

# ESCOPO DO SEQUENCIAMIENTO DE NOVA GERAÇÃO

**Dr. Alejandro Petroni**

Serviço Antimicrobianos

INEI-ANLIS “Dr. Carlos G. Malbrán”

[apetroni@anlis.gob.ar](mailto:apetroni@anlis.gob.ar)

TRABAJANDO  
JUNTOS  
PARA COMBATIR  
LA RESISTENCIA  
A LOS ANTIMICROBIANOS



Organización de las Naciones  
Unidas para la Alimentación  
y la Agricultura



ORGANIZACIÓN MUNDIAL DE SANIDAD ANIMAL  
*Proteger a los animales, preservar nuestro futuro*



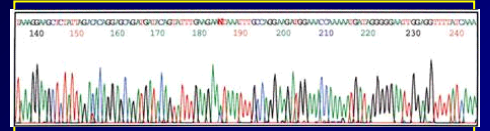
Unión Europea

# SEQUENCIAMENTO

TARGET

1 fragmento de  
sequência CONHECIDA

SANGER



2 eletroferogramas  
(Forward, Reverse),  
máximo: 800-1.000  
bases em cada um

Sequenciamento de 1.<sup>a</sup> geração

TARGET

?

Completo ≠ Variantes

Sequenciamento de 2.<sup>a</sup> geração

Reads curtas: 50 a 600 bases

- ILLUMINA
- ION TORRENT

Diferenças com SANGER:

- Determinação de sequências em **tempo real**
- Dezenas de milhares de leituras de DNA (*reads*) que se sobrepõem e cobrem TODO o target **≥ 30 vezes**

**Cobertura média (coverage, sequencing depth):**  
N.º de vezes, na média, que cada base do target for sequenciada; indica-se como “30x”, “100x”, etc.

## SEQUENCIAMENTO DE NOVA GERAÇÃO

- ✓ Next Generation Sequencing (NGS)
- ✓ Massively parallel sequencing
- ✓ Deep sequencing
- ✓ High-throughput sequencing

Se Target = todo o genoma:

- ✓ Whole Genome Sequencing (WGS)

- PACIFIC BIOSCIENCES (PacBio)

# USO DE NGS EM BACTERIOLOGIA CLÍNICA

- ✓ **EPIDEMIOLOGIA:**
  - **Tipificação molecular de isolamentos:** filogenia por análise de pangenoma, ou por *core genome* MLST (cgMLST): ↑↑↑ poder do que MLST (milhares de genes completos em vez de fragmentos de 6-7 genes). WGS é o padrão-ouro do teste de clonalidade
  - **Estudo de surtos**
  - **Evolução e disseminação espacial de sequências:** genomas, plasmídeos e outros elementos genéticos móveis; fatores de virulência
- ✓ **PERFIS DE RAM:**
  - **Caracterização de novos mecanismos de RAM**
  - **Determinação genotípica de RAM:** desenvolvimento de sistemas de inferência de sensibilidade a partir do genótipo
  - **Dinâmica espaço-temporal de emergência e disseminação de RAM**
- ✓ **METAGENÔMICA:**
  - **detecção e identificação DIRETA de bactérias:** casos de infecção “complexos” (endocardite, infecções polimicrobianas; LCR ou líquido sinovial <sup>c</sup>/cultura negativa)
  - **Caracterização de reservatórios de RAM em hospitais**

# Como é realizado um NGS?

Os esquemas de trabalho das metodologias NGS incluem etapas muito semelhantes

## A) Sequenciamento de 2.<sup>a</sup> geração, Illumina:

Vídeo de sequenciamento por Illumina:  
<https://www.youtube.com/watch?v=womKfikWlxM>

**AMOSTRA** (bactéria isolada: *pellet* celular, cultura ½ líquido)

### 1. Extração automatizada de DNA

~~DNA~~ DNA: determinação de pureza ( $A_{260}$ ,  $A_{280}$ ) e quantificação

### 2. Construção de biblioteca genômica (*library*):

- ✓ Fragmentação enzimática de DNA → conjunto de fragmentos comprimento médio de 800 a 1.000 pb
- ✓ Adaptação de extremos dos fragmentos para realizar o sequenciamento. Inclui a adição a cada amostra de um *barcode ou index*: uma sequência única de 6-8 bases, que permite identificar todas as sequências dessa amostra. Este processo de *barcoding* permite realizar o *multiplexing*: sequenciamento conjunto de diferentes amostras no mesmo dispositivo → alta redução de custos

### 3. Sequenciamento automatizado de DNA:

- ✓ É utilizada uma *flow cell*: dispositivo onde ocorre o processo biológico que determina a sequência
- ✓ Amplificação clonal de <sup>c</sup>/fragmento
- ✓ Sequenciamento por síntese, e em ciclos (50-300): em cada ciclo é incorporado 1 de 4 dNTPs (marcados com diferentes fluoróforos) e a incorporação é detectada por fluorescência
- ✓ Sequenciamento do outro extremo de cada fragmento: *paired-end sequencing* (2 x N.<sup>o</sup> ciclos)

## A) Sequenciamento de 2ª geração, cont':



Sequenciador MiSeq (Illumina)

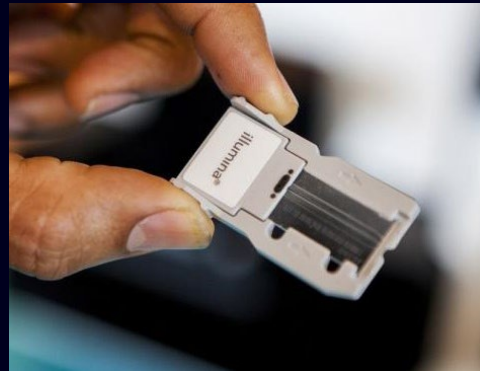
Após a finalização do sequenciamento, um software do sequenciador realiza o **demultiplexing**: utilizando os **barcodes**, classifica e agrupa todas as sequências conforme a amostra de que provêm



### RESULTADOS para cada amostra:

2 arquivos de texto, R1 e R2: contêm todas as **reads** geradas do 1º e 2º extremo de todos os fragmentos sequenciados, respectivamente

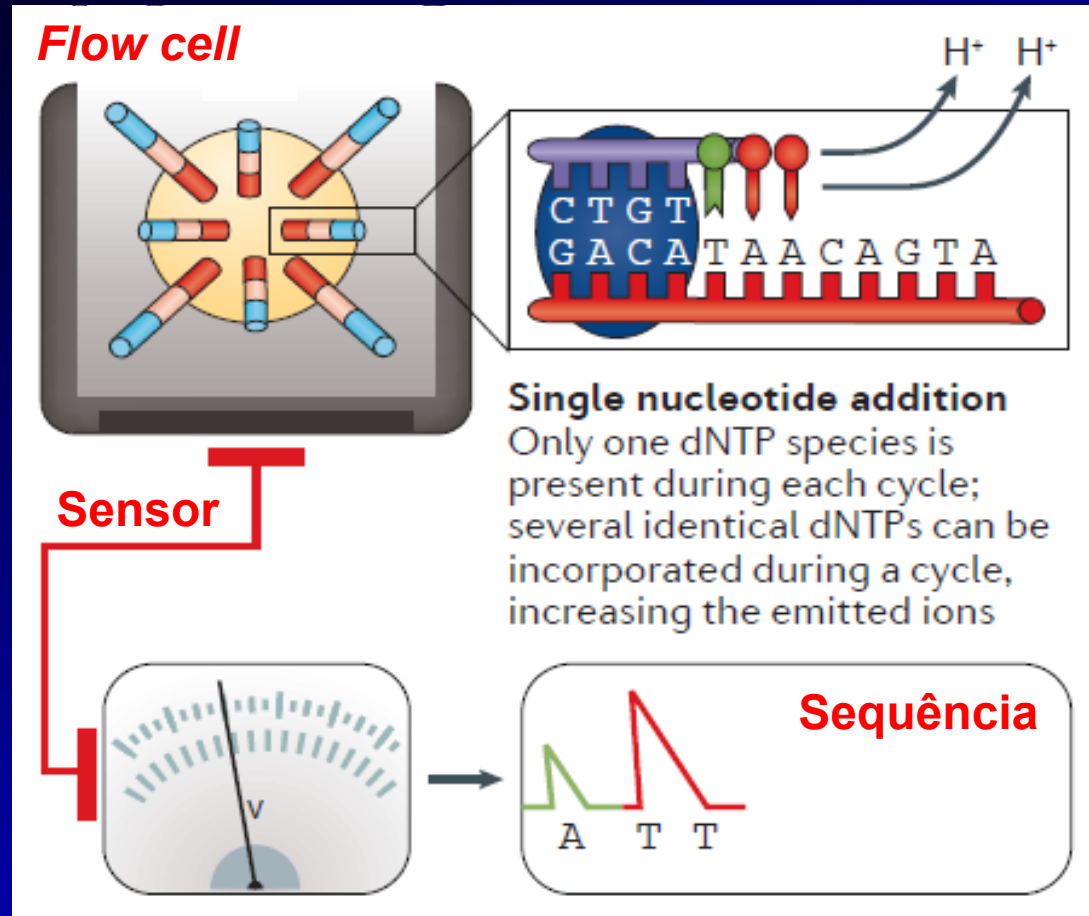
Formato **FASTQ**: em cada **read** é indicada a sequência de bases e os valores de qualidade na determinação de cada base



Flow cell (Illumina)

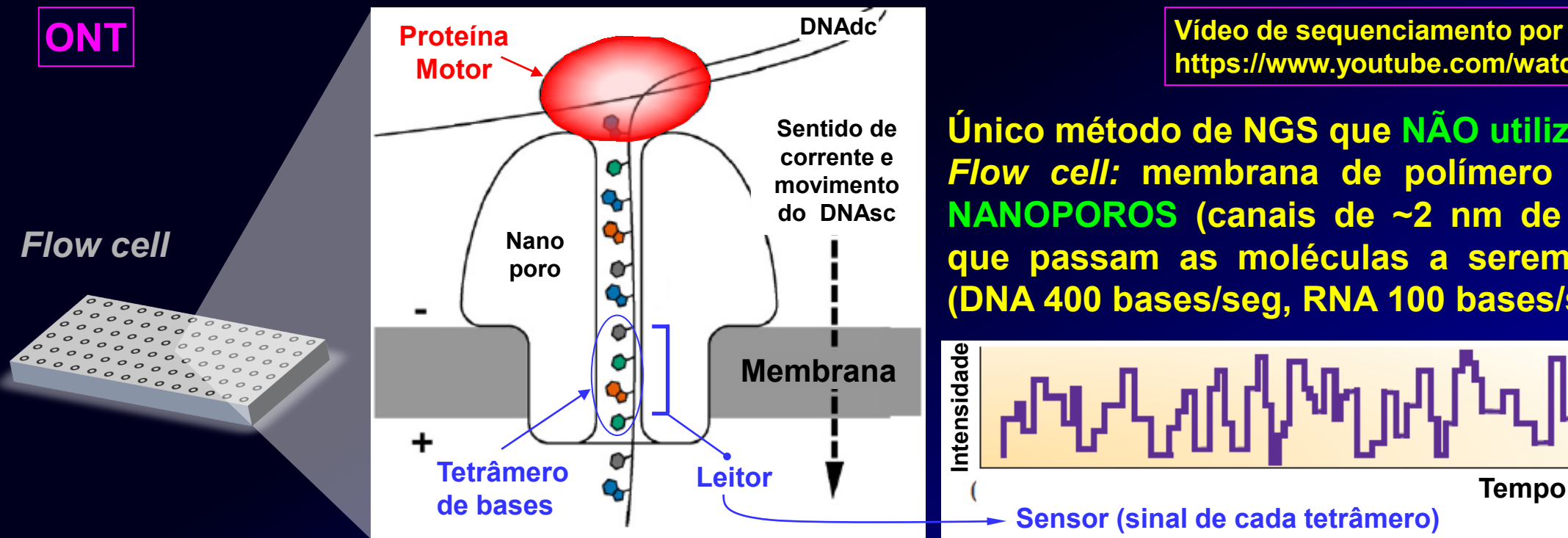
## Ion Torrent (Thermo Fisher Scientific):

- ✓ Sequenciamento por síntese e em ciclos
- ✓ Diferença com Illumina: %dNTP NÃO é detectado por fluorescência, mas pelo íon H<sup>+</sup> liberado durante sua incorporação → mini variação de 0,02 unidades de pH



## B) Sequenciamento de 3.<sup>a</sup> geração

**SMS ou Single-Molecule Real-Time (SMRT) Sequencing:** sequência é determinada sobre uma **molécula ÚNICA**  
→ **NÃO** existe amplificação da molécula a ser sequenciada



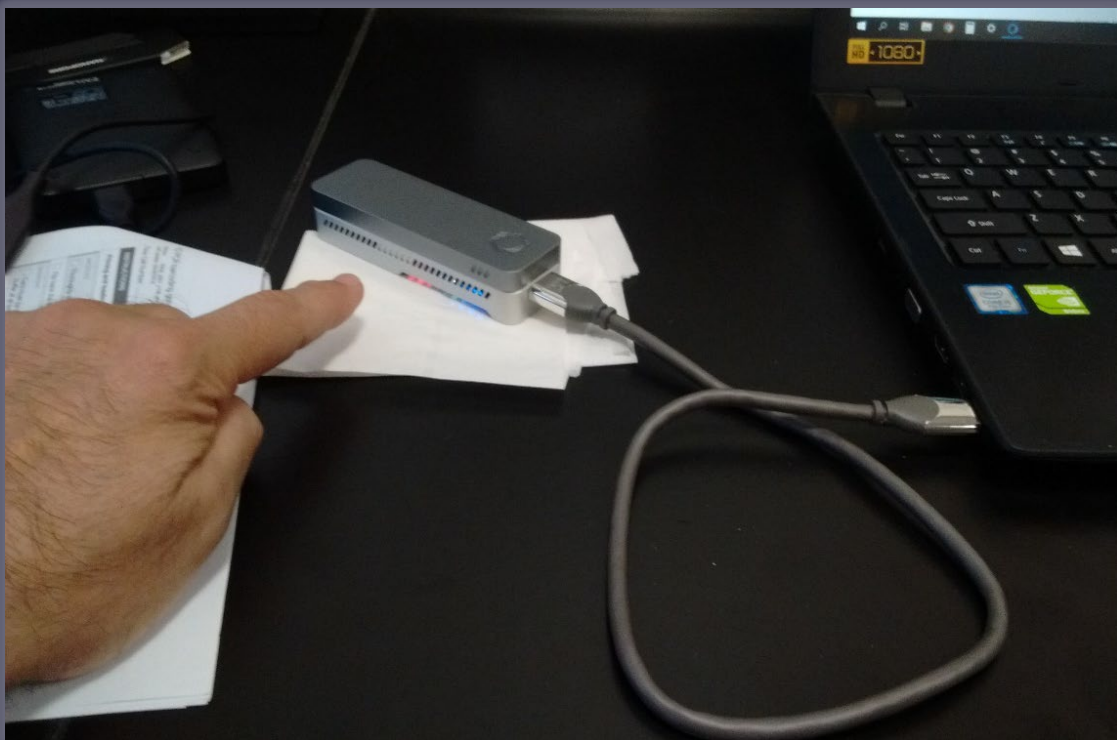
Vídeo de sequenciamento por ONT:  
[https://www.youtube.com/watch?v=iT\\_A\\_ucWMI](https://www.youtube.com/watch?v=iT_A_ucWMI)

**Único método de NGS que NÃO utiliza síntese DNA**  
**Flow cell:** membrana de polímero sintético, com **NANOPOROS** (canais de ~2 nm de diâmetro), por que passam as moléculas a serem sequenciadas (DNA 400 bases/seg, RNA 100 bases/seg)

**Extração DNA/RNA:** manual, automatizada. Determina-se pureza e concentração, bem como **integridade do DNA/RNA:** o comprimento das moléculas a serem sequenciadas deve ser o maior possível  
**Library:** pode incluir *barcoding*, e utilizando o formato 2D é possível sequenciar ambas cadeias de DNA

**Sequenciamento:** cada tetrâmero de bases da molécula interage com um **Leitor** e gera uma variação na microcorrente de íons que atravessa o nanoporo (picoA). A forma da variação (intensidade e tempo) registrada pelo **Sensor** é específica para a sequência desse tetrâmero

## B. Sequenciamento de 3ª geração, cont'



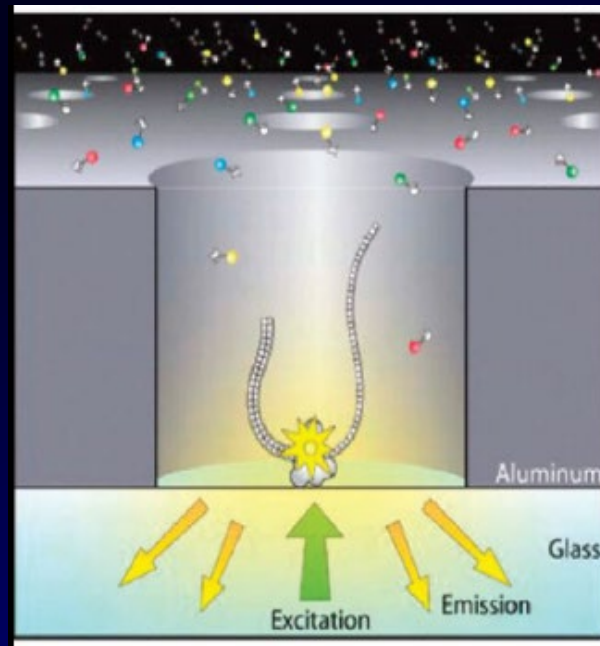
**Sequenciador *MinION Mk1B* (ONT):** portátil (3 x 10 cm); requer conexão a um computador com software necessário. Há 2 modelos novos do *MinION* que incluem estes requerimentos

↘ **Demultiplexing**

**RESULTADO** para cada amostra:  
1 único arquivo FASTQ com todas as sequências

## PacBio

Vídeo de sequenciamento por PacBio:  
<https://dnatech.genomecenter.ucdavis.edu/pacbio-sequencing/>



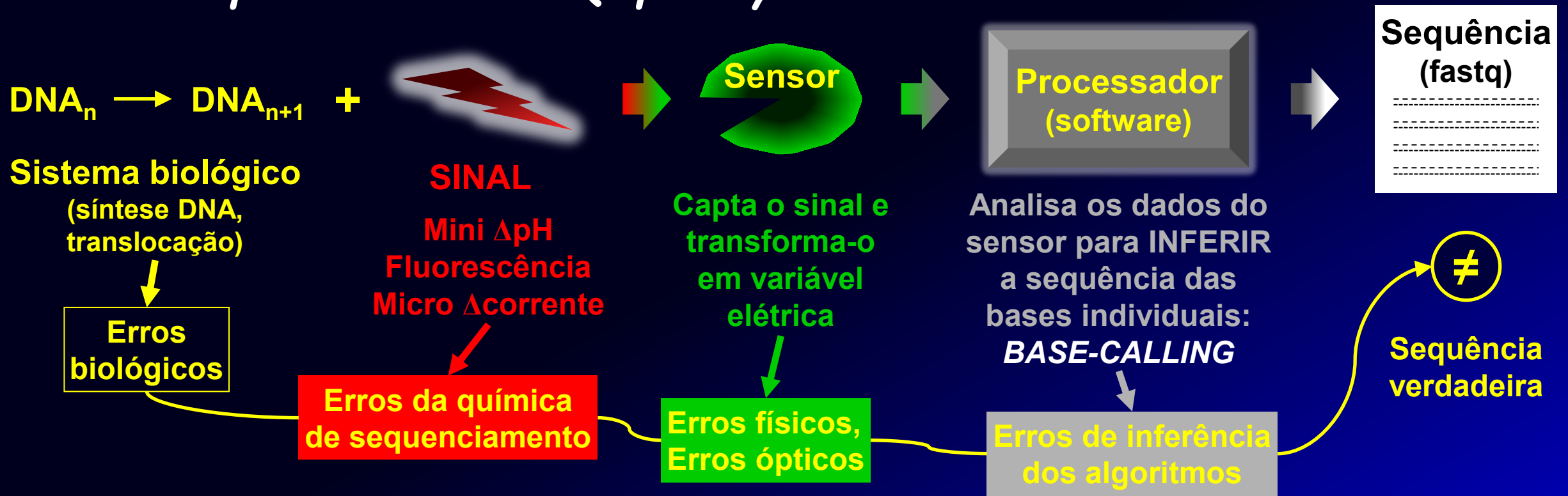
**Flow cell:** lâmina de metal com 1 milhão de poços cilíndricos (70 nm diâmetro x 100 nm alto), fundo de vidro  
**Excitação:** laser 600 nm



Design específico,  
**Zero-mode Waveguide (ZMW):** somente se detecta a luz emitida do fundo do poço

**Sequenciamento por síntese:** 1 molécula de DNAPol unida ao fundo de cada poço incorpora sequencialmente dNTPs, marcados com 4 fluoróforos diferentes. A fluorescência resultante em cada incorporação é captada por um sensor embaixo de cada poço, e o fluoróforo separa-se do dNTP depois da incorporação. Utiliza-se *barcoding* e gera 1 arquivo fastq para cada amostra

# Erro de sequenciamento: Q quality score



Dada uma probabilidade “P” de a base inferida ser ≠ da verdadeira, define-se o *Phred score*, ou *Q score*:

$$Q \text{ quality score} = -10 \times \log_{10} P \quad \leftrightarrow \quad P = 10^{-Q/10}$$

| Q  | Prob. erro de <i>base-calling</i><br>Proporção | P       | Erro %<br>(P x 100) | Exatidão %<br>(100-E%) |
|----|--|---------|---------------------|------------------------|
| 1  | 8 em 10  | 0,8     | 80                  | 20                     |
| 10 | 1 em 10  | 0,1     | 10                  | 90                     |
| 20 | 1 em 100                                       | 0,01    | 1                   | 99                     |
| 30 | 1 em 1.000                                     | 0,001   | 0,1                 | 99,9                   |
| 40 | 1 em 10.000                                    | 0,0001  | 0,01                | 99,99                  |
| 50 | 1 em 100.000                                   | 0,00001 | 0,001               | 99,999                 |

Q score de Seq. de 1.<sup>a</sup>, 2.<sup>a</sup> e 3.<sup>a</sup> geração:

1.<sup>a</sup> geração, Sanger: **Q = 50**  
 2.<sup>a</sup> geração, Illumina: **Q = 30**  
 3.<sup>a</sup> geração, PacBio HiFi: **Q = 33**  
 3.<sup>a</sup> geração, ONT: **Q = 13**    **Q = 21**    **Q = 31**  
    **2018**    **Nov2021**    **2023**

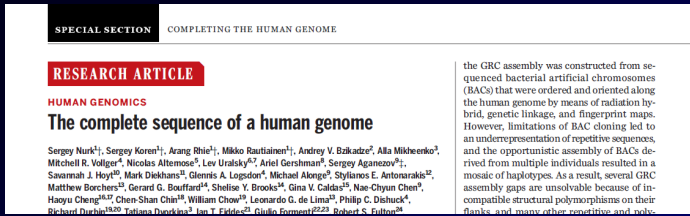


# Massividade, tempo de realização e custo

A exatidão de NGS melhora em muito alto grau com o aumento na cobertura

## Projeto GENOMA HUMANO

- ✓ 20 instituições, 6 países: EUA, Reino Unido, Japão, França, Alemanha, China
- ✓ Duração: 13 anos (1990-2003)
- ✓ Custo: USD 2.700 milhões
- ✓ 92% do genoma resolvido; 8% restante somente pôde ser resolvido mediante NGS: PacBio e ONT



Science 376(6588): 44-53,  
1.º abril 2022

## ILLUMINA

Sequenciador  
*MiSeq*

Output máximo: 25 x 10<sup>6</sup> reads (R1 ou R2)  
química 2 x 300 bases

15 Gb → 30 genomas 5 Mb, 100x

Tempo total: 40 h

Sequenciador  
*NovaSeq 6000*

Output máximo: 10.000 x 10<sup>6</sup> reads (R1 ou R2)  
química 2 x 150 bases

3.000 Gb → 45 genomas 3.300 Mb, 20x

Tempo total: 44 h

## ONT

Sequenciador  
*MinION*

Output máximo: 30 Gb → 60 genomas 5 Mb, 100x

Tempo total: 72 h

Sequenciador  
*PromethION*

Output máximo: 48 flow cells, 200 Gb <sup>c/u</sup>

9.600 Gb → 145 genomas 3.300 Mb, 20x

Tempo total: 72 h

Custo NGS para um genoma humano: ~USD 1.000

# Massividade e cobertura, exemplos de ATB (WGS)

| <b>ILLUMINA</b>      | N.º reads<br>(1 fastq) | Comprimento<br><i>read</i> | N.º total<br>de bases | Genoma *<br>(bases) | <b>Cobertura</b> | Tamanho fastq<br>(MB) |
|----------------------|------------------------|----------------------------|-----------------------|---------------------|------------------|-----------------------|
| <i>E. coli</i>       | 341.982                | 2 x 300                    | 205.189.200           | 4.700.000           | 44x              | 214                   |
| <i>P. stuartii</i>   | 616.442                | 2 x 300                    | 369.865.200           | 4.400.000           | 84x              | 375                   |
| <i>K. pneumoniae</i> | 2.405.979              | 2 x 150                    | 721.793.700           | 5.600.000           | 129x             | 870                   |
| <i>P. mirabilis</i>  | 4.893.932              | 2 x 150                    | 1.468.179.600         | 4.100.000           | 358x             | 1.702                 |

| <b>PACBIO</b>        | N.º reads | N.º total<br>de bases | Compr. médio<br><i>read</i> (bases) | Compr. máx.<br><i>read</i> (bases) | Genoma *<br>(bases) | <b>Cobertura</b> | Tamanho fastq<br>(MB) |
|----------------------|-----------|-----------------------|-------------------------------------|------------------------------------|---------------------|------------------|-----------------------|
| <i>K. pneumoniae</i> | 256.485   | 855.936.460           | 3.337                               | 40.312                             | 5.600.000           | 153x             | 1.683                 |
| <i>K. pneumoniae</i> | 653.727   | 1.852.596.401         | 2.834                               | 37.152                             | 5.600.000           | 331x             | 3.648                 |

| <b>ONT</b>            | N.º reads | N.º total<br>de bases | Comp. médio<br><i>read</i> (bases) | Compr. máx.<br><i>read</i> (bases) | Genoma *<br>(bases) | <b>Cobertura</b> | Tamanho fastq<br>(MB) |
|-----------------------|-----------|-----------------------|------------------------------------|------------------------------------|---------------------|------------------|-----------------------|
| <i>K. pneumoniae</i>  | 70.943    | 677.068.656           | 9.544                              | 297.467                            | 5.600.000           | 121x             | 1.333                 |
| <i>E. coli</i> mínimo | 11.169    | 109.816.285           | 4.963                              | 96.582                             | 4.719.000           | 22x              | 223                   |
| (n=32) ** máximo:     | 126.342   | 1.445.601.847         | 16.550                             | 197.706                            | 5.064.500           | 289x             | 2.915                 |

\* Tamanho estimado de genoma (dados de bibliografia)

\*\* Coleção de 32 cepas de *E. coli* com mesmo cromossomo, mas ≠ plasmídeos; são indicados os valores mínimos e máximos

# O que se fazer com toda essa informação?

DNA / RNA

Sequências

Fastq

↑↑↑ quantidade de dados (arquivos enormes)

Impossibilidade de processamento manual

↑↑ quantidade recurso informático

↑↑ capacidade de armazenamento

## BIOINFORMÁTICA

- ✓ Disponibilidade de servidor informático
- ✓ Uso de **LINUX**: multitarefa, multiusuário, alta capacidade de análise
- ✓ Software e ferramentas bioinformáticas
- ✓ Capacitação de recursos humanos

Controle qualidade de reads: software *FastQC*, *PycoQC*

Tipificação de entidade sequenciada: software *Kraken*

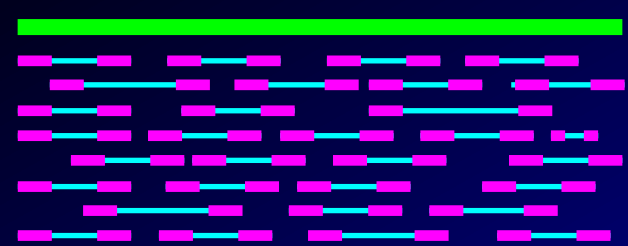
## Garbage in, garbage out

Dados de entrada de baixa qualidade, ou incorretos, produzirão resultados não confiáveis, ou errados

Fastq ↑↑ qualidade

## DUAS ALTERNATIVAS

### Mapeamento contra referência



NÃO se faz montagem, somente alinhamento de reads contra 1 sequência conhecida (referência).  
Epidemiologia: comparações de genoma, estudo de surtos

### Montagem (assembly) de novo

- ✓ NÃO existe referência
- ✓ Referência ≠ de genoma sequenciado
- ✓ É preciso genoma montado: novos mecanismos de RAM, análise de pangenoma, estudo de plasmídeos

# Análise de sequências de NGS (de novo assembly)

**Fastq** ↑↑ qualidade  
(reads)

→ **Deteccção de genes RAM: software + bases de datos (ResFinder, CARD, NCBI, etc.)**



**Software de montagem: reads curtas, ou longas, ou ambas combinadas (montagem híbrida)**



## Montagem

Genoma (+/- plasmídeos) em fragmentos lineares (*contigs*), ou em círculos (sequências “fechadas”)

As sequências da montagem são registradas em 1 só arquivo

- → **Controle de qualidade da montagem (... garbage in, garbage out)**
- → **Deteccção de genes RAM: diversos softwares + bases de datos**
- → **Análise de fatores de virulência**
- → **Comparação de sequências: BLAST**
- → **Análise de sequências de inserção / transpósons**
- → **Análise de replicons de plasmídeos**
- → **Análise de elementos de conjugação**



**Software para anotação de genomas**

**Genoma anotado**

● → **Análise de pangenoma: alinhamento de genes core, presentes em  $\geq 99\%$  dos genomas analisados, permite estabelecer relações filogenéticas entre estes genomas**

# Análise de RAM por NGS, exemplos de ATB

 AMERICAN SOCIETY FOR MICROBIOLOGY Antimicrobial Agents and Chemotherapy® [61](#): e02555-16, 2017

MECHANISMS OF RESISTANCE

***qnrE1*, a Member of a New Family of Plasmid-Located Quinolone Resistance Genes, Originated from the Chromosome of *Enterobacter* Species**

Ezequiel Albornoz,<sup>a</sup> Nathalie Tijet,<sup>b</sup> Denise De Belder,<sup>a</sup> Sonia Gomez,<sup>a</sup> Florencia Martino,<sup>a</sup> Alejandra Corso,<sup>a</sup> Roberto G. Melano,<sup>b</sup> Alejandro Petroni<sup>a</sup>

## Antecedentes:

- ✓ Isolamento clínico de *K. pneumoniae*
- ✓ Resistência de baixo nível a quinolonas, transferível por conjugação (plasmidial)
- ✓ PCRs negativas para todos os genes plasmidial de resistência aos quinolonas conhecidos até agora

Por extração DNA plasmidial, sequenciamento Illumina, montagem *de novo*, anotação de genomas e análise filogenética → identificação de *qnrE1*, uma nova família de genes transferíveis de resistência a quinolonas


## Antecedentes:

- ✓ Isolamento clínico de *K. pneumoniae subsp. quasipneumoniae*
- ✓ Resistência extrema: somente sensibilidade a tigeciclina e colistina

Por causa deste fenótipo, foi feita análise por WGS com Illumina e ONT (montagem híbrida). Em um megaplasmídeo conjugativo de 477 kb, foi identificado o gene silencioso *mcr-9*, que pode se ativar sob certas condições e confere resistência à colistina, portanto constitui um reservatório de resistência a este antibiótico indetectável com métodos fenotípicos


Infection, Genetics and Evolution 81 (2020) 104273

Contents lists available at ScienceDirect

 ELSEVIER

Infection, Genetics and Evolution


journal homepage: [www.elsevier.com/locate/meegid](http://www.elsevier.com/locate/meegid)

 MEEGID Infection, Genetics and Evolution

Short communication

Plasmid carrying *mcr-9* from an extensively drug-resistant NDM-1-producing *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* clinical isolate

Diego Faccione<sup>a,b,1</sup>, Florencia Martino<sup>a,1</sup>, Ezequiel Albornoz<sup>a</sup>, Sonia Gomez<sup>a,b</sup>, Alejandra Corso<sup>a</sup>, Alejandro Petroni<sup>a,\*</sup>

 Check for updates

# Metagenômica

## Definições:

- ✓ **Microbioma:** equivalente ao bioma para um microambiente definido. Abrange um microhábitat INTEIRO, incluindo: **(a)** TODOS os micro-organismos presentes: vírus, arqueas, bactérias, eucariotas; **(b)** TODOS seus genomas, e **(c)** as condições ambientais circundantes
- ✓ **Microbiota:** conjunto de TODOS os micro-organismo presentes em um ambiente definido  
Exemplos: intestino humano ou animal; fezes de animais; águas residuais de um hospital; chão do curral de um feedlot; águas de rios ou lagoas, etc
- ✓ **Metagenoma:** conjunto de genomas e genes dos membros de uma microbiota
- ✓ **Metagenômica:** processo utilizado para caracterizar um metagenoma, a partir do qual se obtém informação sobre as funções potenciais da microbiota



Metagenômica por NGS (*whole-metagenome shotgun sequencing*): independência de cultura e massividade permite identificar micro-organismos NÃO cultiváveis, ou presentes no microbioma em muito baixa proporção

# Metagenômica, análise de RAM na criação intensiva de bovinos

MICROBIAL DRUG RESISTANCE  
121-27. 980-990, 2021  
© Mary Ann Liebert, Inc.  
DOI: 10.1089/mdr.2020.0271

Exploring the Prevalence and Distribution Patterns of Antibiotic Resistance Genes in Bovine Gut Microbiota Using a Metagenomic Approach

**MÉTODOS:** 6 granjas da prov. de Gansu, 5 bovinos por granja:

- ✓ 2 granjas de produção intensiva de carne (Beef), 2 tambos (Dairy), 2 granjas de criação de yaks
- ✓ 1 amostra material fecal fresca (reto) / animais (30 amostras no total)
- ✓ Extração ADN → sequenciamento Illumina (2 x 150); identificação de genes de RAM e de sequências de inserção (ISs)

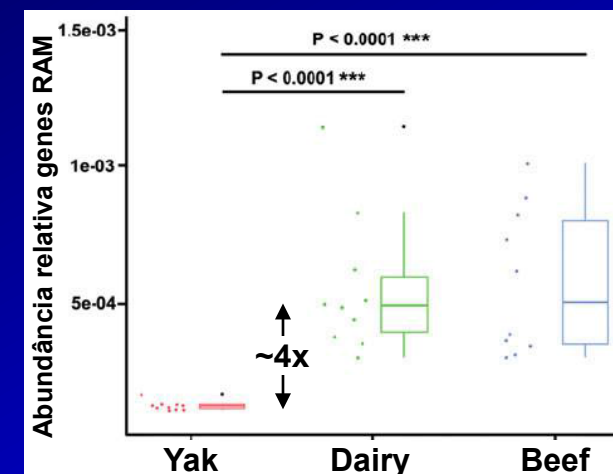
## Resultados:

- ✓ 403 Gb totais (13,4 Gb / amostra) → 4.739 genes RAM; 472 ISs
- ✓ Os genes RAM mais comuns foram: resistência a tetraciclina, seguidos por resistência a bacitracina e a  $\beta$ -lactâmicos
- ✓ Nos yaks, a abundância relativa de genes RAM foi mais de 4 vezes menor do que em Beef ou Dairy ( $p < 0,0001$ ); resultado muito semelhante para abundância de ISs

A diversidade e abundância de genes RAM foram significativamente maiores em bovinos de produção intensiva do que em yaks, com nível de abundância de ISs paralelamente maiores, em concordância com uma maior pressão antibiótica em pecuária intensiva

## Antecedentes:

- ✓ **China: 200.000 toneladas de ATB/ano usadas na indústria pecuária (uso terapêutico, profilaxia ou como promotores de crescimento)**
- ✓ **Os yaks são criados em campos de pastoreio de montanha, com muito pouco uso de ATBs: controle para estudo de impacto de ATBs na pecuária**



# Muito obrigado!!

TRABAJANDO  
JUNTOS  
PARA COMBATIR  
LA RESISTENCIA  
A LOS ANTIMICROBIANOS

