

nf-core/deepmutscan: a community-oriented, modular bioinformatics pipeline for deep mutational scanning by shotgun sequencing

Benjamin Wehnert¹, Taylor L. Mighell¹, Fei Sang², nf-core community, Ben Lehner^{1,2}, Maximilian R. Stammniz¹

¹Centre for Genomic Regulation (CRG), Barcelona, Spain; ²Wellcome Sanger Institute, Cambridge, UK

Experimental strategies for DMS of long ORFs

Deep mutational scanning (DMS) efficiently characterizes variant-phenotype relationships, but its application to long open reading frames remains technically challenging. Established methods to overcome short-read limitations, such as **gene tiling** or **barcoding**, either hinder experimental multiplexing or demand resource-intensive long-read sequencing.

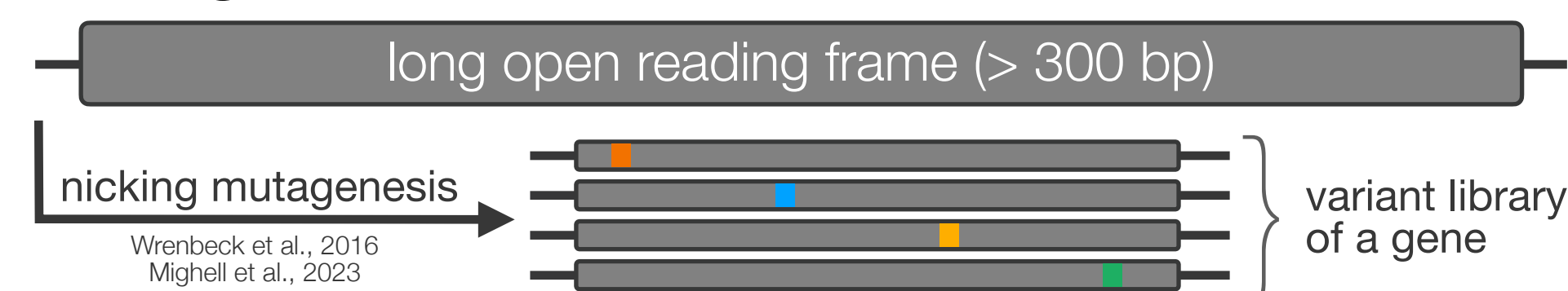
Shotgun sequencing of libraries provides an elegant, underused alternative (Rao *et al.*, 2024). Quantifying variants from randomly fragmented DNA without barcodes scales DMS to larger gene targets while preserving experimental multiplexing and compatibility with standard sequencing platforms.

To standardize this analysis, we present **nf-core/deepmutscan**, an end-to-end pipeline processing raw sequencing files directly into fitness estimations. While optimized for shotgun sequencing, it also supports tiled-based DMS experiments. The pipeline integrates a novel error-correction step to reduce deep sequencing noise. Containerized for effortless deployment across environments, its modular architecture allows low-effort extension with other DMS tools, like fitness calculators, providing a robust, scalable framework.

Rao, Jingyou, *et al.* "Rosace: a robust deep mutational scanning analysis framework employing position and mean-variance shrinkage." *Genome Biology* 25.1 (2024): 138.



Mutagenesis



Experimental strategies

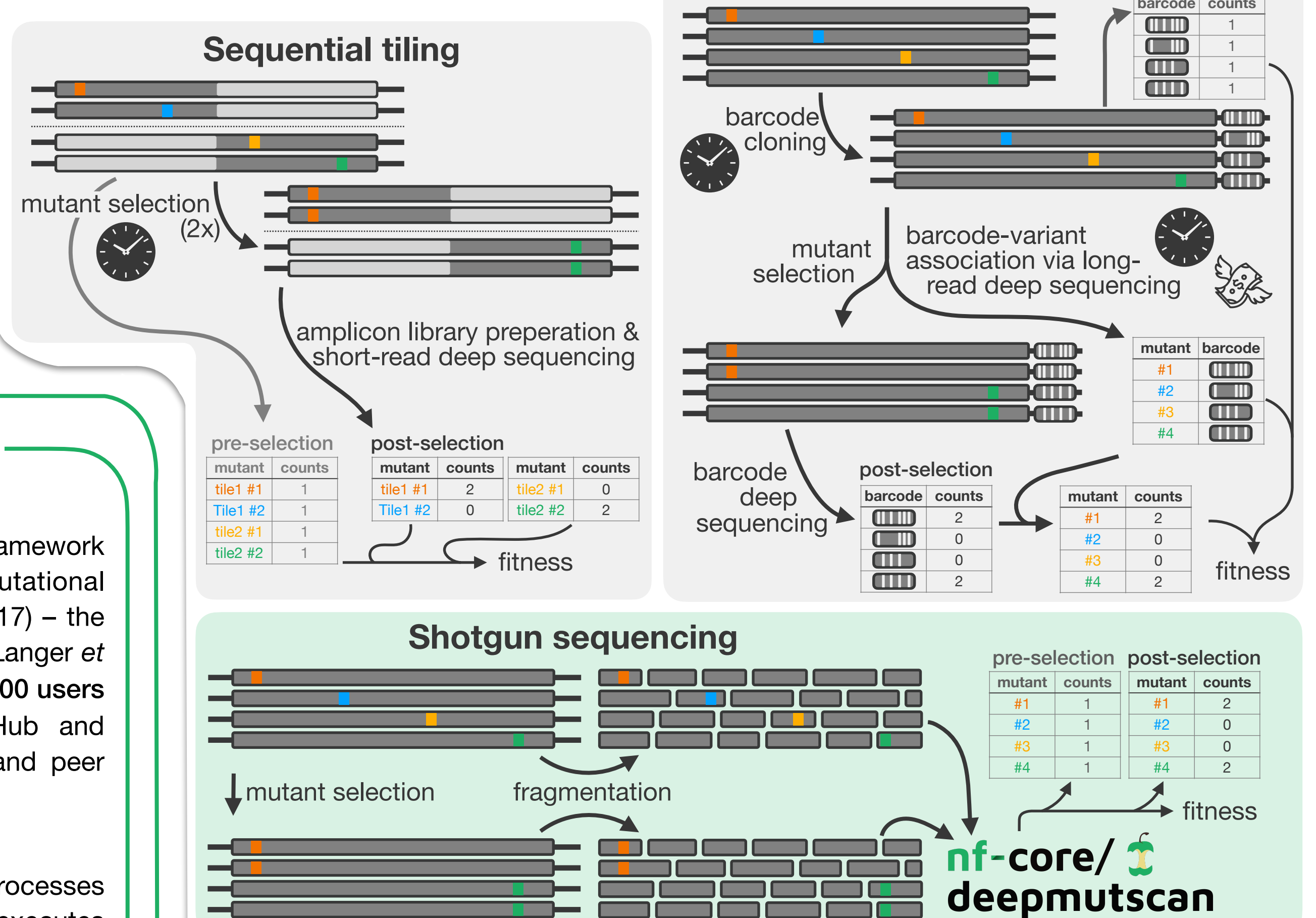
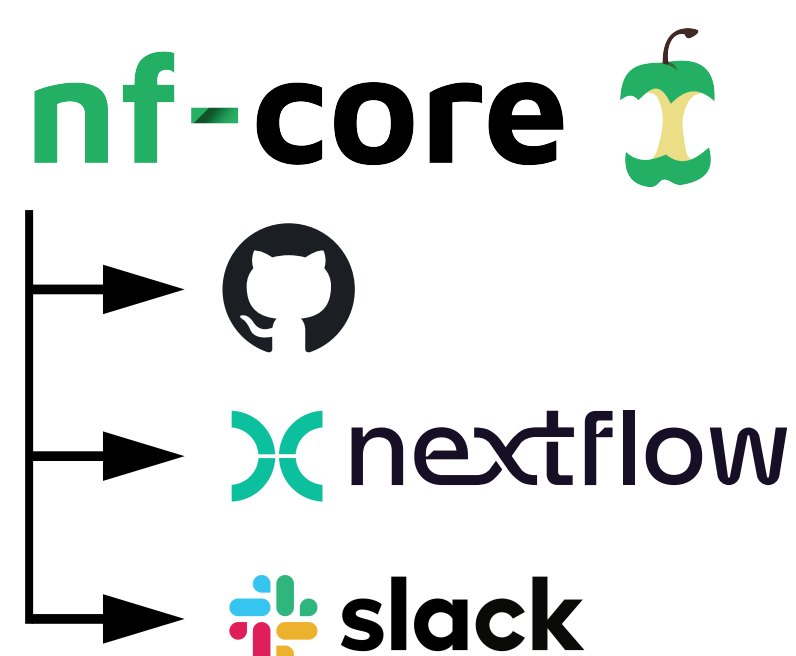


Figure 1: DMS workflows for long ORFs. Following initial mutagenesis, libraries can be processed using tiling, barcoding, or shotgun sequencing, all of which yield the same results.

Workflow management by nf-core



Why we chose nf-core

nf-core offers a standardized, community-driven framework for reproducible, scalable, and portable computational workflows built on Nextflow (Di Tommaso *et al.*, 2017) – the fastest-growing workflow system in bioinformatics (Langer *et al.*, 2024). As of 2025, over 500 developers and 11,000 users maintain ~100 peer-reviewed pipelines via GitHub and Slack. Strict guidelines, modular design, testing, and peer review ensure robustness and interoperability.

How nf-core workflows function

Workflows consist of reusable, containerized processes connected by defined inputs and outputs. Nextflow executes them across local machines, HPC clusters, and cloud systems, ensuring portability and scalability.

Di Tommaso, Paolo, *et al.* "Nextflow enables reproducible computational workflows." *Nature biotechnology* 35.4 (2017): 316-319.
Langer, B.E., Amaral, A., Baudement, M.O. *et al.* Empowering bioinformatics communities with Nextflow and nf-core. *Genome Biol* 26, 228 (2025).

Deepmutscan: from raw reads to variant fitness

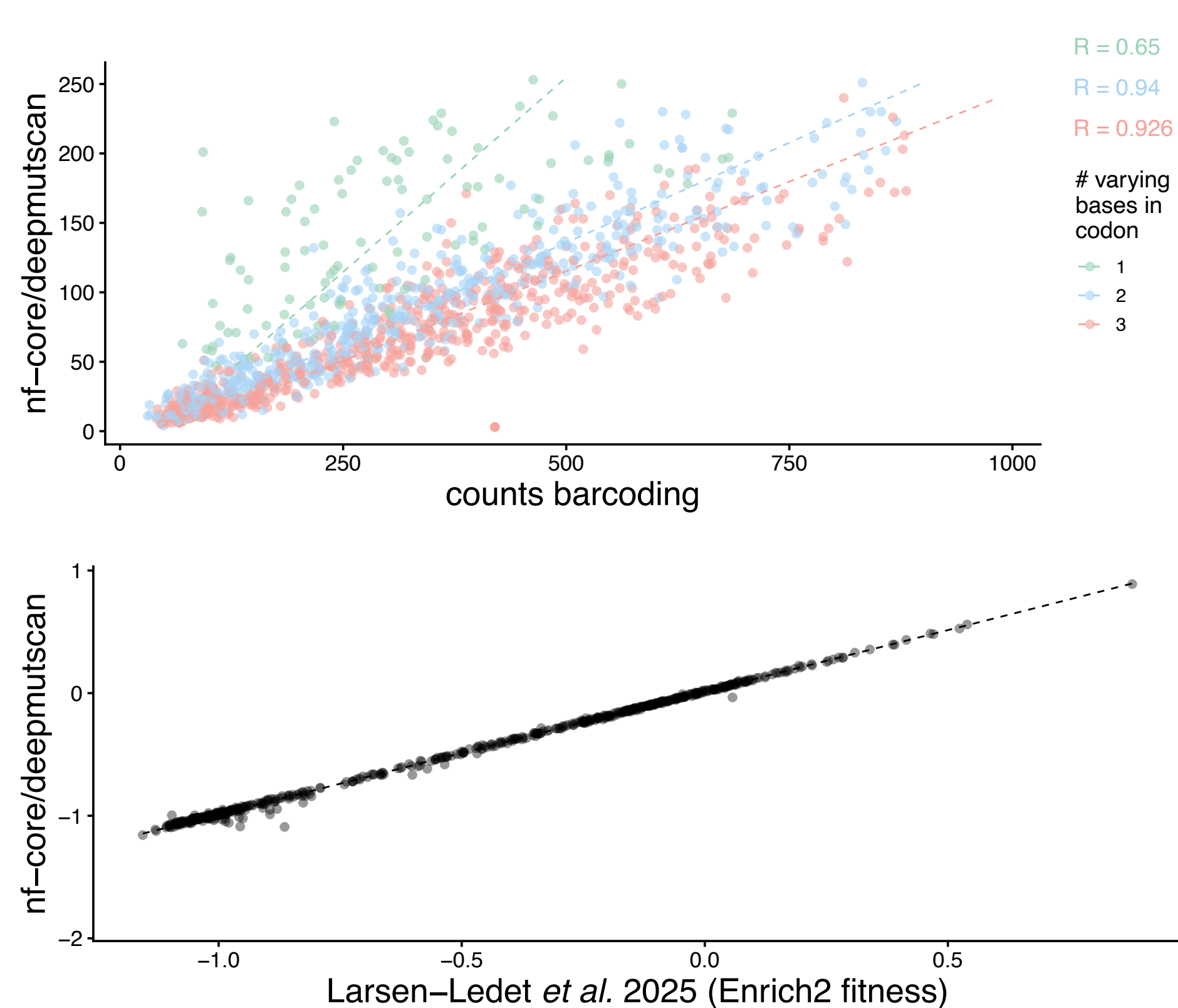


Figure 2: Methodological comparisons. **Top:** Correlation of barcode- and shotgun-based counts across variants of mutagenized MOR. Data provided by Taylor L. Mighell. **Bottom:** Concordance of MLH1 protein abundance fitness scores derived from raw sequencing files using Enrich2 vs. *nf-core/deepmutscan*.

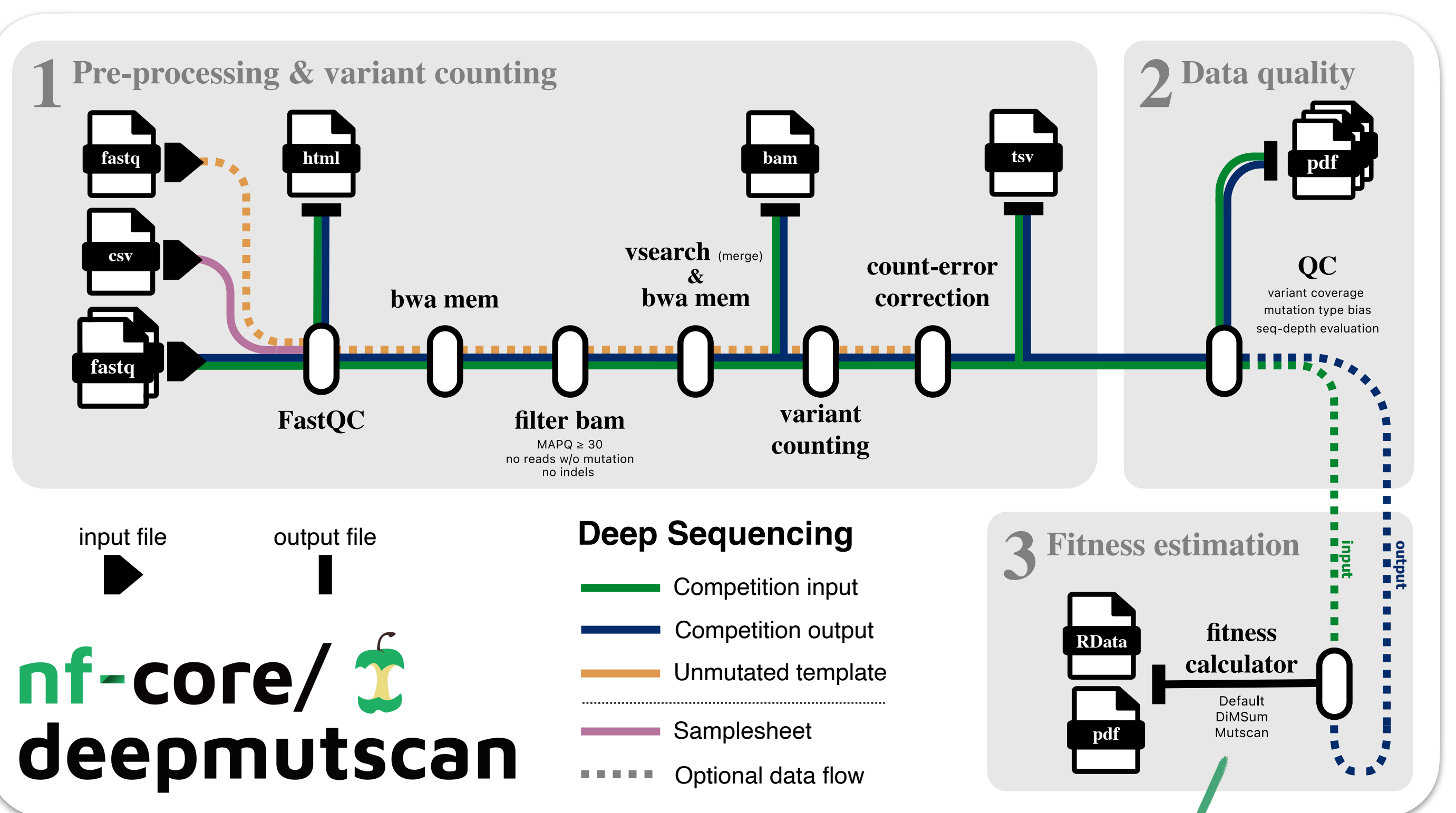


Figure 3: *Deepmutscan* processes shotgun sequencing data in three stages: variant counting, quality control, and optional fitness calculation using multiple tools.

Expandable by your favourite statistics tool

Deepmutscan in action



Report for shotgun-based DMS of GID1A (344 aa), mutagenized via SUNi (11,008 single-codon NNK/NNS variants) and profiled in a yeast growth competition. Over 95% of variants were successfully detected in the input library.

Join us!

To maximize *Deepmutscan*'s flexibility and community impact, we actively seek collaborations to extend compatibility with additional DMS tools. We also aim to diversify the upstream data processing, for example, by integrating additional aligners.



@bwehnert.bsky.social
wehnertbenjamin@gmail.com
@maxstammniz.bsky.social
maximilian.stammniz@crg.eu