

The guinea-pig is not a rodent

**Anna Maria D'Erchia^{*†}, Carmela Gissi^{*†},
Graziano Pesole[‡], Cecilia Saccone^{*§} & Ulfur Arnason[†]**

^{*} Dipartimento di Biochimica e Biologia Molecolare, Università di Bari,
70125 Bari, Italy

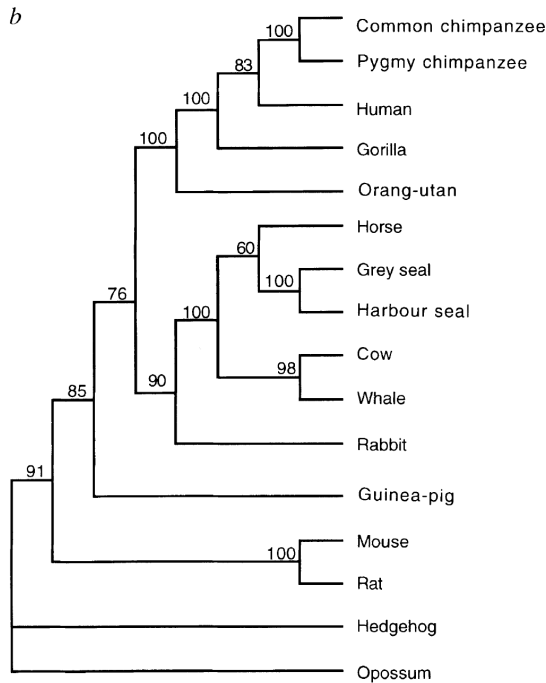
[†] Department of Evolutionary Molecular Systematics, University of Lund,
Sölvegatan 29, S-22362 Lund, Sweden

[‡] Dipartimento di Biologia DBAF, Università della Basilicata, 30100
Potenza, Italy

[§] Centro di Studio sui Mitocondri e Metabolismo Energetico,
Consiglio Nazionale delle Ricerche, 70125 Bari, Italy

IN 1991 Graur *et al.* raised the question of whether the guinea-pig, *Cavia porcellus*, is a rodent¹. They suggested that the guinea-pig and myomorph rodents diverged before the separation between myomorph rodents and a lineage leading to primates and artiodactyls. Several findings have since been reported, both for and against this phylogeny, thereby highlighting the issue of the validity of molecular analysis in mammalian phylogeny. Here we present findings based on the sequence of the complete mitochondrial genome of the guinea-pig, which strongly contradict rodent monophyly. The conclusions are based on the cumulative evidence provided by orthologically inherited genes and the use of three different analytical methods, none of which joins the guinea-pig with myomorph rodents. In addition to the phylogenetic conclusions, we also draw attention to several factors that are important for the validity of phylogenetic analysis based on molecular data.

**D'Erchia et al. 1990
Nature 381, 1996**



The guinea-pig is not a rodent

**Anna Maria D'Erchia^{*†}, Carmela Gissi^{*†},
Graziano Pesole[‡], Cecilia Saccone^{*§} & Ulfur Arnason[†]**

^{*} Dipartimento di Biochimica e Biologia Molecolare, Università di Bari,
70125 Bari, Italy

[†] Department of Evolutionary Molecular Systematics, University of Lund,
Sölvegatan 29, S-22362 Lund, Sweden

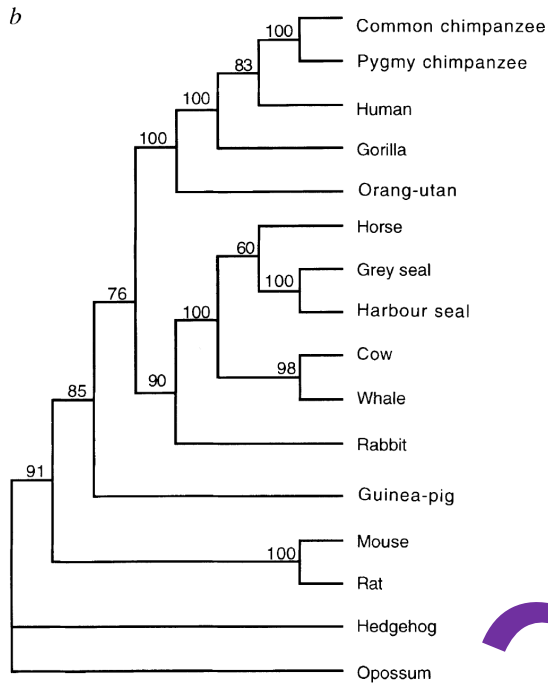
[‡] Dipartimento di Biologia DBAF, Università della Basilicata, 30100
Potenza, Italy

[§] Centro di Studio sui Mitocondri e Metabolismo Energetico,
Consiglio Nazionale delle Ricerche, 70125 Bari, Italy

IN 1991 Graur *et al.* raised the question of whether the guinea-pig, *Cavia porcellus*, is a rodent¹. They suggested that the guinea-pig and myomorph rodents diverged before the separation between myomorph rodents and a lineage leading to primates and artiodactyls. Several findings have since been reported, both for and against this phylogeny, thereby highlighting the issue of the validity of molecular analysis in mammalian phylogeny. Here we present findings based on the sequence of the complete mitochondrial genome of the guinea-pig, which strongly contradict rodent monophyly. The conclusions are based on the cumulative evidence provided by orthologically inherited genes and the use of three different analytical methods, none of which joins the guinea-pig with myomorph rodents. In addition to the phylogenetic conclusions, we also draw attention to several factors that are important for the validity of phylogenetic analysis based on molecular data.



**D'Erchia et al. 1990
Nature 381, 1996**



The guinea-pig is not a rodent

**Anna Maria D'Erchia^{*†}, Carmela Gissi^{*†},
Graziano Pesole[‡], Cecilia Saccone^{*§} & Ulfur Arnason[†]**

^{*} Dipartimento di Biochimica e Biologia Molecolare, Università di Bari,
70125 Bari, Italy

[†] Department of Evolutionary Molecular Systematics, University of Lund,
Sölvegatan 29, S-22362 Lund, Sweden

[‡] Dipartimento di Biologia DBAF, Università della Basilicata, 30100
Potenza, Italy

[§] Centro di Studio sui Mitocondri e Metabolismo Energetico,
Consiglio Nazionale delle Ricerche, 70125 Bari, Italy

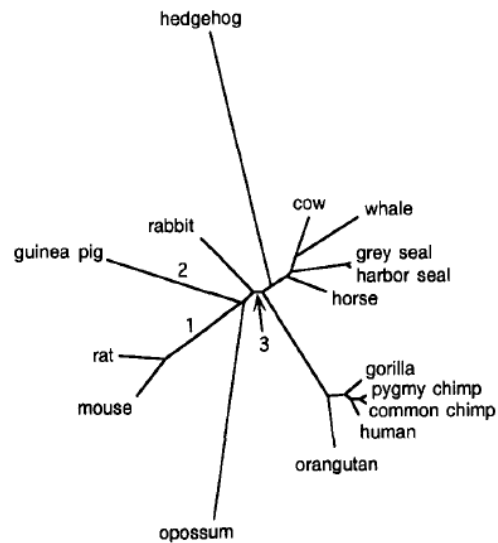
IN 1991 Graur *et al.* raised the question of whether the guinea-pig, *Cavia porcellus*, is a rodent¹. They suggested that the guinea-pig and myomorph rodents diverged before the separation between myomorph rodents and a lineage leading to primates and artiodactyls. Several findings have since been reported, both for and against this phylogeny, thereby highlighting the issue of the validity of molecular analysis in mammalian phylogeny. Here we present findings based on the sequence of the complete mitochondrial genome of the guinea-pig, which strongly contradict rodent monophyly. The conclusions are based on the cumulative evidence provided by orthologically inherited genes and the use of three different analytical methods, none of which joins the guinea-pig with myomorph rodents. In addition to the phylogenetic conclusions, we also draw attention to several factors that are important for the validity of phylogenetic analysis based on molecular data.

**D'Erchia et al. 1990
Nature 381, 1996**



The guinea-pig is ~~not~~ a rodent

B



Journal of Mammalian Evolution, Vol. 4, No. 2, 1997

Are Guinea Pigs Rodents? The Importance of Adequate Models in Molecular Phylogenetics

Jack Sullivan^{1,2} and David L. Swofford¹



Posterior Predictive Simulation Workshop:

Models matter to evolutionary analysis!

Bryan Carstens, Noah Reid, Lyndon Coghill, Anthony Barley, Drew Duckett

Maximum likelihood inference of phylogeny....

$$L_H = \Pr(\text{Data} \mid \text{Hypothesis})$$

$$L_{(\tau)} = P(D \mid \tau, m)$$

$$L_{(\tau)} = \prod_{i=1}^s \Pr(D^i \mid \tau) = \prod_{i=1}^s L^i(\tau)$$

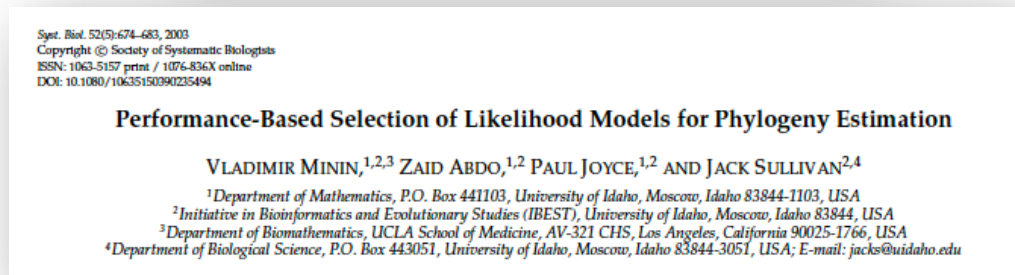
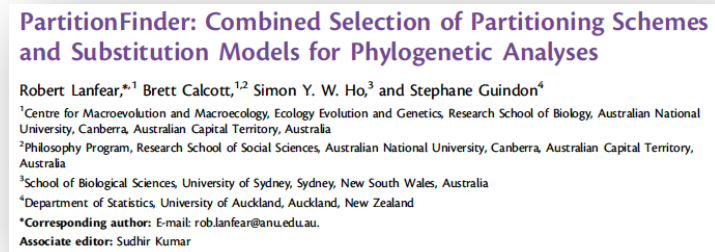
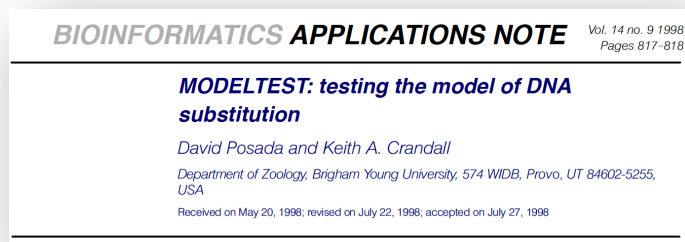
Single-site Likelihood



Maximum likelihood inference of phylogeny....

$$L_H = \Pr (\text{Data} \mid \text{Hypothesis})$$

- formalized model selection (LRTs, AIC, BIC, etc...)
- statistical assessment of model fit



Statistical Tests of Models of DNA Substitution

Nick Goldman

University Museum of Zoology, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

- formalized model selection (LRTs, AIC, BIC, etc...)
- statistical assessment of model fit

Syst. Biol. 49(4):652–670, 2000

Likelihood-Based Tests of Topologies in Phylogenetics

NICK GOLDMAN,¹ JON P. ANDERSON,² AND ALLEN G. RODRIGO³

¹*University Museum of Zoology, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK;
E-mail: N.Goldman@zoo.cam.ac.uk*

²*Department of Molecular Biotechnology, University of Washington, Seattle, Washington USA*

³*School of Biological Sciences, University of Auckland, Auckland, New Zealand*

Comparative Phylogeography of Mesoamerican Highland Rodents: Concerted versus Independent Response to Past Climatic Fluctuations

Jack Sullivan,^{1,*} Elizabeth Arellano,^{2,3} and Duke S. Rogers²

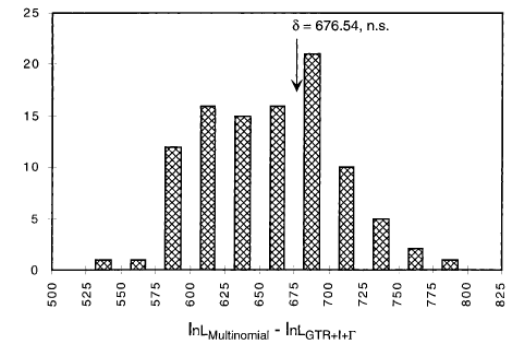


Figure 5: Parametric bootstrap test of the goodness-of-fit between the GTR+I+ Γ model of sequence evolution and the *Reithrodontomys sumichrasti* cyt b data. One hundred replicate data sets were simulated under the GTR+I+ Γ model, and the distribution of the difference between unconstrained (multinomial) lnL and lnL of the ML tree under the model is shown. The difference observed in the real data falls in the middle of the distribution generated under the null hypothesis of a perfect match between model and data; the GTR+I+ Γ model fits the data quite well.

1. simulate sequence data using selected model on ML estimate of phylogeny (match empirical dimensions)
2. estimate phylogenies from simulated data
3. calculate $-\ln L$ of each tree using the selected model of sequence evolution
4. calculate multinomial $-\ln L$
5. construct a null distribution of the difference between the multinomial $-\ln L$ and the $-\ln L$ using the selected model
6. compare this same difference from the empirical data to the distribution

Comparative Phylogeography of Mesoamerican Highland Rodents: Concerted versus Independent Response to Past Climatic Fluctuations

Jack Sullivan,^{1,*} Elizabeth Arellano,^{2,3} and Duke S. Rogers²

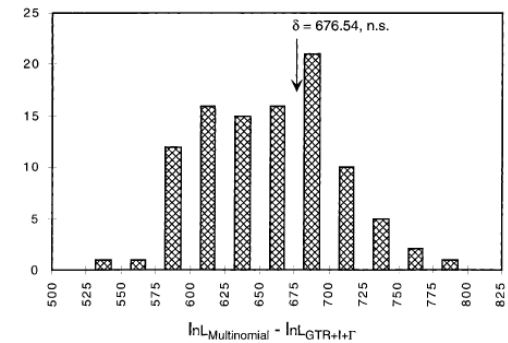


Figure 5: Parametric bootstrap test of the goodness-of-fit between the GTR+I+Γ model of sequence evolution and the *Reithrodontomys sumichrasti* cyt b data. One hundred replicate data sets were simulated under the GTR+I+Γ model, and the distribution of the difference between unconstrained (multinomial) lnL and lnL of the ML tree under the model is shown. The difference observed in the real data falls in the middle of the distribution generated under the null hypothesis of a perfect match between model and data; the GTR+I+Γ model fits the data quite well.

1. simulate sequence data using selected model on ML estimate of phylogeny (match empirical dimensions)
2. estimate phylogenies from simulated data
3. calculate $-\ln L$ of each tree using the selected model of sequence evolution
4. calculate multinomial $-\ln L$
5. construct a null distribution of the difference between the multinomial $-\ln L$ and the $-\ln L$ using the selected model
6. compare this same difference from the empirical data to the distribution

Comparative Phylogeography of Mesoamerican Highland Rodents: Concerted versus Independent Response to Past Climatic Fluctuations

Jack Sullivan,^{1,*} Elizabeth Arellano,^{2,3} and Duke S. Rogers²

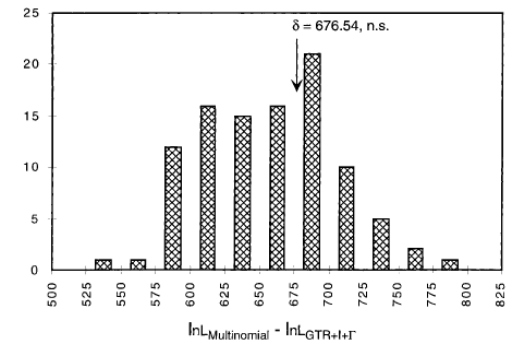


Figure 5: Parametric bootstrap test of the goodness-of-fit between the GTR+I+Γ model of sequence evolution and the *Reithrodontomys sumichrasti* cyt b data. One hundred replicate data sets were simulated under the GTR+I+Γ model, and the distribution of the difference between unconstrained (multinomial) ln likelihoods and ln likelihood of the ML tree under the model is shown. The difference observed in the real data falls in the middle of the distribution generated under the null hypothesis of a perfect match between model and data; the GTR+I+Γ model fits the data quite well.

1. simulate sequence data using selected model on ML estimate of phylogeny (match empirical dimensions)
2. estimate phylogenies from simulated data
3. calculate $-\ln L$ of each tree using the selected model of sequence evolution
4. calculate multinomial $-\ln L$
5. construct a null distribution of the difference between the multinomial $-\ln L$ and the $-\ln L$ using the selected model
6. compare this same difference from the empirical data to the distribution

Comparative Phylogeography of Mesoamerican Highland Rodents: Concerted versus Independent Response to Past Climatic Fluctuations

Jack Sullivan,^{1,*} Elizabeth Arellano,^{2,3} and Duke S. Rogers²

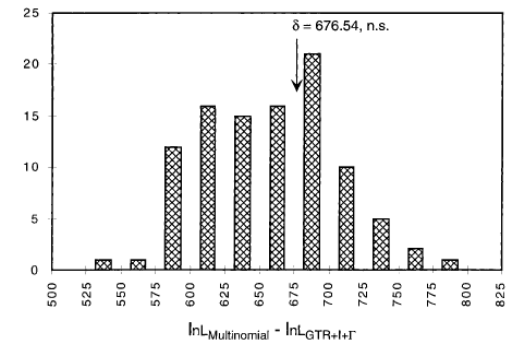


Figure 5: Parametric bootstrap test of the goodness-of-fit between the GTR+I+Γ model of sequence evolution and the *Reithrodontomys sumichrasti* cyt b data. One hundred replicate data sets were simulated under the GTR+I+Γ model, and the distribution of the difference between unconstrained (multinomial) lnL and lnL of the ML tree under the model is shown. The difference observed in the real data falls in the middle of the distribution generated under the null hypothesis of a perfect match between model and data; the GTR+I+Γ model fits the data quite well.

1. simulate sequence data using selected model on ML estimate of phylogeny (match empirical dimensions)
2. estimate phylogenies from simulated data
3. calculate $-\ln L$ of each tree using the selected model of sequence evolution
4. calculate multinomial $-\ln L$
5. construct a null distribution of the difference between the multinomial $-\ln L$ and the $-\ln L$ using the selected model
6. compare this same difference from the empirical data to the distribution

Comparative Phylogeography of Mesoamerican Highland
Rodents: Concerted versus Independent Response
to Past Climatic Fluctuations

Jack Sullivan,^{1,*} Elizabeth Arellano,^{2,3} and Duke S. Rogers²

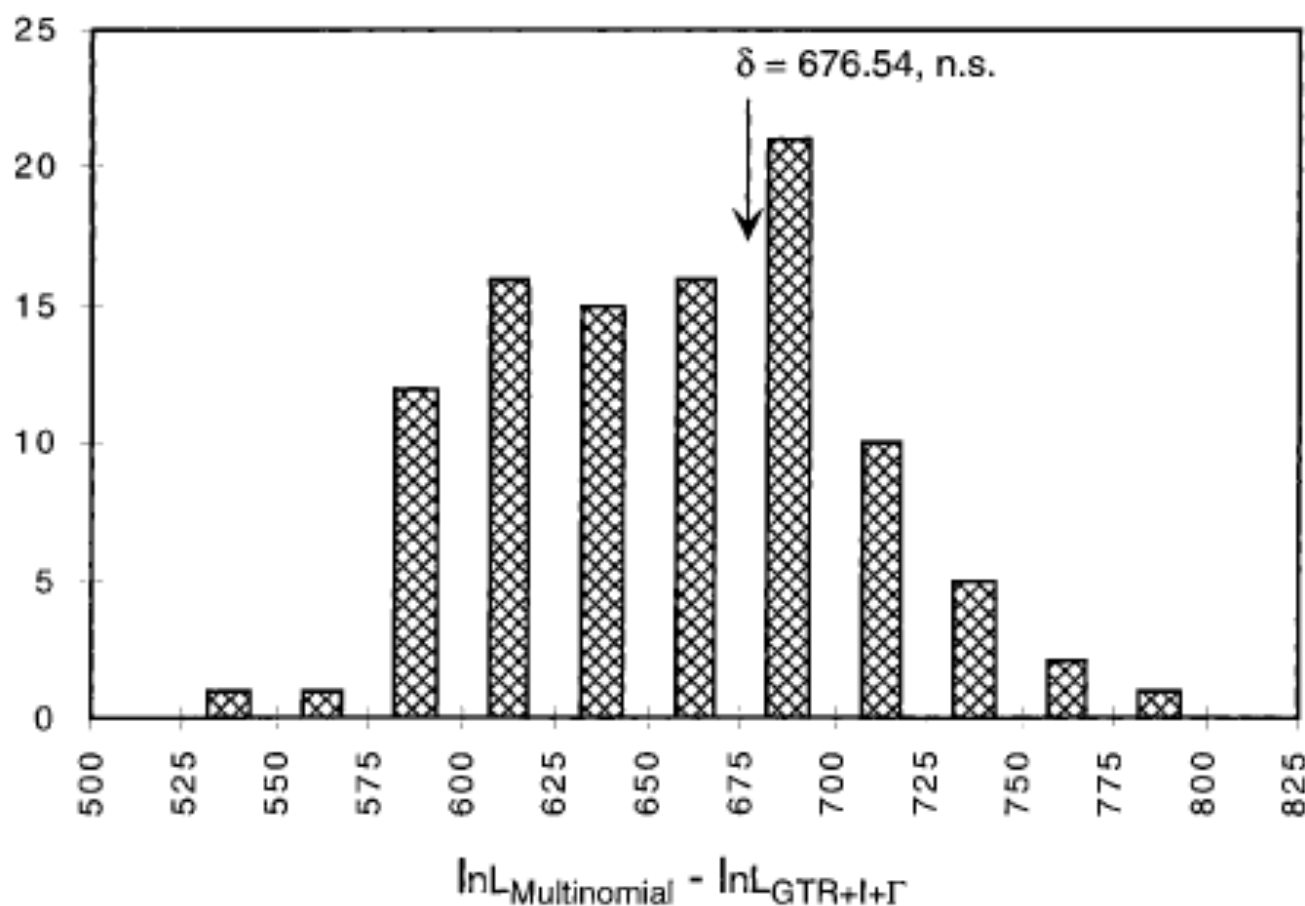


Figure 5: Parametric bootstrap test of the goodness-of-fit between the GTR+I+ Γ model of sequence evolution and the *Reithrodontomys sumichrasti* cyt b data. One hundred replicate data sets were simulated under the GTR+I+ Γ model, and the distribution of the difference between unconstrained (multinomial) ln likelihoods and ln likelihood of the ML tree under the model is shown. The difference observed in the real data falls in the middle of the distribution generated under the null hypothesis of a perfect match between model and data; the GTR+I+ Γ model fits the data quite well.

Parametric bootstraps

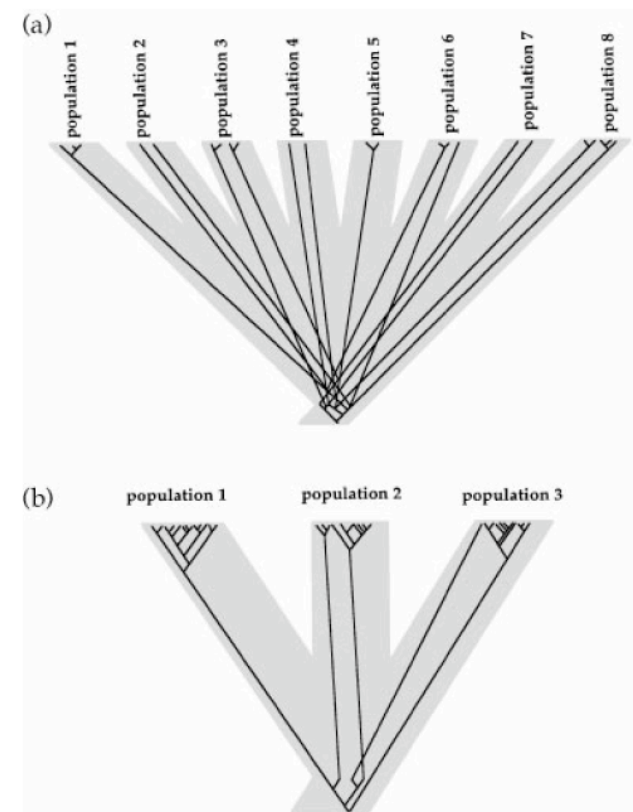
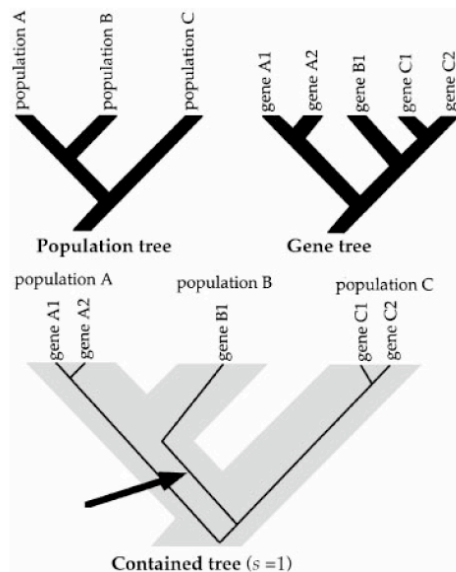
Parametric bootstrapping uses simulation to gather replicates of a parameter of interest in order to build a null distribution of the probability of the parameter given that the null hypothesis is true:

$$\text{Prob}(\Theta \mid H_0 \text{ is true})$$

By comparing the estimate of this parameter from the empirical data to the test distribution, you can then evaluate the null hypothesis.

Parametric bootstraps

“Coalescent” application: *Knowles 2001* - parametric simulations within models of population demographic history.



Molecular Ecology (2001) 10, 691–701

Did the Pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshoppers

L. LACEY KNOWLES*

Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook, New York 11794–5245, USA

Parametric bootstraps

1. simulate sequence data using selected model on ML estimate of phylogeny (match empirical dimensions)
2. estimate phylogenies from simulated data
3. calculate $-\ln L$ of each tree using the selected model of sequence evolution
4. calculate multinomial $-\ln L$
5. construct a null distribution of the difference between the multinomial $-\ln L$ and the $-\ln L$ using the selected model
6. compare this same difference from the empirical data to the distribution

Parametric bootstraps

1. simulate sequence data using selected model on ML estimate of phylogeny (match empirical dimensions)
2. estimate phylogenies from simulated data
3. calculate $-lnL$ of each tree using the selected model of sequence evolution
4. calculate multinomial $-lnL$
5. construct a null distribution of the difference between the multinomial $-lnL$ and the $-lnL$ using the selected model
6. compare this same difference from the empirical data to the distribution

Phylogenetic uncertainty in the topology?

Parametric bootstraps

- Phylogenetic uncertainty in the topology?**
1. simulate sequence data using selected model on ML estimate of phylogeny (match empirical dimensions)
 2. estimate phylogenies from simulated data
 3. calculate $-lnL$ of each tree using the selected model of sequence evolution
 4. calculate multinomial $-lnL$ **$\text{Prob}(\Theta | H_0 \text{ is true})$**
 5. construct a null distribution of the difference between the multinomial $-lnL$ and the $-lnL$ using the selected model
 6. compare this same difference from the empirical data to the distribution



Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology

John P. Huelsenbeck,^{1*} Fredrik Ronquist,² Rasmus Nielsen,³ Jonathan P. Bollback¹

As a discipline, phylogenetics is becoming transformed by a flood of molecular data.

ity of a tree (Fig. 1). Bayes's theorem

$\Pr[\text{Data} | \text{Tree}] \times \Pr[\text{Tree}]$

$$\Pr[\text{Tree} | \text{Data}] = \frac{\Pr[\text{Data} | \text{Tree}] \times \Pr[\text{Tree}]}{\Pr[\text{Data}]}$$



Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology

John P. Huelsenbeck,^{1*} Fredrik Ronquist,² Rasmus Nielsen,³ Jonathan P. Bollback¹

As a discipline, phylogenetics is becoming transformed by a flood of molecular data.

ity of a tree (Fig. 1). Bayes's theorem

$P(\text{Data} | \text{Tree}) \propto P(\text{Tree})$

Posterior predictive simulation

1. draw trees (w/ parameter estimates) at random from the posterior, simulate data on these trees
2. estimate phylogenies from simulated data
3. calculate $-\ln L$ of each tree using the selected model of sequence evolution
4. calculate multinomial $-\ln L$
5. construct a null distribution of the difference between the multinomial $-\ln L$ and the $-\ln L$ using the selected model
6. compare this same difference from the empirical data to the distribution

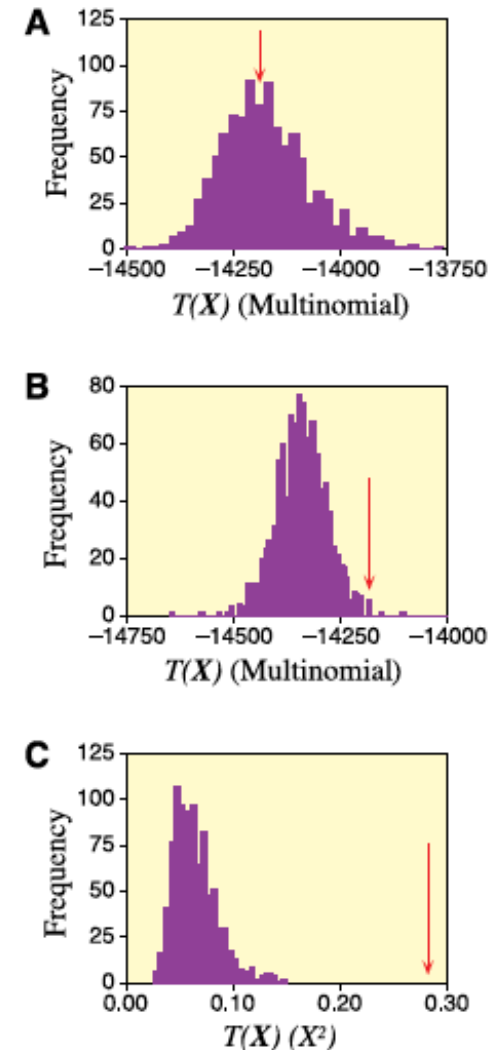


Fig. 3. The posterior predictive distributions for tests of (A) the adequacy of the GTR model, (B) of the adequacy of the Jukes-Cantor model, and (C) the hypothesis of constant nucleotide frequencies over time. The arrows above the distributions show the observed value of the test statistics.

P³: Phylogenetic Posterior Prediction in RevBayes

Sebastian Höhna,^{*,1,2} Lyndon M. Coghill,³ Genevieve G. Mount,³ Robert C. Thomson,⁴ and Jeremy M. Brown³

¹Division of Evolutionary Biology, Ludwig-Maximilians-Universität, München, Germany

²Department of Integrative Biology, University of California, Berkeley, CA

³Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA

⁴Department of Biology, University of Hawai'i, Honolulu, HI

***Corresponding author:** E-mail: sebastian.hoehna@gmail.com.

Associate editor: Keith Crandall

Lyndon Coghill - RevBayes

Syst. Biol. 63(3):322–333, 2014

© The Author(s) 2013. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.

For Permissions, please email: journals.permissions@oup.com

DOI:10.1093/sysbio/syt057

Advance Access publication August 28, 2013

Poor Fit to the Multispecies Coalescent is Widely Detectable in Empirical Data

NOAH M. REID^{1,*}, SARAH M. HIRD¹, JEREMY M. BROWN¹, TARA A. PELLETIER², JOHN D. McVAY¹, JORDAN D. SATLER²,
AND BRYAN C. CARSTENS²

¹*Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA; and*

²*Department of Evolution, Ecology & Organismal Biology, Ohio State University, Columbus, OH 43210, USA*

**Correspondence to be sent to: 202 Life Sciences Building, Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA; E-mail: nreid1@tigers.lsu.edu.*

Received 26 November 2012; reviews returned 2 March 2013; accepted 1 August 2013

Associate Editor: Laura Kubatko

Noah Reid - Detecting fit of the multispecies coalescent model

Syst. Biol. 67(2):269–284, 2018

© The Author(s) 2017. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.

For Permissions, please email: journals.permissions@oup.com

DOI:10.1093/sysbio/syx073

Advance Access publication September 4, 2017

Impact of Model Violations on the Inference of Species Boundaries Under the Multispecies Coalescent

ANTHONY J. BARLEY^{1,*}, JEREMY M. BROWN², AND ROBERT C. THOMSON¹

¹Department of Biology, University of Hawai'i, 2538 McCarthy Mall, Edmondson Hall 216, Honolulu, HI 96822, USA; ²Department of Biological Sciences and Museum of Natural Science, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA

*Correspondence to be sent to: Department of Biology, University of Hawaii, 2538 McCarthy Mall, Edmondson Hall 216, Honolulu, HI 96822, USA; E-mail: ajbarley@hawaii.edu.

Received 15 March 2017; reviews returned 21 June 2017; accepted 31 August 2017

Associate Editor: Laura Kubatko

Anthony Barley – detecting fit to species delimitations under the MSCM

Funding support for Posterior Predictive Simulation workshop:

DBI - 1661029

