BASECLEAR

# NGS **datafile** submission guide

- Before a project can start the client has to visit the Order Portal and submit the order form.
- If you need to submit data files for your project, please deliver these at the very start of the project.
- Data transfer to BaseClear is possible via the following options:
- For up to 1 GB: Upload in Order Portal. Maximum of 200 MB per file, max 5 files.
- For up to 5 GB: Via a file request. Maximum of 100MB per file, no file limit.
- More than 5 GB: Via hard drive postage (SSD or HDD). Password encryption is highly recommended.
- FTP service is not available for upload of data from client to BaseClear.

**Postal address:**
BaseClear B.V.
Attm. Expedition
Galileiweg 4
2333 BE Leiden
The Netherlands.

info@baseclear.com
+31 (0)71 523 3917

## DATA FILE SUBMISSION

### DE NOVO ASSEMBLY USING CUSTOMER SEQUENCE FILES

- In case of extension Illumina data, please provide the two paired-end sequence files separately in FASTQ format (zipped with gzip extension).
- In case of ONT data, please provide the sequence file in FASTQ format (zipped with gz extension).
- In case of both Illumina and ONT data, include an Excel sheet that clearly indicates the correspondence between the Illumina and ONT data files.

### TRANSCRIPTOME ANALYSIS

- For transcriptome analysis submit the annotated reference genome as link to NCBI accession number.
- If this is not available, please submit an annotated GenBank file or GFF+FASTA file.

### qPCR

- Submit the reference genome as a link to the NCBI accession number.
- If this is not available, please submit a GenBank file or FASTA file.
- Provide each reference in a separate file with clear naming.
- It is mandatory to provide an excel table that clearly describes each file and its group, species, strain or gene target for the qPCR assay using the F089-qPCR-01 form.
- Request the F089-qPCR-01 form at info@basclear.com in case you do not have this document.
- For fungal and yeast genomes we may have to use a large scaffold rather than the entire genome. **Note:** Client may provide the scaffold of their choice, alternatively, BaseClear selects the largest scaffold of the genome.

### DIFFERENTIAL GENE EXPRESSION ANALYSIS (DGE)

- Submit a table to indicate clearly which groups should be compared for DGE (Differential Gene Expression analysis). Minimum of 3, preferably biological replicates, samples per group is required for normalisation and reaching sufficient statistical power.

### REFERENCE ALIGNMENT OR VARIANT CALLING ANALYSIS (SNP)

- For both reference alignment or variant calling submit the annotated reference genome as link to NCBI accession number.
- If this is not available, please submit an annotated GenBank file or GFF+FASTA file.
- If multiple references are to be used, provide all of them in separate files with clear file naming.
- In case multiple samples and/or references are to be analysed, it is mandatory to provide an Excel table that clearly describes the pairing between a sample and one (or more) reference sequence(s).