



UNIVERSIDAD NACIONAL DE COLOMBIA
INSTITUTO DE BIOTECNOLOGÍA

Centro de Bioinformática

ALINEAMIENTO MÚLTIPLE

EMILIANO BARRETO H.
Profesor Asociado D.E



Instituto de Biotecnología
Universidad Nacional de Colombia

ALINEAMIENTO MÚLTIPLE DE SECUENCIAS (AMS)

- El alineamiento pareado permite establecer relaciones entre dos secuencias, un AMS puede mostrar como las secuencias de una familia se relacionan unas con otras y en la búsqueda de un ancestro común.

“Los cambios acumulados en una secuencia biológica se producen a una tasa relativamente constante e independiente de parámetros poblacionales”

```

chite  ---ADKPKRPLSAYMLWLNSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat  --DPNPKKRAPS AFFVFMGEFFREEFKQKNPKNKSVA AVGKAAGERWKS LSE
trybr   KKDSNAPKRAMT SFMFFSSDFRS----KHS DLS-IVEMSKAAGA AWKELGP
mouse   -----KPKRPRSAYNIYVSESEFQ-----EAKDDS-AQGK LKLVNEAWKNLSP
          ***.  ::.  .:  . . . . . : . . . * . * : *

chite   AATAKQNYIRALQEYERNGG-
wheat   ANKLGGEYNKAI AAYNKGESA
trybr   AEKDKERYKREM-----
mouse   AKDDRIRYDNEMKSWEEQMAE
          *   : . * . :
    
```

¿ PARA QUÉ ALINEAR SIMULTÁNEAMENTE VARIAS SECUENCIAS?

- **Identificar residuos altamente conservados**

Probables sitios esenciales para estructura/función

Mayor precisión a partir de múltiples secuencias

Mejor predicción de estructura/función, alineamientos pareados

- **Construir familias de genes /proteínas**

Usa regiones conservadas para guiar la búsqueda

- **Base para análisis filogenético**

Inferencia de relaciones evolutivas entre genes

- **Desarrollo de cebadores y sondas**

Uso de regiones conservadas para desarrollar

Cebadores para PCR

Sondas para microarreglos de DNA

CÓMO USAR UN ALINEAMIENTO MÚLTIPLE?

```
chite ---ADKPKRPLSAYMLWLNSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNPKKRAPSAFFVFMGEFREEFKQKNPKNKSVAAVGKAAGERWKSLSLSE
trybr  KKDSNAPKRAMTSFMFFSDFRS----KHSDLS-IVEMSKAAGA AWKELGP
unknown -----KPKRPRSAYNIYVSESFQ-----EAKDDS-AQGKLVNEAWKNLSP
          ***. ::: :. . . . : . . * . *: *

chite  AATAKQNYIRALQEYERNGG-
wheat  ANKLKGEYNKAI AAYNKGESA
trybr  AEKDKERYKREM-----
unknown AKDDRIRYDNEMKSWEEQMAE
          * : .* . :
```

Extrapolación

Homología?

SwissProt

Secuencia desconocida



CÓMO USAR UN ALINEAMIENTO MÚLTIPLE?

```

chite ---ADKPKRPLSAFMLNSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNKPKRAPSAFFVFMGEFREFKQKNPKNKSVAAVGKAAGERWKS LSE
trybr  KKDSNAPKRAMTSMFFSSDFRS----KHS DLS-IVEMSKAAGAAWKE LGP
mouse  ----KPKRPR SAYNIYVSESFQ----EAKDDS-AQGK LKLVNEAWKNLSP
          ***. ::: :. . . . : . . * . *: *

chite  AATAKQNYIRALQEYERNGG-
wheat  ANKLGGEYNKAI AAYNKGES A
trybr  AEKDKERYKREM-----
mouse  AKDDRIRYDNEMKSWEEQMAE
          * : .* . :
    
```

Extrapolación

Patrones Prosite

Apareamiento?

SwissProt

Secuencia desconocida



CÓMO USAR UN ALINEAMIENTO MÚLTIPLE?

```

chite ---ADKPKRPLSAYMLWLNSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNKPKRAPS AFFVFMGEFREEFKQKNPKNKSVAAVGKAAGERWKSLSLSE
trybr  KKDSNAPKRAMTSMFFSSDFRS----KHS DLS-IVEMSKAAGA AWKELGP
mouse  ----KPKRPRSAYNIYVSESFQ----EAKDDS-IQGKLLKLVNEAWK NLSLSP
          ***. ::: .: . . . . : . . . * . * *

chite AATAKQNYIRALQEYERNGG-
wheat ANKLGGEYNKAI AAYNKGESA
trybr  AEKDKERYKREM-----
mouse  AKDDRIRYDNEMKSWEEQMAE
          * : .* . :
    
```

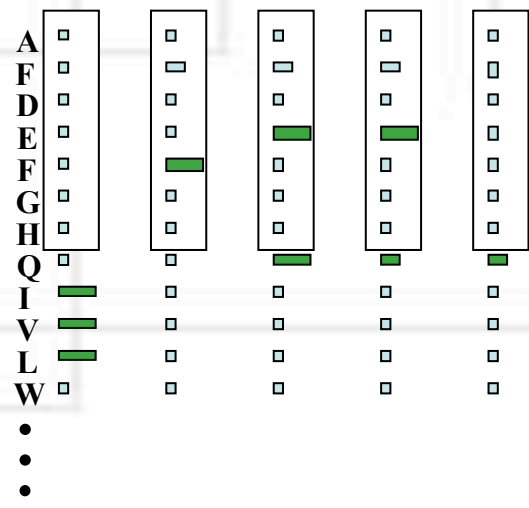
L?

K>R

Extrapolación

Patrones Prosite

Perfiles Prosite



-Mas Sensible
 .. -Mas Especifico

CÓMO USAR UN ALINEAMIENTO MÚLTIPLE?

```

chite  ---ADKPKRPLSAYMLWLNSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat  --DPNKPKRAPSAFFVFMGEFREEFKQKNPKNKSVAAVGKAAGERWKSLSLSE
trybr  KKDSNAPKRAMTSFMFFSSDFRS----KHS DLS-IVEMSKAAGA AWKELGP
mouse  -----KPKRPRSAYNIYVSESFQ-----EAKDDS-AQGK LKLVNEAWKNLSP

```

```

***. ::: .: . . . : . . * . *: *

```

```

chite  AATAKQNYIRALQEYERNGG-
wheat  ANKLGGEYNKAI AAYNKGESA
trybr  AEKDKERYKREM-----
mouse  AKDDRIRYDNEMKSWEEQMAE

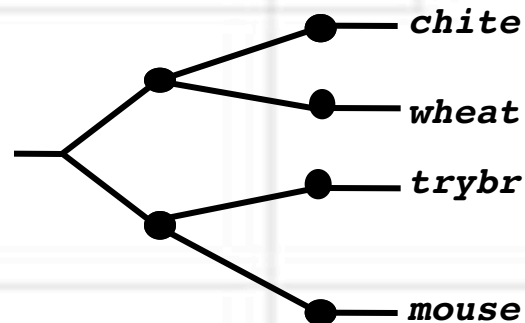
```

```

* : .* . :

```

Filogenia



-Evolución
-Paralogía/Ortología

CÓMO USAR UN ALINEAMIENTO MÚLTIPLE?

```
chite ---ADKPKRPLSAYMLWLNSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNKPKRAPSAFFVFMGEFREFKQKNPKNKSVAAVGKAAGERWKSLSLSE
trybr  KKDSNAPKRAMTSFMFFSSDFRS----KHS DLS-IVEMSKAAGA AWKELGP
mouse  -----KPKRPRSAYNIYVSESFQ----EAKDDS-AQGKLLKLVNEAWKNLSP
      ***. ::: .: .. . : . . * . *: *

chite  AATAKQNYIRALQEYERNGG-
wheat  ANKLGGEYNKAI AAYNKGESA
trybr  AEKDKERYKREM-----
mouse  AKDDRIRYDNEMKSWEEQMAE
      * : .* . :
```

Filogenia

Predicción Estruct.

PhD Para Predicción de estructura secundaria:
75% precisión.

Threading: esta mejorando
Pero aún no es tan bueno.



CÓMO USAR UN ALINEAMIENTO MÚLTIPLE?

```
chite ---ADKPKRPLSAYMLWLNSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNPKRAPSAFFVFMGEFREFFKQKNPKNKSVAAVGKAAGERWKSLSLSE
trybr  KKDSNAPKRAMTSFMFFSSDFRS----KHSMLS-IVEMSKAAGA AWKELGP
mouse  -----KPKRPRSAYNIYVSESFQ----EAKDDS-AQGKLLVNEAWKNLSP
          ***. ::: .: . . . : . . * . *: *
```

```
chite AATAKQNYIRALQEYERNGG-
wheat ANKLLKGEYNKAI AAYNKGESA
trybr  AEKDKERYKREM-----
mouse  AKDDRIRYDNEMKSWEEQMAE
          * : .* . :
```

Filogenia

Predicción Estruct.

Precaución!

Métodos automáticos de
alineamiento múltiple de secuencias
No son siempre perfectos...

EL PROBLEMA...

- Por qué es difícil computar un alineamiento múltiple?

Biología

Qué es **un buen** alineamiento?

```
chite  ---ADKPKRPLSAYMLWLN SARESIKRENPDFK-VTEVAKKG GELWRGLKD
wheat  --DPNKPKRAPSAFFVFMGEF REEFKQKNPKNKSVA AVGKAAGERWKS LSE
trybr   KKDSNAPKRAMTSEFMFFSSDFRS----KHS DLS-IVEMSKAAGA AWKELGP
mouse   -----KPKRPRSAYNIYVSE SFQ----EAKDDS-AQGK LKLVNEAWKNLSP
          ***.  :::  .:  . .  .  :  . .  *  .  *:  *
```

Computación

Cual es **el** buen alineamiento?



EL PROBLEMA...

- Por qué es difícil computar un alineamiento múltiple?

PROBLEMA CIRCULAR...

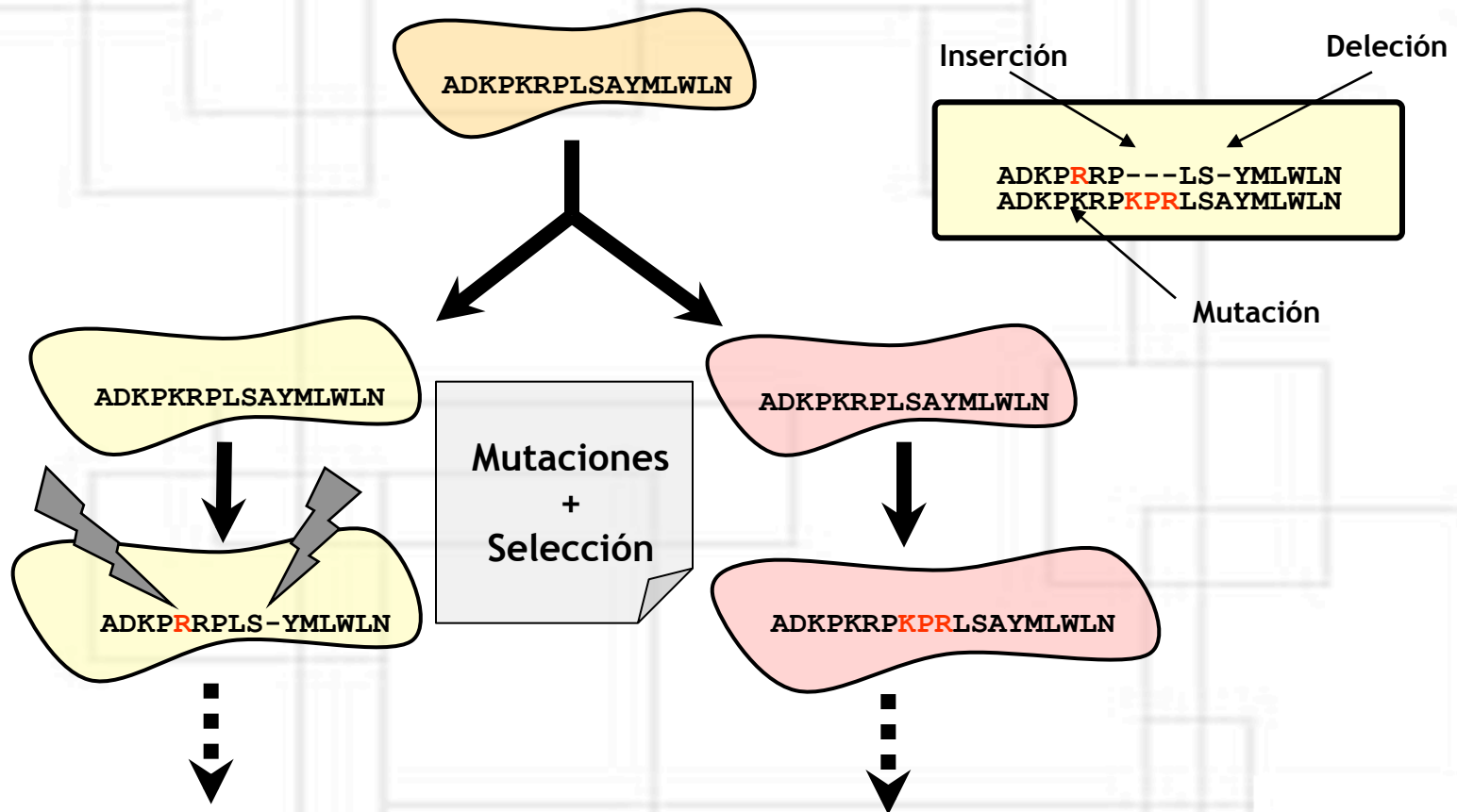
Buenas
secuencias

Buen
Alineamiento

Qué se necesita conocer para hacer un buen alineamiento ?

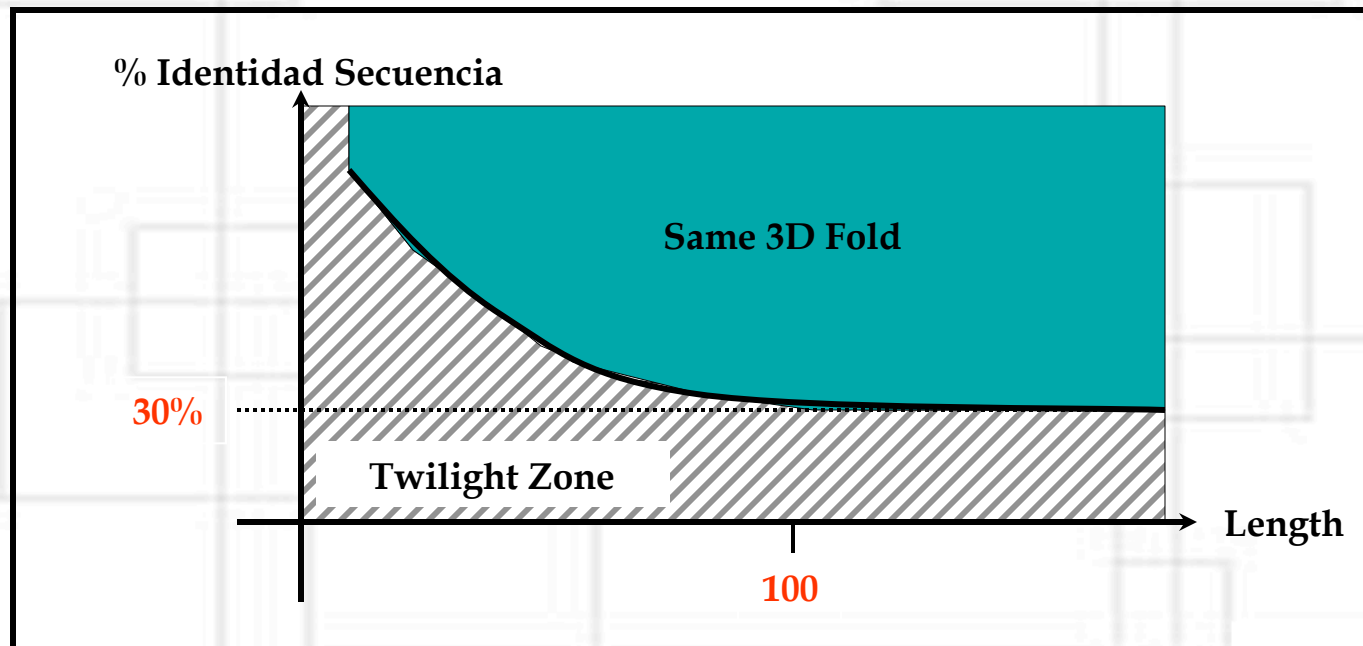
- Cómo evolucionaron las secuencias?
- Cómo alinea las secuencias el computador?
- Cómo escoger las secuencias?
- Cual es el mejor programa?
- Cómo usar el alineamiento?

UN ALINEAMIENTO ES UNA HISTORIA



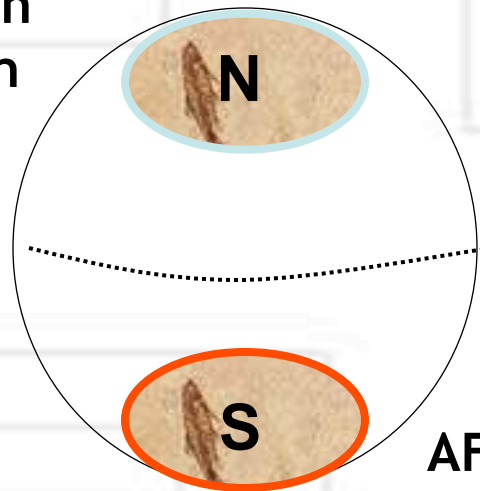
HOMOLOGÍA

- Secuencias Iguales -> Igual origen? -> Igual función? -> Igual plegamiento 3D?



Evolución Convergente

AFGP con (ThrAlaAla)_n
Similar a Trypsinogen

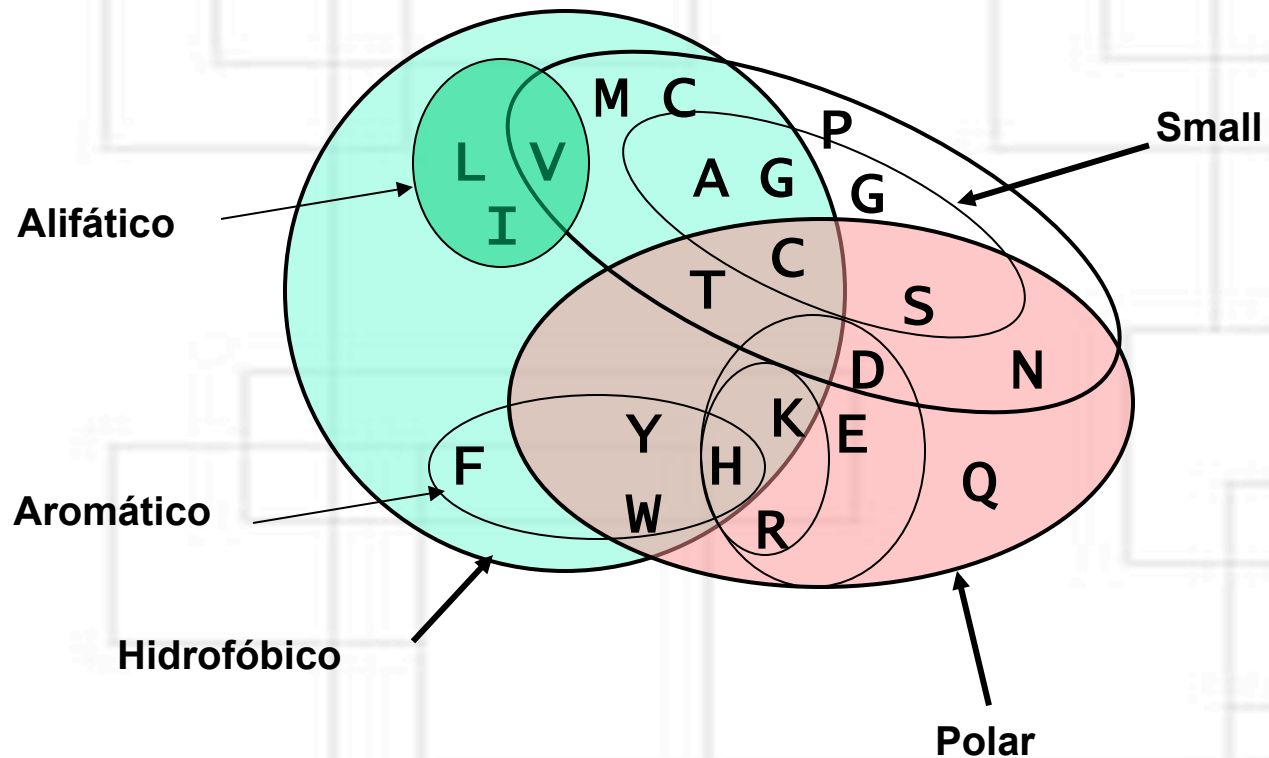


AFGP con (ThrAlaAla)_n
NO
Similar a Trypsinogen

Chen et al, 97, PNAS, 94, 3811-16

RESIDUOS Y MUTACIONES

- Todos los residuos son iguales, pero algunos mas que otros...



CÚAL ES LA MEJOR MATRIZ DE SUSTITUCIÓN?

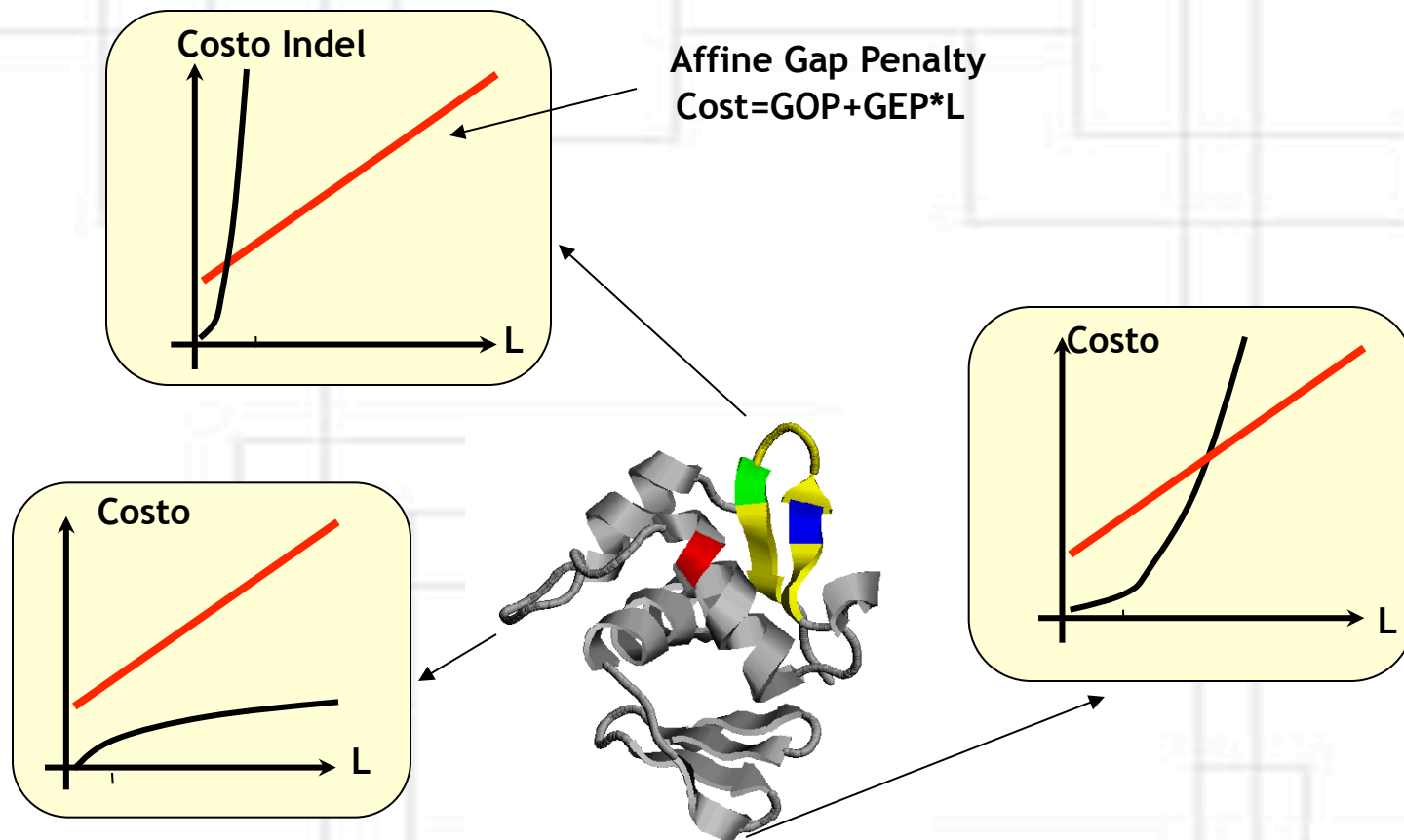
- Las tasas de mutación dependen de la familia proteica

FAMILIA	S	N
Insulin	4.0	0.1
Interleukin I	4.6	1.4
α -Globin	5.1	0.6
Apolipoprot. AI	4.5	1.6
Interferon G	8.6	2.8
Histone3	6.4	0

Tasas en Substituciones/sitio/Billones de años para Ratón Vs Humano (0.08 Billones años)

- Escoger la matriz Correcta puede ser difícil
 - Gonnet250 > BLOSUM62 > PAM250
 - Depende de la familia, el programa usado y su configuración

INSERCIONES Y DELECCIONES?



CÓMO ALINEAR MUCHAS SECUENCIAS?

- Algoritmos exactos consumen mucho tiempo de cómputo
 - Needleman & Wunsch
 - Smith & Waterman

2 Globinas => 1 segundo



3 Globinas => 2 mn



4 Globinas => 5 horas



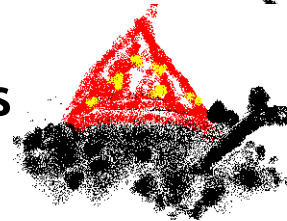
5 Globinas => 3 semanas



6 Globinas => 9 años



7 Globinas => 1000 años



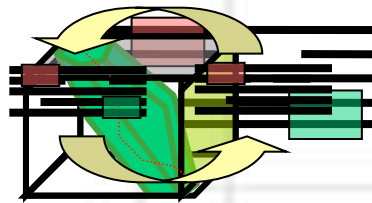
8 Globinas => 150 000 años

HEMO-WHAT?

-> Se requiere un heurístico!

MÉTODOS EXISTENTES

1- Basado en segmentos:



- HMMs, HMMER, SAM.
- MSA, DCA.
- Lento, Algunas veces impreciso
- Pocas secuencias cercanamente relacionadas pueden dar buenos perfiles

4- Progresivo:

- ClustalW, MUSCLE, t_Coffe, Pileup, Multalign...
- rápido y sensible

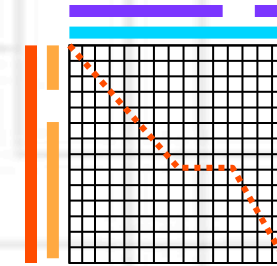
AMS global- Alineamiento Global Progresivo

- **Motivo:**
 - Reducir el costo de construir un alineamiento global incrementalmente
- **Método:**
 1. Computar la distancia entre todos los pares de secuencias
 2. Construir un arbol guia simple reflejando las distancias entre secuencias
 - Usar UPGMA (PILEUP) o Neighbor-joining (CLUSTALW)
 3. Alinear secuencias siguiendo árbol guía, comenzando en las hojas
 - Alinear las secuencias consenso o perfiles
 - Usar algoritmos de alineamiento pareado heurísticos u óptimos
 - Tender a colocar gaps entre regiones conservadas

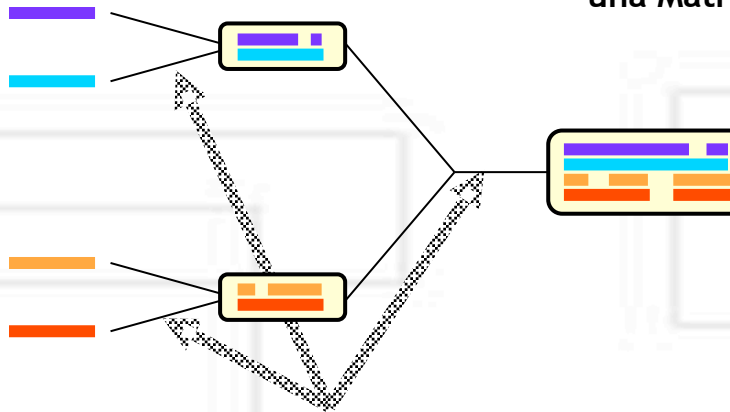


AMS global- Alineamiento Global Progresivo

- Depende de las secuencias **ESCOGIDAS**
- Depende del **ORDEN** de las secuencias (Arbol).
- Depende de los **PARAMETROS**
 - Matriz de Substitución.
 - Penalización (Gop, Gep).
 - Peso Secuencia.
 - Algoritmo construcción arbol.



Programación Dinámica Usando una Matriz de Substitución



•Problemas:

- Aproximación dependiente de alineamientos pareados iniciales
- No puede fijar errores iniciales (gaps no pueden ser removidos)

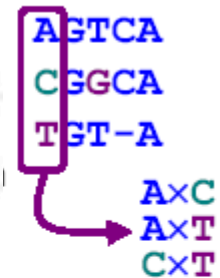
“once a gap, always a gap”

Feng y Dolittle, 1980; Taylor 1981

PUNTUACIÓN DEL ALINEAMIENTO MÚLTIPLE DE SECUENCIAS

- Suma de pares (SP):

- Secuencias Alineadas en columnas
- Dar puntaje a todas las combinaciones pareadas en la columna
- Requiere $\frac{N \times (N-1)}{2}$ comparaciones para N secuencias
- Puntaje total = Σ puntaje de cada columna

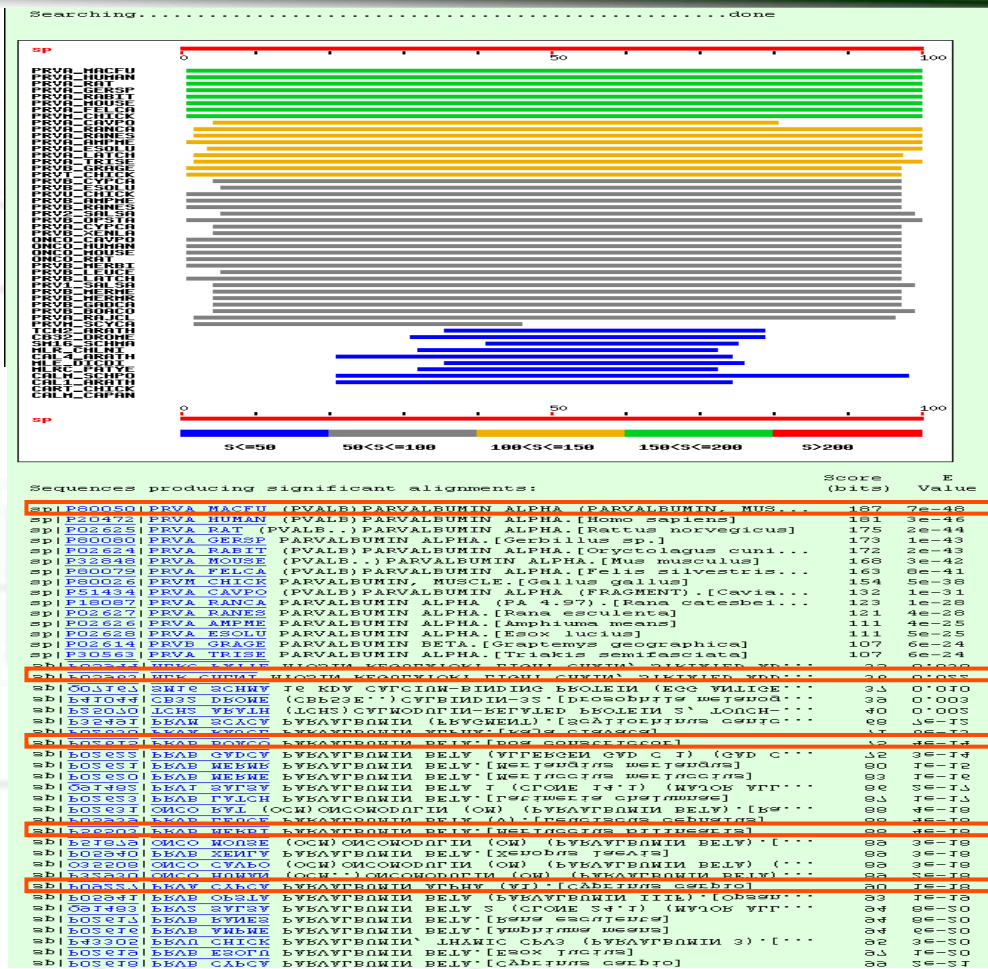


- Condiciones:

- Asume que cada secuencia desciende de N-1 secuencias
- No hay justificación probabilística para el puntaje SP
- Penalización relativa para malos apareamientos disminuye, para mas apareamientos debe aumentar
- Ejemplo: **T** mal apareada

	A	A	A	A	A	A	A	
	A	A	A	A	A	A	A	
	A	T	A	T	A	A	T	
A x A = 4	12	4	24	12	40	24		
A x T = 0	$\Delta=67\%$		$\Delta=50\%$	$\Delta=40\%$				

SELECCIÓN DE SECUENCIAS A PARTIR DE UNA SALIDA BLAST



UN ERROR FRECUENTE...

- **Secuencias cercanamente relacionadas**

```
PRVA_MACFU    SMTDLLNAEDIKKAVGAFSAIDSEFDHKKFFQMVGLKKKSADDVKKVFHILDKDKSGFIEE
PRVA_HUMAN    SMTDLLNAEDIKKAVGAFSATDSEFDHKKFFQMVGLKKKSADDVKKVFHMLDKDKSGFIEE
PRVA_GERSP    SMTDLLSAEDIKKAIGAFAAADSEFDHKKFFQMVGLKKKT PDDVKKVFHILDKDKSGFIEE
PRVA_MOUSE    SMTDVLSAEDIKKAIGAFAAADSEFDHKKFFQMVGLKKKNPDEVKKVFHILDKDKSGFIEE
PRVA_RAT      SMTDLLSAEDIKKAIGAFATAADSEFDHKKFFQMVGLKKKSADDVKKVFHILDKDKSGFIEE
PRVA_RABIT    AMTELLNAEDIKKAIGAFAAAESFDHKKFFQMVGLKKKSTEDVKKVFHILDKDKSGFIEE
:***:*.*.*****:***:* :*****. .:*****:*****

PRVA_MACFU    DELGFILKGFSPDARDLSAKETKTLMAAGDKDGDGKIGVDEFSTLVAES
PRVA_HUMAN    DELGFILKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLVAES
PRVA_GERSP    DELGFILKGFSSDARDLSAKETKTLLAAGDKDGDGKIGVEEFSTLVSES
PRVA_MOUSE    DELGSILKGFSSDARDLSAKETKTLLAAGDKDGDGKIGVEEFSTLVAES
PRVA_RAT      DELGSILKGFSSDARDLSAKETKTLMAAGDKDGDGKIGVEEFSTLVAES
PRVA_RABIT    EELGFILKGFSPDARDLSVKETKTLMAAGDKDGDGKIGADEFSTLVSES
:*** *****.*****.***** *:*****.*****:***
```

- **Secuencias Idénticas no ofrecen información**
- **Alineamientos Múltiple de secuencias son valiosos por la diversidad de secuencias**

SELECCIÓN DE SECUENCIAS DIVERSAS

```

PRVA_MACFU -----SMTDLLN----AEDIKKA
PRVA_HUMAN -----SMTDLLN----AEDIKKA
PRVA_GERSP -----SMTDLLS----AEDIKKA
PRVA_MOUSE -----SMTDVLS----AEDIKKA
PRVA_RAT -----SMTDLLS----AEDIKKA
PRVA_RABIT -----AMTELLN----AEDIKKA
TPCC_MOUSE MDDIYKAAVEQLTEEQKNEFKAAFDIFVLGAEDGCISTKELGKVMRMLGQNPTPEELQEM
: : * . . * : : :
    
```

-Este alineamiento no es informativo acerca de la relación entre TPCC MOUSE y las otras secuencias.

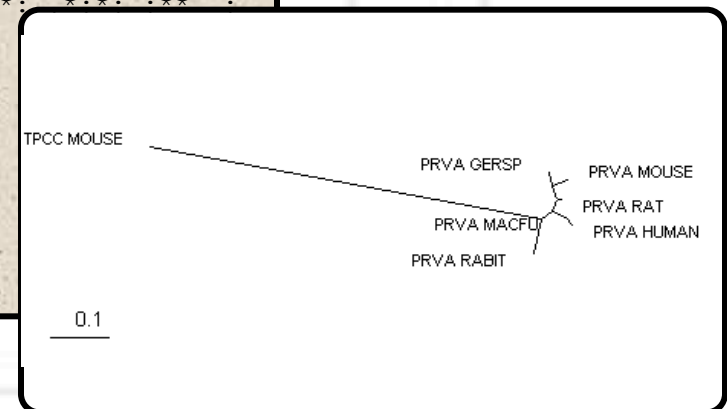
```

PRVA_MACFU VGFSAIDS--FDHKKFFQMGV-----LKKKSADDVKKVFHILDKDKSGFIEEDELGFI
PRVA_HUMAN VGFSAATDS--FDHKKFFQMGV-----LKKKSADDVKKVFHMLDKDKSGFIEEDELGFI
PRVA_GERSP IGAFAAADS--FDHKKFFQMGV-----LKKKTPDDVKKVFHILDKDKSGFIEEDELGFI
PRVA_MOUSE IGAFAAADS--FDHKKFFQMGV-----LKKKNPDEVKKVFHILDKDKSGFIEEDELGSI
PRVA_RAT IGAFATAADS--FDHKKFFQMGV-----LKKKSADDVKKVFHILDKDKSGFIEEDELGSI
PRVA_RABIT IGAFAAAES--FDHKKFFQMGV-----LKKKSTEDVKKVFHILDKDKSGFIEEELGFI
TPCC_MOUSE IDEVDEDGSGTVDFDEFVLMMVRCMKDSSK GKSEEEELSDLFRMFDKNADGYIDLDELKMM
: . . * . * : : * : * * . : : : * : : * * . * * . * *
    
```

-Una mejor extensión de las secuencias es necesaria

```

PRVA_MACFU LKGFSPDARDLSAKETKTLMAAGDKDGDGKIGVDEFSTLVAES-
PRVA_HUMAN LKGFSPDARDLSAKETKMLMAAGDKDGDGKIGVDEFSTLVAES-
PRVA_GERSP LKGFSSDARDLSAKETKTLLAAGDKDGDGKIGVEEFSTLVSES-
PRVA_MOUSE LKGFSSDARDLSAKETKTLLAAGDKDGDGKIGVEEFSTLVAES-
PRVA_RAT LKGFSSDARDLSAKETKTLMAAGDKDGDGKIGVEEFSTLVAES-
PRVA_RABIT LKGFSPDARDLSVKETKTLMAAGDKDGDGKIGADEFSTLVSES-
TPCC_MOUSE LQ---ATGETITEDDIEELMKDGDKNNDGRIDYDEFLEFMKGVE
* : . . : : : * : * * : * * . * * : :
    
```



SELECCIÓN DE SECUENCIAS DIVERSAS

```

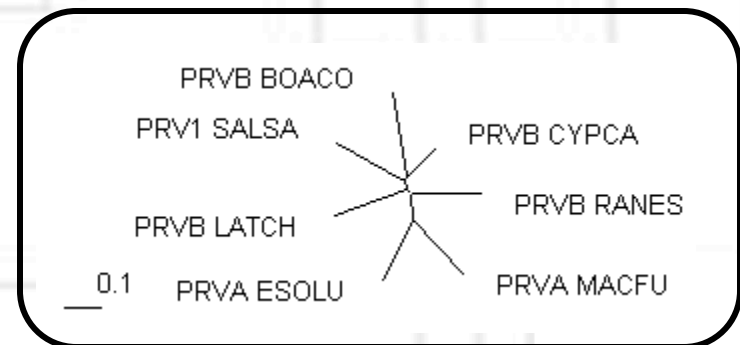
PRVB_CYPKA -AFAGVLNDADIAAALEACKAADSFNHKAFFAKVGLTSSKSAADDVKKAFAIIDQDKSGFIE
PRVB_BOACO -AFAGILSDADIAAGLQSCQAADSFSCKTFFAKSGLHSSKSDQLTKVFGVIDRDKSGYIE
PRV1_SALSA  MACAHLCKEADIKTALEACKAADTFSFKTFFHTIGFASKSADDVKKAFKVIDQDASGFIE
PRVB_LATCH -AVAKLLAAADVTAALEGCKADDSFNHKVFFQKTGLAKKSNEELEAIFKILDQDKSGFIE
PRVB_RANES -SITDIVSEKDIDAALSVKAAGSFNYKIFFQKVLGAGKSAADAKKVFIEILDRDKSGFIE
PRVA_MACFU -SMTDLLNAEDIKKAVGAFSAIDSFDHKKFFQMVGLKKSADDVKKVFHILDKDKSGFIE
PRVA_ESOLU --AKDLLKADDIKKALDAVKAEGSFNHKKFFALVGLKAMSANDVKKVFKAIDADASGFIE
          :  * . : . * . * * * * : * : * . * * * * *
    
```

```

PRVB_CYPKA EDELKLFQNFKADARALTDGETKTFLKAGDSGDGKIGVDEFTALVKA-
PRVB_BOACO EDELKKFLQNFQKARDLTDKETAEFLKEGDTDGDGKIGVEEFVVLVTGK
PRV1_SALSA  VEELKLFQNFQPKARELTDKETAEFLKAGDADGDMIGIDEFAVLVKQ-
PRVB_LATCH DEEELFLQNFQKARTLTKTETETFLKAGDSGDGKIGVDEFQKLVKA-
PRVB_RANES QDELGLFLQNFQKARSVLSAETSAFLKAGDSGDGKIGVEEFQALVKA-
PRVA_MACFU EDELGFILKGFSPDARDLSAKETKTLMAAGDKDGDGKIGVDEFSTLVAES
PRVA_ESOLU EEELKFVLKSFQADGRDLTDKETAEFLKAGDADGDMIGIDEFETLVHEA
          : ** . * . * . * * : ** : : * * * * * * : * * * *
    
```

-Un modelo razonable existe ahora...

-Ir más allá : homologos remotos.



ALINEAMIENTO DE HOMOLOGOS REMOTOS

```

PRVA_MACFU -----SMTDLLNA----EDIKKA
PRVA_ESOLU -----AKDLLKA----DDIKKA
PRVB_CYPKA -----AFAGVLND----ADIAAA
PRVB_BOACO -----AFAGILSD----ADIAAG
PRV1_SALSA -----MACAHLCKE----ADIKTA
PRVB_LATCH -----AVAKLLAA----ADVTA
PRVB_RANES -----SITDIVSE----KDIDAA
TPCS_RABIT -TDQQAEARSYLSEEMIAEFKAAAFDMFDADGG-GDISVKELGTVMRMLGQTPPTKEELDAI
TPCS_PIG -TDQQAEARSYLSEEMIAEFKAAAFDMFDADGG-GDISVKELGTVMRMLGQTPPTKEELDAI
TPCC_MOUSE MDDIYKAAVEQLTEEQKNEFKAAFDIFVLGAEDGCISTKELGKVMRMLGQNPTPEELQEM
: :

```

```

PRVA_MACFU VGAFSAIDS--FDHKKFFQMGV-----LKKKSADDVKKV F H I L D K D K S G F I E E D E I G F I
PRVA_ESOLU L D A V K A E G S -- F N H K K F F A L V G ----- L K A M S A N D V K K V F K A I D A D A S G F I E E E E I K F V
PRVB_CYPKA L E A C K A A D S -- F N H K A F F A K V G ----- L T S K S A D D V K K A F A I I D Q D K S G F I E E D E I K L F
PRVB_BOACO L Q S C Q A A D S -- F S C K T F F A K S G ----- L H S K S K D Q L T K V F G V I D R D K S G Y I E E D E I K K F
PRV1_SALSA L E A C K A A D T -- F S F K T F F H T I G ----- F A S K S A D D V K K A F K V I D Q D A S G F I E V E E I K L F
PRVB_LATCH L E G C K A D D S -- F N H K V F F Q K T G ----- L A K K S N E E L E A I F K I L D Q D K S G F I E D E E I E L F
PRVB_RANES L E S V K A A G S -- F N Y K I F F Q K V G ----- L A G K S A A D A K K V F E I L D R D K S G F I E Q D E I G L F
TPCS_RABIT I E E V D E D G S G T I D F E E F L V M M V R Q M K E D A K G K S E E E L A E C F R I F D R N A D G Y I D A E E I A E I
TPCS_PIG I E E V D E D G S G T I D F E E F L V M M V R Q M K E D A K G K S E E E L A E C F R I F D R N M D G Y I D A E E I A E I
TPCC_MOUSE I D E V D E D G S G T V D F D E F L V M M V R C M K D D S K G K S E E E L S D I F R M F D K N A D G Y I D L D E I K M M
: . . . : . . . * : * : * . * : . * : . * : . * :

```

```

PRVA_MACFU L K G F S P D A R D L S A K E T K T L M A A G D K D G D G K I G V D E F S T L V A E S -
PRVA_ESOLU L K S F A A D G R D L T D A E T K A F L K A A D K D G D G K I G I D E F E T L V H E A -
PRVB_CYPKA L Q N F K A D A R A L T D G E T K T F L K A G D S D G D G K I G V D E F T A L V K A --
PRVB_BOACO L Q N F D G K A R D L T D K E T A E F L K E G D T D G D G K I G V E E F V V L V T K G -
PRV1_SALSA L Q N F C P K A R E L T D A E T K A F L K A G D A D G D G M I G I D E F A V L V K Q --
PRVB_LATCH L Q N F S A G A R T L T K T E T E T F L K A G D S D G D G K I G V D E F Q K L V K A --
PRVB_RANES L Q N F R A S A R V L S D A E T S A F L K A G D S D G D G K I G V E E F Q A L V K A --
TPCS_RABIT F R --- A S G E H V T D E E I E S L M K D G D K N N D G R I D F D E F I K M M E G V Q
TPCS_PIG F R --- A S G E H V T D E E I E S I M K D G D K N N D G R I D F D E F I K M M E G V Q
TPCC_MOUSE L Q --- A T G E T I T E D D I E E L M K D G D K N N D G R I D Y D E F I E F M K G V E
: . . . : . . . * : * : * . * : . * : . * : . * :

```

IR MAS ALLÁ...

General information about the entry

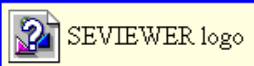
Entry name	TPC_PATYE	
Primary accession number	P35622	

Comments

- FUNCTION:** TROPONIN IS THE CENTRAL REGULATORY PROTEIN OF STRIATED MUSCLE CONTRACTION. TN CONSISTS OF THREE COMPONENTS: TN-I WHICH IS THE INHIBITOR OF ACTOMYOSIN ATPASE, TN-T WHICH CONTAINS THE BINDING SITE FOR TROPOMYOSIN AND TN-C. THE BINDING OF CALCIUM TO TN-C ABOLISHES THE INHIBITORY ACTION OF TN ON ACTIN FILAMENTS.
- MISCELLANEOUS:** THIS PROTEIN BINDS ONE CALCIUM ION PER MOLECULE.
- SIMILARITY:** TO OTHER EF-HAND CALCIUM BINDING PROTEINS.

Features

MOD_RES	<u>1</u>	<u>1</u>	ACETYLTATION.
DOMAIN	<u>22</u>	<u>33</u>	ANCESTRAL CALCIUM SITE 1.
DOMAIN	<u>58</u>	<u>69</u>	ANCESTRAL CALCIUM SITE 2.
DOMAIN	<u>95</u>	<u>106</u>	ANCESTRAL CALCIUM SITE 3.
CA_BIND	<u>131</u>	<u>142</u>	SITE 4.

 [FT table viewer](#)

```
--LKKKSADDVKKVFIILDKDKSGFIEEDLGFII
--LHSKSKDQLTKVFCVIDRDKSGYIEEDLKKFI
--FASKSADDVKKAFKVIDQDASGFIEVEELKLF
EDAKGKSEEEELAEFCRIFDRNADGYIDAEELAEI
EDAKGKSEEEELAEFCRIFDRNMDGYIDAEELAEI
DSKKGSEEEELSDLFRMFDKNADGYIDLDLKM
LKEDLDERELKEAFRVLDKEKKGVIKVIVLRWI
. : * : * : . * * : * .
.DGKIGVDEFSTLVAES--
GDGKIGVEEFVVLVTKG--
PKVI_SALSA LQNFCAKRELTDAAETRAFVLAGDADGDGMIGIDEFAVLVKQ---
TPCS_RABIT FR---ASGEHVTDEEIESIMIDGDKNNDGRIDFDEFKMMEGVQ-
TPCS_PIG FR---ASGEHVTDEEIESIMIDGDKNNDGRIDFDEFKMMEGVQ-
TPCC_MOUSE LQ---ATGETITEDDIEELMIDGDKNNDGRIDYDEFLEFMKGVE-
TPC_PATYE LS---SLGDELTEEEIENMLAETDTDGSSTVDYEEFKCLMMSDA
: . : : : * : . * : : ** : :
```

UN BUEN ALINEAMIENTO...

- Las secuencias mas divergentes
- Muy pocos indels
- Bloques sin gaps separados con indels
- Diferentes clases de residuos dentro de un bloque:
 - Completamente conservado
 - Hidrofilicidad y Tamaño conservados
 - Hidrofilicidad o Tamaño conservados
- La ultima evaluación es cuestión de conocimiento y decisión personal

PARA EVITAR ERRORES...



Centro de
Bioinformática

Instituto de Biotecnología
Universidad Nacional de Colombia

CONSERVAR LA PERSPECTIVA BIOLÓGICA

```
chite ---ADKPKRPLSAYMLWLNSARES IKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNPKKRAPSFAFFVFMGEFREEFKQKNPKNKSVAAVGKAAGERWKSLSSE
trybr  KKDSNAPKRAMTSFMFFSSDFRS----KHS DLS-IVEMSKAAGAAWKELGP
mouse  -----KPKRPRSAYNIYVSESFQ-----EAKDDS-AQGK LKLVNEAWKNLSP
      ***. ::: :. . . : . . . * . *: *
```

```
chite AATAKQNYIRALQEYERNGG-
wheat ANK LKGEYNKAI AAYNKGESA
trybr  AEKDKERYKREM-----
mouse  AKDDRIRYDNEMKSWEEQMAE
      * : .* . :
```

```
chite AD--K----PKR-PLYMLWLNS-ARESIKRENPDFK-VT-EVAKKGGELWRGL-
wheat -DPNK----PKRAP-FFVFMGE-FREEFKQKNPKNKSVA-AVGKAAGERWKSLS
trybr  -K--KDSNAPKR-AMT-MFFSSDFR-S-KH-S-DLS-IV-EMSKAAGAAWKELG
mouse  ----K----PKR-PRYNIYVSESFQEA-K--D-D-S-AQGKL-KLVNEAWKNLS
      *   *** .:: :. . . : * . . . : * . *: *
```

```
chite KSEWEAKAATAKQNY-I--RALQE-YERNG-G-
wheat KAPYVAKANKLKGEY-N--KAIAA-YNK-GESA
trybr  RKVYEEMA EKDKERY----K--RE-M-----
mouse  KQAYIQLAKDDRIRYDNEMKSWEEQMAE-----
      : : * : .* :
```

DIFERENTES
PARAMETROS

NO AFINAR!!!

```
chite ---ADKPKRPLSAYMLWLNSARESIKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNKPKRAPSAFFVFMGEFREEFKQKNPKNKSVAAVGKAAGERWKSLSE
trybr  KKDSNAPKRAMTSFMFFSSDFRS-----KHSDLS-IVEMSKAAGAAWKELGP
mouse  -----KPKRPRSAYNIYVSESFQ-----EAKDDS-AQGKLKLVNEAWKNLSP
      ***. ::: .: . . . : . . . * . *: *
```

```
chite  AATAKQNYIRALQEYERNGG-
wheat  ANKLKGEYNKAIAAAYNKGESA
trybr  AEKDKERYKREM-----
mouse  AKDDRIRYDNEMKSWEEQMAE
      * : .* . :
```

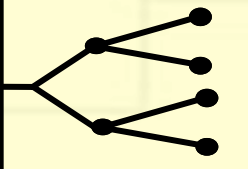
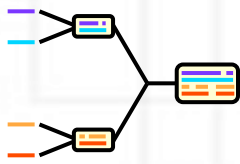
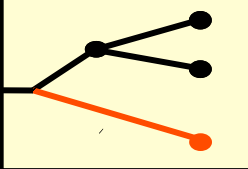
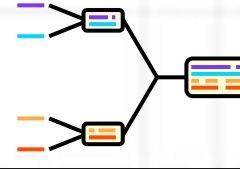
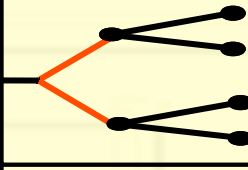
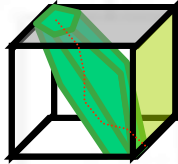
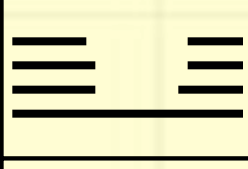
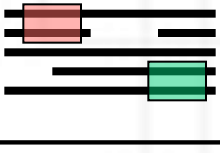
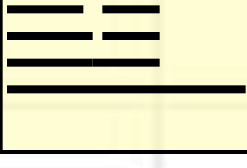
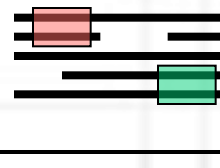
```
chite ---ADKPKRPL-SAYMLWLNSARESIKRENPDFK-VTEVAKKGGELWRGLKD
wheat --DPNKPKRAP-SAFFVFMGEFREEFKQKNPKNKSVAAVGKAAGERWKSLSE
trybr  KKDSNAPKRAMTSFMFFSSDFRS-----KHSDLS-IVEMSKAAGAAWKELGP
mouse  -----KPKRPR-SAYNIYVSESFQ-----EAKDDS-AQGKLKLVNEAWKNLSP
      ***. * .: . . . : . . . * . *: *
```

```
chite  AATAKQNYIRALQEYERNGG-
wheat  ANKLKGEYNKAIAAAYNKGESA
trybr  AEKDKERYKREM-----
mouse  AKDDRIRYDNEMKSWEEQMAE
      * : .* . :
```

No juegue con los
parámetros!
Si conoce el
alineamiento que
desea:
Hagalo usted
mismo!

MÉTODOS DISPONIBLES

Fuente: BaliBase
Thompson et al, NAR, 1999

PROBLEMA	PROGRAMA	METODO
	ClustalW	
	ClustalW	
	MSA	
	DIALIGN II	
	DIALIGN II	

CONCLUSIÓN

- El Mejor Método De Alineamiento:
 - Su cerebro
 - Los datos correctos
- El Mejor Método De evaluación:
 - Sus ojos
 - Información Experimental (SwissProt)
- Que puede concluir?
 - Homología -> extrapolación de información
- Cómo ir más allá?
 - Patrones
 - Perfiles
 - HMMs
 - ...