Gregg Thomas

Email: grthomas@indiana.edu
Website: https://gwct.github.io

EDUCATION

Doctor of Philosophy, Informatics (Bioinformatics track)
Doctor of Philosophy, Ecology, Evolution, and Behavior   August, 2013 - present
Indiana University
Bloomington, IN

Master of Science in Bioinformatics   May, 2013
Indiana University, School of Informatics and Computing
Bloomington, IN

Bachelor of Science in Biology   May, 2010
Purdue University, School of Science
West Lafayette, IN

PUBLICATIONS

Thomas GWC, Wang RJ, Puri A, Harris RA, Raveendran, Hughes DST, Murali SC, Williams LE, Doddapaneni, Muzny DM, Gibbs RA, Abee CR, Galinski MR, Worley KC, Rogers J, Radivojac P, Hahn MW. 2018. Reproductive longevity predicts mutation rates in primates. *Current Biology*. 28(19):3193-3197.

Warren WC et al. 2018. Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nature Ecology and Evolution*. 2:669-679.

Schoville SD et al. 2018. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Scientific Reports*. 8(1931).

Palesch D et al. 2018. Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host. *Nature*. 553:77-81.

Thomas GWC, Ather SA, and Hahn MW. 2017. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Systematic Biology*. 66(6):1007-1018.

Thomas GWC, Hahn MW, and Hahn Y. 2017. The effects of increasing the number of taxa on inferences of molecular convergence. *Genome Biology and Evolution*. 9(1):213-221.

Warren WC, et al. 2015. The genome of the vervet (*Chlorocebus aethiops sabaeus*). *Genome Research*. 25(12):1921-1933.

Thomas GWC and Hahn MW. 2015. Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. *Molecular Biology and Evolution*. 32(5):1232-1236.

Foote AD, Liu Y, Thomas GWC, Vinař T, et al. 2015. Convergent evolution of the genomes of marine mammals. *Nature Genetics*. 47(3):272-275.

Neafsey DE, Waterhouse RM, et al. 2014. Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. *Science*. 347.

Montague MJ, et al. 2014. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *PNAS*. 111(48):17230-17235.

Carbone L, et al. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature*. 513:195-201.

Thomas GWC and Hahn MW. 2014. The human mutation rate is increasing, even as it slows. *Molecular Biology and Evolution*. 31(2):253-257.

Han MV, Thomas GWC, Lugo-Martinez J, and Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution*. 30(8):1987-1997.

PRESENTATIONS

Reproductive longevity predicts mutation rates in primates
Gregg Thomas
Population, Evolutionary, and
Quantitative Genetics Conference, Platform talk
Madison, WI   May 19, 2018

The evolution of the genes and genomes of 76 arthropod species
Gregg Thomas, Matthew Hahn, Stephen Richards
Evolution Meeting, Regular talk
Portland, OR   June 26, 2017

The evolution of the genes and genomes of 76 arthropod species
Gregg Thomas, Matthew Hahn, Stephen Richards
Arthropod Genomics Symposium, Invited talk
South Bend, IN   June 9, 2017

Gene-tree reconciliation with MUL-trees for polyploidy analysis
Gregg Thomas, S. Hussain Ather, Matthew Hahn
Evolution Meeting, Regular talk
Austin, TX   June 19, 2016

Accounting for Sequencing Error in Phylogenetics
Gregg Thomas
Society of Systematic Biologists, Lightning talk
Ann Arbor, Michigan   May 21, 2015

Inferring Molecular Convergence from Genomic Data

Gregg Thomas, Matthew Hahn
Midwest Ecology and Evolution Conference, Contributed talk
Indiana University, Bloomington, Indiana   March 28, 2015

Convergent Evolution of the Genomes of Marine Mammals
Gregg Thomas, Andrew Foote, Marine Mammal Genome Consortium, Matthew Hahn
Society for Molecular Biology and Evolution, Contributed talk
San Juan, Puerto Rico   June 12, 2014

EMPLOYMENT HISTORY
Research Assistant
Hahn Lab, Ecology, Evolution, and Behavior Department
Indiana University, Bloomington, IN   January 2012 - present
- Wrote a program (caferror) which estimates assembly and annotation error in genome assemblies using CAFE 3
- Performed gene family analysis in several genome projects, including cats, mosquitoes, various primates, fish, and birds of paradise
- Collaborated with scientists from other institutions on a project to discover the amount of molecular convergence in the phenotypically convergent marine mammals
- Studied the effects of null model choice and sample size on inferences of convergent evolution
- Established a novel model of mutation rate evolution in mammals given recent observations from whole genome sequencing of human families
- Sequenced several families of owl monkeys (*Aotus nancymaae*) and found that observed mutation rate variation in primates is actually due to sampling families at different ages
- Developed a method and accompanying software package (GRAMPA) to correctly infer the placement and mode of polyploidy events on a phylogeny
- Led part of the Insect 5000 genomes pilot project (i5K) tasked with performing comparative analysis on 27 newly sequenced insect genomes along with 49 previously sequenced ones, including inferring a large species phylogeny for insects and the largest gene family analysis to date
- Wrote the program Referee to assign quality scores to every position in a genome assembly

Associate Instructor
Information Infrastructure II (I211), School of Informatics and Computing
Indiana University, Bloomington, IN   August 2015 - May 2016
- Led and organized a lab to teach students advanced programming skills using the Python language
- Assigned as lead Associate Instructor in a large lecture class tasked with helping students code in group activities and monitoring their progress, attendance, and absences
- Guided the other Associate Instructors in their duties in the class

- Advanced the students beyond the syntax learning of basic programming and into advanced topics such as algorithms, regular expressions, interacting with the Web (HTML and CGI), and databases (SQL)

Associate Instructor
SNP Discovery and Population Genetics (I590), School of Informatics and Computing
Indiana University, Bloomington, IN   August 2014 - December 2014
- Led a lecture to teach graduate students basic programming skills using the Python language
- Met with students weekly to aid in completion of their programming assignments
- Helped students integrate population genetics methods and programming skills to write programs to perform specific tasks, such as calculating nucleotide diversity, detecting positive selection, calculating linkage disequilibrium, and quality filtering

Associate Instructor
Information Representation (I308), School of Informatics and Computing
Indiana University, Bloomington, IN   August 2011 - May 2012
- Taught students about binary representation and operations and basic database skills in a lab environment
- Interacted and aided students with their assignments and projects on a one-on-one basis
- Assisted the professor with creating and grading assignments, projects, and tests

ACTIVITIES

Undergraduate Mentor
Indiana University, Bloomington, IN   August 2015 - present
- Taught basic programming, research, and scholarship skills to three undergraduate students
- Assigned tasks relevant to current ongoing projects in the lab to the students
- Met weekly with the students to discuss the projects and read research papers

Jim Holland Summer Science Research Program Mentor
Indiana University, Bloomington, IN   July 2014
- Mentored and taught a student in a program designed to introduce underrepresented minority high school students to scientific research
- Designed a project to assess divergence time estimates within the scope of the one week program and the capabilities of my student
- Helped the student design and present their work during the program's poster session

Graduate Student Advisor
Indiana University Bioinformatics Club
Indiana University, Bloomington, IN   January 2012 - 2014
- Served as a founding member of the club and its first treasurer
- Kept track of funds and purchases through the club's Student Organization Account

- Facilitated club elections of officers
- Organized bi-weekly club meetings where current events in bioinformatics are discussed
- Guided the officers and members in planning club activities

AWARDS
Genetics, Cellular, and Molecular Sciences Training Grant
Department of Biology
Indiana University, Bloomington, IN   2014 - 2015

SOFTWARE & OTHER CONTENT
Referee: Reference genome quality scores
https://gwct.github.io/referee
- This software uses genotype likelihoods from reads mapped back to their assembly to calculate a quality score for every position in an assembly

Drosophila 25 species phylogeny
http://dx.doi.org/10.6084/m9.figshare.5450602
- As part of a larger project, I inferred the phylogeny of 25 *Drosophila* species and published it standalone on FigShare as a resource for others to use

GRAMPA: Gene-tree Reconciliation Algorithm with MUL-trees for Polyploid Analysis
https://gwct.github.io/grampa.html
- Given a singly-labeled species topology and a set of corresponding gene-trees, this software can infer if any whole genome duplications have occurred and, if so, infer the mode of polyploidization and the placement on the phylogeny

i5K Phylogenomics Website
https://cgi.soic.indiana.edu/~grthomas/i5k-web/main.html
- Given the vast amount of data involved in the i5K pilot project, I developed this website to organize and share the data and results with colleagues

GWCT: Genome-Wide Convergence Tester
https://github.com/gwct/gwct
- Software written to count convergent, divergent, and unique substitutions in amino acid data

caferror
https://hahnlab.github.io/CAFE/
- Part of CAFEv3.1, I wrote this program to utilize CAFE's new error correction function to estimate (via pseudo-maximum likelihood) genome assembly and annotation error