Univeristy of Gothenburg
Victor Lobanov
Carl Skottsbergs gata 22B
405 30 Gothenburg
Sweden

Project code/Order number

122129

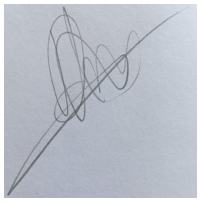
Date of Report

21-Jan-2020

Authorization

Daniël Duijsings

Head of research and development



BaseClear B.V.

Galileiweg 4
2333 BD Leiden
The Netherlands
T +31 71 523 3917
E info@baseclear.com

Next Generation Sequencing project

Next Generation Sequencing project report

Project 122129
Client Victor Lobanov
Date 21-Jan-2020 10:51

1. Introduction

This report contains a detailed overview of your next-generation sequencing experiment. The data and results apply to the following samples:

Sample Name	Illumina demultiplexing	Otu classification
1-1	X	X
1-2	X	X
1-3	X	X
2-1	X	X
2-2	X	X
2-3	X	X
3-1	X	X
3-2	X	X
3-3	X	X
4-1	X	X
4-2	X	X
4-3	X	X
5-1	X	X
5-2	X	X
5-3	X	X
6-1	X	X
6-2	X	X
6-3	X	X
7-1	X	X
7-2	X	X
7-3	X	X
8-1	X	X
8-2	X	X
BF	X	X
HNS	X	X
RAS	X	X
Soil	X	X
WS	X	X

2. Download of data and analysis files

Order placed via the order-portal
The data and analysis results can be downloaded via the order-portal after login with your credentials.
If you would like to download your data via a FTPs client (e.g. FileZilla), you will need to login to the order-portal, navigate to 'Settings', check the 'Make your results also downloadable from our FTP server' box, and save your changes.

FTP connection details:

Login	order-portal username
Password	order-portal password
Host	ftps.baseclear.com
Port	990
Protocol	FTP – File Transfer Protocol
Encryption	Require implicit FTP over TLS

3. Illumina demultiplexing

Single-end or paired-end sequence reads were generated using the Illumina NovaSeq 6000 or MiSeq system. The sequences generated with the MiSeq system were performed under accreditation according to the scope of BaseClear B.V. (L457; NEN-EN-ISO/IEC 17025).
When paired-end sequencing is being performed, the "Number of reads" noted in this report is referring to read pairs. FASTQ read sequence files were generated using bcl2fastq2 version 2.18. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing PhiX control signal were removed using an in-house filtering protocol. In addition, reads containing (partial) adapters were clipped (up to a minimum read length of 50 bp). The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.11.5.

The final quality scores per sample are provided as enclosure.

4. Otu classification

Paired-end sequence reads were collapsed into so-called pseudoreads using sequence overlap with USEARCH version 9.2 (Edgar, 2010).
Classification of these pseudoreads is performed based on the results of alignment with SNAP version 1.0.23 (Zaharia et al., 2011) against the RDP database (Cole et al., 2014) for bacterial organisms, while fungal organisms are classified using the UNITE ITS gene database (Abarenkov et. al, 2010).

The results of the taxonomic classification can be inspected through an interactive online platform, which is an easy way of analysing a metagenomic community. The platform requires no additional software installation and can be accessed through https://genome-explorer.com/. The platform contains a number of unique features among which:

- Phylogenetic tree and OTU bar-chart providing a general overview of the taxonomic composition in all samples.
- Table containing the assigned taxonomies and their relative abundance for each sample.
- Pie-chart (KRONA) which allows interactive interpretation at user-defined taxonomic levels.

Note that all tables and figures can be exported in multiple formats. A free license for the portal is offered for a period of one year (starting on the delivery date mentioned in this report).

To login to the Genome Explorer, you can use the OrderPortal as an authentication method. Navigate to the Genome Explorer and click on the *Sign in with BaseClear's OrderPortal* link. If you have any questions, do not hesitate to contact us at **info@baseclear.com**.

5. Results

A summary of the results is provided below. Sequence and project data are recorded digitally in our secure database and stored for backup purposes only. Data is stored for a period of one month. The result files which have been generated within this project for each analysis are as follows:

Illumina demultiplexing

The files are stored in the "raw_sequences" folder and include:

- The compressed sequence reads in FASTQ format (*.filt.fastq.gz)
- A text file containing the MD5 checksum (fingerprint) of each gzipped FASTQ file ("md5sum_*.txt")

Otu classification

The files are stored in the "otu-classification" folder and include the following files:

- Abundance summary for all taxonomic paths in TXT format (*taxonomic-paths-extended.txt)
- The classification per unique pseudoread in TXT format (*taxonomy-per-unique-read.txt)

Illumina demultiplexing report

Project 122129
Client Victor Lobanov
Date 21-Jan-2020 10:51

Summary of the results - Illumina demultiplexing report

* More information about the average Q-score can be found at the Illumina website (https://www.illumina.com/science/education/sequencing-quality-scores.html)

Statistics after demultiplexing and filtering

Sample name	Number of reads	Yield in mbp	Average quality
1-1	55363	33	34.15
1-2	47050	28	34.77
1-3	48762	29	34.78
2-1	45230	27	34.77
2-2	51750	31	35.02
2-3	49874	29	34.86
3-1	61825	37	34.77
3-2	26311	15	34.75
3-3	44923	26	34.92
4-1	46522	27	34.71
4-2	48638	29	34.81
4-3	46231	27	34.67
5-1	49700	29	34.84
5-2	48289	28	35.02
5-3	47015	28	35.03
6-1	47730	28	34.7
6-2	46390	27	35.01
6-3	45714	27	34.93
7-1	46703	28	34.85
7-2	25633	15	34.81
7-3	49550	29	34.62
8-1	36027	21	34.23
8-2	52204	31	34.8
BF	47778	28	34.93
HNS	50974	30	35.02
RAS	25017	14	34.8
Soil	45189	27	34.95
WS	44696	26	34.74

Otu classification report

Project 122129
Client Victor Lobanov
Date 21-Jan-2020 10:51

Summary of the results - Otu classification report

Taxonomic classification summary statistics

Sample name	Pseudoreads	Pseudoreads percentage	Unique pseudoreads after chimera filtering	Classified reads	Unclassified reads	Chao index	Simpson diversity index	Shannon entropy
1-1	53497.0	96.6	13146	36655	4197	170.4	0.272	3.33
1-2	45762.0	97.3	11668	31609	2698	220.6	0.105	4.40
1-3	47422.0	97.3	11909	32517	3341	213.6	0.0971	4.47
2-1	43890.0	97.0	10354	30435	3030	129.8	0.255	3.27
2-2	50547.0	97.7	11949	35575	1767	149.9	0.107	4.36
2-3	48620.0	97.5	11617	33896	2428	137.7	0.130	3.80
3-1	59917.0	96.9	12907	42199	6371	136.1	0.165	3.62
3-2	25489.0	96.9	7090	15185	1877	118.5	0.239	2.79
3-3	43882.0	97.7	10456	30682	1267	129.3	0.159	3.66
4-1	45151.0	97.1	9346	33196	4737	86.6	0.286	2.61
4-2	47297.0	97.2	9868	35724	4217	67.7	0.191	3.01
4-3	44765.0	96.8	9457	32127	6883	129.4	0.239	2.76
5-1	48496.0	97.6	11694	34347	1397	121.0	0.204	3.55
5-2	47094.0	97.5	12387	33060	749	169.6	0.138	4.29
5-3	45981.0	97.8	10779	34469	419	133.0	0.277	3.17
6-1	46282.0	97.0	10816	33035	3306	95.6	0.154	3.51
6-2	45232.0	97.5	11065	33518	191	148.9	0.227	3.70
6-3	44562.0	97.5	10447	33563	808	144.5	0.289	3.34
7-1	45242.0	96.9	11984	33289	2690	343.2	0.134	4.22
7-2	24944.0	97.3	10402	16377	797	409.4	0.123	4.50
7-3	48269.0	97.4	12582	33224	5125	239.4	0.141	4.09
8-1	34778.0	96.5	10220	24549	5598	361.2	0.120	4.72
8-2	50787.0	97.3	13207	36206	4300	288.5	0.0843	4.71
BF	46558.0	97.4	13263	35957	314	388.6	0.0309	6.10
HNS	49864.0	97.8	10078	39770	397	153.6	0.344	3.03
RAS	24227.0	96.8	9639	17952	435	264.1	0.302	3.25
Soil	44076.0	97.5	17923	34415	1723	908.3	0.0341	7.01
WS	43531.0	97.4	12551	33417	1608	421.0	0.0226	6.48