Hybrid Genome Assembly and Integrative Multi-Omics of *Composting-Residue Isolates* Reveal Bioplastic Degradation Potential

Martina Castellucci
MSc Bioinformatics, University of Bologna

Academic Year 2024/2025

Project budget: €200,000. This project was designed within a fixed budget, allocating resources to hybrid genome sequencing (Nanopore+Illumina), RNA-Seq differential expression, functional and comparative genomics, and dissemination. The main cost drivers are sequencing platforms (long-read flow cells and short-read lanes) and personnel (wet lab & bioinformatics), reflecting the balance between experimental depth and financial constraints.

Abstract

Bioplastics such as polylactic acid (PLA) and polyhydroxyalkanoates (PHA) are marketed as environmentally friendly substitutes for petroleum-derived polymers, their breakdown in composting environments is still incomplete and frequently slow. An integrative multi-omics study on fungal isolates from PLA/PHA-enriched compost is described in this project, along with its design and anticipated results. A subset of the twenty or so isolates that pass the esterase activity screening will undergo comparative genomics, functional annotation, and hybrid genome sequencing (Illumina short reads + Oxford Nanopore long reads). The identification of upregulated hydrolytic enzymes will be accomplished through transcriptomic profiling under PLA, PHA, and control conditions. A molecular framework for fungal-driven bioplastic degradation is anticipated to be provided by the integration of transcriptomic, phylogenomic, and genomic data, which will also highlight potential enzymes (such as lipases, cutinases, and esterases). These findings will support compost-based bioplastic waste valorization strategies and support the objectives of the EU circular bioeconomy.

Study Highlights

- Sampling of PLA/PHA-enriched compost → isolation of ∼20 fungal strains.
- Screening via Rhodamine B assay; ∼5 isolates expected to show strong degradative activity.
- Hybrid genome assemblies (Nanopore + Illumina) predicted to reach high contiguity and completeness (BUSCO >90%).
- Functional annotation (MAKER3, dbCAN3) to reveal CAZyme repertoires including candidate esterases, cutinases, lipases.
- Comparative/phylogenomics to distinguish genes common to degraders but absent in non-degraders.
- RNA-Seq design: 12 libraries (Control, PLA, PHA, Blank × replicates) to capture induction of degradative enzymes.

Keywords: hybrid sequencing; fungal isolates; CAZymes; differential expression; OrthoFinder; compost bioremediation; PLA/PHA degradation.

1 Introduction

1.1 Environmental background and rationale

PLA and PHA have varying decomposition rates in actual composting systems, while being promoted as biodegradable. The shift to a circular bioeconomy is hampered by incomplete mineralization, which causes plastic fragments to persist. Filamentous fungus are interesting biological agents for catalyzing polyester depolymerization because of their capacity to release extracellular hydrolases. However, there are still little genomic and transcriptome data available for fungi that are naturally evolved to habitats with a lot of plastic, which limits the ability to find enzymes systematically and to rationally design bioprocesses for better breakdown.

1.2 Organismal context

One of the most varied sources of possible bioplastic degraders is the fungi that grow in compost that has been enhanced with PLA and PHA leftovers. It is anticipated that a large number of isolates will have multitrophic lifestyles and diverse enzyme repertoires, including lipases, cutinases, esterases, and other hydrolases important for the breakdown of polyester. Novel enzymatic pathways can be found and their regulation in response to synthetic polymers evaluated by integrating hybrid genome sequencing with functional annotation, transcriptomics, and comparative analysis.

1.3 Project objectives

This project aims to: (i) collect and plate compost samples to obtain approximately 20 fungal isolates; (ii) screen them for esterase activity and select ~ 5 with the strongest degradative potential; (iii) generate high-quality hybrid genome assemblies for these candidate isolates; (iv) annotate carbohydrate-active enzymes (CAZymes), with a focus on hydrolytic families; (v) perform RNA-Seq under PLA and PHA exposure to identify inducible candidates; and (vi) integrate comparative genomics and phylogenomics to distinguish features specific to plastic-degrading fungi from those of non-degrading relatives.

2 Materials and Methods

2.1 Sampling, isolation, and preliminary screening

In Bologna, Italy, compost material that is clearly enriched in PLA and PHA pieces will be sampled and homogenized. After plating on PDA medium supplemented with powdered PLA and Rhodamine B, about 20 fungal isolates are anticipated. With its clear fluorescent halos signifying degradative potential, the Rhodamine B assay under UV light will perform as a quick stand-in for extracellular esterase activity. About five of these isolates with the highest activity will be kept for further examination. To confirm taxonomic placement at the genus level, preliminary ITS rDNA sequencing and BLAST-based comparison will be carried out.

2.2 Nucleic acid extraction and QC

An approach based on CTAB and tailored for filamentous fungi will be used to extract high molecular weight genomic DNA. To guarantee fragment sizes more than 20 kb, quality control will use agarose gel electrophoresis, Qubit quantification, and Nanodrop spectrophotometry. Fungal cultures cultivated in four biological replicates under three experimental conditions (Control media, PLA-supplemented medium, and PHA-supplemented medium) will have their total RNA harvested for transcriptome analysis, yielding a total of 12 RNA-Seq libraries. The Bioanalyzer will be used to confirm the integrity of the RNA before the library is prepared.

2.3 Sequencing strategy

To optimize genome accuracy and consistency, a hybrid sequencing strategy will be used. High-depth short reads appropriate for transcriptome sequencing and assembly polishing will be produced by the Illumina NovaSeq PE150. Long readings will be made available by Oxford Nanopore GridION to address structural variations and repetitive elements. It is anticipated that this dual-platform architecture would produce reference-quality assemblies that are complete and consistent.

2.4 Assembly, polishing, and quality control

Flye (v2.9) will be used to assemble the nanopore readings, and then Illumina reads will undergo several iterations of Pilon polishing. QUAST will be used to evaluate assembly quality and contiguity, while BUSCO (fungi_odb10 dataset) will be used to estimate completeness. A high BUSCO score (estimated $> 90\%$) is anticipated given the hybrid technique, which offers assurance in the representation of gene space.

2.5 Structural and functional annotation

RNA-Seq alignments and ab initio predictors (Augustus, GeneMark-ES) will be integrated as supporting evidence for the genome annotation process, which will be carried out using MAKER3. Using dbCAN3, functional annotation will concentrate on CAZymes, allowing for the identification of lipases, cutinases, and carbohydrate esterases. Geneious will be used for manual gene model validation and inspection.

2.6 Comparative genomics and phylogenomics

To determine which gene families are shared and which are isolate-specific, orthogroup inference will be performed using OrthoFinder. We will investigate synteny and collinearity among related species using MCScanX and MAUVE. Multiple single-copy orthologs aligned by MAFFT, curated with Gblocks, concatenated with AMAS, and evaluated with RAxML (maximum likelihood) and MrBayes (Bayesian inference) will be used

for phylogenomic reconstruction. In order to identify gene gains or losses linked to the ability to digest bioplastics, these analyses will compare degraders with non-degraders.

2.7 Transcriptomics: quantification and DEG analysis

Salmon will be used to quantify RNA-Seq reads, and tximport will be used to compile gene-level expression matrices. DESeq2 will be used to perform differential expression analysis, with significant criteria of FDR <0.05 and $|\log_2 \mathrm{FC}| \geq 2$ set. It is anticipated that this research will identify genes, such as those belonging to the hydrolytic CAZyme groups, that are especially elevated under PLA and PHA conditions.

3 Expected Results

The integration of hybrid genome sequencing, annotation, comparative genomics, and transcriptomics is intended to offer a comprehensive assessment of the enzymatic capacity of fungal isolates for bioplastic breakdown. From the $\sim$ 20 compost-derived isolates, $\sim$ 5 will likely be identified as powerful degraders based on Rhodamine B assays and early screening. They should be dependable references for further investigation because their hybrid assemblies should yield highly contiguous genomes with a limited number of contigs, high N50 values, and BUSCO completeness >90%. In accordance with the postulated degradative role, functional annotation is anticipated to reveal a rich CAZyme repertoire, including a notable representation of cutinases, esterases, and lipases, as well as putative PHA depolymerase-like enzymes. Comparative genomics with closely related fungi, including non-degrading isolates, should reveal both shared orthogroups and lineage-specific genes, affording insights into gene gains and losses associated with adaptation to polymer-rich environments. The degraders' evolutionary uniqueness will be further elucidated by their phylogenomic positioning. Numerous differentially expressed genes are anticipated under PLA and PHA treatments, many of which would overlap with the identified CAZyme families, according to transcriptomic analysis of 12 RNA-Seq libraries, which will give dynamic evidence of gene regulation. A logical model in which certain fungi use an inducible enzymatic arsenal for PLA/PHA degradation is expected to be supported by the convergence of transcriptomic induction and genomic predictions. This will highlight both particular candidate enzymes for downstream validation and the ecological strategies underlying fungal adaptation to synthetic polymer niches.

4 Discussion

4.1 Strengths of the project design

This project demonstrates several strong aspects. First, a **hybrid sequencing strategy** (Illumina + Nanopore) is used to optimize accuracy and contiguity, which lays a strong basis for further analysis. We may investigate plastic-degradation potential from complementary perspectives by integrating the **multi-omics layers** (genome, transcriptome, comparative genomics, and phylogenomics). With several conditions and replicates, the RNA-Seq design guarantees reliable identification of genes with differential expression and facilitates the functional interpretation of potential hydrolases. Furthermore, by placing candidate genes within evolutionary links, comparative and phylogenomic techniques reinforce the case for discovering adaptations specific to degradation. The project's applied usefulness beyond academic interest is further reinforced by its alignment with **bioeconomy and EU policy goals**.

4.2 Limitations and future perspectives

The project has inherent limitations despite its resilience. Even while hybrid assembly is probably responsible for the high predicted genome completeness, certain lineage-specific or repetitive elements may still be missed, which could result in an underestimation of the gene content. The requirement for experimental validation is highlighted by the fact that annotation pipelines that rely on domain homology, such as CAZy/dbCAN, can **overpredict enzyme functions**. Biochemical experiments (such as enzyme expression and polymer degradation tests) must be used in conjunction with transcriptomic induction patterns since they are informative but do not demonstrate catalytic activity. The quality and availability of reference genomes from closely related fungi limit the assumptions of comparative genomics. Lastly, although sequencing and analysis are covered by the budget, downstream validation—such as protein engineering or pilot-scale biodegradation assays—that would be required to convert results into commercial applications is not.

4.3 Overall assessment

In conclusion, the project design successfully identifies potential enzymes for PLA/PHA degradation by fusing integrative analysis and high-throughput sequencing. Its key drawbacks are its dependence on computer projections and lack of experimental confirmation, while its advantages are methodological breadth, expected assembly quality, and policy significance. The practical breadth of a genomics-driven discovery pipeline is reflected in this balance of **strengths and weaknesses**, which also offers a solid basis for next bioengineering and applied biodegradation initiatives.

5 Conclusion

This project outlines a genome-to-function pipeline for fungal isolates from PLA/PHA-enriched compost. By integrating hybrid sequencing, comparative genomics, and transcriptomics, it provides a framework to identify candidate enzymes for bioplastic degradation. While results remain predictive and require experimental validation, the approach establishes a strong basis for future

biotechnological applications in sustainable waste management.

Single-Page Table: Project Costs

Table 1: Estimated budget for the project (Euro). Items at 0 € highlight open-source tool usage.

Item	Cost	Description
Wet lab Postdoc salary	45,000	One-year contract for lab activities
Sampling & Isolation	3,000	Compost collection, plating, screening
DNA Extraction & QC	2,000	CTAB reagents, plasticware, gels
Illumina PE150 Sequencing	30,000	Short-read library prep & sequencing
Oxford Nanopore Sequencing	50,000	Flow cells & library kits
RNA extraction (12 libraries)	1,000	TRIzol/silica columns
RNA library preparation (12)	1,500	Illumina TruSeq kits
RNA-Seq Illumina (12)	2,500	PE150 sequencing
Bioinformatics Postdoc	45,000	One-year contract (analysis & reporting)
Genome Assembly & Polishing	0	Flye + Pilon (open-source)
Functional Annotation	0	MAKER3, dbCAN3, Geneious (manual curation)
Comparative Genomics	0	OrthoFinder, MCScanX, MAUVE
Non-open source softwares	5,000	Geneious & other licenses
Reports & Dissemination	15,000	Publications, conferences, outreach
Total	**200,000**	

References

Scientific Articles

1. Menicucci, A., Iacono, S., Ramos, M., Fiorenzani, C., Peres, N. A., et al. (2025). Can whole genome sequencing resolve taxonomic ambiguities in fungi? The case study of *Colletotrichum* associated with ferns. *Frontiers in Fungal Biology*, 6, 1540469. DOI: 10.3389/ffunb.2025.1540469

2. Prasad, P., Varshney, D., & Adholeya, A. (2015). Whole genome annotation and comparative genomic analyses of bio-control fungus *Purpureocillium lilacinum*. *BMC Genomics*, 16(1), 1004. DOI: 10.1186/s12864-015-2229-2

3. Ekanayaka, A. H. et al. (2025). Linking metabolic pathways to plastic-degrading fungi: a comprehensive review. *Journal of Fungal Biology*. DOI: 10.3390/jof11050378

Main Tools Used

Assembly

- **Flye** – Long-read genome assembler. https://github.com/fenderglass/Flye
- **Pilon** – Assembly polishing with Illumina reads. https://github.com/broadinstitute/pilon

Quality Control (QC)

- **QUAST** – Assembly quality assessment. http://quast.sourceforge.net/
- **BUSCO** – Benchmarking Universal Single-Copy Orthologs. https://busco.ezlab.org/

Annotation

- **MAKER3** – Genome annotation pipeline. http://www.yandell-lab.org/software/maker.html
- **dbCAN3** – CAZyme annotation server. https://bcb.unl.edu/dbCAN2/

Comparative Genomics

- **OrthoFinder** – Orthogroup inference. https://github.com/davidemms/OrthoFinder
- **MCScanX** – Synteny/collinearity analysis. https://github.com/wyp1125/MCScanX

Transcriptomics

- **Salmon** – Transcript quantification. https://salmon.readthedocs.io/
- **DESeq2** – Differential expression analysis (R/Bioconductor). https://bioconductor.org/packages/release/bioc/html/DESeq2.html

Visualization

- **R / Python** – Plotting and downstream analysis. https://www.r-project.org/ — https://www.python.org/
- **Geneious** – Manual curation and visualization. https://www.geneious.com/