

# Felipe Vaz Peres

[Visit my website](#) | [LinkedIn](#)

## EXPERIENCE

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

Remote

Guest Researcher

Jan 2024 - Jan 2025

- Designed and optimized a scalable pipeline for metagenomic studies using Oxford Nanopore long-read data, integrating complex microbial community data from 100s of marine sediment samples into MAGs.
- Standardized annotation of Biosynthetic Gene Clusters (BGCs) across MAGs by implementing antiSMASH and BiG-SCAPE, resolving inconsistencies in functional labeling and improving cross-study comparability.
- 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 - Jan 2025

- Applied AI/ML models to automate identification of 8.4M non-coding RNAs (ncRNAs) in sugarcane, creating a pan-RNAome reference for 48 genotypes.
- Analyzed large co-expression networks of protein-coding and ncRNAs (1M+ transcripts), and annotated 250K+ ncRNAs with standardized functional terms, enriching our comprehension of their variability and functional roles.
- Conducted the most comprehensive analysis to date on ncRNA variability in sugarcane, establishing a foundational framework for investigating their biological functions in this complex polyploid crop.

### Genera

São Paulo, Remote

Bioinformatics Intern

Jan 2020 - Mar 2020

- Enhanced human ancestry calculator analyzing 650K+ SNPs.
- Analyzed structure and ancestral origins across global populations, improving resolution for 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 3.7B public NGS short reads (1.1TB data) from 48 sugarcane genotypes, resulting in the identification of 16M RNAs and the construction of the sugarcane pan-transcriptome.
- Developed a Snakemake pipeline to execute from data retrieval to downstream analysis, revealing the variability of 600K proteins within sugarcane, inferring its core, accessory, and exclusive proteins.
- Presented pipeline design and findings at conferences (ISMB), emphasizing computational reproducibility.

## COMPUTATIONAL SKILLS

**Programming:** Python, Bash, R

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

**Pipelines:** Nextflow, Snakemake

**Bioinformatics:** transcriptomics, genomics, metagenomics, pathway analysis, NGS (short and long reads), variant calling, gene expression, co-expression networks, functional annotation, machine learning

## EDUCATION

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

Piracicaba, São Paulo

M.Sc. in Bioinformatics

Jan 2025

- Genomics and Bioinformatics, From Gene to Trait, Mobile Genetic Elements, Bioinformatics and Microbiomes

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

Araras, São Paulo

B.S. in Biotechnology

Jan 2022

- Statistics, Bioinformatics, Molecular Biology, Microbial Genetics, Breeding and Genetics, Genetic Engineering

## HONORS AND AWARDS

---

- **3rd place:** Liga Brasileira de Bioinformática - The largest bioinformatics competition in Latin America *Jun 2025*
- **Graduate Research Fellowship:** National Council for Scientific and Technological Development *April 2022*
- **3rd place:** Developed an automated variant calling pipeline in under 48 hours at *Mendelics Challenge* *Nov 2021*
- **Honorable mention:** International Symposium of Undergraduate Research at University of São Paulo *Oct 2020*
- **Undergraduate Research Fellowship:** São Paulo Research Foundation *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)

*Co-founder*

**Remote**

*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

*Member*

**Araras, São Paulo**

*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)

*Project Management Advisor*

**Araras, São Paulo**

*Jun 2018 - Jun 2021*

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

## SOFTWARE

---

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

### Research projects

- [KAPT](#) - Nextflow pipeline for proteome inference and annotation through de novo transcriptome assembly.
- [R2C](#) - Nextflow pipeline to build (gene) co-expression networks.
- [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.

## 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