ALFRED SIMKIN, PHD

alfred.simkin@gmail.com
www.boxwoodbioinformatics.org
www.github.com/simkin-bioinformatics
www.github.com/alfredsimkin
(508) 925-0354

SUMMARY

I am a research professor and independent consultant with extensive bioinformatics and teaching experience in academia and some experience in industry. I'm looking for opportunities to do freelance bioinformatics consulting which would ideally include both training of public health researchers and building analysis pipelines for partners in industry and academia. I'm also interested in full-time employment in industry or at a bioinformatics core within academia. I enjoy helping people find algorithmically efficient, reproducible, portable, and scalable solutions for bioinformatics problems. I also enjoy learning new skills and training and mentoring others in programming and basic bioinformatics techniques. During my PhD and postdoctoral work, I developed a Python programming bootcamp course with examples drawn from bioinformatics and wrote a simulator of neutral evolutionary processes in small RNA pathways. I gained extensive teaching and wet-lab experience during 4.5 years as a tenure track genetics professor at Elon University, during which I managed an undergraduate genetics lab course. In other roles, I also worked as a bioinformatics research professor running a departmental bioinformatics core at University of Massachusetts Chan Medical School, and worked in industry as a staff scientist at IDEXX in the genomics R&D group. In my current position as a remote assistant research professor at Brown University (physically located in Raleigh, North Carolina), I am working with a large consortium of public health researchers to develop software packages and training students in programming and data analysis techniques. Due to recent funding cuts, my position is now part time and I have much more time available for consulting. My bioinformatics consulting business is Boxwood Bioinformatics, which I started with my brother and business partner, Charlie Simkin, in 2022. I am currently consulting for Deus Ishengoma of the Tanzanian National Institute for Medical Research and Travis Thomson of University of Massachusetts Medical School.

EMPLOYMENT

INDEPENDENT CONSULTANT, SIMKIN BIOINFORMATICS, RALEIGH, NC — Nov 2022-present

BROWN UNIVERSITY (REMOTE), *Providence, RI* — Nov 2022-present
Assistant Research Professor, IDEEL consortium (Pathology department)

IDEXX (REMOTE), *Westbrook, ME* — Feb 2022-Oct 2022
Staff scientist 1, Computational Biology (Genomics R&D group)

UMASS CHAN MEDICAL SCHOOL, *Worcester, MA* — March 2021-Jan 2022
Assistant Research Professor of Bioinformatics (Neurobiology Dept)

ELON UNIVERSITY, *Elon, NC* — Aug 2016-Jan 2021
Assistant Professor of Biology (Tenure track)

CORNELL UNIVERSITY, *Ithaca, NY* Aug 2014-July 2016
Postdoctoral Researcher, (Advisor: Andrew William Grimson)

UMASS CHAN MEDICAL SCHOOL, *Worcester, MA* July 2009-July 2014
PhD student (Advisors: Fen-Biao Gao and Jeffrey D. Jensen)

PUBLICATIONS

Martin AC, Sadler JM, **Simkin A**, Musonda M, Katowa B, Matoba J, Schue J, Simulundu E, Bailey JA, Moss WJ, Juliano JJ, Fola AA. Emergence and Rising Prevalence of Artemisinin Partial Resistance Marker Kelch13 P441L in a Low Malaria Transmission Setting in Southern Zambia. Under review (Journal of Infectious Disease).

Popkin-Hall ZR, Carey-Ewend K, Aghakhanian F, Oriero EC, Seth MD, Kashamuka MM, Ngasala B, Ali IM, Mukomena ES, Mandara CI, Kharabora O, Sendor R, **Simkin A**, Amambua-Ngwa A, Tshefu A, Fola AA, Ishengoma DS, Bailey JA, Parr JB, Lin JT, Juliano JJ. Population Genomics of Plasmodium malariae from Four African Countries. Under review (Nature Microbiology).

M'Angale PG, Lemieux A, Liu Y, Wang S, Zinter M, **Simkin A**, Budnik V, Kelch BA, and Thomson T. 2025. Capsid transfer of the retrotransposon Copia controls structural synaptic plasticity in Drosophila. *PLoS Biology* (in press).

Sadler JM*, **Simkin A***, Tchuenkam VPK, Gyuricza IG, Fola AA, Wamae K, Assefa A, Niaré K, Thwai K, White SJ, Moss WJ, Dinglasan RR, Nsango S, Tume CB, Parr JB, Ali IM, Bailey JA, Juliano JJ. 2025. Application of a new highly multiplexed amplicon sequencing tool to evaluate Plasmodium falciparum antimalarial resistance and relatedness in individual and pooled samples from Dschang, Cameroon. *Frontiers in Parasitology*.
*** indicates co first-authorship**

Carey-Ewend K, Popkin-Hall ZR, **Simkin A**, Muller M, Hennelly C, He W, Moser KA, Gaither C, Niaré K, Aghakanian F, Feleke S, Brhane BG, Phanzu F, Kashamuka MM, Aydemir O, Sutherland CJ, Ishengoma DS, Ali IM, Ngasala B, Kalonji A, Tshefu A, Parr JB, Bailey JA, Juliano JJ, Lin JT. 2024. Population genomics of Plasmodium ovale species in sub-Saharan Africa. *Nature Communications*.

Juliano JJ, Giesbrecht DJ, **Simkin A**, Fola AA, Lyimo BM, Pereus D, Bakari C, Madebe RA, Seth MD, Mandara CI, Popkin-Hall ZR, Moshi R, Mbwambo RB, Niaré K, MacInnis B, Francis F, Mbwambo D, Garimo I, Chacky F, Aaron S, Lusasi A, Molteni F, Njau RJA, Nhiga SL, Mohamed A, Bailey JA, Ishengoma DS. 2024. Prevalence of mutations associated with artemisinin partial resistance and sulfadoxine-pyrimethamine resistance in 13 regions in Tanzania in 2021: a cross-sectional survey. *Lancet Microbe*

Popkin-Hall ZR, Niaré K, Crudale R, **Simkin A**, Fola AA, Sanchez JF, Pannebaker DL, Giesbrecht DJ, Kim IE Jr, Aydemir Ö, Bailey JA, Valdivia HO, Juliano JJ. 2024. High-throughput genotyping of Plasmodium vivax in the Peruvian Amazon via molecular inversion probes. *Nature*

Communications.

Simkin A, McIntyre A, Grimson A. 2020. Evolutionary Dynamics of microRNA target sites across vertebrate evolution. *Plos Genetics*

Geissler R, **Simkin A**, Floss D, Patel R, Fogarty EA, Scheller J, Grimson A. 2016. A widespread sequence-specific mRNA decay pathway mediated by hnRNPs A1 and A2/B1. *Genes and Development*

Simkin A, Gao F-B, Bailey JA, Jensen JD. 2014. Inferring the Evolutionary history of primate microRNA binding sites: overcoming motif counting biases. 2014. *Molecular Biology and Evolution*

Simkin A, Wong A, Poh Y-P, Theurkauf W, Jensen JD. 2013. Recurrent and Recent Selective Sweeps in the piRNA pathway. *Evolution*.

Crisci J, Poh YP, Bean A, **Simkin A**, Jensen JD. 2012. Recent progress in polymorphism-based population genetic inference. *Heredity*.

Yuva-Aydemir Y, **Simkin A**, Gascon E, Gao F-B. 2011. MicroRNA-9: functional evolution of a conserved small regulatory RNA. *RNA Biology*

Poh Y-P, Crisci J, **Simkin A**, Bean A, Jensen JD. 2011. Statistical inference in population genetics. In the book *Evolutionary Developmental Biology*. Intech Open Access Publishers. Rijeka, Croatia.

TEACHING EXPERIENCE

- Bio245, Principles of Genetics (lab and lecture) (11+13 sections over 3.5 years)
- Bio111 Intro. to Cell Biology 111 (lecture) (three sections over one year)
- Bio113 Intro. to Cell Biology 113 (Lab) (two sections in one semester)
- Bio445A Advanced Genetics (lecture and lab) (taught twice, two years apart)
- Introduction to Python mini-course (taught 5 times over 10 years, University of Massachusetts Chan Medical School, Cornell University, Brown University)

BIOINFORMATICS SKILLS

Coding Approach

I try to minimize dependencies in my code, write compact code, and use informative variable names. I enjoy collaborative coding, and I strive to be constantly learning new things. I have some experience automating the creation of virtual machines on AWS and getting them scaled to the correct size. I enjoy coding workflows using Snakemake, which allows me to automate pipelines, appropriately scale individual steps of my workflow in slurm environments, log program run-times, organize my thinking, and visualize workflows.

Coding Languages

Python (extensive, current)
Perl (moderate, several years ago)
R (cursory, mostly with existing packages, graphing)
C++ (basic, many years ago)

Other types of Coding Experience
Unix command line
Docker and Singularity containerization
package managers (pip, conda, apt)
Plotly
JupyterLabs
github and git command line
Ubuntu
Snakemake
AWS
Sysadmin (server configuration, port forwarding, basic Apache configuration)

Major past projects:

- Creating pipelines for African researchers to generate VCF files from Nanopore and Illumina DNA sequencing datasets and convert these into amino acid prevalence tables (ongoing, volunteer work).
- converting MIPTools (a large collection of C++ code, python scripts, shell scripts, and jupyter notebooks for calling haplotypes and drug resistance mutations) into a more memory-efficient, more customizable, less error prone, and easier to run snakemake pipeline.
- Developing pipelines for visualizing genetic variants in targeted amplicon DNA sequencing datasets (in this case using a set of PCR primers that we dubbed Pf-SMARRT).
- Setting up a scalable snakemake workflow for creating AWS virtual machines that are appropriately sized for each step of a pipeline whose steps each have their own unique resource requirements.
- Setting up labs (Umass, IDEXX, Brown University) with private UCSC genome browsers and populating them with data tracks.
- Extracting sequences from multiz files and UCSC tracks to build comparative genome alignments and phylogenies.
- measuring the rates of motif turnover in simulated 3' UTRs.
- measuring the strength of natural selection with programs like PAML, Phylip, and clsw
- quantification of biases associated with RNAseq reads with tools like Polyester
- Finding and performing appropriate statistical tests for associations between motifs and observed gene regulatory shifts.
- Implementation of kmer based search strategies to quickly search genetic sequences (both my own implementations and through programs like blat and lastz)

Illustrative Example programs

- converting a parallelized standalone single slurm cluster job into individual job submissions in parallel using singularity environments.
- Re-writing binomial distribution calculators to be more accurate than R implementations
- sorting large text files that are too large to fit in computer memory
- using sorted files to convert n^2 comparisons of complex lists into 2n time.
- calculating population genetics statistics
- memory efficient conversions between fasta, fastq, and phylip data formats
- functions for reading parameters from and writing parameters to control files for auto documentation.
- A pairwise sequence aligner that approximates multiple sequence alignments in ways that are more informative than MAFFT and MUSCLE in some circumstances.

<u>Aspirational projects</u>:
GPU parallelization - I would like to learn to write code that leverages GPU processors for highly parallel processing with CUDA-compatible GPUs.
Expanded statistical approaches - I would like to find coding problems requiring maximum likelihood, bayesian statistics, and markov chains as a way of understanding these at a deeper level.
machine learning/AI - I have a bit of experience with developing random forest models but would like to take on some projects that require development of these skills

COLLABORATIVE SKILLS

I am currently involved in collaborations with dozens of researchers across the globe in Tanzania, Ethiopia, Ghana, Uganda, Imperial College London, UCSF, Johns Hopkins, UNC Chapel Hill, and Brown University, where I take on a mix of training and software development roles. I have extensive experience interfacing with molecular biologists and fellow programmers. I've trained several undergraduate and graduate students in programming, sending several of the undergraduate students for advanced degrees in bioinformatics. I've also managed a genetics stockroom, training 9 students in molecular biology bench techniques, over a 4 year span. Over the course of 8 months at Umass medical school, I taught a postdoc Python and helped him convert all of his code from matlab to Python, trained a technician to translate biological problems into Python functions, taught Python to ~25 learners, and created an active Python discussion group. At IDEXX, I trained an intern, managed a junior research scientist, developed a random forest model, and led a team of computational researchers in developing AWS pipelines.

COMMUNITY OUTREACH/HOBBIES

I've recently gotten very into 3D printing. I also tinker with solar power, electric bicycles, Raspberry Pi microcomputers, and other energy/computer projects. I’ve built and configured several ssh Linux servers and enjoy tinkering with network connectivity with things like port forwarding, wake on LAN, ftp file servers, jump hosts, and reverse ssh.

People I've helped train in public health bioinformatics techniques

UNC Chapel Hill:

Isabela Gerdes Gyuricza
Kelly Cary-Ewend
Chris Henelly - now training DRC Paluseq team on my methodologies
Ronald Futila - now training DRC Paluseq team on my methodologies
Jacob Sadler
Ashenafi Assefa
Ruthly Francois
Sam White
Wenqiao He
Farhang Aghakhanian
Julia Muller
Emma Sanders

Rwanda:
Lambert Me
Gashema Pierre

Tanzania:
Dativa Pereus
Catherine Bakari

Brown University:
Abebe Fola
Neeva Wernsman-Young
George Tollefson
Jenna Zuromski
Titus Maina

Uganda:
Thomas Katairo
Francis Ddumba Semakuba
Bienvenu Nsengimaana

UCSF:
Oriana Kreutzfeld
Melissa Conrad
Abhishek Kumar

Ethiopia:
Melak Getu
Rajiha Abubeker

Kenya:
Kevin Wamae

Ghana:
Lucindah Fry-Narty

Burkina Faso:
Nassandba Yanogo

Elon RESEARCH MENTEES (BIO 499 research course unless otherwise noted)

Thomas Wilson	(F20)
Morgan Cagarli	(F20)
Joseph Urban	(F20)
Grace Carter	(Lumen 498 F18, S19, F19, SURE Su19, S20)
Keeley Collins	(SURE Su17, F17, F18, F19)
Emily Cooper	(F17, S18)
Margo Greenawald	(S18)
Emily Hughes	(Su18, S19)
Kayla Ervin, Tristan Scarce, Daniel Castillo	(Su18)

RESEARCH GRADUATES

Grace Carter	Broad Institute, Postbaccalaureate research, F20
Keeley Collins	University of Kansas, Computational Biology PhD program, F19
Emily Cooper	University of Colorado, Biostatistics PhD program, F19
Emily Hughes	Boston University, Bioinformatics MA program, F19

REFERENCES

Jeffrey A Bailey (my funding PI currently, at Brown)
Associate Professor of Pathology and Laboratory Medicine
Brown University
jeffrey_bailey@brown.edu

Pradeep Sathyanarayana (manager when I was at IDEXX)
Manager R&D Sr I, Genomics
IDEXX
pradeep-sathyanarayana@idexx.com

Travis Thomson (funding PI when I was a research professor at Umass)
Assistant Professor of Neurobiology
University of Massachusetts Medical School
travis.thomson@umassmed.edu

Antonio Izzo (department chair when I was at Elon)
Associate Professor of Biology
Elon University
aizzo@elon.edu

Jeffrey D Jensen (PhD advisor when I was a PhD student at Umass)
Professor of Population Genetics
Arizona State University
jeffrey.jensen@asu.edu