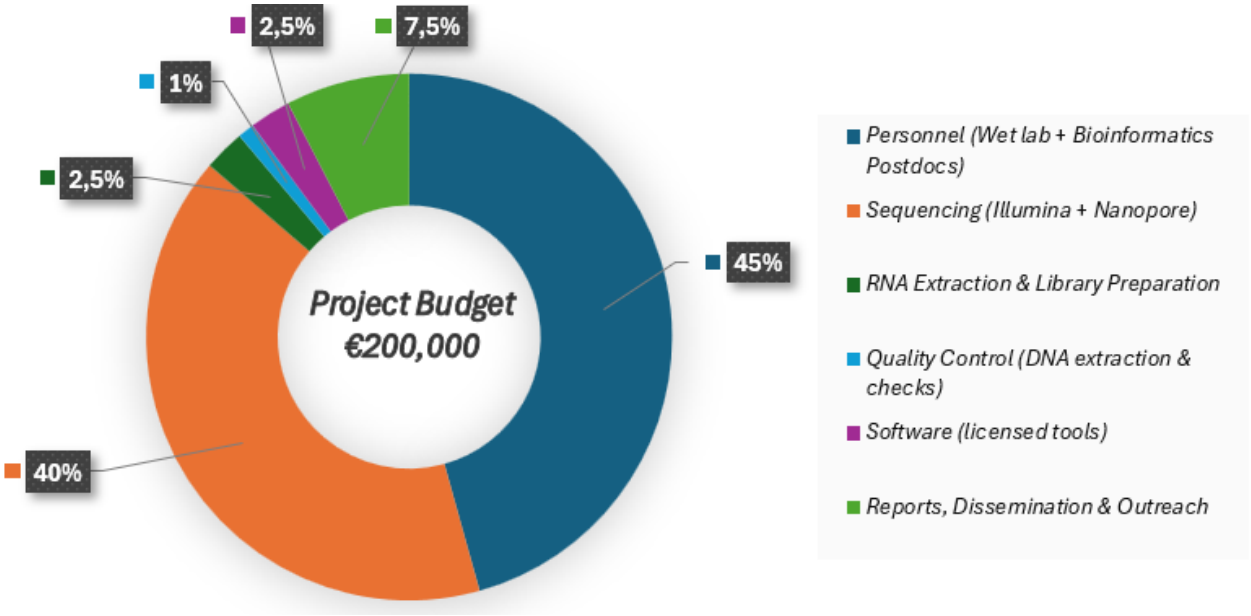
Hybrid Genome Assembly and Integrative Multi-Omics of *Composting-Residue Isolates* Reveal Bioplastic Degradation Potential

Martina Castellucci
MSc Bioinformatics, University of Bologna

Academic Year 2024/2025



This project was designed within a fixed budget (€200,000), 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). 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 [4,6,7,8]. The shift to a circular bioeconomy is hampered by incomplete mineralization, which causes plastic fragments to persist [4,5,9]. Filamentous fungus are interesting biological agents for catalyzing polyester depolymerization because of their capacity to release extracellular hydrolases [3]. 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.

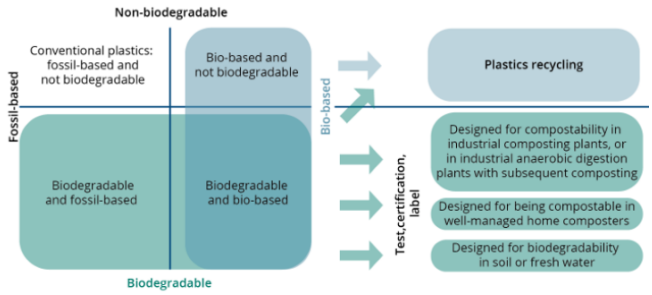


Figure 1: Classification of plastics by origin (fossil-based vs bio-based) and biodegradability. PLA and PHA fall into the *bio-based and biodegradable* category, making them relevant targets for compostability studies. Adapted from the European Environment Agency [4] and European Commission reports [9].

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 [8]. 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 [7]. 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 $\sim$ 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

Compost material clearly enriched in PLA and PHA pieces was sampled and homogenized. After plating on PDA medium supplemented with 0.5% (w/v) powdered PLA and Rhodamine B, about 20 fungal isolates are anticipated. With its clear fluorescent halos under UV light, the Rhodamine B–PLA complex assay will serve as a rapid proxy for extracellular esterase activity [7]. About five of these isolates with the highest activity will be kept for further examination. Preliminary ITS rDNA sequencing and BLAST-based comparison will be carried out to confirm taxonomic placement at the genus level [1].

2.2 Nucleic acid extraction and QC

High molecular weight genomic DNA will be extracted using a CTAB protocol optimized for filamentous fungi. To guarantee fragment sizes above 20 kb, integrity will be checked on 0.8% agarose gels, and DNA yield quantified with Qubit (threshold >50 ng/µL). Purity will be assessed by Nanodrop, with A260/280 ratios expected between 1.8–2.0 for DNA and ∼2.0 for RNA, and A260/230 between 2.0–2.2 as indicator of low contamination. DNA of good integrity will show sharp bands in agarose gel, whereas smearing indicates fragmentation. RNA will be extracted using TRIzol combined with silica column purification. RNA quality will be confirmed with Nanodrop, Qubit, and Bioanalyzer, with RIN scores >8 considered acceptable for RNA-Seq libraries. 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.

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 6000 platform (short reads, high accuracy, sequencing-by-synthesis) [24], which uses sequencing by synthesis chemistry with reversible terminator nucleotides, PE150 configuration, and lane-based high-throughput flow cells. Long reads will be generated using Oxford Nanopore GridION (long reads, real-time nanopore sequencing), which directly sequences native DNA molecules without synthesis, enabling read lengths often exceeding 20 kb, and real-time basecalling with multiple flow cells in parallel [25]. GridION is preferred

over MinION (portable sequencer with lower throughput) because it offers higher output (up to 150 Gb across 5 flow cells) and integrated compute for real-time basecalling, making it more suitable for multi-isolate projects. The recommended ligation sequencing kit SQK-LSK114 will be used, compatible with PCR-free library prep to minimize amplification bias. It is anticipated that this dual-platform architecture will produce reference-quality assemblies.

2.4 Assembly, polishing, and quality control

Nanopore reads will be assembled with Flye (v2.9) (a long-read assembler) [13], followed by multiple iterations of Pilon (short-read correction tool) polishing with Illumina reads [14]. Assembly contiguity and quality will be assessed with QUAST [16], and completeness evaluated using BUSCO (single-copy Benchmarking Universal Single-Copy Orthologs, i.e conserved genes expected in all fungi, used as quality standard) (fungi_odb10 dataset) [15]. Hybrid assemblies aim at coverage of at least 30X Illumina + 40–60X Nanopore, leveraging complementary error profiles (substitution errors in Illumina vs. indel errors in Nanopore). Hybrid assemblies are expected to achieve BUSCO completeness scores above 90%.

2.5 Structural and functional annotation

MAKER3 (annotation pipeline)[17] will integrate ab initio predictors (e.g AUGUSTUS, which predicts genes based only on a sequence features [21]) with Rna-Seq alignments and protein evidence. This combined strategy improves accuracy by reconciling computational predictions with real transcriptional evidence. Repeat masking will be included as a preliminary step to avoid false-positive gene predictions in repetitive regions. Functional annotation will focus on CAZymes using dbCAN3 (database for Carbohydrate-Active Enzymes) [18]. Geneious will be used for manual inspection and validation of gene models.

2.6 Comparative genomics and phylogenomics

Orthogroup inference will be performed with OrthoFinder (orthology assignment software) [19]. Comparisons will rely on single-copy orthologs (genes present once in all species, considered ideal markers for phylogenomics). Synteny and collinearity will be analyzed with MCScanX (tool for detecting conserved gene order)[20] and MAUVE. Phylogenomic reconstruction will rely on concatenated alignments of single-copy orthologs (ensuring robust evolutionary inference): aligned with MAFFT, filtered with Gblocks, concatenated with AMAS, and analyzed with RAxML (maximum likelihood) and MrBayes. MCScanX can also reveal segmental duplications that may underlie CAZyme expansions in degraders compared to non-degraders. These approaches will help identify lineage-specific gene gains or losses associated with PLA/PHA degradation.

2.7 Transcriptomics: quantification and DEG analysis

Salmon will be used to quantify RNA-Seq reads [22], and tximport will be used to compile gene-level expression matrices. DESeq2 will be used to perform differential expression analysis [23], with significance thresholds set at FDR <0.05 and $|\log_2 \mathrm{FC}| \geq 2$. P-values will be adjusted for multiple testing using the Benjamini–Hochberg correction. It is anticipated that CAZyme families (cutinases, esterases, lipases) will show strong induction under PLA and PHA conditions.

2.8 Computational Resources

The bioinformatic analyses will require access to high-performance computing (HPC) facilities. Main needs include:

- **Assembly:** Hybrid assembly (Fyle + Pilon) on nodes with $\geq$128 GB RAM and 32 CPUs; runtime may be several days per isolate.
- **Rna-Seq:** Analyses with Salmon and DESeq2; $\sim$64 GB RAM and multi-threading sufficient for 12 libraries.
- **Comparative genomics:** Tools such as Orthofinder and MCScanX require larger nodes ($\geq$256 GB RAM, 40+ CPUs).
- **Storage:** $\sim$2 TB for raw data and 1-2 TB for processed results.
- **GPU support:** Nanopore basecalling accelerated with NVIDIA GPUs (e.g V100/A100).

Workflows will be managed with Snakemake or Nextflow for scalability and reproducibility.

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, $\sim$5 of the $\sim$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.[10]	Esterase-rich; known PLA degraders; convergence in esterases.
	Fusarium spp.[11]	Cutinases and depolymerases; pathway signatures detectable by comparative genomics.
	Penicillium spp.[11]	Broad CAZyme repertoire; orthogroups distinct between degraders and non-degraders.
	Purpureocillium spp.[2]	Documented polyester degradation; links to biocontrol clade.
Basidiomycota	*Phanerochaete chrysosporium*[12]	White-rot lignin/peroxide system; oxidative auxiliary enzymes.
	Trametes spp.[12]	Laccase-rich; oxidative contribution to polymer breakdown.
Mucoromycota	*Rhizopus* spp.[11]	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.

5. Philp, J.C., Ritchie, R.J., Guy, K. (2013). Biobased plastics in a bioeconomy. *Trends in Biotechnology*, 31(2), 65–67. DOI: 10.1016/j.tibtech.2012.11.009

6. Lambert, S., Wagner, M. (2017). Environmental performance of bio-based and biodegradable plastics: the road ahead. *Chem. Soc. Rev.*, 46, 6855–6871. 10.1039/c7cs00149e

7. Tokiwa, Y., Calabia, B.P. (2006). Biodegradability and biodegradation of poly(lactide). *Appl. Microbiol. Biotechnol.*, 72, 244–251.10.1007/s00253-006-0488-1

8. Emadian, S.M., Onay, T.T., Demirel, B. (2017). Biodegradation of bioplastics in natural environments. *Waste Management*, 59, 526–536. 10.1016/j.wasman.2016.10.006

9. European Commission (2018). *A sustainable bioeconomy for Europe*. Publications Office of the EU. 10.2777/792130

10. Economou, C. N. et al. Enhanced biodegradation of polylactic acid (PLA) films and microparticles by a commercial lipase from Aspergillus oryzae. *Enzymatic Hydrolysis of PLA*, (2023). 10.1016/j.biortech.2025.132807

11. Ekanayaka, A. H. et al. (2022). A review of fungi that degrade plastic. *Fungal Biology*, 10.3390/jof8080772

12. Li, L., Guo, Z., Guo, X., Dong, D., Fan, T., Yao, L., Zha, Y., Kong, J., Dai, Y., Pan, T. (2025). *Phanerochaete chrysosporium* hyphae biocrack, endocytose and metabolize plastic films. *Journal of Hazardous Materials*, 468, 133565. DOI: 10.1016/j.jhazmat.2025.133565

13. Kolmogorov, M. et al. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.*, 37, 540–546. (Flye) 10.1038/s41587-019-0072-8

14. Walker, B.J. et al. (2014). Pilon: an integrated tool for microbial variant detection and assembly improvement. *PLoS One*, 9(11), e112963. 10.1371/journal.pone.0112963

15. Seppey, M., Manni, M., Zdobnov, E.M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol Biol.*, 1962, 227–245. 10.1007/978-1-4939-9173-$0_1$4

16. Gurevich, A. et al. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29, 1072–1075.10.1093/bioinformatics/btt086

17. Cantarel, B.L. et al. (2008). MAKER: an annotation pipeline for emerging model organisms. *Genome Res.*, 18, 188–196. 10.1101/gr.6743907

18. Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., Yin, Y. (2023). dbCAN3: automated carbohydrate-active enzyme annotation server. *Nucleic Acids Research*, 51(W1), W115–W123. 10.1093/nar/gkad328

19. Emms, D.M., Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference. *Genome Biol.*, 20, 238. 10.1186/s13059-019-1832-y

20. Wang, Y. et al. (2012). MCScanX: detection and evolutionary analysis of gene synteny. *Nucleic Acids Res.*, 40(7), e49. 10.1093/nar/gkr1293

21. Stanke, M., Morgenstern, B. (2005). AUGUSTUS: ab initio gene prediction. *Nucleic Acids Res.*, 33, W465–W467. 10.1093/nar/gkl200

22. Patro, R. et al. (2017). Salmon: fast and bias-aware transcript quantification. *Nat. Methods*, 14, 417–419. 10.1038/nmeth.4197

23. Love, M.I., Huber, W., Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15, 550. 10.1186/s13059-014-0550-8

24. Illumina Inc. (2025). NovaSeq 6000 Sequencing System. Product page: https://www.illumina.com/systems/sequencing-platforms/novaseq.html

25. Oxford Nanopore Technologies (2025). GridION sequencing device. Product page: https://nanoporetech.com/products/sequence/gridion