Léo d'Agata
Bioinformatics engineer

Information

- 35 y.o
- Aix-en Provence France

Contact

✉ leo.dagata@gmail.com

📱 0628357801

Introduction

Holder of a Master 2 specialized in Bioinformatics at the University of Nantes, I am bioinformatics consultant in Amaris company. I am currently on assignment for GSK for Vaccines QC. My previous job as an engineer in bioinformatics / biostatistics took place in Life & Soft company and consisted of use my skills in bioinformatics / biostatistics for customers in the field of cosmetics, health, diagnostics and within our internal R&D. My role covers mainly the microbiome analysis and the oncology analysis.

My first experience took place at INRA on the Genoscope site (François Jacob Institute of Biology) within the R&D Bioseq team. My job concerned the NGS data (assembly of reads from DNA sequencing, annotation, taxonomic assignment, phylogeny, search for structural variations). I am thus familiar with several types of sequencing data ranging from short reads (454, Illumina) to long reads (Pacbio, Nanopore).

EXPERIENCES

Bioinformatics consultant

***Amaris** – Montréal*
September 2023 to present – Permanent contract

- Under contract with **GSK** to facilitate the analysis of massive High-throughput sequencing data for **Vaccines** QC on several projects.
- Perform several types of analysis: Adventitious virus **detection** in upstream samples and final containers
- Viral **stability** analysis, investigate **QC material integrity**, bacterial assembly.
- Maintained reproducible, version-controlled tools scalable on HPC clusters and adapted them for **cloud-based HPC** environments.

Bioinformatics engineer/consultant

***Life and Soft** – Bioparc – Fontenay-aux-Roses*
October 2018 to August 2023 – Permanent contract

- Bioinformatics & biostatistics analyses (16S, WGS, RNA-seq) for **l'Oréal, CEA** and **Cerba HealthCare**
- **Pipeline** development for **microbiome** and **oncology** (Illumina/Nanopore)
- **Client interactions**, project scoping, and results presentation
- **Agile** collaboration with developers for web-based result visualization

Bioinformatics engineer

***INRAE** – Genoscope – Evry – France*
October 2014 to September 2018 – Fixed-term contract

- Management of the single-cell genome assembly (**SAGs**) project (processing and choice of data generation, oral communication of results)
- Co-developed a published tool to assess genome assembly quality using **physical/genetic maps**
- Genome assembly and decontamination (Pisum sativum, Saccharomyces cerevisiae S288C strain, marine protists for the TARA Oceans project) from Illumina and Nanopore data
- Delivered scientific reports and gave oral internal and **external presentation**

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)

***INRAE** – 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: Bioinformatics

Faculté des Sciences et des Techniques de Nantes
2012 to 2014

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, genbank 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 **tidyr** libraries
- Use of methods of statistical analysis of microbiome data (Alpha and Beta diversity)
- Differential analyzes using the **DESeq2** tool

Project management

- Agile project management
- Support to bid manager to tenders of different panels of genes related to heart disease
- Client communication