SEMIDÁN ROBAINA, Ph.D.

BIOINFORMATICS SCIENTIST | COMPUTATIONAL BIOLOGIST

github.com/Robaina
linkedin.com/semidan-robaina
contact@semidanrobaina.com
0034-658-865-742 (Tenerife, Spain)

Sr. Bioinformatics Scientist with extensive expertise in

Programming (Python, R, JavaScript), Data Analysis and Visualization, Machine Learning, Project Management, and Communication.

With a solid foundation in computational biology and data science, my career has taken me from kinetic modeling of metabolic systems to in-depth analysis of genome-scale metabolic networks. Prior to my doctorate, I focused on metabolic engineering of bacterial strains. During my PhD, I identified key biological features in genome-scale metabolic networks and developed constraint-based algorithms to incorporate omics data into these models. As a postdoc, I honed my expertise in NGS data analysis, with a particular focus on metagenomic and metatranscriptomic datasets. Throughout these roles, I've built a strong proficiency in scientific programming, particularly within the R and Python ecosystems, equipping me with the tools to drive innovation and efficiency in my work. Fluent in Spanish and English, I bring a collaborative approach to global, diverse teams, underpinned by continuous learning and adaptability.

SKILLS

Tools: **Python**: deployment (pip, **conda**, **docker, codespaces**), **Jupyter** Notebooks, **Google Colaboratory**, pandas, scikit-learn, matplotlib, cobrapy | **R**: RMarkdown, **Tidyverse** (dplyr, ggplot2, …), **DeSeq2** | Bioinformatics: **BASH**, **samtools**, bwa, seqkit, **AlphaFold2** | **Git** & GitHub, **GitHub Actions**

Technical: Experimental design | Data analysis | Data Engineering | Technical reports | Reproducible pipelines | Analytical Modeling | Convex Optimization | Network Analysis | Machine Learning | Logistic regression | Clustering | NGS data analysis | Software development | Automation (CI/CD) | Agile Project Management | GNU/Linux

Transversal: Excellent Communication | Project Management | Problem-Solving | Team Leadership | Mentorship | Resilience to failure | Excellent Ability to Learn and Adapt | Excellent Organizational and Planning skills

EXPERIENCE

Sr. Bioinformatics Scientist | University of La Laguna, Spain | 2021 Mar - present

- Led the **development and deployment** of two **Python** packages to automate function and taxonomy annotation, and analysis of **metagenomic** data, saving my team hundreds of hours of manual labor
- **Taught** Git, GitHub, and Python pipelines to a team of 1 Ph.D. student, 2 technicians, and 2 postdocs without previous experience in the topic
- Led end-to-end **statistical pipelines** of NGS data in BASH and **R** (DeSeq2, gene set enrichment, clustering) to reveal environmental effects on metabolic regulation of key marine prokaryotes, resulting in three scientific publications
- **Mentored** two undergrad students throughout their thesis projects, resulting in a successful thesis publication (one ongoing)

Data Scientist | Ronin Institute - Freelance | 2019 Mar - 2021 Feb

- Led conceptualization and development of a Python pipeline applying **network analysis** and **convex optimization** techniques to uncover novel metabolic network constraints on biochemical reaction fluxes in Escherichia coli, resulting in a scientific publication
- Developed a **Front-End Web** application (HTML5, JavaScript, CSS3, jQuery, Cytoscape.js) to visualize and interact with biological network data resulting in a published web page

- Led the development of **Python** pipelines to analyze and **visualize** demographic data
- Established a **data back-end** through Python and the **Google Sheets API** to automatically update databases deployed with **Power BI**

Sr. Bioinformatics Researcher | University of Potsdam, Germany | 2017 Oct - 2019 Jan

- **Led procurement** of scientific software, saving my team several thousand euros by increasing the share of open-source software alternatives in research pipelines
- Developed **SQL**-empowered **Python** scripts to automate **data wrangling** and analysis of large-scale **multi-omics molecular databases**, saving my team hundreds of hours of manual labor in data collection
- **Managed** course syllabi and taught genome-scale mathematical modeling techniques within the Master in Bioinformatics of the University of Potsdam
- **Mentored** a master's student throughout his thesis project, resulting in a published master's thesis

Research Scientist | Max Planck Institute, Germany | 2013 Oct - 2017 Sep

- Led conceptualization and development of a **MATLAB** tool based on **convex optimization** (Mixed Integer Quadratic Program) and l1 regularization (LASSO) to automate the reconstruction of **context-specific metabolic networks** from multi-omics molecular data sets, resulting in two scientific publications
- Identified key biochemical reactions for maize growth using **network analysis, mathematical modeling,** and multi-omics data sets, resulting in a scientific publication
- Identified critical metabolic pathways involved in plant stomata regulation through **mathematical modeling** and ^{13}C **carbon flux experiments**, resulting in a scientific publication

Research Analyst | University of La Laguna, Spain | 2012 Jun - 2013 Jun

- Led the conceptualization and development of a **mathematical (ODE) model** of *E. coli*'s metabolism, using metabolite **time series** data analysis and **genetic algorithms** to optimize growth in a biotechnological setting, resulting in a master thesis

CERTIFICATIONS

Making Data Science Work for Clinical Reporting - Genentech. Issued March 2023

Good Clinical Practice - The National Institute on Drug Abuse (NIDA). Issued Jan 2023

Getting Started with SAS Programming. SAS. Issued Jan 2023

Databases and SQL for Data Science with Python. IBM. Issue Oct 2022

Google Project Management Specialization. Google. Issued Aug 2022

Google Data Analytics Specialization. Google. Issue Jul 2022

Deep Learning Specialization. DeepLearning.AI. Issue Oct 2022

EDUCATION

Ph.D. in Systems Biology (Magna cum laude) | 2013 - 2017
Max Planck Institute of Molecular Plant Physiology and University of Potsdam, Germany

M. Ed. Teacher Training for Compulsory Secondary Education and Baccalaureate | 2020 - 2021
University of La Laguna, Spain

M. Sc. in Biotechnology | 2012 - 2013
University of La Laguna, Spain

Licenciate in Molecular and Cell Biology (5 academic years) | 2006 - 2012
University of La Laguna, Spain

LANGUAGES

Spanish (native), English (C1-level certification)

PUBLICATIONS

Transcriptional mechanisms of thermal acclimation in Prochlorococcus. *mBio*. Alonso-Sáez L., Palacio A., Cabello A., **Robaina-Estévez S.**, González J., Garczarek L., López-Urrutia A.

A ubiquitous gammaproteobacterial clade dominates expression of sulfur oxidation genes across the mesopelagic ocean. *Nature Microbiology.* 2023. Baltar F., Martínez-Pérez C., Amamo C., Vial M. **Robaina-Estévez, S.**, Reinthaler T., Zhao Z., Logares R., Herndl G., Morales S. González J.

Pynteny: a Python package to perform synteny-aware, profile HMM-based searches in sequence databases. *JOSS* (2023). **Robaina-Estévez S.,** González J.

Flux-based hierarchical organization of *Escherichia coli*'s metabolic network. *Plos Comp. Biol (*2020). **Robaina-Estévez S.**, Nikoloski Z.

Thesis: Context-specific metabolic predictions: computational methods and applications. Universität Potsdam (2017). **Robaina-Estévez S.** https://publishup.uni-potsdam.de/frontdoor/index/index/docId/40136

On the effects of alternative optima in context-specific metabolic model predictions. PLOS Computational Biology (2017). **Robaina-Estévez S**., Nikoloski Z.

Resolving the central metabolism of Arabidopsis guard cells. Scientific Reports (2017). **Robaina-Estévez S**., Daloso D.M., Zhang Y, Fernie A.R., Nikoloski Z.

Metabolic network constraints gene regulation of C 4 photosynthesis: the case of maize. PCB (2016). **Robaina-Estévez S**., Nikoloski Z.

Context-specific metabolic model extraction based on regularized least squares optimization. PLOS ONE (2015). **Robaina-Estévez S**., Nikoloski Z

Generalized framework for context-specific metabolic model extraction Methods. Frontiers in Plant Science (2014). **Robaina-Estévez S**., Nikoloski Z.