



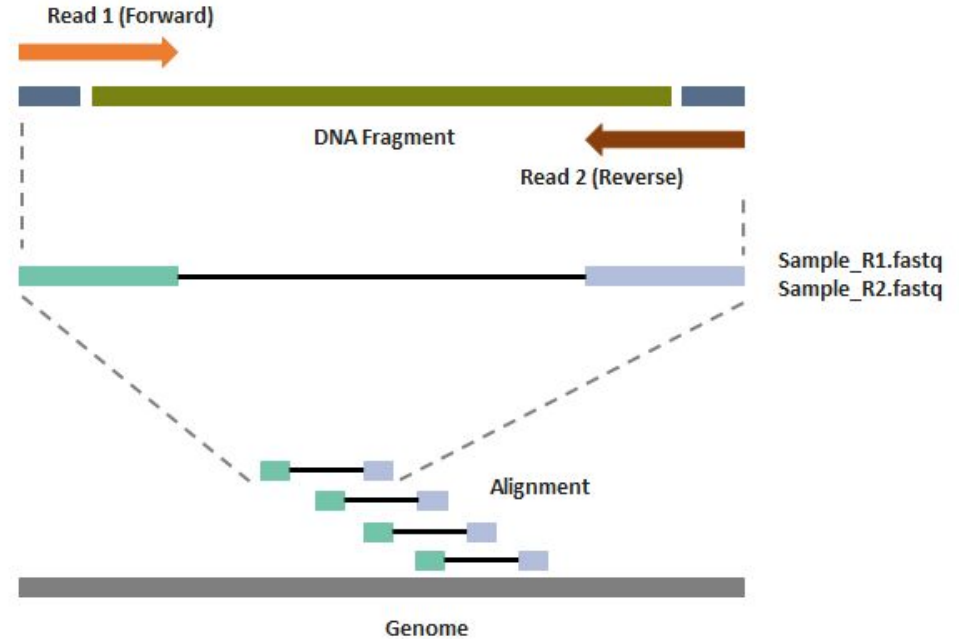
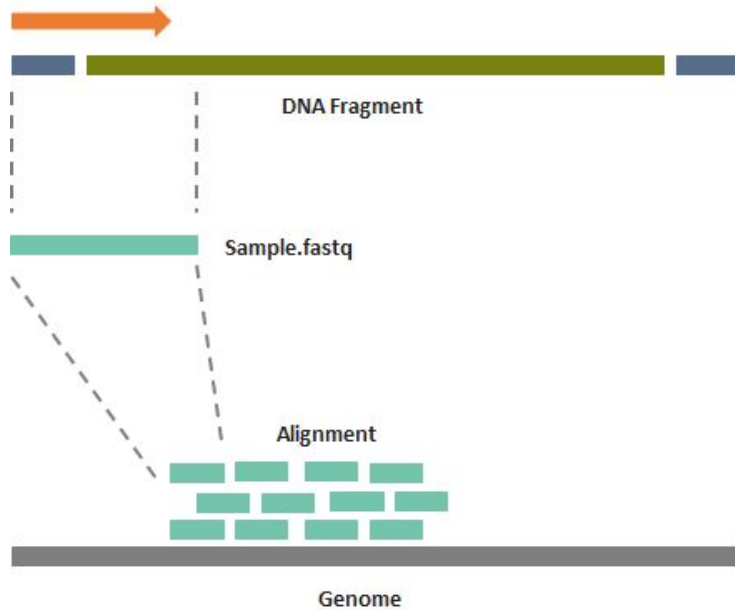
Quality assessment

Raw read QC

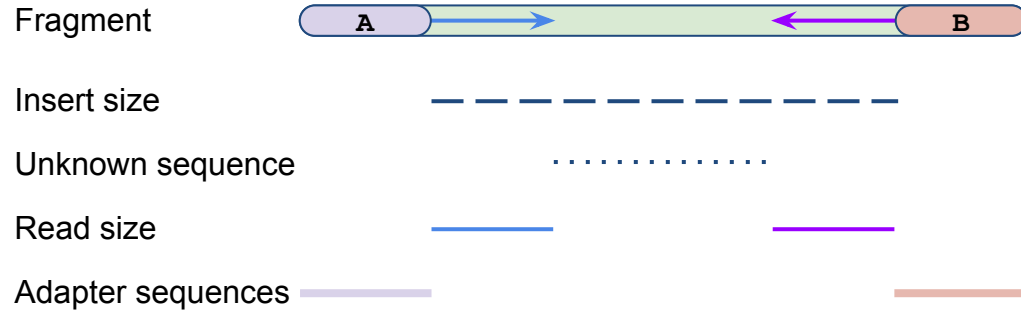
Objectives

- Understand the relation between read sizes and library preparation
- Overview of read count in relation to sequencing depth and error assessment
- Insights into overall read quality on the basis of base quality scoring
- Introduction to assessment of base contents.

A thing about reads...



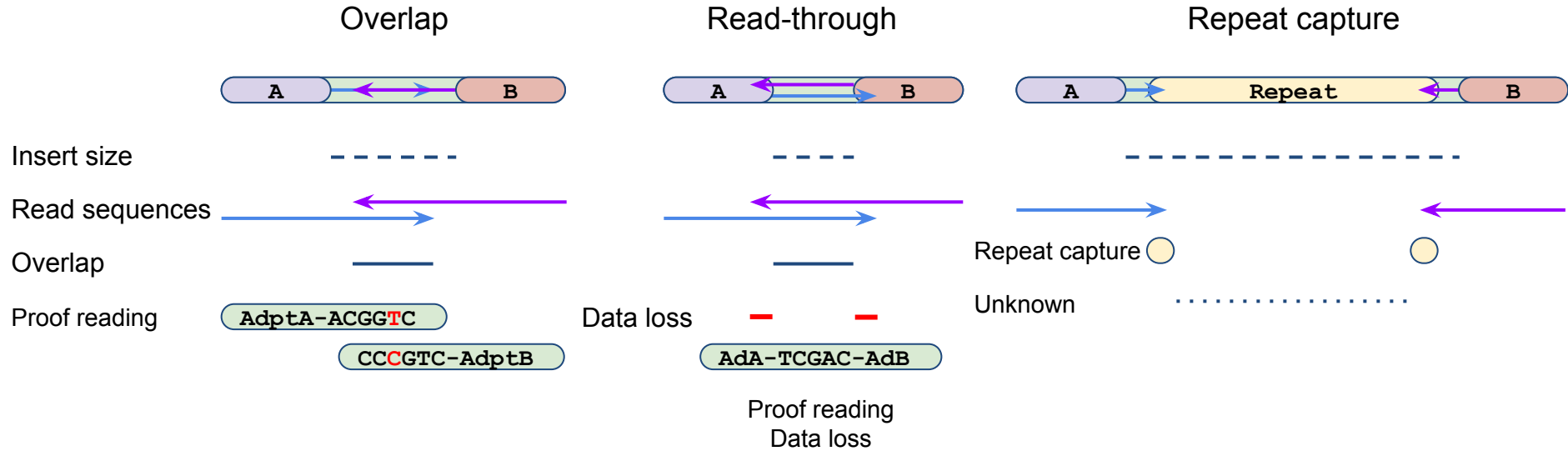
Read types



AdptA-ACGGTCA.....CGTCCGA-AdptB

```
@sample1_Mate1
AdptA-ACGGTCA
...
@sample1_Mate2
AdptB-AGCCTGC
...
```

Scenarios



Illumina read prep

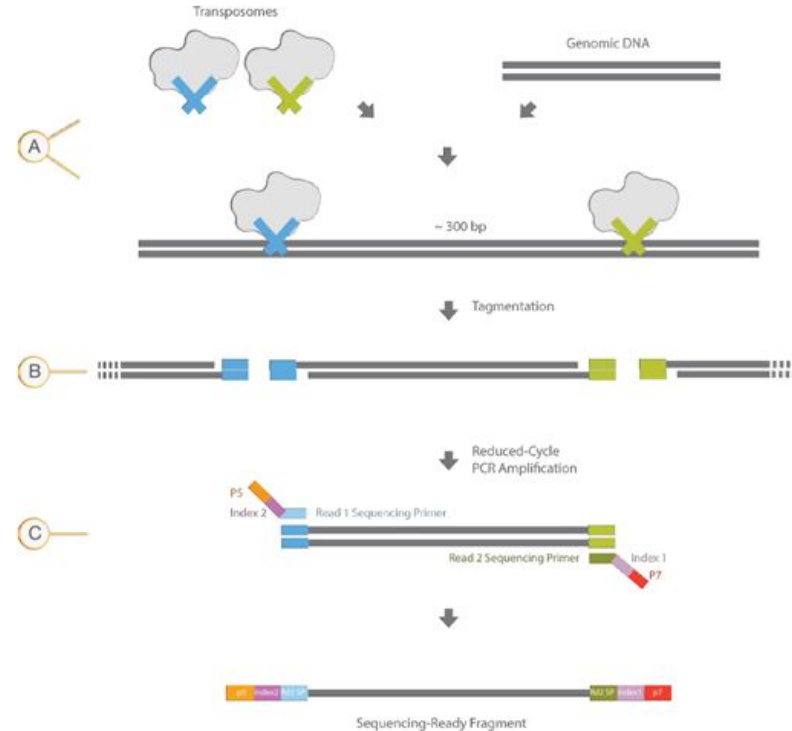
Sample prep

- gDNA extraction
- Pre normalization

Library preparation

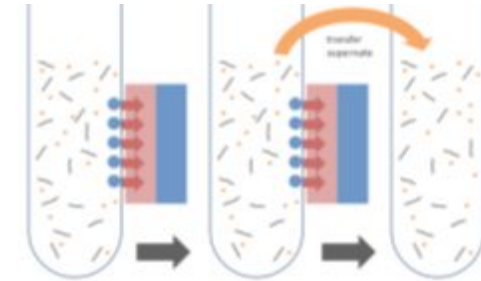
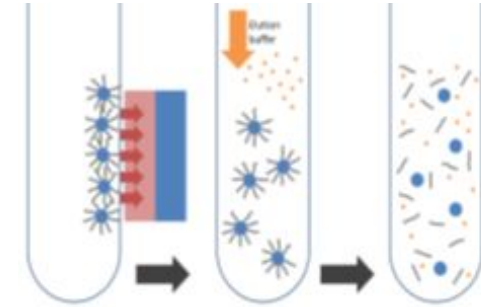
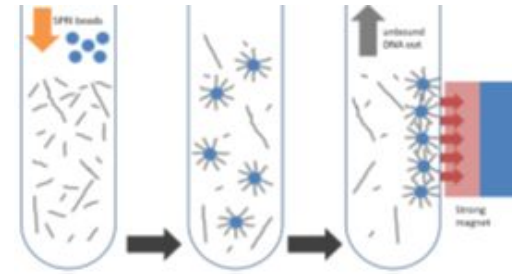
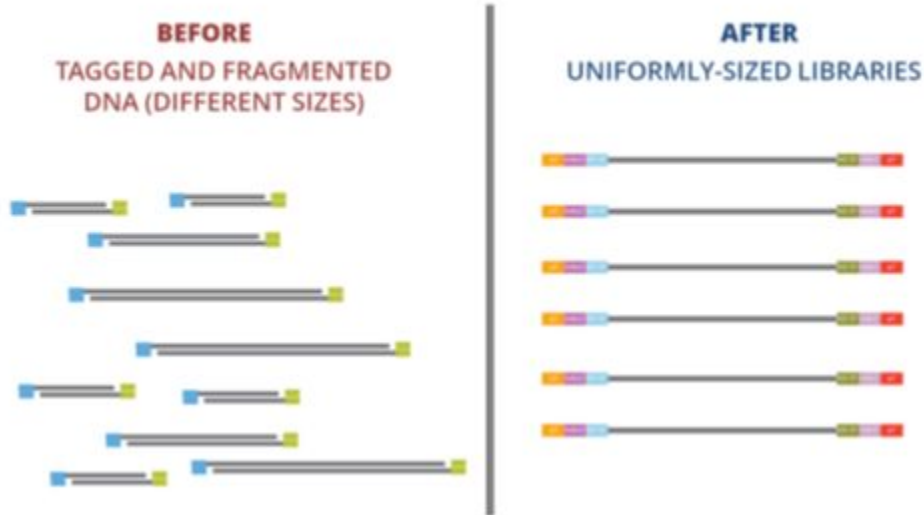
- Tagmentation
- Index PCR
- Normalization and pool

Sequencing

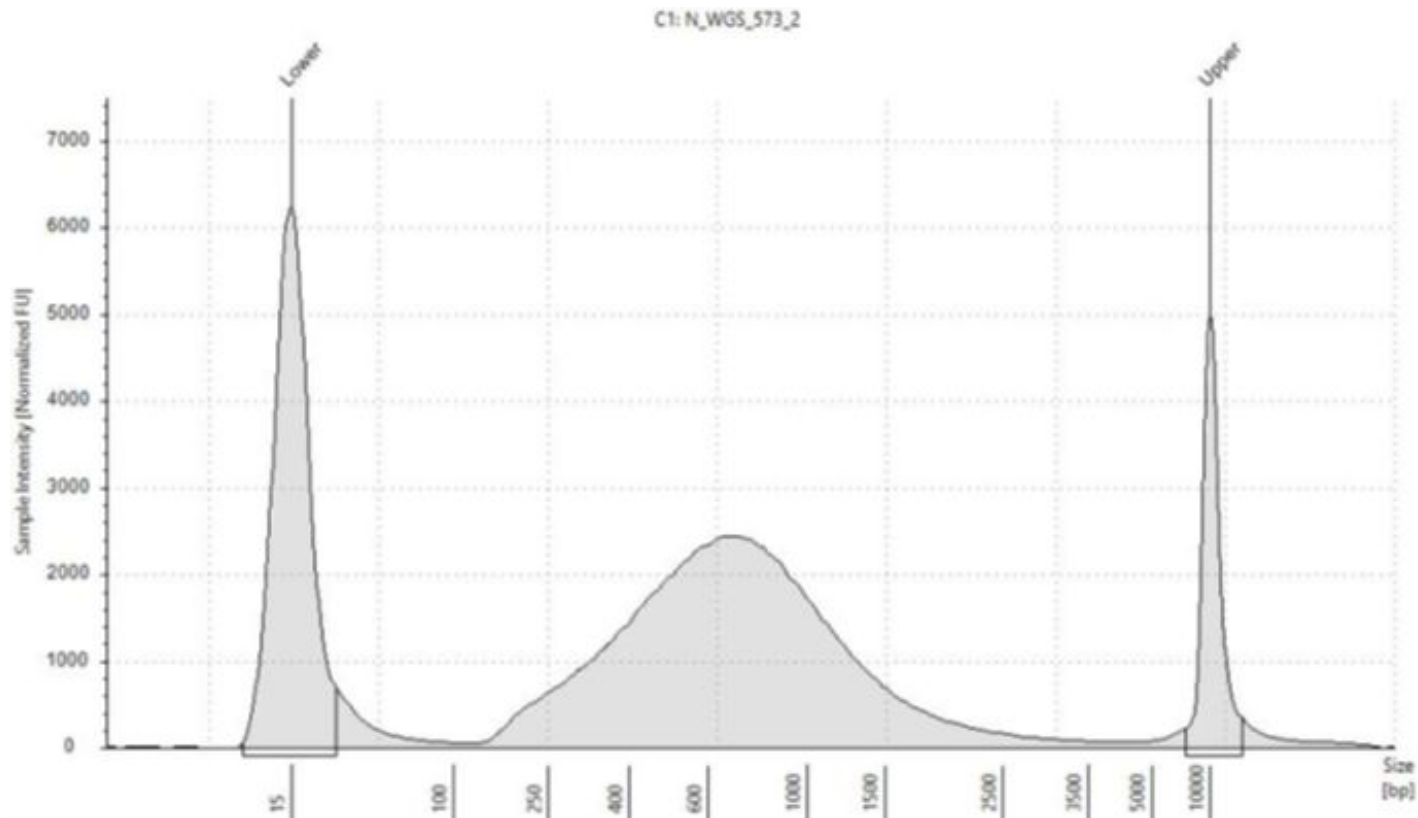


Read size selection

INDEX PCR AND CLEANUP

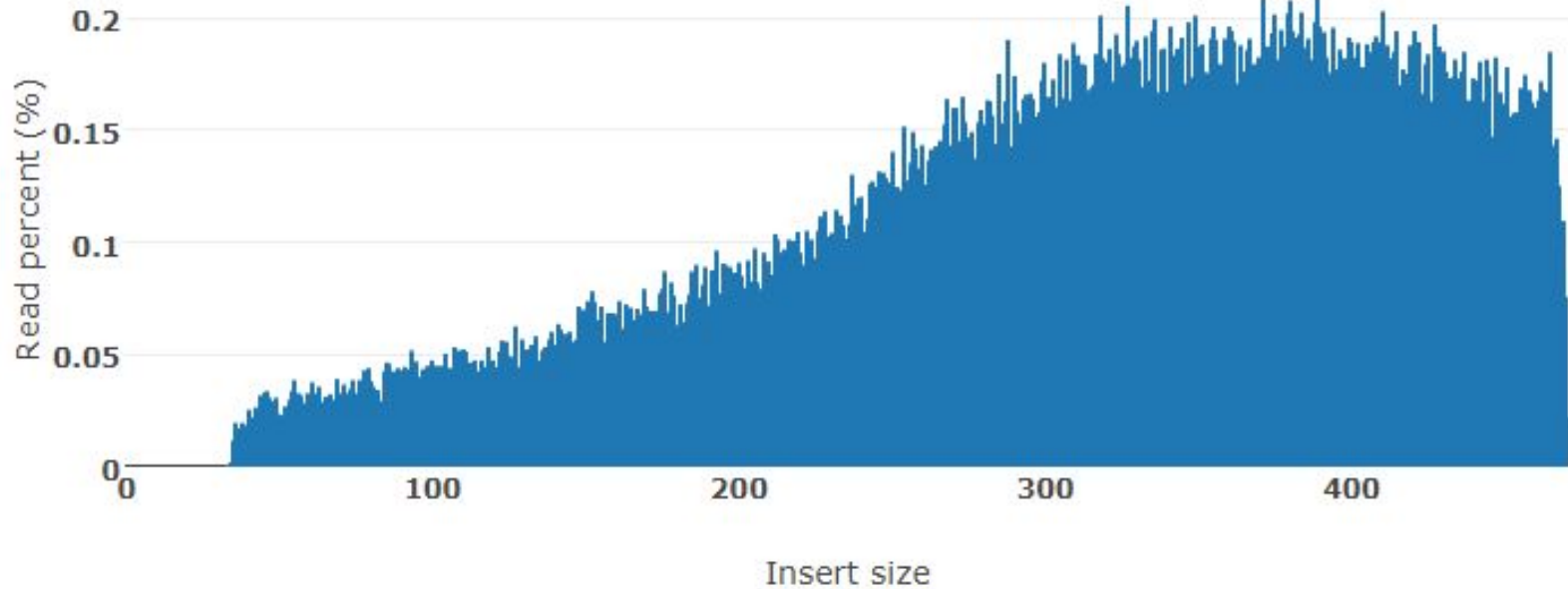


Read sizes



Pop quiz!

What library size do you think is aimed for?

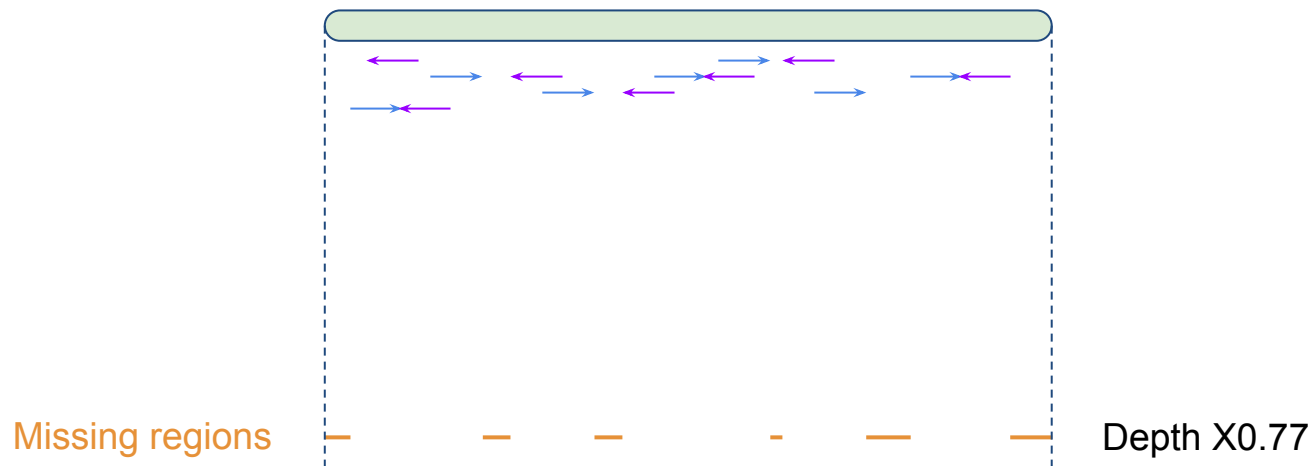


**Read size must
correlate with the
library strategy**

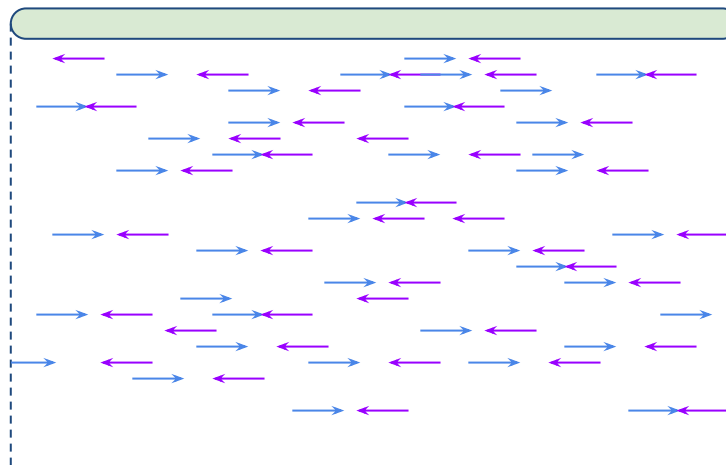
Read **size** must
correlate with the
library **strategy**

Read count

Read count

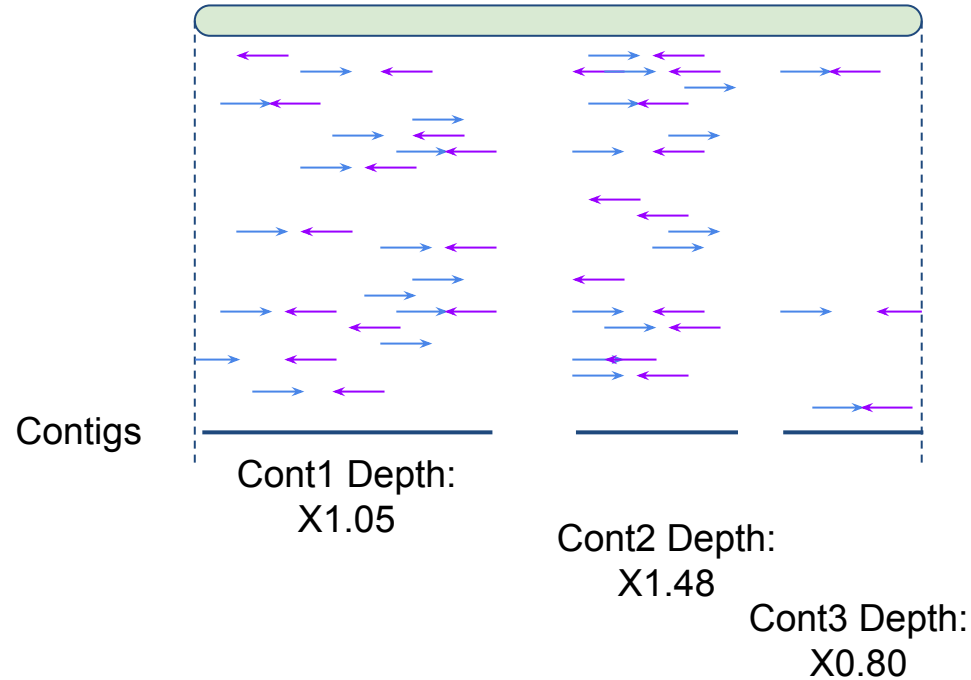


Read count



Depth: X2.8

Read count

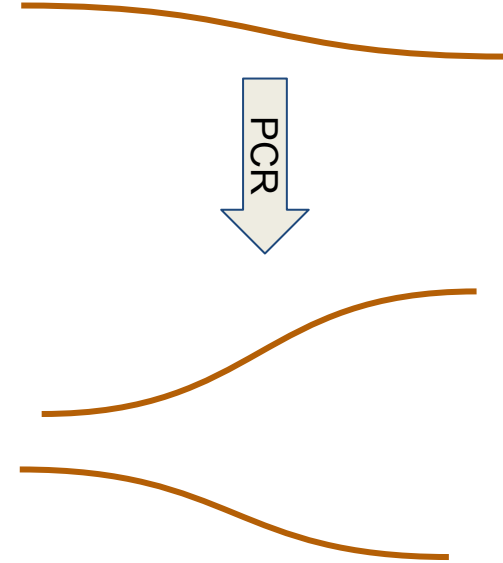
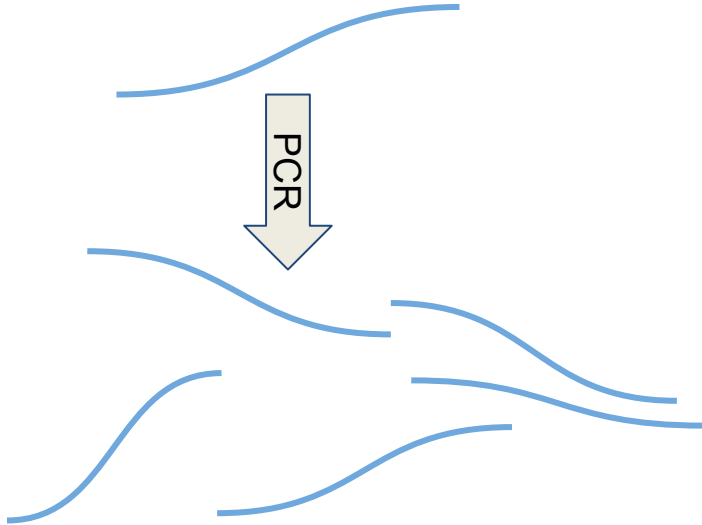


Read counts dictates the sequencing depth

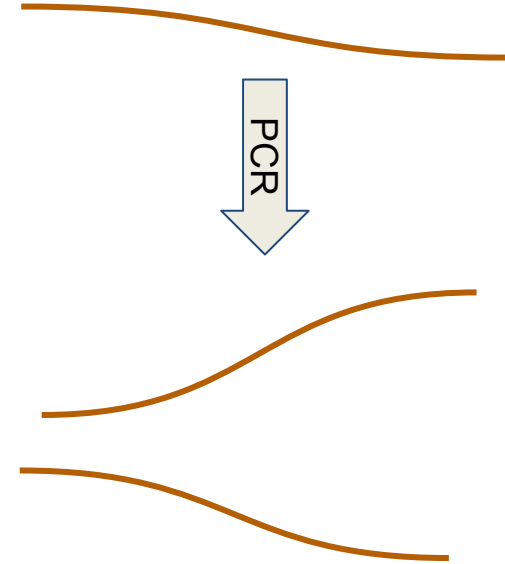
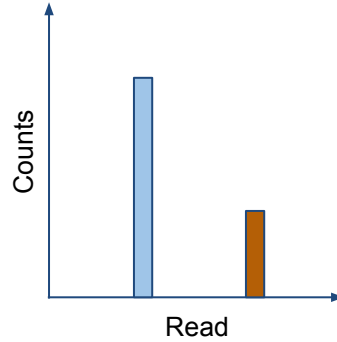
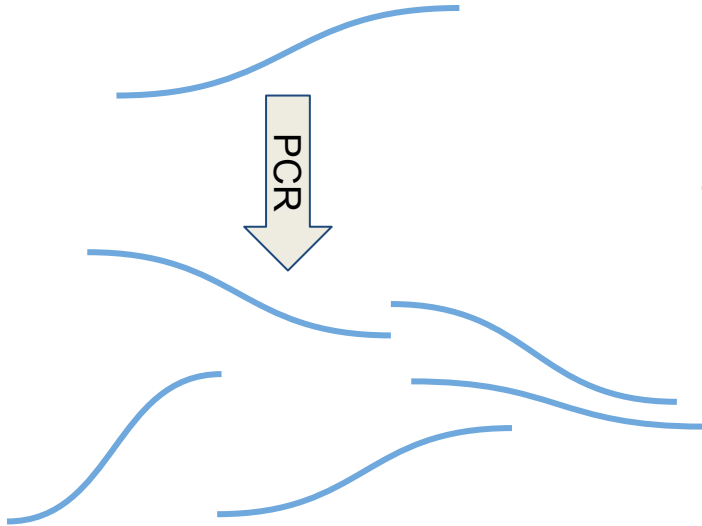
Read **counts** dictates
the sequencing **depth**

Overrepresentation

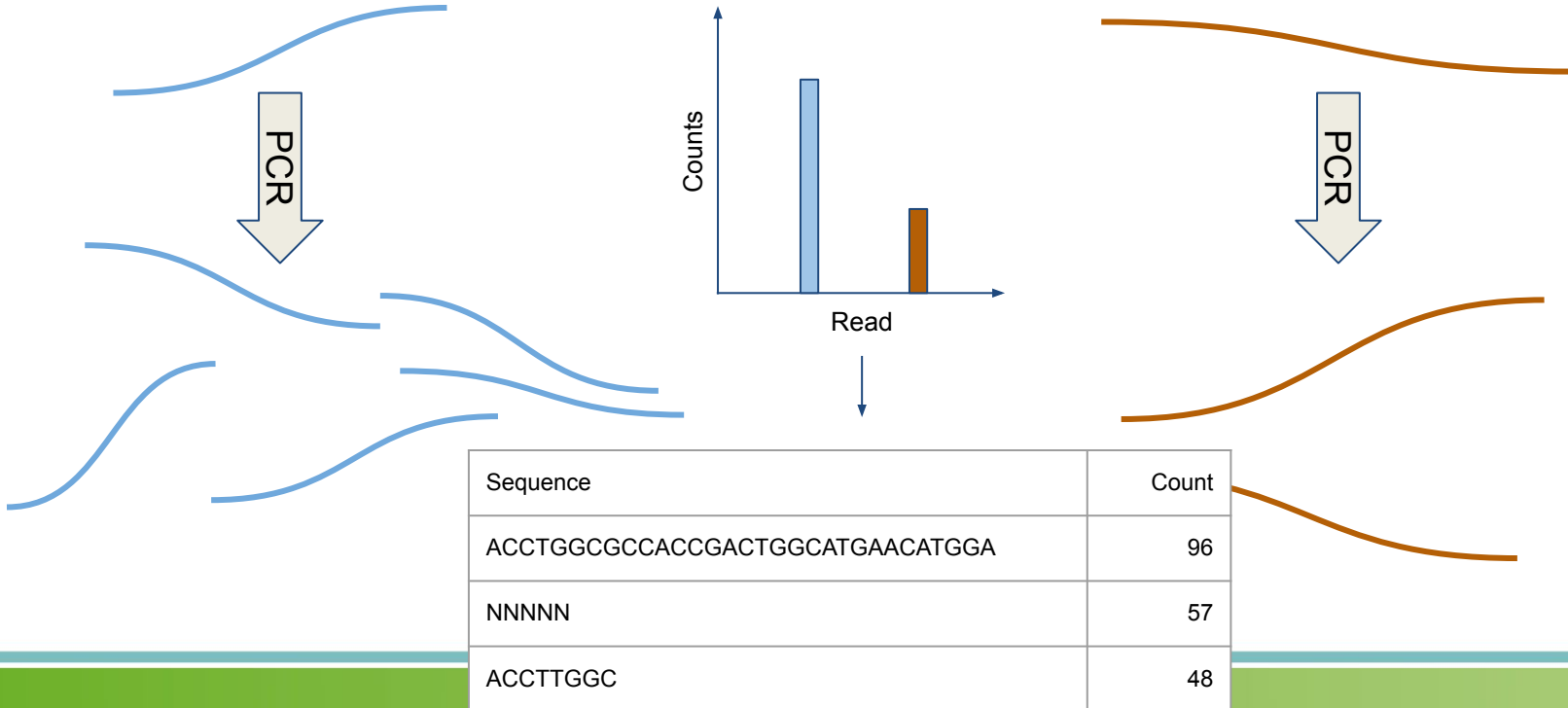
Overrepresentation



Overrepresentation



Overrepresentation



Pop quiz time, yet again!!!

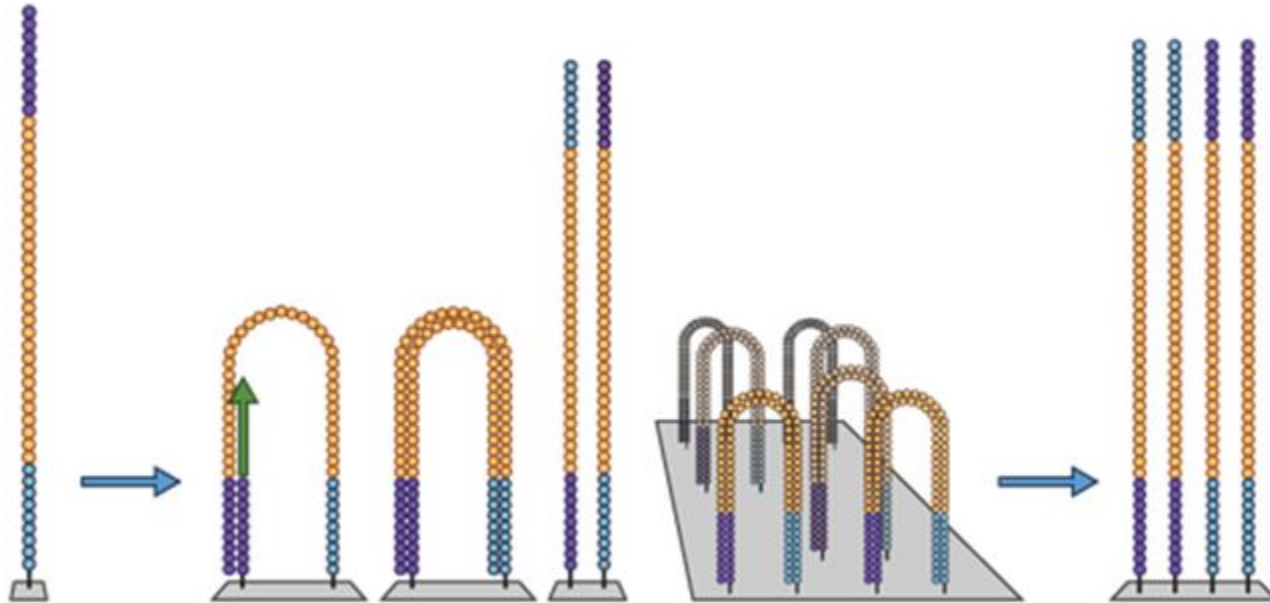
Sequence	Count
ACCTGGCGCCACCGACTGGCATGAACATGGA	96
NNNNN	57
ACCTTGGC	48

- Adapter?
- Biological?
- Technical?

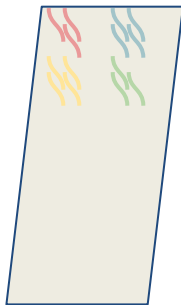
**Overrepresented
reads may indicate
errors in library
preparations or
sequencing.**

**Overrepresented
reads may indicate
errors in library
preparations or
sequencing.**

Cluster formation



Base quality



High quality

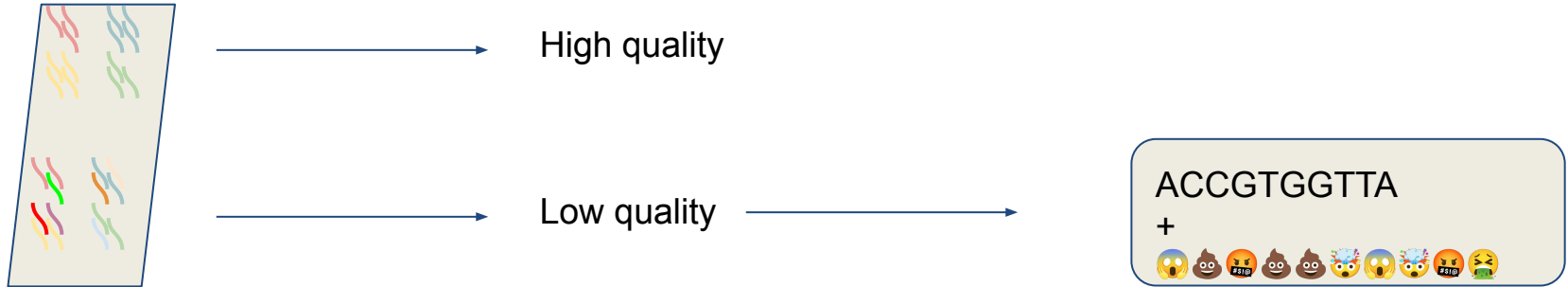


ACCGTGGTTA

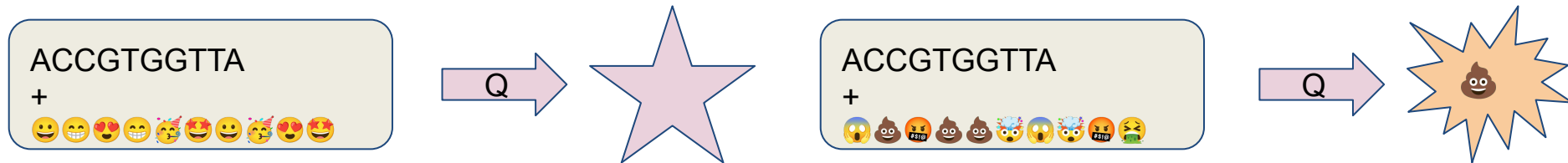
+



Base quality



Overall base quality



EZ quiz

What to do with the crappy bases and reads?

What to do with the crappy bases and reads?

Throw it out!!

**Base- and read -quality
can be applied to filter out
low quality sequencing
data.**

**Base- and read -quality
can be applied to filter out
low quality sequencing
data.**

Base contents

>NC_007795.1 Staphylococcus aureus subsp. aureus NCTC 8325 chromosome, complete genome
CGATTAAAGATAGAAATACACGATGCGAGCAATCAAATTTCATAACATCACCATGAGTTTGGTCCGAAGC
ATGAGTGTTTACAATGTTCGAACACCTTATACAGTTCTTATACATACTTTATAAATTATTTCCCAA

...

Base contents

>NC_007795.1 *Staphylococcus aureus* subsp. *aureus* NCTC 8325 chromosome, complete genome
CGATTAAAGATAGAAATACACGATGCGAGCAATCAAATTTCATAACATCACCATGAGTTTGGTCCGAAGC
ATGAGTGTTTACAATGTTCGAACACCTTATACAGTTCTTATACATACTTTATAAATTATTTCCCAA

...

Base contents

>NC_007795.1 Staphylococcus aureus subsp. aureus NCTC 8325 chromosome, complete genome
CGATTAAAGATAGAAATACACGATGCGAGCAATCAAATTTATAACATCACCATGAGTTTGGTCCGAAGC
ATGAGTGTTTACAATGTTCGAACACCTTATACAGTTCTTATACATACTTTATAAATTATTTCCCAA

...

Base	Count	Percentage
A	49	36.03
C	26	19.12
G	19	13.97
T	42	30.88

GC contents

S. aureus: ~33.09%

E. coli: X%

... Y%

...

Base contents can indicate species and might aid as an early indicator of issues.

Base contents can
indicate species and
might aid as an early
indicator of **issues**.

Objectives

- Understand the relation between read sizes and library preparation
- Overview of read count in relation to sequencing depth and error assessment
- Insights into overall read quality on the basis of base quality scoring
- Introduction to assessment of base contents.

BONUS contents!!!

How do we handle all these params?

- Low quality reads
- Adaptor content
- Reads that are too small
- Remove poly-A-reads (library dependent)

Filtration

Trimmomatic

BBduk

fastp

...

Filtration

Trimmomatic

BBduk

fastp

...

You survived!

Congratz...

Acknowledgements

The creation of this training material was commissioned by ECDC to Statens Serum Institut with the direct involvement of Kasper Thystrup Karstensen
The revision and update of this training material was commissioned by ECDC to Statens Serum Institut with the direct involvement of Kasper Thystrup Karstensen, Astrid Rasmussen, and Søren Hallstrøm