

Statistical Phylogenetics: Comparative methods

Instructor Cecile Ane
Office 1208 Medical Science Center or 341 Birge Hall
Phone 262-3901 or 262 6820
email ane@stat.wisc.edu
Times TR 2:30-3:45 in 5295 MSC
Textbook *The comparative method in Evolutionary Biology*, by Harvey and Pagel.
We will complement this book with readings in the scientific literature.
Web <http://www.stat.wisc.edu/courses/st992-ane/>

- **Prerequisites.** This course will be accessible to graduate students both in statistics and in the biological sciences. **Statistics students** should have completed (or be taking concurrently) Statistics 610 or a more advanced course. **Biology students** should have completed (or be taking concurrently) any two courses in statistics or probability (such as Statistics 571/572), preferably at the graduate level.
- **Summary and Goals.** Comparative biologists ask questions about evolutionary processes. Usually, their observational units are species, which typically do not yield random samples. This is because the sampled species share an evolutionary history, with closely related species usually being more alike than distantly related species, and observations lack independence. In this course, we will cover the major advances in comparative methodology for both discrete and continuous data. All these methods use the genealogical history of the sampled species to overcome their non-independence. The most widely used method is based on modeling the evolution of a character with a Brownian motion on a tree. A more recent model uses the Ornstein-Uhlenbeck process to account for biological selection.

In the second part of the course, we will cover methods for inferring phylogenies (i.e tree-like histories of species) from molecular data, including semi-parametric methods for estimating divergence times. As this course complements other recent courses offered on campus on molecular evolution and phylogenetic inference, I will adapt the second part of the course to the background and interests of the students.

My objective for statistics students is to provide them with a sufficient background in statistical phylogenetics and comparative studies so that they can start exploring their own research questions in the area. My objective for biology students is to provide them with a deeper understanding of the statistical methods available in the area, so that they can do the best choices for their own data, and pull new methodological developments towards what their needs are, through collaborations. For all students, my objective is to give them a taste of fruitful cross-disciplinary work.

- **Course Structure.** I intend to alternate between lectures and discussions. The days when there will be discussion, two students will each present a paper and lead a discussion.
- **Credits.** The course can be taken for 1 to 3 credits. All students will complete a project. Students taking the course for 2 credits will also do either homework assignments **or** a paper presentation, as detailed below. Students taking the course for 3 credits will do both homework assignments and a paper presentation. Grades for students taking the course for 1 credit will be their grade from the project. Grades for students taking the course for 2 credits will be the average of grades from the project and from the other area they have chosen. Grades for students taking the course for 3 credits will be the average of grades from the 3 areas.

- **Project:** All students will complete a project, preferably in **cross-disciplinary** groups of two or three students. A project can be to analyze a data set using one or more methods covered in the course, or a project can be more theoretically focused. For instance, a project could address a question through simulations. I will ask for a written proposal to be handed in by **Thursday Nov. 10**, in which you state in some detail what you intend to do. The due date for the final project write-up will be **Thursday Dec. 8**. Then each group of students will make a short (informal) presentation of its work on **Dec. 13** or **Dec. 15**.
- **Paper presentation:** A sign-up sheet will be available on the first day. Student presenting a paper will explain the main ideas and results of the paper for about 15 minutes, then lead a 15-20 minute discussion. Other students will send questions about the paper in advance. These questions can serve as a basis for stirring up the discussion. Links to papers' pdf will be available on the course webpage (go to "schedule").
- **Homework assignments:** written homework will be assigned approximately once every 3 weeks. I will also assign readings, each time a student will present a paper. Students are expected to read the assignment prior to the class discussion time. They will prepare one question (or more) regarding the reading, and send it to me and to the student presenting the paper. The due date for sending the question(s) will be 2 days before the class discussion time (ex: Sunday midnight if the paper is to be discussed on Tuesday, and Tuesday midnight if the paper is to be discussed on Thursday).

• **Tentative Schedule.** Please refer to the web page for accurate information, as I will update the schedule regularly.

Sept. 6 and 8	Introduction and Examples	[CMJ ⁺ 05, BP05]
Sept. 13 and 15	Methods for categorical inter-specific data	Ch.4, [Mad90, Pag94]
Sept. 20	ANOVA methods for quantitative data	Ch.5: §5.4, 5.5 and 5.8.1, [Bel89]
Sept. 22	Felsenstein's Independent Contrasts	Introduction to Brownian motion, [Fel85], interpretation of regression patterns (§5.9), [Mad91]
Sept. 27	Rescaling, assumption checking	[Gra89, GHI92, DUG96],
Sept. 29	Introduction to the Ornstein-Uhlenbeck process	Micro to macro-evolutionary models: [HM96]
Oct. 4	Generalized linear models	
Oct. 6	PGLS and GEE	[MH97, PC02]
Oct. 11	Randomization tests	[GDJJ93]
Oct. 13	Dealing with unresolved phylogenies	[PG93, GDU99, Mar96a, PMB04]
Oct. 18	PMM	[Lyn91, HML04]
Oct. 20	Comparing comparative methods	[Roh01, MDFH02]
Oct. 25	Ancestral state reconstruction	[SPML97, Pag99, GI00]
Oct. 27	Testing rates of evolution	[Gar92], LRT using BM
Nov. 1 and 3	Testing adaptation, and Model selection (AIC, BIC)	[Han97, BK04]
Nov. 8	(Bret Larget)	
Nov. 10	Tree building methods and estimation of divergence times: as time allows	[TK05]

References

- [Bel89] Graham Bell. A comparative method. *The American Naturalist*, 133(4):553–571, 1989.
- [BK04] Marguerite A. Butler and Aaron A. King. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *The American Naturalist*, 164(6):683–695, 2004.
- [BP05] Daniel Barker and Mark Pagel. Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. *PLoS Computational Biology*, 1(1):e3, 2005.
- [CMJ⁺05] Marcel Cardillo, Georgina M. Mace, Kate E. Jones, Jon Bielby, Olaf R. P. Bininda-Emonds, Wes Sechrest, C. David L. Orme, and Andy Purvis. Multiple causes of high extinction risk in large mammal species. *Science*, 309(5738):1239–1241, 2005.
- [DUG96] Ramon Diaz-Uriarte and Theodore Garland, Jr. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: Sensitivity to deviations from brownian motion. *Systematic Biology*, 45(1):27–47, 1996.
- [Fel85] Joseph Felsenstein. Phylogenies and the comparative method. *The American Naturalist*, 125(1):1–15, 1985.
- [Gar91] Emilia P. Martins; Theodore Garland, Jr. Phylogenetic analyses of the correlated evolution of continuous characters: A simulation study. *Evolution*, pages 534–557, 1991.
- [Gar92] Theodore Garland, Jr. Rate tests for phenotypic evolution using phylogenetically independent contrasts. *he American Naturalist*, 140(3):509–519, 1992.
- [GDJJ93] Theodore Garland, Jr., Allan W. Dickerman, Christine M. Janis, and Jason A. Jones. Phylogenetic analysis of covariance by computer simulation. *Systematic Biology*, 42(3):265–292, 1993.
- [GDU99] Theodore Garland, Jr. and Ramon Diaz-Uriarte. Polytomies and phylogenetically independent contrasts: Examination of the bounded degrees of freedom approach. *Systematic Biology*, 48(3):547–558, 1999.
- [GHI92] Theodore Garland, Jr., Paul H. Harvey, and Anthony R. Ives. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology*, 41(1):18–32, 1992.
- [GI00] Theodore Garland, Jr. and Anthony R. Ives. Using the past to predict the present: Confidence intervals for regression equations in phylogenetic comparative methods. *The American Naturalist*, 155(3):346–364, 2000.
- [Gra89] A. Grafen. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 326(1233):119–157, 1989.
- [Han97] Thomas F. Hansen. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5):1341–1351, 1997.
- [HB87] Raymond B. Huey and Albert F. Bennett. Phylogenetic studies of coadaptation: Preferred temperatures versus optimal performance temperatures of lizards. *Evolution*, 41(5):1098–1115, 1987.
- [HM96] Thomas F. Hansen and Emilia P. Martins. Translating between microevolutionary process and macroevolutionary patterns: The correlation structure of interspecific data. *Evolution*, 50(4):1404–1417, 1996.
- [HML04] Elizabeth A. Housworth, Emlia P. Martins, and Michael Lynch. The phylogenetic mixed model. *The American Naturalist*, 163:84–96, 2004.
- [Los94] Jonathan B. Losos. Integrative approaches to evolutionary ecology: Anolis lizards as model systems. *Annual Review of Ecology and Systematics*, 25:467–493, 1994.

- [Los00] Marguerite A. Butler; Thomas W. Schoener; Jonathan B. Losos. The relationship between sexual size dimorphism and habitat use in greater antillean anolis lizards. *Evolution*, 54(1):259–272, 2000.
- [Lyn91] Michael Lynch. Methods for the analysis of comparative data in evolutionary biology. *Evolution*, 45(5):1065–1080, 1991.
- [Mad90] Wayne P. Maddison. A method for testing the correlated evolution of two binary characters: Are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution*, 44(3):539–557, 1990.
- [Mad91] Wayne P. Maddison. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Systematic Zoology*, 40(3):304–314, 1991.
- [Mar96a] Emilia P. Martins. Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution*, 50(1):12–22, 1996.
- [Mar96b] Emilia P. Martins. Phylogenies, spatial autoregression, and the comparative method: A computer simulation test. *Evolution*, 50(5):1750–1765, 1996.
- [MDFH02] Emilia P. Martins, Jos Alexandre Diniz-Filho, and Elizabeth A. Housworth. Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution*, 56(1):1–13, 2002.
- [MH97] Emilia P. Martins and Thomas F. Hansen. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*, 149(4):646–667, 1997.
- [Pag94] Mark Pagel. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 255(1342):37–45, 1994.
- [Pag99] Mark Pagel. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology*, 48(3):612–622, 1999.
- [PC02] Emmanuel Paradis and Julien Claude. Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology*, 218(2):175–185, 2002.
- [PG93] Andy Purvis and Theodore Garland, Jr. Polytomies in comparative analyses of continuous characters. *Systematic Biology*, 42(4):569–575, 1993.
- [PH89] Mark D. Pagel and Paul H. Harvey. Taxonomic differences in the scaling of brain on body weight among mammals. *Science*, 244(4912):1589–1593, 1989.
- [PMB04] Mark Pagel, Andrew Meade, and Daniel Barker. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, 53(5):673–684, 2004.
- [Roh01] F. James Rohlf. Comparative methods for the analysis of continuous variables: Geometric interpretations. *Evolution*, 55(11):2143–2160, 2001.
- [SPML97] Dolph Schluter, Trevor Price, Arne O. Mooers, and Donald Ludwig. Likelihood of ancestor states in adaptive radiation. *Evolution*, 51(6):1699–1711, 1997.
- [TK05] Jeffrey L. Thorne and Hirohisa Kishino. Estimation of divergence times from molecular sequence data. In *Statistical Methods in Molecular Evolution*, chapter 8, pages 233–256. Springer, 2005.