RICARDO VALENCIA ALBORNOZ

☎ +447596810069 rvalenciaaz ✉ r.g.valencia-albornoz@sms.ed.ac.uk rvalenciaaz 🌐 rvalenciaaz.github.io

SUMMARY

Scientist with a robust foundation in Bayesian optimisation, probabilistic machine learning, bioinformatics of microbial communities, laboratory automation, bioprocesses, metabolomics and genomics. Proven to be a strong independent thinker, pioneering AI for biology and chemistry projects at Edinburgh, but also a flexible team player with interdisciplinary soft and hard skills such as research communication, cybersecurity and quantum technologies. Passionate about building cutting-edge platforms combining high-throughput biology experiments and synthetic biology tools with state-of-the-art machine learning pipelines.

EDUCATION

University of Edinburgh | PhD Quantitative Biology, Biochemistry & Biotechnology 2020 - 2024

PhD project: Bayesian optimisation methods for media optimisation of low-titer bacterial strains (Finishing September 2024)

- Visualization and complex modelling (Gaussian processes, graphical models) for high-throughput metabolomics data using Python and R. Generative AI for decision making in microbe metabolic experiments. Robotic projects are showcased in rvalenciaaz.github.io/portfolio. Supervisors: Prof. Karl Burgess & Dr. Diego Oyarzún

Technical University Federico Santa Maria | BSc Environmental Engineering 2012 - 2018
6-year course equivalent to BSc + MSc on chemical engineering and sustainable chemical and biological applications

Modules: Biotechnology, Chemical Engineering, Unit Operations, Electronics, Environmental Law & Policy, Data Analytics, Project Management

Technical University Federico Santa Maria | Preliminar Programme for Engineering 2011 - Dec 2012
Pre-Bachelor

Modules: Calculus and physics corresponding to the first year curriculum in engineering courses. Attending university lectures while in high-school

EXPERIENCE

University of Edinburgh | Intern, Digital Research Ambassador 2024 April 2024 - July 2024

- (UK) Providing support in statistical analysis, advance modelling and visualization on admission data for medical schools at the Usher Institute, University of Edinburgh.

University of Edinburgh | Demonstrator January 2021 – April 2021

- (UK) Teaching assembly and annotation of bacterial genomes using spades, prokka and RAST. Interpretation of genomic functional data in prokaryotes and eukaryotes using data from InterProScan and EggNog.

Technical University Federico Santa Maria | Research Assistant January 2019 - April 2020

- (Chile) Bioinformatics for the search of natural products in marine strains of *Actinomycetota* phylum. 4 scientific publications in high-quality journals. Phylogeny and diversification of biosynthetic gene cluster mega-enzymes.

Technical University Federico Santa Maria | Research Assistant February 2018 – October 2018

- (Chile) Development of methods for modelling spatio-temporal metagenomic data of hydrocarbon-degradation microbial communities (kriging, auto correlation, community networks). Dynamic of microbial communities at contaminated sites.

Biotecnos | Intern 2016 - 2017

- (Chile) Molecular biology protocols (DNA extraction, PCR) for the detection of bacterial genes involved in hydrocarbon degradation. Orchestration of a manual to perform Risk-Based Corrective Action (RBCA).

PUBLICATIONS

- **Valencia Albornoz R**, O'Loughlin J, Oyarzún D, Burgess K. 2024. Metabolic trade-off discovery using active learning and large language models. In preparation.
- **Valencia Albornoz R**, Oyarzún D, Burgess K. 2024. Optimisation of surfactin yield in *Bacillus* using active learning and high-throughput mass spectrometry. Computational and Structural Biotechnology Journal. doi: 10.1016/j.csbj.2024.02.012.
- **Valencia R**, González V, Undabarrena A, Zamora-Leiva L, Ugalde JA, Cámara B. 2021. An integrative bioinformatic analysis for keratinase detection in marine-derived Streptomyces. Marine Drugs 19:286. doi: 10.3390/md19060286
- Undabarrena A, **Valencia R**, Cumsille A, Zamora-Leiva L, Castro-Nallar E, Barona-Gomez F, Cámara B. 2021. Rhodococcus comparative genomics reveals a phylogenomic-dependent non-ribosomal peptide synthetase distribution: insights into biosynthetic gene cluster connection to an orphan metabolite. Microbial Genomics 7. doi: 10.1099/mgen.0.000621
- Undabarrena A, Salvà-Serra F, Jaén-Luchoro D, Castro-Nallar E, Mendez KN, **Valencia R**, Ugalde JA, Moore ERB, Seeger M, Cámara B. 2018. Complete genome sequence of the marine Rhodococcus sp. H-CA8f isolated from Comau fjord in Northern Patagonia, Chile. Marine Genomics 40:13–17. doi: 10.1016/j.margen.2018.01.004
- Orellana R, Macaya C, Bravo G, Dorochesi F, Cumsille A, **Valencia R**, Rojas C, Seeger M. 2018. Living at the frontiers of life: extremophiles in Chile and their potential for bioremediation. Frontiers in Microbiology 9:2309. doi: 10.3389/fmicb.2018.02309

AWARDS

Darwin Trust of Edinburgh Fellow (PhD Scholarship) 2020 – 2024
Full fee coverage and stipend through PhD course

Runner-up Business Ideas Competition, Edinburgh Innovations 2023
A small fund to develop a business idea (automatic strain selection with robotics)

Federico Santa María award to best graduate 2018
Best graduate BSc Environmental Engineering, class of 2018

PRESENTATIONS

- **16-18 Sept 2024** (Upcoming) 7th RSC-BMCS/RSC-CICAG Artificial Intelligence in Chemistry (Poster). Title: LLM-based preferential Bayesian optimisation on a multi-output kinetic dataset. Cambridge, UK.
- **22 Apr 2024**. Invited to attend OECD Global Forum on Technology: "Building our Biofuture: Policy issues and opportunities for next generation biotechnologies" on the first day of Ministerial Meeting. Paris, France.
- **8-11 Apr 2024**. Microbiology Society Annual Conference 2024 (Talk). Section: Microbial Physiology, Metabolism and Molecular Biology. Session talk on learning metabolism from optimisation. Edinburgh, UK.
- **11-15 Mar 2024**. The Automated Scientist – the future of cell factory engineering (Poster). Novo Nordisk Science Cluster. Favrholm, Denmark.
- **14-16 June 2023**. Darwin Trust Symposium 2023 (Talk). Title: Optimisation of surfactin production in *Bacillus*. Edinburgh, UK.
- **13-14 March 2023**. AI, Engineering Biology & Beyond (Talk). Title: Optimisation of surfactin production in *Bacillus* using active learning and metabolomics. Edinburgh, UK.
- **2-4 November 2022**. IV Latin American Metabolic Profiling Society (LAMPS) Network Meeting (Talk). Title: A metabolomic perspective on bioprocess optimisation. Cartagena, Colombia.

COURSES/ADDITIONAL TRAINING

- **15-19 July 2024** (Upcoming) Accepted for participation at the Cambridge Ellis Unit Summer School on Probabilistic Machine Learning, University of Cambridge.

- **8 May 2024** Invited to attend Cohere Build Day in London. Mini-hackathon testing Cohere LLM models (Command-R+), RAG and agent systems. May 2024.
- **27-28 Mar 2024** Leader of the BOPE-GPT team at the Acceleration Consortium BO Hackathon. Our project integrates LLMs and preferential BO for automated utility function optimisation on a chemical kinetic dataset. Check our Github repo here.
- **24-28 Oct 2022** EMBO Practical Course on Metabolomics Bioinformatics for Life Scientists. Wageningen University and Research. Wageningen, The Netherlands.

Online learning

- Google Advanced Data Analytics Professional Certificate (7 courses). Coursera. Sept 2023; Google Cybersecurity Professional Certificate (8 courses) & IBM Cybersecurity Analyst Professional Certificate (8 courses). Coursera. Sept 2023.
- Quantum Optics 2 - Two photons and more & Introduction to Quantum Information. Coursera. Sept 2023; Quantum Optics 1 : Single Photons. Coursera. July 2020.

TECHNICAL SKILLS

- Bayesian optimisation pipelines using Python and R. Expertise using Botorch, GPyOpt and PyTorch, for custom applications in iterative chemical and biological experimentation, deep neural network hyperparameter tuning and controller tuning. Expertise in advanced Gaussian process algorithms for regression and classification, including deep GPs. Data visualization and statistical analysis of univariate and multivariate data in Python and R, including interactive visualization in matplotlib, seaborn, plotly, bokeh, holoviews, geoviews, and datashader for big data (basic charts, networks, geographic data).
- Automated control of robotic platforms, including Opentrons OT-2 and ChiBio. Building of custom-made robots for experimental applications. Proficiency in electronics theory and practice. Knowledge of PyLabRobot to interconnect labware, and Tensorflow Lite to deploy local ML for decision making. Pipeline consolidation using Apache Airflow. Expertise on AutoCAD Inventor and openSCAD (code-based) for medium complexity 3D modelling for PLA printing.
- Metabolomics: Data analysis of targeted (QqQ-MS) and untargeted data including ion mobility measurements (IM-QToF-MS). Operation of triple quadrupole mass spectrometer (Thermo TSQ Quantiva) and orbitrap equipment (Thermo Q-Exactive). Experience with high performance liquid chromatography (HPLC) systems (ThermoFisher systems).
- Experience in operation of minireactors and bench reactors for high-density culturing of bacteria (Bacillus) for lipopeptide production and CHO cells. Proficient in aseptic technique and bench lab manipulations. ODE and metabolic pathway modelling of bioreactor operation and optimisation of culture conditions using Bayesian optimisation. Proportional-integral-derivative (PID) controllers and model predictive control, applied to chemical engineering and bioprocesses.
- Solid expertise in molecular biology techniques including gel electrophoresis, PCR, DNA and RNA extraction, plasmid prep and protein/metabolite extraction. Electroporation protocols on Bacillus for synthetic gene circuit expression with integrative vectors (BacilloFlex)
- Genomics and metagenomics: Assembly and annotation of genomes from both long-read sequencing data (Pacbio, Nanopore) and short-read sequencing data (Illumina). Tools include canu, unicycler, spades, checkm for genome quality and prokka/bakta for annotation. SNP calling and analysis using breseq and pangenome with Roary. 16s amplicon sequence analysis using qiime2 and functional annotation of WGS using kraken.
- Genome mining: Extraction of gene features from massive datasets using bash and python scripts. Identification of biosynthetic gene clusters (BGC) in bacteria using antiSMASH. Network analysis of BGC using BiGSCAPE. Generation of biological networks using NetworkX and Gephi.
- Phylogeny: Tree construction using Bayesian (mrbayes, BEAST) and Maximum Likelihood (raxml, IQ-TREE) frameworks. Fitting of substitution models using PartitionFinder and modeltest-ng. Orthology analysis of bacterial genomes using OrthoFinder.