

Ricardo R. Pavan
Bioinformatics | Computational Biology | Data Analyst
ricrocha82@gmail.com | 380-250-9540 | Columbus, OH | EU Citizen
LinkedIn: <https://www.linkedin.com/in/ricardorochaps/>
GitHub: <https://github.com/ricrocha82>

PROFESSIONAL SUMMARY

Bioinformatician with PhD-level training in high-throughput data and applied statistics, specializing in statistical modeling, machine learning, pipeline development, and large-scale data analysis. Demonstrated expertise in study design, project management, data visualization, and advanced statistical methods, supported by a strong record of collaboration, peer-reviewed publications, and mentoring. Proven ability to serve as a technical consultant to investigators and stakeholders, translating questions into scalable data architectures and statistically rigorous analyses. Consistently ensures data integrity, clear documentation, and fully reproducible analytical workflows.

SKILLS

Applied Statistics: Generalized mixed-effects models, negative binomial models, differential expression, multivariate analysis, batch correction, clustering, study design, time-series analysis, deep learning, and machine learning.

Genomics & Biomedical Data Analysis: RNA-seq, single-cell & single-nucleus sequencing, genomics and metagenomics, high-dimensional omics integration

Programming & Data Analysis: R (tidyverse, caret, tidymodels, ggplot2), Python (pandas, scikit-learn), SQL, Unix/Linux bash scripting, Pipeline development (Nextflow, Snakemake), Deep Learning Libraries (TensorFlow, Keras, PyTorch)

Data Management and Computing: High-Performance Computing (HPC/SLURM), Git/GitHub, Conda, Docker, reproducible pipelines, Tapis Framework

Communication and Collaboration: manuscripts, technical reports, statistical tables, data visualization, investigator support, cross-disciplinary collaboration

PROFESSIONAL EXPERIENCE

Bioinformatician and Data Analyst

The Ohio State University | Columbus, OH, USA | May 2023 – Present

Managing human and environmental research projects. The role involved close liaison with students, researchers, and analysts to deliver high-impact outcomes on the development of new bioinformatic tools and biological knowledge.

- Actively maintained and extended a scalable, version-controlled bioinformatic pipeline, implementing standardized workflows and logging to ensure reproducibility and data integrity, while collaborating on cross-disciplinary microbiome studies that led to manuscripts under review
- Designed and developed an open-source pipeline distributed freely via GitHub for community use to annotate and compare ssDNA virus genomes
- Led statistical design and benchmarking analyses comparing short-read and long-read viral metagenomics workflows, quantifying effects on genome recovery, bias, and accuracy
- Led statistical analyses of longitudinal and cross-sectional environmental virome datasets, applying multivariate statistics, regression modeling, and time-series analyses to test hypotheses on temporal dynamics and environmental drivers
- Designed and executed statistical analysis plans for a human cohort study (82 pregnant people living with HIV) and a global meta-analysis (1,800+ samples), applying compositional data analysis, regression, non-parametric testing, batch correction, and differential abundance analysis to quantify microbiome-immune associations
- Co-leading the modernization, documentation, and deployment of viral bioinformatics applications on national cyberinfrastructure platforms, supporting NSF-funded pipelines, training, and future grant development.
- Prepare publication-ready tables, figures, and statistical summaries, contributing to 10+ manuscripts

Data Analyst and Machine Learning Specialist

University of Tasmania | Hobart, TAS, Australia | Nov 2018 – Apr 2023

Worked at the Institute for Marine and Antarctic Studies (IMAS), where I identified latent patterns and candidate indicators linked to external system disturbances. I used scalable data pipelines, rigorous validation, and interpretability to translate complex raw data into structured features and quantitative evidence suitable for decision-making in data-driven environments

- Designed statistical frameworks for observational and spatial studies to test whether microbial and protist communities are predictably structured by environmental and anthropogenic gradients in non-randomized systems
- Applied compositional and multivariate analyses (clustering, PCA, db-RDA, PERMANOVA, network analysis) to properly model sequencing-derived count data and quantify the proportion of community variance explained by environmental drivers
- Built and validated regression models (Random Forest and PLS) to assess the extent to which environmental variables explain biological outcomes, achieving R^2 up to 0.87 and demonstrating strong explanatory power
- Developed and evaluated classification models (Random Forest) to test whether community taxonomic and functional profiles could discriminate impacted versus reference conditions, achieving ROC-AUC up to 0.93
- Translated statistical analyses into peer-reviewed publications and technical reports, providing reproducible methods, effect size estimates, and uncertainty quantification for interdisciplinary audiences

EDUCATION

- Advanced Degree: PhD (Doctor of Philosophy) | University of Tasmania, Tasmania, Australia | Nov 2018 – Apr 2023
- Advanced Degree: Master of Business Administration (MBA) | Data Analysis and Statistics | University of São Paulo, SP, Brazil | May 2021 – Feb 2023
- Certificate: Programming Foundations: Fundamentals | Programming | LinkedIn | Feb 2024
- Certificate: Introduction to Career Skills in Software Development | Programming | LinkedIn | Feb 2024
- Certificate: The Complete SQL Bootcamp: Go from Zero to Hero | SQL | Udemy | Dec 2024

LANGUAGES

Portuguese (Native), English (Fluent), Spanish (Intermediate), Italian (Beginner)

PORTFOLIO

CRESENT: <https://github.com/ricrocha82/cresent>

Publications: <https://orcid.org/my-orcid?orcid=0000-0002-7878-2861>

ORCID: <https://orcid.org/0000-0002-7878-2861>