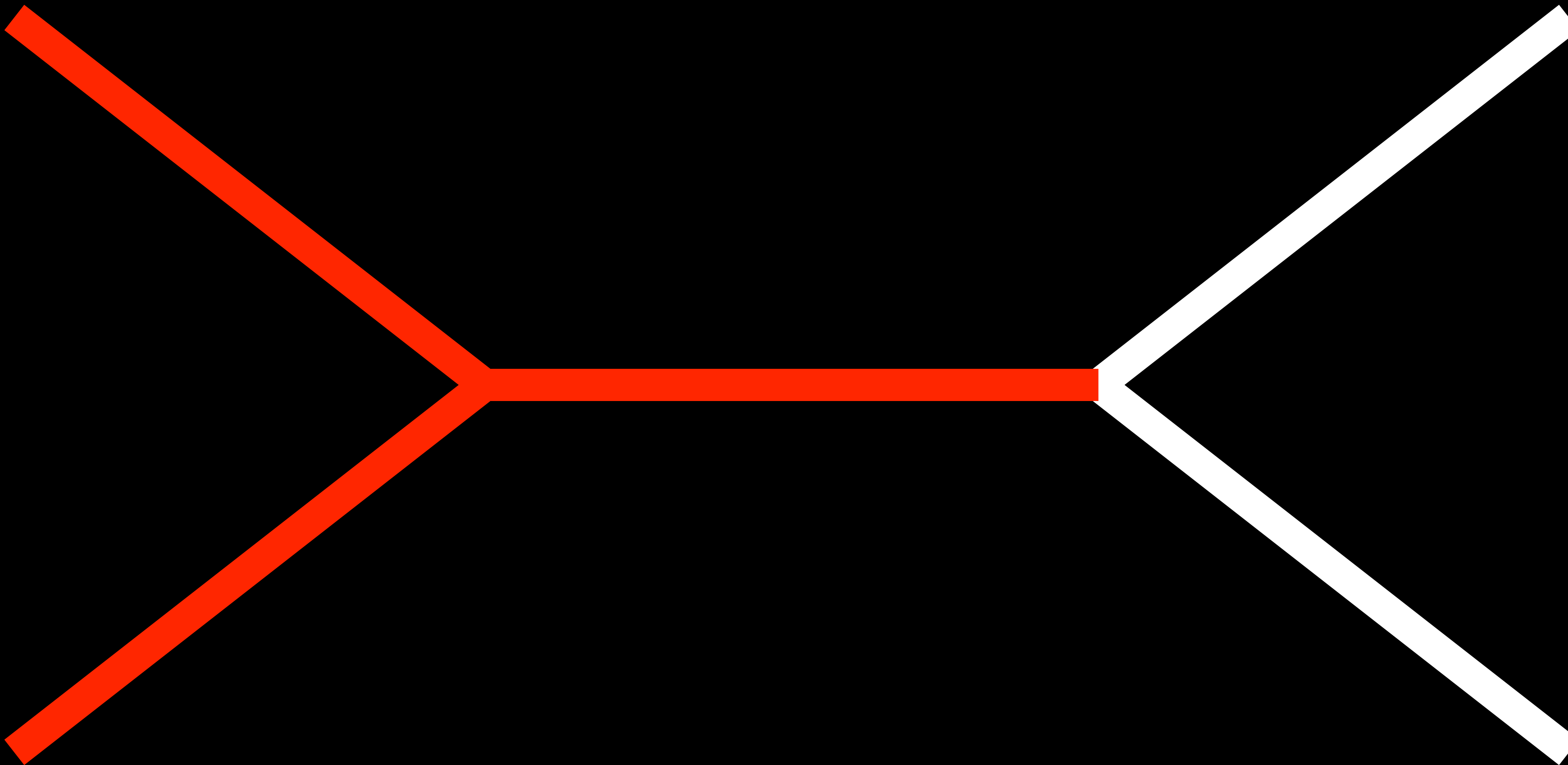


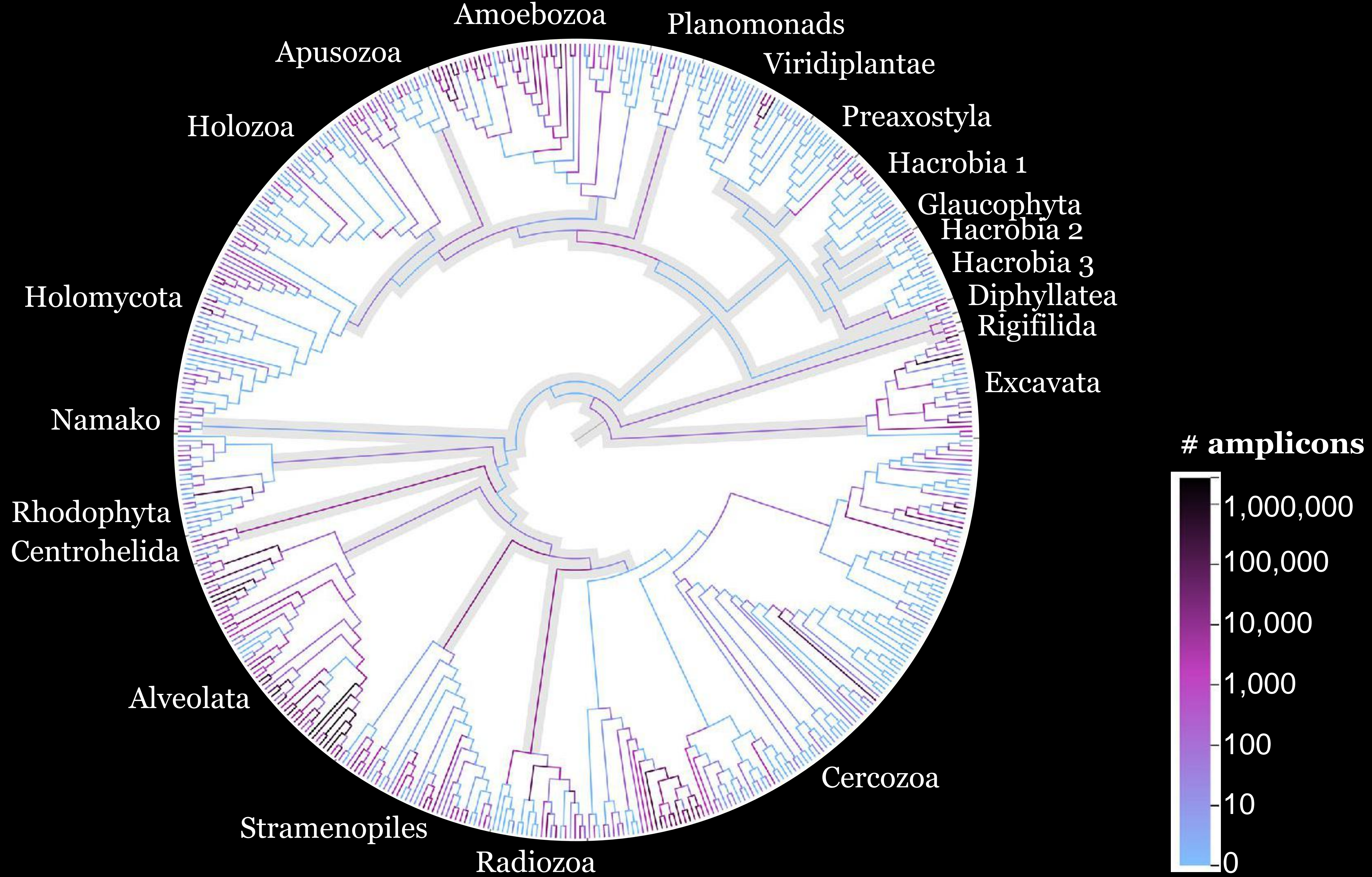
A1

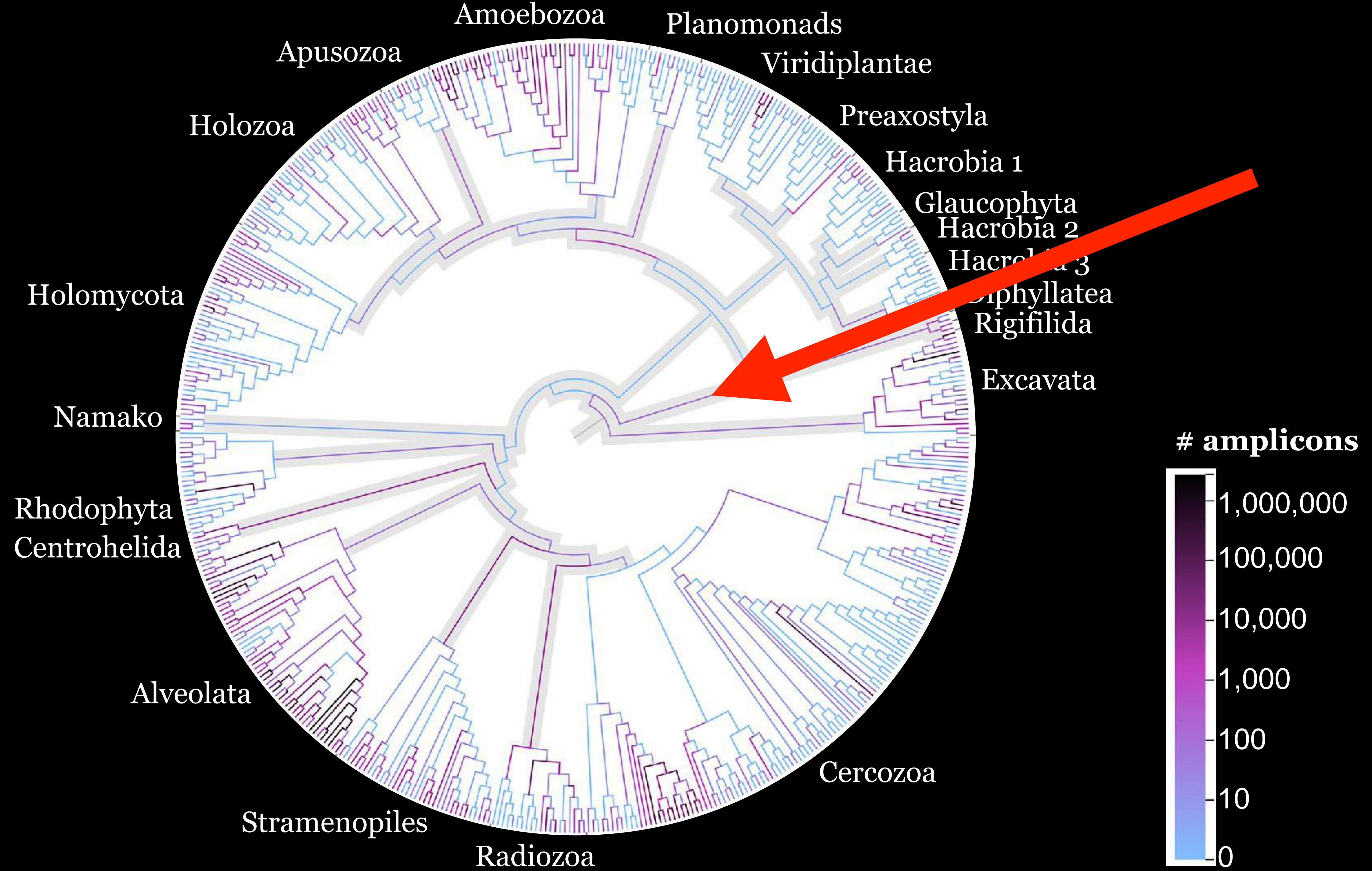
B1

A2

B2







What can it do?

clean your data

taxonomically assign reads/OTUs

2018; 65: 773–782

The Journal of
**Eukaryotic
Microbiology**

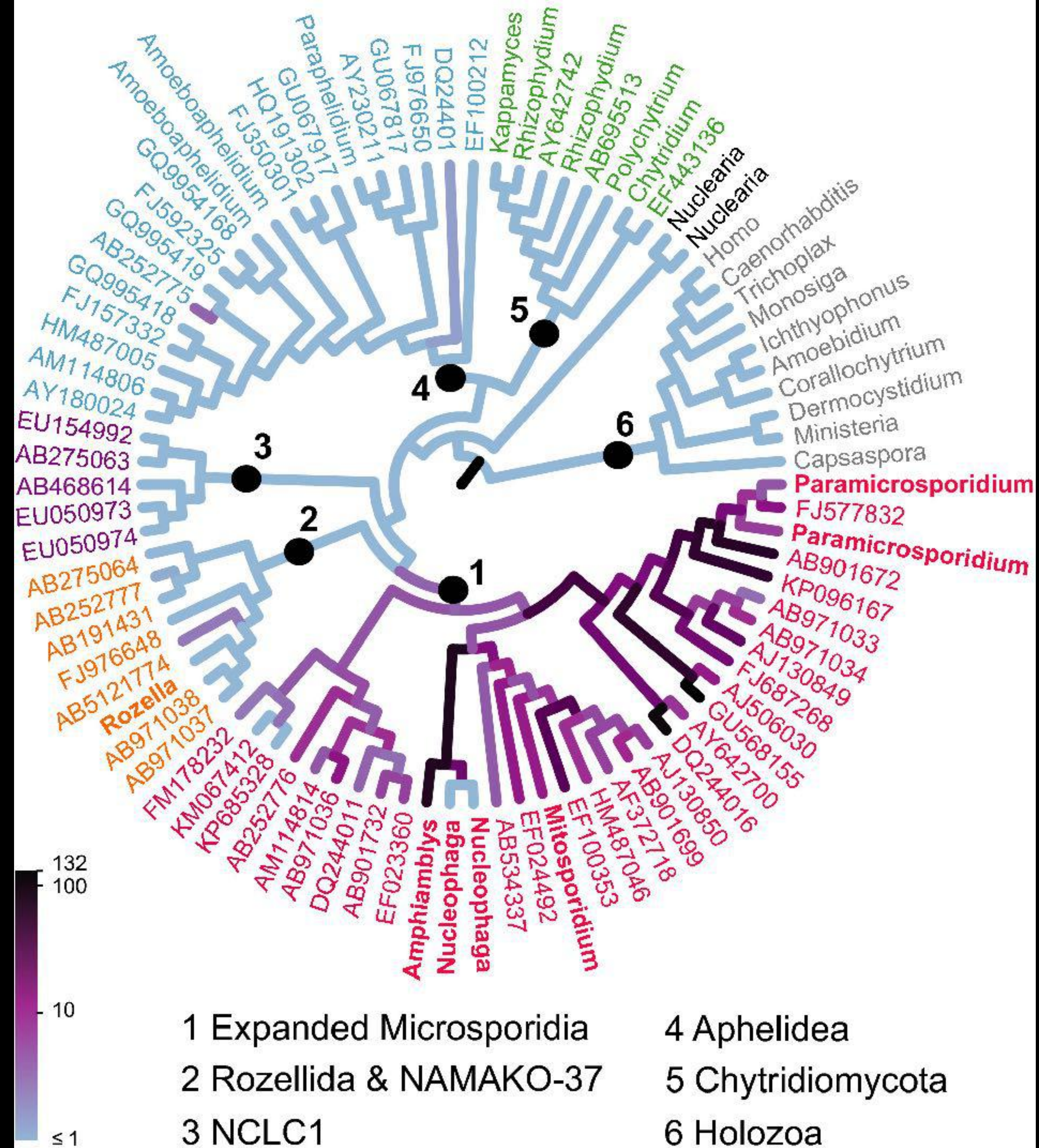
ISOP
Published by the
International Society
of Protistologists

Journal of Eukaryotic Microbiology ISSN 1066-5234

ORIGINAL ARTICLE

Clarifying the Relationships between Microsporidia and Cryptomycota

David Bass^{a,b}, Lucas Czech^c, Bryony A. P. Williams^d, Cédric Berney^e, Micah Dunthorn^f, Frederic Mahé^g,
Guifré Torruella^h, Grant D. Stentiford^a & Tom A. Williamsⁱ



2020; 20: 429–443

DOI: 10.1111/1755-0998.13117





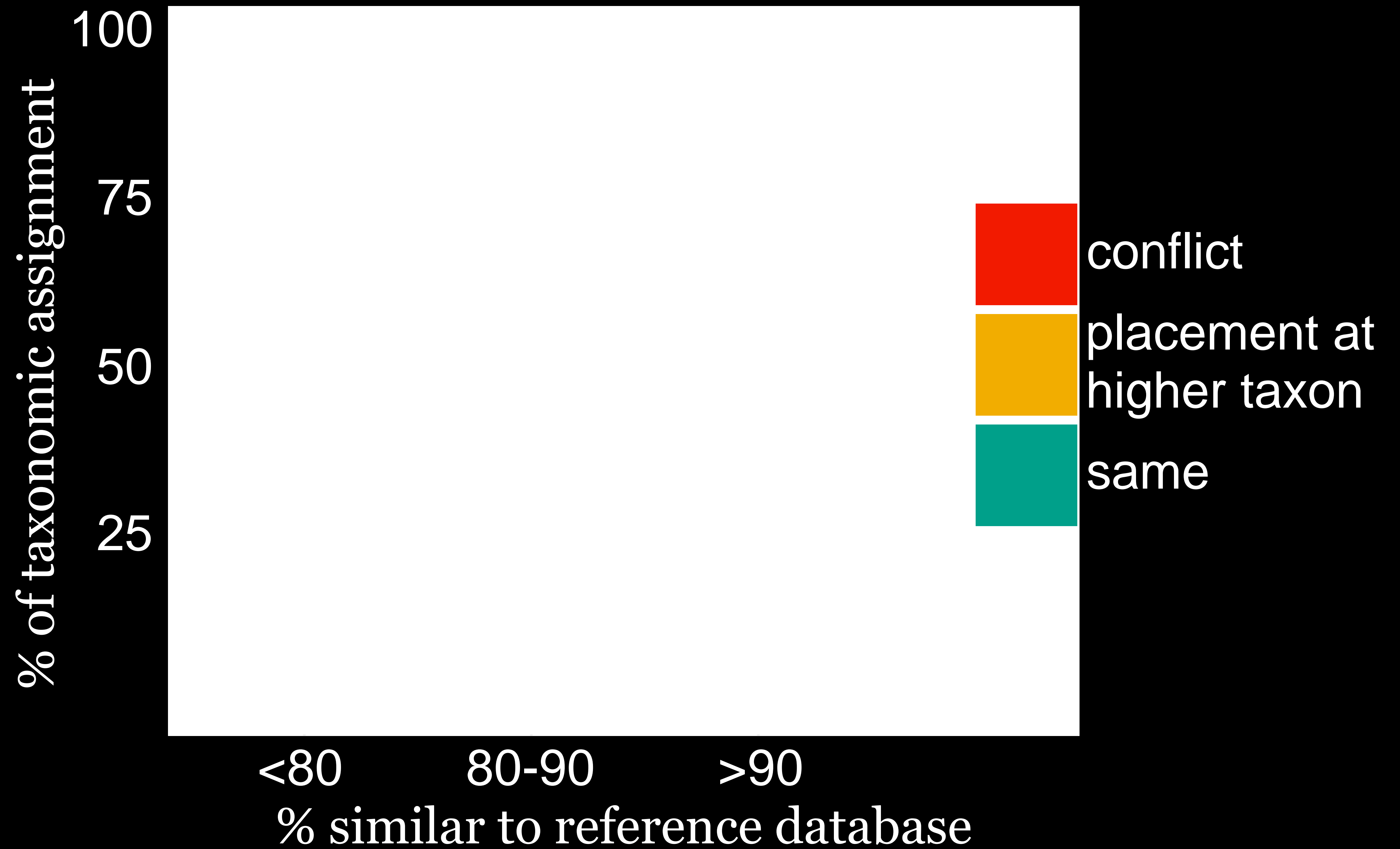
RESOURCE ARTICLE

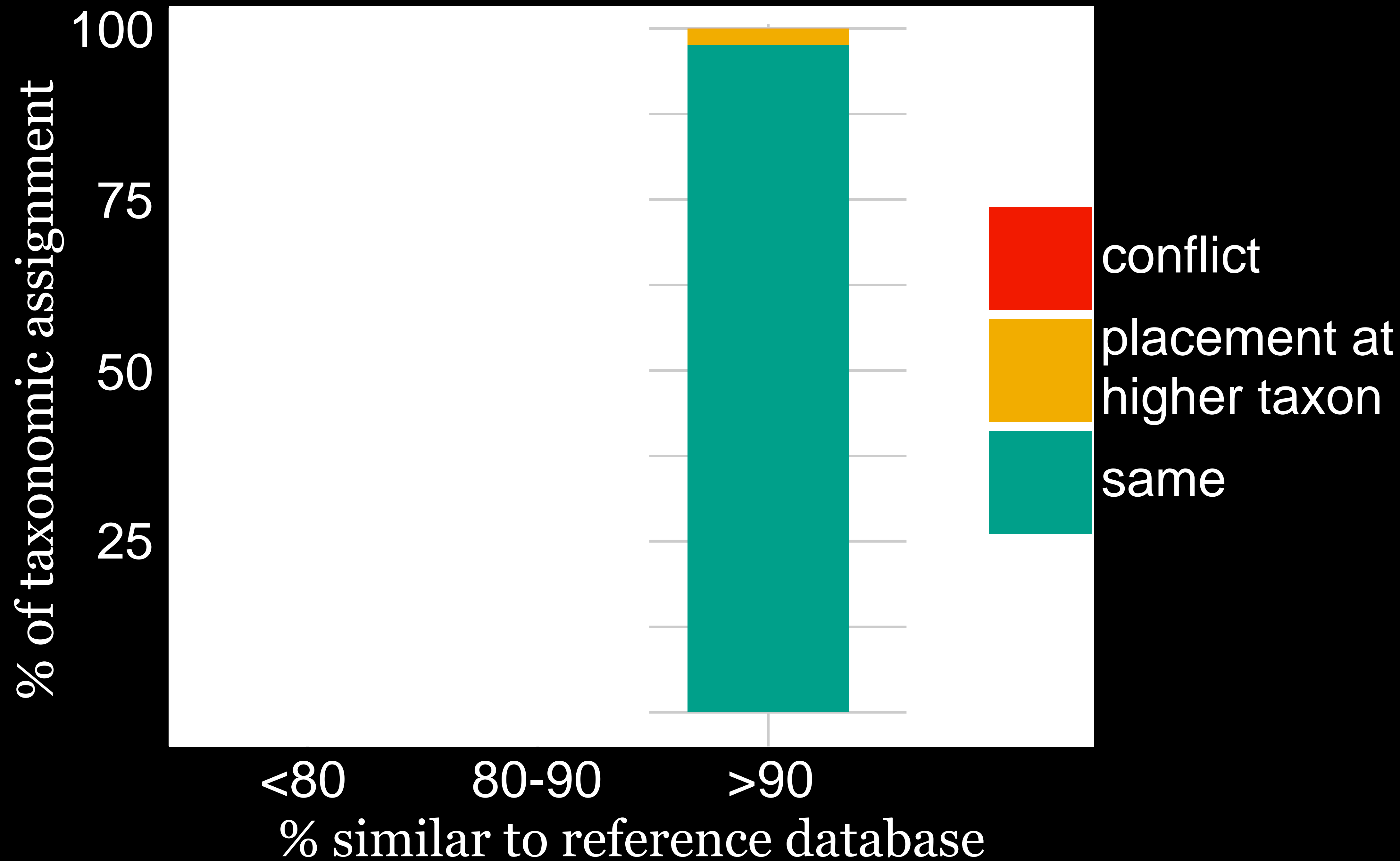
MOLECULAR ECOLOGY
RESOURCES

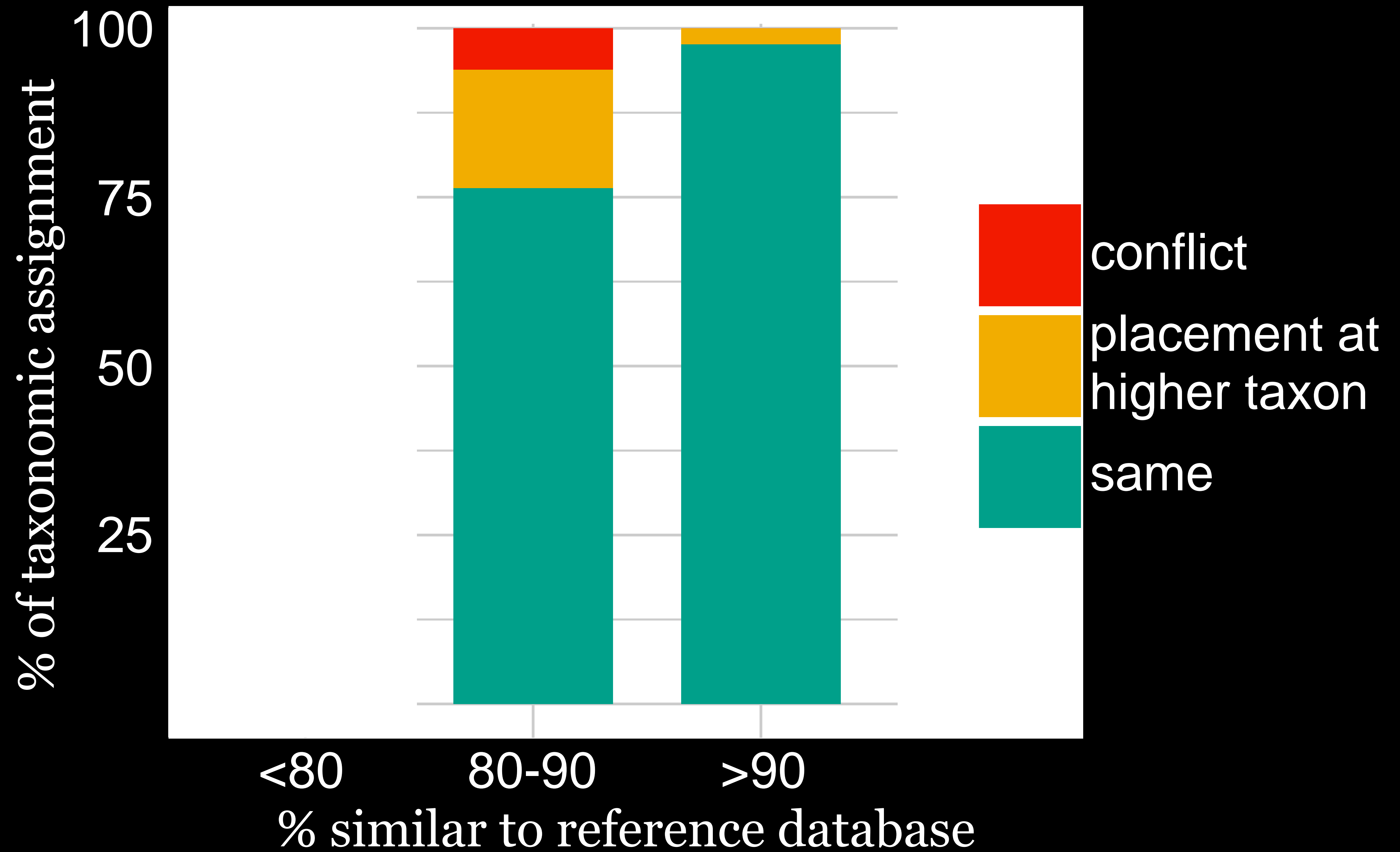
WILEY

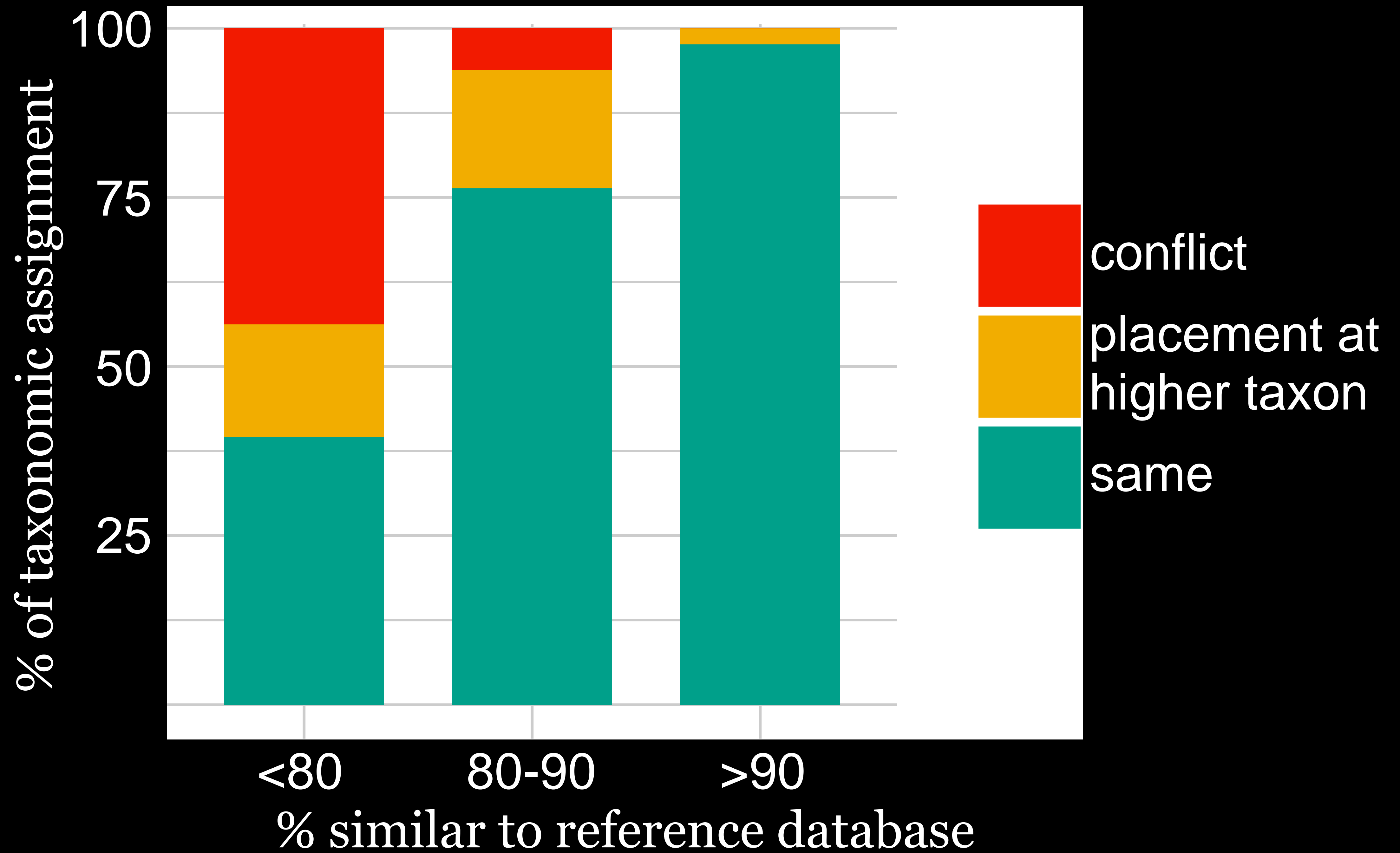
Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity

Mahwash Jamy¹  | Rachel Foster² | Pierre Barbera³ | Lucas Czech³ |
Alexey Kozlov³ | Alexandros Stamatakis^{3,4} | Gary Bending⁵ | Sally Hilton⁵ |
David Bass^{2,6}  | Fabien Burki¹









What can it do?

clean your data

taxonomically assign reads/OTUs

community analyses & diversities

Bioinformatics, 36(10), 2020, 3263–3265

doi: 10.1093/bioinformatics/btaa070

Advance Access Publication Date: 4 February 2020

Applications Note

OXFORD

Phylogenetics

Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data

Lucas Czech ^{1,*}, Pierre Barbera ¹ and Alexandros Stamatakis ^{1,2,*}

¹Computational Molecular Evolution Group, Heidelberg Institute for Theoretical Studies, Heidelberg 69118, Germany and ²Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe 76131, Germany

Received: 5 March 2020 | Revised: 24 July 2020 | Accepted: 25 August 2020

DOI: 10.1111/1755-0998.13255

RESOURCE ARTICLE

**MOLECULAR ECOLOGY
RESOURCES**

WILEY

SCRAPP: A tool to assess the diversity of microbial samples from phylogenetic placements

Pierre Barbera¹  | Lucas Czech¹  | Sarah Lutteropp¹ | Alexandros Stamatakis^{1,2} 

The programs

The programs

pplacer

- no longer updated

Matsen *et al.* *BMC Bioinformatics* 2010, **11**:538
<http://www.biomedcentral.com/1471-2105/11/538>



METHODOLOGY ARTICLE

Open Access

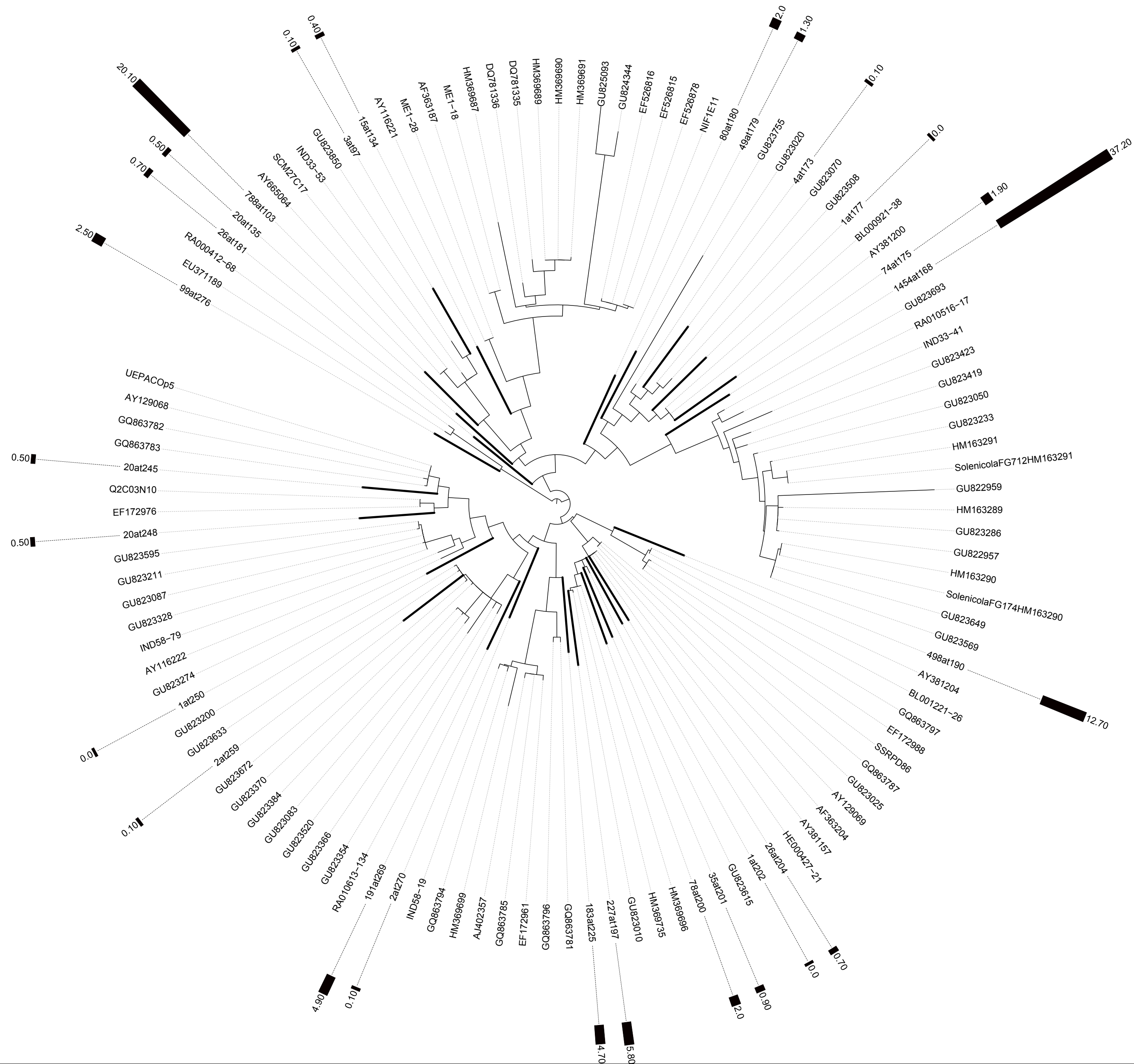
pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree

Frederick A Matsen^{1*}, Robin B Kodner^{2,3}, E Virginia Armbrust²

ORIGINAL ARTICLE

Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing

Ramiro Logares¹, Stephane Audic^{2,3}, Sebastien Santini⁴, Massimo C Pernice¹, Colomban de Vargas³ and Ramon Massana¹



The programs

pplacer

- no longer updated

EPA-ng

- absorbed the power of pplacer

Syst. Biol. 60(3):291–302, 2011

© The Author(s) 2011. Published by Oxford University Press on behalf of Society of Systematic Biologists.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

DOI:10.1093/sysbio/syr010

Advance Access publication on March 23, 2011

Performance, Accuracy, and Web Server for Evolutionary Placement of Short Sequence Reads under Maximum Likelihood

SIMON A. BERGER, DENIS KROMPASS, AND ALEXANDROS STAMATAKIS*

*The Exelixis Lab, Scientific Computing Group, Heidelberg Institute for Theoretical Studies,
Schloss-Wolfsbrunnengasse 35, D-69118 Heidelberg, Germany;*

**Correspondence to be sent to: The Exelixis Lab, Scientific Computing Group, Heidelberg Institute for Theoretical Studies,
Schloss-Wolfsbrunnengasse 35, D-69118 Heidelberg, Germany; E-mail: alexandros.stamatakis@epfl.ch.*

Syst. Biol. 68(2):365–369, 2019

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society of Systematic Biologists.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contactjournals.permissions@oup.com

DOI:10.1093/sysbio/syy054

Advance Access publication September 21, 2018

EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences

PIERRE BARBERA^{1,*}, ALEXEY M. KOZLOV¹, LUCAS CZECH¹, BENOIT MOREL¹, DIEGO DARRIBA^{1,2}, TOMÁŠ FLOURI^{1,3}, AND
ALEXANDROS STAMATAKIS^{1,4}

¹*Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnengasse 35, 69118 Heidelberg, Germany;*

²*Department of Computer Engineering, University of A Coruña, 15071 A Coruña, Spain;*

³*Department of Genetics, Evolution and Environment, University College London, Gower St., Bloomsbury, London WC1E 6BT, UK; and*

⁴*Karlsruhe Institute of Technology, Department of Informatics, Institute of Theoretical Informatics, Postfach 6980, 76128 Karlsruhe, Germany*

**Correspondence to be sent to: Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnengasse 35, 69118 Heidelberg, Germany;*

E-mail: pierre.barbera@h-its.org.

What you need

What you need

fasta file of query sequences

- e.g., OTU representatives, amplicons

>6ddabb21ed47f42bf10ca56c6aa26115a83f5dbe;size=283759; 98.9 Eukaryota|Diaphoretickes|Sar|Alveolata|Ciliophora|Intramacronucleata|CON3P|Colpodea|Platyophryida|g:Platyophrya|*
AGCTCCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTCGAAGTTCTGGCTGGCCCTGTTTCCGGGCTCC
GGTCGACTCAGGGCCTAGTCATCCGTCGGAGAACCCTTGTCGTCCTTCACTGGTCGCTTGGGGGGAAGTACCCTTTACT
TTGAAAAAATTAGAGTGTTTCAGGCAGGCTCTCGCCCGAATACTCCAGCATGGAATAATGGAATAGGACTTTCTTCTATT
TTGTTGGTTATTAGTGGAAGTAATGATTAATAGGGACAGTTGGGGGCATTCGTACTTAATTGTCAGAGGTGAAATTCT
TGGATTTATTAAAGACGAACTAATGCGAAAGCATTTGCCAAGGATGTTTTCA

>1649886e4424f4175304a2a51e72001553dd5564;size=111138; 98.9 Eukaryota|Diaphoretickes|Sar|Alveolata|Ciliophora|Intramacronucleata|CON3P|Colpodea|Platyophryida|g:Platyophrya|*
AGCTCCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTCGAAGTTCTGGCTGGTCCTGTTTCCGGGCTTC
GGTCGACTCAGGACCTAGTCATCCGTCGGAGAACCCTTGTCGTCCTTTACTGGTCGCTTGGGGGGAAGTACCCTTTACT
TTGAAAAAATTAGAGTGTTTCAGGCAGGCTCTCGCCCGAATACTCCAGCATGGAATAATGGAATAGGACTTTCTTCTATT
TTGTTGGTTATTAGTGGAAGTAATGATTAATAGGGACAGTTGGGGGCATTCGTACTTAATTGTCAGAGGTGAAATTCT
TGGATTTATTAAAGACGAACTAATGCGAAAGCATTTGCCAAGGATGTTTTCA

>db560755ofddf5a6703769e3ae284c72aec81f04;size=35834; 96.8 Eukaryota|Diaphoretickes|Sar|Alveolata|Ciliophora|Intramacronucleata|CON3P|Colpodea|Cyrtolophosidida|g:Microdiaphanosoma|Microdiaphanosoma+arcuatum
AGCTCCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGAATTTCTGGCCAATCTCATGCTGCGCTTCT
AGCCGGGTATGCGGTTTGGTCATCCGTAAGGGAAGCCAGTACGAACTTCAATGGTCGGCTGGGGGGAAGTTACAATTT
ACTGTGAAAAAATTAGAGTGTTTCAGGCAGGCAATCGCTTGGATACTCCAGCATGGAATAATGGAATAGGACTTTGG
TCTATTTGTTGGTTTGAGATCAAAGTAATGATTAATAGGGACAGTTGGGGGCATTCGTATTTAATTGTCAGAGGTGAAA
TTCTTGGATTTATTAAAGACGGACTTATGCGAAAGCATTTGCCAAGGATGTTTTCA

What you need

fasta file of query sequences

- e.g., OTU representatives, amplicons

fasta reference alignment

What you need

fasta file of query sequences

- e.g., OTU representatives, amplicons

fasta reference alignment

newick reference tree

Reference alignment

Reference alignment

same gene amplified, but full length

Reference alignment

same gene amplified, but full length

“enough” sequences to easily visualise tree

Reference alignment

same gene amplified, but full length

“enough” sequences to easily visualise tree

at least two exemplars from “every taxon”

Reference alignment

same gene amplified, but full length

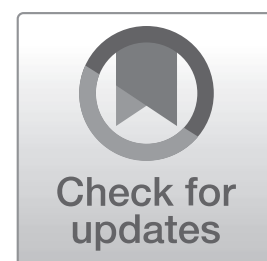
“enough” sequences to easily visualise tree

at least two exemplars from “every taxon”

do not mask amplified region



ELSEVIER



Available online at www.sciencedirect.com

ScienceDirect

European Journal of Protistology 77 (2021) 125747

European Journal of
PROTISTOLOGY

www.elsevier.com/locate/ejop

Colpodean ciliate phylogeny and reference alignments for phylogenetic placements

Lubomír Rajter^{a,*}, Isabelle Ewers^a, Nadine Graupner^a, Peter Vd'ačný^b, Micah Dunthorn^{a,c}

Reference tree

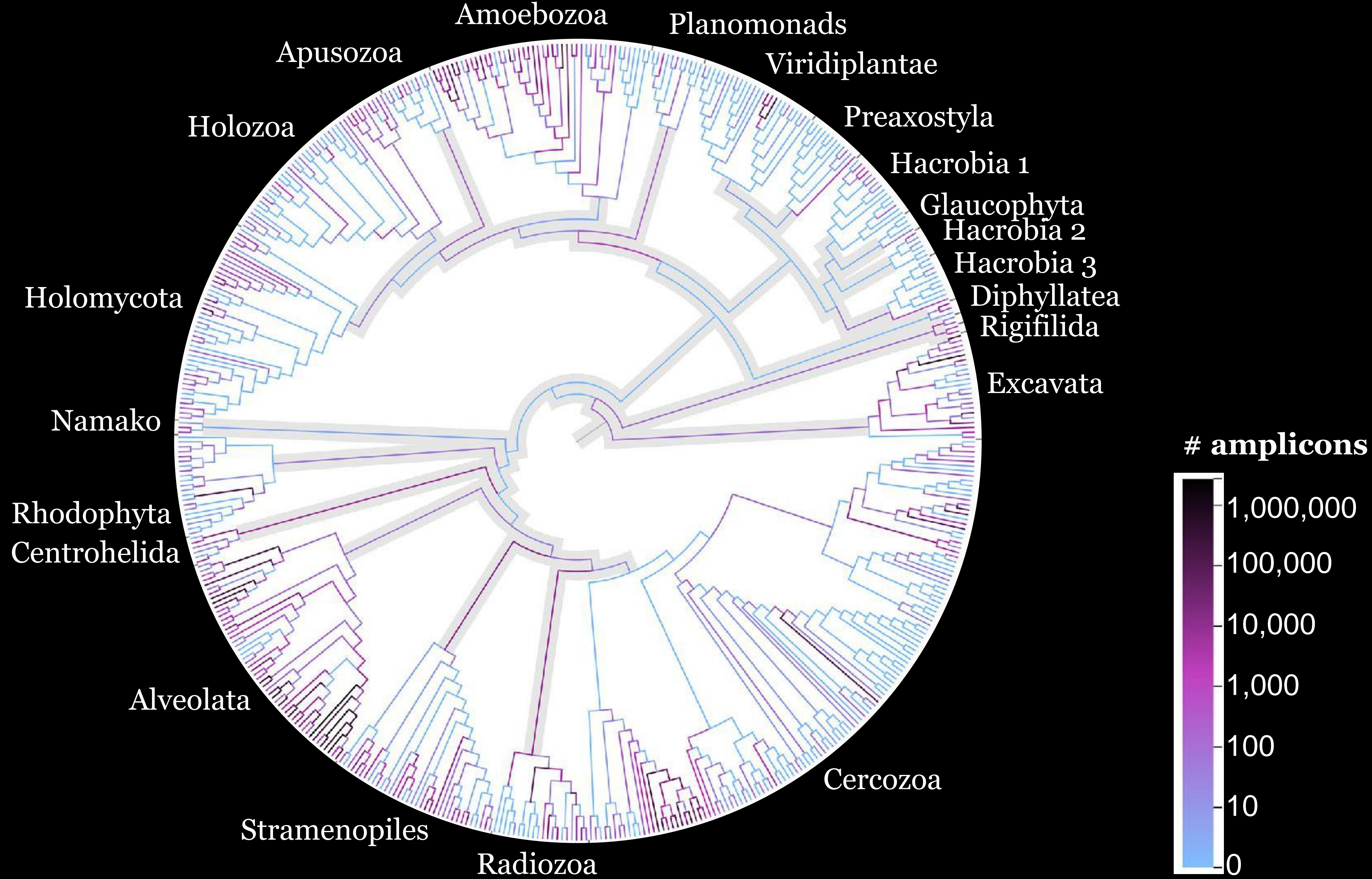
Reference tree

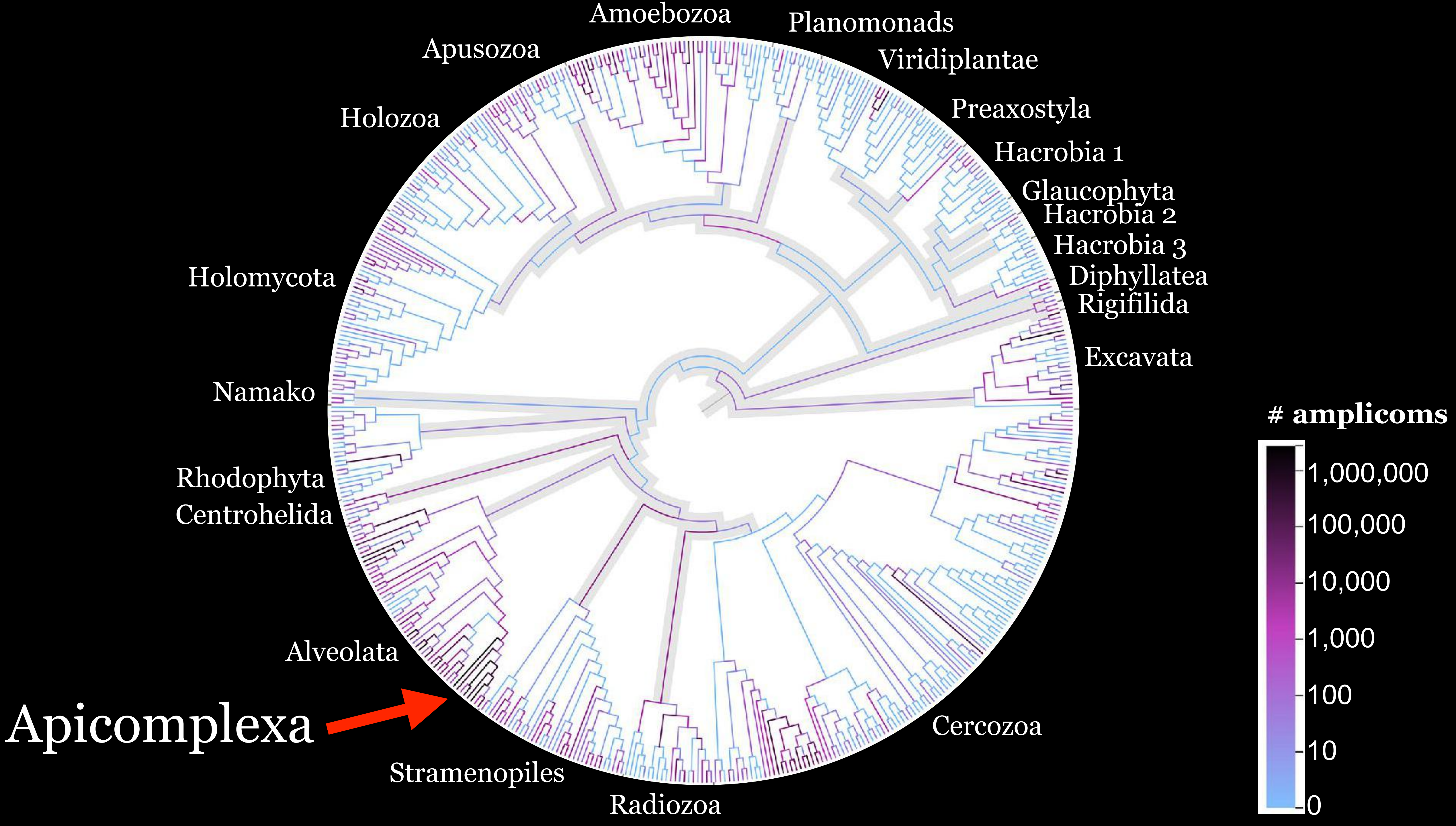
infer with RAxML

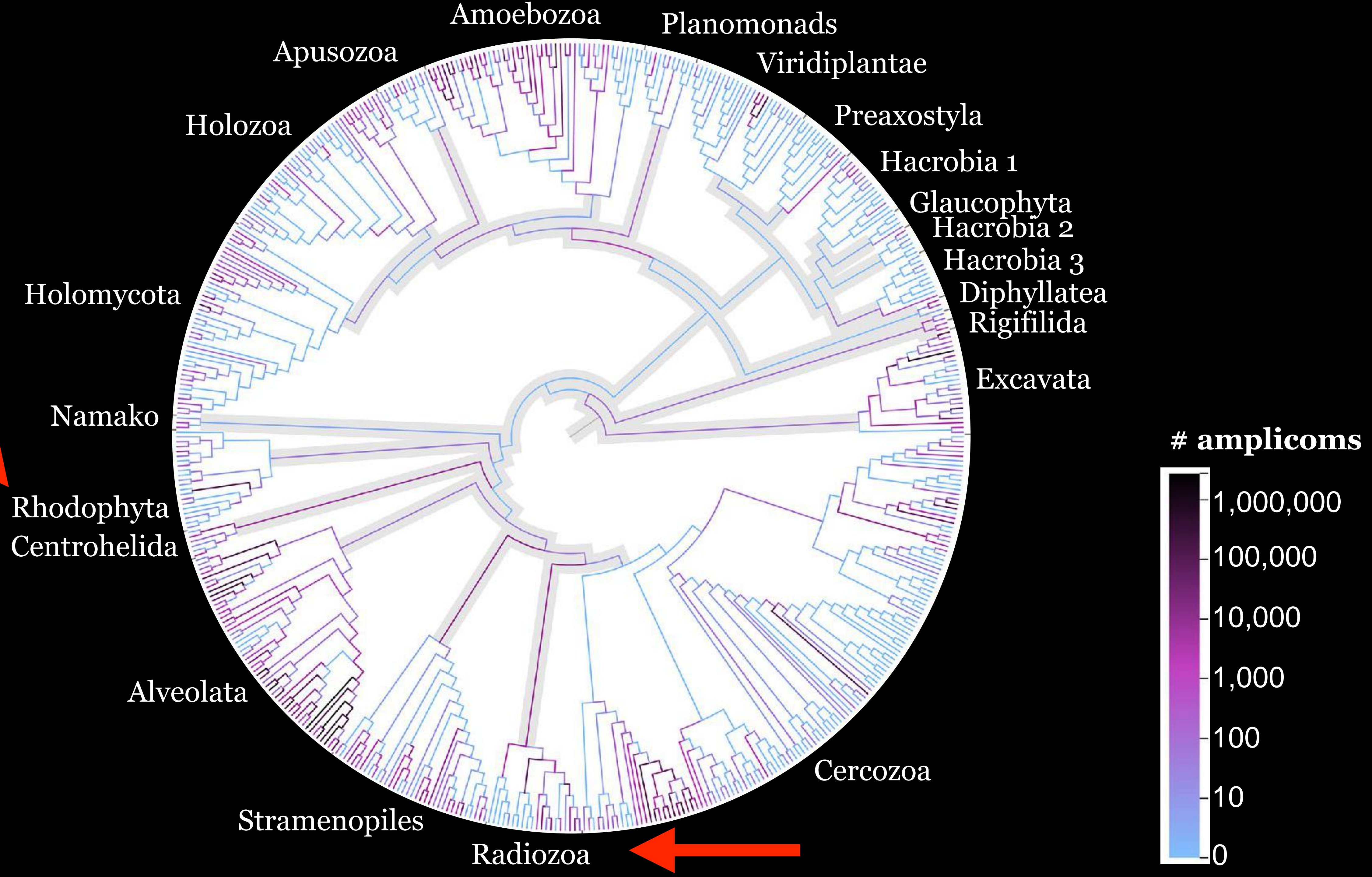
Reference tree

infer with RAxML

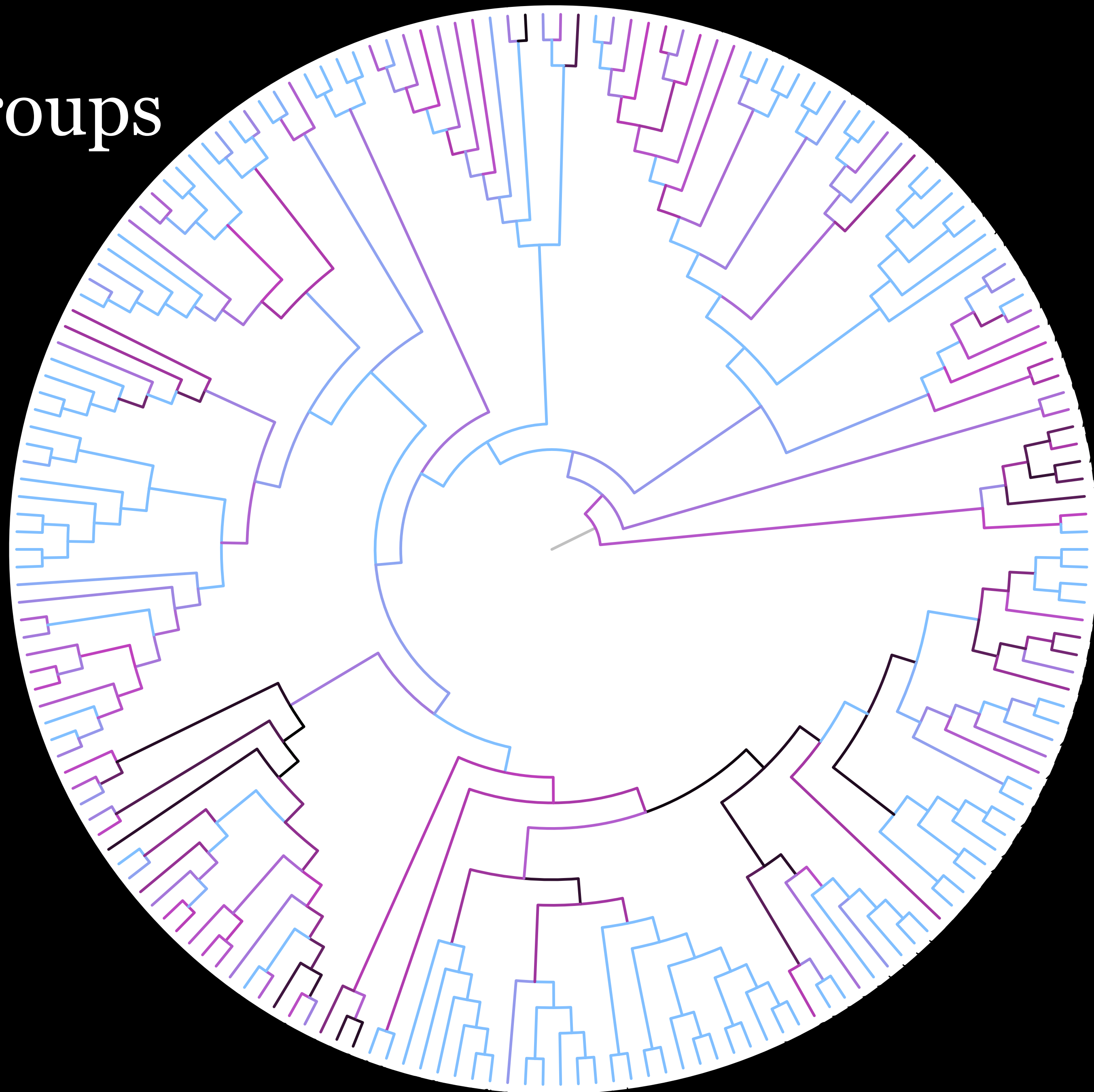
two-step trees to make sure you have
inferred the in-group correctly







Apicomplexa & close out-groups



Reference tree

infer with RAxML

two-step trees to make sure you have
inferred the in-group correctly

rooted before placements

Reference tree

infer with RAxML

two-step trees to make sure you have
inferred the in-group correctly

rooted before placements

can constrain taxa to be monophyletic

