

Research Statement

When I was first considering biology as a career, I had a new study system every day. I wanted to study convergent phenotypes in reptiles, I wanted to examine the comparative morphology of walking fish, I considered using modern birds to study dinosaur locomotion. I realized that what brought together all of these disparate interests was evolution. And fundamental to any comparative study of organisms in an evolutionary framework is phylogeny. I ultimately decided to study phylogenetic methods as a way to engage with disparate organisms, and to research a variety of questions. Below is a description of my current research interests.

Divergence Dating

In 2016, I was awarded an NSF postdoctoral fellowship to work on developing and evaluating new methods for incorporation of fossil data in phylogenetic estimation and divergence time estimation in collaboration with Dr. Tracy Heath at Iowa State University and Dr. Corrie Moreau at the Field Museum. Divergence time estimation allows researchers to date how long it has been since two lineages last shared a common ancestor, providing a time scale for studying the evolution of organisms and traits. This work has both theoretical and empirical components.

Theoretical Issues in Dating Phylogenetic Trees

I am examining model misspecification in divergence dating analyses. New methods for divergence dating treat fossils and genetic data as part of the same process of speciation and extinction, using one cohesive model to estimate divergences for combined morphological and genetic datasets. Several papers have suggested that this new framework results in divergences that are unrealistically ancient, but this may be an artifact of incorrectly specified models. I am using simulations of combined genetic and morphological datasets to assess the impact of incorrectly parameterizing the models used in these analysis. My previous work related to this question has been largely empirical, and this project reflects an expansion of my skills to look at the mathematics of divergence time estimation, in addition to the biology.

Towards a Better Tree of Ants

Ants are a fascinating group for many reasons: they are highly diverse, they are involved in many mutualisms globally, and they have a beautiful fossil record. As part of my fellowship, I will be using molecular and morphological data to estimate a new tree of ants.

Our view of evolutionary history is shaped by the way our data are sampled. Ants form a particularly interesting use case of this: some fossil ant groups are very well-sampled, others are not, and how well-sampled a group is is often a function of their biology and life history. For example, specimens preserved in amber are often very complete and information-rich, while those preserved as imprints are not. This project will produce a new time-scaled tree of ants, but also illuminate many theoretical and mathematical issues relating to the incorporation of fossil information with genetic data.

Future Directions

This is a very ambitious project, and the dataset and software framework I develop during this fellowship will form the basis of my future work. When I have answered some of the questions about how biased sampling of fossils affects our ability to accurately assess phylogenetic relationships and divergence dates, I will move on to making improvements to our methods in order to better incorporate sampling biases. Using the collection of ant data I am currently building and the groundwork of software I am currently laying, I will be able to develop and test new models that more completely incorporate the richness of the fossil record.

Relevant Papers

- Bapst DW, **Wright AM**, Matzke NJ, Lloyd GT. 2016. Tip-dating of fossil theropods analysis recovers Archaeopteryx as a bird, but reveals model-dependence of dates and uncertainty. *Biol. Lett.* 12: 7.
- Matzke NJ, **Wright AM**. 2016. Inferring node dates from tip dates in fossil Canidae: the importance of tree priors. *Biol. Lett.* 12: 8.

Phylogenetic Methods

One of my primary interests is improving methodology for estimating phylogenetic trees. Trees are indispensable for studying evolution, and much of the literature surrounding phylogenetic estimation has focused on using small amounts of genetic information to build trees. We are, however, in an era where more biologists want to make more complete use of fossil record, or of the bountiful data provided by high-throughput sequencing to estimate phylogenies. One focus of my work is on incorporating these datatypes in phylogenetic analysis.

Morphology and Phylogeny:

My doctoral dissertation focused on statistical phylogenetics, both in terms of theory and the use of phylogenetic trees to answer biological questions. My dissertation was supported by a National Science Foundation Doctoral Dissertation Improvement Grant and advanced the methods for the use of morphological data to estimate Bayesian phylogenetic trees from both the fossil record and neontological data. In this research, I evaluated the efficacy of Bayesian statistical approaches for building phylogenetic trees from fossil data, and examined cases of model violation. In the course of conducting my research, I also developed an extensions to the existing open-source software, increasing the number of statistical approaches that can make use of morphological data.

Phylogenomics:

In collaboration with my PhD adviser, Dr. David Hillis, I have completed a phylogenomic study of Eurycea salamanders. These salamanders are endemic to the Edwards Plateau in Texas, and are threatened by human development. We used genome-scale data to understand their population history and assess how many populations are present for conservation purposes. In collaboration with Dr. Mark Holder at University of Kansas, I am testing new corrections for ascertainment bias in phylogenetic data. Ascertainment bias occurs when only certain types of characters (most commonly those that vary within a group of organisms) are collected, and others are not. This bias is prevalent in both paleontological and genomic data, in which whole genomes are often distilled down to their variable sites to use in phylogenetic analysis, or in other analyses of evolution. Ascertainment bias can cause incorrect trees to be favored in phylogenetic analysis. We are implementing statistical corrections for this bias in a new software framework for estimation of phylogenetic trees called RevBayes.

Future Directions

Building trees continues to be a prime interest of mine. Particularly, I am interested in combining genomic-level data with fossils and morphology to estimate phylogenetic trees that reflect both extant and extinct diversity. One aspect of doing this work that is of particular concern is better understanding how to combine phylogenomic data, which is often abundant and rich in phylogenetic signal, with morphological data, which is often sparse. Building on my work in ants and salamanders as well as my work with pure morphology (Wright and Hillis, 2014; Wright, Lloyd and Hillis, 2016), I am especially interested in how using more complex models of morphological evolution can help amplify the signal present in the morphology portion of combined morphology-molecular datasets.

Relevant Papers

- Lanfear R., **Wright AM**, Fransden PB, Senfeldova T and Calcott B. Accepted, Mol. Biol. and Evol. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. and Evol.
- **Wright AM**, Lloyd GT and Hillis DM. 2016. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Syst. Biol. 65: 4.
- **Wright AM**, Lyons KM, Brandley M, Hillis DM. 2015. Which Came First? Robustness in Phylogenetic Reconstruction of Ancestral States. J Exp. Biol. 324: 504-516.
- **Wright AM** and Hillis DM. 2014. Bayesian Analysis Using a Simple Likelihood Model Outperforms Parsimony for Estimation of Phylogeny from Discrete Morphological Data. PLoS ONE 9:e109210.
- Hillis DM, Devitt TJ, **Wright AM** and Cannatella DC. City of Austin Report. Genomic Assessment of Taxonomic Status of Central Texas Eurycea Salamanders

Student Involvement

As part of my research, I have mentored several students. One in particular sticks out in my mind - a student in the Heath lab, Kelly Schiro. Together, we performed divergence dating in bears. During this project, we peer programmed in iPython Notebooks, a programming environment that allows a user to integrate Python code, and verbal and visual explanations of the code in one document. Through these documents, and in-person discussion of them, Kelly became a very proficient programmer. She has since gone on to a bioinformatics internship and a full-time research job in a USDA lab using the same methods she learned working with me to track pathogen evolution.

There are many unanswered questions surrounding phylogenetic estimation, particularly involving fossils and genomes. I will be continuing to improve phylogenetic methods for years to come. This presents exciting opportunities for undergraduate and master's researchers, both in biology and in other disciplines. The two biggest sources of student research with me would be using empirical and simulated datasets to conduct phylogenetic analyses and evaluate the results for biological and geological reasonability, and extending existing software to incorporate new models and data types. My ideal student would be a biology student who wants to learn important computational approaches (simulation, data analysis and visualization, accessing biological databases and processing data, automation of tasks) in a hands-on environment. Educational research supports the importance of pair programming (working with a partner to develop code) and doing frequent code review (checking of code for technical correctness) for novice computational scientists. Therefore, a pair of students who could work together, and meet bi-weekly with me, would be ideal. My

work is mostly computational, and has the potential to result in both theoretical and empirical papers. I also support the communication of research results in alternative formats, such as blogging and conference presentations, and would encourage my students to do this as well.

Other opportunities for student involvement are more densely computational in nature. Biologists with more advanced computer skills, or computer science students who want to learn about biology, could work with me to write new software functions and tools. Educational research also supports peer mentoring as an excellent path to skill mastery, as well as an efficient strategy for retaining women and underrepresented minorities. Students with more advanced computational skills would be encouraged to help others with code review. Students with an interest in statistical modeling of biological systems could also find a place in this research program, as they could get hands-on experience in statistics through the lens of phylogenetics.

July 31, 2018