Genômica Computacional

Introdução e processamento de dados de sequenciamento

Professor: Ricardo A. Vialle

CS31 - Genômica Computacional [14/07-17/07 - 10h00-12h00]

14 de Julho de 2025

Cronograma

Data	Tema
14-Jul	Introdução e processamento de dados de sequenciamento (teórico-prática)
15-Jul	Montagem de genomas (teórico-prática)
16-Jul	Anotação de genomas (teórico-prática)
17-Jul	Analise de variabilidade genética (teórico-prática)

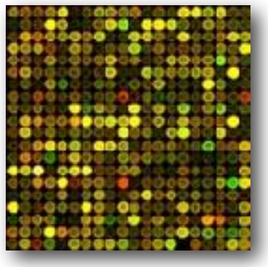
Avaliação

Nota final será determinada de acordo com presença

Genomics: shaped by technology

Sanger DNA sequencing

1977-1990s



DNA Microarrays

Since mid-1990s

2nd-generation DNA sequencing

Since ~2007

3rd-generation & single-molecule DNA sequencing

Since ~2010

These provide very high-resolution snapshots of the world of nucleic acids (not just DNA)

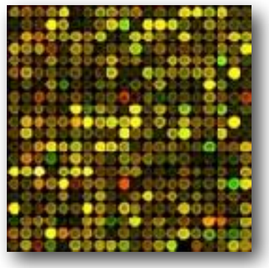
Slide from: Ben Langmead Lab
https://www.cs.jhu.edu/~langmea/resources/lecture_notes/01_genomics_comp_genomics_v2.pdf

		Illumina						
	Sanger	**MiSeq**	**NextSeq**	**HiSeq**	**NovaSeq**	**Ion Torrent**	**Pacific Biosciences**	**Oxford Nanopore**
Throughput range per run (Gb)	*c.* 0.0005	10–15	10–120	1000–1800	2000–6000	1–15	0.5–10	0.1–1
Read length	Up to 1 kb	300	150	150	250	200–400	up to 60 kb	up to 100 kb
Read type	SR	SR, PE	SR, PE	SR, PE	SR, PE	SR	SR	SR
Error rate (%)	0.001–1	0.8	0.8	0.2	0.2–0.8	1–2	13	5–40
Error type	Substitutions	Substitutions	Substitutions	Substitutions	Substitutions	Indels, homopolymers	Indels	Indels, deletions
Advantages	Read accuracy and length	Read length	Throughput	Throughput, low error rate	High throughput	Speed, read length	Speed, read length	Read length, portability

Genomics technology

Sanger DNA sequencing

1977-1990s



DNA Microarrays

Since mid-1990s

2nd-generation DNA sequencing

Since ~2007

3rd-generation & single-molecule DNA sequencing

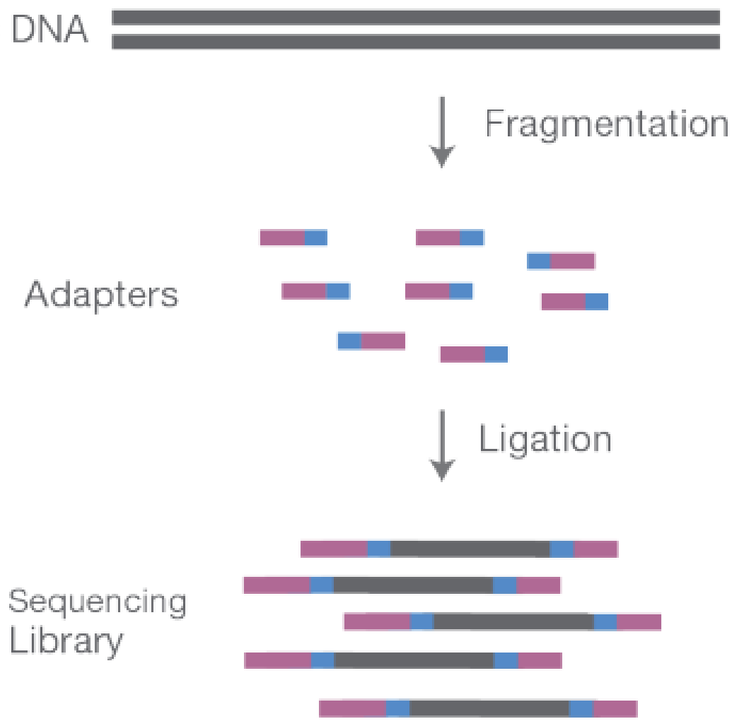
Since ~2010

Solexa

illumina®

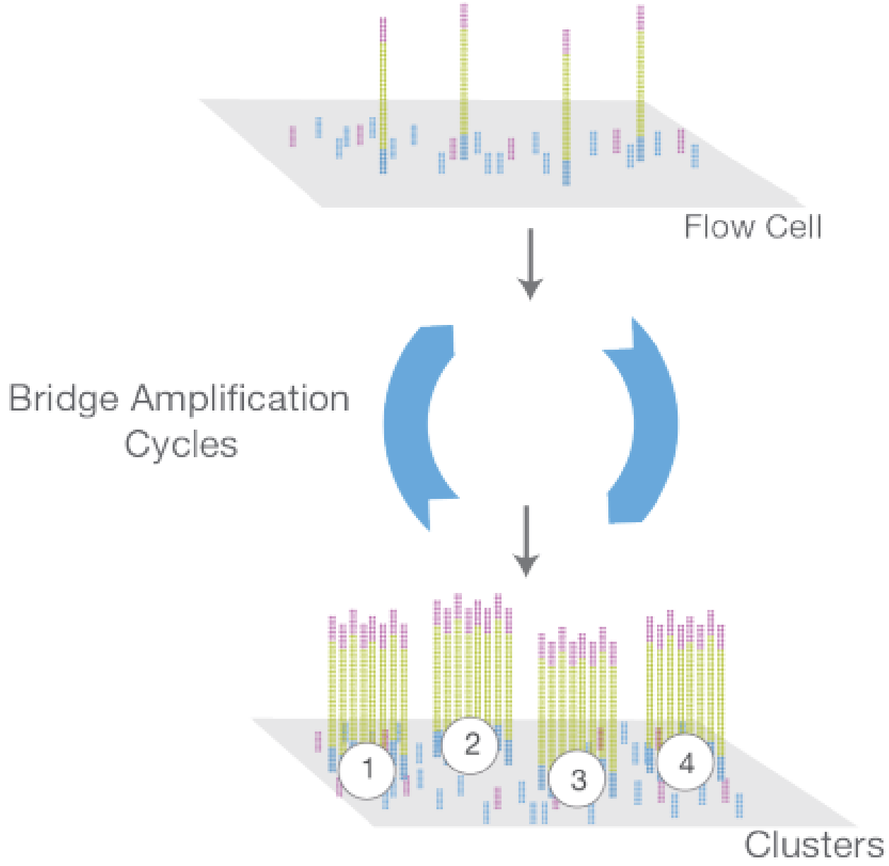
Slide from: Ben Langmead Lab
https://www.cs.jhu.edu/~langmea/resources/lecture_notes/01_genomics_comp_genomics_v2.pdf

A. Library Preparation



NGS library is prepared by fragmenting a gDNA sample and ligating specialized adapters to both fragment ends.

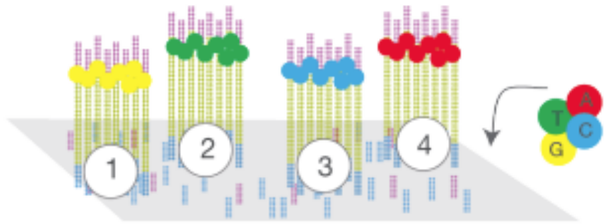
B. Cluster Amplification



Library is loaded into a flow cell and the fragments are hybridized to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.

https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

C. Sequencing



Sequencing Cycles

Digital Image

Data is exported to an output file

```
Cluster 1 > Read 1: GAGT...
Cluster 2 > Read 2: TTGA...
Cluster 3 > Read 3: CTAG...
Cluster 4 > Read 4: ATAC...
```

Text File

Sequencing reagents, including fluorescently labeled nucleotides, are added and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated "n" times to create a read length of "n" bases.

D. Alignment and Data Anaylsis

Reads

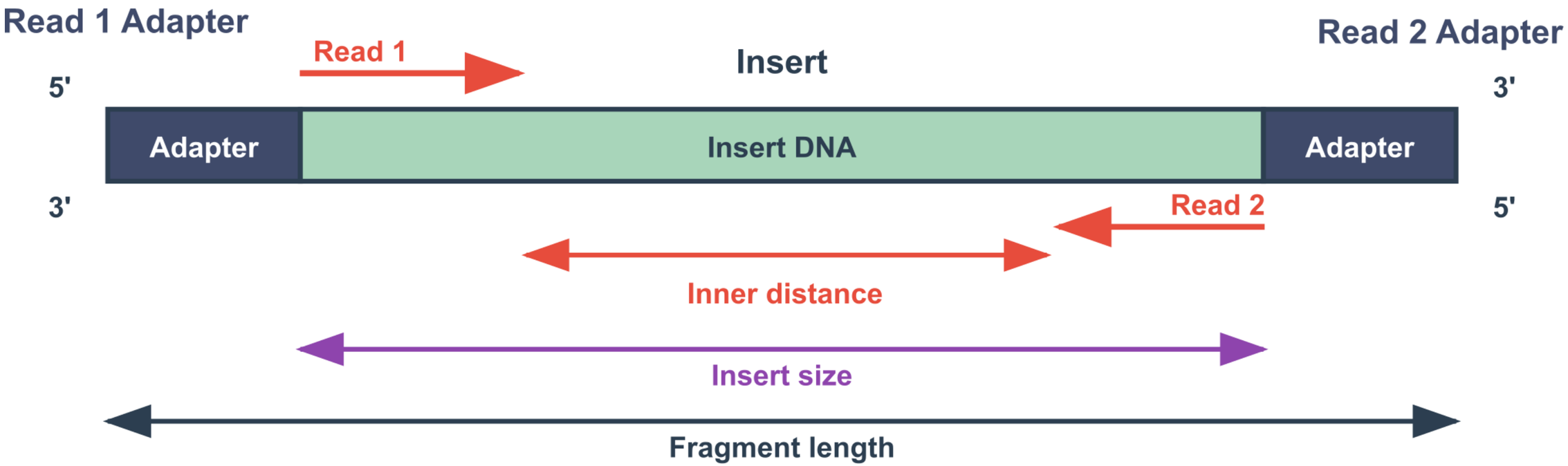
```
ATGGCATTGCAATTTGACAT
  TGGCATTGCAATTTG
AGATGGTATTG
 GATGGCATTGCAA
    GCATTGCAATTTGAC
ATGGCATTGCAATT
AGATGGCATTGCAATTTG
```

Reference Genome

```
AGATGGTATTGCAATTTGACAT
```

Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.

https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf



- Fragment length: Total length including adapters
- Insert size: Length of DNA insert between adapters
- Inner distance: Distance between paired reads

A read in FASTQ format

Name	`@ERR194146.1 HSQ1008:141:D0CC8ACXX:3:1308:20201:3607 2:Y:18:ATCACG`
Sequence	`ACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAAT`
(ignore)	+
Base qualities	`?@@FFBFFDDHHBCEAFGEGIIDHGH@GDHHHGEHID@C?GGDG@FHIGGH@FHBEG:G`

Always starts with “@”
ERR194146.1HSQ1008:141:D0CC8ACXX:3 – Machine, Run, Flowcell, Lane
1308:20201:3607 – Tile, X-pos, Y-pos
2:Y:18:ATCACG – Direction, Filtered?, Control bits, Index/Sample

Base qualities

Bases and qualities line up:

```
AGCTCTGGTGACCCATGGGCAGCTGCTAGGGA
||||||||||||||||||||||||||||||||
IHHHHHHHHHHHHHHGCGC5FEFFFGHHHHHH
```

Base quality is ASCII-encoded version of $Q = -10 \log_{10} p$

Base quality

Probability that base call is incorrect

$Q = 10 \rightarrow$ 1 in 10 chance call is incorrect

$Q = 20 \rightarrow$ 1 in 100

$Q = 30 \rightarrow$ 1 in 1,000

Slide from: Ben Langmead Lab
https://www.cs.jhu.edu/~langmea/resources/lecture_notes/01_genomics_comp_genomics_v2.pdf

ASCII TABLE

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

ASCII TABLE

These character don’t print

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

Value	Character	Value	Character	Value	Character	Value	Character	Value	Character	Value	Character
0	Null	22	Synchronous Idle	44	,	66	B	88	X	110	n
1	Start of Heading	23	End of Transmission Block	45	-	67	C	89	Y	111	o
2	Start of Text	24	Cancel	46	.	68	D	90	Z	112	p
3	End of Text	25	End of Medium	47	/	69	E	91	[	113	q
4	End of Transmission	26	Substitute	48	0	70	F	92	\	114	r
5	Enquiry	27	Escape	49	1	71	G	93	]	115	s
6	Acknowledgement	28	File Separator	50	2	72	H	94	^	116	t
7	Bell (Causes an alert sound)	29	Group Separator	51	3	73	I	95	_	117	u
8	Backspace	30	Record Separator	52	4	74	J	96	@	118	v
9	Horizontal Tab	31	Unit Separator	53	5	75	K	97	a	119	w
10	Line Feed	32	space	54	6	76	L	98	b	120	x
11	Vertical Tab	33	!	55	7	77	M	99	c	121	y
12	Form Feed	34	"	56	8	78	N	100	d	122	z
13	Carriage Return	35	#	57	9	79	O	101	e	123	{

Base 33 (typical)

Bases and qualities line up:

```
AGCTCTGGTGACCCATGGGCAGCTGCTAGGGA
|||||||||||||||||||||||||||||||||
IHHHHHHHHHHHHHGCGC5FEFFFGHHHHHH
```

```
Value - base = Phred
  73  -  33  = 40
```

$Q = -10 \log_{10} p$

Phred	Error Probability	Confidence
0	1 / 1	0%
10	1 / 10	90%
20	1 / 100	99%
30	1 / 1000	99.9%
40	1 / 10000	99.99%
50	1 / 100000	99.999%
60	1 / 1000000	99.9999%

FASTQ

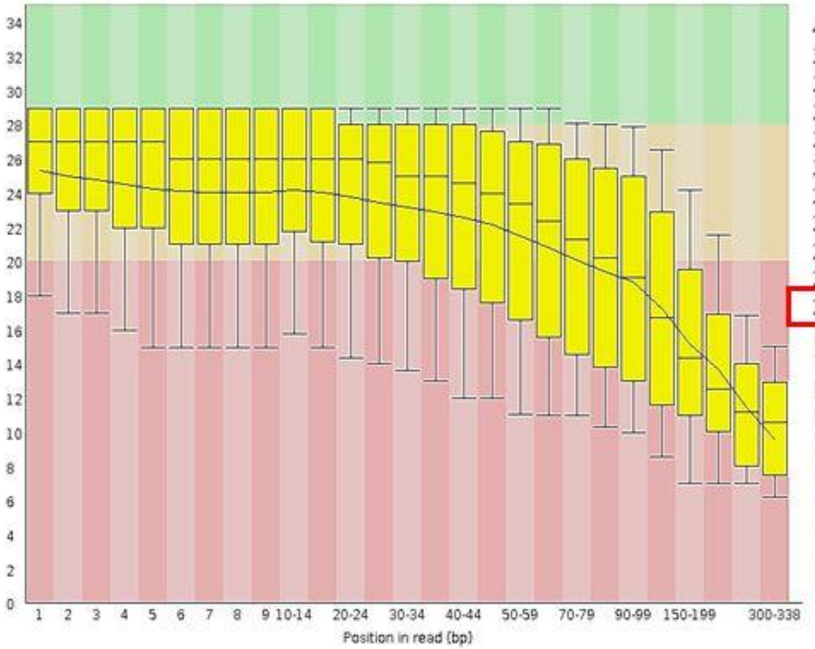
Read 1	Name
	Sequence
	(placeholder)
	Base qualities
Read 2	Name
	Sequence
	(placeholder)
	Base qualities
Read 3	Name
	Sequence
	(placeholder)
	Base qualities
Read 4	Name
	Sequence
	(placeholder)
	Base qualities
Read 5	Name
	Sequence
	(placeholder)
	Base qualities

Sample_S1_L001_R1_001.fastq.gz

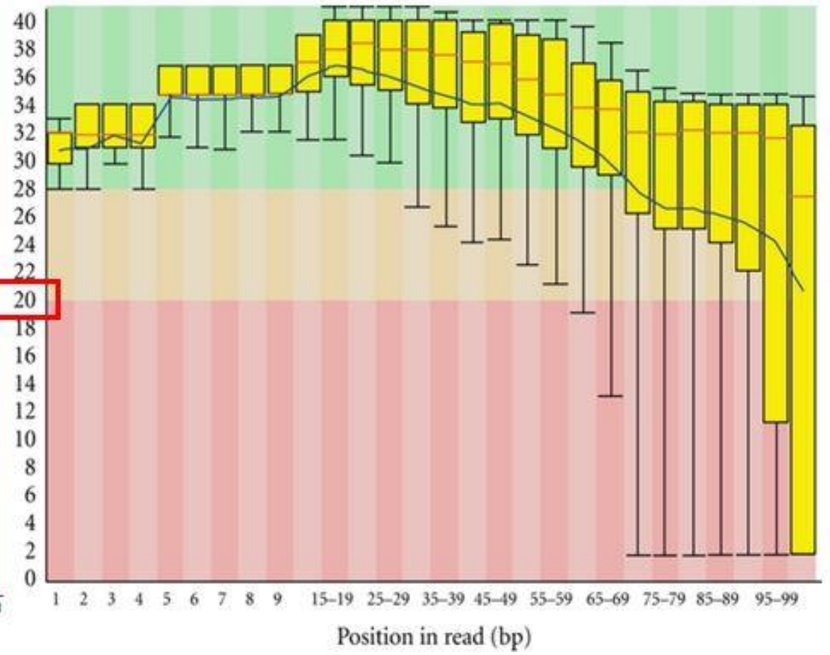
Sample: Sample name
S1: Sample number
L001: Lane number
R1: Read direction (If starts with “I”, Index read direction)
001: File number (always 001 on modern systems)

Comparison in sequencing quality

Ion Torrent PGM



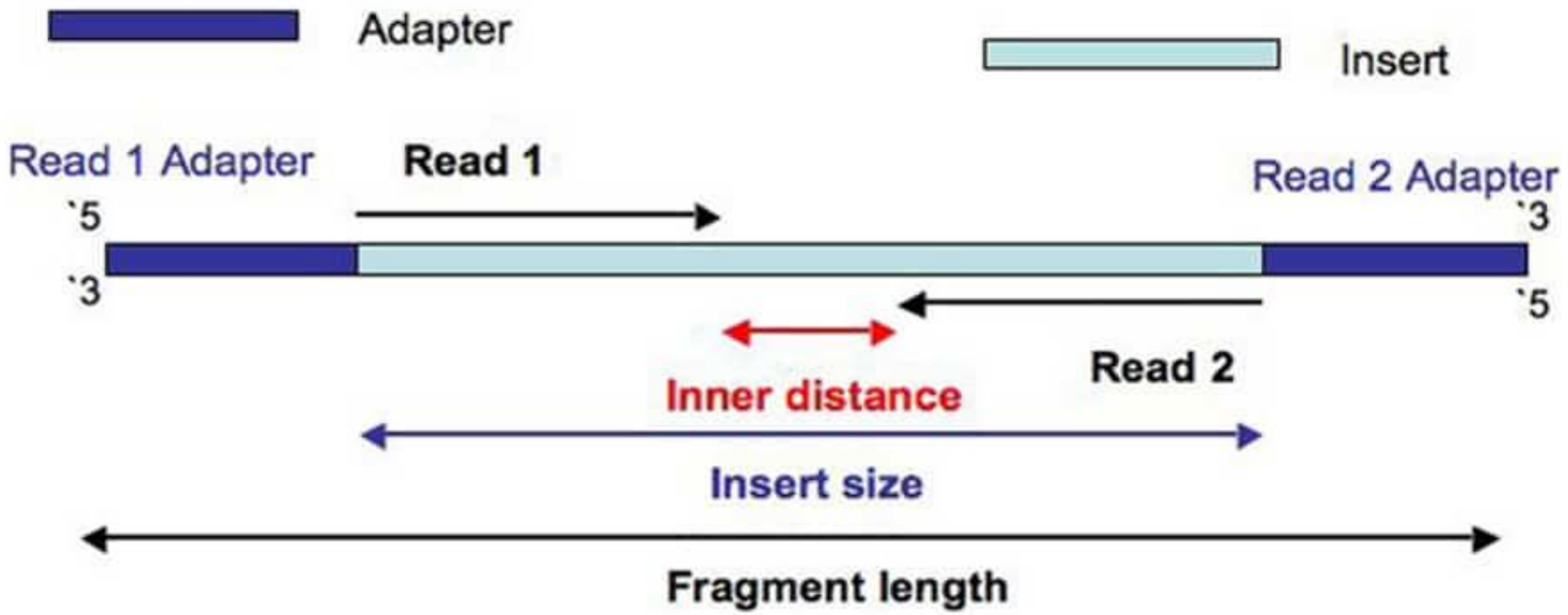
Illumina Hiseq 2000



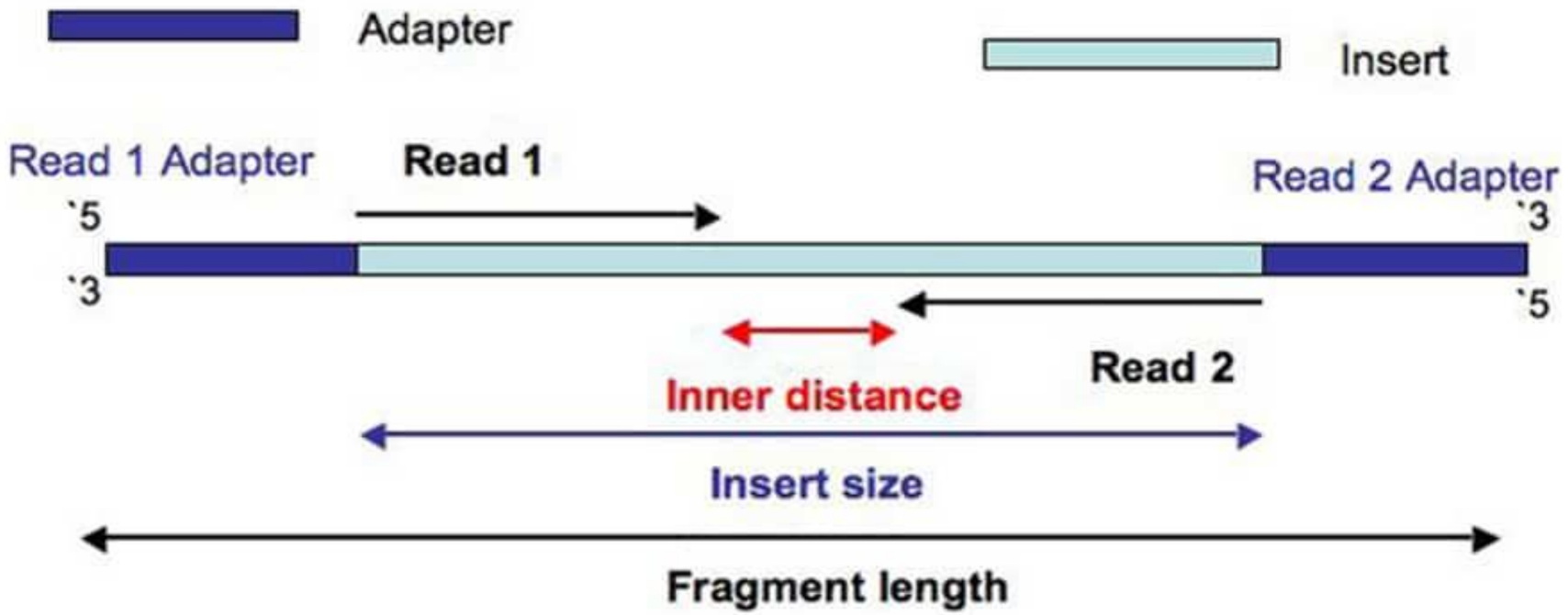
Per base sequence quality of samples generated by FASTQC. The yellow box show the base-calling quality scores across all sequencing reads. The blue line indicates the mean quality score. Q20=99% accuracy. Q30=99.9% accuracy…

https://www.hadriengourle.com/tutorials/qc/

- Trimming
  - Barcode & adapter sequences



- Trimming
  - Barcode & adapter sequences



  - Poor quality sequence at the starts/ends of reads

Which trimming threshold? Examples

RNAseq

- Gentle trimming
- Q > 5 should be enough
- Too aggressive trimming => losing part of the dataset

SNP calling

- You need to be sure of the bases
- Q > 20

MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in genetics*, *5*, 13.

https://www.hadriengourle.com/tutorials/qc/

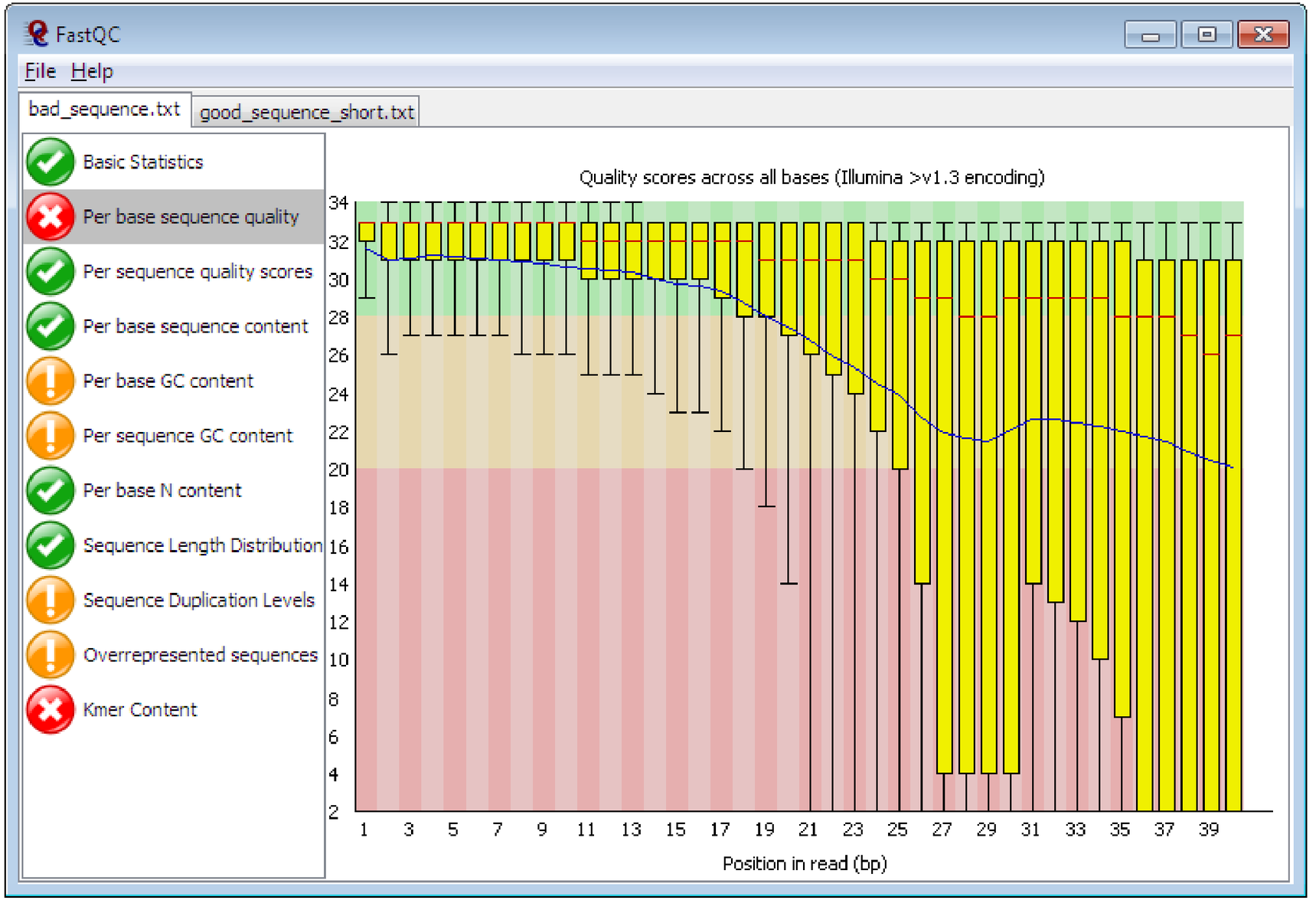
Adapter trimmer [Scythe]

Uses the quality information in a FASTQ entry and a prior to decide whether a 3' substring is adapter.

Very basically, it takes a naïve Bayesian approach to classifying 3'-end contaminants only. Because these are the most poor quality bases and most likely to be contaminated (especially as reads get longer and longer), Scythe is designed to specifically remove these contaminants.

Low quality trimmer [Sickle]

Sickle is a sliding window quality trimmer, designed to be used after Scythe. Unlike *cutadapt* and other tools, this pipeline remove adapter contaminants before quality trimming, as removing poor quality bases throws away any useful information that could be used in identifying a 3'-end adapter contaminant.



FastQC

[See each report detail here]

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

MultiQC
v1.3

- General Stats
- FastQC
  - Sequence Quality Histograms
  - Per Sequence Quality Scores
  - Per Base Sequence Content
  - Per Sequence GC Content
  - Per Base N Content
  - Sequence Length Distribution
  - Sequence Duplication Levels
  - Overrepresented sequences
  - Adapter Content

MultiQC

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2017-11-19, 21:42 based on data in: `/Users/hadrien/Documents/workspace/ar`

General Statistics

Copy table | Configure Columns | Plot | Showing 4/4 rows and 4/5 columns.

Sample Name	% Dups	% GC	Length	M Seqs
SRR957824_500K_R1	16.2%	49%	150 bp	0.5
SRR957824_500K_R2	7.2%	50%	150 bp	0.5
SRR957824_trimmed_R1	2.9%	51%	142 bp	0.4
SRR957824_trimmed_R2	2.7%	51%	136 bp	0.4

FastQC

FastQC is a quality control tool for high throughput sequence data, written by Simon Andrews at the Babraham Institute in Cambridge.

Sequence Quality Histograms 3 1

Toolbox

https://www.hadriengourle.com/tutorials/qc/

Improving quality: toolbox

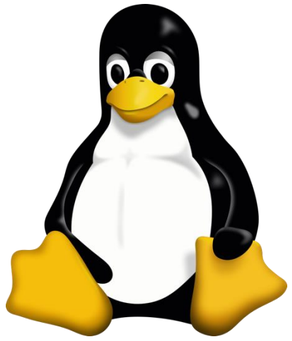
- Trimmomatic
- Cutadapt
- Scythe
- Sickle
- Atropos

Fastp:

https://github.com/OpenGene/fastp

https://www.hadriengourle.com/tutorials/qc/

https://www.hostinger.com/tutorials/linux-commands



1. ls command

The **ls** command lists files and directories in your system. Here's the syntax:

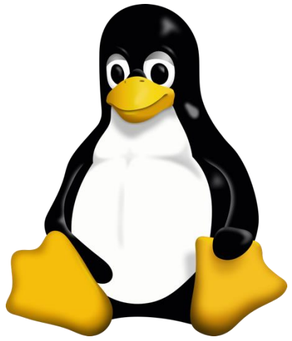
```
ls [/directory/folder/path]
```

```
root@srv:/# ls /directory/folder/path
file1.txt
```

If you remove the path, the **ls** command will show the current working directory's content. You can modify the command using these options:

- **-R** – lists all the files in the subdirectories.
- **-a** – shows all files, including hidden ones.
- **-lh** – converts sizes to readable formats, such as **MB**, **GB**, and **TB**.

https://www.hostinger.com/tutorials/linux-commands



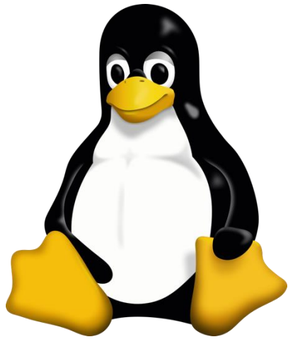
2. pwd command

The **pwd** command prints your current working directory's path, like **/home/directory/path**. Here's the command syntax:

```
pwd [option]
```

It supports two options. The **-L** or **--logical** option prints environment variable content, including symbolic links. Meanwhile, **-P** or **--physical** outputs the current directory's actual path.

```
root@srv:/directory/folder/path# pwd
/directory/folder/path
```



3. cd command

Use the **cd** command to navigate the Linux files and directories. To use it, run this syntax with sudo privileges:

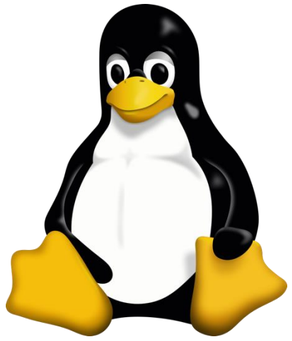
```
cd /directory/folder/path
```

```
root@srv:/# cd /directory/folder/path
root@srv:/directory/folder/path#
```

Depending on your current location, it requires either the full path or the directory name. For example, omit **/username** from **/username/directory/folder** if you are already within it.

Omitting the arguments will take you to the home folder. Here are some navigation shortcuts:

- **cd ~[username]** – goes to another user's home directory.
- **cd ..** – moves one directory up.
- **cd-** – switches to the previous directory.



4. mkdir command

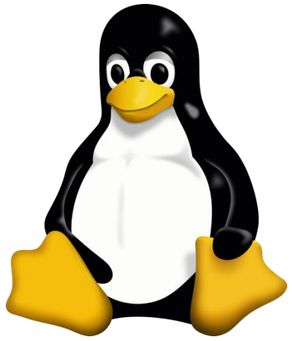
Use the **mkdir** command to create one or multiple directories and set their permissions. Ensure you are authorized to make a new folder in the parent directory. Here's the basic syntax:

```
mkdir [option] [directory_name]
```

To create a folder within a directory, use the path as the command parameter. For example, **mkdir music/songs** will create a **songs** folder inside **music**. Here are several common **mkdir** command options:

- **-p** – creates a directory between two existing folders. For example, **mkdir -p Music/2023/Songs** creates a new **2023** directory.
- **-m** – sets the folder permissions. For instance, enter **mkdir -m777 directory** to create a directory with read, write, and execute permissions for all users.
- **-v** – prints a message for each created directory.

```
root@srv:/# mkdir -v new-folder
mkdir: created directory 'new-folder'
```



6. rm command

Use the rm command to permanently delete files within a directory. Here's the general syntax:

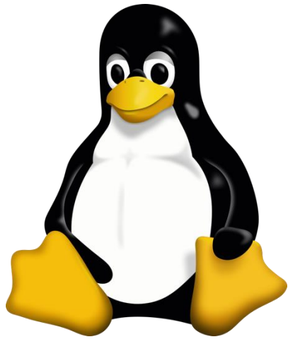
rm [filename1] [filename2] [filename3]

Adjust the number of files in the command according to your needs. If you encounter an error, ensure you have the **write** permission in the directory.

To modify the command, add the following options:

- **-i** – prompts a confirmation before deletion.
- **-f** – allows file removal without a confirmation.
- **-r** – deletes files and directories recursively.

Warning! Use the **rm** command with caution since deletion is irreversible. Avoid using the **-r** and **-f** options since they may wipe all your files. Always add the **-i** option to avoid accidental deletion.



7. cp command

Use the **cp** command to copy files or directories, including their content, from your current location to another. It has various use cases, such as:

- Copying one file from the current directory to another folder. Specify the file name and target path:

```
cp filename.txt /home/username/Documents
```

- Duplicating multiple files to a directory. Enter the file names and the destination path:

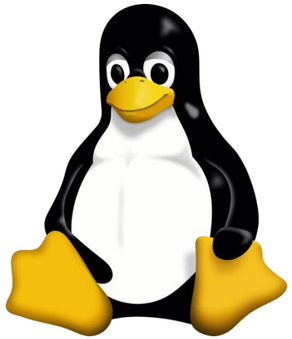
```
cp filename1.txt filename2.txt filename3.txt /home/username/Documents
```

- Copying a file's content to another within the same directory. Enter the source and the destination file:

```
cp filename1.txt filename2.txt
```

- Duplicating an entire directory. Pass the **-R** flag followed by the source and destination directory:

```
cp -R /home/username/Documents /home/username/Documents_backup
```



8. mv command

Use the **mv** command to move or rename files and directories. To move items, enter the file name followed by the destination directory:

```
mv filename.txt /home/username/Documents
```

Meanwhile, use the following syntax to **rename a file in Linux** with the **mv** command:

```
mv old_filename.txt new_filename.txt
```

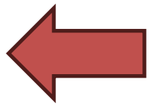
https://github.com/RushAlz/IAMSPE-CS31-Genomica_Computacional

CS31 - Genômica Computacional

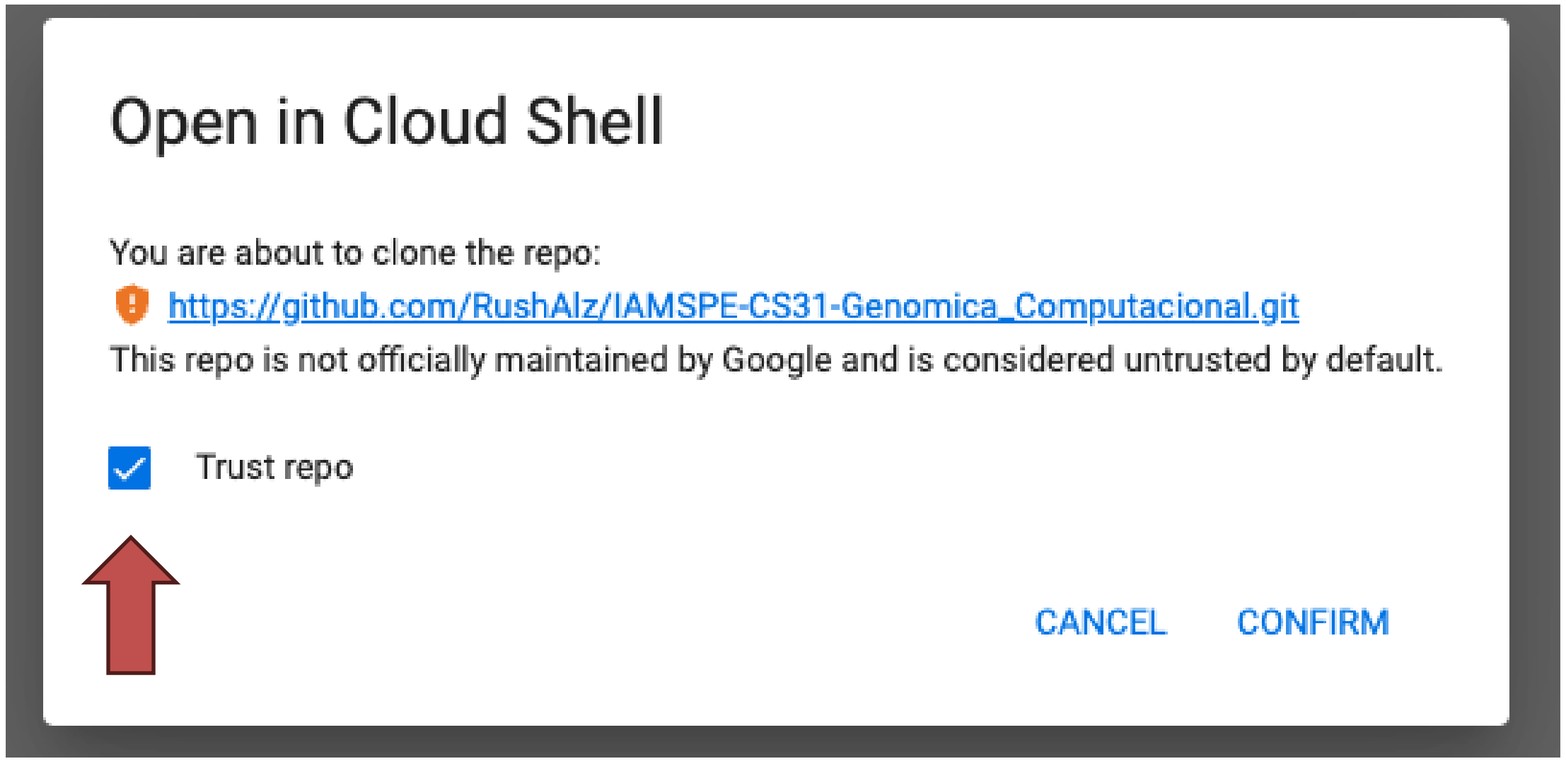
Esse repositório contem materiais de aula para a disciplina de Genômica Computacional da IAMSPE.

Para facilitar execução dos tutoriais, utilizaremos o Google Cloud Shell.

Aula	Data	Tema	Slides	Tutoriais
1	2025-07-14	Introdução e processamento de dados de sequenciamento	Slides	Fastq Quality-Control (QC) tutorial
2	2025-07-15	Montagem de genomas		
3	2025-07-16	Anotacao de genomas		
4	2025-07-17	Mapeando variantes		

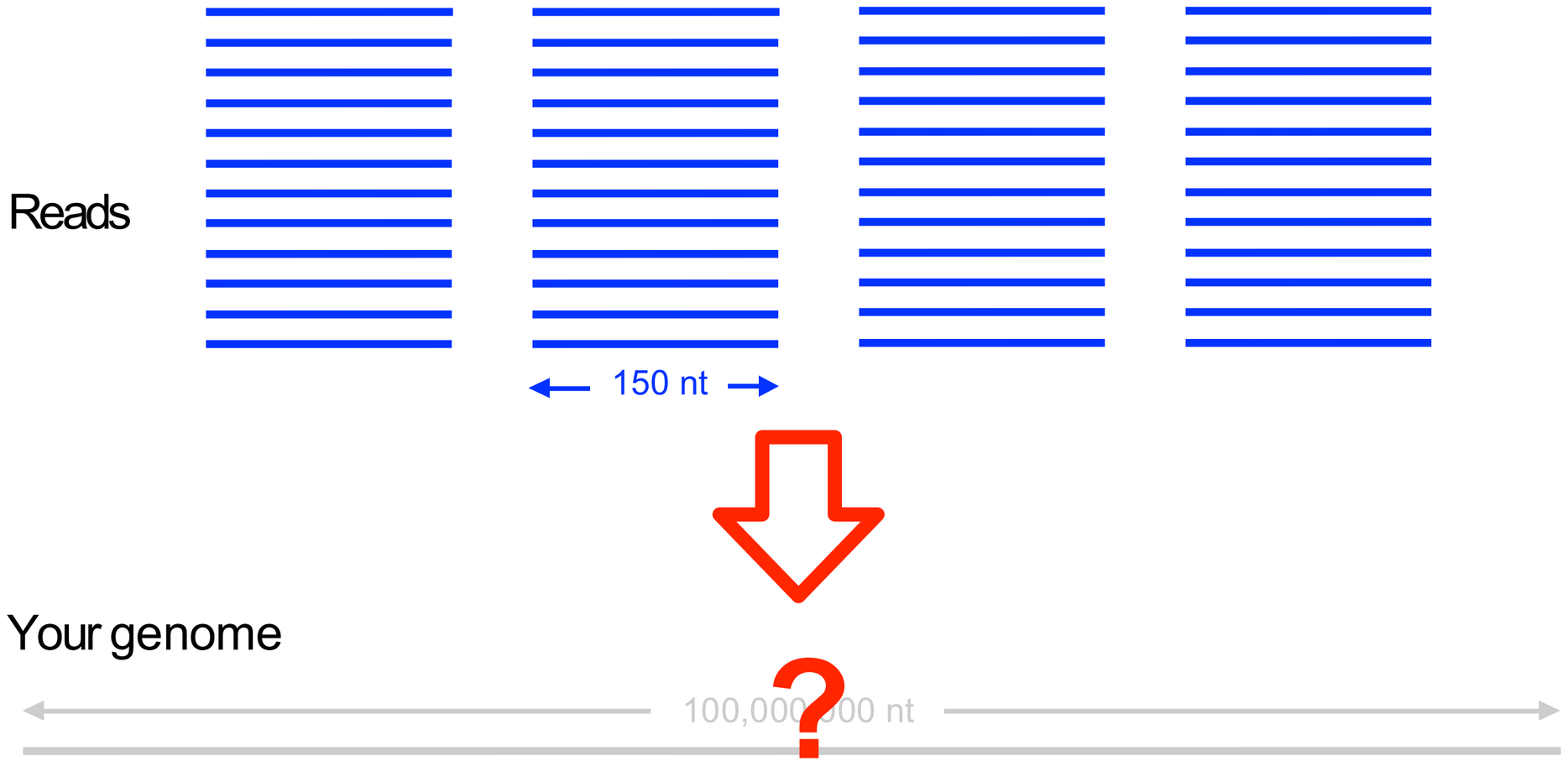


https://github.com/RushAlz/IAMSPE-CS31-Genomica_Computacional



Proxima aula…

Montagem de genomas (teórico-prática)



https://www.cs.jhu.edu/~langmea/resources/lecture_notes/01_genomics_comp_genomics_v2.pdf