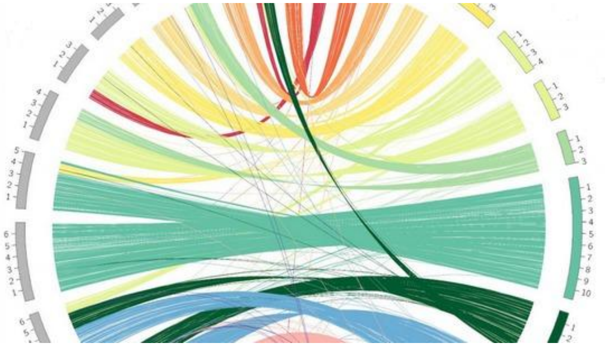
Curso Genómica Evolutiva 2021

Docentes:

Fernando Alvarez
Lucia Balestrazzi
Luisa Berna
Andres Iriarte
Hector Musto
Hector Romero



Docente invitada

Daiana Mir (Centro Universitario Regional del Litoral Norte)

Martes de 15 a 17

Jueves y Viernes de 10 a 12. Prácticos y seminarios

Aprobación del curso

- Concurrencia a los prácticos (se evalúa a través de la entrega de ejercicios)
- Presentación de un articulo en modalidad seminario

Contenido Temático de los teóricos

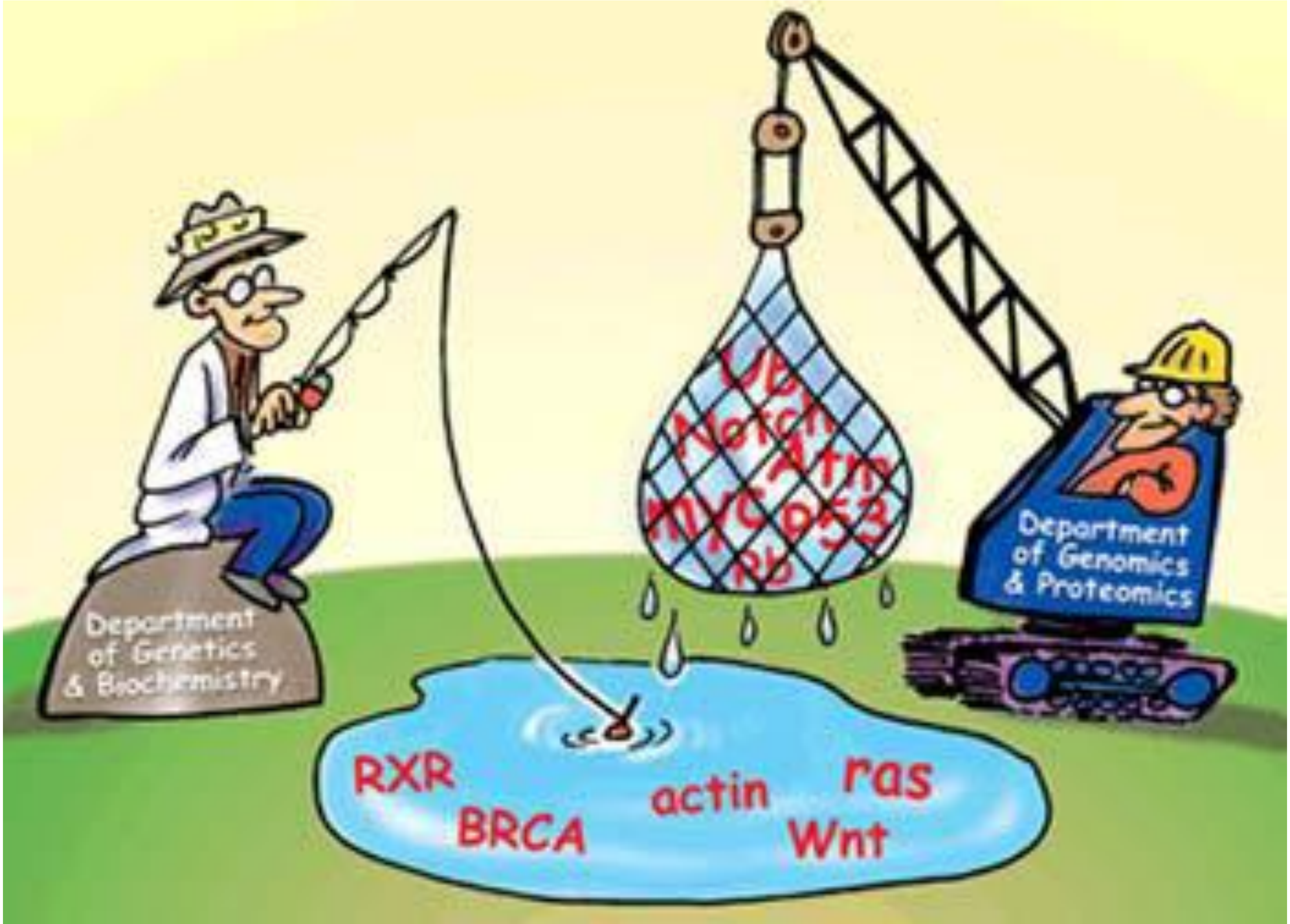
1- Evolución de la complejidad genómica
a - Tamaño genómico y complejidad.
b - Estrategias de compartimentalización genómica: desde virus y operones , [transcription Factories[, isocoros, TADs y LADs.
c - Mecanismos de transmisión de la información genética (más allá de la segregación mendeliana

2- Repetición
a – Duplicación de genoma completo y generacion de diversidad
Repetición basada en replicación del ADN.
b - Elementos móviles.
c - La repetición en las zonas codificantes. Familias multigénicas, nacimiento y muerte de genes.

3- Evolución de las endosimbiosis genómicas
a - Endosimbiosis primarias y origen de los eucariotas.
b - Endosimbiosis secundarias y evolución de eucariotas unicelulares.

4- Temas diversos
a- Origen y evolución del HIV

Genética vs Genómica



GENETICS VS. GENOMICS

GENETICS

- **Genetics** is the study of heredity.
- "Gene" refers to a specific sequence of DNA on a single chromosome.
- Genetics involves the study of functions and composition of the single gene.

GENOMICS

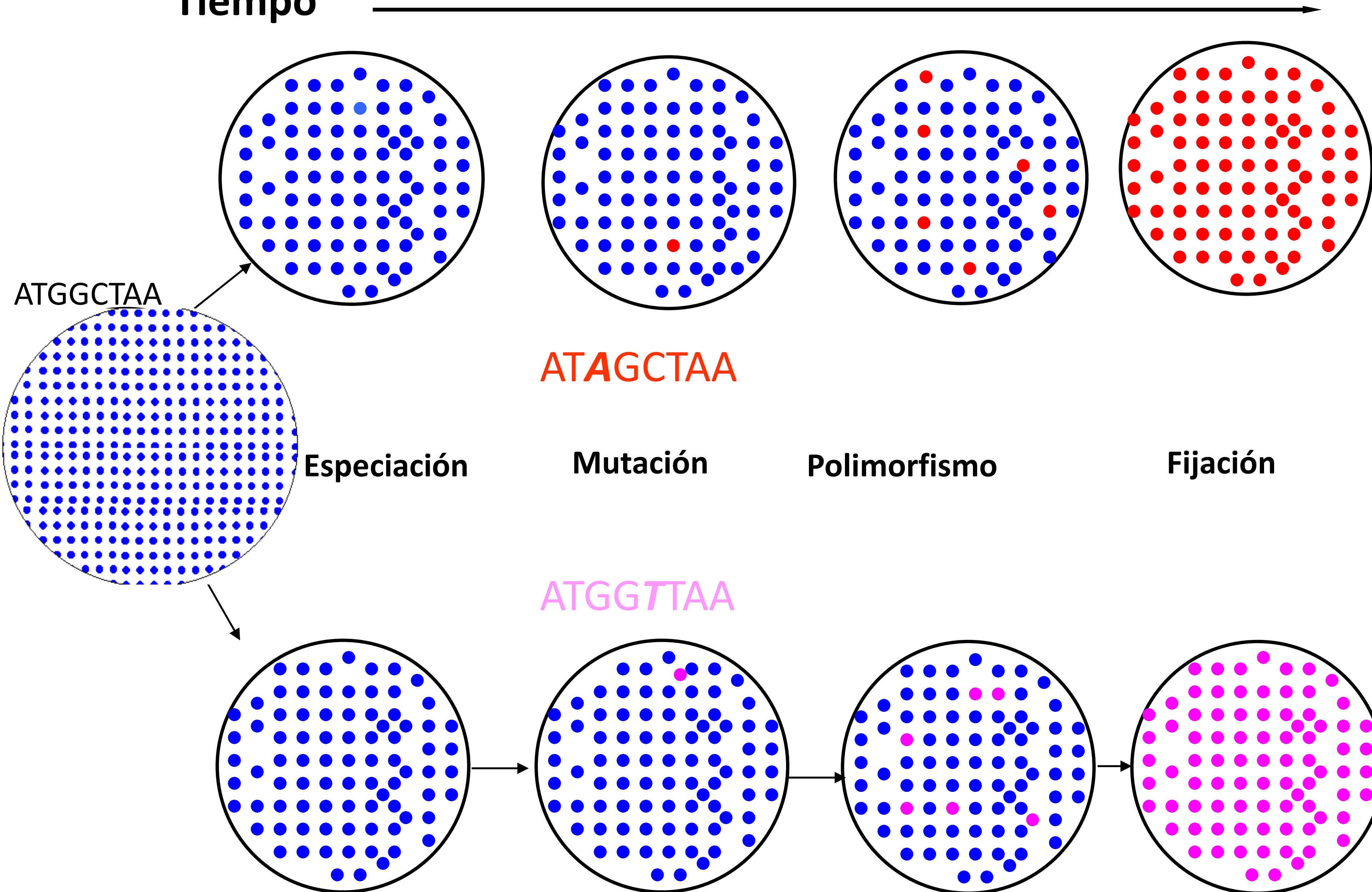
- **Genomics** is the study of the entirety of an organism's genes.
- "Genome" refers to an organism's entire genetic makeup.
- Genomics addresses all genes and their inter relationships.

Genética y Genómica de Poblaciones

- Parte central del estudio de la microevolución (evolución dentro de las especies o poblaciones).

- Es el estudio genético del proceso de la evolución: cambio en las frecuencias alélicas, genotípicas y fenotípicas.

MUTACIONES. POLIMORFISMOS Y SUBSTITUCIONES



Tiempo

ATGGCTAA

AT*A*GCTAA

Especiación

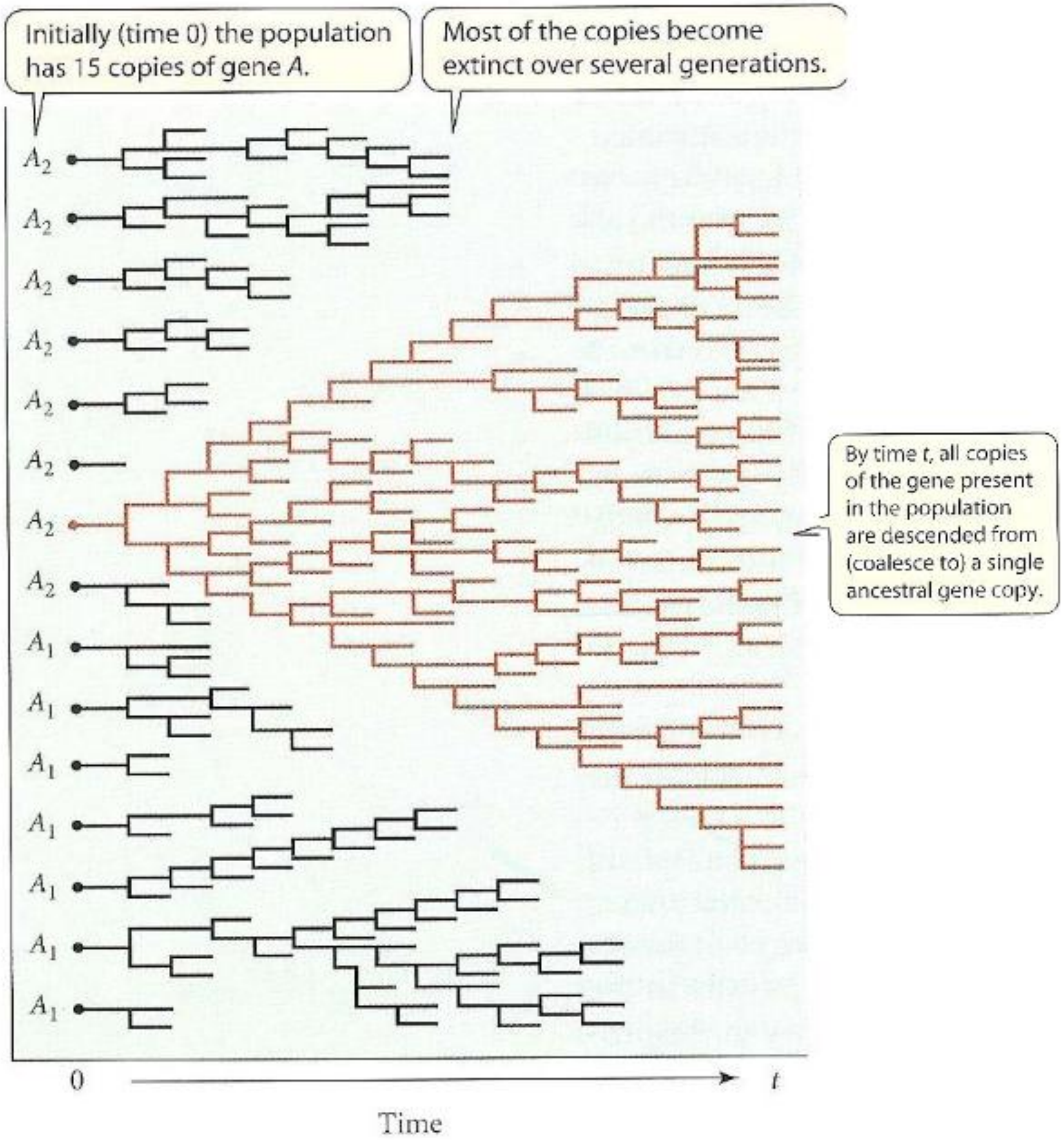
Mutación

Polimorfismo

Fijación

ATGG*T*TAA

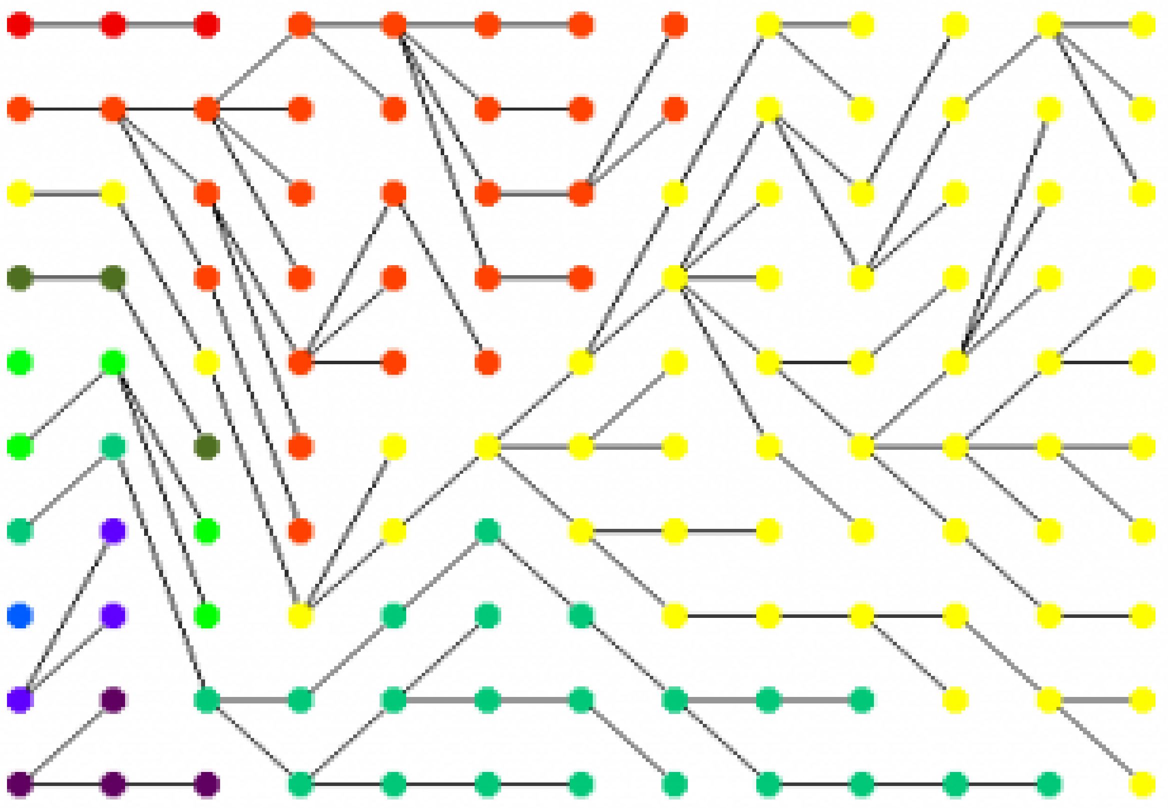
El proceso evolutivo es genealológico


Initially (time 0) the population has 15 copies of gene *A*.

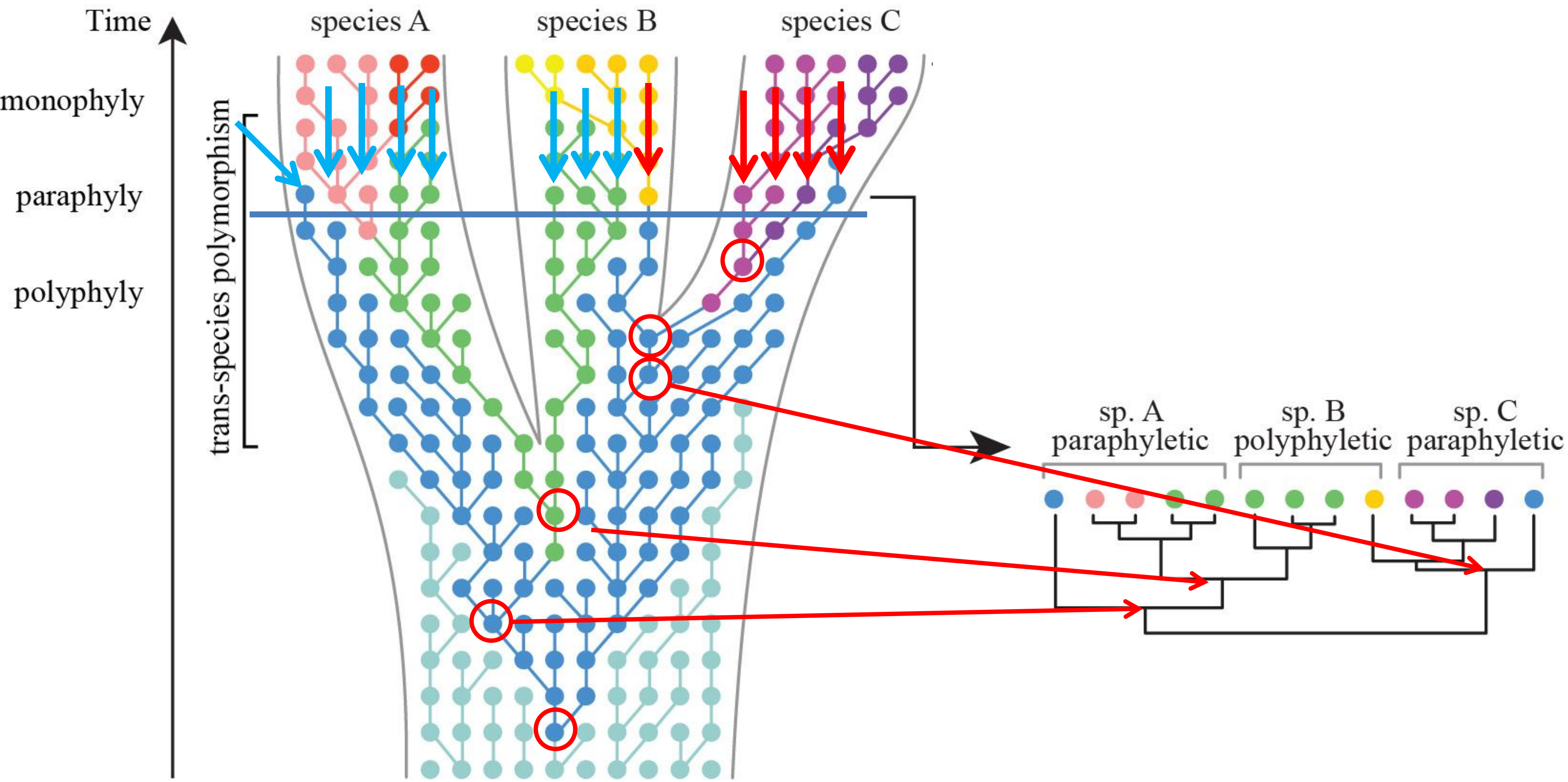
Most of the copies become extinct over several generations.

By time *t*, all copies of the gene present in the population are descended from (coalesce to) a single ancestral gene copy.

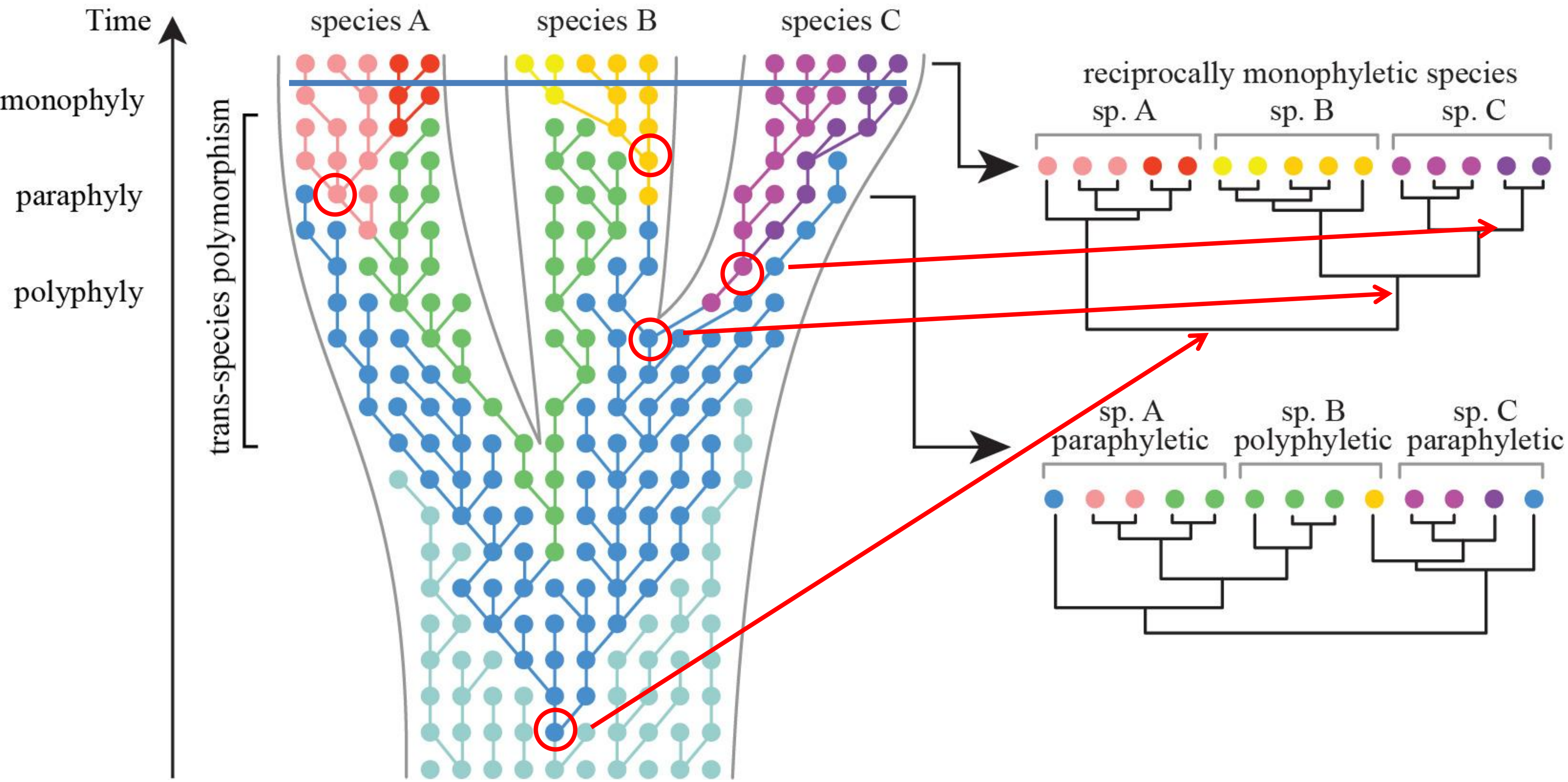
El proceso evolutivo es genealológico

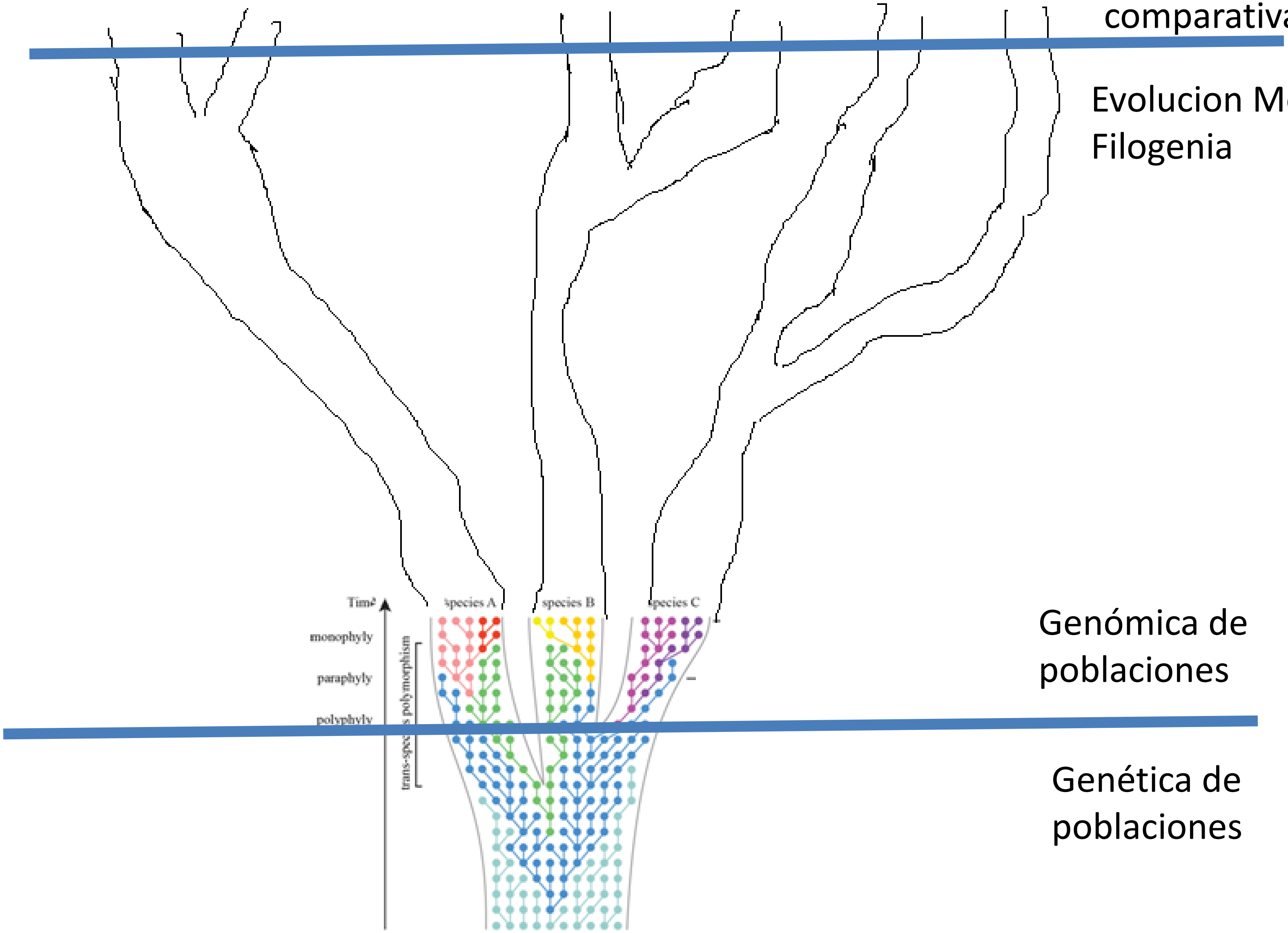


El proceso evolutivo es genealológico y de aparición e independización de lineajes



El proceso evolutivo es genealológico y de aparición e independización de lineajes





Genómica comparativa

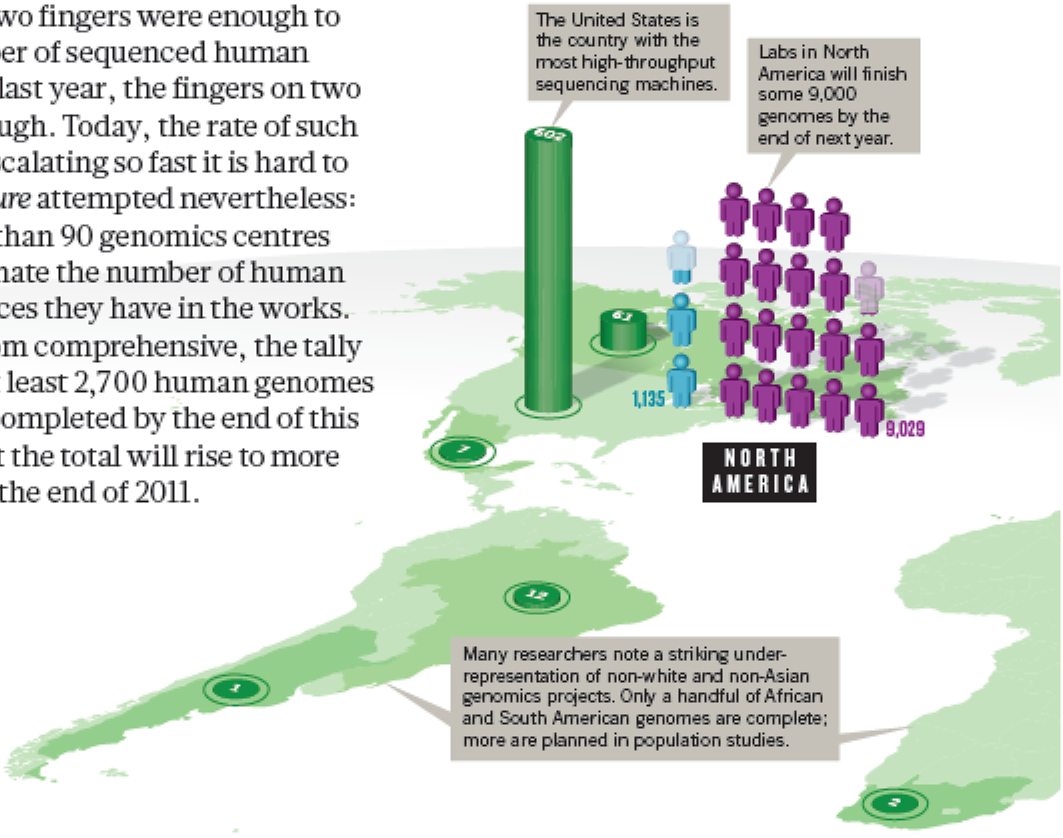
Evolucion Molecular
Filogenia

Genómica de poblaciones

Genética de poblaciones

Genomes by the thousand

Ten years ago, two fingers were enough to count the number of sequenced human genomes. Until last year, the fingers on two hands were enough. Today, the rate of such sequencing is escalating so fast it is hard to keep track. *Nature* attempted nevertheless: we asked more than 90 genomics centres and labs to estimate the number of human genome sequences they have in the works. Although far from comprehensive, the tally indicates that at least 2,700 human genomes will have been completed by the end of this month, and that the total will rise to more than 30,000 by the end of 2011.



The United States is the country with the most high-throughput sequencing machines.

Labs in North America will finish some 9,000 genomes by the end of next year.

Many researchers note a striking under-representation of non-white and non-Asian genomics projects. Only a handful of African and South American genomes are complete; more are planned in population studies.

KEY

500 genomes by the end of October 2010

500 genomes by the end of 2011

Number of high-throughput sequencers in the region

Secuenciación de ADN

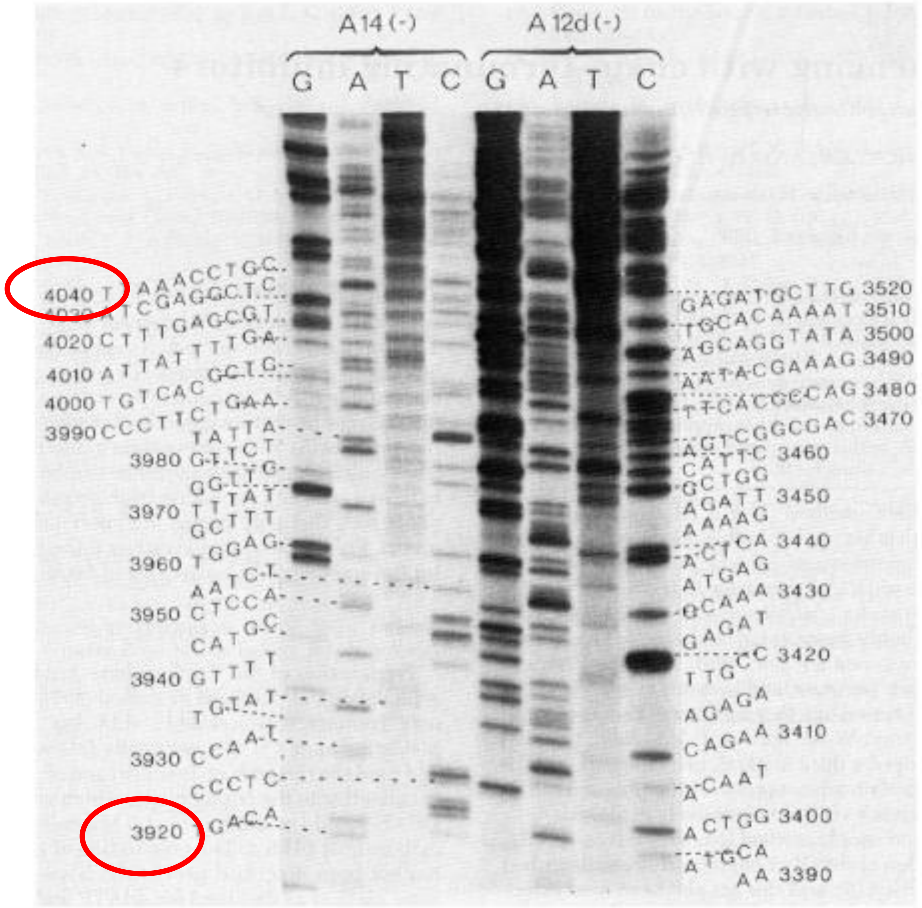
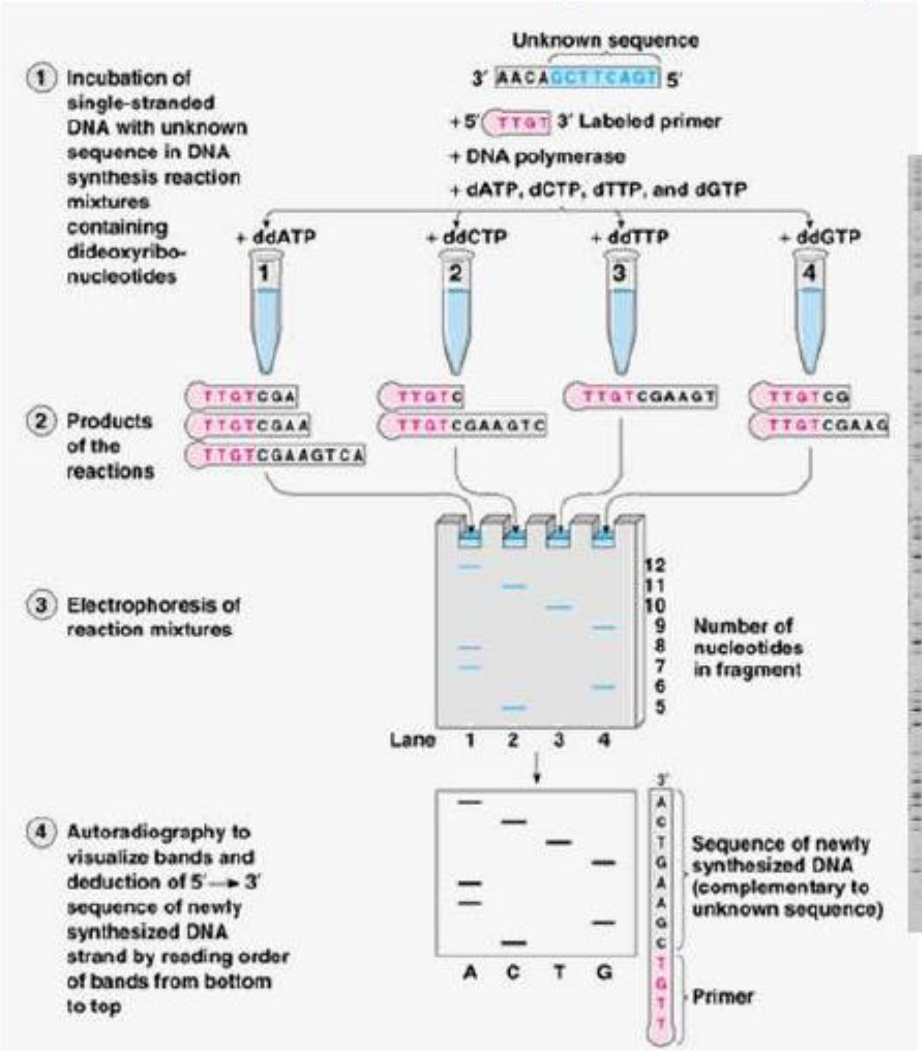
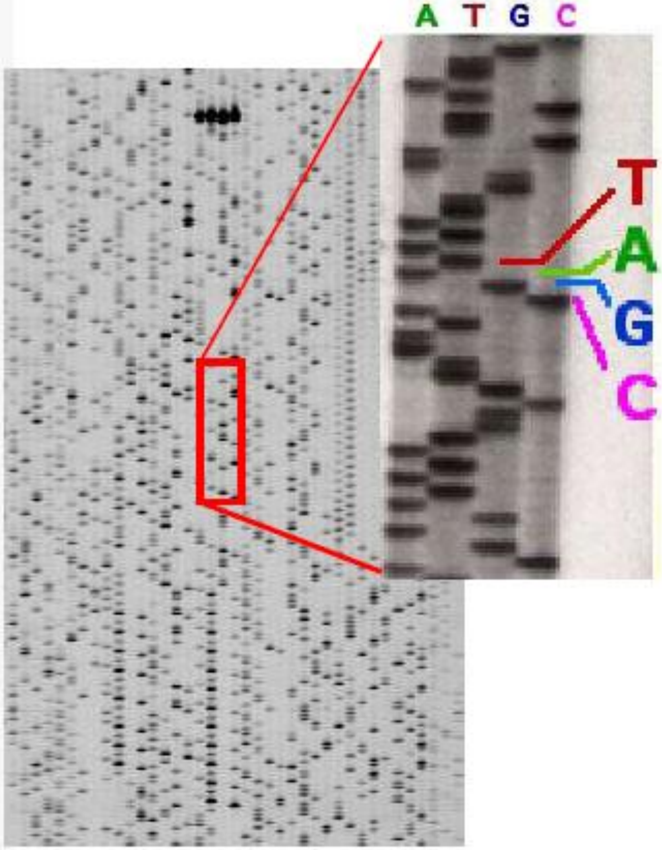


FIG. 1. Autoradiograph of the acrylamide gel from the sequence determination using restriction fragments A12d and A14 as primers on the complementary strand of φX174 DNA. The inhibitors used were (left to right) ddGTP, ddATP, ddTTP, and araCTP. Electrophoresis was on a 12% acrylamide gel at 40 mA for 14 hr. The top 10 cm of the gel is not shown. The DNA sequence is written from left to right and upwards beside the corresponding bands on the radioautograph. The numbering is as given in ref. 2.

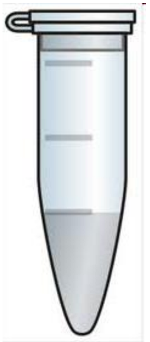
Región secuenciada:
4040-3920 = 120 bases

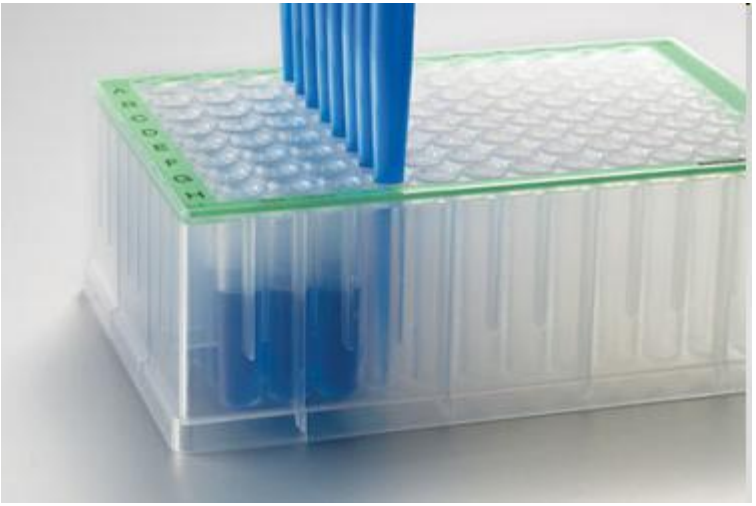
Sanger sequencing





Paralelización en el laboratorio



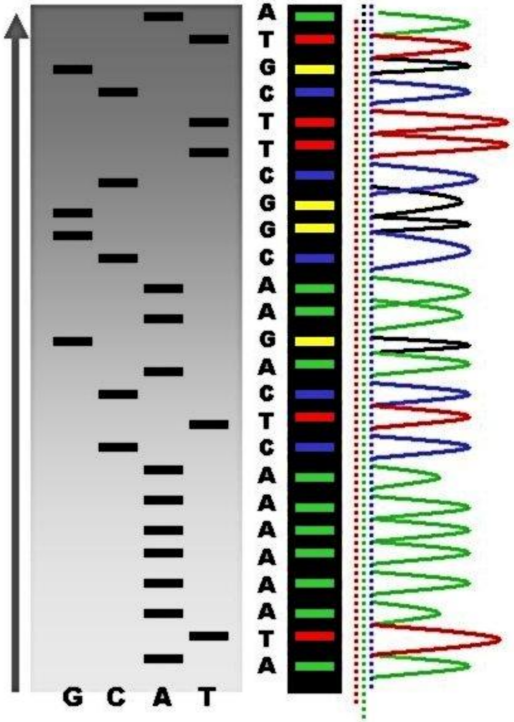


Fluorimetria y Secuenciación

1986:

Aporte de la fluorimetría a la secuenciación de ácidos nucleicos.

Uso de ddNTP's unidos a moléculas fluorescentes



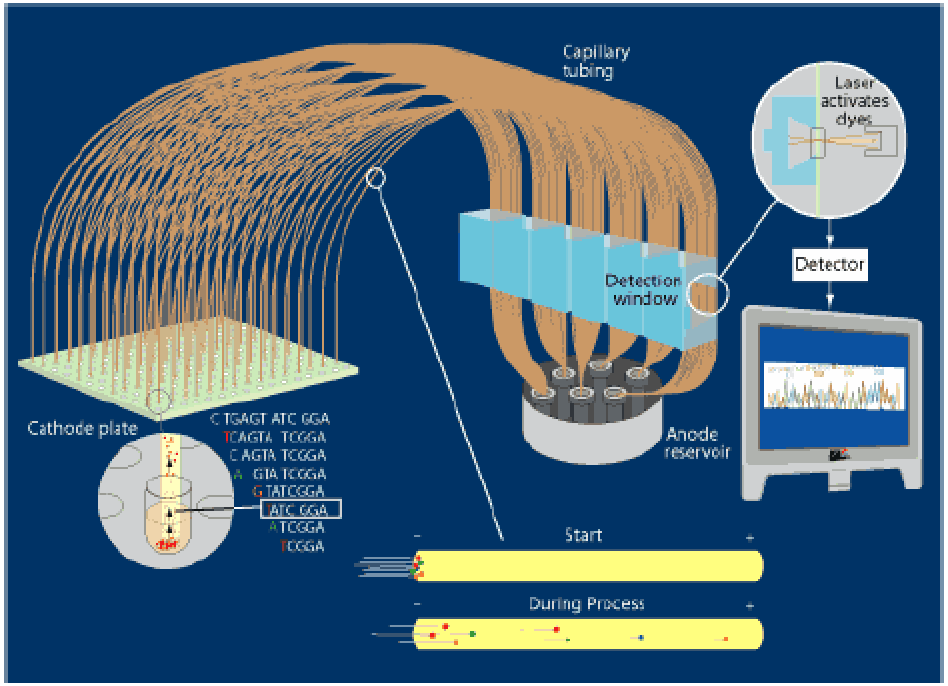
Requerimientos:
Molde a secuenciar
Primer
dNTP's
ddNTP's (cada uno con Dye diferente)

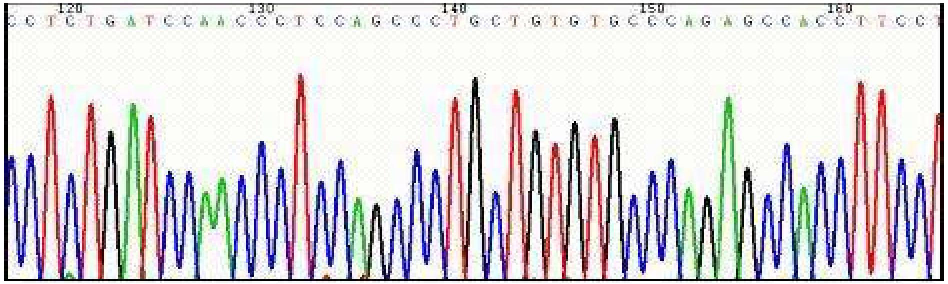
1 reacción

Secuenciador Automático

Nature **321**, 674-679 (12 June 1986)

Sanger sequencing



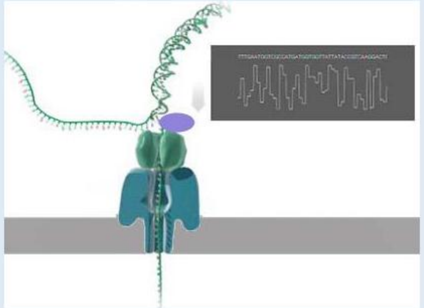


- Can process 96 samples a time
- Increase in automation
- 800 bp read lenght
- 1,6 Mbp/day

3.000.000.000 / 1.600.000 = **1.875 days!**

Años 90 (finales)
Applied Biosystems
ABI 3730XL
Secuenciador 96 capilares
1 Mb / día

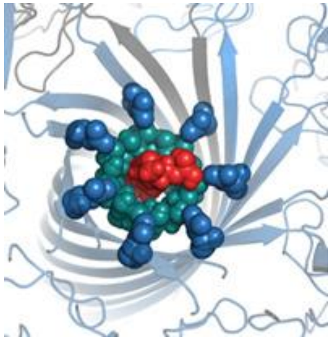
Nanopore -> *De novo* Assembly

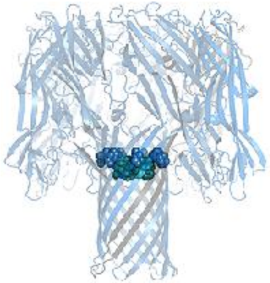


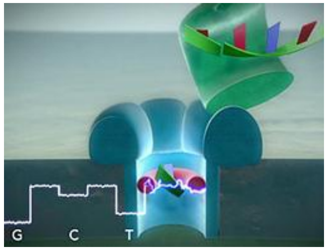
Input DNA	20 – 10 000 ng 10-100kb
Output	Long-read 10-200 kb (ideal coverage?) http://lab.loman.net/2017/03/09/ultrareads-for-nanopore/ -> interesting protocol discussion
Cost	1 run - 1-15Gb – 1000$ ⇒ ~50x for a 100Mb genome
Advantages	Continuous single long molecule (Portable + real time)
Drawbacks	High base-pair error rate (10-15%) False deletions/ homopolymer errors
Bioinformatics	Specialized assembly algorithm - Hybrid error correction (with short-read) (NanoCorr, Spades) if <30x - Self-correction (Canu) ->better if enough coverage - Rapid assembly without correction (Miniasm/Minimap) but need final polishing Polishing: Nanopolish Scaffolding: LINKS

Oxford Nanopore

α-hemolysin *Staphylococcus aureus*
exonuclease





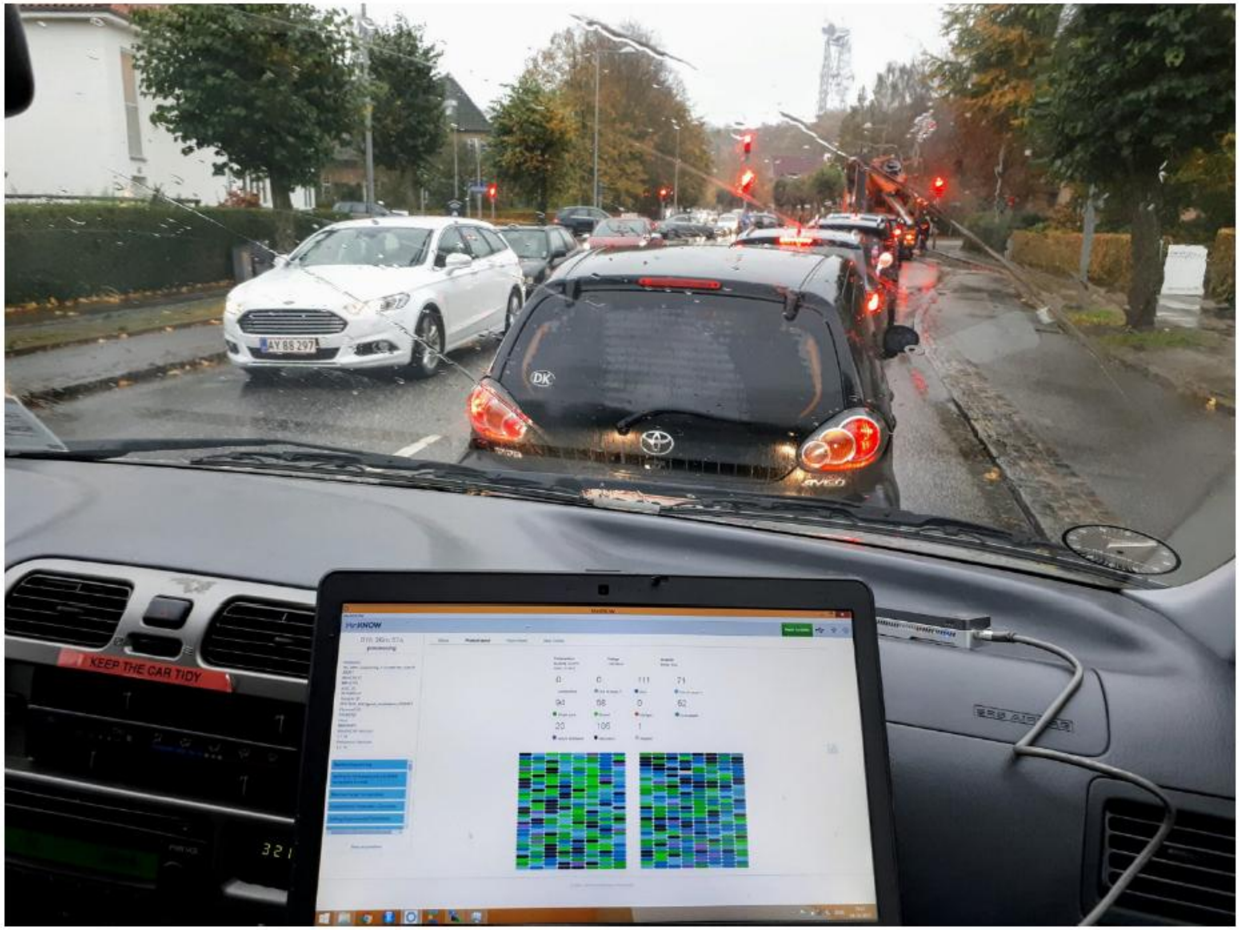


Secuenciación

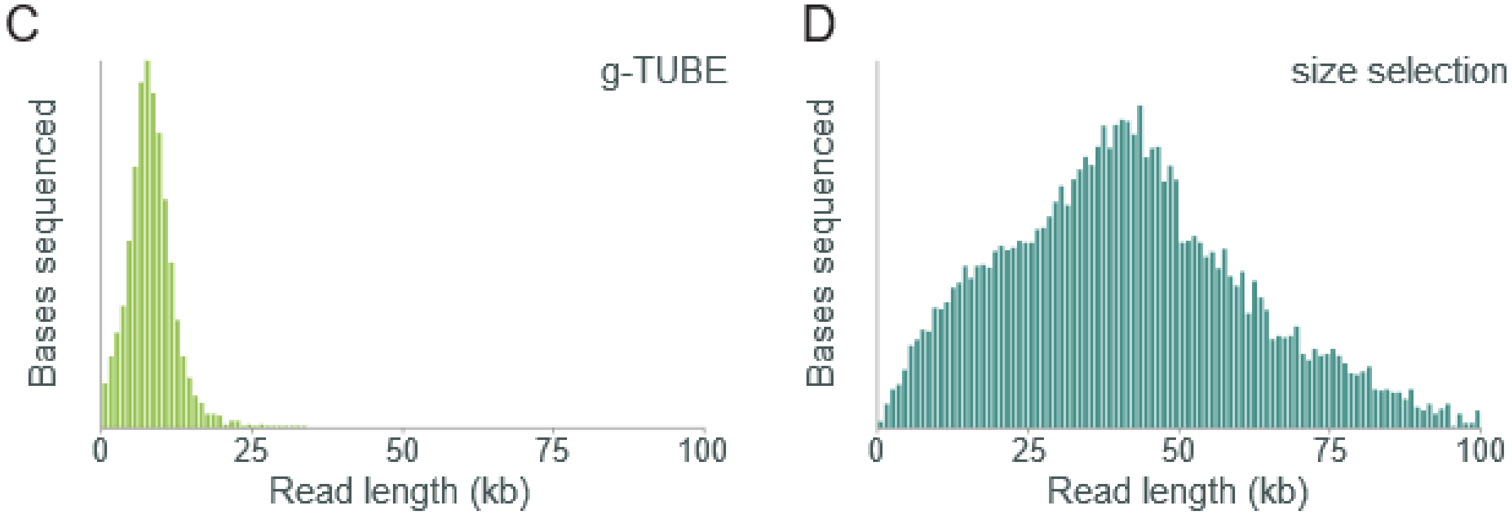
Oxford Nanopore Tech:

Secuenciación

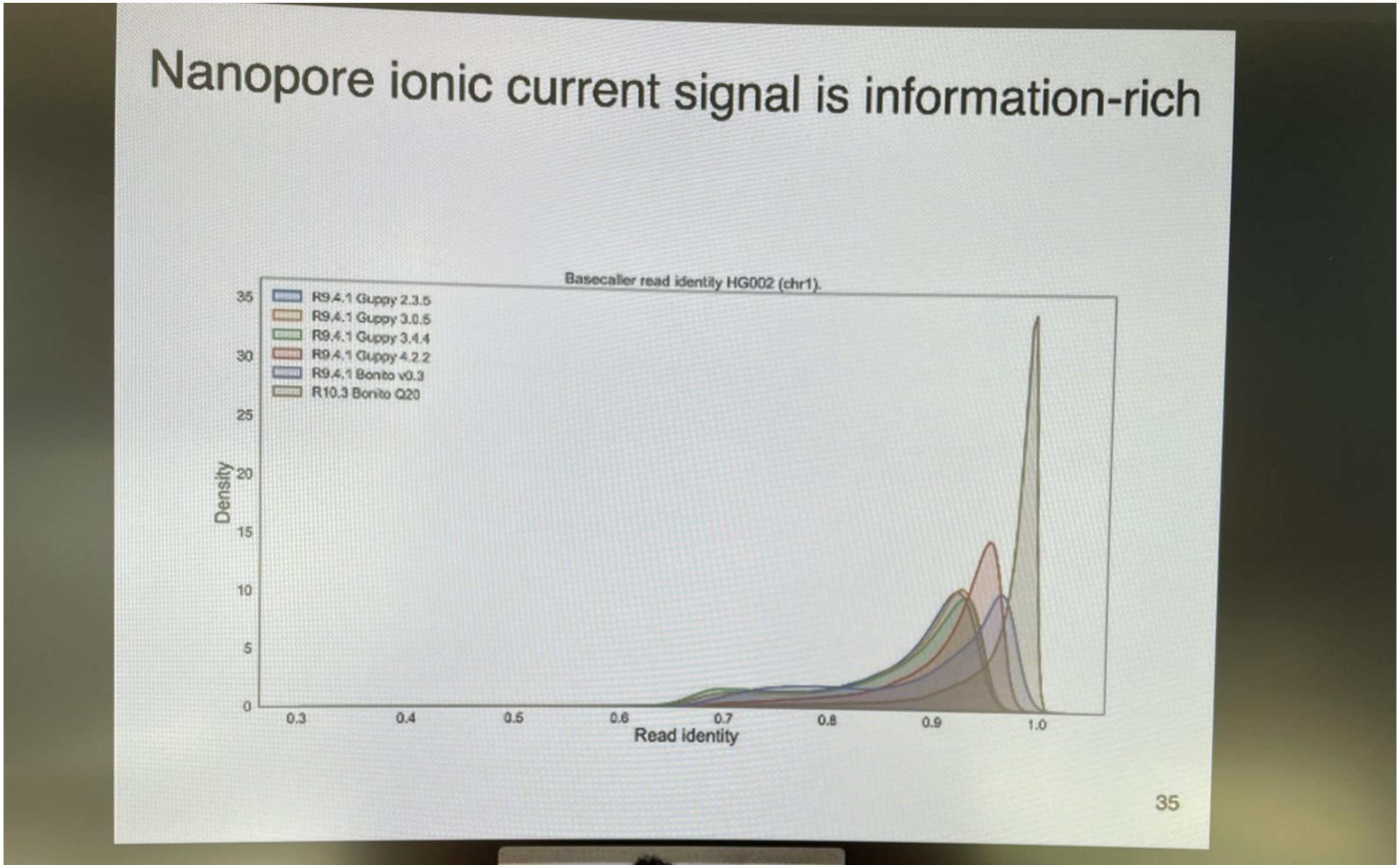
Oxford Nanopore Tech:



How can we go from long contigs level to fully resolved genomes?



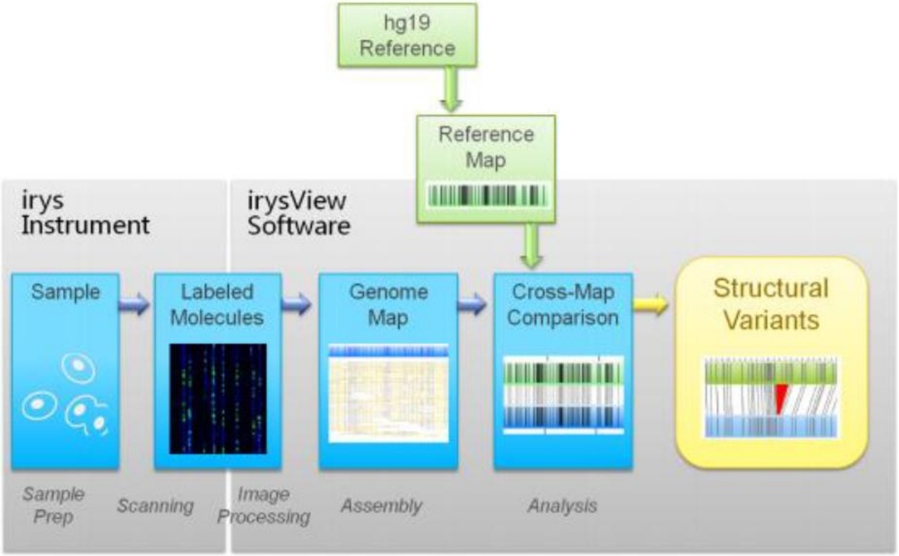
Incremento de la precision de Nanopore



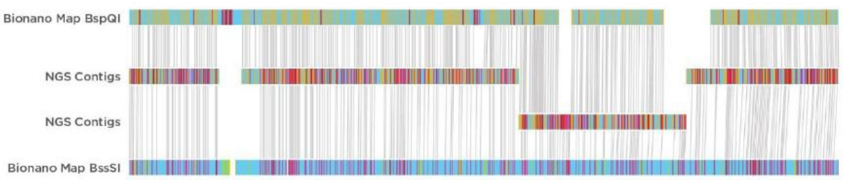
Nanopore ionic current signal is information-rich

BioNano optical mapping -> Scaffolding

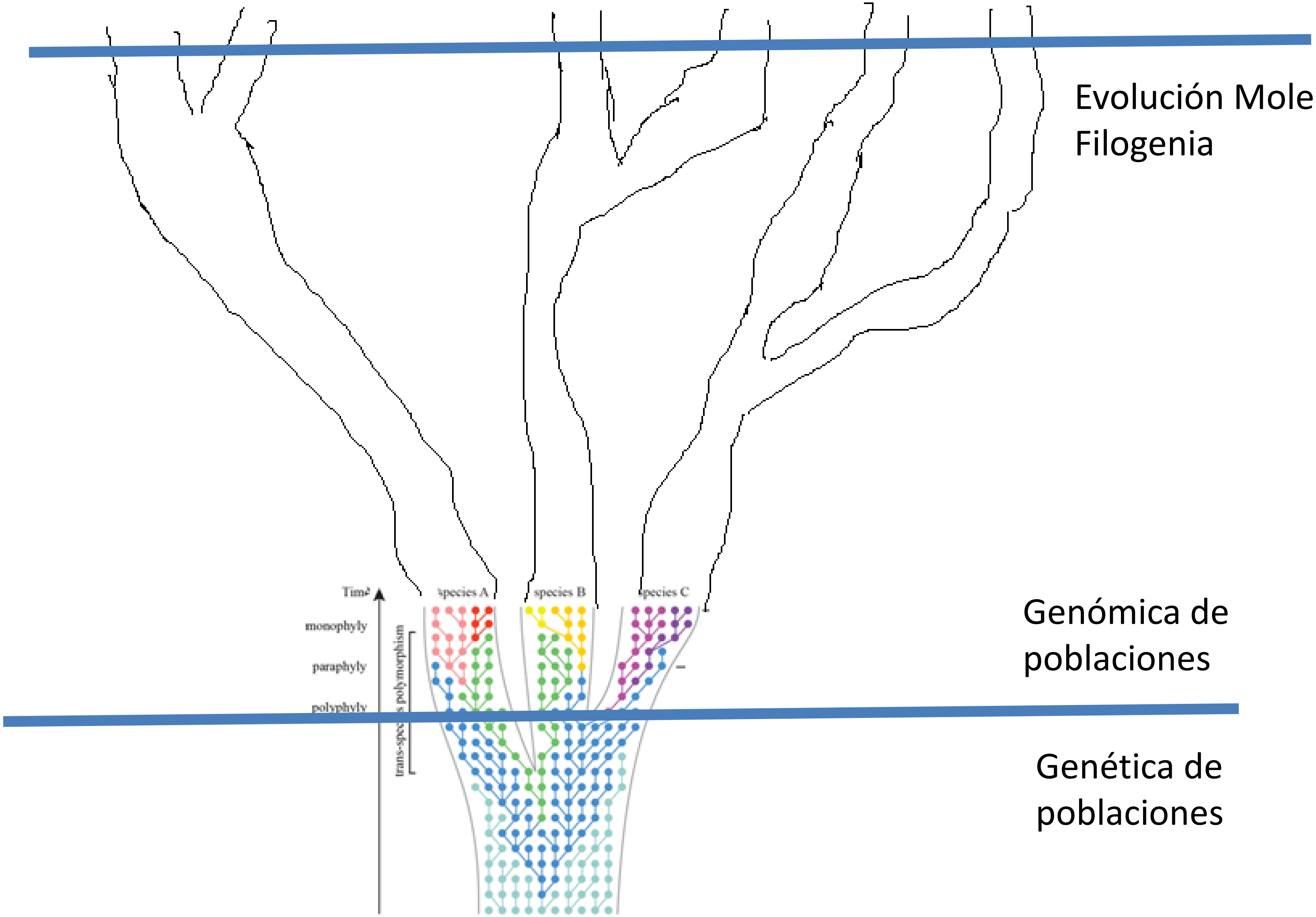
Input DNA	Fresh tissue for 500kb-1Mb DNA fragments
Output	Scaffolded assembly
Cost	8 000$ for 500Mb genome…? (all done at facilities)
Advantages	Estimate gap size
Drawbacks	Need a previous assembly (N50 >50-500kb)
Bioinformatics	Scaffolding: BioNanoAccess



TWO ENZYME HYBRID SCAFFOLDING



Ejemplos de estudios de genómica evolutiva



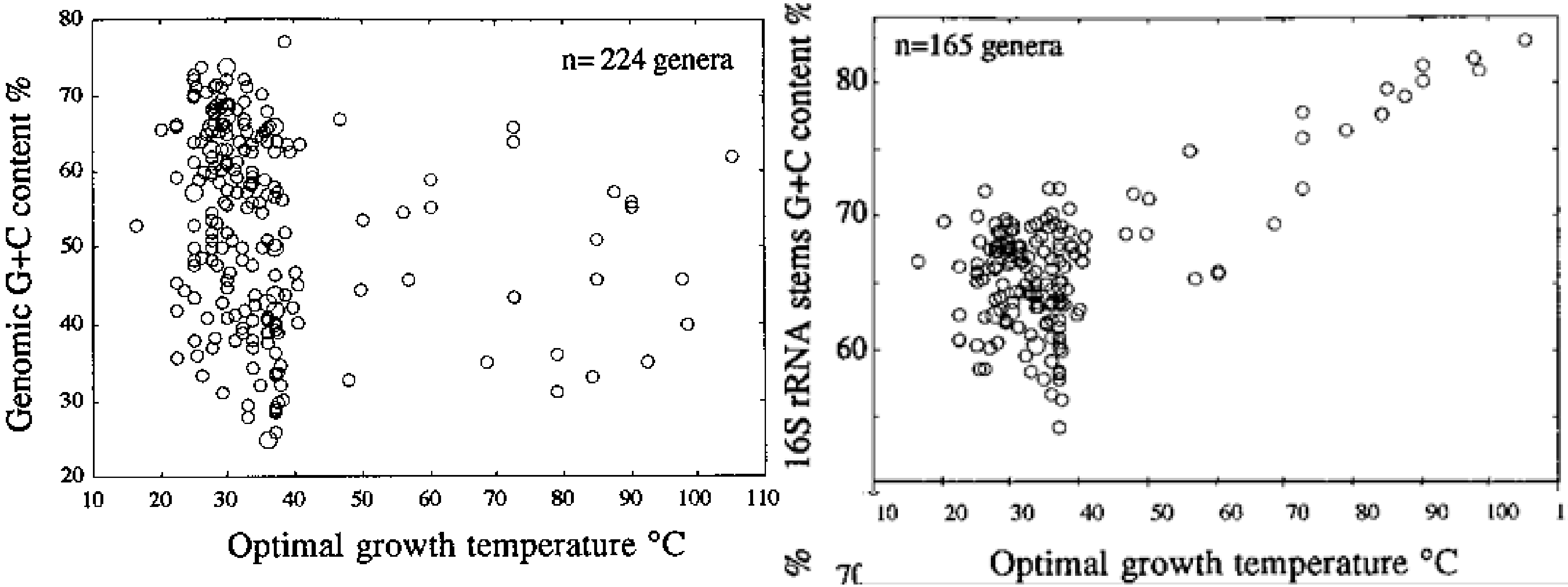
Genómica comparativa

Evolución Molecular
Filogenia

Genómica de
poblaciones

Genética de
poblaciones

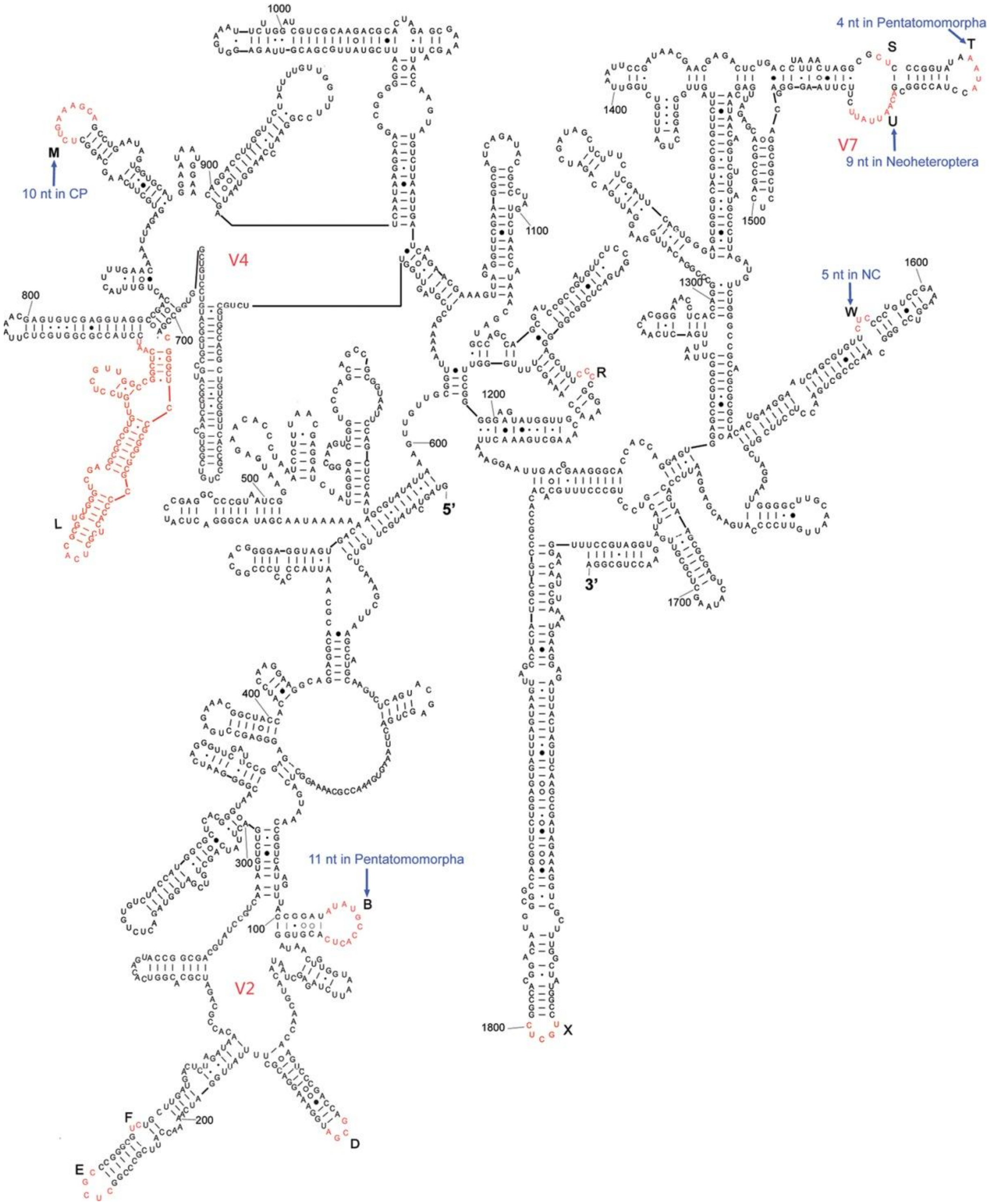
Genómica comparativa



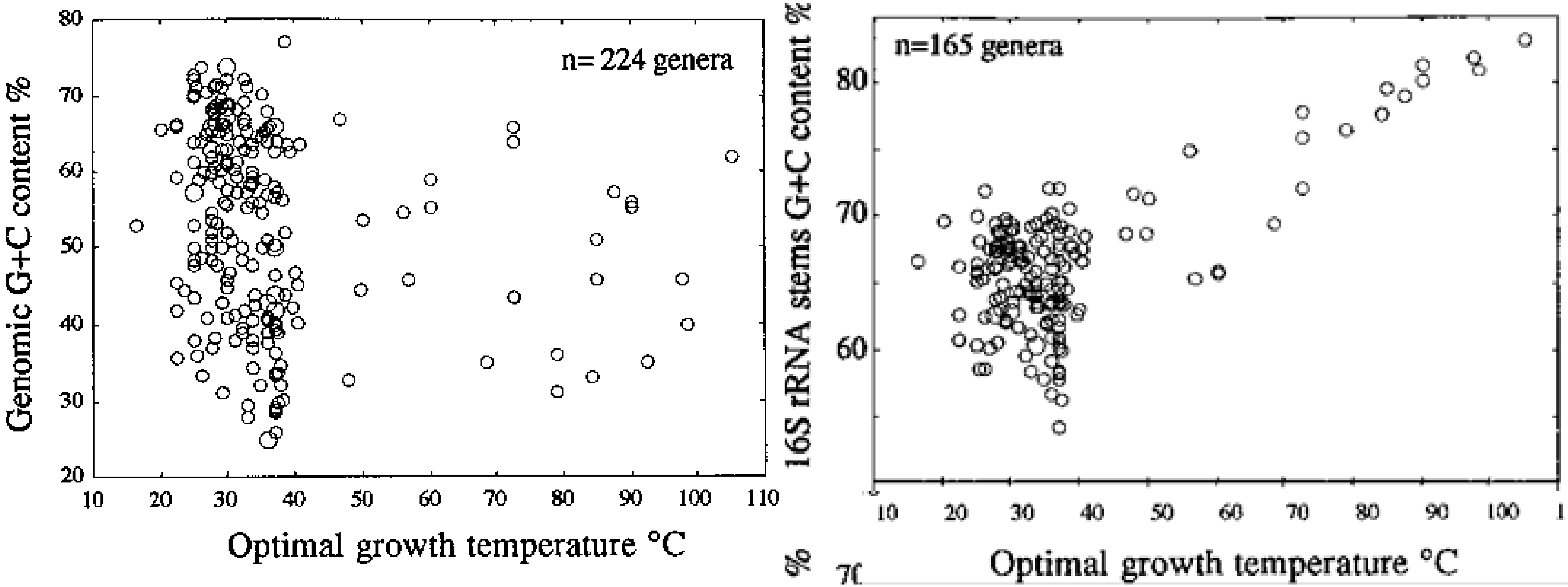
[Relationships Between Genomic G+C Content, RNA Secondary Structures, and Optimal Growth Temperature in Prokaryotes]

Nicolas Galtier, Jean R. Lobry

Estructura secundaria de ARN 16S



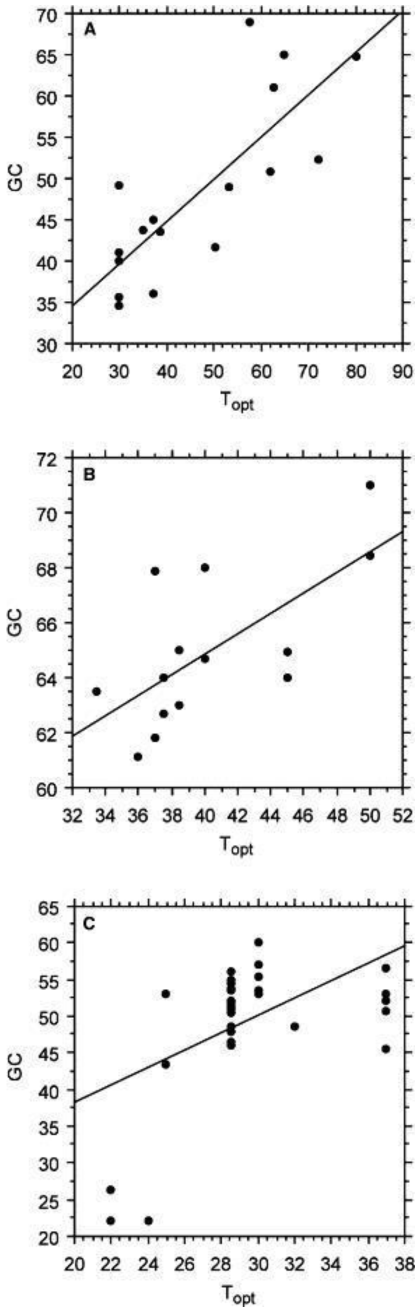
Genómica comparativa



[Relationships Between Genomic G+C Content, RNA Secondary Structures, and Optimal Growth Temperature in Prokaryotes]()

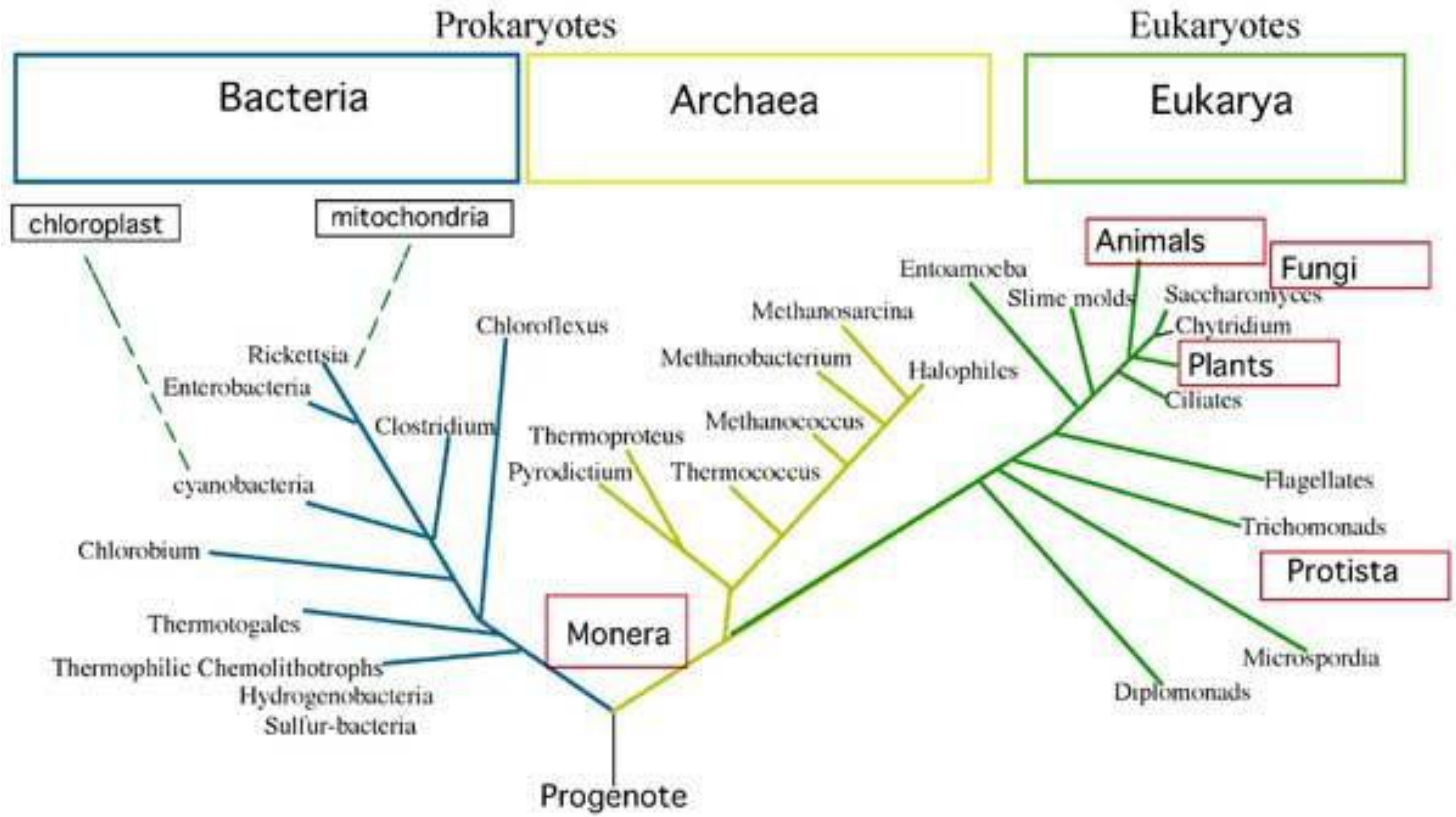
Nicolas Galtier, Jean R. Lobry

Genómica comparativa

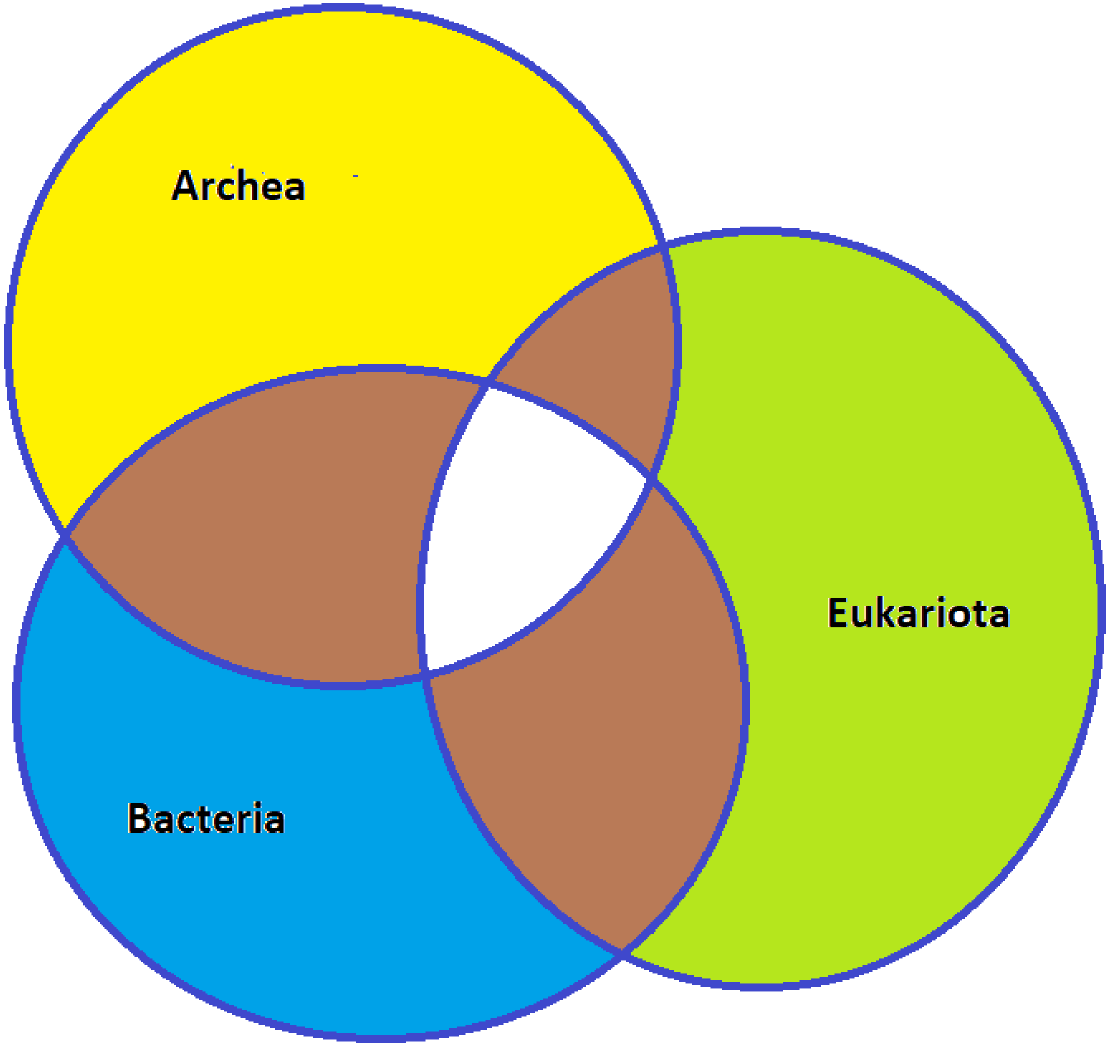


Correlations between genomic GC levels and optimal growth temperatures in prokaryotes
Héctor Musto et al, 2004

Genómica comparativa : core y pangenomes



Genómica comparativa : core y pangenomes



Genómica de poblaciones

LETTER

doi:10.1038/nature11837

The genomic signature of dog domestication reveals adaptation to a starch-rich diet

Erik Axelsson[1], Abhirami Ratnakumar[1], Maja-Louise Arendt[1], Khurram Maqbool[1], Matthew T. Webster[1], Michele Perloski[2], Olof Liberg[3], Jon M. Arnemo[4,5], Åke Hedhammar[6] & Kerstin Lindblad-Toh[1,2]

Secuenciación y Mapeo

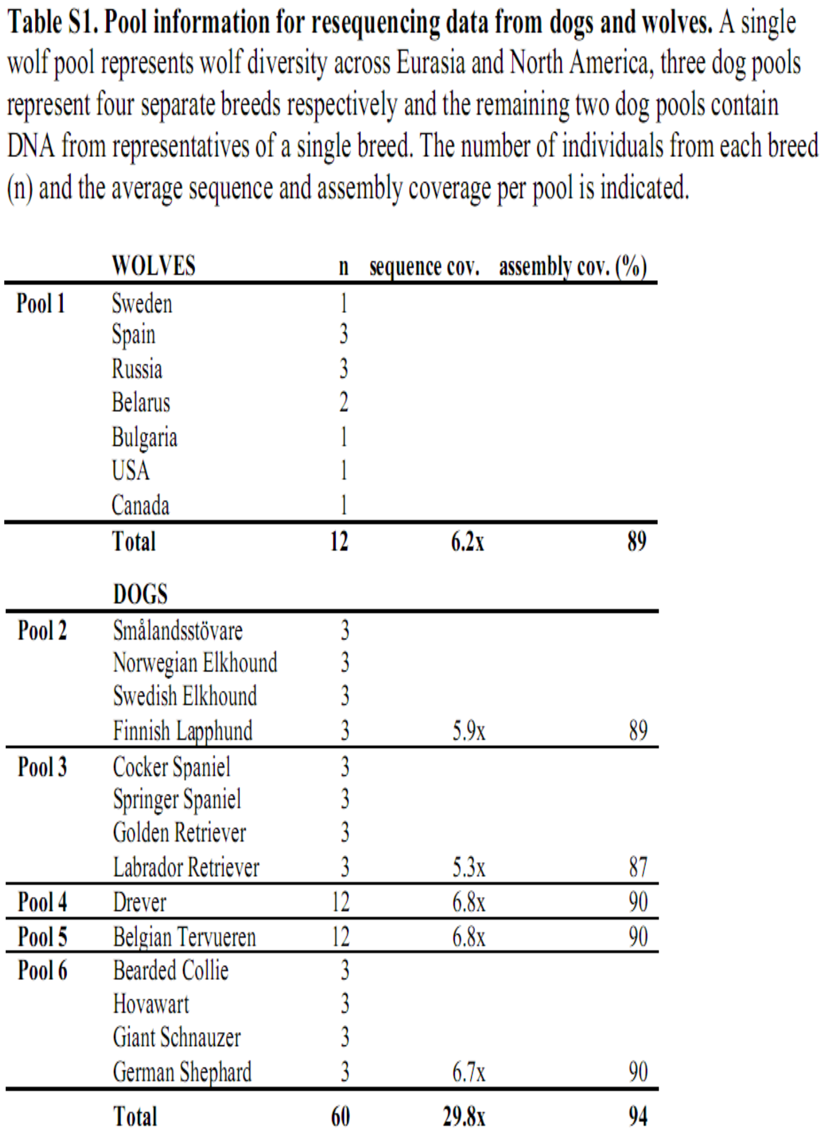


Table S1. Pool information for resequencing data from dogs and wolves. A single wolf pool represents wolf diversity across Eurasia and North America, three dog pools represent four separate breeds respectively and the remaining two dog pools contain DNA from representatives of a single breed. The number of individuals from each breed (n) and the average sequence and assembly coverage per pool is indicated.

	WOLVES	n	sequence cov.	assembly cov. (%)
Pool 1	Sweden	1		
	Spain	3		
	Russia	3		
	Belarus	2		
	Bulgaria	1		
	USA	1		
	Canada	1		
	Total	**12**	**6.2x**	**89**
	DOGS			
Pool 2	Smålandsstövare	3		
	Norwegian Elkhound	3		
	Swedish Elkhound	3		
	Finnish Lapphund	3	5.9x	89
Pool 3	Cocker Spaniel	3		
	Springer Spaniel	3		
	Golden Retriever	3		
	Labrador Retriever	3	5.3x	87
Pool 4	Drever	12	6.8x	90
Pool 5	Belgian Tervueren	12	6.8x	90
Pool 6	Bearded Collie	3		
	Hovawart	3		
	Giant Schnauzer	3		
	German Shephard	3	6.7x	90
	Total	**60**	**29.8x**	**94**

- Hay evidencias que sugieren que la domesticación comenzó hace 33000 años
- Los primeros registros seguros son de hace 11000-12000 años
- Causas de la domesticación:
  - Captura de cachorros de lobos para ayudar en la protección y la caza
  - Alternativamente los lobos se pueden haber acercado solos a los asentamientos humanos para comer los desechos
- Agresividad reducida y cambios de comportamiento social

Genómica de poblaciones


LETTER

doi:10.1038/nature11837

The genomic signature of dog domestication reveals adaptation to a starch-rich diet

Erik Axelsson[1], Abhirami Ratnakumar[1], Maja-Louise Arendt[1], Khurram Maqbool[1], Matthew T. Webster[1], Michele Perloski[2], Olof Liberg[3], Jon M. Arnemo[4,5], Åke Hedhammar[6] & Kerstin Lindblad-Toh[1,2]

Secuenciación y Mapeo

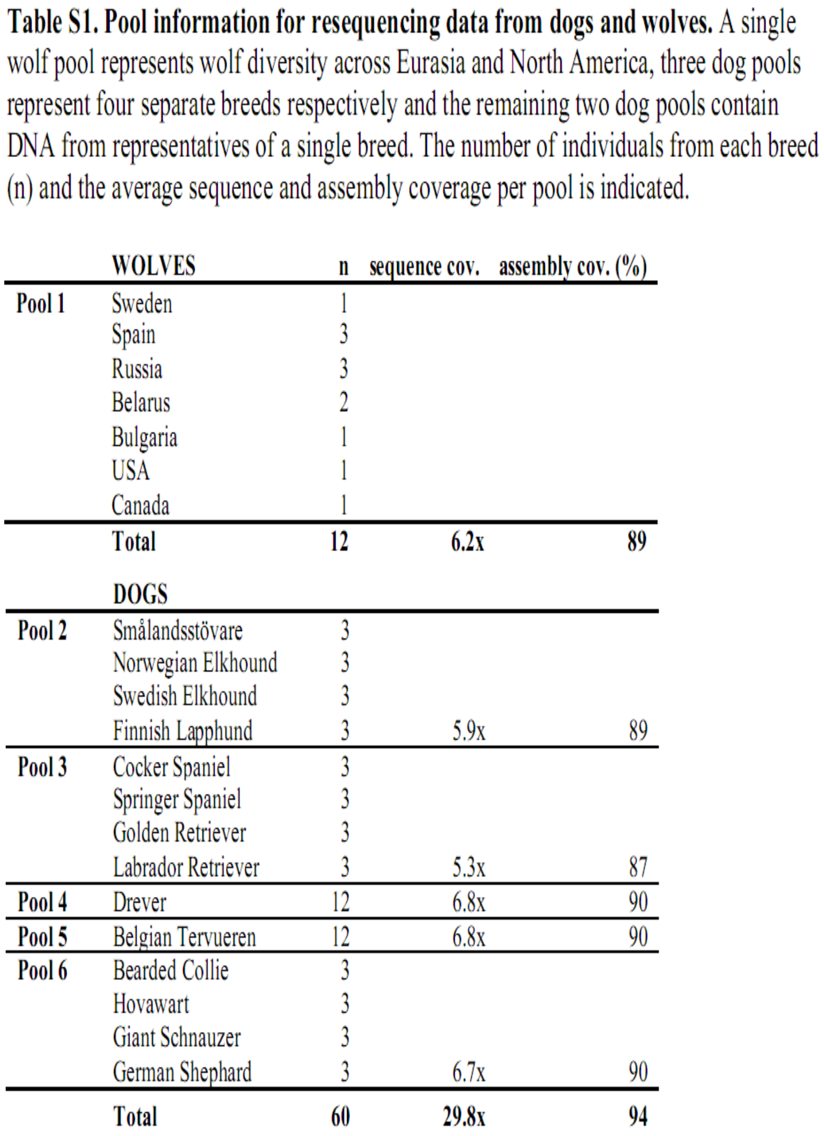
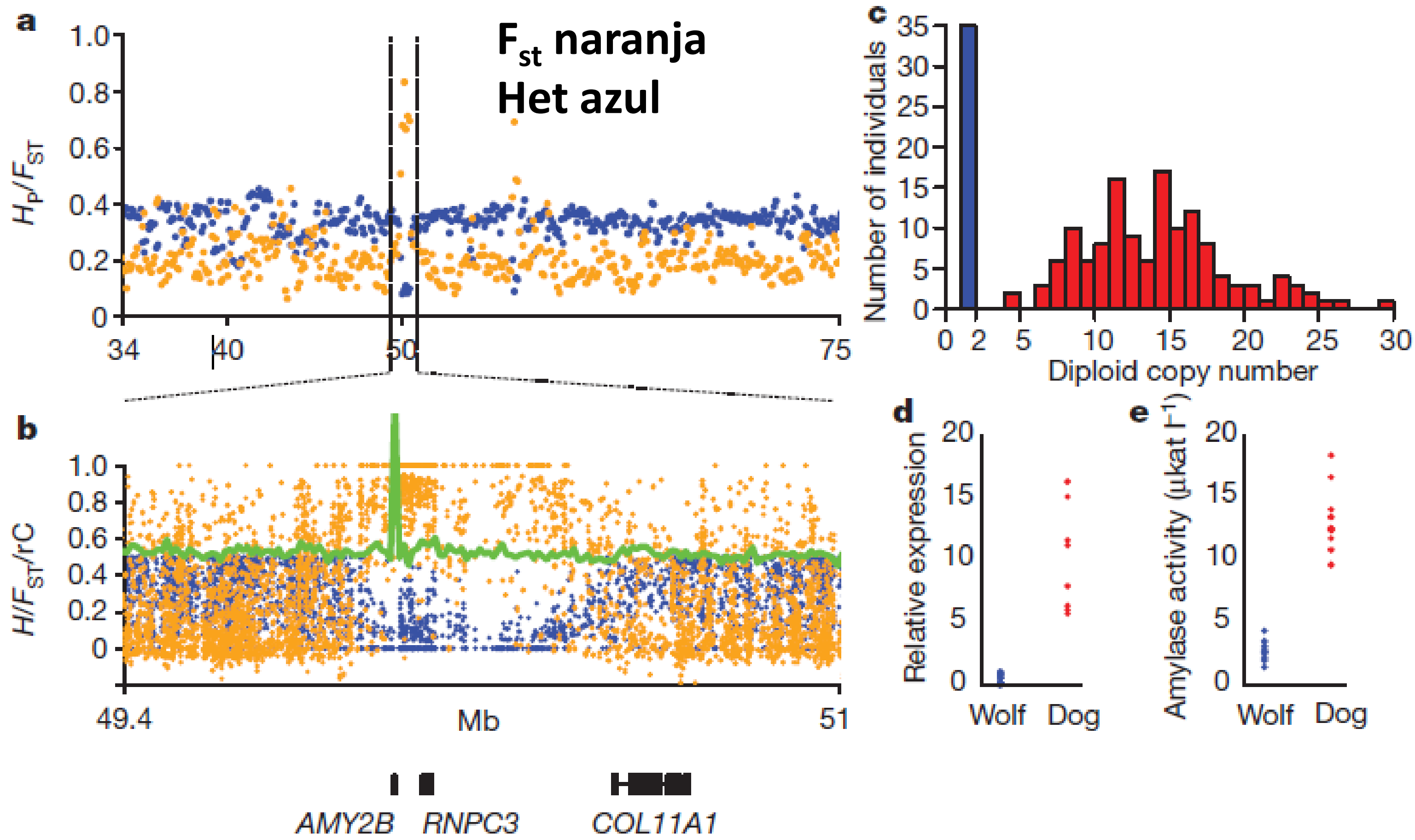


Table S1. Pool information for resequencing data from dogs and wolves. A single wolf pool represents wolf diversity across Eurasia and North America, three dog pools represent four separate breeds respectively and the remaining two dog pools contain DNA from representatives of a single breed. The number of individuals from each breed (n) and the average sequence and assembly coverage per pool is indicated.

	WOLVES	**n**	**sequence cov.**	**assembly cov. (%)**
Pool 1	Sweden	1		
	Spain	3		
	Russia	3		
	Belarus	2		
	Bulgaria	1		
	USA	1		
	Canada	1		
	Total	**12**	**6.2x**	**89**
	DOGS			
Pool 2	Smålandsstövare	3		
	Norwegian Elkhound	3		
	Swedish Elkhound	3		
	Finnish Lapphund	3	5.9x	89
Pool 3	Cocker Spaniel	3		
	Springer Spaniel	3		
	Golden Retriever	3		
	Labrador Retriever	3	5.3x	87
Pool 4	Drever	12	6.8x	90
Pool 5	Belgian Tervueren	12	6.8x	90
Pool 6	Bearded Collie	3		
	Hovawart	3		
	Giant Schnauzer	3		
	German Shephard	3	6.7x	90
	Total	**60**	**29.8x**	**94**

Gen de la amilasa: F_{st} y Heterocigosidad

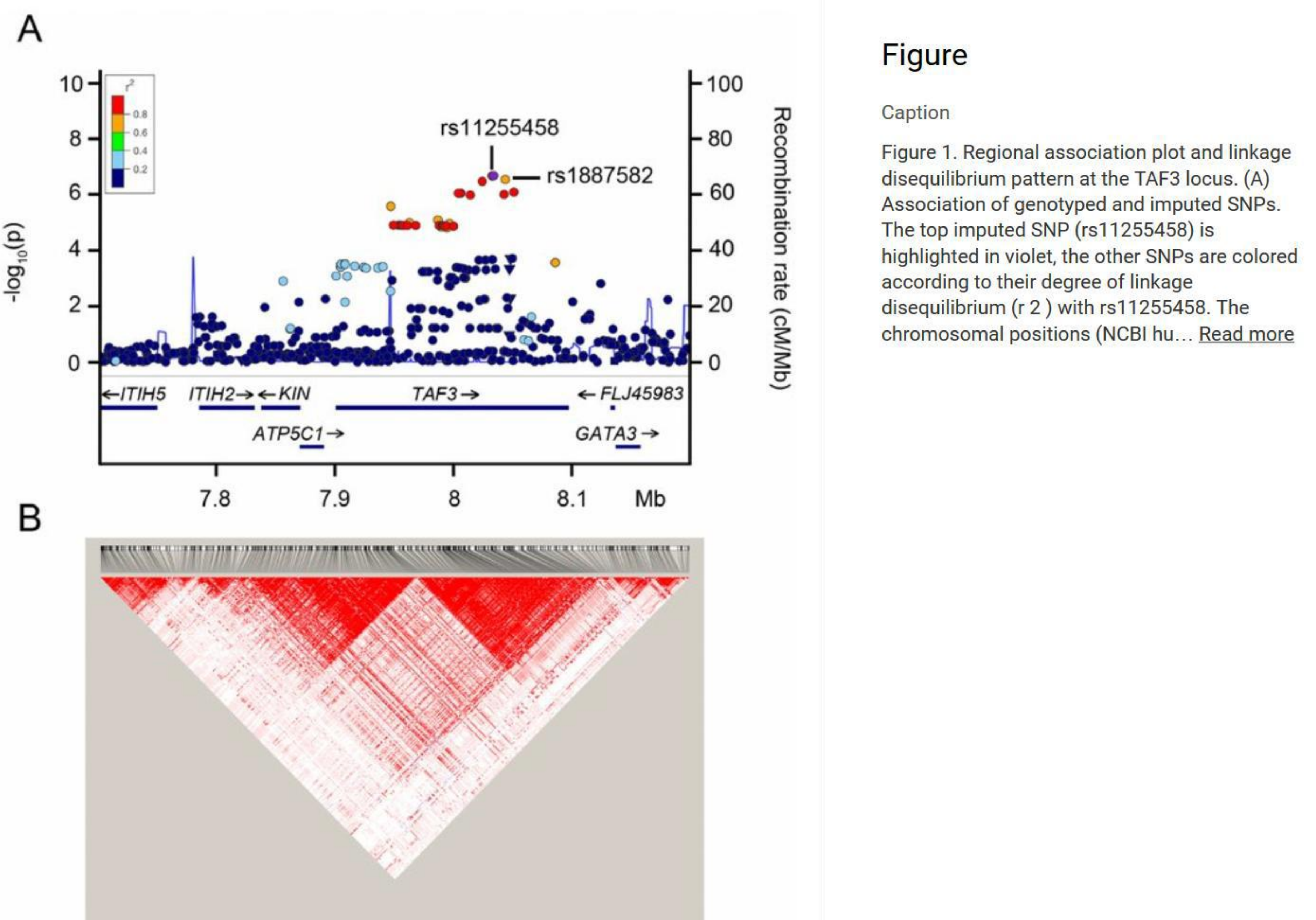


F_{st} naranja
Het azul

2 | **Selection for increased amylase activity.** **a**, Pooled heterozygosity,), and average fixation index, F_{ST} (orange), plotted for 200-kb windows

number in dog. Genes in the region are shown below pane showing the distribution of diploid amylase copy number i

Desequilibrio de ligamiento



Figure

Caption

Figure 1. Regional association plot and linkage disequilibrium pattern at the TAF3 locus. (A) Association of genotyped and imputed SNPs. The top imputed SNP (rs11255458) is highlighted in violet, the other SNPs are colored according to their degree of linkage disequilibrium (r 2) with rs11255458. The chromosomal positions (NCBI hu... Read more

Enriquecimiento de términos de Ontología de genes asociados a la domesticación

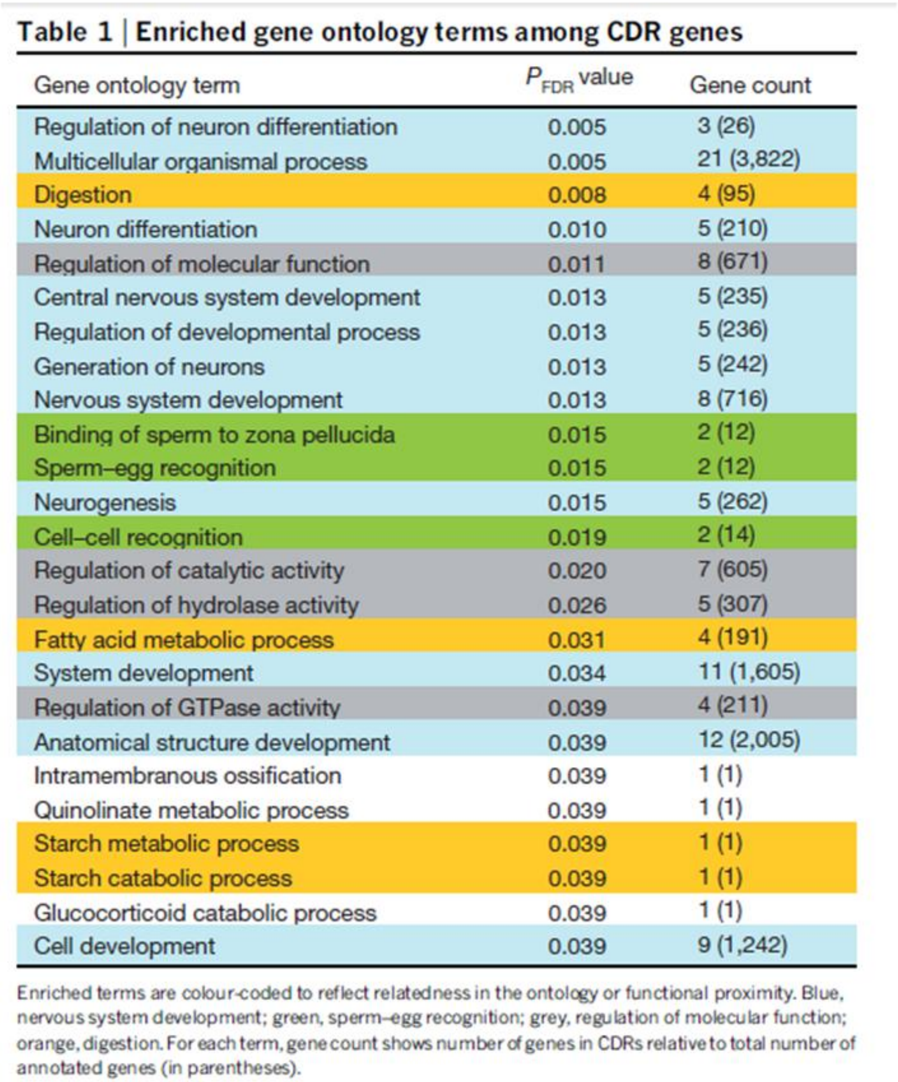


Table 1 | Enriched gene ontology terms among CDR genes

Gene ontology term	P_{FDR} value	Gene count
Regulation of neuron differentiation	0.005	3 (26)
Multicellular organismal process	0.005	21 (3,822)
Digestion	0.008	4 (95)
Neuron differentiation	0.010	5 (210)
Regulation of molecular function	0.011	8 (671)
Central nervous system development	0.013	5 (235)
Regulation of developmental process	0.013	5 (236)
Generation of neurons	0.013	5 (242)
Nervous system development	0.013	8 (716)
Binding of sperm to zona pellucida	0.015	2 (12)
Sperm–egg recognition	0.015	2 (12)
Neurogenesis	0.015	5 (262)
Cell–cell recognition	0.019	2 (14)
Regulation of catalytic activity	0.020	7 (605)
Regulation of hydrolase activity	0.026	5 (307)
Fatty acid metabolic process	0.031	4 (191)
System development	0.034	11 (1,605)
Regulation of GTPase activity	0.039	4 (211)
Anatomical structure development	0.039	12 (2,005)
Intramembranous ossification	0.039	1 (1)
Quinolinate metabolic process	0.039	1 (1)
Starch metabolic process	0.039	1 (1)
Starch catabolic process	0.039	1 (1)
Glucocorticoid catabolic process	0.039	1 (1)
Cell development	0.039	9 (1,242)

Enriched terms are colour-coded to reflect relatedness in the ontology or functional proximity. Blue, nervous system development; green, sperm–egg recognition; grey, regulation of molecular function; orange, digestion. For each term, gene count shows number of genes in CDRs relative to total number of annotated genes (in parentheses).

- Genes de desarrollo del sistema nervioso
- Genes que tienen que ver con la unión del espermatozoide y el ovulo
- Genes de metabolismo de ácidos grasos y de almidón:
  - Ven que genes de todas las etapas están representados


BEFORE