

## Data Analysis Report

Project / Study: Demo

Project specification: Demo\_Sample

Date: February 21, 2019



# Table of Contents

<b>1</b>	<b>De Novo Assembly Report</b>	<b>1</b>
1.1	Assembly Procedure Description . . . . .	1
1.2	Assembly Data . . . . .	2
1.3	Assembly Statistics . . . . .	3

---

# 1 De Novo Assembly Report

## 1.1 Assembly Procedure Description

Prior to the draft genome assembly all reads were adapter and quality trimmed as described in the previous section.

The assembly has been conducted using our in-house pipeline based on the SPAdes assembler software (v3.10.1).

### Citations:

Anton Bankevich, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, Sergey I. Nikolenko, Son Pham, Andrey D. Prjibelski, Alexey V. Pyshkin, Alexander V. Sirotkin, Nikolay Vyahhi, Glenn Tesler, Max A. Alekseyev, and Pavel A. Pevzner. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* 19(5) (2012), 455-477. doi:10.1089/cmb.2012.0021

Sergey Nurk, Anton Bankevich, Dmitry Antipov, Alexey Gurevich, Anton Korobeynikov, Alla Lapidus, Andrey Prjibelsky, Alexey Pyshkin, Alexander Sirotkin, Yakov Sirotkin, Ramunas Stepanauskas, Jeffrey McLean, Roger Lasken, Scott Clingenpeel, Tanja Woyke, Glenn Tesler, Max Alekseyev, and Pavel Pevzner. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. *Lecture Notes in Computer Science* 7821 (2013), pp. 158-170. doi:10.1007/978-3-642-37195-0\_13

## 1.2 Assembly Data

### \* **\_AllContigs.fasta:**

FASTA file of all the consensus basecalled contigs.

### \* **\_LargeContigs.fasta:**

FASTA file with the nucleotide sequences of all "large" contigs contained in AllContigs.fasta (>=1000 bp)

### \* **\_Scaffolds.fasta:**

Contains the scaffold sequences as FASTA file. Spanned gaps between contigs within scaffolds are represented by 'N's. (paired-end data required).

### \* **\_LargeScaffolds.fasta:**

FASTA file with the nucleotide sequences of all "large" scaffolds contained in Scaffolds.fasta (>=1000 bp)

### \* **\_Scaffolds.txt:**

This AGP file indicates how the scaffold contigs are assembled into scaffolds. See also: [http://www.ncbi.nlm.nih.gov/projects/genome/assembly/agp/AGP\\_Specification.shtml](http://www.ncbi.nlm.nih.gov/projects/genome/assembly/agp/AGP_Specification.shtml).

### 1.3 Assembly Statistics

#### Contig/Scaffold Overview

Sample: Demo_Sample				
	AllContigs	LargeContigs	Scaffolds	LargeScaffolds
Total Entries:	11	4	5	3
Total Length:	147,891	144,828	148,126	147,854
Min Length:	63	1,417	63	1,690
Max Length:	89,184	89,184	143,606	143,606
Mean Length:	13,444	36,207	29,625	49,284
N50:	89,184	89,184	143,606	143,606
N90:	45,619	45,619	2,558	2,558
GC Content:	36.2	36.0	36.2	36.2
#A:	42,930	42,216	42,930	42,871
#T:	51,419	50,429	51,419	51,325
#G:	24,514	23,837	24,514	24,446
#C:	29,028	28,346	29,028	28,977
#N:	0	0	235	235

#### Scaffold Survey - Large Scaffolds

Sample: Demo_Sample				
Scaffold	Length	k-mer coverage	GC content	Number of Contigs
scf_1	143,606	80.4152	35.9	3
scf_2	2,558	1.45625	44.3	4
scf_3	1,690	1.50153	44.6	2

Eurofins Genomics' products, services and applications reach the best quality and safety levels. They are carried out under strict QM and QA systems and comply with the following standards:

**ISO 17025** Accredited analytical excellence

**ISO 13485** Oligonucleotides according to medical devices standard

**GLP** The gold standard to conduct non-clinical safety studies

**GCP** Pharmacogenomic services for clinical studies

**cGMP** Products and testing according to pharma and biotech requirements

Eurofins Genomics Europe Sequencing GmbH • Jakob-Stadler-Platz 7 • 78467 Constance • Germany