

# Bioinformatics Overview

# FASTQ

- FASTA with Quality
- [https://en.wikipedia.org/wiki/FASTQ\\_format](https://en.wikipedia.org/wiki/FASTQ_format)

# FASTQ: Filenames

“What’s in a name? that which we call a **FASTQ**

By any other name would smell as sweet”

— *Romeo and Juliet*, William Shakespeare

# FASTQ: Filenames

1C\_TAAGGCGA\_L006\_R1\_001.fastq.gz  
1C\_TAAGGCGA\_L006\_R1\_002.fastq.gz  
1C\_TAAGGCGA\_L006\_R1\_003.fastq.gz  
1C\_TAAGGCGA\_L006\_R1\_004.fastq.gz  
1C\_TAGGCATG\_L003\_R1\_001.fastq.gz  
1C\_TAGGCATG\_L005\_R1\_001.fastq.gz  
2C\_CGTACTAG\_L003\_R1\_001.fastq.gz

# FASTQ: Filenames

Sample Name	Barcode	Lane #	Read #	File Part
1C	TAAGGCGA	L006	R1	001.fastq.gz
1C	TAAGGCGA	L006	R1	002.fastq.gz
1C	TAAGGCGA	L006	R1	003.fastq.gz
1C	TAAGGCGA	L006	R1	004.fastq.gz
1C	TAGGCATG	L003	R1	001.fastq.gz
1C	TAGGCATG	L005	R1	001.fastq.gz
2C	CGTACTAG	L003	R1	001.fastq.gz

# FASTQ: Format

- **FASTA Format**

```
>NC_007779.1 Escherichia coli str. K-12 substr. W3110, complete genome
AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGTCTGATAGCAGCTTCTGAACTG
GTTACCTGCCGTGAGTAAATTTAAATTTTATTGACTTAGGTCACTAAATACTTTAACCAATATAGGCATAGCGCACAGAC
AGATAAAAATTACAGAGTACACAACATCCATGAAACGCATTAGCACCACCATTACCACCACCATCACCATTACCACAGGT
AACGGTGCGGGCTGACGCGTACAGGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTTCGACCAAAGG
TAACGAGGTAACAACCATGCGAGTGTTGAAGTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTTGCCG
ATATTCTGGAAAGCAATGCCAGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGGTG
GCGATGATTGAAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTTTGCCGAACCTTTT
GACGGGACTCGCCGCCGCCAGCCGGGGTTCCCGCTGGCGCAATTGAAAACCTTTCGTCGATCAGGAATTTGCCCAAATAA
AACATGTCCTGCATGGCATTAGTTTGTGGGGCAGTGCCCGGATAGCATCAACGCTGCGCTGATTTGCCGTGGCGAGAAA
```

Header  
Sequence

- **FASTQ Format**

```
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTGCGGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG
+
AAABBBFFFFFGGGGGGGGGGGGHHHHHHHHGHGHGHHHHHHHGGGGGGHHHHGGGGGGGGHHHHGHFFHHHHHHGHGGGGGGGGGGHHHHHHHHHHHHGGG
```

Quality  
Score

# FASTQ: Format

# Quality Scores

- Go to Rmarkdown:

[https://github.com/ibiem-2019/ibiem\\_2019\\_material/blob/master/content/lessons/quality\\_scores.Rmd](https://github.com/ibiem-2019/ibiem_2019_material/blob/master/content/lessons/quality_scores.Rmd)



# FASTQ: Read Files

## Combined\_R1.fastq.gz

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0
TTACGCTAACAGGCGGTAGCCTGGCAGGGTCAGGAAATCAATTAACATCATCGGAAGTGGTGATCTGTTCCATCAAGCGTGCGGCATCGTCAAAACGCCC
+
ABBBABBBBAFFFGGGGGGGGGHGGHGGGCG2GF3FFGHHHHHHGGFGHEHHGGGEHHHHAGGHHGHHHFFDHFHHHGEGGGG@F@H?GHH/GBEFGGG
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTTCGCGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG
+
AAABBFFFFFFFFFFGGGGGGGGGGGGHGGHHHHHHHGHGHHGGHHHHGGGGGGGGGGHGGHHHFFHHHHHHGHGGGGGGGGGGHGGHHHHHHHHHHGGG
```

## Combined\_R2.fastq.gz

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 2:N:0:0
GGAAGATGCGGCGACGGCTGAAATTTCCCGTACCTCGATCTGGCAGTGGATCCATCATCAAAAACGTTGAGCAATGGCAAACCGGTGACCAAAGCCTTGT
+
ABBABFFFFDBDGC??FFGGGHGHFEG3EAEGGFHAE3GFBGGHGGGHHCFGHFGBGHFHHDFFEGGHFHEFH3BFGF0GFEGGGGGHHA/FGHFHHH
@M00698:36:000000000-AFBEL:1:1101:16483:1412 2:N:0:0
GCTTCTTCCGTACTCATGCGGGCATTGAGCAAGCGATCAGCCGTGGCCTGGCGTATGCGCCATATGCTGACCTGGTCTGGTGTGAAACCTCCACGCCGGAT
+
CCCCCFFFBFFGGGGGGGGCECGHHHHHHHHHGGHGGGGHGGCGCHHGFHGGGGHGGGGGGHHHHHHHHHHHHHHHEHGHGGHHHHHGGGGGGG
```

## Combined\_I1.fastq.gz

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0
AGTTCC
+
CCCCDF
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CCTGTC
+
A11>>1
```

# Demultiplexing

**Undetermined S0 L001 R1 001.fastq.gz**

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0
TTACGCTAACAGGCGGTAGCCTGGCAGGGTCAGGAAATCAATTAACATCGGAAGTGGTGATCTGTTCCATCAAGCGTGCGGCATCGTCAAACGCCC
+
ABBBABBBBAFFFGGGGGGGGGGGHGGHGGGCG2GF3FFGHHHHHHGGFGHEHHGGGEHHHHAGGHHGHHHFFDHFHHHGEGGGG@F@H?GHH/GBEFGGG
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0
CTGCCAGTTGAACGACGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTTCGCGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG
+
AAABBFFFFFFFFFGGGGGGGGGGGGGHHHHHHHHGHGHGHHHHHHGHHHGGGGGHHHHGGGGGGGGHHHHGHHFFHHHHHHGHGGGGGGGGGGHHHHHHHHHHHHGGG
```

**Undetermined S0 L001 R2 001.fastq.gz**

```
@M00698:36:000000000-AFBEL:1:1101:14738:1412 2:N:0:0  
GGAAGATGCGGCGACGGCTGAAATTTCCCGTACCTCGATCTGGCAGTGATCCATCATCAAAAACGTTGAGCAATGGCAAACCGGTGACCAAAGCCTTGT  
+  
ABBABFFFFDBDGC??FFGGGHGFEG3EAEGGFHAE3GFBGGHGGGHHCFGHFGBGHFHHDfEGGHFHEFH HHH3BFGF0GFEGGGGGHHA/FGHFHHH  
@M00698:36:000000000-AFBEL:1:1101:16483:1412 2:N:0:0  
GCTTCTTCGTACTCATGCGGGCATTGAGCAAGCGATCAGCCGTGGCCTGGCGTATGCGCCATATGCTGACCTGGTCTGGTGTGAAACCTCCACGCCGGAT  
+  
CCCCFFFFBFfGGGGGGGGCECGHHHHHHHHHHGGHG GGGHGGCGCHHGFHGGGGHHGGGGGHHHHHHHHHHHHHHHEHGHHHGHHHHGGGGGG
```

Sample	Barcode
7A_K	AGTCAA
7B_K	AGTTCC
7C_N	CGTACG
8A_N	GAGTGG
8B_N	CCTGTC
8C_N	ATTCCT
7A_P	CGATGT

**Undetermined S0 L001 I1 001.fastq.gz**

The diagram shows two DNA sequences. The first sequence, on a light blue background, is:  
@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0  
AGTTCC  
+  
CCCCDF  
An arrow points from the 'AGTTCC' sequence to a label '7B\_K' in a white box with a black border. The second sequence, on a light purple background, is:  
@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0  
CCTGTC  
+  
A11>>1  
An arrow points from the 'CCTGTC' sequence to a label '8B\_N' in a white box with a black border.

# Demultiplexing

## **7B K R1 001.fastq.gz**

@M00698:36:000000000-AFBEL:1:1101:14738:1412 1:N:0:0

TTACGCTAACAGGCGGTAGCCTGGCAGGGTCAGGAAATCAATTAACATCATCGGAAGTGGTGATCTGTTCCATCAAGCGTGCGGCATCGTCAAAACGCCC

+

ABBBABBBBAFFFFGGGGGGGGGGHGGHGGGCG2GF3FFGHHHHHHGGFGHEHHGGGEHHHHAGGHHGHHHFFDHFHHHGEGGGG@F@H?GHH/GBEFGGG

## **7B K R2 001.fastq.gz**

@M00698:36:000000000-AFBEL:1:1101:14738:1412 2:N:0:0

GGAAGATGCGGCGACGGCTGAAATTTCCCGTACCTCGATCTGGCAGTGGATCCATCATCAAAAAACGTTGAGCAATGGCAAACCGGTGACCAAAGCCTTGT

+

ABBABFFFFDBDGC??FFGGGGHGHFEG3EAEGGFHAE3GFBGGHGGGHHCFGHFGBGHFHHDFFEGGHFHEFHFFFH3BFGF0GFEGGGGGHHA/FGHFHHH

## **8B N R1 001.fastq.gz**

@M00698:36:000000000-AFBEL:1:1101:16483:1412 1:N:0:0

CTGCCAGTTGAACGACGGCGAGCAGTTATAAGCCAGCAGTTTGCCCGGATATTTGCGGTGGATAGCTTGTGCAAAGCGACGCGCCAGTTCCAGATCCGGCG

+

AAABBFFFFFFFFFGGGGGGGGGGGHGGHHHHHHHGHGHHHHHHHGGGGGGHHHHGGGGGGGGHHHHHGHFFHHHHHHGHGGGGGGGGGGHHHHHHHHHHHGGG

## **8B N R2 001.fastq.gz**

@M00698:36:000000000-AFBEL:1:1101:16483:1412 2:N:0:0

GCTTCTTCGTACTCATGCGGGCATTGAGCAAGCGATCAGCCGTGGCCTGGCGTATGCGCCATATGCTGACCTGGTCTGGTGTGAAACCTCCACGCCGGAT

+

CCCCCFFFBFFGGGGGGGGGCECGHHHHHHHHHGGHGGGGHGGCGCHHGFHGGGGHHGGGGHHHHHHHGHHHHHHHHHHHHEHGHGHGHGHGGGGGGG