

## Filogenómica y Pangenómica

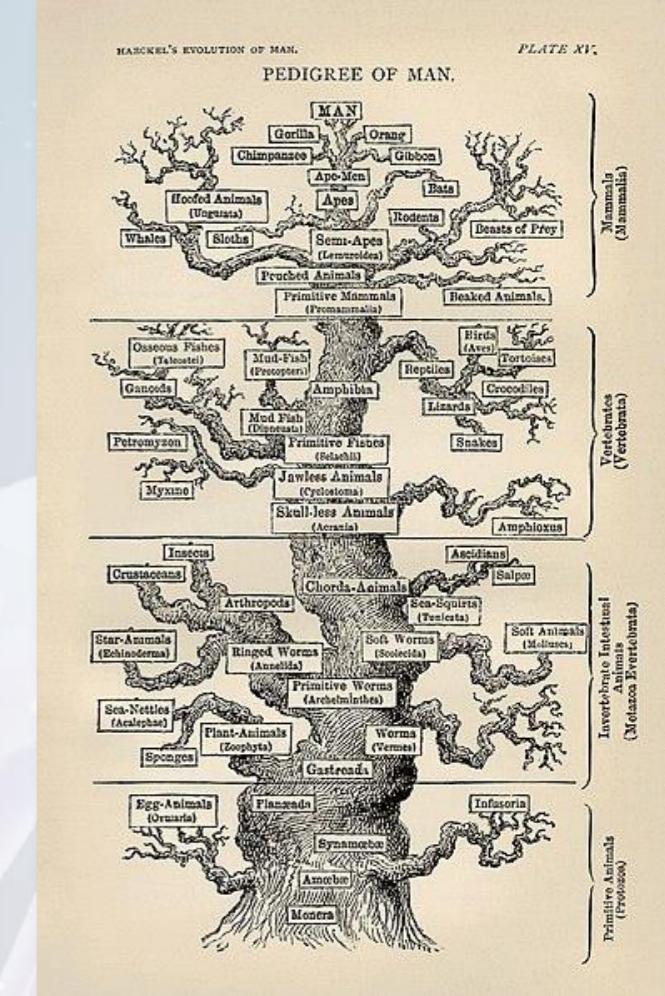
Ascanio Rojas A.

CeCALULA - ULA

[ascanio@ula.ve](mailto:ascanio@ula.ve)



Darwin's notebooks (c. July 1837)

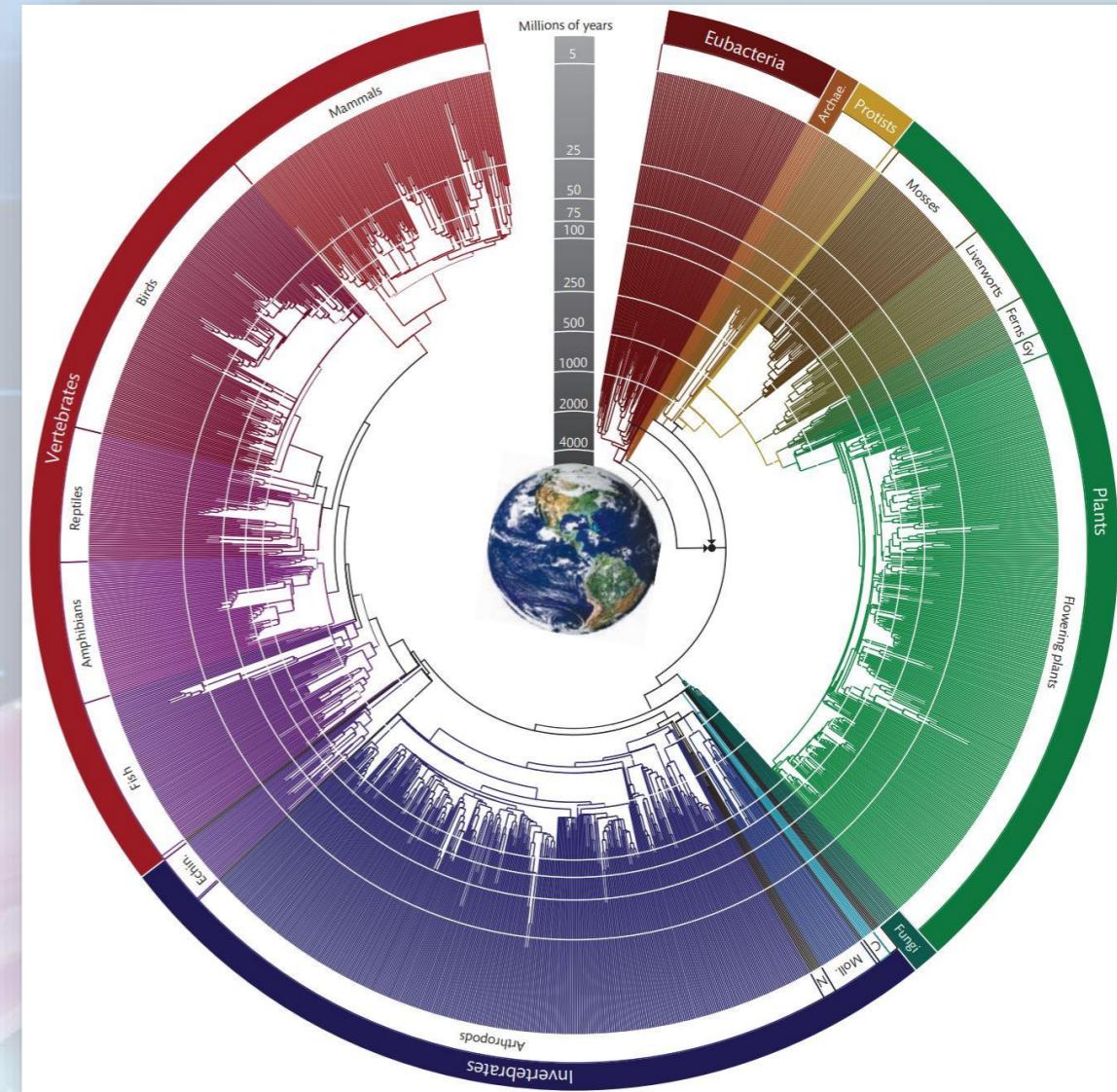


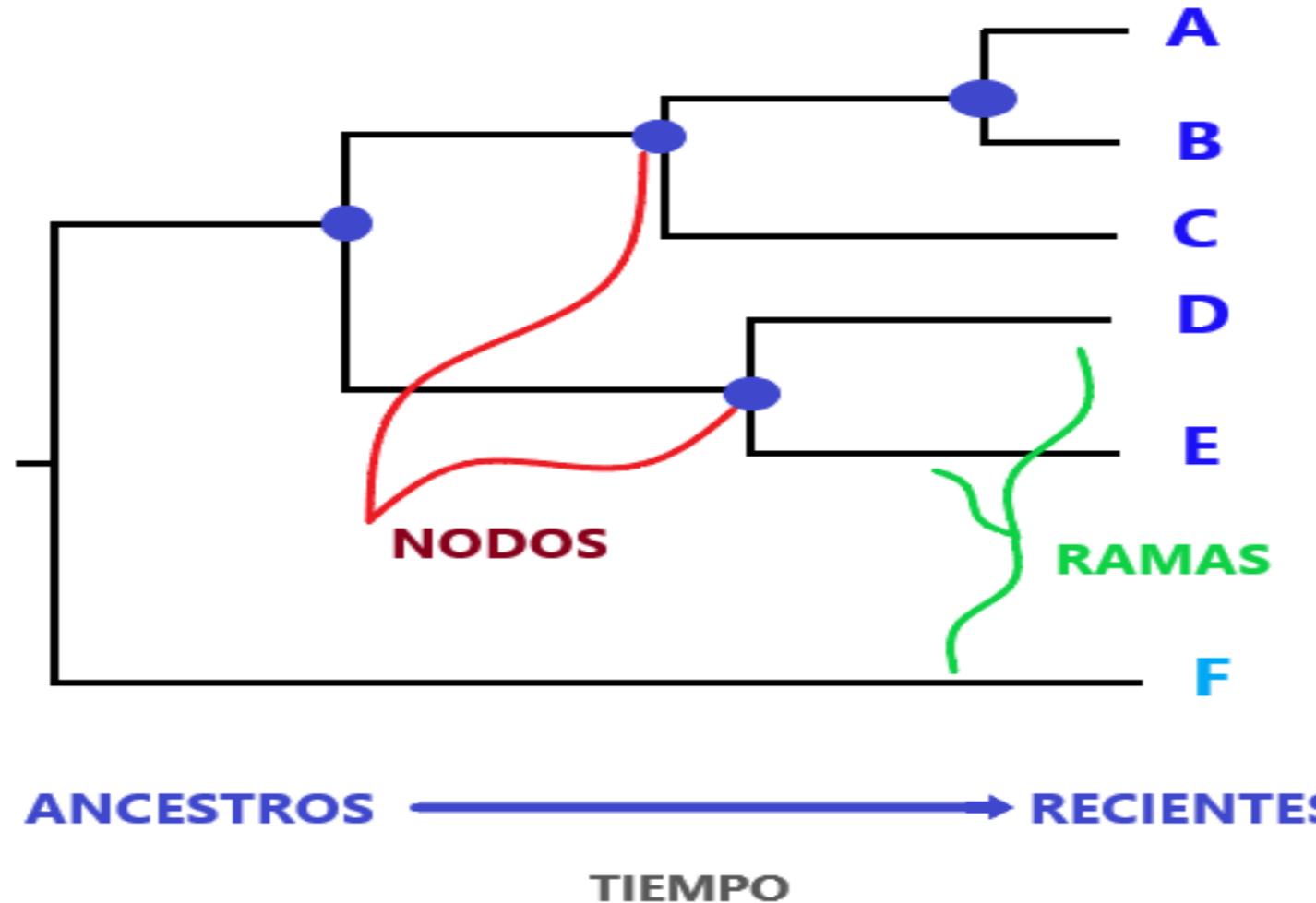
Haeckel in *The Evolution of Man* (1879)

Filogenética es el estudio de las relaciones evolutivas entre organismos genéticamente relacionados, siendo el análisis filogenético el mecanismo para estimar estas relaciones evolutivas.

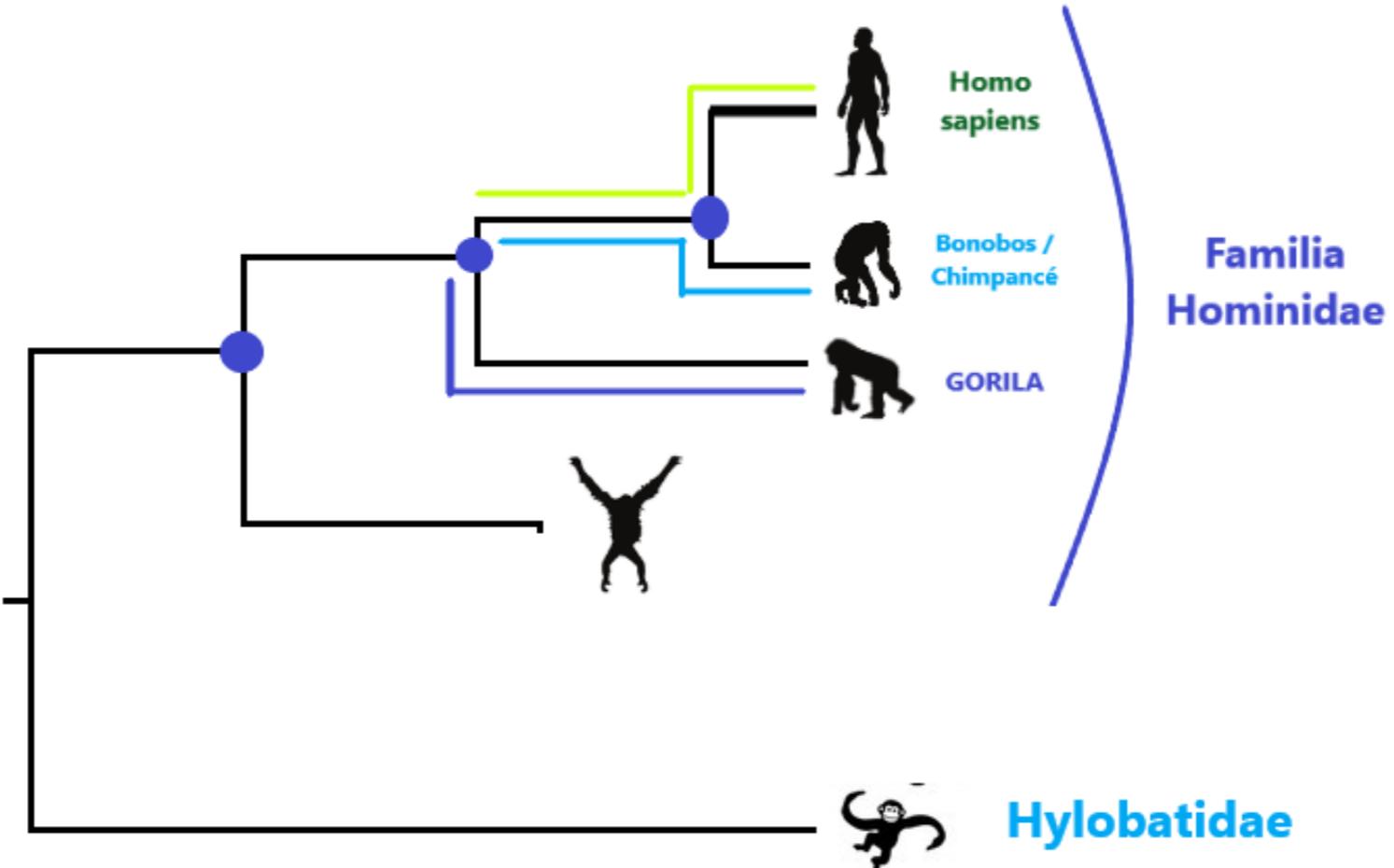
En la filogenética molecular la secuencias genómicas (ARN/ADN) o de proteína se utilizan para evaluar las relaciones entre organismos, siendo el resultado de estos análisis un diagrama de ramificaciones conectadas por **nodos** comúnmente llamado árbol Filogenético.

*The timetree of life. S. Blair Hedges, Sudhir Kumar. Oxford: Oxford University Press. 2009. (ANRNR)*

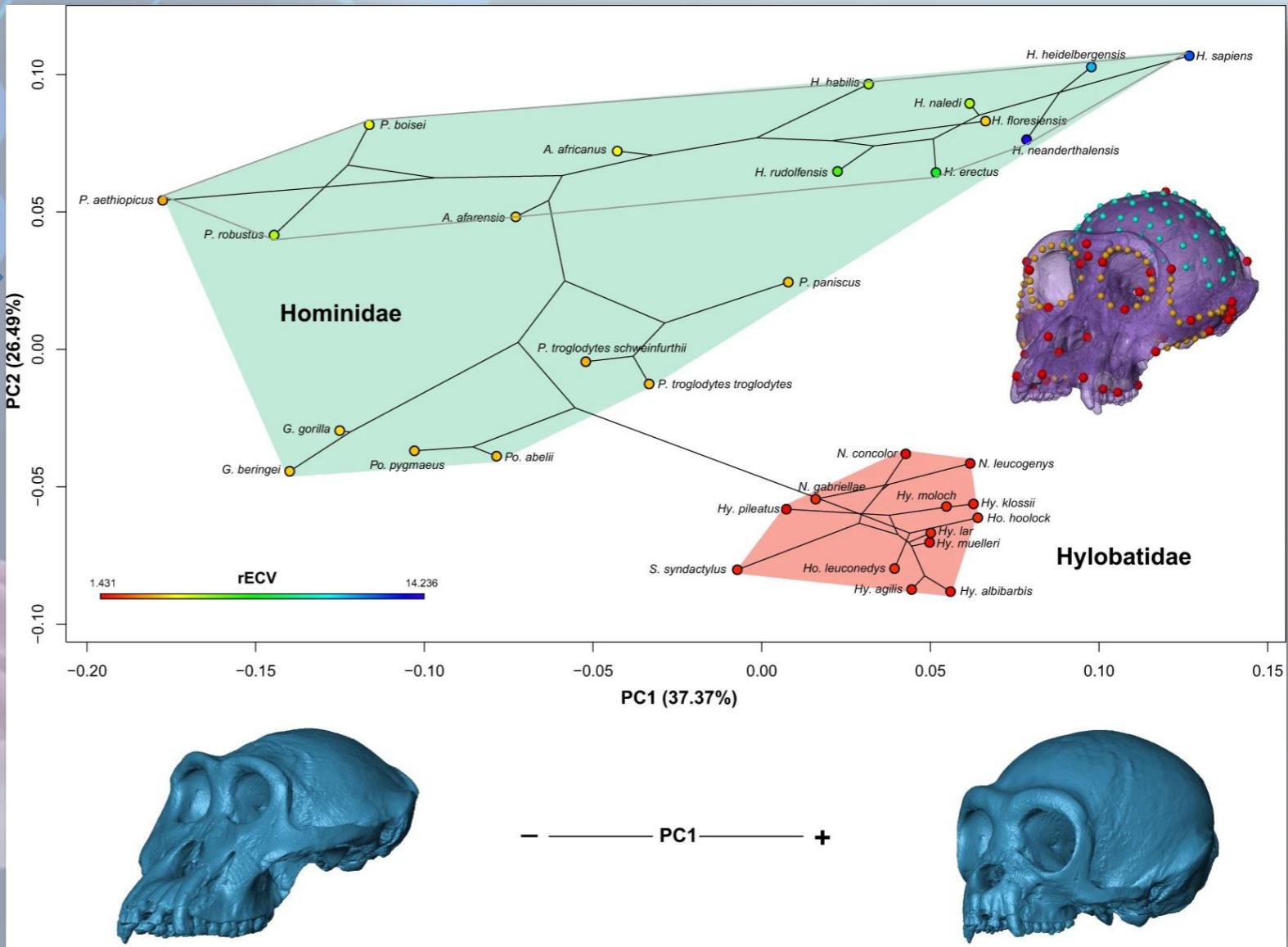
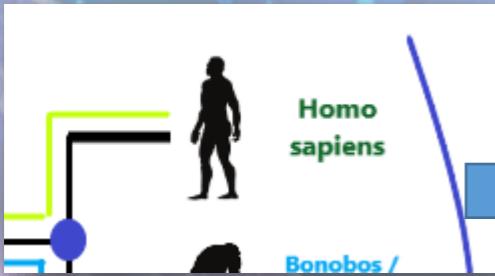




Cursos Internacionales .  
**Secuenciación y Análisis de Datos Genómicos para la Detección Microbiológica  
de Enfermedades Transmitidas por Alimentos y Aguas**



## Filogenias...



Rocatti, G., Perez, S.I. The Evolutionary Radiation of Hominids: a Phylogenetic Comparative Study. *Sci Rep* 9, 15267 (2019). <https://doi.org/10.1038/s41598-019-51685-w>

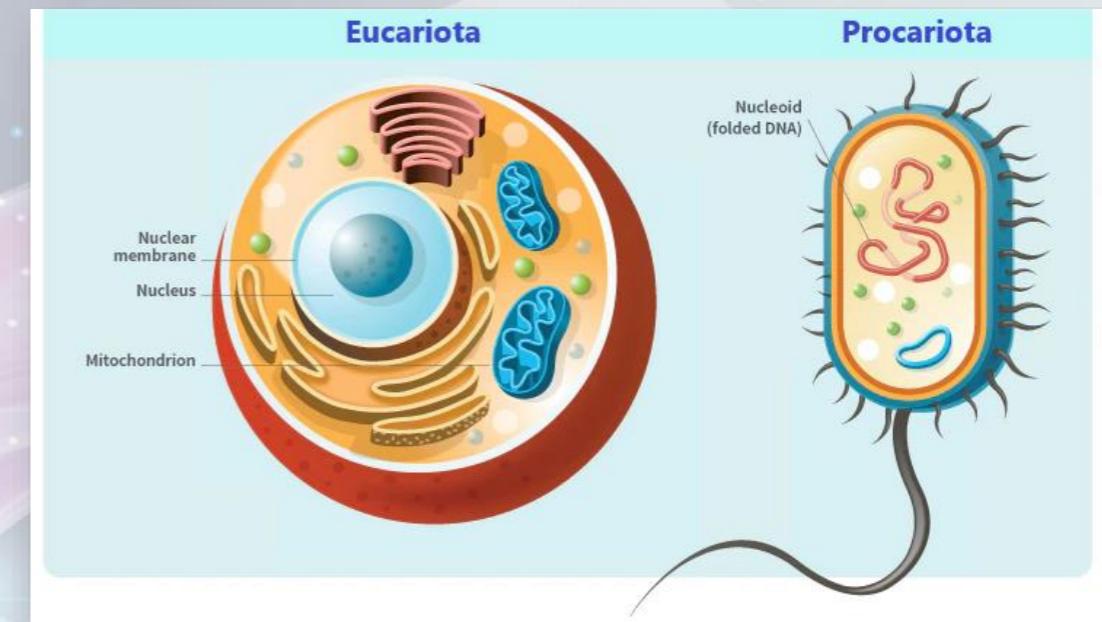
## Filogenias...

Las reconstrucciones filogenéticas moleculares pueden realizarse a partir de **Marcadores “Clásicos”**, provenientes datos morfológicos o estructurales, como lípidos, metabolitos secundarios, cariotipos, reacciones inmunogénicas, electroforesis de proteínas (iso-enzimas) o análisis de RFLP, o a partir de datos provenientes de secuencias de ADN/ARN o Proteínas.

A diferencia de los marcadores morfológicos clásicos las secuencias de **ADN/Proteína** nos permiten comparar mayor cantidad de caracteres, evitando de alguna forma la saturación que suelen sufrir los morfológicos

A nivel molecular tenemos tres fuentes principales de marcadores a utilizar:

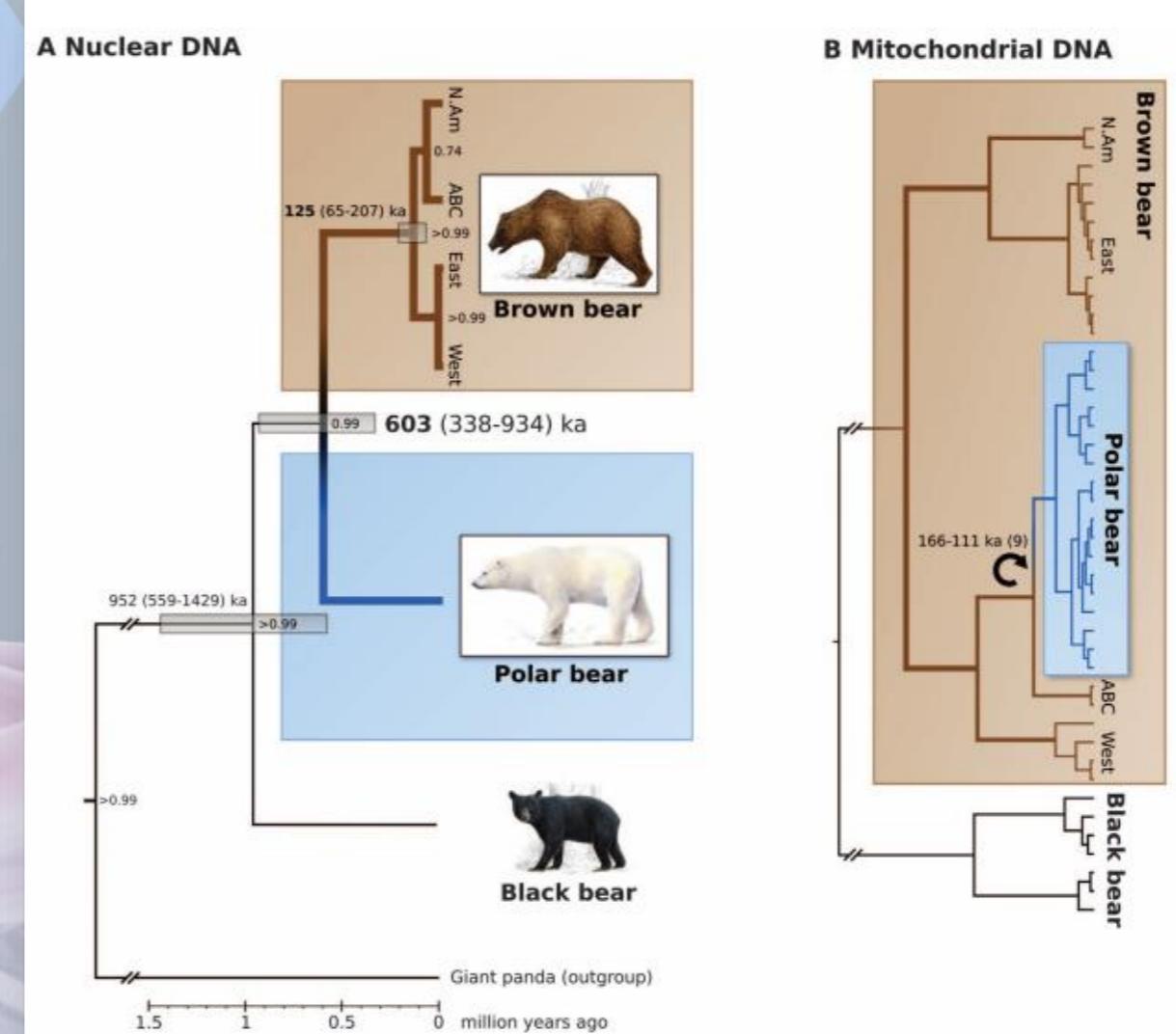
- **ADN Mitocondrial (p. ej. CytB)**
- **ADN Nuclear ((p. ej. SSR)**
- **ADN Ribosomal (p. ej. 16S)**



Filogenias...



*Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and  
Distinct Bear Lineage Frank Hailer et al. Science 336, 344 (2012) DOI:  
10.1126/science.1216424*



## Filogenómica

El término **filogenómica** (Jonathan Eisen 1998 \*) destaca la importancia de los métodos filogenéticos para la anotación de genes sin depender de estudios funcionales. Resaltando que las metodologías para reconstrucciones filogenéticas basadas en genes (anotaciones funcionales) y de especies (reconstrucción del árbol de la vida) solapan en gran medida.

**Filogenómica** en contexto general se usa para referirse al análisis de datos genómicos y reconstrucciones evolutivas. Donde se obtiene información al comparar genomas completos o grandes secciones de uno, y se evalúan sus relaciones evolutivas.

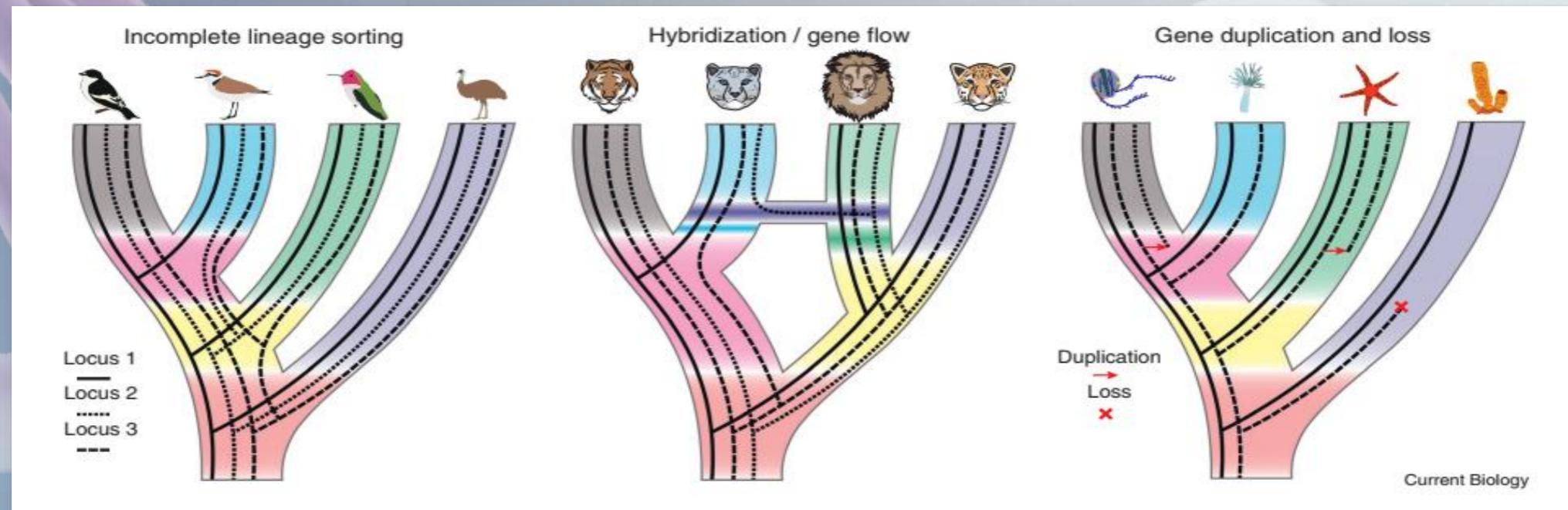
- Predicción de genes funcionales
- Evolución de familias de genes y eventos de transferencia horizontal
- Relaciones evolutivas

- *Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis*  
Genome Research Jonathan A Eisen March 1998 DOI: 10.1101/gr.8.3.163

## Filogenómica

**Predicción de genes funcionales:** para determinar la función de los genes, esta se hacía por la comparación de la secuencias problemáticas contra secuencias a las que se les conocía su función (p. ejm. Haciendo uso de BLAST). Esto podía funcionar para algunos genes, pero en otros la similaridad entre las secuencias no implicaba la misma función.

**Evolución de familias de genes y eventos de transferencia horizontal:** De la misma forma que se determinan la posible función de los genes en filogenómica podemos rastrear eventos de la delección o duplicación de genes, inversión o translocación, incluso detectar eventos de transferencia horizontal. Eventos estos de alta significación evolutiva

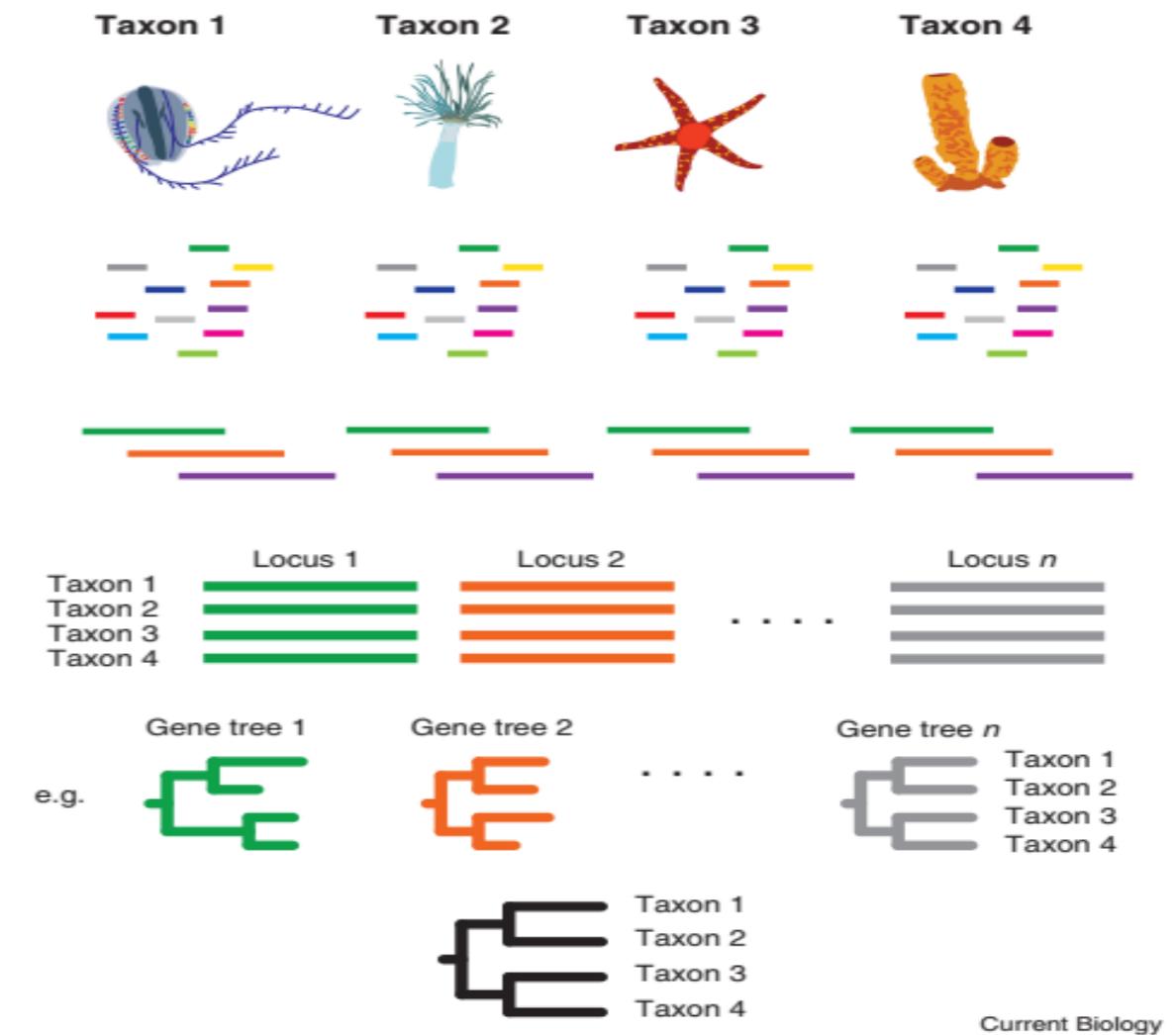
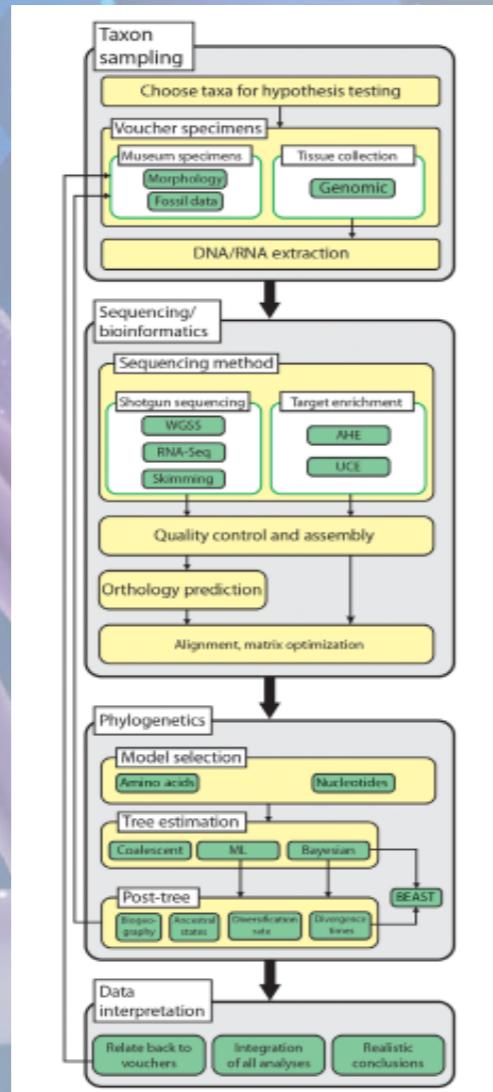


*Phylogenomics David A. Duchêne Current Biology 31, R1141-R1224, October 11, 2021*

# Cursos Internacionales .

## Secuenciación y Análisis de Datos Genómicos para la Detección Microbiológica de Enfermedades Transmitidas por Alimentos y Aguas

### Protocolo básico para Filogenómica



Systematic Entomology  
(2020), 45, 225-247 DOI:  
10.1111/syen.12406

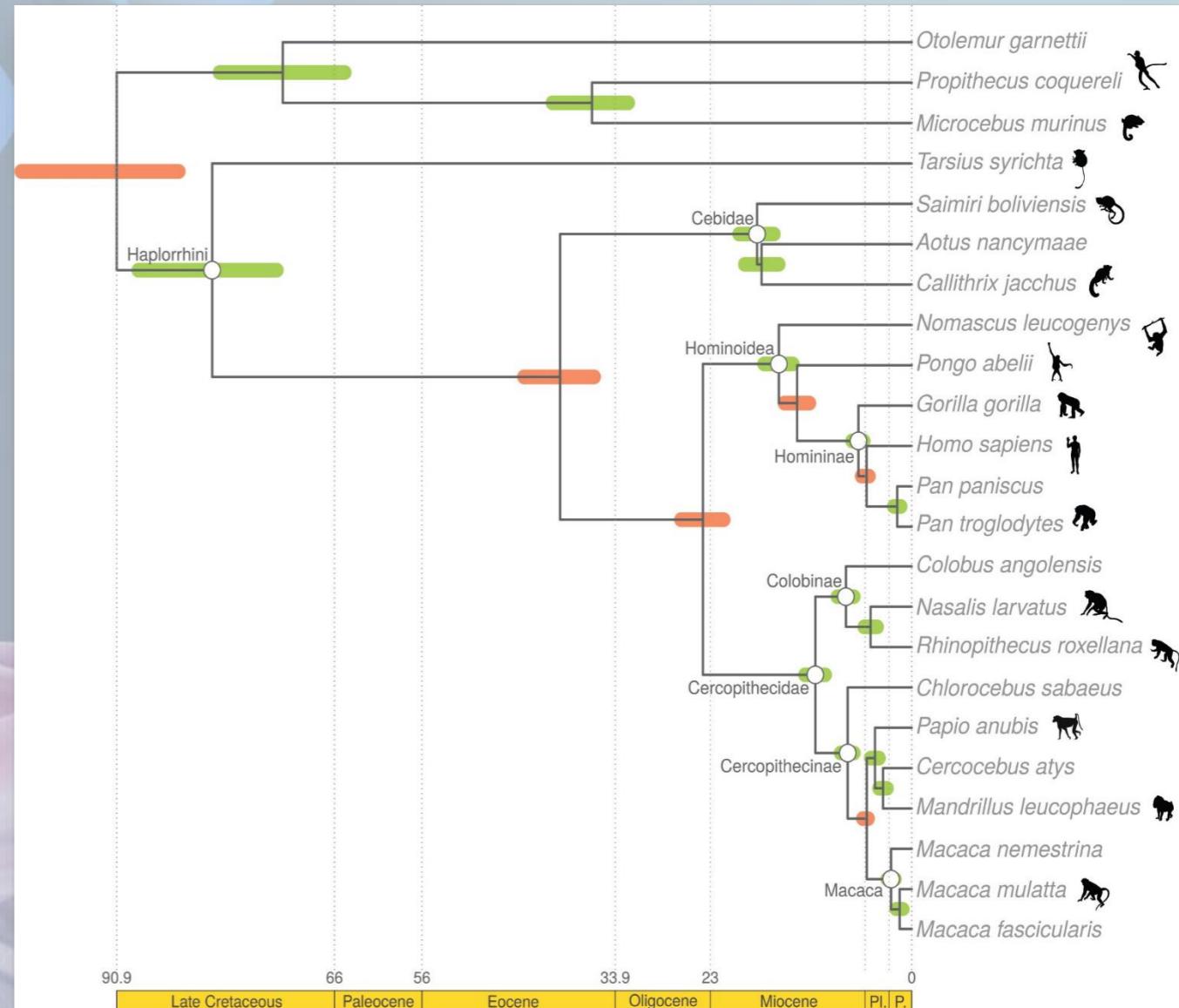
Current Biology 31, R1141-R1224, October 11, 2021

## Filogenómica

En 2018 Collins Y Tomas Hrbek ponen a provea las diferentes aproximaciones que se emplean para Filogenómica sobre 23 genomas completos de primates.

*An In Silico Comparison of Protocols for Dated Phylogenomics*  
 2018 DOI: 10.1093/sysbio/syx089

### Cursos Internacionales . Secuenciación y Análisis de Datos Genómicos para la Detección Microbiológica de Enfermedades Transmitidas por Alimentos y Aguas



## Recursos Webs en Filogenia

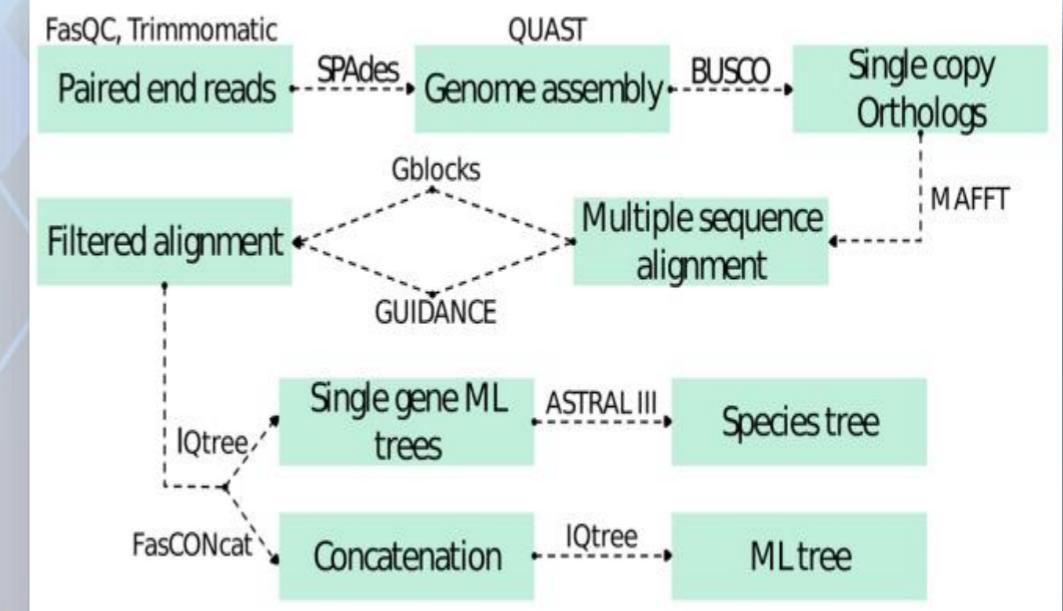
- **RAXML** (Randomized Axelerated Maximum Likelihood) programa para inferencia secuencial por máxima verosimilitud de grandes árboles filogenéticos
- **FastTree** Programa para la inferencia de árboles filogenéticos por aproximación de máxima verosimilitud a partir de alineamientos de secuencias de nucleótidos o proteínas.

## Portales

- <https://www.phylo.org/>
- <https://www.phylogeny.fr/>
- <http://www.atgc-montpellier.fr/phyml/>
- <http://www.trex.uqam.ca/index.php>
- <http://iqtree.cibiv.univie.ac.at/>

## Edición de Arboles

- <https://itol.embl.de/>



- **RAXML** <https://cme.h-its.org/exelixis/web/software/raxml/>
- **FastTree** <http://www.microbesonline.org/fasttree/>
- **CIPRES: Portal Científico que brinda recursos en línea sobre Filogenética, Sistemática y Biología Evolutiva.** <https://www.phylo.org/>

## Concatenar Marcadores moleculares

**FASconCAT** : programa de fácil manejo (Linux) que concatena rápidamente diferentes tipos de datos de secuencias en una supermatriz.

<http://software.zfmk.de/>

**SequenceMatrix**: programa diseñado en Java para facilitar el ensamblaje y el análisis de conjuntos de datos multigénicos.

(a)

SequenceMatrix 1.7.6							
Taxon	Total score	No of charsets	Cytb	COI	COIII	atp6	atp8
Homo sapiens	0.0%	13	(ON TOP)	(ON TOP)	(ON TOP)	(ON TOP)	(ON TOP)
Pan paniscus	0.1%	14	1% #1 (0.7%)	1% #1 (0.44%)	0% #0 (0.51%)	0% #0 (0.58%)	0% #0 (0.47%)
Pan troglodytes	0.66%	14	0% #0 (0.61%)	0% #0 (0.38%)	1% #1 (0.63%)	1% #1 (0.73%)	2% #1 (0.95%)
Gorilla gorilla	4.53%	14	8% #2 (1.66%)	7% #2 (1.1%)	7% #2 (1.27%)	5% #2 (1.31%)	10% #2 (2.38%)
Pongo abelli	17.56%	13	24% #4 (3.85%)	22% #3 (2.7%)	24% #3 (3.18%)	17% #3 (2.93%)	29% #4 (6.19%)
Pongo pygmaeus	18.53%	14	22% #3 (3.59%)	24% #4 (2.85%)	31% #5 (3.95%)	18% #4 (3.07%)	24% #3 (5.23%)
Hylobates lar	24.1%	13	29% #5 (4.46%)	26% #5 (3.15%)	24% #3 (3.18%)	26% #5 (4.25%)	36% #5 (7.61%)
Colobus guereza	44.14%	14	54% #12 (7.8%)	52% #22 (5.85%)	57% #14 (6.88%)	42% #12 (6.45%)	45% #6 (9.52%)
Chlorocebus aethiops	44.24%	14	54% #10 (7.71%)	43% #10 (4.94%)	52% #7 (6.37%)	47% #20 (7.18%)	66% #15 (13.72%)
Chlorocebus sabaeus	44.28%	13	56% #15 (7.97%)	43% #9 (4.87%)	53% #10 (6.5%)	47% #20 (7.18%)	64% #12 (13.23%)
Chlorocebus tantalus	44.4%	14	55% #14 (7.88%)	42% #7 (4.8%)	52% #7 (6.37%)	48% #22 (7.33%)	66% #15 (13.72%)
Trachypithecus obscurus	44.42%	14	54% #10 (7.71%)	49% #20 (5.53%)	55% #12 (6.63%)	40% #9 (6.15%)	60% #10 (12.38%)
Piliocolobus badius	44.73%	14	61% #20 (8.67%)	47% #17 (5.34%)	51% #6 (6.25%)	45% #16 (6.89%)	49% #7 (10.25%)
Chlorocebus pygerythrus	45.0%	14	54% #12 (7.8%)	43% #8 (4.87%)	53% #10 (6.5%)	46% #19 (7.03%)	69% #18 (14.21%)
Presbytis melalopha	45.07%	14	58% #17 (8.28%)	52% #21 (5.79%)	52% #7 (6.37%)	38% #7 (5.86%)	57% #8 (11.9%)
Macaca fascicularis	45.21%	13	49% #6 (7.09%)	45% #11 (5.06%)	56% #13 (6.76%)	41% #11 (6.33%)	71% #21 (14.76%)
Papio hamadryas	45.54%	13	59% #19 (8.41%)	42% #6 (4.79%)	57% #14 (6.88%)	45% #18 (6.9%)	62% #11 (12.85%)
Macaca thibetana	45.58%	13	52% #7 (7.44%)	46% #16 (5.26%)	58% #17 (7.01%)	44% #13 (6.74%)	69% #20 (14.28%)
Nasalis larvatus	45.79%	14	59% #18 (8.36%)	45% #13 (5.08%)	60% #22 (7.27%)	35% #5 (5.57%)	64% #13 (13.33%)
Semnopithecus entellus	45.92%	14	57% #16 (8.1%)	47% #17 (5.34%)	58% #17 (7.01%)	39% #8 (6.01%)	64% #13 (13.33%)
Rhinopithecus roxellana	45.95%	14	61% #21 (8.72%)	45% #15 (5.14%)	59% #20 (7.14%)	40% #9 (6.15%)	59% #9 (12.3%)
Macaca mulatta	46.21%	13	52% #7 (7.44%)	45% #11 (5.06%)	57% #14 (6.88%)	44% #13 (6.74%)	74% #22 (15.23%)
Macaca sylvanus	46.37%	14	52% #7 (7.44%)	45% #13 (5.08%)	58% #17 (7.01%)	45% #16 (6.89%)	69% #18 (14.21%)
Pygathrix nemaeus	46.8%	13	63% #22 (8.89%)	47% #17 (5.34%)	59% #20 (7.14%)	44% #13 (6.74%)	67% #17 (13.8%)
Cebus albifrons	71.02%	14	78% #23 (10.96%)	61% #23 (6.75%)	76% #24 (9.05%)	90% #29 (13.19%)	86% #29 (17.64%)
Saimiri sciureus	75.09%	14	83% #24 (11.57%)	69% #24 (7.59%)	77% #27 (9.18%)	93% #33 (13.63%)	86% #29 (17.64%)

<http://gaurav.github.io/taxondna>

*FASconCAT* <https://doi.org/10.1016/j.ympcv.2010.04.024>

*SequenceMatrix* <https://doi.org/10.1111/j.1096-0031.2010.00329.x>

## Filogenia en epidemiología

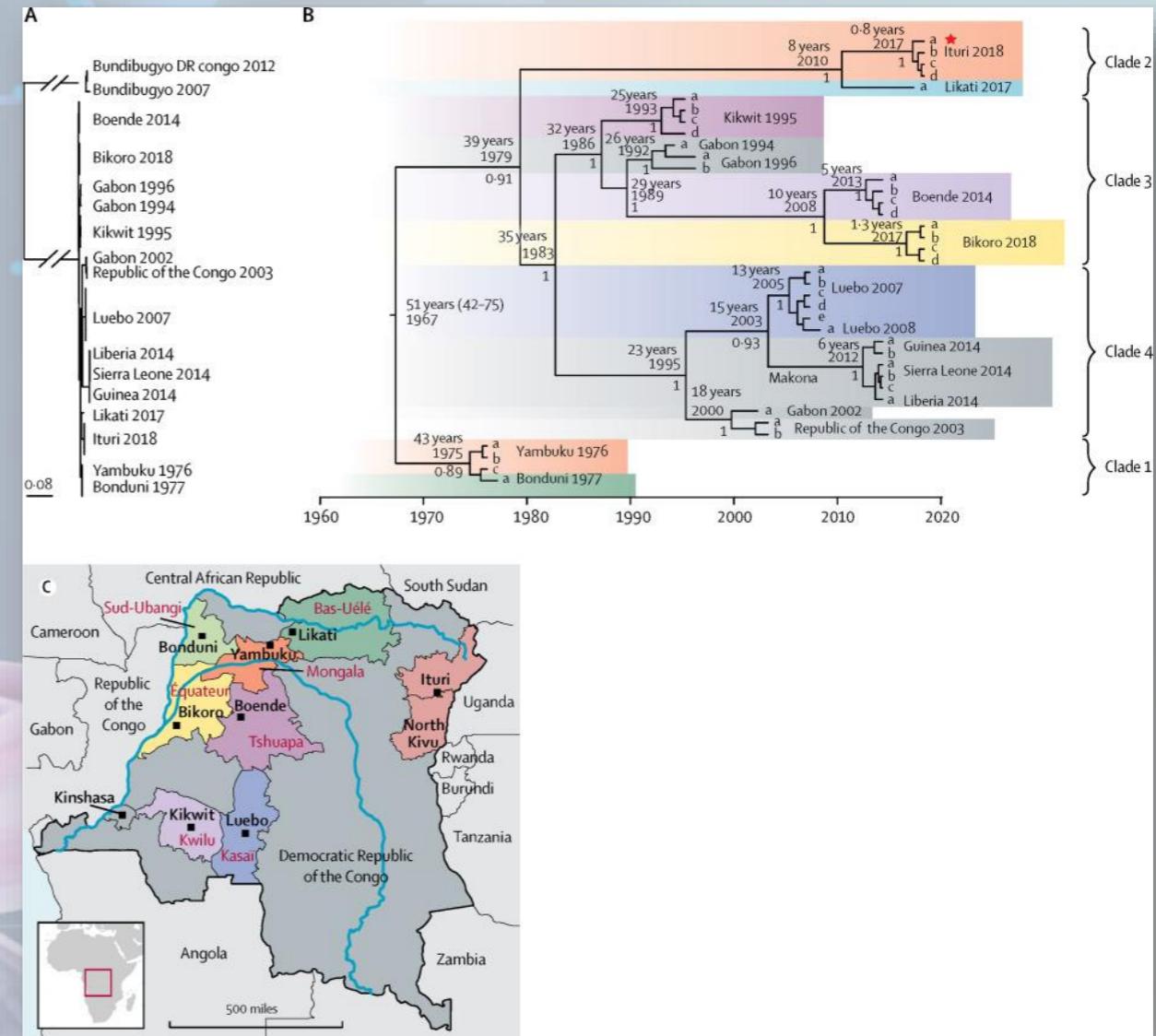
El objetivo de este estudio fue probar la eficacia de las terapias aplicadas y pruebas de diagnóstico con la cepa del brote del virus del Ébola de Ituri brote 2018.

**Characterisation of infectious Ebola virus from the ongoing outbreak to guide response activities in the Democratic Republic of the Congo: a phylogenetic and in vitro analysis**


  
*Laura K McMullan, Mike Flint, Ayan Chakrabarti, Lisa Guerrero, Michael K Lo, Danielle Porter, Stuart T Nichol, Christina F Spiropoulou, César Albarino*

**Summary**  
**Background** The ongoing Ebola virus outbreak in the Ituri and North Kivu Provinces of the Democratic Republic of Congo, which began in July, 2018, is the second largest ever recorded. Despite civil unrest, outbreak control measures and the administration of experimental therapies and a vaccine have been initiated. The aim of this study was to test the efficacy of candidate therapies and diagnostic tests with the outbreak strain Ituri Ebola virus. Lacking a

*Lancet Infect Dis* 2019; 19: 1023–32  
 Published Online  
 July 9, 2019  
[https://doi.org/10.1016/S1473-3099\(19\)30291-9](https://doi.org/10.1016/S1473-3099(19)30291-9)



*Laura K McMullan et al 2019 DOI:[https://doi.org/10.1016/S1473-3099\(19\)30291-9](https://doi.org/10.1016/S1473-3099(19)30291-9)*

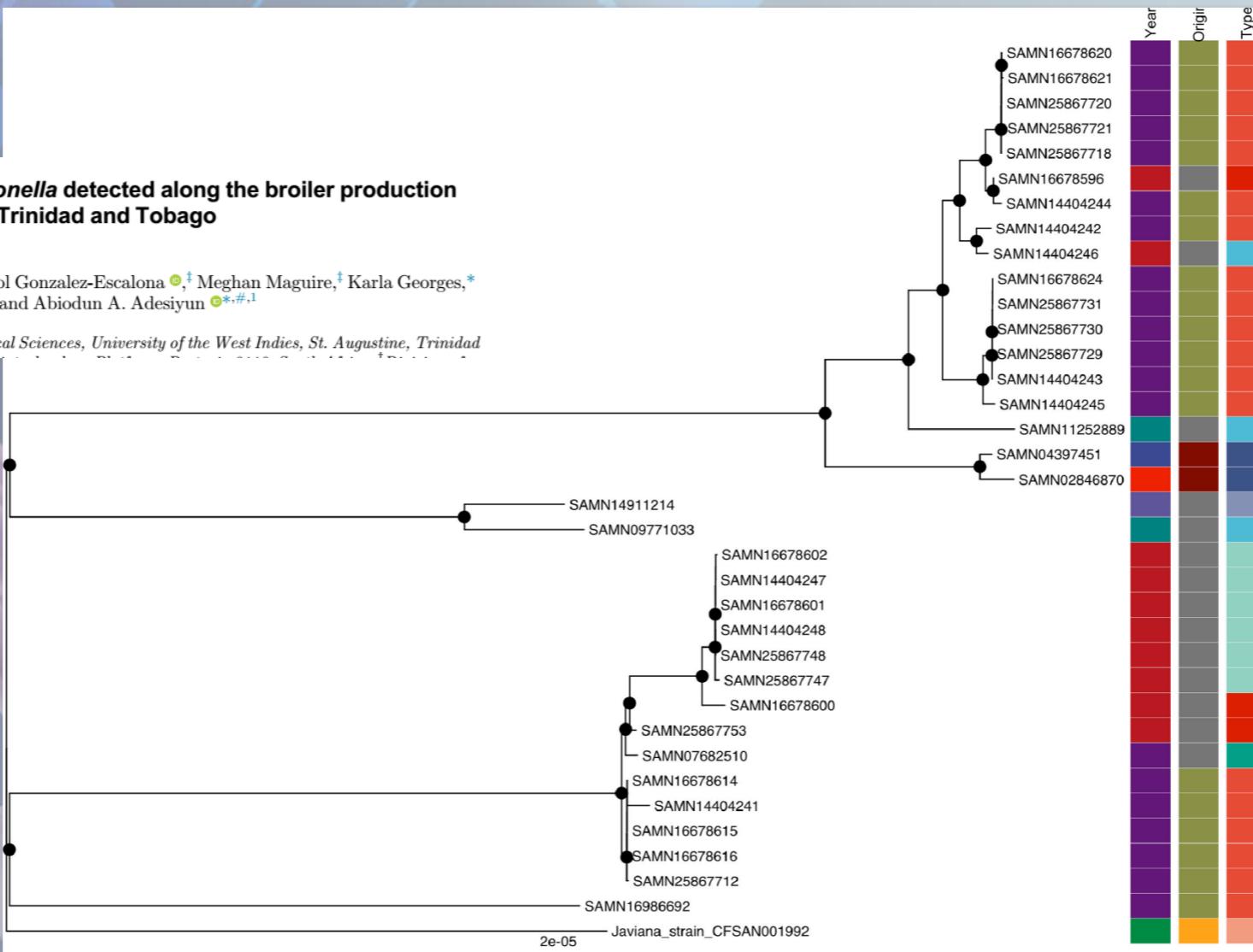
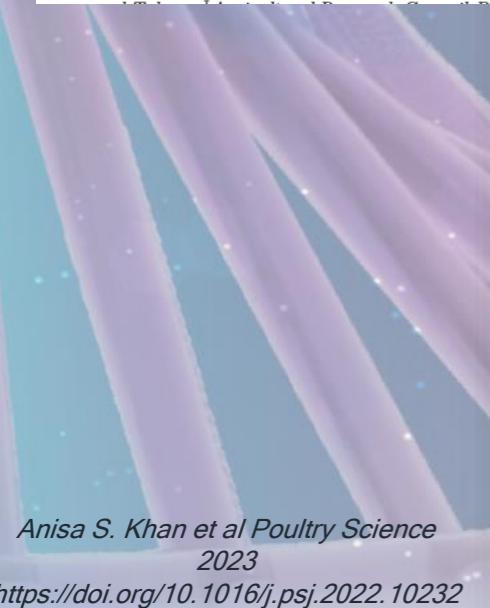
Filogenia

en epidemiología...

### Phylogenetic analyses of *Salmonella* detected along the broiler production chain in Trinidad and Tobago

Anisa S. Khan ,\* Rian E. Pierneef ,<sup>†</sup> Narjol Gonzalez-Escalona ,<sup>‡</sup> Meghan Maguire,<sup>‡</sup> Karla Georges,<sup>\*</sup>  
Woubit Abebe,<sup>§</sup> and Abiodun A. Adesiyun \*,#,<sup>¶</sup>

\*School of Veterinary Medicine, Faculty of Medical Sciences, University of the West Indies, St. Augustine, Trinidad



Type of Sample (Type)

Chicken carcass
Chilled chicken parts
Chilled whole carcass
Cloacal swab
Drag swab litter
Feed
Human faeces
Neck skin
Post-evisceration carcass
Water supply

Year of Isolation (Year)

2007
2008
2014
2016
2018
2019
2021

Farm to Fork (Origin)

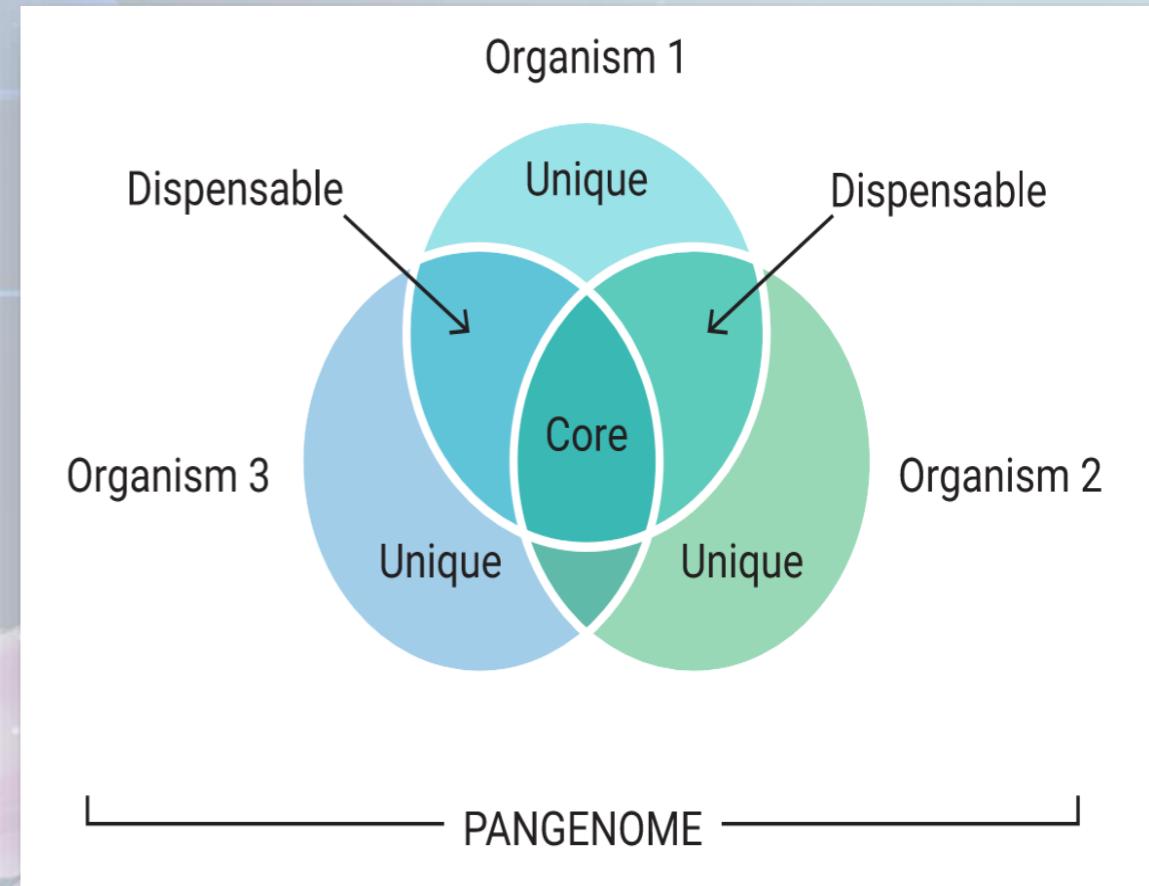
Control
Farm
Processing Plant
Reference

## PANGENOMAS

- El pangenoma representa el conjunto completo de genes de una especie, dividido en **Genoma Central (Core genome)**, un "Pangenoma caparazón" (**Shell Pangenome**) y el genoma "Periférico" (**Disposable genome o cloud pangenome**).
- EN 2005 se realiza el primer estudio de este tipo en seis cepas de *Streptococcus agalactiae*, revelando un genoma central que contenía el 80% de los genes de *S. agalactiae*  
\*.
- Por ejemplo, muchos genes de importancia agronómica en especies vegetales se encuentran con mayor frecuencia en el Genoma Periférico.

**Cursos Internacionales .**

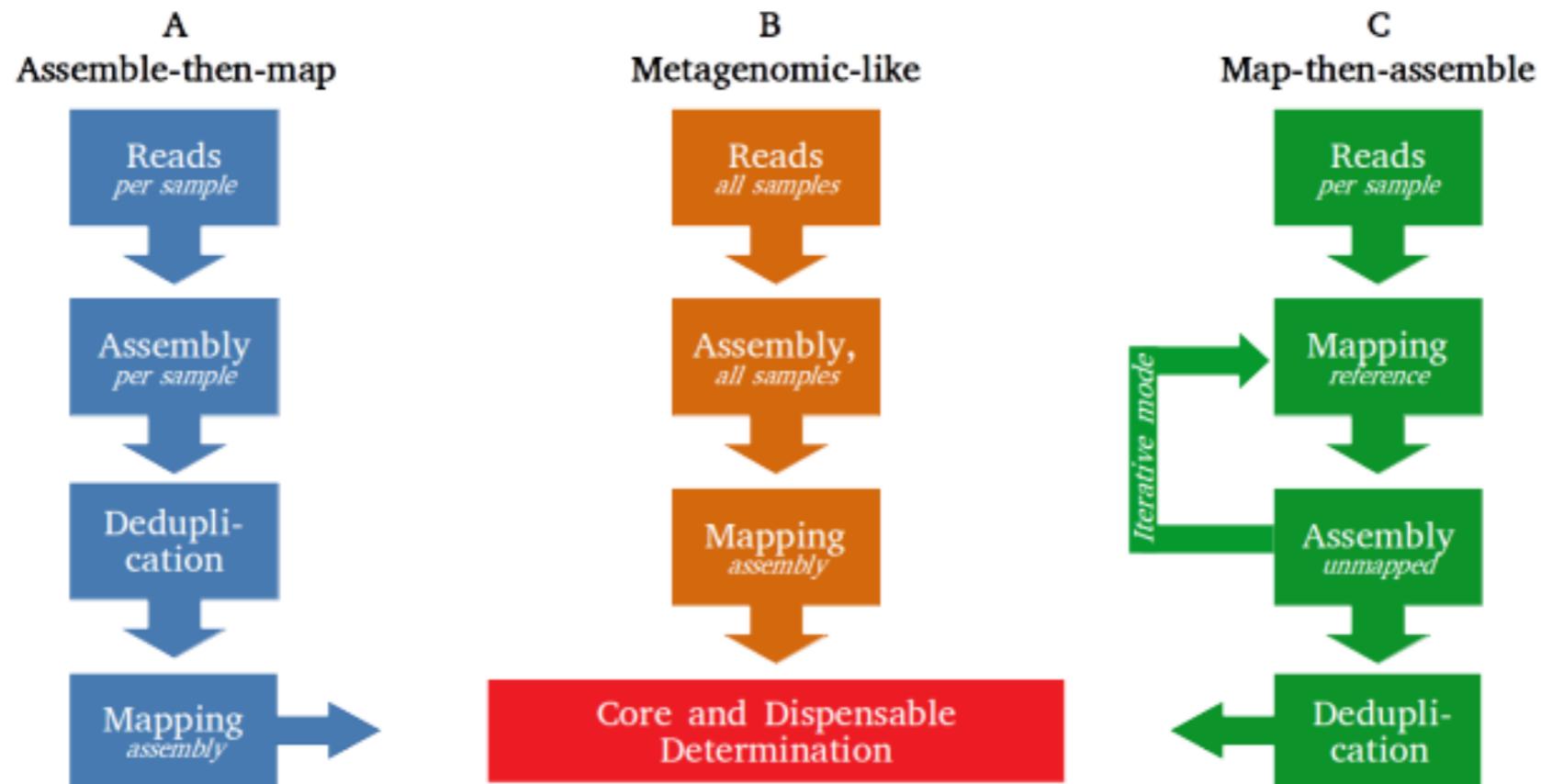
### **Secuenciación y Análisis de Datos Genómicos para la Detección Microbiológica de Enfermedades Transmitidas por Alimentos y Aguas**



\**Proc Natl Acad Sci U S A. 2005 Sep 27; 102(39): 13950-13955.*

*Published online 2005 Sep 19. doi: 10.1073/pnas.0506758102*

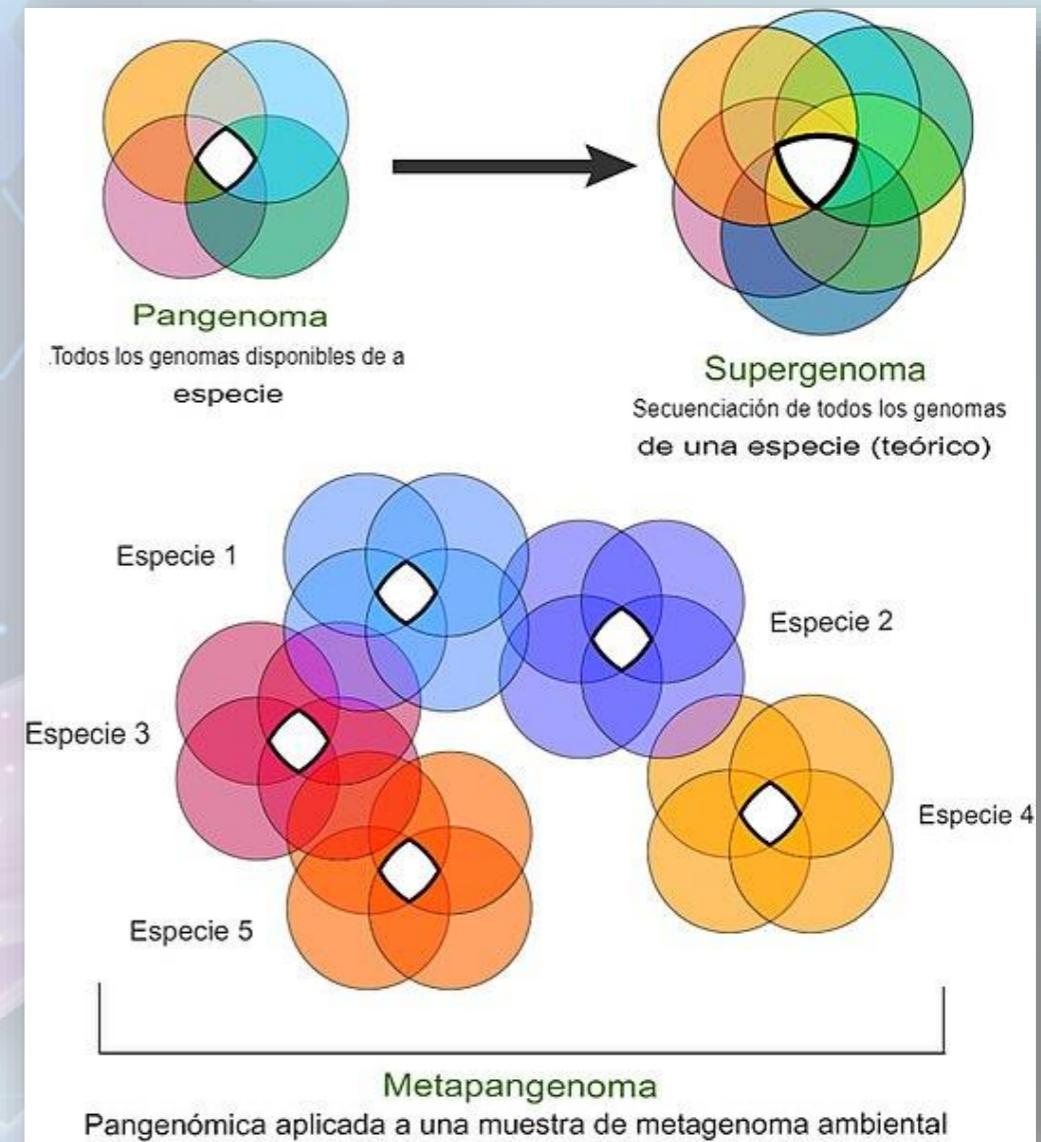
## Estrategias en análisis de PANGENOMAS



**Fig. 4.** The three approaches for pangenome sequence data assembly: left, assemble-then-map; center, metagenomic-like; right, map-then-assemble. For assemble-then-map and map-the-assemble methods, reduction of redundancy is performed (deduplication step) to identify the common sequences of different individuals.

## Aplicaciones de estudios de Pangenoma

- Estudio de Variantes alélicas compartidas y raras en una especie
  - Estudios de asociación del genoma completo (*GWAS*)
  - Variantes genéticas no codificantes
  - Genómica comparativa, Genómica estructural y Genómica funcional
  - Genómica de poblaciones
  - Medicina de precisión



**Article**

# A draft human pangenome reference

<https://doi.org/10.1038/s41586-023-05896-x>  
 Received: 9 July 2022

Accepted: 28 February 2023

Published online: 10 May 2023

Open access

 Check for updates

Wen-Wei Liao<sup>1,2,3,60</sup>, Mobin Asri<sup>4,60</sup>, Jana Ebler<sup>5,6,60</sup>, Daniel Doerr<sup>5,6</sup>, Marina Haukness<sup>4</sup>, Glenn Hickey<sup>4</sup>, Shuangjia Lu<sup>12</sup>, Julian K. Lucas<sup>4</sup>, Jean Monlong<sup>4</sup>, Haley J. Abel<sup>7</sup>, Silvia Buonaiuto<sup>8</sup>, Xian H. Chang<sup>4</sup>, Haoyu Cheng<sup>8,10</sup>, Justin Chu<sup>9</sup>, Vincenza Colonna<sup>8,11</sup>, Jordan M. Eizenga<sup>4</sup>, Xiaowen Feng<sup>8,10</sup>, Christian Fischer<sup>11</sup>, Robert S. Fulton<sup>12,13</sup>, Shilpa G. Cristian Groza<sup>9</sup>, Andrea Guerracino<sup>11,16</sup>, William T. Harvey<sup>17</sup>, Simon Heuimos<sup>18,19</sup>, Kerstin Howe<sup>20</sup>, Miten Jain<sup>21</sup>, Tsung-Yu Lu<sup>22</sup>, Charles Markello<sup>4</sup>, Fergal J. Martin<sup>23</sup>, Matthew W. Mitchell<sup>24</sup>, Katherine M. Munson<sup>17</sup>, Moses Njagi Mwaniki<sup>25</sup>, Adam M. Novak<sup>26</sup>, Hugh E. Olsen<sup>4</sup>, Trevor Pesout<sup>4</sup>, David Porubsky<sup>17</sup>, Pjotr Prins<sup>11</sup>, Jonas A. Sibbesen<sup>26</sup>, Jouni Siirén<sup>4</sup>, Ghad Tomlinson<sup>12</sup>, Flavia Villani<sup>11</sup>, Mitchell R. Volgger<sup>17,27</sup>, Lucinda L. A. Baid<sup>28</sup>, Carl A. Baker<sup>17</sup>, Anastasiya Belyaeva<sup>28</sup>, Joep de Ligt<sup>1</sup>

**PNAS**

## Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”

Hervé Tettelin<sup>a,b</sup>, Vega Masignani<sup>b,c</sup>, Michael J. Cieslewicz<sup>b,d,e</sup>, Claudio Donati<sup>c</sup>, Duccio Medini<sup>c</sup>, Naomi L. Wairu<sup>a</sup>, Samuel V. Angiuoli<sup>a</sup>, Jonathan Crabtree<sup>a</sup>, Amanda L. Jones<sup>g</sup>, A. Scott Durkin<sup>a</sup>, Robert T. DeBoy<sup>a</sup>, Tanja M. Da Marirosa Mora<sup>c</sup>, Maria Scarselli<sup>c</sup>, Immaculada Margarit y Ros<sup>c</sup>, Jeremy D. Peterson<sup>a</sup>, Christopher R. Hauser<sup>a</sup>, Jaideep P. Sundaram<sup>a</sup>, William C. Nelson<sup>a</sup>, Ramana Madupu<sup>a</sup>, Lauren M. Brinkac<sup>a</sup>, Robert J. Dodson<sup>a</sup>, Mary J. Steven A. Sullivan<sup>a</sup>, Sean C. Daugherty<sup>a</sup>, Daniel H. Haft<sup>a</sup>, Jeremy Selengut<sup>a</sup>, Michelle L. Gwinn<sup>a</sup>, Liwei Zhou<sup>a</sup>, Nikhat Zafar<sup>a</sup>, Hoda Khouri<sup>a</sup>, Diana Radune<sup>a</sup>, George Dimitrov<sup>a</sup>, Kisha Watkins<sup>a</sup>, Kevin J. B. O'Connor<sup>h</sup>, Shan Teresa R. Utterback<sup>i</sup>, Owen White<sup>a</sup>, Craig E. Rubens<sup>g</sup>, Guido Grandi<sup>c</sup>, Lawrence C. Madoff<sup>e,j</sup>, Dennis L. Kasper<sup>a</sup>, John L. Telford<sup>c</sup>, Michael R. Wessels<sup>d,e</sup>, Rino Rappuoli<sup>c,k,l</sup>, and Claire M. Fraser<sup>a,b,k,m</sup>

<sup>a</sup>Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850; <sup>b</sup>Chiron Vaccines, Via Fiorentina 1, 53100 Siena, Italy; <sup>c</sup>Divis of Infectious Diseases, Children's Hospital, 300 Longwood Avenue, Boston, MA 02115; <sup>d</sup>Harvard Medical School, Boston, MA 02115; <sup>e</sup>Center of Biotechnology, University of Maryland Biotechnology Institute, 701 East Pratt Street, Baltimore, MD 21202; <sup>f</sup>Children's Hospital and Regional Center, 307 Westlake Avenue N, Seattle, WA 98109; <sup>g</sup>The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218; <sup>h</sup>J. Craig Institute, 5 Research Place, Rockville, MD 20850; <sup>i</sup>Channing Laboratory, Brigham and Women's Hospital, 181 Longwood Avenue, Boston, MA and <sup>m</sup>George Washington University Medical Center, 2300 Eye Street NW, Washington, DC 20037

Contributed by Rino Rappuoli, August 5, 2005

 frontiers | Frontiers in Genetics

TYPE Original Research  
 PUBLISHED: 10 August 2023  
 DOI: 10.3389/fgene.2023.1225248

## Pangenome graphs in infectious disease: a comprehensive genetic variation analysis of *Neisseria meningitidis* leveraging Oxford Nanopore long reads

Lea Guerracino<sup>2,3</sup>, Patrick J. Biggs<sup>4,5</sup>, Nuzla Ismail<sup>6</sup>, Jana Renee Wold<sup>7</sup>, <sup>5,8</sup>, Pjotr Prins<sup>2</sup>, Erik Garrison<sup>2</sup> and Joep de Ligt<sup>1\*</sup>  
<sup>1</sup>Centre for Research, Porirua, New Zealand, <sup>2</sup>Department of Genetics, University of Tennessee Health Science Center, Memphis, TN, United States, <sup>3</sup>Human Technopole, Milan, Italy, <sup>4</sup>Molecular Biosciences Group, School of