



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

EL PERÚ PRIMERO

Análisis Filogenéticos con Secuencias Genéticas Conservadas

Biólogo Rubén Ferro
Asistente de Investigación de la SDIEE

Lima, Setiembre 2020



PERÚ

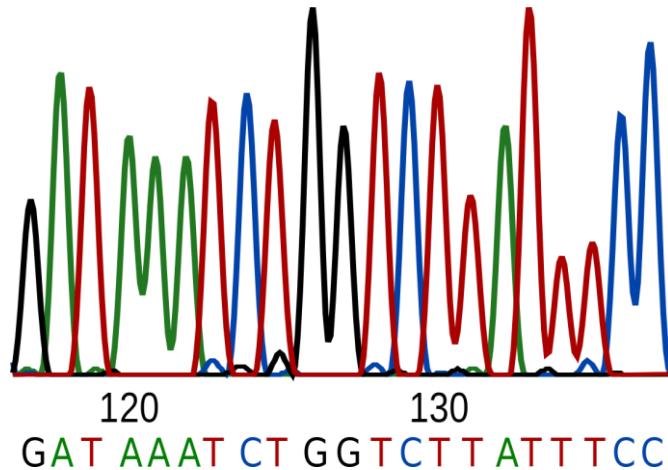
Ministerio
de Agricultura y Riego

OBJETIVOS



Instituto Nacional de Innovación Agraria

- Introducir la teoría práctica de la inferencia filogenética a partir de datos moleculares
- Introducir al uso de programas bioinformáticos para análisis de secuencias de ADN





PERÚ

Ministerio
de Agricultura y Riego



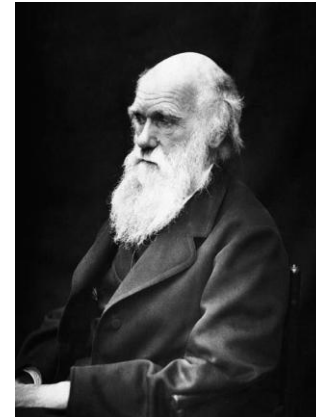
Instituto Nacional de Innovación Agraria

INTRODUCCIÓN

DARWIN

- “Estoy seguro de que llegará el tiempo, aunque yo no viva para verlo, en el que tendremos árboles genealógicos para cada
- Reino de la Naturaleza”

Carta de Darwin a Thomas Huxley, 1857





PERÚ

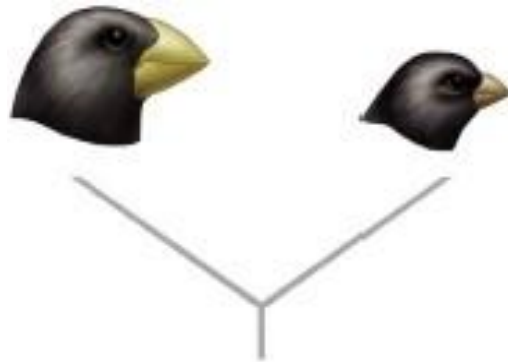
Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

CONCEPTOS EN FILOGENIA

- La **filogenética** es el estudio de las relaciones evolutivas entre grupos de organismos o genes
- Las relaciones evolutivas se representan en forma de árboles filogenéticos
- La filogenética puede utilizar caracteres morfológicos y/o moleculares (en la mayoría de los casos)





PERÚ

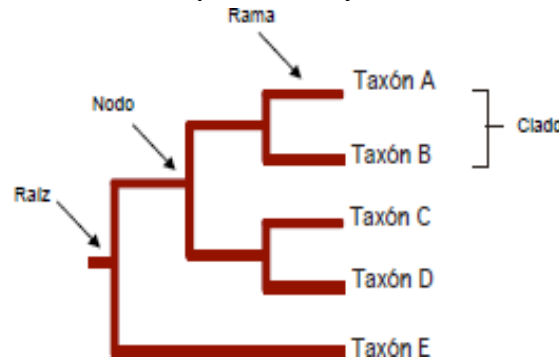
Ministerio
de Agricultura y Riego

CONCEPTOS EN FILOGENIA



Instituto Nacional de Innovación Agraria

- **Taxón:** Elementos cuyas relaciones estamos estudiando, A veces se llaman OTUS. Pueden tratarse de especies, grupos de especies, genes, alelos.
- **Nodo:** Punto de ramificación de una rama (presumiblemente un taxón ancestral)
- **Rama:** Define las relaciones entre taxones en términos de ancestros descendientes.
- **Topología:** Padrón de ramificación del árbol
- **Longitudes de rama:** Representa número de cambios o probabilidad de cambios
- **Raíz:** Ancestro común de todos los taxones
- **Clado:** Grupo de uno o más taxones que incluye al ancestro común y a todos sus descendientes





PERÚ

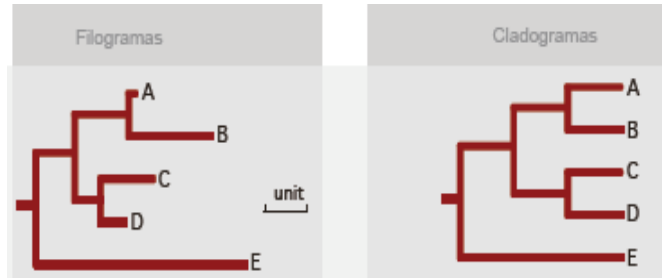
Ministerio
de Agricultura y Riego

CONCEPTOS EN FILOGENIA



Instituto Nacional de Innovación Agraria

- **Dendrograma:** Es un término genérico para la representación diagramática de un árbol filogenético.
- **Fenograma:** Es un dendrograma no enraizado en que se establecen las relaciones de parentesco fenético de los organismos estudiados; surgen de la aplicación de los métodos de taxonomía numérica.
- **Cladograma:** Es un árbol formado usando métodos cladísticos. Este tipo de árbol solo representa un padrón de ramificación, es decir que la longitud de sus ramas no representa el tiempo
- **Filograma:** Es un árbol filogenético que representa explícitamente un número de cambios de rasgos de carácter a lo largo de la longitud de sus ramas, es el resultado de la aplicación de los principios de la sistemática evolutiva.





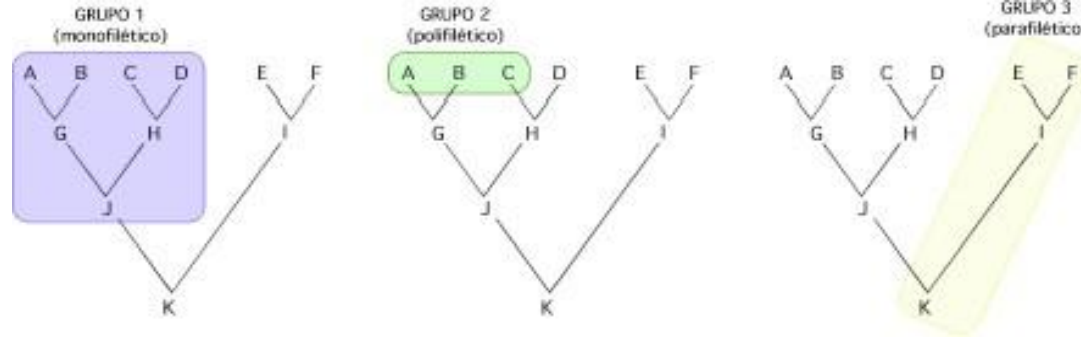
PERÚ

Ministerio
de Agricultura y Riego

CONCEPTOS EN FILOGENIA



Instituto Nacional de Innovación Agraria



- **Monofilético:** Incluye todos los descendientes del taxón ancestral de los miembros del grupo.
- **Polifilético:** No incluye a todos los descendientes del taxón ancestral de los miembros del grupo.
- **Parafilético:** No incluye a todos los descendientes del taxón ancestral de los miembros del grupo, pero sí al mismo ancestro.



PERÚ

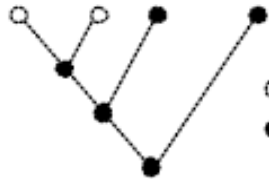
Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

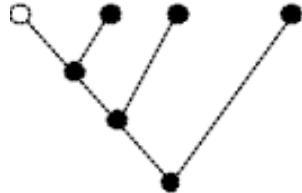
CONCEPTOS EN FILOGENIA

- Terminología relacionada con la reconstrucción de la **historia de cambios en estados de carácter**



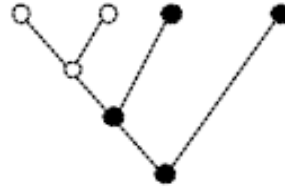
○ **apomorfía** : carácter derivado; estado apomórfico

● **plesiomorfía**: caract. ancestral; estado plesiomórfico o ancestral



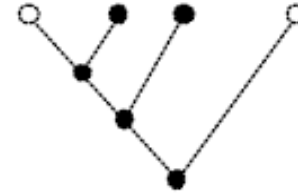
autapomorfía

- carácter derivado único (aut)



sinapomorfía

- carácter derivado compartido (syn)



homoplasia

- carácter compartido no homólogo, es decir, no heredado directamente del ancestro



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

HOMOLOGÍA Y HOMOPLASÍA

- **Homología:** Es la relación entre rasgos de organismos que se comparten como resultado de una herencia común (ex. huesos de los extremidades de los vertebrados).
- **Analogía:** Es la relación entre rasgos de organismos que desempeñan una función similar pero no como resultado de una herencia común (ex. el ala de los insectos y las aves).
- **Homoplasia:** Es la relación entre rasgos similares de organismos que se comparten como resultado de evolución revertida, paralela o convergente (ex. espinas en diversos cactus).



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

HOMOLOGÍA ESTRUCTURAL





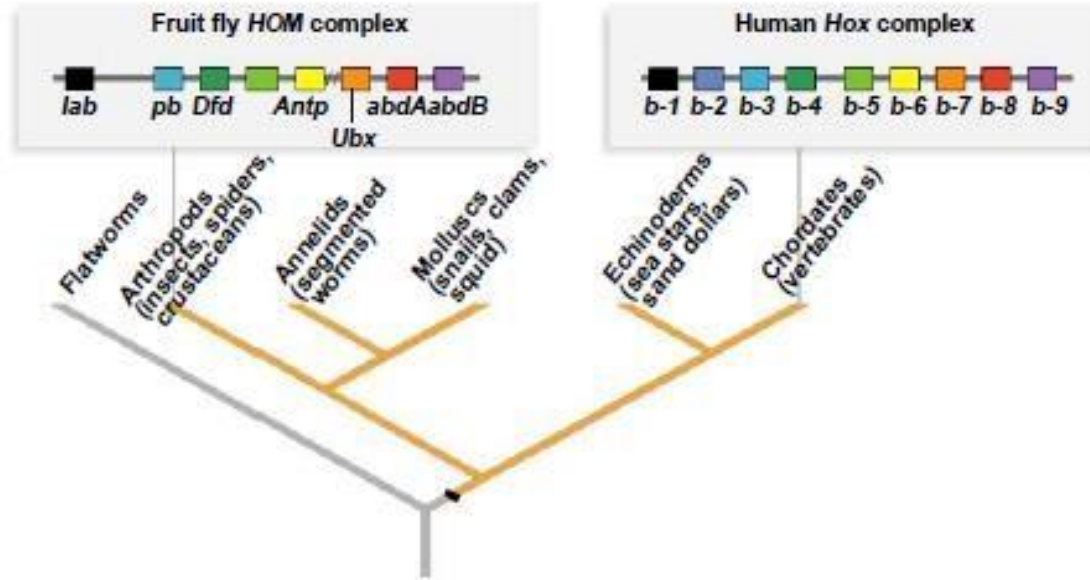
PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

HOMOLOGÍA GENÉTICA



Gene: Aniridia (human)
 eyeless (Fruit fly)

LQRNRTSFT**QEQIE**ALEKEFERTHYPDVFAERERLA**AK**IDLPEARIQVWFSNRRRAKWRREE
 LQRNRTSFT**NDQID**SLEKEFERTHYPDVFAERERLA**GK**IGLPEARIQVWFSNRRRAKWRREE



PERÚ

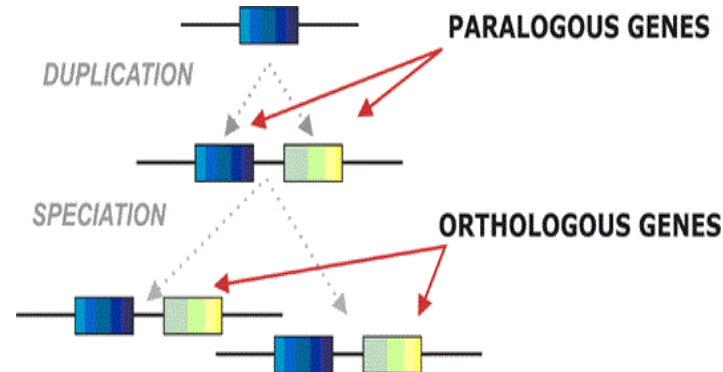
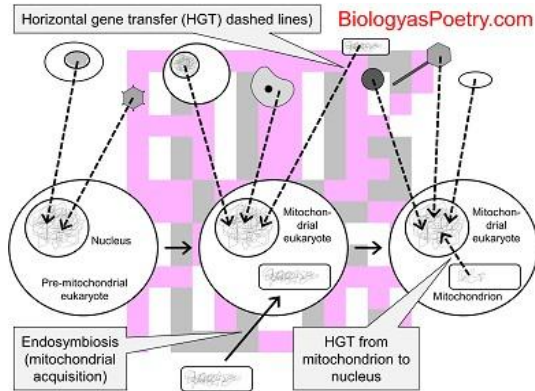
Ministerio
de Agricultura y Riego

TIPOS DE HOMOLOGÍA



Instituto Nacional de Innovación Agraria

- **ORTÓLOGOS:** Homología producida por la especiación. Los ortólogos tienden a tener funciones similares
- **PARÓLOGOS:** Homología producida por duplicación de genes. Los parálogos tienden a tener funciones distintas.
- **XENÓLOGOS:** Homología producida a partir de transferencia genética horizontal entre dos organismos.





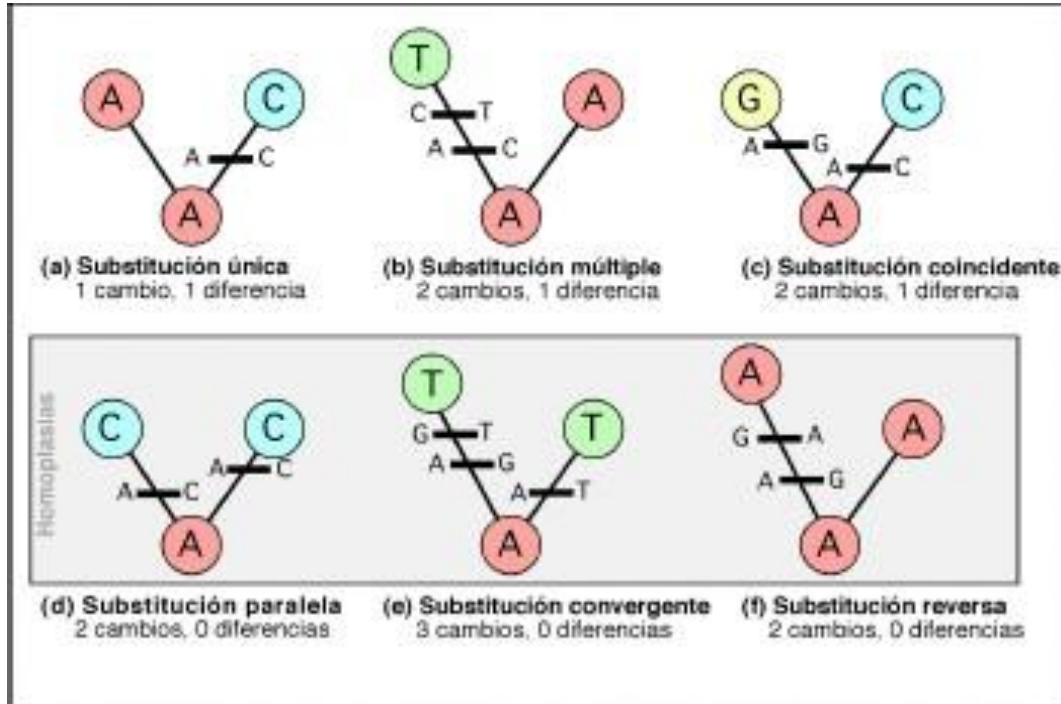
PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

HOMOPLASÍA MOLECULAR





PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

PORQUÉ ESTUDIAR FILOGENÉTICA Y EVOLUCIÓN MOLECULAR?

COROLARIO I

Nada en biología hace sentido excepto a la luz de la evolución

Theodosius Dobzhanski, 1973

The American Biology Teacher 35:125

COROLARIO II

Nada en biología evolucionaria hace sentido excepto a la luz de la filogenia

Jeff Palmer, Douglas Soltis, Mark Chase, 2004 American J. Botany
91:1437-1445



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

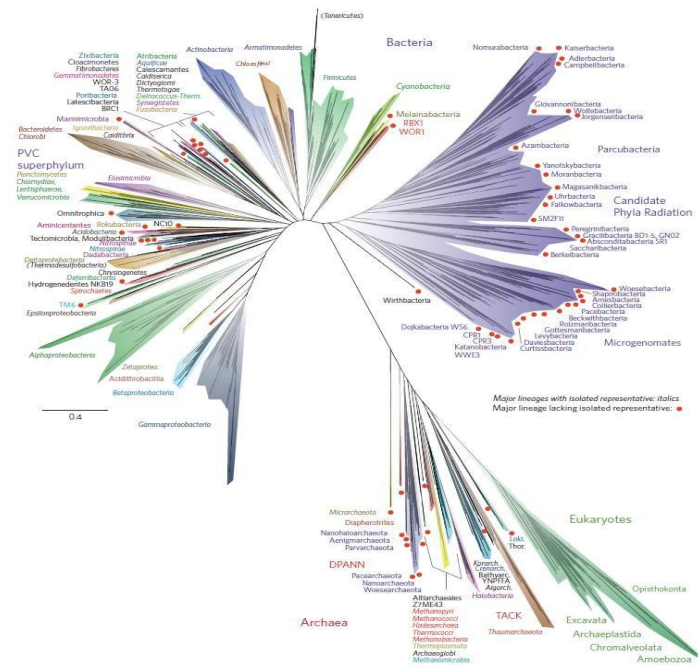
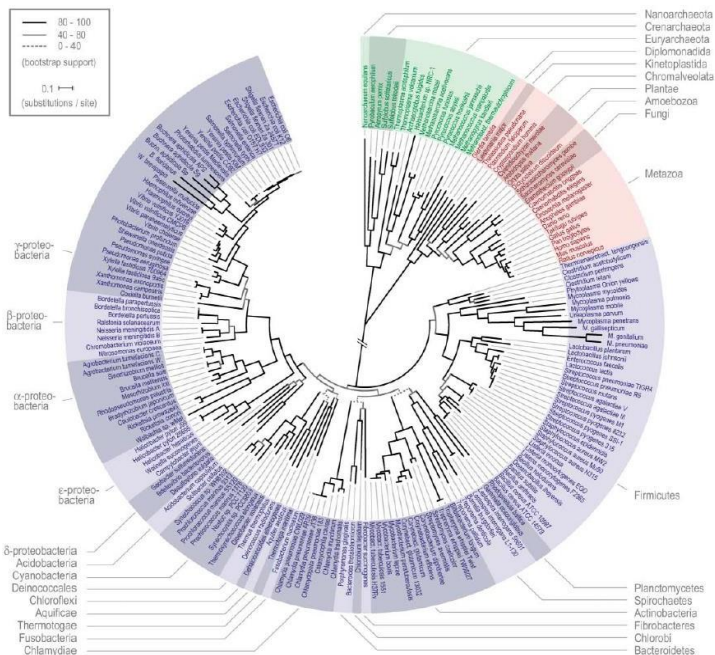
APLICACIONES DE LA FILOGENÉTICA

- Biología evolutiva (ex. Reconstrucción del árbol de la vida)
- Sistemática (ex. Clasificación de grupos)
- Antropológico (ex. Estudios de poblaciones humanas)
- Forense (ex. Transmisión intencional de un virus)
- Parasitológicos (ex. Coevolución de parásitos y hospedadores)
- Epidemiológico (ex. Reconstrucción de la transmisión de una enfermedad)
- Ecológicos (ex. Radiaciones adaptativas)
- Genómica y Proteómica (ex. Comparación con proteínas homólogas)



APLICACIONES DE LA FILOGENÉTICA

Topología filogenética basada en 31 genes ortólogos de 191 genomas



Topología filogenética basada en 2,596 amino ácidos de 3,083 genomas



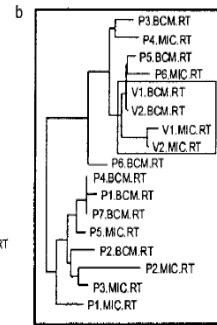
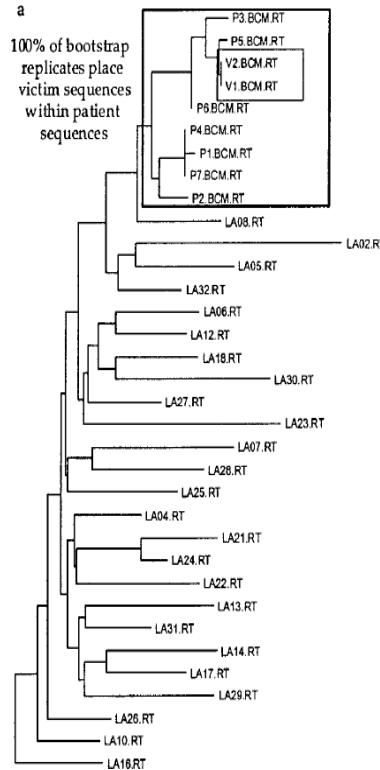
PERÚ

Ministerio
de Agricultura y Riego

APLICACIONES DE LA FILOGENÉTICA



Instituto Nacional de Innovación Agraria



- Usando el gen de la retrotranscriptasa inversa fue posible condenar un médico por crimen en segundo grado de su enamorada, al inocularle sangre positiva con HIV de una de sus pacientes.
- El análisis de las secuencias fueron realizadas por dos instituciones diferentes coincidiendo las muestras de HIV de la enamorada con de la paciente.

Molecular evidence of HIV-1 transmission in a criminal case

Michael L. Metzker^{*†}, David P. Mindell[‡], Xiao-Mei Liu^{*§}, Roger G. Ptak^{||}, Richard A. Gibbs^{*}, and David M. Hillis^{**}

^{*}Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; [‡]Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079; [§]School of Dentistry, Biologic and Materials Sciences, University of Michigan, Ann Arbor, MI 48109; and ^{**}Section of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas, Austin, TX 78712



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

PORQUÉ QUEREMOS ALINEAR SECUENCIAS?

- Asignar funciones a proteínas desconocidas
- Determinar relaciones entre organismos
- Identificar estructuralmente y funcionalmente elementos importantes
- Hacer predicciones acerca de estructuras 3D



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS DE SECUENCIAS

Objetivo general

- Encontrar el máximo grado de semejanza
- Encontrar la mínima distancia evolucionaria



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

TIPOS DE ALINEAMIENTOS DE SECUENCIAS

- Alineamientos global vs local
- Alineamientos espaciados vs no espaciados
- Alineamientos por pares vs multiples



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTO GLOBAL VS LOCAL

Alineamiento Global

- Empareja tanto caracteres como fuere posible de principio a fin
- Encuentra un alineamiento con el más alto score
- Regiones con alta similaridad local pueden ser ignorados en favor de un más alto score total

Example:

```
THIS-ISAGLALALIGNMENT
||  ||  ||  |||  |
THEREISTHEAL--IGN-EDSEQ
```



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTO GLOBAL VS LOCAL

Alineamiento Local

- Encuentra subsecuencias con la más alta densidad de emparejamientos
- Encuentra regiones con el más alto local score
- Secuencias similares pueden extenderse más allá de la local subsecuencia con un más bajo grado de similaridad.

Example:

```
-----LALIGNM-----  
          |||||  
-----EALIGNE-----
```



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS ESPACIADOS VS NO ESPACIADOS

Alineamiento No espaciados

- Comparaciones de secuencias son más o menos proporcional al cuadrado de las longitudes promedios.

MATCHES

||

MAKERS



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS ESPACIADOS VS NO ESPACIADOS

Alineamiento Espaciado

- Si espacios de alguna longitud en alguna posición sería permitido:
- * Computacionalmente muy costoso
- * Alineamientos no serían significativos

MATCHE-S

| | | |

MA--KERS

- Se necesita un número manejable de espacios



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS ESPACIADOS VS NO ESPACIADOS

Penalidades de espacios

- Reduce el número de espacios en el alineamiento
- Asegura un alineamiento más significativo
- Abertura de un espacio es costosa
- Extensión de un espacio es barata

Examples:

Gap opening penalty = - 12

Gap extension penalty = - 1



PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS ESPACIADOS VS NO ESPACIADOS

Penalidades de espacios

Sequence alignment problem

Gap open = -12, Gap extension = -1

- | | | |
|----|------------------------------|------------------------|
| 1. | THESESENTENSESALIGN--NICELY | -12 +(-)1 |
| | | -28 |
| 2. | THESEQUENCE----ALIGNEDNICELY | -12 +(-3) |
| 1. | THESESENTENSESALIGN--NICELY | -12 +(-1) |
| | | -50 |
| 2. | THESE-Q--ENCE-ALIGNEDNICLEY | -12 +(-12)+(-1)+(-12) |
| 1. | THESESENTENSESALIGN--NICELY | -12 +(-1) |
| | | -50 |
| 2. | THE--SEQ-ENCE-ALIGNEDNICLEY | -12 +(-1)+(-12) +(-12) |



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS ENPAREJADOS VS MULTIPLES

- **Alineamiento enparejado**
 - * Necesita dos secuencias
- **Alineamiento múltiple**
 - * Requiere más de dos secuencias
 - * Es un problema computacional es más difícil



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MÉTODOS DE ALINEAMIENTOS ENPAREJADOS

- Dot Matrix (Gibbs and McIntyre)
- Algoritmos de programación dinámica (Needleman- Wunsch, Smith-Waterman)
- Algoritmos heurísticos (BLAST, FASTA)



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Ventajas

- Todos los posibles emparejamientos entre 2 secuencias son mostrados.
- Facilmente revela inserciones y deleciones
- Facilmente identifica repeticiones invertidas directas
- Mismo algoritmo es usado para ADN, ARN y proteínas.

Desventajas

- No muestra un actual alineamiento de secuencia.
- Evaluación cualitativa de los alineamientos
- Significancia estadística de alineamiento no es obvia.



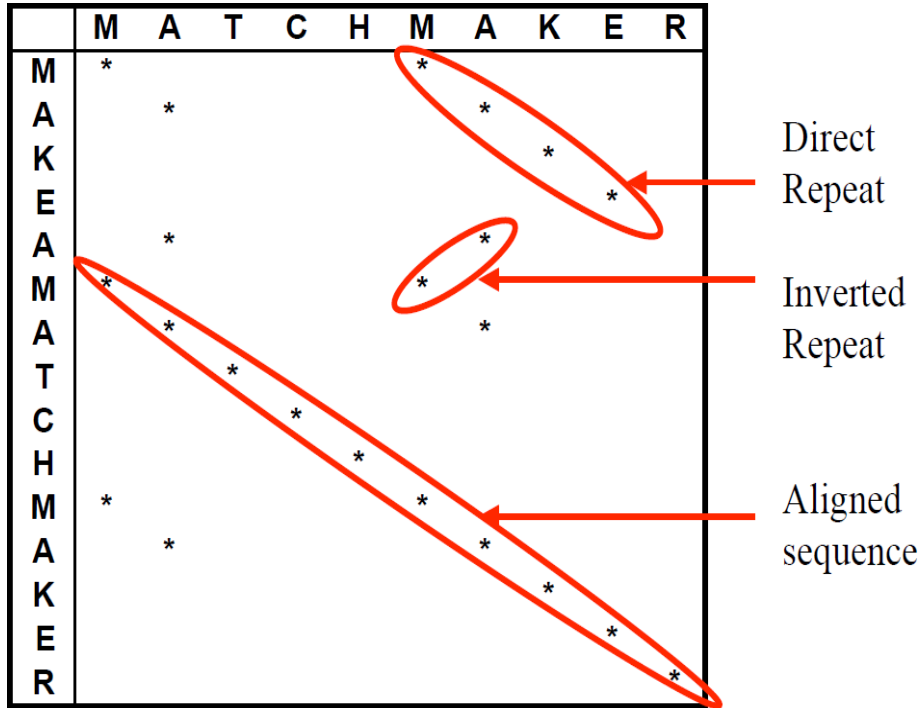
PERÚ

Ministerio de Agricultura y Riego

DOT MATRIX



Instituto Nacional de Innovación Agraria





PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Ventajas

- Todos los posibles emparejamientos entre 2 secuencias son mostrados.
- Facilmente revela inserciones y deleciones
- Facilmente identifica repeticiones invertidas directas
- Mismo algoritmo es usado para ADN, ARN y proteínas.

- **Desventajas**

- No muestra un actual alineamiento de secuencia.
- Evaluación cualitativa de los alineamientos
- Significancia estadística de alineamiento no es obvia.



PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALGORITMOS DE PROGRAMACIÓN DINÁMICA

GLOBAL

Seq. T		<i>j</i>	<i>j</i> +1	<i>n</i>	
		M	A	T	C	H	E	S	
Seq. S		0	-2	-4	-6	-8	-10	-12	-14
<i>i</i>	T	-2	-1	-3	-3	-5	-7	-9	-11
<i>i</i> +1	H	-4	-3	-2	-4	-4	-4	-6	-8
...	A	-6	-5	-2	-3	-5	-5	-5	-7
...	T	-8	-7	-4	-1	-3	-5	-6	-6
...	C	-10	-9	-6	-3	0	-2	-4	-6
...	H	-12	-11	-8	-5	-2	1	-1	-3
...	E	-14	-13	-10	-7	-4	-1	2	0
<i>m</i>	R	-16	-15	-12	-9	-6	-3	0	1

Highest score in last row and last column

LOCAL

Seq. T		<i>j</i>	<i>j</i> +1	<i>n</i>
		M	A	T	C	H	E	S
Seq. S		0	0	0	0	0	0	0
<i>i</i>	T	0	0	0	5	0	0	2
<i>i</i> +1	H	0	0	0	0	2	10	2
...	A	0	0	5	0	0	2	9
...	T	0	0	0	10	2	0	9
...	C	0	0	0	2	23	15	7
...	H	0	0	0	0	15	33	25
...	E	0	0	0	0	7	25	39
<i>m</i>	R	0	0	0	0	0	17	31

Highest score anywhere



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALGORITMOS HEURÍSTICOS (BLAST, FASTA)

Ventajas

- Muy rápido
- Confiable en un sentido estadístico

Desventajas

- Menos sensitivo que la programación dinámica



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS DE MÚLTIPLES SECUENCIAS

Ventajas

- Puede identificar dominios estructurales y funcionales.
- Puede identificar familias de proteínas.

Desventajas

- Aún un difícil problema de algoritmos



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS DE SECUENCIAS

TABLE 1: Types of multiple sequence alignment and corresponding algorithms.

Types of MSA alignment	MSA algorithms
Pairwise alignment	Needleman-Wunsch, k-mer, k-tuple, and Smith-Waterman algorithms.
Progressive alignment	Clustal Omega, ClustalW, MAFFT, Kalign, Probalign, MUSCLE, Dialign, ProbCons, and MSAProbs.
Iterative progressive alignment	PRRP, MUSCLE, DIALIGN, SAGA, and T-COFFEE.
Homology search tools	BLAST, PSI-BLAST, and FASTA.
Structure incorporating alignment	3D-COFFEE, EXPRESSO, and MICAlign.
Motif alignment	PHI-Blast, GLAM2.
Short-read alignment	Bowtie, Maq, and SOAP.



PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ALINEAMIENTOS DE SECUENCIAS





PERÚ

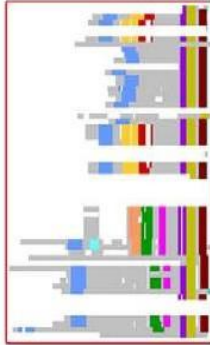
Ministerio
de Agricultura y Riego

ALINEAMIENTOS DE SECUENCIAS

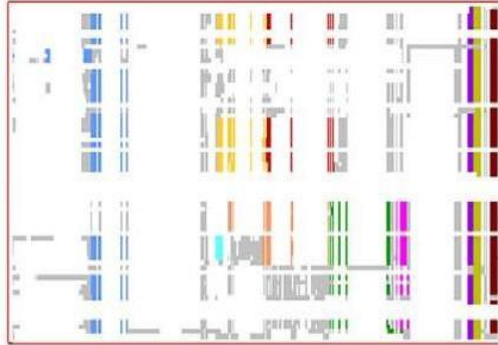


Instituto Nacional de Innovación Agraria

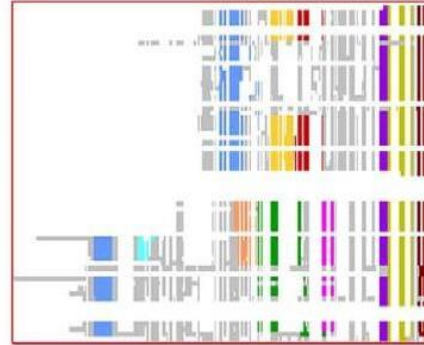
clustalw



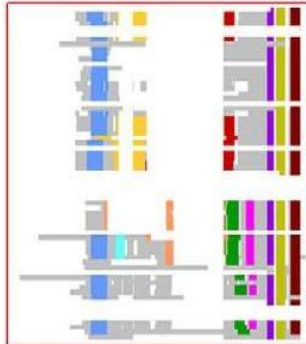
dialign-tx



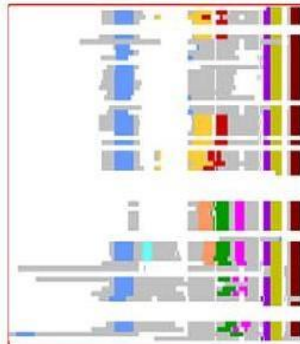
kalign



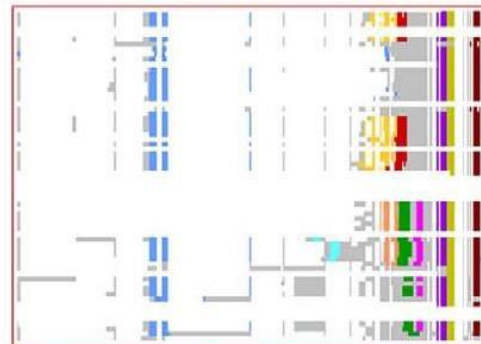
mafft(linsi)



muscle



probcons





PERÚ

Ministerio
de Agricultura y Riego

BASE DE DATOS



Instituto Nacional de Innovación Agraria

- Base de datos es un conjunto de datos relacionados y almacenados para un propósito particular con una estructura lógica.
- Una base de datos de interés biológico, nos permitirá realizar búsqueda de información (autores, títulos, claves, números de acceso), homologías (secuencias iguales o parecidas), patrones (patrones conocidos en mi secuencia) y predicciones (encontrar proteínas parecidas a la mía pero con función conocida).



PERÚ

Ministerio
de Agricultura y Riego

BASE DE DATOS

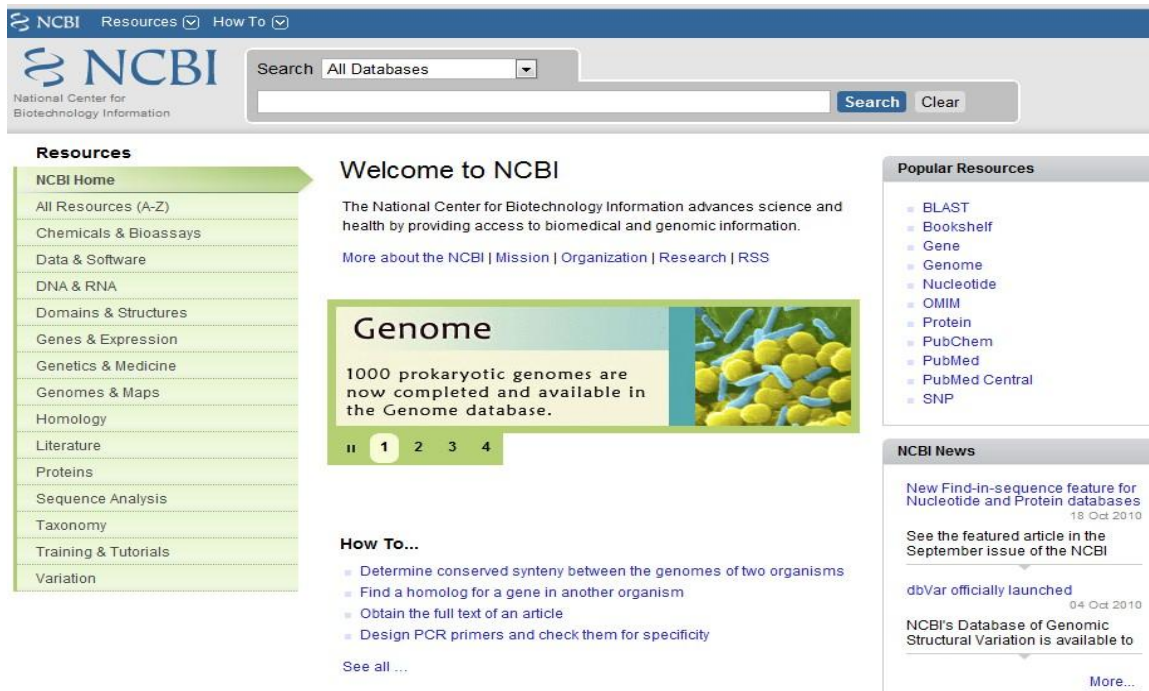


Instituto Nacional de Innovación Agraria

- Existen cientos de BD en número tan elevado que no es práctico enumerarlas.
- Por el tipo de información que contienen distinguimos Bases de datos [bibliográficas](#) Bases de datos [taxonómicas](#) Bases de datos de [nucleótidos](#) Bases de datos [genómicas](#) Bases de datos de [proteínas](#) Bases de datos de [microarrays](#)

BASE DE DATOS

El [National Center for Biotechnology Information \(NCBI\)](http://www.ncbi.nlm.nih.gov) centraliza los bancos de datos y aplicaciones de EEUU.



The screenshot shows the NCBI website interface. At the top, there is a navigation bar with "NCBI Resources" and "How To" dropdown menus. Below this is the NCBI logo and a search bar with a dropdown menu set to "All Databases" and "Search" and "Clear" buttons. The main content area is divided into several sections:

- Resources:** A vertical list of links including "NCBI Home", "All Resources (A-Z)", "Chemicals & Bioassays", "Data & Software", "DNA & RNA", "Domains & Structures", "Genes & Expression", "Genetics & Medicine", "Genomes & Maps", "Homology", "Literature", "Proteins", "Sequence Analysis", "Taxonomy", "Training & Tutorials", and "Variation".
- Welcome to NCBI:** A central text block stating "The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information." and "More about the NCBI | Mission | Organization | Research | RSS".
- Genome:** A featured article with the headline "1000 prokaryotic genomes are now completed and available in the Genome database." and a small image of yellow and blue bacteria.
- How To...:** A list of tasks: "Determine conserved synteny between the genomes of two organisms", "Find a homolog for a gene in another organism", "Obtain the full text of an article", and "Design PCR primers and check them for specificity".
- Popular Resources:** A list of links: "BLAST", "Bookshelf", "Gene", "Genome", "Nucleotide", "OMIM", "Protein", "PubChem", "PubMed", "PubMed Central", and "SNP".
- NCBI News:** A list of news items: "New Find-in-sequence feature for Nucleotide and Protein databases" (18 Oct 2010), "See the featured article in the September issue of the NCBI", "dbVar officially launched" (04 Oct 2010), and "NCBI's Database of Genomic Structural Variation is available to".



PERÚ

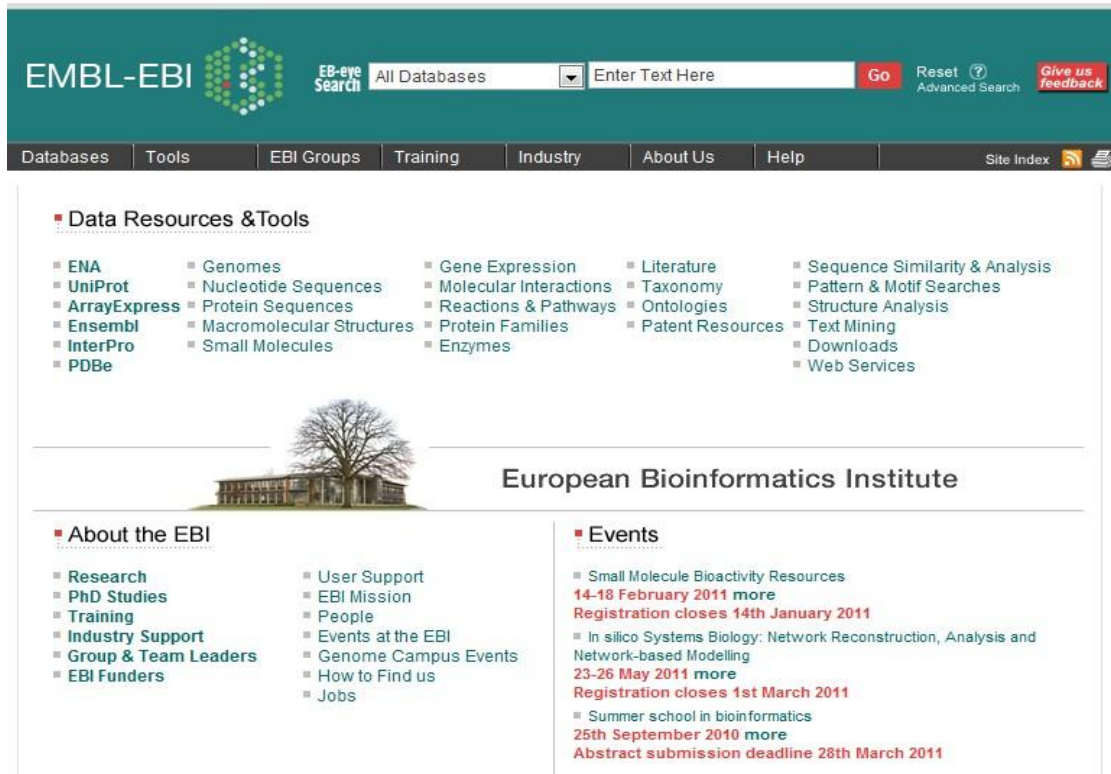
Ministerio de Agricultura y Riego

BASE DE DATOS



Instituto Nacional de Innovación Agraria

El [European Bioinformatics Institute \(EBI\)](#) realiza una función similar en Europa.



The screenshot shows the EMBL-EBI website interface. At the top, there is a search bar with the text "EB-eye Search" and "All Databases" in a dropdown menu. The search input field contains "Enter Text Here". To the right of the search bar are buttons for "Go", "Reset", "Advanced Search", and "Give us feedback". Below the search bar is a navigation menu with links for "Databases", "Tools", "EBI Groups", "Training", "Industry", "About Us", "Help", and "Site Index".

The main content area is titled "Data Resources & Tools" and contains a grid of links:

- ENA
- UniProt
- ArrayExpress
- Ensembl
- InterPro
- PDBe
- Genomes
- Nucleotide Sequences
- Protein Sequences
- Macromolecular Structures
- Small Molecules
- Gene Expression
- Molecular Interactions
- Reactions & Pathways
- Protein Families
- Enzymes
- Literature
- Taxonomy
- Ontologies
- Patent Resources
- Sequence Similarity & Analysis
- Pattern & Motif Searches
- Structure Analysis
- Text Mining
- Downloads
- Web Services

Below the grid is a photograph of a large, modern building with a tree in front of it. To the right of the photograph is the text "European Bioinformatics Institute".

Below the photograph and text are two columns of links:

- About the EBI**
 - Research
 - PhD Studies
 - Training
 - Industry Support
 - Group & Team Leaders
 - EBI Funders
 - User Support
 - EBI Mission
 - People
 - Events at the EBI
 - Genome Campus Events
 - How to Find us
 - Jobs
- Events**
 - Small Molecule Bioactivity Resources
14-18 February 2011 [more](#)
Registration closes 14th January 2011
 - In silico Systems Biology: Network Reconstruction, Analysis and Network-based Modelling
23-26 May 2011 [more](#)
Registration closes 1st March 2011
 - Summer school in bioinformatics
25th September 2010 [more](#)
Abstract submission deadline 28th March 2011



PERÚ

Ministerio de Agricultura y Riego

BASE DE DATOS



Instituto Nacional de Innovación Agraria

GenomeNet reúne bases de datos diversas en Japón.

GenomeNet
KEGG KEGG2 PATHWAY BRITE DISEASE DRUG DBGET

[\[English | Japanese \]](#)

Search for

GenomeNet
[About GenomeNet](#)
[Announcements](#)
[Release notes](#)
[Acknowledgments](#)

DBGET
[Overview](#)
[DB release info](#)
[DB growth curve](#)

KEGG

varDB

Community DBs

Bioinformatics tools
[Other tools](#)

Feedback

GenomeNet Database Resources

DBGET: Integrated Database Retrieval System
[DBGET search](#)
[LinkDB search](#)

KEGG: Kyoto Encyclopedia of Genes and Genomes
[KEGG2 - Table of contents](#)
[KEGG PATHWAY - Systems information: pathways](#)
[KEGG BRITE - Systems information: ontologies](#)
[KEGG DISEASE - Disease information resource](#)
[KEGG DRUG - Drug information resource](#)
[KEGG Organisms - Organism-specific entry points](#)
[KEGG GENES - Genomic information](#)
[KEGG LIGAND - Chemical information](#)
[KEGG GLYCAN - Glycan information resource](#)

varDB: Antigenic variation database

Community Databases
[CYORF - Cyanobacteria annotation database](#)
[BSORF - Bacillus subtilis genome database](#)
[EXPRESSION - Gene expression profile database](#)

GenomeNet Bioinformatics Tools

Sequence Analysis
[BLAST / FASTA - Sequence similarity search](#)
[MOTIF - Sequence motif search](#)
[CLUSTALW / MAFFT / PRRN - Multiple alignment](#)

[KEGG release info](#)
[KEGG update notes](#)
[KEGG pathway maps](#)
[BRITE functional hierarchies](#)
[KEGG Organisms](#)
[\[Animals | Plants \]](#)

Global metabolism map



PERÚ

Ministerio
de Agricultura y Riego

BASE DE DATOS



Instituto Nacional de Innovación Agraria

Expasy es una base de datos multipropósitos de Suiza.

The screenshot shows the ExpASY Bioinformatics Resource Portal. At the top left, there is a logo for SIB (Swiss Institute of Bioinformatics) and the text "ExpASY Bioinformatics Resource Portal". Below this is a search bar with a dropdown menu set to "Query all databases" and a search button. On the left side, there is a navigation menu with categories like "Visual Guidance", "Categories", "Resources A..Z", and "Links/Documentation". The "Categories" section lists various biological fields such as proteomics, genomics, structure analysis, systems biology, evolutionary biology, population genetics, transcriptomics, biophysics, imaging, IT infrastructure, medicinal chemistry, and glycomics. The main content area features a "Featuring today" section with a highlighted article titled "Genome History" about "Gene duplicate identification" with a link to "[details]". At the bottom, there is a "How to use this portal?" section with a question mark icon and a list of links: "Features and updates", "New to ExpASY", and "Experienced ExpASY users: what is different".



PERÚ

Ministerio de Agricultura y Riego

BASE DE DATOS



Instituto Nacional de Innovación Agraria

[Protein Data Bank](#) es una base de datos estructuras 3D de proteínas

RCSB **PDB** PROTEIN DATA BANK

An Information Portal to 131205 Biological Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Go

Advanced Search | Browse by Annotations

PDB-101 | Worldwide PDB | EMDatabank | BIOLOGICAL ACID DATABASE | StructuralBiology Knowledgebase | Worldwide Protein Data Bank Foundation

f t y d

Welcome

- Deposit
- Search
- Visualize
- Analyze
- Download
- Learn

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Zika Illustration Named People's Choice

Vizzies VISUALIZATION CHALLENGE
People's Choice Award Winner
Category: Illustration
Zika Virus
David S. Goodsell

June Molecule of the Month

Adenine Riboswitch in Action



PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

DONDE BUSCAR CON MÁS ESPECIFICIDAD?



Virus Database at University College London
Virus Genomics and Bioinformatics Wolf Vison Centre
Biomolecular Structure and Modelling Group Department of Biochemistry



The GDB Human Genome Database



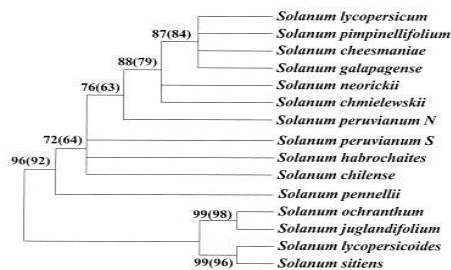
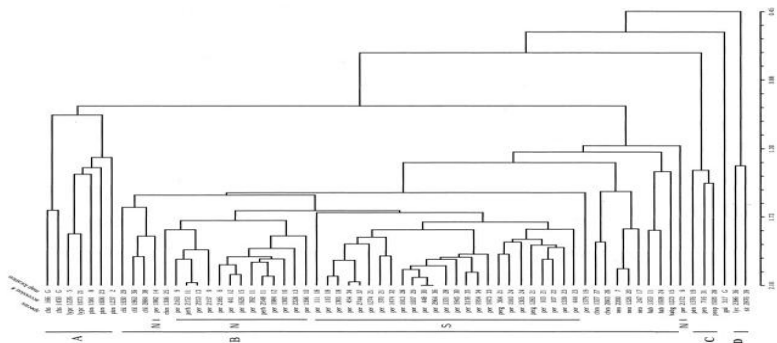
OTROS DATOS USADOS PARA ESTABLECER RELACIONES FILOGENÉTICAS

TABLE 4. Data matrix of cladistic character states corresponding to Table 3.

Species	Characters and Character States																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26		
OUTGROUPS																												
<i>Solanum lycopersicoïdes</i>	1	1	2	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	3	2	1	1	1	
<i>S. sitchensis</i>	?	1	2	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	3	2	1	1	1	
<i>S. ochranthum</i>	1	1	2	2	1	1	2	2	1	1	2	2	1	2	2	2	2	2	2	1	1	1	1	1	1	1	1	
<i>S. juglandifolium</i>	1	1	2	2	1	1	2	2	1	1	2	2	1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	
INGROUPS																												
<i>S. pennellii</i>	3	3	2	2	1	1	1	2	3	2	1	3	2	1	2	2	3	1	3	1	1	1	3	3	2	1	1	
<i>S. habrochaites</i>	2	2	2	1	1	2	3	3	2	1	1	1	1	1	2	3	3	2	1	2	1	3	3	2	1	1		
<i>S. chilense</i>	3	3	2	1	1	1	2	2	1	1	1	1	1	1	2	3	3	2	1	2	1	3	3	2	1	1		
<i>S. peruvianum south</i>	3	3	2	2	1	1	2	2	2	2	1	1	1	1	2	3	3	2	2	2	1	3	3	2	1	1		
<i>S. peruvianum north</i>	3	3	2	2	1	1	2	4	2	1	1	1	2	2	3	3	2	1	3	1	3	3	2	1	1			
<i>S. chmielewskii</i>	3	3	2	2	1	1	2	4	4	2	1	1	2	2	3	3	2	1	3	1	3	3	2	2	2			
<i>S. neorickii</i>	3	3	2	2	1	1	2	4	4	2	1	1	3	2	3	3	2	1	4	1	3	3	2	2	3			
<i>S. cheesmaniae</i>	4	2	1	2	2	1	2	4	4	1	1	1	1	2	2	3	3	2	1	4	2	3	3	2	2	3		
<i>S. galapagense</i>	3	2	1	2	4	2	2	4	4	1	1	1	1	2	2	3	3	2	1	4	2	4	3	2	2	3		
<i>S. pimpinellifolium</i>	3	2	1	2	2	1	2	4	4	1	1	1	1	2	2	3	3	2	1	3	3	3	2	2	2	2		
<i>S. lycopersicum</i>	3	2	1	2	3	1	2	4	4	1	1	1	1	2	2	3	3	2	1	4	3	2	3	2	2	2		

241

FENOGRAMAS



CLADOGRAMAS

Figure 9. Strict consensus tree of the 28 most parsimonious trees (length 54) based on all 26 unordered morphological characters. Percentage of 10,000 bootstrap replicates, followed by percentage of 10,000 jackknife replicates (in parentheses), is given above branches. The same strict consensus topology was obtained in the analysis excluding the two breeding characters, and nearly the same topology (except for *S. peruvianum* N that was part of the basal polytomy with *S. peruvianum* S, *S. habrochaites*, and *S. chilense*) in the analysis of 15 characters (25, 26, Tables 3, 4).

Character number	Character description	che	gal	pim	lyc
12	Leaf length:leaf axis length ratio	1.808 a	1.455 b	2.064 c	2.012 ac
13	Leaf length:length from widest to leaf apex ratio	1.317 a	1.495 b	1.312 a	1.359 a
14	Leaf length:leaf width ratio	1.199 a	1.292 b	1.121 a	1.179 a
15	Leaf length:terminal leaflet length ratio	2.422 a	3.495 b	2.031 a	2.098 a
16	Leaf axis length:petiole length ratio	2.231 a	3.764 b	2.725 a	2.338 a
17	Leaf axis length:terminal leaflet length ratio	1.422 a	2.495 b	1.031 a	1.098 a
18	Length from widest to leaf apex ratio:petiole length ratio	2.929 a	3.666 b	4.268 b	3.464 a
19	Leaf width:leaf axis length ratio	1.350 a	1.000 b	1.586 c	1.533 a
20	Leaf width:terminal leaflet length ratio	2.007 a	2.749 b	1.827 a	1.819 a
21	Terminal leaflet length:petiole length ratio	1.668 a	1.618 a	2.891 b	2.209 b
22	Terminal leaflet:length from widest to leaf apex ratio	0.565 a	0.447 b	0.661 a	0.673 a

Table 4 Means of leaf characters ratios that were found to be statistically significantly different ($P \leq 0.005$) between the taxa, and are indicated with a unique letter. The boxed regions indicate the characters that statistically support *Solanum galapagense* (see Table 2 for character descriptions). che – *S. cheesmaniae*, gal – *S. galapagense*, pim – *S. pimpinellifolium*, lyc – *S. lycopersicum*.



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MUTACIÓN Y SUSTITUCIÓN

- Mutación (μ): cambio de base en la secuencia de DNA
- Substitución (μ): mutación que se fija en una población (o especie)

$$P(\text{substitución}) = N^{\circ}\text{mutaciones} * P(\text{fijación})$$

- Neutralidad: $P(\text{substitución}) = \mu 2N * 1/(2N) = \mu$



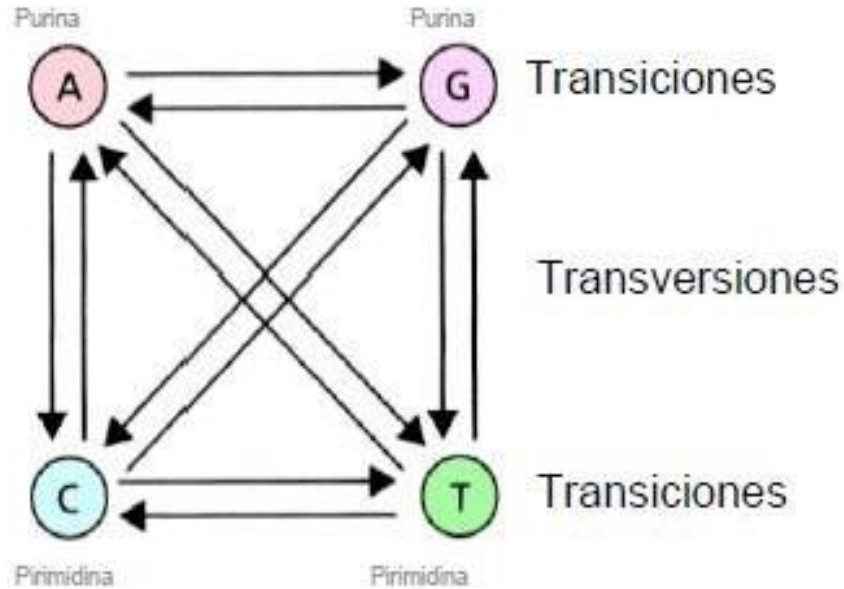
PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

TRANSICIONES Y TRANSVERSIONES





PERÚ

Ministerio de Agricultura y Riego

TASA DE SUBSTITUCIÓN



Instituto Nacional de Innovación Agraria

- Proceso de Markov homogéneo
- Matriz de tasas de sustitución instantánea (Q) (sustituciones por sitio por unidad de tiempo)

$$Q = \begin{bmatrix} -\mu(\pi_C r_{AC} + \pi_G r_{AG} + \pi_T r_{AT}) & \mu r_{AC} \pi_C & \mu r_{AG} \pi_G & \mu r_{AT} \pi_T \\ \mu r_{CA} \pi_A & -\mu(r_{CA} \pi_A + r_{CG} \pi_G + r_{CT} \pi_T) & \mu r_{CG} \pi_G & \mu r_{CT} \pi_T \\ \mu r_{GA} \pi_A & \mu r_{GC} \pi_C & -\mu(r_{GA} \pi_A + r_{GC} \pi_C + r_{GT} \pi_T) & \mu r_{GT} \pi_T \\ \mu r_{TA} \pi_A & \mu r_{TC} \pi_C & \mu r_{TG} \pi_G & -\mu(r_{TA} \pi_A + r_{TC} \pi_C + r_{TG} \pi_G) \end{bmatrix}$$

Substituciones por sitio por unidad de tiempo

μ = tasa de sustitución media instantánea

r_{ij} = tasa de sustitución relativa $i \rightarrow j$

π = frecuencias en el equilibrio, constantes en el tiempo



PERÚ

Ministerio
de Agricultura y Riego

PROBABILIDAD DE CAMBIO

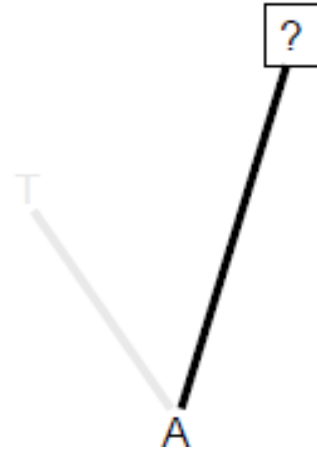


Instituto Nacional de Innovación Agraria

- Podemos calcular la probabilidad de sustitución a lo largo de una rama de longitud t ($\mu \times$ tiempo)

$$P_t = e^{Qt}$$

$$P_t = \begin{bmatrix} P_{AA} & P_{AC} & P_{AG} & P_{AT} \\ P_{CA} & P_{CC} & P_{CG} & P_{CT} \\ P_{GA} & P_{GC} & P_{GG} & P_{GT} \\ P_{TA} & P_{TC} & P_{TG} & P_{TT} \end{bmatrix}$$





PERÚ

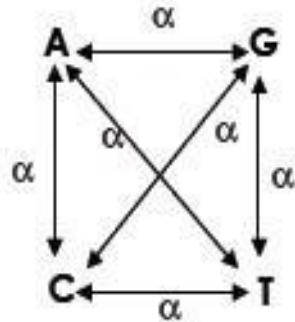
Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MODELO DE JUKES CANTOR (JC69)

1 parámetro (α libres)



$$P_i = \begin{bmatrix} . & \alpha & \alpha & \alpha \\ \alpha & . & \alpha & \alpha \\ \alpha & \alpha & . & \alpha \\ \alpha & \alpha & \alpha & . \end{bmatrix}$$

$$f = \left[\frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4} \right]$$



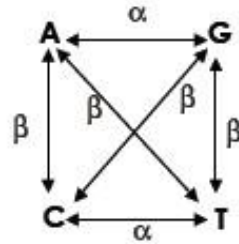
PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

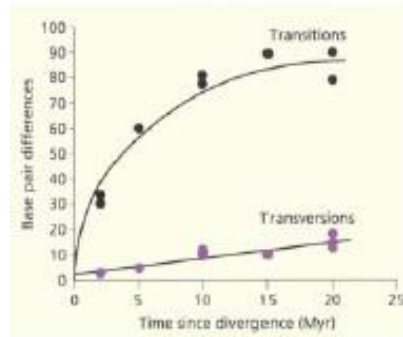
MODELO KIMURA 2 PARAMETROS (K2P O K80)



$$P_t = \begin{bmatrix} . & \beta & \alpha & \beta \\ \beta & . & \beta & \alpha \\ \alpha & \beta & . & \beta \\ \beta & \alpha & \beta & . \end{bmatrix}, \quad \mathbf{f} = \left[\frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4} \right].$$

$$t_i : t_v = \kappa = \alpha / \beta$$

$$D_{K80} = \frac{1}{2} \ln \left(\frac{1}{1 - 2P - Q} \right) + \frac{1}{4} \ln \left(\frac{1}{1 - 2Q} \right)$$



	κ
mtDNA	9.0
12S rRNA	1.75
globinas	0.66
pseudoglobina	2.70



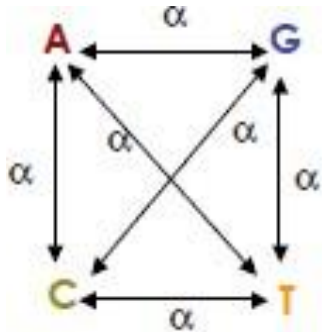
PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MODELO DE FELSENSTEIN 1891 (F81)



$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C\alpha & \pi_G\alpha & \pi_T\alpha \\ \pi_A\alpha & . & \pi_G\alpha & \pi_T\alpha \\ \pi_A\alpha & \pi_C\alpha & . & \pi_T\alpha \\ \pi_A\alpha & \pi_C\alpha & \pi_G\alpha & . \end{bmatrix}, \quad \mathbf{f} = [\pi_A \ \pi_C \ \pi_G \ \pi_T]$$

- Variación en la composición de bases, pero común a todas las secuencias
- Eubacteria (G+C) = 25 - 75%



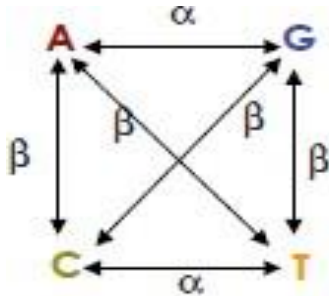
PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MODELO HASEGAWA-KISHINO-YANO (HKY85)



$$P_t = \begin{bmatrix} . & \pi_C \beta & \pi_G \alpha & \pi_T \beta \\ \pi_A \beta & . & \pi_G \beta & \pi_T \alpha \\ \pi_A \alpha & \pi_C \beta & . & \pi_T \beta \\ \pi_A \beta & \pi_C \alpha & \pi_G \beta & . \end{bmatrix}, \quad \mathbf{f} = [\pi_A \ \pi_C \ \pi_G \ \pi_T]$$

- Permite diferentes tasas de transiciones y transversiones
- Permite diferencias en la frecuencia de las bases



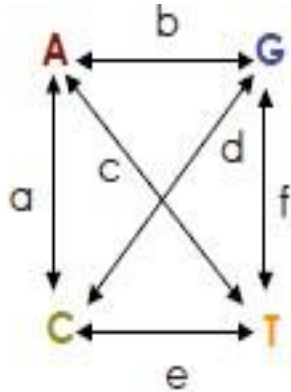
PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MODELO GENERAL DE TIEMPO REVERSIBLE (GTR O REV)



$$\mathbf{P}_t = \begin{bmatrix} . & \pi_C a & \pi_G b & \pi_T c \\ \pi_A a & . & \pi_G d & \pi_T e \\ \pi_A c & \pi_C e & \pi_G f & . \end{bmatrix}, \quad \mathbf{f} = [\pi_A \pi_C \pi_G \pi_T]$$

- Permite 6 tasas diferentes de sustituciones reversibles
- Permite diferencias en la frecuencia de las bases



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

ASUNCIONES DE LOS MODELOS EVOLUTIVOS

- Todos los sitios nucleotídicos cambian de forma independiente
- La tasa de sustitución es constante en el tiempo y entre los distintos linajes
- Las frecuencias de las bases está en equilibrio
- Las probabilidades condicionales de sustitución nucleotídica son las mismas para todos los sitios y no cambian en el tiempo.



PERÚ

Ministerio
de Agricultura y Riego

INFERENCIA FILOGENÉTICA



Instituto Nacional de Innovación Agraria

- Independencia
- Composición de las bases.
- Variación de las tasas entre sitios
- Proporción de sitios invariables
- Distribución gamma



PERÚ

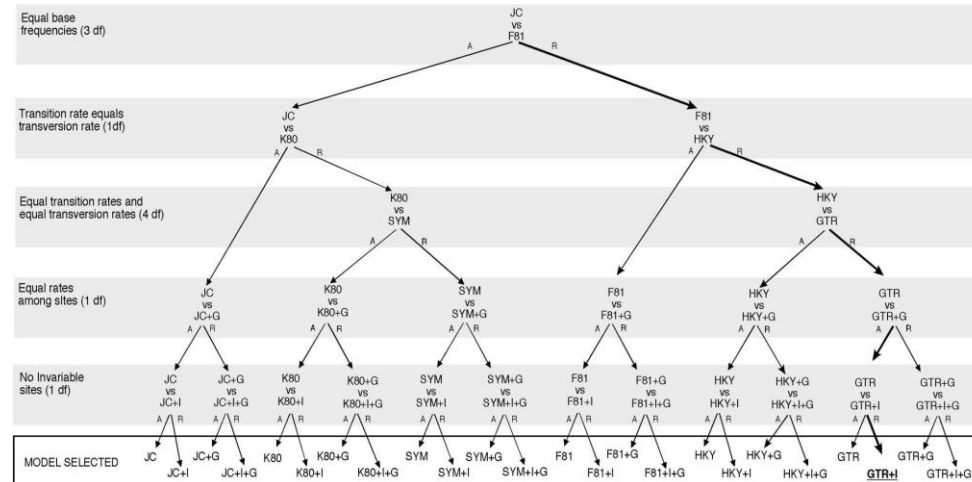
Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

TEST PARA ESCOGER EL MODELO EVOLUTIVO ÓPTIMO

- Test de la razón de verosimilitudes (LRT)
- Test de la razón de verosimilitus jerárquicos (hLRT)
- Criterio de Información de Akaike (AIC)
- Criterio de Información Bayesiano (BIC)





PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MÉTODOS FILOGENÉTICOS

- La inferencia filogenética es un campo *per se* del estudio de la evolución, en continuo movimiento y expansión.
- La inferencia filogenética es un procedimiento de estimación estadística.
- Disponemos de información actual sobre las moléculas (u otros caracteres) e intentamos hipotetizar sobre el pasado
- La pregunta es: ¿cuál es la historia evolutiva de este grupo de taxones?



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

PROPIEDADES DESEABLES DE LOS MÉTODOS FILOGENÉTICOS

- Eficiencia: cuán rápido es el método?
- Poder: ¿cuántos datos se necesitan para producir un resultado razonable?
- Consistencia: ¿convergerá a la solución correcta si proporcionamos suficientes datos?
- Robustez: ¿si se violan algunas asunciones el método sigue proporcionando soluciones razonables?
- Falsabilidad: ¿nos indicará el método cuándo no lo deberíamos estar utilizando?



PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

TIPOS DE MÉTODOS FILOGENÉTICOS

MÉTODO COMPUTACIONAL

		Criterio de optimalidad	Algoritmo de agrupamiento
TIPO DE DATOS	Caracteres	Máxima Parsimonia (MP) Máxima Verosimilitud (ML) Inferencia Bayesiana (BY)	
	Distancias	Evolución mínima (ME) Mínimos cuadrados (LS)	Neighbor-joining (NJ) UPGMA

- Los métodos algorítmicos definen una serie de pasos que conducen a la determinación de un árbol
- Los métodos de optimalidad definen un criterio que permiten comparar diferentes árboles y decidir cuál es mejor (o igual).



PERÚ

Ministerio
de Agricultura y Riego

MÉTODOS ALGORÍTMICOS



Instituto Nacional de Innovación Agraria

- Son bastante “atractivos”
 - Son métodos fáciles de implementar, incluso a mano.
 - El tiempo de computación
 - Suelen resultar en un único árbol
- Tienen limitaciones
 - Pueden depender del orden en el que se añaden los taxones
 - No permiten comparar diferentes árboles
- UPGMA (ya no se usa), Unión del vecino o “Neighbor-joining” (NJ)



PERÚ

Ministerio
de Agricultura y Riego

MÉTODOS DE OPTIMALIDAD



Instituto Nacional de Innovación Agraria

- **Dos pasos:**
 1. Definir el criterio de optimalidad (parsimonia, mínimos cuadrados, verosimilitud, probabilidad posterior)
 2. Buscar el árbol óptimo bajo el criterio en question
- **Proporcionan una forma de establecer un rango de bondad de las diferente soluciones posibles**
- **Estos métodos son mucho más lentos que los métodos algorítmicos, ya que hay que buscar entre varias soluciones posibles**
- **Máxima parsimonia (MP), Mínima evolución (ME), Máxima verosimilitud (ML), Inferencia Bayesiana (Bayes)**



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

BÚSQUEDA DE ÁRBOLES

- Exacta: garantizan todos los óptimos globales (8-20 taxa)
 - Exhaustiva: cuando se evalúan todos los árboles posibles
 - Branch-and-Bound
- Aproximada: se buscan una solución que sea óptima aunque sin garantizar que sea la mejor (¿óptimo local o global?)
 - Heurística: empezamos en un árbol dado y nos movemos a otros árboles intentando mejorar su optimalidad. Cuando no podemos mejorar, paramos.
 - Algoritmos genéticos



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

BÚSQUEDA HEURÍSTICA

1. Hay que empezar en algún sitio: adición secuencial o descomposición de estrella
2. Hay que buscar ("branch swapping"): TBR, SPR, NNI. Conviene hacer varias réplicas de esta búsqueda con diferentes órdenes de adición

MOVIÉNDOSE POR LOS ÁRBOLES

- A partir del árbol inicial, queremos explorar diferentes árboles
- Se efectúan una serie de movimientos y nos quedamos con el mejor árbol
- El proceso se repite hasta que no se pueda mejorar
- Existen varios tipos de movimientos:
 - NNI: intercambio del vecino cercano
 - SPR: poda y regrafo
 - TBR: bisección y reconexión



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MÁXIMA PARSIMONIA

- Ha sido el método más usado ...
- Parsimonia propone simplicidad: evitar hipótesis *ad hoc*
- Para explicar conflictos a menudo recurrimos a las homoplasias
- El método de máxima parsimonia busca minimizar la longitud del árbol, es decir, el número mínimo de cambios entre caracteres



PERÚ

Ministerio
de Agricultura y Riego

MÁXIMA PARSIMONIA



Instituto Nacional de Innovación Agraria

Ventajas

- No depende de algún modelo específico de evolución.
- Da ambos, un árbol y la hipótesis asociada del carácter de evolución.
- Si homoplasia es poco frecuente, da resultados confiables.

Desventajas

- Puede dar resultados no confiables si homoplastia frecuente.
- Subestima longitud de ramas.
- Parsimonia es frecuentemente soportada por argumentos filosóficos en vez de estadísticos.



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Métodos de Distancias

- Alternativa a ML para minimizar el impacto de cambios no observados. Se pueden ver como aproximaciones rápidas a ML.
- Se utilizan distancias corregidas que son estimas de la distancia verdadera entre dos pares de secuencias.
- Se pierde información al transformar los caracteres en distancias.
- Incluyen UPGMA, ME y NJ.



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Métodos de Distancias

Ventajas

- Son muy rápidos.
- Hay muchos modelos por corregir.
- LRT puede ser usado para buscar el mejor modelo.

Desventajas

- Información acerca de evolución de caracteres particulares está perdido.



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

MAXIMA VEROSIMILITUD

- Idealmente, nos gustaría tener un método de parsimonia que
 - Tenga en cuenta las reconstrucciones de estados ancestrales más parsimoniosas pero también las menos parsimoniosas
 - Le diese un peso diferente a los cambios según la longitud de la rama en la que ocurren
 - Le diese un peso diferente a cambios diferentes (p.e., transiciones vs. transversiones)
- Ese método existe. Se llama máxima verosimilitud
 - ... pero requiere un modelo de evolución creíble
 - ...y es computacionalmente intensivo



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Métodos de Distancias

Ventajas

- Cada sitio tiene valor de likelihood.
- Longitud de ramas exactas.
- No hay necesidad de hacer correcciones al modelo.
- Todos los sitios son informativos
- Puede incluir tasa de substitución, frecuencia de caracteres, etc.
- Si el modelo es correcto el árbol obtenido es “correcto”

Desventajas

- Si el modelo es correcto el árbol obtenido es “correcto”
- Demanda capacidad computacional.



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

INFERENCIA BAYESIANA

- La inferencia filogenética Bayesiana está basada en la probabilidad posterior del árbol filogenético, τ .
- La probabilidad posterior de un árbol dado, τ_i , condicionada en una matriz de secuencias de DNA alineadas (X) se obtiene mediante la fórmula de Bayes:

$$f(\tau_i | X) = \frac{f(X | \tau_i) f(\tau_i)}{\sum_{j=1}^{B(s)} f(X | \tau_j) f(\tau_j)}$$

- Generalmente se utiliza una distribución a priori no informativa para los árboles :

$$f(\tau_i) = 1/B(s)$$



PERÚ

Ministerio
de Agricultura y Riego

INFERENCIA BAYESIANA



Instituto Nacional de Innovación Agraria

Ventajas

- Más rápido que Máxima verosimilitud.
- Longitud de ramas exactas.
- No hay necesidad de hacer correcciones al modelo.
- Todos los sitios son informativos
- Puede incluir tasa de sustitución, frecuencia de caracteres, etc.
- Si el modelo es correcto el árbol obtenido es “correcto”
- No hay necesidad de interpretaciones bootstrap

Desventajas

- Cuando saber que convergió
- A que extensión posterior probability es influenciada por la prior probability.



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Metodologías de Remuestreo o Permutaciones

Bootstrap

- Simula repeticiones de variaciones nucleotídicas de la matriz inicial para corroborar la robustez de las ramas
- Mayor capacidad computacional **Jackknife**
- Eliminar datos de variación nucleotídica para corroborar la robustez de las ramas
- Menor capacidad computacional



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

TIPOS DE SELECCIÓN NATURAL

- ✓ Selección Positiva, es un tipo de selección natural que favorece a un solo alelo, y por esto la frecuencia alélica de una población continuamente va en una dirección. Este mecanismo puede conducir adaptaciones.
- ✓ Selección negativa o purificadora, es un tipo de selección natural en el que la diversidad genética decae según un valor particular de carácter.



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

TEORIA NEUTRAL

La gran mayoría de los cambios evolutivos a nivel molecular son causados selectivamente por deriva al azar de mutaciones neutrales o casi neutrales.

Siendo su más importante consecuencia la fijación de mutaciones a una tasa constante dentro de la población



PERÚ

Ministerio
de Agricultura y Riego

Instituto Nacional de Innovación Agraria

Cambios en una secuencia de proteína...

1

89160_mg	MVGERRGTRI	PDLERRGEPR	MTEIPGASTG	RGHVWIVAAG	LNAHRQLLDS	QPDDLTSFQK	IHEGNSDVN
89160_s1	-----	-----	-----	-----MVAAG	LNAHRQLLDS	QPDDLTSFQK	IHEGDGSDFN
89160_s2	-----	-----	MADASRESAG	RGRVRIIAAG	LNAHSQLLDG	QPEDLTSFQK	IHEGDGSDVK

71

89160_mg	VWFASWSNTI	FATDAGLRGI	GHQKVLDNSK	VLLVSPVGDH	DGVLALDSD	GOLYRAEEQT	EQGIVAMVCR
89160_s1	VWFASWSNTI	FATDAGLRSI	GHQKVLDNSK	VLLVSPVGDH	NGVLALDSD	GOLYRAEEQT	EQGIVAMVCO
89160_s2	VWFASWSNTV	FATDAGLRGI	GHQRVLDNSK	VLLVSPVGDH	NGVLALDSD	GOLYRAEEQT	EQGIVAMVCO

141

89160_mg	STDSSPSIGR	LAYAENGRVA	ITIKQAPNGN	LCHVEEFKDL	ETFLRWFQDP	SGDGNYPERH	FMLPGRPKQL
89160_s1	STDLSPSIGR	LAYAENGRVA	ITIKQAPNGN	LCHVEEFKDL	ETFLRWFQDP	SGDGNYPERH	FMLPGRPKQL
89160_s2	STDLSPSIGR	LAYAENGRVA	ITIKQAPNGN	LCHVEEFKDL	EPFLRWFQDP	SGDGNHPERH	FMLPGRPKQL

211

89160_mg	KAGTGIFVLL	MESGQVYTWG	DSRYRSLGRS	VTGDGSKSAD	EPAVLEALDG	LHIKKVDCCG	WMSAALSDDG
89160_s1	EAGTGIFVLL	MESGQVYTWG	DSRYRSLGRS	VTGDGNKSAD	EPAVLEALDG	LHIKKVDCCG	WMSAALSDDG
89160_s2	EAGTGIFVLL	MESGQVYTWG	DPRFRSLGRS	VTGDGNKSAD	EPAVLEALDG	LHIKKVACCG	WMSAALSDDG

281

89160_mg	ALYLWGITSP	SDDVKIGALA	AGEDEEVALV	ELPGDGSEPL	DVVDVALGVE	HIAVLAESGR	LFVTGDKSCG
89160_s1	ALYLWGITSP	SDDVKIRALA	AGEDEEVALV	ELPGDDSEPL	DVVDVALGVE	HIAVLAESGR	LFVTGDNSCG
89160_s2	ALYLWGITSP	SGDVIINALT	AGEDEEIALV	ELPGGGSEPL	DVVDVALGVE	HIAVLGESGR	LFVTGDNSCG



PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Cuantificando variación no sinónima - Una estimación de selección positiva

Mutaciones sinónimas: mutaciones neutrales

Mutaciones no sinónimas: mutaciones no neutrales

		Second Position												
		U		C		A		G						
		code	Amino Acid	code	Amino Acid	code	Amino Acid	code	Amino Acid					
First Position	U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U	Third Position			
		UUC		UCC			UAC		UGC			C		
		UUA	leu	UCA			UAA	STOP	UGA	STOP		A		
		UUG		UCG			UAG	STOP	UGG	trp		G		
	C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U				
		CUC				CCC		CAC				CGC		C
		CUA				CCA		CAA		gln		CGA		A
		CUG				CCG		CAG				CGG		G
	A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U				
		AUC				ACC		AAC		AGC			C	
		AUA				ACA		AAA	lys	AGA		arg	A	
		AUG		met		ACG		AAG		AGG			G	
	G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U				
		GUC				GCC		GAC				GGC		C
		GUA				GCA		GAA		glu		GGA		A
		GUG				GCG		GAG				GGG		G



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Para medir la selección positiva:

Tasa de mutaciones sinónimas Tasa de
mutaciones no sinónimas



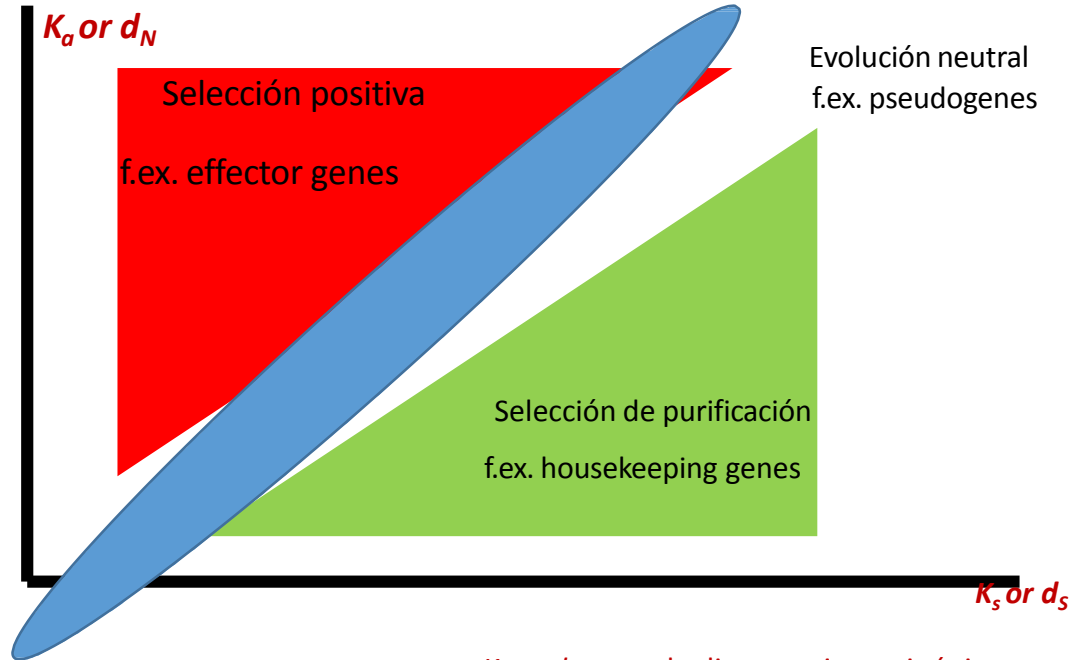
PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Selección positiva entre especies



Evolución entre especies

K_A or d_N : tasa de divergencia no sinónimas

K_S or d_S : tasa de divergencia sinónima



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Estimaciones de polimorfismos y sustituciones no sinónimas y sinónimas proveen perspectiva en los procesos evolucionarios

Analizando divergencia y polimorfismo

- K_A / K_S ratios > 1 indica selección positiva
- K_A / K_S ratios < 1 indica selección negativa
- K_A / K_S ratios $= 1$ indica evolución neutral

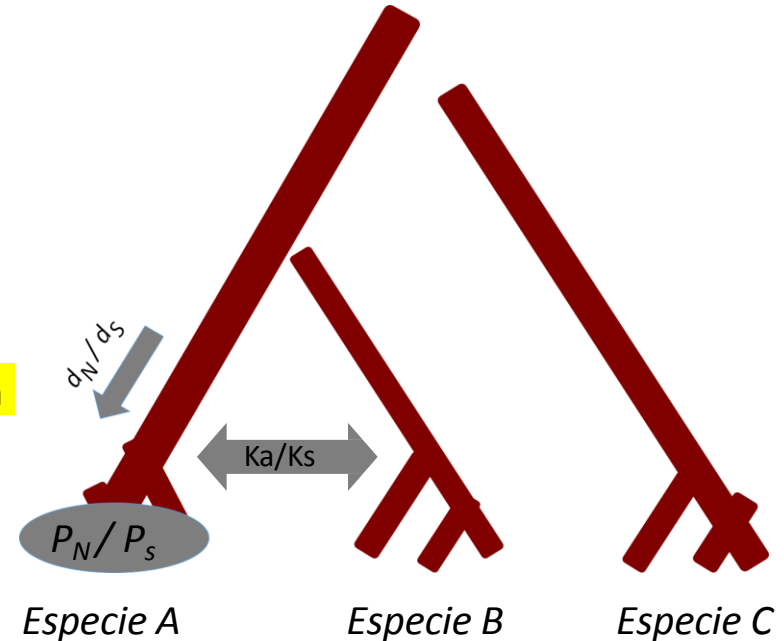
Estimación específico de rama

K_A and d_N : tasa de sustituciones no sinónimos

K_S and d_S : tasa de sustituciones sinónimas P_N :

Cantidad de polimorfismos no sinónimos P_S :

Cantidad de polimorfismos sinónimos





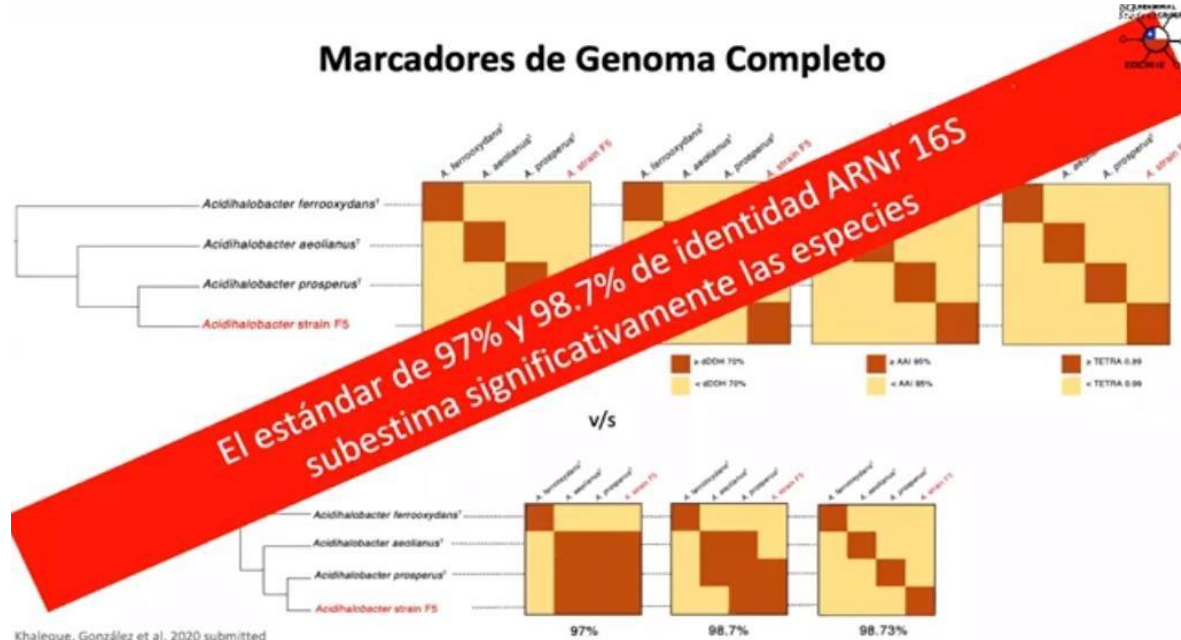
PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

¿Cuánto deben diferir las secuencias para indicar que son especies diferentes?





PERÚ

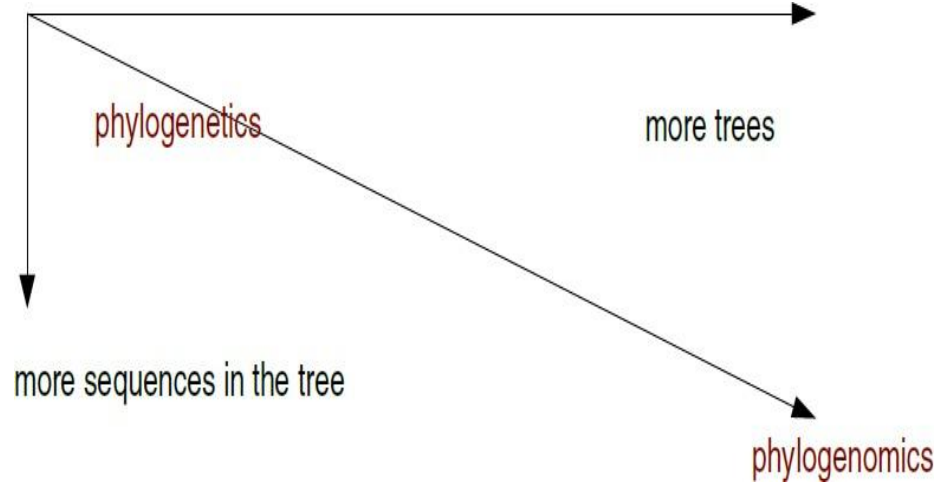
Ministerio
de Agricultura y Riego

FILOGENÓMICA



Instituto Nacional de Innovación Agraria

Es referido a la intersección entre los campos de la evolución y genómica. Por lo tanto, estudios de procesos evolucionarios a escala genómica.





PERÚ

Ministerio
de Agricultura y Riego

FILOGENÓMICA



Instituto Nacional de Innovación Agraria

Requiere gran cantidad de datos.

Recursos computacionales potentes y algoritmos rápidos.

El tiempo es un problema y la velocidad es alcanzada con pérdida de resolución o calidad de los resultados.

Pero es esperado que se mejore con la inclusión de más datos y nuevos métodos

Aplicaciones de Filogenómica

Filogenómica y el árbol de la vida.

Filogenómica como inferencia filogenética a larga escala. Filogenómica para predicción de funciones



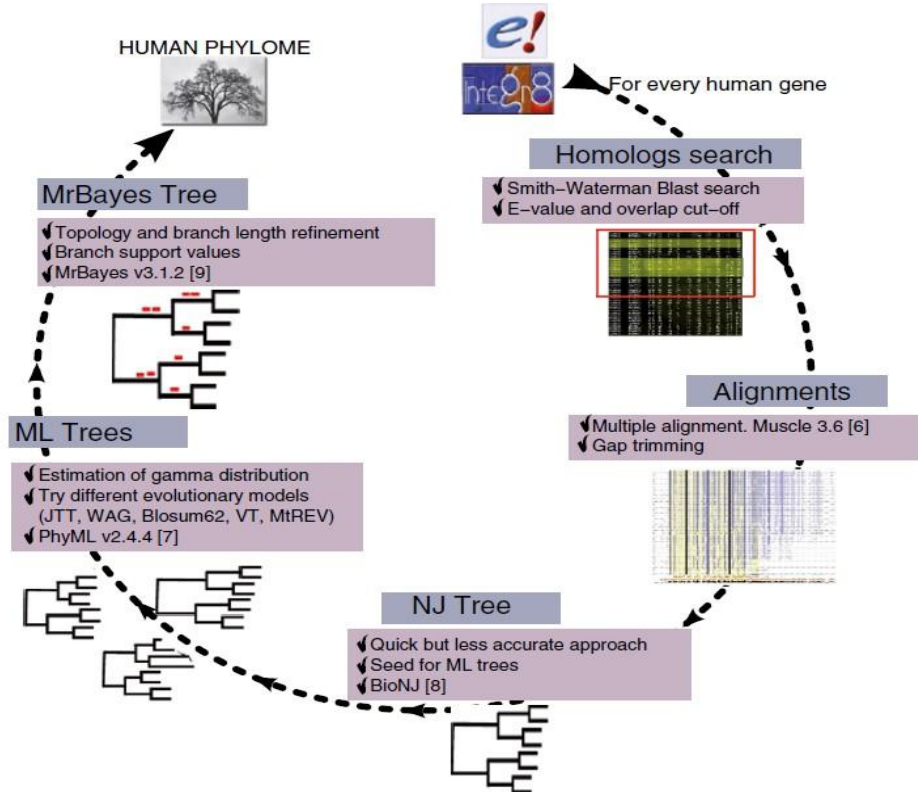
PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria

RECONSTRUCCIÓN DEL FILOMA HUMANO



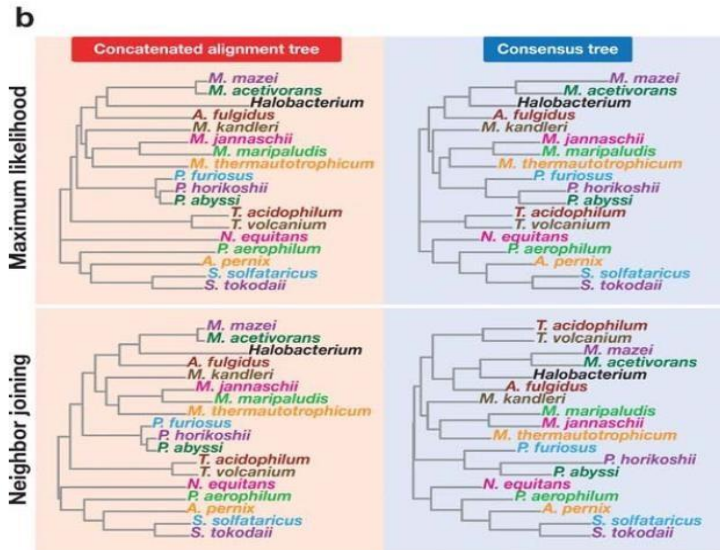
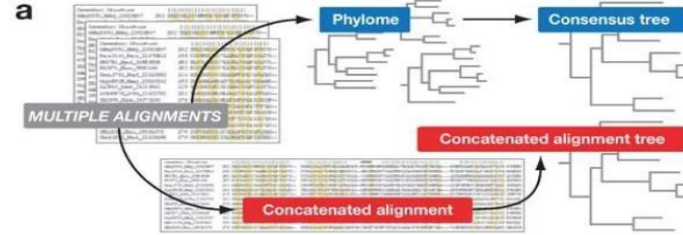


PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria



ÁRBOL CONSENSO VS ÁRBOL DE ALINEAMIENTO CONCATENADO



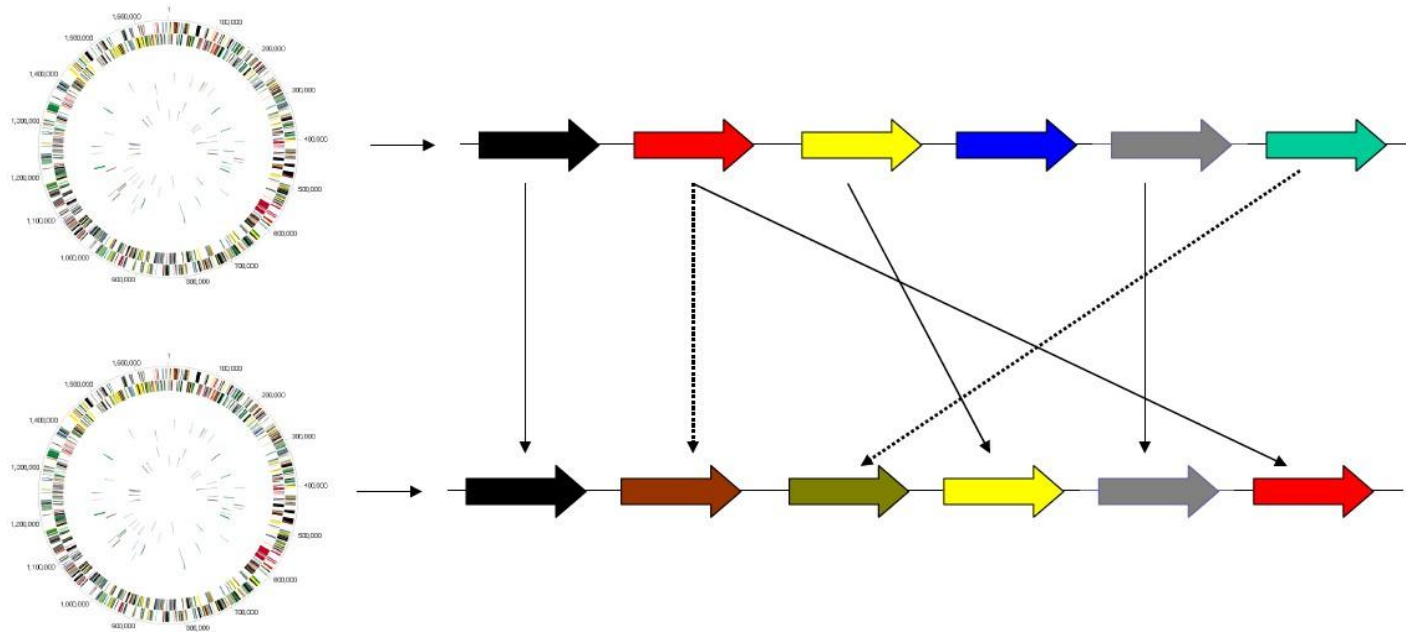
PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Orthology/paralogy: finding equivalent genes in different genomes





PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

Muchas gracias



PERÚ

Ministerio
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

EL PERÚ PRIMERO