



# Léo d'Agata

## Bioinformatics engineer

### EXPERIENCES

#### Bioinformatics consultant

**Amaris** – Montréal

September 2023 to present – Permanent contract



- Under contract with a pharmaceutical client to
- Accelerate the bioinformatic deployment of the **NGS** technology for investigation and **QC testing** of commercial products
- Facilitate the analysis of massive High-throughput sequencing data for **Vaccines QC**

#### Bioinformatics engineer/consultant

**Life and Soft** – Bioparc – Fontenay-aux-Roses

October 2018 to August 2023 – Permanent contract



- Statistical analysis consultant for skin microbiome analysis (16S, 18S, ITS) in the cosmetics industry
- Bioinformatics services for the pharmaceutical industry (plasmid insertion research, automatic updating of reference banks)
- Development of a somatic and germline variant search pipeline in oncology for a diagnostic laboratory
- RNAseq analyzes and analyzes of microbiome data (search for differentially expressed genes or differentially abundant microbes) carried out as part of a collaboration with the CEA with publication:

<https://doi.org/10.3389/fimmu.2022.841723>

#### Bioinformatics engineer

**INRA** – Genoscope– Evry - France

October 2014 to September 2018 – Fixed-term contract



- Research & Development in Bioinformatics for the development of the MaGuS tool (published) to improve and assess the quality of genome assemblies using physical/genetic maps  
<http://www.ncbi.nlm.nih.gov/pubmed/26936254>
- Management of genomic sequencing projects:  
Assembly, decontamination, annotation of genomes (*Pisum sativum*, *Saccharomyces cerevisiae* S288C strain, marine protists for the TARA Oceans project) from Illumina and Nanopore data:  
<https://www.ncbi.nlm.nih.gov/pubmed/28369459>  
<https://doi.org/10.1016/j.xgen.2022.100123>
- Promotion of results in the form of written deliverables and presentations

#### Bioinformatics internship

**CEA** – Genoscope– Evry - France

March 2014 to september 2014



- Development of a tool for the "Evaluation of eukaryotic genome assemblies". This tool was used after the internship to assess the quality of the assemblies made of several species (*Arabidopsis Thaliana*, oak and pea)

**INRA** – Nantes – France

March 2013 to august 2013



- Extraction and processing of molecular data contained in a collection of PDB structural files and recording in a relational database

### EDUCATION

#### Master degree (2<sup>nd</sup> year): Bioinformatics

Faculté des Sciences et des Techniques de Nantes  
2013 à 2014

#### Master degree (1<sup>st</sup> year) : Bioinformatics and Biostatistics

Faculté des Sciences et des Techniques de Nantes  
2012 à 2013

### SKILLS

#### Languages

- French
- English : Scientific and conversational

#### Omics

- Analysis of **NGS** data (illumina, Nanopore)
- Metabarcoding (16S, 18S, ITS) and **Whole genome sequencing (WGS)**.
- Sequences alignment (Blast, **bwa**, bowtie, blat, muscle)
- Genome assemblies (OLC,DBG)
- Taxonomic assignment (**Blast**, **Kraken2**) against SILVA, UNITE and Refseq
- Variants analyses (**Dragen-GATK**, **Mutect2**, **Haplotypecaller**, vardict, **Delly**, **Manta**, **CNVkit**, genefuse)

#### Informatics

- Environments: **Linux**, Windows
- Languages : Bash, Python, Perl, SGBD : PostgreSQL.
- IDE: Visual Studio Code, RStudio
- Versioning : **Git**
- Workflow management: **Snakemake**
- Environment management: **Conda**
- Using the Docker containerization platform
- Use of SLURM in **HPC systems**
- Using AWS Instances

#### Biostatistics

- Most commonly used statistical analysis techniques (Student/Wilcoxon, Fisher, CHI-2)
- Multivariate statistics (PCA, ANOVA, linear regression, mixed models)
- Use of the **R** language and the **ggplot2**, **dplyr** and **tidyverse** libraries
- Use of methods of statistical analysis of microbiome data (Alpha and Beta diversity)
- Differential analyzes using the **DESeq2** tool

#### Project management

- Support to bid manager to tenders of different panels of genes related to heart disease
- Management of the single-cell genome assembly (SAGs) project (processing and choice of data generation, oral communication of results)
- Participation as part of a collaboration in the analyzes of the pea genome sequencing project