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 sustainable substitutes for petroleum-derived polymers; however, their enzymatic depolymerization in composting environments remains inefficient and often rate-limiting. 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) that is designed to generate highly contiguous and complete genome assemblies. The identification of upregulated hydrolytic enzymes will be accomplished through transcriptomic profiling under PLA, PHA, and control conditions. Integrating genomic potential with transcriptomic response is anticipated to provide a powerful gene-to-function framework, directly linking phylogenomic context and inducible gene expression to identify high-priority enzyme candidates (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 × 3 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. A systematic discovery pipeline that links genomic potential to functional expression is therefore crucial to bridge this knowledge map and exploit these fungi for applied bioremediation.

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 genome-wide repertoires, with a dedicated focus on hydrolytic families implicated in polyester breakdown; (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 were grown in three biological replicates under four experimental conditions (Control, PLA, PHA, and Blank), with plastic conditions sampled at 12 h, 24 h, and 36 h, while Control was collected only as a baseline. 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

After assembling the nanopore readings using Flye (v2.9), the Illumina reads will go through multiple Pilon polishing iterations. BUSCO (fungi_odb10 dataset) will be used to estimate completeness, and QUAST will be used to assess assembly quality and contiguity. Given the hybrid technique, which provides assurance in the representation of gene space, a high BUSCO score (estimated $> 90\%$) is expected.

2.5 Structural and functional annotation

MAKER3 will be used to carry out the genome annotation process, and RNA-Seq alignments and ab initio predictors (Augustus, GeneMark-ES) will be incorporated as supporting evidence. This evidence-driven approach integrates ab initio predictions with experimental support (RNA-Seq alignments), significantly enhancing the accuracy and reliability of the gene models compared to ab initio methods alone. Functional annotation will focus on CAZymes using dbCAN3, enabling the identification of lipases, cutinases, and carbohydrate esterases. Geneious will be used for manual inspection and validation of gene models.

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 goal of combining transcriptomics, annotation, comparative genomics, and hybrid genome sequencing is to provide a thorough evaluation of the enzymatic ability of fungal isolates to break down bioplastics. Based on Rhodamine B assays and early screening, ∼5 of the ∼20 isolates derived from compost are probably going to be found to be strong degraders. In addition, enrichment of specific fungal genera is expected in PLA/PHA compost (see Table 2 for a literature-based preview). Given that their hybrid assemblies are expected to produce highly contiguous genomes with a small number of contigs, high N50 values, and BUSCO completeness >90%, these high-quality genomes will serve as unambiguous references for functional annotation and robust comparative genomics. In accordance with their postulated degradative role, functional annotation is expected 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, providing insights into gene gains and losses associated with adaptation to polymer-rich environments. The degraders' evolutionary uniqueness will be further clarified by their phylogenomic positioning. Numerous differentially expressed genes are anticipated under PLA and PHA treatments, many overlapping with the identified CAZyme families. Transcriptomic analysis of 12 RNA-Seq libraries will thus provide dynamic evidence of gene regulation. A coherent model in which certain fungi deploy 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 exhibits a number of strong points. First, to maximize accuracy and continuity, a **hybrid sequencing strategy** (Illumina + Nanopore) is employed, providing a solid foundation for additional analysis. The integrative multi-omics approach is a key strength, enabling a causal link between genomic endowment (what enzymes are present) and transcriptomic response (which are deployed during degradation), all within a phylogenomic context that reveals evolutionary adaptations. The RNA-Seq design allows for the functional interpretation of putative hydrolases and ensures accurate identification of genes with differential expression under a variety of conditions and replicates. Furthermore, comparative and phylogenomic methods support the argument for finding degradation-specific adaptations by situating candidate genes within evolutionary links. The project's alignment with **bioeconomy and EU policy goals** further supports its applied usefulness beyond academic interest.

4.2 Limitations and future perspectives

Despite its resilience, the project has inherent limitations. Although the high predicted genome completeness is likely due to hybrid assembly, some repetitive or lineage-specific elements might still be overlooked, which could lead to an underestimation of the gene content. Annotation pipelines based on domain homology, like CAZy/dbCAN, can generate false-positive predictions of enzyme function, underscoring that in silico findings constitute a prediction that must be confirmed through heterologous expression and enzymatic assays. Since transcriptomic induction patterns are informative but do not show catalytic activity, biochemical experiments (such as enzyme expression and polymer degradation tests) must be used in conjunction with them. Comparative genomics' presumptions are constrained by the caliber and accessibility of reference genomes from closely related fungi. 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 summary, by combining integrative analysis and high-throughput sequencing, the project design effectively finds putative enzymes for PLA/PHA degradation. Methodological breadth, expected assembly quality, and policy significance are its main advantages; its main disadvantages are its reliance on computer projections and lack of experimental confirmation. This balance of **strengths and weaknesses** reflects the practical breadth of a genomics-driven discovery pipeline and provides a strong foundation for future applied biodegradation and bioengineering efforts.

5 Conclusion

The genome-to-function pipeline for fungal isolates from PLA/PHA-enriched compost is described in this project. It offers a framework for identifying potential enzymes for the breakdown of bioplastics by combining transcriptomics, comparative genomics, and hybrid sequencing. The method provides a solid foundation for upcoming biotechnological applications in sustainable waste management, even though the results are still predictive and need experimental validation.

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**	

Table 2: Expected fungal candidates enriched in PLA/PHA-containing environments (literature-based preview).

Phylum	**Representative taxa**	**Key degradative traits / phylogenetic link**
Ascomycota	*Aspergillus* spp.	Esterase-rich; known PLA degraders; convergence in esterases.
	Fusarium spp.	Cutinases and depolymerases; pathway signatures detectable by comparative genomics.
	Penicillium spp.	Broad CAZyme repertoire; orthogroups distinct between degraders and non-degraders.
	Purpureocillium spp.	Documented polyester degradation; links to biocontrol clade.
Basidiomycota	*Phanerochaete chrysosporium*	White-rot lignin/peroxide system; oxidative auxiliary enzymes.
	Trametes spp.	Laccase-rich; oxidative contribution to polymer breakdown.
Mucoromycota	*Rhizopus* spp.	Noted esterase activity; basal lineage with independent esterase expansion.

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

4. European Environment Agency (2020). *Biodegradable and compostable plastics – challenges and opportunities*. EEA Briefing No. 09/2020, Publications Office of the European Union, 10.2800/552241.

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.h

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/