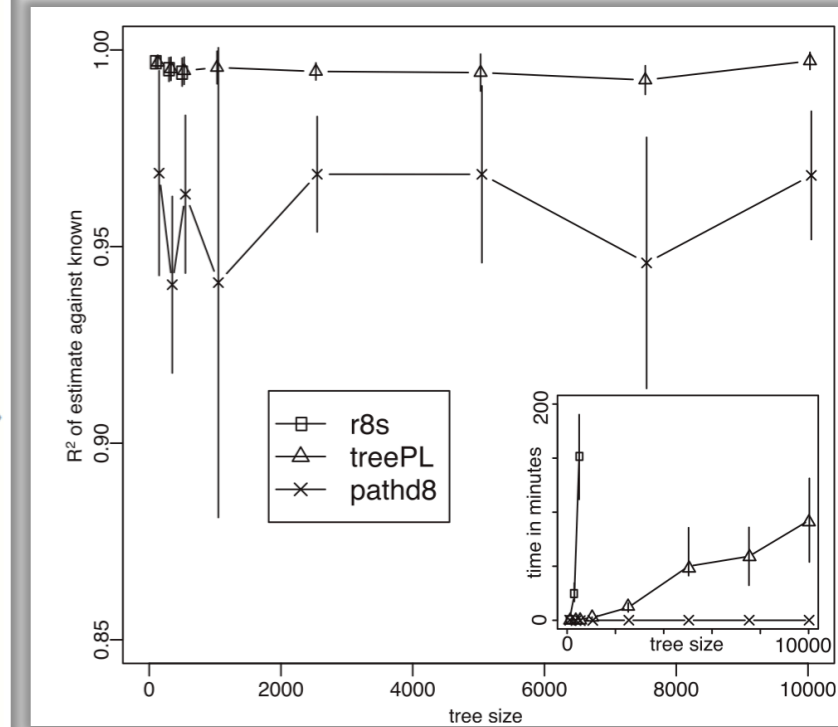
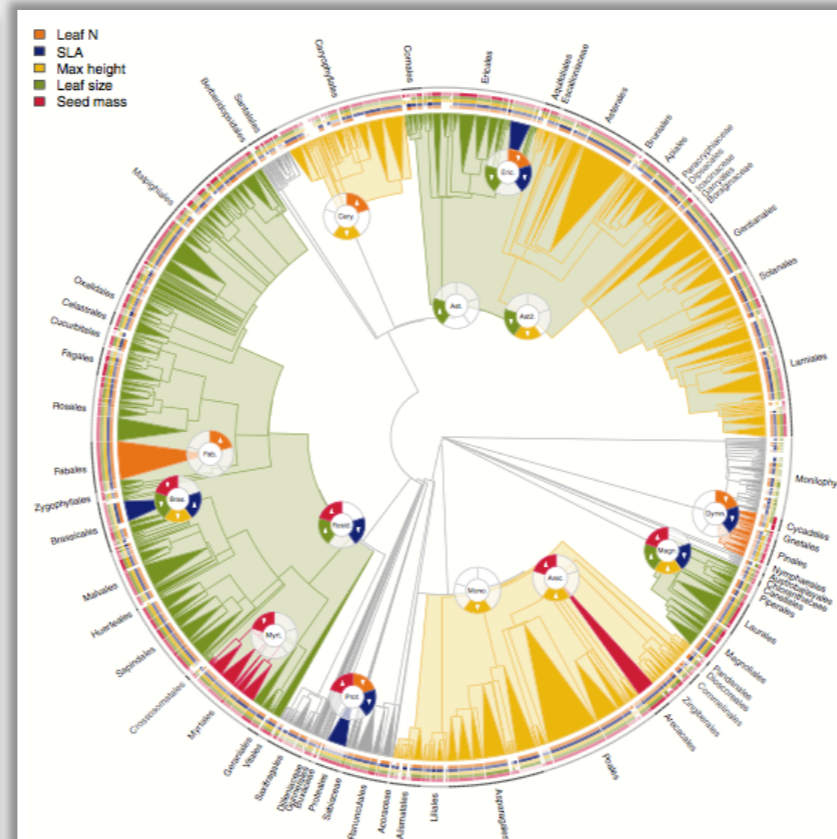
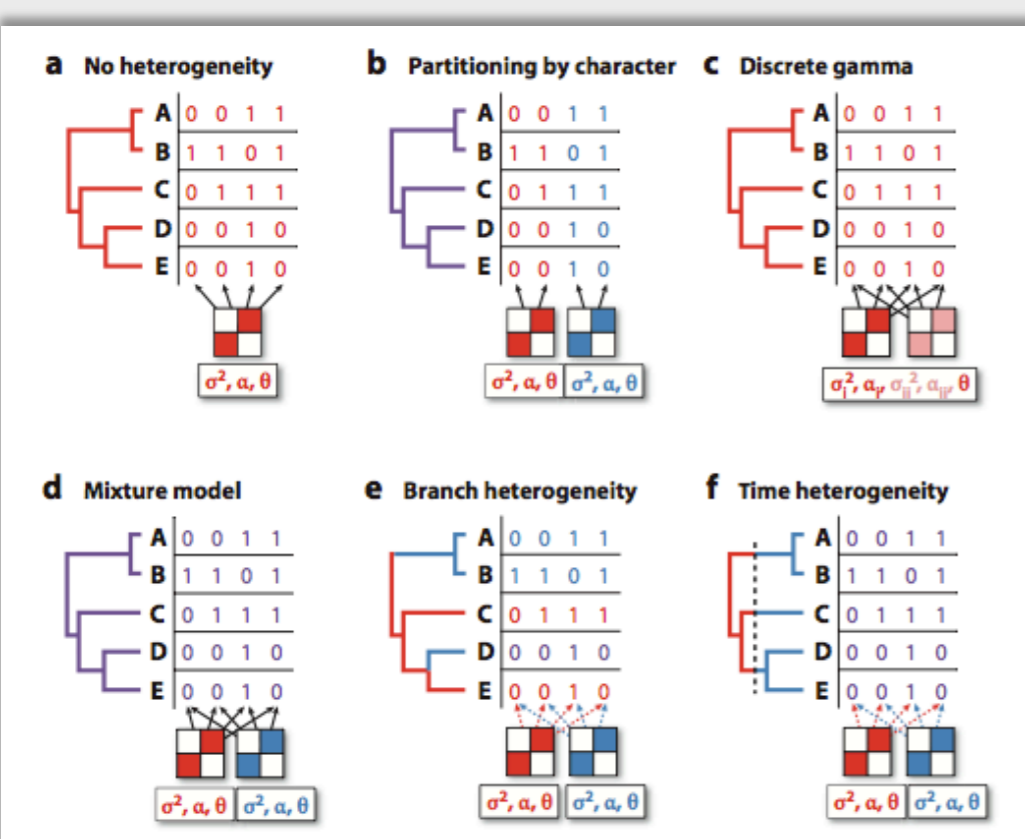


Developing and applying tools to answer outstanding biological questions using phylogenies, II



Brian O'Meara
brianomeara.info
UT Knoxville



Teaching

Service

Research

Teaching

	2014-15	2013-14	2012-13	2011-12	2010-11	2009-10
Speciation	2	1				1
Macroevolution	1	1	1	1	1	1
Core: Evolution*	1	1	1	1	1	1
HOFF*	2	2	2	2		
Bio130: Biodiversity		1		1		
Seminar					1	

- High performance computing for phylogenetics, NIMBioS tutorial. Organizer
- Species delimitation course, Gothenburg, Sweden. Instructor
- eFlower summer school, Vienna, Austria. Instructor
- Evolutionary quantitative genetics, NESCent. Instructor
- Markov processes, Lausanne, Switzerland. Instructor
- Computing in the cloud, NIMBioS tutorial. Co-organizer
- Evolutionary quantitative genetics, NIMBioS tutorial. Instructor

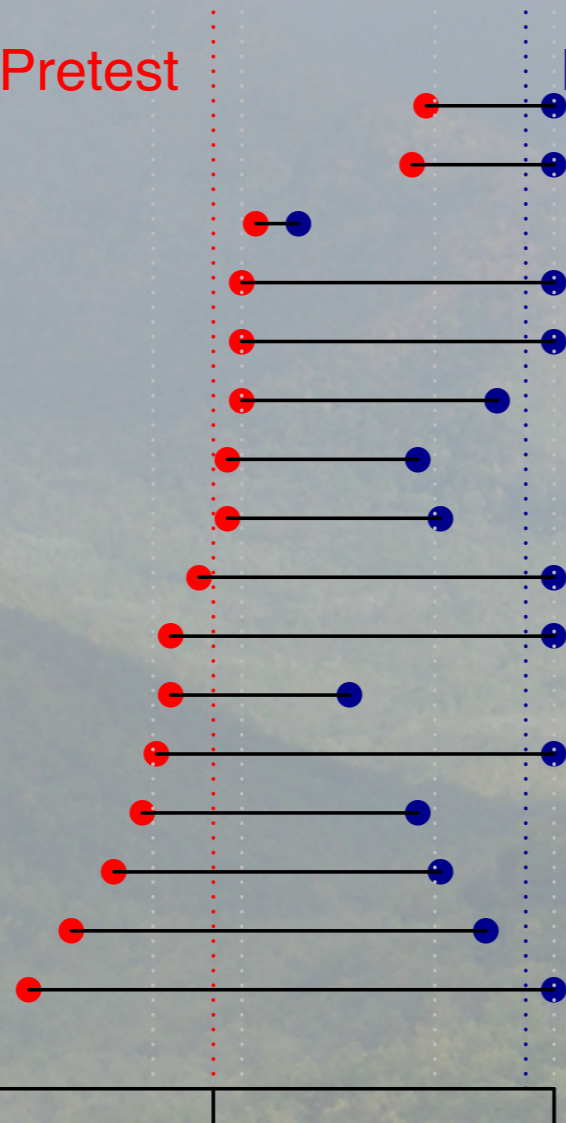
* Team taught courses

Teaching

Assessment

Pretest

Final



Wrong

Poor

Ok

Good

Innovation

Lampyr

Welcome to Lampyr. For more information, see our [webpage](#). There are 128,604 georeferenced species and 8,703,452 points in the database, so it may take a minute to search once you hit the button below.

Find closest species

Options (scroll down)
Use only species with interpret, but fewer s

Lampyr

The 25 closest species:

Ties are broken randomly. Note that these points are where species were at some time in the past. Species migrates, are out down, paved over, etc. Also, sometimes localities are approximate: people might take one GPS measurement at a trail head and use that for all the species they collect on the trail. In short, if a given species is reported as being 0.0 miles away, it might be right there, but it might not. Also note that fewer than 1% of all species are in this database, so the actual diversity surrounding you is far greater.

- Seiurus motacilla*
Louisiana waterthrush
0.2 mi ENE
- Schistocerca damnifica*
Carolina locust
0.4 mi NE
- Nycticorax nycticorax*
black-crowned night-heron
0.5 mi WSW
- Cyprinella galactura*
whitetail shiner
0.8 mi ENE
- Dendroica pensylvanica*
chestnut-sided warbler
2.3 mi WSW

Feedback to Brian

This is a general feedback form allowing you to provide feedback on other long-term interaction. It is another second question is required.

Required

Interaction

What are you complaining/complimenting?

- Macroevolution course
- EEB graduate core
- Research in the lab
- Bio130
- Phylogenetics discussion group
- Other: _____

Brian O'Meara

Home Videos **Playlists** Channels About

Evidence

2013 Macroevolution EEB464

by Brian O'Meara • 32 videos • 138 views • 24 hours

▶ Play all ◀ Share + Save

- Macroevolution: Evidence, 28 Aug 2013** Brian O'Meara
- Macroevolution: Taphonomy, 30 Aug 2013** Brian O'Meara
- Macroevolution: Jargon, 4 Sept 2013** Brian O'Meara
- Macroevolution: Phylogenetics, 6 Sept 2013** Brian O'Meara
- Macroevolution: Empirical Distributions, 9 Sept 2013** Brian O'Meara
- Macroevolution: Biogeography and Phylogeography, 11 Sept 2013** Brian O'Meara

Teaching

- Bio130: Greatest contribution to learning: “The open dialogue we had in every lecture”
- Macroevolution: “I found this class wonderfully stimulating and I really appreciated that Prof. O’Meara would stop in the middle of his lecture to answer any and all questions without it bothering him”
- Core: “The subject is a difficult one, but O’Meara communicated complex topics in a very understandable way. In short, he is a great teacher with many great qualities (organized, attentive, intelligent but relatable)”

People

Students

Sam Borstein

Jenn Bosco

Katie Massana

Orlando Schwery

People

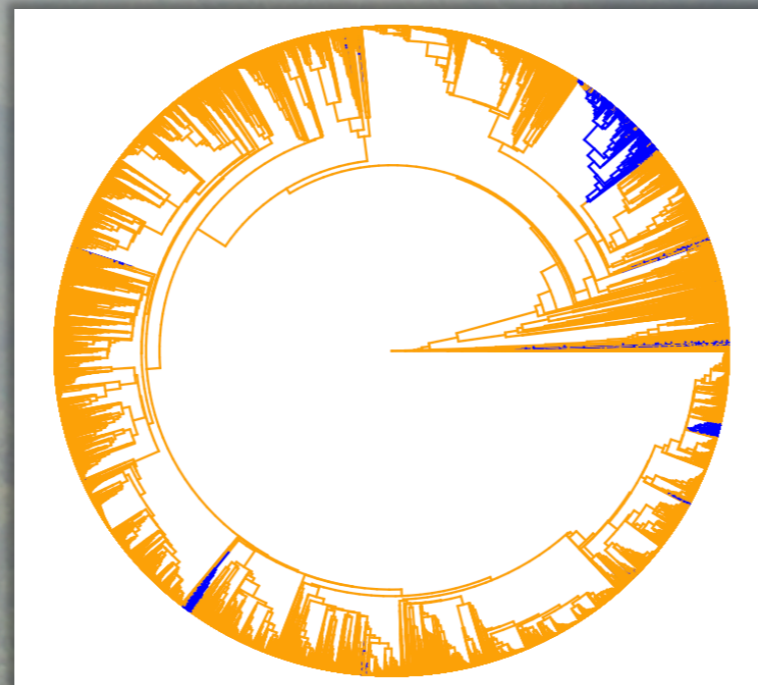
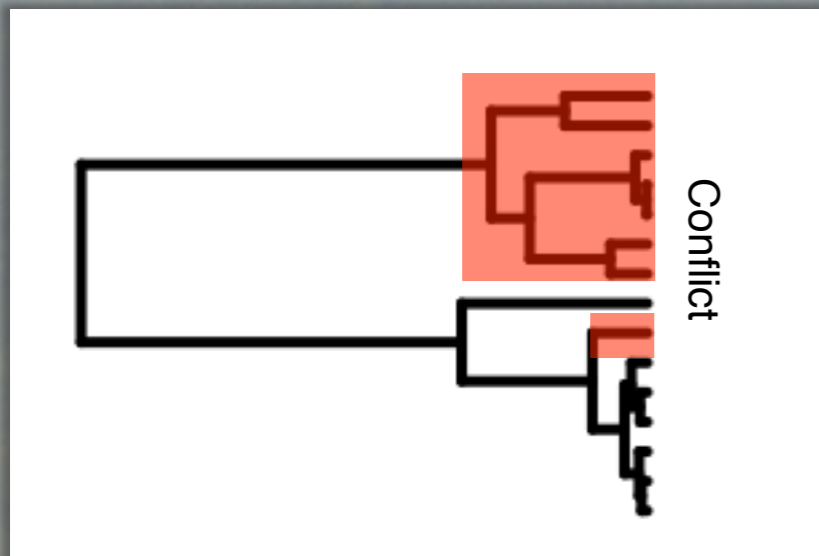
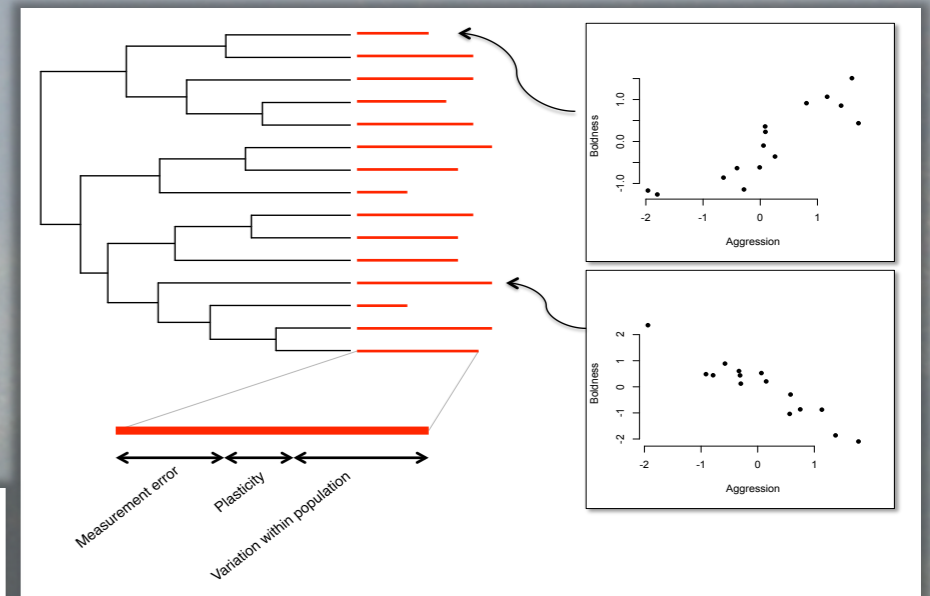
Students

Sam Borstein

Jenn Bosco

Katie Massana

Orlando Schwery



People

Students

Sam Borstein

Jenn Bosco

Katie Massana

Orlando Schwery

Postdocs

Hugo Alamillo

Barb Banbury

Jeremy Beaulieu*

JJ Chai*

Nathan Jackson

Tony Jhwueng*

Sandy Kawano*

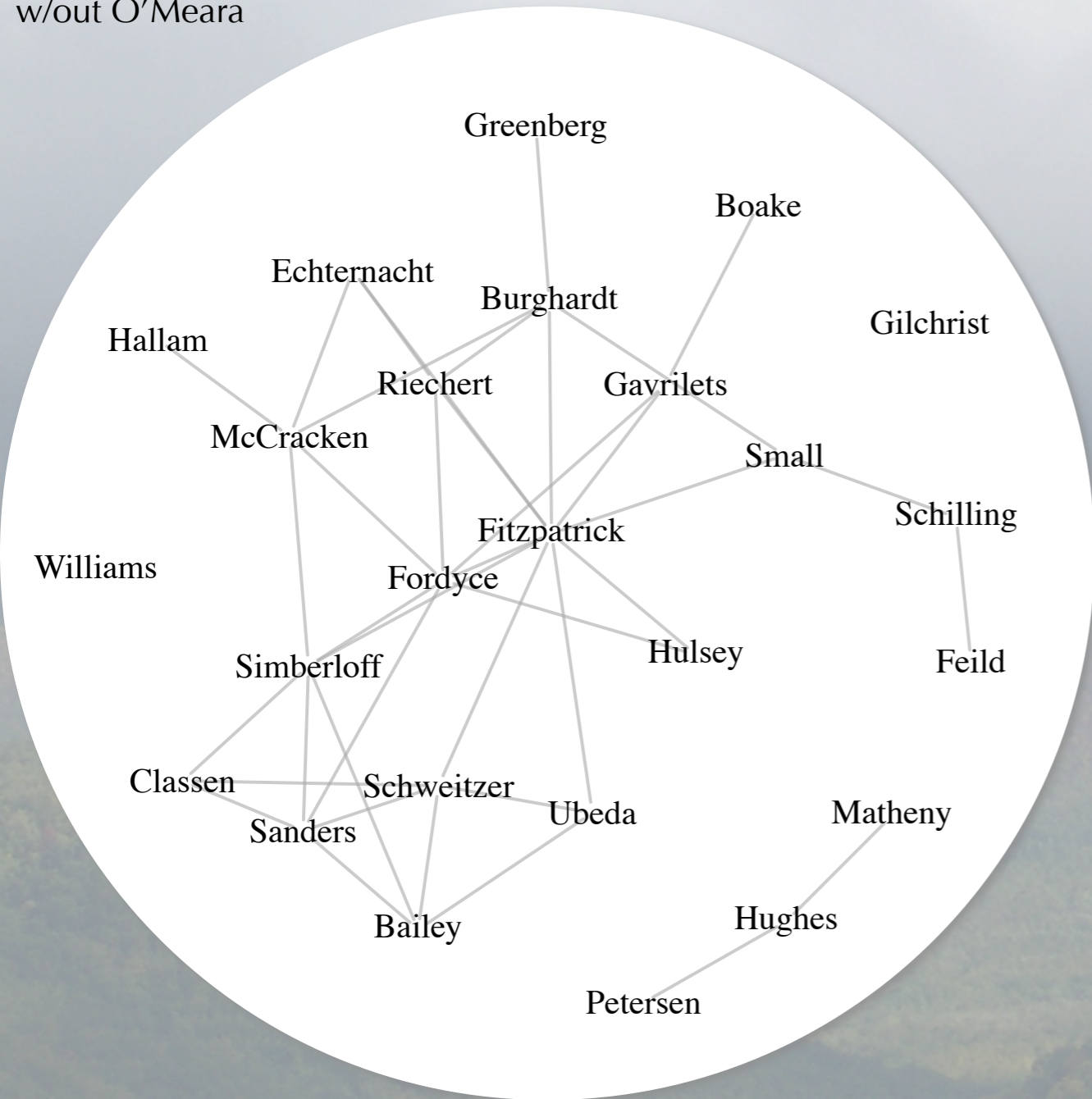
Michelle Lawing*

Ryan Martin*

Nick Matzke*

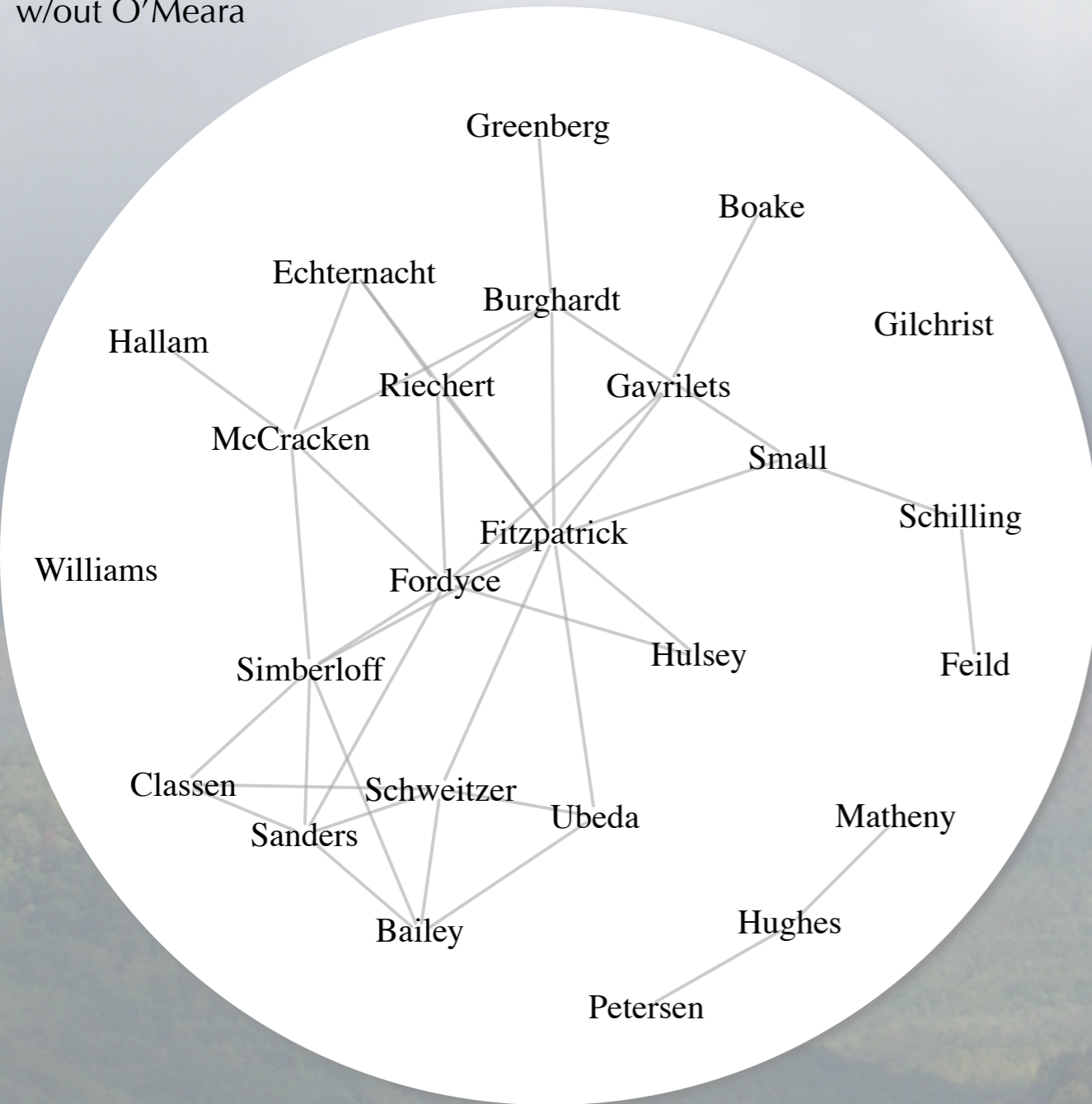
* = NIMBioS postdocs I co-mentored, about 19% of all NIMBioS postdocs

EEB Network,
w/out O'Meara

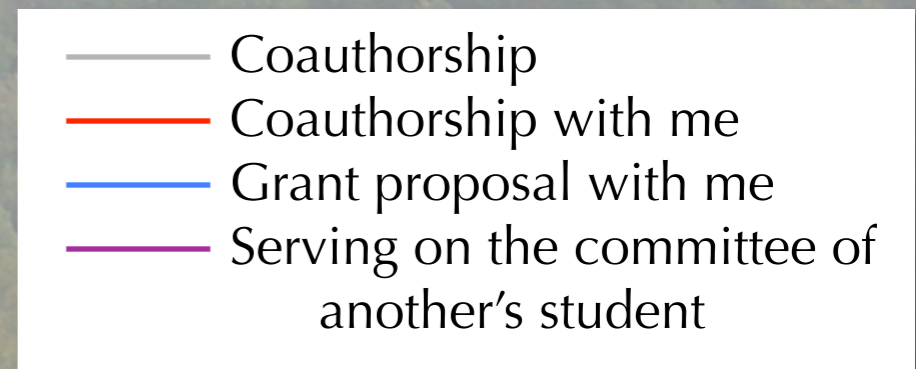
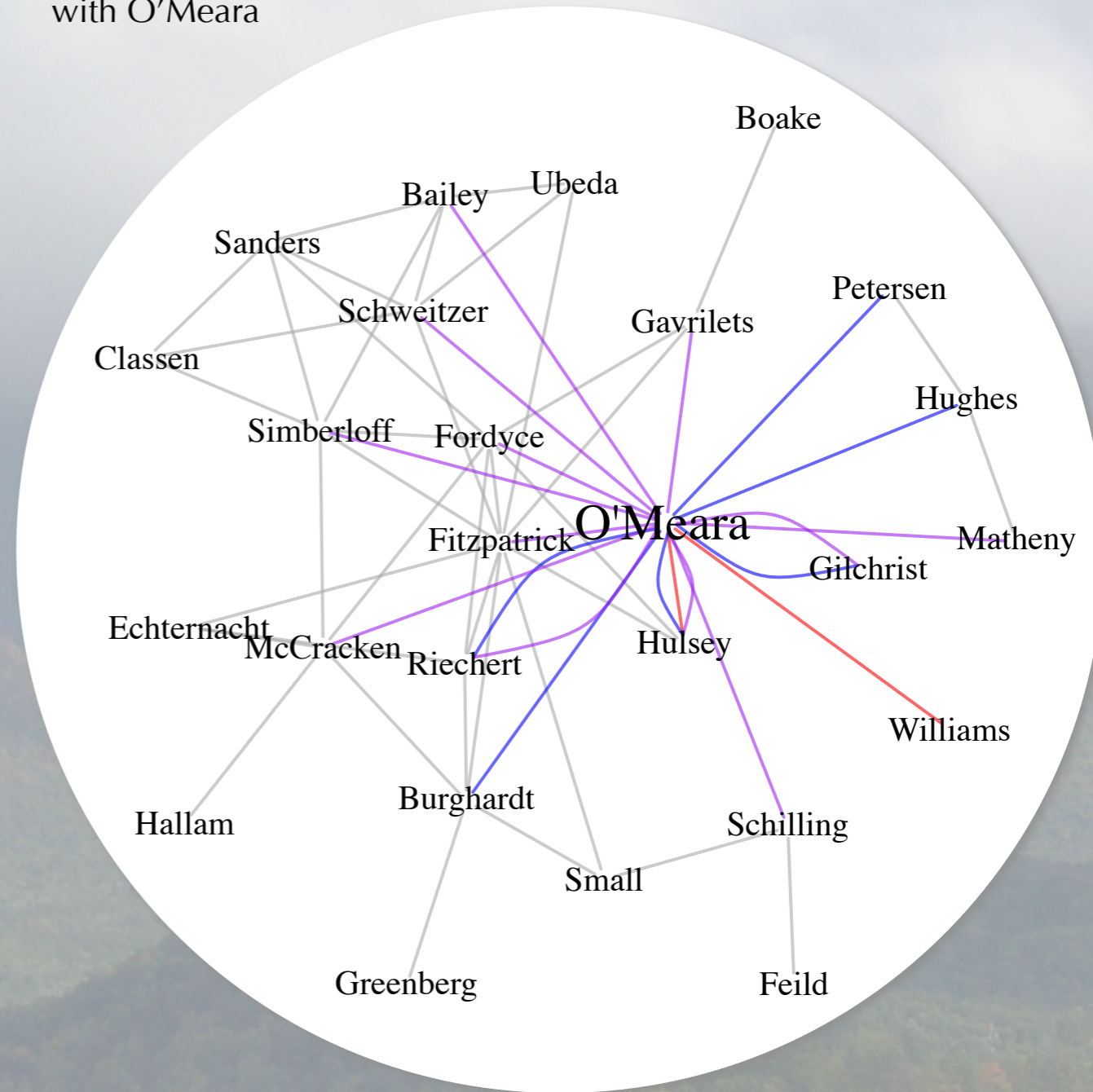


Networks includes all EEB faculty from 2008 – present who are connected to at least one other member on a network

EEB Network,
w/out O'Meara



EEB Network,
with O'Meara



Networks includes all EEB faculty from 2008 – present who are connected to at least one other member on a network



Teaching

Service

Research

Service & Outreach

- Co-Organizer, Evolution 2014 (2000 participants)
- Lightning talk organizer, Evolution 2013
- Phylotastic Leadership Team (hackathons)
- iEvoBio Leadership (satellite conference)
- Elected member of Society of Systematic Biologists council



Service & Outreach

- Co-Organizer, Evolution 2014 (2000 participants)
- Lightning talk organizer, Evolution 2013
- Phylotastic Leadership Team (hackathons)
- iEvoBio Leadership (satellite conference)
- Elected member of Society of Systematic Biologists council
- Darwin Day Tennessee advisor
- Applications Editor, *Methods in Ecology and Evolution*
- Reviewer for *Science*, NSF, *Systematic Biology*, and numerous other journals
- Maintainer of CRAN Task View for Phylogenetics in R
- EEB Dept. web committee, head search committee, undergrad affairs, grad admissions, Dean's Advisory Council, many grad student committees in four departments

Service & Outreach

- Co-organized Women in Science seminar series in first year at UTK, got funds from Haines-Morris and buy-in from multiple departments

Service & Outreach

- Co-organized Women in Science seminar series in first year at UTK, got funds from Haines-Morris and buy-in from multiple departments
- Highlighted twice in *Science Careers* for work helping grad students prepare for NSF proposals during federal shutdown

The screenshot shows the Science Careers website interface. At the top, there is a navigation bar with the Science logo and links for AAAS.ORG, FEEDBACK, HELP, LIBRARIANS, and a search bar. Below this is a secondary navigation bar with links for NEWS, SCIENCE JOURNALS, CAREERS, MULTIMEDIA, and COLLECTIONS. The main content area features a search bar and a list of articles. The featured article is titled "The Latest Shutdown Information for NIH- and NSF-Funded Researchers (UPDATE)" by Jim Austin, dated September 30, 2013. The article text discusses the impact of a potential government shutdown on scientists funded by the National Institutes of Health (NIH) and the National Science Foundation (NSF). It mentions that on Friday, 27 September, the author described the likely impact on scientists and that the government indeed came to pass the following Monday, at midnight. The article includes a date stamp of 10/06/2013 and a list of updates, such as more information from NSF's Web site accessed via Twitter from Brian O'Meara. The article also includes a link to the Internet Archive and a quote from Brian O'Meara dated October 5, 2013. The article concludes with a note that the author may need to be patient because the Wayback Machine is slow, but that it should get the user most of what they need from NSF.

Service & Outreach

- Co-organized Women in Science seminar series in first year at UTK, got funds from Haines-Morris and buy-in from multiple departments
- Highlighted twice in *Science Careers* for work helping grad students prepare for NSF proposals during federal shutdown

Fortunately, there are workarounds. Start at [the Web site of Brian O'Meara](#), an assistant professor at the University of Tennessee, Knoxville, in the Department of Ecology and Evolutionary Biology. O'Meara has located and stashed away some NSF forms likely to be in frequent demand: NSF's [Grant Proposal Guide](#); the program solicitation for Directorate for Biological Sciences [Doctoral Dissertation Improvement Grants](#) (Due date: 10 October); and the program solicitation for the crosscutting [Graduate Research Fellowship Program](#), which has due dates in early November.

Government Shutdown

The Latest Shutdown Information for NIH- and NSF-Funded Researchers (UPDATE)

By Jim Austin
September 30, 2013

Do you have new shutdown-related information that's relevant to NSF- or NIH-funded investigators or those seeking funding from the major agencies? [Send us an email](#).

On Friday, 27 September, [we described the likely impact](#) on scientists funded by the National Institutes of Health (NIH) and the National Science Foundation (NSF) if the government entered a period of partial shutdown. That indeed came to pass the following Monday, at midnight. Here's what we've learned since then:

10/06/2013

- More on accessing information from NSF's Web site--this via Twitter from Brian O'Meara, who, as we mentioned on Wednesday, had archived some critical files on his Web site. O'Meara tweets,

Internet Archive to the rescue: <http://t.co/MSvmF5A6gl> @SciCareerEditor #shutdown @NSF

— Brian O'Meara (@omearabrian) [October 5, 2013](#)

You may need to be patient because it's slow, but that should get you to most of what you need from NSF. (I've been using the Wayback Machine forever, so I'm surprised I didn't think of that earlier. Thanks Brian!)

10/04/2013

Service & Outreach

- Co-organized Women in Science seminar series in first year at UTK, got funds from Haines-Morris and buy-in from multiple departments
- Highlighted twice in *Science Careers* for work helping grad students prepare for NSF proposals during federal shutdown

Fortunately, there are workarounds. Start at [the Web site of Brian O'Meara](#), an assistant professor at the University of Tennessee, Knoxville, in the Department of Ecology and Evolutionary Biology. O'Meara has located and stashed away some NSF forms likely to be in frequent demand: NSF's [Grant Proposal Guide](#); the

progr
date: 10/06/2013

which

- More on accessing information from NSF's Web site--this via Twitter from Brian O'Meara, who, as we mentioned on Wednesday, had archived some critical files on his Web site. O'Meara tweets,

Internet Archive to the rescue: <http://t.co/MSvmF5A6gi> .@SciCareerEditor #shutdown .@NSF

— Brian O'Meara (@omearabrian) [October 5, 2013](#)

You may need to be patient because it's slow, but that should get you to most of what you need from NSF. (I've been using the Wayback Machine forever, so I'm surprised I didn't think of that earlier. Thanks Brian!)

Service & Outreach

- Co-organized Women in Science seminar series in first year at UTK, got funds from Haines-Morris and buy-in from multiple departments
- Highlighted twice in *Science Careers* for work helping grad students prepare for NSF proposals during federal shutdown
- A representative from SSB on joint *American Society of Naturalists*, *Society for the Study of Evolution*, and *Society of Systematic Biologists* group to help gather data for proposal to improve representation of women in those societies
- Co-organizer of symposium at Evolution 2015 on “Women in Science: Pairing Advances in the Theory and Application of Phylogenetic Methods”
- Outreach via social media and web (i.e., Google “Akaike weight”)



Teaching

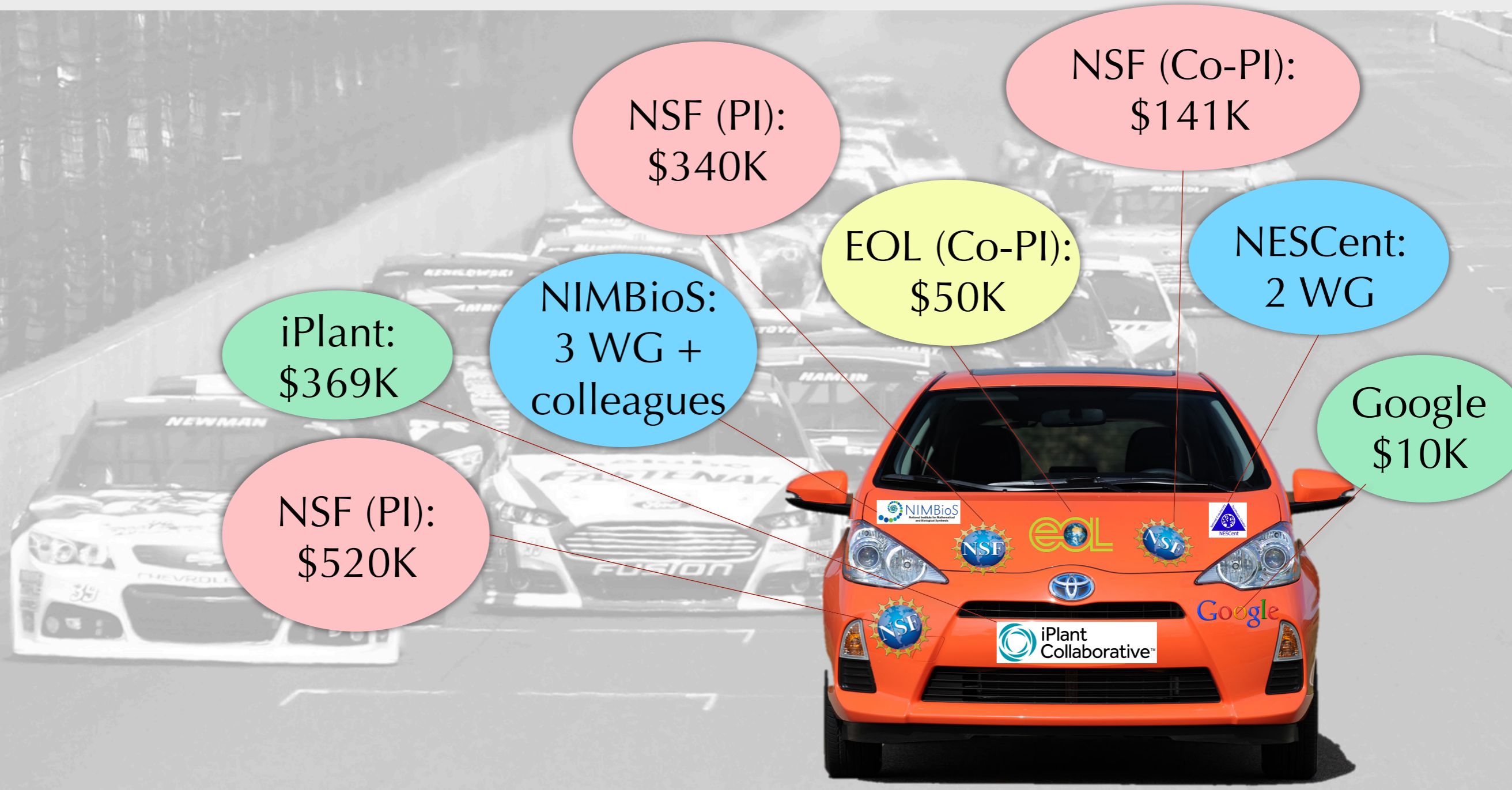
Service

Research

Research: Acknowledgments



Research: Acknowledgments

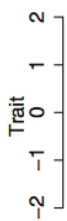


Continuous trait evolution

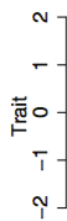
384

B. C. O'Meara and J. M. Beaulieu

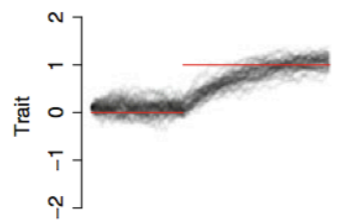
(a) Brownian motion



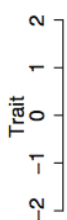
(b) OU, one sigma, one theta, one alpha



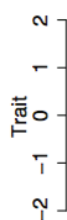
(c) OU, one sigma, two theta, one alpha



(d) OU, two sigma, two theta, one alpha



(e) OU, one sigma, one theta, two alpha



(f) OU, one sigma, two theta, two alpha

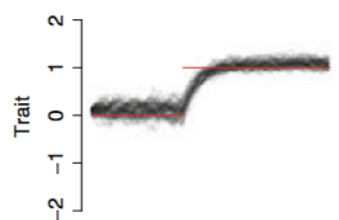


Table 2. The fit of alternative models of genome size evolution in monocots. The best model, based on ΔAIC and Akaike weights, was the OU_{MVA} , which estimated a separate θ , α , and σ^2 for woody and herbaceous monocot lineages.

Model	$-\ln L$	AIC	ΔAIC	w_i
BM1	-227.6	459.1	156.2	<0.01
BMS	-203.0	412.0	109.1	<0.01
OU1	-160.0	326.1	23.2	<0.01
OU_M	-159.2	326.5	23.6	<0.01
OU_{MV}	-147.3	304.6	1.7	0.290
OU_{MA}	-159.2	328.5	25.6	<0.01
OU_{MVA}	-145.3	302.9	0.0	0.678

Cited 249 times

Cited 63 times

Evolution, 60(5), 2006, pp. 922-933

TESTING FOR DIFFERENT RATES OF CONTINUOUS TRAIT EVOLUTION USING LIKELIHOOD

BRIAN C. O'MEARA,¹ CÉCILE ANÉ,² MICHAEL J. SANDERSON,^{3,4} AND PETER C. WAINWRIGHT^{5,6}

¹Center for Population Biology, University of California, Davis, One Shields Avenue, Davis, California 95616
²Department of Statistics, University of Wisconsin-Madison, Medical Science Center, 1300 University Avenue, Madison, Wisconsin 53706-1532
³Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1610
⁴Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1610
⁵Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1610
⁶Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1610

Abstract.—Rates of phenotypic evolution have changed throughout the history of life, producing variation in levels of morphological, functional, and ecological diversity among groups. Testing for the presence of these rate shifts is a key component of evaluating hypotheses about what causes them. In this paper, general predictions regarding changes in phenotypic diversity as a function of evolutionary history and rates are developed, and tests are derived to evaluate rate changes. Simulations show that these tests are more powerful than existing tests using standardized contrasts. The new approaches are distributed in an application called *Brownie* and in *R*.

Key words.—Brownian motion, *Brownie*, comparative method, continuous characters, disparity, morphological evolution, rate.

Received March 8, 2005. Accepted March 4, 2006.

All five extant flamingo species are long-legged filter feeders, whereas their sister group, consisting of twenty species of grebes (Van Tuinen et al. 2001; Chubb 2004; Mayr 2004), feed on prey ranging from fish and squid to minute invertebrates (Fjeldså 1983), and show a variety of body and bill shapes. Methods to test whether the difference in species number between flamingos and grebes arose by chance or reflects differences in diversification rates have been developed (Słowiński and Guyer 1989; Nee et al. 1992; Hey 1992; Harvey et al. 1994). These methods are aimed at discovering factors affecting diversification. But there are also undoubtedly factors that led to the difference in variability of ecologically important traits within these two groups. This paper is concerned with hypotheses about factors that lead to differences between groups in phenotypic and biological diversity, as opposed to species richness.

There are many potential hypotheses regarding factors that can affect the rate of evolution of phenotypic characters (which include morphological, behavioral, physiological, biochemical, and ecological traits). For example, once wings replaced legs as the primary means of locomotion in birds, newly less constrained legs may have begun to evolve new shapes more rapidly (Gates and Middleton 1997). The evolution of asexual reproduction may reduce the rate of genome size evolution. The invasion of a new, competitor-free island may increase the rate of evolution of feeding structures. These hypotheses all attempt to relate a change in some aspect of the biology of the lineage with a change of the rate of evolution of a continuous character based on an idea about how evolution works. Hypotheses can also be generated from observations of patterns of diversity instead of predictions based on a mechanism. Grebes appear to have more interspecific variation in bill dimensions than flamingos; this may reflect a faster rate of bill evolution, or perhaps the grebe species have been evolving independently for more time than the flamingo species.

Habitat use affects morphological diversification in dragon lizards

D. C. COLLAR,¹ J. A. SCHULTE II,¹ B. C. O'MEARA,² & J. B. LOSOS³

¹Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA
²Department of Biology, Clarkson University, Potsdam, NY, USA
³Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA

Keywords:

Agamidae;
 Brownian motion;
 ecomorphology;
 iguanids;
 locomotion;
 phylogenetic comparative method.

Abstract

Habitat use may lead to variation in diversity among evolutionary lineages because habitats differ in the variety of ways they allow for species to make a living. Here, we show that structural habitats contribute to differential diversification of limb and body form in dragon lizards (Agamidae). Based on phylogenetic analysis and ancestral state reconstructions for 90 species, we find that multiple lineages have independently adopted each of four habitat use types: rock-dwelling, terrestriality, semi-arborescence, and arborescence. Given these reconstructions, we fit models of evolution to species' morphological trait values and find that rock-dwelling and arborescence limit diversification relative to terrestriality and semi-arborescence. Models preferred by Akaike information criterion infer slower rates of size and shape evolution in lineages inferred to occupy rocks and trees, and model-averaged rate estimates are slowest for these habitat types. These results suggest that ground-dwelling facilitates ecomorphological differentiation and that use of trees or rocks impedes diversification.

Introduction

One of the great questions in evolutionary biology concerns the causes of differences in diversity among clades. Ecological factors are often implicated to explain this pattern because the ecological circumstances available to the members of a lineage contribute to the mode of natural selection they experience and thus shape ecological divergence, morphological adaptation and the evolution of new species. Although much work has focused on the role of biotic interactions within communities (e.g. competition-driven divergent selection in *Anolis* lizards [Williams, 1972; Losos, 2009], Hawaiian silverswords [Carlquist et al., 2003] and Darwin's finches [Schluter, 1988; Grant & Grant, 2008]), other aspects of a lineage's ecology may also be important for diversification.

In this study, we test the hypothesis that diversity varies as a function of habitat. For many reasons, some habitats may foster greater diversity than others. Some habitat types may be readily subdivided, perhaps because of spatial complexity (e.g. coral reefs [Alfaro et al., 2007]) or geographical area (e.g. garrid habitats in Australia [Rabosky et al., 2007]) and may thereby present evolutionary lineages with many alternative means for microhabitat specialization or local adaptation. Other habitats may impose stringent functional constraints that lead to strong selection resisting ecological and phenotypic divergence away from an adaptive peak (Butler & King 2004; Collar et al., 2009). Habitat types may also contribute differently to diversification because they vary in the number and type of species interactions they present, such as the presence or absence of predators (McPeck & Brown, 2000). In addition, some habitats may provide opportunities if they are variable across space in the strength of species interactions (McPeck, 1996) or in their functional demands.

The consequences of habitat use for diversification have been investigated primarily in the context of

Correspondence: David C. Collar, Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. Tel: +1 617 495 8099; fax: +1 617 495 3667; e-mail: dcollar@oeb.harvard.edu
 Re-use of this article is permitted in accordance with the Terms and Conditions set out at <http://www3.interscience.wiley.com/authors/905525110000.html>

© 2010 THE AUTHORS. J. EVOL. BIOL. 23 (2010) 1033–1049
 JOURNAL COMPILATION © 2010 EUROPEAN SOCIETY FOR EVOLUTIONARY BIOLOGY

1033

ORIGINAL ARTICLE



MODELING STABILIZING SELECTION: EXPANDING THE ORNSTEIN-UHLENBECK MODEL OF ADAPTIVE EVOLUTION

JEREMY M. BEAULIEU,^{1,2} DWUENG-CHWUAN JHWUENG,^{3,4} CARL BOETTIGER,⁵ AND BRIAN C. O'MEARA⁶

¹Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, Connecticut 06520-8106

²E-mail: jeremy.beaulieu@yale.edu

³National Institute for Mathematical and Biological Synthesis, 1534 White Ave, University of Tennessee, Knoxville, Tennessee, 37996-1527

⁴Department of Statistics, Feng-Chia University, Taichung, Taiwan 40724, R.O.C.

⁵Center for Population Biology, University of California, Davis, 1 Shields Avenue, Davis, California, 95616

⁶Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee, 37996-1610

Received September 2, 2011

Accepted February 6, 2012

Comparative methods used to study patterns of evolutionary change in a continuous trait on a phylogeny range from Brownian motion processes to models where the trait is assumed to evolve according to an Ornstein-Uhlenbeck (OU) process. Although these models have proved useful in a variety of contexts, they still do not cover all the scenarios biologists want to examine. For models based on the OU process, model complexity is restricted in current implementations by assuming that the rate of stochastic motion and the strength of selection do not vary among selective regimes. Here, we expand the OU model of adaptive evolution to include models that variously relax the assumption of a constant rate and strength of selection. In its most general form, the methods described here can assign each selective regime a separate trait optimum, a rate of stochastic motion parameter, and a parameter for the strength of selection. We use simulations to show that our models can detect meaningful differences in the evolutionary process, especially with larger sample sizes. We also illustrate our method using an empirical example of genome size evolution within a large flowering plant clade.

KEY WORDS: Brownian motion, comparative method, continuous characters, Hansen model, Ornstein-Uhlenbeck.

Single-rate Brownian motion works reasonably well as a model for evolution of traits. It models drift, drift-mutation balance, and even stabilizing selection toward a moving optimum (Hansen and Martins 1996). However, a single parameter model can certainly not explain the evolution of traits across all life. There have been extensions to the model, such as a single Ornstein-Uhlenbeck (OU) process that has a constant pull toward an optimum value, a multiple mean OU process with different possible means for different groups (Hansen 1997; Butler and King 2004), and multiple rate Brownian motion processes allowing different rates of evolution on different branches (O'Meara et al. 2006; Thomas et al. 2006). These models, while useful, still do not cover all the scenarios biologists want to examine. For example, existing models with a value toward which species are being pulled all have a fixed

strength of pull over the entire history of the group. It is possible to allow the rate of stochastic motion to vary, or the value of the attractor to vary, but not for both to vary. Such restrictions on model complexity may make sense when phylogenies are limited to a few dozen taxa. However, in an era where phylogenies can have over 55,000 taxa (Smith et al. 2011), we may be so bold as to attempt to fit models that vary both rates and means of the evolutionary process. This article develops and implements such models.

Generalizing the Hansen Model

Hansen (1997) described a model where quantitative characters are assumed to evolve according to an OU process. The Hansen model, as it has become known, expresses the amount of change

Chapter 15 Modelling Stabilizing Selection: The Attraction of Ornstein-Uhlenbeck Models

Brian C. O'Meara and Jeremy M. Beaulieu

Abstract. Ornstein-Uhlenbeck models are a generalization of Brownian motion models that allow trait values to evolve to follow optima. They have become broadly popular in evolutionary studies due to their ability to better fit empirical data as well as for the biological conclusions which can be drawn based on their parameter estimates, especially optimum trait values. We include a survey of available software implementing these models in phylogenetics as well as cautions regarding the use of this software.

15.1 Introduction

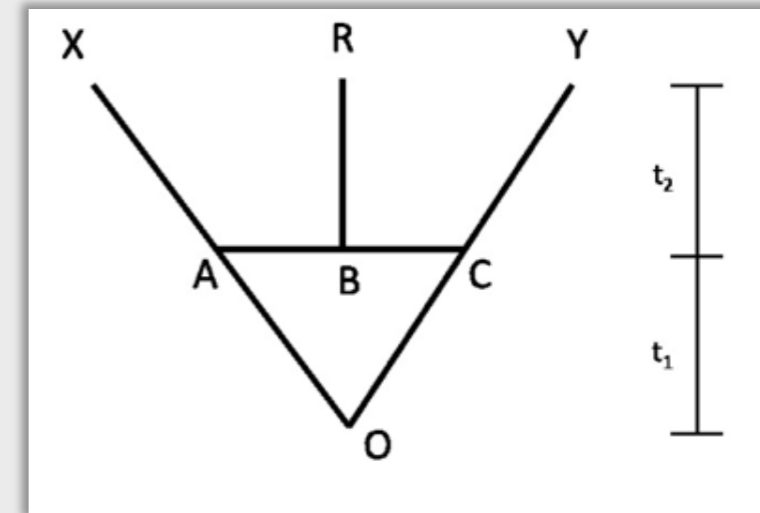
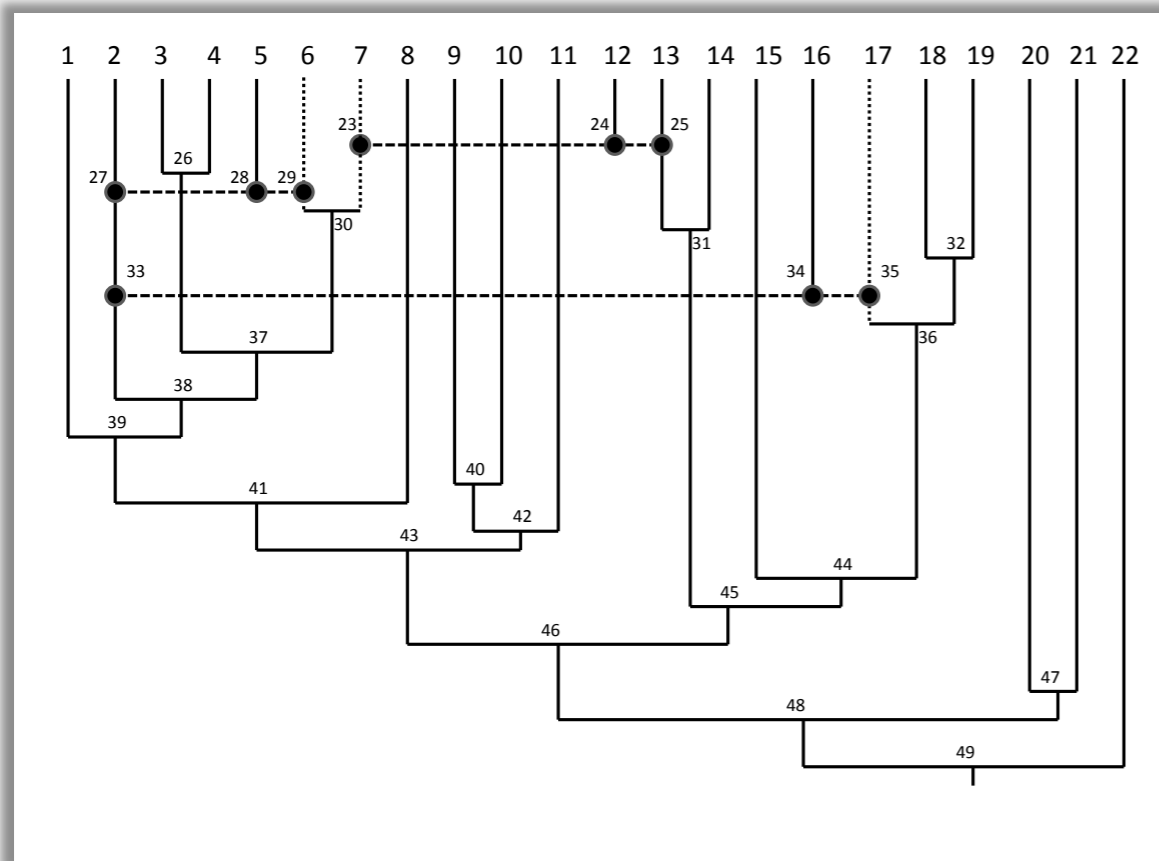
The mean value of a trait in a species is affected by multiple factors: physical constraints on evolution, lack of variation, change due to finite population size, and trade-offs between different optima. From one generation to the next, a trait value could change due to processes such as genetic drift, selection towards an optimum, or mutational pressure. If these movements are independent and identically distributed and have an additive effect through time, by the central limit theorem, evolution will fit a Brownian motion process (if the movements have a multiplicative effect through time, the log of the trait value will be evolving under Brownian motion). An Ornstein-Uhlenbeck (OU) process would better describe the process if these movements tended to be in the direction of a particular trait value (such that species with a trait value larger tend to evolve a smaller trait value).

B. C. O'Meara (✉)
 Department of Ecology and Evolutionary Biology, Knoxville, TN, USA
 e-mail: bomeara@utk.edu

J. M. Beaulieu
 National Institute for Biological and Mathematical Synthesis, University of Tennessee,

Continuous trait evolution, ongoing

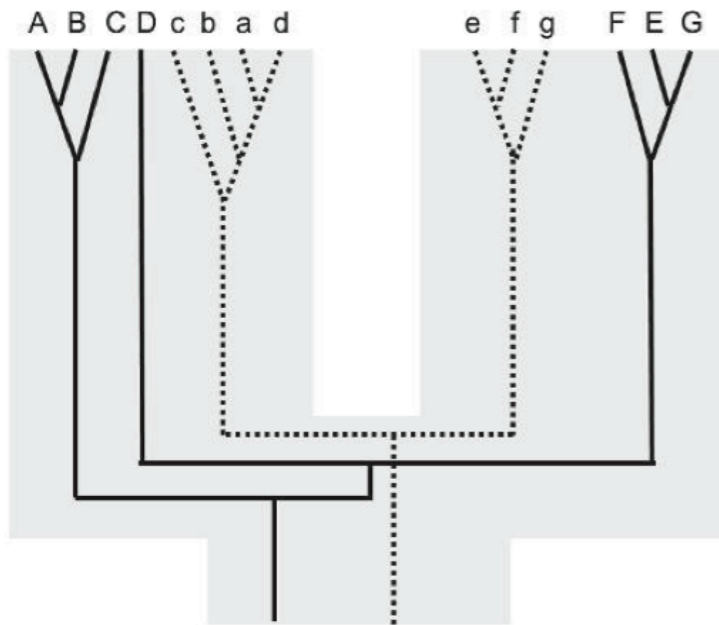
Brownian motion on phylogenetic networks, not trees



$$\mathbf{V}_R = \begin{matrix} X \\ R_1 \\ Y \\ R_2 \\ Z \end{matrix} \begin{bmatrix} 5\sigma^2 & 2\beta_1\sigma^2 & 0 & 0 & 0 \\ 2\beta_1\sigma^2 & (3 + 4\beta_1^2)\sigma^2 + v_{1,H} & 2\beta_1\sigma^2 & 2\beta_1\beta_2\sigma^2 & \beta_1\sigma^2 \\ 0 & 2\beta_1\sigma^2 & 5\sigma^2 & 3\beta_2\sigma^2 & \sigma^2 \\ 0 & 2\beta_1\beta_2\sigma^2 & 3\beta_2\sigma^2 & (2 + 5\beta_1^2)\sigma^2 + v_{2,H} & 3\beta_2\sigma^2 \\ 0 & \beta_1\sigma^2 & \sigma^2 & 3\beta_2\sigma^2 & 5\sigma^2 \end{bmatrix}$$

Jhweng & O'Meara, in revision

Species delimitation



$$\text{Total cost} = (1-wt) \cdot \text{GTP cost} + wt \cdot \text{structure}$$

$$\text{Total cost} = 0.5 \cdot (1 + 0) + 0.5 \cdot (1.009 + 0)$$

$$\text{Total cost} = 1.0045$$

structure

$((a,d),b)$
 $((a,d),c)$
 $((A,B),C) = ((a,b),c)$
 $((b,d),e)$

$((e,f),g) \neq ((E,G),F)$

$((1/0.33143)-1)/2 = 1.009$

$((1/1)-1)/2 = 0$

gene tree parsimony (GTP)

1 duplication (solid) + 0 duplications (dotted)

Syst. Biol. 59(1):59–73, 2010.
 © The Author(s) 2009. Published by Oxford University Press on behalf of Society of Systematic Biologists.
 This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
 DOI:10.1093/sysbio/syp077
 Advance Access publication on November 10, 2009

New Heuristic Methods for Joint Species Delimitation and Species Tree Inference

BRIAN C. O'MEARA*

Department of Ecology & Evolutionary Biology, University of Tennessee, 569 Dabney Hall, Knoxville, TN 37996-1610, USA;
 *Correspondence to be sent to: Department of Ecology & Evolutionary Biology, University of Tennessee, 569 Dabney Hall, Knoxville, TN 37996-1610, USA; E-mail:bomeara@utk.edu.

Received 14 February 2008; revision returned 3 June 2008; accepted 21 September 2009
 Associate Editor: L. Lacey Knowles

Abstract.—Species delimitation and species tree inference are difficult problems in cases of recent divergence, especially when different loci have different histories. This paper quantifies the difficulty of jointly finding the division of samples to species and estimating a species tree without constraining the possible assignments a priori. It introduces a parametric and a nonparametric method, including new heuristic search strategies, to do this delimitation and tree inference using individual gene trees as input. The new methods were evaluated using thousands of simulations and 4 empirical data sets. These analyses suggest that the new methods, especially the nonparametric one, may provide useful insights for systematists working at the species level with molecular data. However, they still often return incorrect results. [Brownie; gene tree parsimony; gene tree species tree; speciation; species delimitation.]

Two of the main goals of systematics are dividing the diversity of life into species and discovering the phylogenetic relationships of these species. Both can be difficult to achieve. Processes such as lineage sorting, introgression, and undetected gene duplication may cause gene trees to disagree with the true tree of species, potentially obscuring the species tree signal (Fitch 1970; Goodman et al. 1979; Avise 1983; Tajima 1983; Pamilo and Nei 1988; Doyle 1992; Hudson 1992; Maddison 1997). For species delimitation, a systematist must choose both a species concept and a criterion to apply this species concept to data. Even if speciation itself is effectively instantaneous, the time required for sufficient evolutionary changes to appear to allow 2 distinct lineages to be recognized will not be (De Queiroz 2007). This causes delimitation of species to be difficult.

These two questions are biologically linked but rarely methodologically coupled. If intervals between speciation events were long enough that all species were monophyletic for all their genes, once the species were correctly delimited, any species could be adequately represented by a single individual on a phylogeny. In reality, putatively independently evolving lineages are often not monophyletic (Funk and Omland 2003). The phylogeny of species, unless they are defined under a strict genealogical species concept (GSC; Baum and Shaw 1995; Hudson and Coyne 2002), will have an assortment of independent evolutionary lineages, which will probably include parphyly for at least some of their genes. Here, I attempt to unite these two questions as the more general one of jointly inferring the species boundaries and the species tree. I calculate the computational complexity of the problem, develop and implement methods for addressing it, and perform simulations and analyses across hundreds of parameter combinations to evaluate the feasibility. I also analyze 4 empirical data sets, *Drosophila* (Machado et al. 2002; Machado and Hey 2003), *Manacus* (Passeriformes) (Brumfield et al. 2008), *Lactarius* fungi (Nuytinck and

Verbeke 2007), and *Melanoplus* grasshoppers (Carstens and Knowles 2007), to evaluate the performance of the new methods.

MATERIALS AND METHODS

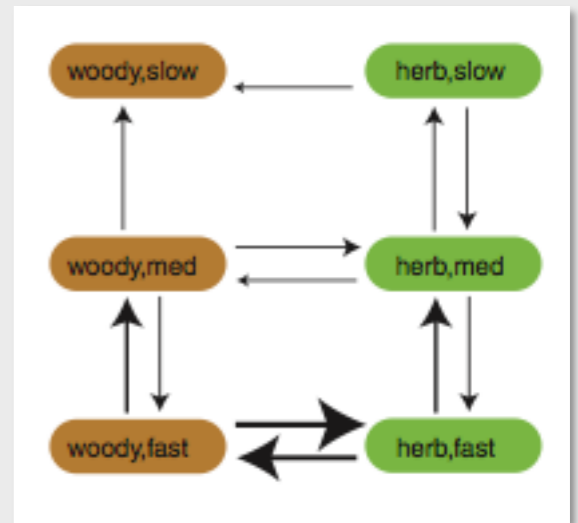
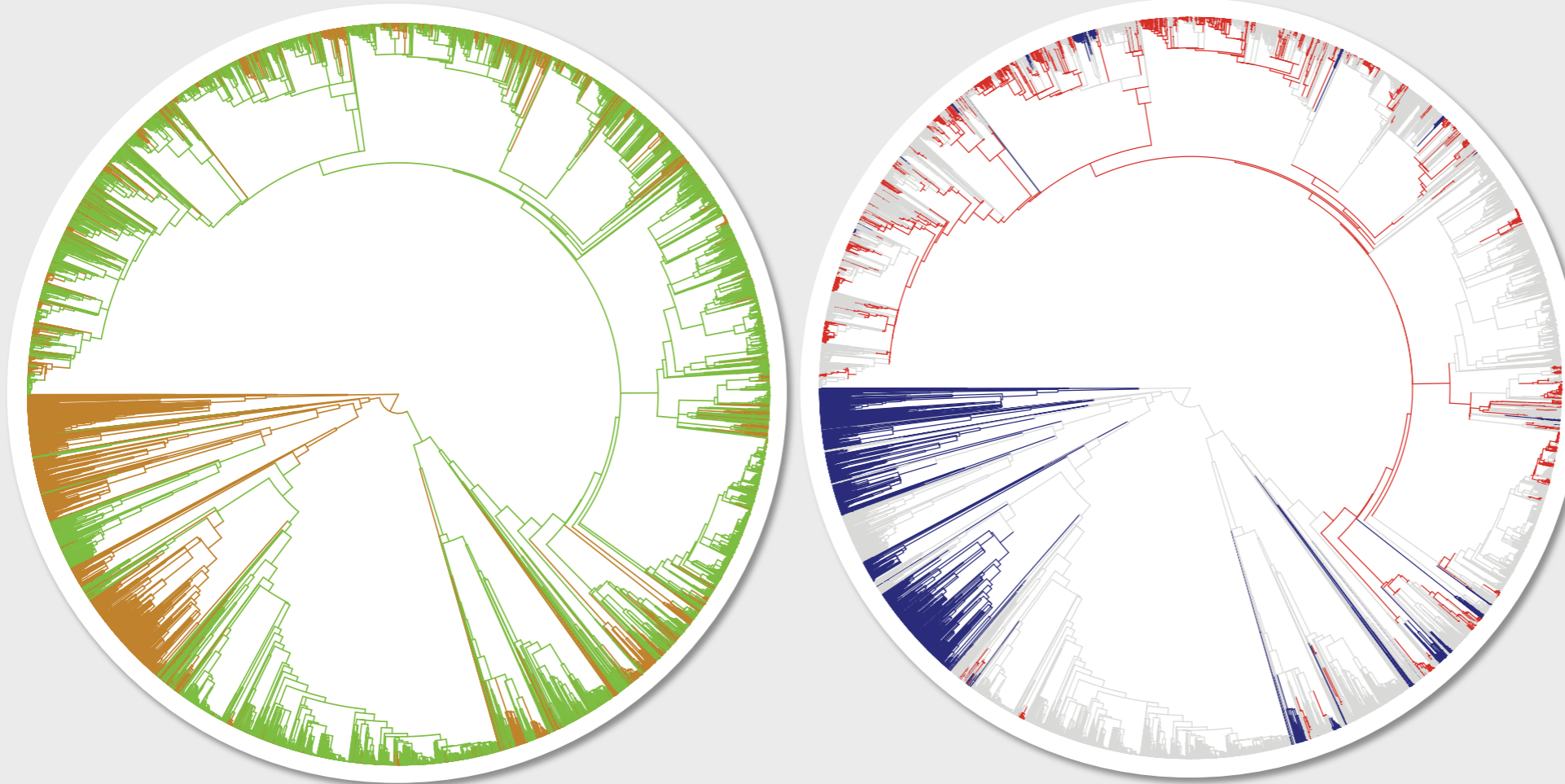
Problem Definition

Given a set of sequences from multiple individuals, the general problem is to allocate those individuals into putative species and estimate the species tree. This solution, the species tree with assignment of samples to species, is termed the “delimited species tree.” Optimally, a method will assign species and estimate the species tree correctly, in a statistically and computationally efficient manner. An estimate of the delimited species tree may differ from the true delimited species tree in topological error and/or through assignment of individuals to the wrong species. The latter might happen by merging 2 species that should be 1, splitting 1 true species into 2, having an individual of 1 species assigned to a different species, or a complex mixture of these. This is a more difficult problem than is typically addressed in DNA barcoding approaches (Hebert et al. 2003; Tautz et al. 2003), where 1 or more unknown individuals are assigned to existing species (Manel et al. 2005; Matz and Nielsen 2005; Abdo and Golding 2007; Zhang et al. 2008).

Most methods in systematics work on restrictions of this general problem, such as assuming that assignments to species are known (Nielsen and Wakeley 2001; Carstens and Knowles 2007; Edwards et al. 2007; Liu and Pearl 2007) or assuming that the gene tree matches the species tree (Hebert et al. 2003; Pons et al. 2006). I follow the approach of several recent authors (Pons et al. 2006; Carstens and Knowles 2007; Knowles and Carstens 2007; Mossel and Roch 2007; Kubatko et al. 2009), in restricting the problem using estimated gene trees as input rather than by integrating across a set of possible gene trees. The restricted problem still makes

Cited 135 times

Hidden rates



Chapter 16 Hidden Markov Models for Studying the Evolution of Binary Morphological Characters

Jeremy M. Beaulieu and Brian C. O'Meara

Abstract Biologists now have the capability of building large phylogenetic trees consisting of tens of thousands of species, from which important comparative questions can be addressed. However, to the extent that biologists have applied these large trees to comparative data, it is clear that current methods, such as those that deal with the evolution of binary morphological characters, make unrealistic assumptions about how these characters are modeled. As phylogenies increase both in size and scope, it is likely that the liability of a binary character will differ significantly among lineages. In this chapter, we describe how a new generalized model, which we refer to as the “hidden rates model” (HRM), can be used to identify different rates of evolution in a discrete binary character along different branches of a phylogeny. The HRM is part of a class of models that are more broadly known as Hidden Markov models because it presupposes that unobserved “hidden” rate classes underlie each observed state and that each rate class represents potentially different transition rates to and from these observed states. As we discuss, the recognition and accommodation of this heterogeneity can provide a robust picture of binary character evolution.

16.1 Introduction

Underlying many important discoveries in ecology, evolution, and behavior is the use of a phylogenetic tree. Phylogenies allow for the non-independence of taxa to be accounted for while also opening up new ways of examining how traits change

Stat. Biol. 42(3):278–292, 2013.
© 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/syt054
Advance Access publication May 14, 2013

Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms

JEREMY M. BEAULIEU^{1,*}, BRIAN C. O'MEARA², AND MICHAEL J. DONOGHUE¹

¹Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208106, New Haven, CT 06620-8106, USA; and ²Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-3030, USA
*Correspondence to be sent to: National Institute for Biological and Mathematical Synthesis, University of Tennessee, 1122 Volunteer Blvd, Ste. 106 Knoxville, TN 37996, USA; E-mail: jbeaulieu@immsion.org

Received 14 January 2013; revision returned 18 March 2013; accepted 6 May 2013
Associate Editor: Mark Fishelson

Abstract.—The growth of phylogenetic trees in scope and in size is promising from the standpoint of understanding a wide variety of evolutionary patterns and processes. With trees comprised of larger, older, and globally distributed clades, it is likely that the liability of a binary character will differ significantly among lineages, which could lead to errors in estimating transition rates and the associated inference of ancestral states. Here we develop and implement a new method for identifying different rates of evolution in a binary character along different branches of a phylogeny. We illustrate this approach by exploring the evolution of growth habit in Campanulidae, a flowering plant clade containing some 35,000 species. The distribution of woody versus herbaceous species calls into question the use of traditional models of binary character evolution. The recognition and accommodation of changes in the rate of growth form evolution in different lineages demonstrates, for the first time, a robust picture of growth form evolution across a very large, very old, and very widespread flowering plant clade. [Binary character; Campanulidae; comparative methods; flowering plants; growth habit; herbaceous; Hidden rates model; woody.]

With the availability of larger and larger phylogenetic trees comes the promise of evaluating more complicated models of evolution. Considerable attention has been focused on continuously varying characters, and it is now possible to apply complex, parameter-rich models for detecting differences on different branches of the tree in the processes associated with phenotypic evolution under Brownian motion (O'Meara et al. 2006; Thomas et al. 2006) or the Ornstein-Uhlenbeck process (Butler and King 2004; Beaulieu et al. 2012a). In contrast, discrete binary characters, which are often the focus of evolutionary studies, have received little attention, and we are continually forced to rely on relatively simple models that apply the same rates of state change to all branches in a tree (but see O'Meara 2007).

Simple models of binary character evolution may make sense for small clades, but they are not likely to adequately explain the evolution of such characters in larger, older, and globally distributed clades. In these instances one might expect the liability of a trait—or the propensity to undergo state changes—to differ significantly among clades. In flowering plants,

55,000 tips; Smith et al. 2011) we should have the power to fit models with more than a few parameters. Here we develop and implement a new method for identifying different rates of evolution in a binary character along different branches of a phylogeny, and we bring this to bear on the inference of ancestral states. We illustrate this approach by exploring the evolution of growth form in Campanulidae, an angiosperm clade containing some 35,000 species, including the familiar composites (sunflowers and relatives), umbels (carrots and relatives), and Dipsacales (honeysuckles and relatives). As we will show, the phylogenetic distribution of woody versus herbaceous species calls into question the use of traditional models of binary character evolution, and the recognition and accommodation of changes in the rate of growth form evolution in different lineages casts a new light on the diversification of campanulids.

Botany
Canadian Botanical Association

A/B CENTENNIAL REVIEW

American Journal of Botany 101(4): 559–571, 2014.

REPEATED EVOLUTION OF TRICELLULAR (AND BICELLULAR) POLLEN¹

JOSEPH H. WILLIAMS^{2,4}, MACKENZIE L. TAYLOR³, AND BRIAN C. O'MEARA²

¹Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996, USA; and ²Department of Biology, Creighton University, Omaha, Nebraska 68178, USA

- Premise of study.**—Male gametophytes of seed plants are sexually immature at the time they are dispersed as pollen, but approximately 30% of flowering plants have tricolpate pollen containing fully formed sperm at anthesis. The classic study of Brewbaker (1967; *American Journal of Botany* 54: 1069–1083) provided a powerful confirmation of the long-standing hypothesis that tricolpate pollen had many parallel and irreversible origins within angiosperms. We addressed the main questions of that study with modern comparative phylogenetic methods.
- Methods.**—We used our own and more recent reports to greatly expand the Brewbaker data set. We modeled trait evolution for 2511 species on a time-calibrated angiosperm phylogeny using (1) Binary State Speciation and Extinction (BISSE), which accounts for the effect of species diversification rates on character transition rates and, (2) the hidden rates model (HRM), which incorporates variation in transition rates across a phylogeny.
- Key results.**—Seventy percent of species had bicellular pollen. BISSE found a 1.9-fold higher bicellularity to tricolpate transition rate than in the reverse direction, and bicellularity lineages had a 1.8-fold higher diversification rate than tricolpate lineages. HRM found heterogeneity in evolutionary rates, with bidirectional transition rates in three of four rate classes.
- Conclusions.**—The tricolpate condition is not irreversible. Pollen cell numbers are maintained at intermediate frequencies because lower net diversification rates of tricolpate lineages are counterbalanced by slower state shifts to the bicellularity condition. The tricolpate lineages diversify slowly and give rise to bicellularity lineages slowly reflects a linkage between the evolution of spermatophyte life cycle and the developmental liability of male gametophytes.

Key words: cell cycle; constraint; diversification rate; Dollo's law; evolution of development; gametogenesis; heterostyly; parallelism; pollen germination; trade-off.

At sexual maturity, the male gametophyte of flowering plants is tricolpate, consisting of a single somatic cell that forms a pollen tube that carries in its cytoplasm two free sperm cells. The first microscopic studies of angiosperm pollen anatomy found that male gametophytes were immature and binucleate at the time of pollen dispersal (Strasburger, 1877; Elfvig, 1879). In such species, the generative nucleus undergoes mitosis to form two sperm nuclei after pollination, within the pollen tube cytoplasm. Hereafter, we refer to male gametic nuclei as generative cells and sperm cells, since it has since been shown that they are not free nuclei within the tube cell cytoplasm (Maheshwari, 1950; Knox, 1984a). The pollen tube itself was first observed in *Portulaca oleracea* (Aitn., 1824), a species that later was found to form its two sperm cells before pollen dispersal (Cooper, 1935). It was soon apparent that species that formed their sperm late (bicellularity pollen) were phylogenetically widespread,

whereas those that formed sperm early (tricolpate pollen) seemed to be restricted to aquatics, grasses, and some herbs (Elfvig, 1879; Strasburger, 1884). Schirrhoff (1926) proposed that tricolpate pollen was independently derived numerous times and that the condition was irreversible. Subsequent surveys of many more species strengthened that hypothesis (Scharf, 1939; Rudenko, 1959), but it was the massive study of J. L. Brewbaker (1967), published in *American Journal of Botany*, that seemed so elegantly and powerfully to settle the matter once and for all. The landmark Brewbaker study stands as one of the first large-scale tests of an evolutionary developmental hypothesis using character mapping onto an explicit phylogenetic tree. The data came from his studies of 812 species and an additional 1096 species from the literature, representing 1219 genera. Based on the species-level data, 265 families were coded as having either bicellularity, tricolpate, or polymorphic pollen dispersal states. Coded families were then placed onto a bifurcating ordinal-level tree constructed from what was known of angiosperm relationships at the time. The resulting two-page phylogeny displayed a pattern in which families with tricolpate pollen repeatedly appeared in nested (or potentially nested) positions in both monocots and dicots, whereas bicellularity families were never clearly nested within tricolpate groups. This distribution was consistent with Schirrhoff's (1926) proposal of many

¹Manuscript received 27 November 2013; revision accepted 4 March 2014.
The authors thank J. Jernstedt for the illustration, A. E. Zamec and co-authors for sharing their “in press” phylogeny, J. Beaulieu for assistance with ancestral state reconstruction, A. Muller for help in the greenhouse, G. Gaskin and P. Oliver for help in the field, and S. Fitch, N. Buckley, and B. Lawing for their efforts on an earlier version of this review. This work was supported by National Science Foundation award BNS 1052924.



corHMM: Analysis of binary character evolution

Fits a hidden rates model that allows different transition rate classes on different portions of a phylogeny by treating rate classes as hidden states in a Markov process and various other functions for evaluating models of binary character evolution.

Version: 1.15
Depends: [ape](#), [nloptr](#)
Imports: [expm](#), [numDeriv](#), [corpcor](#), [phangorn](#), [rgenoud](#), [parallel](#)
Published: 2014-09-18
Author: Jeremy M. Beaulieu, Jeffrey C. Oliver, Brian O'Meara

Empirical papers

Botany

American Journal of Botany 100(5): 916–929. 2013.

PHYLOGENETIC RELATIONSHIPS AND CHARACTER EVOLUTION ANALYSIS OF SAXIFRAGALES USING A SUPERMATRIX APPROACH¹

DOUGLAS E. SOLTIS^{2,3,7}, MARK E. MORT⁴, MARIBETH LATVIS², EVGENY V. MAVRODIEV³, BRIAN C. O'MEARA⁵, PAMELA S. SOLTIS³, J. GORDON BURLEIGH²,

MOLECULAR ECOLOGY RESOURCES

Molecular Ecology Resources (2013) 13, 347–353

doi: 10.1111/1755-0998.12066

Mitochondrial genome primers for Lake Malawi cichlids

C. DARRIN HULSEY, BENJAMIN P. KECK, HUGO ALAMILLO and BRIAN C. O'MEARA
Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA

Abstract

Resolving the evolutionary relationships among the 200+ species of Lake Malawi cichlids is a major challenge. Mitochondrial genome primers for individual loci, the hypothesis for the could have substantial diverse clade.

Keywords: adaptive evolution, Lake Malawi cichlids, mitochondrial DNA, phylogenetics

Received 20 August 2012

Introduction

Reconstructing the evolutionary relationships among the 200+ species of Lake Malawi cichlids is a major challenge. This is largely due to the large number of species and the high degree of genetic divergence among them. We have developed a set of primers for mitochondrial DNA regions that are highly conserved among cichlids and are suitable for phylogenetic analysis.

Components of the mitochondrial genome are frequently used to infer relationships among species.

Correspondence: C. D. Hulsey, E-mail: chulsey@outlook.com

© 2013 Blackwell Publishing Ltd

100 YEARS Journal of Ecology

Journal of Ecology 2014, 102, 345–356

doi: 10.1111/1365-2745.12208

SPECIAL FEATURE

THE TREE OF LIFE IN ECOSYSTEMS

Functional distinctiveness of major plant lineages

William K. Cornwell^{1,2}, Mark Westoby³, Daniel S. Falster³, Richard G. FitzJohn^{3,4}, Brian C. O'Meara⁵, Matthew W. Pennell⁶, Daniel J. McGlenn⁷, Jonathan M. Eastman⁸, Angela T. Moles², Peter B. Reich⁹, David C. Tank⁹, Ian J. Wright⁹, Lonnie Aarssen⁹, Jeremy M. Beaulieu¹⁰, Robert M. Kooyman^{3,11}, Michelle R. Leishman⁹, Eliot T. Miller¹², Ülo Niinemets¹³, Jacek Oleksyn¹⁴, Alejandro Ordóñez¹⁵, Dana L. Royer¹⁶, Stephen A. Smith¹⁷, Peter F. Stevens¹², Laura Warman^{2,18}, Peter Wiif¹⁹ and Amy E. Zanne^{20,21*}

¹Department of Ecological Sciences, Systems Ecology, VU University, de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands; ²Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia; ³Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia; ⁴Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; ⁵Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA; ⁶Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844, USA; ⁷Department of Biology and the Ecology Center, Utah State University, Logan, UT 84322, USA; ⁸Department of Forest Resources, University of Minnesota, St. Paul, MN 55108, USA; ⁹Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada; ¹⁰National Institute for Mathematical & Biological Synthesis, University of Tennessee, Knoxville, TN 37996, USA; ¹¹National Herbarium of New South Wales, Royal Botanic Gardens and Domain Trust, Mrs. Macquaries Rd., Sydney, NSW 2000, Australia; ¹²Department of Biology, University of Missouri – St. Louis, St. Louis, MO 63121, USA; ¹³Department of Plant Physiology, Estonian University of Life Sciences, Kreutzwaldi 1, 51014 Tartu, Estonia; ¹⁴Polish Academy of Sciences, Institute of Dendrology, 62-035 Koronik, Poland; ¹⁵Nelson Institute Center for Climatic Research, Madison, WI 53706, USA; ¹⁶Department of Earth and Environmental Sciences, Wesleyan University, Middletown, CT 06459, USA; ¹⁷Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA; ¹⁸Institute of Pacific Islands Forestry, USDA Forest Service, Hilo, HI 96720, USA; ¹⁹Department of Geosciences, Pennsylvania State University, University Park, PA 16802, USA; ²⁰Department of Biological Sciences, George Washington University, Washington, DC 20052, USA; and ²¹Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, MO 63121, USA

Summary

1. Plant traits vary widely across species and underpin differences in ecological strategy. Despite centuries of interest, the contributions of different evolutionary lineages to modern-day functional diversity remain poorly quantified.
2. Expanding data bases of plant traits plus rapidly improving phylogenies enable for the first time a data-driven global picture of plant functional diversity across the major clades of higher plants. We mapped five key traits relevant to metabolism, resource competition and reproductive strategy onto a phylogeny across 48324 vascular plant species world-wide, along with climate and biogeographic data. Using a novel metric, we test whether major plant lineages are functionally distinctive. We then highlight the trait-lineage combinations that are most functionally distinctive within the present-day spread of ecological strategies.
3. For some trait-clade combinations, knowing the clade of a species conveys little information to neo- and palaeo-ecologists. In other trait-clade combinations, the clade identity can be highly revealing, especially informative clade-trait combinations include Proteaceae, which is highly distinctive, representing the global slow extreme of the leaf economic spectrum. Magnoliidae and Rosidae contribute large leaf sizes and seed masses and have distinctively warm, wet climatic distributions.

*Correspondence author. E-mail: aeazanne@gmail.com

© 2014 The Authors. Journal of Ecology © 2014 British Ecological Society

Abercrombie et al. *EvoDevo* 2011, 2:14
http://www.evodevojournal.com/content/2/1/14



RESEARCH

Open Access

Developmental evolution of flowering plant pollen tube cell walls: callose synthase (*CalS*) gene expression

Jason M Abercrombie

doi:10.1111/j.1420-9101.2010.01971.x

Abstract

Background: A newly discovered early step involved in pollen tube growth is the synthesis and deposition of callose (1,2-β-D-glucan) and septae (callose cross-links) in the cell walls of the pollen tube. The *CalS* gene family encodes the callose synthase enzyme. We investigated the origin of *CalS* orthologues in the angiosperm lineage and determined the expression patterns of *CalS* orthologues in Nymphaeales and other early diverging angiosperm lineages. **Results:** We investigated the origin of *CalS* orthologues in the angiosperm lineage and determined the expression patterns of *CalS* orthologues in Nymphaeales and other early diverging angiosperm lineages. **Conclusion:** The first *CalS* orthologue in the angiosperm lineage is found in Nymphaeales. The expression patterns of *CalS* orthologues in Nymphaeales and other early diverging angiosperm lineages indicate that the pollen tube cell wall is a novel feature of the angiosperm lineage.

Background

The pollen tube is a key feature of seed plants. It is a long-lived and complex structure that is essential for the delivery of sperm cells to the egg cell. The pollen tube cell wall is a novel feature of the angiosperm lineage and is composed of callose and cellulose. The pollen tube cell wall is a complex structure that is essential for the delivery of sperm cells to the egg cell.

Correspondence: joewilliams@biology.utk.edu
Department of Ecology and Evolutionary Biology, Knoxville, TN, USA



Introduction

One of the most important features of the pollen tube is its ability to deliver sperm cells to the egg cell. This is achieved through the growth of the pollen tube through the style. The pollen tube cell wall is a complex structure that is essential for the delivery of sperm cells to the egg cell.

Correspondence: David J. McPherson, E-mail: dmcphe@biology.harvard.edu
Evolutionary Biology and Systematics, Harvard University, Cambridge, MA 02138, USA
Tel: 617 495 9099
Re-use of this article is permitted. Reprints and permissions information is available at www.blackwell-synergy.com/permissions

© 2010 THE AUTHOR
JOURNAL COMPANION

© 2010 THE AUTHOR
JOURNAL COMPANION

¹Department of Biological Sciences, George Washington University, Washington DC 20052, USA; ²Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, Missouri, 63121, USA; ³Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844, USA; ⁴Institute for Bioinformatics and Evolutionary Studies, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada; ⁵Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia; ⁶Department of Biology and the Ecology Center, Utah State University, Logan, Utah 84322, USA; ⁷Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996, USA; ⁸Department of Forest Resources, University of Minnesota, St. Paul, Minnesota 55108, USA; ⁹Hawkesbury Institute for the Environment, University of Western Sydney, Penrith, New South Wales 2751, Australia; ¹⁰Department of Earth and Environmental Sciences, Wesleyan University, Middletown, Connecticut 06459, USA; ¹¹Department of Biology, University of Florida, Gainesville, Florida 32611, USA; ¹²National Institute for Mathematical & Biological Synthesis, University of Tennessee, Knoxville, Tennessee 37996, USA; ¹³Genetics Institute, University of Florida, Gainesville, Florida 32611, USA; ¹⁴Department of Biology, University of Missouri – St. Louis, St. Louis, Missouri 63121, USA; ¹⁵Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada; ¹⁶Department of Biology, College of the Holy Cross, Worcester, Massachusetts 01610, USA; ¹⁷Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom; ¹⁸Polish Academy of Sciences, Institute of Dendrology, 62-035 Koronik, Poland; ¹⁹Department of Plant Physiology and Ecology, Estonian University of Life Sciences, Kreutzwaldi 1, 51014 Tartu, Estonia; ²⁰Department of Biological Sciences, George Washington University, Washington, DC 20052, USA; and ²¹Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, Missouri 63121, USA

LETTER

doi:10.1038/nature12872

Three keys to the radiation of angiosperms into freezing environments

Amy E. Zanne^{1,2}, David C. Tank^{3,4}, William K. Cornwell^{5,6}, Jonathan M. Eastman^{3,4}, Stephen A. Smith⁷, Richard G. FitzJohn^{8,9}, Daniel J. McGlenn¹⁰, Brian C. O'Meara¹¹, Angela T. Moles⁶, Peter B. Reich^{12,13}, Dana L. Royer¹⁴, Douglas E. Soltis^{15,16,17}, Peter F. Stevens¹⁸, Mark Westoby⁹, Ian J. Wright⁹, Lonnie Aarssen⁹, Robert L. Bertin²⁰, Andre Calaminus¹³, Rafaël Govaerts²¹, Frank Hemming⁶, Michelle R. Leishman⁹, Jacek Oleksyn^{12,22}, Pamela S. Soltis^{16,17}, Nathan G. Swenson²³, Laura Warman^{6,24} and Jeremy M. Beaulieu²⁵

Early flowering plants are thought to have been woody species restricted to warm habitats^{1–3}. This lineage has since radiated into almost every climate, with manifold growth forms⁴. As angiosperms spread and climate changed, they evolved mechanisms to cope with episodic freezing. To explore the evolution of traits underpinning the ability to persist in freezing conditions, we assembled a large species-level database of growth habit (woody or herbaceous; 49,064 species), as well as leaf phenology (evergreen or deciduous), diameter of hydraulic conduits (that is, xylem vessels and tracheids) and climate occupancies (exposure to freezing). To model the evolution of species' traits and climate occupancies, we combined these data with an unparalleled dated molecular phylogeny (32,223 species) for land plants. Here we show that woody clades successfully moved into freezing-prone environments by either possessing transport networks of small safe conduits⁵ and/or shutting down hydraulic function by dropping leaves during freezing. Herbaceous species largely avoided freezing periods by senescing cheaply constructed aboveground tissue. Growth habit has long been considered labile⁶, but we find that growth habit was less labile than climate occupancy. Additionally, freezing environments were largely filled by lineages that had already become herbs or, when remaining woody, already had small conduits (that is, the trait evolved before the climate occupancy). By contrast, most deciduous woody lineages had an evolutionary shift to seasonally shedding their leaves only after exposure to freezing (that is, the climate occupancy evolved before the trait). For angiosperms to inhabit novel cold environments; our results suggest that many of these solutions were probably acquired before their foray into the cold.

Flowering plants (angiosperms) today grow in a vast range of environmental conditions, with this breadth probably related to their diverse morphology and physiology⁷. However, early angiosperms are generally thought to have been woody and restricted to warm understory habitats^{1–3}. Debate continues about these assertions, in part because of the paucity of fossils and uncertainty in reconstructing habits for these first representatives^{8–11}. Nevertheless, greater mechanical strength of woody tissue would have made extended lifespans possible at a height necessary to compete for light^{12,13}. A major challenge resulting from increased stature is that hydraulic systems must deliver water at tension

to greater heights: as path lengths increase so too does resistance¹⁴. Among extant strategies, the most efficient method of water delivery is through large-diameter water-conducting conduits (that is, vessels and tracheids) within xylem¹⁵. Early in angiosperm evolution they probably evolved larger conduits for water transport, especially compared with their gymnosperm cousins¹⁴. Although efficient in delivering water, these larger cells would have impeded angiosperm colonization of regions characterized by episodic freezing^{16,17}, as the propensity for freezing-induced embolisms (air bubbles produced during freeze/thaw events that block hydraulic pathways)

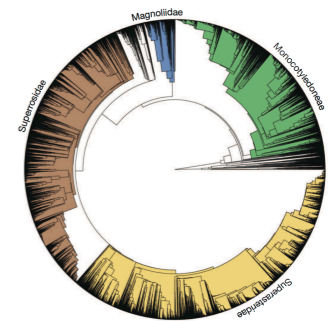


Figure 1 | Time-calibrated maximum-likelihood estimate of the molecular phylogeny for 31,749 species of seed plants. The four major angiosperm lineages discussed in the text are highlighted: Monocotyledoneae (green), Magnoliidae (blue), Superrosidae (brown) and Non-seed plants (yellow). Non-seed plant outgroups (that is, bryophytes, lycophytes and monilophytes) were removed for the purposes of visualization.

¹Department of Biological Sciences, George Washington University, Washington DC 20052, USA; ²Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, Missouri, 63121, USA; ³Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844, USA; ⁴Institute for Bioinformatics and Evolutionary Studies, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada; ⁵Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia; ⁶Department of Biology and the Ecology Center, Utah State University, Logan, Utah 84322, USA; ⁷Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996, USA; ⁸Department of Forest Resources, University of Minnesota, St. Paul, Minnesota 55108, USA; ⁹Hawkesbury Institute for the Environment, University of Western Sydney, Penrith, New South Wales 2751, Australia; ¹⁰Department of Earth and Environmental Sciences, Wesleyan University, Middletown, Connecticut 06459, USA; ¹¹Department of Biology, University of Florida, Gainesville, Florida 32611, USA; ¹²National Institute for Mathematical & Biological Synthesis, University of Tennessee, Knoxville, Tennessee 37996, USA; ¹³Genetics Institute, University of Florida, Gainesville, Florida 32611, USA; ¹⁴Department of Biology, University of Missouri – St. Louis, St. Louis, Missouri 63121, USA; ¹⁵Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada; ¹⁶Department of Biology, College of the Holy Cross, Worcester, Massachusetts 01610, USA; ¹⁷Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom; ¹⁸Polish Academy of Sciences, Institute of Dendrology, 62-035 Koronik, Poland; ¹⁹Department of Plant Physiology and Ecology, Estonian University of Life Sciences, Kreutzwaldi 1, 51014 Tartu, Estonia; ²⁰Department of Biological Sciences, George Washington University, Washington, DC 20052, USA; and ²¹Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, Missouri 63121, USA

Scaling methods for large trees

treePL: divergence time estimation using penalized likelihood for large phylogenies

Stephen A. Smith^{1,*} and Brian C. O'Meara²

¹Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, 48109 and ²Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, 37996 USA

Associate Editor: David Posada

ABSTRACT

Summary: Ever larger phylogenies are being constructed due to the explosion of genetic data and development of high-performance phylogenetic reconstruction algorithms. However, most methods for calculating divergence times are limited to datasets that are orders of magnitude smaller than recently published large phylogenies. Here, we present an algorithm and implementation of a divergence time method using penalized likelihood that can handle datasets of thousands of taxa. We implement a method that combines the standard derivative-based optimization with a stochastic simulated annealing approach to overcome optimization challenges. We compare this approach with existing software including r8s, PATHd8 and BEAST.

Availability: Source code, example files, binaries and documentation for treePL are available at <https://github.com/blackrim/treePL>.

Contact: eebsmith@umich.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on April 14, 2012; revised on July 31, 2012; accepted on August 1, 2012

1 INTRODUCTION

Divergence times make phylogenetic hypotheses easier to interpret in light of other information on geology, biogeography and co-diversification. A number of methods exist for transforming branch lengths to be proportional to time. However, many of these methods do not scale well for phylogenies with thousands of taxa. The number of species included in published phylogenies has exploded (Goloboff *et al.*, 2009; Smith *et al.*, 2009). With the increase in the size of phylogenies, there is the need to develop divergence time methods that are capable of handling larger datasets. Here, we present a method, implemented in the software treePL, for calculating divergence time estimates using penalized likelihood (Sanderson, 2002) on large phylogenies. Penalized likelihood uses a semi-parametric approach that allows for different rates on different branches but has a smoothing parameter, set using cross-validation, that affects how much rate differences over the tree are penalized. Our approach attempts to overcome the problem of local minima, a problem amplified by including more taxa.

2 METHODS

2.1 Penalized likelihood

The optimality criterion used in this article is the penalized likelihood framework described in Sanderson (2002). Note that though this is framed as a likelihood method, it is possible to interpret this as a maximum *a posteriori* estimate in a Bayesian context given particular priors (Thorne and Kishino, 2005). It is also possible, following Sanderson's modifications to r8s, to use a related penalty that uses the differences of log rates rather than differences of untransformed rates (the additive penalty). This has been implemented in treePL, but analyses below use the original additive penalty.

2.2 The algorithm

Divergence time estimation, and penalized likelihood especially, presents a number of optimization challenges. One challenge is the large number of parameters and the ratio of free parameters to observations. This can be dampened by a large penalty function, but still presents a parametric optimization problem. The second challenge is the presence of barriers placed by both the user and the hierarchical nature of phylogenies. We present a first step to overcoming these problems with extremely large phylogenies. Our algorithm includes two main phases: a greedy hill-climbing phase and a stochastic phase. This combination is meant to both speed optimization and avoid local optima.

The greedy phase of the optimization step consists of two gradient-based optimization rounds. The first round of optimization uses gradient values from an exact gradient calculator with math derived from Sanderson (2002). The second round of optimization uses auto-differentiation (Gay, 2005; Griewank *et al.*, 1996). Auto-differentiation, because it has the same complexity as the original function, often continues hill climbing after the exact gradient calculator fails. However, we have found, empirically, that alone it is not as efficient and requires more rounds and restarts than combining exact gradients and auto-differentiation.

We found that a stochastic phase, in this case a partial simulated annealing procedure, is helpful in easing the parameters away from local optima. We consider it partial as the step does not consist of a full simulated annealing run. Instead, the phase consists of thousands of stochastic optimization generations followed by intermittent greedy phases. Although this additional step may lengthen convergence time, the result will often improve.

The performance of optimization methods can vary dramatically by dataset. We have provided a 'priming' step that will run through each optimization option and report those that show the greatest difference in starting and stopping likelihood scores. Although this can help, researchers should replicate analyses and experiment with settings to assure convergence. We have also provided a 'wiggle' option that will report divergence times that are within two log likelihood units of the best likelihood, suggested by Edwards (1992) as a confidence window. This allows for the visualization of nodes that are particularly uncertain, but



The iPlant collaborative: cyberinfrastructure for plant biology

Stephen A. Goff^{1,*}, Matthew Vaughn², Sheldon McKay¹, Eric Lyons¹, Ann E. Stapleton^{3,4}, Damian Gessler¹, Naim Matasci¹, Liya Wang⁵, Matthew Hanlon², Andrew Lenards¹, Andy Muir¹, Nirav Merchant¹, Sonya Lowry¹, Stephen Mock², Matthew Helmke¹, Adam Kubach², Martha Narro¹, Nicole Hopkins¹, David Micklos⁶, Uwe Hilgert⁶, Michael Gonzales², Chris Jordan², Edwin Skidmore¹, Rion Dooley², John Cazes², Robert McLay², Zhenyuan Lu⁵, Shiran Pasternak⁵, Lars Koesterke², William H. Piel⁷, Ruth Grene⁸, Christos Noutsos⁵, Karla Gendler², Xin Feng^{5,9}, Chunlao Tang⁵, Monica Lent¹, Seung-Jin Kim¹, Kristian Kvilekval¹⁰, B. S. Manjunath^{10,27}, Val Tannen¹¹, Alexandros Stamatakis¹², Michael Sanderson¹³, Stephen M. Welch¹⁴, Karen A. Cranston¹⁵, Pamela Soltis¹⁶, Doug Soltis¹⁷, Brian O'Meara¹⁸, Cecile Ane^{19,20}, Tom Bratnell²¹, Daniel J. Kleibenstein²², Jeffery W. White²³, James Leebens-Mack²⁴, Michael J. Donoghue²⁵, Edgar P. Spalding²⁶, Todd J. Vision²⁸, Christopher R. Myers³², David Lowenthal²⁹, Brian J. Enquist¹³, Brad Boyle¹³, Ali Akoglu³⁰, Greg Andrews²⁹, Sudha Ram³¹, Doreen Ware⁵, Lincoln Stein^{5,9} and Dan Stanzione²

¹ BIO5 Institute, University of Arizona, Tucson, AZ, USA

² Texas Advanced Computer Center, University of Texas, Austin, TX, USA

³ Department of Biology, University of North Carolina, Wilmington, NC, USA

⁴ Department of Marine Sciences, University of North Carolina, Wilmington, NC, USA

⁵ Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

⁶ DNA Learning Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

⁷ Yale Peabody Museum, Yale University, New Haven, CT, USA

⁸ Department of Plant Pathology, Physiology and Weed Science, Virginia Tech University, Blacksburg, VA, USA

⁹ Ontario Center for Cancer Research, Toronto, ON, Canada

¹⁰ Center for Bio-image Informatics, University of California, Santa Barbara, CA, USA

¹¹ Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA

¹² Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

¹³ Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

¹⁴ Department of Agronomy, Kansas State University, Manhattan, KS, USA

¹⁵ National Evolutionary Synthesis Center (NESCent), Durham, NC, USA

¹⁶ Florida Museum of Natural History, University of Florida, Gainesville, FL, USA

¹⁷ Department of Biology, University of Florida, Gainesville, FL, USA

¹⁸ Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA

¹⁹ Department of Statistics, University of Wisconsin, Madison, WI, USA

²⁰ Department of Botany, University of Wisconsin, Madison, WI, USA

²¹ Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY, USA

²² Department of Plant Sciences, University of California, Davis, CA, USA

²³ Arid-Land Agricultural Research Center, United States Department of Agriculture-Agricultural Research Service, Maricopa, AZ, USA

²⁴ Department of Plant Biology, University of Georgia, Athens, GA, USA

²⁵ Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

²⁶ Department of Botany, University of Wisconsin, Madison, WI, USA

²⁷ Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA

²⁸ Department of Biology, University of North Carolina, Chapel Hill, NC, USA

²⁹ Department of Computer Science, University of Arizona, Tucson, AZ, USA

³⁰ Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ, USA

³¹ Eller School of Business, University of Arizona, Tucson, AZ, USA

³² Department of Physics, Cornell University, Ithaca, NY, USA

Edited by:

Gane Ka-Shu Wong, University of Alberta, Canada

Reviewed by:

Michael Deyholos, University of Alberta, Canada

Sean W. Graham, University of British Columbia, Canada

*Correspondence:

Stephen A. Goff, iPlant Collaborative,

BIO5 Institute, University of Arizona,

Tucson, AZ 85721, USA.

e-mail: sgoff@iplantcollaborative.org

The iPlant Collaborative (iPlant) is a United States National Science Foundation (NSF) funded project that aims to create an innovative, comprehensive, and foundational cyberinfrastructure in support of plant biology research (PSCIC, 2006). iPlant is developing cyberinfrastructure that uniquely enables scientists throughout the diverse fields that comprise plant biology to address Grand Challenges in new ways, to stimulate and facilitate cross-disciplinary research, to promote biology and computer science research interactions, and to train the next generation of scientists on the use of cyberinfrastructure in research and education. Meeting humanity's projected demands for agricultural and forest products and the expectation that natural ecosystems be managed sustainably will require synergies from the application of information technologies. The iPlant cyberinfrastructure design is based on an unprecedented period of research community input, and leverages developments in high-performance computing, data storage, and cyberinfrastructure for the physical sciences. iPlant is an open-source project with application programming interfaces that allow the community to extend the infrastructure to meet its needs. iPlant is sponsoring community-driven workshops addressing specific

*To whom correspondence should be addressed.

Realistic substitution models

Work with Mike Gilchrist and JJ Chai

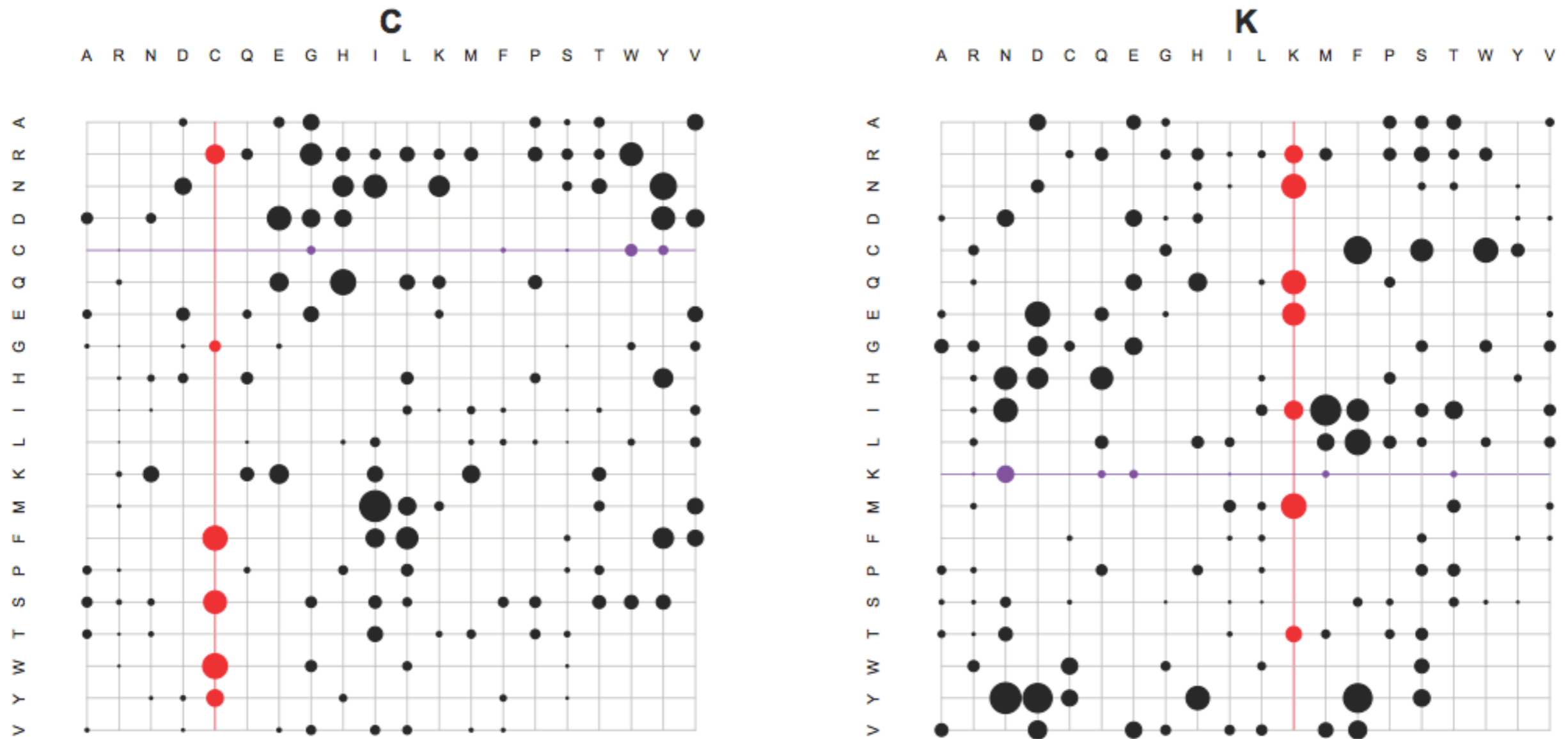


Figure 3: Instantaneous substitution rates from amino acid on row to amino acid in column under our preliminary model. The left panel is when cysteine is the optimal amino acid, the right is when lysine is the optimal amino acid. These are just two of the substitution matrices that are present in a given SAC_{AA} model. Bubble area (not diameter) shows the relative substitution rate from the amino acid listed on the left to the amino acid listed above. Red highlights the substitutions moving towards the optimal amino acid, purple those moving away from the optimal amino acid.

Large scale syntheses of the field

Evolutionary Inferences from Phylogenies: A Review of Methods

Brian C. O'Meara

Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996; email: bomeara@utk.edu, <http://www.brianomeara.info>

Keywords

methods, continuous-time Markov Chain, multivariate normal, birth-death, tree stretching

Abstract

There are many methods for making evolutionary inferences from phylogenetic trees. Many of these can be divided into three main classes of models: continuous-time Markov chain models with finite state space (CTMC-FSS), multivariate normal models, and birth-death models. Numerous approaches are just restrictions of more general models to focus on particular questions or kinds of data. Methods can be further modified with the addition of tree-stretching algorithms. The recent realization of the effect of correlated trait evolution with diversification rates represents an important advance that is slowly revolutionizing the field. Increased attention to model adequacy may lead to future methodological improvements.



ANNUAL REVIEWS Further

Click here for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Annu. Rev. Ecol. Syst. 2012. 43:267-85

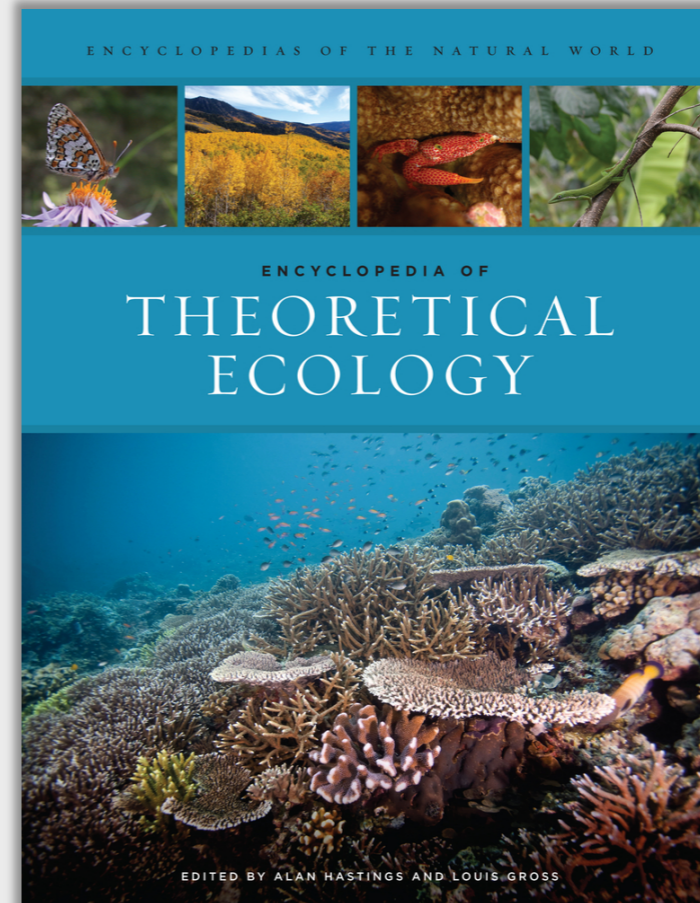
First published online as a Review in Advance on September 4, 2012

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

This article's doi: 10.1146/annurev-ecolsys-110411-160331

Copyright © 2012 by Annual Reviews. All rights reserved

1543-592X/12/1201-0267\$20.00



by adaptation in a fixed (nonplastic) genotype. Under this scenario, individuals having genetic variation for phenotypic plasticity will be favored. Whether the population will adapt by increasing its capacity for phenotypic plasticity ultimately depends on the costs and physiological limits associated with the plastic phenotype. Finally, it is worth noting that already in 1881 (in a letter to Karl Semper) Charles Darwin, clearly before his time, was able to envision the possibility of plastic phenotypes evolving in response to environmental changes: "I speculate whether a species very liable to repeated and great changes of conditions might not assume a fluctuating condition ready to be adapted to either condition."

SEE ALSO THE FOLLOWING ARTICLES

Environmental Heterogeneity and Plants / Evolution of Dispersal / Integrated Whole Organism Physiology / Mutation, Selection and Genetic Drift / Quantitative Genetics

FURTHER READING

Agrawal, A. A. 2001. Phenotypic plasticity in the interactions and evolution of species. *Science* 294: 321-326.
deWitt, T. J., and S. M. Scheiner, eds. 2004. *Phenotypic plasticity: functional and conceptual approaches*. New York: Oxford University Press.
Miner, B. G., S. E. Sultan, S. G. Morgan, D. K. Padilla, and R. A. Releya. 2005. Ecological consequences of phenotypic plasticity. *Trends in Ecology and Evolution* 20: 685-692.

- Pigliucci, M. 2001. *Phenotypic plasticity: beyond nature and nurture*. Syntheses in Ecology and Evolution. Baltimore, MD: Johns Hopkins University Press.
Pigliucci, M. 2005. Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology and Evolution* 20: 481-486.
Pfennig, D. W., M. A. Wund, E. C. Snell-Rood, T. Cruickshank, C. D. Schlichting, and P. Moczek. 2010. Phenotypic plasticity's impact on diversification and speciation. *Trends in Ecology and Evolution* 25: 459-467.
Price, T. D., A. Qvarnström, and D. E. Irwin. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society London B: Biological Sciences* 270: 1433-1440.
Tollrian, R., and C. D. Harvell, eds. 1998. *The ecology and evolution of inducible defenses*. Princeton: Princeton University Press.
West-Eberhard, M. J. 2003. *Developmental plasticity and evolution*. New York: Oxford University Press.

PHYLOGENETIC RECONSTRUCTION

BRIAN C. O'MEARA

University of Tennessee, Knoxville

A phylogeny is a depiction of the evolutionary history of a set of organisms. Typically, this is a branching diagram showing relationships between species, but phylogenies can be drawn for individual genes, for populations, or for other entities.

WHAT DO PHYLOGENIES MEAN?

A phylogeny represents a history of populations. Take the example of Figure 1. Starting at the bottom (root) of the tree, one population splits into two. The population on the left speciates again, but one of the descendant species eventually goes extinct without leaving any descendants. Various other processes occur: population sizes (width of the tree's branches) vary, speciation happens through a gradual rather than instant reduction of gene flow, populations develop and lose subdivision, one species forms as a hybrid of two other species, a few genes introgress from one species to another, and so forth. The history of genes evolving within these populations may be even more complex, with selective sweeps, ancestral polymorphisms persisting across speciation events, gene copies being duplicated and lost within the genome, and recombination shuffling histories within and between genes. All of this complex history is typically summarized by a figure like that of Figure 1, with most of the complex history abstracted away to leave only a simplified history of populations.

It is important to interpret phylogenies correctly. Chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*)

Much more...

Stoltzfus et al. *BMC Research Notes* 2012, 5:574
<http://www.biomedcentral.com/1756-0500/5/574>



DATA NOTE

Open Access

Sharing and re-use of phylogenetic trees (and associated data) to facilitate synthesis

Arlin Stoltzfus^{1*}, Brian O'Meara², Jamie Whitacre³, Ross Mounce⁴, Emily L Gillespie⁵, Sudhir Kumar⁶, Dan F Rosauer⁷ and Rutger A Vos⁸

Abstract

Background: phylogenetic trees are important for understanding the relationships between species. However, sharing and re-use of phylogenetic trees and associated data is difficult. We developed *Reol*, an open-source package for the R environment, which downloads data from the EOL API, searches for and extracts specific information, and builds tables with quantitative data and/or hierarchical classifications. We provide a detailed description how *Reol* can be used as a bridge between the R environment and the EOL API to extract quantitative or hierarchical content. It will be particularly useful for researchers who want information about taxonomic groups of interest (for example: how much research is available for a particular group of species).

Keywords

Keywords: Re-use of phylogenetic trees, data sharing, taxonomic information, open access, R environment, EOL API, quantitative data, hierarchical classifications.

* Correspondence: arlin.stoltzfus@ncsu.edu
¹Biochemical Sciences Division, NC State University, Raleigh, NC 27695-8002, USA
 Full list of author information is available at the end of the article



DE GRUYTER

Stat. Appl. Genet. Mol. Biol. 2014; 13(4): 459-475

Dwueng-Chwan Jhueng, Snehalata Huzurbazar, Brian C. O'Meara and Liang Liu* Investigating the performance of AIC in selecting phylogenetic models

Abstract: The popular likelihood-based model selection criterion, Akaike's Information Criterion (AIC), is a breakthrough mathematical result derived from information theory. AIC is an approximation to Kullback-Leibler (KL) divergence with the derivation relying on the assumption that the likelihood function has finite second derivatives. However, for phylogenetic estimation, given that tree space is discrete with respect to tree topology, the assumption of a continuous likelihood function with finite second derivatives is violated. In this paper, we investigate the relationship between the expected log likelihood of a candidate model, and the expected KL divergence in the context of phylogenetic tree estimation. We find that given the tree topology, AIC is an unbiased estimator of the expected KL divergence. However, when the tree topology is unknown, AIC tends to underestimate the expected KL divergence for phylogenetic models. Simulation results suggest that the degree of underestimation varies across phylogenetic models so that even for large sample sizes, the bias of AIC can result in selecting a wrong model. As the choice of phylogenetic models is essential for statistical phylogenetic inference, it is important to improve the accuracy of model selection criteria in the context of phylogenetics.

Keywords: AIC; Kullback-Leibler divergence; model selection; phylogenetics.

DOI 10.1515/sagmb-2013-0048

1 Introduction

Probabilistic models are fundamental to statistical phylogenetic inference (Johnson and Omland, 2004; Sullivan and Joyce, 2005; Kelchner, 2009). A phylogenetic model assumes that the evolution of molecular sequences follows a substitution process along the branches of a phylogenetic tree. The random process of nucleotide substitutions over time is described probabilistically by a substitution model; over the years many such substitution models have been developed. The parameters in a phylogenetic model include the branch lengths and topology of the phylogenetic tree, as well as the parameters in the substitution model (Bos and Posada, 2005).

One of the major goals of phylogenetic model selection is to select a good substitution model for estimating phylogenetic trees from sequence data (Shapiro et al., 2006). Since statistical approaches for phylogenetic inference are based on particular models, model choice may significantly affect the resulting estimates of the phylogenetic parameters (Buckley and Cunningham, 2002; Posada and Buckley, 2004). Standard model selection criteria have been introduced for selecting phylogenetic models, but the biggest challenge in

*Corresponding author: Liang Liu, Department of Statistics and Institute of Bioinformatics, University of Georgia, 101 Cedar Street, Athens, GA 30606 USA, Phone: +1-706-542-3309, Fax: +1-706-542-3391, e-mail: liu@uga.edu
 Dwueng-Chwan Jhueng: Department of Statistics, Feng-Chia University, Taichung, Taiwan 40724, R.O.C.
 Snehalata Huzurbazar: Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC 27709, USA; Department of Statistics, University of Wyoming, Laramie, WY 82071, USA; and Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA
 Brian C. O'Meara: Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA

Brought to you by | University of Tennessee Knoxville
 Authenticated | 10.248.254.158
 Download Date | 9/2/14 1:00 AM

Ecology and Evolution

Open Access

Reol: R interface to the Encyclopedia of Life

Barbara L. Banbury & Brian C. O'Meara

Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee

Keywords

Application programming interface, Encyclopedia of Life, programmatic access, reproducible research, taxonomy.

Correspondence

Barbara L. Banbury, Department of Biology, University of Washington, Seattle, WA, Tel: (206) 616-2132; Fax: (206) 543-3041; E-mail: bbanbury@uw.edu

Funding Information

Funding came from the Encyclopedia of Life Rubenstein Fellows program.

Received: 14 January 2014; Revised: 18 February 2014; Accepted: 23 April 2014

Ecology and Evolution 2014; 4: 2583

doi: 10.1002/ece3.1109

Availability: *Reol* is a freely-available package with a GPL license. A script can be downloaded through the Comprehensive R Archive Network (<http://cran.r-project.org/web/packages/Reol/index.html>) and the working paper can be found on R-Forge at <https://r-forge.r-project.org/projects/reol/>.

Introduction

The Encyclopedia of Life (EOL) is a resource developed in 2007 to host taxonomic information (for example, taxonomic rankings). It brings together taxonomic information from museum scientists, about all taxonomic groups (for example, scientists, educators, students, and the public). Page content comes in forms such as media, common names, and taxonomic descriptions that have agreed to share their information with EOL. EOL content partners include the U.S. Department of the Interior, the U.S. Department of Education, and the U.S. Department of Justice.

© 2014 The Authors. *Ecology and Evolution* is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original author and source are credited.

Abstract

The Encyclopedia of Life is a website that hosts information about life on Earth. Its mission is to increase awareness and understanding of living nature through a freely accessible digital source. Information is publicly available through graphical webpages (browser interface) or through an application programming interface (API). We developed *Reol*, an open-source package for the R environment, which downloads data from the EOL API, searches for and extracts specific information, and builds tables with quantitative data and/or hierarchical classifications. We provide a detailed description how *Reol* can be used as a bridge between the R environment and the EOL API to extract quantitative or hierarchical content. It will be particularly useful for researchers who want information about taxonomic groups of interest (for example: how much

LETTER

Morphogenera, monophyly, and macroevolution

Jablonski and Finarelli (1) suggest that morphogenera, even when they are nonmonophyletic, serve as good representatives for large-scale evolutionary studies. We feel there are two issues that warrant further discussion. First, the test used to evaluate the effect of using nonmonophyletic groups for macroevolutionary studies was not conservative and thus does not provide strong evidence about the impact of nonmonophyly on evolutionary studies. Their test examined whether the median trait value for the species in a nonmonophyletic genus correlated with the median trait value for those same species plus the additional species needed to make the set monophyletic. Correlation of the median value of a set of measurements with the median

value of a superset of those measurements is to be expected (as is briefly mentioned in the methods section of ref. 1). To demonstrate this, we correlated the median values of *ln* body mass from random sets of 3, 7, and 15 mammal species and 0-30 additional species for 45 simulated genera, using data from ref. 2 (Fig. 1A). Even when the median of 3 species is correlated with those 3 plus 15 additional species, the correlation is significant. Moreover, slightly better than random assignment of species to genera increases the expected correlation. To show this, we compare the results of figure 3A from ref. 1 with simulations where 45 genera of 3 species each were created from species chosen randomly from (i) the same family (Fig. 1B) and (ii) the same order (Fig. 1C) (data from refs. 2 and 3). Correlations between morphogenus median and the smallest clade containing the morphogenera species were significant ($P < 0.0001$). Thus, the observed correlations presented in ref. 1 are not surprising. Second, ref. 1 also appears agnostic as to whether the morphogenera should be used together with a phylogeny when

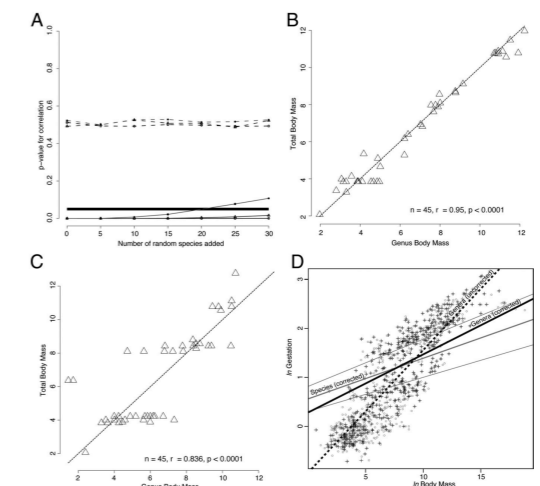


Fig. 1. Correlation analyses of morphogenera and comparative analysis of body mass and gestation. (A) Significance (Spearman correlation, as in ref. 1) of correlation of means of *ln* body mass of a set of *X* species with a set of *X* + *Y* species (solid lines), where *X* is 3 (closed circles), 7 (triangles), or 15 (open circles) and *Y* is 0, 5, 10, 15, 20, 25, 30, where the larger set includes all the species in the smaller set. Dotted lines correspond to correlating a set *Z* + *Y* species, where the size of *Z* and *X* are the same, but the species differ. The dark solid line represents $P = 0.05$. For each point, 2,000 replicates, each with 45 randomly generated sets, were used. The masses came from ref. 2, pruned down to the set of species also in the tree of ref. 3. (B and C) Plots of *ln* body mass of 45 simulated morphogenera, as in figure 3A of ref. 1, where genera consist of 3 species chosen at random from within mammal families (B) or orders (C) (species chosen without replacement within genera but with replacement between genera, and genera simulated until 45 nonmonophyletic ones were recovered). The *x* axis shows the median value for the morphogenera, the *y* axis shows the median for the smallest clade containing all of the species (based on the tree of ref. 3). (D) Generalized least-squares analyses of *ln* gestation period against *ln* body mass for all mammal species and genera. The thick black line shows the fit of these variables for all genera. The grey line shows the fit for all mammal species, and the light grey lines show the 95% confidence interval. The thick dashed line shows the fit for the genera without correcting for the phylogeny. Grey dots represent data using species-level values and black crosses represent generic-level data.