

February 9, 2021

Server Science Samples

An orange speech bubble with a tail pointing towards the bottom left, containing the text "how to stop typing manually?".

how to stop typing
manually?

by Maria Shumilova

During the work with biological data on server I recognized that I had a problem:

I need to type a code in a terminal manually for each sample of my data.

AWFUL!

feel free to smile)

**Let's meet with my
brilliant idea**

PLAN

my script should do the following things:

1

find all samples in a directory_

2

for each sample:

1) find a pair of the sample
(difference between them is _R1_ or _R2_ in the names)

2) change the names of my _R1_ and _R2_ samples for output of a program for paired and unpaired reads

3

indicate the program for terminal

Let's Try!

Step by step

1

find all
samples
in a directory

os.path
for
rglob

```
from pathlib import Path  
import os
```

```
for el in Path('/scratch/mshumilova/filtered_by_tile/').rglob('*'):
```

for each sample:

2

1) find a pair of the sample
(difference between them is _R1_ or _R2_
in the names)

replace

```
from pathlib import Path
import os
```

```
pairname = ''
```

```
for el in Path('/scratch/mshumilova/filtered_by_tile/').rglob('*'):
    if '_R1_' not in str(el):
        continue
    pairname = str(el).replace('_R1_', '_R2_')
```

for each sample:

2

2) change the names of my _R1_ and _R2_ samples for output of a program for paired and unpaired reads

replace

```
from pathlib import Path
import os
```

```
pairname = ''
out1_paired = ''
out1_unpaired = ''
out2_paired = ''
out2_unpaired = ''
```

```
for el in Path('/scratch/mshumilova/filtered_by_tile/').rglob('*'):
    if '_R1_' not in str(el):
        continue
    pairname = str(el).replace('_R1_', '_R2_')
```

```
out1_paired = str(el).replace('/filtered_by_tile/', '/trimm/')
out1_paired = out1_paired.replace('.fq.gz', '.trimm_20_paired_1.fq.gz')
out1_unpaired = out1_paired.replace('_paired', '_unpaired')
```

```
out2_paired = pairname.replace('/filtered_by_tile/', '/trimm/')
out2_paired = pairname.replace('.fq.gz', '.trimm_20_paired_2.fq.gz')
out2_unpaired = out2_paired.replace('_paired', '_unpaired')
```


3

indicate the program for terminal

```
from pathlib import Path
import os
```

```
pairname = ''
out1_paired = ''
out1_unpaired = ''
out2_paired = ''
out2_unpaired = ''
```

```
for el in Path('/scratch/mshumilova/filtered_by_tile/').rglob('*'):
```

```
    if '_R1_' not in str(el):
```

```
        continue
```

```
    pairname = str(el).replace('_R1_', '_R2_')
```

```
    out1_paired = str(el).replace('/filtered_by_tile/', '/trimm/')
    out1_paired = out1_paired.replace('.fq.gz', '.trimm_20_paired_1.fq.gz')
```

```
    out1_unpaired = out1_paired.replace('_paired', '_unpaired')
```

```
    out2_paired = pairname.replace('/filtered_by_tile/', '/trimm/')
    out2_paired = pairname.replace('.fq.gz', '.trimm_20_paired_2.fq.gz')
```

```
    out2_unpaired = out2_paired.replace('_paired', '_unpaired')
```

```
    cmd = f'java -jar Trimmomatic-0.39/trimmomatic-0.39.jar PE -phred33
```

```
{str(el)} {pairname} {out1_paired} {out1_unpaired} {out2_paired}
```

```
{out2_unpaired} LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:60'
```

```
os.system(cmd)
```

What does our output look like?

Let's see

for one sample

OUTPUT

```
java -jar Trimmomatic-0.39/trimmomatic-0.39.jar PE -phred33  
/scratch/mshumilova/filtered_by_tile/FL865_S10_L001_R1_001.fq.gz  
/scratch/mshumilova/filtered_by_tile/FL865_S10_L001_R2_001.fq.gz  
/scratch/mshumilova/trimm/FL865_S10_L001_R1_001.trim_20_paired_1.fq.gz  
/scratch/mshumilova/trimm/FL865_S10_L001_R1_001.trim_20_unpaired_1.fq.gz  
/scratch/mshumilova/trimm/FL865_S10_L001_R2_001.trim_20_paired_1.fq.gz  
/scratch/mshumilova/trimm/FL865_S10_L001_R2_001.trim_20_unpaired_1.fq.gz  
LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:60
```

1

program part

2

pair of samples (_R1_, _R2_)

3

output of the program (_R1_, paired, unpaired)

4

output of the program (_R2_, paired, unpaired)

1

program part

*Thank you for your attention.
I wish you to have a factastic day!*

I've solved my
"manual" problem
and use the script
above for my tasks.
I'm happy)

P.s. recently I found one more problem.
to be continued...