FUNDAÇÃO
UNIVERSIDADE
FEDERAL DE
MATO GROSSO DO SUL

Evolução fenotípica e Métodos Filogenéticos Comparativos

Diogo Borges Provete

Prof. Adjunto A

Campo Grande-MS
Agosto 2019

Avisos

- Disciplina no Moodle
    - Slides
    - Scripts
    - Material leitura/links
    - Aula seguinte liberada no fim da tarde (pegar script e materiais)
- Horários (8h30-12h = 14h-17h)
- Locais para almoçar
- Acesso à Internet no campus
- O que fazer na sexta à tarde?
- Apresentação alunos ☺
- Minha apresentação ☺
- Avaliação final => submetida via Moodle
- ”Nivelamento” em R e inferência filogenética => evitemos monólogos monótonos: Perguntem!

Programa da disciplina

- Segunda
    - Introdução
    - Revisão rápida sobre métodos de inferência filogenética
    - Métodos para lidar com megaphylogenies
    - Base de dados de atributos e topologias
    - Desenho de estudos comparativos
    - Prática com R
- Terça
    - Modelos de evolução de atributos contínuos e categóricos
    - Sinal filogenético para atributos contínuos e categóricos
- Quarta
    - PGLS
    - Estimativa de estado ancestral caracteres contínuos e categóricos

Programa da disciplina

- Quinta
  - Evolução correlacionada (Pagel’s method e threshold)
  - Ajuste matrizes VCV – rate matrix
  - Modelos multi-regime e multi-taxas – prática SURFACE

- Sexta pela manhã
  - Visualização de dados filogenéticos – demonstração (phytools e phylosignal)
  - Métodos para dados multivariados (bem resumido)
  - Dúvidas e discussão de projetos dos participantes

Aula 1

Introdução

Métodos de inferência filogenética

Bases de dados

Desenho de estudos comparativos

Outline

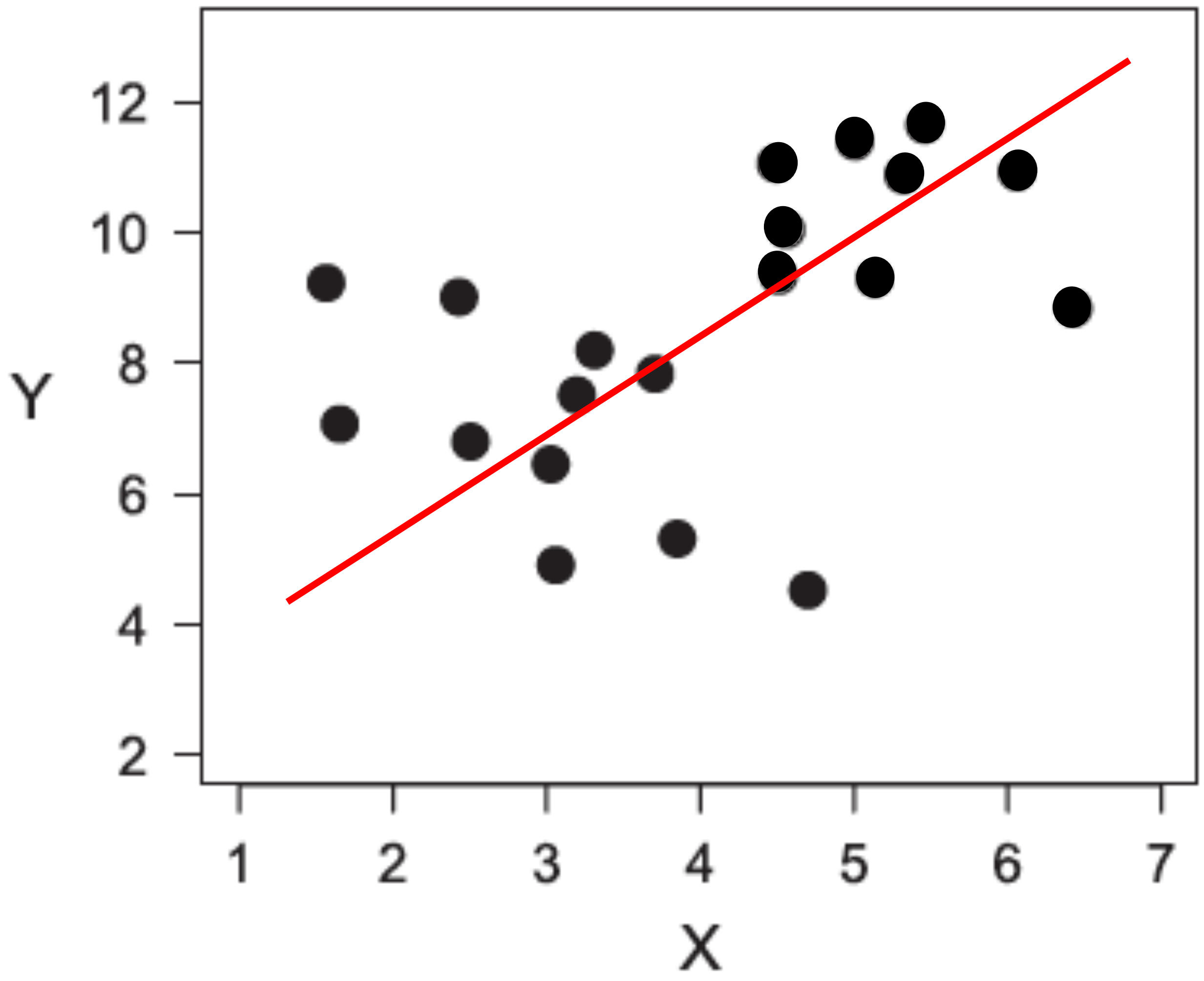
- Breve introdução histórico do desenvolvimento de métodos comparativos;
- métodos para reconstruir hipóteses filogenéticas (incluindo supertree, supermatrix)
- Darwinian e Rauniaerian shortfall.
- Iniciativas recentes parar suprir Darwinian shortfall: SUPERSMART, PASTIS, Open Tree of Life, congruification;
- Como trabalhar com base de dados online de atributos e filogenias.
- Perguntas comuns em estudos de adaptação;
- Como desenhar um estudo comparativo: phyndr e PhyloTargeting


MÉTODOS FILOGENÉTICOS COMPARATIVOS
O QUE SÃO? PARA QUE SERVEM? DE ONDE VIERAM?
VOCÊ VAI VER TUDO AQUI, HOJE NO GLOBO REP...OPS

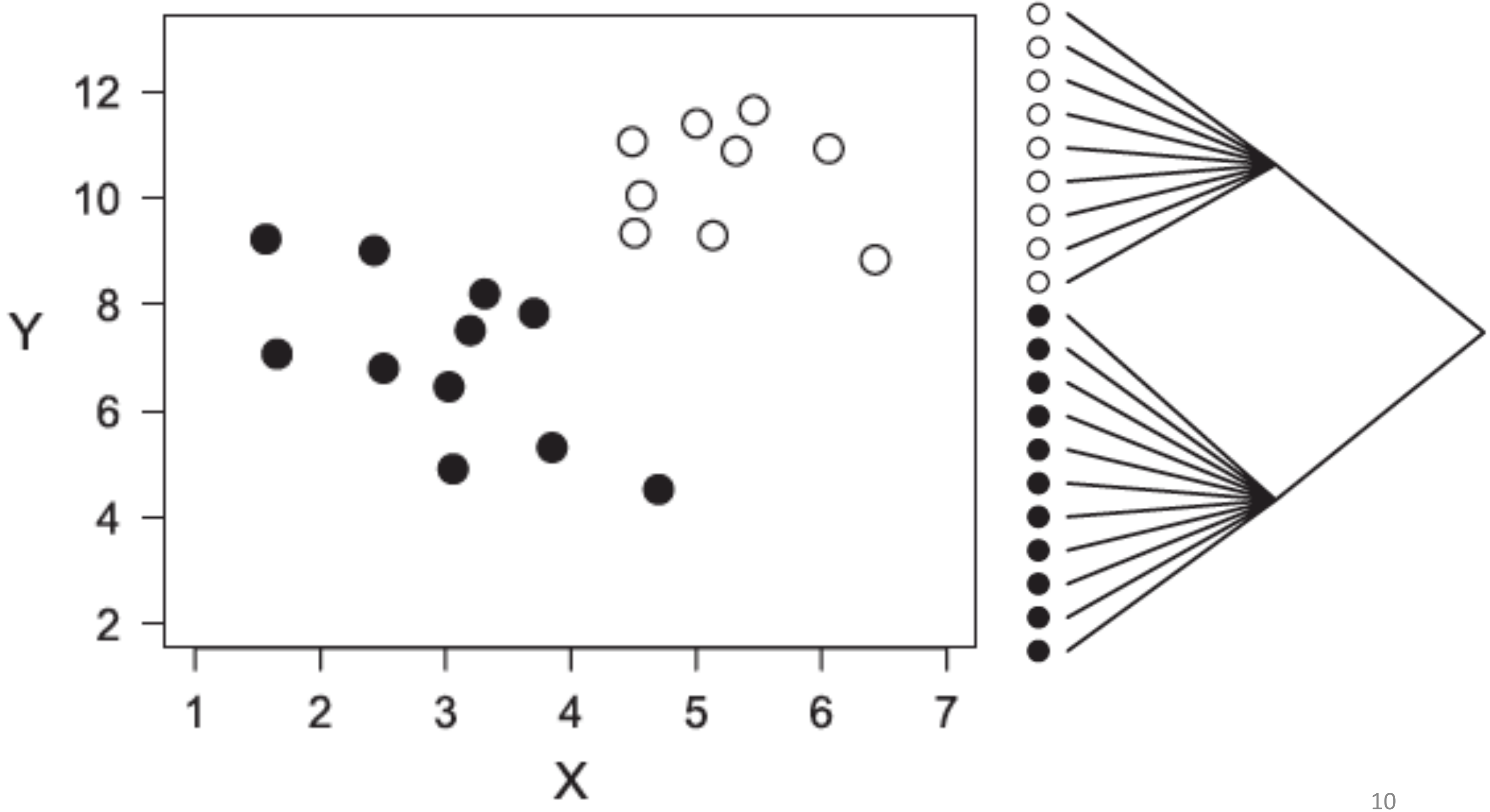
Introdução

- Conjunto de métodos estatísticos que usam a relação de parentesco entre as espécies para testar hipóteses evolutivas sobre seus atributos
- Diferença entre **obter** e **utilizar** uma filogenia
- Perguntas comuns
  - Qual o padrão filogenético de uma dada característica das espécies?
  - Sob que processo e em que velocidade essa característica evolui?
  - Qual a correlação dessa característica com outras ou com a variação ambiental?

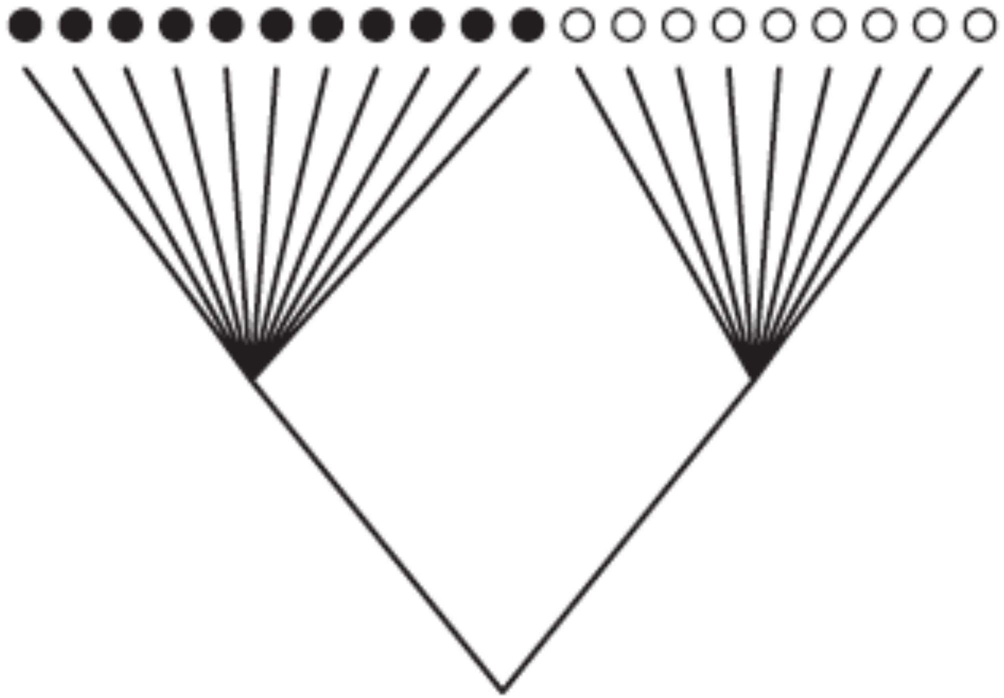
O problema do parentesco comum



O problema do parentesco comum

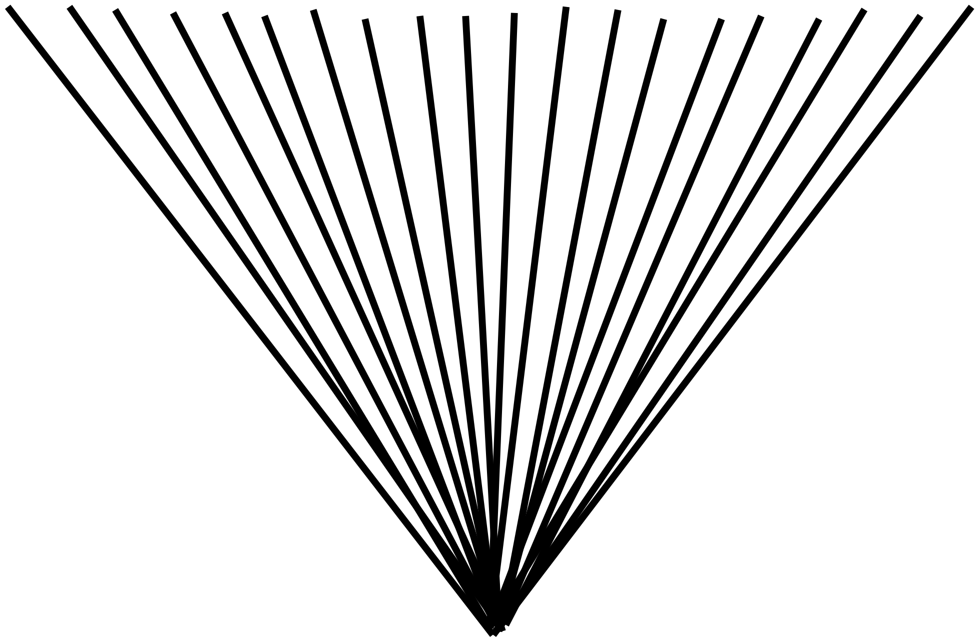


Espécies não são independentes



Filogenia "real"

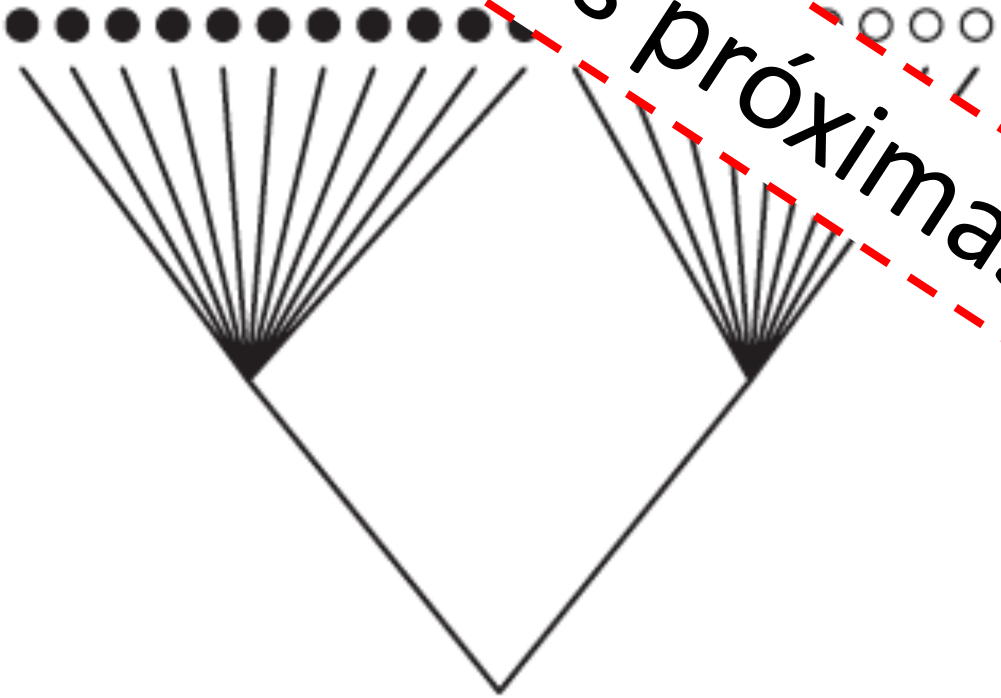
vs.



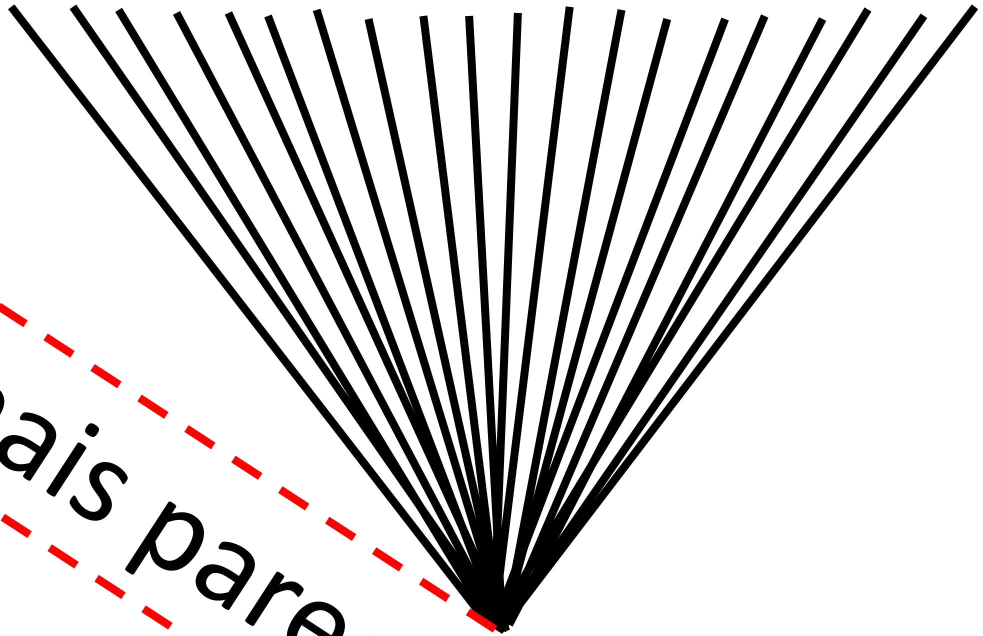
Filogenia em estrela

pécies não são independentes

Espécies próximas são mais parecidas



Filogenia "real"

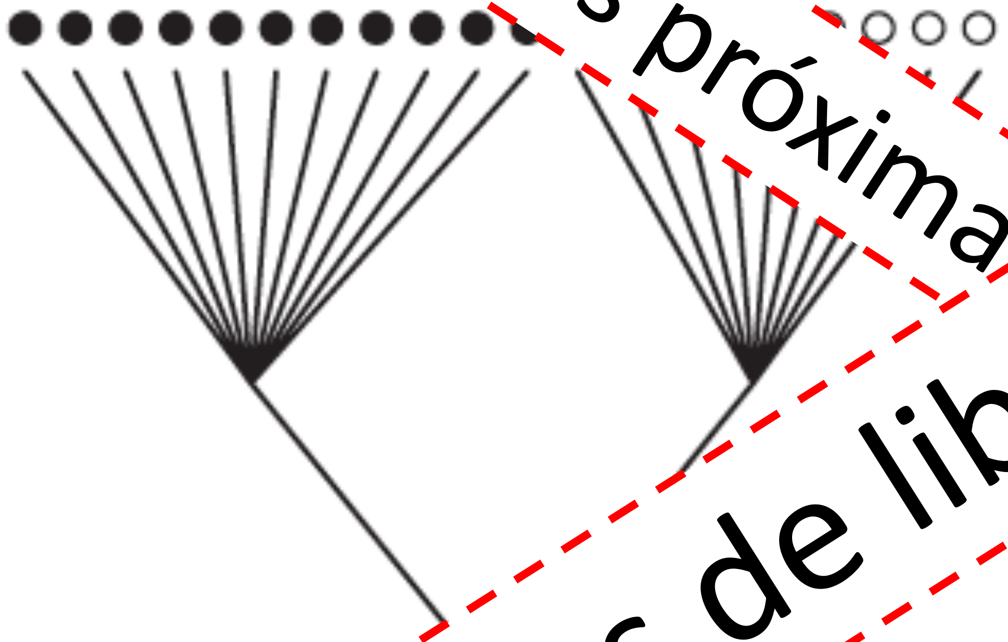


Filoge strela

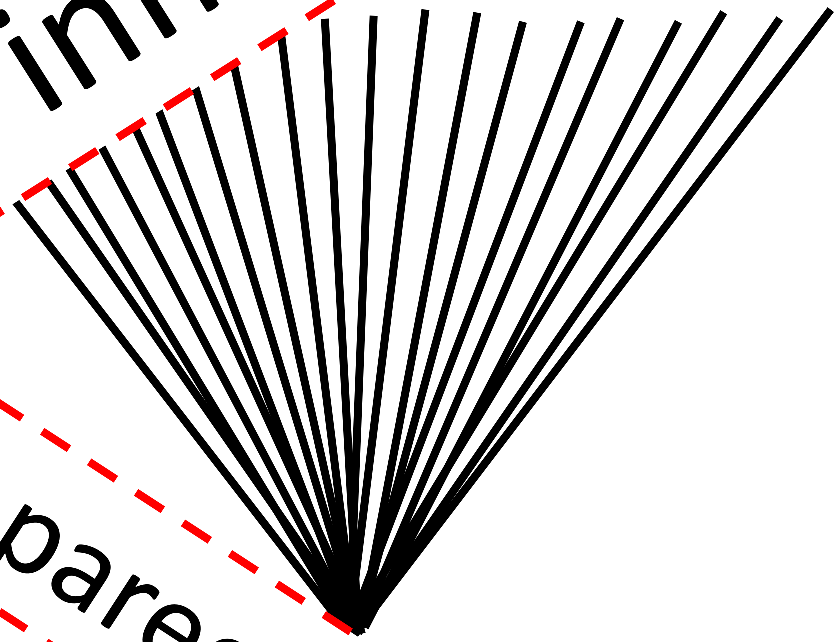
pécies não são independent

Espécies próximas mais parecidas

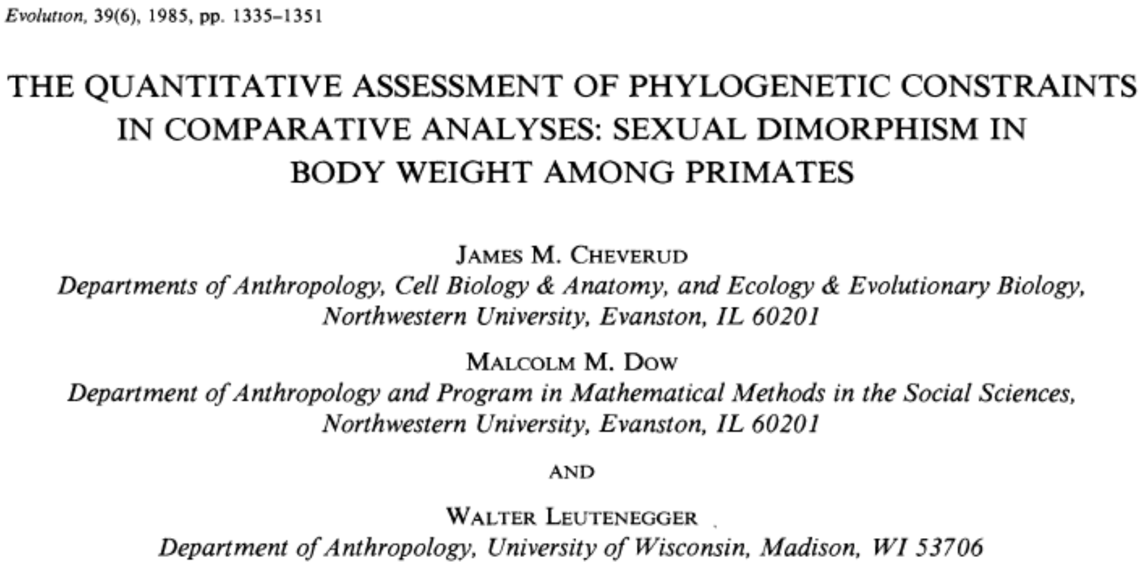
Graus de liberdade inflados



real”



Filoge strela



Evolution, 39(6), 1985, pp. 1335–1351

THE QUANTITATIVE ASSESSMENT OF PHYLOGENETIC CONSTRAINTS IN COMPARATIVE ANALYSES: SEXUAL DIMORPHISM IN BODY WEIGHT AMONG PRIMATES

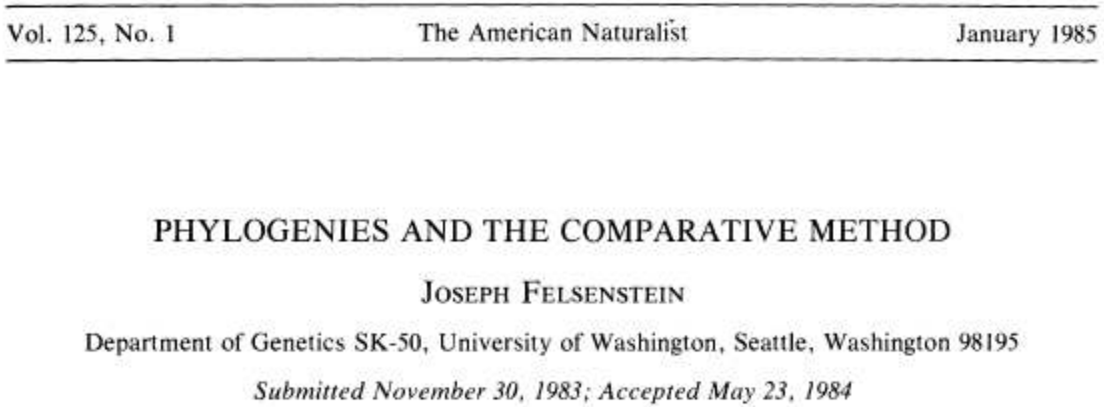
JAMES M. CHEVERUD
Departments of Anthropology, Cell Biology & Anatomy, and Ecology & Evolutionary Biology, Northwestern University, Evanston, IL 60201

MALCOLM M. DOW
Department of Anthropology and Program in Mathematical Methods in the Social Sciences, Northwestern University, Evanston, IL 60201

AND

WALTER LEUTENEGGER
Department of Anthropology, University of Wisconsin, Madison, WI 53706

1985



Vol. 125, No. 1 | The American Naturalist | January 1985

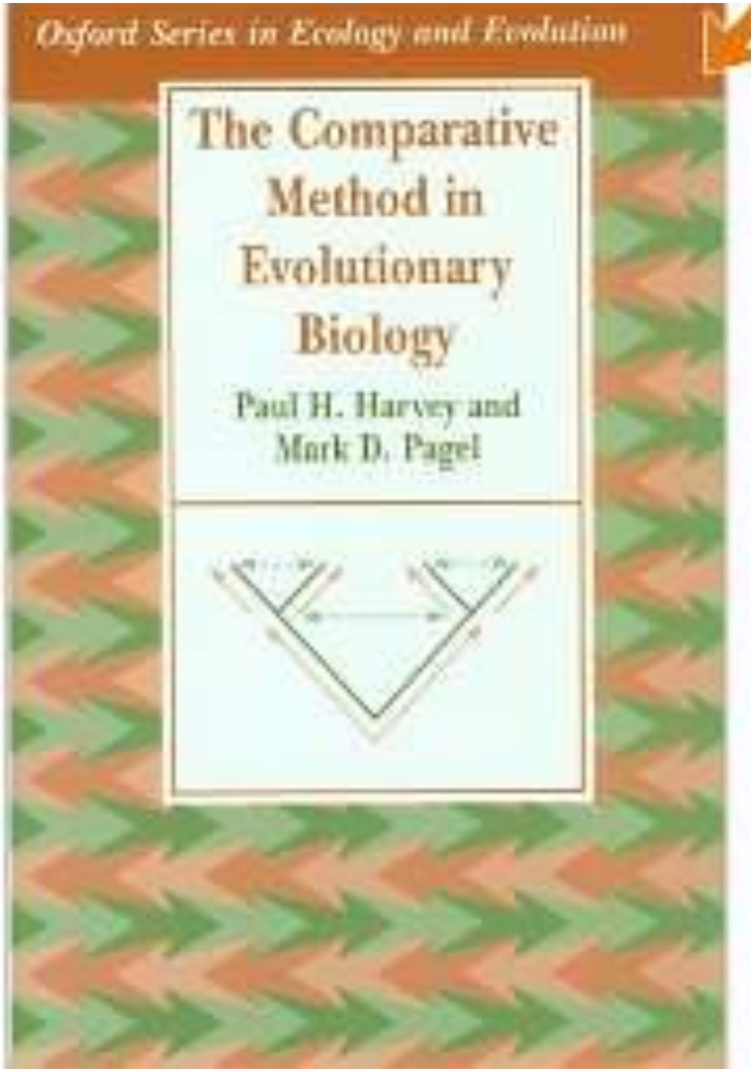
PHYLOGENIES AND THE COMPARATIVE METHOD

JOSEPH FELSENSTEIN

Department of Genetics SK-50, University of Washington, Seattle, Washington 98195

Submitted November 30, 1983; Accepted May 23, 1984

A origem dos métodos filogenéticos comparativos...



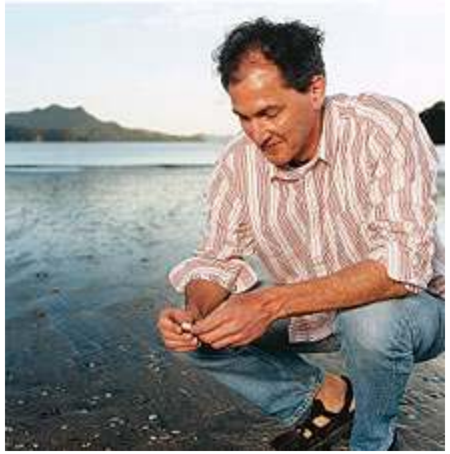
Oxford Series in Ecology and Evolution

The Comparative Method in Evolutionary Biology

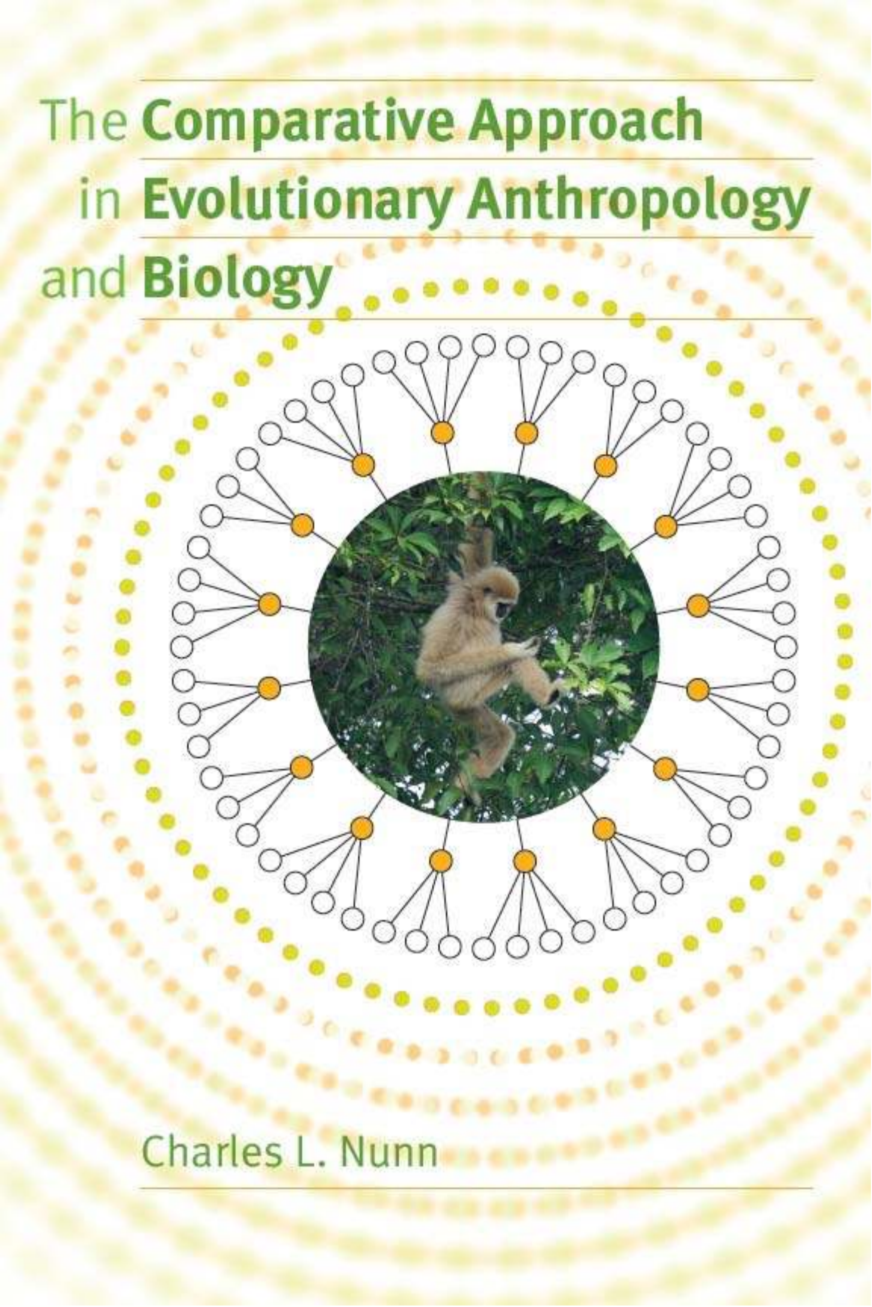
Paul H. Harvey and Mark D. Pagel

1991

Paul Harvey

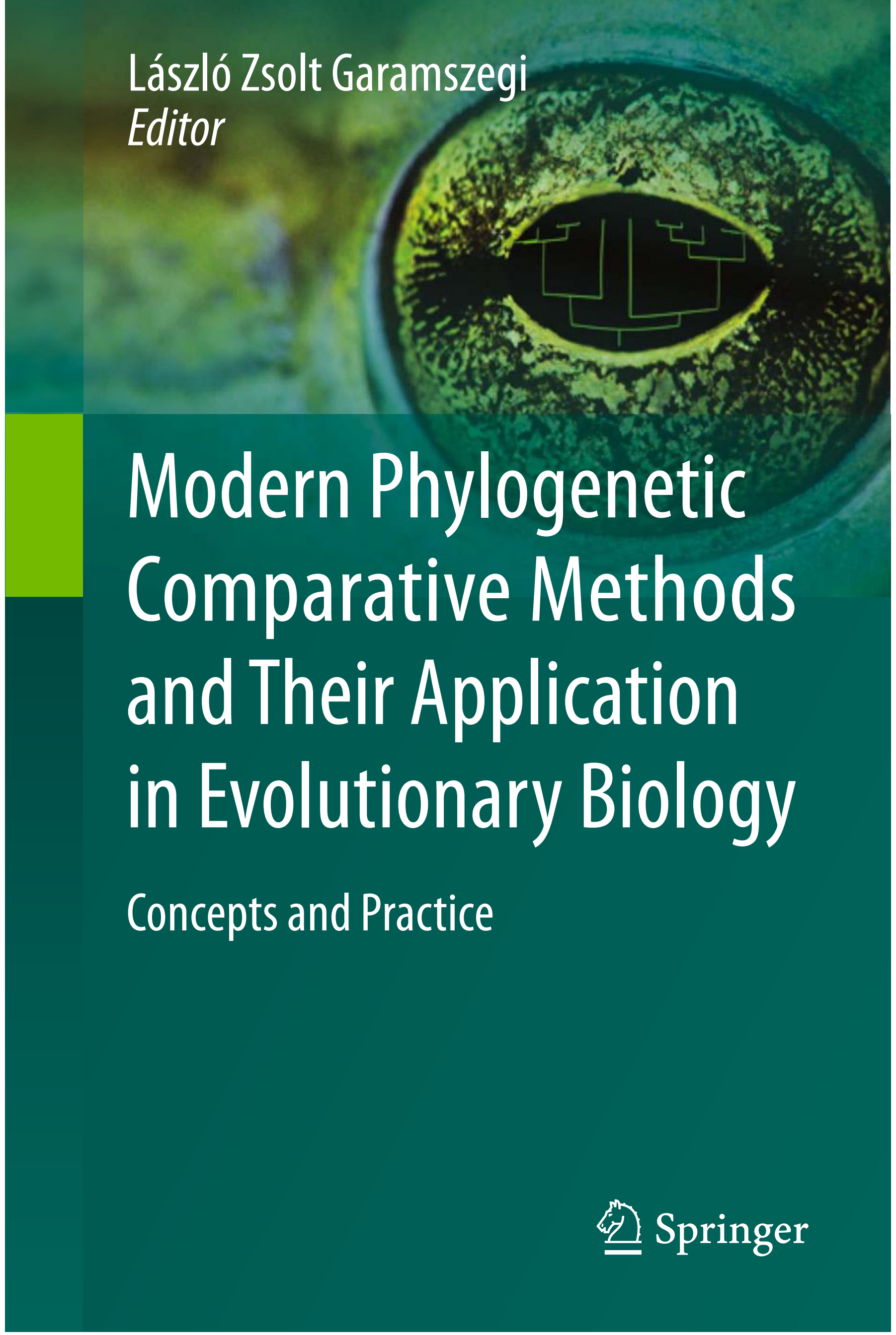


Mark Pagel



2011

Charles Nunn, Duke University



László Zsolt Garamszegi
Editor

Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology

Concepts and Practice

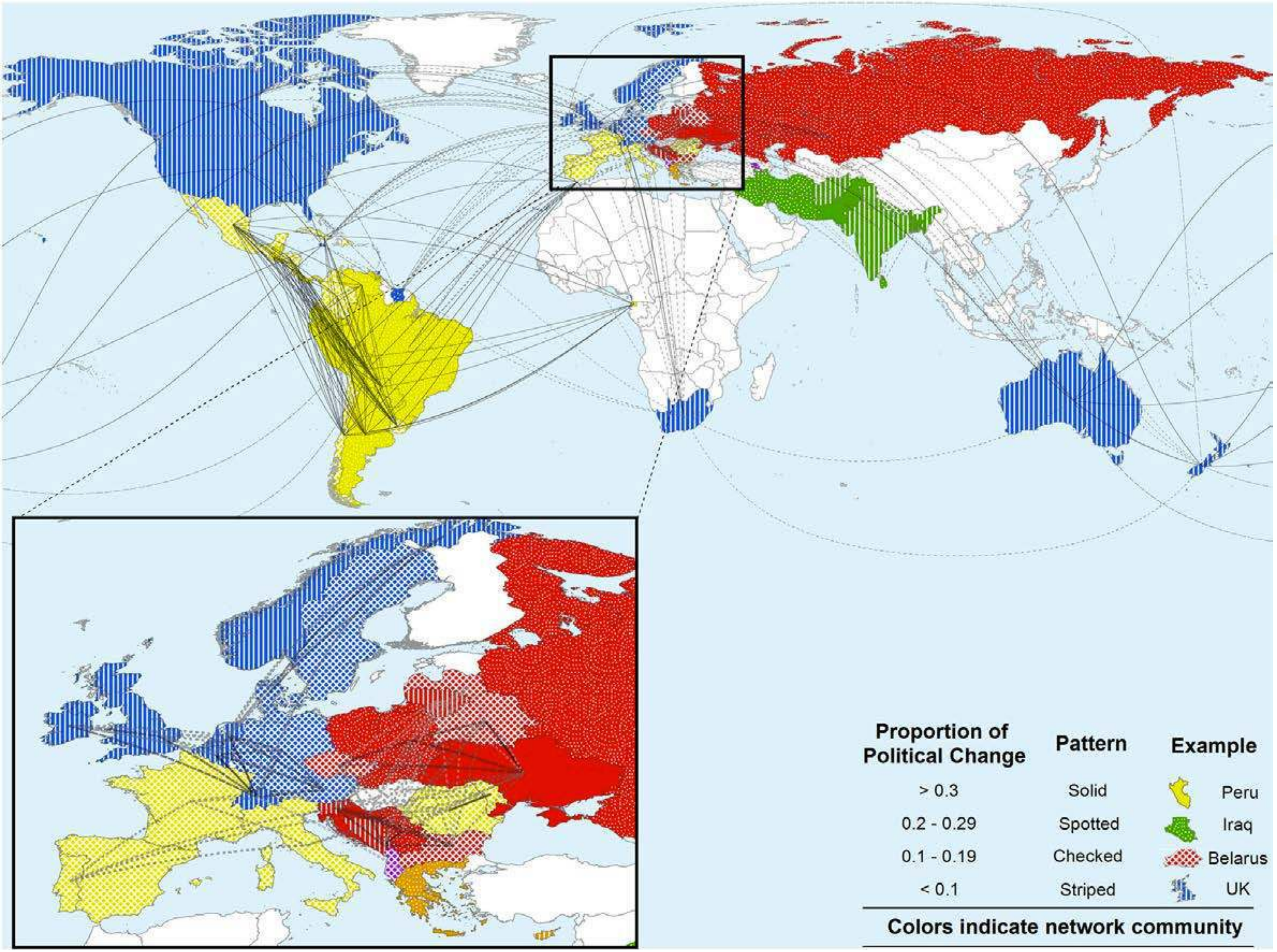
Springer

2015

Sevilha, Espanha, Novembro 2014

http://www.mpcm-evolution.org/practice

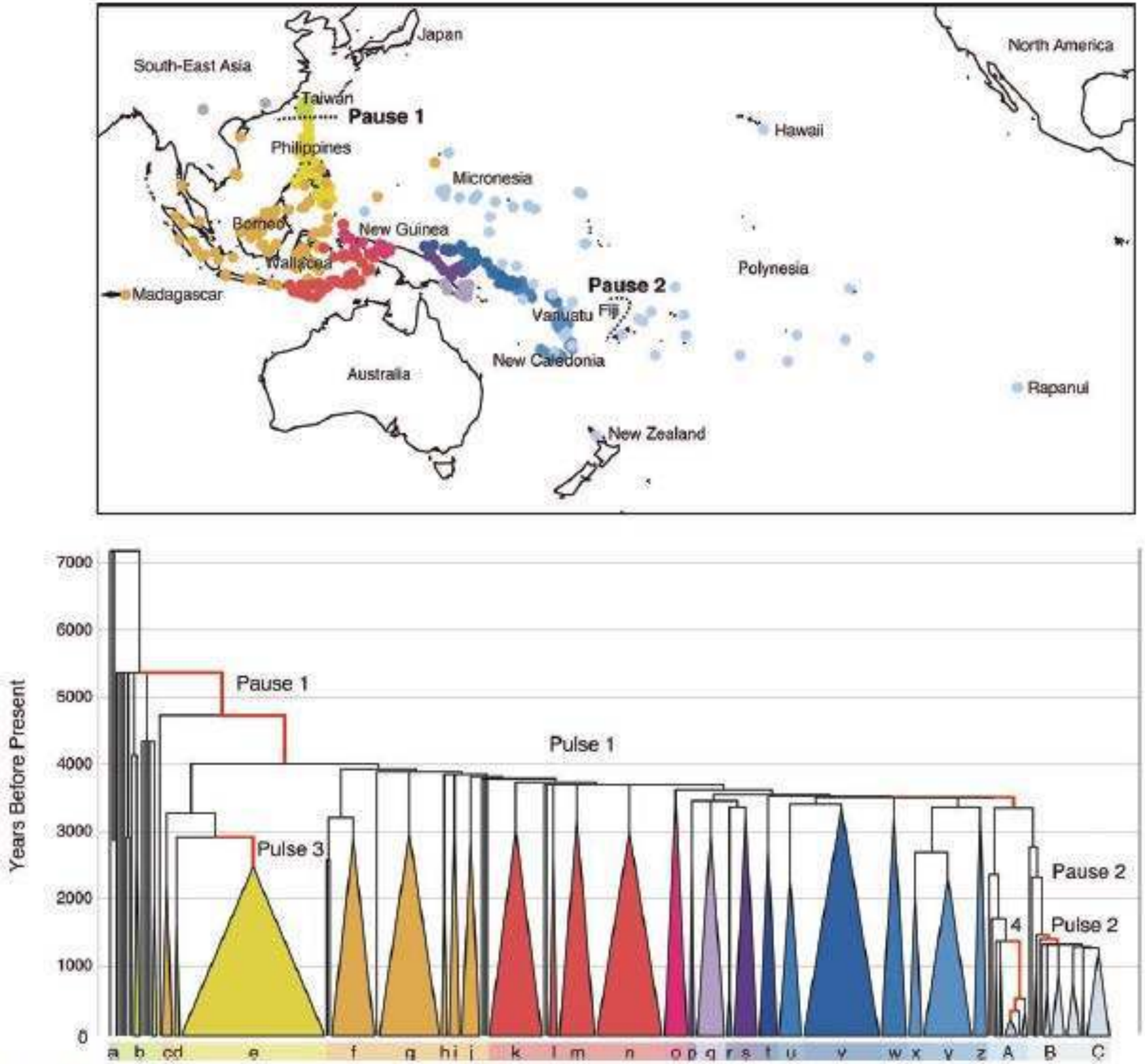
PLOS | ONE

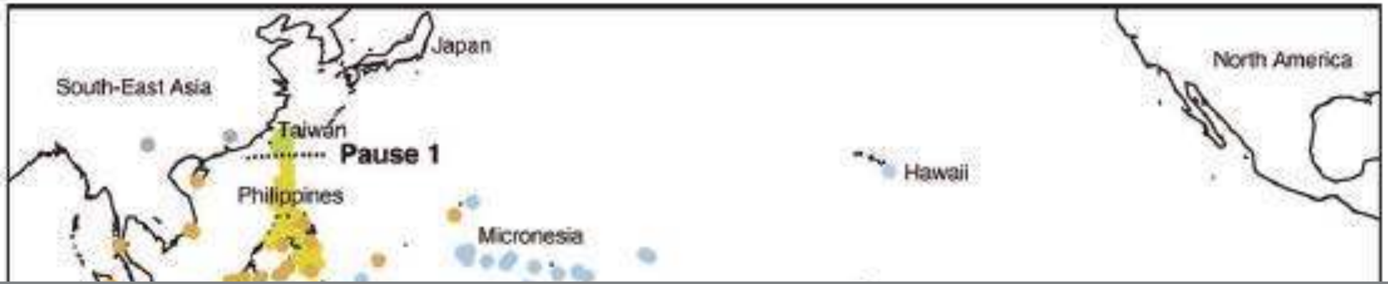


RESEARCH ARTICLE

Shared Cultural History as a Predictor of Political and Economic Changes among Nation States

Luke J. Matthews[1,3]*, Sam Passmore[2], Paul M. Richard[3], Russell D. Gray[2,4], Quentin D. Atkinson[2,4]


Japan
South-East Asia
North America
Taiwan
Pause 1
Philippines
Hawaii
Micronesia
Borneo
New Guinea
Wallacea
Polynesia
Madagascar
Pause 2
Vanuatu
Fiji
New Caledonia
Australia
Rapanui
New Zealand
Years Before Present
7000
6000
5000
4000
3000
2000
1000
0
Pause 1
Pulse 1
Pulse 3
Pause 2
4
Pulse 2
a b cd e f g h i j k l m n o p q r s t u v w x y z A B C



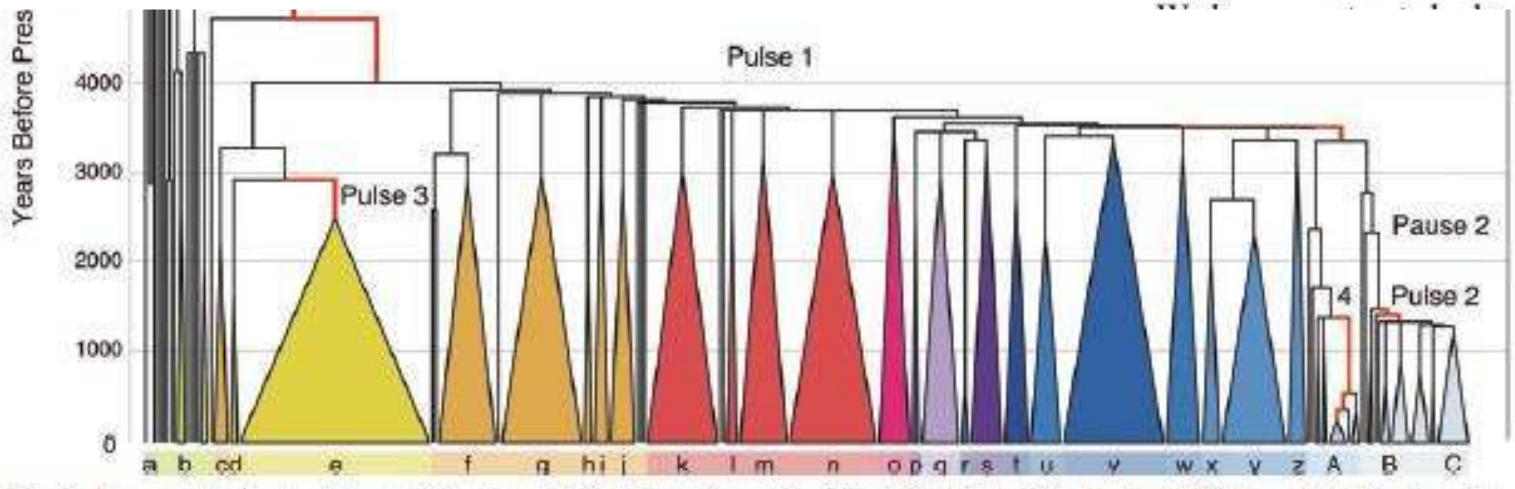
RESEARCH ARTICLE

Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement

R. D. Gray,[1] A. J. Drummond,[2] S. J. Greenhill[1]

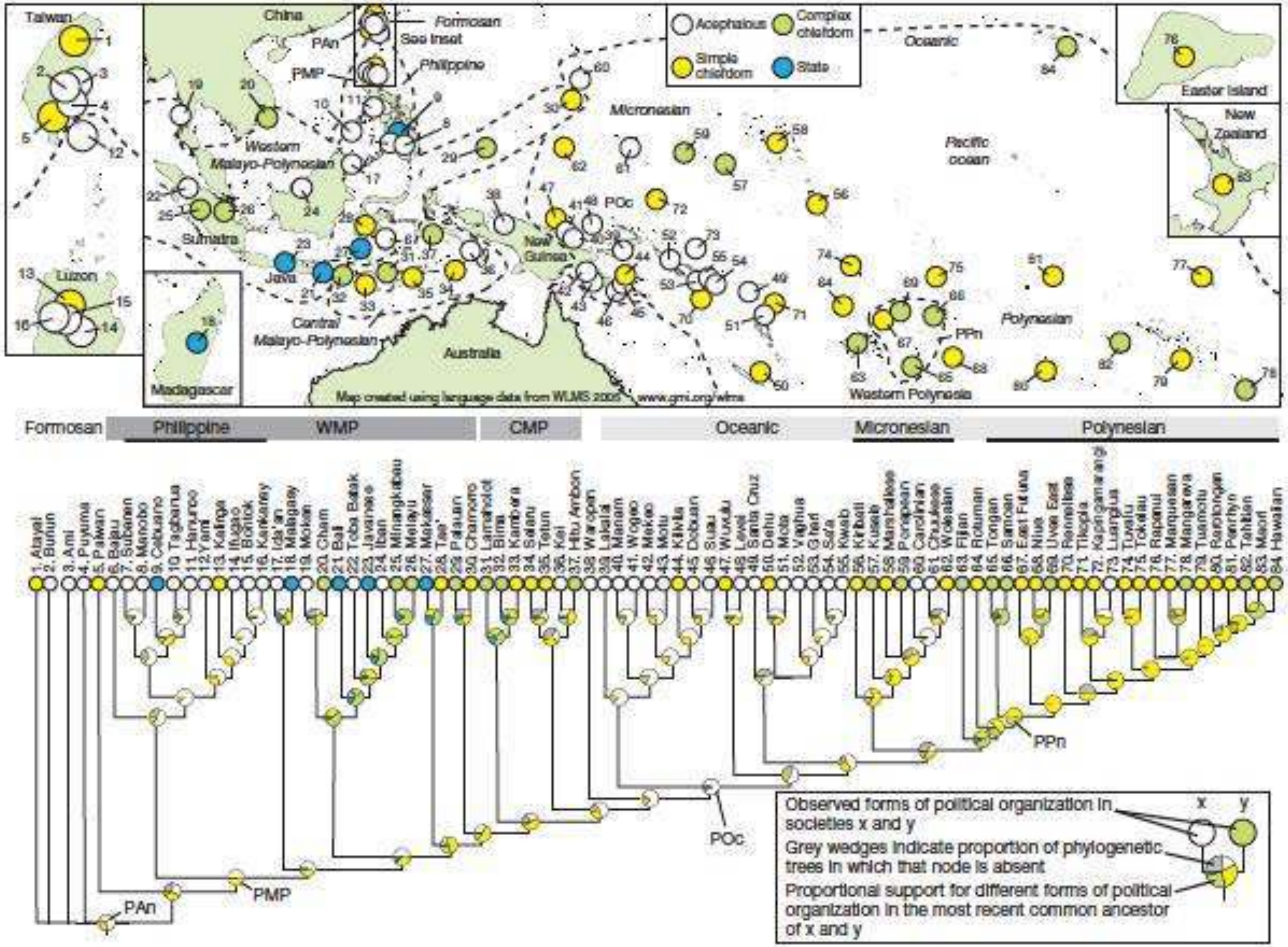
evolves at a rapid rate (*24*). Gray and Jordan's (*25*) previous parsimony analysis of Austronesian lexical data found support for the expansion sequence predicted by the pulse-pause scenario but limitations of the data and methods used meant that the predictions about the timing of Pacific settlement could not be tested.

Lexical data. The Austronesian language family is the one of the largest in the world, with around 1200 languages spread from Taiwan to New Zealand and Madagascar to Easter Island.

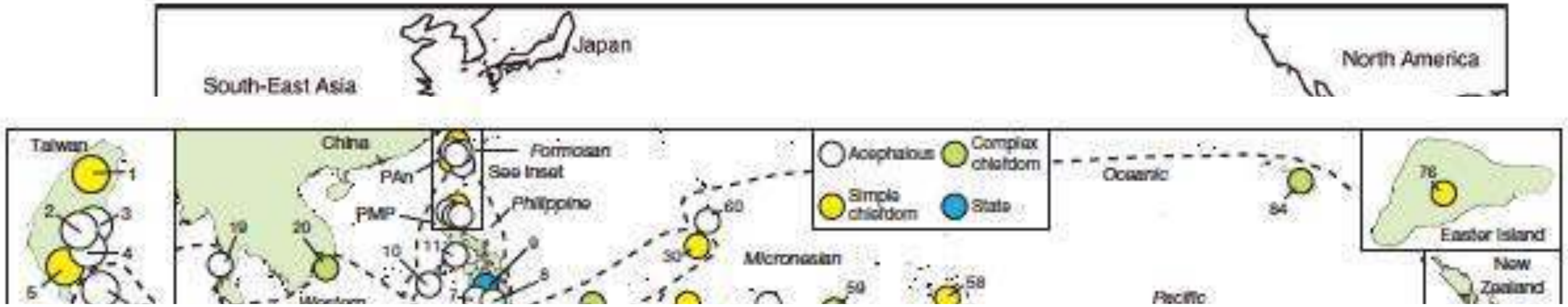


Langu
Expar
Pacifi

R. D. Gray,[1] A



ray and Jordan's
ysis of Austrone-
for the expansion
se-pause scenario
d methods used
ut the timing of
tested.
nesian language
in the world, with
from Taiwan to
to Easter Island.



ICLE

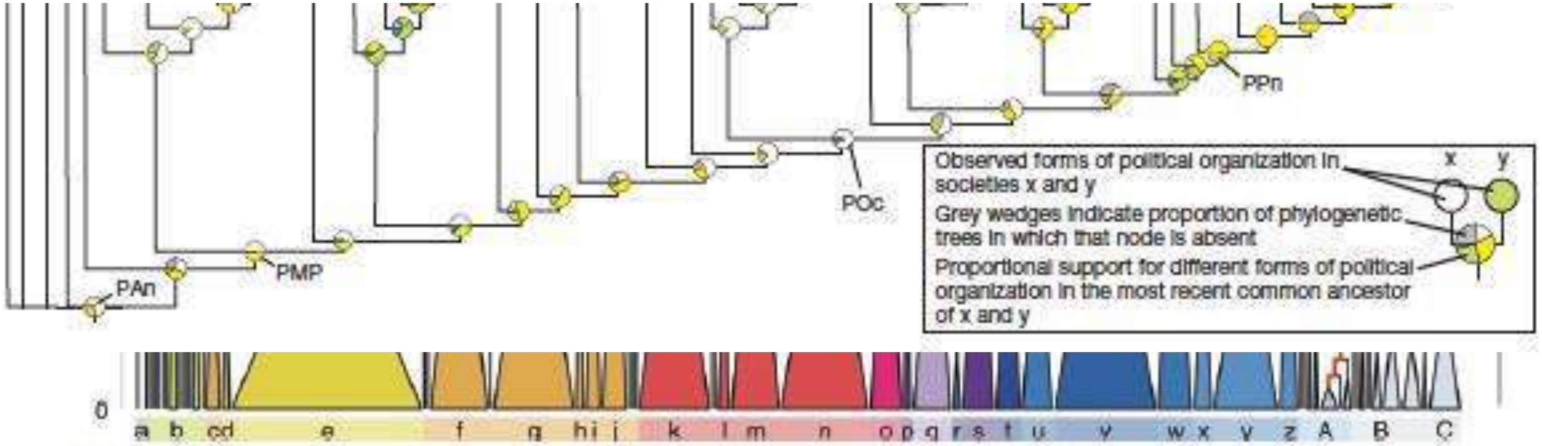
ARTICLE

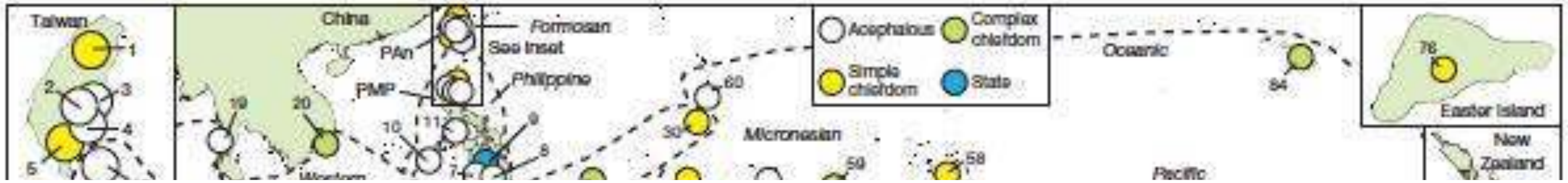
doi:10.1038/nature09461

Rise and fall of political complexity in island South-East Asia and the Pacific

Thomas E. Currie[1,2], Simon J. Greenhill[3,4], Russell D. Gray[3], Toshikazu Hasegawa[1] & Ruth Mace[2]

ay and Jordan's
sis of Austrone-
or the expansion
e-pause scenario
l methods used
it the timing of
ested.
iesian language
i the world, with
from Taiwan to
to Easter Island.





ARTICLE

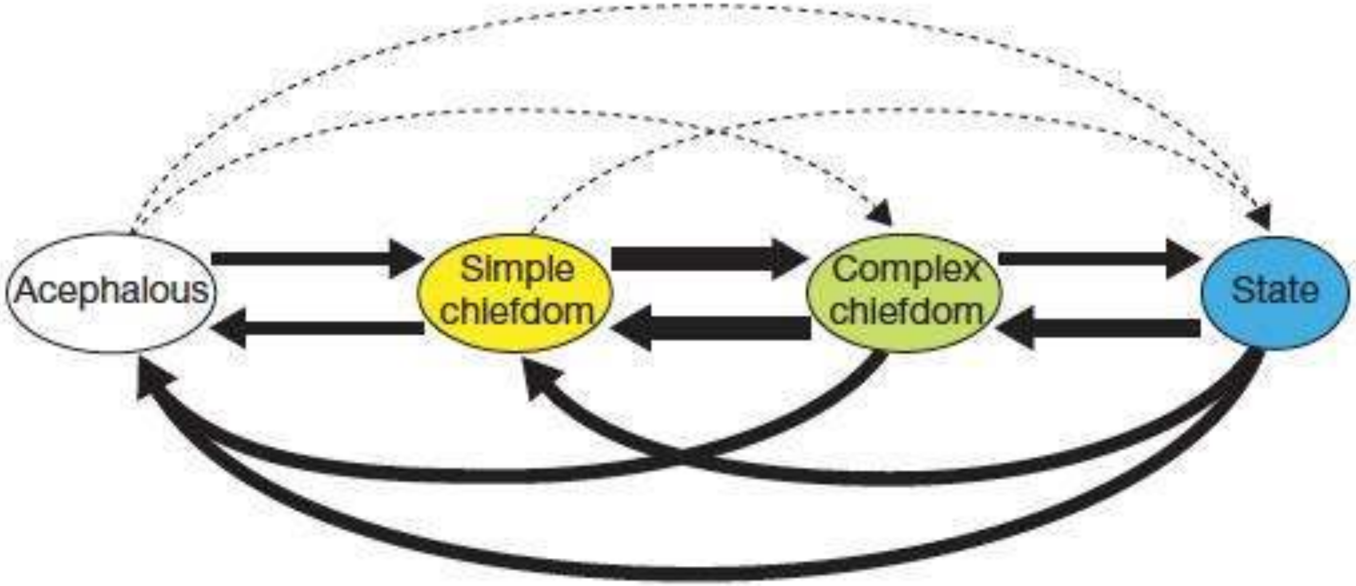
TICLE

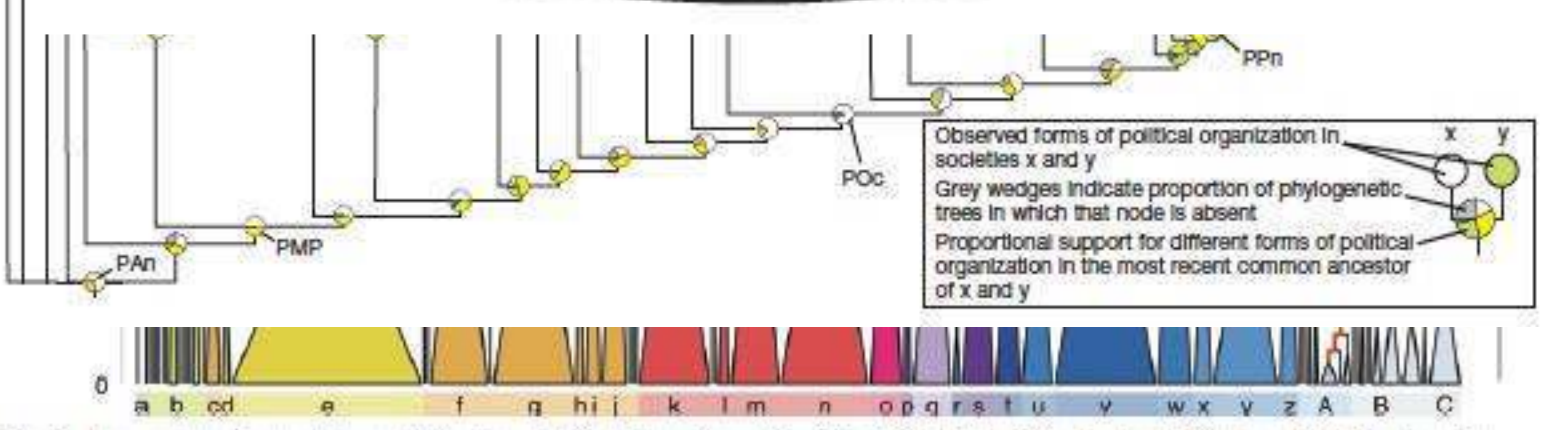
Rise and fa
island Sout

Thomas E. Currie[1,2], Simon J. Gr

e09461

ay and Jordan's
sis of Austrone-
or the expansion
-pause scenario
l methods used
it the timing of
ested.
esian language
the world, with
from Taiwan to
to Easter Island.





Observed forms of political organization in societies x and y

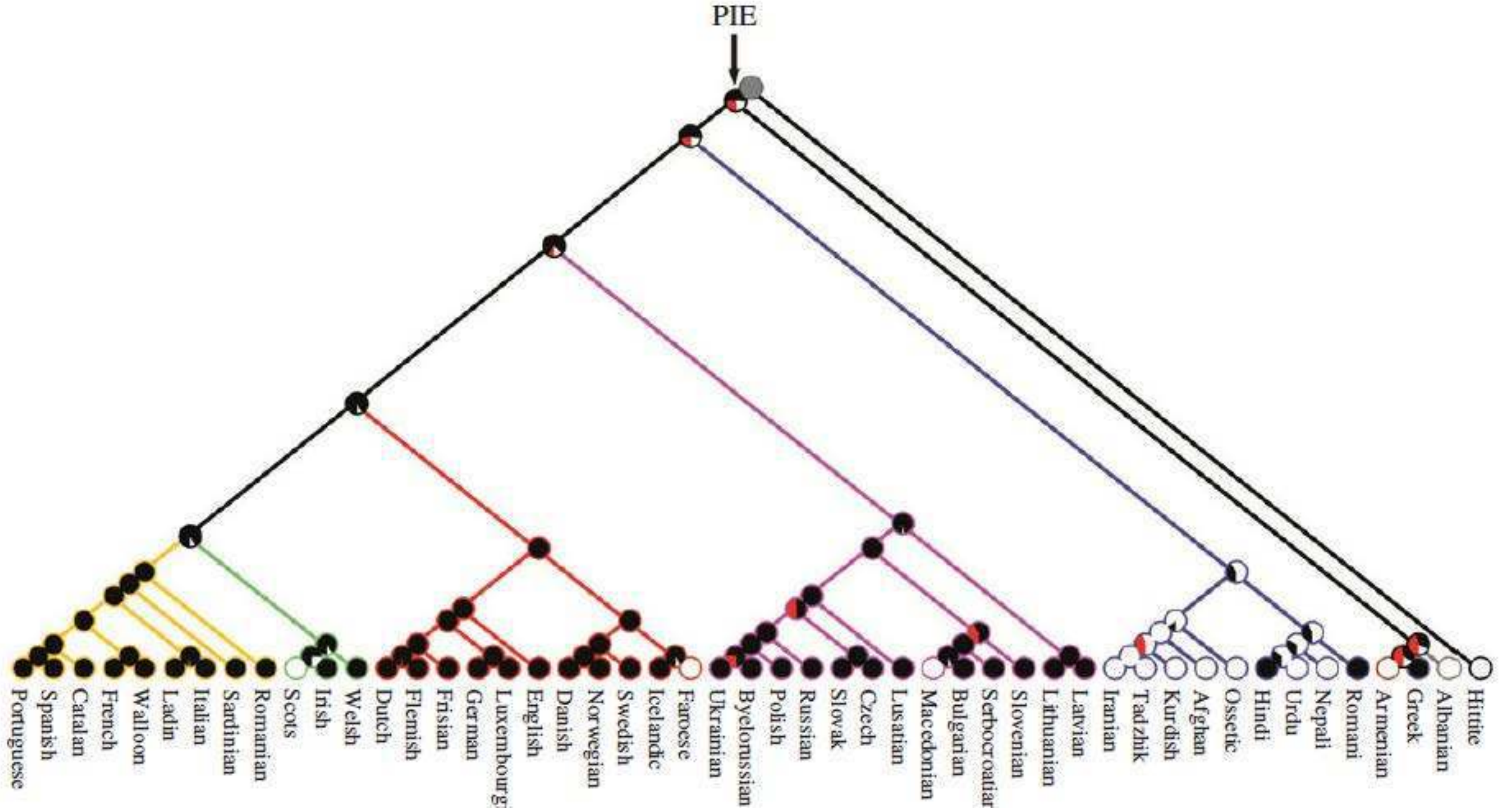
Grey wedges indicate proportion of phylogenetic trees in which that node is absent

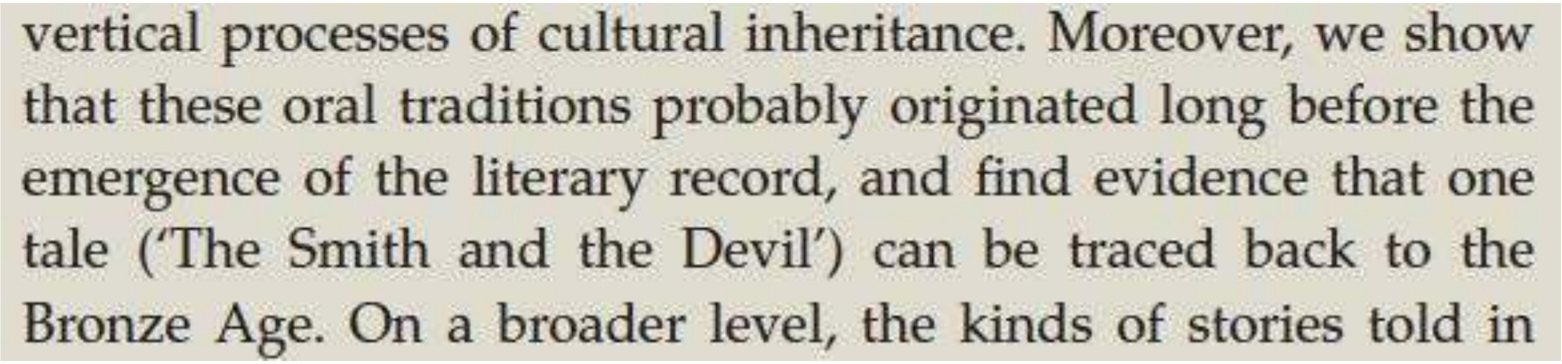
Proportional support for different forms of political organization in the most recent common ancestor of x and y

ROYAL SOCIETY
OPEN SCIENCE

Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales

Sara Graça da Silva[1] and Jamshid J. Tehrani[2]

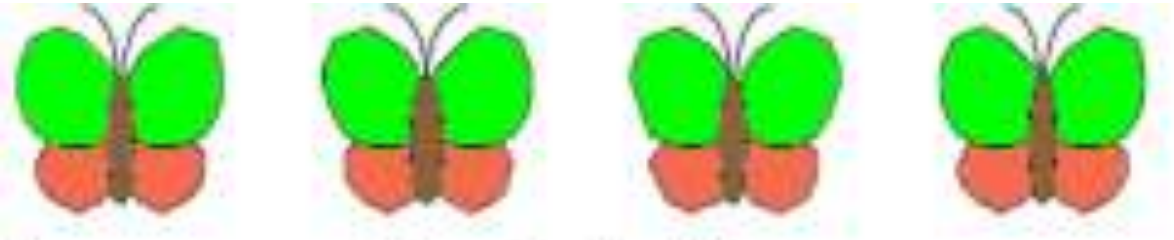




vertical processes of cultural inheritance. Moreover, we show that these oral traditions probably originated long before the emergence of the literary record, and find evidence that one tale ('The Smith and the Devil') can be traced back to the Bronze Age. On a broader level, the kinds of stories told in

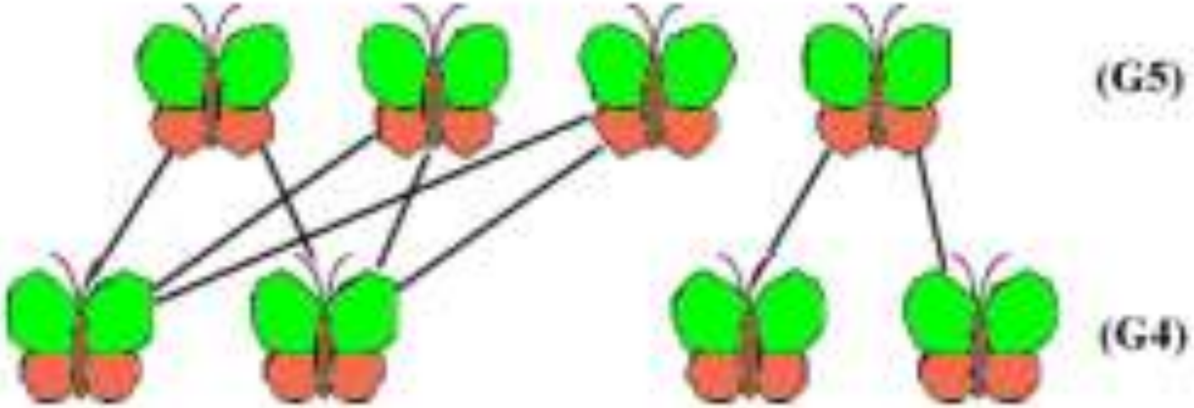
O que é uma filogenia? O que ela representa?

Escala de Parentesco



A Four butterflies

Indivíduos isolados

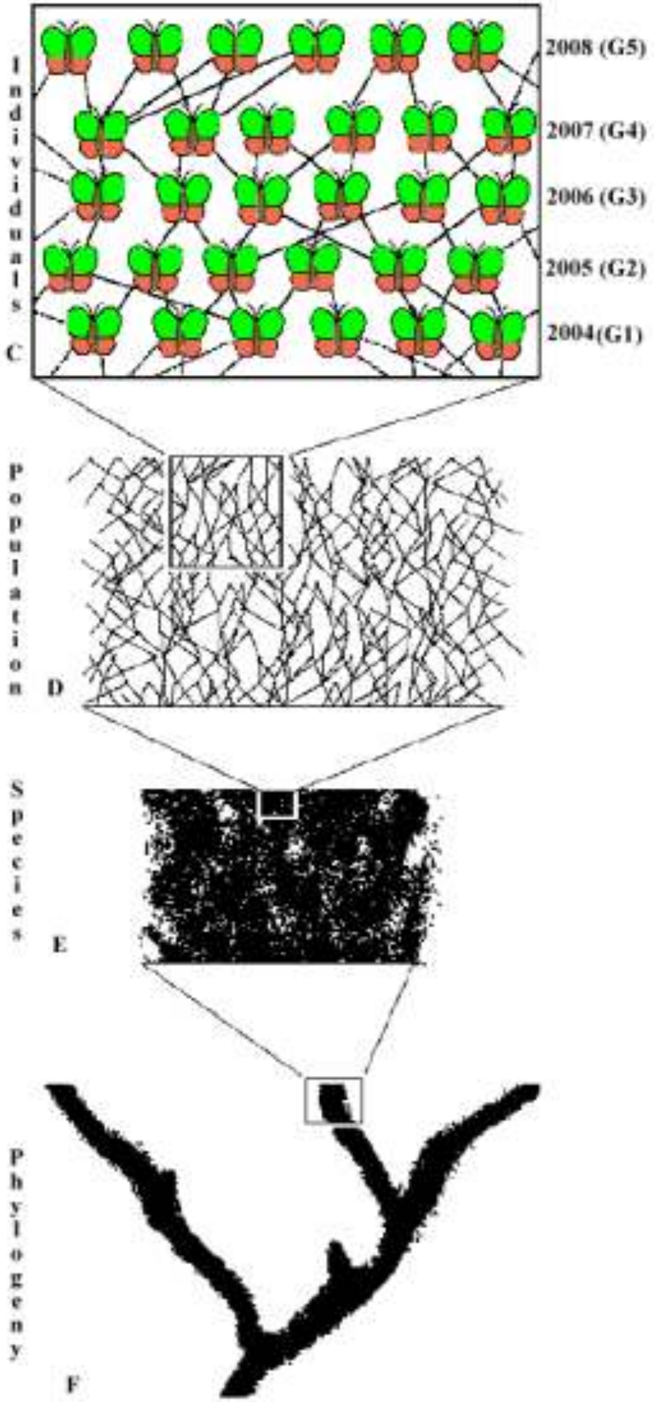


B Four butterflies (top row), connected to their parents (bottom row)

Família

http://www.nature.com/scitable/topicpage/reading-a-phylogenetic-tree-the-meaning-of-41956#

Escala de Parentesco



Famílias

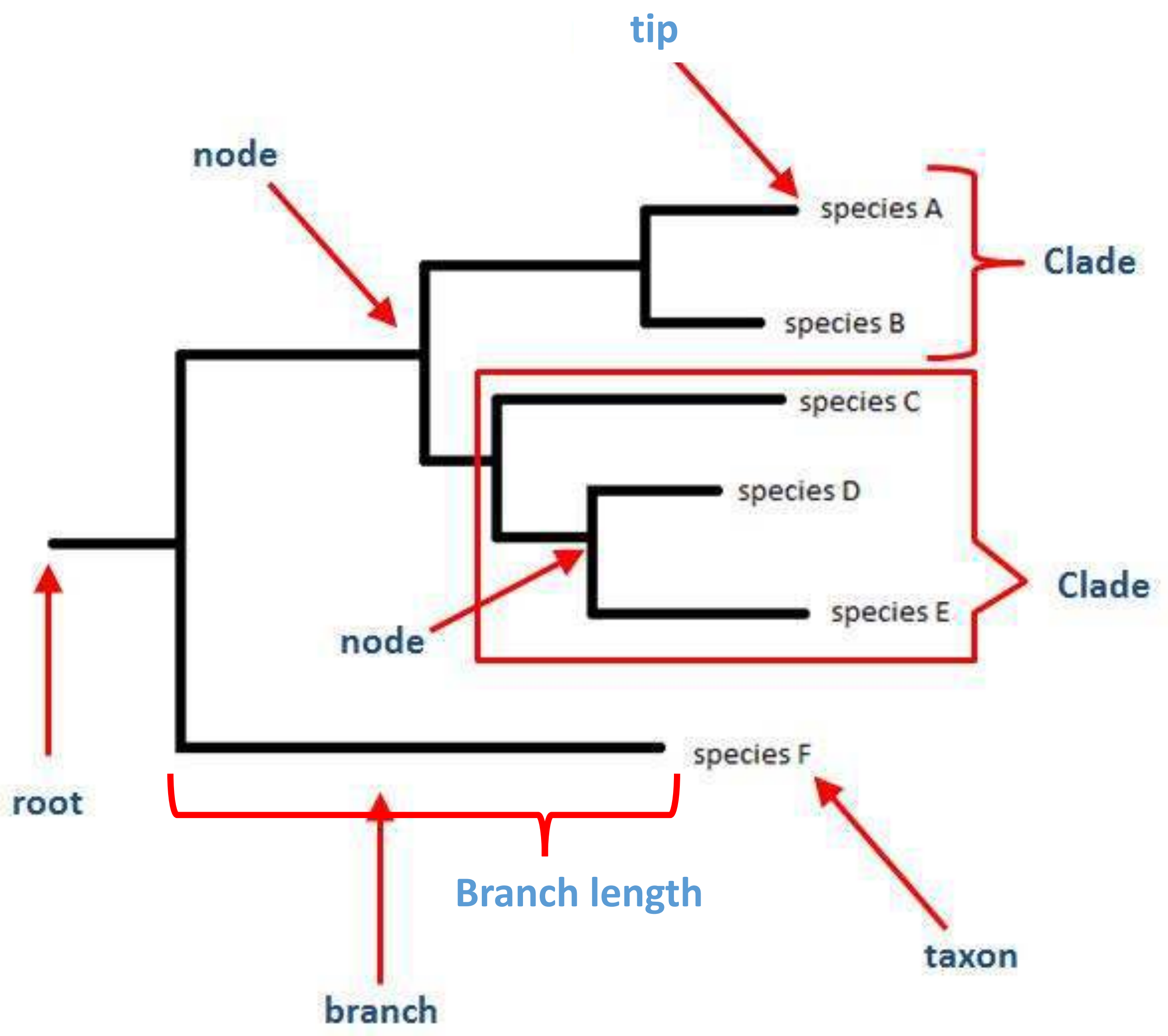
Populações

Espécies

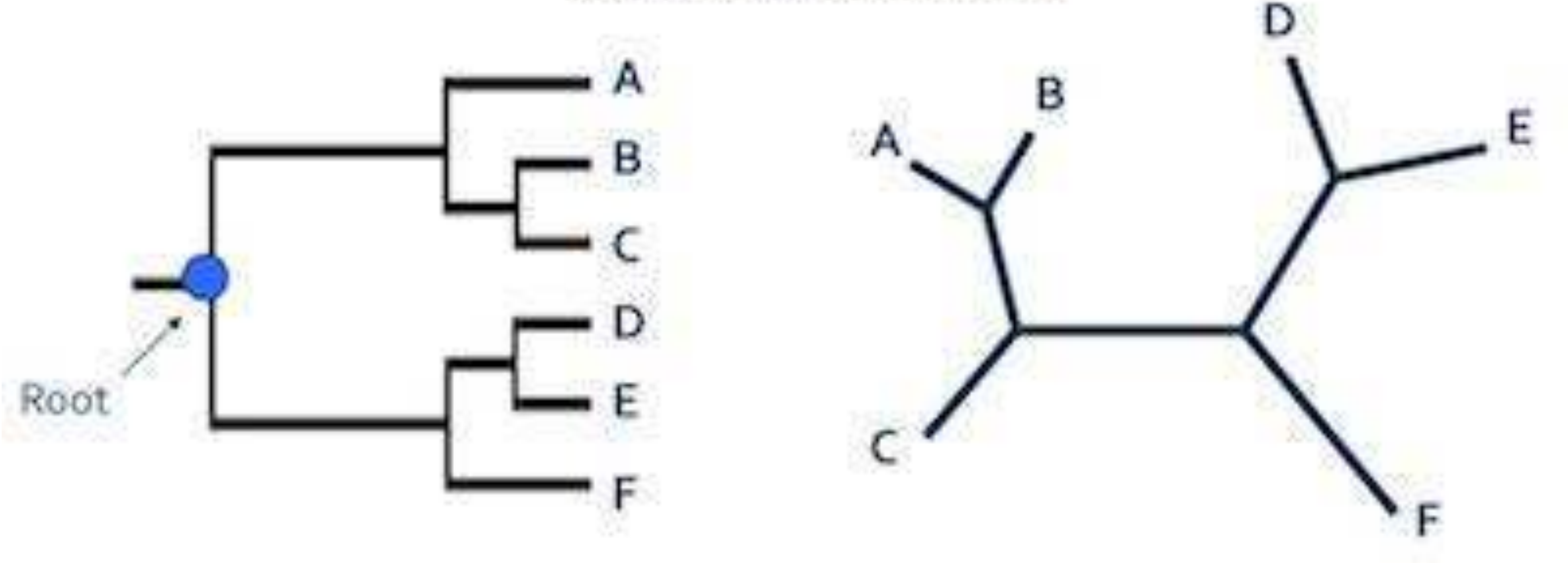
Filogenia (entre espécies)

Conceitos importantes

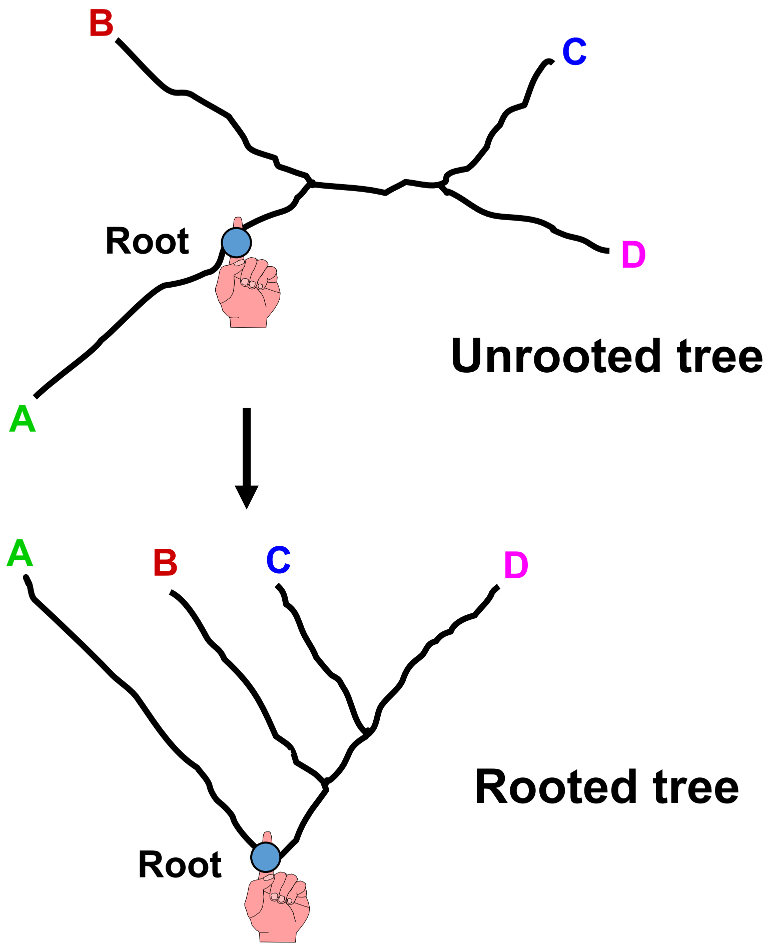
Parts of a phylogenetic tree



Rooted vs. unrooted trees

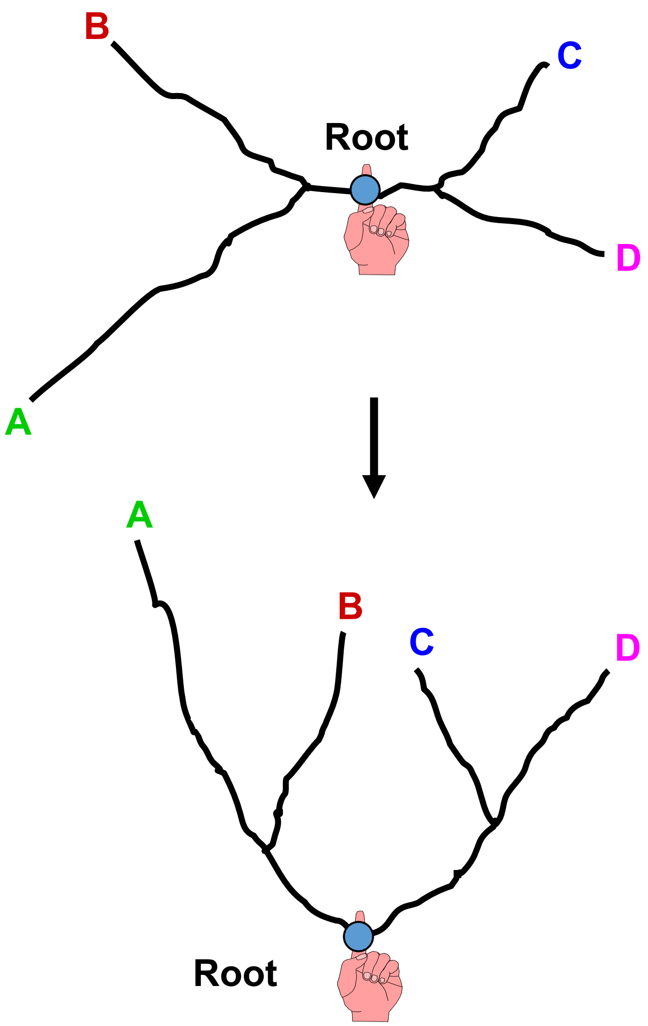


Enraizando a árvore



Taxon A não é mais aparentado com o B que C ou D.

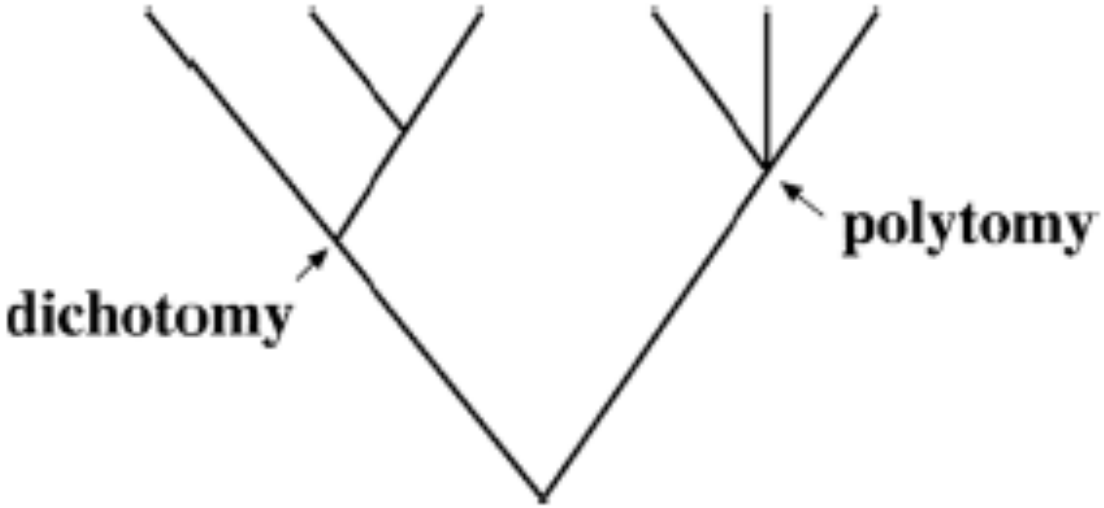
Enraizando em outra posição



Unrooted tree

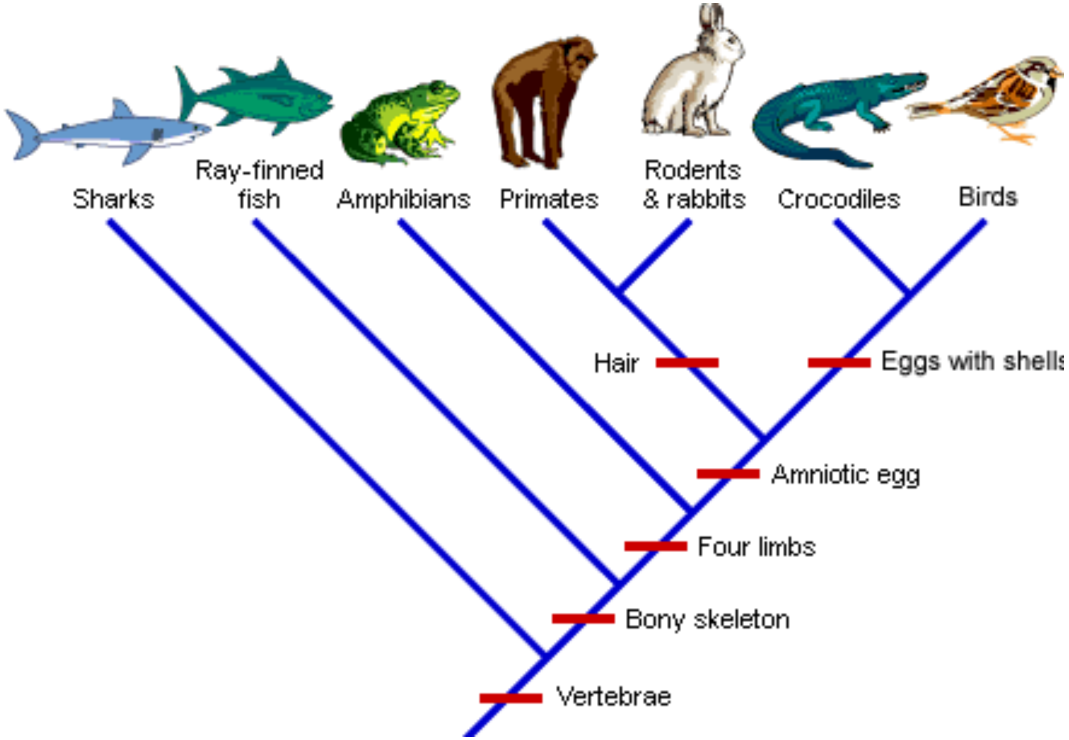
Rooted tree

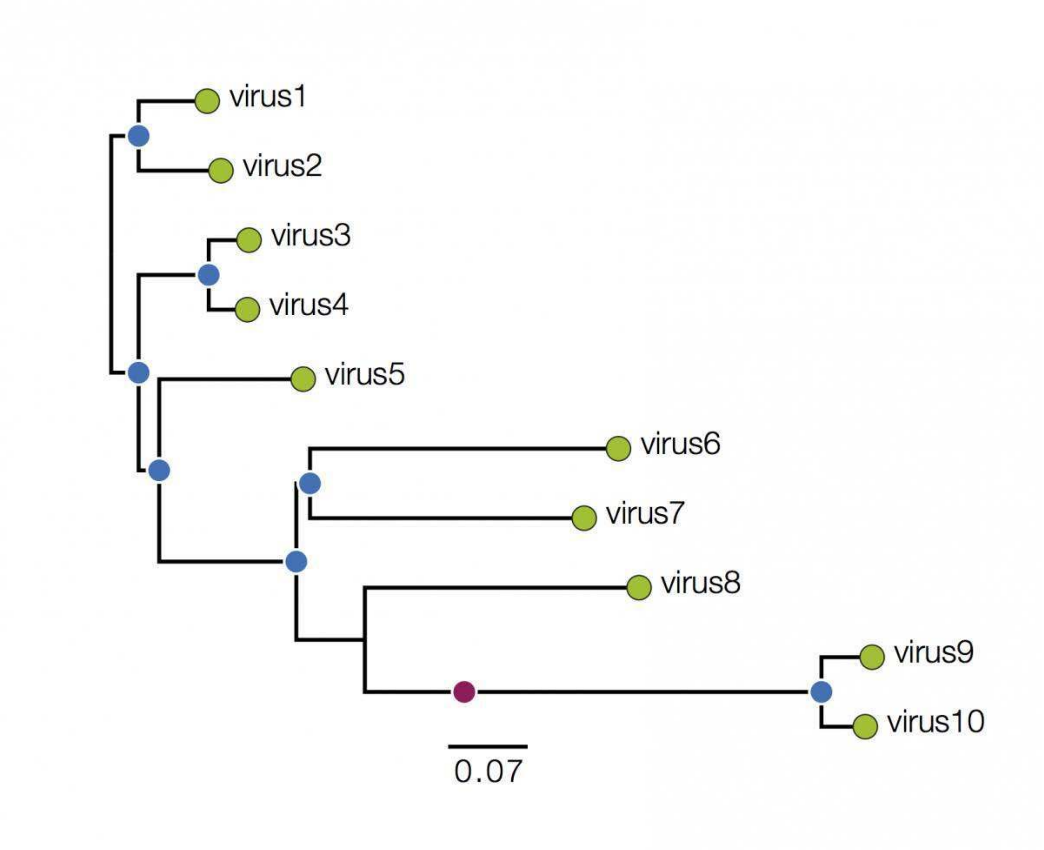
Agora o taxon A é mais próximo de B, e ambos são grupo irmão de C e D.



- **Politomia** – quando um nó (ancestral comum) dá origem a mais de 2 descendentes
    - Soft-polytomy: quando os dados disponíveis não conseguem discriminar a ordem de divergência, devido à rápida divergência
    - Hard-polytomy: quando de fato a politomia é real, usualmente representando uma ”explosão” radiativa de espécies

Tipos de representação de árvores





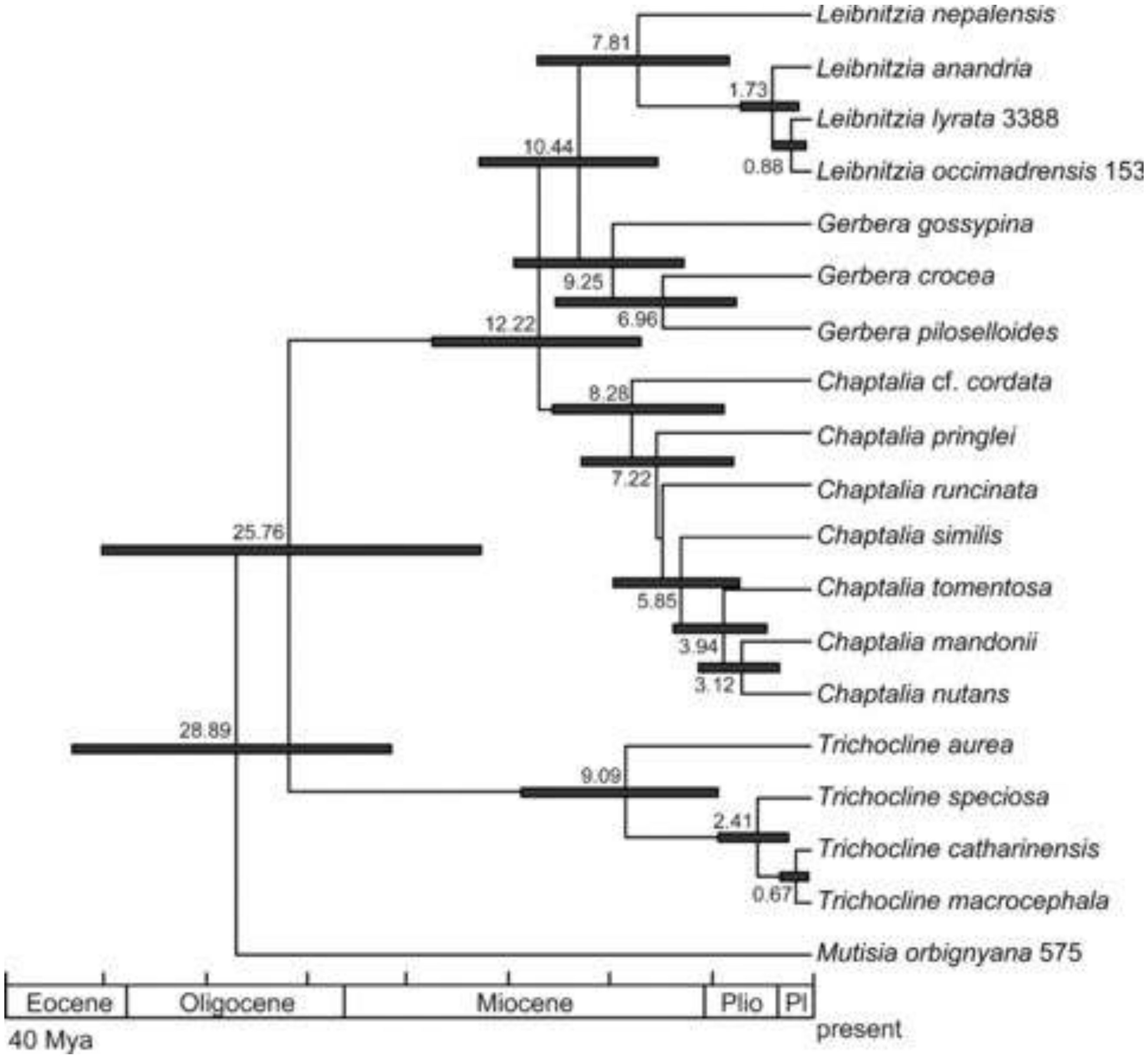
Cladograma

Enfatiza relações entre tips e opcionalmente o compartilhamento de caracteres derivados (sinapomorfias). Comprimentos de ramo não importam

Filograma

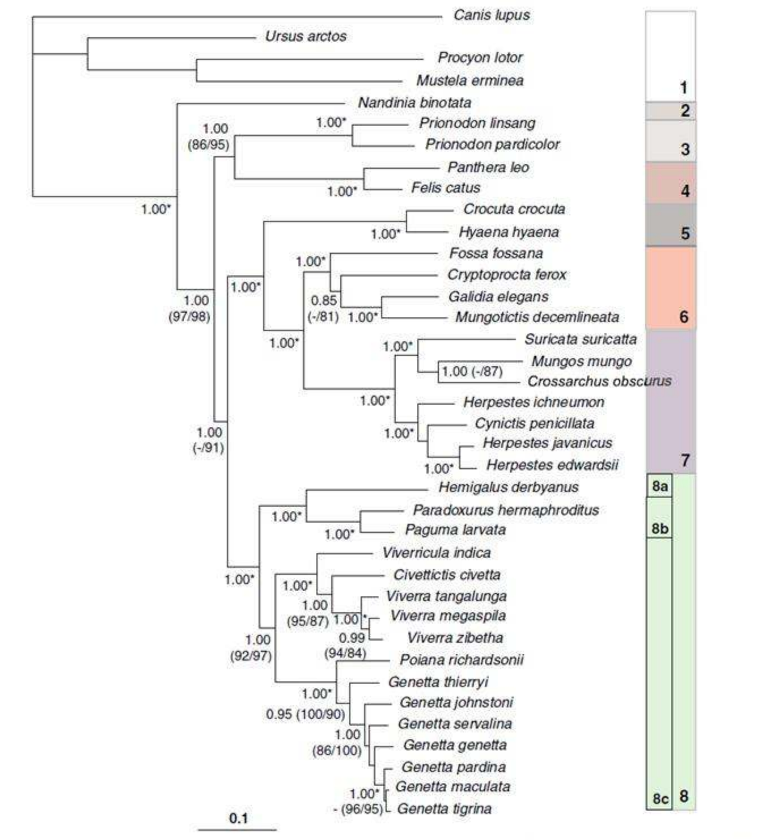
Comprimento de ramo refletem divergência entre espécies

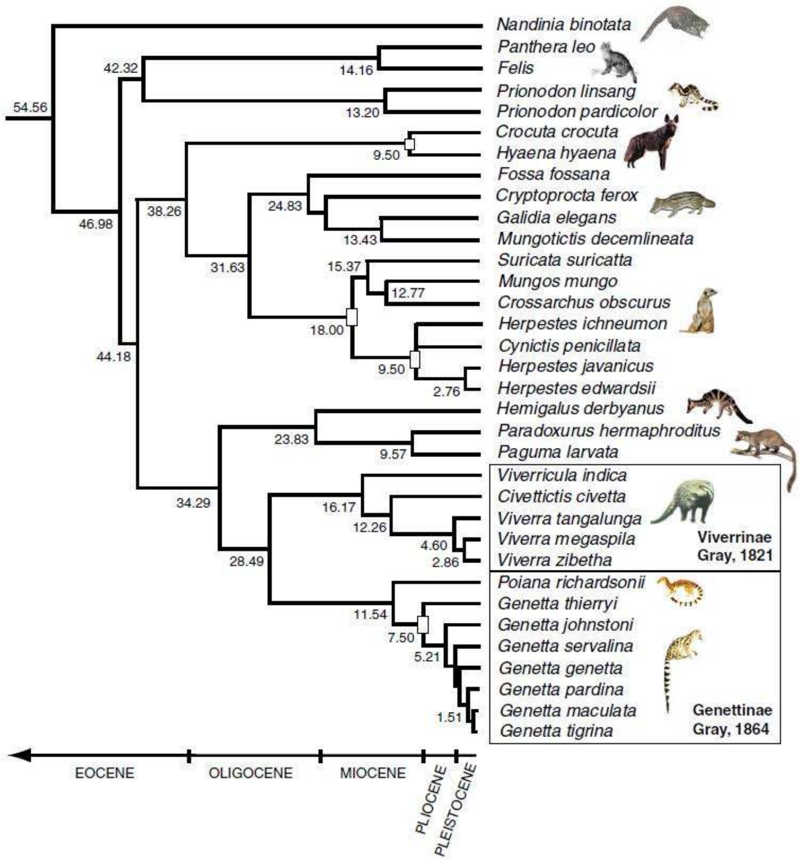
Tipos de representação de árvores



Cronograma

Comprimentos de ramo representam tempo de divergência em Milhões de anos



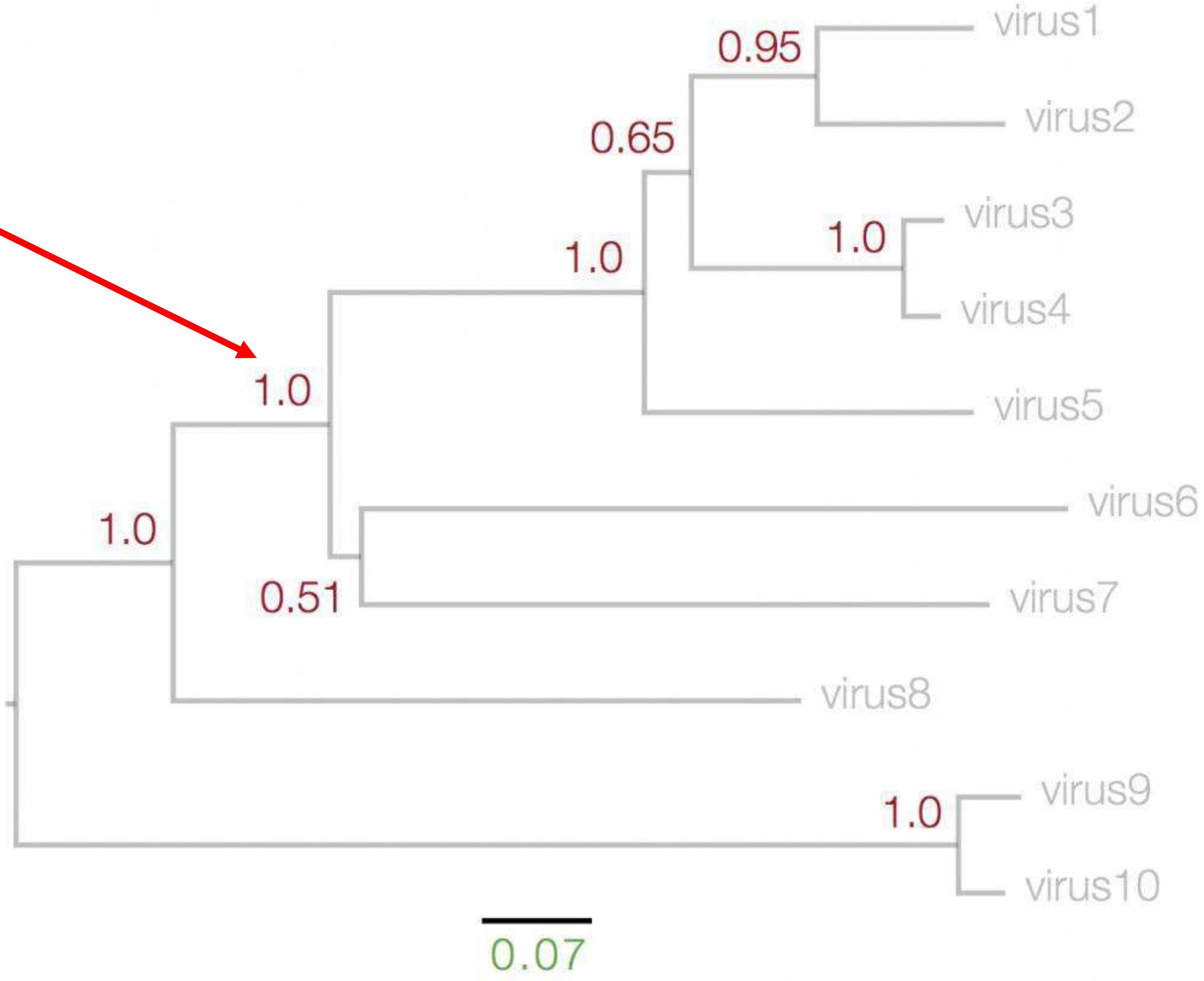


Filograma não ultramétrica

Árvore "ultramétrica"

Transformação de divergência (em termos de distâncias genéticas) em tempo utilizando um relógio molecular

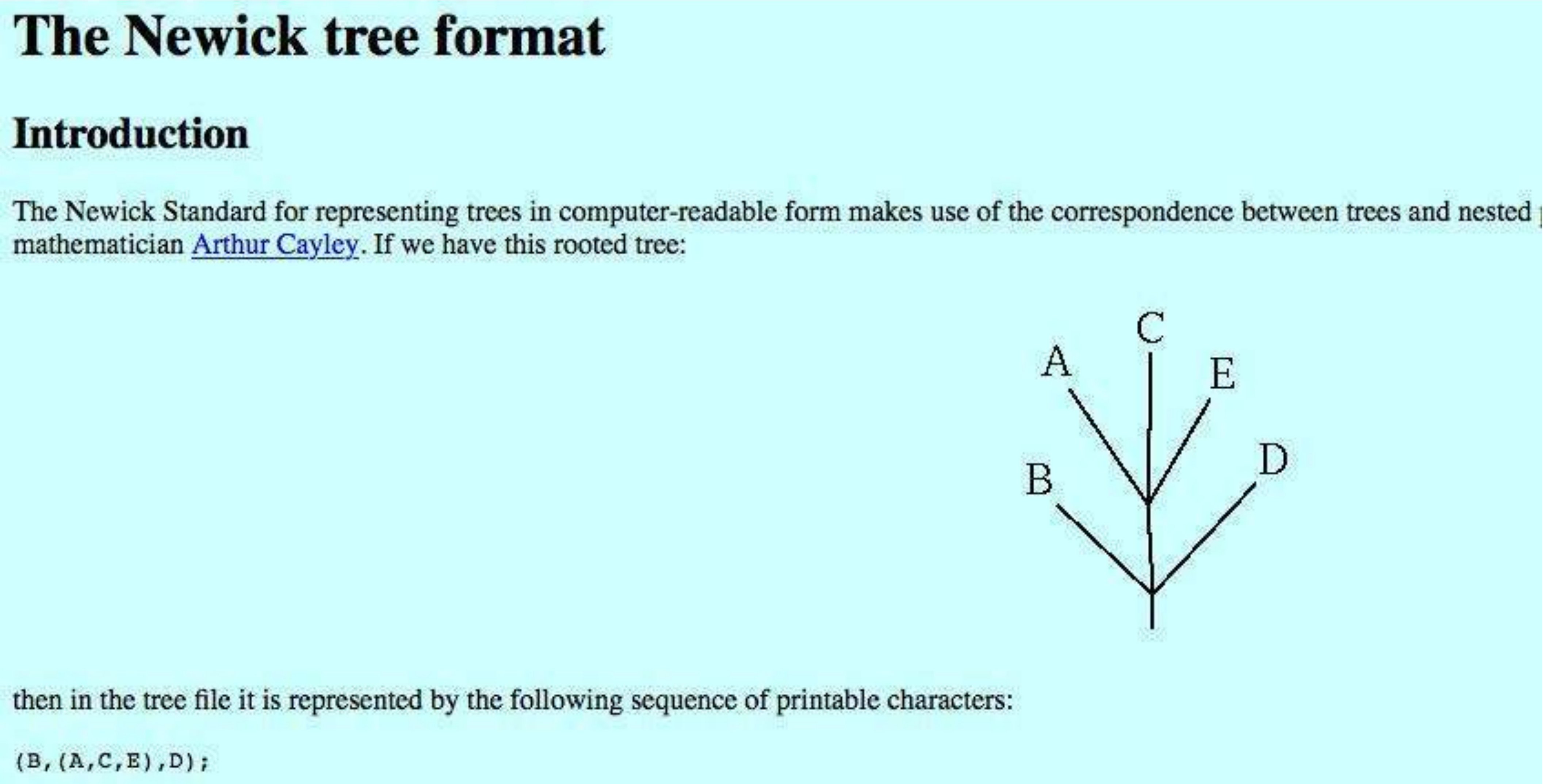
Medida de suporte do nó



Existem várias formas, cada uma mais adequada pra um critério de otimização. P. ex. Bootstrap (usado em qualquer critério), Bremer (usada em MP), Probabilidade Posterior (BI)

Demonstram de maneira geral o grau de suporte para uma dada relação a partir do conjunto de dados usado pra reconstruir a filogenia

Formatos para descrever filogenias



The Newick tree format

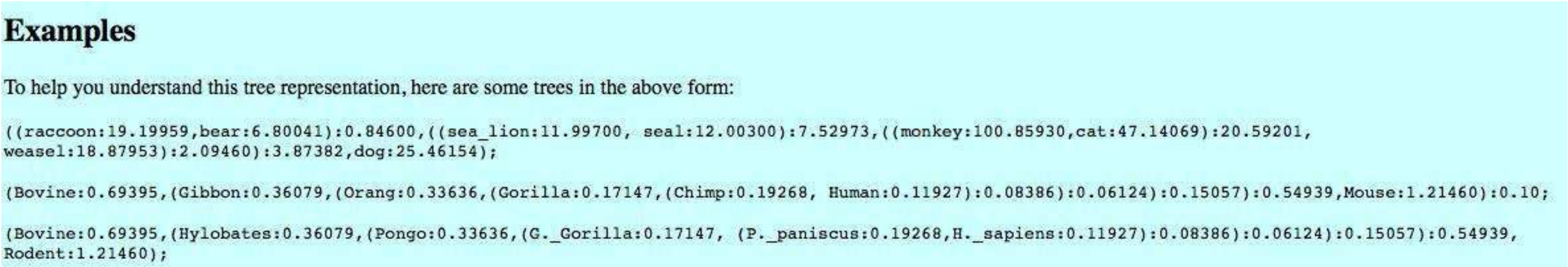
Introduction

The Newick Standard for representing trees in computer-readable form makes use of the correspondence between trees and nested
mathematician Arthur Cayley. If we have this rooted tree:

then in the tree file it is represented by the following sequence of printable characters:

```
(B,(A,C,E),D);
```


NEWICK'S
LOBSTER HOUSE



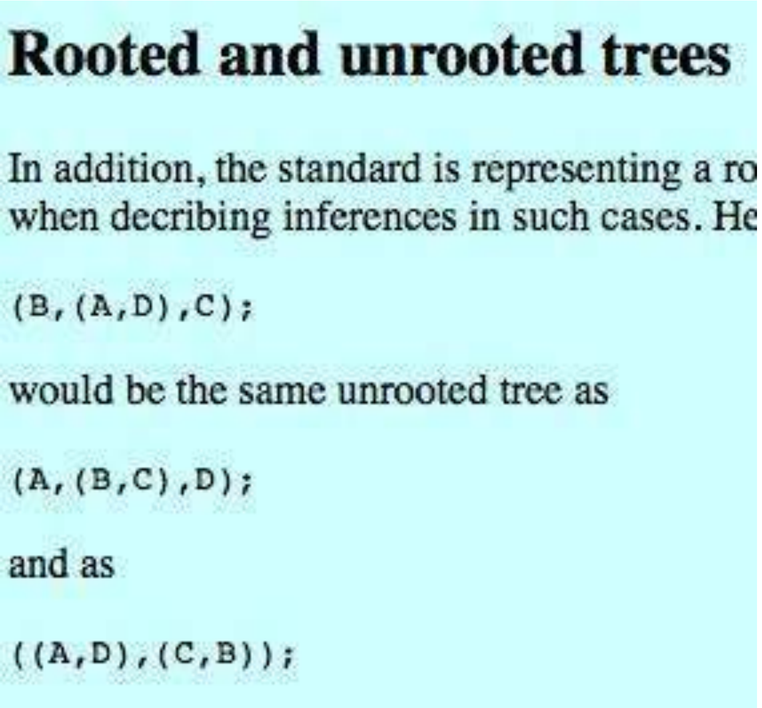
Examples

To help you understand this tree representation, here are some trees in the above form:

```
((raccoon:19.19959,bear:6.80041):0.84600,((sea_lion:11.99700, seal:12.00300):7.52973,((monkey:100.85930,cat:47.14069):20.59201,
weasel:18.87953):2.09460):3.87382,dog:25.46154);

(Bovine:0.69395,(Gibbon:0.36079,(Orang:0.33636,(Gorilla:0.17147,(Chimp:0.19268, Human:0.11927):0.08386):0.06124):0.15057):0.54939,Mouse:1.21460):0.10;

(Bovine:0.69395,(Hylobates:0.36079,(Pongo:0.33636,(G._Gorilla:0.17147, (P._paniscus:0.19268,H._sapiens:0.11927):0.08386):0.06124):0.15057):0.54939,
Rodent:1.21460);
```



Rooted and unrooted trees

In addition, the standard is representing a ro
when decribing inferences in such cases. He

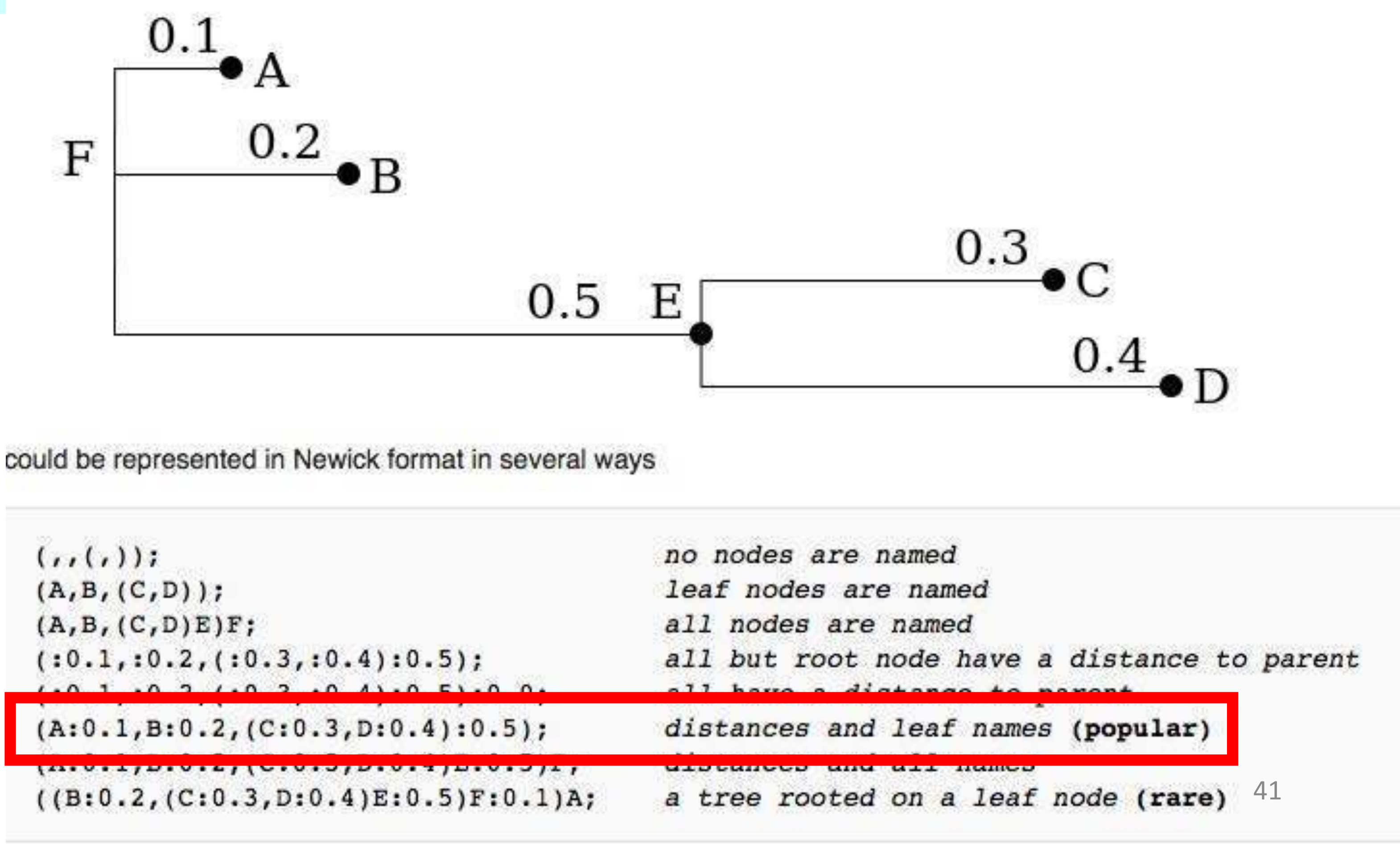
```
(B,(A,D),C);
```

would be the same unrooted tree as

```
(A,(B,C),D);
```

and as

```
((A,D),(C,B));
```

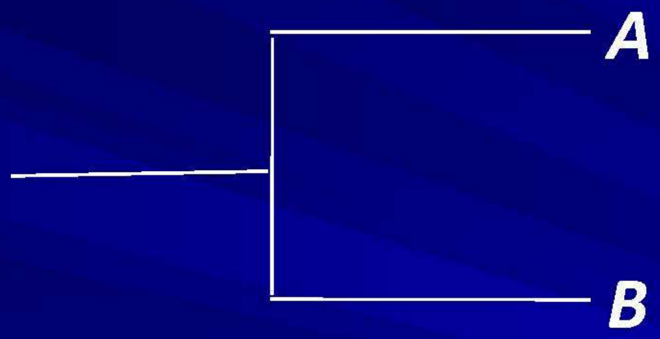


could be represented in Newick format in several ways

```
(,,(,));                              no nodes are named
(A,B,(C,D));                          leaf nodes are named
(A,B,(C,D)E)F;                        all nodes are named
(:0.1,:0.2,(:0.3,:0.4):0.5);          all but root node have a distance to parent
(:0.1,:0.2,(:0.3,:0.4):0.5):0.0;      all have a distance to parent
(A:0.1,B:0.2,(C:0.3,D:0.4):0.5);      distances and leaf names (popular)
(A:0.1,B:0.2,(C:0.3,D:0.4)E:0.5)F;    distances and all names
((B:0.2,(C:0.3,D:0.4)E:0.5)F:0.1)A;   a tree rooted on a leaf node (rare)
```

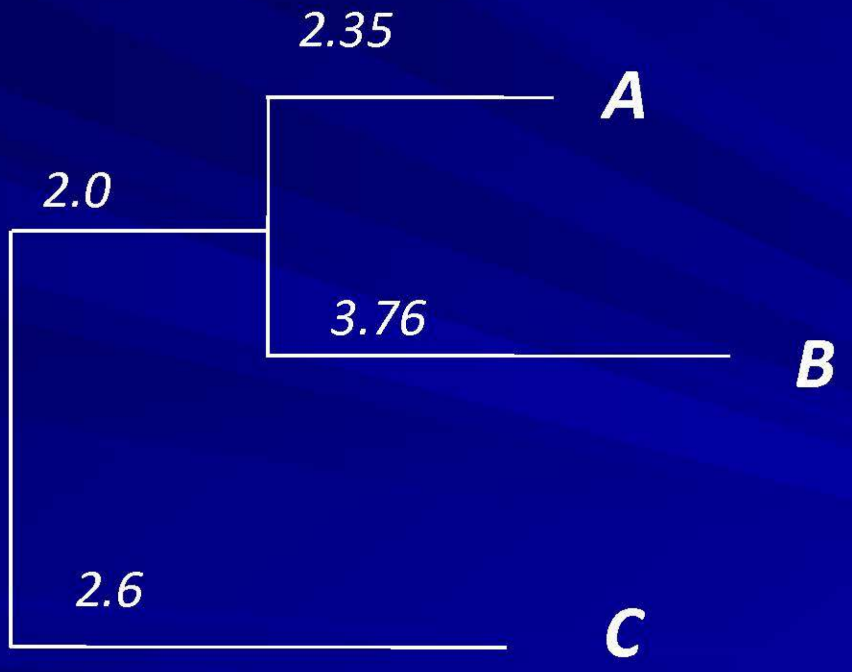
The NEWICK (or New Hampshire) format

(A, B);

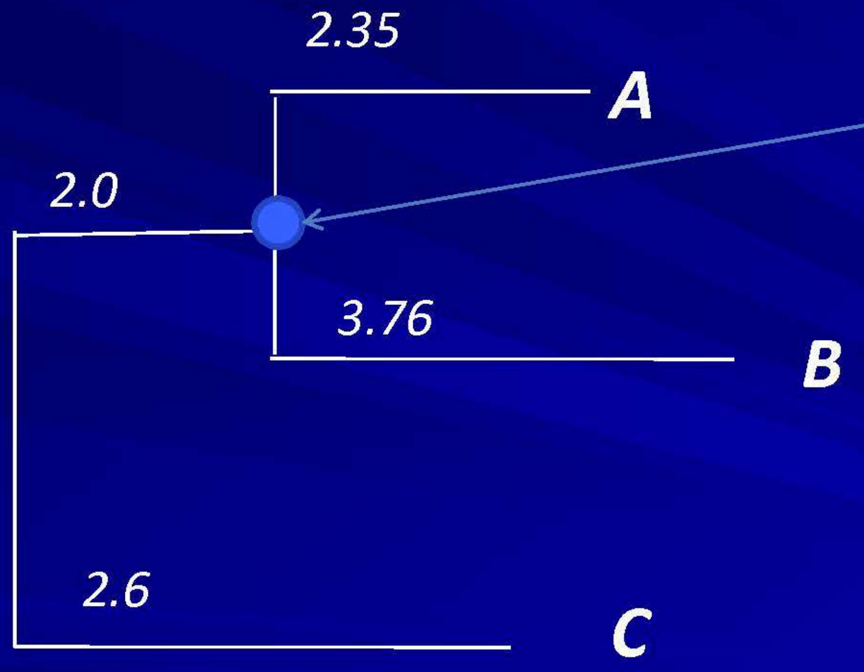


(A:2.35, B:3.76);

((A:2.35, B:3.76): 2.0,C:2.6);

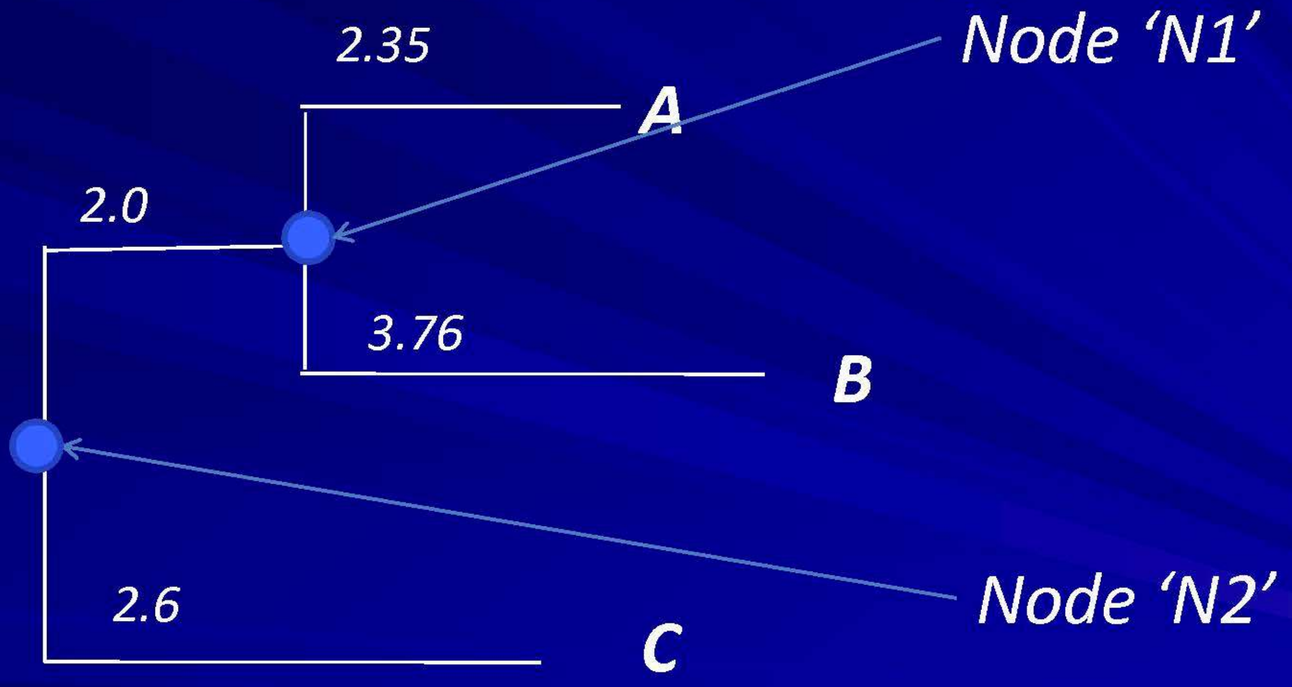


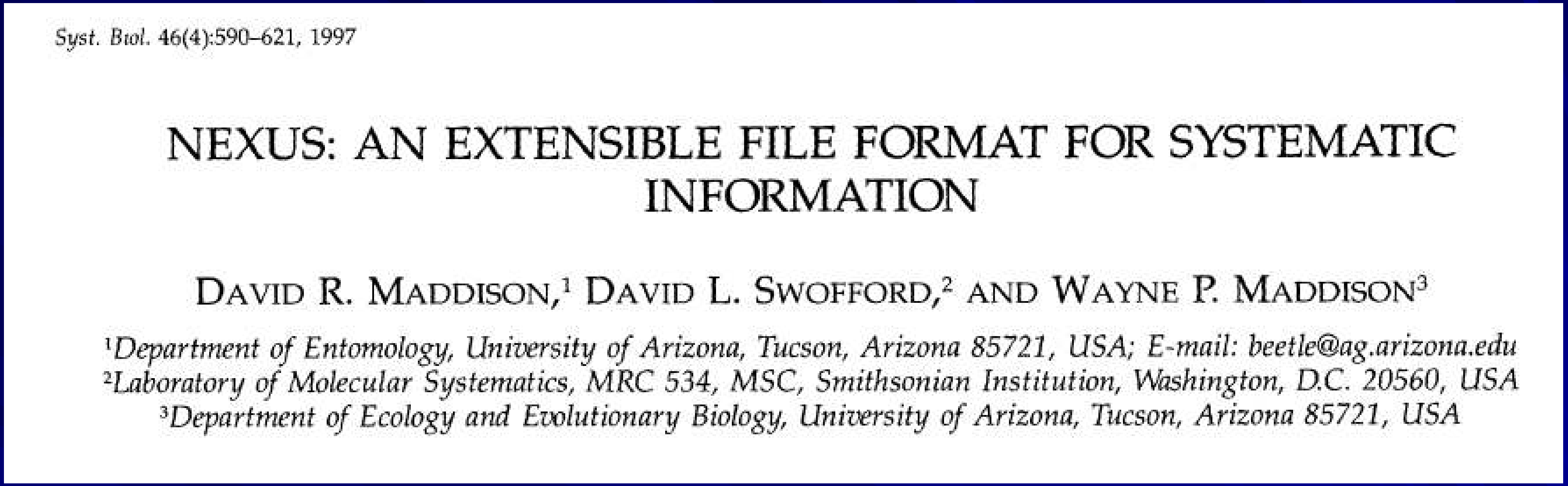
((A:2.35, B:3.76)N1: 2.0,C:2.6);



A NODE CALLED 'N1'

((A:2.35, B:3.76)N1:2.0,C:2.6)N2;





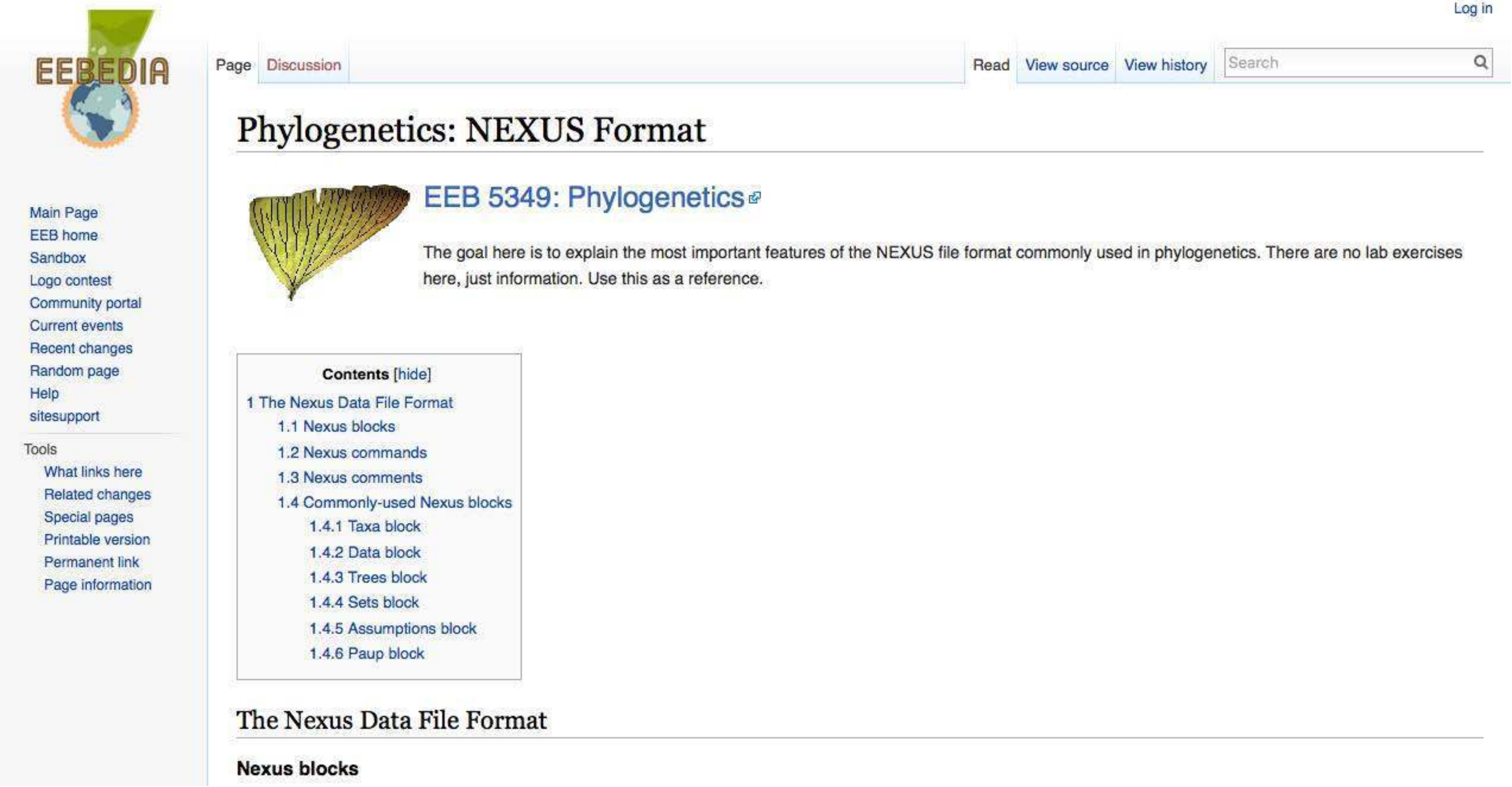
Syst. Biol. 46(4):590–621, 1997

NEXUS: AN EXTENSIBLE FILE FORMAT FOR SYSTEMATIC INFORMATION

David R. Maddison,[1] David L. Swofford,[2] and Wayne P. Maddison[3]

[1]*Department of Entomology, University of Arizona, Tucson, Arizona 85721, USA; E-mail: beetle@ag.arizona.edu*
[2]*Laboratory of Molecular Systematics, MRC 534, MSC, Smithsonian Institution, Washington, D.C. 20560, USA*
[3]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA*

David R Maddison



Phylogenetics: NEXUS Format

EEB 5349: Phylogenetics

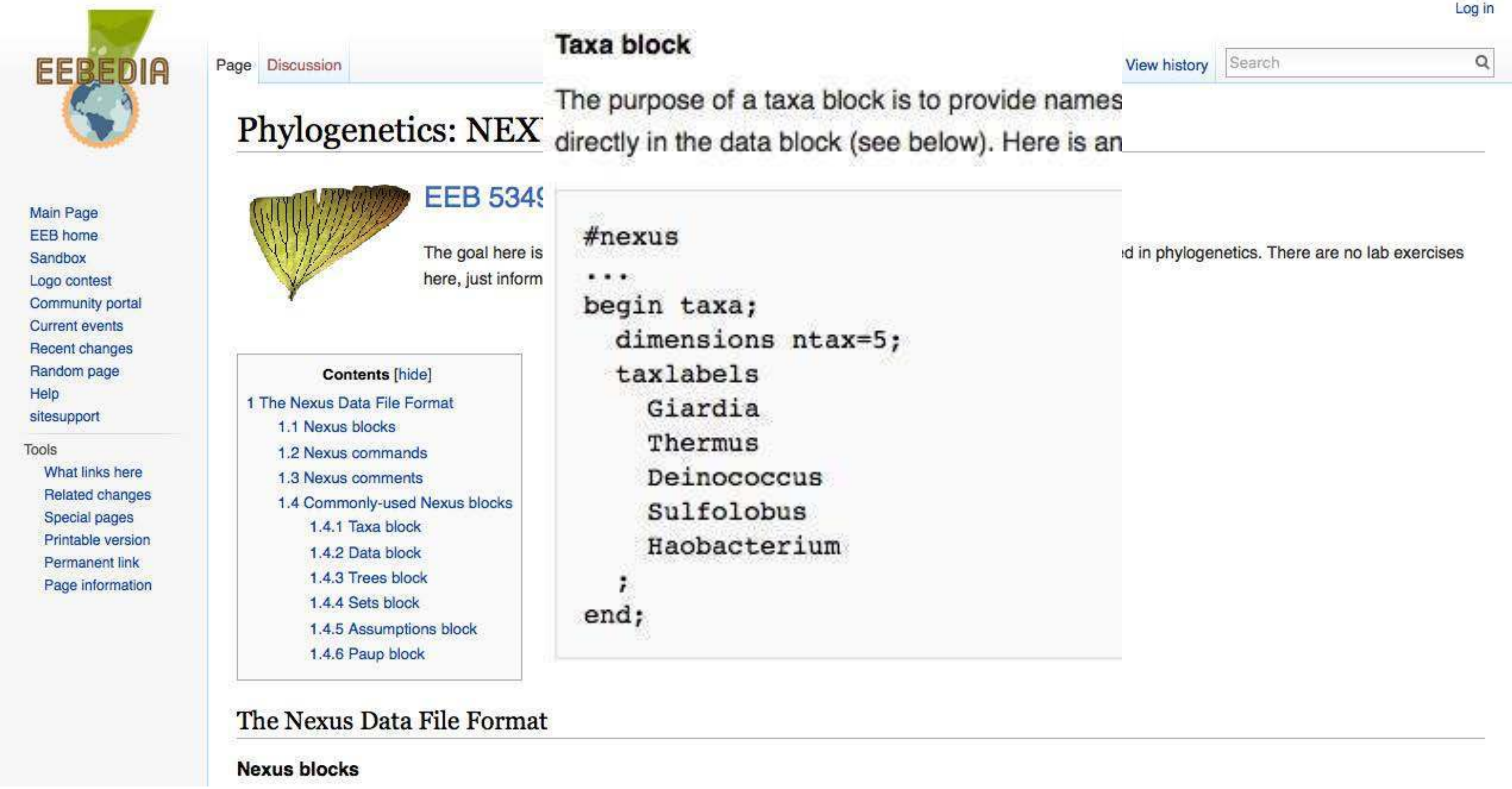
The goal here is to explain the most important features of the NEXUS file format commonly used in phylogenetics. There are no lab exercises here, just information. Use this as a reference.

Contents [hide]

The Nexus Data File Format

Nexus blocks

http://hydrodictyon.eeb.uconn.edu/eebedia/index.php/Phylogenetics:_NEXUS_Format



Log in

Page Discussion View history Search

Phylogenetics: NEX

EEB 534

The goal here is ...d in phylogenetics. There are no lab exercises here, just inform

- Main Page
- EEB home
- Sandbox
- Logo contest
- Community portal
- Current events
- Recent changes
- Random page
- Help
- sitesupport

Tools

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link
- Page information

Contents [hide]

Taxa block

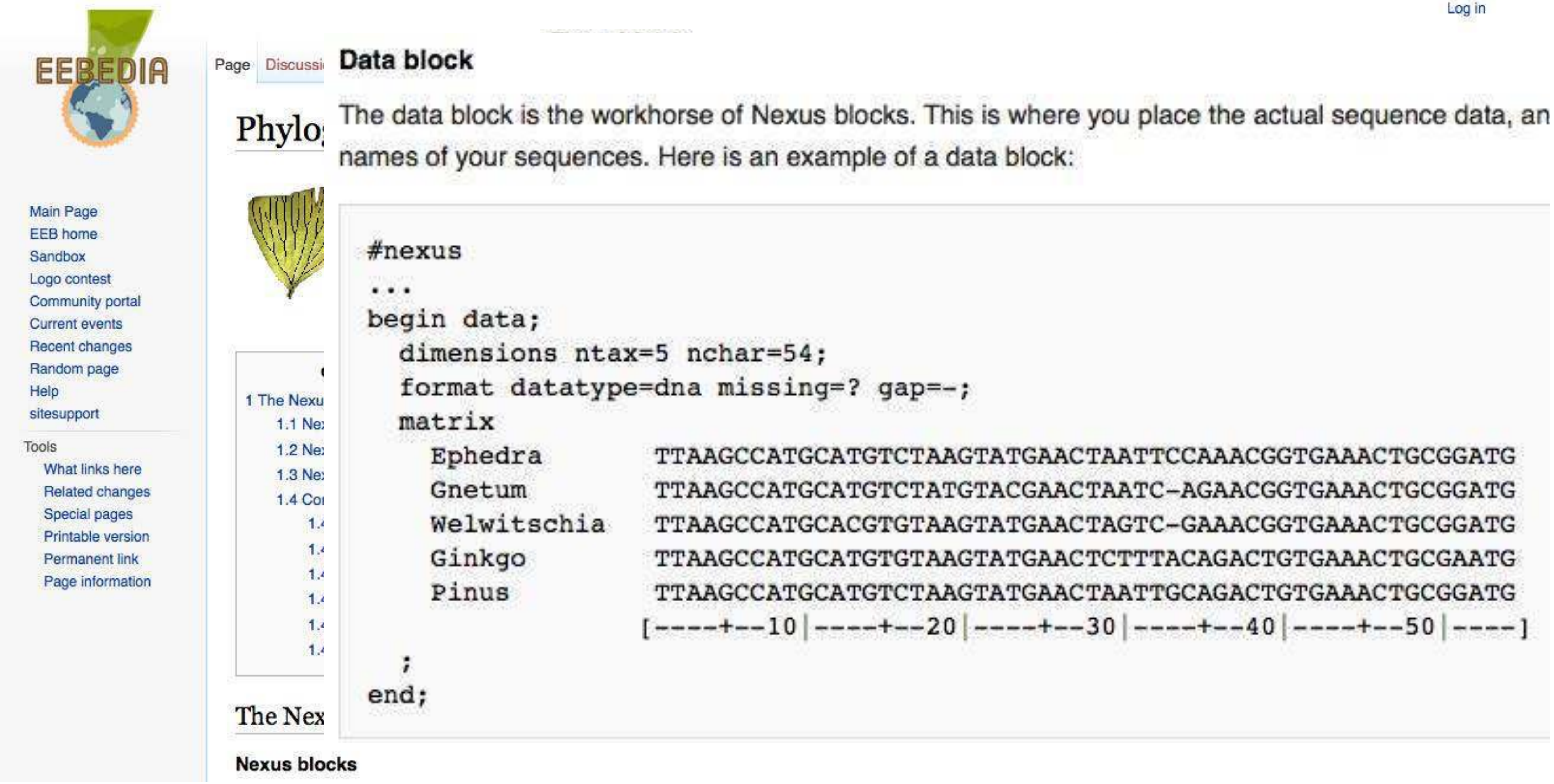
The purpose of a taxa block is to provide names directly in the data block (see below). Here is an

```
#nexus
...
begin taxa;
  dimensions ntax=5;
  taxlabels
    Giardia
    Thermus
    Deinococcus
    Sulfolobus
    Haobacterium
  ;
end;
```

The Nexus Data File Format

Nexus blocks

http://hydrodictyon.eeb.uconn.edu/eebedia/index.php/Phylogenetics:_NEXUS_Format

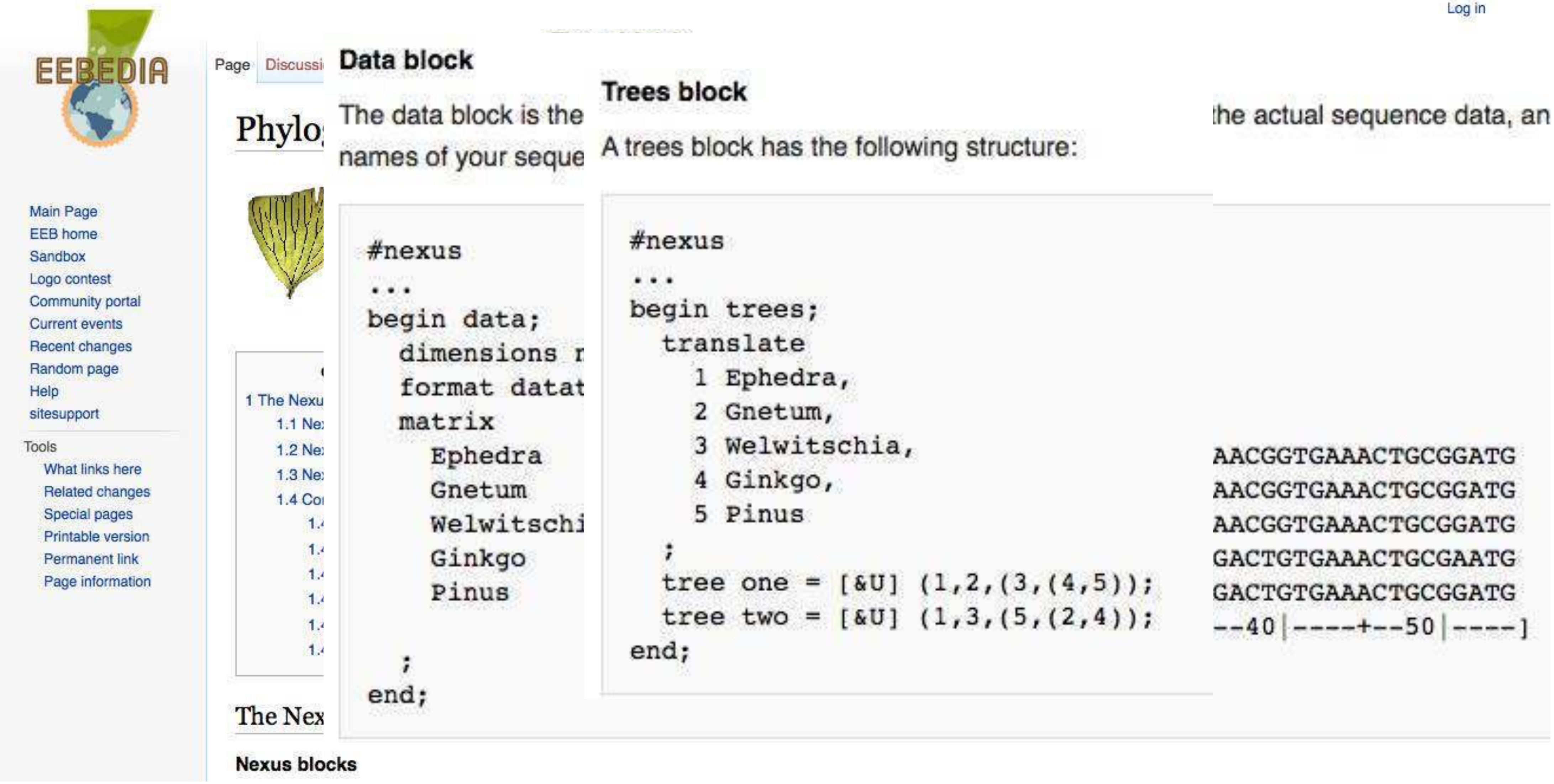


Data block

The data block is the workhorse of Nexus blocks. This is where you place the actual sequence data, an names of your sequences. Here is an example of a data block:

```
#nexus
...
begin data;
  dimensions ntax=5 nchar=54;
  format datatype=dna missing=? gap=-;
  matrix
    Ephedra       TTAAGCCATGCATGTCTAAGTATGAACTAATTCCAAACGGTGAAACTGCGGATG
    Gnetum        TTAAGCCATGCATGTCTATGTACGAACTAATC-AGAACGGTGAAACTGCGGATG
    Welwitschia   TTAAGCCATGCACGTGTAAGTATGAACTAGTC-GAAACGGTGAAACTGCGGATG
    Ginkgo        TTAAGCCATGCATGTGTAAGTATGAACTCTTTACAGACTGTGAAACTGCGAATG
    Pinus         TTAAGCCATGCATGTCTAAGTATGAACTAATTGCAGACTGTGAAACTGCGGATG
                 [----+--10|----+--20|----+--30|----+--40|----+--50|----]
  ;
end;
```

http://hydrodictyon.eeb.uconn.edu/eebedia/index.php/Phylogenetics:_NEXUS_Format



Data block

The data block is the … the actual sequence data, an

names of your seque

```
#nexus
...
begin data;
  dimensions r
  format datat
  matrix
    Ephedra
    Gnetum
    Welwitschi
    Ginkgo
    Pinus

  ;
end;
```

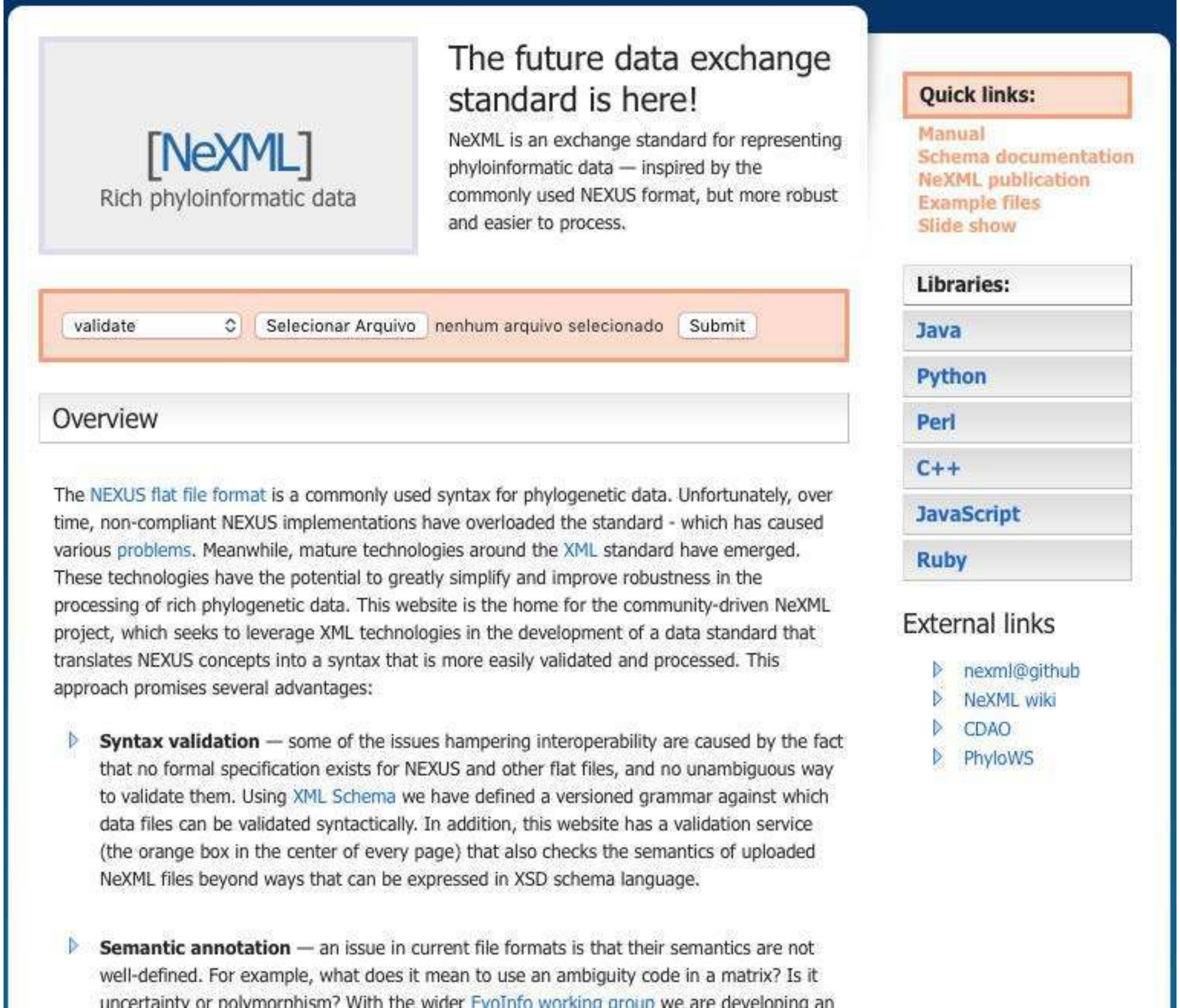
```
AACGGTGAAACTGCGGATG
AACGGTGAAACTGCGGATG
AACGGTGAAACTGCGGATG
GACTGTGAAACTGCGAATG
GACTGTGAAACTGCGGATG
--40|----+--50|----]
```

Trees block

A trees block has the following structure:

```
#nexus
...
begin trees;
  translate
    1 Ephedra,
    2 Gnetum,
    3 Welwitschia,
    4 Ginkgo,
    5 Pinus
  ;
  tree one = [&U] (1,2,(3,(4,5));
  tree two = [&U] (1,3,(5,(2,4));
end;
```

http://hydrodictyon.eeb.uconn.edu/eebedia/index.php/Phylogenetics:_NEXUS_Format



The future data exchange standard is here!

NeXML is an exchange standard for representing phyloinformatic data — inspired by the commonly used NEXUS format, but more robust and easier to process.

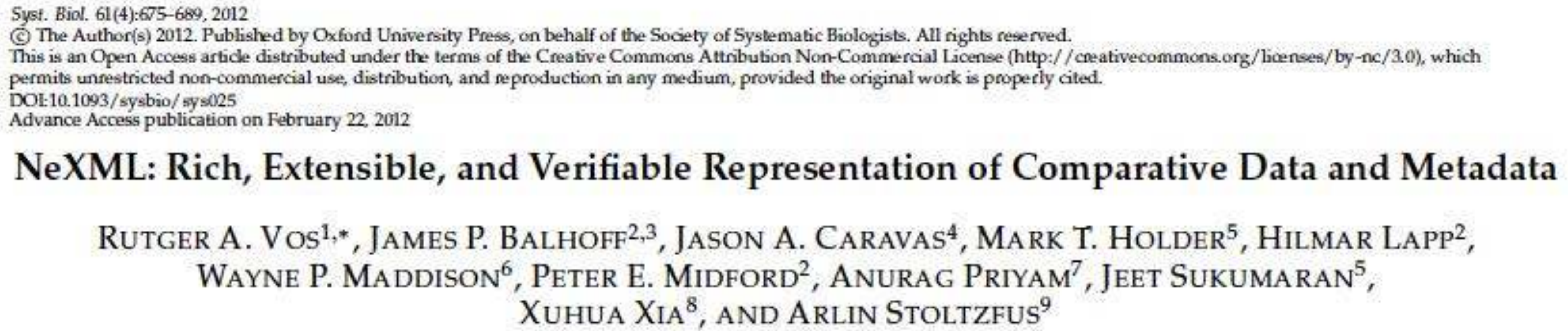
Overview

The NEXUS flat file format is a commonly used syntax for phylogenetic data. Unfortunately, over time, non-compliant NEXUS implementations have overloaded the standard - which has caused various problems. Meanwhile, mature technologies around the XML standard have emerged. These technologies have the potential to greatly simplify and improve robustness in the processing of rich phylogenetic data. This website is the home for the community-driven NeXML project, which seeks to leverage XML technologies in the development of a data standard that translates NEXUS concepts into a syntax that is more easily validated and processed. This approach promises several advantages:

- **Syntax validation** — some of the issues hampering interoperability are caused by the fact that no formal specification exists for NEXUS and other flat files, and no unambiguous way to validate them. Using XML Schema we have defined a versioned grammar against which data files can be validated syntactically. In addition, this website has a validation service (the orange box in the center of every page) that also checks the semantics of uploaded NeXML files beyond ways that can be expressed in XSD schema language.
- **Semantic annotation** — an issue in current file formats is that their semantics are not well-defined. For example, what does it mean to use an ambiguity code in a matrix? Is it uncertainty or polymorphism? With the wider EvoInfo working group we are developing an

The future data exchange standard is here!

NeXML is an exchange standard for representing phyloinformatic data — inspired by the commonly used NEXUS format, but more robust and easier to process.

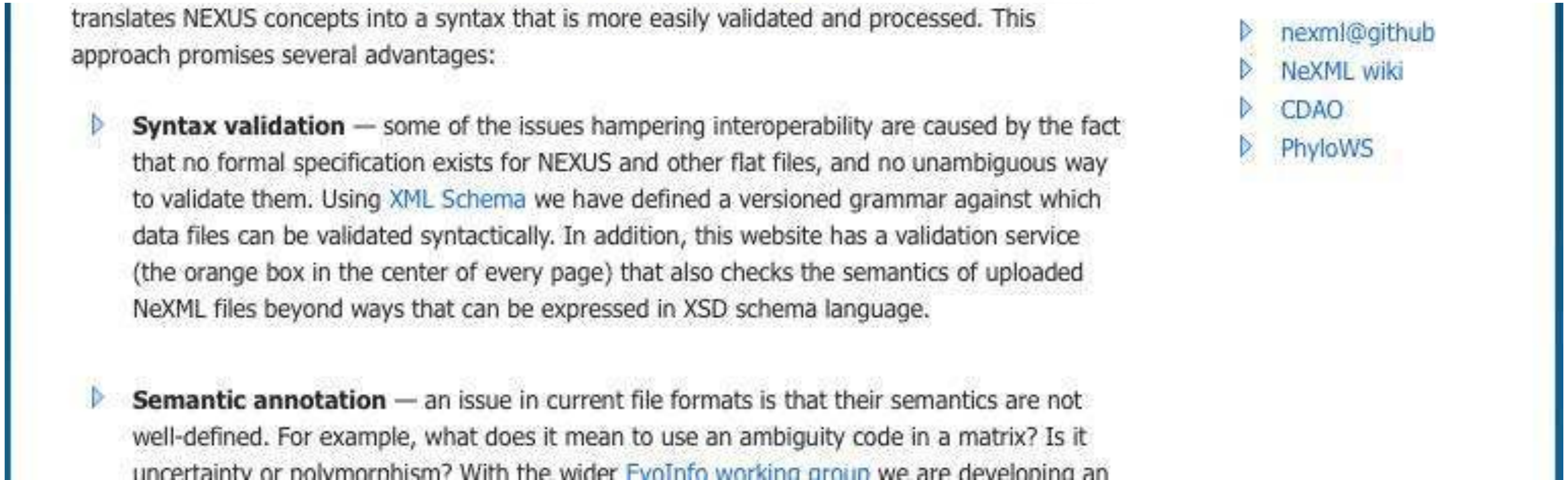


Syst. Biol. 61(4):675–689, 2012

DOI:10.1093/sysbio/sys025
Advance Access publication on February 22, 2012

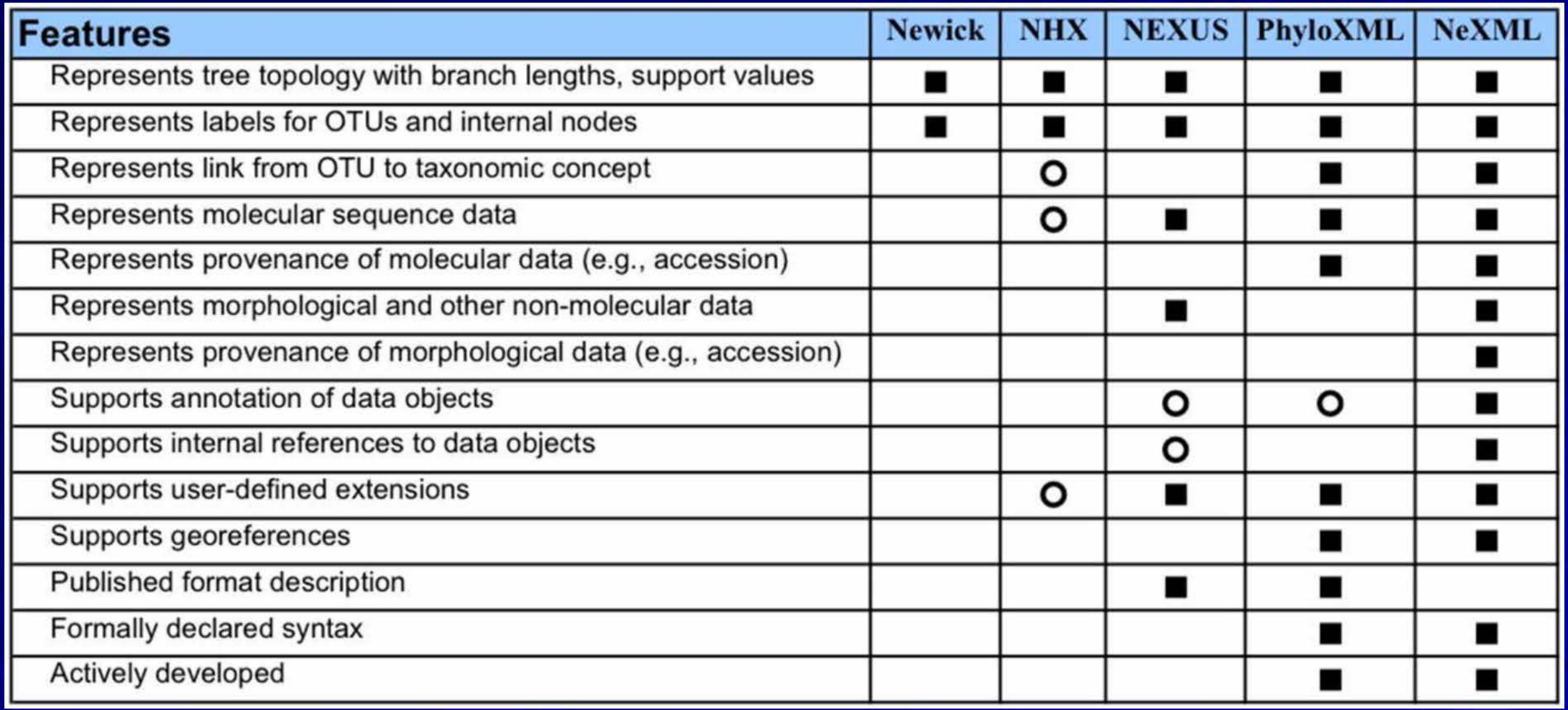
NeXML: Rich, Extensible, and Verifiable Representation of Comparative Data and Metadata

RUTGER A. VOS[1,*], JAMES P. BALHOFF[2,3], JASON A. CARAVAS[4], MARK T. HOLDER[5], HILMAR LAPP[2], WAYNE P. MADDISON[6], PETER E. MIDFORD[2], ANURAG PRIYAM[7], JEET SUKUMARAN[5], XUHUA XIA[8], AND ARLIN STOLTZFUS[9]



translates NEXUS concepts into a syntax that is more easily validated and processed. This approach promises several advantages:

- **Syntax validation** — some of the issues hampering interoperability are caused by the fact that no formal specification exists for NEXUS and other flat files, and no unambiguous way to validate them. Using XML Schema we have defined a versioned grammar against which data files can be validated syntactically. In addition, this website has a validation service (the orange box in the center of every page) that also checks the semantics of uploaded NeXML files beyond ways that can be expressed in XSD schema language.
- **Semantic annotation** — an issue in current file formats is that their semantics are not well-defined. For example, what does it mean to use an ambiguity code in a matrix? Is it uncertainty or polymorphism? With the wider EvoInfo working group we are developing an



Features	Newick	NHX	NEXUS	PhyloXML	NeXML
Represents tree topology with branch lengths, support values	■	■	■	■	■
Represents labels for OTUs and internal nodes	■	■	■	■	■
Represents link from OTU to taxonomic concept		○		■	■
Represents molecular sequence data		○	■	■	■
Represents provenance of molecular data (e.g., accession)				■	■
Represents morphological and other non-molecular data			■		■
Represents provenance of morphological data (e.g., accession)					■
Supports annotation of data objects			○	○	■
Supports internal references to data objects			○		■
Supports user-defined extensions		○	■	■	■
Supports georeferences				■	■
Published format description			■	■	
Formally declared syntax				■	■
Actively developed				■	■

Breve revisão sobre métodos de inferência filogenética

Critérios de otimização

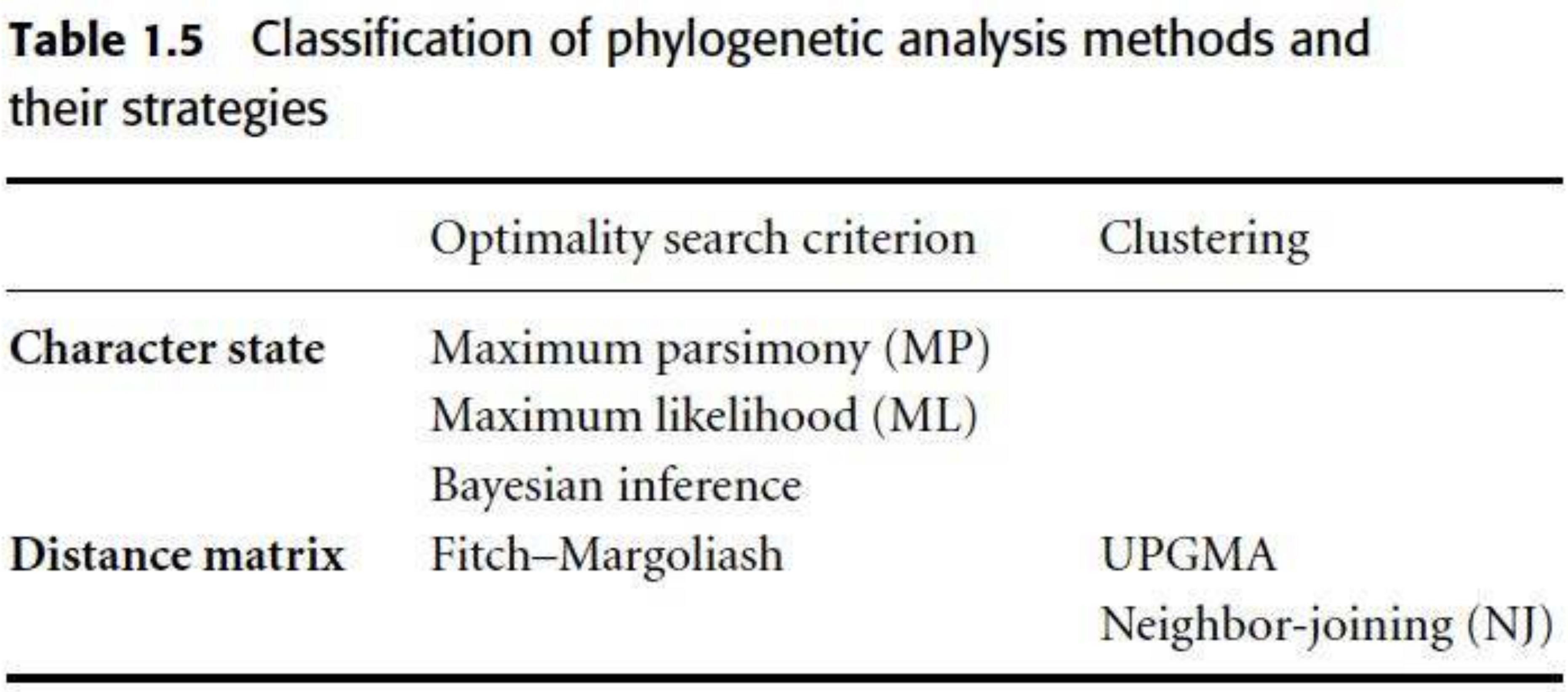
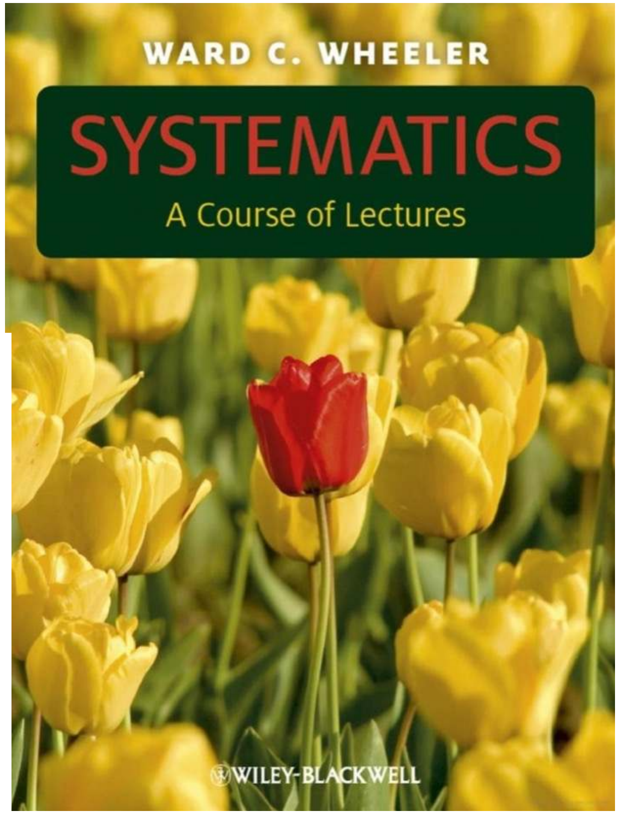


Table 1.5 Classification of phylogenetic analysis methods and their strategies

	Optimality search criterion	Clustering
Character state	Maximum parsimony (MP)	
	Maximum likelihood (ML)	
	Bayesian inference	
Distance matrix	Fitch–Margoliash	UPGMA
		Neighbor-joining (NJ)

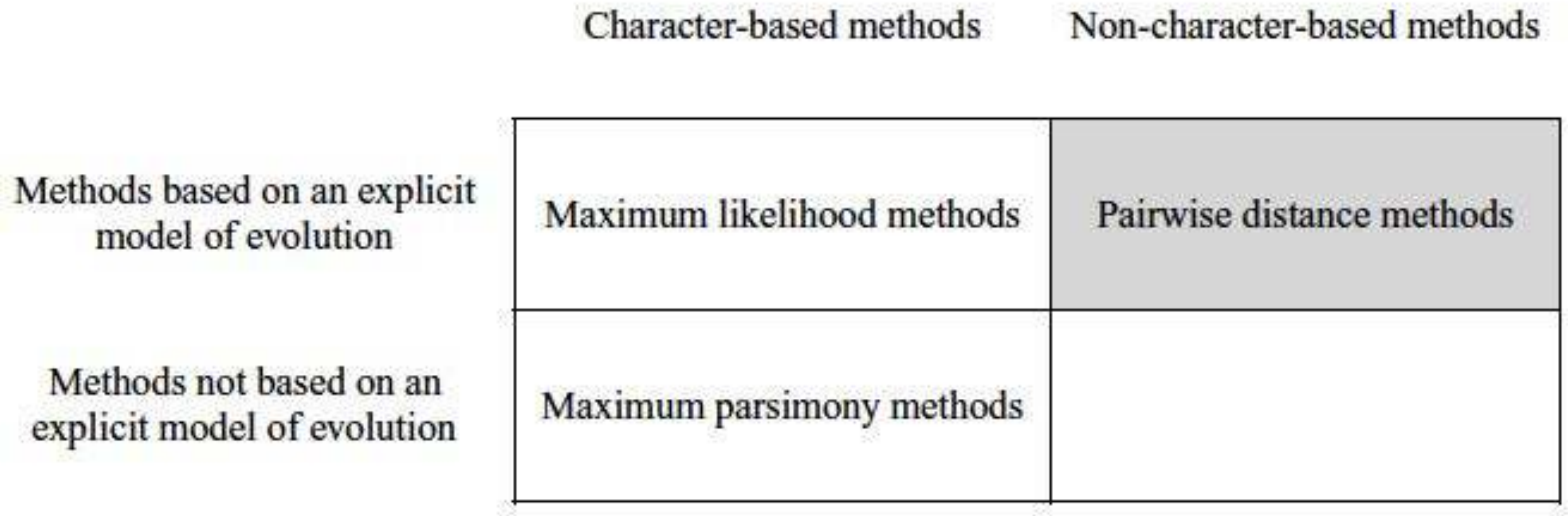


WARD C. WHEELER

SYSTEMATICS

A Course of Lectures

WILEY-BLACKWELL



	Character-based methods	Non-character-based methods
Methods based on an explicit model of evolution	Maximum likelihood methods	Pairwise distance methods
Methods not based on an explicit model of evolution	Maximum parsimony methods	

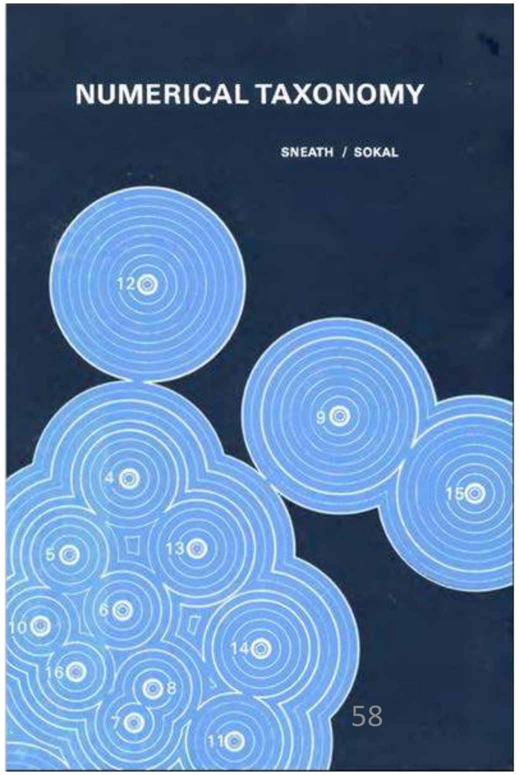
Van der Peer 2009

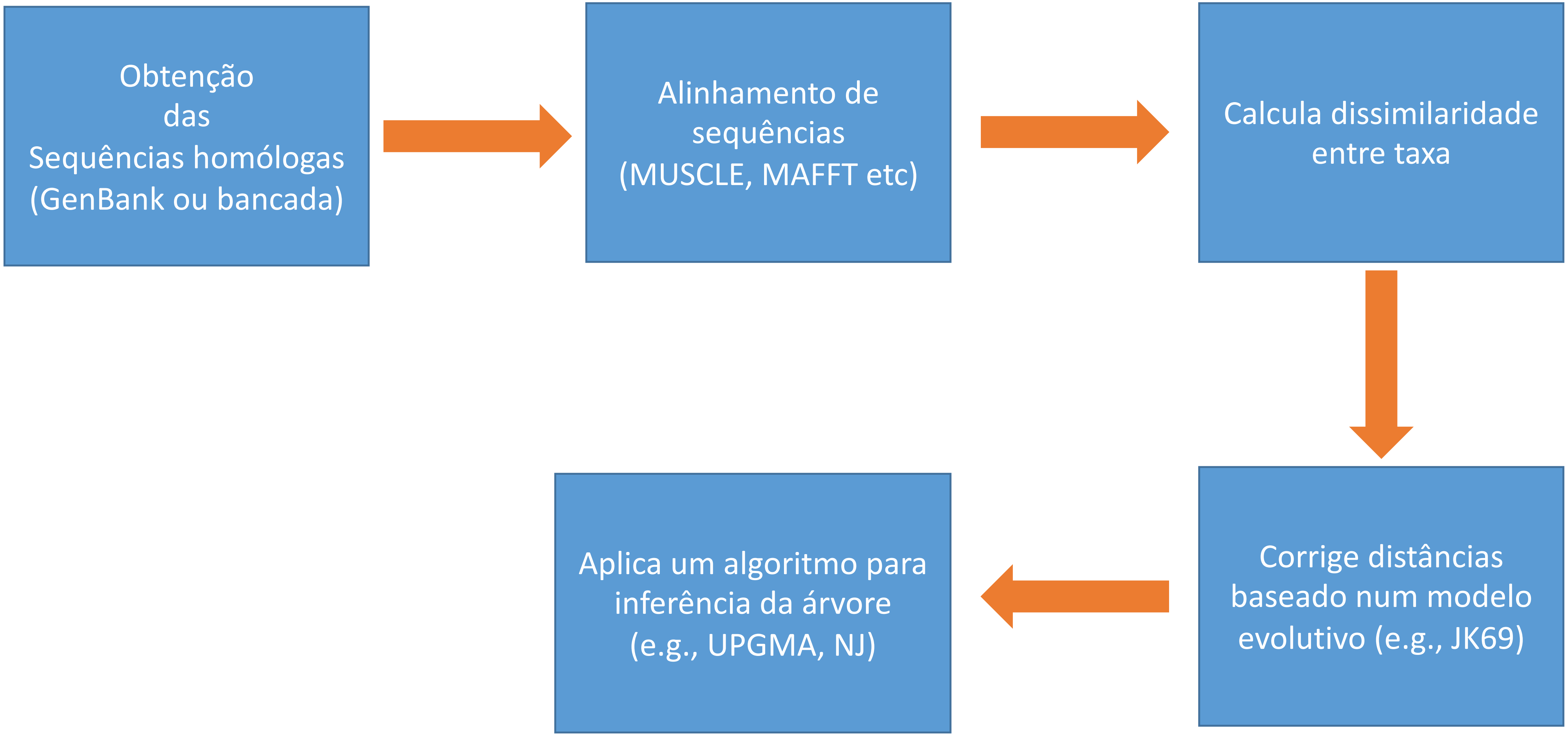
Métodos baseados em distância

- Calculam uma medida de distância para medir a dissimilaridade entre pares de taxa (baseada nas suas sequências alinhadas) para produzir uma matriz de distância par a par, e então inferir as relações entre os taxa

- Posteriormente se corrige essas distâncias utilizando um modelo evolutivo

- Algoritmos mais usados => UPGMA e NJ
  - NJ assume modelo de evolução mínima





Obtenção
das
Sequências homólogas
(GenBank ou bancada)

Alinhamento de
sequências
(MUSCLE, MAFFT etc)

Calcula dissimilaridade
entre taxa

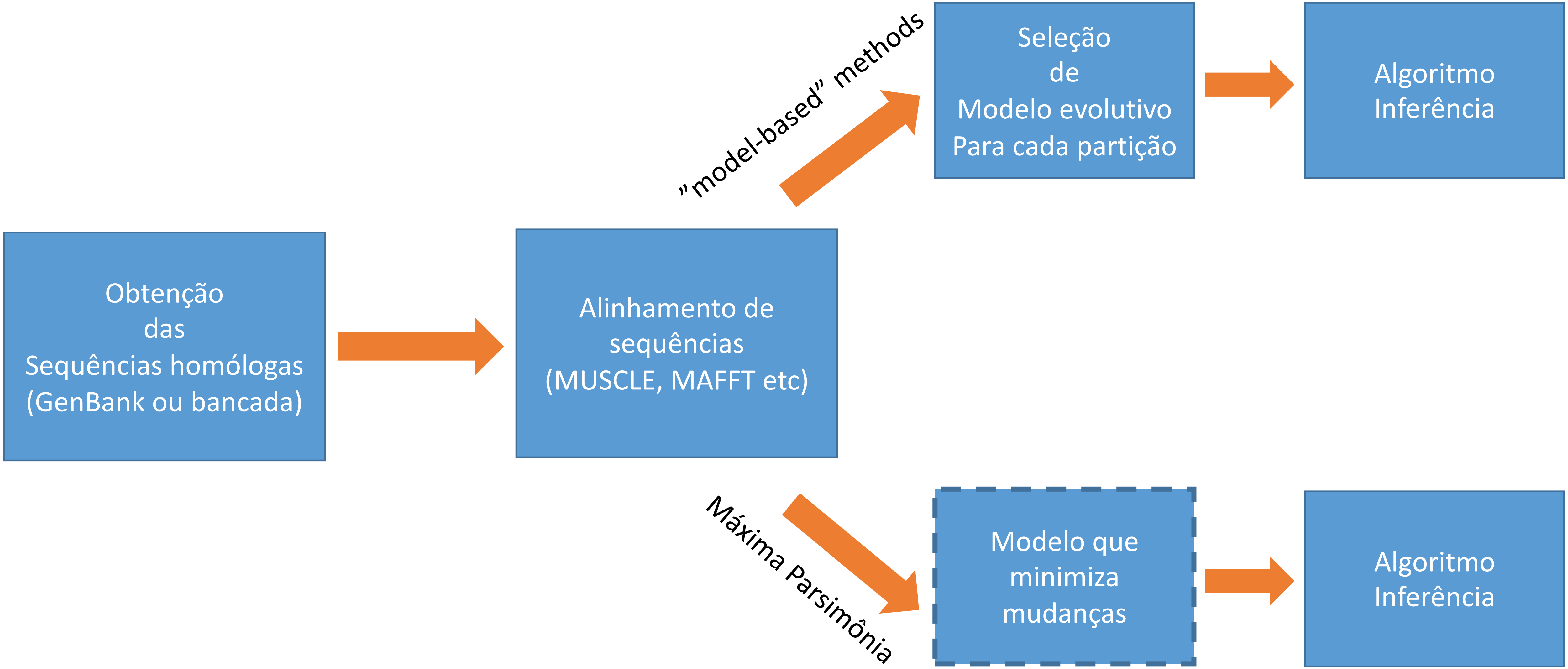
Corrige distâncias
baseado num modelo
evolutivo (e.g., JK69)

Aplica um algoritmo para
inferência da árvore
(e.g., UPGMA, NJ)

Métodos baseados em distância

- Vantagens
  - Rapidez, adequado para um grande conjunto de dados

- Desvantagens
  - Inconsistências na topologia e comprimento de ramo
  - Filosóficas (Fenética) => não está interessado em reconstruir a história dos taxa em termos de compartilhamento de caracteres derivados
  - Não é possível julgar qual árvore é melhor
  - Produzem uma única topologia
  - Não lida bem com heterogeneidade nas taxas de evolução
  - Toda a informação sobre as sequências é perdida, já que é transformada em distância
  - Não estima estados ancestral de caracteres



Obtenção
das
Sequências homólogas
(GenBank ou bancada)

Alinhamento de
sequências
(MUSCLE, MAFFT etc)

”model-based” methods

Seleção
de
Modelo evolutivo
Para cada partição

Algoritmo
Inferência

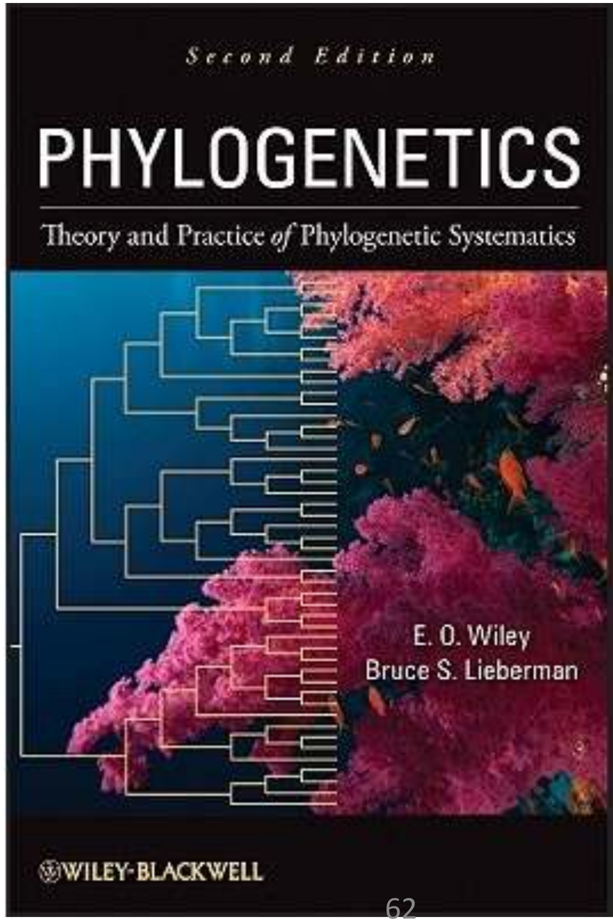
Máxima Parsimônia

Modelo que
minimiza
mudanças

Algoritmo
Inferência

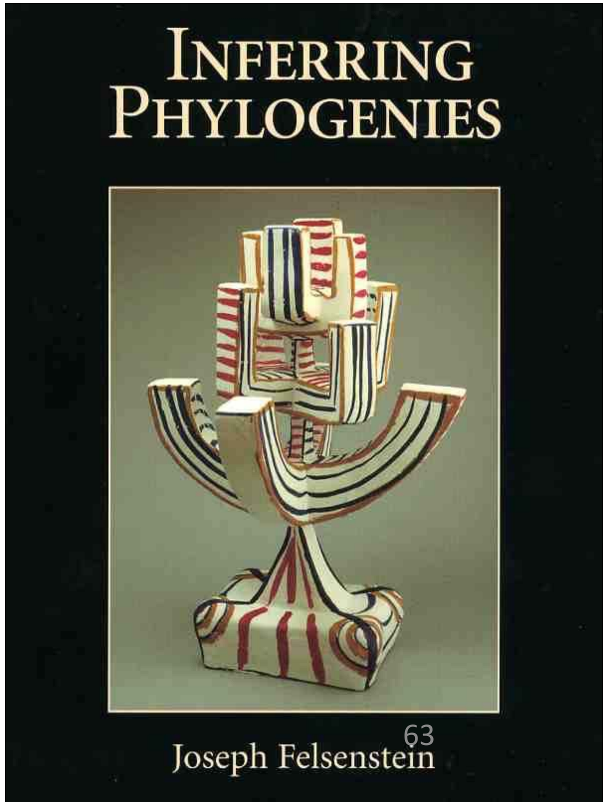
Máxima Parsimônia (MP)

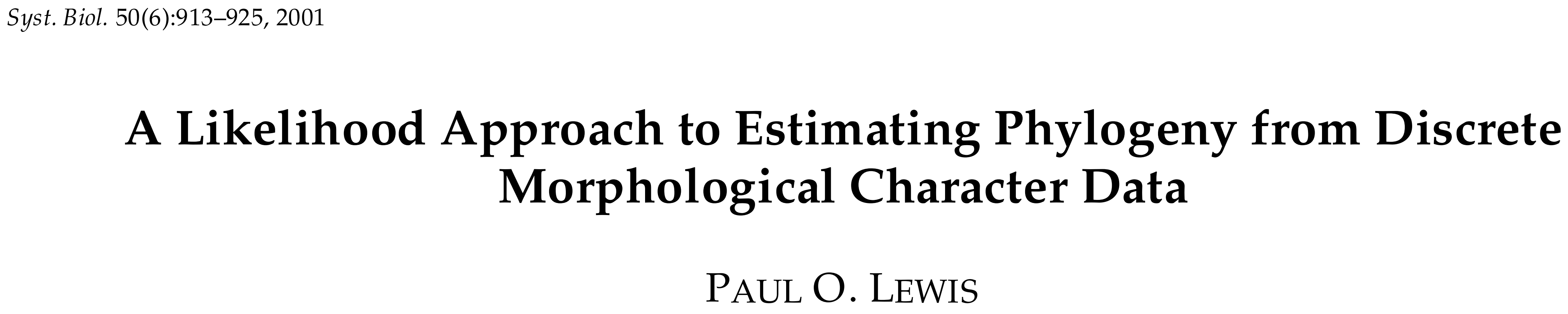
- Procura a topologia que explica um conjunto de sequências alinhadas de forma a minimizar o número de mudanças de caracteres
- Vantagens
  - Simplicidade do critério de otimização
  - Trabalha tanto com dados morfológicos quanto moleculares
  - Rápida para muitos taxa
- Problemas
  - MP assume que um caracter comum é herdado diretamente de ancestral comum. Isso pode subestimar a divergência real entre taxons distantes (*long branch attraction*)
  - Não funciona bem quando a divergência entre sequências é alta
  - Estimativa enviesada de comprimento de ramo



Máxima Verossimilhança (ML)

- Examina diferentes topologias e avalia o suporte relativo somando todas as posições de sequências.
- Sustentada por critérios estatísticos ao invés de filosóficos
- Algoritmos de ML procuram pela topologia e os comprimentos de ramo que maximizam a probabilidade de se observar os estados de caracteres (verossimilhança), dada a topologia e um modelo de evolução
  - Precisa de uma topologia de início e um modelo de evolução explicito
- Vantagens
  - Menos problema com long branch attraction
  - Resultados entre MP e ML podem convergir (”non-common mechanism”)
- Desvantagens
  - Pode fornecer a árvore errada se o modelo de evolução escolhido for o errado
  - Pode ser bem lenta





Syst. Biol. 50(6):913–925, 2001

A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data

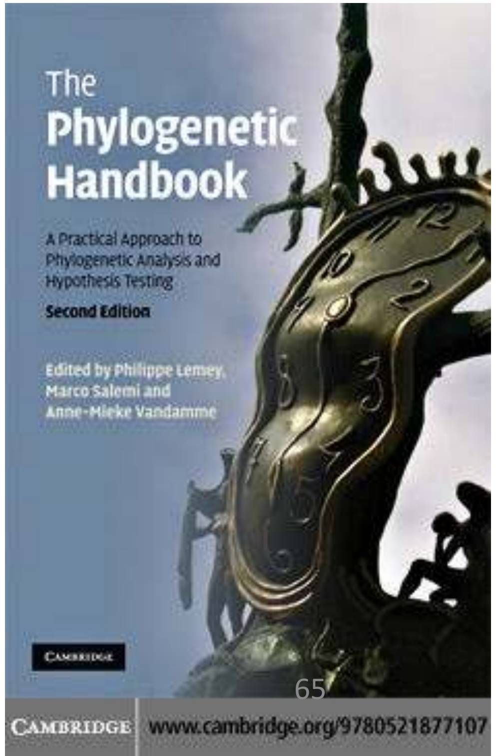
Paul O. Lewis

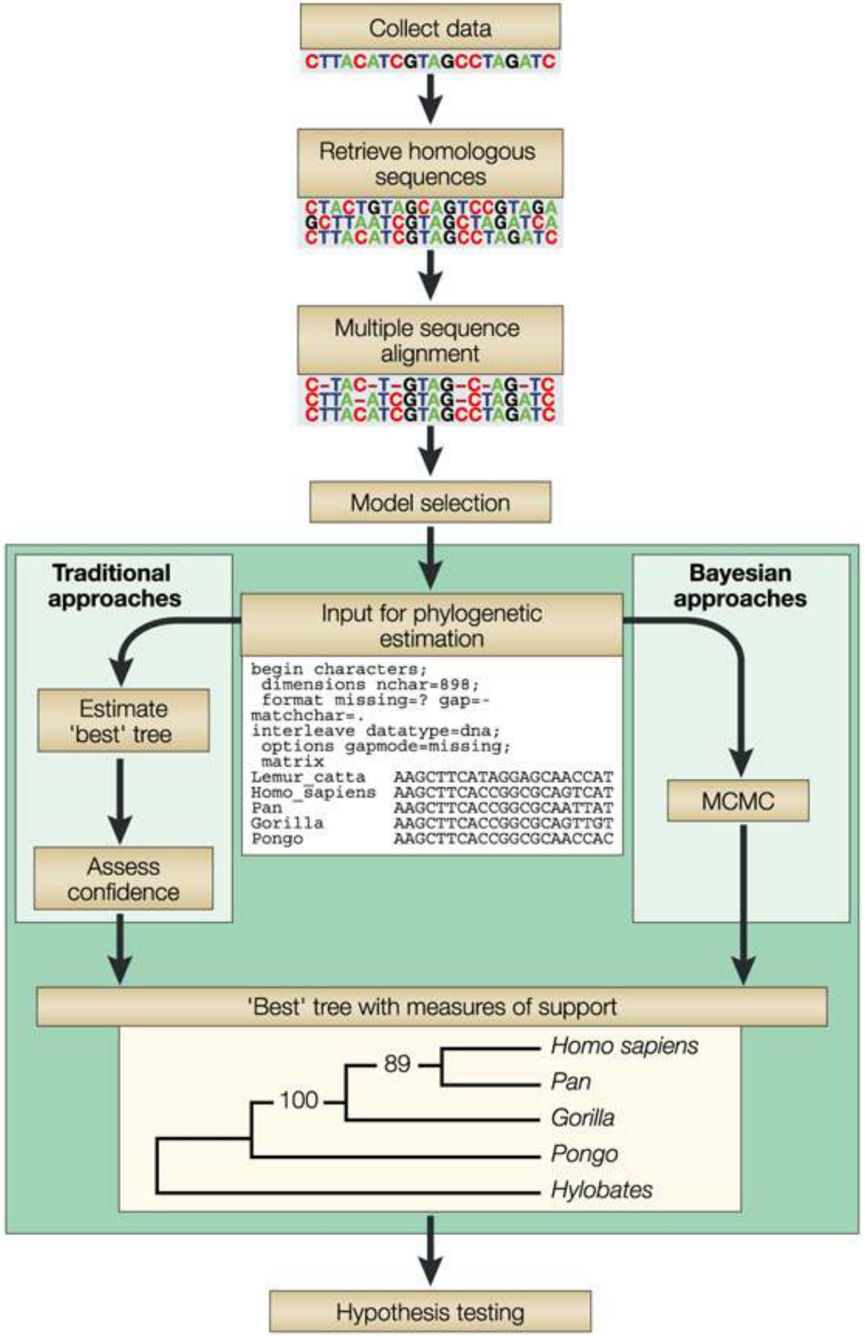
$$Q = \alpha \begin{bmatrix} 1-k & 1 & \dots & 1 \\ 1 & 1-k & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1-k \end{bmatrix}$$

Modelo de Markov
adaptado para
caracteres
morfológicos
discretos (binários)

Inferência Bayesiana (BI)

- Também usam o conceitos de verossimilhança (likelihood), mas não tentam buscar somente pela única melhor árvore. Ao invés disso buscam a distribuição de probabilidade de árvores. Esta probabilidade posterior possui uma estimativa de confidência.
- Distribuições posteriores de árvore são obtidas explorando o espaço de árvores usando um método chamado MCMC
- Normalmente se descarta (”burn-in”) uma parte inicial destas árvores que foram estimadas antes do MCMC convergir
- Precisam de um *prior*, uma informação de crença prévia, que é formalizado pela distribuição estatística anterior dos parâmetros modelo





```
begin characters;
 dimensions nchar=898;
 format missing=? gap=-
matchchar=.
interleave datatype=dna;
 options gapmode=missing;
 matrix
Lemur_catta    AAGCTTCATAGGAGCAACCAT
Homo_sapiens   AAGCTTCACCGGCGCAGTCAT
Pan            AAGCTTCACCGGCGCAATTAT
Gorilla        AAGCTTCACCGGCGCAGTTGT
Pongo          AAGCTTCACCGGCGCAACCAC
```

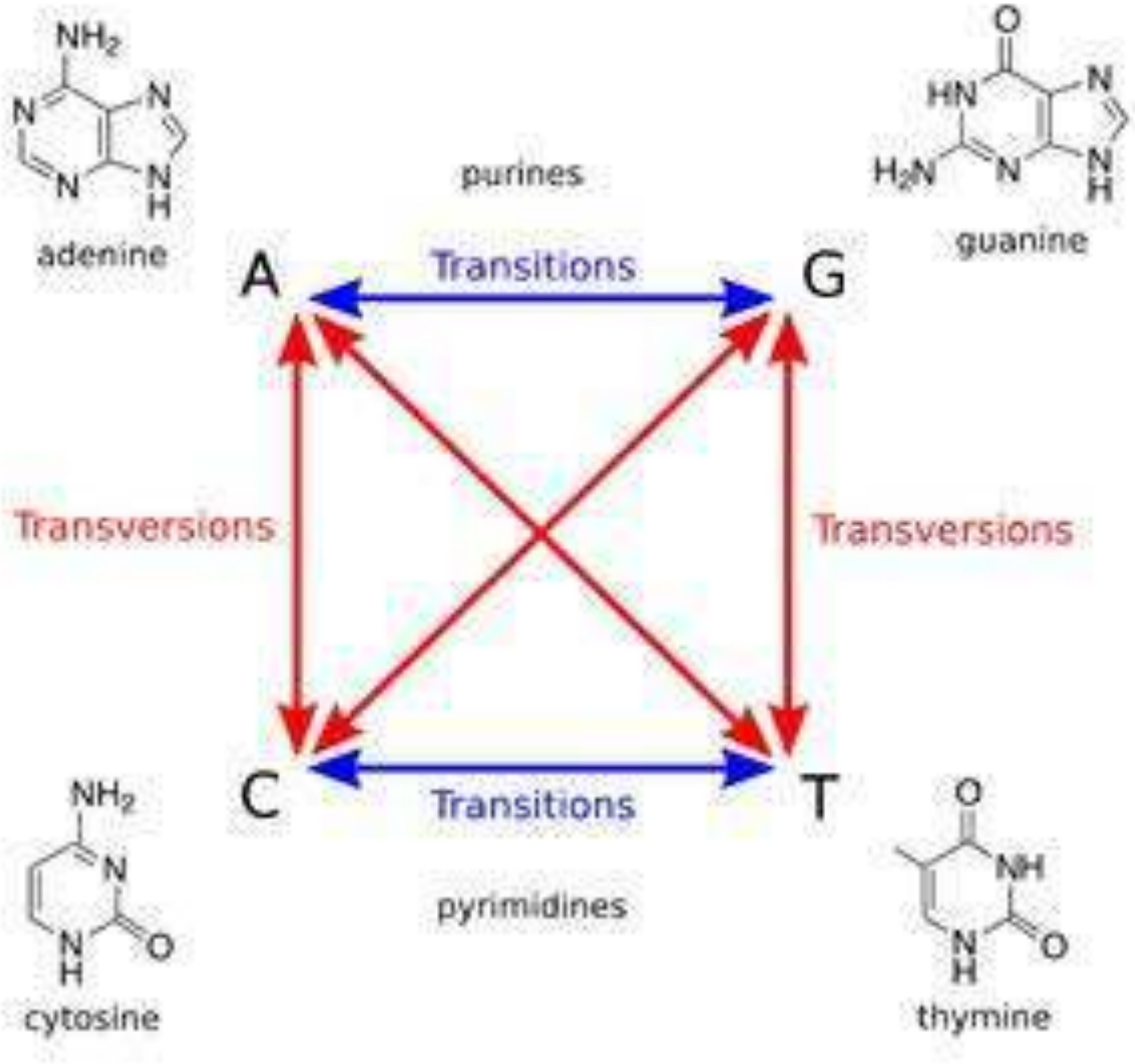
Nature Reviews | Genetics

Holder & Lewis 2003

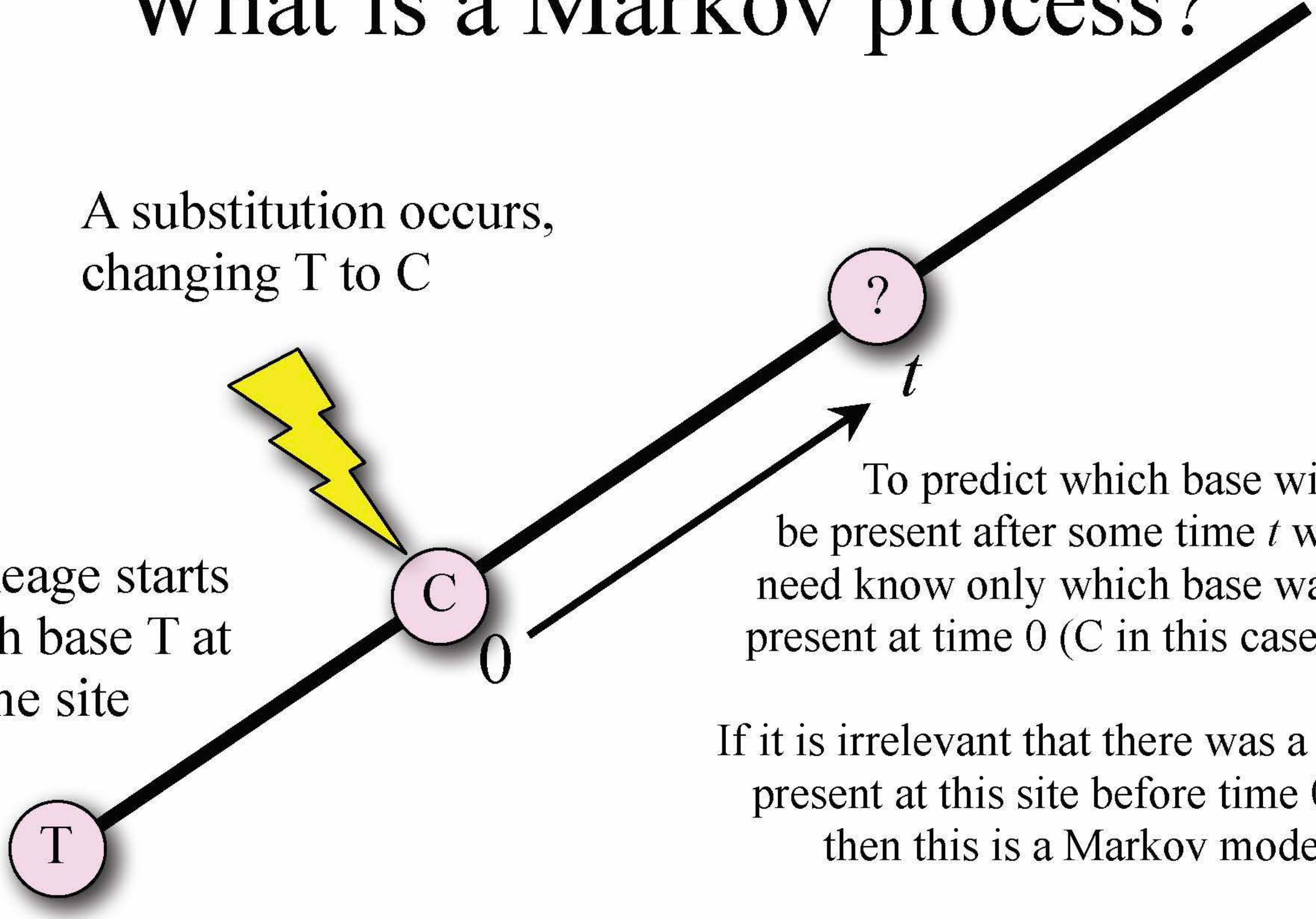
Modelos de evolução de sequências de nucleotídeos

- Processos de Markov homogêneos de tempo contínuo
- Processos de Markov são descritos em matrizes *Q*
  - Matrizes 4x4 na qual cada elemento descreve a taxa instantânea de mudança de uma base para a outra

$$P(t) = \begin{pmatrix} p_{AA}(t) & p_{GA}(t) & p_{CA}(t) & p_{TA}(t) \\ p_{AG}(t) & p_{GG}(t) & p_{CG}(t) & p_{TG}(t) \\ p_{AC}(t) & p_{GC}(t) & p_{CC}(t) & p_{TC}(t) \\ p_{AT}(t) & p_{GT}(t) & p_{CT}(t) & p_{TT}(t) \end{pmatrix}$$


adenine
purines
guanine
A
Transitions
G
Transversions
Transversions
C
Transitions
T
pyrimidines
cytosine
thymine

What is a Markov process?



A substitution occurs,
changing T to C

Lineage starts
with base T at
some site

To predict which base will
be present after some time *t* we
need know only which base was
present at time 0 (C in this case).

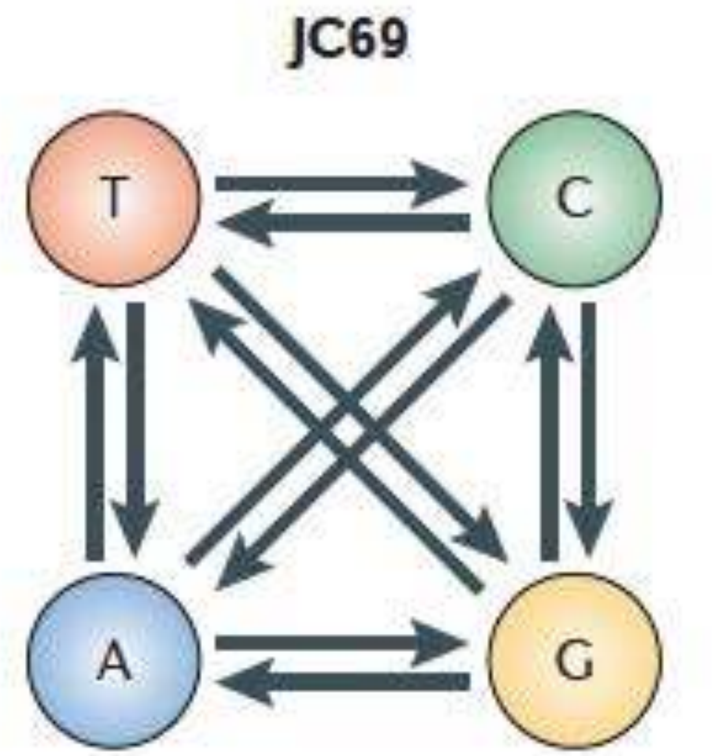
If it is irrelevant that there was a T
present at this site before time 0,
then this is a Markov model.

Modelos de Evolução de DNA

- Pressupostos
  - Em qualquer instante, a taxa de mudança da base *i* para a *j* é independente da informação anterior a i. Dizemos que um processo de Markok tem ”memória curta”
  - Tempo homogêneo/contínuo – taxas de substituição não mudam com o tempo
  - Estacionários– frequência relativa das bases está (π_A,π_C,π_G,π_T) em equilíbrio
  - Muitos modelos são também time-reversible – a taxa de mudança de i para j é o mesmo de j para i.
- Necessários para modelar substituições como um processo estocástico

Modelos de Evolução de sequências

- **Jukes Cantor (JC69) – o modelo mais simples**
- Pressupostos:
  - Frequências de equilíbrio dos nucleotídeos são 25% ($\pi_A=\pi_C=\pi_G=\pi_T=1/4$)
  - Probabilidade iguais para qualquer substituição (a=b=c=d=e=f=1)



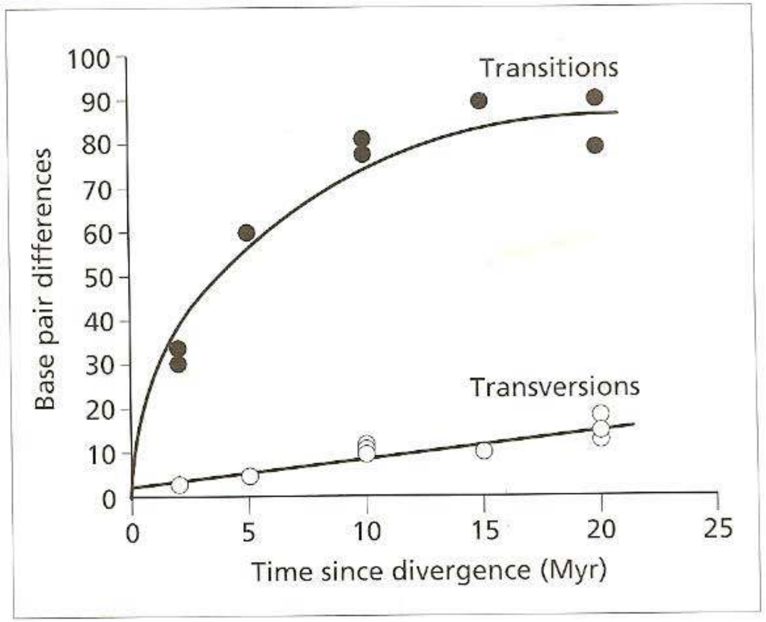
$$Q = \begin{pmatrix} -3/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & -3/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & -3/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & -3/4 \end{pmatrix}$$

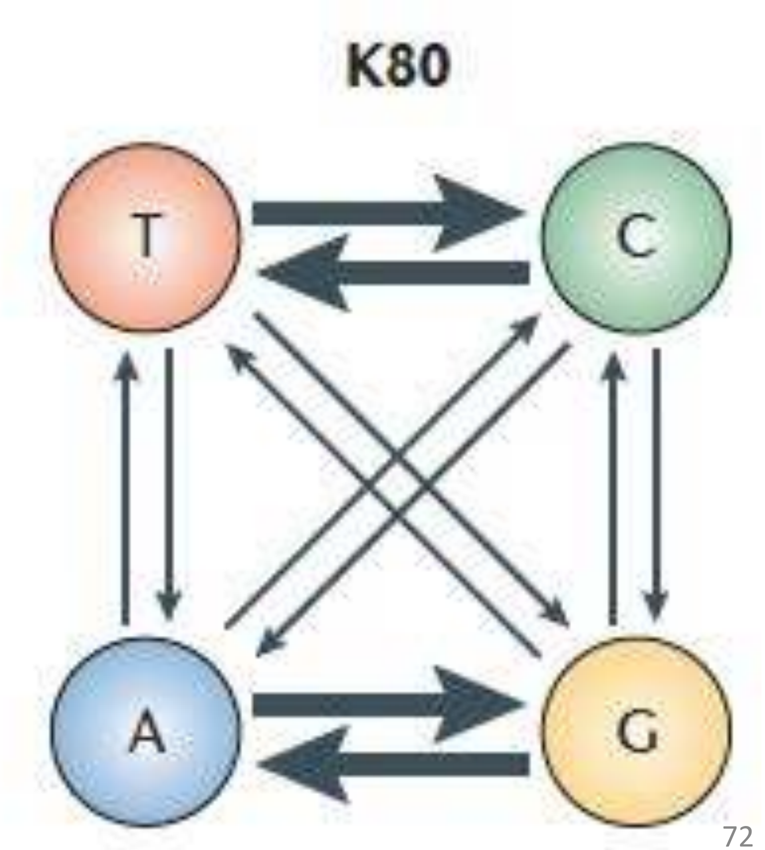
Figure 4.4 Instantaneous rate Q matrix for the Jukes and Cantor model (JC69).

Modelos de Evolução de sequências

- **Kimura 2-parameter (K2P)**
- Geralmente, *transições* ocorrem com taaxs maiores do que *transversões*
- Viola os pressupostos do JC69
- K2P adiciona um parâmetro a mais para modelar cada tipo de substituição
- Assume frequências iguais de nucleotídeos em equlíbrio

Fig. 5.13 The number of transitions and transversions between the same bovid mammal sequences used in Fig. 5.11. Transitions accumulate much more rapidly than transversions and become saturated, whereas transversions accumulate more slowly and show no evidence of saturation.

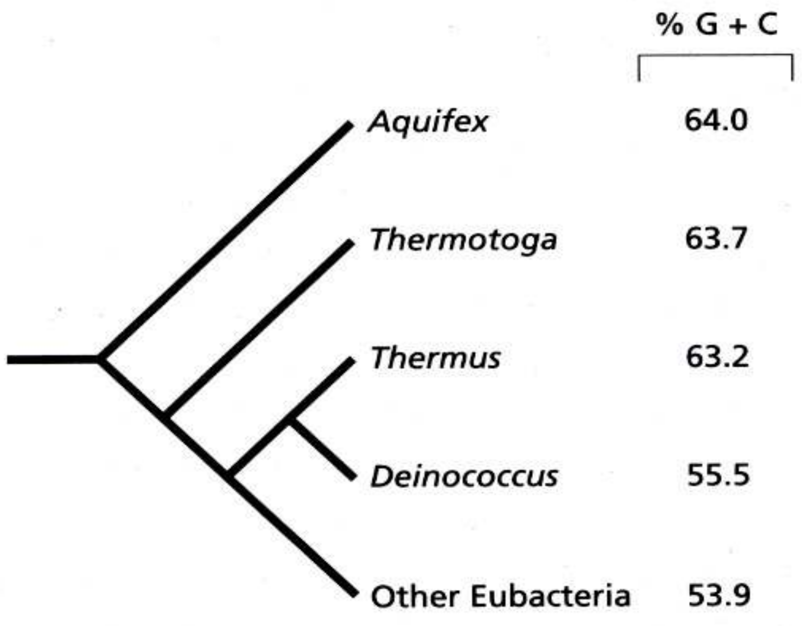




Modelos de Evolução de sequências

- **Felsenstein (1981) - F81**
- Na maioria das spp, A+T ≠ C+G
- Se houver pouco G’s, a taxa de substituição de G para A será menor do que outras substituições

Organism	A	T	C	G
Human	30.9	29.4	19.9	19.8
Chicken	28.8	29.2	20.5	21.7
Grasshopper	29.3	29.3	20.5	20.7
Sea Urchin	32.8	32.1	17.7	17.3
Wheat	27.3	27.1	22.7	22.8
Yeast	31.3	32.9	18.7	17.1
E. Coli	24.7	23.6	26.0	25.7

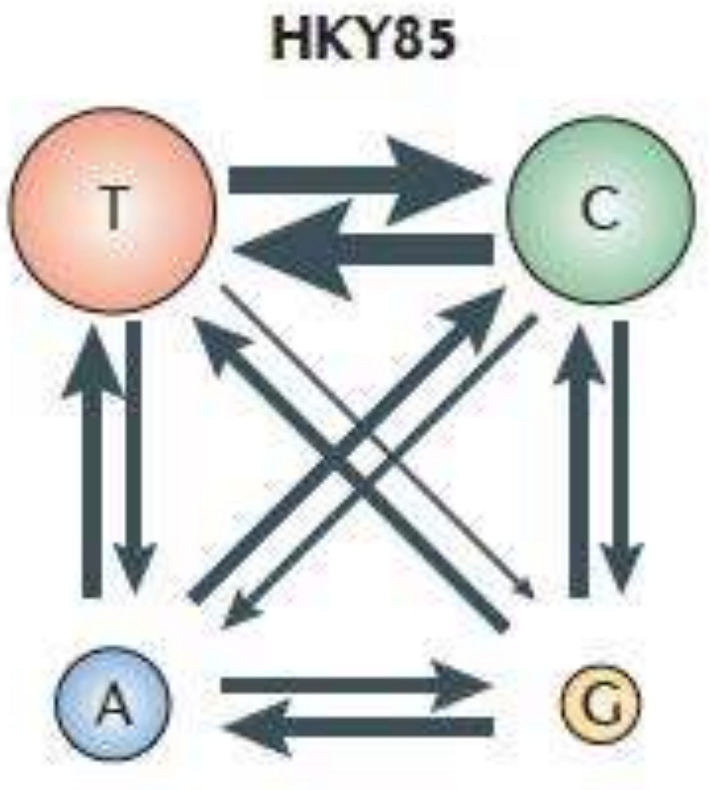


$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C\alpha & \pi_G\alpha & \pi_T\alpha \\ \pi_A\alpha & . & \pi_G\alpha & \pi_T\alpha \\ \pi_A\alpha & \pi_C\alpha & . & \pi_T\alpha \\ \pi_A\alpha & \pi_C\alpha & \pi_G\alpha & . \end{bmatrix}$$

Modelos de Evolução de sequências

- **Hasegawa, Kishino and Yano (HKY85)**
- Combina F81 eK2P

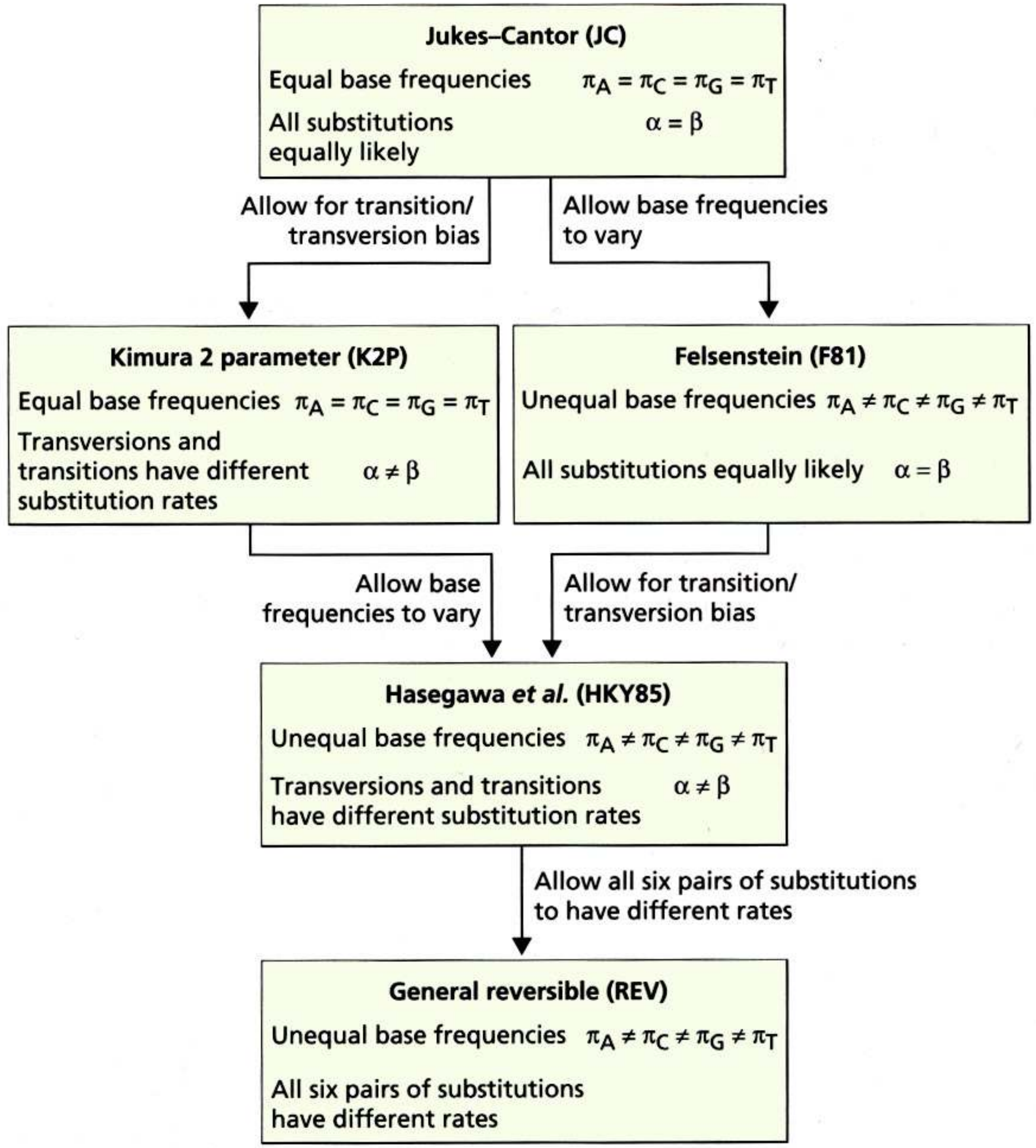
$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C\beta & \pi_G\alpha & \pi_T\beta \\ \pi_A\beta & . & \pi_G\beta & \pi_T\alpha \\ \pi_A\alpha & \pi_C\beta & . & \pi_T\beta \\ \pi_A\beta & \pi_C\alpha & \pi_G\beta & . \end{bmatrix}$$



- **General Time Reversible (GTR)**
- Permite que todos os pares de substituição tenham taxas distintas
- Permite frequências de nucleotídeos diferentes

$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C a & \pi_G b & \pi_T c \\ \pi_A a & . & \pi_G d & \pi_T e \\ \pi_A b & \pi_C d & . & \pi_T f \\ \pi_A c & \pi_C e & \pi_G f & . \end{bmatrix}$$

Modelos de Evolução de sequências



Jukes–Cantor (JC)

Equal base frequencies $\pi_A = \pi_C = \pi_G = \pi_T$

All substitutions equally likely $\alpha = \beta$

Allow for transition/ transversion bias

Allow base frequencies to vary

Kimura 2 parameter (K2P)

Equal base frequencies $\pi_A = \pi_C = \pi_G = \pi_T$

Transversions and transitions have different substitution rates $\alpha \neq \beta$

Felsenstein (F81)

Unequal base frequencies $\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$

All substitutions equally likely $\alpha = \beta$

Allow base frequencies to vary

Allow for transition/ transversion bias

Hasegawa *et al.* (HKY85)

Unequal base frequencies $\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$

Transversions and transitions have different substitution rates $\alpha \neq \beta$

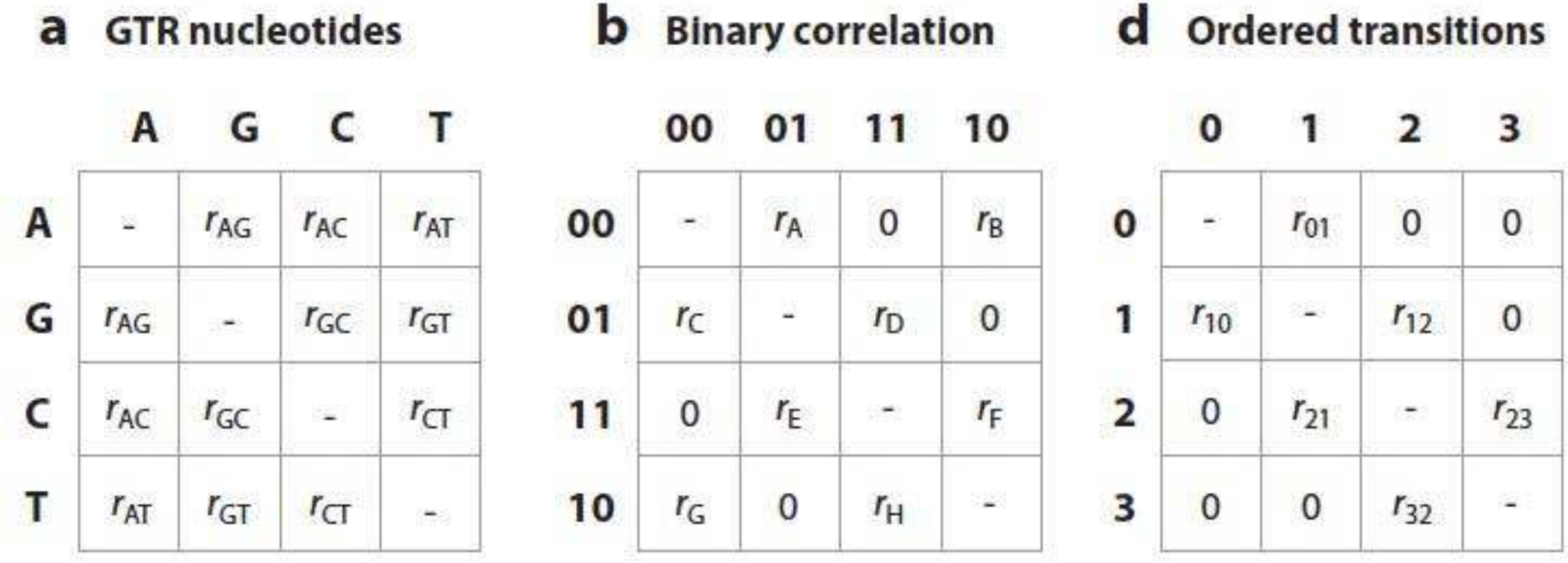
Allow all six pairs of substitutions to have different rates

General reversible (REV)

Unequal base frequencies $\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$

All six pairs of substitutions have different rates

Mas porque isso é importante?



a GTR nucleotides

	A	G	C	T
A	-	r_{AG}	r_{AC}	r_{AT}
G	r_{AG}	-	r_{GC}	r_{GT}
C	r_{AC}	r_{GC}	-	r_{CT}
T	r_{AT}	r_{GT}	r_{CT}	-

b Binary correlation

	00	01	11	10
00	-	r_A	0	r_B
01	r_C	-	r_D	0
11	0	r_E	-	r_F
10	r_G	0	r_H	-

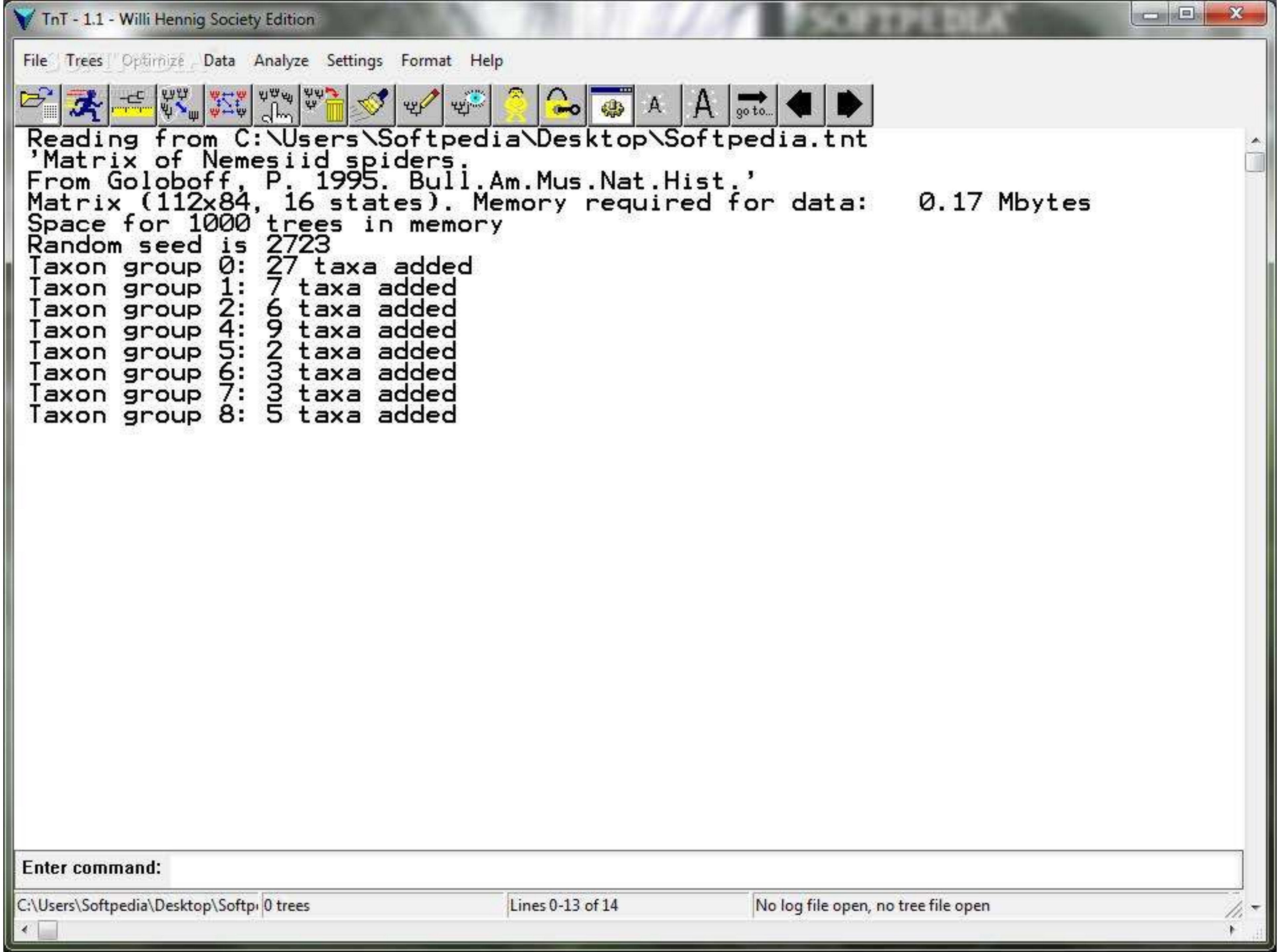
d Ordered transitions

	0	1	2	3
0	-	r_{01}	0	0
1	r_{10}	-	r_{12}	0
2	0	r_{21}	-	r_{23}
3	0	0	r_{32}	-

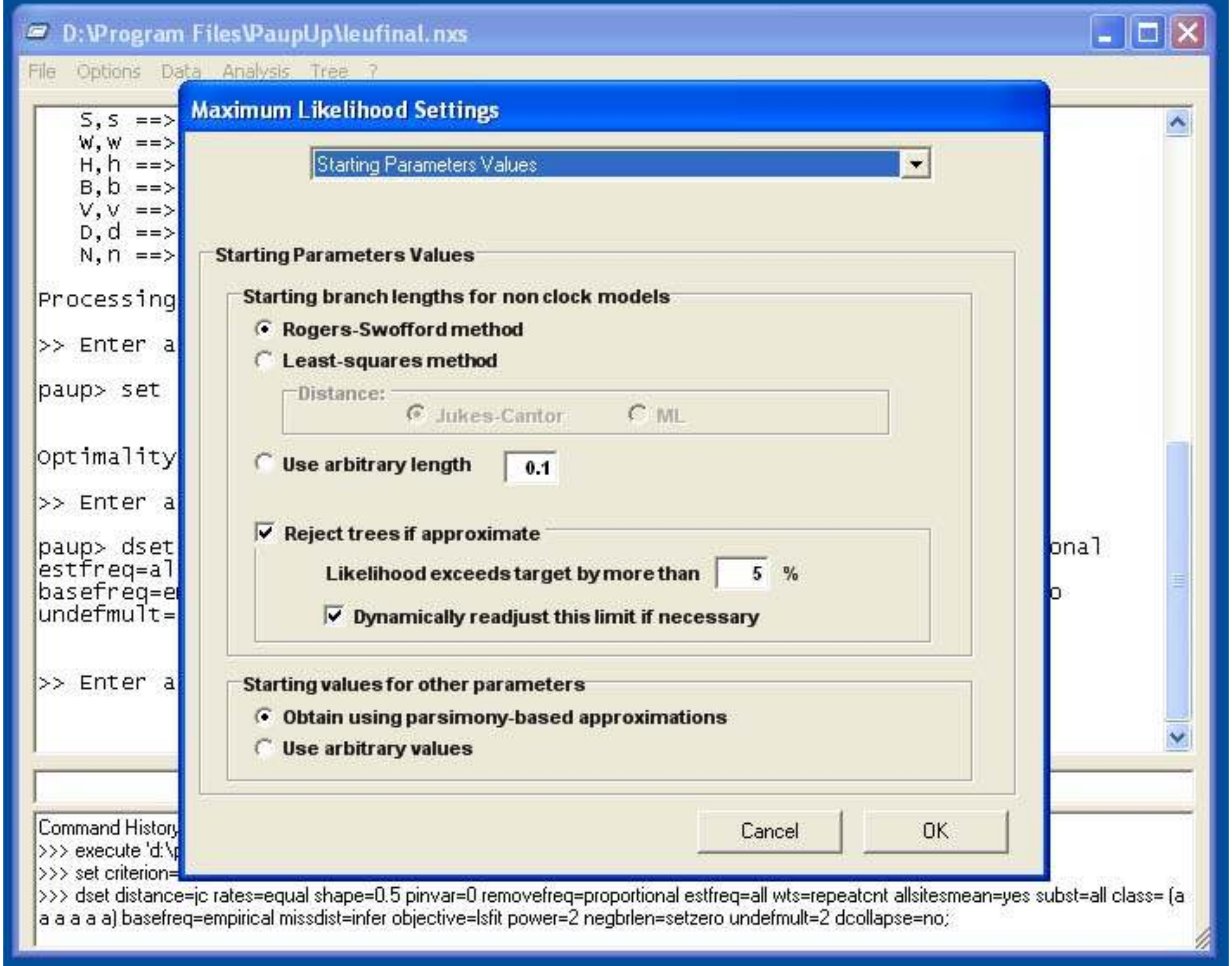
Além de serem necessários em métodos de inferência filogenética ”model-based”, esses modelos de substituição utilizam o mesmo arcabouço (processos de Markov) que métodos comparativos que iremos estudar, tais como correlação de caracteres binários e multiestado

O’Meara 2012

Softwares para inferência filogenética



```
Reading from C:\Users\Softpedia\Desktop\Softpedia.tnt
'Matrix of Nemesiid spiders.
From Goloboff, P. 1995. Bull.Am.Mus.Nat.Hist.'
Matrix (112x84, 16 states). Memory required for data:   0.17 Mbytes
Space for 1000 trees in memory
Random seed is 2723
Taxon group 0: 27 taxa added
Taxon group 1: 7 taxa added
Taxon group 2: 6 taxa added
Taxon group 4: 9 taxa added
Taxon group 5: 2 taxa added
Taxon group 6: 3 taxa added
Taxon group 7: 3 taxa added
Taxon group 8: 5 taxa added
```



```
    S,s ==>
    W,w ==>
    H,h ==>
    B,b ==>
    V,v ==>
    D,d ==>
    N,n ==>

Processing

>> Enter a

paup> set


Optimality

>> Enter a

paup> dset                                                      onal
estfreq=al
basefreq=e                                                      o
undefmult=


>> Enter a
```

Maximum Likelihood Settings

Starting Parameters Values ▾

Starting Parameters Values

Starting branch lengths for non clock models

- ● Rogers-Swofford method
- ○ Least-squares method
  - Distance: ● Jukes-Cantor ○ ML
- ○ Use arbitrary length 0.1

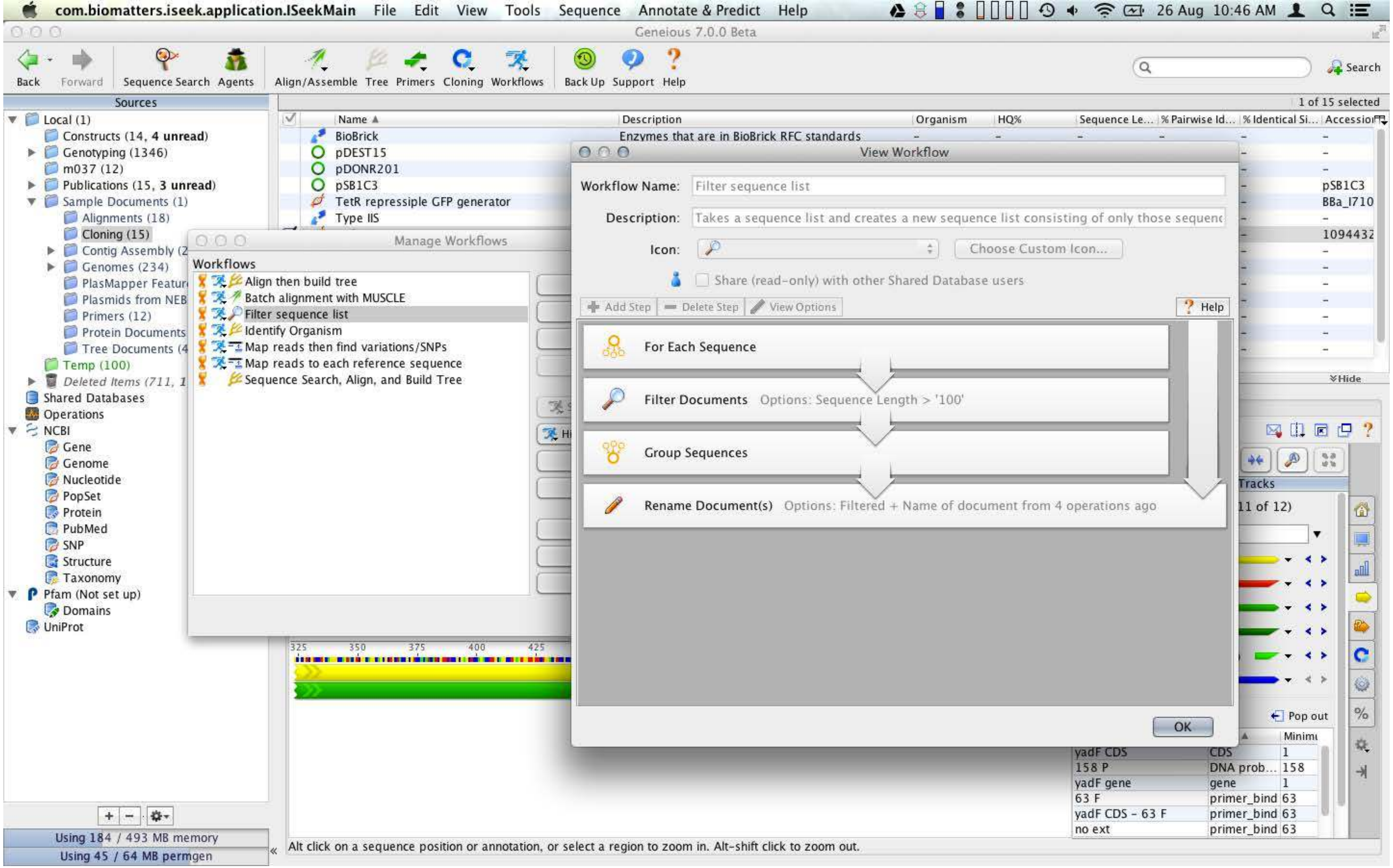
☑ Reject trees if approximate

- Likelihood exceeds target by more than 5 %
- ☑ Dynamically readjust this limit if necessary

Starting values for other parameters

- ● Obtain using parsimony-based approximations
- ○ Use arbitrary values

```
Command History
>>> execute 'd:\p
>>> set criterion=
>>> dset distance=jc rates=equal shape=0.5 pinvar=0 removefreq=proportional estfreq=all wts=repeatcnt allsitesmean=yes subst=all class= (a
a a a a a) basefreq=empirical missdist=infer objective=lsfit power=2 negbrlen=setzero undefmult=2 dcollapse=no;
```



Name	Description	Accession
BioBrick	Enzymes that are in BioBrick RFC standards	-
pDEST15		-
pDONR201		-
pSB1C3		pSB1C3
TetR repressiple GFP generator		BBa_I710
Type IIS		-

Workflows:
- Align then build tree
- Batch alignment with MUSCLE
- Filter sequence list
- Identify Organism
- Map reads then find variations/SNPs
- Map reads to each reference sequence
- Sequence Search, Align, and Build Tree

Workflow Name: Filter sequence list

Description: Takes a sequence list and creates a new sequence list consisting of only those sequenc

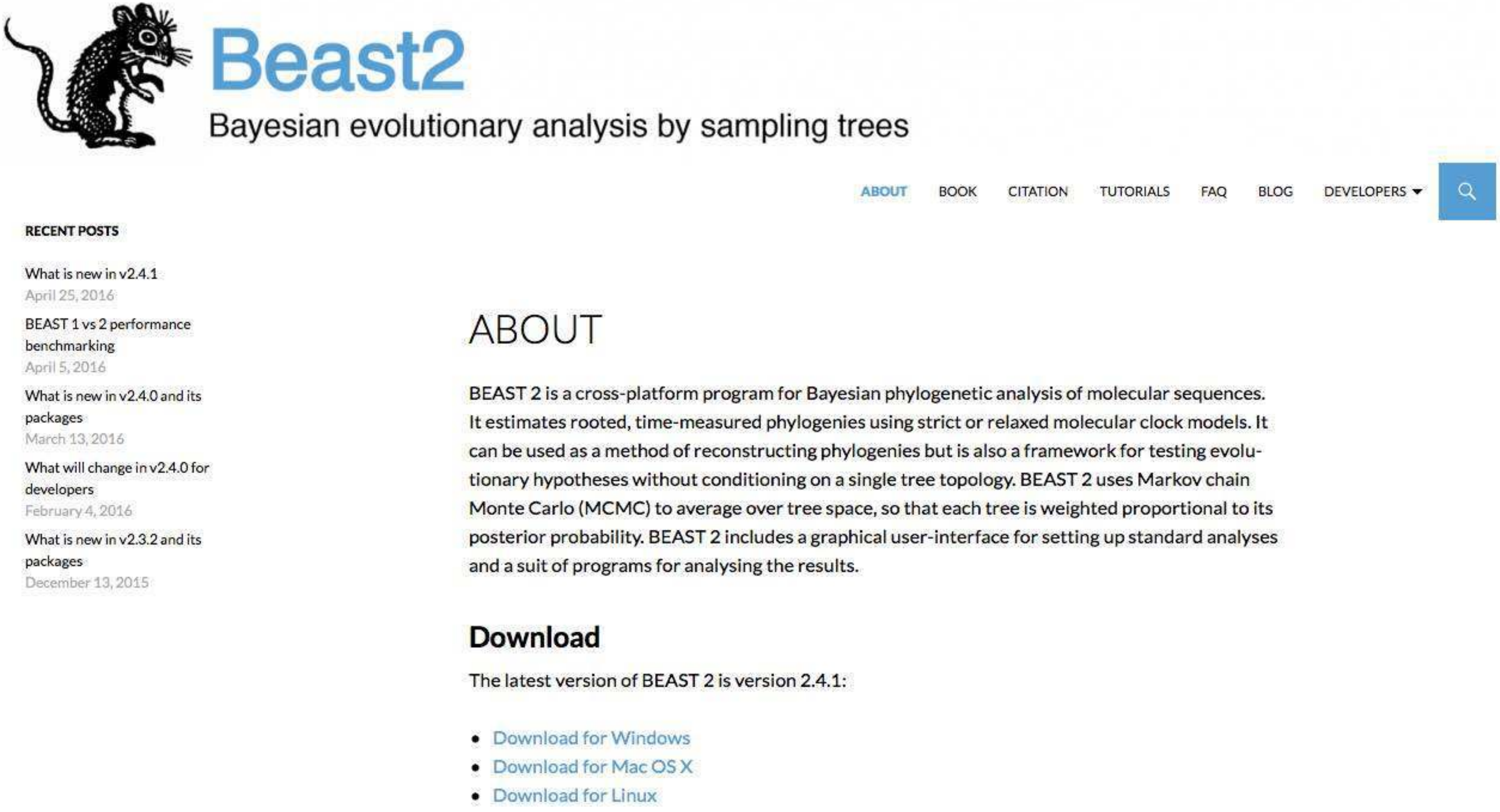
☐ Share (read-only) with other Shared Database users

1. For Each Sequence
2. Filter Documents Options: Sequence Length > '100'
3. Group Sequences
4. Rename Document(s) Options: Filtered + Name of document from 4 operations ago

yadF CDS	CDS	1
158 P	DNA prob...	158
yadF gene	gene	1
63 F	primer_bind	63
yadF CDS - 63 F	primer_bind	63
no ext	primer_bind	63

Alt click on a sequence position or annotation, or select a region to zoom in. Alt-shift click to zoom out.

http://www.geneious.com



Beast2

Bayesian evolutionary analysis by sampling trees

RECENT POSTS

What is new in v2.4.1
April 25, 2016

BEAST 1 vs 2 performance benchmarking
April 5, 2016

What is new in v2.4.0 and its packages
March 13, 2016

What will change in v2.4.0 for developers
February 4, 2016

What is new in v2.3.2 and its packages
December 13, 2015

ABOUT

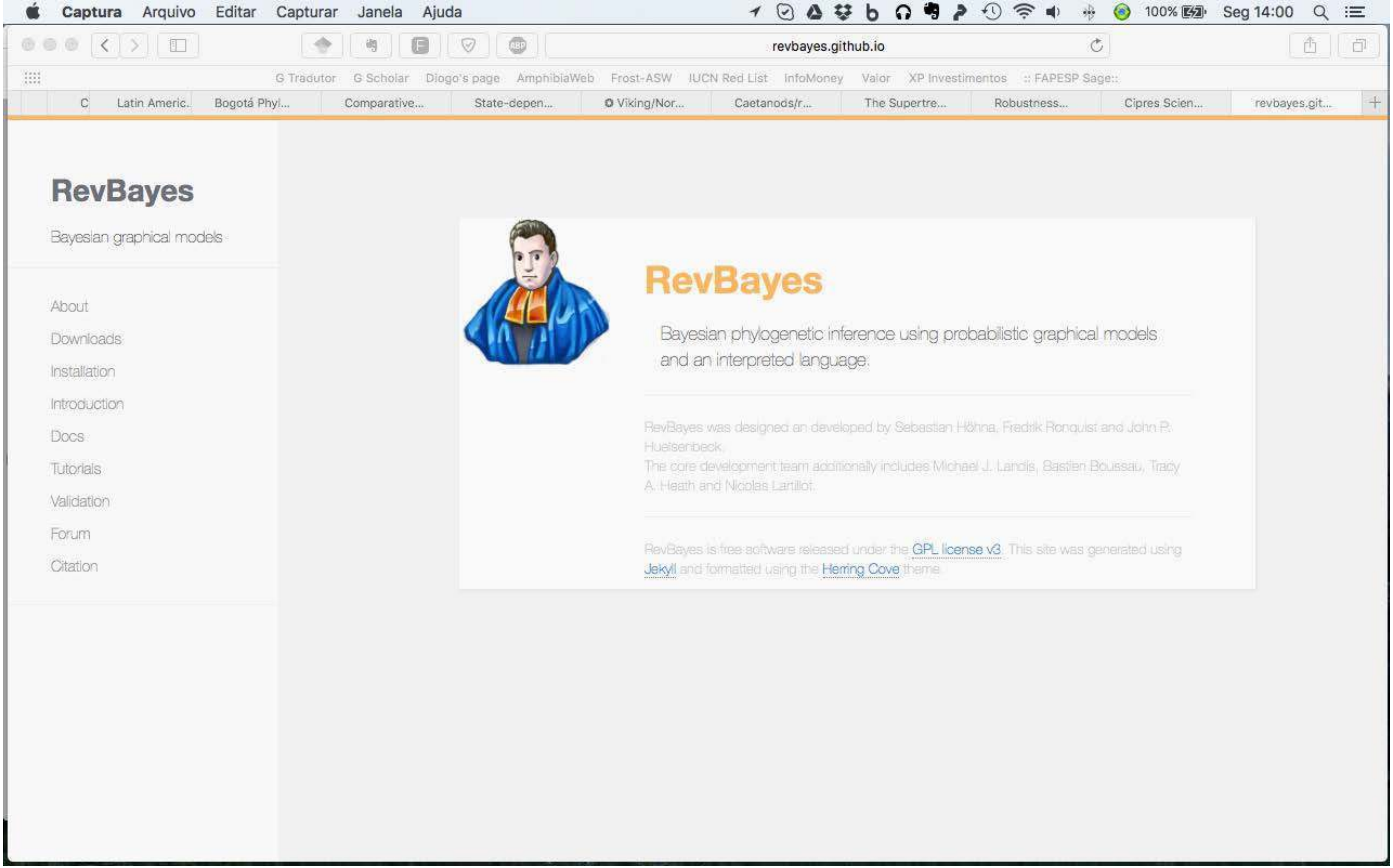
BEAST 2 is a cross-platform program for Bayesian phylogenetic analysis of molecular sequences. It estimates rooted, time-measured phylogenies using strict or relaxed molecular clock models. It can be used as a method of reconstructing phylogenies but is also a framework for testing evolutionary hypotheses without conditioning on a single tree topology. BEAST 2 uses Markov chain Monte Carlo (MCMC) to average over tree space, so that each tree is weighted proportional to its posterior probability. BEAST 2 includes a graphical user-interface for setting up standard analyses and a suit of programs for analysing the results.

Download

The latest version of BEAST 2 is version 2.4.1:

- Download for Windows
- Download for Mac OS X
- Download for Linux

http://www.beast2.org



RevBayes

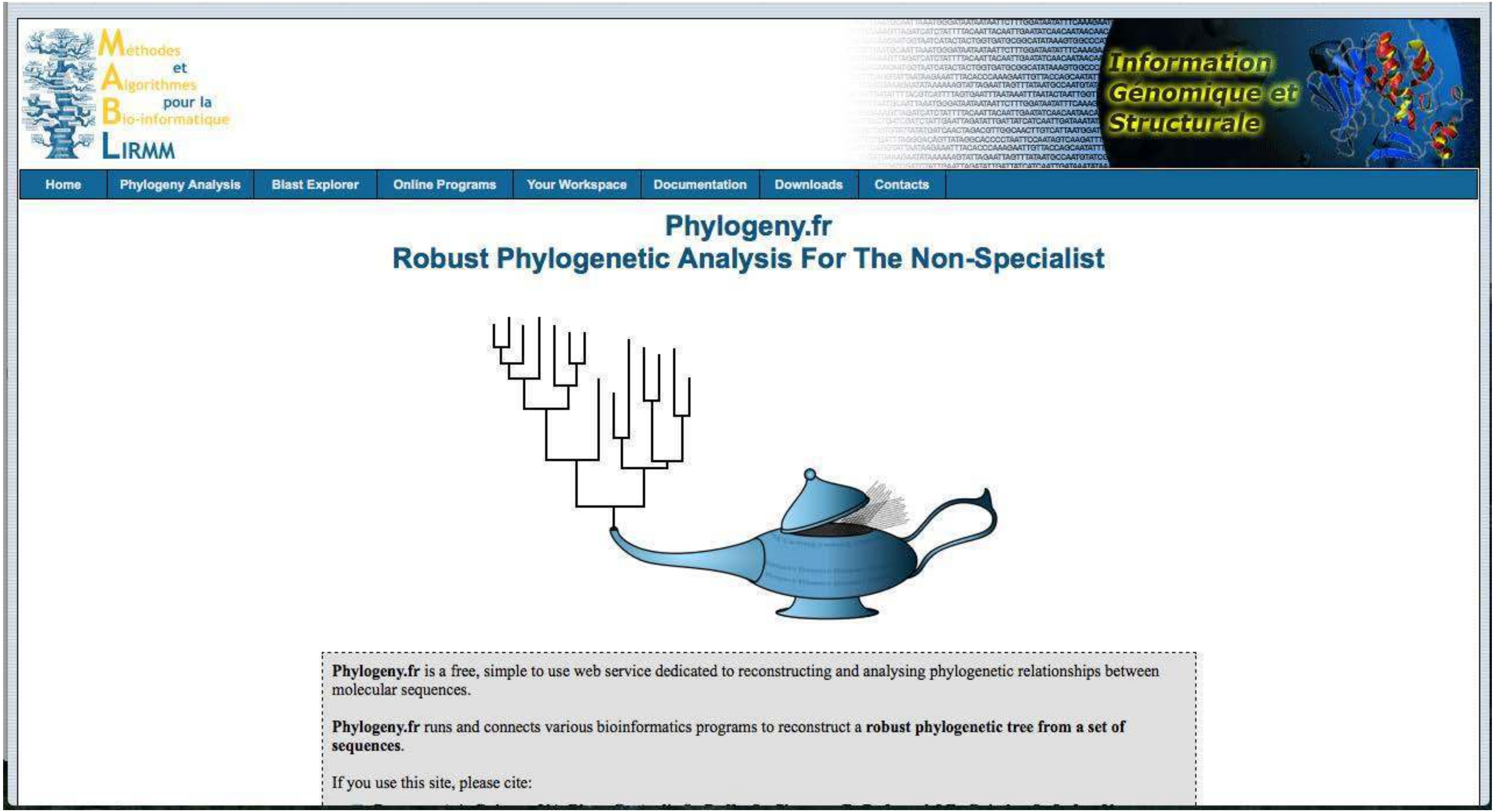
Bayesian graphical models

RevBayes

Bayesian phylogenetic inference using probabilistic graphical models and an interpreted language.

RevBayes was designed an developed by Sebastian Höhna, Fredrik Ronquist and John P. Huelsenbeck.
The core development team additionally includes Michael J. Landis, Bastien Boussau, Tracy A. Heath and Nicolas Lartillot.

RevBayes is free software released under the GPL license v3. This site was generated using Jekyll and formatted using the Herring Cove theme.

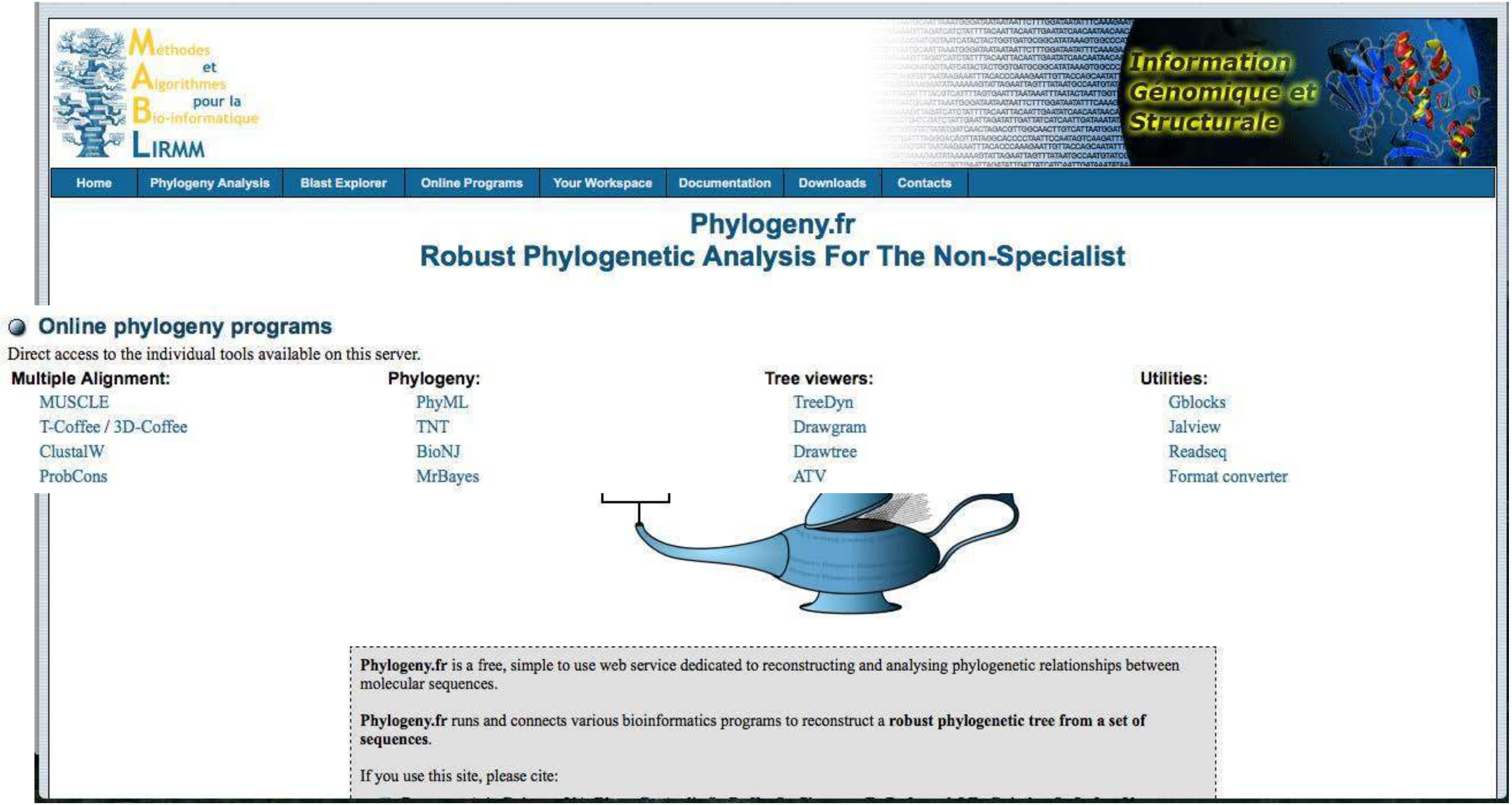


Phylogeny.fr
Robust Phylogenetic Analysis For The Non-Specialist

Phylogeny.fr is a free, simple to use web service dedicated to reconstructing and analysing phylogenetic relationships between molecular sequences.

Phylogeny.fr runs and connects various bioinformatics programs to reconstruct a **robust phylogenetic tree from a set of sequences**.

If you use this site, please cite:



Phylogeny.fr
Robust Phylogenetic Analysis For The Non-Specialist

Online phylogeny programs

Direct access to the individual tools available on this server.

Multiple Alignment:
MUSCLE
T-Coffee / 3D-Coffee
ClustalW
ProbCons

Phylogeny:
PhyML
TNT
BioNJ
MrBayes

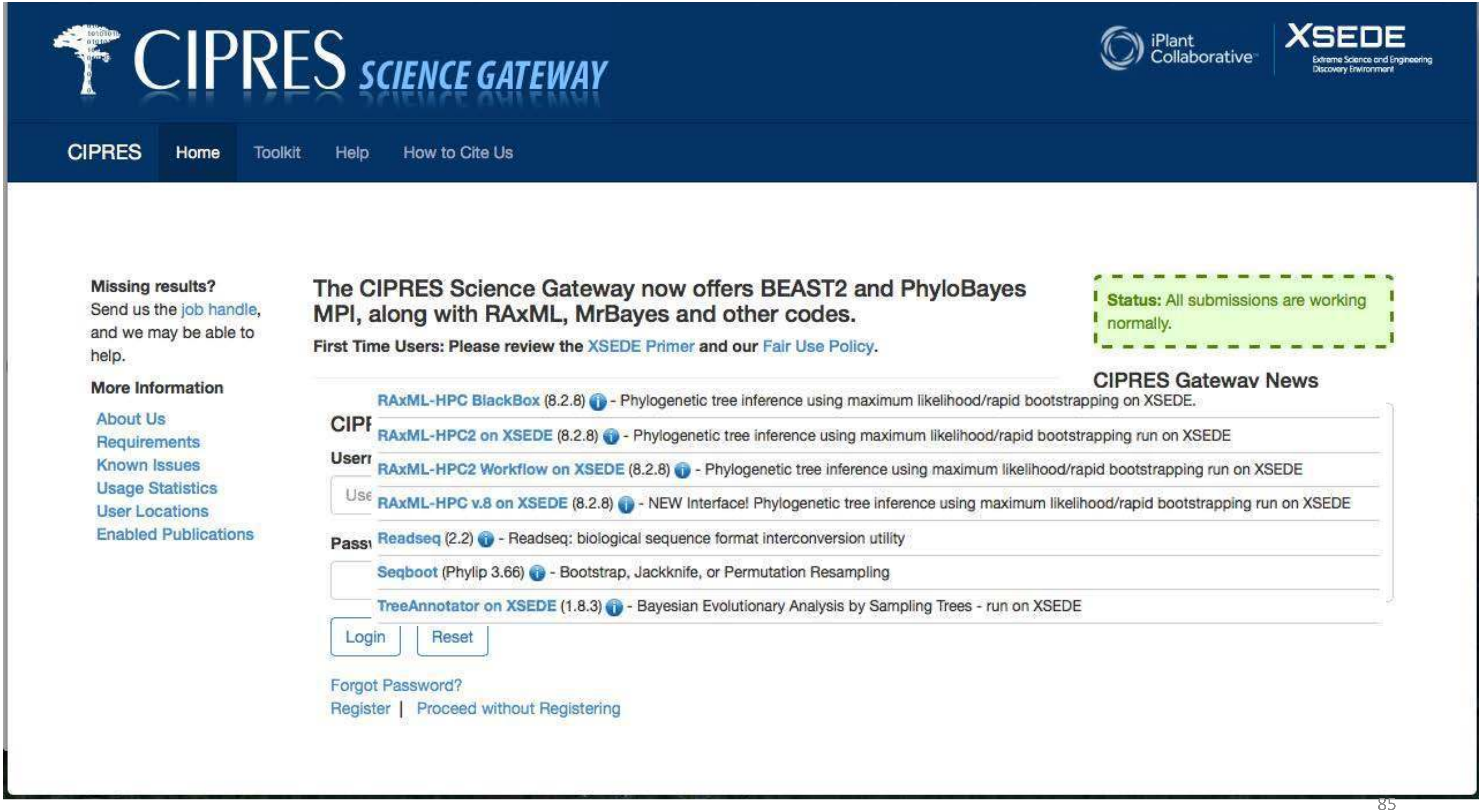
Tree viewers:
TreeDyn
Drawgram
Drawtree
ATV

Utilities:
Gblocks
Jalview
Readseq
Format converter

Phylogeny.fr is a free, simple to use web service dedicated to reconstructing and analysing phylogenetic relationships between molecular sequences.

Phylogeny.fr runs and connects various bioinformatics programs to reconstruct a **robust phylogenetic tree from a set of sequences.**

If you use this site, please cite:



Missing results?
Send us the job handle, and we may be able to help.

More Information

- About Us
- Requirements
- Known Issues
- Usage Statistics
- User Locations
- Enabled Publications

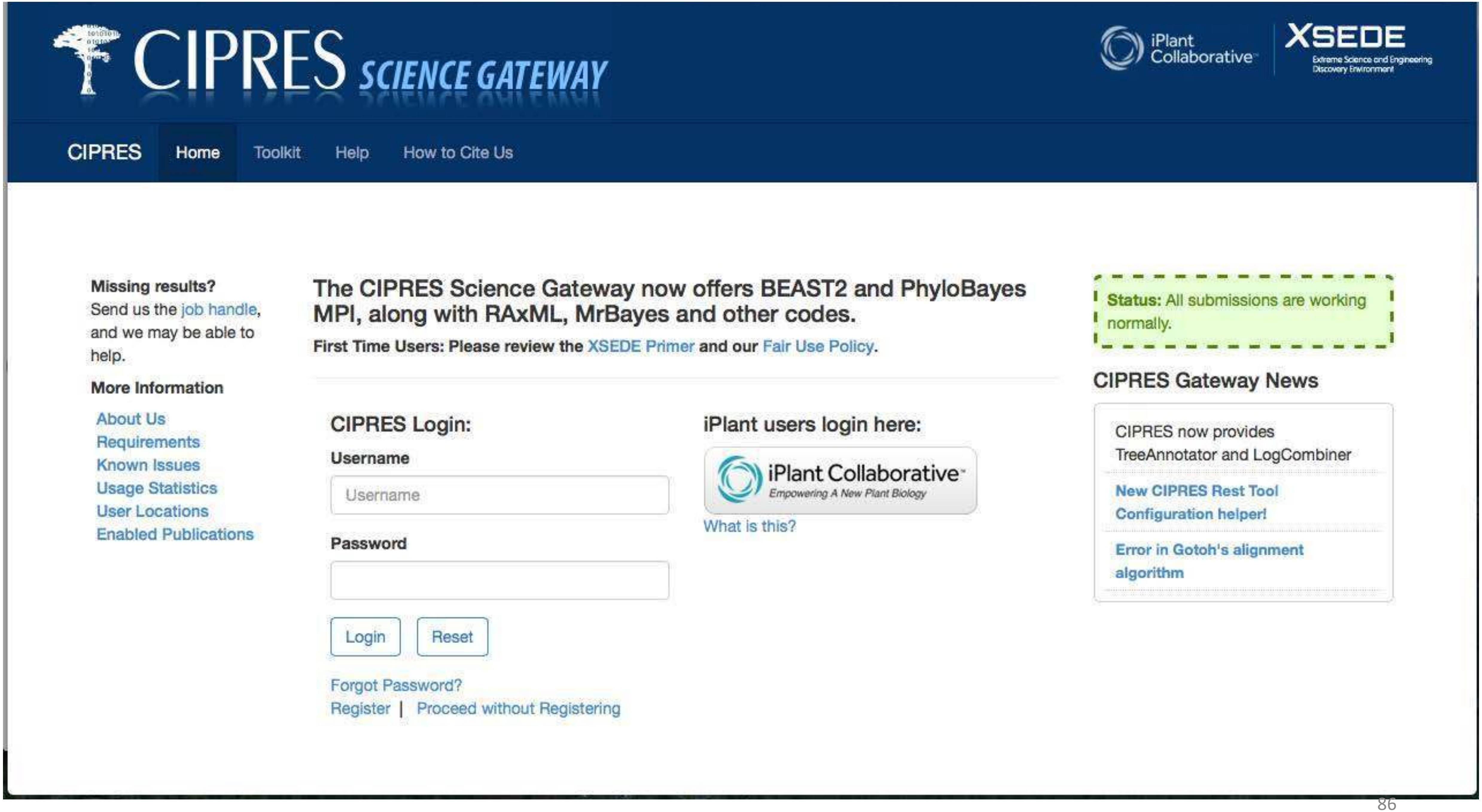
The CIPRES Science Gateway now offers BEAST2 and PhyloBayes MPI, along with RAxML, MrBayes and other codes.

First Time Users: Please review the XSEDE Primer and our Fair Use Policy.

Status: All submissions are working normally.

CIPRES Gateway News

- **RAxML-HPC BlackBox** (8.2.8) ⓘ - Phylogenetic tree inference using maximum likelihood/rapid bootstrapping on XSEDE.
- **RAxML-HPC2 on XSEDE** (8.2.8) ⓘ - Phylogenetic tree inference using maximum likelihood/rapid bootstrapping run on XSEDE
- **RAxML-HPC2 Workflow on XSEDE** (8.2.8) ⓘ - Phylogenetic tree inference using maximum likelihood/rapid bootstrapping run on XSEDE
- **RAxML-HPC v.8 on XSEDE** (8.2.8) ⓘ - NEW Interface! Phylogenetic tree inference using maximum likelihood/rapid bootstrapping run on XSEDE
- **Readseq** (2.2) ⓘ - Readseq: biological sequence format interconversion utility
- **Seqboot** (Phylip 3.66) ⓘ - Bootstrap, Jackknife, or Permutation Resampling
- **TreeAnnotator on XSEDE** (1.8.3) ⓘ - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE



Missing results?
Send us the job handle, and we may be able to help.

More Information

- About Us
- Requirements
- Known Issues
- Usage Statistics
- User Locations
- Enabled Publications

The CIPRES Science Gateway now offers BEAST2 and PhyloBayes MPI, along with RAxML, MrBayes and other codes.

First Time Users: Please review the XSEDE Primer and our Fair Use Policy.

CIPRES Login:

Username

Username

Password

Forgot Password?
Register | Proceed without Registering

iPlant users login here:

What is this?

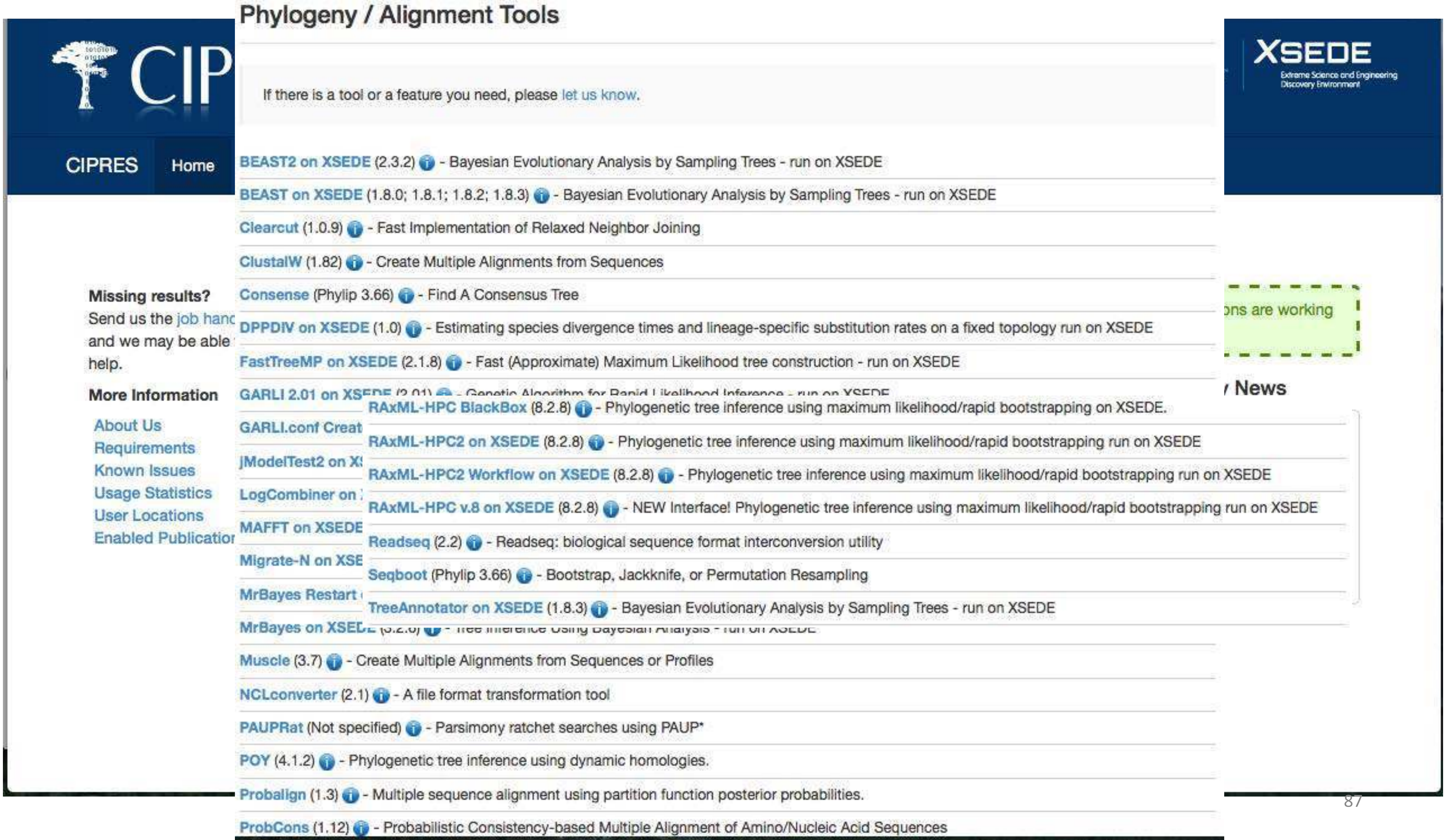
Status: All submissions are working normally.

CIPRES Gateway News

CIPRES now provides TreeAnnotator and LogCombiner

New CIPRES Rest Tool Configuration helper!

Error in Gotoh's alignment algorithm



Missing results?
Send us the job hanc
and we may be able
help.

More Information

- About Us
- Requirements
- Known Issues
- Usage Statistics
- User Locations
- Enabled Publicatior

ons are working

News

Phylogeny / Alignment Tools

If there is a tool or a feature you need, please let us know.

BEAST2 on XSEDE (2.3.2) - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

BEAST on XSEDE (1.8.0; 1.8.1; 1.8.2; 1.8.3) - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

Clearcut (1.0.9) - Fast Implementation of Relaxed Neighbor Joining

ClustalW (1.82) - Create Multiple Alignments from Sequences

Consense (Phylip 3.66) - Find A Consensus Tree

DPPDIV on XSEDE (1.0) - Estimating species divergence times and lineage-specific substitution rates on a fixed topology run on XSEDE

FastTreeMP on XSEDE (2.1.8) - Fast (Approximate) Maximum Likelihood tree construction - run on XSEDE

GARLI 2.01 on XSEDE (2.01) - Genetic Algorithm for Rapid Likelihood Inference - run on XSEDE

GARLI.conf Creat

jModelTest2 on X

LogCombiner on

MAFFT on XSEDE

Migrate-N on XSE

MrBayes Restart

MrBayes on XSEDE (3.2.6) - Tree Inference Using Bayesian Analysis - run on XSEDE

Muscle (3.7) - Create Multiple Alignments from Sequences or Profiles

NCLconverter (2.1) - A file format transformation tool

PAUPRat (Not specified) - Parsimony ratchet searches using PAUP*

POY (4.1.2) - Phylogenetic tree inference using dynamic homologies.

Probalign (1.3) - Multiple sequence alignment using partition function posterior probabilities.

ProbCons (1.12) - Probabilistic Consistency-based Multiple Alignment of Amino/Nucleic Acid Sequences

RAxML-HPC BlackBox (8.2.8) - Phylogenetic tree inference using maximum likelihood/rapid bootstrapping on XSEDE.

RAxML-HPC2 on XSEDE (8.2.8) - Phylogenetic tree inference using maximum likelihood/rapid bootstrapping run on XSEDE

RAxML-HPC2 Workflow on XSEDE (8.2.8) - Phylogenetic tree inference using maximum likelihood/rapid bootstrapping run on XSEDE

RAxML-HPC v.8 on XSEDE (8.2.8) - NEW Interface! Phylogenetic tree inference using maximum likelihood/rapid bootstrapping run on XSEDE

Readseq (2.2) - Readseq: biological sequence format interconversion utility

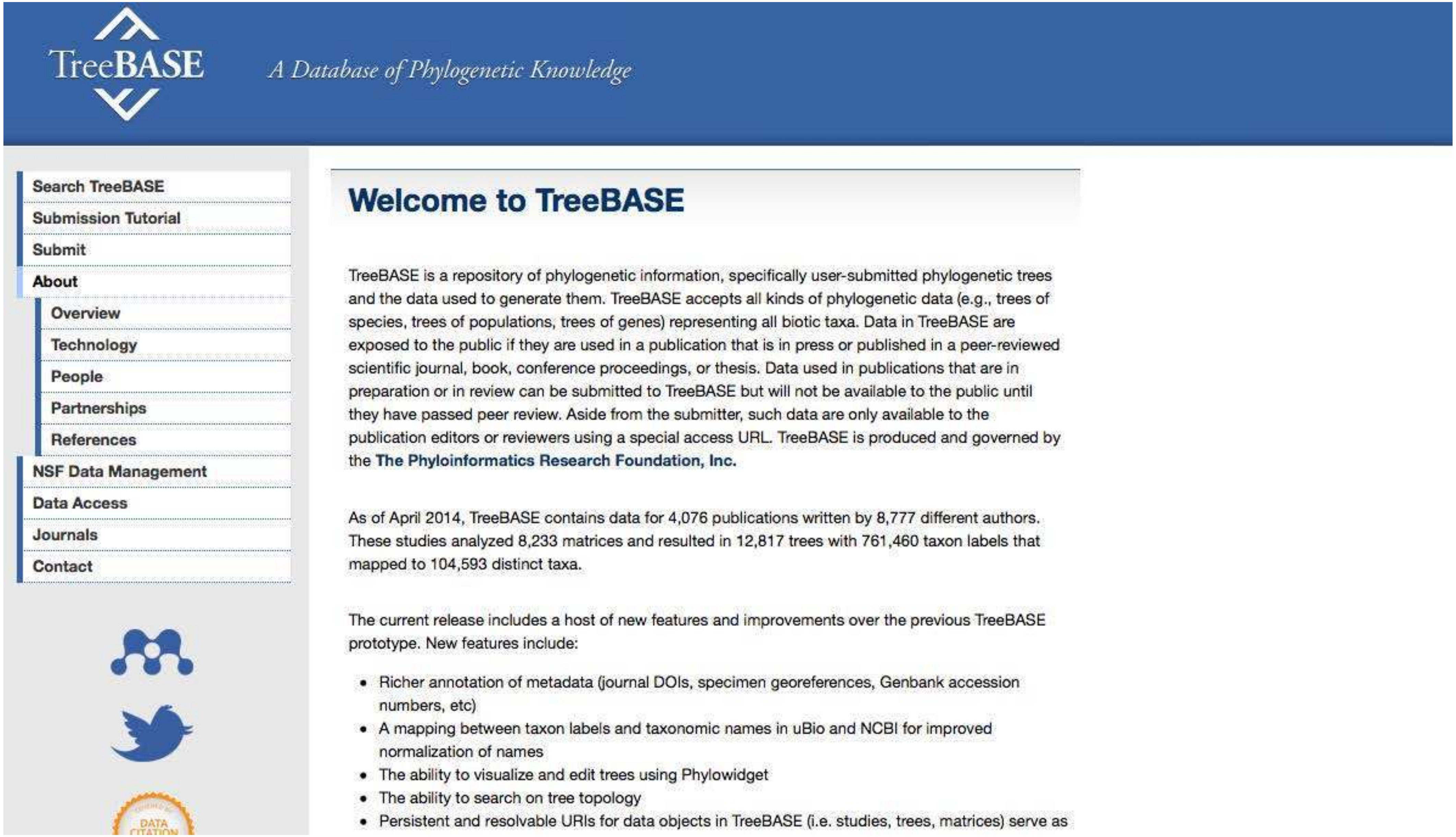
Seqboot (Phylip 3.66) - Bootstrap, Jackknife, or Permutation Resampling

TreeAnnotator on XSEDE (1.8.3) - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

Base de dados para filogenias: como obter uma árvore

- Bases de dados de topologias e tempo de divergência
- Abordagens megaphylogenies
  - Supermatrix
  - Supertree
  - Next-generation sequencing (high-throughput sequencing data)
- Pipelines (semi)automatizadas ”for dummies”

Bases de dados para topologias e tempo de divergência



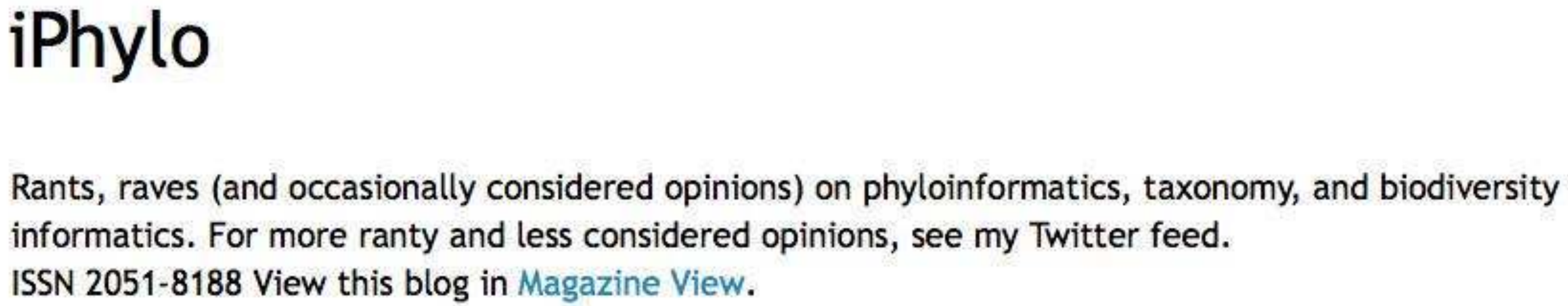
Welcome to TreeBASE

TreeBASE is a repository of phylogenetic information, specifically user-submitted phylogenetic trees and the data used to generate them. TreeBASE accepts all kinds of phylogenetic data (e.g., trees of species, trees of populations, trees of genes) representing all biotic taxa. Data in TreeBASE are exposed to the public if they are used in a publication that is in press or published in a peer-reviewed scientific journal, book, conference proceedings, or thesis. Data used in publications that are in preparation or in review can be submitted to TreeBASE but will not be available to the public until they have passed peer review. Aside from the submitter, such data are only available to the publication editors or reviewers using a special access URL. TreeBASE is produced and governed by the **The Phyloinformatics Research Foundation, Inc.**

As of April 2014, TreeBASE contains data for 4,076 publications written by 8,777 different authors. These studies analyzed 8,233 matrices and resulted in 12,817 trees with 761,460 taxon labels that mapped to 104,593 distinct taxa.

The current release includes a host of new features and improvements over the previous TreeBASE prototype. New features include:

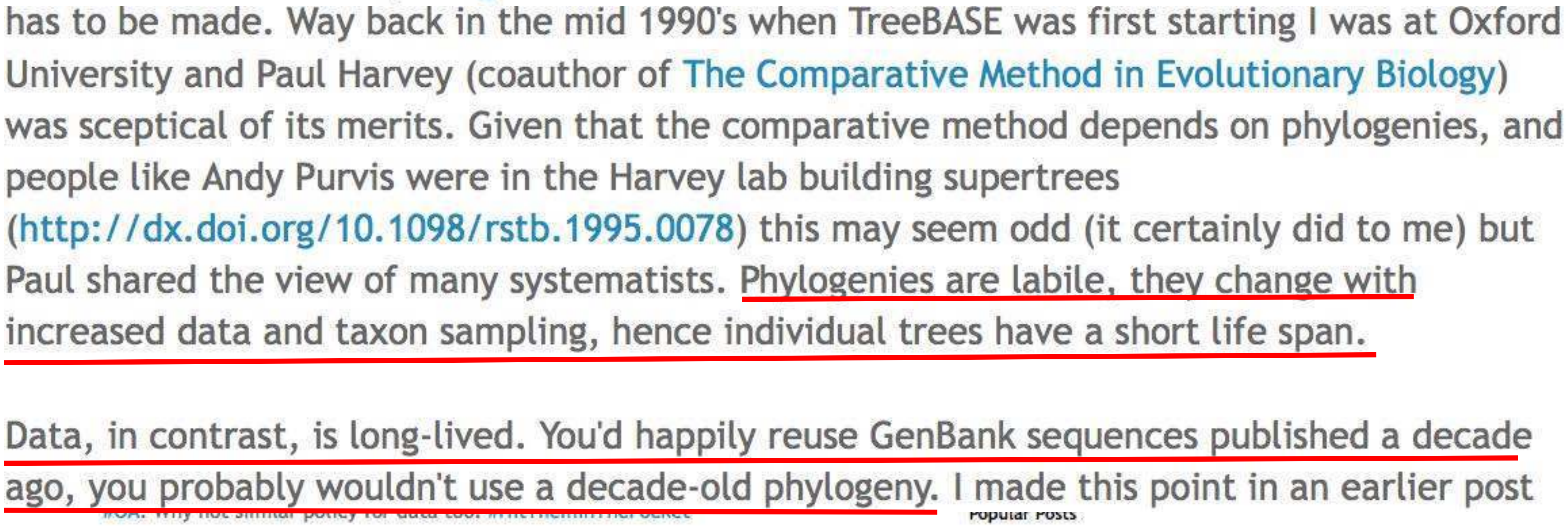
- Richer annotation of metadata (journal DOIs, specimen georeferences, Genbank accession numbers, etc)
- A mapping between taxon labels and taxonomic names in uBio and NCBI for improved normalization of names
- The ability to visualize and edit trees using Phylowidget
- The ability to search on tree topology
- Persistent and resolvable URIs for data objects in TreeBASE (i.e. studies, trees, matrices) serve as



iPhylo

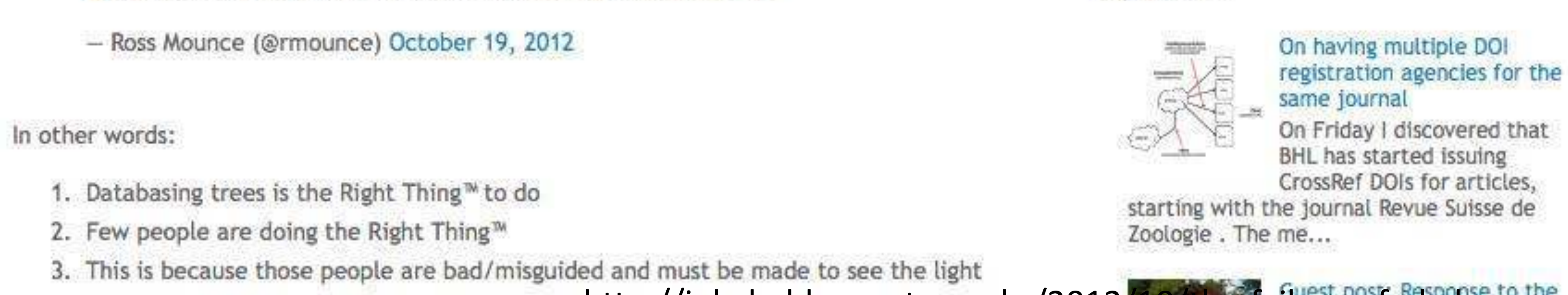
Rants, raves (and occasionally considered opinions) on phyloinformatics, taxonomy, and biodiversity informatics. For more ranty and less considered opinions, see my Twitter feed.

ISSN 2051-8188 View this blog in Magazine View.



has to be made. Way back in the mid 1990's when TreeBASE was first starting I was at Oxford University and Paul Harvey (coauthor of The Comparative Method in Evolutionary Biology) was sceptical of its merits. Given that the comparative method depends on phylogenies, and people like Andy Purvis were in the Harvey lab building supertrees (http://dx.doi.org/10.1098/rstb.1995.0078) this may seem odd (it certainly did to me) but Paul shared the view of many systematists. Phylogenies are labile, they change with increased data and taxon sampling, hence individual trees have a short life span.

Data, in contrast, is long-lived. You'd happily reuse GenBank sequences published a decade ago, you probably wouldn't use a decade-old phylogeny. I made this point in an earlier post



— Ross Mounce (@rmounce) October 19, 2012

In other words:

1. Databasing trees is the Right Thing™ to do
2. Few people are doing the Right Thing™
3. This is because those people are bad/misguided and must be made to see the light

On having multiple DOI registration agencies for the same journal

On Friday I discovered that BHL has started issuing CrossRef DOIs for articles, starting with the journal Revue Suisse de Zoologie . The me...

http://iphylo.blogspot.com.br/2012/10/the-failure-of-phylogeny-databases.html

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution doi: 10.1111/j.2041-210X.2012.00247.x

APPLICATION

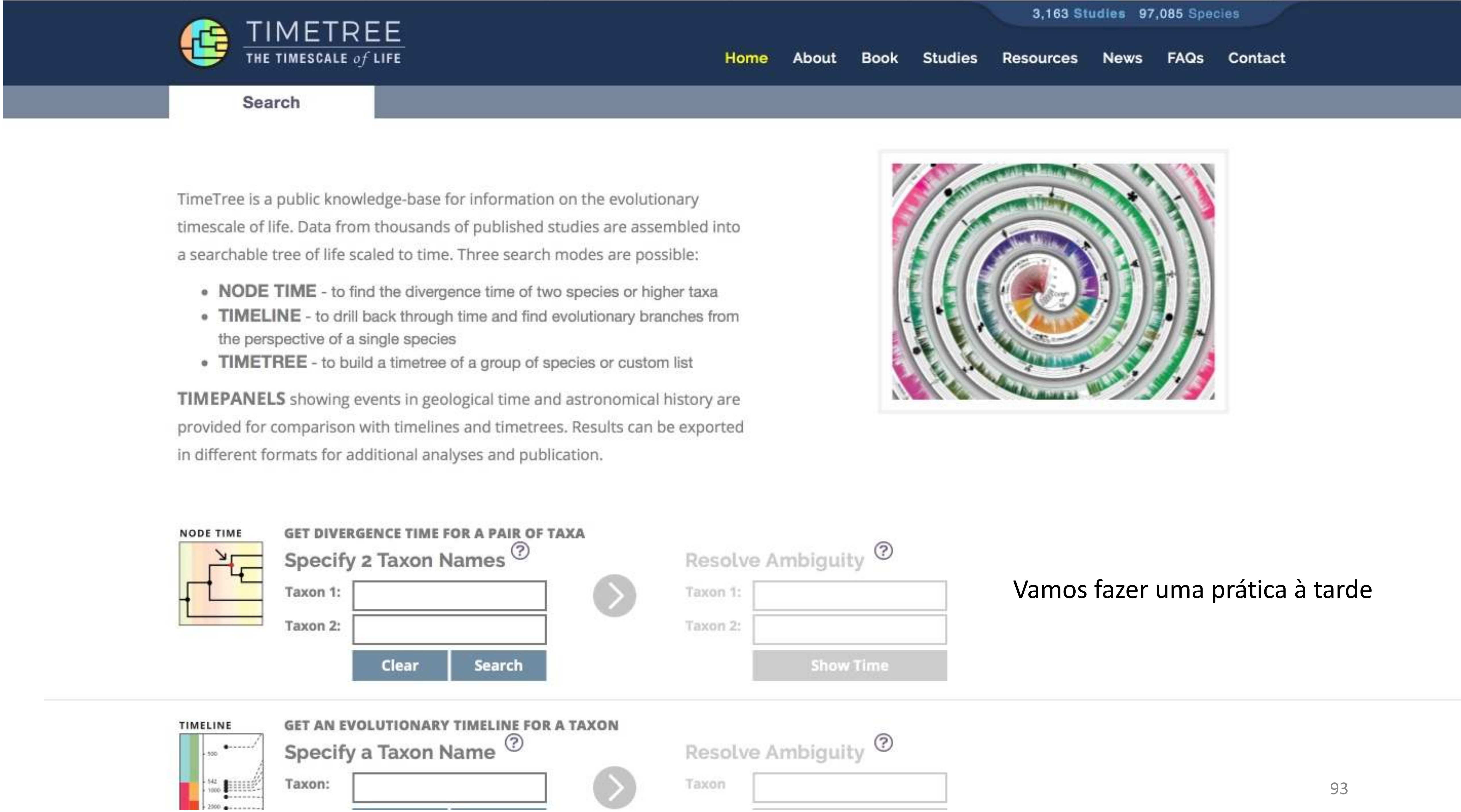
Treebase: an R package for discovery, access and manipulation of online phylogenies

Carl Boettiger[1]* and Duncan Temple Lang[2]

treebase tutorial

for v0.0-7.1

Neste post vamos ilustrar a fucionalidade do pacote com alguns exemplos introdutórios. Graças em parte às novas normas de periódicos como Evolution, Am Nat, e Sys Bio que requerem que os dados sejam depositados e planos de gerenciamento de dados requeridos pela NSF, esperamos que o pacote seja bastante útil para ensino na medida que replica os resultados ou para meta-análises que podem ser atualizadas automaticamente à medida que o repositório cresça. Informações adicionais e reletórios de erro são bem-invos via treebase page.



TimeTree is a public knowledge-base for information on the evolutionary timescale of life. Data from thousands of published studies are assembled into a searchable tree of life scaled to time. Three search modes are possible:

- **NODE TIME** - to find the divergence time of two species or higher taxa
- **TIMELINE** - to drill back through time and find evolutionary branches from the perspective of a single species
- **TIMETREE** - to build a timetree of a group of species or custom list

TIMEPANELS showing events in geological time and astronomical history are provided for comparison with timelines and timetrees. Results can be exported in different formats for additional analyses and publication.

GET DIVERGENCE TIME FOR A PAIR OF TAXA

Specify 2 Taxon Names

Taxon 1:

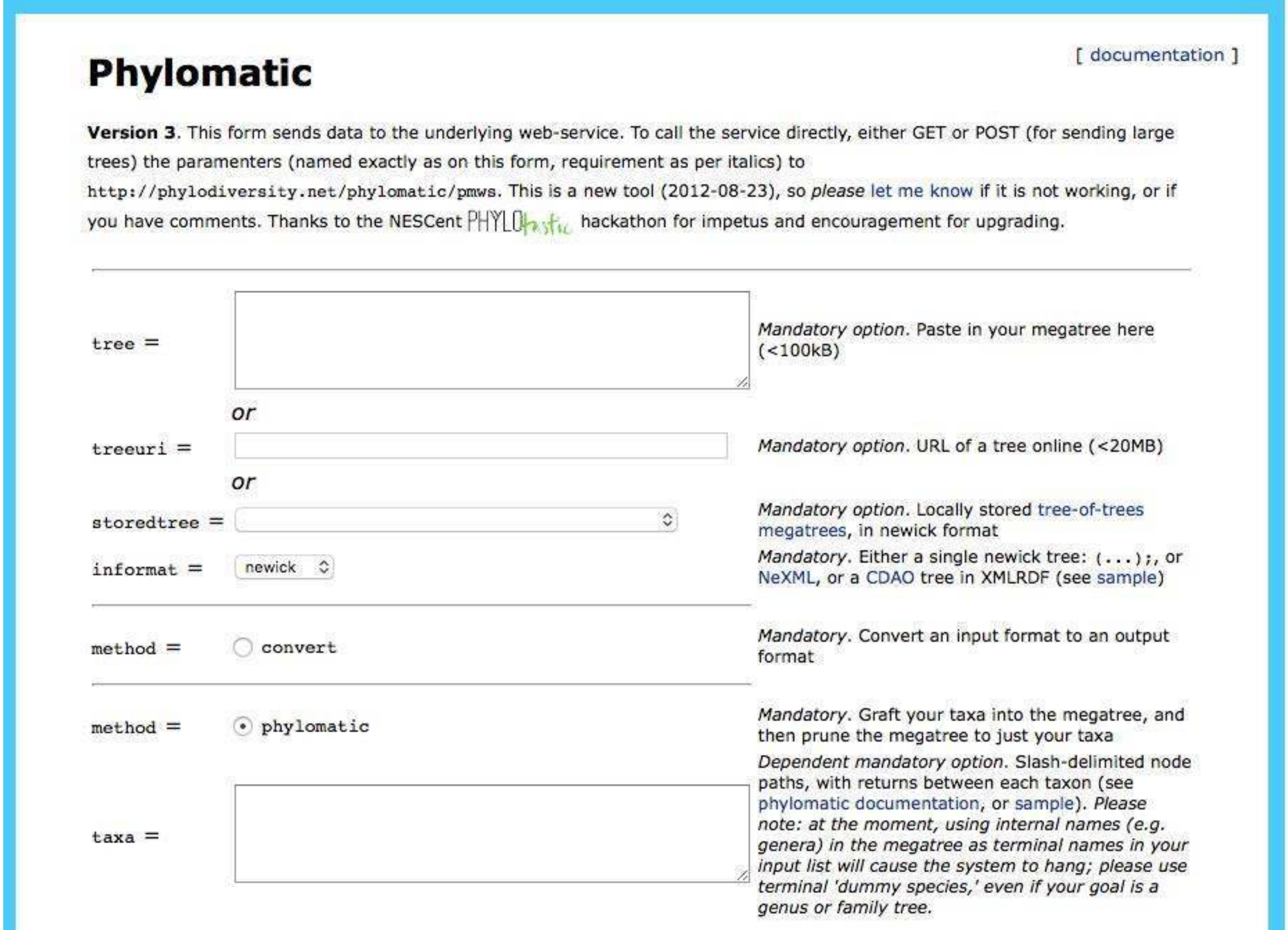
Taxon 2:

GET AN EVOLUTIONARY TIMELINE FOR A TAXON

Specify a Taxon Name

Taxon:

Vamos fazer uma prática à tarde



[documentation]

Phylomatic

Version 3. This form sends data to the underlying web-service. To call the service directly, either GET or POST (for sending large trees) the paramenters (named exactly as on this form, requirement as per italics) to `http://phylodiversity.net/phylomatic/pmws`. This is a new tool (2012-08-23), so *please* let me know if it is not working, or if you have comments. Thanks to the NESCent PHYLOtastic hackathon for impetus and encouragement for upgrading.

`tree =` — *Mandatory option*. Paste in your megatree here (<100kB)

or

`treeuri =` — *Mandatory option*. URL of a tree online (<20MB)

or

`storedtree =` — *Mandatory option*. Locally stored tree-of-trees megatrees, in newick format

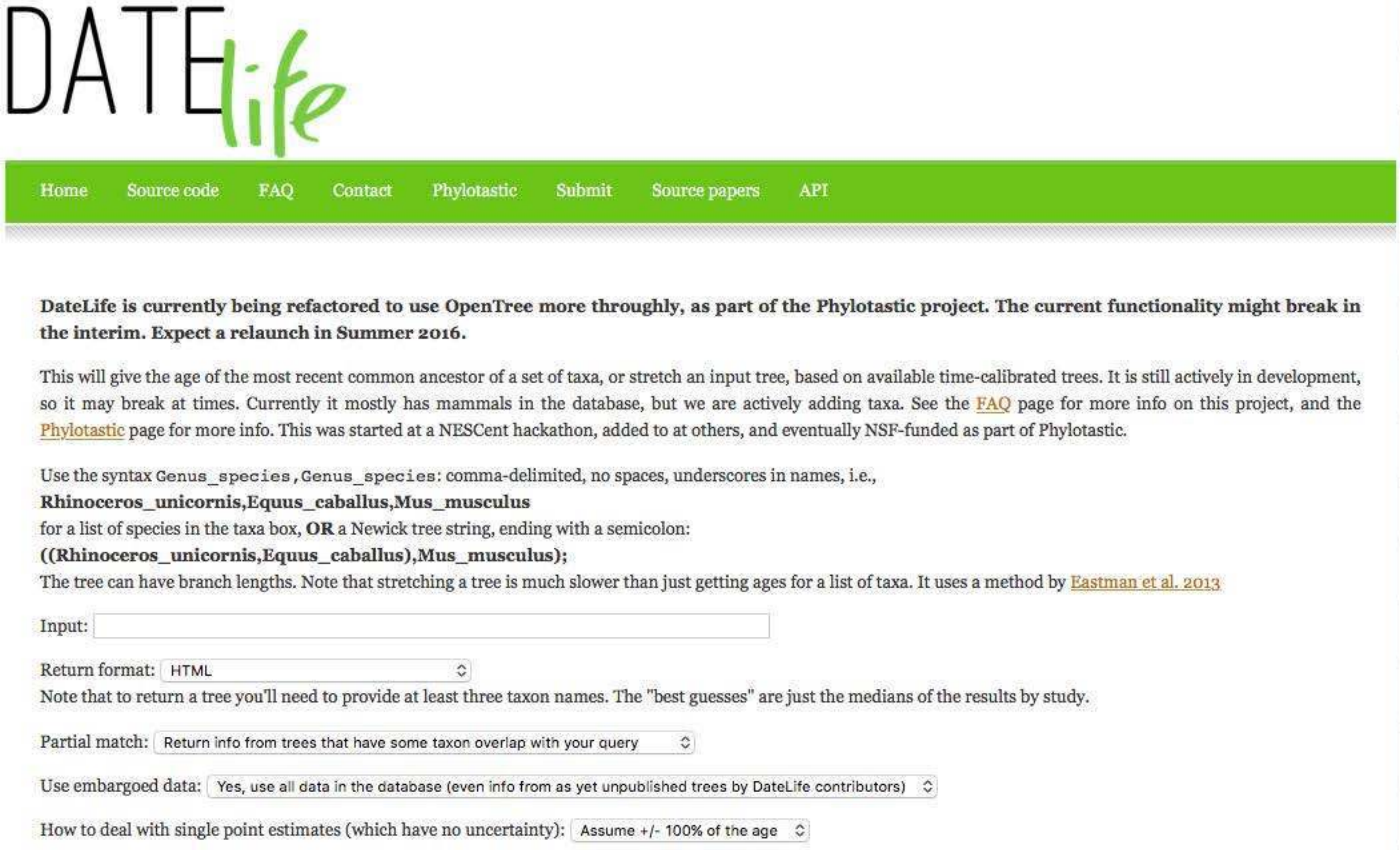
`informat =` newick — *Mandatory*. Either a single newick tree: `(...);`, or NeXML, or a CDAO tree in XMLRDF (see sample)

`method =` ○ `convert` — *Mandatory*. Convert an input format to an output format

---

`method =` ● `phylomatic` — *Mandatory*. Graft your taxa into the megatree, and then prune the megatree to just your taxa

`taxa =` — *Dependent mandatory option*. Slash-delimited node paths, with returns between each taxon (see phylomatic documentation, or sample). *Please note: at the moment, using internal names (e.g. genera) in the megatree as terminal names in your input list will cause the system to hang; please use terminal 'dummy species,' even if your goal is a genus or family tree.*



DateLife is currently being refactored to use OpenTree more throughly, as part of the Phylotastic project. The current functionality might break in the interim. Expect a relaunch in Summer 2016.

This will give the age of the most recent common ancestor of a set of taxa, or stretch an input tree, based on available time-calibrated trees. It is still actively in development, so it may break at times. Currently it mostly has mammals in the database, but we are actively adding taxa. See the FAQ page for more info on this project, and the Phylotastic page for more info. This was started at a NESCent hackathon, added to at others, and eventually NSF-funded as part of Phylotastic.

Use the syntax `Genus_species,Genus_species`: comma-delimited, no spaces, underscores in names, i.e.,
Rhinoceros_unicornis,Equus_caballus,Mus_musculus
for a list of species in the taxa box, **OR** a Newick tree string, ending with a semicolon:
((Rhinoceros_unicornis,Equus_caballus),Mus_musculus);
The tree can have branch lengths. Note that stretching a tree is much slower than just getting ages for a list of taxa. It uses a method by Eastman et al. 2013

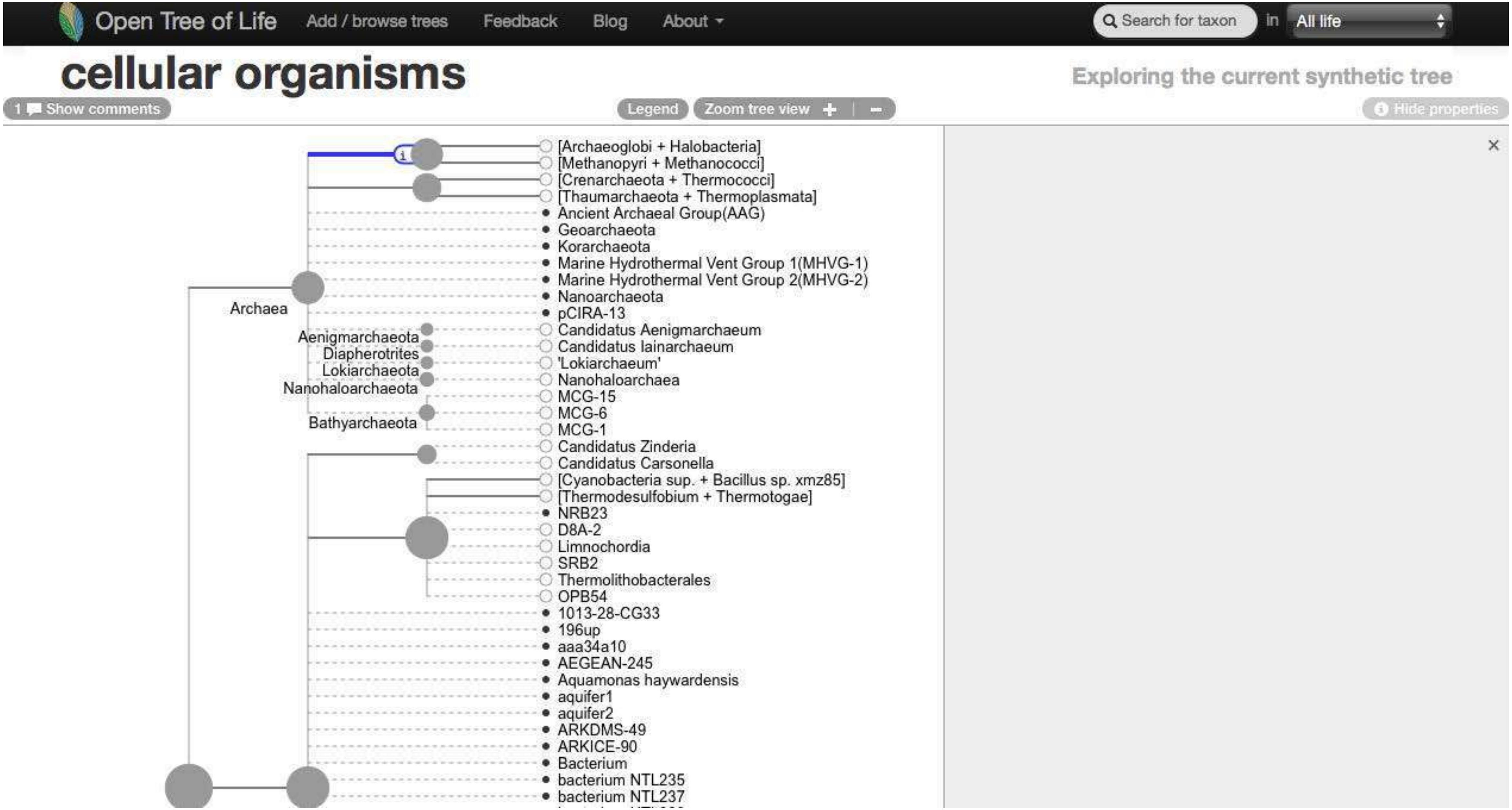
Input:

Return format: HTML
Note that to return a tree you'll need to provide at least three taxon names. The "best guesses" are just the medians of the results by study.

Partial match: Return info from trees that have some taxon overlap with your query

Use embargoed data: Yes, use all data in the database (even info from as yet unpublished trees by DateLife contributors)

How to deal with single point estimates (which have no uncertainty): Assume +/- 100% of the age



https://tree.opentreeoflife.org

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016 doi: 10.1111/2041-210X.12593

APPLICATION

`rotl`: an R package to interact with the Open Tree of Life data

François Michonneau[1,2]*, Joseph W. Brown[3] and David J. Winter[4]

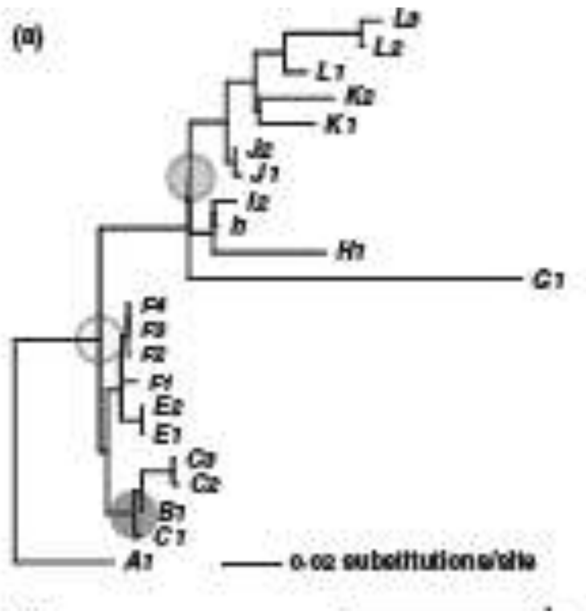
Methods in Ecology and Evolution
British Ecological Society

Methods in Ecology and Evolution 2013 doi: 10.1111/2041-210X.12051

APPLICATION

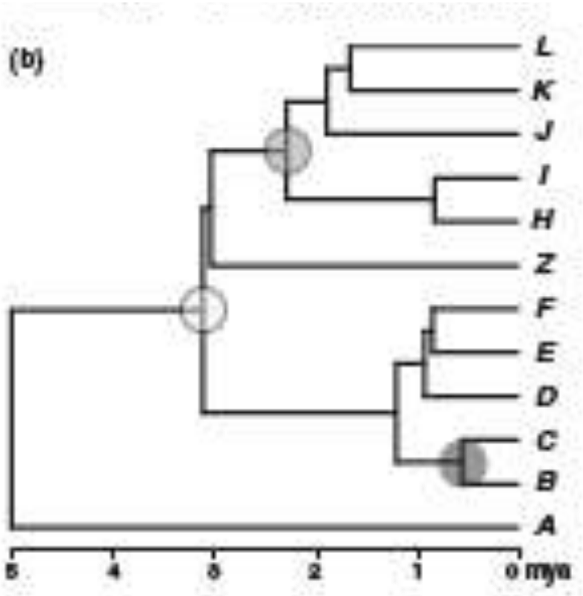
Congruification: support for time scaling large phylogenetic trees

Jonathan M. Eastman[1,2*], Luke J. Harmon[1,2] and David C. Tank[2,3]



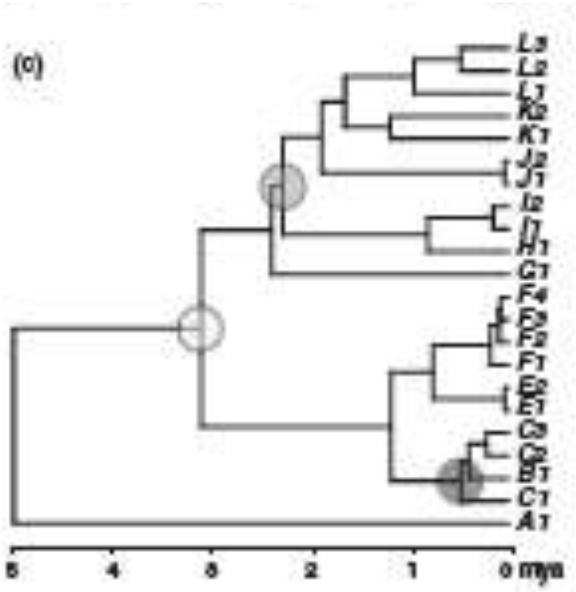
Target tree

+



Reference tree

=



Abordagens de megaphylogeny

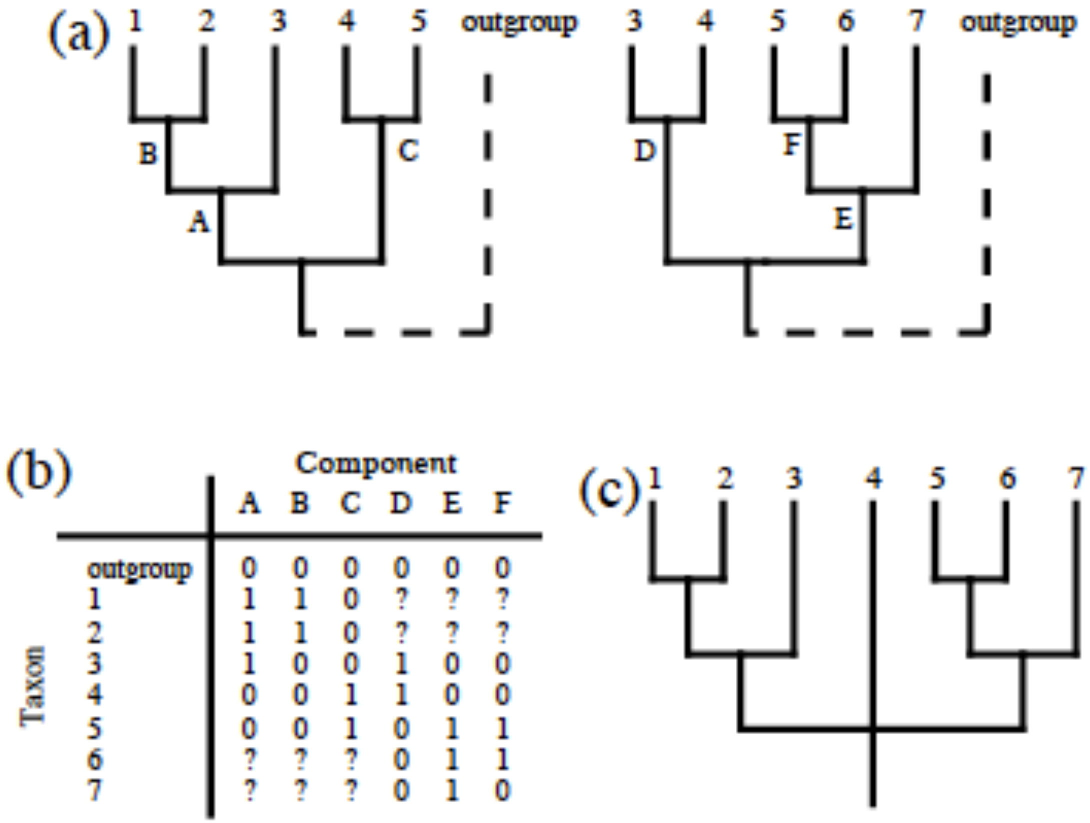
Estimativa de relações filogenéticas para um grande número de espécies

↓

Estudos ecológicos ou comparativos em larga escala, sem preocupação com taxonomia

Supertree

- Filogenias grande escala reconstruídas pela junção de outras menores, que não necessariamente devem ter o mesmo conjunto de espécies, conquanto que haja alguma sobreposição.
- Vantagens
  - Permitem combinar espécies fósseis com atuais
  - São necessários somente as topologias como dados de entrada.
  - Permitem combinar topologias obtidas com diferentes fontes de dado



Taxon	A	B	C	D	E	F
	Component					
outgroup	0	0	0	0	0	0
1	1	1	0	?	?	?
2	1	1	0	?	?	?
3	1	0	0	1	0	0
4	0	0	1	1	0	0
5	0	0	1	0	1	1
6	?	?	?	0	1	1
7	?	?	?	0	1	0

Bininda-Emonds et al. 2002

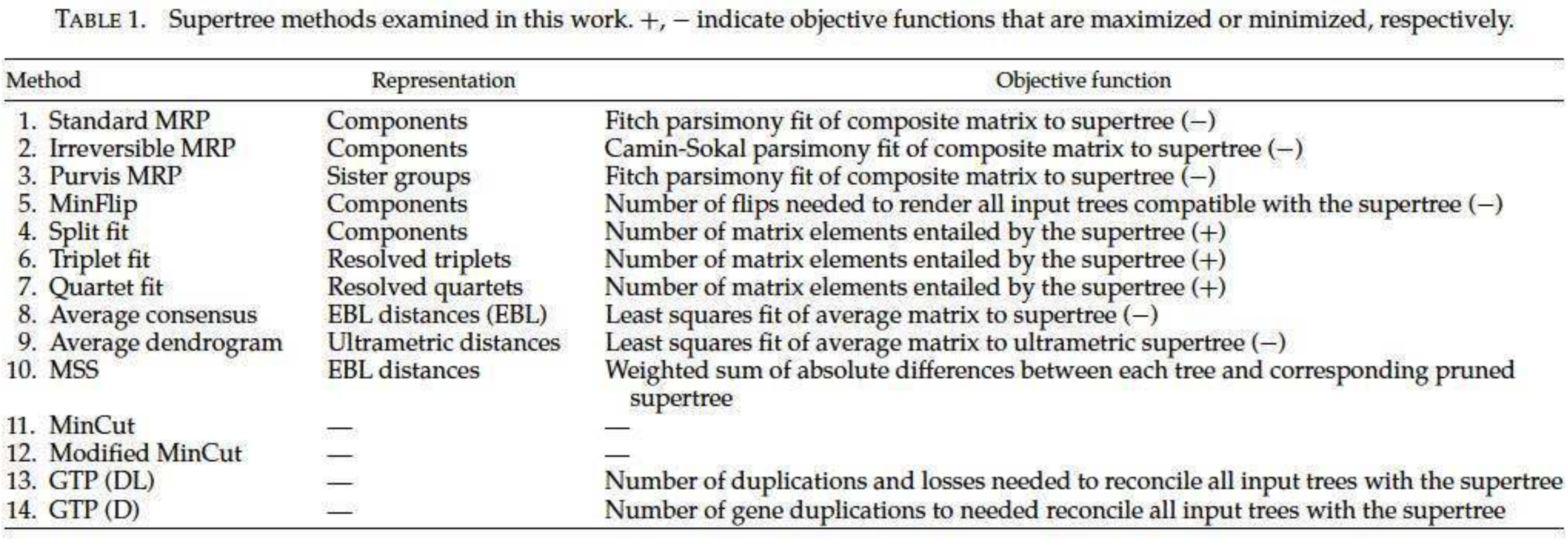


TABLE 1. Supertree methods examined in this work. +, − indicate objective functions that are maximized or minimized, respectively.

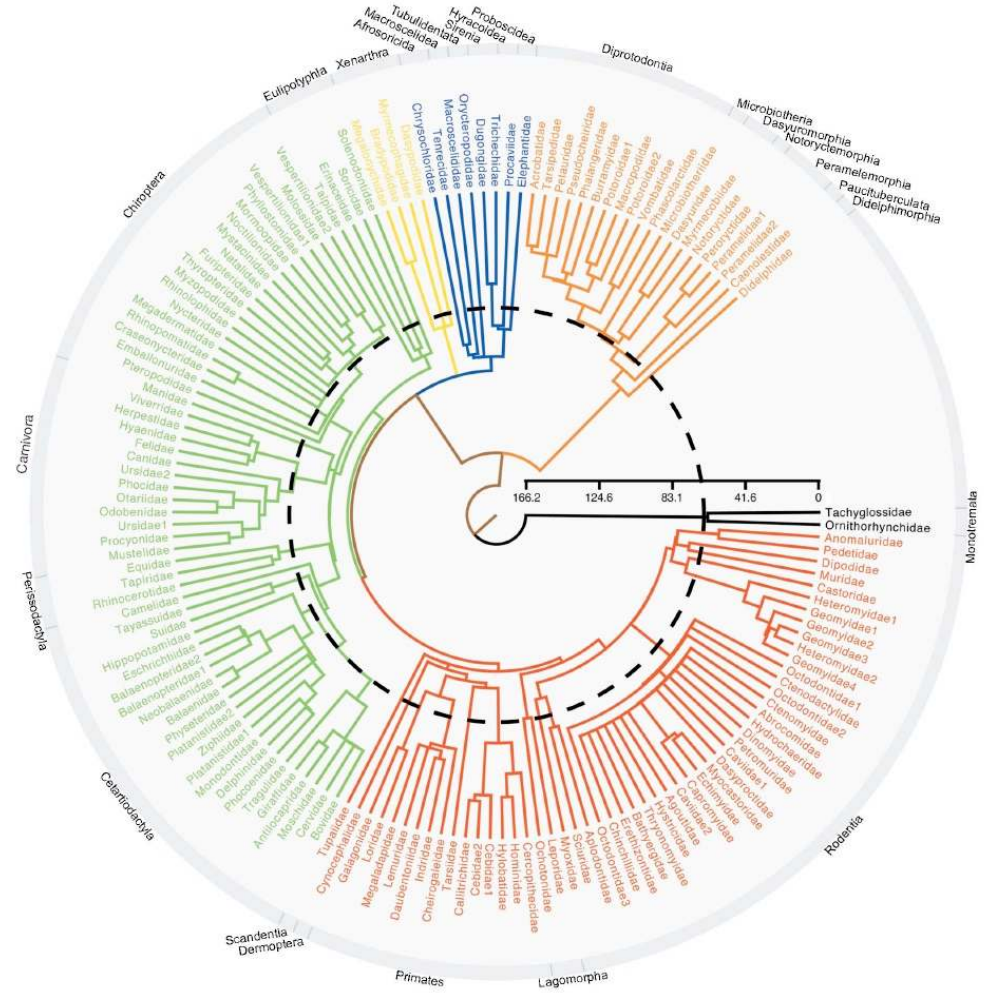
Method	Representation	Objective function
1. Standard MRP	Components	Fitch parsimony fit of composite matrix to supertree (−)
2. Irreversible MRP	Components	Camin-Sokal parsimony fit of composite matrix to supertree (−)
3. Purvis MRP	Sister groups	Fitch parsimony fit of composite matrix to supertree (−)
5. MinFlip	Components	Number of flips needed to render all input trees compatible with the supertree (−)
4. Split fit	Components	Number of matrix elements entailed by the supertree (+)
6. Triplet fit	Resolved triplets	Number of matrix elements entailed by the supertree (+)
7. Quartet fit	Resolved quartets	Number of matrix elements entailed by the supertree (+)
8. Average consensus	EBL distances (EBL)	Least squares fit of average matrix to supertree (−)
9. Average dendrogram	Ultrametric distances	Least squares fit of average matrix to ultrametric supertree (−)
10. MSS	EBL distances	Weighted sum of absolute differences between each tree and corresponding pruned supertree
11. MinCut	—	—
12. Modified MinCut	—	—
13. GTP (DL)	—	Number of duplications and losses needed to reconcile all input trees with the supertree
14. GTP (D)	—	Number of gene duplications to needed reconcile all input trees with the supertree

Wilkinson et al. 2005

The delayed rise of present-day mammals

Olaf R. P. Bininda-Emonds[1]†, Marcel Cardillo[2]†, Kate E. Jones[4], Ross D. E. MacPhee[5], Robin M. D. Beck[6], Richard Grenyer[7], Samantha A. Price[8], Rutger A. Vos[9], John L. Gittleman[10] & Andy Purvis[2,3]

Vol 446|29 March 2007|**doi:10.1038/nature05634** nature



Críticas


von Haeseler *BMC Biology* 2012, **10**:13
http://www.biomedcentral.com/1741-7007/10/13

BMC Biology
incorporating Journal of Biology

COMMENTARY **Open Access**

Do we still need supertrees?

Arndt von Haeseler*

See research article http://www.biomedcentral.com/1741-7007/10/12

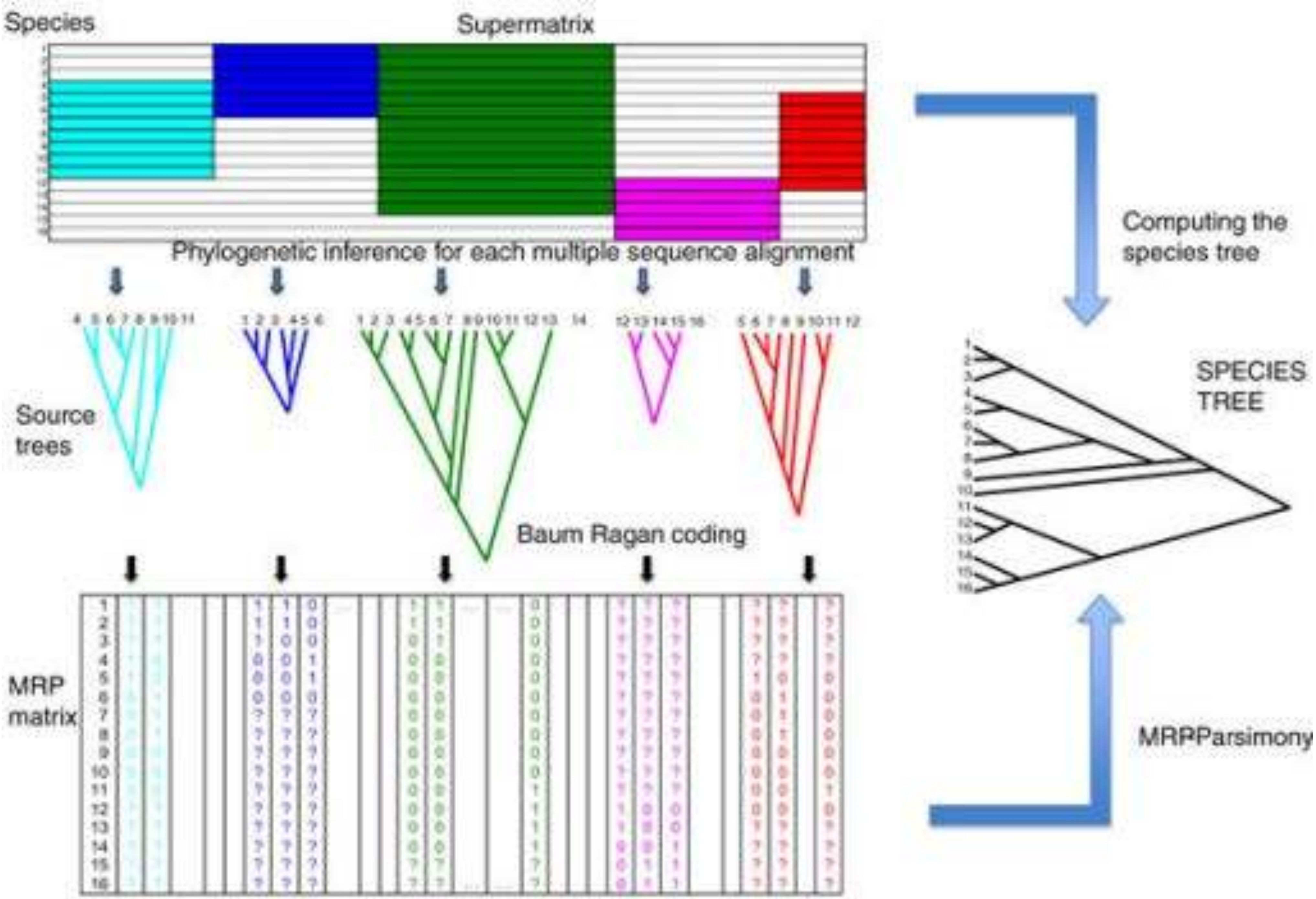
- Lidar com incongruências entre árvores construídas com diferentes métodos e dados não é simples
- Resolução e suporte é limitada às filogenias base menores, já que não distingue entre topologias com alto e baixo suporte. Incongruências acabam como politomias (que são muito comuns)
- Comprimento de ramo não tem significado (problema para análises comparativas)
- Resolução é bem baixa e não é confiável para representar relações muito finas (gênero e espécie)

Supermatrix

- Utiliza sequências de nucleotídeos depositadas no GenBank por estudos anteriores
  - Seleção de genes (marcadores) para minimizar missing data
  - Alinhamento
  - Estimativa de topologia utilizando algoritmos para dados em larga escala
- Limitações
  - GenBank não é padronizado, nomenclatura destoante, muitos dados ruins
  - Garbage in, garbage out
  - Dependendo do grupo existem poucos genes sequenciados
  - Alinhamento difícil ou de baixa qualidade
  - Gasta-se muito tempo filtrando e escolhendo sequências

Supermatrix

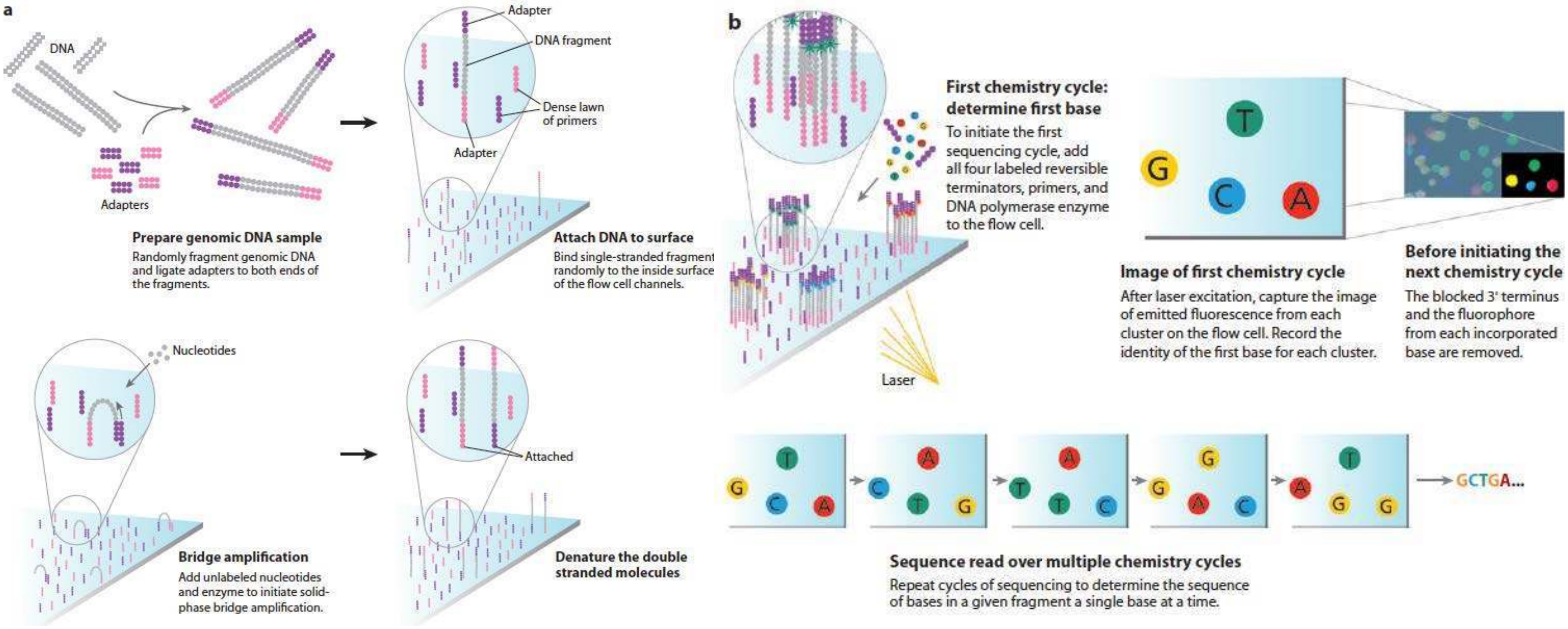
- Algoritmos para inferência de árvores que comportam grande número de terminais
    - RAxML
    - ExaML
    - ExaBayes
- Algoritmos para alinhamento
    - MAFFT
    - MUSCLE
    - ClustalW
    - T-Coffee
- Estimativa de divergência com muitos terminais
    - treePL


Species
Supermatrix
Phylogenetic inference for each multiple sequence alignment
Source trees
Baum Ragan coding
MRP matrix
Computing the species tree
SPECIES TREE
MRPParsimony

Next-generation sequencing e Phylogenomics

- Permitem sequenciar uma grande quantidade DNA e RNA de forma muito mais rápida e barata que o método Sanger comumente usado
- Geram uma quantidade enorme de dados para vários genes (marcadores) ao mesmo tempo (em paralelo)
- Possibilitam aumentar exponencialmente a quantidade de dados utilizados em reconstrução filogenética e filogeográfica
- Atualmente existem 4 plataformas comerciais
  - **Illumina (Solexa) sequencing**
  - Roche 454 sequencing
  - Ion torrent: Proton / PGM sequencing
  - SOLiD sequencing

Workflow Ilumina NGS



Prepare genomic DNA sample
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

Attach DNA to surface
Bind single-stranded fragment randomly to the inside surface of the flow cell channels.

Bridge amplification
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

Denature the double stranded molecules

First chemistry cycle: determine first base
To initiate the first sequencing cycle, add all four labeled reversible terminators, primers, and DNA polymerase enzyme to the flow cell.

Image of first chemistry cycle
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

Before initiating the next chemistry cycle
The blocked 3' terminus and the fluorophore from each incorporated base are removed.

Sequence read over multiple chemistry cycles
Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

Mardis 2008

Pipelines (semi)automatizadas ”for dummies”

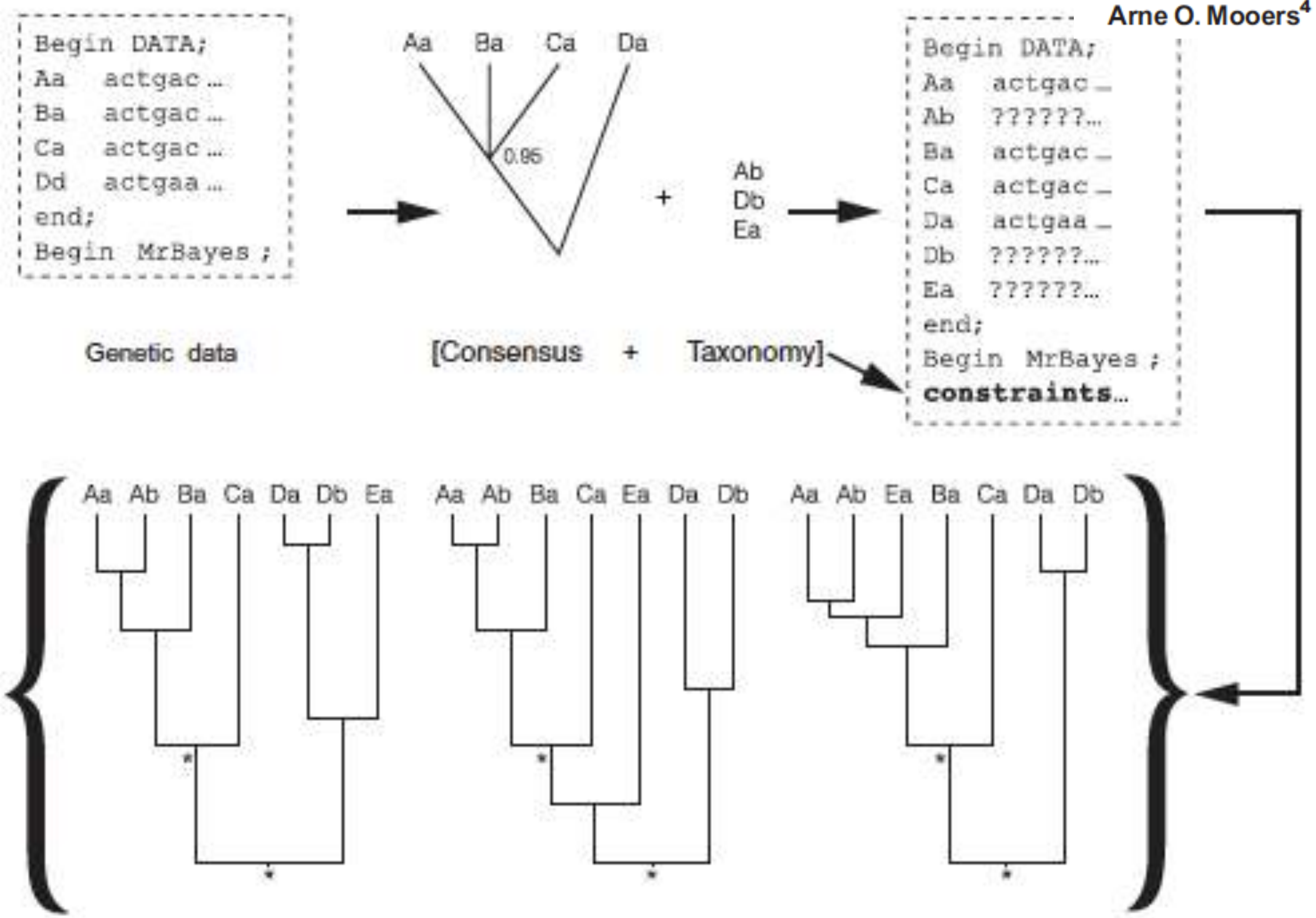
Methods in Ecology and Evolution
British Ecological Society

Methods in Ecology and Evolution 2013, 4, 1011–1017 doi: 10.1111/2041-210X.12117

APPLICATION

PASTIS: an R package to facilitate phylogenetic assembly with soft taxonomic inferences

Gavin H. Thomas[1]*†, Klaas Hartmann[2]†, Walter Jetz[3], Jeffrey B. Joy[4], Aki Mimoto[4] and Arne O. Mooers[4]

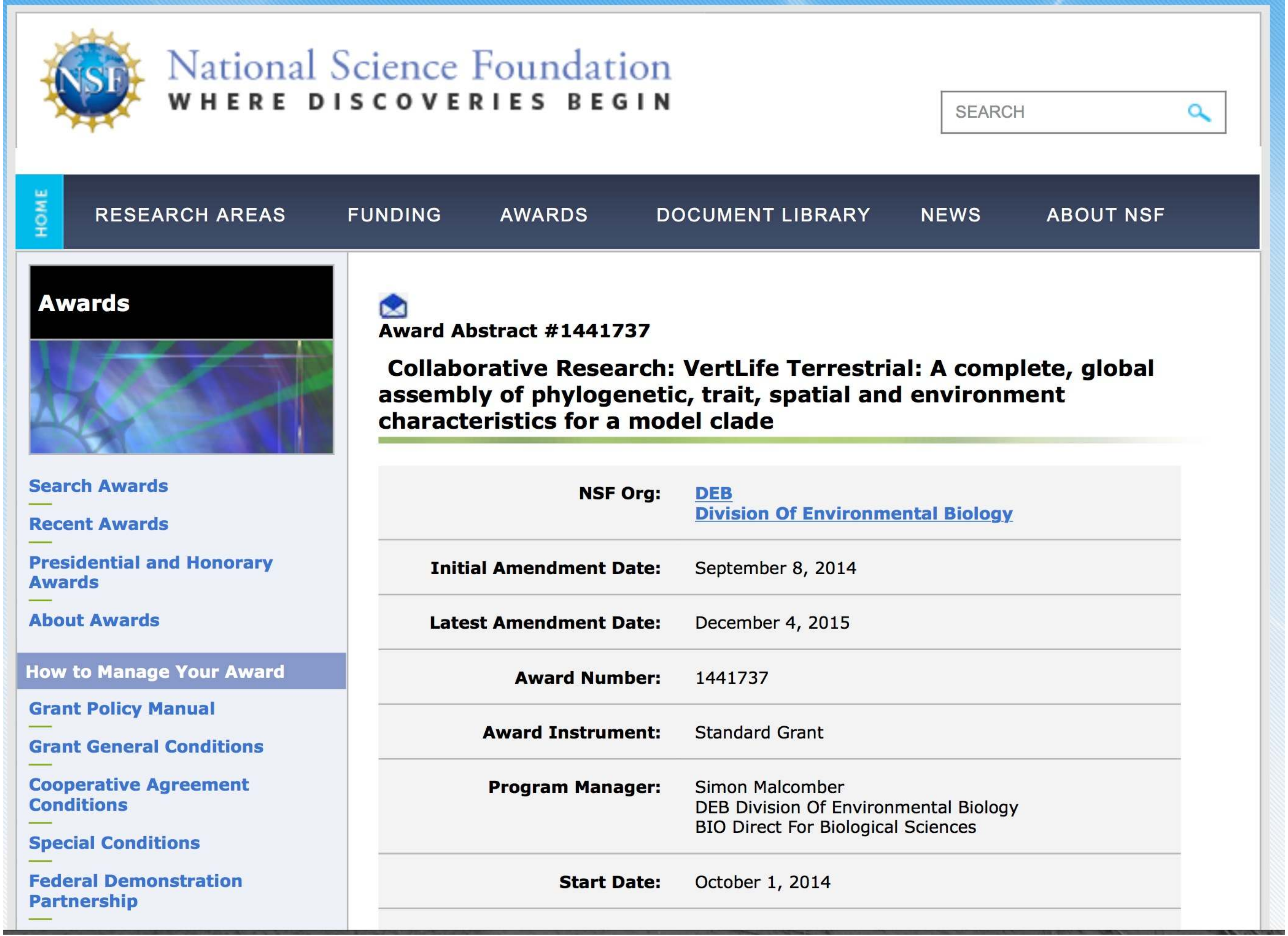


```
Begin DATA;
Aa   actgac …
Ba   actgac …
Ca   actgac …
Dd   actgaa …
end;
Begin  MrBayes ;
```

Genetic data

[Consensus + Taxonomy]

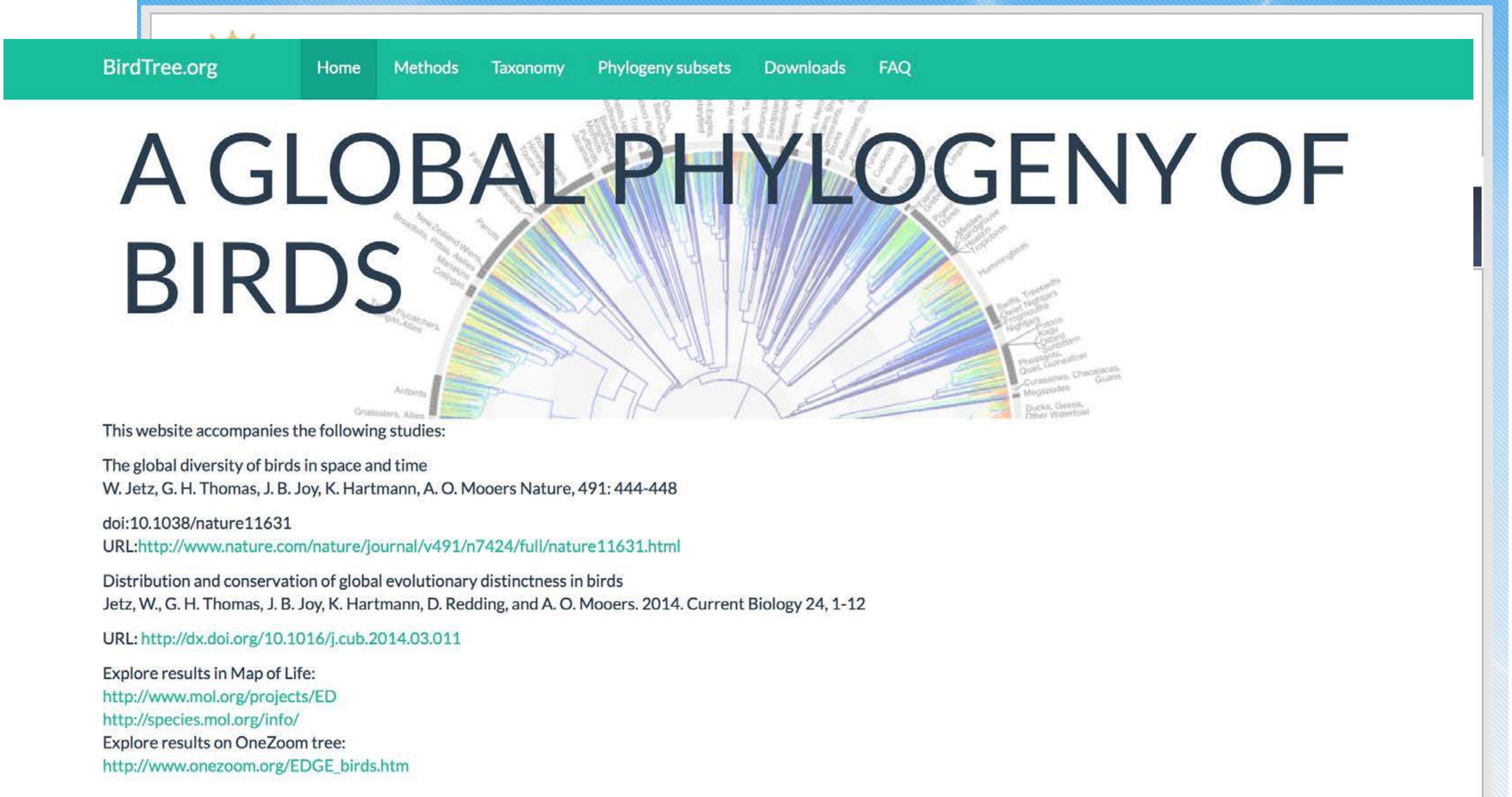
```
Begin DATA;
Aa   actgac …
Ab   ??????…
Ba   actgac …
Ca   actgac …
Da   actgaa …
Db   ??????…
Ea   ??????…
end;
Begin  MrBayes ;
constraints…
```



Award Abstract #1441737

Collaborative Research: VertLife Terrestrial: A complete, global assembly of phylogenetic, trait, spatial and environment characteristics for a model clade

NSF Org:	DEB Division Of Environmental Biology
Initial Amendment Date:	September 8, 2014
Latest Amendment Date:	December 4, 2015
Award Number:	1441737
Award Instrument:	Standard Grant
Program Manager:	Simon Malcomber DEB Division Of Environmental Biology BIO Direct For Biological Sciences
Start Date:	October 1, 2014



A GLOBAL PHYLOGENY OF BIRDS

This website accompanies the following studies:

The global diversity of birds in space and time
W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers Nature, 491: 444-448

doi:10.1038/nature11631
URL:http://www.nature.com/nature/journal/v491/n7424/full/nature11631.html

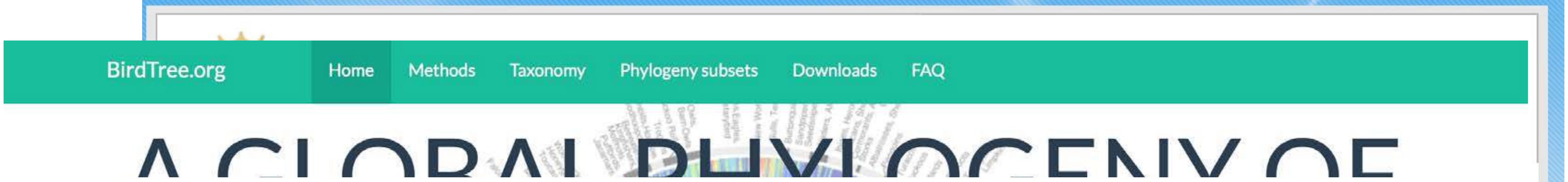
Distribution and conservation of global evolutionary distinctness in birds
Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, D. Redding, and A. O. Mooers. 2014. Current Biology 24, 1-12

URL: http://dx.doi.org/10.1016/j.cub.2014.03.011

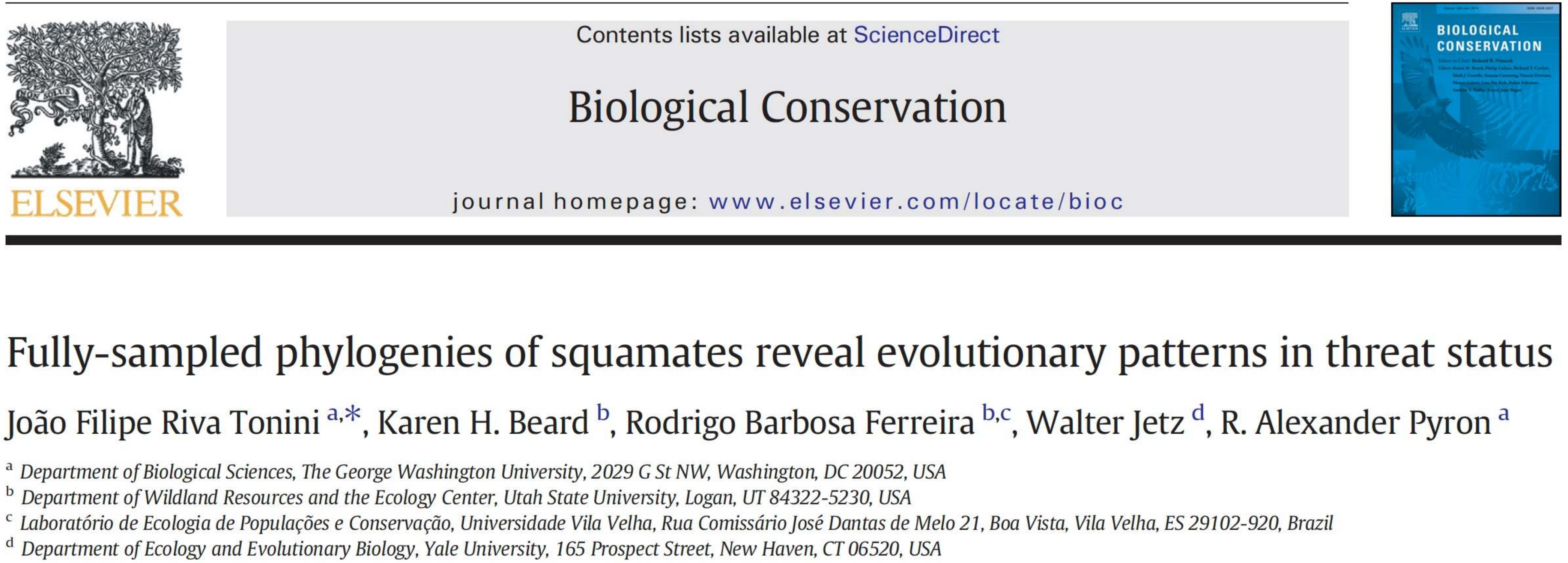
Explore results in Map of Life:
http://www.mol.org/projects/ED
http://species.mol.org/info/
Explore results on OneZoom tree:
http://www.onezoom.org/EDGE_birds.htm

Program Manager: Simon Malcomber
DEB Division Of Environmental Biology
BIO Direct For Biological Sciences

Start Date: October 1, 2014



A GLOBAL PHYLOGENY OF


Contents lists available at ScienceDirect

Biological Conservation

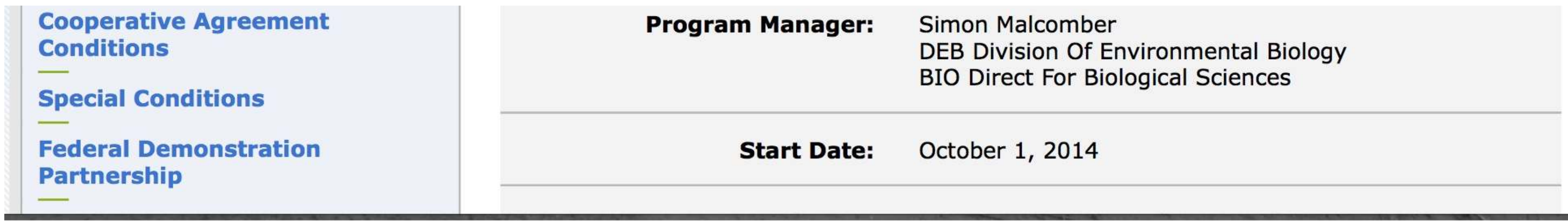
journal homepage: www.elsevier.com/locate/bioc

Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status

João Filipe Riva Tonini [a,*], Karen H. Beard [b], Rodrigo Barbosa Ferreira [b,c], Walter Jetz [d], R. Alexander Pyron [a]

[a] *Department of Biological Sciences, The George Washington University, 2029 G St NW, Washington, DC 20052, USA*
[b] *Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT 84322-5230, USA*
[c] *Laboratório de Ecologia de Populações e Conservação, Universidade Vila Velha, Rua Comissário José Dantas de Melo 21, Boa Vista, Vila Velha, ES 29102-920, Brazil*
[d] *Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 06520, USA*

http://species.mol.org/info/
Explore results on OneZoom tree:
http://www.onezoom.org/EDGE_birds.htm



Program Manager:	Simon Malcomber DEB Division Of Environmental Biology BIO Direct For Biological Sciences
Start Date:	October 1, 2014

Review & synthesis

Ecography 35: 001–014, 2012
doi: 10.1111/j.1600-0587.2012.07773.x

Subject Editor: Ken Kozak. Accepted 23 October 2012

Building megaphylogenies for macroecology: taking up the challenge

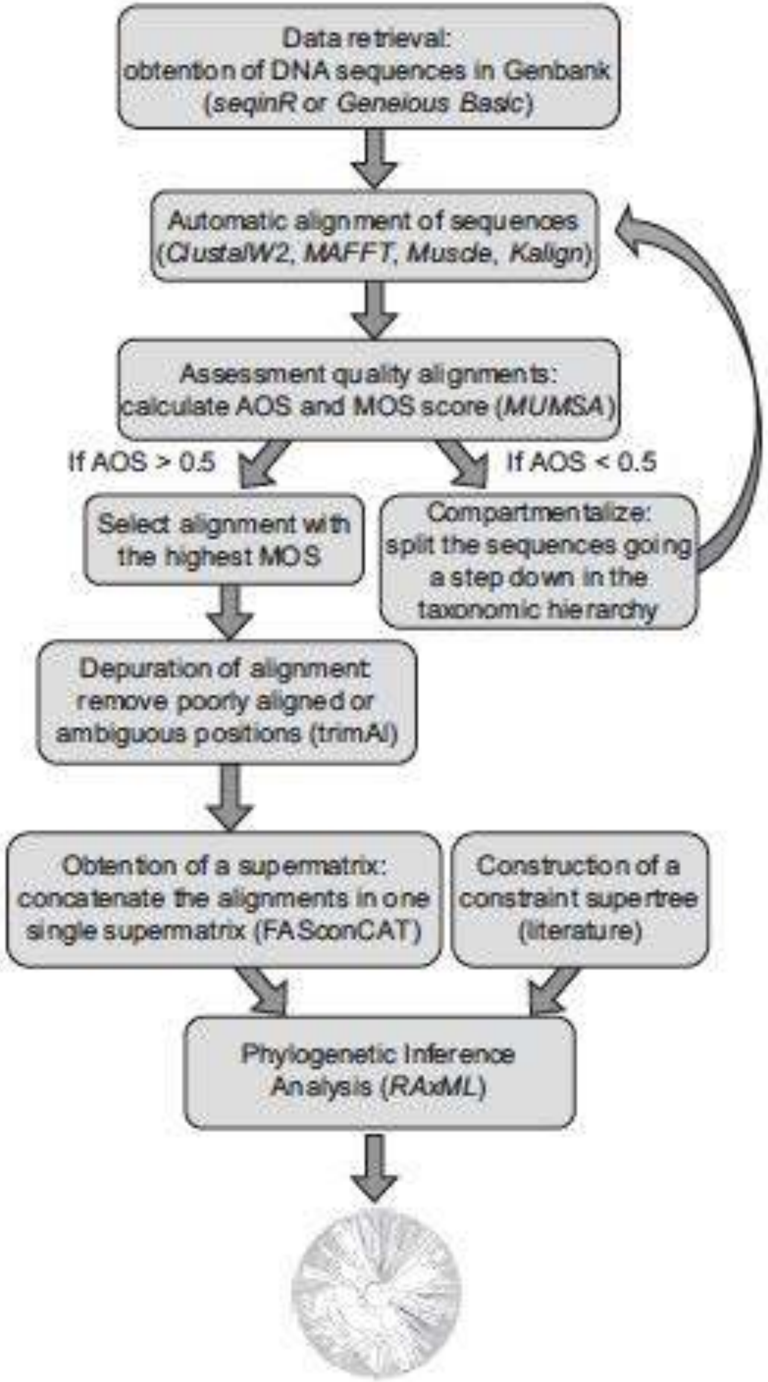
Cristina Roquet, Wilfried Thuiller and Sébastien Lavergne

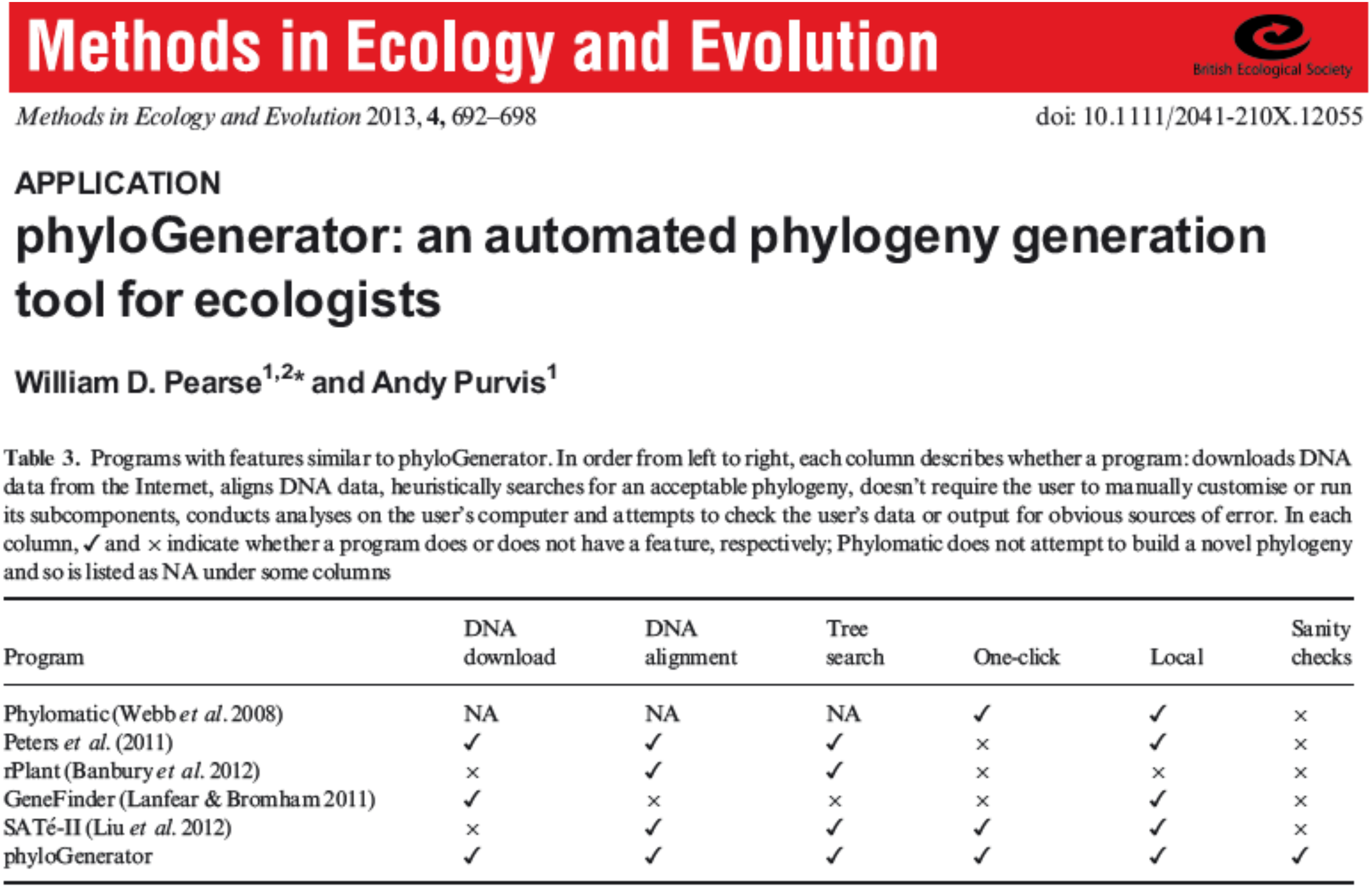
Review & synthesis

Ecography 35: 001–014, 2012
doi: 10.1111/j.1600-0587.2012.07773.x
Authors. Ecography © 2012 Nordic Society Oikos
ect Editor: Ken Kozak. Accepted 23 October 2012

Building megaphylog y: taking up the challenge

Cristina Roquet, Wilfried Thui





Methods in Ecology and Evolution
British Ecological Society

Methods in Ecology and Evolution 2013, **4**, 692–698 doi: 10.1111/2041-210X.12055

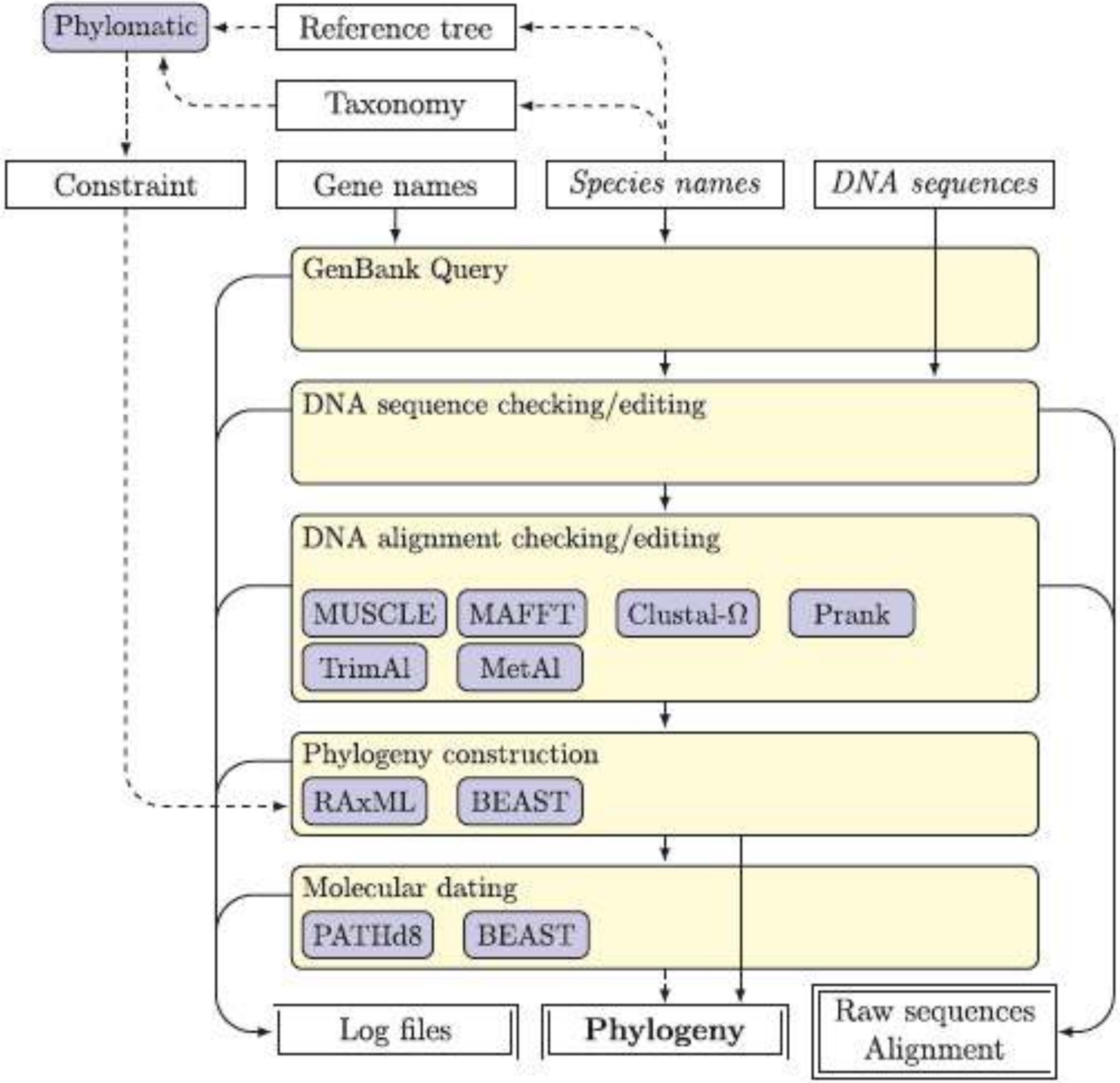
APPLICATION

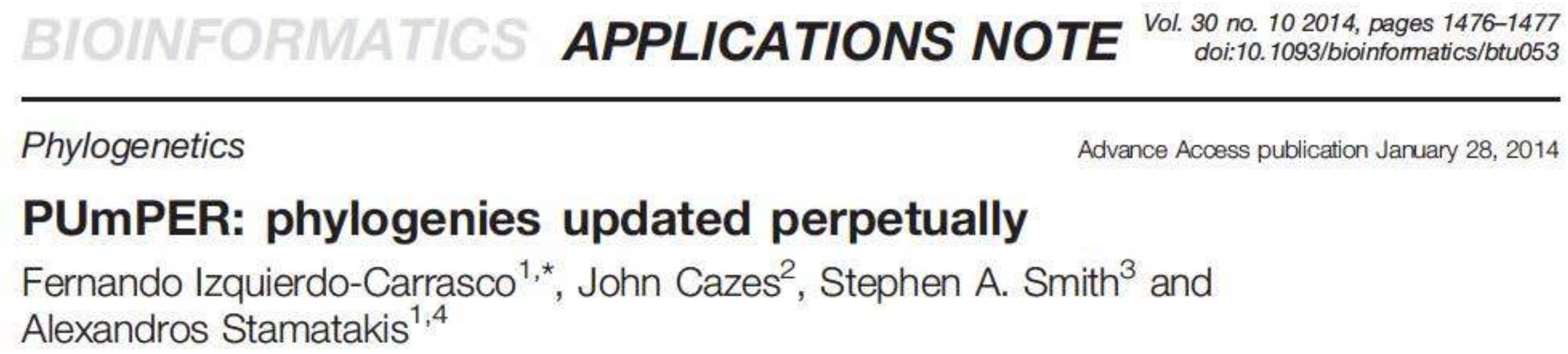
phyloGenerator: an automated phylogeny generation tool for ecologists

William D. Pearse[1,2]* and Andy Purvis[1]

Table 3. Programs with features similar to phyloGenerator. In order from left to right, each column describes whether a program: downloads DNA data from the Internet, aligns DNA data, heuristically searches for an acceptable phylogeny, doesn't require the user to manually customise or run its subcomponents, conducts analyses on the user's computer and attempts to check the user's data or output for obvious sources of error. In each column, ✓ and × indicate whether a program does or does not have a feature, respectively; Phylomatic does not attempt to build a novel phylogeny and so is listed as NA under some columns

Program	DNA download	DNA alignment	Tree search	One-click	Local	Sanity checks
Phylomatic (Webb *et al.* 2008)	NA	NA	NA	✓	✓	×
Peters *et al.* (2011)	✓	✓	✓	×	✓	×
rPlant (Banbury *et al.* 2012)	×	✓	✓	×	×	×
GeneFinder (Lanfear & Bromham 2011)	✓	×	×	×	✓	×
SATé-II (Liu *et al.* 2012)	×	✓	✓	✓	✓	×
phyloGenerator	✓	✓	✓	✓	✓	✓


Phylomatic
Reference tree
Taxonomy
Constraint
Gene names
Species names
DNA sequences
GenBank Query
DNA sequence checking/editing
DNA alignment checking/editing
MUSCLE
MAFFT
Clustal-Ω
Prank
TrimAl
MetAl
Phylogeny construction
RAxML
BEAST
Molecular dating
PATHd8
BEAST
Log files
Phylogeny
Raw sequences
Alignment



BIOINFORMATICS **APPLICATIONS NOTE**
Vol. 30 no. 10 2014, pages 1476–1477
doi:10.1093/bioinformatics/btu053

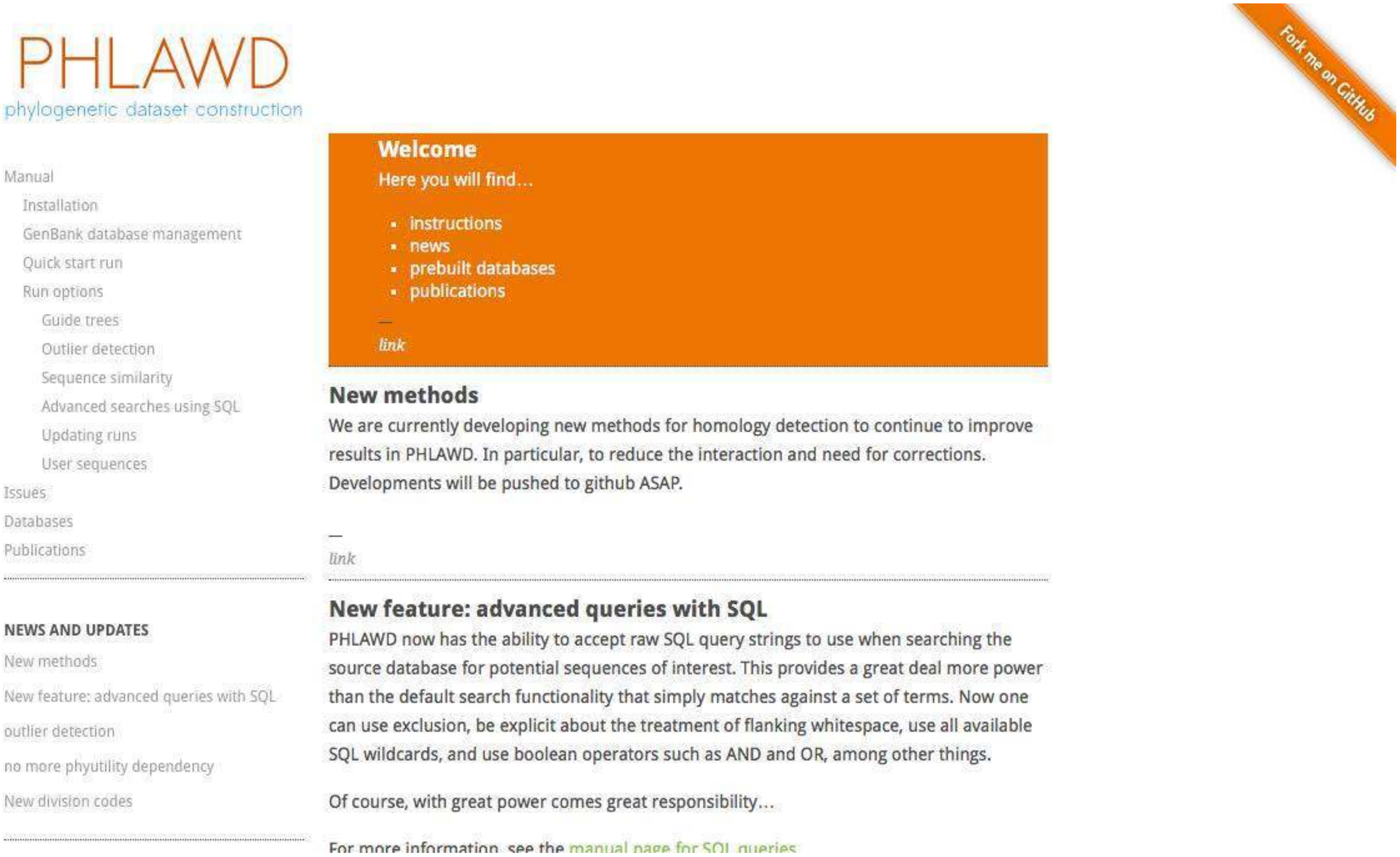
Phylogenetics

Advance Access publication January 28, 2014

PUmPER: phylogenies updated perpetually

Fernando Izquierdo-Carrasco[1,*], John Cazes[2], Stephen A. Smith[3] and Alexandros Stamatakis[1,4]

- Bem complicado de instalar e usar
- Não é atualizado muito frequentemente



PHLAWD
phylogenetic dataset construction

Welcome

Here you will find…

- instructions
- news
- prebuilt databases
- publications

—
link

New methods

We are currently developing new methods for homology detection to continue to improve results in PHLAWD. In particular, to reduce the interaction and need for corrections. Developments will be pushed to github ASAP.

—
link

New feature: advanced queries with SQL

PHLAWD now has the ability to accept raw SQL query strings to use when searching the source database for potential sequences of interest. This provides a great deal more power than the default search functionality that simply matches against a set of terms. Now one can use exclusion, be explicit about the treatment of flanking whitespace, use all available SQL wildcards, and use boolean operators such as AND and OR, among other things.

Of course, with great power comes great responsibility…

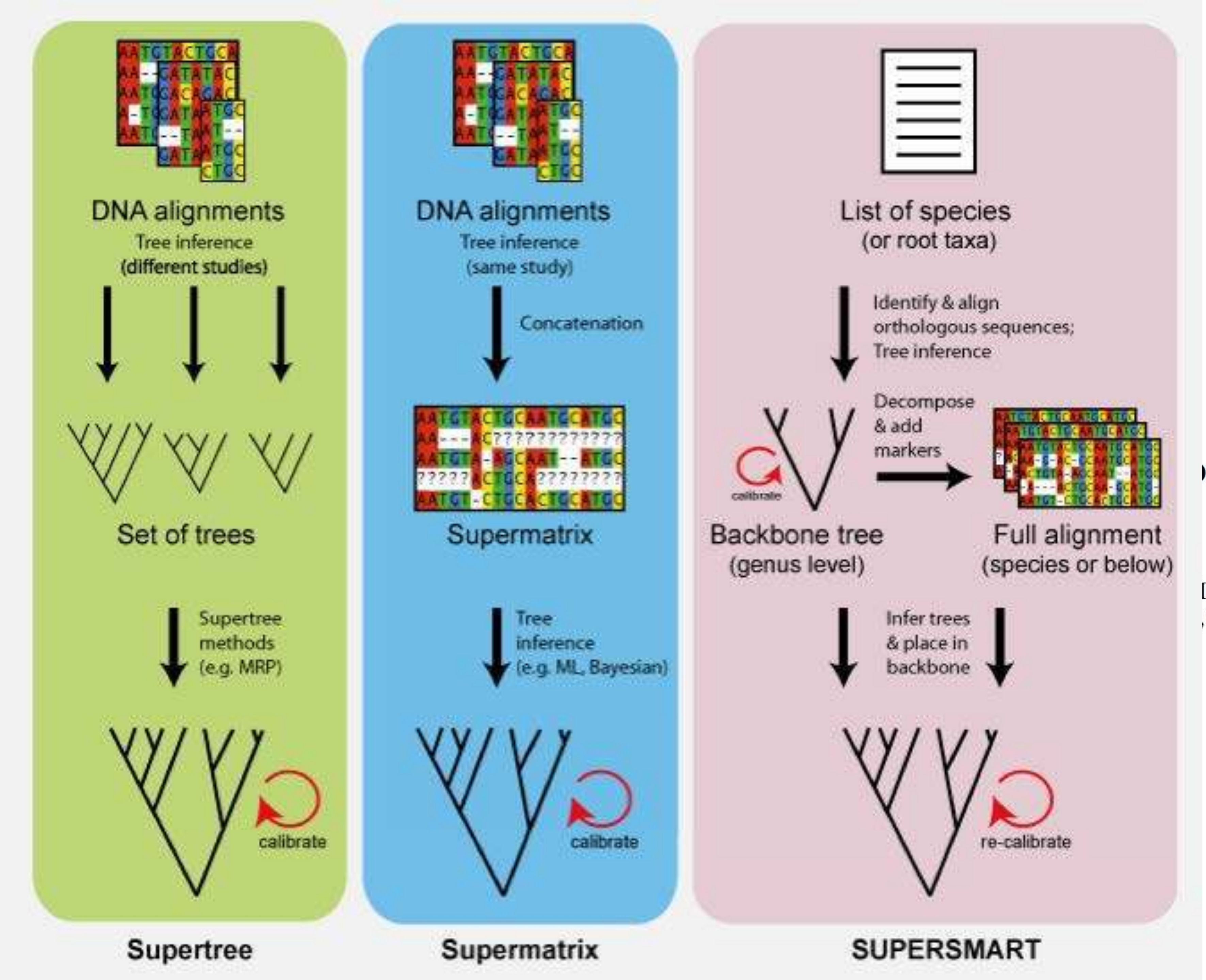
For more information, see the manual page for SQL queries

Syst. Biol. 66(2):152–166, 2017

DOI:10.1093/sysbio/syw066
Advance Access publication August 19, 2016

Toward a Self-Updating Platform for Estimating Rates of Speciation and Migration, Ages, and Relationships of Taxa

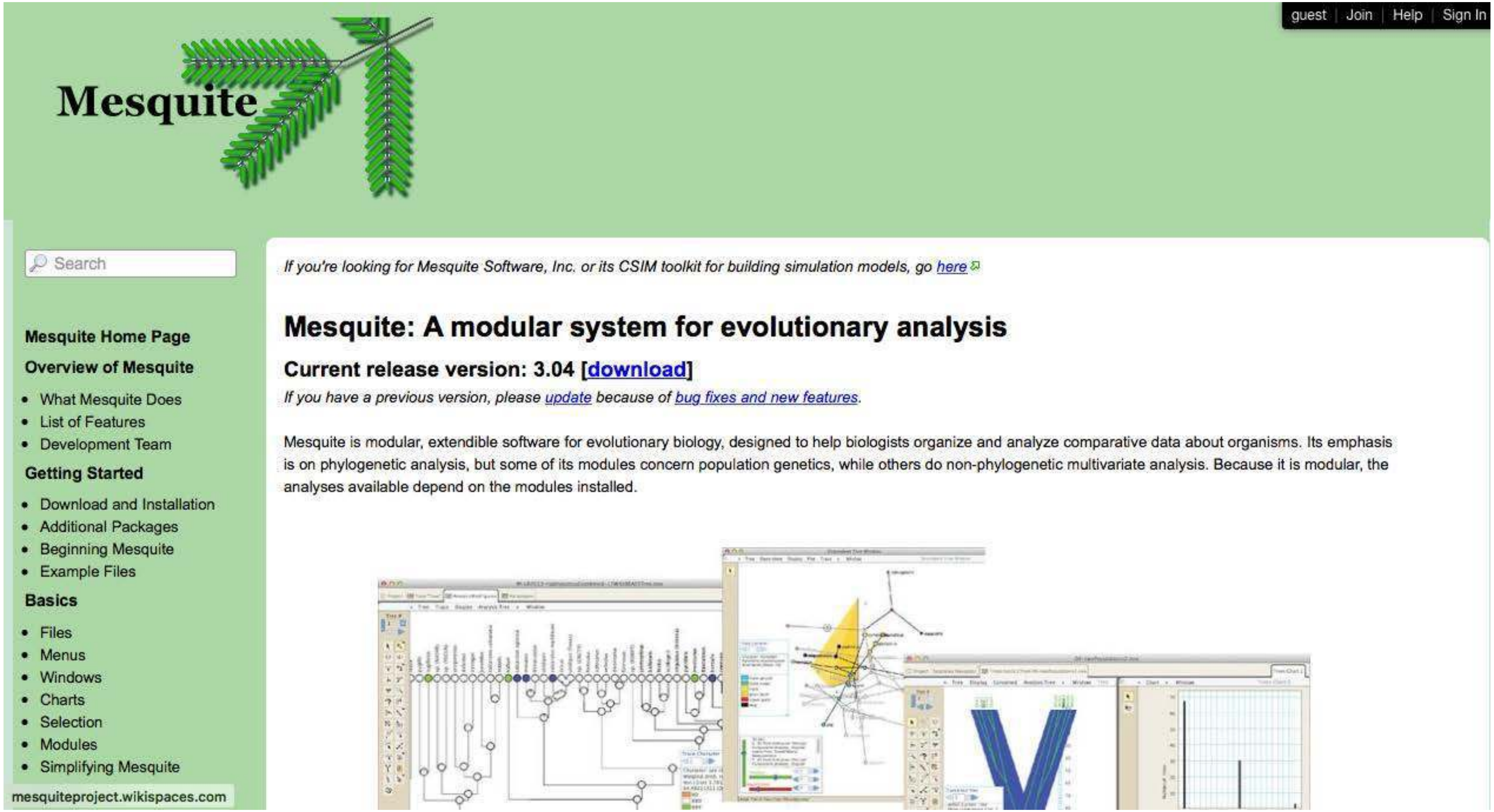
ALEXANDRE ANTONELLI[1,2,*], HANNES HETTLING[3], FABIEN L. CONDAMINE[1,4], KARIN VOS[1], R. HENRIK NILSSON[1], MICHAEL J. SANDERSON[5], HERVÉ SAUQUET[6], RUUD SCHARN[1], DANIELE SILVESTRO[1,7], MATS TÖPEL[8,9], CHRISTINE D. BACON[1], BENGT OXELMAN[1], AND RUTGER A. VOS[3]



Syst. Biol. 66(2):152–166,
© The Author(s) 2016. Pu
This is an Open Access a
which permits unrestrict
DOI:10.1093/sysbio/syw
Advance Access publicat

Toward a S … n, Ages,

ALEXANDRE … ILSSON[1],
MICH … [9],



Mesquite

Mesquite Home Page

Overview of Mesquite

- What Mesquite Does
- List of Features
- Development Team

Getting Started

- Download and Installation
- Additional Packages
- Beginning Mesquite
- Example Files

Basics

- Files
- Menus
- Windows
- Charts
- Selection
- Modules
- Simplifying Mesquite

mesquiteproject.wikispaces.com

If you're looking for Mesquite Software, Inc. or its CSIM toolkit for building simulation models, go here

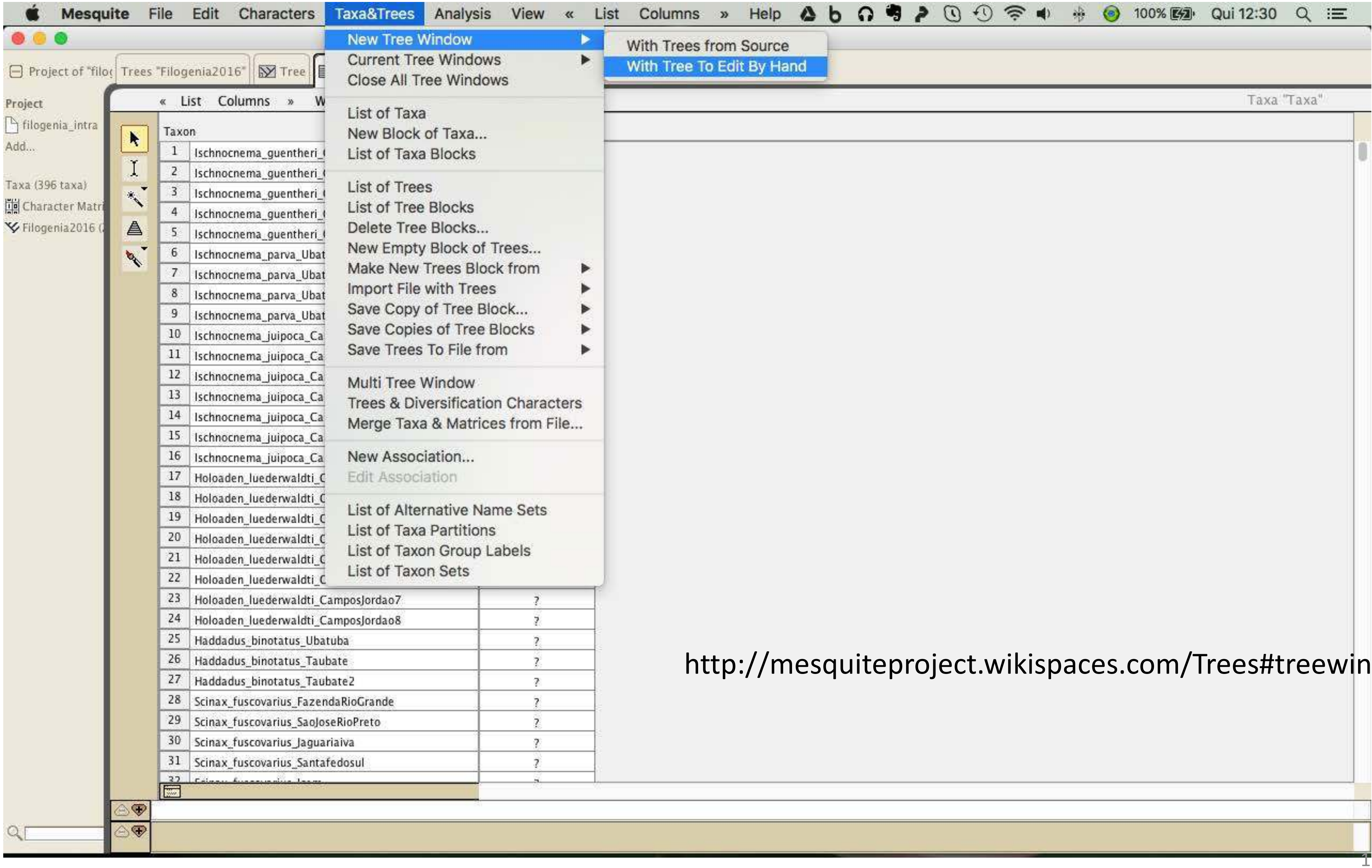
Mesquite: A modular system for evolutionary analysis

Current release version: 3.04 [download]

If you have a previous version, please update because of bug fixes and new features.

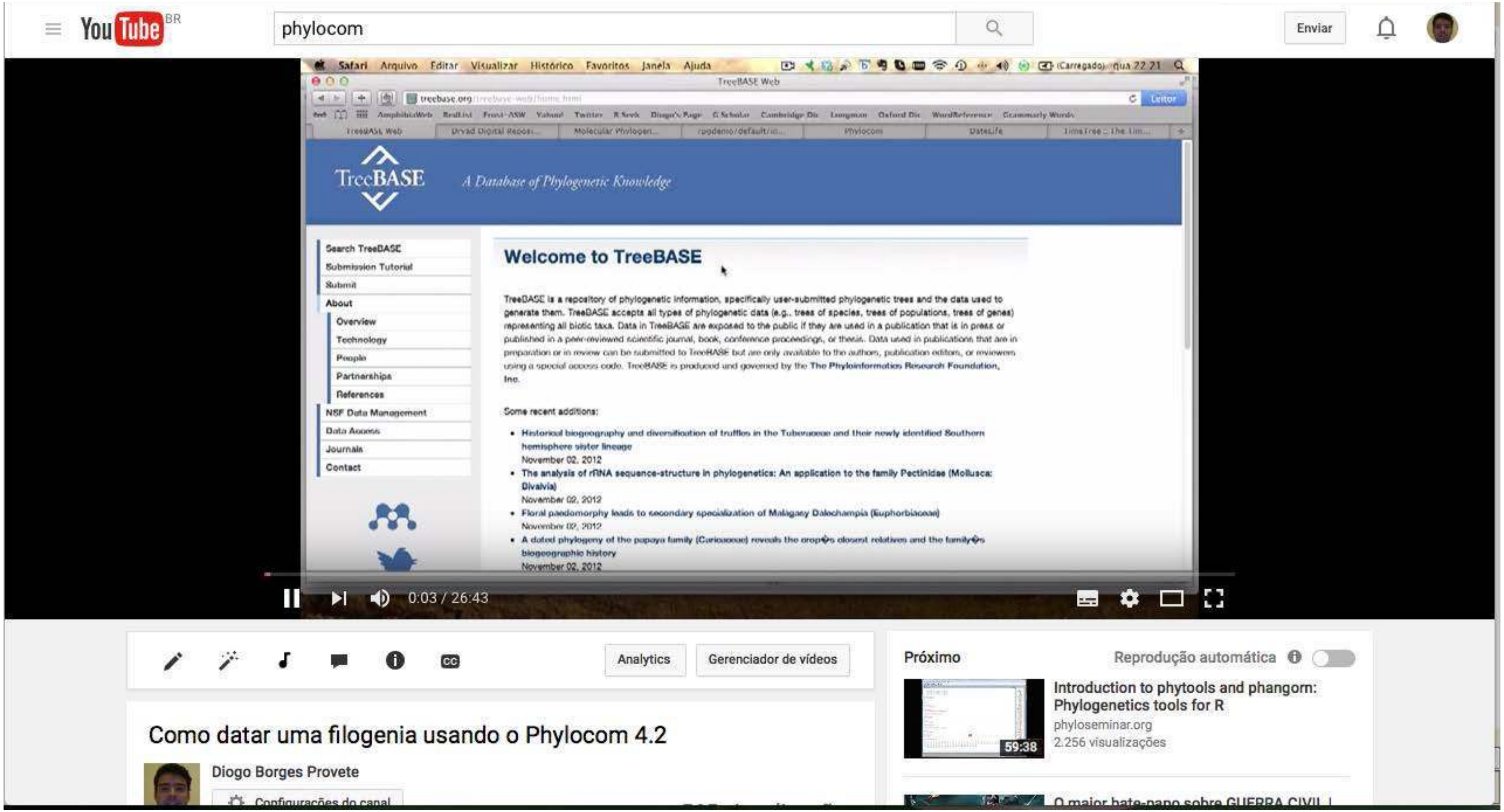
Mesquite is modular, extendible software for evolutionary biology, designed to help biologists organize and analyze comparative data about organisms. Its emphasis is on phylogenetic analysis, but some of its modules concern population genetics, while others do non-phylogenetic multivariate analysis. Because it is modular, the analyses available depend on the modules installed.

Última opção: fazer uma topologia sem comprimento de ramo na mão!



	Taxon	
1	Ischnocnema_guentheri_	
2	Ischnocnema_guentheri_	
3	Ischnocnema_guentheri_	
4	Ischnocnema_guentheri_	
5	Ischnocnema_guentheri_	
6	Ischnocnema_parva_Ubat	
7	Ischnocnema_parva_Ubat	
8	Ischnocnema_parva_Ubat	
9	Ischnocnema_parva_Ubat	
10	Ischnocnema_juipoca_Ca	
11	Ischnocnema_juipoca_Ca	
12	Ischnocnema_juipoca_Ca	
13	Ischnocnema_juipoca_Ca	
14	Ischnocnema_juipoca_Ca	
15	Ischnocnema_juipoca_Ca	
16	Ischnocnema_juipoca_Ca	
17	Holoaden_luederwaldti_C	
18	Holoaden_luederwaldti_C	
19	Holoaden_luederwaldti_C	
20	Holoaden_luederwaldti_C	
21	Holoaden_luederwaldti_C	
22	Holoaden_luederwaldti_C	
23	Holoaden_luederwaldti_CamposJordao7	?
24	Holoaden_luederwaldti_CamposJordao8	?
25	Haddadus_binotatus_Ubatuba	?
26	Haddadus_binotatus_Taubate	?
27	Haddadus_binotatus_Taubate2	?
28	Scinax_fuscovarius_FazendaRioGrande	?
29	Scinax_fuscovarius_SaoJoseRioPreto	?
30	Scinax_fuscovarius_Jaguariaiva	?
31	Scinax_fuscovarius_Santafedosul	?

http://mesquiteproject.wikispaces.com/Trees#treewindow



Welcome to TreeBASE

TreeBASE is a repository of phylogenetic information, specifically user-submitted phylogenetic trees and the data used to generate them. TreeBASE accepts all types of phylogenetic data (e.g., trees of species, trees of populations, trees of genes) representing all biotic taxa. Data in TreeBASE are exposed to the public if they are used in a publication that is in press or published in a peer-reviewed scientific journal, book, conference proceedings, or thesis. Data used in publications that are in preparation or in review can be submitted to TreeBASE but are only available to the authors, publication editors, or reviewers using a special access code. TreeBASE is produced and governed by the **The Phyloinformatics Research Foundation, Inc.**

Some recent additions:

- **Historical biogeography and diversification of truffles in the Tuberaceae and their newly identified Southern hemisphere sister lineage**
  November 02, 2012
- **The analysis of rRNA sequence-structure in phylogenetics: An application to the family Pectinidae (Mollusca: Bivalvia)**
  November 02, 2012
- **Floral pseudomorphy leads to secondary specialization of Malagasy Dalechampia (Euphorbiaceae)**
  November 02, 2012
- **A dated phylogeny of the papaya family (Caricaceae) reveals the crop�s closest relatives and the family�s biogeographic history**
  November 02, 2012

Datação dessa filogenia utilizando o BLADJ a partir de tempos de divergência obtidos em bases de dados (e.g., TimeTree)

Bases de dados de atributos

Bases de dados online de atributos

Ecological Data Wiki

Welcome back, data wiki users

The Ecological Data Wiki has been updated to a newer version of Drupal. If you are a returning user, you may need to request a new password to log in.

Welcome to the Ecological Data Wiki

Ecological data is everywhere, but how do you find what you need and learn how to use it?

The Data Wiki helps scientists collaborate on gathering information about data and how to use it.

The site is a source for finding ecological datasets and quickly figuring out the best ways to use them.

Recent comments

- Can someone point me in the 8 months 1 week ago
- Thanks Jessica. I just 9 months 1 week ago
- Thank you so much for this 9 months 1 week ago

Bases de dados online de atributos

Ecological Data Wiki

ou may need to request a new password to log in.

eed

them.

Recent comments

- Can someone point me in the 8 months 1 week ago
- Thanks Jessica. I just 9 months 1 week ago
- Thank you so much for this 9 months 1 week ago

Quantifying and scaling global plant trait diversity

TRY is a network of vegetation scientists headed by
Future Earth and the Max Planck Institute for Biogeochemistry,
providing a global archive of curated plant traits.
The TRY database is a research platform of iDiv.

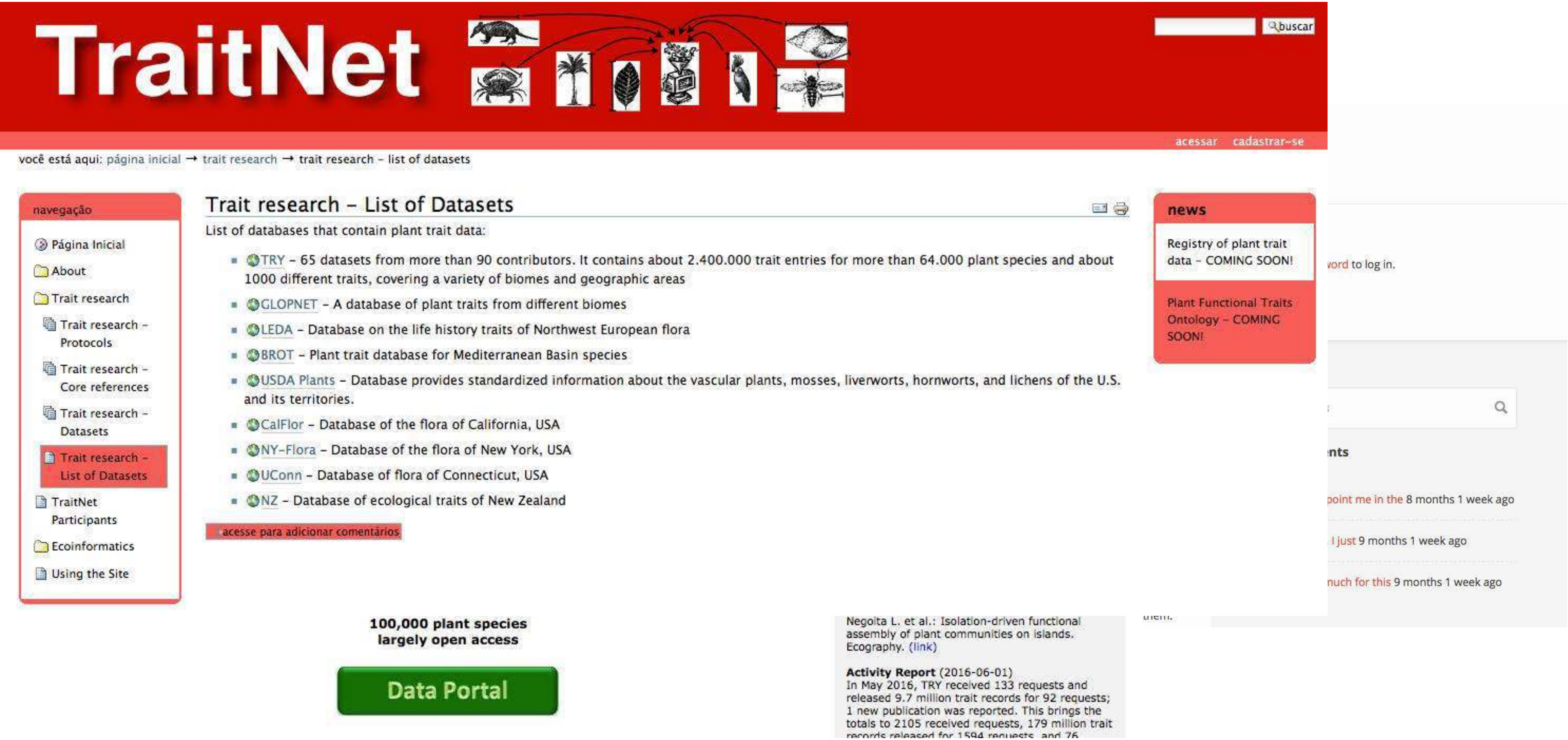
5.6 million trait records
100,000 plant species
largely open access

News

Paper published (2016-06-15)
Hébert F. et al.: Recovery of plant community functional traits following severe soil perturbation in plantations: a case-study, International Journal of Biodiversity Science, Ecosystem Services & Management. (link)

Paper published (2016-06-13)
Negoita L. et al.: Isolation-driven functional assembly of plant communities on islands. Ecography. (link)

Activity Report (2016-06-01)
In May 2016, TRY received 133 requests and released 9.7 million trait records for 92 requests; 1 new publication was reported. This brings the totals to 2105 received requests, 179 million trait records released for 1594 requests, and 76



você está aqui: página inicial → trait research → trait research - list of datasets

Trait research – List of Datasets

List of databases that contain plant trait data:

- TRY – 65 datasets from more than 90 contributors. It contains about 2.400.000 trait entries for more than 64.000 plant species and about 1000 different traits, covering a variety of biomes and geographic areas
- GLOPNET – A database of plant traits from different biomes
- LEDA – Database on the life history traits of Northwest European flora
- BROT – Plant trait database for Mediterranean Basin species
- USDA Plants – Database provides standardized information about the vascular plants, mosses, liverworts, hornworts, and lichens of the U.S. and its territories.
- CalFlor – Database of the flora of California, USA
- NY-Flora – Database of the flora of New York, USA
- UConn – Database of flora of Connecticut, USA
- NZ – Database of ecological traits of New Zealand

Registry of plant trait data - COMING SOON!

Plant Functional Traits Ontology - COMING SOON!

vord to log in.

nts

point me in the 8 months 1 week ago

I just 9 months 1 week ago

nuch for this 9 months 1 week ago

100,000 plant species largely open access

Negoita L. et al.: Isolation-driven functional assembly of plant communities on islands. Ecography. (link)

Activity Report (2016-06-01)
In May 2016, TRY received 133 requests and released 9.7 million trait records for 92 requests; 1 new publication was reported. This brings the totals to 2105 received requests, 179 million trait records released for 1594 requests, and 76

TraitNet

você está aqui: página inicial → trait research → trait research – list of datas

Trait research – List of Da

List of databases that contain plant trait

- TRY – 65 datasets from more th 1000 different traits, covering a va
- GLOPNET – A database of plant
- LEDA – Database on the life his
- BROT – Plant trait database for
- USDA Plants – Database provid and its territories.
- CalFlor – Database of the flora
- NY-Flora – Database of the flor
- UConn – Database of flora of C
- NZ – Database of ecological tra

100,000 pla
largely ope

SCALES

Securing the Conservation of biodiversity across Administrative Levels and spatial, temporal, and Ecological Scales

www.scales-project.net

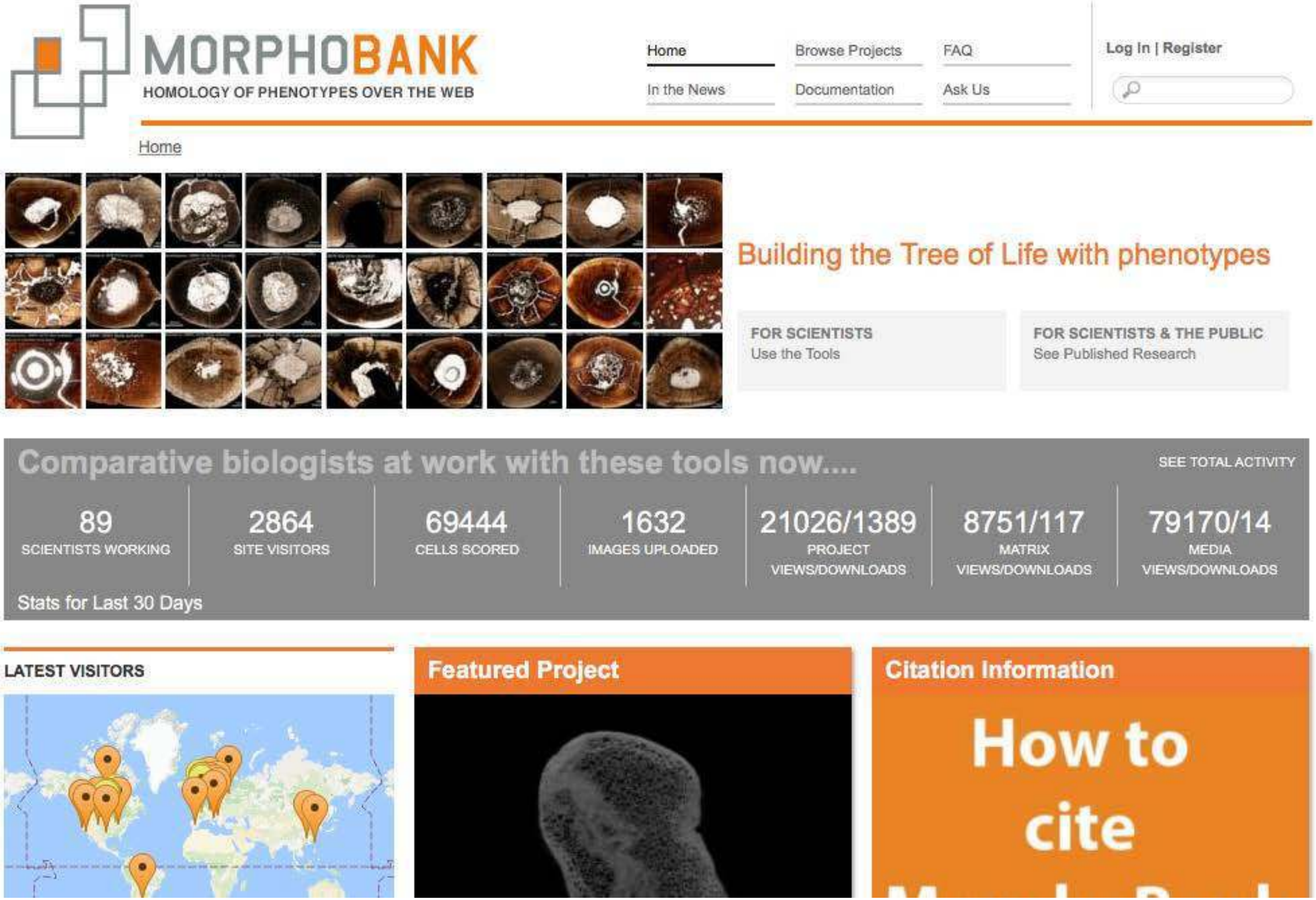
Species traits databases

Species traits describe characteristics of species and are relevant to population dynamics across space and time. They affect ecosystem processes indirectly through abiotic control and directly through changes in biotic control (e.g. predators, pathogens). On the other hand, they are also affected by other species traits and environmental conditions which are highly sensitive to global change pressures, such as land use changes and climate change. Since species traits seem to play one of the most important roles in ecosystem processes and are thus highly relevant for conservation planning, it is inevitable to have overviews and compilations for different species available.

For that purpose, existing databases and review literature on species traits and, in the following, dispersal related traits, were compiled. This compilation also includes derived traits like functional traits, area requirements, and habitat selectivity to dispersal and functional connectivity. These data provide an excellent overview on species traits, which can be used for parameterizing model inputs, and will help to assess and visualize scaling effects across taxa and spatial scales. Databases for plants, insects, reptiles, and birds across Europe are either available online or upon author request.

Ecography. (link)

Activity Report (2016-06-01)
In May 2016, TRY received 133 requests and released 9.7 million trait records for 92 requests; 1 new publication was reported. This brings the totals to 2105 received requests, 179 million trait records released for 1594 requests, and 76



MORPHOBANK
HOMOLOGY OF PHENOTYPES OVER THE WEB

Building the Tree of Life with phenotypes

FOR SCIENTISTS
Use the Tools

FOR SCIENTISTS & THE PUBLIC
See Published Research

Comparative biologists at work with these tools now....

89	2864	69444	1632	21026/1389	8751/117	79170/14
SCIENTISTS WORKING	SITE VISITORS	CELLS SCORED	IMAGES UPLOADED	PROJECT VIEWS/DOWNLOADS	MATRIX VIEWS/DOWNLOADS	MEDIA VIEWS/DOWNLOADS

Stats for Last 30 Days

LATEST VISITORS

Featured Project

Citation Information

How to cite

TraitBank

TraitBank® is a ***searchable***, comprehensive, open digital repository for organism traits, measurements, interactions and other facts for all taxa across the tree of life.

TraitBank currently features over ***11 million records*** related to more than ***330 attributes*** for ***1.7 million taxa*** obtained from over ***50 data sources***.

Become part of the EOL community!
Join EOL now
Already a member? Sign in

Mirrors : fishbase.org | fishbase.us | fishbase.de | fishbase.fr | fishbase.se | fishbase.tw | fishbase.cn | fishbase.sa | fishbase.ca

English | Español | Português (Br , Pt) | Français | Deutsch | Italiano | Nederlands | 简体中文 | 繁體中文 | 日本語 [More...]

FishBase

ver. (01/2016)

(33200 Species, 317200 Common names, 56900 Pictures, 52600 References, 2230 Collaborators, 700000 Visits/Month)

Home | FishBase Book | Best Photos | Hints | Guest Book | Download | Links | Fish Forum | Fish Quiz |

FishWatcher | Ichthyology Course| LarvalBase | Team | Collaborators | Quick Identification | Services

Common Name

is (e.g. rainbow trout) Search

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

中文 العربية Русский 日本語 हिन्दी Ελληνικά More scripts...

Scientific Name

☐ Advanced Match

Genus is (e.g. Rhinodon) Search

Species is (e.g. typus) ○ Random Species

Genus + Species Sp. ID Search

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Why name assessments may be different between FishBase and the independent Catalog of Fishes (Eschmeyer, 2014)

rganism traits, measurements, interactions and other

Search TraitBank

n *330 attributes* for *1.7 million taxa* obtained from

****** Please note that all manuscripts submitted after 1 January 2016 follow separate guidelines for supporting information!***

****** Ecological Archives through the end of 2015 will be hosted on FigShare once the transition to publishing with Wiley is completed. Thereafter, supplemental material may be hosted on Wiley Online, and/or data deposited with FigShare, Dryad, and other repositories.***

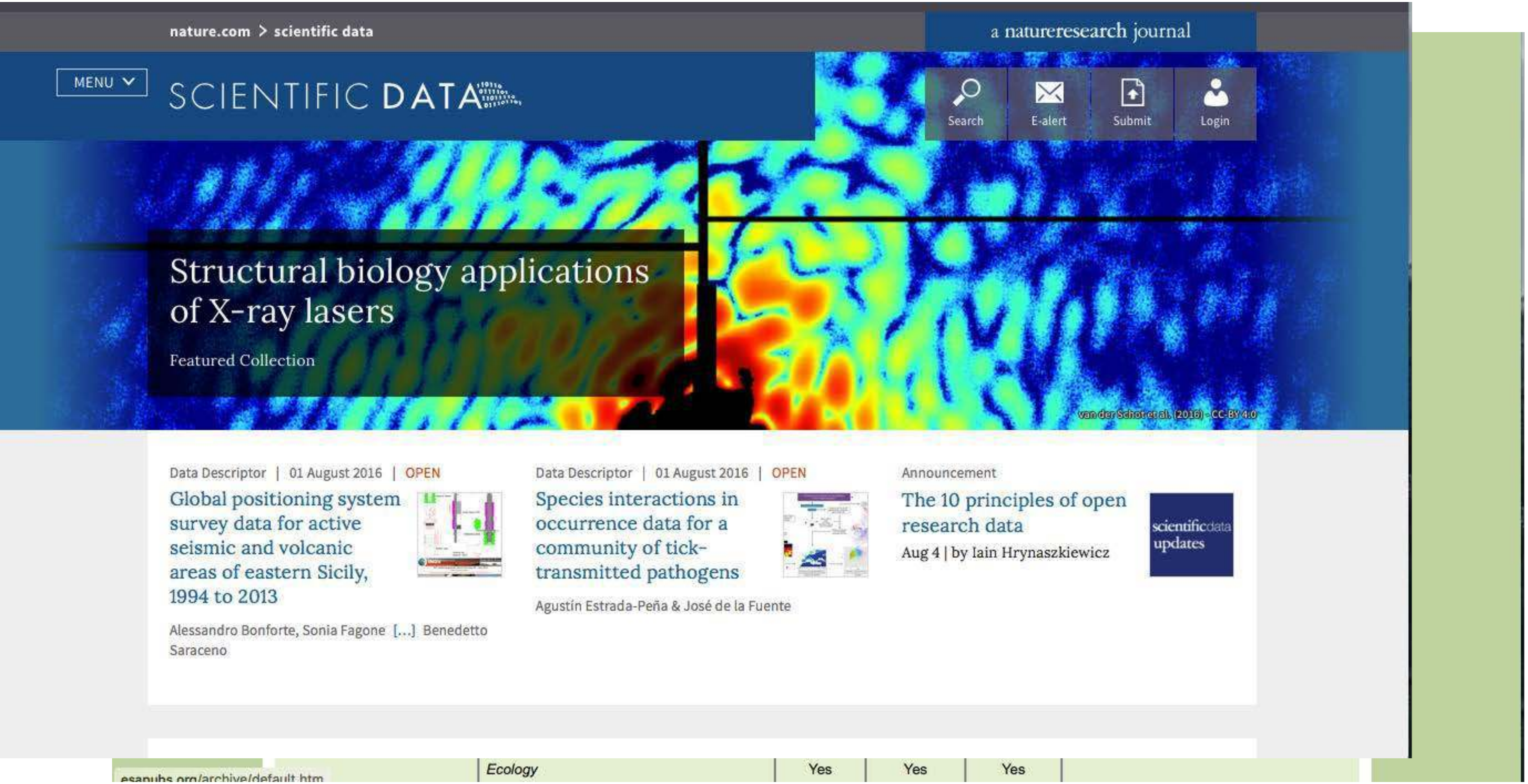
What is *Ecological Archives*?

Ecological Archives publishes materials that are supplemental to articles that appear in the ESA journals (*Ecology*, *Ecological Applications*, *Ecological Monographs*, *Ecosphere*, *Ecosystem Health and Sustainability* and *Bulletin of the Ecological Society of America*), as well as peer-reviewed data papers with abstracts published in the printed journals. *Ecological Archives* is published in digital, Internet-accessible form.

Three kinds of publications appear in *Ecological Archives*: appendices, supplements, and data papers. **However**, these types *vary* among the journals.

TABLE 1. *Ecological Archives* categories published by journal.

	Appendices	Supplements	Data Papers
Ecosphere	Yes†	Yes	No
Ecology	Yes	Yes	Yes



Structural biology applications of X-ray lasers

Featured Collection

Data Descriptor | 01 August 2016 | OPEN

Global positioning system survey data for active seismic and volcanic areas of eastern Sicily, 1994 to 2013

Alessandro Bonforte, Sonia Fagone [...] Benedetto Saraceno

Data Descriptor | 01 August 2016 | OPEN

Species interactions in occurrence data for a community of tick-transmitted pathogens

Agustín Estrada-Peña & José de la Fuente

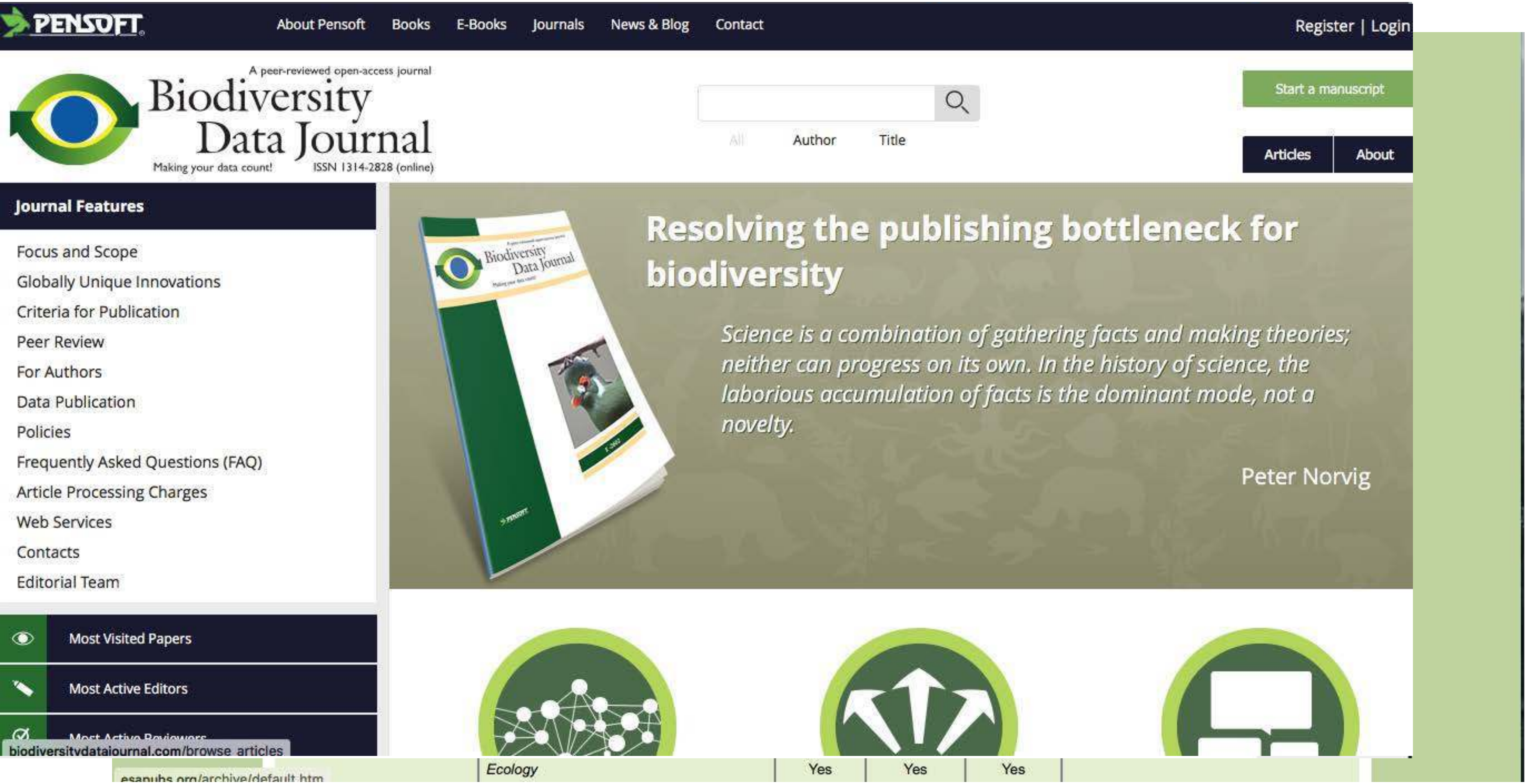
Announcement

The 10 principles of open research data

Aug 4 | by Iain Hrynaszkiewicz

esapubs.org/archive/default.htm

Ecology	Yes	Yes	Yes



A peer-reviewed open-access journal

Biodiversity Data Journal

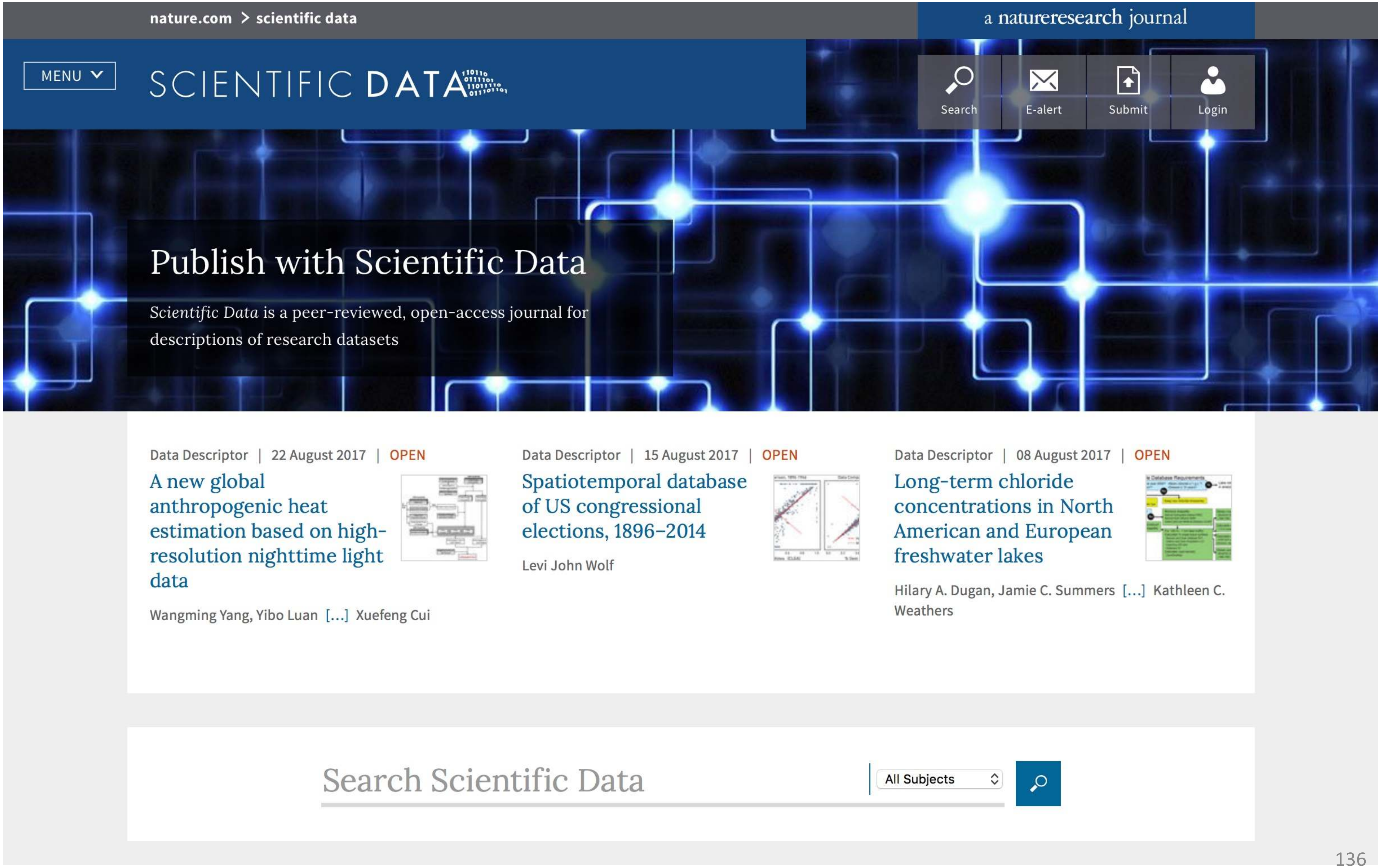
Making your data count!

ISSN 1314-2828 (online)

Resolving the publishing bottleneck for biodiversity

Science is a combination of gathering facts and making theories; neither can progress on its own. In the history of science, the laborious accumulation of facts is the dominant mode, not a novelty.

Peter Norvig



Publish with Scientific Data

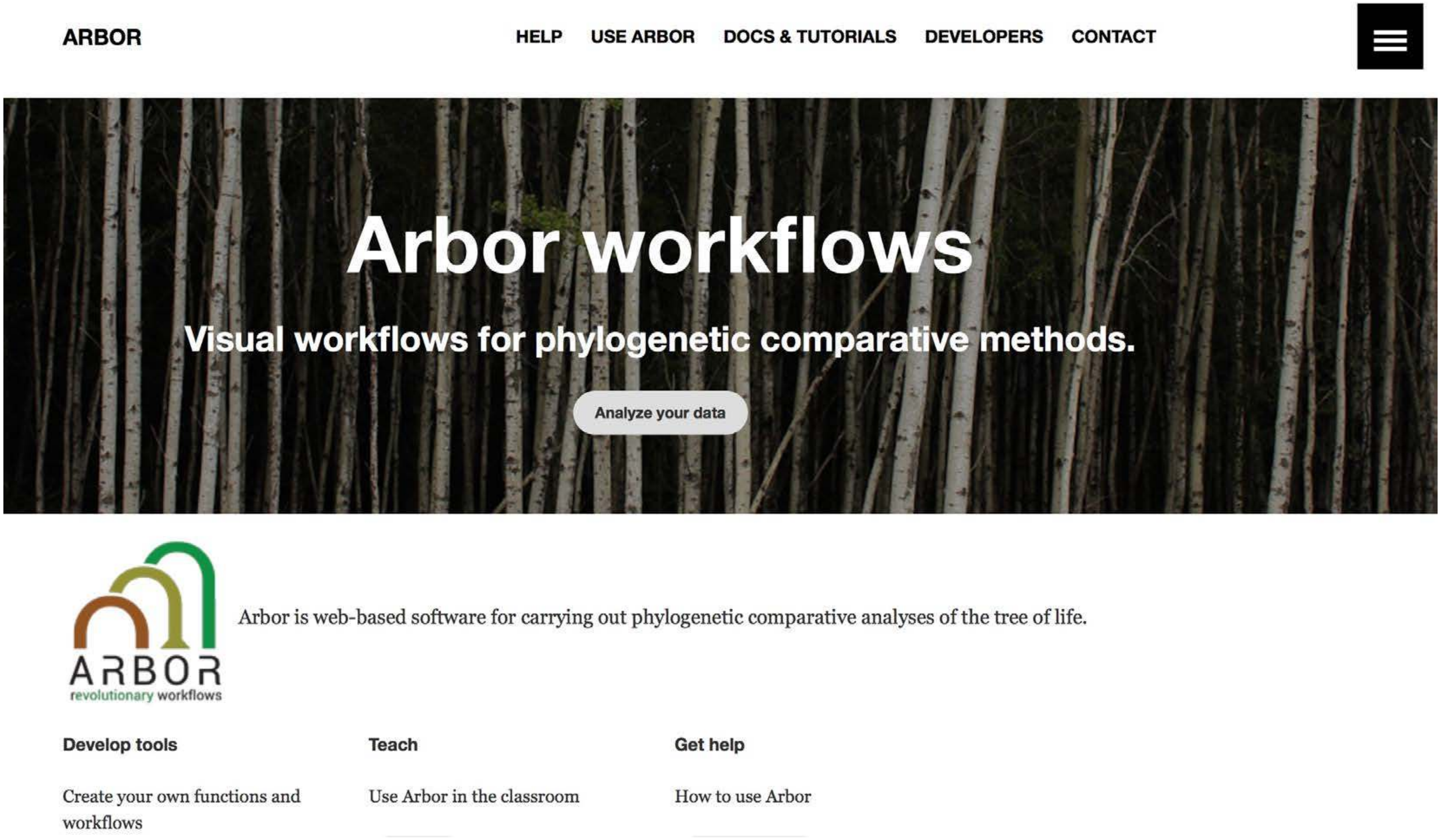
Scientific Data is a peer-reviewed, open-access journal for descriptions of research datasets

Data Descriptor | 22 August 2017 | OPEN
A new global anthropogenic heat estimation based on high-resolution nighttime light data
Wangming Yang, Yibo Luan [...] Xuefeng Cui

Data Descriptor | 15 August 2017 | OPEN
Spatiotemporal database of US congressional elections, 1896–2014
Levi John Wolf

Data Descriptor | 08 August 2017 | OPEN
Long-term chloride concentrations in North American and European freshwater lakes
Hilary A. Dugan, Jamie C. Summers [...] Kathleen C. Weathers

Realizando análises comparativas na nuvem



Arbor workflows

Visual workflows for phylogenetic comparative methods.

Arbor is web-based software for carrying out phylogenetic comparative analyses of the tree of life.

Develop tools

Create your own functions and workflows

Teach

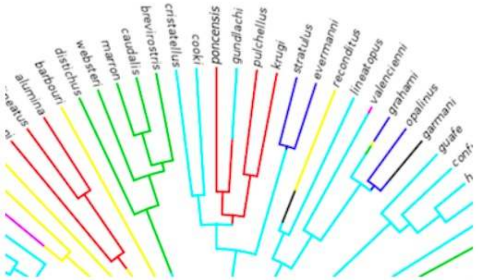
Use Arbor in the classroom

Get help

How to use Arbor

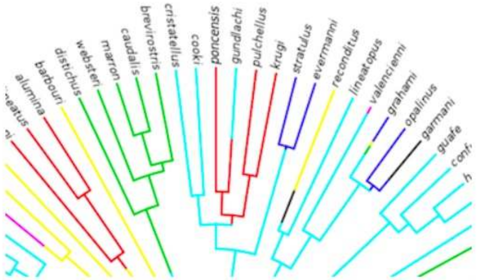
Arbor Apps

Comparative methods using a drag-and-drop interface



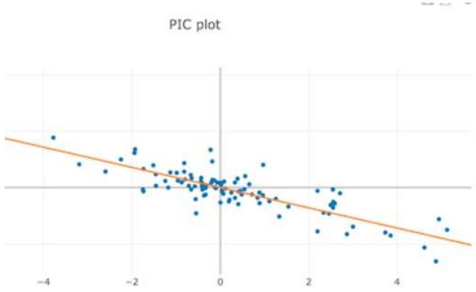
May 31, 2017

Plot tree using d3



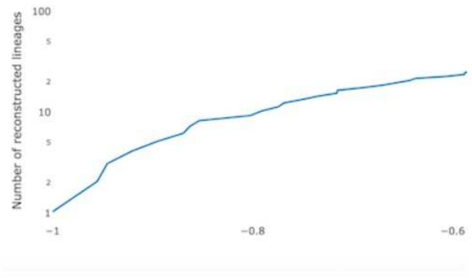
May 31, 2017

Stochastic character mapping using simmap



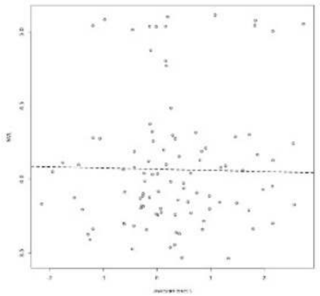
May 31, 2017

Phylogenetic independent contrasts



May 31, 2017

Lineage-through-time plots



May 30, 2017

PGLS - two variables only (for now!)

May 29, 2017

Ancestral state reconstruction

Reconstruct character evolution on a phylogenetic tree.

Browse or drop files here

Tree: (none)

Table: (none)

Load data above to proceed

Go!

May 28, 2017

Phylogenetic signal

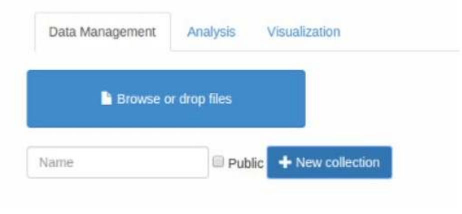
Tutorials

A collection of tutorials for Arbor.

January 08, 2017

Web services 2

What are web services?



January 06, 2017

Creating new collections and functions

How to make your own collection of Arbor functions and workflows

January 05, 2017

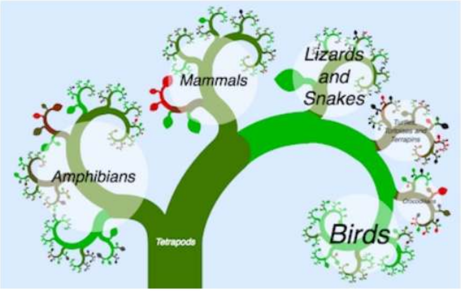
Using the Arbor webapp

How to run functions for comparative analyses in Arbor

April 18, 2016

Timetrees

Creating timetrees



April 18, 2016

Megaphylogenies

What is a megaphylogeny?

March 11, 2016

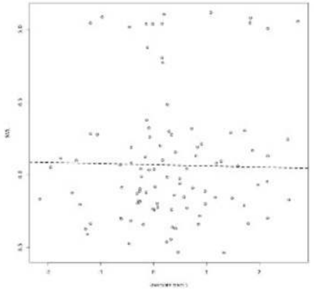
Web services 1

What are web services?

March 03, 2016

Arbor apps

What are Arbor apps?



July 17, 2015

PGLS

Use PGLS to test for a correlation

Desenho de estudos comparativos

Como maximizar a amostragem e lidar com *mismatch* de dados

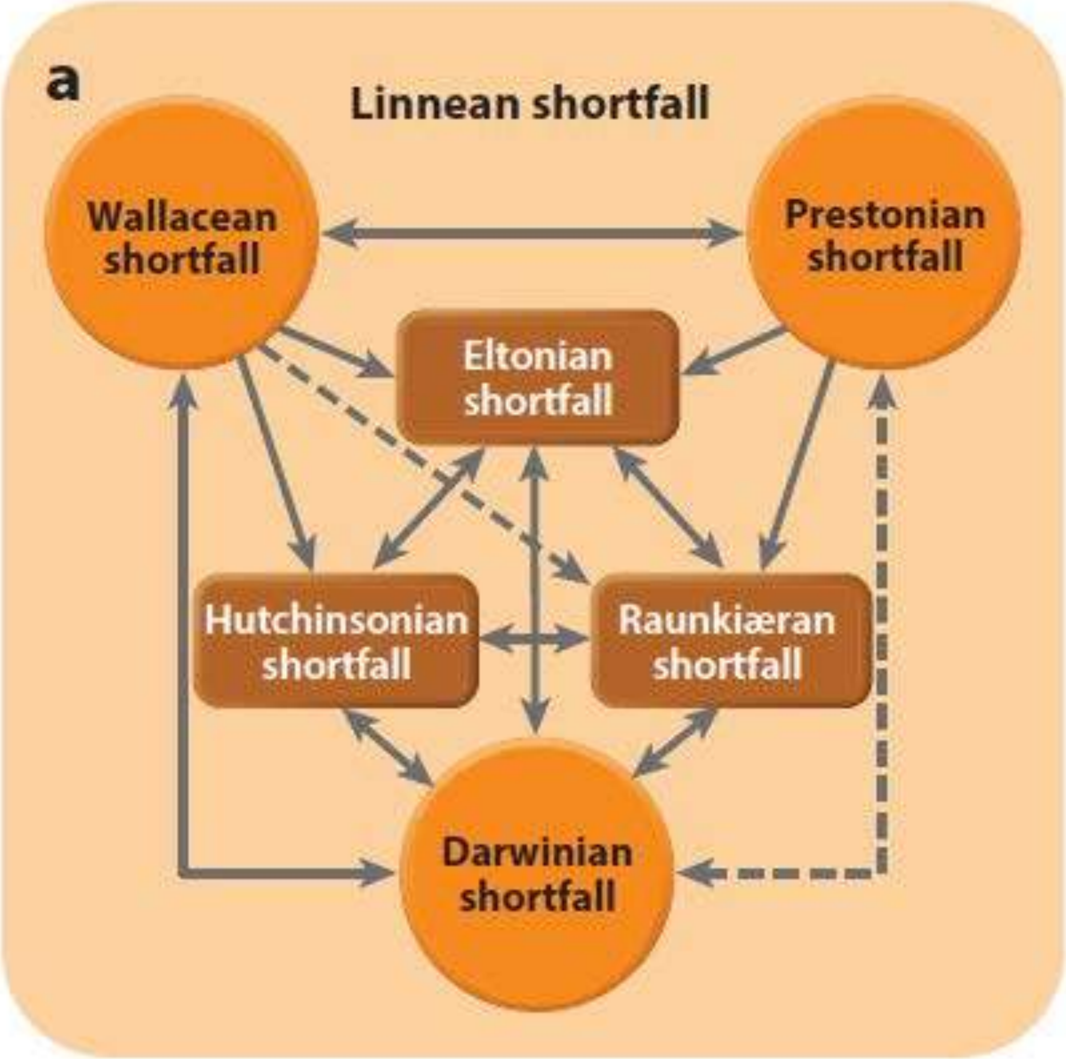
Problemas mais comuns

- Falta completa de dados (filogenéticos e traits)

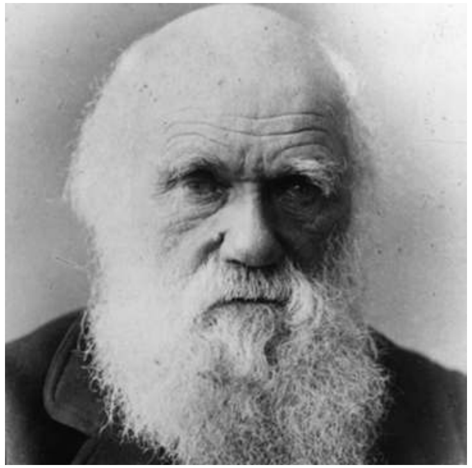
- **Descasamento (*mismatch*) entre disponibilidade de dados de traits e de filogenias**

- Dados existem mas não estão acessíveis

- Déficits de conhecimento sobre biodiversidade: Darwiniano e Raunkiaeriano
  - Dificultam ganho de conhecimento sobre evolução de atributos
  - Impedimento para estudos comparativos



Déficit Raunkiaeriano
Falta de informações sobre atributos das espécies



Déficit Darwiniano
Falta de informações sobre relações entre espécies

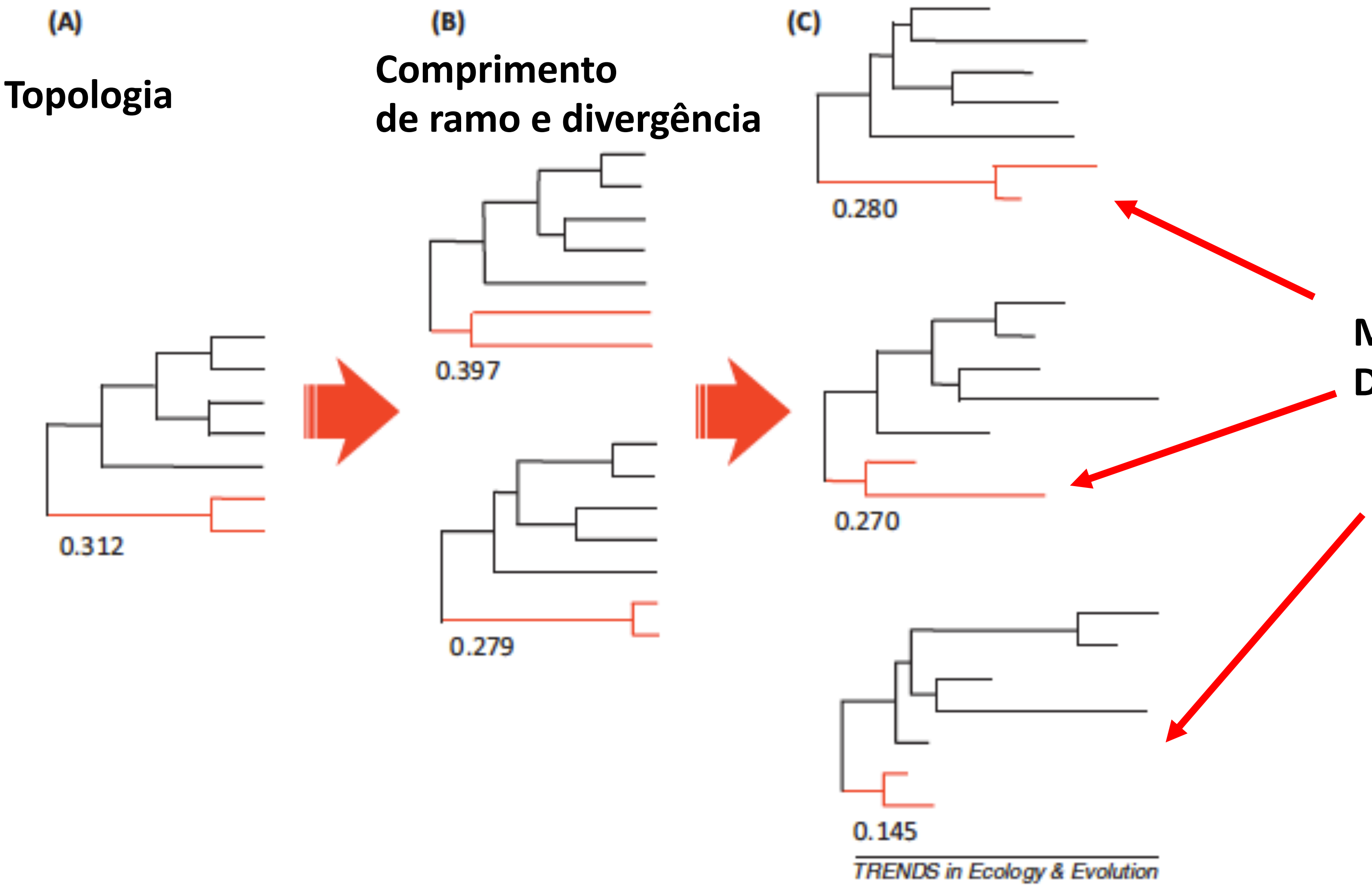
Hortal et al. 2015

Opinion

Cell PRESS

Darwinian shortfalls in biodiversity conservation

José Alexandre F. Diniz-Filho[1], Rafael D. Loyola[1], Pasquale Raia[2], Arne O. Mooers[3], and Luis M. Bini[1]



TRENDS in Ecology & Evolution

Topologia

Comprimento
de ramo e divergência

Ignorâncias

Modelo evolutivo
De atributos


Meu deus o que eu faço agora? Meu projeto foi pras cucuias? Perdi meu doutorado? Quem poderá me socorrer?!...Oh oh!!

Meu deus
Meu proje
Perdi meu
poderá me

Pacotes phyndr e taxonlookup

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015 doi: 10.1111/2041-210X.12517

APPLICATION

A simple approach for maximizing the overlap of phylogenetic and comparative data

Matthew W. Pennell[1,2†], Richard G. FitzJohn[3†] and William K. Cornwell[4,5†]

- Maximiza a amostragem trocando espécies na árvore, para os quais não há dados de atributos, por espécies ”filogeneticamente equivalentes” sem alterar a estrutura da filogenia
  - Baseado em ranking taxonômico (ou só os nomes dos gêneros) ou uma árvore topológica
  - Alternativa a métodos de aleatorização ou de imputação de dados faltantes

Ecology and Evolution

Open Access

Filling the gap in functional trait databases: use of ecological hypotheses to replace missing data

Simon Taugourdeau[1,2,3], Jean Villerd[1,2], Sylvain Plantureux[1,2], Olivier Huguenin-Elie[3] &
Bernard Amiaud[4,5]

Ecography 37: 105–110, 2014
doi: 10.1111/j.1600-0587.2013.00528.x

Subject Editor: Nathan J. Sanders. Accepted 26 July 2013

Phylogenetic imputation of plant functional trait databases

Nathan G. Swenson

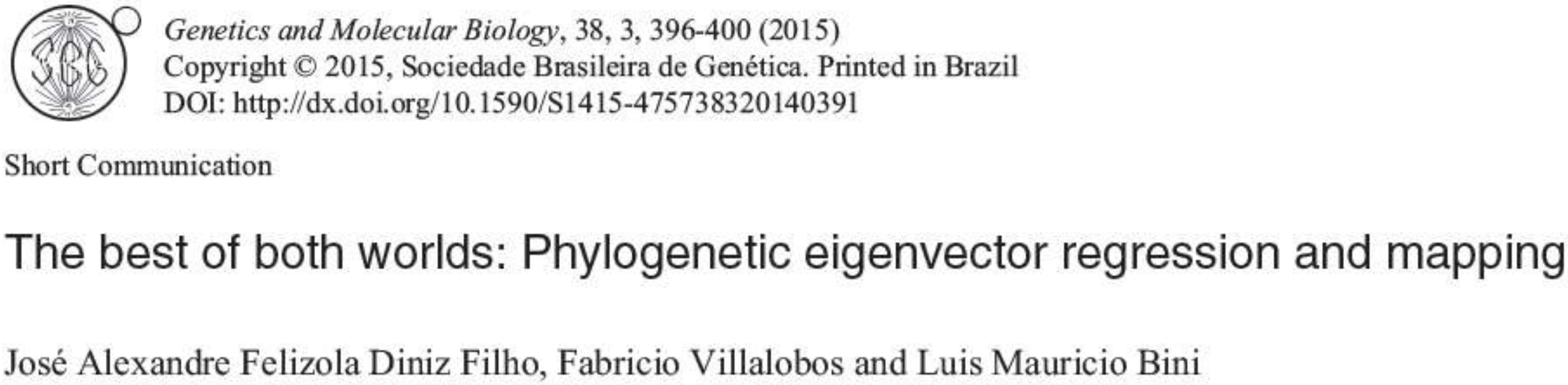
Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2013, **4**, 1120–1131 doi: 10.1111/2041-210X.12111

Phylogenetic eigenvector maps: a framework to model and predict species traits

Guillaume Guénard[1]*, Pierre Legendre[1] and Pedro Peres-Neto[2]

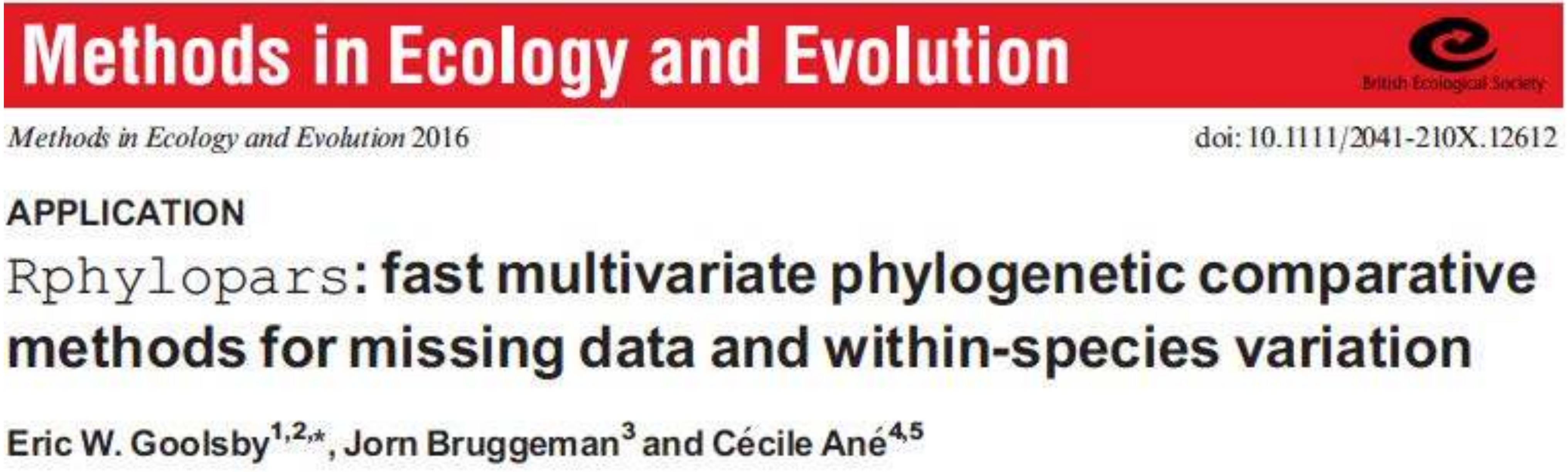


Genetics and Molecular Biology, 38, 3, 396-400 (2015)
Copyright © 2015, Sociedade Brasileira de Genética. Printed in Brazil
DOI: http://dx.doi.org/10.1590/S1415-475738320140391

Short Communication

The best of both worlds: Phylogenetic eigenvector regression and mapping

José Alexandre Felizola Diniz Filho, Fabricio Villalobos and Luis Mauricio Bini



Methods in Ecology and Evolution

Methods in Ecology and Evolution 2016 doi: 10.1111/2041-210X.12612

APPLICATION

`Rphylopars`: fast multivariate phylogenetic comparative methods for missing data and within-species variation

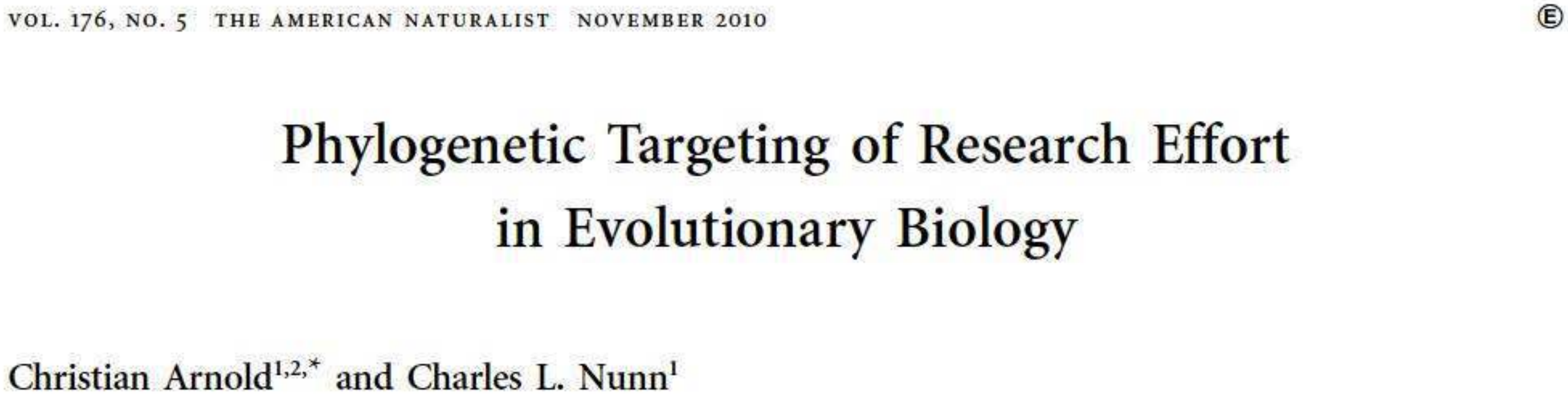
Eric W. Goolsby[1,2,*], Jorn Bruggeman[3] and Cécile Ané[4,5]

Methods in Ecology and Evolution

Methods in Ecology and Evolution 2014 doi: 10.1111/2041-210X.12232

Imputation of missing data in life-history trait datasets: which approach performs the best?

Caterina Penone[1*], Ana D. Davidson[2,3], Kevin T. Shoemaker[2], Moreno Di Marco[4], Carlo Rondinini[4], Thomas M. Brooks[5], Bruce E. Young[3], Catherine H. Graham[2] and Gabriel C. Costa[1]

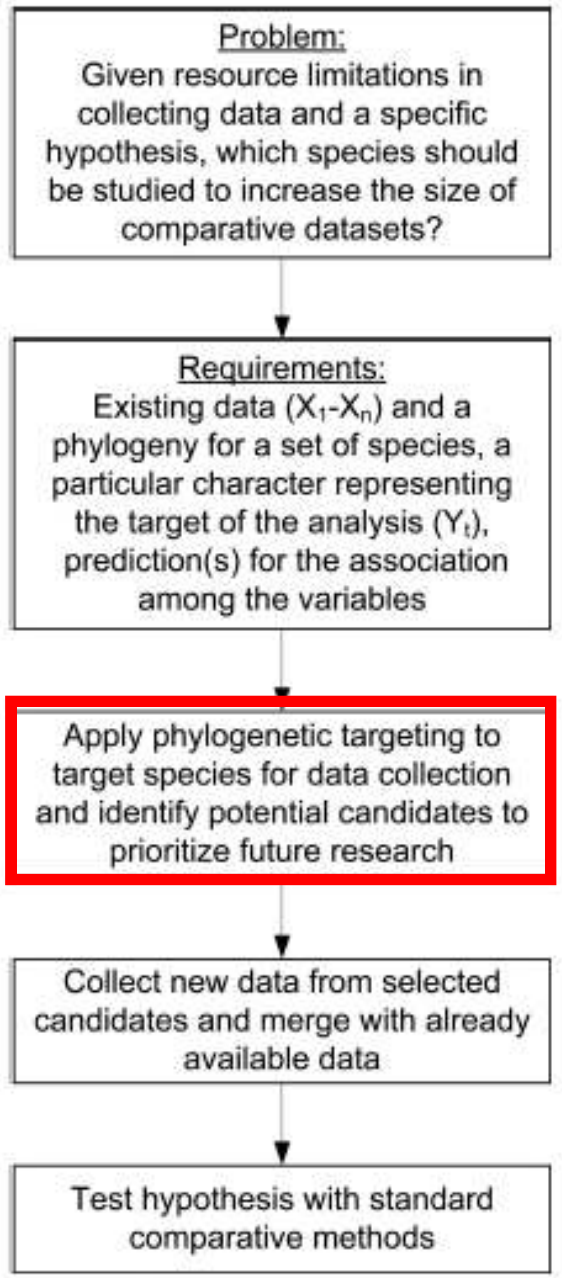


VOL. 176, NO. 5 THE AMERICAN NATURALIST NOVEMBER 2010 Ⓔ

Phylogenetic Targeting of Research Effort in Evolutionary Biology

Christian Arnold[1,2,*] and Charles L. Nunn[1]

Workflow PhyloTargeting



Problem:
Given resource limitations in collecting data and a specific hypothesis, which species should be studied to increase the size of comparative datasets?

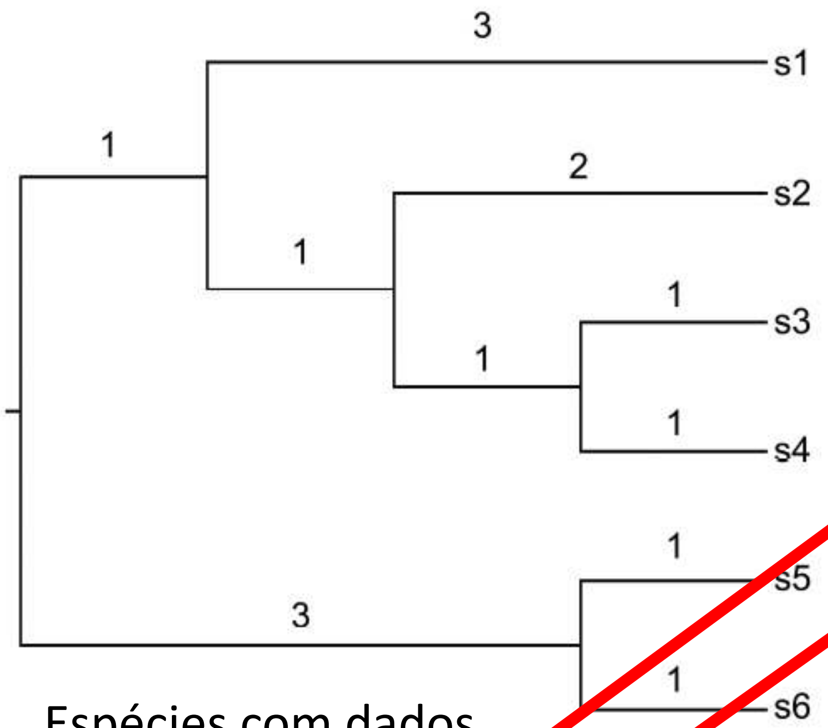
Requirements:
Existing data (X_1-X_n) and a phylogeny for a set of species, a particular character representing the target of the analysis (Y_t), prediction(s) for the association among the variables

Apply phylogenetic targeting to target species for data collection and identify potential candidates to prioritize future research

Collect new data from selected candidates and merge with already available data

Test hypothesis with standard comparative methods

3 Atributos contínuos



	X_1	X_2	Y_t	B_t
s1	-0.4	10.5	5.5	1
s2	0.1	7.5	?	0
s3	0.4	9.0	?	0
s4	0.8	7.8	?	0
s5	0.6	25.3	6.9	1
s6	0.2	20.1	?	0

Espécies com dados faltantes
Para variável resposta

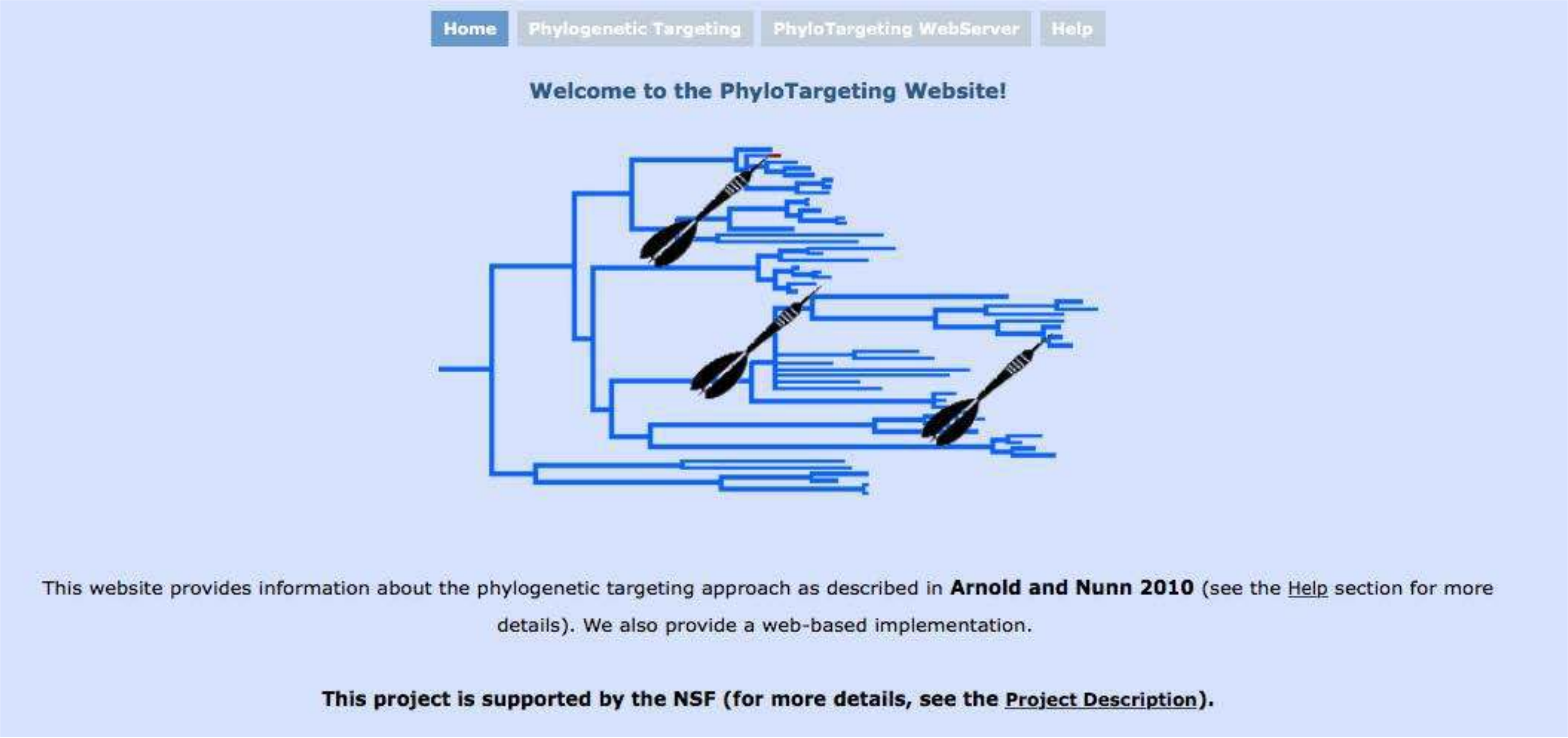
Como otimizar o levantamento de dados?

Como o Phylotargeting funciona?

- O que é preciso?
    - Identificação de variáveis resposta e preditoras
    - Hipótese
    - Filogenia
- Objetivo: Identificar espécies para as quais devem ser coletadas dados da variável resposta (Y_t) usando informação de variáveis preditoras de espécies já estudadas
- Atribui um *score* pra cada par de espécies que incorpora filogenia e valores de variáveis relevantes para testar hipóteses, envolvendo especificamente as variáveis preditoras e resposta
- Espécies com scores altos devem ser preferidas para coleta de novos dados

Como o Phylotargeting funciona?

- Espécies que diferem mais na(s) variável(is) resposta são as preferidas, de maneira a ampliar o *range* de variação nesta variável, aumentando o poder do teste
- Assume que as variáveis estão relacionadas linearmente
- Permite variáveis discretas e contínuas
- Uma ou mais preditoras (covariáveis)
- Mais adequado pra desenhos e perguntas que envolvem modelos lineares (PGLS, correlação, ANOVA, PGLMM etc)

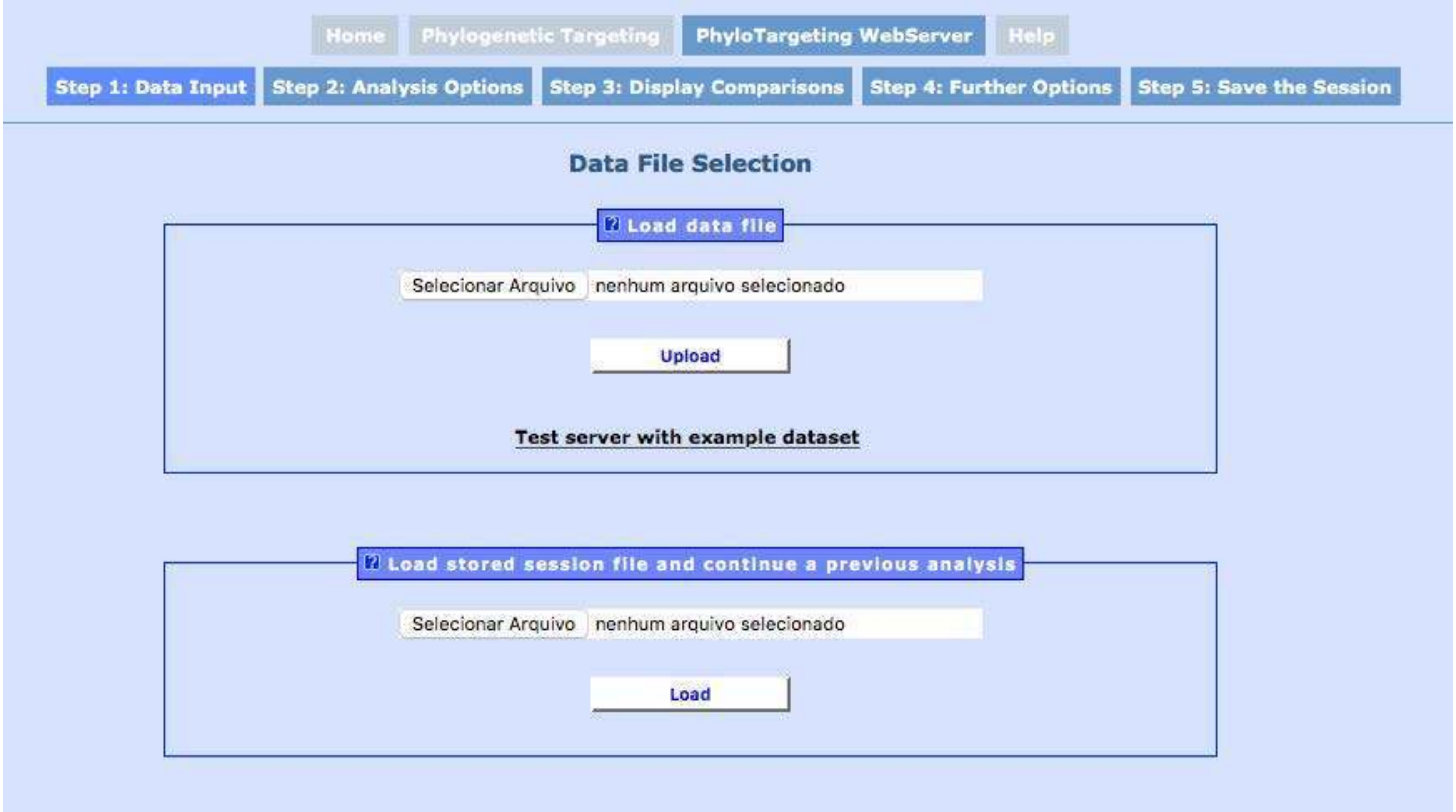


Welcome to the PhyloTargeting Website!

This website provides information about the phylogenetic targeting approach as described in **Arnold and Nunn 2010** (see the Help section for more details). We also provide a web-based implementation.

This project is supported by the NSF (for more details, see the Project Description).

http://phylotargeting.nunn-lab.org/index.html


Home
Phylogenetic Targeting
PhyloTargeting WebServer
Help
Step 1: Data Input
Step 2: Analysis Options
Step 3: Display Comparisons
Step 4: Further Options
Step 5: Save the Session
Data File Selection
Load data file
Selecionar Arquivo
nenhum arquivo selecionado
Upload
Test server with example dataset
Load stored session file and continue a previous analysis
Selecionar Arquivo
nenhum arquivo selecionado
Load


YES!
AGORA BORA BOTAR A MÃO NA MASSA!

Prática

- Introdução ao R
  - Importação e manipulação de dados
  - Importação e manipulação de ≠ formatos de filogenias
- Timetree.org
- Phyndr
- Phylotargeting
- rotl
- Rphylopars