Curro Campuzano Jiménez

Aarhus, Denmark | campuzanocurro@gmail.com | github.com/currocam/ | twitter.com/currocam

Who am I

I am an MSc. degree student in bioinformatics at Aarhus University, and I am passionate about data science and open-source software. I am thrilled at the prospect of future large-scale bioinformatics challenges. During the last year, I have applied this enthusiasm to the fields of meta-transcriptomics and microbial ecology. I have experience in high-performance computing for data analysis and scripting in R, Julia, and Python. I really care about reproducible research and understanding the state-of-the-art methods in statistics and machine learning development.

Experience

September 2022 – Present

Bioinformatician | Student programmer | Environmental Sciences Department Aarhus University

I have automatized most of the bioinformatics pipelines of the section. To achieve this, I have used the framework Snakemake, isolated environments using Conda, Docker and Singularity, Bash, Python, and R scripting. Specifically, I have worked on a pipeline of large-TotalRNA meta-transcriptomic, long-reads amplicon sequencing, and whole-genome sequencing and genome mining (you can find a poster I did about it). Finally, I contribute to the data analysis of various research projects, some of which have been published:

- Exploring microbial diversity in Greenland Ice Sheet supraglacial habitats through culturing-dependent and -independent approaches.

- Abrupt permafrost thaw triggers activity of copiotrophs and microbiome predators.

February 2023 - July 2023

Bioinformatician | Internship | Bioinformatics Research Center, Aarhus University

I worked with my supervisor, Thomas Bataillon, as part of my MSc. program. We aimed to use Pacbio HiFi reads to obtain a chromosome-level assembly of Mediterranean thyme. We improved the highly fragmented *de novo* assembly by scaffolding utilizing the genome of a closely related species from N50=1.87 Mb (n=133) to N50=48.92 Mb (n=8). This work allowed me to work with PacBio-HiFi reads, use the GenomeDK cluster, and dig into Mediterranean thyme's genetic and ecological diversity (slides available for more information).

February 2022 - July 2022

Data Scientist | Internship | DaSCI, Andalusian Institute in Artificial Intelligence

I developed [HMMERutils](#), an R library for preprocessing and analyzing homologous sequences using hidden Markov model profiles during my internship. Using this library, I collected the train data set I later used to improve the Fascin protein annotation (a cancer therapeutic target). I trained a Random Forest classifier based on physicochemical properties and indexes calculated from the amino acid sequence. Using LIME and Shapley Values, I also found that my model paid particular attention to the presence of specific motifs in the binding sites (without priori information about it).

July 2021 - July 2022

Bioinformatician | Student programmer | Physical Chemistry Department University of Granada

I focused on structural bioinformatics and, primarily, protein coevolution. By applying a technique called Statistical Coupling Analysis to a set of homologous sequences of the Fascin protein, I could identify two groups of coevolving residues that are potentially responsible for the allosteric of the protein.

Education

September 2022 - July 2024

University of Aarhus, Aarhus | MSc in Bioinformatics

- Highest grade in the courses [Evolutionary Thinking](#), [Data Science in Bioinformatics](#), [Statistical and Machine Learning](#), and [Algorithms in Bioinformatics.](#)

- PhD course Next Generation Sequencing Data Analysis.

September 2018 - July 2022

University of Granada, Granada | Degree in Biotechnology

- Research initiation grant.

- I graduated with honors in my bachelor's thesis in Bioinformatics.

Skills

- Programming: R (Tidyverse and Bioconductor), Python, Bash, Julia, Rust, and Stan, among others.

- Workflow management system: Snakemake, Docker, Singularity, Conda.

- Version control: Git, GitHub, and GitHub Actions.

- Languages: Advanced Level of Certified English (C1) and Spanish native

Other skills

- Experience working in interdisciplinary environments, bridging the gap between professionals from different backgrounds.

- Good communication and outreach skills.

- I work both independently and as part of a team.

References

- Thanassis Zervas (az@envs.au.dk)

- Thomas Bataillon (tbata@birc.au.dk)

- Coral del Val Muñoz (delval@decsai.ugr)