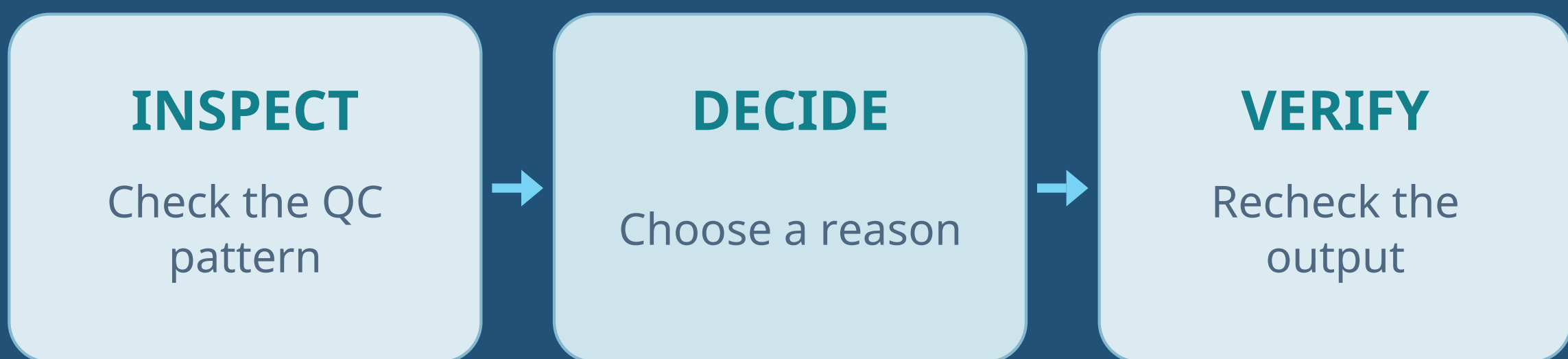


Should we trim every read?

When should you trim, filter, or keep reads unchanged?

FASTQ quality control is a decision point.



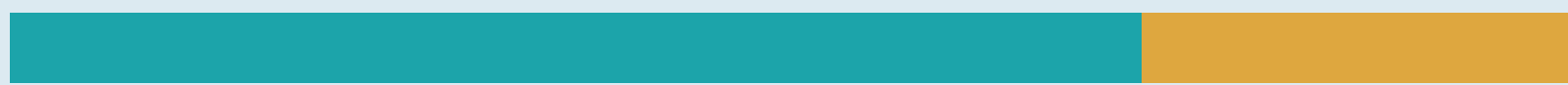
KEY IDEA

More trimming is not always better.

Trimming vs filtering

These operations change a FASTQ file in different ways.

TRIMMING



Remove unwanted bases; keep a shorter read.

Example: poor-quality bases at the read end.

FILTERING



Discard the whole read if it fails a chosen rule.

Example: too short after trimming.

REMEMBER

Both operations can discard useful information.

Read the warning first

The FastQC pattern and your experiment guide the action.

1

Known adapter sequence

Trim the unwanted adapter, then check the result.

2

Low-quality read ends

Consider tail trimming only when it matters downstream.

3

Expected GC or duplication bias

Investigate the library design. Do not trim automatically.

RULE

Do not trim just to turn FastQC flags green.

Set up the practice

One small tool, three synthetic FASTQ datasets.

1 Install fastp in Ubuntu (once only)

```
sudo apt update  
sudo apt install fastp -y  
fastp --version
```

2 Prepare the Part 6 folder

```
mkdir -p ~/bioinformatics_training/part6  
cd ~/bioinformatics_training/part6
```

3 Copy the four practice files here

```
explorer.exe .
```

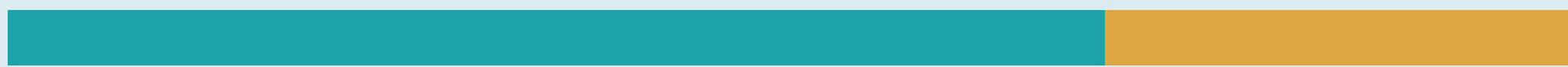
From the downloaded ZIP, copy:

good.fq.gz tail.fq.gz adapter.fq.gz a.fa

Exercise A: trim a poor tail

Keep the reliable part of the read when appropriate.

BEFORE



70 high-quality bases + 30 low-quality bases

AFTER



Run in the Part 6 folder:

```
fastp -i tail.fq.gz -o tail.clean.fq.gz -r -l 50
```

What the options do

-r trims low-quality right-side regions. -l 50 excludes reads shorter than 50 bases.

Exercise B: trim an adapter

We know the adapter sequence used in this practice file.

READ BEFORE



Biological insert (67 bases) + adapter (33 bases)

READ AFTER



Run with the supplied adapter file:

```
fastp -i adapter.fq.gz -o a.clean.fq.gz --adapter_fasta a.fa
```

WHAT IS a.fa?

It contains the known adapter sequence in FASTA format. Check the real library before choosing adapters.

Quality control again

Always compare the original reads with the processed reads.

Run FastQC on both originals and outputs:

```
fastqc *.fq.gz
```

Open the HTML reports in Windows:

```
explorer.exe .
```

1

Quality profiles

Did the low-quality tail become shorter or disappear?

2

Adapter Content

Is less unwanted adapter sequence detectable now?

Was the trade-off worth it?

Cleaner reads are helpful only if you keep enough useful data.

1

Read length

Did trimming remove too much sequence?

2

Reads retained

How many reads passed filtering?

3

QC patterns

Did the targeted warning improve?

4

Biological question

Is the remaining data suitable for mapping?

Your turn

Compare the evidence, not just the colours on FastQC.

1

Run the two fastp commands.

2

Re-run FastQC on original and cleaned reads.

3

What changed in the quality and adapter plots?

4

Would you trim good.fq.gz? Why or why not?

NEXT

Part 7: Map reads against a reference genome.