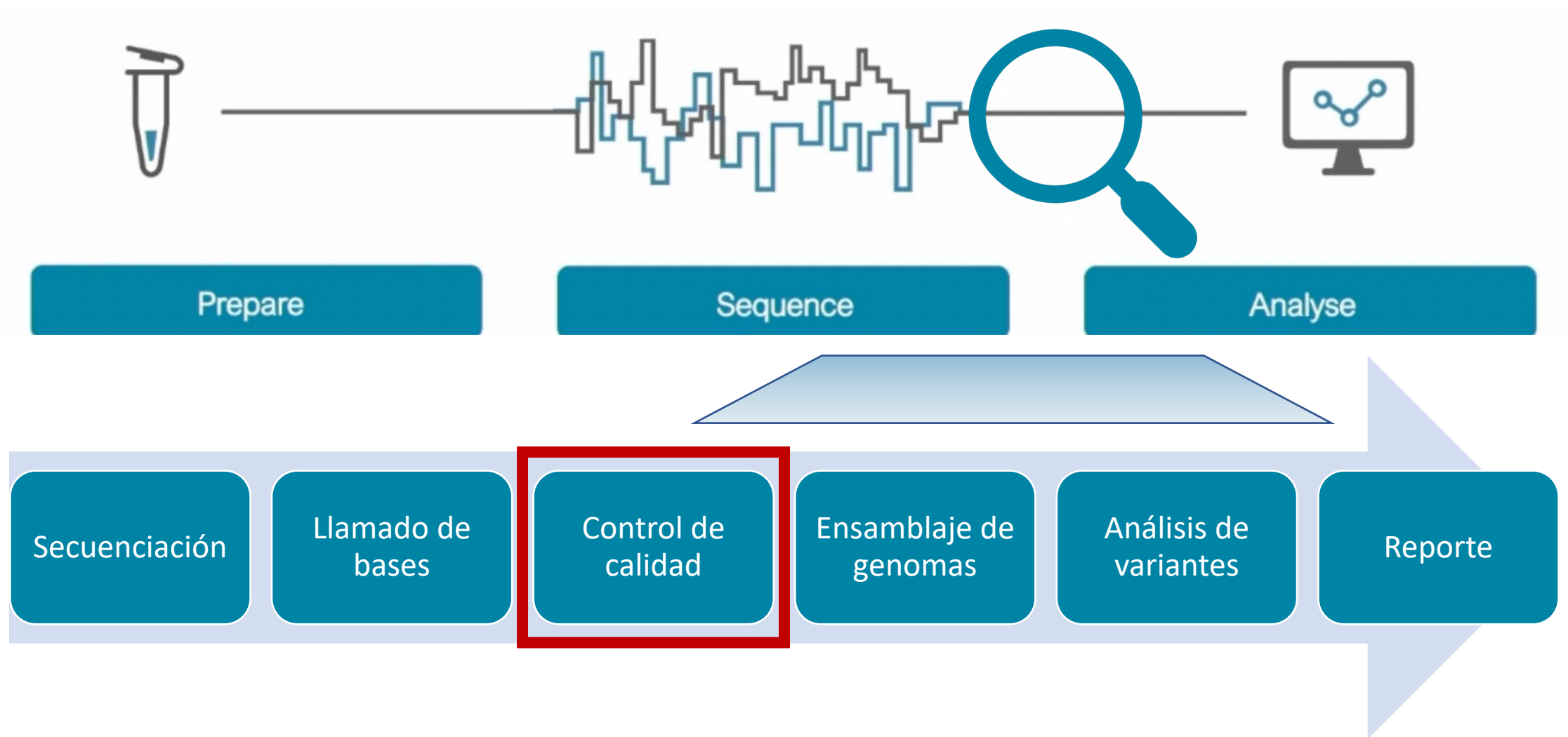


Control de calidad de lecturas generadas por ONT

Daniel Bautista – Universidad del Magdalena – Santa Marta, Colombia
2021

Pasos Procesamiento Datos Genómicos



Control de Calidad

- ¿Qué certeza tenemos que el resultado de nuestra secuenciación corresponde a la composición nucleotídica de nuestra muestra original?
- Instrumentos de secuenciación no realizan una observación directa de las bases:
 - Illumina, PacBio: secuenciación por síntesis, registro óptico.
 - Nanopore: paso por nanoporo, alteraciones corriente.

Phred Score

- Asignarle una probabilidad de qué tan preciso fue la secuenciación.
- Hacerlo por cada base secuenciada.
- Tiene en cuenta las limitaciones del instrumento.
- Empíricamente calibrado comparando secuencias obtenidas por llamado de bases con secuencias conocidas.

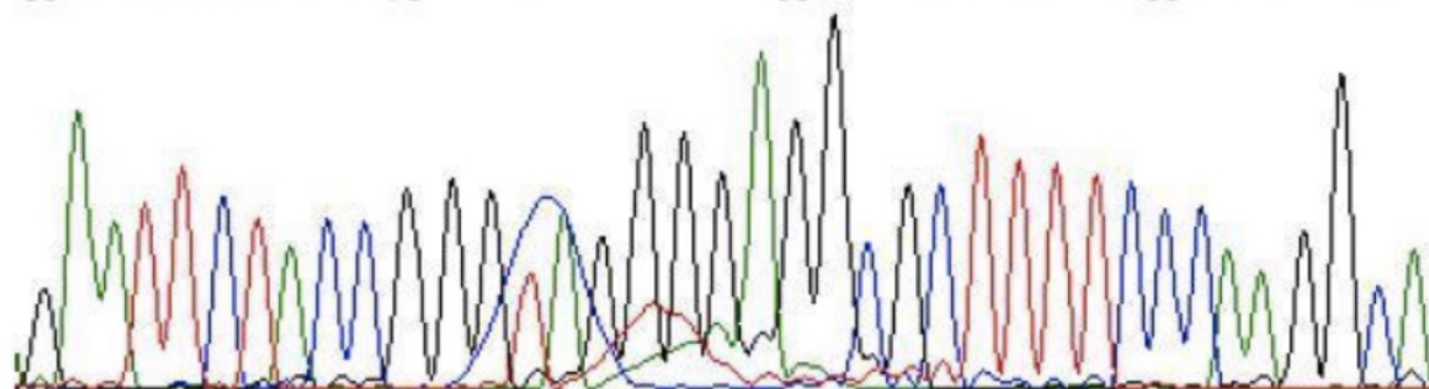
Quality Score	Error Probability
Q40	0.0001 (1 in 10,000)
Q30	0.001 (1 in 1,000)
Q20	0.01 (1 in 100)
Q10	0.1 (1 in 10)

An example of a base that has been given a very high Phred score of 50, indicating that there is 99.999% probability that this base has been correctly assigned.

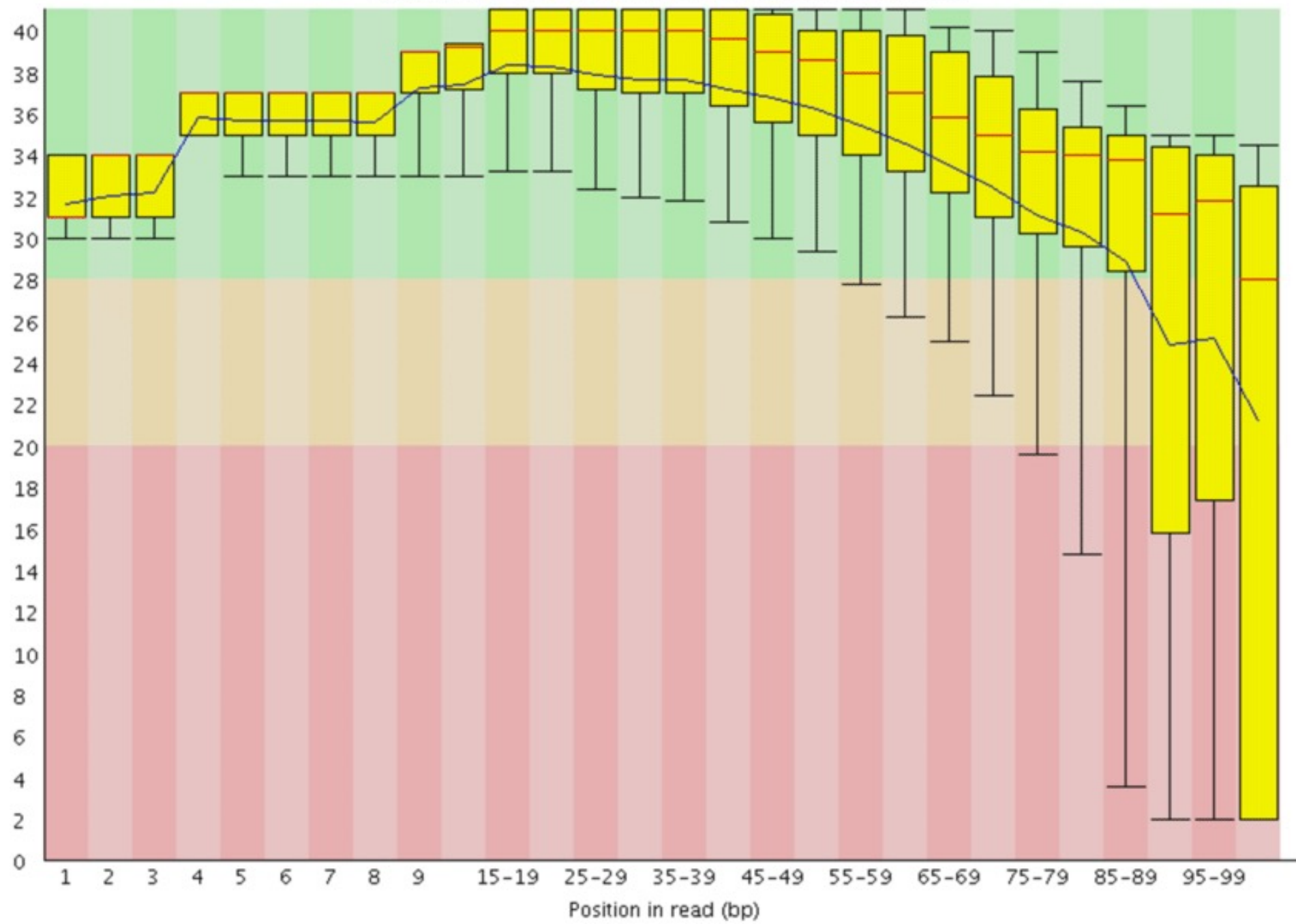
An example of a base that has been given a Phred score of 10, indicating that there is only a 90% probability that this base has been correctly assigned.

An example of a base for which no Phred score could be calculated,, since the sequencer could not determine which base was present (therefore, an 'N' was designated in the sequence).

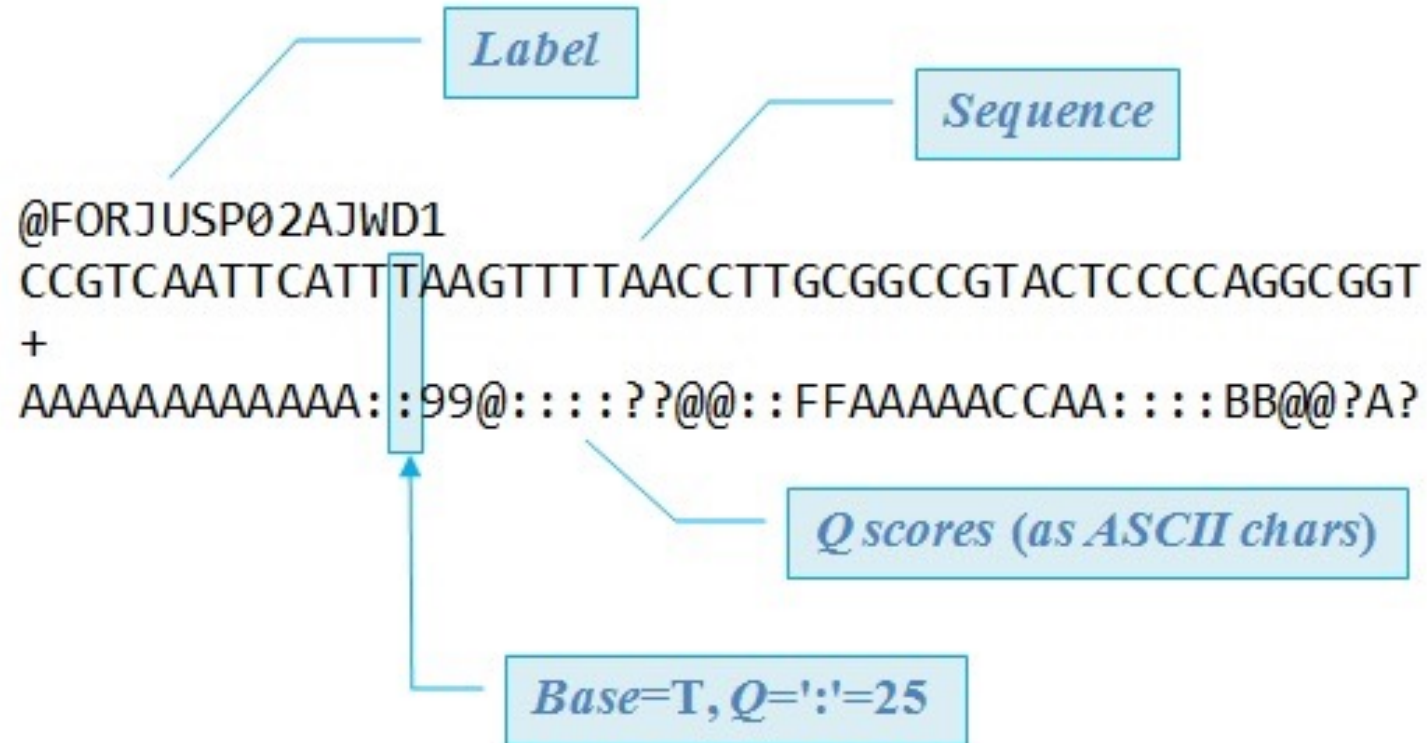
Phred score=20 →



Quality scores across all bases (Sanger / Illumina 1.9 encoding)



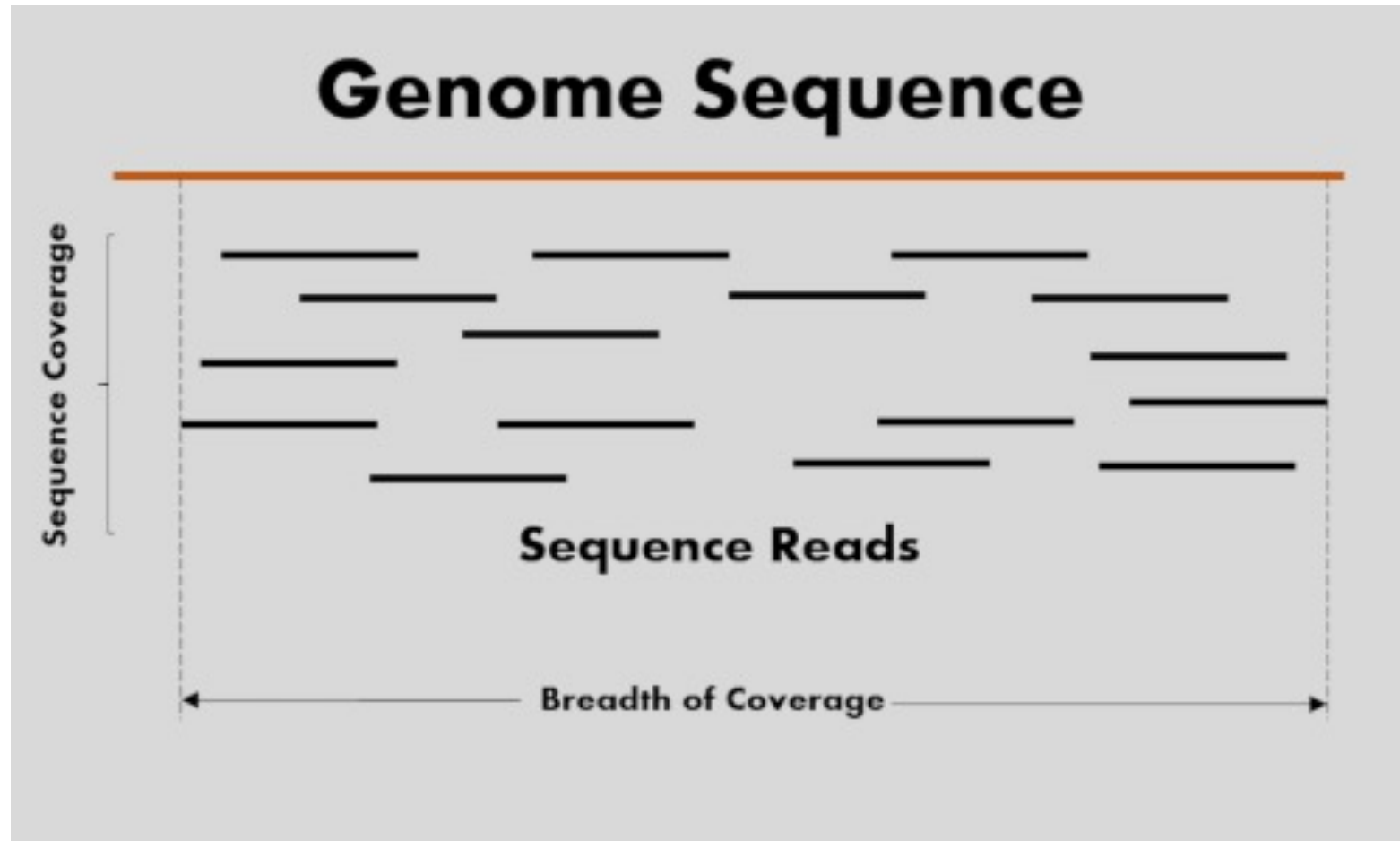
fastq



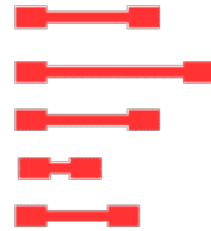
Vamos a visualizar nuestros resultados

- Diferente calidad y longitud.
- Tener en cuenta el kit.
- Lecturas muy largas a veces no muy buenas.

Tener en cuenta



(A)



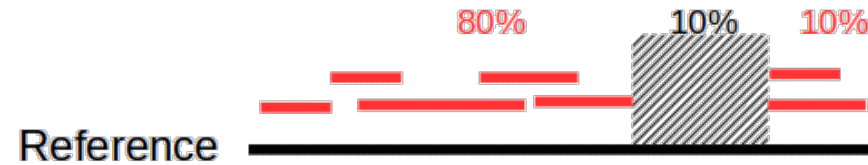
Reads



Reference

$$C = \frac{\begin{array}{l} \text{\# sequenced bases}^1 \\ (= \text{\# bases of all mapped reads}) \end{array}}{\text{\# bases of reference}}$$

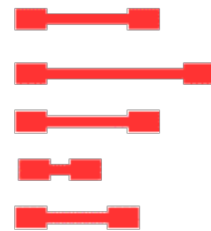
(B)



Reference

$$C = \frac{\text{\# area covered by reads}}{\text{\# reference area}}$$

(C)



Reads for mapping

Sequencing depth= total read number

