Pablo González de la Rosa

Bioinformatician | Oxford Nanopore Technologies

Email: pgonzale60@gmail.com • GitHub: github.com/pgonzale60 • Remote-friendly (Oxford, UK)

PROFESSIONAL SUMMARY

Bioinformatics workflow engineer specialising in high-throughput genomic analysis, pipeline automation, and verification. Maintainer of the `wf-pgx` Nextflow workflow, delivering reproducible pharmacogenomics analyses across regulated environments. Experienced in containerised deployments, Git-centric collaboration (code review, CI/CD, testing), and ensuring robustness and scalability for verification workflows. Passionate about building infrastructure that keeps complex scientific software traceable, testable, and production-ready. Comfortable blending Python, Rust, and Nextflow to turn research ideas into robust production pipelines. Longstanding focus on automation (from GNU Make and PHP to modern CI/CD) and continuously integrating AI/NLP to streamline text-heavy tasks such as matching user preferences against free-text from hundreds of adverts.

CORE COMPETENCIES

- **Workflow Engineering**: Nextflow DSL2, nf-core patterns, schema-driven parameters, ingress/egress design, resource profiling.
- **Containerisation & Build Automation**: Docker/OCI, Singularity, Whalefish (automated SHA-tagged containers from Git repos), multi-arch support.
- **Software Collaboration**: Git workflows, code review, design documentation, pair debugging, changelog management.
- **CI/CD & Verification**: GitHub Actions, Jenkins, regression suites, smoke tests, container provenance, automated docs (`parse_docs`).
- **Cloud & HPC**: AWS Batch, Slurm, caching strategies, cost/performance tuning.
- **Data Handling**: Python (pandas, NumPy), R (tidyverse, Shiny), Bash, Rust for performance-critical glue.
- **Observability & Reporting**: Coverage metrics (mosdepth), VCF statistics, HTML report generation (`workflow-glue`).

PROFESSIONAL EXPERIENCE

Oxford Nanopore Technologies — Bioinformatician

Oxford, UK • Nov 2023 – Present

- Built Nextflow pipelines for human targeted sequencing applications using software development best practices (Git, containers, CI).
- Maintained and extended `epi2me-labs/wf-pgx` (pharmacogenomics workflow) with container provenance, CI smoke tests, and schema-driven parameters.
- Built custom Python and Rust utilities for variant detection and analysis across human, cattle, and bacterial datasets.
- Curated genomic datasets to validate analytical performance and ensure high-quality results.
- Collaborated with cross-functional teams to improve pipeline efficiency, accuracy, and reproducibility.

- Replaced fragmented PharmCAT/Chinook artefacts with a unified `star_alleles.tsv` while retaining raw match JSON for audit trails.
- Structured documentation with `parse_docs`, schema/output definitions, and automated README generation.
- Participated actively in a bioinformatics tools journal club; presented three sessions in the past year.

Winter Genomics — Bioinformatics Intern

Mexico City, Mexico • Aug 2013 – Nov 2015

- Performed microarray-based expression profiling and genotyping.
- Developed a MySQL database with a web interface for genomic variants.
- Built internal R analysis scripts and supported pipeline automation for host–pathogen studies.

Independent Bioinformatics Consultant

Remote • 2019 – 2022

- Migrated academic genomics scripts to reproducible Nextflow pipelines; delivered Singularity/Apptainer-enabled HPC deployments.
- Established Git-based collaboration workflows (branching, review templates, CI smoke tests) for university labs.

SELECTED PROJECTS

- **wf-pgx Evolution (ONT)**: Designed and now maintain flagship pharmaco-genomics workflow; added automated container builds, regression suites, and robust documentation pipeline.
- **Verification Workflow Suite**: Reusable Nextflow modules for coverage reporting, variant metrics, and report diffs to validate releases across datasets.
- **Chromosome-scale Genome Annotation (MSc)**: Nextflow workflow to annotate multi-megabase animal genomes; integrated gene prediction, repeat masking, and evidence-driven refinement for mid-scale HPC.
- **Assembly Completeness Assessment (PhD)**: Automated pipeline to assess completeness of newly assembled multi-megabase animal genomes using orthology- and k-mer–based metrics.
- **Whalefish Integration**: Wired Nextflow builds into Whalefish so every git commit produces an SHA-tagged container traceable in production.
- **GEO Differential Expression Automation (BSc thesis)**: Automated microarray-based differential expression across dozens of GEO experiments using GNU Make; downloaded datasets and Brainarray annotations, inferred experimental designs, executed group-vs-group contrasts, and summarised results in a single table.
- **Regulatory Gene Visualisation (BSc training)**: Automated gene visualisation for a gene regulation website using PHP and available regulatory annotations.
- **AI-assisted Preference Matching (personal R&D)**: Prototyped NLP/AI workflows to match user preferences against large volumes of free-text (e.g., adverts), reducing manual review effort.

PUBLICATIONS (SELECTED)

- Stevens, L., *et al.* (2024). *The genome of Litomosoides sigmodontis illuminates the origins of Y chromosomes in filarial nematodes.* PLoS Genetics, 20(1), e1011116.

- Stevens, L., Martínez-Ugalde, I., *et al.* (2023). *Ancient diversity in host-parasite interaction genes in a model parasitic nematode.* Nature Communications, 14(1), 7776.
- Gonzalez de la Rosa, P. M., *et al.* (2021). *A telomere-to-telomere assembly of Oscheius tipulae and the evolution of rhabditid nematode chromosomes.* G3, 11(1), jkaa020.

Full publication list available on request.

EDUCATION

- **PhD Evolutionary Biology**, University of Cambridge & Wellcome Sanger Institute (2020 – 2023)

Chromosome evolution and programmed DNA elimination in Rhabditina nematodes.

- Developed an automated pipeline to assess completeness of newly assembled multi-megabase animal genomes (orthology/k-mer metrics, reproducible runs).
- **MSc Integrative Biology**, LANGEBIO-CINVESTAV, Mexico (2016 – 2018)

Host-specific gene expression in monarch butterflies.

- Built a Nextflow workflow to annotate multi-megabase animal genomes (repeat/structural annotation, evidence integration, HPC scheduling).
- **BSc Genomic Sciences**, Center for Genomic Sciences, UNAM, Mexico (2011 – 2015)
- Thesis: Automated GEO differential expression pipeline with GNU Make (data/annotation retrieval, design inference, batch contrasts, unified summary table).

TECHNICAL TOOLKIT

- **Languages**: Python, R, Rust, Bash, SQL, PHP.
- **Workflow Tools**: Nextflow DSL2, nf-core modules, Snakemake (legacy support), GNU Make.
- **Containers**: Docker, Singularity/Apptainer, Whalefish, container registries.
- **CI/CD**: GitHub Actions, Jenkins, pre-commit, changelog automation.
- **Infra**: AWS Batch, HPC schedulers (Slurm), S3, Artefact repositories.
- **AI/NLP**: Embeddings, semantic search/matching, LLM-assisted extraction, prompt design; practical experience applying LLMs (OpenAI/Gemini) to automate text-heavy triage and matching tasks.
- **Dataviz & Reporting**: matplotlib, Plotly, Shiny, custom HTML reports via `workflow-glue`.

PROFESSIONAL DEVELOPMENT & COMMUNITY

- Nextflow Summit attendee; contributor to nf-core discussions and module reviews.
- Internal trainer for reproducible workflow practices, Whalefish usage, and code review standards.
- Active in cross-functional initiatives linking wet-lab requirements with informatics pipelines.

REFERENCES

Available upon request.