Department of Scientific Computing

CONTENTS

Lemmon enriches genome research

For years, phylogeneticists have attempted to reconstruct the earth's evolutionary origins by assembling the Tree of Life (TOL), a comparative and predictive framework for all life forms over the past 3.5 billion years. This work is important, complex and challenging, with such difficulties as confirming phylogenetic estimates based on relatively small numbers of markers, filling in the gaps in taxon sampling, resolving more difficult branches in the tree, and combining results of dozens of independent research groups into a single well-resolved phylogeny.

As the research becomes more comprehensive and intricate, continued progress on the TOL requires correspondingly sophisticated models and methods to reveal the increasingly complex biological relationships between species. One aspect of SC Assistant Professor Alan Lemmon's research addresses gene sequencing in species lacking a reference genome, or nonmodel organisms. "There are very few species that are completely sequenced – for example less than 0.05% of animals have been sequenced. One of the big problems is that people who aren't working on model systems couldn't access the genes they need to do their research. Genomes are very large, and without a reference genome, they couldn't fish out the things they needed. If you don't know what you're looking for, you can't get it out, so it would be difficult to extract the information," says Lemmon.

To address this problem, Lemmon developed a novel method for rapidly capturing large data sets to resolve species relationships by using existing genomic sequences, target enrichment techniques and high throughput sequencing for nonmodel species, without the need for additional primer development or testing. The new approach is called anchored enrichment.

"What we did is we developed a way to take existing genomic resources – such as human genome and mouse genome, etc. – and leverage those resources to give us access to any vertebrate species. It takes millions of dollars to sequence and assemble one genome, so we used those few million dollar investments as a jumping off point. We're doing this by comparing the genome to the species that exist, and identifying regions we can use as anchor points."

Lemmon's anchored enrichment method uses a set of targeted probes within specific genomic regions to capture data within vertebrates. The probes capture genomic fragments in regions that are highly conserved, flanked by less-conserved sites, unique within the genome and widely distributed throughout the genome. The method can be easily customized to fit the phylogenetic scale of interest, permitting deep and shallow level phylogenetic studies. Once a capture probe set is fully developed, it can be reused for any collection of organisms within that taxonomic group. In addition, anchored enrichment is a faster and more economical approach to generating hundreds of variable, single-copy nuclear loci for any system using next-generation sequencing.

"Our technique is a method for designing

Message from the Chair

DSC CHAIR MAX GUNZBURGER

Our students who finished their degrees in the last year have not only been able to obtain excellent positions, but have also unanimously expressed a very high level of satisfaction with the training and mentoring they received from us.

We start the Fall 2013 semester with our largest incoming class of graduate students and with a steadily growing cohort of undergraduate students. Perhaps more important, our students who finished their degrees in the last year have not only been able to obtain excellent positions, but have also unanimously expressed a very high level of satisfaction with the training and mentoring they received from us. We have also received unsolicited comments from employers praising not only the quality of the students we have graduated, but also about the broad and deep training in computational science that those students received in our degree programs.

But, we are not resting on our laurels. When we designed and implemented our degree programs,

especially the B.S. program, we had very few programs (none at the undergraduate level) elsewhere to use as models. Although the faculty thought long and hard about what our programs should look like, we knew that we would not get everything 100% perfect. Our degree programs, both at the undergraduate and graduate levels, have been around long enough for us to learn what can be improved. Feedback we sought and obtained from students has been extremely valuable in forming our ideas for what needs to be changed to better our programs. Thus, in the coming year, we will be again thinking long and hard about how we can improve all aspects of our B.S., M.S. and Ph.D. degree programs.

tiny probes or pieces of DNA from known genome sequences. First, we determine a region that's fairly conserved, so it could be a region involved in making a backbone for example. We put these probes that we already know the genome sequence of - a human, mouse, chicken, etc. - in a solution in a test tube. These probes act as little hooks which allow us to catch pieces of DNA in a test tube from any species. We can take the DNA from any vertebrate species for which we don't have the genome sequence, break the DNA into pieces and put them into a tube with the probes. Then we can pull the most useful fragments out of the tube and sequence it. We submitted a patent application because it could have important implications - especially for medical and agricultural research."

The technique has been highly successful, causing an immediate and intense shift in the Lemmon lab structure, scope and work flow. A surge of pent up global interest in comparing phylogenetic diversity across species emerged from scientists who have questions they've wanted to answer for a long time, and private companies are interested in the method's commercial potential.

"We're getting so many contracts that our current method of processing the samples is overwhelmed, so we identified a robot that would help us increase the rate at which we are processing samples. The robot has 96 pipettes on a single head, so it can suck up 96 liquids at the same time from different wells. This will dramatically increase our ability to process samples. So instead of processing 100 samples in two weeks, this robot will help us sample 100 samples in about one day.

That will help us increase more of everything - more contracts for the lab, more revenue for the university, and more research published. This will allow us to increase the flow by quite a bit. We do the sequencing here at the College of Medicine, and we do laboratory work here at our lab. It's a nice process. Right now, we have about $500K in contracts and about $500K on deck to come in."

"We have roughly three dozen collaborators right now, and we're working with people all over the world. We have a big contract with a researcher in Brazil to sequence all the snakes in Brazil. We're sequencing all the squamates in Australia - squamates are lizards and snakes, skinks, scaled reptiles - things like that. We're setting up a contract to do a large number of fish - around a couple thousand fish species. We have contracts with people at Yale and Berkeley to do birds. And we're just now getting into the insects. We have people with all flavors of insects - beetles, butterflies, grasshoppers should get started soon, cicadas, you name it, we're doing it. It's very exciting because it's accelerating the rate at which we can process the data by a factor of 100, and we're also reducing the cost by a factor of 100."

High impact potential for the method has many applications in academic science, in corporate research studies, and in other areas of Lemmon's research. "Because we saw a potential commercial application and submitted a patent application, I got some funding from the GAP program on campus, then that led to a six month NSF I-CORPS training grant to help me and other academics think about the business side of the operation. I attended a workshop in DC to get some training on entrepreneurship and that gave

Comparison of Genomic Partitioning Strategies
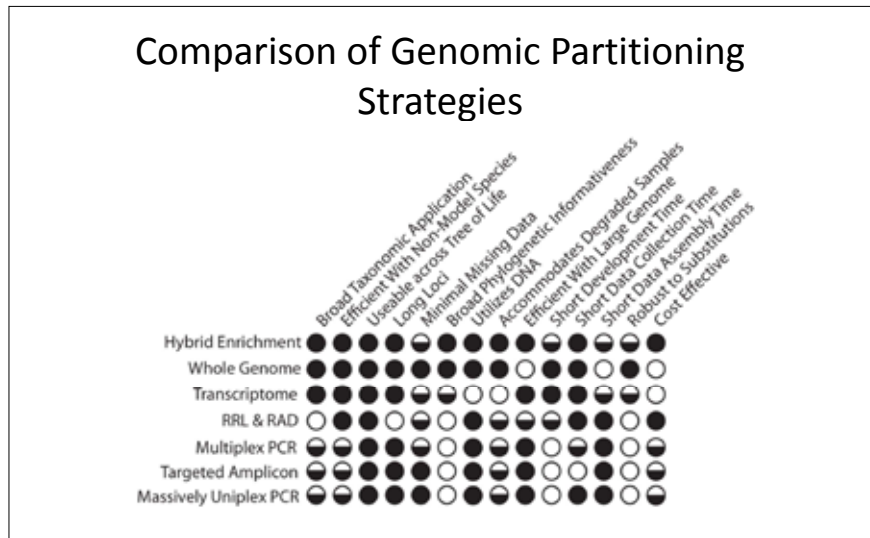
Comparison of attributes of genomic partitioning approaches for phylogenetic data collection, Lemmon and Lemmon Annual Review of Ecology, Evolution, and Systematics, Vol. 44, 2013

me an opportunity to make inroads into potential markets. We talked to 85 different companies with the ICOR program funding.

"We're also starting to use this method for targeting classes of genes that are important for studying evolution or important for economic issues. For example, I'm a co-PI on a $900K snake venom grant to study the evolution of snake venom – how snake venom evolved, mainly in North American species. The slow way of doing that is to extract the venom out of the organism which can require capturing, then killing or sedating the organism, in order to get the RNA. That gives us an idea of what genes are expressed in the venom, but that process takes a long time. Using our anchored enrichment approach, we can take information we've obtained from the RNA of a few specimens and develop a capture kit to capture all the snake venom genes from the genome so we can use genomic DNA from a tail clip or road kill. We can use this process and our new robot to process thousands of

snake species studying across species variation in the snake venom.

"The cool thing is this approach we developed requires quite a bit of bioinformatics upstream to identify these target regions and develop the probes. When the data comes back from the sequencing matching, we get back approximately 30 billion nucleotides of data - about 100 GB per project - and we begin to analyze data. We're doing a lot of networking with people and collaborators, and we have been looking for new avenues, new ways of using this type of tool. It's exciting to think that any vertebrate we discover tomorrow, we can sequence these genes from the genome, and we can do it quickly and economically."

For more information, go to:
www.anchoredphylogeny.com
www.evotutor.org
www.phylo.gor/atol/
www.annualreviews.org/journal/ecolsys
sc.fsu.edu

Fun at Oak Ridge: Scientific Computing out by the pool for a bit of relaxation (Above). John Burkardt meets a future SC alum (Below left). SC joins the FSU sciences in recruiting (Center). Recent grad Serena Pham prepares to present her research (Right).
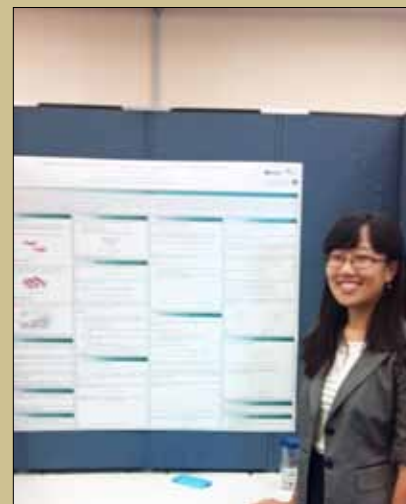
Spotlight on SC Alum Ian Johnson

Before beginning his graduate studies, Ian Johnson was crystal clear on what he wanted to accomplish as a professional, and believed that a computational science degree would give him the necessary skills. "I knew that I wanted to code for a living, and that I wanted to work on things that touch people's lives directly. A major reason¬ I joined the department was to improve my coding and math skills, which I saw as a pre-requisite for understanding advanced algorithms."

Johnson completed his MS degree in August 2011, and accepted a position at visual.ly, a community platform for data visualization and infographics. He found out about the position at a hackathon while doing an internship at Lawrence Berkeley National Lab.

"I joined a random team and got along well with one of my teammates. She knew about the position, and encouraged me to apply for the job. Hackathons are really easy to get involved in; you just show up! Usually there is some time at the beginning to find teams, the most fun (and productive) thing is to join a team with people you don't know. I find an open mind and curiosity are the most important things in order to have fun. A lot of people are starting to treat hackathons as a free resource for development so you do have to think a little critically about which one to take part in. Luckily there are so many to choose from it's not hard to have a good time."

Johnson left his position as programmer at visual.ly in 2012 to follow his entrepreneurial instincts, and started freelance programming work with a group of colleagues. Their target clients: any organization that had data they needed to communicate to people through visualization.

Johnson freelanced primarily in the San Francisco Bay Area, doing contract work for several consulting companies and at Stanford University where he worked for a lab. Lever, one of the companies where he consulted, offered him a full-time position which he decided to accept. "Lever is a startup that builds software to help people stay on top of the hiring process. I met one of the founders at a meetup a while back, and we became friends over a year ago. My title is simply Engineer, but I'm the first employee they hired full-time."

It's working with the data and its possibilities that is the common denominator and is most attractive for Johnson. "I write code to help people visualize data. It's not data itself that fascinates me, but rather the potential for insight that is often hidden within the data that is interesting. By collecting and analyzing data, we can find patterns in our behavior as well as the behavior of the systems around us. Data visualization is an essential tool in discovering these patterns and communicating them to other people. I don't have specific goals to accomplish in this vein, but I enjoy helping other people understand more about the world around them."

"This was a big focus at SC while I was in the visualization lab. Although the details are a little different, I find that many of the skills I learned apply directly. Honestly, the most lasting thing I took from the department is the scientific discipline Gordon [Erlebacher] tried to instill in me. I can still hear him questioning assumptions I made; it helps me to be a better programmer as well as businessman. The work I am doing is not scientific by nature, but we are collecting large amounts of data and are in the position to make scientific inquiries. I also hold myself and my code to a higher standard thanks to Tomek's [Plewa]V&V class. Too many people out here trust their own and others ability far too easily.

"Because of the training I had in numerical analysis, linear algebra, signal processing and machine learning in my various classes, I can easily understand and evaluate the simpler algorithms being used in my field currently. With the data in hand I am able to apply and analyze more sophisticated techniques."

SC Alum Ian Johnson

Although he's happy where he is now, Johnson is already thinking entrepreneurially about his future. He plans to develop and operate educational programs in technology and art to help young people create things. "I'd like to organize large open source and community initiatives that empower people to create with digital tools. My hope is to be able to independently fund these kinds of activities, or at least have enough community support to focus on them full time. It's hard for me to think too far in the future; the tools I'm using and building change so fast that the only thing I know for certain is that I will be a lot more productive by the time I'm ready."

More info is available about Johnson and his endeavors at enja.org.

Department welcomes new researchers

Dongwei Gui

Peigui Liu

Aretha Teckentrup

Dongwei Gui comes to Scientific computing with an undergraduate in Computer Applications from Xinjiang Normal University and a Masters in Oasis Ecology from Xinjiang University. He received the Ph.D. in Oasis Ecology from Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences in Wulumuqi, China in June 2010. Gui's research areas are agriculture management, water use and management, hydrologic modeling, soil quality and sustainable development. He is published in multiple journals and has been honored for his research and academic achievement.

In 2010,Gui worked at the Xinjiang Institute of Ecology and Geography as a Research Assistant. He also worked at the Secondary Vocational Technology School of Xinjiang as a teacher and lecturer of computing applications.

Gui enjoys staying active by playing table tennis and basketball, and he and his wife Yan Qin have a one-year-old son, Shiyu. He is working with Ming Ye to study the use of irrigated water in urban landscapes.
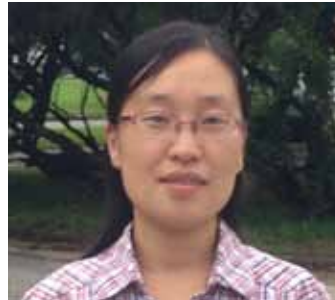
Peigui Liu received her B.S. degree in Water Resources and Hydrology Engineering in July 2004 from Shandong Agricultural University, and was awarded the Ph.D. in Hydrology in December 2008 by Hohai Unviersity. Since 2009, Liu has been a lecturer in the School of Civil Engineering at HeFei University of Technology in Eastern China.

Liu has studied water resources broadly, and has performed studies of water resources in mining areas, assessed groundwater resources in well fields, investigated the environmental impacts of mine drainage, and researched groundwater thermal transport in shallow aquifers. She is published in multiple journals, including the Journal of Hydrologic Engineering, Geotechnical Investigation and Surveying, and Environmental Earth Sciences.

Liu is a movie buff and an avid climber. She is married to Manting Shang, and has a three year old son, Hanzhang Shang. Liu works with Ming Ye.

Born in Cologne, Germany, Aretha Teckentrup grew up in the south of Norway, then studied in the United Kingdom at the University of Bath. She completed her Ph.D. in mathematics in 2013. Teckentrup's doctoral research was on the development and analysis of multilevel Monte Carlo methods for model problems where the coefficients are rough and highly oscillatory. These multilevel methods greatly reduce the cost of the classical Monte Carlo estimators, and allow researchers to compute quantities of interest to a high accuracy in a feasible amount of time.

In her spare time, she enjoys running, reading, sports, and travel. She will work with Max Gunzburger.

Doctoral grad begins new position at Google

At 8:04 a.m., Michal Palczewski is waiting for the movers to arrive. He's not sure when they'll show up; he just knows they're coming some time today. "I'm not allowed to pack anything. After they get here, I'm not supposed to touch a single item because of the liability clause."

Things have moved quickly for Palczewski in the last half year; he finished his doctoral thesis, and worked as an intern and as a full time employee. It looked like he made a smooth transition into a stable career, but then Google called. Now he's headed to California to work for one of the world's premier technology companies.

The interview process began when Google representatives contacted Palczewski using information they found on his LinkedIn profile, which showed he was a candidate in Scientific Computing for the Ph.D., and that he was developing mobile applications for the iPhone and Android mobile devices. They also noticed Palczewski was President and founder of the Tallahassee Android Developers Club. Android is an open source Linux-based operating system for touchscreen mobile devices owned by Google.

"One day I came home, and I had a message from a recruiter saying they wanted to talk to me. I was pretty happy with my position at Datamaxx, but I thought I'd go ahead and talk to them, so I called them back and we scheduled an interview. I had my first phone interview a week later. A week after that, I had another phone interview, and a week after that, they flew me out to Mountain View for an in-person interview. They made all the arrangements, and when I arrived in Mountain View, I had a rental car and a hotel room waiting for me."

"The next morning I showed up at the Google campus, and it was quite a day! I had three separate interviews before lunch, then after lunch I had two more interviews. All told, I had five interviews with five different engineers from the company. "

Each interview was conducted in a room that had a large white board and markers. The questions asked were technical, and required creative thinking and targeted problem solving. Over and over again, Palczewski had to write programming code on the fly to answer the questions he was asked. He was able to select the programming language he used from a predetermined set of languages. "They let me pick the programming language that I used to code. I could have used any mainstream language like Java, Javascript, C++, or Python. I did mine in Java. Java was my first programming language, and lately I've been doing a lot of programming in Java so I decided to use that language."

In the two week window between the telephone interviews and his flight to California, Palczewski practiced writing code, and read everything he could that he thought might help him. One of the books he read was Programming Pearls, by Jon Bentley, a book with practical programming and fundamental design principles. "I didn't know whether I'd get the job, but I decided to pull out all the stops - to try as hard as I could to do as well as I could in the interview."

Two and a half weeks later, Google called him with an offer. Palczewski thinks they were impressed with several things including his Ph.D. and his Datamaxx experience, but mostly with his extensive programming background. "They do have a lot of Ph.D. grads at Google. One of the guys that interviewed me said they really like Ph.D.s but only those who have the ability to code. I think they were impressed that I had the Ph.D. and the coding background."

Palczewski was recruited to SC's

"Mike always had great ideas about how to code complex concepts and develop algorithms for our projects. "
PETER BEERLI

7

Palczewski continued from p 7

Ph.D. program by his advisor, Peter Beerli. Several years ago, Palczewski was an undergraduate programmer in a lab at the University of Washington where Beerli worked as a postdoctoral researcher. Beerli knew he'd be an excellent candidate for the Ph.D.

"Mike always had great ideas about how to code complex concepts and to develop algorithms for our projects," said Beerli. "An important contribution to our research was a proof that allowed calculating quantities of a test statistic in parallel on a computer cluster, opening the door for massively parallel evaluations of genomic data."

Palczewski is glad about all the changes, but is looking forward to things slowing down a bit. "I got hired on full time at Datamaxx, I got a Ph.D. and now I'm moving to California to work at Google. I love everything that has happened, but at the same time I'm ready for things to settle down. I just want to do one thing for a while."

To see some Google interview questions and viable answers, go to www.careercup.com/page?pid=google-interview-questions. For more info about Palczewski and the Beerli lab, go to people.sc.fsu.edu/~pbeerli/Beerli_Lab/Peter.html.

SC.FSU.EDU facebook twitter

Palczewski snaps a selfie during first day training