LGC Genomics GmbH
Ostendstraße 25, 12459 Berlin, Germany
t +49 (0)30 5304 2200 **f** +49 (0)30 5304 2201
e info.de@lgcgroup.com **w** lgcgroup.com biosearchtech.com

BIOSEARCH™ TECHNOLOGIES
GENOMIC ANALYSIS BY LGC

Project NGS3582 data delivery, 24 Apr 2023:

Samples:	45
Sequencing amount:	2.5 million reads
Sequencing type:	300bp paired-end read (Illumina MiSeq V3)
Amplicon type:	Bacteria 16S (341F-785R)

Delivery contents:

- 'RAW': raw sequencing data after basecalling and demultiplexing in compressed FASTQ format
- 'AdapterClipped': compressed FASTQ files containing sequencing adapter clipped reads
- 'PrimerClipped': compressed FASTQ files containing primer sorted reads
- 'Combined': compressed FASTQ files containing consensus sequences after overlap combination of forward and reverse reads

FastQC [1] reports, containing read quality metrics, are stored along with the FASTQ files.

Data analysis:

- Demultiplexing of all libraries for each sequencing lane using the Illumina bcl2fastq v2.20 software [2] (folder 'RAW'):
  - 1 or 2 mismatches or Ns were allowed in the barcode read when the barcode distances between all libraries on the lane allowed for it
- Sorting of reads by amplicon inline barcodes (folder RAW):
  - 1 mismatch was allowed per barcode
  - the barcode sequence was clipped from the sequence after sorting
  - reads with missing barcodes, one-sided barcodes or conflicting barcode pairs were discarded
- Clipping of sequencing adapter remnants from all reads (folder 'AdapterClipped'):
  - reads with final length < 100 bases were discarded
- Primer detection and clipping (folder 'PrimerClipped'):
  - 3 mismatches were allowed per primer
  - pairs of primers (Fw-Rev or Rev-Fw) had to be present in the sequence fragments
  - if primer-dimers were detected, the outer primer copies were clipped from the sequence
  - the sequence fragments were turned into forward-reverse primer orientation after removing the primer sequences

LGC Genomics GmbH
Bank: HSBC Trinkaus & Burkhardt AG
IBAN: DE94300308800005189004
BIC: TUBDDEDD

UST-ID-Nr.: DE 811547771 | Steuernr.: 37/083/46000
Amtsgericht Berlin-Charlottenburg | HRB 48332
Geschäftsführer: Steffi Ulrich | Lucas-Caspar Schwinner-Strachwitz

LGC

LGC Genomics GmbH
Ostendstraße 25, 12459 Berlin, Germany
t +49 (0)30 5304 2200 **f** +49 (0)30 5304 2201
e info.de@lgcgroup.com **w** lgcgroup.com biosearchtech.com

- Combination of forward and reverse reads using BBMerge v34.48 [3] (folder 'Combined'):
  - the consensus sequence of combinable fragments are named "*joined-SR*", uncombinable read pairs sequences end up in the "*R1*" and "*R2*" files
- Creation of FastQC reports for all FASTQ files
- Generation of read_counts.xlsx, containing all read counts for all samples at a glance
- The PCR_NGS3582.xlsx file contains information about inline barcodes and primers used for the samples, as well as quality control information regarding the PCRs

If you have any questions related to your data or some steps of the data analysis, do not hesitate to contact us directly:
Email: bioinformatics@lgcgroup.com
Tel.: +49 (0)30 5304 2202

References

[1] Simon Andrews. *FastQC A Quality Control tool for High Throughput Sequence Data*. URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[2] Illumina. *bcl2fastq2 Conversion Software*. URL: https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html.

[3] Brian Bushnell. *BBTools*. URL: http://jgi.doe.gov/data-and-tools/bbtools/.

LGC Genomics GmbH
Bank: HSBC Trinkaus & Burkhardt AG
IBAN: DE94300308800005189004
BIC: TUBDDEDD

UST-ID-Nr.: DE 811547771 | Steuernr.: 37/083/46000
Amtsgericht Berlin-Charlottenburg | HRB 48332
Geschäftsführer: Steffi Ulrich | Lucas-Caspar Schwinner-Strachwitz