Sample Submission

Customers can either submit extracted DNA eluted in 10 mM Tris-HCl pH 8.0 or nuclease free water, or provide us with strains. For strain submissions, pure cultures of each strain are grown by the customer, on plates or in broth. Cells are harvested and resuspended in a tube with DNA/RNA Shield (Zymo Research, USA) following MicrobesNG strain submission procedures, and sent to MicrobesNG for sequencing.

DNA Processing

Five to forty microlitres of the cell suspension are lysed with 120 µL of TE buffer containing lysozyme (MPBio, USA) metapolyzyme (Sigma Aldrich, USA) and RNase A (ITW Reagents, Spain), incubated for 25 min at 37°C.

Proteinase K (VWR Chemicals, Ohio, USA) (final concentration 0.1mg/mL) and SDS (Sigma-Aldrich, Missouri, USA) (final concentration 0.5% v/v) are added and incubated for 5 min at 65°C. Genomic DNA is purified using an equal volume of SPRI beads and resuspended in EB buffer (10mM Tris-HCl, pH 8.0).

DNA extracted by MicrobesNG, or sent directly by the customer is then quantified with the Quant-iT dsDNA HS (ThermoFisher Scientific) assay in an Eppendorf AF2200 plate reader (Eppendorf UK Ltd, United Kingdom) and diluted as appropriate.

Illumina Sequencing (Short and Hybrid)

Genomic DNA libraries are prepared using the Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: input DNA is increased 2-fold, and PCR elongation time is increased to 45 seconds.

DNA quantification and library preparation are carried out on a Hamilton Microlab STAR automated liquid handling system (Hamilton Bonaduz AG, Switzerland). Libraries are sequenced on an Illumina NovaSeq 6000 (Illumina, San Diego, USA) using a 250 bp paired end protocol. For projects predating April, 2020, libraries were sequenced on an Illumina HiSeq 2500 using a 250 bp paired end protocol (Illumina, San Diego, USA).

Reads are adapter trimmed using Trimmomatic version 0.30 [1] with a sliding window quality cutoff of Q15. De novo assembly is performed on samples using SPAdes version 3.7 [2], and contigs are annotated using Prokka 1.11 [3].

Variant Calling (Short and Hybrid)

Reads are aligned to the reference using BWA mem v0.7.17 and are then processed using SAMtools v1.9. Variants are called using VarScan v2.4.0 with two thresholds, sensitive and specific, where the variant allele frequency is greater than 90% and 10% respectively. The effects of the variants are predicted and annotated using SnpEff v4.3.

ONT Sequencing (Long and Hybrid)

For long and hybrid projects completed after November, 2023, DNA libraries are prepared with Oxford Nanopore Technologies (ONT) SQK-RBK114.96 kit (ONT, United Kingdom) using 200-400 ng of High Molecular weight (HMW) DNA. Barcoded samples are pooled together into a single sequencing library and loaded in a FLO-MIN114 (R.10.4.1) flow cell in a GridION (ONT, United Kingdom).

For hybrid (formerly known as Enhanced) projects that were completed before November 2023 DNA libraries were prepared with Oxford Nanopore SQK-LSK109 kit with Native Barcoding EXP-NBD104/114 (ONT, United Kingdom) using 400-500 ng of HMW DNA. Barcoded samples were pooled together into a single sequencing library and loaded in a FLO-MIN106 (R.9.4.1) or FLO-MIN111 (R10.3) flow cell in a GridION (ONT, United Kingdom).

Hybrid assembly is performed using Unicycler version 0.4.0 [4], and contigs are annotated using Prokka version 1.11 [3].

Details on Long read methodology, including precise versions can be found in your project portal under the “Methods” heading.

Submitting data to the SRA/ENA

When submitting data to public repositories, you'll need to fill out some metadata to indicate the origin and strategies of your sequencing data. You will have most of this data, such as sample name, collection date, location etc.

Here's a quick and easy reference for the bits we're responsible for; Illumina data is provided with our Short, Hybrid, Scout and Pioneer services, and Nanopore data is provided for our Long Read and Hybrid services.

	Illumina	Nanopore
Instrument Model	NovaSeq 6000	GridION
Library Source	Genomic	Genomic
Library Selection	Random PCR	Random
Library Strategy	WGS	WGS
Library Layout	Paired	Single

[1] Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120.
[2] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology, 19(5), 455–477.
[3] Seemann T (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30(14):2068-9
[4] Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology 13(6): e1005595.