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 with 8+ years building reproducible genomic analysis pipelines. Strong in Nextflow, containerisation, and CI/CD; maintainer of `wf-pgx` (research-use only).

I turn research into robust systems with Python and R, and an automation-first mindset (from GNU Make to modern CI). I prototype LLM-assisted workflows to streamline text-heavy tasks when useful, while following quality practices aligned with ISO-style accreditation goals.

CORE COMPETENCIES

- **Workflow Engineering**: Nextflow DSL2, nf-core patterns, schema-driven parameters, clear I/O contracts.
- **Containerisation & Builds**: Docker/OCI, Singularity/Apptainer; reproducible environments.
- **Software Collaboration**: Git workflows, code review, design documentation, changelog management.
- **CI/CD & Verification**: GitHub Actions/Jenkins, smoke tests, automated documentation (`parse_docs`).
- **HPC & Cloud**: Slurm; working familiarity with AWS (S3, EC2 basics).
- **Data Handling**: Python (pandas, NumPy), R (tidyverse, Shiny), Bash; Rust/PHP (basic familiarity).
- **Reporting**: Coverage metrics (mosdepth), VCF stats, HTML report generation.

PROFESSIONAL EXPERIENCE

Oxford Nanopore Technologies — Bioinformatician

Oxford, UK • Nov 2023 – Present

- Built and maintained Nextflow pipelines for targeted sequencing (schema-driven parameters, containerised modules, CI smoke tests).
- Maintained and evolved `epi2me-labs/wf-pgx` (research-use only) with representative datasets and automated documentation.
- Delivered Python utilities for variant calling/analysis across human, cattle, and bacterial datasets; integrated into pipelines and CI.
- Consolidated PharmCAT/Chinook artefacts into a unified `star_alleles.tsv`, preserving raw JSON for traceability.
- Improved documentation quality and consistency via `parse_docs`, schema/output definitions, and templated READMEs.
- Speaking & engagement: active in bioinformatics tools journal club (3 sessions/year).

Winter Genomics — Bioinformatics Intern

Mexico City, Mexico • Aug 2013 – Nov 2015

- Performed microarray-based expression profiling and genotyping; automated common analyses and QC.
- Designed a MySQL-backed variant database with a web interface to support collaborators' queries and curation.
- Built internal R scripts and contributed to 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.
- Rolled out Git-based collaboration workflows (branching, reviews, CI smoke tests) to improve team velocity and code quality.

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.
- **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 LLM-assisted workflows to match user preferences against free-text (e.g., adverts) to reduce manual screening.

Context and outcomes (selected):

- wf-pgx: improved clarity via schema-driven params and docs automation; easier review for research users.
- Verification Suite: standardised release checks and metrics diffs; faster validation cycles.
- AI matching: accelerated manual screening tasks using LLM-assisted triage and ranking prototypes.

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 (advanced), R (advanced), Bash (proficient), SQL (proficient), Rust/PHP (familiar)
- **Workflow Tools**: Nextflow DSL2 (nf-core patterns), schema-driven parameters, reproducible containers (Docker/Singularity)
- **CI/CD & Quality**: GitHub Actions/Jenkins, smoke tests, docs automation; familiarity with ISO-aligned quality practices
- **HPC & Cloud**: Slurm; AWS familiarity (S3, EC2 basics)
- **AI/LLM Prototyping**: Prompting and LLM-assisted extraction for text triage (early-stage prototypes)
- **Reporting**: matplotlib, Plotly, Shiny, custom HTML reports

---

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.