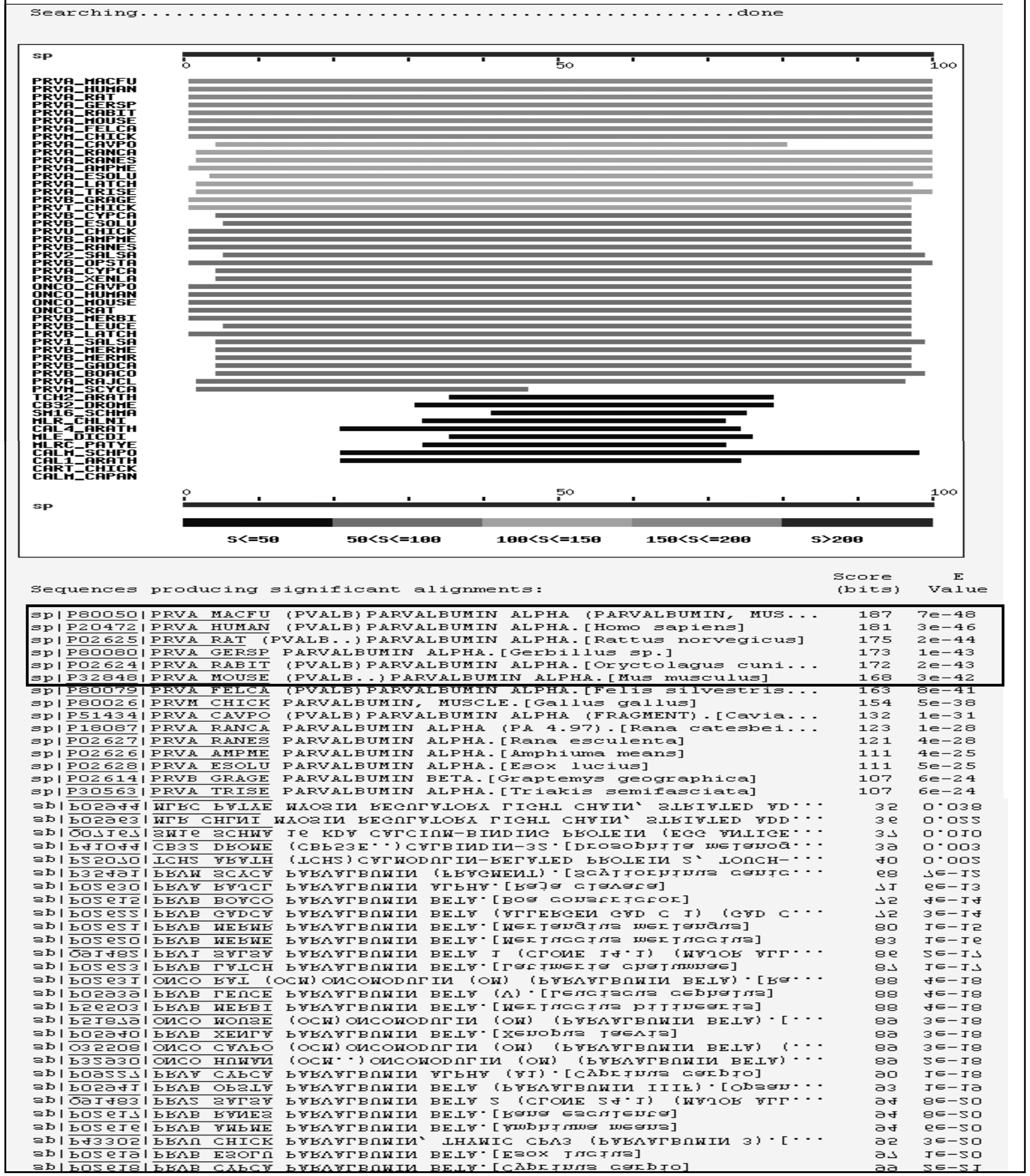
# Gathering Sequences: BLAST



```
Searching..................................................done

                                                                   Score    E
Sequences producing significant alignments:                        (bits)  Value

sp|P80050|PRVA_MACFU (PVALB)PARVALBUMIN ALPHA (PARVALBUMIN, MUS...   187   7e-48
sp|P20472|PRVA_HUMAN (PVALB)PARVALBUMIN ALPHA.[Homo sapiens]         181   3e-46
sp|P02625|PRVA_RAT (PVALB..)PARVALBUMIN ALPHA.[Rattus norvegicus]    175   2e-44
sp|P80080|PRVA_GERSP PARVALBUMIN ALPHA.[Gerbillus sp.]               173   1e-43
sp|P02624|PRVA_RABIT (PVALB)PARVALBUMIN ALPHA.[Oryctolagus cuni...   172   2e-43
sp|P32848|PRVA_MOUSE (PVALB..)PARVALBUMIN ALPHA.[Mus musculus]       168   3e-42
sp|P80079|PRVA_FELCA (PVALB)PARVALBUMIN ALPHA.[Felis silvestris...   163   8e-41
sp|P80026|PRVM_CHICK PARVALBUMIN, MUSCLE.[Gallus gallus]             154   5e-38
sp|P51434|PRVA_CAVPO (PVALB)PARVALBUMIN ALPHA (FRAGMENT).[Cavia...   132   1e-31
sp|P18087|PRVA_RANCA PARVALBUMIN ALPHA (PA 4.97).[Rana catesbei...   123   1e-28
sp|P02627|PRVA_RANES PARVALBUMIN ALPHA.[Rana esculenta]              121   4e-28
sp|P02626|PRVA_AMPME PARVALBUMIN ALPHA.[Amphiuma means]              111   4e-25
sp|P02628|PRVA_ESOLU PARVALBUMIN ALPHA.[Esox lucius]                 111   5e-25
sp|P02614|PRVB_GRAGE PARVALBUMIN BETA.[Graptemys geographica]        107   6e-24
sp|P30563|PRVA_TRISE PARVALBUMIN ALPHA.[Triakis semifasciata]        107   6e-24
sp|P05944|MLRC_PATYE MYOSIN REGULATORY LIGHT CHAIN, STRIATED AD...    35   0.038
sp|P05963|MLR_CHLNI MYOSIN REGULATORY LIGHT CHAIN, STRIATED ADD...    36   0.022
sp|Q07167|SM16_SCHMA 16 KDA CALCIUM-BINDING PROTEIN (EGG ANTIGE...    37   0.010
sp|P41044|CB32_DROME (CBP53E..)CALBINDIN-32.[Drosophila melanog...    39   0.003
sp|P25070|TCH2_ARATH (TCH2)CALMODULIN-RELATED PROTEIN 2, TOUCH-...    40   0.002
sp|P35491|PRVM_SCYCA PARVALBUMIN (FRAGMENT).[Scyliorhinus canic...    68   7e-12
sp|P02630|PRVA_RAJCL PARVALBUMIN ALPHA.[Raja clavata]                 71   6e-13
sp|P02615|PRVB_BOACO PARVALBUMIN BETA.[Boa constrictor]               75   4e-14
sp|P02622|PRVB_GADCA PARVALBUMIN BETA (ALLERGEN GAD C 1) (GAD C...    75   3e-14
sp|P02621|PRVB_MERMR PARVALBUMIN BETA.[Merlangius merlangus]          80   1e-15
sp|P02620|PRVB_MERME PARVALBUMIN BETA.[Merluccius merluccius]         83   1e-16
sp|Q91482|PRV1_SALSA PARVALBUMIN BETA 1 (CLONE 14.1) (MAJOR ALL...    86   2e-17
sp|P02623|PRVB_LATCH PARVALBUMIN BETA.[Latimeria chalumnae]           87   1e-17
sp|P02631|ONCO_RAT (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA).[Ra...    88   4e-18
sp|P05939|PRVB_LEUCE PARVALBUMIN BETA (V).[Leuciscus cephalus]        88   4e-18
sp|P56503|PRVB_MERBI PARVALBUMIN BETA.[Merluccius bilinearis]         88   4e-18
sp|P51879|ONCO_MOUSE (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA).[...    89   3e-18
sp|P05940|PRVB_XENLA PARVALBUMIN BETA.[Xenopus laevis]                89   3e-18
sp|O35508|ONCO_CAVPO (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA) (...    89   3e-18
sp|P32930|ONCO_HUMAN (OCM..)ONCOMODULIN (OM) (PARVALBUMIN BETA)...    89   2e-18
sp|P09227|PRVA_CYPCA PARVALBUMIN ALPHA (AI).[Cyprinus carpio]         90   1e-18
sp|P05941|PRVB_OPSTA PARVALBUMIN BETA (PARVALBUMIN IIIB).[Opsan...    93   1e-19
sp|Q91483|PRV2_SALSA PARVALBUMIN BETA 2 (CLONE 24.1) (MAJOR ALL...    94   8e-20
sp|P02617|PRVB_RANES PARVALBUMIN BETA.[Rana esculenta]                94   8e-20
sp|P02616|PRVB_AMPME PARVALBUMIN BETA.[Amphiuma means]                94   6e-20
sp|P43305|PRVU_CHICK PARVALBUMIN, THYMIC CPV3 (PARVALBUMIN 3).[...    95   3e-20
sp|P02619|PRVB_ESOLU PARVALBUMIN BETA.[Esox lucius]                   97   1e-20
sp|P02618|PRVB_CYPCA PARVALBUMIN BETA.[Cyprinus carpio]               99   2e-21
```

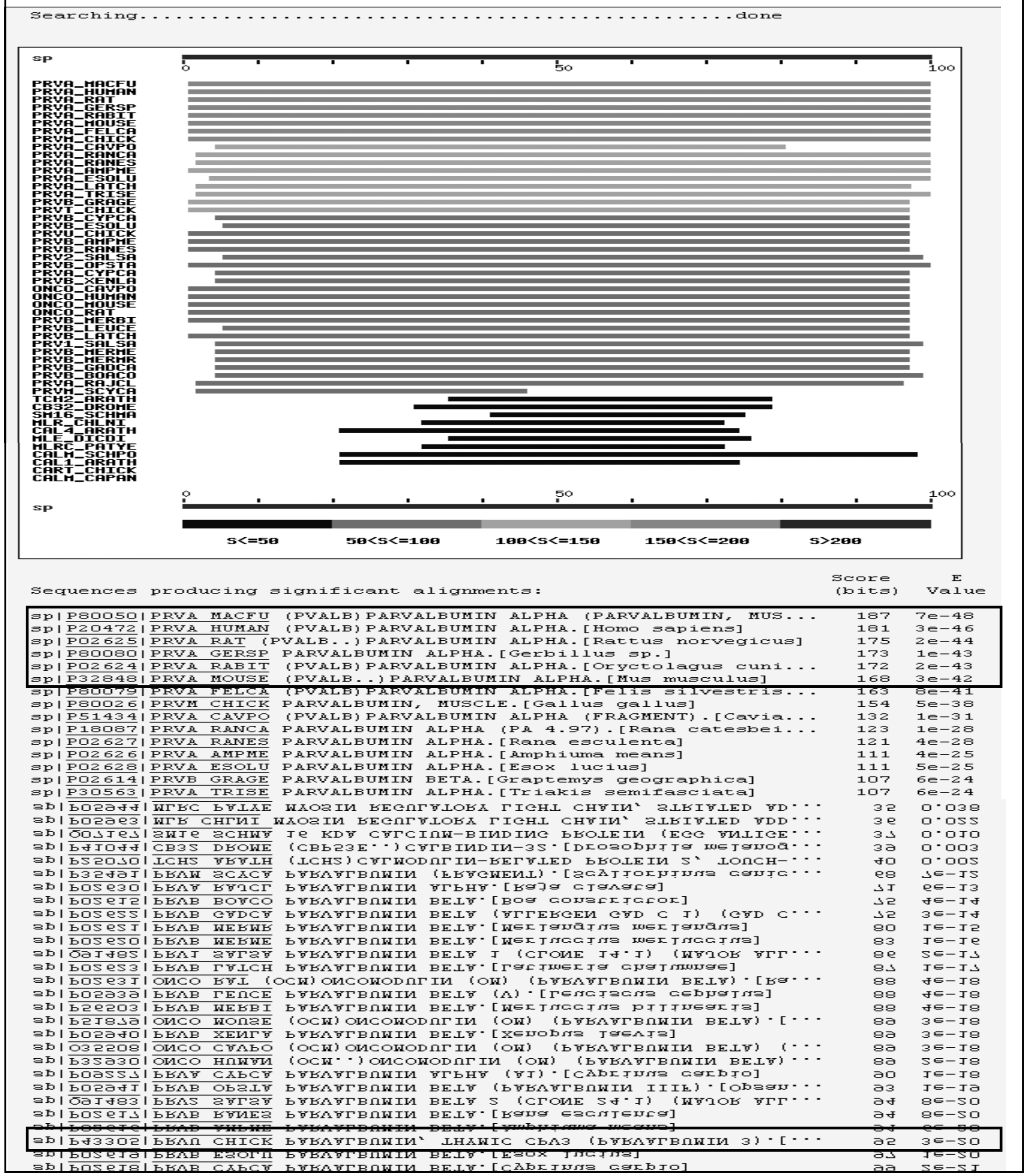
# Common Mistake: Sequences Too Closely Related

```
PRVA_MACFU   SMTDLLNAEDIKKAVGAFSAIDSFDHKKFFQMVGLKKKSADDVKKVFHILDKDKSGFIEE
PRVA_HUMAN   SMTDLLNAEDIKKAVGAFSATDSFDHKKFFQMVGLKKKSADDVKKVFHMLDKDKSGFIEE
PRVA_GERSP   SMTDLLSAEDIKKAIGAFAAADSFDHKKFFQMVGLKKKTPDDVKKVFHILDKDKSGFIEE
PRVA_MOUSE   SMTDVLSAEDIKKAIGAFAAADSFDHKKFFQMVGLKKKNPDEVKKVFHILDKDKSGFIEE
PRVA_RAT     SMTDLLSAEDIKKAIGAFTAADSFDHKKFFQMVGLKKKSADDVKKVFHILDKDKSGFIEE
PRVA_RABIT   AMTELLNAEDIKKAIGAFAAAESFDHKKFFQMVGLKKKSTEDVKKVFHILDKDKSGFIEE
             :**::*.*******:***:*  :*****************..::******:***********

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

-IDENTICAL SEQUENCES BRING NO INFORMATION FOR THE MULTIPLE SEQUENCE ALIGNMENT

-MULTIPLE SEQUENCE ALIGNMENTS THRIVE ON DIVERSITY…

# Selecting Diverse Sequences (Opus I)



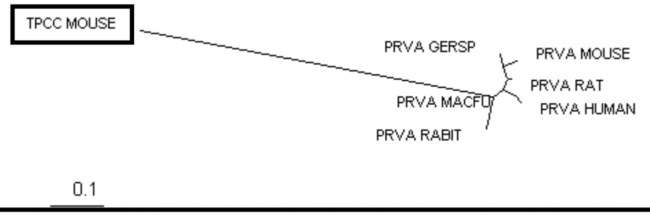
```
Searching..................................................done

                                                                 Score     E
Sequences producing significant alignments:                      (bits)  Value

sp|P80050|PRVA MACFU  (PVALB)PARVALBUMIN ALPHA (PARVALBUMIN, MUS...   187   7e-48
sp|P20472|PRVA HUMAN  (PVALB)PARVALBUMIN ALPHA.[Homo sapiens]         181   3e-46
sp|P02625|PRVA RAT (PVALB..)PARVALBUMIN ALPHA.[Rattus norvegicus]     175   2e-44
sp|P80080|PRVA GERSP  PARVALBUMIN ALPHA.[Gerbillus sp.]               173   1e-43
sp|P02624|PRVA RABIT  (PVALB)PARVALBUMIN ALPHA.[Oryctolagus cuni...   172   2e-43
sp|P32848|PRVA MOUSE  (PVALB..)PARVALBUMIN ALPHA.[Mus musculus]       168   3e-42
sp|P80079|PRVA FELCA  (PVALB)PARVALBUMIN ALPHA.[Felis silvestris...   163   8e-41
sp|P80026|PRVM CHICK  PARVALBUMIN, MUSCLE.[Gallus gallus]             154   5e-38
sp|P51434|PRVA CAVPO  (PVALB)PARVALBUMIN ALPHA (FRAGMENT).[Cavia...   132   1e-31
sp|P18087|PRVA RANCA  PARVALBUMIN ALPHA (PA 4.97).[Rana catesbei...   123   1e-28
sp|P02627|PRVA RANES  PARVALBUMIN ALPHA.[Rana esculenta]              121   4e-28
sp|P02626|PRVA AMPME  PARVALBUMIN ALPHA.[Amphiuma means]              111   4e-25
sp|P02628|PRVA ESOLU  PARVALBUMIN ALPHA.[Esox lucius]                 111   5e-25
sp|P02614|PRVB GRAGE  PARVALBUMIN BETA.[Graptemys geographica]        107   6e-24
sp|P30563|PRVA TRISE  PARVALBUMIN ALPHA.[Triakis semifasciata]        107   6e-24
sp|P02618|PRVB CYPCA  PARVALBUMIN BETA.[Cyprinus carpio]               99   2e-21
sp|P02619|PRVB ESOLU  PARVALBUMIN BETA.[Esox lucius]                   97   1e-20
sp|P43305|PRVU CHICK  PARVALBUMIN, THYMIC CPV3 (PARVALBUMIN 3).[...    95   3e-20
sp|P02616|PRVB AMPME  PARVALBUMIN BETA.[Amphiuma means]                94   8e-20
sp|P02617|PRVB RANES  PARVALBUMIN BETA.[Rana esculenta]                94   8e-20
sp|Q91483|PRV2 SALSA  PARVALBUMIN BETA 2 (CLONE 24.1) (MAJOR ALL...    94   8e-20
sp|P02641|PRVB OPSTA  PARVALBUMIN BETA (PARVALBUMIN IIIB).[Opsan...    93   1e-19
sp|P09227|PRVA CYPCA  PARVALBUMIN ALPHA (AI).[Cyprinus carpio]         90   1e-18
sp|P32930|ONCO HUMAN  (OCM..)ONCOMODULIN (OM) (PARVALBUMIN BETA)...    89   2e-18
sp|O35508|ONCO CAVPO  (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA) (...    89   3e-18
sp|P02640|PRVB XENLA  PARVALBUMIN BETA.[Xenopus laevis]                89   3e-18
sp|P51879|ONCO MOUSE  (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA).[...    89   3e-18
sp|P56503|PRVB MERBI  PARVALBUMIN BETA.[Merluccius bilinearis]         88   4e-18
sp|P02639|PRVB LEUCE  PARVALBUMIN BETA (A).[Leuciscus cephalus]        88   4e-18
sp|P02631|ONCO RAT  (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA).[Ra...    88   4e-18
sp|P02623|PRVB LATCH  PARVALBUMIN BETA.[Latimeria chalumnae]           87   1e-17
sp|Q91482|PRV1 SALSA  PARVALBUMIN BETA 1 (CLONE 14.1) (MAJOR ALL...    86   2e-17
sp|P02620|PRVB MERME  PARVALBUMIN BETA.[Merluccius merluccius]         83   1e-16
sp|P02621|PRVB MERMR  PARVALBUMIN BETA.[Merlangius merlangus]          80   1e-15
sp|P02622|PRVB GADCA  PARVALBUMIN BETA (ALLERGEN GAD C 1) (GAD C...    75   3e-14
sp|P02615|PRVB BOACO  PARVALBUMIN BETA.[Boa constrictor]               75   4e-14
sp|P02630|PRVA RAJCL  PARVALBUMIN ALPHA.[Raja clavata]                 71   6e-13
sp|P32491|PRVM SCYCA  PARVALBUMIN (FRAGMENT).[Scyliorhinus canic...    68   7e-12
sp|P25070|TCH2 ARATH  (TCH2)CALMODULIN-RELATED PROTEIN 2, TOUCH-...    40   0.002
sp|P41044|CB32 DROME  (CBP53E..)CALBINDIN-32.[Drosophila melanog...    39   0.003
sp|Q07167|SM16 SCHMA  16 KDA CALCIUM-BINDING PROTEIN (EGG ANTIGE...    37   0.010
sp|P05963|MLR CHLNI  MYOSIN REGULATORY LIGHT CHAIN, STRIATED ADD...    36   0.022
sp|P05944|MLRC PATYE  MYOSIN REGULATORY LIGHT CHAIN, STRIATED AD...    35   0.038
```

# Respect Information!

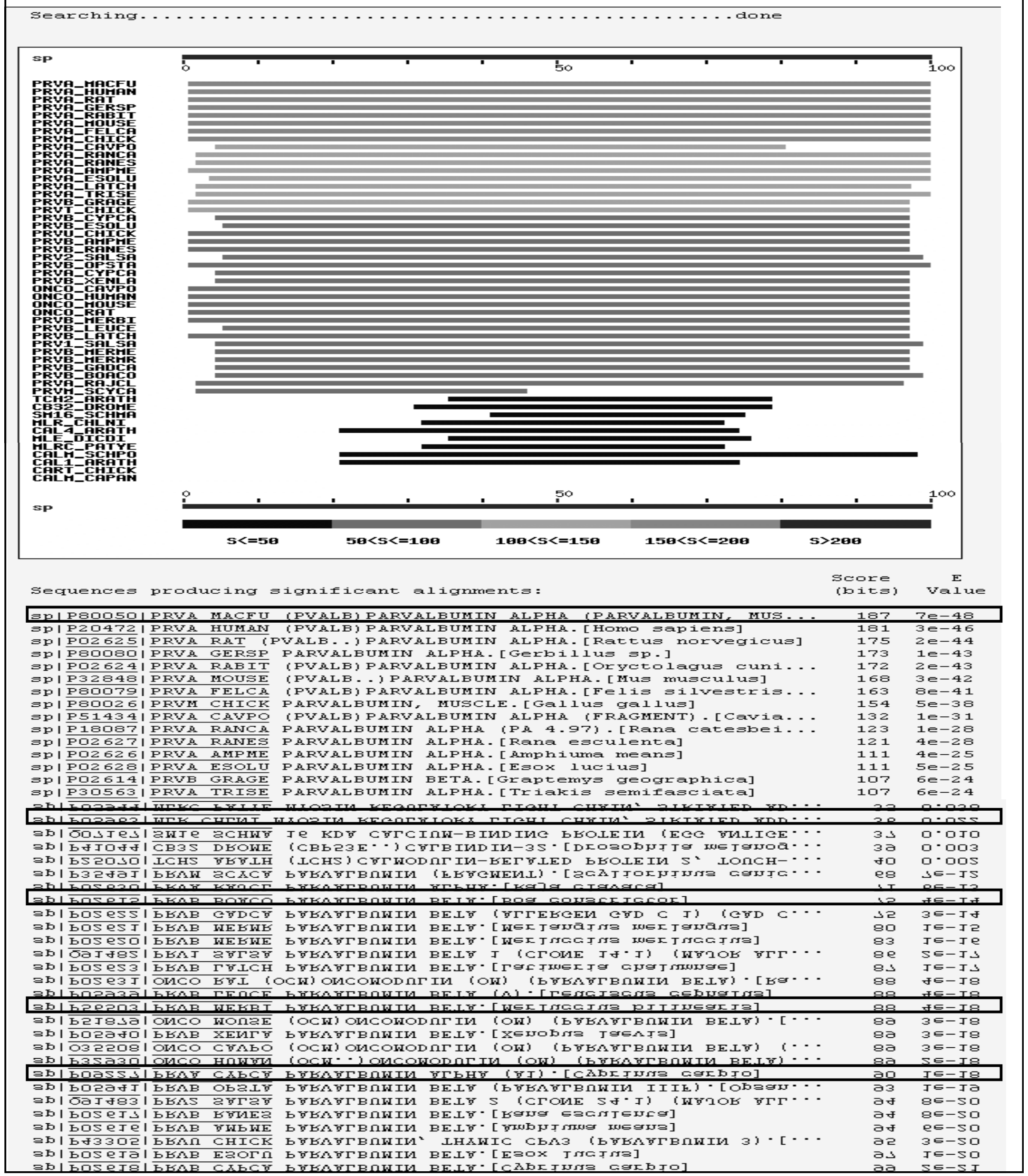
```
PRVA_MACFU   ----------------------------------------SMTDLLN----AEDIKKA
PRVA_HUMAN   ----------------------------------------SMTDLLN----AEDIKKA
PRVA_GERSP   ----------------------------------------SMTDLLS----AEDIKKA
PRVA_MOUSE   ----------------------------------------SMTDVLS----AEDIKKA
PRVA_RAT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

PRVA_MACFU   VGAFSAIDS--FDHKKFFQMVG------LKKKSADDVKKVFHILDKDKSGFIEEDELGFI
PRVA_HUMAN   VGAFSATDS--FDHKKFFQMVG------LKKKSADDVKKVFHMLDKDKSGFIEEDELGFI
PRVA_GERSP   IGAFAAADS--FDHKKFFQMVG------LKKKTPDDVKKVFHILDKDKSGFIEEDELGFI
PRVA_MOUSE   IGAFAAADS--FDHKKFFQMVG------LKKKNPDEVKKVFHILDKDKSGFIEEDELGSI
PRVA_RAT     IGAFTAADS--FDHKKFFQMVG------LKKKSADDVKKVFHILDKDKSGFIEEDELGSI
PRVA_RABIT   IGAFAAAES--FDHKKFFQMVG------LKKKSTEDVKKVFHILDKDKSGFIEEEELGFI
TPCC_MOUSE   IDEVDEDGSGTVDFDEFLVMMVRCMKDDSKGKSEEELSDLFRMFDKNADGYIDLDELKMM
             :.  .     *  .*..:*: *:          * *. ::::..:*:::**:  .*:*: :**  :

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

-This Alignment Is not Informative about the relation Betwwen TPCC MOUSE and the rest of the sequences.

-A better Spread of the Sequences is needed



# Selecting Diverse Sequences (Opus II)



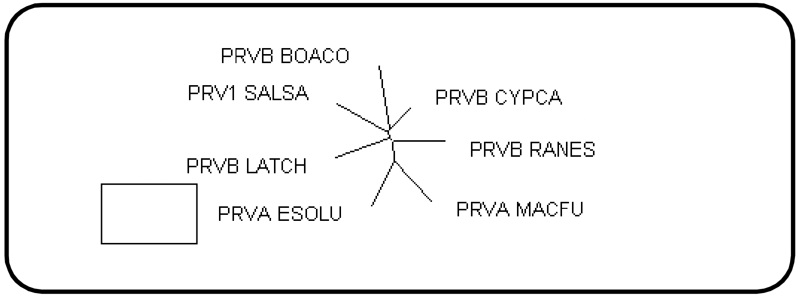
```
Searching.................................................done

                                                                  Score     E
Sequences producing significant alignments:                       (bits)  Value

sp|P80050|PRVA MACFU  (PVALB)PARVALBUMIN ALPHA (PARVALBUMIN, MUS...   187   7e-48
sp|P20472|PRVA HUMAN  (PVALB)PARVALBUMIN ALPHA.[Homo sapiens]          181   3e-46
sp|P02625|PRVA RAT  (PVALB..)PARVALBUMIN ALPHA.[Rattus norvegicus]     175   2e-44
sp|P80080|PRVA GERSP  PARVALBUMIN ALPHA.[Gerbillus sp.]                173   1e-43
sp|P02624|PRVA RABIT  (PVALB)PARVALBUMIN ALPHA.[Oryctolagus cuni...    172   2e-43
sp|P32848|PRVA MOUSE  (PVALB..)PARVALBUMIN ALPHA.[Mus musculus]        168   3e-42
sp|P80079|PRVA FELCA  (PVALB)PARVALBUMIN ALPHA.[Felis silvestris...    163   8e-41
sp|P80026|PRVM CHICK  PARVALBUMIN, MUSCLE.[Gallus gallus]              154   5e-38
sp|P51434|PRVA CAVPO  (PVALB)PARVALBUMIN ALPHA (FRAGMENT).[Cavia...    132   1e-31
sp|P18087|PRVA RANCA  PARVALBUMIN ALPHA (PA 4.97).[Rana catesbei...    123   1e-28
sp|P02627|PRVA RANES  PARVALBUMIN ALPHA.[Rana esculenta]               121   4e-28
sp|P02626|PRVA AMPME  PARVALBUMIN ALPHA.[Amphiuma means]               111   4e-25
sp|P02628|PRVA ESOLU  PARVALBUMIN ALPHA.[Esox lucius]                  111   5e-25
sp|P02614|PRVB GRAGE  PARVALBUMIN BETA.[Graptemys geographica]         107   6e-24
sp|P30563|PRVA TRISE  PARVALBUMIN ALPHA.[Triakis semifasciata]         107   6e-24
sp|P05944|MLRC PATYE  MYOSIN REGULATORY LIGHT CHAIN, STRIATED AD...     35   0.039
sp|P05963|MLR CHLNI  MYOSIN REGULATORY LIGHT CHAIN, STRIATED ADD...     36   0.022
sp|Q07167|SM16 SCHMA  16 KDA CALCIUM-BINDING PROTEIN (EGG ANTIGE...     37   0.010
sp|P41044|CB32 DROME  (CBP53E..)CALBINDIN-32.[Drosophila melanog...     39   0.003
sp|P25070|TCH2 ARATH  (TCH2)CALMODULIN-RELATED PROTEIN 2, TOUCH-...     40   0.002
sp|P35491|PRVM SCYCA  PARVALBUMIN (FRAGMENT).[Scyliorhinus canic...     68   7e-12
sp|P02630|PRVA RAJCL  PARVALBUMIN ALPHA.[Raja clavata]                  71   9e-13
sp|P02615|PRVB BOACO  PARVALBUMIN BETA.[Boa constrictor]                75   4e-14
sp|P02622|PRVB GADCA  PARVALBUMIN BETA (ALLERGEN GAD C 1) (GAD C...     75   3e-14
sp|P02621|PRVB MERMR  PARVALBUMIN BETA.[Merlangius merlangus]           80   1e-15
sp|P02620|PRVB MERME  PARVALBUMIN BETA.[Merluccius merluccius]          83   1e-16
sp|Q91482|PRV1 SALSA  PARVALBUMIN BETA 1 (CLONE 14.1) (MAJOR ALL...     86   2e-17
sp|P02623|PRVB LATCH  PARVALBUMIN BETA.[Latimeria chalumnae]            87   1e-17
sp|P02631|ONCO RAT  (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA).[Ra...     88   4e-18
sp|P02939|PRVB LEUCE  PARVALBUMIN BETA (A).[Leuciscus cephalus]         88   4e-18
sp|P56503|PRVB MERBI  PARVALBUMIN BETA.[Merluccius bilinearis]          88   4e-18
sp|P51879|ONCO MOUSE  (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA).[...     89   3e-18
sp|P02640|PRVB XENLA  PARVALBUMIN BETA.[Xenopus laevis]                 89   3e-18
sp|O35208|ONCO CAVPO  (OCM)ONCOMODULIN (OM) (PARVALBUMIN BETA) (...     89   3e-18
sp|P32930|ONCO HUMAN  (OCM..)ONCOMODULIN (OM) (PARVALBUMIN BETA)...     89   2e-18
sp|P09227|PRVA CYPCA  PARVALBUMIN ALPHA (AI).[Cyprinus carpio]          90   1e-18
sp|P02641|PRVB OPSTA  PARVALBUMIN BETA (PARVALBUMIN IIIB).[Opsan...     93   1e-19
sp|Q91483|PRV2 SALSA  PARVALBUMIN BETA 2 (CLONE 24.1) (MAJOR ALL...     94   8e-20
sp|P02617|PRVB RANES  PARVALBUMIN BETA.[Rana esculenta]                 94   8e-20
sp|P02616|PRVB AMPME  PARVALBUMIN BETA.[Amphiuma means]                 94   6e-20
sp|P43305|PRVU CHICK  PARVALBUMIN, THYMIC CPV3 (PARVALBUMIN 3).[...     95   3e-20
sp|P02619|PRVB ESOLU  PARVALBUMIN BETA.[Esox lucius]                    97   1e-20
sp|P02618|PRVB CYPCA  PARVALBUMIN BETA.[Cyprinus carpio]                99   2e-21
```

# Selecting Diverse Sequences (Opus II)

```
PRVB_CYPCA    -AFAGVLNDADIAAALEACKAADSFNHKAFFAKVGLTSKSADDVKKAFAIIDQDKSGFIE
PRVB_BOACO    -AFAGILSDADIAAGLQSCQAADSFSCKTFFAKSGLHSKSKDQLTKVFGVIDRDKSGYIE
PRV1_SALSA    MACAHLCKEADIKTALEACKAADTFSFKTFFHTIGFASKSADDVKKAFKVIDQDASGFIE
PRVB_LATCH    -AVAKLLAAADVTAALEGCKADDSFNHKVFFQKTGLAKKSNEELEAIFKILDQDKSGFIE
PRVB_RANES    -SITDIVSEKDIDAALESVKAAGSFNYKIFFQKVGLAGKSAADAKKVFEILDRDKSGFIE
PRVA_MACFU    -SMTDLLNAEDIKKAVGAFSAIDSFDHKKFFQMVGLKKKSADDVKKVFHILDKDKSGFIE
PRVA_ESOLU    --AKDLLKADDIKKALDAVKAEGSFNHKKFFALVGLKAMSANDVKKVFKAIDADASGFIE
                     :    *:  .:  . *  :*  * **    *:   *  :    *  :* * **:**

PRVB_CYPCA    EDELKLFLQNFKADARALTDGETKTFLKAGDSDGDGKIGVDEFTALVKA-
PRVB_BOACO    EDELKKFLQNFDGKARDLTDKETAEFLKEGDTDGDGKIGVEEFVVLVTKG
PRV1_SALSA    VEELKLFLQNFCPKARELTDAETKAFLKAGDADGDGMIGIDEFAVLVKQ-
PRVB_LATCH    DEELELFLQNFSAGARTLTKTETETFLKAGDSDGDGKIGVDEFQKLVKA-
PRVB_RANES    QDELGLFLQNFRASARVLSDAETSAFLKAGDSDGDGKIGVEEFQALVKA-
PRVA_MACFU    EDELGFILKGFSPDARDLSAKETKTLMAAGDKDGDGKIGVDEFSTLVAES
PRVA_ESOLU    EEELKFVLKSFAADGRDLTDAETKAFLKAADKDGDGKIGIDEFETLVHEA
               :**  .*:.*    .* *:  **   ::  .* **** **::**  **
```

-A REASONABLE Model Now Exists.

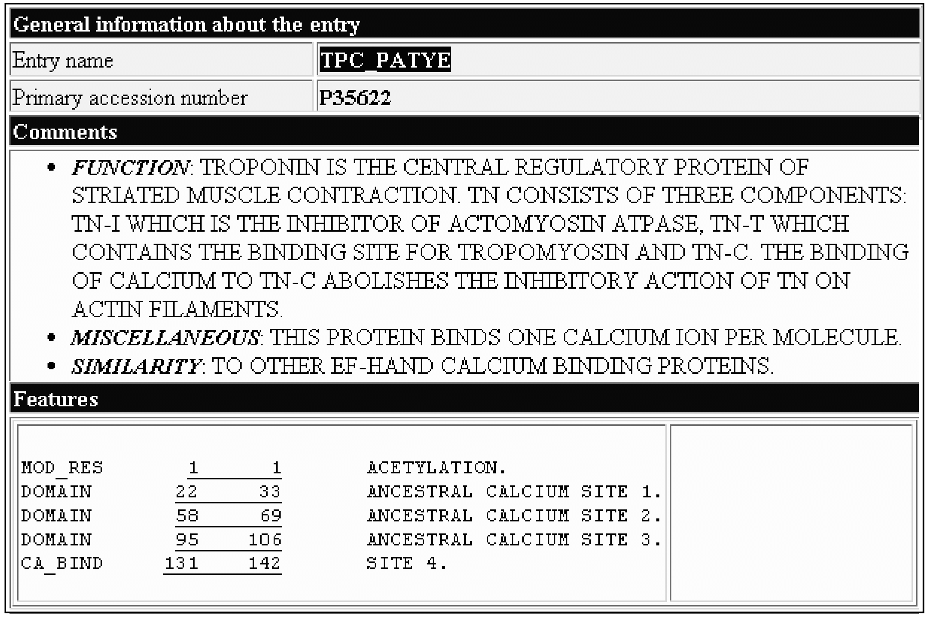
-Going Further:Remote Homologues.



# Aligning Remote Homologues

```
PRVA_MACFU    ------------------------------------------SMTDLLNA----EDIKKA
PRVA_ESOLU    -------------------------------------------AKDLLKA----DDIKKA
PRVB_CYPCA    ------------------------------------------AFAGVLND----ADIAAA
PRVB_BOACO    ------------------------------------------AFAGILSD----ADIAAG
PRV1_SALSA    -----------------------------------------MACAHLCKE----ADIKTA
PRVB_LATCH    ------------------------------------------AVAKLLAA----ADVTAA
PRVB_RANES    ------------------------------------------SITDIVSE----KDIDAA
TPCS_RABIT    -TDQQAEARSYLSEEMIAEFKAAFDMFDADGG-GDISVKELGTVMRMLGQTPTKEELDAI
TPCS_PIG      -TDQQAEARSYLSEEMIAEFKAAFDMFDADGG-GDISVKELGTVMRMLGQTPTKEELDAI
TPCC_MOUSE    MDDIYKAAVEQLTEEQKNEFKAAFDIFVLGAEDGCISTKELGKVMRMLGQNPTPEELQEM
                                                               :        ::

PRVA_MACFU    VGAFSAIDS--FDHKKFFQMVG------LKKKSADDVKKVFHILDKDKSGFIEEDELGFI
PRVA_ESOLU    LDAVKAEGS--FNHKKFFALVG------LKAMSANDVKKVFKAIDADASGFIEEEELKFV
PRVB_CYPCA    LEACKAADS--FNHKAFFAKVG------LTSKSADDVKKAFAIIDQDKSGFIEEDELKLF
PRVB_BOACO    LQSCQAADS--FSCKTFFAKSG------LHSKSKDQLTKVFGVIDRDKSGYIEEDELKKF
PRV1_SALSA    LEACKAADT--FSFKTFFHTIG------FASKSADDVKKAFKVIDQDASGFIEVEELKLF
PRVB_LATCH    LEGCKADDS--FNHKVFFQKTG------LAKKSNEELEAIFKILDQDKSGFIEDEELELF
PRVB_RANES    LESVKAAGS--FNYKIFFQKVG------LAGKSAADAKKVFEILDRDKSGFIEQDELGLF
TPCS_RABIT    IEEVDEDGSGTIDFEEFLVMMVRQMKEDAKGKSEEELAECFRIFDRNADGYIDAEELAEI
TPCS_PIG      IEEVDEDGSGTIDFEEFLVMMVRQMKEDAKGKSEEELAECFRIFDRNMDGYIDAEELAEI
TPCC_MOUSE    IDEVDEDGSGTVDFDEFLVMMVRCMKDDSKGKSEEELSDLFRMFDKNADGYIDLDELKMM
              :   .  .:  .. . *:              *  :    *  :* : .*:*: :**  .

PRVA_MACFU       LKGFSPDARDLSAKETKTLMAAGDKDGDGKIGVDEFSTLVAES-
PRVA_ESOLU       LKSFAADGRDLTDAETKAFLKAADKDGDGKIGIDEFETLVHEA-
PRVB_CYPCA       LQNFKADARALTDGETKTFLKAGDSDGDGKIGVDEFTALVKA--
PRVB_BOACO       LQNFDGKARDLTDKETAEFLKEGDTDGDGKIGVEEFVVLVTKG-
PRV1_SALSA       LQNFCPKARELTDAETKAFLKAGDADGDGMIGIDEFAVLVKQ--
PRVB_LATCH       LQNFSAGARTLTKTETETFLKAGDSDGDGKIGVDEFQKLVKA--
PRVB_RANES       LQNFRASARVLSDAETSAFLKAGDSDGDGKIGVEEFQALVKA--
TPCS_RABIT       FR---ASGEHVTDEEIESLMKDGDKNNDGRIDFDEFLKMMEGVQ
TPCS_PIG         FR---ASGEHVTDEEIESIMKDGDKNNDGRIDFDEFLKMMEGVQ
TPCC_MOUSE       LQ---ATGETITEDDIEELMKDGDKNNDGRIDYDEFLEFMKGVE
                 ::     .. ::  :   ::  .* :.** *. :**  ::
```

# Going Further…



**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 | 1 | 1 | ACETYLATION. |
| DOMAIN | 22 | 33 | ANCESTRAL CALCIUM SITE 1. |
| DOMAIN | 58 | 69 | ANCESTRAL CALCIUM SITE 2. |
| DOMAIN | 95 | 106 | ANCESTRAL CALCIUM SITE 3. |
| CA_BIND | 131 | 142 | SITE 4. |

```
PRVA_MACFU    VGAFSAIDS--FDHKKFFQMVG------LKKKSADDVKKVFHILDKDKSGFIEEDELGFI
PRVB_BOACO    LQSCQAADS--FSCKTFFAKSG------LHSKSKDQLTKVFGVIDRDKSGYIEEDELKKF
PRV1_SALSA    LEACKAADT--FSFKTFFHTIG------FASKSADDVKKAFKVIDQDASGFIEVEELKLF
TPCS_RABIT    IEEVDEDGSGTIDFEEFLVMMVRQMKEDAKGKSEEELAECFRIFDRNADGYIDAEELAEI
TPCS_PIG      IEEVDEDGSGTIDFEEFLVMMVRQMKEDAKGKSEEELAECFRIFDRNMDGYIDAEELAEI
TPCC_MOUSE    IDEVDEDGSGTVDFDEFLVMMVRCMKDDSKGKSEEELSDLFRMFDKNADGYIDLDELKMM
TPC_PATYE     SDEMDEEATGRLNCDAWIQLFER---KLKEDLDERELKEAFRVLDKEKKGVIKVDVLRWI
                  .   :  .. . ::              .  :    *  :* : .* *. : *  .

PRVA_MACFU    LKGFSPDARDLSAKETKTLMAAGDKDGDGKIGVDEFSTLVAES--
PRVB_BOACO    LQNFDGKARDLTDKETAEFLKEGDTDGDGKIGVEEFVVLVTKG--
PRV1_SALSA    LQNFCPKARELTDAETKAFLKAGDADGDGMIGIDEFAVLVKQ---
TPCS_RABIT    FR---ASGEHVTDEEIESLMKDGDKNNDGRIDFDEFLKMMEGVQ-
TPCS_PIG      FR---ASGEHVTDEEIESIMKDGDKNNDGRIDFDEFLKMMEGVQ-
TPCC_MOUSE    LQ---ATGETITEDDIEELMKDGDKNNDGRIDYDEFLEFMKGVE-
TPC_PATYE     LS---SLGDELTEEEIENMIAETDTDGSGTVDYEEFKCLMMSSDA
              :      .  ::  :   ::   * :..* :. :**  ::
```

# WHAT MAKES A GOOD ALIGNMENT…

-THE MORE DIVERGEANT THE SEQUENCES, THE BETTER

-THE FEWER INDELS, THE BETTER

-NICE UNGAPPED BLOCKS SEPARATED WITH INDELS

-DIFFERENT CLASSES OF RESIDUES WITHIN A BLOCK:

- Completely Conserved
- Conserved For Size *and* Hydropathy
- Conserved For Size *or* Hydropathy

-THE ULTIMATE EVALUATION IS A MATTER OF PERSONNAL JUDGEMENT AND KNOWLEDGE.

# DO NOT OVERTUNE!!!

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

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

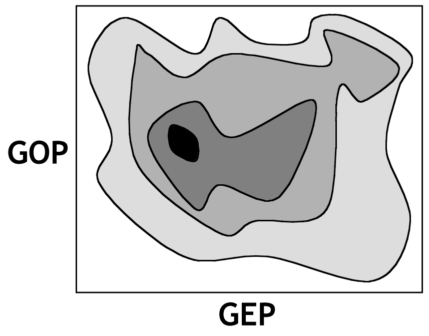
DO NOT PLAY WITH PARAMETERS IF YOU KNOW THE ALIGNMENT YOU WANT: MAKE IT YOURSELF!

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

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

# TUNING or NOT TUNING?

-PARAMETERS TO TUNE USUALLY INCLUDE:

- GOP/ GEP
- MATRIX
- SENSITIVITY Vs SPEED



Substitution Matrices
(Etzold and al. 1993)

| | |
|---|---|
| Gonnet | 61.7 |
| Blosum50 | 59.7 |
| Pam250 | 59.2 |

-MOST METHODS ARE TUNED FOR WORKING WELL ON AVERAGE

-PARAMETERS BEHAVIOUR DO NOT NECESSARILY FOLLOW THE THEORY (i.e. Substitution Matrices).

-A GOOD ALIGNMENT IS USUALLY ROBUST(i.e. Changes little).

-TUNE IF YOU WANT TO CONVINCE YOURSELF.

# KEEP A BIOLOGICAL PERSPECTIVE

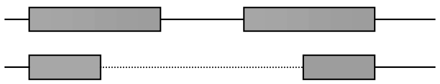
```
chite  ---ADKPKRPLSAYMLWLNSARESIKRENPDFK-VTEVAKKGGELWRGLKD
wheat  --DPNKPKRAPSAFFVFMGEFREEFKQKNPKNKSVAAVGKAAGERWKSLSE
trybr  KKDSNAPKRAMTSFMFFSSDFRS----KHSDLS-IVEMSKAAGAAWKELGP
mouse  -----KPKRPRSAYNIYVSESFQ----EAKDDS-AQGKLKLVNEAWKNLSP
             ***. ::: .: ..  .    :  . .      *  .  *: *

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

DIFFERENT PARAMETERS

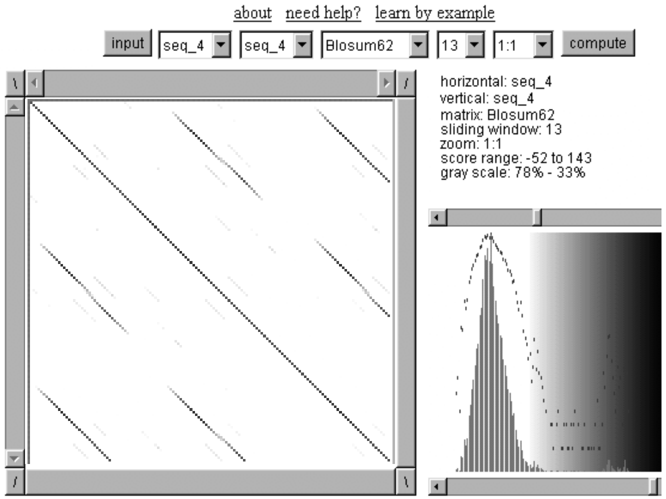
```
chite AD--K----PKR-PLYMLWLNS-ARESIKRENPDFK-VT-EVAKKGGELWRGL-
wheat -DPNK----PKRAP-FFVFMGE-FREEFKQKNPKNKSVA-AVGKAAGERWKSLS
trybr -K--KDSNAPKR-AMT-MFFSSDFR-S-KH-S-DLS-IV-EMSKAAGAAWKELG
mouse ----K----PKR-PRYNIYVSESFQEA-K--D-D-S-AQGKL-KLVNEAWKNLS
          *    *** .:: ::...  :   *  . . .     : *  .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

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

## REPEATS

THERE IS A PROBLEM WHEN TWO SEQUENCES DO NOT CONTAIN THE SAME NUMBER OF REPEATS



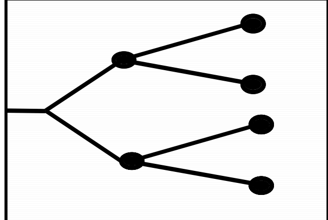
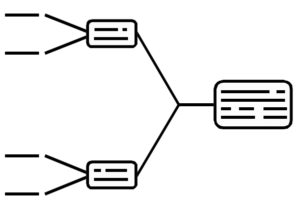
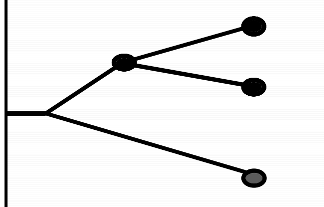
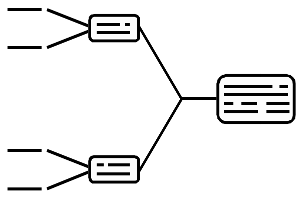
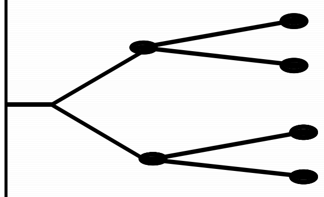
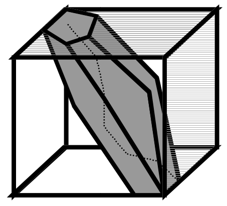
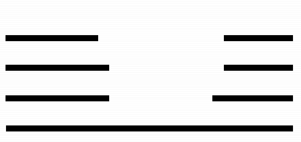
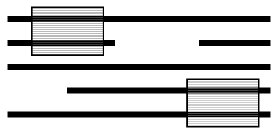
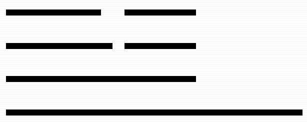
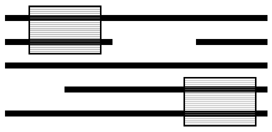
IT IS THEN BETTER TO MANUALLY EXTRACT THE REPEATS AND TO ALIGN THEM. INDIVIDUAL REPEATS CAN BE RECOGNIZED USING DOTTER



horizontal: seq_4
vertical: seq_4
matrix: Blosum62
sliding window: 13
zoom: 1:1
score range: -52 to 143
gray scale: 78% - 33%

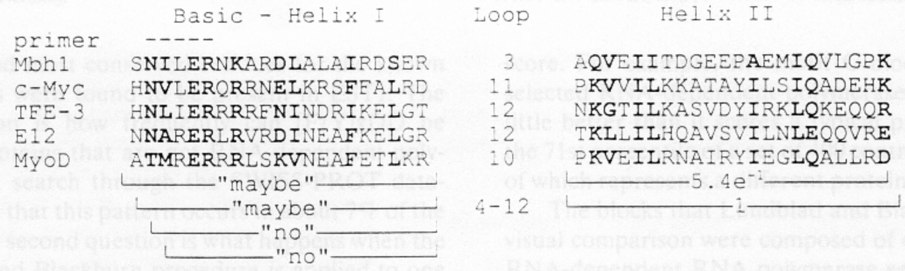
## Choosing The Right Method

Source: BaliBase, Thompson et al, NAR, 1999

| PROBLEM | PROGRAM | METHOD |
|---|---|---|
|  | ClustalW |  |
|  | ClustalW |  |
|  | MSA |  |
|  | DIALIGN II |  |
|  | DIALIGN II |  |

# Examples of Mistakes

# Playing With Blocks: Mbh1

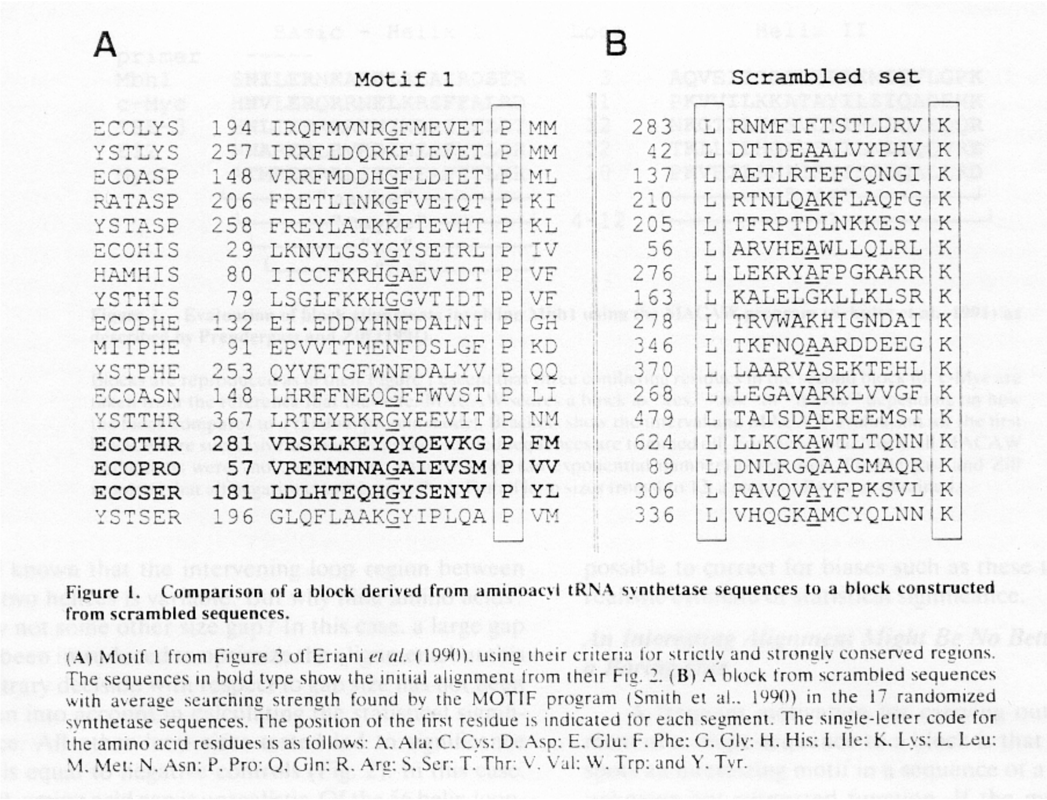


```
                 Basic - Helix I       Loop           Helix II
primer     -----
Mbh1       SNILERNKARDLALAIRDSER        3    AQVEIITDGEEPAEMIQVLGPK
c-Myc      HNVLERQRRNELKRSFFALRD       11    PKVVILKKATAYILSIQADEHK
TFE-3      HNLIERRRRFNINDRIKELGT       12    NKGTILKASVDYIRKLQKEQQR
E12        NNARERLRVRDINEAFKELGR       12    TKLLILHQAVSVILNLEQQVRE
MyoD       ATMRERRRLSKVNEAFETLKR       10    PKVEILRNAIRYIEGLQALLRD
          └──────"maybe"──────┘             └──────5.4e-1──────┘
            └─────"maybe"─────┘       4-12   └────────1─────────┘
              └─────"no"──────┘
                └────"no"─────┘
```

**Figure 2. Evaluation of block alignments involving Mbh1 using the MACAW program (Schuler et al., 1991) as described by Prendergast and Ziff (1991).**

Blocks are reproduced as in their Figure 1 except that three conflicting residues in the second block for c-Myc are taken from the reference that they cite. MACAW scores a block as "yes," "no," or "maybe" depending on how the block compares to a random protein model. Brackets show the interval and MACAW evaluation for the first block, where successive columns of primer-related sequences are trimmed off. For the second block, all MACAW evaluations were "no", so that the actual P-score (an exponential number) is reported. Prendergast and Ziff reported that all negative controls gave P = 1. For all loop sizes from 4 to 12, a score of P = 1 was obtained.

## Playing With Blocks: tRNA Synthases

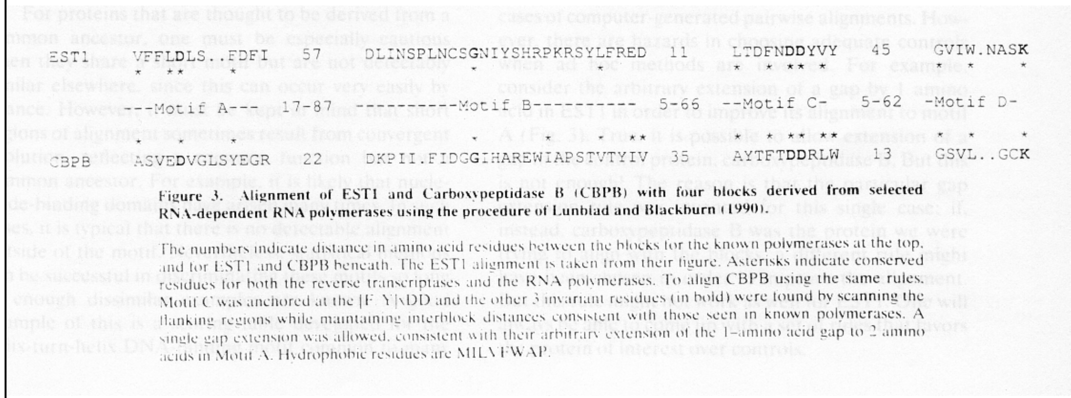


```
A                                           B
                  Motif 1                         Scrambled set

ECOLYS  194 IRQFMVNRGFMEVET P MM     283 L RNMFIFTSTLDRV K
YSTLYS  257 IRRFLDQRKFIEVET P MM      42 L DTIDEAALVYPHV K
ECOASP  148 VRRFMDDHGFLDIET P ML     137 L AETLRTEFCQNGI K
RATASP  206 FRETLINKGFVEIQT P KI     210 L RTNLQAKFLAQFG K
YSTASP  258 FREYLATKKFTEVHT P KL     205 L TFRPTDLNKKESY K
ECOHIS   29 LKNVLGSYGYSEIRL P IV      56 L ARVHEAWLLQLRL K
HAMHIS   80 IICCFKRHGAEVIDT P VF     276 L LEKRYAFPGKAKR K
YSTHIS   79 LSGLFKKHGGVTIDT P VF     163 L KALELGKLLKLSR K
ECOPHE  132 EI.EDDYHNFDALNI P GH     278 L TRVWAKHIGNDAI K
MITPHE   91 EPVVTTMENFDSLGF P KD     346 L TKFNQAARDDEEG K
YSTPHE  253 QYVETGFWNFDALYV P QQ     370 L LAARVASEKTEHL K
ECOASN  148 LHRFFNEQGFFWVST P LI     268 L LEGAVARKLVVGF K
YSTTHR  367 LRTEYRKRGYEEVIT P NM     479 L TAISDAEREEMST K
ECOTHR  281 VRSKLKEYQYQEVKG P FM     624 L LLKCKAWTLTQNN K
ECOPRO   57 VREEMNNAGAIEVSM P VV      89 L DNLRGQAAGMAQR K
ECOSER  181 LDLHTEQHGYSENYV P YL     306 L RAVQVAYFPKSVL K
YSTSER  196 GLQFLAAKGYIPLQA P VM     336 L VHQGKAMCYQLNN K
```

**Figure 1. Comparison of a block derived from aminoacyl tRNA synthetase sequences to a block constructed from scrambled sequences.**

(A) Motif 1 from Figure 3 of Eriani *et al.* (1990), using their criteria for strictly and strongly conserved regions. The sequences in bold type show the initial alignment from their Fig. 2. (B) A block from scrambled sequences with average searching strength found by the MOTIF program (Smith et al., 1990) in the 17 randomized synthetase sequences. The position of the first residue is indicated for each segment. The single-letter code for the amino acid residues is as follows: A. Ala: C. Cys: D. Asp: E. Glu: F. Phe: G. Gly: H. His: I. Ile: K. Lys: L. Leu: M. Met: N. Asn: P. Pro: Q. Gln: R. Arg: S. Ser: T. Thr: V. Val: W. Trp: and Y. Tyr.

## Playing With Blocks:RTase



```
EST1   VFFLDLS..FDFI   57   DLINSPLNCSGNIYSHRPKRSYLFRED  11   LTDFNDDYVY  45   GVIW.NASK
        *  **    *          *        *                        *   * ** *         * *     *

    ---Motif A---  17-87  ----------Motif B----------  5-66  --Motif C-   5-62  -Motif D-

        *   *    *          *        *     *  **              *   * ** **        * *     *
CBPB   ASVEDVGLSYEGR   22   DKPII.FIDGGIHAREWIAPSTVTYIV  35   AYTFTDDRLW  13   GSVL..GCK
```

**Figure 3. Alignment of EST1 and Carboxypeptidase B (CBPB) with four blocks derived from selected RNA-dependent RNA polymerases using the procedure of Lunblad and Blackburn (1990).**

The numbers indicate distance in amino acid residues between the blocks for the known polymerases at the top, and for EST1 and CBPB beneath. The EST1 alignment is taken from their figure. Asterisks indicate conserved residues for both the reverse transcriptases and the RNA polymerases. To align CBPB using the same rules. Motif C was anchored at the [F Y]xDD and the other 3 invariant residues (in bold) were found by scanning the flanking regions while maintaining interblock distances consistent with those seen in known polymerases. A single gap extension was allowed, consistent with their arbitrary extension of the 1-amino acid gap to 2 amino acids in Motif A. Hydrophobic residues are MILVFWAP.

# Conclusion

The Best Alignment Method:

- Your Brain
- The Right Data

The Best Evaluation:

- Your Eyes
- Experimental Information (SwissProt)

What Can I Conclude:

- Homology=> Information Extrapolation

How Can I go Further?:

- PrositePatterns.
- PrositeProfiles.