



UNIVERSIDAD DE
COSTA RICA

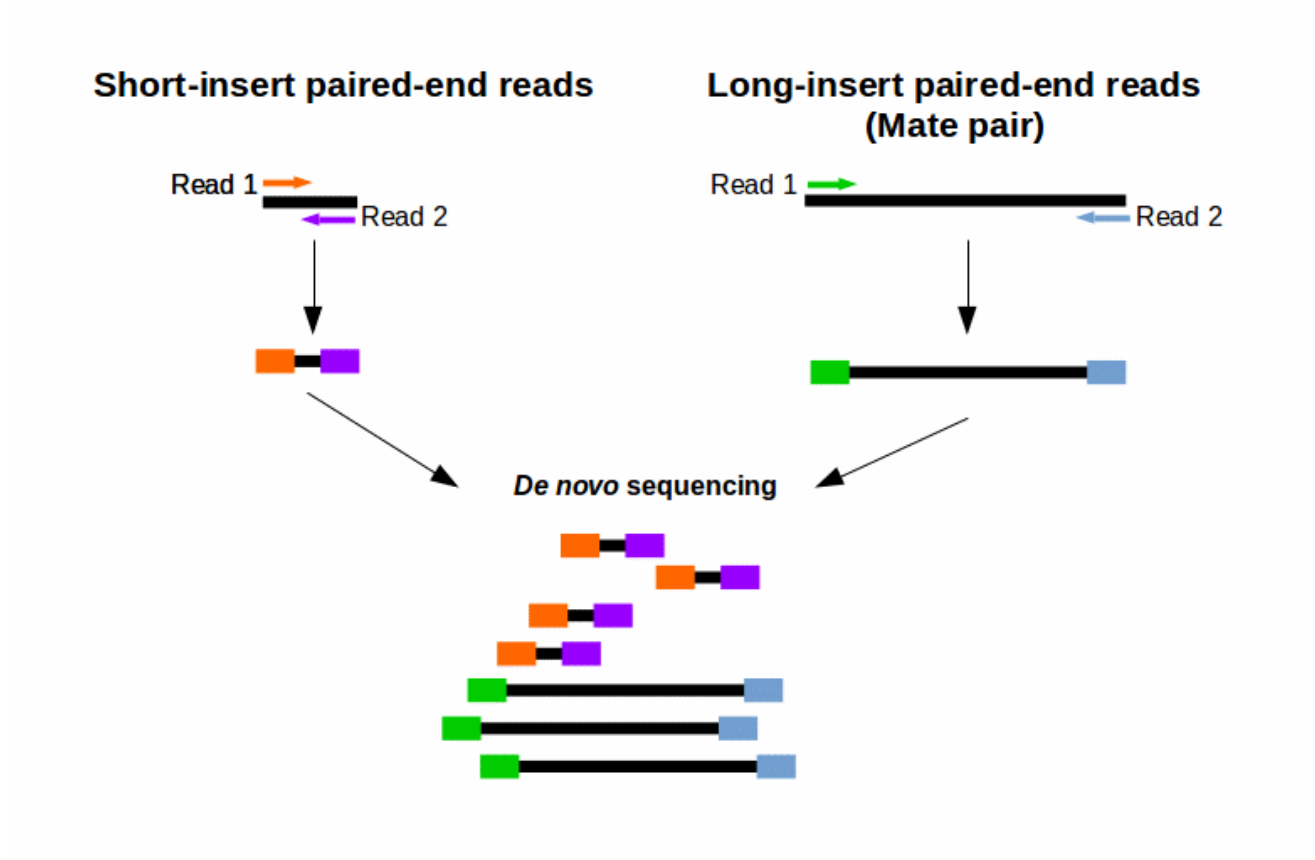
Ensamblaje de genomas

Curso de Genómica de procariontes

Bradd Mendoza

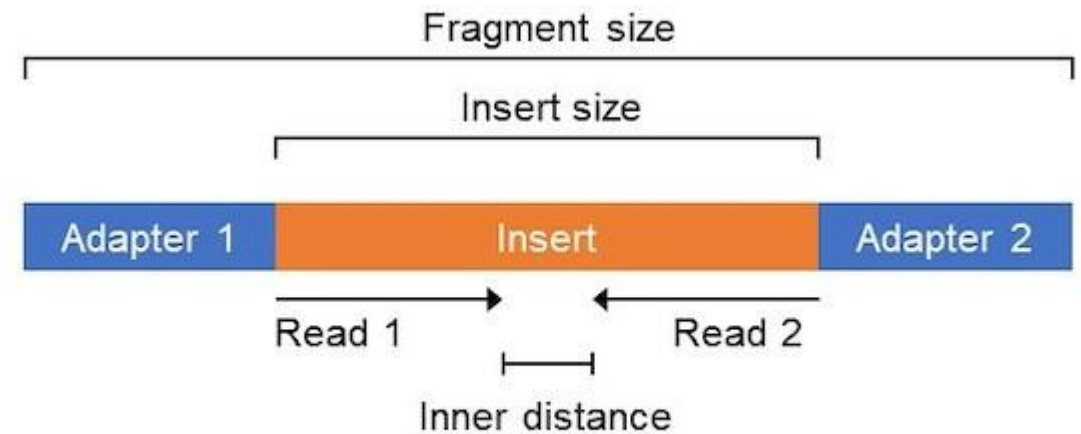
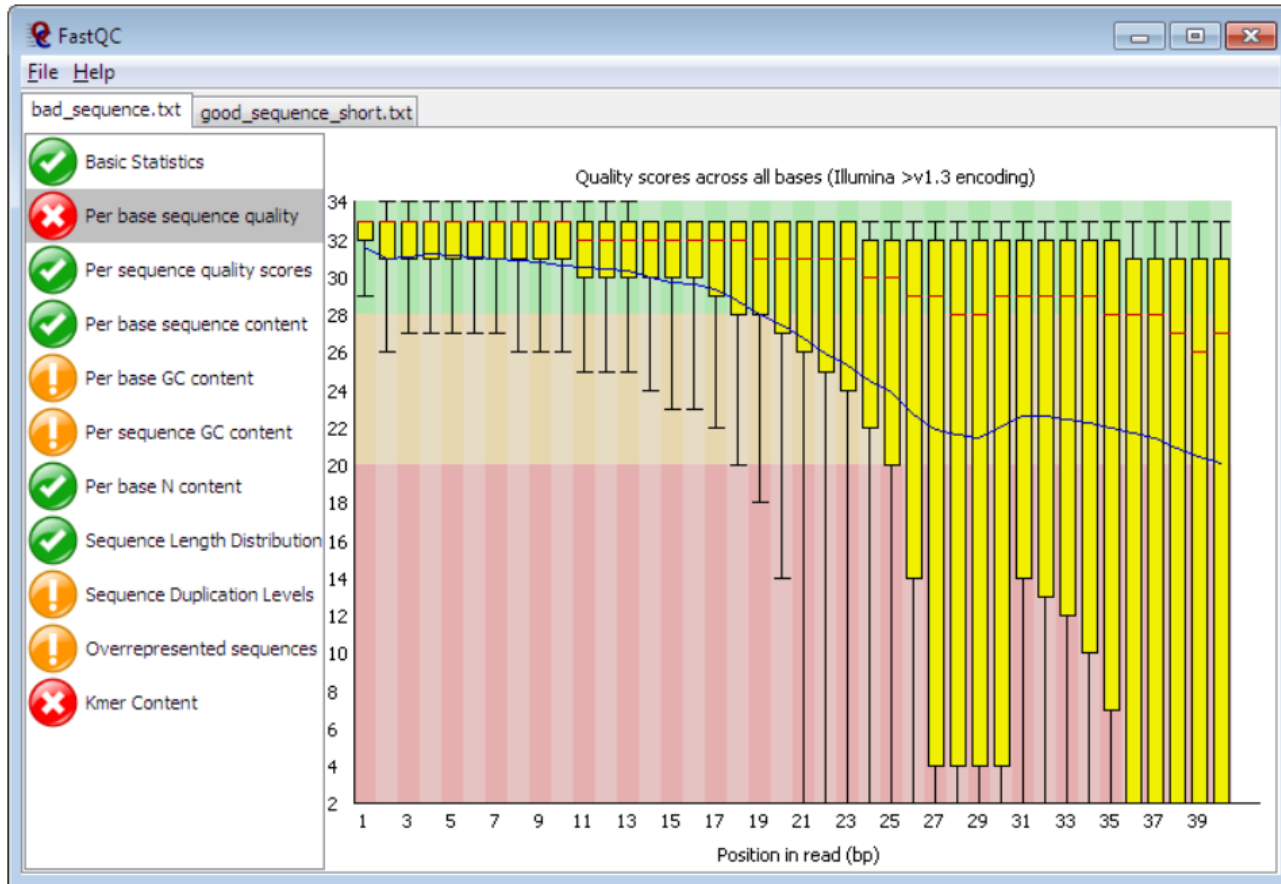


Reads

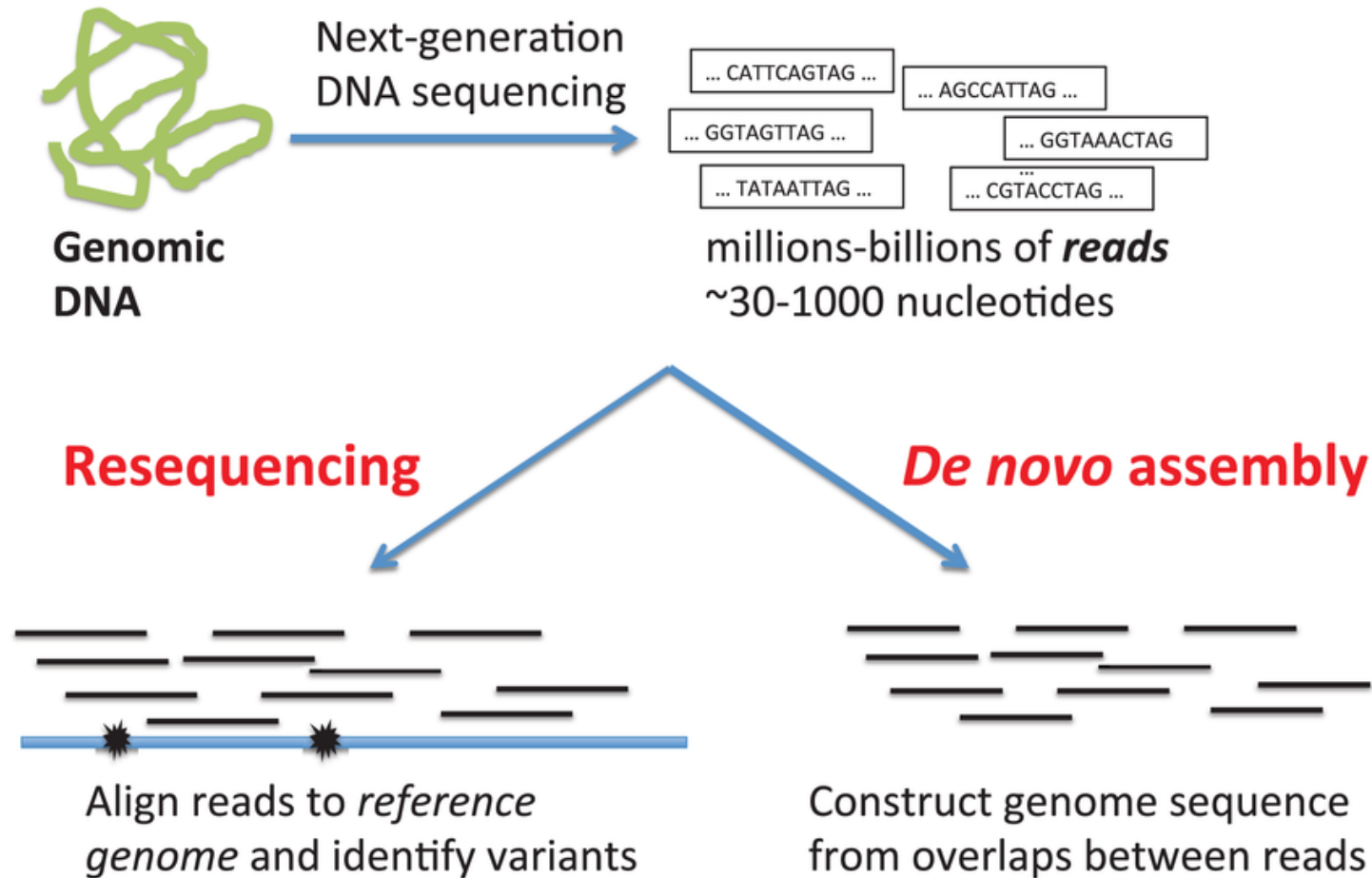




Control de calidad de los reads

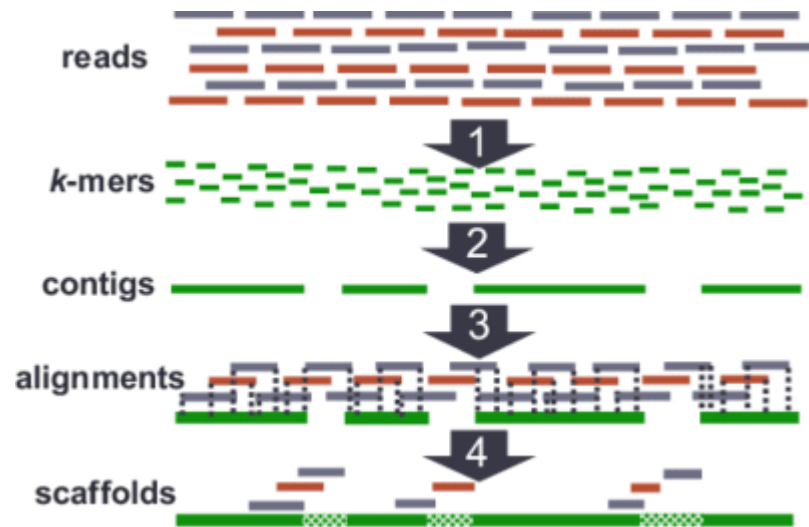


Ensamblaje





De novo assembly



A. Short read to k -mers ($k=4$)

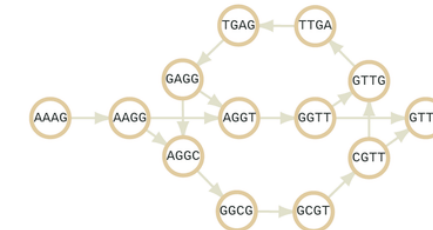
AAAGGCGTTGAGGTT

AAAG
AAGG
AGGC
GGCG
GCGT
CGTT
GTTG
TTGA
TGAG
GAGG
AGGT
GGTT

B. Eulerian de Bruijn graph



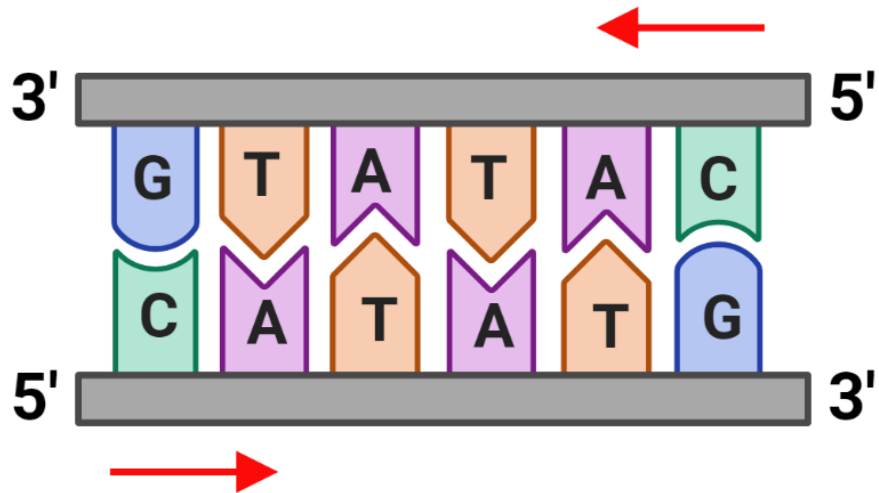
C. Hamiltonian de Bruijn graph





Entendiendo los k-mers

- K-mers siempre impares?
 - Secuencias palindrómicas



TTGACACTTACCGA

Read

TTGACACTTACC
TGACACTTACCG
GACACTTACCGA

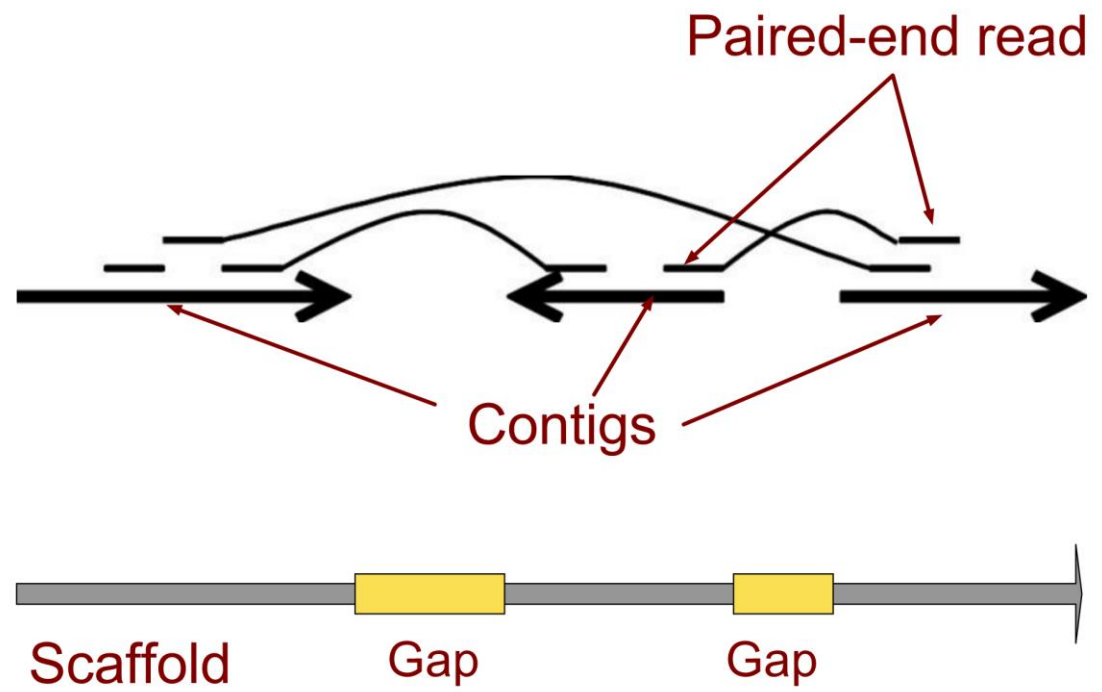
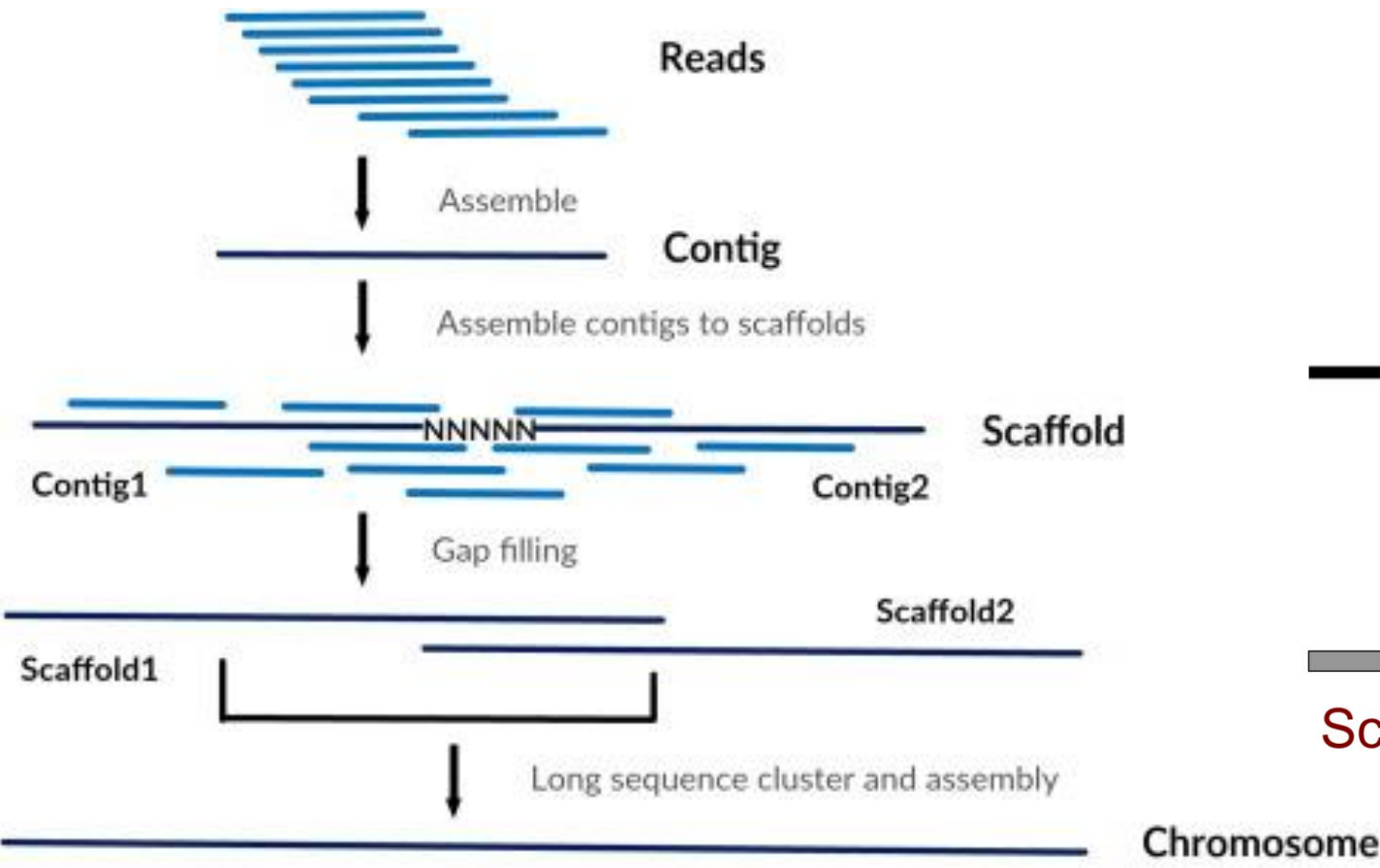
k-mers for k=12

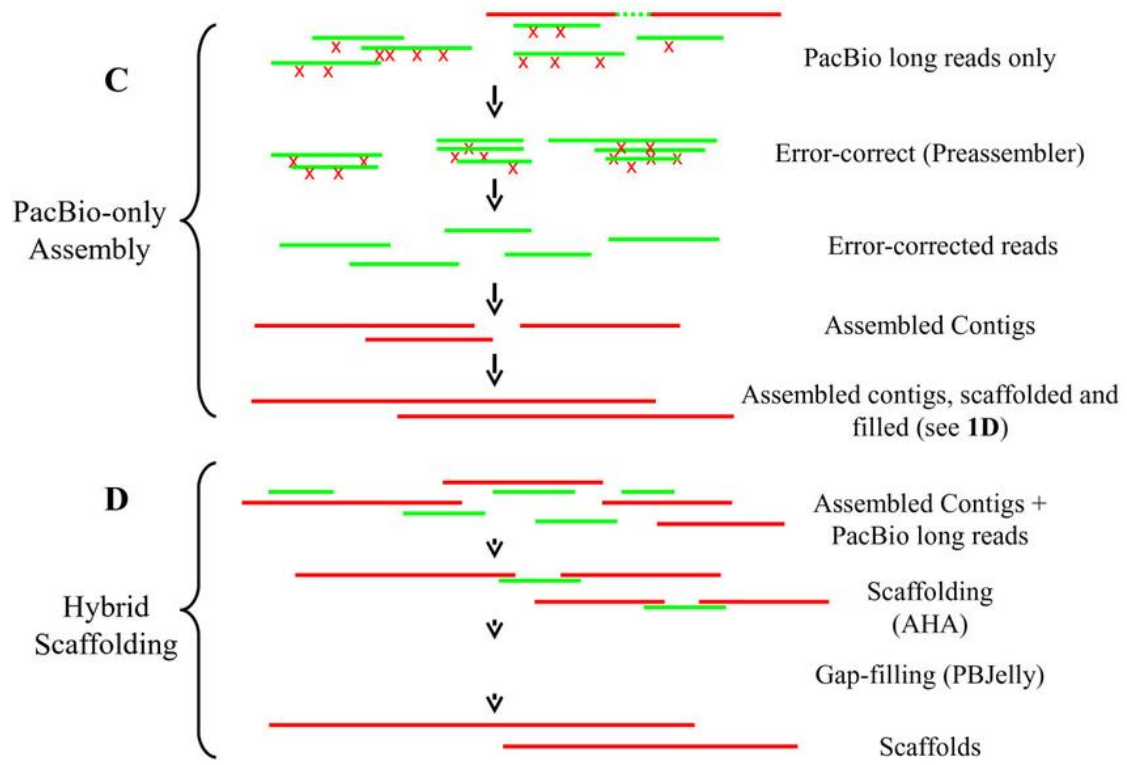
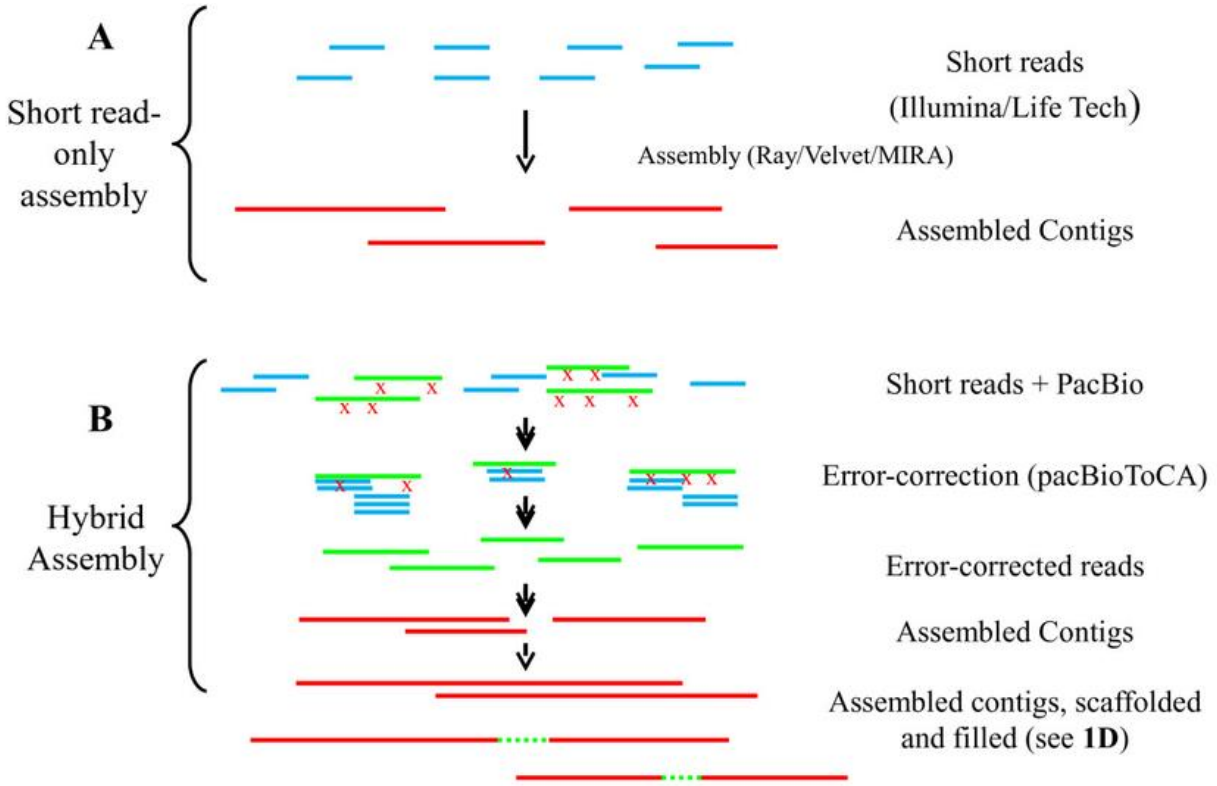
Menor probabilidad de ensamblaje pero mayor confianza de los contigs

TTGAC
TGACA
GACAC
ACACT
CACTT
ACTTA
CTTAC
TTACC
TACCG
ACCGA

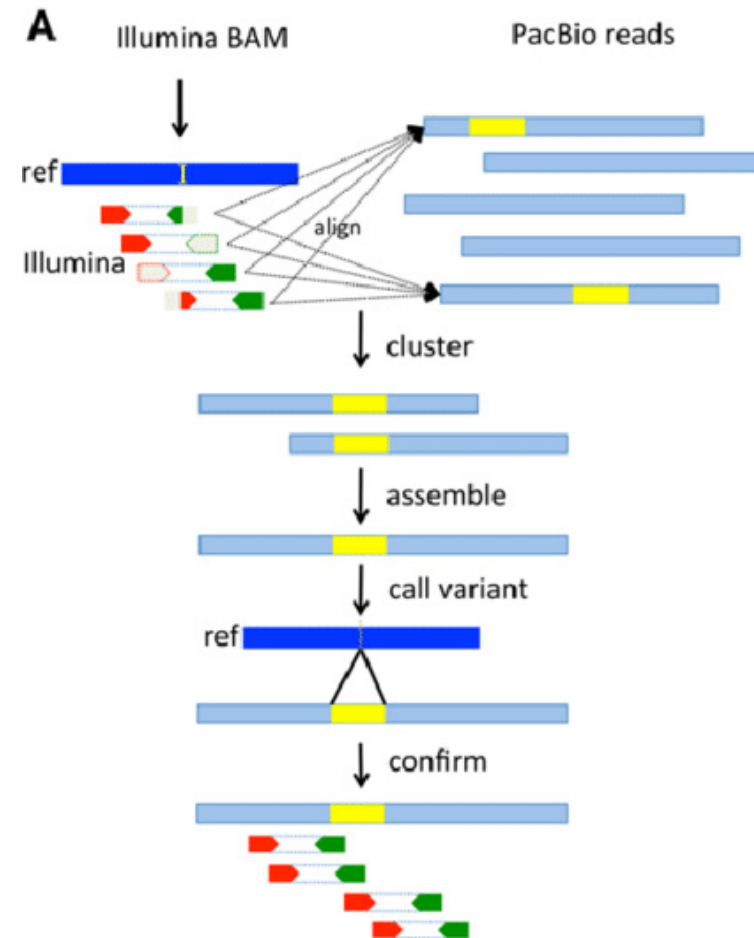
k-mers for k=5

Contigs más largos pero menos fiables



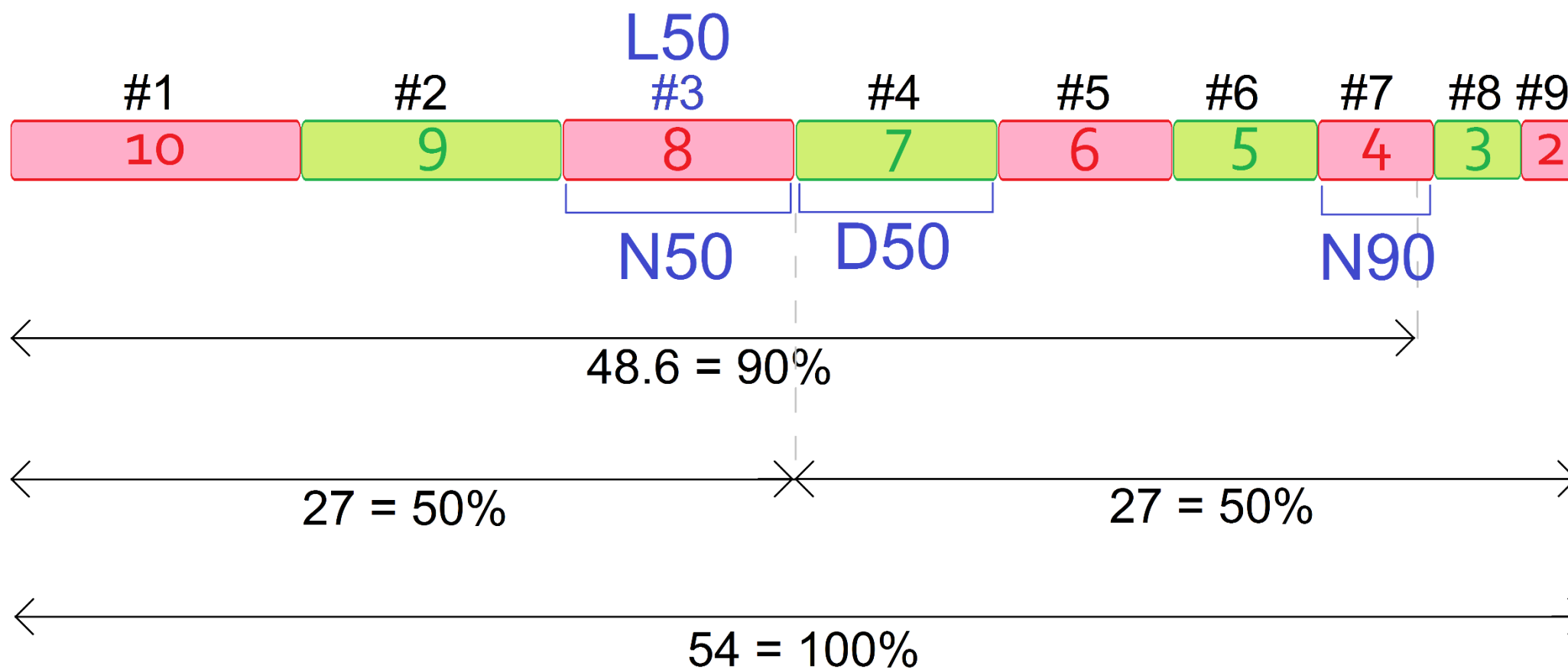


Ensamblaje híbrido



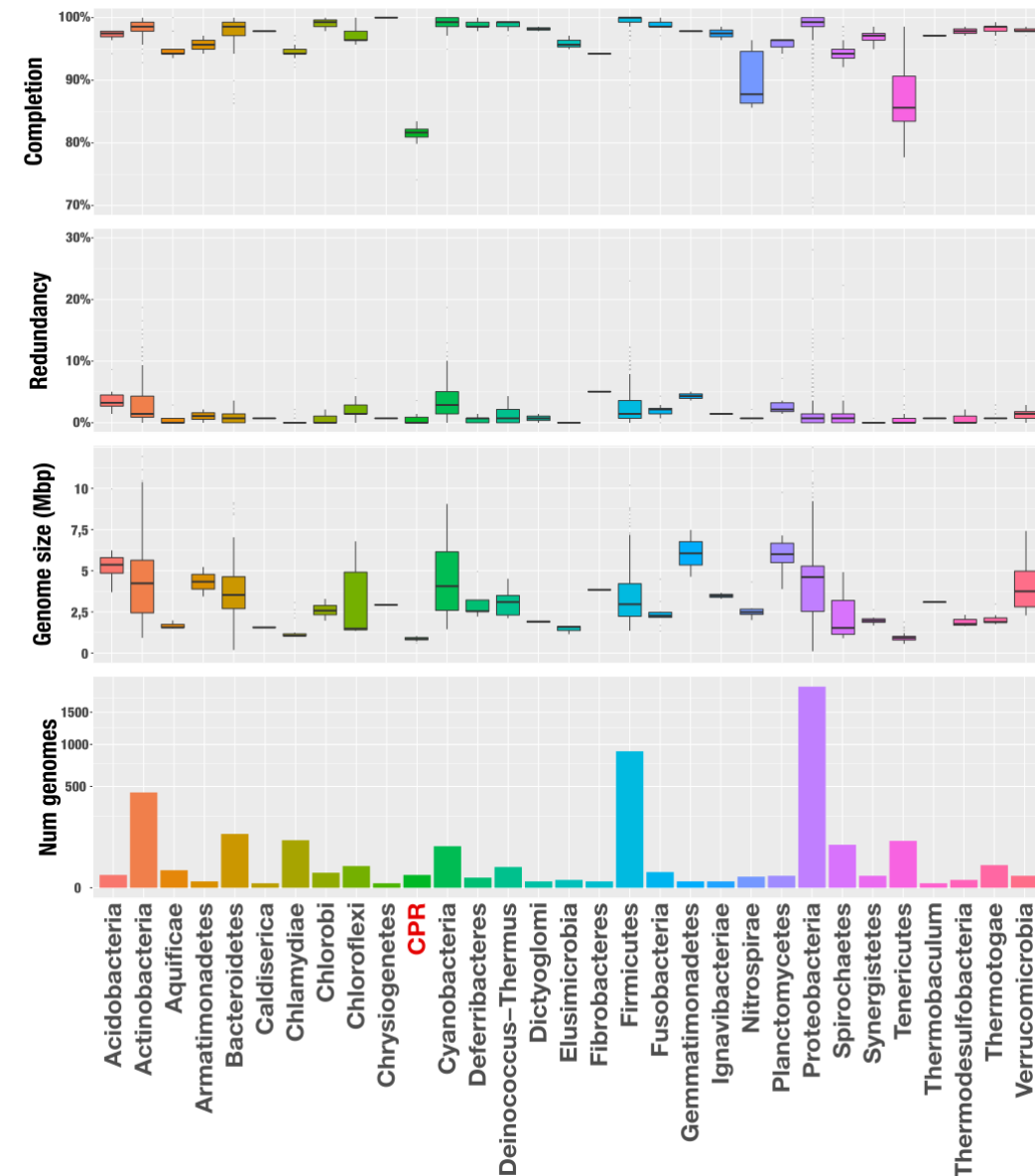
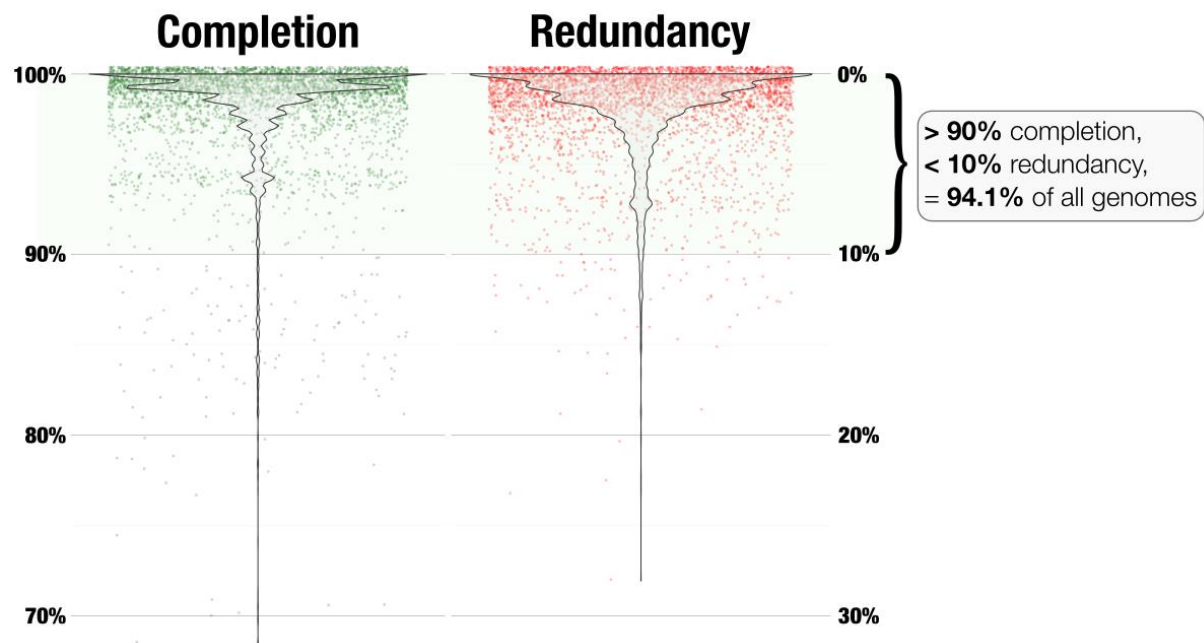


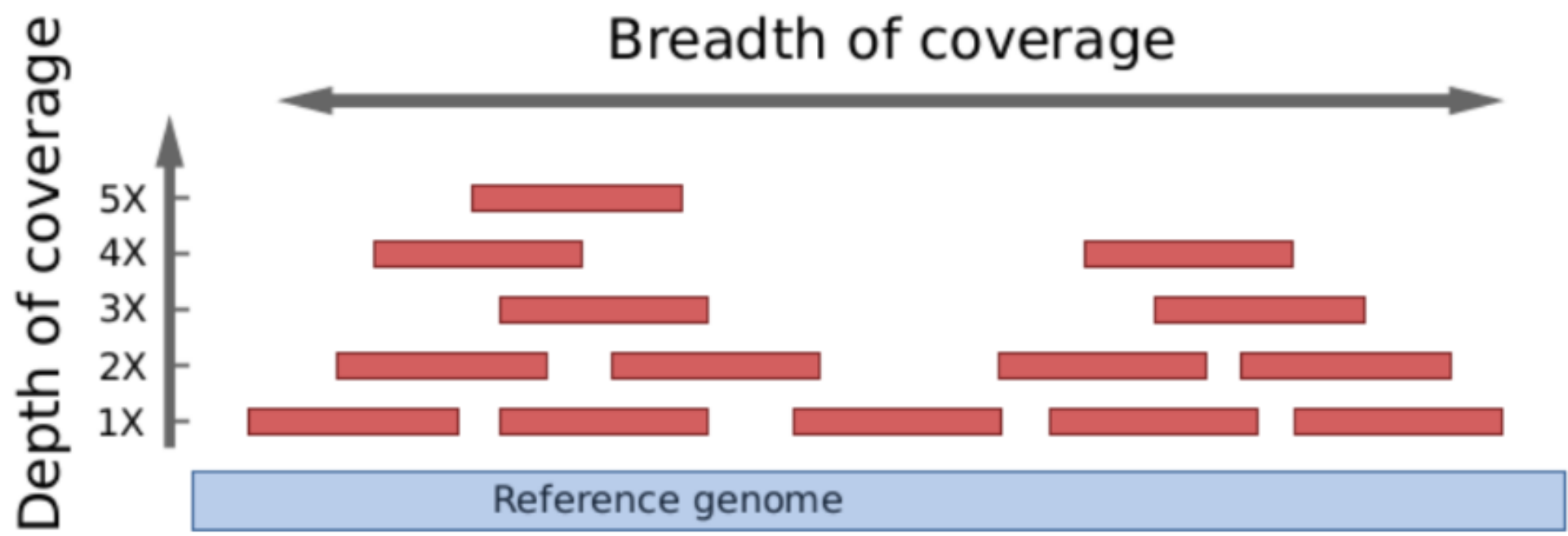
Evaluación de calidad





- Número de Single Copy Core Genes







Control de calidad y ensamblaje de genomas

