

## ***Список использованных команд с #комментариями***

#Making symbolic links

```
ln -s /usr/share/data-minor-bioinf/assembly/oil_R1.fastq
```

```
ln -s /usr/share/data-minor-bioinf/assembly/oil_R2.fastq
```

```
ln -s /usr/share/data-minor-bioinf/assembly/oilMP_S4_L001_R1_001.fastq
```

```
ln -s /usr/share/data-minor-bioinf/assembly/oilMP_S4_L001_R2_001.fastq
```

#Substracting samples

```
seqtk sample -s113 oil_R1.fastq 5000000 > ~/HT1/samples/sub_oil_R1.fq
```

```
seqtk sample -s113 oil_R2.fastq 5000000 > ~/HT1/samples/sub_oil_R2.fq
```

```
seqtk sample -s113 oilMP_S4_L001_R1_001.fastq 1500000 > ~/HT1/sam-  
ples/sub_oilMP_R1.fq
```

```
seqtk sample -s113 oilMP_S4_L001_R2_001.fastq 1500000 > ~/HT1/sam-  
ples/sub_oilMP_R2.fq
```

#Making fastqc reports on samples

```
mkdir fastqc
```

```
ls *.fq | xargs -P 4 -tI{ } fastqc -o fastqc { }
```

#Making cumulative report

```
mkdir multiqc
```

```
multiqc -o multiqc fastqc
```

#Copy report from server to local host (command performed on local host)

```
scp -P 5222 sakomlev@92.242.58.92:~/HT1/samples/multiqc/multiqc_report.html
```

~/

#Trim adapters from paired-end

```
platanus_trim sub_oil_R1.fq sub_oil_R2.fq
```

#Trim adapters from mate-pair

```
platanus_internal_trim sub_oilMP_R1.fq sub_oilMP_R2.fq
```

#Making fastqc reports on trimmed samples

mkdir trim\_fastqc

ls \*trimmed | xargs -P 4 -tI{ } fastqc -o trim\_fastqc { }

#Making cumulative report on trimmed samples

mkdir trim\_multiqc

multiqc -o trim\_multiqc trim\_fastqc

#Copy report from server to local host (command performed on local host)

scp -P 5222 sakomlev@92.242.58.92:~/HT1/samples/trimmed/trim\_multiqc/multiqc\_report.html ~/

#Delete not trimmed sequences (trimmed sequences are stored in another dir, so rm won't affect them)

rm \*.fq\*

#Platanus assemble

time platanus assemble -o Poil -t 4 -n 20 -f ../trimmed/sub\_oil\_R1.fq.trimmed ../trimmed/sub\_oil\_R2.fq.trimmed 2> assemble.log

#Copy contigs from server to local host (command performed on local host)

scp -P 5222 sakomlev@92.242.58.92:~/HT1/samples/assemble/Poil\_contig.fa ~/minor

#Platanus scaffold

time platanus scaffold -o Poil -t 4 -c ../assemble/Poil\_contig.fa -IP1 ../trimmed/sub\_oil\_R1.fq.trimmed ../trimmed/sub\_oil\_R2.fq.trimmed -OP2 ../trimmed/sub\_oilMP\_R1.fq.int\_trimmed ../trimmed/sub\_oilMP\_R2.fq.int\_trimmed 2> scaffold.log

```
#Copy scaffolds from server to local host (command performed on local host)
scp -P 5222 sakomlev@92.242.58.92:~/HT1/samples/scaffolds/Poil_scaffold.fa
~/minor
```

```
#Use screen
screen -S gap_close_sakomlev
```

```
#Platanus gap_close
time platanus gap_close -o Poil -c ../scaffolds/Poil_scaffold.fa -IP1
../trimmed/sub_oil_R1.fq.trimmed ../trimmed/sub_oil_R2.fq.trimmed -OP2
../trimmed/sub_oilMP_R1.fq.int_trimmed ../trimmed/sub_oilMP_R2.fq.int_trimmed 2>
gap_close.log
```

```
#Make fasta file with longest scaffold
echo scaffold1_cov232 > _tmp.txt
seqtk subseq Poil_gapClosed.fa _tmp.txt > longest.fasta
```




```
#Delete trimmed fastq files
rm *trimmed*
```

# Скриншоты и статистика из файлов multiQC

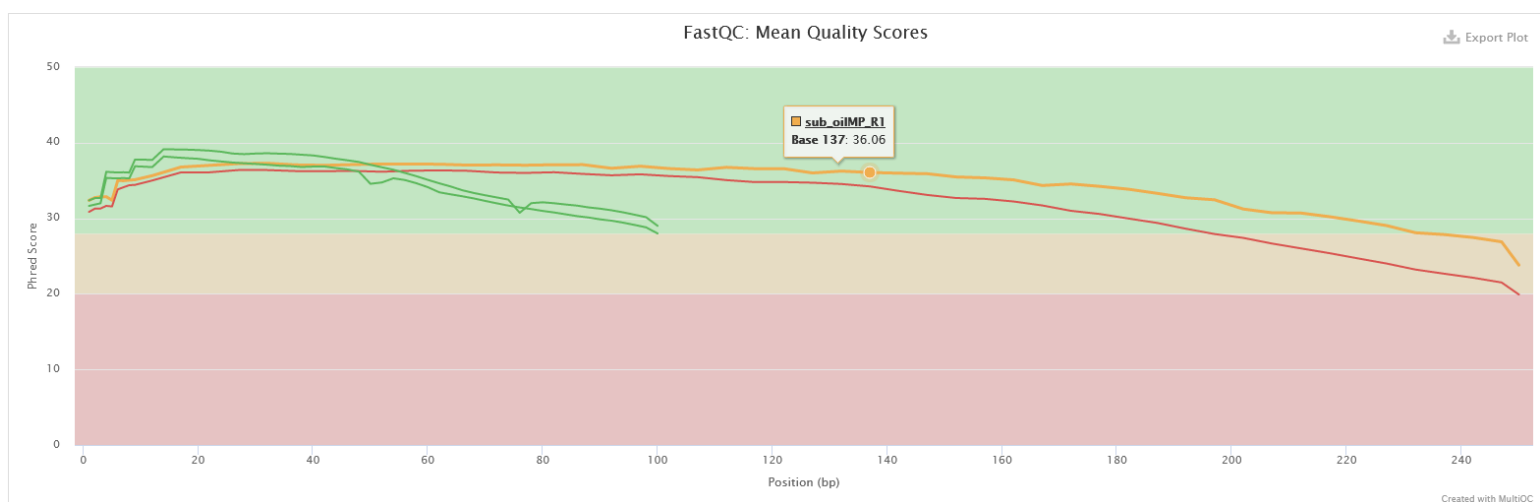
## 1. До подрезания чтений (удаления адаптеров)

Базовая статистика. Стоит обратить внимание на среднюю длину чтений.

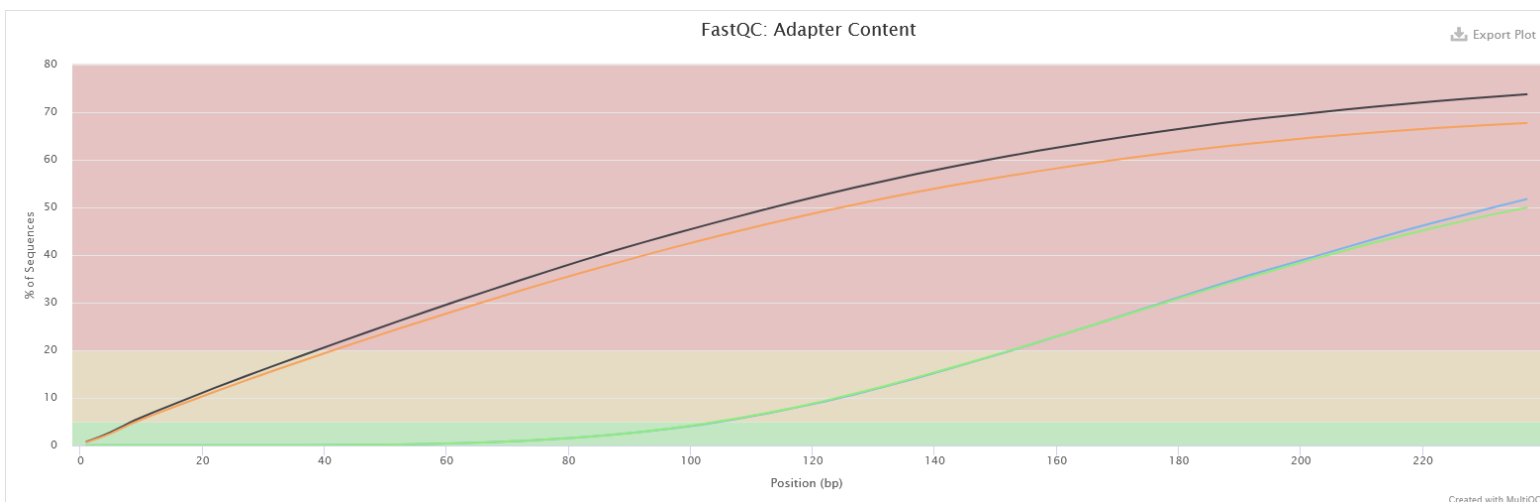
### General Statistics

 Copy table	 Configure Columns	 Plot	Showing $\frac{1}{4}$ rows and $\frac{1}{5}$ columns.	
Sample Name	% Dups	% GC	Length	M Seqs
sub_oilMP_R1	4.5%	44%	251 bp	1.5
sub_oilMP_R2	3.9%	44%	251 bp	1.5
sub_oil_R1	32.3%	46%	101 bp	5.0
sub_oil_R2	30.7%	46%	101 bp	5.0

Качество mate-pairs опускается до неприемлемых значений под конец чтений.



Содержание адаптеров в чтениях высокое



## 2. После подрезания чтений (удаления адаптеров)

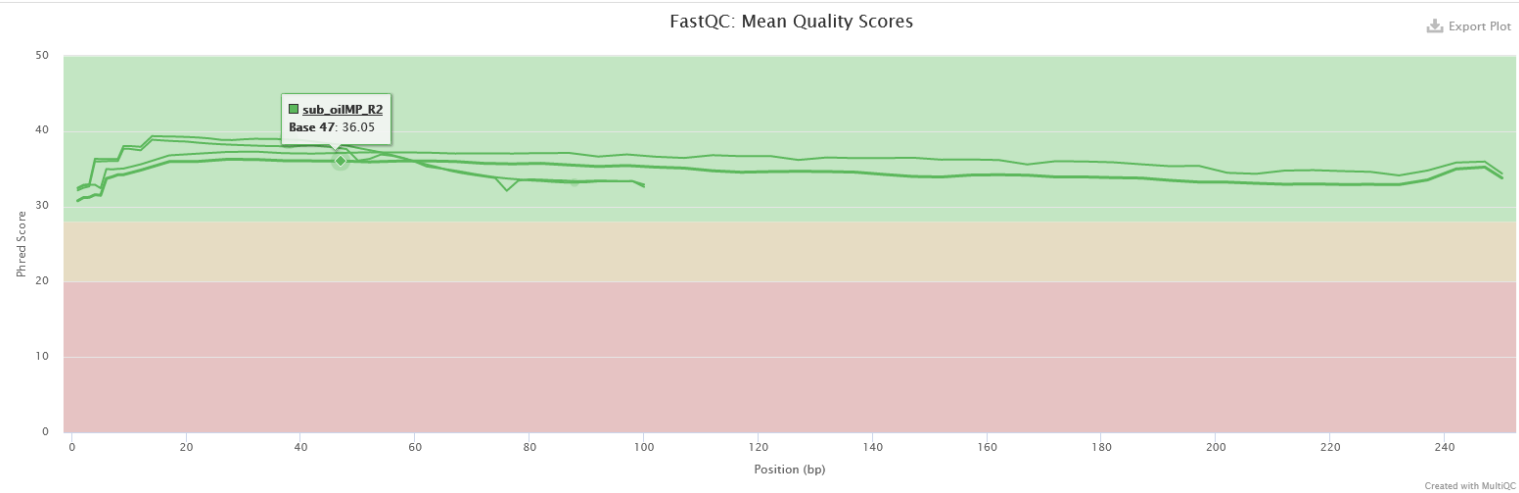
Средняя длина чтений и количество последовательностей заметно сократились.

### General Statistics

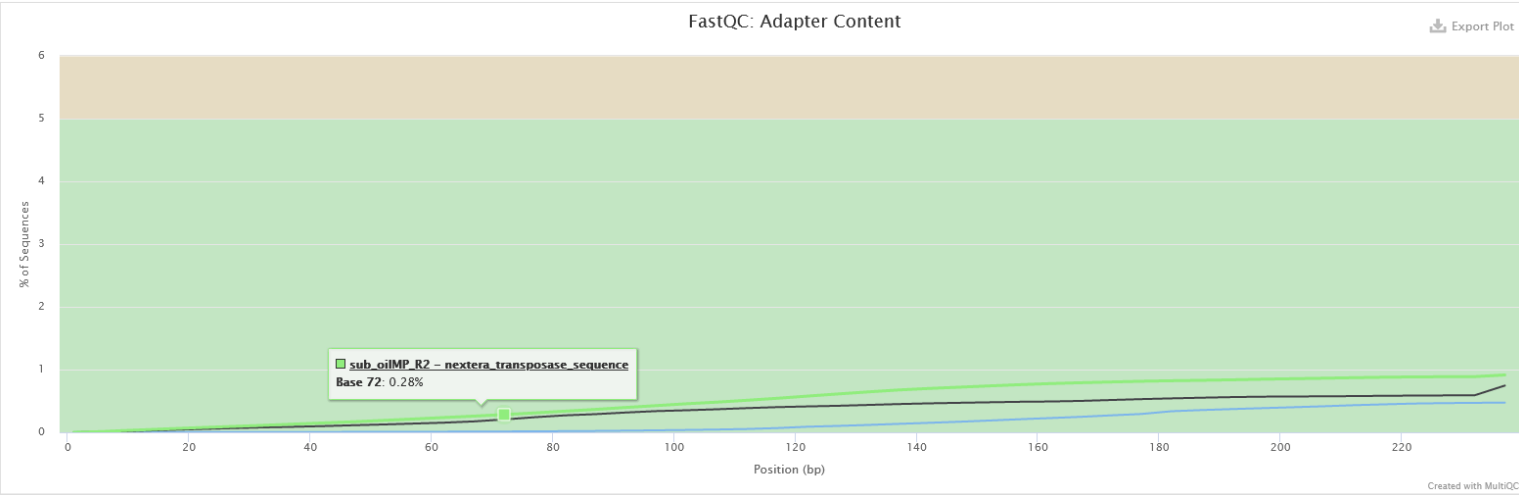
Copy tableConfigure ColumnsPlotShowing 4/4 rows and 4/5 columns.

Sample Name	% Dups	% GC	Length	M Seqs
sub_oilMP_R1	2.4%	46%	111 bp	0.9
sub_oilMP_R2	2.0%	45%	105 bp	0.9
sub_oil_R1	30.6%	46%	97 bp	4.8
sub_oil_R2	29.5%	46%	96 bp	4.8

Качество чтений приемлемое на протяжении всей длины последовательности.



Содержание адаптеров в чтениях не превышает 1%.



## Результаты полученные в Jupyter ноутбуке

### 1. Анализ полученных контигов и скаффолдов

Анализ полученных контигов

```
Ввод [126]: count_basic_assemble_statistics(get_contig_sizes('/home/doctor/minor/HT1/Poil_contig.fa'))  
  
Общее кол-во контигов (скаффолдов): 603  
Общая длина контигов (скаффолдов): 3925361  
Длина самого длинного контига (скаффолда): 179304  
N50: 53989
```

Анализ полученных скаффолдов

```
Ввод [127]: count_basic_assemble_statistics(get_contig_sizes('/home/doctor/minor/HT1/Poil_scaffold.fa'))  
  
Общее кол-во контигов (скаффолдов): 73  
Общая длина контигов (скаффолдов): 3874648  
Длина самого длинного контига (скаффолда): 3836186  
N50: 3836186
```

### 2. Подсчёт гэпов

Подсчёт гэпов до убирания гэпов

```
Ввод [130]: count_gaps(get_longest_scaffold('/home/doctor/minor/HT1/Poil_scaffold.fa'))  
  
Суммарная длина гэпов (количество N): 7336  
Количество гэпов: 61
```

Подсчёт гэпов после убирания гэпов

```
Ввод [131]: count_gaps(get_longest_scaffold('/home/doctor/minor/HT1/Poil_gapClosed.fa'))  
  
Суммарная длина гэпов (количество N): 2851  
Количество гэпов: 11
```