

João Vitor F. Cavalcante

BIOINFORMATICS DEVELOPER

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

About me

I'm a bioinformatician fascinated by open science and free software. I have experience with software development in Python and R, NGS (Next Generation Sequencing) data analysis, as well as experience in managing and developing workflows for large scale data processing in tools such as Nextflow and Snakemake. The bulk of my work has been focused towards workflow development and data analysis, with a focus on user-friendliness, through documentation and automated reports, customization and reproducibility. I'm a MSc researcher at BioME/UFRN, in Brazil, where I maintain bioinformatics pipelines and participate in systems biology research projects using metagenomics and RNA-Seq data. In my free time I like to make generative art, contribute to open source software and edit the free knowledge base [Wikidata](#).

Experience

Dalhousie University

NEXTFLOW DEVELOPER

Remote

February 2023–March 2024

- Worked for Dr. Robert Beiko's lab in the Faculty of Computer Sciences on the development of custom pipelines using Nextflow, primarily beiko-lab/ARETE, implementing them on SLURM HPC environments and guaranteeing their function through automated unit tests, using nf-test and GitHub Actions.
- | Nextflow | Docker | Singularity | Git | GitHub |

Freelance

BIOINFORMATICS CONSULTANT

Remote

July 2022–January 2023

- Worked with freelance bioinformatics projects in the field of workflow development, working with clients to re-implement old pipelines into newer technologies, such as Nextflow and WDL, as well as building Python and R scripts to enhance parts of the data processing.
- | Python | R | Nextflow | WDL |

BioME/UFRN

RESEARCHER

Natal/Brazil

December 2018–ongoing

- I have performed analyses in RNA-Seq, Metagenomics and genomics data, with a focus on network and community analysis, primarily using R, Python and Shell scripting. I have also extensively worked developing custom data processing pipelines for my research group, in both Snakemake and Nextflow.
- | RNA-seq/Metagenomics data analysis | Network analysis | R/Bioconductor | Data visualization | Snakemake | Nextflow |

R Project for Statistical Computing

GSOC STUDENT DEVELOPER

Remote

June 2021–August 2021

- Worked in analysing social network data, primarily from websites as X (formerly Twitter) and Meetup, generating interactive dashboards and focusing on gaining insights regarding the R community as a whole and how we could best take advantage of trends in the community to organize new events and focus efforts on particular projects.
- | R/Tidyverse | Dashboards and data visualization | GitHub Actions for CI/CD | R/Shiny |

Skills

Programming

PYTHON, R, BASH, JAVASCRIPT

Workflows

NEXTFLOW, SNAKEMAKE

Query

SPARQL, SQL

Markup

RMARKDOWN, LATEX, HTML

Tools

GIT/GITHUB, GITHUB ACTIONS, SLURM, DOCKER, SINGULARITY, R/SHINY

Software

ARETE

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

- Maintainer

Nextflow

reconciler

INTEGRATES PANDAS AND W3C RECONCILIATION SERVICES

- Author and maintainer

Python

EURYALE

A PIPELINE FOR TAXONOMIC CLASSIFICATION AND FUNCTIONAL ANNOTATION OF METAGENOMIC READS

- Author and maintainer

Nextflow

MicroView

A REPORTING TOOL FOR AGGREGATING RESULTS FROM TAXONOMIC CLASSIFICATION ANALYSES

- Author and Maintainer

Python

pytrials

PYTHON INTERFACE TO THE CLINICALTRIALS.GOV API

- Author and maintainer

Python

go2cell

INTEGRATES GENE ONTOLOGIES (GO) WITH WIKIDATA CELL TYPE ITEMS

- Author and maintainer

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.

Education

Federal University of Rio Grande do Norte

MSC IN BIOINFORMATICS

Natal

2023-ongoing

Federal University of Rio Grande do Norte

BACHELOR IN BIOMEDICAL SCIENCE

Natal

2018-2022

Publications

- **Ancestry analysis indicates two different sets of essential genes in eukaryotic model species.** 2021. <https://doi.org/10.1007/s10142-021-00794-9>
- **Reverse engineering of the pediatric sepsis regulatory network and identification of master regulators.** 2021. <https://doi.org/10.3390/biomedicines9101297>
- **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>
- **Metagenomic analyses reveal the influence of depth layers on marine biodiversity on tropical and subtropical regions.** 2023. <https://doi.org/10.3390/microorganisms11071668>
- **Revealing metastatic castration-resistant prostate cancer master regulator through lncRNAs-centered regulatory network.** 2023. <https://doi.org/10.1002/cam4.6481>
- **What Powers Trastuzumab's Cardiotoxicity? Decoding Mitochondrial-Related Gene Expression Through Integrative Review and Meta-Analysis in Cardiomyocytes.** 2024. <https://doi.org/10.1089/omi.2024.0004>