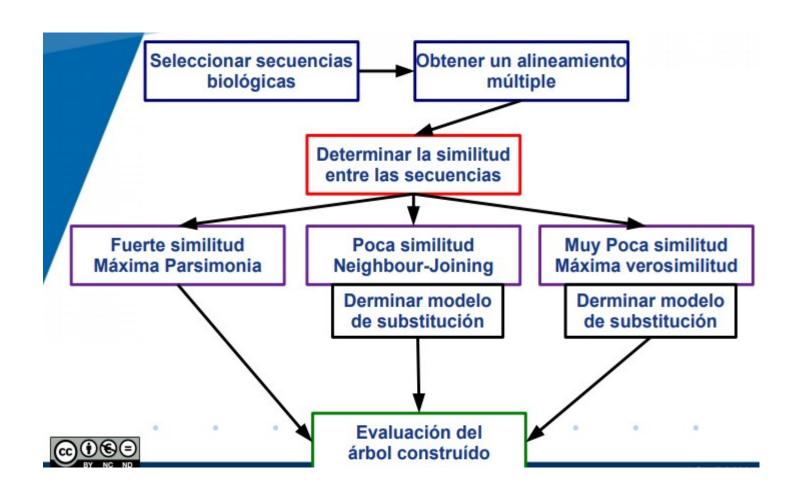
Qué Metodo de construcción de árboles seguir?



Filogenia:

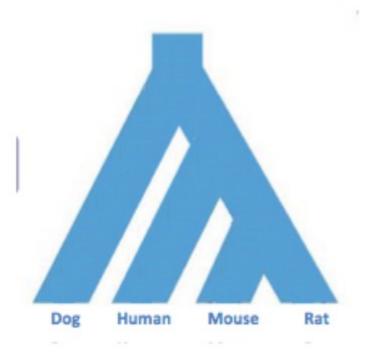
Filogenia

Reconstrucción filogenética Inferencia de filogenias

Arboles de Genes Arboles de Especies Reconciliación de Arboles

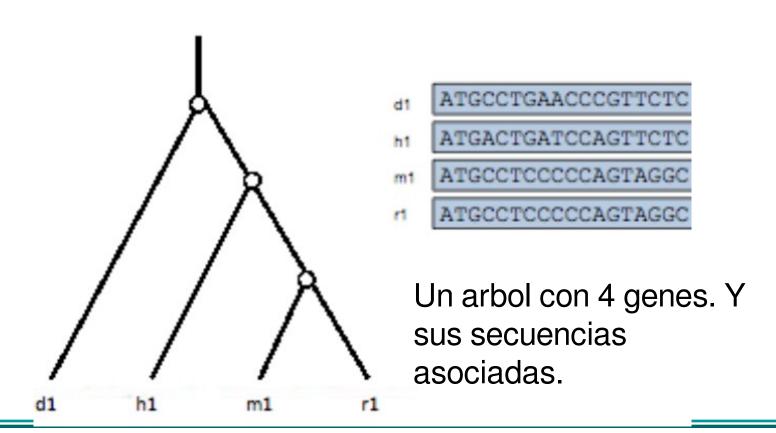
Filogenia: Arboles de Especies

Arboles de Especies: Muestran Cómo diferentes especies evolucionan una de otra. Estos Arboles se crean usando caracteres morfológicos, evidencia fosil, etc. Las hojas de cada arbol se etiquetan como especies, y el resto del arbol muestra como estas especies estan relacionadas. Una Especie se puede pensar como una "Bolsa de Genes", es decir el grupo de genes comunes a una especie.



Filogenia: Arboles de Genes

Arboles de Genes: Son árboles que tienen información de diferentes genes en diferentes especies. Las hojas de los árboles de genes se etiquetan con secuencias de genes o con Identificadores de Genes asociados con secuencias especificas.

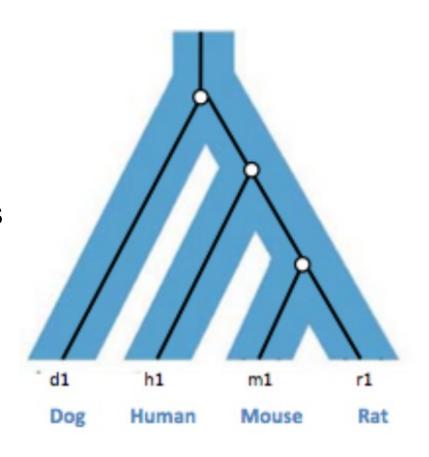


Filogenia:

Evolucion de los Arboles de Genes:

Los árboles de genes evolucionan dentro de arboles de especies.

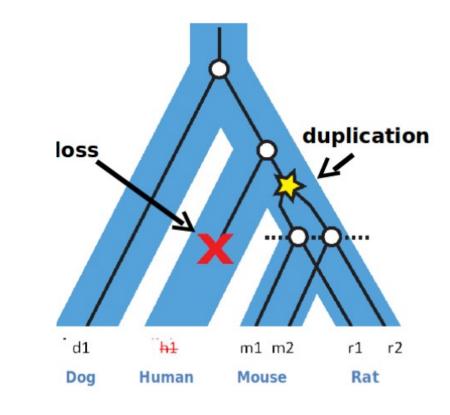
Los métodos de Reconciliacion permiten hacer coincidir arboles de genes dentro de arboles de especies.



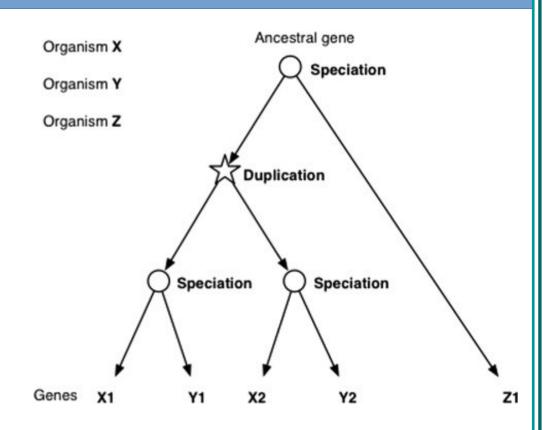
Reconciliacion:
Es un algoritmo ayuda a
compararar arboles de genes
con arboles de especies. La
idea radica en hacer coincidir
un arbol de genes dentro de
un arbol de especies.

Genes Ortólogos: Dos Genes son ortólogos si su mas reciente ancestro Comun (MRCA) es una especiación (División en 2 diferentes especies)
Genes Parálogos: Genes cuyo MRCA es una Duplicación

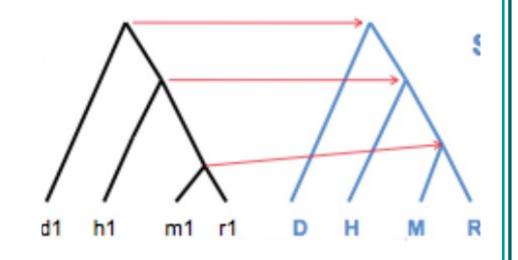




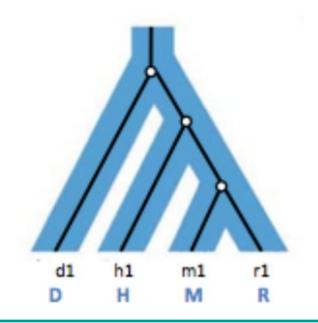
Las hojas del árbol representan los genes y los nodos internos corresponden a los eventos evolutivos. X1 tiene dos nodos ancestrales asociados con eventos de especiación; uno es el último ancestro común con Y1, y el otro es común con Z1. Por lo tanto, Y1 y Z1 son ortólogos de X1. Por otro lado, X2 e Y2 son parálogos para X1, ya que su último antecesor común tiene asociado un evento de duplicación.



Un Diagrama de Mapeo es un diagrama que muestra el mapeo de nodos de un árbol de genes a un árbol de especies

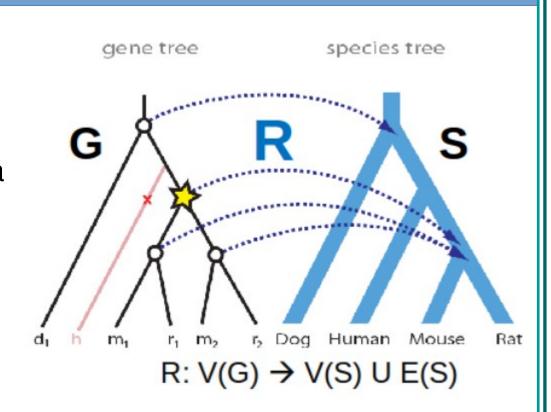


Un diagrama de anidación muestra como un arbol de genes puede ser anidado dentro de una arbol de especies.



Algoritmo de Reconciliacion con Maxima Parsimonia (MPR)

Es un Algoritmo encuentra la reconciliación que calza un arbol de genes dentro de un árbol de especies, mientras minimiza el número de duplicaciones y de borrados



Algoritmo de Reconciliacion con Maxima Parsimonia (MPR). Consiste en hacer coincidir las hojas del árbol de genes con las hojas del arbol de especies. El algoritmo progresa hacia los vértices del arbol de genes, trazando una relacion entre el MRCA de todas las hojas dentro del sub-arbol de un determinado vértice y el MRCA correspondiente en un árbol de especies

```
Solve recursively:
```

```
•R[v] = species of v
                                                  if v \in L(G)
\cdot R[v] = LCA(R[right(v)], R[left(v)])
                                                if v \in I(G)

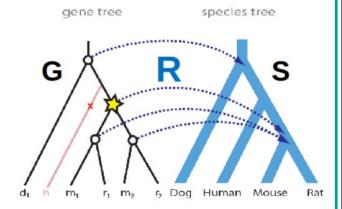
    LCA = "least common ancestor"

       (also called "most recent common ancestor")
```

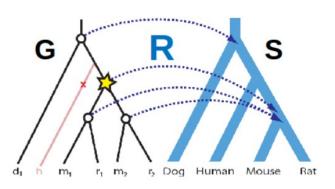
Labeling events:

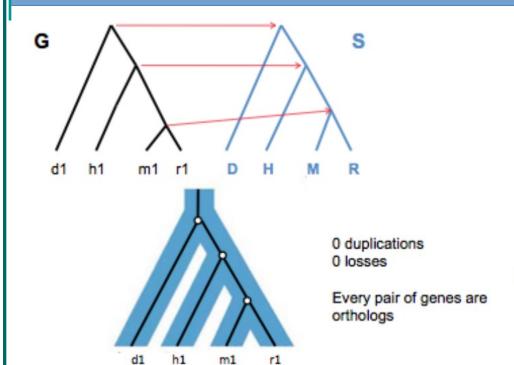
•v is a dup if R[v] = R[right(v)] or R[left(v)]

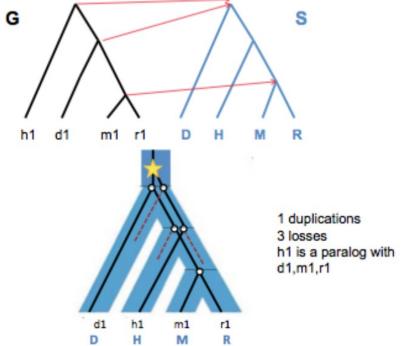
Branch above v has at least one loss if R[parent(v)] != R[v] or parent[R[v]]



Algoritmo de Reconciliacion con Maxima Parsimonia (MPR). El algoritmo avanza recursivamente de abajo hacia arriba, empezando por las hojas. Por cuanto se toman genes de especies conocidas para construir el arbol de genes, existe un mapeo directo entre las hojas del arbol de genes y las hojas del arbol de especies. Para mapear los ancestros, para cada nodo (en la direccion hacia arriba) buscamos el hijo de la derecha y de la izquierda y tomamos el el ancestro comun (Least comon ancestor - LCA) de la especie a la que ellos se mapean. SI un nodo mapea a su hijo de la derecha o izquierda, sabremos que hay una duplicacion. Una rama que no exista significa una perdida (borrado).

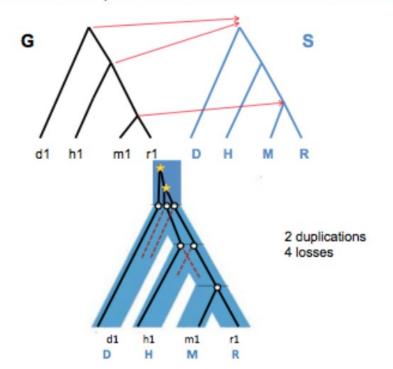


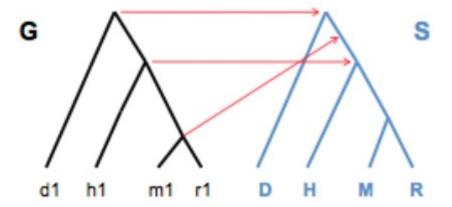




Filogenia: Reconciliacion no MPR e invalida

does a non-parsimonious reconciliation look like?

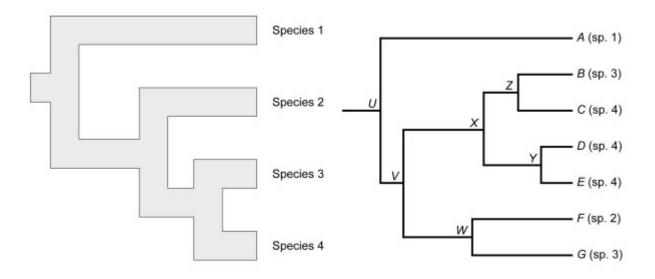




Valid reconciliations must respect descendentancestor relationships.

If a < b in G, then R[a] <= R[b] in S

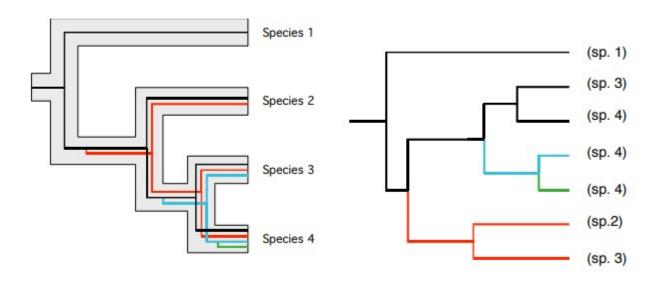
Reconciliation



A reconciliation is a map between a gene tree and a species tree with gene duplications and losses being postulated to explain any incongruence between the trees.

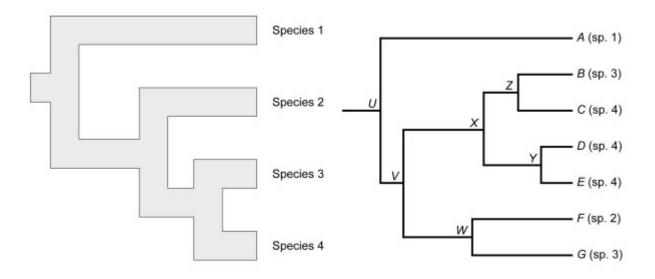
6

Homologs: unrecognized paralogy



Suppose that certain homologs are lost or not yet found, and that sp. 2 only appears once, as above. We will think red 2 and red 3 are orthologs.

Reconciliation



A reconciliation is a map between a gene tree and a species tree with gene duplications and losses being postulated to explain any incongruence between the trees.

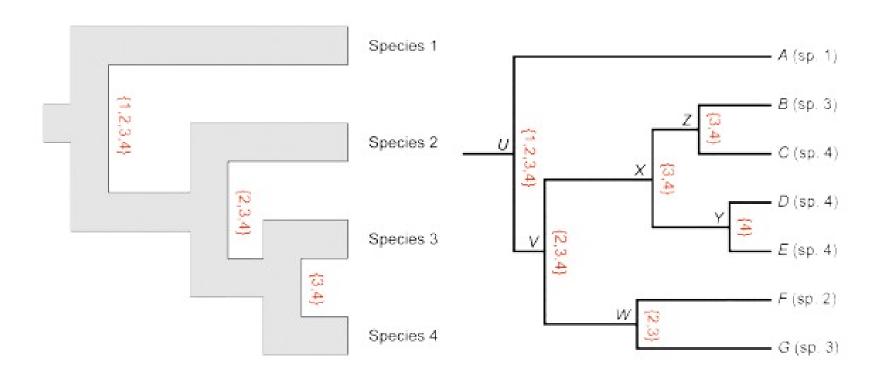
6

Algoritmo

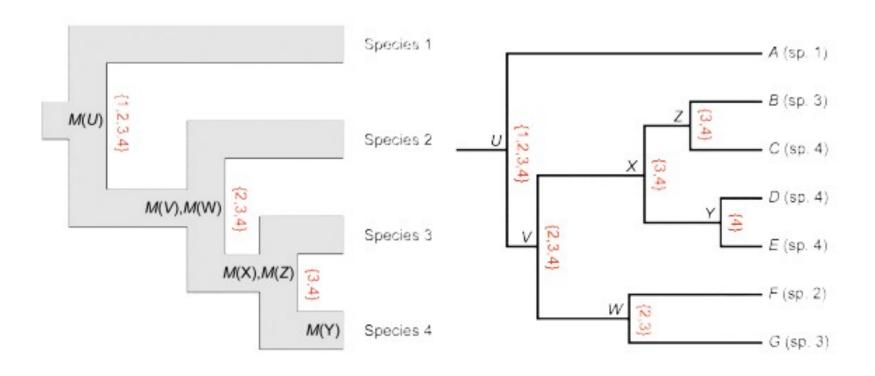
Sean S y G conjuntos de Nodos del arbol de especies y el arbol de genes. (Ambos Arboles se asumen enraizados y binarios) Para g \in G se define $\sigma(g)$ como el conjunto de especies que estan en el sub arbol que empieza en el nodo g. Para s \in S se define $\sigma(s)$ en forma similar.

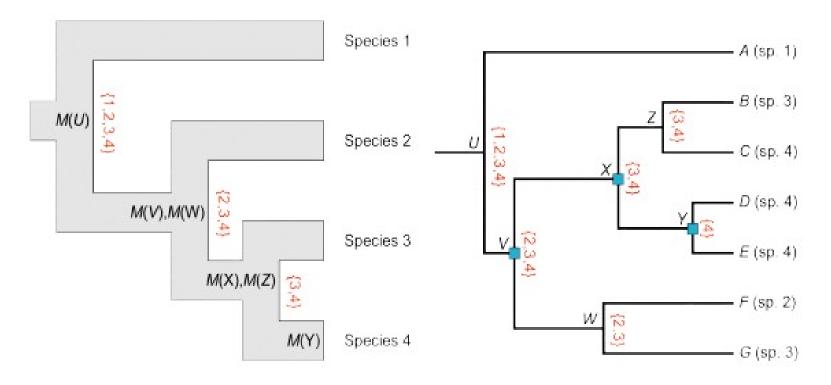
Un mapeo de G a S: Para cada $g \in G$, sea M(g) el mas reciente $s \in S$ para el cual $\sigma(g) \subseteq \sigma(s)$ Para cualquier nodo interno $g \in G$ con hijos g_1 y g_2 Inferimos que g representa un evento de duplicacion, si solo si M(g) es igual a $M(g_1)$ ó $M(g_2)$ es decir, si el nodo g es mapeado en la misma posicion del arbol de especies a la cual mapea uno (o ambos) de sus hijos

$\{\sigma(s):s\in S\}$ and $\{\sigma(g):g\in G\}$



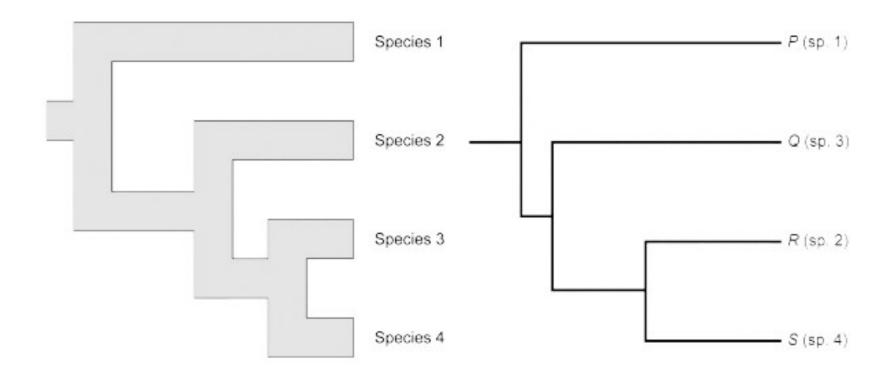
Reconciliation mapping $M:G \rightarrow S$





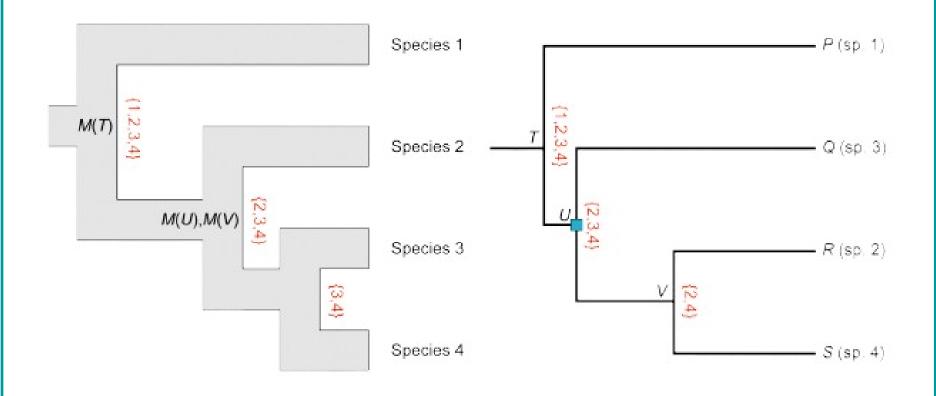
Ahora se podrían inferir genes perdidos o no encontrados todavía, completando la parte del arbol de especies que queda en cada duplicación

Inconsistencies between S and G

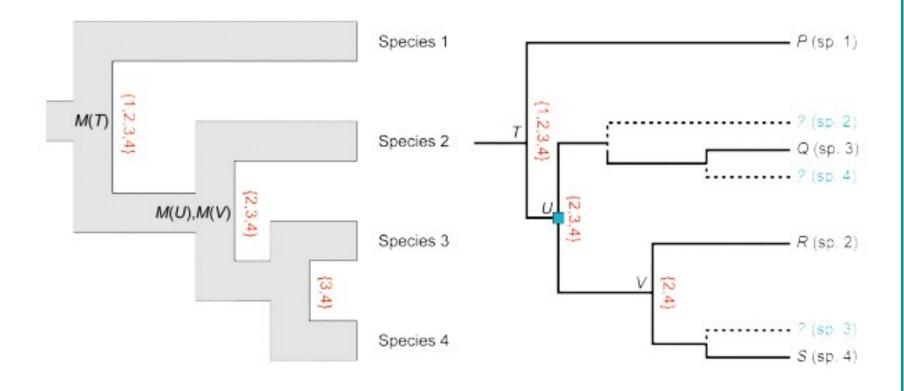


Inconsistencias se pueden resolver de la misma manera

Inconsistencies between S and G



Inconsistencies between S and G



Insertando Genes perdidos o no encontrados todavía se reconcilian los árboles

Los Arboles de genes cuando estan reconciliados con los arboles de especies, ofrecen información signficante sobre eventos de la evolución – especificamente sobre duplicaciones y perdidas.

Duplicacion indica que el mismo gen se encuentra en un sitio separado (ej m2 o r2) en este caso es un mecanismo para crear nuevos genes y funciones. Las consecuencias de esto caen en tres categorias:

No Funcionalización: Una de las copias No funciona.

Neo – Funcionalizacion: Una de las copias desarrolla una funcion enteramente nueva

Sub – Funcionalizacion: Las copias retienen diferentes partes (dividen el trabajo) y en conjunto, realizan la misma funcion.

Mapeo de Secuencias

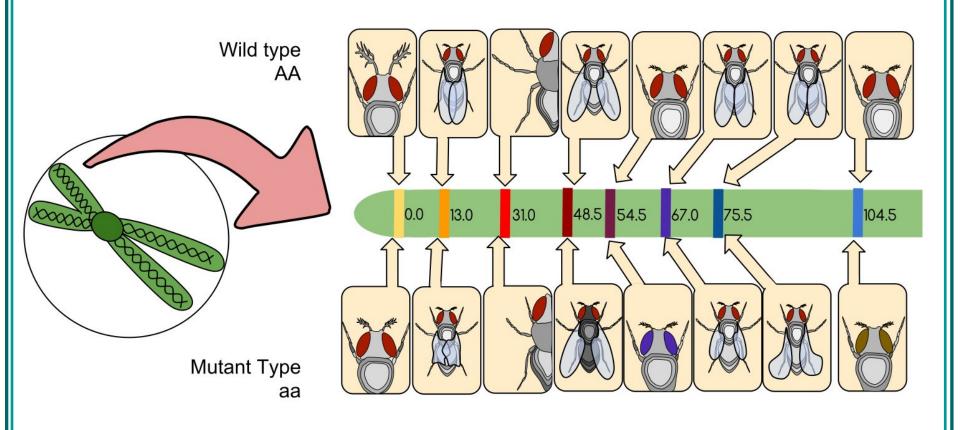
- 1. El Problema de Double Digest y Partial Digest.
- 2. Tecnicas utilizadas en el mapeo de secuencias.
- 3. Mapeo con non Unique Probes
- 4. Mapeo con Unique Probes

Mapeo de Secuencias:

El mapeo genético - también llamado mapeo de vinculación - puede ofrecer pruebas firmes de que una enfermedad transmitida de padres a hijos está vinculada a uno o más genes. El mapeo también proporciona pistas sobre qué cromosoma contiene el gen y precisamente donde el gen se encuentra en ese cromosoma.

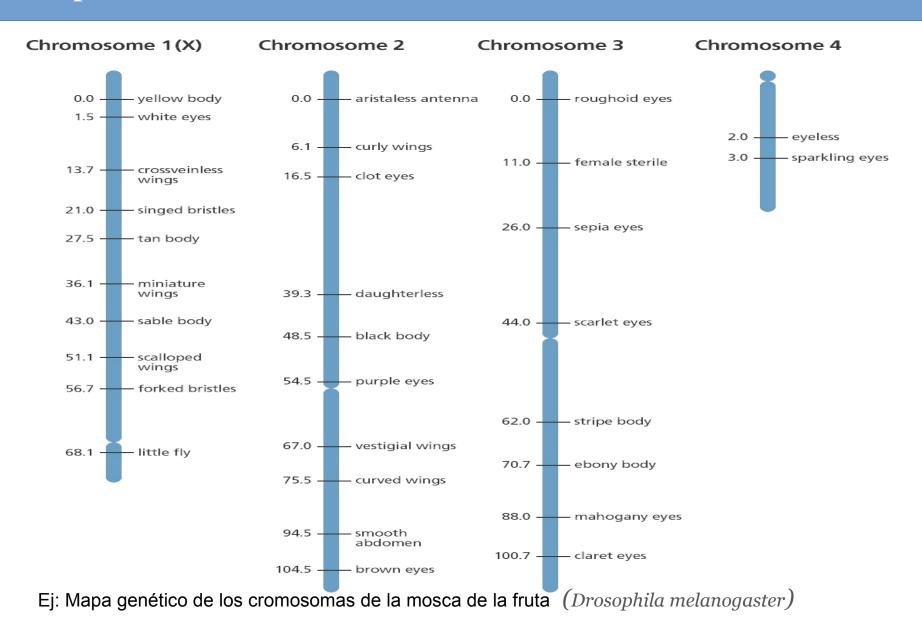
Los mapas genéticos se han utilizado con éxito para encontrar el gen responsable de los trastornos heredados de un solo gen, relativamente raros, como la fibrosis quística y la distrofia muscular de Duchenne. Los mapas genéticos también son útiles para guiar a los científicos a los muchos genes que se cree que juegan un papel en el desarrollo de trastornos más comunes como el asma, las enfermedades del corazón, la diabetes, el cáncer y las condiciones psiquiátricas.

Mapeo de Secuencias



El mapeo del Genoma se utiliza para identificar y registrar la ubicación de los genes y las distancias entre los genes en un cromosoma.

Mapeo de Secuencias:



Mapeo de Secuencias: Marcador Genético

Un marcador genético es un marcador molecular que puede ser gen o una secuencia de ADN con una localización conocida en un cromosoma, que puede usarse para identificar individuos o especies. Se puede describir como una variación (que puede surgir debido a la mutación o alteración en posiciones conocidas o loci genómicos) que se puede observar. Un marcador genético puede ser una secuencia corta de ADN, tal como una secuencia que rodea un solo cambio de par de bases.

Ejemplo de marcadores más utilizados :

RFLP (del inglés *Restriction Fragment Length Polymorphism*) se refiere a secuencias específicas de <u>nucleótidos</u> en el <u>ADN</u> que son reconocidas y cortadas por las <u>enzimas de restricción</u> (también llamadas <u>endonucleasas</u> de restricción)

Mapeo de Secuencias

Discovering Restriction Enzymes



Werner Arber

Daniel Nathans

Hamilton Smith

Werner Arber – discovered restriction enzymes

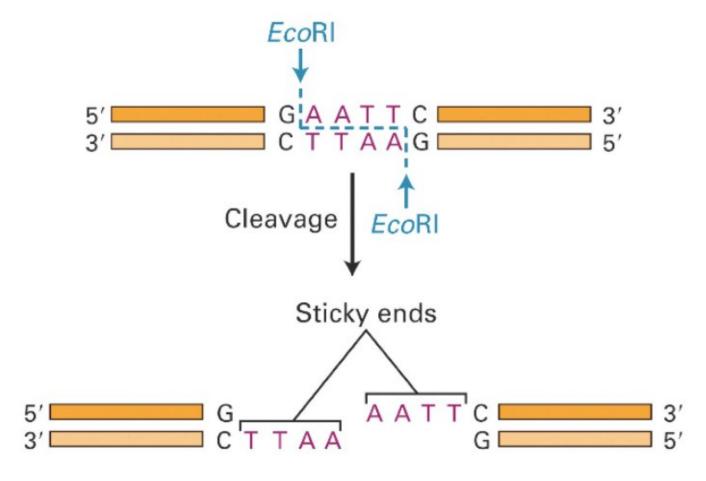
Daniel Nathans - pioneered the application of restriction for the construction of genetic maps

Hamilton Smith - showed that restriction enzyme cuts DNA in the middle of a specific sequence

My father has discovered a servant who serves as a pair of scissors. If a foreign king invades a bacterium, this servant can cut him in small fragments, but he does not do any harm to his own king. Clever people use the servant with the scissors to find out the secrets of the kings. For this reason my father received the Nobel Prize for the discovery of the servant with the scissors".

Daniel Nathans' daughter (from Nobel lecture)

Mapeo de Secuencias



Un mapa de restricción mostrará todos los lugares donde ocurre esta secuencia GAATTC

Mapeo de Secuencias: Enzimas de Restricción

Enzimas de Restricción y las secuencias que reconocen

GAC/GTC

Zral

Enzyme	Recognition Sequence	Enzyme	Recognition Sequence
EcoRI §	G/AATTC	Haelli	GG/CC
EcoRI-HF®	G/AATTC	Hgal	GACGC(5/10)
EcoRV §	GAT/ATC	Hhal	GCG/C
EcoRV-HF®	GAT/ATC	HincII	GTY/RAC
Fatl	/CATG	HindIII §	A/AGCTT
Fnu4HI	GC/NGC	HindIII-HF®	A/AGCTT
Fsel	GGCCGG/CC	Zral	GAC/GTC
Fspl	TGC/GCA		

Mapeo de Secuencias: Enzimas de Restricción

Otros usos de las enzimas de retricción

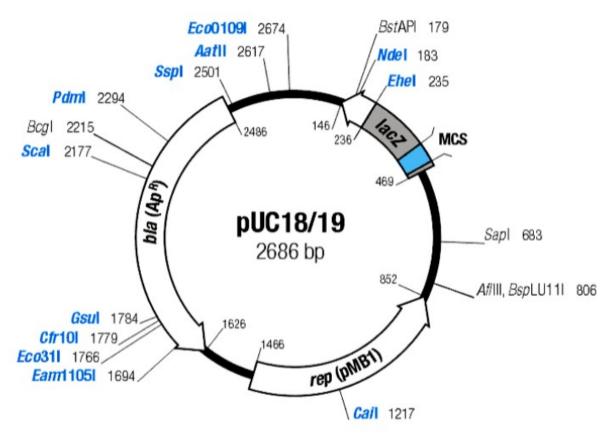
- Recombinant DNA technology
- Cloning
- cDNA/genomic library construction
- DNA mapping

El ADN complementario (ADNc) es ADN sintetizado a partir de un molde de ARN monocatenario (por ejemplo, ARN mensajero (ARNm) o microARN en una reacción catalizada por la enzima transcriptasa inversa. El cDNA se usa a menudo para clonar genes eucariotas en procariotas.

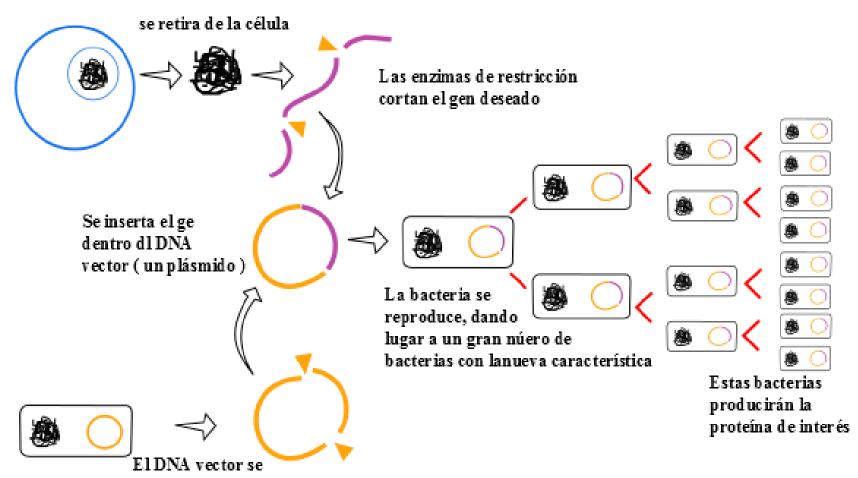
Una biblioteca de ADNc es una combinación de fragmentos de ADNc clonado (ADN complementario) insertados en una colección de células hospedadoras, que juntas constituyen una porción del transcriptoma del organismo y se almacenan como una "biblioteca". El cDNA se produce a partir de ARNm completamente transcrito que se encuentra en el núcleo y, por lo tanto, contiene solo los genes expresados de un organismo.

Mapeo de Secuencias: Enzimas de Restricción

- A map showing positions of restriction sites in a DNA sequence
- If DNA sequence is known then construction of restriction map is a trivial exercise
- In early days of molecular biology DNA sequences were often unknown
- Biologists had to solve the problem of constructing restriction maps without knowing DNA sequences



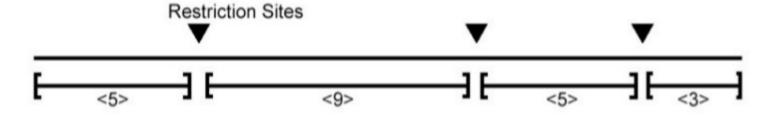
Mapeo de Secuencias: Mapeo con señales de restricción



toma de la bacteria y se corta con las mismas enzimasde restricción

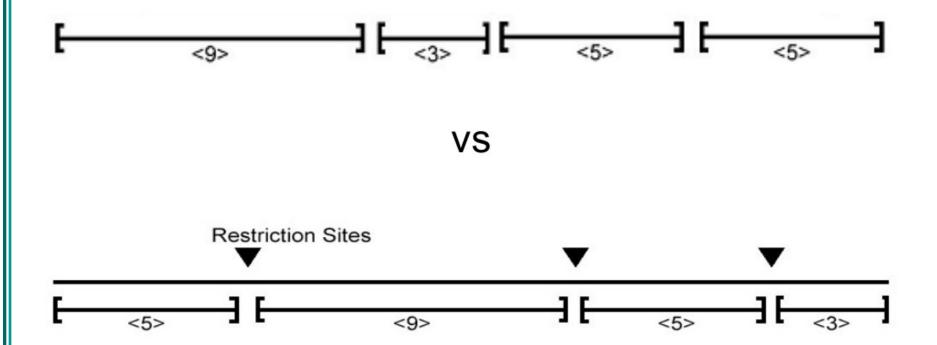
Mapeo de Secuencias: Mapeo con señales de restricción

Cortar el ADN en cada sitio de restricción crea muchos fragmentos de restricción:



Será posible reconstruir el orden de los fragmentos a partir de sus tamaños? {3,5,5,9}

Ordenamiento alternativo de los fragmentos de restricción – multiples soluciones



Midiendo la longitud de los fragmentos de restricción

Las enzimas de restricción cortan el ADN en segmentos de restricción.

La Electroforésis en Gel es un proceso para separar el ADN por su longitud, y para medir los tamaños de los fragmentos de restricción

Puede separar fragmentos de ADN que difieren en longitud desde un nucleótido hasta fragmentos de 500 nucleótidos de longitud.

Electroforésis en GEL

Los Fragmentos de ADN se inyectan en un Gel que se encuentra en un cámpo electrico

ADN esta negativamente cargado.

El componente formado por la Ribosa – en el ARN y la desoxi ribosa en el ADN- en conjunto con el grupo fosfato tiene una carga total negativa.

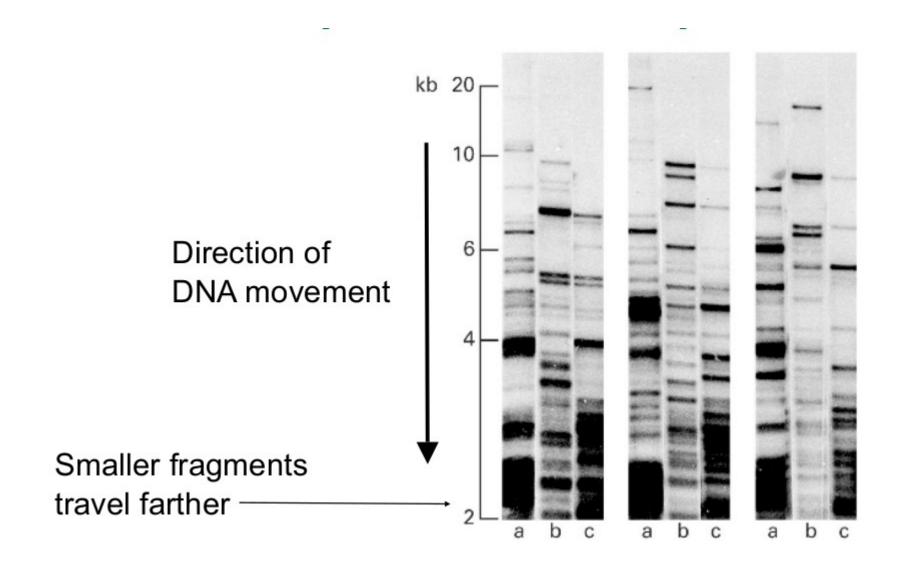
Las moléculas de ADN se mueven hacia el electrodo positivo.

Electroforésis en GEL

Los Fragmentos de ADN de diferentes longitudes son separados de acuerdo a su tamaño

Las moléculas de ADN más pequeñas de se mueven atraves de la matriz del gel más rapido que las moléculas más grandes.

La matriz del gel restringe una difusión aleatoria, asi que moléculas de diferentes tamaños se agrupan en bandas diferentes



Autoradiografia

Una forma de visualizar las bandas separadas en un gel es la autoradiografia.

El ADN se marca radioactivamente.

El Gel se coloca sobre una hoja de material fotosensitivo en la oscuridad.

La placa fotográfica presentará las posiciones donde el material de ADN esta presente.

Detectando el ADN: Fluorescencia

Otra forma de visualizar las bandas separadas de ADN en un gel es la fluorescencia.

El gel es preparado con una solución que contiene el tinte fluorescente ethidium

Ethidium se adhiere a las moléculas de ADN

El ADN se muestra cuando es expuesto a luz ultravioleta.

Partial restriction digest

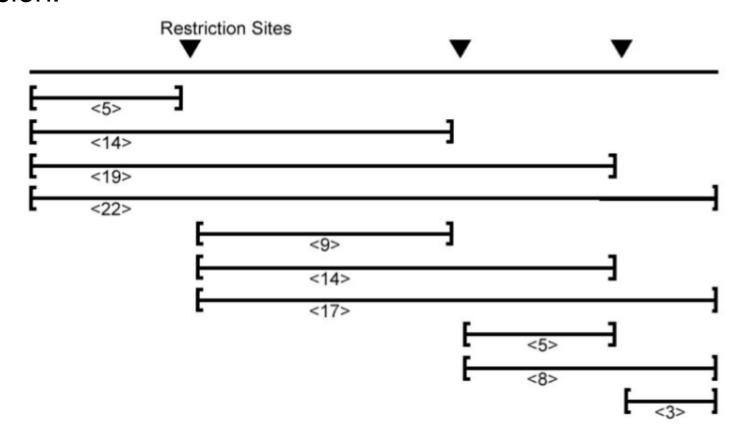
Una muestra de ADN se expone a la enzima de restricción por un tiempo limitado para prevenir que sea cortada en todos los sitios de restricción.

El experimento genera un conjunto de todos los posibles fragmentos de restricción entre dos cortes.

El conjunto de longitudes de estos fragmentos se usa para determinar las posiciones de los sitios de restricción en la secuencia de ADN

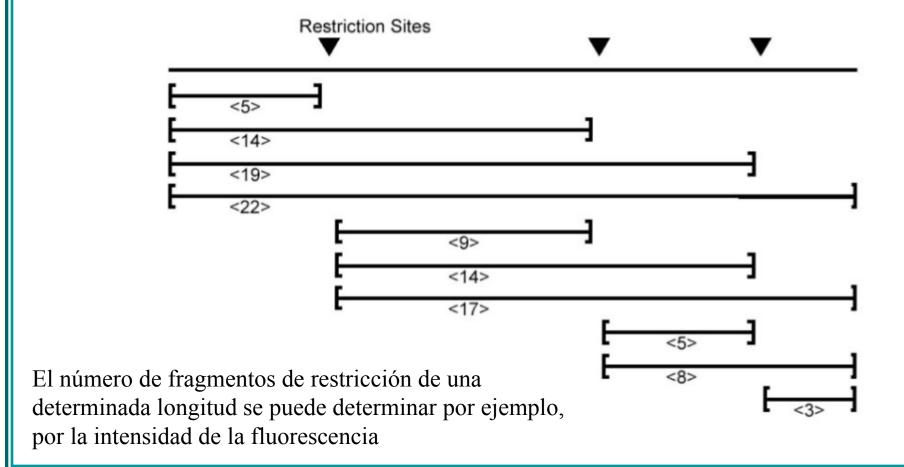
Ejemplo de Partial digest

Un Partial Digest resulta en los siguientes 10 fragmentos de restricción:



Multiset de fragmentos de restricción

Multiset: {3, 5, 5, 8, 9, 14, 14, 17, 19, 22}



Partial Digest

Sea X: el conjunto de los **n** enteros que representan la locacion de todos los cortes en el mapa de restricción incluyendo el del comienzo y el final. Ej: X= {0, 2, 4, 7, 10}

Sea n: el numero de cortes total Sea DX: el multiset de enteros que representan las longitudes de cada uno de los fragmentos producidos por un partial digest.

X	О	2	4	7	10
0		2	4	7	10
2			2	5	8
4				3	6
7					3
10					

Multiset DX: {2, 2, 3, 3, 4, 5, 6, 7, 8, 10}

El problema del Partial Digest

Sea X: el conjunto de los n enteros				•	-10
que representan la locacion de todos	0	2	4	7	10
los cortes en el mapa de restricción incluyendo el del comienzo y el final.	2		2	5	8
Ej: X= {0, 2, 4, 7, 10}	4			3	6
Sea n: el numero de cortes total Sea DX: el multiset de enteros que	7				3
representan las longitudes de cada	10				

10 |

Multiset DX: {2, 2, 3, 3, 4, 5, 6, 7, 8, 10}

Dadas todas las longitudes de los segmentos de restricción – n(n-1)/2 - enteros, Determinar el conjunto X de enteros que corresponden a la locacion de los cortes en el mapa de restricción

uno de los fragmentos producidos

por un partial digest.

El problema del Partial Digest

No siempre es posible reconstruir un conjunto X si nos basamos solamente en DX

```
Ej: X = \{0, 2, 5\} y X + 10 = \{10, 12, 15\}
Ambos producen DX = \{2, 3, 5\}
```

Otro ejemplo. Menos trivial:

```
{0,1,2,5,7,9,12} y {0,1,5,7,8,10,12}. Ambos forman el partial digest: {1,1,2,2,2,3,3,4,4,5,5,5,6,7,7,7,8,9,10,11,12}
```

El problema del Partial Digest

Otro ejemplo. Menos trivial:

{0,1,2,5,7,9,12} y {0,1,5,7,8,10,12}. Ambos forman el partial digest:

{1,1,2,2,2,3,3,4,4,5,5,5,6,7,7,7,8,9,10,11,12}

	0	1	2	5	7	9	12
0		1	2	5	7	9	12
1			1	4	6	8	11
2				3	5	7	10
5					2	4	7
7						2	5
9							3
12							

	0	1	5	7	8	10	12
0		1	5	7	8	10	12
1			4	6	7	9	11
5				2	3	5	7
7					1	3	5
8						2	4
10							2
12							

El problema del Partial Digest : Ejemplo

Dado el set de distancias de fragmentos de restriccion:

$$L = \{2,2,3,3,4,5,6,7,8,10\}$$

Encontrar el Mapa con los puntos de corte. Empezamos con

$$X = \{0\}$$

El problema del Partial Digest : Ejemplo

```
Dado L = \{2,2,3,3,4,5,6,7,8,10\}
Encontrar X, Empezamos con X=\{0\}
```

Removemos 10 de L y lo insertamos en X, por ser el fragmento más largo.

```
L = \{2,2,3,3,4,5,6,7,8,10\} X = \{0,10\}
```

0

El problema del Partial Digest : Ejemplo

$$L = \{2,2,3,3,4,5,6,7,8,10\}$$

 $X=\{0,10\}$

Tomamos 8 de L y verificamos sus distancias (\mathbf{y}) a los elementos que ya estan en X (Puede ser 10-8=2 u 8-0=8). Como los dos casos son simetricos asumimos \mathbf{y} =2.

Las distancias de y a los elementos de X: $D(y,X) = \{8,2\}$ Entonces removemos $\{8,2\}$ de L y añadimos 2 a X

$$L = \{2,2,3,3,4,5,6,7,8,10\}$$
$$X = \{0,2,10\}$$

El problema del Partial Digest : Ejemplo

```
L = \{2,2,3,3,4,5,6,7,8,10\}
X=\{0,2,10\}
Tomamos 7 de L y calculamos las distancias a los elementos de X: D(\mathbf{y},X)=\{7,5,3\}
Entonces removemos \{7,5,3\} de L y añadimos 7 a X
```

$$L = \{2,2,3,3,4,5,6,7,8,10\}$$
$$X = \{0,2,7,10\}$$



El problema del Partial Digest : Ejemplo

```
L = \{2,2,3,3,4,5,6,7,8,10\}X=\{0,2,7,10\}
```

Tomamos 6 y calculamos las distancias: $D(y,X)=\{6,4,1,4\}$. Como estas distancias no son un subconjunto de L, no se toman en cuenta.

Tomamos 4 Y calculamos las distancias: $D(y,X)=\{4,2,3,6\}$.

Entonces removemos {4,2,3,6} de L y añadimos 4 a X:

$$L = \{2,2,3,3,4,5,6,7,8,10\}$$

$$X = \{0,2,4,7,10\}$$

Como L queda vacia, es una solución encontrada.

Otras posibles soluciones se pueden encontrar haciendo back tracking, explorando otras opciones en pasos anteriores p.ej: en vez de escoger 4 escogemos 3 de L

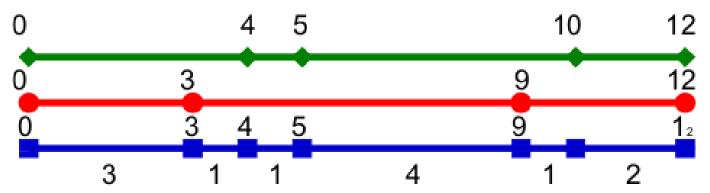


El problema del Double Digest

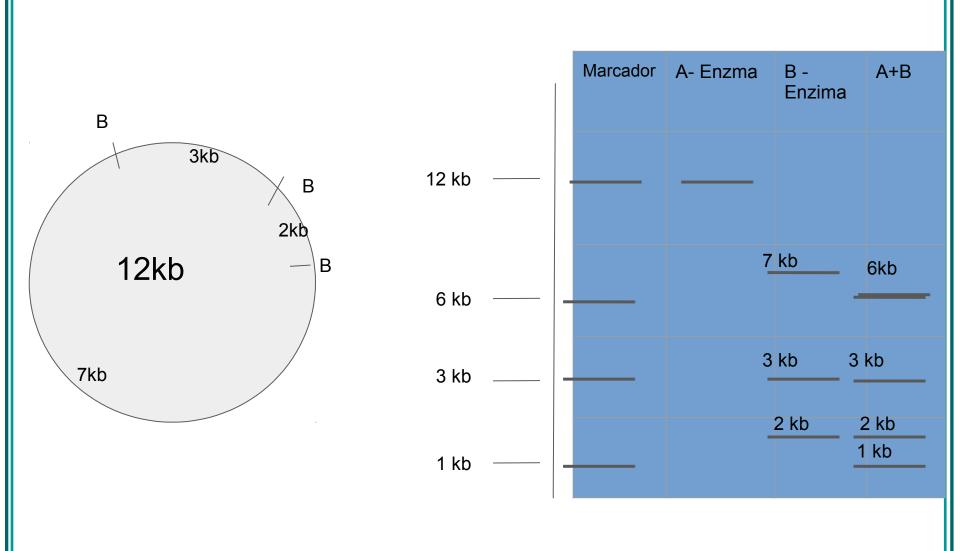
Un método alternativo es el método del Double Digest.

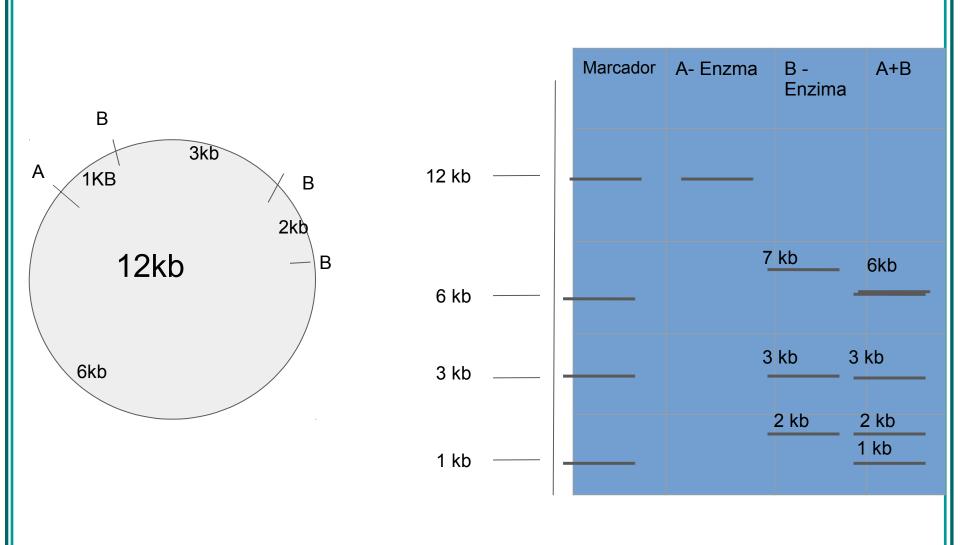
Consiste en utilizar 2 enzimas en tres pasos.

- 1° se hace el Partial digest con la enzima de restricción A
- 2° se hace el Partial digest con la enzima de restricción B
- 3° se hace el Partial digest con ambas enzimas A y B.



Las entradas son las longitudes de 3 conjuntos de fragmentos de restricción [1,2,4,5], [3,3,6],[1,1,1,2,3,4]





Problema: Un fragmento de ADN se corta con PstI y HindIII por separado.

Posteriormente, se utiliza una mezcla de ambas enzimas obteniéndose los fragmentos indicados a continuación:

Pstl: 3Kb y 4Kb

HindIII: 3Kb y 4Kb

PstI+HindIII: 1Kb, 2Kb

Dibujar el mapa de restricción de este segmento de ADN

Problema: Un fragmento de ADN se corta con HindIII y Smal por separado.

Posteriormente, se utiliza una mezcla de ambas enzimas obteniéndose los fragmentos indicados a continuación:

HindIII: 1Kb, 2Kb, 4Kb, 5Kb

Smal: 4Kb y 6Kb

Smal+HindIII: 1Kb, 2Kb, 3Kb

Dibujar el mapa de restricción de este segmento de ADN

HindIII: 1Kb, 2Kb, 4Kb, 5Kb

Smal: 4Kb y 6Kb

Smal+HindIII: 1Kb, 2Kb, 3Kb

