

Sandra Godinho Silva

📍 Portugal ✉ sandragodinhosilva@gmail.com ☎ +351 925990343 🔗 sandragodinhosilva.github.io
 in sandragodinhosilva 🌐 sandragodinhosilva

Ph.D. in Biotechnology and Biosciences, specializing in Data Science and AI applied to healthcare. Experienced in working with real-world health datasets, including federated learning in precision oncology and medical ontology management. Skilled in scalable data analysis, automation, and omics analysis. Passionate about applying ML and AI techniques to improve healthcare outcomes, bridging pharmaceutical sciences with programming and research. Proven track record in delivering production-grade data pipelines and thriving in interdisciplinary research environments.

Education

Lisbon Data Science Academy

Oct 2024 – Jun 2025

- Coursework: Binary Classification; Data Wrangling; Time Series; Text Classification; Recommender Systems; Data Science in the real world.
- Capstone project: built a predictive time-series model to forecast retailer prices using anonymized real data.

Doctoral Program in Biotechnology and Biosciences

2018 – 2023

Instituto Superior Técnico (IST) - University of Lisbon

- Thesis topic: [Secondary metabolite biosynthesis in *Aquimarina* species: antimicrobial properties and bioactivities from the rare marine biosphere](#)
- Jury final classification: **Pass with Distinction and Honour**

Master Degree in Pharmaceutical Sciences

2012 – 2017

Faculty of Pharmacy - University of Lisbon

- Thesis topic: [Multi-approach analysis of the metagenome of a marine sponge containing latrunculin A](#)

Experience

Data Scientist

Jul 2025 – Present

IQVIA

- Contributed to [DigiONE](#), a European pilot powering privacy-preserving federated learning across cancer centers using standardized OMOP data, generating real-world oncology insights to support precision medicine research.

Statistical Programmer

Oct 2022 – Jun 2025

IQVIA

- Progressed from supporting to lead programmer in multi-country real-world evidence studies for major pharmaceutical companies (e.g., Bayer, Merck, Novartis, AstraZeneca).
- Designed and executed analyses on Electronic Medical Records, national health registries, and large-scale health databases, including survival analysis and longitudinal data modeling.
- Built robust data preprocessing and transformation workflows in SAS and R, ensuring datasets were analysis-ready, compliant with regulatory standards (FDA, EMA) for better flow.

Teaching Assistant

2020 - 2022

Bioengineering Department, Instituto Superior Técnico

- Lectures and hands-on sessions on bioinformatics principles, genome annotation, and secondary metabolism for the Microbiome (master level) course.
- Informal supervision of several Master students.

Bioinformatics Assistant

2018 - 2022

Microbial Ecology and Evolution Laboratory, IST

- Provided bioinformatics support for multiple research projects, enhancing the efficiency and accuracy of *in-silico* analyses.
- Migrated all bioinformatics workflows to a High-Performance Computing cluster, reducing analysis time and enabling large-scale genomic studies.
- Automated key processes with Python, R, and Bash, streamlining data handling and improving reproducibility.

Visiting Student

Oct 2021 - Feb 2022

Institute of Microbiology, ETH Zürich

Switzerland

- Discovered and isolated bioactive microbial compounds; elucidated their chemical structures.
- Analyzed genomic and metabolomic data to identify promising targets for natural product discovery.

Visiting Student

Mar - Dec 2020

Helmholtz-Centre for Environmental Research, UFZ Leipzig

Germany, Remote

- Conducted large-scale comparative genomics, including use of ML to identify the most important features distinguishing different bacterial genera.
- Designed and implemented [Melange](#), an automated Snakemake pipeline for genome annotation.

Visiting Student

Feb - Apr 2017

Institute of Pharmaceutical Biology, University of Bonn

Germany

- Applied bioinformatics and sequence-based screening to identify biosynthetic gene clusters (PKS/NRPS) and profile microbial communities from marine sponge metagenomes, linking genetic potential to natural product discovery.

Selected Publications

For a complete list, please visit my [Google Scholar page](#).

Natural product biosynthetic potential reflects macroevolutionary diversification within a widely distributed bacterial taxon.

2023

Silva SG, Homsí MN, Keller-Costa T, Nunes da Rocha U, Costa R. [mSystems](#).

Machine learning-assisted identification of bioindicators predicts medium-chain carboxylate production performance of an anaerobic mixed culture.

2022

Liu B, Sträuber H, Saraiva J, Harms H, Silva SG, Kasmanas JC, Kleinstaub S, Rocha UN. [Microbiome](#).

MarineMetagenomeDB: a public repository for curated and standardized metadata for marine metagenomes.

2022

Nata'ala MK, Santos APA, Kasmanas JC, Bartholomäus A, Saraiva JP, Silva SG, Keller-Costa T, Costa R, Gomes NCM, Carvalho ACPLF, Stadler PF, Sanches DS, Nunes da Rocha U. [Environmental Microbiome](#).

Metagenomics-resolved genomics provides novel insights into chitin turnover, metabolic specialization, and niche partitioning in the octocoral microbiome.

2022

Keller-Costa T, Kozma L, Silva SG, Toscan R, Gonçalves J., Lago-Lestón A, Kyrpides NC, da Rocha UN, Costa R. [Microbiome](#).

Comparative genomics reveals complex natural product biosynthesis capacities and carbon metabolism across host-associated and free-living *Aquimarina* (*Bacteroidetes*, *Flavobacteriaceae*) species.

2019

Silva SG, Blom J, Keller-Costa T, Costa R. [Environmental Microbiology](#).

Certifications

Carpentries Instructor

June 2022

Certified by [The Carpentries](#) to teach foundational coding and data science skills, including R programming, using inclusive and evidence-based teaching practices.

Skills

- **Languages:** Portuguese (native), English (fluent, Cambridge English Advanced (CAE), C1)
- **Programming Languages:** Python (expert), R (proficient), Bash (proficient), SAS (expert)
- **Tools:** Git, Docker, VS Code, HPC environments, Jupyter
- **ML:** scikit-learn, XGBoost, LightGBM
- **Data Science:** Pandas, NumPy, Matplotlib, SQL