Environmental metagenomics

Quality control and trimming

Igor S. Pessi & Antti Karkman

Quality control

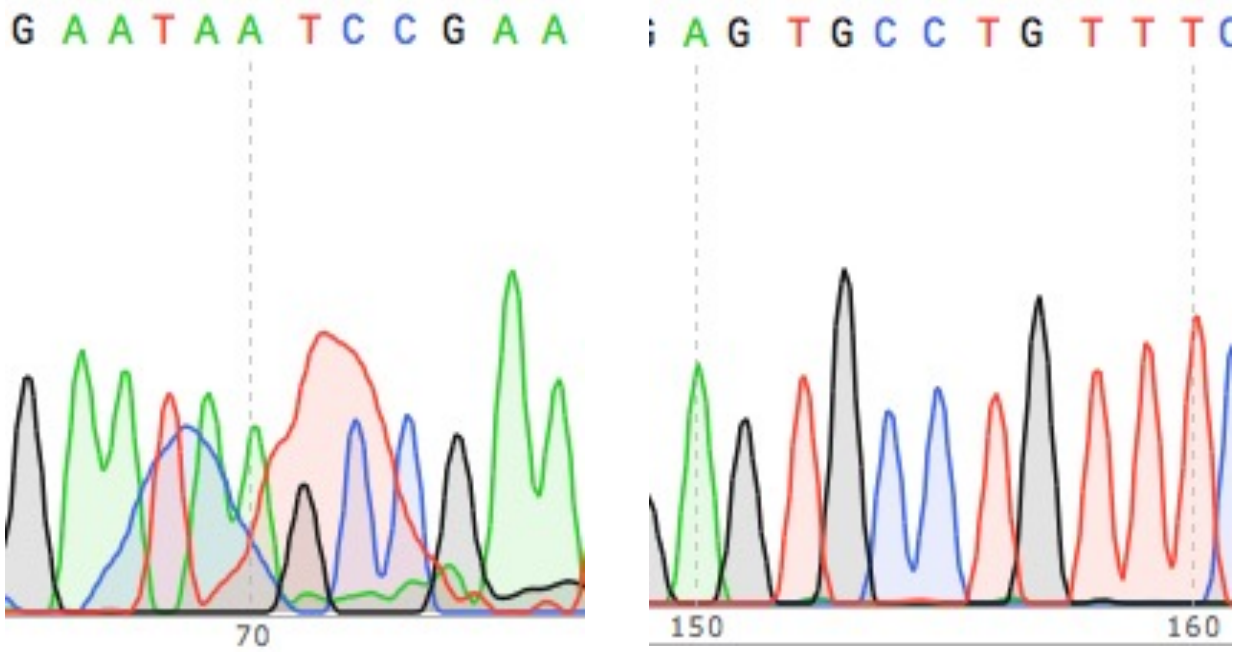
- Things to consider
  - Sequence quality
  - Adapter contamination (Illumina data)

Sequence quality

- PCR – what happens in PCR, stays in PCR

Sequence quality

- PCR – what happens in PCR, stays in PCR
- Sequencing

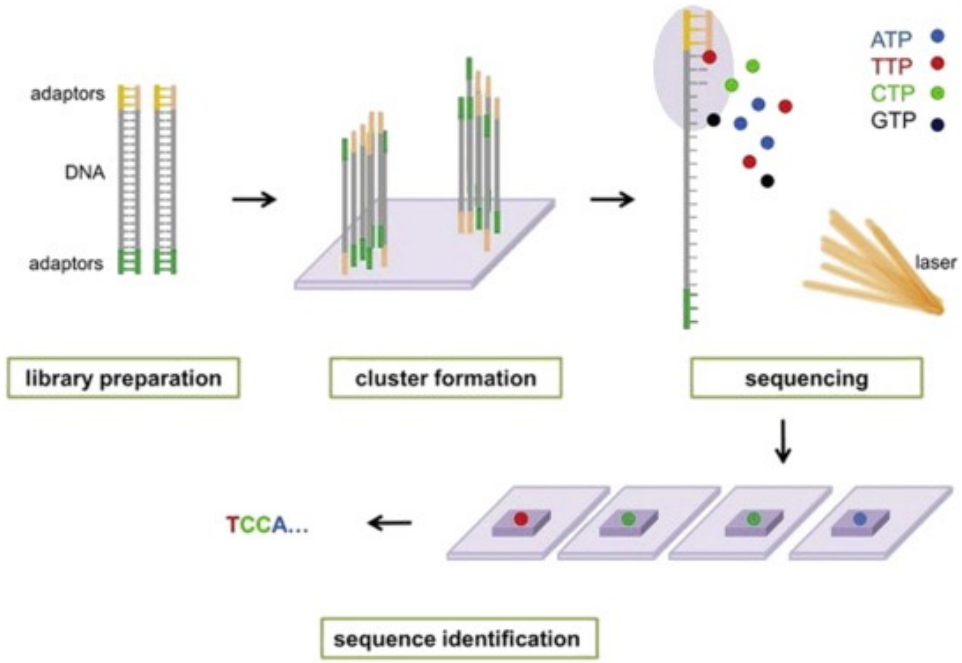


https://bitesizebio.com/10709/how-to-analyze-dna-sequencing-results-properly/

Sequence quality

- PCR – what happens in PCR, stays in PCR
- Sequencing

```
@A00464:69:H7L3KDSXX:1:1101:7726:1000 1:N:0:GTCTAATG+NGTGGTCA
GNATTGCCATCGGGTAAAGCGTCAGGAAGCCGAGCAGCAGGATCAGCAGGGCGATGCCGGCAG
GCCGGCCGAAGCGGCGGAGGCCGCGCGGCAGCGCAGCCGTCGGAAGAGTGACGGCAACCGTCA
TCGGCAGCCGCCCATCGTCAATCGT
+
F#FFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:F
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFF
```



Zhou et al. 2015. Atlas Oral Microbiol. https://doi.org/10.1016/B978-0-12-802234-4.00002-1

Sequence quality

- PCR – what happens in PCR, stays in PCR
- Sequencing

```
@A00464:69:H7L3KDSXX:1:1101:7726:1000 1:N:0:GTCTAATG+NGTGGTCA
GNATTGCCATCGGGTAAAGCGTCAGGAAGCCGAGCAGCAGGATCAGCAGGGCGATGCCGGCAG
GCCGGCCGAAGCGGCGGAGGCCGCGCGGCAGCGCAGCCGTCGGAAGAGTGACGGCAACCGTCA
TCGGCAGCCGCCCATCGTCAATCGT
+
F#FFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:F
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFF
```

- PHRED scores

$$Q = -10\, log_{10} P$$

$$P = 10^{\frac{-Q}{10}}$$

ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	|
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

Sequence quality

- PCR – what happens in PCR, stays in PCR
- Sequencing

```
@A00464:69:H7L3KDSXX:1:1101:7726:1000 1:N:0:GTCTAATG+NGTGGTCA
GNATTGCCATCGGGTAAAGCGTCAGGAAGCCGAGCAGCAGGATCAGCAGGGCGATGCCGGCAG
GCCGGCCGAAGCGGCGGAGGCCGCGCGGCAGCGCAGCCGTCGGAAGAGTGACGGCAACCGTCA
TCGGCAGCCGCCCATCGTCAATCGT
+
F#FFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:F
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFF
```

- PHRED scores

$$Q = -10\, log_{10} P$$

$$P = 10^{\frac{-Q}{10}}$$

ASCII 33

Q = 0

P = 1

ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	|
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

Sequence quality

- PCR – what happens in PCR, stays in PCR
- Sequencing

```
@A00464:69:H7L3KDSXX:1:1101:7726:1000 1:N:0:GTCTAATG+NGTGGTCA
GNATTGCCATCGGGTAAAGCGTCAGGAAGCCGAGCAGCAGGATCAGCAGGGCGATGCCGGCAG
GCCGGCCGAAGCGGCGGAGGCCGCGCGGCAGCGCAGCCGTCGGAAGAGTGACGGCAACCGTCA
TCGGCAGCCGCCCATCGTCAATCGT
+
F#FFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:F
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFF
```

- PHRED scores

$$Q = -10\, log_{10} P$$

$$P = 10^{\frac{-Q}{10}}$$

ASCII 33

Q = 37

P = 0.0002

ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	|
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

Sequence quality

- PCR – what happens in PCR, stays in PCR
- Sequencing

```
@SRR11673985.11 11 length=2169
TAAAGGATAAACTGTTTCACTATTTTCATTAATATCTTTTATCCTCATCAAAAGTTTCATATTTTACACAATATTCTTATTTCTATCAAAAATTGCTGATAAAGAGGTTACTACAGCATCTTCTTCTATGGATTTTTAGTTTTCTCGCCAATA
CATTGATTTTTATAATCTCTTGTATCATATCATTATTATAAACAAAAAATGGTGCATTACTTTTTAACAAGGGCTATTCACGCCTGACTTTATGGGGATCTTCTGACGCTTATTGAAATACCCAATGTAATAGTAATAAAAGCCGATATCAAC
TCTTCAAATATCAATTGATTATATTTTTAAGAAGTTTCTAAGCCGTAGGCTCTCGTATATTTTGGTCCGTCATTTTTATCACAACGTAAACGTTACCATATAGATTCTATAAATATGAACACTATATTTTGATGAAAATTCATTCAATTTTGC
TTCGAGTTAAACTAGGCTAGTTGAATTGTAGTTCAAGATCAACATCATCAAAAAGTCAATATCGAAATTAAGCAGGTTGATACAAAATCATCTTCTTTATGTCGTATTTCTTATATCTATTTTTAAATTCTGTTAATTCATCATAATAATTTA
TTTTGAATCTCTTTTCGTTTCTTGAACATATCAAAGTTTTTGTTGTTTCAACTTGGTCCAGTAAATGAATTAAACTGTGTTGCTAAAAACCTATTTCATCCTTTGATTTAACTATTTTTCAAAATCGAGTTATTTTGCCTTTTTCAACTTCAG
AAAACATTTTATTAAATTGGAATAAATAACTGGCAATAACATTGAATTATTAACAAAACTCAAACAAAGCACATTAATCATTAGAATTAATTCACAATAAGATGAAGAGTTAACAACATAAATAGAAAGATACATATACCAATCCCAATTTTT
GGTTGTATTCAACCGTAAATTTTTTGATAAACATAATTTACCGGGGTCCTCAACCTTTGTTCACGGTCGCCATATAGGGTTTCTTTTTATTATAATGATTTTAGTTTATTAATAATTGTTTACCCATTGGGCGCATTAATTTAGATGTTTCAA
CATTTTGAAATTCAGGAGCAGATTCAAAAATAATCTTTGAATCAAAAATAATATGTATATTCAAGCCTATTGCAAAACTTTGTGACCAATATTCTGTCTCTTTTTTAAATCATCTTTATTTTCGATGCGCGATTTGTATTTATAATTGAAACT
ACTTTATTAATTATTGTTGCAGGTTTCATCTCCGTATCCGGTCAAATAACTATTTACTTTAGGCGATAAAAAGACTAAAAAGATAATGTCTTAAAACTATTGGCAATAACCAACAGATTTTTAAGAAATGTCCTTATTCTTTCCACGCTTACA
ATCTACAATGCTTTATTGCTATTTTTGACAAATCTTACATTTTGATTTAATAGTTGCTGAAGTAATGCAACCATCTCTTTCCCATAACATTACAACAGGTTGAACCCATTTCAAAAATATCCTAAACATTCACCTTTTGATATTTAAGGTATT
TTAGTAGTATTACAGCTTTCCTCTCTTTTTGCAATTTGTTGTTTACTCTATTTCAAATTAAATACCATTTGATAGCATTGTCTAACAAAATACCATAAAGATTTTTCATTGCAACACATTCTCAAAATAACTCTTTCGACATTAAAAAGGGTT
CGGCTCTTTATTTAATTTTTAGATTTACAGGATAAGTTTCCCCTTGGATTTAATGGATTAATTTTAATAACTTAAAATCTGGGTGTATTATGCTGTGATCATCTAGGTGATAAATAGAAATTCATAAAAGAACCATCTTTTATTTTCTCAAAA
TTTGATAGAATAATAATAAGTTATACAACTCTTTAGCGCTTATATTCAACACAACACATTGCTTAATGCCATATCTGTTTGAATTTGTCCACACATTGTGTGATTAAACTATCAGTTGGAGAAATGTCCACTTGAGTTCAATGTCTCCTCTTT
TTATGATTAACTCTCTTGTAAAAATAAATCATTTAAAGATTTGTAATATCTCGCATTTTTAAATTCACTCATATTAAAGCCCATTAACGTAACTAGAATTTATTAAGTACGTATGGAGATTGGAAGGAAATGCACCTTGAACTTTCCAAAATA
TGAAGAGATAAGATTTGTATGCATTAA
+SRR11673985.11 11 length=2169
35401.(%&)'+21CDD9*%869<DG;$$&$$/252//685//3:@>A@D?<;<FIH8,((---)%#)'764445.*+&&)0()20>B</+%%&%#%+,/000%('&&).754+637>>:@6:?B+;755<5=;::<'8)9@B;98$$$0,))
*)*+8:J?@?72==;:953('%(($)*+-77'''$#%%%11$2<A?:8210(&%&$')$(****+$%%*,+().//0025$%01/04:78947((()),955:<.%)##%$%*))*%%*$%%%&%872(&0.*.,/0=>F?4502128;4.2%'&
'('($))'+,)((((')%#%(&%'))))((&+7%)(*,,%$$%'$$$$$$%(12.)$%,156?<:5.*%%'%'%%$%%%('+)%%((&'/04+.00?)12*)($('''&(%%%&'$$&$%&%$&$%-''1565((/-.-.+,,1---%%%%--..$$
%'&$$$$&&)%$$&$$%'$&),$$$$%$%%)2%(($%%&'&$%%,21:4?9.(+01268/(:$&&$+,-6'+*%$12,0..21676>5:<755.--)*(((''%(')**%$$)=@;<@+<?@A?@>5::#'(78$&%&:830'$)&$%(&0.)
+//.%$'%%$%&''%$(,,,$#$&$$#$$#08(516H7332&$&/25?=5<55666%%()$345444489:=@:==130:;+,:5F=?+4.<1>=<59345543,(+6+0,)/+15554'-.)()(#')789<:::7544=BFG21476::63
1-+'%*+++*%%&'&#$&%%&$)*'&'%#$$$()''(&#%&.$&&$&%%)$$$&,../06&(&&&(%#$%#%''%(&%&*+&(&+'())(1.$&/1000/2)(''&$%&(%$$*(+5/55.--,1/(&%)'#455.15557**($&100//0
30;;)003497...+*,/;<:998311,-0&%-,02./'&%%%6/4&&*')'$#$%+-+&((**-/16/21+*-%%$%'(((+10452&&'%$&'((%*,+)$(./##%&'37==9;/156/8+('%'$)&%#$%%(%&#'&&&4<:;4-,0,
++468:68:453+*##$&''1$*,0.-8;66=732%(0/.>:=A=44**-.),$%)()*%,-('$',-'&'%#.021//12'%%$#$$$#$$$)**45944;AAA87AAGE:;?LL865:@@;))'$#$&,&46*&((55&;76+12%)40.0
+-895--./5',,+,%''%%%(,011(-43&:)7:>))+&'5(+.<?D>544--**&'.&*0/765456++))))*+13/,++$%&%%,+*+-,)'$%%&&$$$**())'))&%&&*50.41%%$,*%%%##)%'%())(,-('$#$#$$%%-
56<>2131/*$%-$,,-414'1/---,))&.-63359<>EA@=34G973.'%*#&%%&$$$%&'''&/3&..73439>>EH=80>=511142:;0)),*:9+**776)+8;$(*==6:;--89<9*64<;D6@<>;:-7.%%32%&+%'%$**
*%&&''%*&%&$%,$*++.0)*()+**)&%$#%%)'$(*$&'$$&0,)$)(($&%$%&%&%%&'+++).)$&%%&-.&'%#$%%)+-'&(%%%$(,*$$:<===<,,.%*&&$%%(-0.%&'*)&%&(+-.-..($'&&#%261)&10/)(
&%''()(.97<42**$'(()/446,)*%%>9/%%&%1344'('&'&&##%&00%%7$45675765;>885++++9979**)&*%%'&''+($#&&$()(&(%&%%*,/0033;411,<@;C@?6@@DCB64;;;=>:IEB86;<<A@E;,2)$
%&&&$%$%%&'/75?A=@6066)%%&))))()&'&%#$#$%(/;;@@;&($%)(&&&$'12*&&',.;9<<<7689%%9940//9:(3/.0%)+2377'-'./''($'))176=)2***061453%%)$%%$$$%#-**)$%'&)0//-43672
229?>BBC7+%,-)+)*,2:644,69=?@BCHH<;931167.7.8<;<677:,08?@@?;3,)(-;&788)&&'()*()*)))))*%#%(*'187(-101(&#$%$%#$%.34$$&$%$#%%%(/%$0+*(&#$#$#$*))$369?.JDD2//
/62@45?@<==>9760'%(&'2,78?=
```

Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
32	20	[SPACE]	64	40	@	96	60	`
33	21	!	65	41	A	97	61	a
34	22	"	66	42	B	98	62	b
35	23	#	67	43	C	99	63	c
36	24	$	68	44	D	100	64	d
37	25	%	69	45	E	101	65	e
38	26	&	70	46	F	102	66	f
39	27	'	71	47	G	103	67	g
40	28	(	72	48	H	104	68	h
41	29	)	73	49	I	105	69	i
42	2A	*	74	4A	J	106	6A	j
43	2B	+	75	4B	K	107	6B	k
44	2C	,	76	4C	L	108	6C	l
45	2D	-	77	4D	M	109	6D	m
46	2E	.	78	4E	N	110	6E	n
47	2F	/	79	4F	O	111	6F	o
48	30	0	80	50	P	112	70	p
49	31	1	81	51	Q	113	71	q
50	32	2	82	52	R	114	72	r
51	33	3	83	53	S	115	73	s
52	34	4	84	54	T	116	74	t
53	35	5	85	55	U	117	75	u
54	36	6	86	56	V	118	76	v
55	37	7	87	57	W	119	77	w
56	38	8	88	58	X	120	78	x
57	39	9	89	59	Y	121	79	y
58	3A	:	90	5A	Z	122	7A	z
59	3B	;	91	5B	[	123	7B	{
60	3C	<	92	5C	\	124	7C	|
61	3D	=	93	5D	]	125	7D	}
62	3E	>	94	5E	^	126	7E	~
63	3F	?	95	5F	_	127	7F	[DEL]

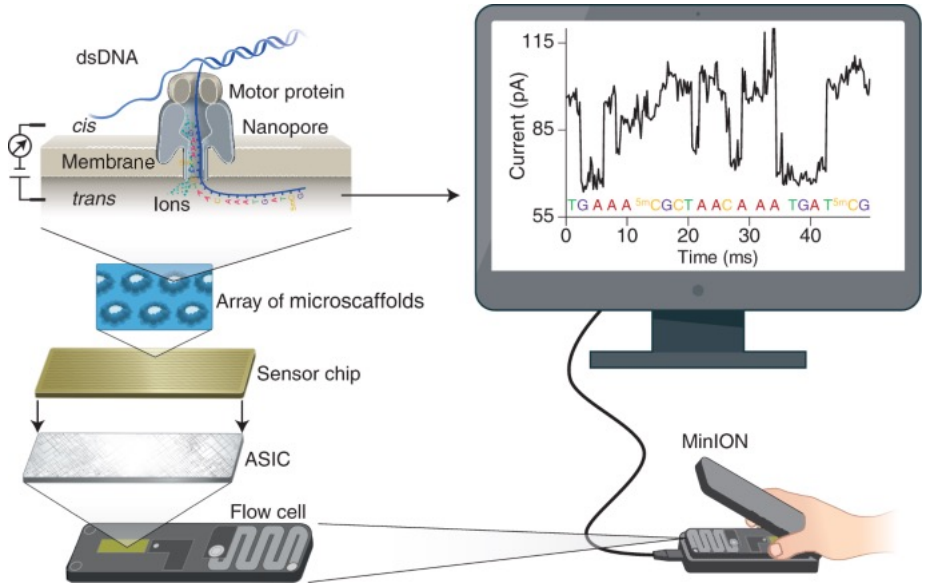
Sequence quality

- PCR – what happens in PCR, stays in PCR
- Sequencing
- Illumina:
  - Occurrence: substitutions > indels
  - Quality scores: substitutions < indels
  - Overall quality: R1 > R2; beginning > end

> ”G seems to be preferentially incorporated if an A, C or T is sequenced and if G is sequenced a T is falsely incorporated for the majority of substitutions.”
>
> Schirmer et al. 2016. BMC Bioinformatics

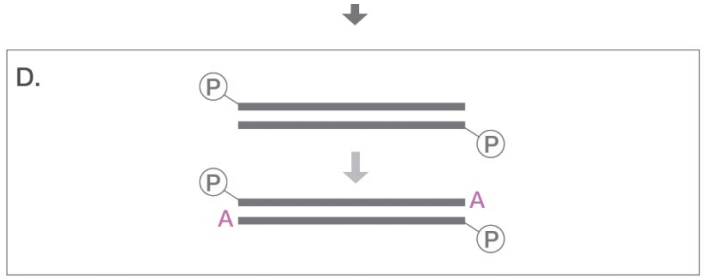
Sequence quality

- PCR – what happens in PCR, stays in PCR
- Sequencing
- Illumina:
  - Occurrence: substitutions > indels
  - Quality scores: substitutions < indels
  - Overall quality: R1 > R2; beginning > end
- Nanopore:
  - Substitutions
  - Problems with homopolymers (AAAA, GGGGG)

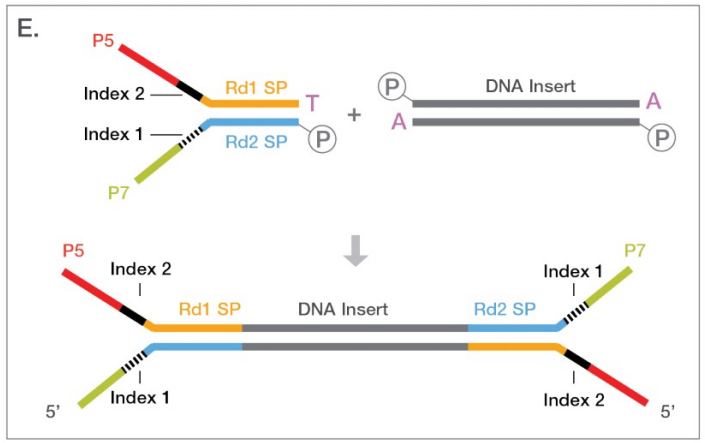


Wang et al. 2021. Nat Biotechnol. https://doi.org/10.1038/s41587-021-01108-x

Adapter contamination

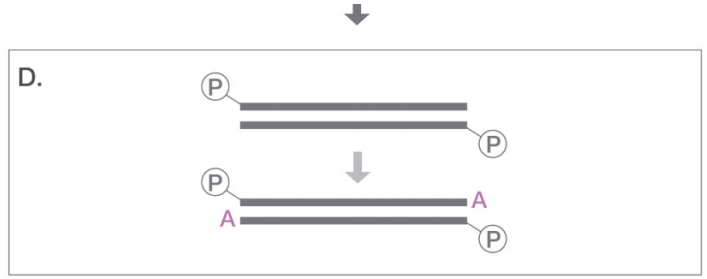


A-base is added.

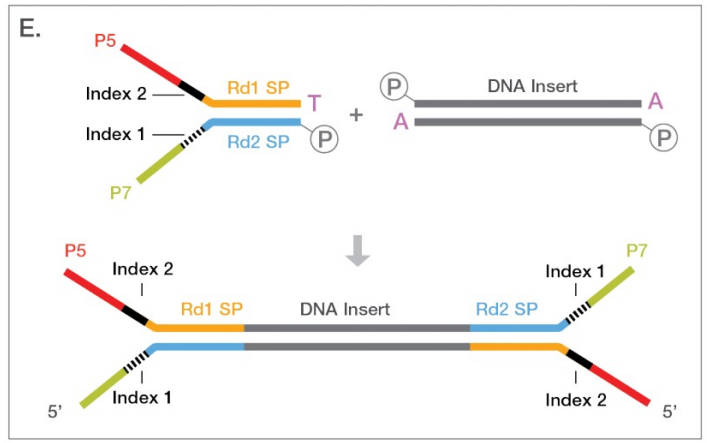


Dual-index adapters are ligated to the fragments* and final product is ready for cluster generation.

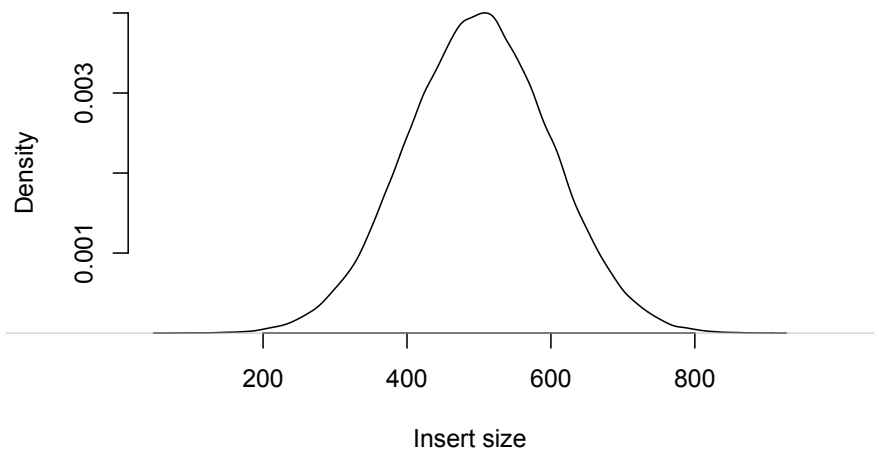
Adapter contamination



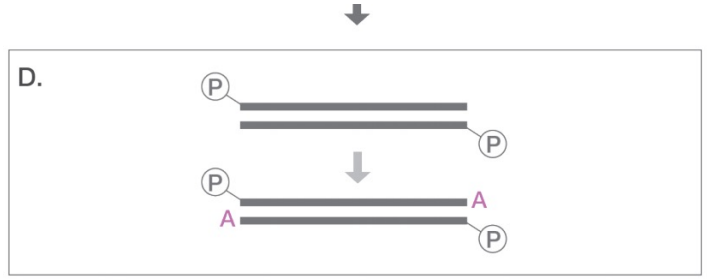
A-base is added.



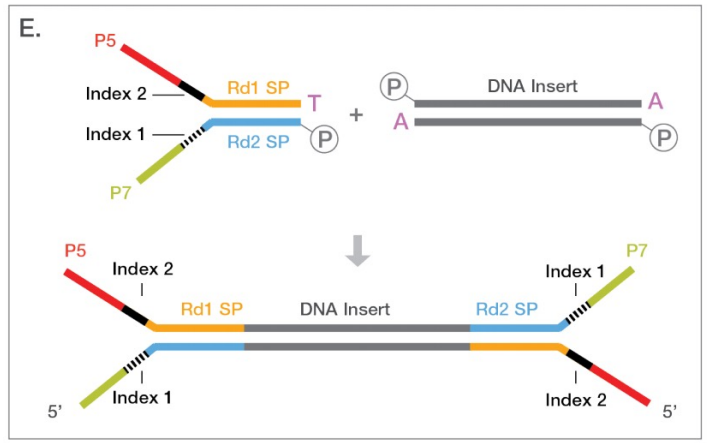
Dual-index adapters are ligated to the fragments* and final product is ready for cluster generation.



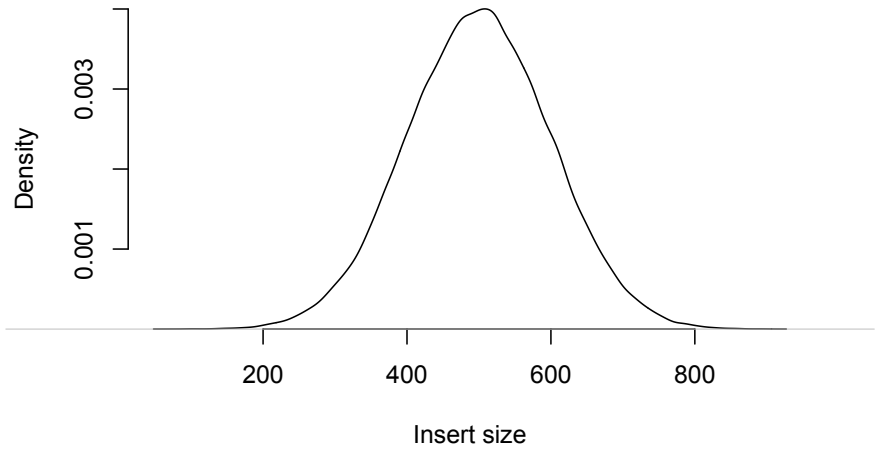
Adapter contamination

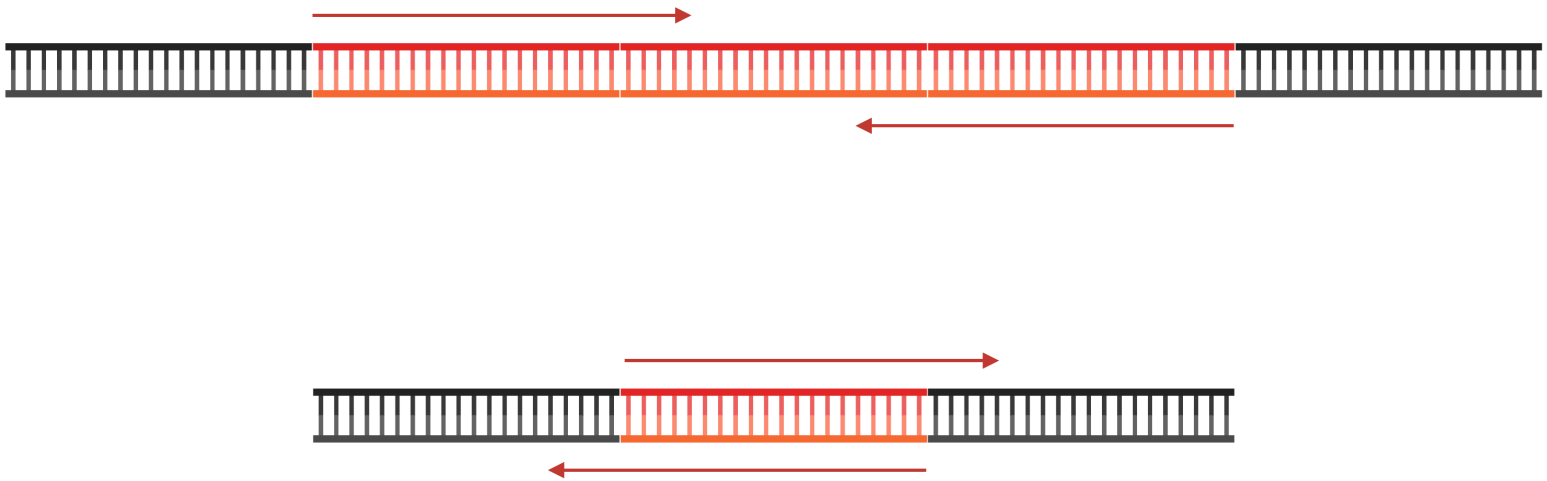


A-base is added.



Dual-index adapters are ligated to the fragments* and final product is ready for cluster generation.





Trimming

- What to do then?