

CURRENT APPOINTMENT

Bioinformatics Scientist

Faculty of Arts and Sciences
Harvard University, Cambridge, MA

2021 – present

EDUCATION

Doctor of Philosophy*	Bioinformatics and Evolutionary Biology	2019
Indiana University Bloomington, IN		
Master of Science	Bioinformatics	2013
Indiana University Bloomington, IN		
Bachelor of Science	Biology	2010
Purdue University West Lafayette, IN		

* = University-wide Distinguished Ph.D. Dissertation Award

PUBLICATIONS

** = undergraduate mentee

1. **Thomas GWC**, Wang RJ, Nguyen J**, Harris RA, Raveendran M, Rogers J, Hahn MW. Origins and long-term patterns of copy-number variation in rhesus macaques. (2020). *Molecular Biology and Evolution*. msaa303.
2. Sun C, ..., **Thomas GWC**, ..., Mueller RL. Genus-wide characterization of bumblebee genomes reveals variation associated with key ecological and behavioral traits of pollinators. (2020). *Molecular Biology and Evolution*. msaa240.
3. Wang RJ, **Thomas GWC**, Raveendran M, Harris RA, Doddapaneni H, Muzny DM, Capitanio JP, Radivojac P, Rogers J, Hahn MW. (2020). Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. *Genome Research*. 30:826-834.
4. **Thomas GWC**, Dohmen E, Hughes ST, Murali SC, Poelechau M, Glastad K, ..., Chipman AD, Waterhouse RM, Bornberg-Bauer E, Hahn MW, Richards S. (2020). The genomic basis of Arthropod diversity. *Genome Biology*. 21(15).

5. Bentz AB, **Thomas GWC**, Rusch DB, Rosvall KA. (2019). Tissue-specific expression profiles and positive selection analysis in the tree swallow (*Tachycineta bicolor*) using a *de novo* transcriptome assembly. *Scientific Reports*. 9:15849.
6. **Thomas GWC** and Hahn MW. (2019). Referee: reference assembly quality scores. *Genome Biology and Evolution*. 11(5):1483-1486.
7. Rogers J, ..., **Thomas GWC**, ..., Jolly CJ, Gibbs RA, Worley KC. 2019. The comparative genomics and complex population history of *Papio* baboons. *Science Advances*. 5(1).
8. Da Lage J-L, **Thomas GWC**, Bonneau M, Courtier-Orgogozo V. 2019. Evolution of salivary glue genes in *Drosophila* species. *BMC Evolutionary Biology*. 19(36).
9. Prost S, Armstrong EE, Nylander J, **Thomas GWC**, Suh A, Petersen B, Dalen L, Benz BW, Blom MPK, Palkopoulou E, Ericson PGP, Irestedt M. 2019. Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. *GigaScience*. 8(5).
10. **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.
11. Warren WC, García-Pérez R, ..., **Thomas GWC**, ..., Schartl M. 2018. Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nature Ecology and Evolution*. 2:669-679.
12. Schoville SD, Chen YH, ..., **Thomas GWC**, ..., Richards S. 2018. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Scientific Reports*. 8(1931).
13. Palesch D, Bosinger SE, ..., **Thomas GWC**, ..., Silvestri G. 2018. Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host. *Nature*. 553:77-81.
14. **Thomas GWC**, Ather SA**, and Hahn MW. 2017. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Systematic Biology*. 66(6):1007-1018.
15. **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.
16. Warren WC, ..., **Thomas GWC**, ..., Freimer NB. 2015. The genome of the vervet (*Chlorocebus aethiops sabaeus*). *Genome Research*. 25(12):1921-1933.
17. **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.
18. Foote AD, Liu Y, **Thomas GWC**, Vinař T, ..., Gibbs RA. 2015. Convergent evolution of the genomes of marine mammals. *Nature Genetics*. 47(3):272-275.

19. Neafsey DE, Waterhouse RM, ..., **Thomas GWC**, ..., Besansky NJ. 2014. Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. *Science*. 347.
20. Montague MJ, ..., **Thomas GWC**, ... Warren WC. 2014. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. *Proc Natl Acad Sci USA*. 111(48):17230-17235.
21. Carbone L, ... **Thomas GWC**, ..., Gibbs RA. 2014. Gibbon genome and the fast karyotype evolution of small apes. *Nature*. 513:195-201.
22. **Thomas GWC** and Hahn MW. 2014. The human mutation rate is increasing, even as it slows. *Molecular Biology and Evolution*. 31(2):253-257.
23. 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

1. **Pedigree sequencing and mutation rate variation in primates**
Meeting of the American Association of Biological Anthropologists, virtual
Invited talk 2021
2. **Patterns of Molecular Evolution in Arthropods**
Arthropod Genomics Symposium, virtual
Invited talk 2020
3. **Reproductive longevity predicts mutation rates in primates**
Population, Evolutionary, and Quantitative Genetics Conference, Madison, WI
Platform talk 2018
4. **The evolution of the genes and genomes of 76 arthropod species**
Evolution Meeting, Portland, OR
Regular talk 2017
5. **The evolution of the genes and genomes of 76 arthropod species**
Arthropod Genomics Symposium, Notre Dame University, South Bend, IN
Invited talk 2017
6. **Gene-tree reconciliation with MUL-trees for polyploidy analysis**
Evolution Meeting, Austin, TX
Regular talk 2016
7. **Accounting for sequencing error in phylogenetics**
Society of Systematic Biologists, University of Michigan, Ann Arbor, MI
Lightning Talk 2015

8. **Inferring molecular convergence from genomic data**
Midwest Ecology and Evolution Conference, Indiana University, Bloomington, IN 2015
Contributed talk
9. **Convergent evolution of the genomes of marine mammals**
Society for Molecular Biology and Evolution, San Juan, Puerto Rico 2014
Contributed talk

RESEARCH EXPERIENCE

Postdoctoral research associate

Laboratory of Jeffrey Good 2019 – 2021
Division of Biological Sciences
University of Montana, Missoula, MT

- Lead a comparative project to study molecular evolution and phylogenetics in murine rodents using whole exome sequences from 210 species and whole genome sequences from dozens of species.
- Devised best-practices for assembly and annotation of a large sample of exomes.
- Update and maintain software released by the lab to automate reference-guided genome assembly through iterative mapping (pseudo-it).
- Analyze the phylogenetic relationships of newly sequenced rodent genomes using the reference genomes and genetic resources from the well annotated mouse and rat systems to build an empirical landscape of phylogenetic discordance across chromosomes.
- Applied for grants (NIH NRSA) to expand whole genome sampling of murine rodents to study phylogenetic discordance, patterns of molecular evolution, convergent evolution, and sex chromosome rearrangements.
- Administrator of lab's computational resources, including two 32 core 200GB servers and one 192TB NAS, and the lab github account.
- Mentor and guide graduate and undergraduate students in the lab regarding computational methods in genomics, phylogenetics, and molecular evolution.

Research Assistant

Laboratory of Matthew Hahn 2012 – 2019
School of Informatics, Computing, and Engineering
Department of Biology
Indiana University, Bloomington, IN

- Developed a method to estimate genome assembly and annotation error from gene count data using CAFE's error model function (cafererror).
- Studied patterns of convergent evolution in marine mammals and echolocating mammals and devised best practices for identifying molecular convergence.
- Devised a method to infer the presence and mode of polyploidy from gene tree topologies (GRAMPA).
- Modeled and observed mutation rate patterns in primates, including single nucleotide mutations and structural variants, by sequencing families of owl monkeys and macaques.
- Led the comparative phylogenetic portion of the i5K pilot project which involved analyzing the genomes of 76 arthropods.

- Wrote software to annotate genomes with quality scores (Referee).
- Participated in several collaborations by performing comparative analyses, such as phylogeny reconstruction and assessment, gene family analysis, and positive selection scans.

TEACHING EXPERIENCE

Instructor

Conservation Genetics and Population Genomics course (ConGen) 2020
 Virtual course
 University of Montana

- Gave keynote lecture on genome sequencing and assembly.
- Designed hands-on activities for a 2-hour workshop on genome assembly and read mapping.
- Met one-on-one with students during office hours to discuss and give advice about their data.

Student Mentor

School of Informatics, Computing, and Engineering, 2014 – 2019
 Department of Biology
 Indiana University, Bloomington, IN

Provided guidance to high school and undergraduate students in conceptualizing evolution by involving them in various computational projects, providing a basis in programming, data analysis, and scholarship.

- CEWiT Research Experience for Undergraduate Women 2018-2019
- Computer Science Independent Study 2017-2018
- Computer Science Independent Study 2016-2017
- Jim Holland Summer Science Research Program 2014

Teaching Assistant

School of Informatics, Computing, and Engineering, 2011 – 2016
 Department of Biology
 Indiana University, Bloomington, IN

Taught lab sessions, led class discussions, graded assignments, and met with students individually to assist them.

- INFO-I211: Information Infrastructure 2014, 2016
- BIOL-Z620/INFO-I590: SNP Discovery and Population Genetics 2014
- INFO-I308: Information Representation 2011–2012

PROFESSIONAL SERVICE

Graduate Student Advisor

Indiana University Bioinformatics Club 2012–2014

Indiana University, Bloomington, IN

Served as a co-founding member and treasurer (2012 only) to raise awareness of bioinformatics and associated opportunities for undergraduate and graduate students by facilitating group projects and discussions, tours, and social events.

Peer Review

- *G3* (2x)
- *PLoS ONE*
- *Molecular Biology and Evolution* (2x)
- *New Phytologist*
- *Pacific Symposium of Biocomputing* (2x)
- *Genes*
- *Nature Communications*
- *Systematic Biology* (2x)
- *Evolution Letters*
- *Science Advances*
- *Genome Biology and Evolution* (3x)
- *Society of Systematic Biologists Graduate Student Research Award*
- *Molecular Ecology Resources*
- *Molecular Ecology*

GRANTS

Genetics, Cellular, and Molecular Sciences Training Grant

Department of Biology
Indiana University, Bloomington, IN

2014–2015

Sandy Ostroy Summer Research Award for Undergraduates

Department of Biology
Purdue University, West Lafayette, IN

2008

AWARDS

Distinguished Ph.D. Dissertation Award

The University Graduate School
Indiana University, Bloomington, IN

2020

SOFTWARE & RESOURCES

Pseudo-it: Pseudo-genome assembly with iterative mapping

<https://github.com/goodest-goodlab/pseudo-it>

- This software iteratively maps reads to generate a pseudo-assembly to reduce reference bias. I re-wrote this software to modularize it and speed it up.

ConGen 2020 genome assembly workshop

<https://gwct.github.io/congen/>

- I built this website as a workshop resource for students during the Conservation Genomics Course.

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 the assembled genome.

***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://arthrofam.org>

- With the vast amount of data involved in the i5K pilot project, I developed this website to organize and share the phylogenetic and comparative results with colleagues.

GWCT: Genome-Wide Convergence Tester

<https://github.com/gwct/gwct>

- Software written to count convergent, divergent, and unique substitutions in sequence data.

cafererror

<https://hahnlab.github.io/CAFE/>

- Part of CAFE version 3, I wrote this program to use CAFE's error modeling function to estimate genome assembly and annotation error.