Applied Genomics Project: Compliance Verification

Confirmation of Alignment with AG Course Requirements

Martina Castellucci
University of Bologna

August 12, 2025

The project **"Hybrid Genome Assembly and Integrative Multi-Omics of *Purpureocillium lilacinum* PLA-C1"** is fully aligned with the key theoretical and practical topics of the *Applied Genomics* course. It covers the complete bioinformatics and wet-lab pipeline: from targeted environmental sampling, through NGS-based genome and transcriptome analysis, to functional annotation and applied biotechnology outcomes. The table below links each project component to course topics and learning objectives.

1. Choice of Organism and Sampling Rationale

- **Project:** Isolation of *P. lilacinum* PLA-C1 from PLA/PHA-enriched compost, chosen for its extracellular esterase activity and potential in bioplastic degradation.
- **Course Topics:** Environmental metagenomics, targeted organism selection, industrial biotechnology applications.
- **Justification:** Selecting an organism from a polymer-rich ecological niche increases the likelihood of capturing unique hydrolases and polymer-degrading pathways.

2. DNA Extraction and Quality Control

- **Project:** High-molecular-weight genomic DNA extraction using a CTAB protocol optimised for filamentous fungi; QC with Nanodrop, Qubit, and agarose gel electrophoresis.
- **Course Topics:** Nucleic acid integrity, quantification, and purity checks as prerequisites for NGS.
- **Justification:** DNA quality directly affects read length, basecalling accuracy, and downstream assembly contiguity.

3. Hybrid Sequencing Strategy

- **Project:** Illumina NovaSeq PE150 ($\sim$101$\times$ coverage) for short reads; Oxford Nanopore GridION ($\sim$35$\times$ coverage) for long reads.

- **Course Topics:** Trade-offs between sequencing platforms; hybrid assembly advantages.
- **Justification:** Short reads provide high accuracy for polishing; long reads resolve repeats and structural variation, enabling high-contiguity assemblies.

4. Genome Assembly and Assessment

- **Project:** Flye (Nanopore) + Pilon polishing (Illumina, 3× rounds). QUAST and BUSCO (*fungi_odb10*) metrics: 38.6 Mb genome, N50 = 5.3 Mb, 10 contigs, GC = 58.5%, BUSCO completeness = 76.3%.
- **Course Topics:** Assembly metrics (N50, L50, GC%), completeness benchmarking.
- **Justification:** BUSCO informs on completeness relative to a lineage dataset, identifying gaps potentially linked to niche-specific adaptation.

5. Functional Genome Annotation

- **Project:** MAKER3 integrating Augustus and GeneMark-ES; CAZyme prediction with dbCAN3; manual curation in *Geneious*. Identified 272 CAZymes, including cutinases, esterases, and PHA depolymerases.
- **Course Topics:** ORF prediction, domain annotation, enzyme family classification, manual curation.
- **Justification:** Combining automated pipelines with manual inspection ensures accuracy and functional relevance of gene models.

6. Transcriptomic Analysis and Functional Validation

- **Project:** RNA-Seq (12 libraries; Control, PLA, PHA) quantified with Salmon, summarised via tximport, and analysed in DESeq2. Results: 84 PLA-specific DEGs, 51 PHA-specific DEGs, 29 shared.
- **Course Topics:** Experimental design for RNA-Seq, differential expression, integration with genome annotation.
- **Justification:** Validates predicted enzymatic functions by linking gene presence to substrate-induced expression.

7. Comparative Genomics and Phylogenomics

- **Project:** OrthoFinder orthogroup inference, MCScanX/Mauve synteny mapping. Phylogeny from 338 single-copy orthologs (MAFFT, Gblocks, AMAS, RAxML, MrBayes, MEGA11), curated in *Geneious*.
- **Course Topics:** Orthology inference, structural genome conservation, phylogenetic reconstruction.
- **Justification:** Comparative context reveals unique gene content and evolutionary relationships, supporting claims of niche adaptation.

8. Applied and Policy-Relevant Outcomes

- **Project:** Enzymatic potential of PLA-C1 linked to EU EEA concerns on slow bioplastic degradation; data applicable for composting optimisation.
- **Course Topics:** Translational genomics, environmental policy, bioeconomy.
- **Justification:** Aligns with Circular Economy Action Plan goals; provides molecular basis for eco-innovation strategies.

Integrated Checklist

- Targeted environmental sampling and organism selection
- High-quality DNA extraction with NGS-grade QC
- Hybrid short- and long-read sequencing
- De novo assembly with robust quality metrics
- Structural + functional genome annotation with manual curation
- RNA-Seq differential expression validation
- Orthology, synteny, and phylogenomic analysis
- Policy-relevant application in bioeconomy and environmental genomics