

SIMON URIBE, PH.D.

SAN FRANCISCO, CA

uribe.convers@gmail.com

Google Scholar: [goo.gl/Hg7De4](https://scholar.google.com/citations?user=Hg7De4)

github.com/uribe-convers

simonuribe.com

[linkedin.com/in/simon-uribe](https://www.linkedin.com/in/simon-uribe)

PROFILE

Data scientist specializing in bioinformatics and evolutionary genetics with over 10 years of experience in scientific research and data analysis. Effective communicator and relationship builder who enjoys working in cross-functional teams. Skilled in Agile/Scrum software development with a strong focus on user-driven design. Extensive experience designing and building pipelines and applications to handle high-throughput sequence data (NGS) from multiple sources (amplicon, shotgun, hybridization capture, GBS/RADSeq) and platforms (Illumina: MiSeq, NextSeq, HiSeq).

SKILLS

Programming Languages: Python, SQL, Bash, R.

Software Development: Git, GitHub, Jupyter, ETL, AWS (EC2, RDS, S3), PostgreSQL, SQLAlchemy, Pycopg2, Requests. Familiar with AI algorithms (neural nets, random forests) and NLP in PyTorch.

Bioinformatics: GATK, Bioconductor, IGV, FastQC, MultiQC, BLAST, BLAT, samtools, bedtools, VCFtools, SPAdes, Bowtie2, BWA, Trinity, Cutadapt, Trimmomatic, DADA2, USEARCH, CD-HIT, Primer3.

Evolutionary Biology: phylogenomics, population genetics, sequence alignment, MAFFT, RAxML, MrBayes, BEAST, PAUP*, ASTRAL, Geneious, STRUCTURE.

Molecular Biology: DNA extraction, DNA quantification, library prep, primer design, PCR, cloning, microfluidic PCR, hybridization capture probe design, transcriptomics, metagenomics, Sanger and Illumina sequencing and QC.

Management: Early-stage startups, hiring, strategic planning, Agile, Scrum, Jira, Airtable, scientific writing.

Languages: English (fluent), Spanish (native), German (fluent).

EDUCATION

Ph.D. Evolutionary Biology, University of Idaho, Moscow, ID, USA 2009 - 2014

B.Sc. Biology (Honors Thesis), Universidad de los Andes, Bogota, Colombia 2001 - 2008

PROFESSIONAL EXPERIENCE

DATA SCIENTIST

Phylagen; San Francisco, CA

Feb 2018 - Present

- o Designed, built, and maintained pipelines and applications (Python, SQL, Bash, R, AWS) to process, curate, analyze, and visualize NGS data. These applications were used by the lab, commercial, and informatics teams.
- o Wrote and maintained API integrations (Python) with the LIMS, NGS, and sample metadata databases (PostgreSQL).
- o Led bioinformatics operations by routinely processing NGS data while monitoring and reporting quality control metrics. This included collaborating with the lab team to troubleshoot difficult samples or sequencing runs.
- o Built a bioinformatic pipeline (Python, Bash, AWS) to analyze a large (~800Gb) metagenomic dataset to identify candidate biomarkers. Identified >1,100 genomic targets, designed PCR primers in Primer3, and coordinated with the lab to prioritize targets according to product requirements.
- o Managed computing infrastructure and information systems locally (Linux) and in the cloud (AWS: EC2, RDS, S3).
- o Led a weekly Scrum meeting to prioritize software development to meet the lab team's requirements for LIMS, monitoring data quality control, and providing data science (Python: pandas, matplotlib) support for R&D projects.

BIOINFORMATICS POSTDOCTORAL FELLOW

University of Michigan; Ann Arbor, MI

Jul 2017 - Jan 2018

- o Built the largest phylogeny to date for all fungi with publicly available sequence data (~30,000 species).
- o Mined online repositories (GenBank) for publicly available data with Python scripts.
- o Developed bioinformatic pipelines (Python, Bash) to analyze large genomic and phylogenomic datasets.
- o Studied wood evolution on a phylogeny containing all species of plants (~350,000 species).

BIOINFORMATICS POSTDOCTORAL FELLOW

University of Missouri; St. Louis, MO

Oct 2015 - Jul 2017

- o Led experiments and analyses of multiple types of NGS data (shotgun libraries, hybridization capture, and transcriptomes). This included designing probes for hybridization capture, de novo and reference-based assemblies, genomic annotation, and evolutionary and phylogenomic analyses.
- o Generated 500 times the amount of data available for a quarter of the original cost.
- o Designed, built, and executed analytic models and statistical analyses for a number of diverse projects.
- o Supervised and mentored personnel working in molecular laboratory and greenhouse projects.

RESEARCH ASSOCIATE

Missouri Botanical Garden, St. Louis, MO

Oct 2015 - Jul 2017

- o Investigated the genomic differences among cultivated and wild species of grapes to identify novel genetic variation for future breeding efforts.
- o Led bioinformatic and evolutionary analyses of genotyping by sequencing data with thousands of variant sites.

GRADUATE RESEARCH ASSISTANT

University of Idaho, Moscow, ID

Aug 2009 - Aug 2014

- o Developed and published two novel approaches to generate large genomic datasets quickly and cost effectively using either microfluidic PCR or long-PCR. These methods are now widely used in the field of phylogenomics.
- o Led a wide range of projects in genomics, phylogenetics, and macroevolution.
- o Authored several publications and disseminated the results in presentations at international conferences.
- o Formulated and authored grant proposals that completely funded my Ph.D.
- o Taught the Dendrology laboratory, teaching students how to identify over 200 species of trees.

RESEARCH ASSISTANT

Johannes Gutenberg University, Mainz, Germany

Aug 2004 - Jul 2005

- o Designed and performed molecular biology experiments with AFLP and Sanger sequence data. This included DNA extraction, amplification, ligation, and sequencing.
- o Studied the phylogenetic and hybridization histories of a group of plants in northeastern Africa.
- o Co-authored a scientific publication as an undergraduate student.

ADDITIONAL TRAINING

AWS Certified Solutions Architect Associate, A Cloud Guru (in progress)	2019
Fundamentals of Deep Learning using PyTorch, University of San Francisco, San Francisco, CA	2019
Data Scientist with Python Career, DataCamp	2017
Python for Genomic Data Science, Coursera	2017
Bioconductor for Genomic Data Science, Coursera	2017
Bodega Bay Applied Phylogenetics Workshop, University of California, Davis, CA	2010

SELECTED PUBLICATIONS

-
- McKain MR, Johnson MG, **Uribe-Convers S**, Eaton D, Yang Y. 2018. Practical considerations for plant phylogenomics. *Applications in Plant Sciences* 6 (3): e1038. **All authors contributed equally.*
- Uribe-Convers S**, Carlsen M, Lagomarsino LP, Muchhala N. 2017. Phylogenetic relationships of *Burmeistera* (Campanulaceae): Combining whole plastome with targeted loci data to improve resolution in a recent radiation. *Molecular Phylogenetics and Evolution* 107: 551-563.
- Uribe-Convers S**, Settles ML, Tank DC. 2016. A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the South American species of *Bartsia* L. (Orobanchaceae). *PLoS ONE*. 11(2): e0148203. doi:10.1371/journal.pone.0148203.
- Uribe-Convers S**, Duke JR, Moore JM, Tank DC. 2014. A long-PCR based method for chloroplast genome enrichment and phylogenomics in angiosperms. *Applications in Plant Sciences* 2: 1300063. Featured in *EurekAlert!* (goo.gl/C4rAeh) and in *ScienceDaily* (goo.gl/zdFuZ4).