

# Felipe Vaz Peres

[felipevzps.github.io](https://felipevzps.github.io) | [linkedin.com/in/felipevzps/](https://www.linkedin.com/in/felipevzps/) | <https://github.com/felipevzps>

## EXPERIENCE

### Center for Microbial Secondary Metabolites, Technical University of Denmark

Remote

Guest Researcher

Jan 2024 - Nov 2024

- Collaborated with wet-lab scientists to design, implement, and optimize bioinformatics analysis from complex microbial communities in marine sediments, utilizing Oxford Nanopore Technology long reads.
- Developed a complete pipeline to analyze metagenomics data for Biosynthetic Gene Clusters identification in Metagenome-Assembled Genomes.
- Developed and documented pipeline benchmarks, and presented research findings at lab meetings.

### Computational, Evolutionary and Systems Biology Laboratory, University of São Paulo

Piracicaba, SP

Graduate Researcher

Jan 2022 - Nov 2024

- Applied AI/ML algorithms to identify over 8 million non-coding RNAs in the sugarcane pan-transcriptome.
- Analyzed large co-expression networks of protein-coding and non-coding RNAs (over 1 million transcripts), revealing the variability of non-coding RNAs in sugarcane.
- Annotated over 250.000 non-coding RNAs, enriching our comprehension of their variability and functional roles.

### Genera

São Paulo, Remote

Bioinformatics Intern

Jan 2020 - Mar 2020

- Enhanced human ancestry calculator analyzing over 650.000 SNPs.
- Analyzed population structure and ancestral origins across 70+ diverse population groups and geographical regions.
- Development of Genera's ancestry calculator documentation.

### Computational, Evolutionary and Systems Biology Laboratory, University of São Paulo

Piracicaba, SP

Undergraduate Researcher

May 2019 - May 2021

- Analyzed over 3.7 billion public NGS short reads (exceeding 1.1TB of raw data) from 48 sugarcane genotypes, resulting in the identification of over 16 million RNAs and the inference of the sugarcane pan-transcriptome.
- Developed a Snakemake pipeline to execute from data retrieval to downstream analysis, revealing the variability of over 600.000 proteins within sugarcane and inferring its core, accessory, and exclusive proteins.
- Presented research findings at lab meetings and computational biology conferences (e.g. ISMB).

## COMPUTATIONAL SKILLS

**Programming:** Python, Bash, R

**Development:** Unix/Linux, HPC cluster (SGE, PBS, LSF), git

**Pipelines:** Nextflow, Snakemake

**Bioinformatics:** transcriptomics, genomics, metagenomics, pathway analysis, familiar with short and long reads, gene expression, co-expression networks, annotation

## EDUCATION

### University of São Paulo, Center of Nuclear Energy in Agriculture

Piracicaba, São Paulo

M.Sc. in Bioinformatics

Nov 2024

- Relevant Coursework: Genomics and Bioinformatics, From Gene to Trait (advanced biotechnology pipeline)

### Federal University of São Carlos, Center of Agrarian Sciences

Araras, São Paulo

B.S. in Biotechnology

Jan 2022

## HONORS AND AWARDS

- **CNPq:** National Council for Scientific and Technological Development - *Graduate Research Fellowship* April 2022
- **3rd place:** Developed an automated variant calling pipeline in under 48 hours - *Mendelics Challenge* Nov 2021
- **Honorable mention:** International Symposium of Undergraduate Research at University of São Paulo Oct 2020
- **FAPESP:** São Paulo Research Foundation - *Undergraduate Research Fellowship* April 2019
- **2nd place:** Synthetic Biology Hackathon at University of São Carlos Aug 2018

## PROJECTS AND LEADERSHIP

### Inteligência Não Artificial - INA (Artificial Intelligence Study Consortium)

Remote

Co-founder

Aug 2024 - Present

- Engaged in regular meetings on AI, covering ML, neural networks, and real-world applications.
- Actively contributed to hackathons and competitions, such as the Genomes to Fields (G2F), applying AI techniques to complex challenges in predictive modeling and data analysis.

### Study Group in Bioinformatics and Molecular Biology

Araras, São Paulo

Member

Jun 2018 - Nov 2021

- Conducted monthly meetings with 10+ bioinformatics enthusiasts, emphasizing programming languages, computational biology and bioinformatics, and the replication of scientific paper pipelines.

### Sustec Junior (Junior Enterprise in Biotechnology)

Araras, São Paulo

Project Management Advisor

Jun 2018 - Jun 2021

- Implementation of new biotechnological solutions; collaboration in project planning and execution; and monitoring customer feedback.

## PUBLICATIONS

*in preparation:* **Multi-genotype analyses of long non-coding RNAs in sugarcane**

**Peres, F.V.**, Riaño-Pachón, D.M.

*in preparation:* **CoNekT Grasses, a web platform for comparative transcriptomics of grasses**

Santos, R.A.C., Cardoso, A.S.O., Pastos, B.S., **Peres, F.V.**, Ferraz, D.T., Pérez, J.M.M., Furlan, G., Riaño-Pachón, D.M.

*in preparation:* **Inference and annotation of the sugarcane pan-transcriptome**

Hellen, **Peres, F.V.**, Pérez, J.M.M., Riaño-Pachón, D.M.

2021: **ContFree-NGS: Removing reads from contaminating organisms in next generation sequencing data**

**Peres, F.V.**, Riaño-Pachón, D.M. [doi.org/10.1007/978-3-030-91814-9\\_6](https://doi.org/10.1007/978-3-030-91814-9_6)

## SOFTWARE

Additional work can be seen on my GitHub profile: <https://github.com/felipevzps>

*Research projects*

- [sugarcane RNAome](#) - Multi-genotype analyses of long-ncRNAs in sugarcane.
- [SCPT](#) - Inference and annotation of the sugarcane pan-transcriptome.
- [ContFree-NGS](#) - Open source software that removes sequences from contaminating organisms in NGS datasets.
- [YAATAP](#) - Snakemake pipeline for downloading raw RNA-seq datasets and performing transcriptome assemblies.
- [seabed-symphony](#) - Pipeline for BGCs identification in microbial community of marine sediments.

*Side projects*

- [cavebot-fiesta](#) - A software that automates repetitive actions in the game Tibia using real-time computer vision.
- [pesca-a-dor](#) - Automated robot to play PokeXGames using OpenCV and PyAutoGUI.
- [indeed\\_jobs](#) - Automated web scraping using Selenium and BeautifulSoup to extract job listings from Indeed.

## SELECTED PRESENTATIONS

### 2022 Python: From zero to your first software

Presented at workshop - Python Workshop for Biological Data, Online Conference - UNICAMP

### An introduction to programming for bioscientists

Presented at workshop - Genetics and Molecular Biology Meeting, Online Conference - UNICAMP

### 2021 Removing reads from contaminating organisms in NGS datasets

Poster - Brazilian Symposium on Bioinformatics, Online Conference - Federal University of Minas Gerais

### 2020 Analysis of *de novo* transcriptome assemblies in sugarcane

Selected Talk (Top 15%) - International Symposium of Undergraduate Research, University of São Paulo