

John St. John

UC Santa Cruz Bioinformatics PhD Candidate

UC Santa Cruz
1156 High St
Mail Stop SOE-Grads
Santa Cruz, CA 95064
☎ 831 824 4020
✉ jstjohn@soe.ucsc.edu

EDUCATION

- 2009 – Current **PhD**, *University of California Santa Cruz*, Santa Cruz, *Degree in Progress*.
- Department of Biomolecular Engineering, Bioinformatics Emphasis
- 2004 – 2009 **B.S.**, *University of Oregon*, Eugene, *3.75 Upper Division GPA*.
- Computer Science Major, Bioinformatics Track
 - Departmental Honors
 - Dean's List
 - Biology and Chemistry Minor

TEACHING ASSISTANTSHIPS

- 03/2010 – 06/2010 **Computational Biology Tools**, *BME 110 / BIOL 181*, University of California Santa Cruz.
<http://www.soe.ucsc.edu/classes/bme110/Spring10>
- Enrollment of 50 students
 - Held office hours twice per week
 - Lead lab section once per week
 - Graded tests and homework assignments

RESEARCH EXPERIENCE

GRADUATE RESEARCH

- 9/'10 – Present **Graduate Research Assistant**, *Ed Green, Professor of Biomolecular Engineering*, University of California Santa Cruz.
- Alligator Genome Assembly and Analysis

PHD ROTATIONS

- 9/'09 – 12/'09 **David Haussler, Professor of Biomolecular Engineering**, University of California Santa Cruz.
- Compare repeat masking strategies in the cactus graph generation pipeline.
 - Wrote software to run RepeatMasker on the cluster in parallel
- 1/'10 – 3/'10 **Mark Akeson, Professor of Biomolecular Engineering**, University of California Santa Cruz.
- Wrote code to analyze nanopore signals to detect when single nucleotide shifts through the nanopore are likely to have occurred
- 3/'10 – 6/'10 **Todd Lowe, Professor of Biomolecular Engineering**, University of California Santa Cruz.
- Extracted probable tRNA expression information from sRNA-seq datasets from Tumor and Normal patient tissue samples
- 6/'10 – 9/'10 **Ed Green, Associate Professor of Biomolecular Engineering**, University of California Santa Cruz.
- Examined the reference allele bias in mapped short read data

UNDERGRADUATE RESEARCH

- 6/'08 – 10/'08 **Bruce Bowerman, Professor of Biology, Institute of Molecular Biology**, University of Oregon.
- Automated a powerful bioinformatics pipeline for mining exons of interest from a trace database
 - Used the program on GATA transcription factors of several unannotated genomes to make novel phylogenetic inferences
- 9/'07 – 7/'09 **Joseph W. Thornton, Associate Professor of Biology, Center for Ecology and Evolutionary Biology**, University of Oregon.
- Evaluated performance of statistical model selection techniques on maximum likelihood phylogenetic inference, incorporating evolutionary rate heterogeneity

- Set up, and ran large scale phylogenetics simulations

6/'07 – 9/'07 **Robert Schofield, Adjunct Sr. Research Associate, Department of Physics, University of Oregon.**

- Conducted biomaterials research examining metal ion utilization in chitin
- Developed an educational website geared toward educating, and intriguing the general public about research in Biomaterials. This website can be found online at <http://www.uoregon.edu/~rmss/homepage/>

WORK EXPERIENCE

INTERNSHIPS

6/'06 – 9/'06 **Management Intern, New Horizons In-Home Care, Eugene.**

- Scheduled employees to fill open time slots
- Assisted the Human Resources department in the process of hiring, firing, and managing employees' files
- Assisted with the management of a satellite office located in another city
- Created an updated New Horizon's Resource Guide for the Marketing department
- Assisted the Accounting department in organizing file archives
- Shadowed and assisted the Chief Executive Officer

PROGRAMMING EXPERIENCE

functional	Erlang, Haskell, R	web	PHP, HTML5, JavaScript, CSS3
imperitive	C, Java, C++, Matlab	typesetting	L ^A T _E X, Sweave, Sage
scripting	Python, Bash	database	MySQL

SCIENTIFIC INTERESTS

- Genome Assembly
- Evolution
- Making Science Accessible
- High Performance Computing
- Whole genome quantitate RNA seq
- Phylogenetics
- Machine Learning
- Algorithm Parallelization

PROFESSIONAL MEMBERSHIP

2009 – Present International Society of Computational Biology
 2007 – 2009 Association for Computing Machinery

UNDERGRADUATE THESIS

title *Selecting an Accurate Model of Evolutionary Rate Heterogeneity*

supervisors John Conery, Joseph W. Thornton

abstract Phylogenetic analysis provides the framework for all comparative inferences and historical reconstructions in evolutionary and developmental biology. Modern phylogenetic methods generally produce highly accurate results when the statistical model captures the important features of the evolutionary process, but they can be strongly biased when the wrong model is used. Our laboratory has shown that site-specific evolutionary rate heterogeneity (“heterotachy”) can be a major cause of phylogenetic error. We have developed a mixed branch length model which incorporates heterotachy by allowing sites to evolve over multiple sets of branch lengths on a single topology. We have shown that this approach can dramatically improve phylogenetic accuracy when the correct number of branch length sets is incorporated in the model. In reality, however, this quantity is never known in advance. Here we evaluate the accuracy of several common model evaluation techniques for selecting mixed branch-length phylogenetic models. Using a simulation-based experimental approach, we show that the Akaike Information Criterion had the highest success in selecting the model with the correct number of branch length categories and was dramatically more accurate than the Bayes Information Criterion or the likelihood ratio test. Our findings suggest that a strategy consisting of the mixed branch-length model and AIC has the potential to improve the reliability of phylogenetic analysis.

Publications

Dent A Earl, Keith Bradnam, John St. John, Aaron Darling, Dawei Lin, Joseph Faas, Hung On Ken Yu, Buffalo Vince, Daniel R Zerbino, Mark Diekhans, Ngan Nguyen, Pramila Nuwantha, Ariyaratne Wing-Kin Sung, Zemin Ning, Matthias Haimel, Jared T Simpson, Nuno A Fonseca, Inanc Birol, T Roderick Docking, Isaac Y Ho, Daniel S

Rokhsar, Rayan Chikhi, Dominique Lavenier, Guillaume Chapuis, Delphine Naquin, Nicolas Maillet, Michael C Schatz, David R Kelly, Adam M Phillippy, Sergey Koren, Shiaw-Pyng Yang, Wei Wu, Wen-Chi Chou, Anuj Srivastava, Timothy I Shaw, J Graham Ruby, Peter Skewes-Cox, Miguel Betegon, Michelle T Dimon, Victor Solovyev, Petr Kosarev, Denis Vorobyev, Ricardo Ramirez-Gonzalez, Richard Leggett, Dan Maclean, Fangfang Xia, Ruibang Luo, Zhenyu L, Yinlong Xie, Binghang Liu, Sante Gnerre, Iain MacCallum, Dariusz Przybylski, Filipe J Ribeiro, Shuangye Yin, Ted Sharpe, Giles Hall, Paul J Kersey, Richard Durbin, Shaun D Jackman, Jarrod A Chapman, Xiaoqiu Huang, Joseph L Derisi, Mario Caccamo, Yingrui Li, David B Jaffe, Richard Green, David Haussler, Ian Korf, and Benedict Paten. Assemblathon 1: A competitive assessment of de novo short read assembly methods. *Genome Research*, September 2011.

William Q Gillis, John St John, Bruce Bowerman, and Stephan Q Schneider. Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. *BMC Evolutionary Biology*, 9(1):207, 2009.

Benedict Paten, Mark Diekhans, Dent Earl, John St John, Jian Ma, Bernard Suh, and David Haussler. Cactus Graphs for Genome Comparisons. *Research in Computational Molecular Biology*, 6044:410–425, 2010.

John St. John. Identifying single nucleotide shifts in nanopore signals. Poster Presentation at the Graduate Student Research Symposium 2010, UC Santa Cruz, California, May 2010.

John St. John and Richard E. Green. In search of rare alleles in short read re-sequencing data. Poster Presentation at the Biology of Genomes 2011, Cold Spring Harbor Labs, New York, May 2011.

John St. John, B. Kolaczowski, and Joseph W. Thornton. Selecting an accurate model of evolutionary rate heterogeneity. Poster Presentation at Integrating Evolution Development and Genomics 2008, Berkeley California, May 2008.

 **UPDATED**

October 9, 2011