João Vitor F. Cavalcante

BIOINFORMATICS DEVELOPER

✉ jvfecav@gmail.com | Ⓞ 0000-0001-7513-7376 | ⦿ jvfe | in joao-v-cavalcante

About me

I'm a bioinformatician passionate about open science and making data analysis more accessible. My expertise lies in workflow development and large-scale data analysis, with a strong focus on reproducibility 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.

Experience

Sepal AI   *Remote*

BIOINFORMATICIAN   *August 2025–ongoing*

- Data curation and coding for evaluation of AI models in bioinformatics scenarios.
- | R | Nextflow |

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, reflog 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

Tools

Git/GitHub, GitHub Actions, Slurm, Docker, Singularity

Software

ARETE _Nextflow_

Analysis pipeline for profiling the genomic repertoire and evolutionary dynamics of microorganisms

EURYALE _Nextflow_

A pipeline for taxonomic classification and functional annotation of metagenomic reads

MicroView _Python_

A reporting tool for aggregating results from taxonomic classification analyses

pytrials _Python_

Python interface to the ClinicalTrials.gov API

go2cell _R_

Integrates Gene Ontologies (GO) with Wikidata cell type items

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 _Natal_

PhD in Bioinformatics _2024–ongoing_

Federal University of Rio Grande do Norte _Natal_

MSc in Bioinformatics _2023–2024_

Federal University of Rio Grande do Norte _Natal_

Bachelor in Biomedical Science _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`
- **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`
- **Causeway: A pipeline for genome-wide effector gene screening with mendelian randomization and colocalization**. 2025. `https://doi.org/10.1093/bioadv/vbaf110`