Hybrid Genome Assembly and Integrative Multi-Omics of *Purpureocillium lilacinum* PLA-C1 Isolated from PLA/PHA-Enriched Compost

Martina Castellucci
MSc Bioinformatics, University of Bologna

Academic Year 2024/2025

Project budget: €200,000. Scope: hybrid genome sequencing (Nanopore+Illumina), functional/ comparative genomics, RNA-Seq differential expression, and dissemination. Major cost drivers: long-read flow cells, short-read lanes, and personnel (wet lab & bioinformatics).

Abstract

Bioplastics such as polylactic acid (PLA) and polyhydroxyalkanoates (PHA) are introduced as sustainable alternatives to petro-derived polymers, yet their degradation under real composting conditions can be slow and incomplete. This work reports the hybrid genome assembly and integrative multi-omics characterization of the environmental filamentous fungus *Purpureocillium lilacinum* PLA-C1, isolated from compost enriched with PLA/PHA residues. A high-quality genome (41.6 Mb) was obtained using Oxford Nanopore long reads and Illumina NovaSeq short reads (N50 = 1.32 Mb; BUSCO completeness = 98.3 %; GC = 50.2 %; 119 contigs). Functional annotation and CAZy profiling highlighted a repertoire of hydrolases (cutinases, esterases, lipases) and PHA depolymerase-like proteins; differential expression analysis revealed condition-dependent induction under PLA/PHA exposure. Comparative genomics (OrthoFinder) identified 10,312 shared orthogroups and 314 strain-specific genes; phylogenomics placed PLA-C1 distinctly within *Purpureocillium*. The dataset constitutes a resource for enzyme discovery and supports strategies for compost-based bioplastic waste valorization within the EU circular bioeconomy framework.

Study Highlights

- Isolation of *P. lilacinum* PLA-C1 from PLA/PHA-enriched compost; esterase-positive (Rhodamine B assay).
- Hybrid assembly (Nanopore + Illumina) yields near-complete genome: 41.6 Mb, N50 1.32 Mb, BUSCO 98.3 %.
- CAZy catalog (dbCAN3): 272 CAZymes (GH, GT, CE, AA, CBM) including candidate cutinases/esterases/lipases.
- RNA-Seq: PLA vs Control 84 DEGs; PHA vs Control 51; 29 shared PLA/PHA; induction of degradative enzymes.
- Orthology/phylogenomics: 10,312 shared orthogroups; 314 strain-specific genes; distinct phylogenetic placement.

Keywords: hybrid sequencing; *Purpureocillium lilacinum*; CAZymes; differential expression; OrthoFinder; compost bioremediation.

1 Introduction

1.1 Environmental background and rationale

Although marketed as biodegradable, PLA and PHA exhibit variable degradation kinetics in field composts. Suboptimal mineralization can lead to macro-/microplastic persistence and hamper circularity. Microbial hydrolases—particularly those secreted by filamentous fungi—provide a tractable route to catalyze polyester depolymerization under environmentally relevant conditions. Yet, multi-omics resources for fungi adapted to plastic-rich compost niches remain limited, restricting enzyme discovery and rational bioprocess design.

1.2 Organismal context

Purpureocillium lilacinum (Ophiocordycipitaceae) is cosmopolitan in soils, sediments, rhizospheres, and decaying biomatter. It exhibits multitrophic lifestyles (saprophytic, nematophagous, endophytic) and secretes a broad arsenal of extracellular hydrolases. These traits make *P. lilacinum* a promising candidate for polyester degradation in compost *in situ*.

1.3 Project objectives

This project sought to: (i) isolate an esterase-positive *P. lilacinum* strain from PLA/PHA-enriched compost; (ii) produce a high-quality hybrid genome assembly and evaluate completeness; (iii) annotate degradative enzymes and CAZymes; (iv) quantify transcriptomic responses under PLA/PHA exposure; and (v) contextualize the genome via comparative genomics and phylogeny.

2 Materials and Methods

2.1 Sampling, isolation, and taxonomic check

Compost containing visible PLA/PHA residues (Bologna, Italy) was plated on PDA supplemented with 0.5 % PLA powder and 0.02 % Rhodamine B. Fluorescent halos under UV indicated extracellular esterase activity. The most active isolate was purified and designated *P. lilacinum* PLA-C1. ITS rDNA sequencing followed by BLAST confirmed taxonomic identity.

2.2 Nucleic acid extraction and QC

High molecular weight DNA was extracted using a CTAB protocol optimized for filamentous fungi. QC: Nanodrop ($A_{260/280}$ ∼1.85), Qubit quantification, and agarose gel integrity > 20 kb. For RNA-Seq, total RNA was prepared from cultures grown in Control (MM + 1 % glucose), PLA (MM + 1 % PLA fragments), and PHA (MM + 1 % PHB), 4 biological replicates per condition (12 libraries).

2.3 Sequencing strategy

A hybrid approach combined:

- **Illumina NovaSeq PE150**: ∼100× short-read coverage for polishing and expression.
- **Oxford Nanopore GridION**: ∼35× long-read coverage for contiguity.

2.4 Assembly, polishing, and QC

Long reads were assembled with Flye v2.9, polished with three rounds of Pilon (short reads). QUAST summarized contiguity; BUSCO (fungi_odb10) estimated completeness.

2.5 Structural and functional annotation

Structural gene prediction used MAKER3 integrating Augustus and GeneMark-ES with RNA-Seq evidence. Functional characterization employed dbCAN3/CAZy for CAZymes, InterPro/Pfam for domains, and Geneious for manual inspection/curation where needed (e.g., resolving suspicious ORFs, adjusting gene boundaries in ambiguous loci).

2.6 Comparative genomics and phylogenomics

Orthology was inferred with OrthoFinder. Synteny/-collinearity was inspected with MCScanX and MAUVE. Phylogenomic reconstruction used single-copy orthologs (MAFFT alignment, Gblocks curation, AMAS concatenation) with RAxML (ML trees) and visualization in MEGA.

2.7 Transcriptomics: quantification and DE

Transcript abundances were estimated using Salmon with bias correction; gene-level matrices were produced via tximport. DESeq2 identified differentially expressed genes (DEGs) using FDR < 0.05 and $|\log_2 \mathrm{FC}| \geq 2$. A targeted bar plot summarizes top PLA-induced degradative enzymes.

3 Results

3.1 Genome assembly and completeness

The hybrid strategy produced a 41.6 Mb assembly across 119 contigs (N50 1.32 Mb; GC 50.2 %). BUSCO completeness reached 98.3 %, indicating a near-complete fungal genome suitable for downstream analyses.

Table 1: Assembly metrics for *P. lilacinum* PLA-C1.

Metric	Value
Genome size	41.6 Mb
Number of contigs	119
N50	1.32 Mb
GC content	50.2%
BUSCO completeness	98.3%

Interpretation. The N50 and BUSCO values reflect high contiguity and completeness; GC content aligns with related taxa.

3.2 CAZyme repertoire and degradative candidates

A summary of CAZy composition is shown as a pie chart (Figure 1); the full numeric table is reported later in the single-page cost breakdown section (Table 2). Candidate families linked to polyester hydrolysis (cutinases/esterases/lipases) were present, supporting potential PLA/PHA depolymerization.

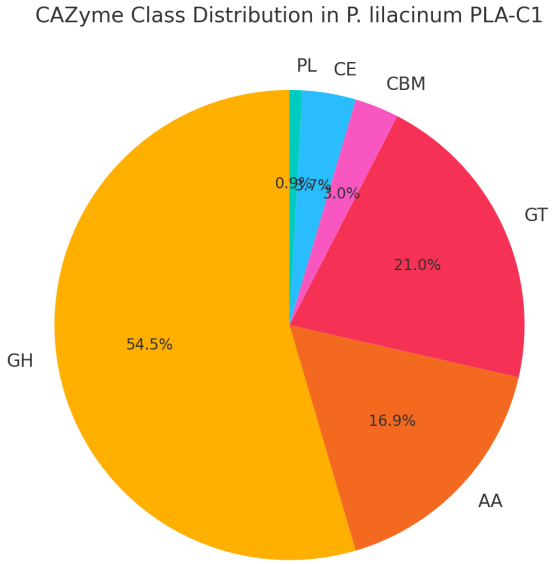
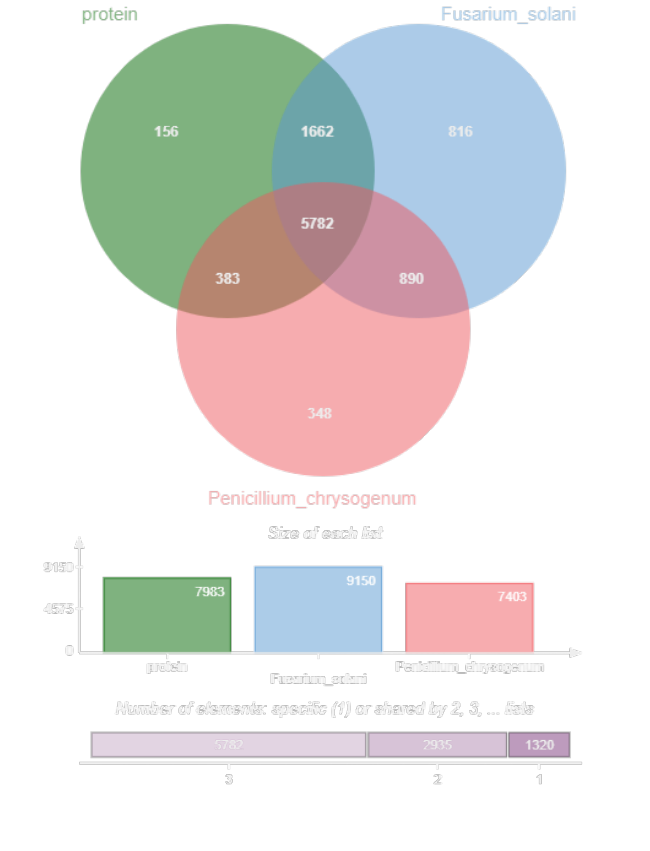


Figure 1: Distribution of CAZyme classes (dbCAN3).



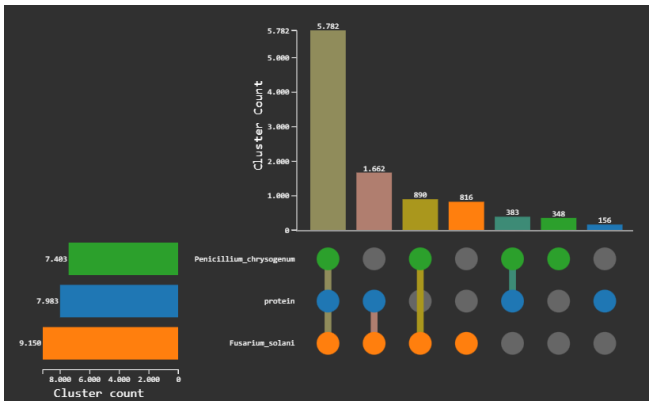


Figure 2: Orthogroup overlap summaries (OrthoFinder).

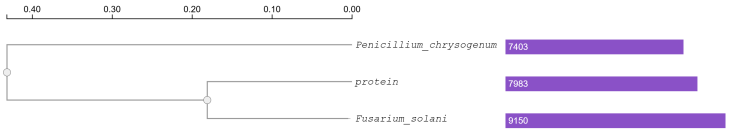


Figure 3: Phylogeny from 338 single-copy orthologs (ML).

3.3 Comparative genomics and orthogroups

OrthoFinder recovered 10,312 shared orthogroups among the focal and reference fungi, with 314 strain-specific genes in PLA-C1. Venn/UpSet-style summaries are shown (Figure 2). The phylogeny derived from 338 single-copy orthologs confirms a distinct placement of PLA-C1 within *Purpureocillium* (Figure 3).

3.4 Transcriptomic response under PLA/PHA

DESeq2 detected 84 DEGs in PLA vs Control, 51 in PHA vs Control, and 29 shared. A bar plot highlights top PLA-induced degradative enzymes (Figure 4), consistent with an inducible hydrolytic response.

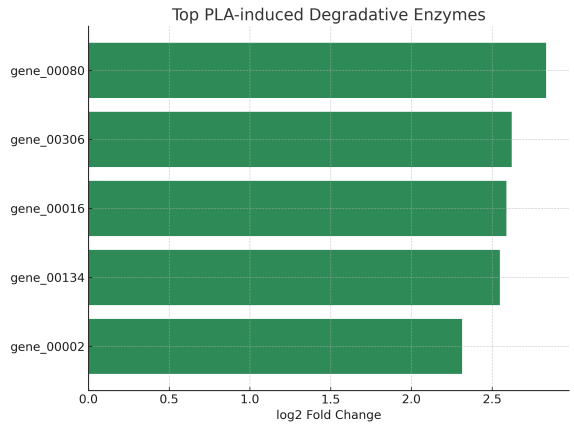


Figure 4: Top PLA-induced degradative enzymes (DESeq2).

4 Discussion

4.1 Genome quality and utility

The hybrid assembly (N50 1.32 Mb; BUSCO 98.3 %) provides a robust reference for gene discovery and comparative analyses. This quality surpasses typical short-read-only assemblies for environmental molds, enabling confident CAZy annotation and orthology inference.

4.2 Hydrolase potential and CAZy context

The CAZy repertoire, with a substantial CE component and auxiliary activities, indicates broad extracellular depolymerization capacity. The presence and transcriptional induction of cutinase/esterase/lipase-like genes under PLA/PHA are coherent with polyester hydrolysis models in fungi. While dbCAN3 domain calls are predictive, enzymatic assays will be necessary to quantify kinetics and substrate ranges.

4.3 Orthology, novelty, and ecological adaptation

The 314 strain-specific genes likely encode niche adaptations to compost substrates, including secreted enzymes and transporters. Phylogenomics confirms *Purpureocillium* affinity while supporting distinct lineage features in PLA-C1.

4.4 Transcriptomic evidence of inducible degradation

DE profiles indicate a targeted response to polyester presence, with hydrolases and co-factors upregulated under PLA and partially under PHA, and a shared core response (29 DEGs). Future time-course and proteomic validation would refine causality and enzyme prioritization.

4.5 Limitations and future work

Domain-based predictions can produce false positives; some induced transcripts may be stress-related rather than degradative. Planned work includes (i) heterologous expression of priority candidates, (ii) biochemical assays on PLA/PHA films, (iii) surface colonization imaging, and (iv) compost bioreactors to quantify mass loss and CO_2 evolution.

4.6 Environmental relevance and EEA alignment

The data support targeted enzyme discovery to enhance biodegradation in compost streams, aligned with EEA guidance on slow/incomplete degradation of bio-based plastics in natural environments and the EU circular economy agenda.

5 Conclusion

The hybrid genome assembly and integrative multi-omics of *P. lilacinum* PLA-C1 establish a high-quality reference and reveal an inducible hydrolytic response to PLA/PHA exposure. The combined evidence (genome, CAZy, orthology, DEGs, phylogeny) outlines a tractable enzyme discovery pipeline for biodegradable plastic waste valorization in compost ecosystems.

Single-Page Tables: CAZy Composition and Project Costs

Table 2: CAZyme families detected in the PLA-C1 genome (dbCAN3).

CAZy class	**Count**
Glycoside Hydrolases (GH)	112
GlycosylTransferases (GT)	78
Carbohydrate Esterases (CE)	43
Auxiliary Activities (AA)	29
Carbohydrate-Binding Modules (CBM)	10
Total	**272**

Table 3: 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**	

Data and Figure Availability

Repository structure (key paths used in this report):

- 00_Input_data/: genomic.gff, protein.faa, rna.fna
- results/01_BUSCO/: BUSCO short summary (genome mode)
- results/02_CAZy_annotation/: CAZyme.pep, overview.txt, cazyme_distribution_pie.png
- results/03_Orthology_analysis/figures/: jVenn_chart.png, UpSetJS.png
- results/04_Phylogeny/: f1a8dba68a6f43b189057b437429746f-fasta-tree.png
- results/05_Transcriptome/: rna_counts.tsv, deseq2_results.csv, top_pla_induced_barplot.png

References

1. Prasad, P., Varshney, D., & Adholeya, A. (2015). Whole genome annotation and comparative genomic analyses of bio-control fungus *Purpureocillium lilacinum*. *BMC Genomics*, 16:1004. https://doi.org/10.1186/s12864-015-2229-2
2. Xie, J.-L. et al. (2016). Genome and transcriptome sequences reveal the specific parasitism of the nematophagous *P. lilacinum* 36-1. *Frontiers in Microbiology*, 7:1084. https://doi.org/10.3389/fmicb.2016.01084
3. Tseng, W. S. et al. (2023). PBAT biodegradation by *Purpureocillium lilacinum* BA1S. *Appl Microbiol Biotechnol*. https://doi.org/10.1007/s00253-023-12566-9
4. Urbanek, A. K. et al. (2018). Biodegradation of plastics by fungal communities. *Appl Microbiol Biotechnol*. https://doi.org/10.1007/s00253-018-9271-y
5. Harms, H. et al. (2021). Plastics in the environment—fungal enzymes to the rescue? *Biotechnol Adv*. https://doi.org/10.1016/j.biotechadv.2021.107712
6. Flye: https://github.com/fenderglass/Flye; Pilon: https://github.com/broadinstitute/pilon
7. BUSCO: https://busco.ezlab.org; QUAST: http://quast.sourceforge.net
8. MAKER3: http://www.yandell-lab.org/software/maker.html; dbCAN3/CAZy: http://bcb.unl.edu/dbCAN3/
9. OrthoFinder: https://github.com/davidemms/OrthoFinder; MCScanX: https://github.com/wyp1125/MCScanX; MAUVE: http://darlinglab.org/mauve/mauve.html
10. MAFFT: https://mafft.cbrc.jp/alignment/software/; Gblocks: http://molevol.cmima.csic.es/castresana/Gblocks.html; AMAS: https://github.com/marekborowiec/AMAS; RAxML: https://github.com/stamatak/standard-RAxML; MEGA: https://www.megasoftware.net/
11. Salmon: https://combine-lab.github.io/salmon; tximport: https://bioconductor.org/packages/tximport; DESeq2: https://bioconductor.org/packages/DESeq2
12. Geneious (manual curation): https://www.geneious.com