|  |  |
| --- | --- |
| **Gregg Thomas** | [greggwct@gmail.com](mailto:greggwct@gmail.com)  <https://gwct.github.io> |

|  |  |
| --- | --- |
| **EDUCATION** |  |
| Doctor of Philosophy, Informatics (Bioinformatics track)  Doctor of Philosophy, Ecology, Evolution, and Behavior  Indiana University  Bloomington, IN | July, 2019 |
| Master of Science in Bioinformatics  Indiana University  Bloomington, IN | May, 2013 |
| Bachelor of Science in Biology  Purdue University  West Lafayette, IN | May, 2010 |

|  |  |  |
| --- | --- | --- |
| **PUBLICATIONS** | |  |
|  | **Thomas GWC**, Dohmen E, Hughes ST, Murali SC, Poelechau M, Glastad K, …, Chipman AD, Waterhouse RM, Bornberg-Bauer E, Hahn MW, Richards S. (in review). The genomic basis of Arthropod diversity. | |
|  | Bentz A, **Thomas GWC**, Rusch DB, Rosvall KA. (in review). *De novo* transcriptome assembly of the tree swallow (*Tachycineta bicolor*): tissue-specific expression profiles and positive selection analysis. | |
|  | **Thomas GWC** and Hahn MW. (in press). Referee: reference genome quality scores. *Genome Biology and Evolution*. | |
|  | 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). | |
|  | 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. | |
|  | 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*. | |
|  | **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, 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. | |
|  | 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). | |
|  | 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. | |
|  | **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, …, **Thomas GWC**, …, Freimer NB. 2015. The genome of the vervet (*Chlorocebus aethiops sabaeus*). G*enome 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, …, Gibbs RA. 2015. Convergent evolution of the genomes of marine mammals. *Nature Genetics*. 47(3):272-275. | |
|  | Neafsey DE, Waterhouse RM, …, **Thomas GWC**, …, Besansky NJ. 2014. Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. *Science*. 347. | |
|  | 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. | |
|  | Carbone L, … **Thomas GWC**, …, Gibbs RA. 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**  Population, Evolutionary, and Quantitative Genetics Conference, Madison, WI  Platform talk | May 19, 2018 |
|  | **The evolution of the genes and genomes of 76 arthropod species**  Evolution Meeting,Portland, OR  Regular talk | June 26, 2017 |
|  | **The evolution of the genes and genomes of 76 arthropod species**  Arthropod Genomics Symposium,Notre Dame University, South Bend, IN  **Invited talk** | June 9, 2017 |
|  | **Gene-tree reconciliation with MUL-trees for polyploidy analysis**  Evolution Meeting,Austin, TX  Regular talk | June 19, 2016 |
|  | **Accounting for sequencing error in phylogenetics**  Society of Systematic Biologists,University of Michigan, Ann Arbor, MI  Lightning Talk | May 21, 2015 |
|  | **Inferring molecular convergence from genomic data**  Midwest Ecology and Evolution Conference,Indiana University, Bloomington, IN  Contributed talk | March 28, 2015 |
|  | **Convergent evolution of the genomes of marine mammals**  Society for Molecular Biology and Evolution,San Juan, Puerto Rico  Contributed talk | June 12, 2014 |

|  |  |
| --- | --- |
| **RESEARCH EXPERIENCE** |  |
| ***Research Assistant***  Laboratory of Matthew Hahn  School of Informatics, Computing, and Engineering  Department of Biology  Indiana University, Bloomington, IN | 2012 – present |
| * Developed a method to estimate genome assembly and annotation error from gene count data using CAFE’s error model function (caferror). * Studied patterns of convergent evolution in marine mammals and echolocating mammals. * 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 arthopods. * 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** |  |
| ***Student Mentor***  School of Informatics, Computing, and Engineering  Department of Biology  Indiana University, Bloomington, IN | 2014 –2019 |
| 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.   * Jelena Nguyen, Indiana University: CEWiT Research Experience for Undergraduate Women (Fall 2018 to Spring 2019). * Arthi Puri, Indiana University: Computer Science Independent Study (Fall 2017 to Spring 2018). * S. Hussain Ather, Indiana University: Computer Science Independent Study (Spring 2016 to Spring 2017). * Nana Addo, Indiana University: Jim Holland Summer Science Research Program (Summer 2014). | |
| ***Teaching Assistant***  School of Informatics, Computing, and Engineering  Department of Biology  Indiana University, Bloomington, IN | 2011 – 2016 |
| Taught lab sessions, led class discussions, graded assignments, and met with students individually to assist them.   * INFO-I211: Information Infrastructure (Fall 2014, Spring 2016). * BIOL-Z620/INFO-I590: SNP Discovery and Population Genetics (Fall 2014). * INFO-I308: Information Representation (Fall 2011, Spring 2012). | |

|  |  |
| --- | --- |
| **PROFESSIONAL SERVICE** |  |
| ***Graduate Student Advisor***  Indiana University Bioinformatics Club  Indiana University, Bloomington, IN | 2012 –2014 |
| 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. | |
| ***Reviewer***   * Molecular Biology and Evolution * New Phytologist * Pacific Symposium on Biocomputing, 2019 * PLoS One | |

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

|  |  |
| --- | --- |
| **SOFTWARE** |  |
| **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://i5k.gitlab.io/ArthroFam](https://i5k.gitlab.io/ArthroFam/)   * 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. | |
| **caferror**  <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. | |