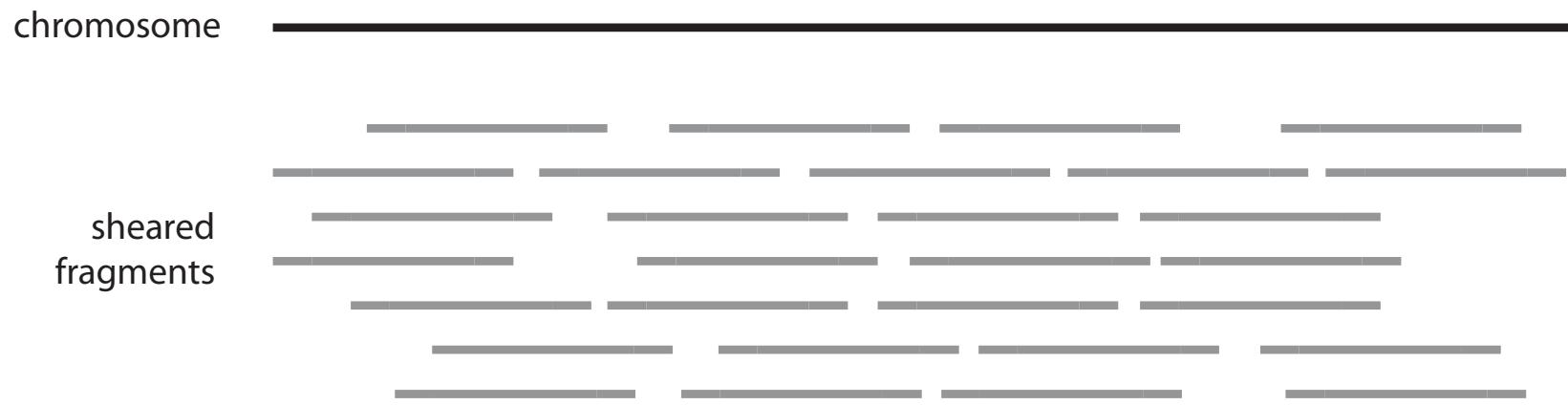


Short read assembly

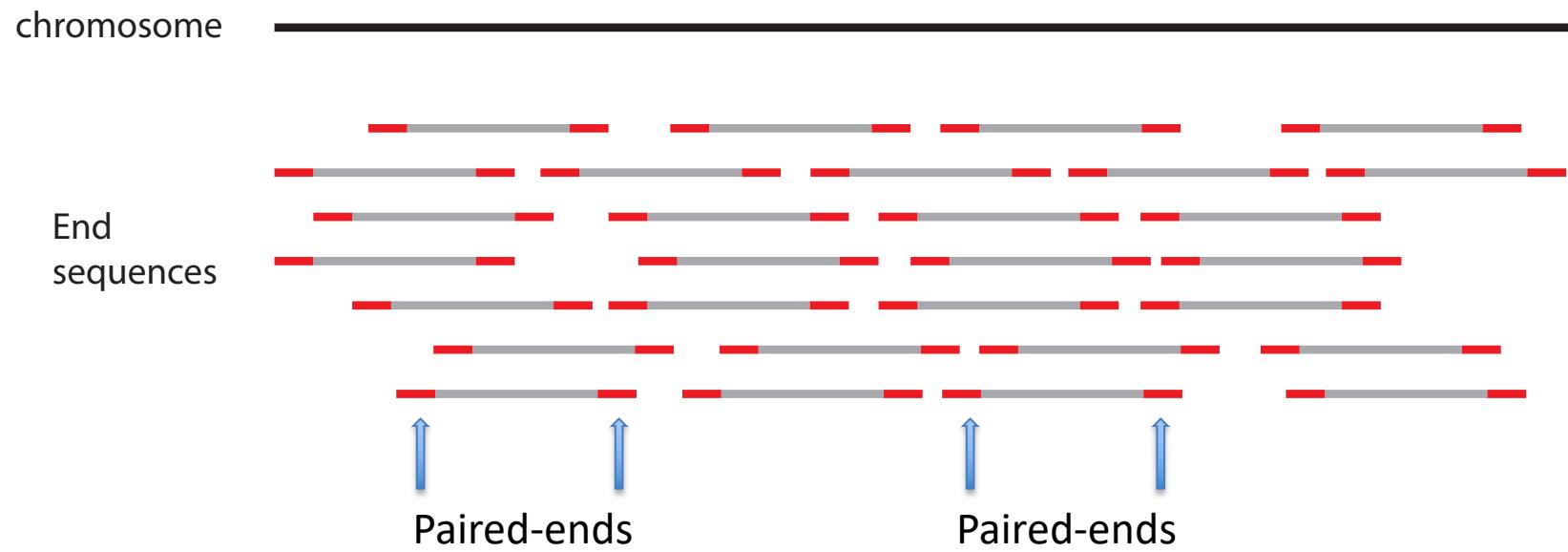
chromosome



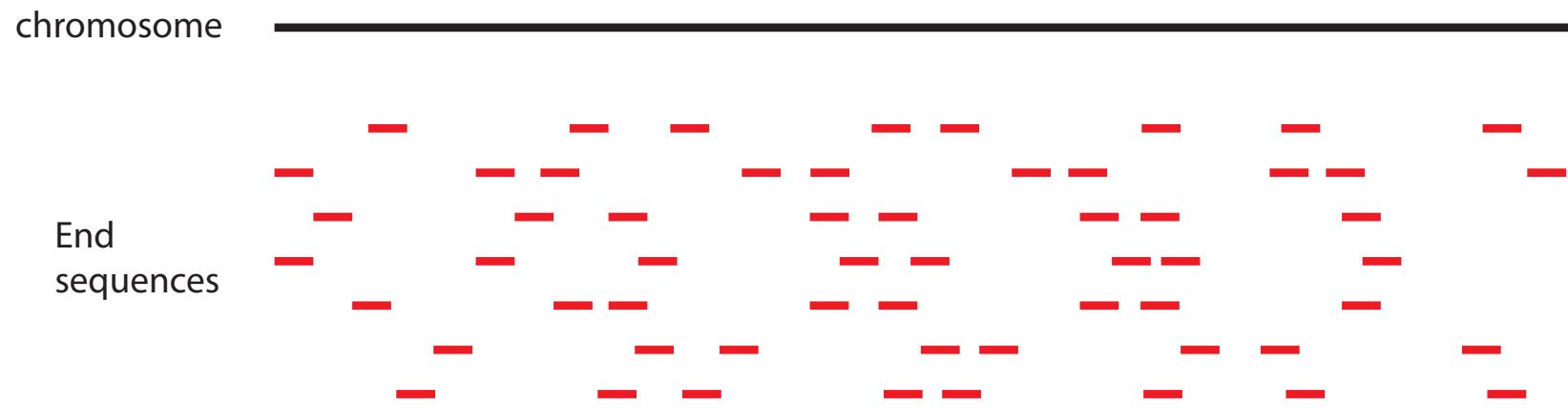
Shear genomic DNA



Sequence both ends of each fragment



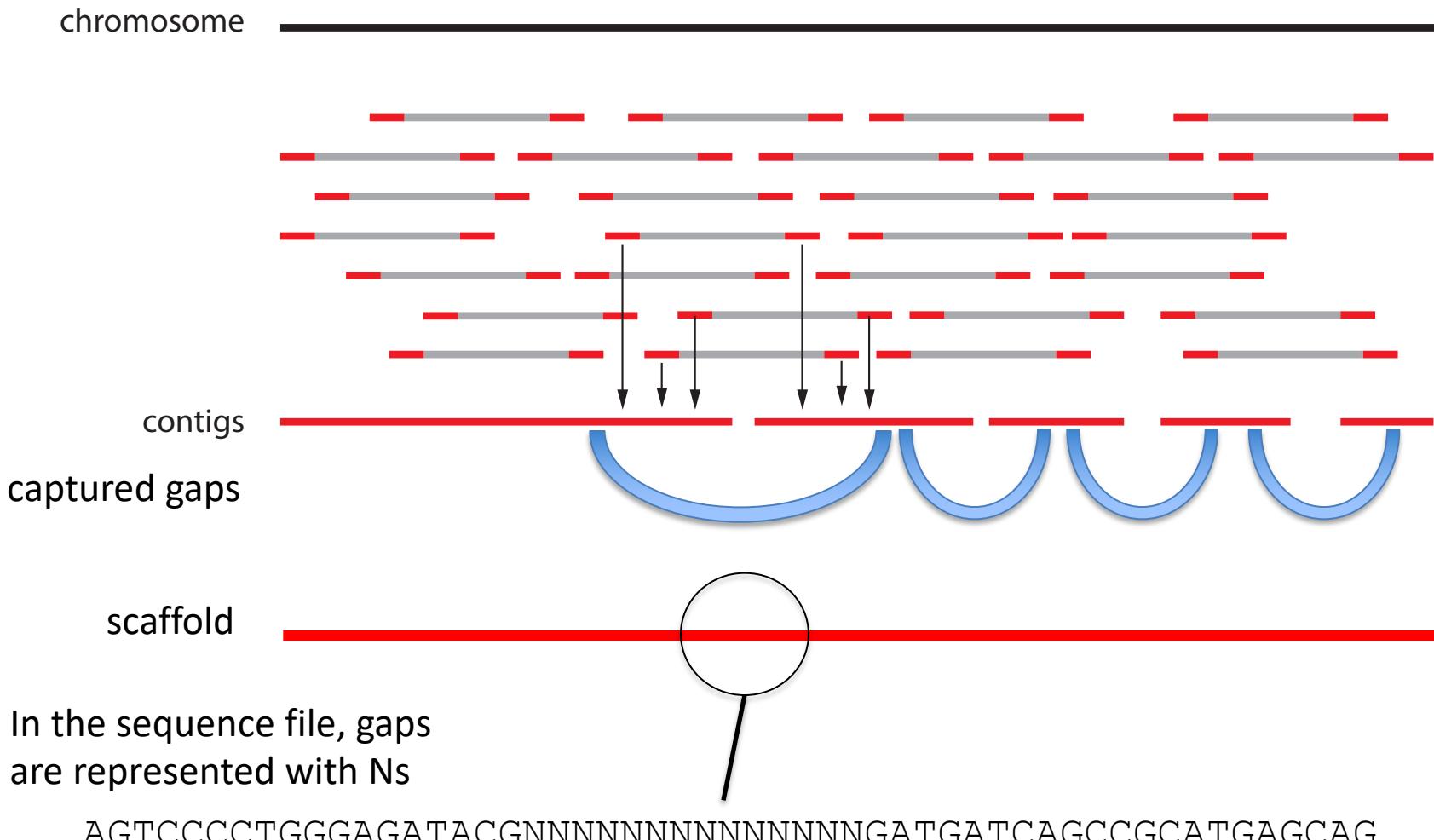
Sequence both ends of each fragment



Align sequence reads to form contigs



Paired ends allow linking of contigs into scaffolds



Assembly metrics

- No. of scaffolds/contigs
- Largest scaffold/contig
- N50 scaffold/contig size
 - 50% of genome contained in scaffolds/contigs of size \geq N50
- L50
 - Minimum number of scaffolds/contigs with summed length \geq 50% of genome
- Genome coverage (read coverage)
 - Each base represented by an average of X reads

Scaffold statistics

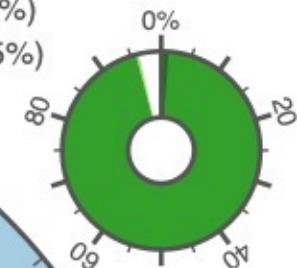
- Log₁₀ scaffold count (total 86)
- Scaffold length (total 1 GB)
- Longest scaffold (92.9 MB)
- N50 length (56.2 MB)
- N90 length (39.2 MB)

“snail plot”



BUSCO (n = 5,286)

- Comp. (95.9%)
- Dup. (1.2%)
- Frag. (0.5%)

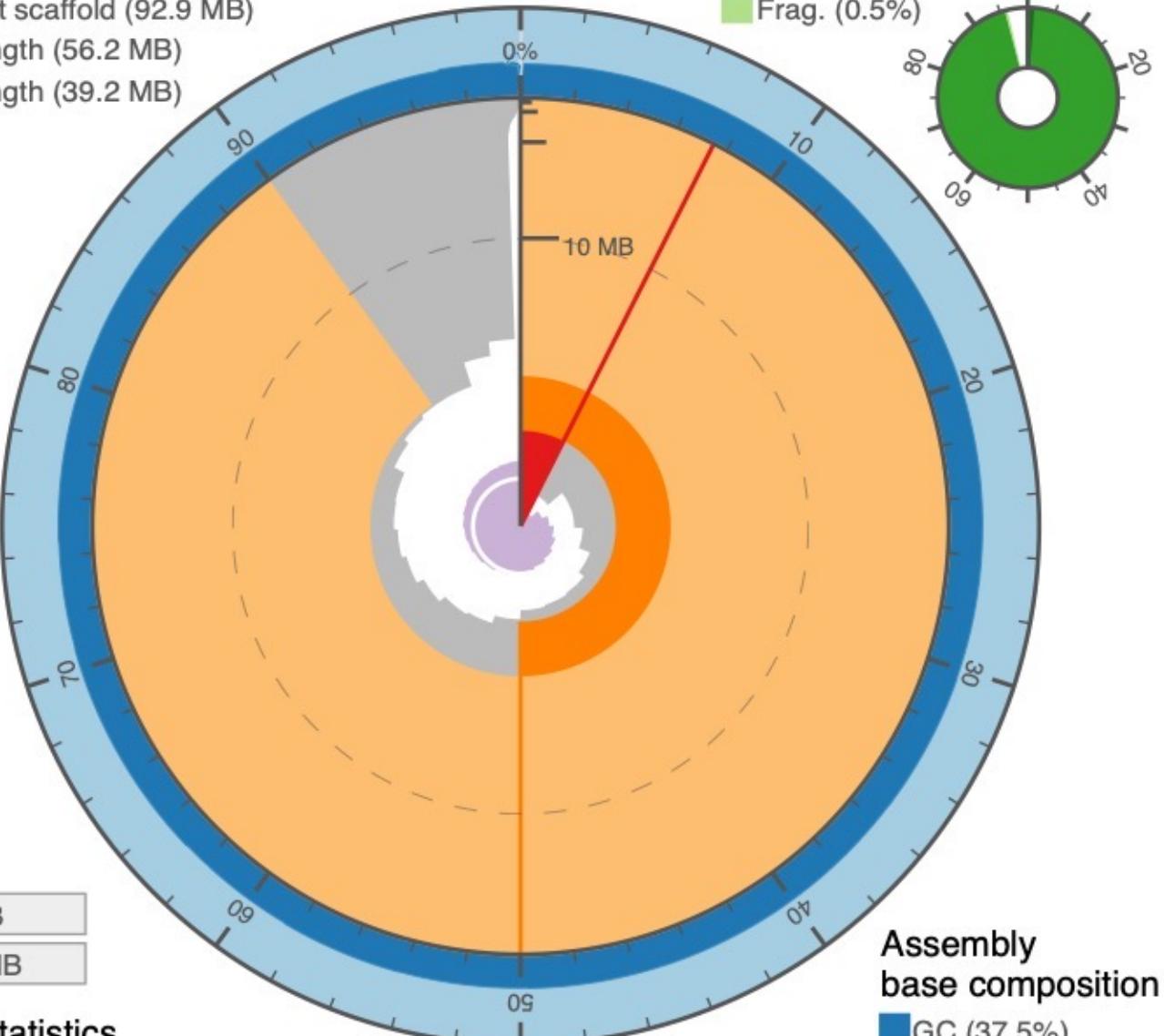


Scale

- 1.2 GB
- 92.9 MB

Contig statistics

- Log₁₀ contig count (total 1,067)
- Contig length



Assembly base composition

- GC (37.5%)
- AT (62.4%)
- N (0.0%)

GNU nano 2.9.8

SRR22045704.paired.sam

sam format

Sam format

^G Get Help **^O Write Out** **^W Where Is** **^K Cut Text** **^J Justify** **^C Cur Pos** **M-U Undo** **M-A Mark Text** **M-]** To Bracket **M-▲ Previous** **^B Back**
^X Exit **^R Read File** **^| Replace** **^U Uncut Text** **^T To Spell** **^** Go To Line **M-E Redo** **M-G Copy Text** **M-W WhereIs Next** **M-▼ Next** **^F Forward**

A

Coor	12345678901234	10	20	30	40	
ref	AGCATGTTAGATAA**GATAGCTGTGCTAGTAGGCAGTCAGGCCAT					
+r001/1	TTAGATAAAGGATA*CTG					
+r002	aaaAGATAA*GGATA					
+r003	gcctaAGCTAA					
+r004	ATAGCT.....TCAGC					
-r003	ttagctTAGGC					
-r001/2	CAGCGGCAT					

sam format

B

QHD VN:1.5 SO:coordinate
@SQ SN:ref LN:45

Header section

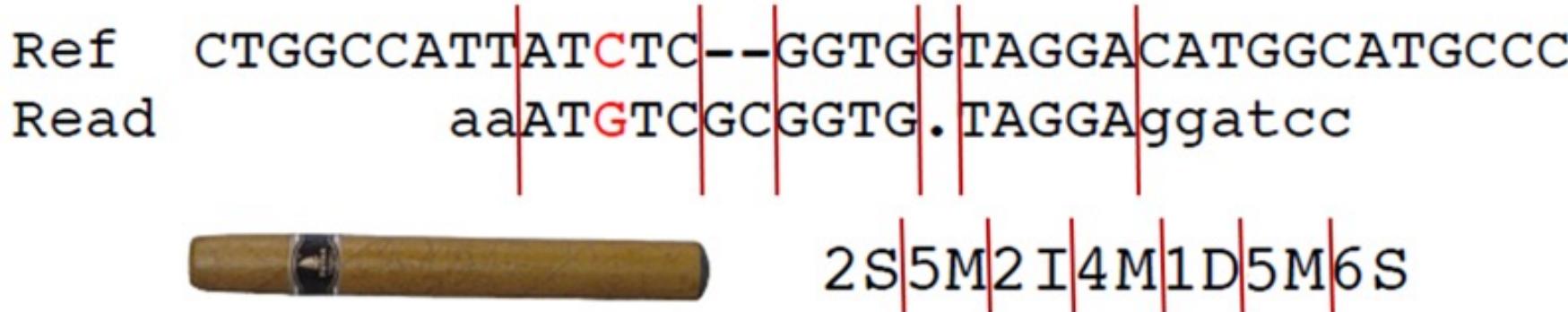
QUAL (read quality; * meaning such information is not available)

r001	99	ref	7	30	8M2I4M1D3M	=	37	39	TTAGATAAAGGATACTG	*
r002	0	ref	9	30	3S6M1P1I4M	*	0	0	AAAAGATAAGGATA	*
r003	0	ref	9	30	5S6M	*	0	0	GCCTAACGCTAA	* SA:Z:ref,29,-,6H5M,17,0;
r004	0	ref	16	30	6M14N5M	*	0	0	ATAGCTTCAGC	*
r003	2064	ref	29	17	6H5M	*	0	0	TAGGC	* SA:Z:ref,9,+,5S6M,30,1;
r001	147	ref	37	30	9M	=	7	-39	CAGCGGCAT	* NM:i:1

Alignment

QNAME	FLAG	RNAME	POS	MAPQ	CIGAR	RNEXT	PNEXT	TLEN	SEQ	Optional fields in the format of TAG:TYPE:VALUE
(query template name, aka. read ID)	(indicates alignment information about the read, e.g. paired, aligned, etc.)	(reference sequence name, e.g. chromosome /transcript id)	(1-based position)	(mapping quality)	(summary of alignment, e.g. insertion, deletion)	(reference sequence name of the primary alignment of the NEXT read; for paired-end sequencing, NEXT read is the paired read; corresponding to the RNAME column)	(Position of the primary alignment of the NEXT read in the template; corresponding to the POS column)	(the number of bases covered by the reads from the same fragment. In this particular case, it's 45 - 7 + 1 = 39 as highlighted in Panel A). Sign: plus for leftmost read, and minus for rightmost read	(read sequence)	

CIGAR string



Op	BAM	Description
M	0	alignment match (can be a sequence match or mismatch)
I	1	insertion to the reference
D	2	deletion from the reference
N	3	skipped region from the reference
S	4	soft clipping (clipped sequences present in SEQ)
* H	5	hard clipping (clipped sequences NOT present in SEQ)
* P	6	padding (silent deletion from padded reference)
* =	7	sequence match
* X	8	sequence mismatch

"N" indicates splicing event in

RNAseq BAMs

*Rarer / newer

CIGAR = "Concise Idiosyncratic Gapped Alignment Report"

fastq format

```
@M01478:6:00000000-A40C5:1:1101:16859:1439 1:N:0:1
```

```
ATCGTTCGGAGCAAGGCAACTGTNTCAGGCACCATGAAGTTGAGCTATTCTACTGCGCCAACCTTGCGAGA  
GCCNTNNTTATCANCGTCAATTGGAANTCAGATGTGCCACCNAAN
```

```
+
```

```
ABBAABFBBBBGE GGFGGGGGHF#AAFF2AGFGHGHHHHHFHFFDGF GHHHHGHEGGGGCGGGHFABEEGF  
BFG#?##??FFH#??FEFGHHEHHG#??FFEDGGGFHFH##??#
```

- Machine name
- Run number
- Flowcell ID
- Flowcell lane
- Tile in flowcell
- X-coordinate in tile
- Y-coordinate in tile
- Member of read pair (1/2)
- Read filtered? (Y/N)
- Control bits on (0 or even number)
- Sample number (when multiplexing)

```
[jdu282@mcc-login002 Bdor_pop_WGS_SRA_data]$ zcat SRR22045704_pass_1.fastq.gz | head  
@SRR22045704.1.1 1 length=150  
CNCCGGCTTCGAAGCTTCTCCCTATTATAAAAAAAAATTACAAACTAAAATTTTGTCCTAGCTGTATCAAACGATA  
ATTGTGCTACGGTTGGCTTGATAAGCAAAATAATGTATTGGAACAGACGGAGATGGTT  
+SRR22045704.1.1 1 length=150  
F#FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF  
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF  
@SRR22045704.2.1 2 length=150  
TNGCGATCAGTTCCGCATGCCAGAACAGCAATTGCCACATTCTCGCTGTTCTATTGCTATTAAATGCCGGTCCA  
CCATCTGACCACCATTACACCCATACTCCGTGAGGATGTGCATTGCTCGCGATAAGAATTG  
+SRR22045704.2.1 2 length=150  
F#FFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFF  
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFF  
@SRR22045704.3.1 3 length=150  
TCCATTATTAGCAATCTTAGTTCTAAGAAAAACAAGTTGTCTATAAATTATTTCAAATTAAAATGAATATAAACAACTTT  
ATTGCATATGGATTTCTTATTAAATTAAACAACTAAGTAATATTATAATAATAT  
[jdu282@mcc-login002 Bdor_pop_WGS_SRA_data]$ zcat SRR22045704_pass_2.fastq.gz | head  
@SRR22045704.1.2 1 length=150  
TTAAGATTATTATAATTAGGATAAAAAACAGTGGTTAGTACGCAAAACTTAGTTGTATTAAGTATATATTGCATTTGT  
TAACATTAATACATTTATATAACAATGTTCTGTAATTTAGGTTATAACTATGTAAG  
+SRR22045704.1.2 1 length=150  
FFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFF:FF  
FFFFFFFFFFFF:FFF:FFFFFF:FF:FFFFFF:FFF::FFFF:  
@SRR22045704.2.2 2 length=150  
ACCACCAATAGCATCGGAAATACACTTGACACACCGAATATCATGTTGCCAGCTGGTGGCTAACACGTAGAGCTGCGGCCG  
CAACGAAAATGACCACCGAGAGCAACATAAGCATAGCCGGCACGCTGAAGGCGAGTGAAAAACA  
+SRR22045704.2.2 2 length=150  
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF  
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFF  
@SRR22045704.3.2 3 length=150  
GTGGAATGATAAAATTAACTAACTTACATTCAAAGTACATAATTCACTCAGGAGAACGGTGCAGGTAGCTATTAAAT  
AAACACAAATAAGAAAATCTATGATAATATCAAAATTATATAATTGGTACTGAAGT
```

Phred quality score

A quality value Q is an integer representation of the probability p that the corresponding base call is incorrect.

$$Q = -10 \log_{10} P \quad \longrightarrow \quad P = 10^{\frac{-Q}{10}}$$

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

ASCII TABLE

Decimal	Hexadecimal	Binary	Octal	Char	Decimal	Hexadecimal	Binary	Octal	Char	Decimal	Hexadecimal	Binary	Octal	Char
0	0	0	0	[NULL]	48	30	110000	60	0	96	60	1100000	140	`
1	1	1	1	[START OF HEADING]	49	31	110001	61	1	97	61	1100001	141	a
2	2	10	2	[START OF TEXT]	50	32	110010	62	2	98	62	1100010	142	b
3	3	11	3	[END OF TEXT]	51	33	110011	63	3	99	63	1100011	143	c
4	4	100	4	[END OF TRANSMISSION]	52	34	110100	64	4	100	64	1100100	144	d
5	5	101	5	[ENQUIRY]	53	35	110101	65	5	101	65	1100101	145	e
6	6	110	6	[ACKNOWLEDGE]	54	36	110110	66	6	102	66	1100110	146	f
7	7	111	7	[BELL]	55	37	110111	67	7	103	67	1100111	147	g
8	8	1000	10	[BACKSPACE]	56	38	111000	70	8	104	68	1101000	150	h
9	9	1001	11	[HORIZONTAL TAB]	57	39	111001	71	9	105	69	1101001	151	i
10	A	1010	12	[LINE FEED]	58	3A	111010	72	:	106	6A	1101010	152	j
11	B	1011	13	[VERTICAL TAB]	59	3B	111011	73	;	107	6B	1101011	153	k
12	C	1100	14	[FORM FEED]	60	3C	111100	74	<	108	6C	1101100	154	l
13	D	1101	15	[CARRIAGE RETURN]	61	3D	111101	75	=	109	6D	1101101	155	m
14	E	1110	16	[SHIFT OUT]	62	3E	111110	76	>	110	6E	1101110	156	n
15	F	1111	17	[SHIFT IN]	63	3F	111111	77	?	111	6F	1101111	157	o
16	10	10000	20	[DATA LINK ESCAPE]	64	40	1000000	100	@	112	70	1110000	160	p
17	11	10001	21	[DEVICE CONTROL 1]	65	41	1000001	101	A	113	71	1110001	161	q
18	12	10010	22	[DEVICE CONTROL 2]	66	42	1000010	102	B	114	72	1110010	162	r
19	13	10011	23	[DEVICE CONTROL 3]	67	43	1000011	103	C	115	73	1110011	163	s
20	14	10100	24	[DEVICE CONTROL 4]	68	44	1000100	104	D	116	74	1110100	164	t
21	15	10101	25	[NEGATIVE ACKNOWLEDGE]	69	45	1000101	105	E	117	75	1110101	165	u
22	16	10110	26	[SYNCHRONOUS IDLE]	70	46	1000110	106	F	118	76	1110110	166	v
23	17	10111	27	[END OF TRANS. BLOCK]	71	47	1000111	107	G	119	77	1110111	167	w
24	18	11000	30	[CANCEL]	72	48	1001000	110	H	120	78	1111000	170	x
25	19	11001	31	[END OF MEDIUM]	73	49	1001001	111	I	121	79	1111001	171	y
26	1A	11010	32	[SUBSTITUTE]	74	4A	1001010	112	J	122	7A	1111010	172	z
27	1B	11011	33	[ESCAPE]	75	4B	1001011	113	K	123	7B	1111011	173	{
28	1C	11100	34	[FILE SEPARATOR]	76	4C	1001100	114	L	124	7C	1111100	174	
29	1D	11101	35	[GROUP SEPARATOR]	77	4D	1001101	115	M	125	7D	1111101	175	}
30	1E	11110	36	[RECORD SEPARATOR]	78	4E	1001110	116	N	126	7E	1111110	176	~
31	1F	11111	37	[UNIT SEPARATOR]	79	4F	1001111	117	O	127	7F	1111111	177	[DEL]
32	20	100000	40	[SPACE]	80	50	1010000	120	P					
33	21	100001	41	!	81	51	1010001	121	Q					
34	22	100010	42	"	82	52	1010010	122	R					
35	23	100011	43	#	83	53	1010011	123	S					
36	24	100100	44	\$	84	54	1010100	124	T					
37	25	100101	45	%	85	55	1010101	125	U					
38	26	100110	46	&	86	56	1010110	126	V					
39	27	100111	47	'	87	57	1010111	127	W					
40	28	101000	50	(88	58	1011000	130	X					
41	29	101001	51)	89	59	1011001	131	Y					
42	2A	101010	52	*	90	5A	1011010	132	Z					
43	2B	101011	53	+	91	5B	1011011	133	[
44	2C	101100	54	,	92	5C	1011100	134	\					
45	2D	101101	55	-	93	5D	1011101	135]					
46	2E	101110	56	.	94	5E	1011110	136	^					
47	2F	101111	57	/	95	5F	1011111	137	_					

Google “phred+33 quality score”

ASCII BASE=33 Illumina, Ion Torrent, PacBio and Sanger

Q	P_error	ASCII									
0	1.00000	33 !	11	0.07943	44 ,	22	0.00631	55 7	33	0.00050	66 B
1	0.79433	34 "	12	0.06310	45 -	23	0.00501	56 8	34	0.00040	67 C
2	0.63096	35 #	13	0.05012	46 .	24	0.00398	57 9	35	0.00032	68 D
3	0.50119	36 \$	14	0.03981	47 /	25	0.00316	58 :	36	0.00025	69 E
4	0.39811	37 %	15	0.03162	48 0	26	0.00251	59 ;	37	0.00020	70 F
5	0.31623	38 &	16	0.02512	49 1	27	0.00200	60 <	38	0.00016	71 G
6	0.25119	39 '	17	0.01995	50 2	28	0.00158	61 =	39	0.00013	72 H
7	0.19953	40 (18	0.01585	51 3	29	0.00126	62 >	40	0.00010	73 I
8	0.15849	41)	19	0.01259	52 4	30	0.00100	63 ?	41	0.00008	74 J
9	0.12589	42 *	20	0.01000	53 5	31	0.00079	64 @	42	0.00006	75 K
10	0.10000	43 +	21	0.00794	54 6	32	0.00063	65 A			

```
[jdu282@mcc-login002 Bdor_pop_WGS_SRA_data]$ zcat SRR22045704_pass_1.fastq.gz | head  
@SRR22045704.1.1 1 length=150  
CNCCGGCTTCGAAGCTTCTCCCTATTATAAAAAAAAATTACAAACTAAAATTTTGTCCTCTAGCTGTATCAAACGATAT  
ATTGTGCTACGGTTGGCTTGAATAAGCAAAATAATGTATTGGAACAGACGACGGAGATGGTT  
+SRR22045704.1.1 1 length=150  
F#FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF  
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF  
@SRR22045704.2.1 2 length=150  
TNGCGATCAGTTCCGCATGCCAGAACAGCAAGCAATTGCCACATTCTCGCTGTTCTATTGCTATTAAATGCCGGTTCCA  
CCATCTGACCACCATTACACCCATACTCCGTGAGGATGTGCATTGCTCGCGATAAGAATTG  
+SRR22045704.2.1 2 length=150  
F#FFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFF:FFFFFFFFFFFFFFFFFFFFFFFF  
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFF  
@SRR22045704.3.1 3 length=150  
TCCATTATTAGCAATCTTAGTTCTAAGAAAAACAAGTTGTCTATAAATTATTCAAATTAAAATGAATATAAACATCTTC  
ATTCCGATATGGATTTTCTTATTTAATTAATAAGCAAGTAAGTAATTTATATAATAATA
```

Quality trimming & masking

>Sequence1

16	16	21	9	10	13	14	12	8	8	9	16	24	21	19	19	19	25	25	33	35	35	34	34	34	
34	34	34	34	40	45	45	56	56	56	51	51	40	45	37	37	37	40	40	40	40	40	40	39	39	
39	40	40	40	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	45	51	45	39	39
39	39	39	39	39	39	39	39	39	39	39	39	40	51	51	51	51	51	56	56	56	56	56	56	56	56
40	35	35	35	35	35	39	40	40	45	45	45	45	45	51	56	40	39	39	39	39	39	39	45	45	45
45	40	40	40	39	35	35	40	40	40	40	40	40	38	38	38	39	25	23	18	10	8	9	23	31	
51	51	45	43	43	43	43	43	43	43	43	43	43	43	56	56	56	56	56	56	56	56	56	56	43	43
43	43	43	43	45	45	45	51	51	56	51	51	51	51	51	56	51	45	43	43	56	56	56	56	56	56
56	56	51	51	51	45	45	45	40	40	40	44	42	38	38	40	40	40	51	56	56	56	56	56	56	46
46	51	51	51	51	56	56	56	51	40	45	45	40	40	40	42	42	42	42	45	45	45	42	42	42	42
56	42	42	40	40	34	37	33	40	40	40	44	48	48	48	29	29	29	26	32	29	32	32	32	32	32
33	44	48	56	40	40	40	40	40	40	40	40	40	37	34	34	37	40	40	40	40	37	34	34	48	
40	32	28	25	25	25	34	48	48	48	40	40	32	29	24	25	29	40	40	40	40	40	40	33	33	
37	40	40	40	43	43	42	42	42	44	44	56	56	56	56	40	35	34	33	33	40	40	40	40	40	
40	40	29	29	34	29	29	29	40	29	34	25	27	23	23	21	23	18	20	25	25	25	32	32	32	
32	32	29	18	20	14	16	16	17	17	16	22	20	18	25	19	14	16	15	26	27					

Trimmed & masked sequence

>Sequence1

CCAGAAACTACGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCGGGCTGCAGATCGTC
CGCCAGACTAAAGAAGTCCAAGAGTTGGCTGCCAAAACGCGCTAAAAACGCAAAAAGCGG
CGACCAGTAGANNNNAGGCGAGGCAGGAAGAACAGCCAACCTTTGGGGTTAACGACTATG
TTTCGTCAAGAAAAAAGGGTTCCGACGACCGCACCGACGACCAGATTGGATTACAGTG
GACCGGACCATGGCAGATTCTAGAAGAACGAGGATATAGCTATGTTGGACGTACCTGAA
TCGTTAAAGGAAAAAAATTGTTCCACGCAGACCGCCTCCGCAAAGCCGAATGGACCCAT
TACCACAACAGAAAAGAGAGGCCGCCTCCGCCAGAACGAGATCAACGCCAGAGTTGTGGTCG
ATAAGTTTAGCGTCCGATTATTGCCGGAGTAAGATATTGCAATACCAGGTGCGCATG
GCAAGGATGTGATCCAGACGACACGTGGTACCCGGCTGAAAACTTCAAGAATTAGCGACA
GCCCTTGACGACTCCACAAGAAGTAC

fasta format

>header
[sequence]

>geneXYZ
ATTGCGATAGCTAGCTCGATCGATCG

>geneXYZ
ATTGCGATAGCTAGC
TCGATCGATCG

fasta format

```
[jdu282@login003 6_fastas]$ head OG0008033_73-548_s163_1.renamed.fasta
>fraterculus_Mexico_KST_002 OG0008033_73-548_s163_1 matched_4 trashed_0
ATGTCGATCTTACGATATGCAAGTACAAAACCATTGAACAAATGAAGTATTCAAGTGTAGCGCATGCCGTGGCTTGGCGACTCAAACGTAAAT
ACCACTCAGAATCTCGGGAGCAAATCGAAGAAGAATATGAAAAGTACAAAAACTCGCATTGATTCAATCGCTTACGTACATTGAAAGATCTGGTATC
GATATAGACTTCTTGATAATATCGAGGCAGAAATGAAAAATTCAATTGACTCGCGTCTTCAGGAGCATTAAAGCAATAATCTAAATTGATAGAGAAA
CTGCGGCCACGCAACATGAACGACTCTCAGCCACTGCCAGCACCTGGCTATGTACCGCAAGCTAGTGCGAAGAAGTGCAT
>fraterculus_Mexico_KST_003 OG0008033_73-548_s163_1 matched_1 trashed_1
ATGTCGATCTTACGATATGCAAGTACAAAACCATTGAACAAATGAAGTATTCAAGTGTAGCGCATGCCGTGGCTTGGCGACTCAAACGTAAAT
ACCACTCAGAATCTCGGGAGCAAATCGAAGAAGAATATGAAAAGTACAAAAACTCGCATTGATTCAATCGCTTACGTACATTGAAAGATCTGGTATC
GATATAGACTTCTTGATAATATCGAGGCAGAAATGAAAAATTCAATTGACTCGCGTCTTCAGGAGCATTAAAGCAATAATCTAAATTGATAGAGAAA
CTGCGGCCACGCAACATGAACGACTCTCAGCCACTGCCAGCACCTGGCTATGTACCGCAAGCTAGTGCGAAGAAGTGCAT
>fraterculus_Brazil_KST_004 OG0008033_73-548_s163_1 matched_1 trashed_0
ATGTCGATCTTACGATATGCAAGTACAAAACCATTGAACAAATGAAGTATTCAAGTGTAGCGCATGCCGTGGCTTGGCGACTCAAACGTAAAT
ACCACTCAGAATCTCGGGAGCAAATCGAAGAAGAATATGAAAAGTACAAAAACTCGCATTGATTCAATCGCTTACGTACATTGAAAGATCTGGTATC
GATATAGACTTCTTGATAATATCGAGGCAGAAATGAAAAATTCAATTGACTCGCGTCTTCAGGAGCATTAAAGCAATAATCTAAATTGATAGAGAAA
CTGCGGCCACGCAACATGAACGACTCTCAGCCACTGCCAGCACCTGGCTATGTACCGCAAGCTAGTGCGAAGAAGTGCAT
>fraterculus_Brazil_Vacaria_KST_006 OG0008033_73-548_s163_1 matched_1 trashed_0
ATGTCGATCTTACGATATGCAAGTACAAAACCATTGAACAAATGAAGTGTCAAGTGTAGCGCATGCCGTGGCTTGGCGACTCAAACGTAAATA
AATACCACTCAGAATCTCGGGAGCAAATCGAAGAAGAATATGAAAAGTACAAAAACTCGCATTGATTCAATCGCTTACGTACATTGAAAGATCTGGT
ATCGATATAGACTTCTTGATAATATCGAGGCAGAAATGAAAAATTCAATTGACTCGCGTCTTCAGGAGCATTAAAGCAATAATCTAAATTGATAGAG
AAACTGCGCCTACGCAGCATGAGAGACTCTCAGCCACTGCCAGCACCTGGCTATGTACCGCAAGCTAGTGCGAAGAAGTGCAT
>fraterculus_Peru_Libertad_KST_016 OG0008033_73-548_s163_1 matched_1 trashed_0
ATGTCGATCTTACGATATGCAAGTACAAAACCATTGAACAAATGAAGTATTCAAGTGTAGCGCATGCCGTGGCTTGGCGACTCAAACGTAAAT
ACCACTCAGAATCTCGGGAGCAAATCGAAGAAGAATATGAAAAGTACAAAAACTCGCATTGATTCAATCGCTTACGTACATTGAAAGATCTGGTATC
GATATAGACTTCTTGATAATATCGAGGCAGAAATGAAAAATTCAATTGACTCGCGTCTTCAGGAGCATTAAAGCAATAATCTAAATTGATAGAGAAA
CTGCGGCCACGCAACATGAACGAGACTCTCAGCCACTGCCAGCACCTGGCTATGTACCGCAAGCTAGTGCGAAGAAGTGCAT
```