

Camila Duitama González

BIOINFORMATICIAN · RESEARCH SOFTWARE ENGINEER

Paris, France

☎ (+33)749927193 | ✉ | 🏠 | 📧 | 📷 | 🎓

Summary

Research engineer and bioinformatician with a PhD in k-mer-based methods for ancient metagenomics and a track record spanning sequence bioinformatics, palaeogenomics, and deep learning for large-scale genomic data. Author of four peer-reviewed open-source tools (DIANA, MUSET, decOM, aKmerBroom), each developed from independently formulated research questions and distributed through Bioconda. Experienced in Python and C++ scientific software development, HPC pipeline design, reproducible workflows, supervision of students and engineers, collaborative grant writing, and bioinformatics training.

Skills

Bioinformatics & Genomics	Metagenomics, paleogenomics, k-mer methods, assembly, Nextflow, Snakemake, Bioconda
Programming & Development	Python, C++, R, Bash, Git, pytest, GitHub Actions
Machine Learning & AI	PyTorch, deep learning, multi-task networks, Optuna, predictive modelling
Infrastructure & Compute	HPC, Slurm, GPU computing, Docker, Singularity, Conda, AWS
Statistics & Phylogenetics	Bayesian inference, MCMC, BEAST, IQ-TREE, benchmarking, phylodynamics
Data Engineering	Pandas, Polars, NumPy, SQL, Spark
Languages	Spanish (Native), English (C2), French (B2), German (B1)

Experience

Laboratory of Evolutionary Dynamics of Infectious Diseases, Institut Pasteur

Paris, France

RESEARCH SOFTWARE ENGINEER

May 2025 - Present

- Lead engineer for the group, owning software engineering and AI tool implementation for infectious disease research.
- Define the group's IT and compute setup, including HPC cluster usage guidelines, and advise on investments of over €100K in GPU infrastructure.
- Implement best practices for code management, version control, code review, CI/CD and reproducible research.
- Supervise Bachelor's students as main supervisor, leading their research projects and reviewing their code from design to benchmarking.

Sequence Bioinformatics Group, Institut Pasteur

Paris, France

POSTDOCTORAL RESEARCHER

January 2024 - March 2025

- Designed, developed and benchmarked reproducible data pipelines for genotype-phenotype association studies.
- Co-supervised a team of research engineers on a collaborative software development project in sequence bioinformatics, from design and implementation to benchmarking and publication (*Bioinformatics*, 2025).
- Mentored 1 Master's intern.

Max Planck Institute for Biology of Ageing

Cologne, Germany

RESEARCH ASSISTANT/PYTHON DEVELOPER

Jan-Dec 2018, Apr-Aug 2020

Max Planck Institute of Biochemistry

Munich, Germany

INTERN

March 2017 - July 2017

Open-Source Software Tools

2026	DIANA , Deep learning framework for ancient metagenomics	Python, PyTorch
2025	MUSET , Abundance unitig matrix construction utility	C++
2023	aKmerBroom , Ancient DNA decontamination toolkit	Python
2023	decOM , Microbial source tracking tool	Python

Publications

5 peer-reviewed articles and 2 preprints since 2020 — 3 as first author, 2 as last author. * Intern supervised by Camila Duitama.

Preprint

BioRxiv

2026

- **Duitama González, C.**, Lopopolo, M., Nishimura, L., Faure, R., & Duchene, S. (2026). DIANA: Deep Learning Identification and Assessment of Ancient DNA. *Preprint*. doi:10.64898/2026.04.09.717429

Journal Article

Nature Microbiology

2026

- You, Y., et al. (incl. **Duitama González, C.**) (2026). Emergence of distinct *Streptococcus pyogenes* emm1 and emm12 lineages in China. *Nature Microbiology*. doi:10.1038/s41564-026-02324-4

Journal Article

Molecular Biology and Evolution

2026

- Weber, A., Kende, J., **Duitama González, C.**, Översti, S., & Duchene, S. (2026). The Phylodynamic Threshold of Measurably Evolving Populations. *Molecular Biology and Evolution*, 43(6), msag133. doi:10.1093/molbev/msag133

Preprint

BIORxiv

2026

- Fosse, S., Duchene, S., & **Duitama González, C.** (2026). Benchmarking BEAGLE to find optimal parameters for BEAST X. *Preprint*. doi:10.64898/2026.03.10.710534

Journal Article

OXFORD BIOINFORMATICS

2025

- Vicedomini, R., Andreace, F., Dufresne, Y., Chikhi, R., & **Duitama González, C.** (2025). MUSE: Set of utilities for the construction of abundance unitig matrices from sequencing data. *Bioinformatics*. doi:10.1093/bioinformatics/btaf054

Journal Article

ISCIENCE

2023

- **Duitama González, C.**, Rangavittal, S., Vicedomini, R., Chikhi, R., & Richard, H. (2023). aKmerBroom: Ancient oral DNA decontamination using Bloom filters on k-mer sets. *iScience*, 26(11). doi:10.1016/j.isci.2023.108057

Journal Article

MICROBIOME

2023

- **Duitama González, C.**, Vicedomini, R., Lemane, T., Rascovan, N., Richard, H., & Chikhi, R. (2023). decOM: similarity-based microbial source tracking of ancient oral samples using k-mer-based methods. *Microbiome*, 11(1), 243. doi:10.1186/s40168-023-01670-3

Education

Sorbonne Université - Institut Pasteur

Paris, France

PH.D. INFORMATION AND COMMUNICATION SCIENCE AND TECHNOLOGY

Oct 2020 - Jan 2023

Bonn University

Bonn, Germany

M.SC. LIFE SCIENCE INFORMATICS

Oct 2017 - Oct 2019

Universidad Nacional de Colombia

Medellín, Colombia

B.ENG. BIOLOGICAL ENGINEERING

Feb 2012 - Sept 2017

Teaching & Training

Scalable Genomics and Pangenomics, Wellcome Connecting Science

Wellcome Genome Campus,

Hinxton, UK

TEACHING ASSISTANT

Oct 2026

- Support instruction on scalable, k-mer-based genomics, pangenome graph representations, data indexing/compression, and workflow design for population-scale analysis.

Statistics, Bioinformatics & Image Analysis Training Program, Institut Pasteur

Paris, France

TEACHING ASSISTANT

Nov 2021

- Assisted in the R Programming and Statistics track: introduction to R, hypothesis testing, linear models, and multivariate analyses.

Presentations

- 2026 **Invited speaker**, Symposium “Artificial Intelligence in Biology and Health”, Sequences session — “DIANA” *Institut Pasteur*
- 2024 **Presenter**, Genome Informatics — “MUSE” *WGC, Hinxton, UK*
- 2023 **Presenter**, RECOMB-Seq — “aKmerBroom” *Istanbul, Turkey*
- 2022 **Paper & Poster**, LMRL Workshop, NeurIPS *Virtual*

References

Head of the SeqBio group, Institut Pasteur

PhD/Postdoc supervisor

DR. RAYAN CHIKHI

rchikhi@pasteur.fr

Head of the Evolutionary Dynamics of Infectious Diseases Group, Institut Pasteur

Current supervisor

DR. SEBASTIAN DUCHENE

sduchene@pasteur.fr

Associate Director, Data Science & Advanced Analytics, Grünenthal Group

Master thesis supervisor

DR. ASHAR AHMAD

ashar799@gmail.com