

Comparison of burn-in controls

Lambda phage (48.5kb)

	<i>Nb reads</i>	<i>Nb 2D reads</i>	<i>Mean (2D-consensus) Reads Length (bp)</i>	<i>Identical bases per 100 aligned bases (including indels)</i>
Lambda Juis (2014)	20818	1579	5527	32
Lambda Julien (2015)	5560	3069	6834	19
Lambda Daniel (2016)	4952	3107	9541	19
PacBio 2Kb (2015)	50662	n/a	1374	16

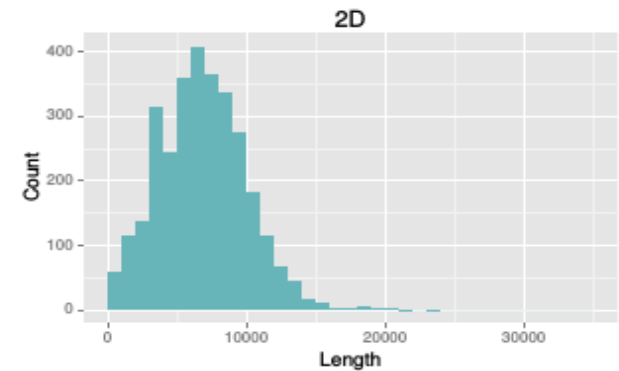
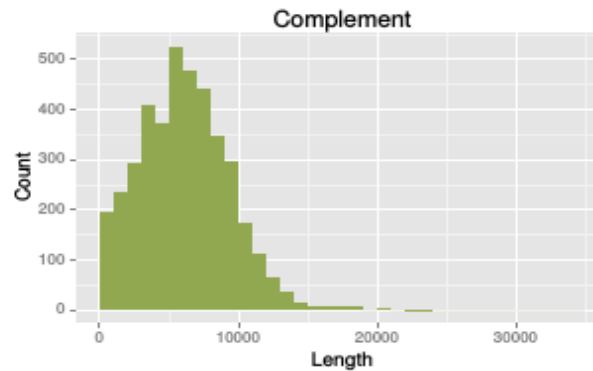
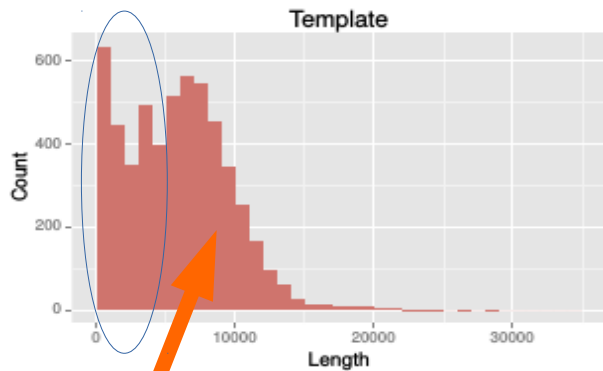
Burn-in June 2015

Lambda phage (48.5kb)

Unexpected longest read

Read lengths

Type	NumReads	TotalBases	Mean	Longest	Shortest	N50
Template	5406	31523281	5831.17	145552	12	7967
Complement	4028	24573708	6100.72	23218	18	7601
2D	3068	20968580	6834.61	23496	192	8127

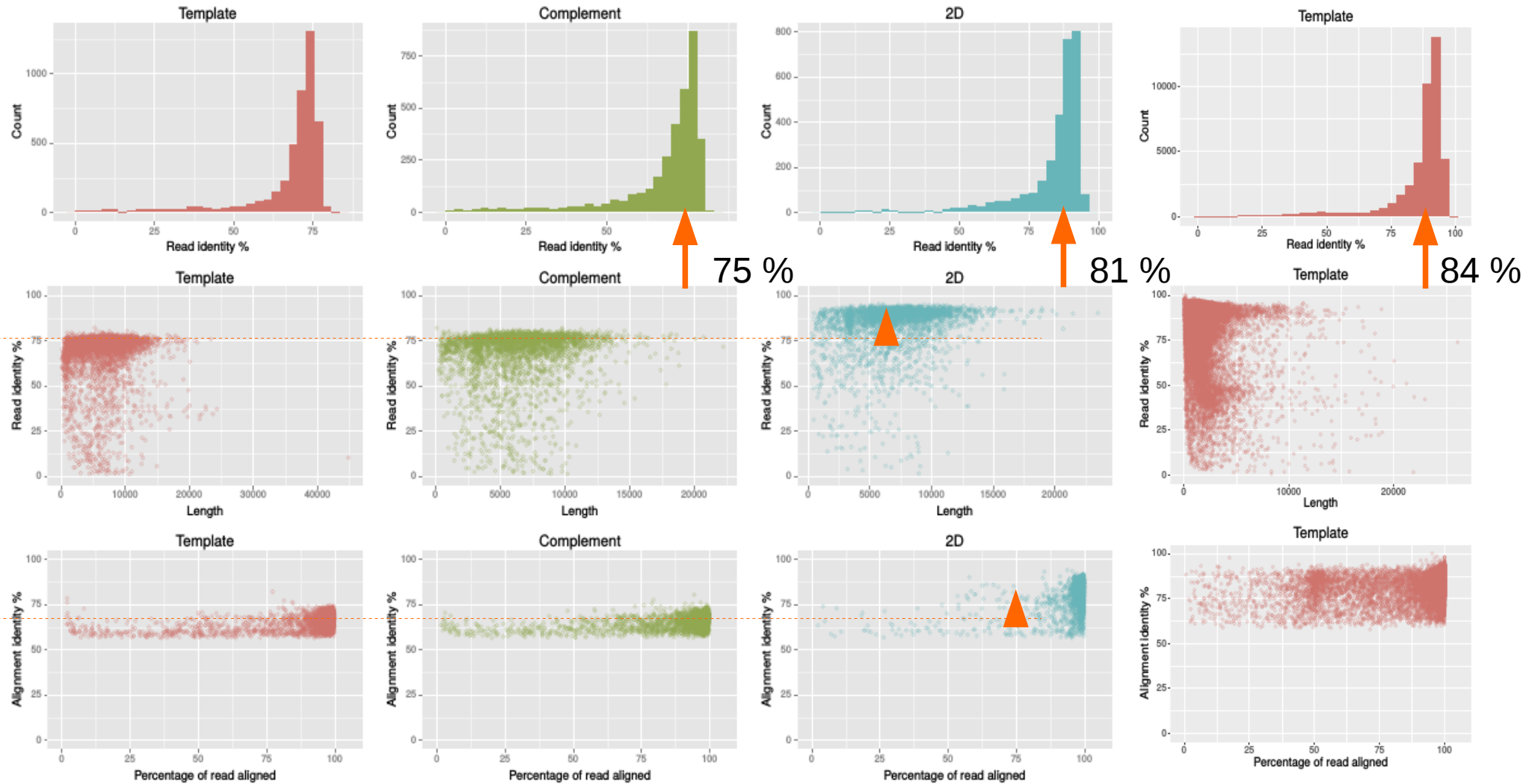


54 % able to go through 2D

27 % are short reads - template only -
(started and crashed?)

Identity versus reference

PacBio



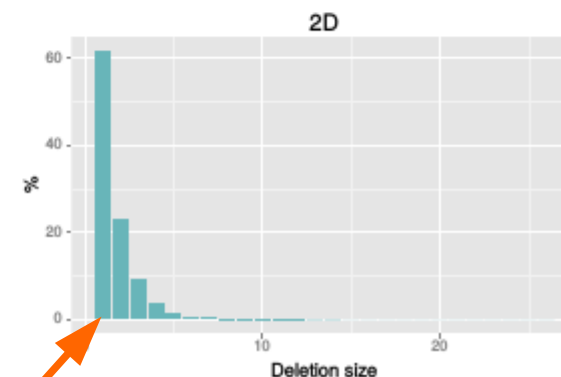
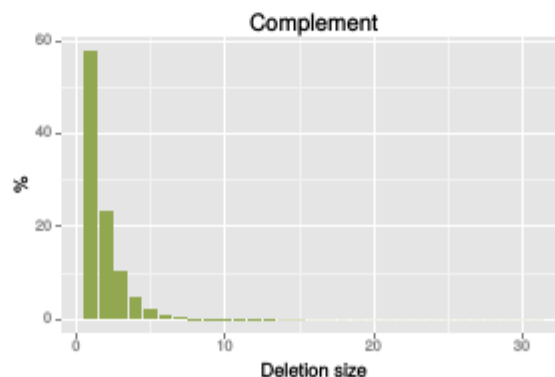
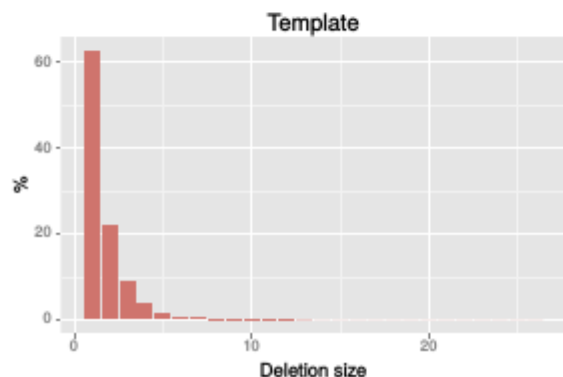
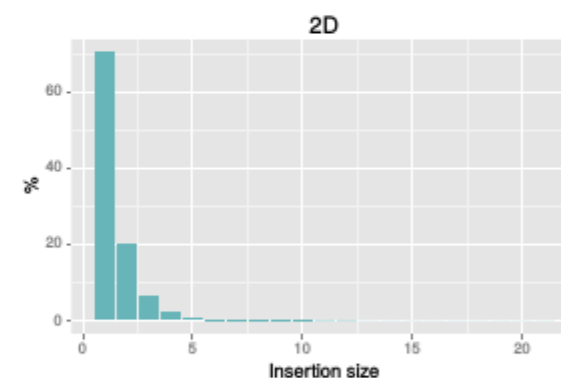
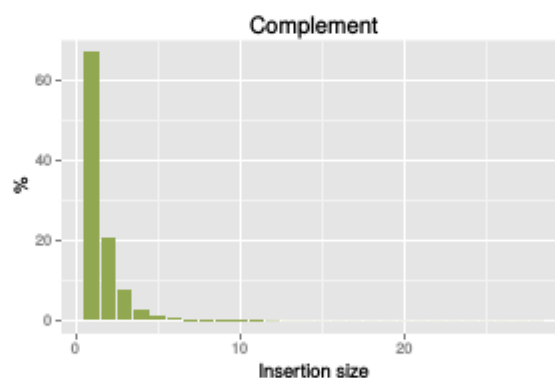
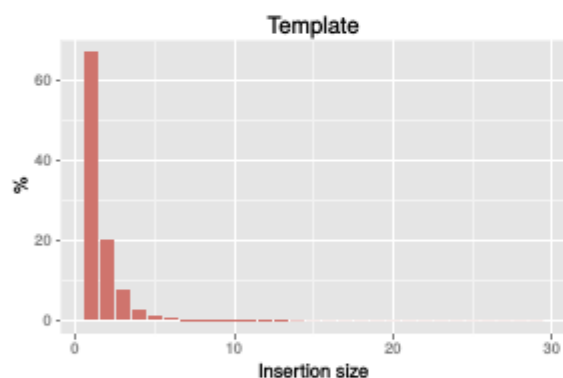
burn-in error analysis

Error type

	Template	Complement	2D
Overall base identity (excluding indels)	67.50%	68.16%	84.55%
Aligned base identity (excluding indels)	79.44%	80.71%	91.39%
Identical bases per 100 aligned bases (including indels)	67.01%	66.58%	81.81%
Inserted bases per 100 aligned bases (including indels)	5.04%	4.49%	2.90%
Deleted bases per 100 aligned bases (including indels)	10.61%	13.02%	7.58%
Substitutions per 100 aligned bases (including indels)	17.34%	15.91%	7.71%
Mean insertion size	1.56	1.54	1.45
Mean deletion size	1.64	1.78	1.65

→ 19 % error rate

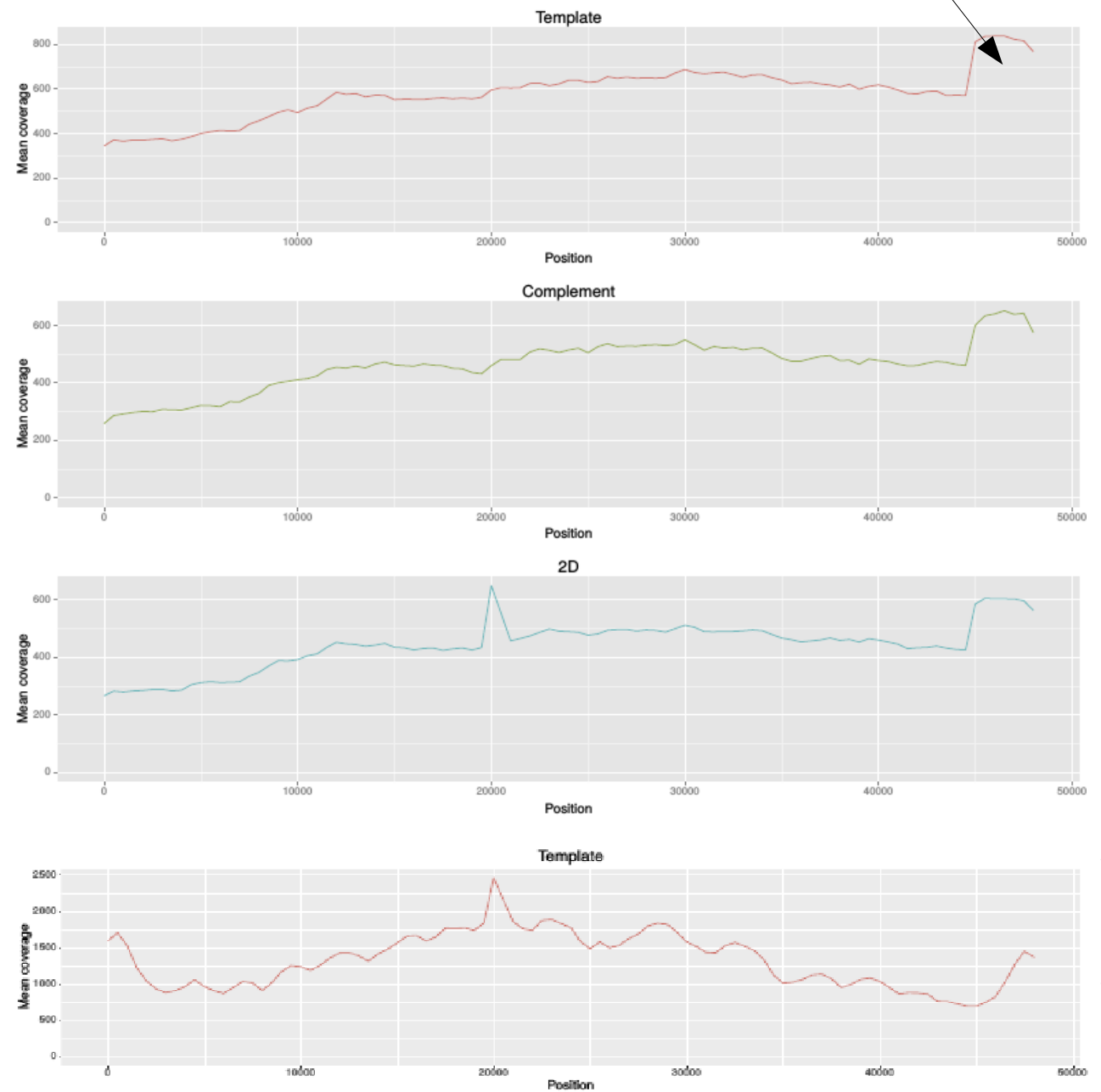
Mean Indel size :
1.56 (Minion) to 1.16 (PacBio)



Short indels : majority is 1 to 2

coverage

Internal control sequence



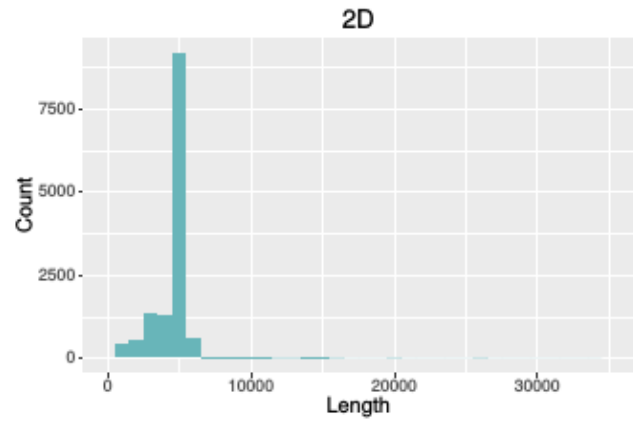
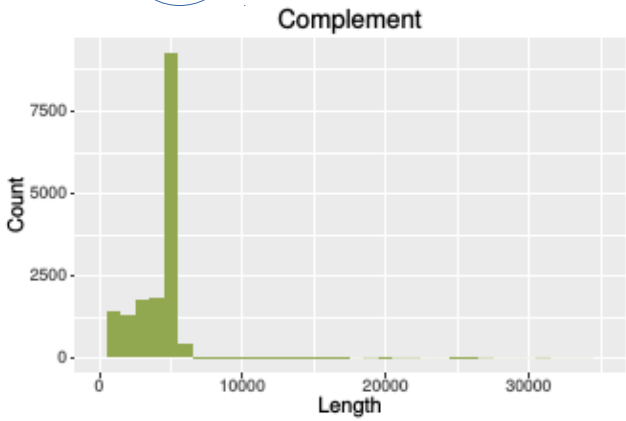
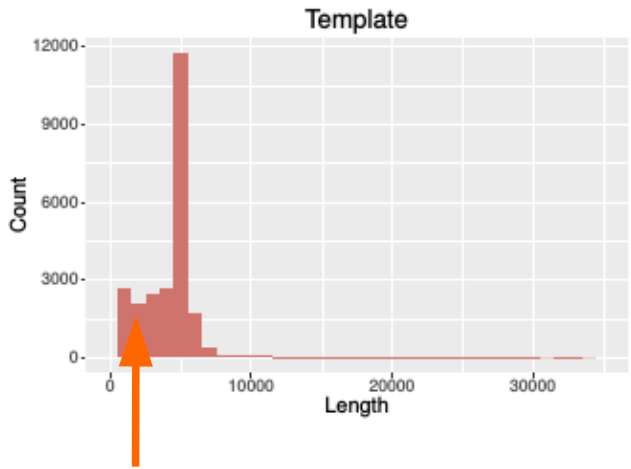
Daniel's Plasmid (5.3kb)

	Nb reads	Nb 2D reads	Mean (2D) Reads Length (bp)	Identical bases per 100 aligned bases (including indels)
Plasmid Daniel (2016)	27261	13585	4648	19.8

Read lengths

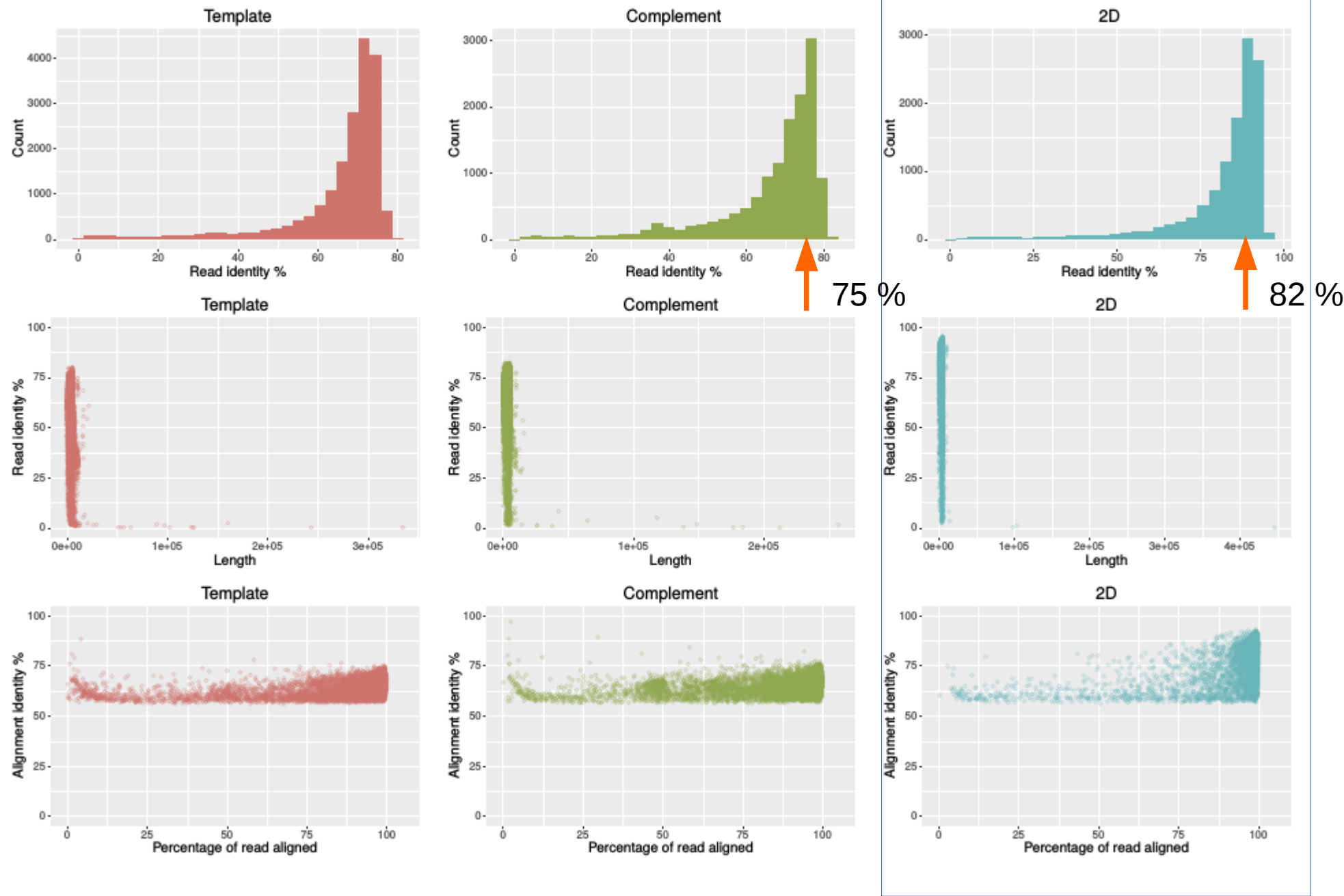
Unexpected longest read

Type	NumReads	TotalBases	Mean	Longest	Shortest
Template	26843	114867933	4279.25	945362	6
Complement	17123	70150773	4096.87	257350	10
2D	13585	63144394	4648.10	446163	118



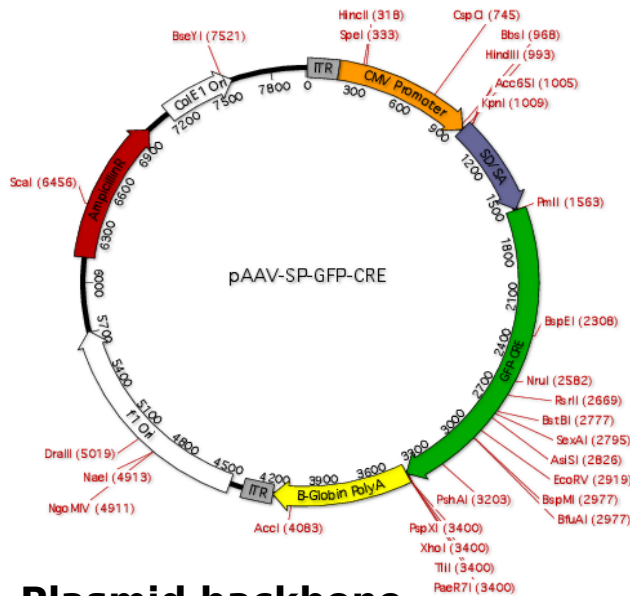
Fewer proportion of short read

AAV CMV GFP SfoI digest linearized read identity

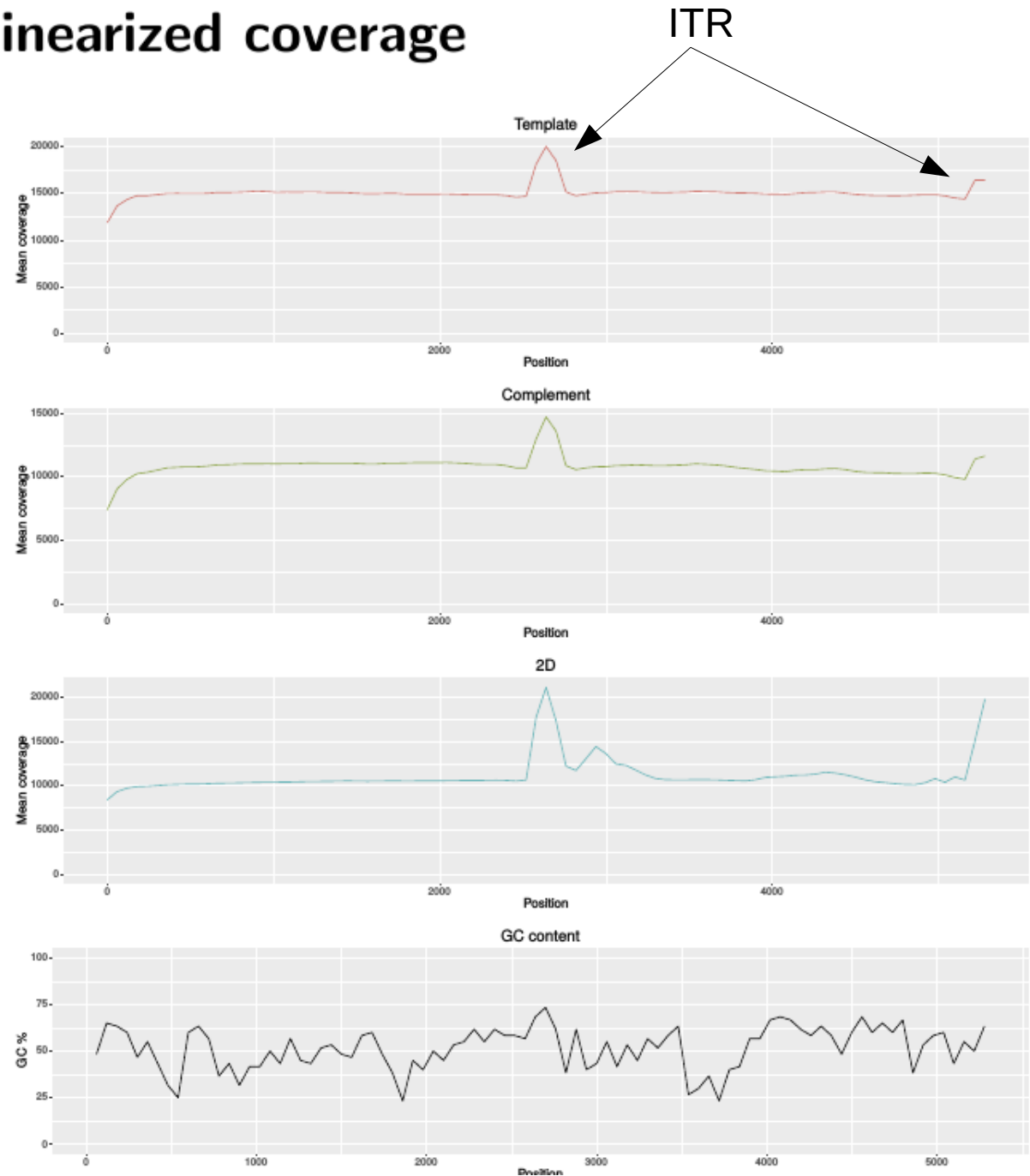


AAV CMV GFP SfoI digest linearized coverage

Very high coverage ~10000x
We might consider to
use more than one
Library : sequentially/pooled ?



Plasmid backbone



Julien's Drosophila cDNAs

genome : 172Mb

– 14 000 genes

– refseq 34 042 sequences

	<i>Nb reads</i>	<i>Nb 2D reads</i>	<i>Mean (2D) Reads Length (bp)</i>	<i>Identical bases per 100 aligned bases (including indels)</i>
cDNA Julien (2015)	159 974	113 250		

**MAPPING STILL ON
PROCESS
2+3days on Vital-IT**