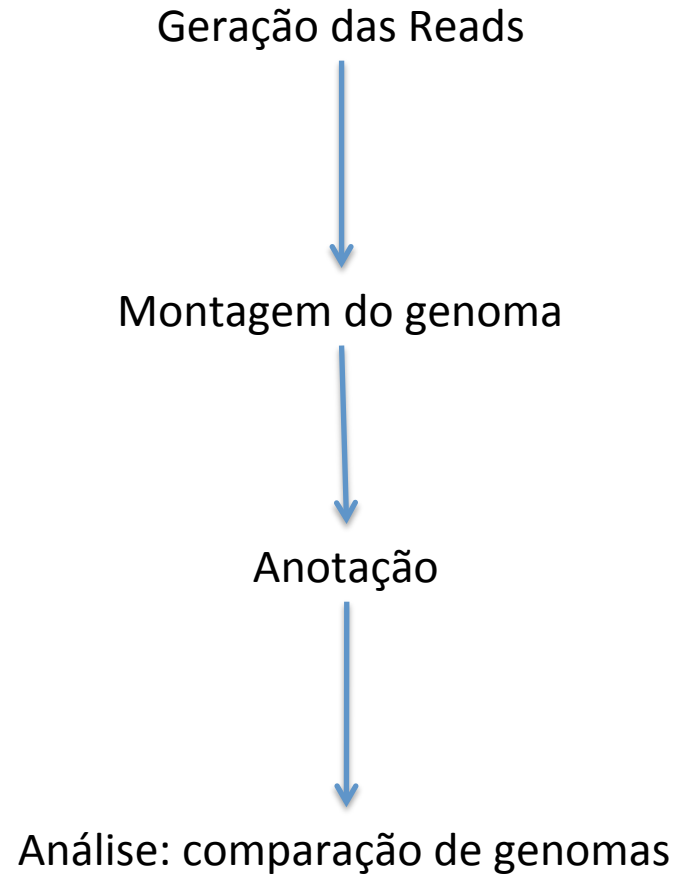


Montagem de Genomas

- Next-generation sequencing -

Tiago Mendes
Outubro/2015

Sequenciamento de genomas:



Sequenciamento de genomas:

Geração das Reads



Montagem do genoma

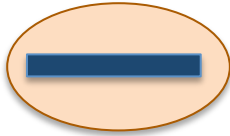


Anotação

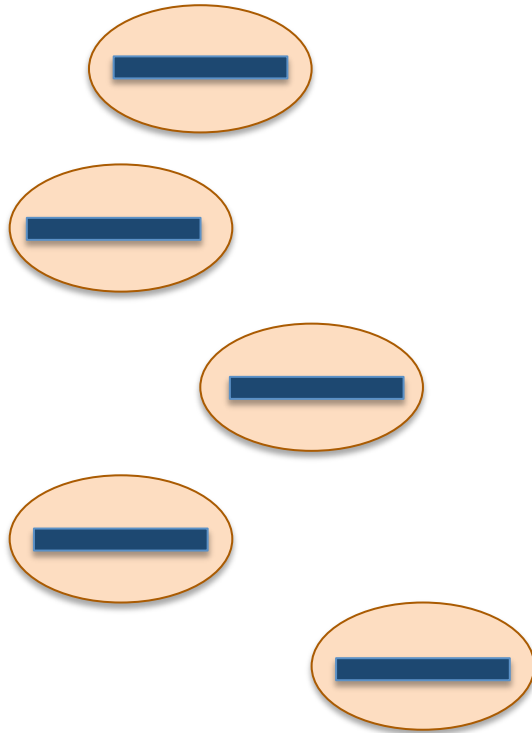


Análise: comparação de genomas

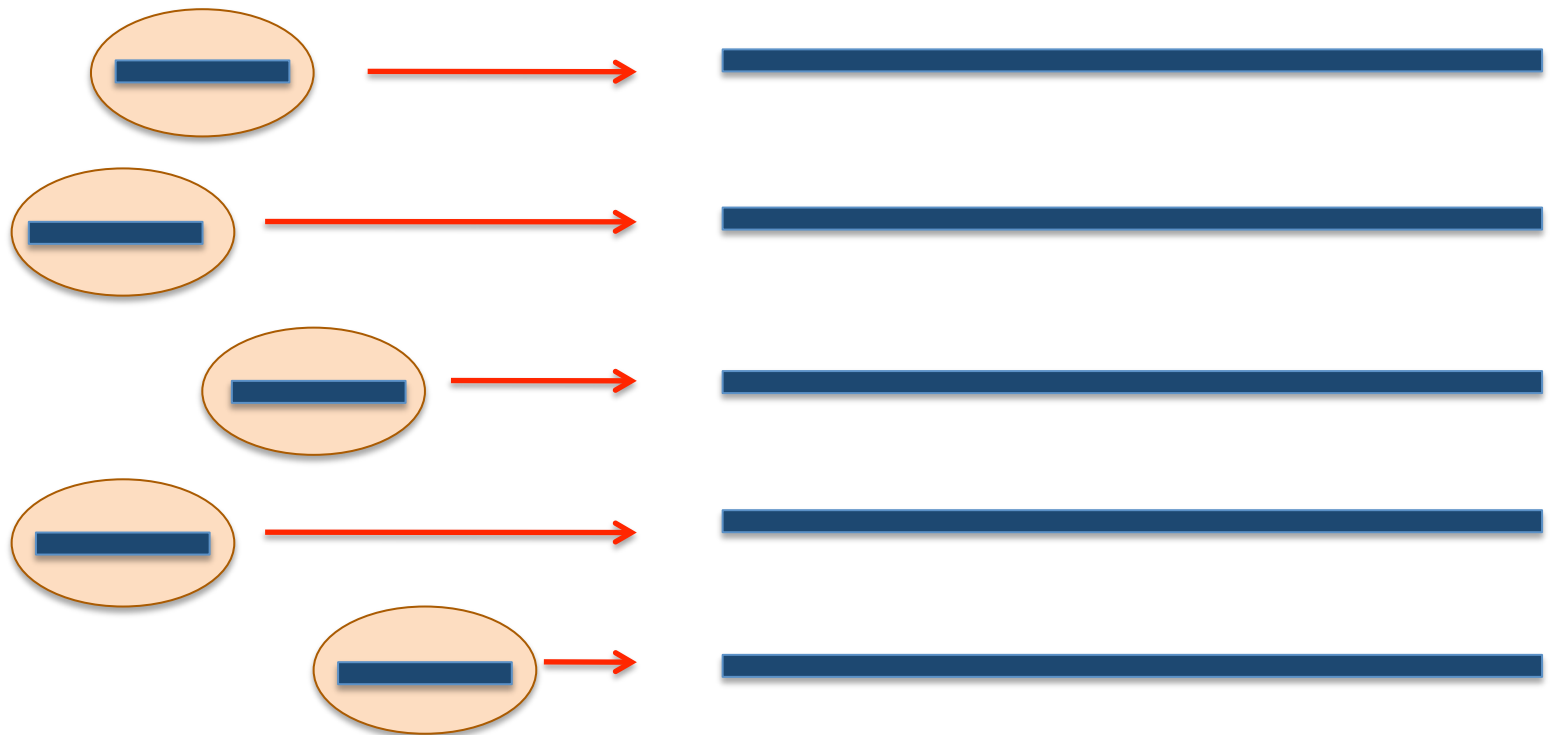
Geração de reads:



Geração de reads:



Geração de reads:



Geração de reads:



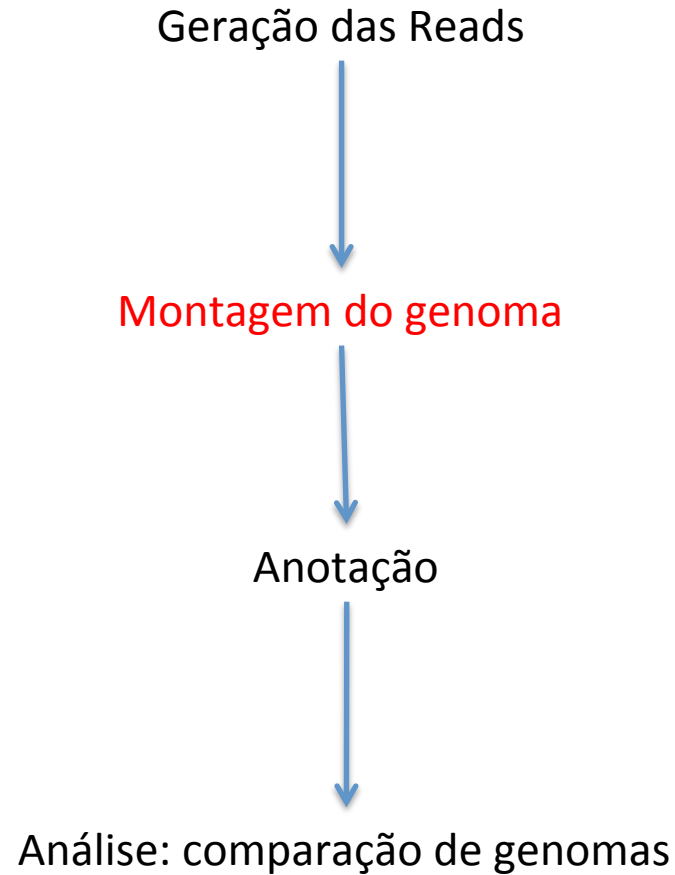
Geração de reads:



SEQUENCIAMENTO



Sequenciamento de genomas:



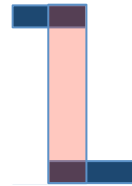
Montagem do genoma:



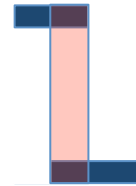
Montagem do genoma:



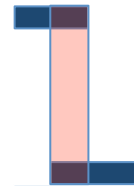
Montagem do genoma:



Montagem do genoma:



Montagem do genoma:



Aligned reads

```
ACGCGATTCAGGTTACCACG
GCGATTCAGGTTACCACGCG
GATTCAGGTTACCACGCGTA
TTCAGGTTACCACGCGTAGC
CAGGTTACCACGCGTAGCGC
GGTTACCACGCGTAGCGCAT
TTACCACGCGTAGCGCATT
ACCACGCGTAGCGCATTACA
CACGCGTAGCGCATTACACA
CGCGTAGCGCATTACACAGA
CGTAGCGCATTACACAGATT
TAGCGCATTACACAGATTAG
```

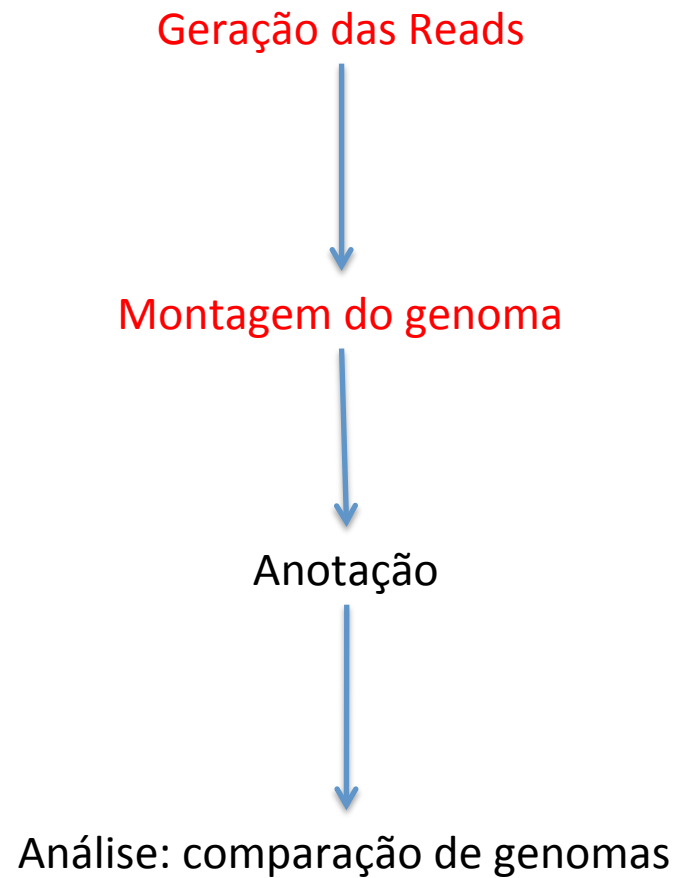
Consensus contig

```
ACGCGATTCAGGTTACCACGCGTAGCGCATTACACAGATTAG
```

Montagem do genoma:

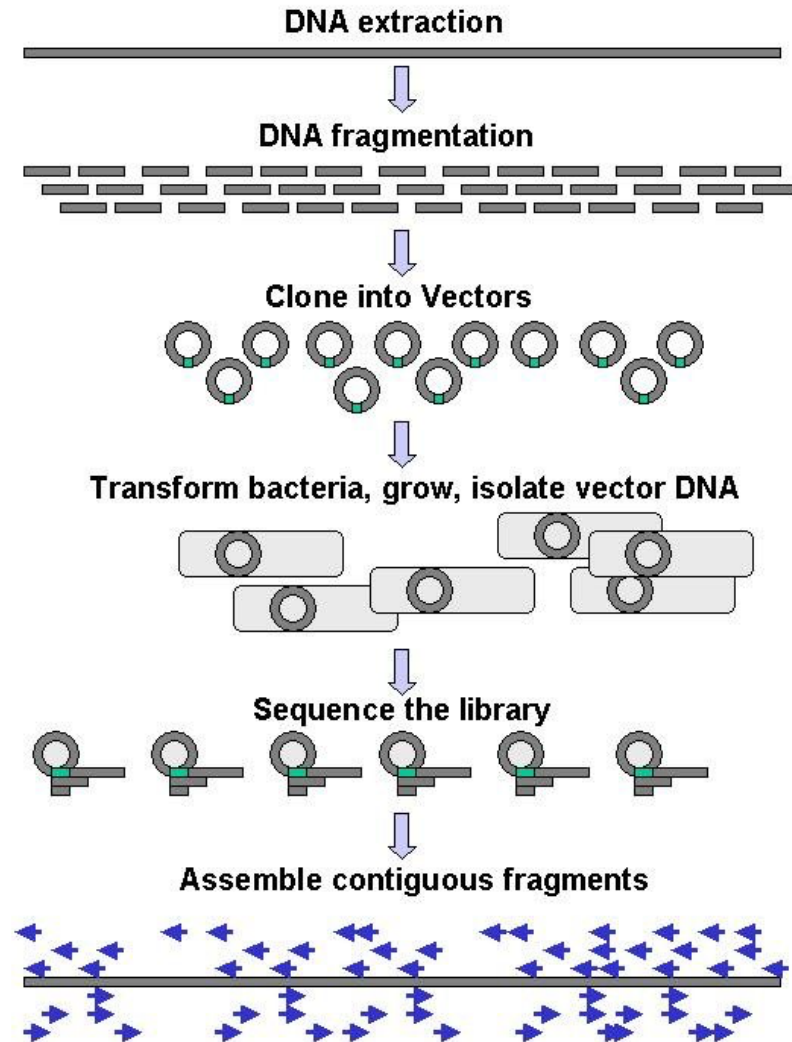


Sequenciamento de genomas:

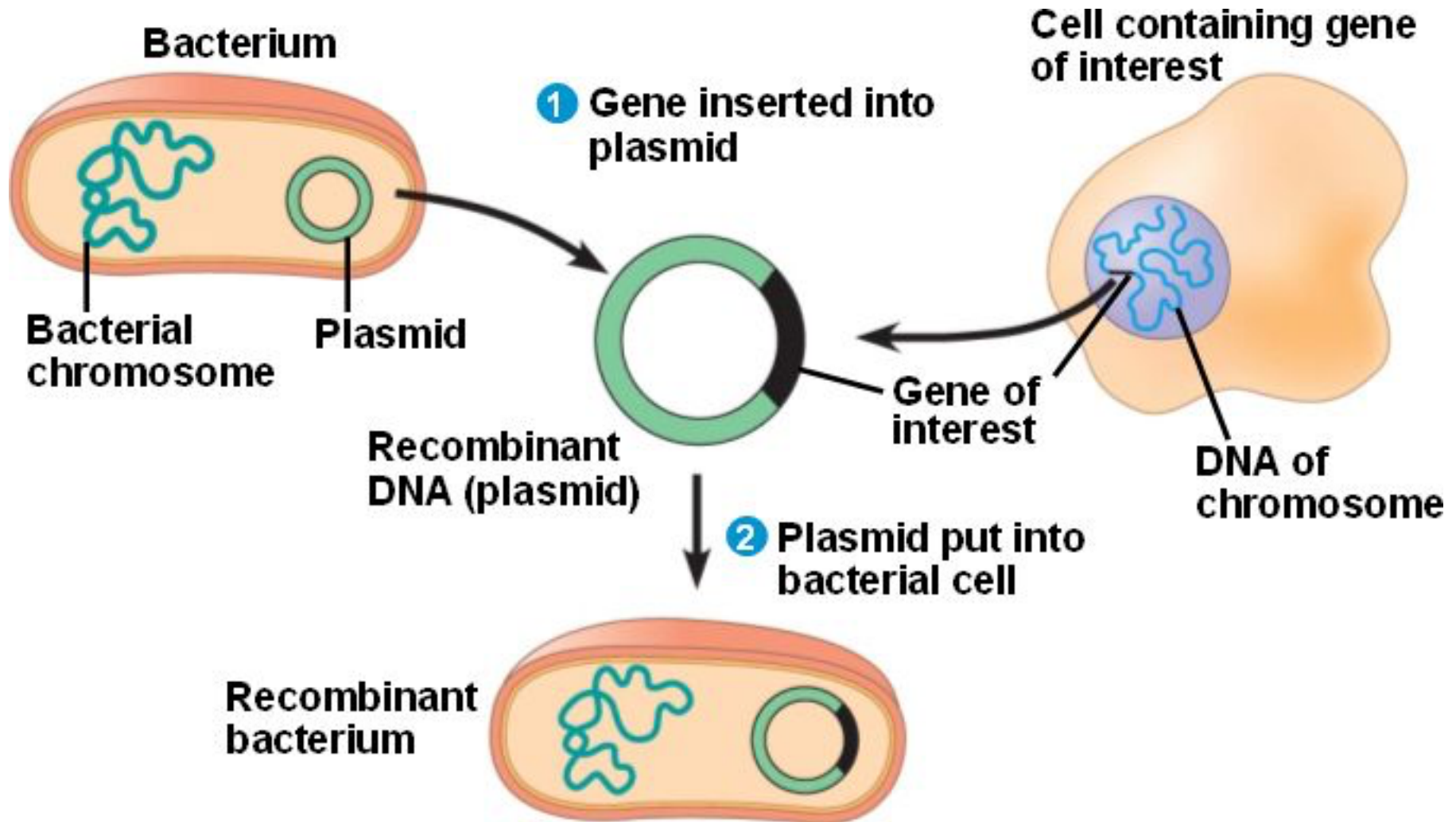


Sanger

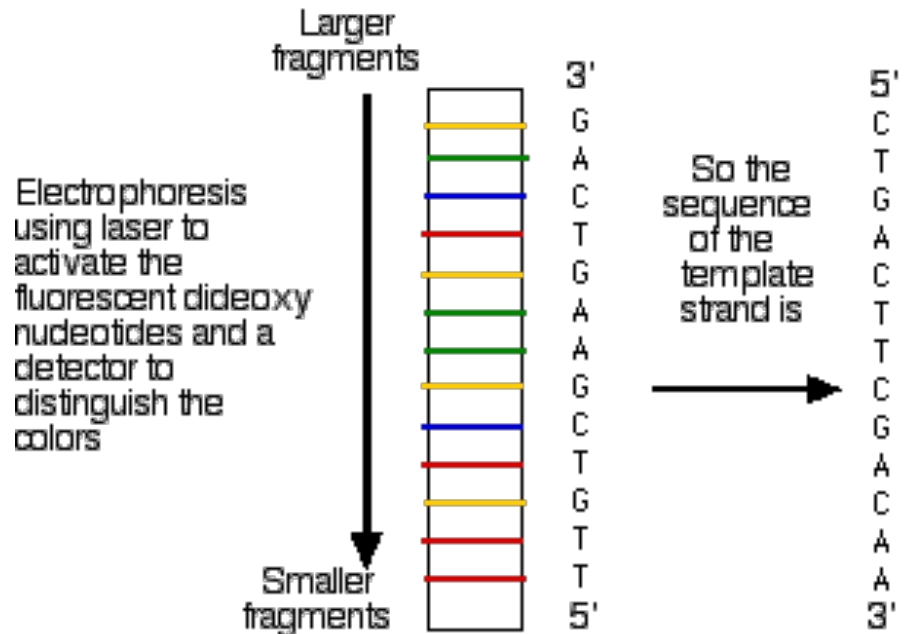
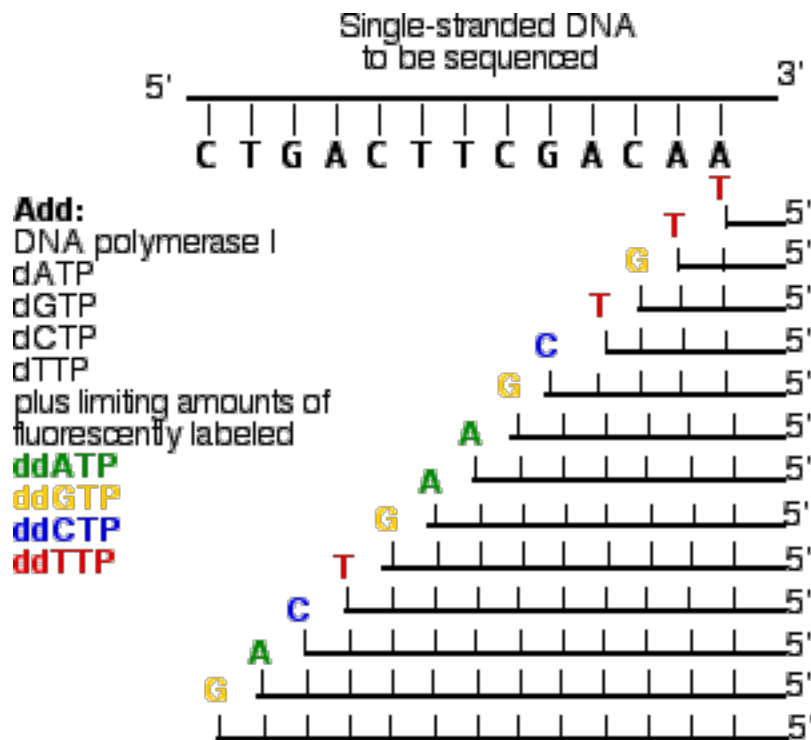
Sequenciamento de genomas:



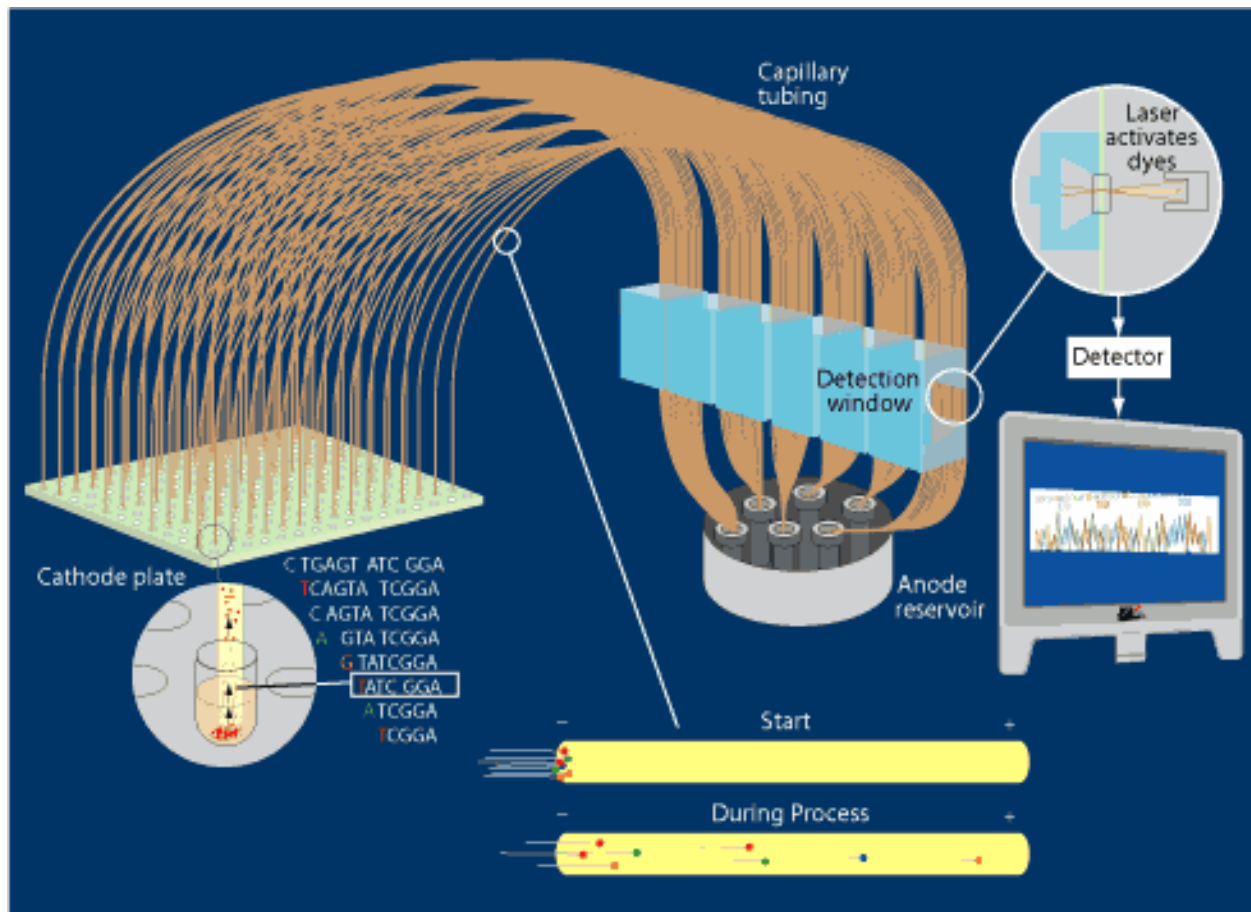
Sequenciamento de genomas:



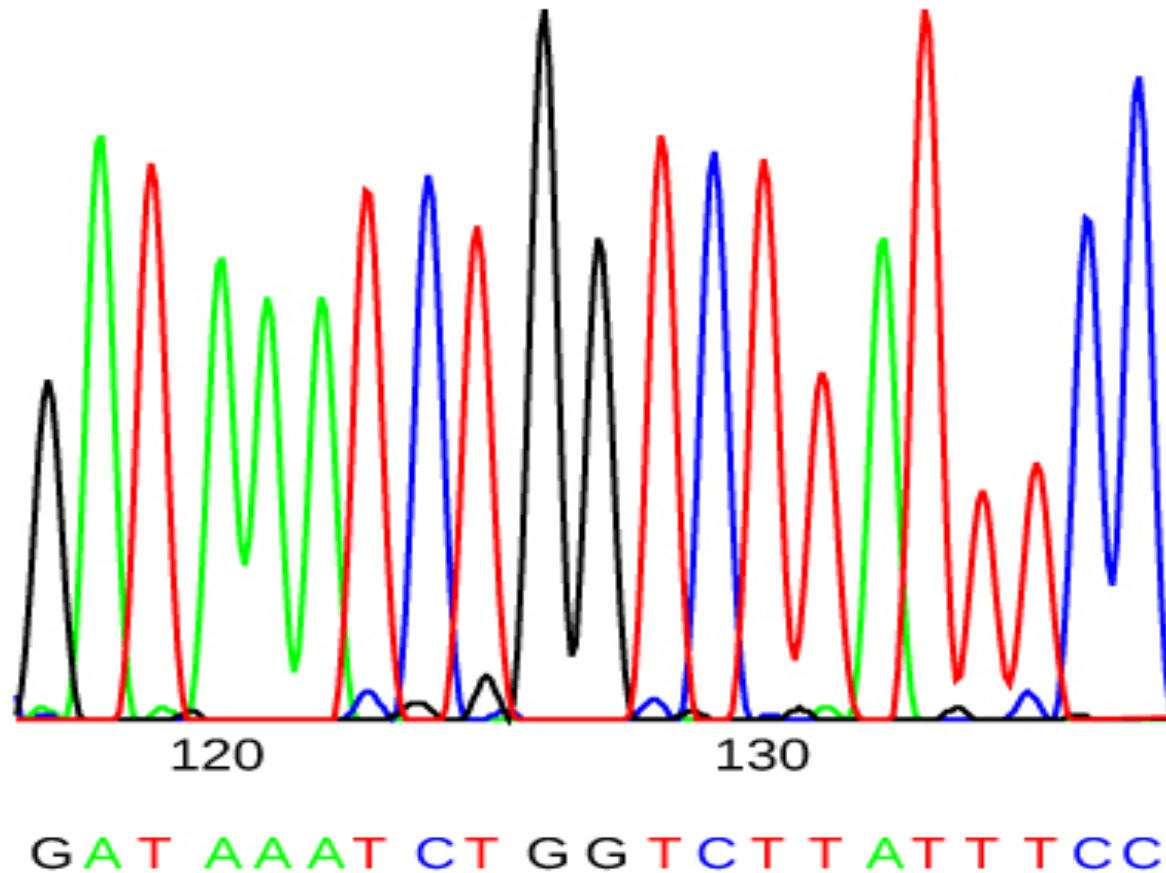
Sequenciamento de genomas:



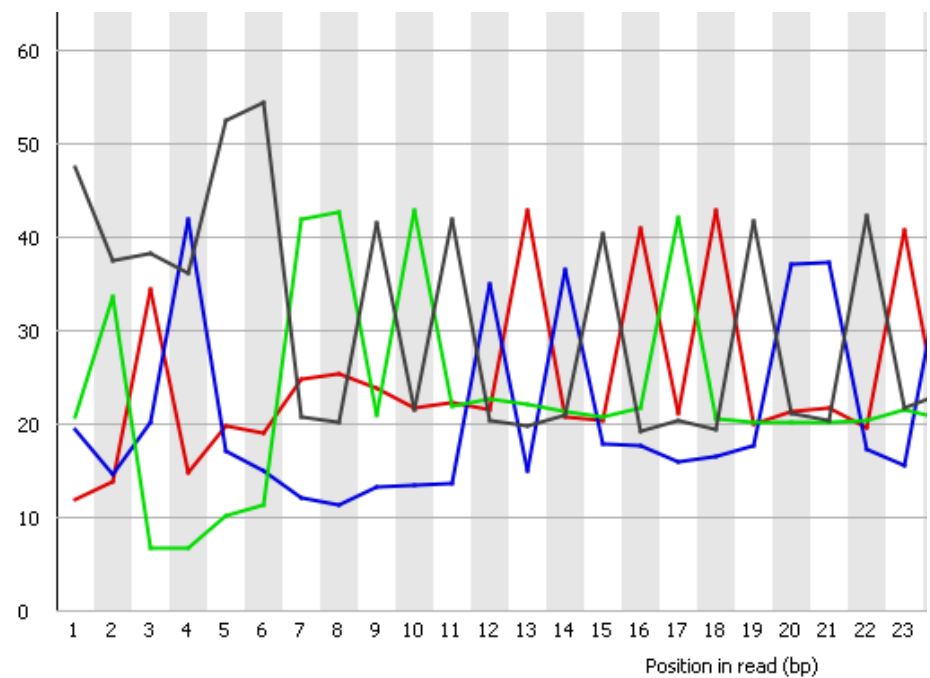
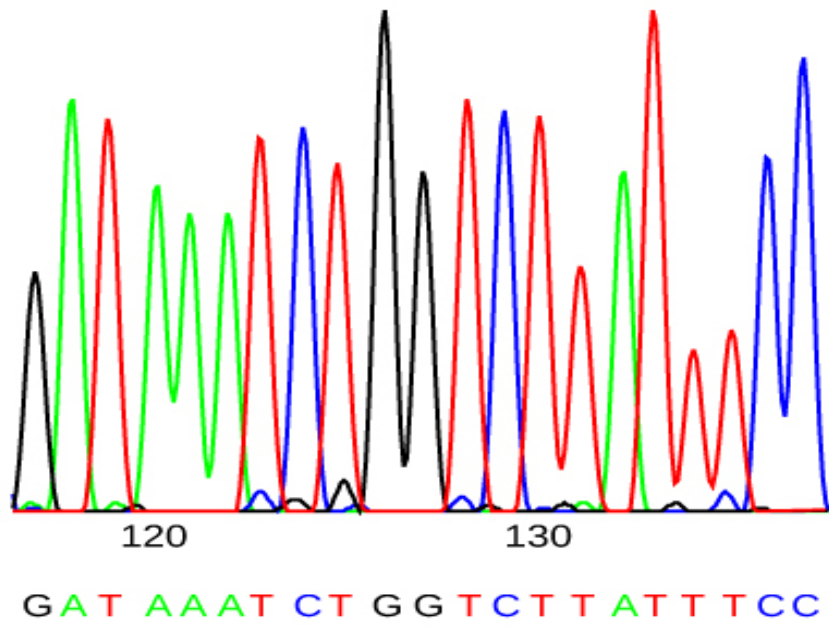
Sequenciamento de genomas:



Sequenciamento de genomas:



Sequenciamento de genomas:



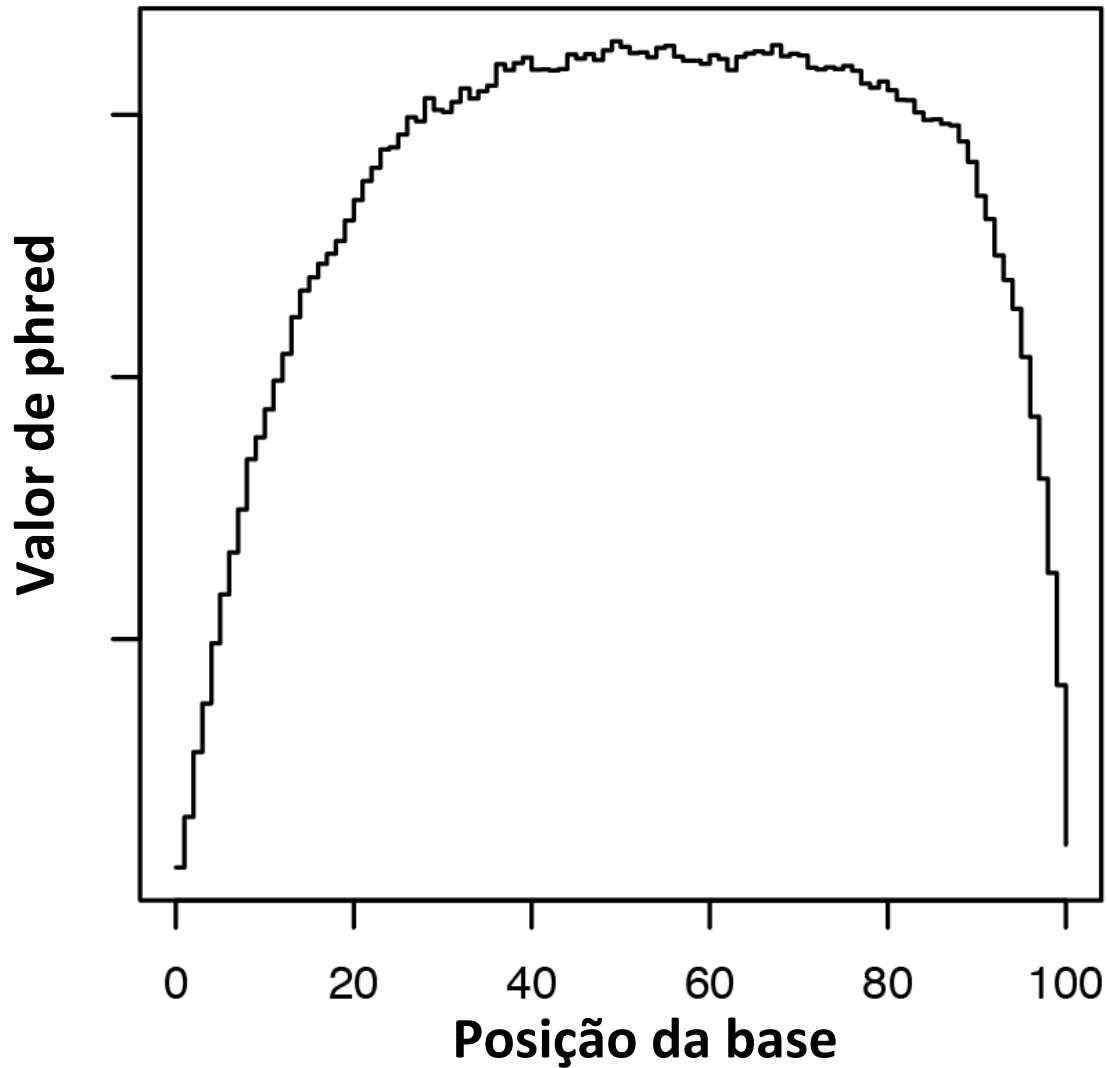
Sequenciamento de genomas:

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

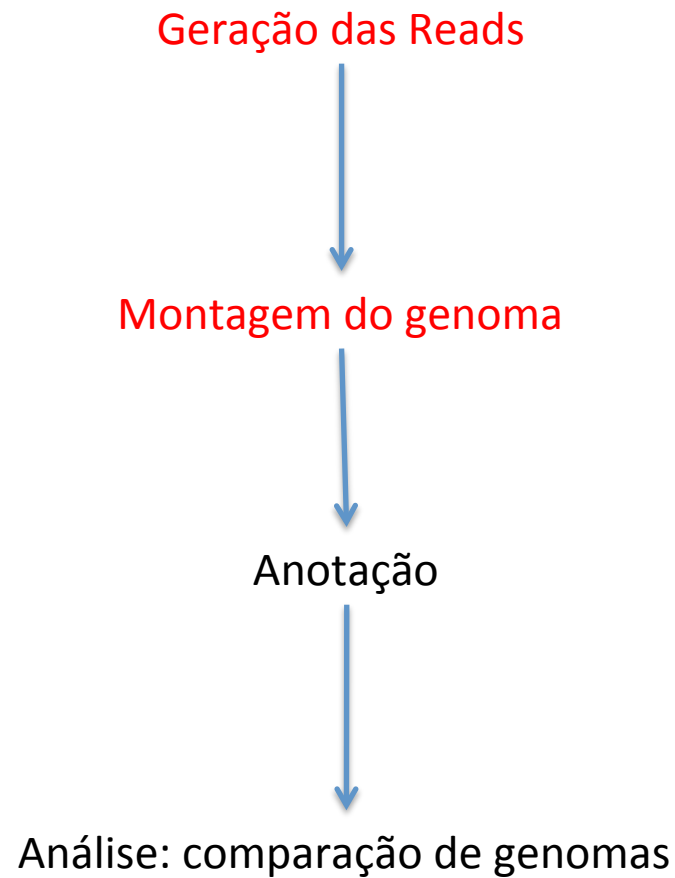
Phred quality scores are logarithmically linked to error probabilities

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 10,000	99.99%
50	1 in 100,000	99.999%
60	1 in 1,000,000	99.9999%

Sequenciamento de genomas:

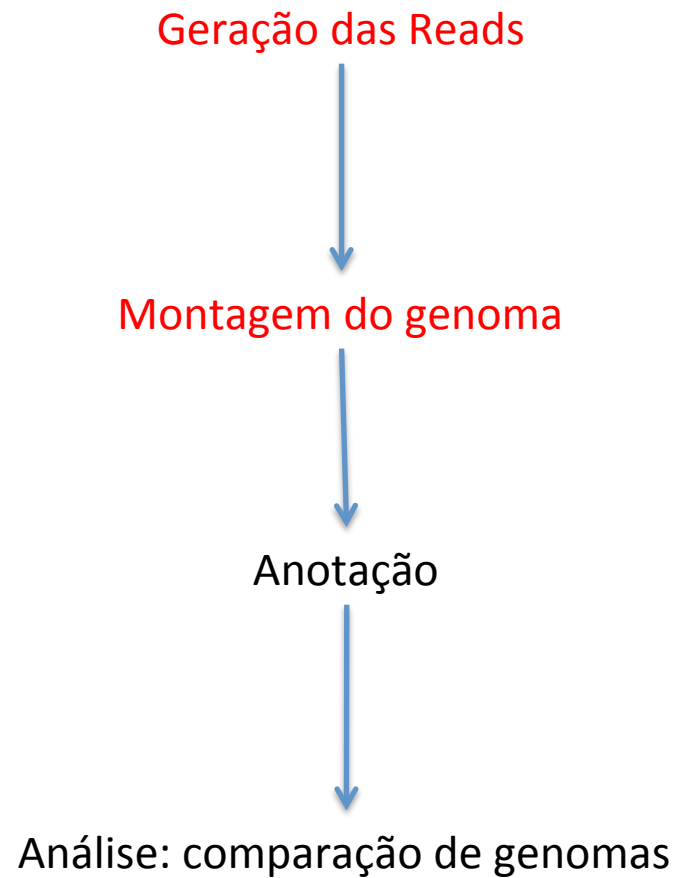


Sequenciamento de genomas:



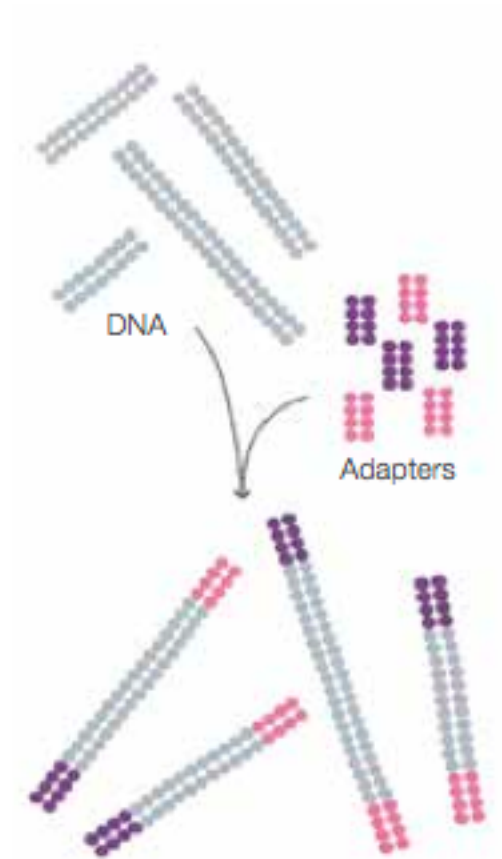
Solexa/Illumina

Sequenciamento de genomas:



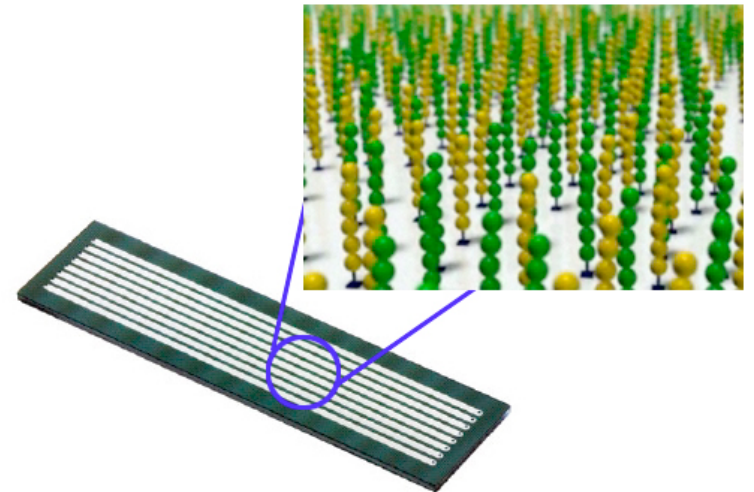
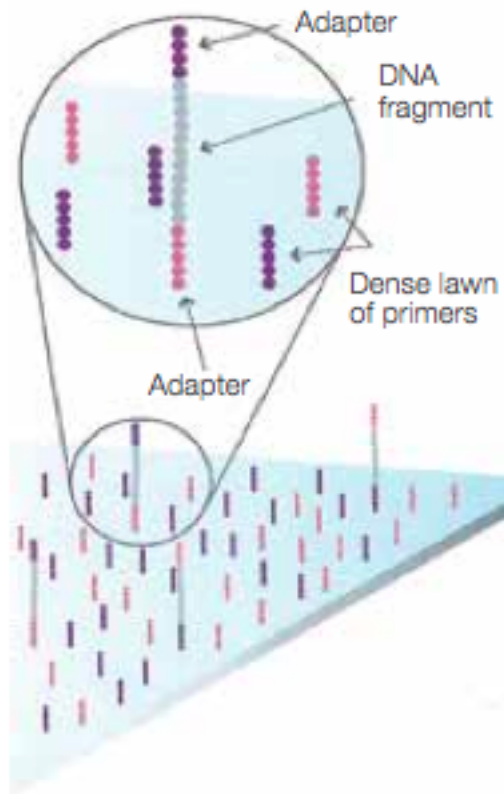
Solexa/Illumina

Ligamento de adaptadores:



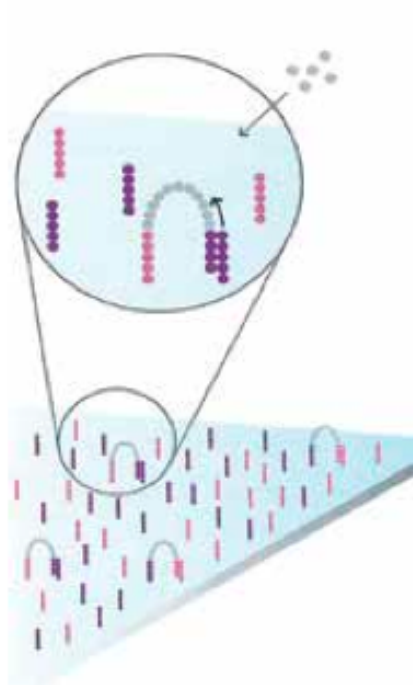
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

Ligação do DNA:

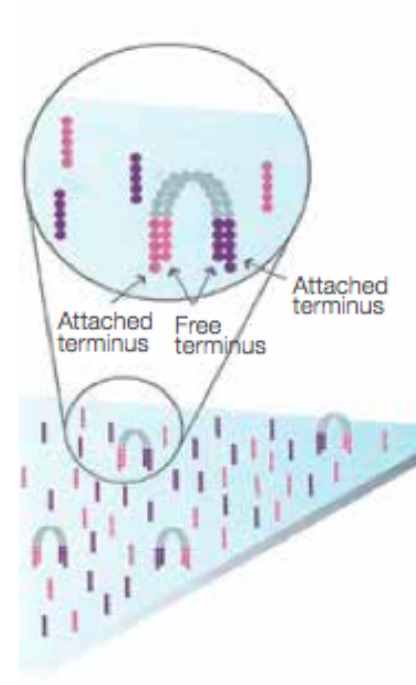


Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

Amplificação por ponte:

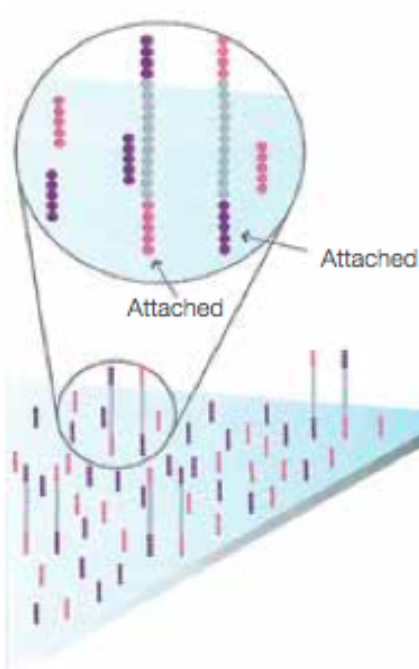


Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

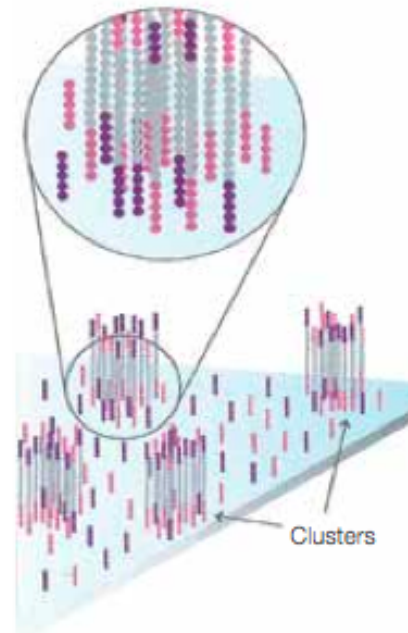


The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

Amplificação por ponte:

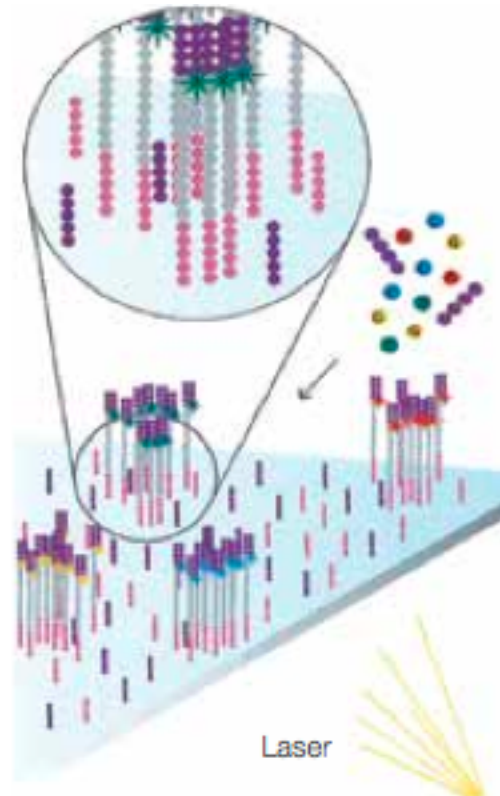


Denaturation leaves single-stranded templates anchored to the substrate.



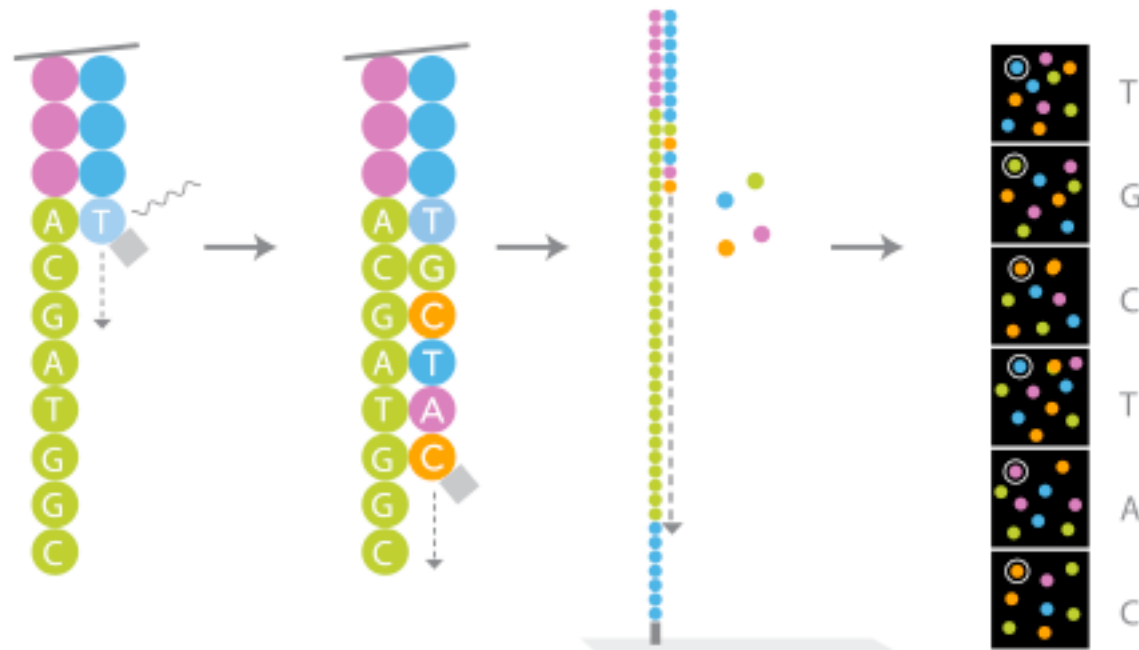
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

Reação de sequenciamento:

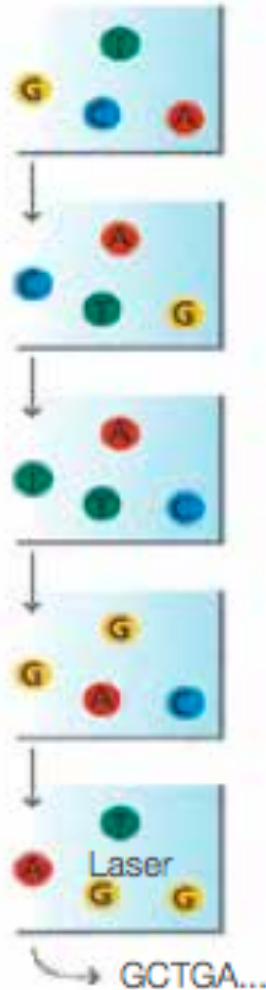


The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

Reação de sequenciamento:



Reação de sequenciamento:



Next
Generation
Sequencing
an animation by Aidan Flynn

FASTQ:

FASTA e QUAL:

```
>SRR014849.1 EIXKN4201CFU84 length=93
GGGGGGGGGGGGGGGGGGCTTTTTTTGTTTGGAAACCGAAAGG
GTTTTGAATTTCAAACCCTTTTCGGTTTCCAACCTTCCAA
AGCAATGCCAATA
```

and as a QUAL entry holding the PHRED scores:

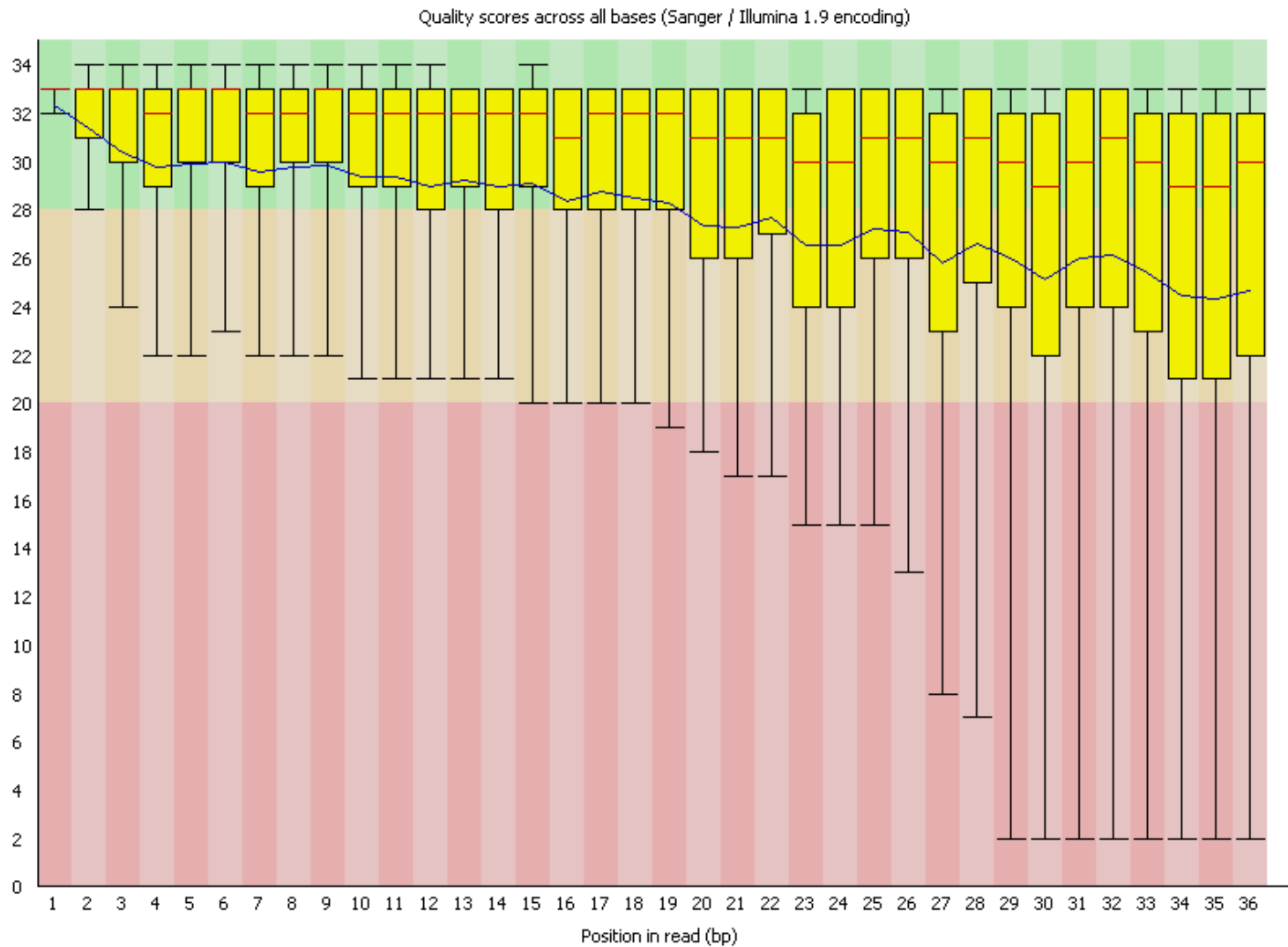
```
>SRR014849.1 EIXKN4201CFU84 length=93
18 10 5 3 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 22 37
31 22 16 11 6 1 26 34 30 11 33 26 30 21
33 26 25 36 32 16 36 32 16 36 32 20 6
24 33 25 30 25 2 24 36 32 15 35 31 17
36 32 20 6 25 29 20 30 25 4 32 26 32 23
32 26 30 24 33 26 35 31 14 28 27 30 22
28 24 27 17 32 23 28 28
```

FASTQ:

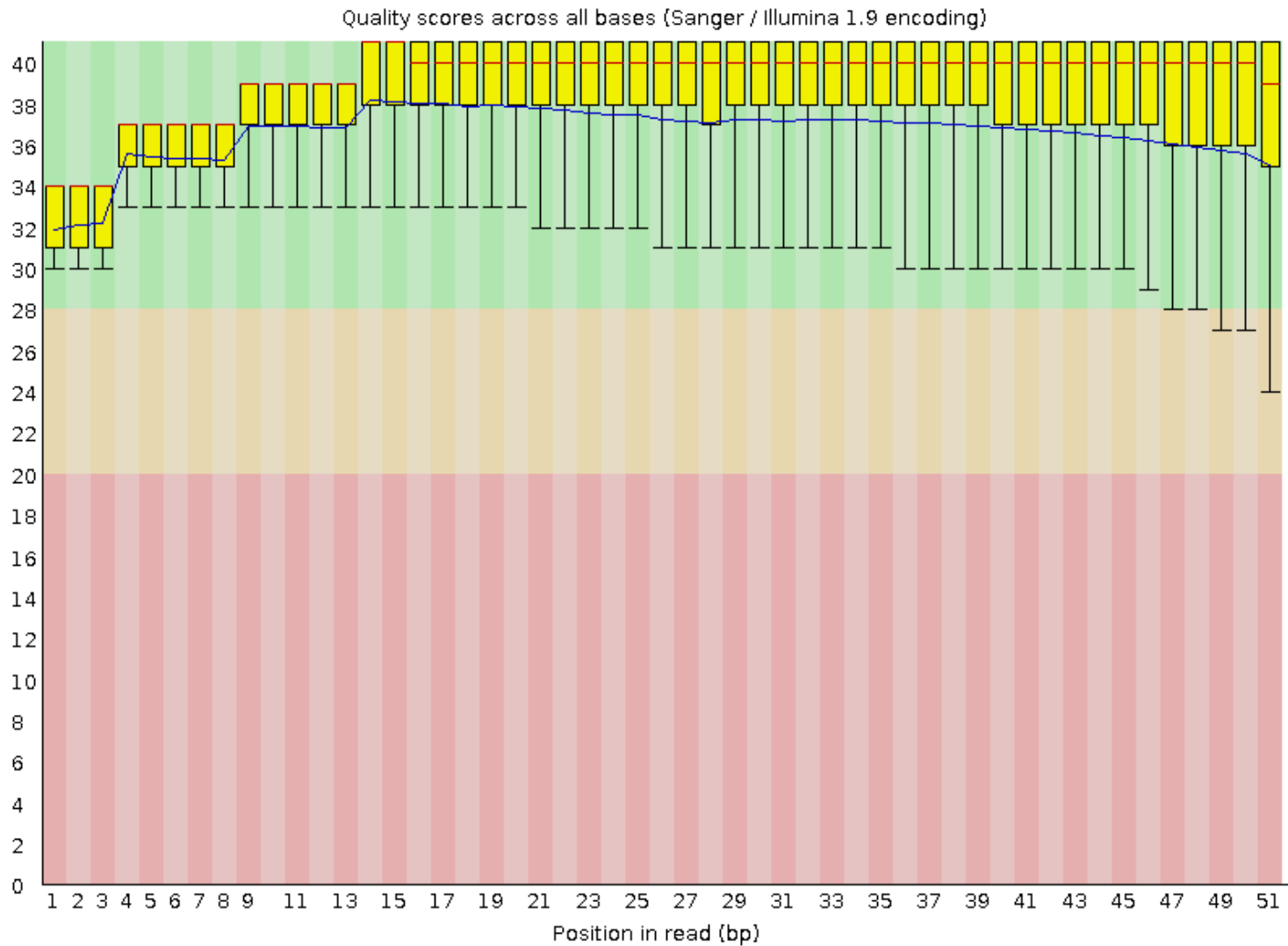
```
@SRR014849.1 EIXKN4201CFU84 length=93
GGGGGGGGGGGGGGGGGGCTTTTTTTGTTTGGAAACCGAAAGG
GTTTTGAATTTCAAACCCTTTTCGGTTTCCAACCTTCCAA
AGCAATGCCAATA
+SRR014849.1 EIXKN4201CFU84 length=93
3+&$#"7F@71, '";C?,B;?6B;:EA1EA
1EA5'9B?:#9EA0D@2EA5':>5?:%A;A8A;?9B;D@
/=<?7=9<2A8==
```

*@title and optional description
sequence line(s)
+optional repeat of title line
quality line(s)*

FASTQ:

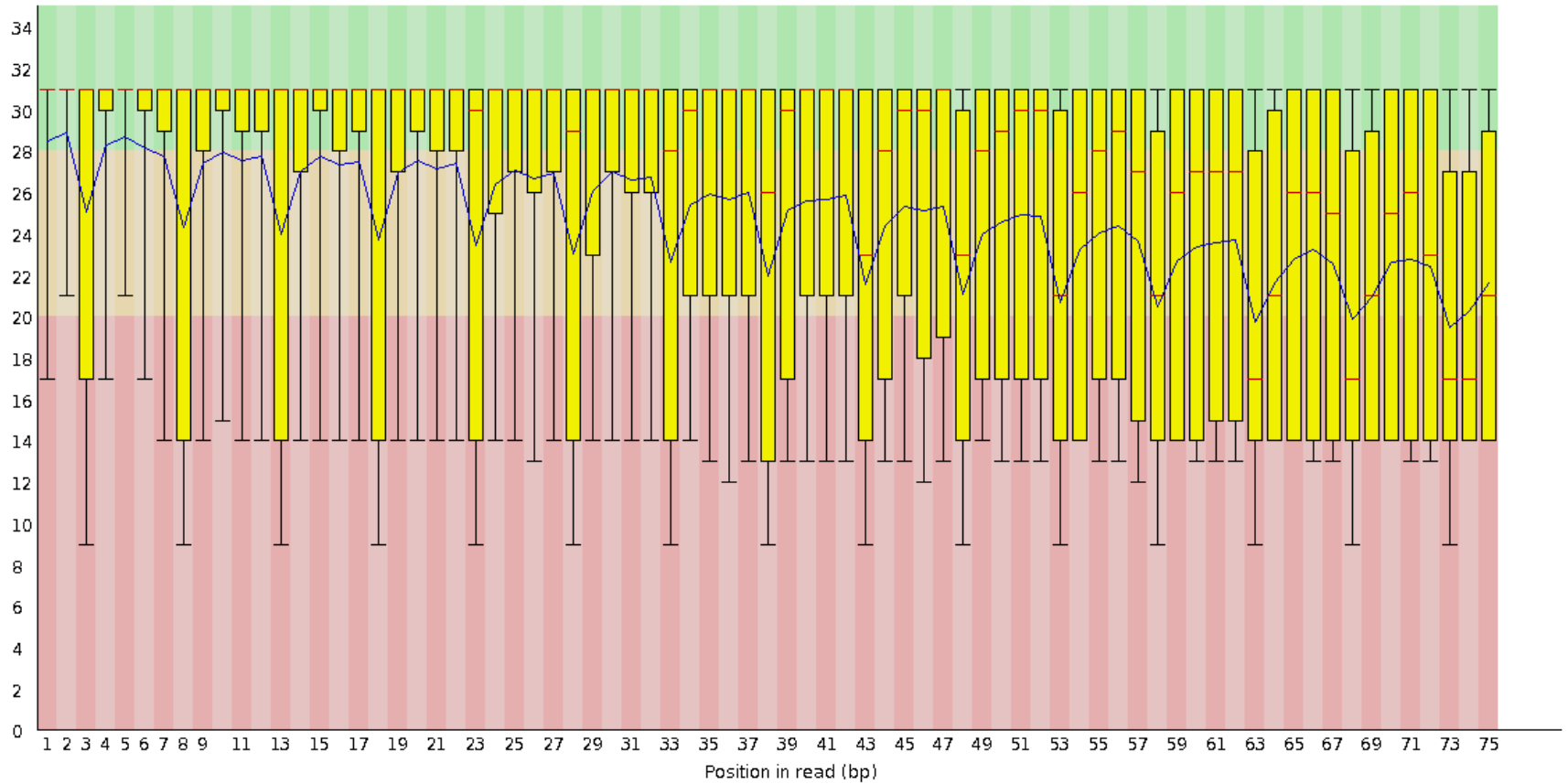


FASTQ:

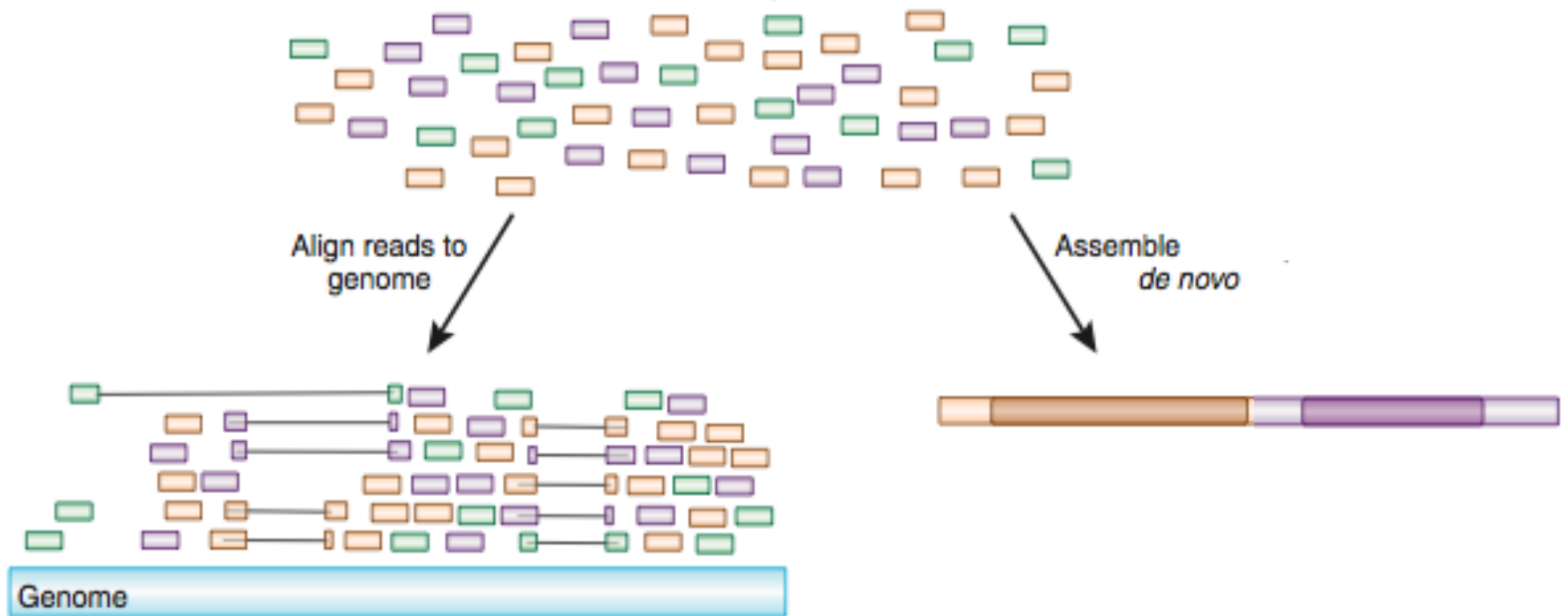


FASTQ:

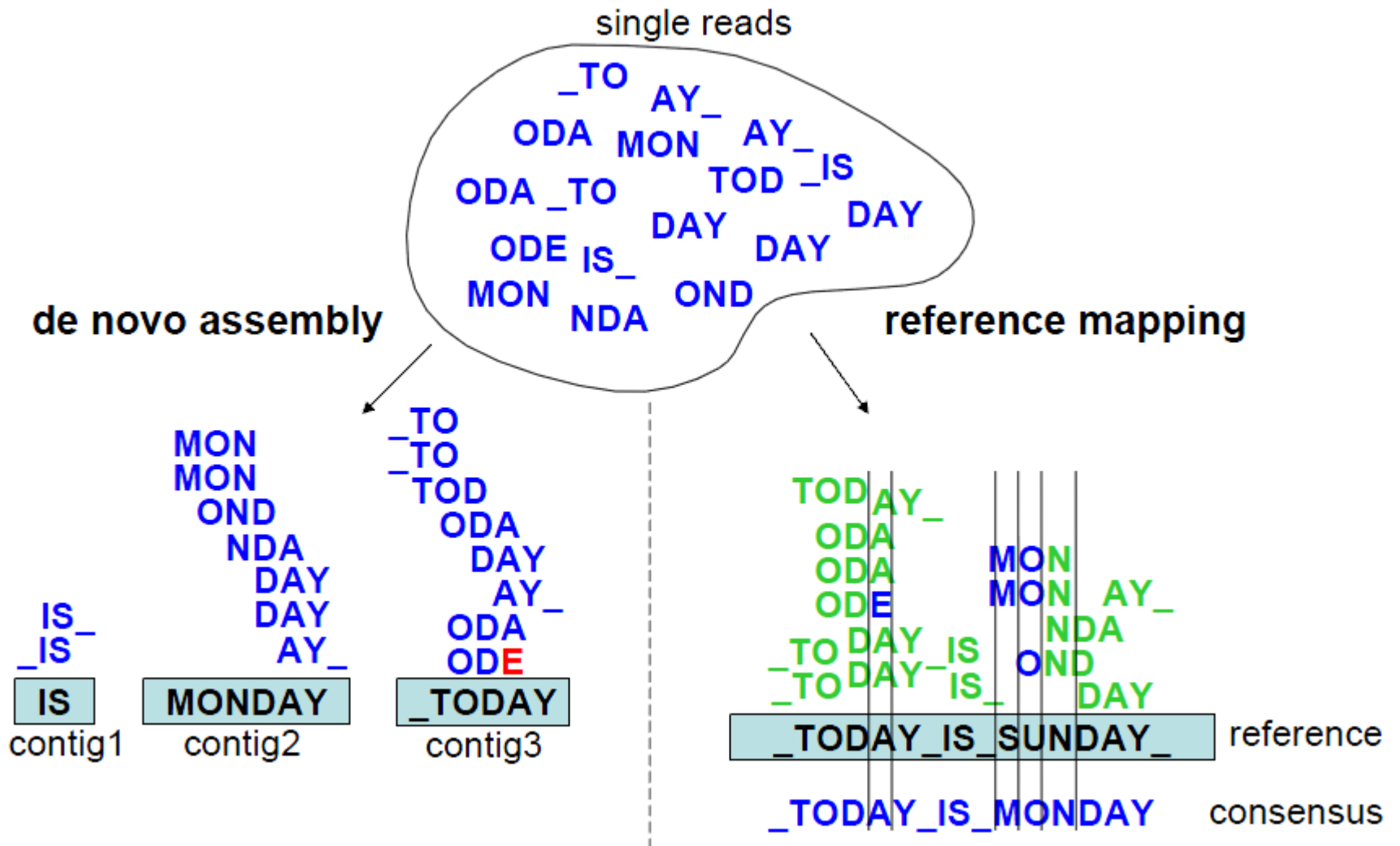
Quality scores across all bases (Sanger / Illumina 1.9 encoding)



Tipos de montagem: De Novo x Referência



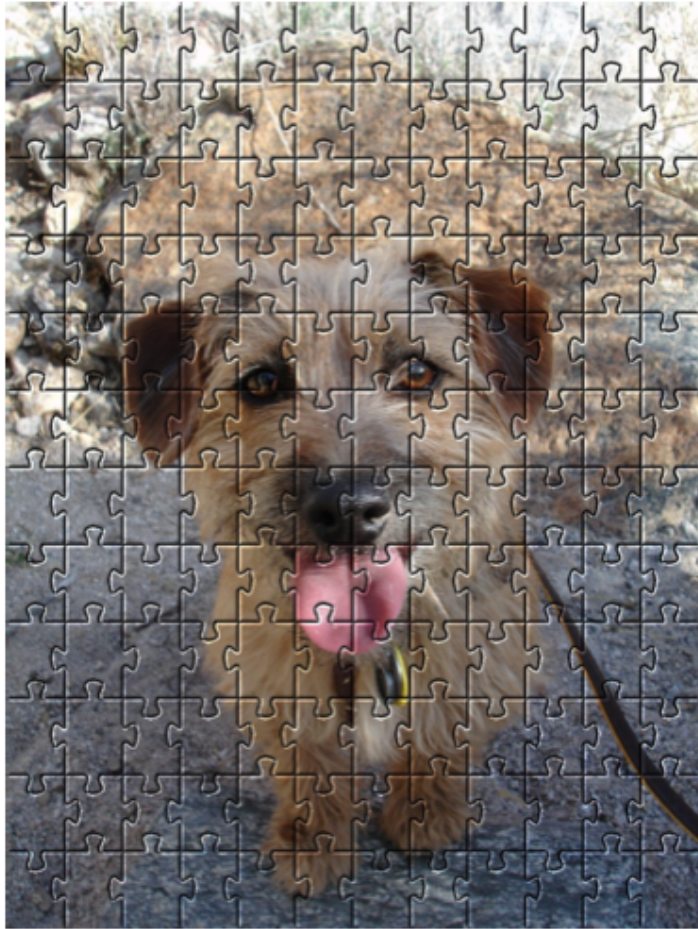
Tipos de montagem: De Novo x Referência



Problemas de montagem: Short reads



Problemas de montagem: Short reads



Short Reads



Long Reads

Problemas de montagem: Short reads



Estratégias de sequenciamento:

Único segmento

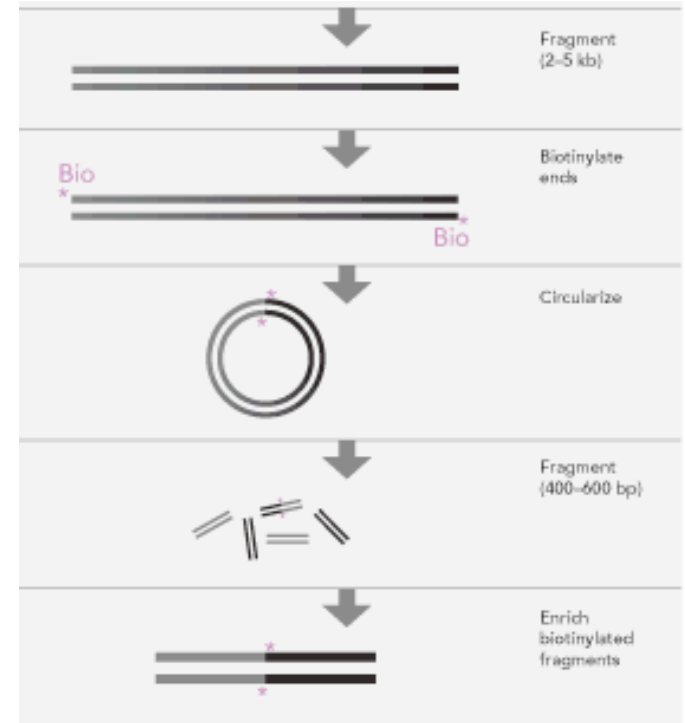


Paired-end



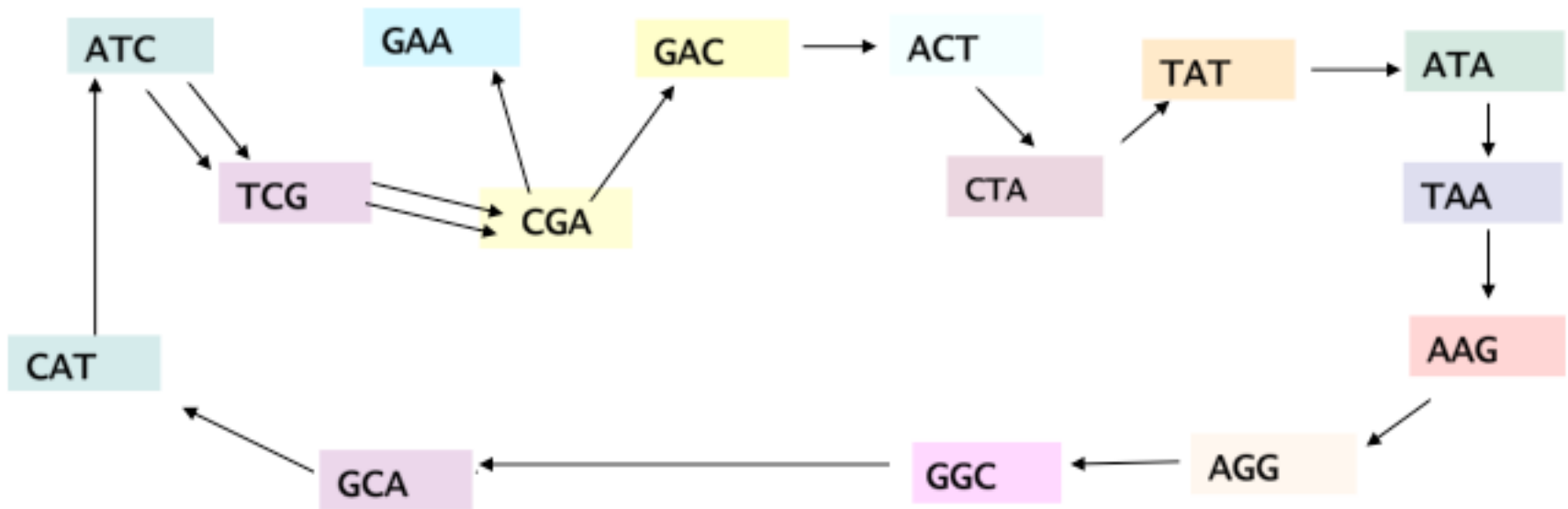
Reads

Mate-pair



Montador De Novo: Velvet

ATCGACTATAAGGCATCGAA



Montador por referência: Bowtie

The screenshot displays the Bowtie genome browser interface. The top toolbar includes sections for Data (Open Assembly, Import Features), Layout Style (Enhanced, Classic, Packed, Stacked, Sort), Adjust (Zoom, Variants), Navigate (Page Left, Page Right, Jump to Base), and Options. The main area is divided into a left sidebar with a contig table and a main alignment view.

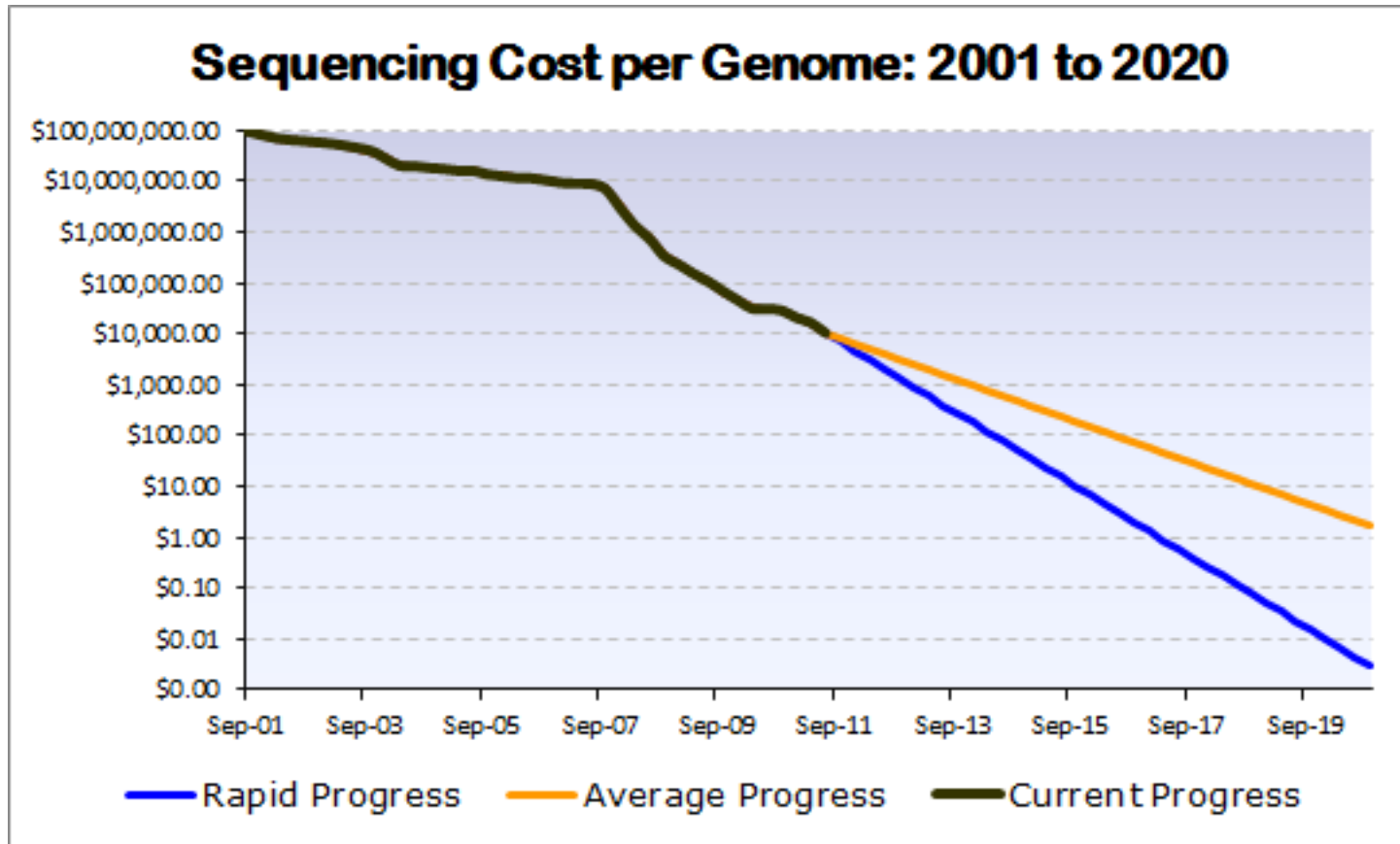
Contigs (33,518): Features (100):

Contig	Leng...	R...	Fea...
CL1Contig5...	4454	6016	78
CL1Contig37	7128	4897	140
CL1Contig5...	2424	2423	70
CL1Contig5...	1522	2297	64
CL1Contig9...	2674	2254	208
CL1Contig4...	943	2165	68
CL1Contig8...	9046	1682	100
CL1Contig6...	1266	1532	58
CL1Contig7...	1561	1450	38
CL1Contig5...	964	1280	62
CL5Contig2	2175	1275	43
CL6Contig3	337	1250	21
CL4Contig10	2167	1182	54
CL1Contig9...	802	1120	113
CL1Contig6...	1819	1118	70
CL1Contig8...	1635	1101	55
CL1Contig1...	1723	1092	63
CL8Contig3	2854	1075	66
CL1Contig3...	8783	1003	108
CL1Contig9...	1768	996	51
CL9Contig1	1934	995	43
CL1Contig6...	2332	984	30
CL10Contig3	1737	964	33

Feature tooltip: s36_FUBM92F02G4TOP
From: 4,747 U4,489 to 4,996 U4,716
Length: 250 U229

Tablet Tip: Navigate around an alignment by clicking and dragging on either the overview display area or the main display area

Preço de um genoma



Preço de um genoma

**First European service provider to provide
Complete Human Genome DNA sequencing**



**Discover
our SPECIAL
OFFER at 7500€
for 30x coverage!!**