

Comparación de dos secuencias El alineamiento

Bioinformática
Máster en Biotecnología

¿Para que comparar dos secuencias?

La comparación de dos secuencias mediante su alineamiento es uno de los pilares de la bioinformática y es la base de muchos métodos diferentes

1) Comparación de secuencias homólogas para hacer un análisis evolutivo

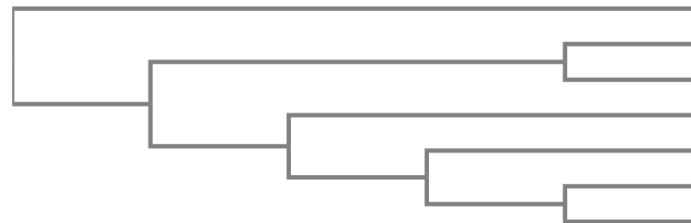
- Detectar regiones conservadas
- Determinar el grado de parentesco entre varias secuencias (y de las especies que las portan)

```
INS_chicken  MALWIRSLPLLALLVFSGPGTSYAAANQHLCGSHLVEALYLVCGERGFFYSPKARRDVEQ
INS2_mouse   MALWMRFLPLLALLFLWESHPTQAFVKQHLGSHLVEALYLVCGERGFFYTPMSRREVED
INS2_rat     MALWIRFLPLLALLILWEPRPAQAFVKQHLGSHLVEALYLVCGERGFFYTPMSRREVED
INS_cow      MALWTRLAPLLALLALWAPAPARAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEG
INS_dog      MALWMRLLPLLALLALWAPAPTRAFVNQHLCGSHLVEALYLVCGERGFFYTPKARREVED
INS_human    MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED
INS_chimp    MALWMRLLPLLALLALWGPDPASAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAED
**** *  **.* :      : * .:*****:*****:* :*:.*
```

```
INS_chicken  PLVSSPL-RGEAGV--LPFQQEYKVKRGIVEQCCHNTCSLYQLENYCN
INS2_mouse   PQVAQLELGGGPGAGDLQTLALEVAQKRGIVDQCCTSICSLYQLENYCN
INS2_rat     PQVAQLELGGGPGAGDLQTLALEVARQKRGIVDQCCTSICSLYQLENYCN
INS_cow      PQVGALELAGGPGAGGL----EGPPQKRGIVEQCCASVCSLYQLENYCN
INS_dog      LQVRDVELAGAPGEGGLQPLALEGALQKRGIVEQCCTSICSLYQLENYCN
INS_human    LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
INS_chimp    LQVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN
*          * * * * *      *****:*** . *****
```

Regiones conservadas

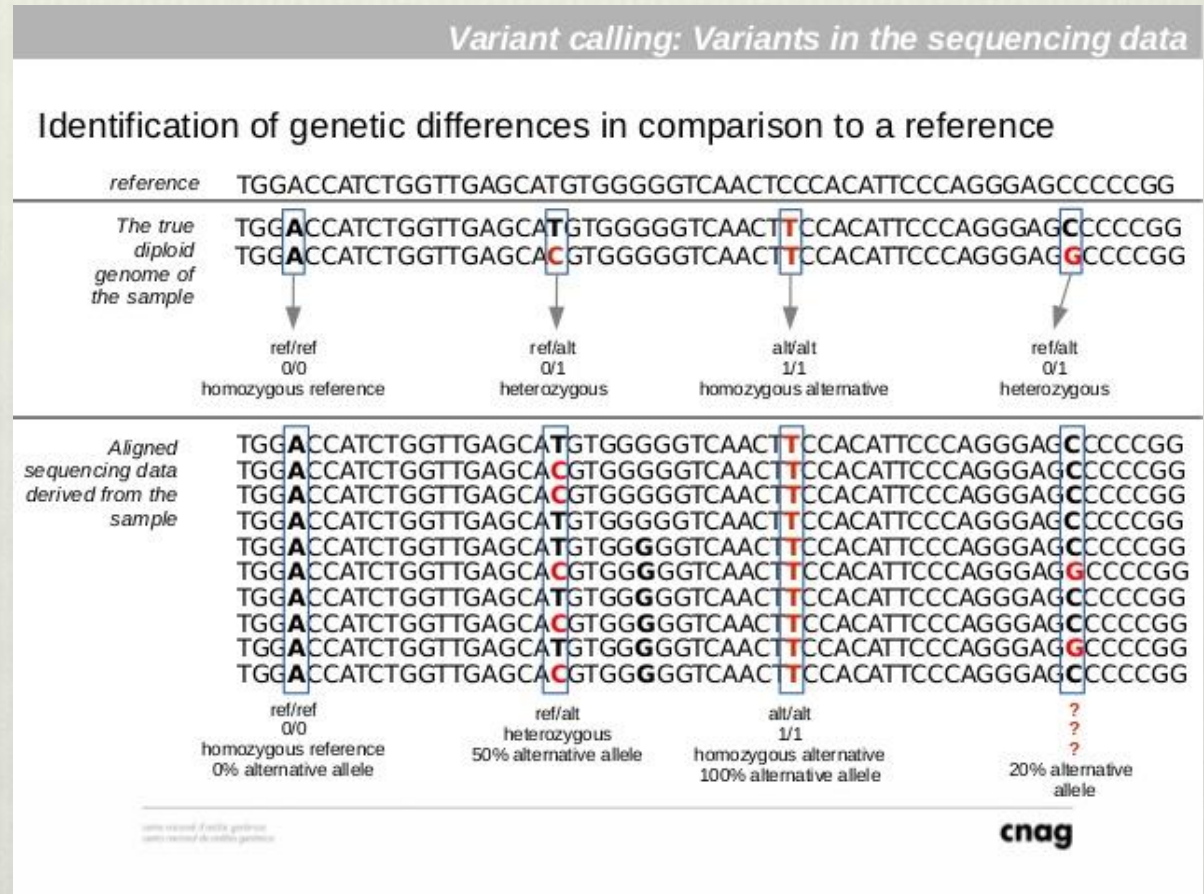
Árbol filogenético



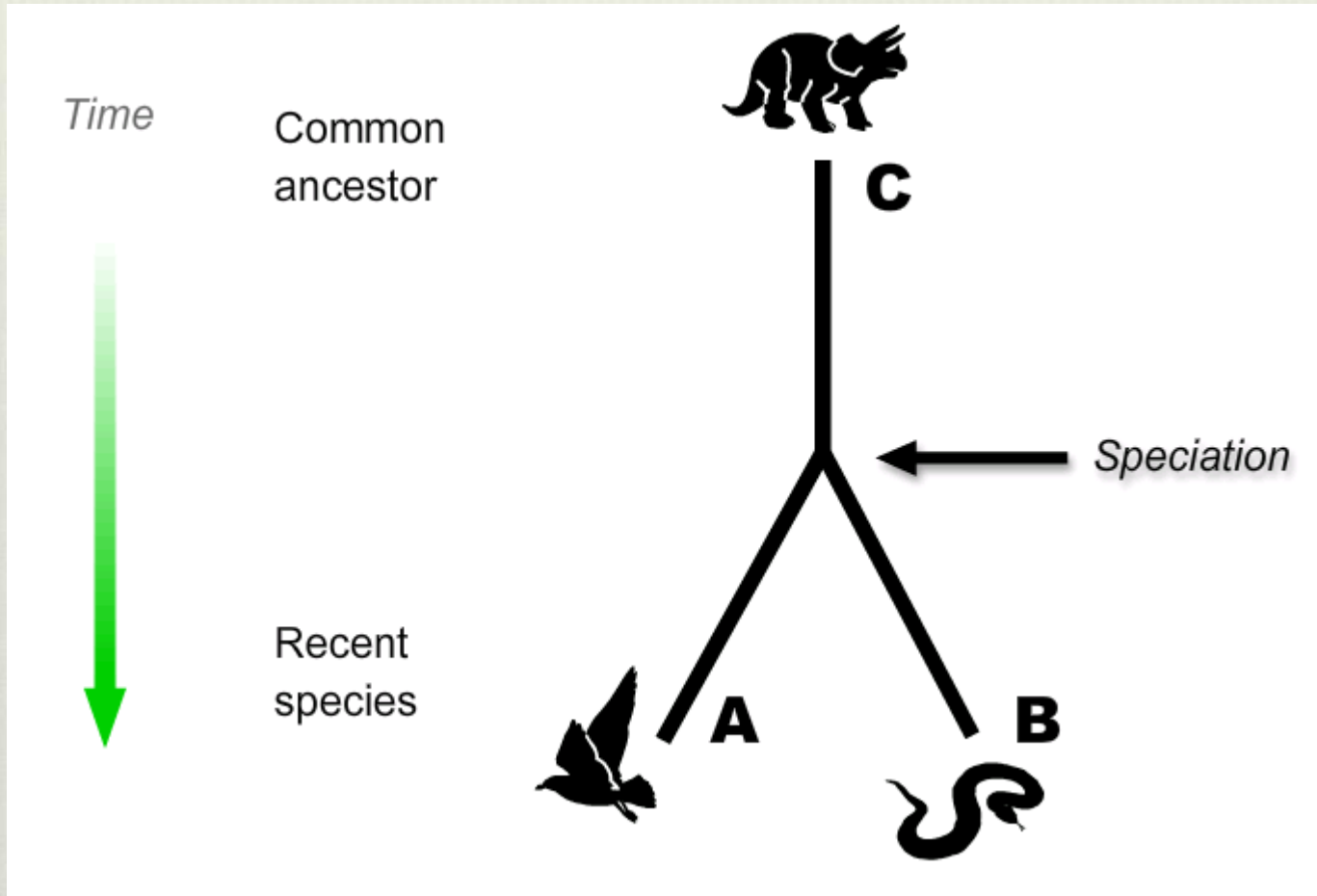
INS_chicken 0.215565
INS2_mouse 0.0306801
INS2_rat 0.0306801
INS_cow 0.0862504
INS_dog 0.067093
INS_human 0.00994053
INS_chimp 0.00994053

¿Para que comparar dos secuencias?

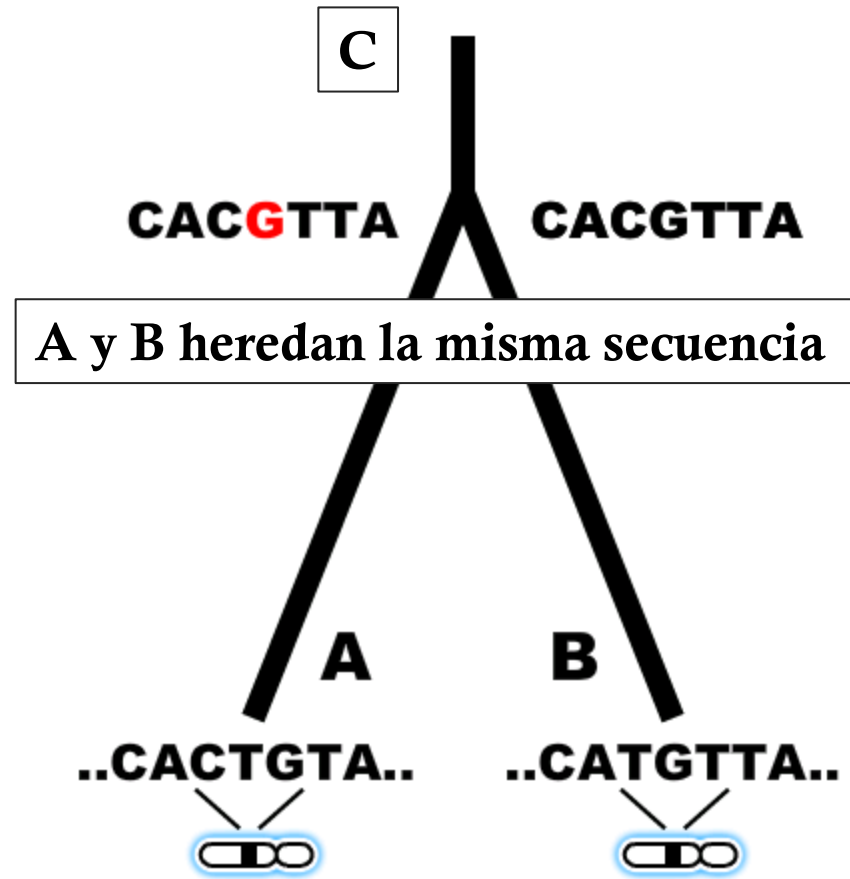
5) Resequenciar un genoma para encontrar variantes de secuencia y cambios estructurales



El alineamiento en un contexto evolutivo



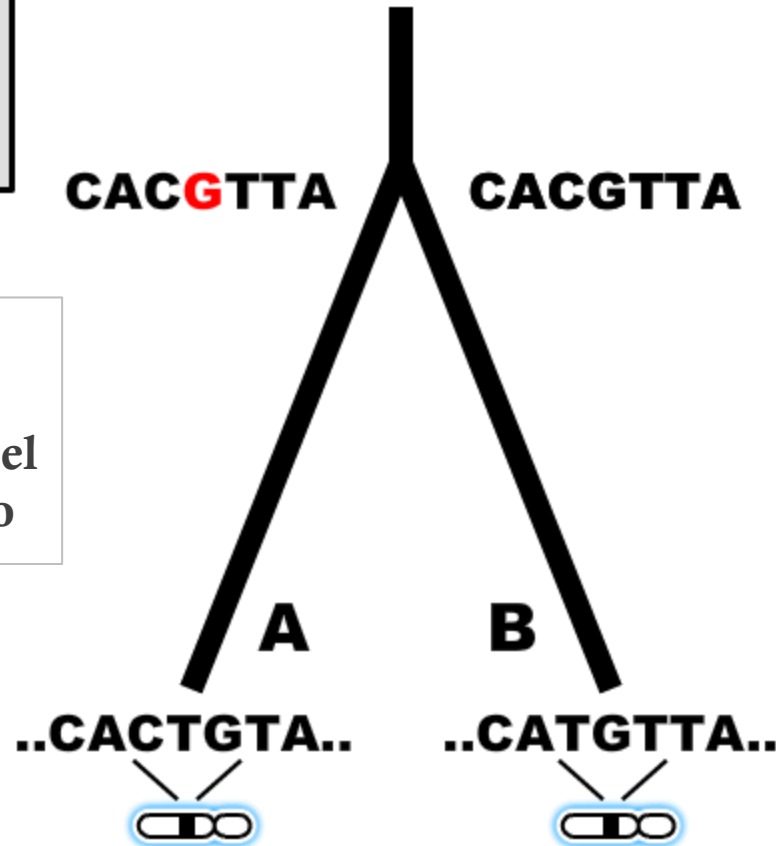
Cambios evolutivos y el alineamiento



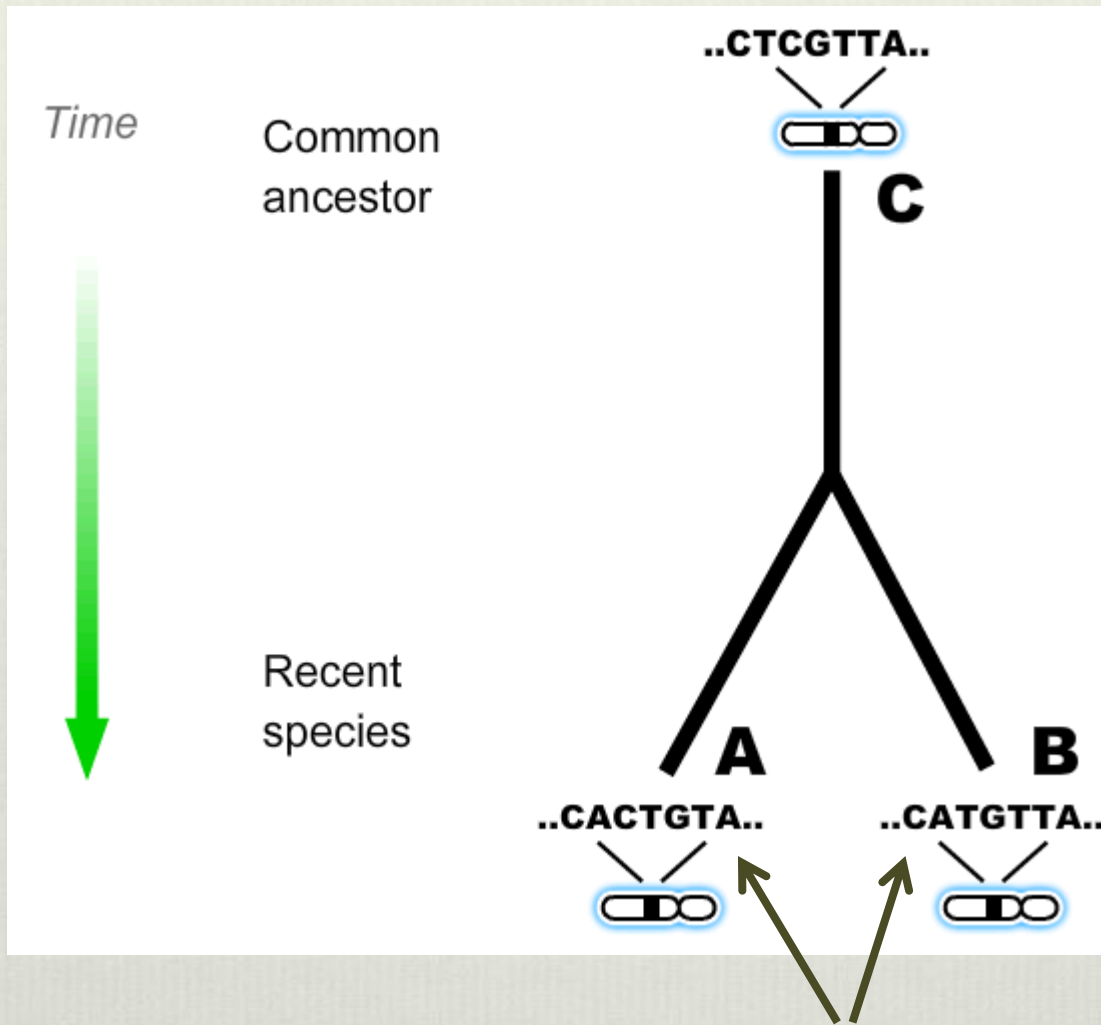
Cambios evolutivos y el alineamiento



En el momento de la especiación, las dos secuencias son idénticas y el 'alineamiento' es perfecto



El alineamiento en un contexto evolutivo

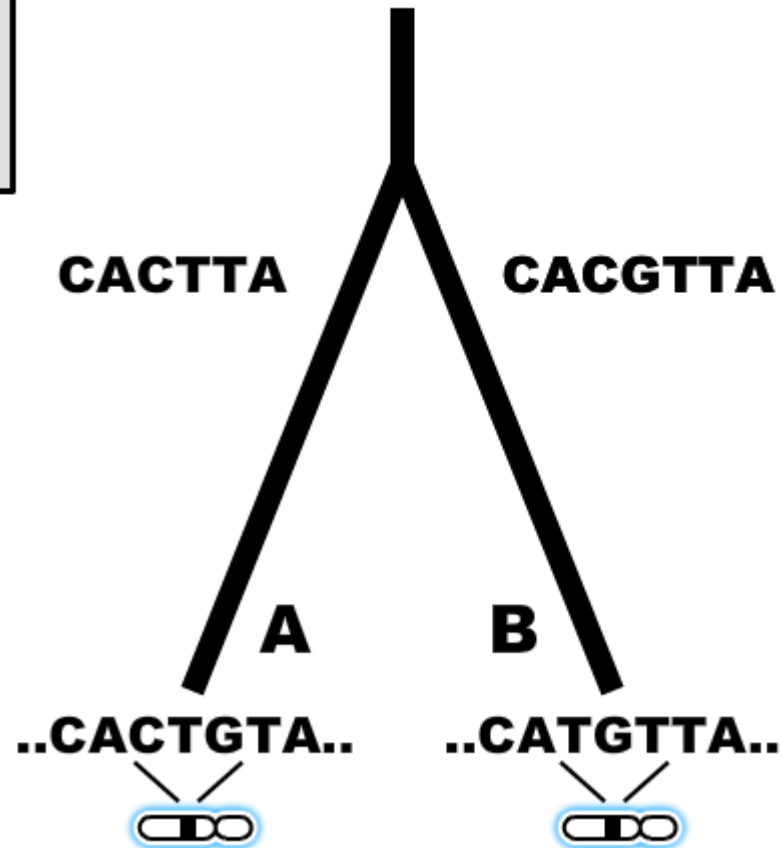


Estas dos secuencias son similares debido a que derivan de un ancestro común y no debido al azar

Cambios evolutivos y el alineamiento



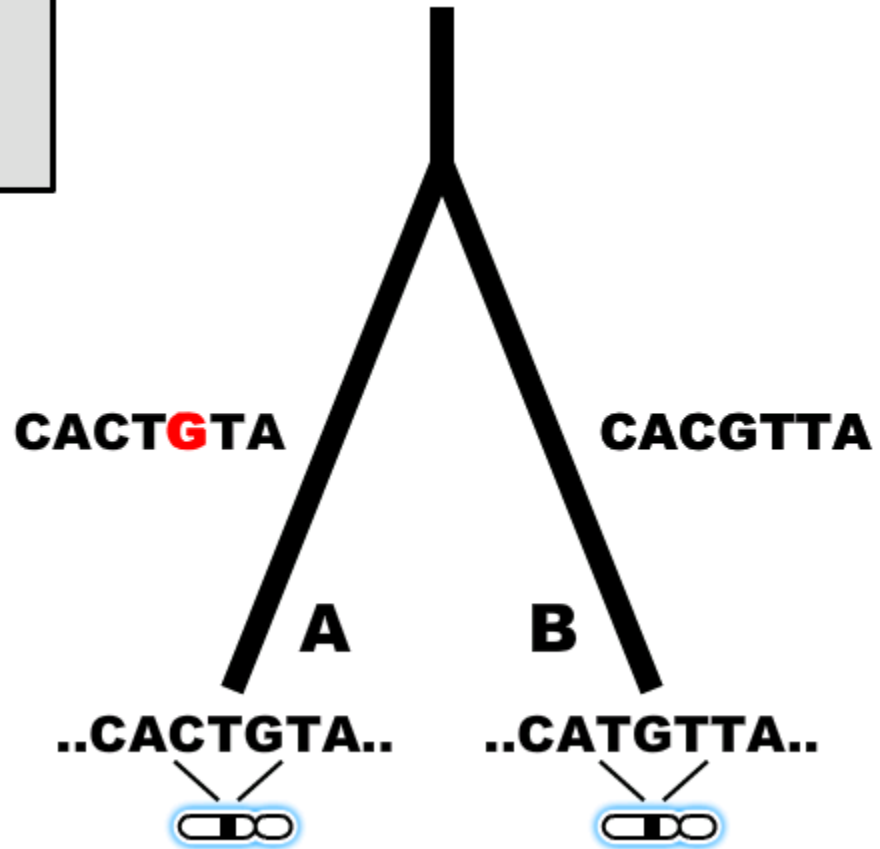
Deletion



Cambios evolutivos y el alineamiento

Sequence 1
CAC-TGTA
Sequence 2
CACGT-TA

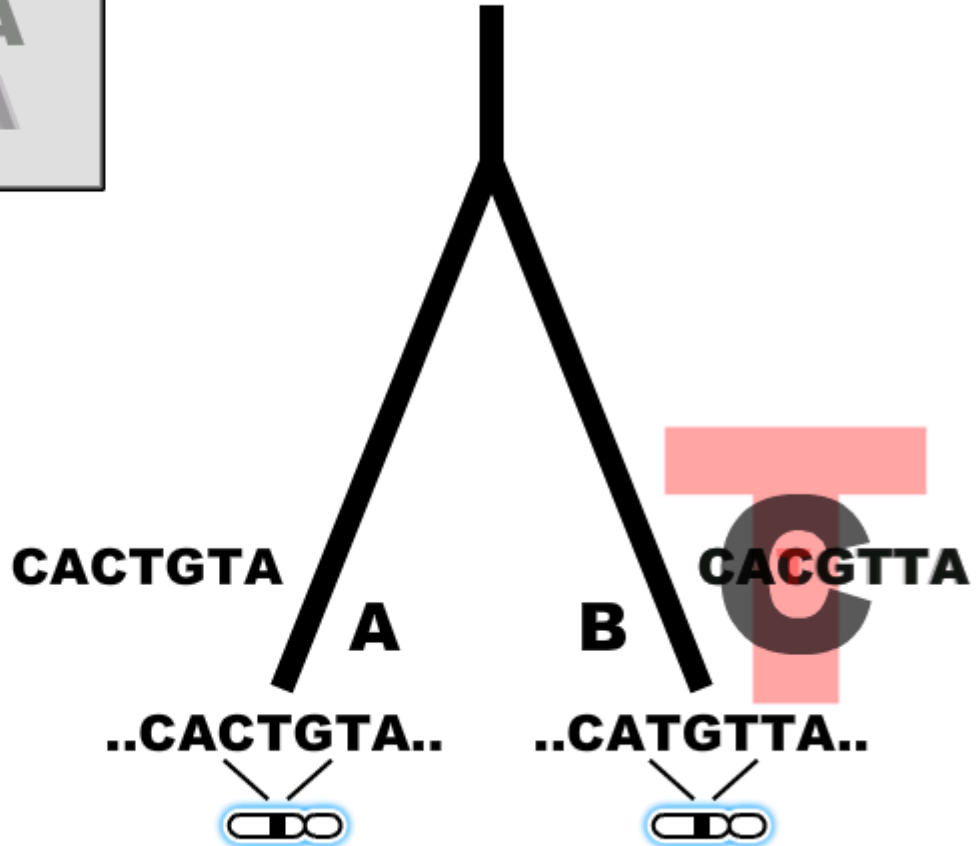
Insertion



Cambios evolutivos y el alineamiento

Sequence 1
CAC-TGTA
CAC**CGT**-TA
Sequence 2

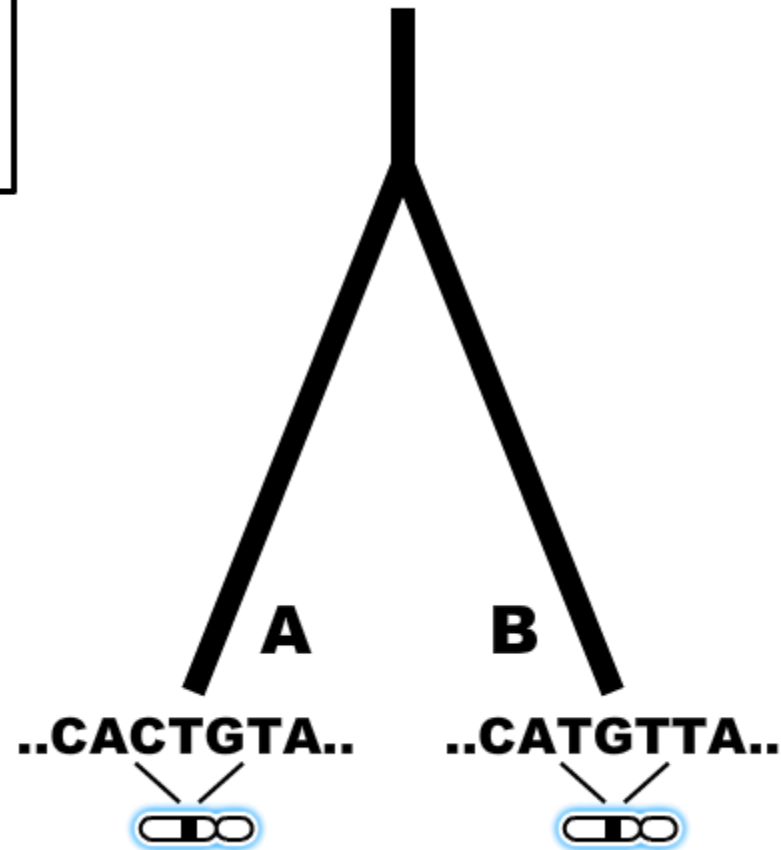
Substitution



Cambios evolutivos y el alineamiento

Sequence 1
CAC-TGTA
CATGT-TA
Sequence 2

Este alineamiento refleja
la 'historia' evolutiva de
las dos secuencias



Cambios evolutivos y el alineamiento

Original alignment

CAC-TGTA
CATGT-TA

Alineamiento
correcto

CACTGTA
CATGTTA

Alternative
alignments

CACTGT-A
CA-TGTTA

OBJETIVO:

Encontrar el alineamiento
que refleja la 'verdadera'
historia evolutiva de las
secuencias

Puntuar los
alineamientos con el
esquema de puntuación
adecuado

Puntuar el alineamiento

Punto de partida:
Dos o mas secuencias

Sec1	ACGTATAGCG
Sec2	ACGTAGCG



Alinear las
secuencias

ACGTATAGCG
ACGTA--GCG

Dos huecos/
gaps

ACGTATAGCG
ACGTAGCG

Dos desemparejamientos



¿Cual de los dos
alineamientos es mas
probable?

Puntuar el alineamiento

```
ACGTATAGCG
|||||  |||
ACGTA--GCG
```

```
ACGTATAGCG
|||||  |
ACGTAGCG
```

Puntuar los alineamientos:

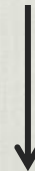
Emparejamiento: +1

Desemparejamiento: -1

Hueco: -5

Puntuación
/Score: 3

Puntuación
/Score: 4



El alineamiento mas probable dado el sistema de puntuación

Puntuar el alineamiento

```
ACGTATAGCG
|||||  |||
ACGTA--GCG
```

```
ACGTATAGCG
|||||  |
ACGTAGCG
```

Puntuar los alineamientos:

Emparejamiento: +1

Desemparejamiento: -2

Hueco: -5

Puntuación
/Score:

Puntuación
/Score:

Puntuar el alineamiento

```
ACGTATAGCG
|||||  |||
ACGTA--GCG
```

```
ACGTATAGCG
|||||  |
ACGTAGCG
```

Puntuar los alineamientos:

Emparejamiento: +1

Desemparejamiento: -2

Hueco: -5

Puntuación
/Score: 3

Puntuación
/Score: 2



El alineamiento mas probable dado el sistema de puntuación

Puntuar el alineamiento

ACGTATAGCG

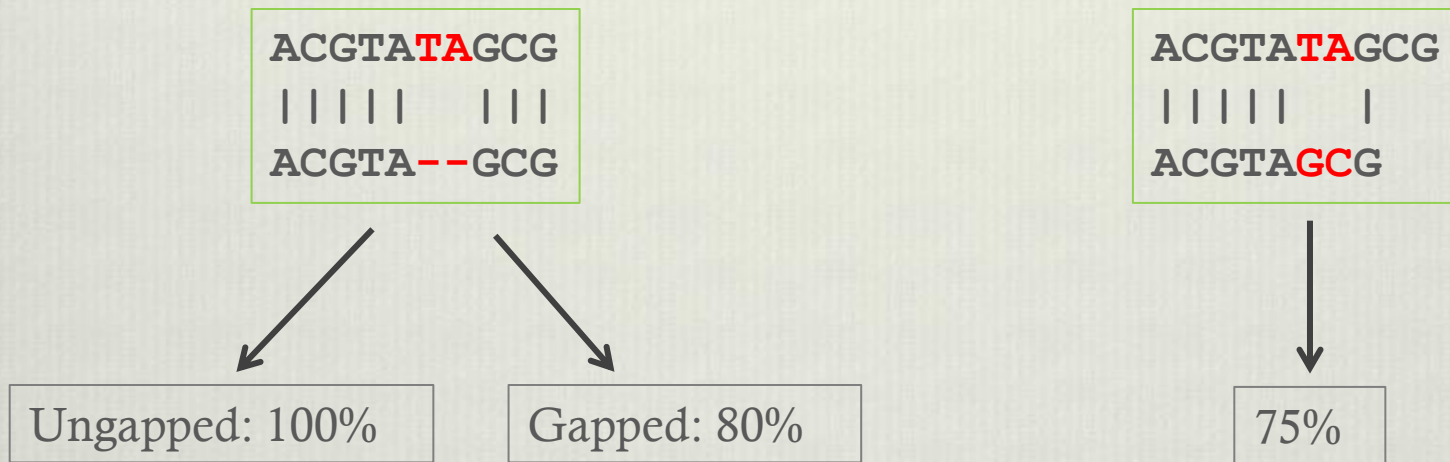
ACGTATAGCG

¡¡El alineamiento mas probable depende del sistema de puntuación utilizado!!

El esquema de puntuación elegido depende del conocimiento previo que tengamos de nuestras secuencias y/o la finalidad concreta del alineamiento

Similitud

Similitud de secuencia = número de coincidencias / longitud del alineamiento



Similitud y homología



La similitud de secuencia **NO ES LO MISMO** que homología

Similitud de secuencia:

grado de similitud que hay entre dos secuencias

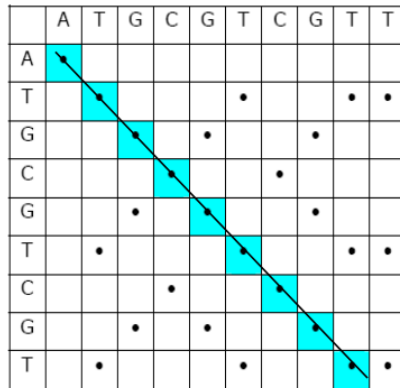
Homología:

las dos secuencias derivan de la misma secuencia ancestral presente en el ancestro común

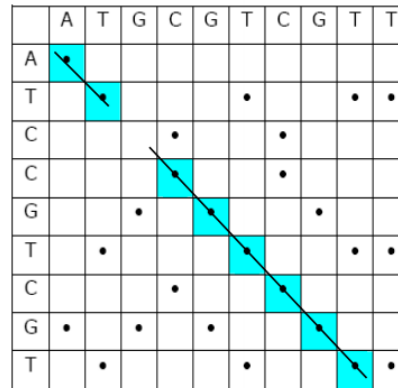
Si observamos un alto grado de similitud de secuencia podemos deducir que las dos secuencias son homólogas
(son similares debido a la procedencia común y no debido al azar)

La matriz de puntos

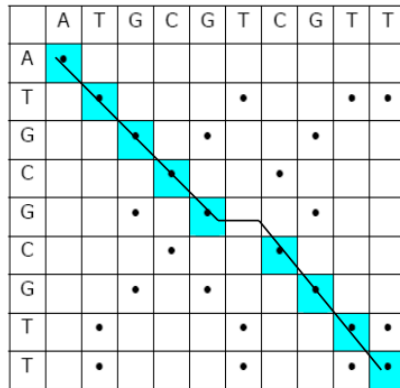
a)



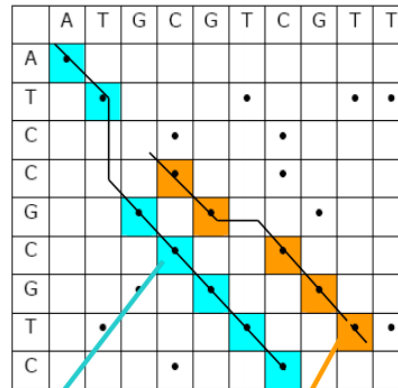
b)



c)



d)



Alineamiento 1:

AT--GCGTCGTT
ATCCGCGTC---

Alineamiento 2:

ATCGTCGTT
ATCG-CGTT

- Las dos secuencias son idénticas en la parte alineada.
- Las dos secuencias muestran un desemparejamiento debido a una sustitución; la posición (3,3) se queda en blanco.
- Las dos secuencias difieren por una inserción/delección (indel), dando lugar a un hueco o gap; nótese el quiebro o zig-zag de la diagonal principal.
- Dos posibles alineamientos mostrando desemparejamientos y huecos. El alineamiento 1 supondría en total cinco huecos (o un hueco de dos nucleótidos y otro hueco terminal de tres nucleótidos) y ningún desemparejamiento, mientras que el alineamiento 2 supondría un hueco y dos desemparejamientos.

Matriz de puntuación (BLOSUM62)

Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4
Ala		Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val

Matriz de puntuación (PAM250)

PAM = Point Accepted Mutations

PAM 250	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	
A	2	-2	0	0	-2	0	0	1	-1	-1	-2	-1	-1	-3	1	1	1	-6	-3	0	0	0	0	0
R	-2	6	0	-1	-4	1	-1	-3	2	-2	-3	3	0	-4	0	0	-1	2	-4	-2	-1	0	0	-1
N	0	0	2	2	-4	1	1	0	2	-2	-3	1	-2	-3	0	1	0	-4	-2	-2	2	1	0	0
D	0	-1	2	4	-5	2	3	1	1	-2	-4	0	-3	-6	-1	0	0	-7	-4	-2	3	3	-1	-1
C	-2	-4	-4	-5	12	-5	-5	-3	-3	-2	-6	-5	-5	-4	-3	0	-2	-8	0	-2	-4	-5	-3	-3
Q	0	1	1	2	-5	4	2	-1	3	-2	-2	1	-1	-5	0	-1	-1	-5	-4	-2	1	3	-1	-1
E	0	-1	1	3	-5	2	4	0	1	-2	-3	0	-2	-5	-1	0	0	-7	-4	-2	3	3	-1	-1
G	1	-3	0	1	-3	-1	0	5	-2	-3	-4	-2	-3	-5	0	1	0	-7	-5	-1	0	0	-1	-1
H	-1	2	2	1	-3	3	1	-2	6	-2	-2	0	-2	-2	0	-1	-1	-3	0	-2	1	2	-1	-1
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5	2	-2	2	1	-2	-1	0	-5	-1	4	-2	-2	-1	-1
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6	-3	4	2	-3	-3	-2	-2	-1	2	-3	-3	-1	-1
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5	0	-5	-1	0	0	-3	-4	-2	1	0	-1	-1
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6	0	-2	-2	-1	-4	-2	2	-2	-2	-1	-1
F	-3	-4	-3	-6	-4	-5	-5	-5	-2	1	2	-5	0	9	-5	-3	-3	0	7	-1	-4	-5	-2	-2
P	1	0	0	-1	-3	0	-1	0	0	-2	-3	-1	-2	-5	6	1	0	-6	-5	-1	-1	0	-1	-1
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	2	1	-2	-3	-1	0	0	0	0
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-3	0	1	3	-5	-3	0	0	-1	0	0
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17	0	-6	-5	-6	-4	-4
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	-2	-3	-4	-2	-2
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4	-2	-2	-1	-1
B	0	-1	2	3	-4	1	3	0	1	-2	-3	1	-2	-4	-1	0	0	-5	-3	-2	3	2	-1	-1
Z	0	0	1	3	-5	3	3	0	2	-2	-3	0	-2	-5	0	0	-1	-6	-4	-2	2	3	-1	-1
X	0	-1	0	-1	-3	-1	-1	-1	-1	-1	-1	-1	-1	-2	-1	0	0	-4	-2	-1	-1	-1	-1	-1

Nomenclatura

One letter code	Three letter code	Amino acid	Possible codons
A	Ala	Alanine	GCA, GCC, GCG, GCT
B	Asx	Asparagine or Aspartic acid	AAC, AAT, GAC, GAT
C	Cys	Cysteine	TGC, TGT
D	Asp	Aspartic acid	GAC, GAT
E	Glu	Glutamic acid	GAA, GAG
F	Phe	Phenylalanine	TTC, TTT
G	Gly	Glycine	GGA, GGC, GGG, GGT
H	His	Histidine	CAC, CAT
I	Ile	Isoleucine	ATA, ATC, ATT
K	Lys	Lysine	AAA, AAG
L	Leu	Leucine	CTA, CTC, CTG, CTT, TTA, TTG
M	Met	Methionine	ATG
N	Asn	Asparagine	AAC, AAT
P	Pro	Proline	CCA, CCC, CCG, CCT
Q	Gln	Glutamine	CAA, CAG
R	Arg	Arginine	AGA, AGG, CGA, CGC, CGG, CGT
S	Ser	Serine	AGC, AGT, TCA, TCC, TCG, TCT
T	Thr	Threonine	ACA, ACC, ACG, ACT
V	Val	Valine	GTA, GTC, GTG, GTT
W	Trp	Tryptophan	TGG
X	X	any codon	NNN
Y	Tyr	Tyrosine	TAC, TAT
Z	Glx	Glutamine or Glutamic acid	CAA, CAG, GAA, GAG
*	*	stop codon	TAA, TAG, TGA

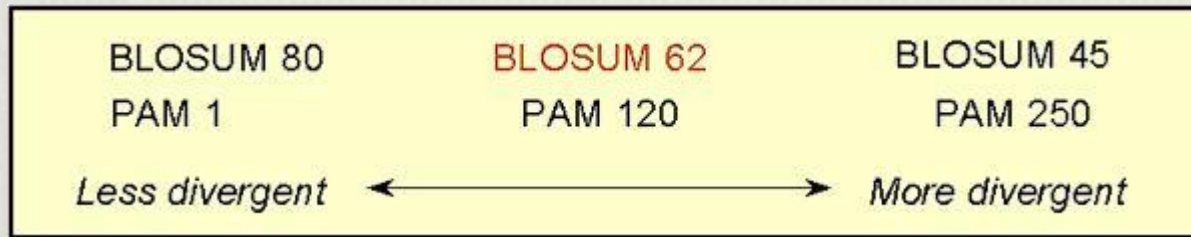
Propiedades de los aminoácidos

Property		Amino acids
small		Ala, Gly
acidic / amide		Asp, Glu, Asn, Gln
charged	negative	Asp, Glu
	positive	Lys, Arg
polar		Ala, Gly, Ser, Thr, Pro
hydrophobic		Val, Leu, Ile, Met
size	big	Glu, Gln, His, Ile, Lys, Leu, Met, Phe, Trp, Tyr
	small	Ala, Asn, Asp, Cys, Gly, Pro, Ser, Thr, Val
aliphatic		Ile, Leu, Val
aromatic		His, Phe, Tyr, Trp

Comparación de las matrices

PAM1: 1% de los residuos han cambiado (secuencias altamente relacionadas)

PAM250: Sustituciones múltiples (2.5 por residuo) para secuencias remotamente relacionadas



Percent Accepted Mutation PAM	Blocks Substitution Matrix
Based on explicit evolutionary model	Based on empirical frequencies
Represents a specific evolutionary distance	Always a blend of distances as seen in the database and PROSITE
Ranges from identical to completely random	Narrower range than PAM matrix

Tipos de alineamientos

Global: se fuerza un alineamiento de las dos secuencias en su longitud completa

Local: búsqueda de subsecuencias con alto grado de similitud entre dos secuencias mas largas que pueden estar muy divergidas

Exacto: Un método exacto nos garantiza encontrar el mejor alineamiento entre dos secuencias (Smith-Waterman, Needleman- Wunsch): **LENTO**

Heurístico: Solución aproximada; no garantiza encontrar el mejor alineamiento (BLAST; BLAT): **MAS RAPIDO**

```
Global FTFTALILLAVAV
      F--TAL-LLA-AV
```

```
Local FTFTALILL-AVAV
     --FTAL-LLAAV--
```

Illustration of global and local alignments demonstrating the 'gappy' quality of global alignments that can occur if sequences are insufficiently similar

```
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:   90/149 (60.4%)
# Gaps:         9/149 ( 6.0%)
# Score: 292.5
#
#
#=====
HBA_HUMAN      1 MV-LSPADKTNVKAANGKVGAGHAGEYGAELERMFLSFPTTKTYFPHF-D    48
                || |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
HBB_HUMAN      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD    48
```

Secuencias de DNA: alfabeto y formato

A = adenina

C = citosina

G = guanina

T = timina

R = G A (purina)

Y = T C (pirimidina - *pyrimidine*)

K = G T (ceto, en inglés *keto*)

M = A C (amino)

S = G C (enlaces fuertes, en inglés *strong*)

W = A T (enlaces débiles, en inglés *weak*)

B = G T C (cualquiera excepto A)

D = G A T (cualquiera excepto C)

H = A C T (cualquiera excepto G)

V = G C A (cualquiera excepto T)

N = A G C T (cualquiera)

ID / nombre

Descripción

>gi|4558520|gb|AF033819.3|

HIV-1, complete genome

```
GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCC
TCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGA
TCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACCTGAAAGCGAA
AGGGAACCCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGG
CGGCGACTGGTGAGTACGCCAAAAATTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCG
TCAGTATTAAGCGGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAAGAAAAAAT
ATAAATTAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTTCGAGTTAATCCTGGCCTGTTAGA
AACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAATT
AGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAGG
AAGCTTTAGACAAGATAGAGGAAGAGCAAAAACAAAAGTAAGAAAAAAGCACAGCAAGCAGCAGCTGACAC
AGGACACAGCAATCAGGTGAGCCAAAATACCCTATAGTGCAGAATCCAGGGGCAAATGGTACATCAG
GCCATATCACCTAGAACTTTAAATGCATGGGTAAGTAGTAGAAGAGAAGGCTTTTCAGCCCAGAAGTGA
TACCCATGTTTTTCAGCATTATCAGAAGGAGCCACCCACAAGATTTAAACACCATGCTAAACACAGTGGG
GGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCAATGAGGAAGCTGCAGAATGGGATAGAGTG
CATCCAGTGCATGCAGGGCCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAAGTGACATAGCAGGAA
CTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCCCAGTAGGAGAAATTTA
TAAAAGATGGATAATCCTGGGATTAATAAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATA
AGACAAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAGCAAG
CTTCACAGGAGGTAAAAAATGGATGACAGAAACCTTGTGGTCCAAAATGCGAACCCAGATTGTAAGAC
TATTTTAAAAGCATTGGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGA
CCCGGCCATAAGGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATTCAGCTACCATAATGATGC
AGAGAGGCAATTTTAGGAACCAAAGAAAGATTGTTAAGTGTTCATTGTGGCAAAGAAGGGCACACAGC
CAGAAATTCAGGGCCCTAGGAAAAAGGGCTGTTGAAATGTGAAAGGAAGGACACCAAATGAAAGAT
TGTACTGAGAGACAGGCTAATTTTTTAGGGAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTC
TTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTCTGGGGTAGAGACAACAATCC
CCCTCAGAAGCAGGAGCCGATAGACAAGGAAGTATCCTTTAACTCCCTCAGGTCACTCTTTGGCAAC
GACCCCTCGTCACAATAAAGATAGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATAC
AGTATTAGAAGAAATGAGTTTGCCAGGAAGATGGAACCAAAAATGATAGGGGAATTGGAGTTTTATC
AAAGTAAGACAGTATGATCAGATACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAG
GACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGCACCTTTAAATTTTCCAT
```


Secuencias de proteínas: alfabeto y formato

One letter code	Three letter code	Amino acid
A	Ala	Alanine
B	Asx	Asparagine or Aspartic acid
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
X	X	any codon
Y	Tyr	Tyrosine
Z	Glx	Glutamine or Glutamic acid
*	*	stop codon

```
>gi|28872819|ref|NP_057849.4| Gag-Pol [Human immunodeficiency virus 1]
MGARASVLSGGELDRWEKIRLRPGGKKKYKLVKHIWASRELERFAVNPGLLETSEGCRQILGQLQPSLQT
GSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAADTGHNSNQVSQNYPIVQNIQG
QMVHQAISPRTLNAAVVKVVEEKAFSPEVIMPFSAALSEGATPQDLNNTMLNTVGGHQAAMQMLKETINEEAA
EWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNPPPIPVGEIYKRWIILGLNKKIVRMYSP
SILDIRQGPKPEFRDYVDRFYKTLRAEQASQEVKNWMTETLLVQANANPDCKTILKALGPAATLEEMMTAC
QGVGGPGHKARVLAEAMSQVNTSATIMMQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGKEG
HQMKDCTERQANFLREDLAFLLQGKAREFSSEQTRANSPTRRELQVWGRDNNSPSEAGADRQGTVSFNFPQ
VTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWPKPKMIGGIGGFIVRQYDQILIEICGHKAI
GTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPMDGPKVKQWPLTEEKIKALVEICTEMEK
EGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGD
AYFSVPLDEDFRKYTAFTIPI SINNETPGIRYQYNVLPQGWKGSFAIFQSSMTKILEPFRKQNPDIVIYQY
MDDLIVGSDLEIGQHRKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPKDSWT
VNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEALELELAENREILKEPVHGVYDPSK
DLIAEIQKQGQGWTYQIYQEPFKNLKTGKYARMGAHTNDVKQLTEAVQKITTESIWIWGTKPKFKLPI
QKETWETWWTEYWQATWIPEWEFVNTPLVVLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTNRGR
QKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSESELVNIIEQLIKKEKVYL
AWVPAHKGIGGNEQVDKLVSAGIRKVLFLDGDIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKC
QLKGEAMHGQVDCSPGIWQLDCTHLEGKVIIVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKT
HTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIH
NFKRKGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRNPWLKGPALKLLWKGEAVVIQD
NSDIKVVPRRKAKIIRDYGKQMGDDCVASRQDED
```


Coordenadas del alineamiento
relativas a la secuencia de
entrada

Coordenadas cromosómicas
del alineamiento

Score/puntuación

Similitud de secuencia

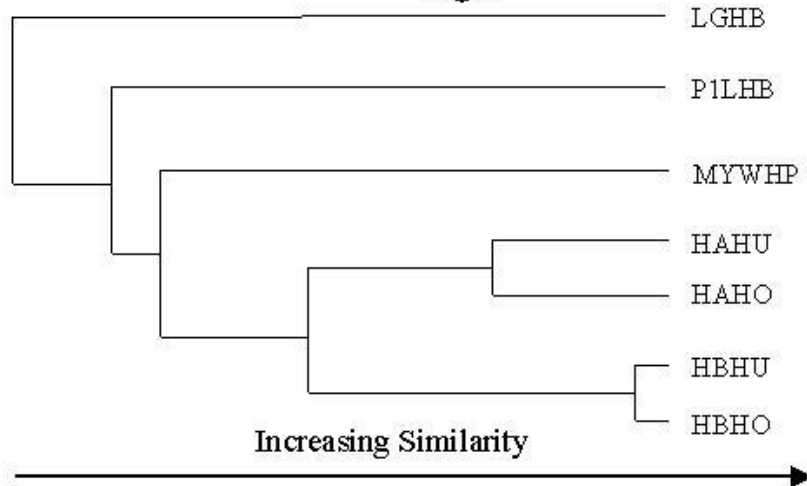
Longitud de la región
en el cromosoma

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	NM_000346.4	3929	1	3931	3931	100.0%	chr17	+	72121020	72126416	5397
browser details	NM_000346.4	244	586	912	3931	91.0%	chr16	+	982121	983836	1716
browser details	NM_000346.4	211	571	803	3931	95.3%	chr22	-	37983357	37983589	233
browser details	NM_000346.4	127	690	902	3931	82.3%	chr20	-	64048851	64049259	409
browser details	NM_000346.4	108	685	902	3931	74.8%	chr13	+	112067809	112068026	218
browser details	NM_000346.4	100	684	887	3931	74.6%	chr3	+	181712480	181712683	204
browser details	NM_000346.4	82	686	857	3931	73.9%	chr20	+	326043	326214	172
browser details	NM_000346.4	81	681	858	3931	90.2%	chr6	+	21594705	21594882	178
browser details	NM_000346.4	73	820	906	3931	92.0%	chr8	+	54459087	54459173	87
browser details	NM_000346.4	62	690	791	3931	80.4%	chr8	-	10730195	10730296	102
browser details	NM_000346.4	55	686	830	3931	69.0%	chr3	+	137764807	137764951	145
browser details	NM_000346.4	42	675	717	3931	100.0%	chr13	-	94711996	94712044	49
browser details	NM_000346.4	35	819	859	3931	92.7%	chr17	-	7589358	7589398	41
browser details	NM_000346.4	32	686	719	3931	97.1%	chr1	+	204123700	204123733	34
browser details	NM_000346.4	31	3414	3503	3931	94.3%	chr19	+	48279421	48279551	131
browser details	NM_000346.4	30	3411	3442	3931	96.9%	chr3	-	116220181	116220212	32
browser details	NM_000346.4	29	2527	2563	3931	84.4%	chr16	-	71518534	71518568	35

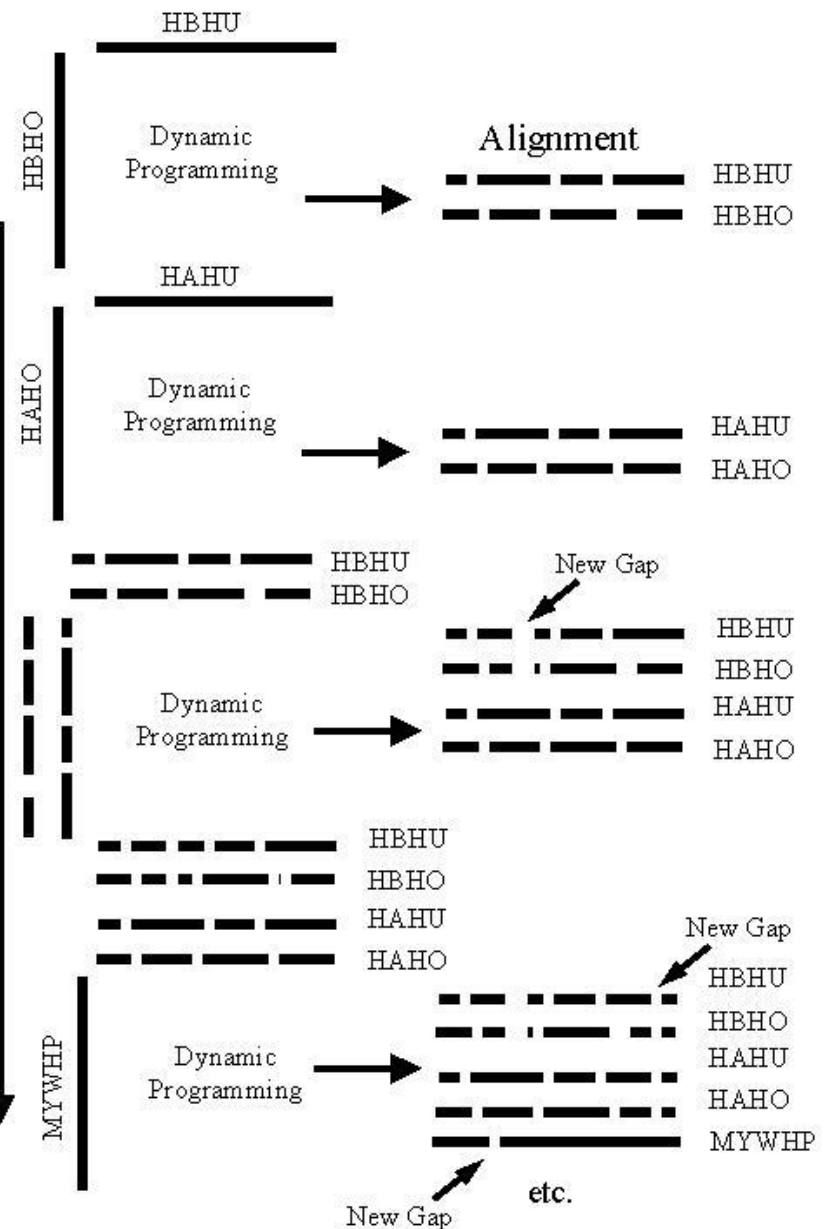
Alineamiento múltiple

	HAHU	HBHU	HAHO	HBHO	MYWHP	P1LHB	LGHB
HAHU							
HBHU	21.1						
HAHO	32.9	19.7					
HBHO	20.7	39.0	20.4				
MYWHP	11.0	9.8	10.3	9.7			
P1LHB	9.3	8.6	9.6	8.4	7.0		
LGHB	7.1	7.3	7.5	7.4	7.3	4.3	

Cluster Analysis



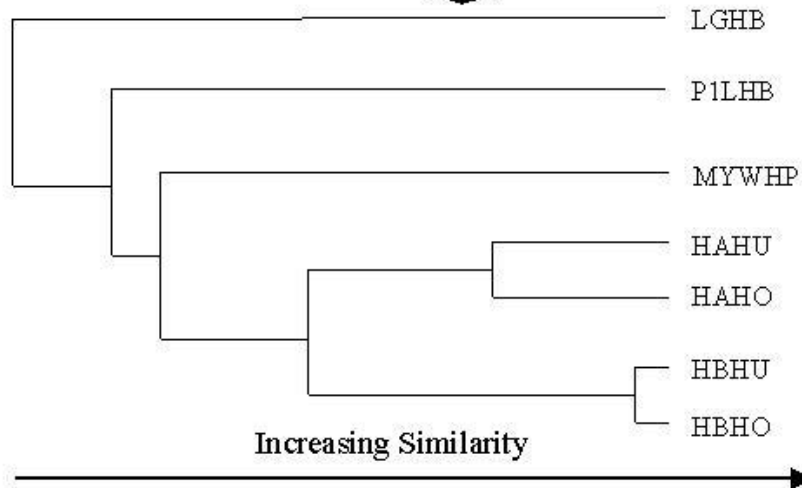
Multiple Alignment



UPGMA (Unweighted Pair Group Method with Arithmetic Mean)

	HAHU	HBHU	HAHO	HBHO	MYWHP	P1LHB	LGHB
HAHU							
HBHU	21.1						
HAHO	32.9	19.7					
HBHO	20.7	39.0	20.4				
MYWHP	11.0	9.8	10.3	9.7			
P1LHB	9.3	8.6	9.6	8.4	7.0		
LGHB	7.1	7.3	7.5	7.4	7.3	4.3	

Cluster Analysis



Multiple Alignment

