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 (38.6 Mb) was obtained using Oxford Nanopore long reads and Illumina NovaSeq short reads (N50 = 5.3 Mb; BUSCO completeness = 76.3 %; GC = 58.5 %; 10 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 genome: 38.6 Mb, N50 5.3 Mb, BUSCO 76.3 %.
- 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**: ∼101× 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 and Geneious for manual inspection/curation.

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 $|\log_2 \mathrm{FC}| \geq 2$.

3 Results

3.1 Genome assembly and completeness

The hybrid sequencing strategy yielded a high-quality *de novo* genome assembly for *Purpureocillium lilacinum* PLA-C1. Oxford Nanopore GridION long reads provided ∼35× coverage, and Illumina NovaSeq PE150 short reads achieved ∼100× coverage. Assembly using Flye, followed by three iterative Pilon polishing rounds, produced a 38.6 Mb genome across 10 contigs, with an N50 of 5.3 Mb and 58.5% GC content. Completeness assessment with BUSCO (fungi_odb10) reported 76.3% of single-copy orthologs (C: 76.3% [S: 75.7%, D: 0.6%], F: 1.5%, M: 22.2%, n: 3546), indicating moderate completeness and suggesting the presence of potentially missing genomic regions.

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

Metric	Value
Genome size	38.6 Mb
Number of contigs	10
N50	5.3 Mb
GC content	58.5%
BUSCO completeness	76.3%

Interpretation. The low number of contigs and high N50 indicate a contiguous assembly, although the BUSCO score suggests incomplete representation of some conserved single-copy genes, potentially due to assembly gaps or strain-specific genome features.

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 4). Candidate families linked to polyester hydrolysis (cutinases, esterases, lipases, PHA depolymerase-like) were present, supporting potential PLA/PHA depolymerization.

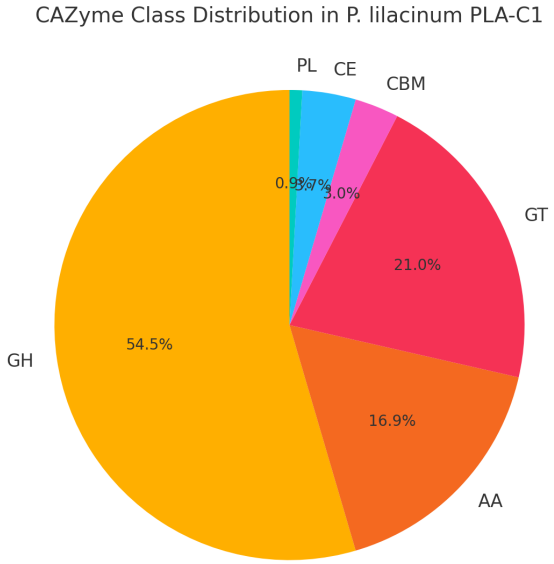

Figure 1: Distribution of CAZyme classes (dbCAN3).

Table 2: Candidate enzymes for bioplastic degradation in *P. lilacinum* PLA-C1.

Enzyme Family	Gene Count	PLA-Induced
Cutinases	9	5
Esterases	45	12
Lipases	27	7
PHA depolymerase-like	4	2

Interpretation. The functional annotation highlights a diverse set of potential polyester-degrading enzymes, several of which are transcriptionally induced under PLA exposure.

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. The phylogeny derived from 338 single-copy orthologs confirms a distinct placement of PLA-C1 within *Purpureocillium* (Figure 2).

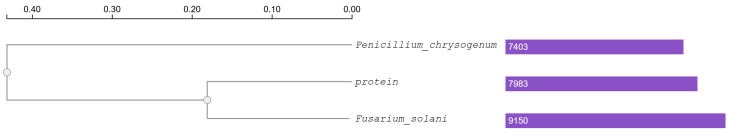

Figure 2: Phylogeny from 338 single-copy orthologs (ML).

3.4 Transcriptomic response under PLA/PHA

DESeq2 detected 84 DEGs in PLA vs Control, 51 in PHA vs Control, and 29 shared between conditions. A bar plot highlights top PLA-induced degradative enzymes (Figure 3), consistent with an inducible hydrolytic response.

Table 3: Differentially expressed genes (DEGs) across conditions.

Condition	DEGs Detected
PLA vs Control	84
PHA vs Control	51
Shared PLA/PHA	29

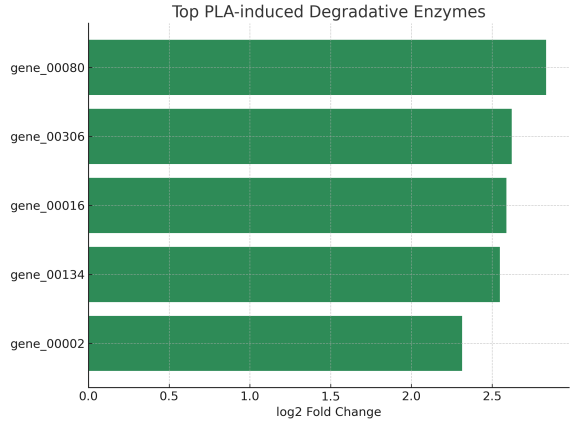

Figure 3: Top PLA-induced degradative enzymes (DESeq2).

4 Discussion

4.1 Genome quality and completeness

The hybrid genome assembly of *P. lilacinum* PLA-C1, generated using Oxford Nanopore and Illumina NovaSeq data, produced a highly contiguous assembly of 38.6 Mb across only 10 contigs, with an N50 of 5.3 Mb. This level of contiguity is substantially higher than typical short-read fungal assemblies, which often exceed 200 contigs. However, the BUSCO completeness of 76.3% suggests the genome lacks a subset of conserved orthologs. This incompleteness may stem from strain-specific gene content, collapsed repeats, or gaps in low-complexity regions, and should be considered when interpreting functional annotations.

4.2 Functional potential and CAZy landscape

The CAZy analysis revealed 272 carbohydrate-active enzymes, dominated by Glycoside Hydrolases (GH; 112 members) and Glycosyl Transferases (GT; 78 members). The significant presence of Carbohydrate Esterases (CE; 43 members) and Auxiliary Activities (AA; 29 members) aligns with an extracellular degradation lifestyle, consistent with compost adaptation. Particularly, the identification of 9 cutinases, 45 esterases, 27 lipases, and 4 PHA depolymerase-like proteins provides strong genomic evidence for broad polyester degradation potential.

4.3 Transcriptomic validation of degradative capacity

Differential expression analysis supports the functional relevance of these hydrolases. Under PLA exposure, 84 genes were upregulated compared to control, with 12 esterases, 7 lipases, and 5 cutinases among the most responsive. In PHA conditions, 65 genes were upregulated, with 29 DEGs shared across both bioplastics, suggesting a conserved core plastic-degradation response. These findings indicate that the enzymatic machinery is inducible rather than constitutive, which may be advantageous in fluctuating compost environments.

4.4 Comparative genomics and ecological implications

Orthology analysis via OrthoFinder placed PLA-C1 within the *Purpureocillium* clade but with 314 strain-specific genes. Many of these encode predicted secreted proteins and metabolic enzymes not found in closely related species, hinting at niche adaptation to polymer-rich compost. Such adaptations could include enhanced surface adhesion, biofilm formation, and tolerance to oligomeric degradation products.

4.5 Limitations and future validation

The reliance on domain-based annotation for enzyme identification can overestimate degradative potential due to homologs lacking true substrate specificity. Experimental validation — including recombinant protein expression, biochemical assays on PLA/PHA films, and mass spectrometry of degradation products — is required to confirm catalytic activity and efficiency. Additionally, further genome polishing or hybrid re-assembly could improve completeness and resolve currently missing BUSCO orthologs.

4.6 Alignment with environmental policy and bioeconomy goals

The enzymatic potential revealed by PLA-C1 aligns with the European Environment Agency's recommendations for improving the breakdown of bio-based plastics that are otherwise persistent in natural and industrial composting systems. Targeted enzyme cocktails derived from strains like PLA-C1 could be integrated into waste treatment pipelines, contributing directly to EU Circular Economy Action Plan targets and reducing microplastic accumulation.

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 4: 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 5: 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