ALFRED SIMKIN

alfred.simkin@gmail.com
(508) 925-0354

SUMMARY

I am a computational biologist with PhD and postdoctoral experience modeling the statistical parameters of simulated evolutionary processes in small RNA pathways, and experience analyzing RNA and DNA sequencing datasets (Illumina, PacBio, and Nanopore). I have 4.5 years of teaching experience as a tenure track genetics professor, where I also coordinated a genetics lab and extended my familiarity with many basic wet-lab techniques. When the pandemic started, I began searching for positions that would give me an opportunity to do more cutting edge computational biology research and mentor committed computational students. I spent 10 months as a bioinformatics research professor running a departmental bioinformatics core, 9 months working as a staff scientist at IDEXX, and am currently a research professor at Brown University, working with a consortium of malaria and infectious disease researchers that includes a large Chapel Hill contingent (which is where I'm physically located). I also have a consulting business, Simkin Bioinformatics, LLC, that I run with my business partner, Charlie Simkin. 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.

EMPLOYMENT

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)

EDUCATION

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

UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL, *Worcester, MA*   July 2014
PhD in Molecular Evolution, (Advisors: Jeffery Jensen and Fen-Biao Gao)

EARLHAM COLLEGE, *Richmond, IN*   May 2009
Bachelor of Arts in Biology, Phi Beta Kappa

PRESENTATIONS

16th annual teaching and learning conference at Elon University   August 2019
        Incorporating Metacognitive Awareness into STEM Courses

to Build Motivation and Resilience in Problem-solving
Elon, North Carolina

20th annual Evolutionary Biology Meeting   Sep. 2016
    Recurrent Selective Sweeps in piRNA Pathway Proteins, and an
    Alternative to Classic 'Arms Race' Dynamics
    Marseilles, France

Society for Molecular Biology and Evolution Annual Meeting   June 2014
    Inferring the Evolutionary History of Primate microRNA Binding
    Sites: Overcoming Motif Counting Biases.
    Puerto Rico, USA

Boston Evolutionary Genomics 2012 Annual Retreat   August 2012
    miRNA-Target Interactions: Is it Possible to be Conserved Through Drift?
    Boston, MA

Society for Molecular Biology and Evolution Annual Meeting   July 2011
    Patterns of Recurrent and Recent Selective Sweeps within piRNA
    Pathway Proteins in the Host/Transposon Genomic Conflict.
    (Travel Fellowship Award Recipient)
    Kyoto, Japan

PUBLICATIONS

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.

POSTERS

SURF 2018 Faculty Poster   April 2018
    Alfred Simkin. The ever-changing genome: understanding
    the hidden forces and parasites that allow genomes to shrink,
    grow, and adapt to new circumstances
    Elon, North Carolina

Society For Molecular Biology and Evolution Annual Meeting   June 2013
    Alfred Simkin, Jeffrey A. Bailey, Fen-Biao Gao and Jeffrey
    Jensen. A Revised Approach to Motif Turnover Rates Reveals
    Inherent Bias Toward Gains
    Chicago, Illinois

RiboClub annual meeting   September 2012
    Alfred Simkin, Fen-Biao Gao, and Jeffrey Jensen.
    miRNA Target Sites Appear to be Evolving Neutrally.
    (Travel Fellowship Award Recipient)
    Quebec, Canada

Society For Molecular Biology and Evolution Annual Meeting   June 2012
    Alfred Simkin, Fen-Biao Gao, and Jeffrey Jensen.
    miRNA Target Sites Appear to be Evolving Neutrally.
    Dublin, Ireland.

52nd annual Drosophila Research Conference   March 2011
    Alfred Simkin, William Theurkauf, and Jeffrey Jensen.
    Searching for Signatures of an "Evolutionary Arms Race" between
    transposons and piRNAs.
    San Diego, CA.

University of Massachusetts Medical School Woods Hole annual retreat   October 2010
    Alfred Simkin, William Theurkauf, and Jeffrey Jensen.
    Searching for Signatures of an "Evolutionary Arms Race" between
    transposons and piRNAs.
    Woods Hole, MA.

57th annual American Society of Tropical Medicine and Hygiene meeting   December 2008
    Will Vincent, **Alfred Simkin** and Peter Blair.
    A bioinformatic approach to refining the annotations of *P. yoelii.*
    New Orleans, LA.

TEACHING EXPERIENCE
- COR110 The Global Experience (first year humanities seminar) (1 section F20)
- Bio245, Principles of Genetics (lecture) (1 section S17, 2 sections F17, 1 section S18, 2 sections F18, 2 section F19, 1 section S20, 2 sections F20)
- Bio246 Principles of Genetics (lab) (2 sections S17, 3 sections F17, 3 sections F18, 3 sections F19, 1 section S20, 1 section F20)
- Bio105 Reconstructing Dinosaur DNA (1 section W19)
- Bio111 Intro. to Cell Biology 111 (lecture) (2 sections F16, 1 section S17)
- Bio113 Intro. to Cell Biology 113 (Lab) (2 sections F16)
- Bio445A Advanced Genetics (lecture) (1 section S18, 1 section S20)
- BioL445A (lab) (1 section S18, 1 section S20)
- Created and taught a 12 hour 'mini-course' to teach Cornell students Python (Fall 2015)
- Created and taught six rounds of a twelve hour introductory programming course structured around simple bioinformatics exercises in the Python language, (January 2014, February 2014, April 2014, July 2014, and September 2015)
- Founded the 'bootstrappers' to create graduate student taught courses for graduate students at the University of Massachusetts Medical School (May 2013)
- University of Massachusetts Medical School Sequence Analysis Bootcamp Teaching Assistant (Spring 2014)

BIOINFORMATICS SKILLS

Coding Approach

I like to write my own versions of existing packages, both as a way of understanding them and extending their functionality. 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. Lately I've become enamored with Snakemake, and enjoy using this in slurm environments as a way of automating pipelines, appropriately scaling individual steps of my workflow, logging program run-times, organizing my thinking, and visualizing workflows.

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

Major past projects:
- Setting up a scalable snakemake workflow for creating and 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) with private UCSC genome browsers and populating them with data tracks.
- Extracting sequences from multiz files and UCSC tracks to build 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 two 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.

Aspirational projects:
collaborative coding with git - I would like to to code collaboratively more, with version control.
GPU parallelization - I would like to learn to use more integrated GPU solutions for parallel processing.
workflow management - I use snakemake extensively, and would like to start using nextflow.
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.
code isolation I have used Docker and Conda moderately but am attempting to build my own environments and reduce dependencies.
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 have extensive experience interfacing with molecular biologists and fellow programmers. I've made several close bioinformatician friends over the years. I've taught a Python workshop to ~250 learners at 4 institutions, and I've mentored several undergraduate students in programming, bringing them from no background to extensive bioinformatics skills. I've also managed a genetics stockroom, with 3 TAs

working under me in molecular biology techniques, over a 4 year span (training a total of ~9 students in these techniques). 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, created an active Python discussion group, and taught two grad students to intersect UCSC genes from BED files with coordinates of interest. At IDEXX, I trained an intern, developed a random forest model with CS engineers, and led a team of computational researchers in developing AWS pipelines.

COMMUNITY OUTREACH/HOBBIES

Triad electric vehicle association, unofficial bioinformatics algorithms club at Elon, wind turbine building workshop, ukulele club, tectonic plates. I also tinker with solar power, wood heat, electric bicycles, Raspberry Pi microcomputers, and other sustainability 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.

COMMITTEES

Elon Faculty Research and Development Committee (Chair)	ASY 2020/21
Elon Faculty Research and Development Committee	ASY 2019/20
Elon Train committee	ASY 2018/19
Elon Biology Department Curriculum Committee	ASY 2018/19
Science for non-majors tenure track search committee	ASY 2017/18
Elon Phi Beta Kappa member	ASY 2016/18
Elon Sustainability Fellow	ASY 2016/17
Data Analytics Fellow	ASY 2017/18

SERVICE

Science Unsolved seminar organization group	F19, S20
Majors Mondays (Biology majors coffee)	S19, F19, S20, F20
Genetics Lab coordinator	F17-F20
co-advisor for STEM residence life	F18-F20
Engaged learning sessions (first year orientation)	F17, F18
NCUR (national conferences on undergraduate research) faculty participant	S18
Undergraduate Programming discussion group	F17, F18, F19
MakerHub project ("Make Gadgets with Raspberry Pi and Pizza Pie")	F17
Maker Takeover	S17, S18, S19
STEM Workshops for teachers workshop ("Introducing Students to the world of programming and building things with the Raspberry Pi")	Su17

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