

# Curriculum vitæ of Alessandro Fuschi

## Personal Informations

Date of birth: 13/06/1991

Birth Place: Orvieto (TR)

Domicile: Via Murri 110, 40137, Bologna

Cel.: +39 334 1114474

E-Mail Unibo: alessandro.fuschi2@unibo.it

E-Mail Personal: fuschi.alessandro@gmail.com

Web Site: <https://www.unibo.it/sitoweb/alessandro.fuschi2>



## Professional and Academic Background

02/2024 - Present **Researcher in the TRIGGER Project.**

*Project Focus:* The TRIGGER project aims to understand the impacts of climate change on health through genomic and metagenomic techniques to sample subjects and environments. The goal is to analyze metagenomic data to describe pathologies and their evolutions, identify diagnostic markers, or indicators of environmental changes due to climate change.

*Responsibilities:*

- Development of mathematical models based on the theory of compositional data to analyze metagenomic data.
- Implementation and application of bioinformatics pipelines for the analysis of metagenomic.
- Application of methods derived from network theory to characterize communities of organisms, and mathematical models to study the evolution of these networks or their features.

*Interdisciplinary Collaboration:* The project requires interaction with experts in virology/bacteriology, epidemiology, medical/clinical sciences, and environmental physics.

11/2020 - 01/2024 **PhD in Applied Physics** at the University of Bologna, Co-Financed by the European VEO Project (Versatile Emerging Infectious Disease Observatory) <https://www.veo-europe.eu/>.

*Research Focus:* Data analysis and physical modeling on metagenomics data. Specially, explored compositional aspects of sequencing data, specializing in the development of network analyses tailored to this type of data. Proficient in crafting bioinformatics pipelines optimized for High-Performance Computing (HPC) systems, with a strong emphasis on massively parallel computation. Utilized Conda and Snakemake for the efficient processing and management of complex bioinformatics workflows.

02/2023-04/2023 **Period abroad in Denmark**, hosted at the Technical University of Denmark (DTU), under the supervision of Dr. Frank Aerestrup and Patrick Munk.

*Exchange Focus:* Development of a bioinformatics pipeline and subsequent analysis, aimed at preventing emerging diseases through the monitoring of wastewater in European cities.

2018-2020 **Master's Degree in Applied Physics** at Alma Mater Studiorum - University of Bologna, grade 110/110 with honors. Thesis title: "Compositional data analysis applied to human microbiome network reconstruction" - Supervisor: Prof. D. Remondini.

- 2013-2017 **Bachelor's Degree in Physics** with grade 89/110 at Alma Mater Studiorum - University of Bologna. Thesis title: "Calculation of holograms with matlab and octave" - Supervisor: Prof. M. Cuffiani.
- 2005-2010 **Scientific Diploma** with grade 90/100 at the Liceo Ettore Majorana, Orvieto (Italy).

## Competitive and Auxiliary Projects

- 09/2022-11/2022 **Co-supervisor of a Bachelor's Thesis in Physics** entitled *Analysis of cell migration in vitro in the presence and absence of an electric field* with supervisor Prof. Gastone Castellani.
- 08/2022-10/2022 **CINECA ISCRA-C**. Responsible for the writing and the programming implementation for the project *MEtaGenomics Assembly for AntiMicrobial Resistance*. The goal was the identification of antibiotic resistance genes within de novo assembled genomes sourced from environmental samples collected on livestock farms. The project required management of important computational resources, specifically 75000 core hours on the Galileo100 High-Performance Computing (HPC) system. This system featured 528 nodes with 384GB memory and Intel CascadeLake 8260 CPUs.
- 09/2021-12/2021 **Didactic Tutor** at the course of *Numerical analysis* [cod. 04524], Degree in Computer Science for Management, supervised by Professor Giulio Casciola.  
<https://www.unibo.it/it/didattica/insegnamenti/insegnamento/2021/375171>
- 09/2021-12/2021 **Didactic Tutor** at the laboratory of the course of *Numerical methods for calculation* [cod. 69176], Degree in Computer Science for Management, supervised by Professor David Palitta.  
Task: Assist students in developing programs in Matlab.  
<https://www.unibo.it/it/didattica/insegnamenti/insegnamento/2019/375737>
- 09/2020-12/2020 **Didactic Tutor** at the laboratory of the course of *Numerical methods for calculation* [cod. 69176], Degree in Computer Science for Management, supervised by Professor Giulio Casciola.  
Task: Assist students in developing programs in Matlab.  
<https://www.unibo.it/it/didattica/insegnamenti/insegnamento/2019/375737>
- 10/2019-02/2020 **Didactic Tutor** to the *OFA preparation for Mathematics* [cod. 74566], Degree in Computer Science for Management, supervised by Professor Chiara Guidi.  
Task: Exercises in the class of Mathematical Analysis concerning goniometric functions, limits, study of functions, taylor expansions, derivatives, integrals and generalized integrals. Correction in class of the exams taken by students in the winter session.  
<https://www.unibo.it/it/didattica/insegnamenti/insegnamento/2019/367006>
- 11/2018-03/2019 **Co-supervisor of a Bachelor's Thesis in Physics** entitled *Measurement of amplitude and phase in digital holography* with supervisor Prof. Marco Cuffiani.
- 04/2018-06/2018 **Didactic Tutor** at the laboratory of the course of *Informatic skills for chemistry* [cod. 84639], Degree in Chemistry and Chemistry of Materials supervised by Professor Marco Lombardo.  
Task: To assist students in the development of programs in Fortran 95.  
<https://www.unibo.it/it/didattica/insegnamenti/insegnamento/2018/422231>

## Conferences with Contributions

- 12-14/06/2023 Engaged as a speaker at the *Versatile Epidemiological Observatory (VEO) Symposium*

2023 in Rotterdam, complemented by a research poster presentation.  
<https://www.aanmelder.nl/143737>

16-18/05/2022 Presentation of a scientific poster and speaker at the young section to the *Versatile Epidemiological Observatory* (VEO) Symposium 2022 held in The Hague.  
<https://www.aanmelder.nl/143737>

## Publications

In submission **Correlation measures in metagenomic data: the blessing of dimensionality:** Conducted a comprehensive study investigating potential biases in correlations due to inherent properties of metagenomics data, such as compositionality and data sparsity. This was achieved through the generation of synthetic data using the 'Normal To Anything' approach. Additionally, the study entailed a comparative analysis of state-of-the-art methods applied to real-world data sets. <https://doi.org/10.1101/2023.06.20.545665>.

In submission **Time-series metagenomic analyses reveals bacterial communities with annual periodicity:** Led a study on metagenomic samples from sewage systems across five European cities, with a specific focus on time series analysis of bacterial abundances. This research revealed distinct bacterial communities exhibiting annual trends, identified through an innovative network-based approach.

In submission **An extensive metagenomic dataset of sewage samples across five European cities: A resource for analyzing microbial communities, antimicrobial resistance and pathogens:** Developed and documented a data descriptor paper on the study of metagenomic samples from five European sewage systems. In this study, we employed various methods to reconstruct bacterial abundances, with a special focus on the co-assembly approach. This technique allowed for a detailed reconstruction of the genomes present in the environment, a process that involved substantial computational efforts.

## Skills

Languages Italian (native language)  
English (B1)

Informatics Excellent mastership of the R programming language;  
Excellent knowledge of Linux operating systems;  
Excellent mastership of Matlab;  
Excellent mastership virtual environments using Conda;  
Excellent knowledge of snakemake for building bioinformatics pipelines;  
Proficient in Bash scripting language. Good knowledge of C ++ programming language, integrated with the boost libraries;  
Good knowledge of the High-Performance Computing system;  
Good knowledge of Gnuplot for the creation of graphics.  
Good knowledge of the OpenCV computer vision library;  
Good knowledge of Windows and Mac operating systems;  
Good knowledge of Microsoft Office, OpenOffice, LibreOffice packages;  
Good knowledge of python and Fortran95 languages.

## Research Area with Specific Skills

My research primarily centered on metagenomic data analysis, pivotal for elucidating microbial biodiversity and ecosystem dynamics. I possess comprehensive expertise in standard metagenomic analytical techniques, including alpha and beta diversity assessments, advanced methods in dimensionality reduction, and differential abundance analysis. Using next-generation sequencing (NGS) techniques, I analyzed microbial DNA from various natural environments, yielding a thorough view of microbial diversity.

A key aspect of my research is the exploration and adoption of advanced methodologies, with a special emphasis on network analysis in metagenomic data. I began by mastering widely-used network reconstruction techniques such as SPIEC-EASI, sparCC, and rho proportionality, which laid the groundwork for developing my own unique approaches to network reconstruction and analysis. A significant innovation in my method was incorporating the signs of relationships between species. For example, I implemented a signed version of modularity to define communities as clusters of nodes mainly connected by positive links, enhancing the depiction of cooperative and antagonistic interactions within microbial communities and yielding novel insights into their behavior. Furthering this work, I have developed an R package designed to streamline the handling and network analysis of metagenomic data. This package, built on the S4 class object system in R, is currently in its alpha version and is accessible at <https://github.com/Fuschi/mgnet>.

Throughout my research, a consistent focus has been on recognizing metagenomic data as inherently compositional, delving deeply into the theoretical framework established by Aitchison's studies. This exploration encompassed not only the selection of appropriate metrics but also a thorough analysis of potential spurious correlations that may emerge when assessing inter-species relationships during network reconstruction. My work meticulously addressed these challenges, ensuring more robust and accurate interpretations of metagenomic data, particularly in the context of network analysis. In this context, my research extended to include sophisticated analyses involving the generation of synthetic data. This was crucial for evaluating spurious correlations. A key part of this process was the implementation of the 'Normal to Anything' method, enabling the creation of data with a pre-selected covariance structure and a specific desired distribution. This approach not only allowed for a more precise assessment of correlations within metagenomic data but also provided a robust framework for testing and validating network reconstruction methodologies under controlled conditions, further enhancing the reliability and accuracy of my findings in the realm of metagenomic research.

I had the fortunate opportunity to apply my expertise in two distinct and enriching collaborative settings. Firstly, as a part of my Ph.D. scholarship co-funded by the VEO project, I engaged in a close collaboration with Frank Aarestrup's group in Copenhagen. This collaboration centered around investigating the wastewater of five major European cities. Our primary discovery was the identification of distinct bacterial communities exhibiting clear seasonal trends within these urban sewer systems. This finding indicated the existence of intricate inter-bacterial relationships and the formation of complex ecological systems even in urban wastewater environments. I also partnered with George Weinstock from the Jax Laboratory (USA) and Jethro Johnson from the Kennedy Institute in Oxford for a study on the ecology of intestinal bacteria. In this collaboration, I successfully integrated data on species abundances with their associated metabolic pathways. This comprehensive analysis has yielded promising preliminary results. These findings potentially offer insights into environmental factors linked to varying diets. Our work is at the forefront of understanding how microbial communities present in the gut can reflect broader environmental and lifestyle patterns.

In both collaborative projects, I played a crucial role in developing the bioinformatics pipelines essential for generating final data sets for statistical analysis.

A significant portion of my project involved the intricate process of generating wastewater data. Under the guidance of Dr. Patrick Munk from DTU's FOOD institute, we implemented cutting-edge pipelines for reconstructing bacterial genomes from shotgun sequencing data. Given the complexity of wastewater samples, which contain hundreds, if not thousands, of mostly unknown bacterial species, our approach

was designed to minimize information loss. We adopted a computationally intensive yet effective strategy, achieving an average alignment of 70 – 80% of the original reads across all samples. Essentially, we created a bespoke reference database from raw data. This involved assembling reads with Megahit and grouping them by city, a process known as co-assembly, which yielded more complete contigs at the cost of enormous computational resources. For instance, co-assembling a single city's data typically required an entire 'fat node' of the DTU HPC called Computerome, utilizing 40 parallel cores, over a terabyte of RAM, and nearly a month of continuous processing. All this followed by the binning phase with the metabat2 software and the validation of the quality of the bins obtained with checkm2. To finalize the pipeline, we employed the dRep tool to select the best representative genome for each species, addressing the issue of multiple reconstructions of a single species. This step was crucial in achieving a non-redundant reference database, ensuring accuracy and efficiency in our analyses.

In the development of these bioinformatics pipeline, I leveraged the power of Snakemake for workflow management and Conda for managing virtual environments, ensuring seamless tool integration and efficient processing on the HPC with Slurm as the workload manager.

Finally, in the collaborative project with George Weinstock, I was entrusted with the critical task of enriching the informational depth of the 16S intestinal microbiota data. This involved applying picrust2 to over 800 samples in the dataset, a process that I executed on the High-Performance Computing (HPC) Open Physics Hub (OPH) at the University of Bologna.

Throughout these diverse research projects, I have consistently collaborated with both biologists and bioinformaticians, honing my ability to engage effectively in interdisciplinary work. This experience has not only enriched my technical and analytical skills but has also enhanced my capacity for teamwork, communication, and integrating diverse perspectives. My role often bridged the gap between complex computational methods and biological understanding, emphasizing the importance of collaborative synergy in advancing scientific research.

Bologna, April 2024.