

Computational and Theoretical Phylogenetics

Miller Fellow Focus: Sebastian Höhna

Phylogenies depict the evolutionary relationships among species. Although the field has roots that go back to Darwin's Origin of Species, the field largely languished for a century or more. Today, however, the field is thriving; there is probably more interest in estimating phylogenies today than at any time in the past. Much of the current interest in phylogenetics comes from advances in DNA sequencing technologies which now allow the phylogenetic comparison of hundreds or thousands of genes. The field is remarkable in that it combines areas of biology, statistics, mathematics, and computer science. Here I discuss connections between these areas and how my research program contributes to this field.

Phylogenetics

Although evolution by natural selection is widely accepted as fact, the detailed evolutionary history of specific organisms remains largely unknown. Phylogenetic research seeks to discover these evolutionary histories by allowing us to address biological questions, such as: Is species A more closely related to species B or to species C? When did species A diverge from species B? Did group A diversify more rapidly/slowly than other groups? Did group A experience a mass-extinction event? Which genes are evolving more rapidly (i.e., are under positive selection) or are more highly conserved (i.e., evolve under negative selection) than the back-



ground rate of evolution? Which genes are (in-)congruent with the evolution of the species tree due to gene transfers, hybridization or other mechanisms? Is a given phenotype correlated with mutation (i.e., substitution), speciation or other rates?

Most answers to these questions can be obtained from an estimated phylogenetic tree and mapped traits or rates along its branches (see Figure 1). A major challenge in phylogenetics is to reconstruct the events that occurred million of years ago while only having data from present day species. Complicating matters, evolution only happened once which means we cannot use replicated experiments to learn about the evolutionary tree of life. Nevertheless, new mathematical and statistical methods push the boundaries of what we can learn.

[CONTINUED ON PAGE 2 >](#)

Inside this edition:

Miller Fellow Focus	1-3
Recapping our 60th Anniversary	1, 5
Gifts to the Miller Institute	4
Professorship Awards	5
60th Anniversary in Photographs	6, 7
In the News, Next Steps & Birth Announcements	8

Recapping our - 60th Anniversary -

The Miller Institute family enjoyed a 60th Anniversary Celebration from January 15th to the 17th starting with a reception Friday evening, a day of talks spanning the sciences on Saturday, and local field trips on Sunday. As measured in scientific progress, the last 60 years has witnessed more stunning scientific progress than any comparable interval in history.

When the Miller Institute began, Sputnik had not launched, the Hale telescope at Palomar

[MORE ON PAGE 5 >](#)

"The most exciting conferences have experts in one science, of all ages and stages, presenting, without jargon, to experts in other sciences - and then talking for hours over meals. Imagine that you could get to attend such a conference each week, even if just for one hour. That's the Miller Institute."

- Saul Perlmutter, 2009 Miller Senior Fellow, Professor of Physics, University of California Berkeley; Nobel Prize in Physics 2011, Breakthrough Prize in Fundamental Physics 2015.



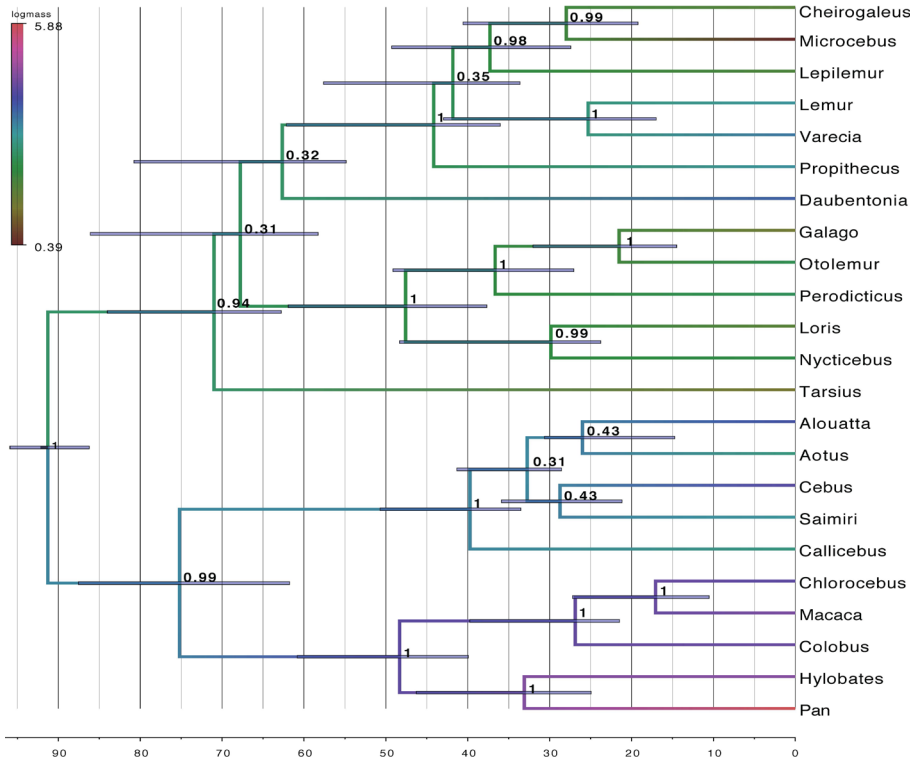


Figure 1: A phylogeny of primates estimated from a cartoon analysis. The phylogeny shows the relationship and timing of the evolution with the posterior probabilities of the clades at the nodes. Bars at the nodes represent uncertainties in the estimated dates and the color of the branches corresponds to the estimated body mass.

Mathematical Models

Address the research questions outlined above by casting the relevant hypotheses as mathematical models. These mathematical models define the types of possible events that can occur and associate them with probabilities. For example, the diversification of species can be modeled as a stochastic birth-death process where each species has some probability of giving birth to a new species and some probability of going extinct. During my Ph.D. studies, I extended standard birth-death models to include diversification rate variation over time, mass-extinction events and/or diversification rate variation among species. In a recent study, I used this new approach to show that the diversity of conifers was impacted by one major mass-extinction event approximately 23 million years ago. This coincides with the known timing of the increase of more arid, grassland ecosystems.

Similarly, continuous time Markov models are used to describe the evolution of molecular sequences. Thus, the probability of the observed molecular characters (i.e., DNA or amino acid states) can be computed once the homology of character has been established (i.e., which characters share a common ancestor). Insertion and deletion events required to

align the sequences are commonly treated simply as unobserved states. During my time as a Miller Fellow, I initiated work on mathematical models that describe the joint processes of insertion, deletion, and substitution. With such a model, we can harness the information of rare events such as insertions or deletions and circumvent the need to establish homology between characters a priori. Interestingly, because insertions and deletions of molecular characters can be considered as birth and death events, I use approaches similar to those I used in my Ph.D. work on species diversification. The specific models allow for insertion and deletion-rate heterogeneity with the aim to identify genes or branches in the phylogeny and link these to adaptation.

In my research, all aspects of evolution are represented by stochastic processes. These processes span a broad range (e.g., branching processes, discrete state Markov processes, Brownian motion and diffusion models, and general processes defined by differential equations), but ultimately

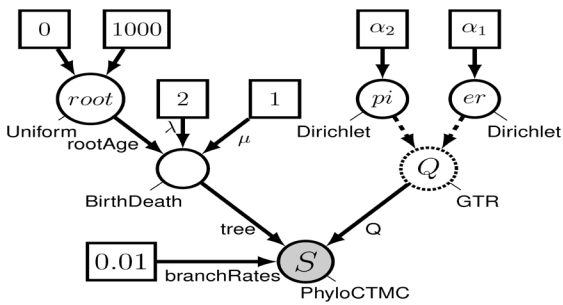
constitute a mathematical model quantifying the probability of the observed data. The mathematically explicit models allow robust inference of parameters while clearly stating the assumptions of the models.

Statistical Inference

In my work I use Bayesian statistics which provides a natural framework to account for nuisance parameters and to quantify uncertainty in parameter estimates by means of posterior probabilities. Specifically, I develop Markov chain Monte Carlo (MCMC) algorithms that simulate samples from the posterior distribution. Efficient MCMC algorithms are a major challenge especially for exploring the vast space of phylogenetic trees. I tackled this problem during my Masters studies, although the research is ongoing.

During my Ph.D. studies, we introduced the concept of probabilistic graphical models to phylogenetics. A probabilistic graphical model consists of vertices (the parameters/variables of the model) and edges (the dependencies between parameters; see Figure 2). Vertices (parameters) are associated with conditional probability distributions or parameter transformations, thus, a larger model is decomposed into smaller, modular pieces. This model representation has the advantage of being easily extendable to more complex (i.e., accurate) models. Furthermore, teaching statistics to empirical biologists is simplified by probabilistic graphical models because model assumptions





```

root ~ dnUniform(0,1000.0)
psi ~ dnBirthDeath(lambda=2.0, mu=1.0,
                  rootAge=root, taxa)

alpha1 <- v(1,1,1,1,1,1)
alpha2 <- v(1,1,1,1)

er ~ dnDirichlet(alpha1)
pi ~ dnDirichlet(alpha2)
Q_mol := fnGTR(er, pi)

seq ~ dnPhyloCTMC(tree=psi, Q=Q_mol, branchRates=0.01, type="DNA")
seq.clamp(data)

```

Figure 2: A simple phylogenetic model depicted in graphical-model notation (left) and the corresponding specification in the Rev language (right).

are made explicit and parameter dependencies are readily visualized (see Figure 2).

Computer Software

In a collaborative project, I combined and implemented the mathematical and statistical theory, described above, in a new computer program called RevBayes. RevBayes is an open source project (currently more than 200,000 lines of code) with many contributors (of which I am currently the lead developer). The program is based on probabilistic graphical models and provides its own interactive model specification language, which is similar to the popular statistical programming languages of R and Bugs. The strength of RevBayes is generality and extensibility which allows users and developers to easily apply many existing models and to implement new models. The similarity between many models (e.g., continuous trait, discrete trait and molecular sequence models) supports the re-usability of code, which, in turn, simplifies development and increases the robustness of the software. Other important features of RevBayes are the intrinsic ability to jointly estimate, for example, the phylogeny, substitution rates, diversification rate and ancestral traits using hierarchical models. This approach naturally accommodates uncertainties in the parameters of the model, e.g., the phylogeny, instead of assuming that the estimated phylogenetic tree is known. Although the initial development was more than challenging and included several complete re-writes during the first years, the effort is starting to pay-off by having this incredibly powerful framework available for new, exciting research questions.

Outlook

The amount of data available now in the era of genomics is both an opportunity and a curse. Cheap sequencing of whole genomes allows us to fill gaps in previously neglected study groups and provides more data to increase the power of statistical inferences. However, genomic data also present many new challenges. First, the increased amount of data demands

faster and better computer software. Due to the exorbitant amount of data most studies have resorted to simpler and approximate methods to provide fast results. The amount of data enables these studies, often based on simple summary statistics, to yield new insights but full and exact models can shed more light to disentangle the plethora of intertwined processes.

Second, and more importantly, it is well established that different regions of the genome evolve differently and thus the genetic process of evolution is rather heterogeneous than homogeneous. Assumptions that were reasonable for carefully chosen genes and small study groups are violated for whole genomes and hence bias our estimates. For example, for studies with single or few genes we could neglect recombination, gene duplication and transfer but this is not true for genome studies. Similarly, the heterogeneity of the evolutionary process among branches, e.g., the average percentage of GC nucleotides, is more prevalent in large phylogenies (e.g., including all 10,000 birds) and hence needs to be modeled adequately.

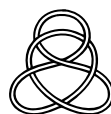
Currently, genomic data pose a big challenge for robust phylogenetic inference. Over the next few years it will yield many new insights into old and new questions in evolution and genetics. It is therefore a wonderful time to be a computer scientist, mathematician and statistician working in this thriving field on these exciting questions.

Sebastian Höhna (Miller Fellow 2014–2017) completed his undergraduate studies (German Diplom Ingenieur) in computer science at the Berufsakademie Berlin (now Berlin School of Economics) while also working as a software engineer for Schering AG. (The Berufsakademie provides a dual studies where one part consists of classes and the other of practical experience.) He then completed a M.Sc. in Computer Science at the University of Auckland and a Ph.D. in Mathematical Statistics at Stockholm University. His interest in phylogenetics was sparked by Alexei Drummond and deepened by John Huelsenbeck, who acted as a mentor during several stages of his career. Now in California, Sebastian owns his first car but stays true to his roots and uses it for camping, even when there is snow in Yosemite.

Gifts to the Miller Institute

The Miller Institute gratefully acknowledges the following contributors to the Miller Institute programs in 2015. These generous donations help support both the Miller Research Fellowship program and the general programs of the Institute. (* = 5 years of giving)

Make A Gift



Donations can be made by going to:
Give to CAL for the Miller Institute
miller.berkeley.edu/gift

Patrons of the Miller Knot (\$5000 - \$9999)

Nancy Steinhaus
(in memory of Edward A. Steinhaus)

Leadership Circle (\$2500 - \$4999)

*William Craig
Barbara Meyer (The Byron R. Meyer Living Trust)

Director's Circle (\$1000 - \$2499)

Mark Laidre
*Michael Manga & Susan Storch
*Jasper Rine
*Randy Schekman & Nancy Walls
Fausta Segre Walsby
Christine Shannon
Jeremy Thorner
*Peter & Gloria Yu

Miller Partners (\$500 - \$999)

*Rachel Akeson
Lutz Duembgen
*Craig Evans
*Thomas Juenger
Kam-Biu Luk
Dustin Rubenstein
Raman Sanyal & Vanessa Kaeaeab-Sanyal
Rich Saykally & Christine Read
Ryan & Lauren Trainor
Theodore & Janice van Duzer
Taizan Watari

Miller Associates (\$250 - \$499)

W. Craig Carter
Matthew Good
Carole Hickman
*Ron Hoy & Margaret Nelson
**Ronald Johnson
Judith Klinman
Joel Moore
Martin Scharlemann & Barbara Wagner
*Jesse Thaler
Patricia Zambryski

Miller Advocates (up to \$249)

Anna Behrensmeyer
Adrian Bejan
Justin Brown
Itay Budin
Roland Burgmann
Beth Burnside
*William Clemens
Morgan Conn
*Kathryn Day
Christopher Douglas
David & Monika Eisenbud
Sandra Faber
Claude-Andre Faucher-Giguere
Alex Filippenko
Mary K. Gaillard
Britt Glausinger
Steven Glickman
Matthew Good
Alberto & Dolores Grunbaum
Richard Harland
John Harte
John Hunter & Amy Roussel
Raymond Jeanloz
*Takahiro Kawai
Eric King
Mimi Koehl
Stephen Leone
Milo Lin
Michael Manga
Geoff Marcy and Susan Kegley
Xanthippi Markenscoff
Maryam Modjaz
Akino Nose
Rachel Pepper
Jasper Rine
Richard & Jean Roberts
Stephen Suh
John Taylor
Jeremy Thorner
Salil Vadhan
*Dan-Virgil Voiculescu
Ken Wachter
Benjamin Weiss
David Wemmer
John West

For More Information: miller.berkeley.edu

Questions? Kathryn Day: 510-642-4088 | millerinstitute@berkeley.edu



Miller Research Competitions: Professorship Awards

On December 7, 2015, the Advisory Board of the Miller Institute met to select next year's Professorship awards. The Board is comprised of four advisors external to UCB: Roger Blandford (Stanford University), David Botstein (CALICO), Yun Song (University of Pennsylvania) and Steven Block (Stanford University); and three internal Executive Committee members: Executive Director Michael Manga (Earth & Planetary Science), Jasper Rine (Genetics and Developmental Biology) and Stephen Leone (Chemistry/Physics). The Board is chaired by Chancellor Nicholas Dirks.

The Miller Institute is proud to announce the awards for Professorship terms during the Academic Year 2016-2017. These outstanding scientists pursue their research, following promising leads as they develop. The Visiting Miller Professors join faculty hosts on the Berkeley campus for collaborative research interactions.

Mina Aganagic, Miller Professorship, UC Berkeley Departments of Physics/Mathematics.

Roberto Bassi, Visiting Miller Professor in the Dept. of Plant & Microbial Biology. Host: Krishna Niyogi. Home institution: University of Verona.

Inez Fung, Miller Professorship, UC Berkeley Departments of Earth & Planetary Science / ESPM.

Zahid Hasan, Visiting Miller Professor in the Dept. of Physics. Host: Alessandra Lanzara. Home institution: Princeton University.

Chuan He, Visiting Miller Professor in the Dept. of Chemistry. Host: Christopher Chang. Home institution: University of Chicago.

Miaki Ishii, Visiting Miller Professor in the Dept. of Earth & Planetary Science. Host: Richard Allen. Home institution: Harvard University.

Mark Johnson, Visiting Miller Professor in the Dept. of Chemistry. Host: Daniel Neumark. Home institution: Yale University.

Arash Komeili, Miller Professorship, UC Berkeley Department of Plant & Microbial Biology.

Vivek Malhotra, Visiting Miller Professor in the Dept. of Molecular & Cell Biology. Host: Randy Schekman. Home institution: Center for Genomic Regulation, Barcelona, Spain.

Susan Marqusee, Miller Professorship, UC Berkeley Department of Molecular & Cell Biology.

David Nadler, Miller Professorship, UC Berkeley Department of Mathematics.

Harold Poor, Visiting Miller Professor in the Dept. of EECS. Host: Venkat Anantharam. Home institution: Princeton University.

Thomas Richards, Visiting Miller Professor in the Dept. of Molecular & Cell Biology. Host: Nicole King. Home institution: University of Exeter.

Hans-Walter Rix, Visiting Miller Professor in the Dept. of Astronomy. Host: Mariska Kriek. Home institution: Max Planck Institute for Astronomy.

Steven Wofsy, Visiting Miller Professor in the Dept. of Chemistry. Host: Ronald Cohen. Home institution: Harvard University.

Ahmet Yildiz, Miller Professorship, UC Berkeley Dept. of Physics.

Bin Yu, Miller Professorship, UC Berkeley Dept. of Statistics/EECS.

Eric Zaslow, Visiting Miller Professor in the Dept. of Mathematics. Host: Vivek Shende. Home institution: Northwestern University.

60th Anniversary: Looking forward to 2031!

> CONTINUED FROM PAGE 1

was a leading window into space, plate tectonics was decades away, and the genetic code had not been cracked. For the last 60 years, members of the Institute have played leading roles in the progress that has profoundly shaped the world we live in. The meeting provided the 100+ registrants the opportunity to reconnect with the breadth of scientific interests that we all had as students, before the pressures of professional specialization narrowed our focus. The talks spanned multiple scales of size and time, from the origin of marine tetrapods 250 million years ago, to femtosecond spectroscopy; from the cosmos to the gut; and a sobering insights into earthquakes and their survival that left us "shaken, not stirred". The Miller Institute, through the efforts of the Fellows, Miller Professors and Visiting

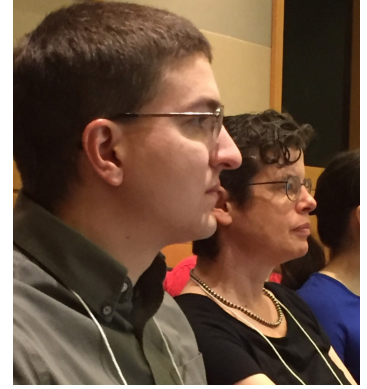
Miller Professors, will continue to provide a unique ecosystem of cross-disciplinary research and communication. While predictions regarding the next 60 years are of dubious reliability, we can be sure that the 75th Anniversary celebration will be at least as exciting as the 60th. So carve out some time in your 2031 calendar, and we look forward to seeing you then.

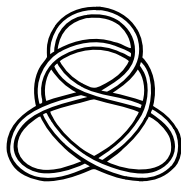


Jasper Rine
Chair of the Organizing Committee

Celebrating 60 years of Science







Non-Profit
Organization
U.S. Postage
PAID
University of
California,
Berkeley

University of California, Berkeley
Miller Institute for Basic Research in Science
2536 Channing Way, #5190
Berkeley, CA 94720-5190
510.642.4088
miller.berkeley.edu



Miller Institute News :: Winter 2016
Please send address corrections to:
miller_adm@berkeley.edu

In the News

(see current Miller Institute News: miller.berkeley.edu/news)

Professor Paul Alivisatos (Miller Professor 2001-2002, Chemistry) has been appointed as Berkeley's next Vice Chancellor, Research, effective March 1, 2016.

Roger Blandford (Visiting Miller Professor 2013, Astronomy) and Roy Kerr have been awarded the 2016 Crafoord Prize in Astronomy "for fundamental work concerning rotating black holes and their astrophysical consequences".

Philip Hopkins (Miller Fellow 2008-2011, Astronomy) won this year's Warner Prize from the American Astronomical Society, which recognises an early-career astronomer for significant contributions to astronomy in the five years preceding the award.

Paul Alivisatos (Miller Professor 2001-2002, Chemistry) and **Simon Levin** (Visiting Miller Professor Spring 2003, Integrative Biology) are among this year's nine recipients of the National Medal of Science, the nation's highest honor for achievement and leadership in advancing the fields of science and technology.

Aziz Sancar (Visiting Miller Professor Spring 2002, Chemistry) shares the 2015 Nobel Prize in Chemistry for mechanistic studies of DNA repair.

Next Steps

The Miller Institute congratulates the Miller Fellows on their next endeavors:

Rebekah Dawson (Assistant Professor @ Penn State University)

Birth Announcements

Ryan Trainor (Miller Fellow 2014-2017) & and his wife Lauren welcomed son, Silas Dominic Francis Trainor born 9/26/15.

Chris Martin (Miller Fellow 2013-2015) & and his wife Erica Morehouse announced the birth of their daughter, Kai Maya Morehouse Martin, born 10/3/15.



The Miller Institute is "dedicated to the encouragement of creative thought and the conduct of research and investigation in the field of pure science and investigation in the field of applied science in so far as such research and investigation are deemed by the Advisory Board to offer a promising approach to fundamental problems."

For More Information:

+ Staff: Kathryn Day, Donata Hubert, Erin Lyman & Emily Birman