

CONTENTS

Beerli to pursue genomic research



2

Research on frog calls



4

Summer courses



8

New equipment for genetic research



Recently, Professor Peter Beerli was awarded a \$750,000 grant from the National Science Foundation to develop a mathematical framework for investigating genomic sequences that respond differently to local environmental stresses. This framework will compare mutation patterns from samples taken at different locations to estimate possible genealogies of these samples; these genealogies will be used to estimate parameters of population genetic models.

“These parameters will be able to tell us about potential selection differences among geographical locations. This new mathematically rigorous framework has the promise to supersede current ad hoc and inadequate summary statistics,” said Beerli.

The framework has many potential applications, such as improving interventions to diseases (e.g., individualized responses for HIV patients), or refining our understanding of which gene regions are responsible for long-term survival in harsh environments. The framework will be publicly available in stand-alone computer software that can be run on small computers or large computing clusters.

“This research draws from multiple science and technology disciplines (biology, computational science, and statistics) and thus will provide a great basis to mentor undergraduate, graduate, and postdoctoral

students in a field that desperately needs more training opportunities.”

The building blocks for this framework are rooted in coalescence theory, a branch of theoretical population genetics which studies the shapes of genealogies of individuals, and bayesian statistics. The research involves evaluating different scenarios and integrating possible solutions using Markov Chain Monte Carlo technology. The data will be genomic sequences which will, of course, contain errors; to successfully differentiate among gene regions that are under selection for particular environments, these errors will need to be taken into account. Additionally, samples from different geographical locations (for example different patients,

See Beerli, page 3



Peter Beerli [right] discussing research with Sachin Shanbhag

Frog call studies show geographic variation in vocalization patterns



SC grad student Alex Townsend

Vocalizations play an important role in the social structure and development of many species of animals including hominins, apes and humans. They evolve and diversify to become complex languages. Studying how vocalizations change and evolve is a central pursuit in understanding the social structure, behavior and evolution of many animals, including humans. One component of vocalization research which is remarkably pronounced and may play a role in speciation via reproductive isolation is spatial variation.

Alex Townsend studies how these vocalizations change and evolve. “I’m currently working on a project with the morph lab that investigates patterns of geographic variation in animal vocalizations. Essentially, I want to know if animals have regional accents like humans do,” Townsend said.

Using a Cornell University repository, Townsend downloaded a sample of 187 mating calls of Cope’s Gray Tree-

frog (*Hyla chrysoscelis*) with GPS coordinates of recording locations in Universal Transverse Mercator or UTM format, a common coordinate format for GIS applications. After obtaining the samples, Townsend removed the noise from each sample, extracted a single pulse from each frog call, then aligned them using a modified form of Generalized Procrustes Analysis. He then clustered the frog calls into groups using a form of Hierarchical Clustering. By using both real and simulated test data, Townsend’s research demonstrates a reliable and powerful way to quantitatively analyze the geographic variation in vocalization patterns.

Townsend has been active this year in presenting his vocalization research and his list of research interests is extensive and includes acoustic ecology, bioinformatics, machine learning, high-dimensional statistics, time series analysis, geometric morphometrics, computational and data-driven medicine, computational statistics and digital signal and image processing. In addition to presenting his vocalization research

at the national conference of American Association of Physical Anthropologists and the Southeastern Ecology and Evolution Conference this year, Townsend is creating software that analyzes spatial signals. Dennis Slice is his supervising professor.

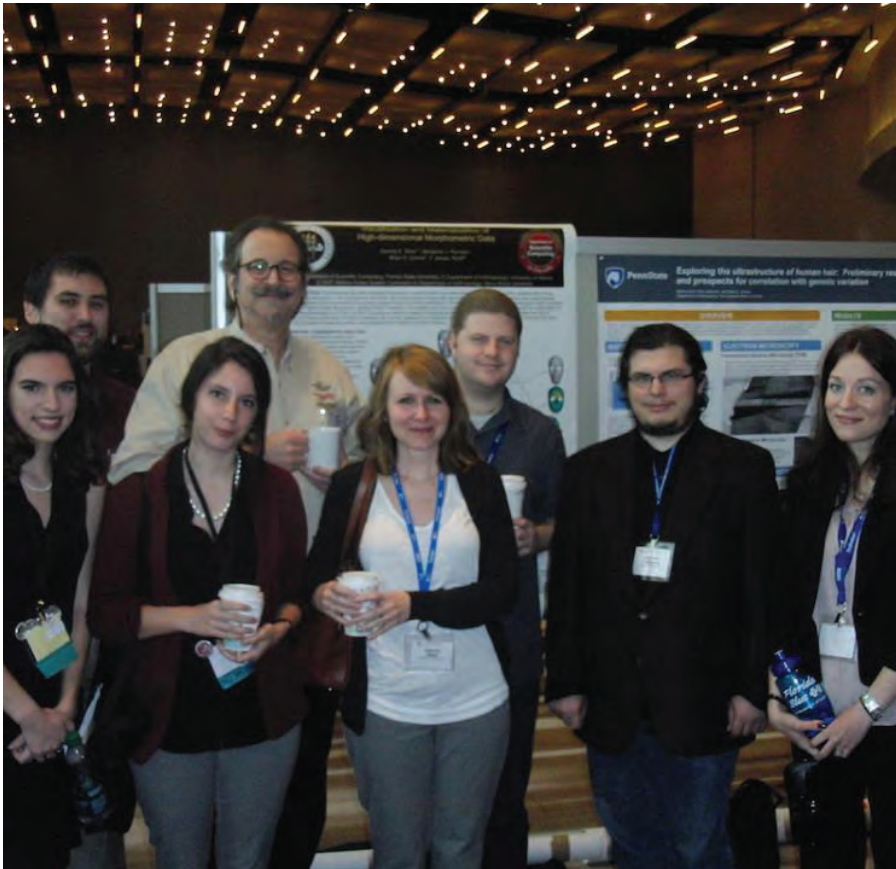
You can see more of Townsend’s research online. To see a video of a talk Townsend did on his research at GEOSET, go to the department’s YouTube page at www.youtube.com/FSUSciComp.

You can review research by Townsend and Slice at <http://people.sc.fsu.edu/~sat10c/posters/>. Once you get to the directory, click on SEEC_2016.pdf and AAPA_2016_XX.pdf.



Slice lab travels to Atlanta

The Morphometrics Lab, led by Professor Dennis Slice, traveled in April to the American Association of Physical Anthropologists Conference to present their research and to hear Lee Berger, Slice collaborator and leading scholar in morphometrics and physical anthropology.



Slice Morphometrics Lab at AAPA Conference. Above, Back Row, Left: Ben Pomidor, Dennis Slice, Cameron Berkley. Front Row, Left: Alexa Pennavaria, Devin Geraci, Kathryn Miyar, Alex Townsend, Detelina Stoyanova. Top: James Soda. Bottom Photo: Dinner with featured speaker Lee Berger, collaborators and colleagues in the field.

Beerli continued from page 1

different islands, or different habitats) can be grouped in different ways, which requires that the framework be capable of delivering optimality criteria that help to order different scenarios.

Beerli will be assisted in the project by a graduate student and a research assistant. The grant began May 15, 2016 and will continue through May 2019; progress and the final work will be documented at <http://peterbeerli.com> and <https://popgen.sc.fsu.edu>.

Summer offers extended to students

Cresswell selected for Princeton Neuroscience Institute program

Evan Cresswell will attend Princeton University's newly developed summer intensive course designed to introduce students with quantitative training in the sciences to the concepts and research methodologies of modern neuroscience. The course is unique in its focus on neural dynamics at several scales of complexity and will include lectures and laboratory exercises. The capstone of this course is a one-week student-designed research project which integrates concepts and methodologies encountered during the initial formal lectures and laboratory exercises.

"Through this course, I can gain a stronger grasp of the link between the mathematics I've been taught and the biology that drives it, so my research will have a stronger connection with the underlying experiments. This understanding would also help me guide further experiments or reject the results from modeling effort. In addition, a better understanding of experimental neurobiology will make more of that research accessible to me than is possible with mathematics alone," said Cresswell.

Lei receives Los Alamos internship

Doctoral student Hongzhan Lei will intern at Los Alamos National Laboratory in the Computational Earth Sciences Division of Earth and Environmental Sciences Group. He will work with staff scientists Zhiming Lu and Velimir Vesselinov to identify heterogeneities in porous media. Lei will be at Los Alamos through August.

Bystricky to attend August lecture series

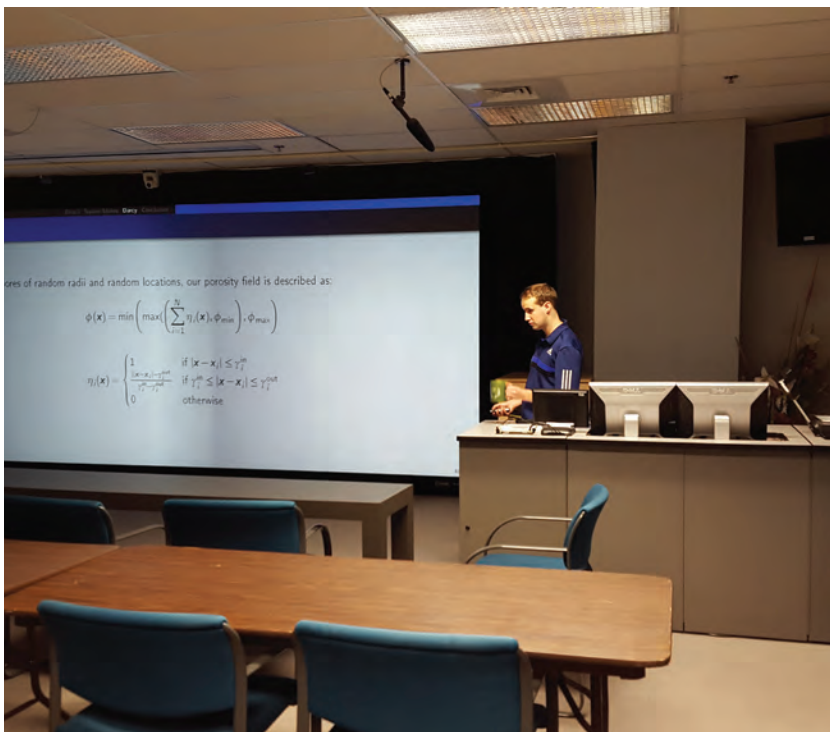
Lukas Bystricky was selected to attend the SAMSI Summer School in Optimization, to be held throughout the month of August at the Hammer Conference Center in Research Triangle Park, NC. The Summer School, which is part of the 2016-17 SAMSI Research Program on Optimization, will introduce graduate students and early-career researchers in the mathematical and statistical sciences to cutting-edge approaches in Optimization and its applications.

Lecturers for the series will include Eric Chi, a statistician from North Carolina State, Vanja Dukic, an applied mathematician at CU-Boulder, Noemi Petra, a mathematician at UC-Merced, Volker Schulz from the University of Trier in Germany), and Hua Zhou, associate professor of biostatistics at UCLA.

Potential topics to be covered include: expectation-maximization algorithms, optimization transfer algorithms, sparse regression; computational methods for PDE-constrained optimization; Bayesian inference; statistical and mathematical inverse problems; optimization under uncertainty; and robust optimization.

For more on SAMSI, go to www.samsi.info.

To see notes for Noemi Petra's lecture entitled "Hippylib: An Extensible Software Framework for Large-Scale Deterministic and Linearized Bayesian Inverse Problems," go to the SAMSI website.



Lukas Bystricky presenting research in March 2016

Conry to attend international summer workshop

A Scientific Computing graduate student has been selected to attend a well-respected international learning program this summer. Michael Conry was chosen to participate in the International High Performance Computing (HPC) Summer School on HPC Challenges in Computational Sciences. The primary focus of this year's workshop is parallel programming and HPC methodologies in life sciences.

Conry was selected for the International HPC Summer School after competing with science and engineering students from all fields. Preference was given to applicants with parallel programming experience and a research plan that demanded high performance computing systems. When Conry learned his background and expertise were the focus of this year's workshop, he was especially enthusiastic.

"This year, they were specifically looking for students who had a background in biological sciences and programming so I knew I had a good shot at getting in. Also I knew Michelle Perry had gone a few years ago and mentioned it somewhere in my application. I was reminded about the workshop because of the fliers that Dr. Burkardt hung around the department."

As a Computational Science student, Conry's coursework has helped him recognize the importance of parallel programming techniques when it comes to reducing the effort needed in solving difficult scientific problems. He hopes to use the knowledge acquired at the summer school to parallelize existing and future projects and optimize code.

In his application, Conry cited the projects he has planned for his masters and doctoral research, and explained how he would apply the skills he learns. Both his thesis and dissertation topics are biologically and computationally based.

"The project I am currently working on for my master's thesis is a simulation study of DNA Recombination in population genetics. We are hoping to determine the often overlooked impact that meiotic recombination rates and distances between exons on a chromosome may have on phylogenetic trees. I am writing java software that can simulate and visualize many thousand individuals mating and their alleles crossing over. The application still requires multi-threading.



Michael Conry

"In the fall, I will begin working on my doctoral dissertation with a cell therapeutics synthetic biology group. We will be creating simulation, analytic, and visualization tools (in python) to predict the behavior of custom made cells that have been programmed using DNA plasmids to do specific tasks such as sensing a chemical signal or releasing a drug. I will likely be the main programmer and the software will require parallel programming."

Conry believes more students should feel confident about applying for future workshops. "I have always been interested in meeting students from varying backgrounds, learning about their approach to their research and how their thinking can influence my own. The application didn't take very long, maybe a few hours. Others students should definitely be encouraged to go in the next few years."

The workshop will be held June 26 - July 1, 2016 in Ljubljana, the capital and largest city in Slovenia. The International HPC Summer School is targeted to leading European, Canadian, Japanese, and American computational scientists and HPC technologists. The week-long workshop will have lectures on different areas of science and how to apply parallel program concepts.

For more information on the International HPC Summer School, go to <http://ihpcss2016.hpc.fs.uni-lj.si/agenda>.

Mechtley selected for Google Summer of Code



SC grad student and Google Summer of Code participant Alisha Mechtley

You were selected for the Google Summer of Code!

Who will you work with, and what is your project?

AM: I am working with the Open Bioinformatics Foundation (OBF). My mentors are Jaime Huerta-Cepa and Renato Aves, two researchers from EMBL (European Molecular Biology Institute). The third mentor is Francois Serra who is working on a similar project for chromosomes at the Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Universitat Pompeu Fabra (UPF) in Barcelona, Spain.

Why did you decide to apply?

I decided to apply to GSOC thanks to a flyer put up by John [Burkardt] next to the elevator. He personally mentioned it to me as we were both heading out of the building one day. He mentioned the \$5,000 stipend and the opportunity to work with experts in the field.

How did you select an organization?

I went to the GSOC organizations website and immediately did a search on the term “bioinformatics”. Several projects

popped up. I talked to my advisor about the top ten that I was interested in and he narrowed it down to two that he thought were the best. I immediately contacted the organizations I wanted to work with.

What happened along the way? Were there any setbacks or obstacles?

Late Application:

The people at OBF mentioned that it was a bit late to have them look over a proposal. Projects had only been approved by Google the week before and the application window had literally just opened up so it seemed early to me, but students were encouraged to contact the organizations they were interested in working with and begin coming up with proposals MONTHS in advance according to the GSOC student handbook. I didn't let that deter me though; I submitted a GoogleDoc the next day with the hopes that they would leave a few quick comments, and they did. I took an additional day to make the changes.

Proof of enrollment:

My proof of enrollment was rejected. My Spring courses (dissertation credits) did not appear on my unofficial transcript as I expected (unlike other universities, FSU only

shows courses for which you have already received a grade). You have to submit an enrollment verification form instead. Luckily, Google sent me an email and gave me a few days to resubmit the form.

Any idea what the competitive landscape looked like? How many people applied? How many were accepted?

There were 1,206 student proposals accepted of the 7,543 proposals submitted from 5,107 students. Each student can only be accepted for one project. That means that about 1 in every 4 to 5 students that applied was accepted. The chances of being accepted are likely to be higher for students pursuing advanced degrees in a computational field. I believe four students in our department applied and two were accepted. The accepted students are currently displayed on the project pages (replacing the project descriptions written by the mentors with the proposal abstracts of the students, which is unfortunate because you can't compare the original project description to the abstracts). They may also be announced on the organization sites, as mine was for OBF.

Do you have any advice for other students who might be interested next year?

Some things I did to improve my chances of being accepted:

1) Submitted a GoogleDoc to the organization early to get feedback.

2) Submitted more than one proposal (not to the same organization though).

3) Reviewed all of the links the organization sent me and tried to incorporate something from every resource.

4) Reviewed their current code on GitHub, fixed a line of code, and submitted a pull request (GitHub is the way that most open source projects share code and by submitting a pull request, I am basically contributing to the project before I am selected - a clever trick that I don't think most students are brave enough to try but certainly will set you apart).

5) Mentioned the lab I work with, the resources I have available to me, my education, and anything that sets me apart from the other candidates (anything even remotely relevant to the project) in the proposal.

I also submitted another proposal that was not accepted. Among the many projects that were approved by Google initially, only seven were listed in the end. The project I wrote a proposal for called "Galaxy Pages Overhaul" for OGI (Open Genome Informatics), was not listed among the final accepted projects. In other words, another student didn't write a better proposal for that project and get selected for the project instead of me. The project was likely not weighted as being as important as the other projects in the end (either by Google or OGI) and no one was chosen to work on it.

Mechtley's project extends through the summer and ends August 23rd. For more information on Google Summer of Code, go to <https://summerofcode.withgoogle.com>.

Abstract of Mechtley's project

In phylogenetics, trees are used to illustrate genetic relationships among species. The ETE (Environment for Tree Exploration) is a Python toolkit for the analysis and visualization of phylogenies. The toolkit has various methods for comparing trees, such as the Robinson-Foulds Distance, but these methods compare the overall difference between two trees and do not allow for custom pattern matching.

In mathematics, a standard way of representing graphical trees with edge lengths is the Newick format which uses a specific syntax (such as parentheses and commas). One way of defining a tree pattern could be to extend the Newick format to include rules and filters with a regular-expression-like vocabulary. These patterns could then be searched for using a tree traversal algorithm. Once a method is in place, common patterns could then be extracted from a group of trees (e.g., using machine learning). Parallelization of the method would extend search capabilities to be used on thousands of trees. If time permits, I would also like to further extend the features to include a visualization component.

Department of Scientific Computing
400 Dirac Science Library
P. O. Box 3064120
Tallahassee, FL 32306-4120
www.sc.fsu.edu

First Class
US Postage
PAID
Tallahassee FL
Permit 55

The department's mission is to be the focal point of science and computation at Florida State University. Gordon Erlebacher is the Chair of the Department of Scientific Computing. He can be reached at 850.644.0143. Newsletters are issued three times each year. Subscriptions and single copies are available by calling 850.644.0196. This publication is available in an alternative format on request.

Lemmon awarded CRC grant

SC Associate Professor Alan Lemmon was recently awarded a Council on Research & Creativity (CRC) Equipment and Infrastructure Enhancement Grant from the Office of the Vice President for Research.



Alan Lemmon

Lemmon, along with biology professor and close collaborator Emily M. Lemmon are co-directors of the Center for Anchored Phylogenomics; their research in genetic sequencing has attracted over \$1.6 million in contracts and collaborative projects.

The grant will fund purchase of a new PippinHT, an instrument which assists in size selection of libraries of genetic data (DNA fragment length). The Pippin is a high-throughput DNA sizing platform for sequencing workflows. The instrument assists the professors, their graduate students and staff in

testing new protocols, exploring new directions in research, and generating high quality data for collaborators and contracts. The award amount is \$21,500.

Lemmon's lab sequences samples of enriched genomic data and performs the bulk of downstream bioinformatics including assembly, orthology, alignment and preliminary phylogeny estimation. These sequencing methods have provided the locus and taxon sampling required to study vertebrates and Lemmon has worked with over 100 collaborators to study the Tree of Life.

For more on Lemmon's research, go to these links.

www.anchoredphylogeny.com
www.research.fsu.edu/crc/