Proficiency test number 28

Whole Genome Sequencing of *Campylobacter*

Bo Segerman

Ásgeir Ástvaldsson

EURL-*Campylobacter* Workshop 2021

PT28 Objectives

- To assess the performance of DNA extraction and whole genome sequencing (WGS) of *Campylobacter*

- Quantify differences between whole genome sequence (WGS) data from *Campylobacter*, produced at different laboratories

- Participants received 2 lyophilized strains of *Campylobacter* and two gDNA samples from the same strains

- Participants extracted gDNA from the lyophilized strains and sequenced all 4 samples own methods

- Reporting was done through a Questback questionnaire and raw sequencing data was uploaded to the EURL

    11 NRLs met the deadline for submission

    5 NRLs failed the deadline for submission

- Results from all participants compared in terms of different QC for raw sequence data and assembly metric

EURL-*CAMPYLOBACTER*

REPORT

PROFICIENCY TEST NUMBER 28

Whole genome sequencing of *Campylobacter*

PT28 final report published in March 2021

Accessible through the EURL-*Campylobacter* website:

https://www.sva.se/en/about-us/eurl-campylobacter/proficiency-tests/

Individual reports prepared with QC metrics and comments on performance

Sent out December 2020 and September 2021

Available PT28 data

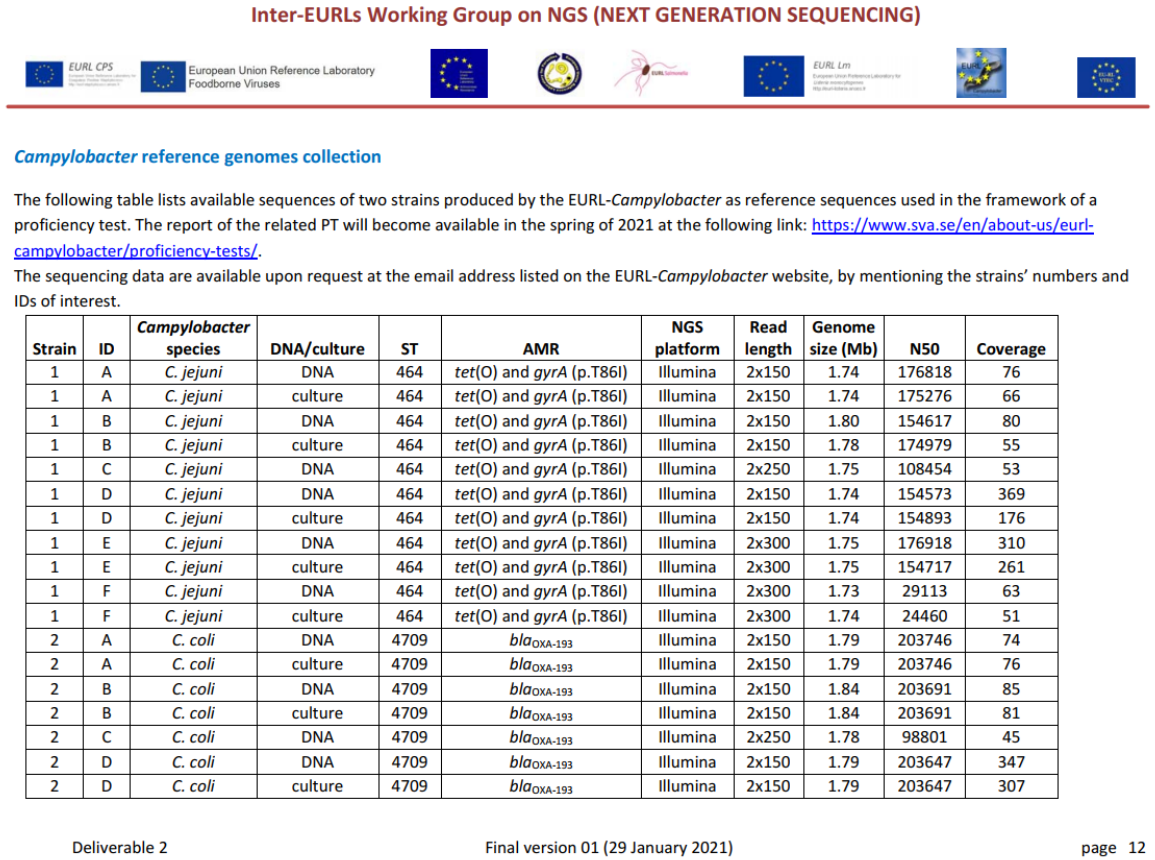
Sequencing datasets from six participants available upon request

C. *jejuni*: 11 datasets

C. *coli*: 11 datasets

https://www.sva.se/en/about-us/eurl-campylobacter/laboratory-procedures/inter-eurls-working-group-on-ngs/

> Reference Whole Genome Sequencing collection – curated by EURL-*Salmonella*



Inter-EURLs Working Group on NGS (NEXT GENERATION SEQUENCING)

EURL CPS | European Union Reference Laboratory Foodborne Viruses | EURL Lm

***Campylobacter* reference genomes collection**

The following table lists available sequences of two strains produced by the EURL-*Campylobacter* as reference sequences used in the framework of a proficiency test. The report of the related PT will become available in the spring of 2021 at the following link: https://www.sva.se/en/about-us/eurl-campylobacter/proficiency-tests/.

The sequencing data are available upon request at the email address listed on the EURL-*Campylobacter* website, by mentioning the strains' numbers and IDs of interest.

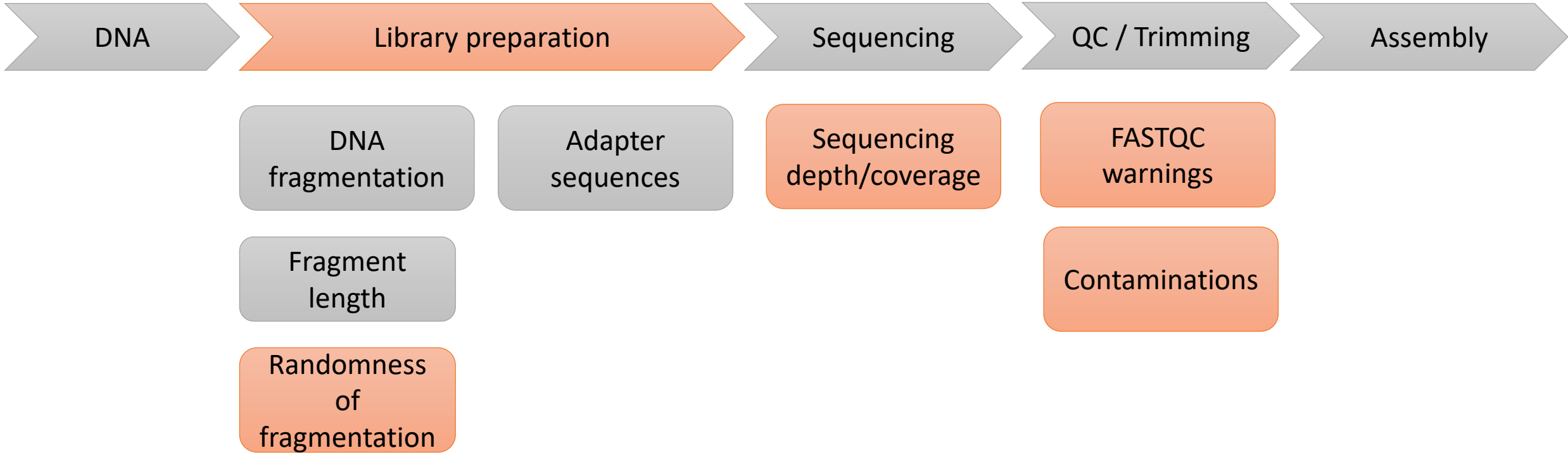
Strain	ID	*Campylobacter* species	DNA/culture	ST	AMR	NGS platform	Read length	Genome size (Mb)	N50	Coverage
1	A	*C. jejuni*	DNA	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x150	1.74	176818	76
1	A	*C. jejuni*	culture	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x150	1.74	175276	66
1	B	*C. jejuni*	DNA	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x150	1.80	154617	80
1	B	*C. jejuni*	culture	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x150	1.78	174979	55
1	C	*C. jejuni*	DNA	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x250	1.75	108454	53
1	D	*C. jejuni*	DNA	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x150	1.74	154573	369
1	D	*C. jejuni*	culture	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x150	1.74	154893	176
1	E	*C. jejuni*	DNA	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x300	1.75	176918	310
1	E	*C. jejuni*	culture	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x300	1.75	154717	261
1	F	*C. jejuni*	DNA	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x300	1.73	29113	63
1	F	*C. jejuni*	culture	464	*tet*(O) and *gyrA* (p.T86I)	Illumina	2x300	1.74	24460	51
2	A	*C. coli*	DNA	4709	$bla_{OXA-193}$	Illumina	2x150	1.79	203746	74
2	A	*C. coli*	culture	4709	$bla_{OXA-193}$	Illumina	2x150	1.79	203746	76
2	B	*C. coli*	DNA	4709	$bla_{OXA-193}$	Illumina	2x150	1.84	203691	85
2	B	*C. coli*	culture	4709	$bla_{OXA-193}$	Illumina	2x150	1.84	203691	81
2	C	*C. coli*	DNA	4709	$bla_{OXA-193}$	Illumina	2x250	1.78	98801	45
2	D	*C. coli*	DNA	4709	$bla_{OXA-193}$	Illumina	2x150	1.79	203647	347
2	D	*C. coli*	culture	4709	$bla_{OXA-193}$	Illumina	2x150	1.79	203647	307

Reference genomes

- Complete, gap-free genomes generated for both PT28 strains
- WGS using both Illumina MiSeq and Oxford Nanopore Flongle
- Hybrid assemblies generated using Trycycler and Unicycler assemblers
- Assemblies and raw datasets available at DDBJ/ENA/NCBI(GenBank) websites

	Study ID	Sample ID	Assembly Acc. No.
PT28-1 *C. jejuni*	PRJEB45600	SAMEA8911315	GCA_912579705
PT28-2 *C. coli*		SAMEA8911316	GCA_912579715

Reflections from the PT28 WGS dataset

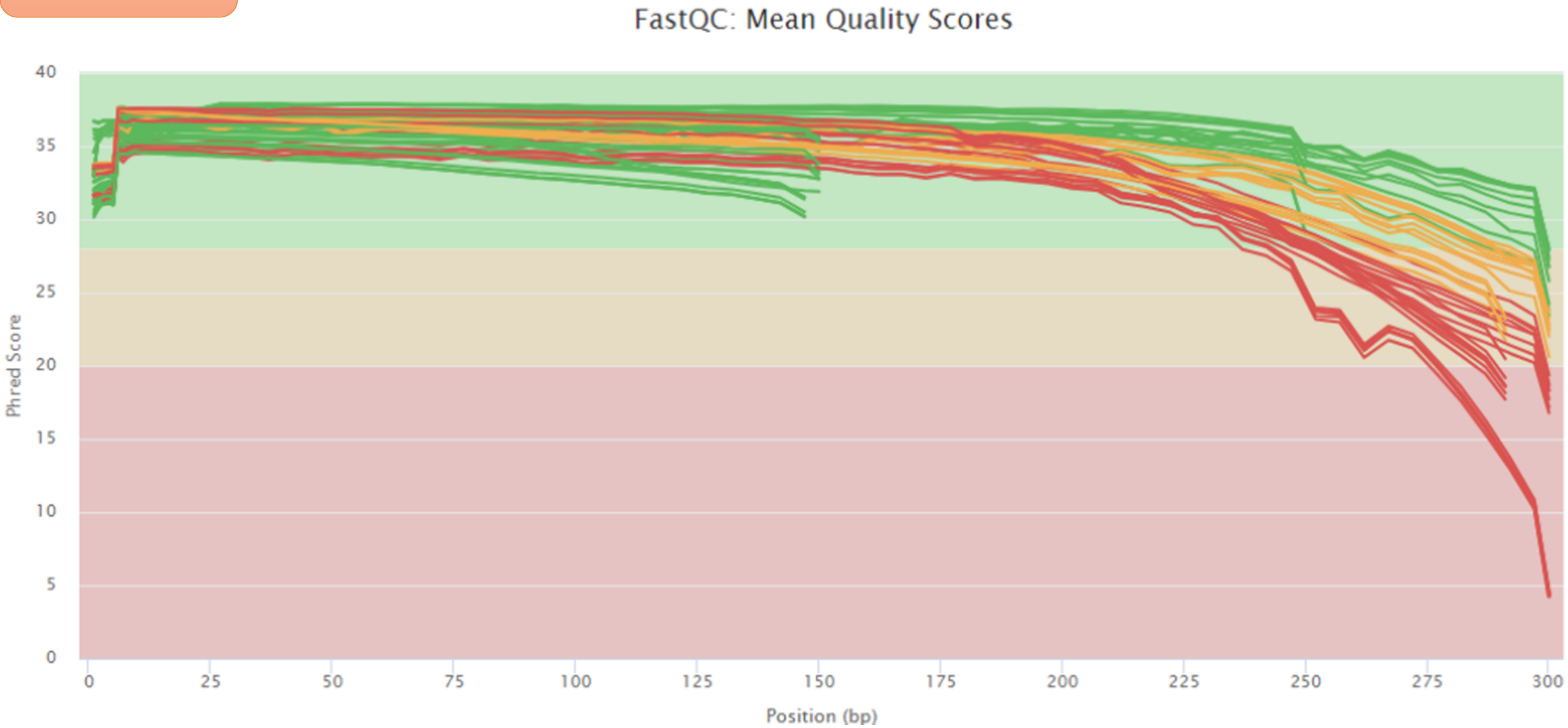


FASTQC warnings

Quality control step (FASTQC status check)

Trimming related

FASTQC warnings



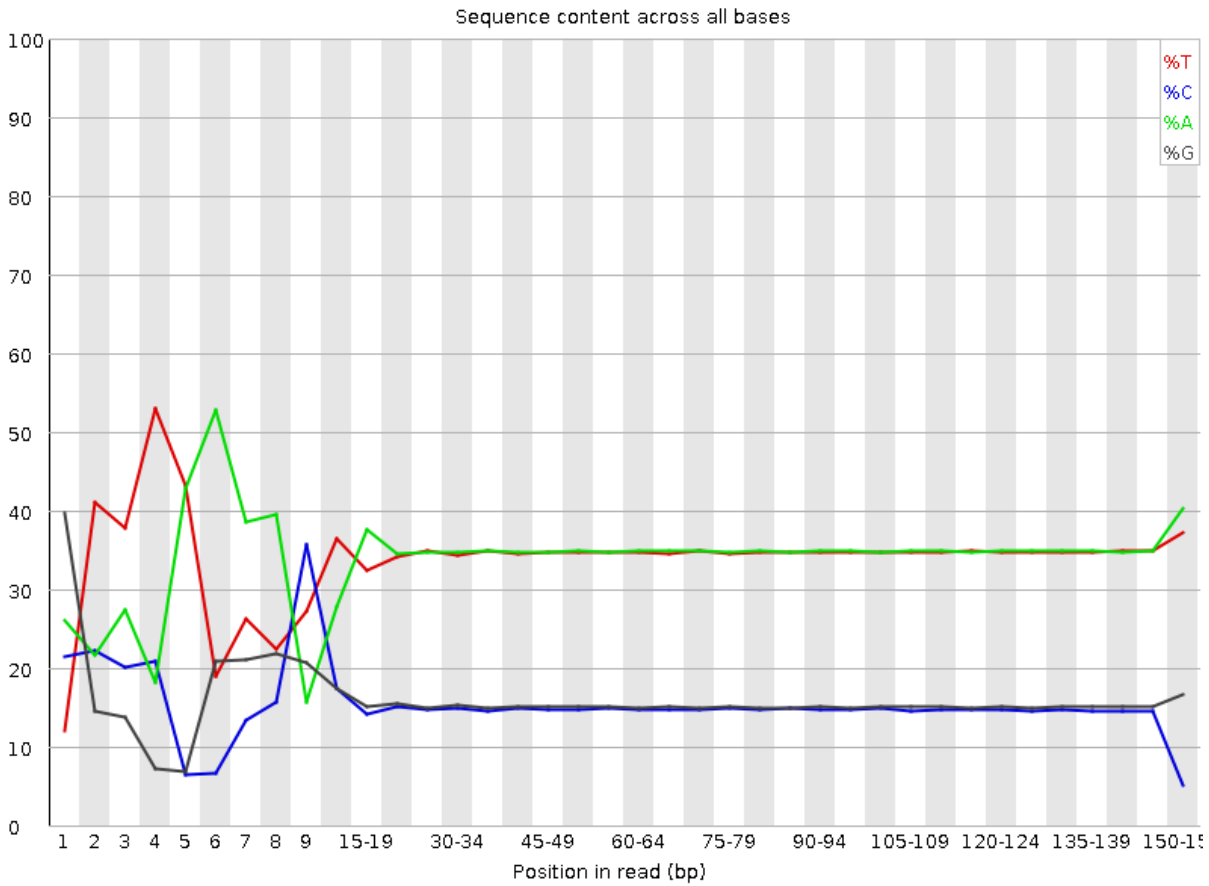
FASTQC warnings

Quality control step (FASTQC status check)

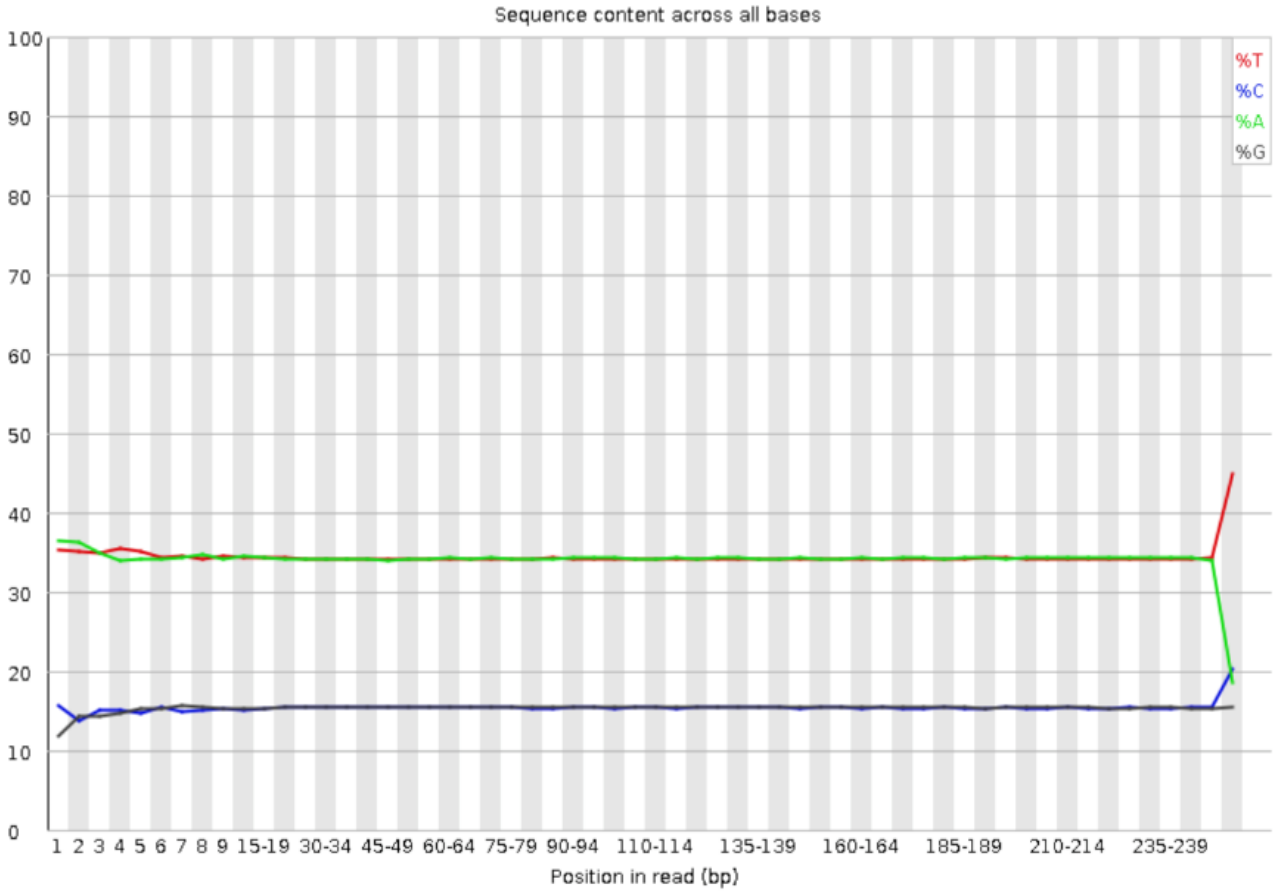
Library prep Nextera related

Randomness of fragmentation

Library kit by: **Nextera**

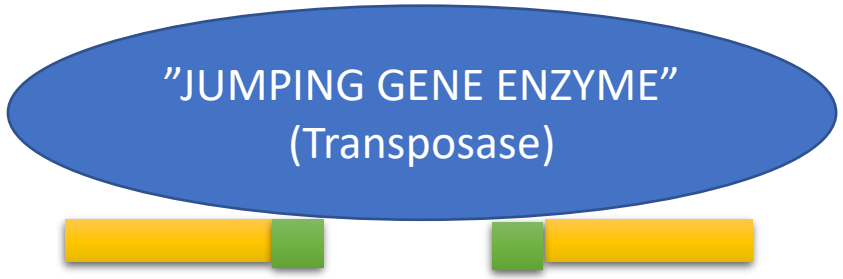


TrueSeq



Randomness of fragmentation

DNA Fragmentation / Tagmentation Nextera



Adapter Adapter

Not entirely random
Preference for certain sequence motifs

”Tagged with adapters + fragmentation = ”Tagmentation”

Adapter Adapter Adapter Adapter

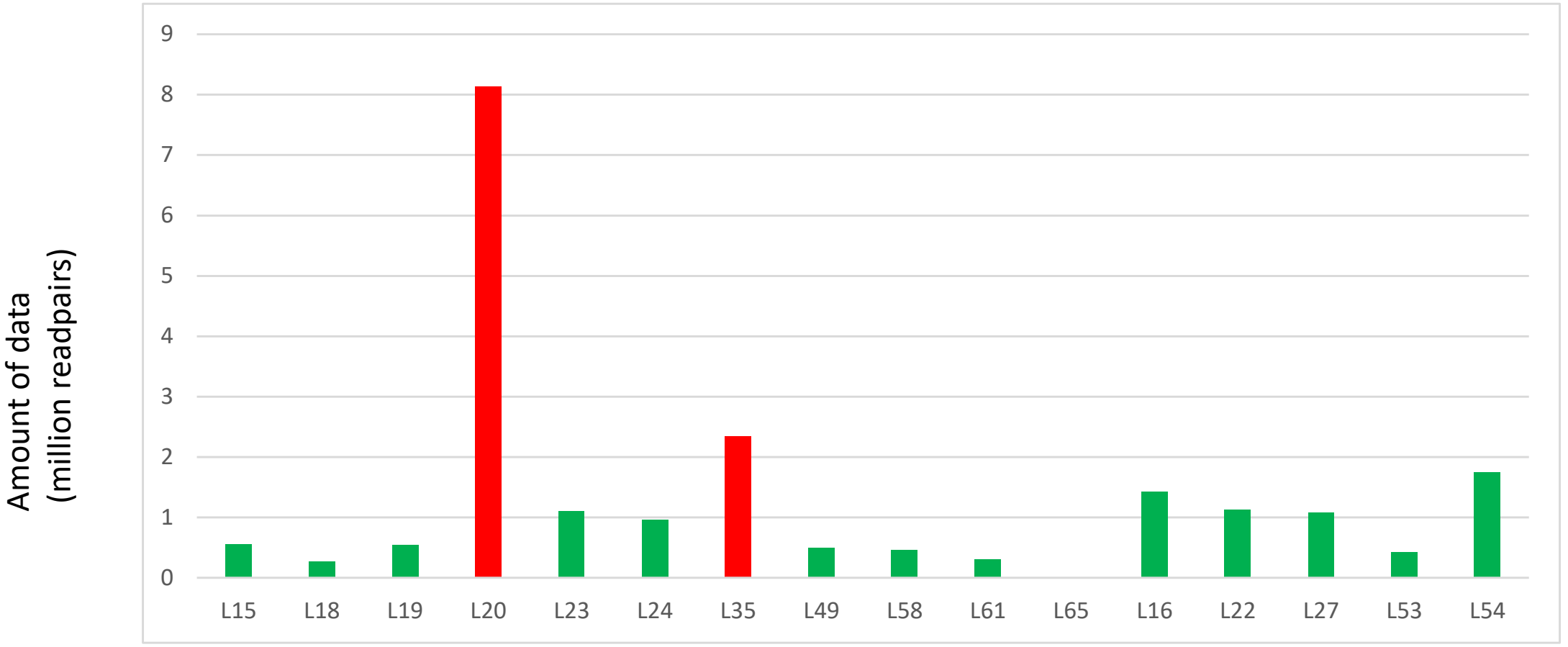
FASTQC warnings

Quality control step (FASTQC status check)

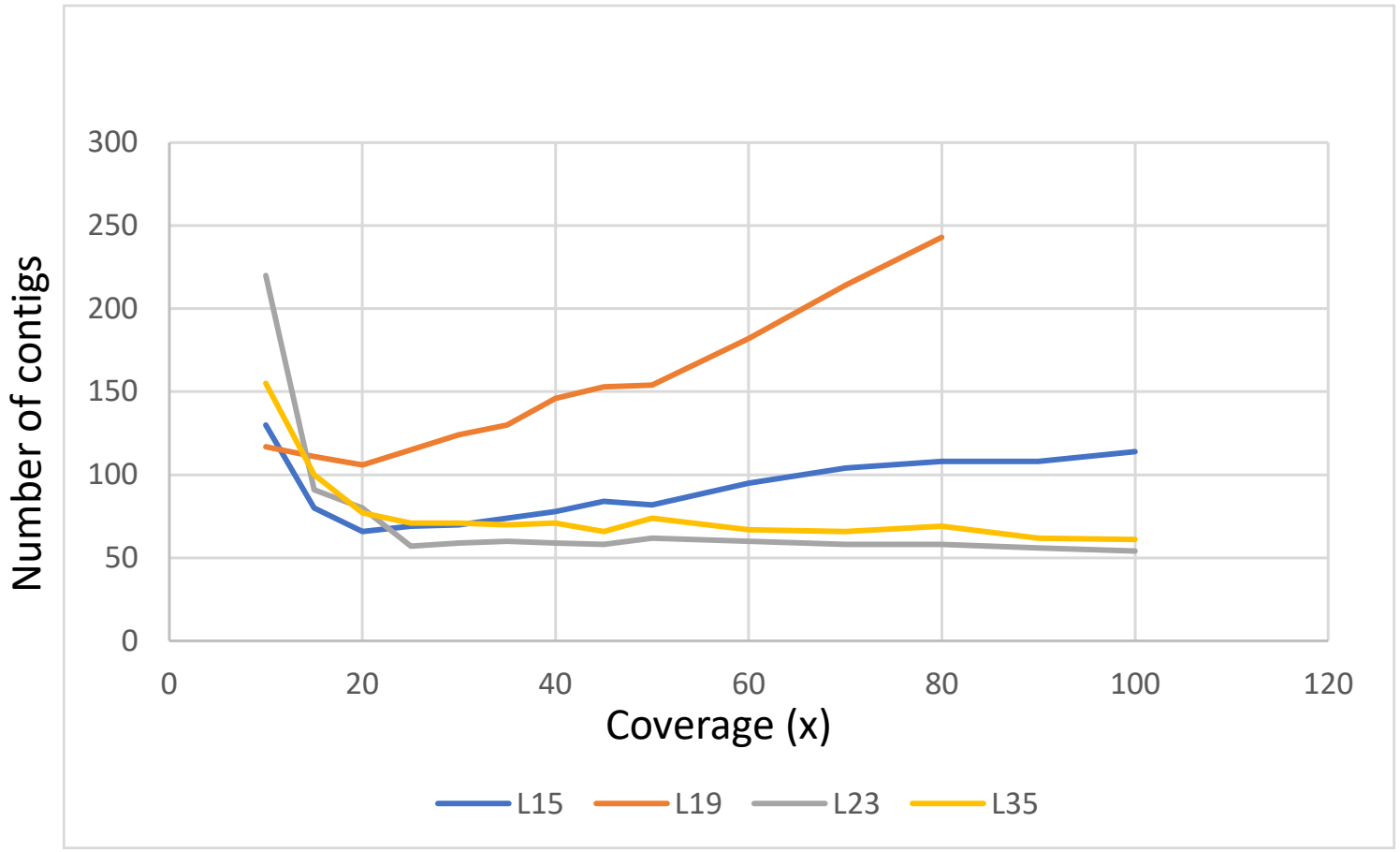
Duplicated reads

FASTQC warnings

Duplication warnings/error flags
Large amount of data in combination with selective (Nextera) fragmentation
Usually not a problem

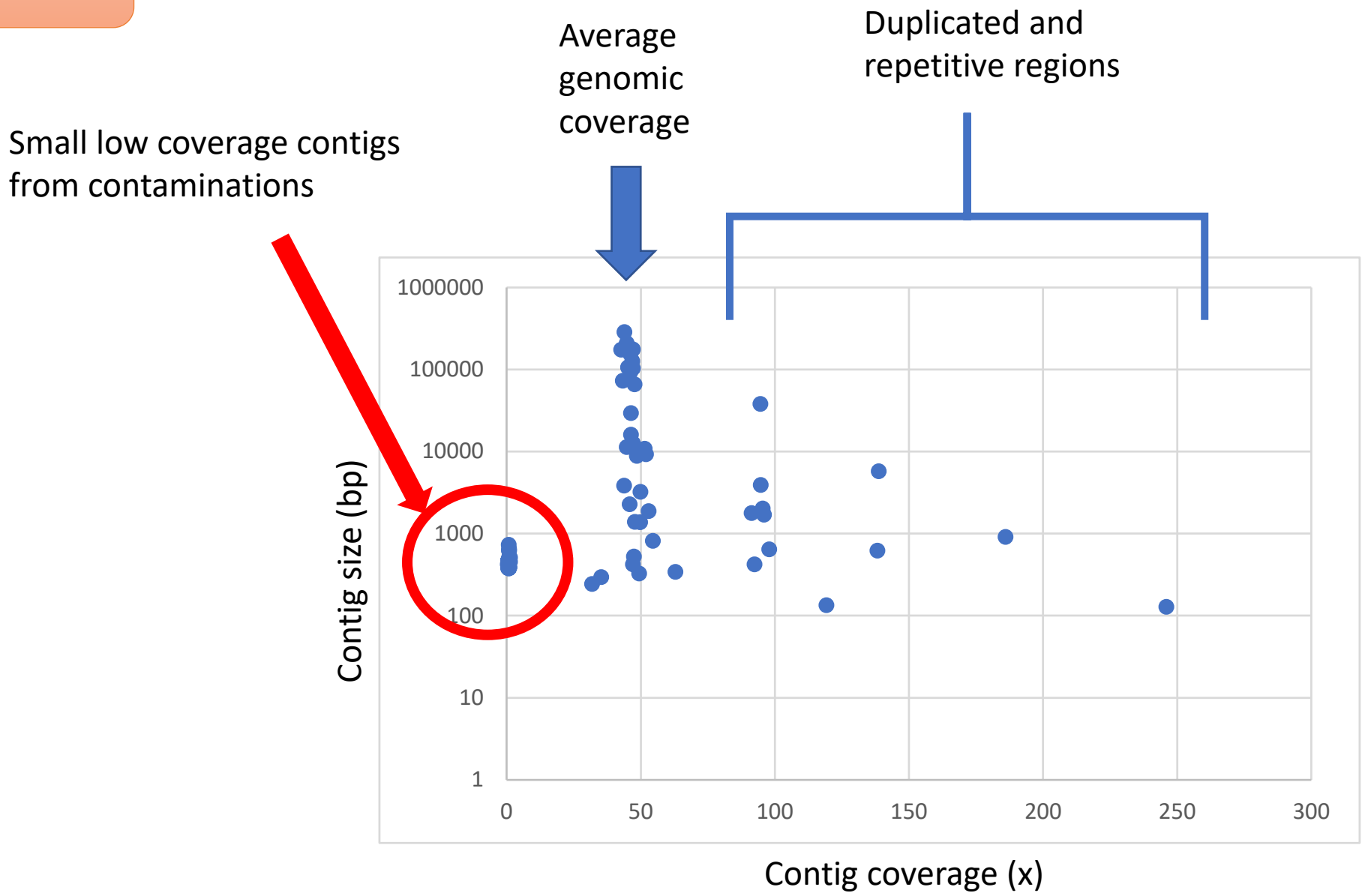


Contaminations



Small levels of contaminations in raw data leads to an increasing number of small contamination-related contigs if the amount of data (coverage) is increased when doing assembly

Contaminations



Library preparation

Randomness of fragmentation

Nextera XT has higher bias in the fragmentation step
Thus requires higher coverage for acceptable assembly