Genómica Comparativa

Genómica para bioinformática INB320
18 abril 2016
Eduardo Castro-Nallar, PhD
www.castrolab.org

"…from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved"
Charles Darwin

¿Qué llevamos hasta ahora?

- Técnicas de secuenciamiento masivo
- Construcción de genotecas
- Alineamiento de secuencias
- Ensamblaje y anotación de genomas

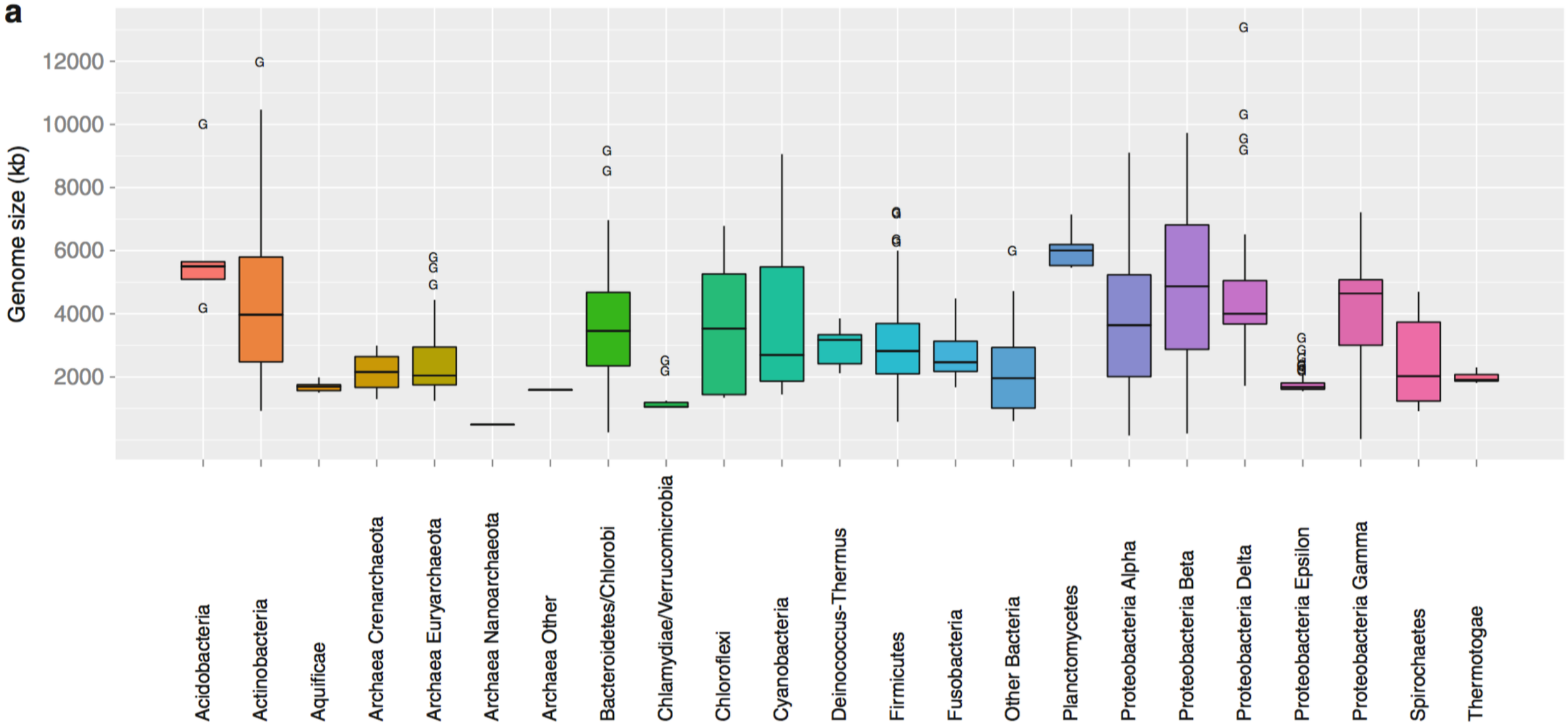
Preprocesamiento

- Genómica comparativa
- Filogenética
- Pangenomas
- Metagenómica

Acción

Todo en genómica es comparativo

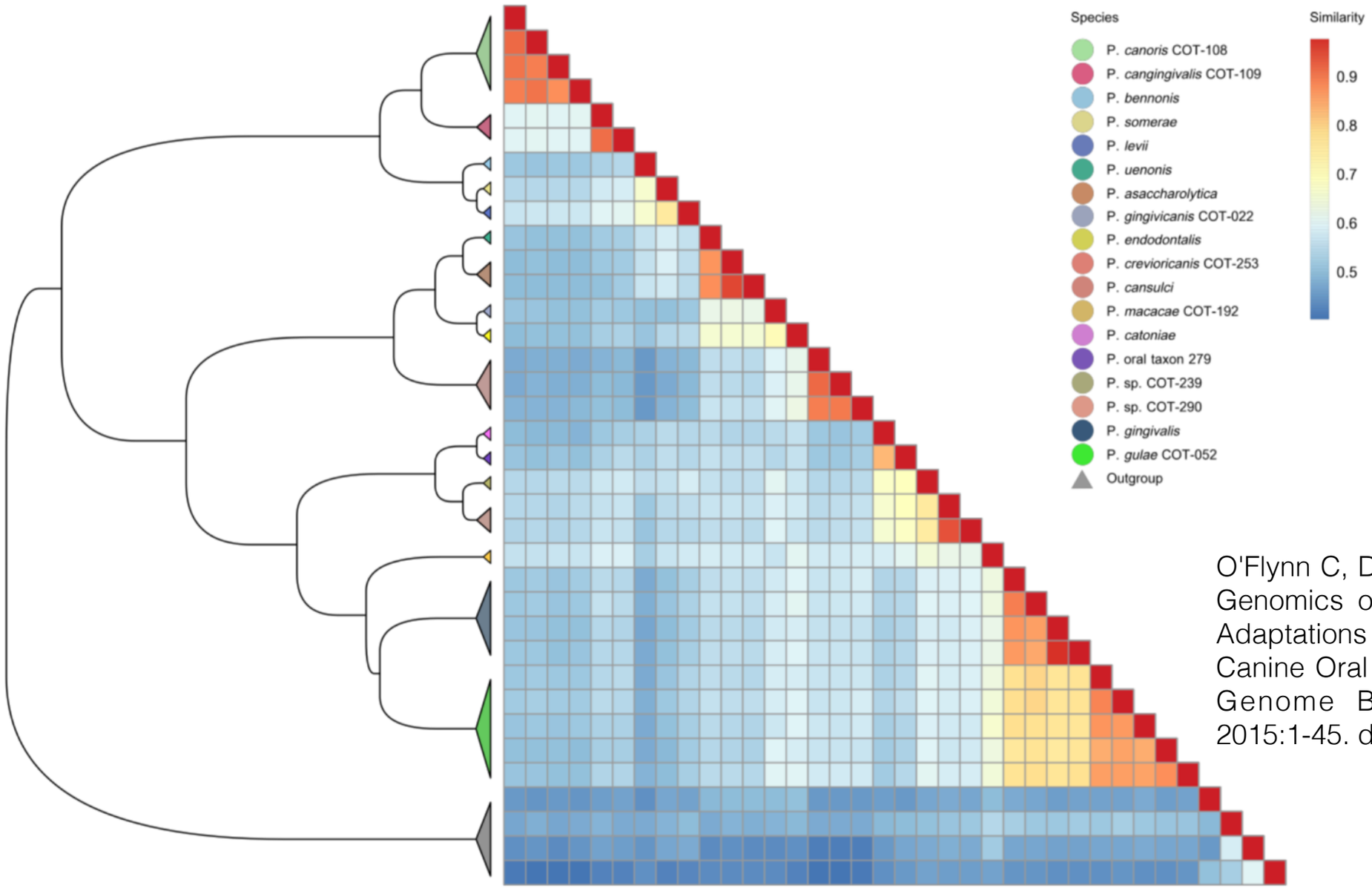
- Genómica de poblaciones en bacterias, virus, multicelulares (e.g., humanos). Virus y bacteria mayoritariamente



Ozen AI, Vesth T, Ussery DW. Comparative Genomics. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, eds. The Prokaryotes. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013:209-227. doi:10.1007/978-3-642-30194-0_11.

Ejemplos en procariontes

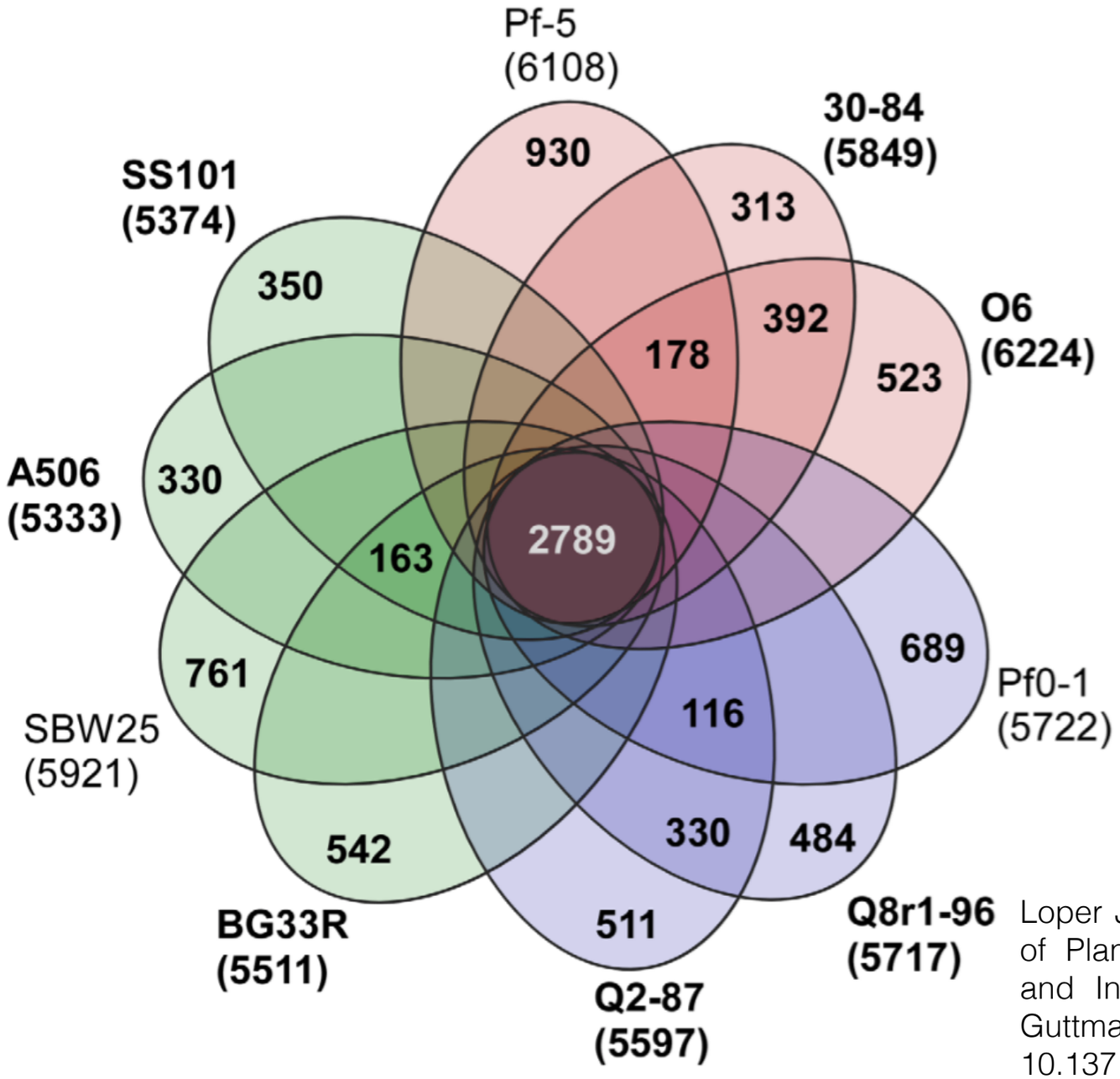
- Similitud y evolución de genomas



O'Flynn C, Deusch O, Darling A, et al. Comparative Genomics of the Genus Porphyromonas Identifies Adaptations for Heme Synthesis within the Prevalent Canine Oral Species Porphyromonas cangingivalis. Genome Biology and Evolution. November 2015:1-45. doi:10.1093/gbe/evv220.

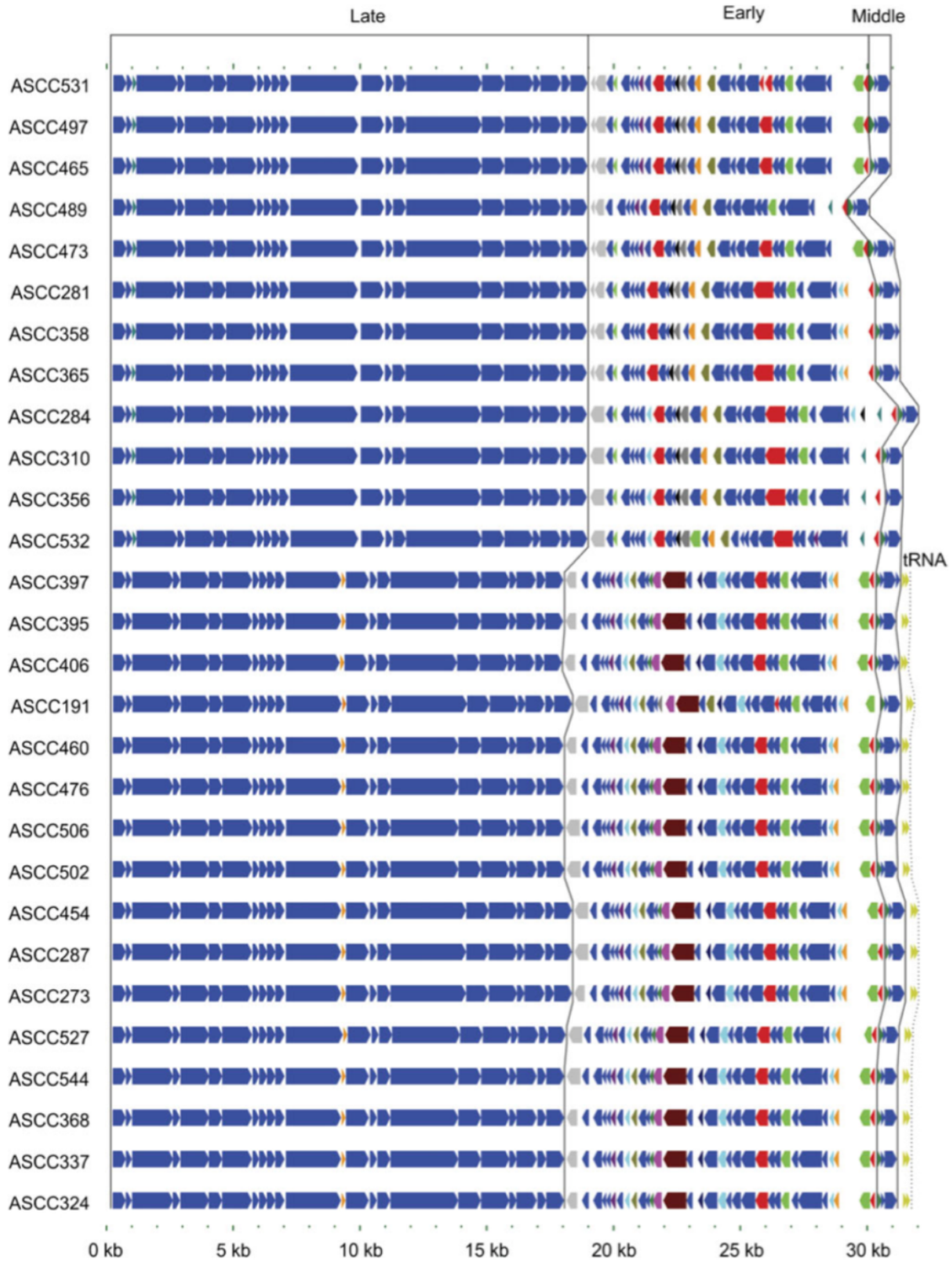
Ejemplos en procariontes

- Genes homólogos entre cepas diferentes - pangenoma



Loper JE, Hassan KA, Mavrodi DV, et al. Comparative Genomics of Plant-Associated Pseudomonas spp.: Insights into Diversity and Inheritance of Traits Involved in Multitrophic Interactions. Guttman DS, ed. PLoS Genet. 2012;8(7):e1002784–27. doi: 10.1371/journal.pgen.1002784.

Ejemplos en procariontes

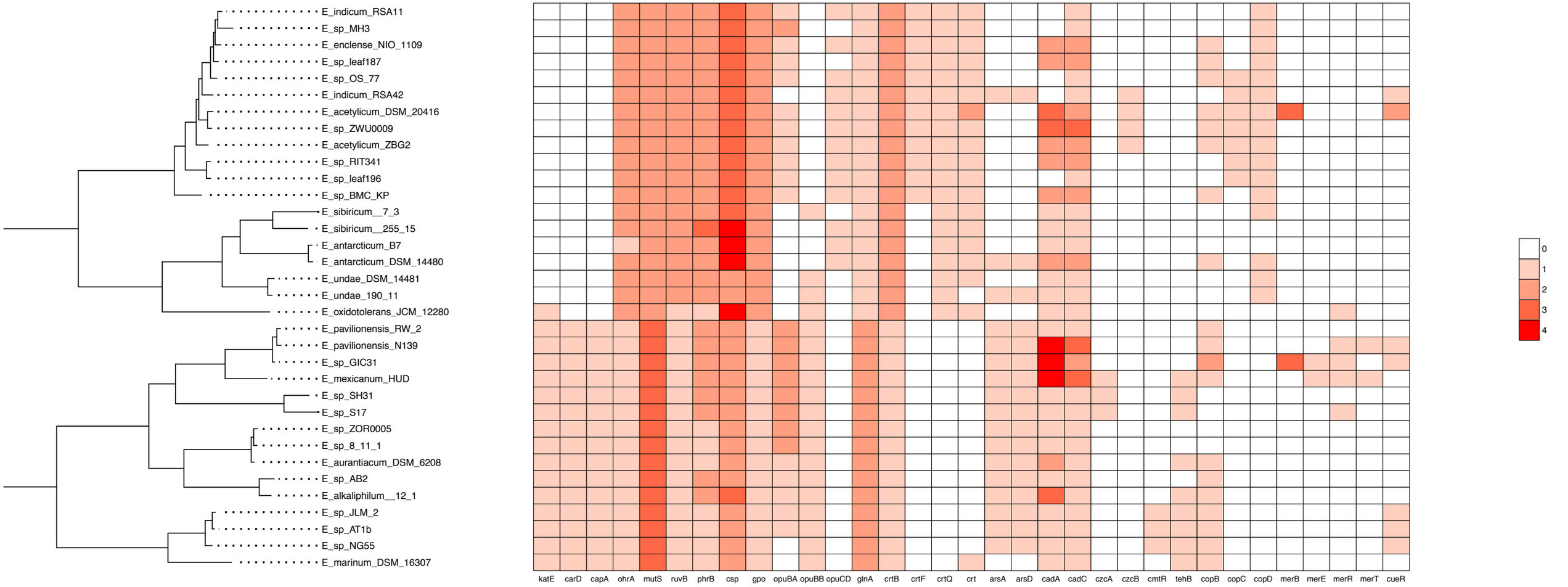


- Estructura de genomas, funcional y sintenia

Castro-Nallar E, Chen H, Gladman S, et al. Population Genomics and Phylogeography of an Australian Dairy Factory Derived Lytic Bacteriophage. Genome Biology and Evolution. 2012;4(3):382-393. doi: 10.1093/gbe/evs017.

Ejemplos en procariontes

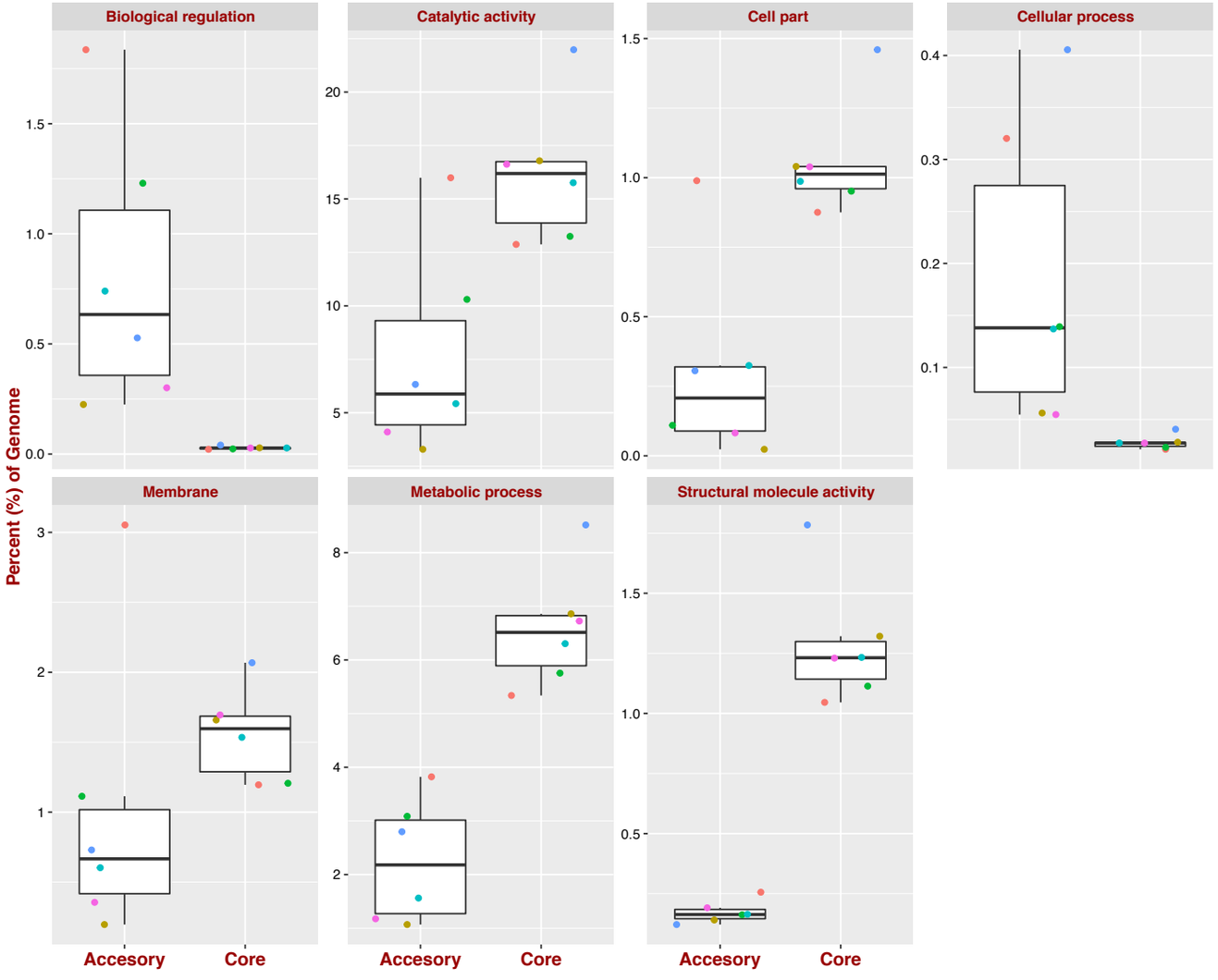
- Relaciones filogenéticas y número de copias de familias génicas

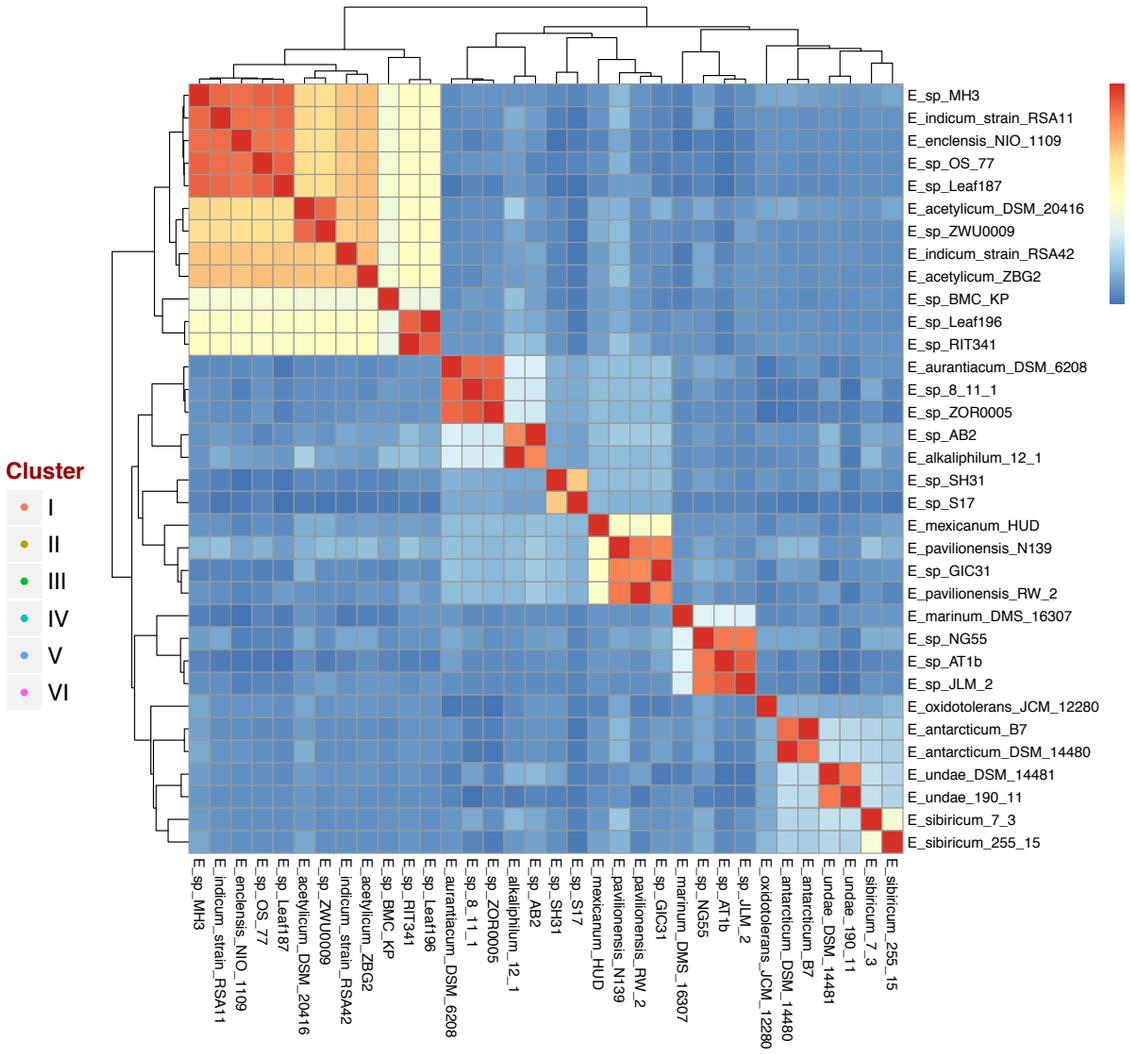


Castro-Severyn, et al. Comparative genomics of Exiguobacterium. Unpublished results

Ejemplos en procariontes

- Procesos biológicos y genomas

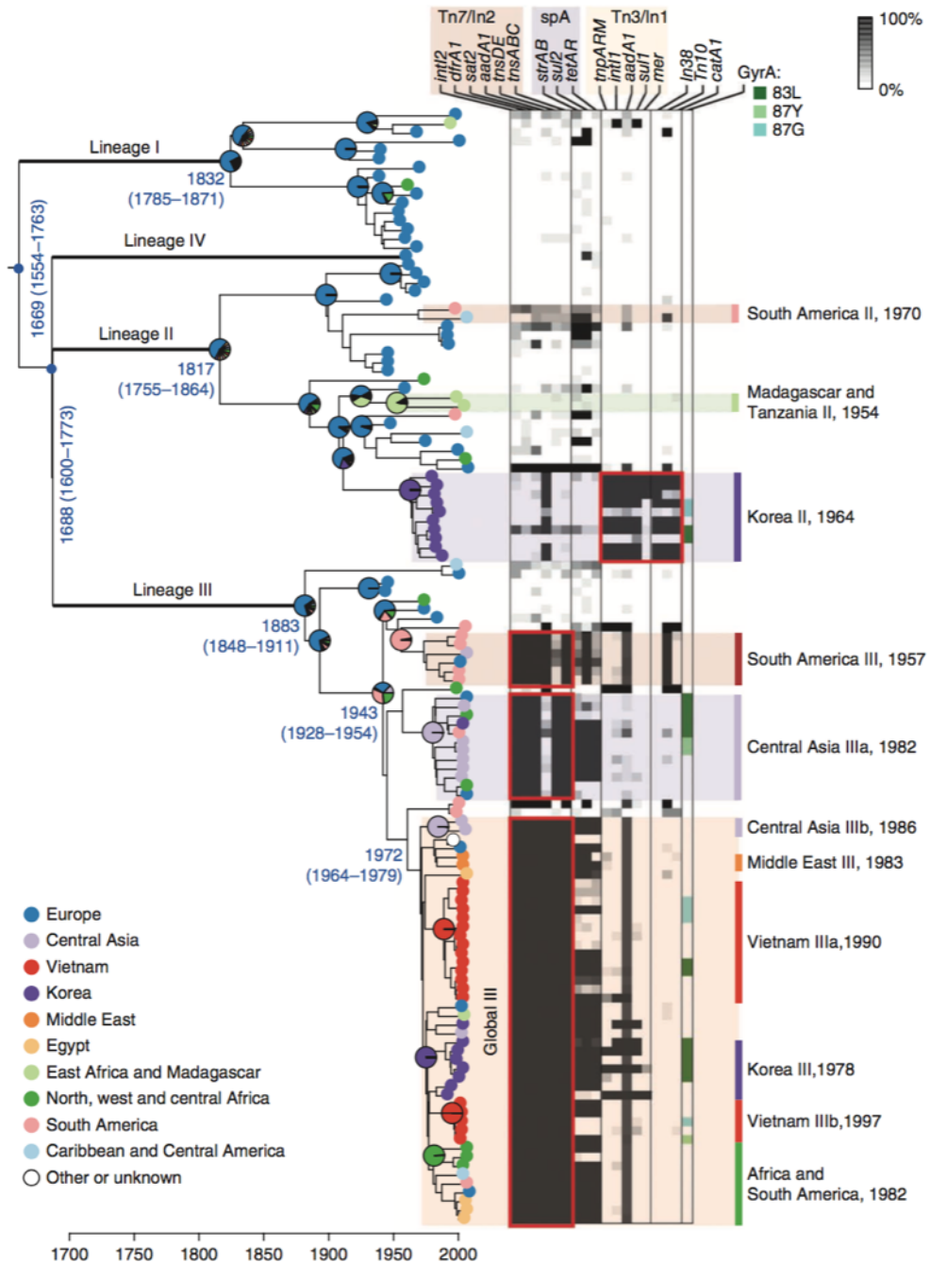


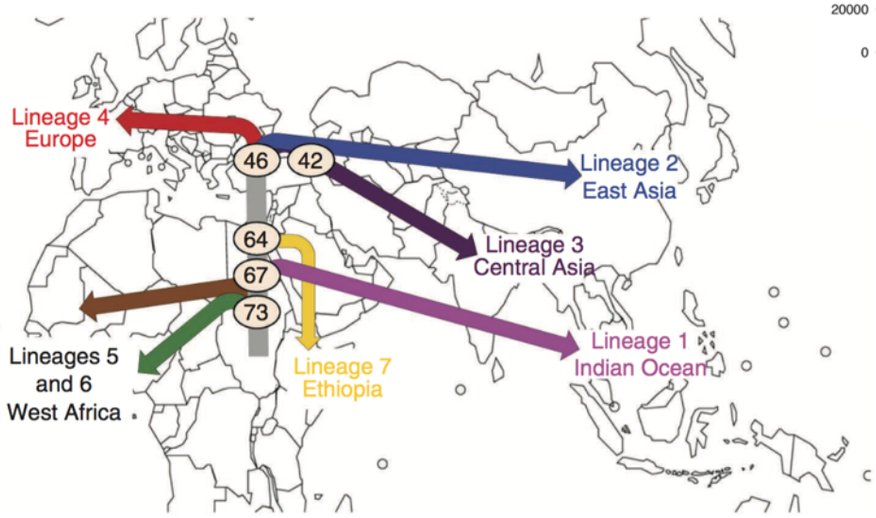


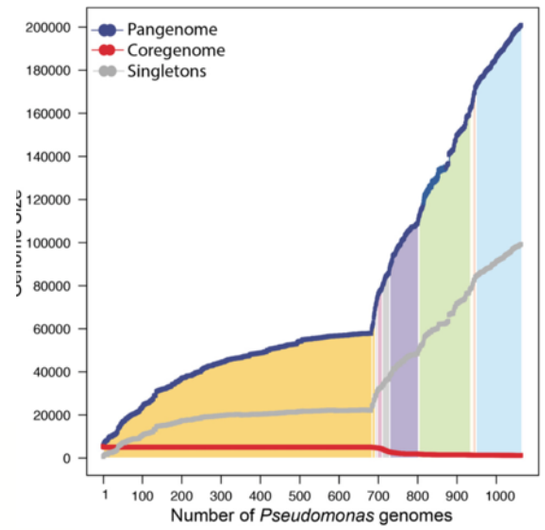
Castro-Severyn, et al. Comparative genomics of Exiguobacterium. Unpublished results

Ejemplos en procariontes

- Relojes moleculares - filogeografía - pangenomas —> cientos o miles de genomas







Comas I, Coscolla M, Luo T, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nature Genetics. September 2013. doi:10.1038/ng.2744.

Jun S-R, Wassenaar TM, Nookaew I, et al. Diversity of Pseudomonas Genomes, Including Populus-Associated Isolates, as Revealed by Comparative Genome Analysis. Kivisaar M, ed. Applied and Environmental Microbiology. 2015;82(1):375-383. doi:10.1128/AEM.02612-15.

Holt KE, Baker S, Weill F-X, et al. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nature Genetics. 2012;44(9):1056-1059. doi:10.1038/ng.2369.

También en eucariontes

The UK10K project identifies rare variants in health and disease

The UK10K Consortium*

The contribution of rare and low-frequency variants to human traits is largely unexplored. Here we describe insights from sequencing whole genomes (low read depth, 7×) or exomes (high read depth, 80×) of nearly 10,000 individuals from population-based and disease collections. In extensively phenotyped cohorts we characterize over 24 million novel sequence variants, generate a highly accurate imputation reference panel and identify novel alleles associated with levels of triglycerides (*APOB*), adiponectin (*ADIPOQ*) and low-density lipoprotein cholesterol (*LDLR* and *RGAG1*) from single-marker and rare variant aggregation tests. We describe population structure and functional annotation of rare and low-frequency variants, use the data to estimate the benefits of sequencing for association studies, and summarize lessons from disease-specific collections. Finally, we make available an extensive resource, including individual-level genetic and phenotypic data and web-based tools to facilitate the exploration of association results.

También en eucariontes

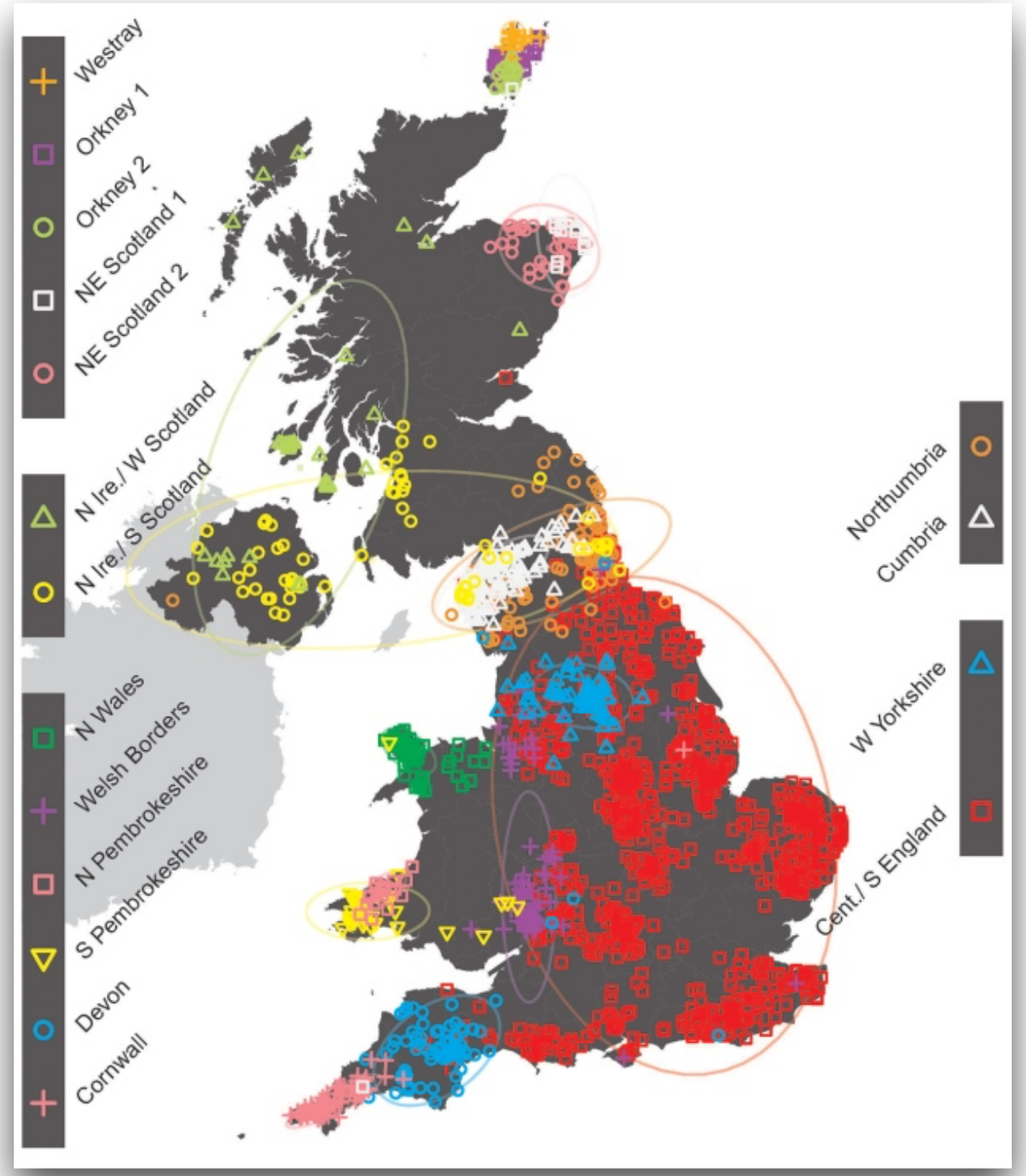
- No solo para relaciones genotipo enfermedad
- Migración humana, mestizaje

ARTICLE

doi:10.1038/nature14230

The fine-scale genetic structure of the British population

Stephen Leslie[1,2,3]*, Bruce Winney[3]*, Garrett Hellenthal[4]*, Dan Davison[5], Abdelhamid Boumertit[3], Tammy Day[3], Katarzyna Hutnik[3], Ellen C. Royrvik[3], Barry Cunliffe[6], Wellcome Trust Case Control Consortium 2†, International Multiple Sclerosis Genetics Consortium†, Daniel J. Lawson[7], Daniel Falush[8], Colin Freeman[9], Matti Pirinen[10], Simon Myers[11], Mark Robinson[12], Peter Donnelly[9,11]§ & Walter Bodmer[3]§

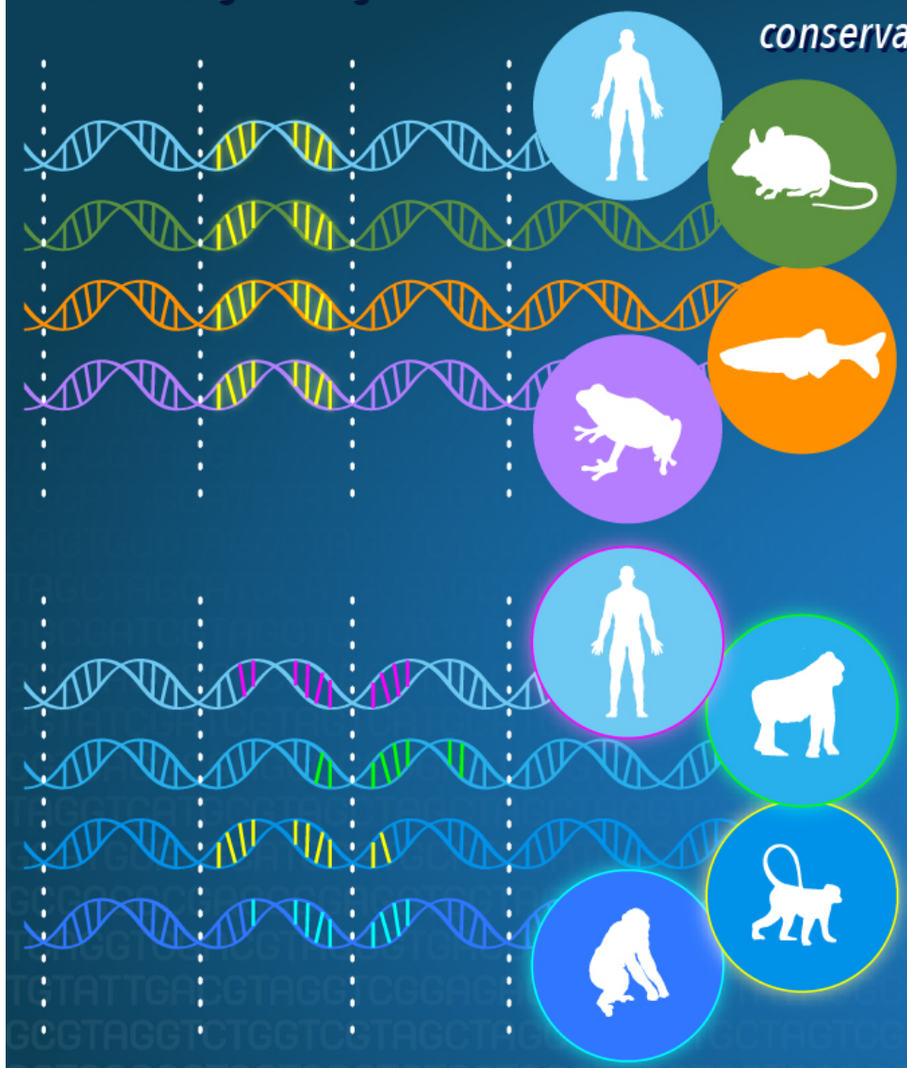


COMPARATIVE GENOMICS

NHGRI FACT SHEETS

genome.gov

Researchers choose the appropriate time-scale of evolutionary conservation for the question being addressed.



Common features of different organisms such as humans and fish are often encoded within the DNA evolutionarily conserved between them.

Looking at **closely related species** such as humans and chimpanzees shows which genomic elements are unique to each.

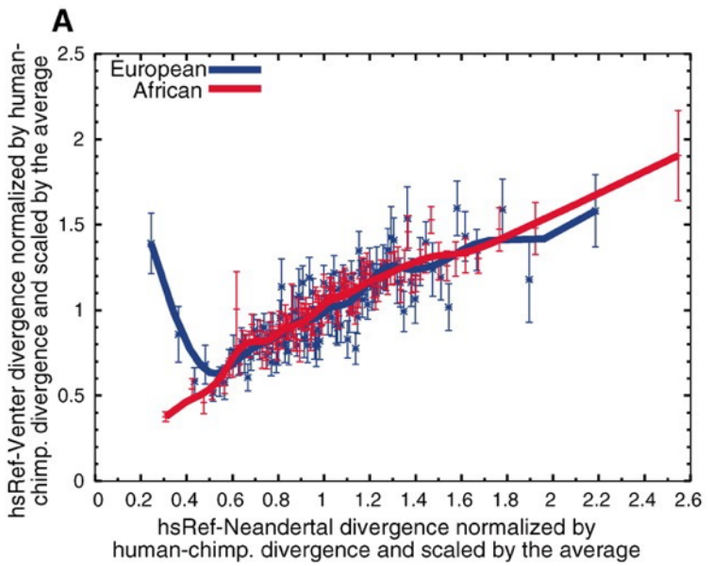
Genetic differences **within one species** such as our own can reveal variants with a role in disease.

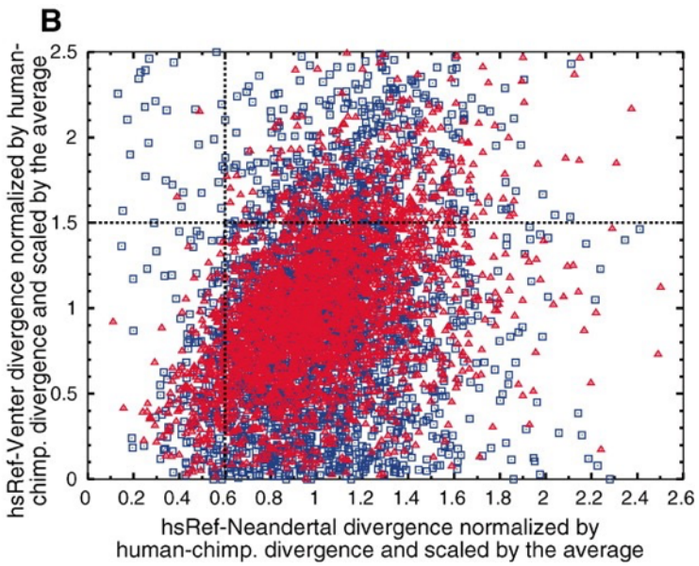
NIH National Human Genome Research Institute

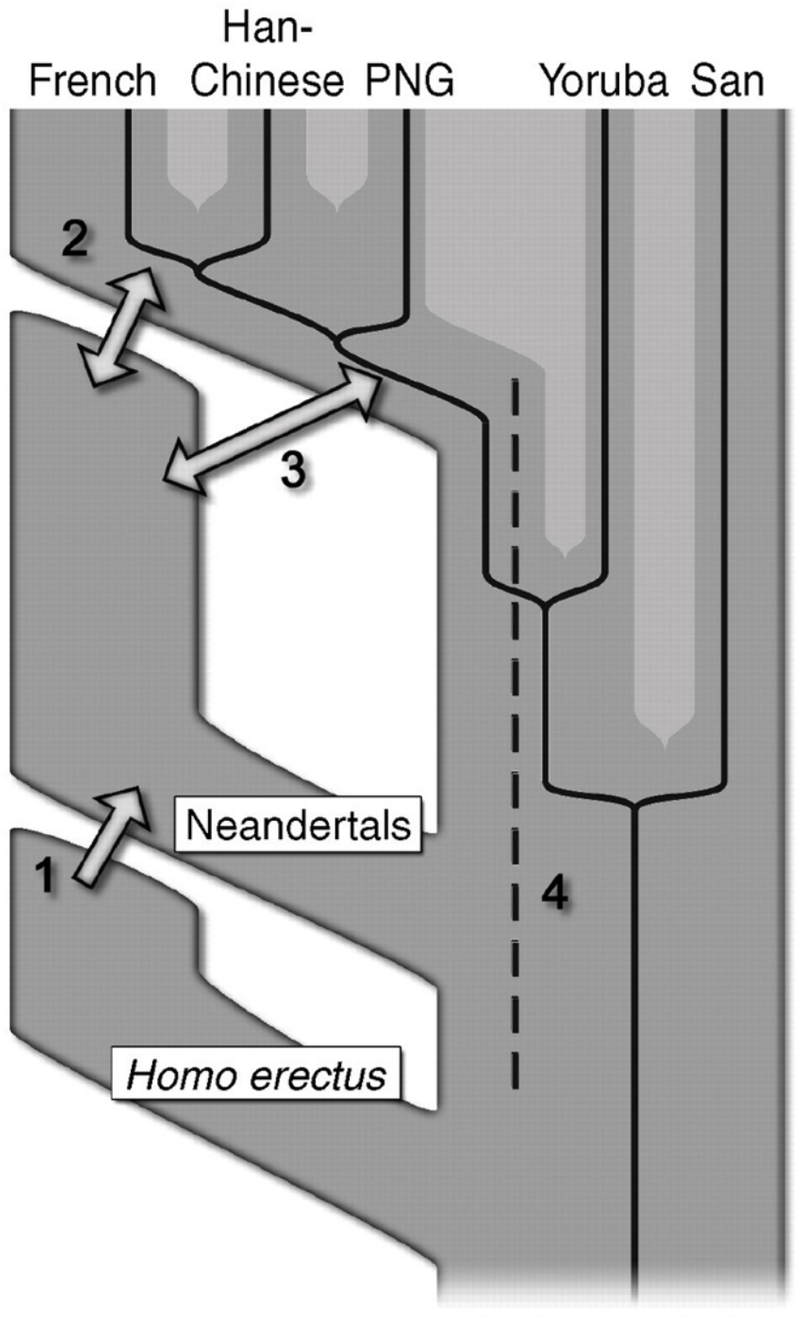
Ancestros: profundos y recientes

- Neandertales son más parecidos a europeos que a africanos

- Modelo de especiación y flujo génico entre neandertales y humanos modernos



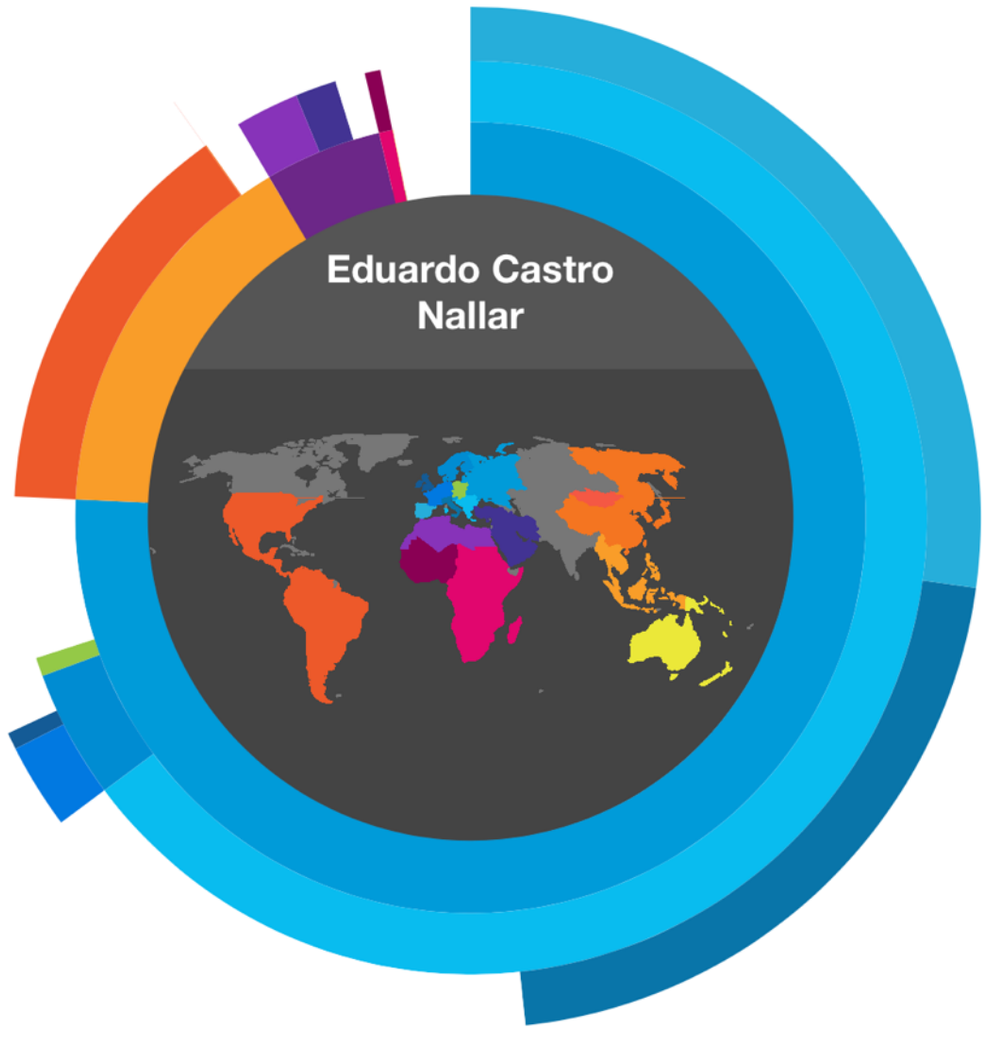




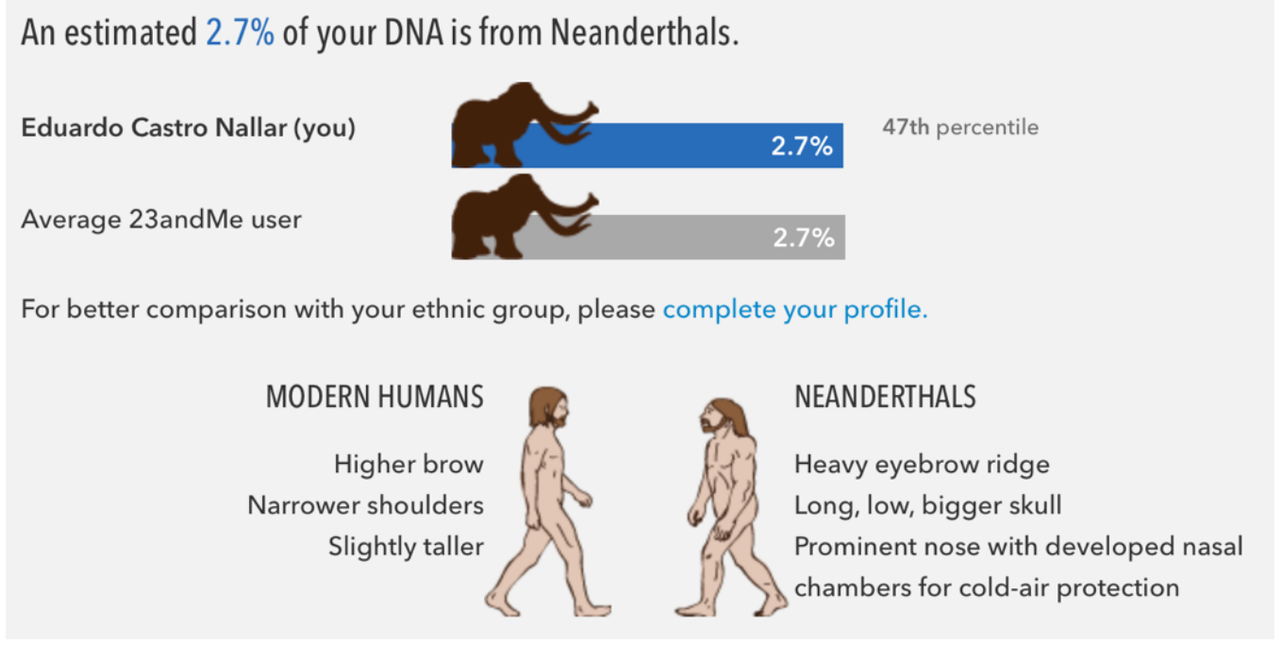
Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y. and Hansen, N.F., 2010. A draft sequence of the Neandertal genome. *science*, *328*(5979), pp.710-722.

Ancestros: profundos y recientes

- Composición genómica —> impacto en predisposición a enfermedades genéticas



- Mestizaje entre humanos modernos y neandertales



23andme.com

¿Qué es la genómica comparativa?

Comparative genomics

From Wikipedia, the free encyclopedia

Comparative genomics is a field of biological research in which the genomic features of different organisms are compared.[2][3] The genomic features may include the DNA sequence, genes, gene order, regulatory sequences, and other genomic structural landmarks.[3] In this branch of genomics, whole or large parts of genomes resulting from genome projects are compared to study basic biological similarities and differences as well as evolutionary relationships between organisms.[2][4][5] The major principle of comparative genomics is that common features of two organisms will often be encoded within the DNA that is evolutionarily conserved between them.[6] Therefore, comparative genomic approaches start with making some form of alignment of genome sequences and looking for orthologous sequences (sequences that share a common ancestry) in the aligned genomes and checking to what extent those sequences are conserved. Based on these, genome and molecular evolution are inferred and this may in turn be put in the context of, for example, phenotypic evolution or population genetics.[7]

https://en.wikipedia.org/wiki/Comparative_genomics

Comparison of whole genome sequences provides a highly detailed view of how organisms are related to each other at the genetic level. How are genomes compared and what can these findings tell us about how the overall structure of genes and genomes have evolved?

http://www.nature.com/scitable/knowledge/library/comparative-genomics-13239404

Estrategias y preguntas

- Alineamientos genómicos - establecer una hipótesis de homología a nivel de genoma

- Filogenética* - establecer relaciones evolutivas entre genomas a comparar

- Pangenomas - establecer genes comunes y accesorios en un grupo de genomas

- Metagenomas* - saltar un nivel jerárquico y establecer estructura, relaciones, y características genéticas de interés: familias de genes

Alineamientos genómicos - Mauve

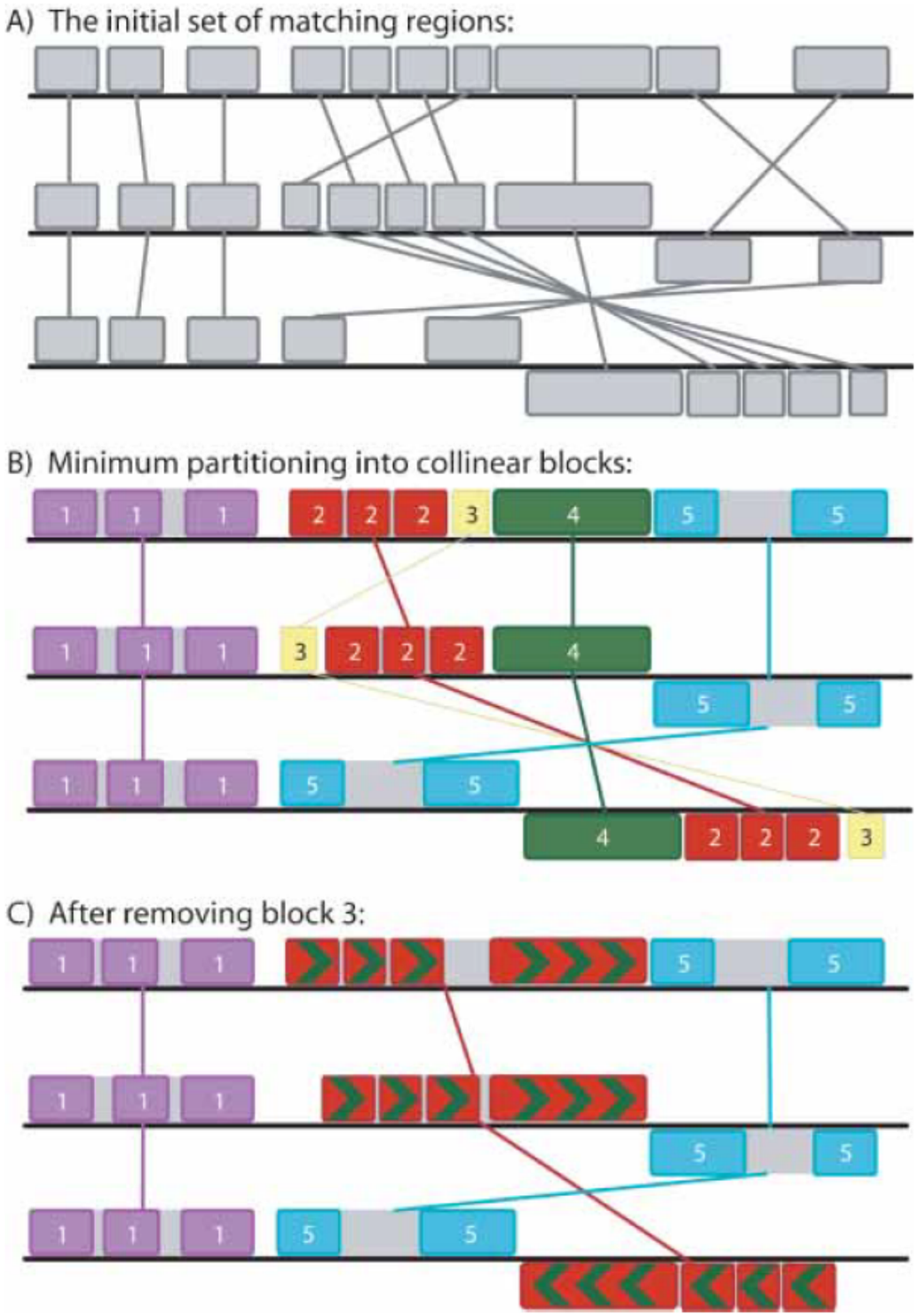
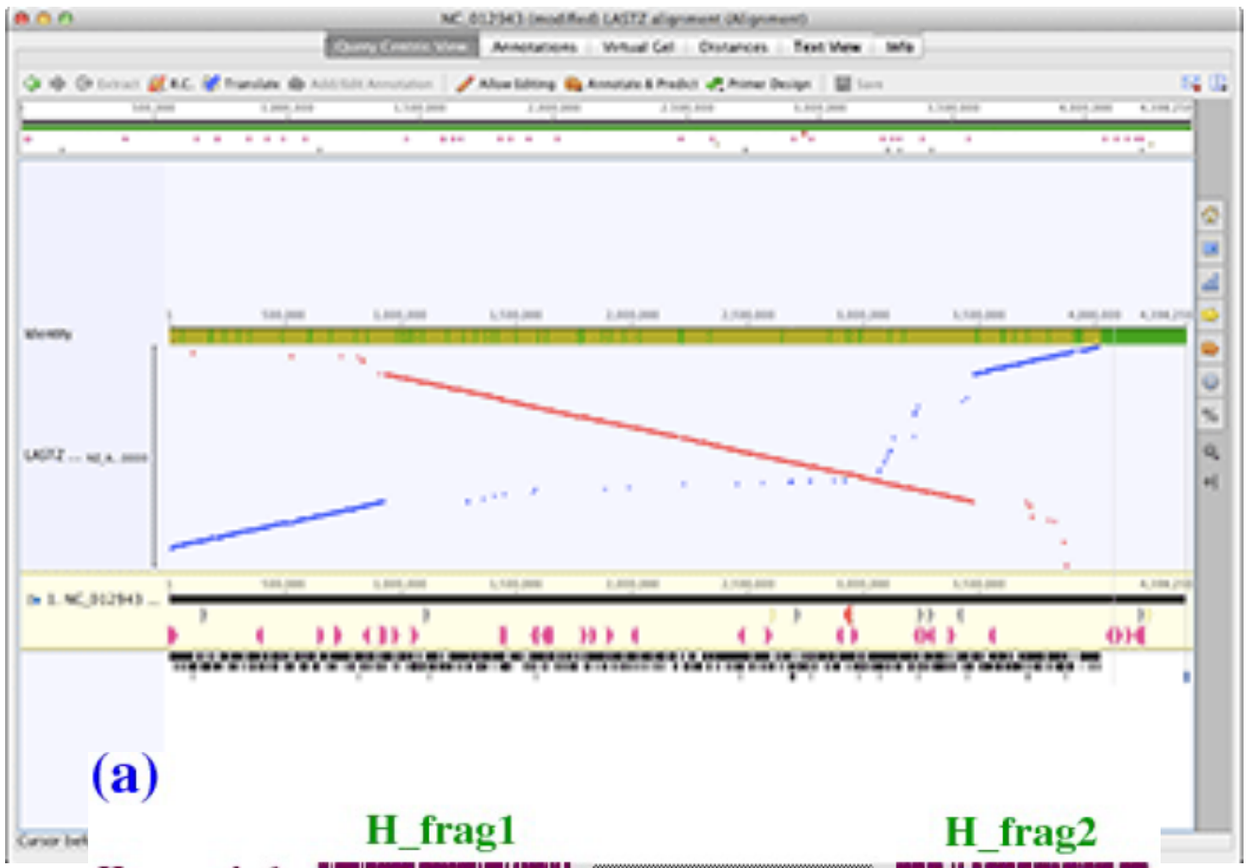
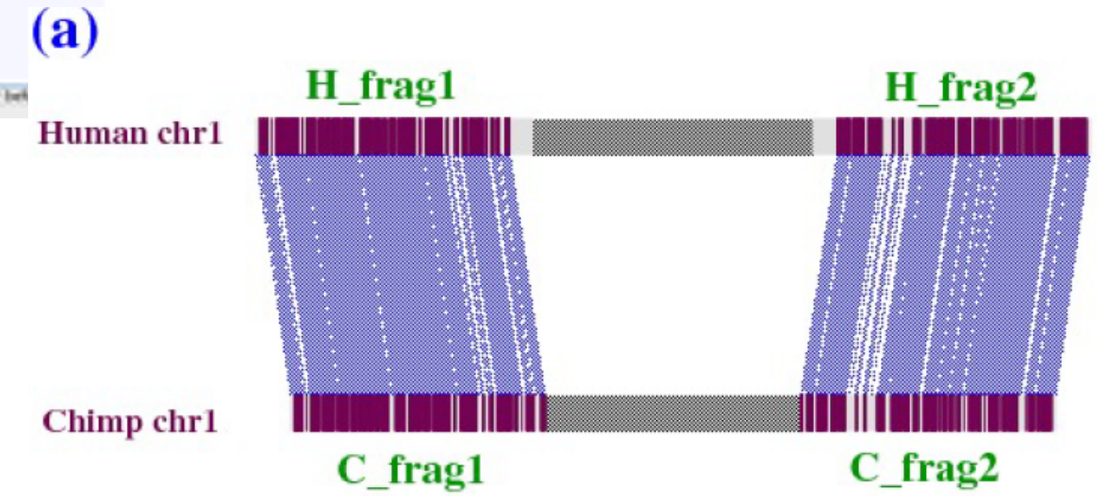


Figure 1 A pictorial representation of greedy breakpoint elimination in three genomes. (*A*) The algorithm begins with the initial set of matching regions (multi-MUMs) represented as connected blocks. Blocks below a genome's center line are inverted relative to the reference sequence. (*B*) The matches are partitioned into a minimum set of collinear blocks. Each sequence of identically colored blocks represents a collinear set of matching regions. One connecting line is drawn per collinear block. Block 3 (yellow) has a low weight relative to other collinear blocks. (*C*) As low-weight collinear blocks are removed, adjacent collinear blocks coalesce into a single block, potentially eliminating one or more breakpoints. Gray regions within collinear blocks are targeted by recursive anchoring.

Darling, A.C., Mau, B., Blattner, F.R. and Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, *14*(7), pp.1394-1403.

Alineamientos genómicos - LASTZ





MASKING: Both genomes have to be repeatmasked and masked Tandem Repeat Finder (trf) first

ALIGNING: The two genomes are aligned with LASTZ. This generates lav-files, which have to be converted to psl (lavToPsl)

CHAINING: Two matching alignments next to each other are joined into one fragment if they are close enough (axtChain). As every genomic fragment can match with several others, we keep only the longest chains : first do axtSort then filter with axtBest

NETTING: Group blocks of chained alignments into longer stretches of synteny (netChain)

MAF'ING: From the synteny-files (positions), get the sequences and re-create alignments

http://www.bx.psu.edu/miller_lab/dist/README.lastz-1.02.00/README.lastz-1.02.00a.html

http://arxiv.org/pdf/1407.3895.pdf

Filogenética

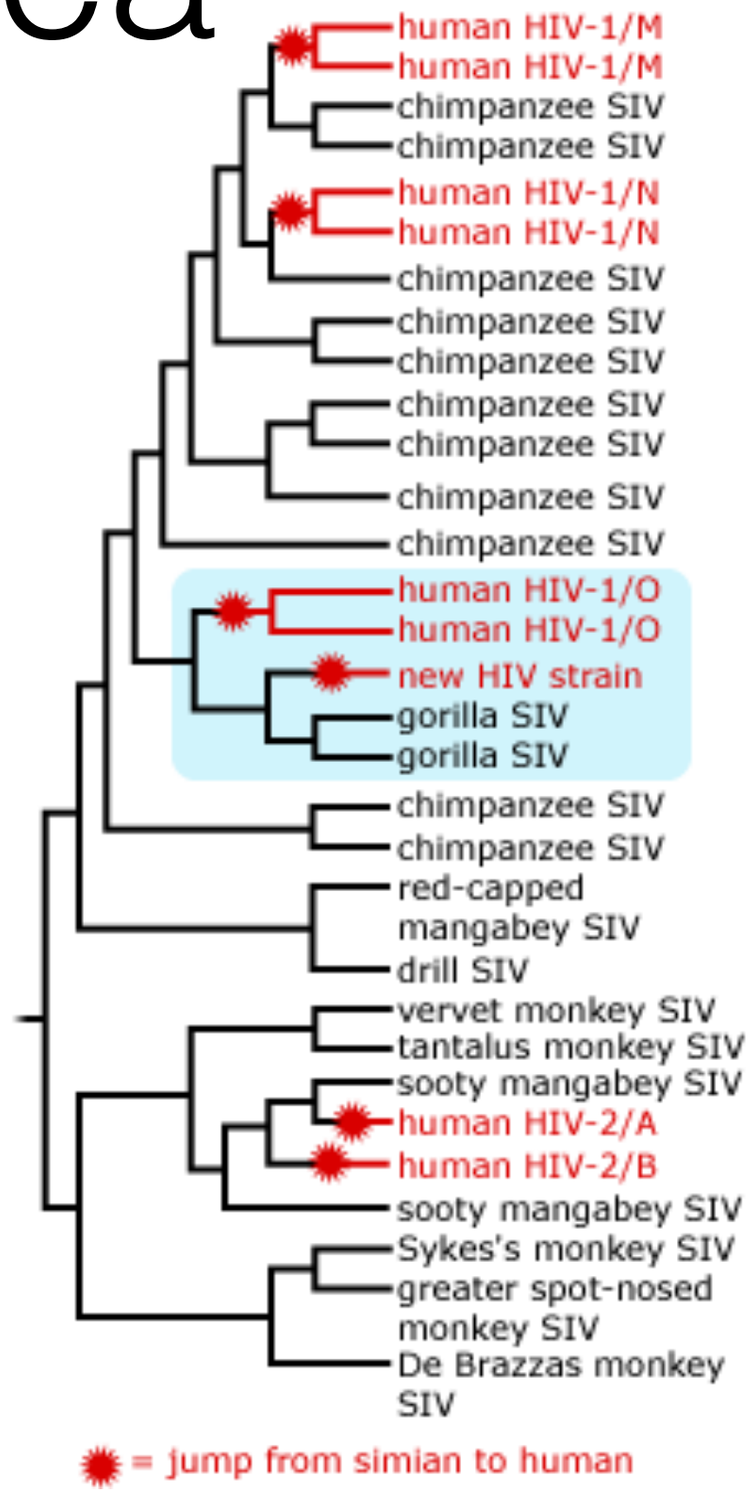
- Central para establecer un marco conceptual respecto de los genomas que están siendo comparados

- Una explicación de cómo las secuencias han evolucionado, sus relaciones genealógicas, y por lo tanto de cómo han llegado a ser de la forma que son hoy día

https://www.ebi.ac.uk/training/online/course/introduction-phylogenetics/what-phylogeny
https://www.ebi.ac.uk/training/online/course/introduction-phylogenetics/why-phylogenetics-important

Filogenética

- Secuencias homólogas
- Alineamiento
- Modelo de evolución
- Criterio de optimalidad



https://www.uni-ulm.de/universitaet/veranstaltungen/jahrestag/45-jahrestag-2012/promotionspreise-uug/dr-biol-hum-daniel-sauter.html

Filogenética

- Métodos basados en distancia = UPGMA, NJ
- Máxima Parsimonia
- Máxima Verosimilitud
- Inferencia Bayesiana
- <u>Conflicto</u>

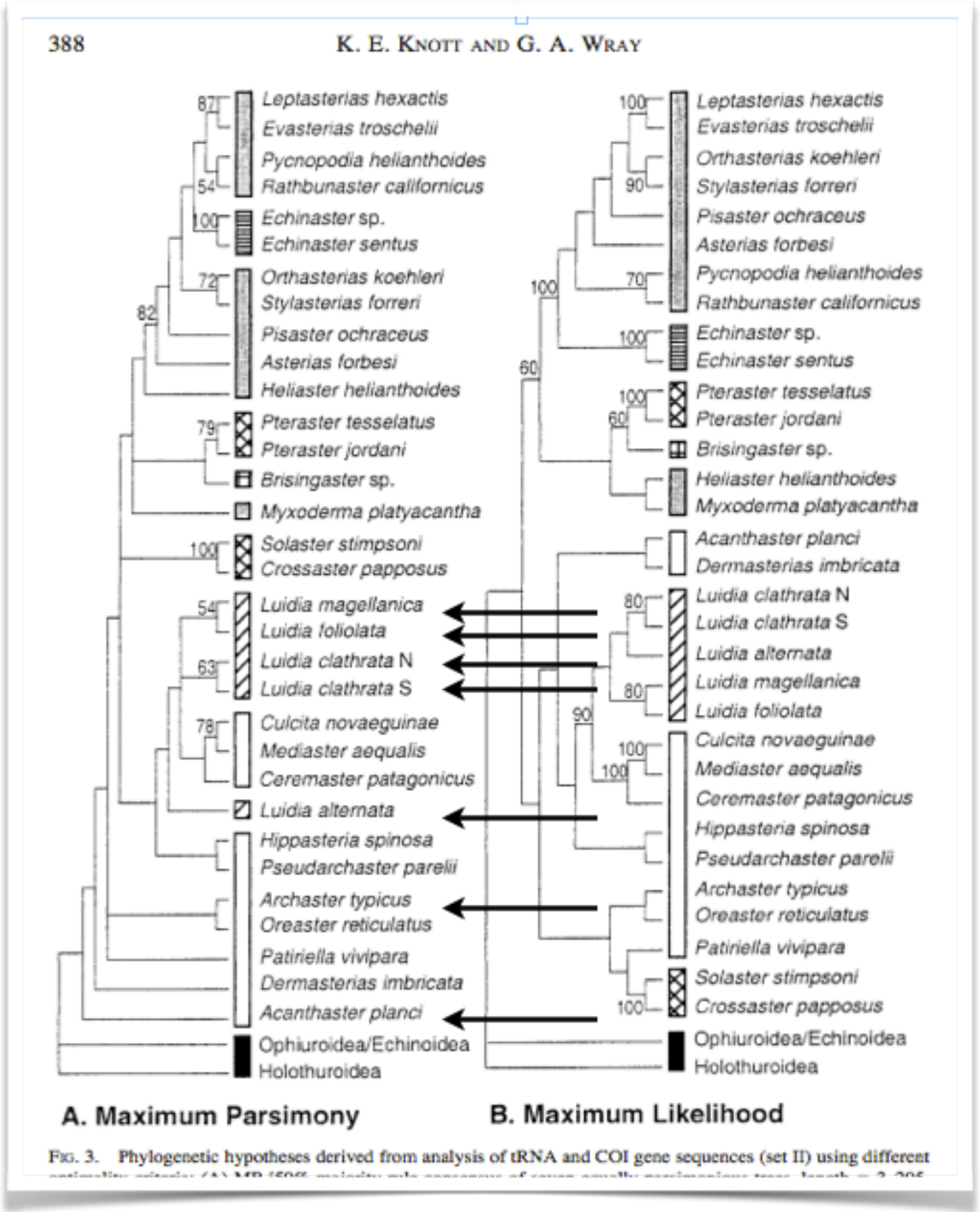
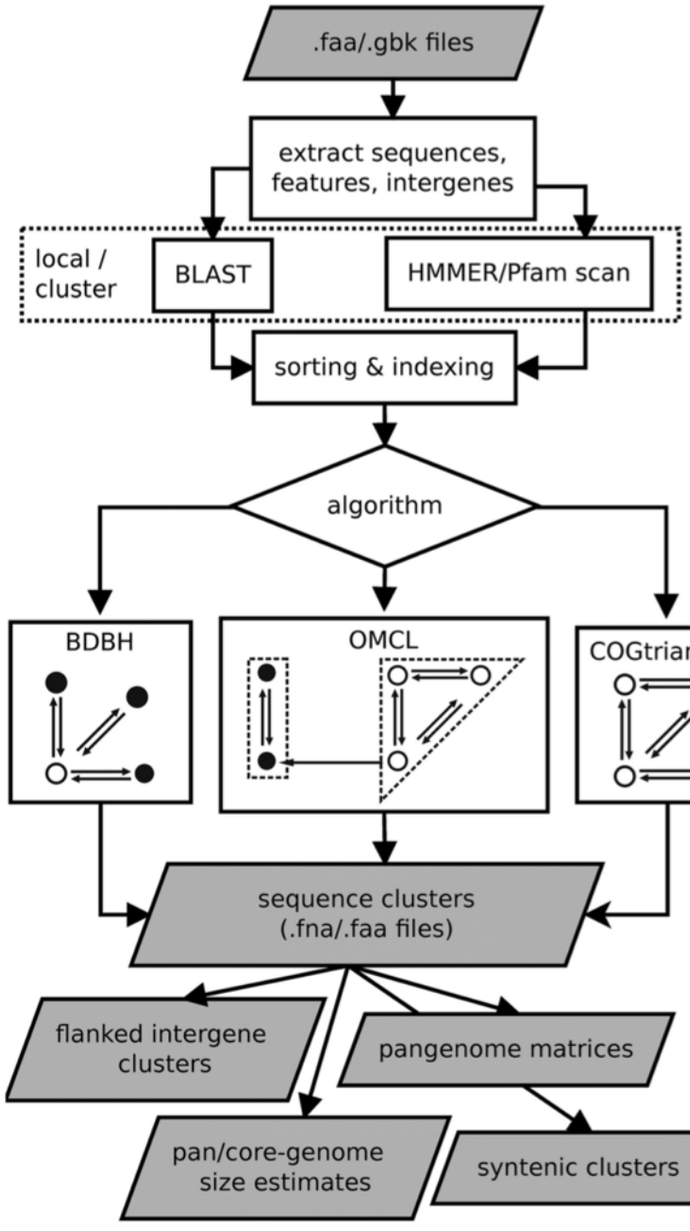
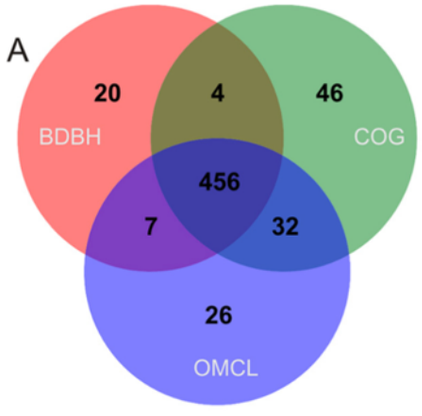


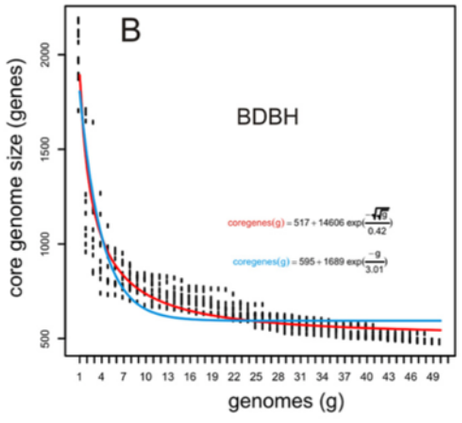
FIG. 3. Phylogenetic hypotheses derived from analysis of tRNA and COI gene sequences (set II) using different

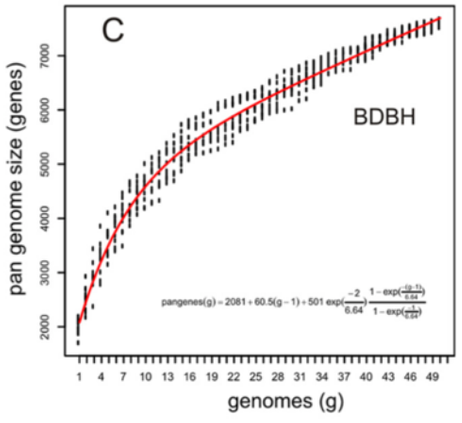
Pangenomas

- Un genoma no representa a una especie
- Genes en común y genes accesorios
- Genes comunes = funciones básicas
- Genes accesorios = adaptaciones locales



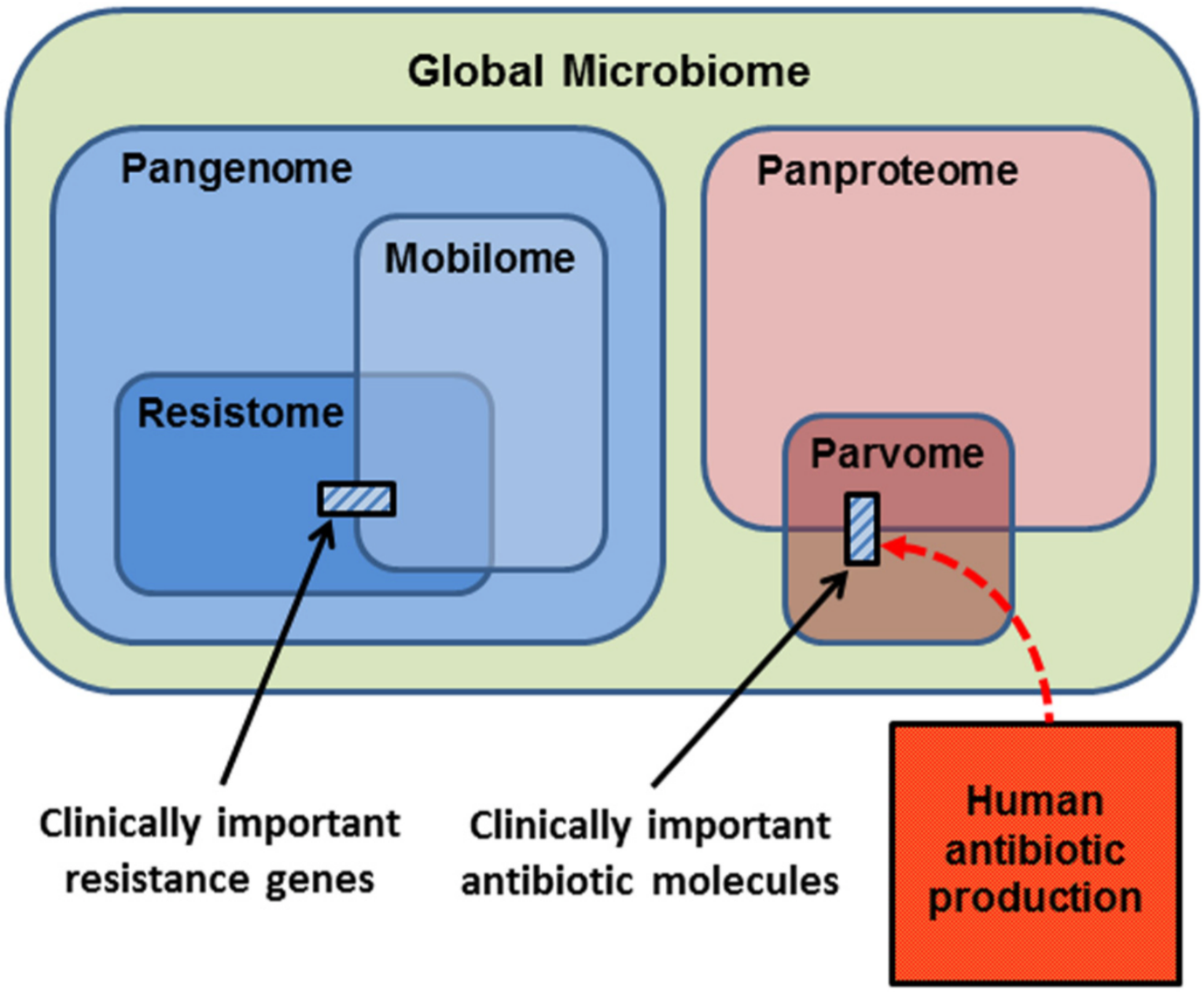






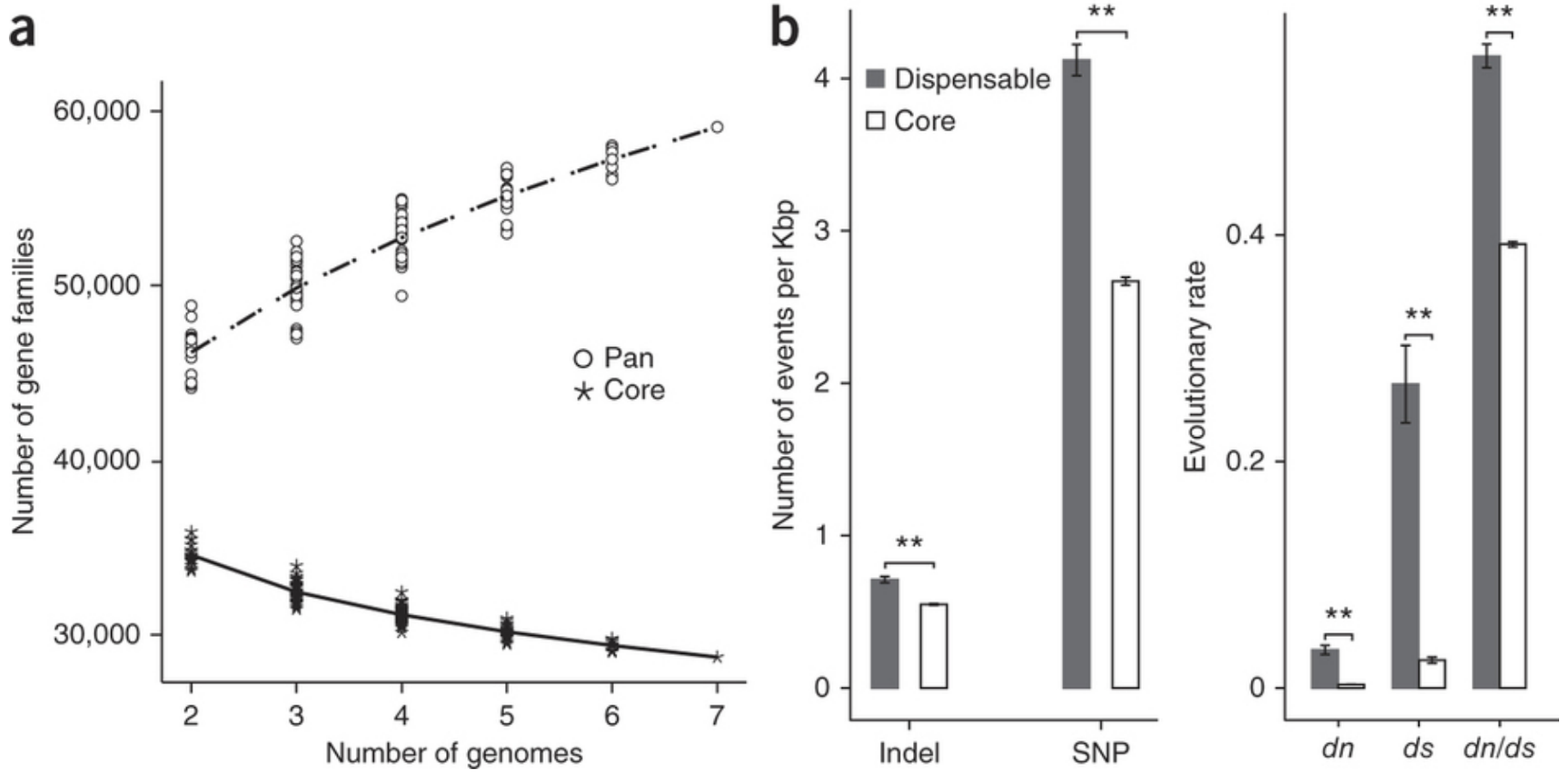
Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. Applied and Environmental Microbiology. 2013;79(24):7696-7701. doi:10.1128/AEM.02411-13.

Representación conceptual del pangenoma



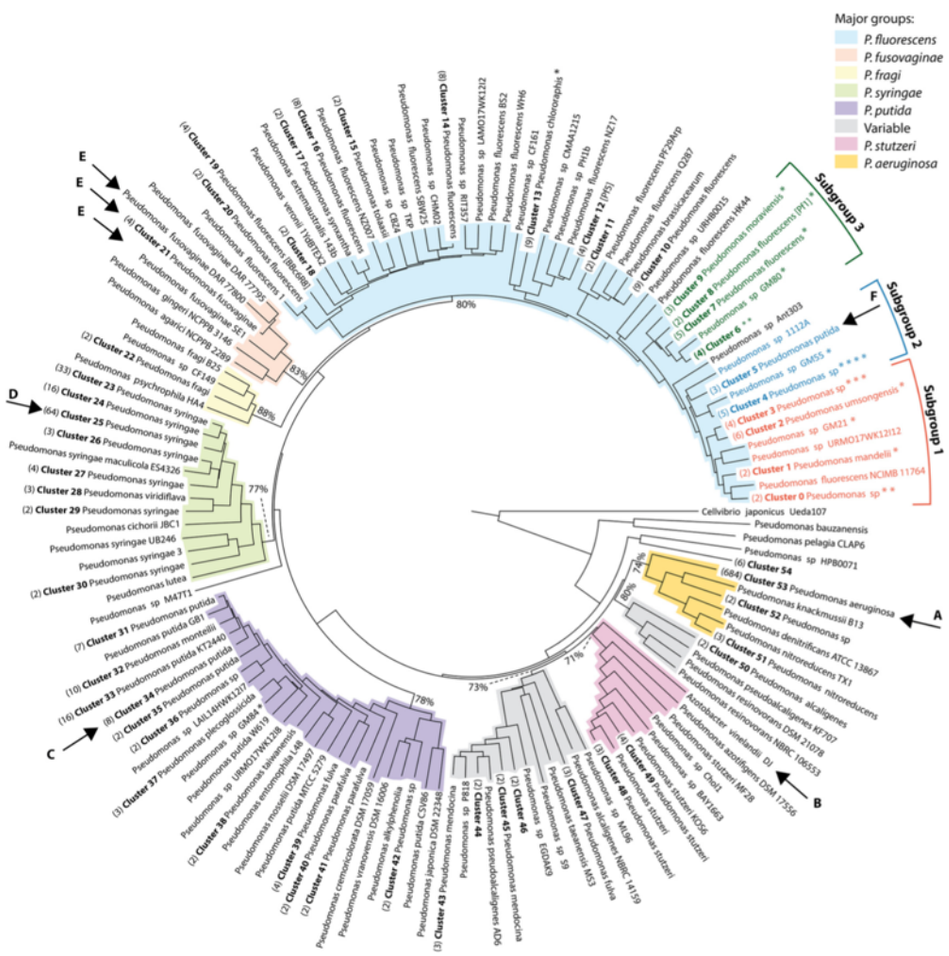
Pangenomas

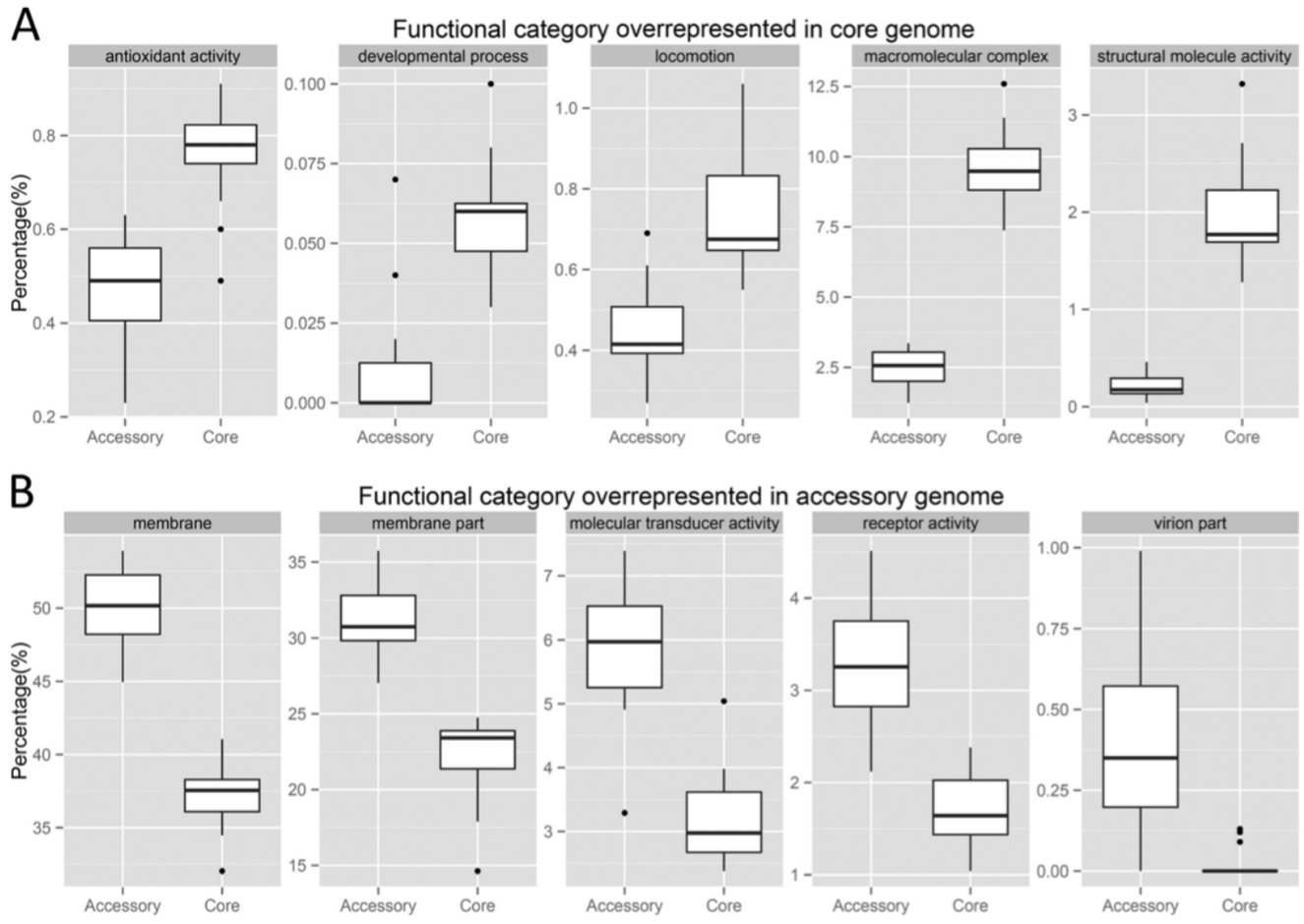
- “Thus, the de novo construction of a pan-genome for a species, consisting of a core genome shared among individuals and individual-specific or partially shared dispensable genome, is necessary to capture the majority of genetic diversity within a species”



Li, Y.H., Zhou, G., Ma, J., Jiang, W., Jin, L.G., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L. and Zhang, S.S., 2014. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature biotechnology*, *32*(10), pp.1045-1052.

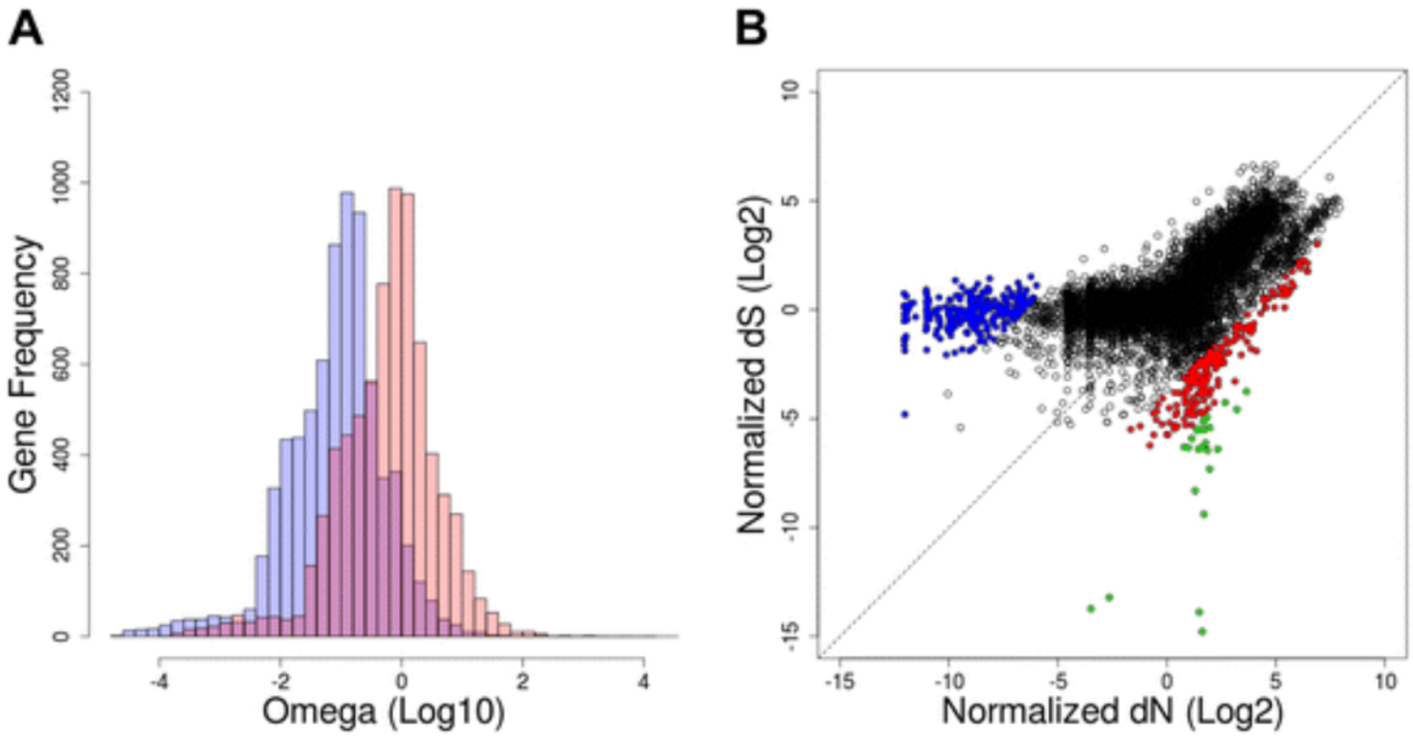
Pangenomas





Jun S-R, Wassenaar TM, Nookaew I, et al. Diversity of Pseudomonas Genomes, Including Populus-Associated Isolates, as Revealed by Comparative Genome Analysis. Kivisaar M, ed. Applied and Environmental Microbiology. 2015;82(1):375-383. doi:10.1128/AEM.02612-15.

Pangenomas y selección positiva



Gene family[a]	Function[b]	Omega (ω)	Length (bp)	Strain frequency[c]
3333	Chitin binding protein	108	1170	98.9 % (179)
3675	Flagellar basal-body rod protein FlgF	5884	750	99.5 % (180)
4766	Predicted branched-chain amino acid permease AzlC	86	763	96.7 % (175)
5348	Unknown function	32	573	99.5 % (180)

Mosquera-Rendón, J., Rada-Bravo, A.M., Cárdenas-Brito, S., Corredor, M., Restrepo-Pineda, E. and Benítez-Páez, A., 2016. Pangenome-wide and molecular evolution analyses of the Pseudomonas aeruginosa species. *BMC genomics*, *17*(1), p.1.

Buen lugar para comenzar

HOSTED BY

ELSEVIER

Genomics, Proteomics & Bioinformatics

Volume 13, Issue 1, February 2015, Pages 73–76

Open Access

Resource Review

A Brief Review of Software Tools for Pangenomics

Jingfa Xiao [a], Zhewen Zhang[b], Jiayan Wu[c], Jun Yu[d]

Show more

doi:10.1016/j.gpb.2015.01.007

ELSEVIER

Current Opinion in Microbiology

Volume 23, February 2015, Pages 148–154

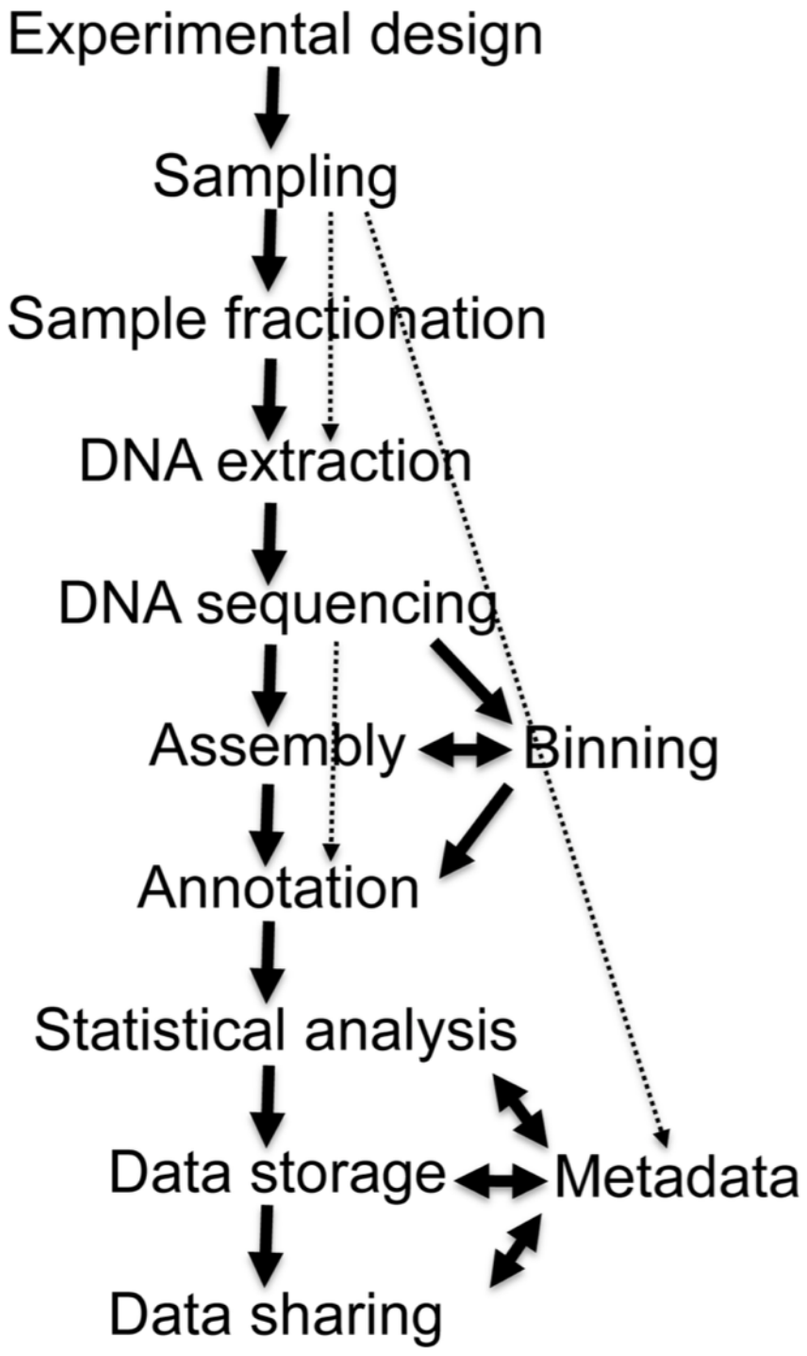
Host–microbe interactions: bacteria • Genomics

Ten years of pan-genome analyses

George Vernikos[1], Duccio Medini[2], David R Riley[3], Hervé Tettelin[3],

Metagenómas

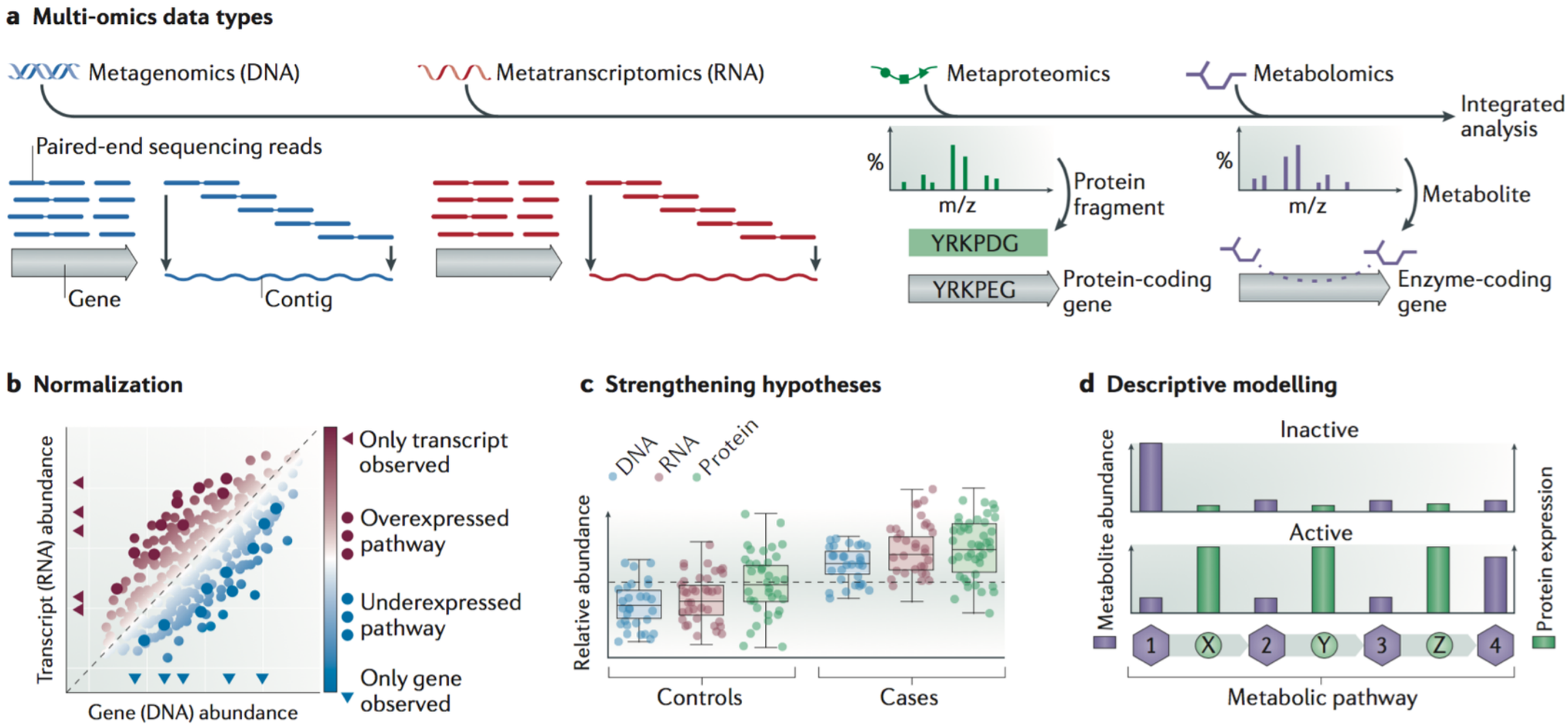
- Capturar todo el complemento de genes y genomas de una muestra
- Entender su estructura, diversidad, y distribución para inferir procesos biológicos
- RNA, DNA, marcadores



Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. Microbial Informatics and Experimentation. 2012;2(1):3. doi:10.1186/2042-5783-2-3.

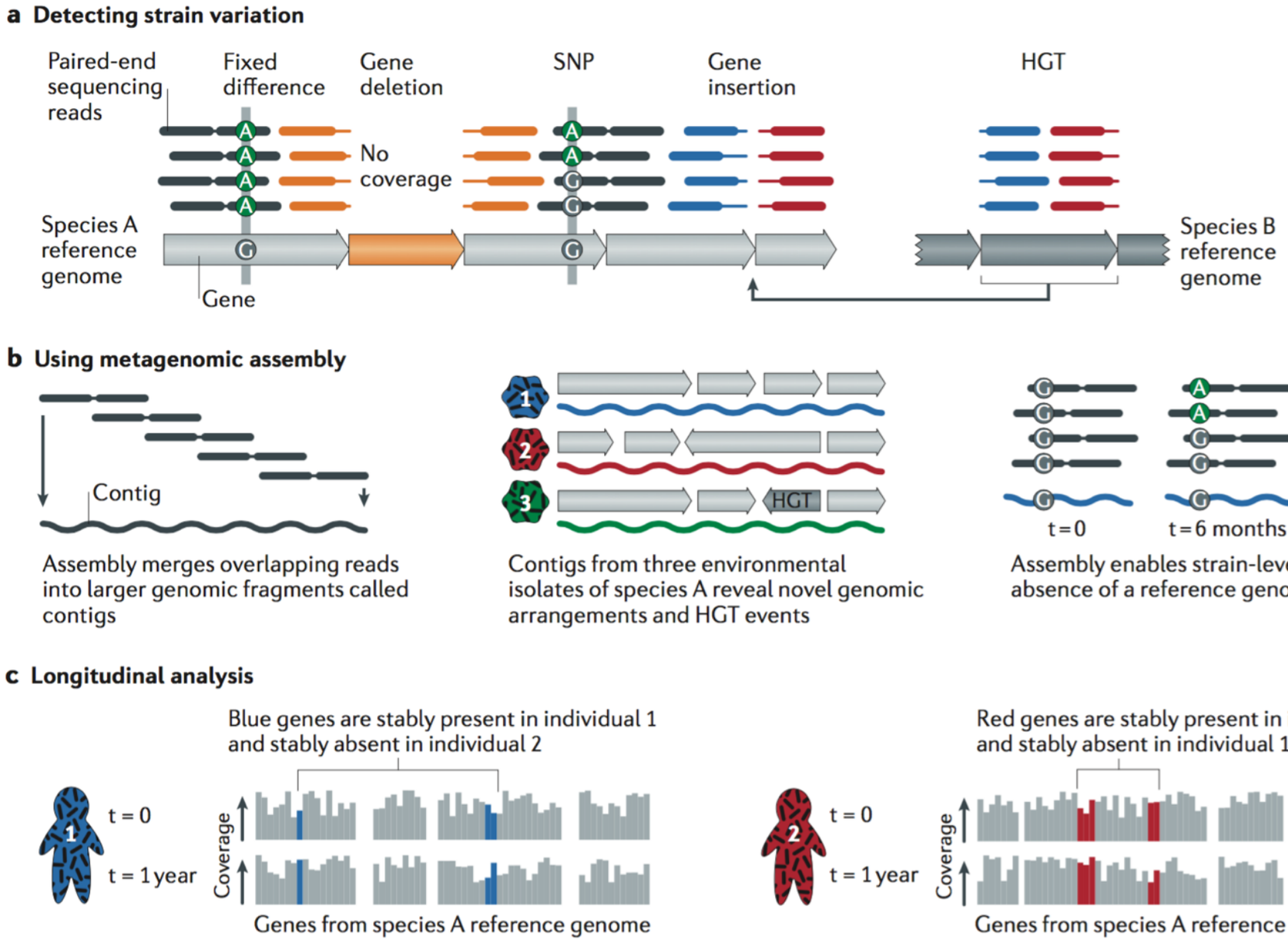
Metagenómas

- Integrar ómicas para pintar una imagen completa, desde genotipo hasta ruta metabólica



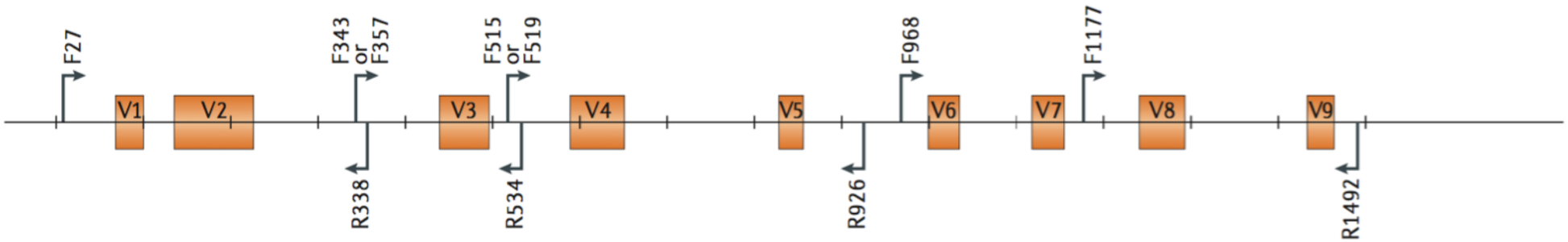
Franzosa EA, Hsu T, Sirota-Madi A, et al. Sequencing and beyond: integrating molecular “omics” for microbial community profiling. Nat Rev Micro. 2015;13(6):360-372. doi:10.1038/nrmicro3451.

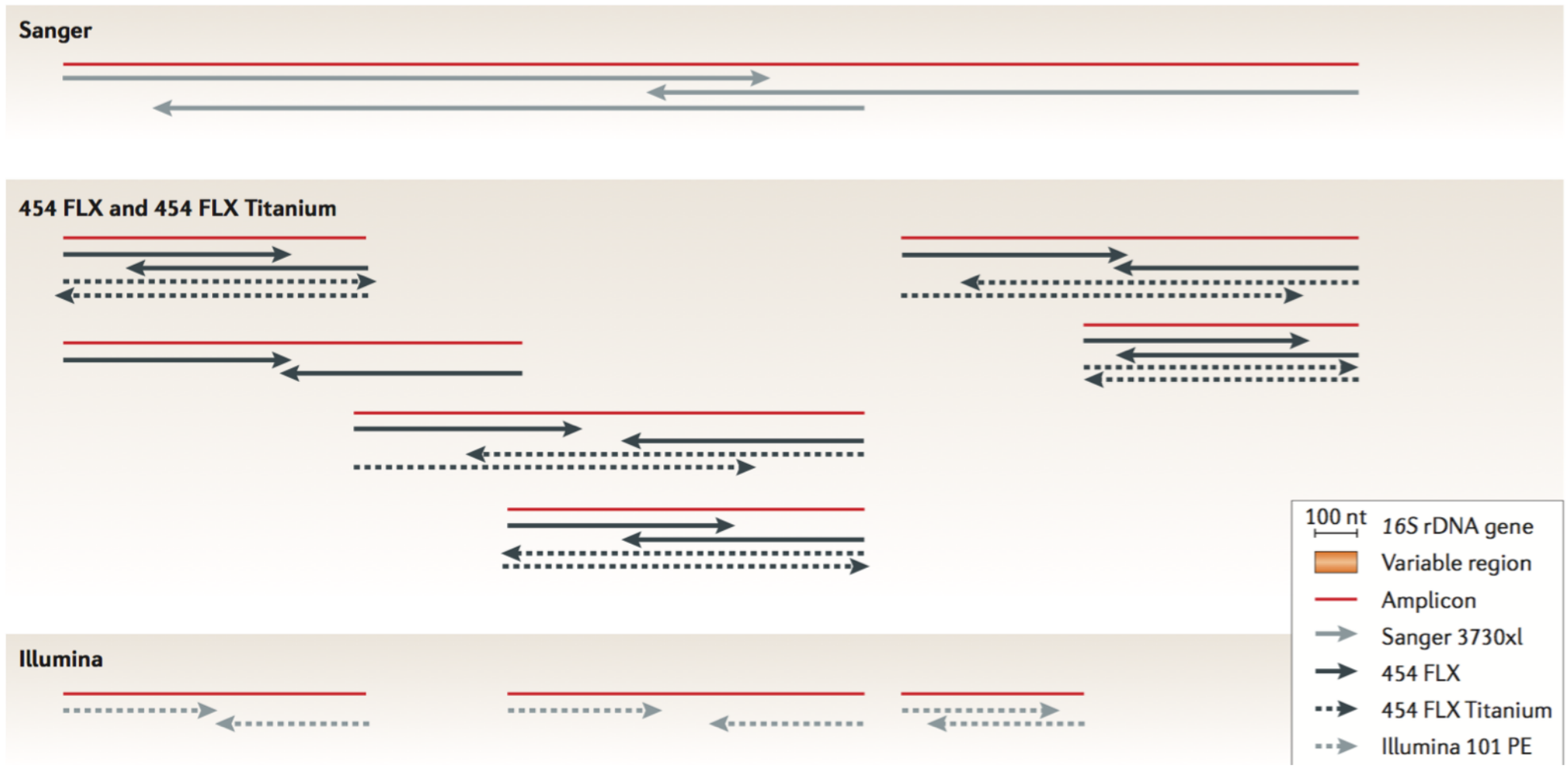
Metagenómas



Franzosa EA, Hsu T, Sirota-Madi A, et al. Sequencing and beyond: integrating molecular “omics” for microbial community profiling. Nat Rev Micro. 2015;13(6):360-372. doi:10.1038/nrmicro3451.

Metataxonomia con 16S rRNA





Kuczynski J, Lauber CL, Walters WA, et al. Experimental and analytical tools for studying the human microbiome. Nat Rev Genet. 2011;13(1):47-58. doi:10.1038/nrg3129.

Buen lugar para comenzar

The human microbiome: at the interface of health and disease

Ilseung Cho[1,2] and Martin J. Blaser[1,2,3,4]

Microbial Metagenomics: Beyond the Genome

Jack A. Gilbert[1,2,3] and Christopher L. Dupont[4]

[1]Plymouth Marine Laboratory, Plymouth PL1 3DH, United Kingdom

[2]Argonne National Laboratory, Argonne, Illinois 60439

[3]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637; email: gilbertjack@gmail.com

[4]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, California 92121; email: cdupont@jcvi.org

Marchesi and Ravel *Microbiome* (2015) 3:31
DOI 10.1186/s40168-015-0094-5

Microbiome

EDITORIAL Open Access

The vocabulary of microbiome research: a proposal

Julian R. Marchesi[1,2] and Jacques Ravel[3,4*]

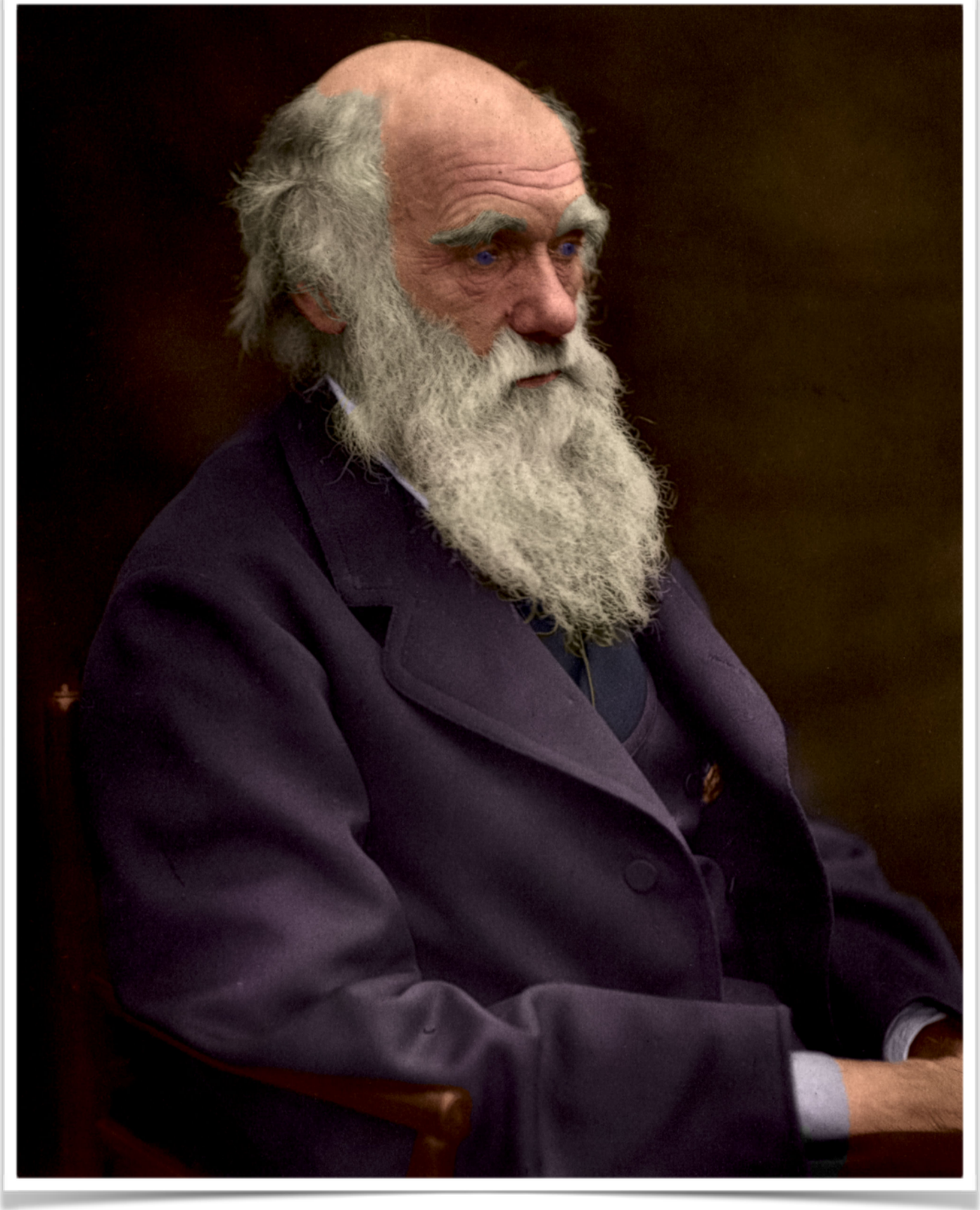
Conducting a Microbiome Study

Julia K. Goodrich,[1,2] Sara C. Di Rienzi,[1,2] Angela C. Poole,[1,2] Omry Koren,[1,2,9] William A. Walters,[3] J. Gregory Caporaso,[4,5] Rob Knight,[6,7,8] and Ruth E. Ley[1,2,*]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[2]Department of Microbiology, Cornell University, Ithaca, NY 14853, USA
[3]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA
[4]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA
[5]Institute for Genomics and Systems Biology, Argonne National Laboratory, Argonne, IL 60439, USA
[6]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
[7]BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA
[8]Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA
[9]Present address: Faculty of Medicine, Bar Ilan University, Ramat Gan 52900, Israel
*Correspondence: rel222@cornell.edu
http://dx.doi.org/10.1016/j.cell.2014.06.037

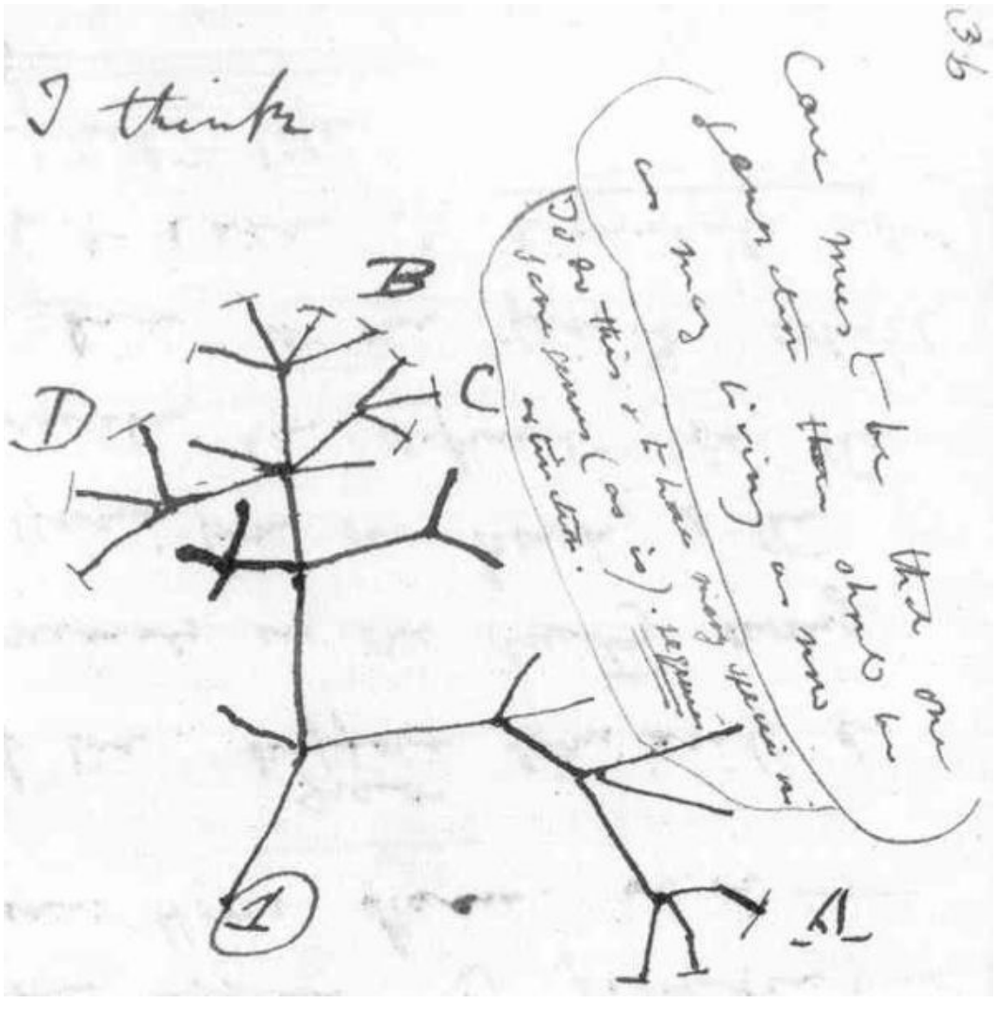
Experimental and analytical tools for studying the human microbiome

Justin Kuczynski[1], Christian L. Lauber[2], William A. Walters[1], Laura Wegener Parfrey[3], José C. Clemente[3], Dirk Gevers[4] and Rob Knight[3,5]



…from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved

The origin of species, 1859



de un cuaderno de Darwin de 1837