November 1, 2014
Search Committee Chair
Department of Biology
Temple University

Dear Search Committee Chair,

I am writing to apply for the position of Assistant Professor in Genomics at Temple University. I am a computational biologist with a track record of combing theoretical and empirical analysis to understand the impact of evolutionary forces on variation from the level of DNA to the level of whole organisms. I studied with Montgomery Slatkin at the University of California, Berkeley while pursuing my Ph.D. and I am currently funded by the NSF Postdoctoral Fellowship in Biology to work with Joshua Akey and Jonathan Wakefield at the University of Washington.

During my graduate work, I developed novel theory and methods for analyzing both micro- and macro-evolutionary datasets. I then applied these methods to empirical datasets to gain insights about the evolutionary history of organisms such as yeasts, pigs and humans. I am continuing my work on population genetics and phenotypic evolution during my postdoctoral fellowship. In addition, I am conducting in functional genomics and the genetic basis of complex traits.

Temple has proven itself a remarkable force in computational genomics. By recently hiring several top scientists from across the country, iGEM@Temple is proving that it will be a fantastic place to pursue biological research. My interest in phylogenomics, particularly inferences of phylogenetic relationships in the context of gene flow, complements an already strong core of phylogenomics groups at Temple (e.g. Kumar, Hedges, and Liberles groups). In addition, my background in theoretical population genetics will strengthen the core of the Hey group by providing new insights into how genes evolve within populations. Finally, I will bring extensive experience in functional genomics and molecular phenotypes that will ensure that iGEM@Temple is ready for the next generation of high throughput functional biology.

Please find attached my curriculum vitae as well as my research and teaching statements. I have asked for letters of recommendation to be sent by Montgomery Slatkin, Joshua Akey, Steven Evans, and Michael Turelli. Thank you for your time and consideration of my application. I look forward to hearing back from you.

Sincerely,

Joshua G. Schraiber
Foege Building S-303
Seattle, WA 98195
Phone: (206) 897-1846
e-mail: schraib@uw.edu
http://scholar.google.com/citations?user=lCFcY7AAAAAJ&hl=en&oi=ao

Joshua G. Schraiber *Curriculum Vitae*

Department of Genome Sciences
S-303 South Foege Building
3720 15th Ave. NE
Seattle, WA 98195-5065
schraib@uw.edu

CURRENT POSITION

NSF Postdoctoral Fellow July 2014-present
University of Washington
Advisers: Joshua Akey and Jonathan Wakefield

EDUCATION

Ph.D. Integrative Biology Filed July 2014
Designated emphasis in Computational Biology
University of California, Berkeley
Adviser: Montgomery Slatkin

B.S. Genetics, with highest honors 2009
Minor in Mathematics
University of California, Davis

AWARDS

1. **NSF Postdoctoral Fellow** 2014-2016
   NSF Postdoctoral Fellowship in Biology
2. **Trainee** 2011-2013
   National Institutes of Health Training Grant
3. **Graduate Student Travel Award** 2012
   Society for Molecular Biology and Evolution Annual Meeting
4. **Best Graduate Student Talk** 2011
   Computational Biology Retreat, University of California, Berkeley
5. **Graduate Student Travel Award** 2011
   Society for Molecular Biology and Evolution Annual Meeting
6. **Honorable mention** 2011
   National Science Foundation Graduate Research Fellowship
7. **Honorable mention** 2010
   National Science Foundation Graduate Research Fellowship
8. **Chancellor's Fellowship** 2009-2011
   University of California, Berkeley
9. **Departmental Citation** 2009
   Molecular and Cellular Biology, University of California, Davis
10. **Member** 2009
    Phi Beta Kappa

PEER-REVIEWED PUBLICATIONS

1. Racimo, F., **J.G. Schraiber**. Approximation to the distribution of fitness effects across functional categories in human segregating polymorphisms. PLoS Genetics, *in press*.
2. Lazaridis, I., …, **J.G. Schraiber**, …, *et al*. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409-413.
3. **Schraiber, J.G.** A path integral formulation of the Wright-Fisher process with genic selection. Theoretical Population Biology 92:30-35.
4. Frantz, L.A.F., **J.G. Schraiber**, O. Madsen, H.-J. Megens, M. Bosse, Y. Paudel, G. Semiadi, E. Meijaard, N. Li, R.P.M.A. Crooijmans, A.L. Archibald, M. Slatkin, L.B. Schook, G. Larson and M.A.M. Groenen. 2013. Genome sequencing reveals fine-scale diversification and reticulation history during speciation. Genome Biology 14:R107.
5. **Schraiber J.G.**, Y. Mostovoy, T.Y. Hsu, and R.B. Brem. 2013. Inferring evolutionary histories of pathway regulation from transcriptional profiling data. PLoS Computational Biology 9:e1003255.
6. **Schraiber, J.G.**, R.C. Griffiths, and S.N. Evans. 2013. Analysis and rejection sampling of Wright-Fisher diffusion bridges. Theoretical Population Biology 89:64-74.
7. **Schraiber**, **J.G**, S. Shih, and M. Slatkin. 2012. Genomic tests of variation in inbreeding among individuals and among chromosomes. Genetics 192:1477-1482.
8. Landis, M.J.*, **J.G. Schraiber***, and M. Liang. 2013. Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. Systematic Biology 62:193-204.
9. Groenen, M.A.M., …, **J.G. Schraiber**, …, *et al.* 2012. Analyses of pig genomes provide insights into porcine demography and evolution. Nature 491:393-398.
10. Meyer, M.*, M. Kircher*, …, **J.G. Schraiber**, …, *et al.* 2012. A high-coverage genome sequence from an archaic Denisovan individual. Science 338:222-226.
11. Martin, H.C.*, J.I. Roop,* **J.G. Schraiber***, T.Y. Hsu, and R.B. Brem. 2012. Evolution of a membrane protein regulon in *Saccharomyces*. Molecular Biology and Evolution 29:1747-1756.
12. **Schraiber, J.G.**, A.N. Kaczmarczyk, R. Kwok, M. Park, R. Silverstein, F.U. Rutaganira, T. Aggarwal, M.A. Schwemmer, C.L. Hom, R.K. Grosberg, and S.J. Schreiber. 2011. Constraints on the use of lifespan-shortening *Wolbachia* to control dengue fever. Journal of Theoretical Biology 29:26-32.
13. Riely, B.K., H. He, M. Venkateshwaran, B. Sarma, **J.G. Schraiber**, J.M. Ané, and D.R. Cook. 2011. Identification of legume RopGEF gene families and characterization of a *Medicago truncatula* RopGEF mediating polar growth of root hairs. The Plant Journal 65:230-243.

PUBLICATIONS IN REVIEW

1. **Schraiber J.G.**, M.J. Landis. Sensitivity of quantitative traits to mutational effects, number of loci and population history. Genetics, *in revision*. bioRxiv: 008540
2. Frantz, L.A.F., **J.G. Schraiber**, O. Madsen, H.-J. Megens, A. Cagan, M. Bosse, Y. Paudel, R.P.M.A. Crooijmans, G. Larson, M.A.M Groenen. Analysis of Eurasian wild and domestic pig genomes reveal long-term gene flow during domestication. PNAS, *submitted*. bioRxiv: 010959

PUBLICATIONS IN PREPARATION

1. Matzke, N.J., **J.G. Schraiber**, M. Collard, M. Dembo, M.A. Yang. Tighter estimation of hominoid divergence times using Bayesian joint analysis of dated fossil morphology and incompletely sorted genes. *In prep.*
2. **Schraiber J.G.**, S.N. Evans, M. Slatkin. Bayesian inference of natural selection from allele frequency time series. *In prep*.

** These authors contributed equally*

PRESENTATIONS

1. **Schraiber, J.G.** Detecting non-neutral evolution in transcriptome-wide gene expression measurements. Society for Molecular Biology and Evolution meeting 2011, Kyoto, Japan.
2. **Schraiber, J.G.**, E.Y. Durand and M. Slatkin. Sequentially Markov admixture inference. UC Berkeley Computational Biology retreat 2011, Tomales Bay, California.
3. **Schraiber, J.G.**, E.Y. Durand and M. Slatkin. Sequentially Markov admixture inference. Society for Molecular Biology and Evolution meeting 2012, Dublin Ireland.
4. **Schraiber J.G**, R.C. Griffiths and S.N. Evans**.** Theory and applications of Wright-Fisher diffusion bridges. Bay Area Population Genomics meeting, February 2013, Stanford Univeristy.
5. **Schraiber J.G.,** Matzke N.J., M. Collard, M. Dembo, M.A. Yang. Dating the great ape phylogeny using phylogenomics and fossil tip dating. CEHG Evolgen Seminar, August 2013, Stanford University.
6. **Schraiber J.G.,** Landis, M.J. Quantitative trait evolution when mutations have large effects. New Directions in Probabilistic Models of Evolution, May 2014, Simons Institute for the Theory of Computing, UC Berkeley.

PUBLIC LECTURES

7. **Schraiber J.G.** Studying Neandertals using ancient DNA. UC Berkeley Osher Life-Long Learning Institute. 2013.

TEACHING EXPERIENCE

Lecturer Summer 2012, Summer 2013
QB3 Python short course
Prepared and conducted lessons to teach students to use Python

Teaching Assistant Spring 2012
Molecular and Cellular Biology 247: Genome Project Lab
Instructed students on how to use Python for bioinformatics

Teaching Assistant
Integrative Biology 164: Human Genetics and Genomics Fall 2010
Responsible for computer laboratory, office hours, and grading

MENTORSHIP

Stephannie Shih (undergraduate student) 2011-2012
Resulted in a publication with undergraduate coauthor

OUTREACH

- **Berkeley High School**, Berkeley, CA 2013
  Taught high school students about using ancient DNA to understand evolution
- **Community Resources for Science**, Oakland, CA 2010-2013
  Led activities that teach middle school students about community ecology

Schraiber, Joshua G.
Research Statement

I am a computational biologist dedicated to understanding the evolutionary forces that shape variation from the molecular to the macroscopic level. I utilize high-powered computing to build theoretical models and develop novel statistical methods that are then tested by rigorous empirical analysis. My work is highly interdisciplinary and I work in a range of model systems, from yeast to pigs. Empirically and theoretically, I am primarily interested in the intersection of population genetics with the evolution of quantitative and complex traits.

Population genetics

Population genetics is the lens through which we look at patterns of genetic variation to make inferences about evolutionary history. Both demography and natural selection leave signatures in the genome, and I work on inferring and disentangling these patterns in high-throughput sequencing data.

Allele frequencies in ancient DNA and experimental evolution. Modern advancements in ancient DNA (aDNA) and experimental evolution have opened a long-shut window into the temporal dynamics of evolution. Because all the information about natural selection acting to change allele frequency is contained in the temporal dynamics of the allele, we can now gain unprecedented insight into the history of alleles subject to natural selection. I am driven to use this data to make inferences about the strength of positive selection in nature, and to ask about the frequency of natural selection acting on standing genetic variation.

Past work: I attacked the mathematical difficulty of analyzing selection by building novel inference strategies based on the probabilities of individual allele frequency paths. In doing so, I quantified the impact of weak natural selection on allele frequency changes and found that weak selection is probably impossible to detect, even with extremely densely sampled time series.

Current work: Building on my theoretical results, I developed a Bayesian method to reconstruct allele age and selection intensity from allele frequency trajectories. I applied this framework to aDNA from humans and horses. Notably, while I found the expected signal of strong positive selection in favor of lactase persistence in Europeans, I also inferred that the lactase allele substantially predates the onset of agriculture, suggesting that selection acted on standing genetic variation.

Future work: Utilizing diversity at linked, neutral sites provides the most powerful tool for detecting recent selection without using aDNA. Inspired by this fact, I will develop a Bayesian method to incorporate linked neutral diversity into inferences from allele frequency time series. This approach fits naturally within the path probability framework that I developed. By conditioning on the path of the selected allele and the genealogy at the linked neutral sites, I will develop a Markov chain Monte Carlo algorithm to improve inferences about natural selection. Using whole genome aDNA data from ancient humans, horses, dogs and pigs, I will apply this method genome-wide to gain insight into the role of natural selection in shaping diversity across a wide variety of species. Because this methodology will make use of all the available information contained in the data, I will be able to definitively answer questions about the frequency of selection on standing variation.

Admixture and natural selection in natural populations. High-throughput sequencing lets us see into the evolutionary history of both model and non-model organisms. Interbreeding among local populations (or even different species) of organisms has important implications for understanding both the evolutionary past, as well as the future in the face of human interference with natural habitats. Similarly, we can gain insight into both past and future by understanding how natural selection has shaped organisms.

Past work: With Martien Groenen's group, I sought to understand the impact of geological and human activity on diversity within and between pig species. Using high-coverage, full-genome data, we found that pigs in island Southeast Asia had experienced multiple periods of secondary contact due to rising and falling sea levels. We also found a substantial role of human translocation in shaping the genomes of wild suids.

Current work: Because of the importance of domestication in shaping human history, I am examining the genomic signatures of domestication in pigs. We find that the process of domestication was diffuse, and that interbreeding between domestic pigs and wild boars was common in both Asia and Europe. Despite the rampant signatures of admixture, we found islands of domestication, including indications of parallel targets of selection in European and Asian domestic pigs.

Future work: Continuing my collaboration with Martien Groenen and Greger Larson, I will utilize aDNA to further refine our understanding of pig domestication. In particular, we will analyze the timescales of gene flow between wild and domestic pigs. Doing so will require novel methodological advances to take full advantage of the available aDNA. Moreover, in collaboration with my postdoctoral co-adviser, Joshua Akey, I will examine demography and admixture in budding yeast. The demographic and selective history of wild yeast is understudied compared to its importance as a model organism in molecular biology. Using a new, full-genome dataset from worldwide samples of yeast, we will analyze population structure and look for signatures of local adaptation, which has only rarely been identified in wild yeasts. We will also look for the genomic signal of domestication in yeast, and compare it to the signal found in multicellular organisms.

Quantitative genetics and complex traits

Many traits closely related to organismal fitness have a complex genetic basis. In the century since Fisher reconciled Mendelian genetics with the observations of biometricians, we have made significant strides in understanding the genetic basis of complex traits. However, we still struggle to understand how evolution at the molecular level maps onto evolution at the phenotypic level. To bridge this gap, I work both within traditional quantitative genetic modeling as well as forging novel methods to jointly model sequence and phenotypic evolution.

Comparative phenomics. Evolution gives us the opportunity to see how closely related species respond to different environmental conditions. By comparing these species phenotypically, we gain insight into the forces that shape evolution at the phenotypic level. I develop novel phylogenetic comparative methods, both to analyze more traditional phenotypes such as body size and to analyze functional genomic phenotypes such as RNA expression.

Past work: One of the great debates in evolutionary biology is the relative importance of punctuated vs. gradual evolution. I co-developed an approach to address this question with comparative data using models known as Lévy processes. With this method, we found evidence that the ancestor of great apes had experienced a burst of brain size evolution. Furthermore, working with Rachel Brem, I developed a method to detect signatures of natural selection in RNA-sequencing datasets. Because these datasets typically contain small numbers of species, I leveraged the power of agglomerating the signal across genes in predefined functional categories. I applied this method and found a strong signature of stabilizing selection acting on gene expression across the *Saccharomyces sensu stricto* species, and several examples of lineage-specific natural selection.

Current work: Working with Joshua Akey, I am generating RNA sequencing data from *Saccharomyces* species and strains subjected to different environmental stresses. I found that a core set of genes act in a conserved stress response pathway; however, each species examined also has a lineage-specific stress response. Interestingly, we find that the genes involved in lineage specific responses show evidence of accelerated evolution at both the level of RNA expression and DNA sequence.

Future work: I will continue to address questions about punctuated evolution. Specifically, I will apply my method to several large comparative datasets, and ask about the frequency of punctuated evolution across a variety of clades. Because the debate about punctuated vs. gradual evolution has largely been one of paleontologists vs. neonatologists, I will augment the method to analyze phenotypic data from fossil taxa. By rigorously analyzing comparative data from both extant and extinct species, I will shed light on long-standing questions about the tempo and mode of evolution at the phenotypic level.

Evolution of DNA sequence and phenotype. I want to explain how natural selection acting at the level of phenotype impacts the variation we see at the molecular level. To approach this question, I have begun to develop joint models of sequence and phenotypic evolution.

Past work: I proposed a model of phenotypic evolution that is explicitly based on mapping mutations onto genealogies at loci that control the trait in question. We found that the details of the mutational effect distribution leave characteristic signatures in the distribution of a quantitative trait, suggesting that is possible to make inferences about mutational effects from quantitative trait variation in natural populations.

Current work: I am conducting research funded by the NSF Postdoctoral Fellowship in Biology with Joshua Akey and Jonathan Wakefield to assay within and between species variation in DNA sequence, RNA expression, and chromatin accessibility in the *Saccharomyces sensu stricto* species. We will be assaying genome sequence, RNA expression and chromatin accessibility in population samples from 5 different *Saccharomyces* species. Ultimately, we will use patterns of polymorphism and divergence to assess the signatures of selection acting at these three distinct levels of phenotype. I am also developing a Bayesian approach to infer the parameters of mutational effects on RNA expression levels. This will allow us to understand the raw mutational input upon which natural selection acts. We will apply this method to analyze several population-level expression datasets to ask if the parameters of mutational input are similar in different taxa.

Future work: Like much work in theoretical population genetics, I began by assuming that quantitative traits evolve neutrally. An important refinement of this model will be to incorporate the effects of weak selection on the trait in question. I will use a perturbative approach based on the ancestral selection graph to understand how weak selection changes the distribution of a quantitative trait in a population. Once this is computed, I will extend the approach I am currently developing with Jonathan Wakefield to infer the parameters of natural selection acting on RNA expression. A full, mechanistic understanding of how selection operates from molecules to phenotypes will require the development of population genetic models that relate sequence evolution to phenotypic divergence. By combining my experience developing models of sequence and phenotypic evolution, I will develop first-principles models that jointly describe the evolution of genotype and phenotype. I will then test these models by applying them to publically available datasets of molecular phenotypes, including assays of chromatin accessibility and RNA expression

Schraiber, Joshua G.
Teaching Statement

Sometimes, people grow up without noticing the distinction between biology and stamp collecting. They think that biology is about memorizing binomial nomenclature, anatomy, or amino acids, and that these words are to be memorized for a test and then quickly forgotten. But this view could not be further from the truth. Biology is a tangled web of interconnected concepts, theories and hypotheses about how living things work. Evolution is the thread that ties that web together, and my primary goal in teaching, mentoring and outreach is to help both students and the general public see this.

My diverse background ensures that I am qualified to teach a variety of courses. My research exists at the intersection of mathematics, statistics, computer science, and biology. Hence, I would be able to teach introductory biology courses as well as more advanced courses in statistics, computational biology, population genetics, or evolution. I see laboratory exercises as an essential part of any curriculum, because they promote active learning. In particular, computational laboratory exercises should force the students to come up with their own solution to problems, to ensure that they truly understand the underlying concepts. This will also assist them in their futures, where computational skills are becoming more important in both industry and academia.

As a graduate student, I was able to teach at both the undergraduate and graduate levels. During my teaching, I stressed the importance of seeing connections between seemingly disparate areas of biology to my students. To do this, I crafted assignments that would engage the students, and force them to utilize all of the information they learned during the course. For instance, while teaching undergraduate human genetics, we created an assignment that required students to read a paper about disease gene mapping using natural variation. This required them to bring together their knowledge of both classical and molecular genetics and interpret it within a population genetics framework. I also co-taught a two-week-long bootcamp to introduce biologists to computational techniques; we met five days a week, for eight hours a day. This required me to adapt and develop different teaching strategies than those I would use in a normal classroom setting.

In addition to teaching at the university level, I also went into local schools to ensure that students from diverse socioeconomic backgrounds could have the opportunity to be as excited about evolution as I am. I believe it is important to show children from a young age that scientists are people, too, so I worked with Bay Area Scientists In Schools to teach community ecology and predator-prey relationships to middle school students by using interactive activities. In addition, with other graduate students and postdocs, we developed a series of lectures and activities to teach high school students about evolution, phylogeny and natural selection. During my postdoctoral fellowship, I am continuing my outreach by working with HiveBio to develop courses for the general public on inferring phylogenies from genetic data.

Classroom learning is essential for turning students into productive scientists, but receiving mentoring from more senior individuals is critical, as well. I was fortunate to mentor an extremely bright undergraduate (Stephannie Shih) during my graduate work, and our collaboration resulted in a publication. She is now pursuing a Ph.D. in epidemiology at Brown University. Additionally, I worked with several younger graduate students in a mentorship role. Several of these collaborations resulted in publications in which no faculty members were involved, and I took on the duties of providing guidance and direction for the research.