

# Fastq quality data.

Your Name here

October 27, 2020

## Contents

<b>1</b>	<b>Project characteristics</b>	<b>1</b>
<b>2</b>	<b>Global summaries</b>	<b>1</b>
2.1	Project names and read numbers . . . . .	2
<b>3</b>	<b>Nucleotide patterns</b>	<b>2</b>
3.1	N nucleotides . . . . .	3
3.2	GC content . . . . .	4
3.3	Nucleotide frequencies . . . . .	5
<b>4</b>	<b>Phred qualities</b>	<b>7</b>
<b>5</b>	<b>Hierarchical clustering</b>	<b>8</b>

## 1 Project characteristics

---

Project characteristics

Contact

Phone

Institute

Mail

Start date

---

## 2 Global summaries

Input data: Summarized data on FASTQ files.

```
[fastqq] File ( 1/2) '/private/tmp/Rtmp8bv3WJ/Rinst154357ba1379b/seqTools/extdata/g4
```

```
[fastqq] File ( 2/2) '/private/tmp/Rtmp8bv3WJ/Rinst154357ba1379b/seqTools/extdata/g5
```

Printout of Fastqq object:

```
> fqq
```

```
Class      :      Fastqq
nFiles     :          2
maxSeqLen  :        101
k (Kmer len):         4

nReads     :        200
nr  N   nuc :          2
Min seq len :        101
Max seq len :        101
```

## 2.1 Project names and read numbers

```
> dfr<-data.frame(file=basename(fileName(fqq)),
+                 sample=probeLabel(fqq),
+                 reads=format(nReads(fqq), big.mark=Sys.localeconv()[7]))
> print(dfr)
```

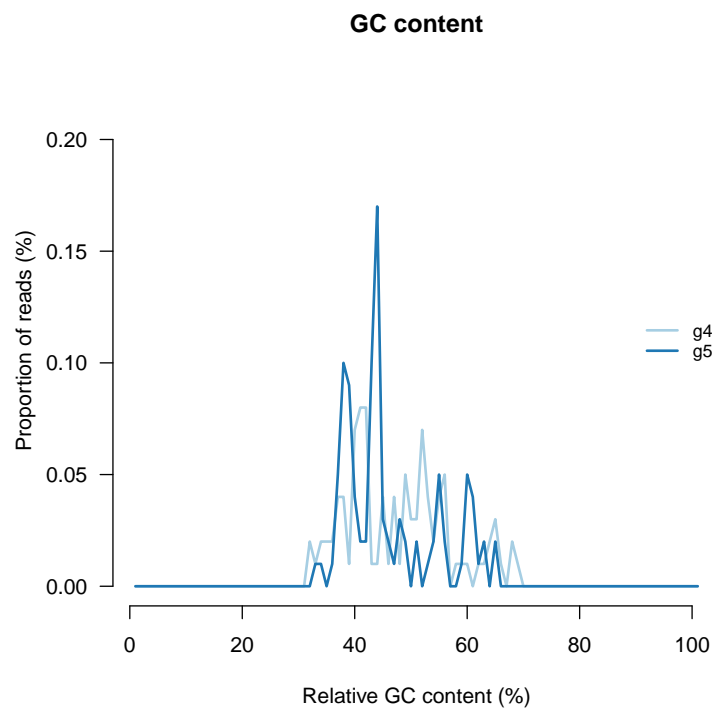
```
      file sample reads
1 g4_l101_n100.fq.gz    g4   100
2 g5_l101_n100.fq.gz    g5   100
```

## 3 Nucleotide patterns

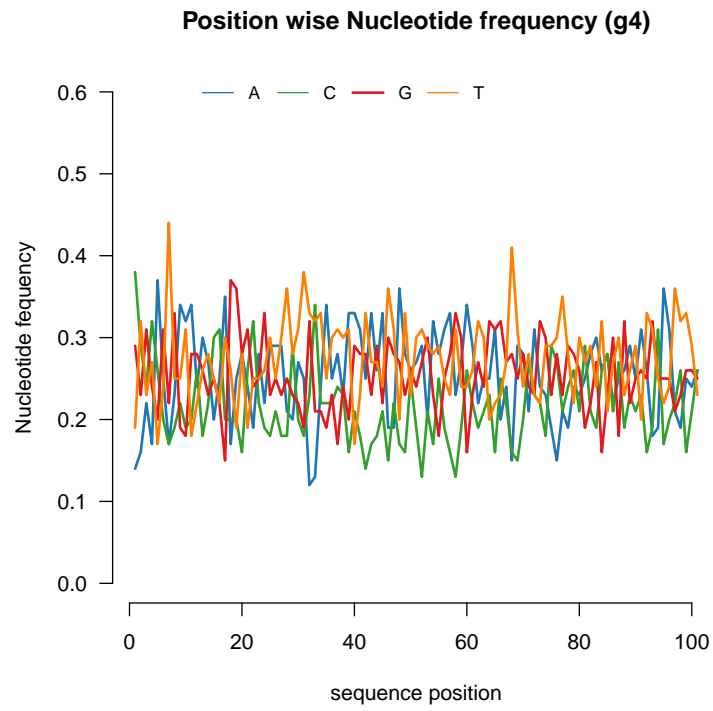
### 3.1 N nucleotides



## 3.2 GC content



### 3.3 Nucleotide frequencies





## 4 Phred qualities





## 5 Hierarchical clustering

1_g4	1
2_g5	2



