

## AVAILABLE POSITIONS

|                                 |                              |
|---------------------------------|------------------------------|
| <b>Principal Investigator</b>   | <b>GRAZIANO PESOLE</b>       |
| <b>Institute of Affiliation</b> | University of Bari Aldo Moro |

| PROJECT INFO                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title of the proposed project:   | <b>A Galaxy-Integrated Framework for Multi-Layer Long-Read Genome Analysis: Genome Assembly, Structural Variation, Haplotype Resolution and Epigenome Profiling</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Short description of the project | <p>Despite remarkable advances in sequencing technologies, the emerging complexity of eukaryotic genomes remains difficult to resolve using conventional short-read approaches, which systematically miss large structural variants, collapse repetitive regions, and fail to capture haplotype-resolved variation and DNA methylation in a single experiment. Third-generation long-read platforms - particularly Oxford Nanopore Technologies (ONT) and PacBio HiFi - have fundamentally changed this landscape by enabling the generation of highly contiguous genome assemblies, the resolution of complex structural rearrangements, and the simultaneous profiling of base modifications at single-molecule resolution. However, the translation of these capabilities into accessible, reproducible, and standardized analytical workflows remains a critical bottleneck, particularly for research groups lacking dedicated bioinformatic expertise or high-performance computing infrastructure.</p> <p>This PhD project aims to develop an integrated computational framework that enables multi-layer analysis of long-read genomic data, bringing together complementary dimensions of genome information within a single, standardized and reproducible analytical environment. The central hypothesis is that the joint interpretation of genome assembly, structural variation, haplotype architecture, epigenetic information and family-based inheritance will substantially improve the characterization of complex genomes compared with current fragmented analytical approaches, while making advanced long-read analyses accessible to a broader scientific community.</p> <p>The proposed analysis framework will be implemented within the Galaxy platform, providing an intuitive, modular and fully reproducible environment that supports the complete analytical workflow from raw sequencing reads to biologically interpretable results. Both reference-guided and de novo assembly strategies will be supported through decision-aware workflows that can be adapted to different sequencing platforms, genome complexities and experimental designs. Rather than functioning as independent analytical modules, the different components of the framework will be tightly integrated, allowing information generated at each analytical level to enhance the interpretation of the others. Structural variants will be interpreted within phased haplotypes, methylation profiles will be assigned to individual alleles whenever possible, and pedigree information will facilitate</p> |

## AVAILABLE POSITIONS

|                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|--------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                      | <p>the identification of inherited and de novo genomic variants through family-based analyses.</p> <p>The workflow will be systematically benchmarked using publicly available datasets and validated through representative applications in both human and non-human genomics. In particular, the project will explore its application to rare disease genomics, where the integration of structural variants, phased genomes and epigenetic information may improve molecular diagnosis, as well as to population and evolutionary genomics, where long-read sequencing enables a more accurate characterization of genomic diversity. The project is also well aligned with emerging international initiatives such as the Human Pangenome Reference Consortium, and Genome of Europe, which increasingly require standardized, interoperable and reproducible analytical infrastructures for large-scale long-read genomic data. By integrating multiple dimensions of genome analysis within a single Galaxy-based environment, this project aims to establish a new paradigm for the analysis of long-read sequencing data. The resulting framework will be openly distributed through the Galaxy ecosystem together with training materials and best-practice workflows, promoting reproducibility, accessibility and broad adoption by the international genomics community.</p> |
| Main research area for the project   | Computational biology                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Second research area for the project | Genomics                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 3 key words for the project          | Genome assembly, Epigenomics, Variant prioritization                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

| LAB INFO                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Main topic/s of the lab               | Computational biology and Molecular Biology                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Short description of the lab activity | <p>The laboratory led by Prof. Graziano Pesole spans the University of Bari Aldo Moro and the CNR Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), of which he is research fellow. The group develops algorithms, software and reproducible workflows for high-throughput and long-read sequencing data, with long-standing expertise in comparative genomics, genome annotation and molecular biodiversity. It coordinates ELIXIR-IT, the Italian node of the ELIXIR European research infrastructure, and operates the Galaxy-based Laniakea@ReCaS environment on the ReCaS-Bari HPC/cloud facility, providing the reproducible, standardized analytical infrastructure on which this PhD project builds. The lab is deeply engaged in the European initiatives that this project targets: as Italian coordinator within the Genomic Data Infrastructure (GDI) and through Italy's central role in Genome of Europe (GoE), whose data are being made accessible via the GDI infrastructure. It also contributes to reference-genome efforts such as the European Reference Genome Atlas (ERGA),</p> |

## AVAILABLE POSITIONS

|                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                     | <p>reinforcing its focus on federated, interoperable and reproducible analysis of large-scale genomic data.</p> <p>The PhD student will design and develop an integrated Galaxy-based computational framework for multi-layer analysis of long-read sequencing data, combining genome assembly, structural variant discovery, haplotype phasing, native DNA methylation profiling and pedigree-aware analyses. The project will include the implementation and optimization of modular analytical workflows, systematic benchmarking against state-of-the-art methods using public long-read datasets, and validation through applications in human rare disease and population genomics. The student will contribute to the dissemination of the developed tools by integrating them into the Galaxy ecosystem and producing documentation and training materials, promoting reproducible, accessible and interoperable long-read genomic analyses for the international research community.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Recent bibliography | <p>1: Tangaro MA, Chiara M, Pesole G, Zambelli F. Federated, governed, and interoperable? The emerging architecture of public human genomic data infrastructures: a European perspective. <i>Front Genet.</i> 2026 Apr 1;17:1819270. doi: 10.3389/fgene.2026.1819270. PMID: 41988645;</p> <p>2: Fonzino A, Fosso B, Visci G, Gissi C, Pesole G, Picardi E. Ab initio detection of multiple epitranscriptomic modifications from Oxford nanopore technology direct RNA sequencing data. <i>Brief Bioinform.</i> 2026 Jan 7;27(1):bbaf709. doi: 10.1093/bib/bbaf709. PMID: 41603648;</p> <p>3: Cox SN, Varvara AS, Pesole G. MitSorter: a standalone tool for accurate discrimination of mtDNA and NuMT ONT reads based on differential methylation. <i>Bioinform Adv.</i> 2025 Jul 10;5(1):vbaf135. doi: 10.1093/bioadv/vbaf135. PMID: 40688360;</p> <p>4: Mager S, et al. Genomic diversity in time and space in the toxic diatom <i>Pseudo-nitzschia multistriata</i>. <i>Harmful Algae.</i> 2025 Feb;142:102791. doi: 10.1016/j.hal.2024.102791. Epub 2024 Dec 22. PMID: 39947866.</p> <p>5: Traversa D, Pazzani C, D'Addabbo P, Trisolini L, Chiara M, Oliva M, Marzella A, Mandorino C, Calia C, Chimienti G, Manzari C, Pesole G, Scrascia M. De Novo Assembly of the Polyhydroxybutyrate (PHB) Producer <i>Azohydromonas lata</i> strain H1 Genome and Genomic Analysis of PHB Production Machinery. <i>Microorganisms.</i> 2025 Jan 10;13(1):137. doi: 10.3390/microorganisms13010137. PMID: 39858905;</p> <p>6: Mc Cartney AM, et al. The European Reference Genome Atlas: piloting a decentralised approach to equitable biodiversity genomics. <i>NPJ Biodivers.</i> 2024 Sep 17;3(1):28. doi: 10.1038/s44185-024-00054-6. Erratum in: <i>NPJ Biodivers.</i> 2024</p> |

## AVAILABLE POSITIONS

|                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                         | <p>Oct 15;3(1):31. doi: 10.1038/s44185-024-00065-3. PMID: 39289538;</p> <p>7: Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J, Malukiewicz J, Mouton A, Oomen RA, Paez S, Palsbøll PJ, Pampoulie C, Ruiz-López MJ, Svardal H, Theofanopoulou C, de Vries J, Waldvogel AM, Zhang G, Mazzoni CJ, Jarvis ED, Bálint M; European Reference Genome Atlas (ERGA) Consortium. The era of reference genomes in conservation genomics. Trends Ecol Evol. 2022 Mar;37(3):197-202. doi: 10.1016/j.tree.2021.11.008. Epub 2022 Jan 24. PMID: 35086739;</p> <p>8: Tangaro MA, Mandreoli P, Chiara M, Donvito G, Antonacci M, Parisi A, Bianco A, Romano A, Bianchi DM, Cangelosi D, Uva P, Molineris I, Nosi V, Calogero RA, Alessandri L, Pedrini E, Mordenti M, Bonetti E, Sangiorgi L, Pesole G, Zambelli. Laniakea@ReCaS: exploring the potential of customisable Galaxy on-demand instances as a cloud-based service. BMC Bioinformatics. 2021 Nov 8;22(Suppl 15):544. doi: 10.1186/s12859-021-04401-3. PMID: 34749633;</p> |
| Group composition       | 5 fellows, 2 PhD, 10 postdocs                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Institutional page link | <a href="http://www.uniba.it">www.uniba.it</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Lab website link        | <a href="https://www.uniba.it/docenti/pesole-graziano">https://www.uniba.it/docenti/pesole-graziano</a>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Social media links      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| Video link              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |