

Data Wrangling and Processing for Genomics - Parte 3

Evidencia:

```
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq$ cp ~/src/Trimmomatic-0.38/adapters/NexteraPE-PE.fa .
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq$ trimmomatic PE SRR2589044_1.fastq.gz SRR2589044_2.fastq.gz \
> SRR2589044_1.trim.fastq.gz SRR2589044_1un.trim.fastq.gz \
> SRR2589044_2.trim.fastq.gz SRR2589044_2un.trim.fastq.gz \
> SLIDINGWINDOW:4:20 MINLEN:25 ILLUMINACLIP:NexteraPE-PE.fa:2:40:15
trimmomatic: command not found
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq$ java -jar ~/src/Trimmomatic-0.38/trimmomatic-0.38.jar PE SRR2589044_1.fastq.gz SRR2589044_2.fastq.gz SRR2589044_1.trim.fastq.gz SRR2589044_1un.trim.fastq.gz SRR2589044_2.trim.fastq.gz SRR2589044_2un.trim.fastq.gz SLIDINGWINDOW:4:20 MINLEN:25 ILLUMINACLIP:NexteraPE-PE.fa:2:40:15
TrimmomaticPE: Started with arguments:
SRR2589044_1.fastq.gz SRR2589044_2.fastq.gz SRR2589044_1.trim.fastq.gz SRR2589044_1un.trim.fastq.gz SRR2589044_2.trim.fastq.gz SRR2589044_2un.trim.fastq.gz SLIDINGWINDOW:4:20 MINLEN:25 ILLUMINACLIP:NexteraPE-PE.fa:2:40:15
Multiple cores found: Using 4 threads
Using PrefixPair: 'AGATGTGTATAAGAGACAG' and 'AGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTCCGAGCCACGAGAC'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTGACGCTGCCGACGA'
ILLUMINACLIP: Using 1 prefix pairs, 4 forward/reverse sequences, 0 forward only sequences, 0 reverse only sequences
Exception in thread "main" java.io.FileNotFoundException: SRR2589044_2.fastq.gz (No such file or directory)
    at java.base/java.io.FileInputStream.open0(Native Method)
    at java.base/java.io.FileInputStream.open(FileInputStream.java:219)
    at java.base/java.io.FileInputStream.<init>(FileInputStream.java:157)
    at org.usadellab.trimmomatic.fastq.FastqParser.parse(FastqParser.java:135)
    at org.usadellab.trimmomatic.TrimmomaticPE.process(TrimmomaticPE.java:268)
    at org.usadellab.trimmomatic.TrimmomaticPE.run(TrimmomaticPE.java:555)
    at org.usadellab.trimmomatic.Trimmomatic.main(Trimmomatic.java:80)
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq$
```

Se puede ver una Excepción y tuve que bajar de nuevo UNA sola secuencia que al parecer no se descargó con el curl, talvez por no poner el comando sudo al inicio.

```

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq$ java -jar ~/src/Trimmomatic-0
.38/trimmomatic-0.38.jar PE SRR2589044_1.fastq.gz SRR2589044_2.fastq.gz SRR2589044_1.trim.fast
q.gz SRR2589044_1un.trim.fastq.gz SRR2589044_2.trim.fastq.gz SRR2589044_2un.trim.fastq.gz SLID
INGWINDOW:4:20 MINLEN:25 ILLUMINACLIP:NexteraPE-PE.fa:2:40:15
TrimmomaticPE: Started with arguments:
  SRR2589044_1.fastq.gz SRR2589044_2.fastq.gz SRR2589044_1.trim.fastq.gz SRR2589044_1un.trim.fas
tq.gz SRR2589044_2.trim.fastq.gz SRR2589044_2un.trim.fastq.gz SLIDINGWINDOW:4:20 MINLEN:25 IL
LUMINACLIP:NexteraPE-PE.fa:2:40:15
Multiple cores found: Using 4 threads
Using PrefixPair: 'AGATGTGTATAAGAGACAG' and 'AGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTCCGAGCCCACGAGAC'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTGACGCTCCGACGA'
ILLUMINACLIP: Using 1 prefix pairs, 4 forward/reverse sequences, 0 forward only sequences, 0 r
everse only sequences
Quality encoding detected as phred33
Input Read Pairs: 1107090 Both Surviving: 885220 (79.96%) Forward Only Surviving: 216472 (19.5
5%) Reverse Only Surviving: 2850 (0.26%) Dropped: 2548 (0.23%)
TrimmomaticPE: Completed successfully

```

```

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq$ for infile in *.1.fastq.gz; do base=$(basename ${infile} .1.fastq.gz); java -jar ~/src/trimmomatic-0.38/trimmomatic-0.38.jar PE ${infile}
) ${base}.2.fastq.gz ${base}.1.trim.fastq.gz ${base}.1un.trim.fastq.gz ${base}.2.trim.fastq.gz ${base}.2un.trim.fastq.gz SLIDINGWINDOW:4:20 MINLEN:25 ILLUMINACLIP:NexteraPE-PE.fa:2:40:15; done
TrimmomaticPE: Started with arguments:
  SRR2584863_1.fastq.gz SRR2584863_2.fastq.gz SRR2584863_1.trim.fastq.gz SRR2584863_1un.trim.fastq.gz SRR2584863_2.trim.fastq.gz SRR2584863_2un.trim.fastq.gz SLIDINGWINDOW:4:20 MINLEN:25 ILLUMINACLIP:Ne
xteraPE-PE.fa:2:40:15
Multiple cores found: Using 4 threads
Using PrefixPair: 'AGATGTGTATAAGAGACAG' and 'AGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTCCGAGCCCACGAGAC'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTGACGCTCCGACGA'
ILLUMINACLIP: Using 1 prefix pairs, 4 forward/reverse sequences, 0 forward only sequences, 0 reverse only sequences
Quality encoding detected as phred33
Input Read Pairs: 1553259 Both Surviving: 1271924 (81.89%) Forward Only Surviving: 273488 (17.61%) Reverse Only Surviving: 3861 (0.25%) Dropped: 3986 (0.26%)
TrimmomaticPE: Completed successfully
TrimmomaticPE: Started with arguments:
  SRR2584866_1.fastq.gz SRR2584866_2.fastq.gz SRR2584866_1.trim.fastq.gz SRR2584866_1un.trim.fastq.gz SRR2584866_2.trim.fastq.gz SRR2584866_2un.trim.fastq.gz SLIDINGWINDOW:4:20 MINLEN:25 ILLUMINACLIP:Ne
xteraPE-PE.fa:2:40:15
Multiple cores found: Using 4 threads
Using PrefixPair: 'AGATGTGTATAAGAGACAG' and 'AGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTCCGAGCCCACGAGAC'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTGACGCTCCGACGA'
ILLUMINACLIP: Using 1 prefix pairs, 4 forward/reverse sequences, 0 forward only sequences, 0 reverse only sequences
Quality encoding detected as phred33
Input Read Pairs: 2768398 Both Surviving: 2053874 (74.19%) Forward Only Surviving: 578725 (20.90%) Reverse Only Surviving: 131060 (4.73%) Dropped: 4739 (0.17%)
TrimmomaticPE: Completed successfully
TrimmomaticPE: Started with arguments:
  SRR2589044_1.fastq.gz SRR2589044_2.fastq.gz SRR2589044_1.trim.fastq.gz SRR2589044_1un.trim.fastq.gz SRR2589044_2.trim.fastq.gz SRR2589044_2un.trim.fastq.gz SLIDINGWINDOW:4:20 MINLEN:25 ILLUMINACLIP:Ne
xteraPE-PE.fa:2:40:15
Multiple cores found: Using 4 threads
Using PrefixPair: 'AGATGTGTATAAGAGACAG' and 'AGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTCCGAGCCCACGAGAC'
Using Long Clipping Sequence: 'CTGTCTCTTATACACATCTGACGCTCCGACGA'
ILLUMINACLIP: Using 1 prefix pairs, 4 forward/reverse sequences, 0 forward only sequences, 0 reverse only sequences
Quality encoding detected as phred33
Input Read Pairs: 1107090 Both Surviving: 885220 (79.96%) Forward Only Surviving: 216472 (19.55%) Reverse Only Surviving: 2850 (0.26%) Dropped: 2548 (0.23%)
TrimmomaticPE: Completed successfully
estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq$

```

Y los movemos a una capeta nueva de ya trimado:

TrimmomaticPE: Completed successfully

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq\$ ls

```
NexteraPE-PE.fa      SRR2584866_1.fastq.gz      SRR2589044_1.trim.fastq.gz
SRR2584863_1.fastq.gz  SRR2584866_1.trim.fastq.gz  SRR2589044_1un.trim.fastq.gz
SRR2584863_1.trim.fastq.gz  SRR2584866_1un.trim.fastq.gz  SRR2589044_2.fastq.gz
SRR2584863_1un.trim.fastq.gz  SRR2584866_2.fastq.gz      SRR2589044_2.trim.fastq.gz
SRR2584863_2.fastq.gz      SRR2584866_2.trim.fastq.gz  SRR2589044_2un.trim.fastq.gz
SRR2584863_2.trim.fastq.gz  SRR2584866_2un.trim.fastq.gz
SRR2584863_2un.trim.fastq.gz  SRR2589044_1.fastq.gz
```

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq\$ cd ~/dc_workshop/data/untrimmed_fastq

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq\$ mkdir ../trimmed_fastq

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq\$ mv *.trim* ../trimmed_fastq

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/untrimmed_fastq\$ cd ../trimmed_fastq

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/trimmed_fastq\$ ls

```
SRR2584863_1.trim.fastq.gz  SRR2584866_1.trim.fastq.gz  SRR2589044_1.trim.fastq.gz
SRR2584863_1un.trim.fastq.gz  SRR2584866_1un.trim.fastq.gz  SRR2589044_1un.trim.fastq.gz
SRR2584863_2.trim.fastq.gz  SRR2584866_2.trim.fastq.gz  SRR2589044_2.trim.fastq.gz
SRR2584863_2un.trim.fastq.gz  SRR2584866_2un.trim.fastq.gz  SRR2589044_2un.trim.fastq.gz
```

estuardo8u14@LAPTOP-5IN4BIR3:~/dc_workshop/data/trimmed_fastq\$