

ALCANCES DE LA SECUENCIACIÓN DE NUEVA GENERACIÓN

Dr. Alejandro Petroni
Servicio Antimicrobianos
INEI-ANIS “Dr. Carlos G. Malbrán”
apetroni@anlis.gob.ar

TRABAJANDO
JUNTOS
PARA COMBATIR
LA RESISTENCIA
A LOS ANTIMICROBIANOS



SECUENCIACIÓN



SANGER



Secuenciación de 1^{era} Generación

Secuenciación de 2^{da} Generación

Reads cortas: 50 a 600 bases

- ILLUMINA
- ION TORRENT

Diferencias con SANGER:
a) Determinación de secuencias en tiempo real
b) Decenas de miles de lecturas de ADN (reads), que se superponen y cubren TODO el target **≥30 veces**

SECUENCIACIÓN DE NUEVA GENERACIÓN

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

Si Target = todo el genoma:

- ✓ Whole Genome Sequencing (WGS)

Secuenciación de 3^{era} Generación

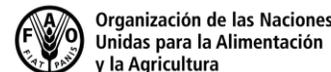
Cobertura promedio (coverage, sequencing depth): N° de veces, en promedio, que se secuencia cada base del target; se indica como "30x", "100x", etc

- OXFORD NANOPORE TECHNOLOGIES (ONT)
- PACIFIC BIOSCIENCES (PacBio)

USO DE NGS EN BACTERIOLOGÍA CLÍNICA

- ✓ **EPIDEMIOLOGÍA:**
 - **Tipificación molecular de aislamientos:** filogenia por análisis de pangenoma, o por *core genome* MLST (cgMLST): ↑↑↑ poder que MLST (miles de genes completos en lugar de fragmentos de 6-7 genes). WGS es el *gold standard* del testeo de clonalidad
 - **Estudio de brotes**
 - **Evolución y diseminación espacial de secuencias:** genomas; plásmidos y otros elementos genéticos móviles; factores de virulencia
- ✓ **PERFILES DE RAM:**
 - **Caracterización de nuevos mecanismos de RAM**
 - **Determinación genotípica de RAM:** desarrollo de sistemas de inferencia de sensibilidad a partir del genotipo
 - **Dinámica espacio-temporal de emergencia y diseminación de RAM**
- ✓ **METAGENÓMICA:**
 - **Detección e identificación DIRECTA de bacterias:** casos de infección “complejos” (endocarditis, infecciones polimicrobianas; LCR o líquido sinovial °/cultivo negativo)
 - **Caracterización de reservorios de RAM en hospitales**

TRABAJANDO
JUNTOS
PARA COMBATIR
LA RESISTENCIA
A LOS ANTIMICROBIANOS



¿Cómo se realiza una NGS?

Los esquemas de trabajo de las metodologías NGS incluyen etapas muy similares

A) Secuenciación de 2^{da} Generación, Illumina:

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

MUESTRA (bacteria aislada: *pellet* celular, cultivo ½ líquido)

1. Extracción automatizada de ADN

ADN: determinación de pureza (A_{260} , A_{280}) y cuantificación

2. Construcción de biblioteca genómica (*library*):

- ✓ Fragmentación enzimática de ADN → conjunto de fragmentos longitud promedio de 800 a 1.000 pb
- ✓ Adaptación de extremos de los fragmentos para realizar la secuenciación. Incluye la adición a cada muestra, de un *barcode ó index*: una secuencia única de 6-8 bases, que permite identificar a todas las secuencias de esa muestra. Este proceso de *barcoding* permite realizar el *multiplexing*: secuenciación conjunta de diferentes muestras en el mismo dispositivo → alta reducción de costos

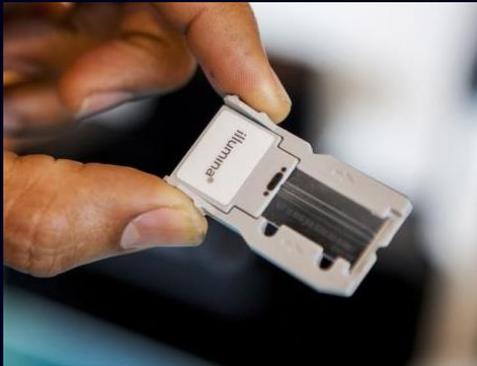
3. Secuenciación automatizada de ADN:

- ✓ Se usa una *flow cell*: dispositivo donde se produce el proceso biológico que determina la secuencia
- ✓ Amplificación clonal de ^c/fragmento
- ✓ Secuenciación por síntesis, y en ciclos (50-300): en cada ciclo se incorpora 1 de 4 dNTPs (marcados con distintos fluoróforos) y la incorporación se detecta por fluorescencia
- ✓ Secuenciación desde el otro extremo de cada fragmento: *paired-end sequencing* (2 x N^o ciclos)

A. Secuenciación de 2^{da} Generación, cont'



Secuenciador MiSeq (Illumina)



Flow cell (Illumina)

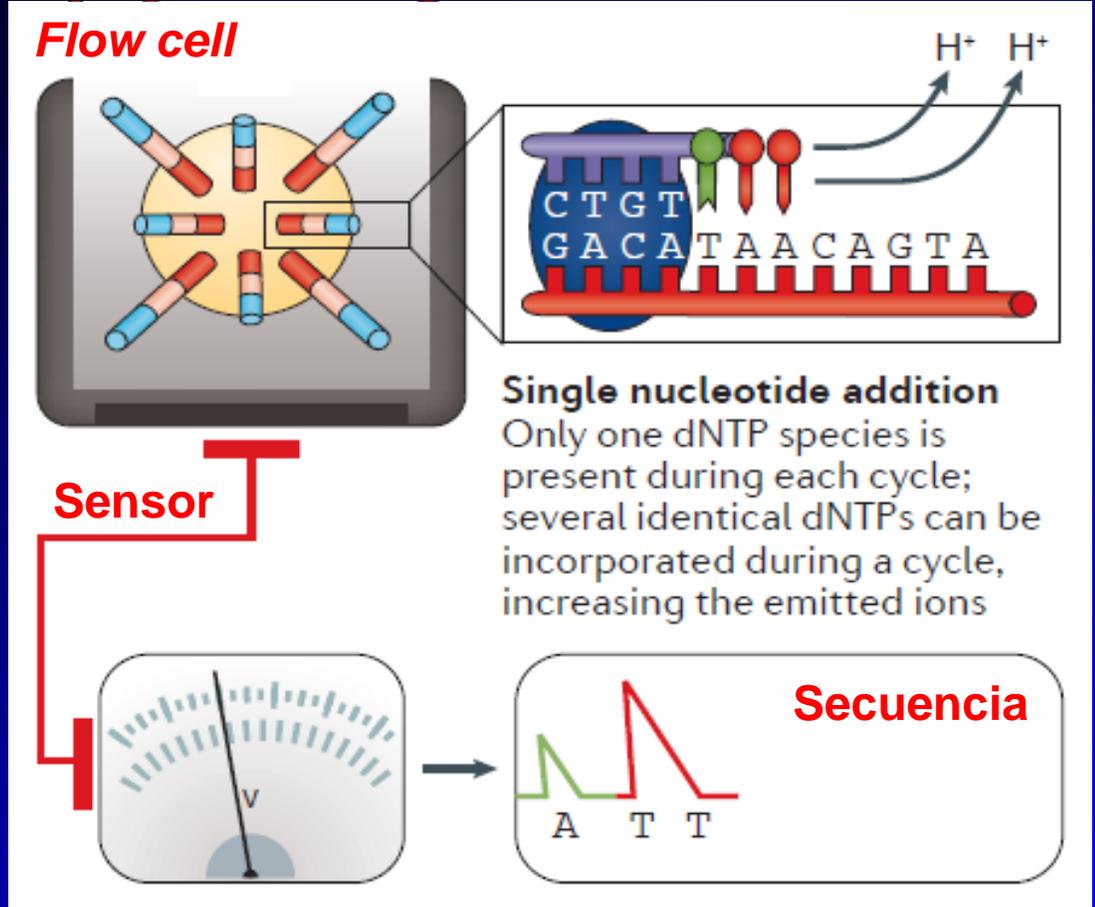
Luego de finalizada la secuenciación, un software del secuenciador realiza el **demultiplexing**: utilizando los **barcodes**, clasifica y agrupa todas las secuencias de acuerdo a la muestra de la cual provienen



RESULTADOS para cada muestra:
2 archivos de texto, R1 y R2: contienen todas las **reads** generadas desde el 1^{er} y 2^{do} extremo de todos los fragmentos secuenciados, respectivamente
Formato FASTQ: en cada **read** se indica la secuencia de bases y los valores de calidad en la determinación de cada base

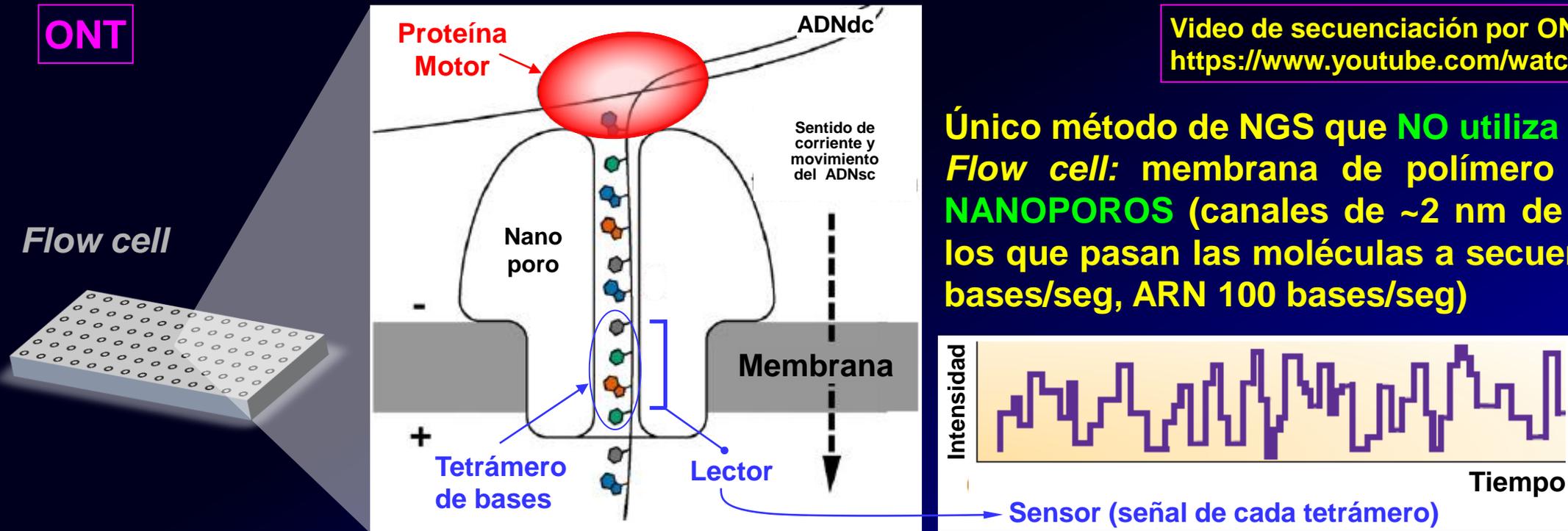
Ion Torrent (ThermoFisher Scientific):

- ✓ Secuenciación por síntesis, y en ciclos
- ✓ Diferencia con Illumina: ^c/dNTP NO se detecta por fluorescencia, sino por el ión H⁺ liberado durante su incorporación → mini variación de 0,02 unidades de pH



B) Secuenciación de 3^{era} Generación

SMS ó Single-Molecule Real-Time (SMRT) Sequencing: secuencia se determina sobre una molécula ÚNICA
→ NO hay amplificación de la molécula a secuenciar



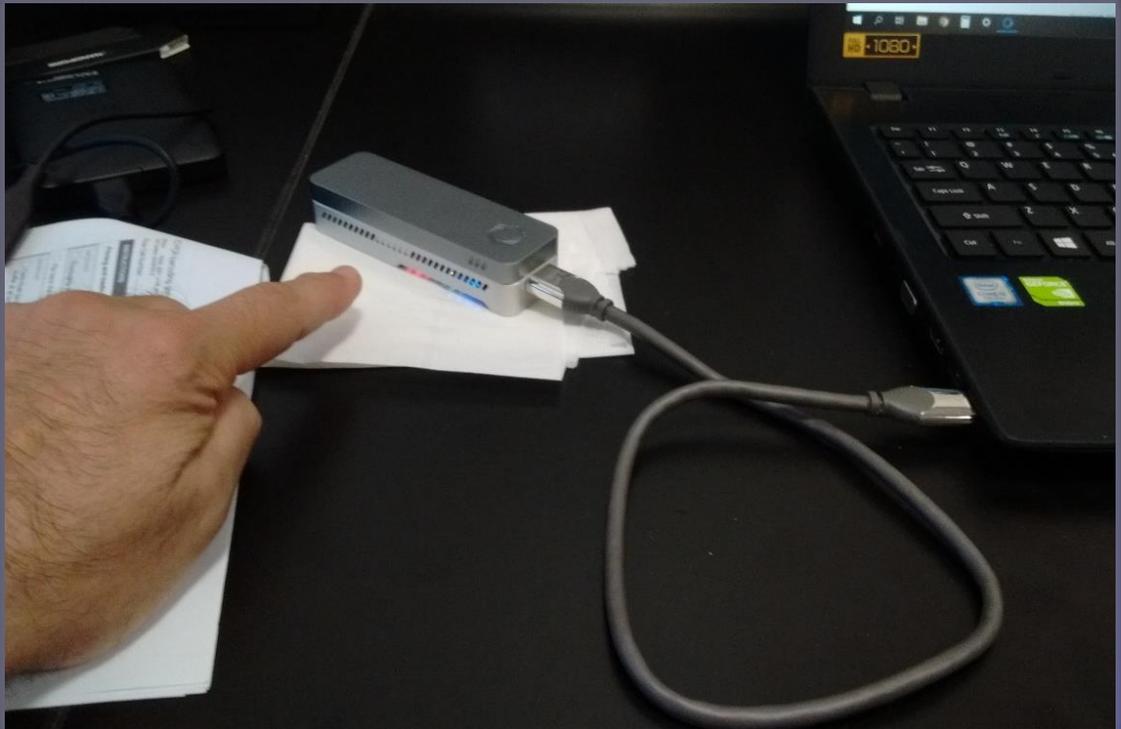
Único método de NGS que NO utiliza síntesis ADN
Flow cell: membrana de polímero sintético, con **NANOPOROS** (canales de ~2 nm de diámetro), por los que pasan las moléculas a secuenciar (ADN 400 bases/seg, ARN 100 bases/seg)

Extracción ADN/ARN: manual, automatizada. Se determina pureza y concentración, y también **integridad del ADN/ARN:** la longitud de las moléculas a secuenciar debe ser la mayor posible

Library: puede incluir **barcoding**, y utilizando el formato 2D se pueden secuenciar ambas cadenas de ADN

Secuenciación: cada tetrámero de bases de la molécula interacciona con un **Lector** y genera una variación en la micro-corriente de iones que atraviesa el nanoporo (picoA). La forma de la variación (intensidad y tiempo) que registra el **Sensor** es específica para la secuencia de ese tetrámero

B. Secuenciación de 3^{era} Generación, cont'



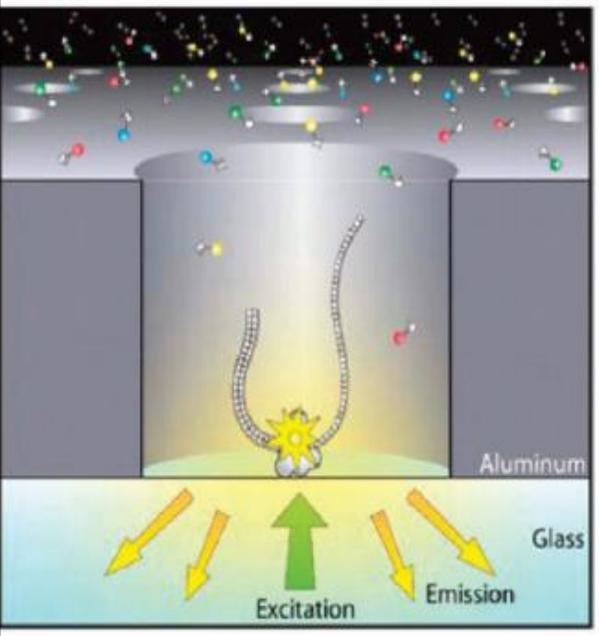
Secuenciador *MinION Mk1B* (ONT): portátil (3 x 10 cm); requiere conexión a una computadora con software necesarios. Hay 2 modelos nuevos de *MinION* que incluyen estos requerimientos

Demultiplexing

RESULTADO para cada muestra:
1 único archivo FASTQ con todas las secuencias

PacBio

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

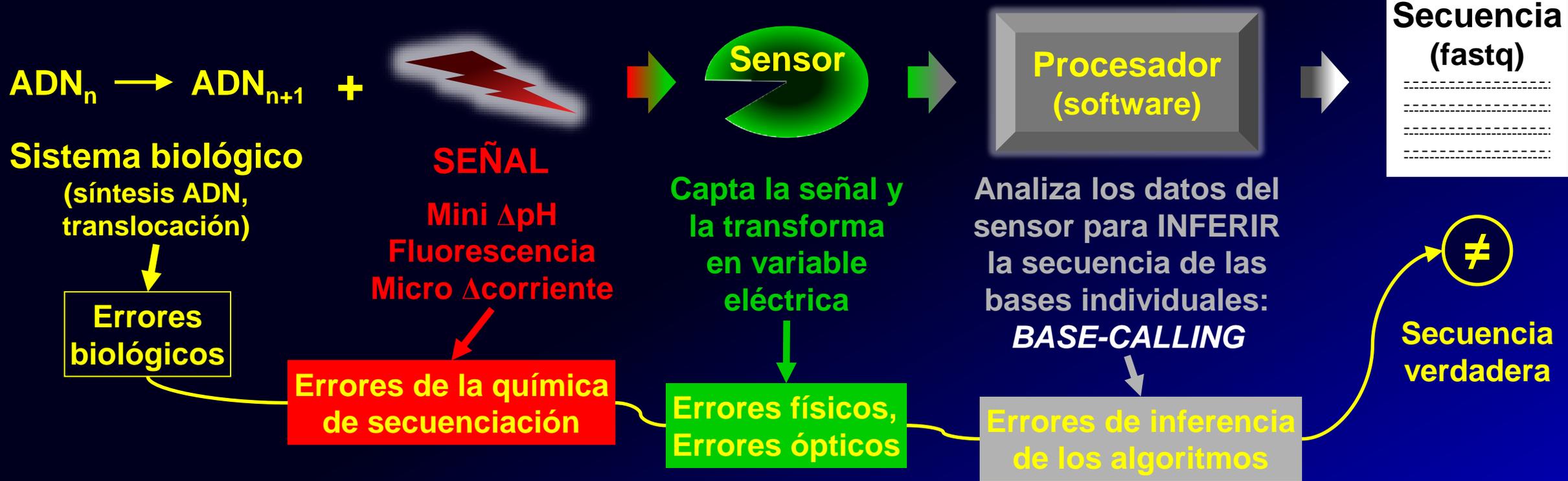


Flow cell: lámina de metal con 1 millón de pocillos cilíndricos (70 nm diámetro x 100 nm alto), fondo de vidrio
Excitación: láser 600 nm

↓
Diseño específico,
Zero-mode Waveguide (ZMW): solo se detecta la luz emitida desde el fondo del pocillo

Secuenciación por síntesis: 1 molécula de ADNpol unida al fondo de pocillo incorpora secuencialmente dNTPs, marcados con 4 fluoróforos diferentes. La fluorescencia resultante en cada incorporación se capta por un sensor debajo de cada pocillo, y el fluoróforo se separa del dNTP después de la incorporación. Se utiliza *barcoding* y genera 1 archivo fastq para cada muestra

Error de secuenciación: Q quality score



Dada una probabilidad “P” de que la base inferida sea ≠ a la verdadera, se define el *Phred score*, ó *Q score*:

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

Q	Prob. error de <i>base-calling</i> Proporción	P	Error % (P x 100)	Exactitud % (100-E%)
1	8 en 10	0,8	80	20
10	1 en 10	0,1	10	90
20	1 en 100	0,01	1	99
30	1 en 1.000	0,001	0,1	99,9
40	1 en 10.000	0,0001	0,01	99,99
50	1 en 100.000	0,00001	0,001	99,999

Q score de Sec. de 1^{era}, 2^{da} y 3^{era} Generación:

- 1^{era} Generación, Sanger: **Q = 50**
- 2^{da} Generación, Illumina: **Q = 30**
- 3^{era} Generación, PacBio HiFi: **Q = 33**
- 3^{era} Generación, ONT: **Q = 13** (2018) **Q = 21** (Nov2021) **Q = 31** (2023)

Masividad, tiempo de realización y costo

La exactitud de NGS mejora en muy alto grado con el incremento en la cobertura

Proyecto GENOMA HUMANO

- ✓ 20 instituciones, 6 países: USA, Reino Unido, Japón, Francia, Alemania, China
- ✓ Duración: 13 años (1990-2003)
- ✓ Costo: USD 2.700 millones
- ✓ 92% del genoma resuelto; 8% restante solo pudo resolverse por NGS: PacBio y ONT



ILLUMINA

Secuenciador *MiSeq*

Output máximo: 25 x 10⁶ reads (R1 ó R2) química 2 x 300 bases

15 Gb → 30 genomas 5 Mb, 100x

Tiempo total: 40 hs

Secuenciador *NovaSeq 6000*

Output máximo: 10.000 x 10⁶ reads (R1 ó R2) química 2 x 150 bases

3.000 Gb → 45 genomas 3.300 Mb, 20x

Tiempo total: 44 hs

ONT

Secuenciador *MinION*

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

Tiempo total: 72 hs

Secuenciador *PromethION*

Output máximo: 48 flow cells, 200 Gb ^o/_u

9.600 Gb → 145 genomas 3.300 Mb, 20x

Tiempo total: 72 hs

Costo NGS para un genoma humano: ~USD 1.000



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

Masividad y cobertura, ejemplos de ATB (WGS)

ILLUMINA	N° reads (1 fastq)	Longitud read	N° total de bases	Genoma * (bases)	Coverage	Tamaño fastq (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

PACBIO	N° reads	N° total de bases	Long. prom. read (bases)	Long. máx. read (bases)	Genoma * (bases)	Coverage	Tamaño fastq (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

ONT	N° reads	N° total de bases	Long. prom. read (bases)	Long. Máx. read (bases)	Genoma * (bases)	Coverage	Tamaño fastq (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

* Tamaño estimado de genoma (datos de bibliografía)

** Colección de 32 cepas de *E. coli* con mismo cromosoma, pero ≠ plásmidos; se indican los valores mínimos y máximos

¿Qué se hace con toda esta información?

ADN / ARN

Secuencias

Fastq

↑↑↑ cantidad de info (archivos enormes)

Imposibilidad de procesamiento manual

↑↑ cantidad recurso informático

↑↑ capacidad de almacenamiento

BIOINFORMÁTICA

- ✓ Disponibilidad de servidor informático
- ✓ Uso de **LINUX**: multi-tarea, multi-usuario, alta capacidad de análisis
- ✓ Software y herramientas bioinformáticas
- ✓ Capacitación de recursos humanos

Control calidad de reads: software *FastQC*, *PycoQC*

Tipificación de entidad secuenciada: software *Kraken*

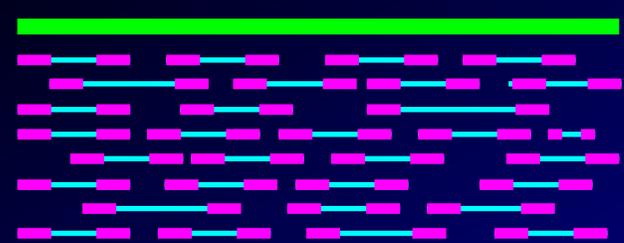
Garbage in, garbage out

Datos de entrada de baja calidad, o incorrectos, producirán resultados no confiables, o erróneos

Fastq ↑↑ calidad

DOS ALTERNATIVAS

Mapeo contra referencia



NO se realiza ensamblado, solo alineamiento de reads contra 1 secuencia conocida (referencia).
Epidemiología: comparaciones de genomas, estudio de brotes

Ensamble (*assembly*) de novo

- ✓ NO existe referencia
- ✓ Referencia ≠ de genoma secuenciado
- ✓ Se requiere genoma ensamblado: nuevos mecanismos de RAM, análisis de pangenoma, estudio de plásmidos

Análisis de secuencias de NGS (de novo assembly)

Fastq ↑↑ **calidad**
(reads)

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



Software ensamblador: reads cortas, o largas, o ambas combinadas (ensamble híbrido)



Ensamblado

Genoma (+/- plásmidos) en fragmentos lineales (contigs), o en círculos (secuencias "cerradas")

Las secuencias del ensamblado se registran en 1 solo archivo

- → **Control de calidad del ensamblado (... garbage in, garbage out)**
- → **Detección de genes RAM: diversos software + bases de datos**
- → **Análisis de factores de virulencia**
- → **Comparación de secuencias: BLAST**
- → **Análisis de secuencias de inserción / transposones**
- → **Análisis de replicones de plásmidos**
- → **Análisis de elementos de conjugación**



Software para anotación de genomas

Genoma anotado

● → **Análisis de pangenoma: alineamiento de genes core, presentes en $\geq 99\%$ de los genomas analizados, permite establecer relaciones filogenéticas entre estos genomas**

Análisis de RAM por NGS, ejemplos 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,^a Nathalie Tijet,^b Denise De Belder,^a Sonia Gomez,^a Florencia Martino,^a Alejandra Corso,^a Roberto G. Melano,^b Alejandro Petroni^a

Antecedentes:

- ✓ Aislamiento clínico de *K. pneumoniae*
- ✓ Resistencia de bajo nivel a quinolonas, transferible por conjugación (plasmídica)
- ✓ PCRs negativas para todos los genes plasmídicos de resistencia a quinolonas conocidos hasta el momento

Por extracción ADN plasmídico, secuenciación Illumina, ensamble *de novo*, anotación de genomas y análisis filogenético → identificación de *qnrE1*, una nueva familia de genes transferibles de resistencia a quinolonas

Antecedentes:

- ✓ Aislamiento clínico de *K. quasipneumoniae* subsp. *quasipneumoniae*
- ✓ Resistencia extrema: solo sensibilidad a tigeciclina y colistina

Debido a este fenotipo, se analizó por WGS con Illumina y ONT (ensamble híbrido). En un megaplásmido conjugativo de 477 kb, se identificó el gen silencioso *mcr-9*, que puede

activarse en ciertas condiciones, confirmando resistencia a colistina, por lo que constituye un reservorio de resistencia a este antibiótico indetectable con 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



Short communication

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

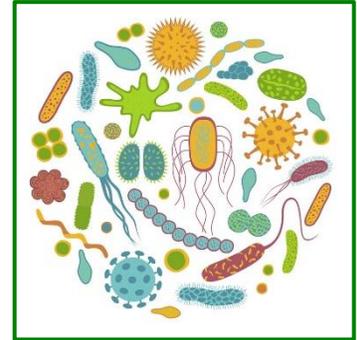
Diego Faccione^{a,b,1}, Florencia Martino^{a,1}, Ezequiel Albornoz^a, Sonia Gomez^{a,b}, Alejandra Corso^a, Alejandro Petroni^{a,*}



Metagenómica

Definiciones:

- ✓ **Microbioma:** equivalente al bioma para un microambiente definido. Comprende un microhábitat ENTERO, incluyendo: **(a)** TODOS los microorganismos presentes: virus, arqueas, bacterias, eucariotas; **(b)** TODOS sus genomas, y **(c)** las condiciones ambientales circundantes
- ✓ **Microbiota:** conjunto de TODOS los microorganismos presentes en un ambiente definido
Ejemplos: intestino humano, o animal; heces de animales; aguas residuales de un hospital; suelo del corral de un feedlot; aguas de ríos o lagunas; etc
- ✓ **Metagenoma:** conjunto de genomas y genes de los miembros de una microbiota
- ✓ **Metagenómica:** proceso utilizado para caracterizar un metagenoma, a partir del cual se obtiene información sobre las funciones potenciales de la microbiota



Metagenómica por NGS (*whole-metagenome shotgun sequencing*): independencia de cultivo y masividad permite identificar microorganismos NO cultivables, o presentes en el microbioma en muy baja proporción

Metagenómica, análisis de RAM en cría intensiva de bovinos

MICROBIAL DRUG RESISTANCE
27(7): 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 de prov. de Gansu, 5 bovinos por granja:

- ✓ 2 granjas de producción intensiva de carne (Beef), 2 tambos (Dairy), 2 granjas de cría de yaks
- ✓ 1 muestra material fecal fresca (recto) / animal (30 muestras en total)
- ✓ Extracción ADN → secuenciación Illumina (2 x 150); identificación de genes de RAM y de secuencias de inserción (ISs)

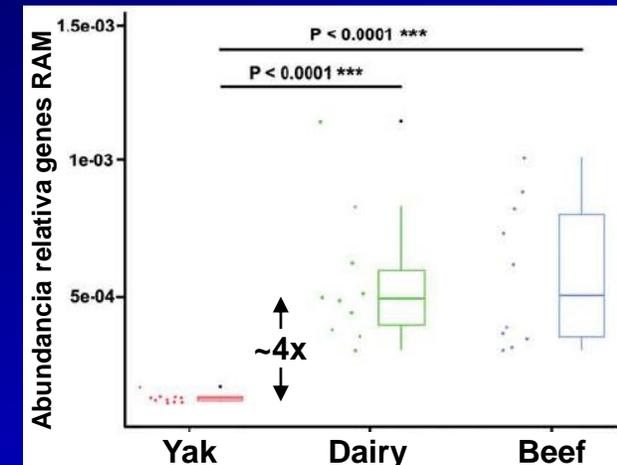
Resultados:

- ✓ 403 Gb totales (13,4 Gb / muestra) → 4.739 genes RAM; 472 ISs
- ✓ Los genes RAM más comunes fueron: resistencia a tetraciclina, seguidos por resistencia a bacitracina y a β -lactámicos
- ✓ En yaks, la abundancia relativa de genes RAM fue más de 4 veces menor que en Beef o Dairy ($p < 0,0001$); resultado muy similar para abundancia de ISs

La diversidad y abundancia de genes RAM fueron significativamente mayores en bovinos de producción intensiva que en yaks, con niveles de abundancia de ISs paralelamente mayores, en concordancia con una mayor presión antibiótica en ganadería intensiva

Antecedentes:

- ✓ China: 200.000 toneladas de ATB/año usadas en industria ganadera (uso terapéutico, profilaxis o como promotores de crecimiento)
- ✓ Los yaks se crían en campos de pastoreo de montaña, con muy poco uso de ATBs: control para estudio de impacto de ATBs en ganadería



Muchas gracias!!

TRABAJANDO
JUNTOS
PARA COMBATIR
LA RESISTENCIA
A LOS ANTIMICROBIANOS

