

Computational biology; towards understanding of complex biological systems

Javier Correa Alvarez, PhD.
Sergio Pulido Tamayo cPhD.
2015

“To use this flood of [sequence] knowledge, which will pour across the computer networks of the world, biologists not only must become computer literate, but also change their approach to the problem of understanding life” [Walter Gilbert](#), 1994

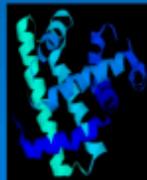
“I spent many hours persuading people that computing was not only going to be the essential tool for biological research but would also provide models for analyzing complexity...The development of sequencing techniques and their widespread application has generated enormous databases of information, and the need for computers is no longer questioned” [Sydney Brenner](#)

What Computation Has Done for Biology

- Numerous bioinformatics tools have been developed to derive functional/structural/interactional information from *omic* data



Gene finding, function prediction



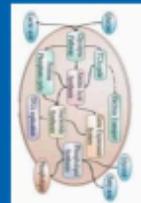
Structure prediction, modeling



Molecular Interaction prediction



Biological network prediction, modeling



Cell-level modeling, simulation



Mechanical stress to cell signaling modeling



Modeling of ecosystems

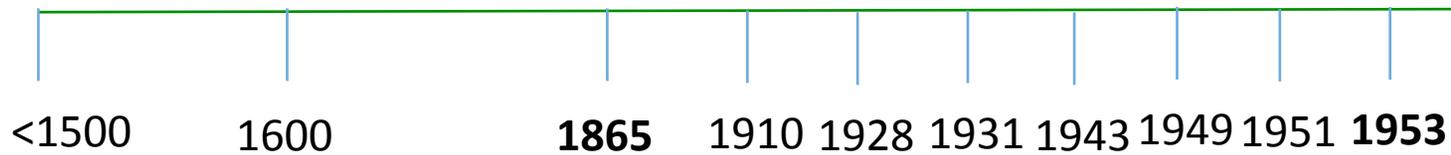
Principales razones para formarse en biología computacional/bioinformática

0. La computación es la clave para la biología del siglo XXI
1. Las habilidades computacionales son altamente transferibles. Sólo el 14% de los PhD en biología la posee, mayoría en posiciones académicas.
2. Publicará mucho más y con mayor impacto.
3. Podrá ser más flexible en sus prácticas de trabajo.
4. Mejorará el costo beneficio de sus investigaciones.
5. Usted entenderá porque las listas deben empezar en 0

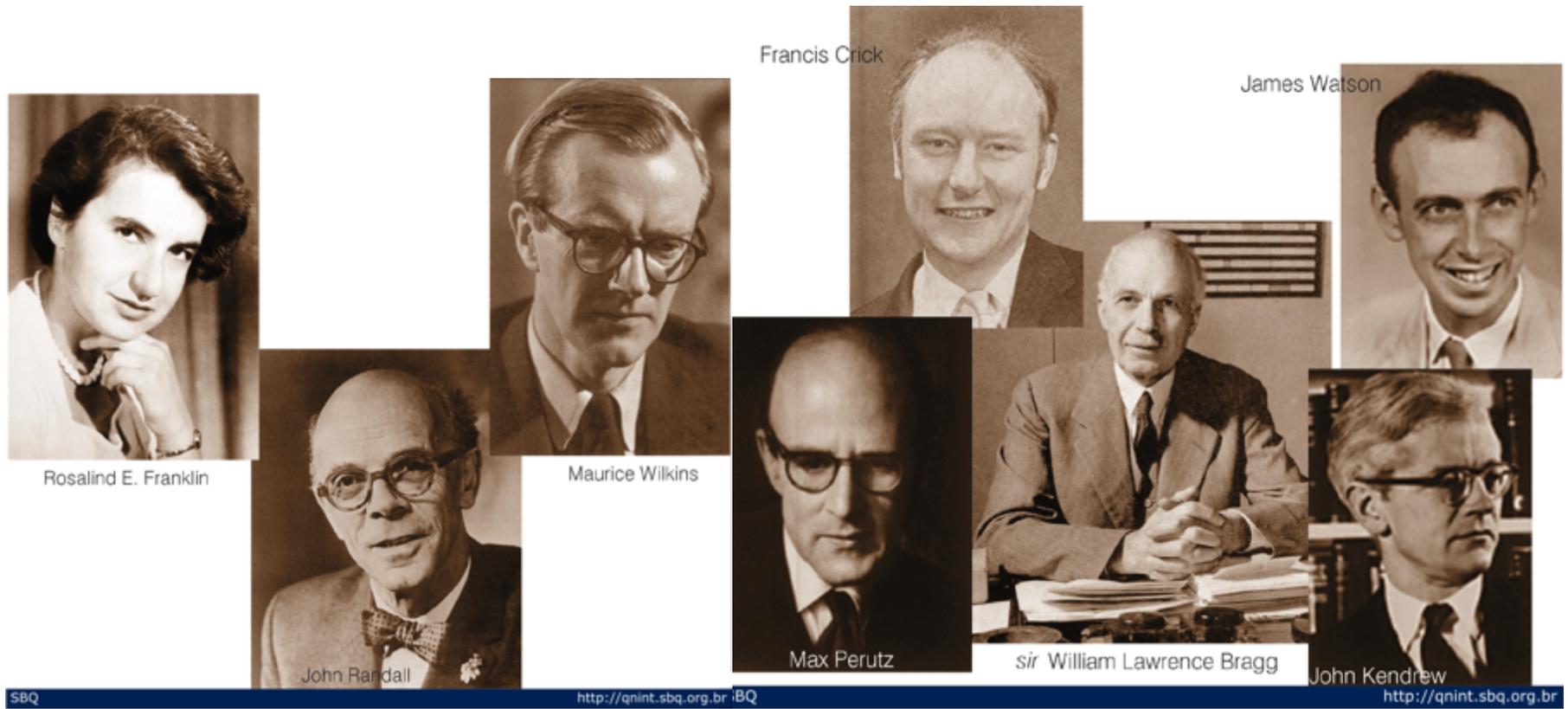
Breve historia de las bases moleculares de la vida

Genética Clásica

Genética Molecular



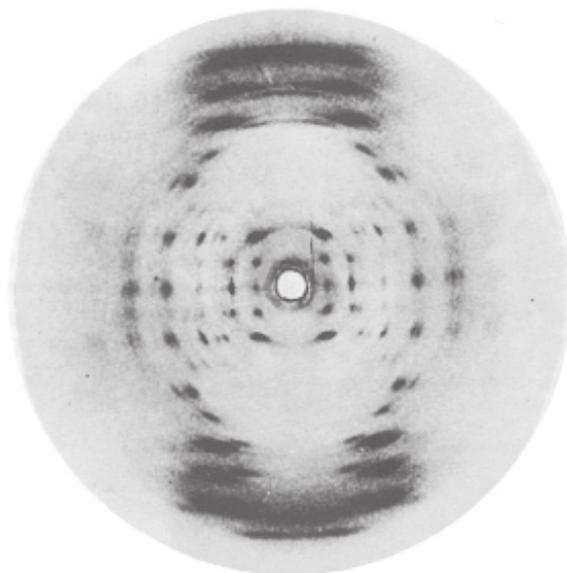
Estructura del DNA



SBQ <http://qnint.sbq.org.br> iBQ <http://qnint.sbq.org.br>

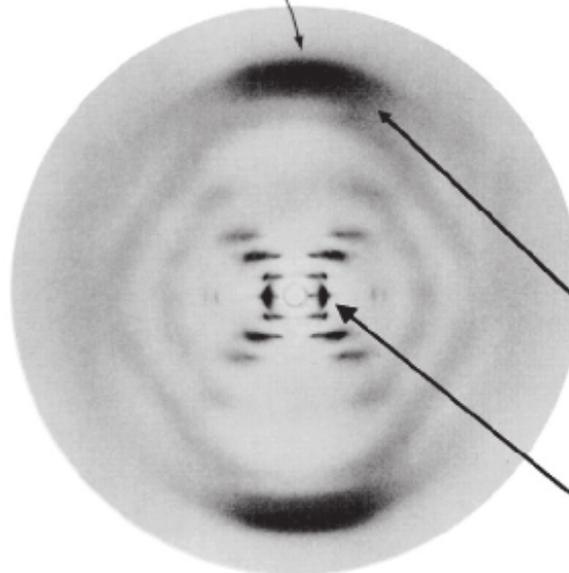
Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]



DNA tipo "A"

espaçamento de 3,4 Å

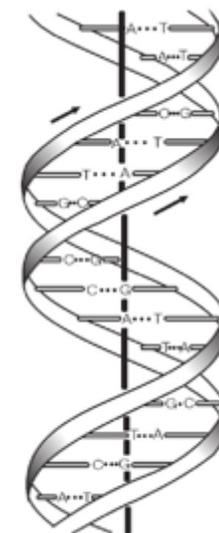
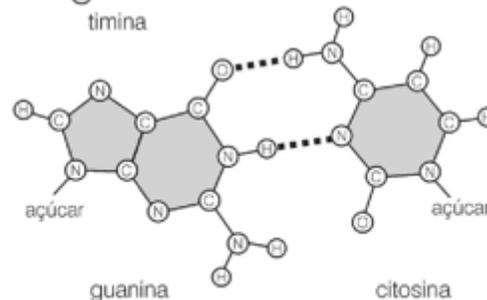
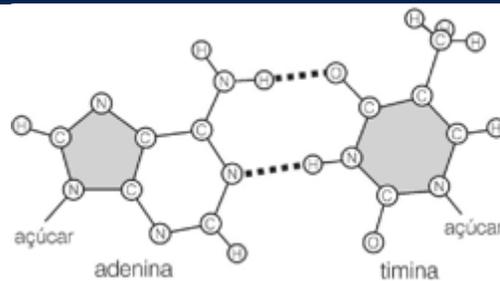


DNA tipo "B"

DNA tipo "B": forma hidratada e pouco cristalina; a foto com um "X" é a assinatura de estruturas em hélices.

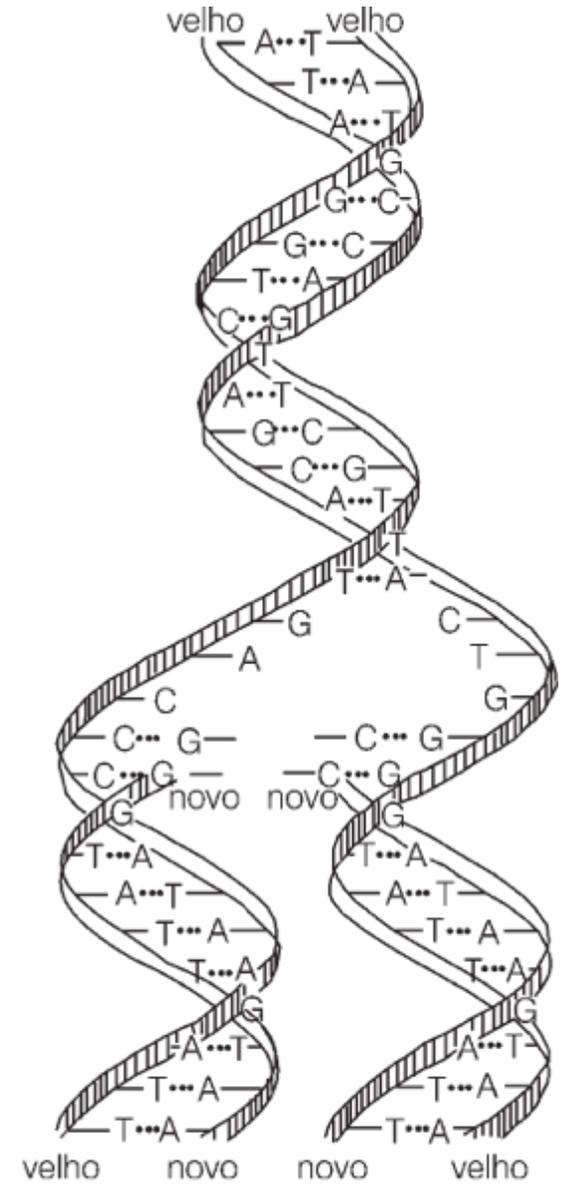
As reflexões do "X" indicam que a periodicidade da hélice é de ~ 34 Å

A reflexão axial indica diâmetro da hélice da ordem de 20 Å





SBQ

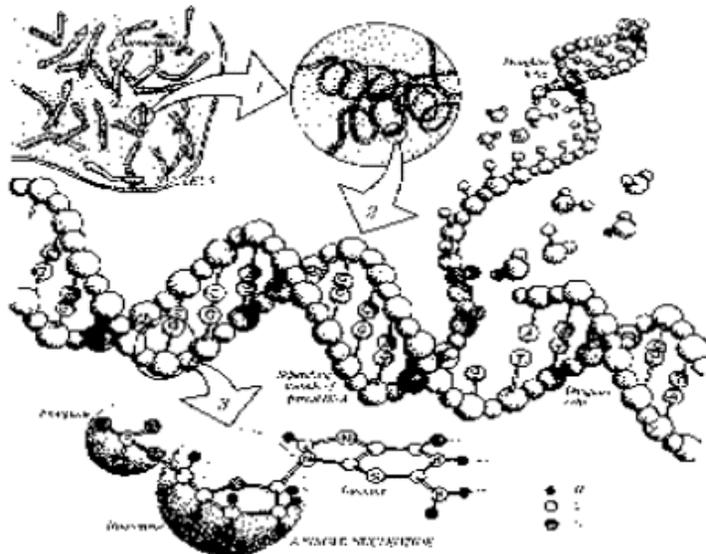


<http://qnint.sbq.org.br>

Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]

Genoma

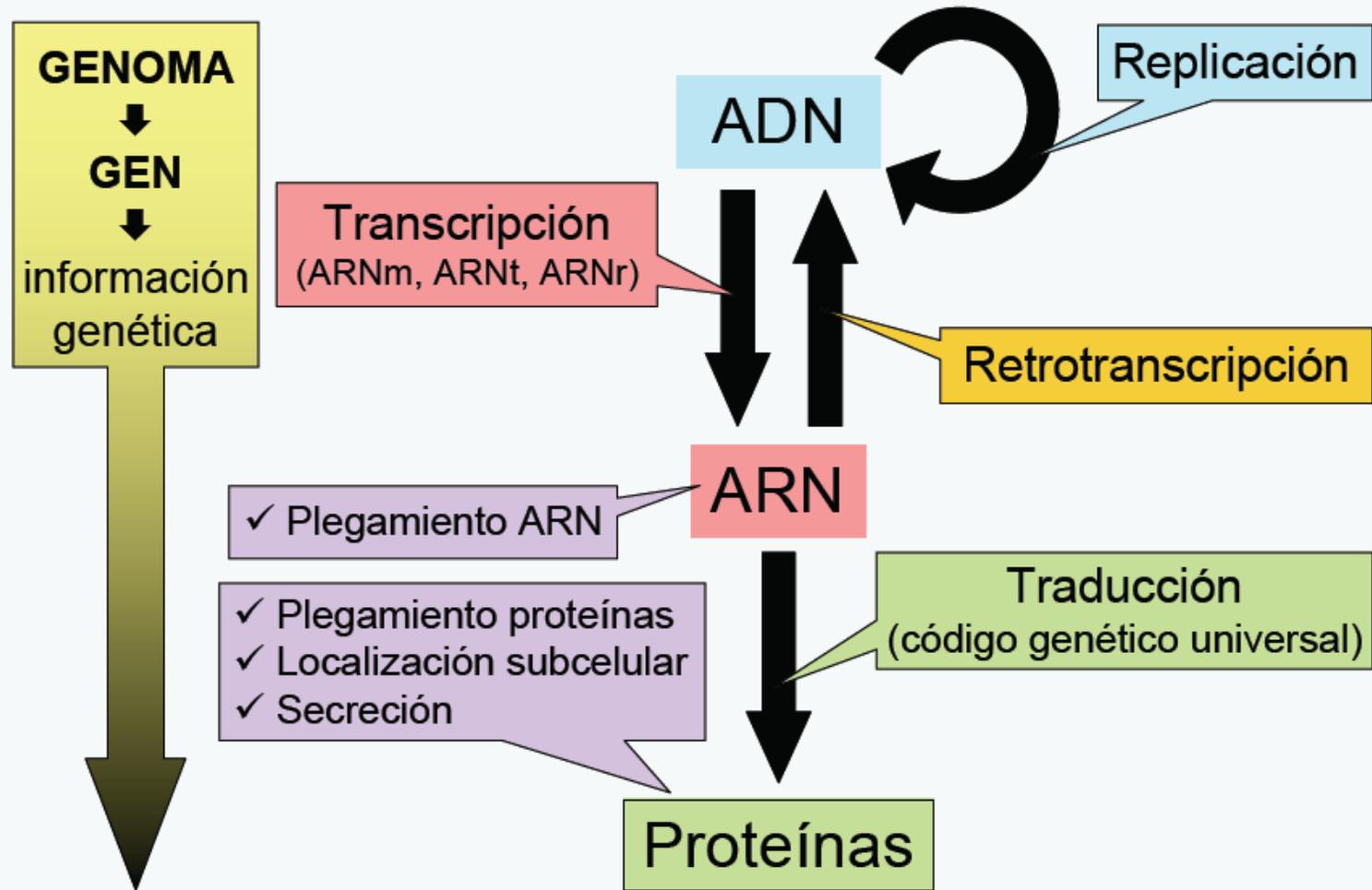


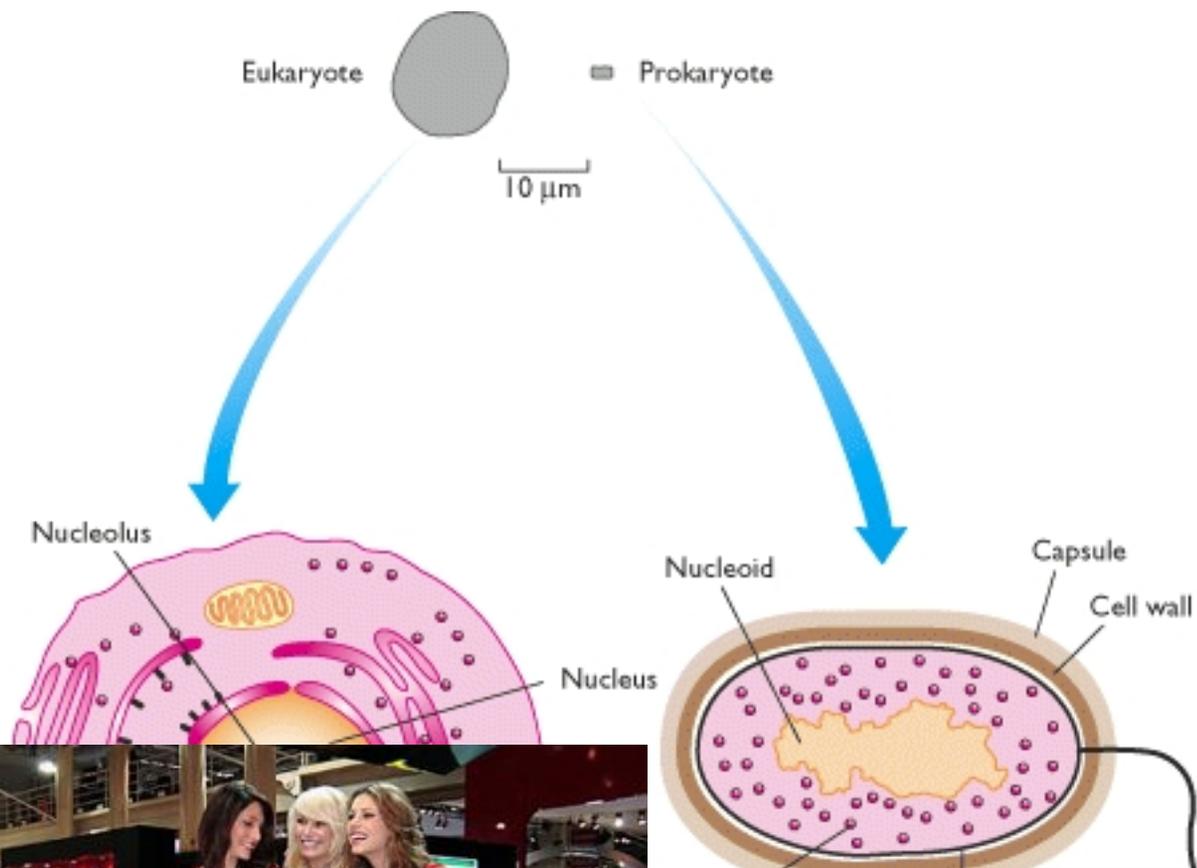
- **Genoma:**

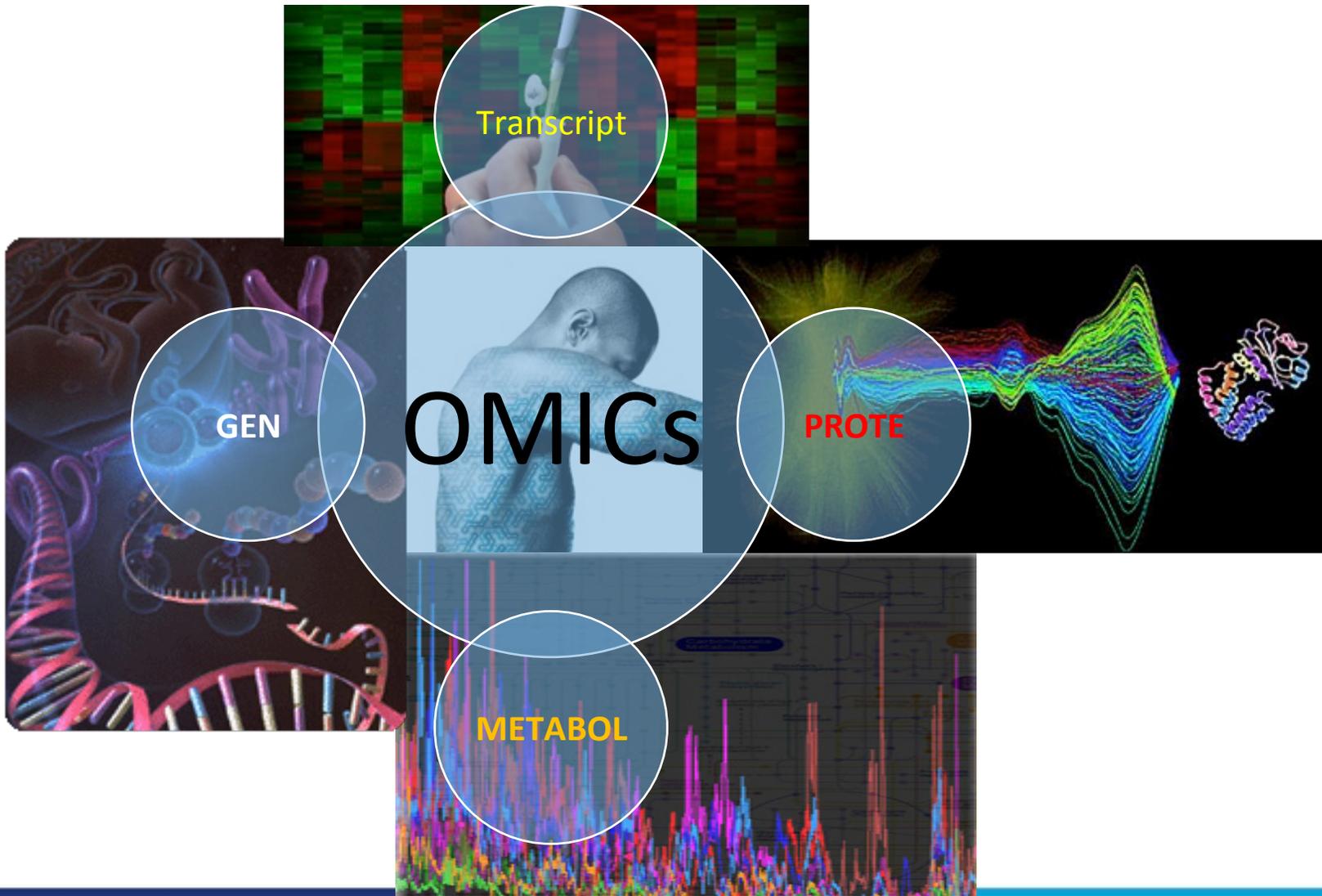
- La información hereditaria de un organismo codificada en su ADN y encerrada en una sola célula. La información contenida en el ADN de un solo organismo es su genoma.
- La molécula de ADN puede ser vista como una secuencia muy larga de nucleótidos o bases.

$$\Sigma = \{A, T, C, G\}$$

Dogma Central de la Biología Molecular





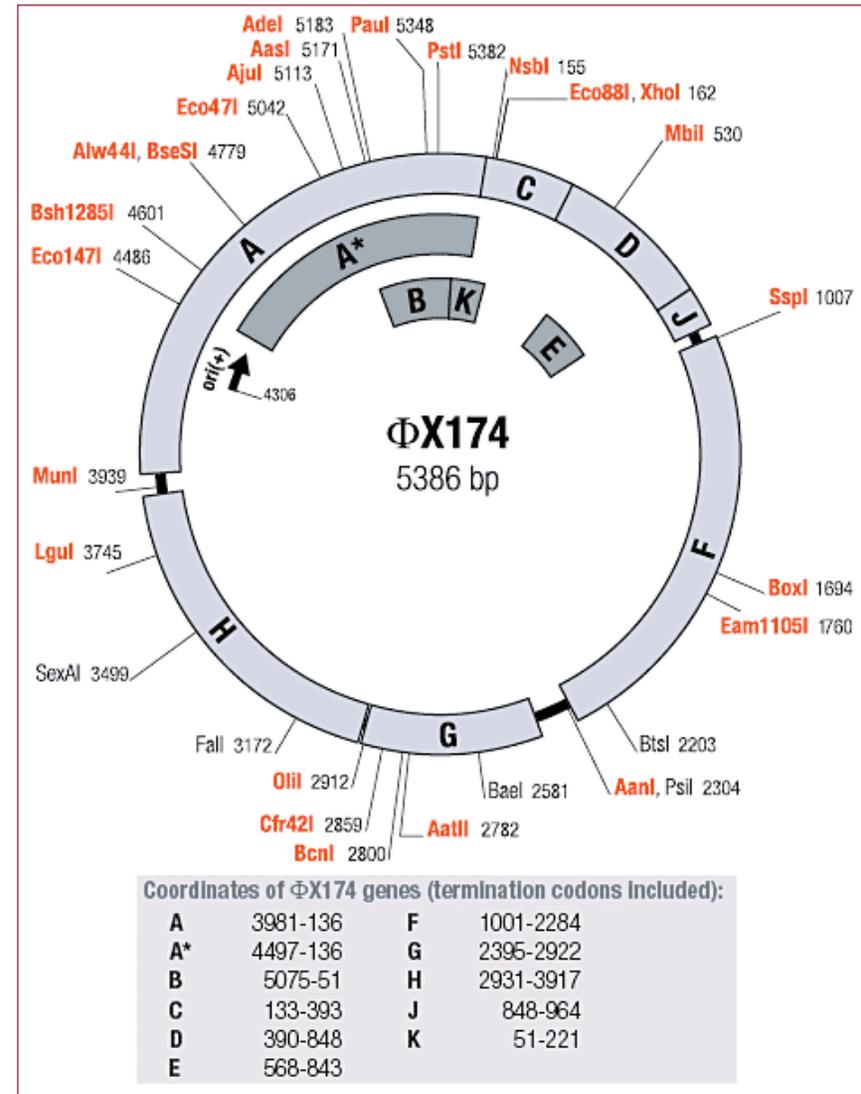
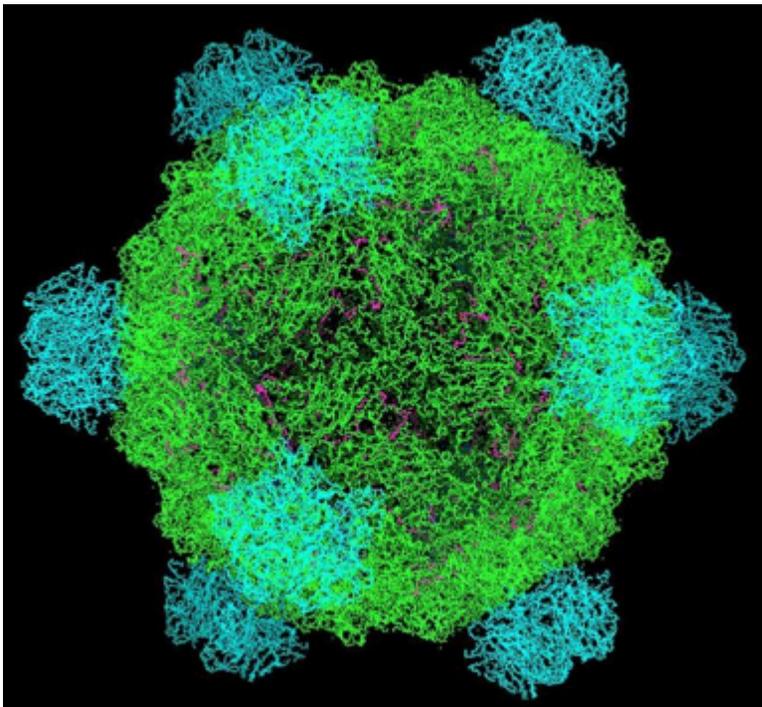


Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]

- **Fred Sanger 1977**
Phi X 174 (5386 nt, 11 prot)

- 2003 – Craig Venter
Primer genoma sintético

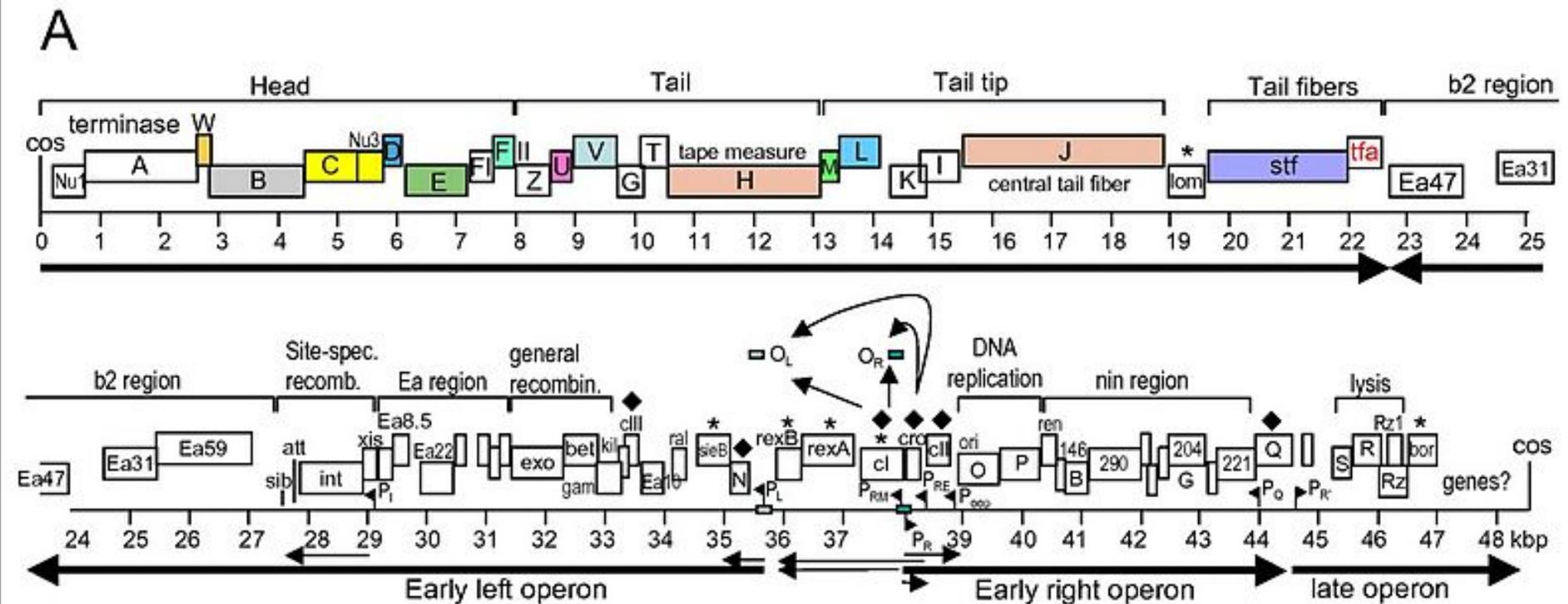


- **Sanger 1982**

Enterobacteria phage λ – Esther Lederberg 1950

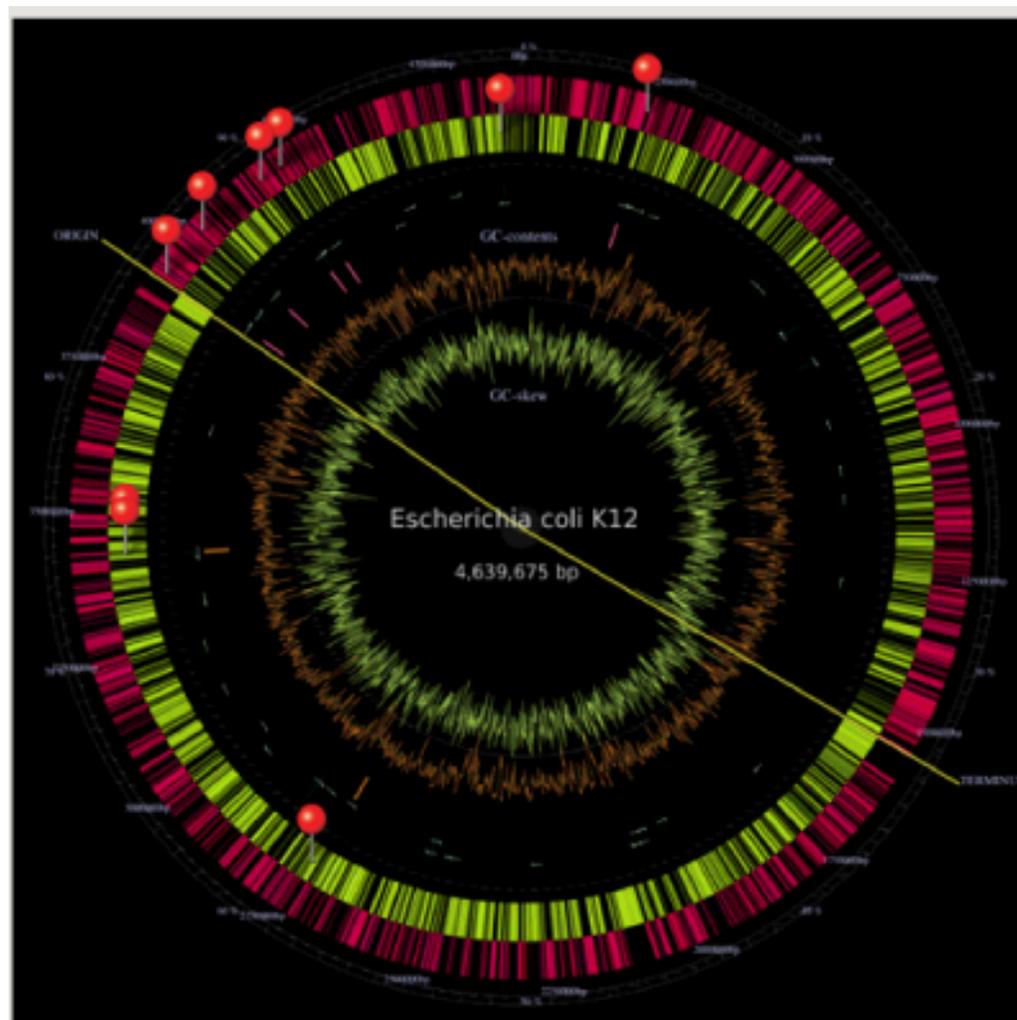
48 Kb – 20 genes, 12 -14 proteínas

Figure 1.





E. Coli – Tamaño 1-2 um



Genoma 1997 – 4331 genes

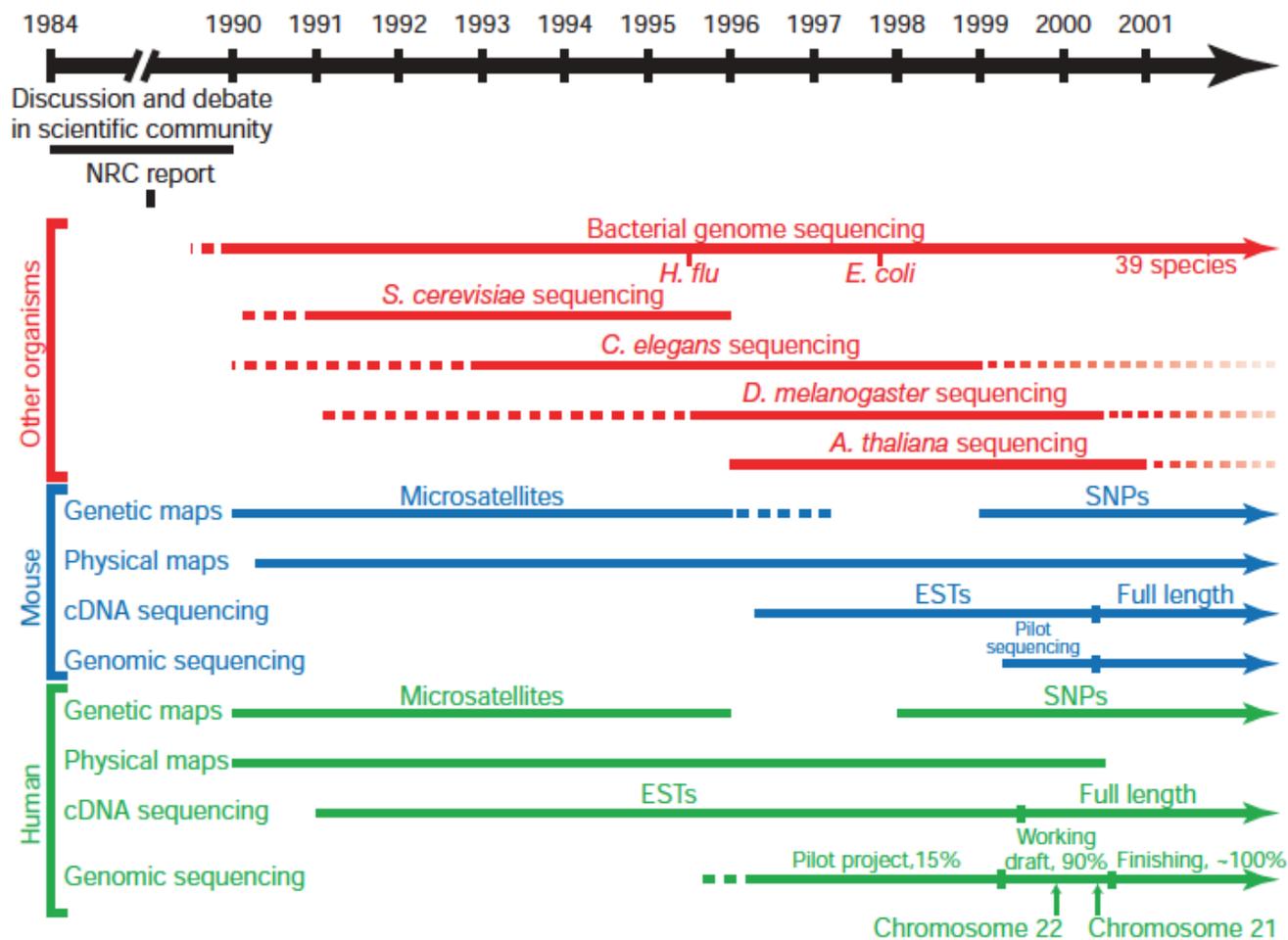
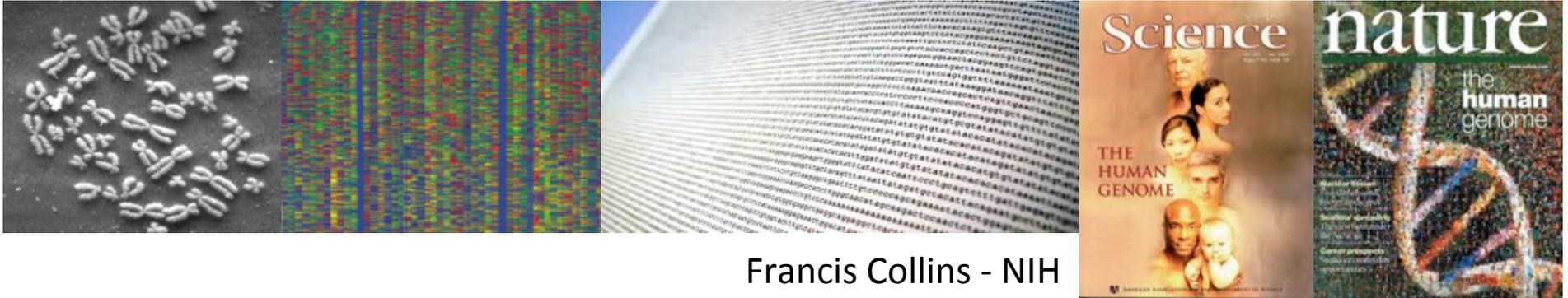


Figure 1 Timeline of large-scale genomic analyses. Shown are selected components of (green) from 1990; earlier projects are described in the text. SNPs, single nucleotide



Francis Collins - NIH

J. Craig Venter – Celera Genomics

Proyecto Genoma Humano – (1986 - 2001)

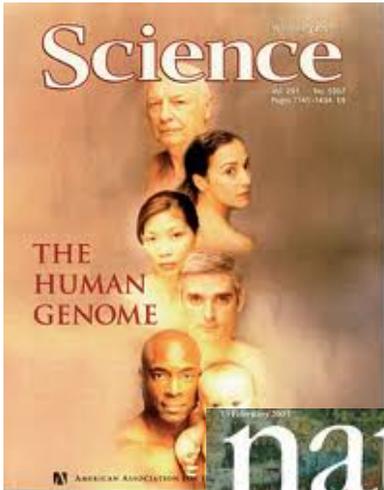
- ¿Qué tan grande es el genoma humano?
3 billones de letras – 1000 libros de 500 paginas, un DVD de capacidad.
Secuenciar 1 base = 1 US\$. Quien lo pagaría?? NIH (USA) 25%
- ¿Qué se ganaría con saber el genoma?
Saber que hacen y susbtibilidades en genes (~100.000)
Curar enfermedades congénitas y el Cancer
Predecir enfermedades que aún no se conocen (la clave para el futuro del hombre)
Conocer nuestro pasado – Cómo evolucionamos hace 5 millones de años

Genomas eucarióticos comparados

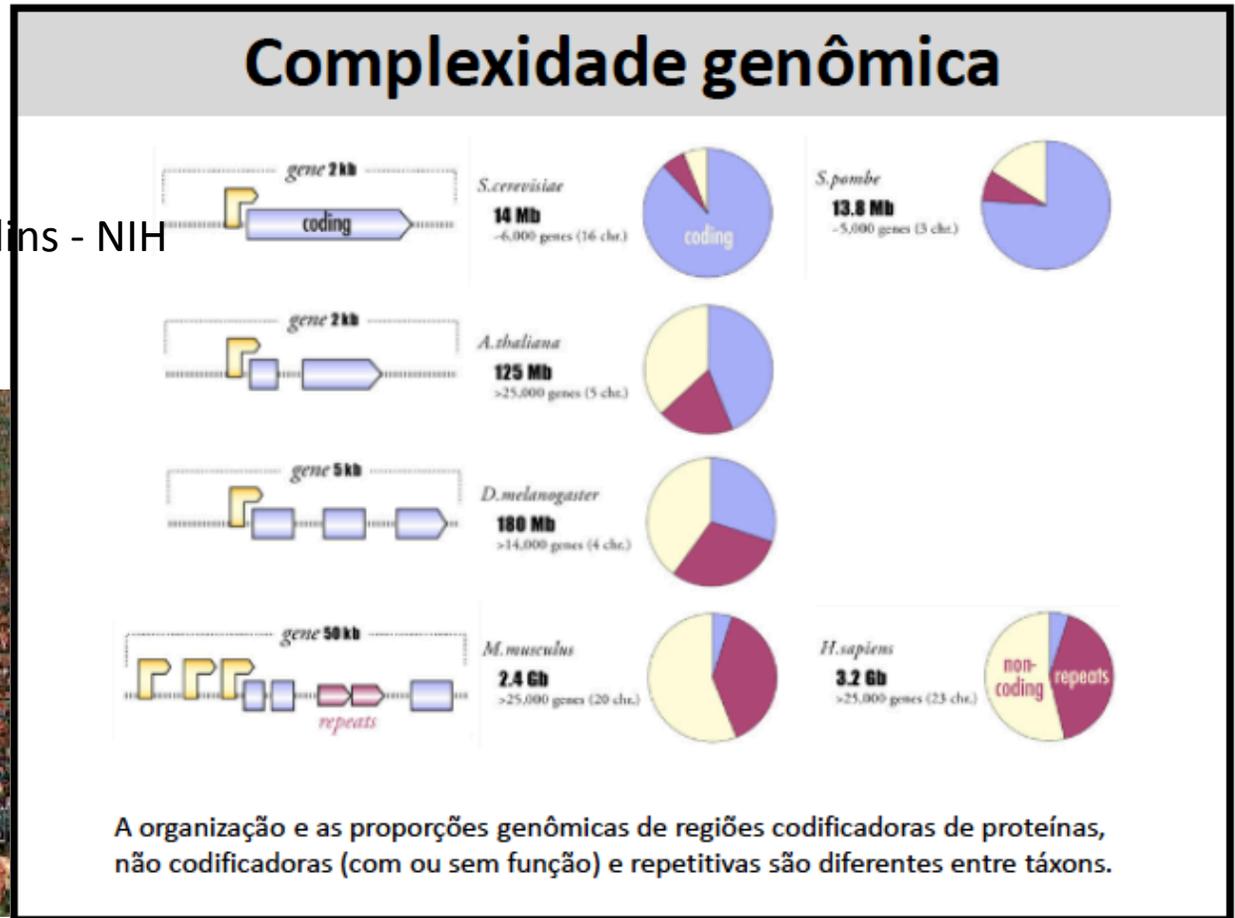


Homem vs..	Chimpanzé	Camundongo	Gambá	Baiacu
Tamanho (Gpb)	3,0	2,5	4,2	0,4
Tempo desde a divergência	~5-7 MAA	> 75 MAA	> 150 MAA	~450 MAA
Conservação em regiões codificadoras	>99%	~80%	~70-75%	~65%
Permite identificar:	Mudanças genômicas e/ou adaptativas exclusivamente humanas	Sequências codificadoras e não codificadoras	Sequências codificadoras e não codificadoras	Sequências codificadoras

Y qué nos dicen los genomas???



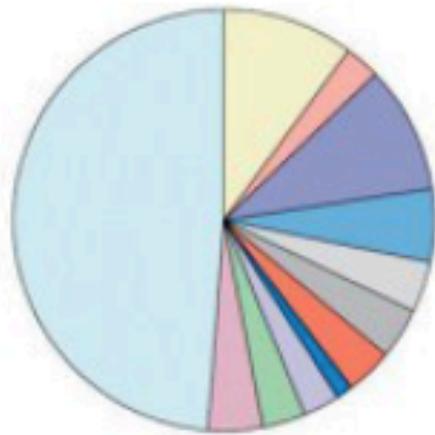
Francis Collins - NIH



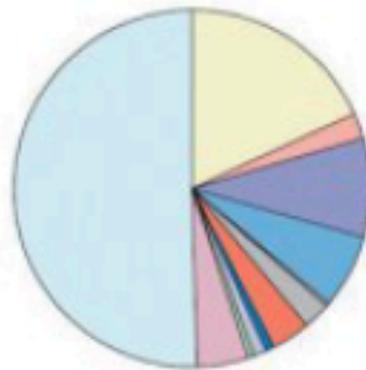
J. Craig Venter – Celera Genomics

Inspira Crea Transforma

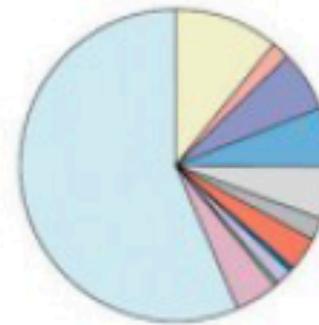
UNIVERSIDAD
EAFIT



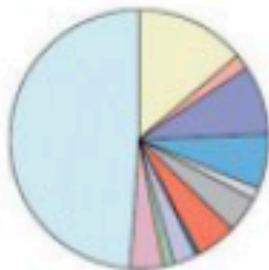
Organism Human
Genes ~32,000



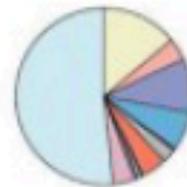
Organism *Arabidopsis* (plant)
Genes 25,706



Organism *C. elegans* (roundworm)
Genes 18,266



Organism *Drosophila* (fly)
Genes 13,338



Organism *Saccharomyces* (yeast)
Genes ~6000

Grande parte dos genes eucarióticos não tem função conhecida (*unknown*)

- | | | |
|------------------------------|---------------------------------|------------------------|
| Metabolism | Cell-cell communication | Cytoskeleton/structure |
| DNA replication/modification | Protein folding and degradation | Defense and immunity |
| Transcription/translation | Transport | Miscellaneous function |
| Intracellular signaling | Multifunctional proteins | Unknown |

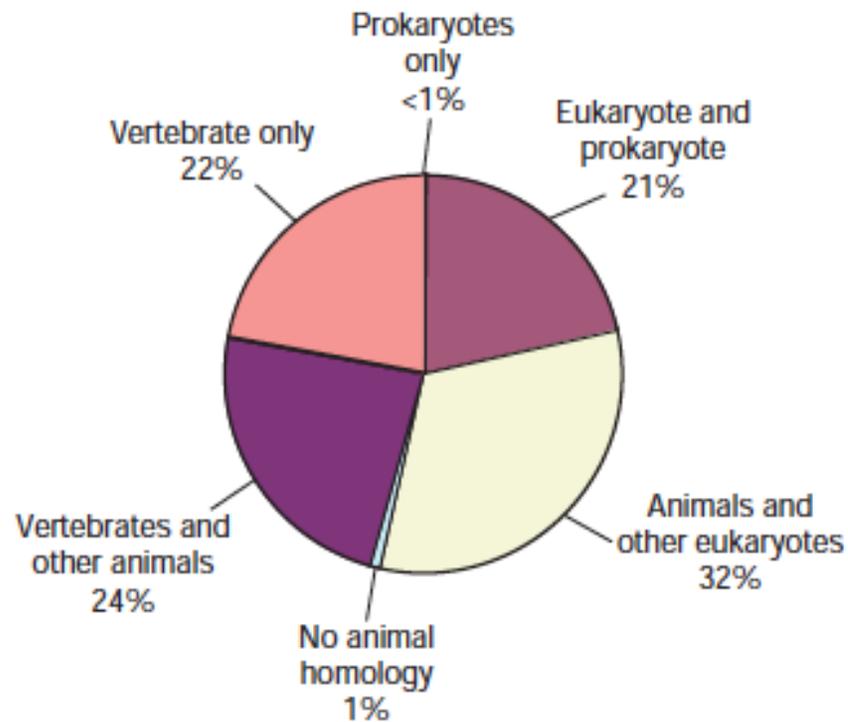


Figure 38 Distribution of the homologues of the predicted human proteins. For each protein, a homologue to a phylogenetic lineage was considered present if a search of the NCBI nonredundant protein sequence database, using the gapped BLASTP program, gave a random expectation (E) value of ≤ 0.001 . Additional searches for probable homologues with lower sequence conservation were performed using the PSI-BLAST program, run for three iterations using the same cut-off for inclusion of sequences into the profile³²⁸.

Homología de los genes

- Estructura típica a los vertebrados, sin novedad evolutiva.
- 1% genes no conocidos. De la especie??

Y todo ese DNA “basura”, no codificante para que sirve???

Funciones estructurales y reguladoras:

- RNAs funcionales: miRNA, snRNA, RNAi, etc..
- Regiones Conservadas no codificadoras: Intrones, Promotores, Enhancers...
- Secuencias altamente repetitivas (STR) 100 pb asociadas con heterocromatina.
- Secuencias Moderadamente repetitivas tRNA, rRNA...
- Mayor parte son: Elementos Transponibles
 - LINE (elementos retrotransponibles via RNA. Largos)
 - SINE (Retrotransposones cortos)
 - ERV (retrovirus endógenos)
 - Transposones (Elementos transponibles via DNA)

Tecnologías de primeira generación



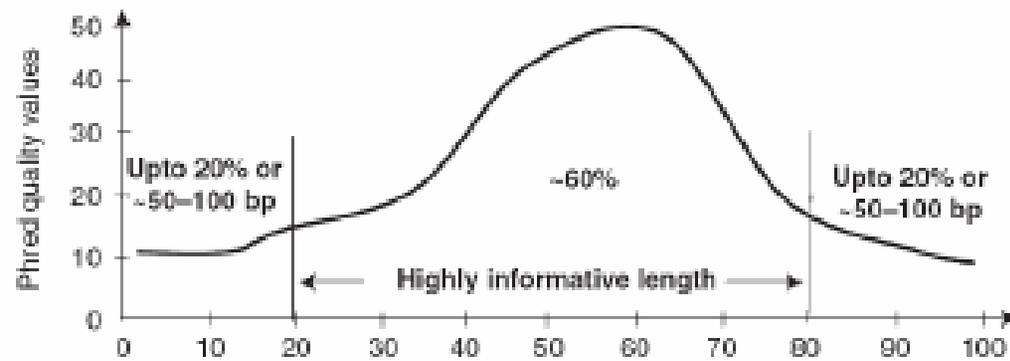
Frederick Sanger
Prêmio Nobel em Química
(1980)



Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]

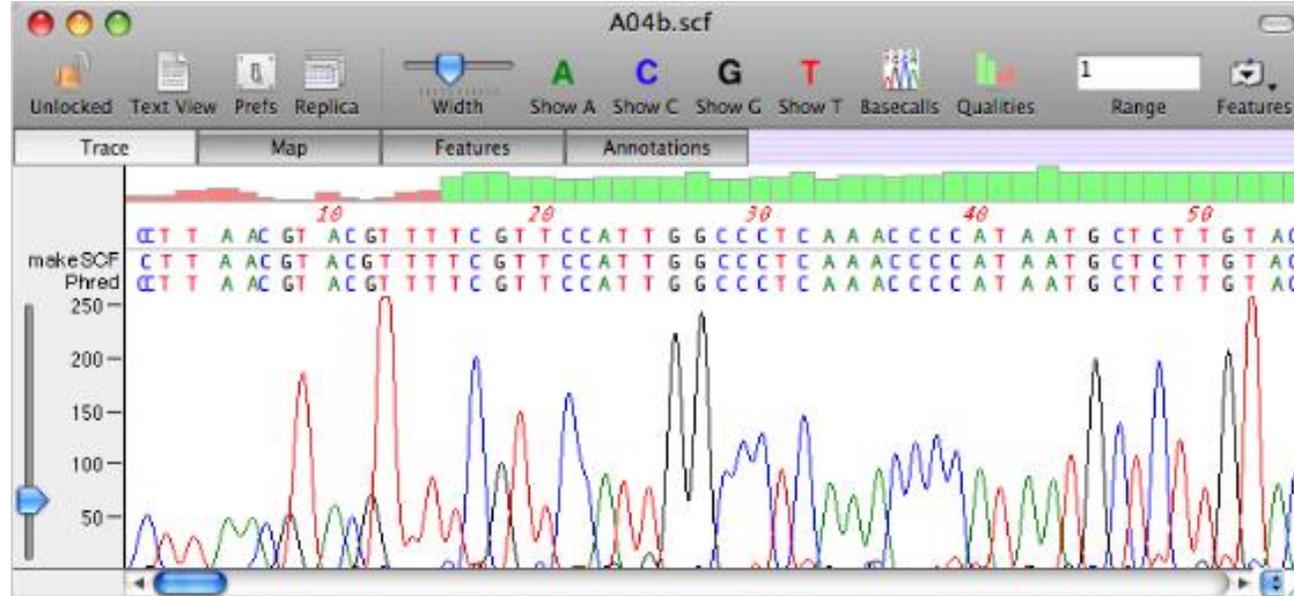




$$q = -10 \times \log_{10}(P)$$

Donde q es la valor phred y P es la probabilidad de encontrar una base errada:

- valor phred = 20 => 1 base errada a cada 100 (99%)
- valor phred = 30 => 1 base errada a cada 1000 (99.9%)



Phred quality score	Probability that the base is called wrong	Accuracy of the base call
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

Características del secuenciamiento - Sanger



Características	
Tamaño de los reads	500-900 nt
Total de reads	96 - 384
Total de bases	50 - 350 kb
Tiempo de corrida	15 horas
Costo	mas de USD 1.000/Mb

Tecnologías de segunda generación



Genome Sequencer FLX Titanium
Roche



SOLiD 3
Applied Biosystems



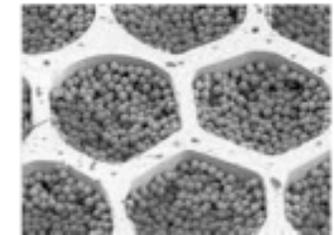
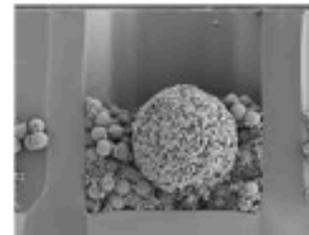
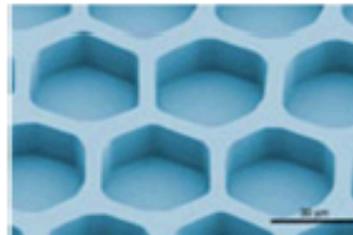
Genome Analyzer II_x
Illumina

454 GS FLX Titanium



Características

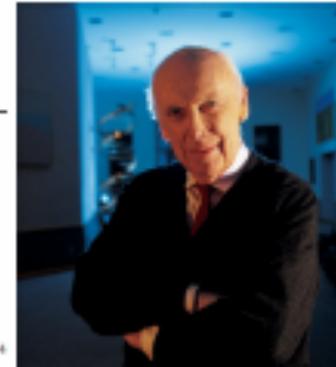
Tamanho dos <i>reads</i>	300 – 400 nt
Total de <i>reads</i>	1 milhão
Total de bases	300 – 400 Mb
Tempo de corrida	10 horas
Custo	USD 22.50 /Mb



LETTERS

The complete genome of an individual by massively parallel DNA sequencing

David A. Wheeler^{1*}, Maithreyan Srinivasan^{2*}, Michael Egholm^{2*}, Yufeng Shen^{1*}, Lei Chen¹, Amy McGuire², Wen He², Yi-Ju Chen², Vinod Makhijani², G. Thomas Roth², Xavier Gomes², Karrie Tartaro^{2†}, Faheem Niazi², Cynthia L. Turcotte², Gerard P. Irzyk², James R. Lupski^{1,3,4}, Craig Chinault², Xing-zhi Song¹, Yue Liu¹, Ye Yuan¹, Lynne Nazareth¹, Xiang Qin¹, Donna M. Muzny¹, Marcel Margulies², George M. Weinstock^{1,4}, Richard A. Gibbs^{1,4} & Jonathan M. Rothberg^{2†}



James Watson

2 meses para el secuenciamiento
100 veces más barato

Secuenciamiento tipo SOLEXA

Illumina Hiseq 2000

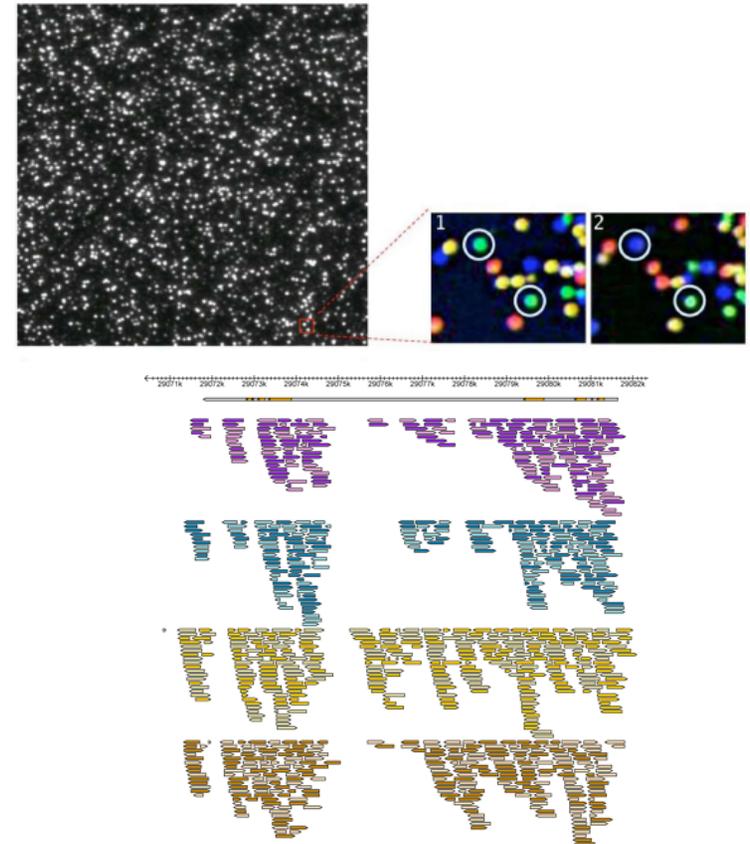


Características	
Tamaño de los reads	Hasta 150 nt
Total de reads	3 billones
Total de bases	100 – 200 Gb
Tiempo de corrida	Entre 1.5 y 8 días
Costo	Menos de USD 0.1/Mb

Combina la química SBS con terminadores y tecnología de grupos (del inglés, “cluster”). La compañía fue adquirida por Illumina en 2007, produciendo el “Genome Analyzer Sequencing System” Este tipo de tecnología genera diez veces más lecturas que la 454, pero con solamente 35 pb o menos en longitud.

Principales Aplicaciones del Secuenciamiento Illumina

- DNA Sequencing
- Gene Regulation Analysis
- Sequencing-Based Transcriptome Analysis
- SNP Discovery and Structural Variation
- Cytogenetic Analysis
- DNA-Protein Interaction Analysis (ChIP-Seq)
- Sequencing-Based Methylation Analysis
- Small RNA Discovery
- Personal Genomics
- Pharmacogenomics



Cobertura vs Costo

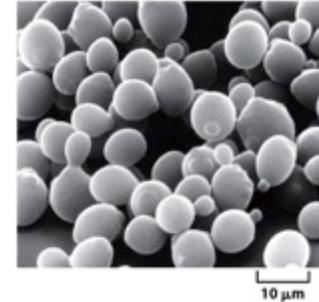
1 Línea (USD 2000) = 10 Gb



Arabidopsis (80X)



E. coli (2000X)



Saccharomyces (700X)



Drosophila (75X)



Humanos (3X)

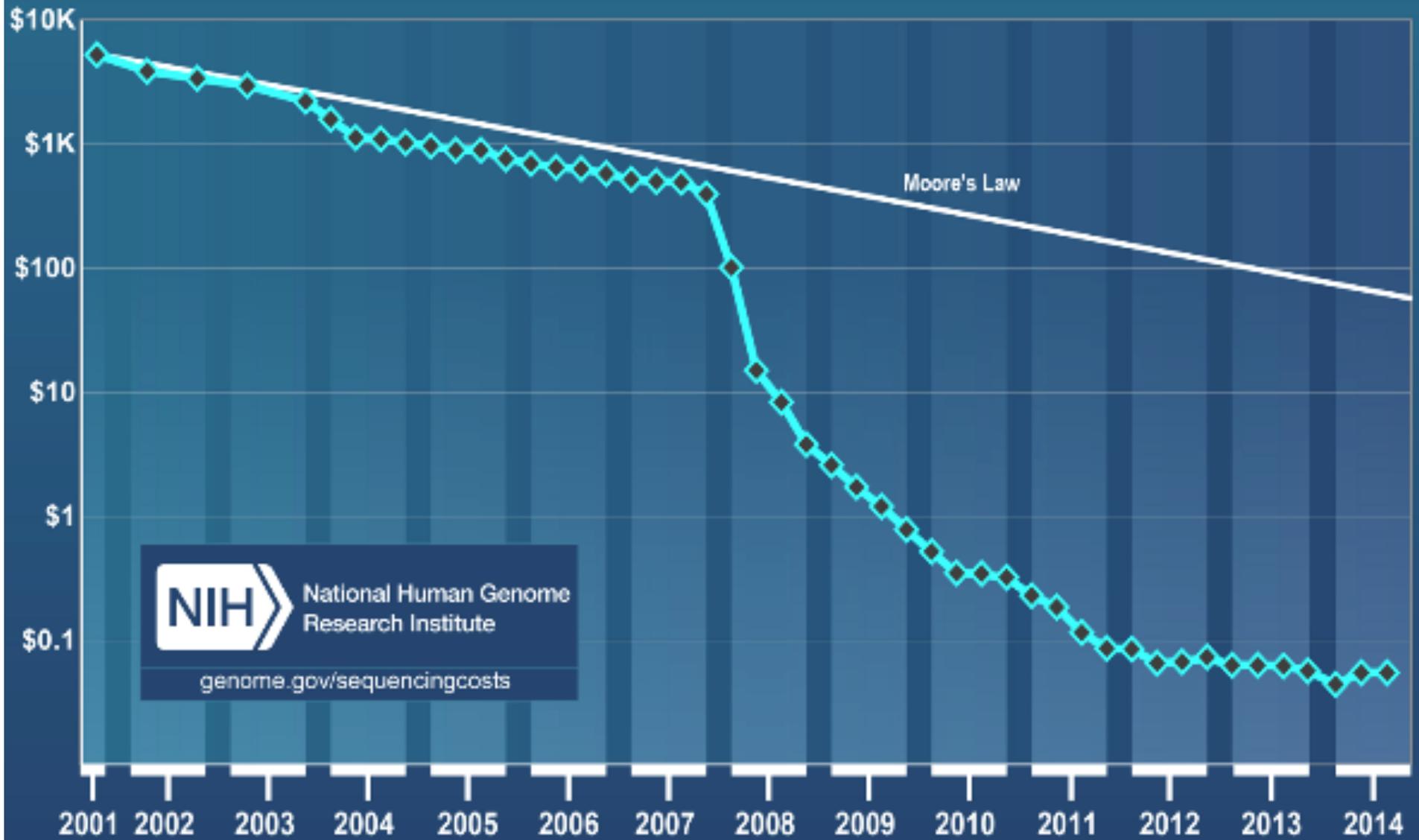


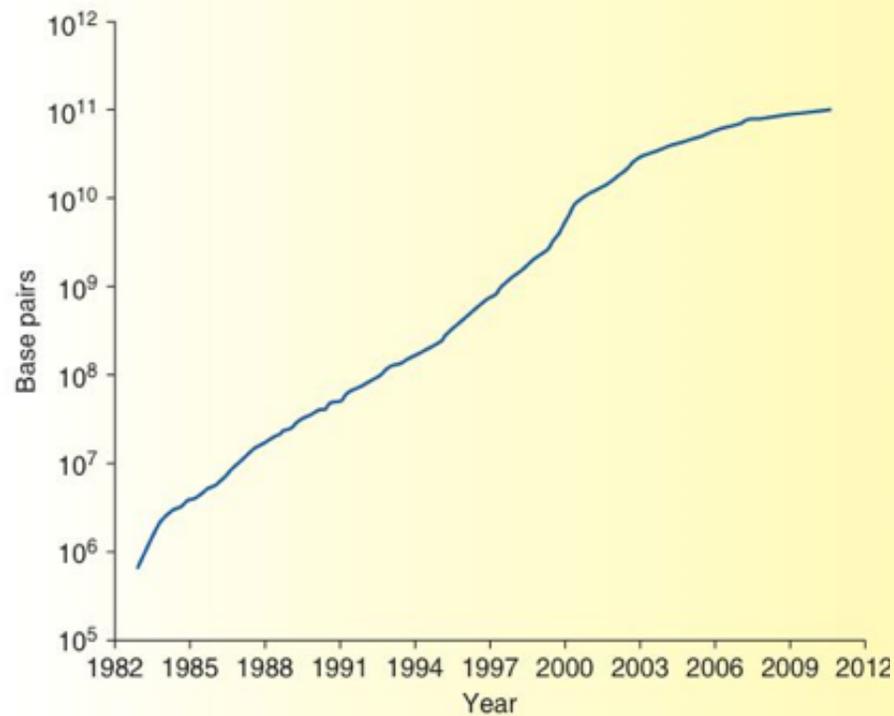
M. pernicioso (200X)

Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]

Cost per Raw Megabase of DNA Sequence





Courtesy of NCBI'S GenBank, October 2007

Figure B10.1. Growth in DNA sequencing The amount of sequenced DNA continues to increase rapidly. The number of base pairs deposited in the GenBank database each year is increasing exponentially.

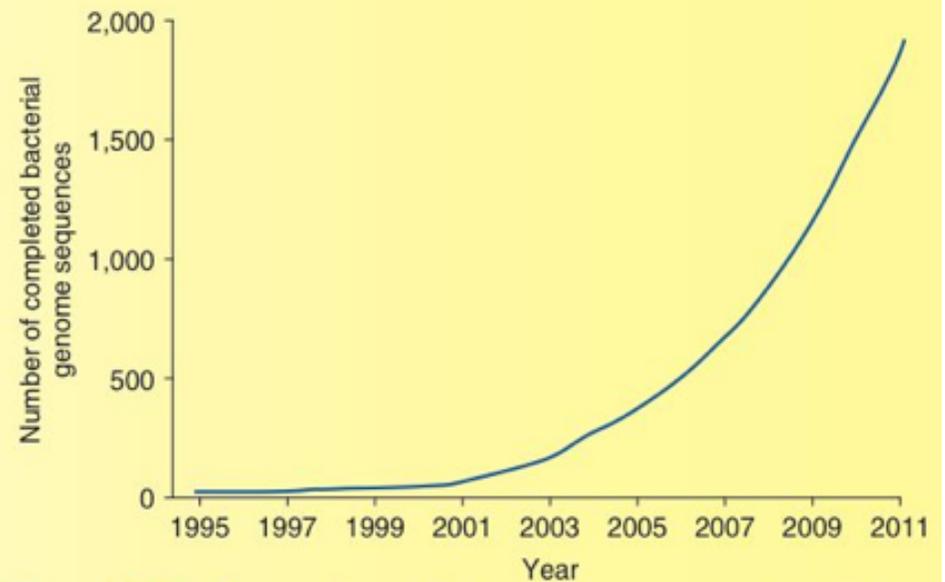
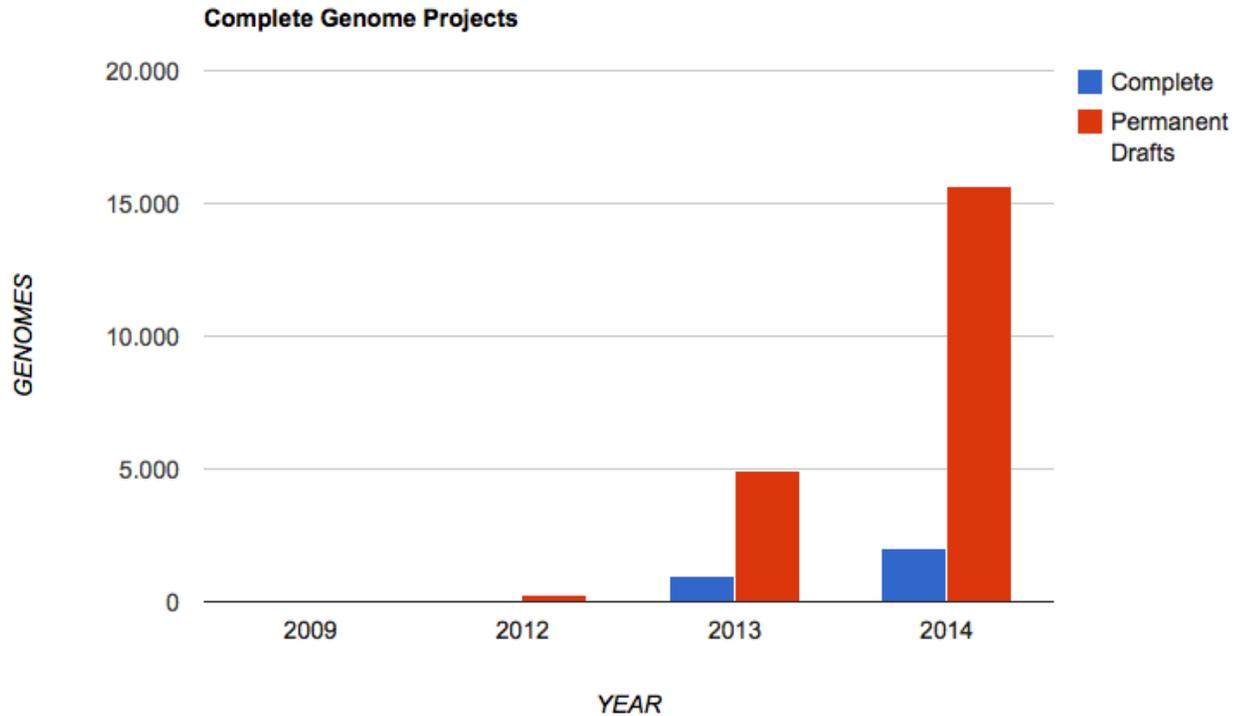


Figure B10.2. Sequencing of bacterial genomes Since the late 1990s, the number of sequenced bacterial genomes has been increasing exponentially, doubling about every 20 months.

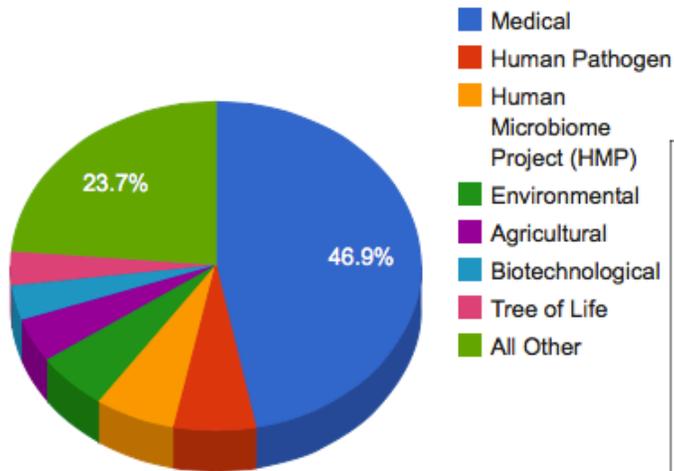
Statistics

Complete and Permanent Draft Genome Totals in GOLD (by year and status)



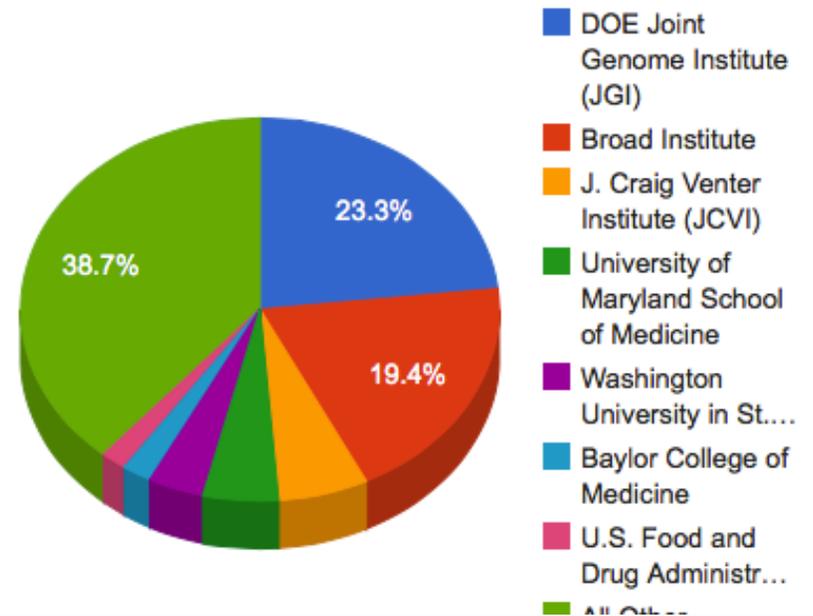
Project Relevance of Bacterial Genome Projects

Project Relevance of Bacterial Projects

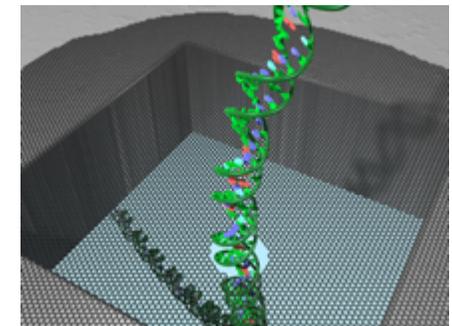
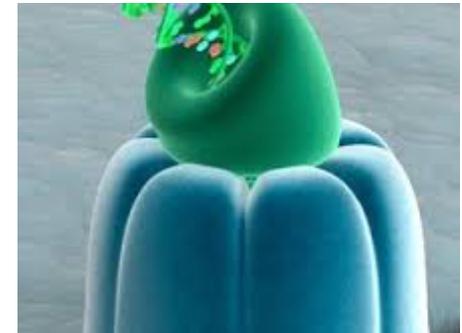


Projects By Major Sequencing Centers

Projects by Sequencing Center



Tercera generación de secuenciadores



Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]

Published Online May 20 2010
Science 2 July 2010:
Vol. 329 no. 5987 pp. 52–56
DOI: 10.1126/science.1190719

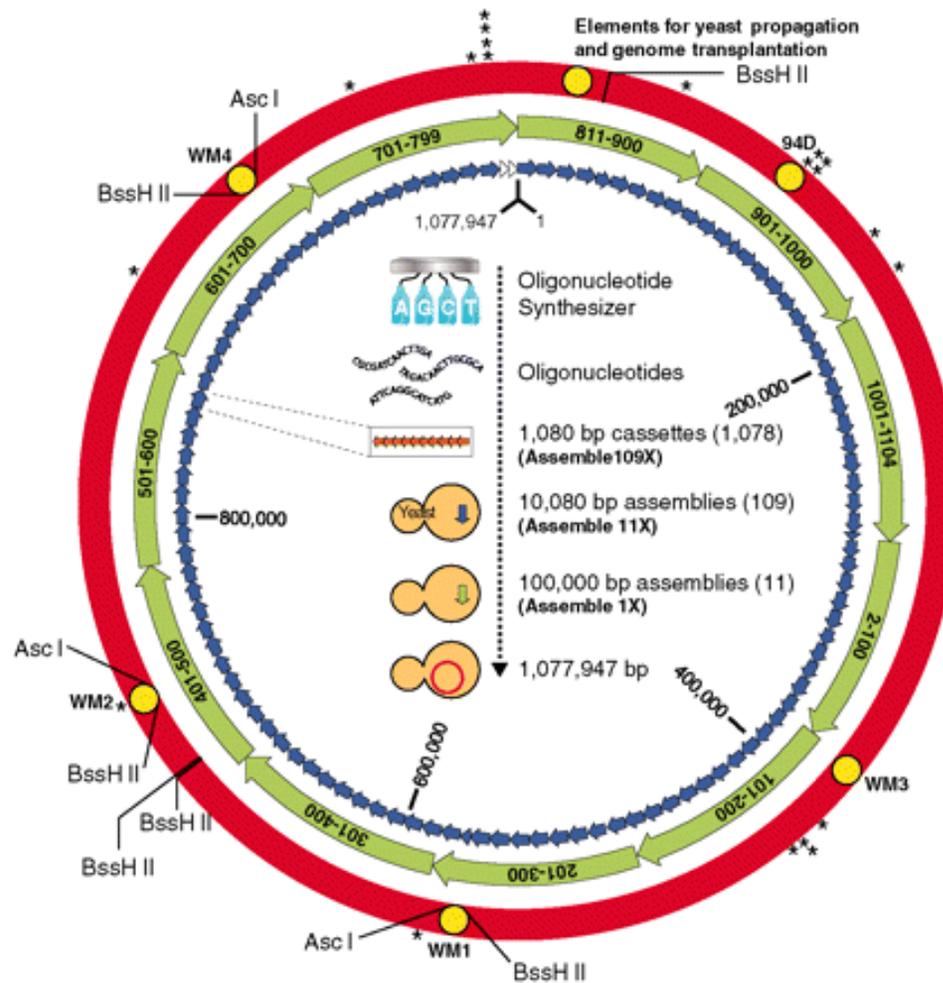
[< Prev](#) | [Table of Contents](#) | [Next >](#)

RESEARCH ARTICLE

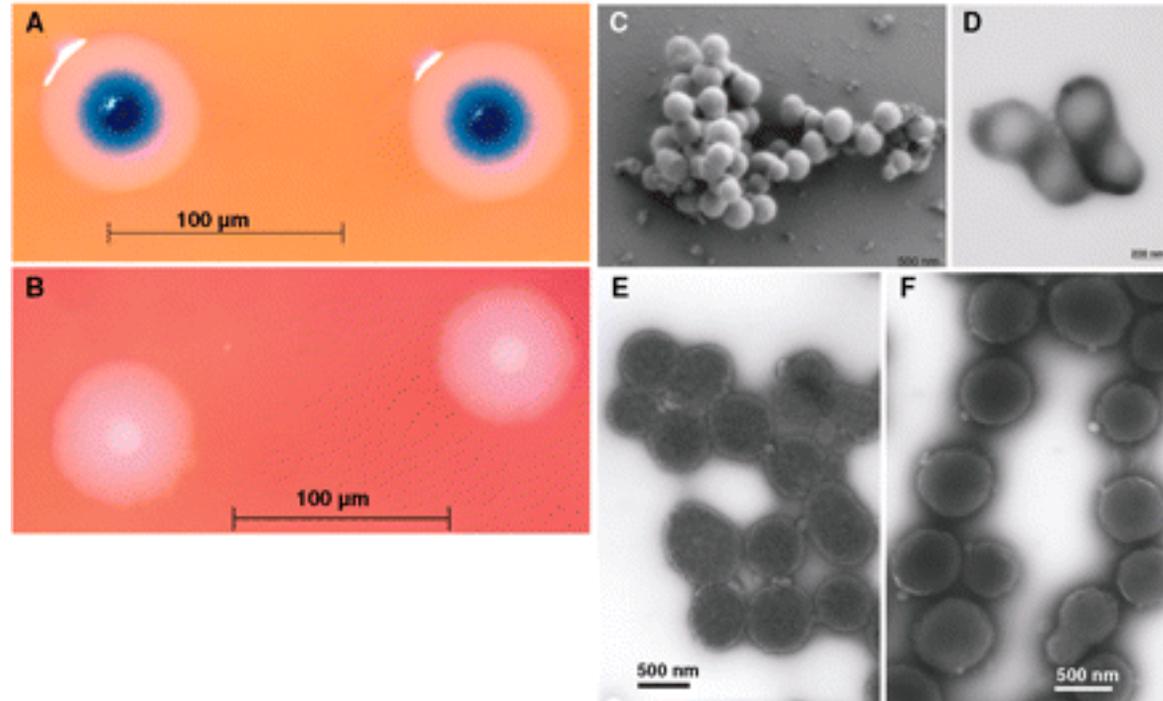
Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome

Daniel G. Gibson¹, John I. Glass¹, Carole Lartigue¹, Vladimir N. Noskov¹, Ray-Yuan Chuang¹,
Mikkel A. Algire¹, Gwynedd A. Benders², Michael G. Montague¹, Li Ma¹, Monzia M. Moodie¹,
Chuck Merryman¹, Sanjay Vashee¹, Radha Krishnakumar¹, Nacyra Assad-Garcia¹,
Cynthia Andrews-Pfannkoch¹, Evgeniya A. Denisova¹, Lei Young¹, Zhi-Qing Qi¹,
Thomas H. Segall-Shapiro¹, Christopher H. Calvey¹, Prashanth P. Parmar¹, Clyde A. Hutchison III²,
Hamilton O. Smith², J. Craig Venter^{1,2,*}

The assembly of a synthetic *M. mycoides* genome in yeast



Sintética



Wild-type

Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]

¿Qué hacemos en
investigación en genómica?

New Hydrocarbon Degradation Pathways in the Microbial Metagenome from Brazilian Petroleum Reservoirs

Isabel Natalia Sierra-García^{1*}, Javier Correa Alvarez², Suzan Pantaroto de Vasconcellos³, Anete Pereira de Souza⁴, Eugenio Vaz dos Santos Neto⁵, Valéria Maia de Oliveira¹

1 Microbial Resources Division, Research Center for Chemistry, Biology and Agriculture (CPQBA), University of Campinas - UNICAMP, Campinas, Brazil, **2** Laboratory of Genomics and Expression, University of Campinas - UNICAMP, Campinas, Brazil, **3** Federal University of São Paulo - UNIFESP, São Paulo, Brazil, **4** Center of Molecular Biology and Genetic Engineering – CBMEG/UNICAMP, Rio de Janeiro, Brazil, **5** PETROBRAS/R&D Center, Rio de Janeiro, Brazil

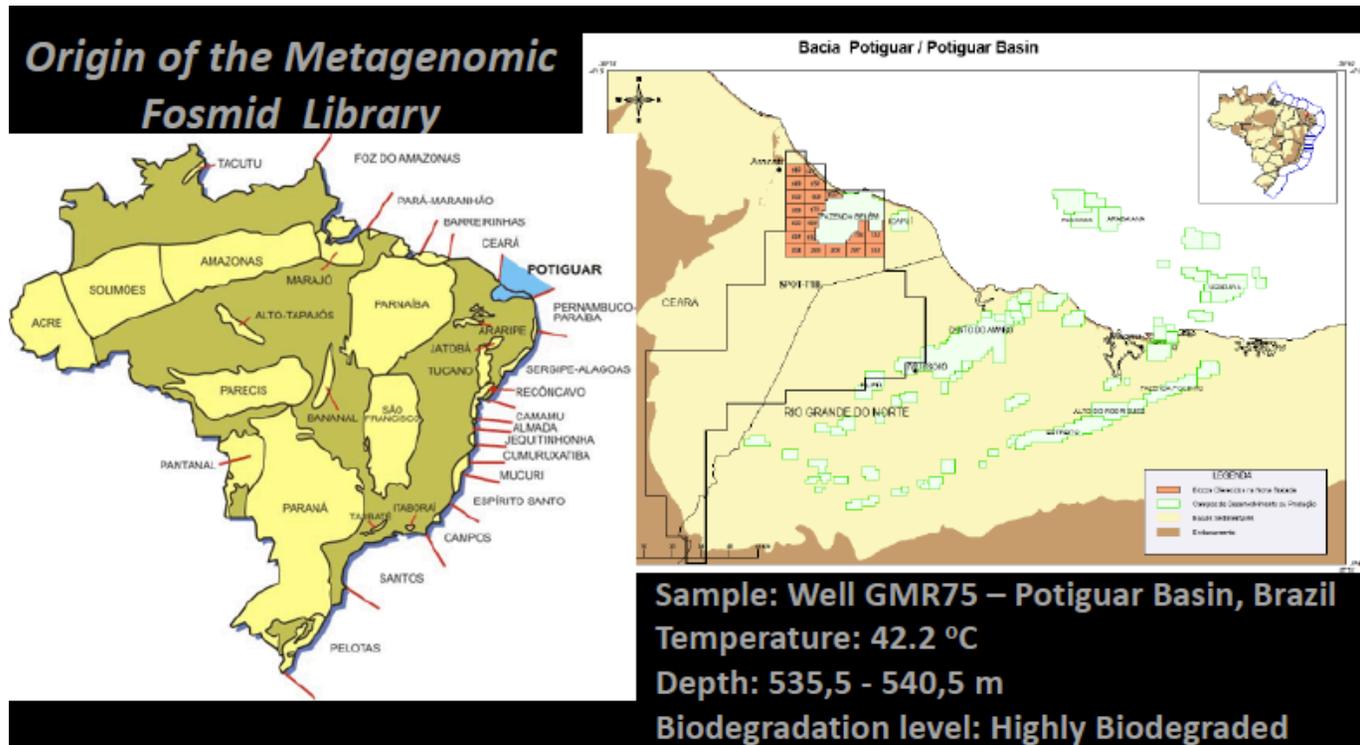
Abstract

Current knowledge of the microbial diversity and metabolic pathways involved in hydrocarbon degradation in petroleum reservoirs is still limited, mostly due to the difficulty in recovering the complex community from such an extreme environment. Metagenomics is a valuable tool to investigate the genetic and functional diversity of previously uncultured microorganisms in natural environments. Using a function-driven metagenomic approach, we investigated the metabolic abilities of microbial communities in oil reservoirs. Here, we describe novel functional metabolic pathways involved in the biodegradation of aromatic compounds in a metagenomic library obtained from an oil reservoir. Although many of the deduced proteins shared homology with known enzymes of different well-described aerobic and anaerobic catabolic pathways, the metagenomic fragments did not contain the complete clusters known to be involved in hydrocarbon degradation. Instead, the metagenomic fragments comprised genes belonging to different pathways, showing novel gene arrangements. These results reinforce the potential of the metagenomic approach for the identification and elucidation of new genes and pathways in poorly studied environments and contribute to a broader perspective on the hydrocarbon degradation processes in petroleum reservoirs.

Citation: Sierra-García IN, Correa Alvarez J, Pantaroto de Vasconcellos S, Pereira de Souza A, dos Santos Neto EV, et al. (2014) New Hydrocarbon Degradation Pathways in the Microbial Metagenome from Brazilian Petroleum Reservoirs. PLoS ONE 9(2): e90087. doi:10.1371/journal.pone.0090087

Tratamiento y Montaje de genoma

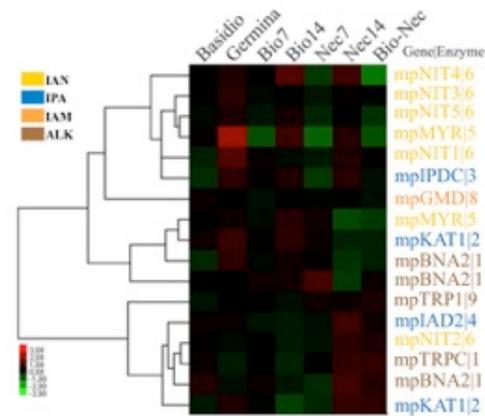
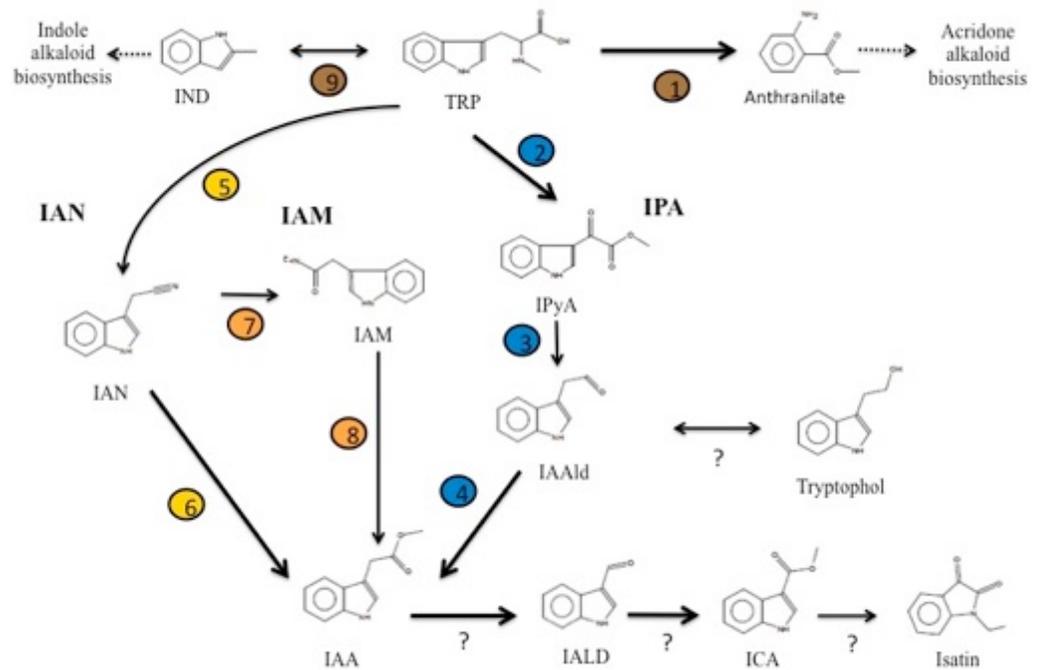
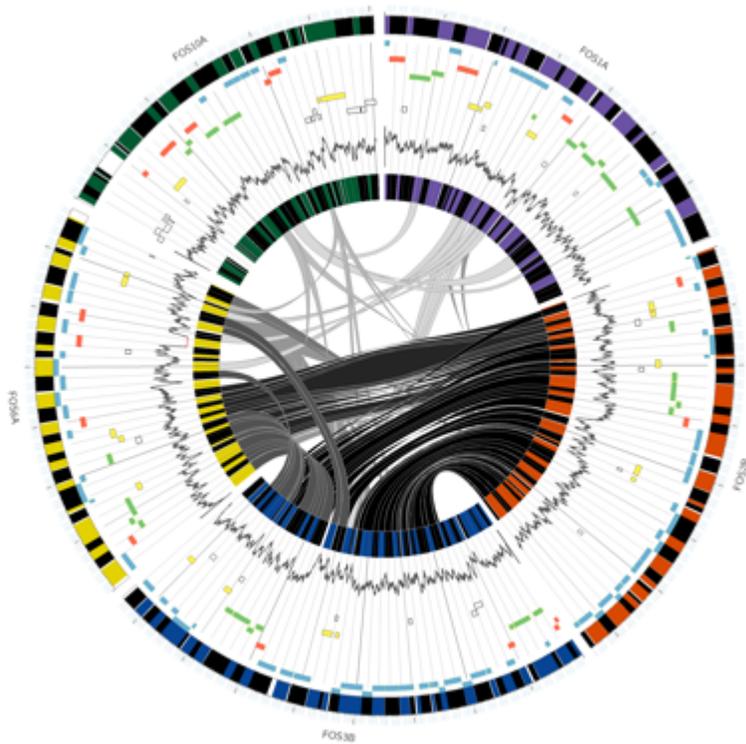
Naturaleza de los datos: Metagenómica



Vasconcellos et al. 2010 - [dx.doi.org/10.1016/j.orggeochem.2010.03.014](https://doi.org/10.1016/j.orggeochem.2010.03.014).

Inspira Crea Transforma

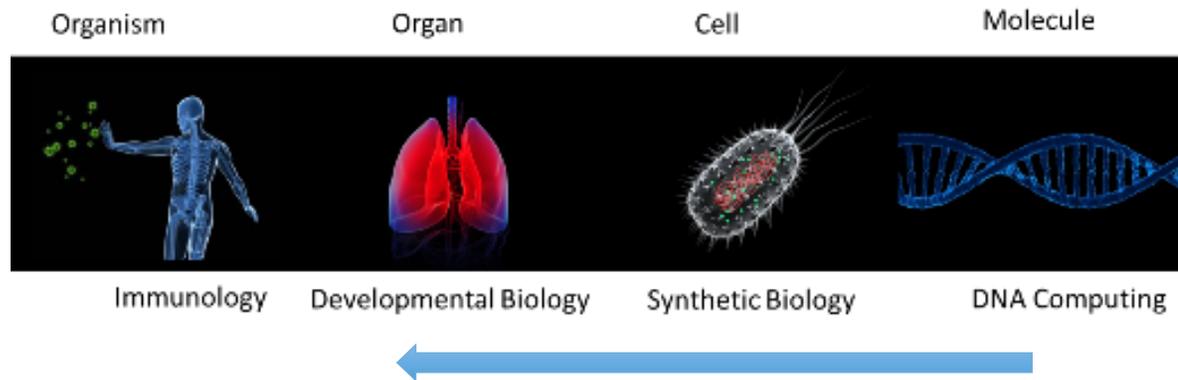
UNIVERSIDAD
EAFIT[®]



Alvarez et, al. 2015. FGB in-pross

Genómica, Transcriptómica y Metabolómica
 -> Interactómica -> Biología de sistemas

¿Qué viene para el “futuro” en la biología?



Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]

Functional genomic hypothesis generation and experimentation by a robot scientist

Ross D. King¹, Kenneth E. Whelan¹, Ffion M. Jones¹, Philip G. K. Reiser¹, Christopher H. Bryant², Stephen H. Muggleton³, Douglas B. Kell⁴ & Stephen G. Oliver⁵

1. Department of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK

2. School of Computing, The Robert Gordon University, Aberdeen AB10 1FR, UK

3. Department of Computing, Imperial College, London SW7 2AZ, UK

4. Department of Chemistry, UMIST, P.O. Box 88, Manchester M60 1QD, UK

5. School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Manchester M13 9PT, UK

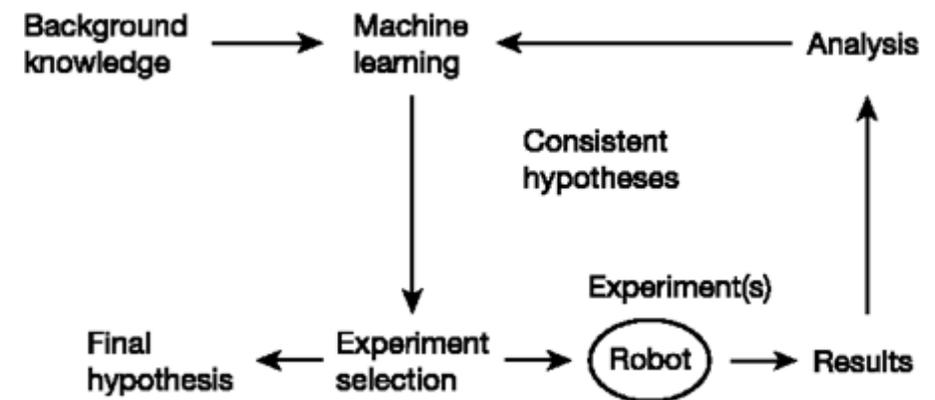
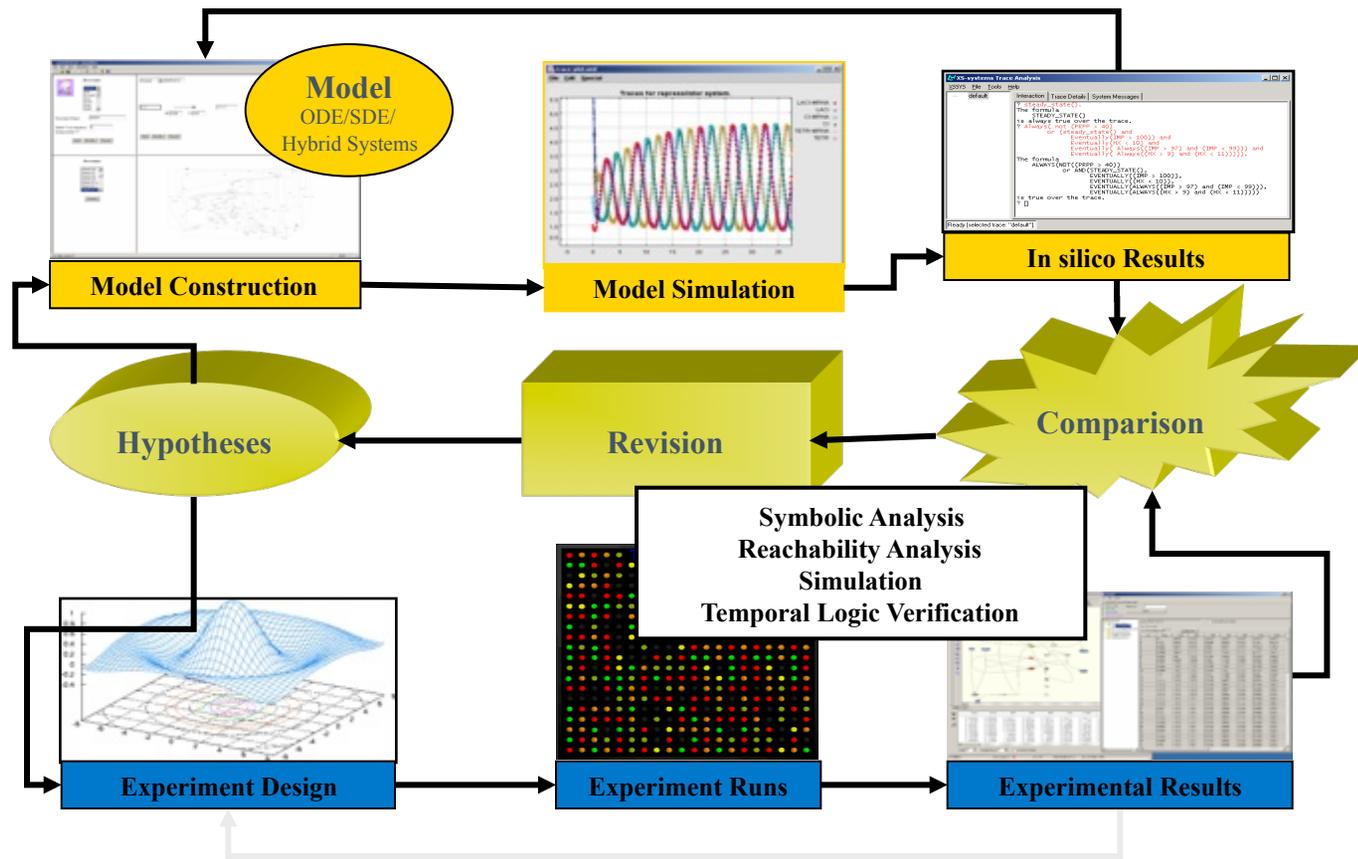


Figure 1 The Robot Scientist hypothesis-generation and experimentation loop.

Razonamiento y experimentación, el futuro en los sistemas biológicos



High-Throughput Screening en la plataforma Hamilton STARlet

Drug discovery

Quimiogenómica



Inspira Crea Transforma

**UNIVERSIDAD
EAFIT[®]**

Anotación de genomas

Descubierta de genes que codifican “enzimas”

20 genes candidatos, 12 confirmados en HT

3 genes verificados manualmente



The Automation of Science

Ross D. King *et al.*

Science **324**, 85 (2009);

DOI: 10.1126/science.1165620

WIRED SCIENCE

NEWS FOR YOUR NEURONS



[PREVIOUS POST](#)

First Robot Scientist Makes Gene Discovery

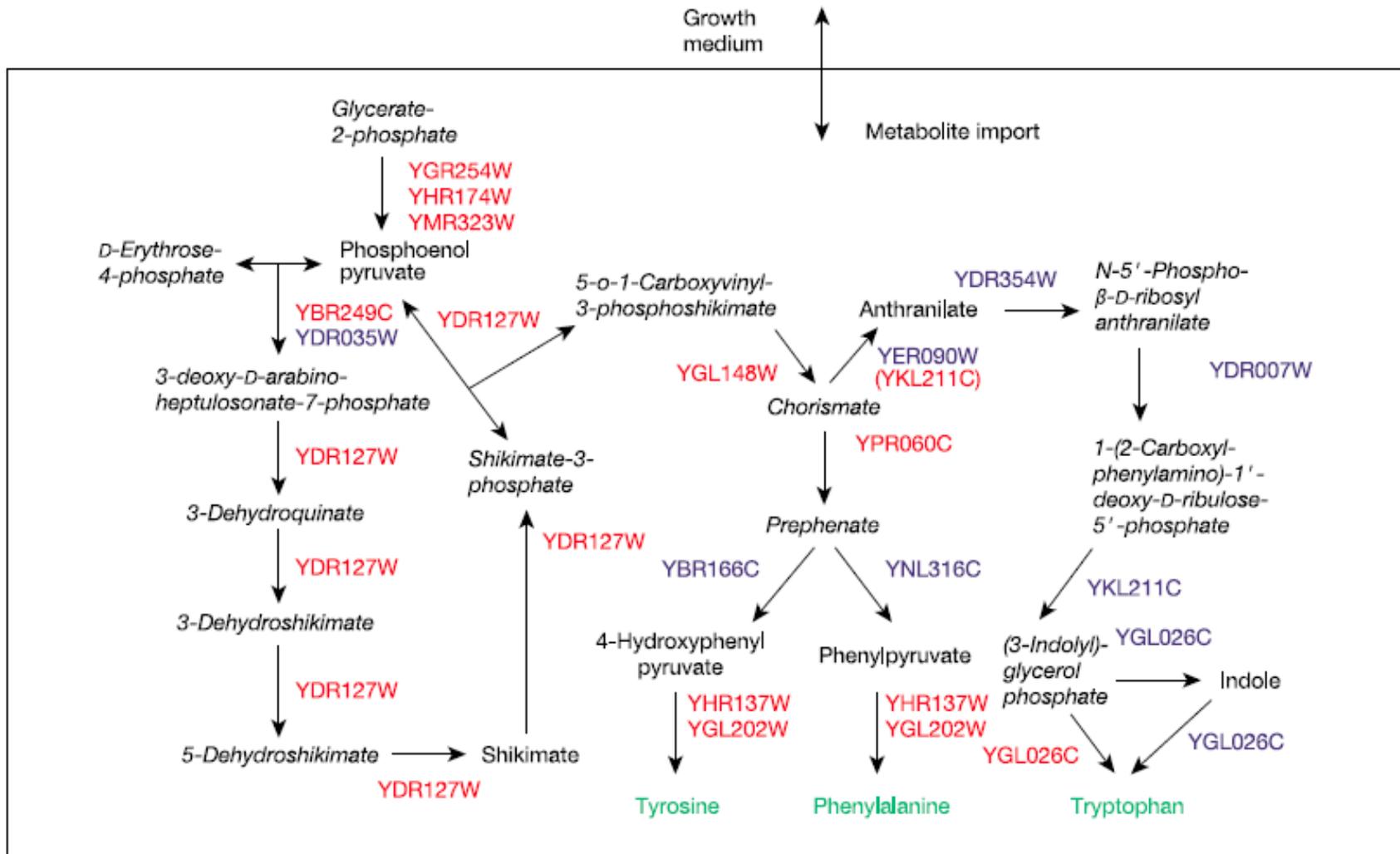
Kate Ravilious
for [National Geographic News](#)
April 2, 2009

Robot Makes Scientific Discovery All by Itself

By [Lizzie Buchen](#) April 2, 2009 | 12:15 pm | Categories: [Tech](#)

Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]



¿Dónde vemos al doctorado en Ing.
matemáticas contribuir a la biología
computacional?

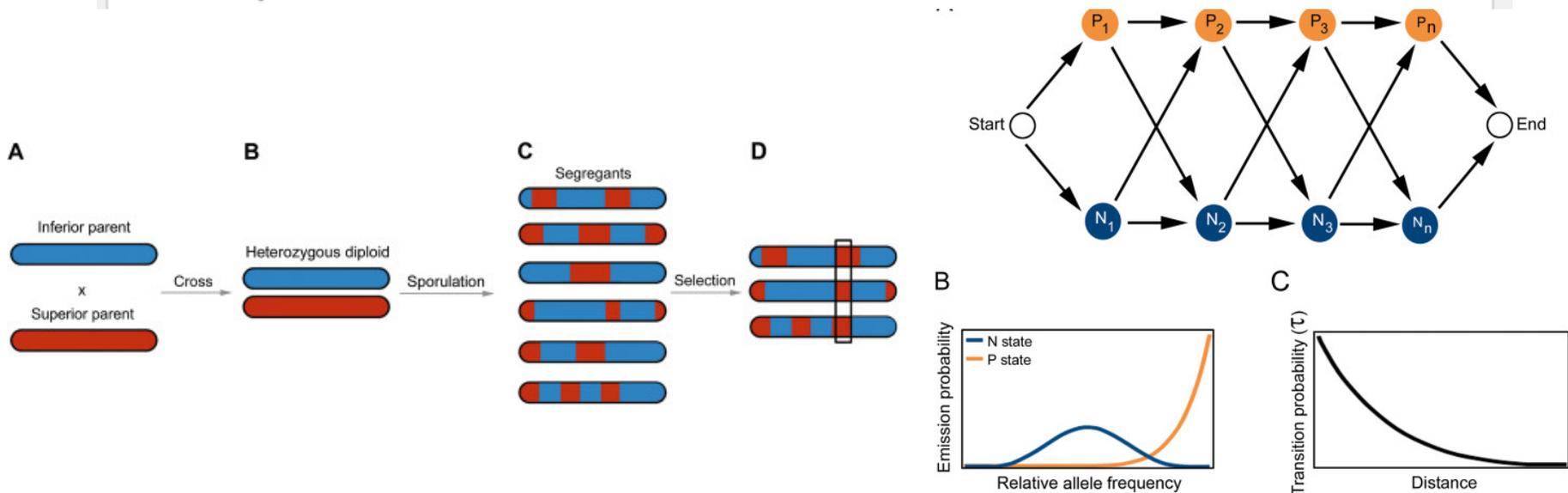
Análisis de datos por medio de
Modelos Gráficos
(e.g. Hidden Markov Models)

Methodology article

Open Access

Improved linkage analysis of Quantitative Trait Loci using bulk segregants unveils a novel determinant of high ethanol tolerance in yeast

Jorge Duitama¹, Aminaél Sánchez-Rodríguez², Annelies Goovaerts³, Sergio Pulido-Tamayo^{2,4,5}, Georg Hubmann³, María R Foulquié-Moreno³, Johan M Thevelein^{3*}, Kevin J Verstrepen^{1*} and Kathleen Marchal^{2,4,5*}

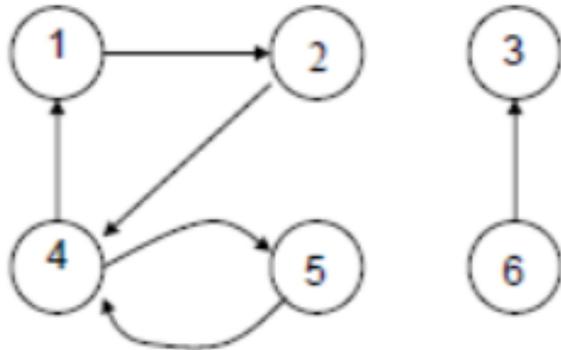


Análisis de datos por medio de
redes de interacción
(e.g. Interactoma)

Redes (grafos)

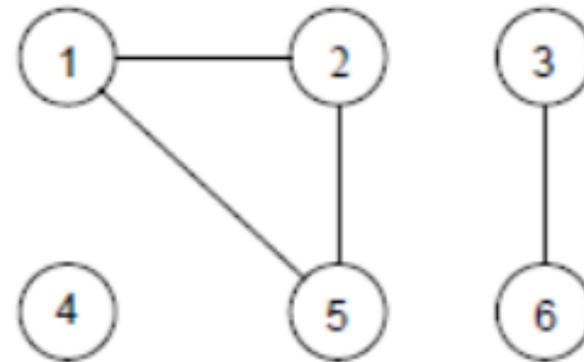
Informalmente, un grafo es un conjunto de vértices (V) conectados por un conjunto de aristas (E).

Grafo direccionado



$V = \{1, 2, 3, 4, 5, 6\}$
 $E = \{(1,2), (4,1), (2,4), (4,5), (5,4), (6,3)\}$

Grafo no direccionado



$V = \{1, 2, 3, 4, 5, 6\}$
 $E = \{(1,2), (1,5), (2,5), (3,6)\}$

Redes (grafos)

El grado o la valencia de un nodo (vértice) es el número de aristas que tiene conectados.

Una propiedad importante de las redes es la distribución de grados.

Por ejemplo:

En una red aleatoria cada nodo tiene una probabilidad p de estar conectado con cada uno de los demás nodos.

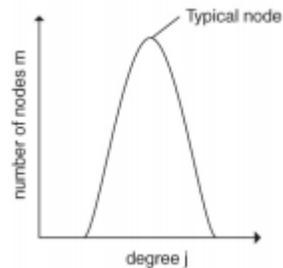
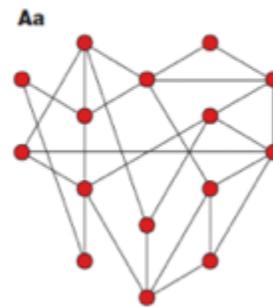
$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \text{ (binomial)}$$

Redes (grafos)

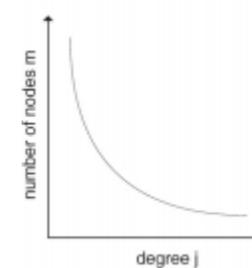
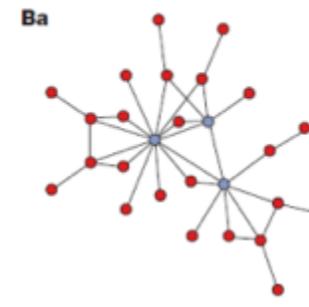
Tipos de redes

- Random
- Scale free
- Hierarchical

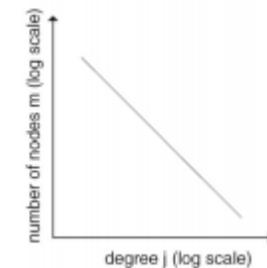
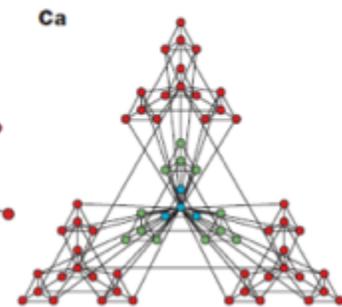
A Random network



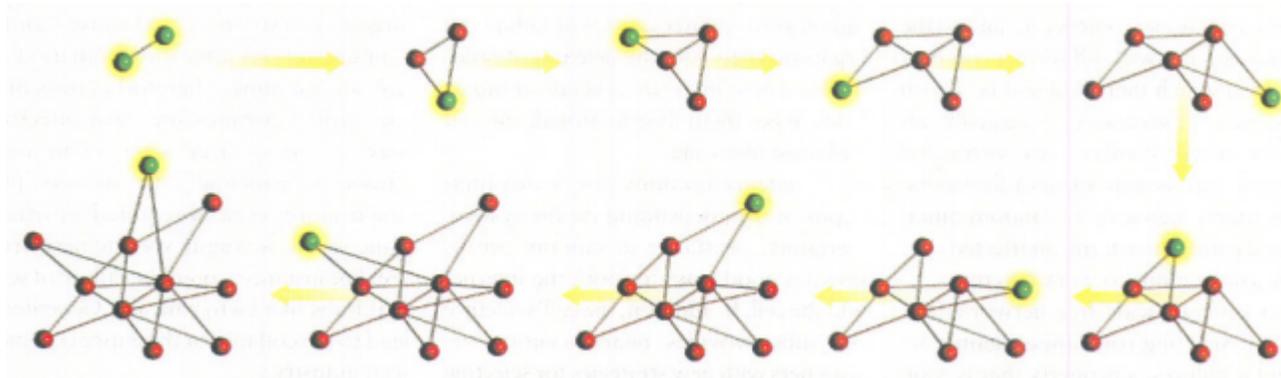
B Scale-free network



C Hierarchical network



Redes Libres de Escala (scale free networks)



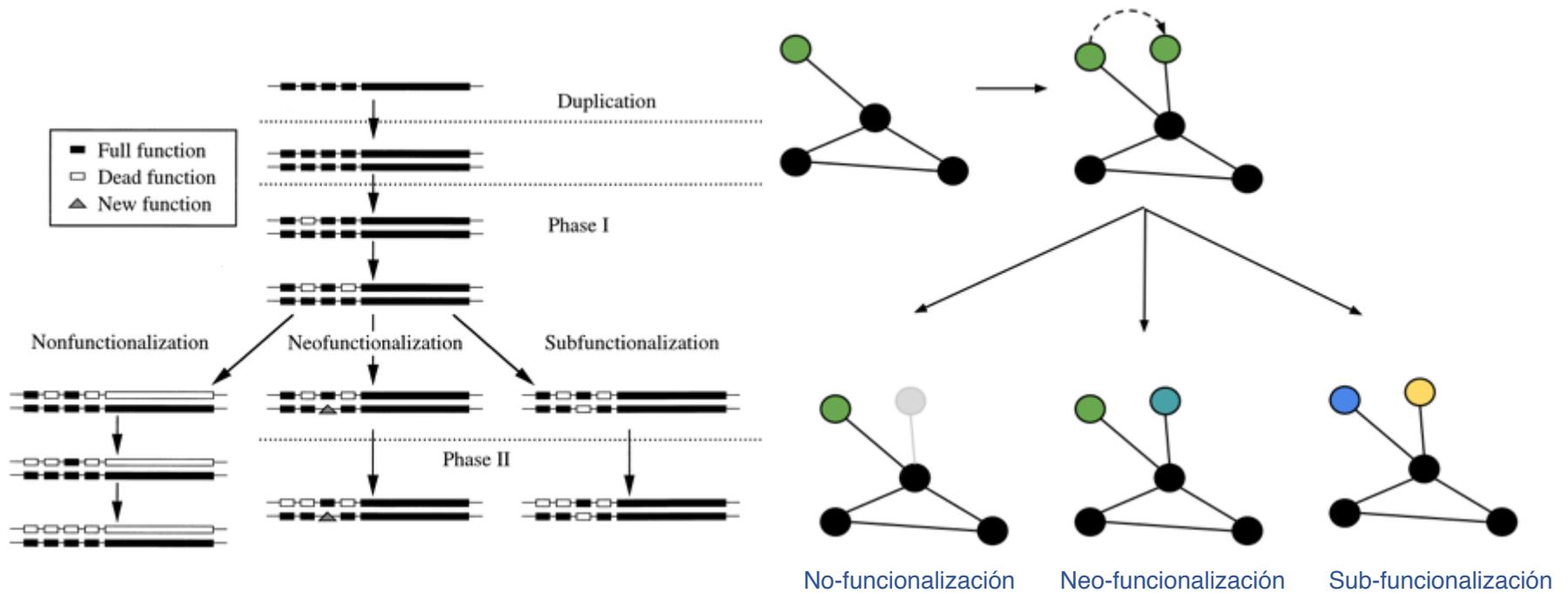
Nuevos nodos tienden a unirse con nodos ya existentes.

Las redes de interacción de proteínas son redes libres de escala

... por qué?

Redes Libres de Escala (scale free networks)

Evolución por duplicación de genes:
Nuevos nodos en la red aparecen gracias a funciones diferentes a los mismos genes de siempre



Enlaces en las redes biológicas

Genomic Context



High-throughput Experiments



(Conserved) Coexpression



Previous Knowledge



Neighborhood
 Gene Fusion
 Cooccurrence
 Coexpression
 Experiments
 Databases
 Textmining
 [Homology]

Contexto Genómico

- Elementos funcionales
- Homología
- “Guilt by association”

Experimental

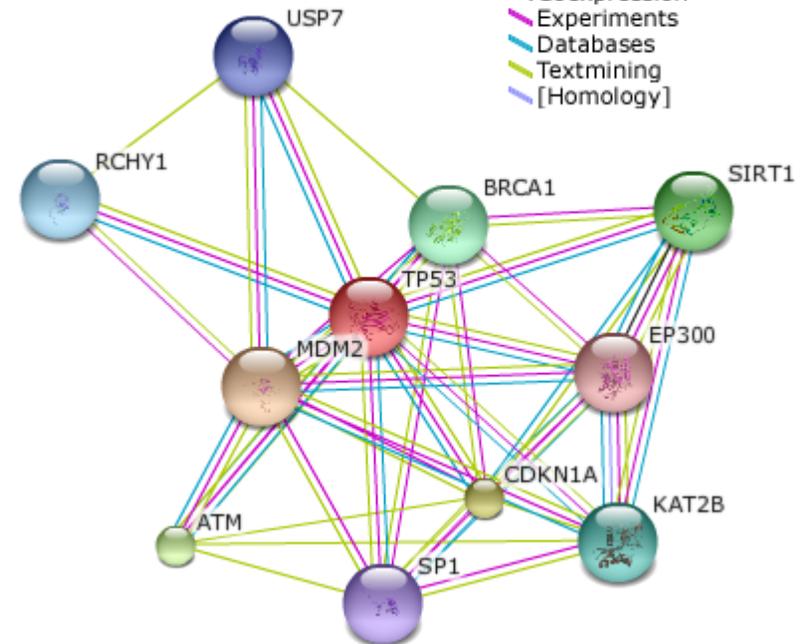
- Reacción física entre proteínas
- Y2H, etc

Coexpresión

- Genes que siempre se expresan o regulan juntos
- Deben trabajar juntos

Text Mining

- Genes que se investigaron juntos
- Se infiere relación de lo escrito en publicaciones



Preguntas sin resolver:

- Cómo traducir todo este conocimiento a un peso (weight) de la arista?
- Como inferir nuevas conexiones en redes?

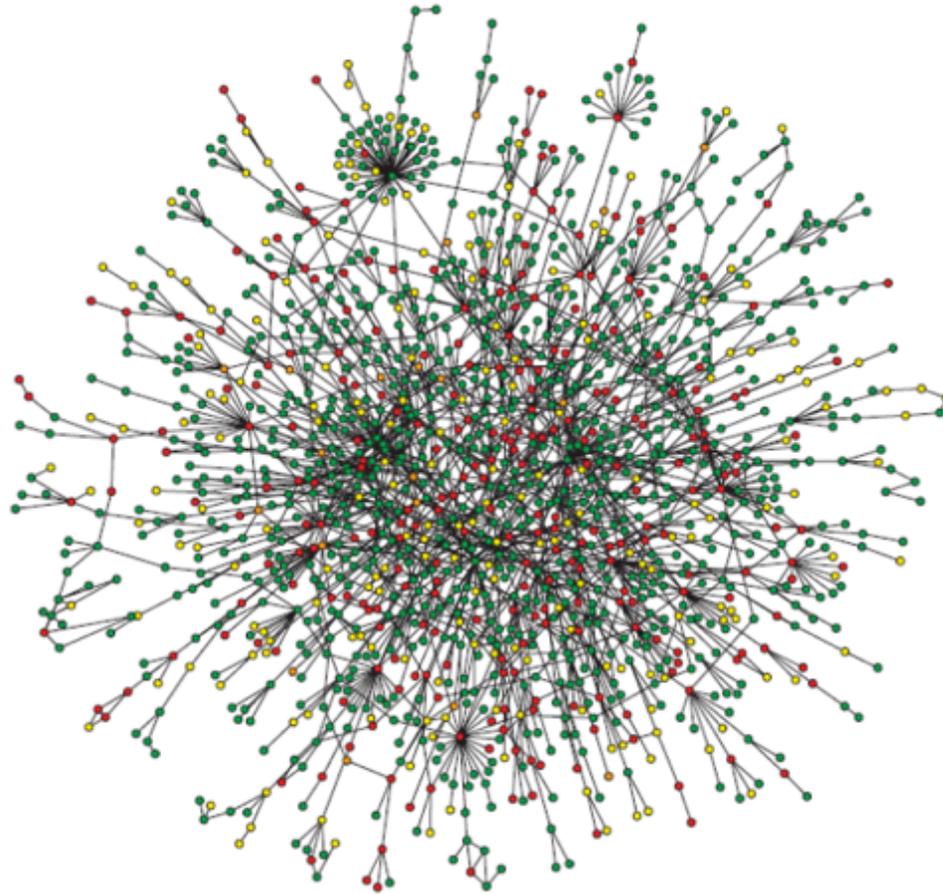
<http://string-db.org/>

Inspira Crea Transforma



UNIVERSIDAD
EAFIT

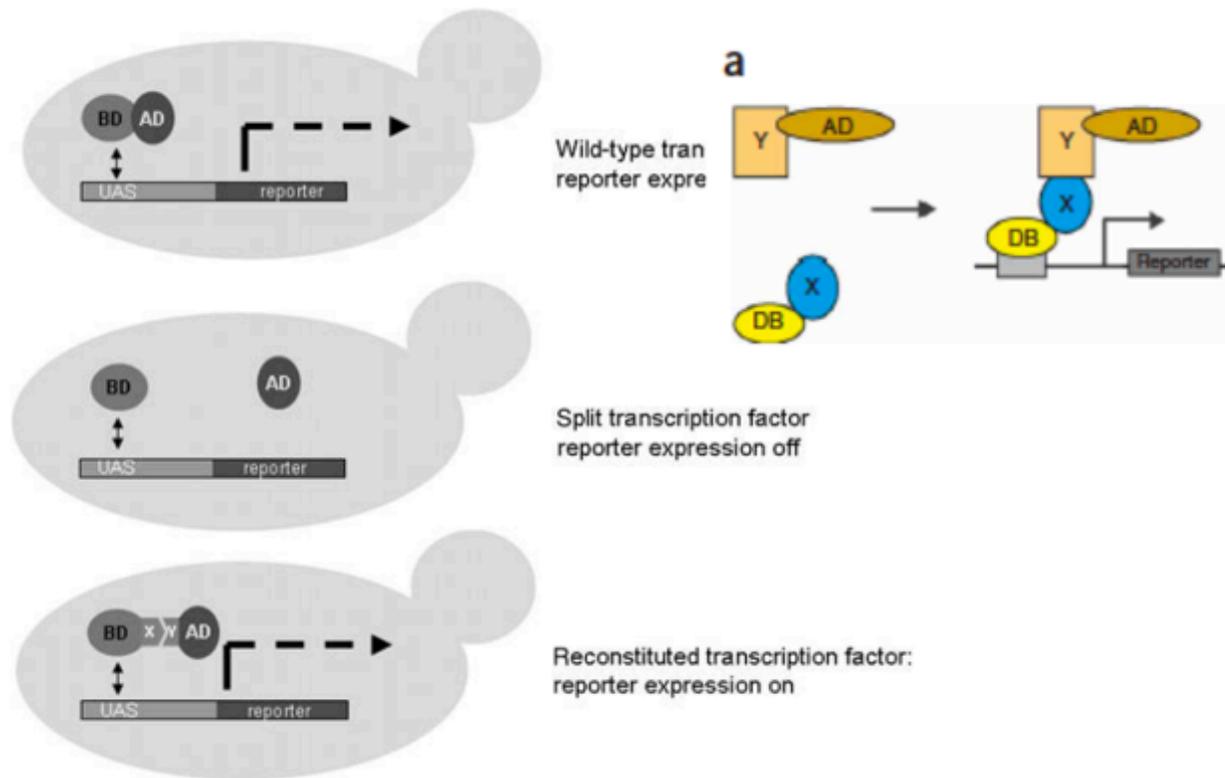
Creación de redes de Interacción



Inspira Crea Transforma

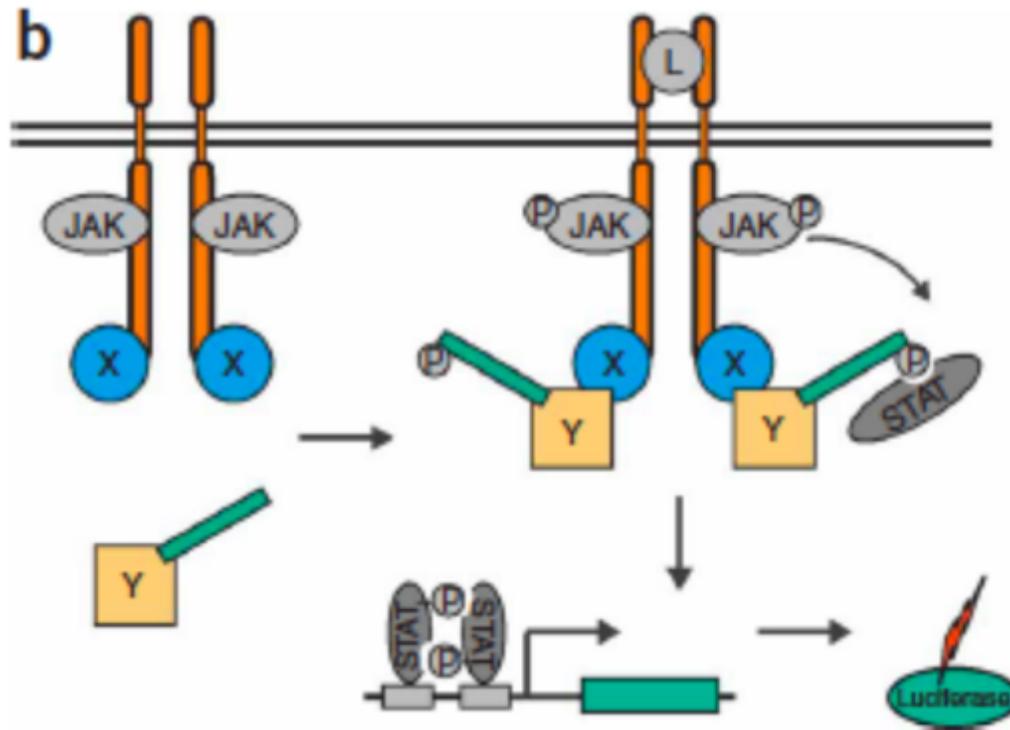
Creación de redes de Interacción

Y2H (Yeast 2 Hybrid)



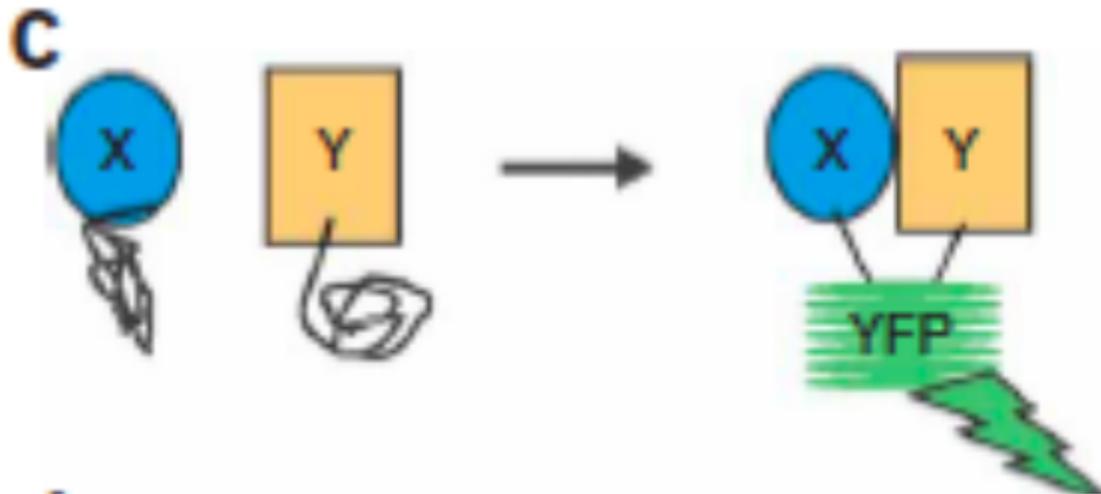
Creación de redes de Interacción

MAPPIT



Creación de redes de Interacción

PCA - YFP

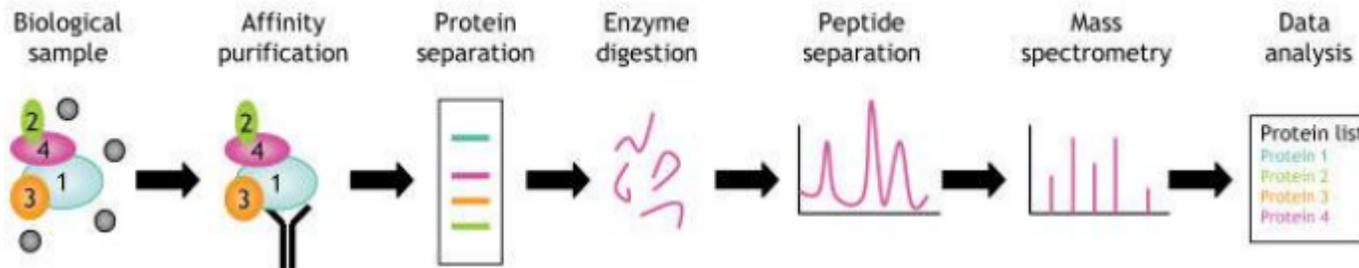


Inspira Crea Transforma

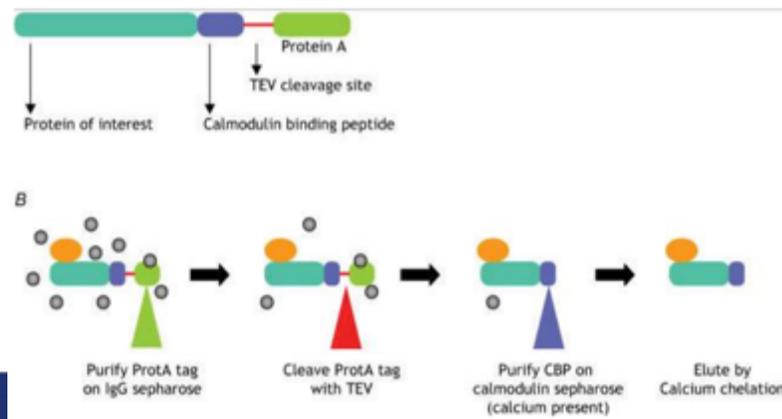
UNIVERSIDAD
EAFIT[®]

Creación de redes de Interacción

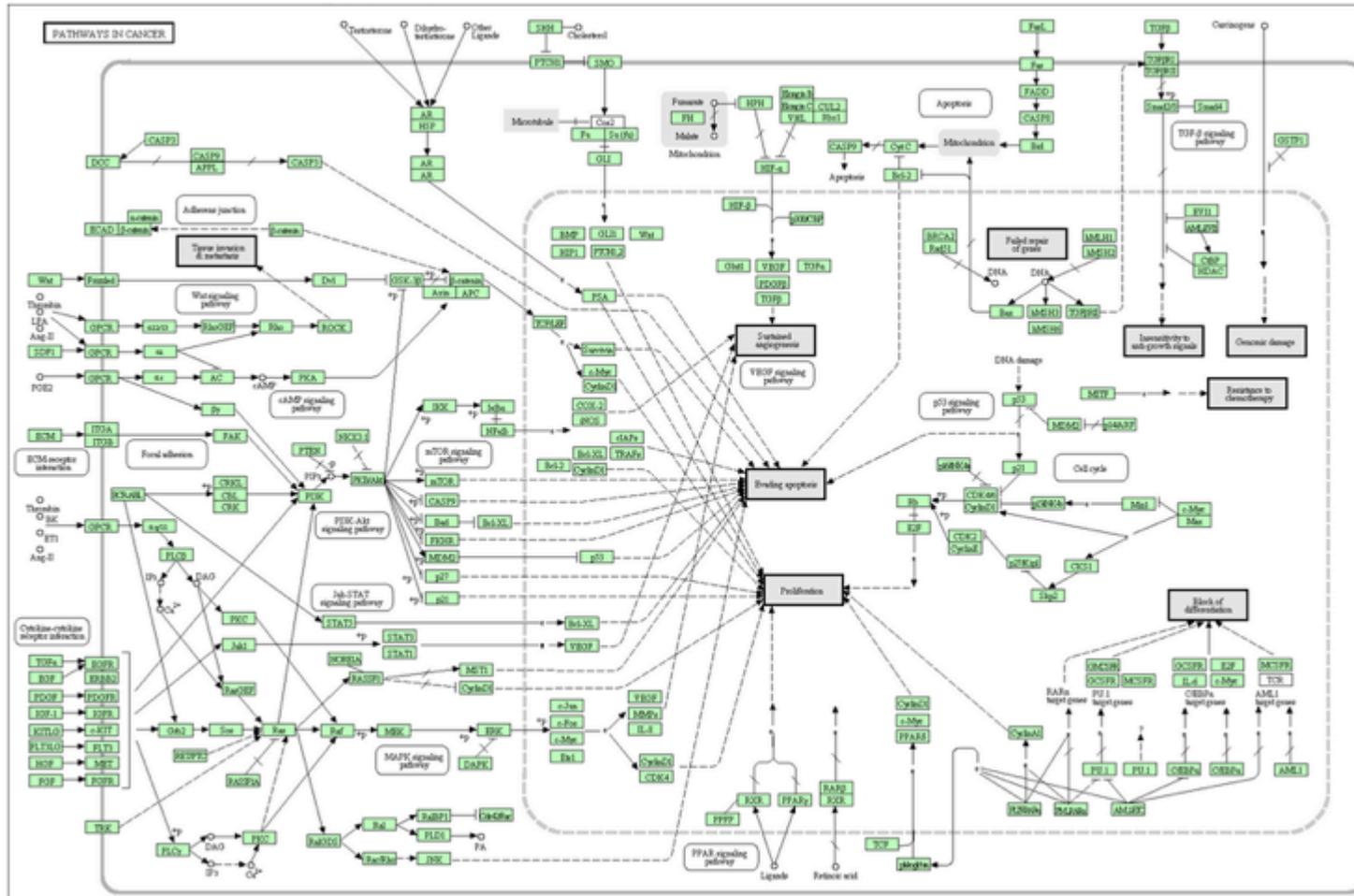
Affinity purification



Tandem AP

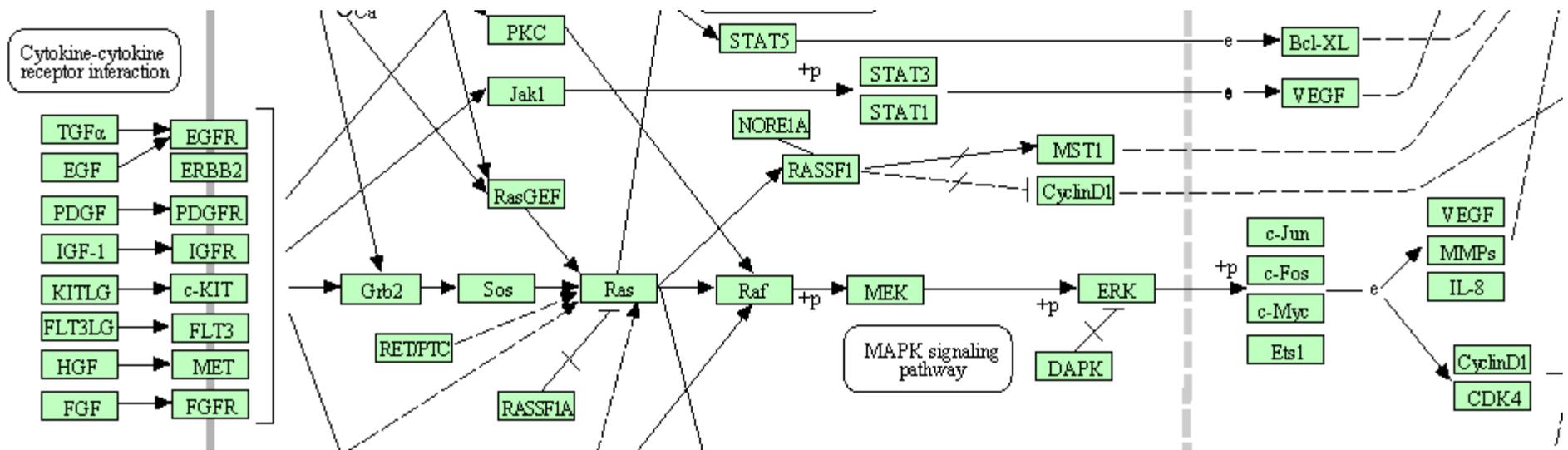


Pathways

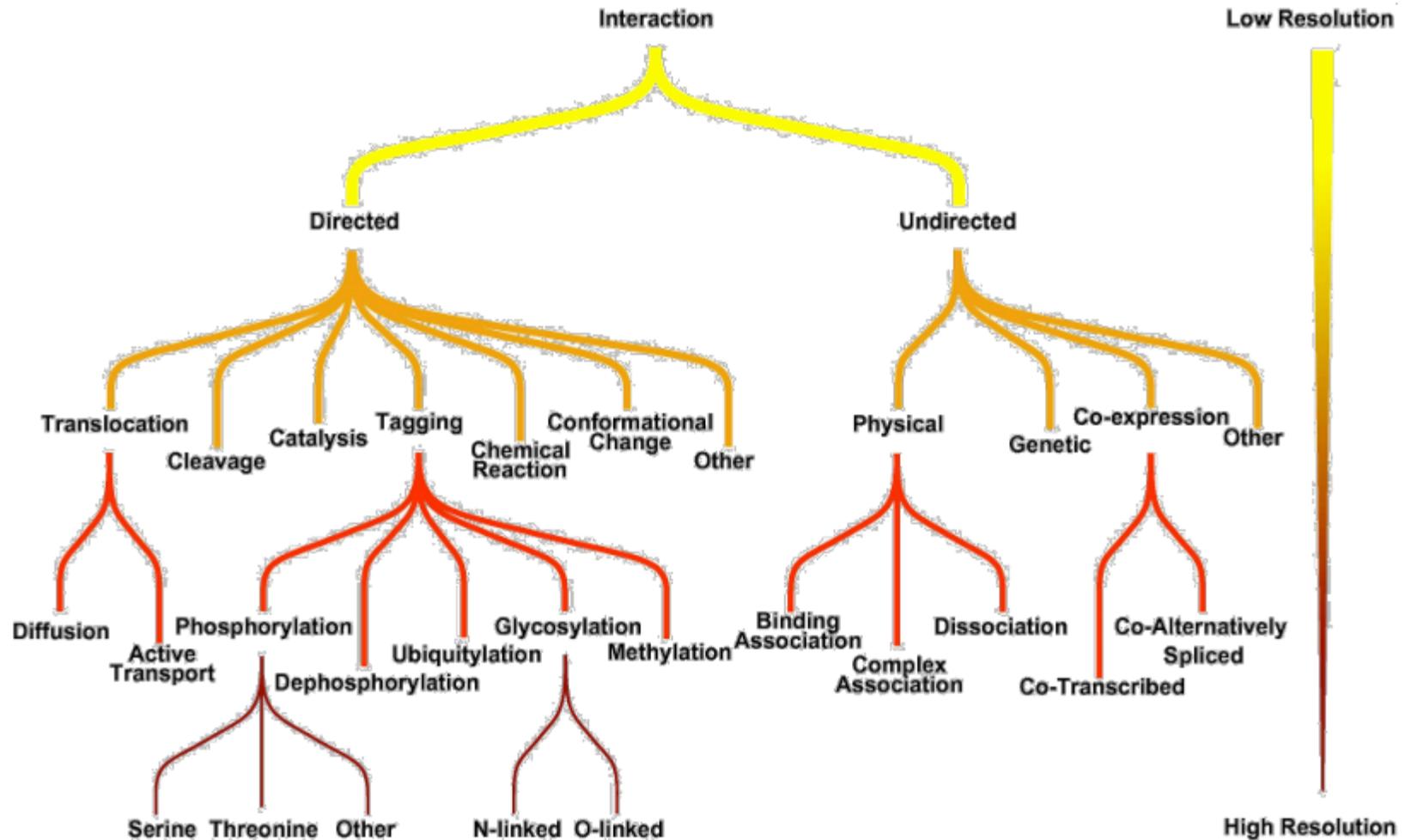


Inspira Crea Transforma

Pathways



Tipos de interacción biológica



Análisis biológicos - Difusión

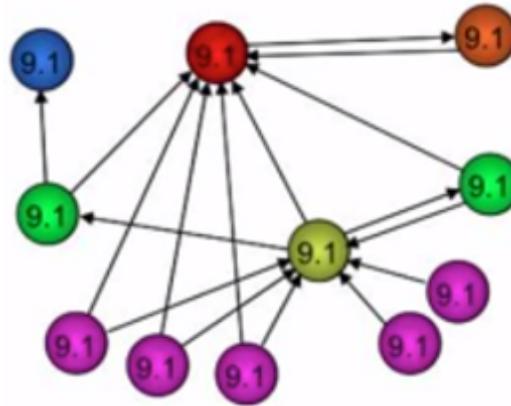
Page Rank de Google:

$PR(x)$ = Page Rank de página x

N = Número de páginas ($N=11$)

λ = posibilidad de usar un link ($\lambda=0.82$)

Inicialice $PR(x)=100\%/N \forall x$



Análisis biológicos - Difusión

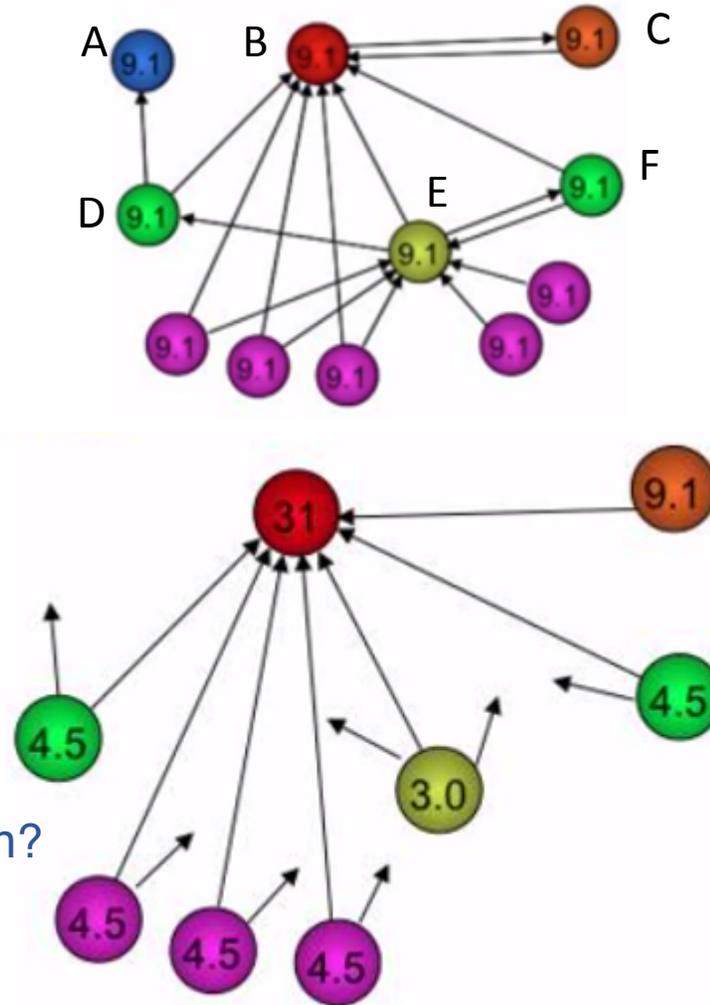
Page Rank de Google:

Primera iteración:

$$PR(x) = 1 - \lambda/N + \lambda \sum_{y \rightarrow x} \frac{PR(y)}{\text{out}(y)}$$

$$PR(B) = 0.18 \times 9.1 + 0.82 (PR(E) + PR(D)/2 + \dots)$$

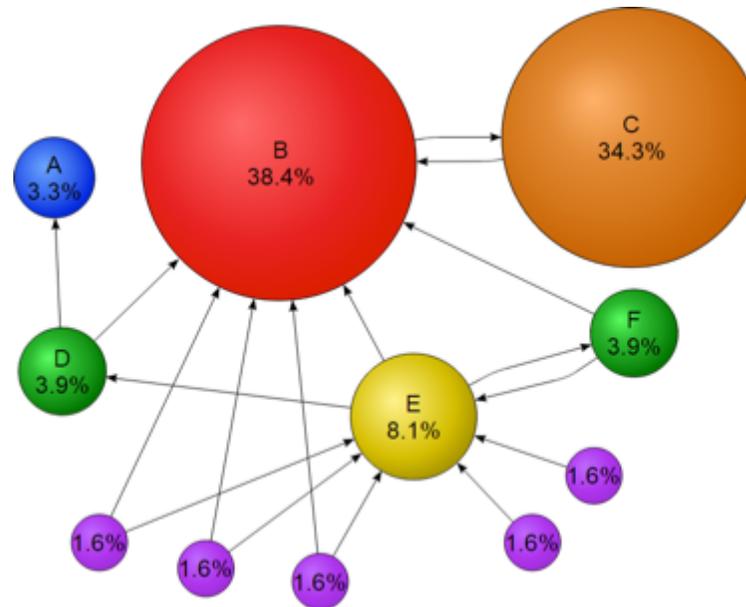
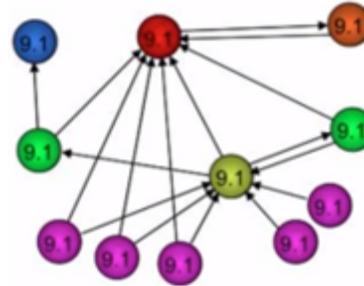
Cuál es el $PR(C)$ en la segunda iteración?



Análisis biológicos - Difusión

Page Rank de Google:

Iterar hasta que el problema converja.

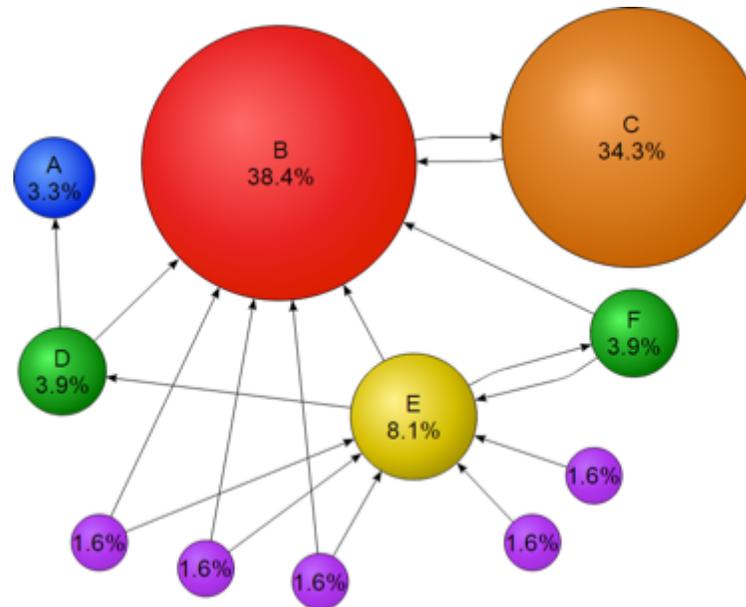
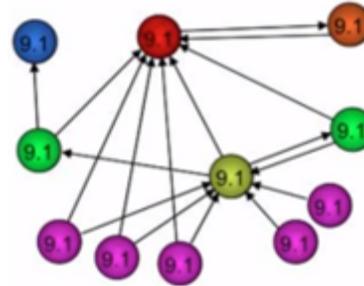


Análisis biológicos - Difusión

Page Rank de Google:

No distribuable. Alta complejidad.

Solución: Random walks.



Análisis biológicos - Difusión

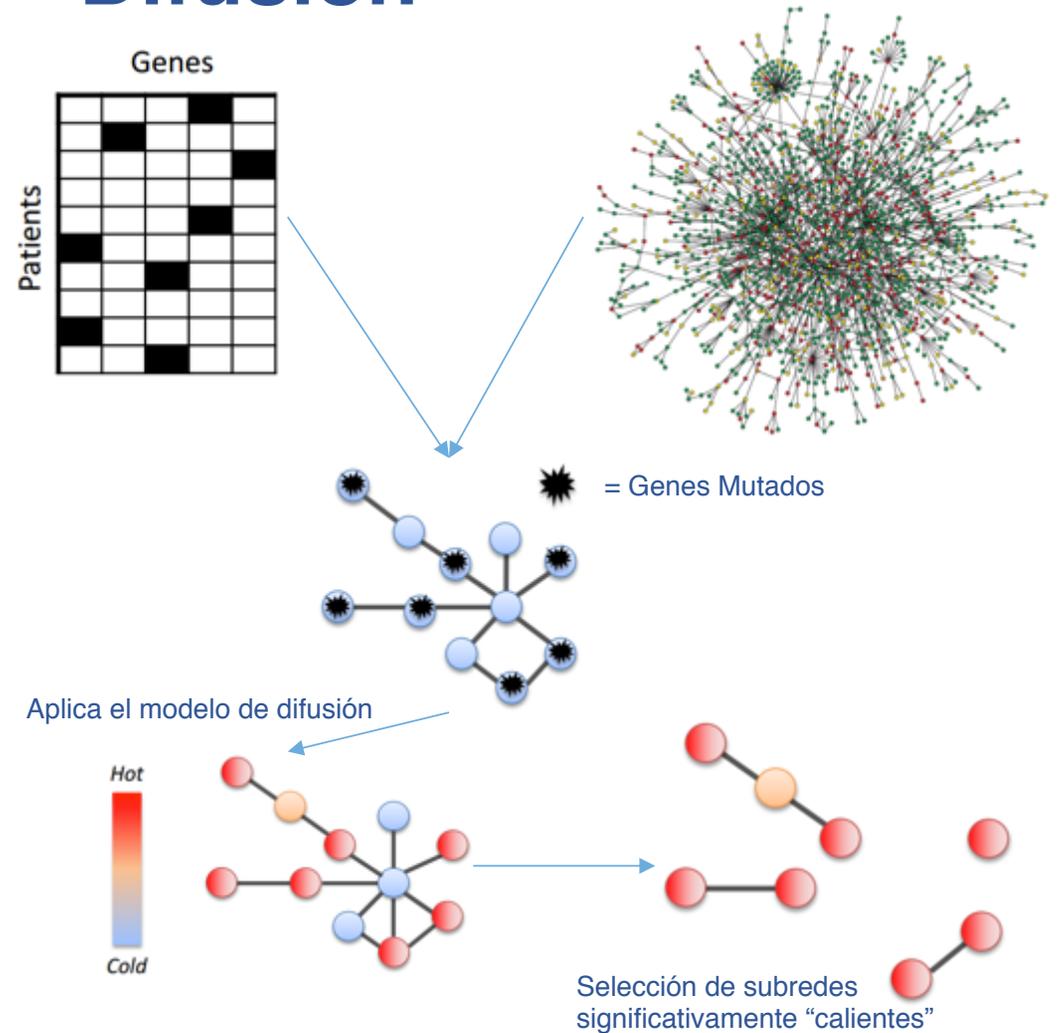
Hotnet:

Modelo de difusión de calor.

El número de mutaciones en un gen representa que tan “caliente” esta.

Un gen caliente, difunde un poco de su calor a los genes cercanos en la red.

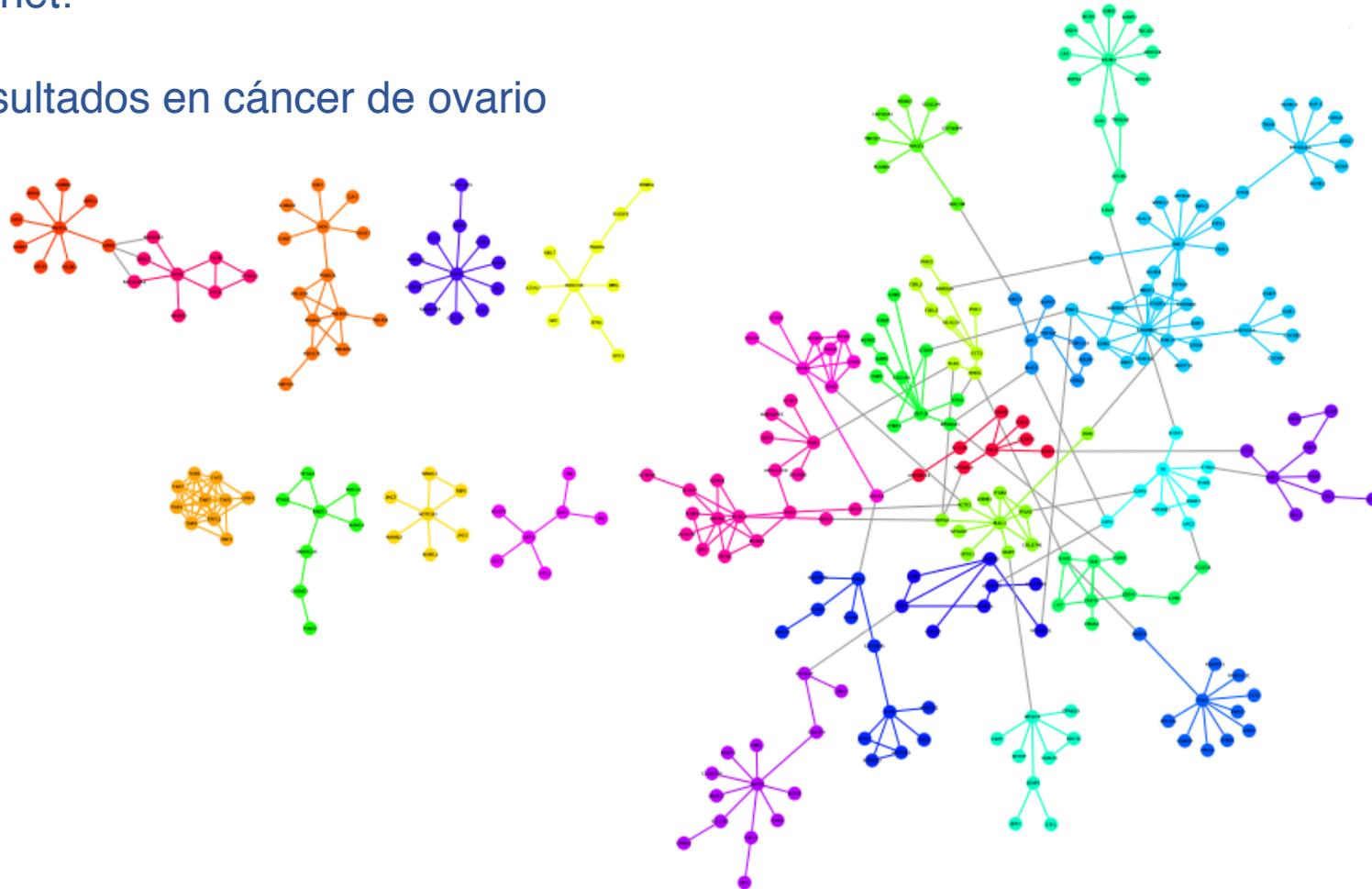
Sub redes calientes son seleccionadas como drivers.



Análisis biológicos - Difusión

Hotnet:

Resultados en cáncer de ovario

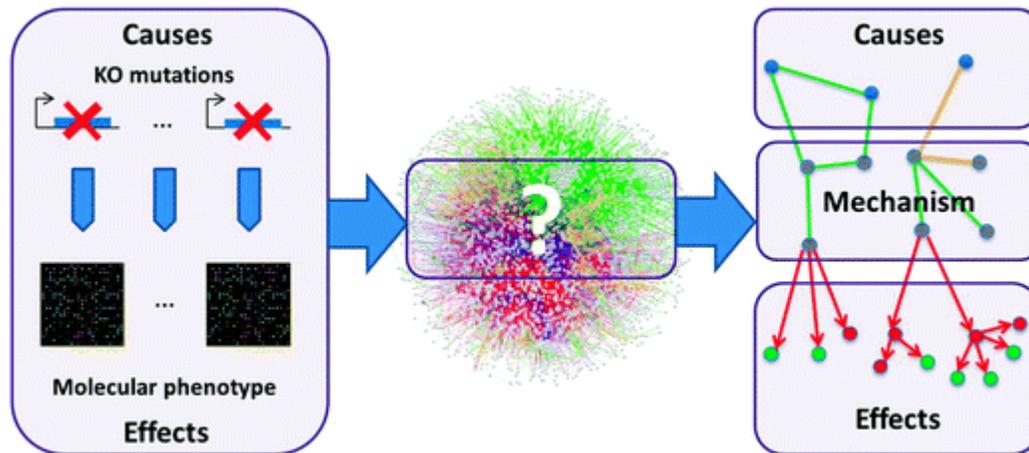


Inspira Crea Transforma

UNIVERSIDAD
EAFIT[®]

Análisis biológicos – Análisis de Vías

PheNetic: Decisión Theoretic (DT) problem.
Interpretación de listas de genes por medio de redes



Fenotipo: Resistencia al acido

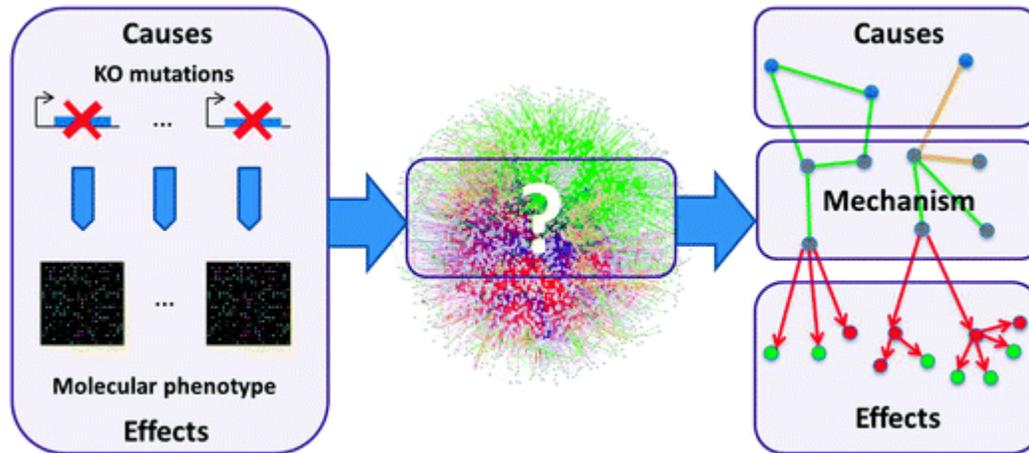
Célula: un Sistema dinámico.

KO: Perturbación al Sistema.

Expresión de Genes: Output del Sistema.

Análisis biológicos – Análisis de Vías

PheNetic: Decisión Theoretic (DT) problem.
 Interpretación de listas de genes por medio de redes



Una ganancia igual a la probabilidad por cada camino que conecte una causa x con un efecto y proporcional a la diferencia de expression del efecto $f \downarrow r$

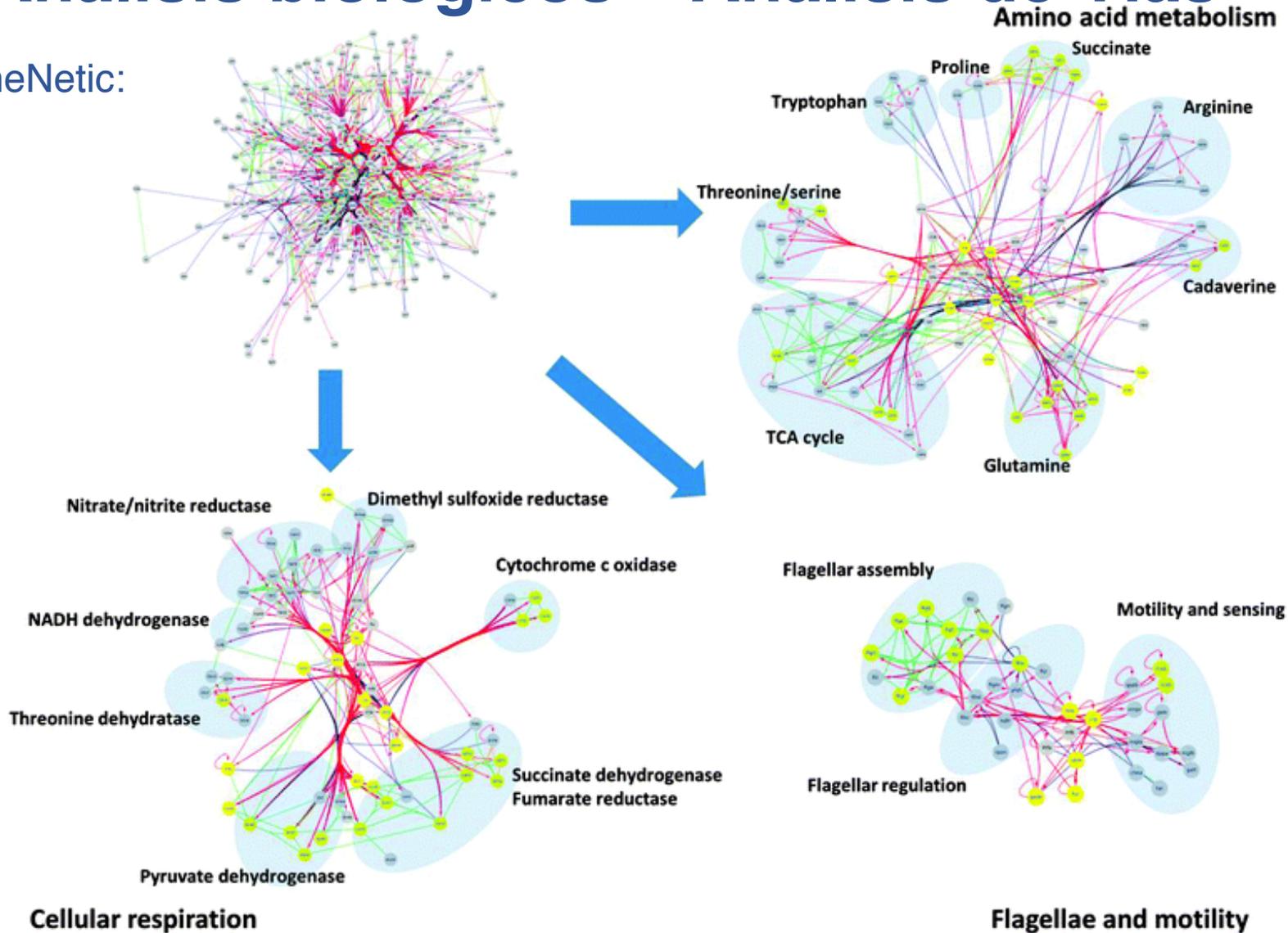
$$\max \{ \sum_{(x,y) \in I \uparrow \downarrow} f \downarrow r \times P(\text{path}(x,y) | E) - |E| x \downarrow c \}$$

Un costo $x \downarrow c$ por cada nodo en la subred E

↓
↓
 Ganancia Costo

Análisis biológicos – Análisis de Vías

PheNetic:



Inspira Crea Transforma

Flagellae and motility