

João Vitor F. Cavalcante

BIOINFORMATICS DEVELOPER

✉ jvfecav@gmail.com | ☎ 0000-0001-7513-7376 | 🌐 jvfe | 📄 joao-vitor-cavalcante

About me

I'm a bioinformatician passionate about open science, free software, and making data analysis more accessible. My expertise lies in workflow development and large-scale data analysis, with a strong focus on reproducibility, automation, and user-friendly design. I have experience in Python and R for bioinformatics, NGS data analysis, and workflow management using Nextflow and Snakemake. Most of my work involves developing and maintaining bioinformatics pipelines, improving documentation and testing, and automating reports to enhance usability and reproducibility. Currently, besides being a researcher at Dalmolin Systems Biology Group (Brazil), where I work in projects involving metagenomics and scRNA-Seq data, I also collaborate with the Sanger Institute, as part of the BioDev Network, organizing Bioinformatics-related events and activities.

Experience

Future Innovators - Wellcome Sanger Institute

Remote

MENTEE

June 2024–December 2024

- As part of the first cohort of the Future Innovators program, based in the Sanger Institute, I've developed a project related to the nf-core toolkit, implementing features in this toolkit that facilitated development of in-house Nextflow pipelines at the Darwin Tree of Life project, such features not only helped enhance data processing, but increased potential collaboration with other organisations, within and beyond the Sanger Institute.
- | Python | Nextflow |

Anaconda Learning

Remote

INSTRUCTOR

August 2024–September 2024

- Developed and taught an advanced Git/GitHub course for the Anaconda Learning platform, focusing on concepts typically not covered in most Git courses, such as worktrees, reflow and Git-LFS.
- | Git | GitHub |

Dalhousie University

Remote

NEXTFLOW DEVELOPER

February 2023–March 2024

- Worked for Dr. Robert Beiko's lab in the Faculty of Computer Sciences. Developed and optimized Nextflow pipelines, primarily beiko-lab/ARETE, for SLURM HPC environments. Integrated automated testing with nf-test and GitHub Actions to ensure workflow reliability, improving scalability and reproducibility in large-scale bioinformatics analyses.
- | Nextflow | Docker | Singularity | Git | GitHub | HPC |

Freelance

Remote

BIOINFORMATICS CONSULTANT

July 2022–January 2023

- Re-implemented legacy pipelines using Nextflow and WDL, enhancing workflow efficiency and maintainability. Developed custom Python and R scripts to optimize data processing for various clients.
- | Python | R | Nextflow | WDL |

BioME/UFRN

Natal/Brazil

RESEARCHER

December 2018–ongoing

- Performed RNA-Seq, metagenomics, and genomics analyses with a focus on network and community analysis using R, Python, and Shell scripting. Designed and maintained Snakemake and Nextflow pipelines for research group projects.
- | RNA-seq/Metagenomics data analysis | Network analysis | R/Bioconductor | Data visualization | Snakemake | Nextflow |

R Project for Statistical Computing

Remote

GSoC STUDENT DEVELOPER

June 2021–August 2021

- Analyzed social network data from X (formerly Twitter) and Meetup, generating interactive dashboards to identify trends in the R community and optimize event organization and project focus.
- | R/Tidyverse | Dashboards and data visualization | GitHub Actions for CI/CD | R/Shiny |

Skills

Programming

PYTHON, R, JAVASCRIPT, BASH

Data Pipelines

NEXTFLOW, SNAKEMAKE, WDL

Web

VUE.JS, FLASK, MATERIALIZE, BOOTSTRAP

Query

SPARQL, SQL

Tools

GIT/GITHub, GitHub ACTIONS, SLURM, DOCKER, SINGULARITY

Software

ARETE

ANALYSIS PIPELINE FOR PROFILING THE GENOMIC REPERTOIRE AND EVOLUTIONARY DYNAMICS OF MICROORGANISMS

Nextflow

EURYALE

A PIPELINE FOR TAXONOMIC CLASSIFICATION AND FUNCTIONAL ANNOTATION OF METAGENOMIC READS

Nextflow

MicroView

A REPORTING TOOL FOR AGGREGATING RESULTS FROM TAXONOMIC CLASSIFICATION ANALYSES

Python

pytrials

PYTHON INTERFACE TO THE CLINICALTRIALS.GOV API

Python

go2cell

INTEGRATES GENE ONTOLOGIES (GO) WITH WIKIDATA CELL TYPE ITEMS

R

I frequently contribute to Open Source software projects, such as: nf-core/modules, pyp5js and BioPython. Other projects can be found in my GitHub.

Distinctions

Nextflow Ambassador

SEQERA'S OUTREACH PROGRAM FOR SUPPORTERS OF THE NEXTFLOW COMMUNITY

Future Innovators Mentee

SELECTED BY THE SANGER INSTITUTE FROM A GLOBAL COHORT OF YOUNG RESEARCHERS

Education

Federal University of Rio Grande do Norte

PHD IN BIOINFORMATICS

Natal

2024-ongoing

Federal University of Rio Grande do Norte

MSC IN BIOINFORMATICS

Natal

2023-2024

Federal University of Rio Grande do Norte

BACHELOR IN BIOMEDICAL SCIENCE

Natal

2018-2022

Publications

Selected publications. For the full list, see the [ORCID](#) record.

- **Complex Portal 2022: new curation frontiers.** 2021. <https://doi.org/10.1093/nar/gkab991>
- **BioProv - a provenance library for bioinformatics workflows.** 2021. <https://doi.org/10.21105/joss.03622>
- **MEDUSA: A pipeline for sensitive taxonomic classification and flexible functional annotation of metagenomic shotgun sequences.** 2022. <https://doi.org/10.3389/fgene.2022.814437>
- **Bridging the gaps in meta-omic analysis: Workflows and reproducibility.** 2023. <https://doi.org/10.1089/omi.2023.0232>
- **Revealing metastatic castration-resistant prostate cancer master regulator through lncRNAs-centered regulatory network.** 2023. <https://doi.org/10.1002/cam4.6481>
- **Sex-specific gene expression differences in the prefrontal cortex of major depressive disorder individuals.** 2024. <https://doi.org/10.1016/j.neuroscience.2024.09.012>