

Ingeniería Genética II

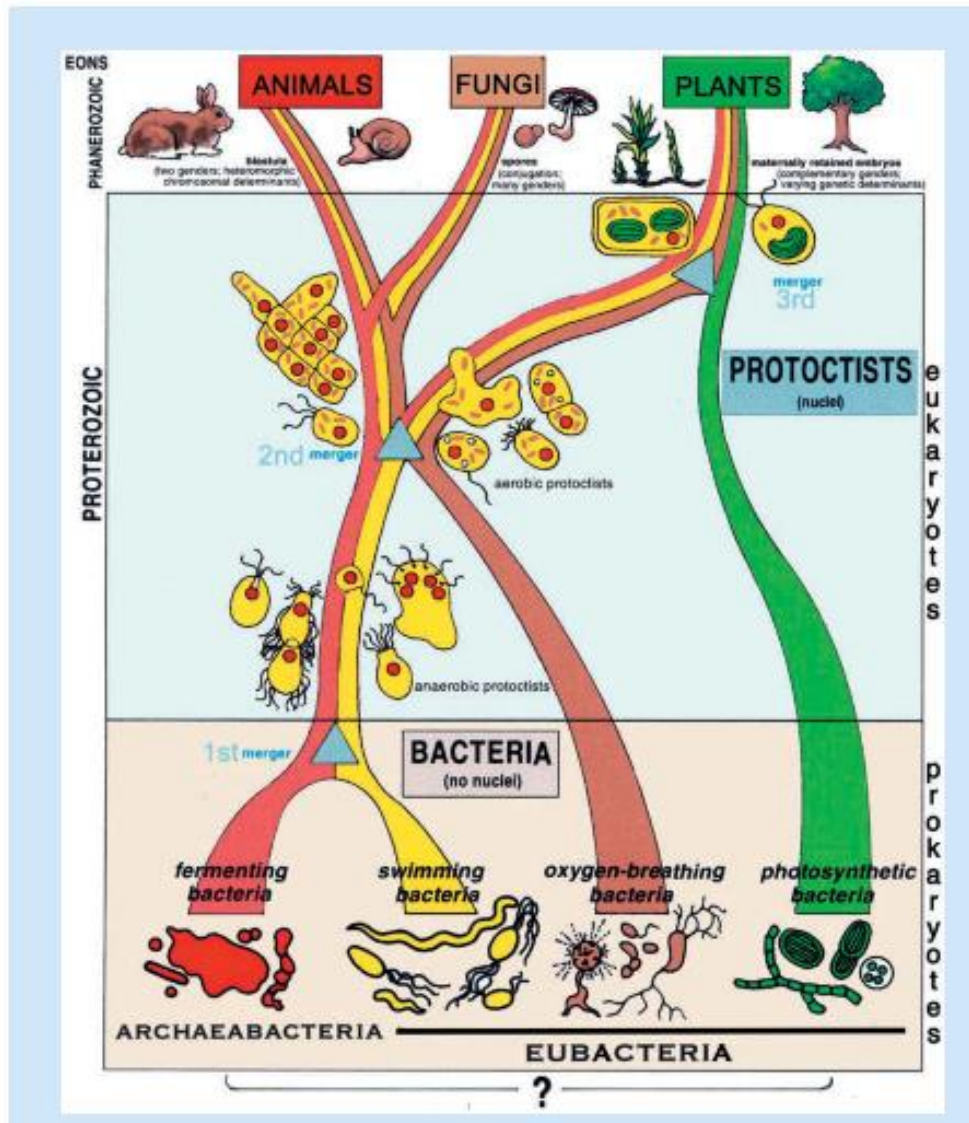
Unidad III

Nuevas tecnologías de secuenciación

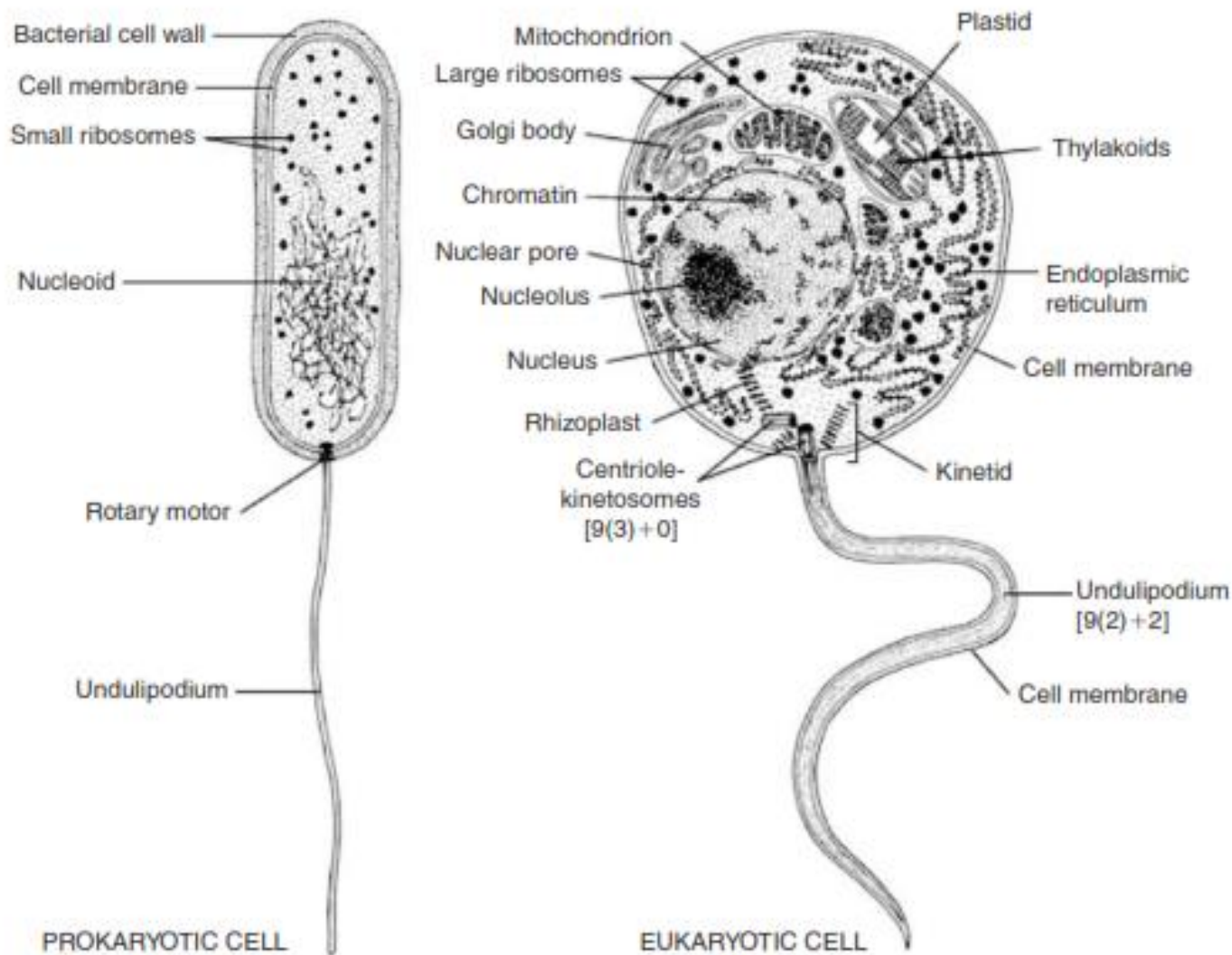
Genética e Ingeniería Genética



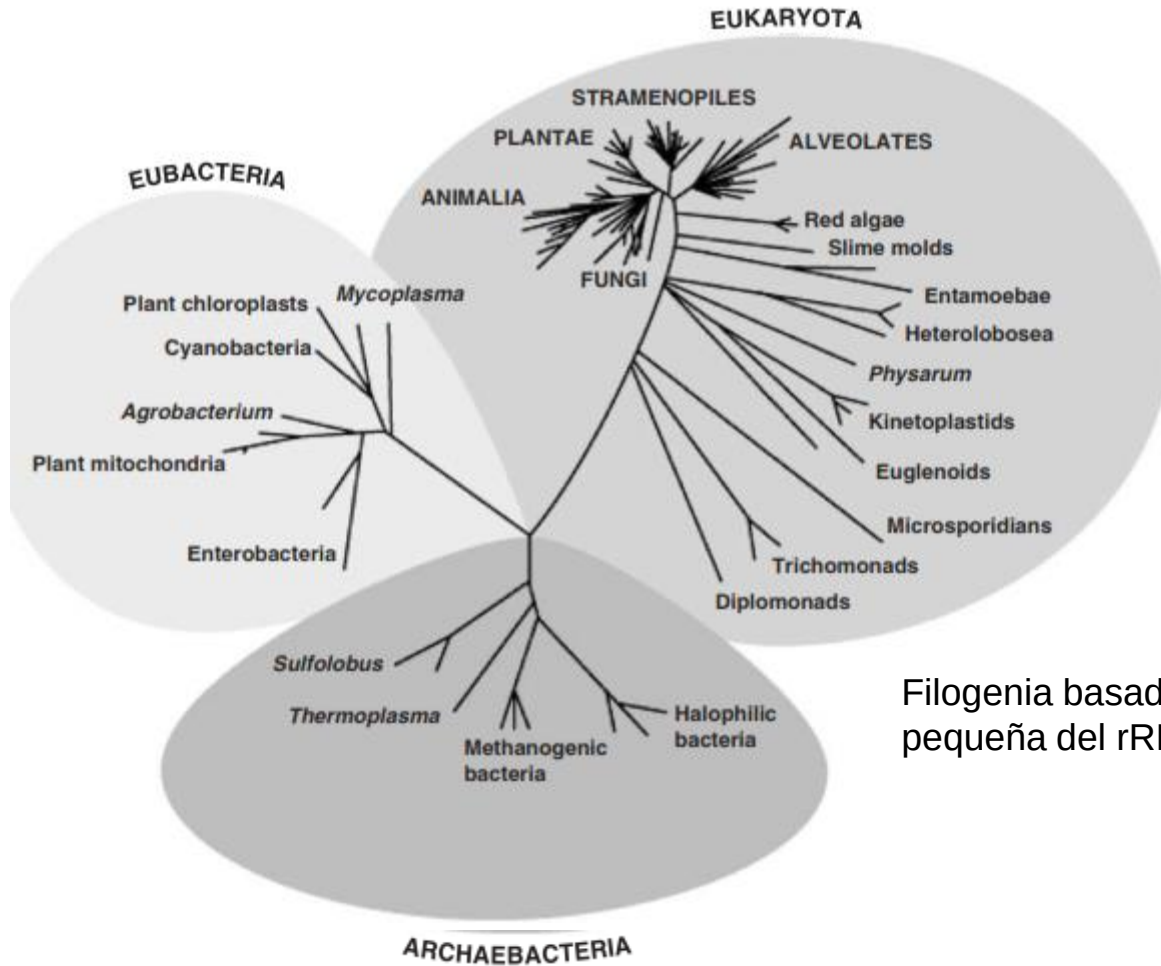
La materia viva



La materia viva



La materia viva



Filogenia basada en la subunidad pequeña del rRNA

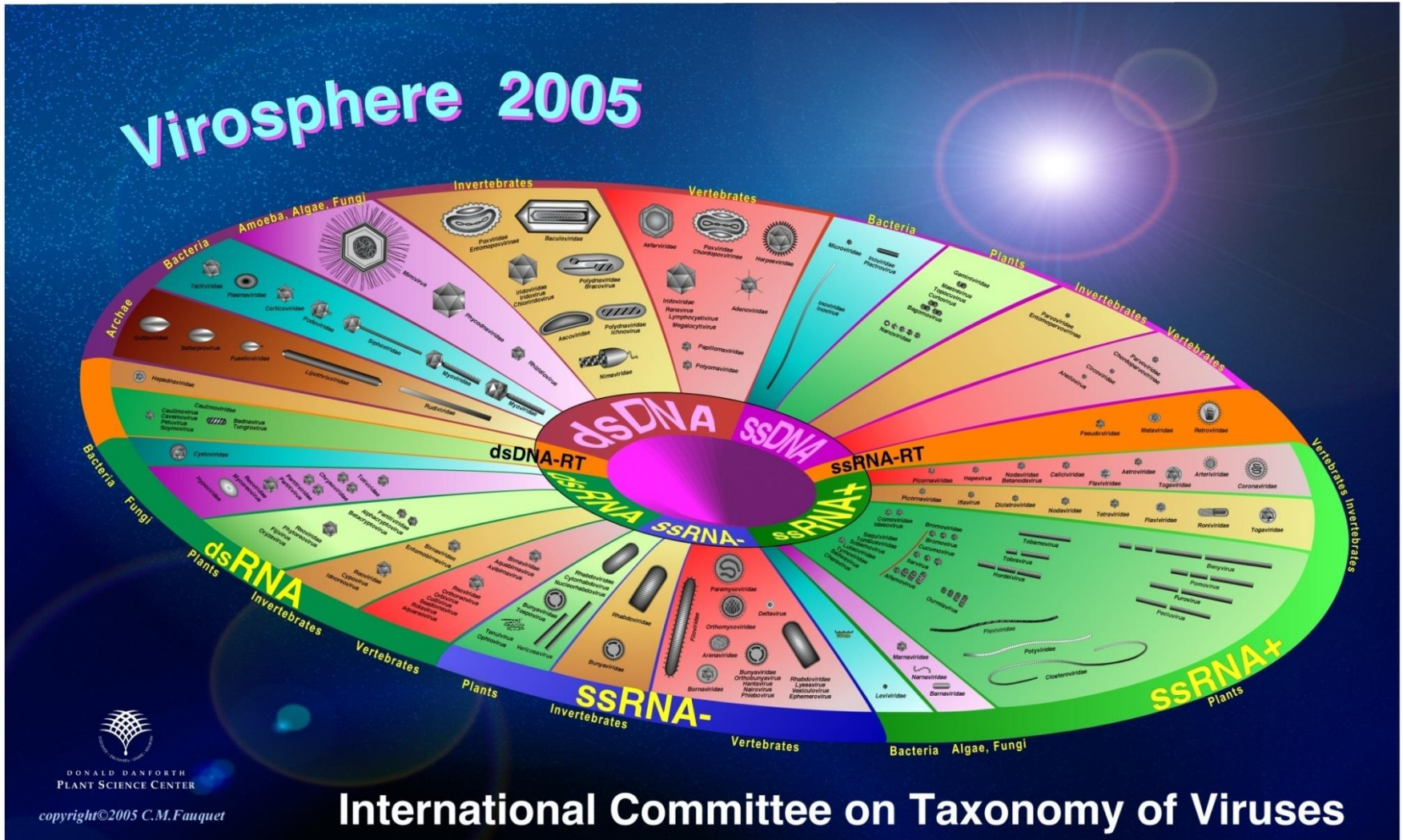
La materia viva: funciones básicas

La materia viva es un sistema interdependiente de macromoléculas, que posee las siguientes funciones mínimas esenciales:

- *Transformación y almacenamiento de energía*
- *Síntesis de materia (sostenimiento)*
- *Capacidad de perpetuación (con cambios)*
- *Capacidad de respuesta al entorno*

+ *Funciones auxiliares*

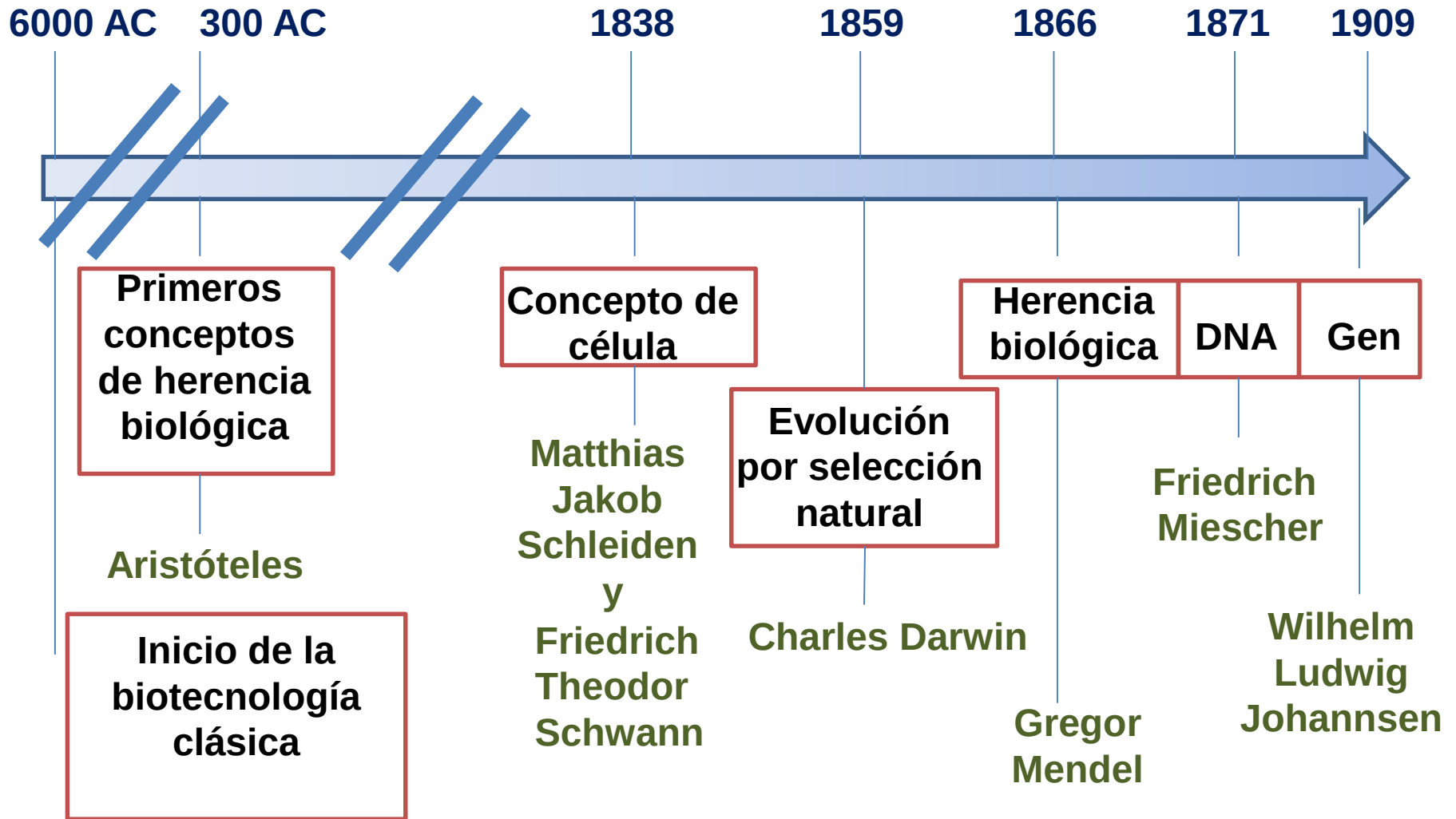
Virus



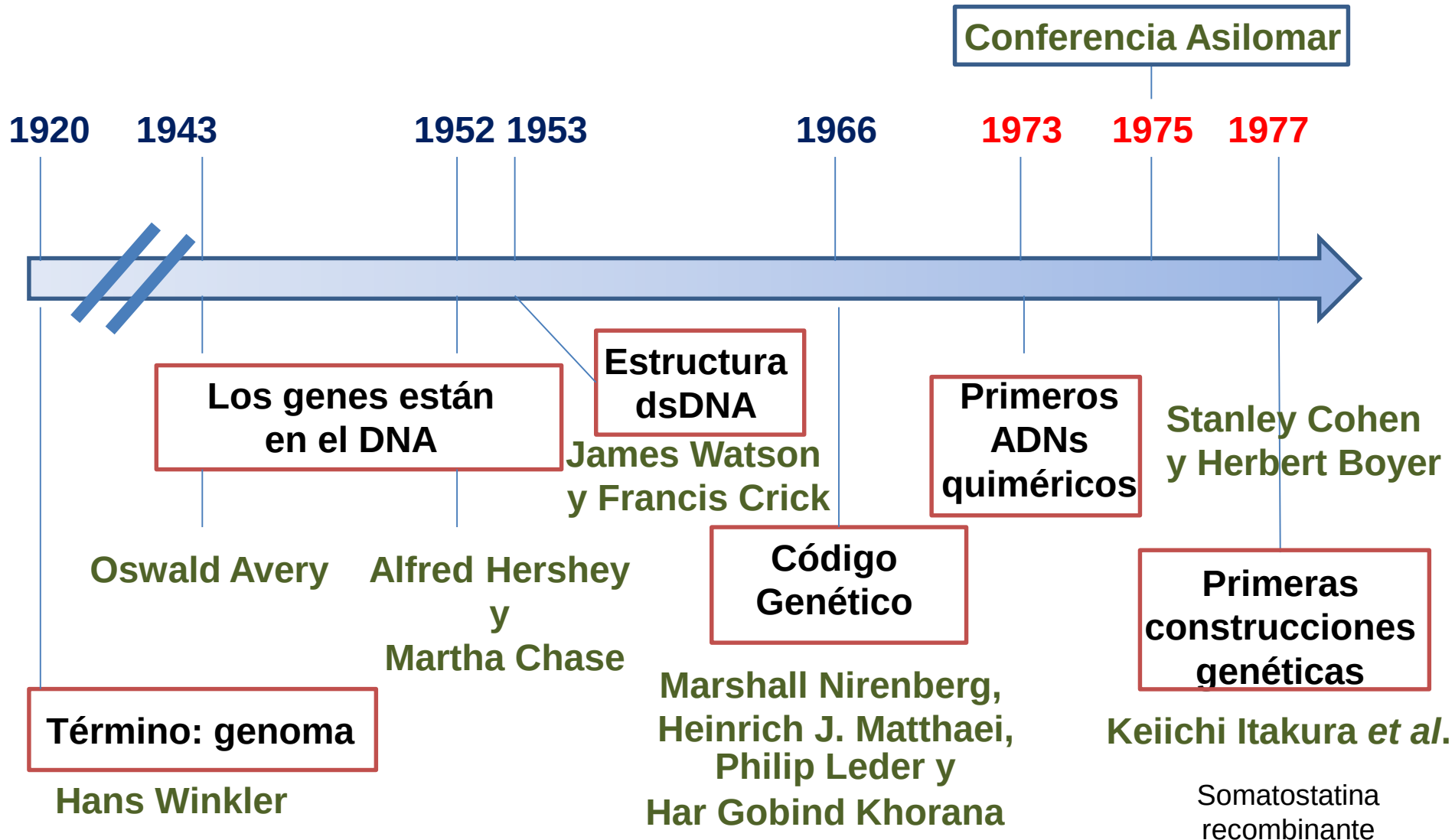
Entidades que son **parásitos obligados de la vida** (al menos carecen de la capacidad de transformar la energía).

Son considerados **miembros del Moviloma**.

Hacia el camino de entender la información biológica



Hacia el camino de entender la información biológica



Ciencias y Tecnologías de la vida

GENÉTICA

Estudio de:

- *...la información contenida en la materia viva;*
- *su sintaxis;*
- *y sus mecanismos de transmisión.*



CONOCIMIENTO

INGENIERÍA GENÉTICA

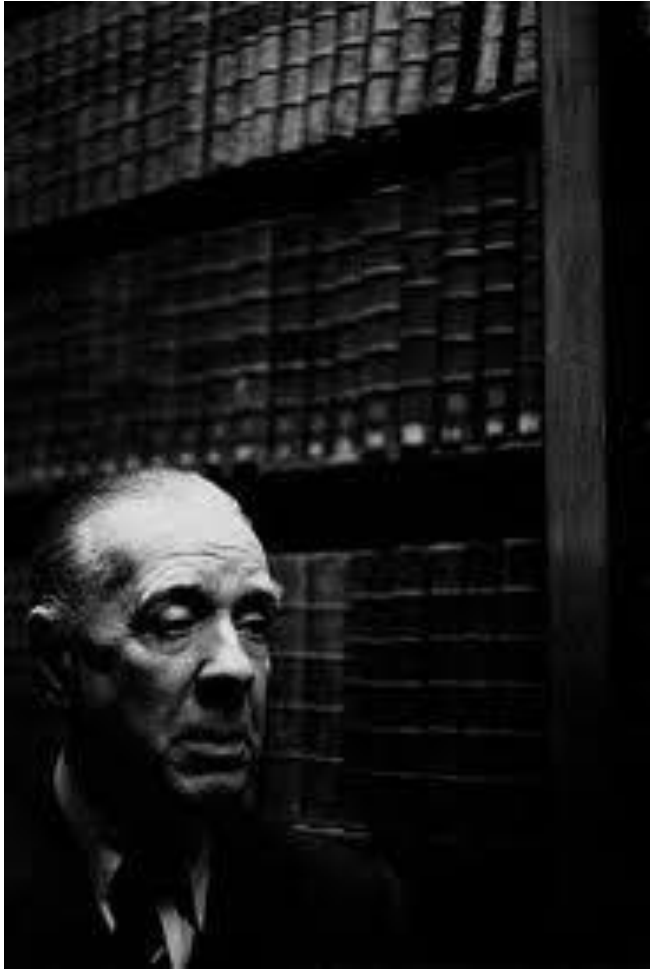
Aplicación basada en la:

- *...creación o recreación de información genética;*
- *respetando sintaxis;*
- *e implantándola en materia viva.*



BIEN/SERVICIO

Necesidad de conocer secuencias



*Que otros se jacten de las páginas que han escrito;
a mí me enorgullecen las que he leído.*

*No habré sido un filólogo,
no habré inquirido las declinaciones, los modos, la laboriosa
mutación de las letras,
la de que se endurece en te,
la equivalencia de la ge y de la ka,
pero a lo largo de mis años he profesado
la pasión del lenguaje.*

*Mis noches están llenas de Virgilio;
haber sabido y haber olvidado el latín
es una posesión, porque el olvido
es una de las formas de la memoria, su vago sótano,
la otra cara secreta de la moneda.*

*Cuando en mis ojos se borraron
las vanas apariencias queridas,
los rostros y la página,
me di al estudio del lenguaje de hierro
que usaron mis mayores para cantar
espadas y soledades,*

*y ahora, a través de siete siglos,
desde la última Thule,
tu voz me llega, Snorri Sturluson.*

*El joven, ante el libro, se impone una disciplina precisa
y lo hace en pos de un conocimiento preciso;
a mis años, toda empresa es una aventura
que linda con la noche.*

*No acabaré de descifrar las antiguas lenguas del Norte,
no hundiré las manos ansiosas en el oro de Sigurd;
la tarea que emprendo es ilimitada
y ha de acompañarme hasta el fin,
no menos misteriosa que el universo
y que yo, el aprendiz.*

Jorge Luis Borges

Adquisición de información genómica

- Uno de los **objetivos** primarios de la **Ingeniería Genética** es la necesidad de **conocer información genómica**.
- Todos los **organismos y virus** poseen **material genético** basado en moléculas de **ácidos nucleicos** (dsDNA en los organismos, ssDNA/dsDNA/ssRNA/dsRNA en los virus).
- La **secuencia primaria** contenida en esas moléculas y su posterior **interpretación** es una **f fuente de conocimiento y de potenciales múltiples aplicaciones**.

Adquisición de información genómica

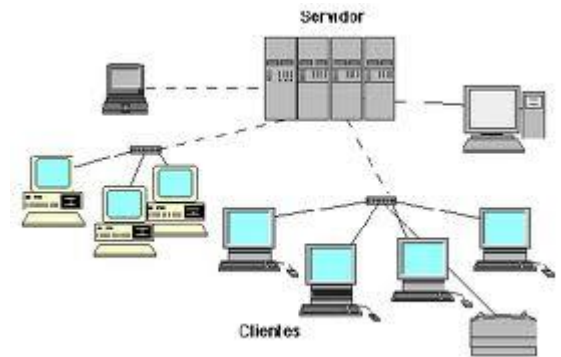
Entidades biológicas



Proyectos genomas



Bases de Datos



La información genómica

Algunos conceptos

GENOMA

*Conjunto de la **información** que soporta a **una materia viva** y que es **verticalmente transmisible**.*

PANGENOMA

*Información genómica presente en el **conjunto** de los **individuos** de una misma **especie**.*

METAGENOMA

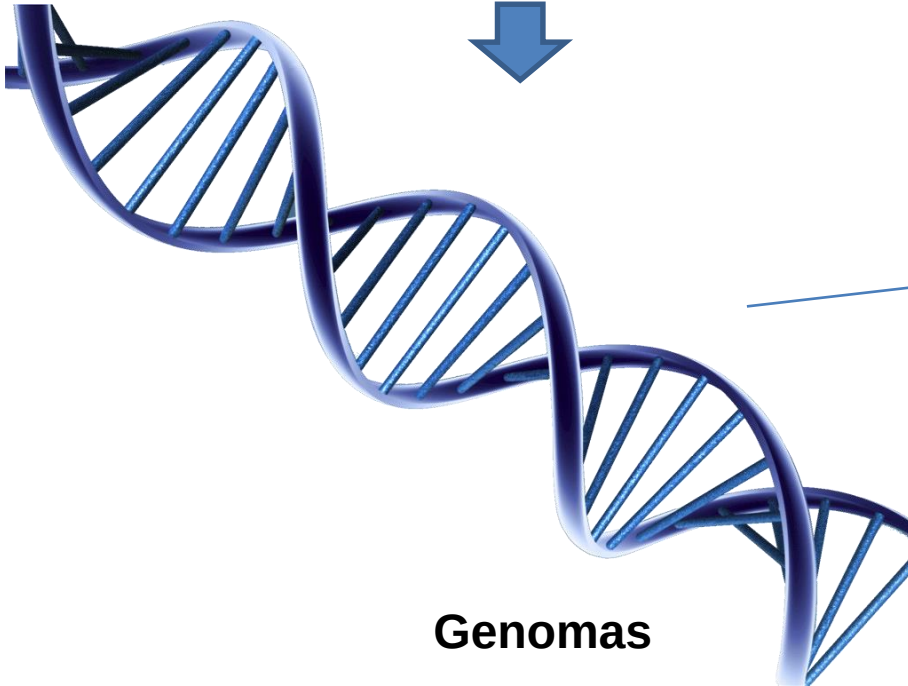
*Información genómica presente en el **conjunto** de los **individuos** del **conjunto** de **especies** que se encuentran presentes en un **mismo espacio físico**.*

EPIGENOMA

Información genómica disponible producto del grado de compactación de la cromatina en los eucariotas.

La información genómica

Materia viva



Genomas



Información

La información genómica



**Información
genómica**

- **Secuencias Andamio**

(Orígenes de replicación, secuencias de segregación, secuencias de protección de extremos)

- **Secuencias que expresan las funciones celulares**

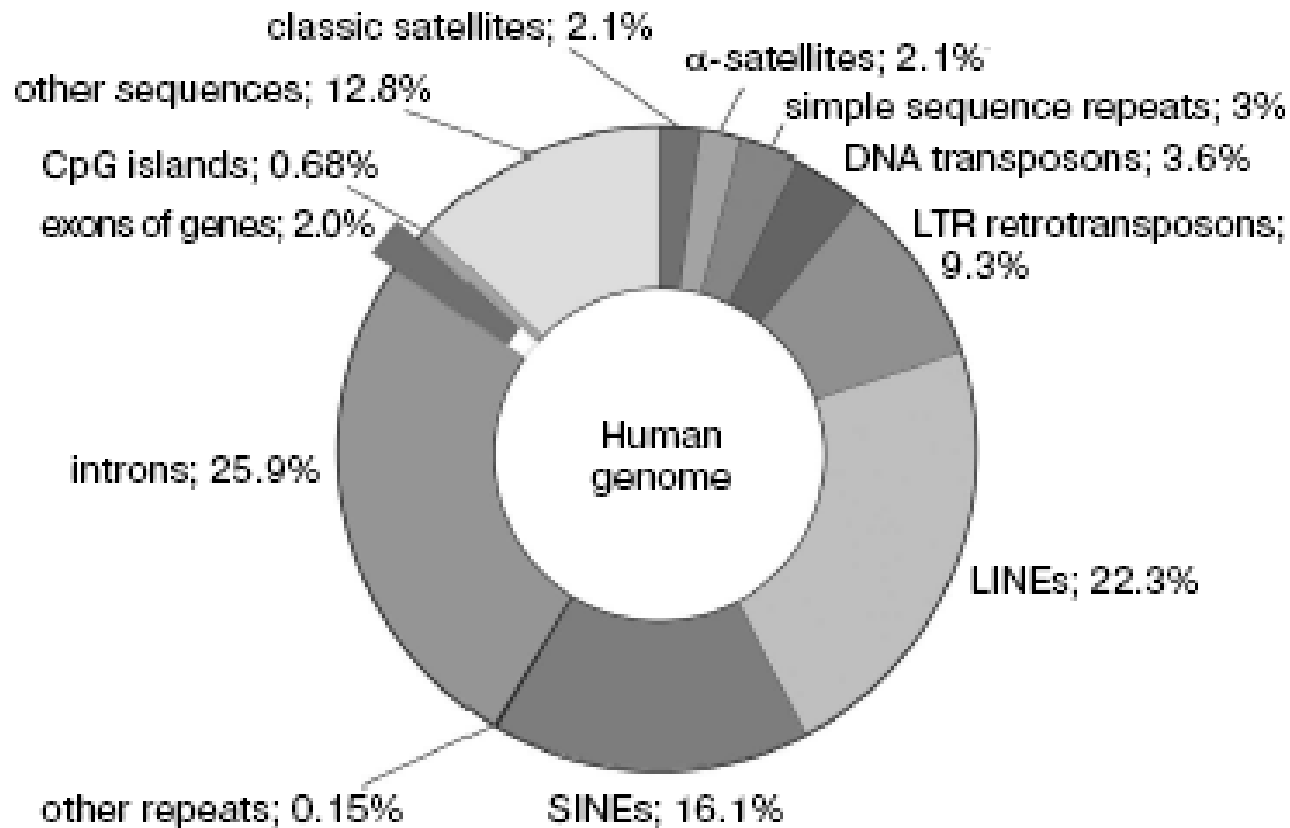
(Genes codificantes de proteínas, Genes productores de RNAs)

- **Relictos secuenciales no-funcionales y Moviloma**

(Pseudogenes, EVEs –*Endogenous Viral Elements*-, Transposones, Plásmidos integrados)

La información genómica

Contenido del genoma humano



La información genómica

Contenido de diversos genomas

Transposon and gene content depending on genome size (after [24, 369])

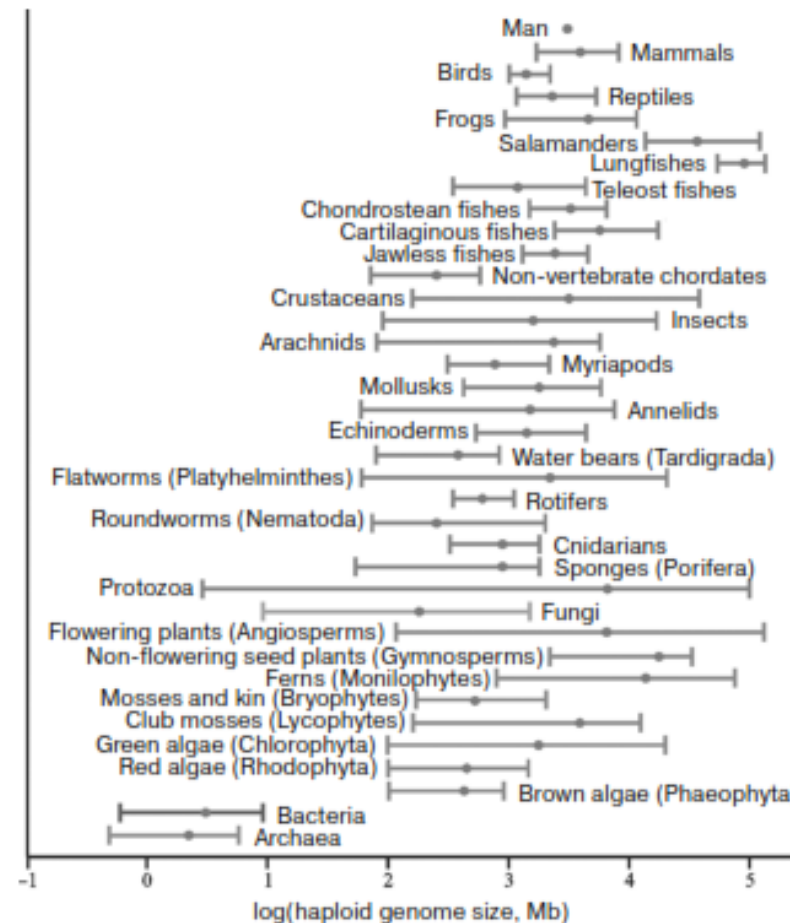
Species	Common name	Genome size, pg	% TEs	Gene number
<i>Fritillaria assyriaca</i>	lily	127.4	95-99	
<i>Rana esculenta</i>	frog	5.6-8.0	77	
<i>Homo sapiens</i>	human	3.5	45	23,000
<i>Xenopus laevis</i>	frog	3.5	37	
<i>Mus musculus</i>	mouse	3.4	40	35,000
<i>Zea mays</i>	maize	2.5	60	
<i>Gallus domesticus</i>	hen	1.25	27	20,000
<i>Tetraodon nigroviridis</i>	fish	0.51	0.14	22,000
<i>Takifugu rubripes</i>	fish	0.4	2	31,000
<i>Anopheles gambiae</i>	malaria mosquito	0.28	16	14,000
<i>Drosophila melanogaster</i>	fruit fly	0.18	15-22	14,039
<i>Ciona intestinalis</i>	ascidian	0.16	10	15,500
<i>Arabidopsis thaliana</i>	arabidopsis	0.16	14	26,000
<i>Caenorhabditis elegans</i>	worm	0.1	12	20,060
<i>Saccharomyces cerevisiae</i>	yeasts	0.012	3-5	6,680
<i>Escherichia coli</i>	bacterium	0.0046	0.3	4,500

1 pg = 978 Mpb

TE = Transposable Elements

La información genómica

Dimensión comparativa de los genomas



La información genómica

Genoma codificante de proteínas

Fracción de gDNA codificante para polipéptidos.

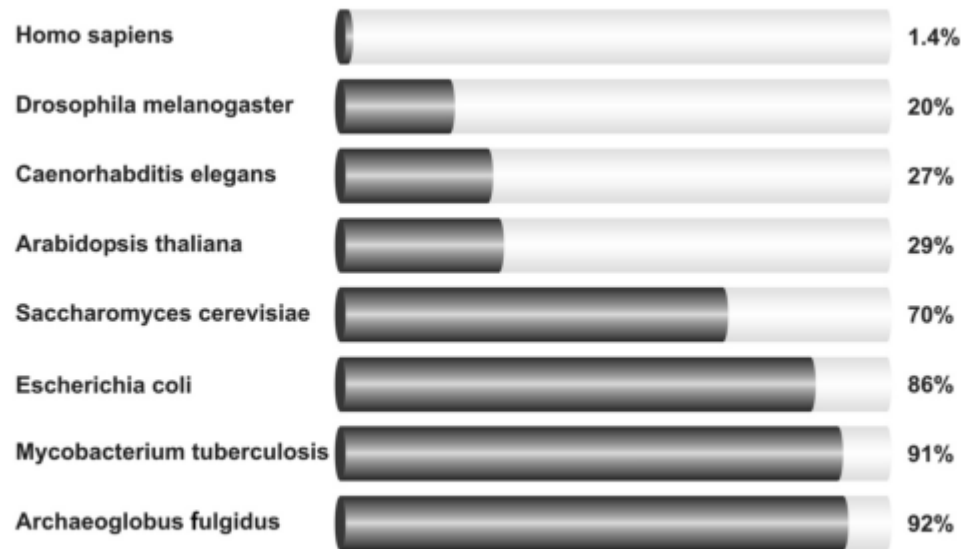


Figure 1 The percentage of protein-coding genes sequences in several eukaryotic and bacterial genomes.

A pesar que menos del **2%** del **gDNA humano** es **codificante para proteínas**, se ha visto que alrededor del **90%** es **activamente transcrito**, lo cual muestra la importancia del transcriptoma no codificante.

La información: deconstrucción



TEXTO



PÁRRAFO



ORACION



PALABRA



LETRA



GENOMA



**CIRCUITO
GÉNICO**



GEN



**BIOBRICK
GÉNICO**



NUCLEÓTIDO



**SISTEMA
OPERATIVO**



RUTINA



SUBROUTINA



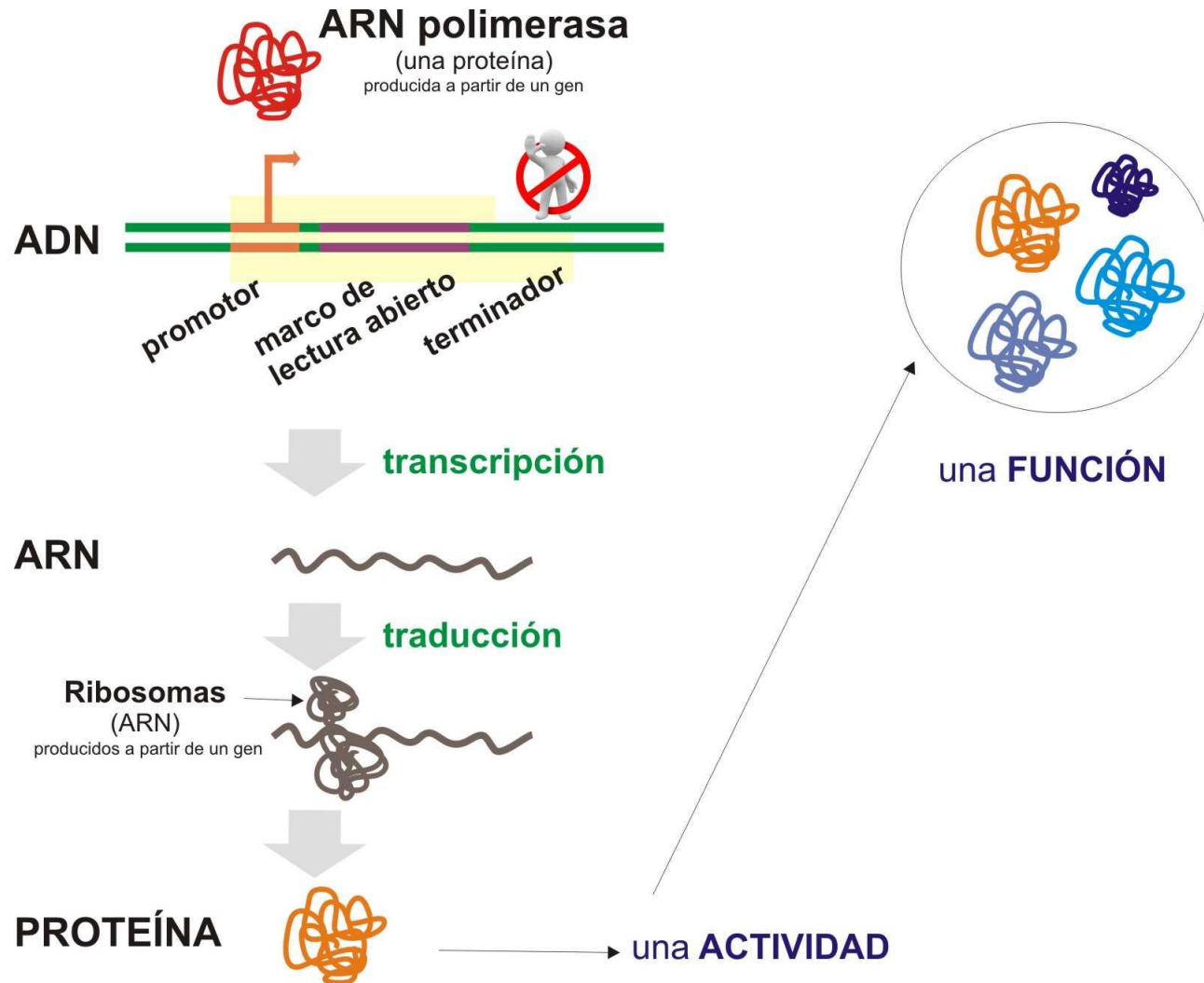
COMANDO



LETRA

La unidad de la información

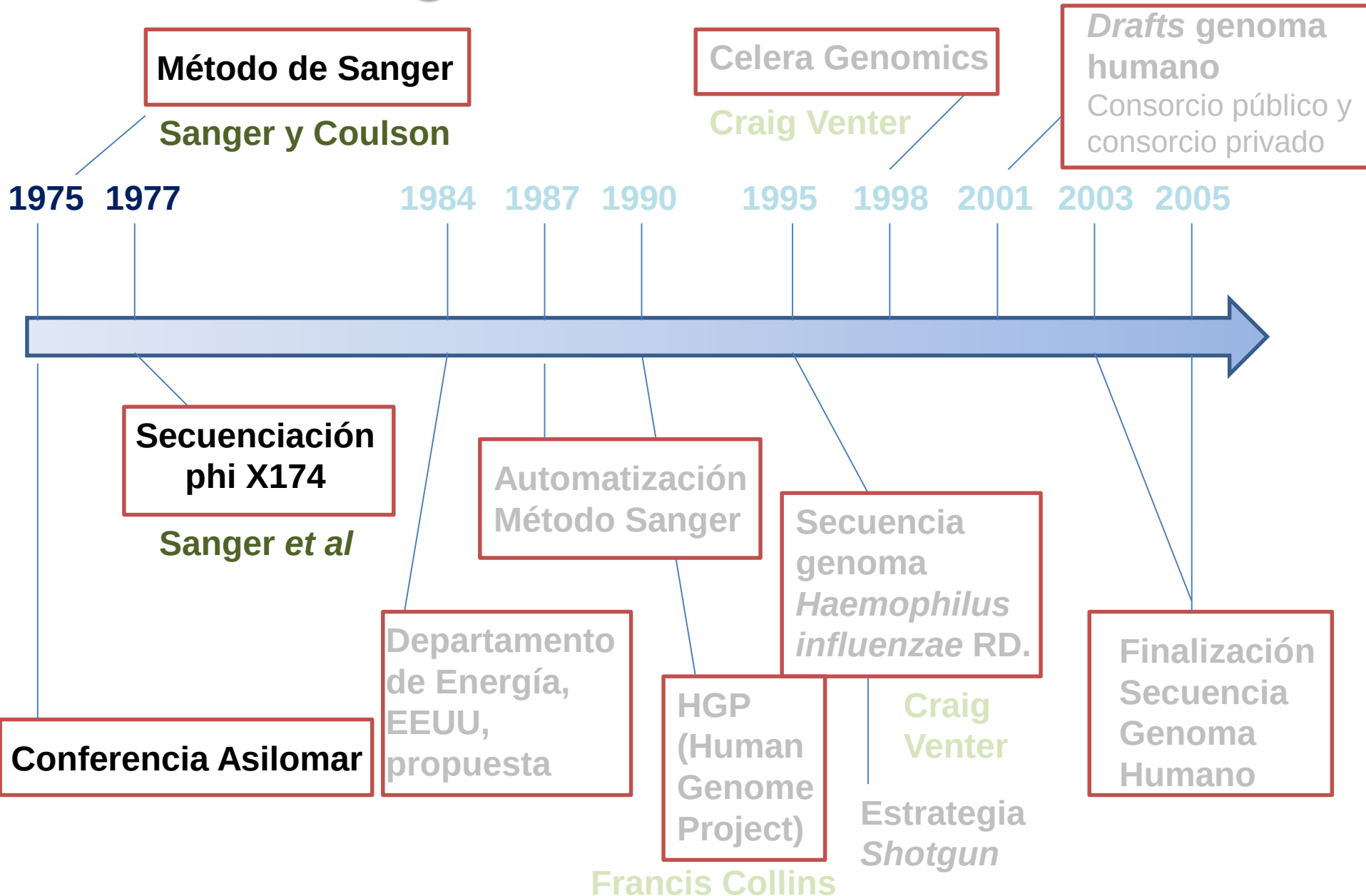
El gen



Tecnologías de secuenciación



Tecnologías de secuenciación: 1G



Tecnologías de secuenciación: 1G

J. Mol. Biol. (1975) **94**, 441–448

A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase

F. SANGER AND A. R. COULSON

*Medical Research Council
Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England*

(Received 20 December 1974)

A simple and rapid method for determining nucleotide sequences in single-stranded DNA by primed synthesis with DNA polymerase is described. It depends on the use of *Escherichia coli* DNA polymerase I and DNA polymerase from bacteriophage T4 under conditions of different limiting nucleoside triphosphates and concurrent fractionation of the products according to size by ionophoresis on acrylamide gels. The method was used to determine two sequences in bacteriophage ϕ X174 DNA using the synthetic decanucleotide A-G-A-A-A-T-A-A-A-A and a restriction enzyme digestion product as primers.

Tecnologías de secuenciación: 1G

articles

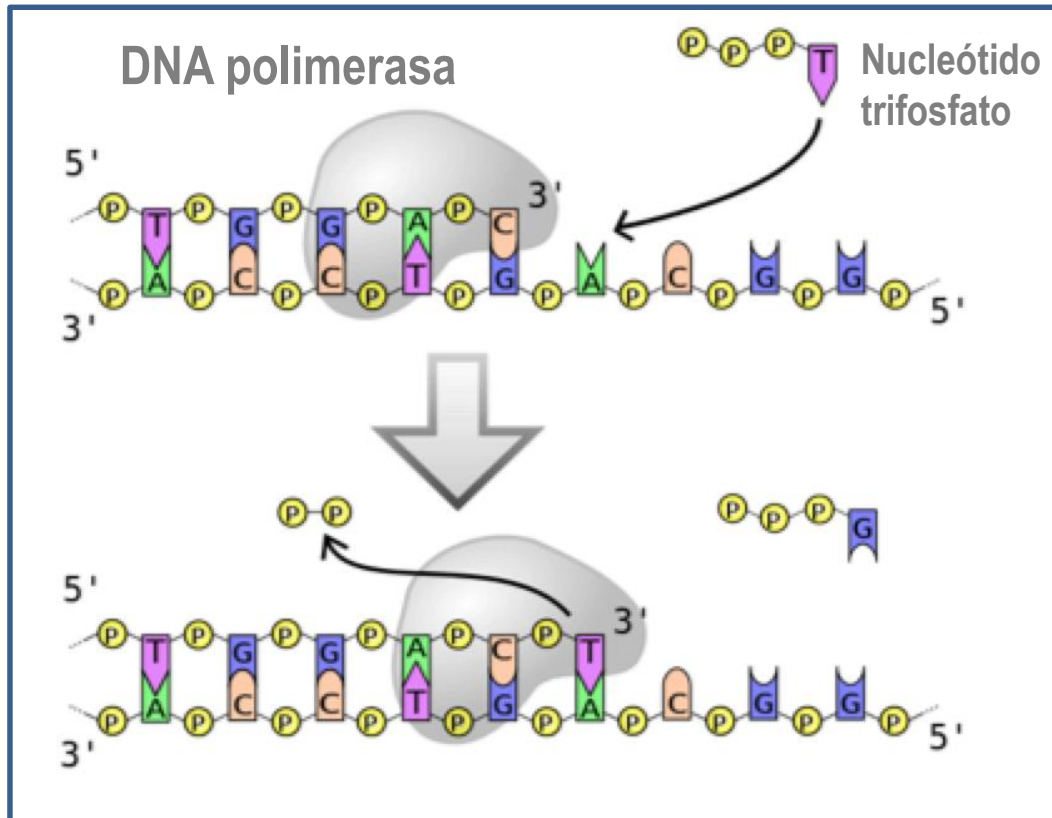
Nucleotide sequence of bacteriophage Φ X174 DNA

F. Sanger, G. M. Air*, B. G. Barrell, N. L. Brown†, A. R. Coulson, J. C. Fiddes, C. A. Hutchison III‡, P. M. Slocombe§ & M. Smith*

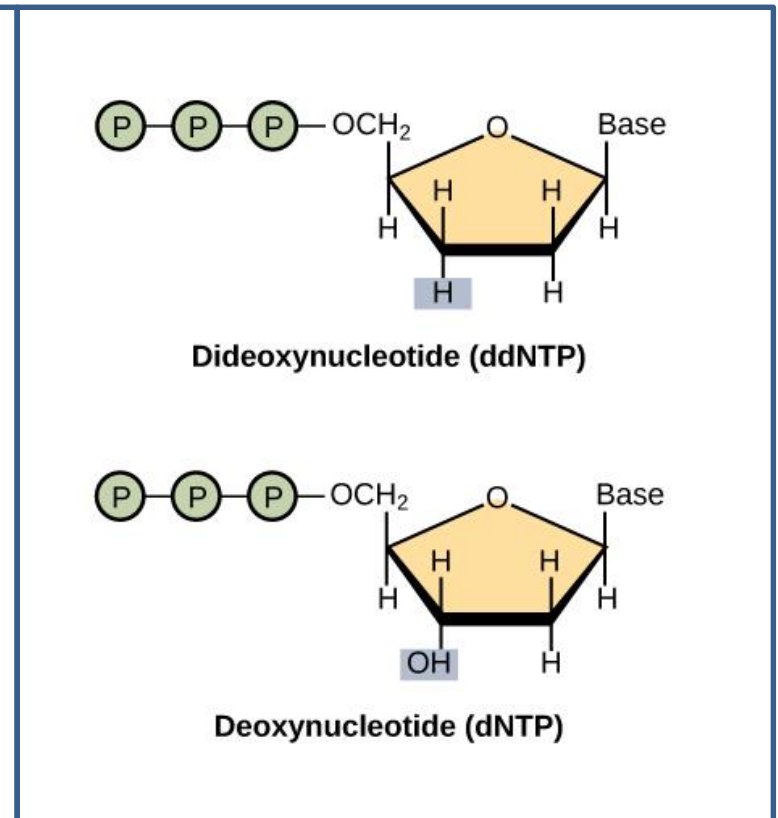
MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

A DNA sequence for the genome of bacteriophage Φ X174 of approximately 5,375 nucleotides has been determined using the rapid and simple 'plus and minus' method. The sequence identifies many of the features responsible for the production of the proteins of the nine known genes of the organism, including initiation and termination sites for the proteins and RNAs. Two pairs of genes are coded by the same region of DNA using different reading frames.

Tecnologías de secuenciación: 1G

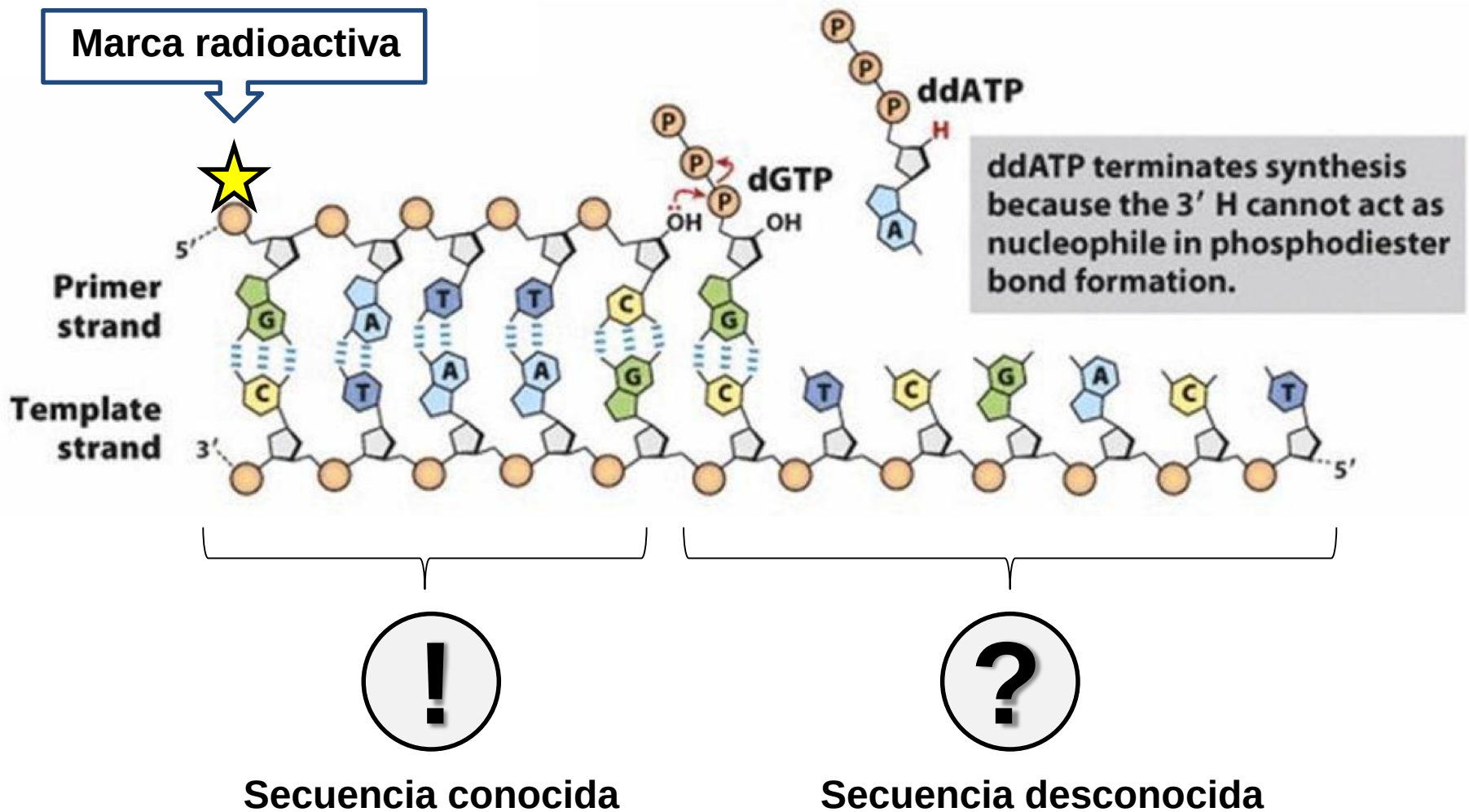


Actividad de la DNA polimerasa

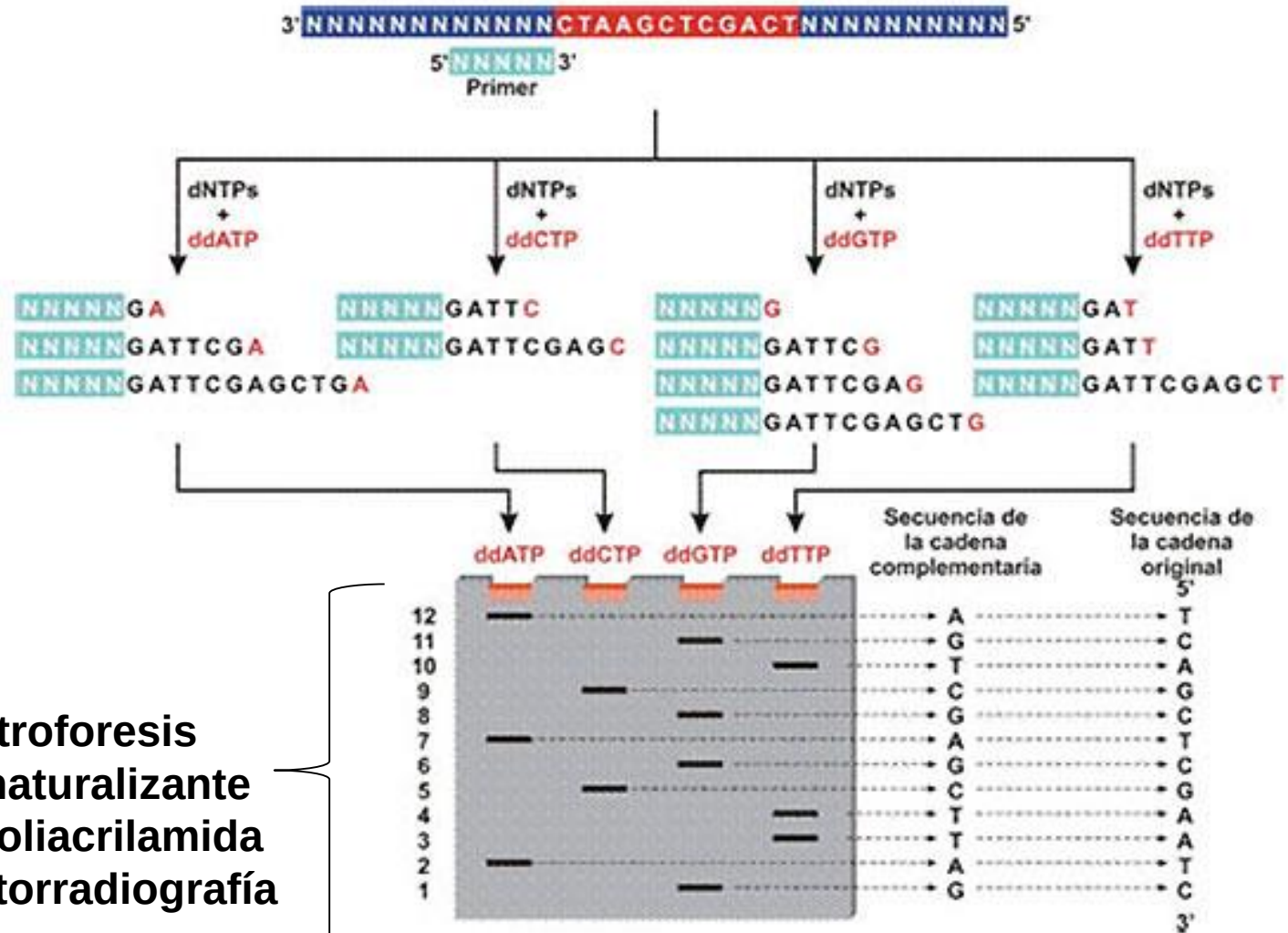


Terminador de cadena

Tecnologías de secuenciación: 1G

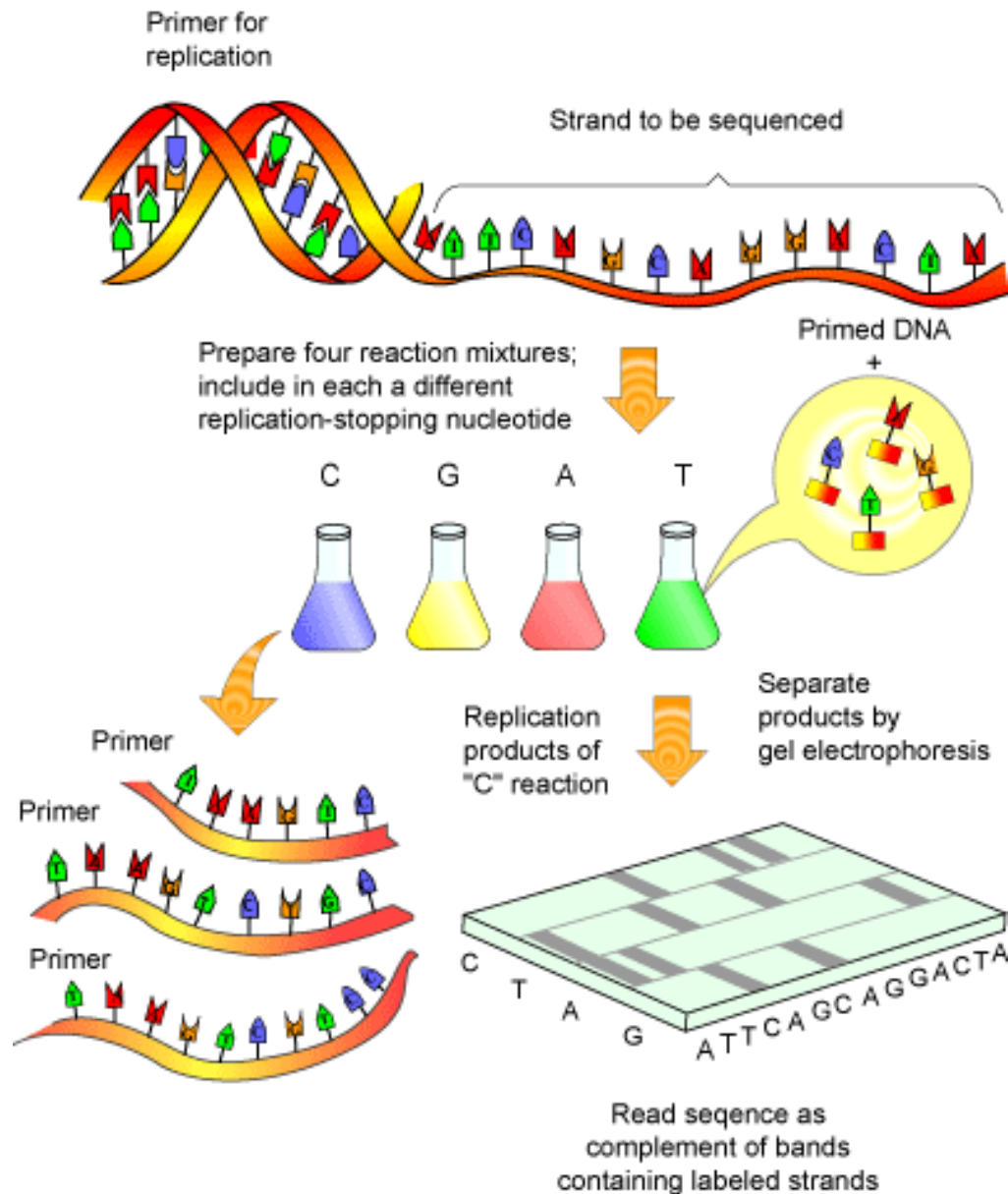


Tecnologías de secuenciación: 1G

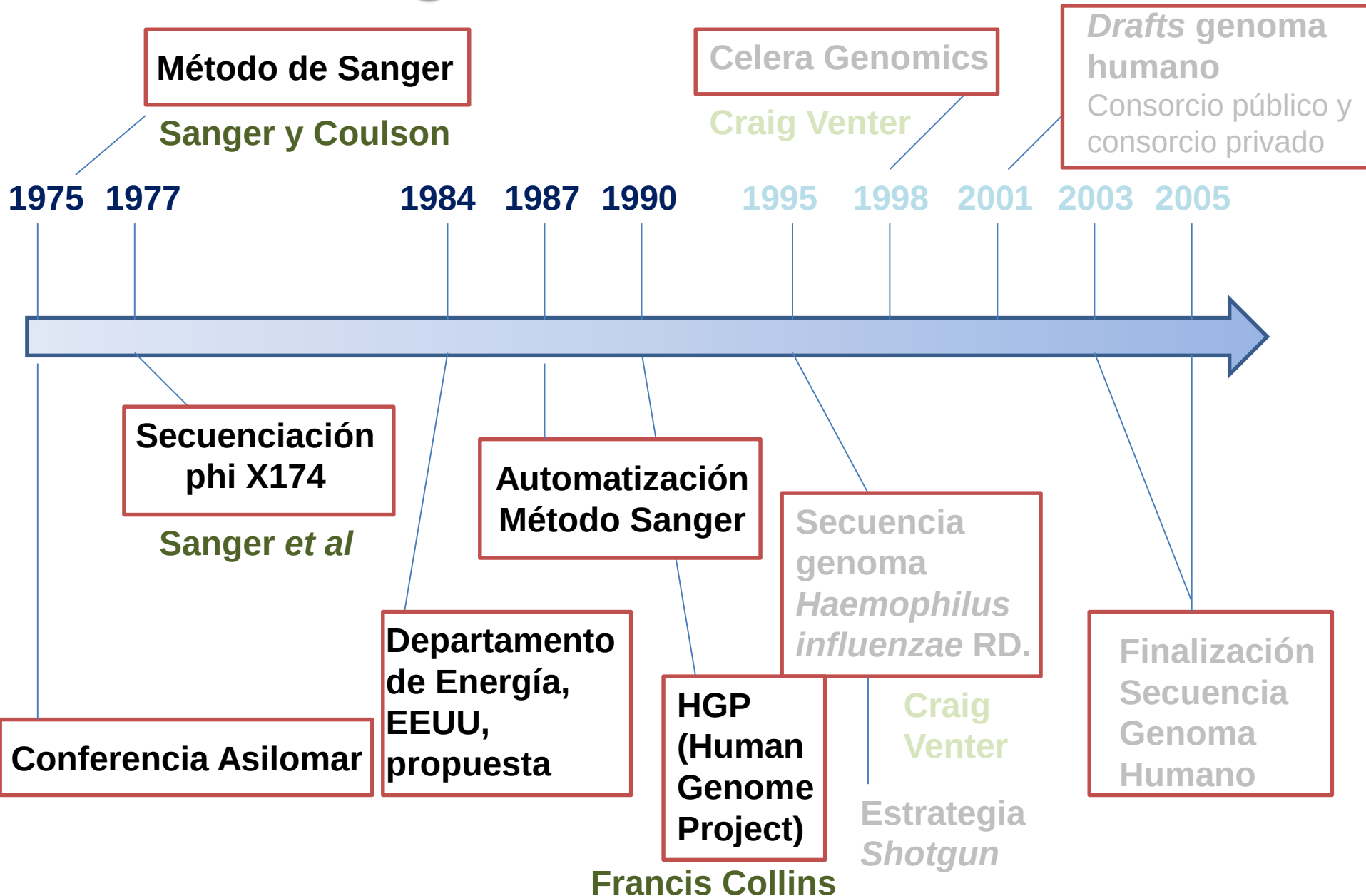


Electroforesis desnaturalizante en poliacrilamida y autorradiografía

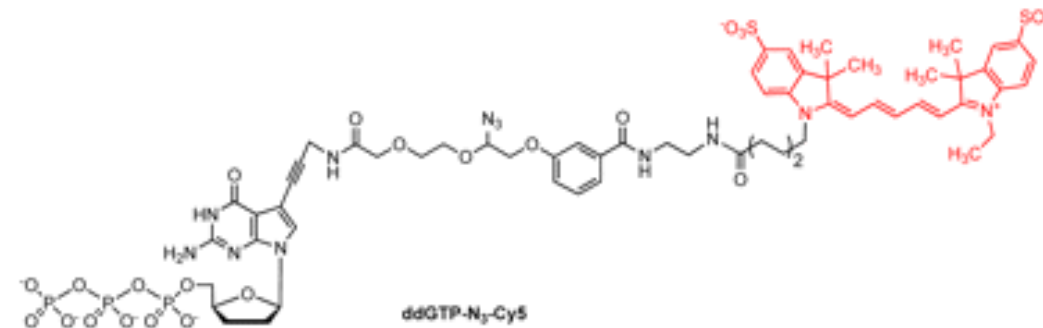
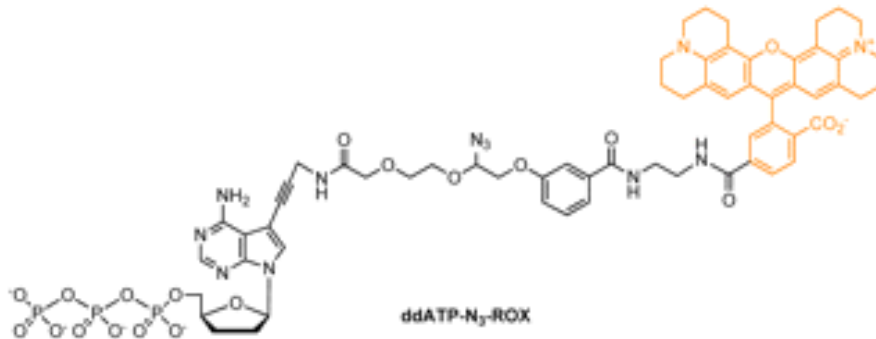
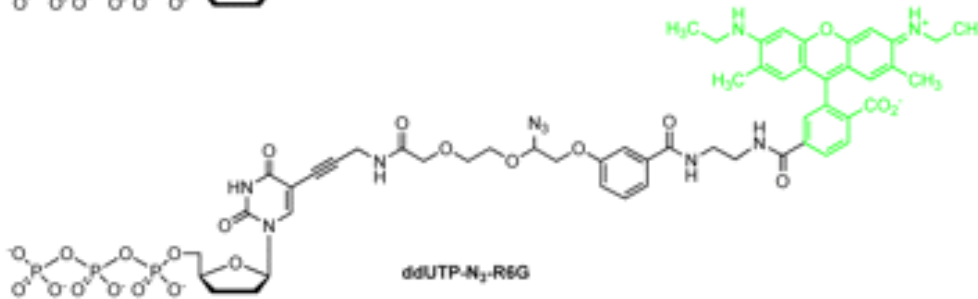
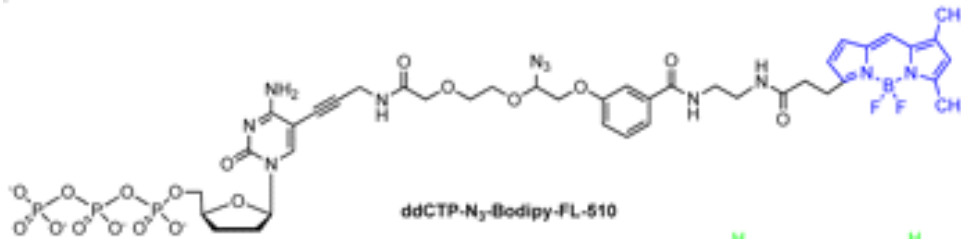
Tecnologías de secuenciación: 1G



Tecnologías de secuenciación: 1G

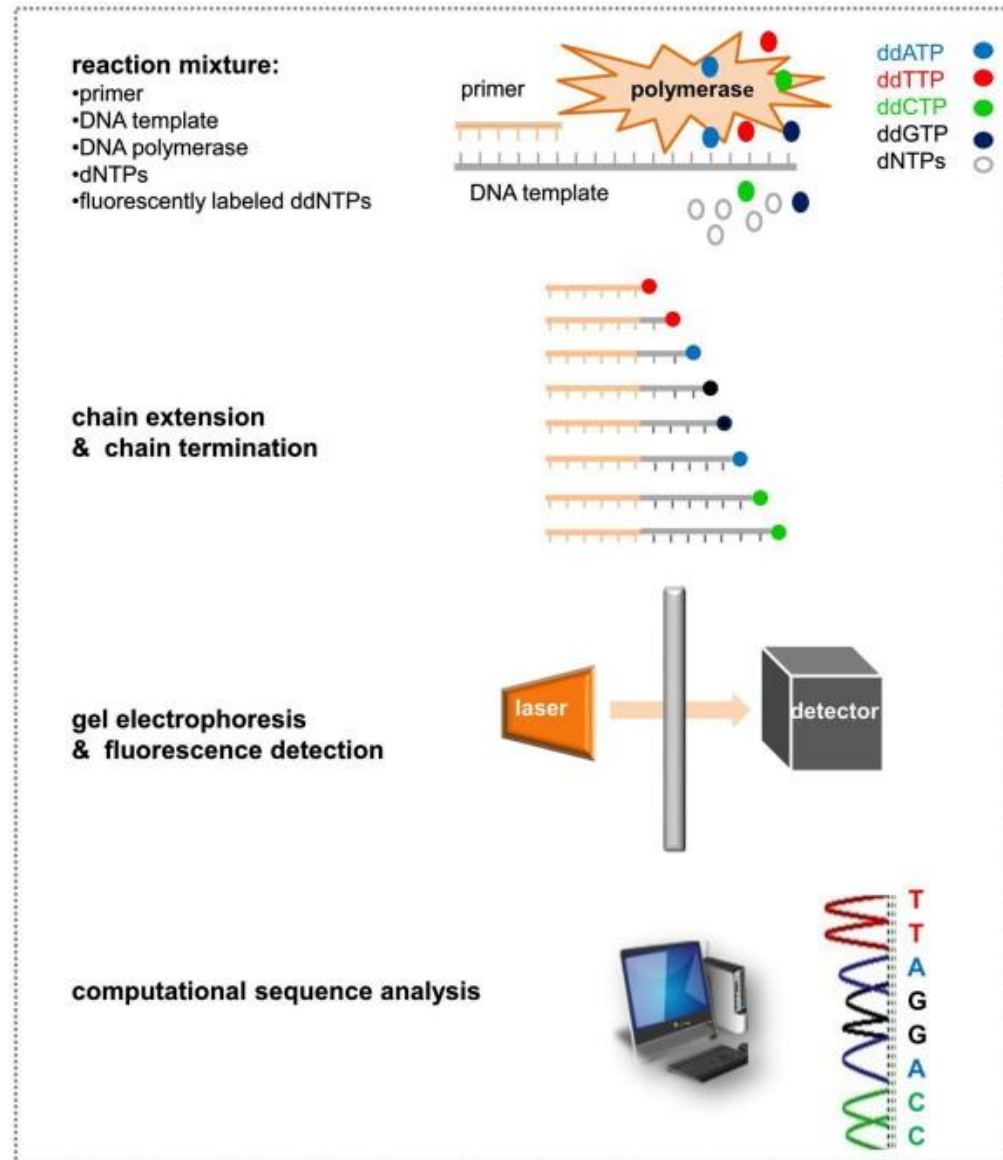


Tecnologías de secuenciación: 1G



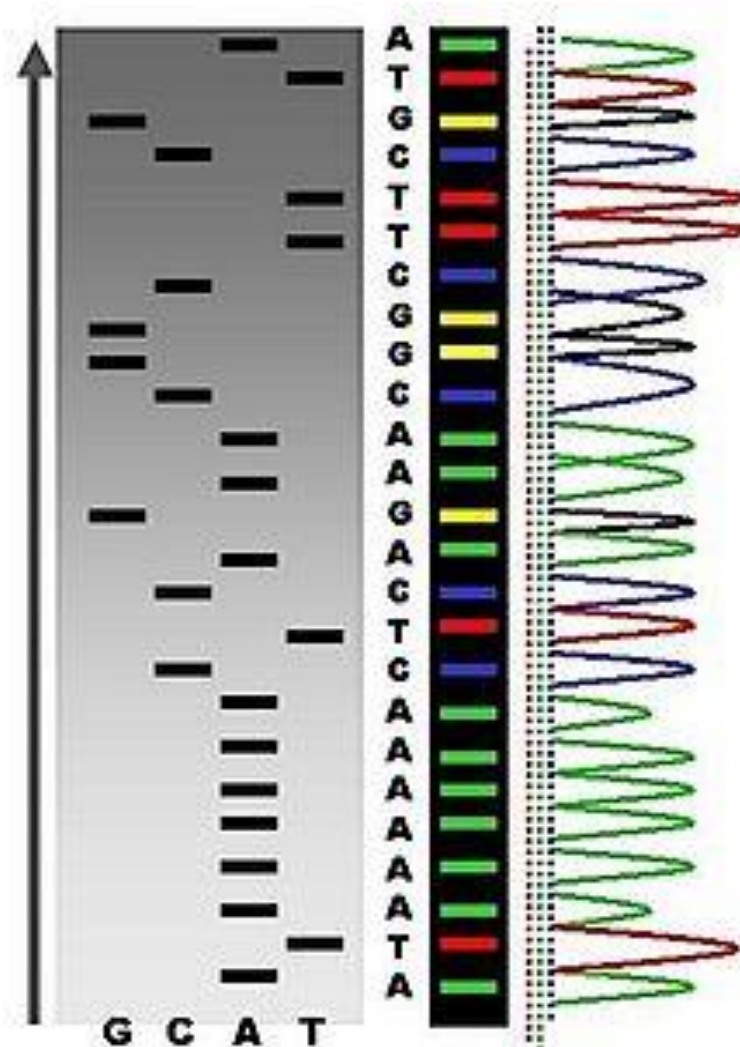
**ddNTPs
con
fluorocromos**

Tecnologías de secuenciación: 1G



Tecnologías de secuenciación: 1G

PAGE
Autorradiografía



PAGE
Fluorimetría

Tecnologías de secuenciación: 1G

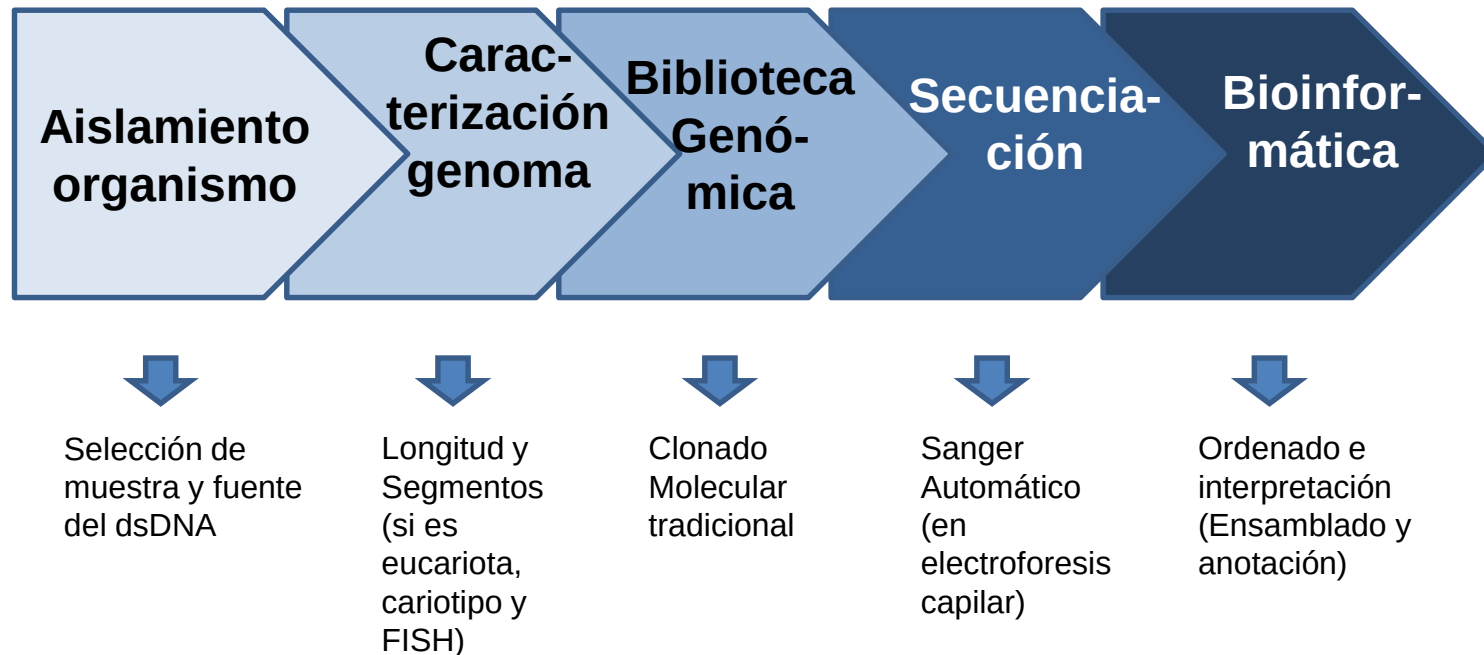
1987 - Primera Secuenciadora

- ABI 370



Tecnologías de secuenciación: 1G

La adquisición de secuencias genómicas suele seguir el siguiente flujo de trabajo...



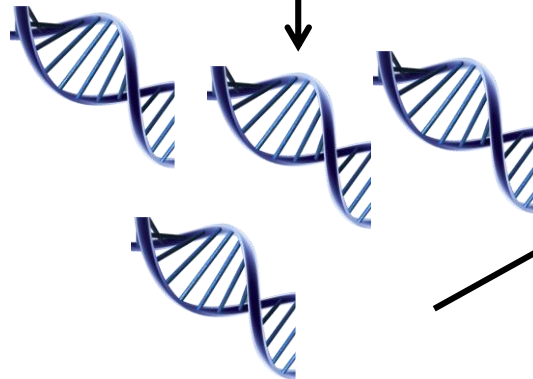
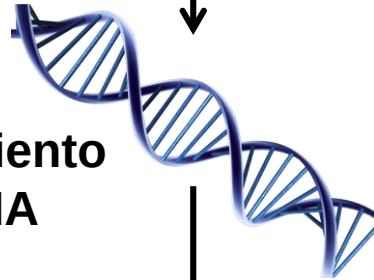
Tecnologías de secuenciación: 1G

Materia viva

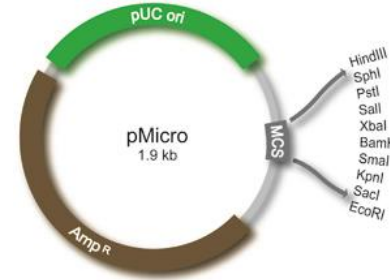


Biblioteca Genómica

Aislamiento
gDNA

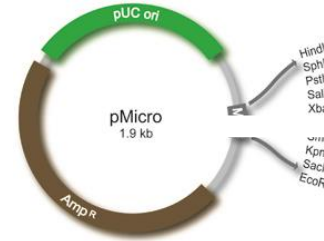


Fragmentación
del
gDNA

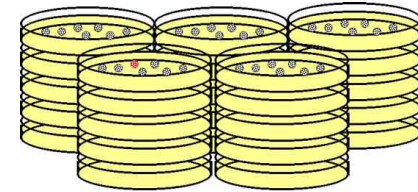


Vector de clonado
(plásmido)

Apertura (Enzimas de
Restricción)



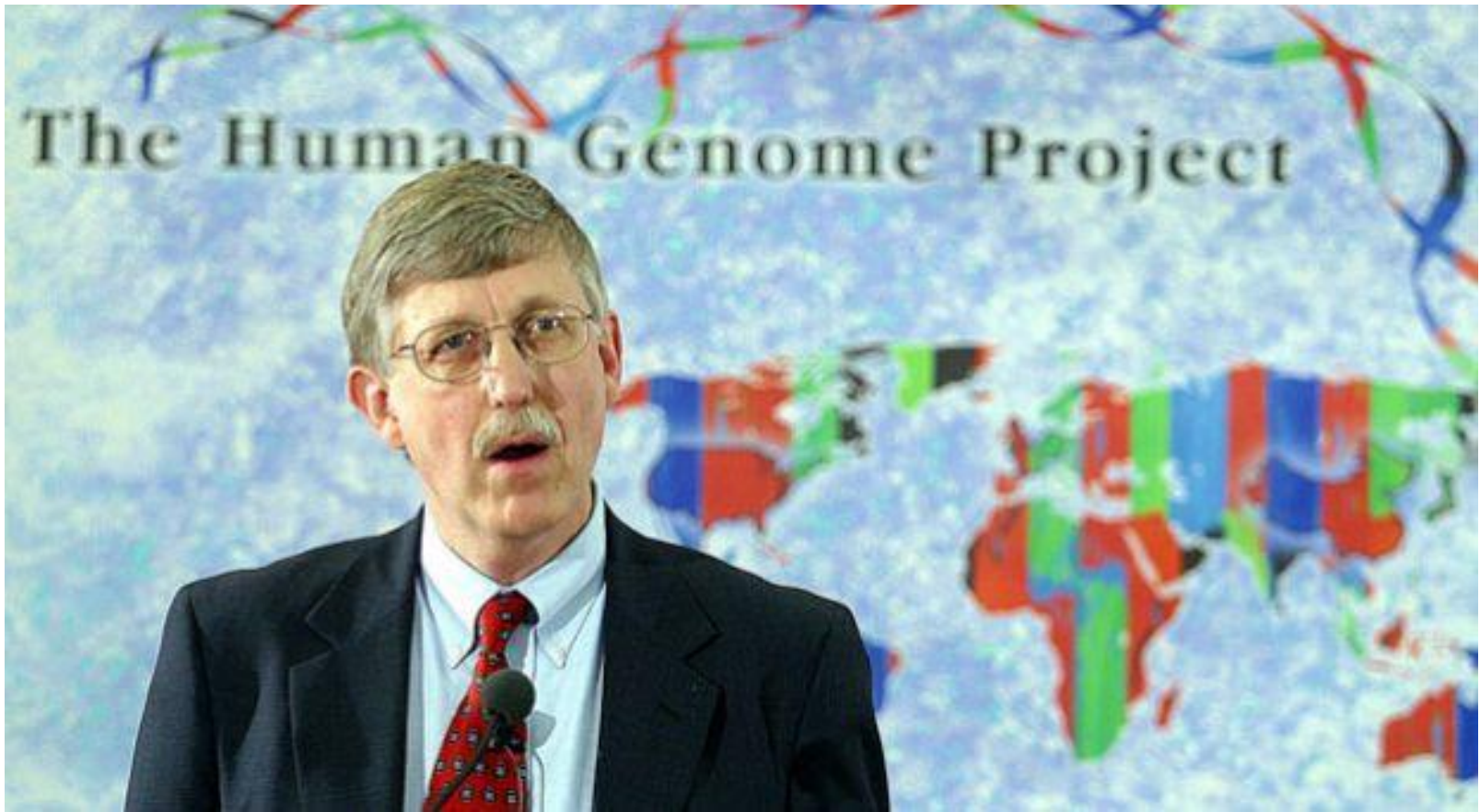
Ligación
(DNA ligasa)



Transformación
Escherichia coli



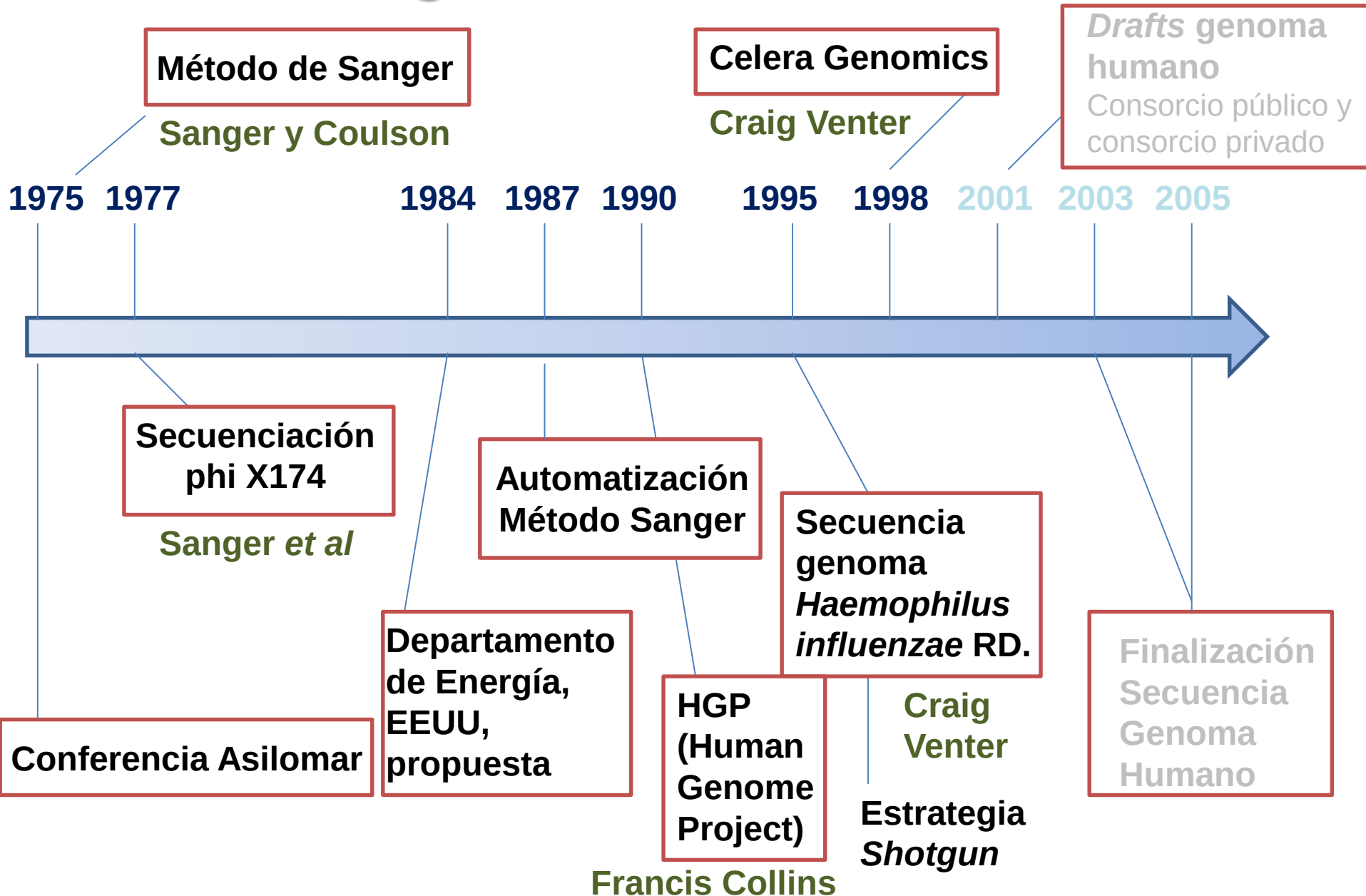
Tecnologías de secuenciación: 1G



Francis Collins

Se origina en 1990 y planificó finalizarse en 2005

Tecnologías de secuenciación: 1G



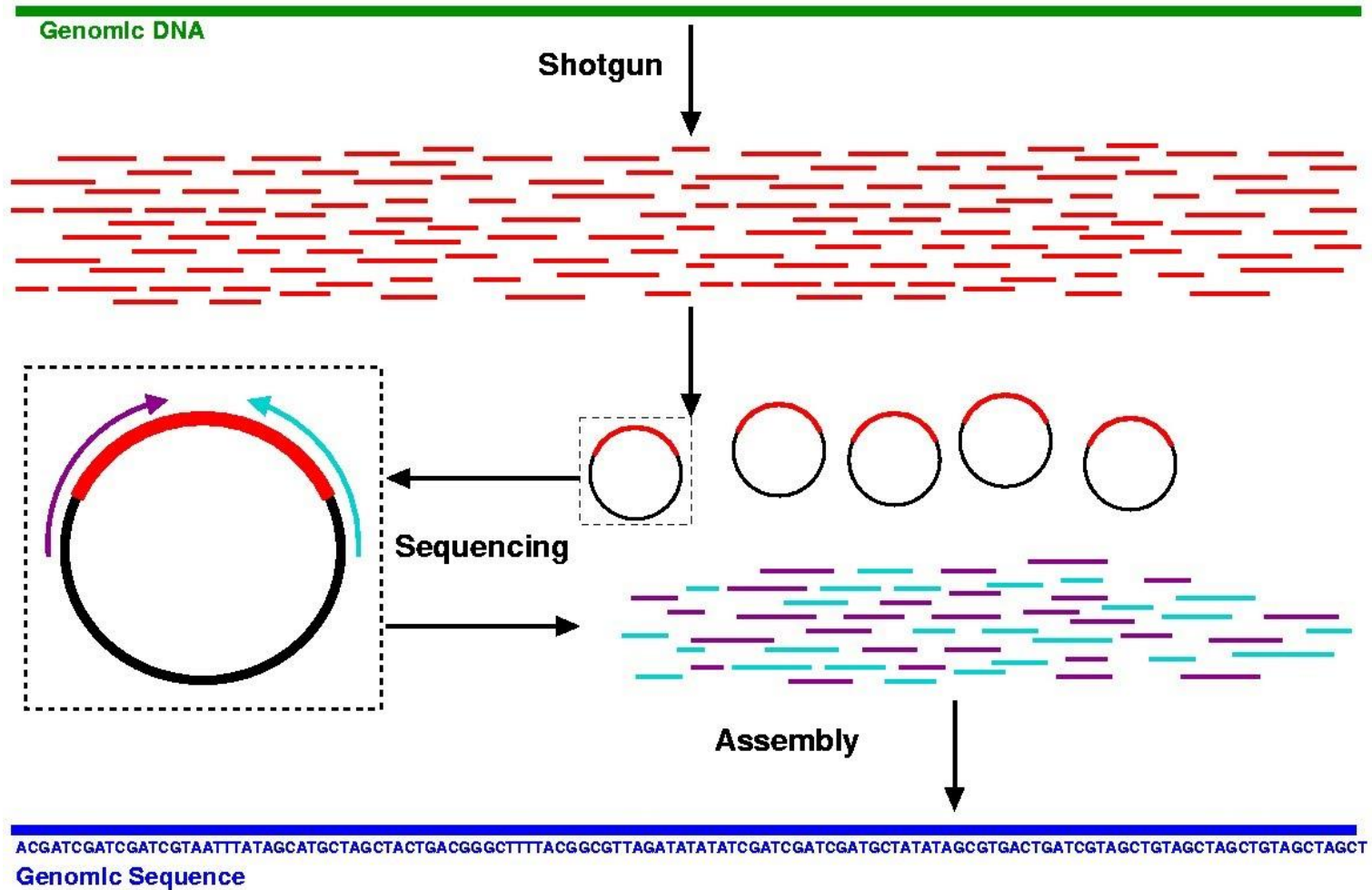
Tecnologías de secuenciación: 1G

Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Saudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghagen, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Venter†

An approach for genome analysis based on sequencing and assembly of unselected pieces of DNA from the whole chromosome has been applied to obtain the complete nucleotide sequence (1,830,137 base pairs) of the genome from the bacterium *Haemophilus influenzae* Rd. This approach eliminates the need for initial mapping efforts and is therefore applicable to the vast array of microbial species for which genome maps are unavailable. The *H. influenzae* Rd genome sequence (Genome Sequence DataBase accession number L42023) represents the only complete genome sequence from a free-living organism.

Tecnologías de secuenciación: 1G



Esto fue posible gracias al desarrollo de algoritmos bioinformáticos.

Tecnologías de secuenciación: 1G



Craig Venter

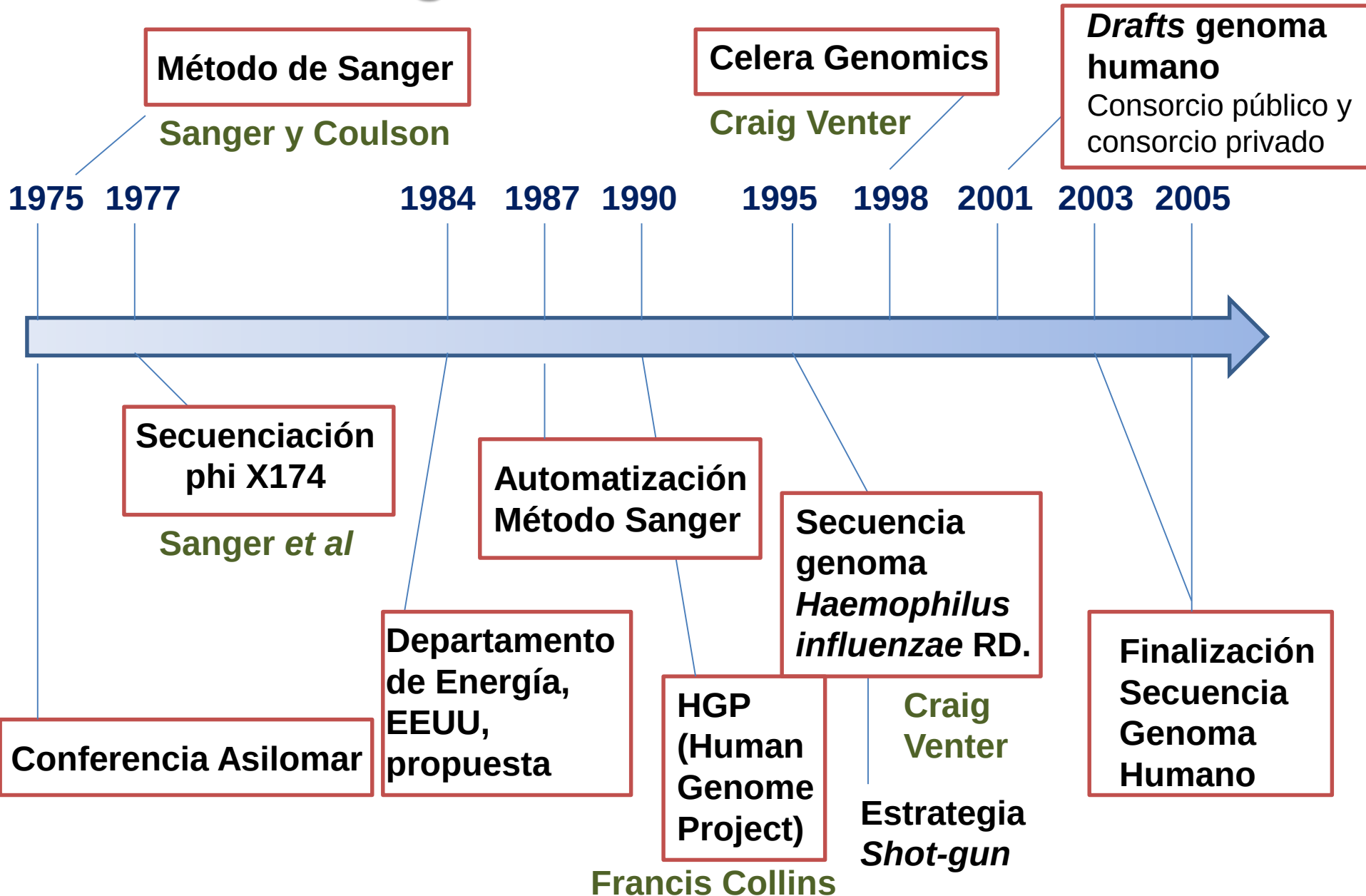
Se involucra en secuenciar el genoma humano desde 1998

Tecnologías de secuenciación: 1G



- **ABI 310**
- Equipo introducido en **1996**
- Primer secuenciador de capilares
- Se incrementa la capacidad de secuencias por corrida

Tecnologías de secuenciación: 1G



Tecnologías de secuenciación: 1G

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

** A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.*

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

Tecnologías de secuenciación: 1G

THE HUMAN GENOME

The Sequence of the Human Genome

J. Craig Venter,^{1*} Mark D. Adams,¹ Eugene W. Myers,¹ Peter W. Li,¹ Richard J. Mural,¹ Granger G. Sutton,¹ Hamilton O. Smith,¹ Mark Yandell,¹ Cheryl A. Evans,¹ Robert A. Holt,¹ Jeannine D. Gocayne,¹ Peter Amanatides,¹ Richard M. Ballieu,¹ Daniel H. Huson,¹ Jennifer Russo Wortman,¹ Qing Zhang,¹ Chinappa D. Kodra,¹ Xiangqun H. Zheng,¹ Lin Chen,¹ Marlan Skupnik,¹ Gangadharan Subramanian,¹ Paul D. Thomas,¹ Jinghui Zhang,¹ George L. Gabor Miklos,² Catherine Nelson,³ Samuel Broder,¹ Andrew G. Clark,⁴ Joe Nadeau,⁵ Victor A. McKusick,⁶ Norton Zinder,⁷ Arnold J. Levine,⁸ Richard J. Roberts,⁹ Mel Simon,⁹ Carolyn Slayman,¹⁰ Michael Hunkapiller,¹¹ Randall Bolanos,¹² Arthur Delcher,¹ Ian Dew,¹³ Daniel Fasulo,¹⁴ Michael Flanigan,¹ Lilliana Flores,¹ Aaron Halpern,¹ Sridhar Hannenhalli,¹ Saul Kravitz,¹ Samuel Levy,¹ Clark Mobarry,¹ Knut Reinert,¹ Karin Remington,¹ Jane Abu-Threideh,¹ Ellen Beasley,¹ Kendra Biddick,¹ Vivien Bonazzi,¹ Rhonda Brandon,¹ Michele Cargill,¹ Ishwar Chandramoulewaran,¹ Rosane Charlab,¹ Kabir Chaturvedi,¹ Zuoming Deng,¹ Valentina Di Francesco,¹ Patrick Dunn,¹ Karen Eilbeck,¹ Carlos Evangelista,¹ Andrei E. Gabrielian,¹ Weiniu Gan,¹ Wangmao Ge,¹ Fangcheng Gong,¹ Zhiping Gu,¹ Ping Guan,¹ Thomas J. Helman,¹ Maureen E. Higgins,¹ Rui-Ru Ji,¹ Zhaoxi Ke,¹ Karen A. Ketchum,¹ Zhongwu Lai,¹ Yiding Lei,¹ Zhenya Li,¹ Jayin Li,¹ Yong Liang,¹ Xiaoying Lin,¹ Fu Lu,¹ Gennady V. Merkulov,¹ Natalia Milshina,¹ Helen M. Moore,¹ Ashwinikumar K. Naik,¹ Valthav A. Narayan,¹ Beena Neelam,¹ Deborah Nusskern,¹ Douglas B. Rusch,¹ Steven Salzberg,¹² Wei Shao,¹ Bixiong Shue,¹ Jingtao Sun,¹ Zhen Yuan Wang,¹ Aihui Wang,¹ Xin Wang,¹ Jian Wang,¹ Ming-Hui Wei,¹ Ron Wildes,¹⁵ Chundin Xiao,¹ Chunhua Yan,¹ Allison Yao,¹ Jane Ye,¹ Ming Zhan,¹ Weiqing Zhang,¹ Hongyu Zhang,¹ Qi Zhao,¹ Liansheng Zheng,¹ Fei Zhong,¹ Wenyang Zhong,¹ Shiaoqing C. Zhu,¹ Shaying Zhao,¹² Dennis Gilbert,¹ Suzanna Baumhueter,¹ Gene Spler,¹ Christine Carter,¹ Anibal Cravchik,¹ Trevor Woodage,¹ Feroze Ali,¹ Huijin An,¹ Aderonke Awe,¹ Danita Baldwin,¹ Holly Baden,¹ Mary Barnstead,¹ Ian Barrow,¹ Karen Beeson,¹ Dana Buzam,¹ Amy Carver,¹ Angela Center,¹ Ming Lai Cheng,¹ Liz Curry,¹ Steve Danaher,¹ Lionel Davenport,¹ Raymond Dealata,¹ Susanne Dietz,¹ Kristina Dodson,¹ Lisa Doup,¹ Steven Fiertera,¹ Neha Garg,¹ Andreas Gluecksmann,¹ Brit Hart,¹ Jason Haynes,¹ Charles Haynes,¹ Cheryl Helner,¹ Suzanne Hladun,¹ Damon Hostin,¹ Jarratt Houck,¹ Timothy Howland,¹ Chinyere Ibegwam,¹ Jeffery Johnson,¹ Francis Kalush,¹ Lesley Kline,¹ Shashi Koduru,¹ Amy Love,¹ Felecia Mann,¹ David May,¹ Steven McCawley,¹ Tina McIntosh,¹ Ivy McMullan,¹ Hee Moy,¹ Linda Moy,¹ Brian Murphy,¹ Keith Nelson,¹ Cynthia Pfannkuch,¹ Eric Pratt,¹ Vinita Puri,¹ Hina Qureshi,¹ Matthew Reardon,¹ Robert Rodriguez,¹ Yu-Hui Rogers,¹ Deanna Romblad,¹ Bob Ruhfel,¹ Richard Scott,¹ Cynthia Sitter,¹ Michelle Smallwood,¹ Erin Stewart,¹ Renee Strong,¹ Ellen Suh,¹ Reginald Thomas,¹ Ni Ni Tint,¹ Sukyee Tse,¹ Claire Vech,¹ Gary Wang,¹ Jeremy Wetter,¹ Sherita Williams,¹ Monica Williams,¹ Sandra Windoor,¹ Emily Winn-Deen,¹ Kerliellen Wolfe,¹ Jayshree Zaveri,¹ Karana Zaveri,¹ Josep F. Abril,¹⁴ Roderic Guigó,¹⁴ Michael J. Campbell,¹ Kimmen V. Sjolander,¹ Brian Karlak,¹ Anish Kejarwal,¹ Hualyu Mi,¹ Betty Lazareva,¹ Thomas Hutton,¹ Apurva Narechania,¹ Karen Diemer,¹ Anushya Muruganujam,¹ Nan Guo,¹ Shinji Sato,¹ Vineet Bafna,¹ Sorin Istrail,¹ Ross Lippert,¹ Russell Schwartz,¹ Brian Walenz,¹ Shibu Yooseph,¹ David Allen,¹ Anand Basse,¹ James Saxendale,¹ Louis Blick,¹ Marcelo Caminha,¹ John Carnes-Stine,¹ Parris Casik,¹ Yen-Hui Chiang,¹ My Coyne,¹ Carl Dahlke,¹ Anne Deslattes Mays,¹ Maria Dombroski,¹ Michael Donnelly,¹ Dale Ely,¹ Shiva Eparham,¹ Carl Foster,¹ Harold Gire,¹ Stephen Glanowski,¹ Kenneth Glasser,¹ Anna Glodek,¹ Mark Gorokhov,¹ Ken Graham,¹ Barry Gropman,¹ Michael Harris,¹ Jeremy Hell,¹ Scott Henderson,¹ Jeffrey Hoover,¹ Donald Jennings,¹ Catherine Jordan,¹ James Jordan,¹ John Kasha,¹ Leonid Kagan,¹ Cheryl Kraft,¹ Alexander Levitsky,¹ Mark Lewis,¹ Xiangjun Liu,¹ John Lopez,¹ Daniel Ma,¹ William Majoros,¹ Joe McDaniel,¹ Sean Murphy,¹ Matthew Newman,¹ Trung Nguyen,¹ Ngoc Nguyen,¹ Marc Nodell,¹ Sue Pan,¹ Jim Peck,¹ Marshall Peterson,¹ William Rowe,¹ Robert Sanders,¹ John Scott,¹ Michael Simpson,¹ Thomas Smith,¹ Arlan Sprague,¹ Timothy Stockwell,¹ Russell Turner,¹ Eli Venter,¹ Mei Wang,¹ Melyuan Wen,¹ David Wu,¹ Mitchell Wu,¹ Ashley Xia,¹ Ali Zandieh,¹ Xiaohong Zhu¹

THE HUMAN GENOME

A 2.91-billion base pair (bp) consensus sequence of the euchromatic portion of the human genome was generated by the whole-genome shotgun sequencing method. The 14.8-billion bp DNA sequence was generated over 9 months from 27,271,853 high-quality sequence reads (5.11-fold coverage of the genome) from both ends of plasmid clones made from the DNA of five individuals. Two assembly strategies—a whole-genome assembly and a regional chromosome assembly—were used, each combining sequence data from Celera and the publicly funded genome effort. The public data were shredded into 350-bp segments to create a 2.9-fold coverage of those genome regions that had been sequenced, without including biases inherent in the cloning and assembly procedure used by the publicly funded group. This brought the effective coverage in the assemblies to eightfold, reducing the number and size of gaps in the final assembly over what would be obtained with 5.11-fold coverage. The two assembly strategies yielded very similar results that largely agree with independent mapping data. The assemblies effectively cover the euchromatic regions of the human chromosomes. More than 90% of the genome is in scaffold assemblies of 100,000 bp or more, and 25% of the genome is in scaffolds of 10 million bp or larger. Analysis of the genome sequence revealed 26,388 protein-encoding transcripts for which there was strong corroborating evidence and an additional ~12,000 computationally derived genes with mouse matches or other weak supporting evidence. Although gene-dense clusters are obvious, almost half the genes are dispersed in low G+C sequence separated by large tracts of apparently noncoding sequence. Only 1.1% of the genome is spanned by exons, whereas 24% is in introns, with 75% of the genome being intergenic DNA. Duplications of segmental blocks, ranging in size up to chromosomal lengths, are abundant throughout the genome and reveal a complex evolutionary history. Comparative genomic analysis indicates vertebrate expansions of genes associated with neuronal function, with tissue-specific developmental regulation, and with the hemostasis and immune systems. DNA sequence comparisons between the consensus sequence and publicly funded genome data provided locations of 2.1 million single-nucleotide polymorphisms (SNPs). A random pair of human haploid genomes differed at a rate of 1 bp per 1250 on average, but there was marked heterogeneity in the level of polymorphism across the genome. Less than 1% of all SNPs resulted in variation in proteins, but the task of determining which SNPs have functional consequences remains an open challenge.

Tecnologías de secuenciación: 1G

- La **estrategia** del genoma haploide de la especie humana se denomina ***Clone by Clone*** e involucra un **alto consumo de tiempo, insumos y recursos humanos**. Está basada en el **método Sanger**, incapaz de leer más de 900-1500 nt por evento de secuenciación.
- Una **adaptación** fue la estrategia ***Shot gun*** (utilizada para el genoma de un individuo), similar a la anterior pero fragmentando mucho más los genomas o los clones de una biblioteca, sin tantos controles para revelar su calidad y orden.
- La **secuencia resultante** de las estrategias basadas en Sanger da como producto final la **secuencia consenso** (en función de la muestra) **de una especie** de organismos o virus.
- Bajo este escenario se **secuenciaron numerosos genomas**, incluyendo el del ser humano (**HGP**). Este fue concluido en el 2004* (como especie) y en el 2007** se aplicó para un individuo.

*International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945 (2004).

**Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. The diploid genome sequence of an individual human. PLoS Biol. 2007; 5:e254.

Tecnologías de secuenciación: 1G

articles

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium*

** A list of authors and their affiliations appears in the Supplementary Information*

The sequence of the human genome encodes the genetic instructions for human physiology, as well as rich information about human evolution. In 2001, the International Human Genome Sequencing Consortium reported a draft sequence of the euchromatic portion of the human genome. Since then, the international collaboration has worked to convert this draft into a genome sequence with high accuracy and nearly complete coverage. Here, we report the result of this finishing process. The current genome sequence (Build 35) contains 2.85 billion nucleotides interrupted by only 341 gaps. It covers ~99% of the euchromatic genome and is accurate to an error rate of ~1 event per 100,000 bases. Many of the remaining euchromatic gaps are associated with segmental duplications and will require focused work with new methods. The near-complete sequence, the first for a vertebrate, greatly improves the precision of biological analyses of the human genome including studies of gene number, birth and death. Notably, the human genome seems to encode only 20,000–25,000 protein-coding genes. The genome sequence reported here should serve as a firm foundation for biomedical research in the decades ahead.

Tecnologías de secuenciación: 1G

OPEN ACCESS Freely available online

PLoS BIOLOGY

The Diploid Genome Sequence of an Individual Human

Samuel Levy^{1*}, Granger Sutton¹, Pauline C. Ng¹, Lars Feuk², Aaron L. Halpern¹, Brian P. Walenz¹, Nelson Axelrod¹, Jiaqi Huang¹, Ewen F. Kirkness¹, Gennady Denisov¹, Yuan Lin¹, Jeffrey R. MacDonald², Andy Wing Chun Pang², Mary Shago², Timothy B. Stockwell¹, Alexia Tsiamouri¹, Vineet Bafna³, Vikas Bansal³, Saul A. Kravitz¹, Dana A. Busam¹, Karen Y. Beeson¹, Tina C. McIntosh¹, Karin A. Remington¹, Josep F. Abril⁴, John Gill¹, Jon Borman¹, Yu-Hui Rogers¹, Marvin E. Frazier¹, Stephen W. Scherer², Robert L. Strausberg¹, J. Craig Venter¹

1 J. Craig Venter Institute, Rockville, Maryland, United States of America, **2** Program in Genetics and Genomic Biology, The Hospital for Sick Children, and Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada, **3** Department of Computer Science and Engineering, University of California San Diego, La Jolla, California, United States of America, **4** Genetics Department, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

Presented here is a genome sequence of an individual human. It was produced from ~32 million random DNA fragments, sequenced by Sanger dideoxy technology and assembled into 4,528 scaffolds, comprising 2,810 million bases (Mb) of contiguous sequence with approximately 7.5-fold coverage for any given region. We developed a modified version of the Celera assembler to facilitate the identification and comparison of alternate alleles within this individual diploid genome. Comparison of this genome and the National Center for Biotechnology Information human reference assembly revealed more than 4.1 million DNA variants, encompassing 12.3 Mb. These variants (of which 1,288,319 were novel) included 3,213,401 single nucleotide polymorphisms (SNPs), 53,823 block substitutions (2–206 bp), 292,102 heterozygous insertion/deletion events (indels)(1–571 bp), 559,473 homozygous indels (1–82,711 bp), 90 inversions, as well as numerous segmental duplications and copy number variation regions. Non-SNP DNA variation accounts for 22% of all events identified in the donor, however they involve 74% of all variant bases. This suggests an important role for non-SNP genetic alterations in defining the diploid genome structure. Moreover, 44% of genes were heterozygous for one or more variants. Using a novel haplotype assembly strategy, we were able to span 1.5 Gb of genome sequence in segments >200 kb, providing further precision to the diploid nature of the genome. These data depict a definitive molecular portrait of a diploid human genome that provides a starting point for future genome comparisons and enables an era of individualized genomic information.

Tecnologías de secuenciación: 1G

- Los **principales limitantes** de la tecnología ***Clone by clone*** están centrados en el **consumo de recursos** y en que el resultado es un consenso para una especie.

- Esta tecnología transforma en **impracticable la secuenciación de genomas de individuos**, perdiéndose en consecuencia el conocimiento correspondiente a la variabilidad poblacional de las especies (detección de SNPs* –*Single Nucleotide Polymorphisms*; SNVs* –*Single Nucleotide Variants*-, Indels o DIPs –inserciones o deleciones <1000 pb, *Deletion Insertion Polymorphisms*-, variantes estructurales –inserciones o deleciones >1000 pb). .
- Esta situación estimuló a la comunidad científica a **innovar** sobre este objetivo primario de la biología actual.

* SNPs vs SNVs depende de la frecuencia de ocurrencia

Adquisición de información genómica

El estímulo comienza con financiación...



by Maria Anderson

NEWS ANALYSIS

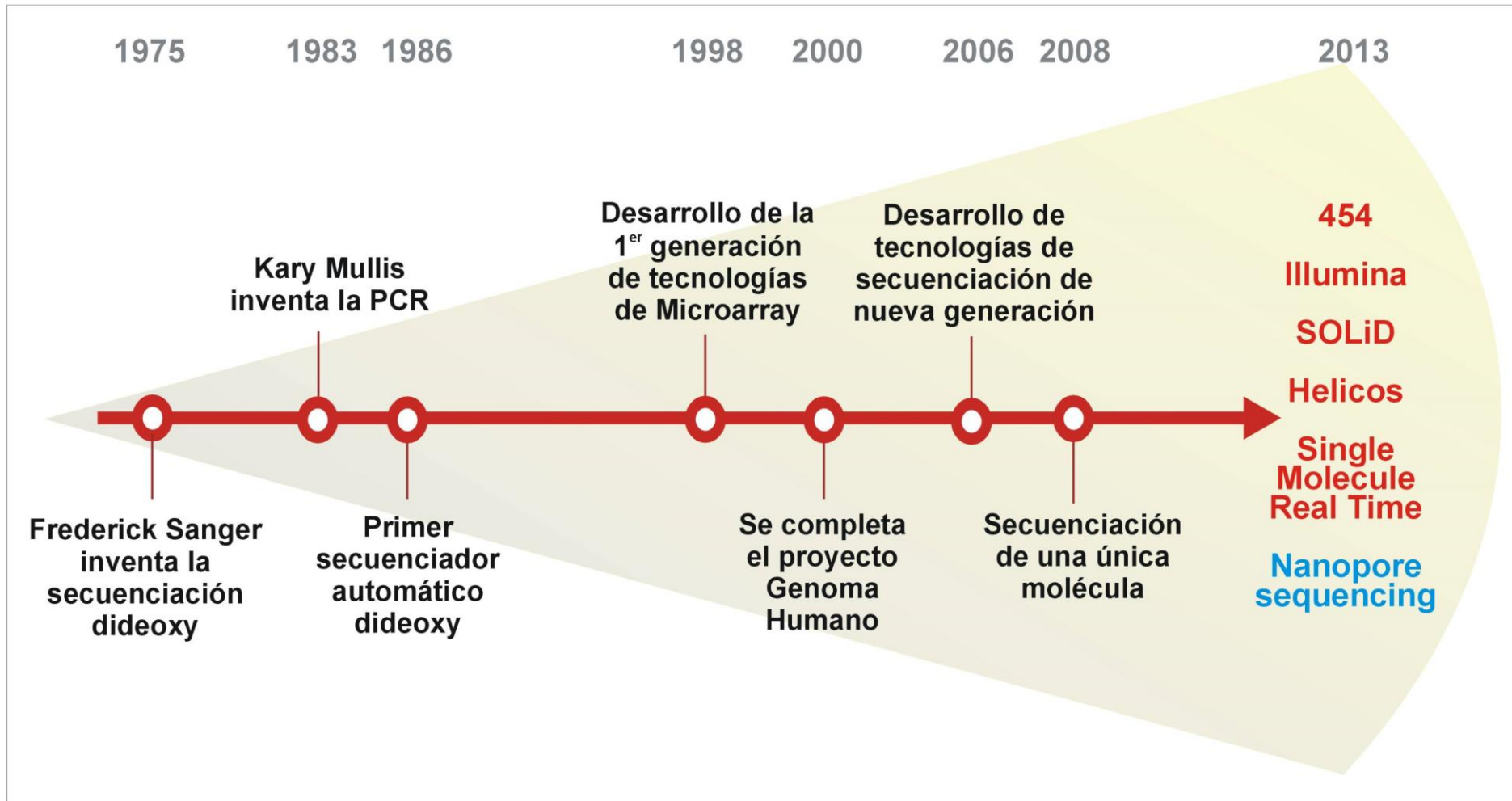
NIH offers \$1000 genome grant

Email: Maria Anderson - manderson@the-scientist.com
ews from The Scientist 2004, 5(1):20040223-05

Published 23 February 2004

The National Institutes of Health (NIH) is now soliciting proposals for funding to work toward the much-vaunted \$1000 genome. Earlier this month (February 12), the NIH published a request for applications (RFA) for grants to develop low-cost genome-sequencing technologies.

Adquisición de información genómica



Adquisición de información genómica

Next Generation Sequencing

- Los nuevos métodos que surgieron se denominan en su conjunto **tecnologías NGS** (*Next Generation Sequencing*) y son adecuadas para la realización de **WGS** (*Whole Genome Sequencing*), tanto *de novo* como resecuenciamiento.
- Las tecnologías **NGS** han revolucionado las ciencias biológicas **disminuyendo** dramáticamente el **consumo de recursos** (dinero, insumos, personal, tiempo) y posibilitando **secuenciar genomas de individuos**.
- Incluso, las tecnologías **NGS** permiten **secuenciar genomas** de entidades biológicas **sin** necesidad de su **previo aislamiento**, y **también** adquirir información genómica de **restos fósiles**.

Adquisición de información genómica

Next Generation Sequencing

- Las tecnologías **NGS** son del tipo **Shot gun**, dado que las lecturas de secuencias se realizan sobre fragmentos pequeños y si un orden pre-establecido.
- En general, las **lecturas** de las tecnologías **NGS** son **más cortas** que las logradas por el método de **Sanger** (50-500 nt vs 900-1500 nt).
- Existen varias **tecnologías NGS** comercialmente disponibles, tales como:

Tecnologías
de 2^{da}
generación

- Roche GS-FLX 454 Genome Sequencer (originalmente 454 sequencing).
- Illumina Genome Analyzer (originalmente Solexa technology).
- ABI SOLiD analyzer.
- Ion Torrent

- PACBIO
- Nanopore Technologies

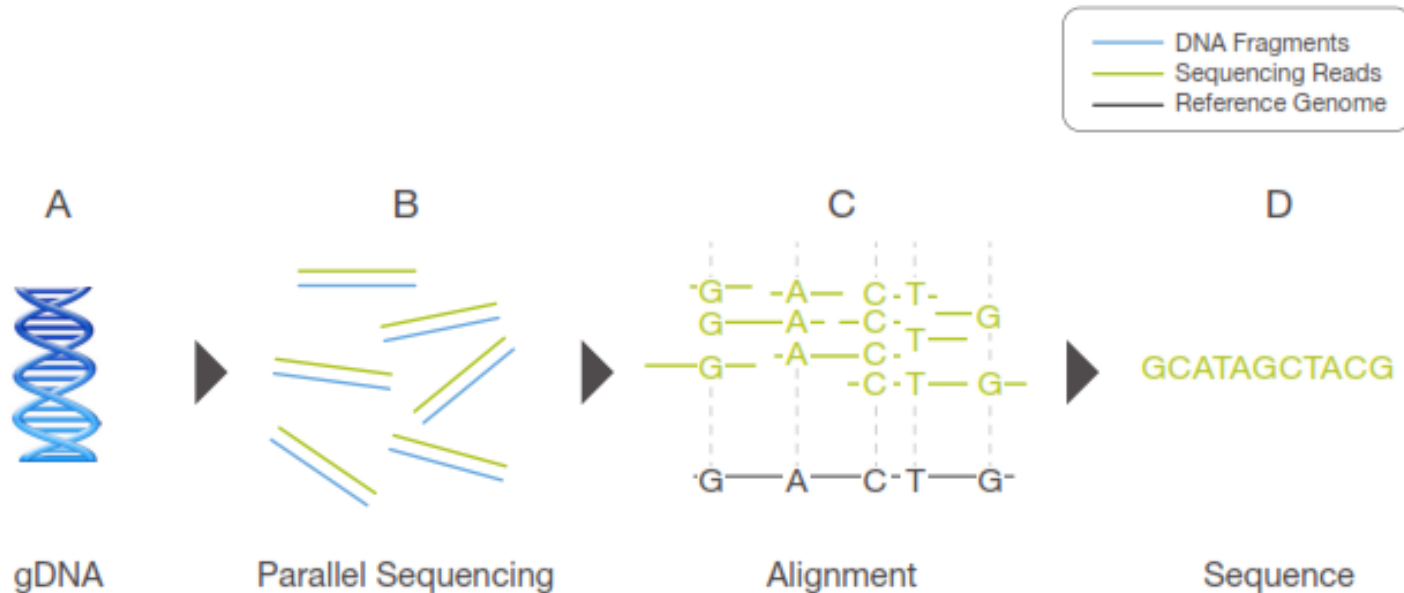
Tecnologías
de 3^{ra}
generación

→ **SMRT seq**
(Single Molecule Real
Time sequencing)

Adquisición de información genómica

Next Generation Sequencing

Figure 1: Conceptual Overview of Whole-Genome Resequencing

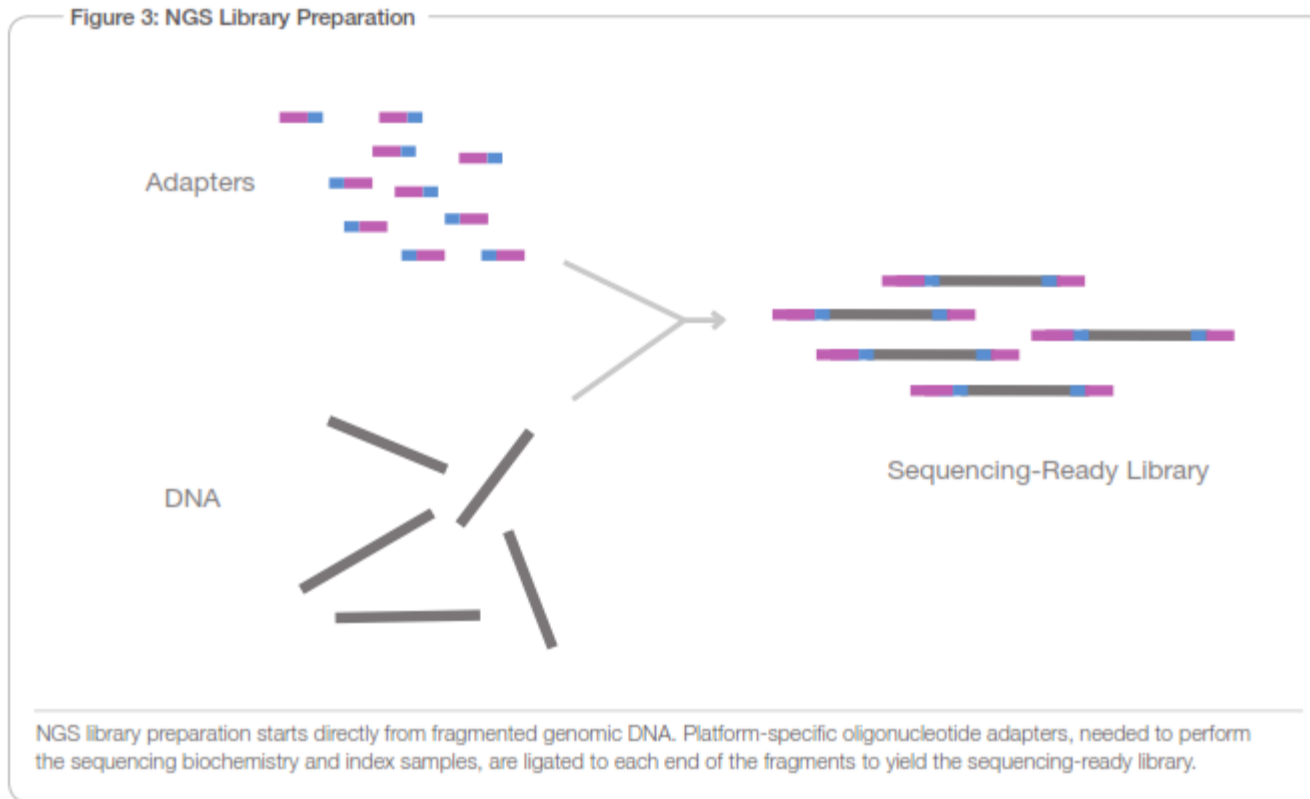


- A. Extracted gDNA.
B. gDNA is fragmented into a library of small segments that are each sequenced in parallel.
C. Individual sequence reads are reassembled by aligning to a reference genome.
D. The whole-genome sequence is derived from the consensus of aligned reads.

Adquisición de información genómica

Next Generation Sequencing

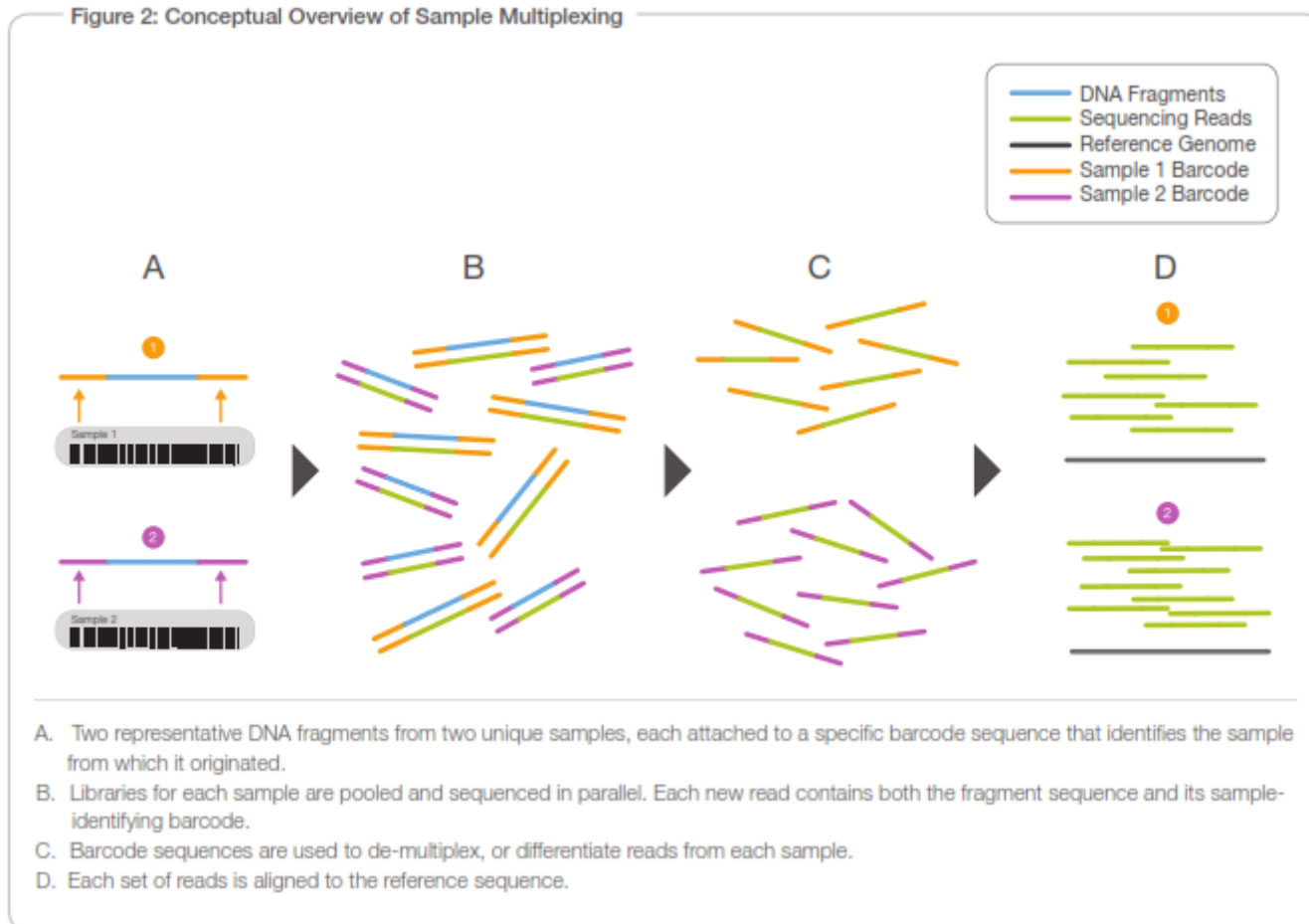
Estas **estrategias** también se basan en la generación de una **biblioteca**, pero *in vitro*.



Adquisición de información genómica

Next Generation Sequencing

Y también, permiten secuenciar varios genomas al mismo tiempo.



Descripción de los métodos más utilizados clasificados dentro de las tecnologías NGS

Tecnología 454

Tecnologías de secuenciación: 2G

NGS: tecnología 454

- La **tecnología 454** está basada en **pirosecuenciación** y es comercializada por la empresa Roche (<http://www.454.com/>).
- La **pirosecuenciación** fue desarrollada por **Mustafá Ronaghi*** e involucra la lectura de información nucleotídica a partir de síntesis enzimática.
- La **tecnología 454** fue la **primera NGS** en estar comercialmente disponible. Fue la primera de estas tecnologías aplicada para secuenciar el genoma de un individuo humano (James Watson)**.

*Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M., and Nyren, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**: 84

**Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008; 452:872–876.

Tecnologías de secuenciación: 2G

NGS: tecnología 454



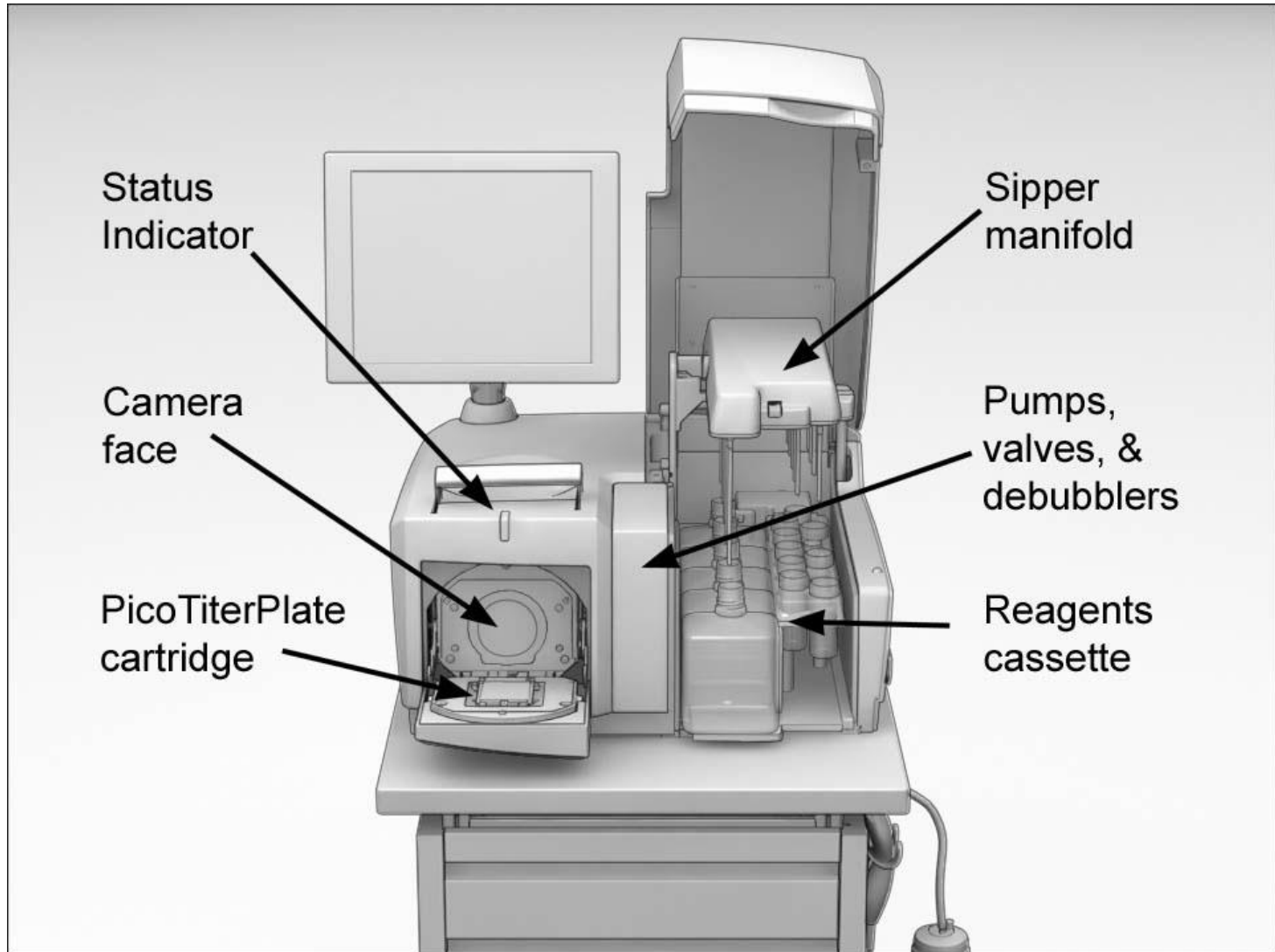
GS FL System



GS Junior System

Tecnologías de secuenciación: 2G

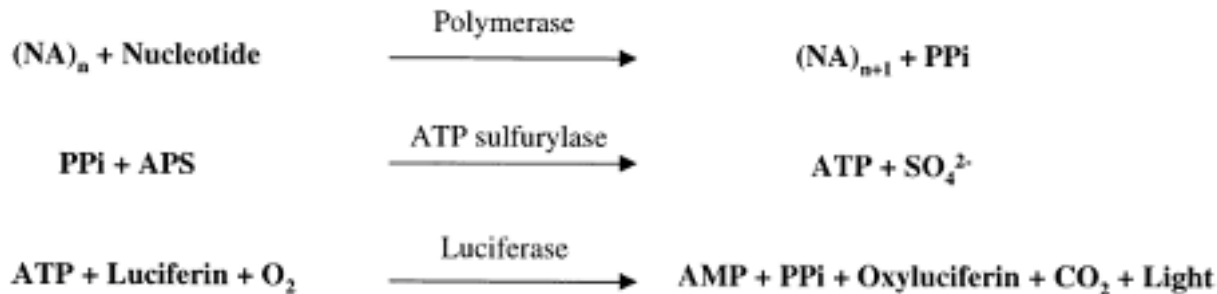
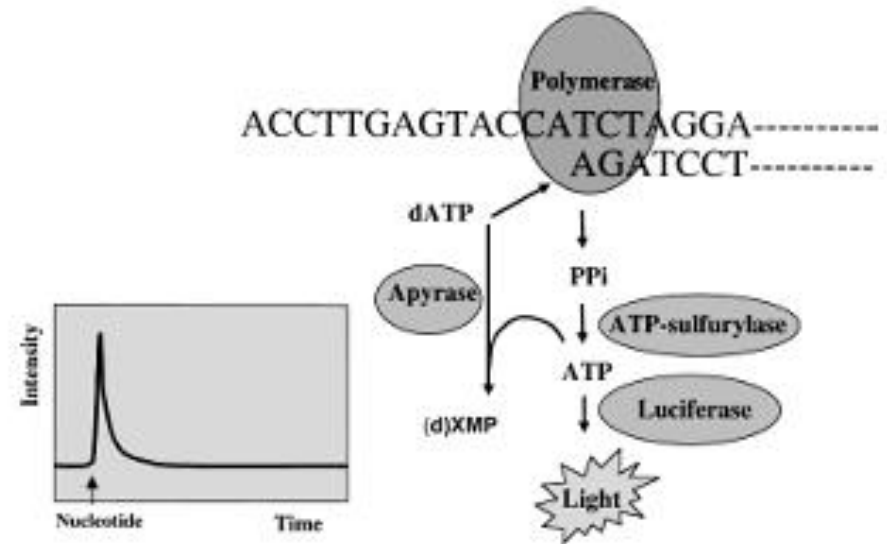
NGS: tecnología 454



Adquisición de información genómica

NGS: tecnología 454

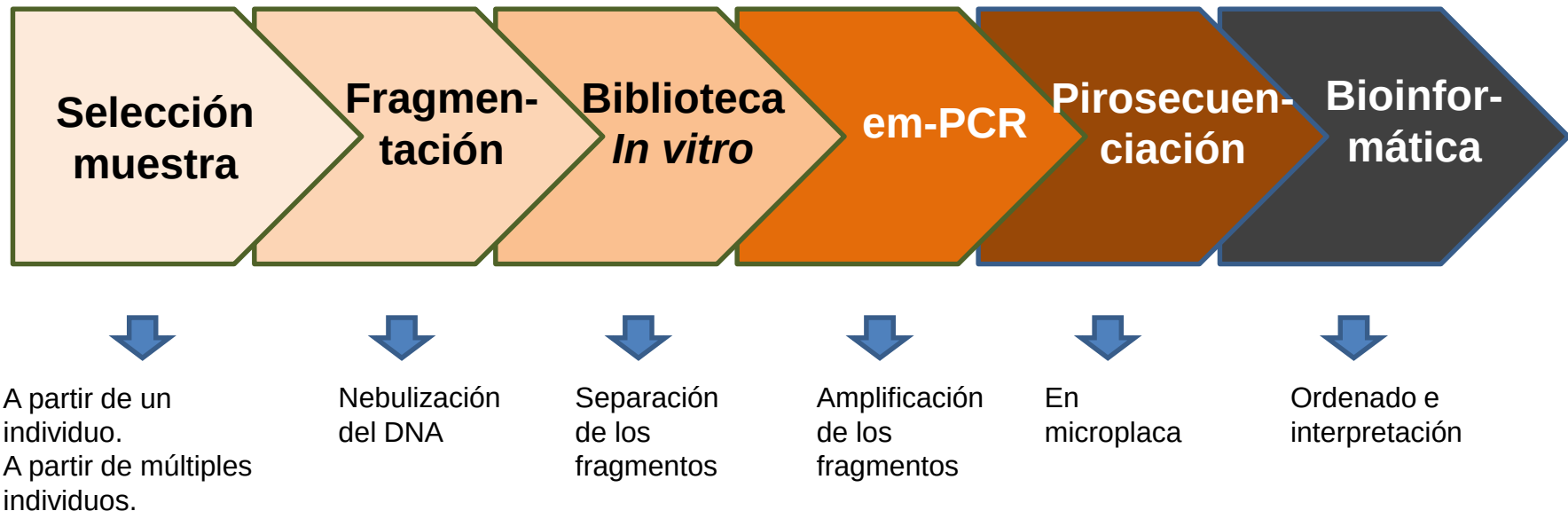
Pirosecuenciación



Tecnologías de secuenciación: 2G

NGS: tecnología 454

Etapas involucradas en la tecnología 454



Tecnologías de secuenciación: 2G

NGS: tecnología 454

Etapas involucradas en la tecnología 454



Selección muestra

La **muestra** original debe ser mayor a **500 ng** de DNA y derivar de uno o más individuos de una especie.

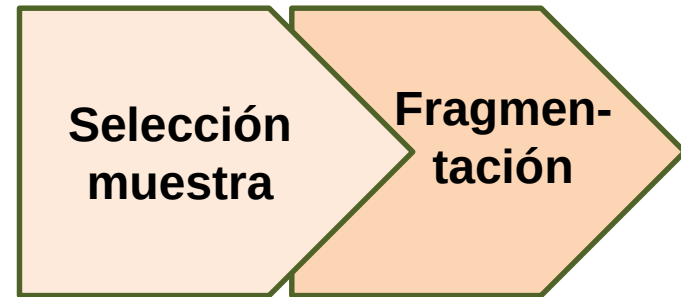


A partir de un individuo.
A partir de múltiples individuos.

Tecnologías de secuenciación: 2G

NGS: tecnología 454

Etapas involucradas en la tecnología 454



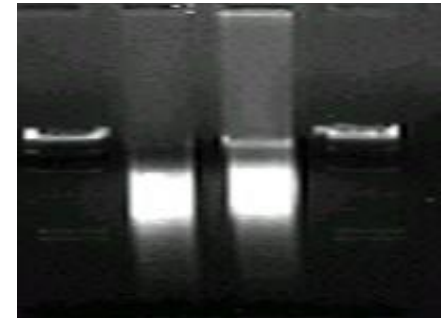
↓

A partir de un individuo.
A partir de múltiples individuos.

↓

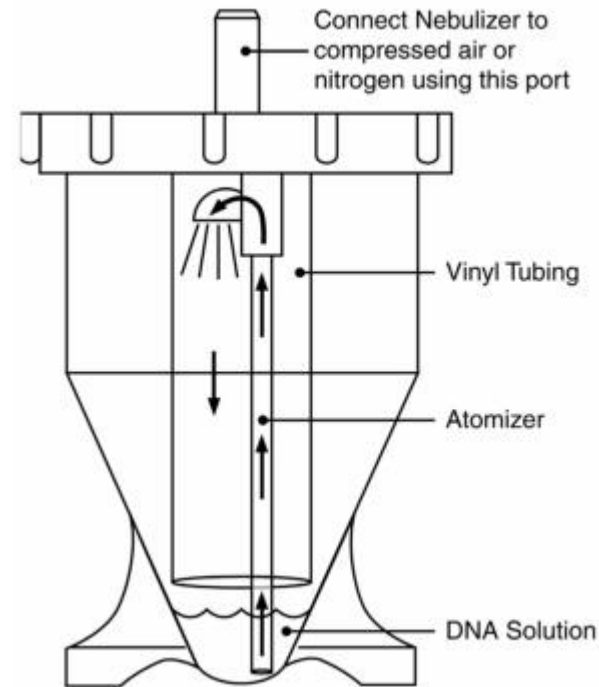
Nebulización del DNA

Realizada mediante nebulización de la solución de DNA en fragmentos de 400-800 pb



Tecnologías de secuenciación: 2G

NGS: tecnología 454

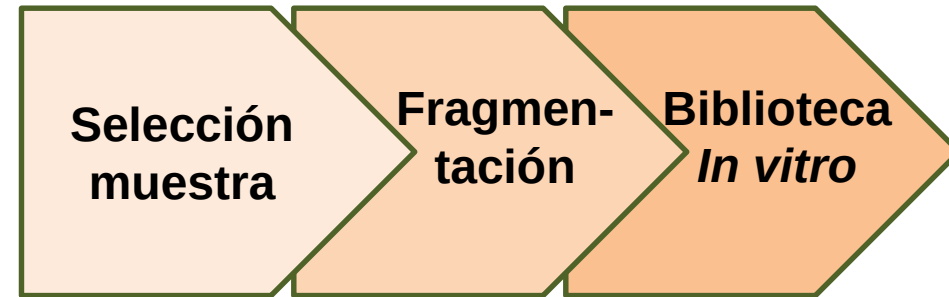


Nebulizador de DNA

Tecnologías de secuenciación: 2G

NGS: tecnología 454

Etapas involucradas en la tecnología 454



↓

A partir de un individuo.
A partir de múltiples individuos.

↓

Nebulización del DNA

↓

Separación de los fragmentos

Los fragmentos de **dsDNA** son **modificados** en sus extremos mediante la ligación de **adaptadores**, y luego separados entre sí para formar una colección de **ssDNAs unidos a microesferas**.

Tecnologías de secuenciación: 2G

NGS: tecnología 454

- El **dsDNA fragmentado** de la muestra es **unido** a **dos adaptadores** (**A** y **B**, este **último biotinilado**). Los adaptadores están desfosforilados para evitar ligación entre ellos.
- El dsDNA con los adaptadores es **luego** tratado con **PNK** y **DNA ligasa** (reparación de *nicks*).
- **Después**, el dsDNA es tratado con **T4 DNA Polimerasa**, **Klenow** y **PNK** (para transformar los extremos en romos).
- **Posteriormente**, el **dsDNA biotinilado** es **capturado** con **microesferas** con **avidina**.
- **A partir de las microesferas**, se liberan los **ssDNA** **flanqueados** por oligonucleótidos **A** y **B** derivados de los adaptadores.

Tecnologías de secuenciación: 2G

NGS: tecnología 454

Ejemplo de secuencia de los adaptadores...

MID-1 Adaptor A:

```
5' -CCATCTCATCCCTGCGTGTCTCCGACTCAGACGAGTGCGT-3'   Ti-MID1-A
                        3' -AGAGGCTGAGTCTGCTCACGCA-5'   Ti-MID1-Aprime
```

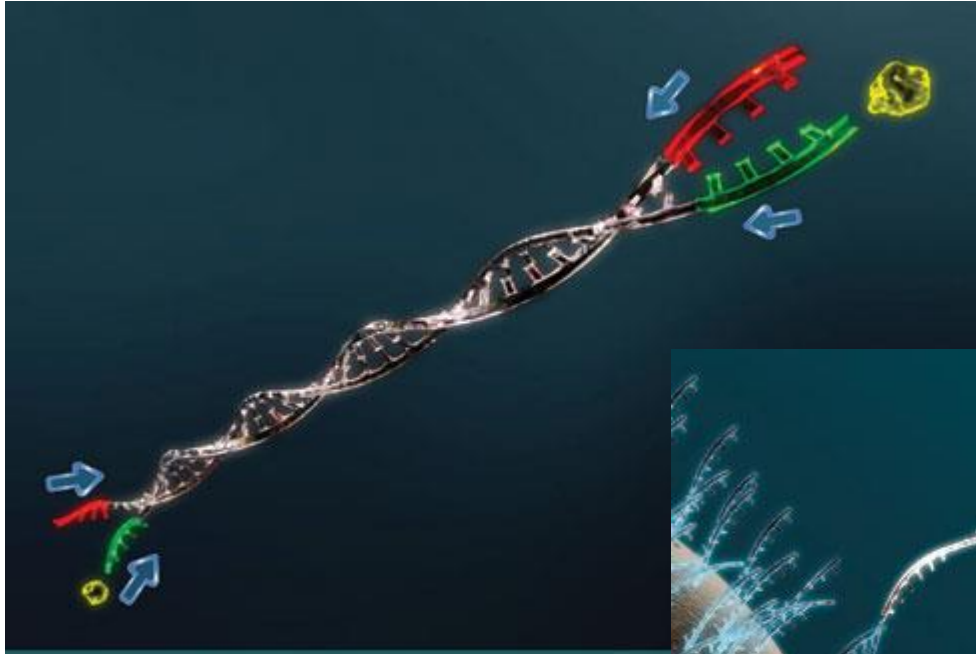
MID Adaptor B (Common):

```
5' Biotin-TEG-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG-3'   Ti-MID-B
                        3' - ACCGTCAGAGTC-5'   Ti-MID-Bprime
```

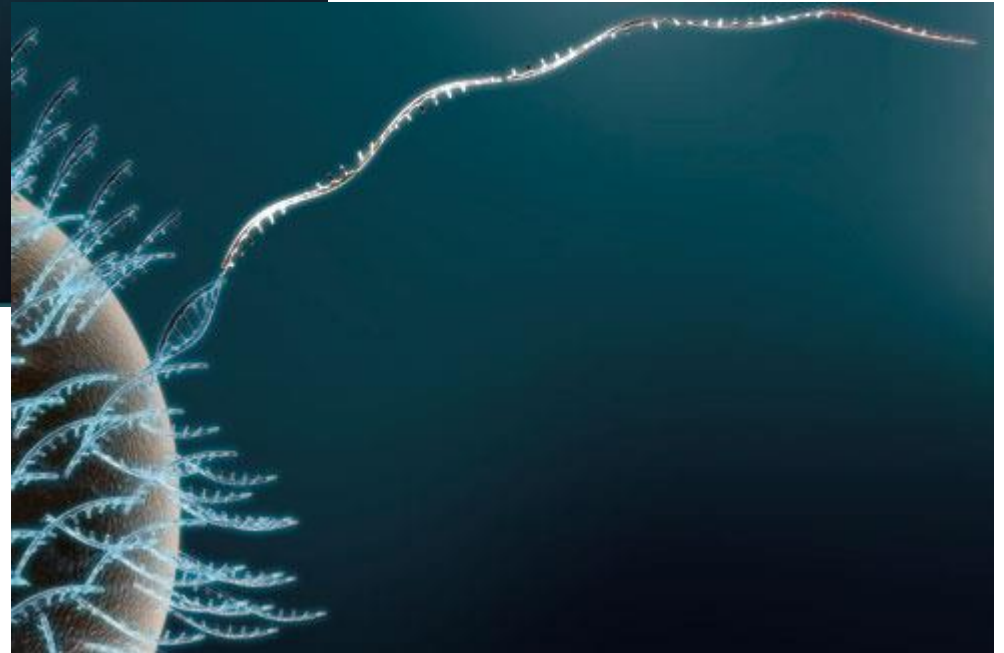
- Phosphorothioate-modified bases are not shown in this figure for ease of sequence alignment
- Sequencing Key is indicated in **GREEN**. 10 bp MID sequence is shown in **YELLOW**.

Tecnologías de secuenciación: 2G

NGS: tecnología 454



Fragmentos de dsDNA
modificados en sus
extremos

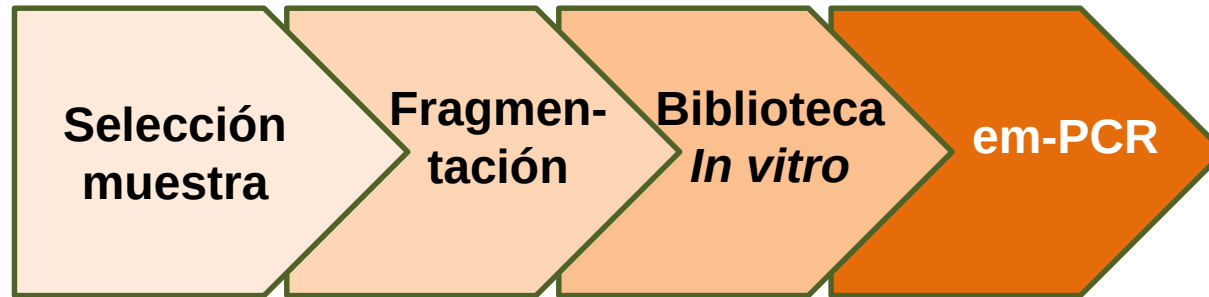


Colección de ssDNA aislados

Tecnologías de secuenciación: 2G

NGS: tecnología 454

Etapas involucradas en la tecnología 454



↓

A partir de un individuo.
A partir de múltiples individuos.

↓

Nebulización del DNA

↓

Separación de los fragmentos

↓

Amplificación de los fragmentos

Los ssDNA aislados son **multiplicados** mediante **PCR** en **emulsión** y retenidos en microesferas.

Tecnologías de secuenciación: 2G

NGS: tecnología 454

Los ssDNA flanqueados por los oligonucleótidos A y B son capturados por microesferas conteniendo oligonucleótido anti-B. Luego, son emulsionados en una mezcla de **solución acuosa** (mix de PCR) y **aceite**. Finalmente, se realiza un **ciclado térmico** para multiplicar los DNA contenidos en cada microesfera.



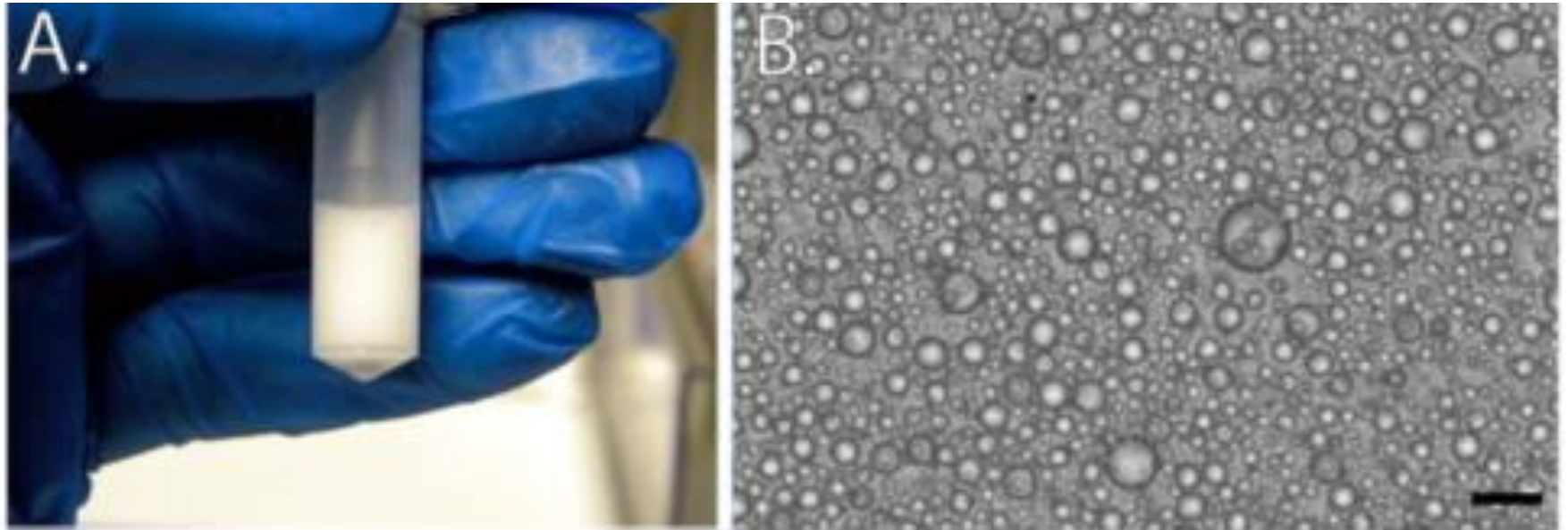
PCR en emulsión



Tecnologías de secuenciación: 2G

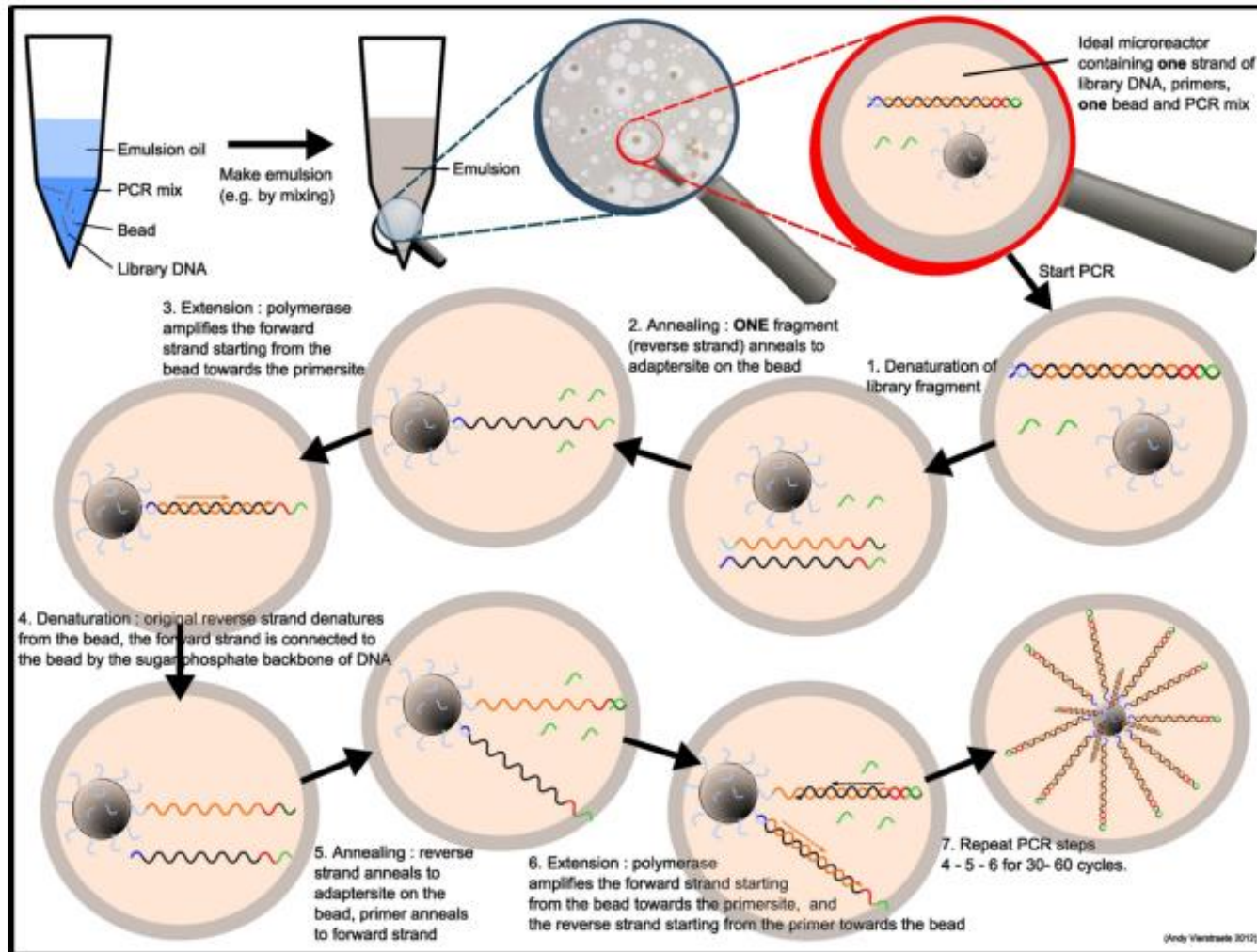
NGS: tecnología 454

Apariencia de una emulsión para PCR.



Tecnologías de secuenciación: 2G

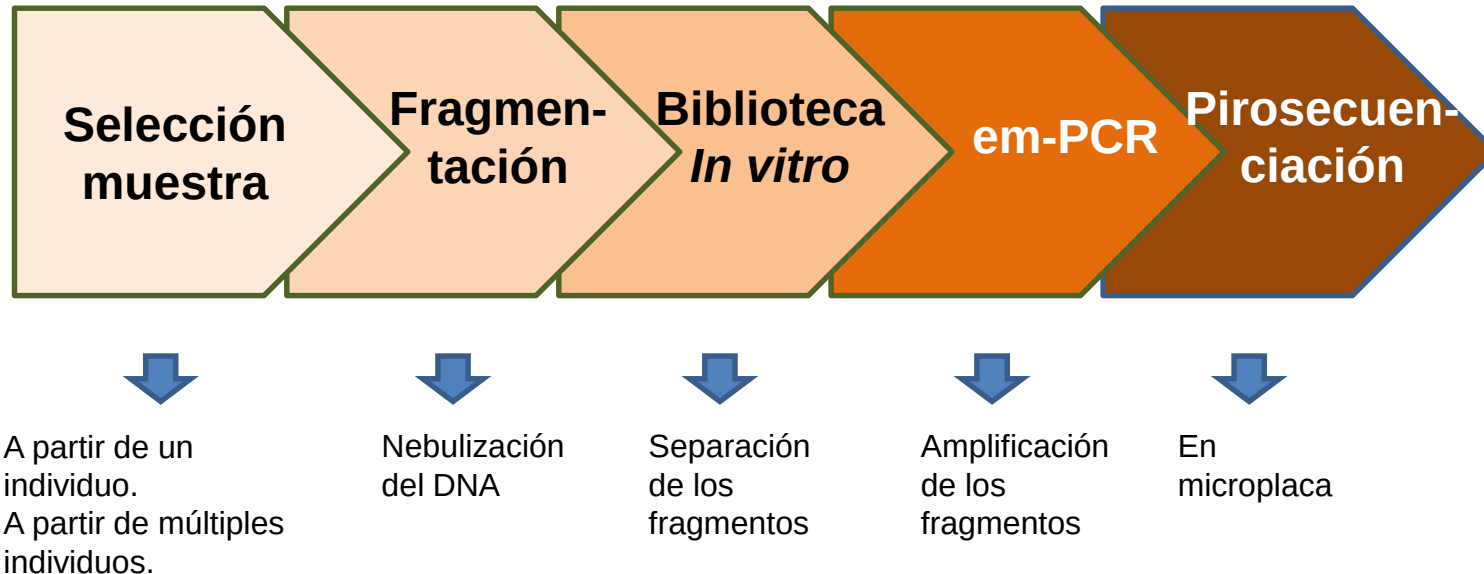
NGS: tecnología 454



Tecnologías de secuenciación: 2G

NGS: tecnología 454

Etapas involucradas en la tecnología 454



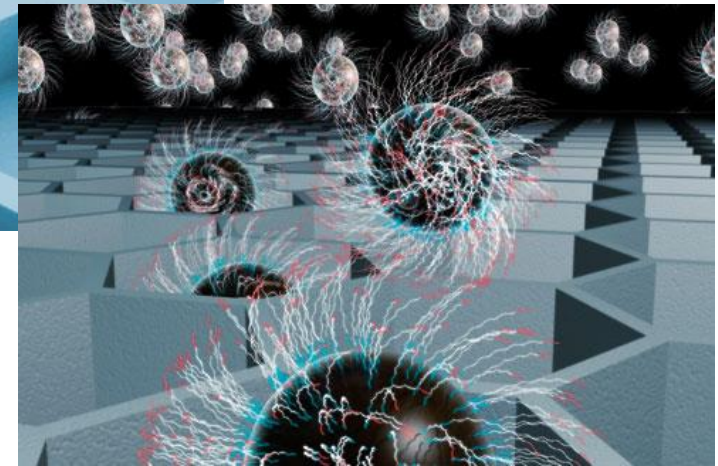
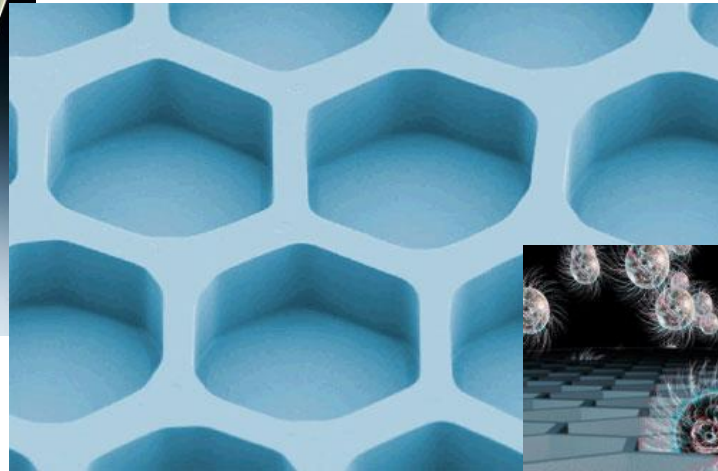
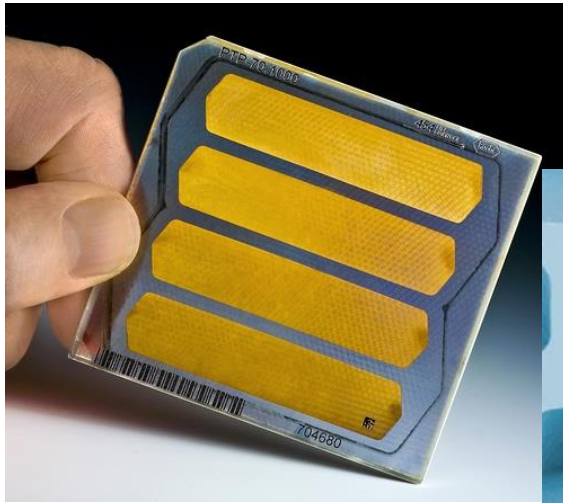
Tecnologías de secuenciación: 2G

NGS: tecnología 454

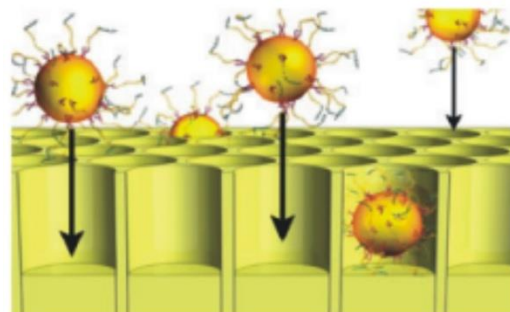
- Las **microesferas** conteniendo los **productos de PCR** son **tratados** con **solución alcalina** para que queden sólo los **ssDNA** a ser secuenciados.
- Luego, las **microesferas** se disponen en una **microplaca** para proceder a la pirosecuenciación.
- En la **microplaca**, se **agregan microesferas** más pequeñas con las enzimas ***Sulfurilasa*** y ***Luciferasa***.
- Luego, se agregan de a **pulsos mezcla de síntesis** (DNA polimerasa, 1 dNTP, APS, luciferina).
- Si el **nucleótido** es **incorporado**, se **produce PPi**, lo cual derivará en la posterior producción de ATP, oxiluciferina y **luz**.
- Para continuar, se **“limpia”** cada reacción con la enzima **Apirasa** y se repite el ciclo.

Tecnologías de secuenciación: 2G

NGS: tecnología 454



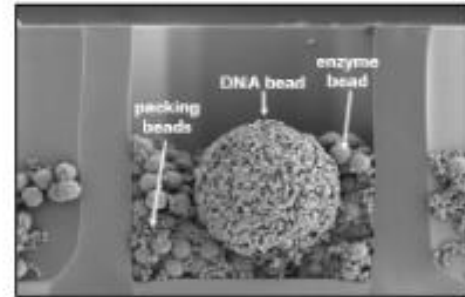
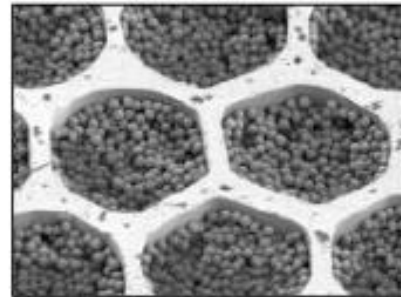
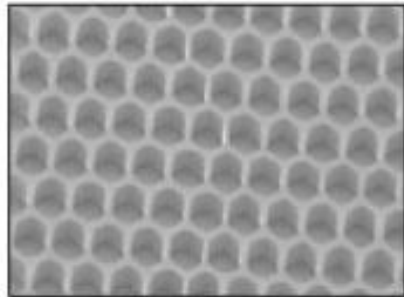
Microplacas donde se disponen las microesferas



Tecnologías de secuenciación: 2G

NGS: tecnología 454

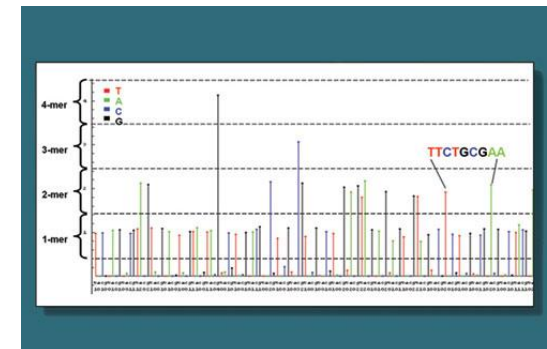
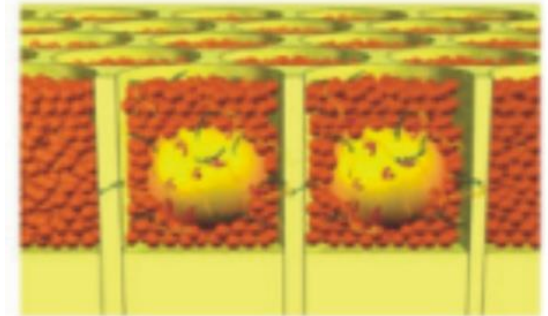
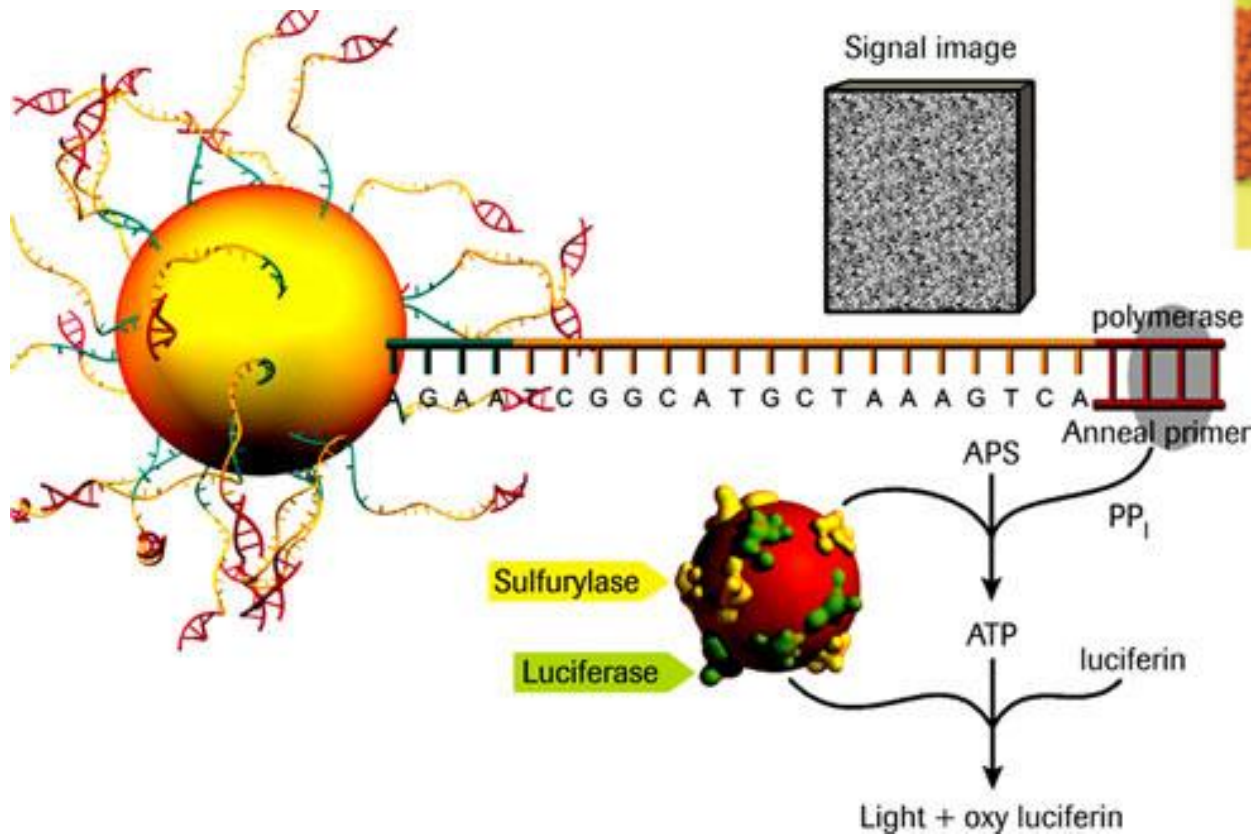
Detalle de las celdas de la microplaca conteniendo las microesferas.



Tecnologías de secuenciación: 2G

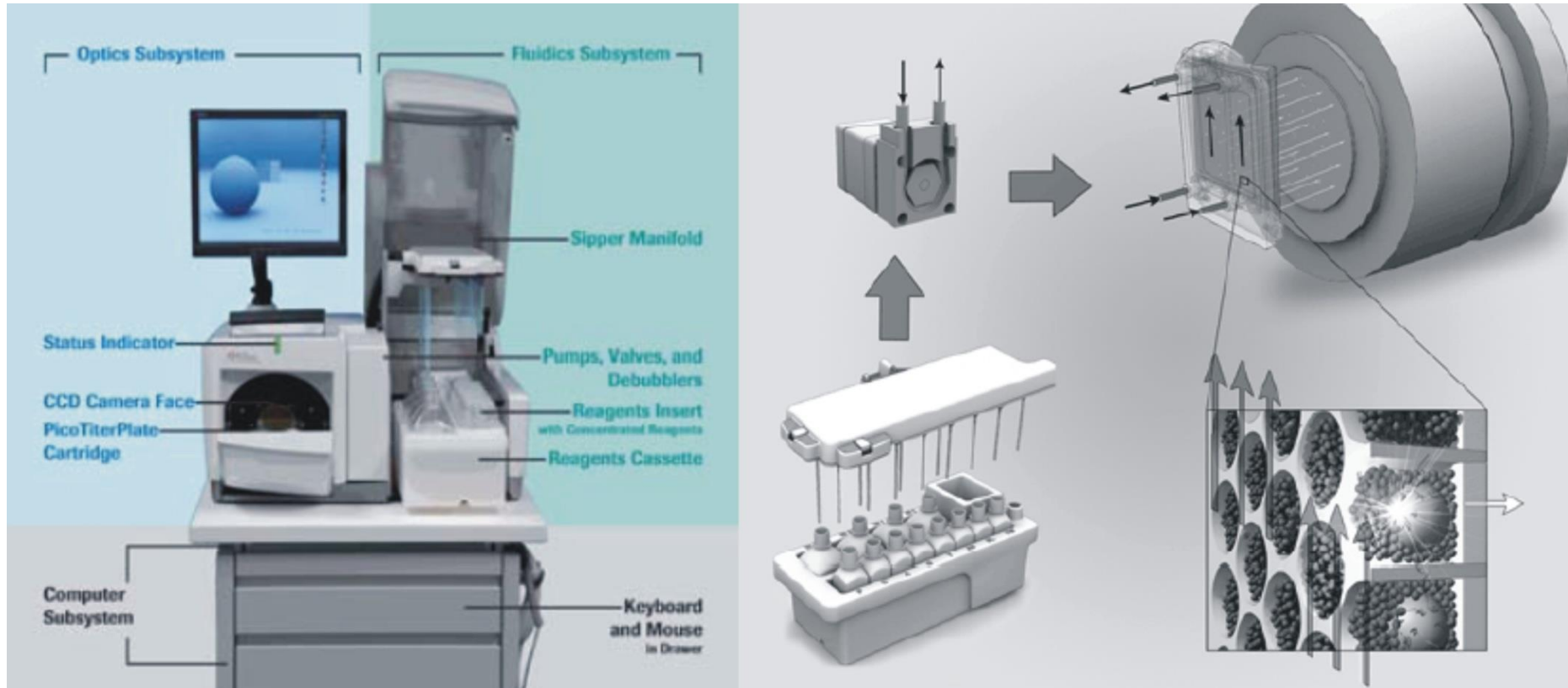
NGS: tecnología 454

Reacción de secuenciación



Tecnologías de secuenciación: 2G

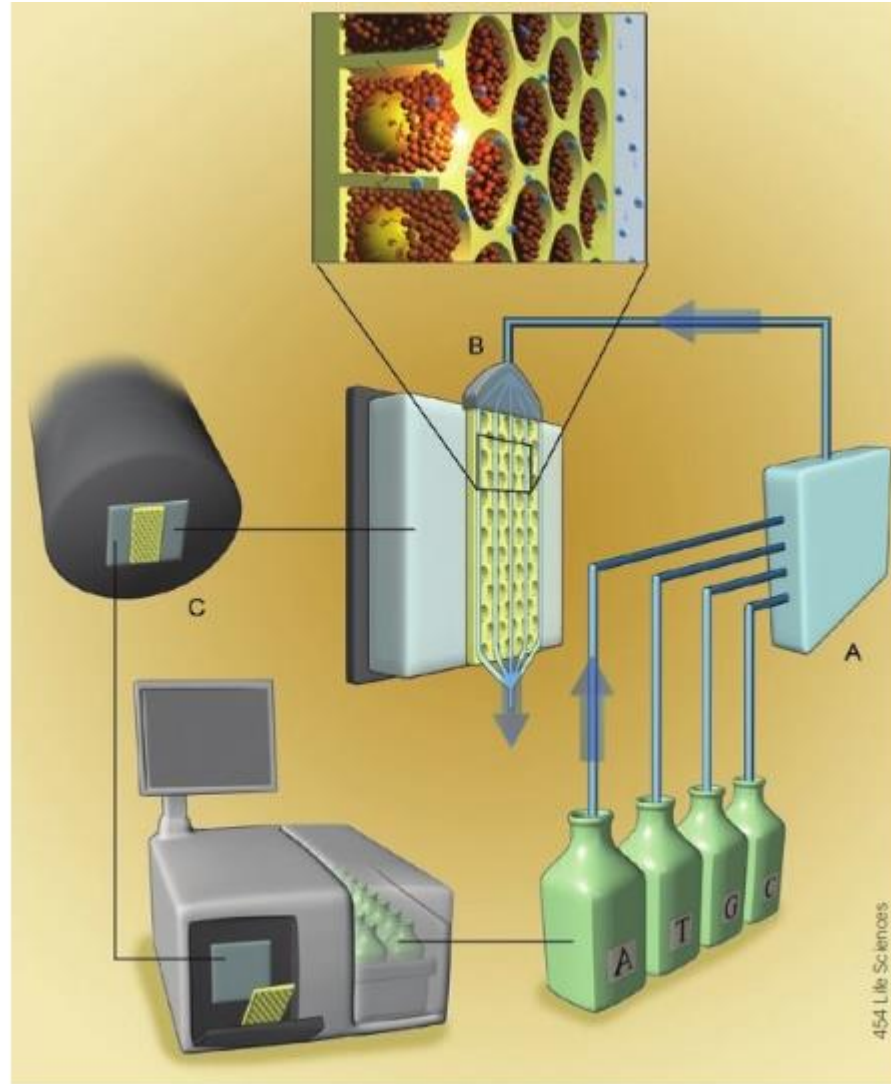
NGS: tecnología 454



Equipamiento por dentro

Tecnologías de secuenciación: 2G

NGS: tecnología 454



Equipamiento por dentro

Descripción de los métodos más utilizados clasificados dentro de las tecnologías NGS

Tecnología *Illumina*

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

- La **tecnología Illumina** fue desarrollada por **Solexa**, compañía que fue adquirida por la primera en 2007.
- Este procedimiento de secuenciación se basa en ***reversible terminator chemistry****, involucrando síntesis de DNA.
- Los **terminadores** de síntesis son *39-O-azidometil 29-deoxinucleosido trifosfatos* (A, C, G y T) marcados con **fluoróforos distintos y reversibles**.
- Las **secuencias** se determinan sobre ***clusters*** de moléculas, semejantes a colonias bacterianas (**polonias**).

*Bentley, David R.; et al. (6 November 2008). "Accurate whole human genome sequencing using reversible terminator chemistry". *Nature* **456** (7218): 53–59.

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

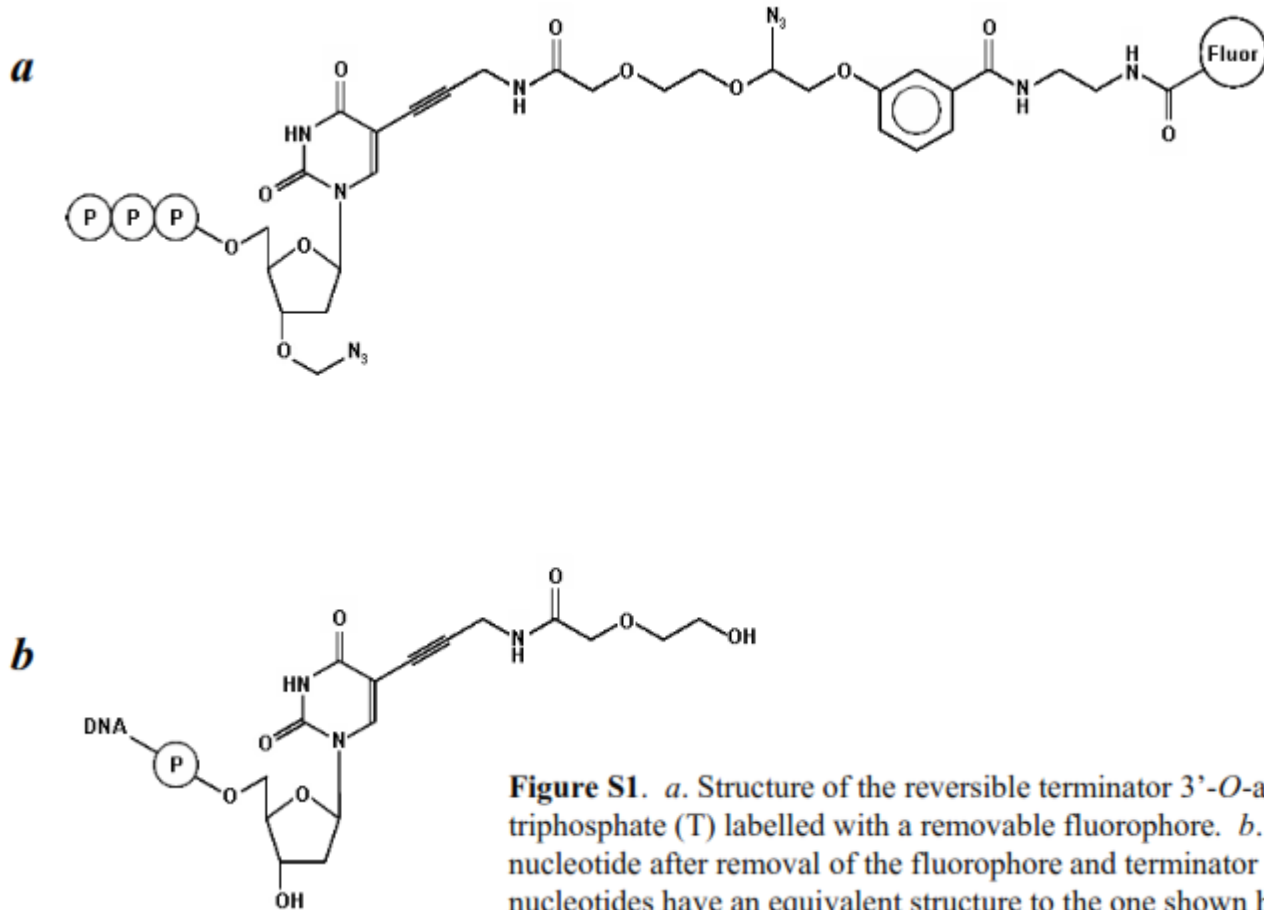


Figure S1. *a.* Structure of the reversible terminator 3'-O-azidomethyl 2'-deoxythymine triphosphate (T) labelled with a removable fluorophore. *b.* Structure of the incorporated nucleotide after removal of the fluorophore and terminator group. Each of the four nucleotides have an equivalent structure to the one shown here, except for the different base and a corresponding base-specific fluor.

Adquisición de información genómica

NGS: tecnología Illumina

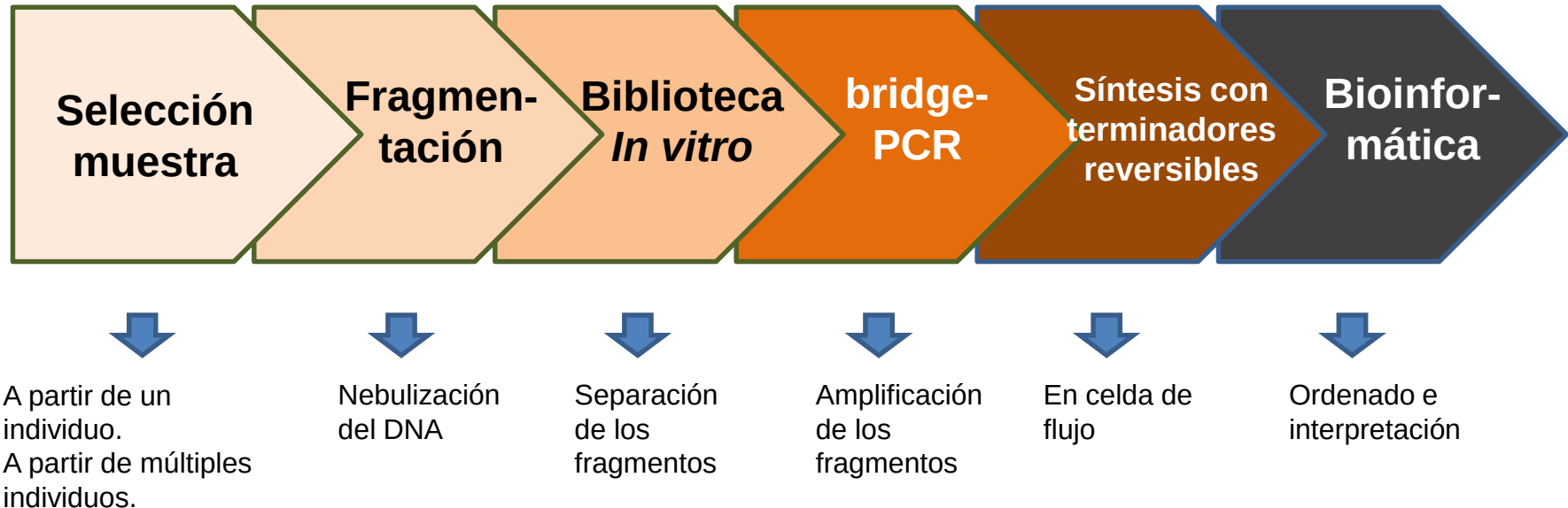


Equipamiento que utiliza el sistema de terminadores reversibles de cadena sintetizada

Adquisición de información genómica

NGS: tecnología Illumina

Etapas involucradas en la tecnología Illumina



Adquisición de información genómica

NGS: tecnología Illumina

Etapas involucradas en la tecnología Illumina

Selección muestra

La **muestra** original debe ser mayor a **50 ng** de DNA y derivar de uno o más individuos de una especie.

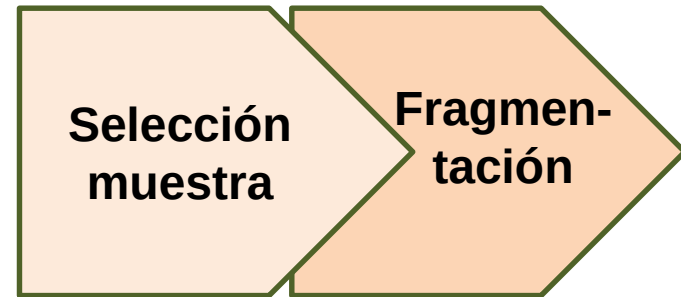


A partir de un
individuo.
A partir de múltiples
individuos.

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

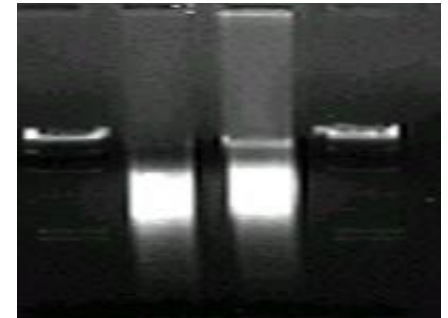
Etapas involucradas en la tecnología Illumina



A partir de un individuo.
A partir de múltiples individuos.

Nebulización del DNA

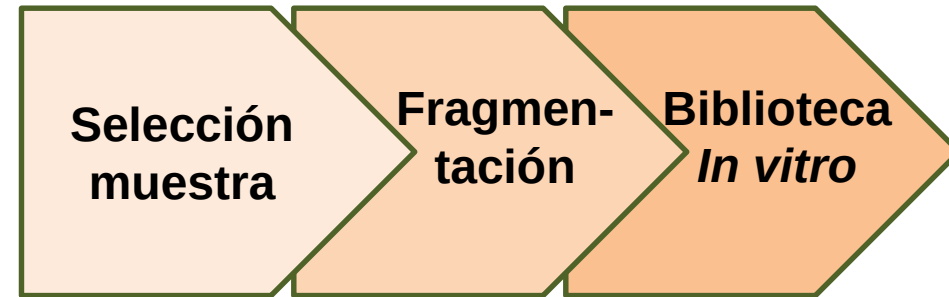
Realizada mediante **nebulización** de la solución de DNA en fragmentos de 200-400 pb



Adquisición de información genómica

NGS: tecnología Illumina

Etapas involucradas en la tecnología Illumina



↓

A partir de un individuo.
A partir de múltiples individuos.

↓

Nebulización del DNA

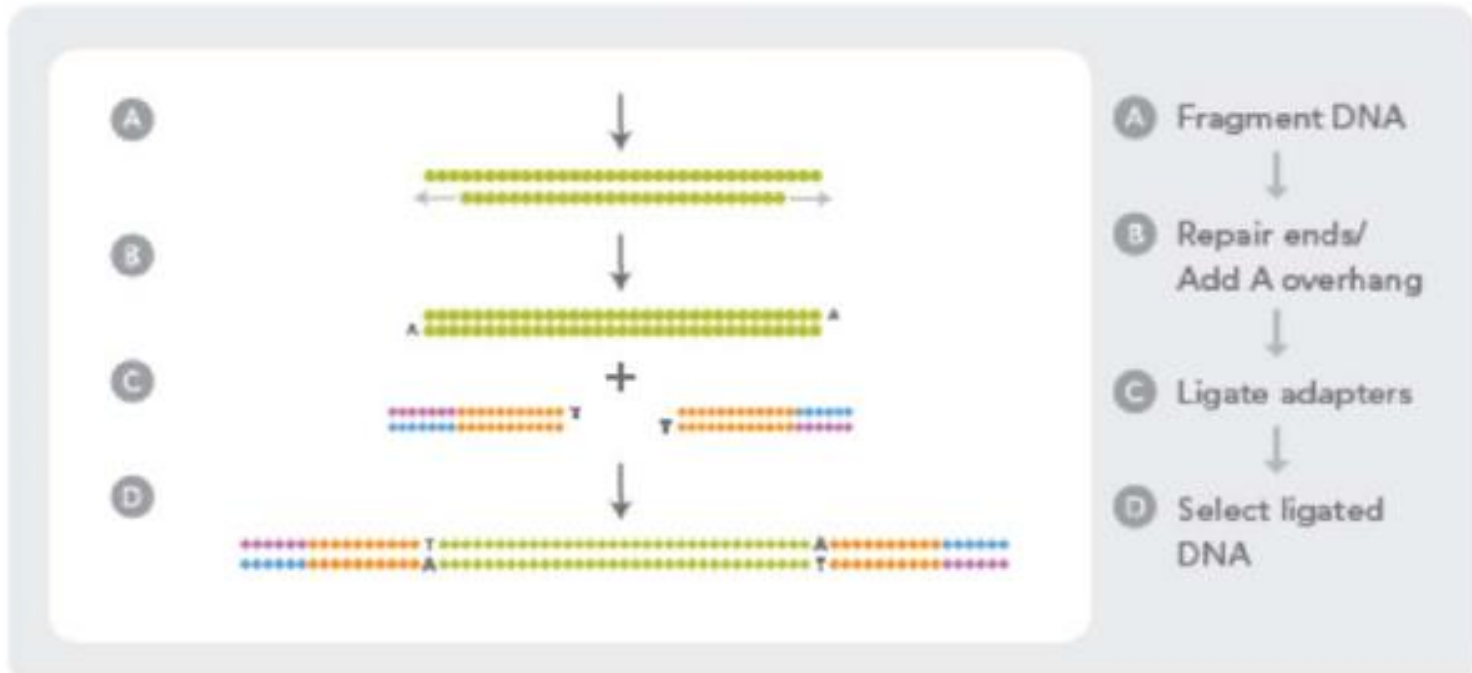
↓

Separación de los fragmentos

Los fragmentos de **dsDNA** son **modificados** en sus extremos mediante la **ligación de adaptadores**, y luego separados entre sí para formar una colección de ssDNAs unidos a una celda de flujo.

Adquisición de información genómica

NGS: tecnología Illumina

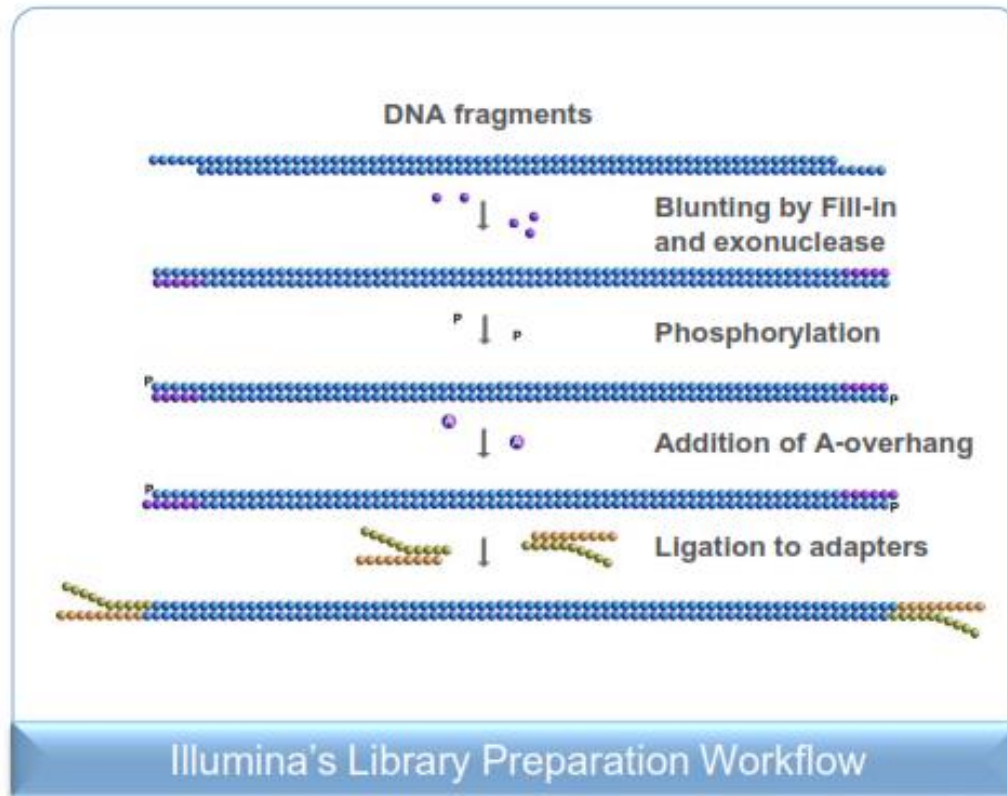


Procedimiento de reparación de extremos y ligación de adaptadores.

La mezcla de reparación contiene *T4 DNA polimerasa*, *Klenow* y *Polinucleótido quinasa (PNK)*. La introducción de Aes 3' extendidas se realiza con *Klenow*.

Tecnologías de secuenciación: 2G

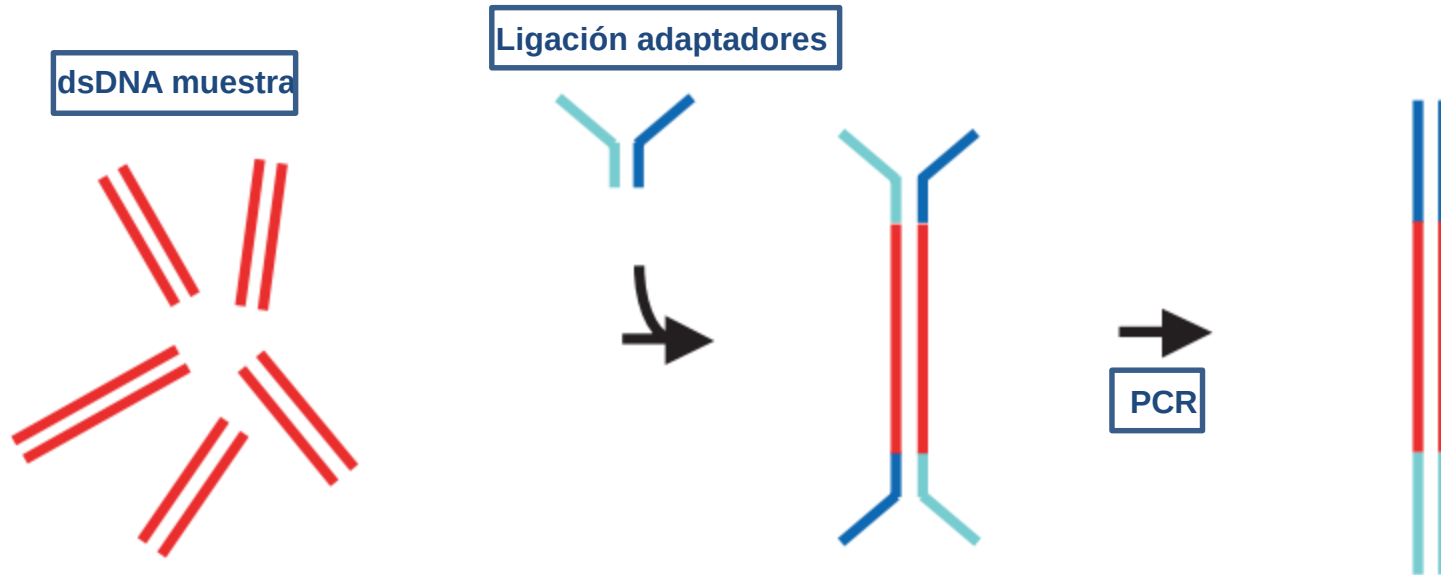
NGS: tecnología Illumina



Los **adaptadores** poseen **secuencias no complementarias** en los **extremos**.
Luego de esto puede realizarse una PCR (6-18 ciclos) con *primers* que hibridan en los adaptadores. Esto permite la secuenciación posterior de cada una de las cadenas.

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

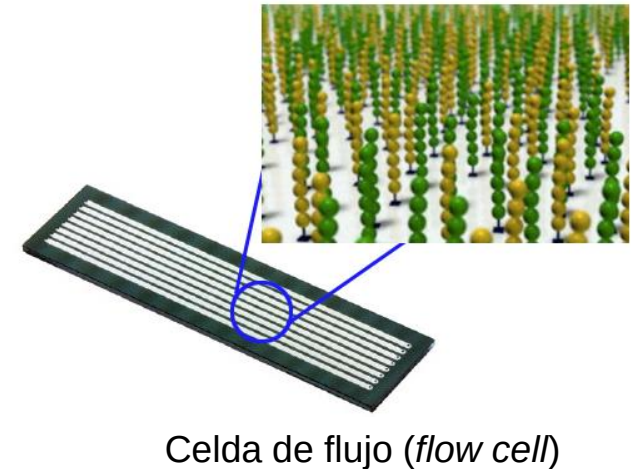
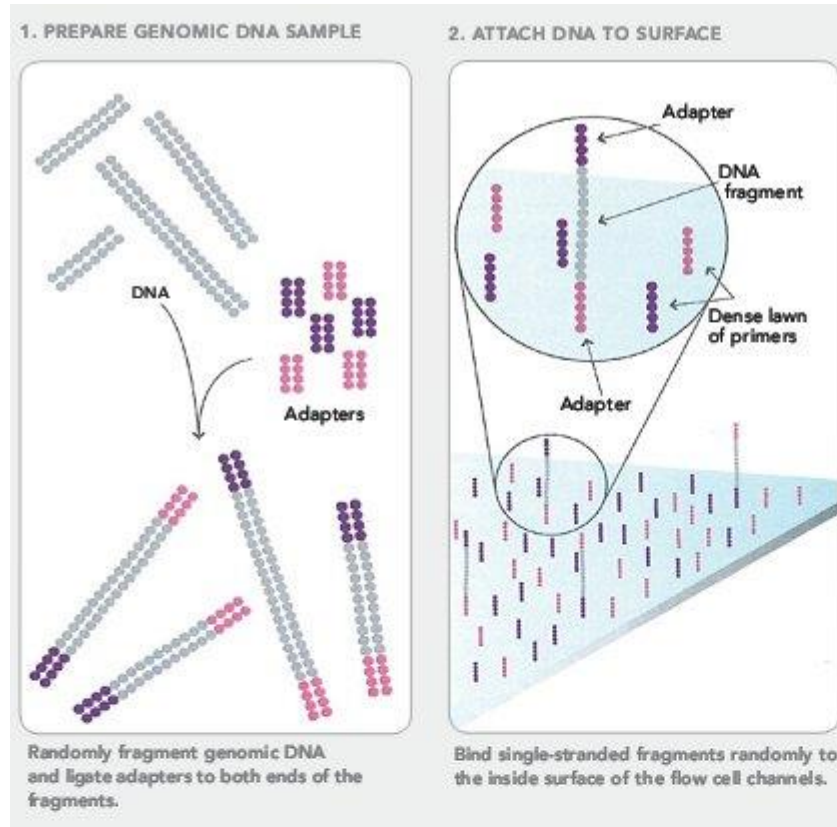


Resumen de los pasos involucrados en la preparación de la colección de fragmentos con las modificaciones en sus extremos.

Todos los fragmentos amplificados quedan con secuencias diferenciales en los extremos.

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

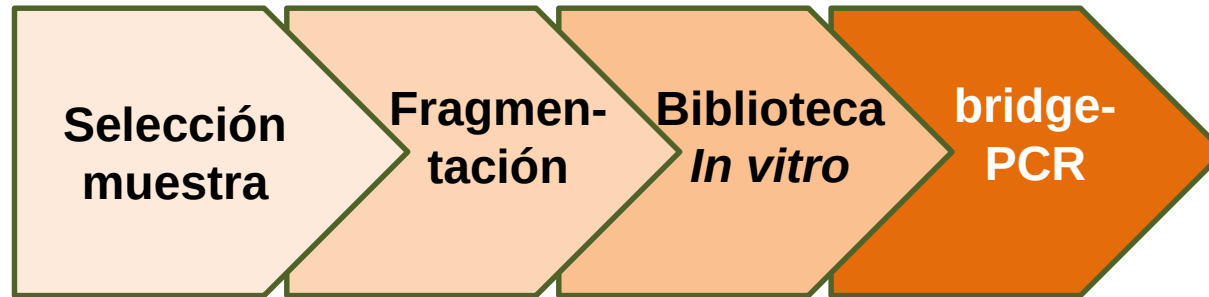


Los **dsDNA se desnaturalizan e hibridan en una celda**. Primero se realiza un tratamiento alcalino, luego dilución en solución salina –SSC 5X más Tween 20- y finalmente se agregan a la celda de flujo a alta temperatura (96°C). Posteriormente, la temperatura se reduce lentamente hasta 40°C para favorecer el *annealing* con oligonucleótidos complementarios a los adaptadores que se encuentran inmovilizados en una celda.

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

Se realiza una amplificación tipo *solid-phase bridge* o *bridge-PCR*



↓

A partir de un individuo.
A partir de múltiples individuos.

↓

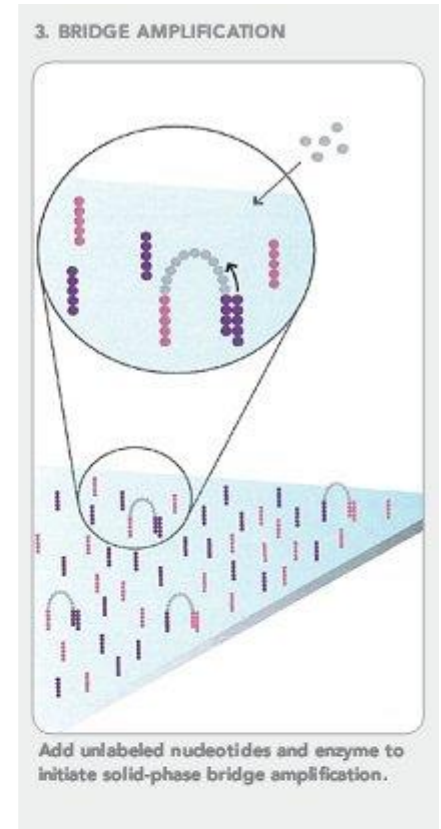
Nebulización del DNA

↓

Separación de los fragmentos

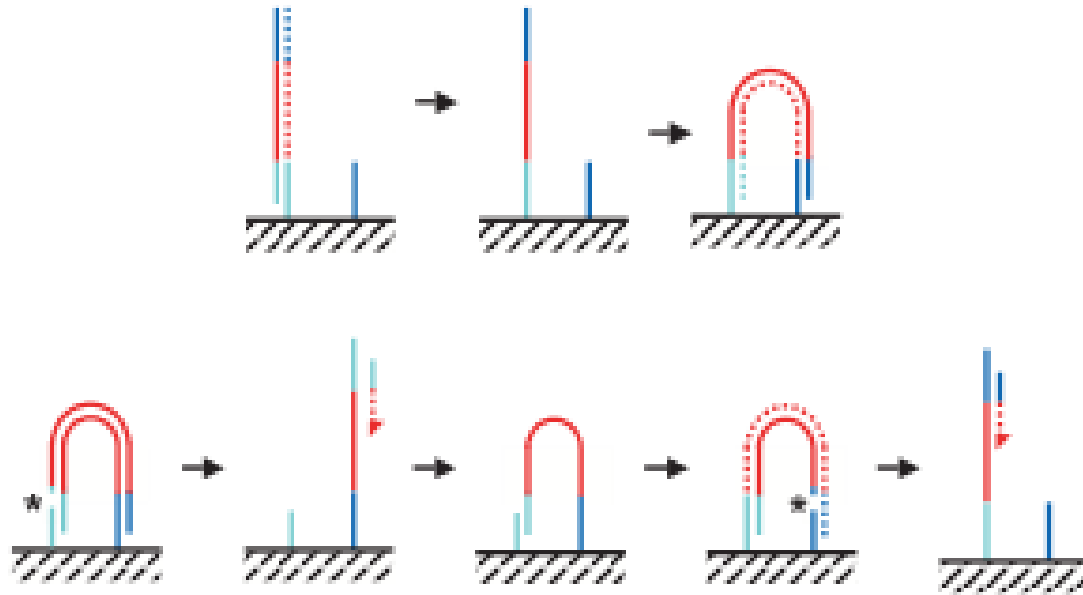
↓

Amplificación de los fragmentos



Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

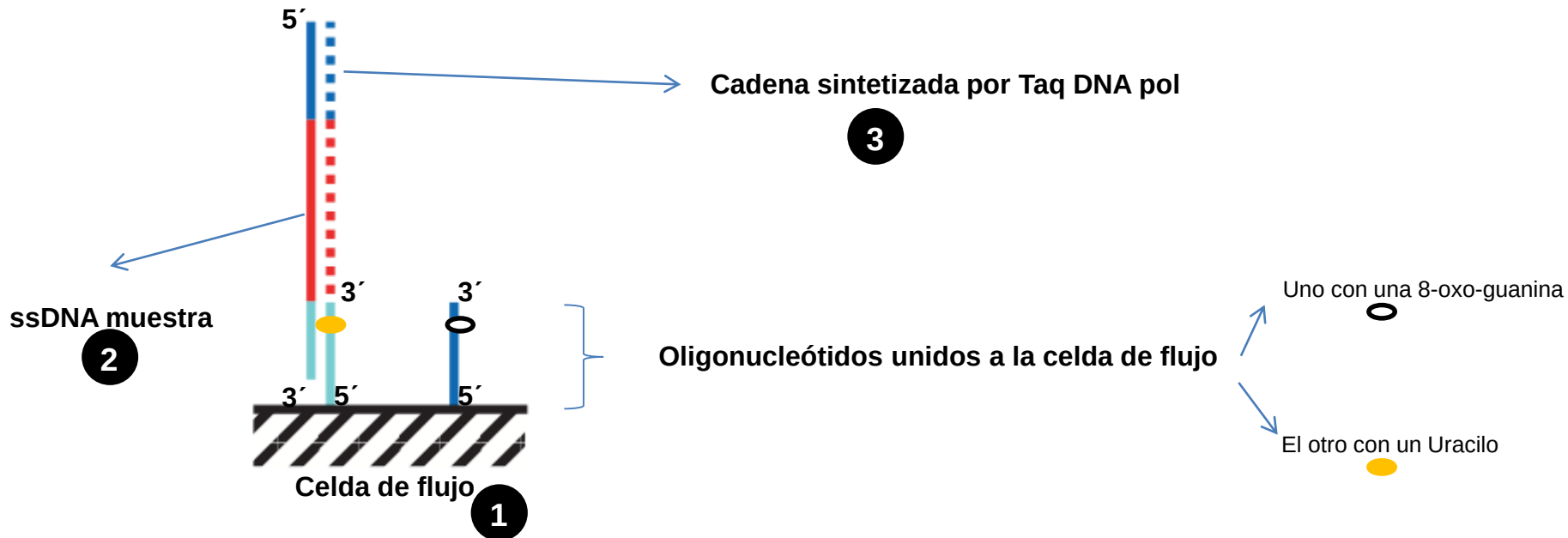


Bridge-PCR

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

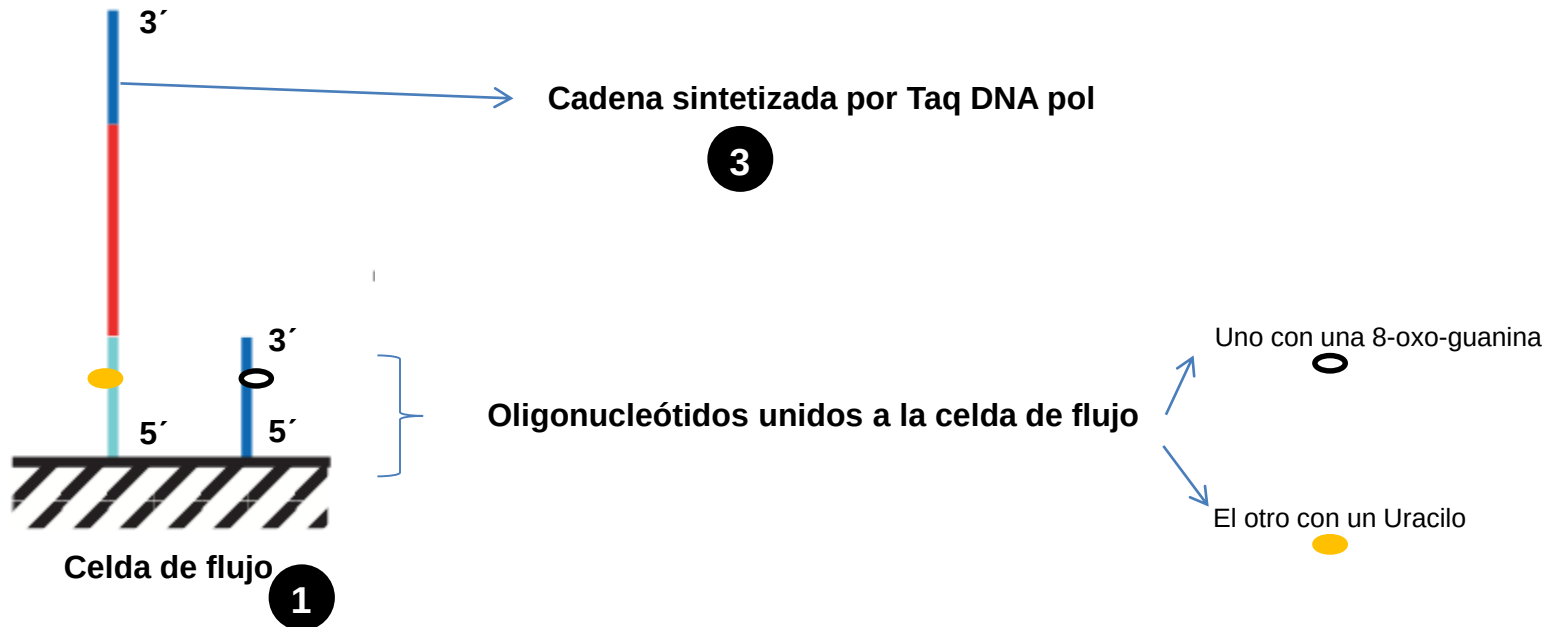
- Los **ssDNA** derivados de la muestra e hibridados a los oligonucleótidos complementarios presentes en la celda son **extendidos** mediante el uso de **TaqDNA pol.**



Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

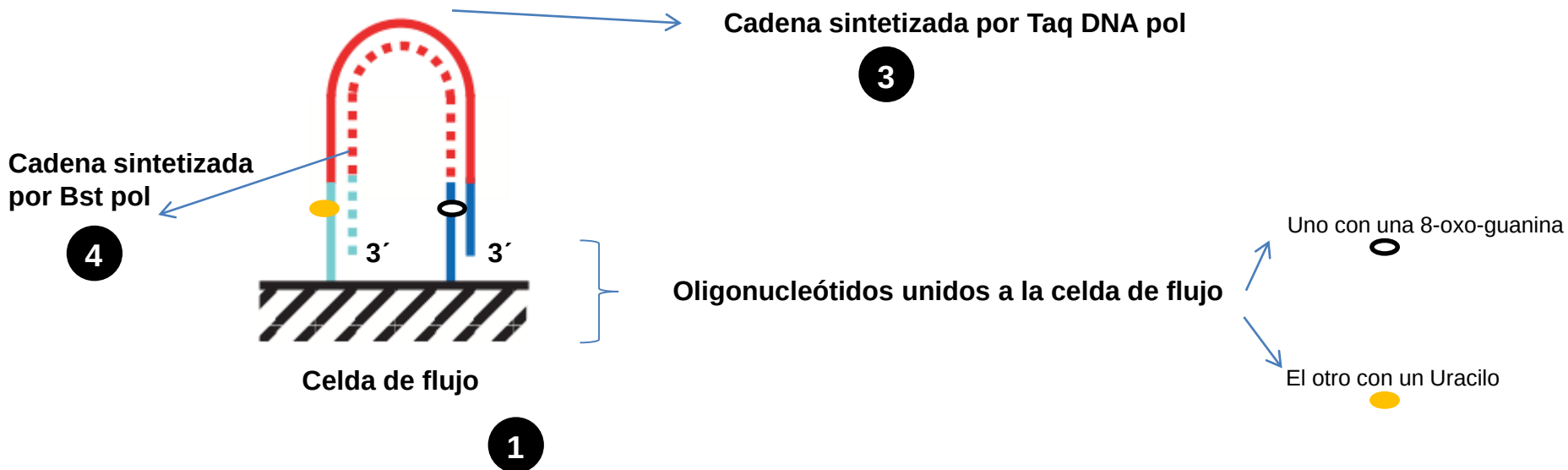
- **Luego** se hace un **tratamiento desnaturalizante** con Formamida para remover la cadena original (indicada previamente con el número 2).



Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

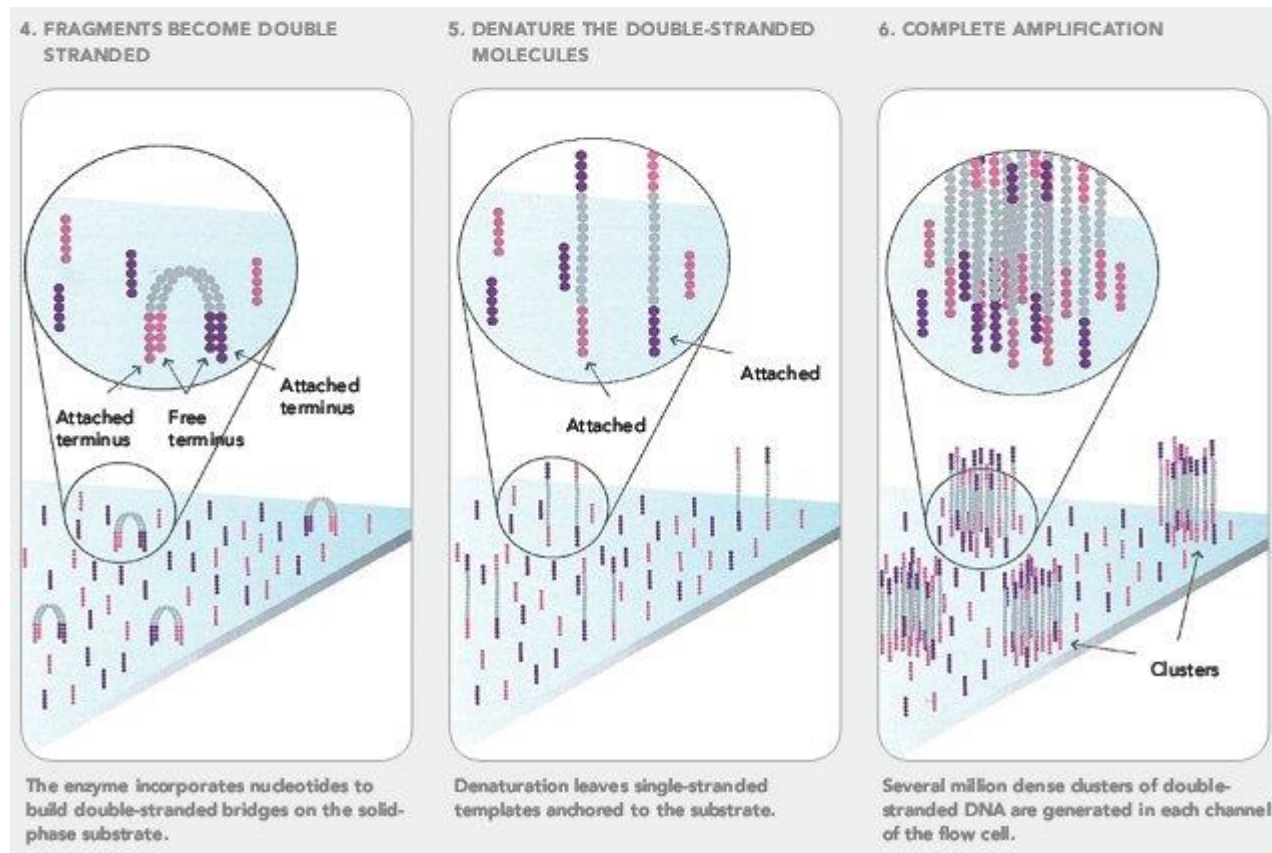
- Posteriormente se realiza una **amplificación isotérmica** a 60°C con *DNA polimerasa Bst* (35 ciclos). La desnaturalización entre-ciclos se realiza con formamida.



La *DNA pol Bst* deriva de *Bacillus stearothermophilus* y posee actividad *DNA* helicasa.

Tecnologías de secuenciación: 2G

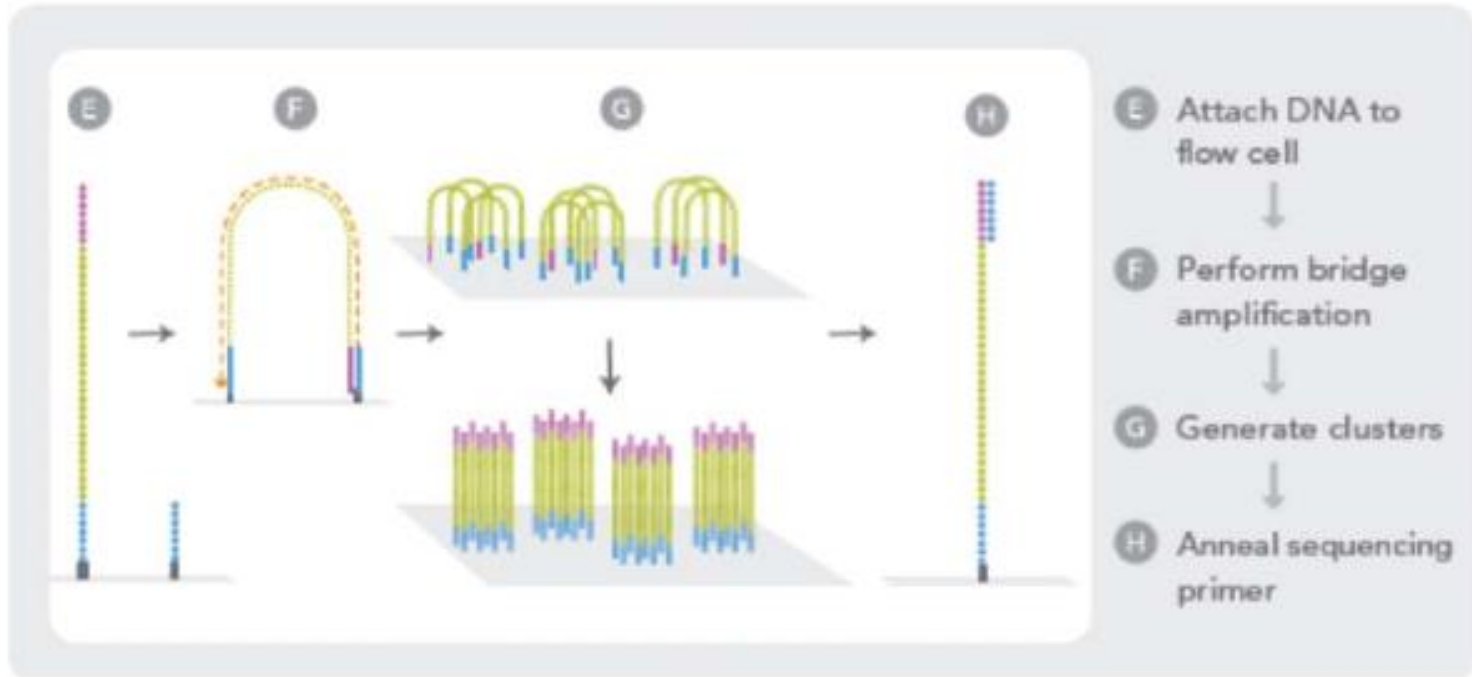
NGS: tecnología Illumina



Luego de la amplificación, se generan *clusters* de ssDNA, o *PCR colonies* denominadas ***polonies*** (polonias). Se generan hasta 40 millones de polonias.

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

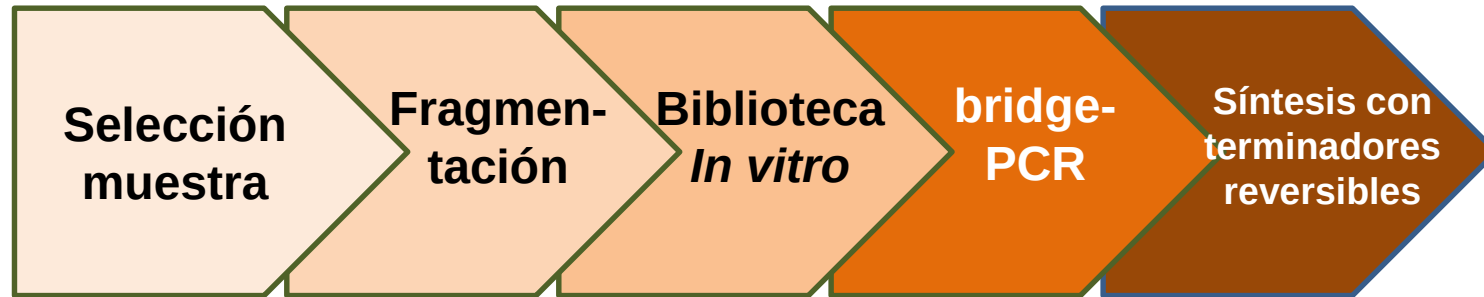


Se generan hasta **1000 copias** de cada fragmento en un diámetro de **1 micrón**.

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

Etapas involucradas en la tecnología 454



↓

A partir de un individuo.
A partir de múltiples individuos.

↓

Nebulización del DNA

↓

Separación de los fragmentos

↓

Amplificación de los fragmentos

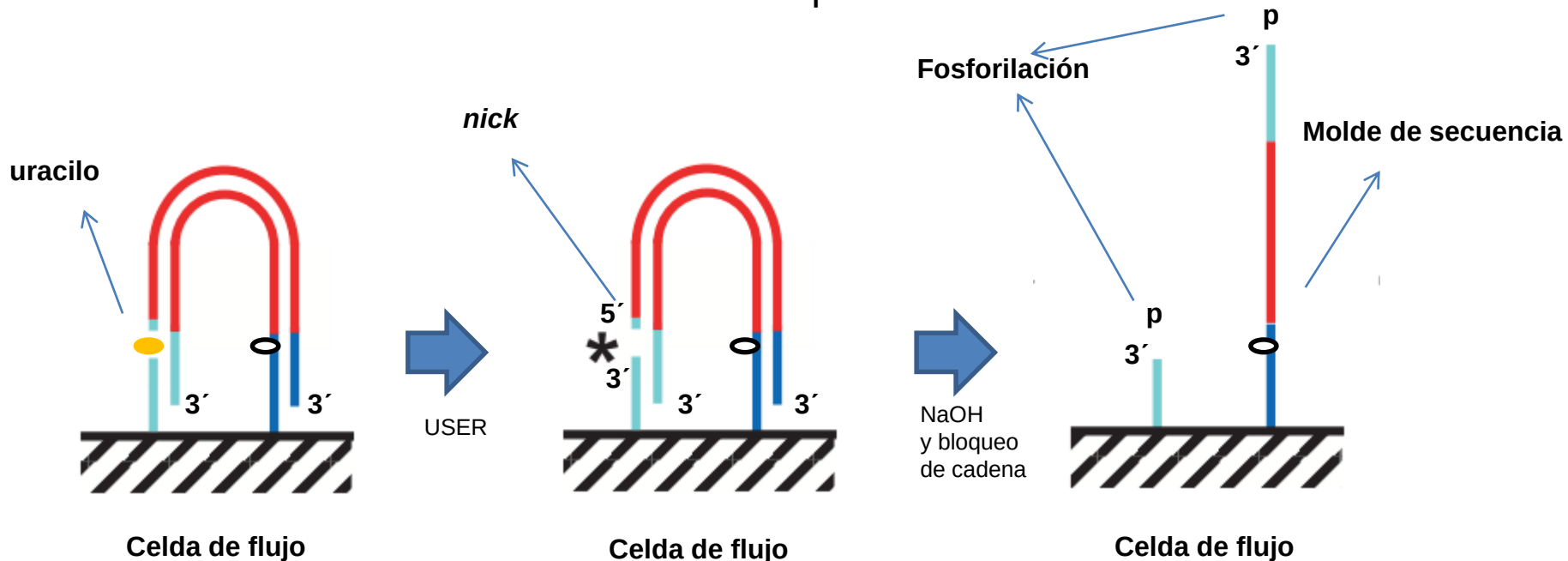
↓

En celda de flujo

Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

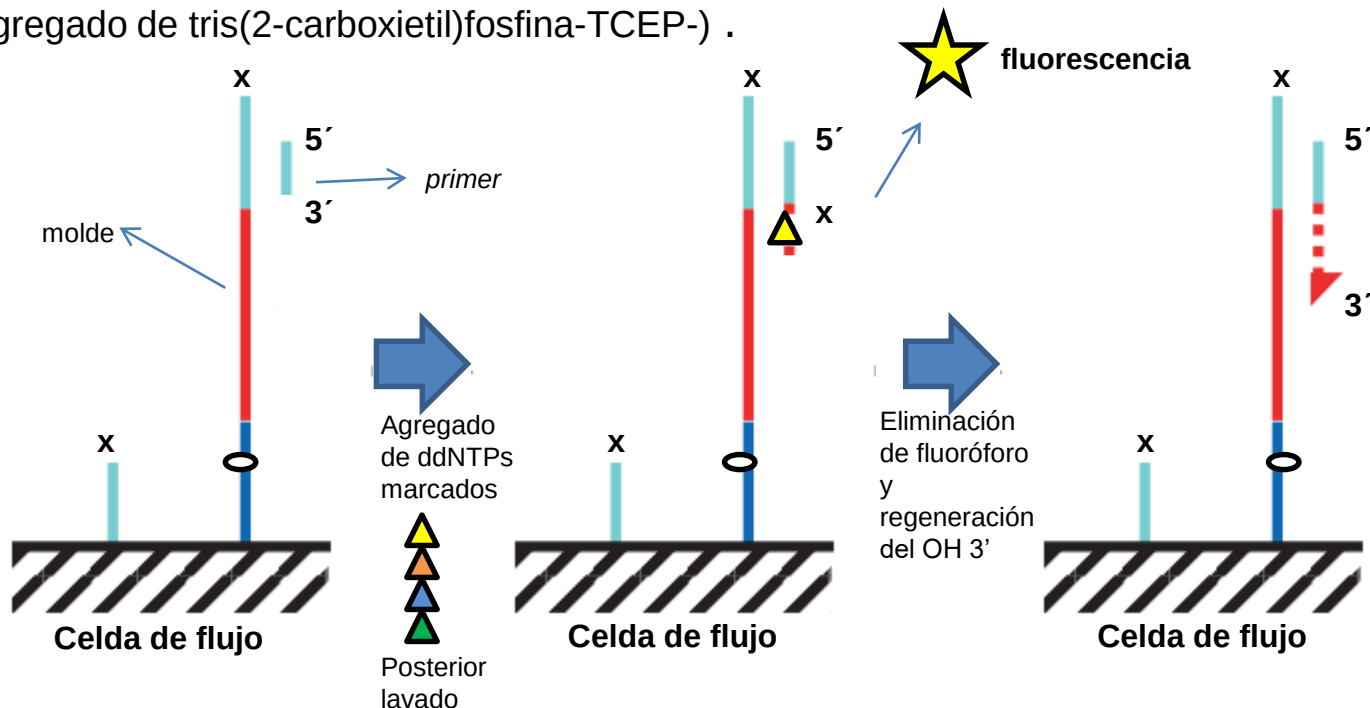
- Los **DNA** **amplificados** sobre la **celda de flujo** son **tratados** con la enzima **USER** (*Uracil-Specific Excision Reagent*; mezcla de Uracil DNA glicosilasa –UDG– y la DNA endonucleasa glicosilasa-liasa VIII), lo cual **provocará un *nick*** en uno de los oligonucleótidos unidos a la celda de flujo. **Luego** se desnaturaliza con **álcali** para dejar expuesta una sola cadena de DNA y se introduce un fosfato en el extremo 3' con una quinasa.



Tecnologías de secuenciación: 2G

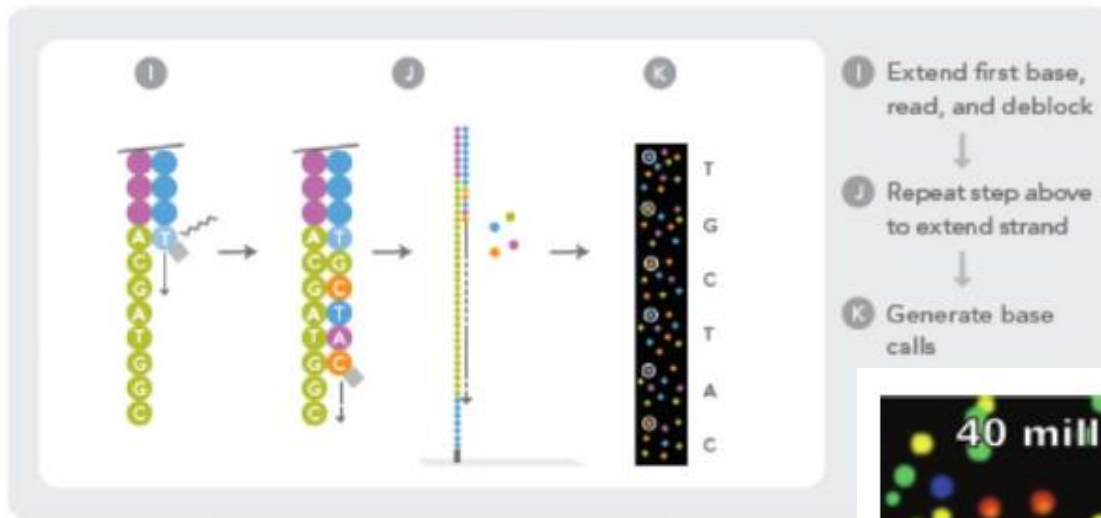
NGS: tecnología Illumina

- Los **ssDNA** luego son **secuenciados** mediante **síntesis**. Para ello se incorpora un **primer específico** para la **cadena en estudio**, *DNA polimerasa DNA dependiente*, y los **4 ddNTPs** marcados con fluoróforos (en formato reversible) agregados en conjunto. **Después de lavar los ddNTPs no incorporados**, se **excita** la celda y se **lee la emisión de fluorescencia**. **Posteriormente**, químicamente se **elimina el fluoróforo** y se **revierte la capacidad terminadora de síntesis** para repetir el ciclo (agregado de tris(2-carboxietil)fosfina-TCEP-).

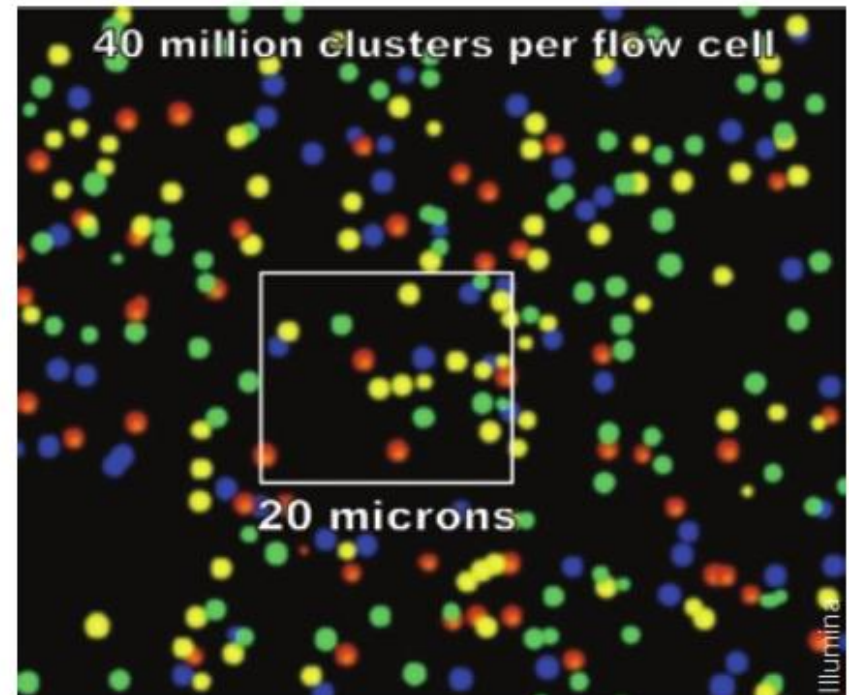


Tecnologías de secuenciación: 2G

NGS: tecnología Illumina



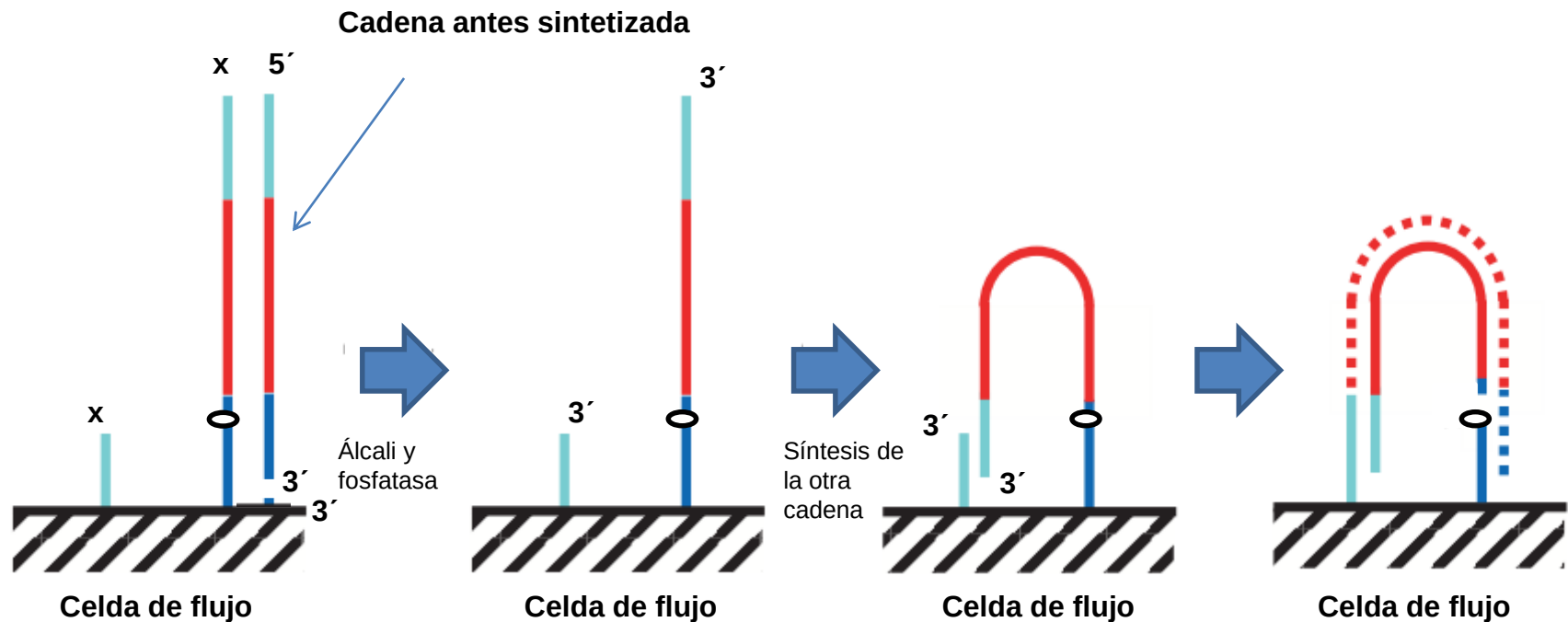
Visión ampliada de las **polonias** emitiendo **fluorescencia** en función del DNA presente en el *cluster*



Tecnologías de secuenciación: 2G

NGS: tecnología Illumina

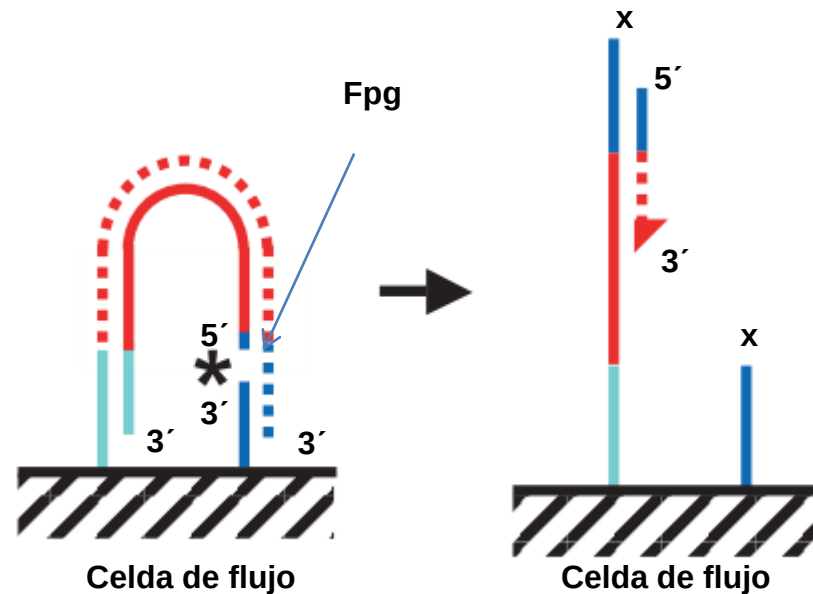
- Una vez **concluida** la **secuencia** de una de las **cadena**s, se **desnaturaliza** la muestra para **eliminar** la **cadena** recientemente **sintetizada**. Luego se **desfosforilan** los **extremos 3' libres** (con fosfatasa) y se **realiza** una reacción de síntesis tipo **bridge**.



Tecnologías de secuenciación: 2G

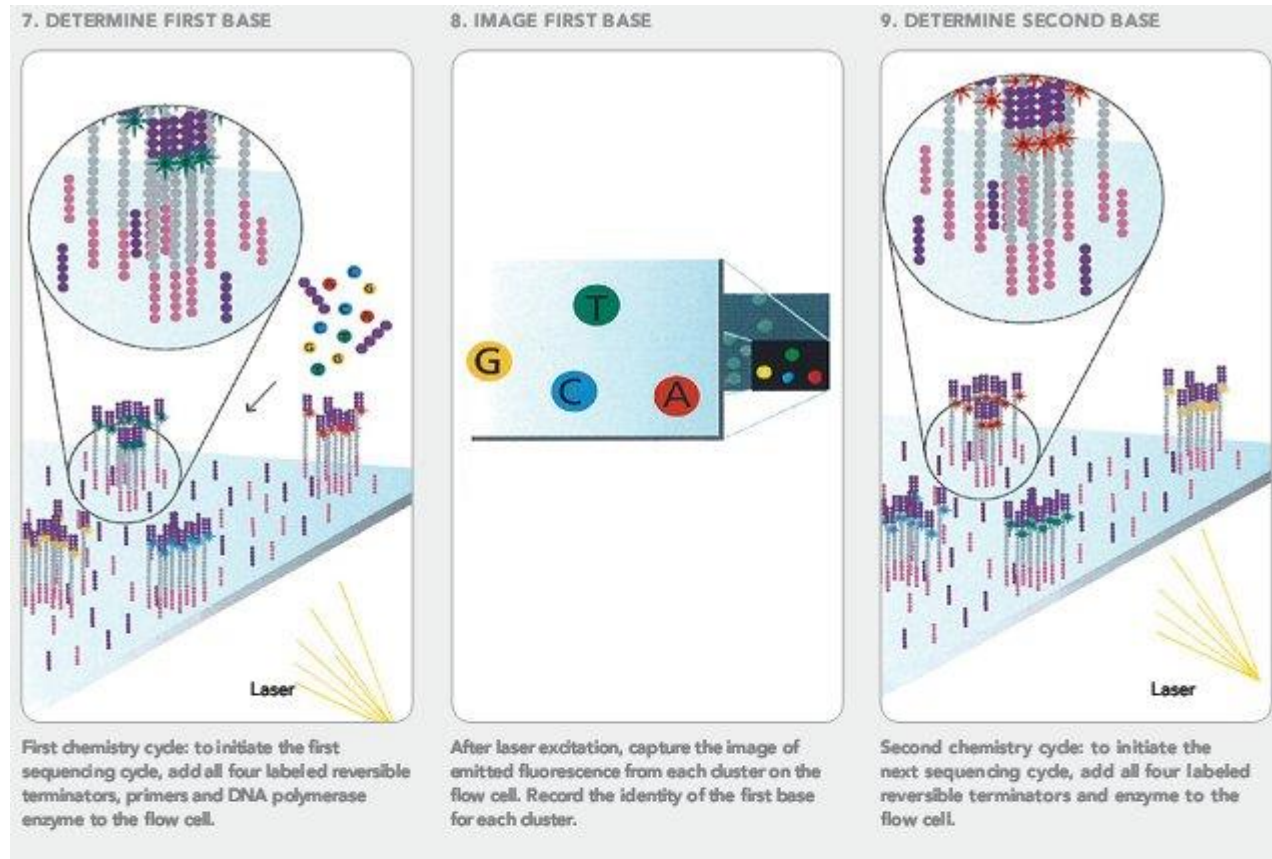
NGS: tecnología Illumina

- Finalmente, se elimina la cadena ya secuenciada mediante tratamiento con **Fpg** (*formamido-pirimidina DNA glicosidasa*) y posterior **desnaturalización**. La **cadena** que queda **unida a la celda** es **secuenciada** del mismo modo que la primera.



Tecnologías de secuenciación: 2G

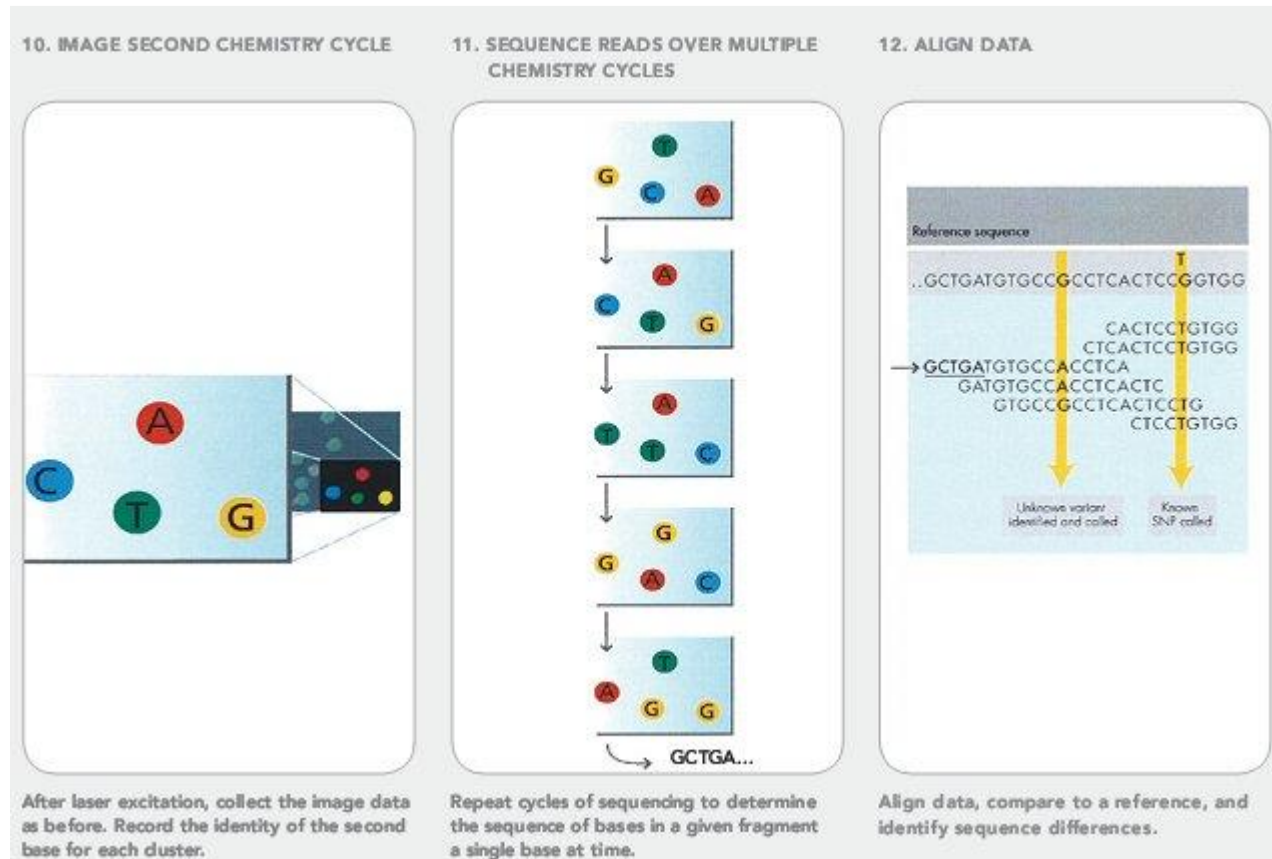
NGS: tecnología Illumina



La reacción de **secuenciación** se realiza por **síntesis e incorporación** de **terminadores de cadena** (cada uno marcado con un fluoróforo diferente). Los mismos son **reversibles**, y por ende se puede continuar con la síntesis.

Tecnologías de secuenciación: 2G

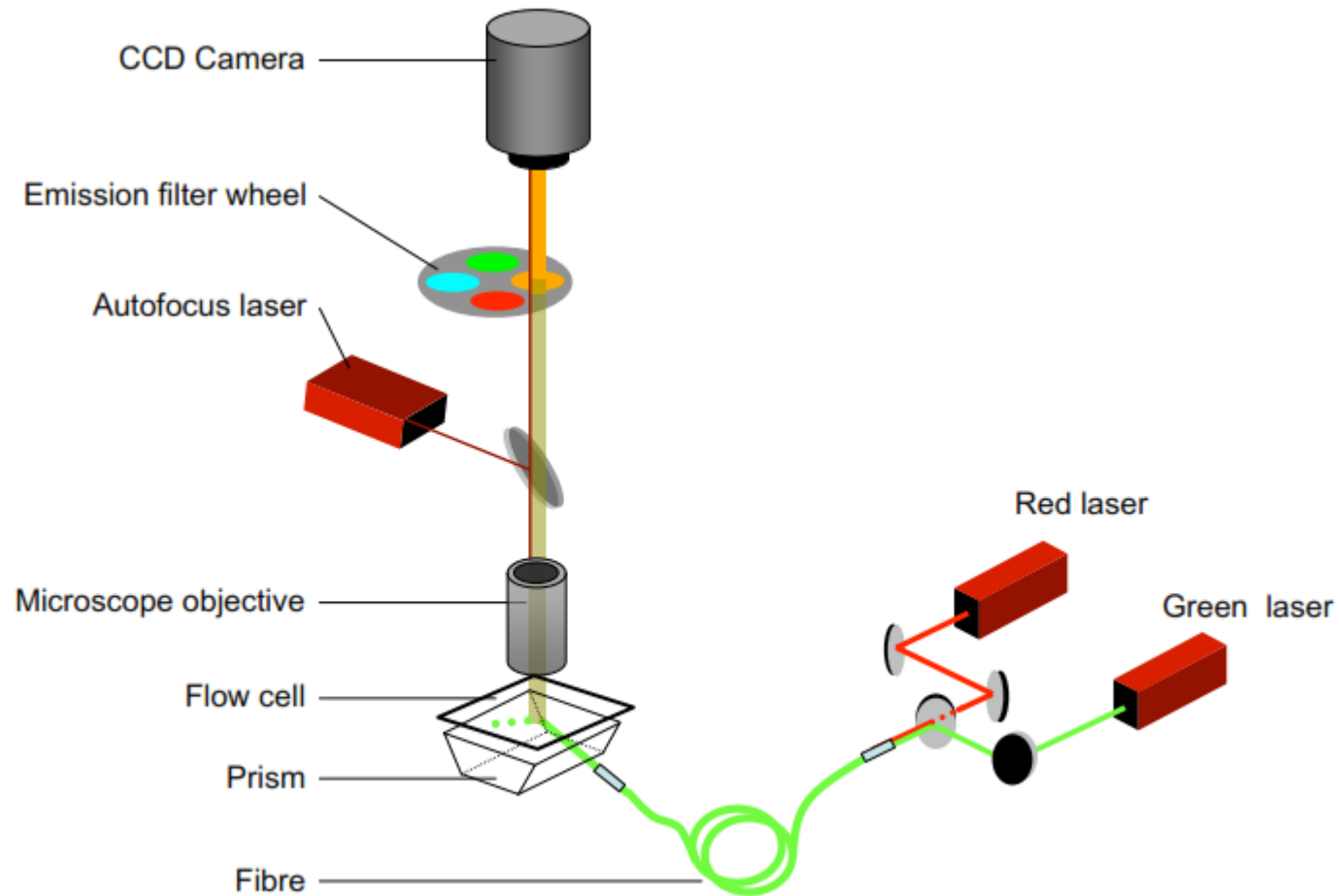
NGS: tecnología Illumina



Una **cámara** va registrando las **señales** derivadas de cada terminador.

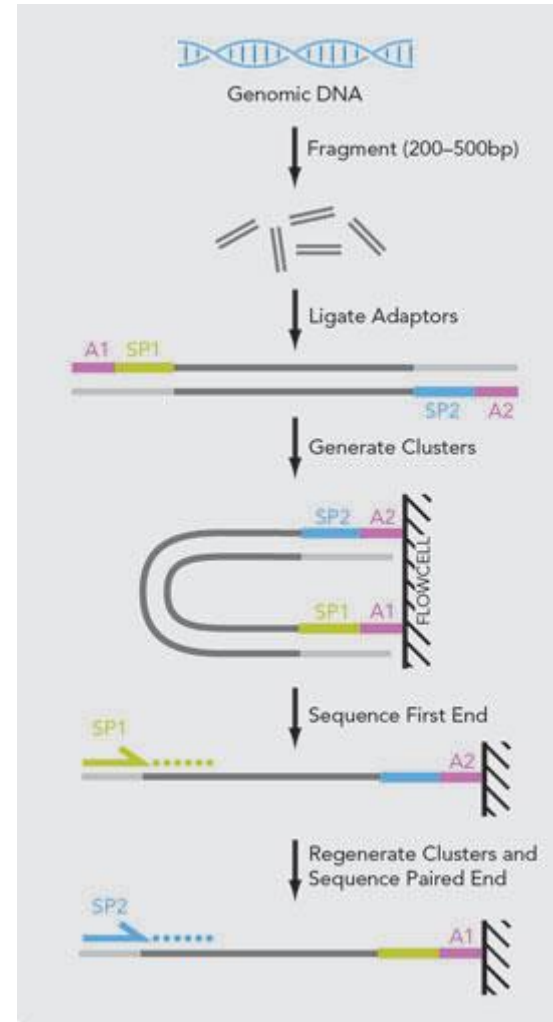
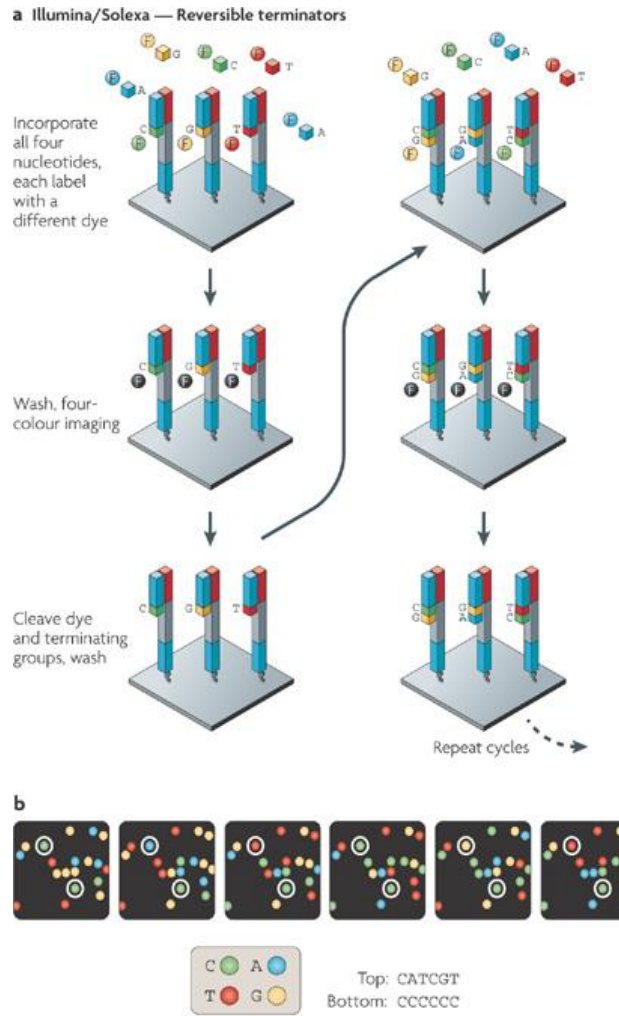
Tecnologías de secuenciación: 2G

NGS: tecnología Illumina



Tecnologías de secuenciación: 2G

NGS: tecnología Illumina



Resumen de la tecnología

Descripción de los métodos más utilizados clasificados dentro de las tecnologías NGS

Tecnología ABI SOLID

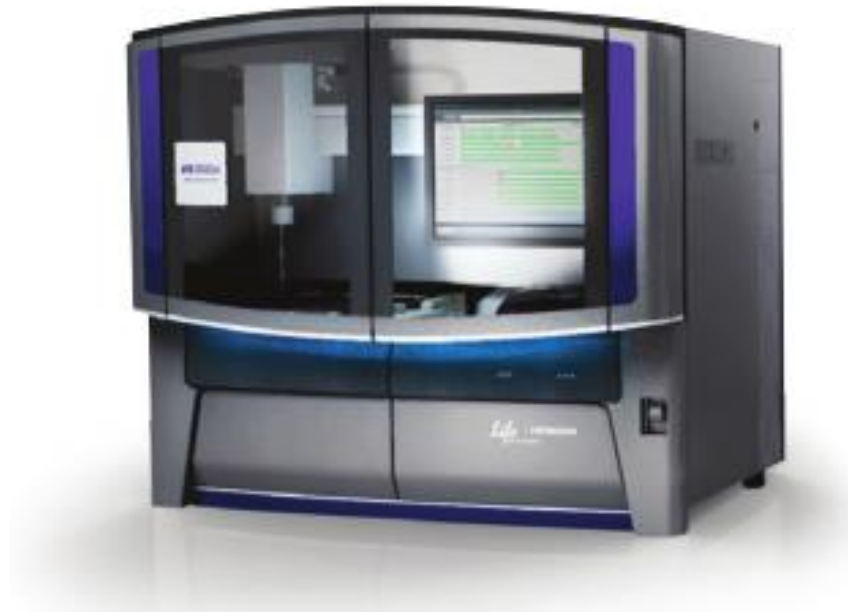
Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

- La **tecnología SOLID** (*Sequencing by Oligonucleotide Ligation and Detection*) disponible desde **2008** ha sido desarrollada por **Applied Biosystems**, la empresa que aportó la tecnología automatizada de Sanger para secuenciar el genoma de la especie humana durante la década de los 90 del siglo XX.
- Desde el año 2008 **Applied Biosystems** se fusionó con **Invitrogen** creando la empresa biotecnológica denominada **Life Technologies**.
- El sistema de **secuenciamiento** de **SOLID** también es conocido como **2 base encoding** y está basado en **reacciones de ligación**.
- Este método **no involucra síntesis** como la pirosecuenciación (tecnología 454) o la química reversible de terminadores fluorescentes (tecnología *Illumina*).

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

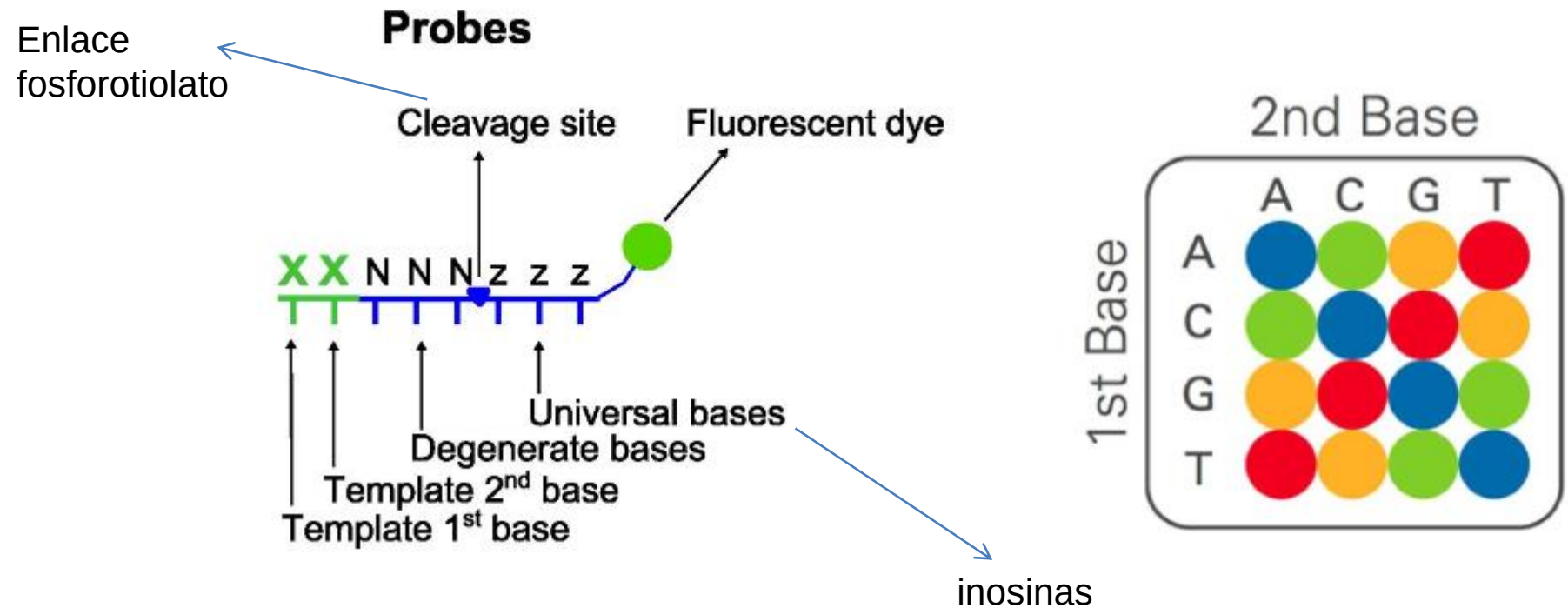


Equipamiento SOLID

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

El método de secuenciación involucra el uso de sondas

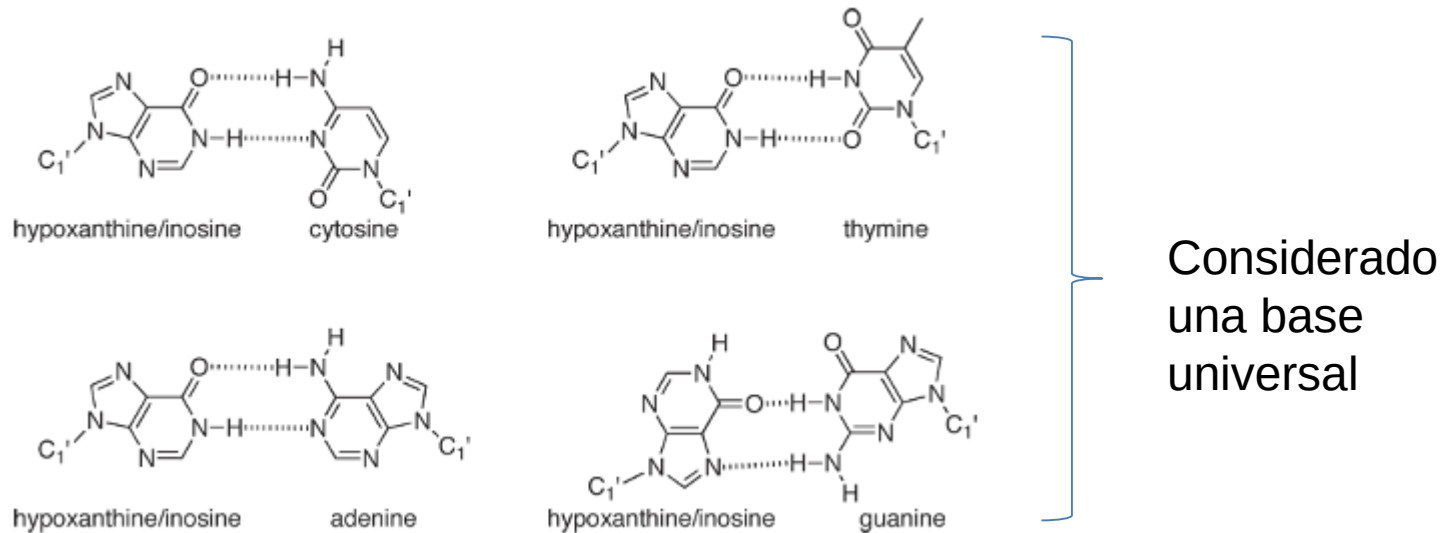


Son **256 sondas** por cada color.

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

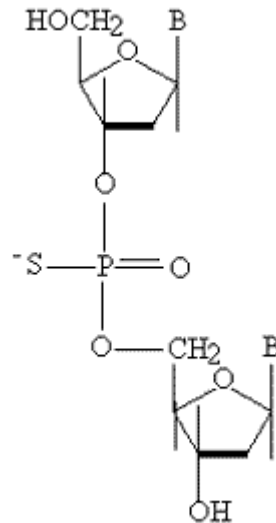
- Apareamiento de **inosina** con el resto de las bases nitrogenadas



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

- Los **enlaces fosforotiolatos** dan resistencia a los oligonucleótidos de la acción de exonucleasas, y también permiten su **hidrólisis selectiva** mediante tratamiento con **iones de plata**.



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

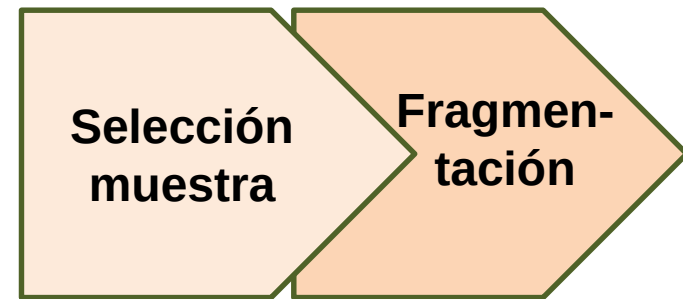
Etapas involucradas en la tecnología SOLID (a)



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

Etapas involucradas en la tecnología SOLID



A partir de un individuo.
A partir de múltiples individuos.

Nebulización del DNA

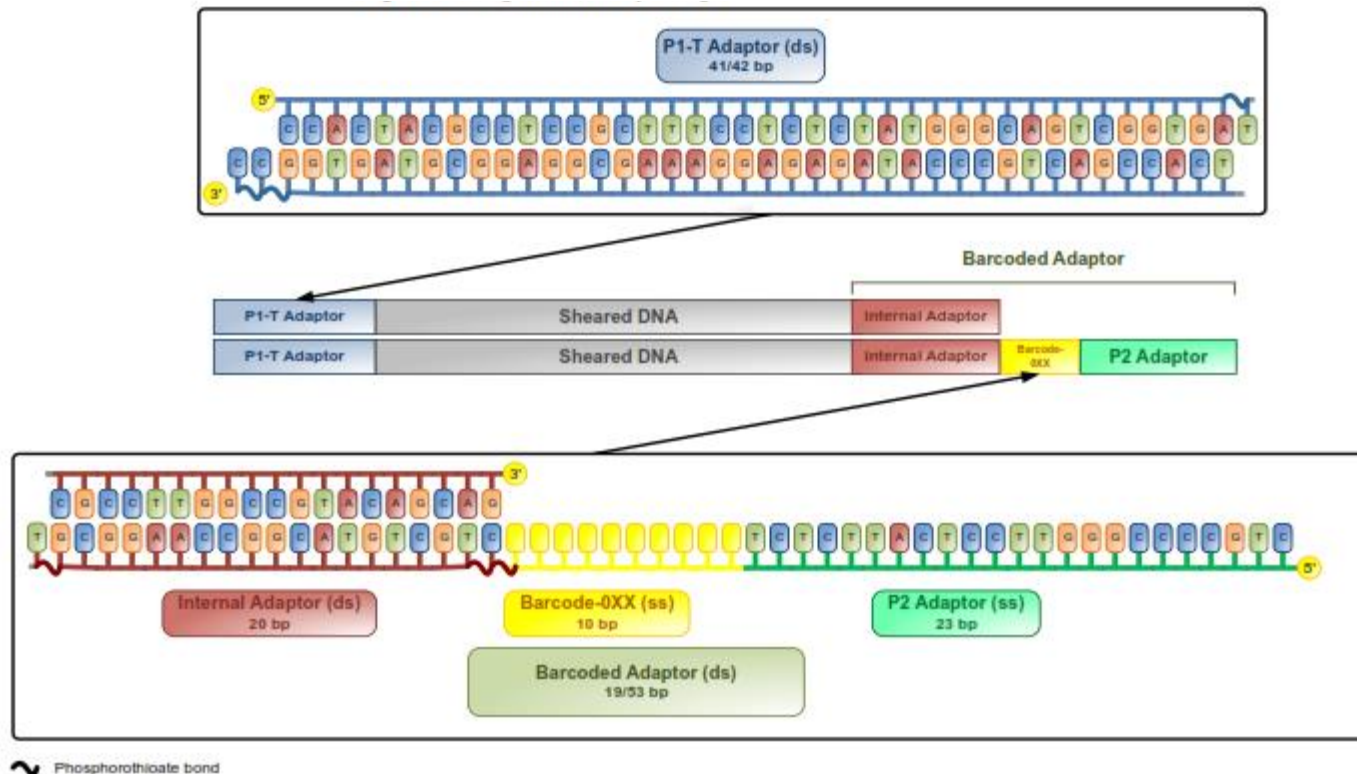
Se puede utilizar como mínimo **10 ng** de **gDNA**. Los fragmentos nebulizados y reparados en sus extremos deben tener entre **100-250 pb**.

A los mismos se les **liga adaptadores** de secuencia conocida, comunmente llamados **P1 y P2**.

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

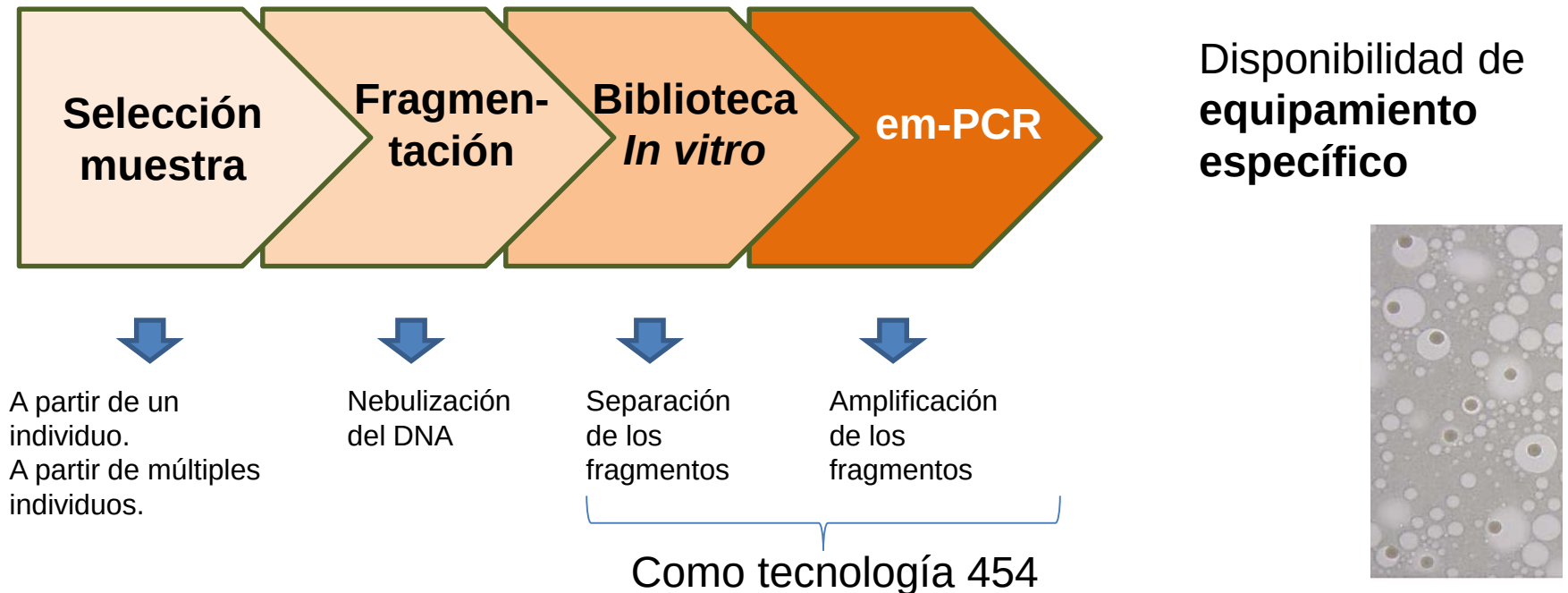
Ejemplos de adaptadores que pueden ligarse a los fragmentos del gDNA.



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

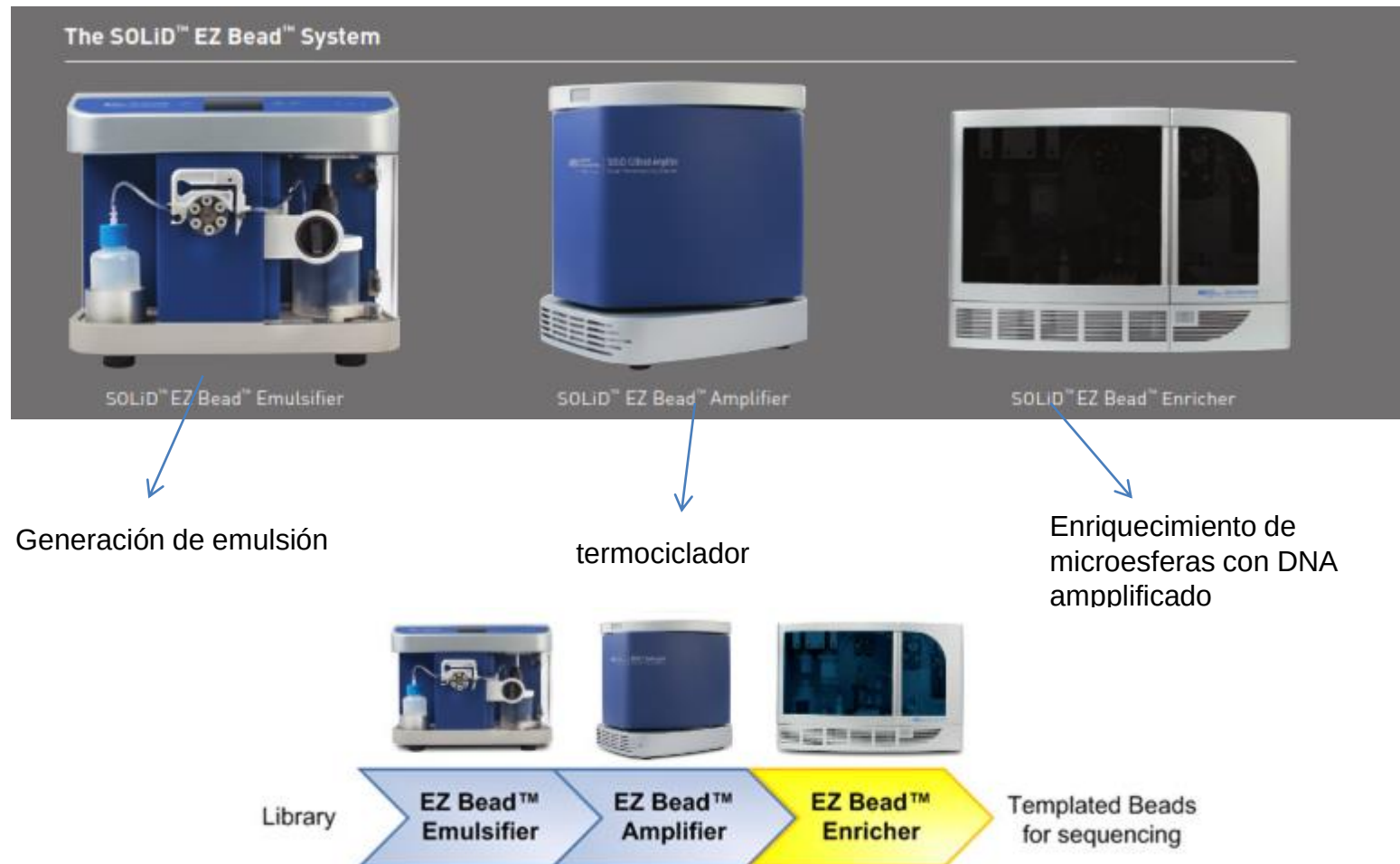
Etapas involucradas en la tecnología SOLID (a)



Tecnologías de secuenciación: 2G

NGS: tecnología SOLiD

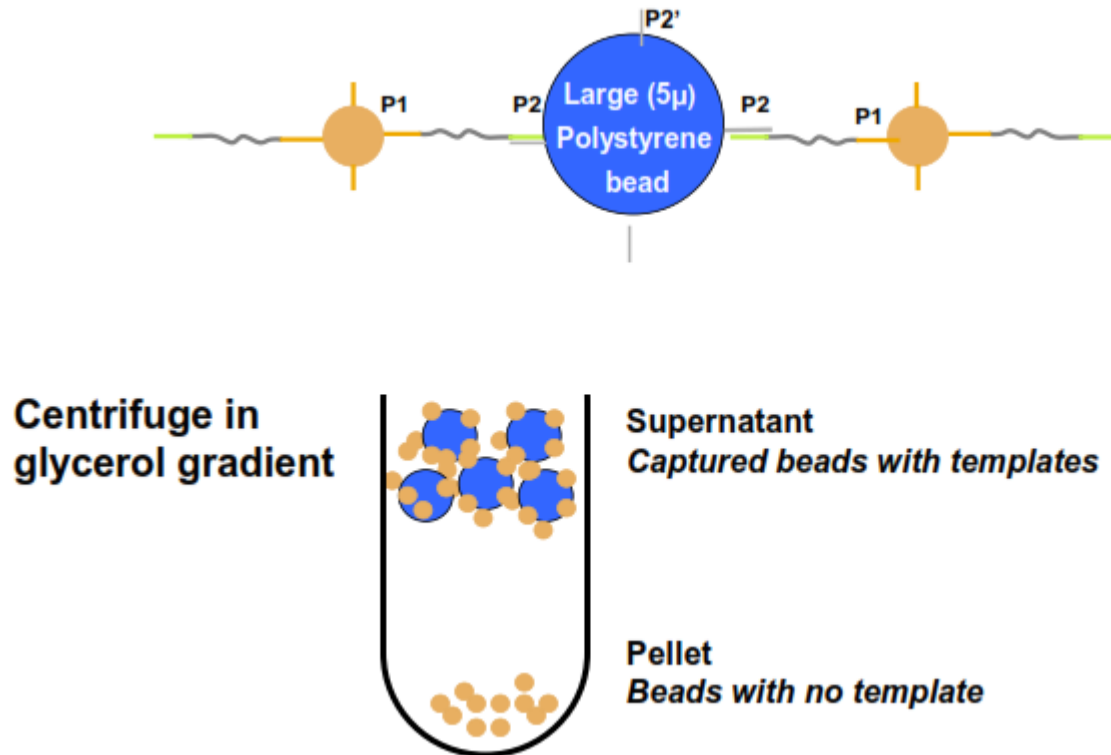
Equipamiento disponible para la construcción de la biblioteca *in vitro* utilizando microesferas



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

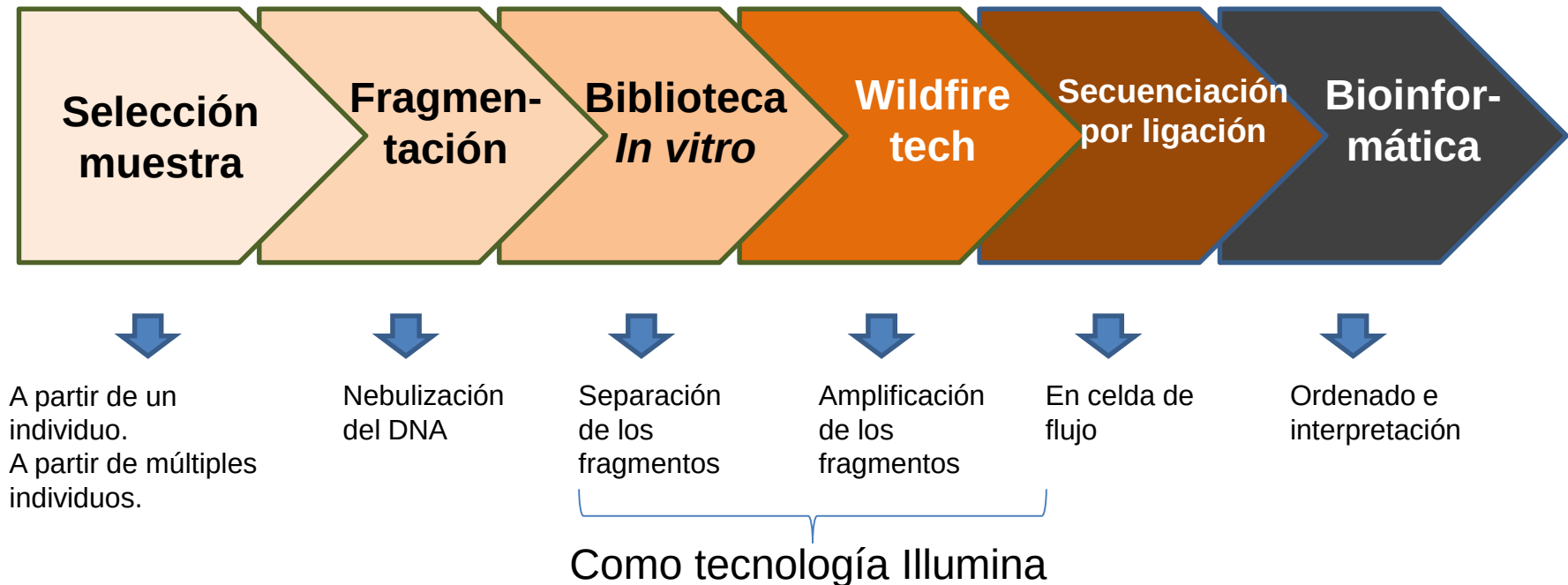
- El **enriquecimiento** de las **microesferas con amplificación** se realiza mediante su **captura** con **microesferas** más grandes de **poliestireno** y con **secuencia complementaria al otro adaptador**.



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

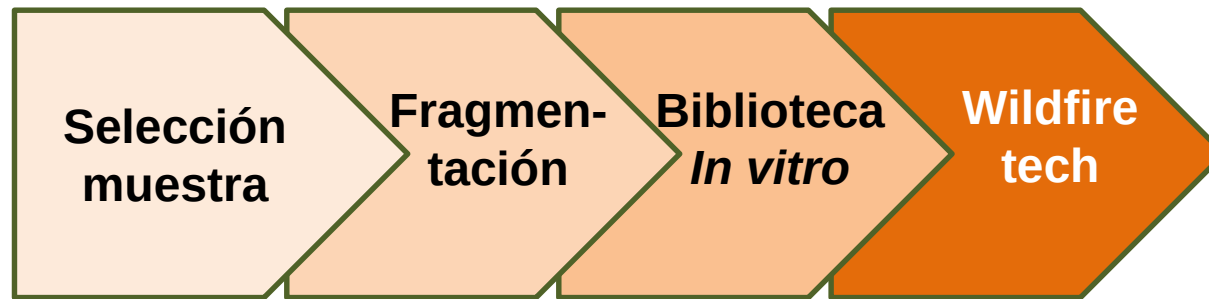
Etapas involucradas en la tecnología SOLID (b)



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

Etapas involucradas en la tecnología SOLID (b)



Disponibilidad de la **tecnología *Wildfire*** en la generación de colonias.

↓
A partir de un individuo.
A partir de múltiples individuos.

↓
Nebulización del DNA

↓
Separación de los fragmentos

↓
Amplificación de los fragmentos

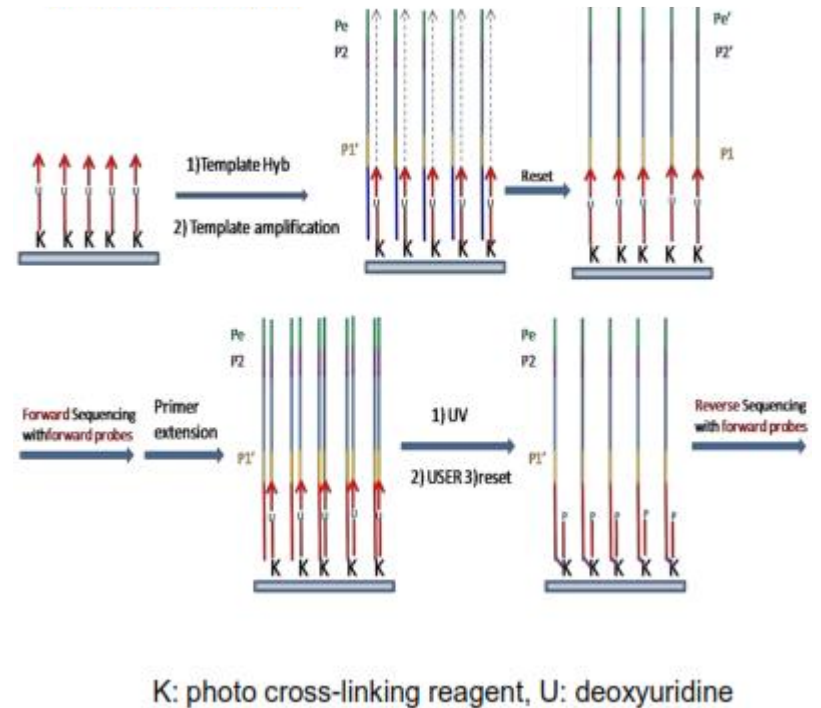
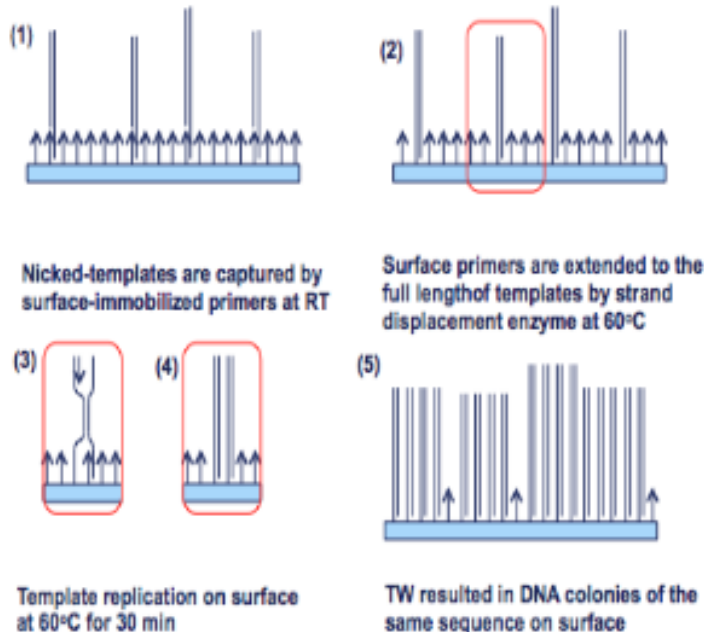
Como tecnología Illumina

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

Para la generación de las **colonias** la empresa desarrolló la **tecnología Wildfire**

Isothermal Template Walking (TW)

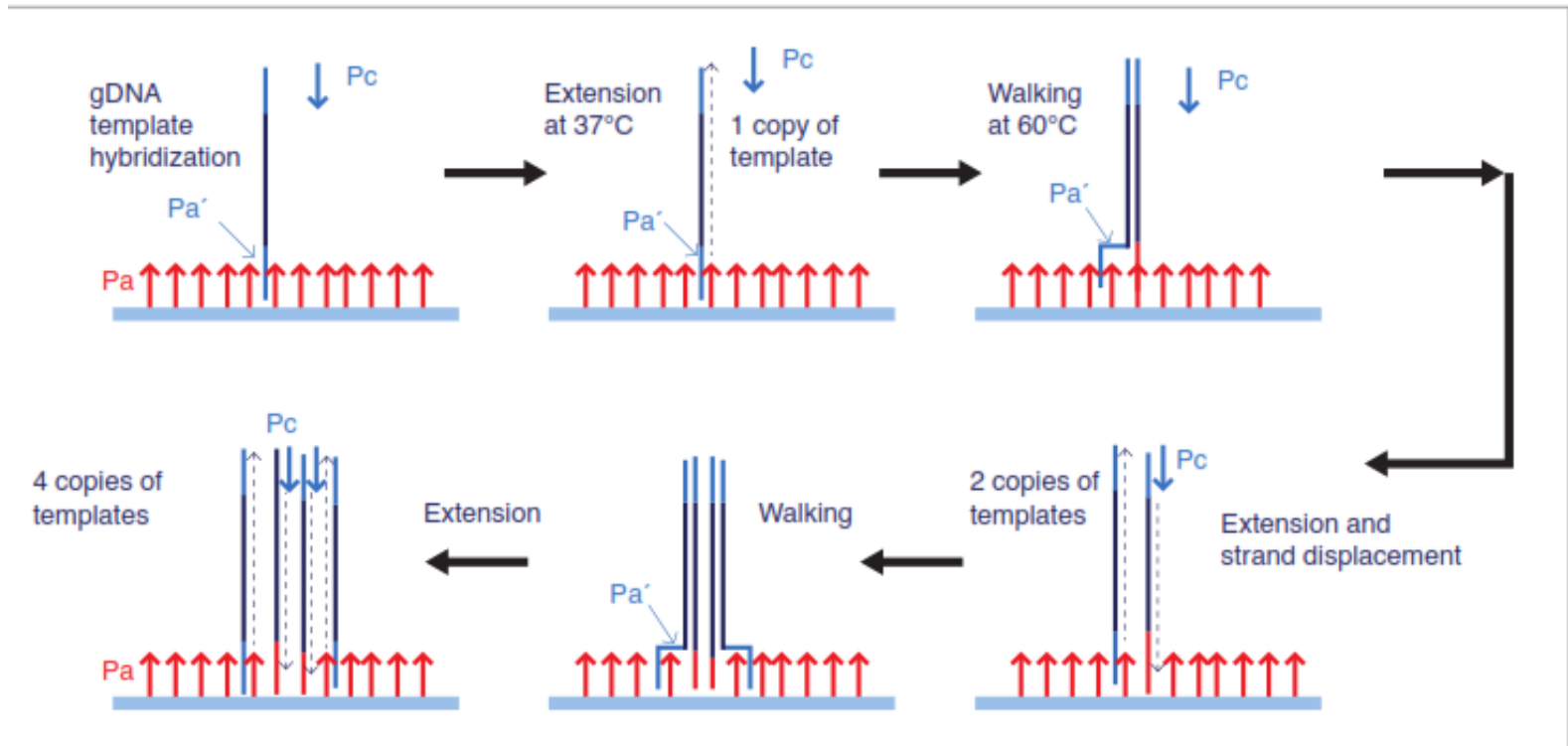


Se generan **1,5 millones de colonias** por **mm²** de superficie

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

Tecnología *wildfire*

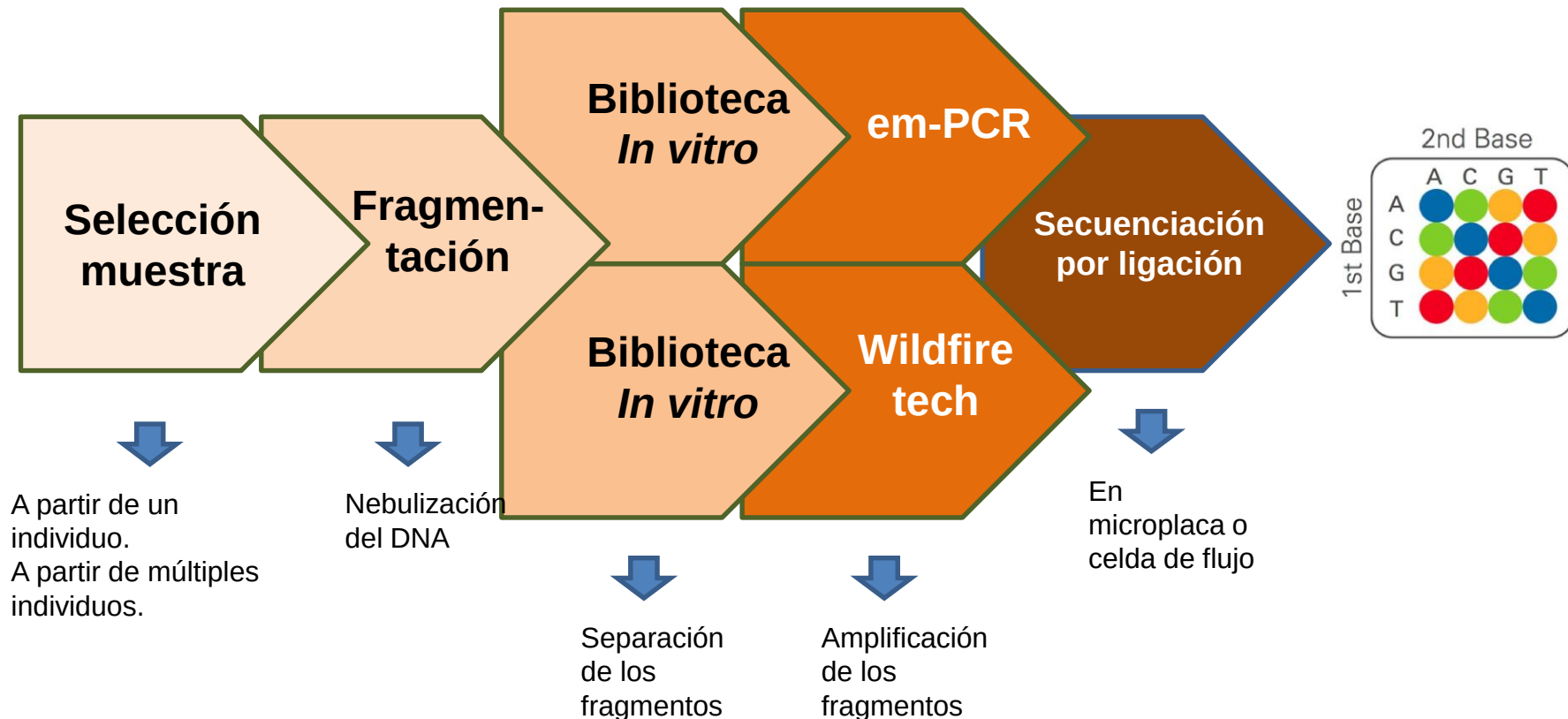


El procedimiento de amplificación dura aproximadamente 30 minutos. El crecimiento de la colonia se detiene cuando se topa con otra.

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

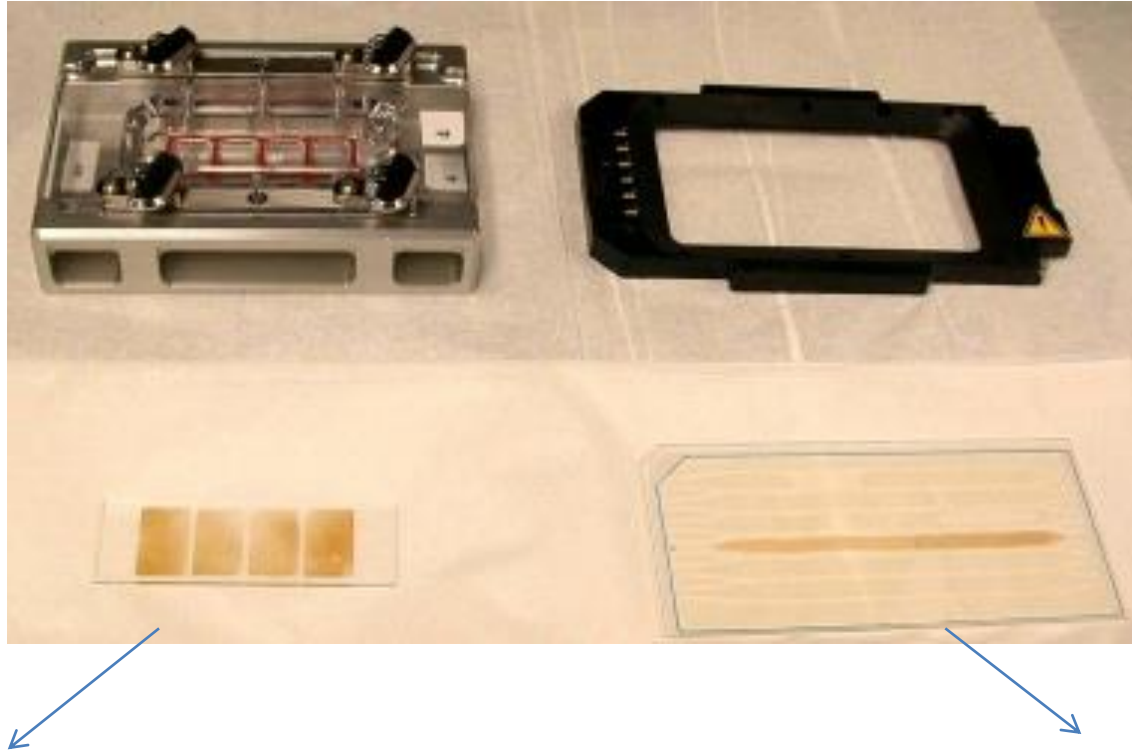
Etapas involucradas en la tecnología SOLID



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

- Si se usa **em-PCR** en microesferas...



Portaobjetos de vidrio utilizados para depositar las microesferas (aproximadamente 600 millones)

Celda de flujo utilizada para depositar las microesferas (aproximadamente 1080 millones)

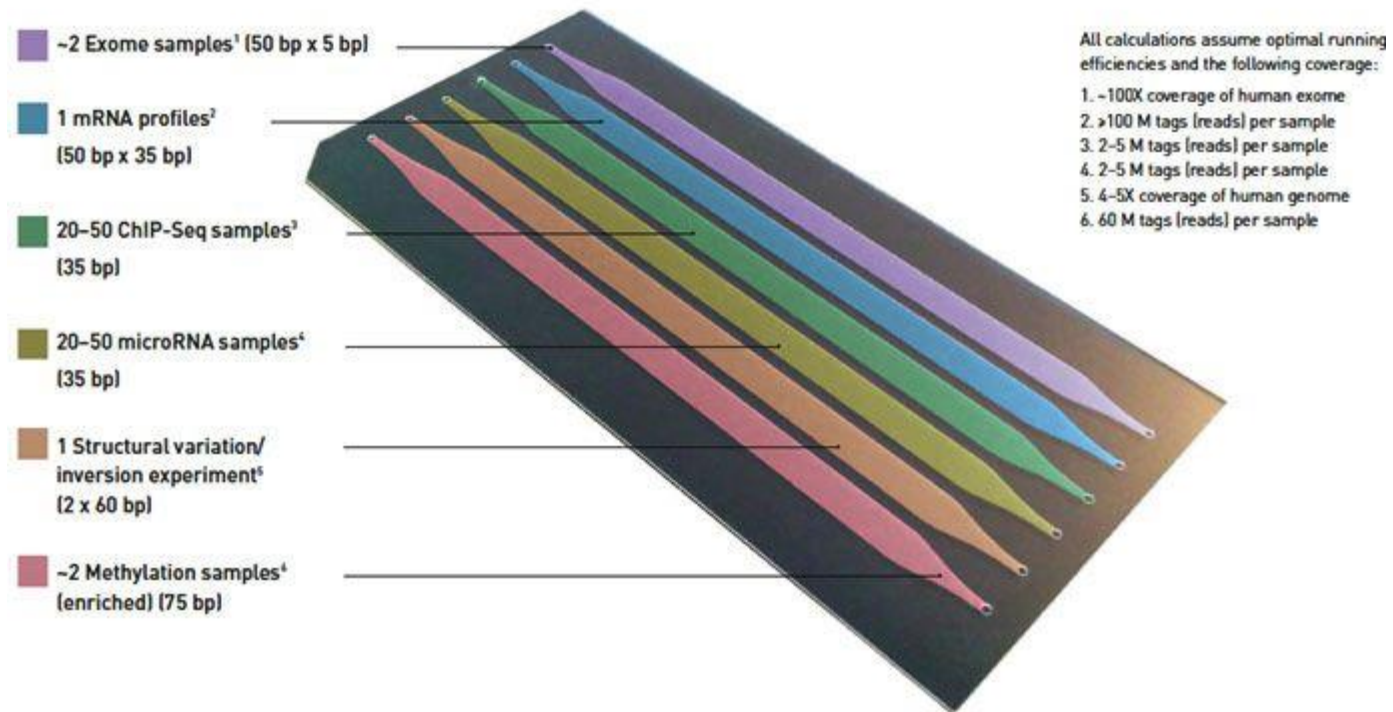
La disposición de las microesferas será al azar, y quedarán retenidas por una unión covalente entre el extremo 3' de algunos ssDNA y la superficie del vidrio

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

- Si se usa **generación de colonias**...

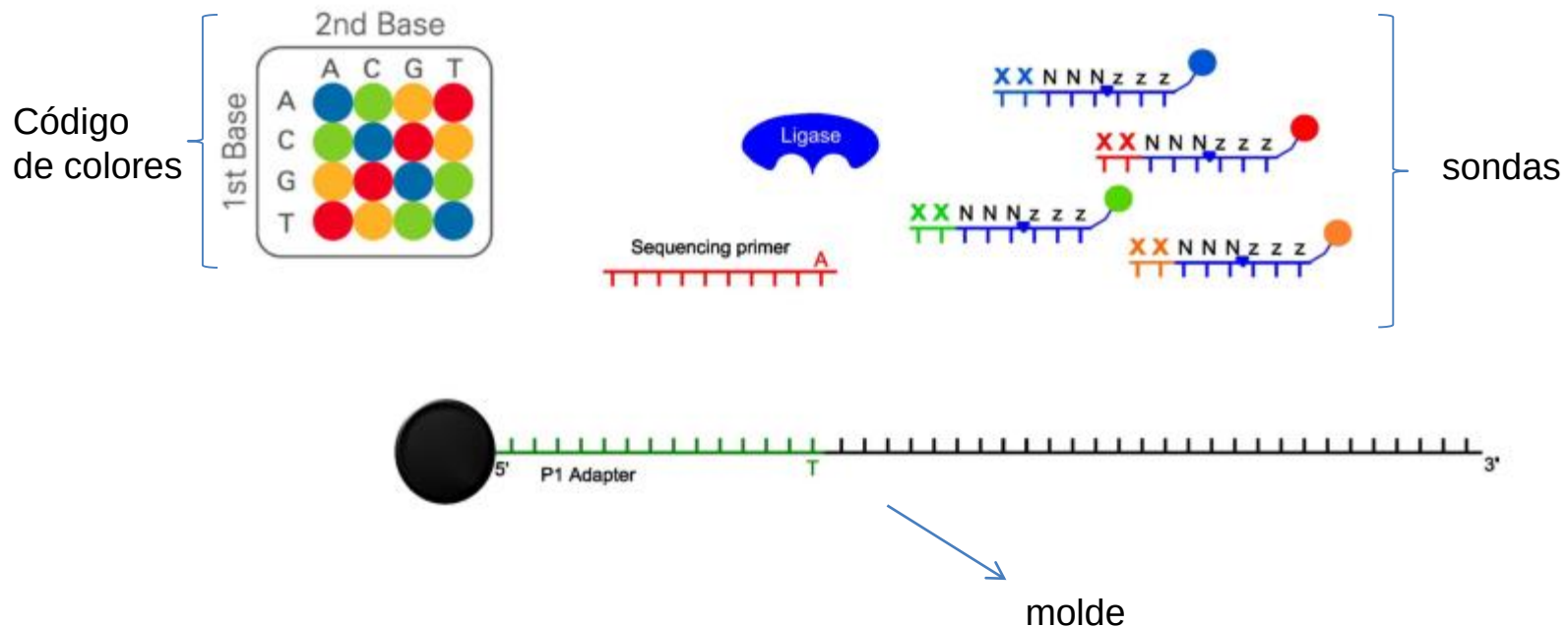
Multiple applications on a single FlowChip with different read lengths and chemistries.



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

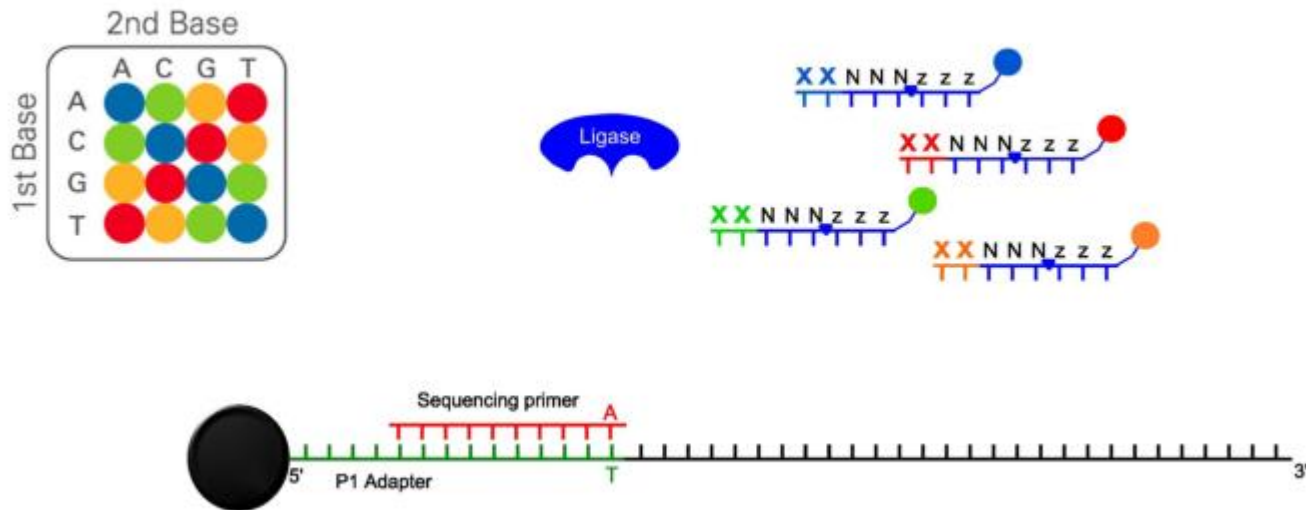
- **Primera Ronda de secuenciación.** Se introduce un **primer 1** específico que hibrida sobre la secuencia adaptadora (incorporada luego de nebulizar las muestras de gDNA), **DNA ligasa** y las **sondas fluorescentes** con diferentes dinucleótidos.



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

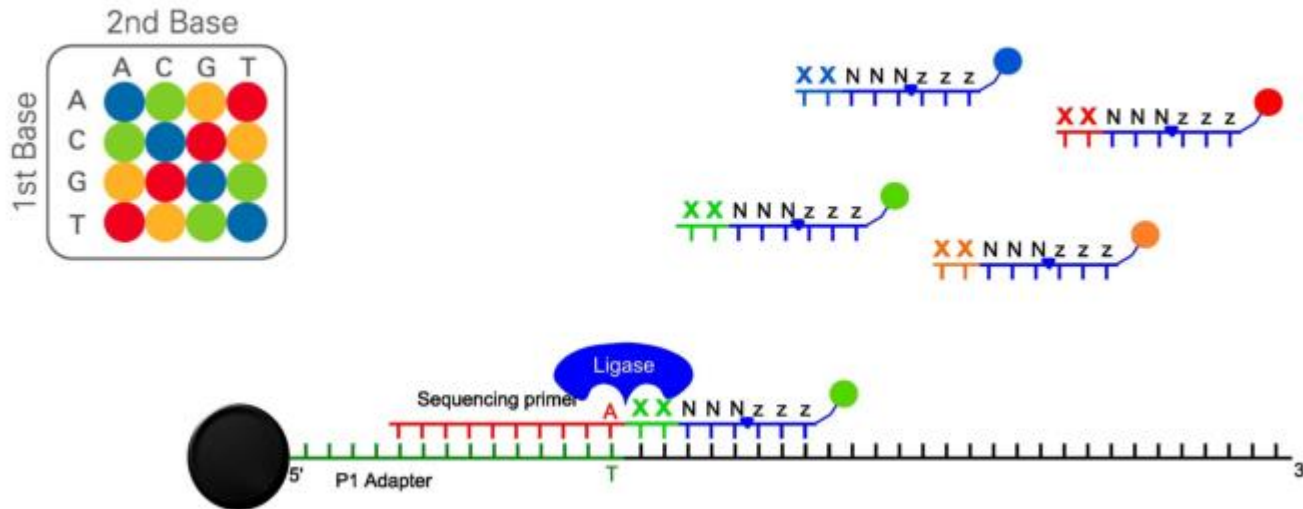
- **Primera Ronda de secuenciación.** El *primer* hibrida ...



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

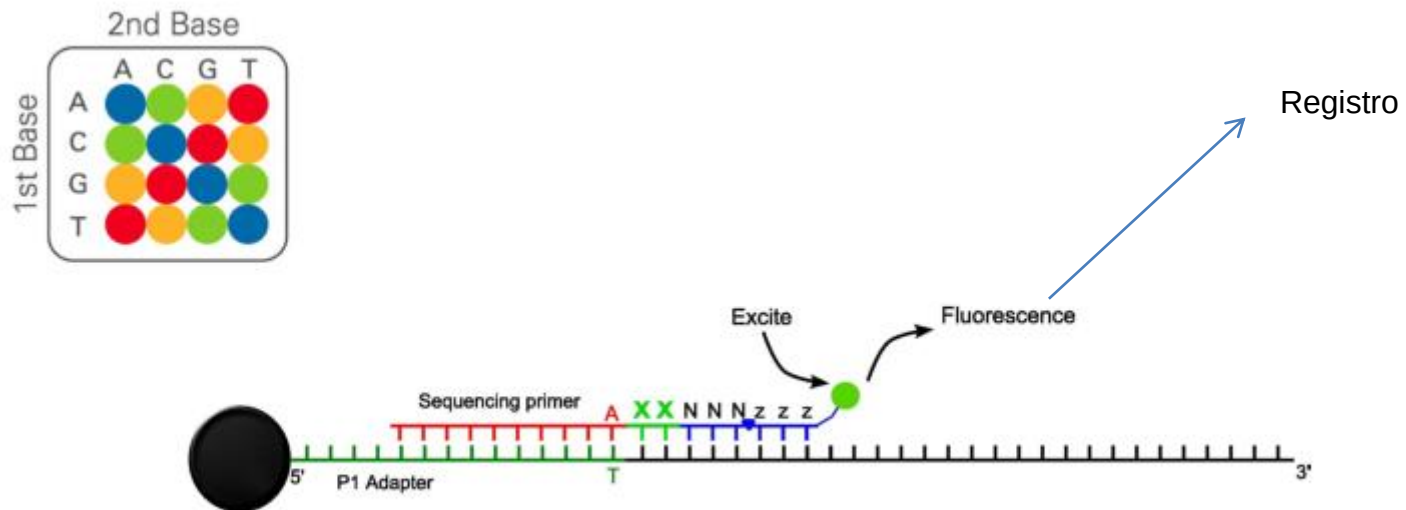
- **Primera Ronda de secuenciación.** El *primer* hibrida y la sonda que tiene los dos nucleótidos complementarios **también** lo hace. **Luego**, la **DNA ligasa** cataliza la formación del enlace fosfodiéster.



Tecnologías de secuenciación: 2G

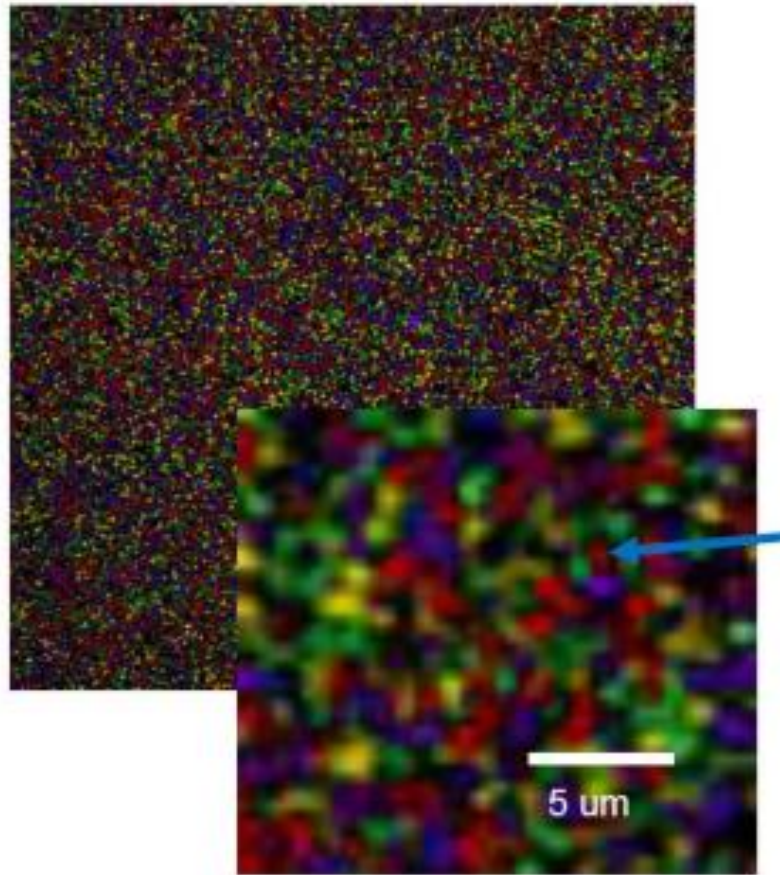
NGS: tecnología SOLID

- **Primera Ronda de secuenciación.** El *primer* hibrida y la sonda que tiene los dos nucleótidos complementarios **también** lo hace. **Luego**, la **DNA ligasa** cataliza la formación del enlace fosfodiéster. Se **lava** y se **excita** con luz para así detectar emisión de **fluorescencia**.



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

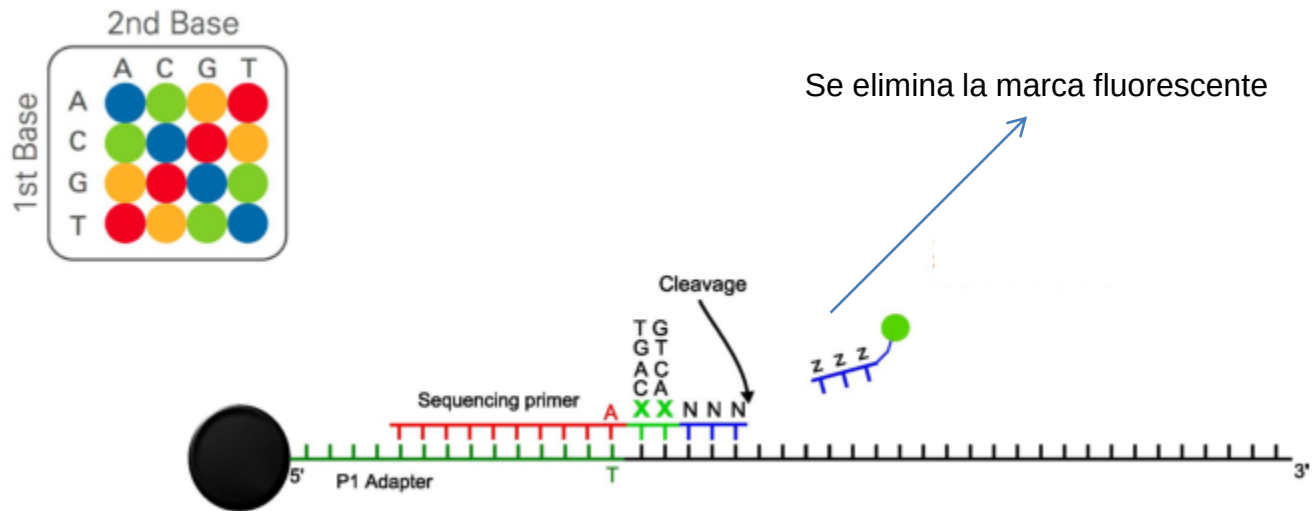


Registro de
fluorescencia de
cada una de las
colonias o
microesferas

Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

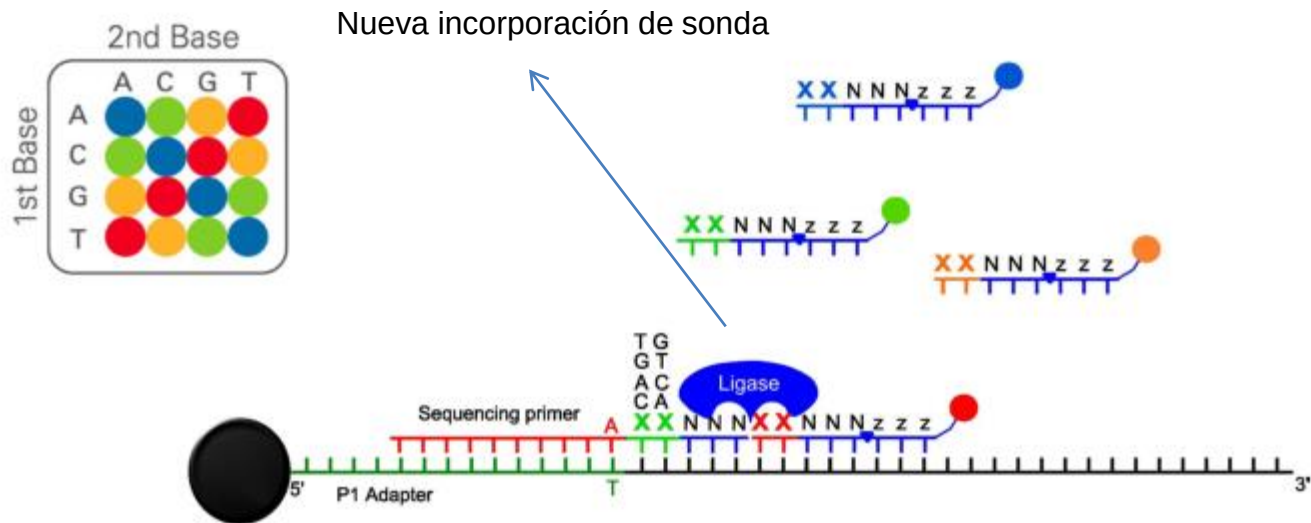
- **Primera Ronda de secuenciación.** El **primer** hibrida y la **sonda** que tiene los dos nucleótidos complementarios **también** lo hace. **Luego**, la **DNA ligasa** cataliza la formación del enlace fosfodiéster. Se **lava** y se **excita** con luz para así detectar emisión de **fluorescencia**. Después, se realiza el **clivaje** de la **sonda** hibridada y un lavado.



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

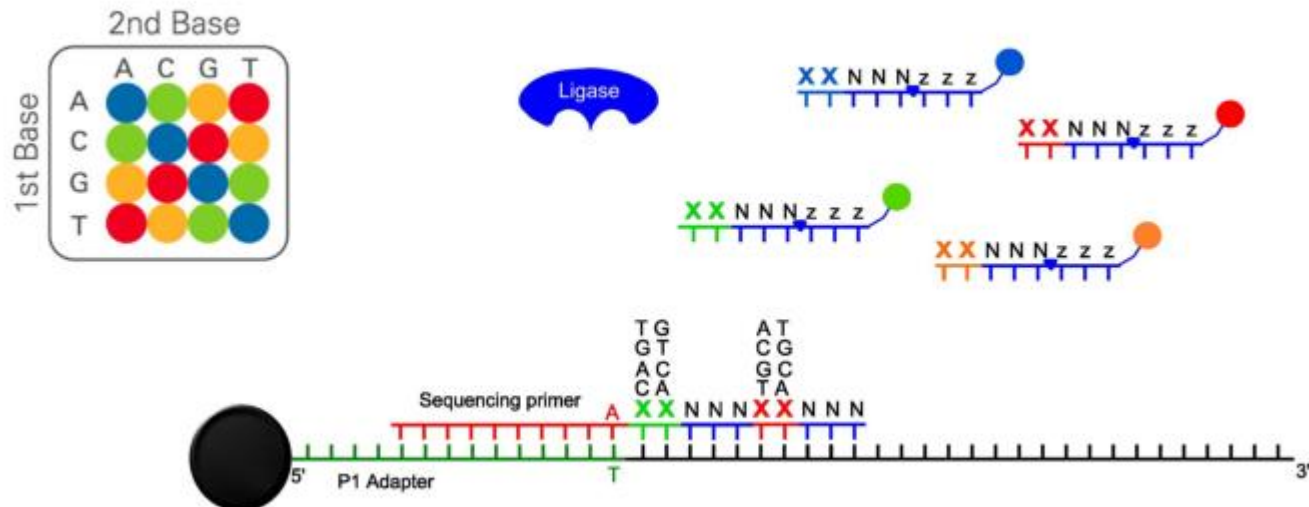
- **Primera Ronda de secuenciación.** El **primer** hibrida y la **sonda** que tiene los dos nucleótidos complementarios **también** lo hace. **Luego**, la **DNA ligasa** cataliza la formación del enlace fosfodiéster. Se **lava** y se **excita** con **luz** para así detectar emisión de **fluorescencia**. Después, se realiza el **clivaje** de la **sonda** hibridada y un lavado. Se **repite** el procedimiento agregando nuevamente las **sondas**.



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

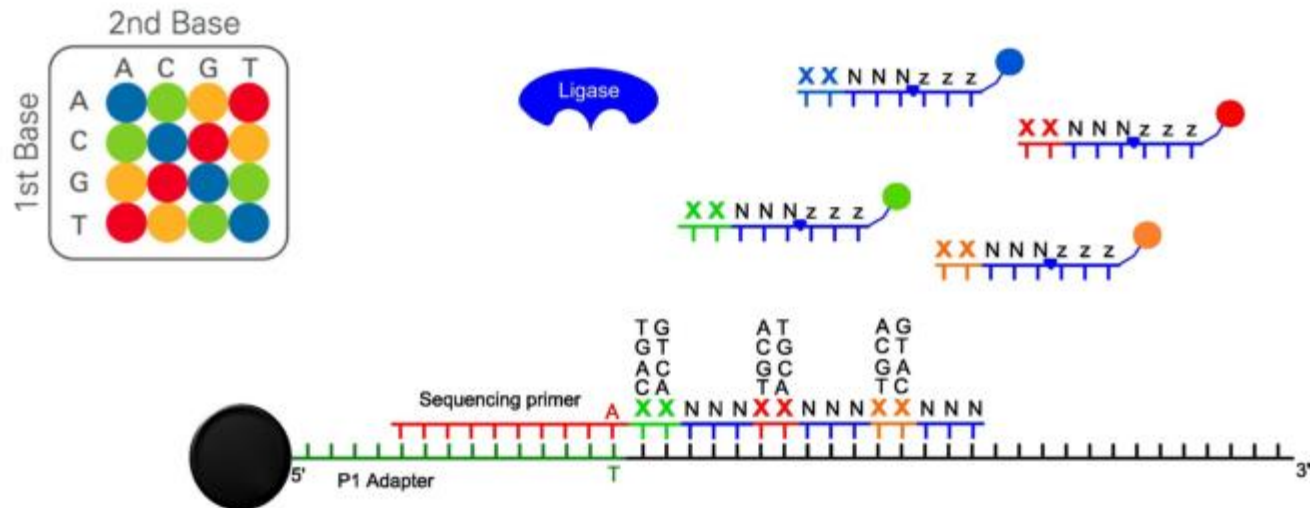
- **Primera Ronda de secuenciación.** ... Así, otra sonda es incorporada y se repite otra vez el procedimiento....



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

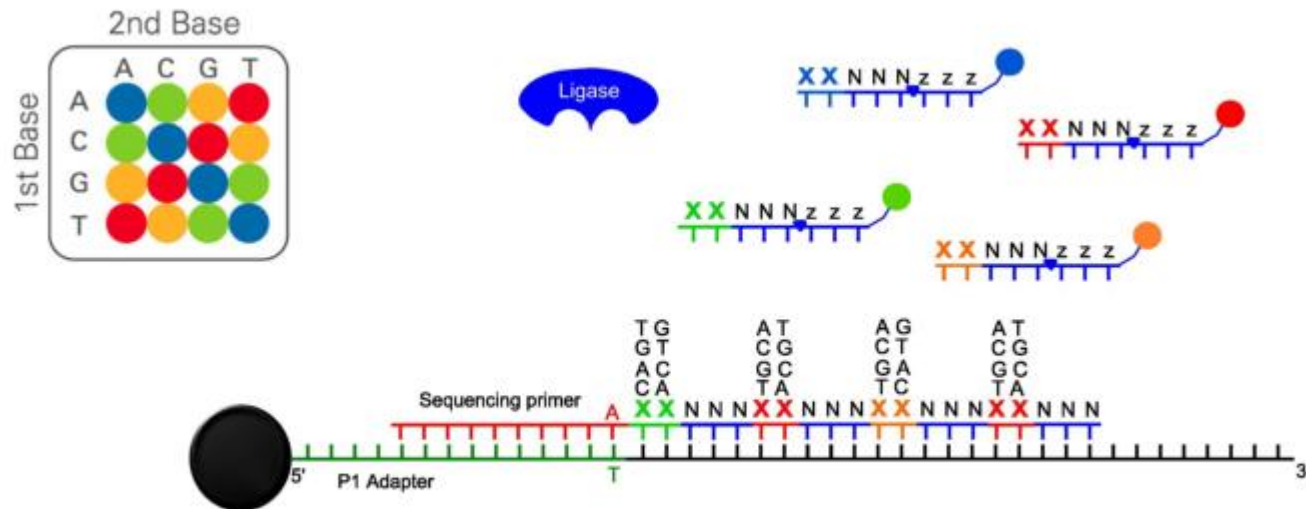
- **Primera Ronda de secuenciación.** ... Así, otra sonda es incorporada y se repite el procedimiento, introduciéndose una más....



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

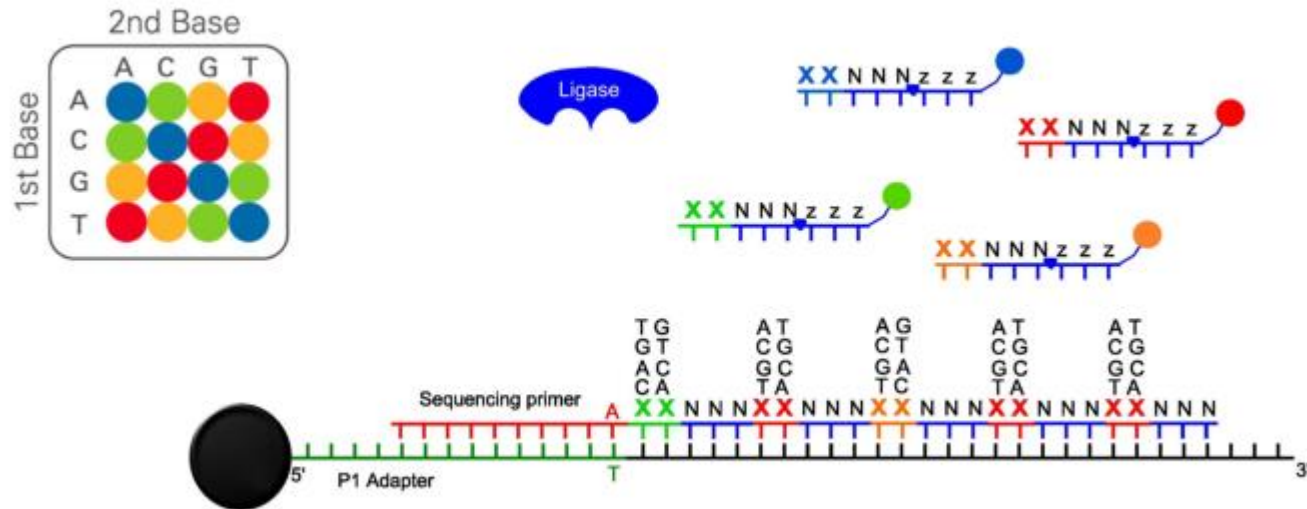
- **Primera Ronda de secuenciación.** ... Así, otra sonda es incorporada y se repite el procedimiento, introduciéndose una más, y otra....



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

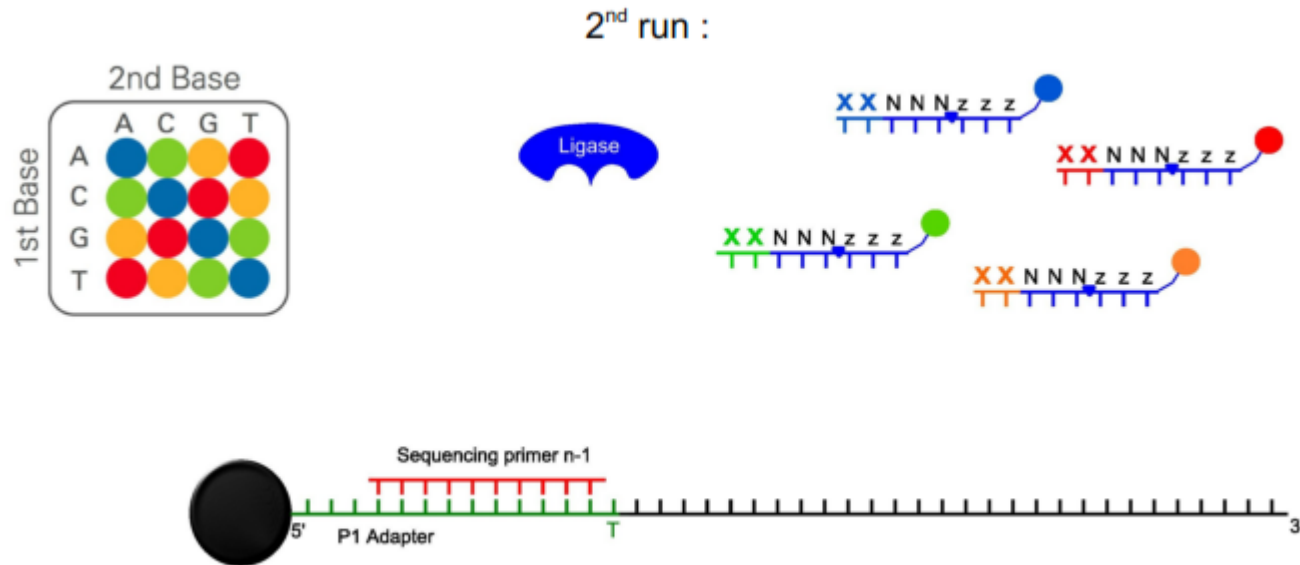
- **Primera Ronda de secuenciación.** ... Así, otra sonda es incorporada y se repite el procedimiento, introduciéndose una más, y otra hasta completar el molde....



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

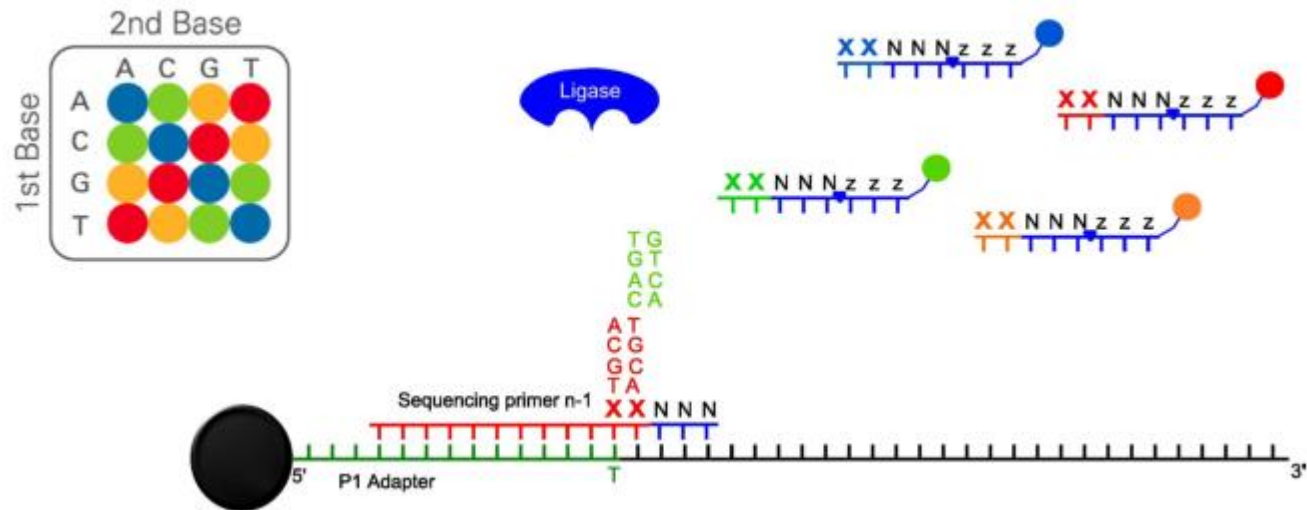
- **Segunda Ronda de secuenciación:** Finalmente, se **desnaturaliza** todo el **dsDNA** y se **inicia una nueva ronda de secuenciación** con un **nuevo primer** (hibrida en la secuencia adaptadora pero en una posición -1 respecto del *primer* utilizado en la primera ronda).



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

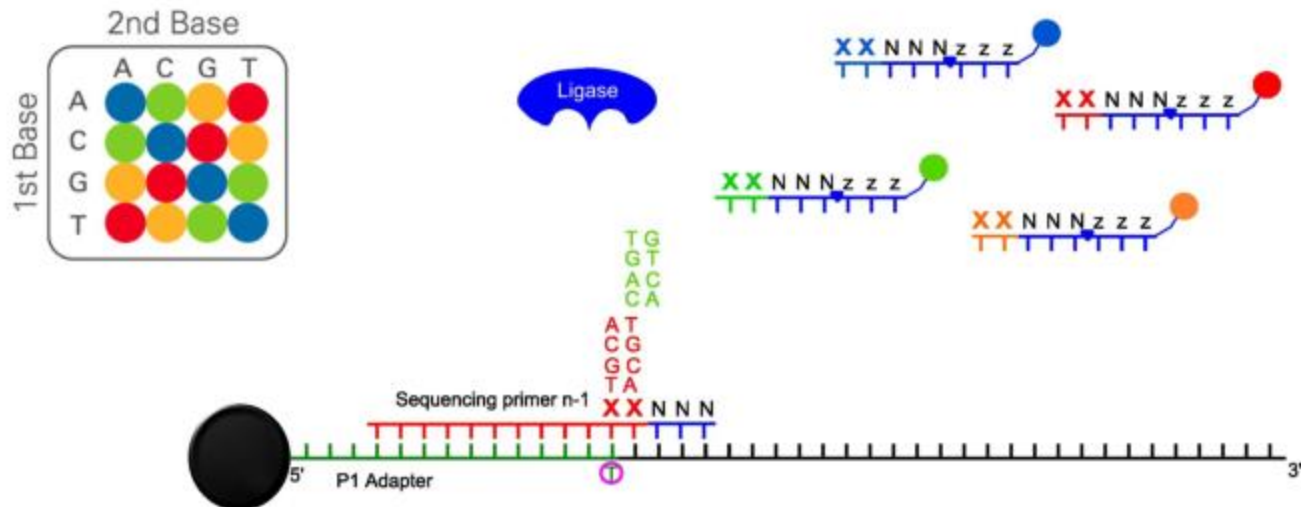
- **Segunda Ronda de secuenciación:** Cuando se **incorpora** una nueva sonda y se registra la **fluorescencia**...



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

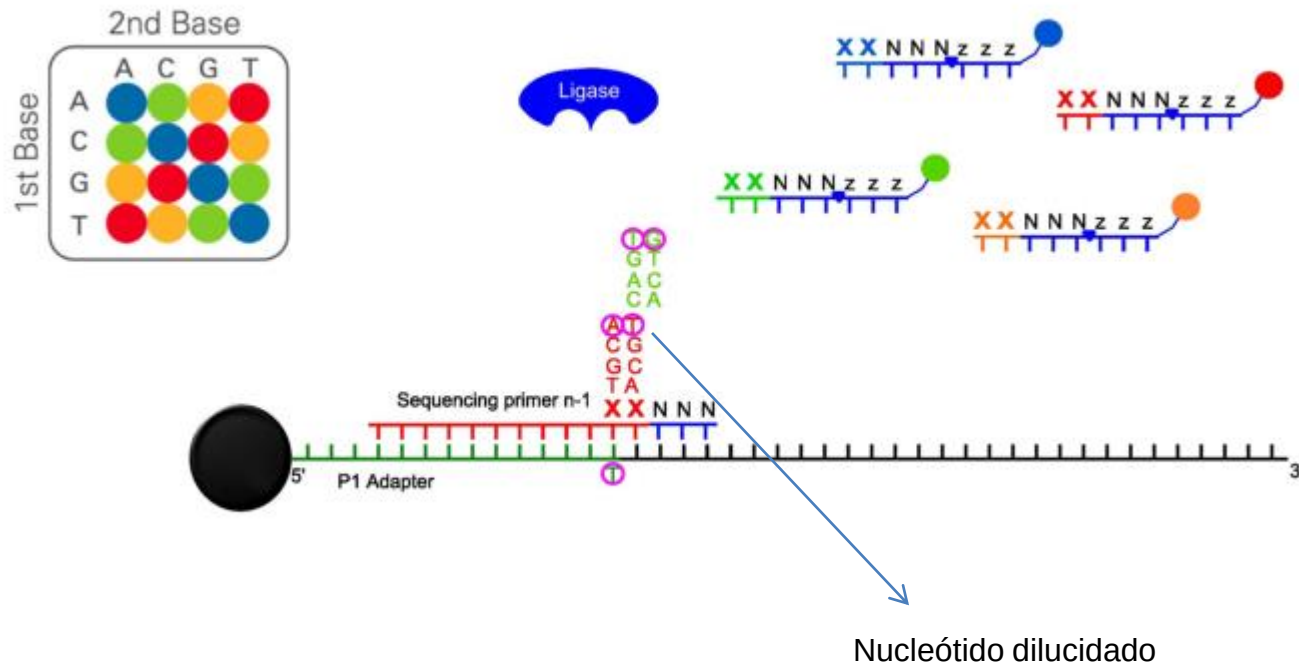
- **Segunda Ronda de secuenciación:** ... el **software** asociado al equipamiento **dilucida** cuál **nucleótido** se encuentra en el **DNA molde**...



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

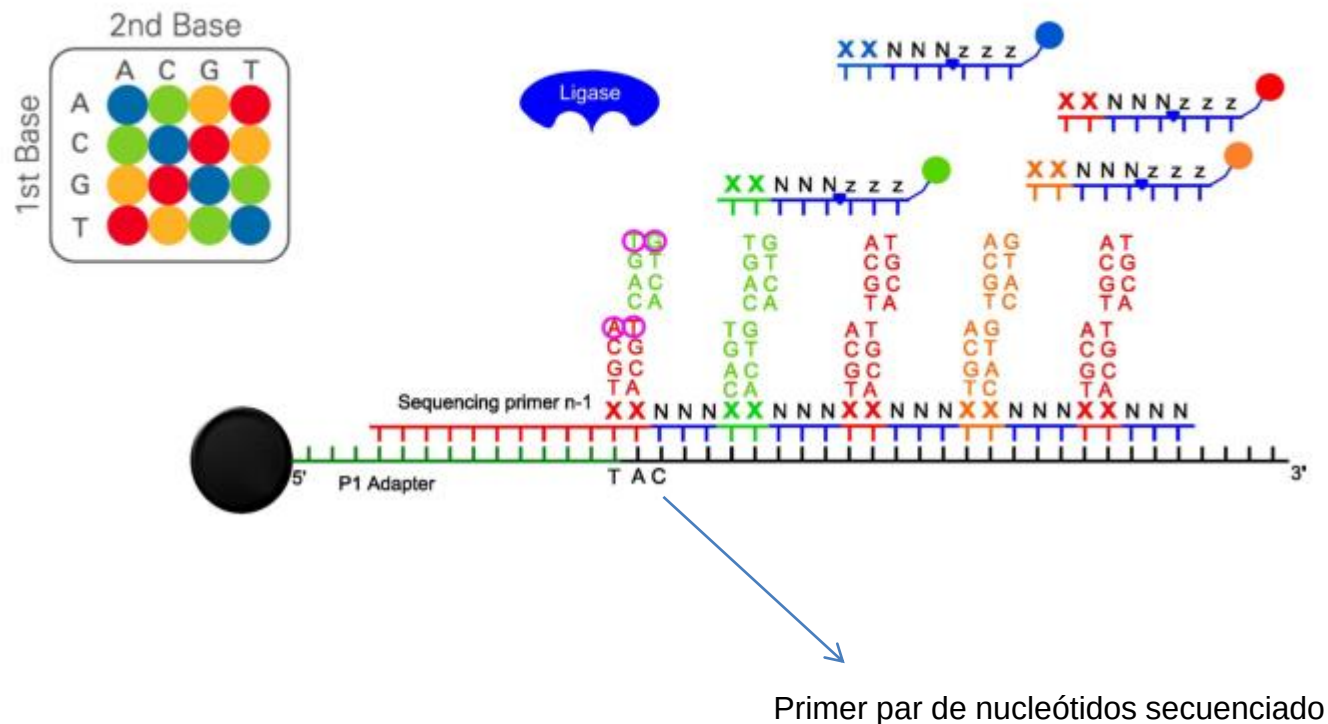
- **Segunda Ronda de secuenciación:** ... conociéndose así los primeros datos de **secuencia** del fragmento...



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

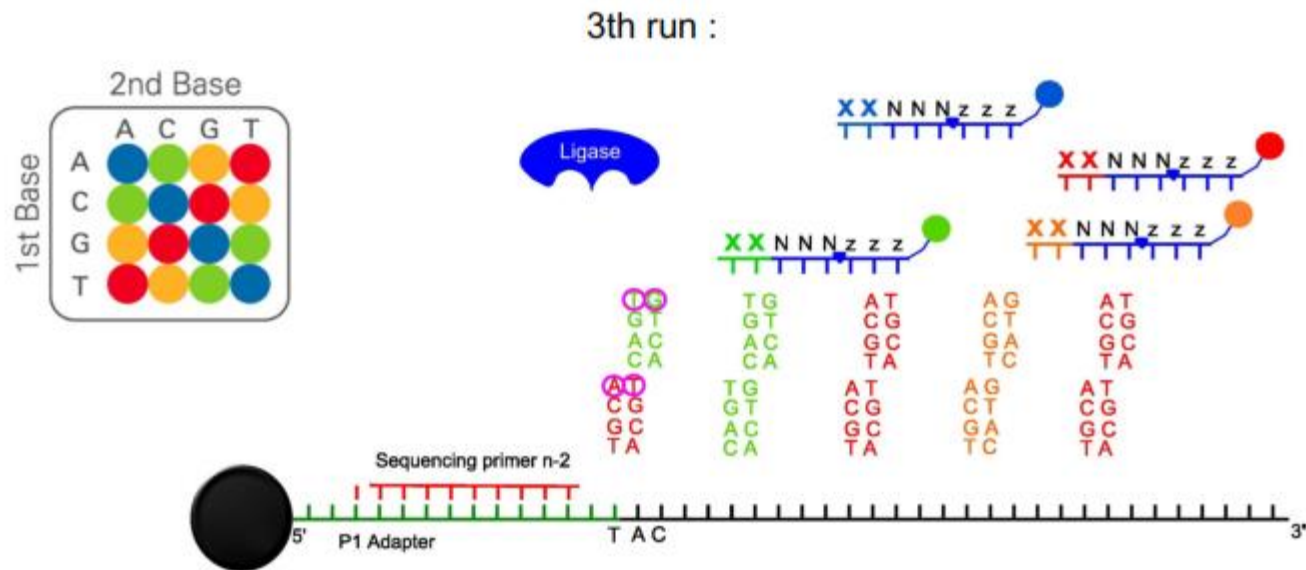
- **Segunda Ronda de secuenciación:** ... conociéndose así los primeros datos de **secuencia** del fragmento...



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

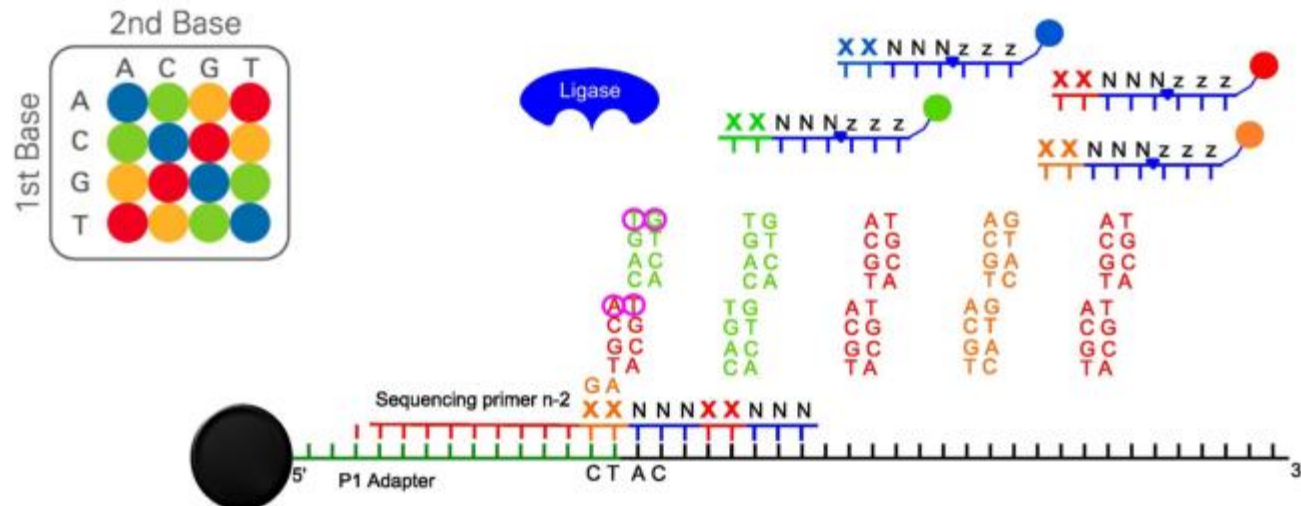
- **Tercera Ronda de secuenciación:** Nuevamente, se **desnaturaliza** todo el **dsDNA** y se **inicia** una **nueva ronda de secuenciación** con un **nuevo primer** (hibrida en la secuencia adaptadora pero en una posición -2 respecto del *primer* utilizado en la primera ronda).



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

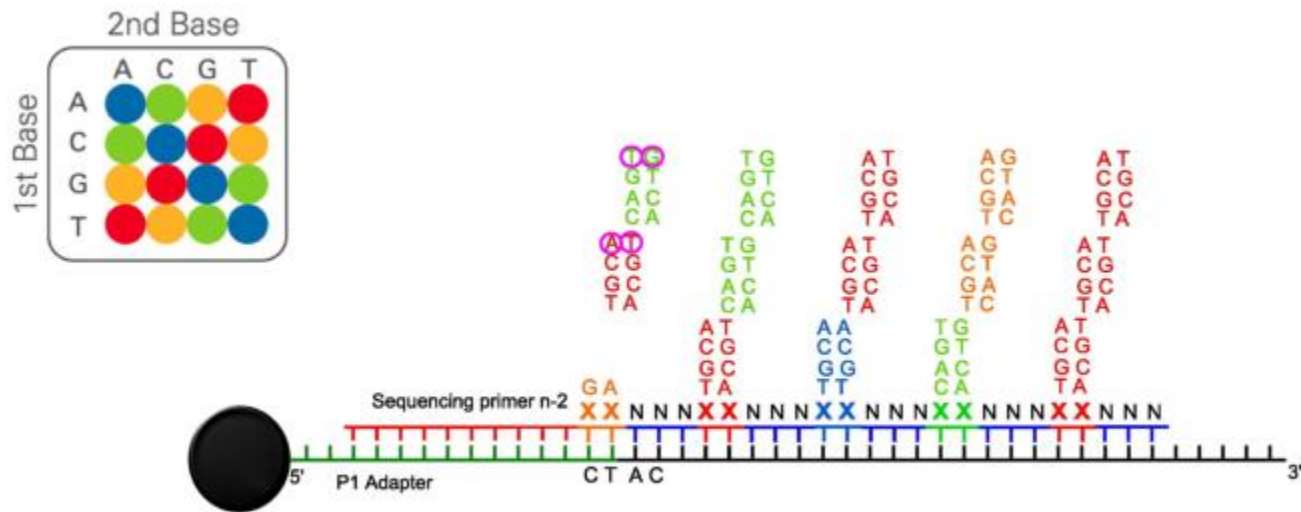
- **Tercera Ronda de secuenciación:** El procedimiento se desarrolla del mismo modo que en las rondas anteriores...



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

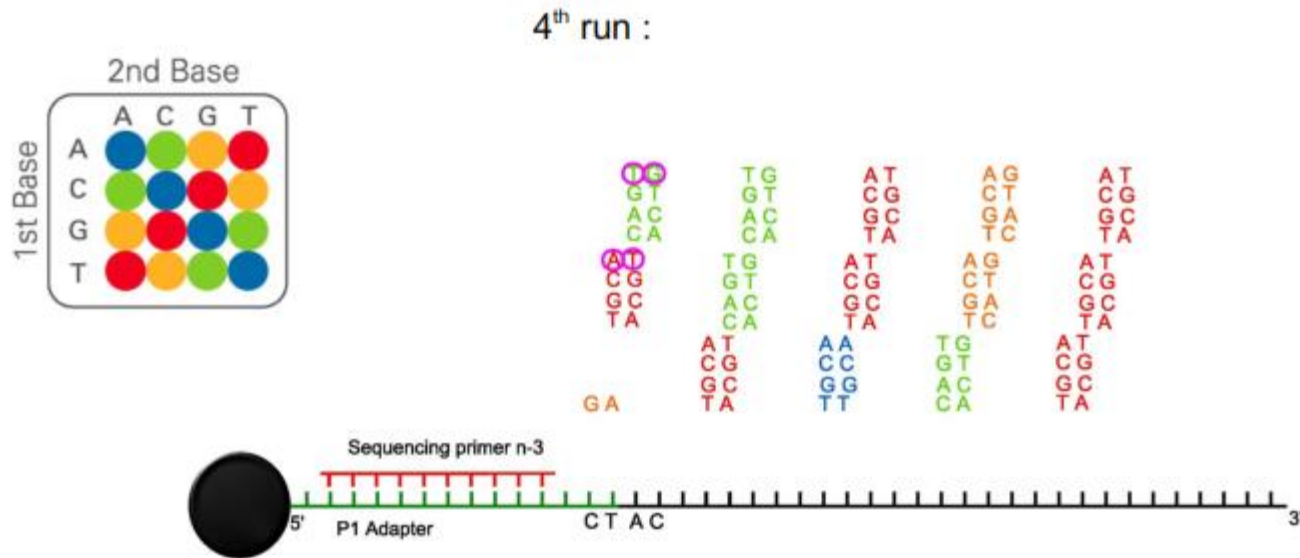
- **Tercera Ronda de secuenciación:** ...hasta **completar el molde...**



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

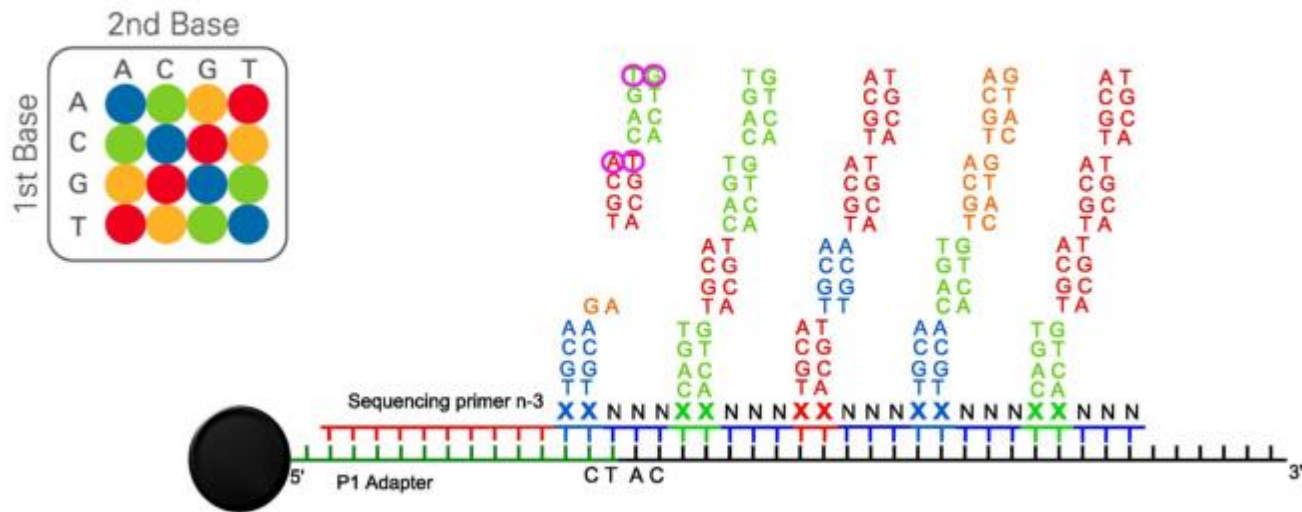
- **Cuarta Ronda de secuenciación:** Se **desnaturaliza** todo el **dsDNA** y se **inicia** una **nueva ronda de secuenciación** con un **nuevo primer** (hibrida en la secuencia adaptadora pero en una posición -3 respecto del *primer* utilizado en la primera ronda).



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

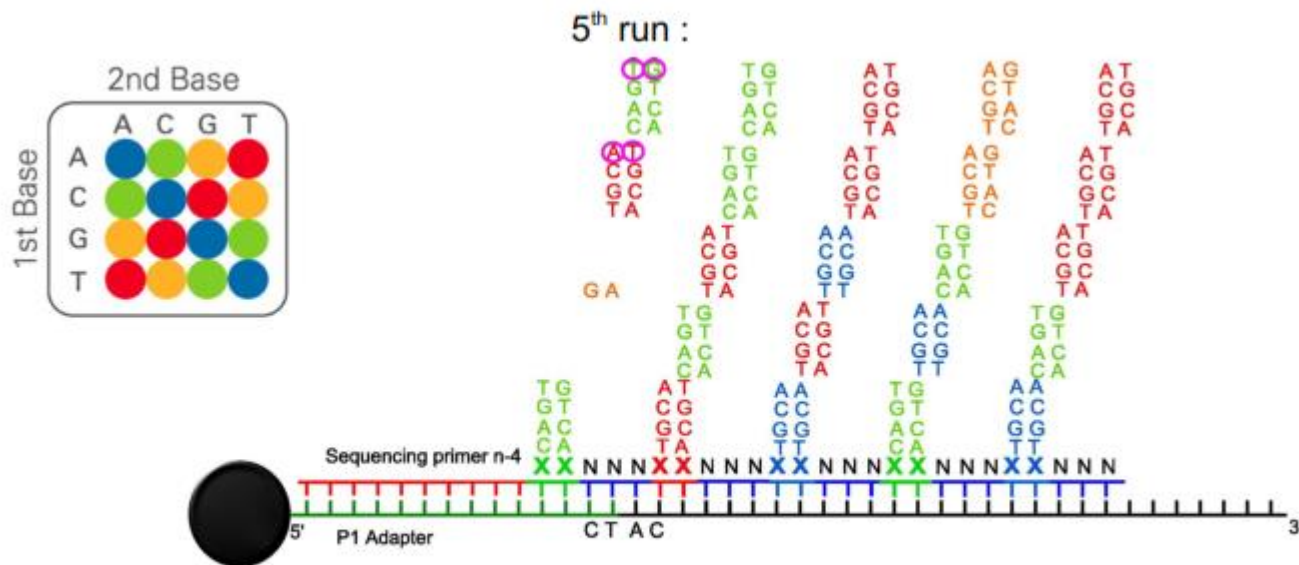
- **Cuarta Ronda de secuenciación:** ...repetiendo todo el procedimiento...



Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

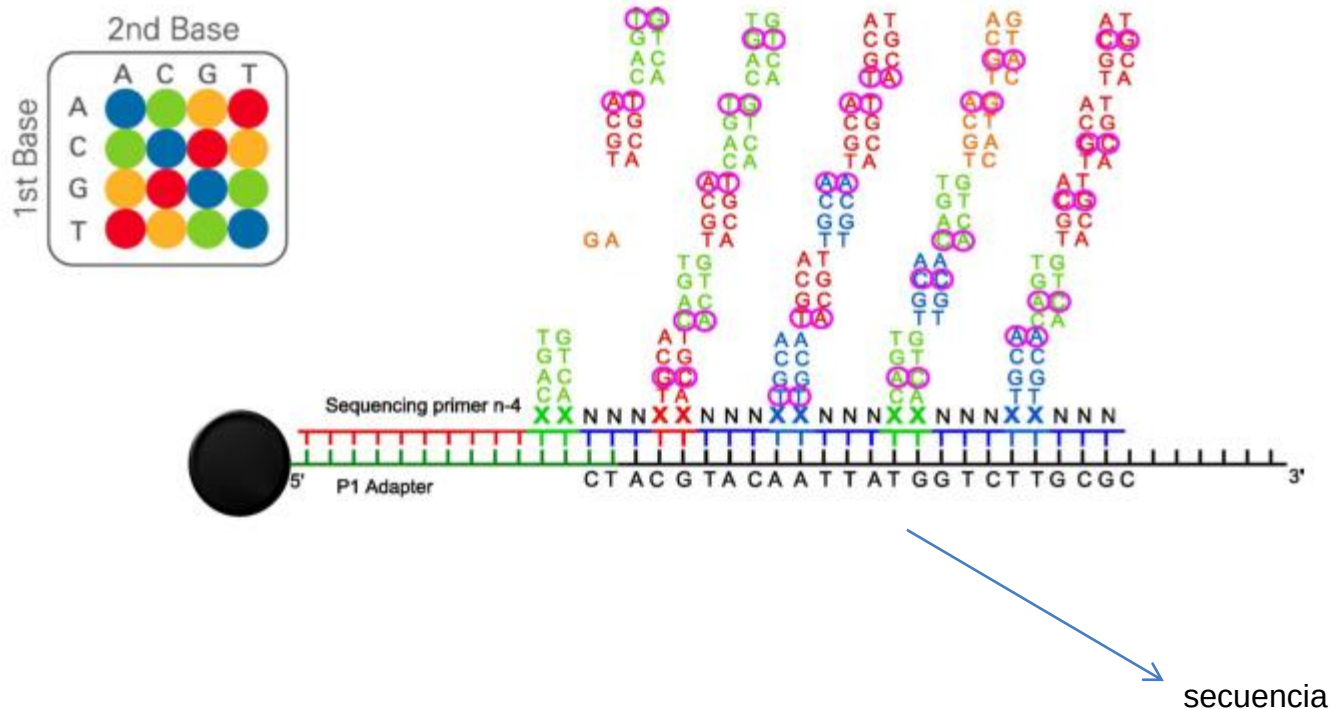
- **Quinta Ronda de secuenciación:** Del mismo modo, **siguen las** rondas con primers desplazados (en este caso *primer -4*)...



Tecnologías de secuenciación: 2G

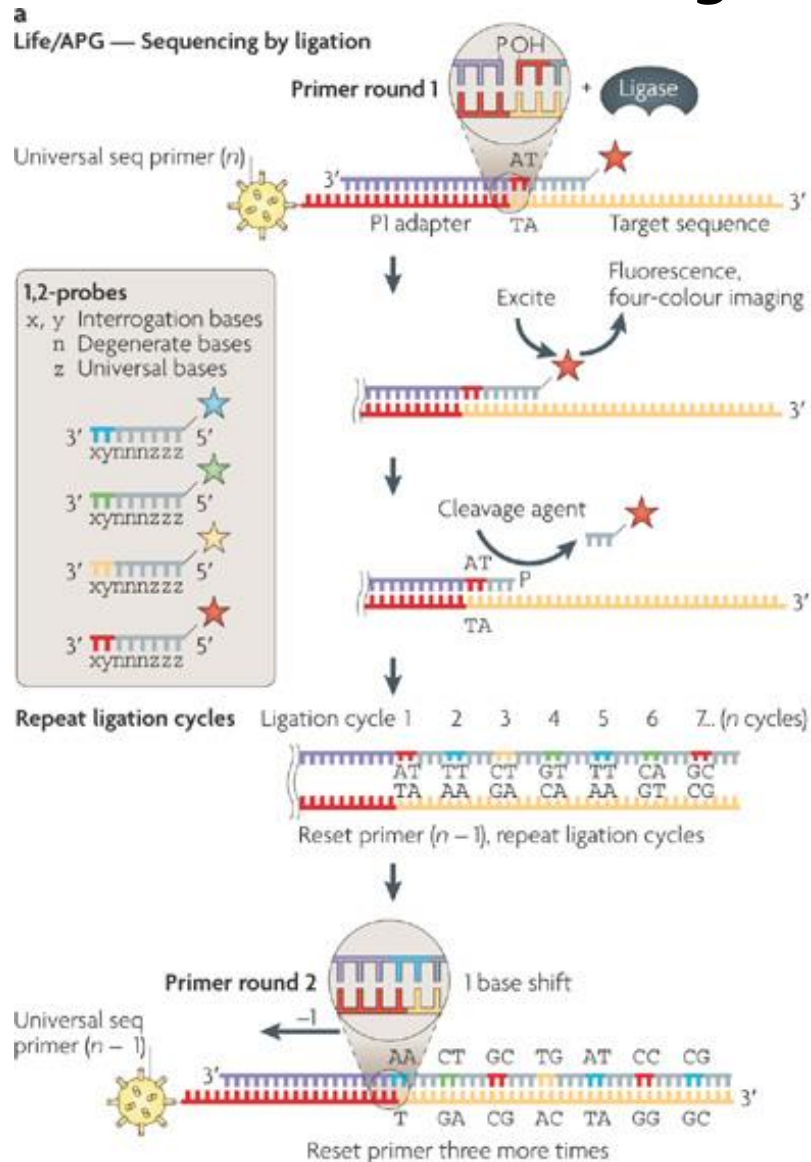
NGS: tecnología SOLID

- **Quinta Ronda de secuenciación:** ... continuándose interpretando cuál es la **secuencia del DNA molde**.



Tecnologías de secuenciación: 2G

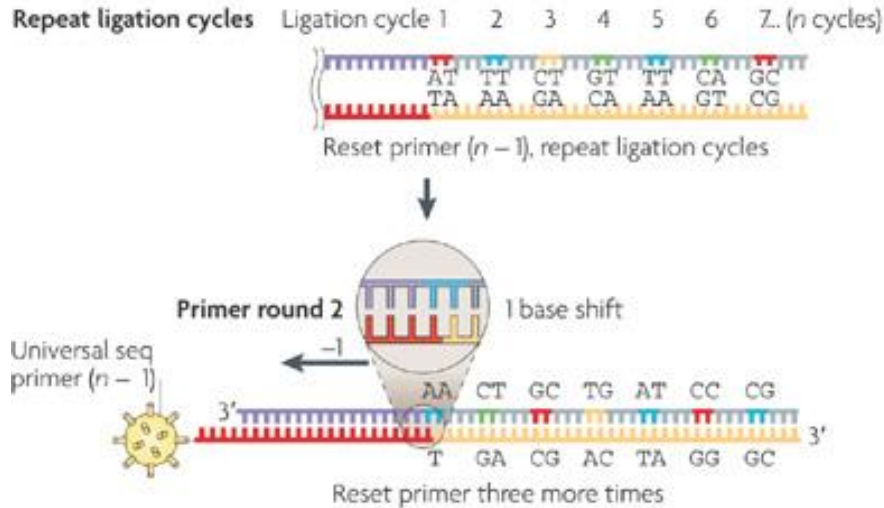
NGS: tecnología SOLID



Resumen de la
reacción de
secuenciación
mediada por
ligación de
sondas
fluorescentes

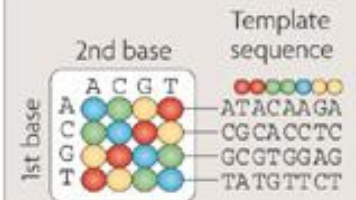
Tecnologías de secuenciación: 2G

NGS: tecnología SOLID

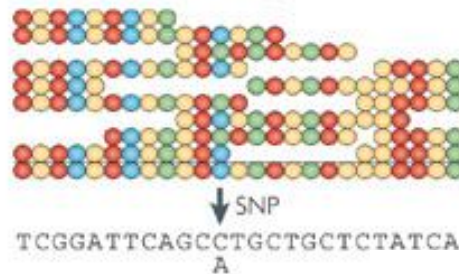


b

Two-base encoding: each target nucleotide is interrogated twice



Alignment of colour-space reads to colour-space reference genome



Read Position		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1	Universal seq primer (n) 3'																																				
2	Universal seq primer (n-1) 3'																																				
3	Universal seq primer (n-2) 3'																																				
4	Universal seq primer (n-3) 3'																																				
5	Universal seq primer (n-4) 3'																																				

Resumen de la
reacción de
secuenciación
mediada por
ligación de
sondas
fluorescentes

Descripción de los métodos más utilizados clasificados dentro de las tecnologías NGS

Tecnología Ion Torrent

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

- La **tecnología Ion Torrent** se basa en un método de **secuenciación** basado en **semiconductores de iones**.
- Esta metodología **detecta** los **iones hidrógeno** liberados **luego** de un evento de **polimerización**. Por ende, es una **secuenciación** basada en **síntesis**.
- Esta tecnología **no utiliza** nucleótidos modificados ni está basada en **propiedades ópticas** de compuestos.
- La tecnología fue licenciada en el año **2010** y hoy en día es comercializada por ***Life Technologies***.

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

PGM (Personal Genome Machine)



Proton



Equipamiento

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent



W1 bottle : 350 μ l 100 mM NaOH
W2 bottle : 2 liter W2 solution (contains detergent)
W3 bottle : 50 ml W3 (= buffer)
4 tubes with 20 μ l of the different dNTP's



Tecnologías de secuenciación: 2G

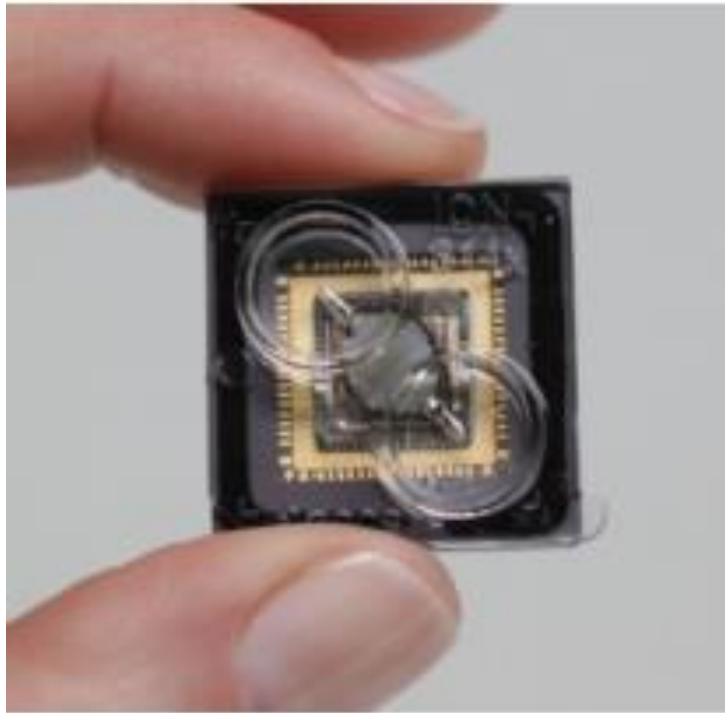
NGS: tecnología Ion Torrent

- Cuando se produce la **incorporación** de un **nucleótido** específico mediante **síntesis** en una reacción de polimerización es retenido en la cadena naciente el nucleósido monofosfato (base de la secuenciación automatizada de Sanger y de la aplicada por Illumina), y **se libera pirofosfato** (base de la pirosecuenciación y de la tecnología 454) y **un ión hidrógeno**.
- El **ión hidrógeno cambiará** en consecuencia el **pH** del entorno, señal que si es registrada permitirá identificar que el ssDNA molde que está siendo secuenciado posee un nucleótido complementario al incorporado.
- Mediante un **ISFET** (*Ion Sensitive Field Effect Transistor*) es posible registrar tales sutiles cambios del pH reportando una señal eléctrica.
- Si a su vez se dispone un **CMOS** (*Complementary metal-oxide-semiconductor*), se termina conformando un **chip** similar a los utilizados en la industria de la electrónica.

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

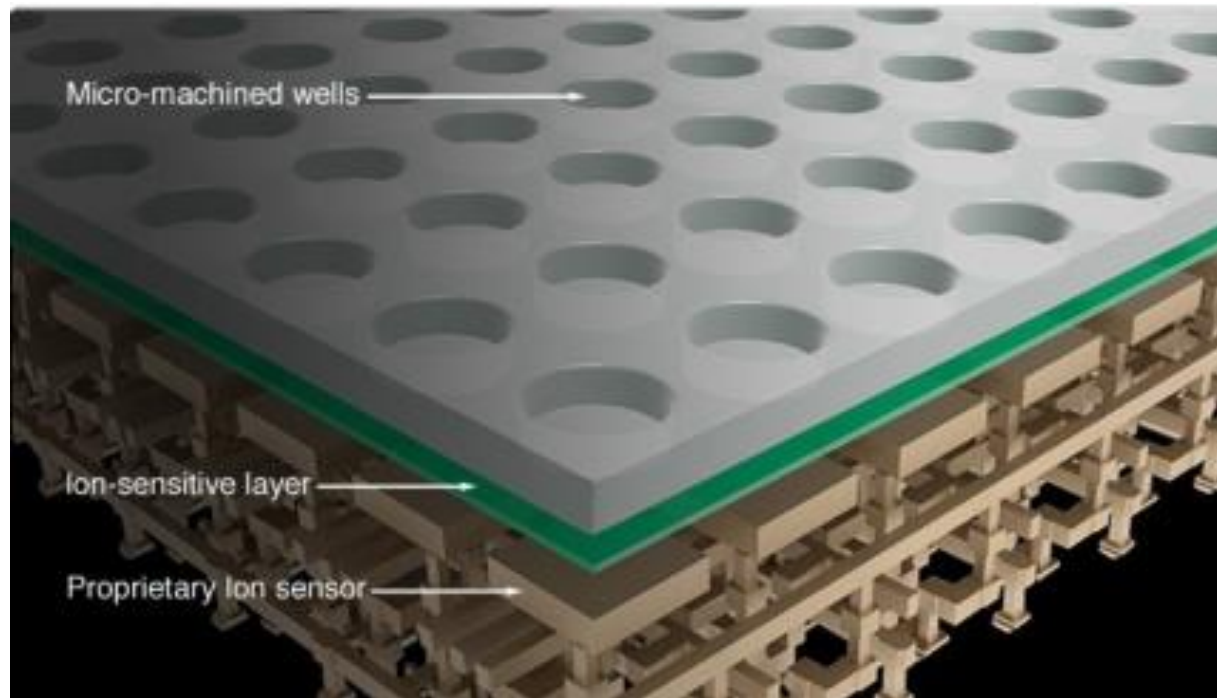
Dimensiones del *chip*



Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

Características del *chip*...

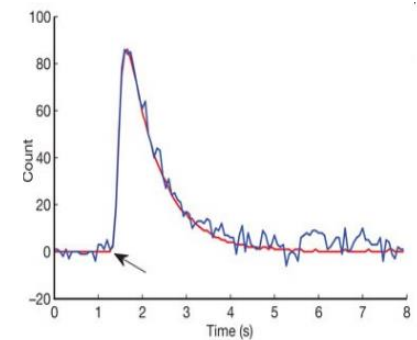
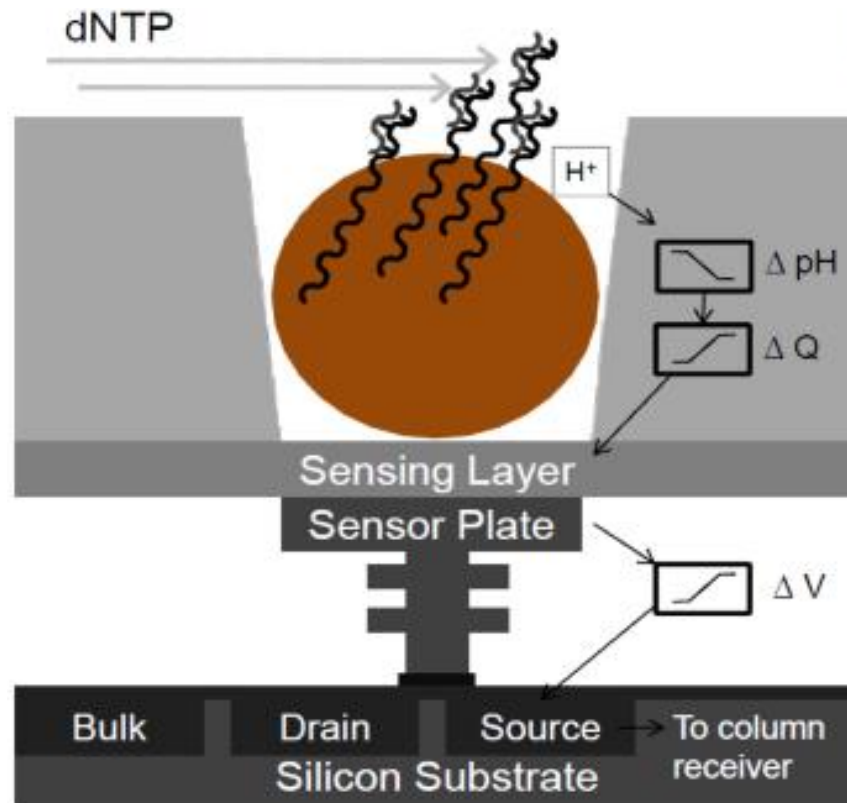


Esta **tecnología** está confeccionada para utilizar **bibliotecas** construidas en **microesferas**, las cuales serán incorporadas a cada *micro-well*.

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

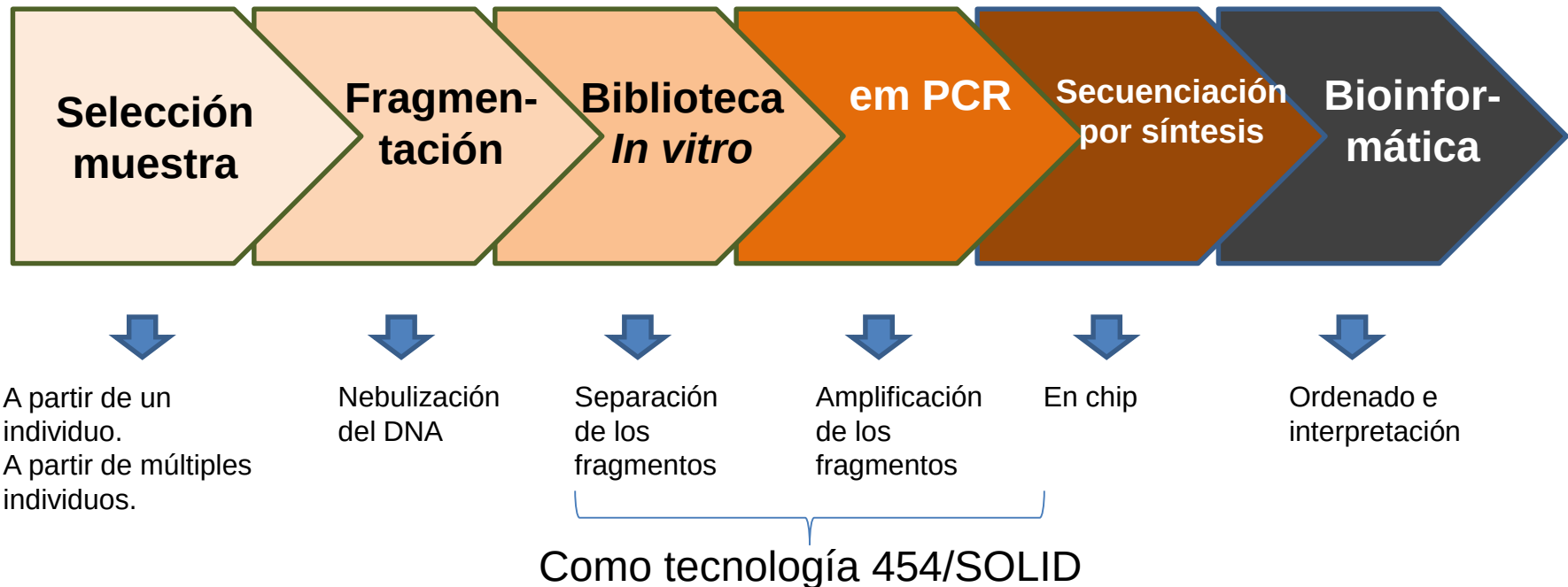
Bases de la detección de un nucleótido incorporado...



Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

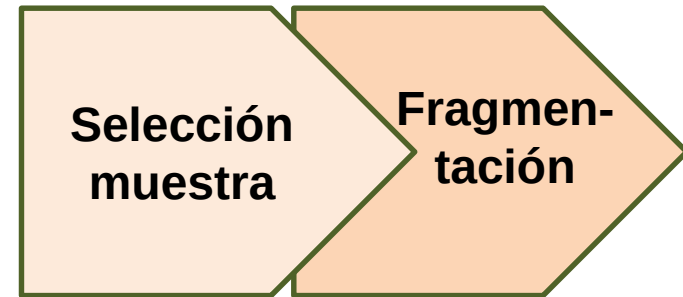
Etapas involucradas en la tecnología *Ion Torrent*



Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

Etapas involucradas en la tecnología *Ion Torrent*



A partir de un individuo.
A partir de múltiples individuos.

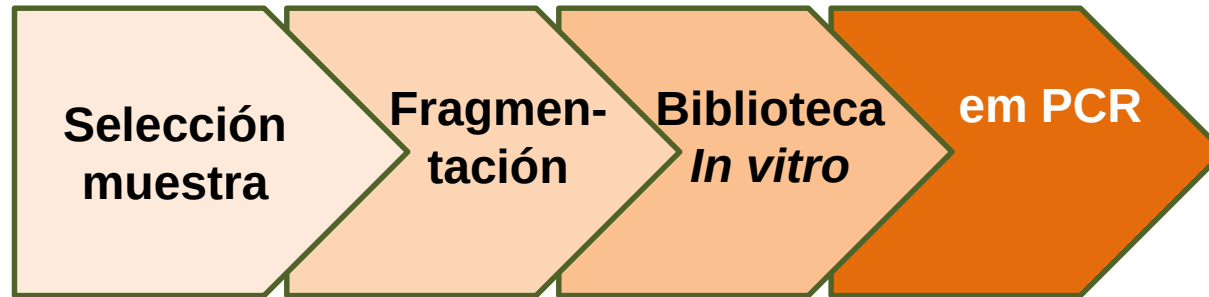
Nebulización del DNA

Se puede partir de **10 ng de gDNA**.
Los fragmentos deben tener un tamaño de entre **100-200 pb**.
A los fragmentos se le ligan **adaptadores**.

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

Etapas involucradas en la tecnología *Ion Torrent*



Se utilizan procedimientos similares a los realizados para 454 y SOLID

↓
A partir de un individuo.
A partir de múltiples individuos.

↓
Nebulización del DNA

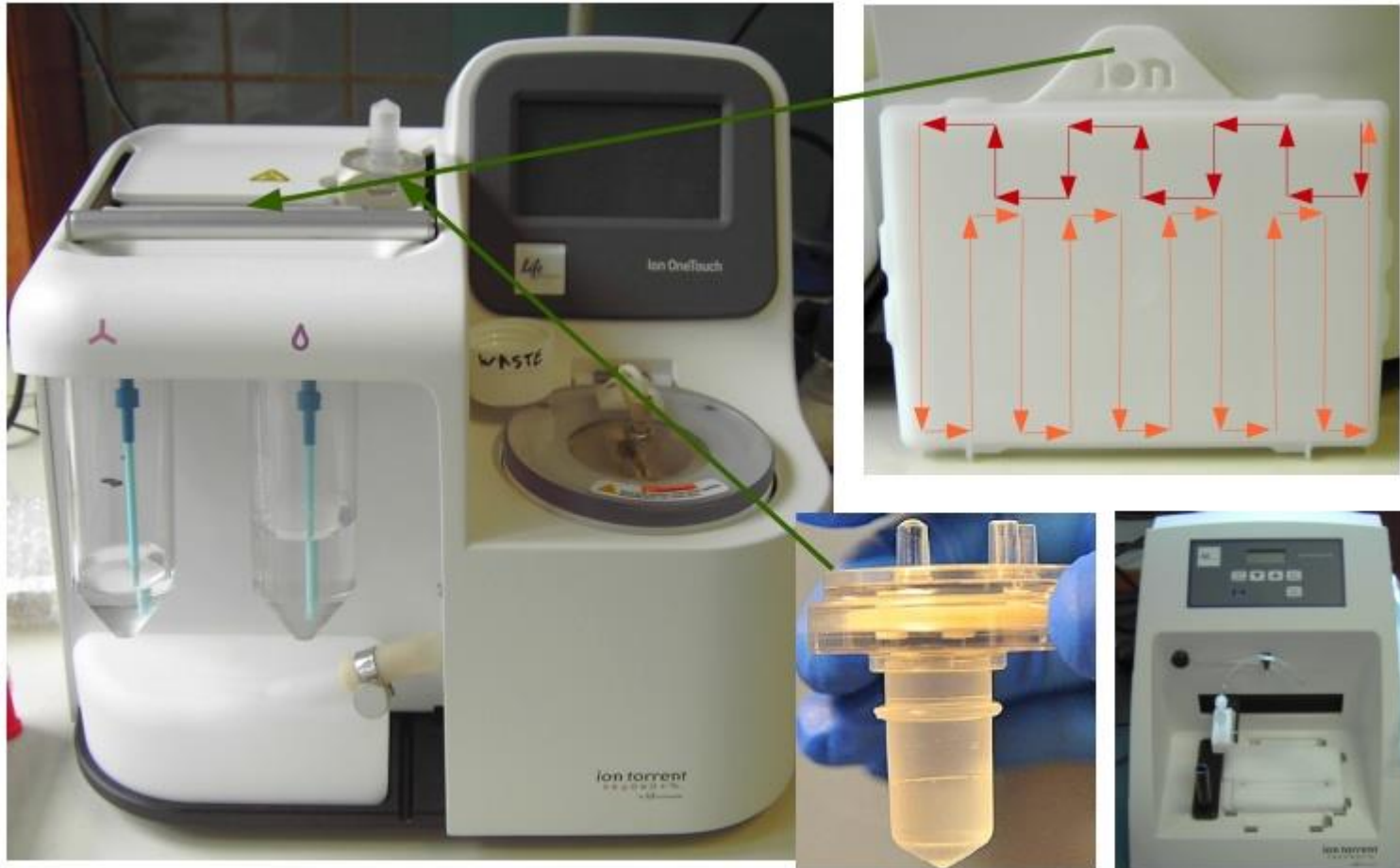
↓
Separación de los fragmentos

↓
Amplificación de los fragmentos

Como tecnología 454/SOLID

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent



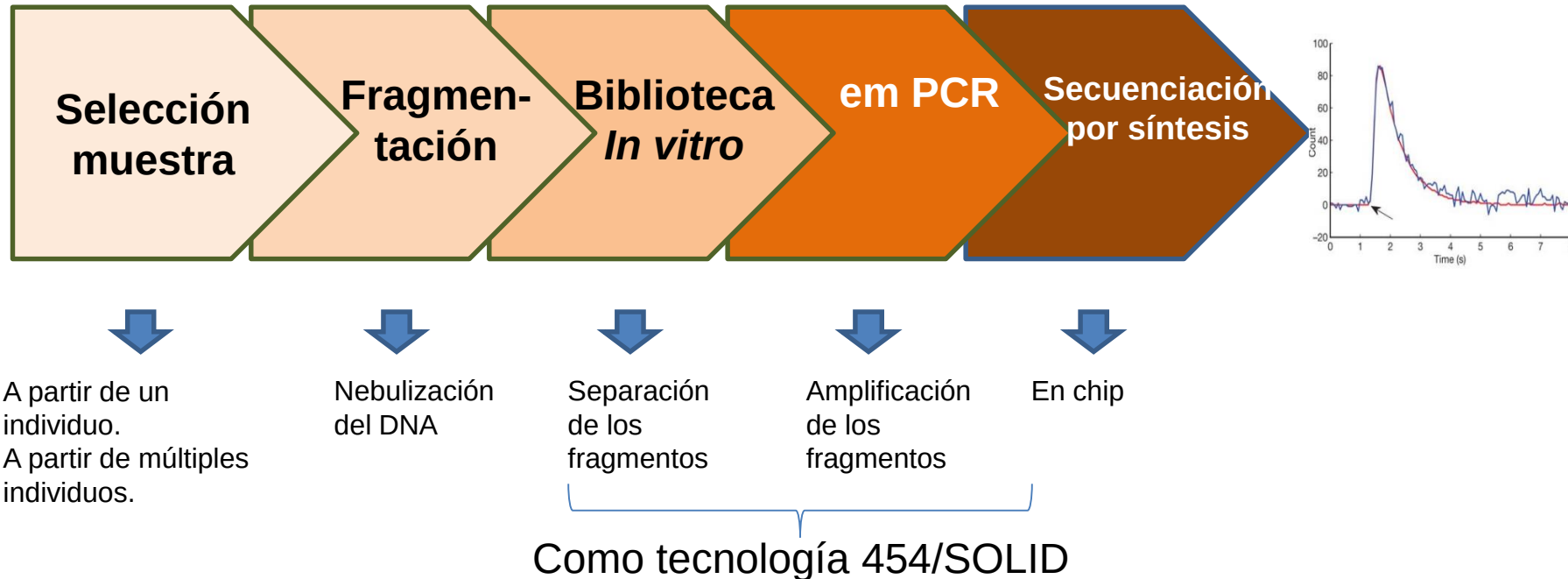
15 min hands-on; 4 hours amplification; 35 min enrichment

Equipamiento para **em-PCR** para tecnología Ion Torrent

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

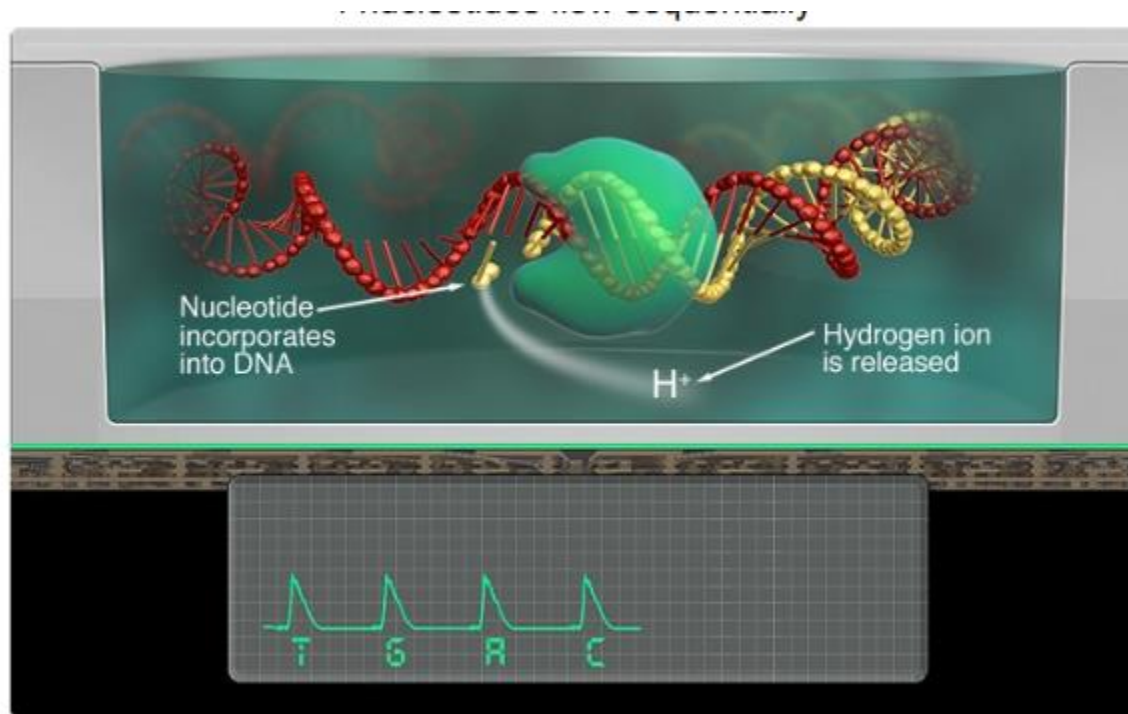
Etapas involucradas en la tecnología *Ion Torrent*



Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

- Sobre el **chip** circula una **mezcla de síntesis** (DNA polimerasa, *primer*, 1 dNTP) en pulsos. Si el **nucleótido se incorpora**, se libera **H⁺**. Esto altera el **pH**, lo cual es **sensado** por el **IFSET**, transmitiéndose un voltaje por el semiconductor.

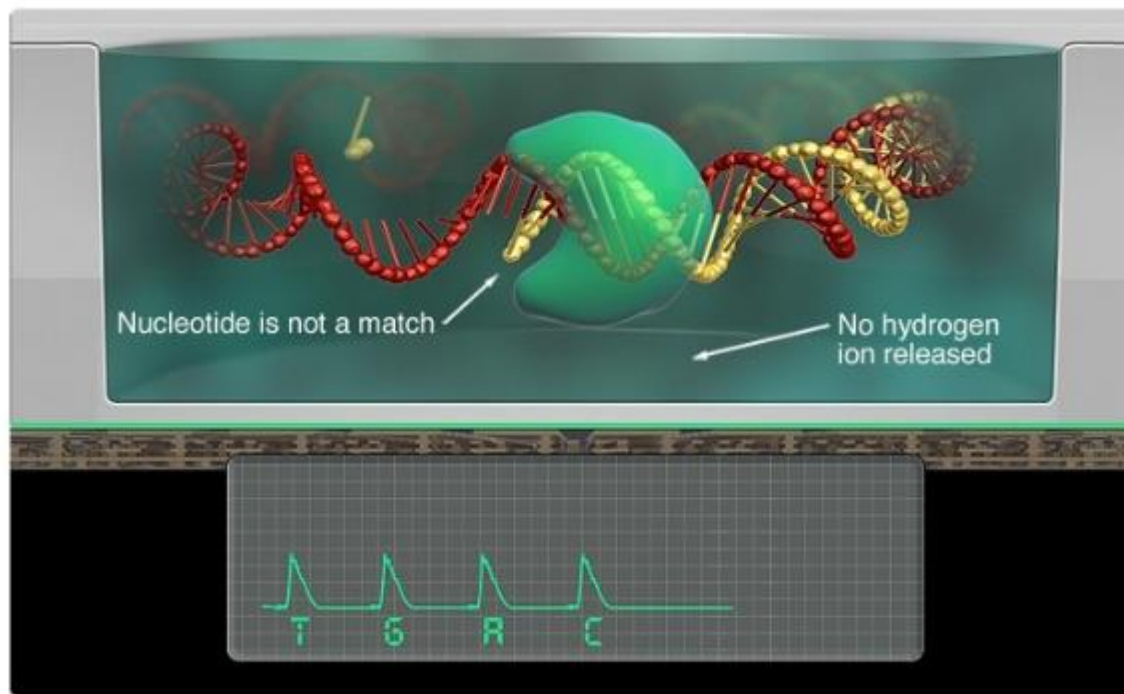


No camera, just a pH sensor

Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

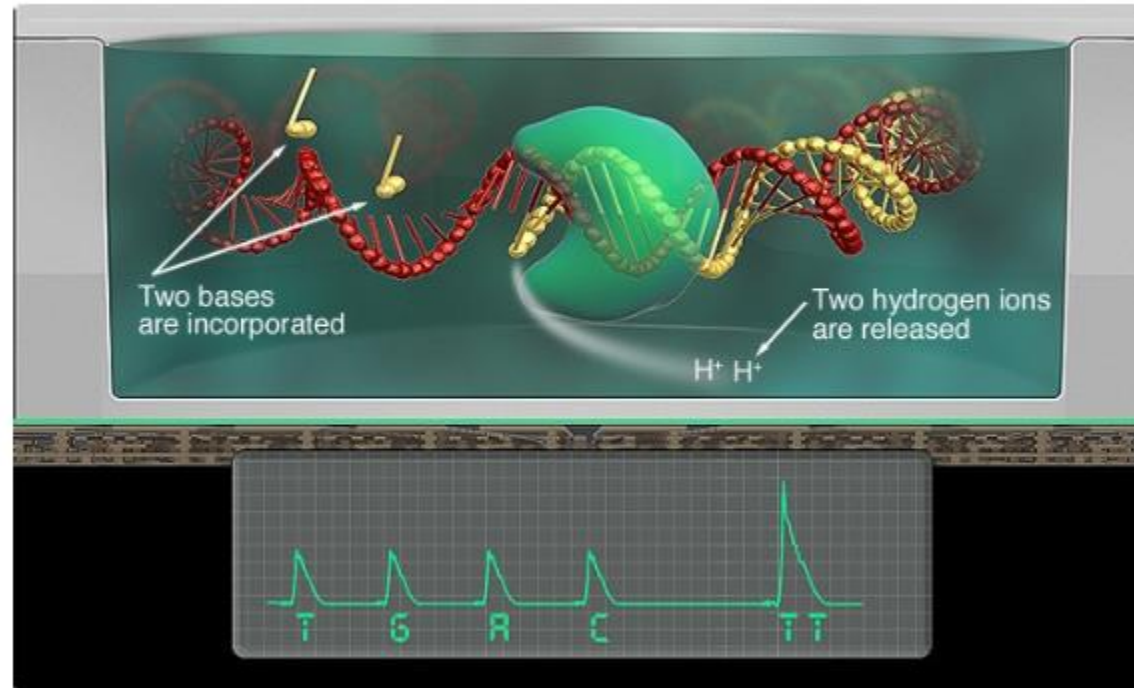
- Cuando se introduce un **dNTP** que no se incorpora, no hay cambio de pH y por ende no existe señal eléctrica en dicho *microwell*..



Tecnologías de secuenciación: 2G

NGS: tecnología Ion Torrent

- Y cuando se introduce un **dNTP** que se incorpora dos veces, el cambio de **pH** es mayor y por ende la señal eléctrica en dicho *microwell* también lo será, interpretándose correctamente.



Descripción de los métodos clasificados dentro de las tecnologías NGS de tercera generación

Tecnología PacBio

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

- **Simple Molecule Real Time sequencing (SMRT)** son procedimientos de secuenciación de moléculas de DNA en tiempo real **sin pasos de amplificación**.
- Las **tecnologías SMRT** se consideran **NGS de tercera generación**, y la primera disponible comercialmente fue la desarrollada por **Pacific Biosciences** en **2011**.
- Esta tecnología se denomina **PacBio**, y está **basada** en **síntesis** con **dNTPs** marcados con **fluoróforos** y en la **lectura** usando la **guía de ondas en modo cero**, o **ZMW** (*zero-mode waveguide*)*.
- La **ZMW** es una **estructura nanofotónica** que **reduce** el **volumen de observación** en más de tres órdenes de magnitud relativo a la microscopia confocal (volúmenes de zeptolitros, 10^{-21} L), posibilitando que no existan pasos previos de amplificación.

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

- Las **tecnologías SMRT** secuencian **moléculas originales** de la muestra.
- Esto es una **ventaja** dado que la **lectura** se realiza sobre la **molécula sintetizada *in vivo***.
- Pero por otro lado, la **principal restricción** está dada en que la **muestra de dsDNA** debe ser de **muy alta calidad**.
- Esto es, **libre de contaminantes** (ssDNA, RNA, proteínas, sales, solventes orgánicos) y con **alto grado de integridad** (ausencia de *nicks*, de dímeros entre bases).
- La **muestra de dsDNA** no debe ser sometida a **altas temperaturas** (>65°C), ni **pHs extremos** (menores que 6 o mayores que 9). **No debe tener intercalantes** ni haber sido expuesta a **radiación UV**.

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

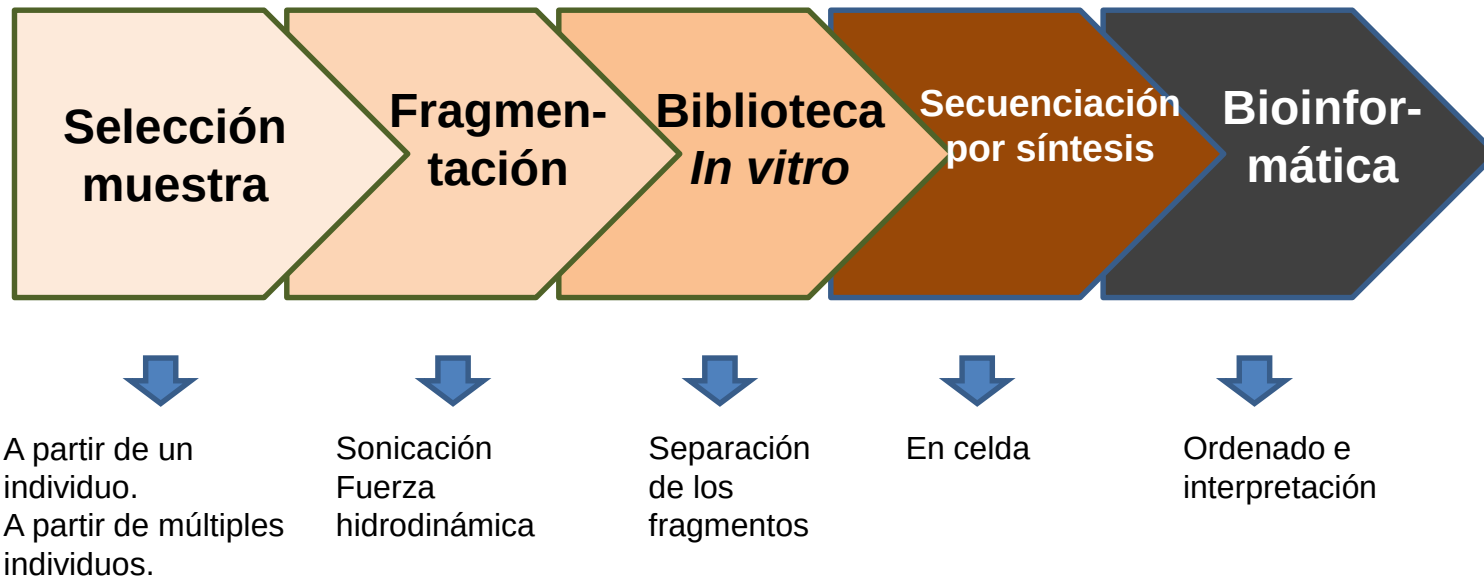


Equipamiento (PacBio RSII)

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

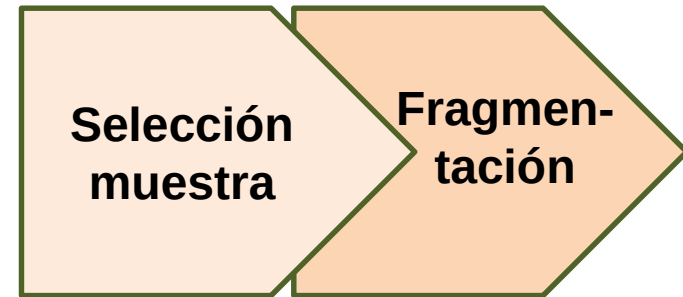
Etapas involucradas en la tecnología *PacBio*



Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

Etapas involucradas en la tecnología *PacBio*



A partir de un individuo.
A partir de múltiples individuos.

Sonicación
Fuerza
hidrodinámica

El gDNA debe tener una masa de 500 ng (para fragmentos de 500 pb), de 1 μ g (para fragmentos de 1 kpb o de 2 kpb), de 4 μ g (para fragmentos de 5 kpb), de 10 μ g (para fragmentos de 10 kpb).

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

- La **fragmentación** del gDNA debe ser **muy controlada** y sin generarse cambios drásticos de temperatura. Además, los fragmentos serán de mayor longitud que los utilizados por las NGS de segunda generación (mayores a 800 pb).
- Ante esta situación, en preferencia a la nebulización, la tecnología *PacBio* requiere otros procedimientos, basados en **sonicación**.



M220 ultrasonicator

ultrasonido



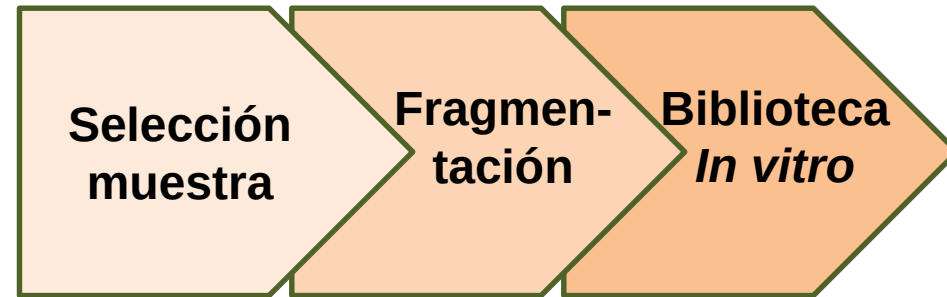
HydroShear

Fuerzas
hidrodinámicas

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

Etapas involucradas en la tecnología *PacBio*



↓

A partir de un individuo.
A partir de múltiples individuos.

↓

Sonicación
Fuerza
hidrodinámica

↓

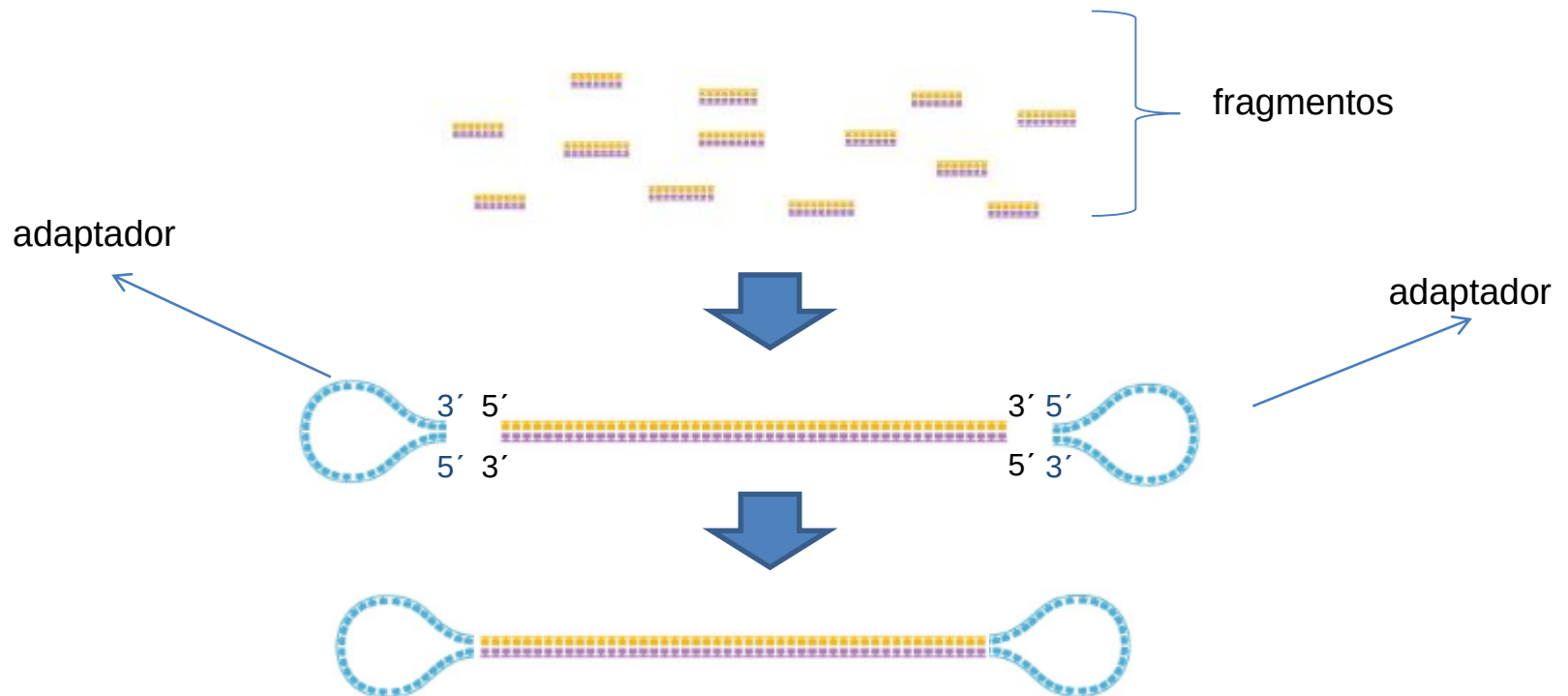
Separación
de los
fragmentos

El gDNA fragmentado (generalmente de 2 kpb) es reparado enzimáticamente en sus extremos para transformarlos en romos o apto para ligación TA.

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

- Luego de disponer de los fragmentos de dsDNA de alrededor de 2 kpb y con extremos romos, se procede a la **ligación de adaptadores**.
- Los **adaptadores** tienen **estructura de horquilla** (*hairpin*).



Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

Ejemplo de secuencias adaptadoras...

5-GTACGGGTCTCACCCGGTTAACTGCACCTGCATTAA-3

5-CCTAAGGTCTCGGAAGGAAATTATTTTCGAAAAAAGA-3

Tecnologías de secuenciación: 3G

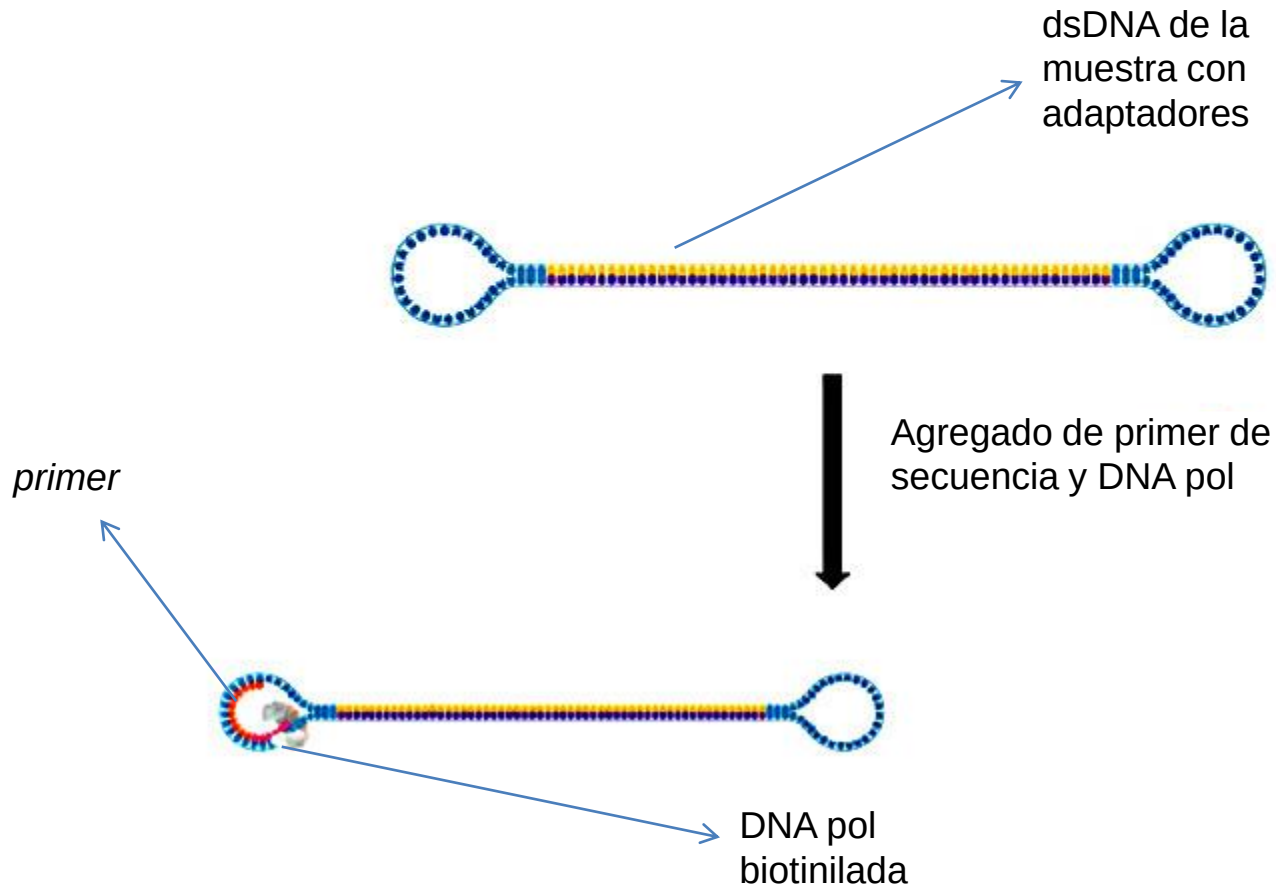
NGS: tecnología PacBio

- **Luego** de tener los fragmentos con los extremos modificados, se debe **agregar** el ***primer*** de secuencia para que suceda el *annealing* con el adaptador.
- **Posteriormente**, se agrega la **DNA polimerasa (ϕ 29)** para que se forme el complejo, la cual posee unida **biotina**.
- La polimerasa **ϕ 29** es codificada por el fago del mismo nombre que multiplica en ***Bacillus subtilis***.
- Esta polimerasa es muy **procesiva**, tiene **actividad correctora de pruebas** y tiene **mayor preferencia** por el **ssDNA**, simulando una actividad helicasa en el proceso de síntesis.

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

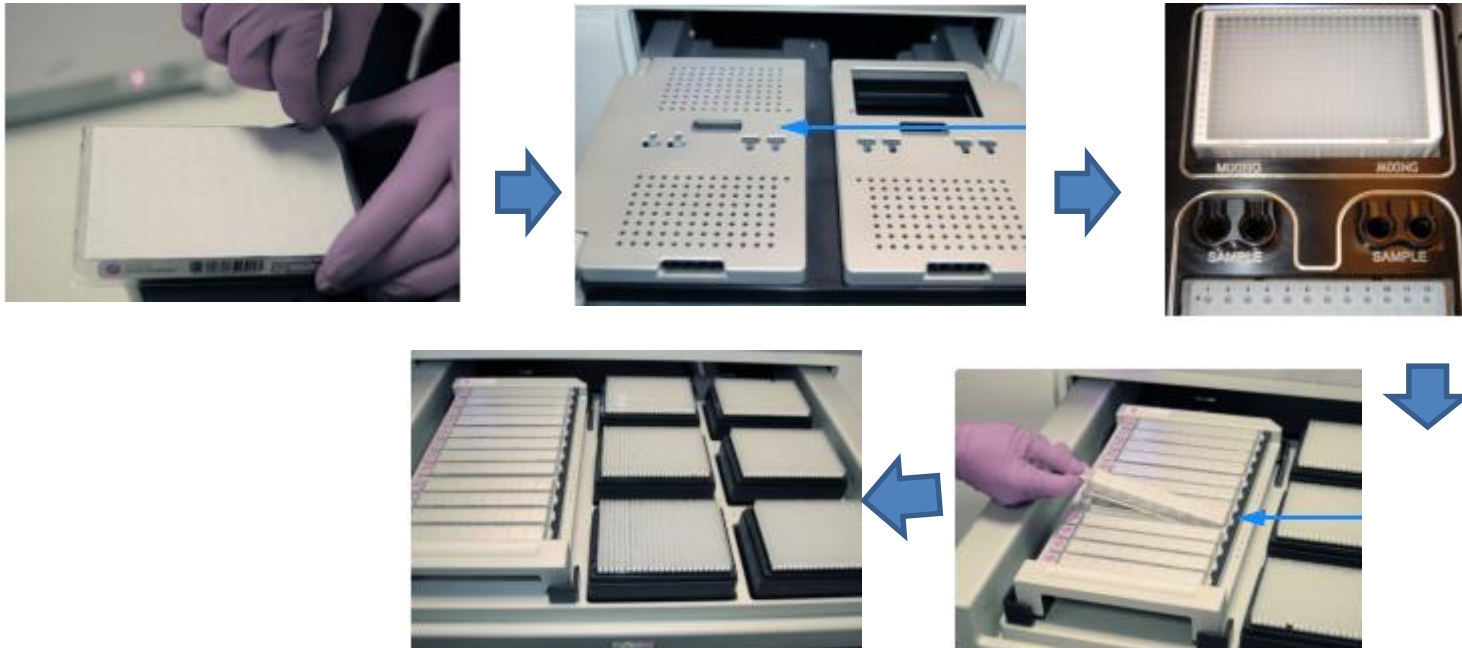
- La **DNA pol** se une al DNA, previamente complejoado con el ***primer***.



Tecnologías de secuenciación: 3G

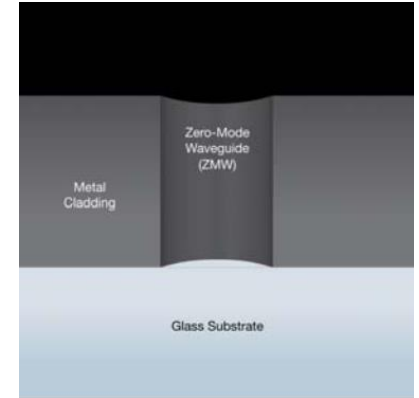
NGS: tecnología PacBio

- La **muestra** así preparada (sin dNTPs) es **depositada** en **placas de 96 wells**.
- **Luego** cada una de las placas **son dispuestas** en el **equipo**, el cual se encarga de disponer los **complejos de DNA/Enzima** en pequeños orificios (**ZMW**) de una **celda**. Solo queda **un complejo por celda**.



Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

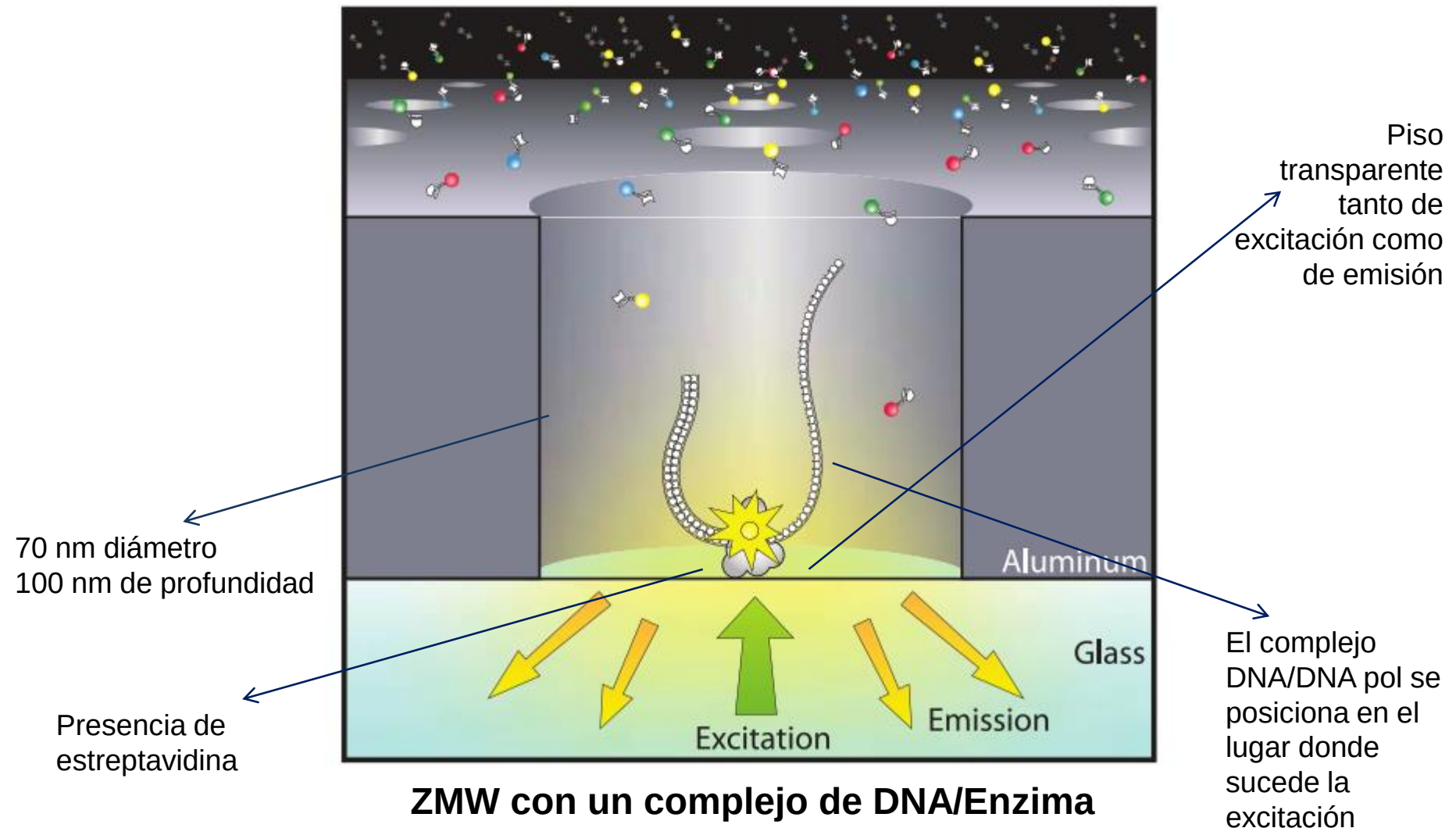


ZMW

Celdas con ZMWs (alrededor de 150000)

Tecnologías de secuenciación: 3G

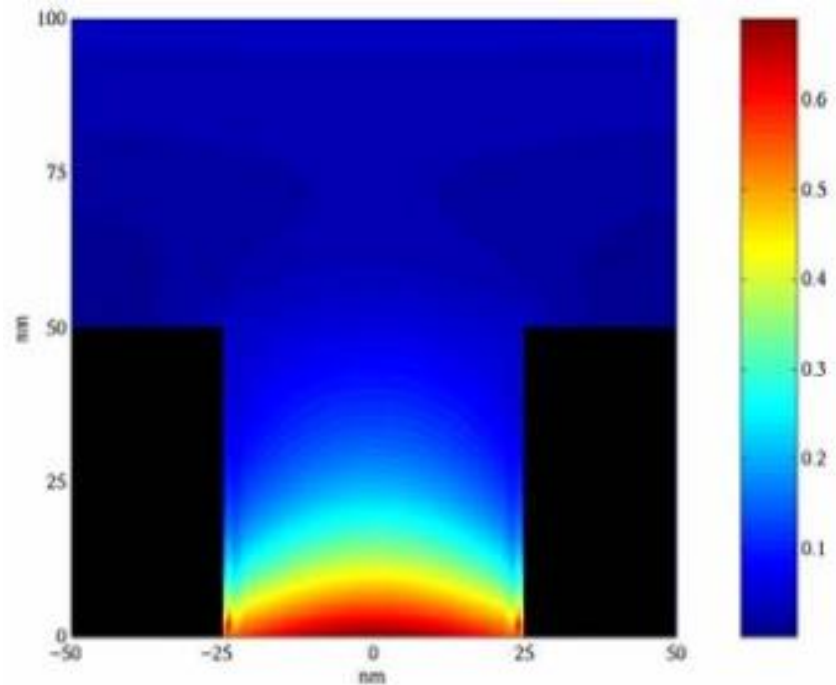
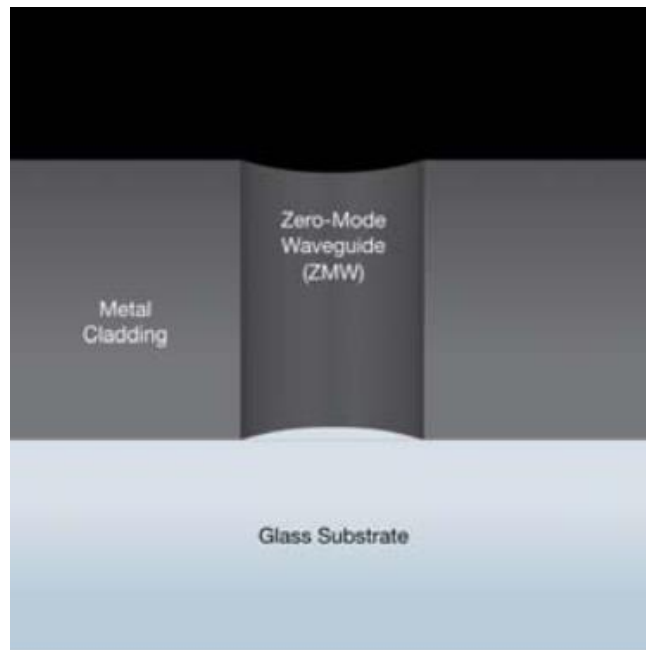
NGS: tecnología PacBio



Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

- Los **orificios** realizados sobre el aluminio **son tan pequeños**, que cuando la **luz** **ingresa sólo penetra los 20-30 nm** de la parte inferior.

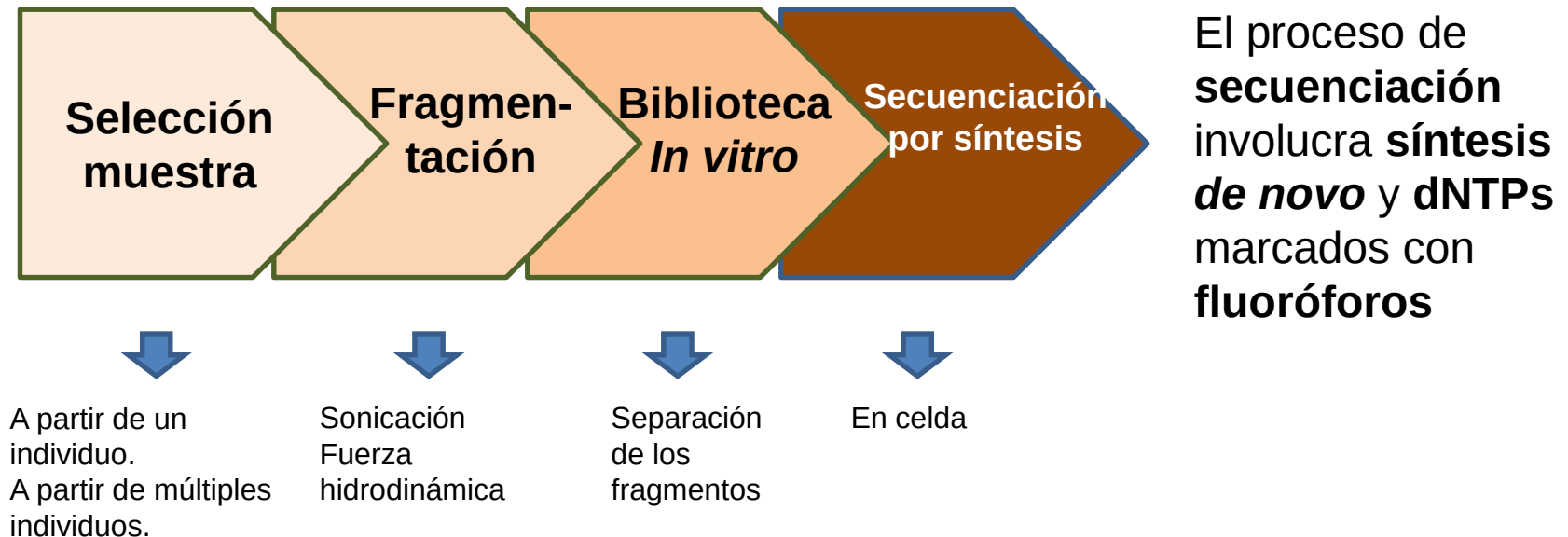


- Esto deriva que el **volumen de detección** sea de **20 zeptolitros**.

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

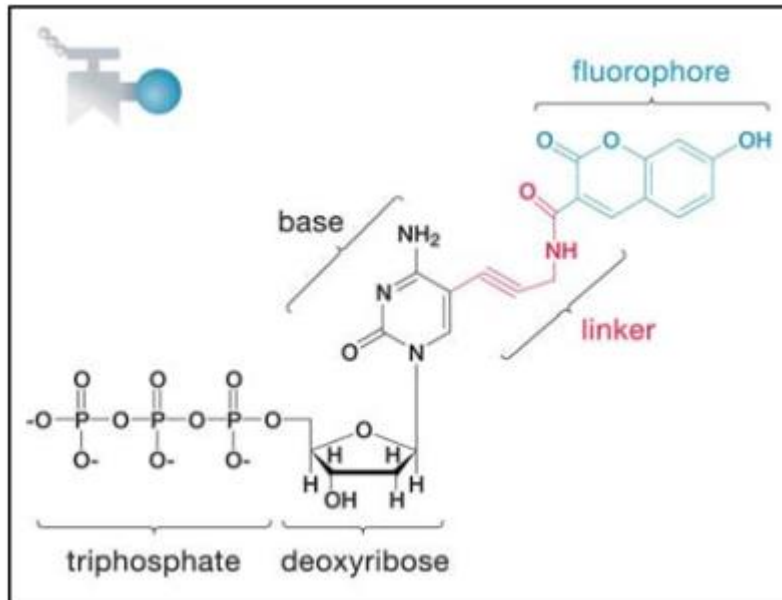
Etapas involucradas en la tecnología *PacBio*



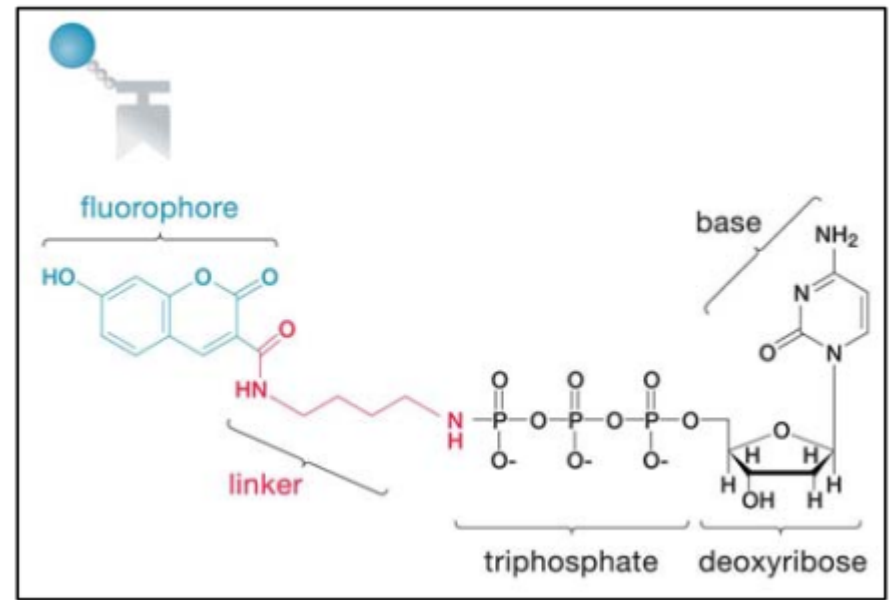
Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

- La **secuenciación comienza** con la **introducción de los 4 dNTPs** marcados con 4 **fluoróforos** distintos. A diferencia del método *Sanger automático* y de *Illumina*, la **marca no se encuentra en la base nitrogenada**.



dNTP marcado en la base

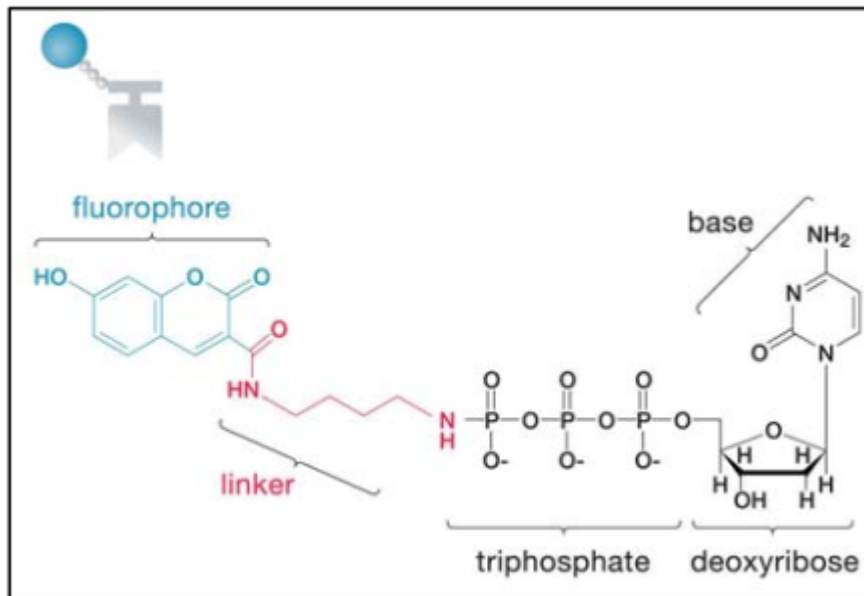


dNTP marcado en los fosfatos

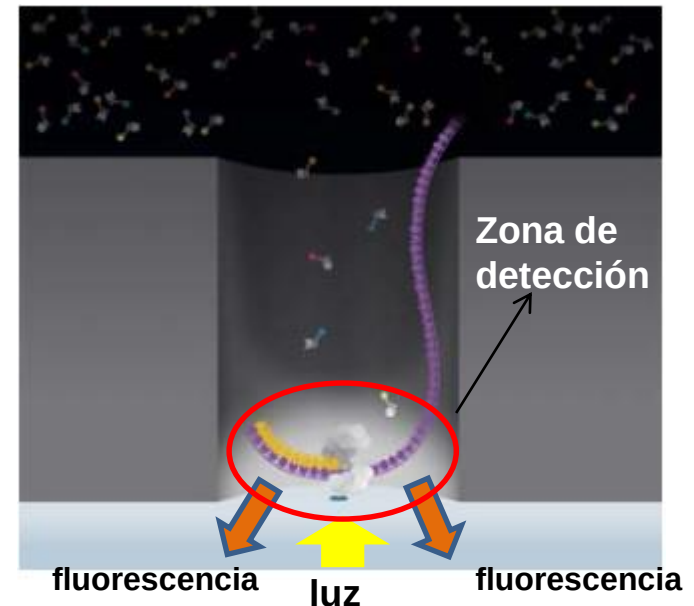
Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

- Esta situación permite que la **fluorescencia no quede en la molécula sintetizada**, dado que se elimina en el pirofosfato.
- De este modo, detectando el dNTP que ingresa al sitio activo de la enzima en el proceso de síntesis, es posible realizar la lectura de la secuencia.



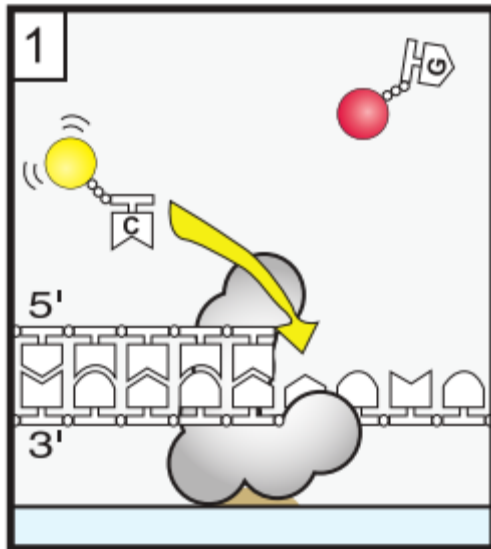
dNTP marcado en los fosfatos



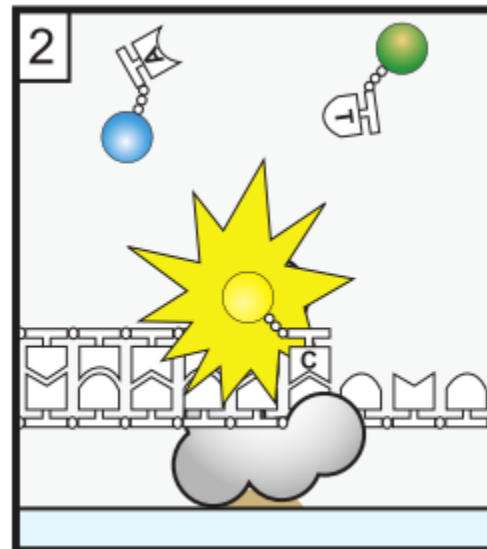
Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

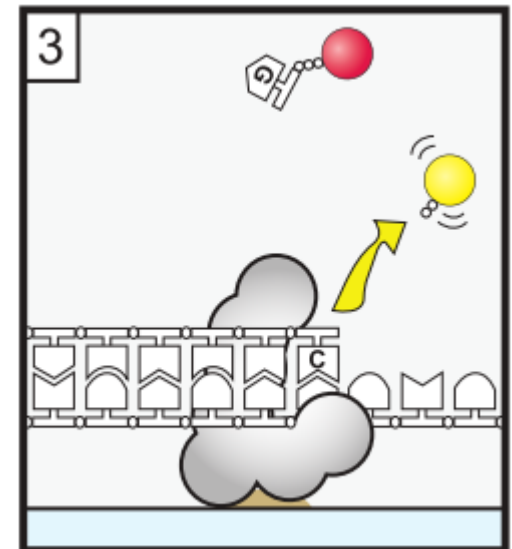
- Durante 10 milisegundos el nucleótido complementario a la cadena molde es retenido por la enzima para su incorporación, tiempo suficiente para que quede en la zona de detección y **emita fluorescencia**, lo cual es **registrado en tiempo real**.



El dNTP
complementario
se acerca



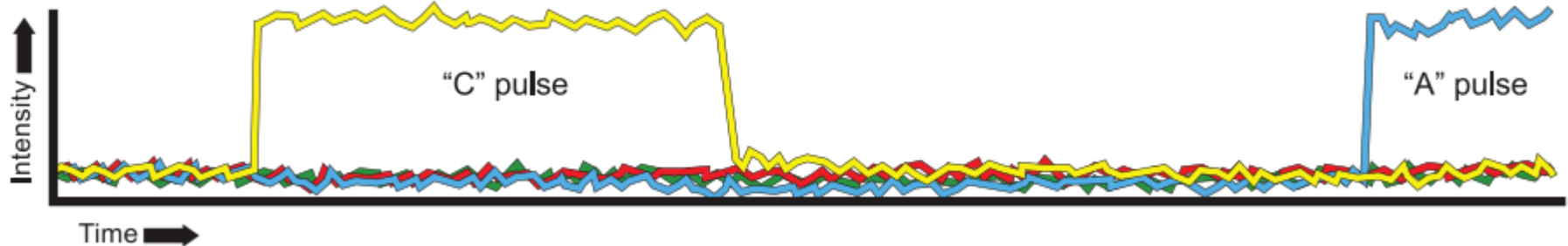
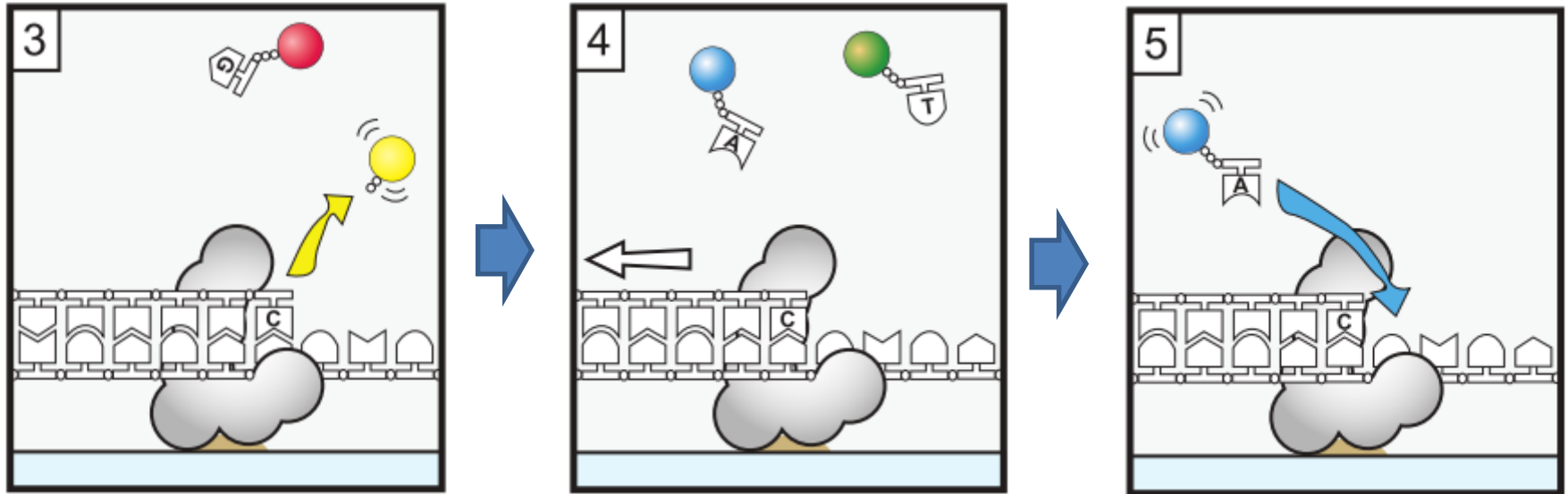
El dNTP
complementario
se incorpora



El pirofasfato
con el fluoróforo
es liberado

Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

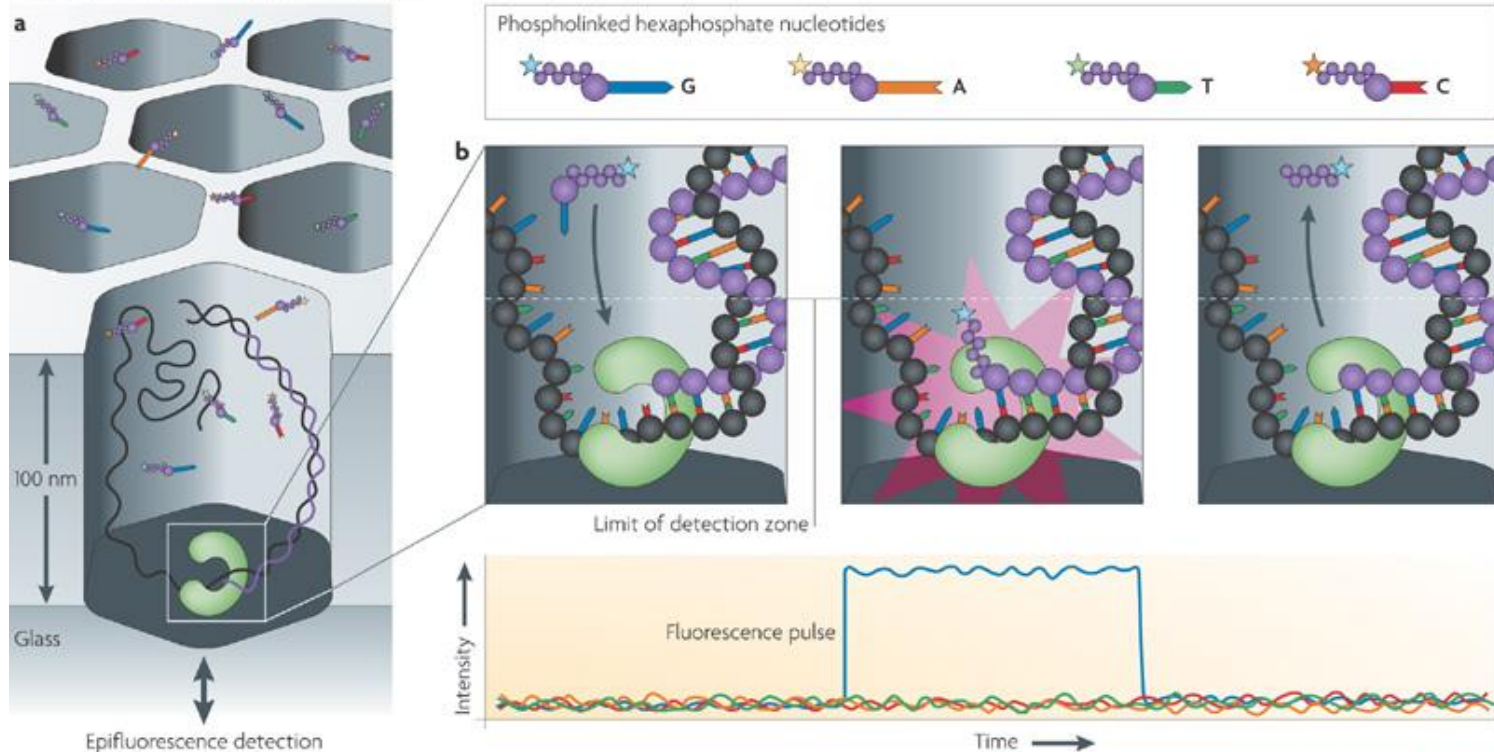


Tecnologías de secuenciación: 3G

NGS: tecnología PacBio

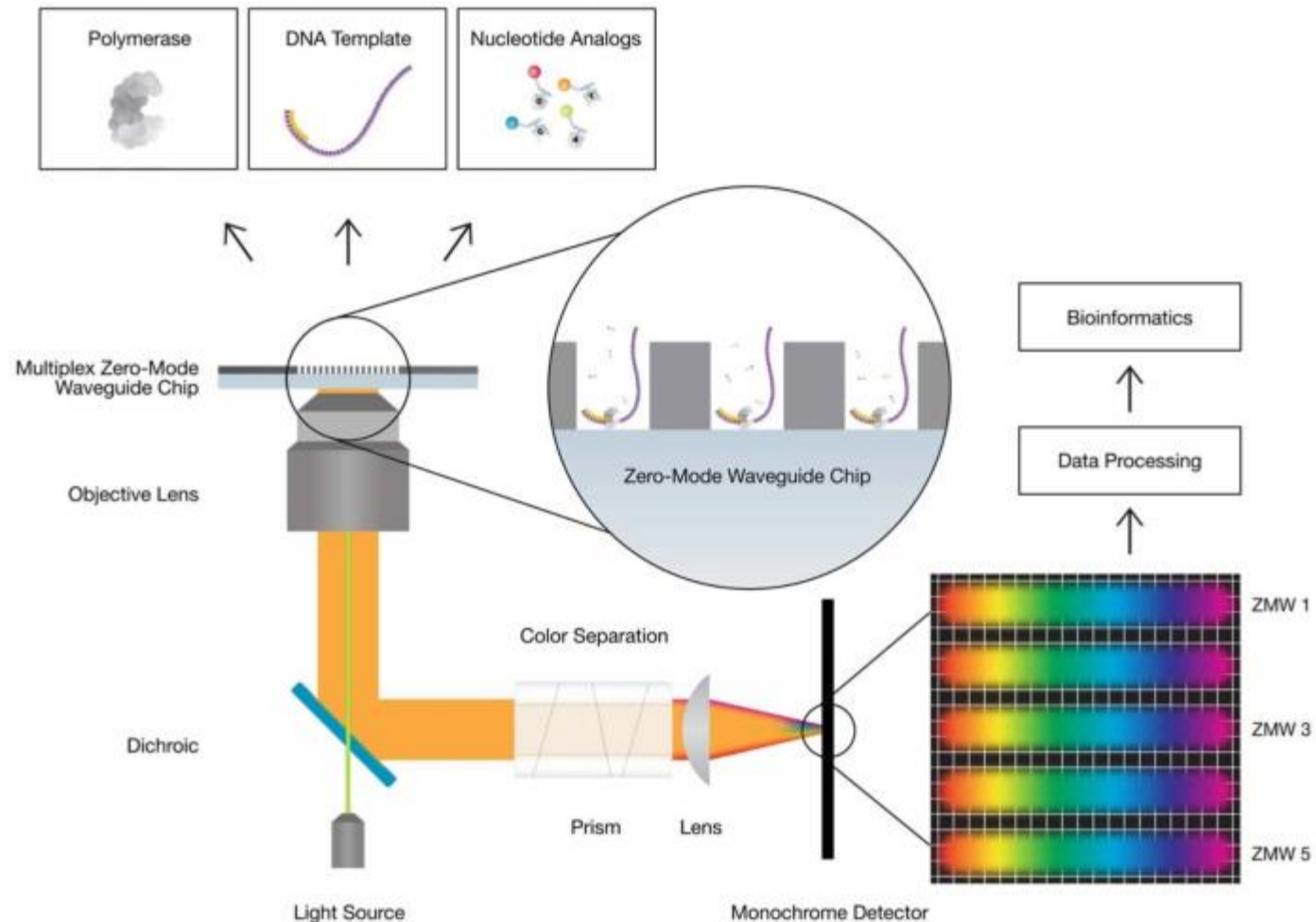
Resumen de la tecnología PacBio

Pacific Biosciences — Real-time sequencing



Tecnologías de secuenciación: 3G

NGS: tecnología PacBio



Descripción de los métodos más utilizados clasificados dentro de las tecnologías NGS

Tecnología Nanopore

Tecnologías de secuenciación: 3G

NGS: tecnología Nanopore

- **Oxford Nanopore Technologies** es una empresa que comercializa un sistema de secuenciación en tiempo real a partir de 2015.
- El **equipo** más representativo se denomina **MinION**, es **portable** y soporta **secuenciaciones de longitud larga** (>10 kb; a mayor escala está el **PromethION**, el **GridION**, y a menor, el **SmidgION**).
- La **tecnología** se basa en **poros** que poseen dimensiones en **escala nanométrica**, los cuales son de **naturaleza proteica** (por ejemplo alpha-hemoslisina).
- Los **nanoporos** son dispuestos en **membranas conductoras** de la electricidad, de modo tal de que cuando son **atravesados** por moléculas de **ssDNA**, en función de sus **nucleótidos**, ocurren **cambios de potencial** que pueden ser detectados.

Tecnologías de secuenciación: 3G

NGS: tecnología Nanopore



MINION



- Posee 2048 nanoporos.

Tecnologías de secuenciación: 3G

NGS: tecnología Nanopore



PromethION



- Cada **celda** poseen **3000** poros.
- Se **estima** que puede producir hasta **6 Tb por día** (2000 genomas humanos, 30X).

Tecnologías de secuenciación: 3G

NGS: tecnología Nanopore

GridION



Tecnologías de secuenciación: 3G

NGS: tecnología Nanopore

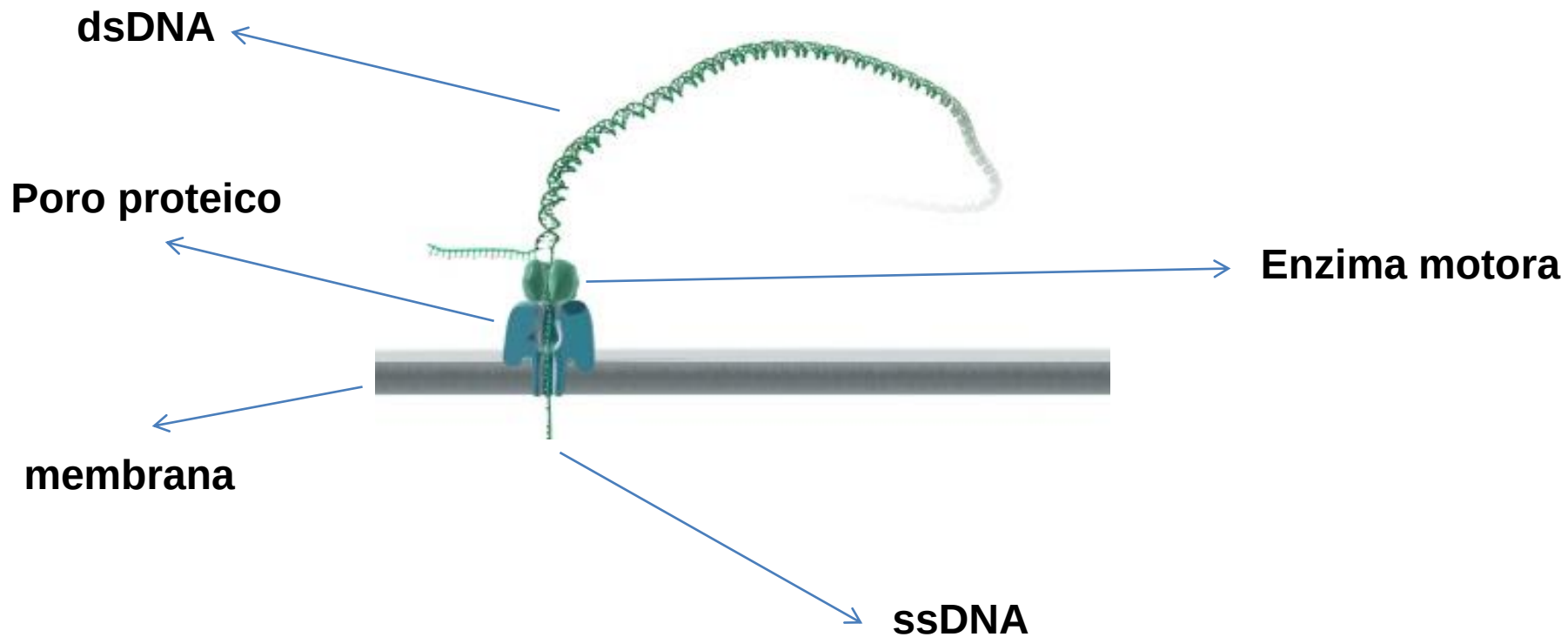


SmidgION



Tecnologías de secuenciación: 3G

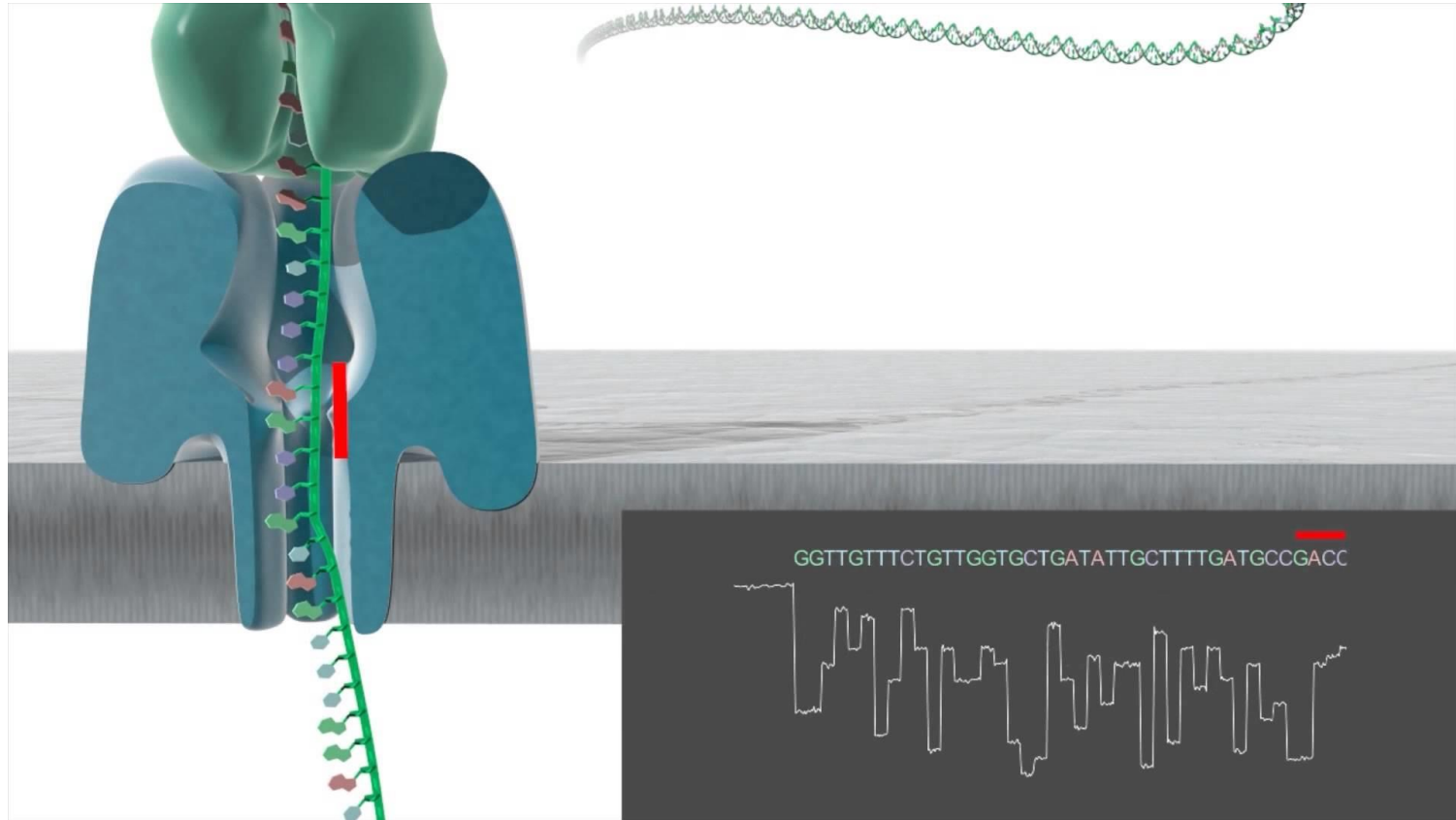
NGS: tecnología Nanopore



2048 nanoporos en un MinION

Tecnologías de secuenciación: 3G

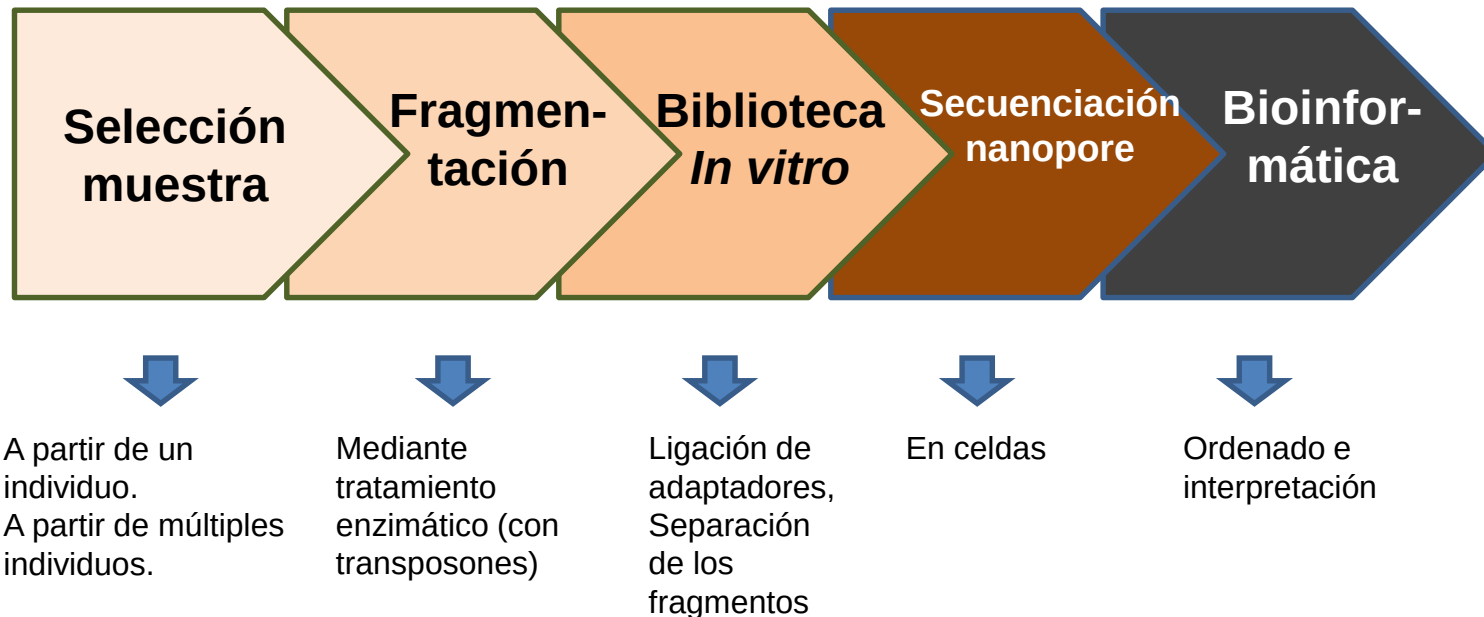
NGS: tecnología Nanopore



Tecnologías de secuenciación: 3G

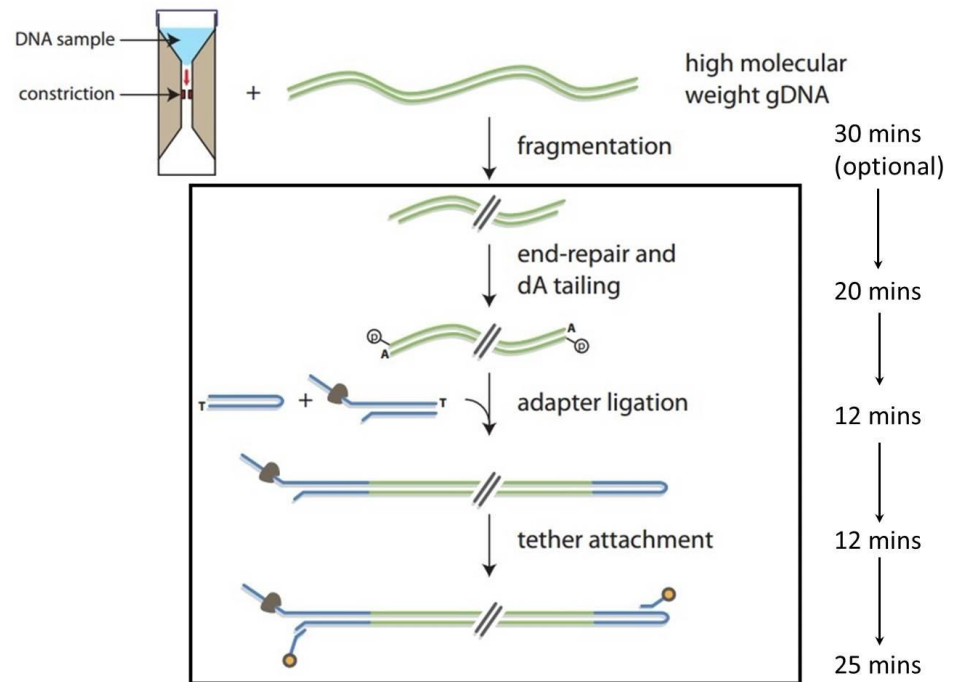
NGS: tecnología Nanopore

Etapas involucradas en la tecnología *PacBio*



Tecnologías de secuenciación: 3G

NGS: tecnología Nanopore



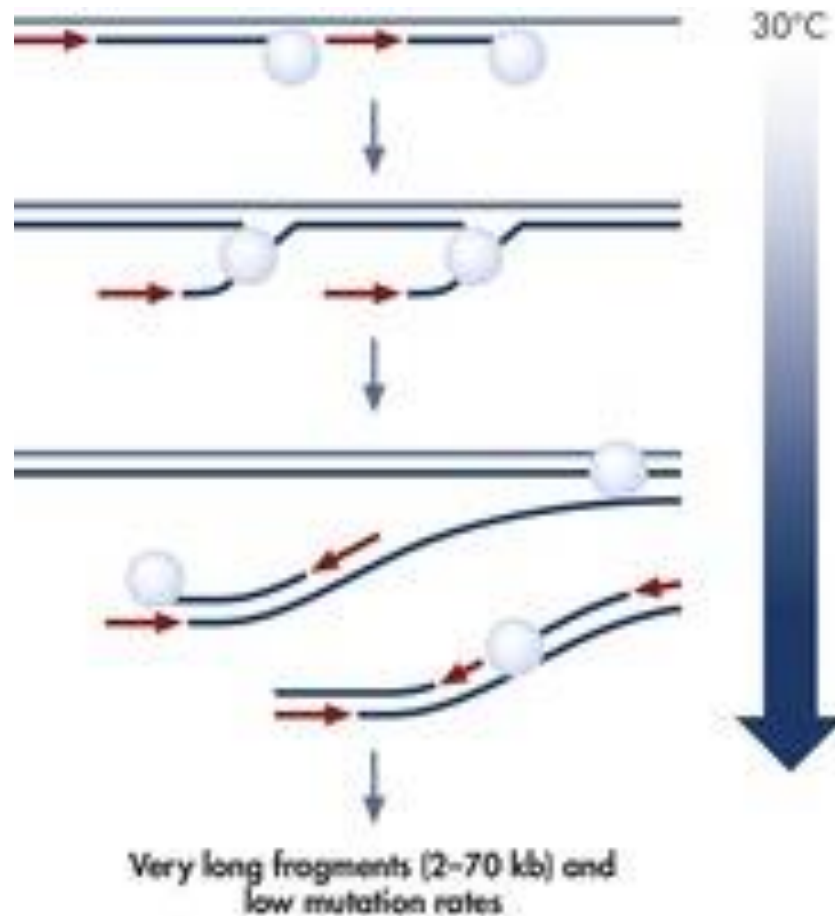
Amplificaciones previas a la secuenciación

Procesos de amplificación previos

- Cuando se parte de **muy poca cantidad de gDNA** es posible realizar una **amplificación previa** a la construcción de la biblioteca.
- Para ello, se suelen utilizar **DNA polimerasas** con **actividad de desplazamiento de cadena** y **hexámeros al azar** como cebadores, para realizar reacciones isotérmicas sin sesgos composicionales.
- Estos **procedimientos** se denominan **WGA** (*Whole Genome Amplification*), y la metodología más utilizada es la conocida como **SDA** (**Strand Displacement Amplification**), reacción isotérmica dirigida por la **DNA polimerasa** del fago $\phi 29$.

Procesos de amplificación previos

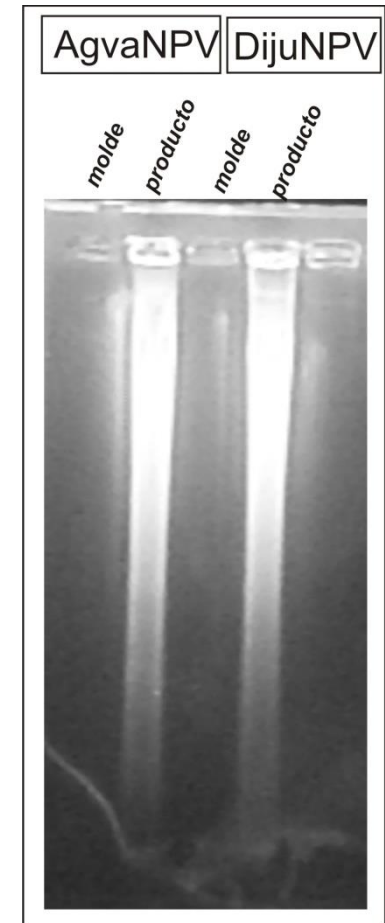
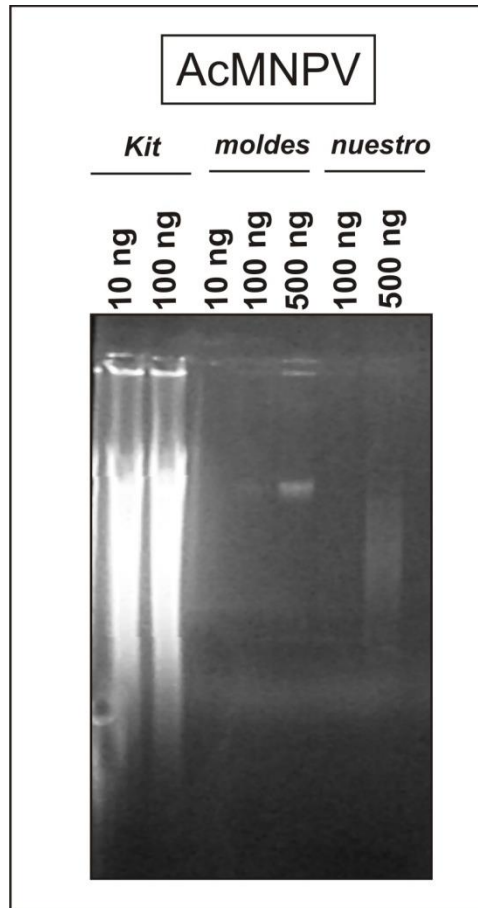
Esquema de la amplificación por SDA



Procesos de amplificación previos

Resultados

- Amplificación *in vitro* de genomas baculovirales como ejemplo de la aplicación de la técnica



Comparación y aplicaciones de los métodos NGS

Adquisición de información genómica

Next Generation Sequencing

Algunas comparaciones de los métodos comerciales...

Table 1. Properties of currently available next-generation sequencing technologies.

Platform	Read length	Throughput/ run	Approximate time/run	Machine cost (US\$)	Reagent cost (US\$)	Reagent cost/Gb (US\$)	Primary error	Base error rates
HiSeq™ 2000	100 bp	600 Gb	11 days	690,000	23,470	40	Substitution	~1–2% over 100 bp
SOLiD™ 4	75 bp	100 Gb	12 days	475,000	8128	<110	A–T bias	0.06%
SOLiD™ 4hq	75 bp	300 Gb	14 days	595,000	10,503	70	A–T bias	0.01%
SOLiD™ PI	75 bp	77 Gb	8 days	349,000	6101	80	A–T bias	0.01%
454 GS FLX Titanium XL+	700 mean bp, ≤1000 bp	700 Mb	23 h	500,000	6200	7000	Indel	0.5%
IonTorrent™ PGM™ 316	200 bp	100 Mb	~2 h	50,000	750	<7500	Indel	1.2% over 150 bp
IonTorrent™ PGM™ 318	200 bp	1 Gb	~2 h	50,000	925	<925	Indel	1.2% over 150 bp
MiSeq®	150 bp	>1 Gb	27 h	125,000	750	740	Substitution	~1–2% over 100 bp
454 GS Junior	400 mean bp	35 Mb	12 h	108,000	1100	22,000	Indel	1.00%
PacBio RS (early 2012)	2700 mean bp, ≤5000 bp	90 Mb per cell	<1 day (?)	695,000	110–1700	11,000– 340,000	CG deletion	13.00%

Currently available next-generation sequencing technologies. All specifications are liable to rapid change. Cost information is taken from Glenn [18], but can vary widely by country and individual deal. The read lengths given are taken from manufacturers websites [101–105], and are the maximum over all currently available protocols with the system in question. The same holds for throughput, and times are the shortest for the given throughput. Base error rate, from the same sources, should be taken as a rough indicator only as the details of errors vary. It is an average for the read, or for a fixed number of bases from the start of the read (which is stated in this case).

Adquisición de información genómica

Next Generation Sequencing

Algunas comparaciones de los métodos comerciales...

method	Pac Bio	Ion torrent	Pyroseq	Seq by synthesis	Seq by ligation	Sanger
Company	Pac Bio	Life tech	Roche	Illumina	Applied bio	
Read length	2900 bp	200 bp	700 bp	150 bp	50 bp	800 bp
Reads/run	$75 * 10^5$	$5 * 10^6$	$1 * 10^6$	$3 * 10^9$	$1,2 * 10^9$	1
Time /run	2 hours	2 hours	24 hours	10 days	14 days	2 hours
\$ $1*10^6$ bases	\$2	\$1	\$10	\$0.1	\$0.13	\$2400

Adquisición de información genómica

Next Generation Sequencing: Aplicaciones

